Pseudo Bias Corrected Ensemble QPF (ensqpfbc) Described & Verified

Keith Brill

Spring 2011

Historical Perspective

- Began as the “smart” ensemble QPF with downscaling (ensqpf, late 2005)
- Intended to support HPC Day 2—3 6-h QPFs
- Installed as an interim automated backup for HPC QPF at SPC (early 2006)
- Used for “early” Day 3 for WR points (late 2007)
- Listed still as a selection in nmap2
- Updated with the pseudo bias correction (ensqpfbc, April 2008)

Presentation Overview

- Describe the "smart" ensemble concept
- Describe the pseudo bias correction applied to the "smart" ensemble 6-h QPF
- Describe the PRISM multiplicative down-scaling applied after pseudo bias correction
- Present the verification experiment
- Summarize and enumerate future work

The “Smart” Ensemble Concept

- Use a multi-model approach
  - Deterministic runs (NAM, GFS, ECMWF) **+**
  - CMC and ensembles: SREF or GEFS
- Counter low dispersion by including ensemble maximum as another member
- Measure uncertainty using normalized spread
- Compute the “smart” mean:
  - Larger uncertainty -> lean toward mean of the full ensemble
  - Lesser uncertainty -> lean toward the mean of deterministic runs

Deterministic Runs

Model	Resolution of Output	Cold Season (Oct-Apr) Weight	Warm Season Weight
NAM	20-km	0.2	1 / 3
GFS	1 X 1 degree	0.3	1 / 3
ECMWF	1 X 1 degree	0.5	1 / 3

1) Subjective weights are applied in computing the deterministic run ensemble mean for forecast hours 06—84.
2) Equal weights are used beyond 84 hours after NAM drops out.

The Full Ensemble

1. NAM + GFS + ECMWF +
2. The SREF (21 members) or GEFS (20 members) +
3. The Canadian GEM (CMC) +
4. MAX (NAM,GFS,ECM,SREF/GEFS,CMC)

- ✓ 06—84 hours: 3 + 21 + 1 + 1 = 26 members
- ✓ 90+ hours: 2 + 20 + 1 + 1 = 24 members

The "Smart" Ensemble Mean

1. Compute the full ensemble mean and spread
2. Compute spread per unit adjusted mean:

$$\hat{\sigma} = \frac{\sigma}{(\mu + .025)} \qquad \text{(at each grid point)}$$

3. Find the maximum $\hat{\sigma}$ on the domain
4. Compute the weight for the full ensemble:

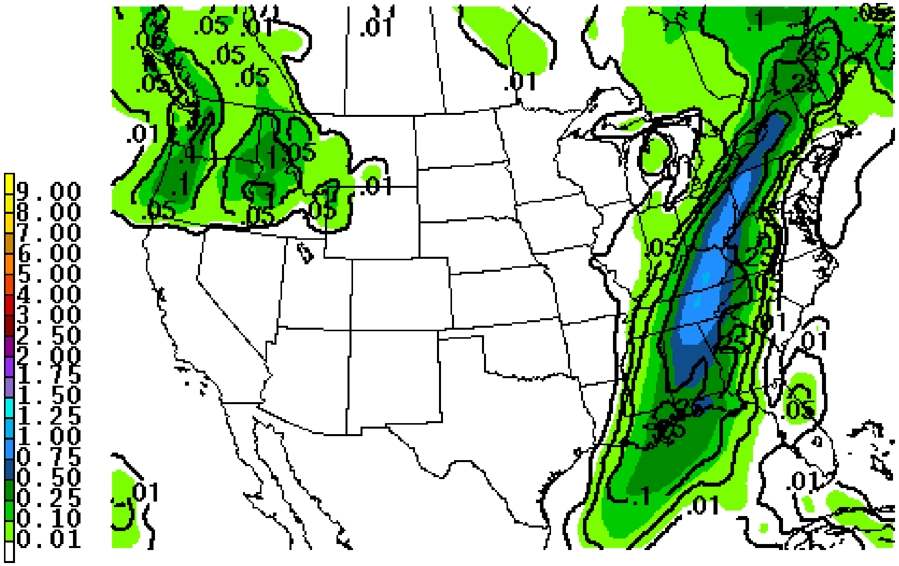
$$w = \hat{\sigma} / \hat{\sigma}_{max}$$

5. Compute the "smart" ensemble mean:

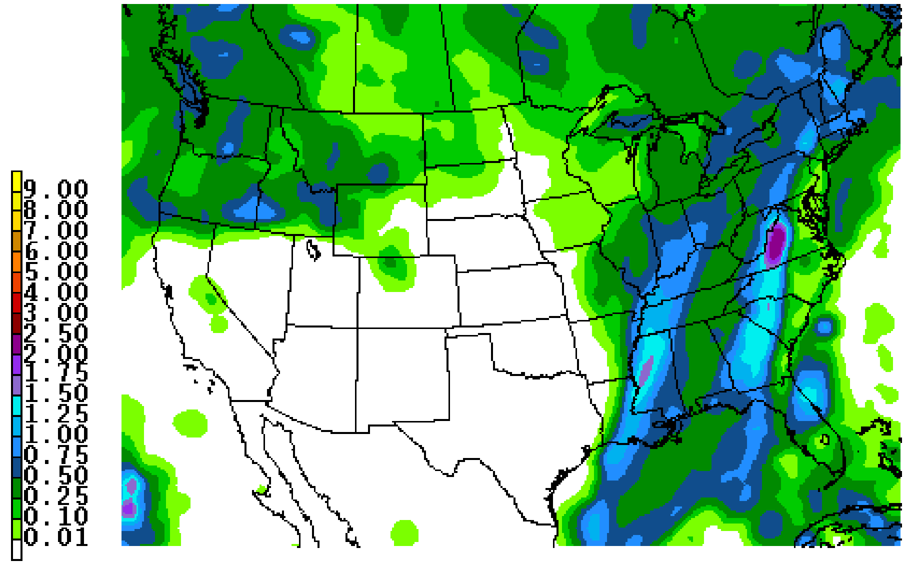
$$\mu_s = w\mu + (1 - w)\mu_{dr},$$

where μ_{dr} is mean of the deterministic runs

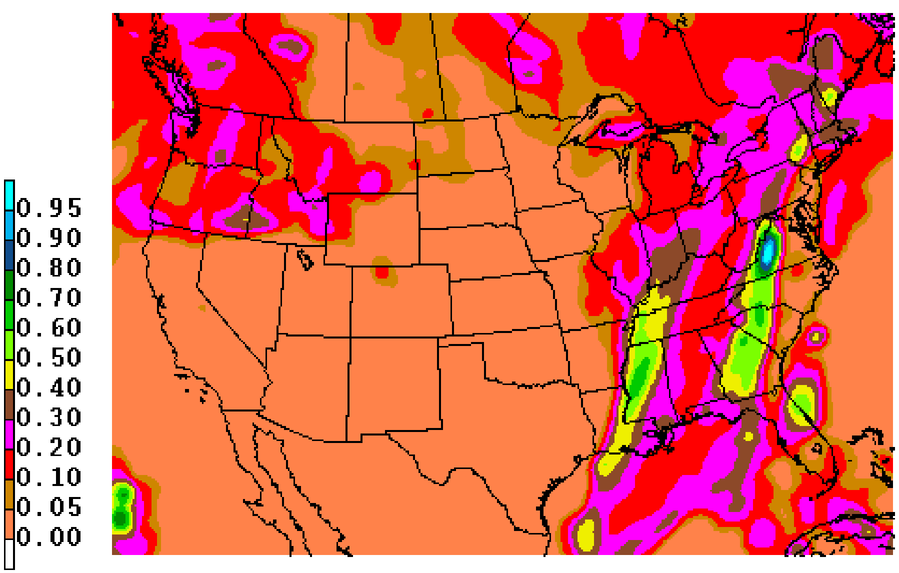
Consider examples (6-h QPFS) …



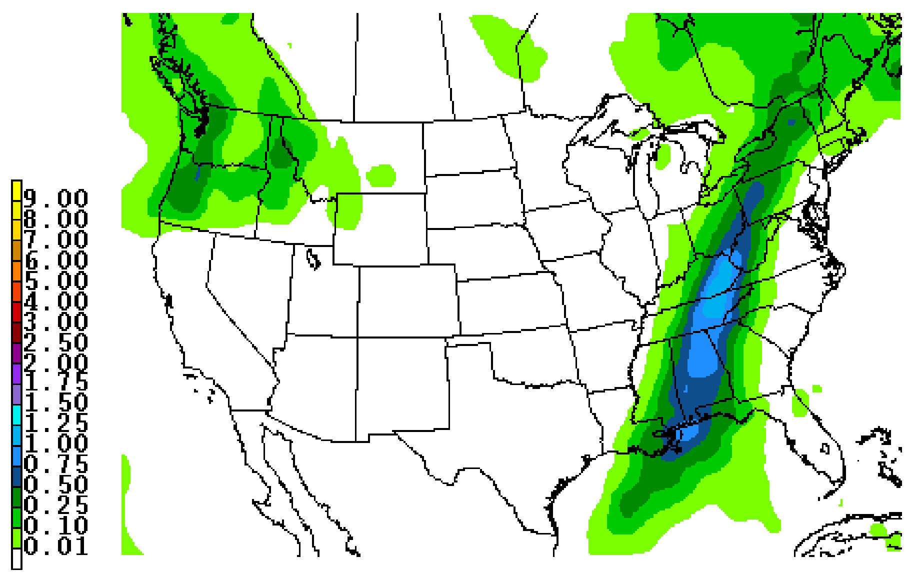
110405/0600V012 MEAN (FILL) AND SPREAD (CONTOURS)



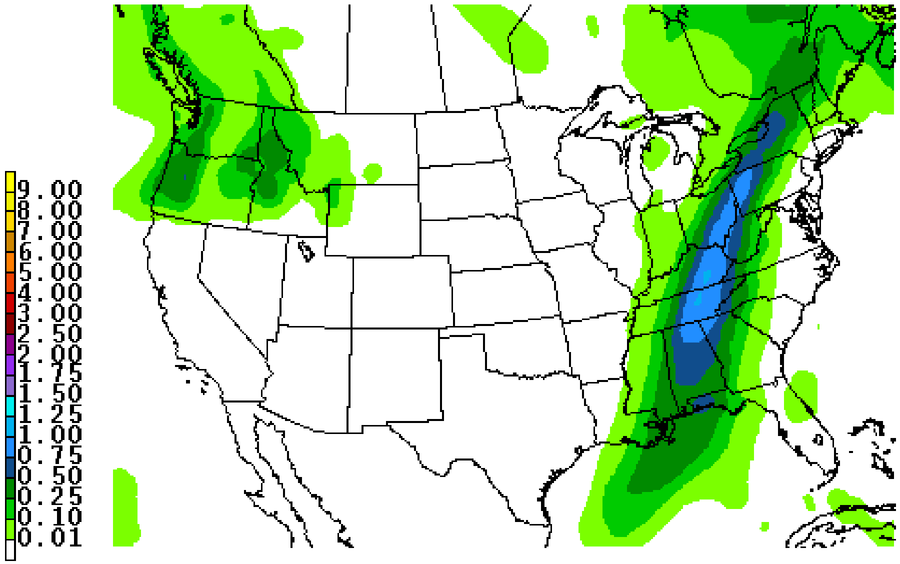
110405/0600V012 SPREAD PER UNIT MEAN



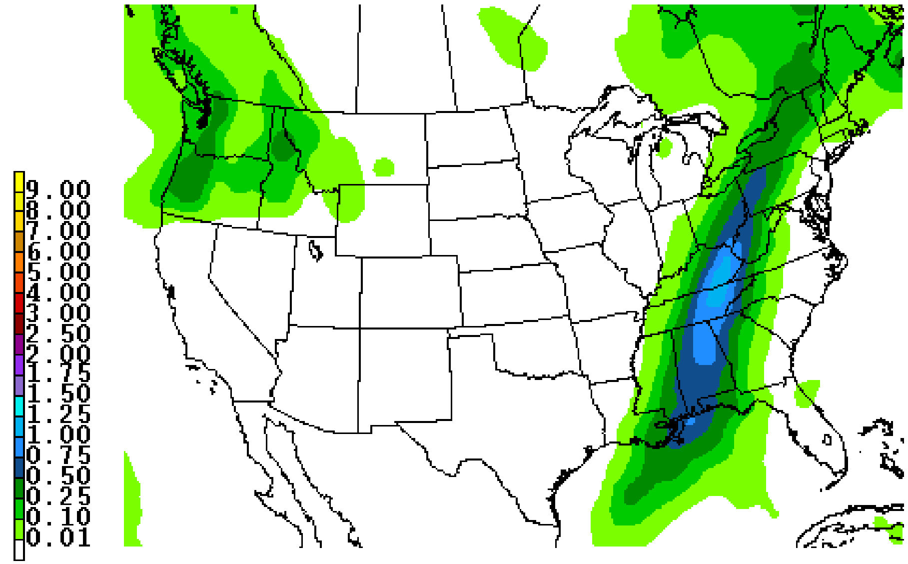
110405/0600V012 WEIGHT FOR FULL ENS MEAN



