An evaluation of the probabilistic information in multi-model ensembles

David Unger

Huug van den Dool , Malaquias Pena, Peitao Peng, and Suranjana Saha

30th Annual Climate Prediction and Diagnostics Workshop

October 25, 2005

Objective

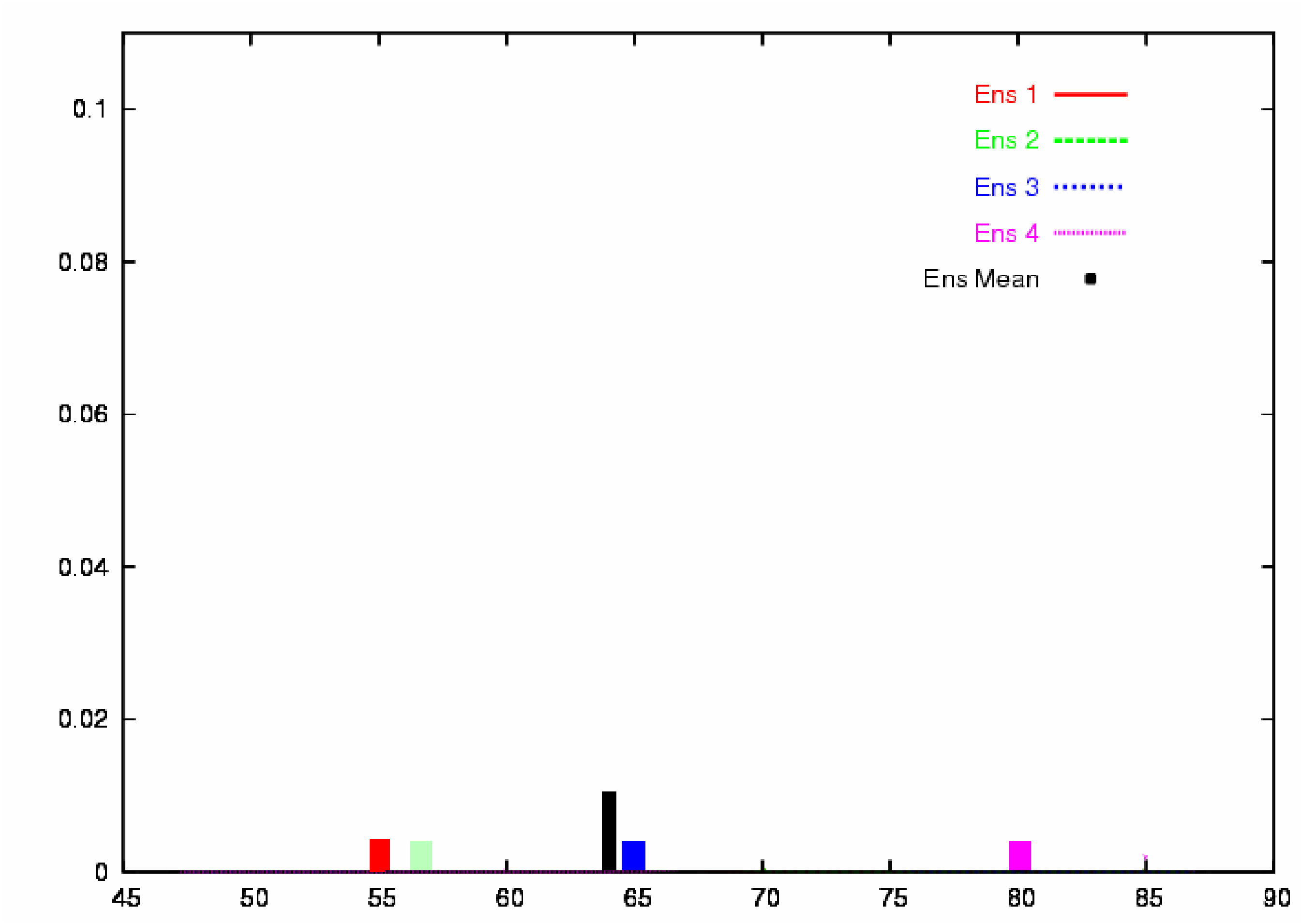
- Produce a Probability Distribution Function (PDF) from the ensembles.

Challenge

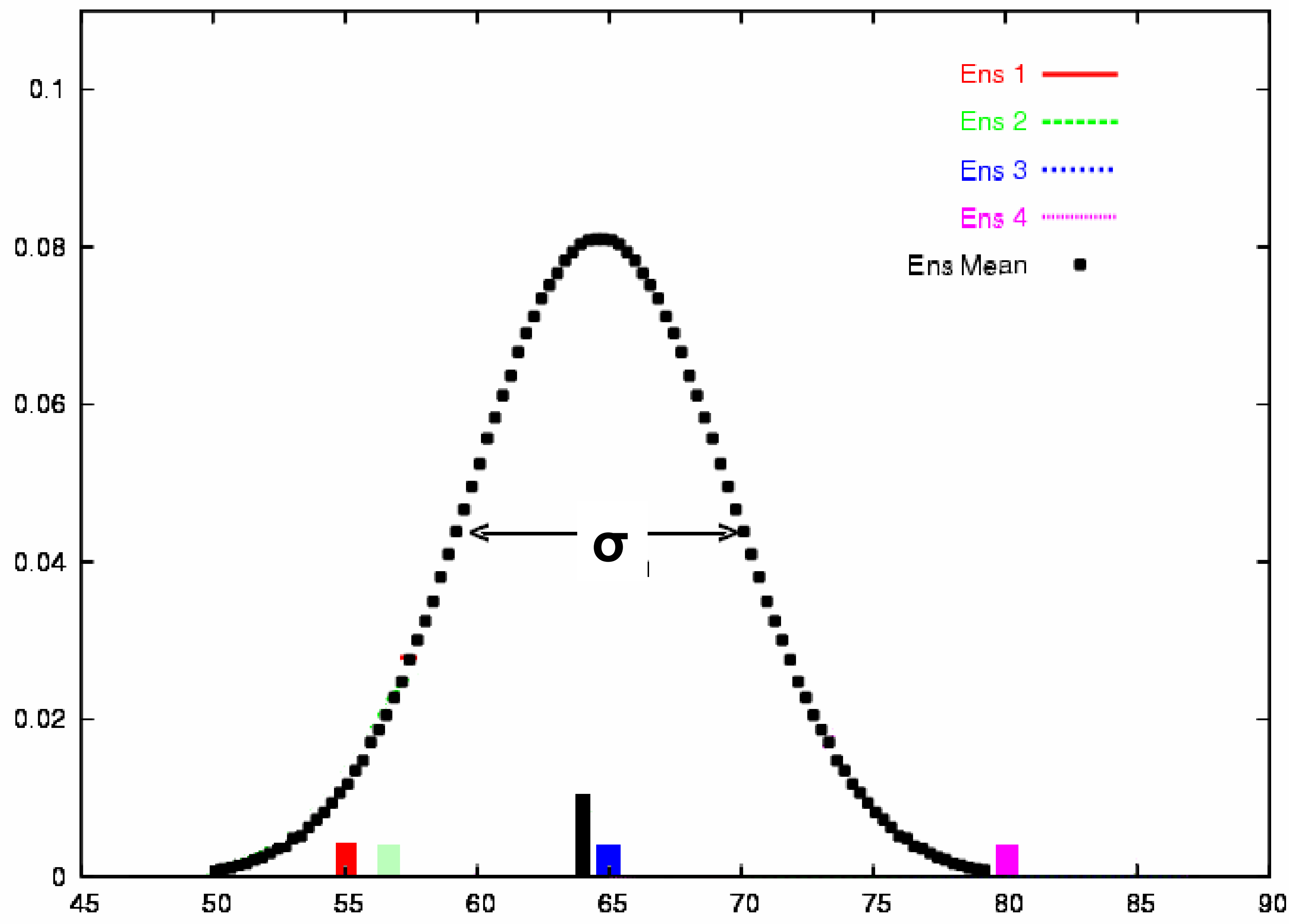
- Calibration
- Account for skill
- Retain information from ensembles

  (Or not if no skill)

