A Consolidated Seasonal Temperature Forecast based on the NMME

David Unger

Climate Prediction Center

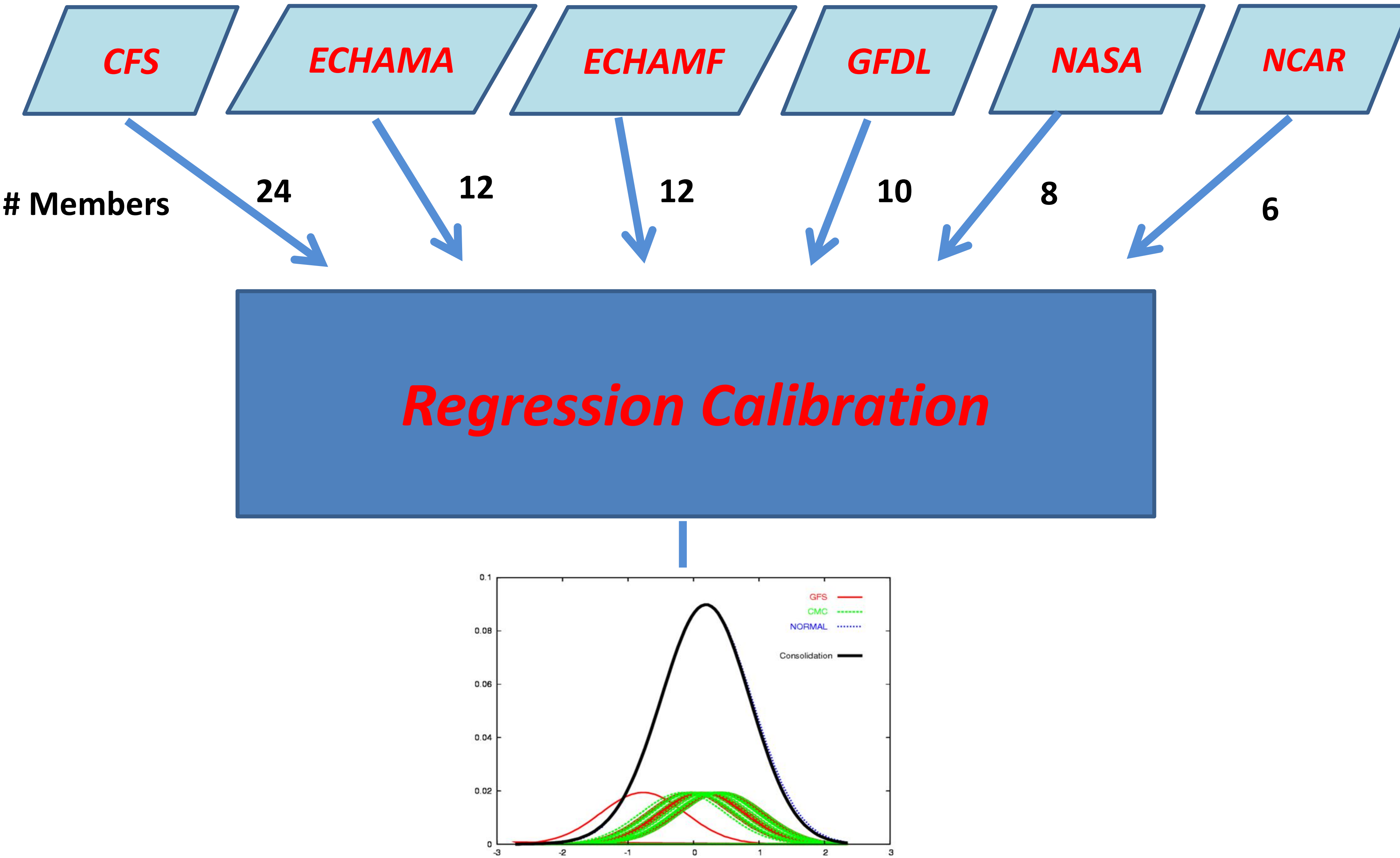
NOAA/NCEP/NWS

Acknowledgements: Qin Zhang

OBJECTIVES

- **An objective method for weighting and calibration of multi-model ensembles**
- **Theoretically sound regression based treatment.**
- **Able to handle models with differing skills**
- **Able to handle models with different ensemble sizes**
- **Provide forecast diagnostics**
- **Make use of individual ensemble members where possible.**

Overview


CFS
ECHAMA
ECHAMF
GFDL
NASA
NCAR
Members
24
12
12
10
8
6
Regression Calibration
GFS
CMC
NORMAL
Consolidation
0.1
0.08
0.06
0.04
0.02
0
-3
-2
-1
0
1
2
3

ENSEMBLE REGRESSION

- **Given:** **An ensemble forecast for a single event**

ASSUMPTIONS

1) We know the probability of each member being closest to the observation **(usually assumed equal for any given model)**

