Evaluation of Ensemble Forecast

- Performance metrics

Yuejian Zhu

EMC/NCEP/NOAA

5th NAEFS workshop, Mexican City

May 18 2010

Ensemble forecast metrics

- NCEP GEFS
  - NCEP GFS based ensemble forecast system
  - Half resolution of GFS forecast
  - ETR initial perturbation method
- NAEFS
  - Combine NCEP/GEFS, CMC/GEFS (current)
  - Adding FNMOC/GEFS (by Q4FY2010 or Q1FY2011)
  - About 60 members for 00UTC and 12UTC
  - Output every 6 hours out to 16 days
  - Exchange about 80 variables
  - Post products
    - Bias correction (49 variables)
    - Downscaling (8 variables)

NCEP/GEFS current evaluations (1)

- Variables:
  - 500hPa and 1000hPa heights
  - 850hPa and 2-meter temperatures
  - 10-meter U and V
  - 850hPa and 200hPa U and V
- Domains:
  - Globally, NH (20N-80N), SH (20S-80S) and TROP (20N-20S)
  - NA (Northern American), EU (Europe) and AS (Asia)
- Resolution:
  - Based on 2.5*2.5 degree resolution
- Frequency:
  - Every 12/24 hours, out to 16 days
- Against:
  - Analysis (or best analysis), and observations (possible for future)
- Verifying:
  - PAC, RMS errors, ABS errors for ensemble mean
  - Ensemble spread and histogram
  - BS (in terms of reliability and resolution)
  - RPSS and CRPS
  - ROC and EV (economic values)
- Skills:
  - Use NCEP/NCAR 40y reanalysis as references

Example of score cards for ensembles evaluation

NAEFSb (40 members) vs NAEFSb+FNMOCb (56 members): NH-Z500 in Spring 2009

Days	1	2	3	4	5	6	7	8	9	10
AC										
CRPS										
Rel										
Res										

NAEFSb (40 members) vs NAEFSb+FNMOCb (56 members): NH-T850 in Spring 2009

Days	1	2	3	4	5	6	7	8	9	10
AC										
CRPS										
Rel										
Res										

NAEFSb (40 members) vs NAEFSb+FNMOCb (56 members): NH-Z1000 in Spring 2009

Days	1	2	3	4	5	6	7	8	9	10
AC										
CRPS										
Rel										
Res										

NAEFSb (40 members) vs NAEFSb+FNMOCb (56 members): NH-T2M in Spring 2009

Days	1	2	3	4	5	6	7	8	9	10
AC										
CRPS										
Rel										
Res										

NAEFSb (40 members) vs NAEFSb+FNMOCb (56 members): NH-U10M in Spring 2009

Days	1	2	3	4	5	6	7	8	9	10
AC										
CRPS										
Rel										
Res										

NAEFSb (40 members) vs NAEFSb+FNMOCb (56 members): NH-V10M in Spring 2009

Days	1	2	3	4	5	6	7	8	9	10
AC										
CRPS										
Rel										
Res										

- Using 95% confidence interval (2.5%-97.5%), BLUE means NAEFSb+FNMOCb is significantly better than NAEFSb, RED means otherwise.
- The reliability (Rel) and resolution (Res) are from Brier Score decomposition.

NAEFSb (40 members) vs NAEFSb+FNMOCb (56 members): NH-Z500 in Winter 0809

Days	1	2	3	4	5	6	7	8	9	10
AC										
CRPS										
Rel										
Res										

NAEFSb (40 members) vs NAEFSb+FNMOCb (56 members): NH-T850 in Winter 0809

Days	1	2	3	4	5	6	7	8	9	10
AC										
CRPS										
Rel										
Res										

NAEFSb (40 members) vs NAEFSb+FNMOCb (56 members): NH-Z1000 in Winter 0809

Days	1	2	3	4	5	6	7	8	9	10
AC										
CRPS										
Rel										
Res										

NAEFSb (40 members) vs NAEFSb+FNMOCb (56 members): NH-T2M in Winter 0809

Days	1	2	3	4	5	6	7	8	9	10
AC										
CRPS										
Rel										
Res										

NAEFSb (40 members) vs NAEFSb+FNMOCb (56 members): NH-U10M in Winter 0809

Days	1	2	3	4	5	6	7	8	9	10
AC										
CRPS										
Rel										
Res										

NAEFSb (40 members) vs NAEFSb+FNMOCb (56 members): NH-V10M in Winter 0809

Days	1	2	3	4	5	6	7	8	9	10
AC										
CRPS										
Rel										
Res										

- Using 95% confidence interval (2.5%-97.5%), BLUE means NAEFSb+FNMOCb is significantly better than NAEFSb, RED means otherwise.
- The reliability (Rel) and resolution (Res) are from Brier Score decomposition.

NCEP/GEFS current evaluations (2)

- Precipitation
- CONUS only
- Based on 1*1 degree resolution
- Every 24 hours, out to 15 days
  - From 1200UTC – 1200UTC
- Verify against: observations (Gauges)
  - About 8000-10000 reports for every 24 hours
- Verifying statistics for:
  - RMS errors, ABS errors for ensemble mean
  - Ensemble spread
  - CRPS (and reliability)
  - ETS, TSS and Bias for ensemble mean
    - different thresholds (0.01, 0.2, 2.0, 5.0, 10.0, 15.0, 25.0, 35.0 and 50.0 mm/days)

NCEP/GEFS current evaluations (3)

- Tropical storm tracks (or storm tracks)
- Domains:
  - Globally, Atlantic basin, East Pacific basin and West Pacific Basin
- Based on model output – surface pressure and other variables at 1.1 degree
- Out to 7 days (will be extended to 10 days)
- Against: observed position (best position)
- Verifying statistics for:
  - Track errors for ensemble mean
  - Ensemble spread
- Plotting:
  - Case by case for individual ensemble member
  - Statistics for case average
  - Statistics for seasonal average

NCEP GEFS Probabilistic Evaluation Metrics (drafted plan)

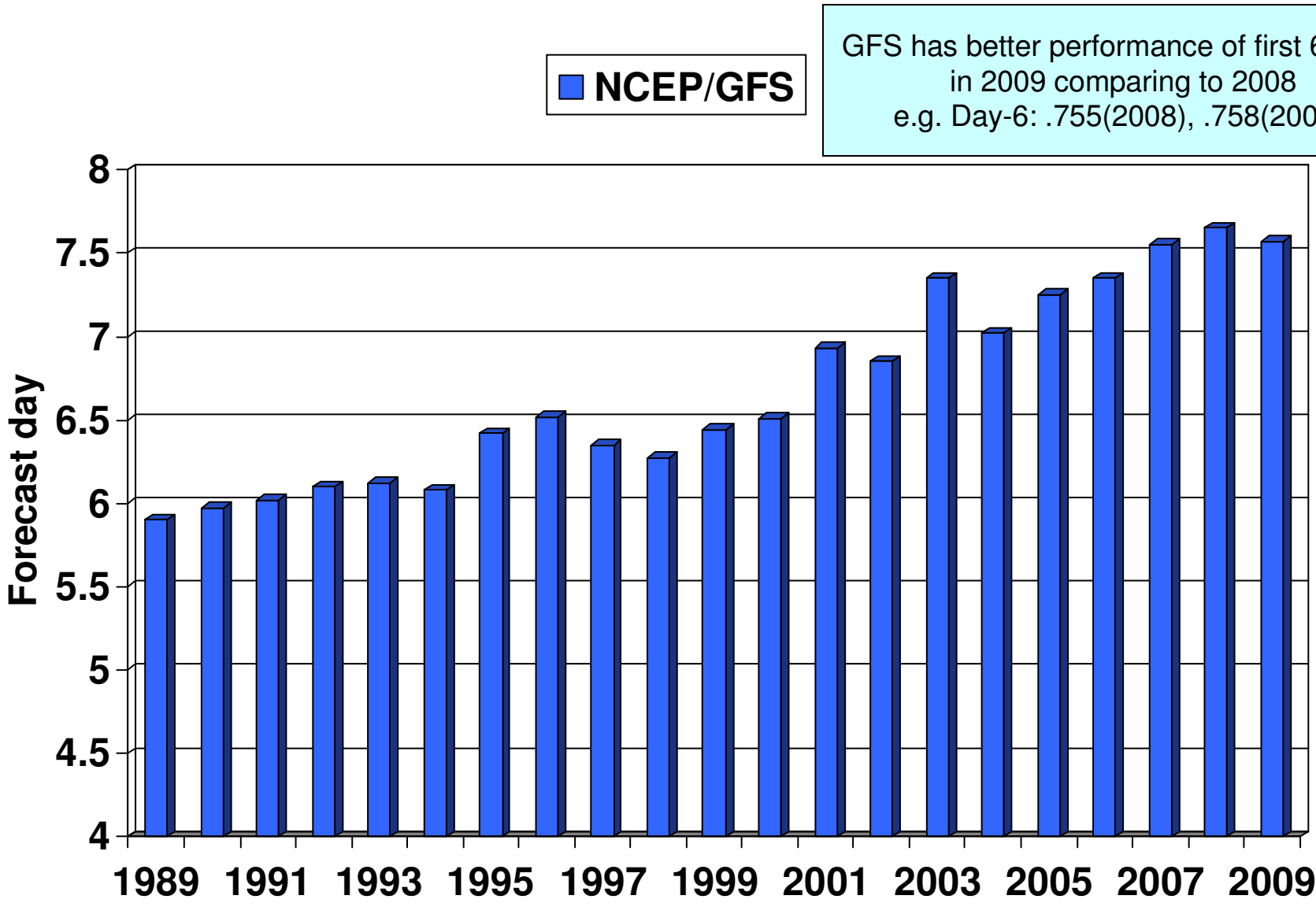
Fields	Levels	Area	Variable	Measures	Lead-time	Index
Grid	500hPa	NH	Height	AC/RMS/sprd	3,5,8 days	3
Grid	500hPa	NH	Height	Skillful fcst	Days	1
Grid	500hPa	NH	Height	CRPS	3,5,8 days	3
Grid	1000hPa	NH	Height	CRPS	3,5,8 days	3
Grid	850hPa	NH	Temperature	CRPS	3,5,8 days	3
Grid/OBS	Surface	NH/Tropical	Pressure/winds	Track Error	1,3,5 days	3
Grid	850hPa	Tropical	Winds	CRPS	3,5 days	2
Grid	200hPa	Tropical	Winds	CRPS	3,5 days	2
Grid/OBS	Surface	CONUS	Precipitation	ETS/CRPS	1,3,5 days	3
Grid	Surface	NH	Temperature	CRPS	3 days	1
Grid	Surface	NH	Winds	CRPS	3 days	1

