Ensemble Verification

Yuejian Zhu

Environmental Modeling Center
NOAA/NWS/NCEP

Acknowledgements:
Zoltan Toth ***EMC***

Outlines

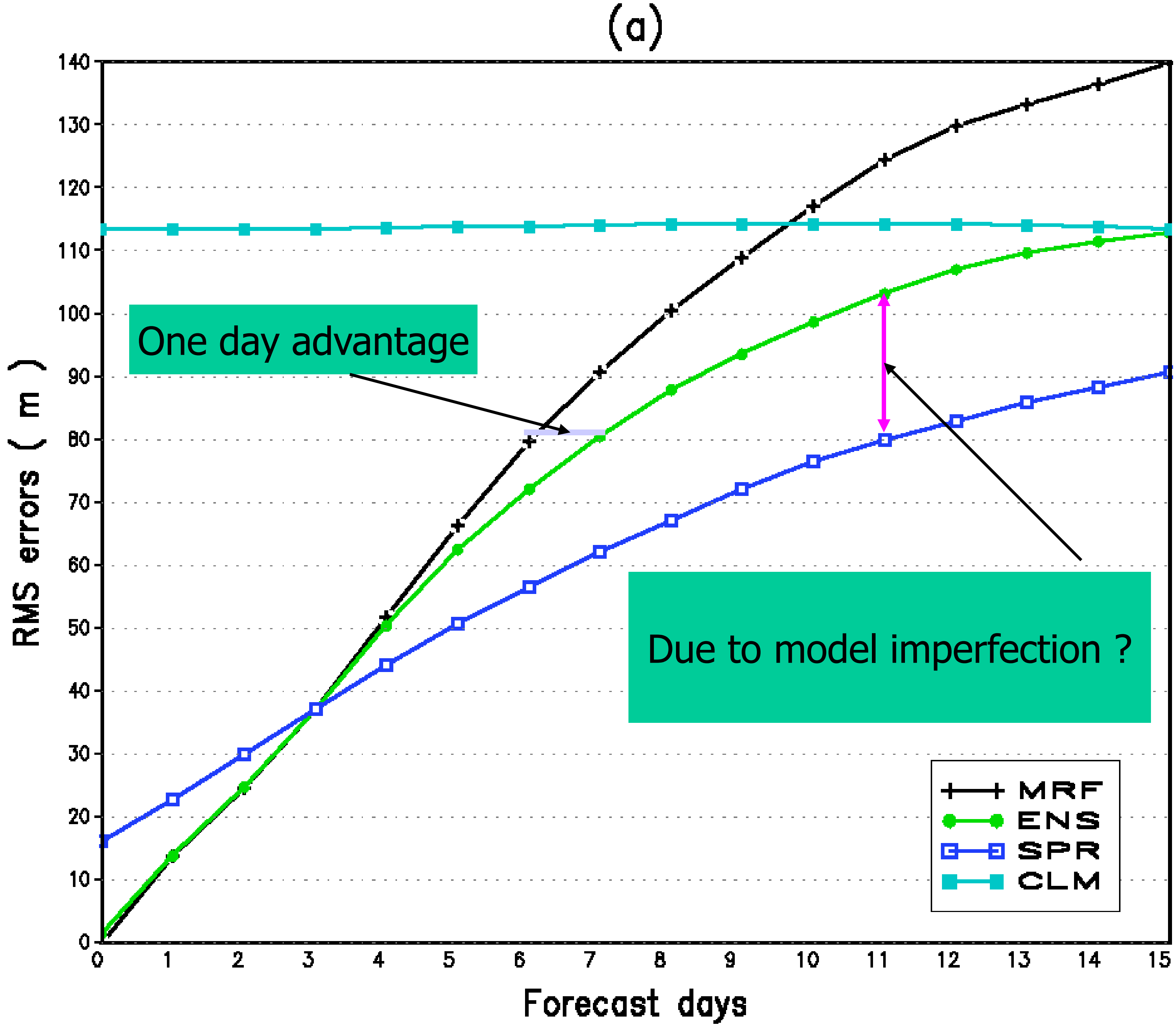
- ❑ Climatological Data
- ❑ Verify Analysis (proxy truth)
- ❑ RMS and Spread
- ❑ Mean Error and Absolute Error
- ❑ Histogram and Outlier
- ❑ RPS and RPSS
- ❑ CRPS and CRPSS
- ❑ BSS (Resolution and Reliability)
- ❑ ROC (Hit Rate and False Alarm Rate)
- ❑ Economic Value (cost-loss analysis)

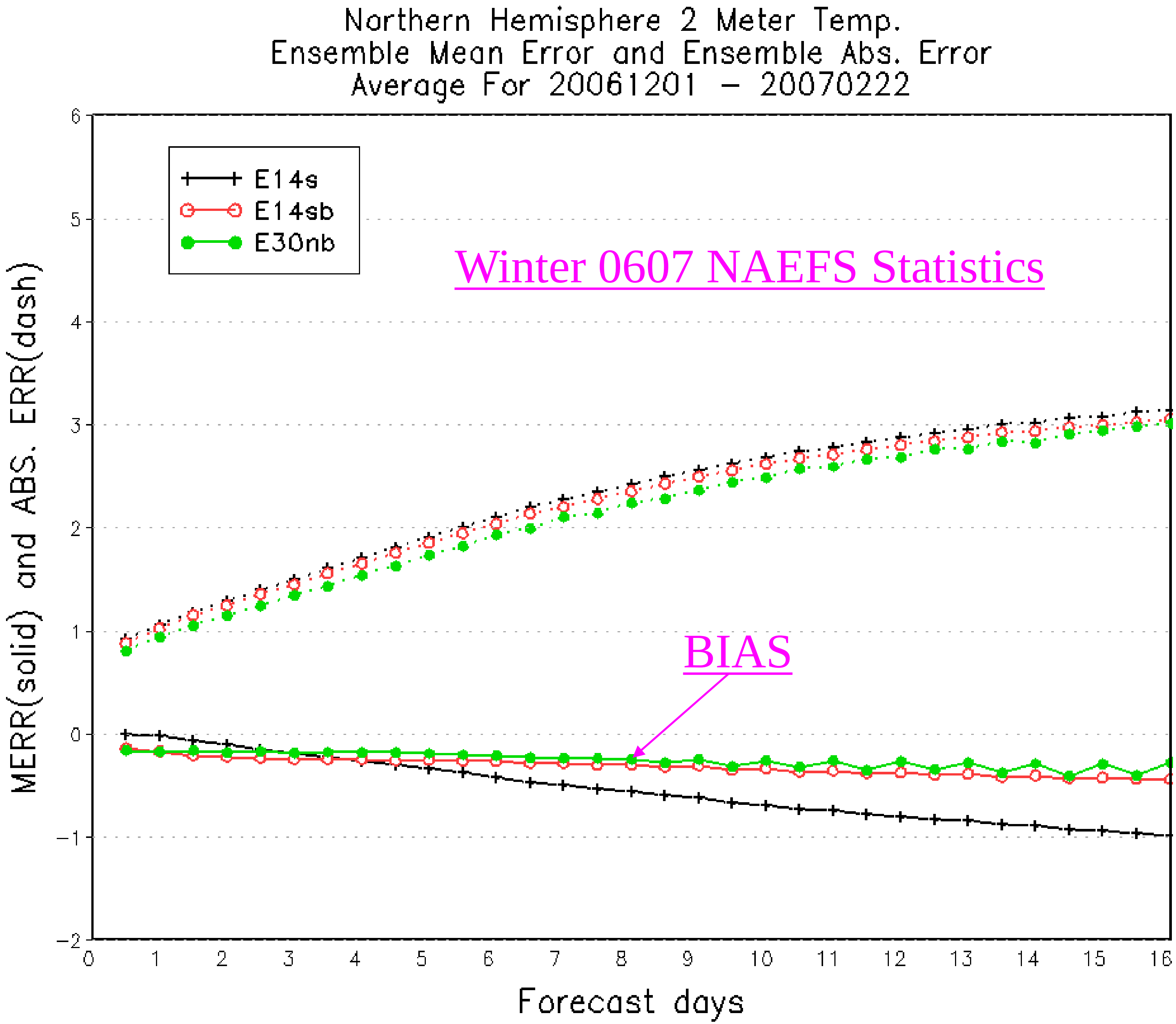
Climatological Data

- NCEP/NCAR 40 years (1958-1997) reanalysis
- Monthly Sampling
  - For example: 40*30=1200
- 10 equally-a-likely, based on sampling
- Projected to verify date
- All forecast skills will base on 10 equally-a-likely climatological bins.

Verify Analysis (proxy truth)

- All following deterministic and probabilistic verification are based on 2.5*2.5 grid forecast, analysis and climatology in globally
- NCEP best analysis (GSI) is our best reference (proxy truth) to apply all NCEP forecast verifications.
- Other model forecast verification is using their own available analysis (proxy truth)
- For jointed ensemble (or multi-model ensemble), it is using NCEP analysis (as truth) in practice.


