NOAA Earth System Research Laboratory

Reforecasting and probabilistic weather-climate prediction

Tom Hamill

NOAA Earth System Research Lab

Boulder, CO

tom.hamill@noaa.gov

Motivation

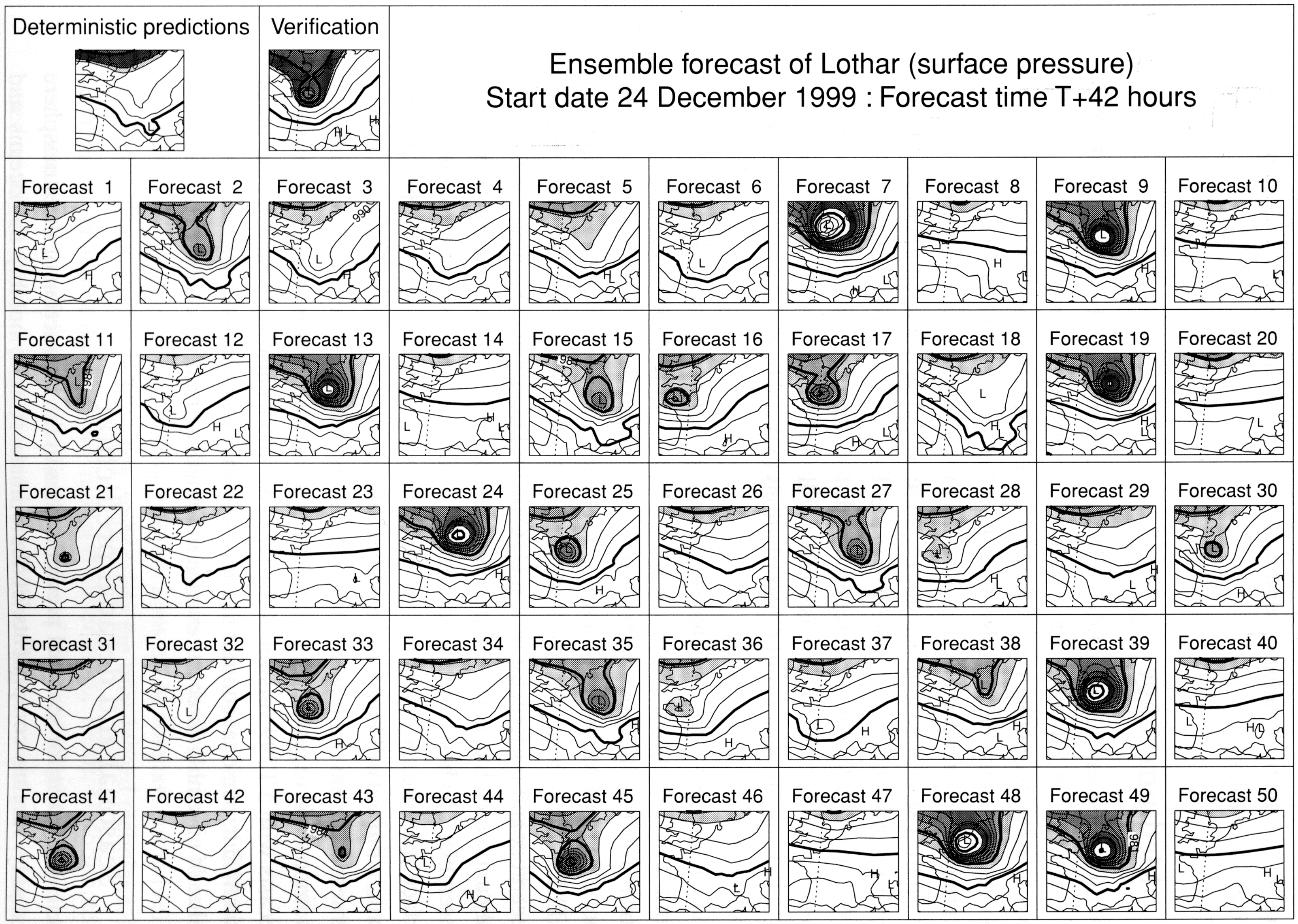
"*Hydrometeorological services in the United States are an Enterprise effort. Therefore, effective incorporation of uncertainty information will require a fundamental and coordinated shift by all sectors of the Enterprise. Furthermore, it will take time and perseverance to successfully make this shift. As the Nation's public weather service, NWS has the responsibility to take a leading role in the transition to widespread, effective incorporation of uncertainty information into hydrometeorological prediction*."

- From finding 1 of 2006 NRC report "Completing the Forecast"

Ensemble forecasts: the backbone

Multiple simulations of the weather from slightly different initial conditions, perhaps different forecast models

Deterministic forecast totally misses damaging storm over France; some ensemble members forecast it well.

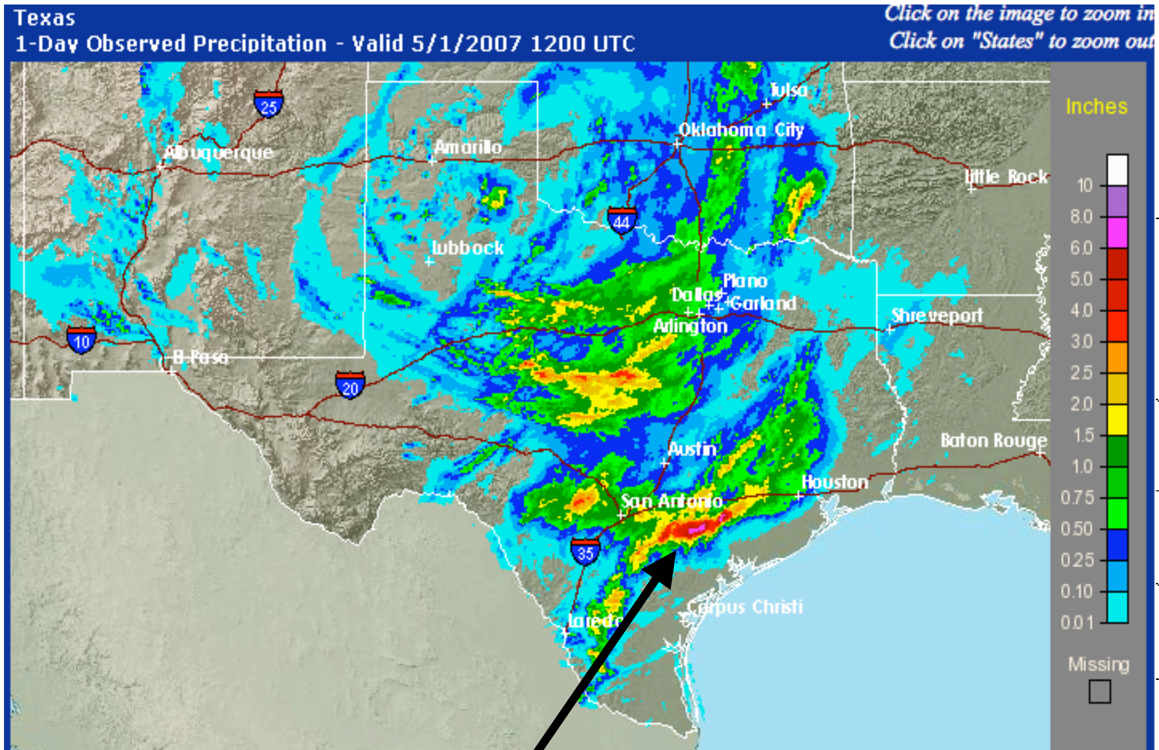


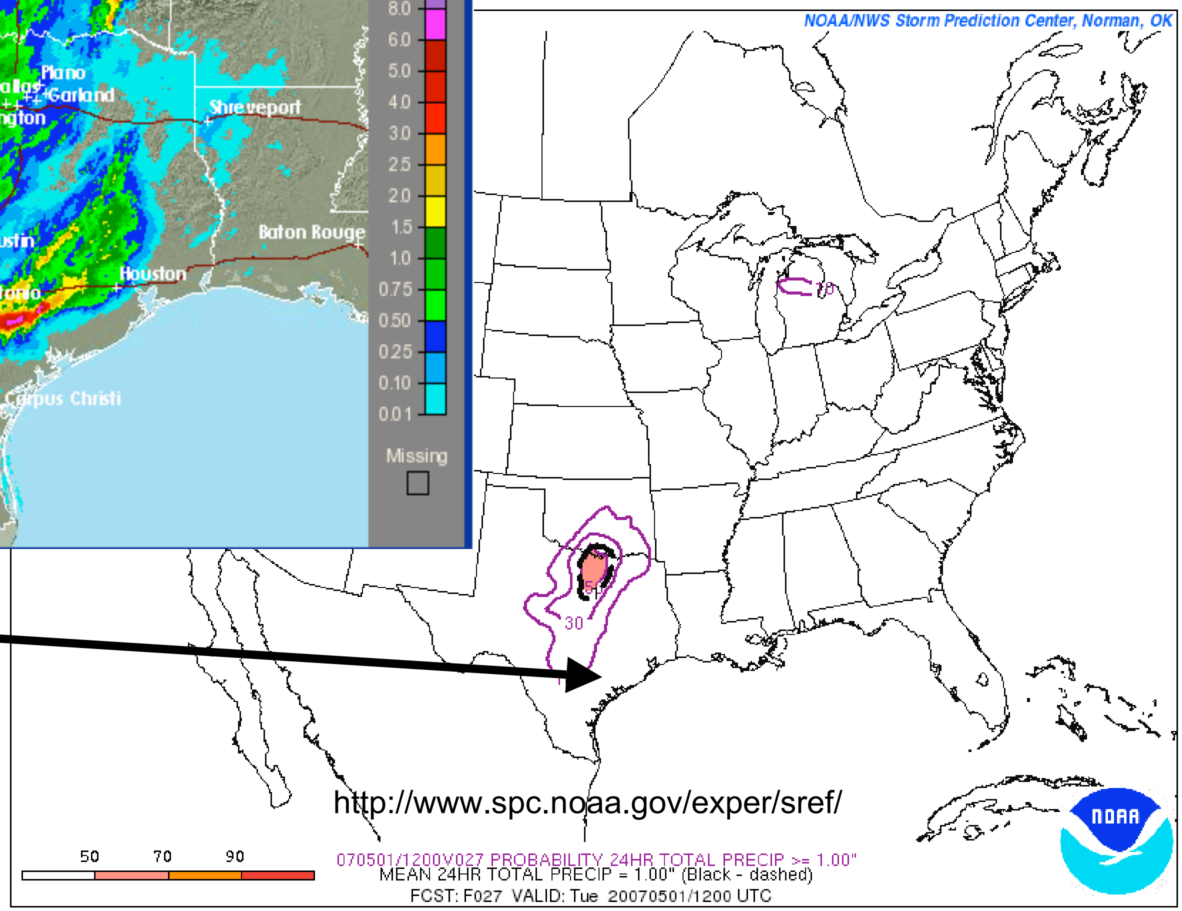
Probabilities commonly estimated from frequency of event in the ensemble.

from Tim Palmer's book chapter, 2006.

Problem with current ensemble forecast systems

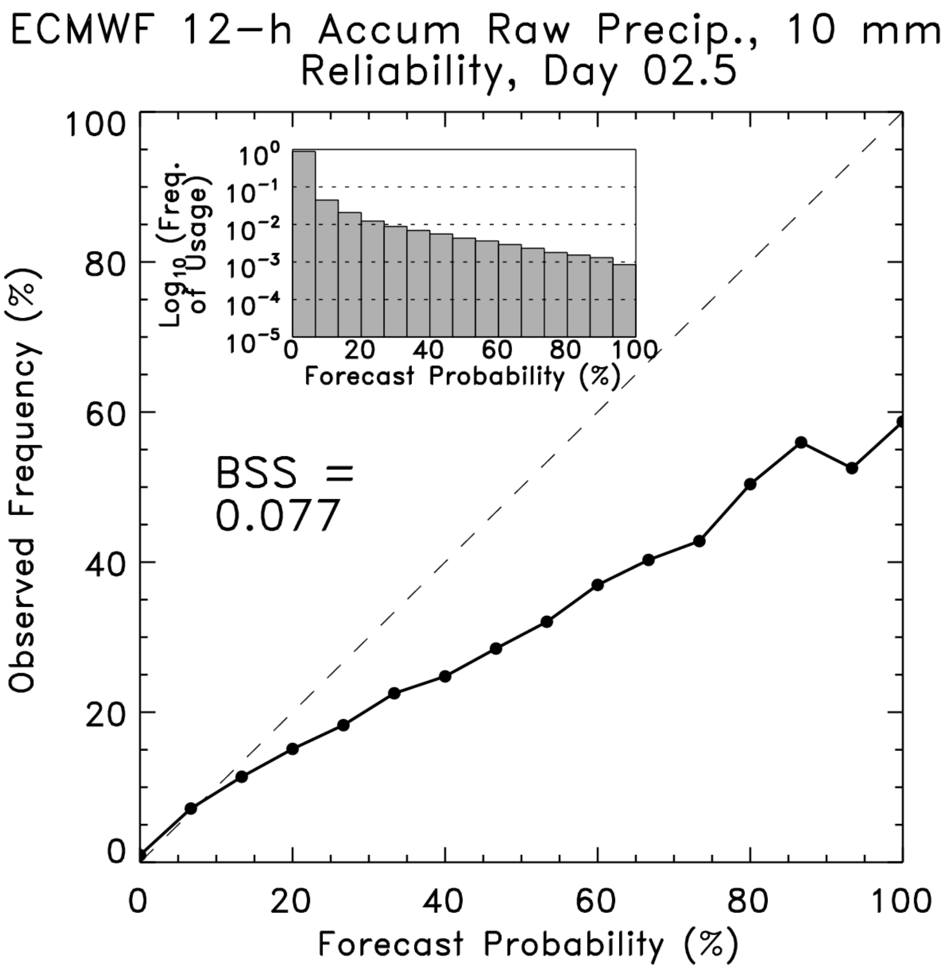
Forecasts may be biased and/or deficient in spread; probabilities are mis-estimated.





Heavy rain in an area where none of the ensemble members predicted it.

⇒ Unreliable forecasts



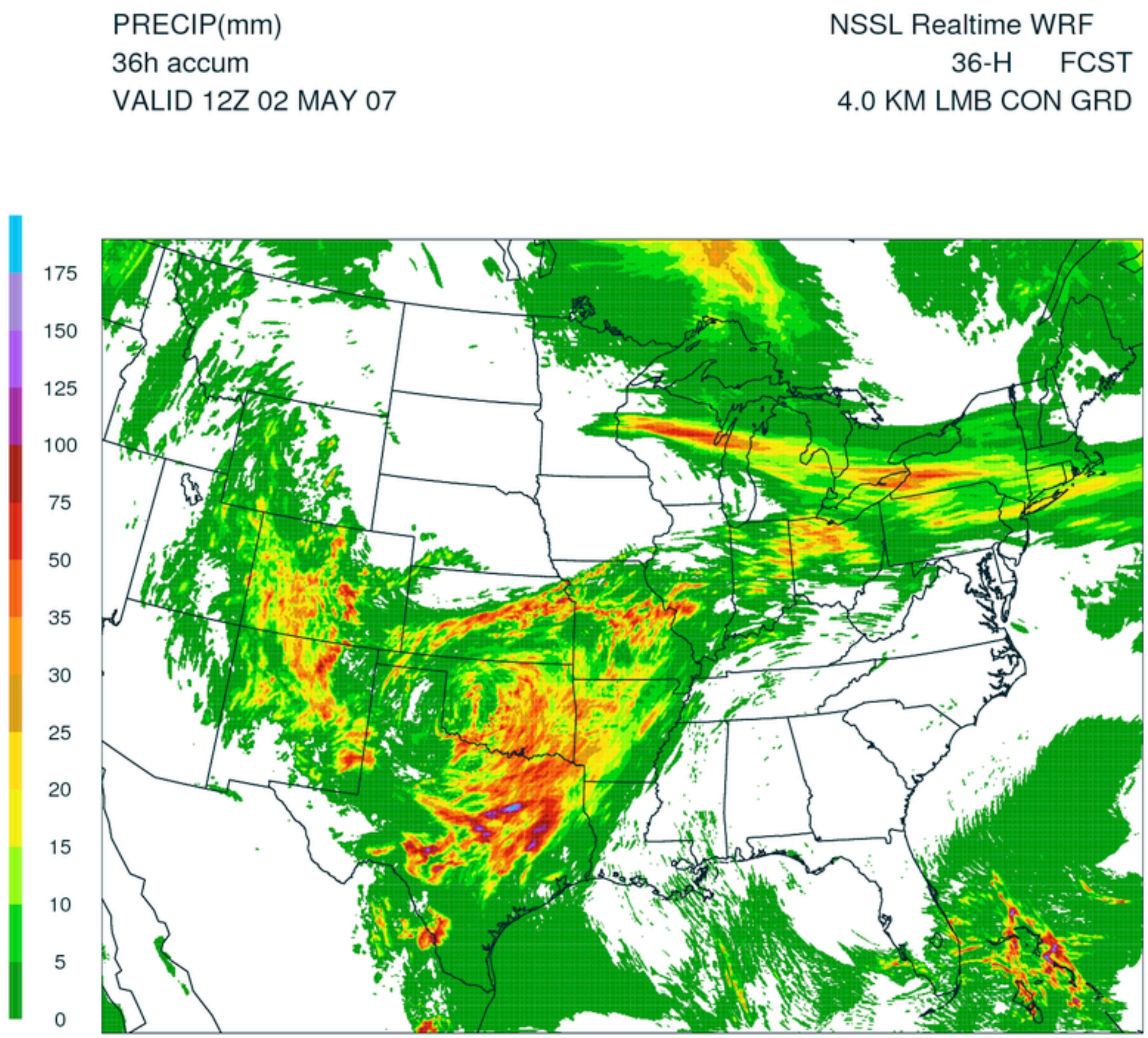
Probabilistic forecasts from raw ensembles are not very reliable, due to deficiencies in forecast model, ensemble methods.

Users want “sharp” and “reliable” forecasts.

NRC vision:
NWS to make all products probabilistic. How?

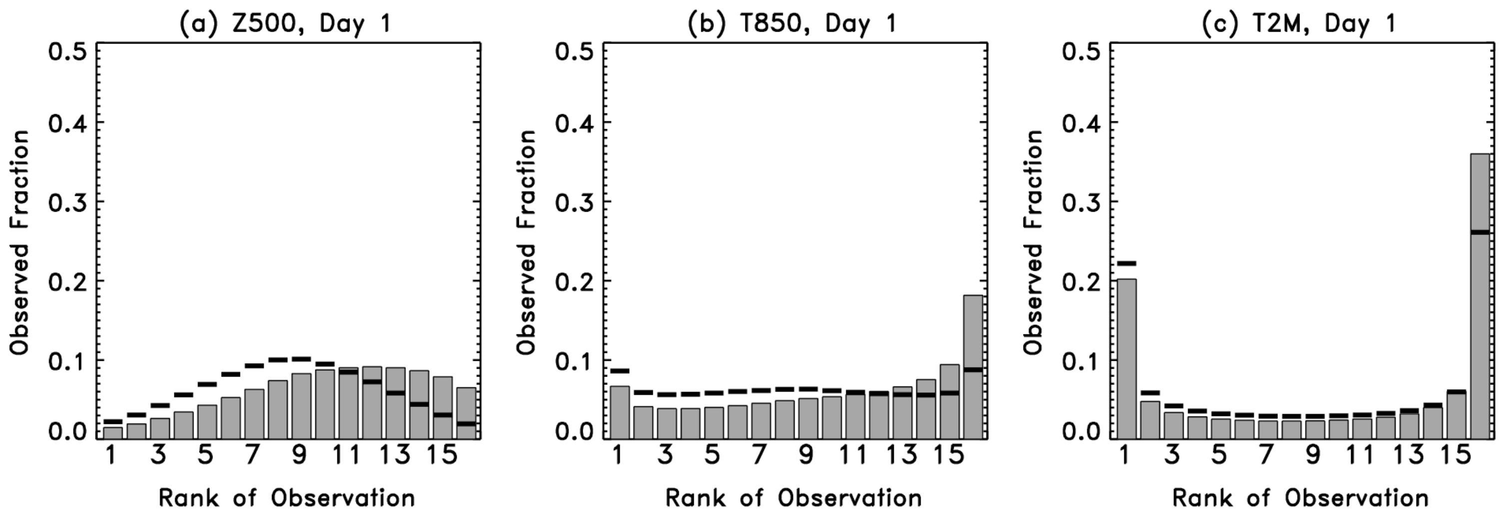
- **General option 1**: Work harder at current vision of developing hi-resolution models and ensembles. Probabilistic products based on these, perhaps QC'ed by humans.

Models ARE improving dramatically, and with them, ensemble forecasts.



We now have models with explicit convection that produce forecasts that look, for the first time, like radar images of precipitation.

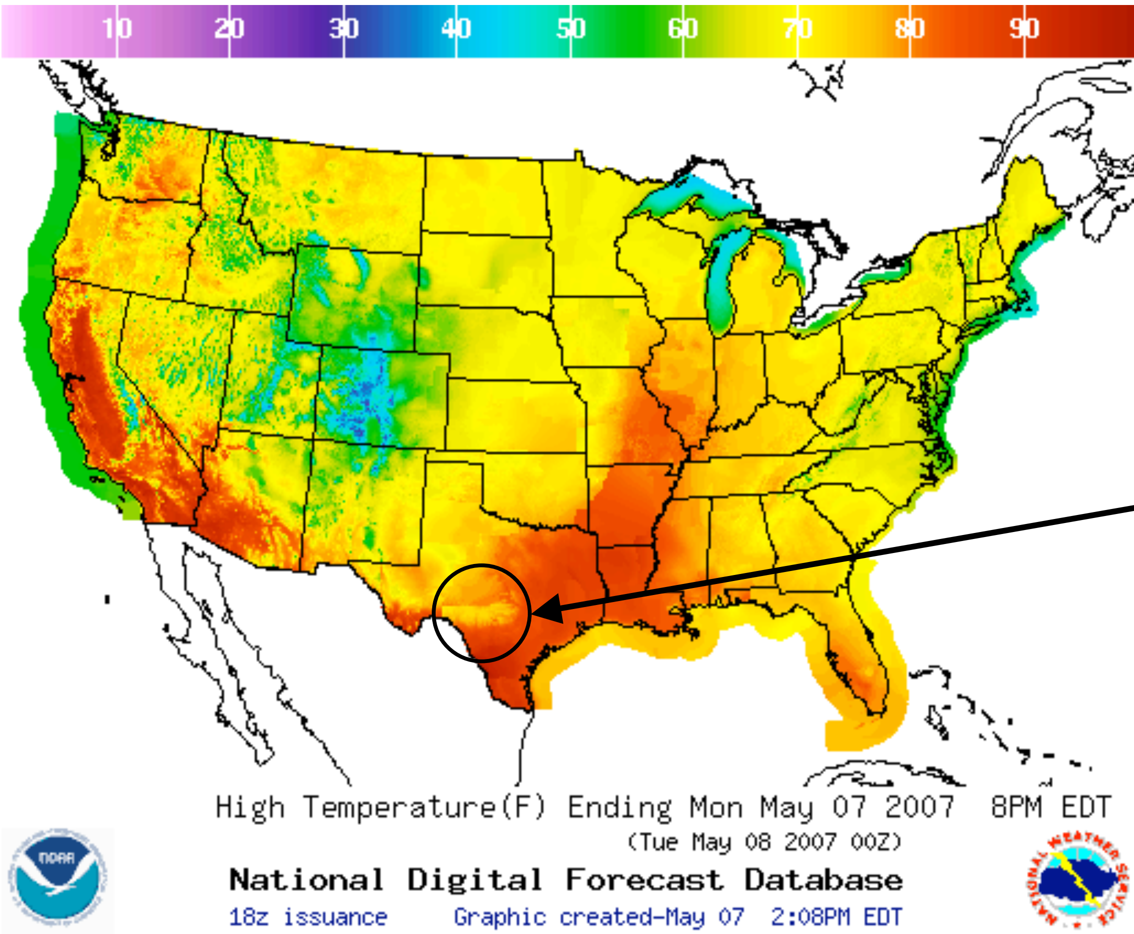
Still, a formidable list of ensemble deficiencies …



Ensembles are least useful for the variables we care the most about, such as surface temperature and precipitation.