110405/0600V012 MEAN OF DETERMINISTIC RUNS

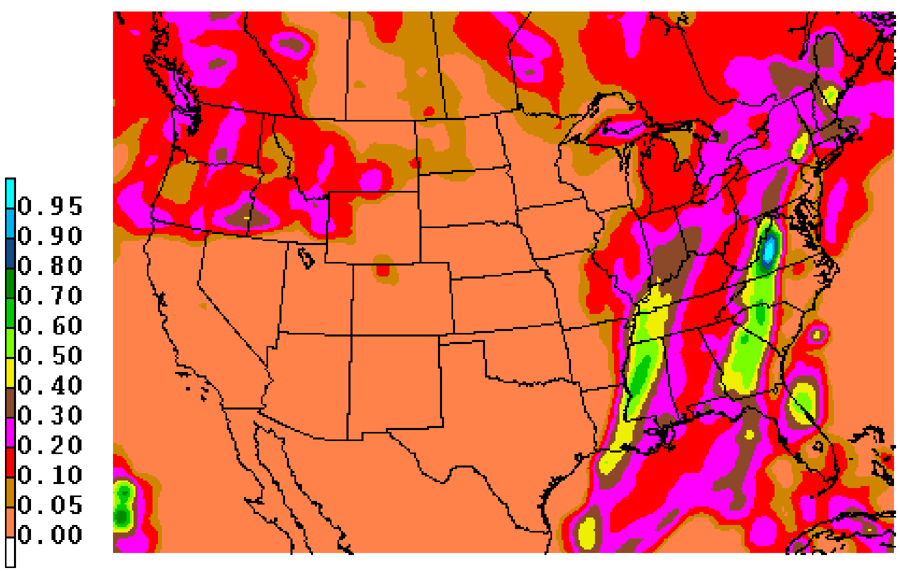


110405/0600V012 FULL ENSEMBLE MEAN

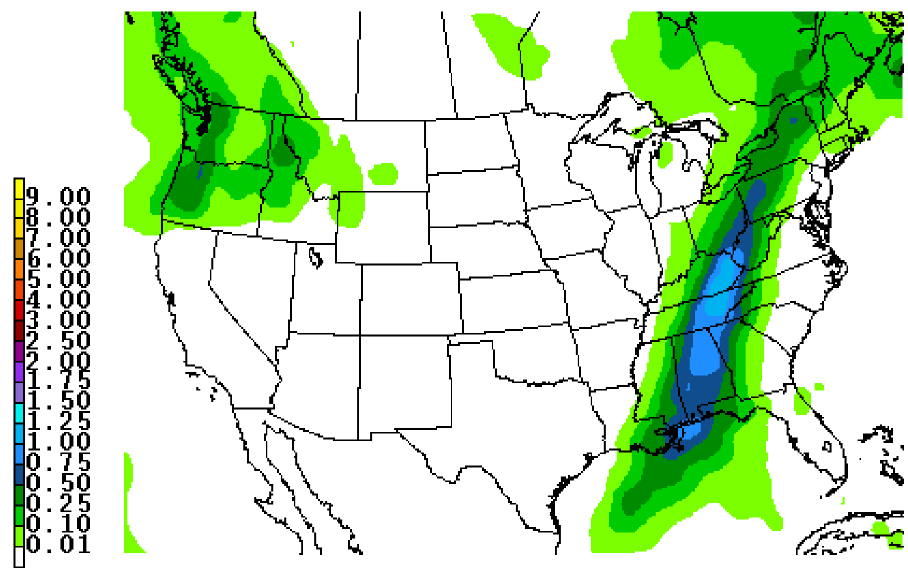


110405/0600V012 "SMART" ENSEMBLE MEAN

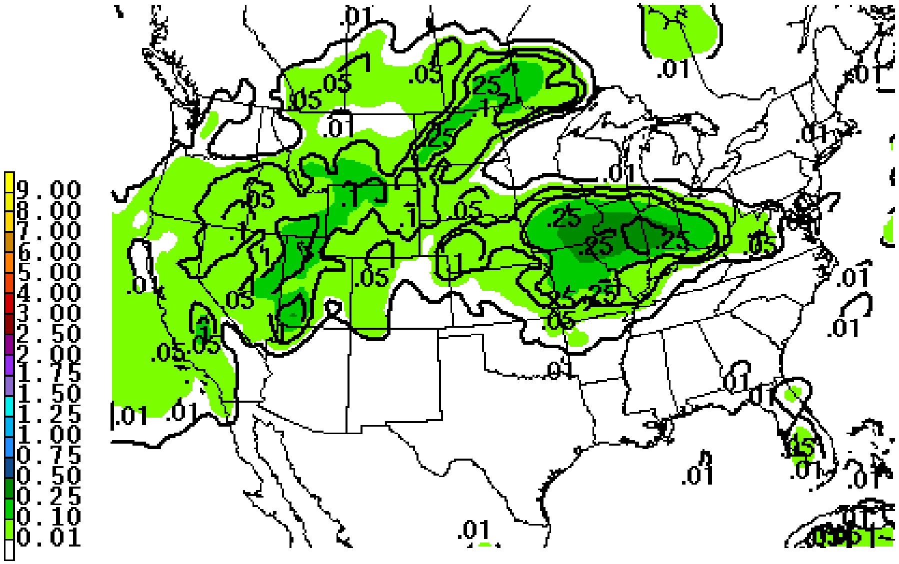
$$\mu_s = w\mu + (1-w)\mu_{dr}$$



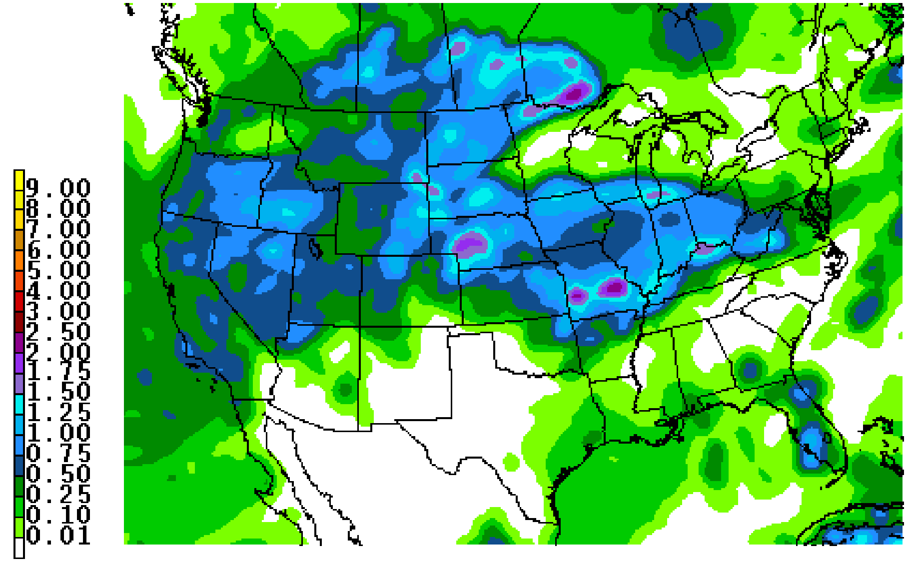
110405/0600V012 WEIGHT FOR FULL ENS MEAN



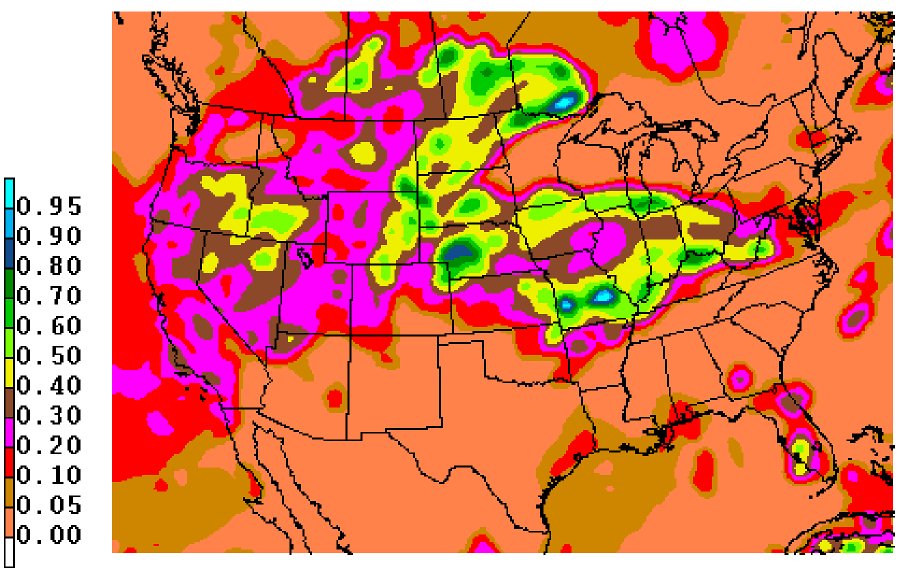
110405/0600V012 MEAN OF DETERMINISTIC RUNS



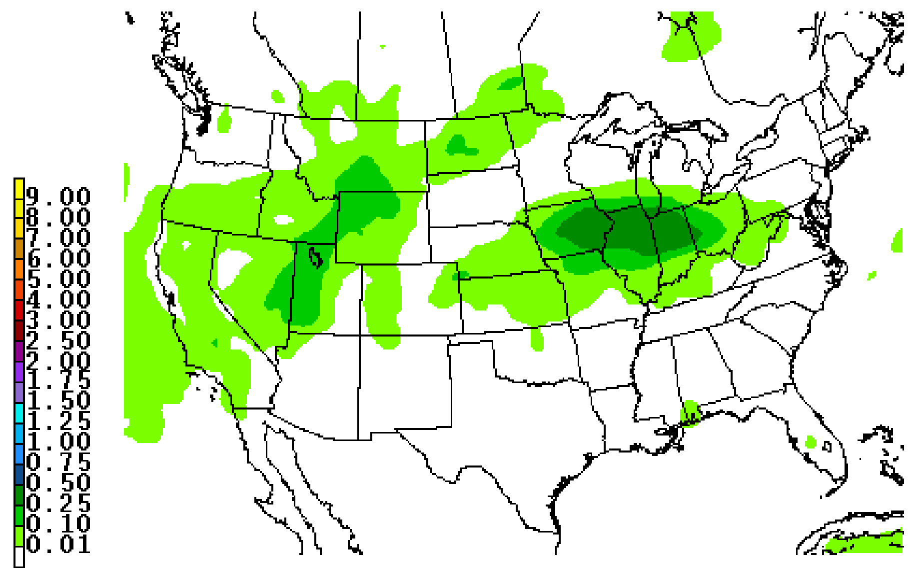
110408/0600V030 MEAN (FILL) AND SPREAD (CONTOURS)



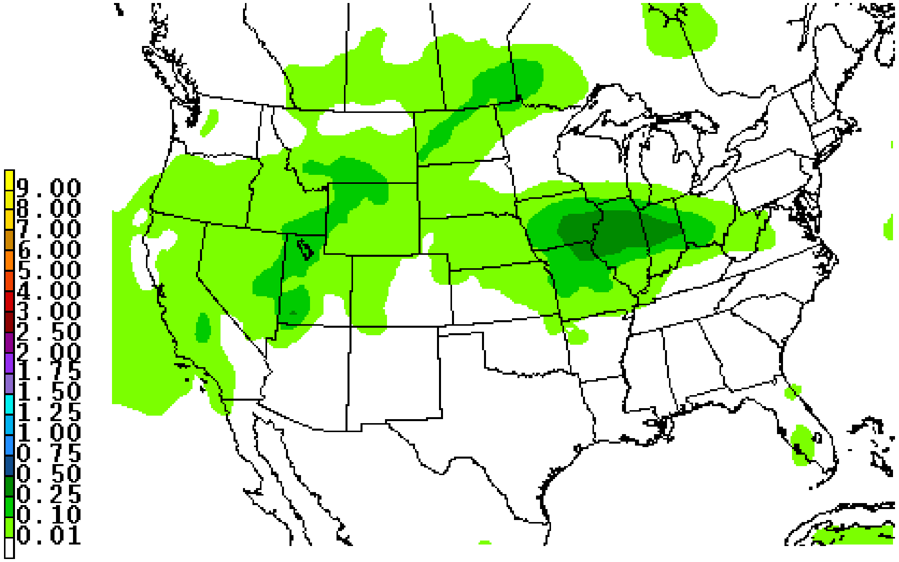
110408/0600V030 SPREAD PER UNIT MEAN



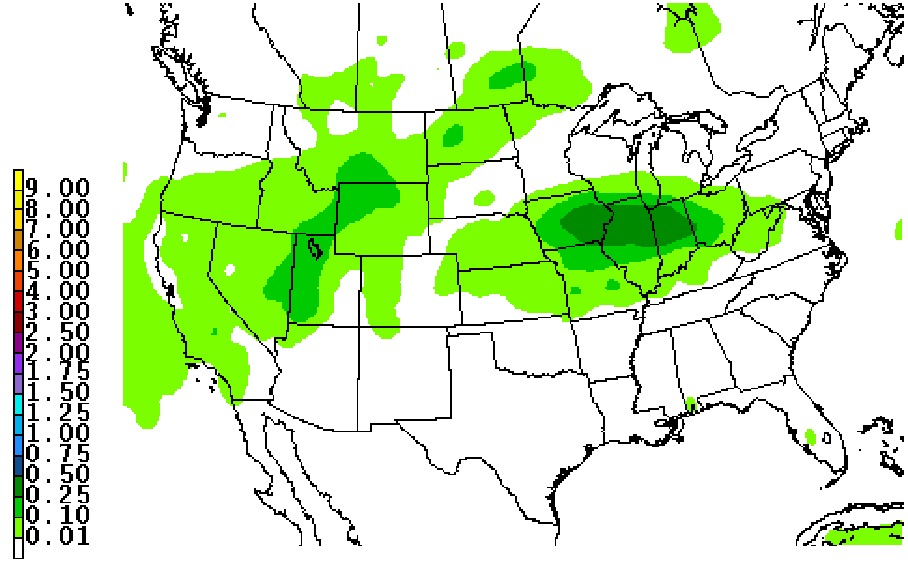
110408/0600V030 WEIGHT FOR FULL ENS MEAN