(a)
RMS errors (m)
Forecast days
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
One day advantage
Due to model imperfection ?
MRF
ENS
SPR
CLM


Northern Hemisphere 2 Meter Temp.
Ensemble Mean Error and Ensemble Abs. Error
Average For 20061201 – 20070222
E14s
E14sb
E30nb
Winter 0607 NAEFS Statistics
BIAS
MERR(solid) and ABS. ERR(dash)
Forecast days

Prob. Evaluation ***(simple measurement)***

1. Talagrand Distribution (histogram distribution):

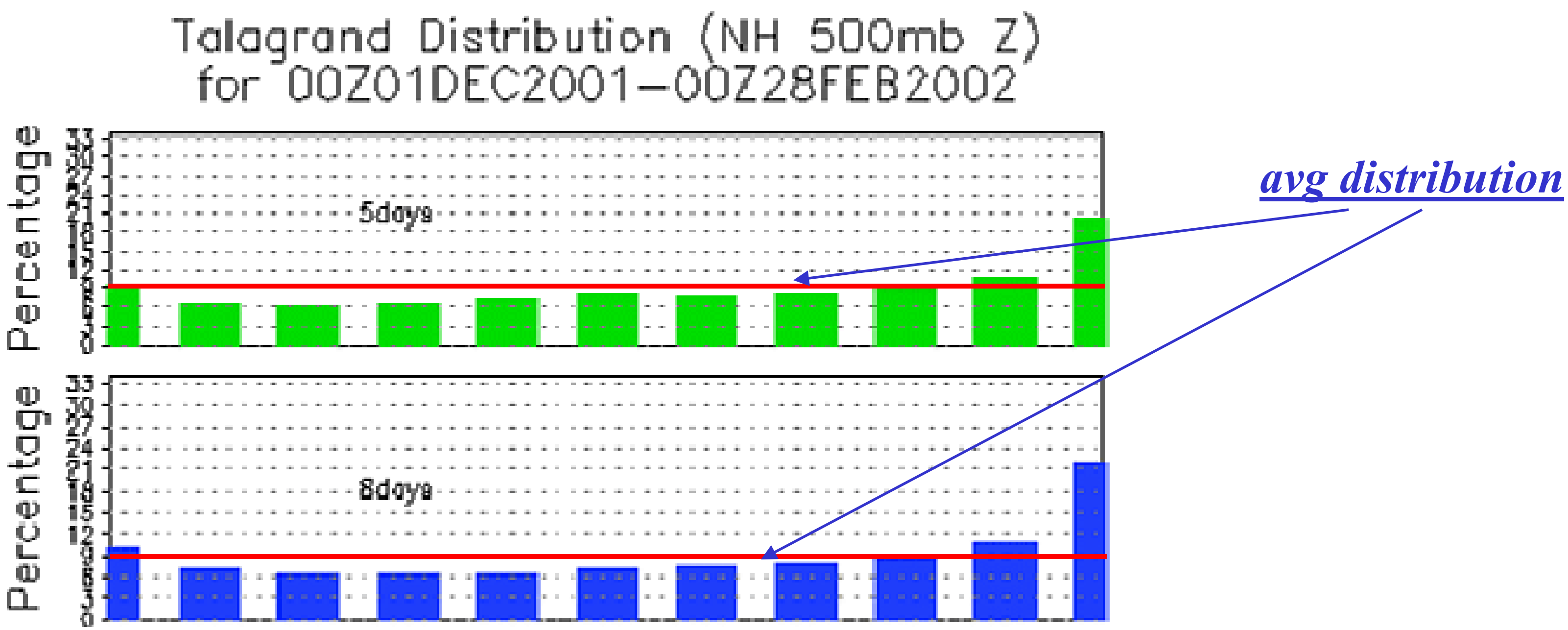
Sorting forecast in order, to check where the analysis is falling

Reliability measurement, system bias detected**.**

positive/negative biased for forecasting model,

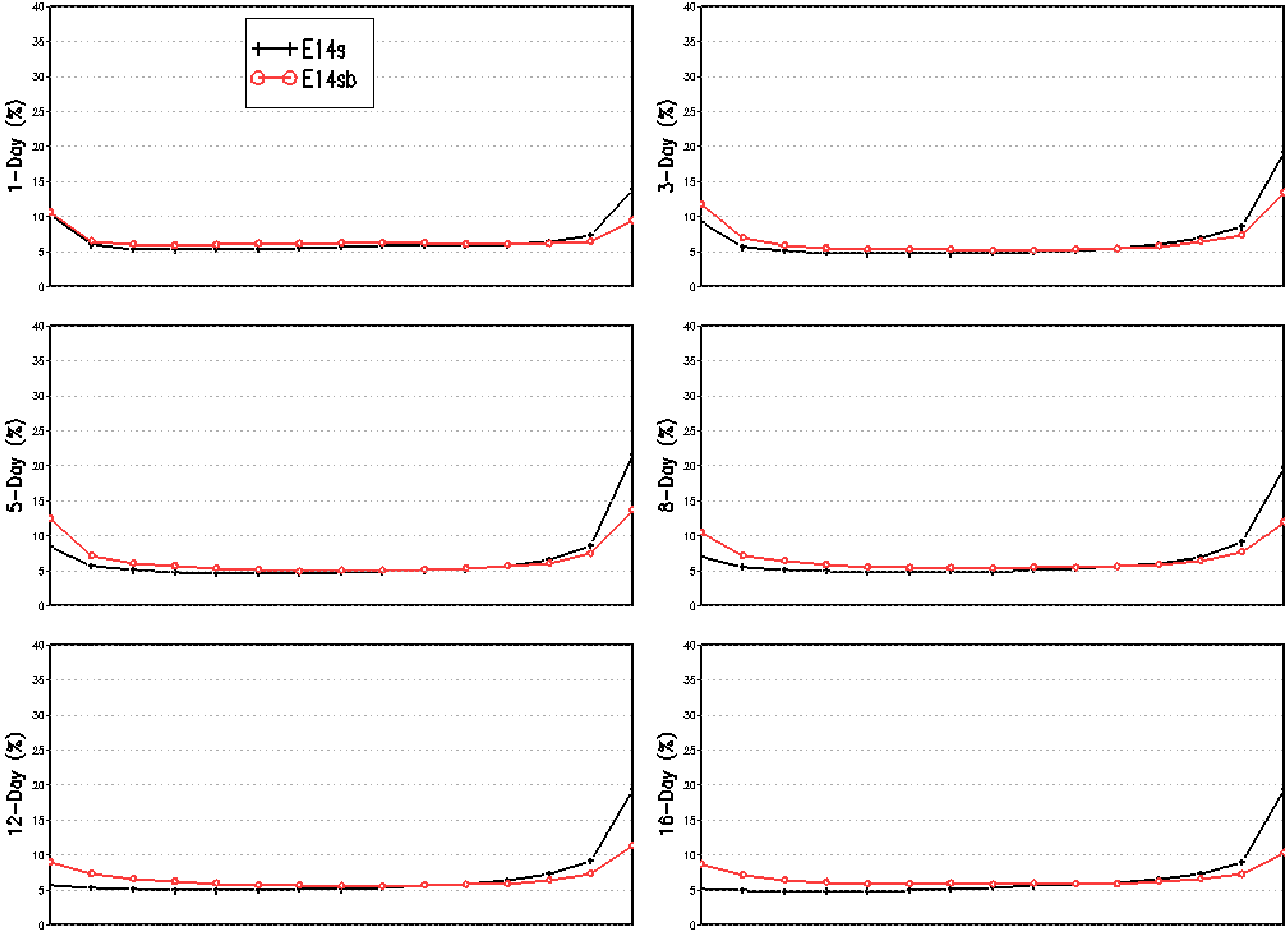
example of these forecasts --> cold bias,

