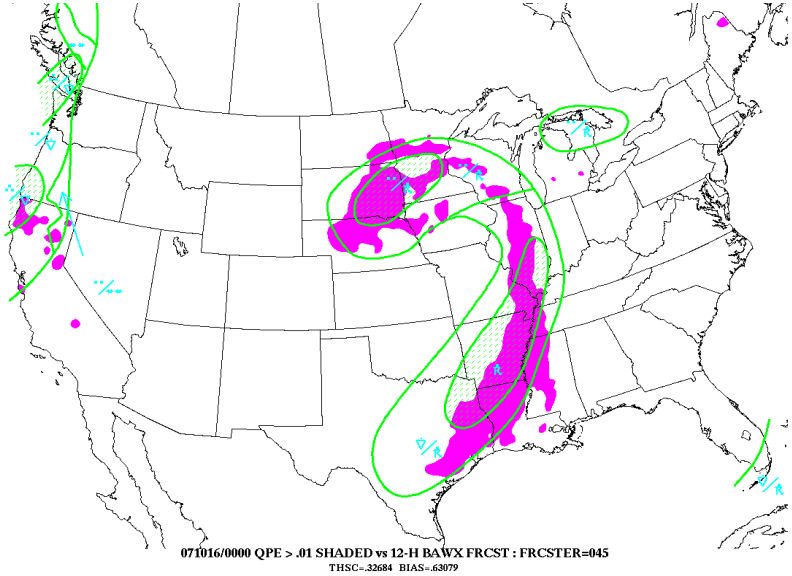
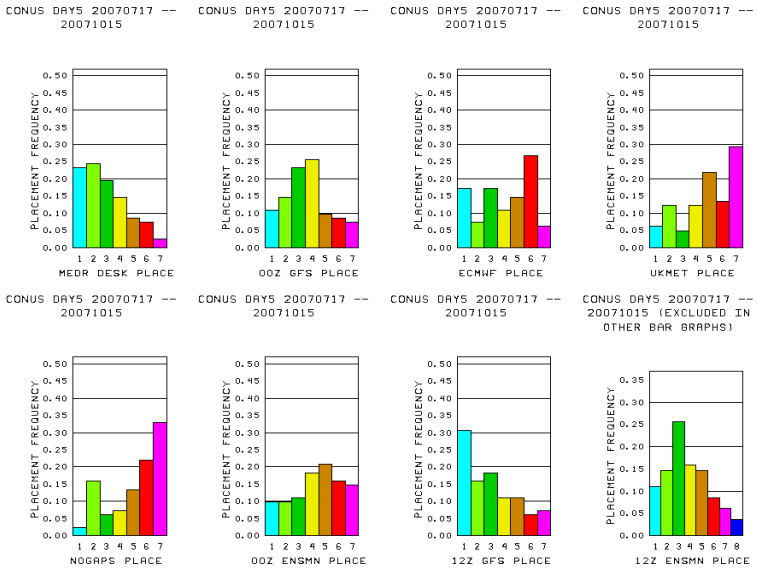
Verification of Weather Forecasts



Keith F. Brill



NCEP/Hydrometeorological Prediction Center (HPC)

February 2008

GOALS

- Introduce basic concepts and terminology related to forecast verification
- Demonstrate the broad scope of the topic of forecast verification by way of examples

Overview

- Introductory information on forecast verification
- Objective verification of
  - Non-probabilistic or deterministic forecasts
  - Probabilistic forecasts

What is “forecast verification” ??

Verification: the comparison of forecast events to the corresponding observed events for the purpose of assessing the quality or usefulness of the forecast.

Why Verify??

- Forecasters' tool—knowing past performance can help forecasters improve their forecasts
- Managers' tool—performance directs and justifies expenditures
- Users' tool—knowing past performance informs planning based on current forecasts

General Process of Forecast Verification

1. Acquire verification data
2. Transform verification data, e.g.,
   (a) Objective analysis (data assimilation)
   (b) Remapping
   (c) Fourier decomposition
3. Transform forecast data if necessary
4. Match forecast data to verification data
5. Compute and display performance measures

Accuracy vs Skill

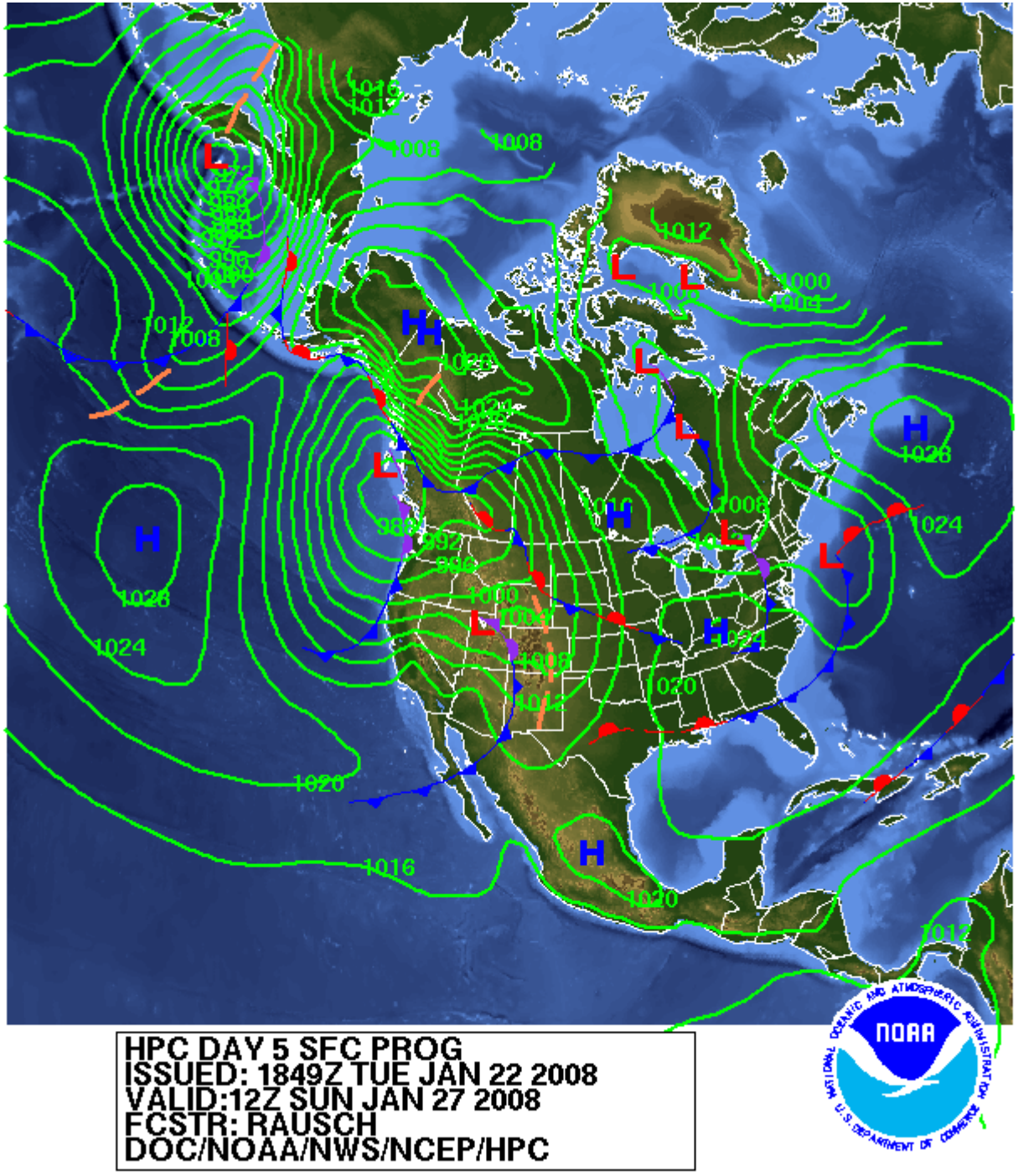
- Accuracy measures how well forecasts match observations
  - positively oriented accuracy metrics increase for better forecasts
  - negatively oriented accuracy metrics decrease for better forecasts
- Skill assesses graphically or quantitatively the relative accuracy of forecasts compared to benchmark forecasts such as
  - Random chance
  - Persistence
  - Climatology
  - Other forecasting systems

$$Skill\ \ Score = \frac{A - A_{benchmark}}{A_{perfect} - A_{benchmark}}$$

Objective Verification of Deterministic Forecasts

HPC DAY 3—7 PMSL Forecasts

- Issued three times per day:
  - Early issuance by 0900 UTC
  - Preliminary issuance by 1315 UTC
  - Final issuance by 1900 UTC
- Valid at 12 UTC
- Verifying data: HPC manual PMSL analysis
- Performance accuracy measures:
  - Anomaly correlations (positively oriented)
  - Phase errors (both orientations)

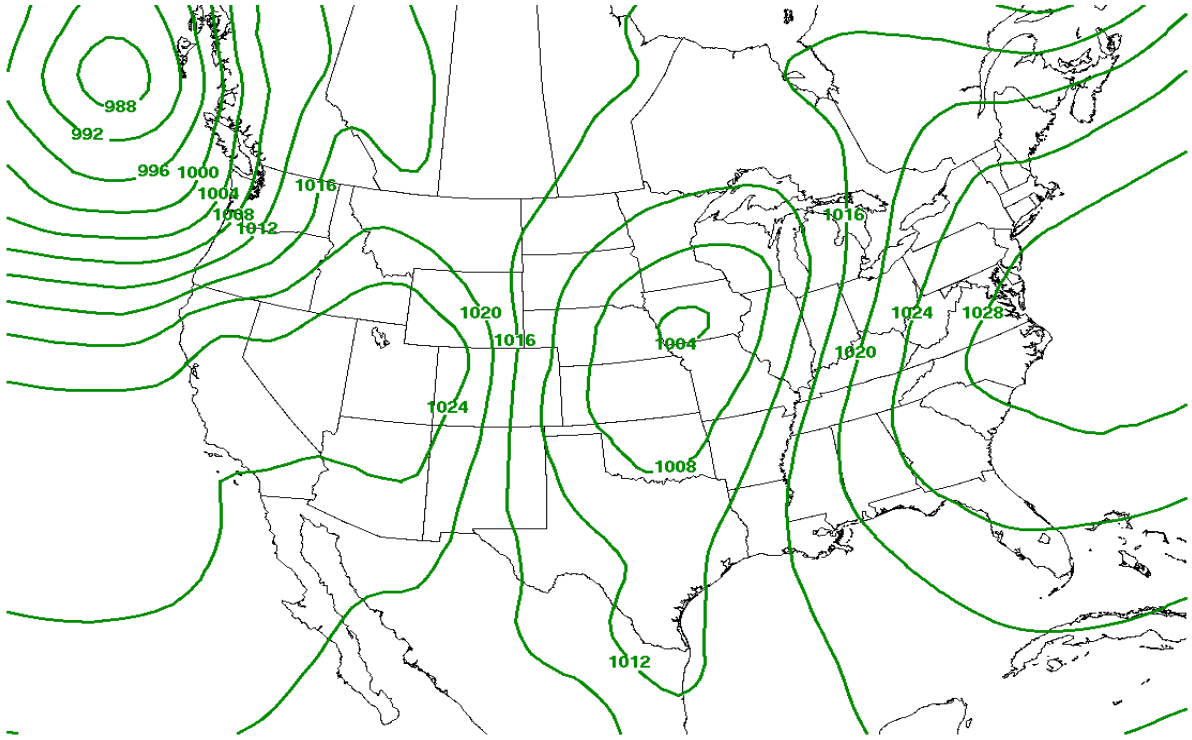


Example HPC Day 5 MSLP forecast with fronts.