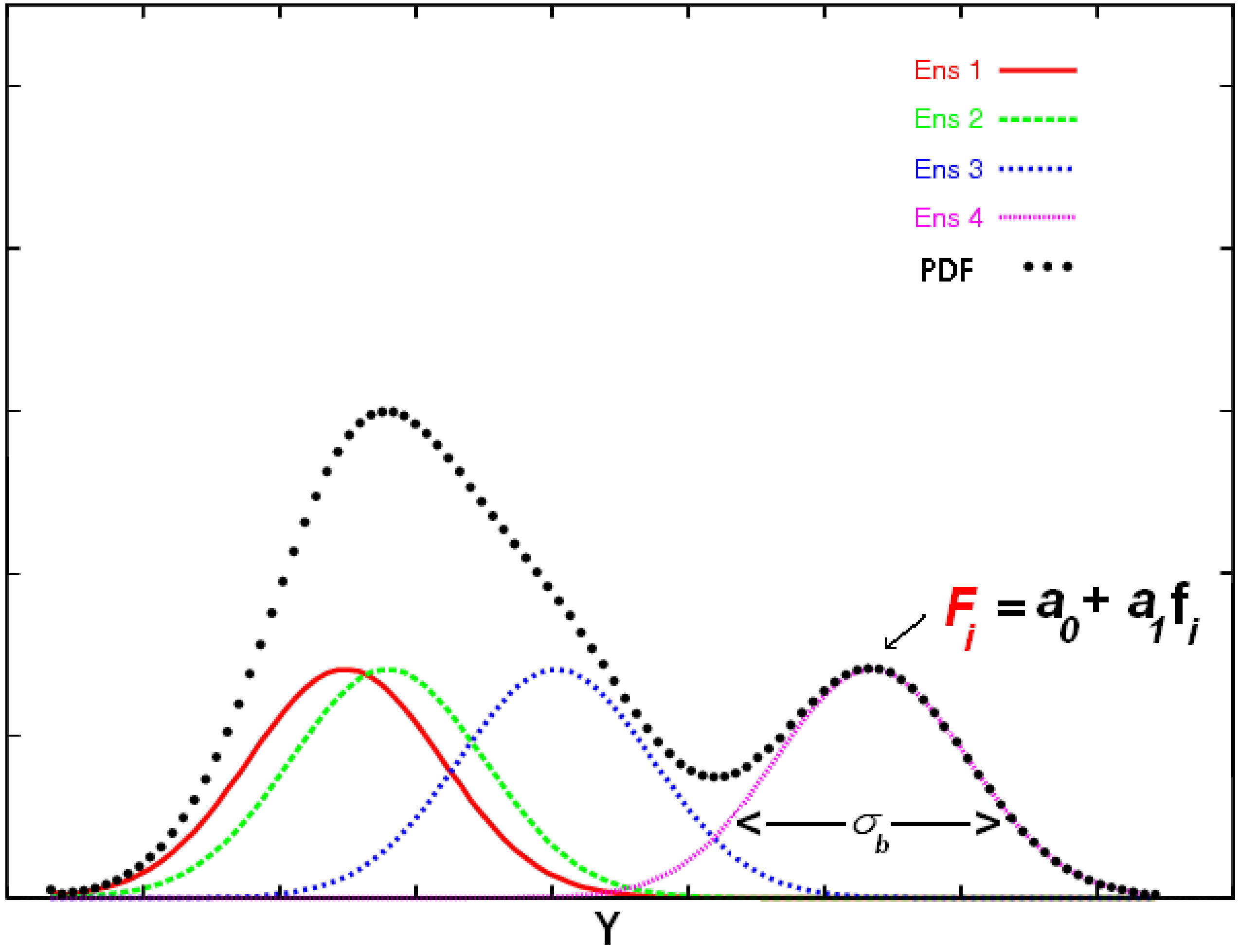
2) The residual errors about the best model are Gaussian

3) Linear relationship between the BEST forecast and obs.

THEN

We can derive a regression equation for the *EXPECTED VALUE* of least squares solution between the best member and the observation, together with an expected residual error distribution.