110408/0600V030 MEAN OF DETERMINISTIC RUNS

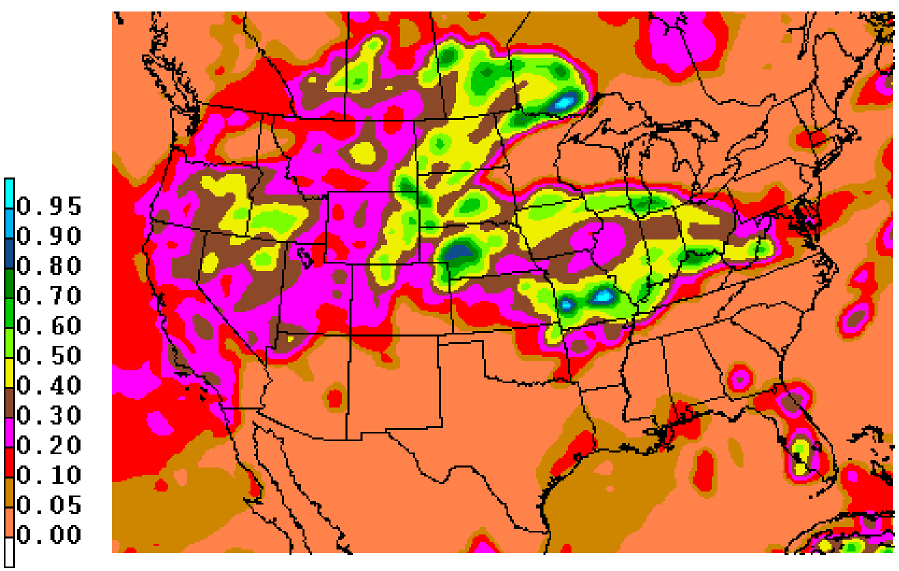


110408/0600V030 FULL ENSEMBLE MEAN

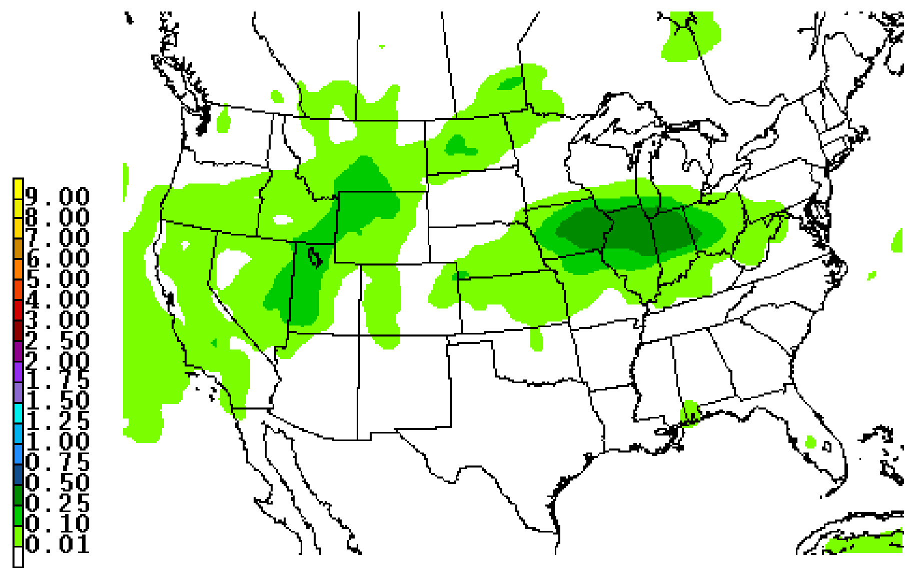


110408/0600V030 "SMART" ENSEMBLE MEAN

$$\mu_s = w\mu + (1-w)\mu_{dr}$$



110408/0600V030 WEIGHT FOR FULL ENS MEAN



110408/0600V030 MEAN OF DETERMINISTIC RUNS

We could stop and be content with the "smart" ensemble mean, but . . . there is a frequency bias issue.

What is frequency bias?

$$\text{Frequency Bias} = \frac{\text{Frequency of forecast event occurrence}}{\text{Frequency of observed event occurrence}}$$

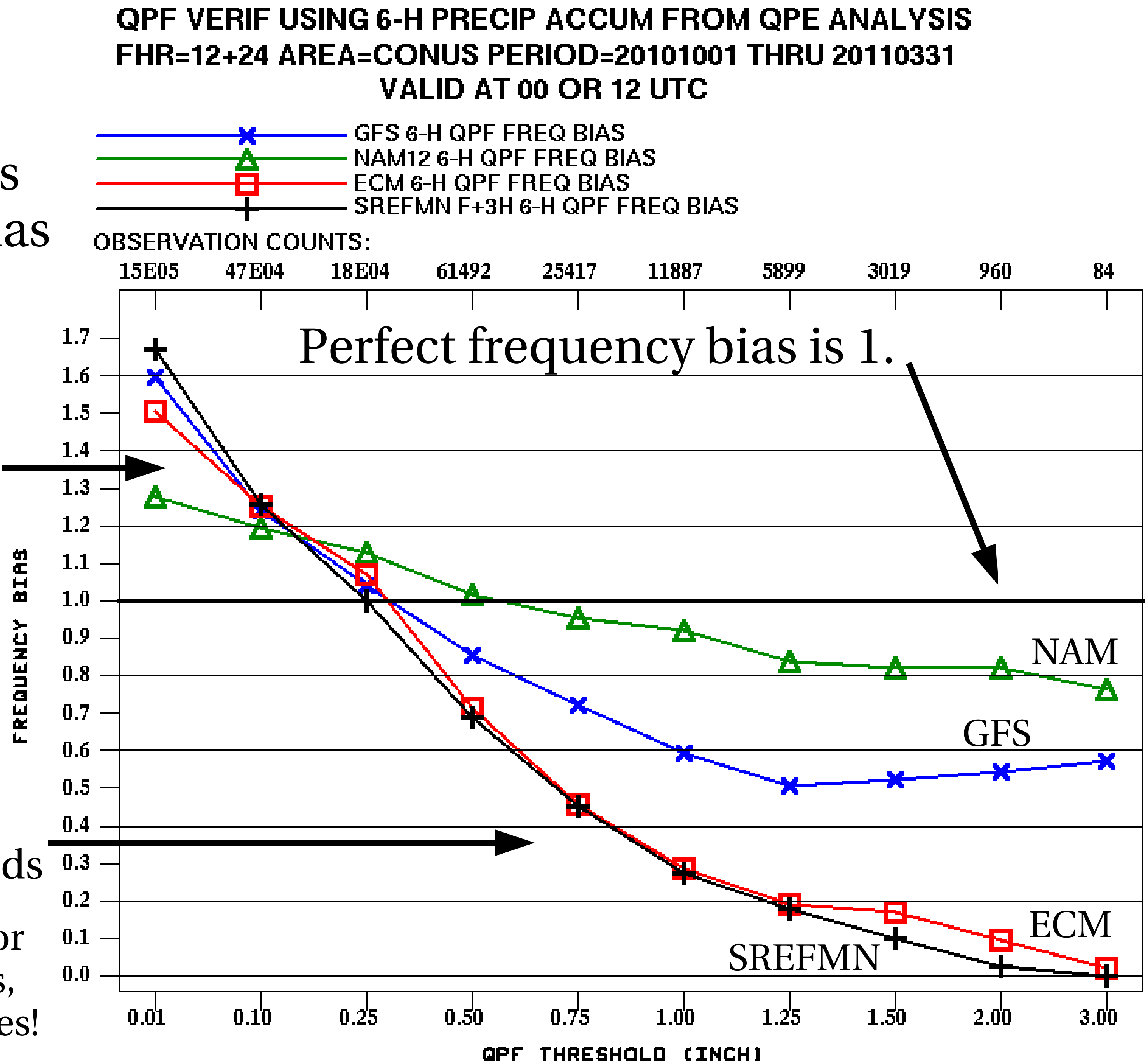
What is an event?

An event is accumulation of precipitation exceeding some specified threshold.

Perfect frequency bias has the value 1.

Frequency bias does not measure forecast skill, but it is important to consider it when assessing skill.

What is the nature of this frequency bias issue?



High bias at low thresholds

Low bias at higher thresholds

Especially true for ensemble means, even “smart” ones!

Approaches to QPF Bias Correction

- Traditional approaches: objectively based on observed data for past 30 or more days
  1. Compute a multiplicative mass balance adjustment factor (McCollor & Stull 2008), OR
  2. Pull the adjusted value from the corresponding position in the observed CDF (NCEP/EMC)
- NEW approach: subjectively based on how bias depends on thresholds for deterministic model runs having better bias characteristics
  - Compute a multiplicative correction factor
  - Enhance a target QPF using the ensemble maximum

The new approach accounts for case-to-case bias changes.

Pseudo Bias Corrected QPF (PBCQPF)

1. Initialize PBCQPF using the “smart” ensemble mean
2. Initialize an “observed” target QPF field, $q_t = \mu_{dr}$
3. Ascend through a series of thresholds*, applying a multiplicative correction factor based on a volumetric form for frequency bias (inverted):

   $$PBCQPF_{cor} = PBCQPF \times \beta, \text{ where } \beta = V_{O \geq T} / V_{F \geq T},$$

   the ratio of observed to forecast volume for amounts exceeding a given threshold, T
4. Nudge the target QPF toward the full ensemble maximum as the threshold increases because bias decreases with threshold for the deterministic runs
   - *The bias correction is computed by treating the mean of the deterministic runs enhanced by the ensemble maximum as “observed” data.*

*Thresholds used: 0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00, 1.25, 1.50, 1.75, 2.00, 2.50, and 3.00 inches

Red area represents volume of QPF *where* the 1 inch threshold is exceeded.