Schematic illustration

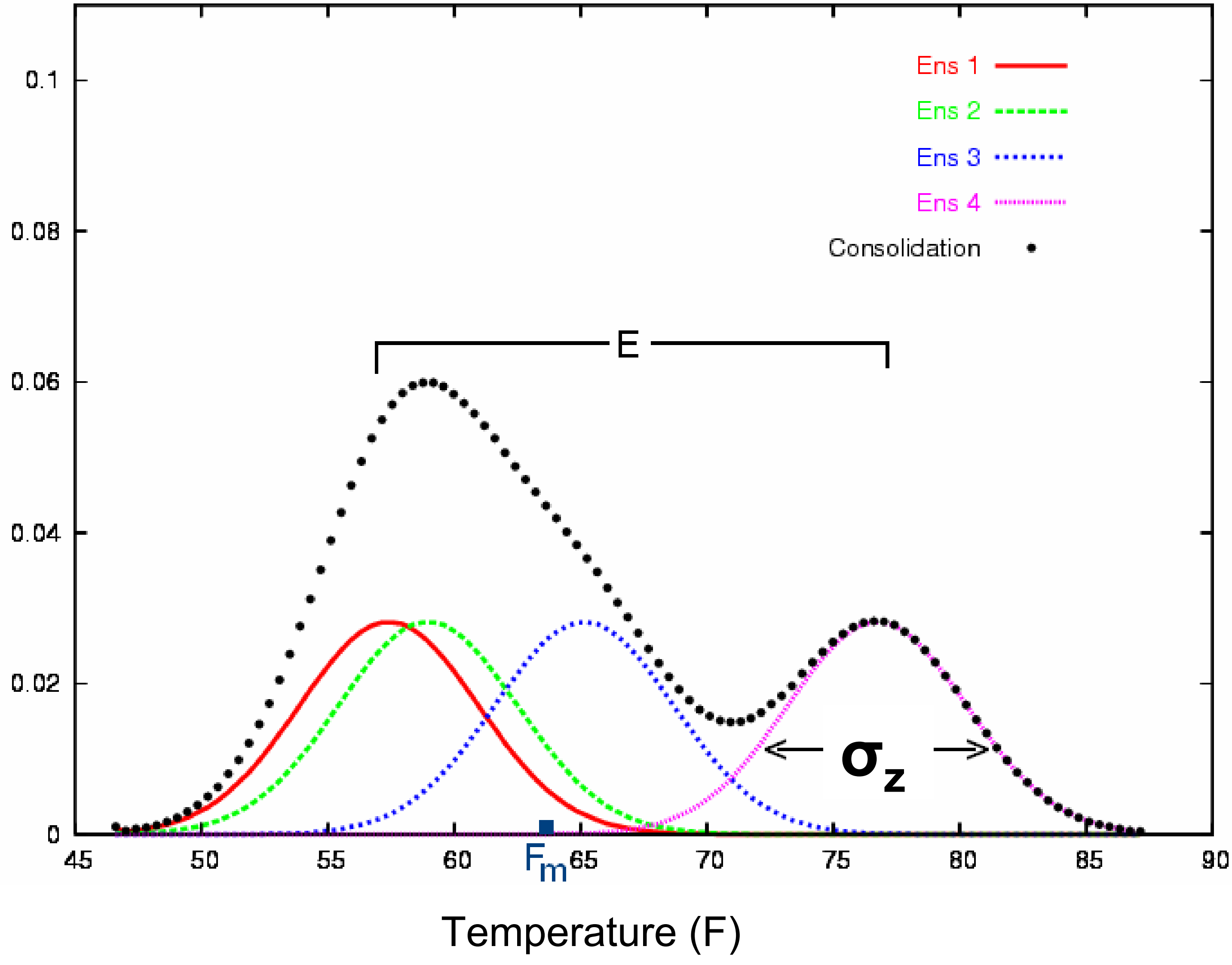


Temperature

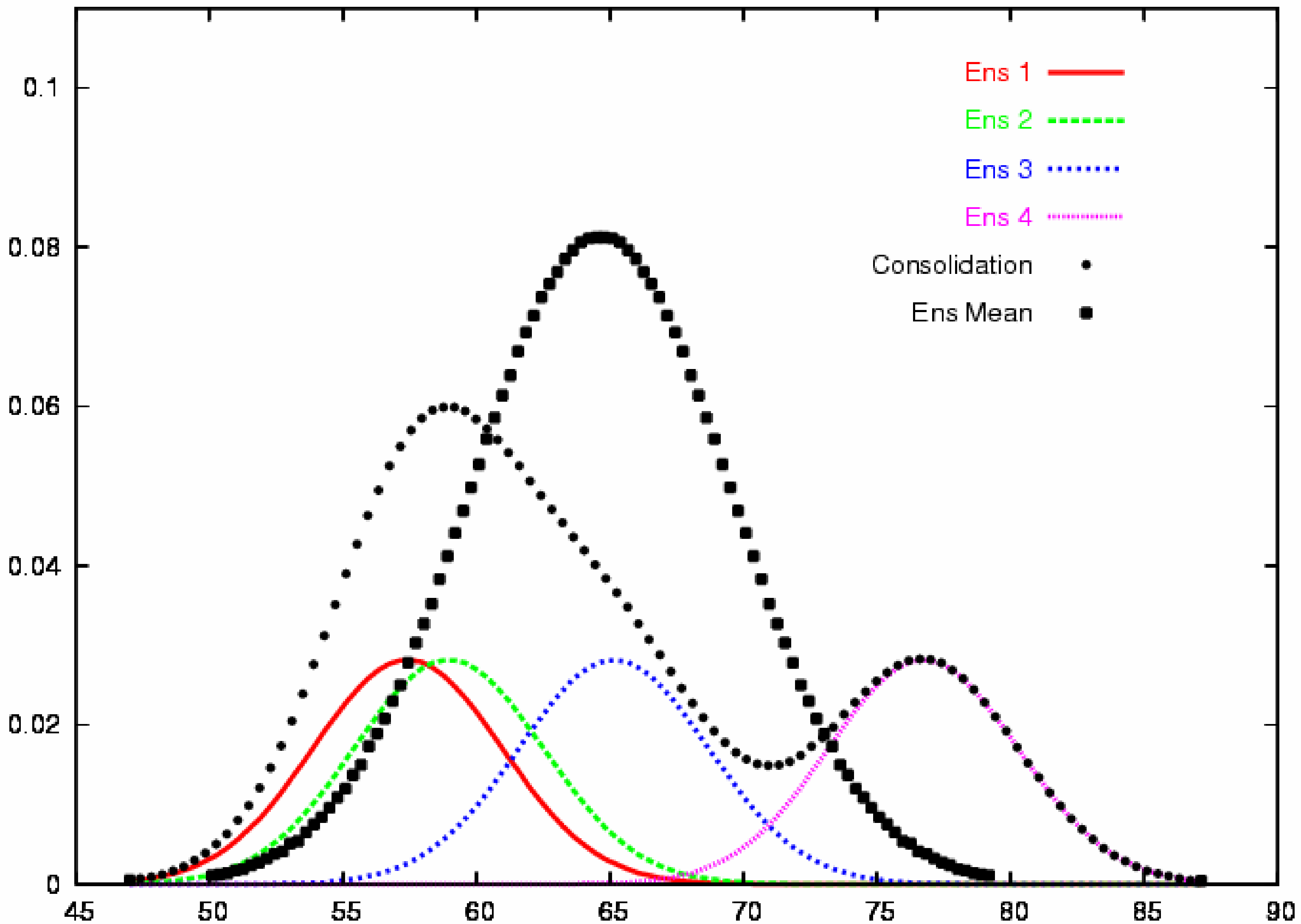


Ens 1
Ens 2
Ens 3
Ens 4
Ens Mean
σ
0.1
0.08
0.06
0.04
0.02
0
45
50
55
60
65
70
75
80
85
90

Schematic example



Kernel vs. Mean



Regression

Step 1. Standardization

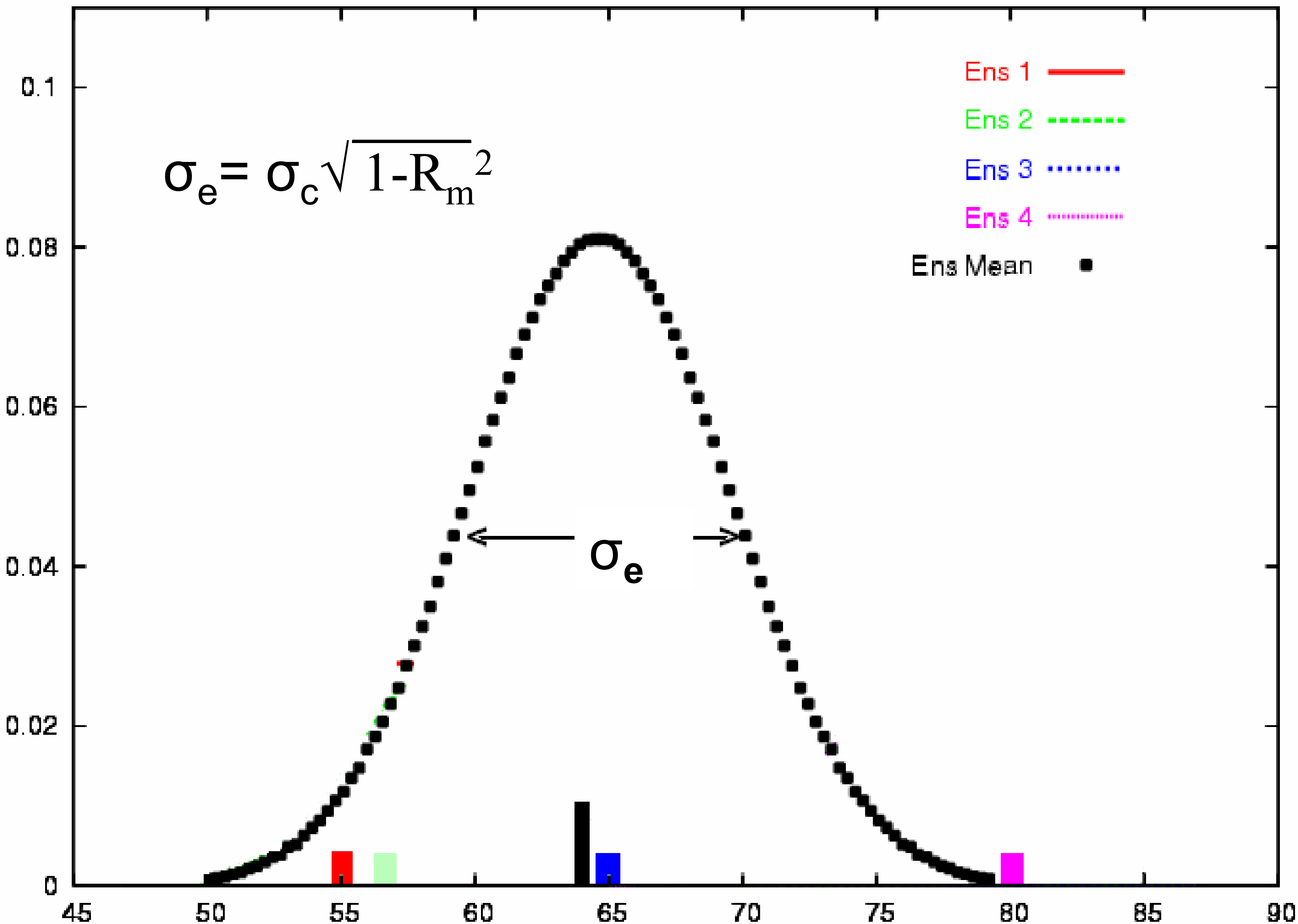
$$Z = \frac{(F - \overline{F})}{\sigma_F}$$

Step 2. Skill Adjustment

$$\hat{Z} = RZ$$

Step 3. Make the forecast

$$\hat{F} = \sigma_C \hat{Z} + C$$



σe= σc√1-Rm2
Ens 1
Ens 2
Ens 3
Ens 4
Ens Mean
σe
0.1
0.08
0.06
0.04
0.02
0
45
50
55
60
65
70
75
80
85
90

Analysis of Ensemble Variance

V = Variance

Total Variance	**= Explained Variance**	**+ Unexplained Variance**

$$\sigma_c^2 = V_e + V_u$$

$$\sigma_c^2 = \sigma_{\hat{F}m}^2 + \sigma_e^2$$

$$\sigma_c^2 = \sigma_{\hat{F}m}^2 + (E^2 + \sigma_z^2)$$

Analysis of Ensemble Variance (Continued)

With help of some relationships commonly used in linear regression:

$$\langle \hat{E}^2 \rangle = (R_m^2 - R_i^2)\, \sigma_c^2$$

$$\sigma_z^2 = (1 - 2R_m^2 + R_i^2)\, \sigma_c^2$$

$$R_z^2 = 2R_m^2 - R_i^2 \qquad\qquad R_z \leq 1.$$

Ensemble Calibration

Step 1. Standardization

$$Z_i = \frac{(F_i - \overline{F})}{\sigma_F}$$

Step 2. Ensemble Spread Adjustment

$$Z_i' = K(Z_i - Z_m) + Z_m$$

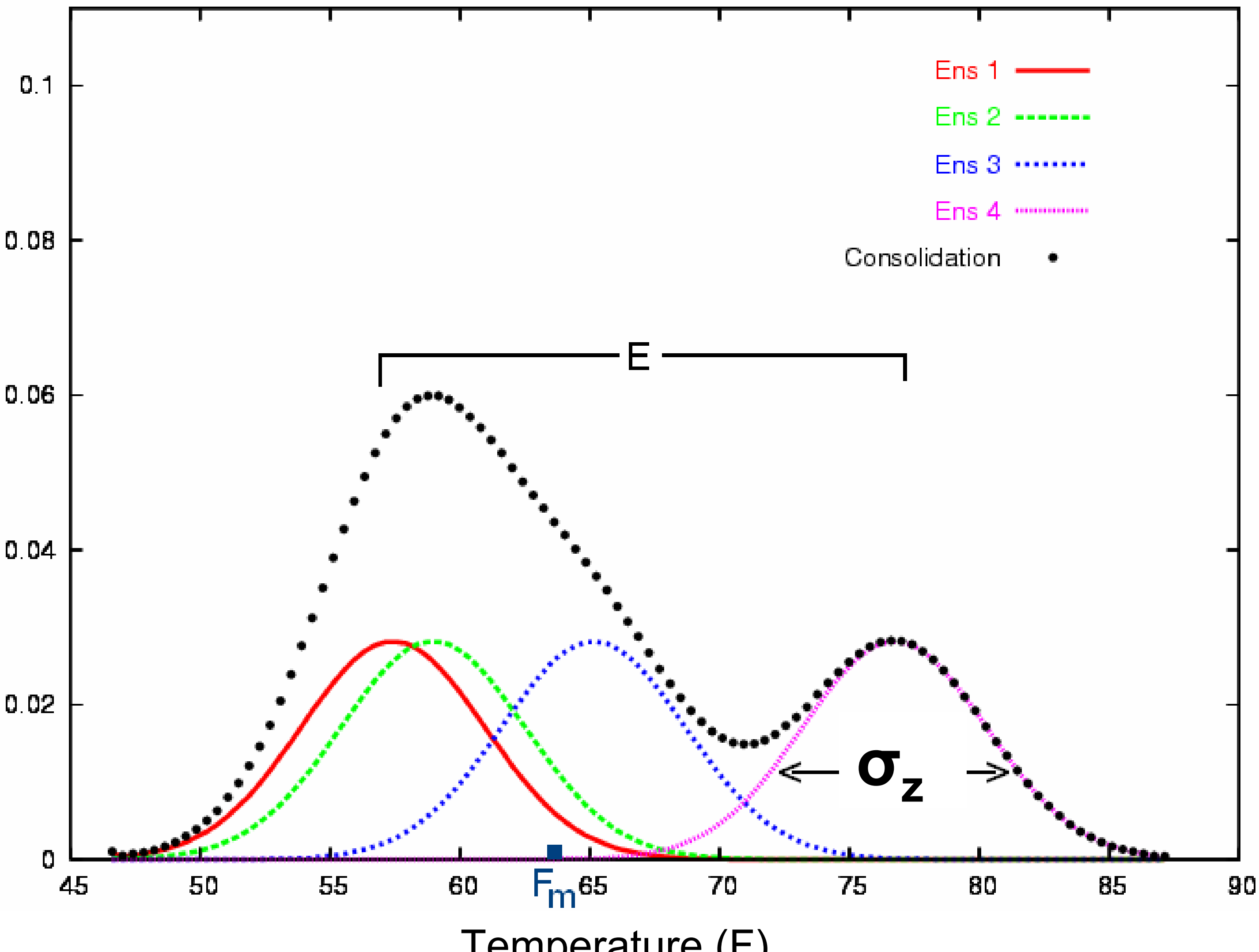
Step 3. Skill Adjustment

$$\hat{Z}_i = R_z Z_i'$$

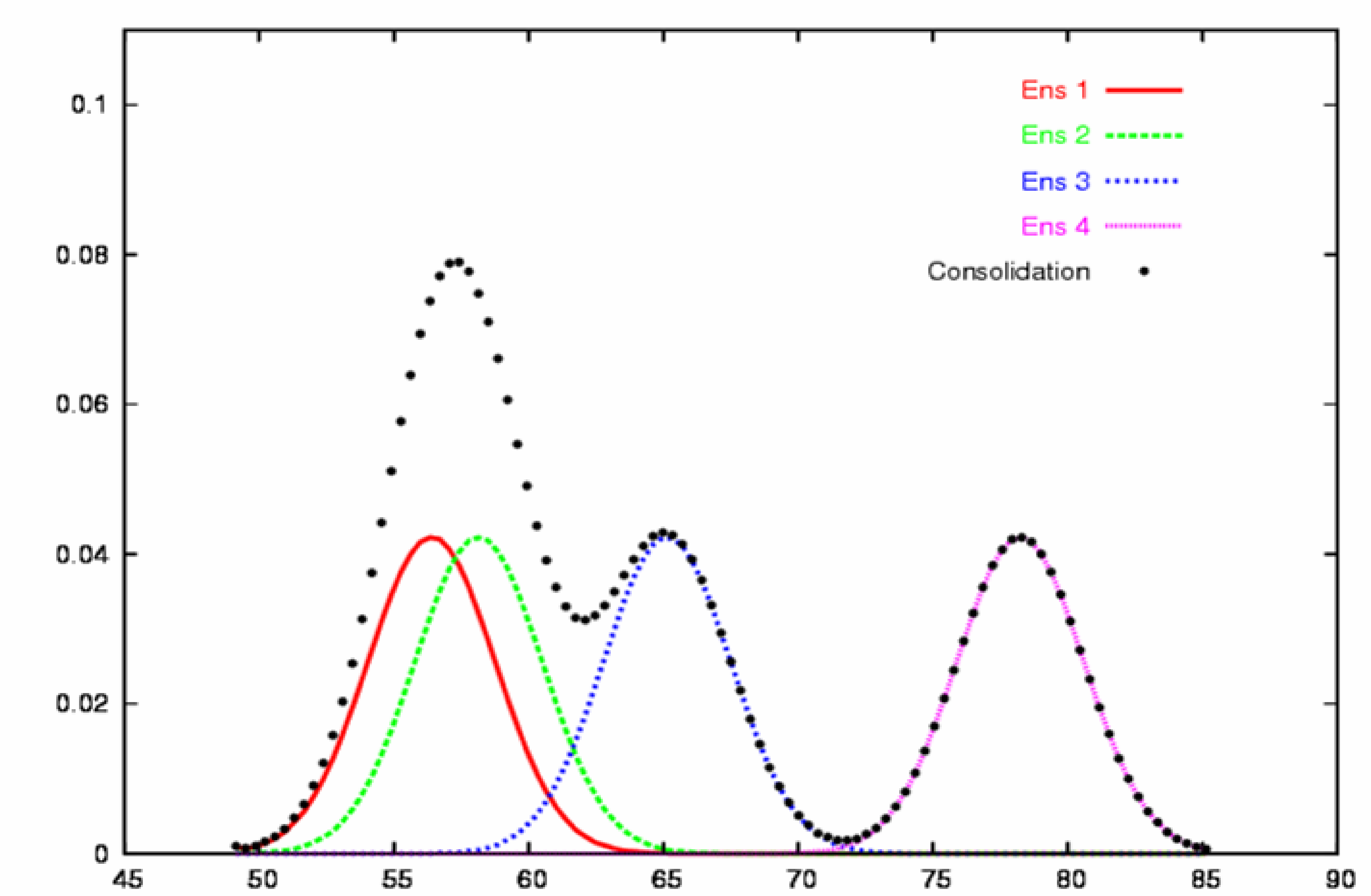
Step 4. Make the Forecast