ENSEMBLE REGRESSION EXAMPLE



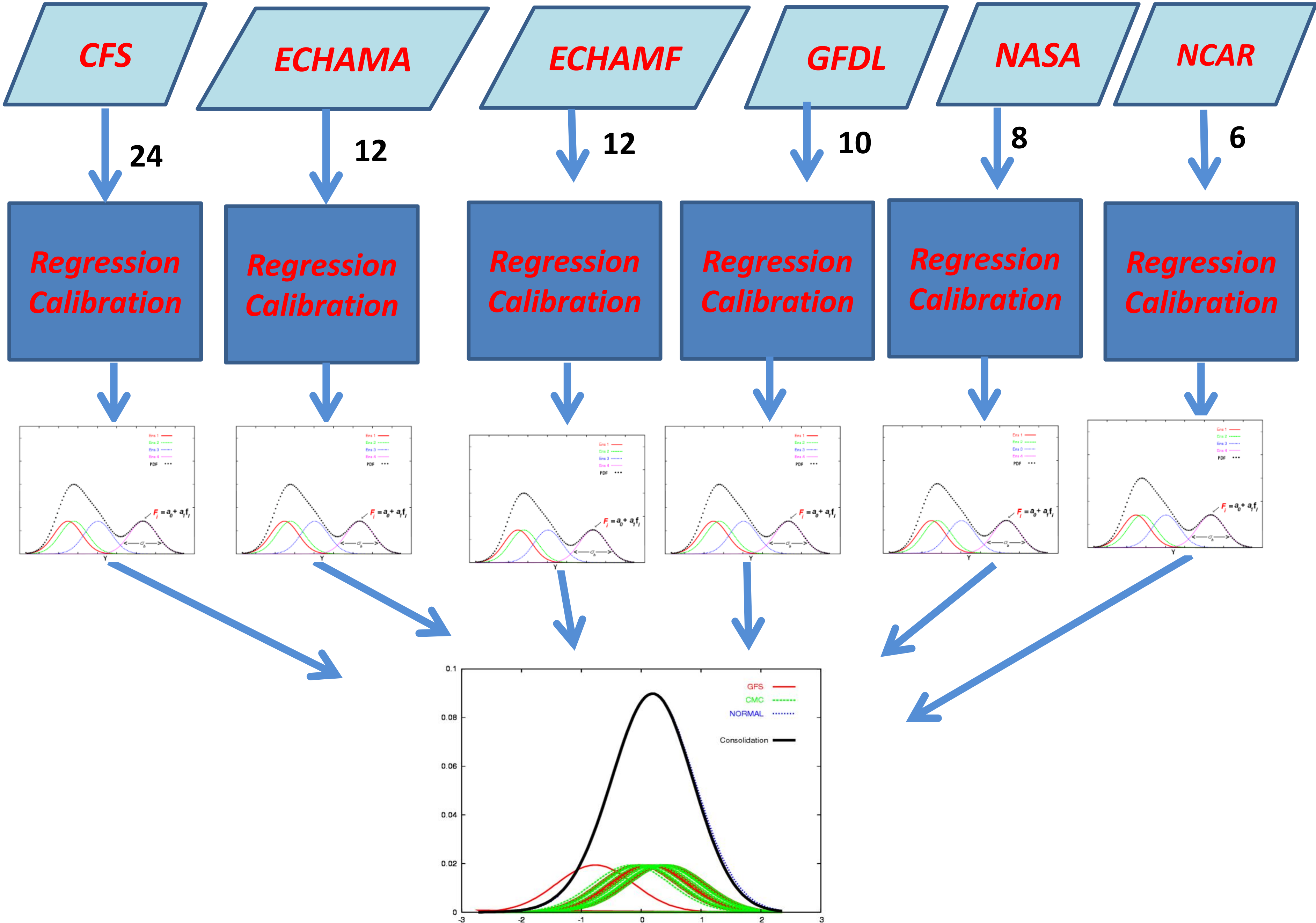
First step: Process individual models

- **Assume each member equally likely to be best**
- **Use regression theory to produce a probability density function (PDF) that represents the entire ensemble set.**

Step 2: EVALUATE THE ENSEMBLE FOR EACH CASE AND EACH MODEL

- **Use Regression to obtain a translate the forecast into a PDF**
- **The PDF reflects the skill of the forecast**
- **Equalizes for number of members.**

Overview



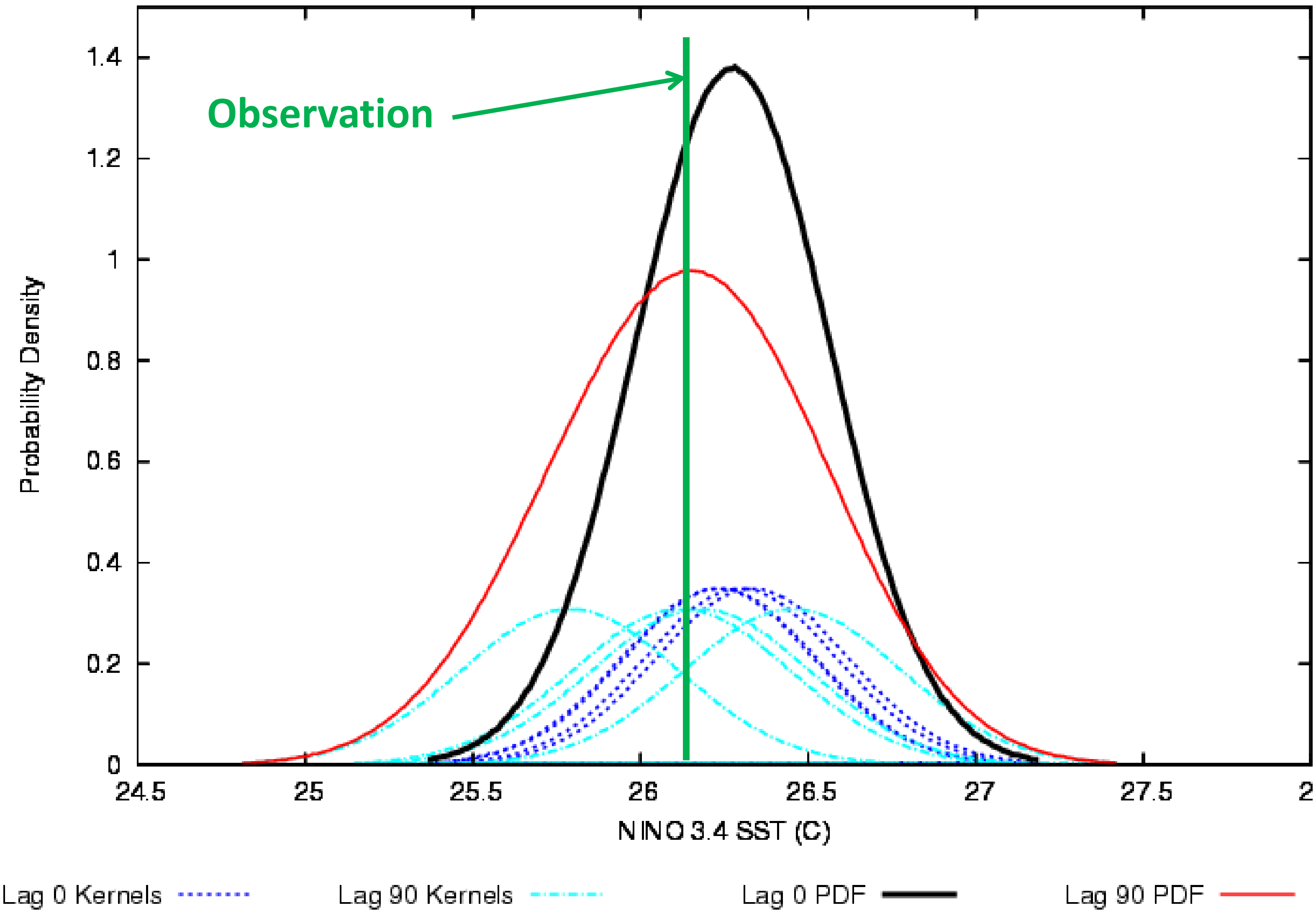
CFS
ECHAMA
ECHAMF
GFDL
NASA
NCAR
24
12
12
10
8
6
Regression Calibration
Regression Calibration
Regression Calibration
Regression Calibration
Regression Calibration
Regression Calibration
GFS
CMC
NORMAL
Consolidation

