



Using reforecasts to calibrate weather predictions

Tom Hamill and Jeff Whitaker

tom.hamill@noaa.gov

jeffrey.s.whitaker@noaa.gov

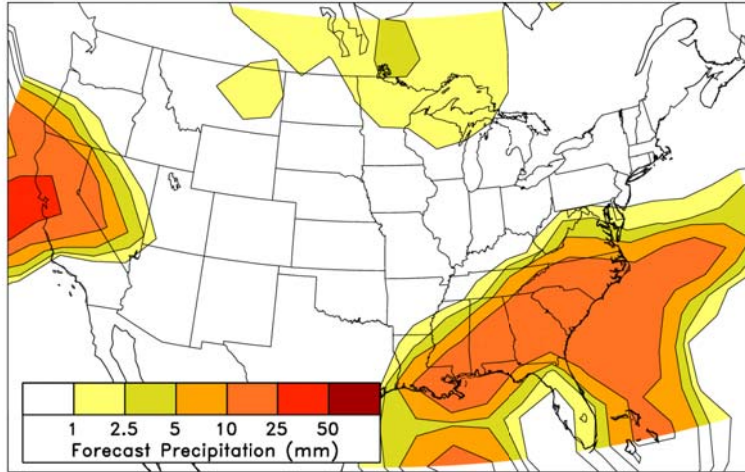
Improving probabilistic weather forecasts

- Better ensembles
 - More members
 - Improved initial conditions
 - Higher resolution
 - Improved forecast models
- Statistical corrections of the NWP forecasts

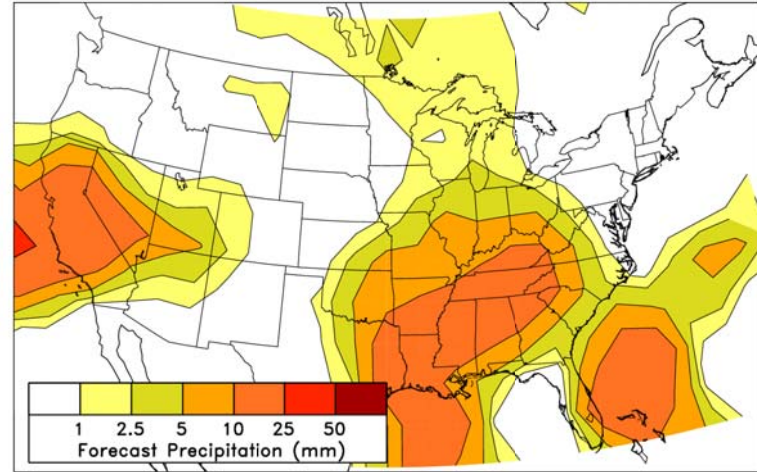
Ensemble-base probabilistic forecasts: problems we'd like to correct

Forecast Initial Time = 0000 UTC 02 Jan 1988

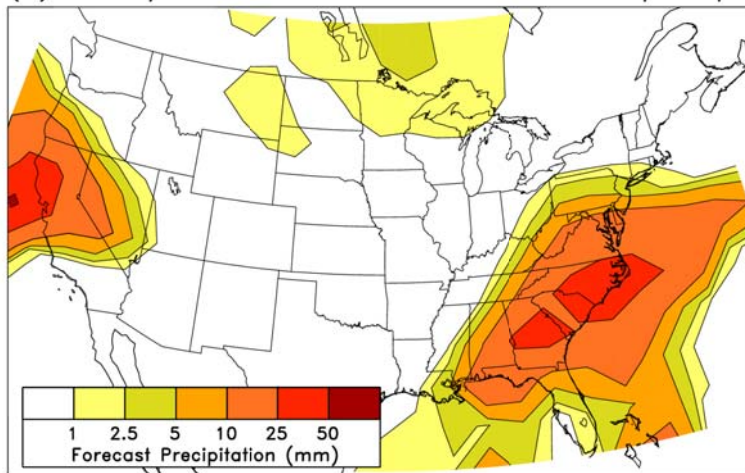
(a) 2-day fcast 24-h accum. member 1 precip



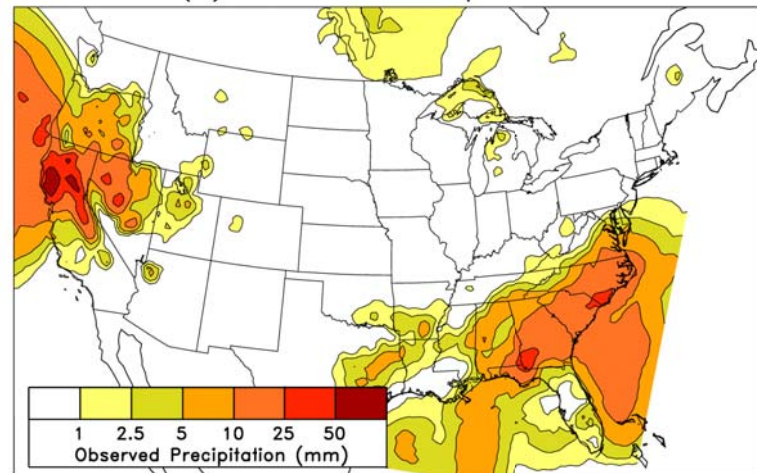
(b) 2-day fcast 24-h accum. member 2 precip



(c) 2-day fcast 24-h accum. member 3 precip



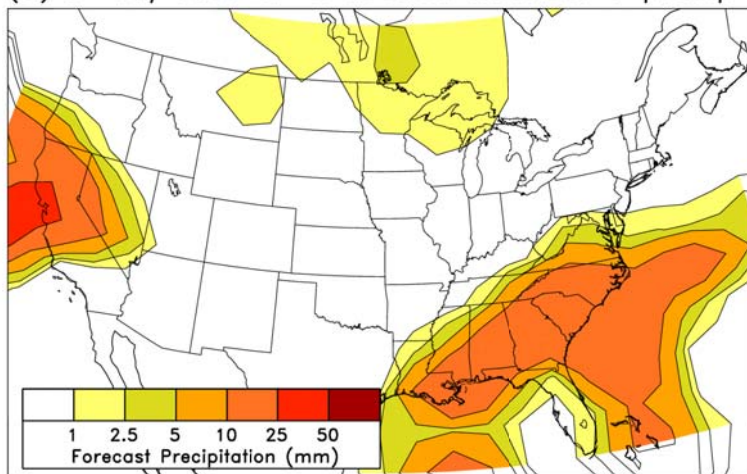
(d) Observed Precipitation



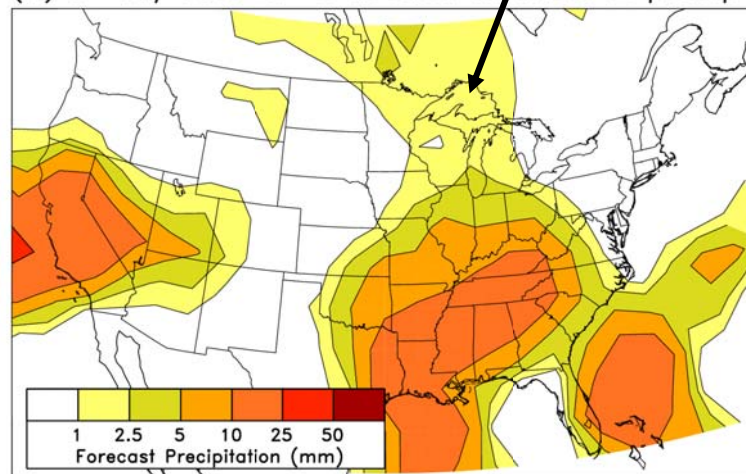
(1) bias (drizzle over-forecast)

Forecast Initial Time = 0000 UTC 02 Jan 1988

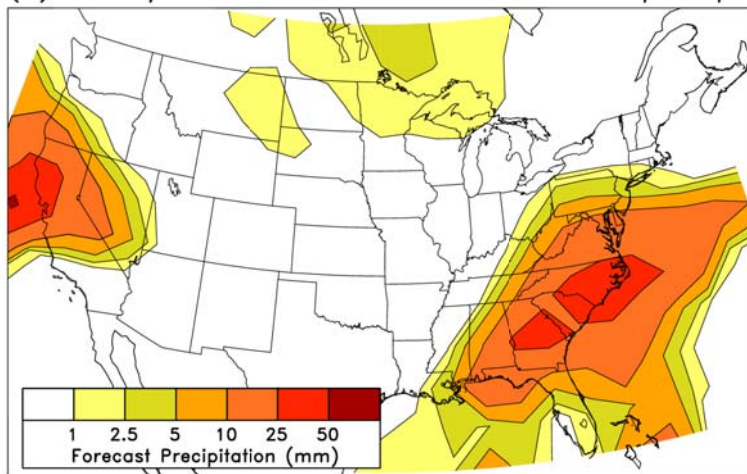
(a) 2-day fcast 24-h accum. member 1 precip