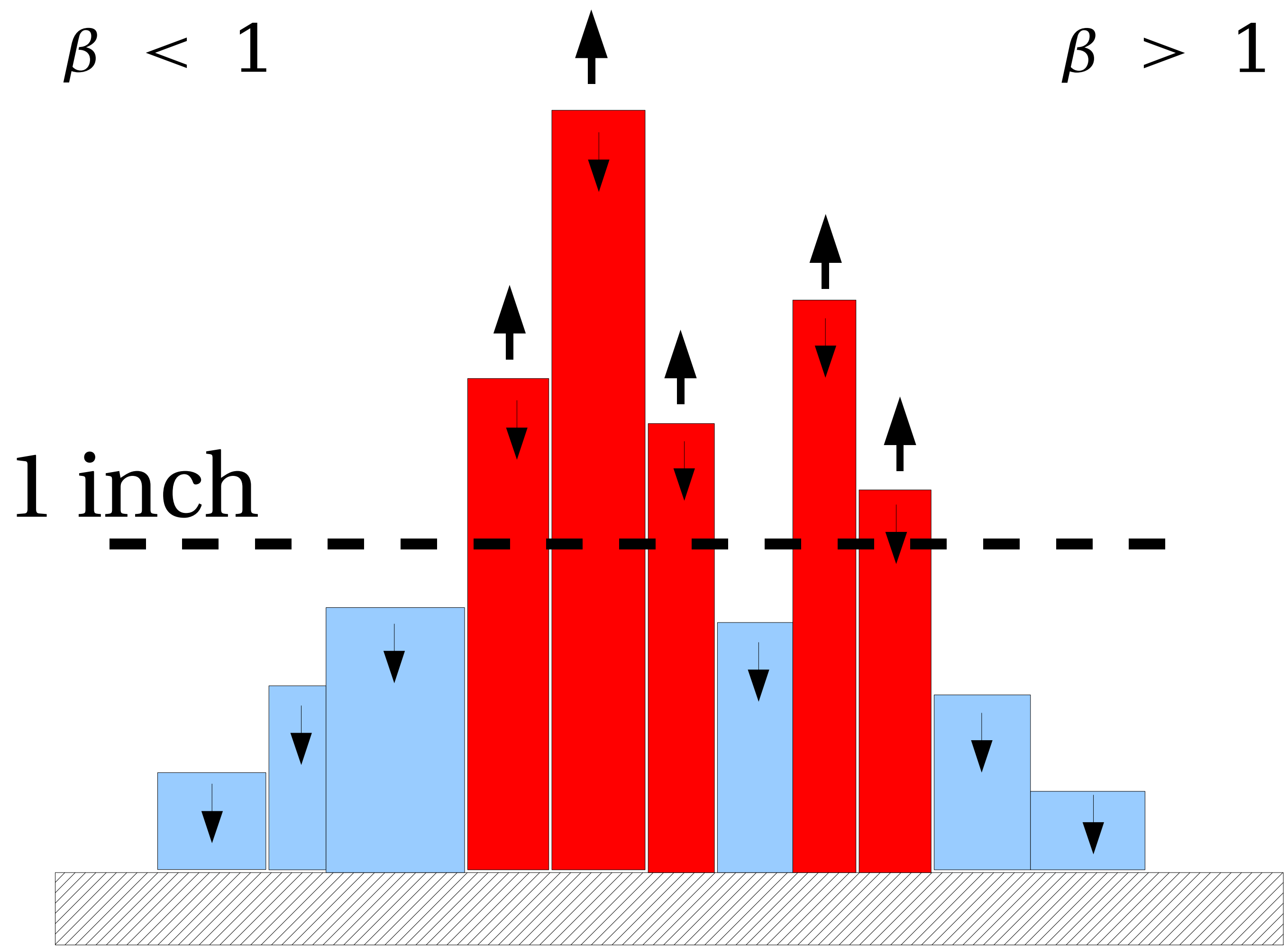
↓ Change direction to correct over bias

$\beta < 1$

OR

↑ Change direction to correct under bias

$\beta > 1$



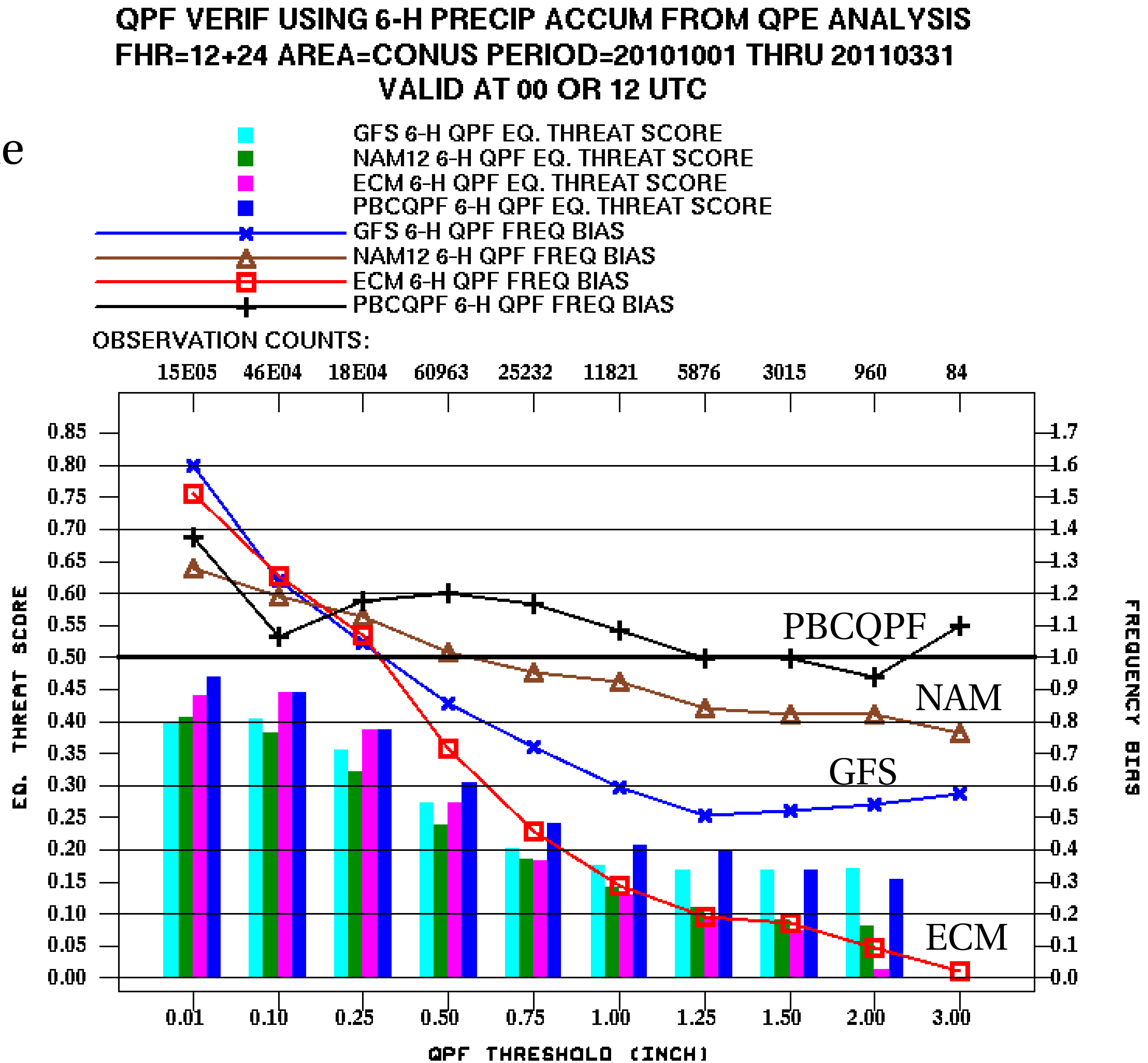
Adjusting the target QPF:

- Why? Even the mean of the deterministic runs suffers from low bias at higher thresholds.
- How? Nudge the target QPF, q_t, toward the full ensemble maximum QPF by applying the following for each threshold above .10 inch:

$$q_t = (1-\alpha) q_t + \alpha q_{max}, \ \textit{where} \ \alpha = .05$$

Thresholds used: 0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00, 1.25, 1.50, 1.75, 2.00, 2.50, and 3.00 inches

How does the PBCQPF perform in terms of bias & ETS?

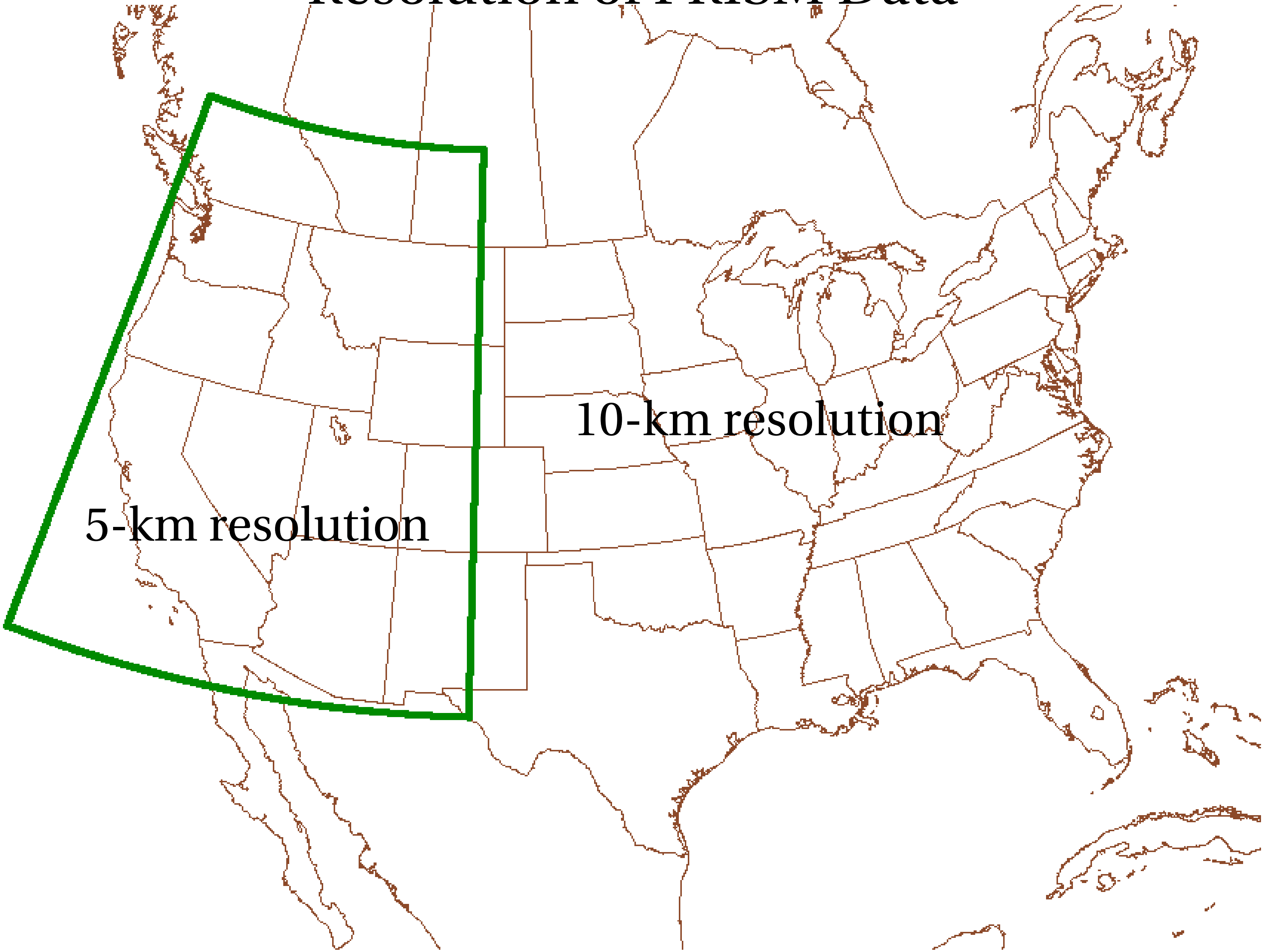


The Downscaling

- Based on the Parameter-elevation Regressions on Independent Slopes Model (PRISM) monthly precipitation data
  - 5-km resolution over western U.S.
  - 10-km resolution over eastern U.S.
- Applied on the 32-km HPC QPF grid
- Uses a multiplicative downscaling factor:

$$PBCQPF_{ds} = PBCQPF \times f_{ds}$$

Resolution of PRISM Data



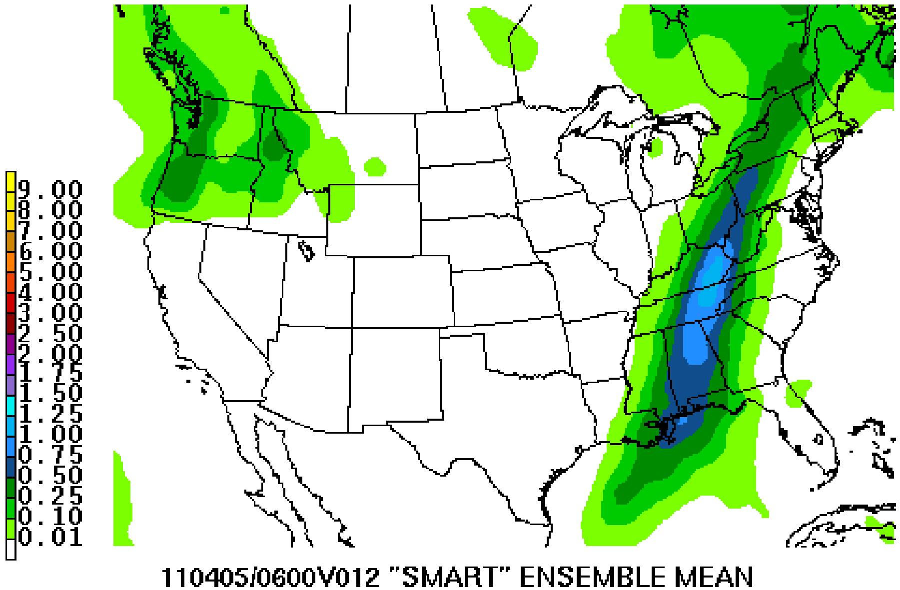
WESTERN 5-KM PRISM ZONE

Determining the Downscaling Factor

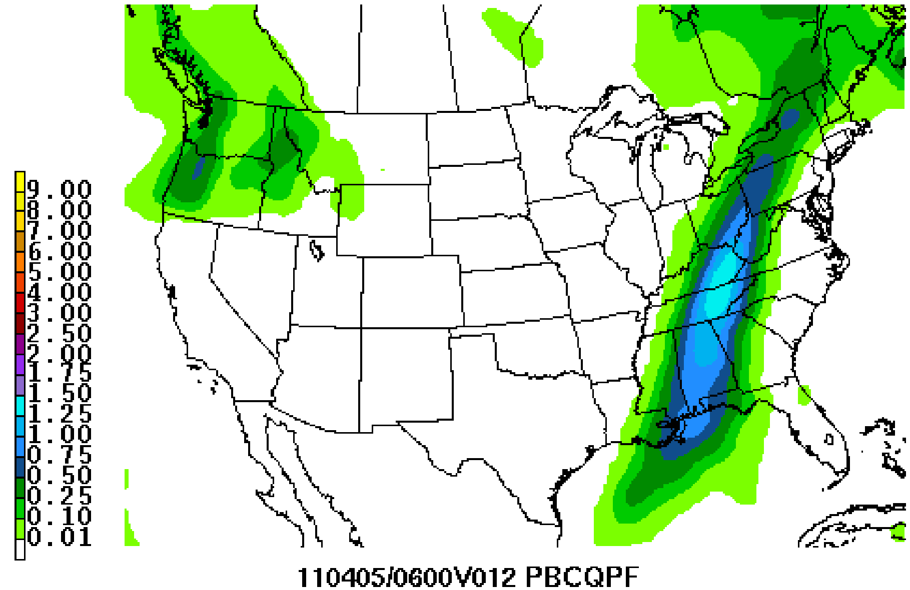
- Remap the monthly PRISM data to the 32-km QPF grid using the GEMPAK area average preserving remapping
- Remap the 32-km PRISM data back to its original high-resolution grid (effectively a bi-linear interpolation)
- Compute the ratio of the original PRISM data on the high-resolution grid to the back-interpolated data
- Transfer these ratio values from the high-resolution grid to the 32-km QPF grid ***using nearest-point assignment***—these are the downscaling factors
- Impose **lower bounds** on the downscaling factors by month (reduce rain shadowing effects during warm season):

Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
0.3	**0.3**	**0.4**	**0.4**	**0.5**	**0.7**	**0.9**	**0.9**	**0.7**	**0.5**	**0.4**	**0.3**

Example showing pseudo bias correction and downscaling . . .