Step 3: EVALUATE THE PDF FOR EACH MODEL AT THE OBSERVATION

- **This provides a relative likelihood that the observation is represented by the model.**

  (A bit like Bayesian analysis)

BEST MEMBER EXAMPLE



Step 4: EVALUATE THE PDF STATS ON HISTORICAL DATA

- **Find the model with the highest likelihood of being the “Best” for this forecast case.**

(The model with the highest PDF at the observation).

- **This provides the ensemble regression with the information needed to combine different models.**

WEIGHTING

- **Recall the regression theory requires the probability that each member will be best.**

P(best) = The probability that each MODEL contributed the best member (among all models) From the historical performance on PDF's)

$$\frac{P(best)}{N} = W$$

Where W is the *individual* member weighting

FINAL CONSOLIDATION

$\text{Member}_i \quad W_i \quad i=1, (N_1+N_2+ \ldots +N_m)$

Enter Weighted Super Ensemble in a Second-Pass Ensemble Regression (All kernel widths the same)

Enter Weighted Kernels into a consolidated PDF (Kernel widths vary)

TESTS

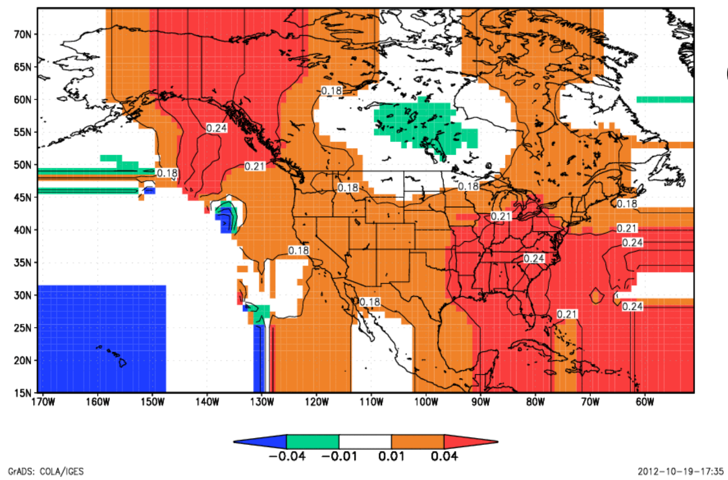
- **6 Models: CFS, ECHAMA, ECHAMF, GFDL, NASA, NCAR**
- **1982-2010 Hindcasts.**
- **October Initial Conditions**
- **DJF target period. (Lead 2)**
- **One degree Resolution – North America**
- **Observed 2-meter Temperatures from Reanalysis.**
- **Regressions initialized with a exact derivation of coefficients 1982-1994. Smoothed by a centered 11x11 degree square.**
- **Weights determined 1982-1994 statistics.**
- **Validation Statistics from 1982-2010 (1982-1994 dependent data).**

WEIGHT GRAPHS

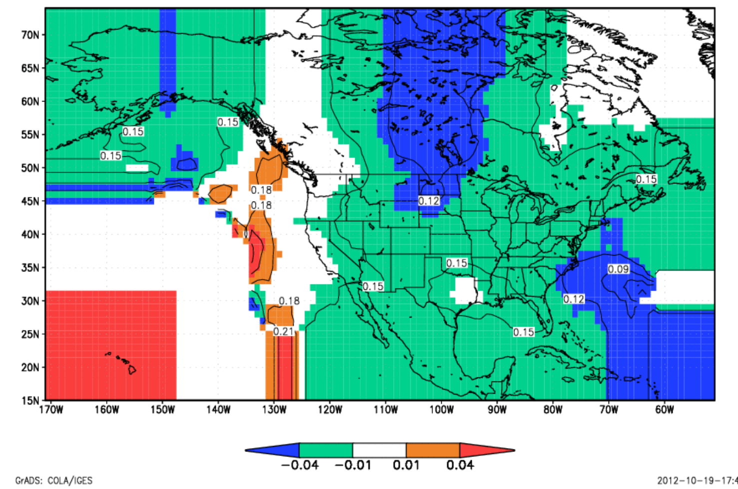
- **Colors indicate excess or deficit relative to an equally weighted ensemble**