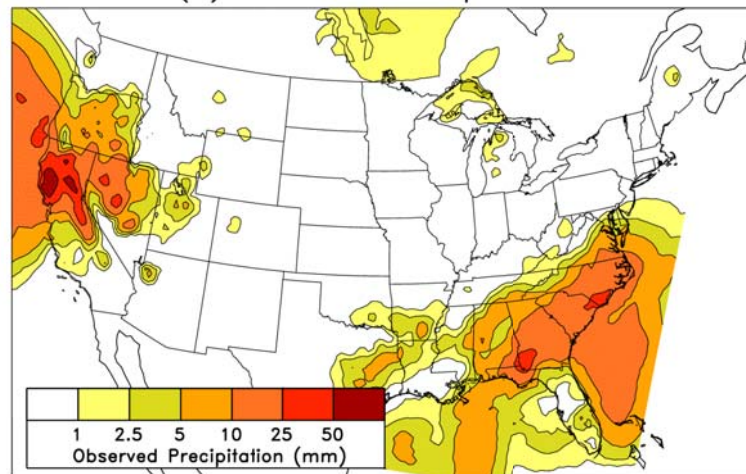
(b) 2-day fcast 24-h accum. member 2 precip



(c) 2-day fcast 24-h accum. member 3 precip



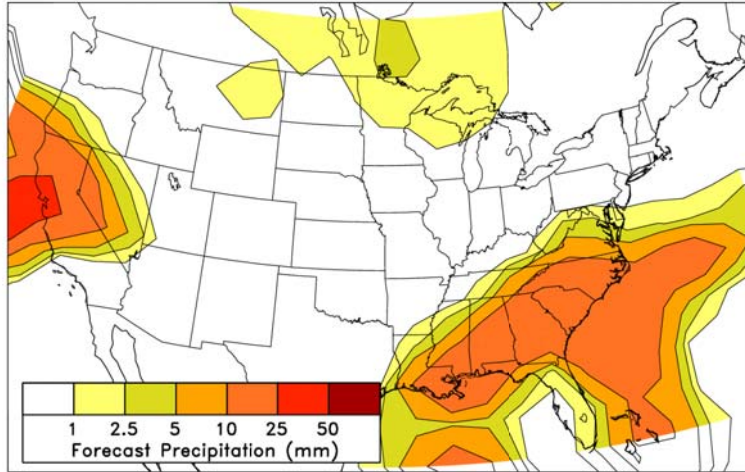
(d) Observed Precipitation



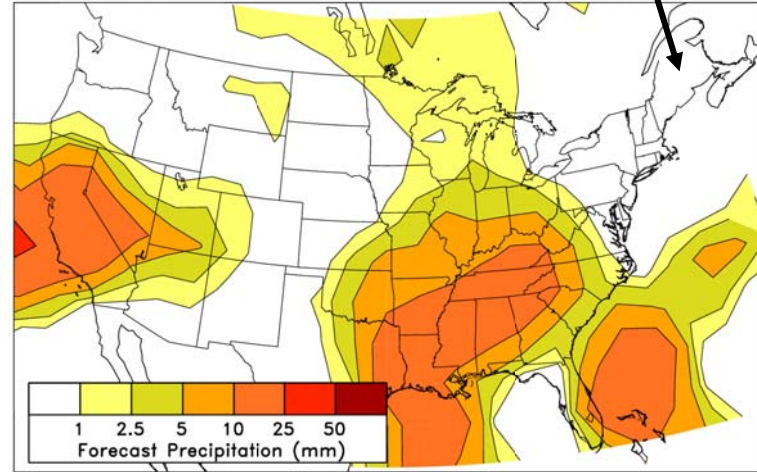
(2) ensemble members too similar to each other.

Forecast Initial Time = 0000 UTC 02 Jan 1988

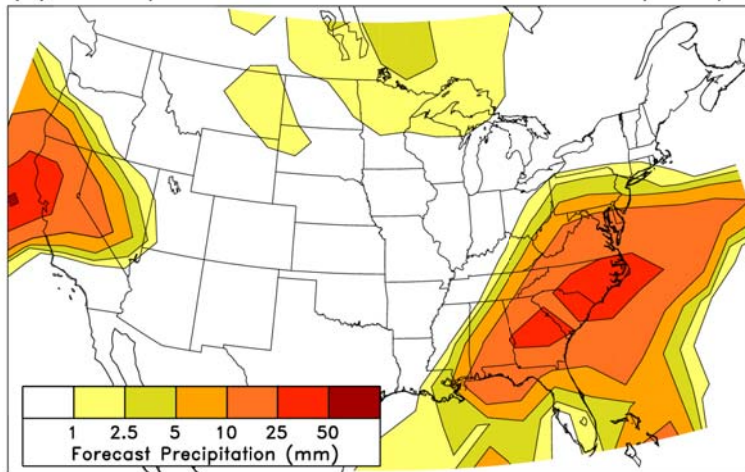
(a) 2-day fcast 24-h accum. member 1 precip



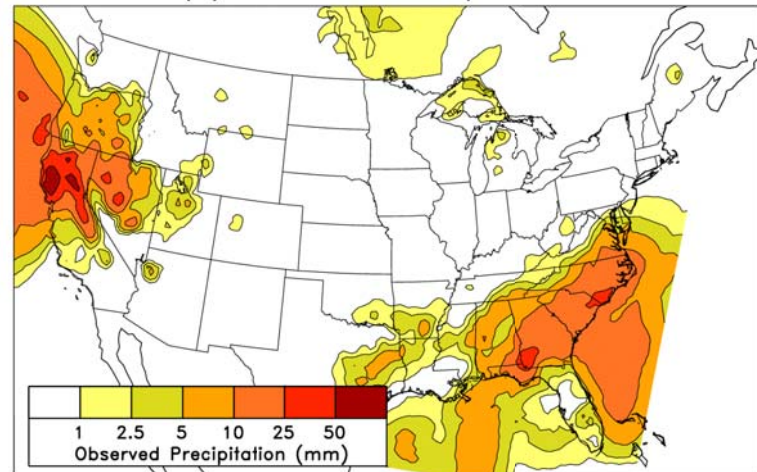
(b) 2-day fcast 24-h accum. member 2 precip



(c) 2-day fcast 24-h accum. member 3 precip



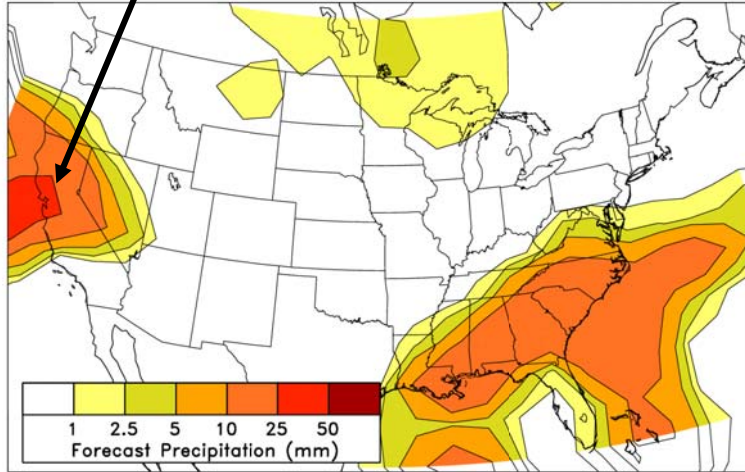
(d) Observed Precipitation



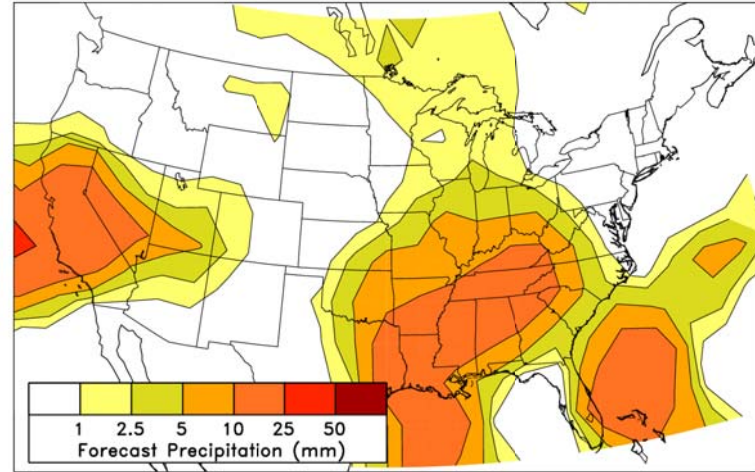
(3) Ensembles are too smooth, not capturing intense local precipitation due to orographic forcing. *Downscaling* needed.