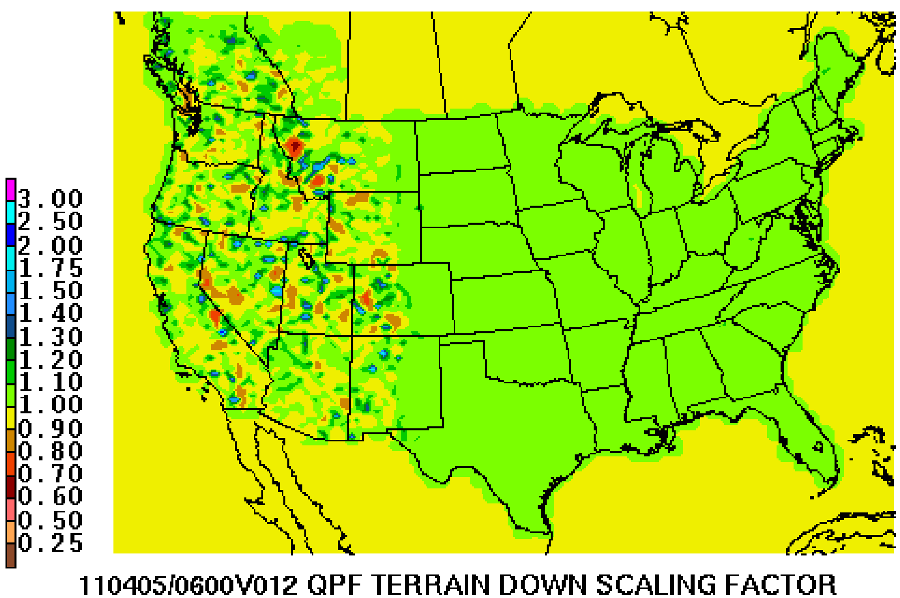


“smart” ensemble mean

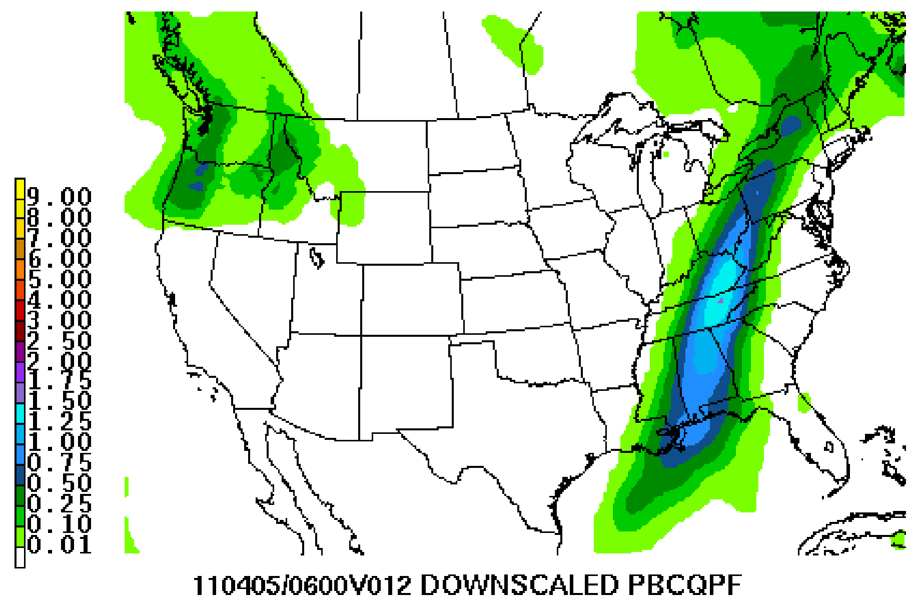


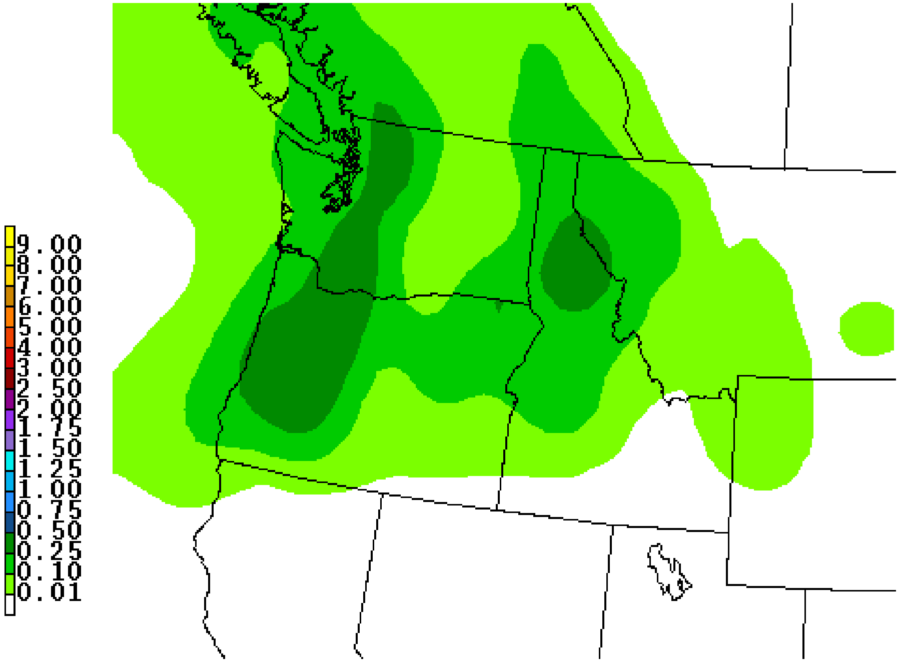
pseudo bias corrected mean

downscaling factor

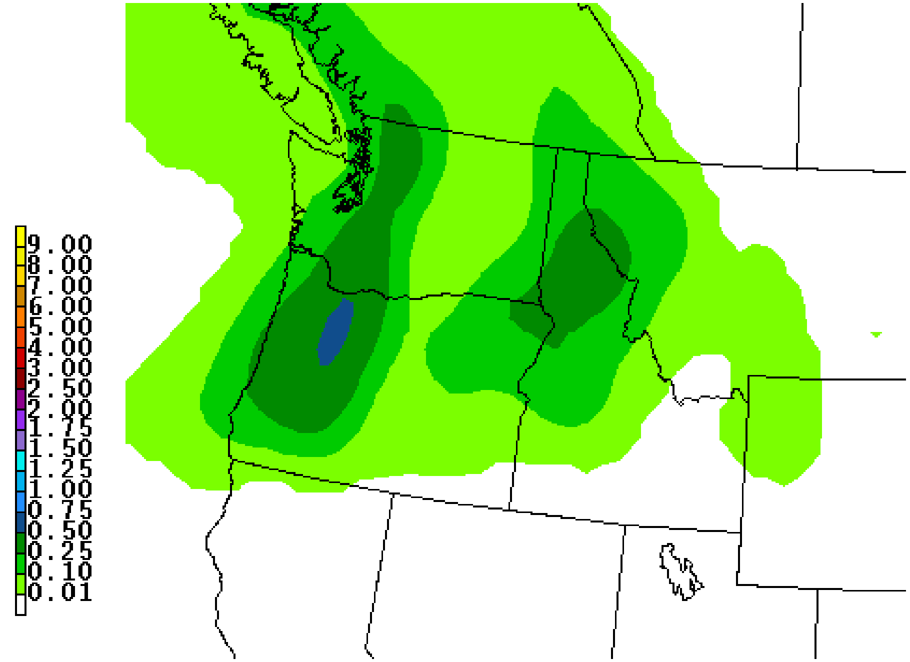


downscaled pseudo bias corrected mean

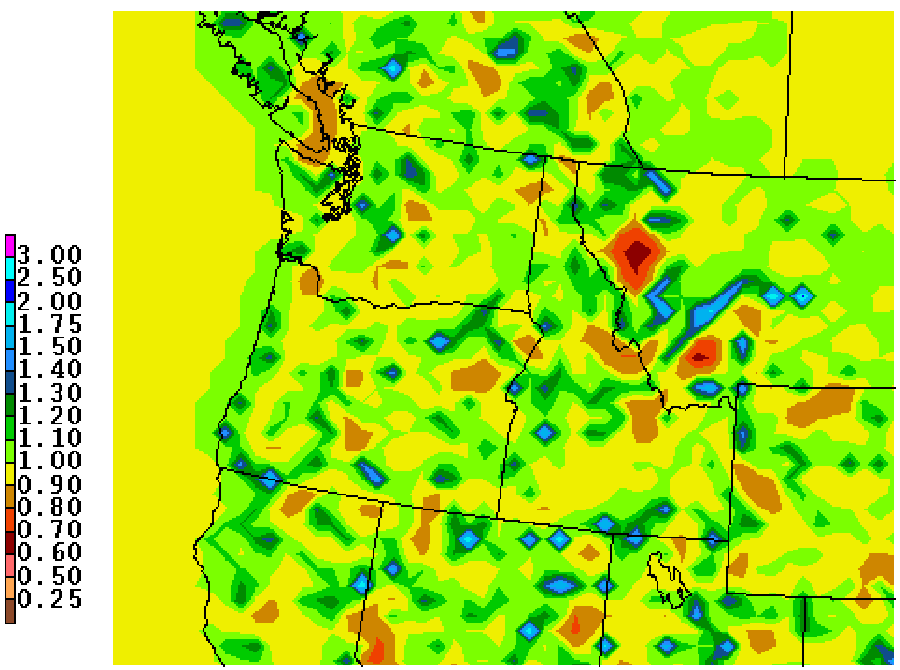




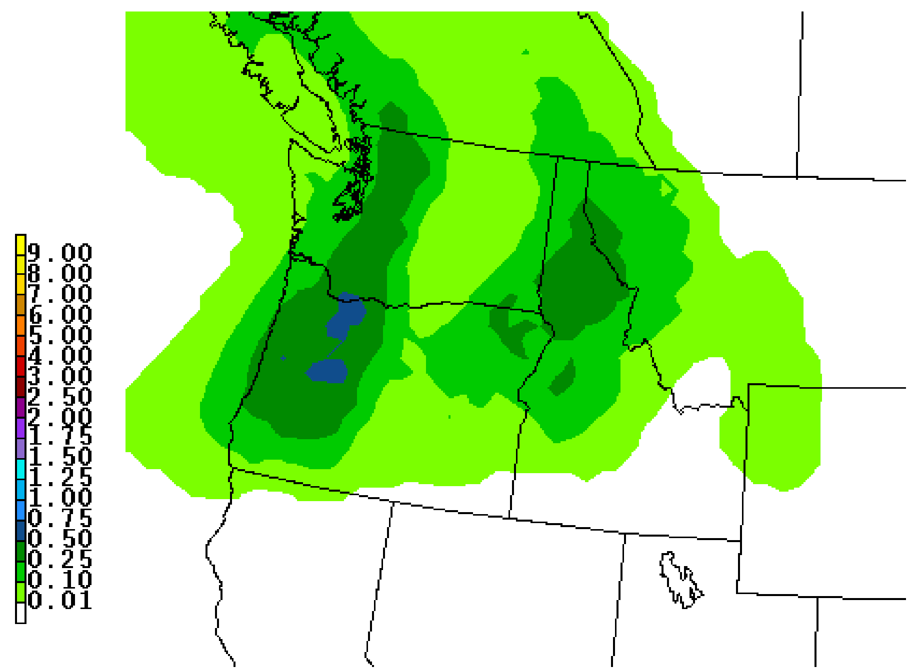
110405/0600V012 "SMART" ENSEMBLE MEAN



110405/0600V012 PBCQPF



110405/0600V012 QPF TERRAIN DOWN SCALING FACTOR



110405/0600V012 DOWNSCALED PBCQPF

The Verification Experiment

- Addresses this question: **Can an automated QPF using the latest available models and issued prior to operational deadlines be as good as the human forecast?**
- Breaks with the HPC's typical practice of comparing model-based forecasts against human forecasts with a time lag to allow human consideration
- Setup the experiment for the two “final” cycles at 00 and 12 UTC for $PBCQPF_{ds}$ vs HPC QPF

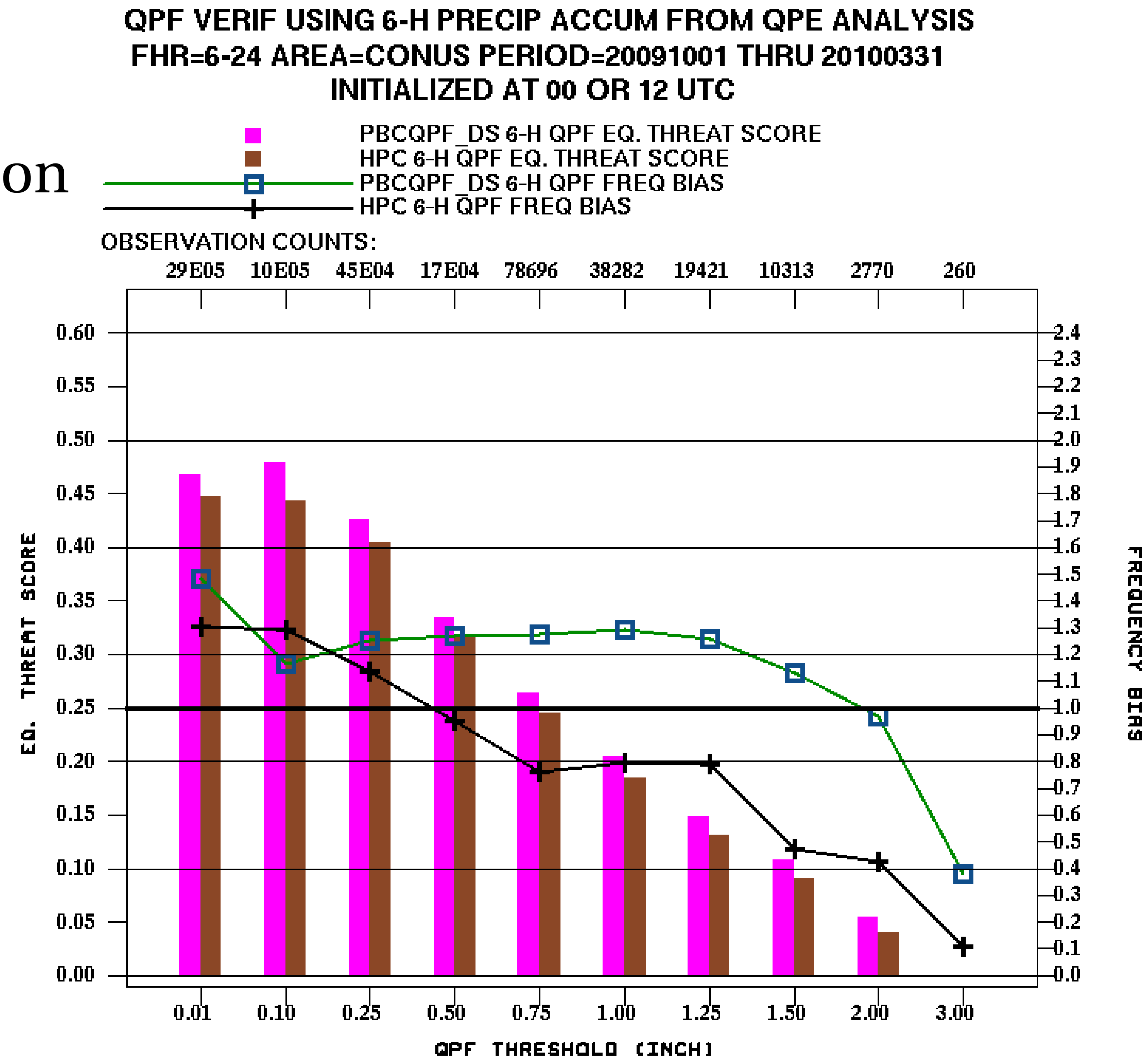
Experiment Product Summary Table for 12 UTC (add 12 hours for 00 UTC)