  Red (ish) = Weighted more than its share

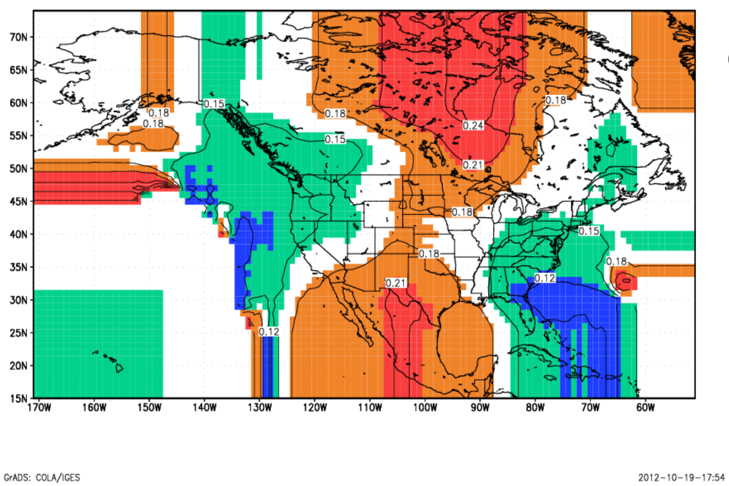
  Blue or green = Weighted less than its share