assume analysis is bias-free (perfect). Common -"U" sharp



Northern Hemisphere 500hPa Height Histogram Distribution

Average For 20061201 – 20070222



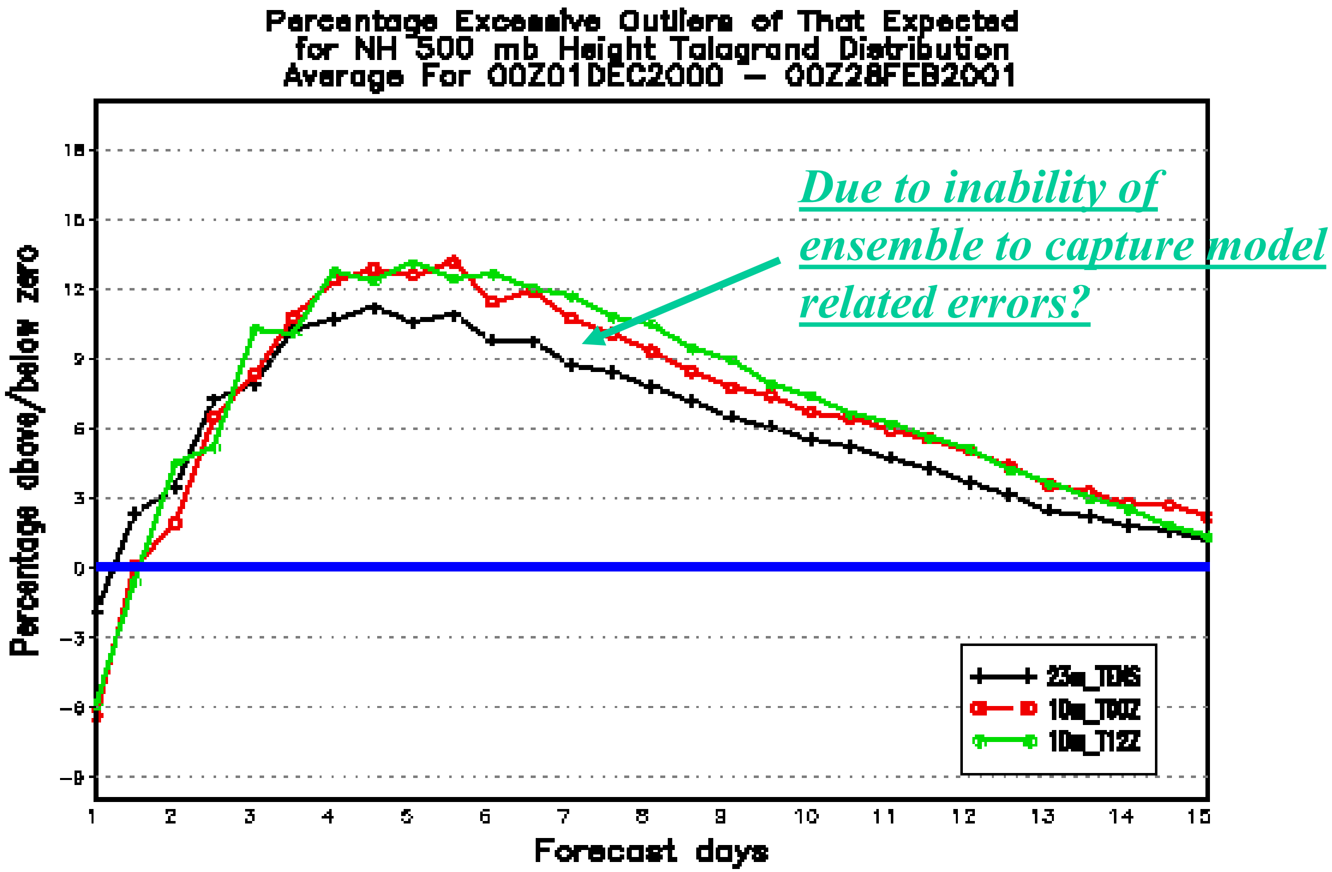
Prob. Evaluation (simple measurement)

1. Talagrand distribution (continue).

   . Outlier evolution by different leading time

   .. Adding up two outliers subtract the average.

   … Ideal forecasts will have zero outliers.



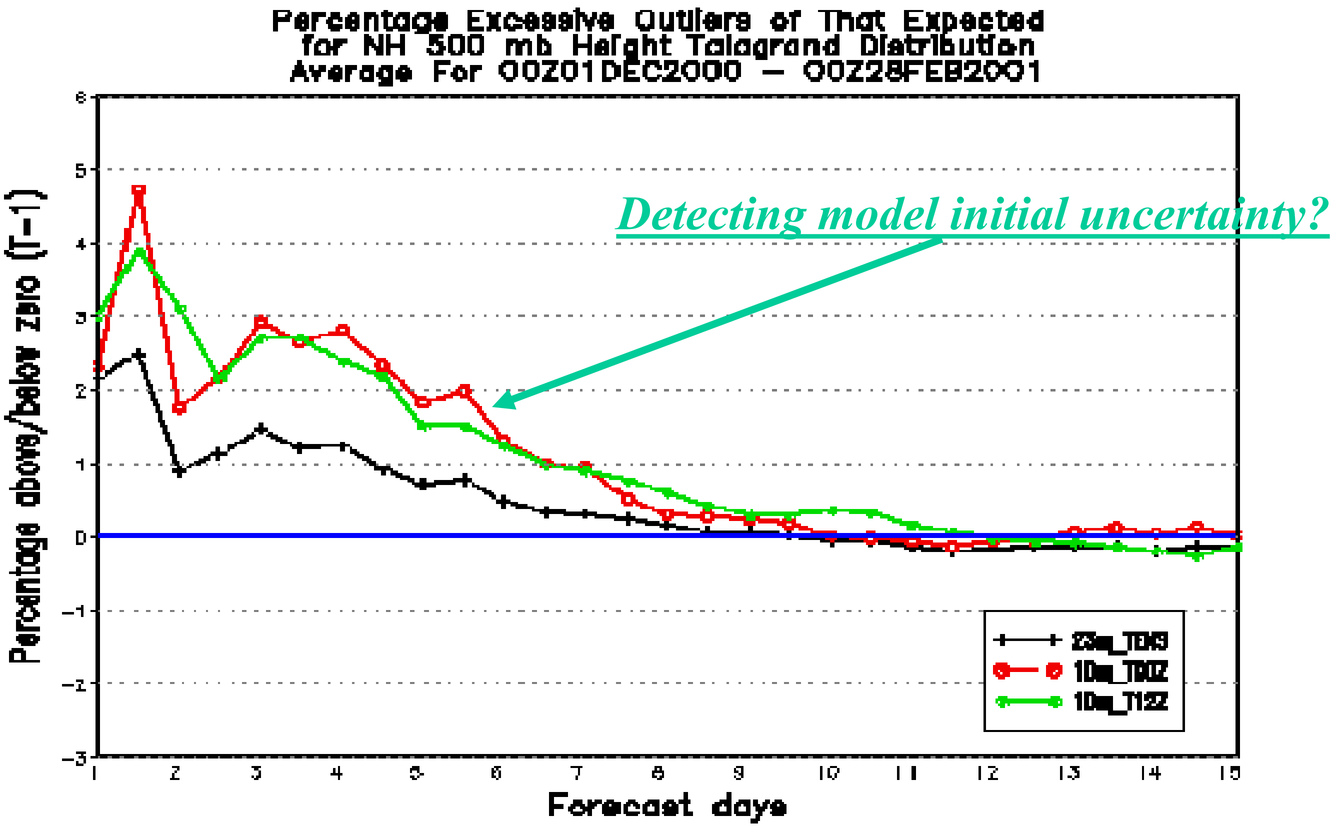
Prob. Evaluation (simple measurement)

Outlier --> diagnostic

forecasts .vs. next forecasts (f+24hrs valid at same time)

assume forecasting model is perfect, f+24.

perfect forecast system will expect the outliers are zero.



Prob. Evaluation (multi-categories)

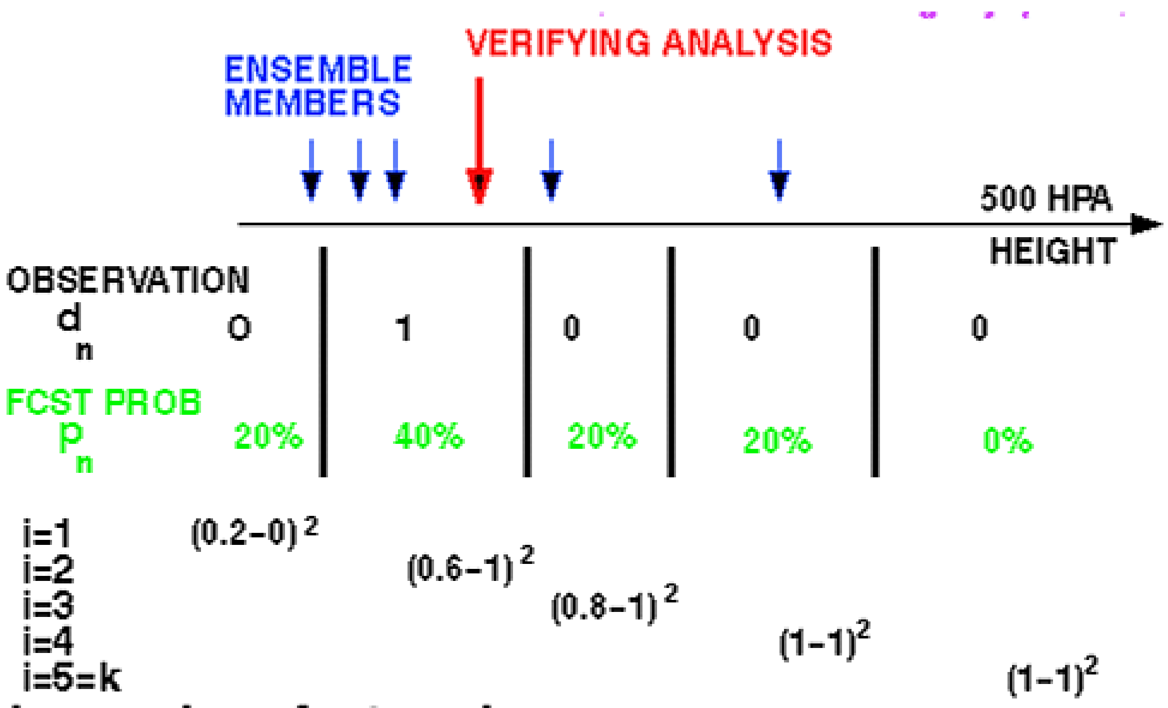
Based on climatological equally likely bins (for example. 5 bins)

For verifying multi-category probability forecasts.