Source	Projection	Deadline	Primary Guidance
HPC	Day 1	1000 UTC	00 & 06 Z NAM; 00 Z GFS; D-1 12 & 00 Z ECM; D-1 21 & 03 Z SREF
PBCQPF	Day 1	0900 UTC	06 Z NAM12; 00 Z GFS; 00 Z ECM; 03Z SREF
HPC	Day 2	0600 UTC	00 Z NAM; 00 Z GFS; D-1 12 Z ECM; D-1 21 Z SREF
PBCQPF	Day 2	0415 UTC	00 Z NAM; 00 Z GFS; D-1 12 Z ECM; D-1 21 Z SREF
HPC	Day 3	0800 UTC	00 Z NAM; 00 Z GFS; D-1 12 Z ECM; D-1 21 Z SREF
PBCQPF	Day 3	0415 UTC	00 Z NAM; 00 Z GFS; D-1 12 Z ECM; D-1 21 Z SREF

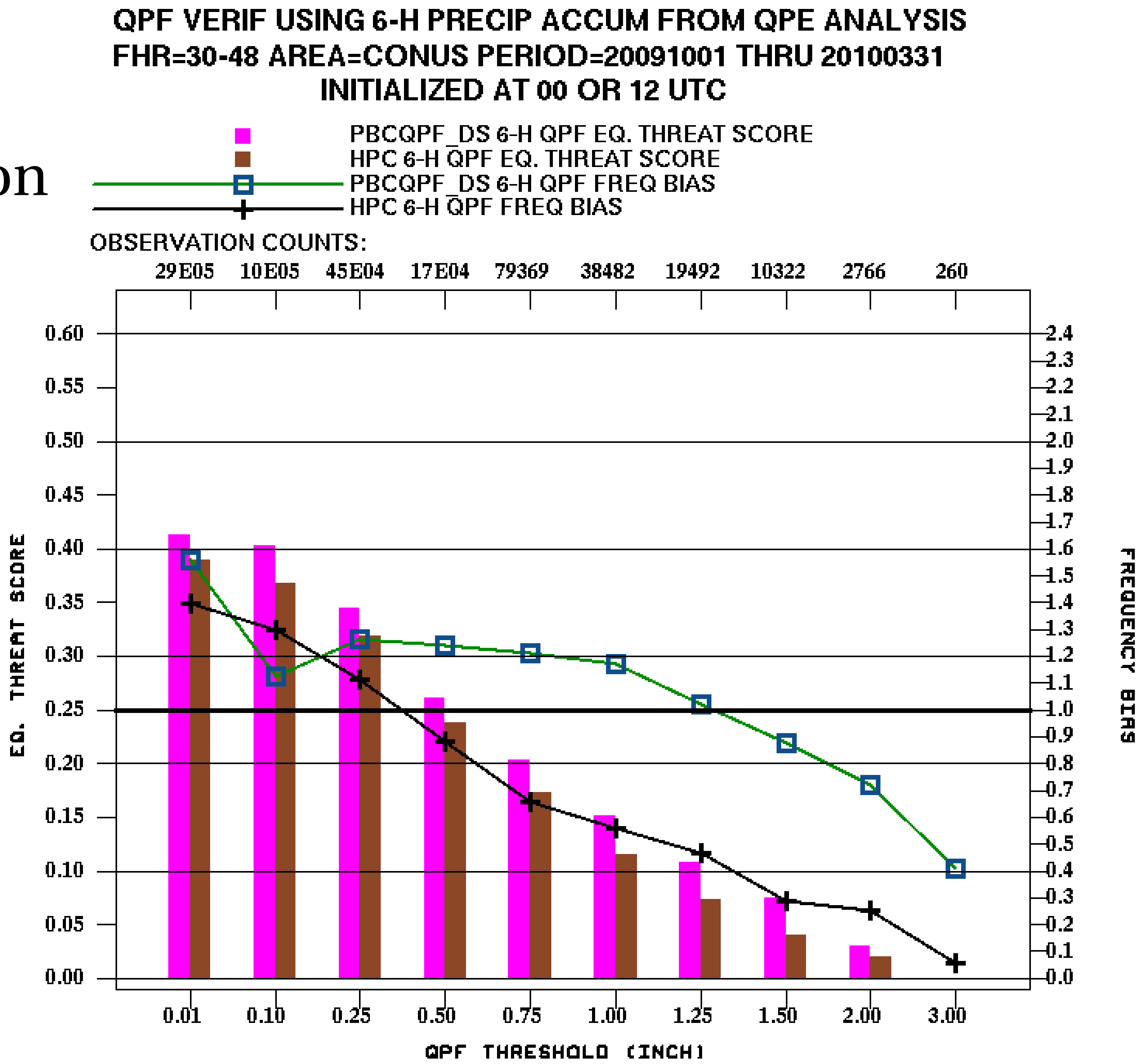
Verification of the 6-h QPFs

- Verify against the **QC'ed 6-h RFC QPE analyses** from the NPVU remapped to 32-km resolution
- Examine *equitable threat score* and frequency bias as a function of thresholds
- Combine forecasts initialized from 00 or 12 UTC
  - For FY10 cold season: October 2009—March 2010
    - Little diurnal variation
    - Combine all four 6-h forecasts for each projection day 1—3
  - For FY10 warm season: April—September 2010
    - Noticeable diurnal variation for Day 1; so, look at separate combinations of 12-h + 24-h and 06-h + 18-h forecasts
    - Combine all four 6-h forecast for each projection day 2—3