Forecast Initial Time = 0000 UTC 02 Jan 1988

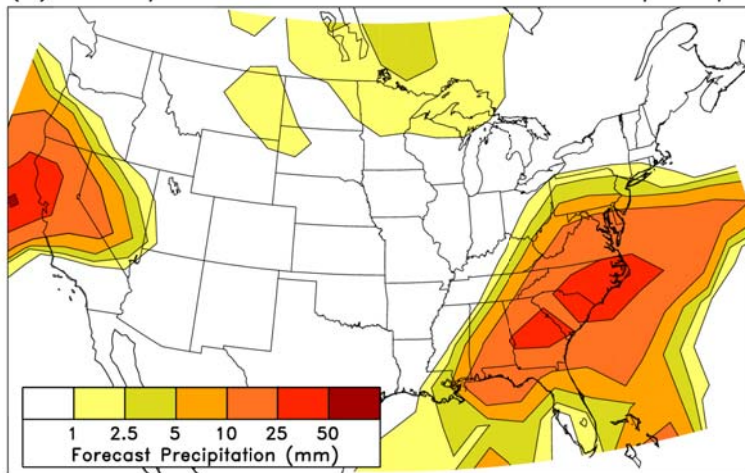
(a) 2-day fcast 24-h accum. member 1 precip



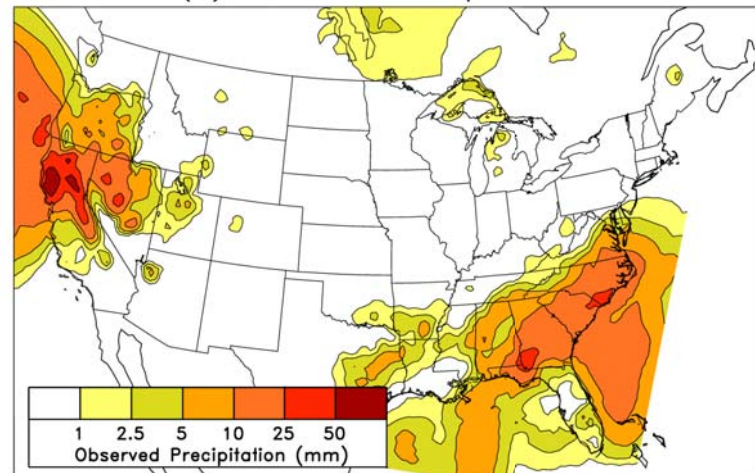
(b) 2-day fcast 24-h accum. member 2 precip



(c) 2-day fcast 24-h accum. member 3 precip



(d) Observed Precipitation



CDC Reforecast Data Set

- **Definition:** a data set of retrospective numerical forecasts using the same model to generate real-time forecasts
- **Model:** T62L28 NCEP global forecast model, circa 1998 (<http://www.cdc.noaa.gov/people/jeffrey.s.whitaker/refcst> for details).
- **Initial States:** NCEP-NCAR reanalysis plus 7 +/- bred modes (Toth and Kalnay 1993).
- **Duration:** 15 days runs every day at 00Z from 1978/11/01 to now. (<http://www.cdc.noaa.gov/people/jeffrey.s.whitaker/refcst/week2>).
- **Data:** Selected fields (winds, hgt, temp on 5 press levels, precip, t2m, u10m, v10m, pwat, prmsl, rh700, heating). NCEP/NCAR reanalysis verifying fields included (Web form to download at <http://www.cdc.noaa.gov/reforecast>).

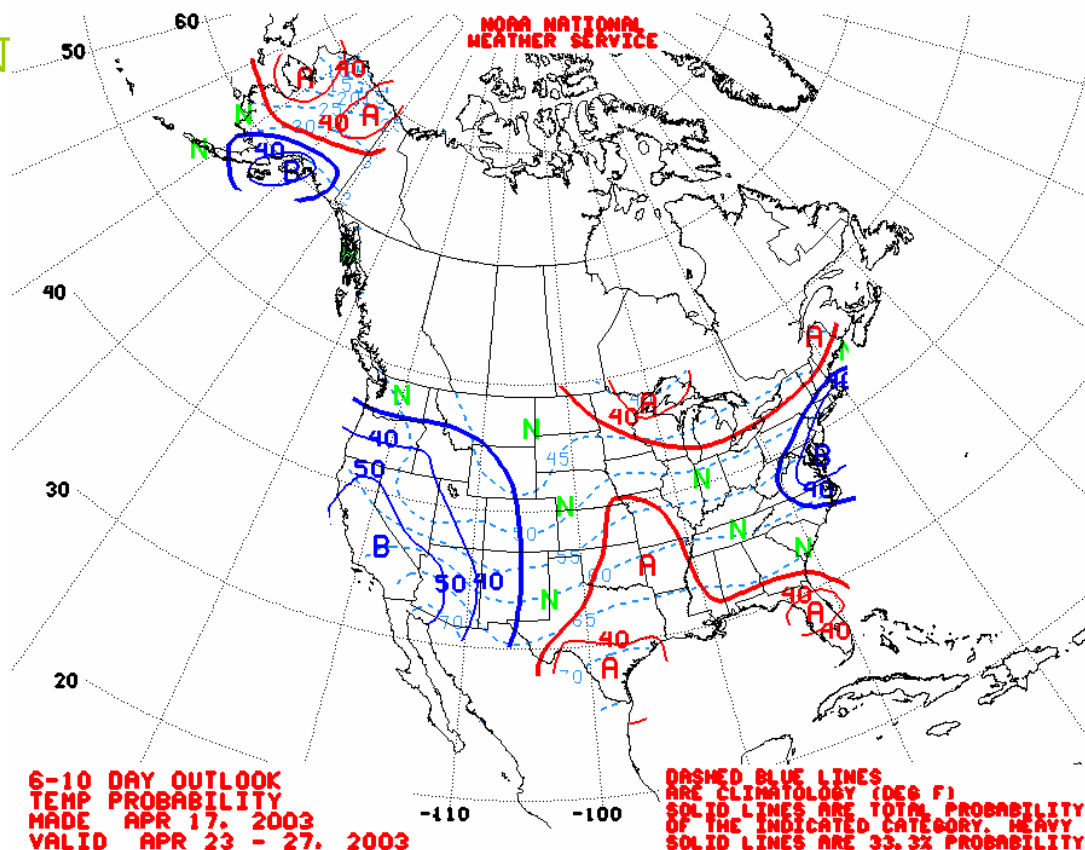
Application: tercile probability forecasts

Climatological distribution split into 3 equally likely bins.

These categories are often called

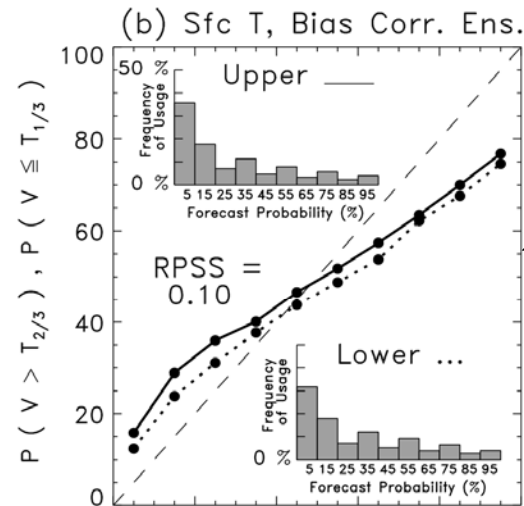
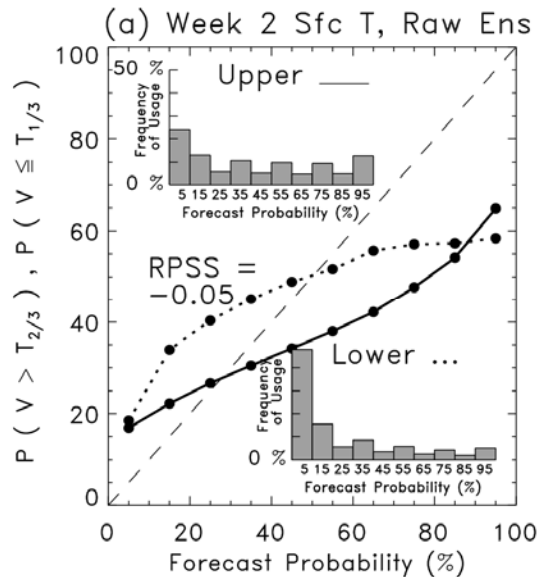
Below/N

NCEP
Climate
Prediction
Center
(CPC)
operational
product

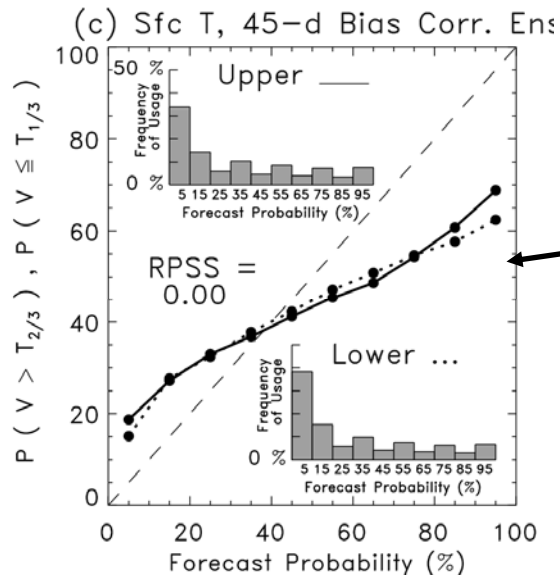


There's more to making a good probabilistic forecast than just correcting model bias

