Methods of forecast verification

Kiyotoshi Takahashi

Climate Prediction Division
Japan Meteorological Agency

Outline

1. Purposes of verification
2. Verification methods
   - For deterministic forecasts
   - For probabilistic forecasts
   - Standardised Verification System for Long-Range Forecasts (SVSLRF)
3. Results on the TCC web page

1. Purposes of verification

Forecast verification is a process of assessing quality of forecasts.

- to monitor forecast quality
  - how accurate are forecasts and are they improving?
- to guide forecasters and users
  - help forecasters understand model characteristics
  - help us provide higher value forecasts to users
- to guide future developments of system
  - identify model faults and improve systems
  - help us compare and evaluate different forecasts (model or guidance)

2. Verification methods

- For deterministic forecast

  ACOR, RMSE, Bias (ME) ,MSSS (ROC)

- For probabilistic forecast

  ROC, Reliability diagram,

  BSS (Brel, Bres)

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(F_i - O_i\right)^2}$$

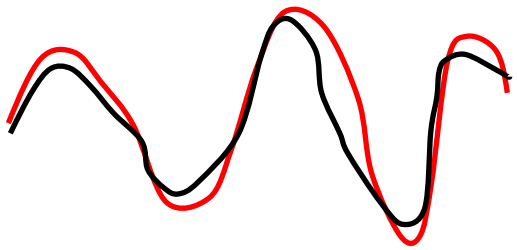
F : **forecast**

O : **observation**

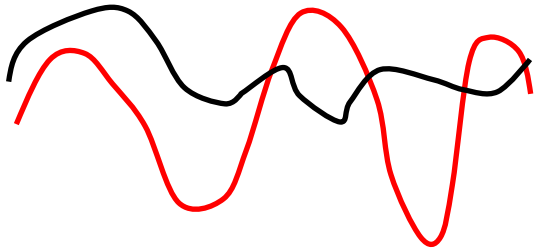
N : **sample size**

Range: 0 to infinity, Perfect score: 0.

- RMSE measures absolute magnitude of the forecast error.
- It does not indicate the direction the error.



RMSE small



RMSE large

Mean Error (ME)

$$ME = \frac{1}{N}\sum_{i=1}^{N}\left(F_i - O_i\right)$$

F : **forecast**

O : **observation**

N : **sample size**

Range: variable, Perfect score: 0.

- ME measures average magnitude of the forecast error.
- It indicates the direction the error.

$$RMSE^2 = ME^2 + \sigma_e^{\,2} \qquad \sigma_e^{\,2} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - a_i - ME\right)^2$$

RMSE can be divided into ME(systematic error) and random error (σ_e).

Anomaly Correlation (AC)

$$AC = \frac{\sum_{i=1}^{N}(F_i - C_i)(O_i - C_i)}{\sqrt{\sum_{i=1}^{N}(F_i - C_i)^2}\sqrt{\sum_{i=1}^{N}(O_i - C_i)^2}}$$

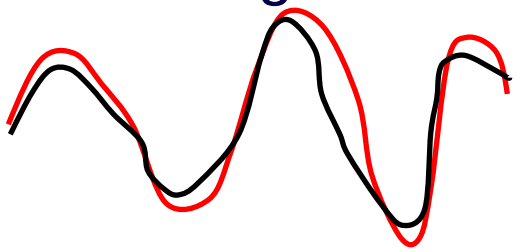
F : **forecast**

O : **observation**

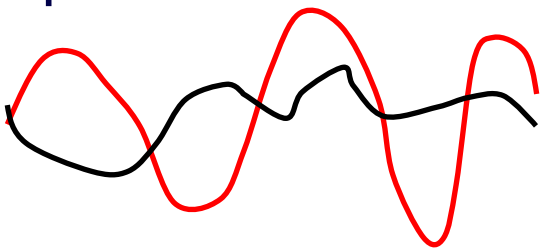
C : **climatology**

Range: -1 to 1. Perfect score: 1.

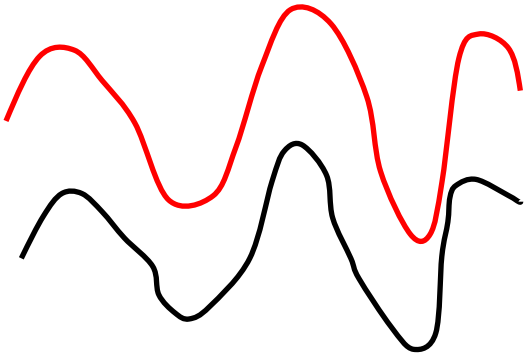
- AC measures correspondence or phase difference between forecast and observation, subtracting out the climatological mean at each point.



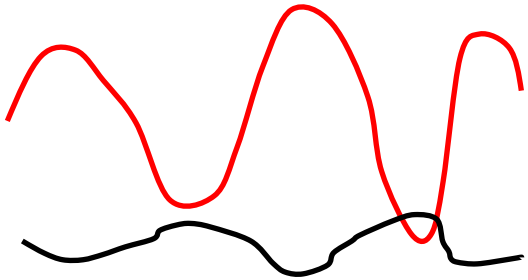
AC ≒ 1



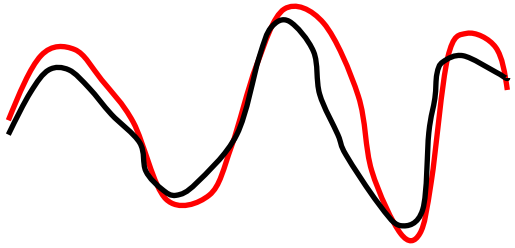
AC <<1



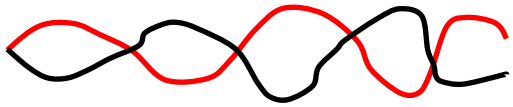
AC positive (≒1)
RMSE large



AC negative
RMSE large



AC positive(≒1)
RMSE small



AC negative
RMSE small

Mean Squared Skill Score (MSSS)

$$MSSS = 1 - \frac{MSE}{MSE_c}$$

Perfect score: 1 (when MSE=0)
Climatology forecast score: 0

where *MSE* is the mean squared error

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2$$

F :**forecast**
O :**observation**

and MSE_c is the MSE of climatology forecast.