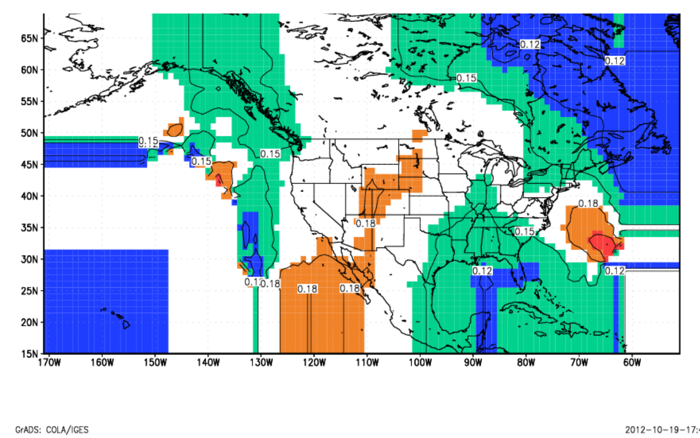
CFSv2



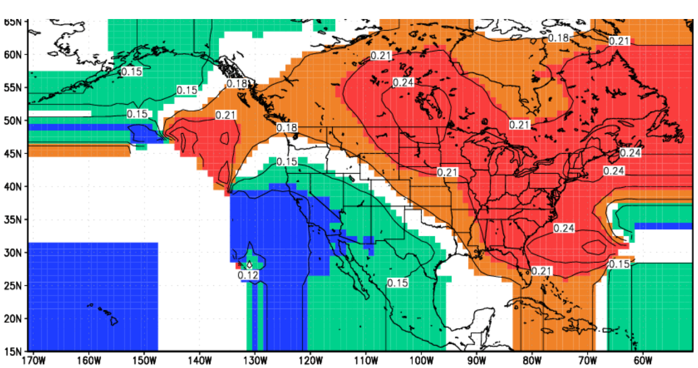
ECHAMF



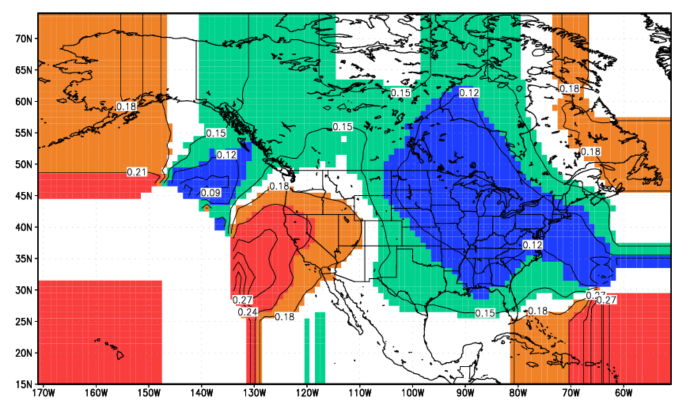
GFDL



ECHAMA



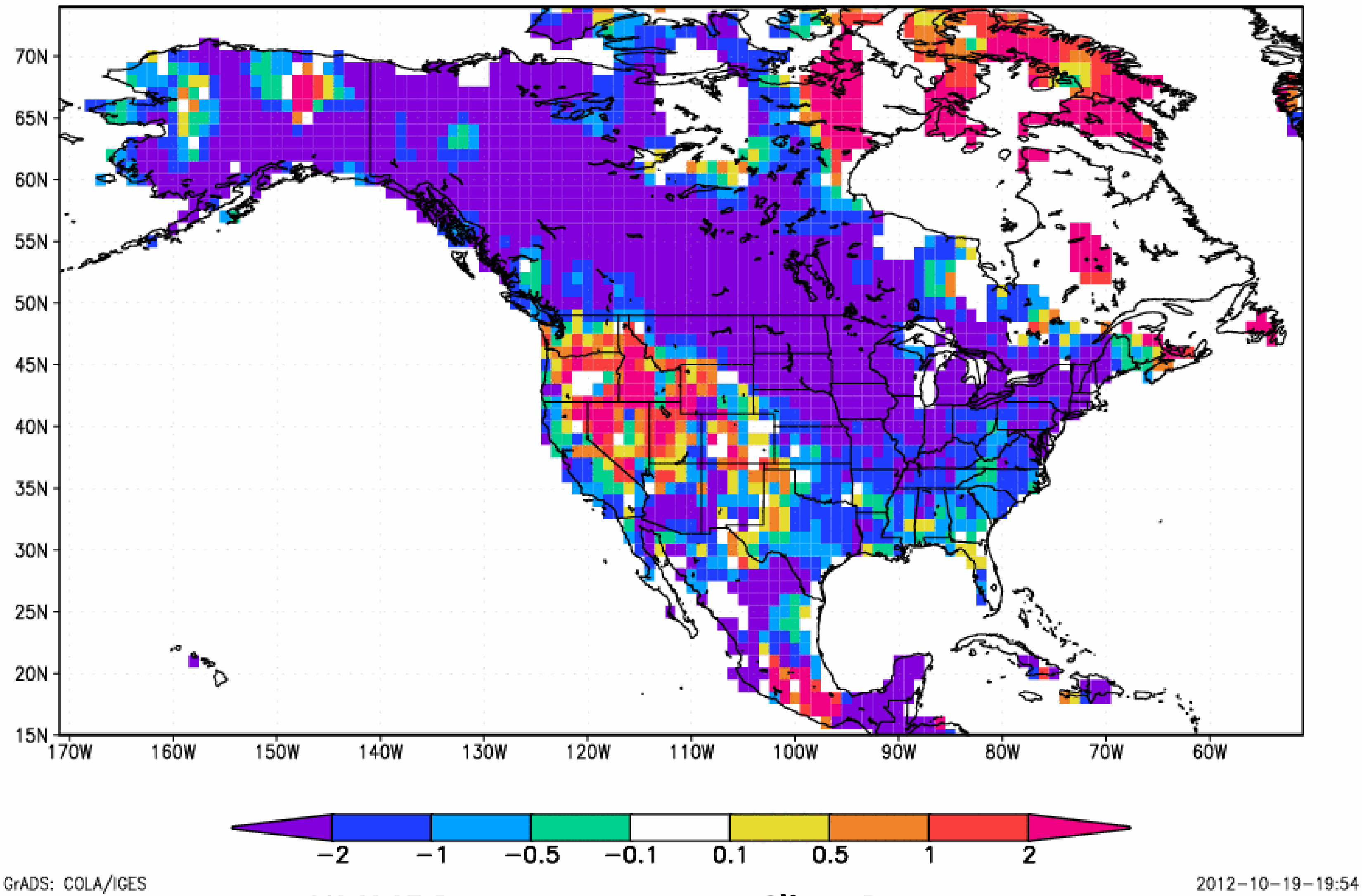
NASA



NCAR

FORECAST PERFORMANCE – CRPS

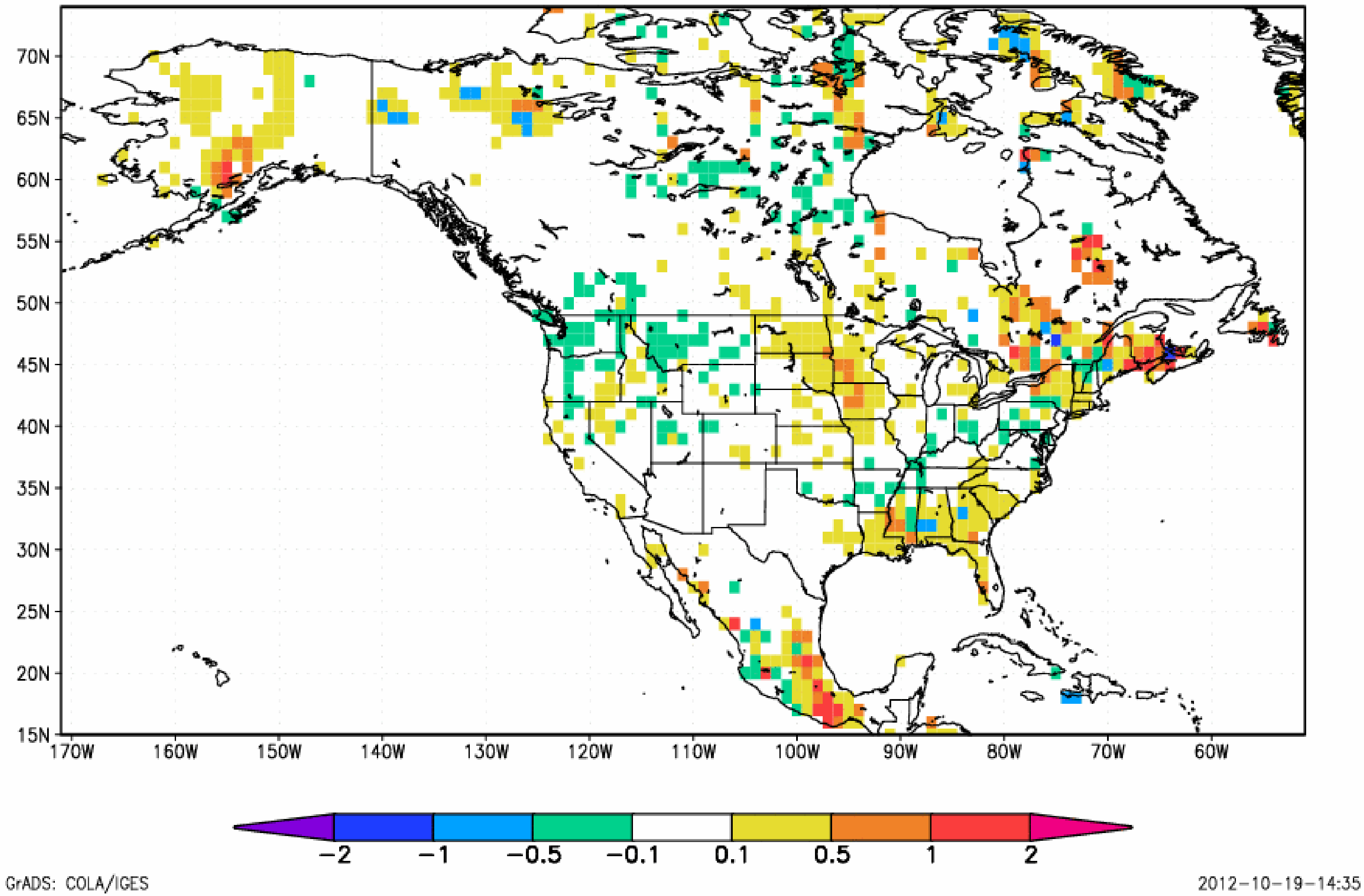
Multi-model vs Climatology



NMME Better **Climo Better**

FORECAST PERFORMANCE – CRPS

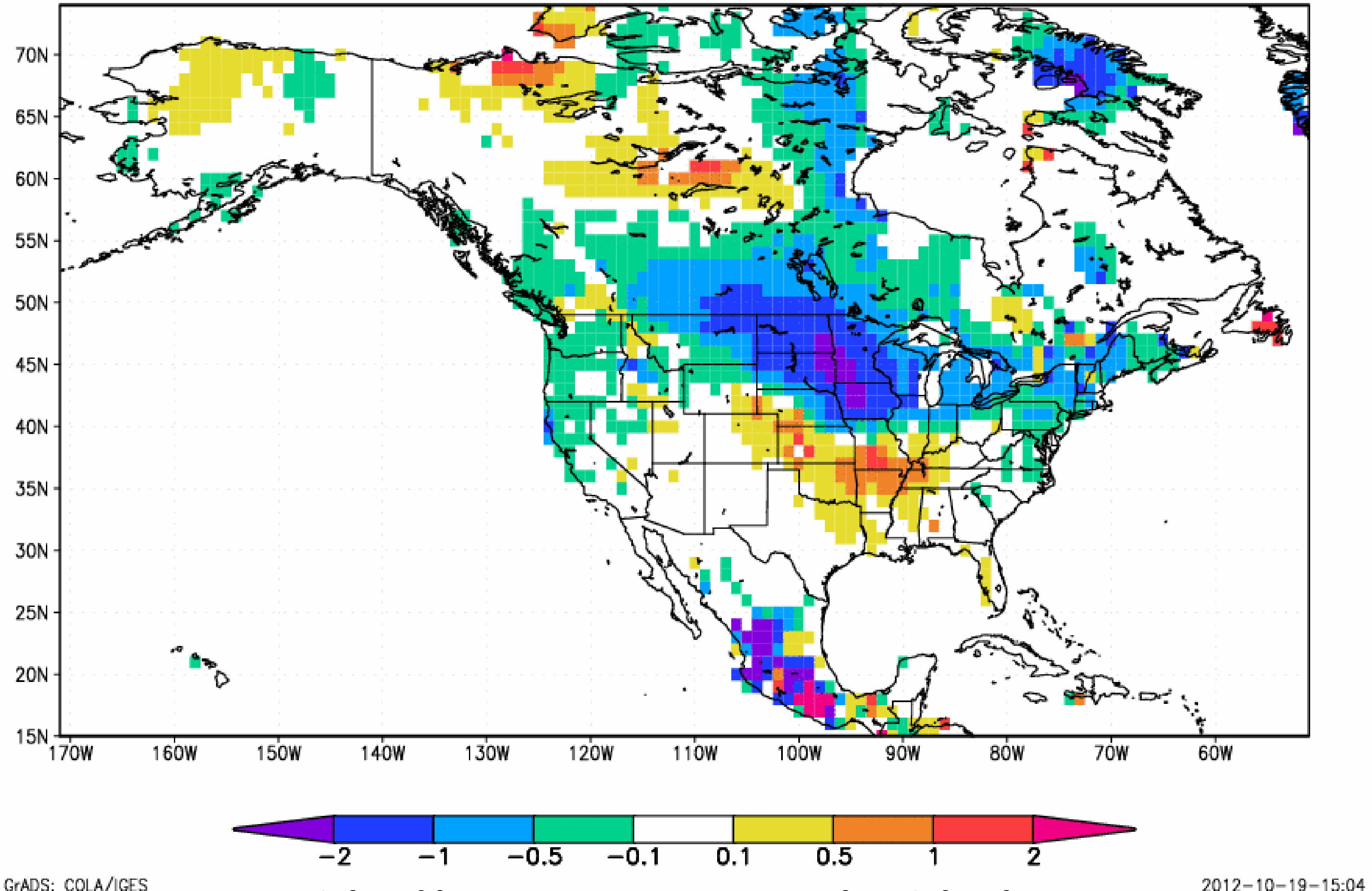
Kernel fit vs. Ensemble mean



Kernels better **Ens. Mean better**