Probabilities from raw ensemble

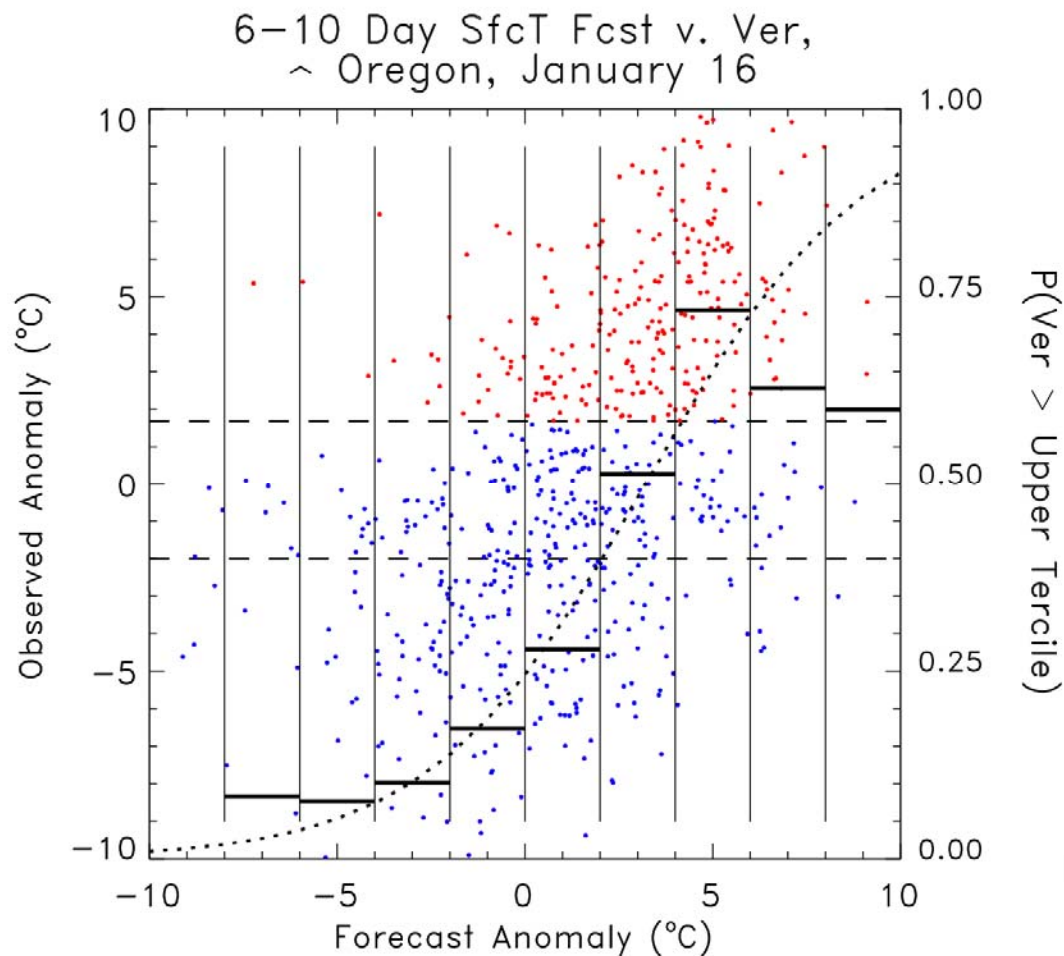


Correction of biases estimated from full 22 years of forecast data



Correction of biases estimated from last 45 days of data

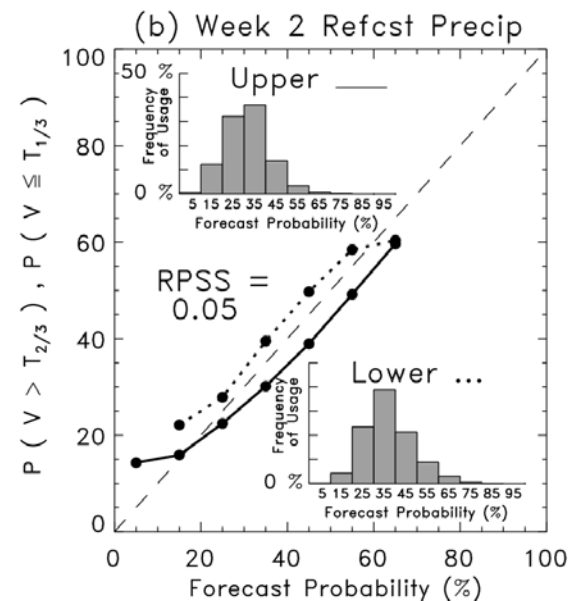
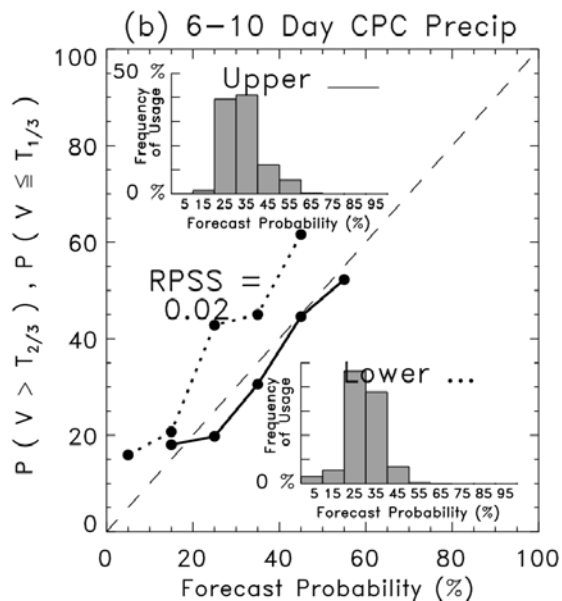
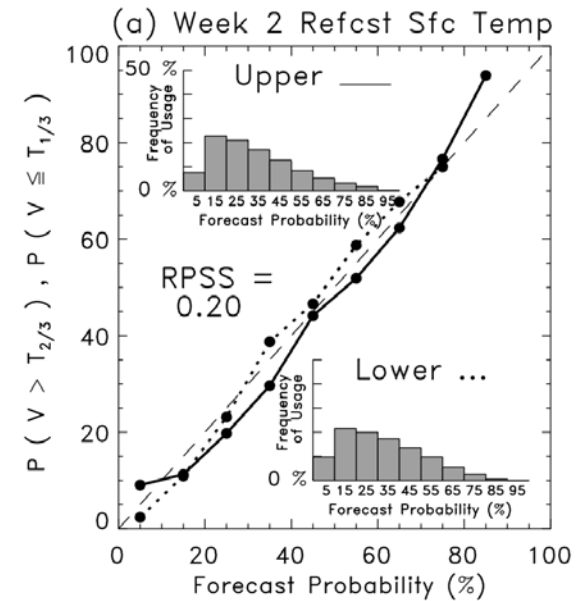
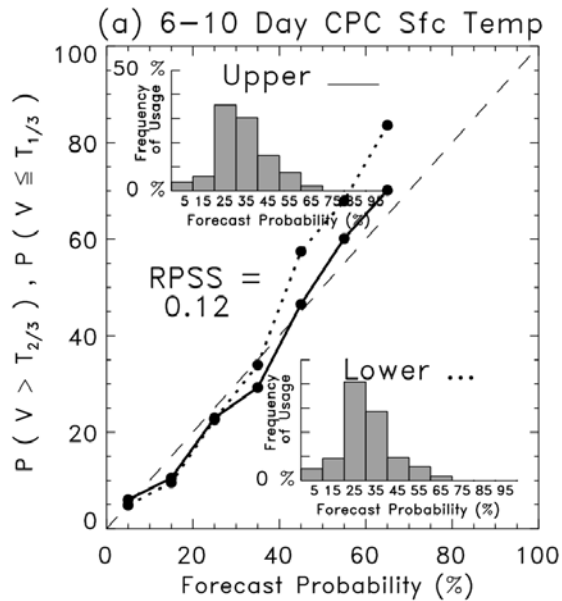
What can we do with a long data set of observed and forecast anomalies?



With our reforecasts, we have 23 years of data. Let's use old data in a 31-day window around the date of interest to make statistical corrections.

Dashed lines: tercile boundaries
Red points: samples above upper tercile
Blue points: samples below upper tercile
Solid bars: probabilities by bin count
Dotted line: a fitted model, TBD

Comparison against NCEP / CPC forecasts at 155 stations, 100 days in winter 2001-2002