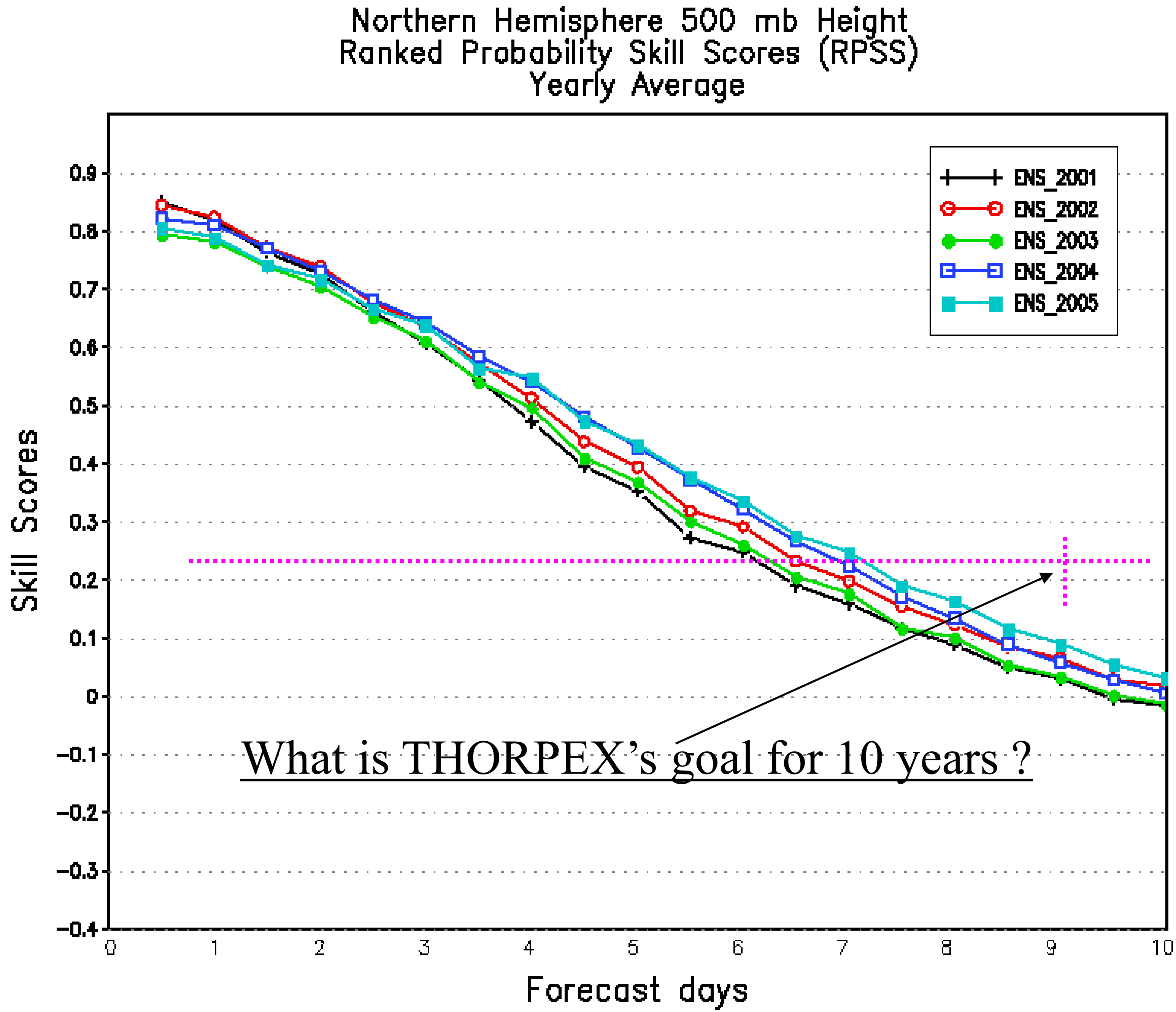
measure both reliability and resolution.

1. Ranked (ordered) probability score (RPS) and RPSS

RPSS=(RPSf - RPSc)/(1 - RPSc)



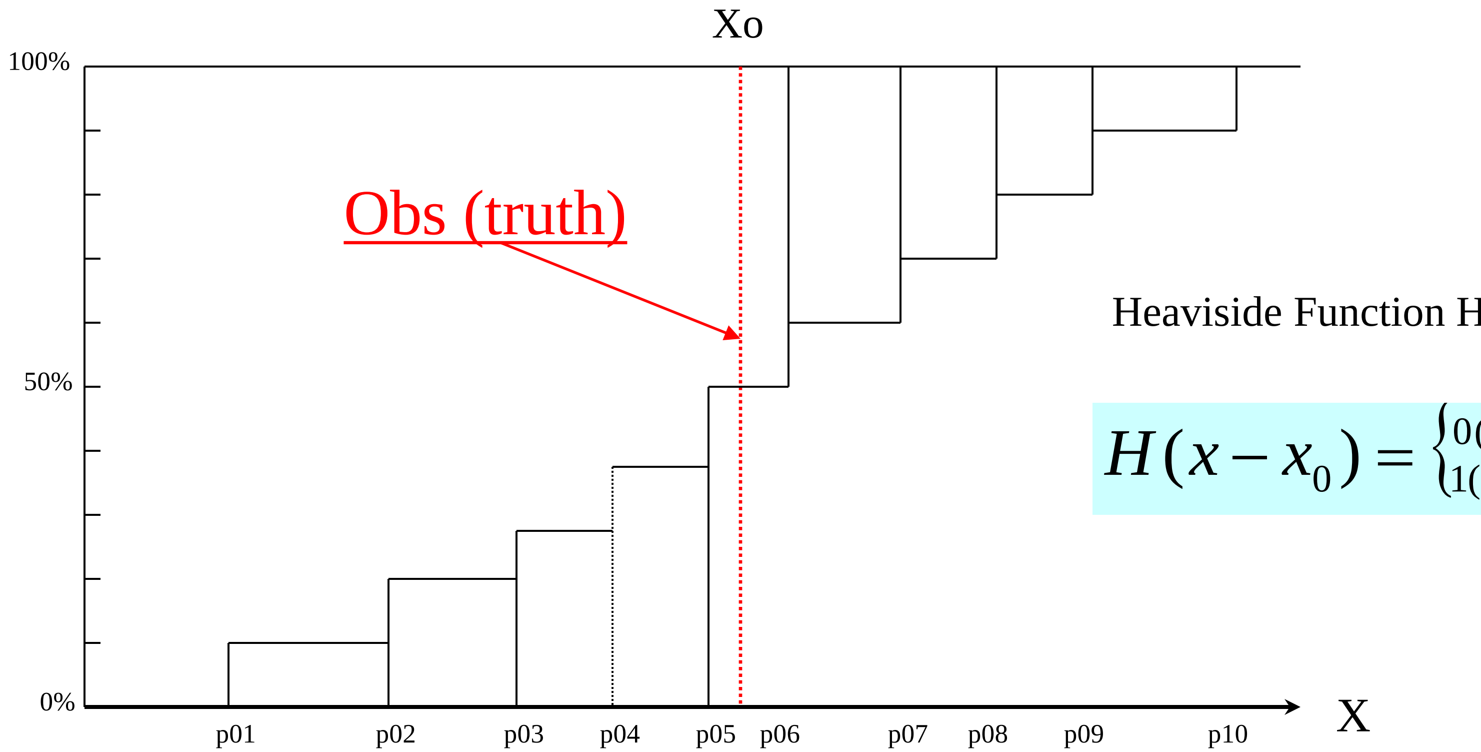
$$\mathrm{RPS}(p,d) = 1 - \frac{1}{k-1}\left[\sum_{i=1}^{k}\left(\sum_{n=1}^{i} P_n - \sum_{n=1}^{i} d_n\right)^2\right]$$


Northern Hemisphere 500 mb Height
Ranked Probability Skill Scores (RPSS)
Yearly Average
Skill Scores
Forecast days
ENS_2001
ENS_2002
ENS_2003
ENS_2004
ENS_2005
What is THORPEX's goal for 10 years ?

Continuous Rank Probability Score

$$CRPS = \int_{-\infty}^{+\infty} [F(x) - H(x - x_0)]^2 \, dx$$

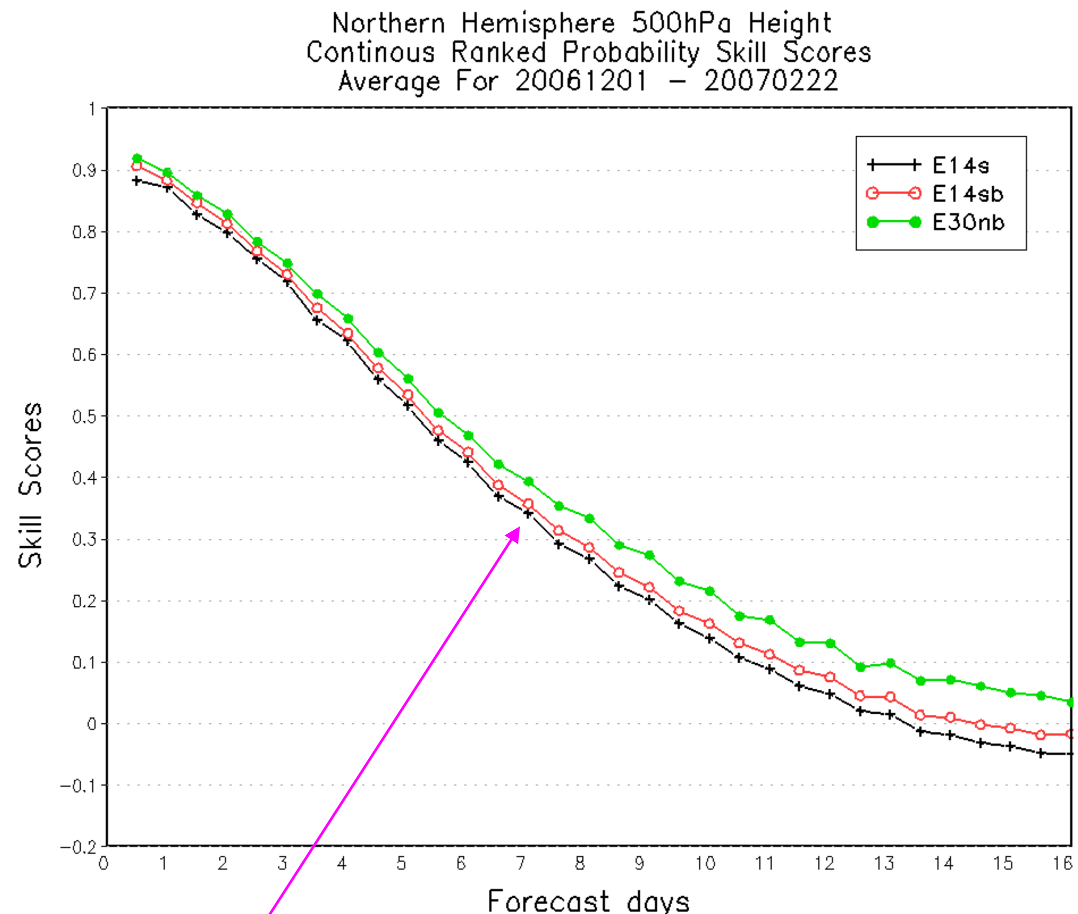
$$CRPSS = \frac{CRPS_c - CRPS_f}{CRPS_c}$$