$$\hat{F}_i = \sigma_C \hat{Z}_i + C$$

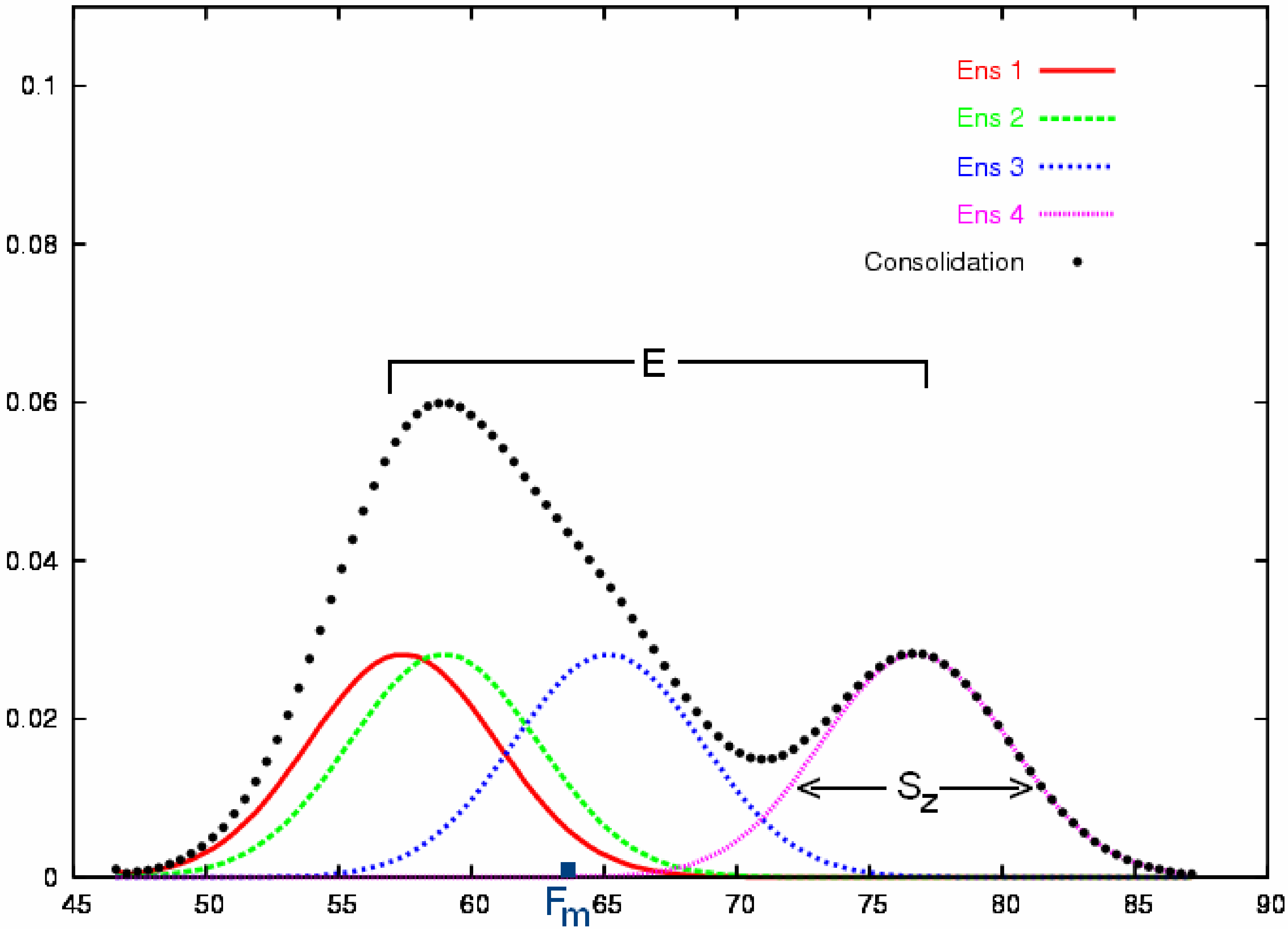
Schematic example



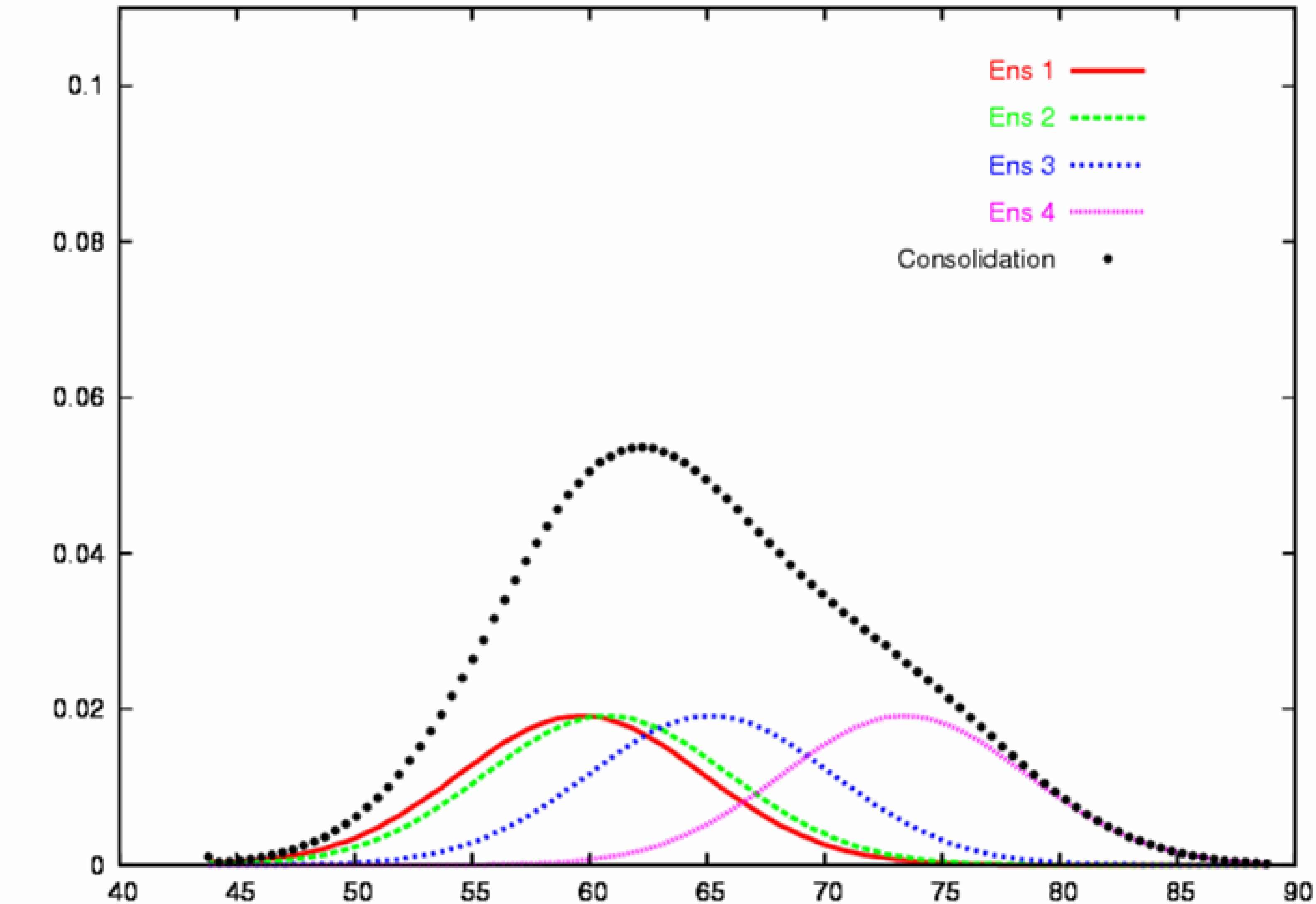
R_z=.97, R_{fm}=.94, R_i=.90



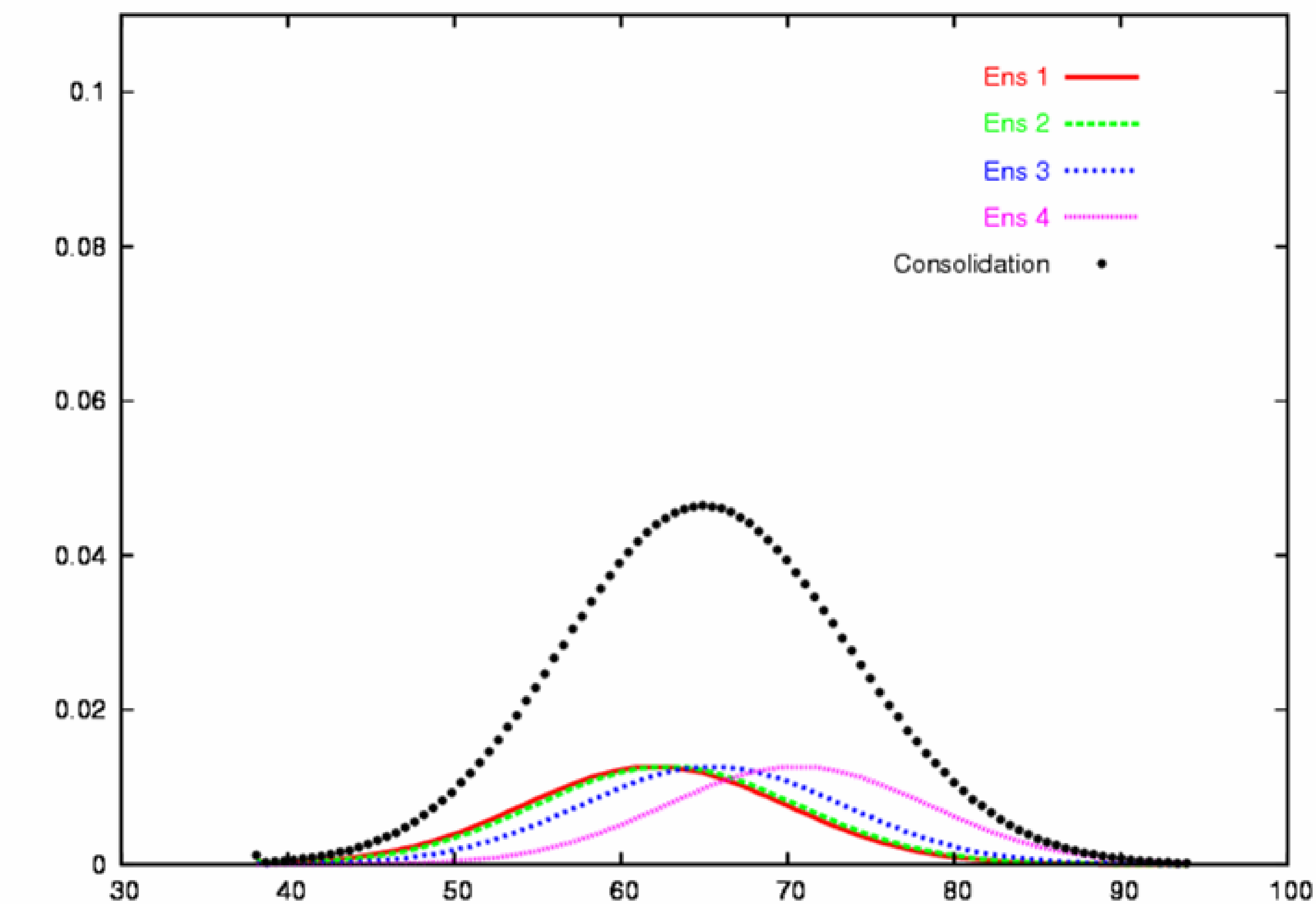
R_z=.93, R_{fm}=.87, R_f=.30



R_z=.85, R_{fm}=.67, R_i=.41



R_z=.62, R_{fm}=.46, R_f=.20



Weighting

$$w_i = \frac{R_i}{(1 - R_i)}$$

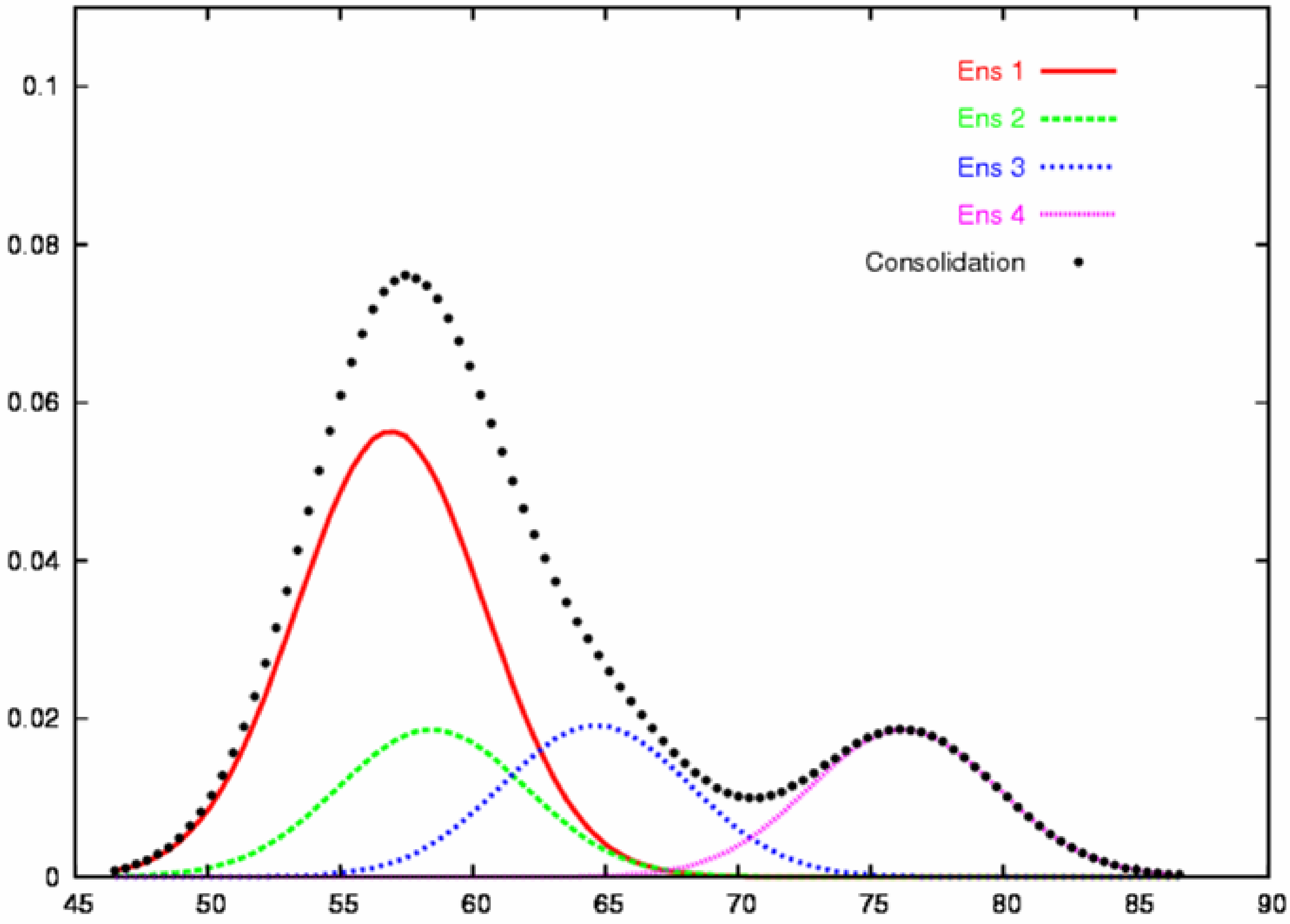
$$wt_i = \frac{w_i}{(\sum(w_i))}$$

$$R = .9: \quad 9 = \frac{.9}{(1 - .9)} \quad , \quad R = .8: \quad 5 = \frac{.8}{(1 - .8)}$$