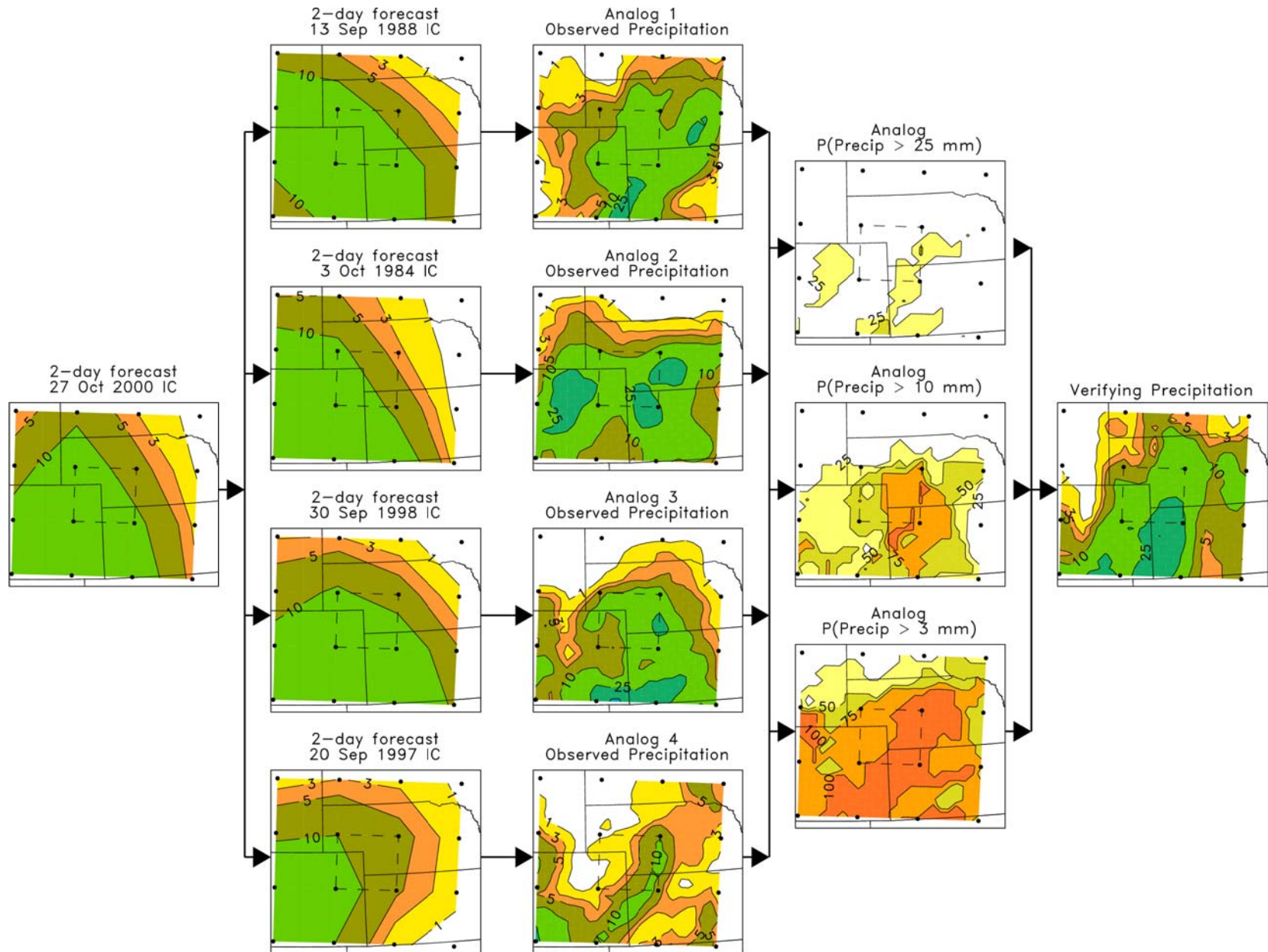


MOS-based
Week 2
forecasts
using low-res
T62 model
more skillful
than
operational
NCEP/CPC
6-10 day!

[link: training
set size](#)

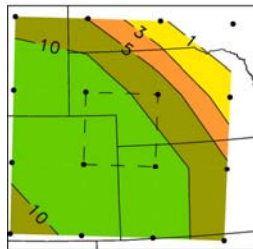
[link:
results from
full 23 years](#)

Analog high-resolution precipitation forecast technique

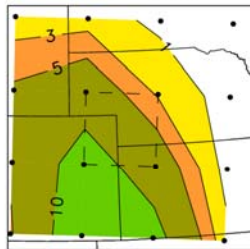


20 closest forecast analogs

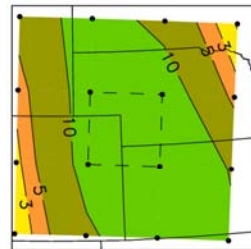
13 Sep 1988 IC
SNR = 3.19



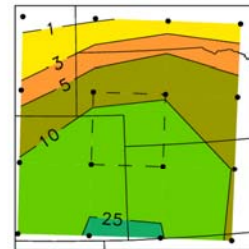
20 Oct 1986 IC
SNR = 1.72



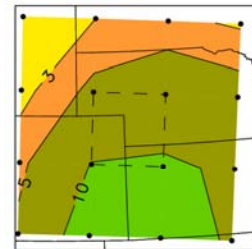
17 Sep 1996 IC
SNR = 1.61



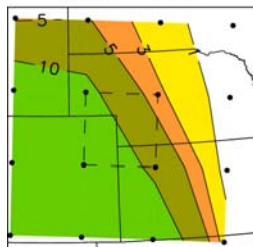
20 Sep 1985 IC
SNR = 1.48



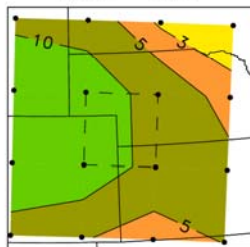
17 Oct 1983 IC
SNR = 1.38



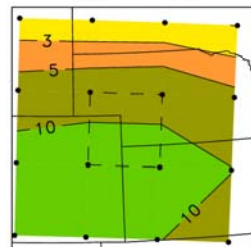
3 Oct 1984 IC
SNR = 2.84



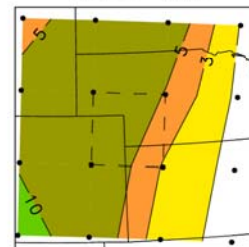
14 Nov 1991 IC
SNR = 1.67



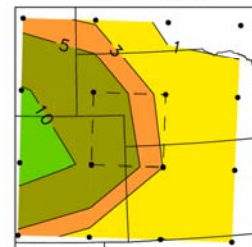
19 Sep 1997 IC
SNR = 1.55



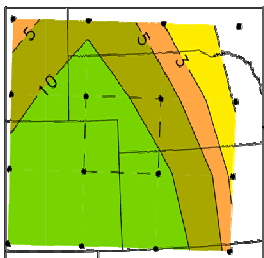
11 Nov 1994 IC
SNR = 1.46



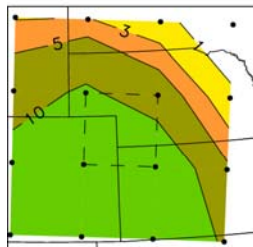
21 Nov 1999 IC
SNR = 1.35



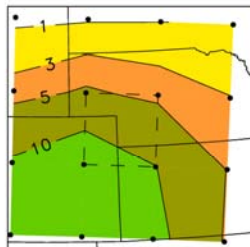
2-day forecast
27 Oct 2000 IC



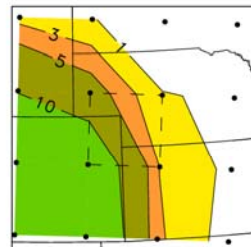
30 Sep 1998 IC
SNR = 2.06



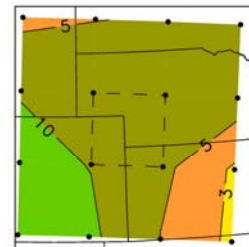
28 Sep 1990 IC
SNR = 1.65



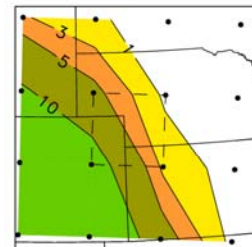
26 Oct 1998 IC
SNR = 1.52



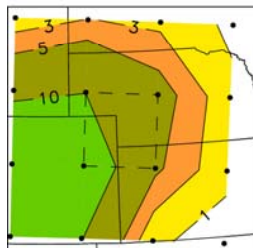
21 Sep 1997 IC
SNR = 1.45



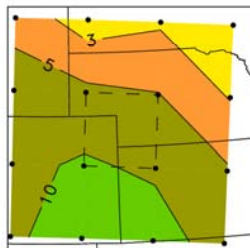
19 Oct 1986 IC
SNR = 1.33



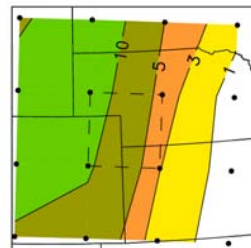
20 Sep 1997 IC
SNR = 1.98



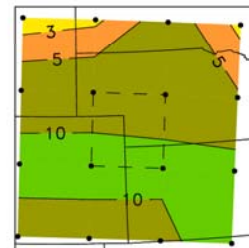
21 Oct 2000 IC
SNR = 1.64



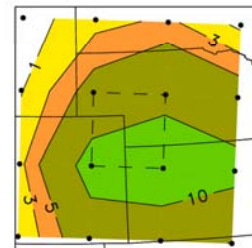
25 Oct 1989 IC
SNR = 1.51



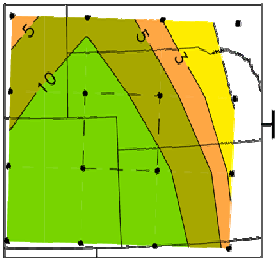
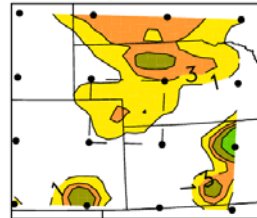
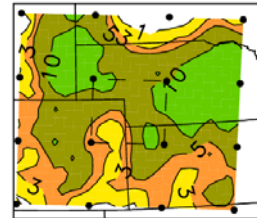
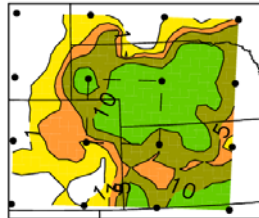
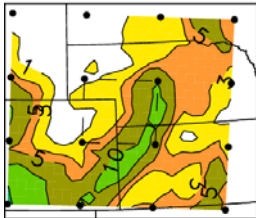
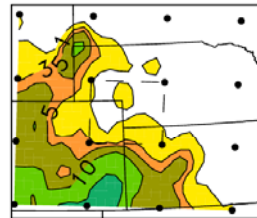
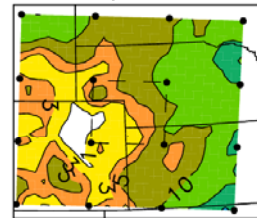
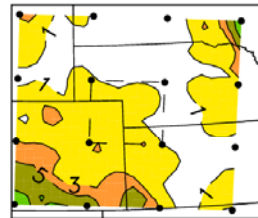
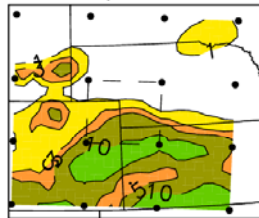
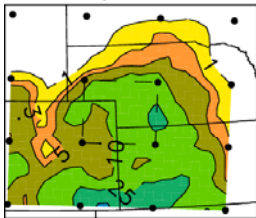
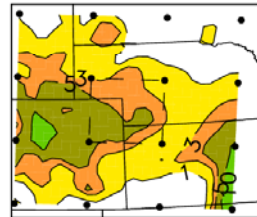
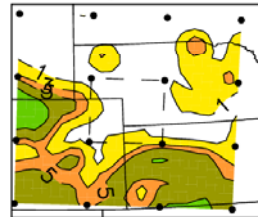
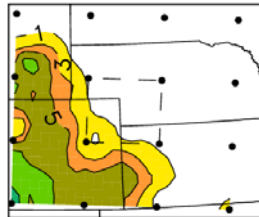
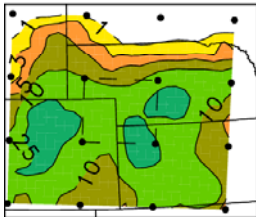
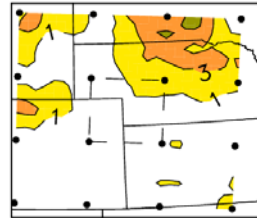
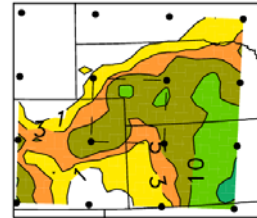
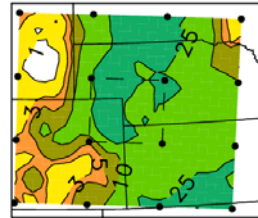
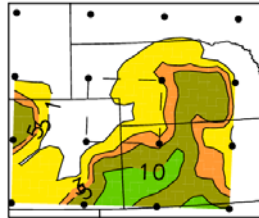
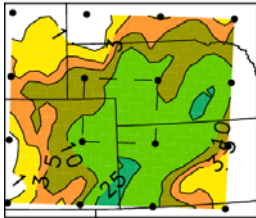
4 Oct 1984 IC
SNR = 1.41