MSSS can be expanded (Murphy,1988) as

$$MSSS = \left\{ 2\frac{s_f}{s_o} r_{fo} - \left(\frac{s_f}{s_o}\right)^2 - \left(\frac{\bar{f} - \bar{o}}{s_o}\right)^2 + \frac{2n-1}{(n-1)^2} \right\} \Bigg/ \left\{ 1 + \frac{2n-1}{(n-1)^2} \right\}$$

① ② ③

s : squared root of variance
r : correlation
f : forecast
o : observation

$$r_{fO} = \frac{\frac{1}{n}\sum_{i=1}^{n}(F_i - \bar{F})(O_i - \bar{O})}{S_f S_O}$$

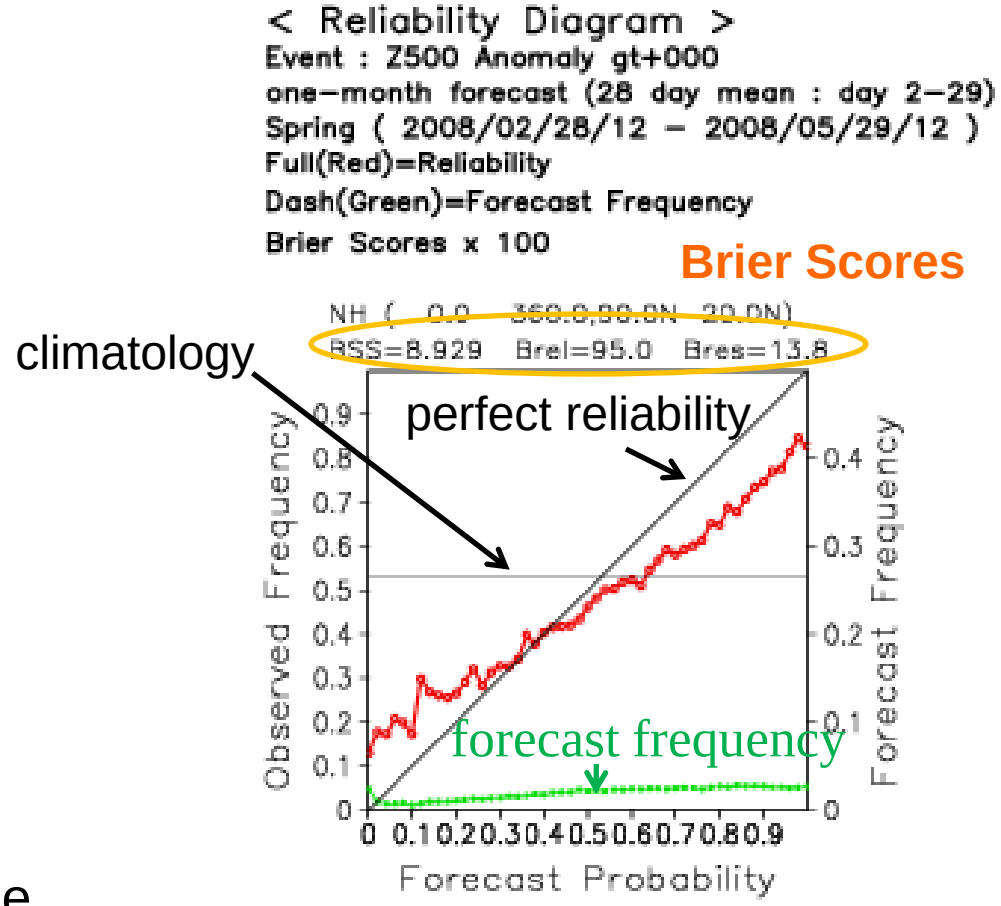
The first 3 terms are related to
① phase error (through the correlation)
②amplitude errors (through the ratio of the forecast to observed variances)
③bias error

Reliability diagram

The reliability diagram plots the observed frequency(Y-axis) against the forecast probability(X-axis).

The diagonal line indicates perfect reliability (observed frequency equal to forecast probability for each category).

Points below(above) the diagonal line indicate overforecasting (underforecasting).



Brier (skill) score

- Brier score measures mean squared error of the probability forecasts.

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2$$

p_i : **forecast probability**

o_i : **observed occurrence (0 or 1)**

N : **sample size**

Range: 0 to 1. Perfect score: 0 Climatology: $\bar{o}(1-\bar{o})$
Random:1/3

- Brier skill score measures ***skill*** relative to a reference forecast (usually climatology).

$$BSS = 1 - \frac{BS}{BS_{reference}}$$

Range: minus infinity to 1. BSS=0 indicates no skill when compared to the reference forecast. Perfect score: 1.

Decomposition of the Brier score

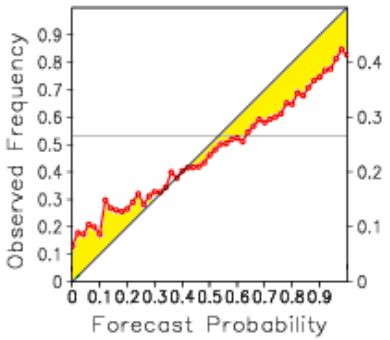
Murphy(1973) showed that the Brier score could be partitioned into three terms(for K probability classes and N samples). These terms are shown separately to attribute sources of error.

$\bar{o}$:**climatological occurrence**

$$BS = \frac{1}{N}\sum_{k=1}^{K} n_k (p_k - \overline{o_k})^2 - \frac{1}{N}\sum_{k=1}^{K} n_k (\overline{o_k} - \bar{o})^2 + \bar{o}(1-\bar{o})$$

reliability (brel)

the mean squared difference between the forecast probability and the observed frequency.
Perfect score: 0



resolution (bres)

the mean squared difference between the observed frequency and climatological frequency.
- indicates the degree to which the forecast can separete different situations.

climatologial forecast score:0
Perfect score: $\bar{o}(1-\bar{o})$

uncertainty (bunc)

measures the variability of the observations.