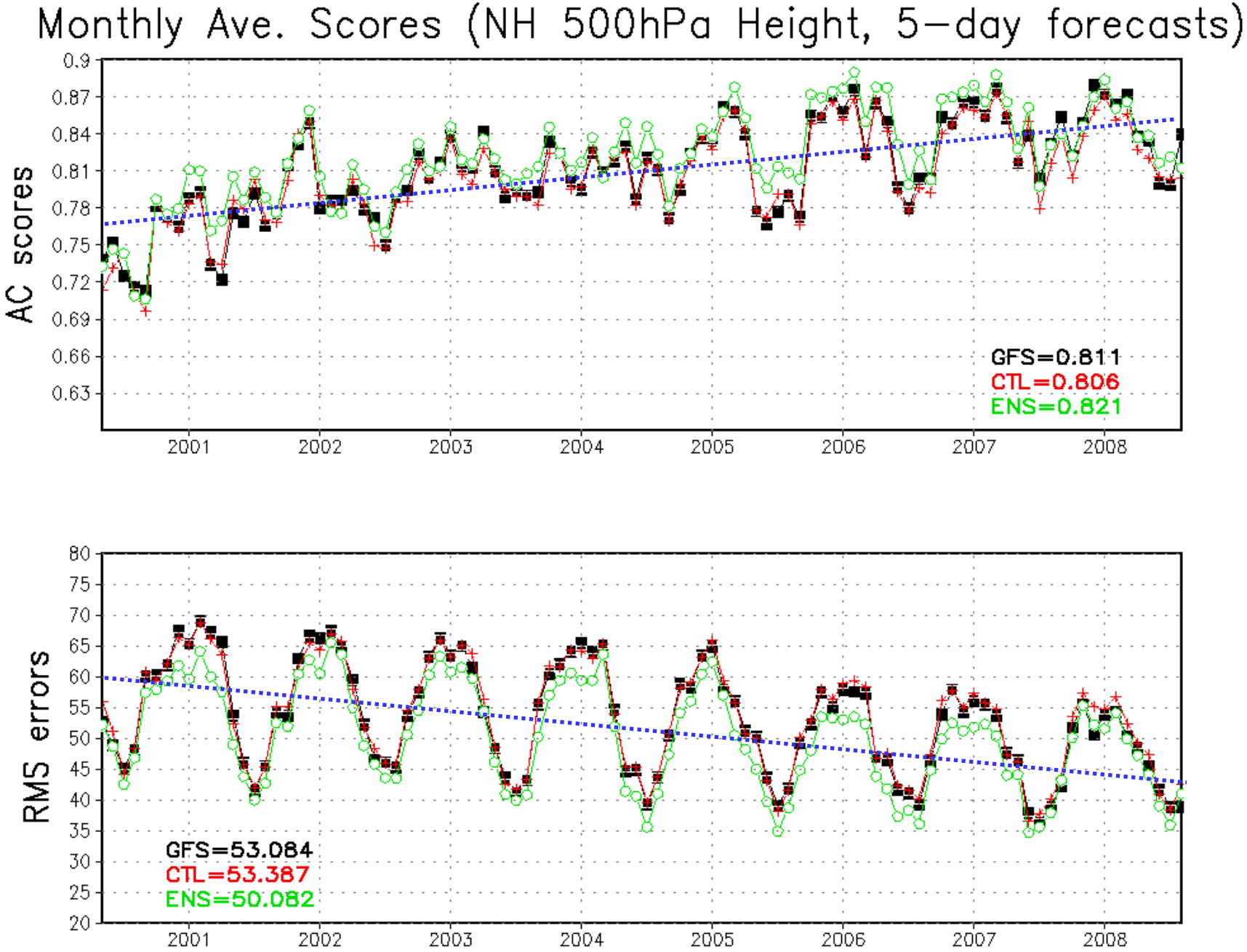
GFS and GEFS performance

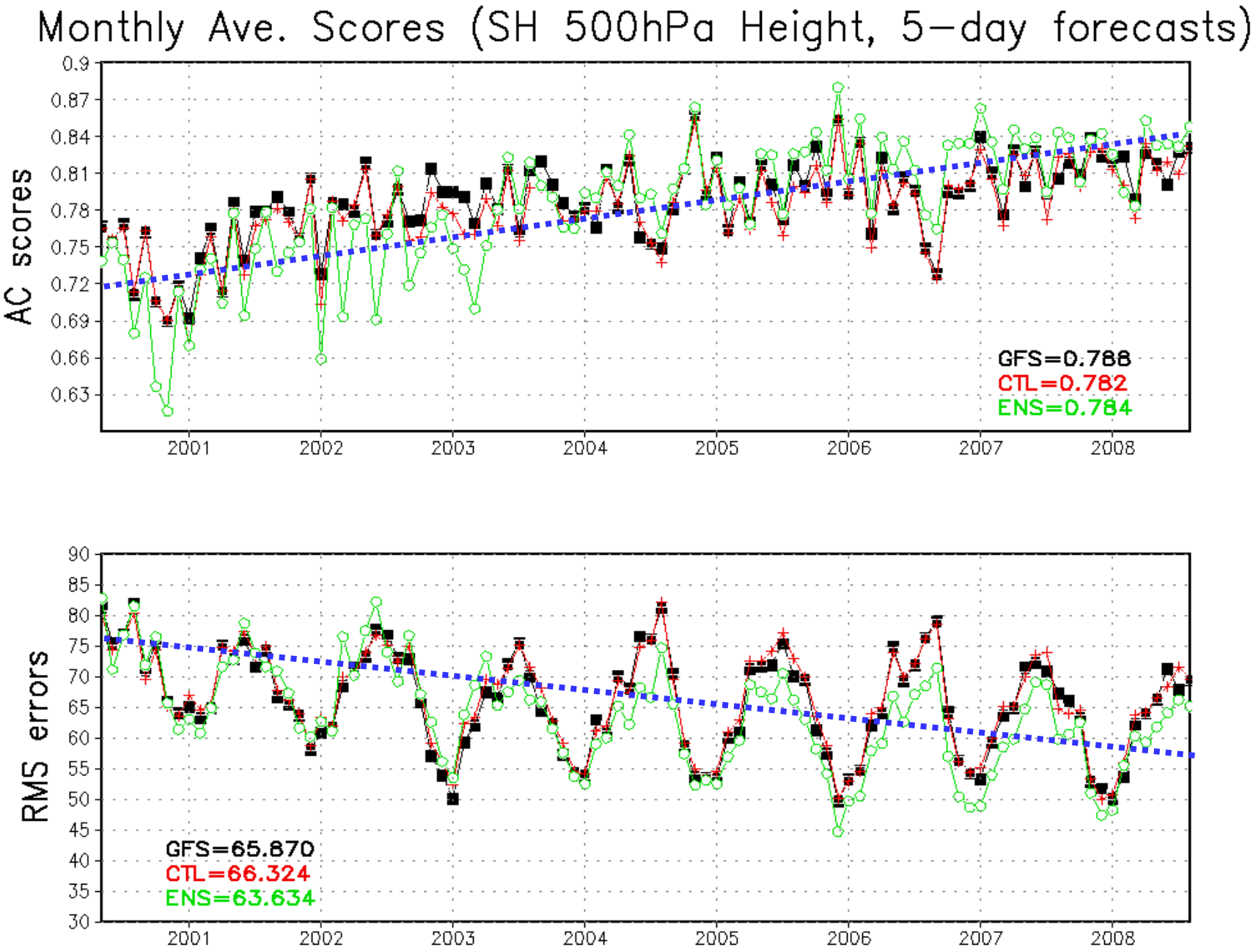
- Skillful forecast

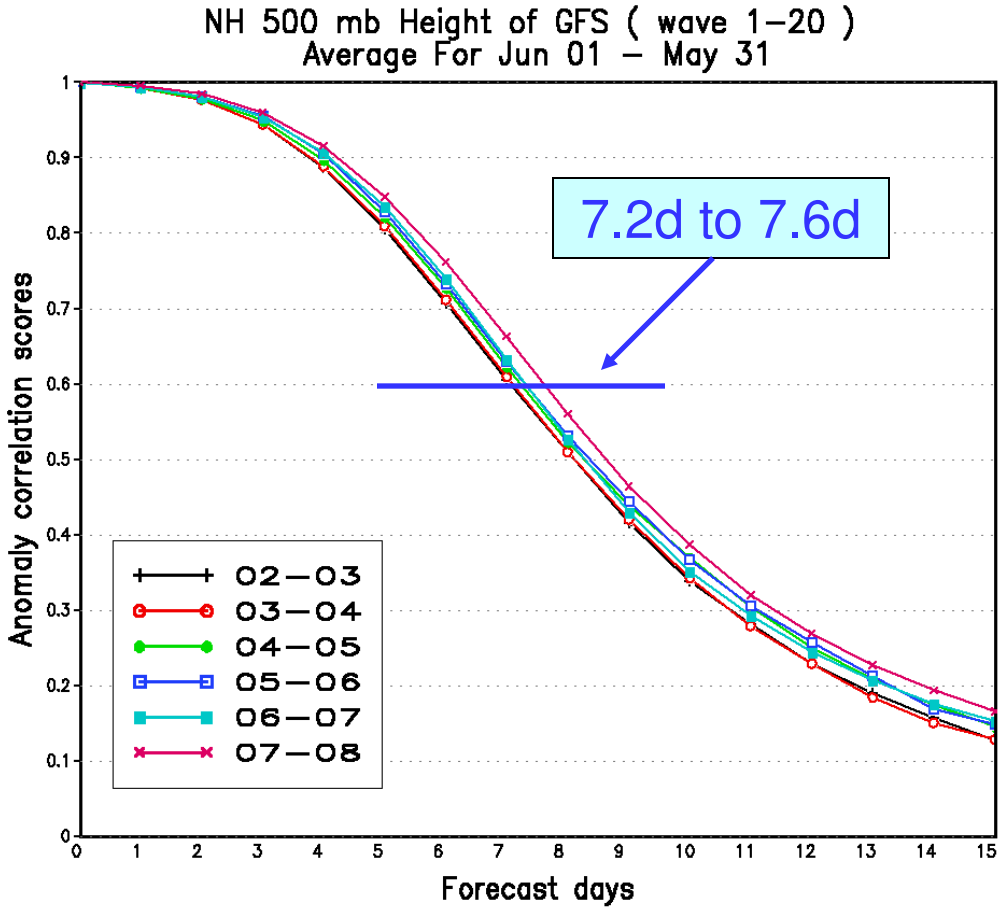
Day at which fcst loses useful skill (AC=0.6)
N. Hemisphere calendar year means

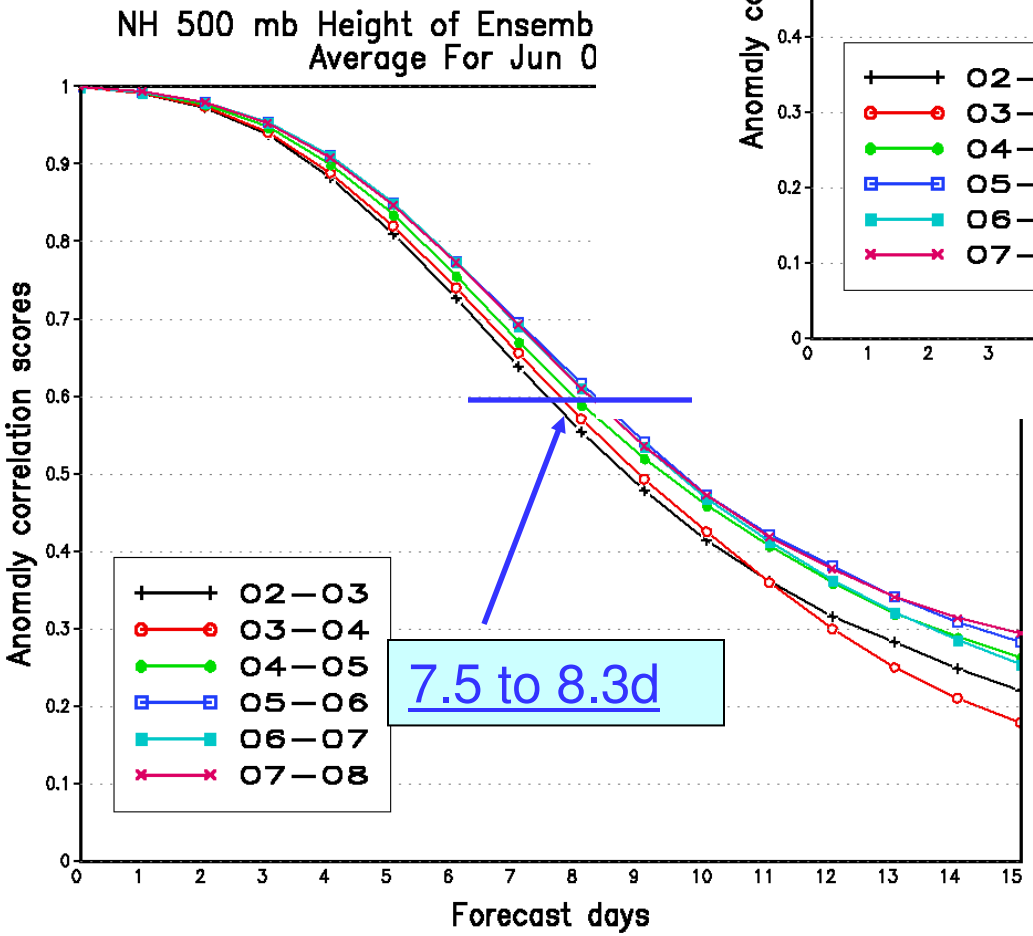
GFS has better performance of first 6 days in 2009 comparing to 2008
e.g. Day-6: .755(2008), .758(2009)