Much work still needed in: (1) methods for developing sets of initial conditions; (2) ways of sampling the uncertainty due to forecast model deficiencies; (3) development of better, higher-resolution NWP models.

Manually QC new probabilistic products? Tough task.



Different WFOs have different ideas about what corrections to make, leading to discontinuities. Expect probabilistic QC even tougher.

In most circumstances, it may be a better use of forecasters' time to focus on the shorter-range, more severe-weather problems.

NRC vision: NWS to make all products probabilistic. How?

- **General option 2**: Automated probabilistic forecast products, or "Ensemble MOS"; ensemble forecast system + computer-based statistical post-processing using "*reforecasts*" (past forecasts from same system used operationally).

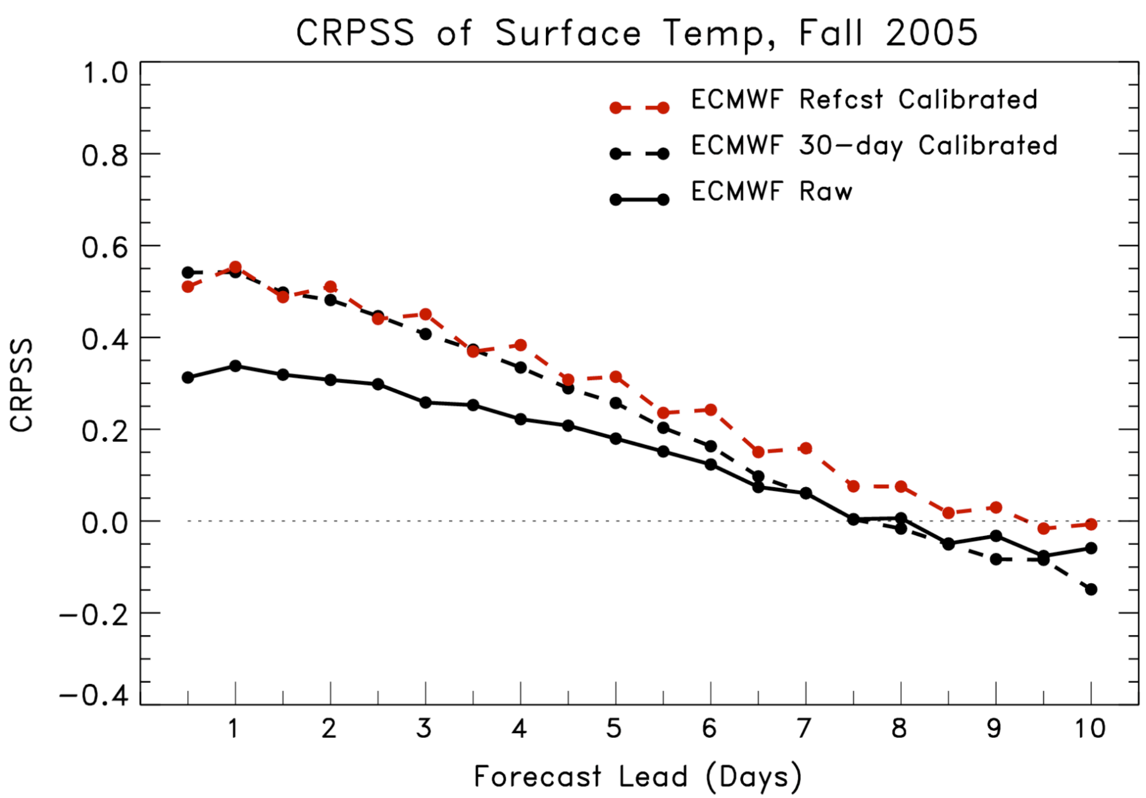
NOAA’s reforecast data set

- **Model**: T62L28 NCEP GFS, circa 1998
- **Initial Conditions**: NCEP-NCAR Reanalysis II plus 7 +/- bred modes.
- **Duration**: 15 days runs every day at 00Z from 19781101 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*).
- **Real-time** probabilistic precipitation forecasts: http://www.cdc.noaa.gov/reforecast/narr

ECMWF's reforecast data set

- **Model**: 2005 version of ECMWF model; T255 resolution.
- **Initial Conditions**: 15 members, ERA-40 analysis + singular vectors
- **Dates of reforecasts**: 1982-2001, Once-weekly reforecasts from 01 Sep - 01 Dec, 14 total. So, 20*14 ensemble reforecasts = 280 samples.
- **Data** sent to NOAA / ESRL : T_{2M}, precip. ensemble over most of North America, excluding Alaska. Saved on 1-degree lat / lon grid. Forecasts to 10 days lead.

Good news: for some variables a few prior forecasts are adequate to calibrate.



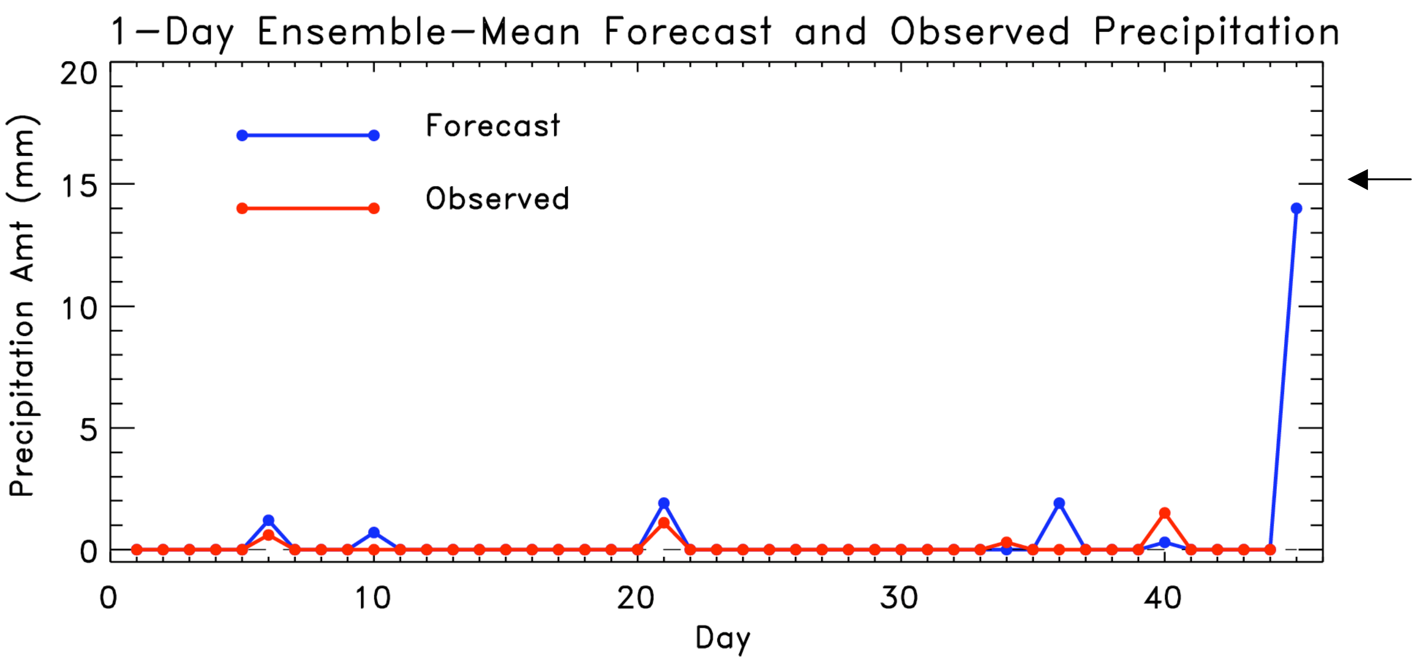
Statistically correcting a probabilistic surface temp. forecast with 30-day training data set and 20-year, once-weekly training data set.

At short leads, there is no advantage to a long training data set. At longer leads, there is an advantage.

Short-lead temperature bias in this 2005 version of ECMWF model is relatively consistent from one day to the next.

Tougher news: for other problems such as calibrating heavy precipitation, larger training data sets are necessary.

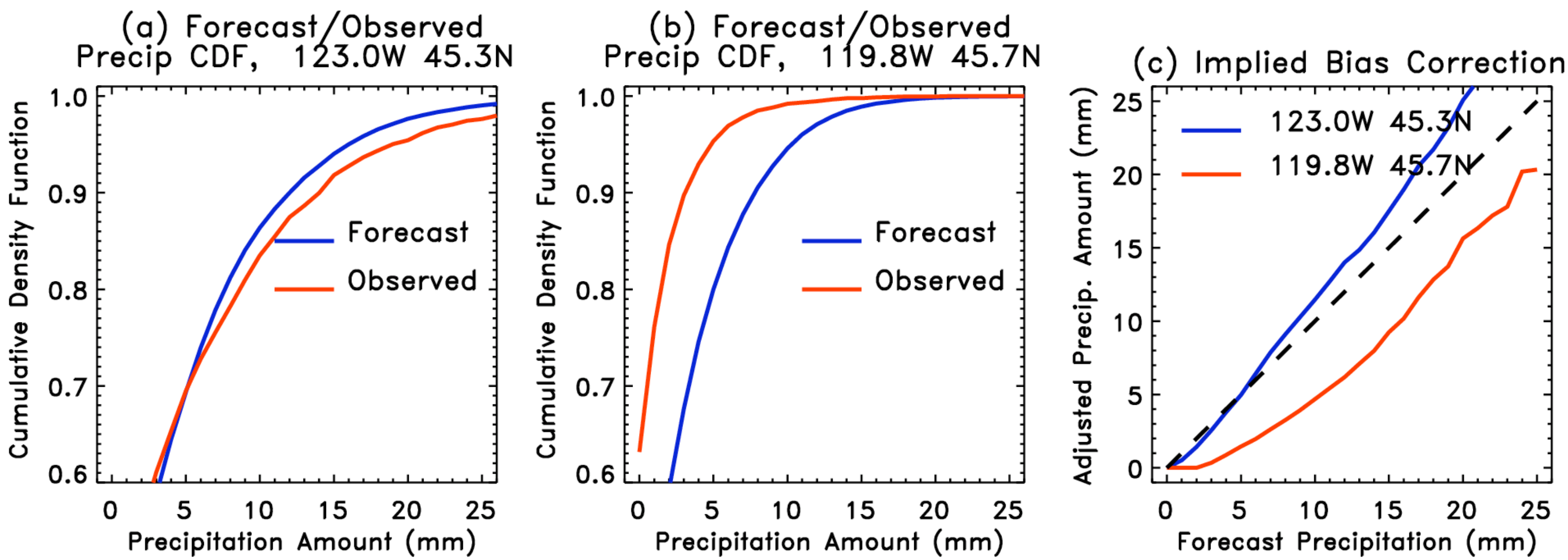
Consider training with a short sample in a climatologically dry region. How could you calibrate this latest forecast?



You'd like enough training data to have some similar events at a similar time of year to this one.

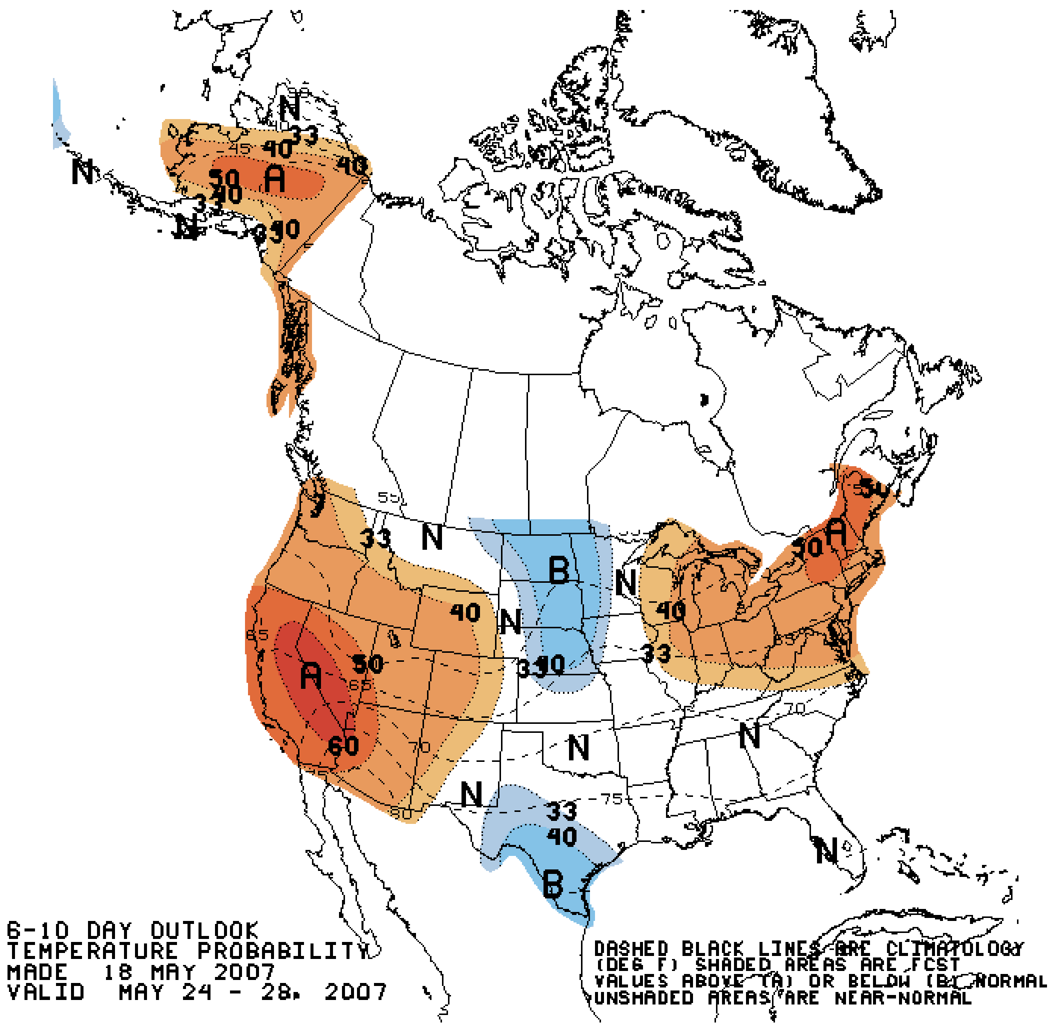
Boost sample size in statistical calibration by compositing statistics over different locations?

A good idea, if done with care.
However, even nearby grid points may have different forecast errors.



Panels (a) and (b) provide the cumulative density function (CDF) of 1-day forecasts of precipitation for 1 January (CDFs determined from reforecast data and observations in Dec-Jan). Panel (a) is for a location near Portland, Oregon, and panel (b) is in north central Oregon, east of the Cascades. Panel (c) provides the implied function for a bias correction from the forecast amount to a presumed observed amount. Note the very different corrections implied at two nearby locations.

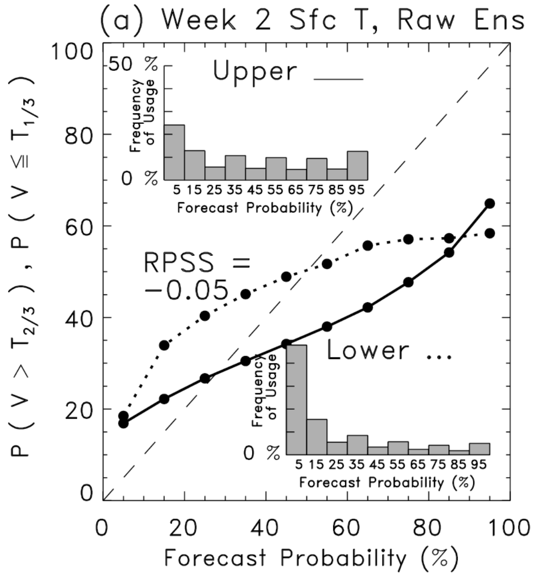
Calibrating Week 2 and 6-10 day probability forecasts

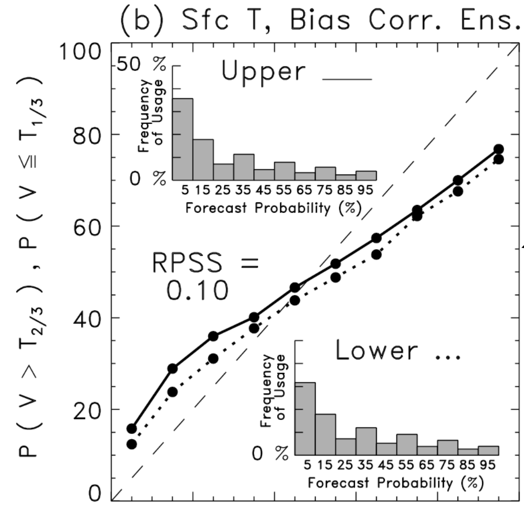


An example of the operational 6-10 day temperature forecast produced by NCEP/CPC.