Objective Verification

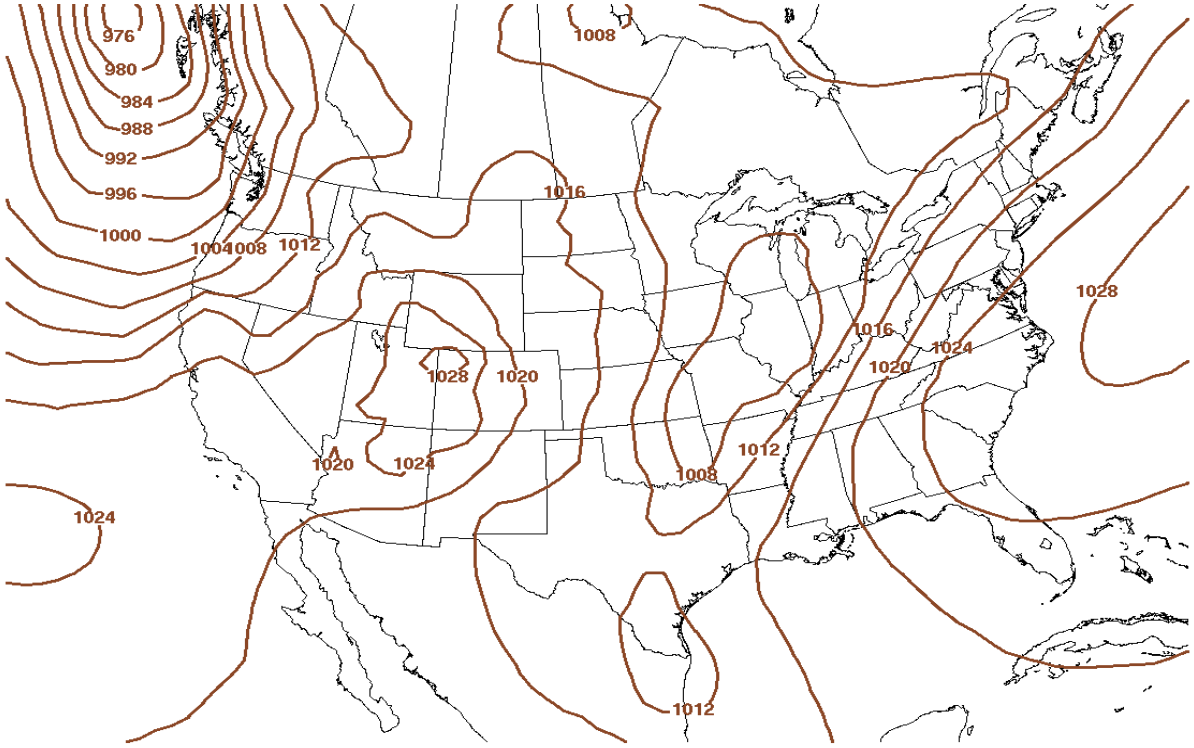
Deterministic HPC MSLP forecasts for days 3—7: verification using anomaly correlation calculations.

Example PMSL Forecast



HPC PMSL DAY 6 FORECAST VALID TUE 080108/1200V144

Verifying Analysis



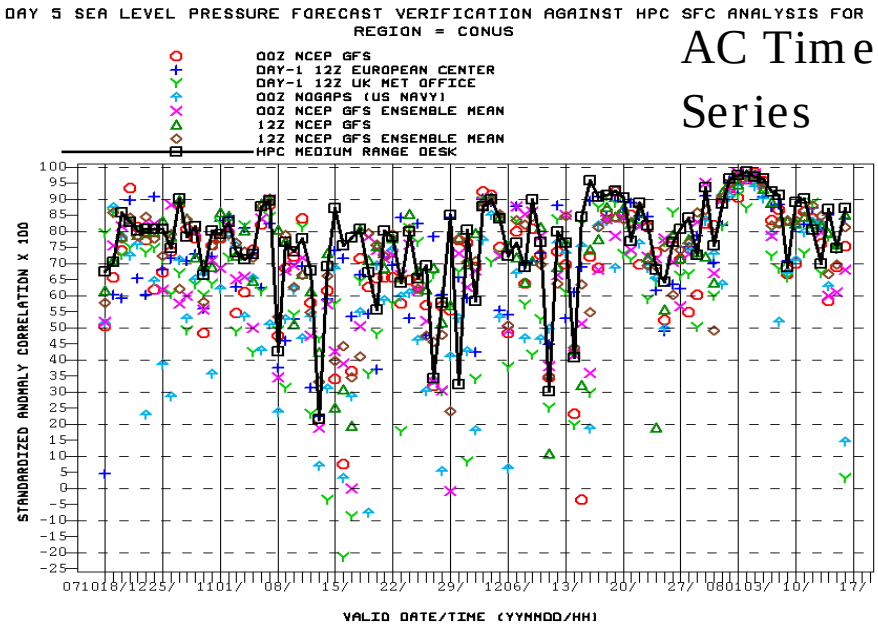
PMSL FROM HPC SFC ANALYSIS DESK 080108/1200F000

Anomaly Correlation (AC)

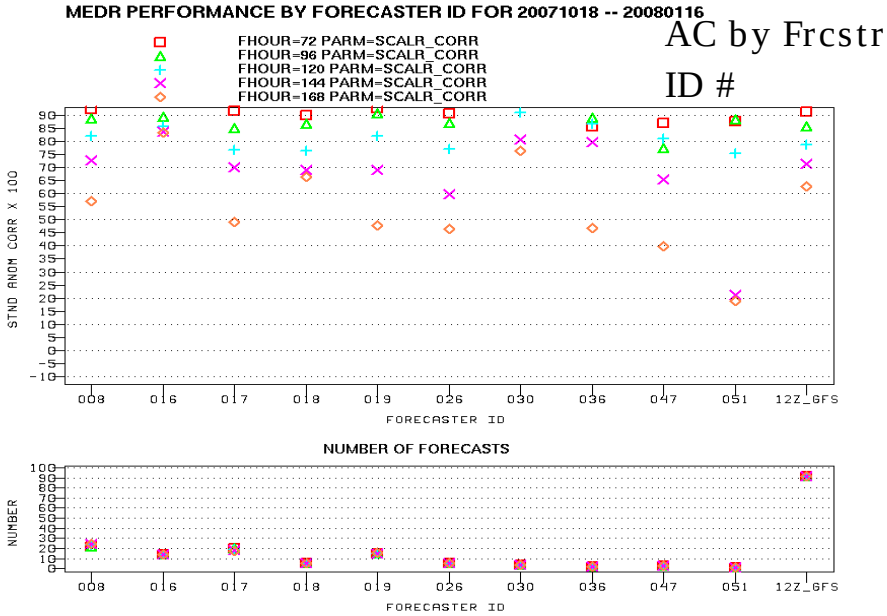
- Transform data: Standardized MSLP Anomaly = $P' = \frac{(P - \bar{P}_{CLIMO})}{\delta_{CLIMO}}$
  - Use climatological values from 41 years (1958—1998) of Global Reanalysis
  - Apply transformation to both analysis and forecast grid point values.
- Compute accuracy measure, AC, Pearson product-moment correlation:

$$AC = \frac{\sum_{m=1}^{N} (f_m' - \bar{f}')(o_m' - \bar{o}')}{\sqrt{\sum_{m=1}^{N} (f_m' - \bar{f}')^2 \sum_{m=1}^{N} (o_m' - \bar{o}')^2}}$$

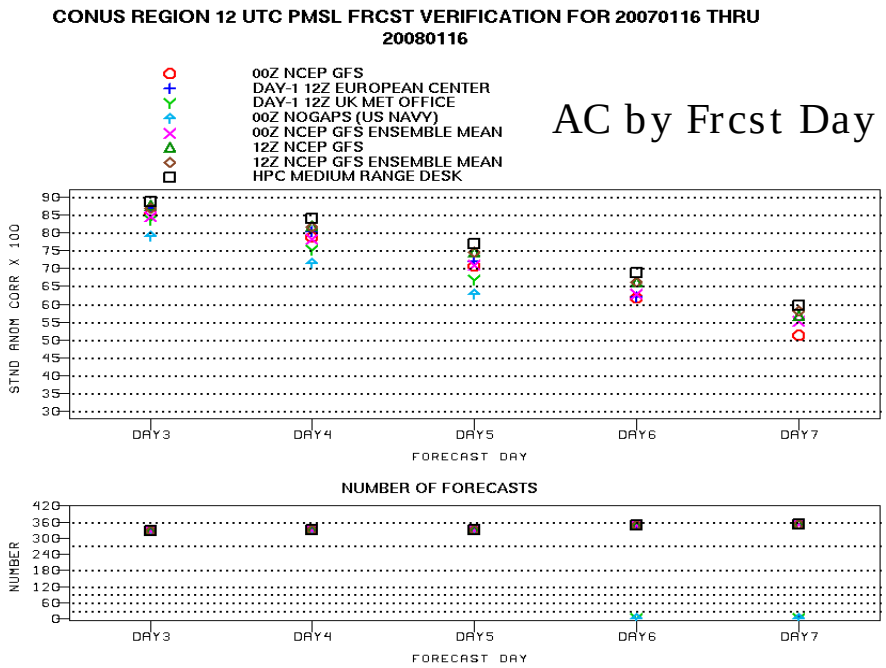
Various Displays of Anomaly Correlations