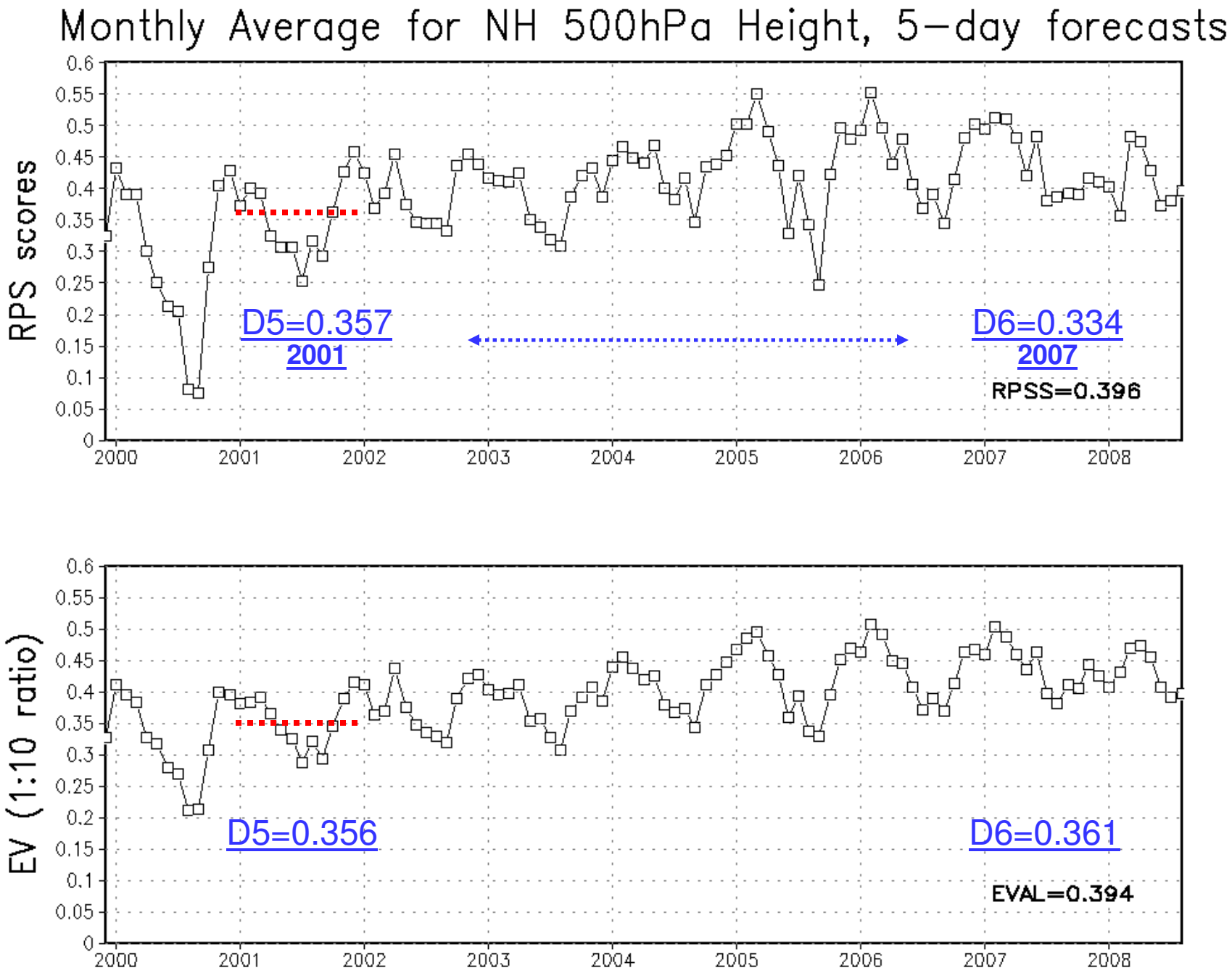



Monthly Ave. Scores (NH 500hPa Height, 5-day forecasts)
AC scores
0.9
0.87
0.84
0.81
0.78
0.75
0.72
0.69
0.66
0.63
GFS=0.811
CTL=0.806
ENS=0.821
2001
2002
2003
2004
2005
2006
2007
2008
RMS errors
80
75
70
65
60
55
50
45
40
35
30
25
20
GFS=53.084
CTL=53.387
ENS=50.082
2001
2002
2003
2004
2005
2006
2007
2008


Monthly Ave. Scores (SH 500hPa Height, 5-day forecasts)
AC scores
0.9
0.87
0.84
0.81
0.78
0.75
0.72
0.69
0.66
0.63
GFS=0.788
CTL=0.782
ENS=0.784
2001
2002
2003
2004
2005
2006
2007
2008
RMS errors
90
85
80
75
70
65
60
55
50
45
40
35
30
GFS=65.870
CTL=66.324
ENS=63.634
2001
2002
2003
2004
2005
2006
2007
2008


NH 500 mb Height of GFS (wave 1-20)
Average For Jun 01 - May 31
Anomaly correlation scores
Forecast days
02-03
03-04
04-05
05-06
06-07
07-08
7.2d to 7.6d


NH 500 mb Height of Ensemb
Average For Jun 0
Anomaly correlation scores
Forecast days
02-03
03-04
04-05
05-06
06-07
07-08
7.5 to 8.3d

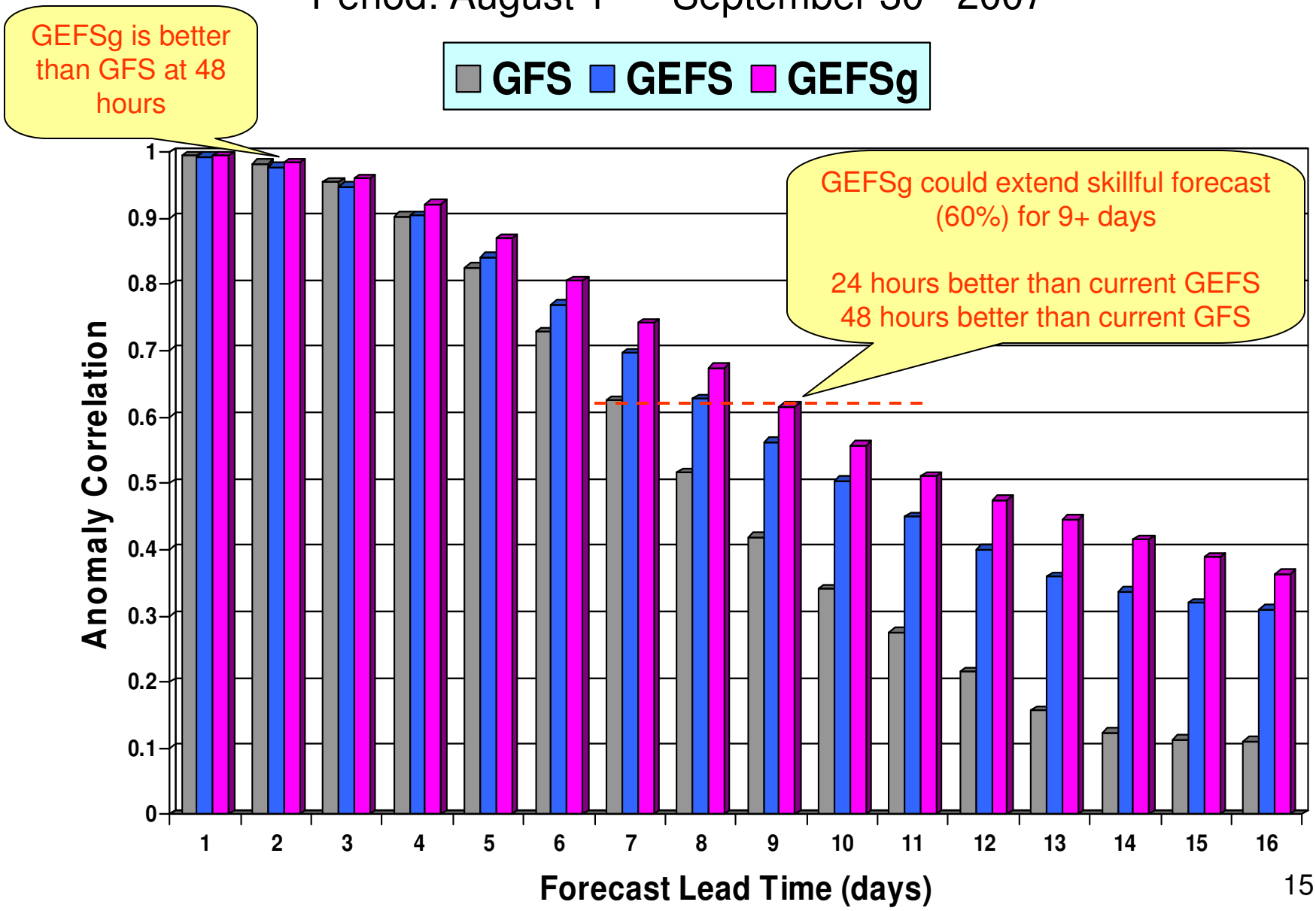

Monthly Average for NH 500hPa Height, 5-day forecasts
RPS scores
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
2000
2001
2002
2003
2004
2005
2006
2007
2008
D5=0.357
2001
D6=0.334
2007
RPSS=0.396
EV (1:10 ratio)
D5=0.356
D6=0.361
EVAL=0.394

General evaluations for ensemble mean

- Comparison for implementation

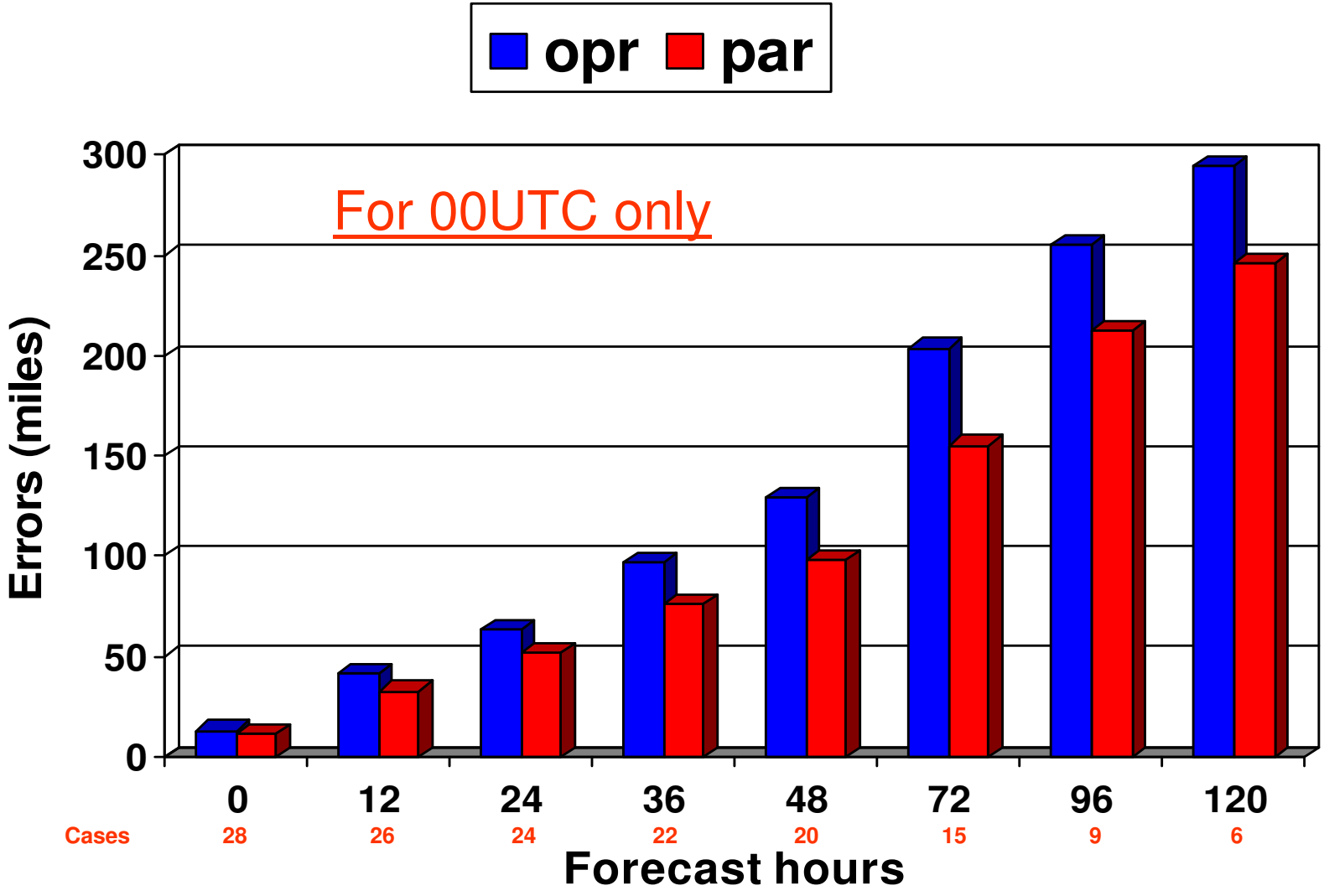
NH Anomaly Correlation for 500hPa Height