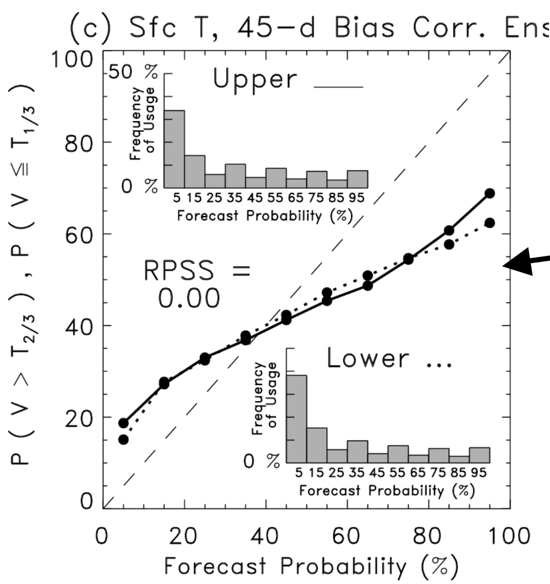
Week-2 Temperature Forecasts

Probabilities from raw ensemble



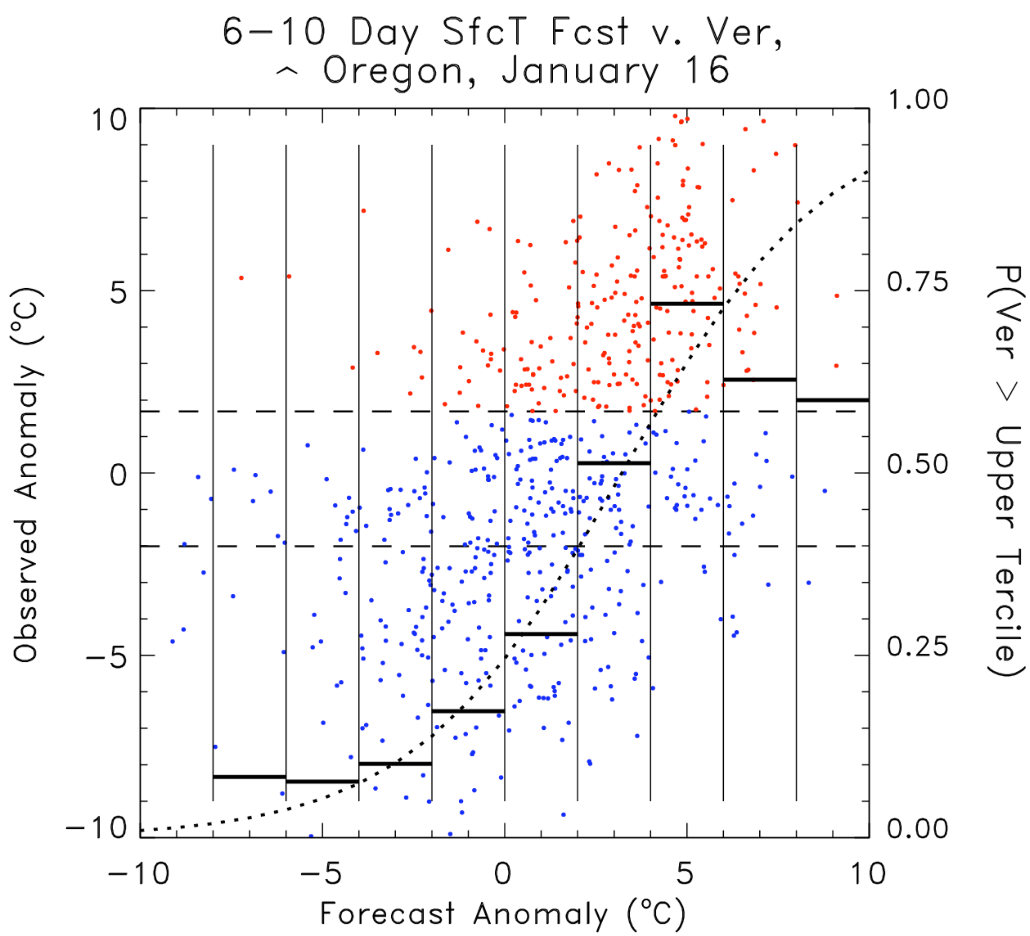


Correction of biases estimated from full 22 years of forecast data



Correction of biases estimated from last 45 days of data

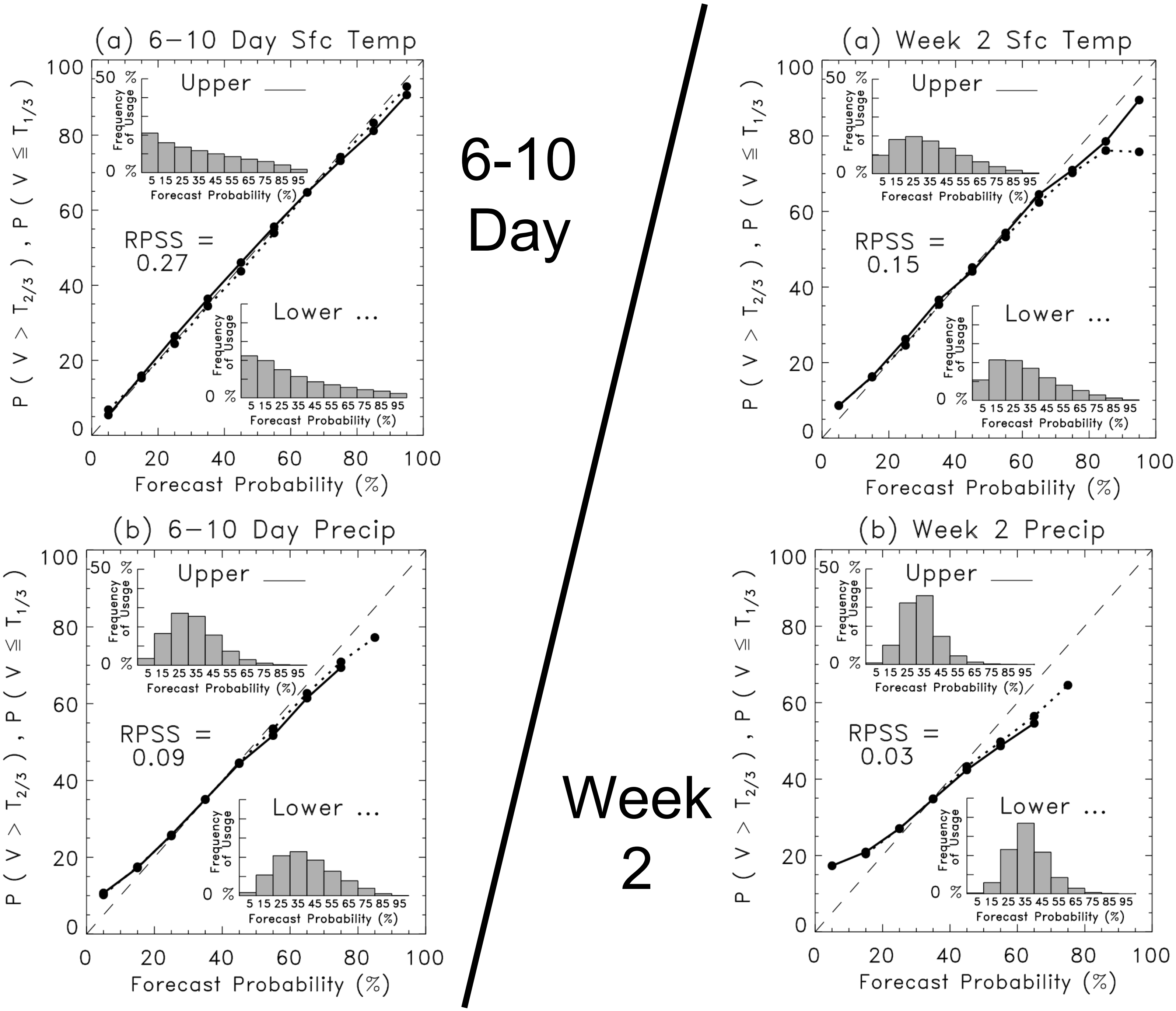
Calibration using a long data set of observed and forecast anomalies



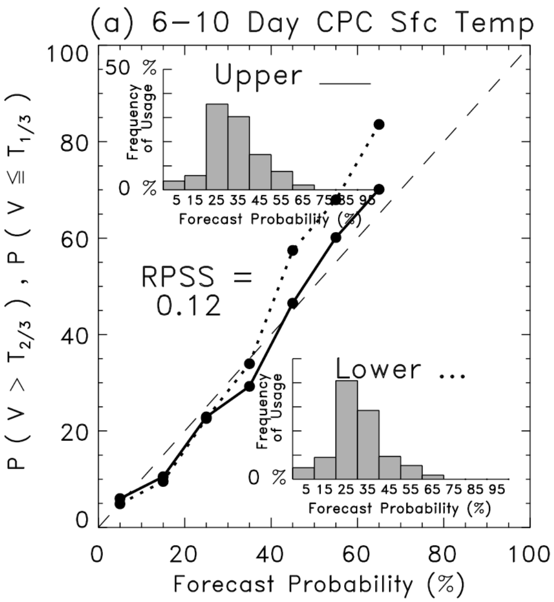
With our reforecasts, we have 20+ years of data. Let's use old data in a 31-day window around the date of interest to make statistical corrections.

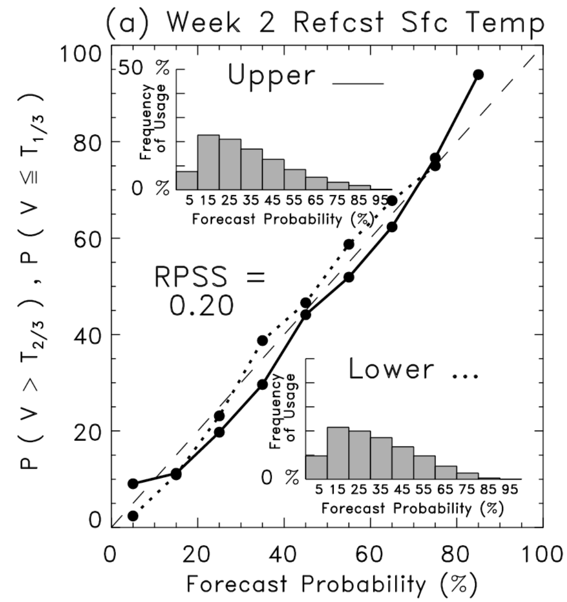
Dashed lines: tercile boundaries
Red points: samples above upper tercile
Blue points: samples below upper tercile
Solid bars: probabilities by bin count
Dotted line: a fitted logistic regression

$$P = 1 - \frac{1}{1 + exp\left(\beta_0 + \beta_1 x\right)}$$


6-10 Day
Week 2
(a) 6–10 Day Sfc Temp
RPSS = 0.27
(b) 6–10 Day Precip
RPSS = 0.09
(a) Week 2 Sfc Temp
RPSS = 0.15
(b) Week 2 Precip
RPSS = 0.03
Upper ——
Lower ...
Frequency of Usage
Forecast Probability (%)
P (V > T2/3) , P (V ≤ T1/3)

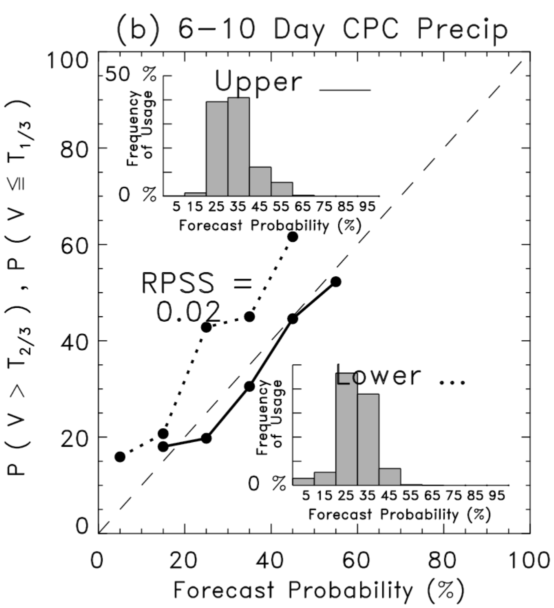
Comparison against NCEP / CPC forecasts at 155 stations, 100 days in winter 2001-2002

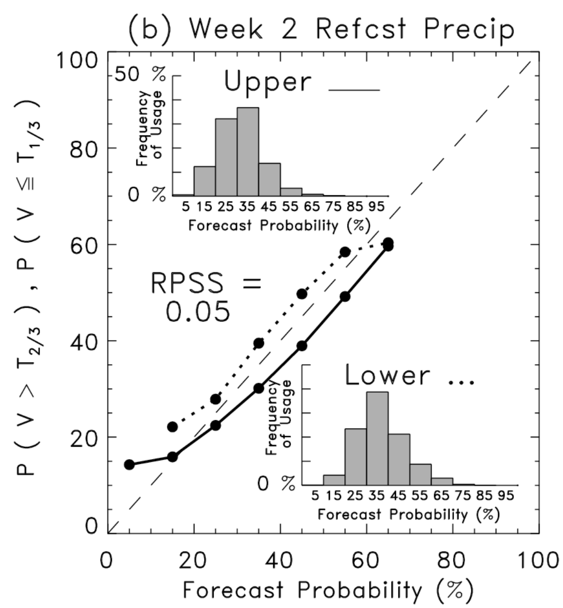




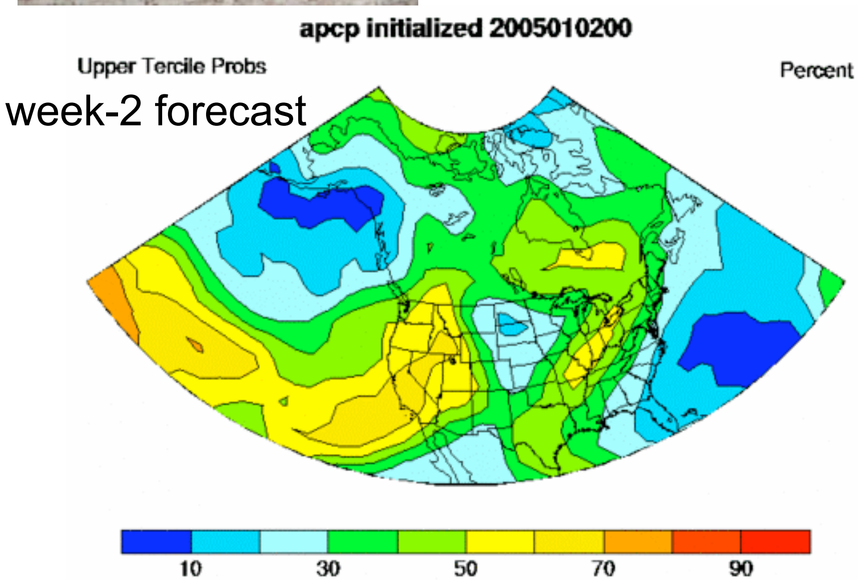
MOS-based **Week 2** forecasts using low-res T62 model more skillful than operational NCEP/CPC **6-10 day**.

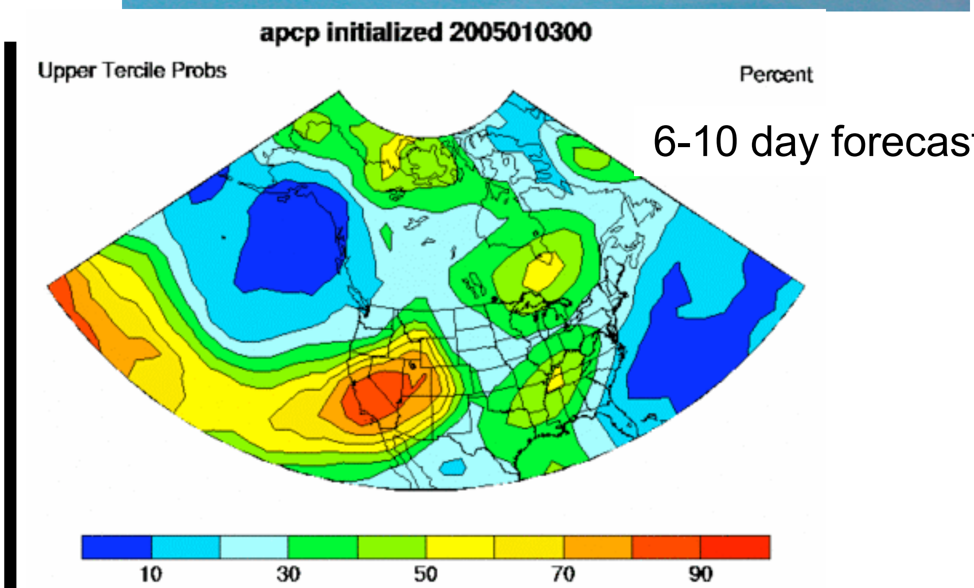
(NCEP now heavily utilizes reforecasts in these products)





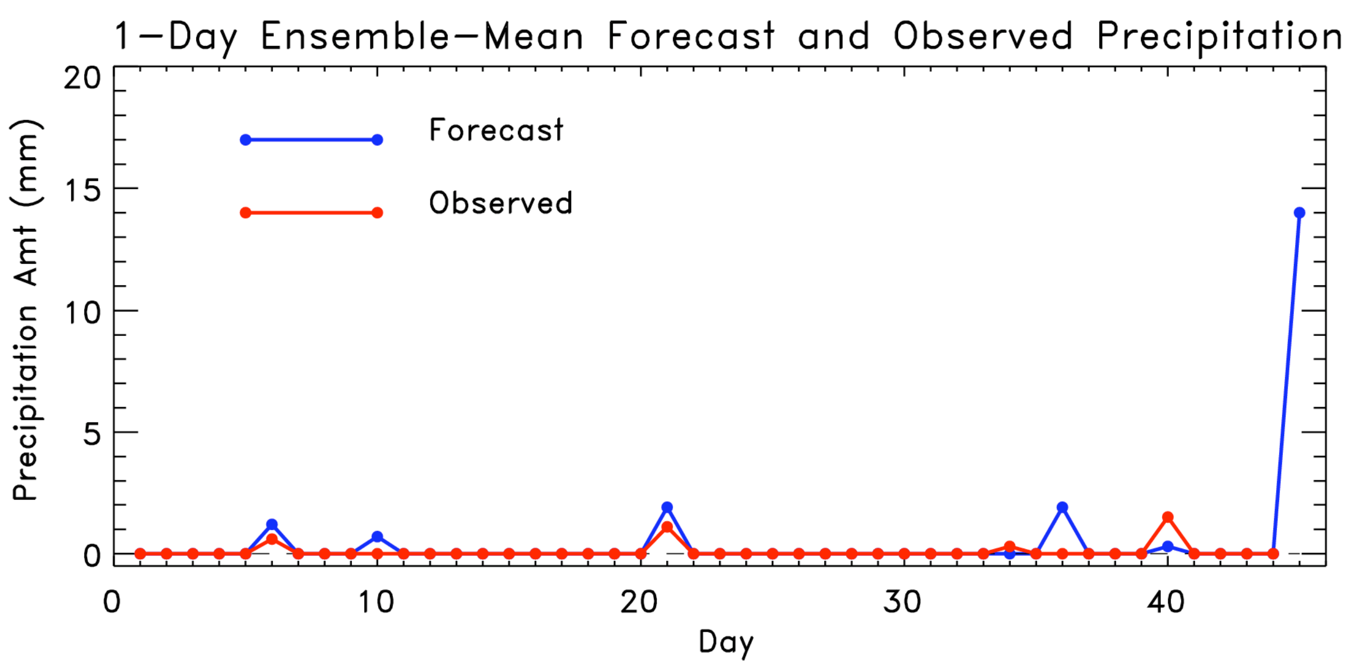
Example: floods causing La Chonchita, CA mudslide, 12 Jan 2005



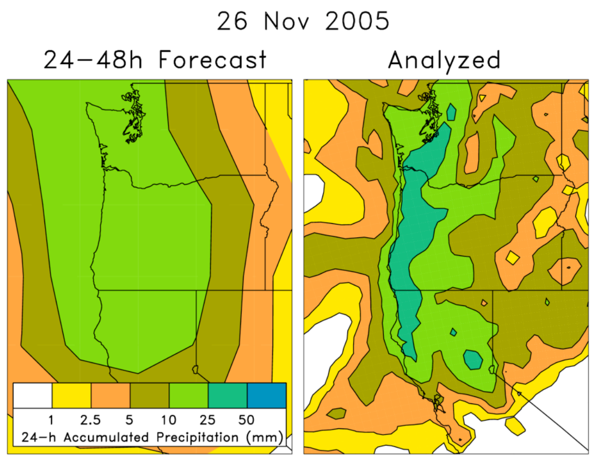


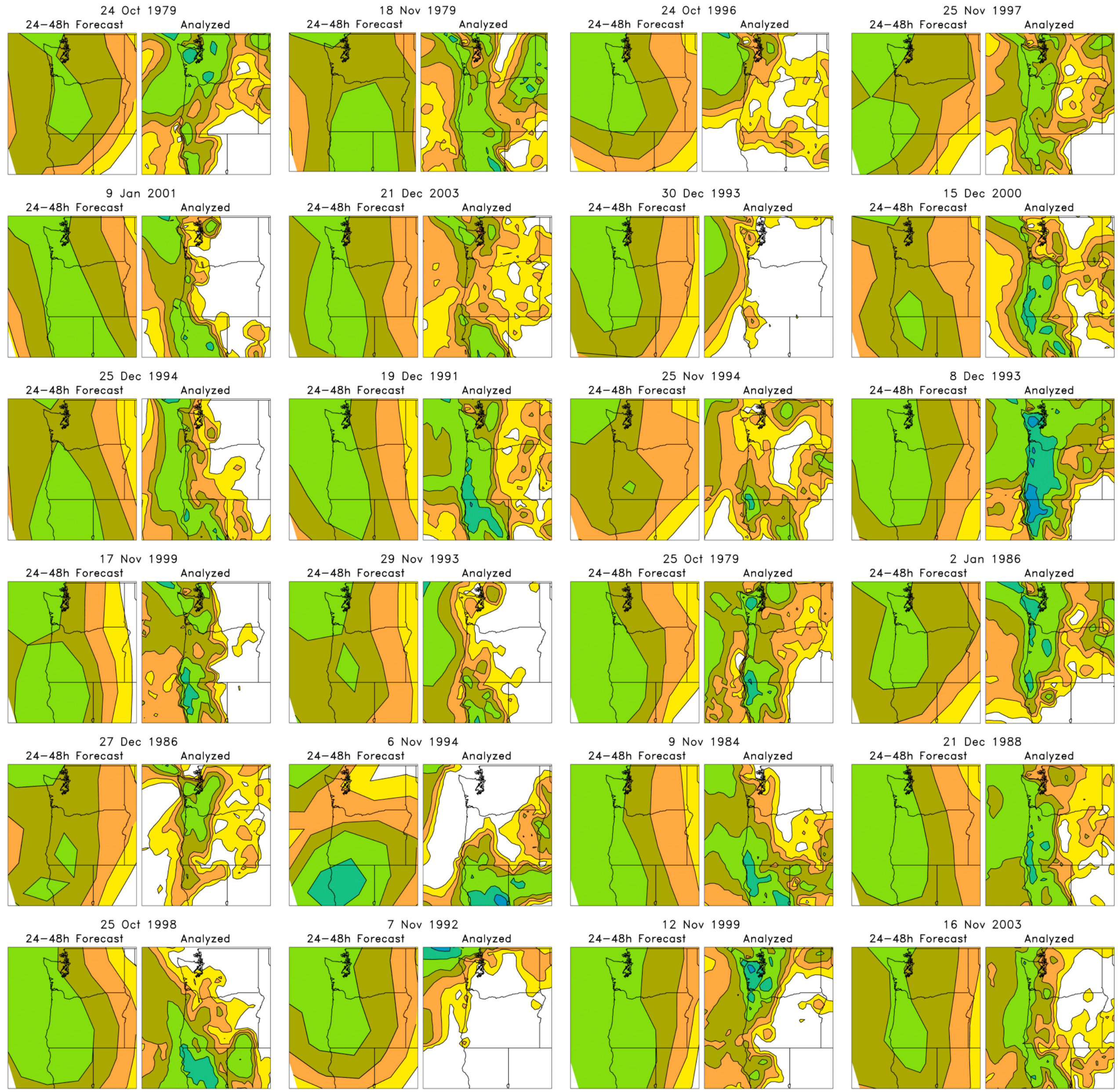
Calibration of PQPF & rare events

Want lots of old forecast cases that were similar to today's forecast. Then the difference between the observed and forecast on those days can be used to calibrate today's forecast.