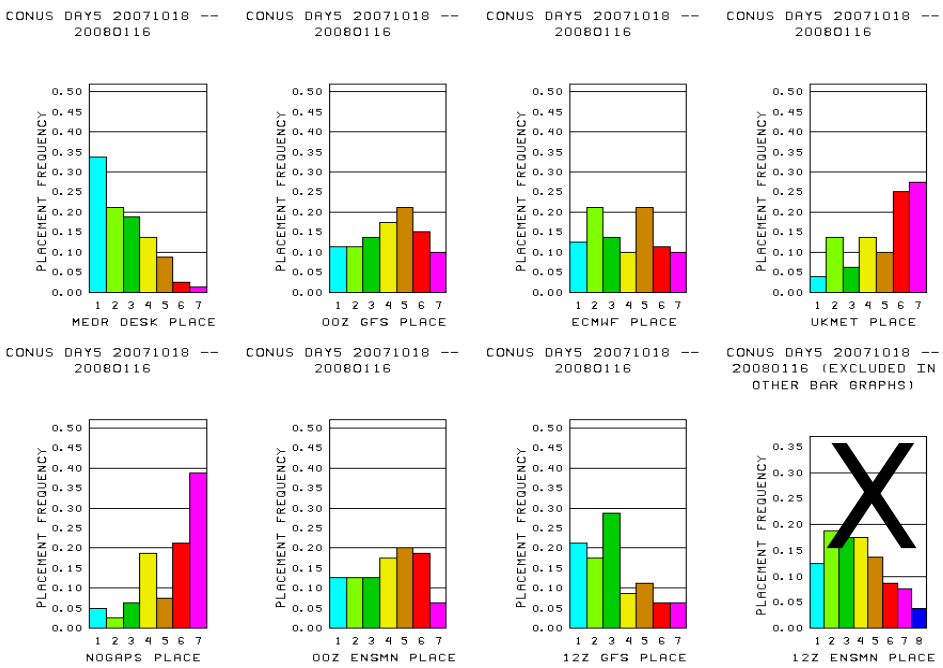
AC Time Series



AC by Frcstr ID #



AC by Frcst Day

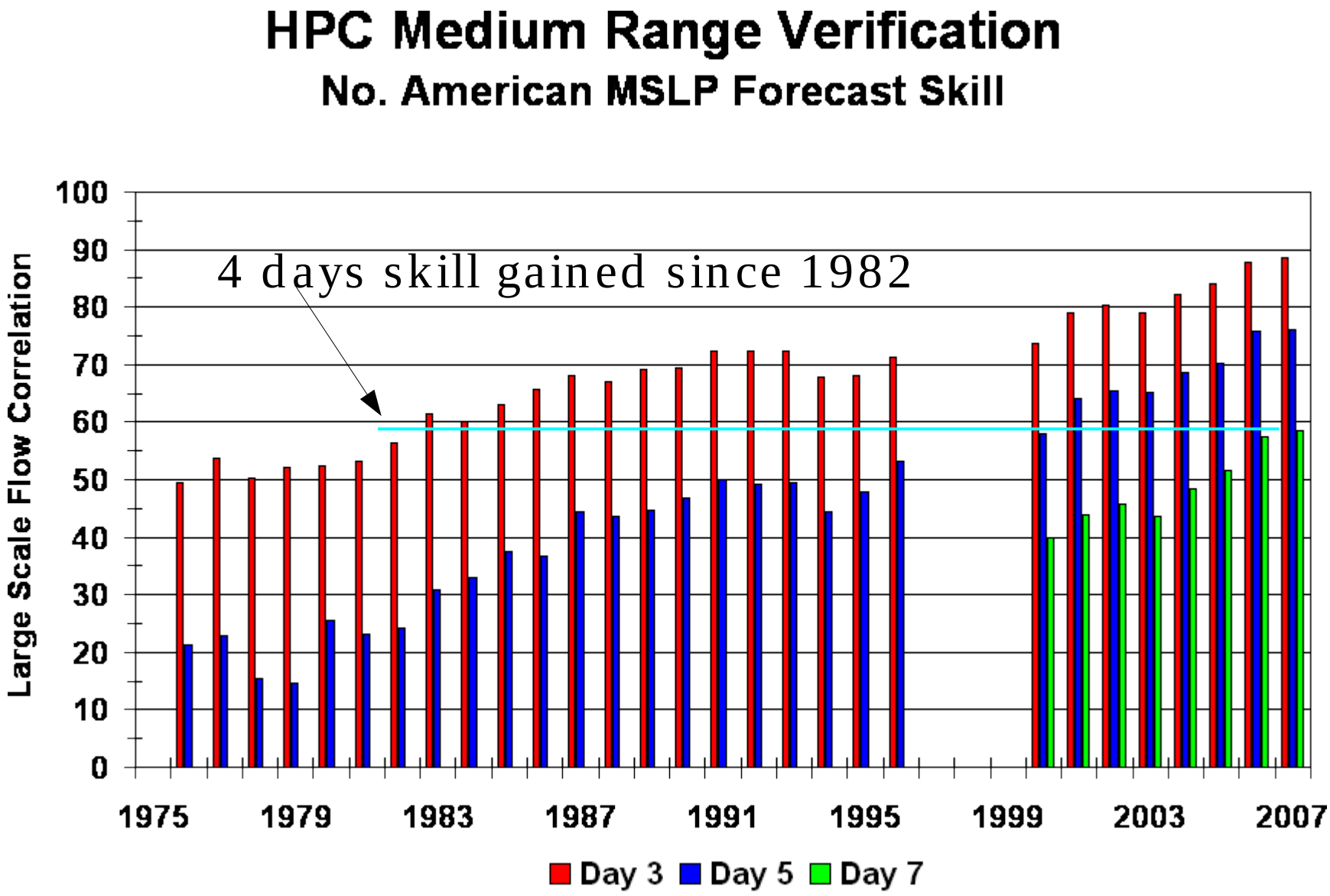


Significance Test: Does HPC beat the 12Z ENS mean?

STATISTICAL SIGNIFICANCE ANALYSIS FOR PAIRED DATA FROM TWO TRACES

```
STAT=SSAL1L2 PARAM=SLP FHOUR=120 V_ANL=HPC/SFC V_RGN=MRDG VHHMM=1200
                    200710181200 TO 200801161200

                TRACE #  1 MODEL=MEDR/ SCALR_CORR
                TRACE #  2 MODEL=ENSMN SCALR_CORR

Null Hypothesis (trace1 = trace2):
   The mean of the differences of the paired values from the
   two traces is zero.

Alternative Hypothesis (trace1 > trace2):
   The mean of the paired differences exceeds zero.

Gaussian probability of wrongly rejecting null hypothesis:   0.002145

        Test Level   Decision (one-sided Gaussian)
          0.100       REJECT null hypothesis
          0.050       REJECT null hypothesis
          0.010       REJECT null hypothesis
          0.005       REJECT null hypothesis
          0.001       ACCEPT null hypothesis

ZRAW =   2.83104E+00  ZTST =   2.85600E+00  N =       91  Lag1 Corr = -0.008778
```

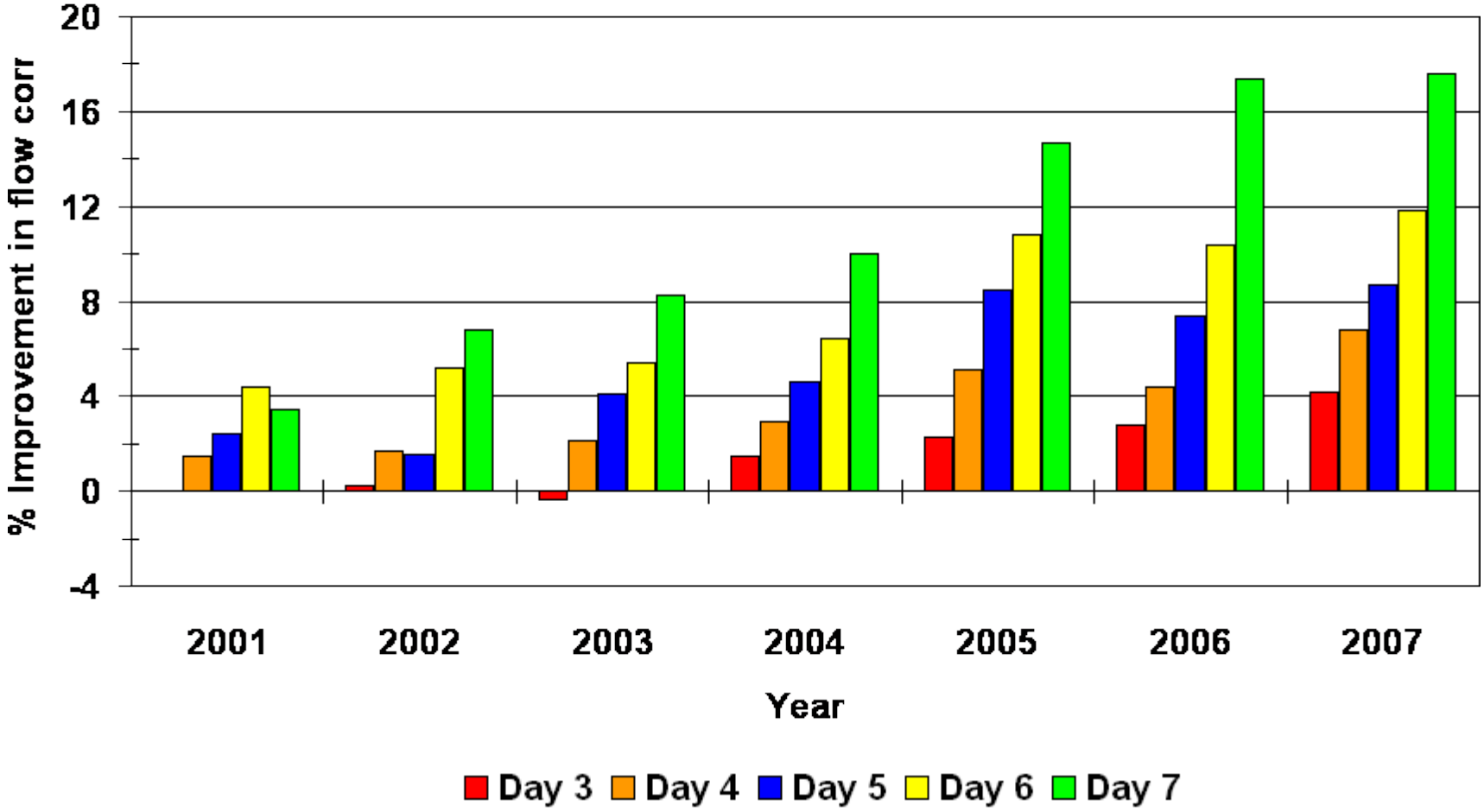
No significant difference at strictest test level