Brier skill score

= the relative skill of the probabilistic forecast to the climatology

Brier skill score

$$BSS = \frac{BS - BS_{clim}}{0 - BS_{clim}} = 1 - \frac{BS}{BS_{clim}}$$

$$Skill\ score = \frac{score_{forecast} - score_{reference}}{score_{perfect\ forecast} - score_{reference}}$$

$$BS_{clim} = bunc$$

Range: minus infinity to 1. Perfect score: 1
BSS=0 indicates no skill when compared to the climatology.
BSS>0 : better than clim.

Reliability skill score

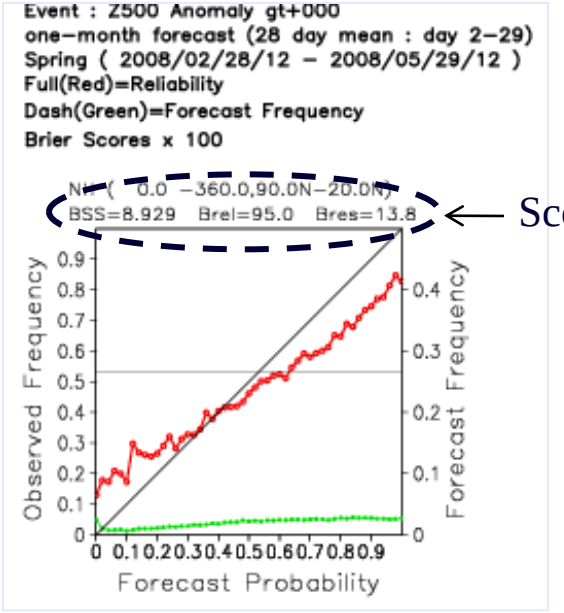
$$Brel = 1 - \frac{brel}{BS_{clim}}$$

Perfect score: 1

Resolution skill score

$$Bres = \frac{bres}{BS_{clim}}$$

Perfect score: 1

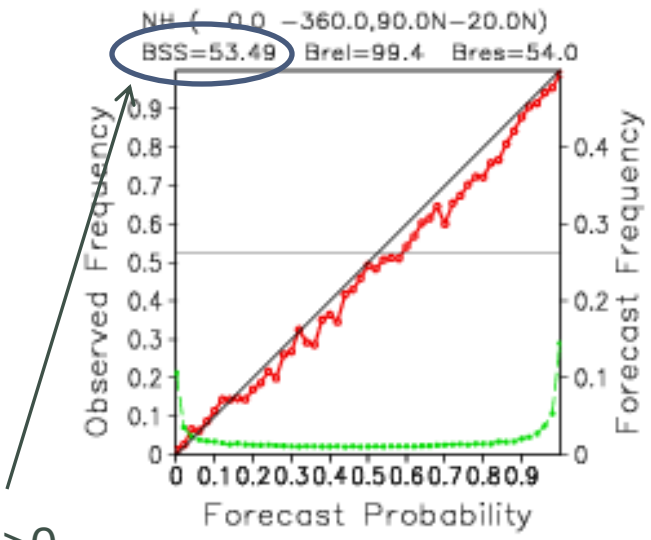


The larger these skill scores are, the better.

Interpretation of Reliability diagram and BSS

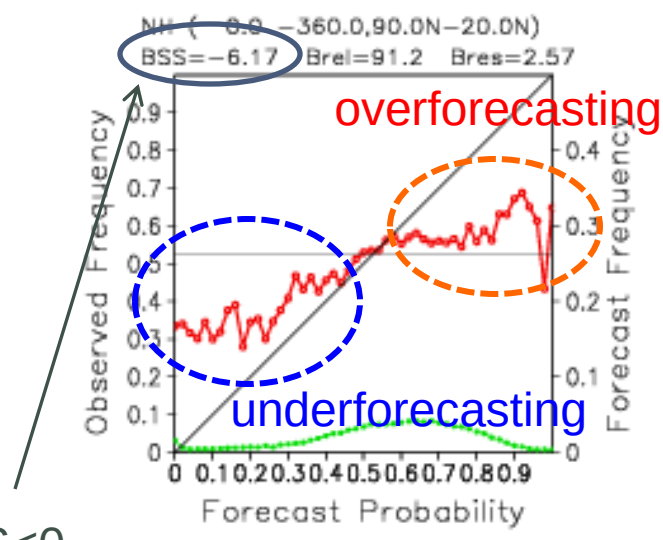
Event : Z500 Anomaly > 0
Northern Hemisphere
Spring of 2008 (2008/2/28 ～2008/5/29)

First week forecast
(day 2-8)



BSS>0
better than climatology

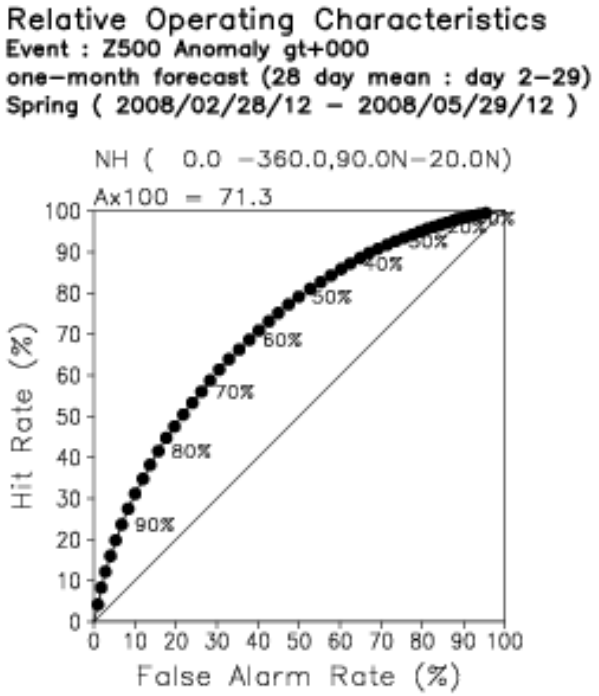
Third and fourth week
(day 16-29)



BSS<0
inferior to climatology

Relative Operating Characteristic (ROC)

- ROC is created by plotting the hit rate(Y-axis) against the false alarm rate(X-axis) using increasing probability thresholds to make the yes/no decision.