$$9 + 5 = 14$$

$$.64 = \frac{9}{14} \qquad .36 = \frac{5}{14}$$

Wgts: 50% Ens. 1, 17%Ens 2, 3, 4



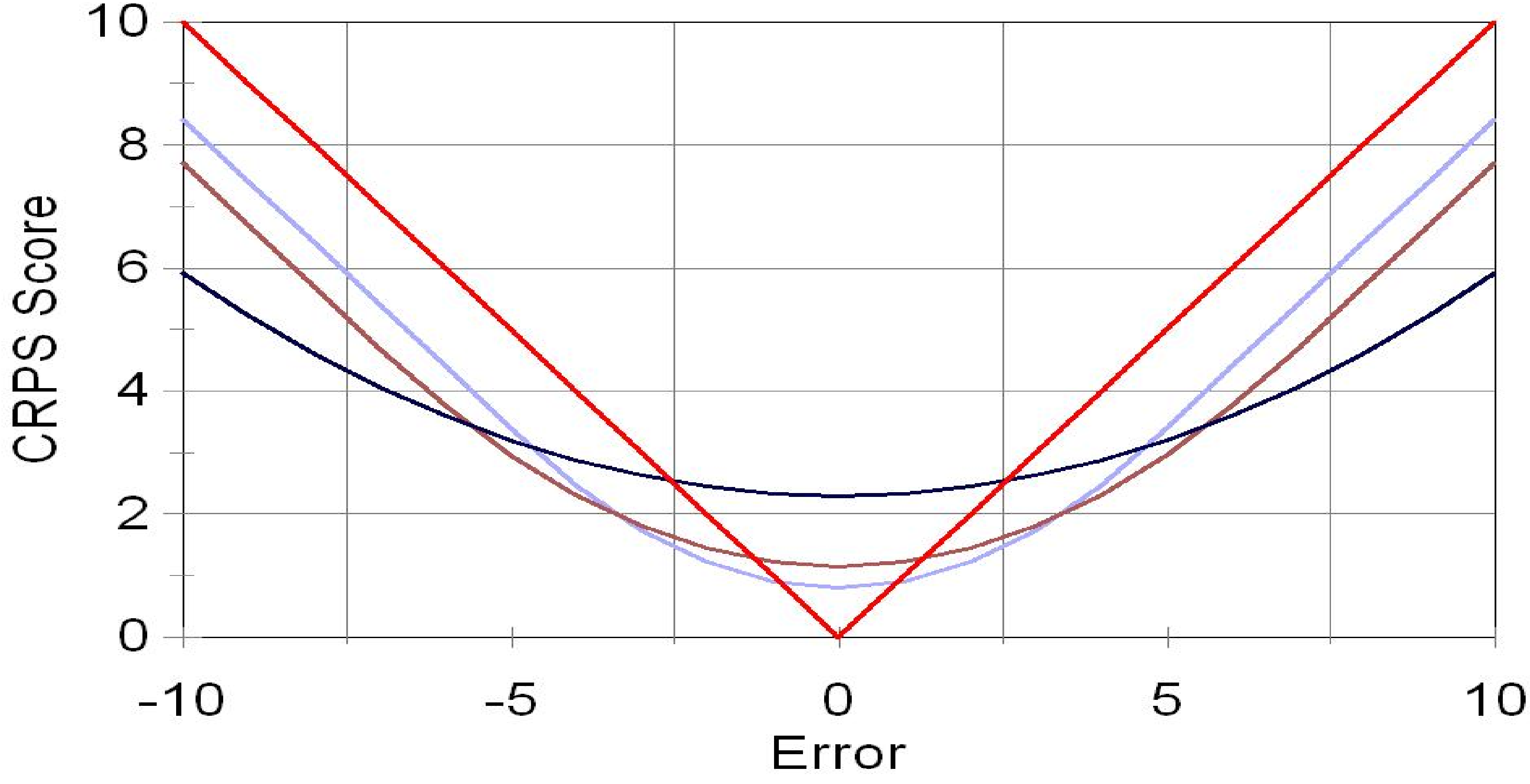
Real time system

- Time series estimates of Statistics.
  - Exponential filter

$$F_{T+1} = (1-\alpha)F_T + \alpha f_{T+1}$$

  - Initial guess provided from 1956-1981 CA statistics

Continuous Ranked Probability Score



Some Results

- ***Nino 3.4 SSTs***

Operational system

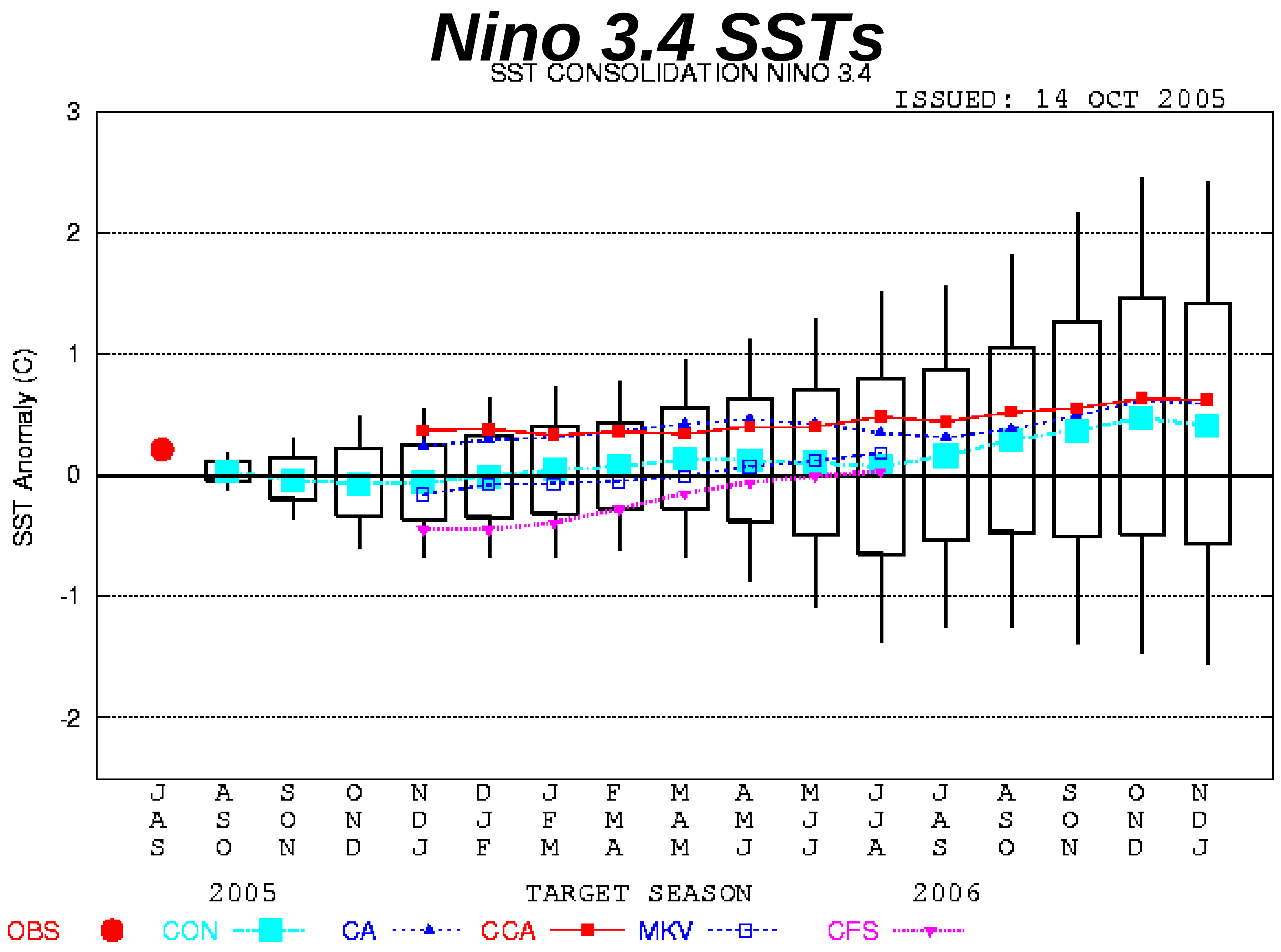
15 ***CFS*** ,12 ***CA,*** 1 ***CCA*** , 1 ***MKV***

- **Demeter Data**

9 ***CFS***, 12 ***CA***, 1 ***CCA***, 1 ***MKV***

9 ***UKM***, 9 ***MFR***, 9 ***MPI***,

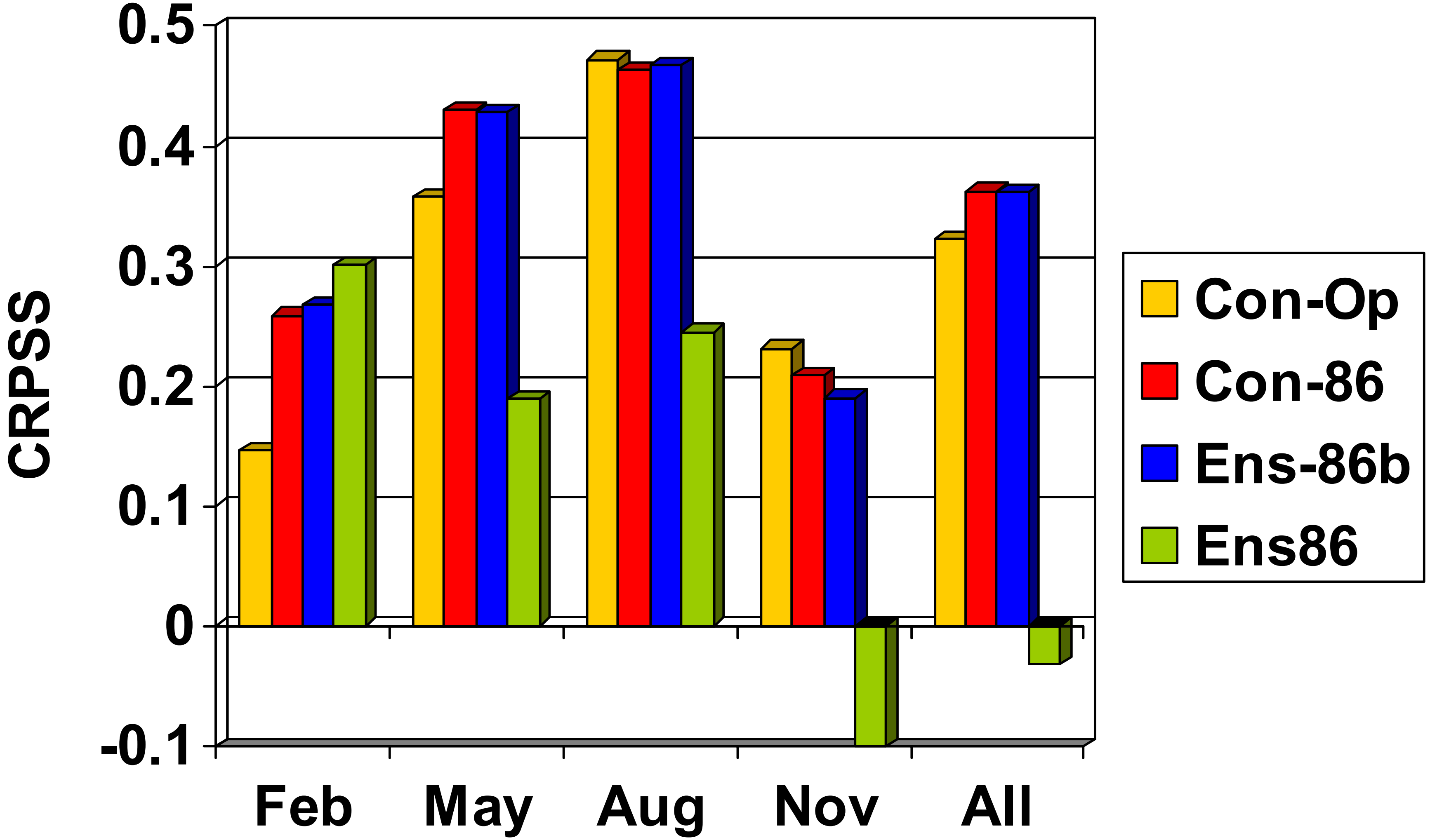
9 ***ECM***, 9 ***ING***, 9 ***LOD***, 9 ***CER***,



Nino 3.4 SSTs
SST CONSOLIDATION NINO 3.4
ISSUED: 14 OCT 2005
SST Anomaly (C)
3
2
1
0
-1
-2
JAS
ASO
SON
OND
NDJ
DJF
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
2005
TARGET SEASON
2006
OBS
CON
CA
CCA
MKV
CFS