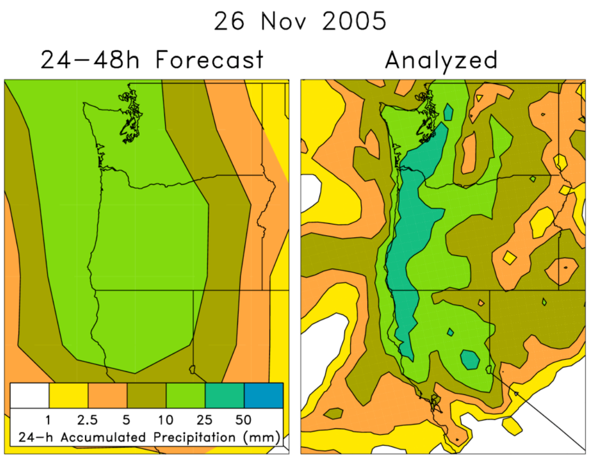
Producing a distribution of observed given forecast using analogs

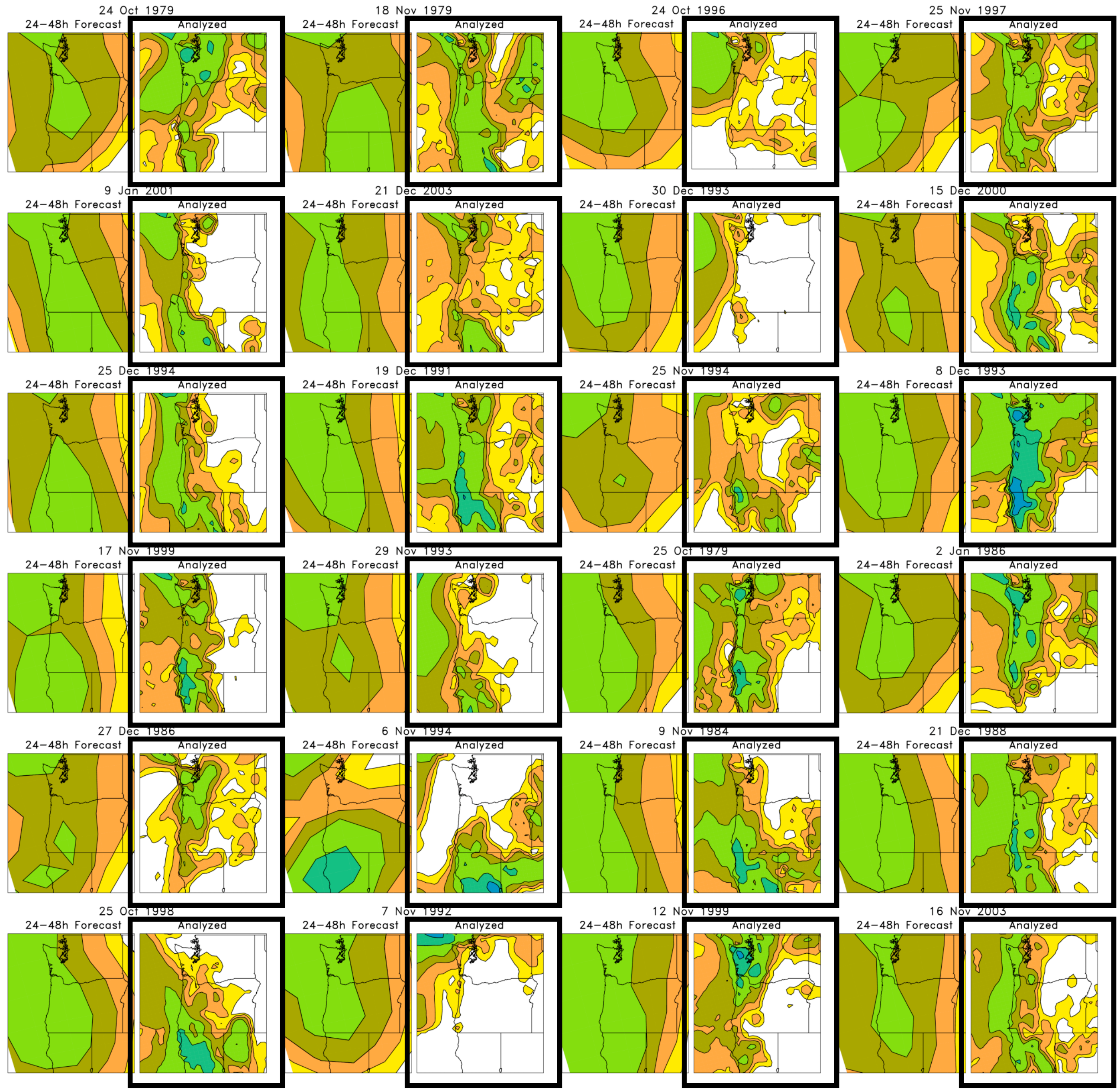




On the left are old forecasts similar to today's ensemble-mean forecast. For making probabilistic forecasts, form an ensemble from the accompanying analyzed weather on the right-hand side.

Producing a distribution of observed given forecast using analogs



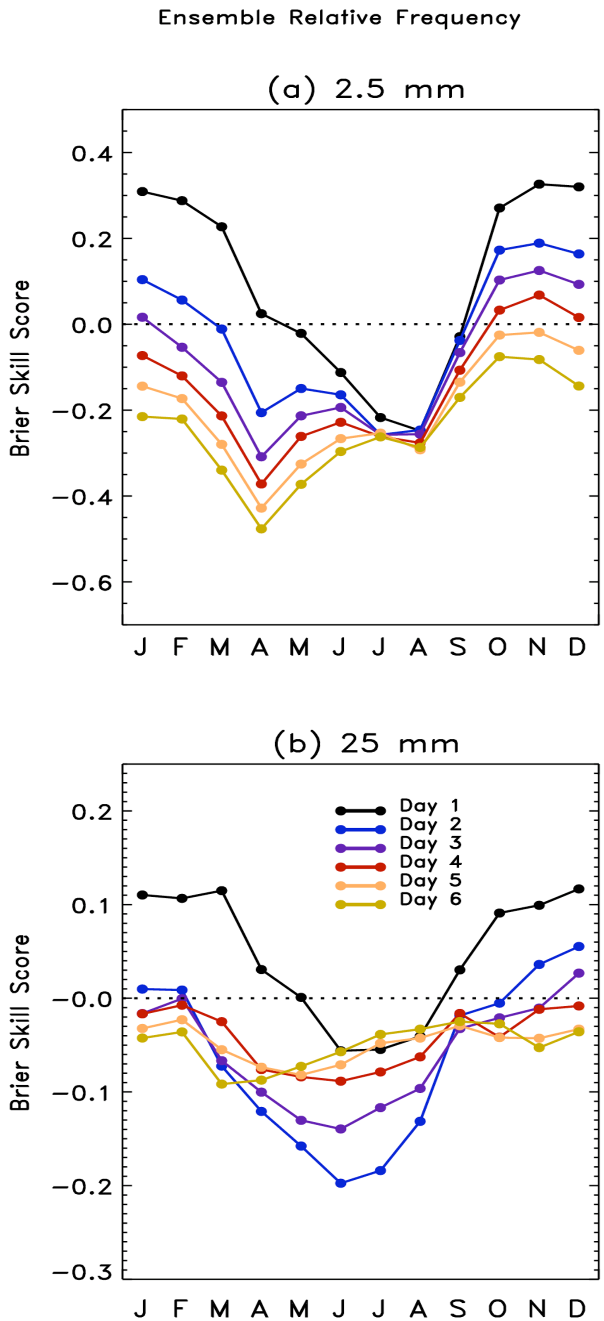


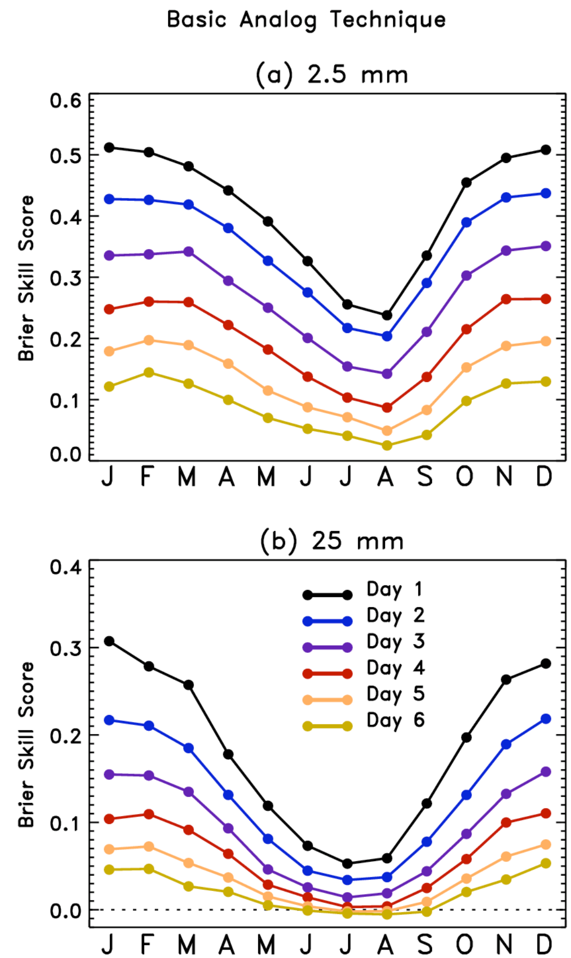
On the left are old forecasts similar to today's ensemble-mean forecast. For making probabilistic forecasts, form an ensemble from the accompanying analyzed weather on the right-hand side.

Asymptotic behavior of analog technique

- Q: What happens as correlation(F,O) $\rightarrow$ 0 ?
  A: Ensemble of observed analogs becomes random draw from climatology.

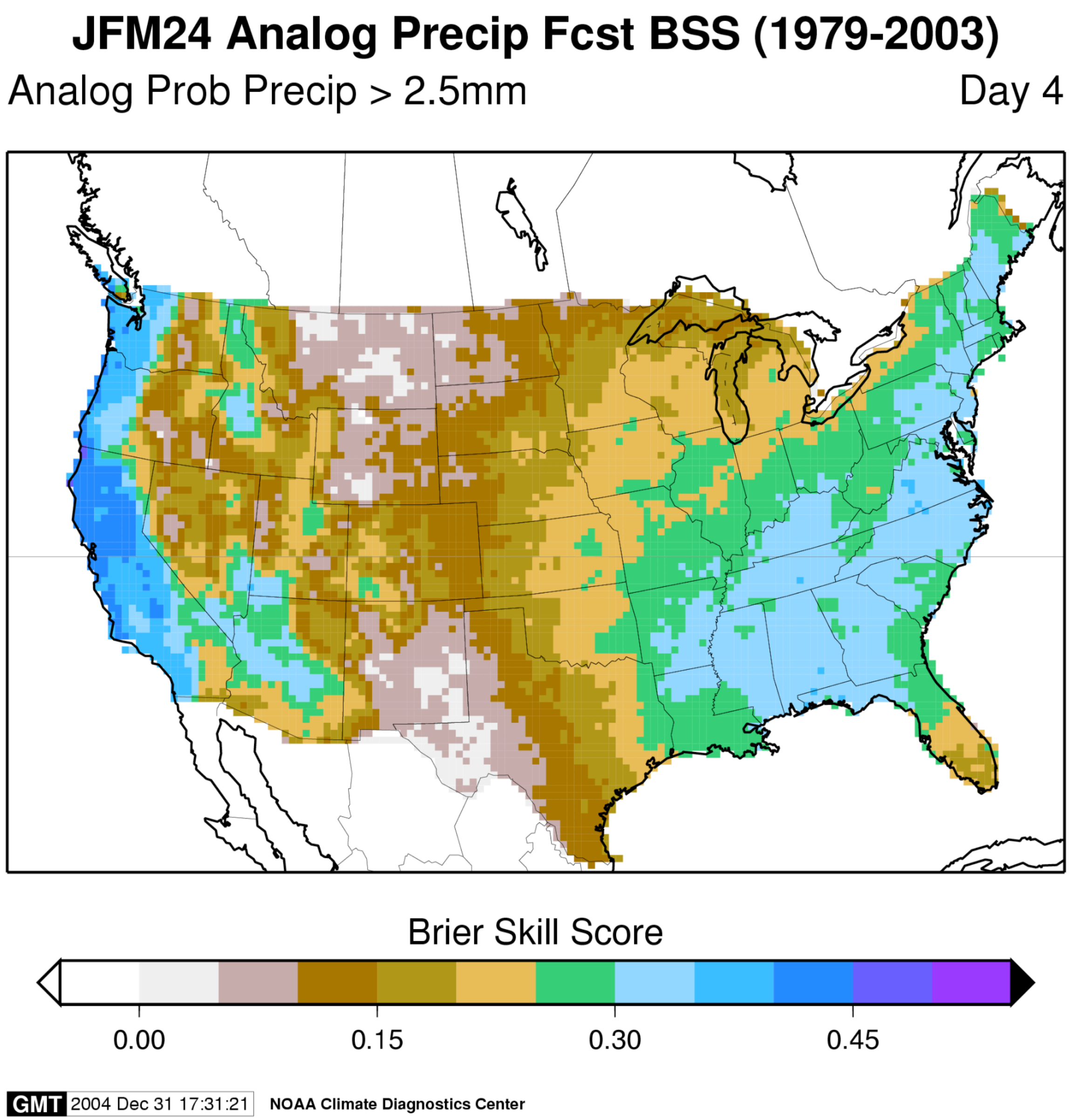
- Q: What happens as correlation(F,O) $\rightarrow$ 1 ?
  A: Ensemble of observed analogs looks just like today's forecast. Sharp, skillful forecasts



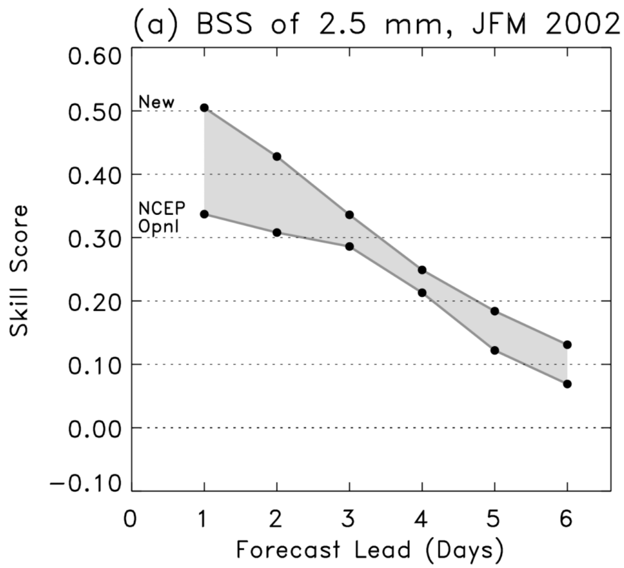


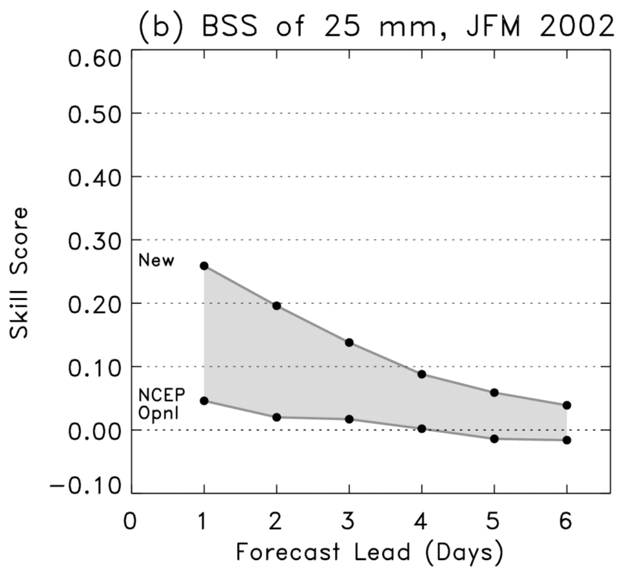
Verified over 25 years of forecasts; skill scores use conventional method of calculation which may overestimate skill
(Hamill and Juras 2006, QJRMS, Oct).

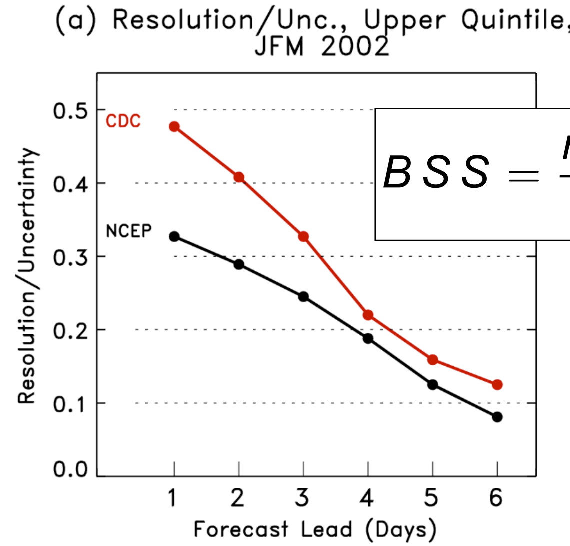
Skill as function of location



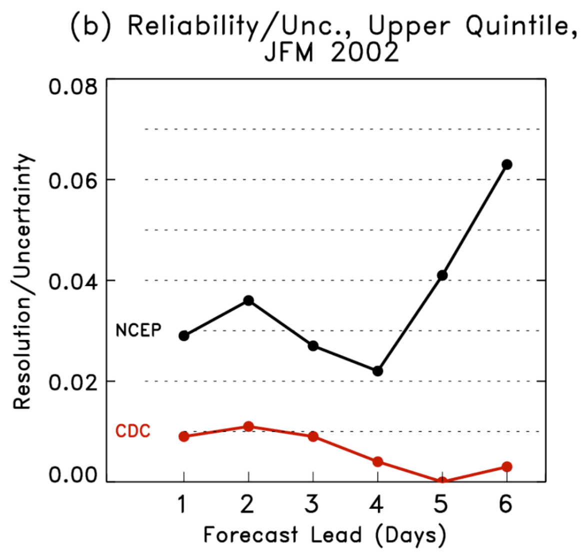
Comparison against NCEP medium-range T126 ensemble, ca. 2002



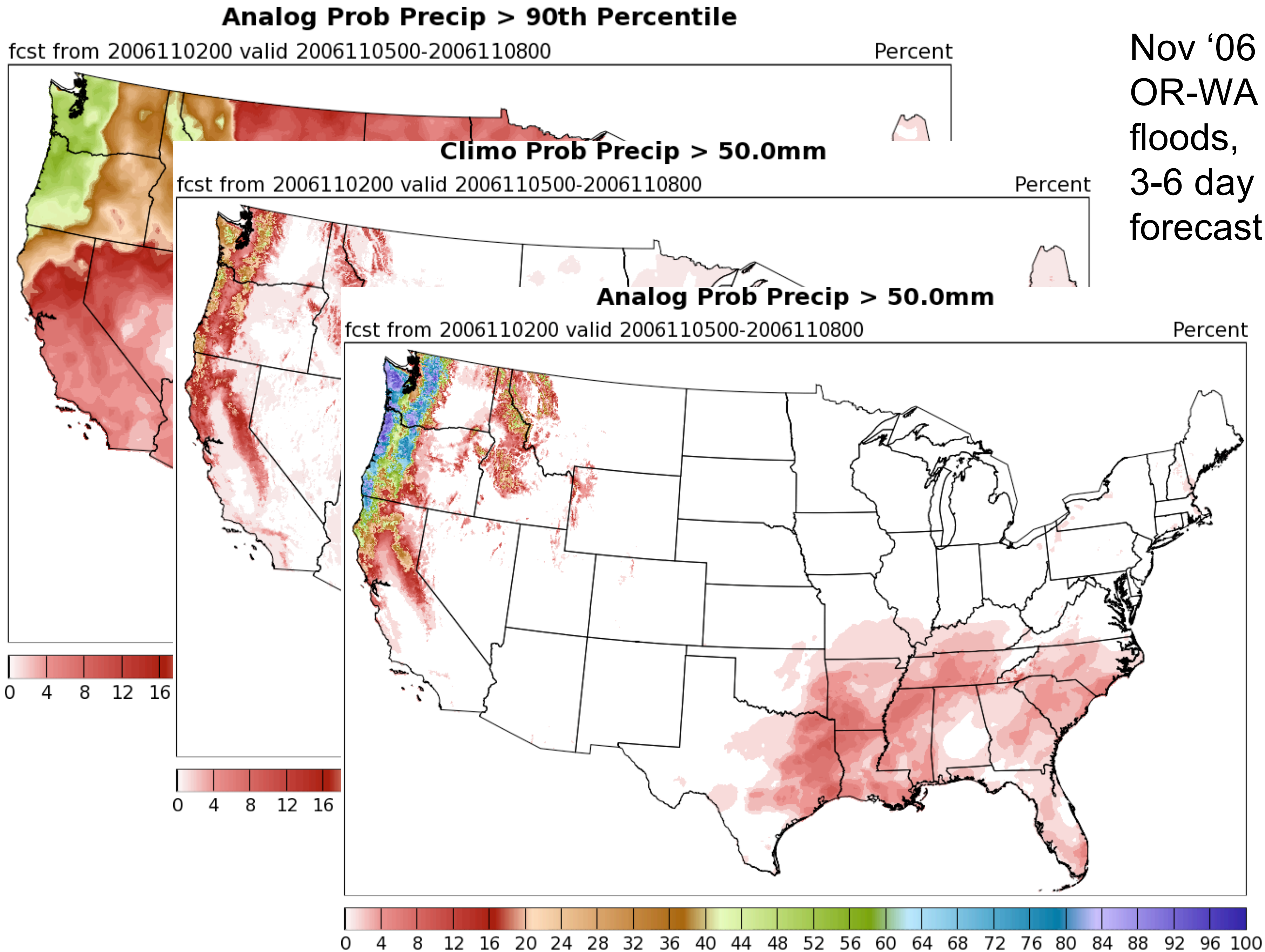




$$BSS = \frac{resolution - reliability}{uncertainty}$$

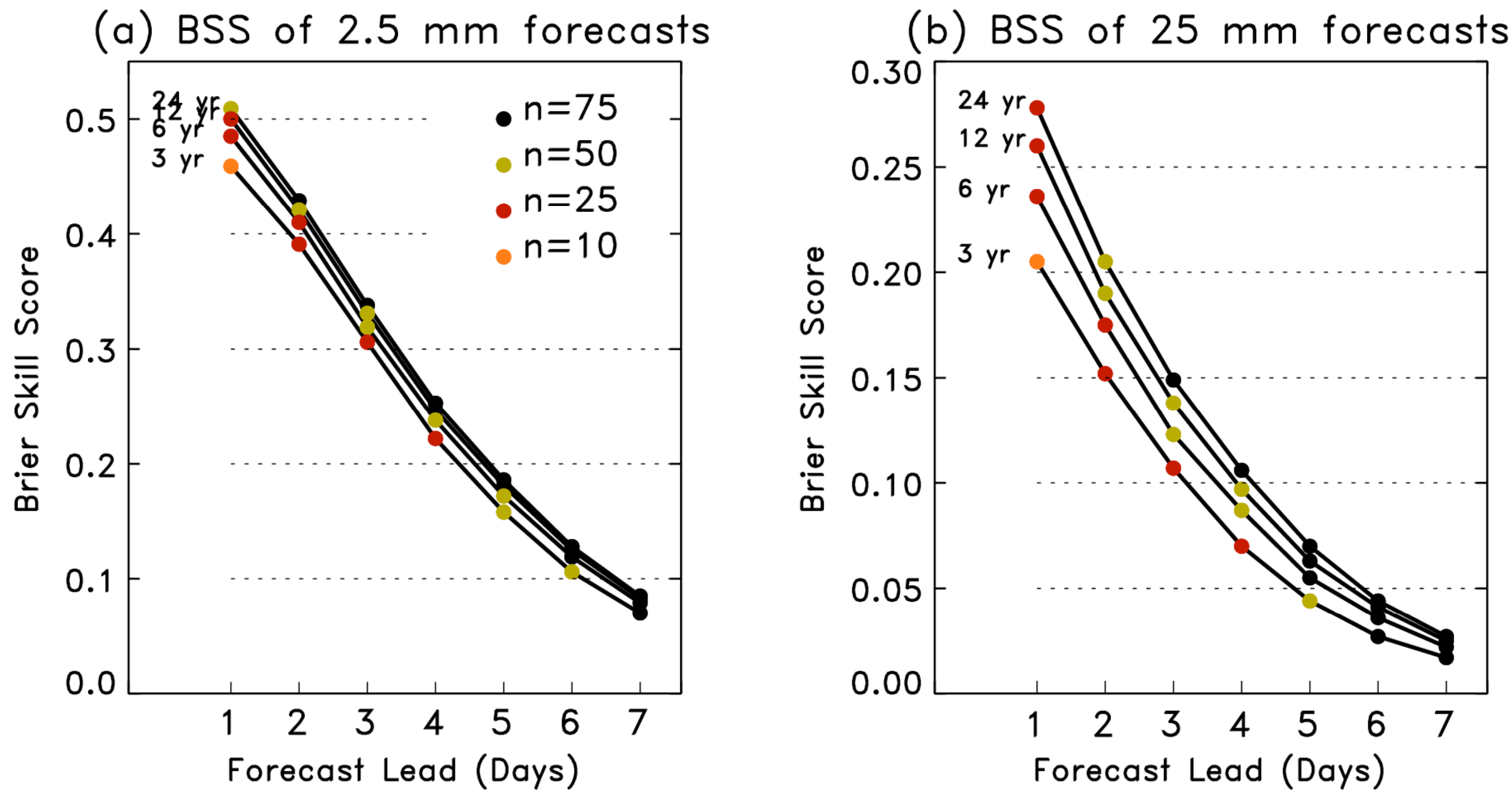


the improvement is a little bit of increased reliability, a lot of increased resolution.