  - The area under the ROC curve (=ROC area) is frequently used as a score.

Steps for making ROC diagram

1. For each forecast probability category, count the number of hits,misses, false alarms, and correct non-events
2. Compute the hit rate and false alarm rate in each category k

   hit $rate_k$= $hits_k$/ ($hits_k$+ $misses_k$)

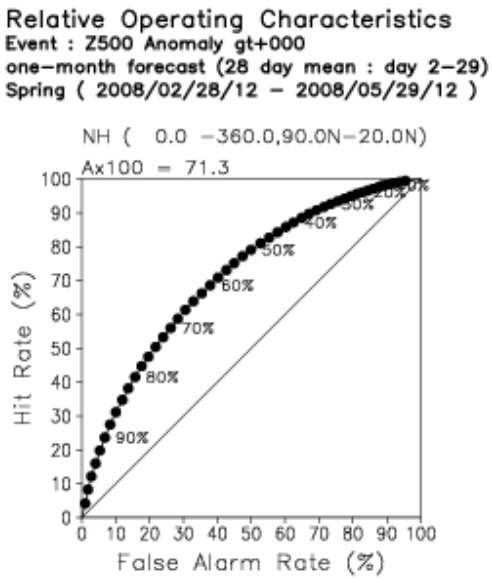
   false alarm $rate_k$= false $alarms_k$/ (false $alarms_k$+ correct non-$events_k$)
3. Plot hit rate vs false alarm rate
4. ROC area is the integrated area under the ROC curve

Observed

Forecast	yes	no
yes	hits	false alarms
no	misses	correct non-events
total	Observed yes	Observed no

Contingency Table

Forecast probability	Hit rate	False alarm rate
≥ 0.0		
≥ 0.02		
≥ 0.04		
·		
·		
·		
·		
·		
·		
·		
≥ 1.0		



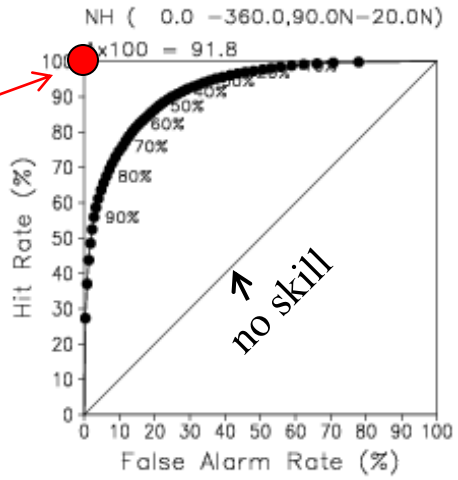
Interpretation of ROC curves

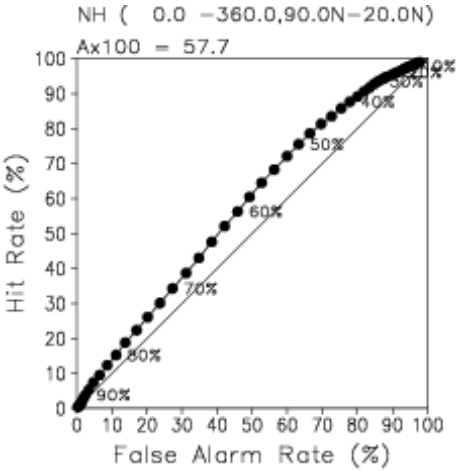
Event: Z500 anomaly > 0
Spring 2008

1st week forecast

3rd and 4th week forecast

Perfect performance





high resolution
(high potential skill)

low resolution
(low potential skill)

- ROC is not sensitive to bias in forecasts. Forecasts with bias may have a good ROC curve if resolution is still good. In this sense, the ROC can be considered as a measure of potential usefulness.
- On the other hand, reliability diagram is sensitive to the bias. It is needed to see both the ROC and the reliability diagram.

SVSLRF

(Standard Verification System for Long-Range Forecast)

- WMO standard tool to verify skill in seasonal models
- It was introduced by the Commission for Basic Systems (CBS) of the World Meteorological Organization (WMO) in December, 2002.
- Users can appropriately evaluate forecast skill with common measures.

Outline of SVSLRF

Mandatory part

	Parameters	Verification regions	Deterministic forecasts	Probabilistic forecasts
Level 1	T2m anomaly Precipitation anomaly (Nino3.4 Index)	Tropics(20S-20N) Northern extratropics(20N-90N) Southern extratropics(20S-90S) (N/A)	MSSS	ROC curves ROC areas Reliability diagrams Frequency histograms
Level 2	T2m anomaly Precipitation anomaly (SST anomaly)	Grid-point verification on a 2.5° by 2.5° grid	MSSS and its three-term decomposition at each grid-point	ROC areas at each grid-point
Level 3	T2m anomaly Precipitation anomaly (SST anomaly)	Grid-point verification on a 2.5° by 2.5° grid	3 by 3 contingency tables at each grid-point	ROC reliability tables at each grid-point