13 Sep 1984 IC
SNR = 1.30

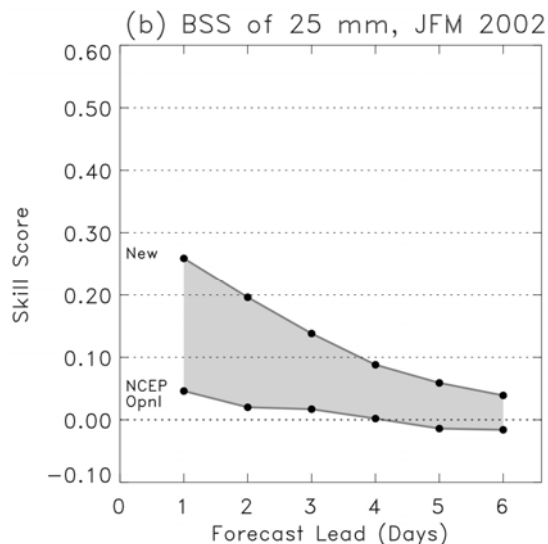
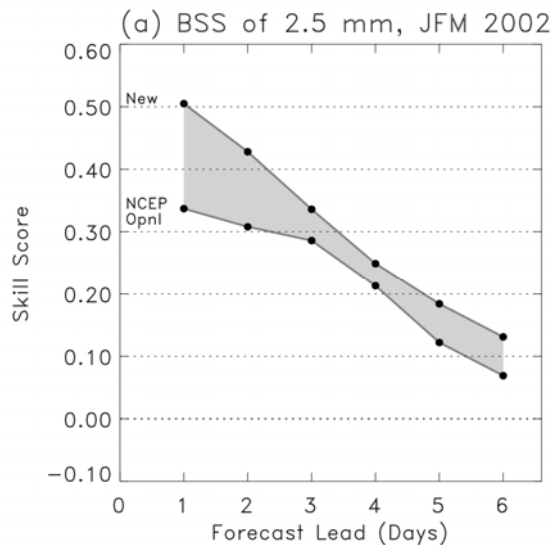


Associated 20 precipitation analyses



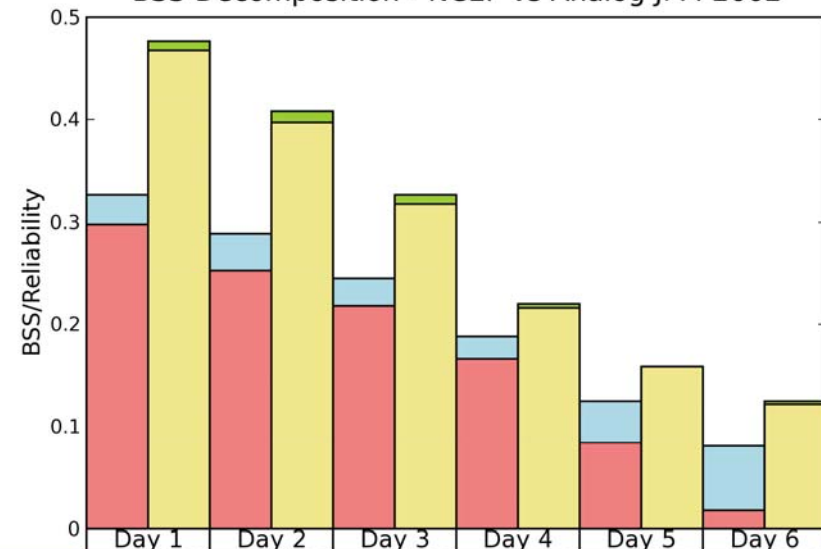
Comparison against NCEP ensemble

$$BSS = \frac{\text{resolution} - \text{reliability}}{\text{uncertainty}}$$



Upper Quintile

BSS Decomposition - NCEP vs Analog JFM 2002



NCEP BSS	0.298	0.253	0.218	0.166	0.084	0.018
Analog BSS	0.468	0.397	0.318	0.216	0.159	0.122
NCEP Resolution	0.327	0.289	0.245	0.188	0.125	0.081
Analog Resolution	0.477	0.408	0.327	0.220	0.159	0.125
NCEP Reliability	0.029	0.036	0.027	0.022	0.041	0.063
Analog Reliability	0.009	0.011	0.009	0.004	0.000	0.003

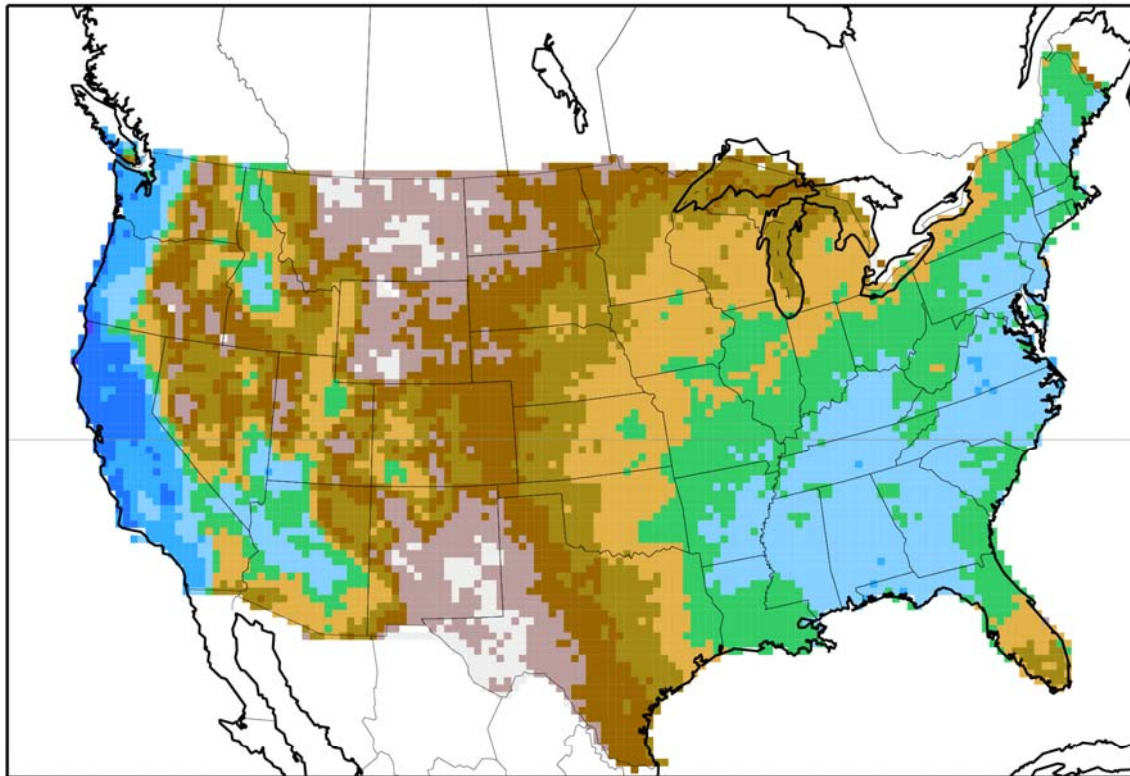
the improvement is a little bit of increased reliability, a lot of increased resolution.