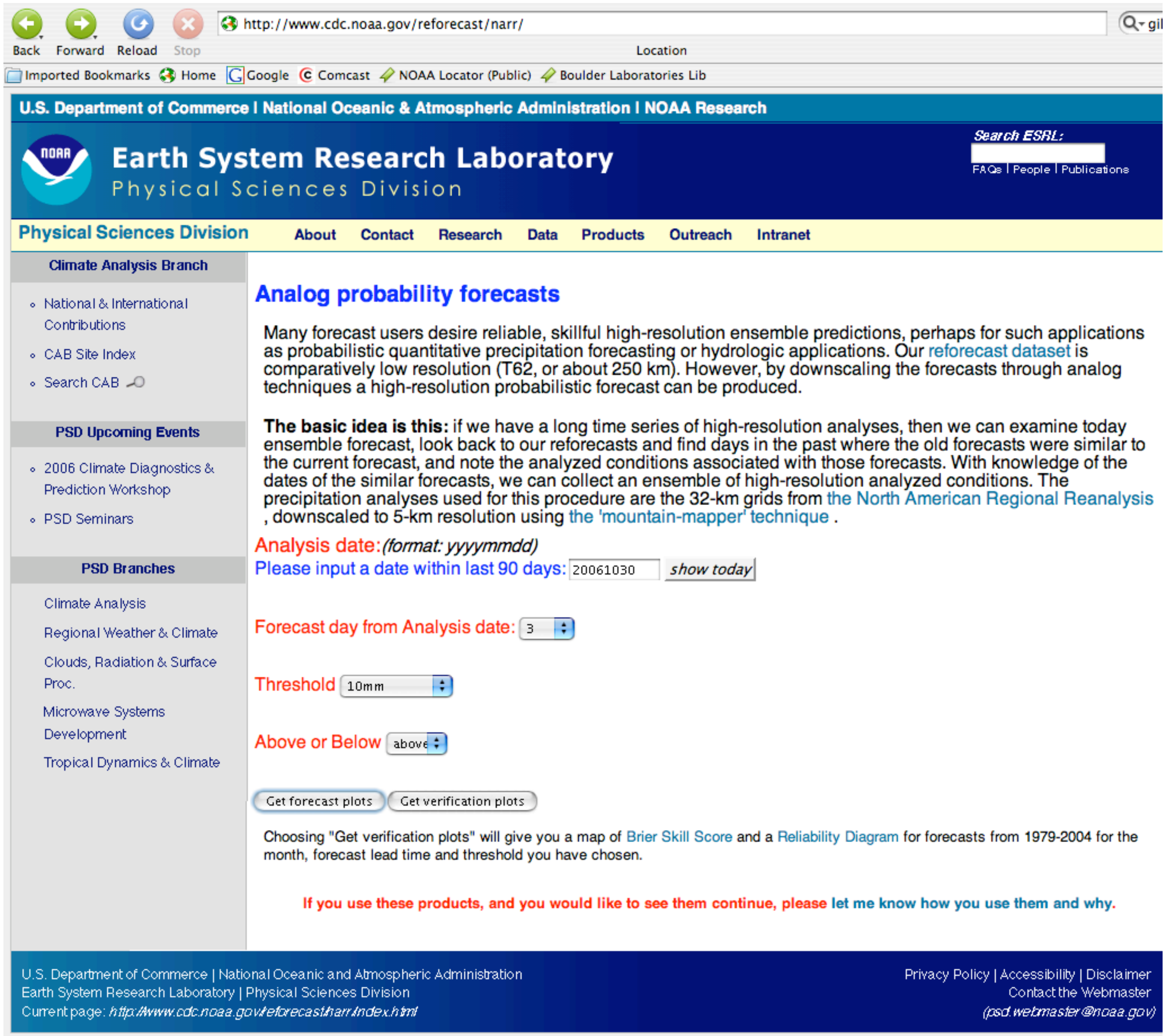
Nov '06 OR-WA floods, 3-6 day forecast

Effect of training sample size



colors of dots indicate which size analog ensemble provided the largest amount of skill.

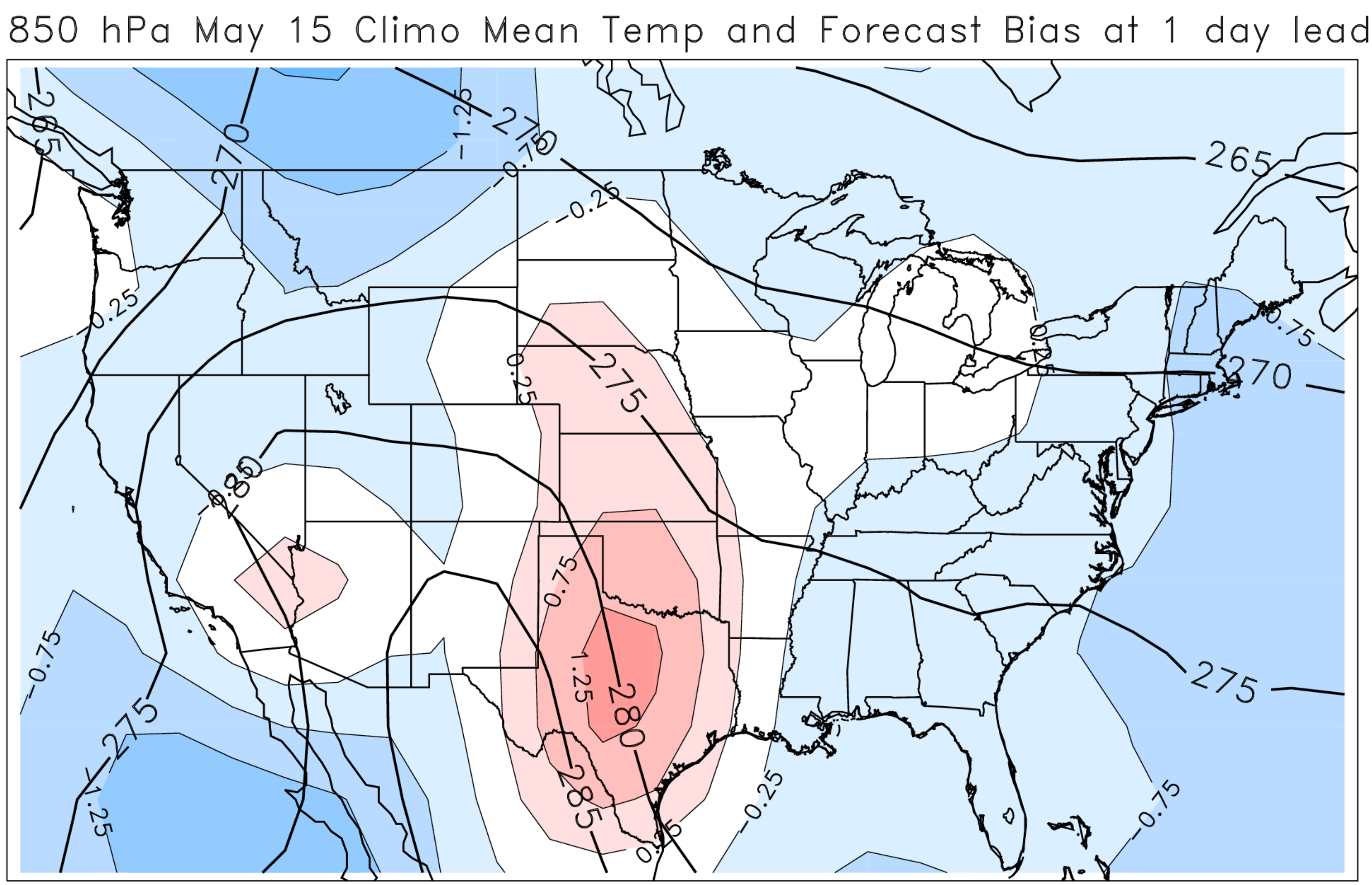
Real-time products



Probabilistic Calibrated Tornado Forecasts?

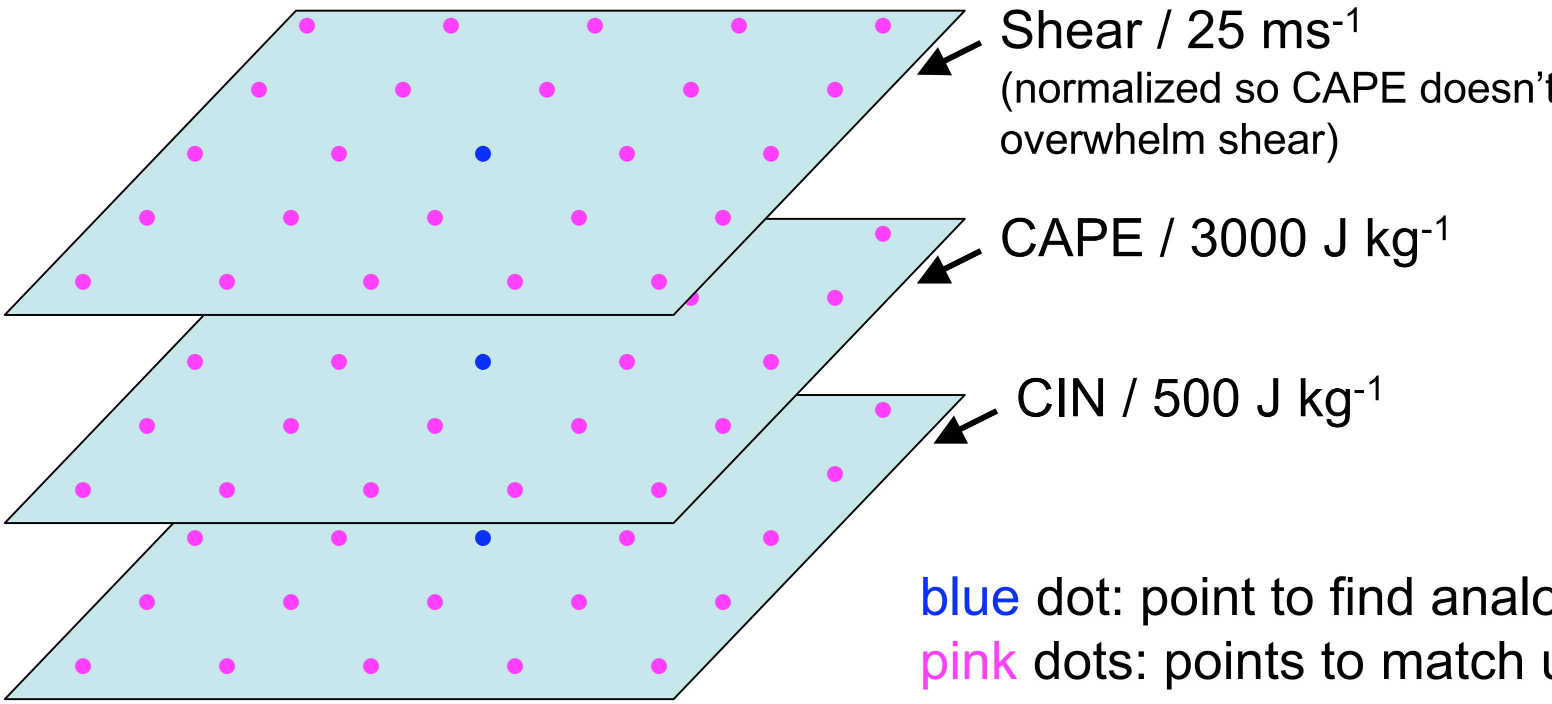
- CAPE, CIN, shear are known useful predictors of severe weather.
- Following the analog approach, can we:
  - (1) Examine today's forecast CAPE, shear
  - (2) Find old cases with similar forecast CAPE, shear
  - (3) Determine probabilities from frequency of tornado occurrence on dates with similar forecast?

Our assumption: regional forecast biases would harm our ability to find good analogs from location x for location y, i.e., to composite forecast data between locations



Technique for finding tornado forecast analogs

(1) For a given grid point, match today's scaled *ensemble mean* fields with past scaled forecast fields. (2) Find dates of *n* closest analog. (3) Determine tornado frequency from percentage of *n* dates with tornadoes.



blue dot: point to find analogs for

pink dots: points to match up current forecasts with old ones.

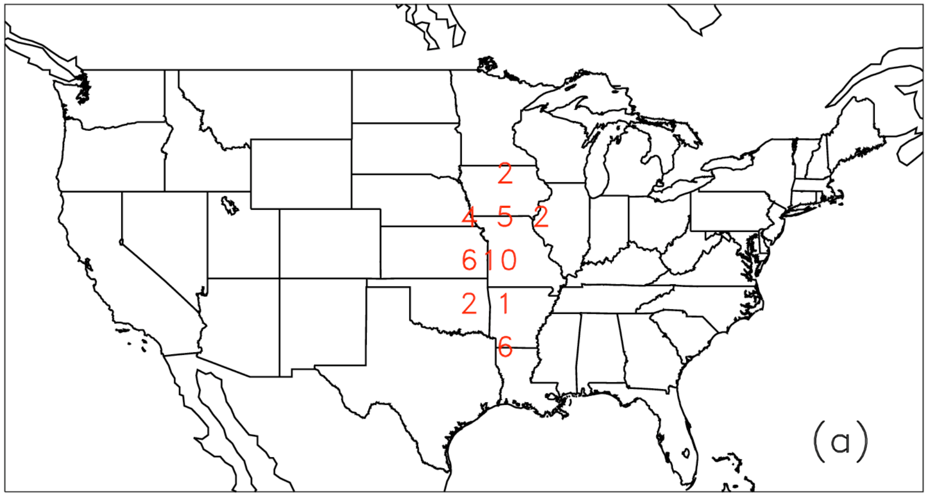
Note: can vary weights horizontally

Algorithm outputs:

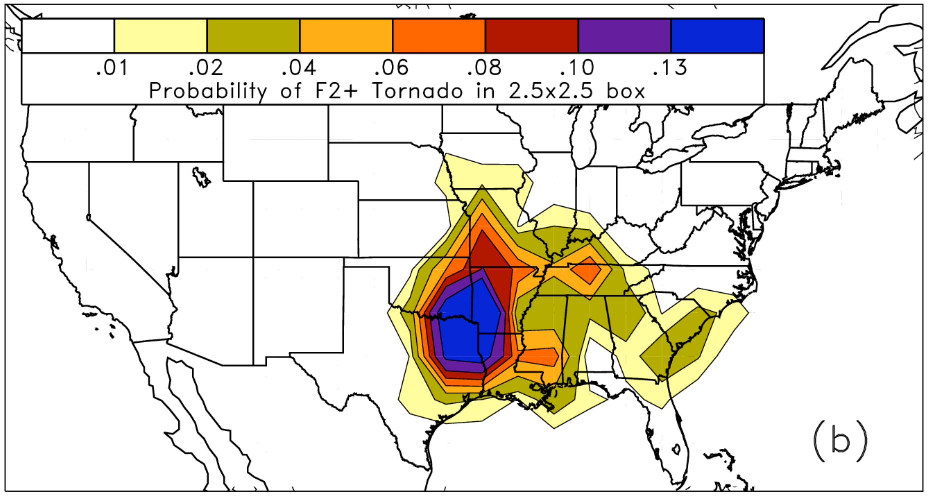
1) Dates of *n* analogs
2) Numerical quantification of how good the pattern match is for each of *n*.

First look? Day-1 forecast not bad.

Observed F2+ Tornado Counts in 12–hour Window
Centered on 0000 UTC 27 Apr 1991

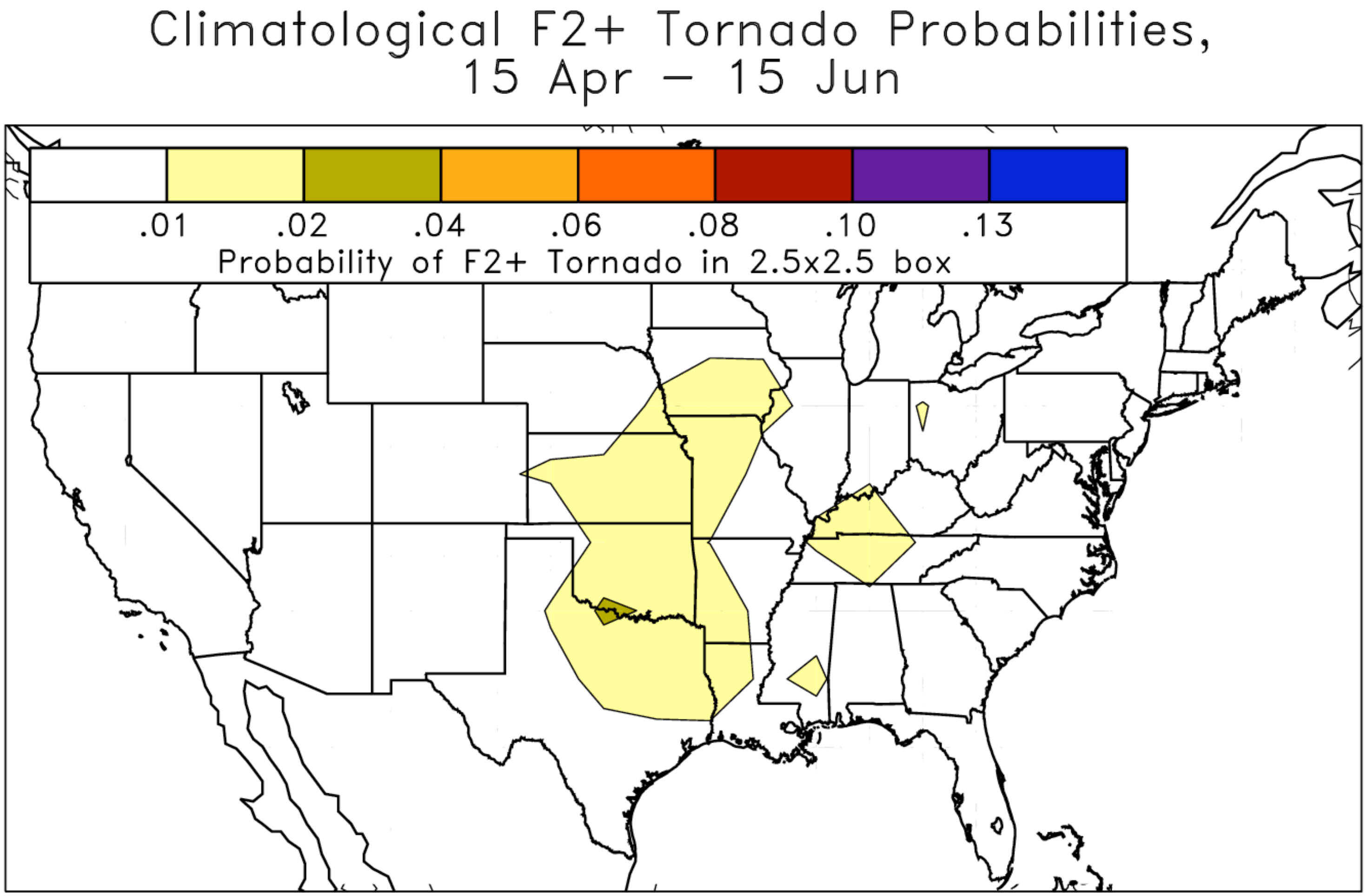


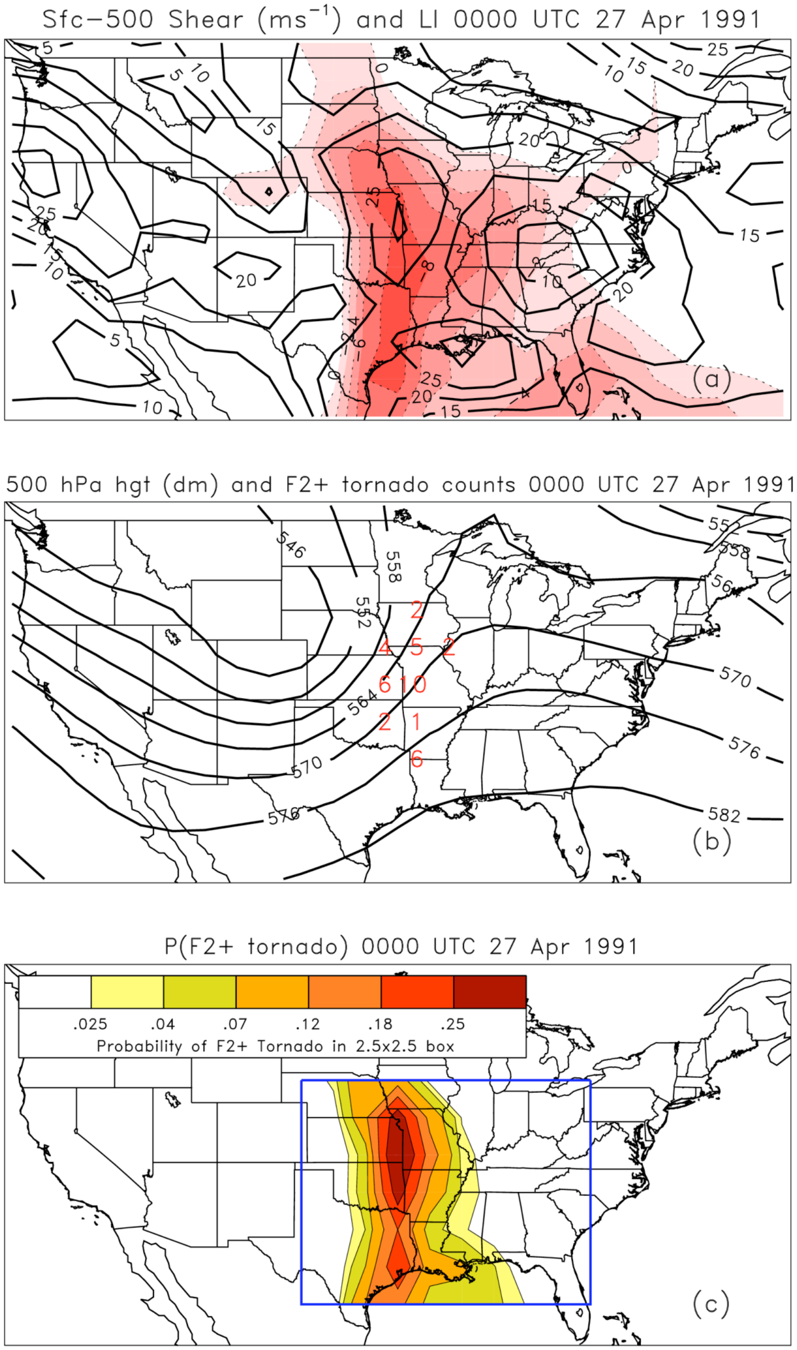
Tornado Probabilities for
01–day Forecast from 26 Apr 1991



Tornado probabilities don't extend back into KS, but this was very fast-moving system, and by 00Z front had moved through central KS.