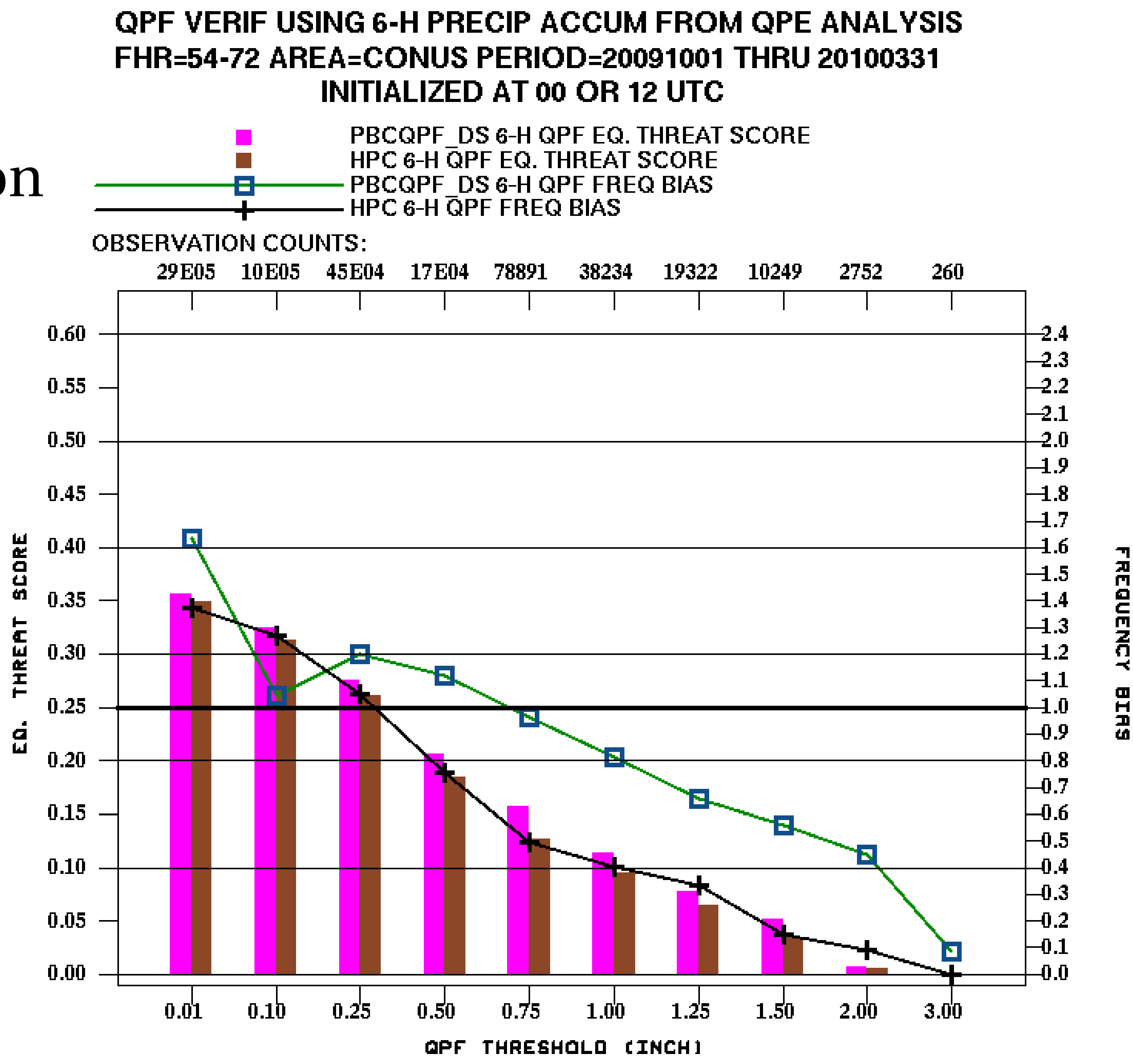
Day 1
Cold Season



Day 2
Cold Season

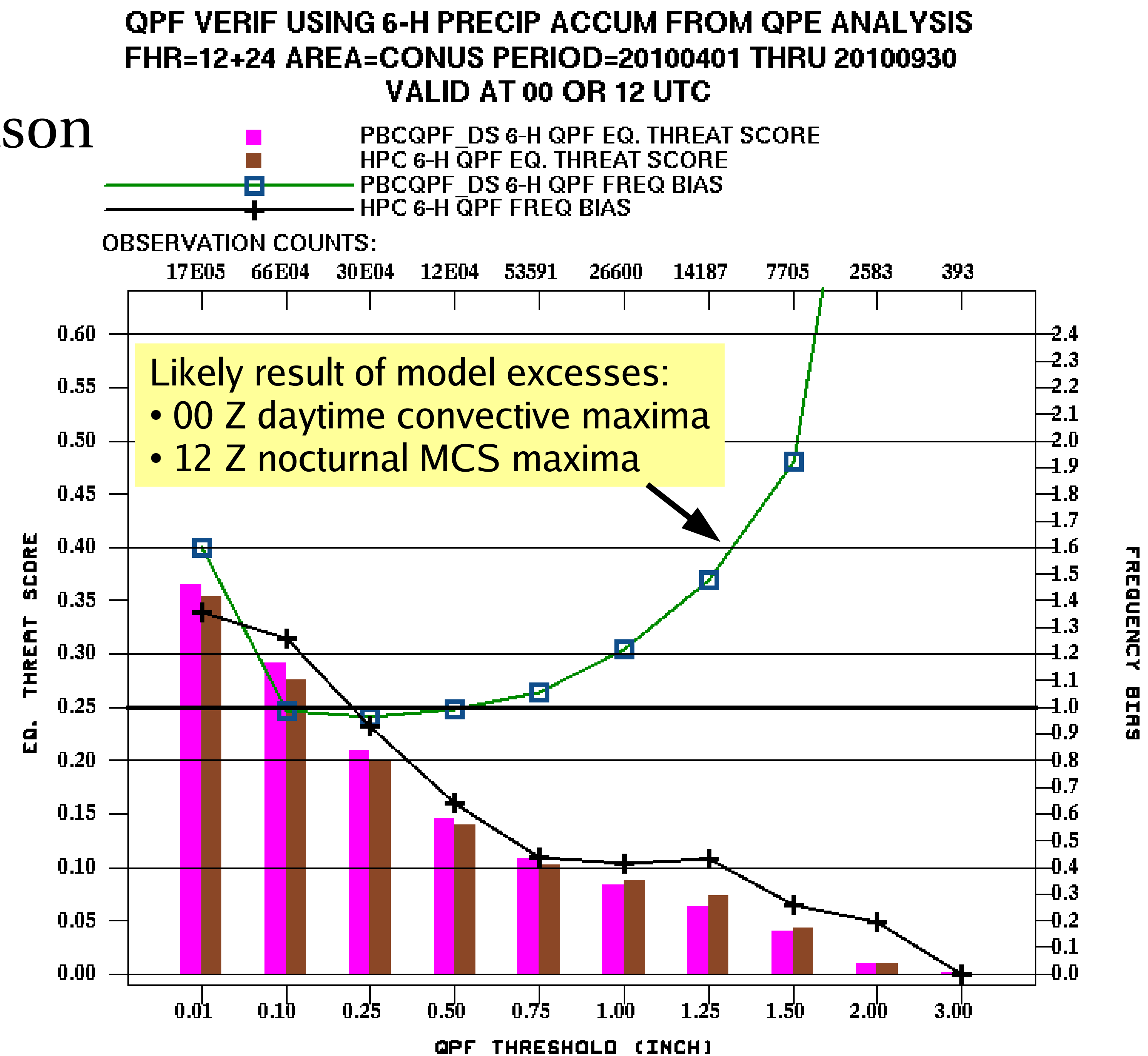


Day 3
Cold Season



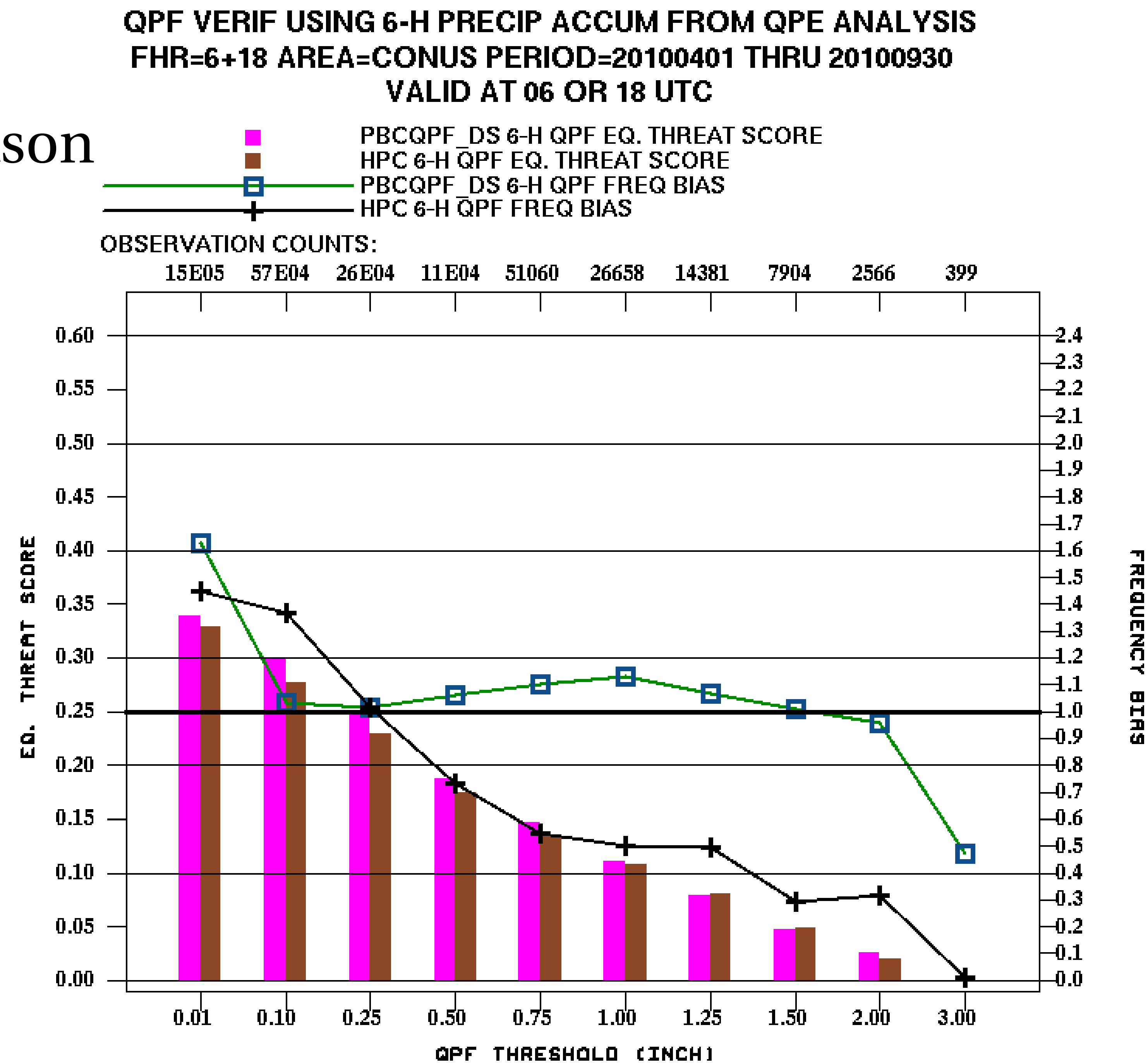
Day 1
Warm Season

Forecasts valid
at 00 & 12 Z

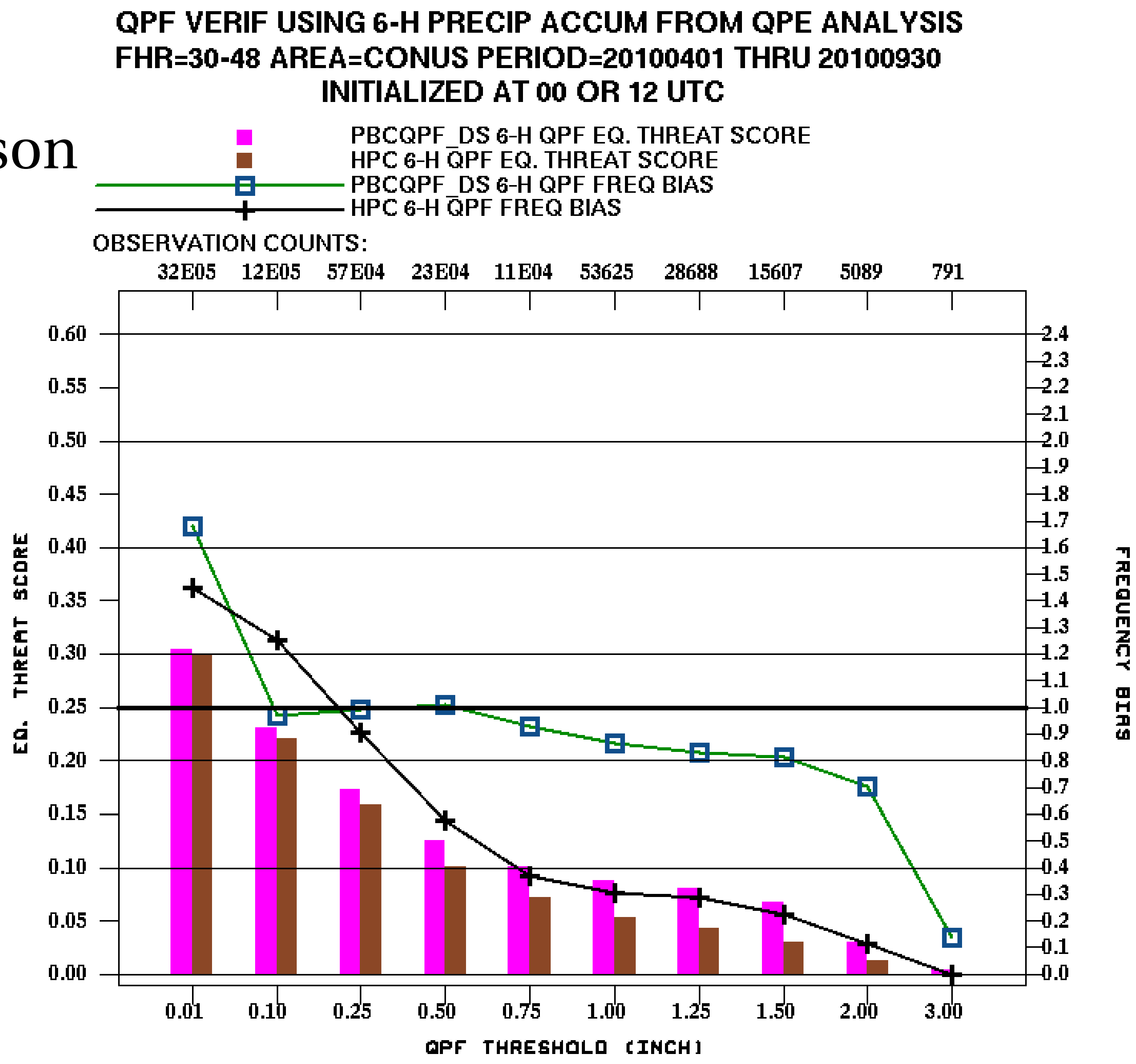


Day 1
Warm Season

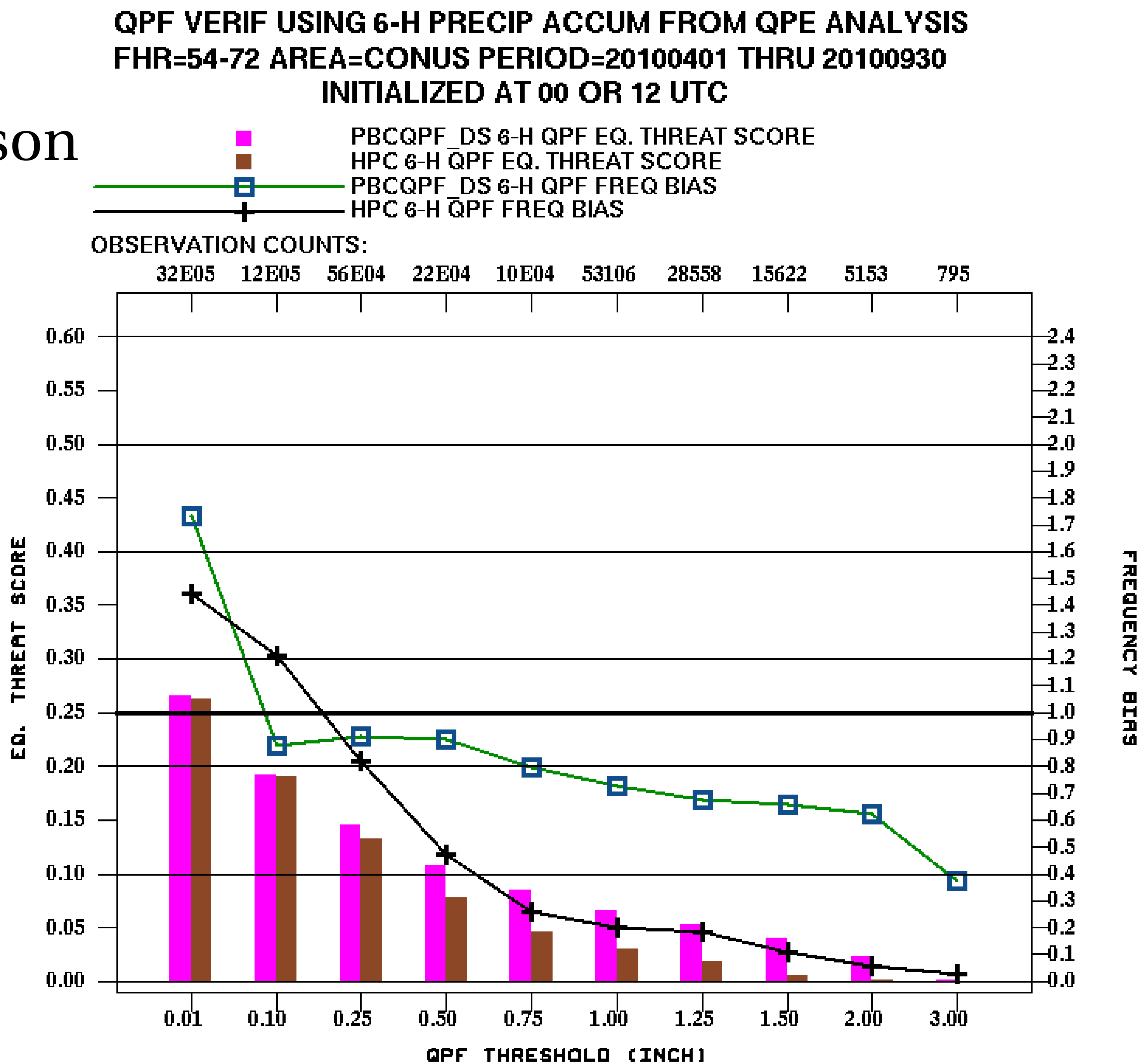
Forecasts valid
at 06 & 18 Z



Day 2
Warm Season



Day 3
Warm Season



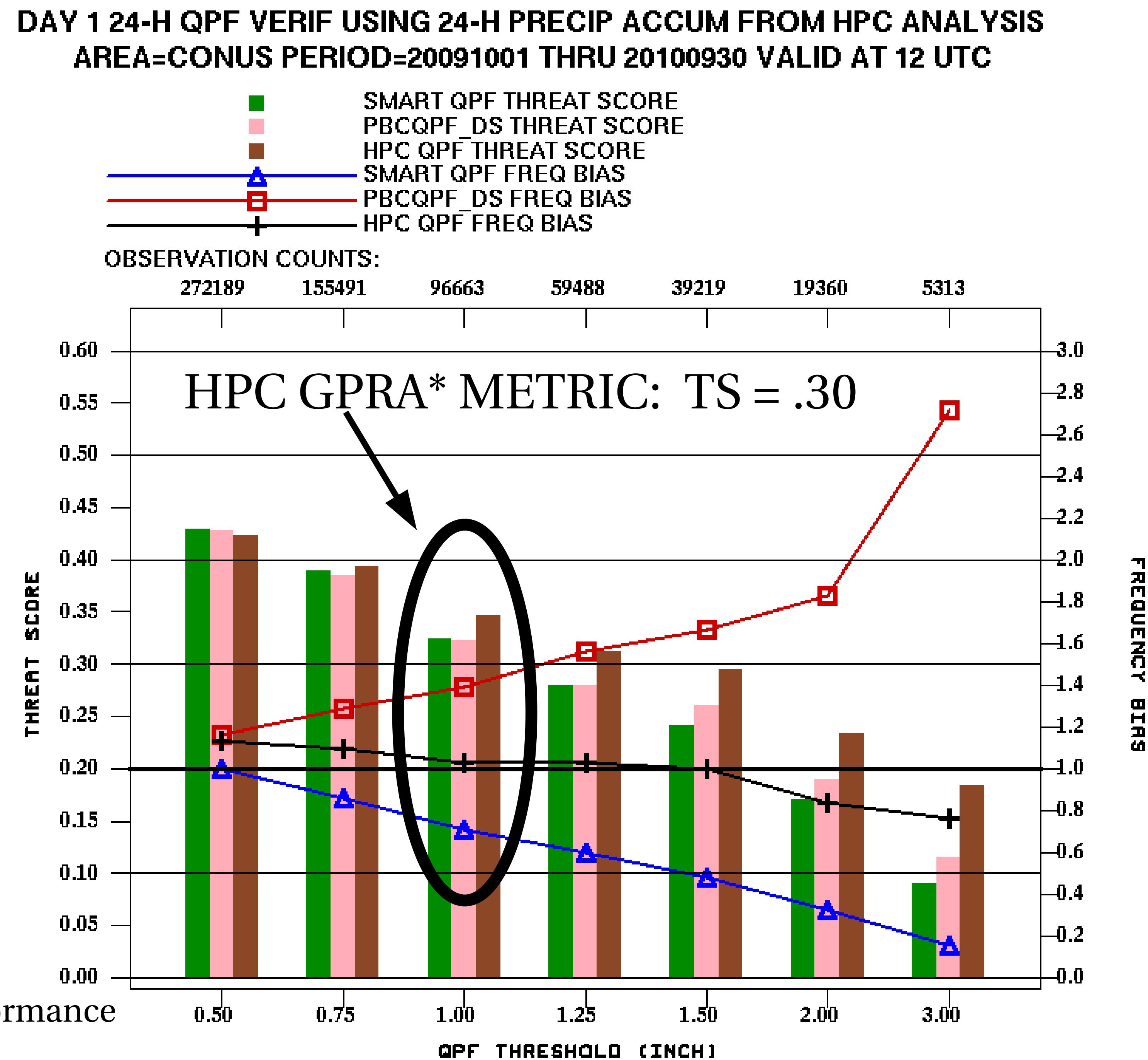
6-h QPF Verification Results

- Cold Season
  - PBCQPF_DS has consistently higher ETS versus HPC QPF
  - PBCQPF_DS bias is excessive for Day 1
  - HPC QPF is under biased for thresholds **0.50”** and above
- Warm Season
  - PBCQPF_DS exhibits excessive bias for QPF valid at 00 & 12 Z on Day 1
  - HPC QPF has higher ETS versus PBCQPF_DS only at higher thresholds for Day 1; otherwise, PBCQPF_DS is superior
  - HPC QPF is under biased for thresholds **0.25”** and above
- Both Seasons (with exceptions noted above)
  - Both QPFs tend toward under bias at the high thresholds
  - Both QPFs tend toward over bias at the lower thresholds

Verification of 24-h QPF

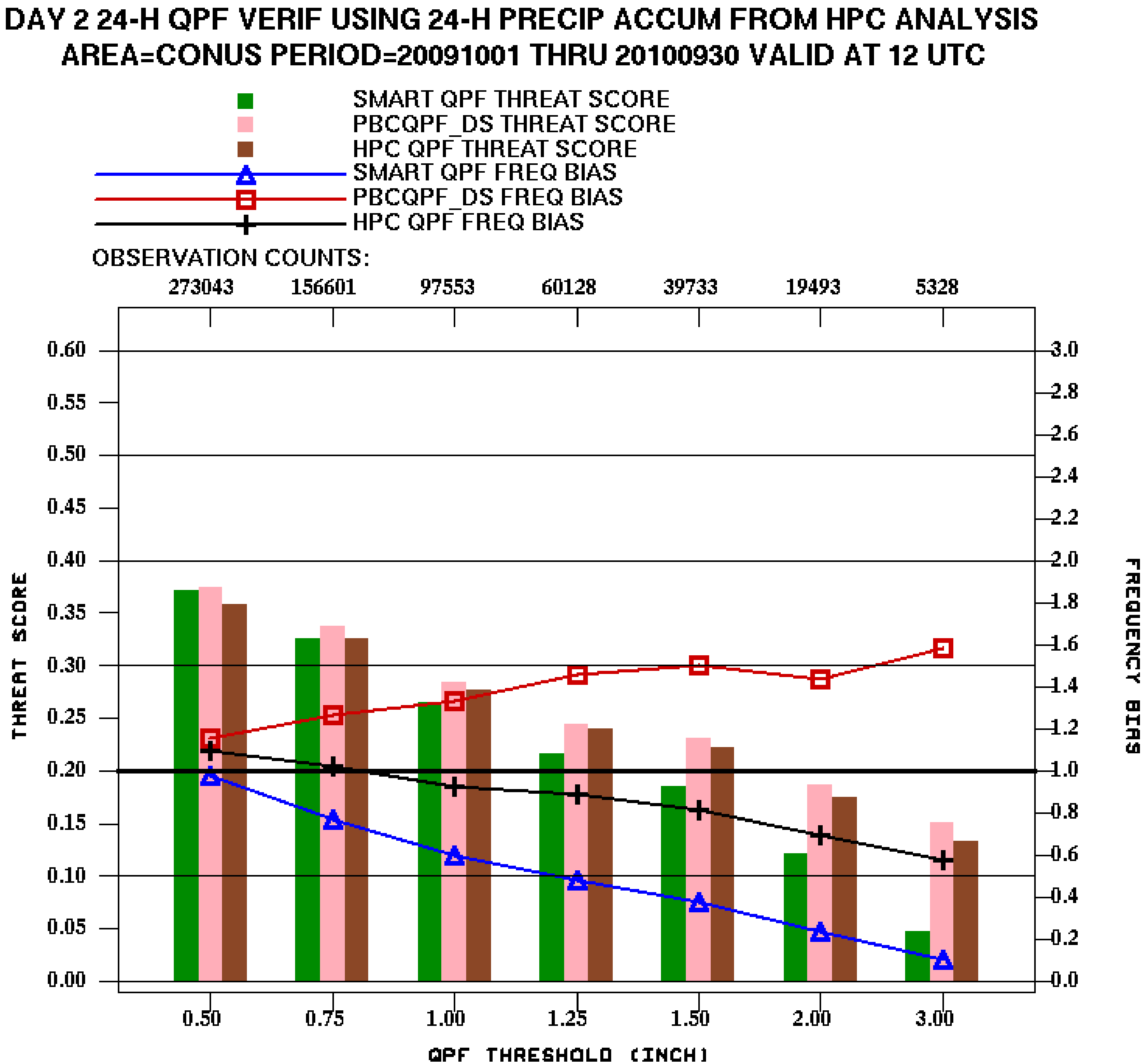
- Verify 24-h accumulations of 6-h QPFs against the **HPC manual analysis** having 32-km resolution valid at 12 Z
- Examine *threat score* and frequency bias for thresholds .50 inch and higher
- Combine contingency tables over one year: FY10—01 October 2009 through 30 September 2010
- Compare these three QPFs:
  (1) "Smart" ensemble mean (ensqpf in nmap2)
  (2) PBCQPF_DS (ensqpfbc in nmap2)
  (3) HPC QPF

Day 1

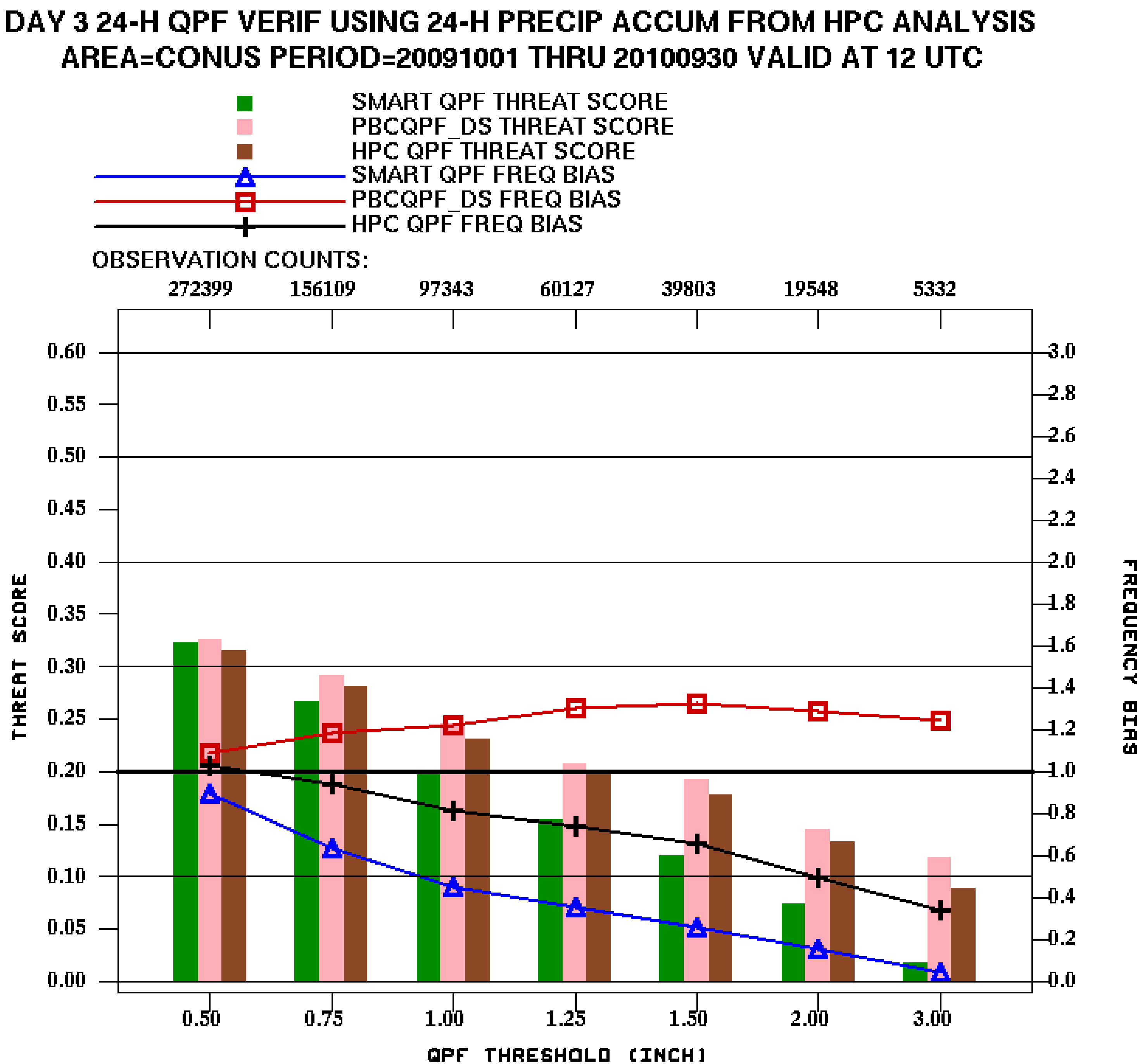


*Government Performance and Results Act

Day 2


DAY 2 24-H QPF VERIF USING 24-H PRECIP ACCUM FROM HPC ANALYSIS
AREA=CONUS PERIOD=20091001 THRU 20100930 VALID AT 12 UTC
SMART QPF THREAT SCORE
PBCQPF_DS THREAT SCORE
HPC QPF THREAT SCORE
SMART QPF FREQ BIAS
PBCQPF_DS FREQ BIAS
HPC QPF FREQ BIAS
OBSERVATION COUNTS:
273043
156601
97553
60128
39733
19493
5328
THREAT SCORE
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
FREQUENCY BIAS
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.50
0.75
1.00
1.25
1.50
2.00
3.00
QPF THRESHOLD (INCH)

Day 3


DAY 3 24-H QPF VERIF USING 24-H PRECIP ACCUM FROM HPC ANALYSIS
AREA=CONUS PERIOD=20091001 THRU 20100930 VALID AT 12 UTC
SMART QPF THREAT SCORE
PBCQPF_DS THREAT SCORE
HPC QPF THREAT SCORE
SMART QPF FREQ BIAS
PBCQPF_DS FREQ BIAS
HPC QPF FREQ BIAS
OBSERVATION COUNTS:
272399
156109
97343
60127
39803
19548