FORECAST PERFORMANCE – CRPS

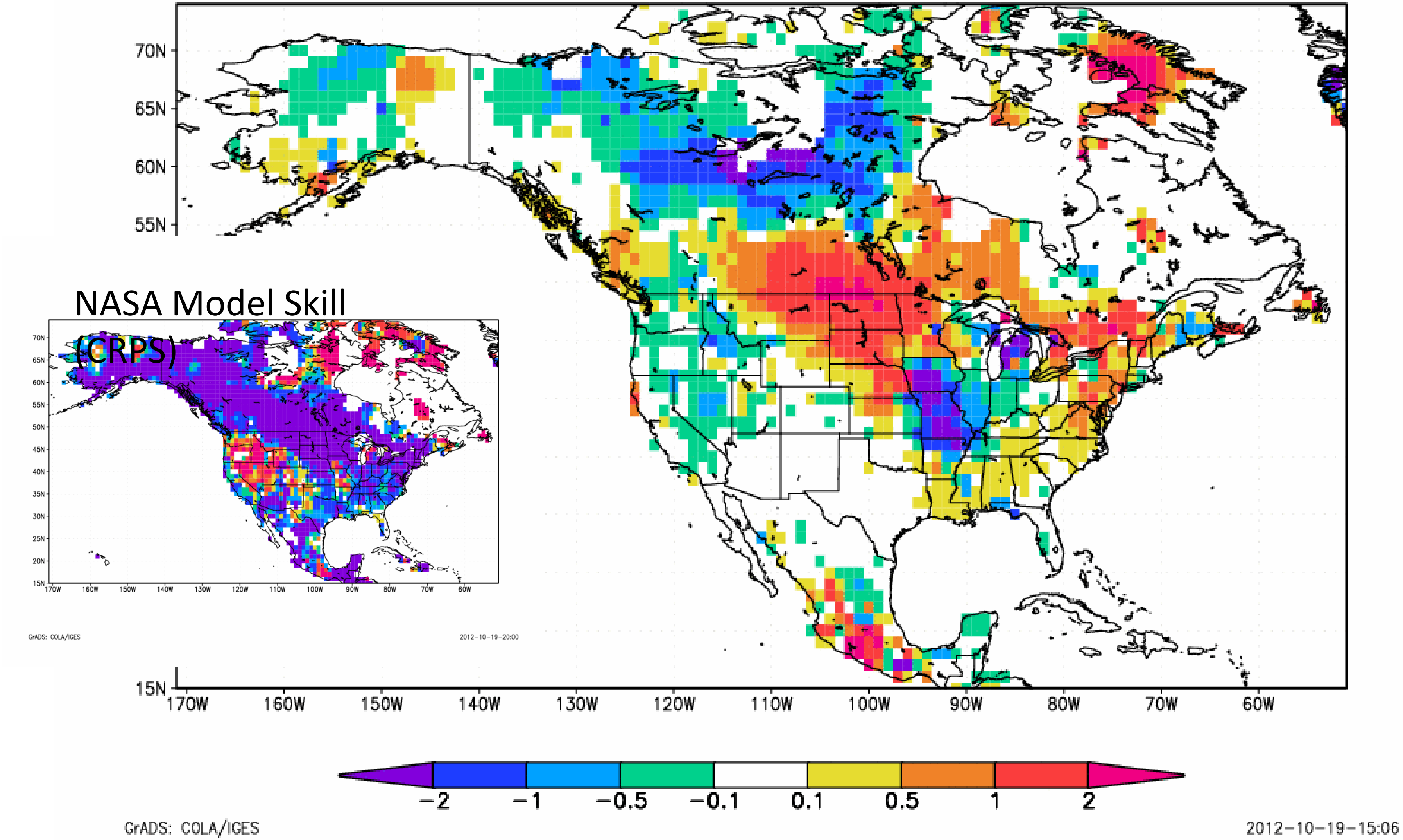
Weighted vs. Equally Weights



Weighted better **Equal weights better**

FORECAST PERFORMANCE – CRPS

Multi-model vs. Individual Model



NMME Better NASA Model Better

CONCLUSIONS

The Good

- The methodology produces reasonable weighting (Models are not easily eliminated)
- Regional skill differences are accounted for.
- “Best PDF” based weighting accounts for model redundancy and differing ensemble size.

The not so good

- Difficult to eliminate really bad models
- Best models can occasionally better the NMME in places.
- Low skill models revert to climatological distributions, which is a good contender in the weighting scheme.

FUTURE WORK

- **Fully adaptive system (Adaptive regression and weighting done together in a integrated and recursive fashion)**
- **A better model elimination procedure needs to be developed**
- **Spatial consistency may need to be addressed.**