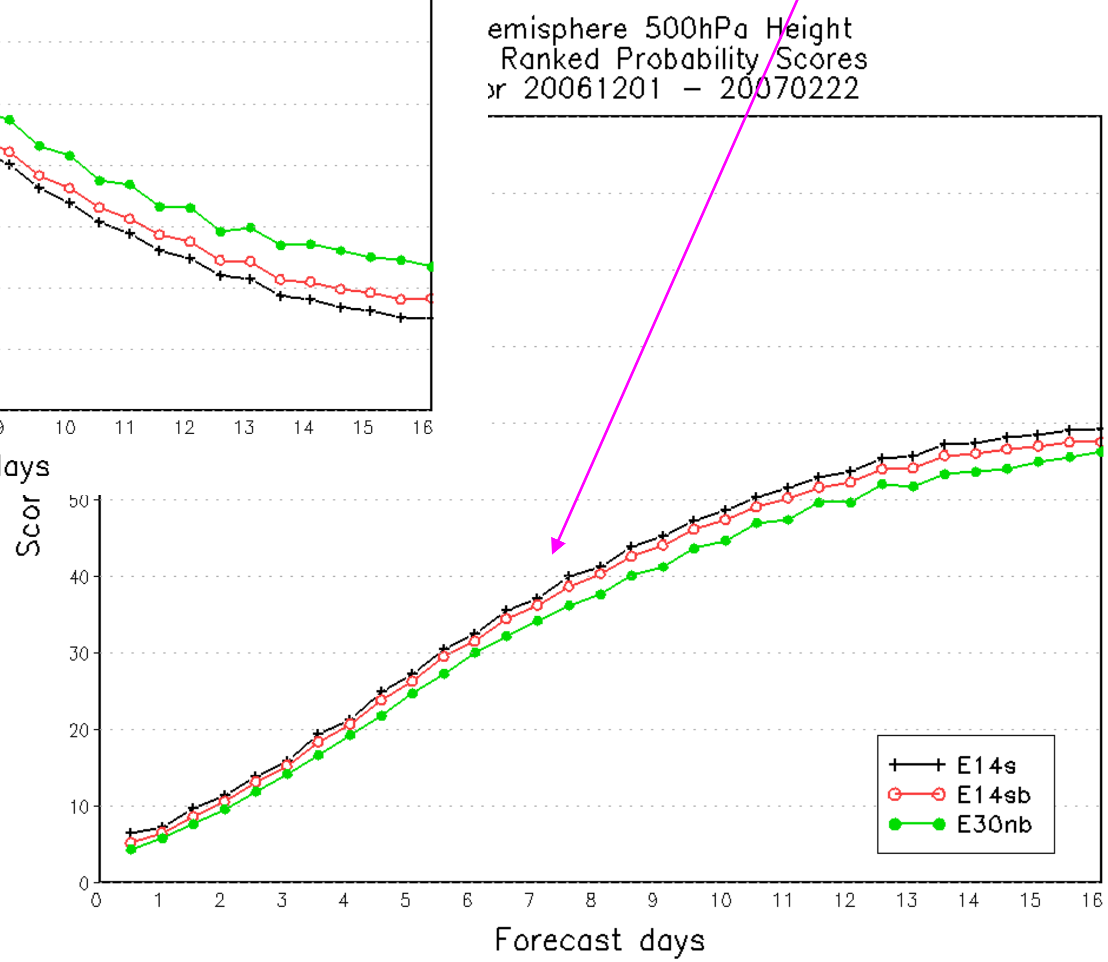


$$H(x - x_0) = \begin{cases} 0 (x \le x_o) \\ 1 (x > x_o) \end{cases}$$

Order of 10 ensemble members (p01, p02,…,p10)