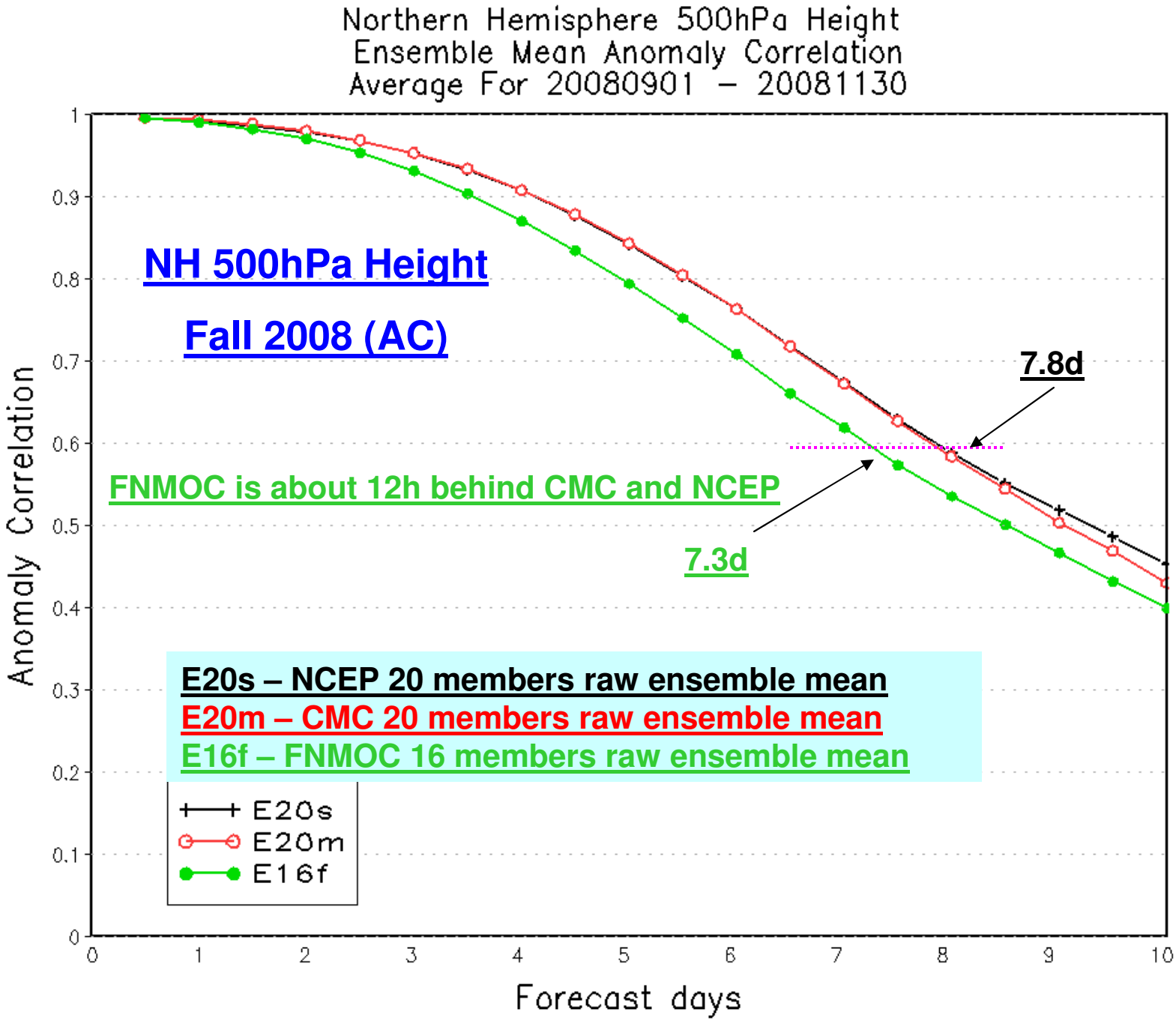
Period: August 1st – September 30th 2007

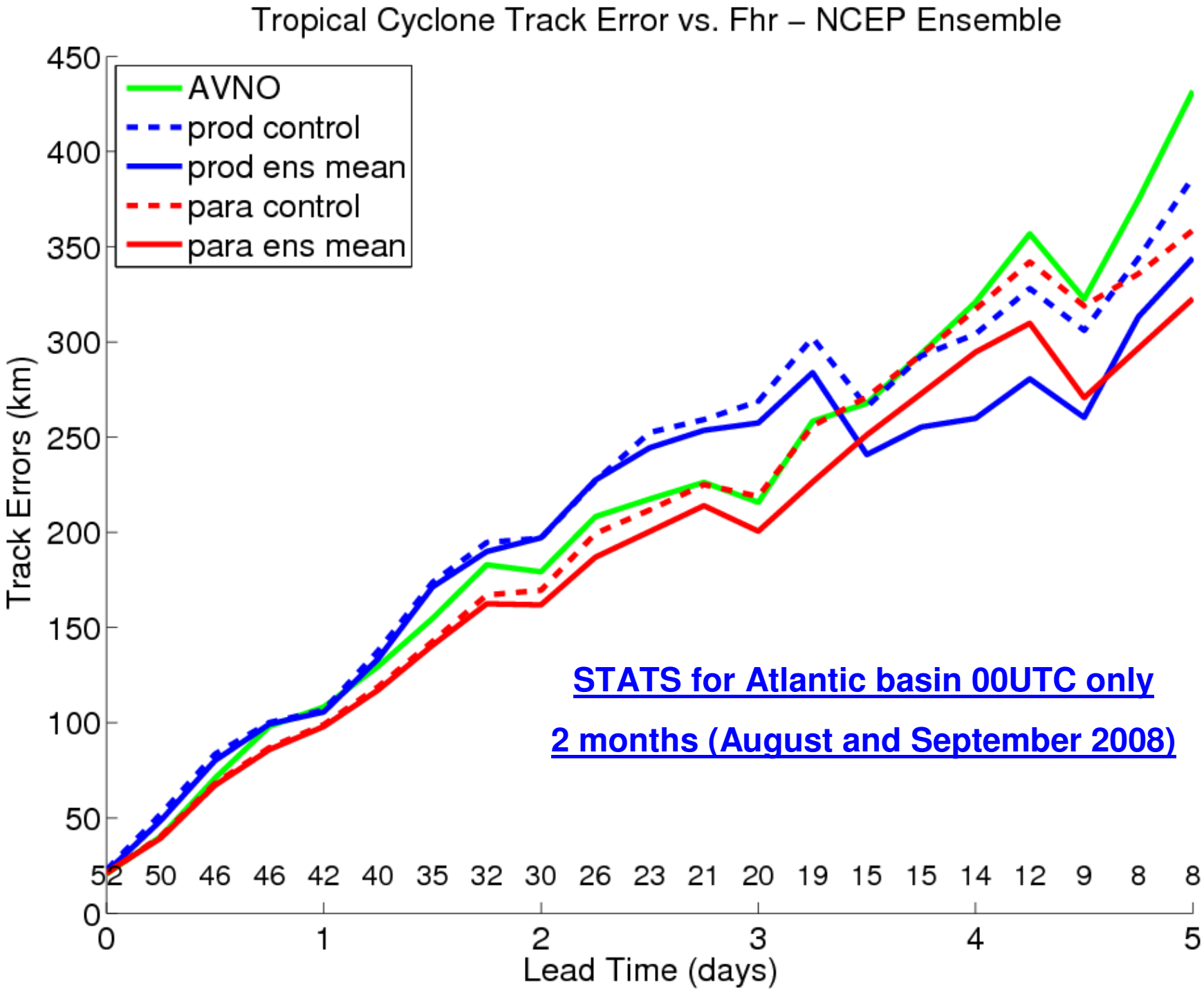


Summary of the important cases of Bill Jimena, Rick and Ida

TS track errors (2009)