Climatology of F2+ tornadoes

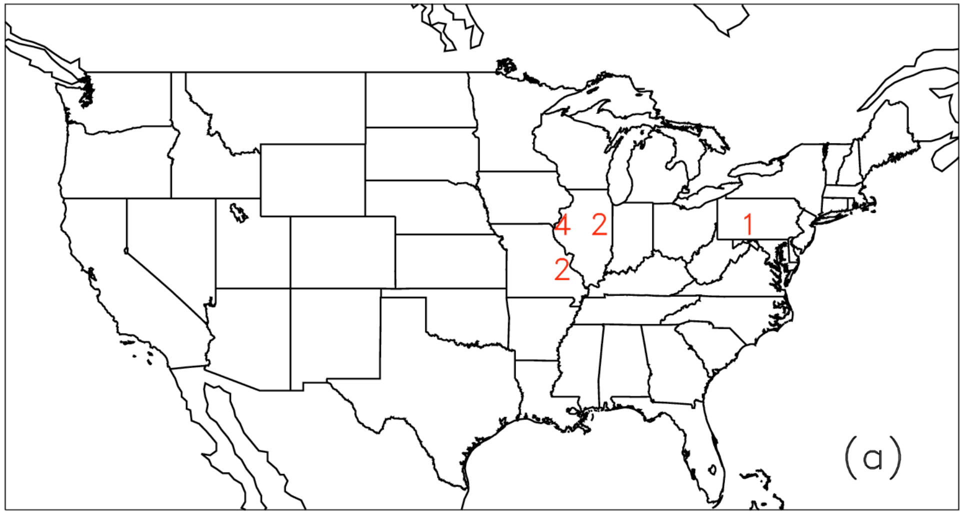




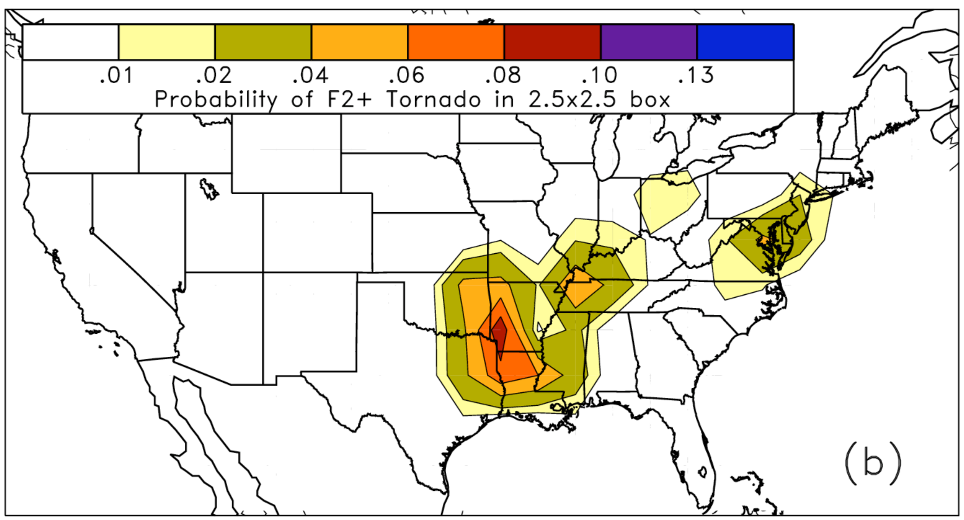
Observed
shear and LI
from NCEP-NCAR
reanalysis

Nowcast tornado probabilities
based on analyzed shear, LI
(see BAMS article on May ‘03
outbreak, Apr ‘05 issue)

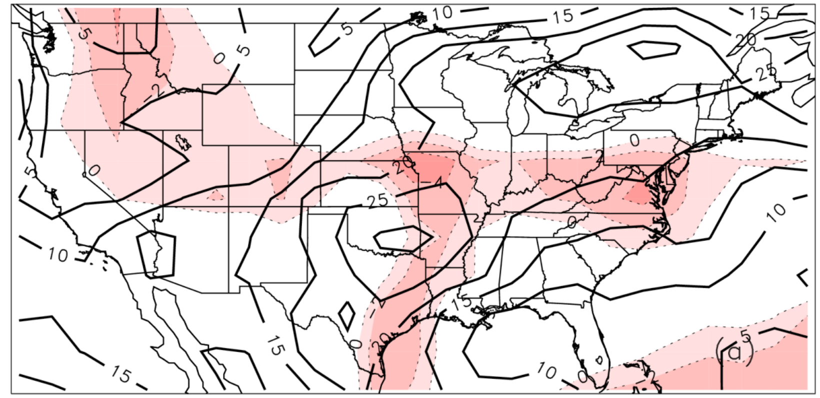
Observed F2+ Tornado Counts in 12–hour Window Centered on 0000 UTC 13 May 1980



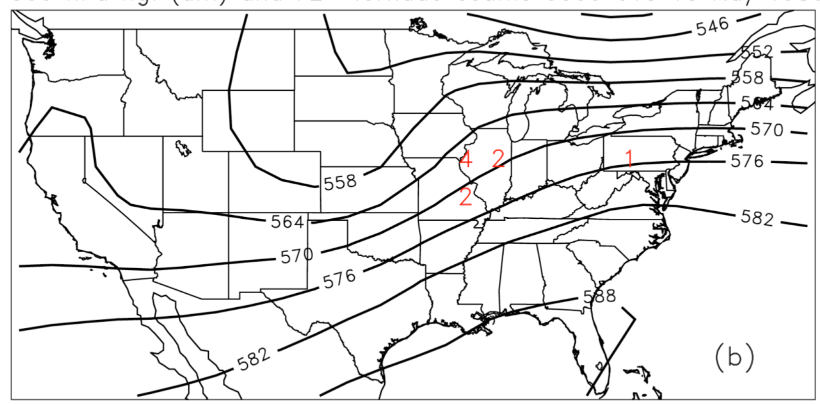
Tornado Probabilities for 01–day Forecast from 12 May 1980



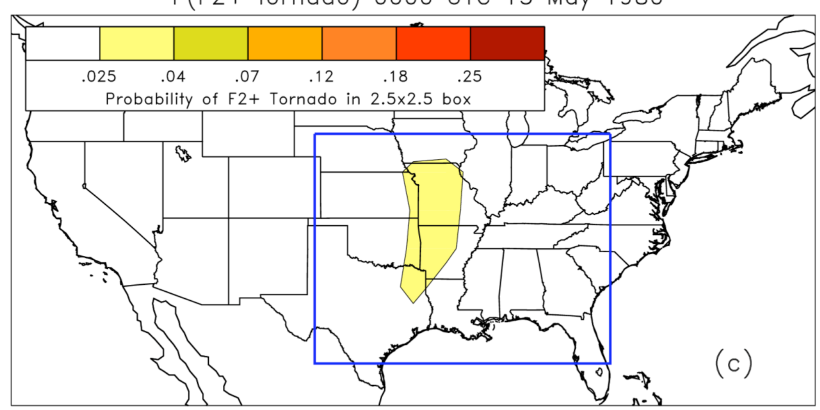
Sfc–500 Shear (ms^{-1}) and LI 0000 UTC 13 May 1980

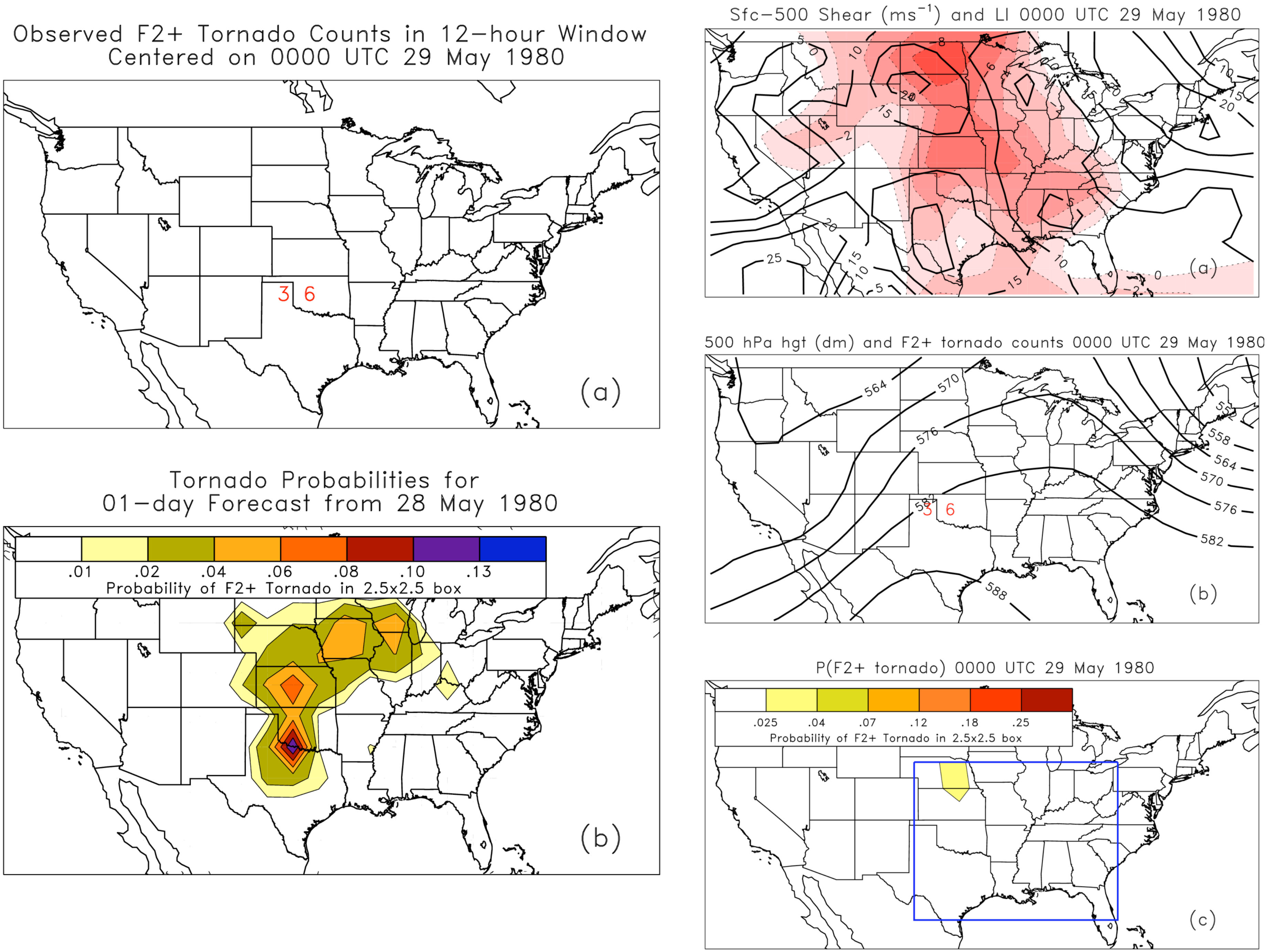


500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 13 May 1980

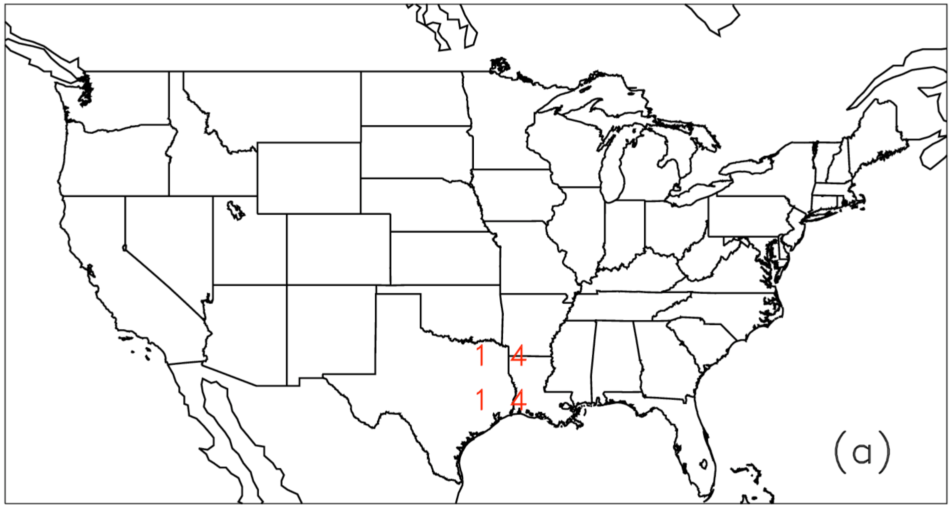


P(F2+ tornado) 0000 UTC 13 May 1980

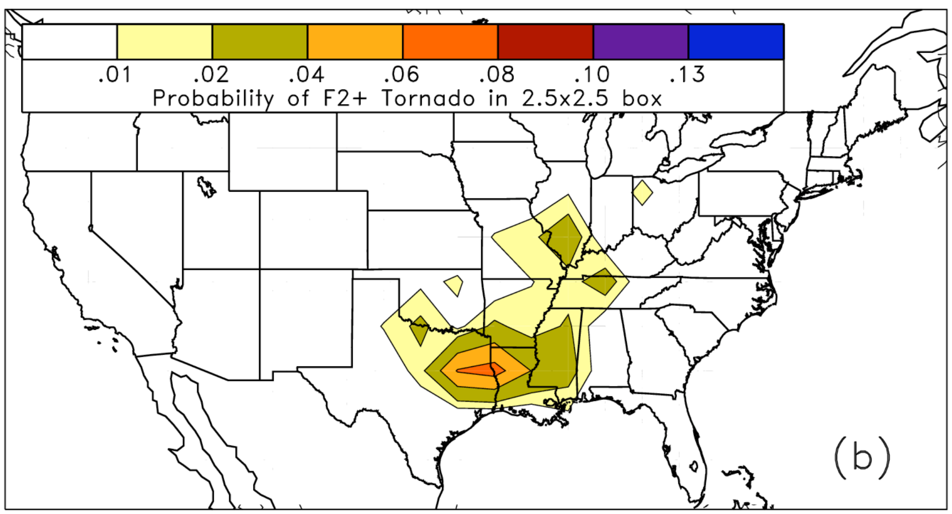



Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 29 May 1980
3 6
(a)
Tornado Probabilities for
01-day Forecast from 28 May 1980
.01 .02 .04 .06 .08 .10 .13
Probability of F2+ Tornado in 2.5x2.5 box
(b)
Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 29 May 1980
(a)
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 29 May 1980
564
570
576
582
588
558
552
3 6
(b)
P(F2+ tornado) 0000 UTC 29 May 1980
.025 .04 .07 .12 .18 .25
Probability of F2+ Tornado in 2.5x2.5 box
(c)

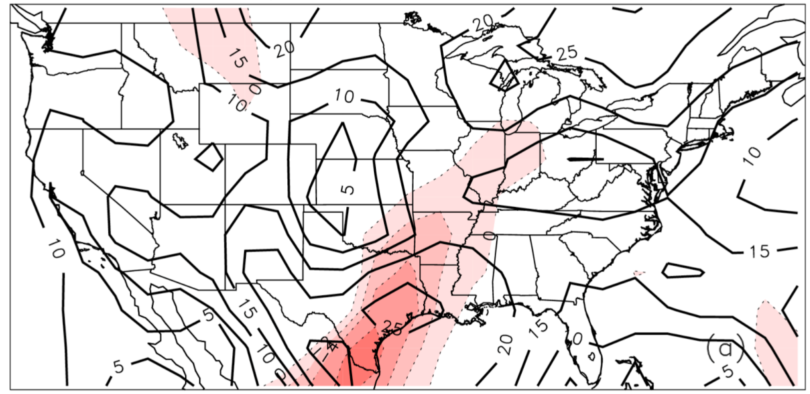
Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 10 May 1981



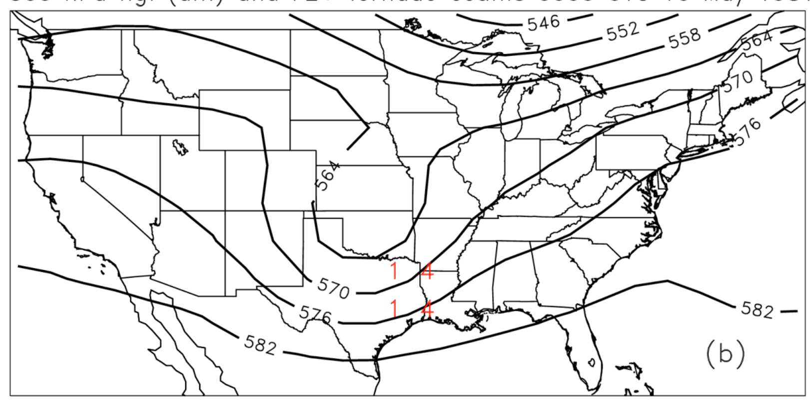
Tornado Probabilities for
01-day Forecast from 09 May 1981



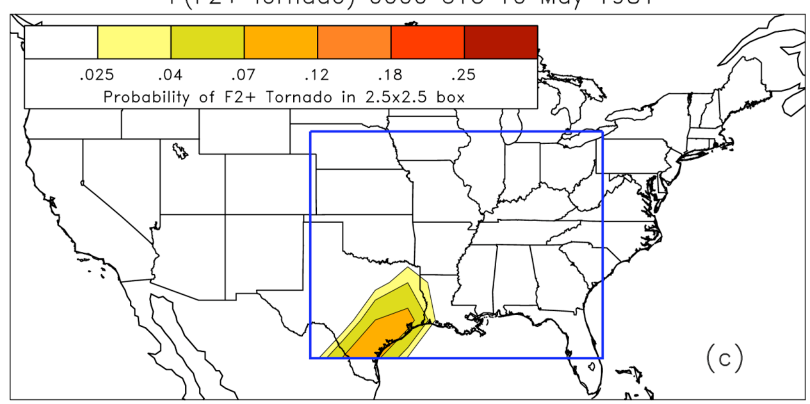
Sfc–500 Shear (ms^{-1}) and LI 0000 UTC 10 May 1981