LC-SVSLRF

Lead Centre for the Long-Range Forecast Verification System

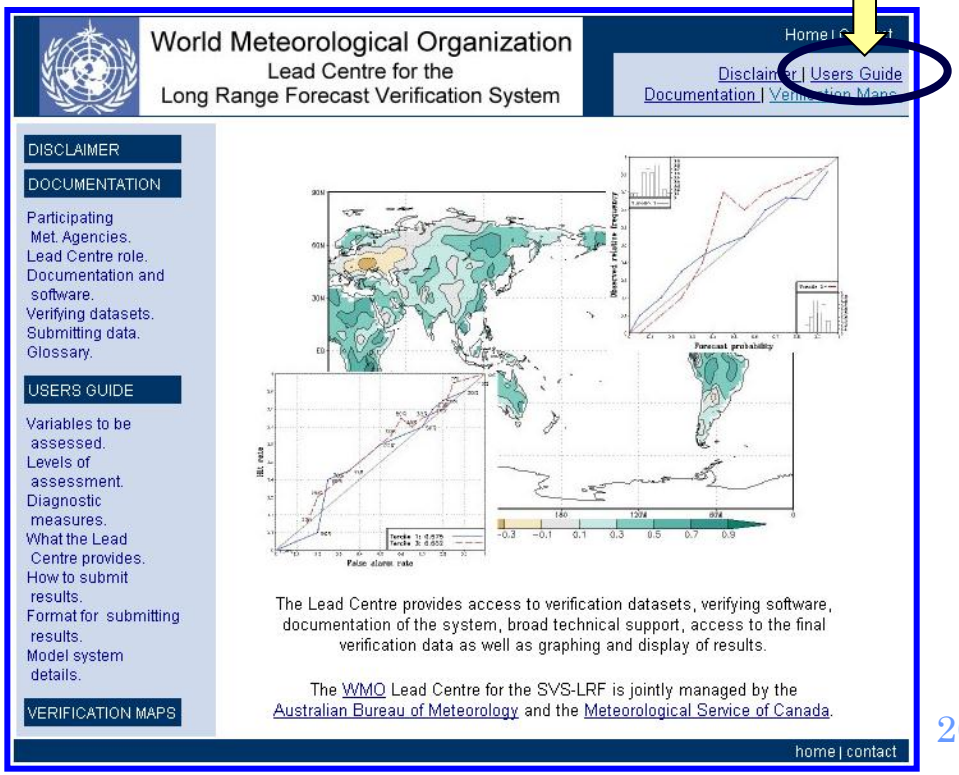
- Australian Bureau of Meteorology (BOM)
- Meteorological Service of Canada (MSC)

URL: http://www.bom.gov.au/wmo/lrfvs/

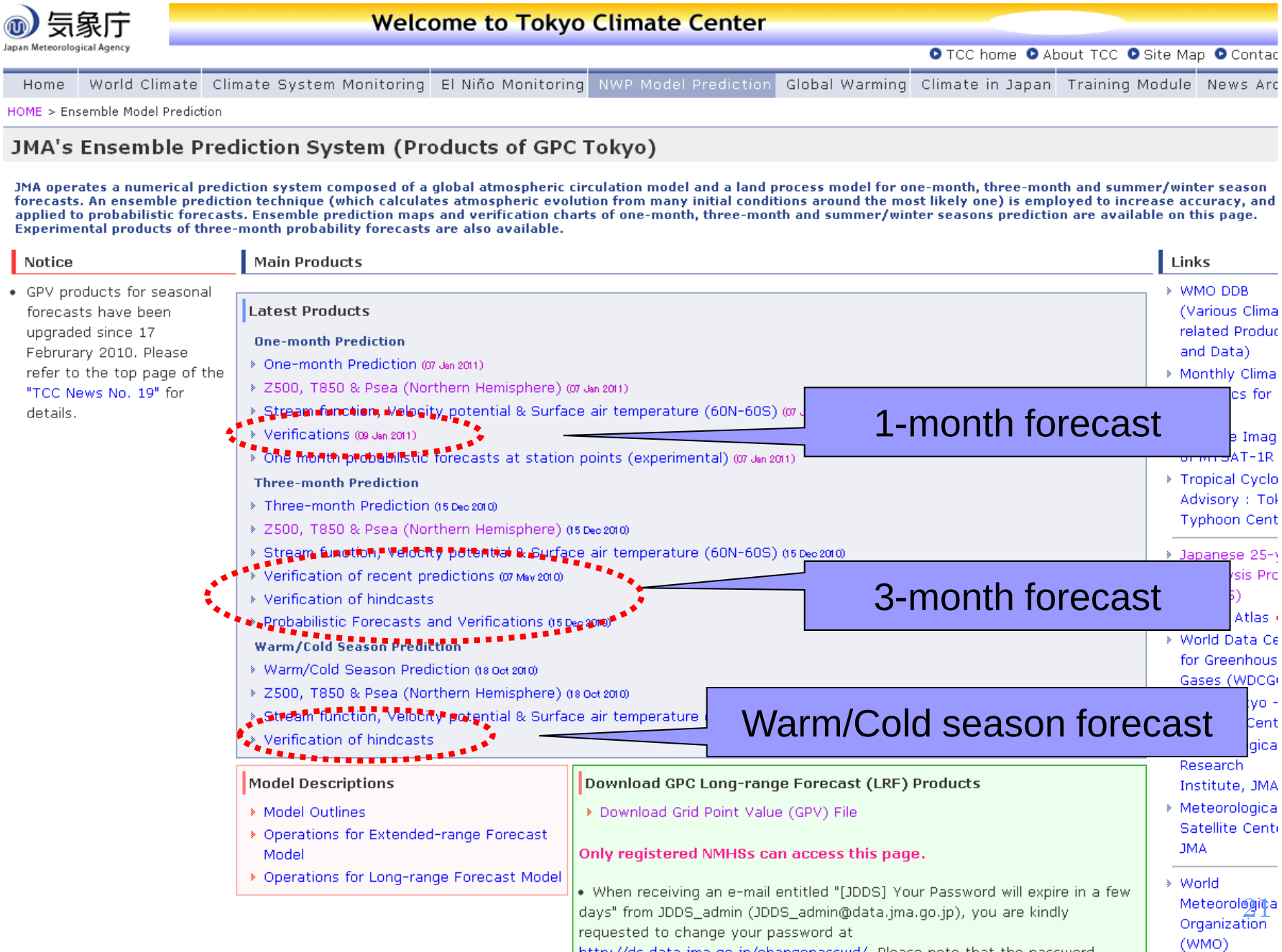
The Manual is here

Lead Center web site

You can get scores of several centers on Lead Centre website.