Skill as function of location

JFM24 Analog Precip Fcst BSS (1979-2003)

Analog Prob Precip > 2.5mm

Day 4



Brier Skill Score

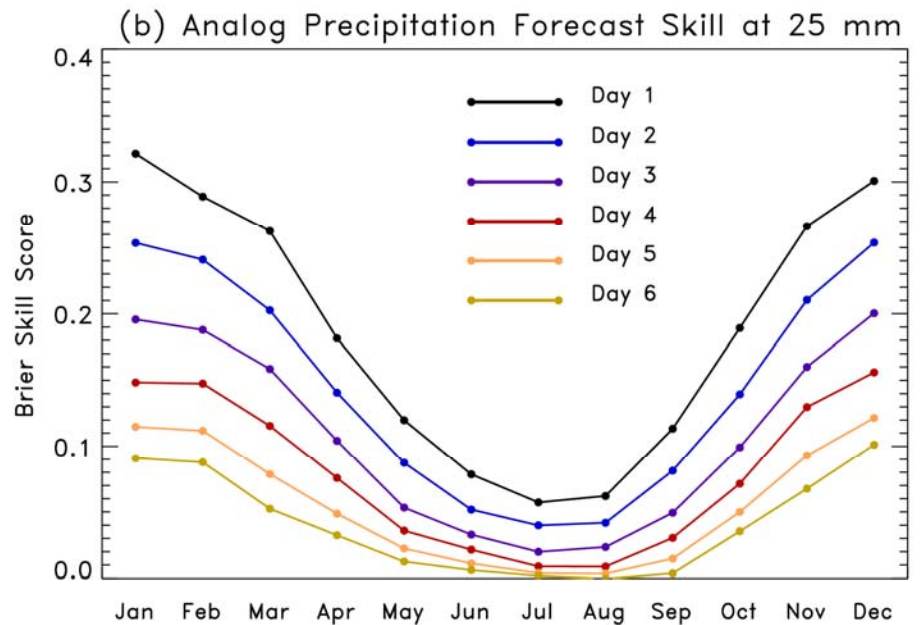
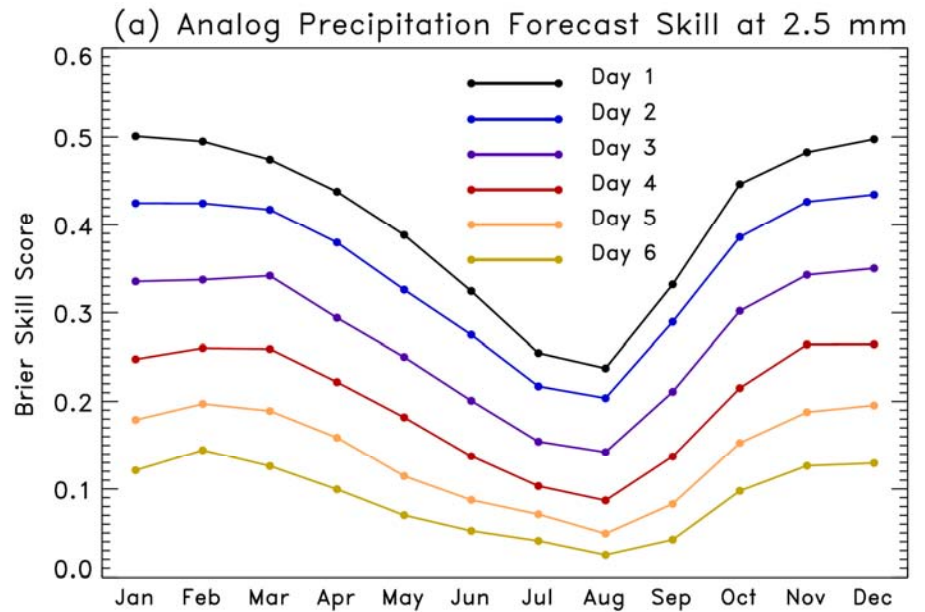


Notes:

(1) Less skill where it's dry (climatological forecasts better here).

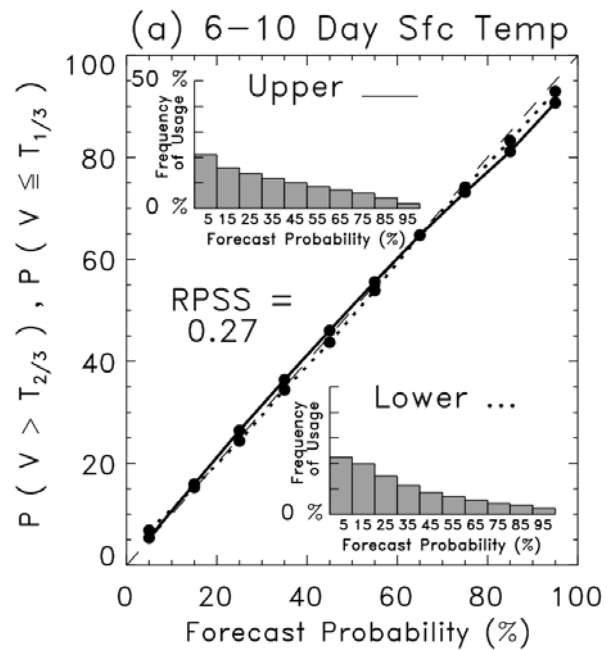
(2) Regions where precipitation analyses are poor are less skillful (snowy regions, poor coverage by gages & Doppler)

Skill as function of time of year

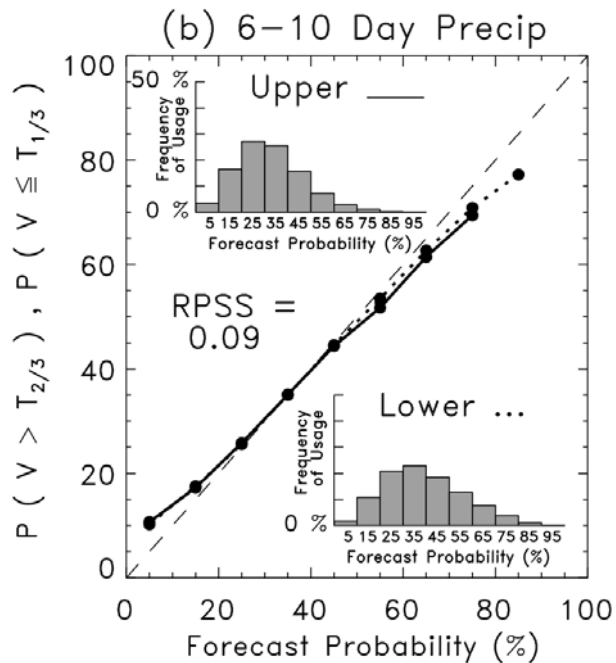
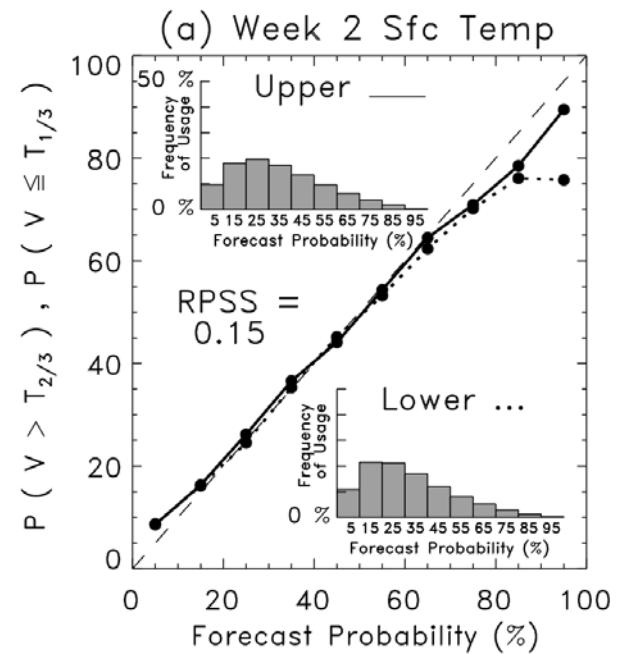


Conclusions

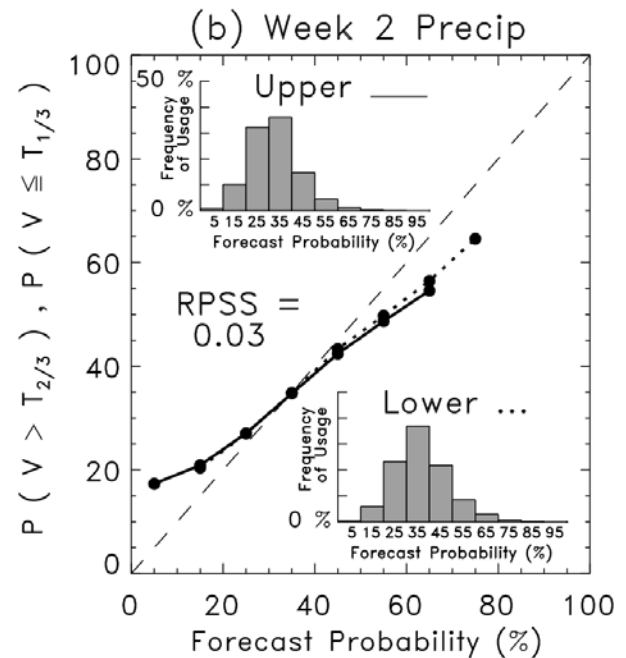
- Great promise for improving forecasts. We've shown that with **numerical forecasts from old, out-of-date model** at low resolution + **statistical corrections** based on extensive reforecasts -> **forecasts competitive with state-of-the-art** model forecasts.
- Potential for **even more improvement when reforecasts used with newer, higher-resolution forecast models**
- Real-time and archived data freely available for your use.
- Future work:
 - Further improvement of reforecast techniques, applications to specific user problems.
 - Foster broader testing of reforecasts by other groups.
 - Get output into hands of forecasters for evaluation.
 - Try higher-resolution reforecasts.
 - Assuming success, plan path to make an expanded suite of reforecast-based products a part of operational NWP in NWS.