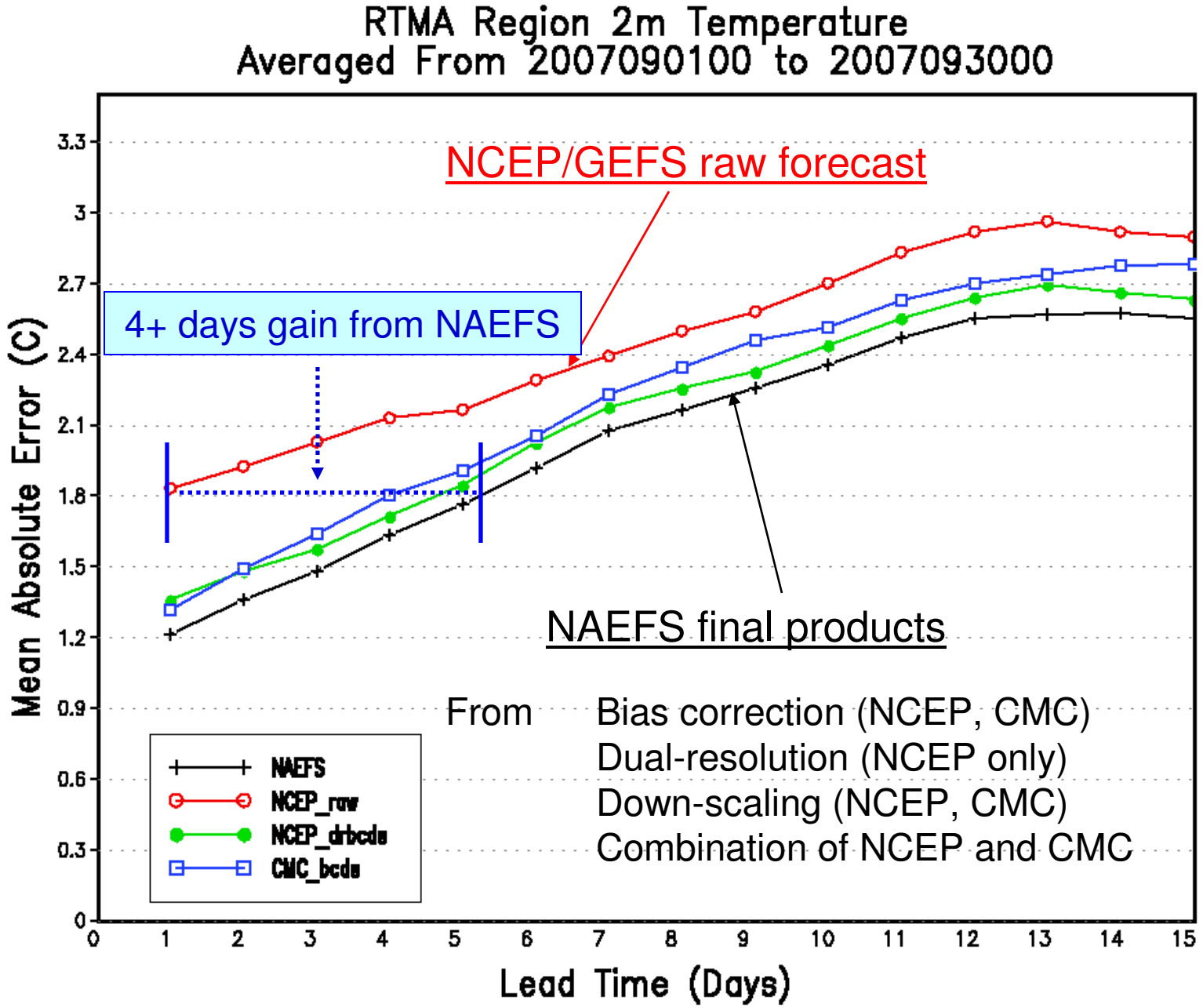



Northern Hemisphere 500hPa Height
Ensemble Mean Anomaly Correlation
Average For 20080901 – 20081130
NH 500hPa Height
Fall 2008 (AC)
7.8d
FNMOC is about 12h behind CMC and NCEP
7.3d
E20s – NCEP 20 members raw ensemble mean
E20m – CMC 20 members raw ensemble mean
E16f – FNMOC 16 members raw ensemble mean
E20s
E20m
E16f
Anomaly Correlation
Forecast days
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
1
2
3
4
5
6
7
8
9
10

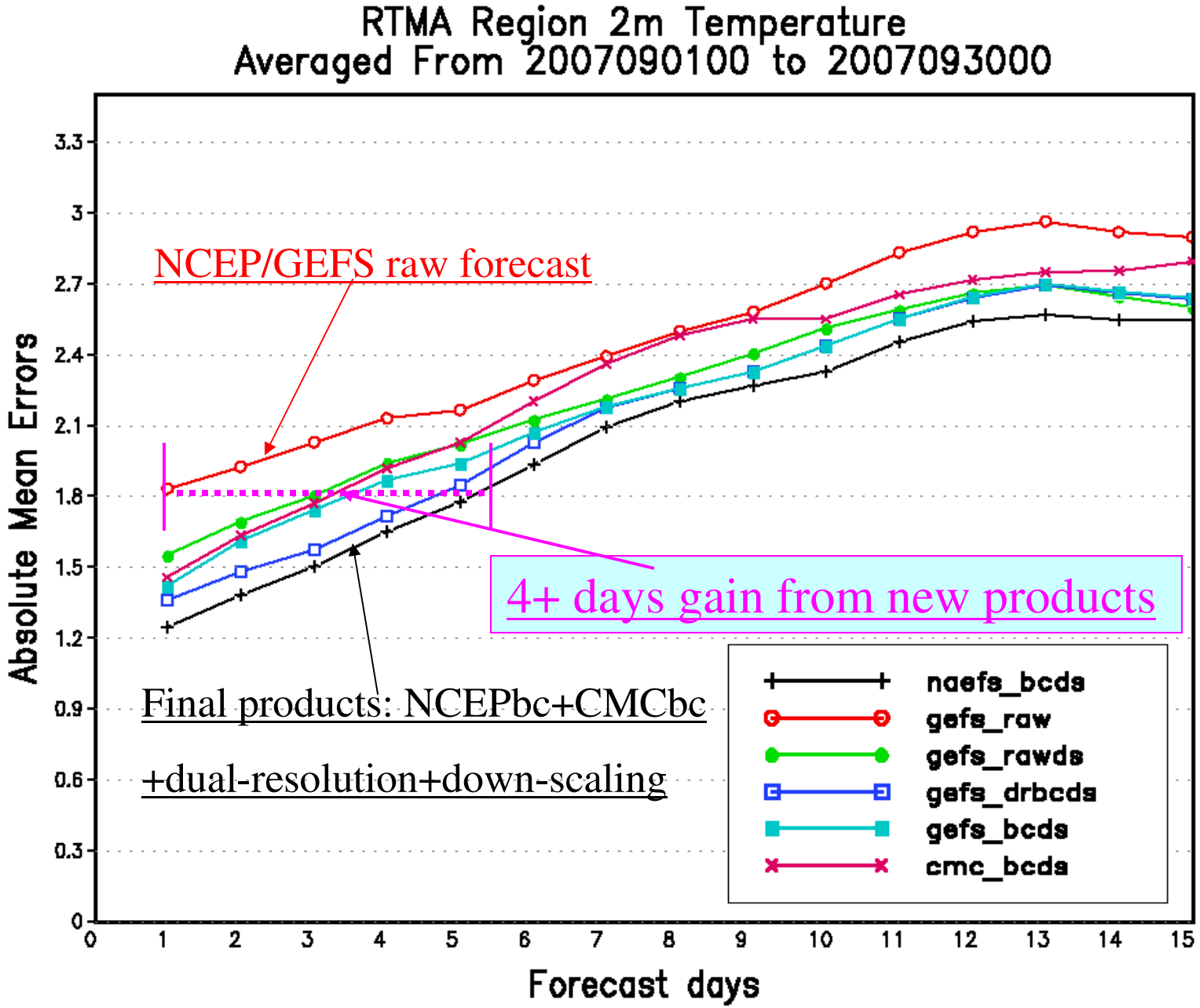

Tropical Cyclone Track Error vs. Fhr – NCEP Ensemble
AVNO
prod control
prod ens mean
para control
para ens mean
Track Errors (km)
450
400
350
300
250
200
150
100
50
0
52 50 46 46 42 40 35 32 30 26 23 21 20 19 15 15 14 12 9 8 8
0
1
2
3
4
5
Lead Time (days)
STATS for Atlantic basin 00UTC only
2 months (August and September 2008)