3. Verification results on TCC web page



Verification of operational 1-month forecasts

verification of one month forecast

Maps

- Error maps for every forecast (updated every week)
  - Z500, T850 and PSEA
    * Systemtic error is removed.(Bias based on hindcas
    Climatological normals were calculated with NCEP/N 94-2000).
  - Stream Function and Velocity Potential
    * Model normals based on hindcast from 1982 to 200
    Observed climatology were calculated with ERA-15(1979-1993).
- Reliability diagrams for each season
- ROC curves for each season

Scores

- Score in each season
- Score in each year

Summary of verification in 2001
Summary of verification in 2002
Summary of verification in 2003
Summary of verification in 2004
Summary of verification in 2005
Summary of verification in 2006
Summary of verification in 2007

Deterministic forecast

Ensemble mean forecast error maps, RMSE and Anomaly Correlation

Probabilistic forecast

Time sequence of RMSE and AC

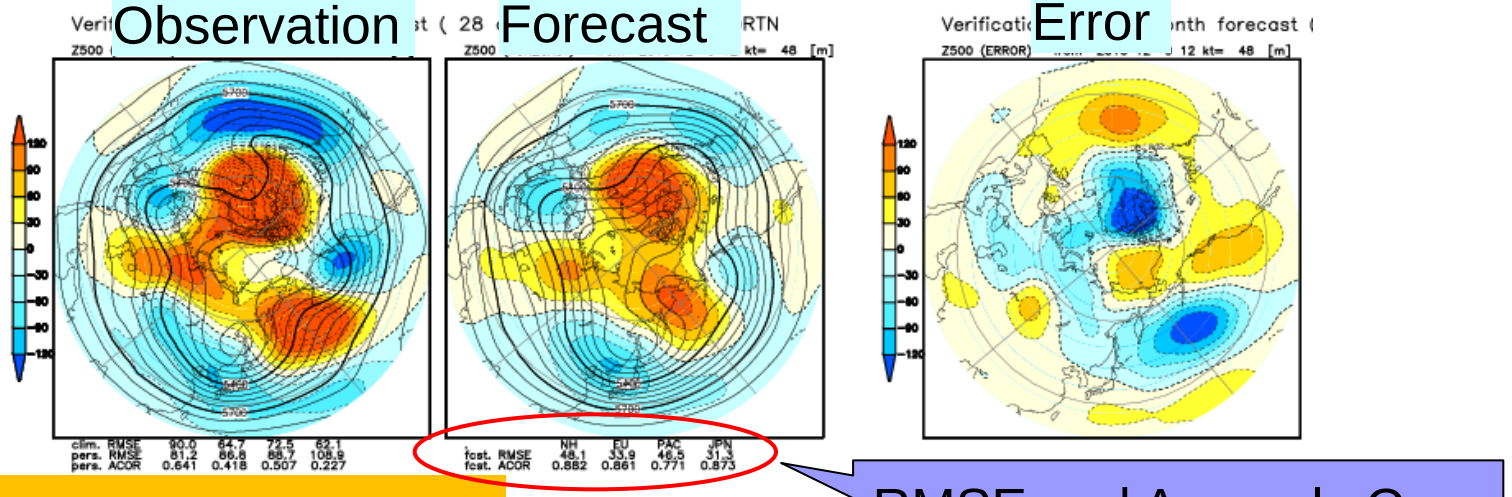
Verification of 1-month Ensemble mean forecast maps (Deterministic)

Z500 over the Northern Hemisphere

Observation | Forecast | Error

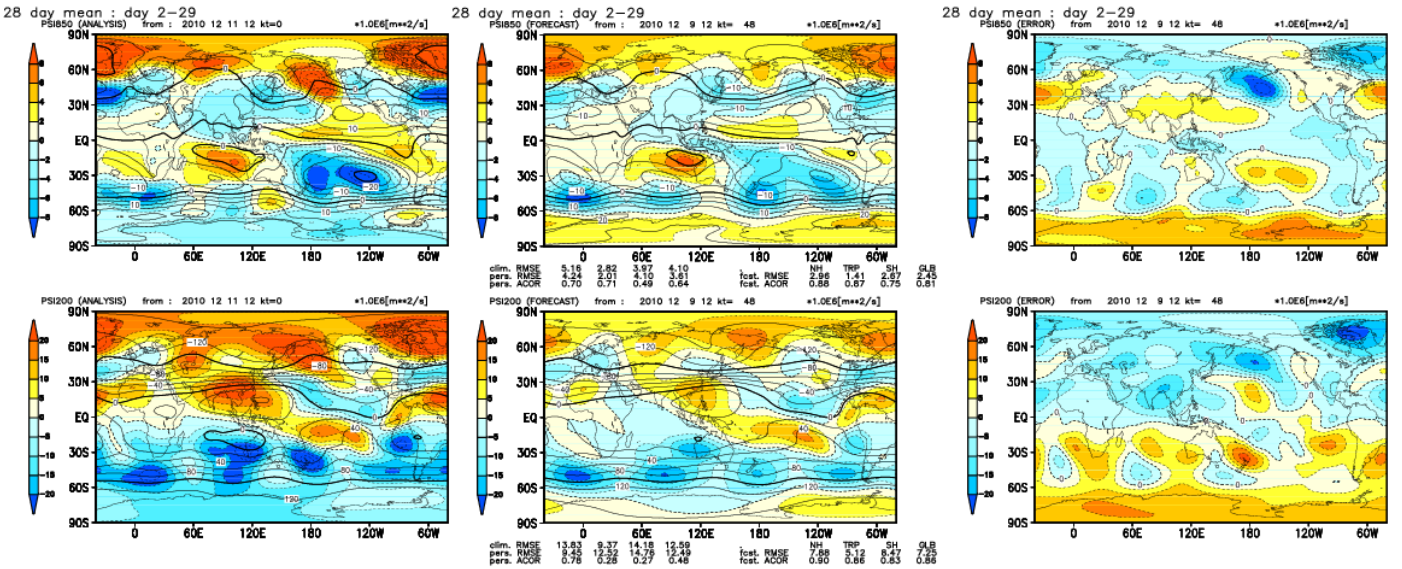
forecast period
28 days mean
initial date
2010.12.09.12Z
corresponding recent forecast

(from left to right)
left : analysis (Shaded patterns show anomalies.)
middle : ensemble mean forecast (Shaded patterns show anomalies.)