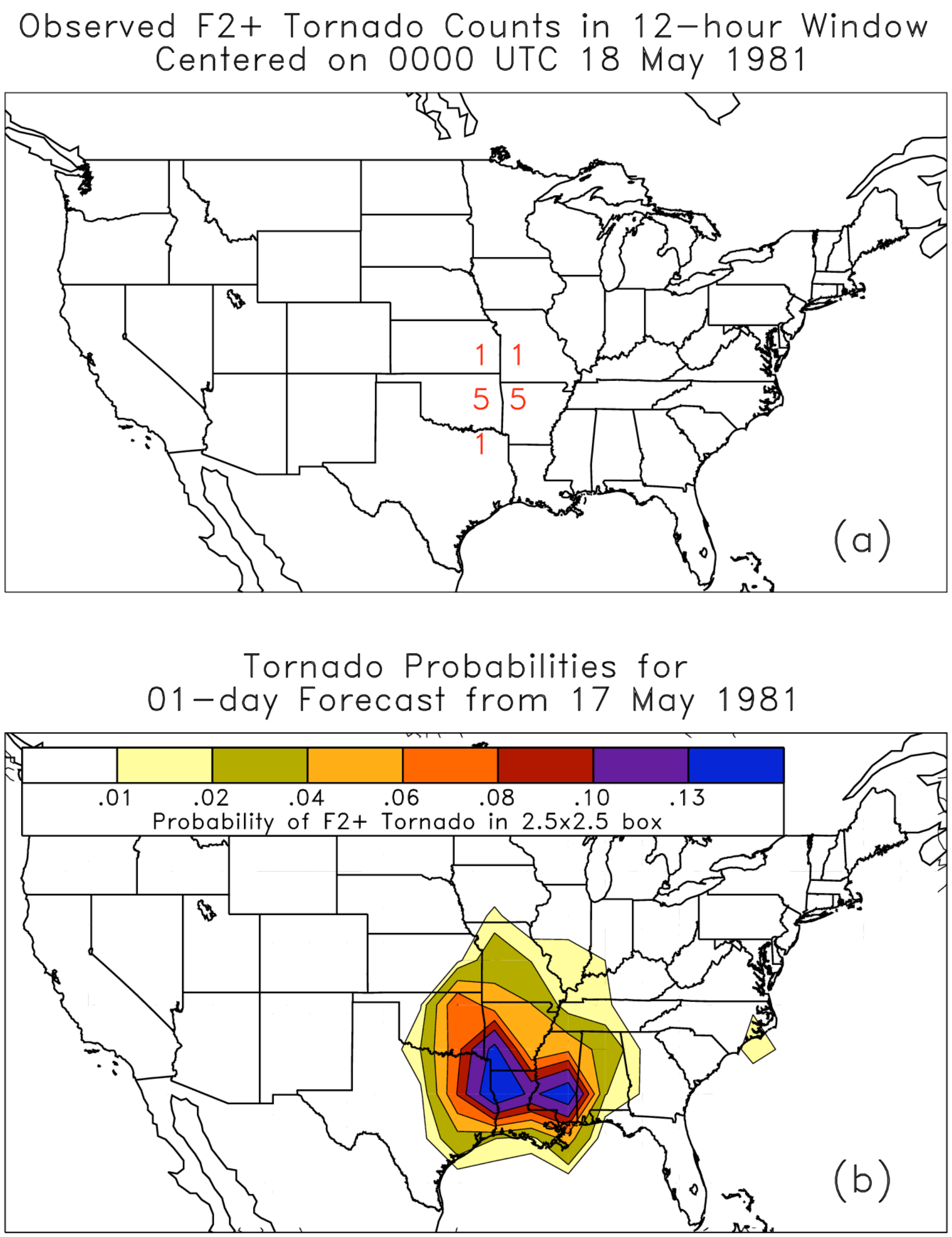


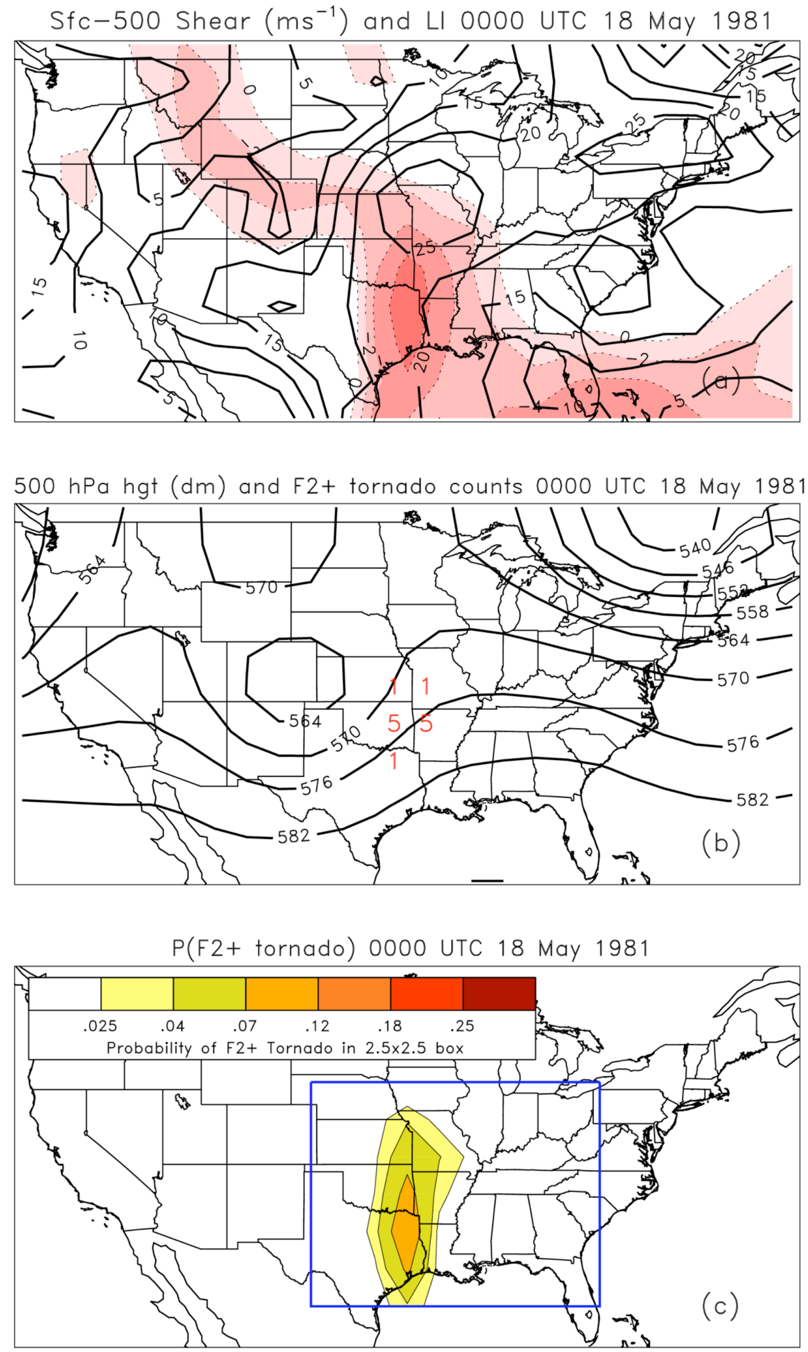
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 10 May 1981



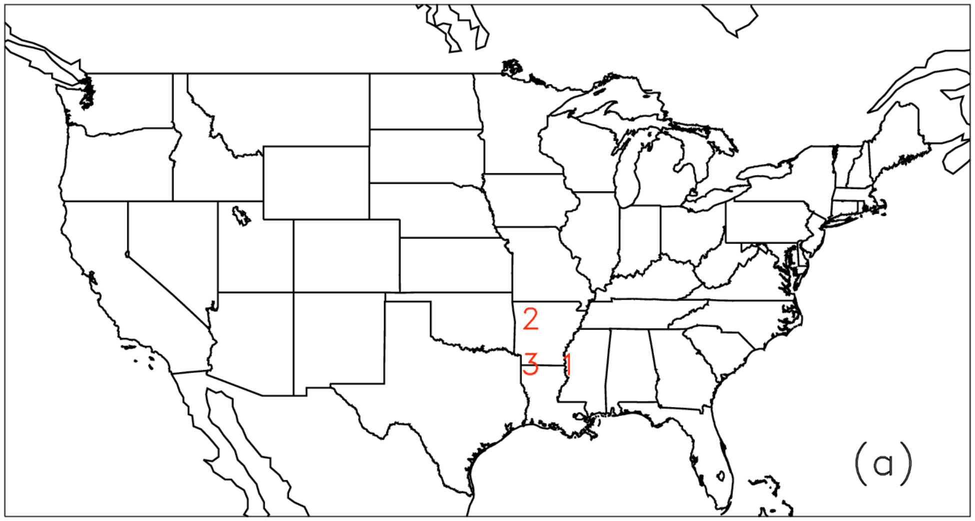
P(F2+ tornado) 0000 UTC 10 May 1981



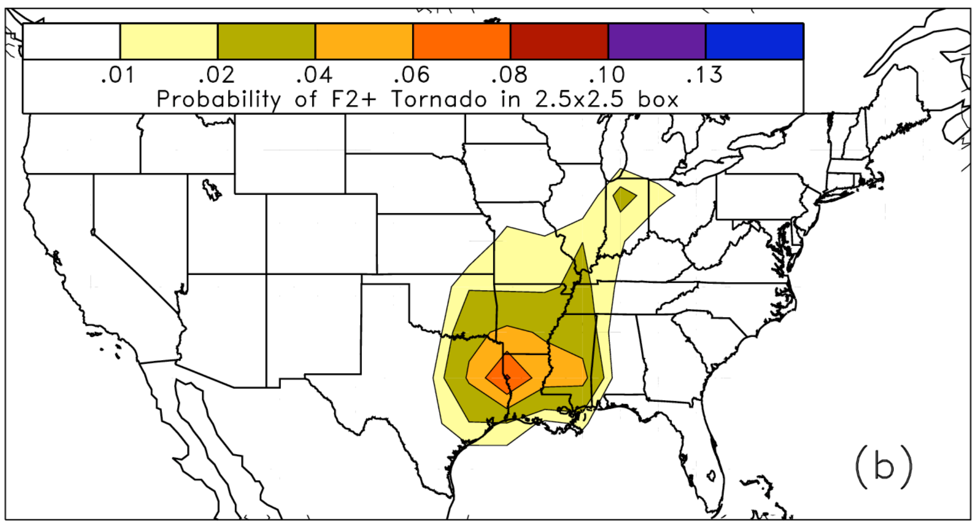

Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 18 May 1981
(a)
Tornado Probabilities for
01-day Forecast from 17 May 1981
.01 .02 .04 .06 .08 .10 .13
Probability of F2+ Tornado in 2.5x2.5 box
(b)


Sfc-500 Shear (ms-1) and LI 0000 UTC 18 May 1981
(a)
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 18 May 1981
(b)
P(F2+ tornado) 0000 UTC 18 May 1981
.025 .04 .07 .12 .18 .25
Probability of F2+ Tornado in 2.5x2.5 box
(c)

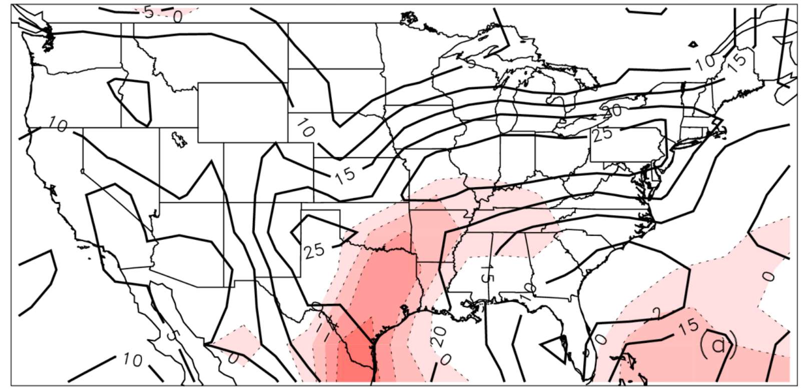
Observed F2+ Tornado Counts in 12−hour Window
Centered on 0000 UTC 14 May 1981



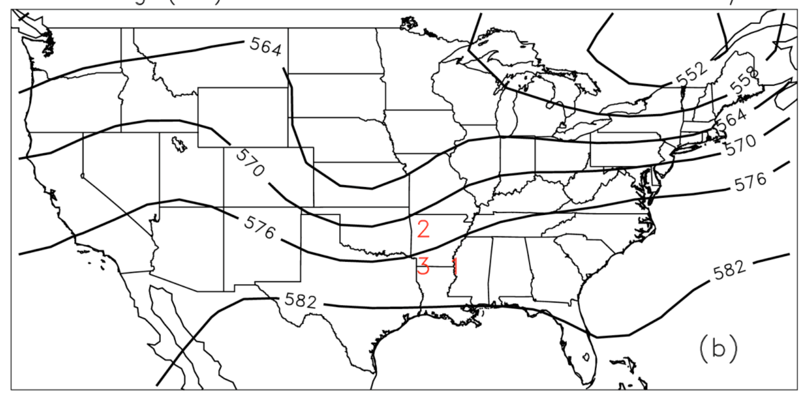
Tornado Probabilities for
01−day Forecast from 13 May 1981



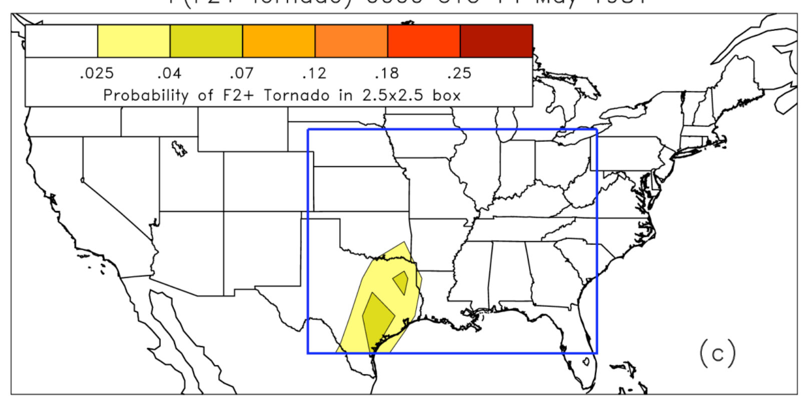
Sfc−500 Shear (ms^{-1}) and LI 0000 UTC 14 May 1981

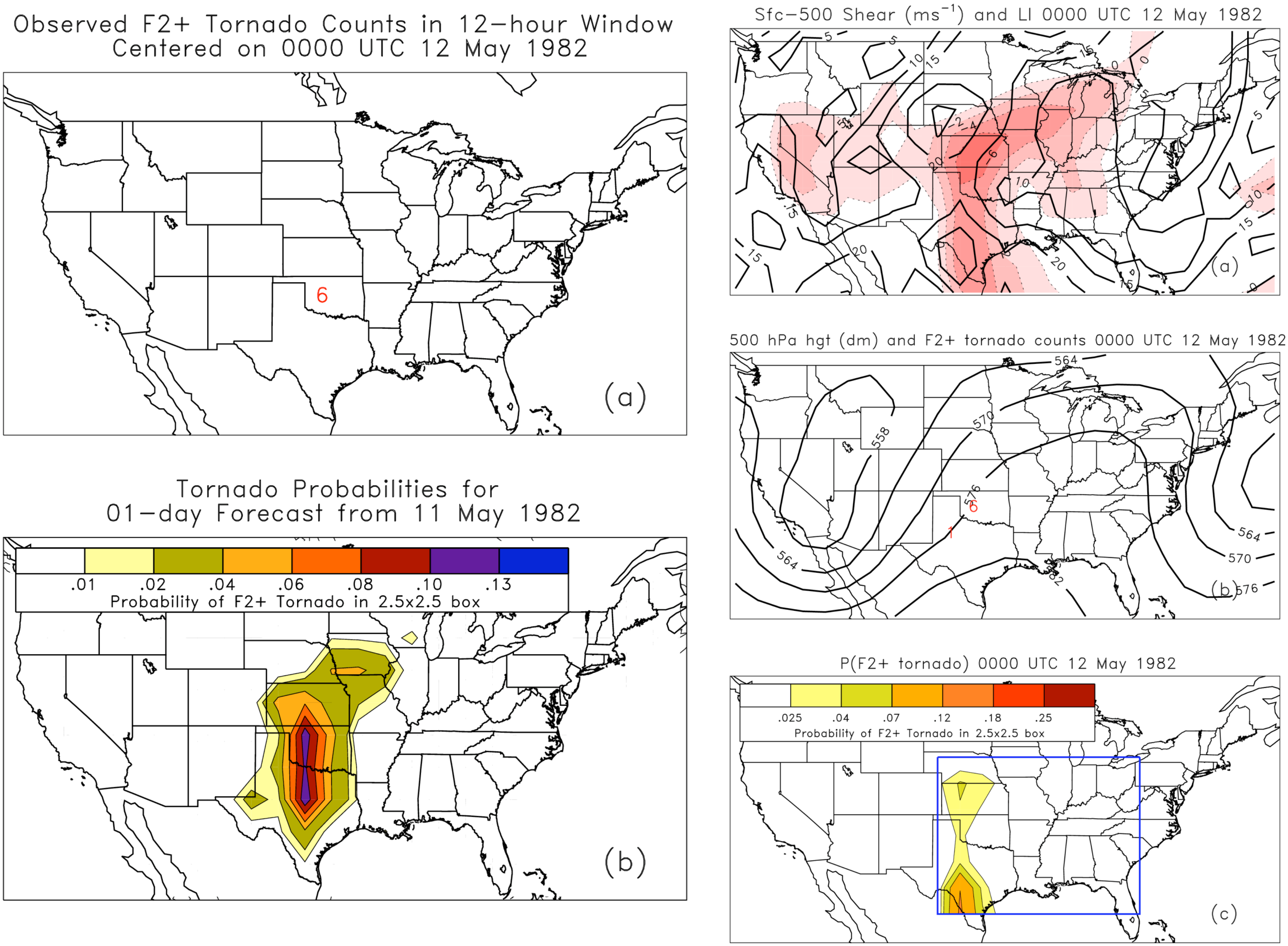


500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 14 May 1981



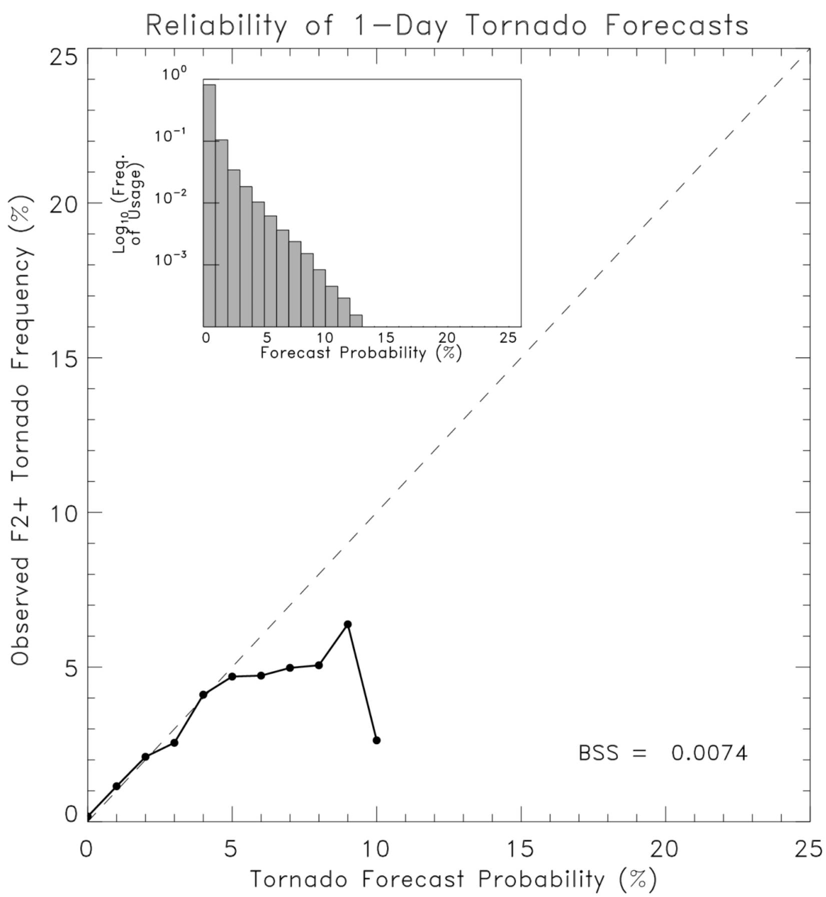
P(F2+ tornado) 0000 UTC 14 May 1981

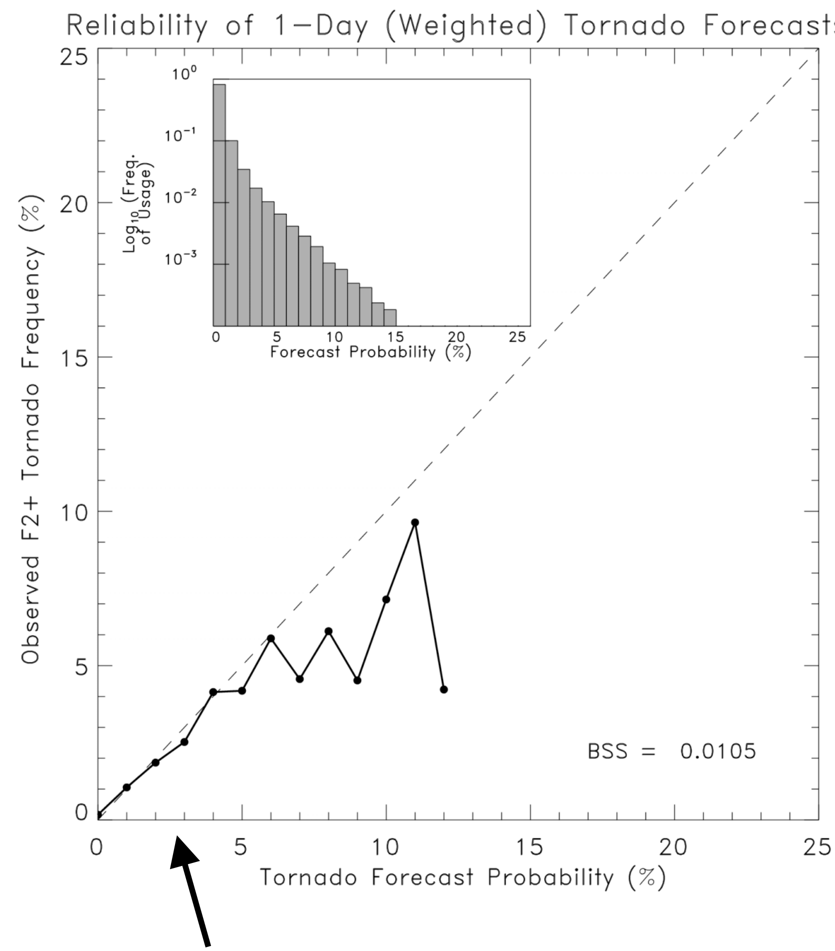




Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 12 May 1982
6
(a)
Tornado Probabilities for
01-day Forecast from 11 May 1982
.01 .02 .04 .06 .08 .10 .13
Probability of F2+ Tornado in 2.5x2.5 box
(b)
Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 12 May 1982
(a)
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 12 May 1982
564
570
576
558
582
6
1
(b)
P(F2+ tornado) 0000 UTC 12 May 1982
.025 .04 .07 .12 .18 .25
Probability of F2+ Tornado in 2.5x2.5 box
(c)

Reliability and Skill





here, closer analogs are weighted more than further ones

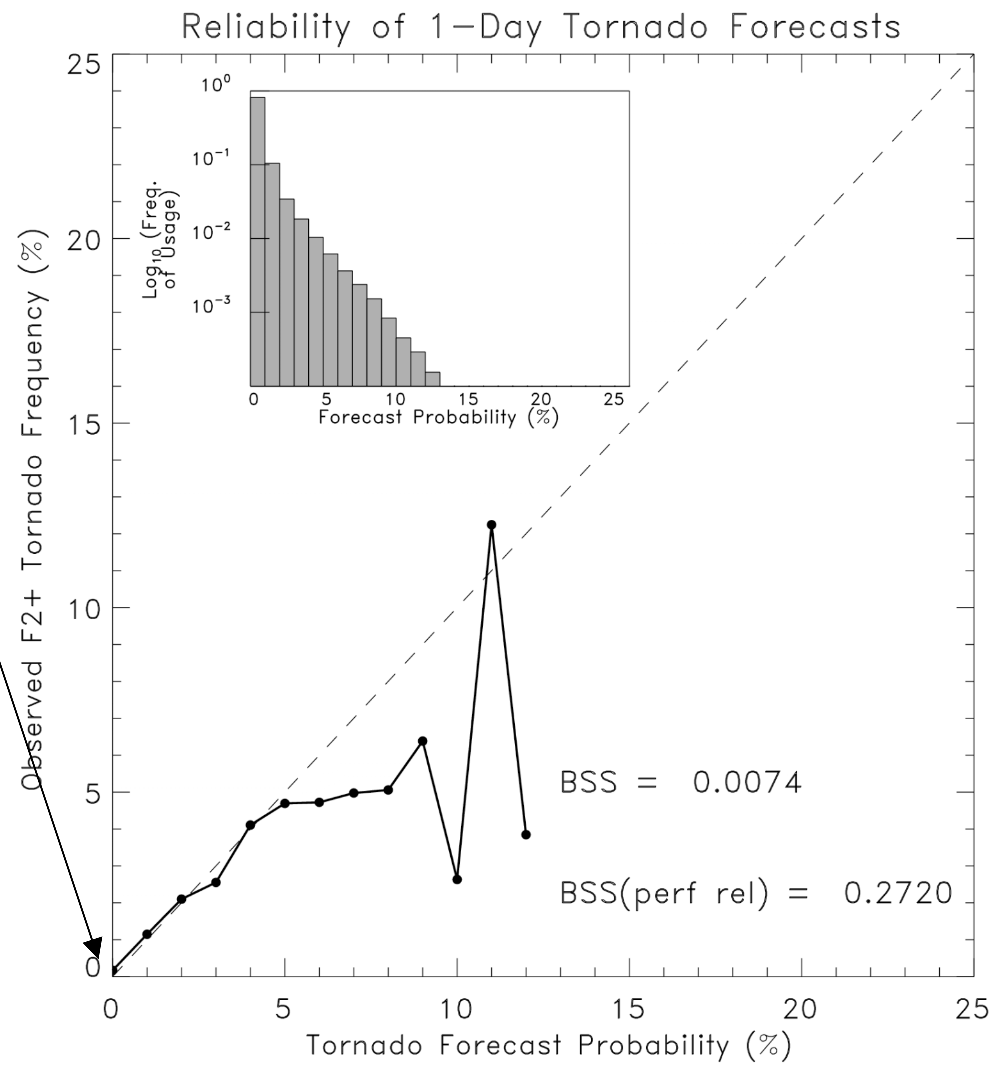
What’s the cause of low BSS?