CRPS for winter 0607

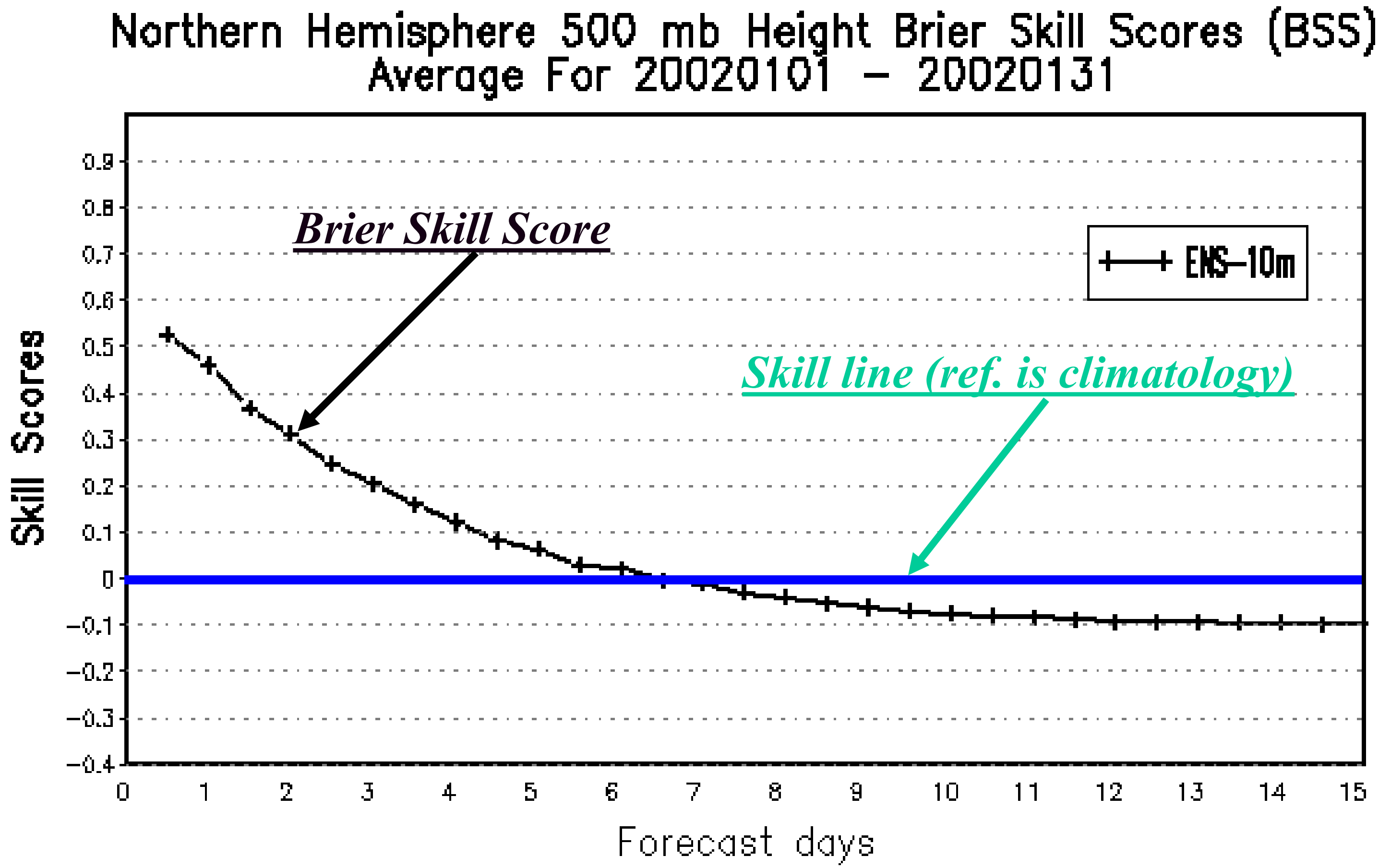




CRPSS for winter 0607

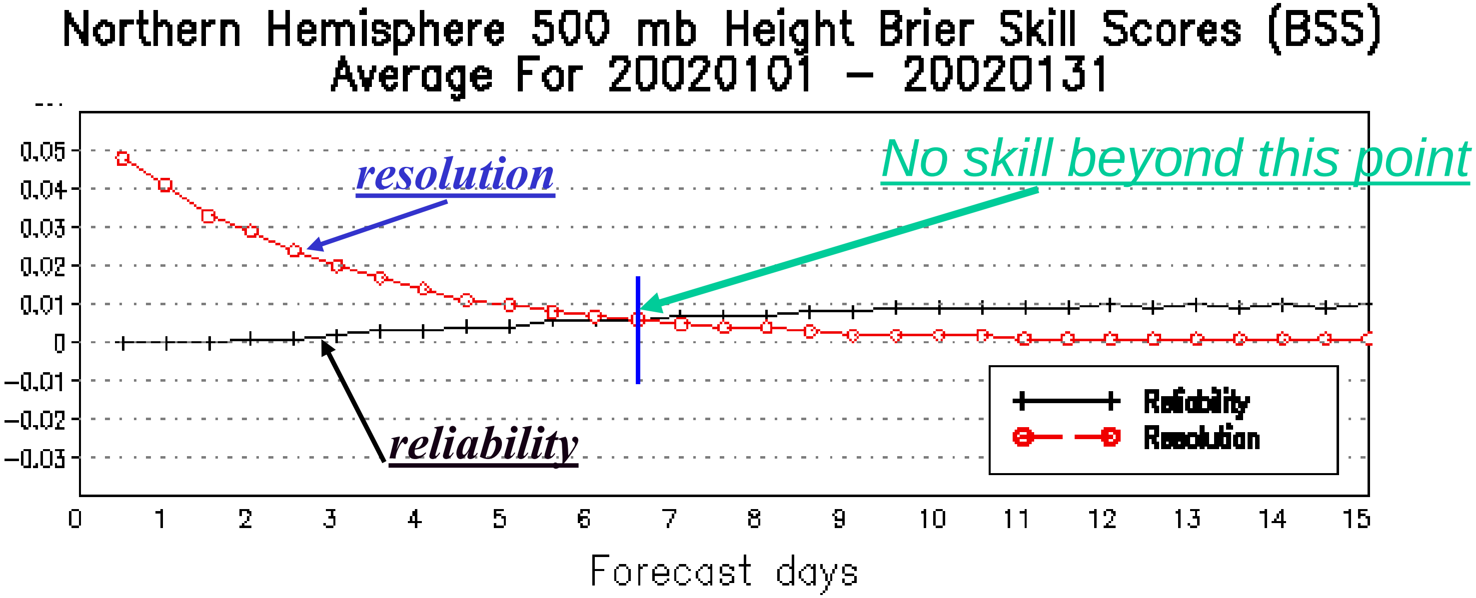
Prob. Evaluation (multi-categories)

2. Brier Score(BS, non-ranked), Brier Skill Score(BSS).
   from two categories to multi-categories/probabilistic
   ----measure both reliability and resolution



Prob. Evaluation (multi-categories)

3. Decomposition of Brier Score:
   consider ***sub-sample*** and ***overall-sample***
   reliability, resolution and uncertainty.
   for reliability: 0 is perfectly reliable
   for resolution: 0 is no resolution (= climatology)
   when resolution = reliability → no skill
   example of global ensemble:

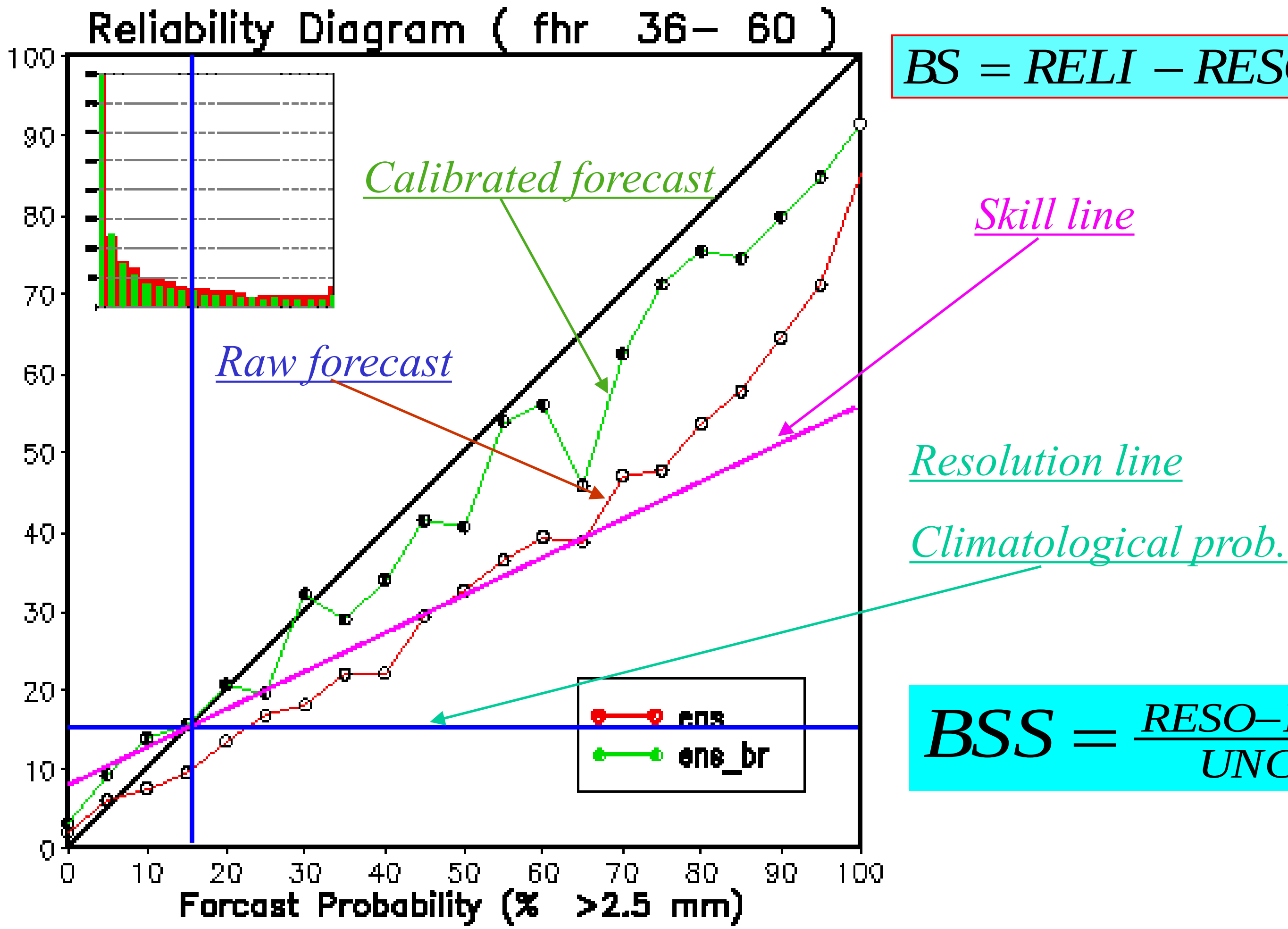


Prob. Evaluation (multi-categories)