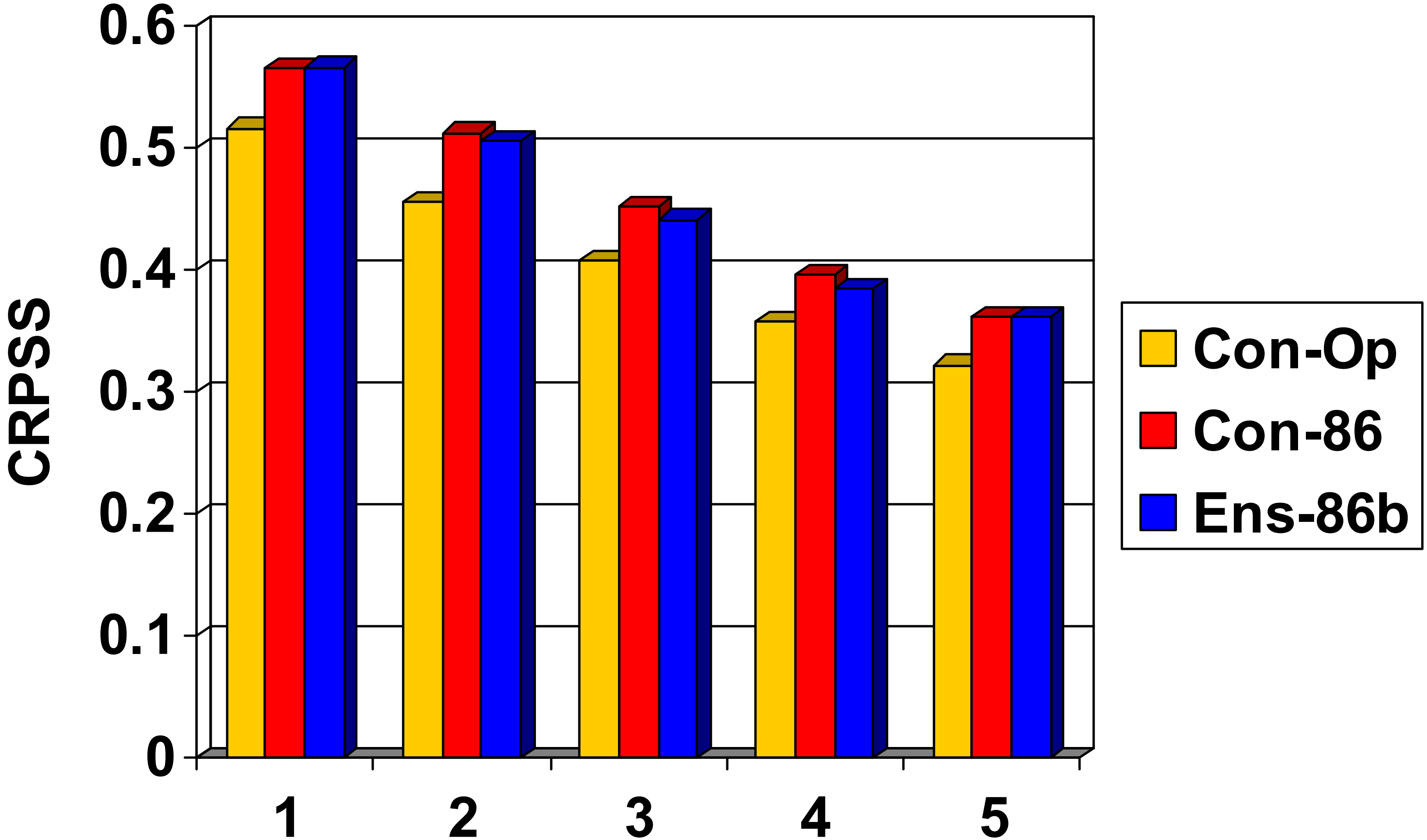
Nino 3.4 SST

5-month lead by initial time 1982-2001



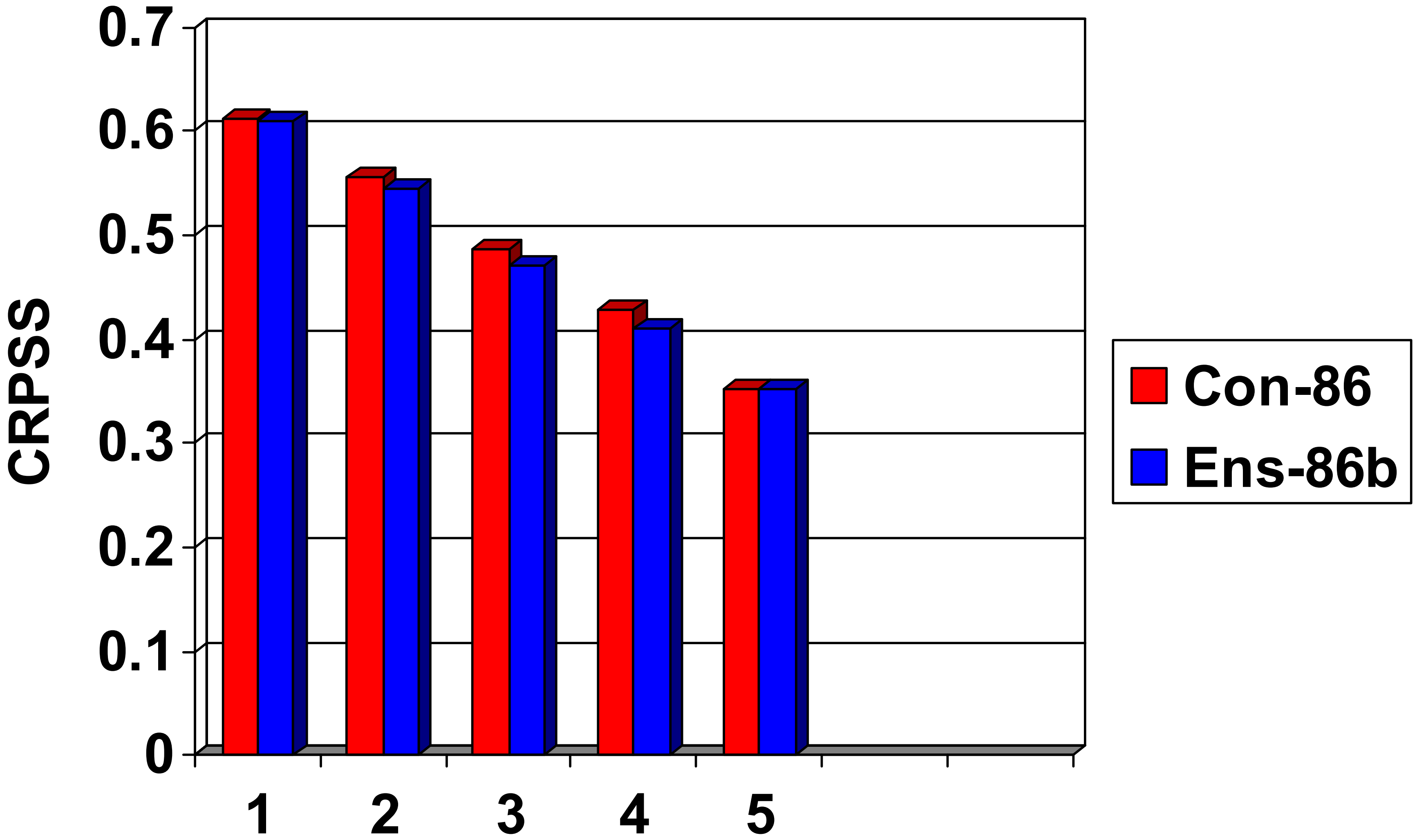
CRPSS – Nino 3.4 SSTs

All Initial times   1982-2001



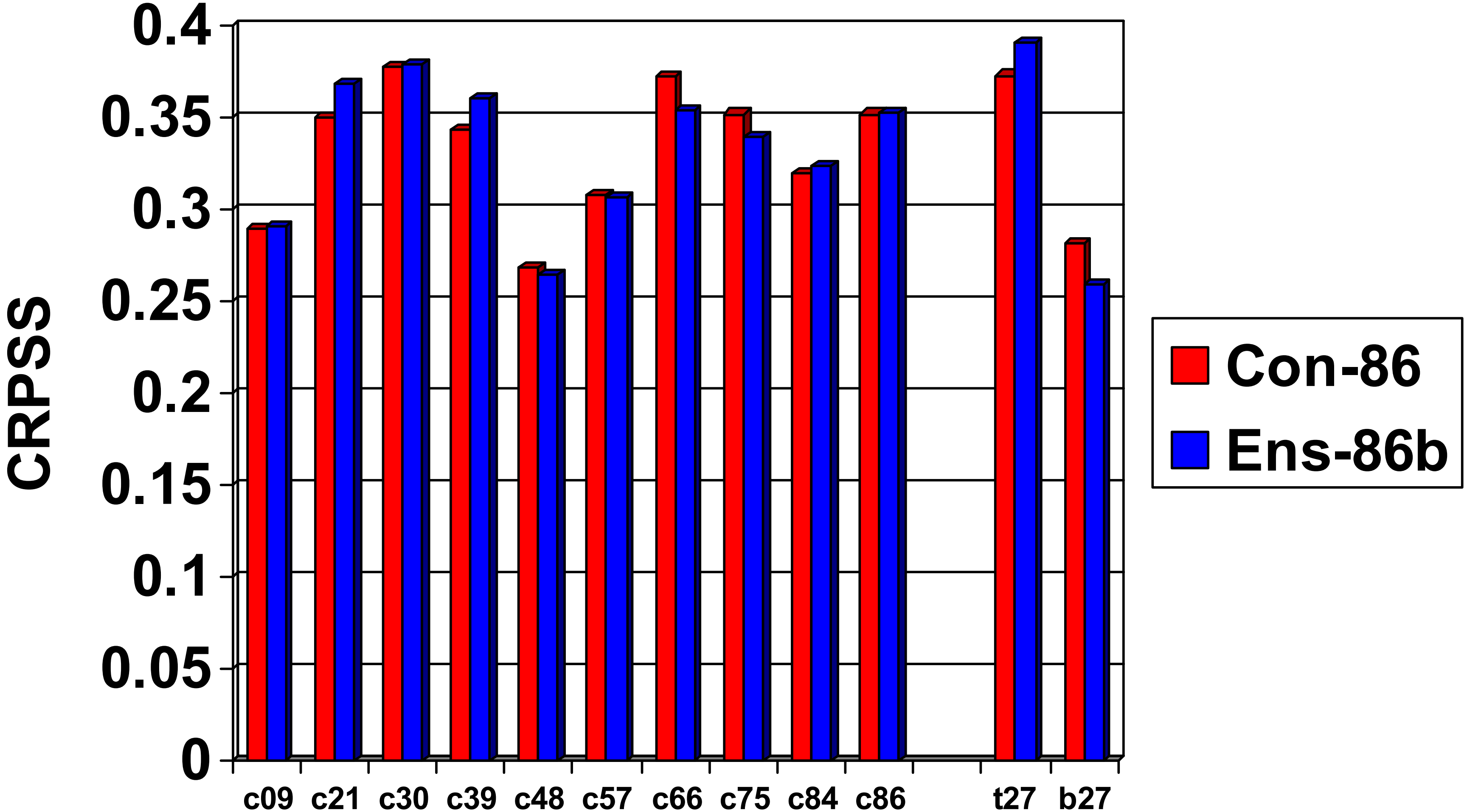
CRPSS – Nino 3.4 SSTs

All Initial times   1990-2001 (Independent)