Evaluations for bias correction and downscaling

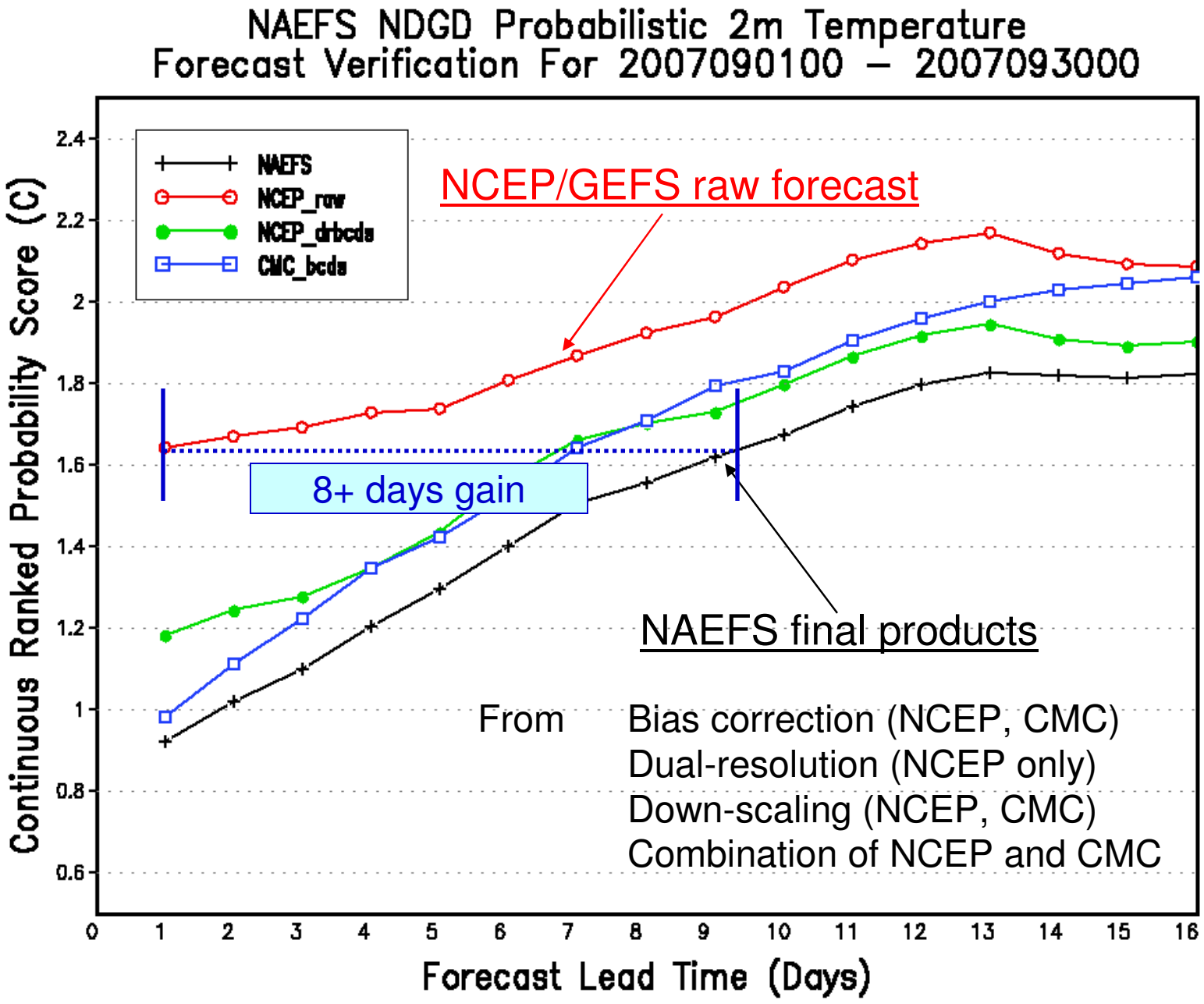
- Demonstrate the values added



BO CUI, GCWMB/EMC/NCEP/NOAA

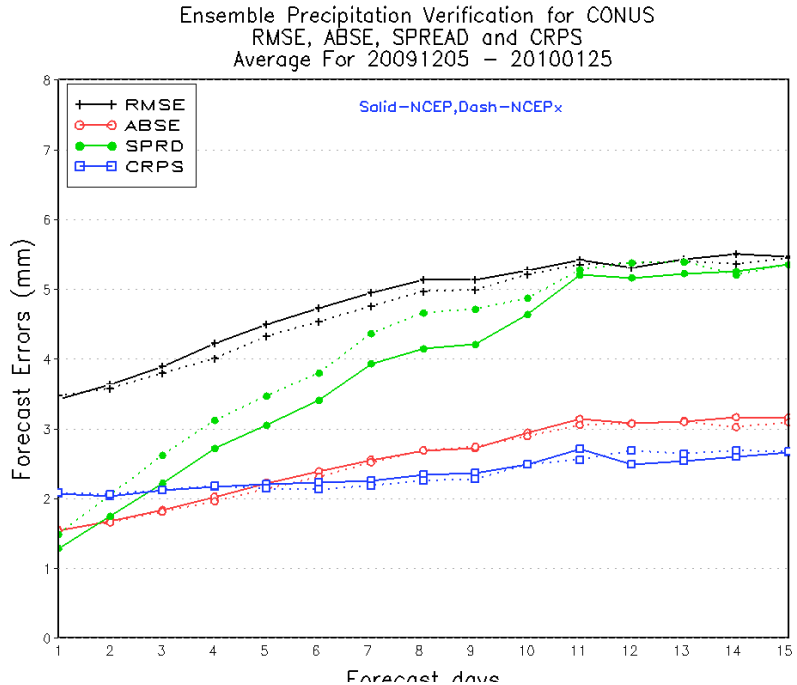


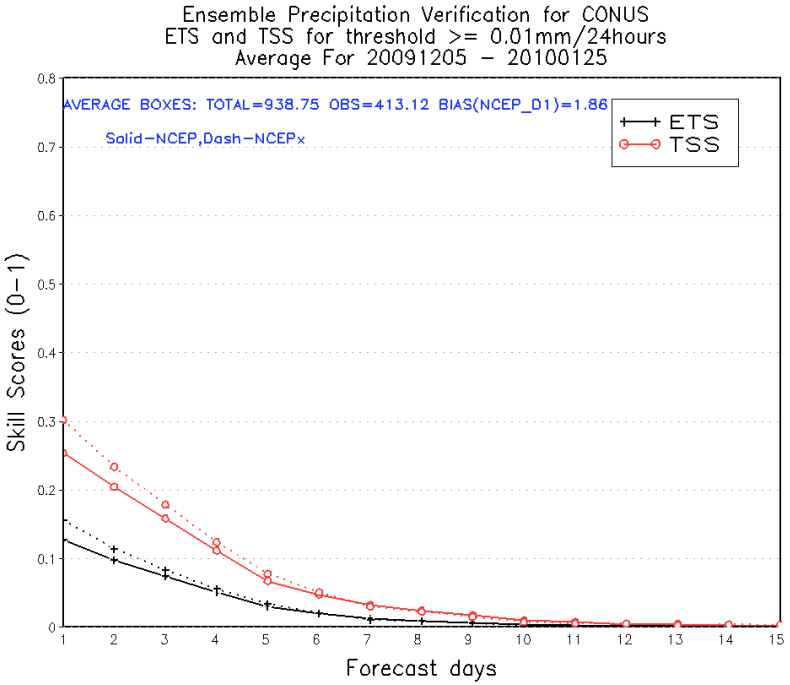
BO CUI, GCWMB/EMC/NCEP/NOAA

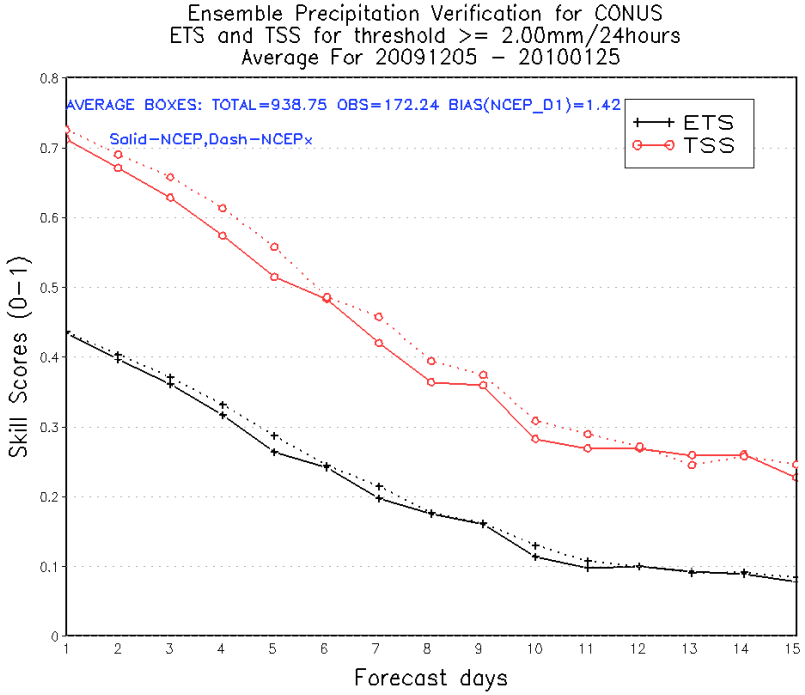

NAEFS NDGD Probabilistic 2m Temperature
Forecast Verification For 2007090100 – 2007093000
NAEFS
NCEP_raw
NCEP_drbcds
CMC_bcds
NCEP/GEFS raw forecast
8+ days gain
NAEFS final products
From
Bias correction (NCEP, CMC)
Dual-resolution (NCEP only)
Down-scaling (NCEP, CMC)
Combination of NCEP and CMC
Continuous Ranked Probability Score (C)
Forecast Lead Time (Days)
BO CUI, GCWMB/EMC/NCEP/NOAA

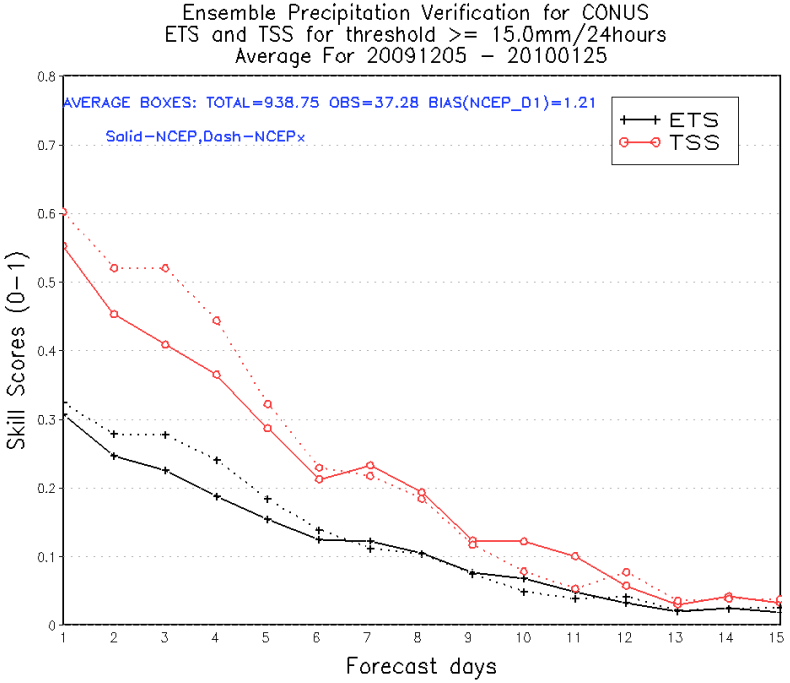
General evaluations for precipitation

- Ensemble mean and distribution


Ensemble Precipitation Verification for CONUS
RMSE, ABSE, SPREAD and CRPS
Average For 20091205 - 20100125
Solid-NCEP,Dash-NCEPx
RMSE
ABSE
SPRD
CRPS
Forecast Errors (mm)
Forecast days


Ensemble Precipitation Verification for CONUS
ETS and TSS for threshold >= 0.01mm/24hours
Average For 20091205 - 20100125
AVERAGE BOXES: TOTAL=938.75 OBS=413.12 BIAS(NCEP_D1)=1.86
Solid-NCEP,Dash-NCEPx
ETS
TSS
Skill Scores (0-1)
Forecast days


Ensemble Precipitation Verification for CONUS
ETS and TSS for threshold >= 2.00mm/24hours
Average For 20091205 - 20100125
AVERAGE BOXES: TOTAL=938.75 OBS=172.24 BIAS(NCEP_D1)=1.42
Solid-NCEP,Dash-NCEPx
ETS
TSS
Skill Scores (0-1)
Forecast days

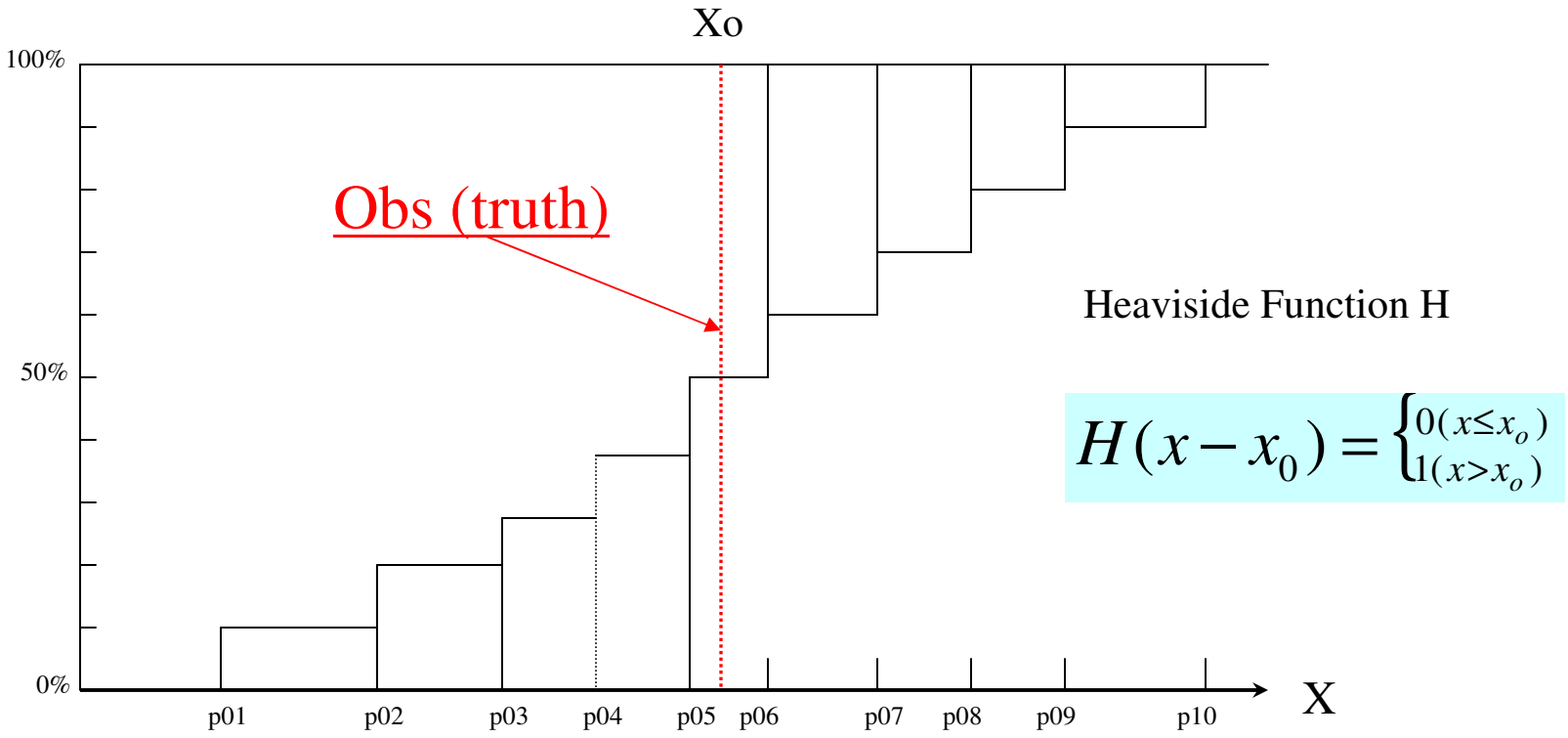

Ensemble Precipitation Verification for CONUS
ETS and TSS for threshold >= 15.0mm/24hours
Average For 20091205 - 20100125
AVERAGE BOXES: TOTAL=938.75 OBS=37.28 BIAS(NCEP_D1)=1.21
Solid-NCEP,Dash-NCEPx
ETS
TSS
Skill Scores (0-1)
Forecast days