4. Reliability and possible calibration (remove bias):

For period precipitation evaluation

$$BS = RELI - RESO + UNCE$$

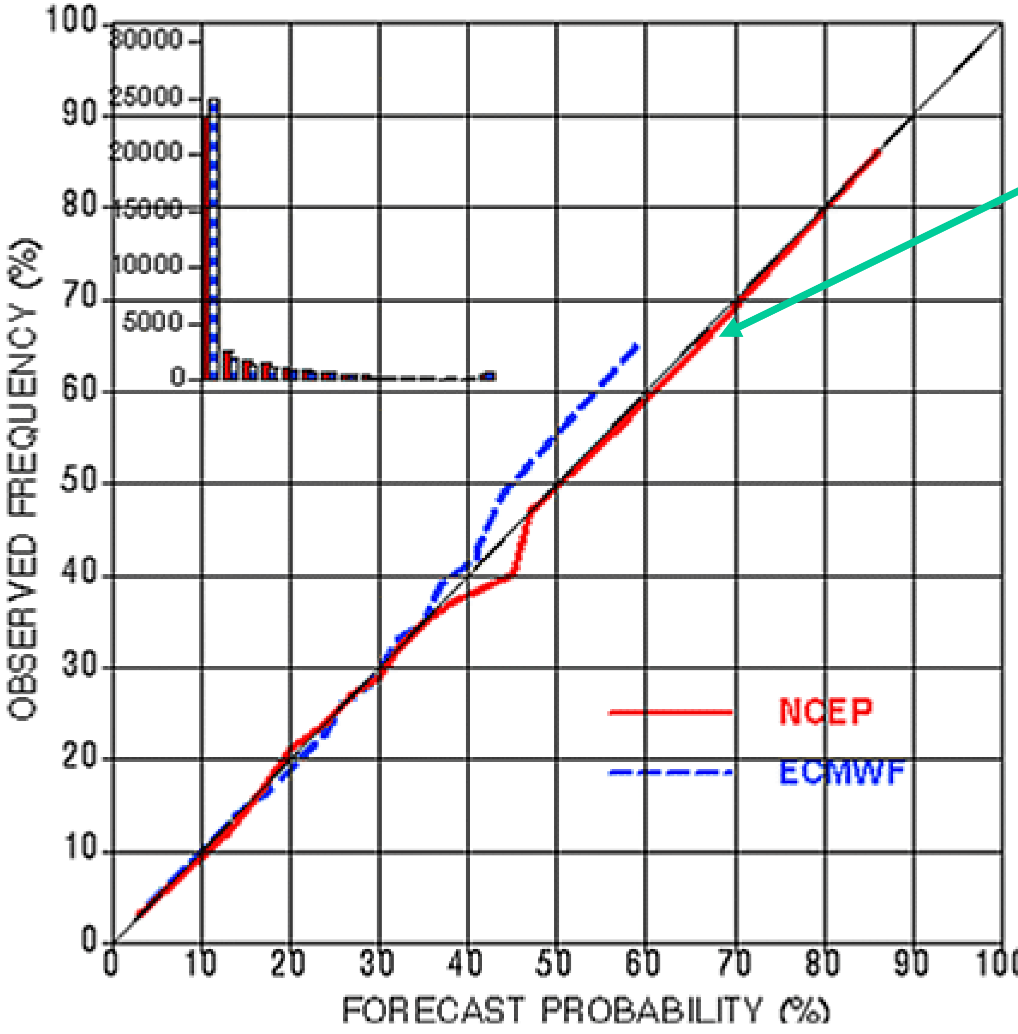


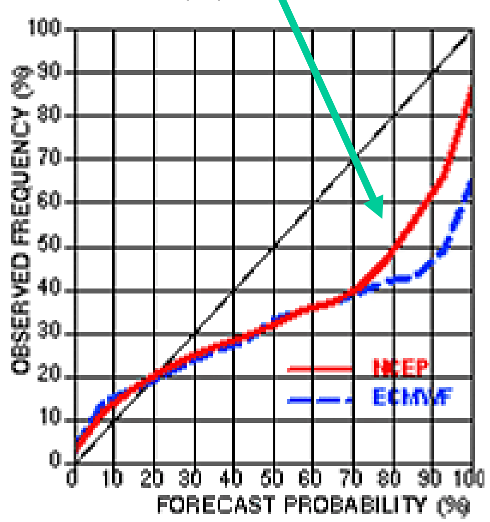
$$BSS = \frac{RESO - RELI}{UNCE}$$

Prob. Evaluation (multi-categories)

4. Reliability and possible probabilistic calibration:

re-label fcst prob by obs frequency associated with fcst





Prob. Evaluation (cost-loss analysis)

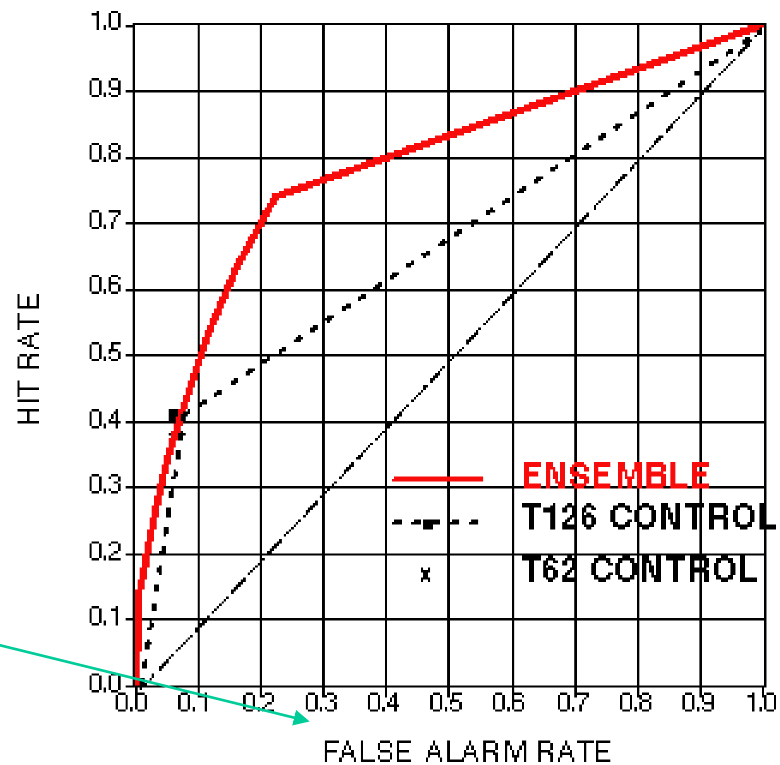
Based on hit rate (HR) and false alarm (FA) rate.

1. Relative Operating Characteristics (ROC) area - ***Appl. of signal detection theory for measuring discrimination between two alternative outcome.***

ROCarea = Intergrated area * 2 (0-1 normality)

h/(h+m)

f/(h+f)

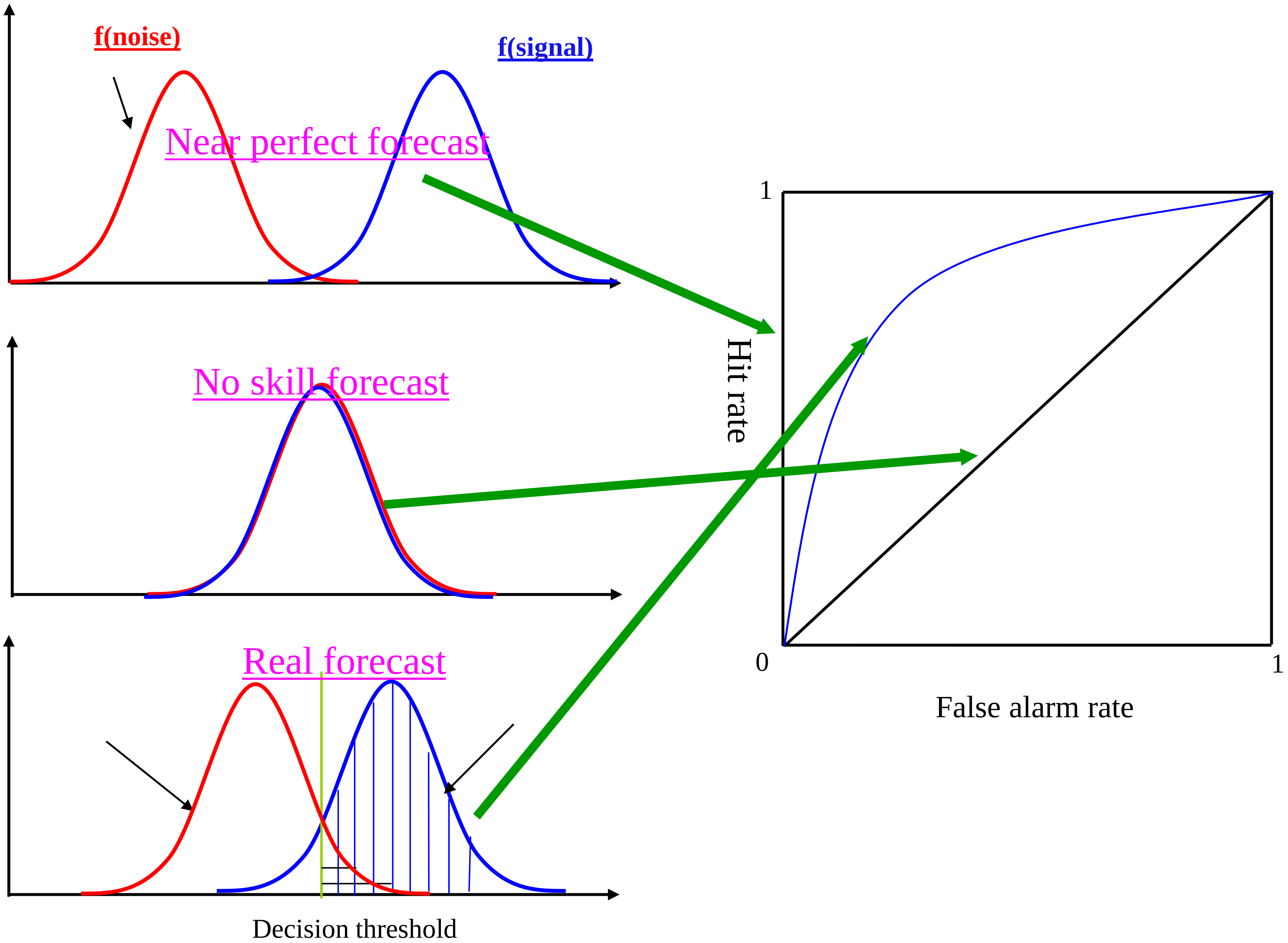


Relative Operating Characteristics

o\f	y(f)	n(f)
y(o)	h	m
n(o)	f	c

*ROC (Relative Operating Characteristics) curve for a 10-member T62 ensemble of forecasts and for T126 and T62 control forecasts for the 500 hPa height, **NH extratropics**, March-May 1997. The closer a curve is to the upper left hand corner, the more ability the forecasting system has in delineating between cases when a certain event (in this case, the occurence of one of 10 climatologically equally likely bins) did or did not occur.*

Relative Operating Characteristics area (ROC area)



USER NEEDS – PROBABILISTIC FORECAST INFORMATION FOR MAXIMUM ECONOMIC BENEFIT

ECONOMIC VALUE OF FORECASTS

Given a particular forecast, a user either does or does not take action (eg, protects its crop against frost) *Mylne & Harrison, 1999*