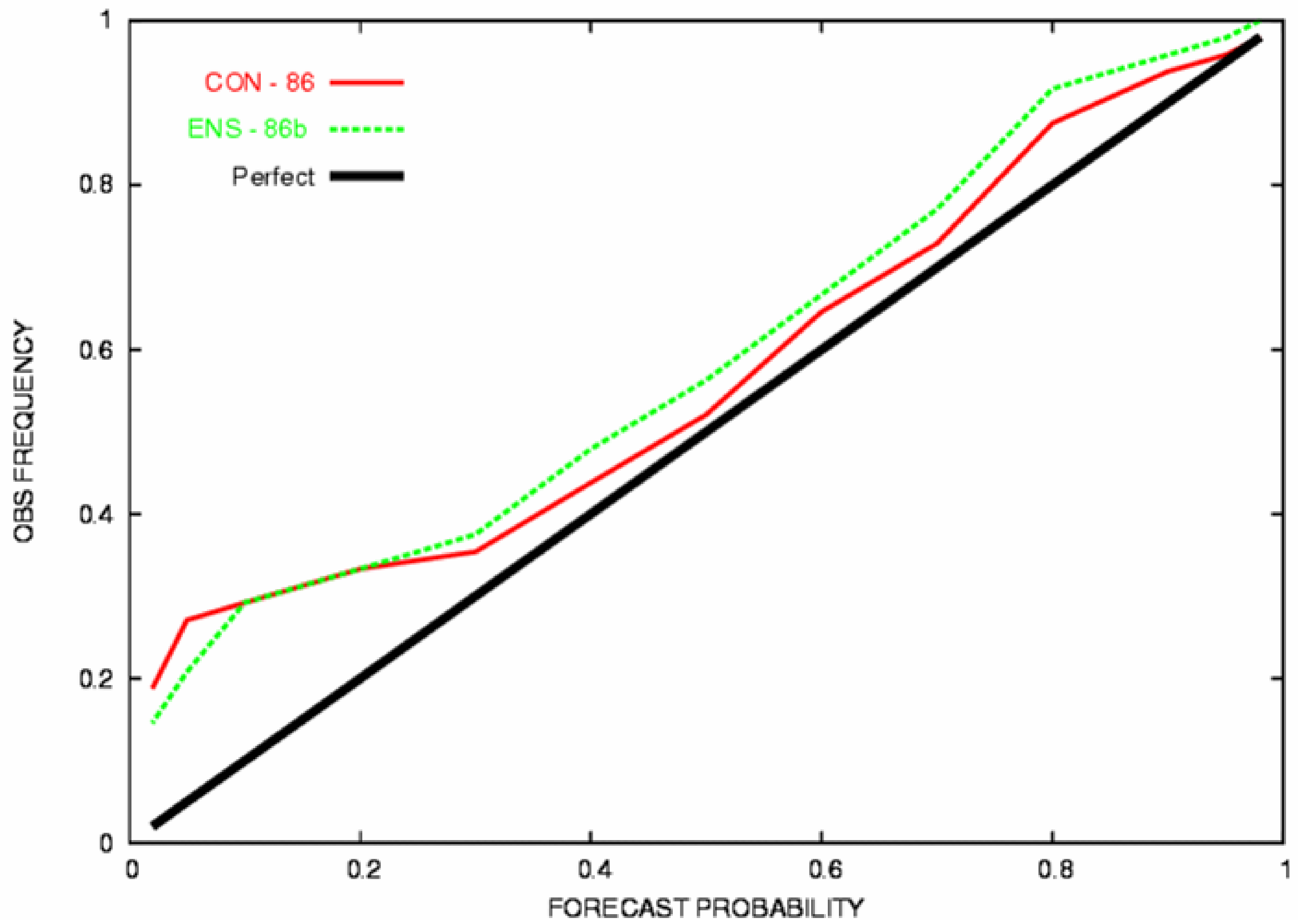


CRPSS – Nino 3.4 SSTs

5-month Lead, All Initial times 1990-2001
(Independent data)



Reliability Nino 3.4 SST (1990-2001)



U.S. Temperature and Precipitation Consolidation

- 15 CFS
- 1 CCA
- 1 SMLR

Trends are removed from models

Statistics and distribution are computed

Trend added to end result.

Trend Problem

- **Should a skill mask be applied? How much?**

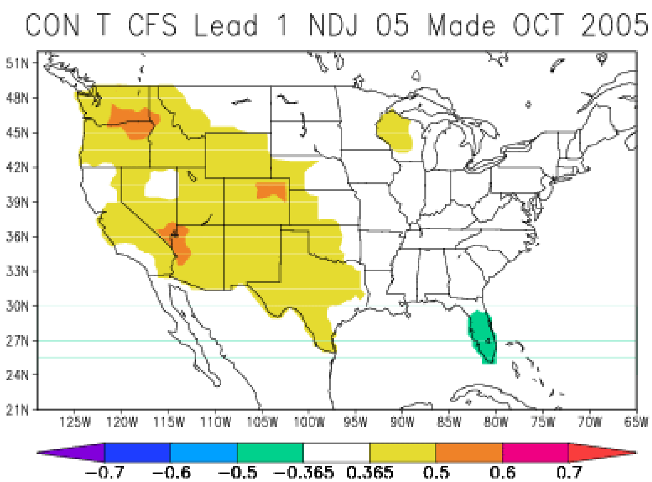
  - This technique requires a quantitative estimate of the trend.

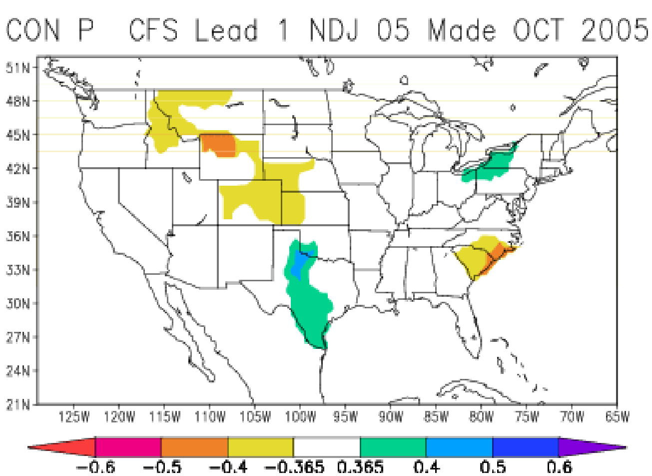
- **Component models sometimes “learn” trends, making bias correction difficult. – Doubles trends.**