5332
THREAT SCORE
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
FREQUENCY BIAS
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.50
0.75
1.00
1.25
1.50
2.00
3.00
QPF THRESHOLD (INCH)

24-h QPF Verification Results

- HPC QPF outperforms PBCQPF_DS only for Day 1
- Day 3 is a solid win for PBCQPF_DS
- PBCQPF_DS excessive bias for Day 1 is likely due in part to model excesses
- The "smart" ensemble mean performs weakest and is under biased
- Excessive bias indicated for PBCQPF_DS and increased HPC QPF bias at higher thresholds suggests a tendency for *under bias* in the HPC analysis considering the results for 6-h QPF

Verification Summary

- Can an automated QPF using the latest available models and issued prior to operational deadlines be as good as the human forecast?*
  - For Day 3 – Yes!
  - For Day 2 – Maybe?
  - For Day 1 – Not yet.
- Gives an indication of adjustments that could make the PBCQPF_DS better . . .

*Imposition of forecast continuity by human forecasters is not considered in this evaluation.

Possible pseudo bias correction adjustments:

- Change initialization procedures for PBCQPF and target QPF to reduce bias at lowest thresholds
- Nudge toward the full ensemble 90^{th} percentile rather than the maximum (avoid grid-scale precip "bombs" or convective excesses)
- Begin the nudging at a higher value of the threshold (e.g., at .25" rather than .15")
- Compute a separate bias correction factor for western third of CONUS. . . may be tricky

Summary

- A normalized spread measures uncertainty.
- The “smart” ensemble mean weights deterministic runs more where uncertainty is relatively low.
- The bias correction is computed by treating the mean of the deterministic runs enhanced by the ensemble maximum as “observed” data.
- Verification of 6-h and 24-h forecasts indicates comparable or superior skill compared to human forecasts in most cases, but with too much bias at lowest several thresholds
- Verification information points to possible improvements to the pseudo bias correction.

Future Work

- Refine the pseudo bias correction as described earlier
- ✓ Update the deterministic runs to use the finest resolution now available
- Derive a dynamic downscaling approach using high-resolution (e.,g., 4-km) model runs in combination with ensemble information
- Investigate the properties of the HPC 24-h manual precipitation accumulation analysis—cursory examination suggests a low bias at thresholds beyond 1 inch