		FORECAST	
		YES	*NO*
OBSERVATION	YES	H(its) ***Mitigated Loss***	M(isses) ***Loss***
	NO	F(alse alarms) ***Cost***	C(orrect rejections) ***No Cost***

$$\textit{Mean Expense}_{fc} = hML + mL + fC$$

$$\textit{Mean Expense}_{perf} = oML$$

$$ME_{cl} = \min[oL,\ oML + (1-o)C]$$

$$Value = \frac{ME_{cl} - ME_{fc}}{ME_{cl} - ME_{perf}}$$

o=climatological frequency

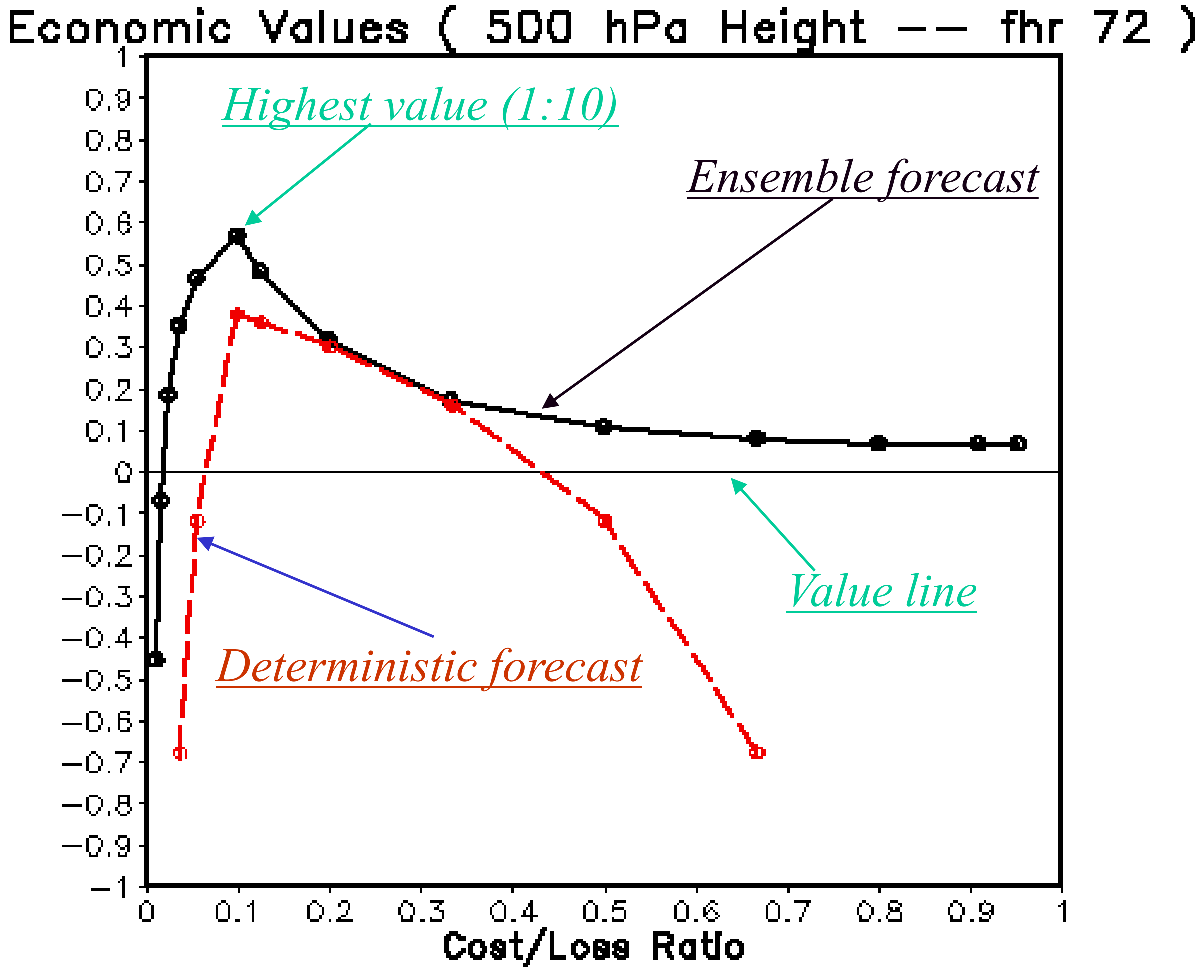
Optimum decision criterion for user action: P(weather event)=C/L

(Murphy 1977)

Prob. Evaluation (cost-loss analysis)

2. Economic Value (EV) of forecasts.

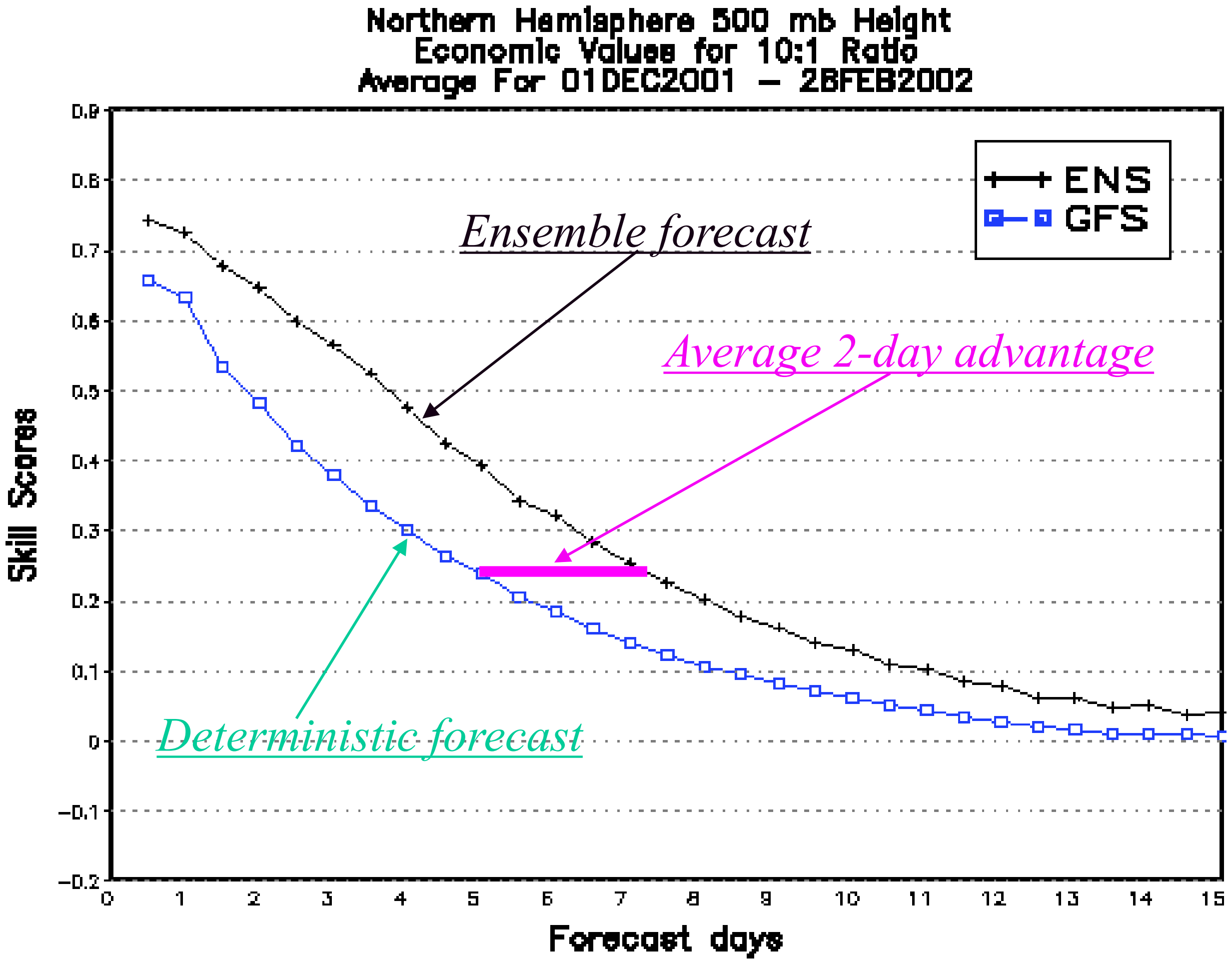
Given a particular forecast, a user either does or does not take action



Prob. Evaluation (cost-loss analysis)

Based on hit rate (HR) and false alarm (FA) analysis

.. Economic Value (EV) of forecasts



Decision Theory Example

Critical Event: sfc winds > 50kt

Cost (of protecting): $150K

Loss (if damage): $1M

Observed?	Forecast? YES	Forecast? NO
YES	*Hit* $150K	***Miss*** **$1000K**
NO	*False Alarm* $150K	*Correct Rejection* $0K

Case	Deterministic Forecast (kt)	Observation (kt)	Cost ($K)	Probabilistic Forecast	Cost ($K) by Threshold for Protective Action: 0%	20%	40%	60%	80%	100%
1	65	**54**	150	42%	150	150	150	**1000**	**1000**	**1000**
2	58	**63**	150	71%	150	150	150	150	**1000**	**1000**
3	73	**57**	150	95%	150	150	150	150	150	**1000**
4	55	37	150	13%	150	0	0	0	0	0
5	39	31	0	3%	150	0	0	0	0	0
6	31	**55**	**1000**	36%	150	150	**1000**	**1000**	**1000**	**1000**
7	62	**71**	150	85%	150	150	150	150	150	**1000**
8	53	42	150	22%	150	150	0	0	0	0
9	21	27	0	51%	150	150	150	0	0	0
10	52	39	150	77%	150	150	150	150	0	0
		Total Cost:	$ 2,050		$1,500	$1,200	$1,900	$2,600	$3,300	$5,000

Optimal Threshold = 15%