Continuous Rank Probability Score

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

CRP Skill Score is

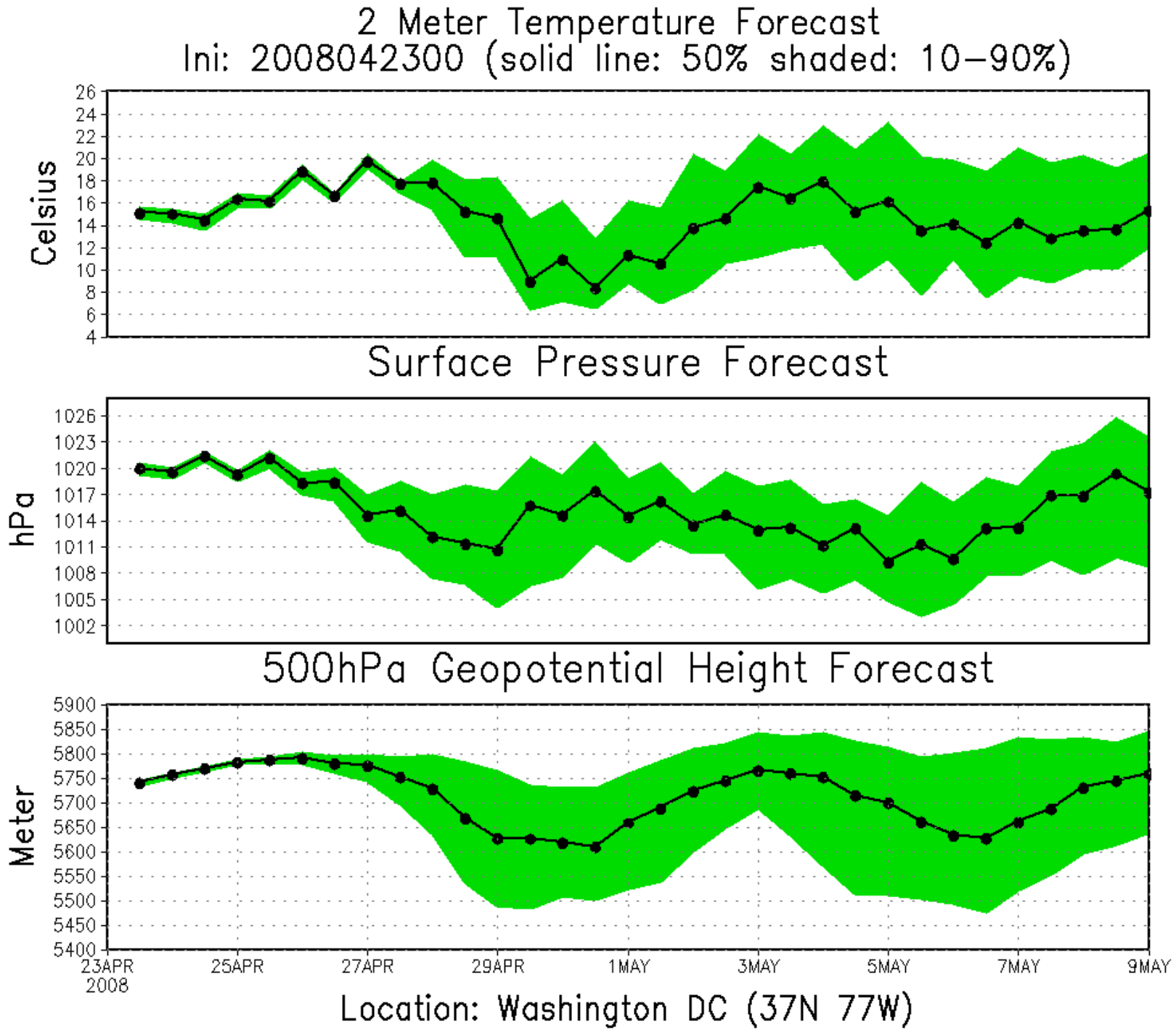
$$CRPSS = \frac{CRPS_c - CRPS_f}{CRPS_c}$$



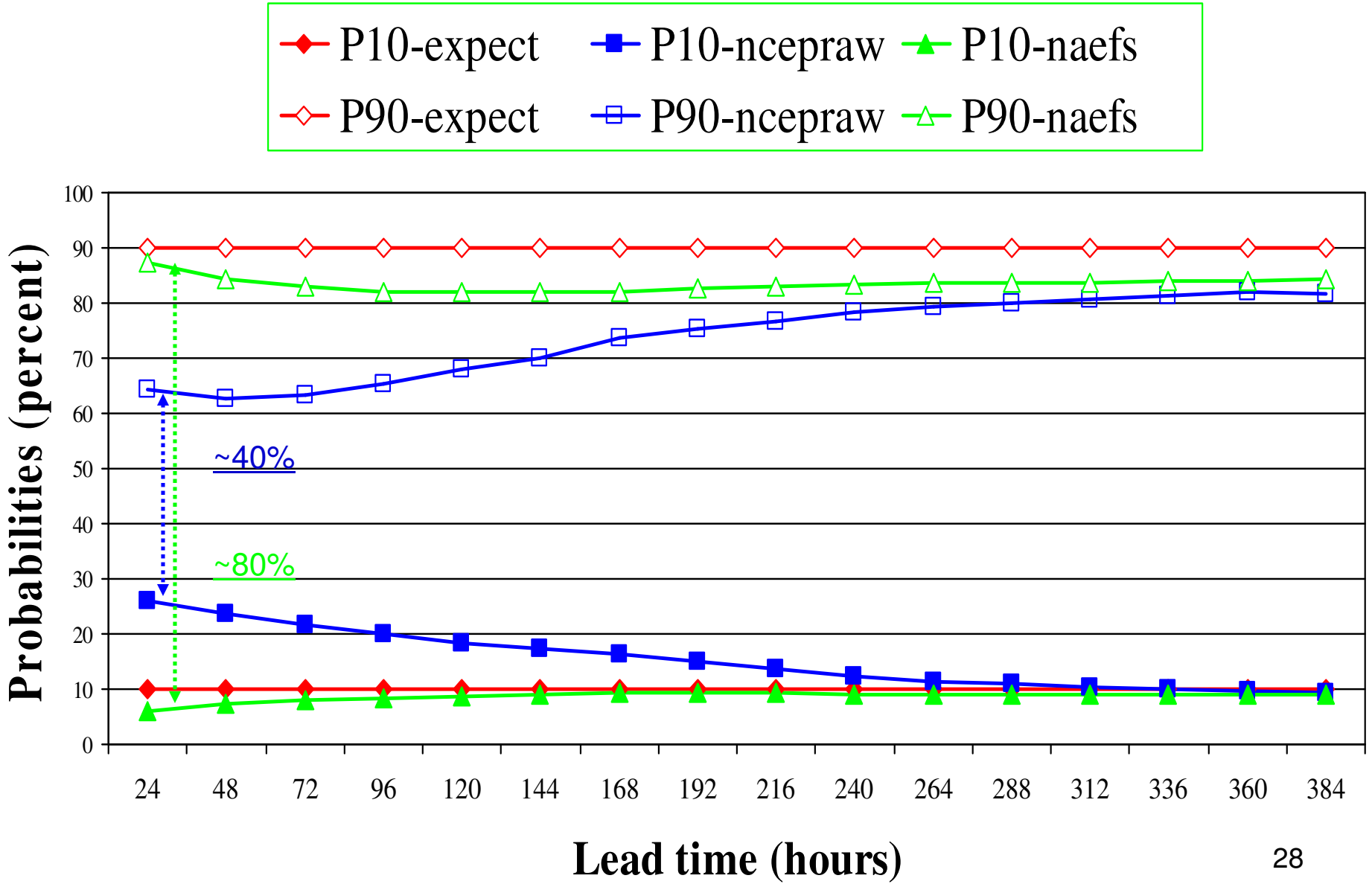
Order of 10 ensemble members (p01, p02,…,p10)

Evaluation for probabilistic forecast

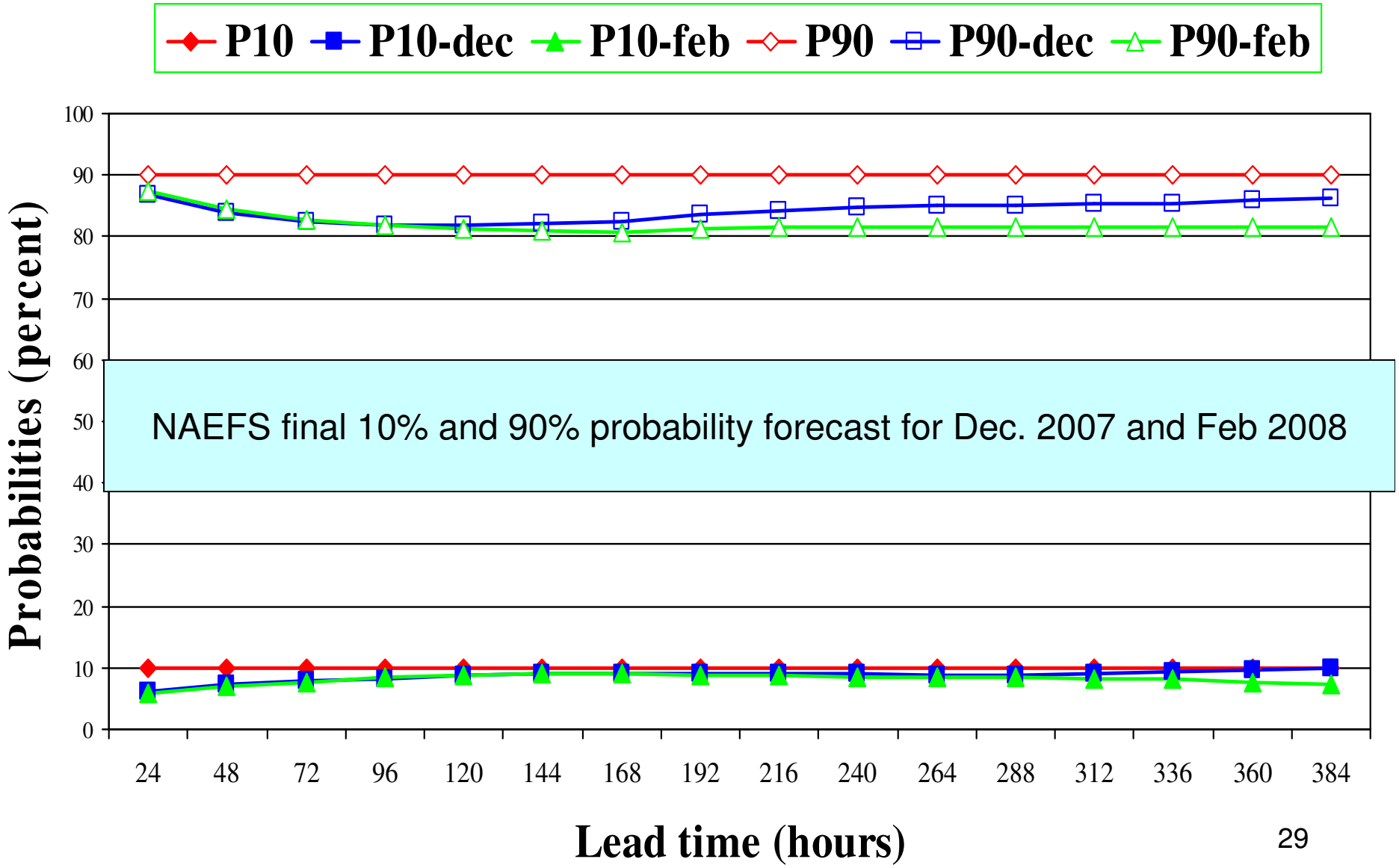
- Still working on ……


2 Meter Temperature Forecast
Ini: 2008042300 (solid line: 50% shaded: 10-90%)
Celsius
26
24
22
20
18
16
14
12
10
8
6
4
Surface Pressure Forecast
hPa
1026
1023
1020
1017
1014
1011
1008
1005
1002
500hPa Geopotential Height Forecast
Meter
5900
5850
5800
5750
5700
5650
5600
5550
5500
5450
5400
23APR
2008
25APR
27APR
29APR
1MAY
3MAY
5MAY
7MAY
9MAY
Location: Washington DC (37N 77W)

2-meter temperature 10/90 probability forecast verification
Northern Hemisphere, period of Dec. 2007 – Feb. 2008