Historical AC Performance

HPC Medium Range Verification

% improvement in flow correlation over GFS 2001-2007



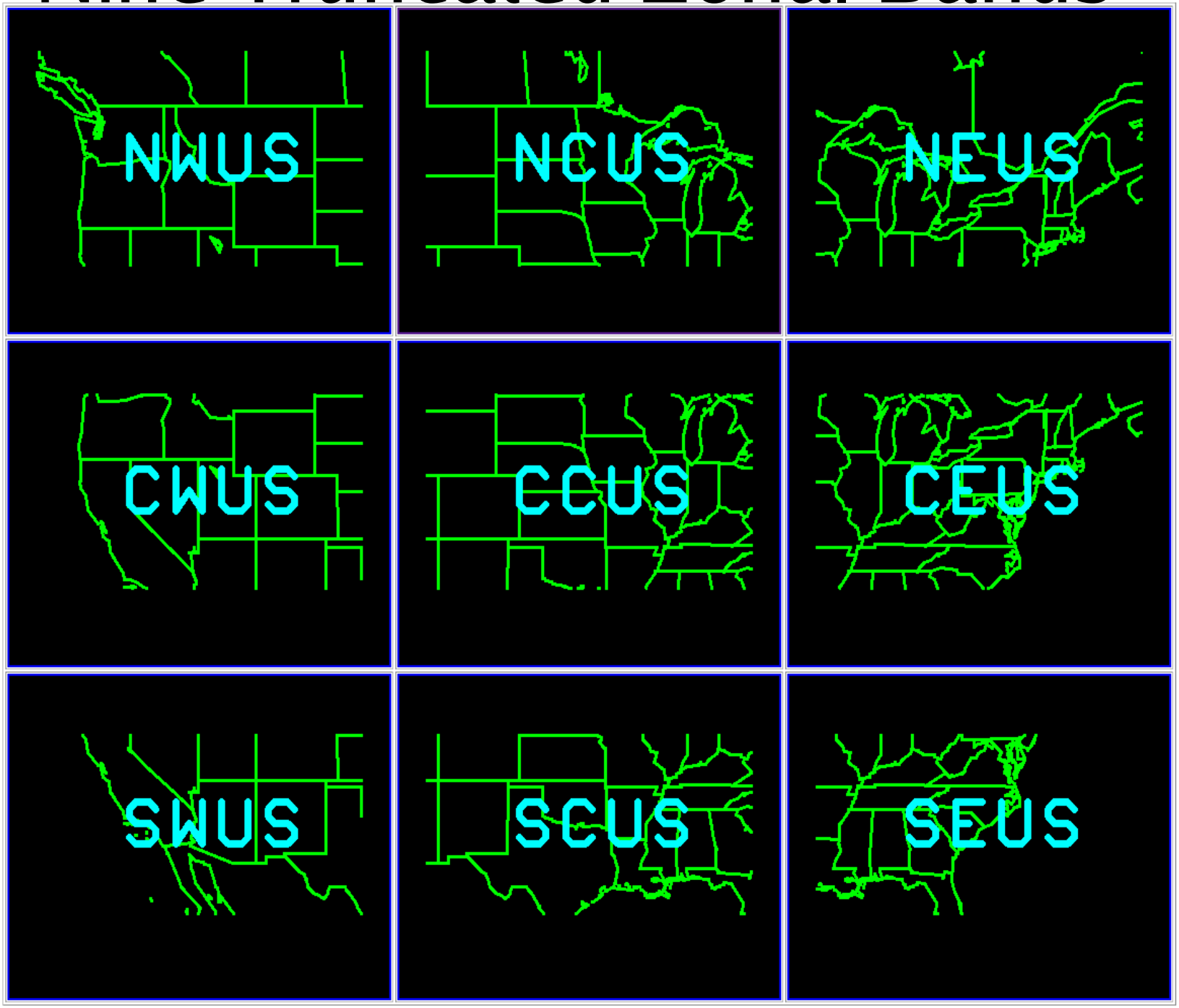
Objective Verification

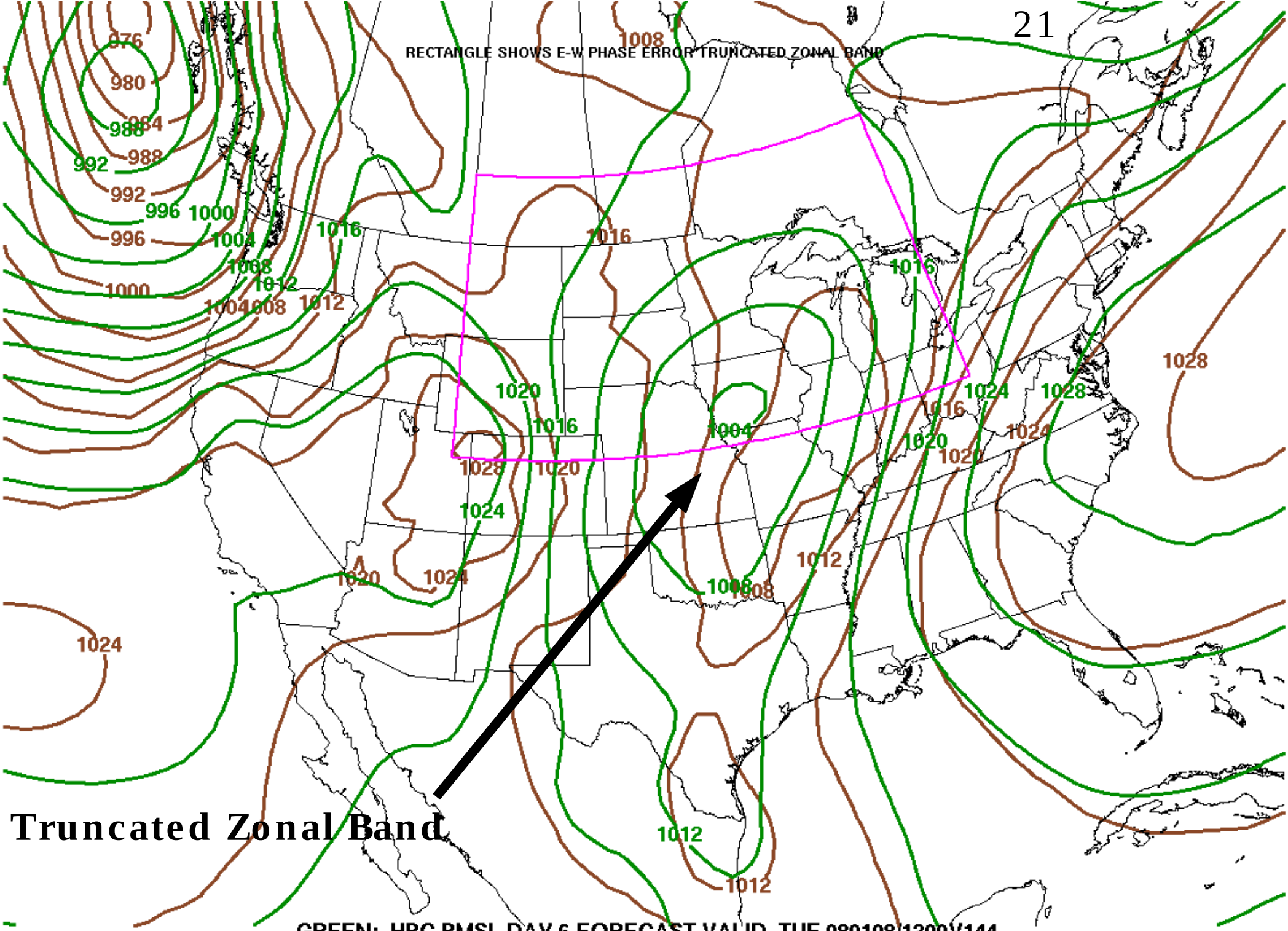
Deterministic HPC MSLP forecasts for days 3—7: verification by computing east-west phase error.

Data Transformation for East-West Phase Error

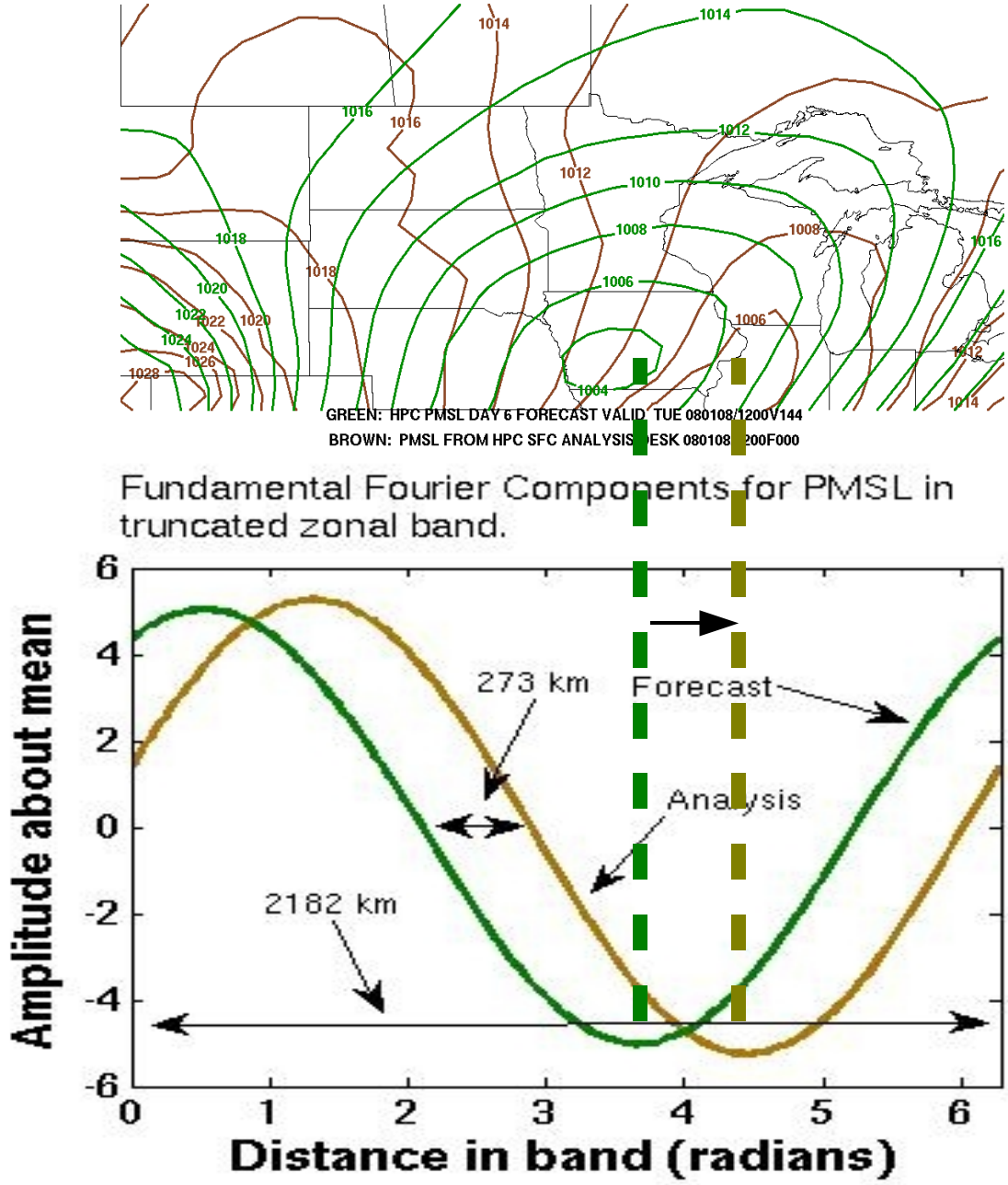
- Average meridionally to get series of PMSL values in zonal direction in truncated zonal bands
- Do Fourier Transform in truncated zonal bands
- Use the first harmonic or fundamental of cosine series
- Convert difference in phase angle between forecast and analysis to phase error distance