RMSE and Anomaly Corr.

Stream function (850, 200hPa)



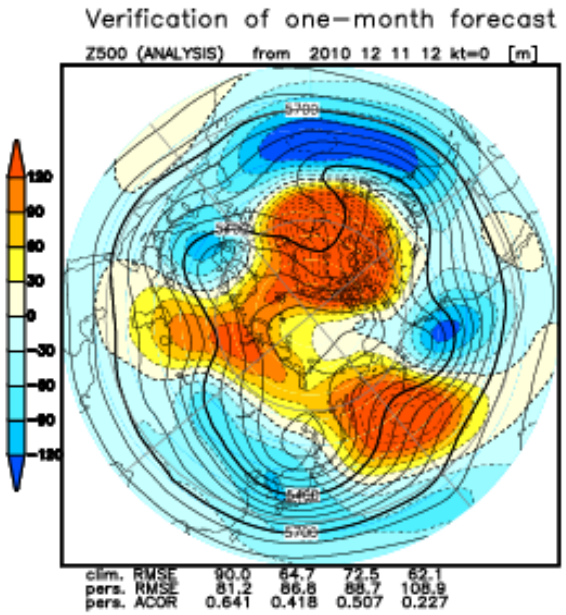
Verification of 1-month Ensemble mean forecast maps (Deterministic)

Z500 over the Northern Hemisphere

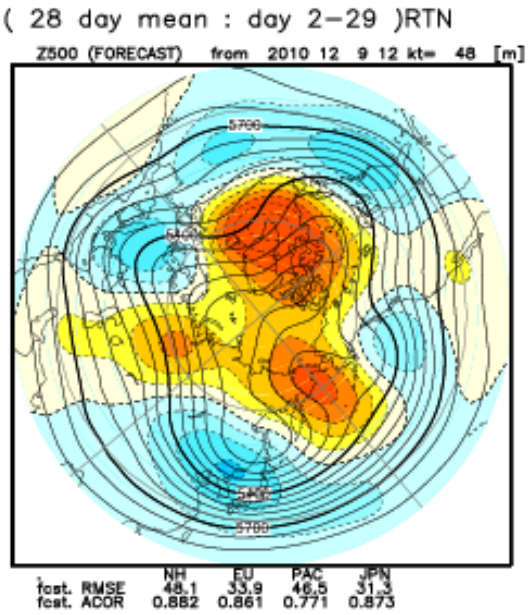
forecast period
28 days mean
initial date
2010.12.09.12Z
corresponding recent forecast

(from left to right)
left : analysis (Shaded patterns show anomalies.)
middle : ensemble mean forecast (Shaded patterns show anomalies.)

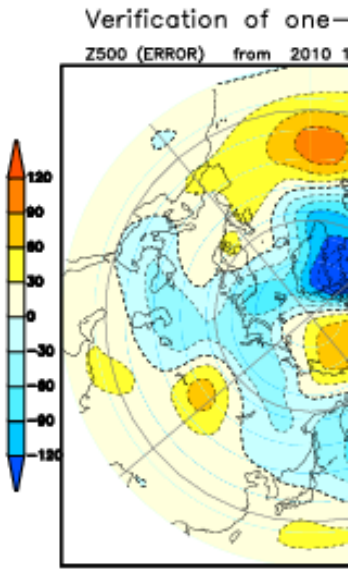
Observation



Forecast



Error



ANL-Clim (JRA-25/JCDAS)

FCT-Clim (model)

	NH	EU	PAC	JPN
fcst. RMSE	1.76	1.56	1.76	1.43
fcst. ACOR	0.905	0.938	0.915	0.955

Verification of operational 1-month forecasts

verification of one month forecast

Maps

- Error maps for every forecast (updated every week)
  - Z500, T850 and PSEA

    * Systemtic error is removed.(Bias based on hindcast from 1982 to 2001.)
    Climatological normals were calculated with NCEP/NCAR reanalysis-1(1971-1978),ERA-15(1979-1993) and GANAL(1994-2000).
  - Stream Function and Velocity Potential

    * Model normals based on hindcast fr
    Observed climatology were calculated
- Reliability diagrams for each season
- ROC curves for each season

Scores

- Score in each season
- Score in each year

Summary of verification in 2001
Summary of verification in 2002
Summary of verification in 2003
Summary of verification in 2004
Summary of verification in 2005
Summary of verification in 2006
Summary of verification in 2007

Probabilistic forecast

・**Reliability diagrams** and **Brier skill scores**

・**ROC curves and area** for each season

気象庁
Japan Meteorological Agency

Welcome to Tokyo Climate Center

TCC home | About TCC

Home | Climate in the World | Climate System Monitoring | El Niño Monitoring | NWP Model Prediction | Global Warming | Climate in Japan | Training

HOME > Ensemble Model Prediction

JMA's Ensemble Prediction System

JMA operates a numerical prediction system composed of a global atmospheric circulation model and a land process model for one-month, three-month and summer/winter season forecasts. An ensemble pr calculates atmospheric evolution from many initial conditions around the most likely one) is employed to increase accuracy, and is applied to probabilistic forecasts. Ensemble prediction maps and verificati month and summer/winter seasons prediction are available on this page. Experimental products of three-month probability forecasts are also available.

Notice

- JMA's one-month prediction model was upgraded on 21 March 2008. Available products remain the same. Verification maps of one-month probabilistic forecasts at station points have been updated accordingly.
- JMA's extended ensemble prediction systems (EPS) was updated on 9 March 2007. Please refer to the "TCC News No.7" for details.
- JMA's extended ensemble prediction systems (EPS) (for three-month and warm/cold season predictions) was updated on 12 September 2007. Please refer to the "TCC News No.9" for details.