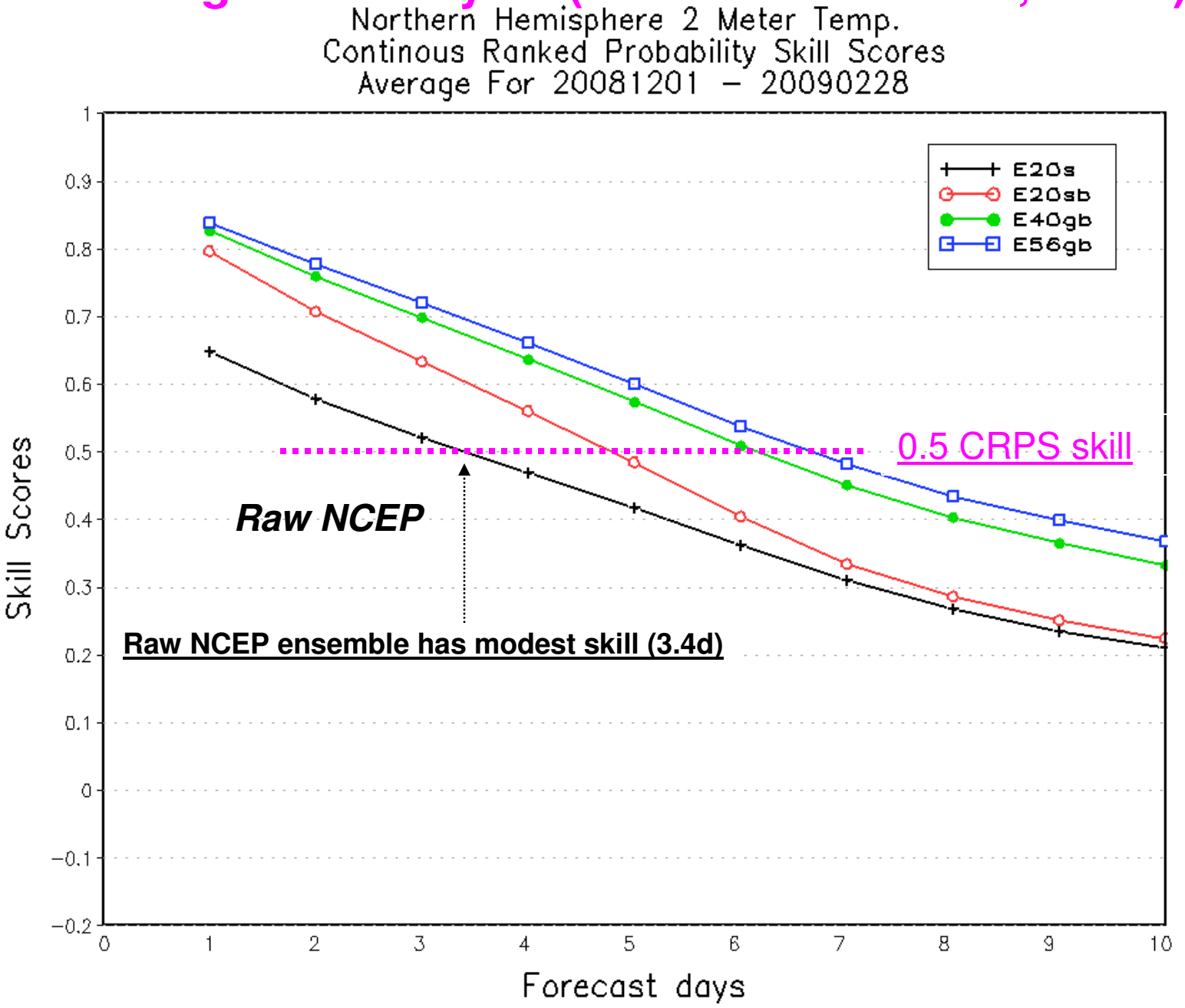
2-meter temperature 10/90 probability forecast verification Northern Hemisphere, seasonal variation for NAEFS



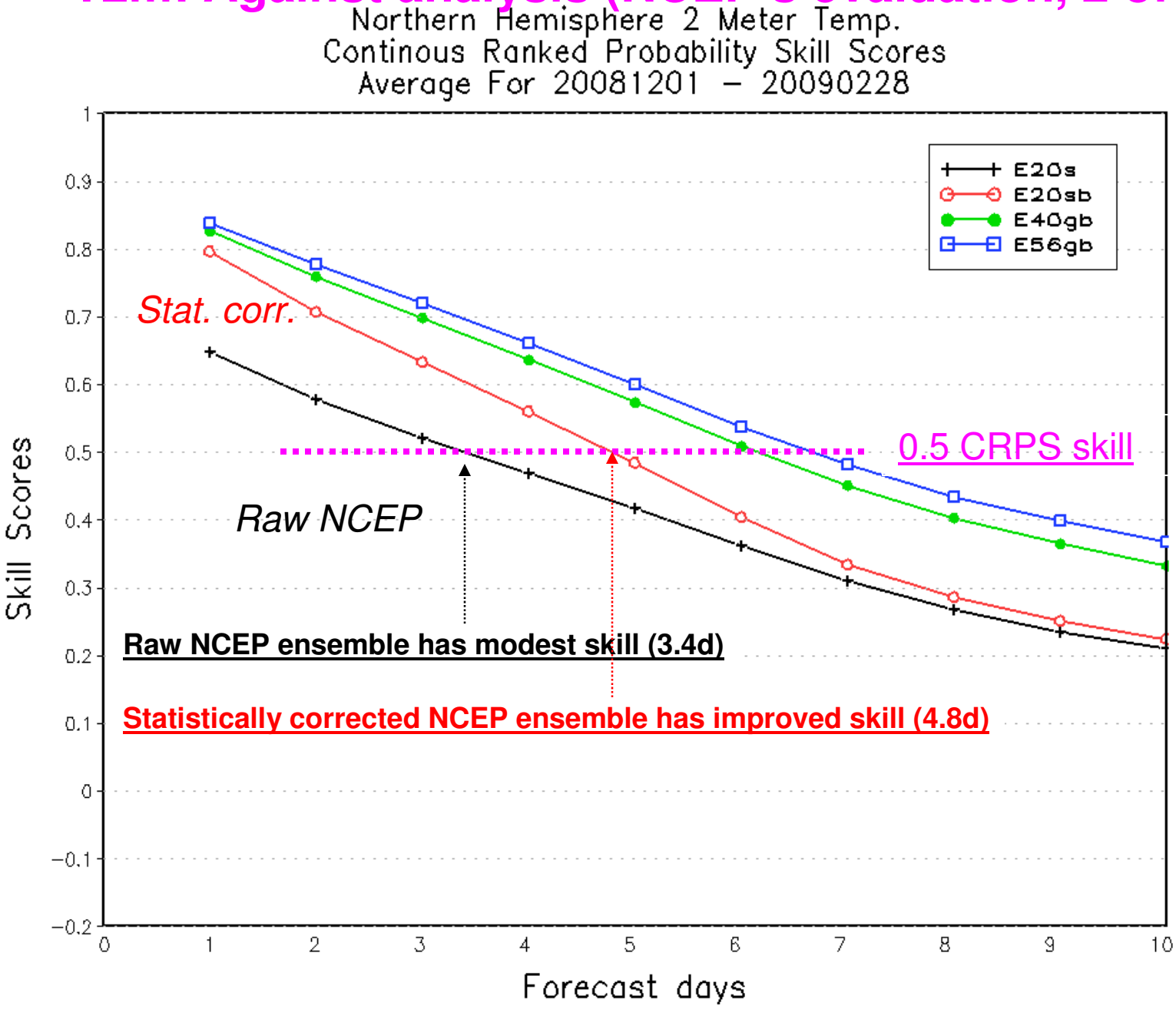
Values from adding additional ensembles

- For decision making

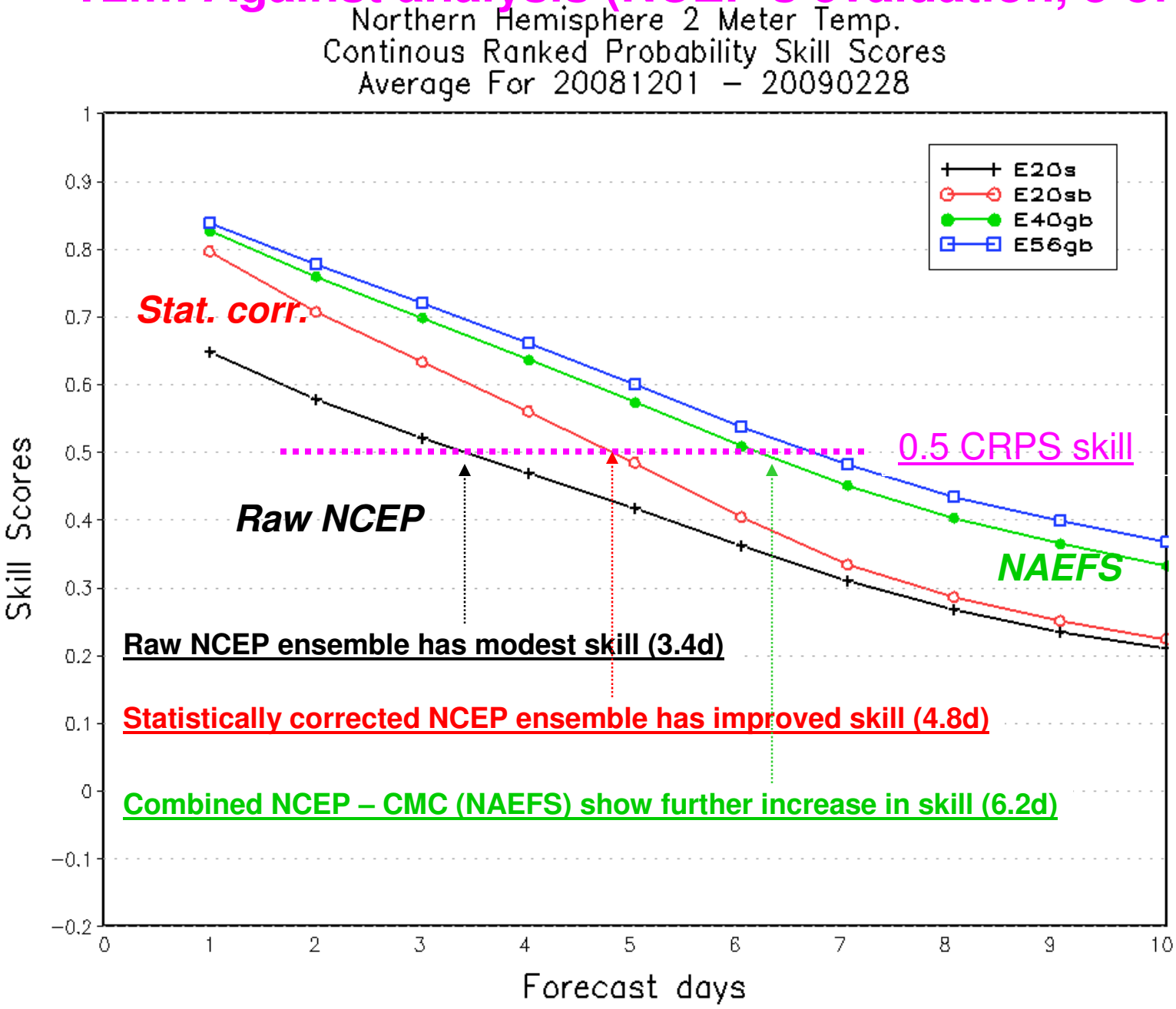
Value-added by including FNMOC ensemble into NAEFS
T2m: Against analysis (NCEP's evaluation, 1 of 4)



Value-added by including FNMOC ensemble into NAEFS
T2m: Against analysis (NCEP's evaluation, 2 of 4)



Value-added by including FNMOC ensemble into NAEFS
T2m: Against analysis (NCEP's evaluation, 3 of 4)



Value-added by including FNMOC ensemble into NAEFS
T2m: Against analysis (NCEP's evaluation, 4 of 4)