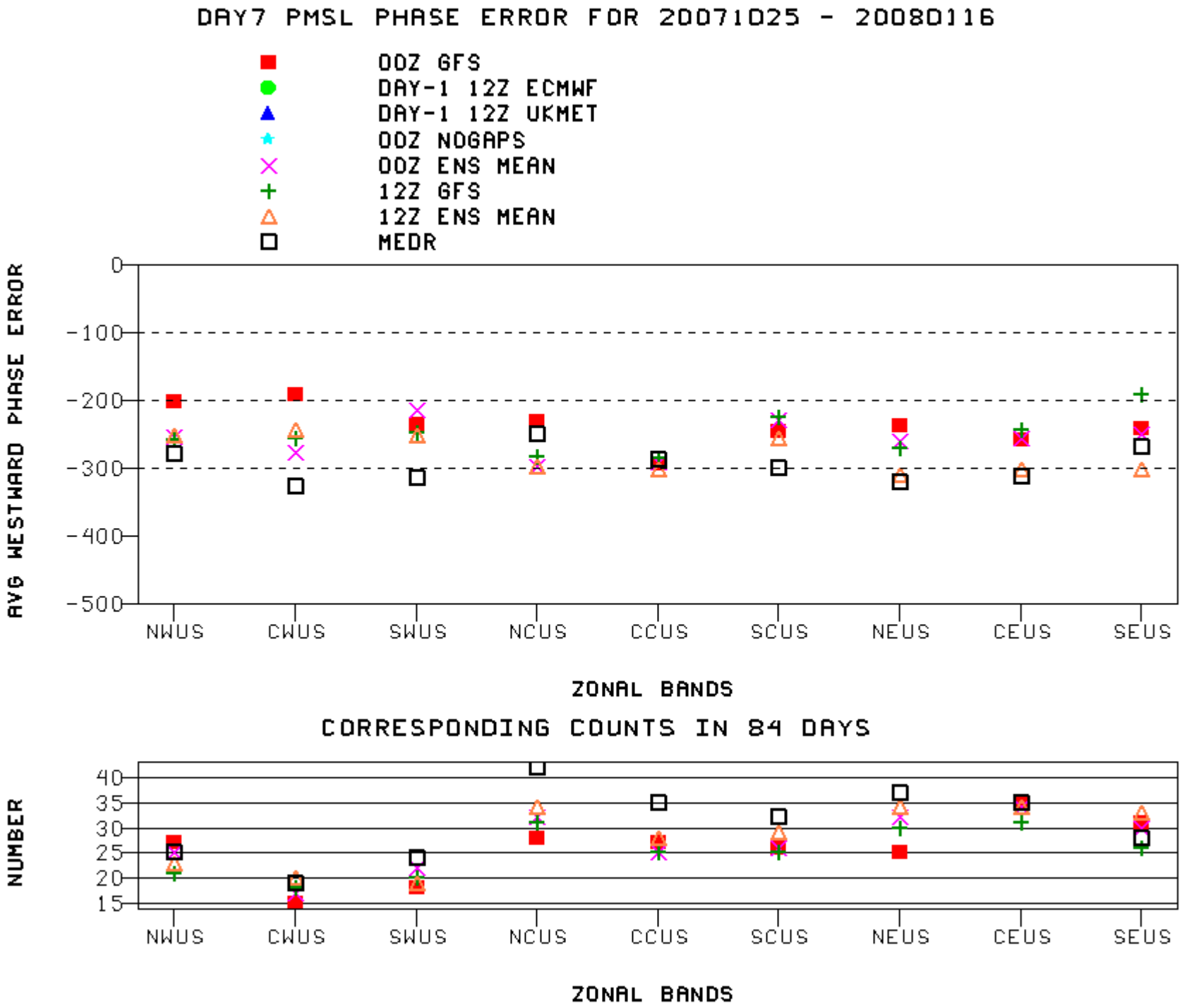
Nine Truncated Zonal Bands

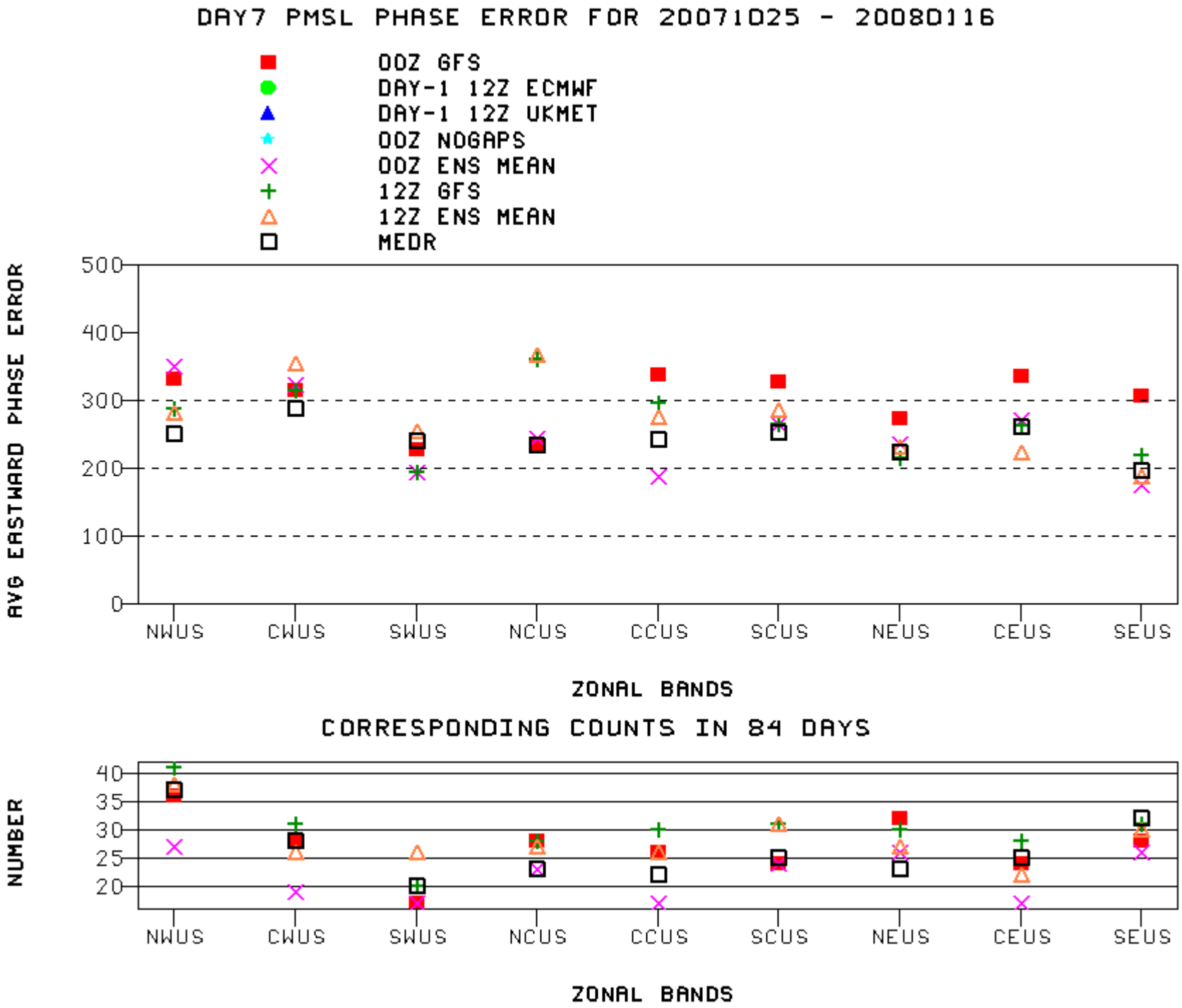



RECTANGLE SHOWS E-W PHASE ERROR TRUNCATED ZONAL BAND
Truncated Zonal Band
GREEN: HPC PMSL DAY 6 FORECAST VALID TUE 080108/1200V144
BROWN: PMSL FROM HPC SFC ANALYSIS DESK 080108/1200F000