- **Errors in estimating high frequency model forecasts**

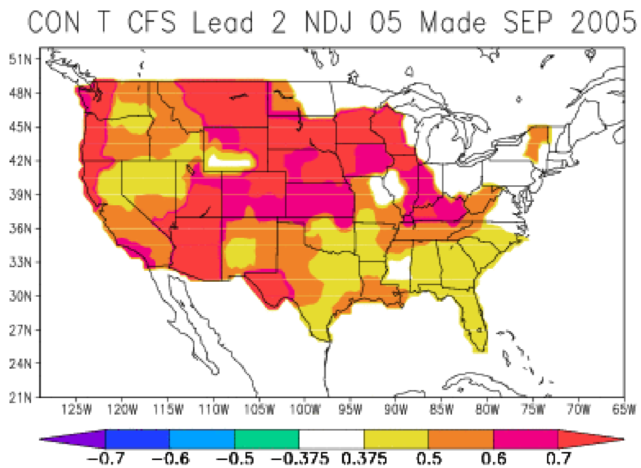
U.S. T and P consolidation

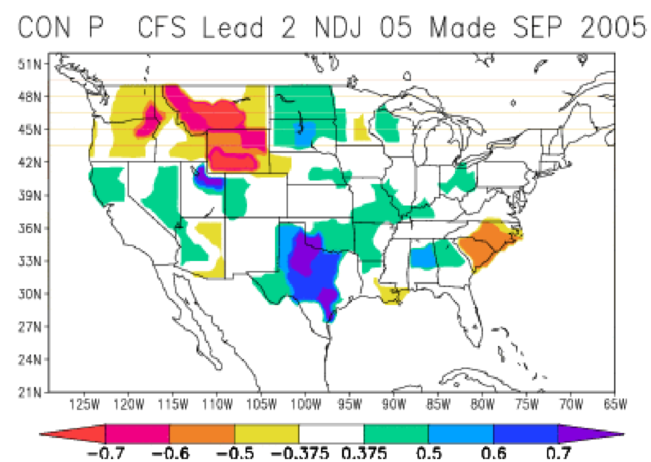
Skill Mask on Trends



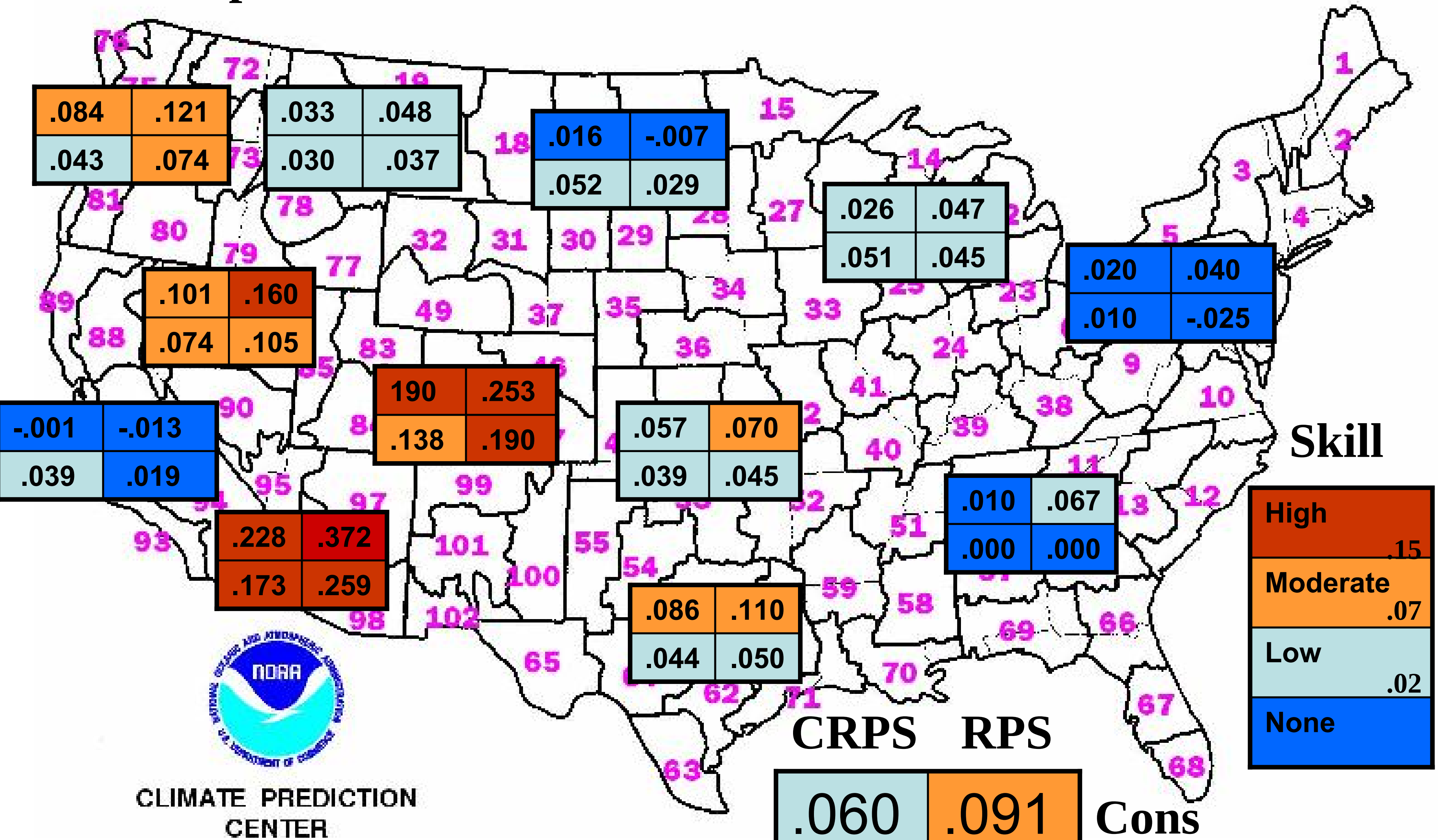


No Skill Mask on Trends





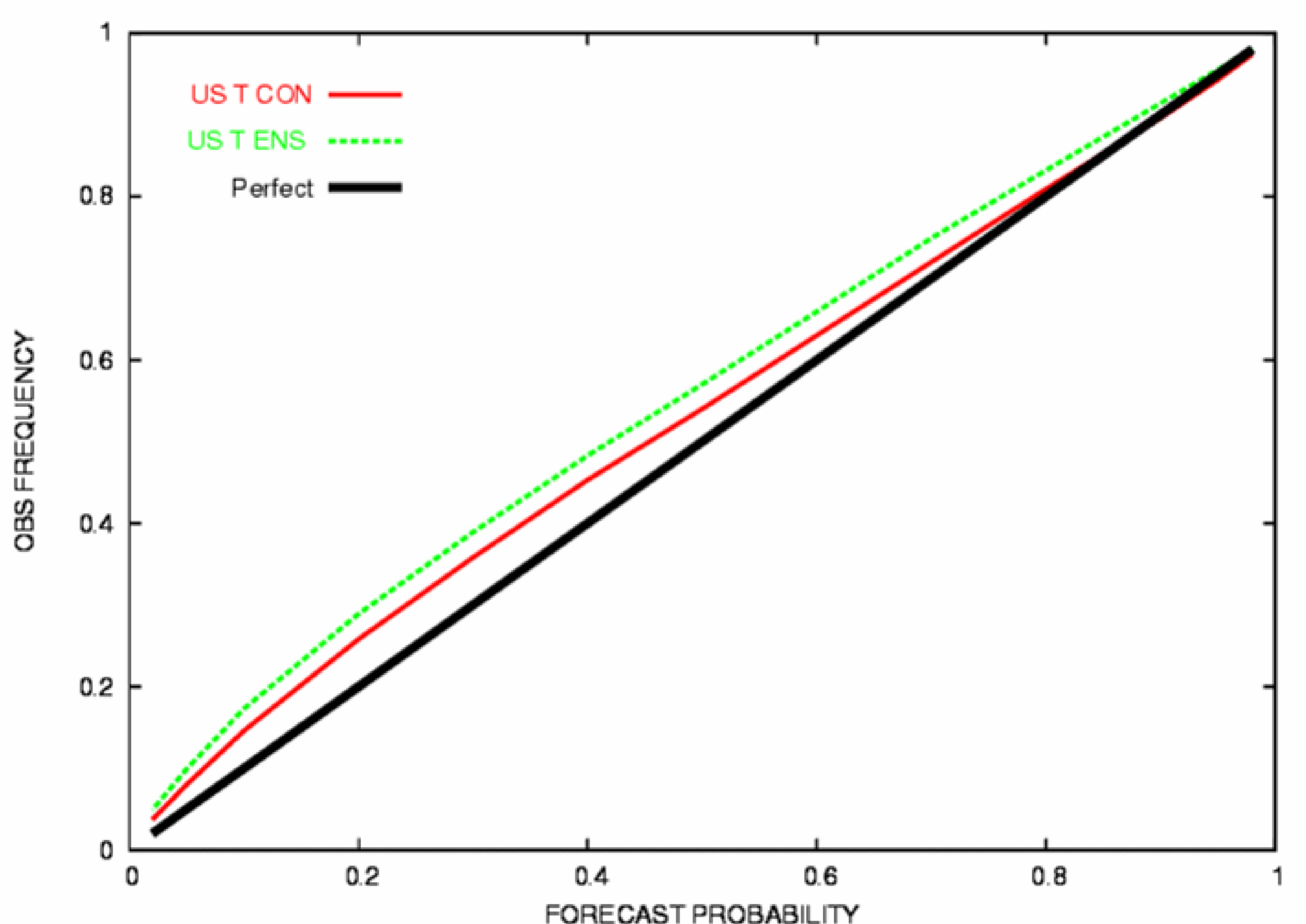
CRPS and RPS-3 (BNA) Skill Scores: Temperature



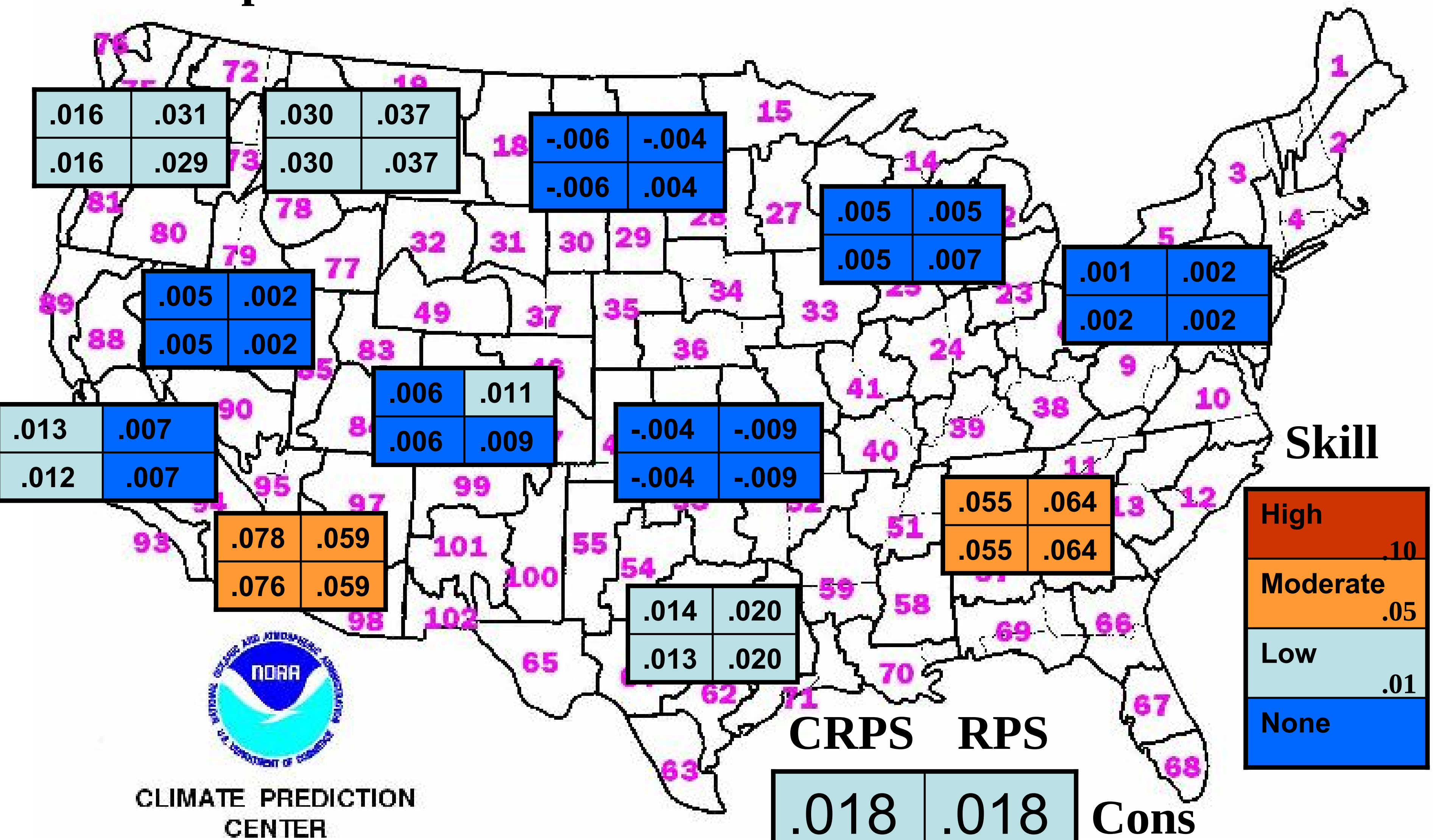
1-Month Lead, All initial times

	CRPS	RPS
Cons	.060	.091
Ensm	.053	.063

Reliability U. S. Temperatures
(1995-2003)



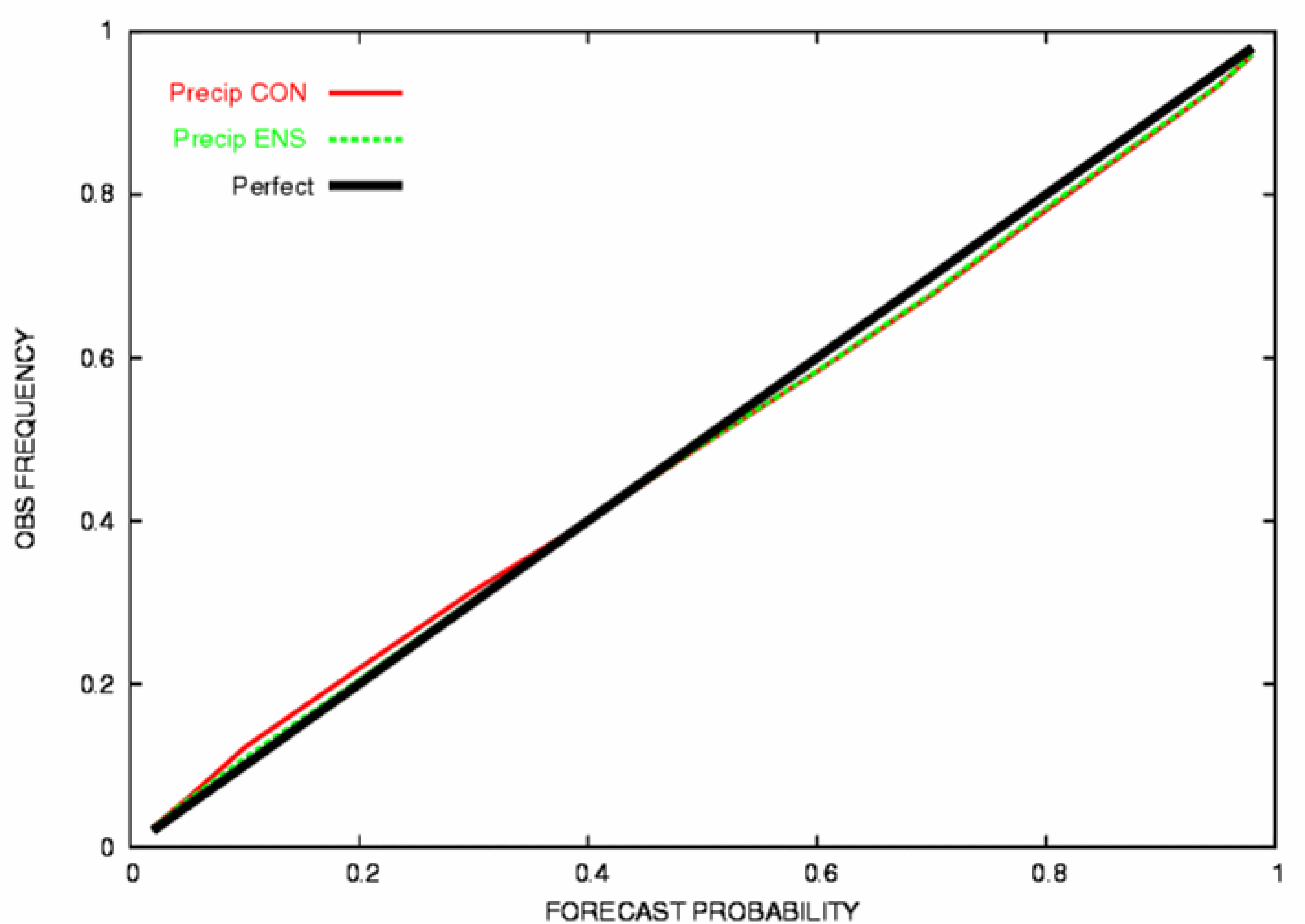
CRPS and RPS-3 (BNA) Skill Scores: Precipitation



	CRPS	RPS
Cons	.018	.018
Ensm	.017	.018

1-Month Lead, All initial times

Reliability U.S. Precipitation
(1995-2003)



Conclusions

- **Calibrated ensemble and ensemble means score very closely (by CRPS)**

  Calibrated ensembles seem to have a slight edge.

- **No penalty for including many ensembles (but not much benefit either)**
- **Considerable penalty for including less skillful ensembles – Weighting is critical.**
- **Probabilistic predictions are reliable (when looked at in terms of a continuous PDF)**

Conclusions (Continued)

- **Calibrated ensembles tend to be slightly overconfident**
- **Trends are a major problem – and are outside the realm of consolidation (but they are critically important for seasonal temperature forecasting).**

6-10 day Forecasts (based on Analogs)