6-10
Day



Week
2



Web Form for Downloading Data

Reforecast Ensemble Data

<http://www.cdc.noaa.gov/reforecast/>



Jump to: [Map Room Weather Products](#)

Search for: Go!

You are at: [CDC Home](#) > Reforecast Ensemble Data

[Home](#) | [Search](#) | [Site index](#) | [Privacy policy](#) | [Disclaimer](#) | [Contact us](#)

Download Reforecast Ensemble Data

The web download form is currently out of order. You can still download the netCDF forecast files directly from the [ftp server](#).

Select variable and associated level:

Variable Field:

Level(mb):

Select date range (available from 1978110100 to 2003123100):

Start Year:

Start Month:

Start Day:

End Year:

End Month:

End Day:

download all the forecasts within the chosen time period

Example: if the start date is "1979 Jan 1st" and the end date is "1990 Feb 28th", you will get all the forecasts between these dates.

download forecast subsets within the chosen time period

Example: if the start date is "1979 Jan 1st" and the end date is "1990 Feb 28th", you will get the dates between Jan 1 and Feb 28 for all the years between 1979 and 1990 (Jan1-Feb28 1979, Jan1-Feb28 1980, Jan1-Feb28 1990)

Forecast hour range:

Start Hour:

End Hour:

individual time step

Example: if start hour=12, end hour=48, you will get the individual forecast lead times between 12hr and 48hr (i.e. 12h, 24h, 36h and 48h forecasts).

average for this time period

Example: To obtain week2 mean value, select 180h (start hour), 336h (end hour)

Select download type:

☐ all ensemble members
☐ ensemble mean

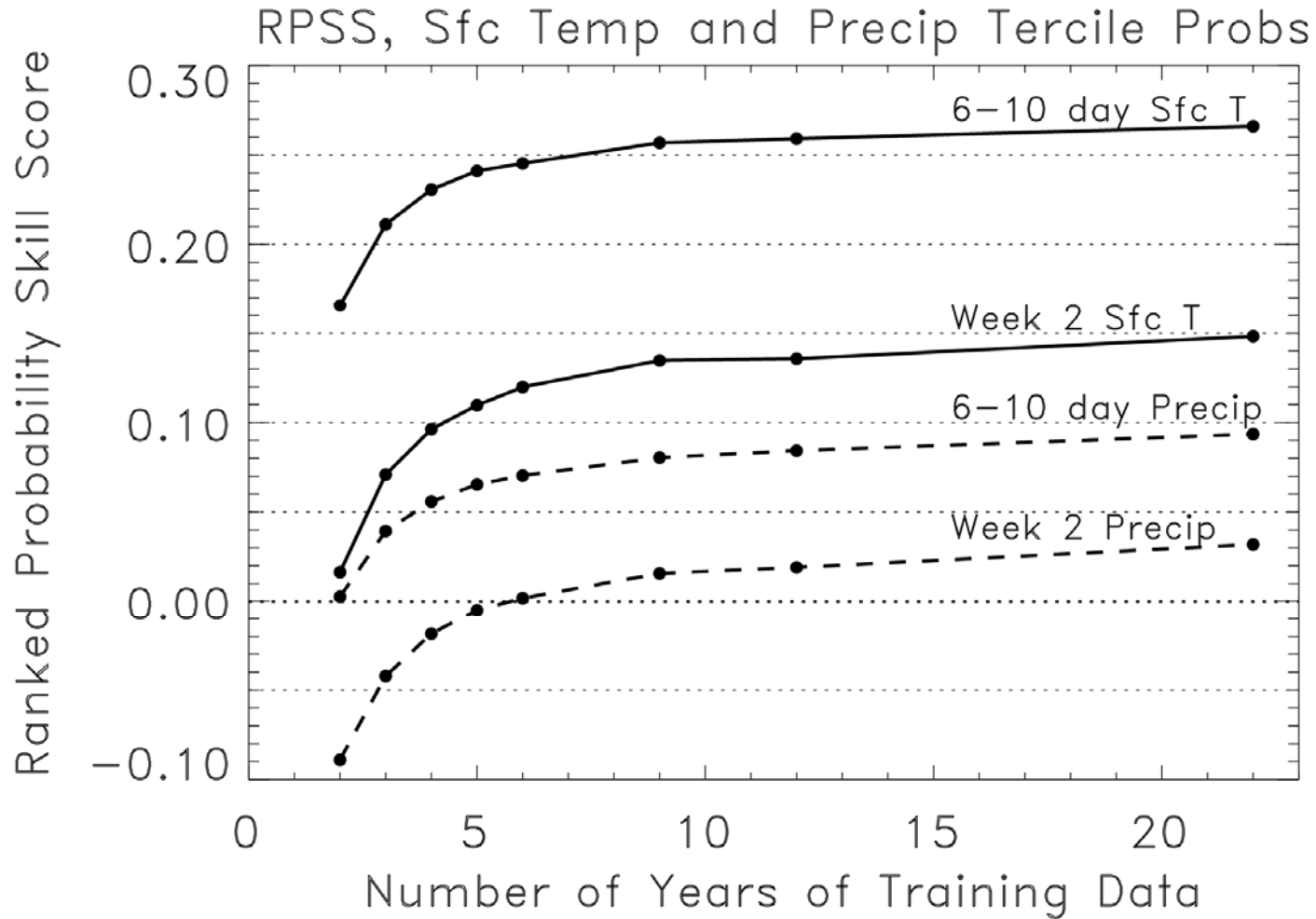
☐ control run
☐ verification data

e-mail address to notify when file is ready: Create the file!

[NOAA-CIRES Climate Diagnostics Center](#)
Document maintained by [Xue Wei \(xue.wei@noaa.gov\)](#)
Updated: May 25, 2004 09:31:19 MDT
<http://www.cdc.noaa.gov/reforecast/index.html>

[back](#)

How many re-forecasts do we really need?



[\[back\]](#)

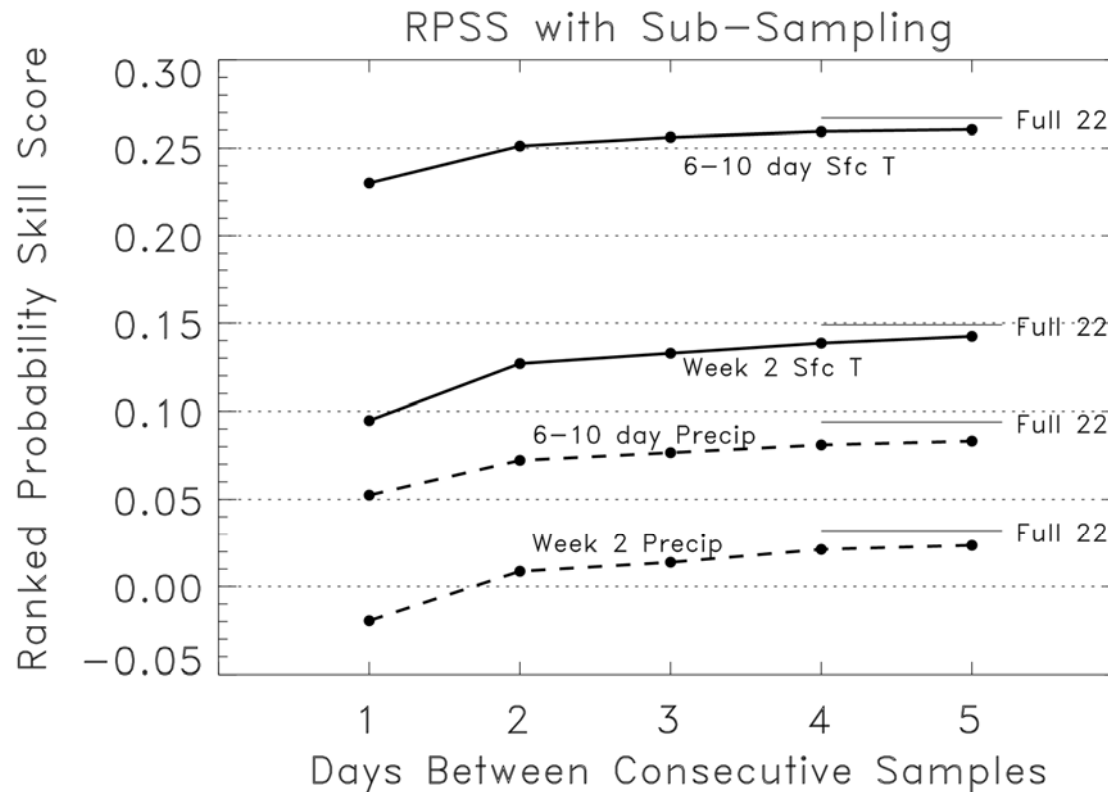
for this application, most of the benefit with 10 years

[link:](#)
[skipping days](#)

How much data do we need, continued

Errors of consecutive daily forecasts are not independent;
what if we skip days?

4 years of training data, skill as a function of days between forecasts



for this application, 4 years of data may be enough - if ensembles are run every 5 days over 20 years. (Still need control run & bred cycle every day)

Forecast calibration with MOS: logistic regression

- For each grid point (or station) let x = continuous predictor data (ens. mean forecast value), y = binary predictand data (1.0 if predicted event happened, 0.0 if not).
- Problem: Compute $P(y=1.0 \mid x)$ as a continuous function of x .
- Logistic Regression:
$$P = \frac{1}{1 + \exp(\beta_0 + \beta_1 x)}$$