Computing phase error in NCUS Zonal Band

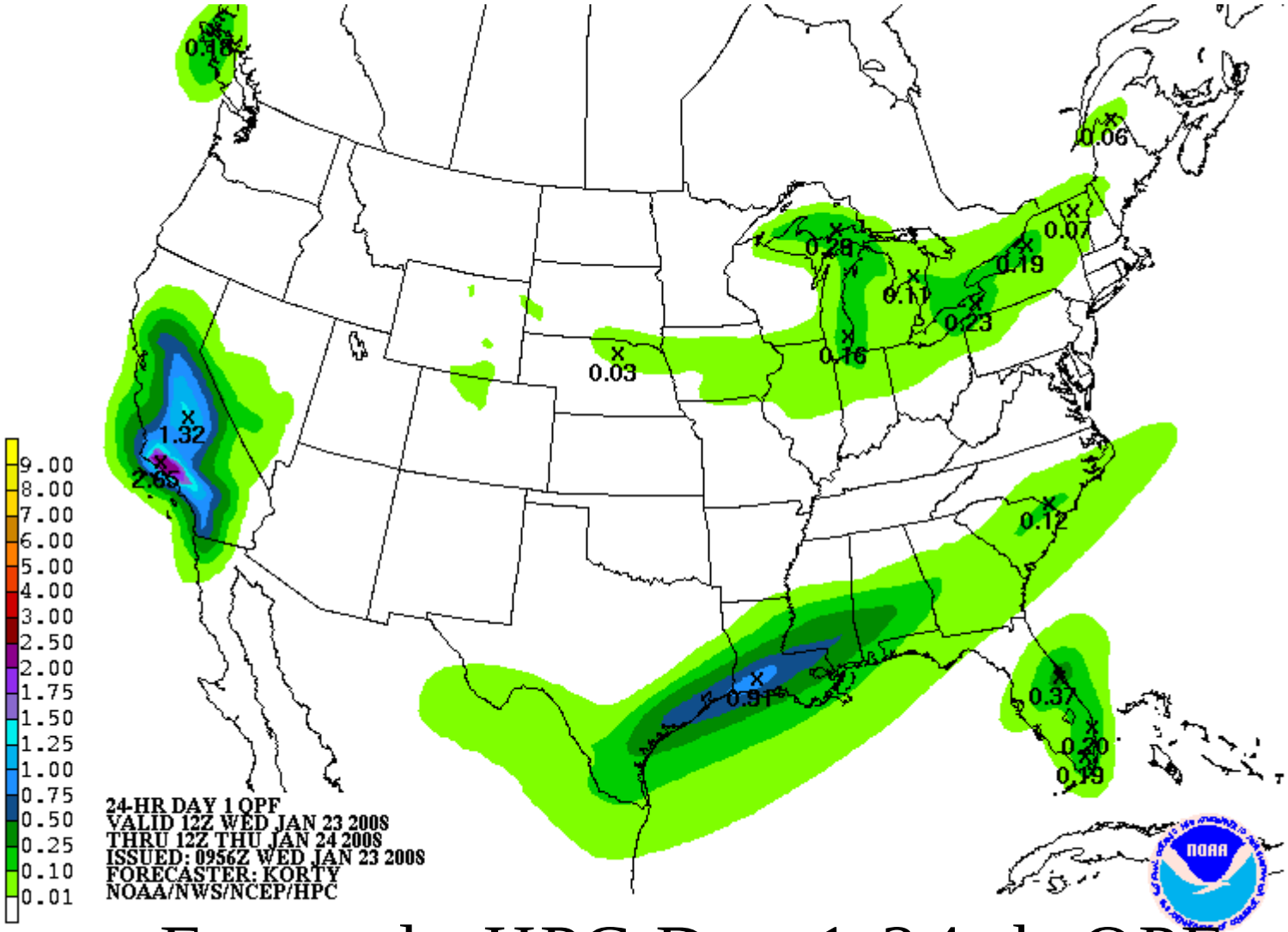



DAY7 PMSL PHASE ERROR FOR 20071025 - 20080116
00Z GFS
DAY-1 12Z ECMWF
DAY-1 12Z UKMET
00Z NOGAPS
00Z ENS MEAN
12Z GFS
12Z ENS MEAN
MEDR
AVG WESTWARD PHASE ERROR
0
-100
-200
-300
-400
-500
NWUS
CWUS
SWUS
NCUS
CCUS
SCUS
NEUS
CEUS
SEUS
ZONAL BANDS
CORRESPONDING COUNTS IN 84 DAYS
NUMBER
40
35
30
25
20
15
NWUS
CWUS
SWUS
NCUS
CCUS
SCUS
NEUS
CEUS
SEUS
ZONAL BANDS


DAY7 PMSL PHASE ERROR FOR 20071025 - 20080116
00Z GFS
DAY-1 12Z ECMWF
DAY-1 12Z UKMET
00Z NOGAPS
00Z ENS MEAN
12Z GFS
12Z ENS MEAN
MEDR
AVG EASTWARD PHASE ERROR
500
400
300
200
100
0
NWUS
CWUS
SWUS
NCUS
CCUS
SCUS
NEUS
CEUS
SEUS
ZONAL BANDS
CORRESPONDING COUNTS IN 84 DAYS
NUMBER
40
35
30
25
20
NWUS
CWUS
SWUS
NCUS
CCUS
SCUS
NEUS
CEUS
SEUS
ZONAL BANDS

More Objective Verification

Verification of the deterministic HPC quantitative precipitation forecasts (QPF) for day 1 by constructing a 2X2 contingency table for accumulations exceeding a threshold.

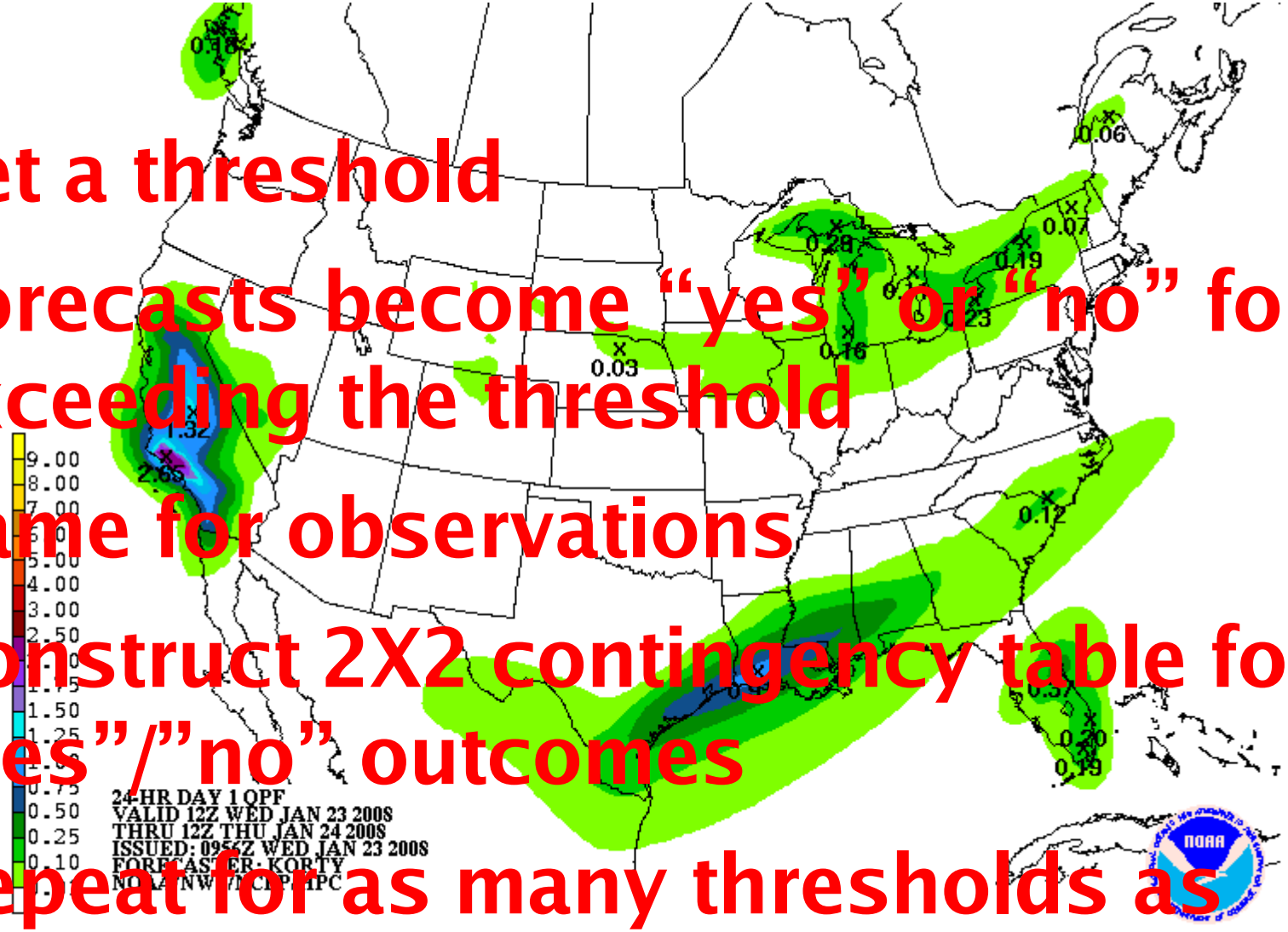


Example HPC Day 1 24-h QPF

Verified against the HPC manual analysis of gage data

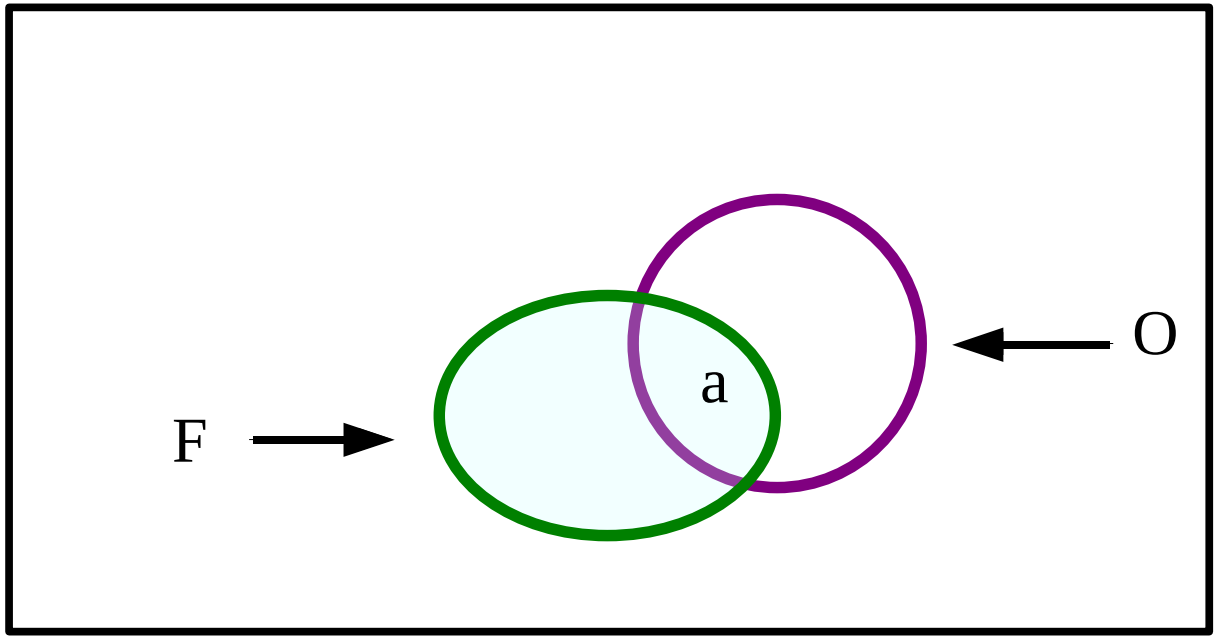
A Simple Transformation

- Set a threshold
- Forecasts become "yes" or "no" for exceeding the threshold
- Same for observations
- Construct 2X2 contingency table for "yes"/"no" outcomes
- Repeat for as many thresholds as desired



2X2 Contingency Table

Event	“Yes” Observed	“No” Observed	Totals
“Yes” forecasts	a	b	a + b = F
“No” forecasts	c	d	c + d = ~F
Totals	a + c = O	b + d = ~O	a + b + c + d