Main Products

Latest Products

One-month Prediction

- One-month Prediction (05 Sep 2008)
- Z500, T850 & Psea (Northern Hemisphere) (05 Sep 2008)
- Stream function, Velocity potential & Surface air temperature (60N-60S) (05 Sep 2008)
- Verifications (07 Sep 2008)
- One month probabilistic forecasts at station points (experimental) (06 Jun 2008) NEW

Three-month Prediction

- Three-month Prediction (01 Sep 2008)
- Z500, T850 & Psea (Northern Hemisphere) (01 Sep 2008)
- Stream function, Velocity potential & Surface air temperature (60N-60S) (01 Sep 2008)
- Verification of recent predictions (01 Sep 2008)
- Verification of hindcasts
- Probabilistic forecasts and verifications (19 Aug

Warm/Cold Season Prediction

- Warm/Cold Season Prediction (20 Apr 2008)
- Z500, T850 & Psea (Northern Hemisphere) (20
- Stream function, Velocity potential & Surface
- Verification of hindcasts

Verification of hindcast based on SVSLRF

Model Descriptions

- Model Outlines
- Operations for Extended-range Forecast Model
- Operations for Long-range Forecast Model

Download GPV (Grid Point Value)

- Download GPV file

Only registered NMHSs can access this page.

Hindcast verification methods based on Standardised Verification System for Long-Range Forecasts (SVSLRF)

Verification of deterministic forecasts

- Mean Square Skill Score
  RAIN | T2m | PSEA | Z500 | T850
  Dependence of MSSS on Initial date
- 3 by 3 contingency tables (in Japan)
  (Category : Below Normal, Near Normal, Above Normal)
  RAIN | T2m | PSEA | Z500 | T850

Verification of Probabilistic forecasts

- Reliability diagrams (Aggregated verification)
  (Anomaly > 0, Below Normal, Near Normal, Ab
  RAIN | T2m | PSEA | Z500 | T850
- Relative Operating Characteristics
  - ROC curves, ROC areas (Aggregated verificatio
    (Anomaly > 0, Below Normal, Near Normal, Above Normal)
    RAIN | T2m | PSEA | Z500 | T850
    Dependence of ROC areas on Initial date
  - ROC areas (Grid point verification)
    (Anomaly > 0, Below Normal, Near Normal, Above Normal)
    RAIN | T2m | PSEA | Z500 | T850

Verification data

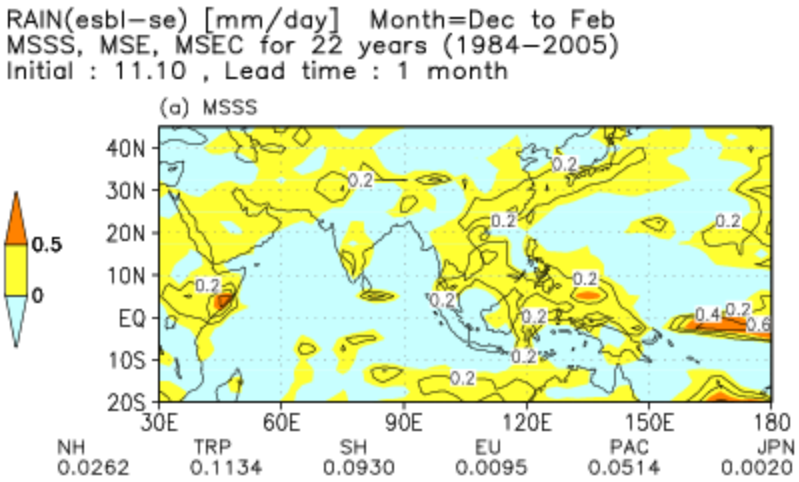
Verification of deterministic forecasts
- Mean Square Skill Score (MSSS)
- Contingency tables

Verification of probabilistic forecasts
- Reliability diagrams
- ROC curves and ROC areas

Examples of MSSS for precipitation

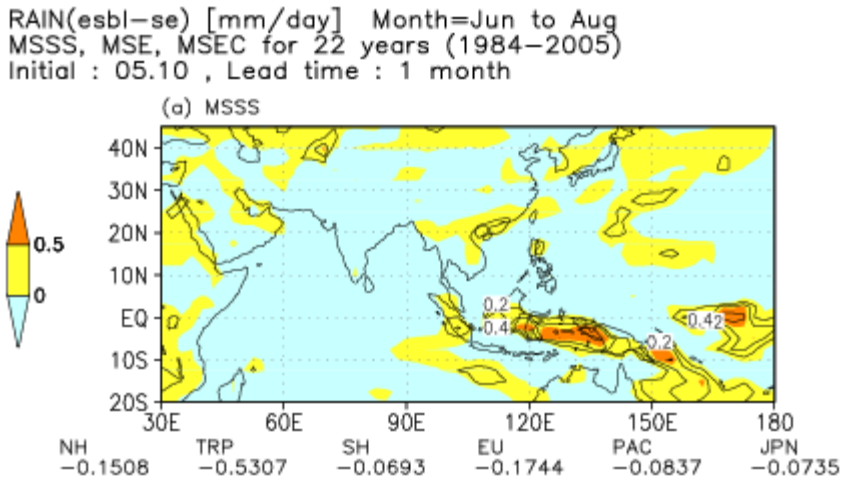
Dec-Feb

started on 10 Nov.



Jun-Aug

started on 10 May.



- Positive MSSS indicates that the forecast is better than climatological forecast.

Summary

Deterministic

Index	Random	Climatology	Perfect
RMSE	>0	・・・	0
ME	・・・	・・・	0
AC	-1	・・・	+1
MSSS	・・・	0	+1
MSE	>0	>0	0

Probabilistic

Index	Random	Climatology	Perfect
Reliability diagram			Fit to the diagonal line
BS	1/3	bunc=$\bar{o}(1-\bar{o})$	0
BSS(x100)	100>	0	+100
Brel(x100)	100>	100>	+100
Bres(x100)	100>	0	+100
Roc area(x100)	100>	50	+100

References

Murphy, A.H., 1973: A new vector partition of the probability score. J. Appl. Meteor., 12, 595-600.

Murphy, A.H., 1988: Skill scores based on the mean square error and their relationships to the correlation coefficient. Mon. Wea. Rev., 16, 2417-2424.

http://www.bom.gov.au/wmo/lrfvs/index.html

http://www.ecmwf.int/newsevents/meetings/workshops/2007/jwgv/index.html