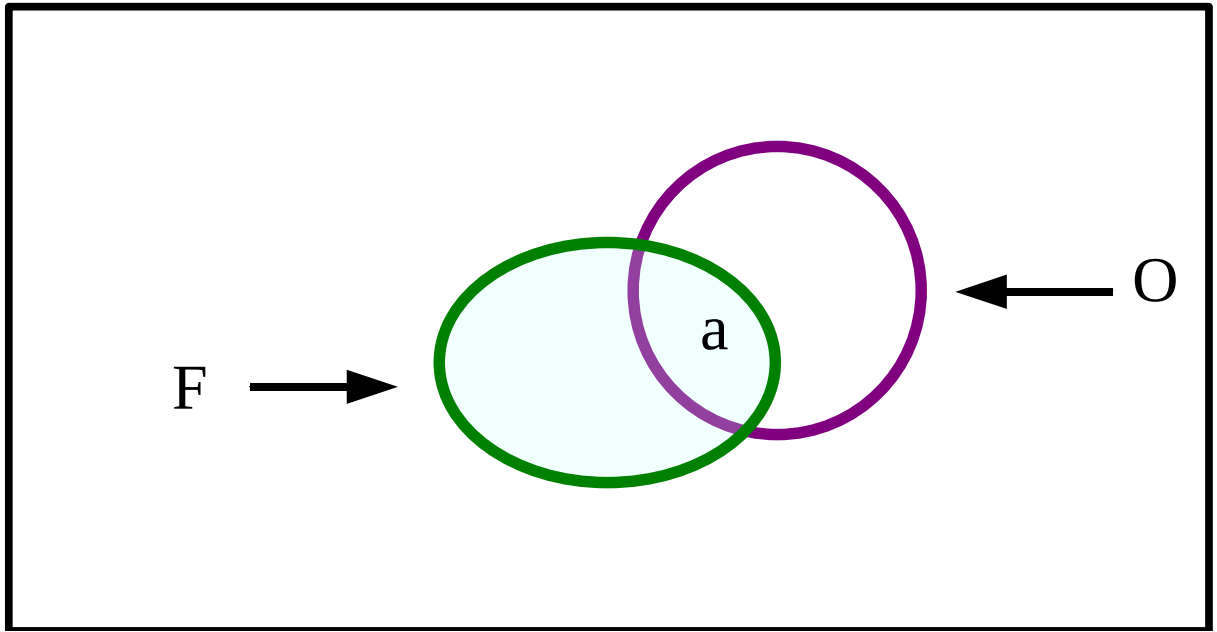
Accuracy Measures

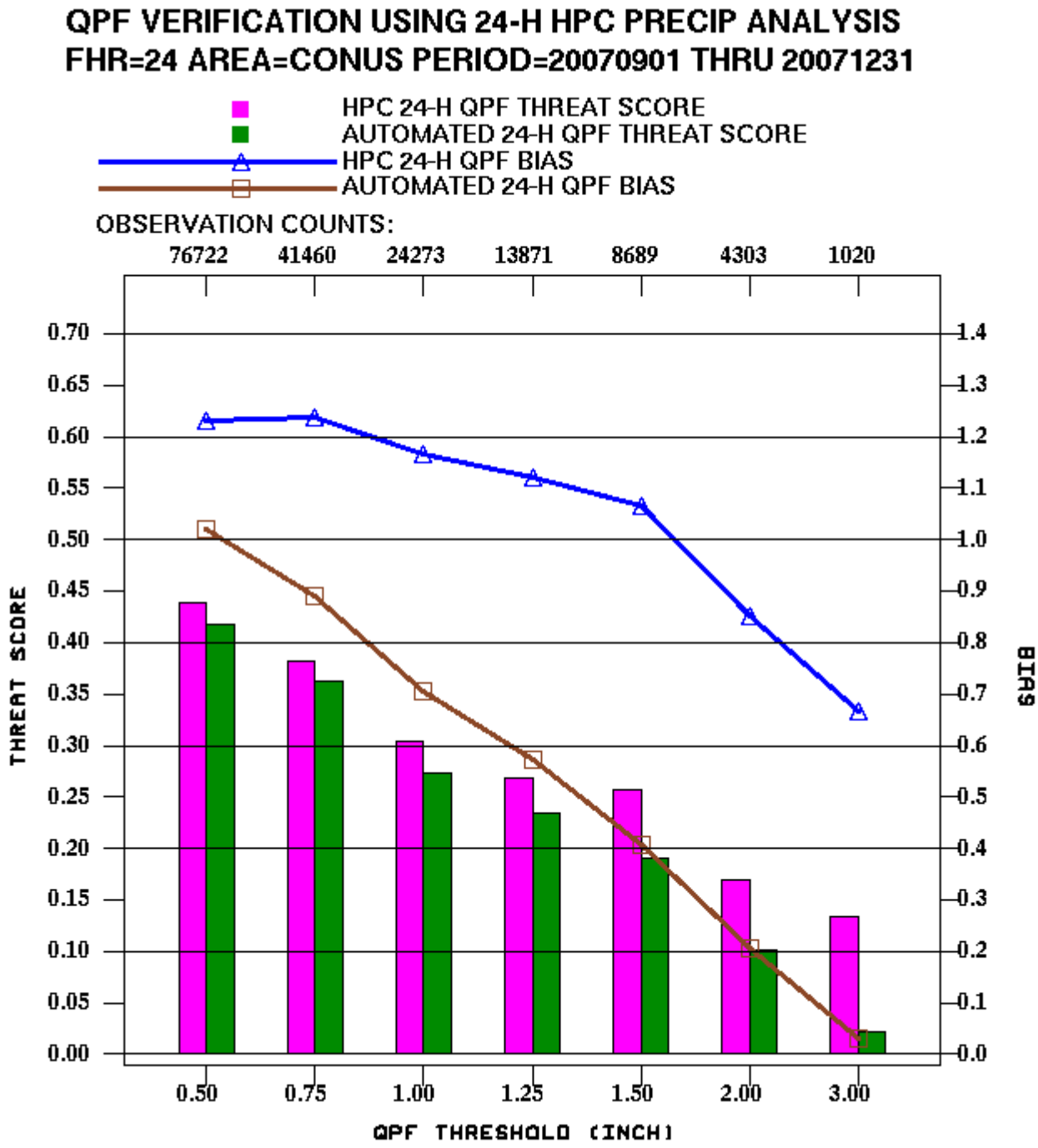
Threat score = area correct / (area forecast + area observed) =

$$\frac{a}{(F+O-a)}$$

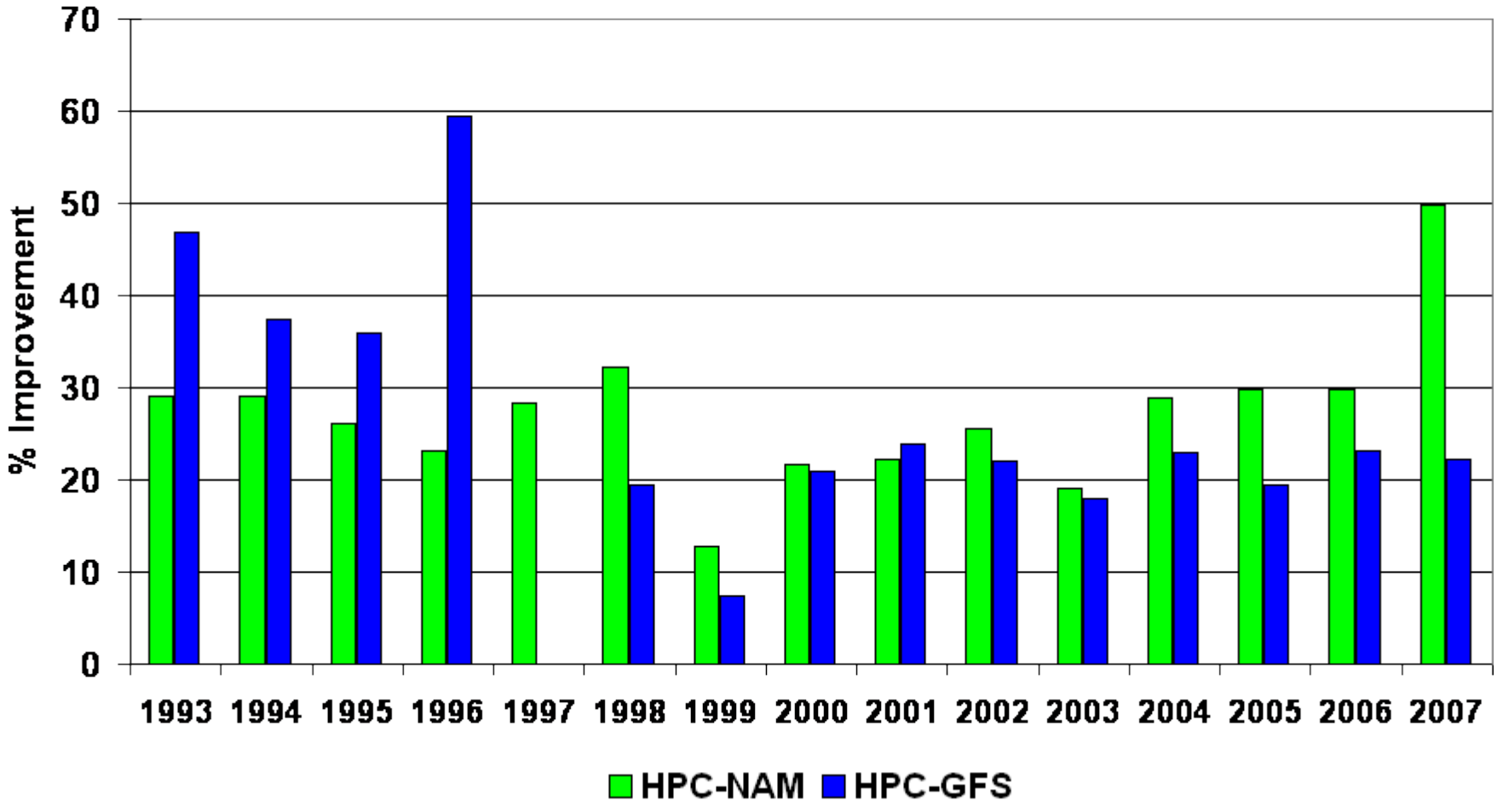
Threat score ranges from 0 to 1 (perfect forecast) and is positively oriented.

Bias = F / O; ideally B = 1.




QPF VERIFICATION USING 24-H HPC PRECIP ANALYSIS
FHR=24 AREA=CONUS PERIOD=20070901 THRU 20071231
HPC 24-H QPF THREAT SCORE
AUTOMATED 24-H QPF THREAT SCORE
HPC 24-H QPF BIAS
AUTOMATED 24-H QPF BIAS
OBSERVATION COUNTS:
76722
41460
24273
13871
8689
4303
1020
THREAT SCORE
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
BIAS
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.50
0.75
1.00
1.25
1.50
2.00
3.00
QPF THRESHOLD (INCH)

HPC % Improvement to NCEP models
1-Inch Day 1 QPF Forecast



Verification of Probabilistic Forecasts

Some Terminology

- Reliability: Does the event occur X% of the time when the forecast is for X% chance?
- Resolution: Do forecasts of different probabilities separate the observations into groups for which the observed frequencies differ and differ from climatology?
- Uncertainty: What challenge is posed by the climatological probability of the observed event?

Performance Assessment

- Construct attribute diagrams
- Compute Brier score = BS = $\frac{1}{N}\sum_{m=1}^{N}(p_F - p_O)^2$
  - p_O is either 1 or 0 – the event either happens or not.
  - Mean squared error in probability; so, perfection is BS=0 (negatively oriented).
  - Decomposition of BS:

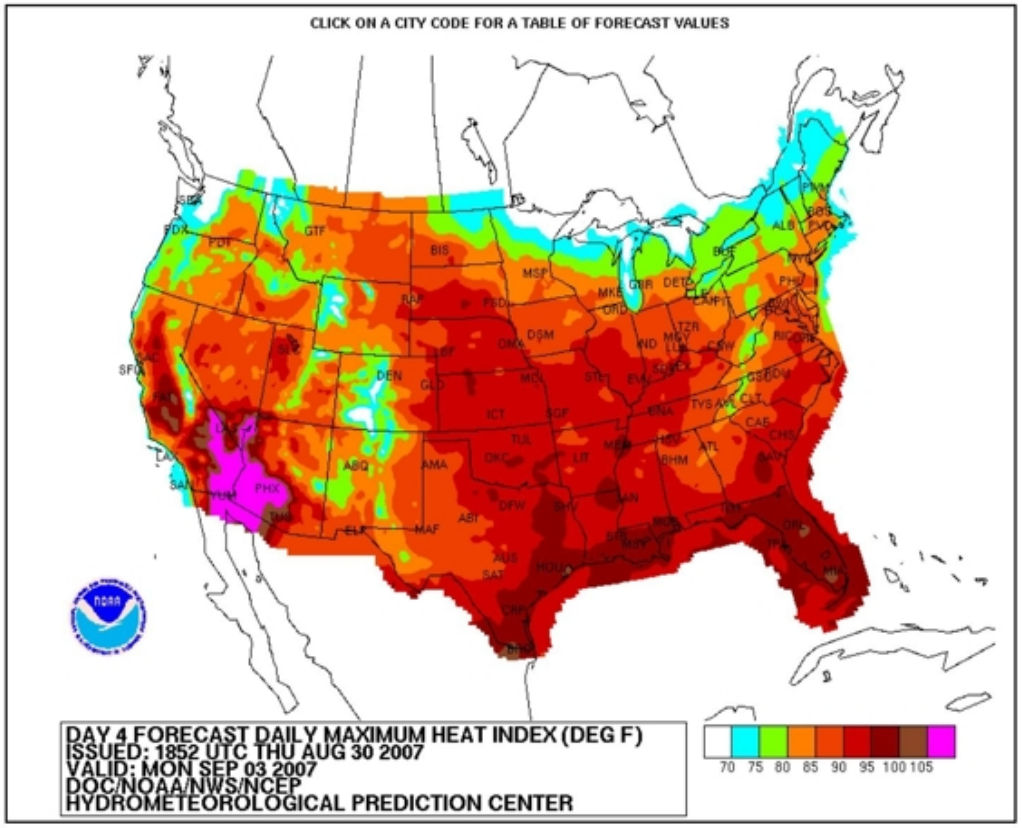
    BS = Reliability – Resolution + Uncertainty

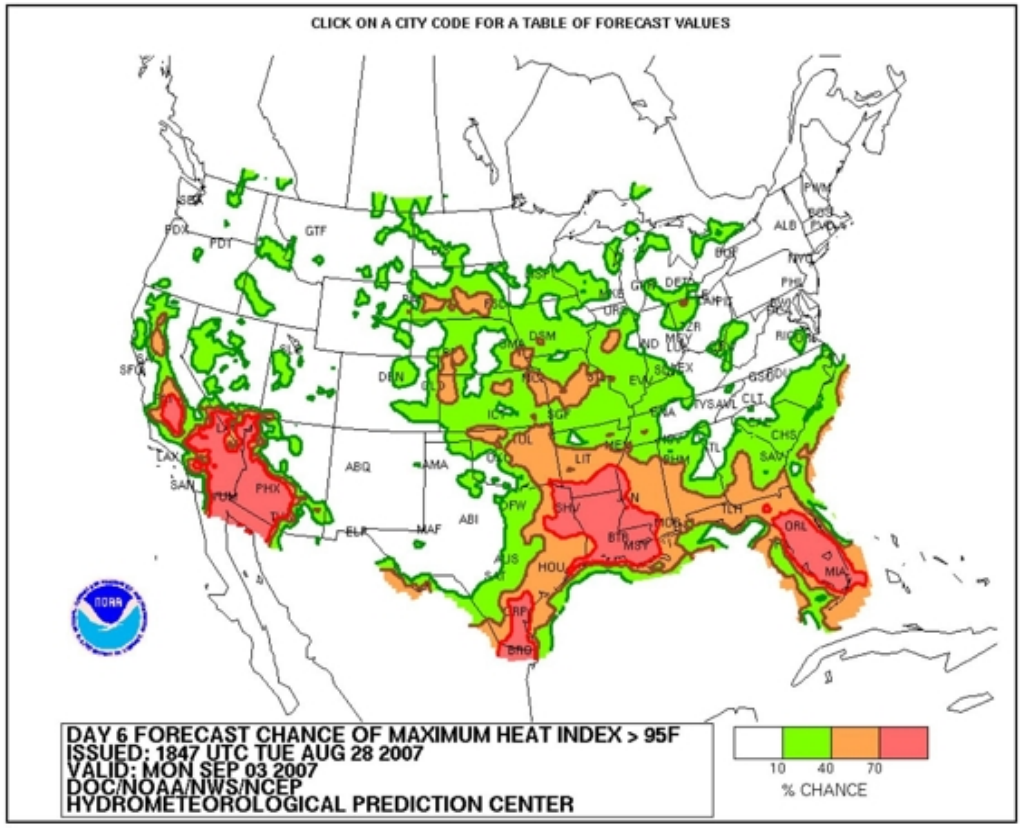
- Brier Skill Score: skill w.r.t. sample climatology

  BSS= (Resolution – Reliability)/ Uncertainty

  BSS=1 for perfect forecasts

HPC Heat Index Guidance Products

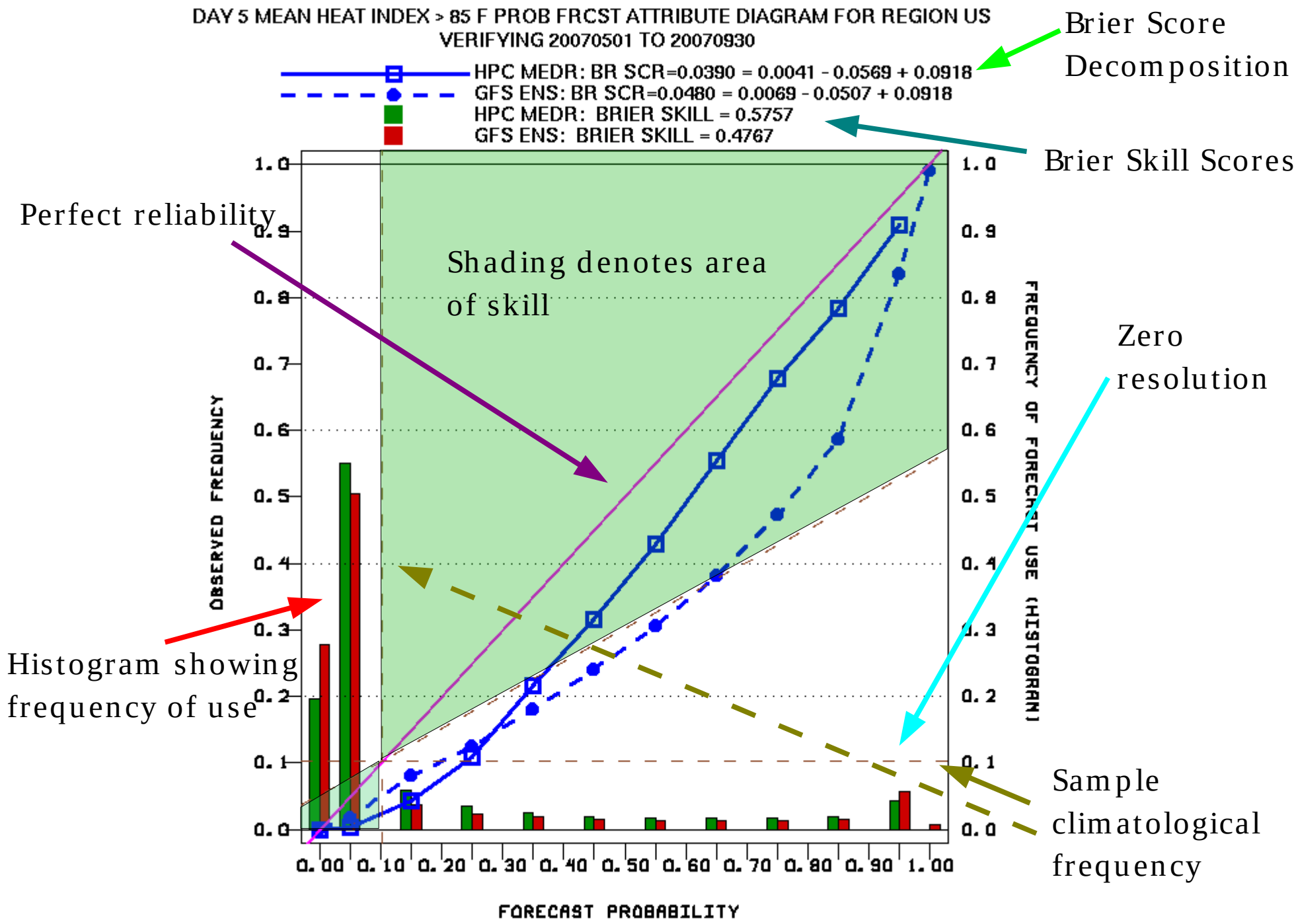




Deterministic HI forecast verification generates L1L2 norm statistics.

Probabilistic HI forecast verification generates Partitioned Brier Score (PBS) statistics.

Verified using Rapid Update Cycle (RUC) Data


DAY 5 MEAN HEAT INDEX > 85 F PROB FRCST ATTRIBUTE DIAGRAM FOR REGION US
VERIFYING 20070501 TO 20070930
HPC MEDR: BR SCR=0.0390 = 0.0041 - 0.0569 + 0.0918
GFS ENS: BR SCR=0.0480 = 0.0069 - 0.0507 + 0.0918
HPC MEDR: BRIER SKILL = 0.5757
GFS ENS: BRIER SKILL = 0.4767
Brier Score Decomposition
Brier Skill Scores
Perfect reliability
Shading denotes area of skill
Zero resolution
Histogram showing frequency of use
Sample climatological frequency
OBSERVED FREQUENCY
FREQUENCY OF FORECAST USE (HISTOGRAM)
FORECAST PROBABILITY
0.00 0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 1.00
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

The Cutting Edge of Verification

- Use “fuzzy logic” to identify and match forecast and observed features to get displacement errors
- Use re-sampling techniques to assess statistical significance of verification outcomes
- Use observational error to put error bars on accuracy measurements

The End