Suppose we could simply change contingency table to enforce perfect reliability; then BSS goes up to 0.27.

However, MOST of the improvement is simply changing the few cases at 0.0 probability.

Our intuition is that in locations where climo probability is ~0.0, for days with high CAPE/shear, it’s tough to find other analog days where tornadoes occurred.

!

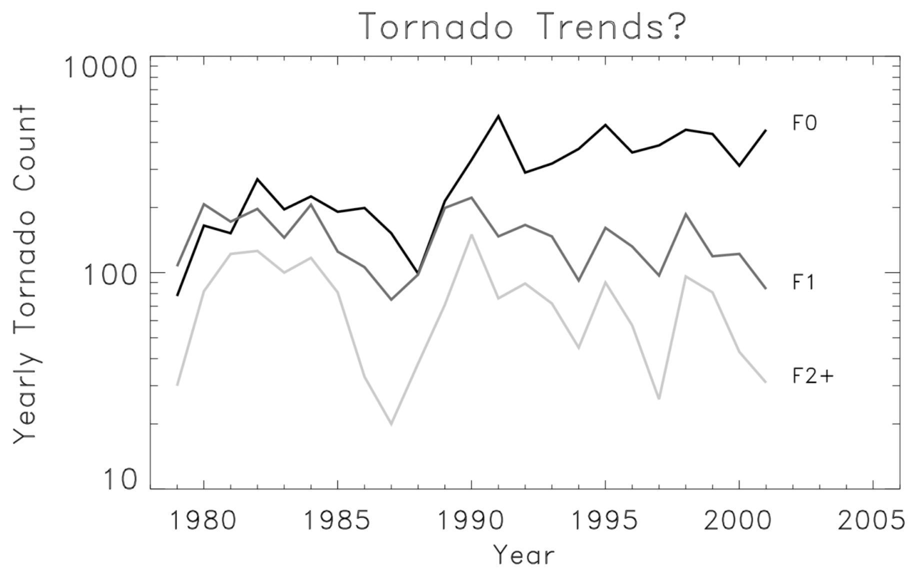


Other results

- CIN ~ useful as predictor.
- *n*=100 analogs much better than *n*=50.
- Sfc-650 shear not as good a predictor as Sfc-500 shear
- No skill (yet) beyond day 2.

Possible ways to improve

- Rarity of events part of the problem; use F1+, not F2+. But need stationary climatology.



- We didn't composite samples from different locations together, given regional nature of biases. Possibly next time, do 2-step procedure: (1) bias correct fields, (2) composite samples from different locations when doing calibration.
- Get a next-generation reforecast data set!

Can we do both hi-res model development and reforecasting, or a compromise?

- **Alternative 1**: Continue development of high-res. models. Do reforecasting with inexpensive, low-res. model, so operations are impacted minimally.
  - Suppose operational T300, 60-layer, 50-member ensemble forecast system.
  - Reforecast T150, 40 layer, 5-member ensemble :
    - Operational cost: 120x less
    - 120 days of reforecasts for one day of operational forecast, so a 20-year reforecast for the cost of 60 days of operational model forecasts.
    - If new reforecast model implemented once, say, every 4 years, minimal impact to operations integrated over time.

- **Alternative 2**: Continue development of high-res. models. Do reforecasting offline, on non-operational computer system.
  - ~ $700K would buy a computer system that could do a T170L42, 5-member reforecast out to 10 days in ~ 1 year wall time.

What's next for reforecasting?

- Growing interest from NWP centers worldwide
  - ECMWF exploring once-weekly ensemble reforecasts (with my participation)
  - Canadians planning 5-year ensemble reforecasts
  - NCEP envisioning 1-member, real-time reforecast for bias correction.
- Possibility that NOAA/ESRL may get money to do a more complete, 2nd-generation reforecast data set for NOAA.
- Being discussed in NOAA's strategic planning.

Research questions

- Given computational expense of reforecasts, how do we best:
  - Limit the number of reforecasts that we need to do (fewer ensemble members, not every day, etc.)
  - Can we do things like composite the data across different locations to boost sample size?
  - Do we need a new reanalysis every time we do a new reforecast?
  - Do the benefits of reforecasts propagate down to users like hydrological forecasters?
- We welcome your thoughts and requirements for next-generation reforecast system.

References

Hamill, T. M., J. S. Whitaker, and X. Wei, 2003: Ensemble re-forecasting: improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev*., **132**, 1434-1447.
http://www.cdc.noaa.gov/people/tom.hamill/reforecast_mwr.pdf

Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2005: Reforecasts, an important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc*., **87**, 33-46.
http://www.cdc.noaa.gov/people/tom.hamill/refcst_bams.pdf

Whitaker, J. S, F. Vitart, and X. Wei, 2006: Improving week two forecasts with multi-model re-forecast ensembles. *Mon. Wea. Rev*., **134**, 2279-2284.
http://www.cdc.noaa.gov/people/jeffrey.s.whitaker/Manuscripts/multimodel.pdf

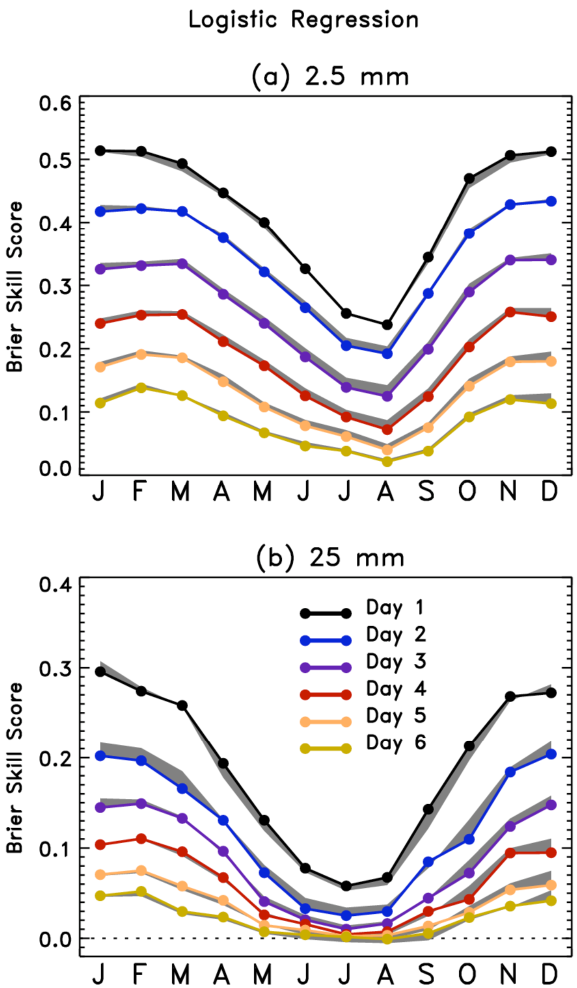
Hamill, T. M., and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application. *Mon. Wea. Rev*., in press.
http://www.cdc.noaa.gov/people/tom.hamill/reforecast_analog_v2.pdf

Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: is it real skill or is it the varying climatology? *Quart. J. Royal Meteor. Soc*., in press. http://www.cdc.noaa.gov/people/tom.hamill/skill_overforecast_QJ_v2.pdf

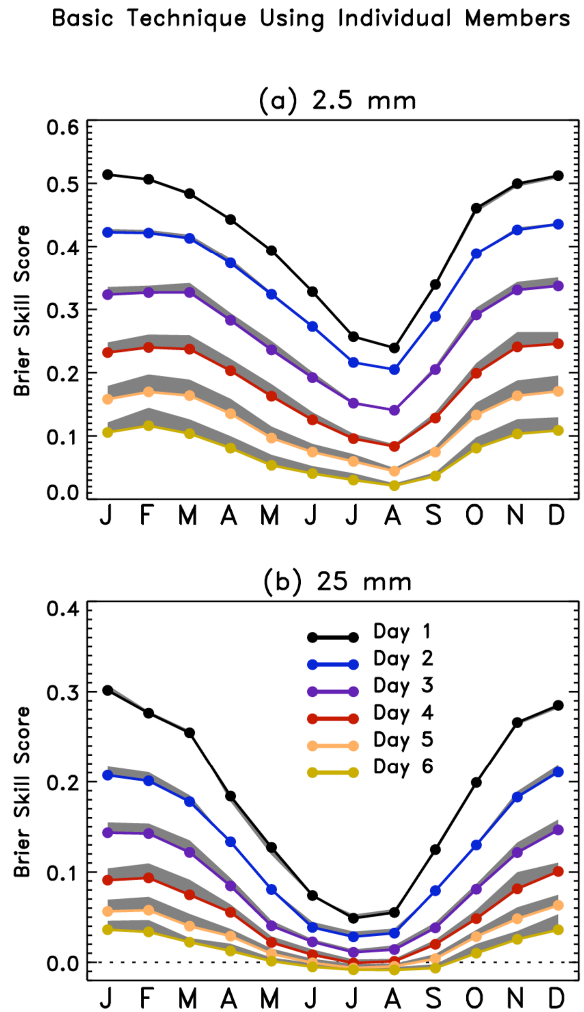
Wilks, D. S., and T. M. Hamill, 2006: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev*., in press. http://www.cdc.noaa.gov/people/tom.hamill/WilksHamill_emos.pdf

Hamill, T. M. and J. S. Whitaker, 2006: White Paper. "Producing high-skill probabilistic forecasts using reforecasts: implementing the National Research Council vision." Available at
http://www.cdc.noaa.gov/people/tom.hamill/whitepaper_reforecast.pdf .

Some other tests

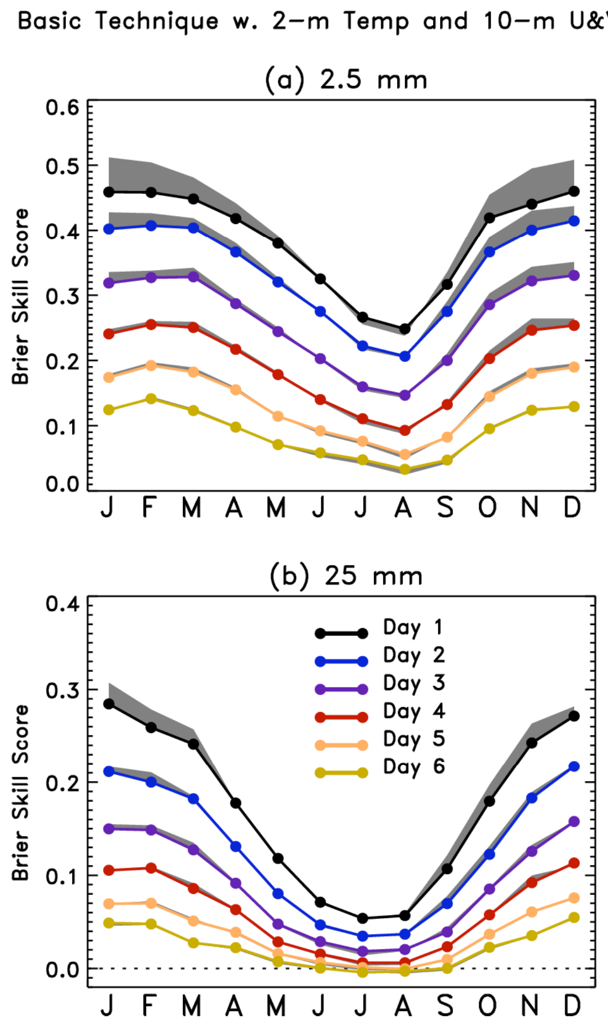


Mixed results when probabilistic forecasts generated using logistic regression approach.

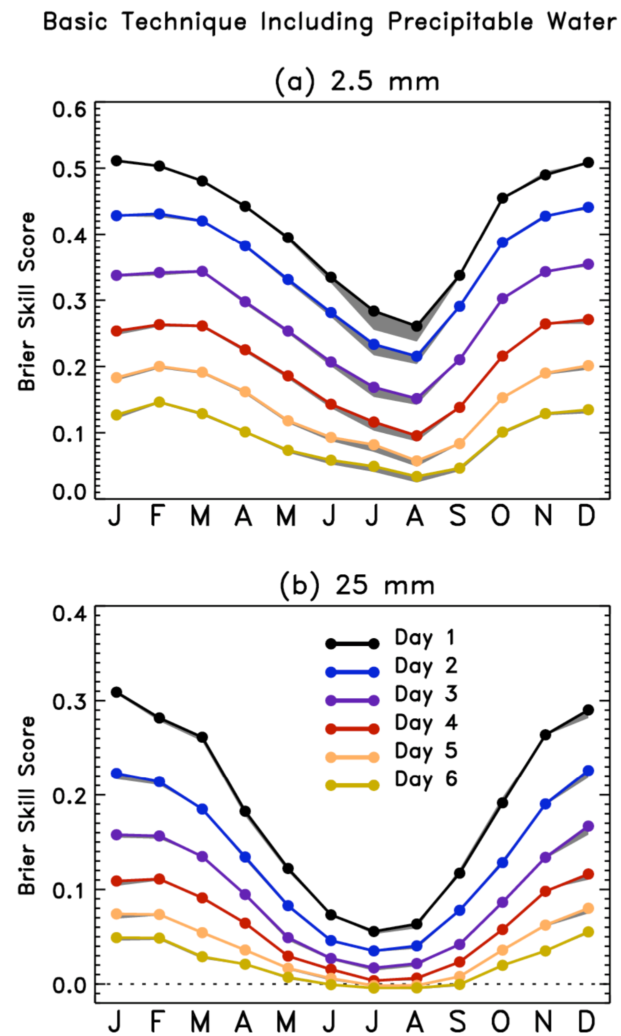


Worse skill when attempting to fit individual members.

Other tests, continued

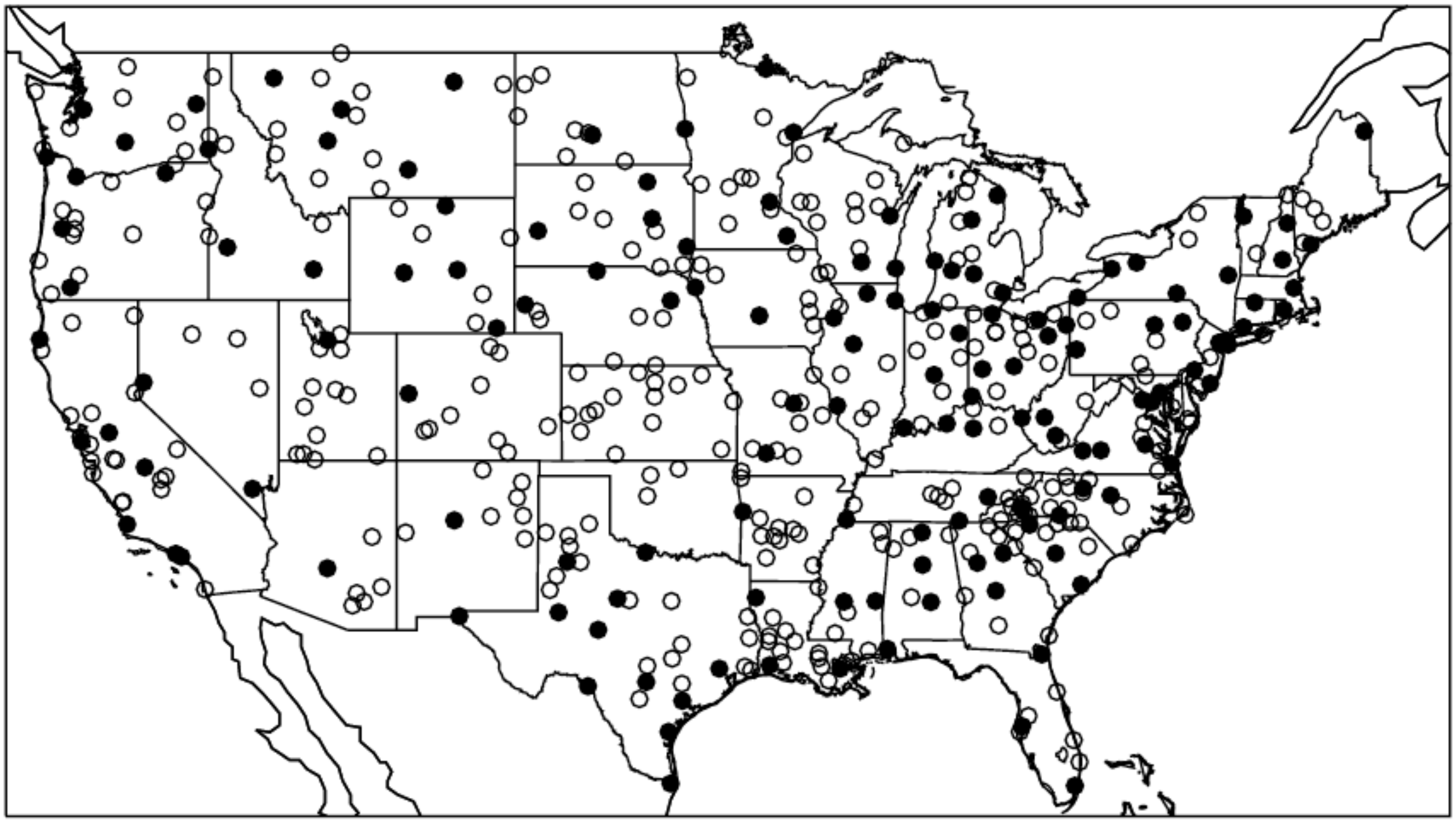


Worse skill when basing analogs on precip/U/V/T fit.



Some skill improvement in the summer when adding precipitable water as predictor.

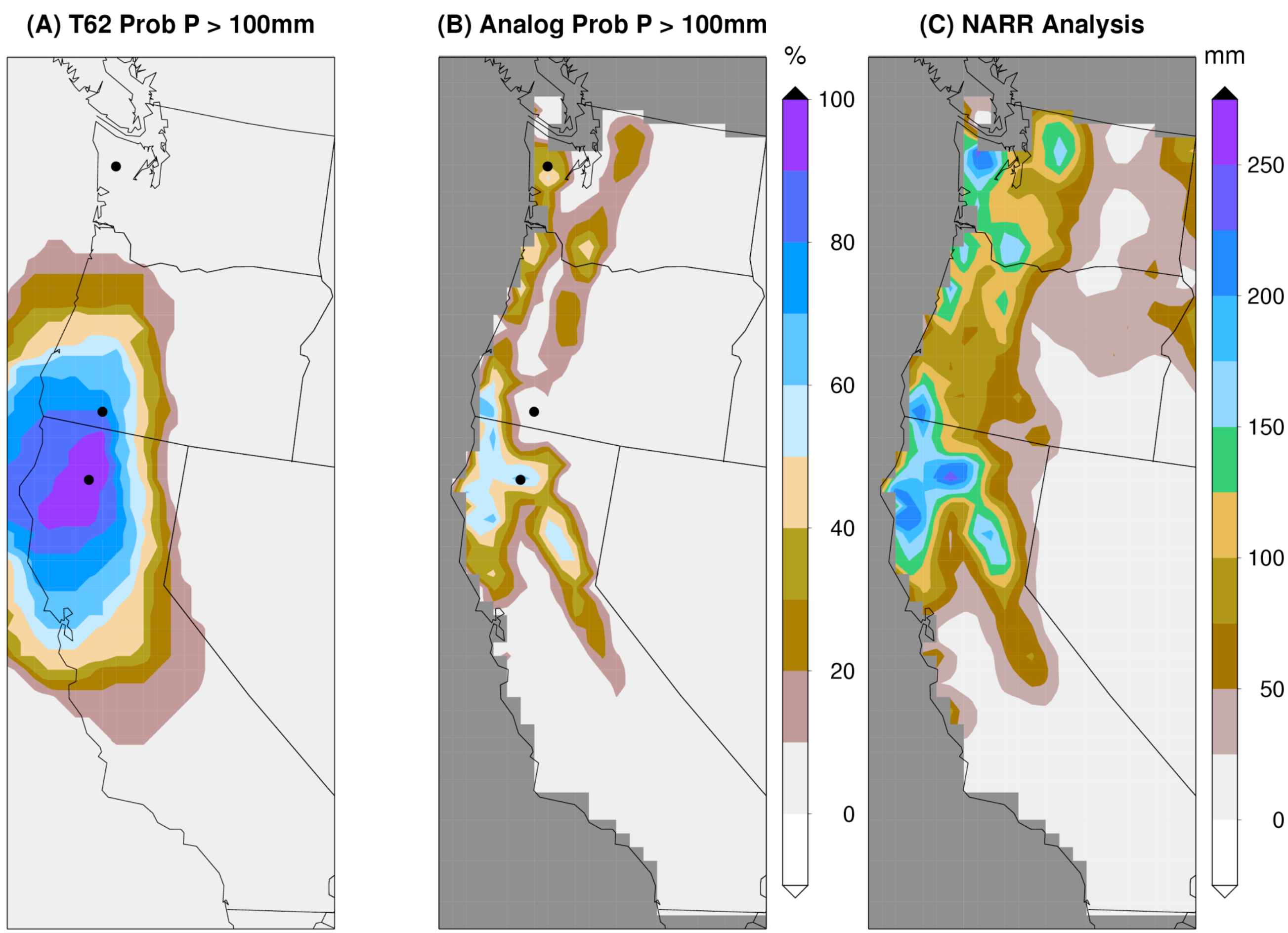
Station Locations in CONUS



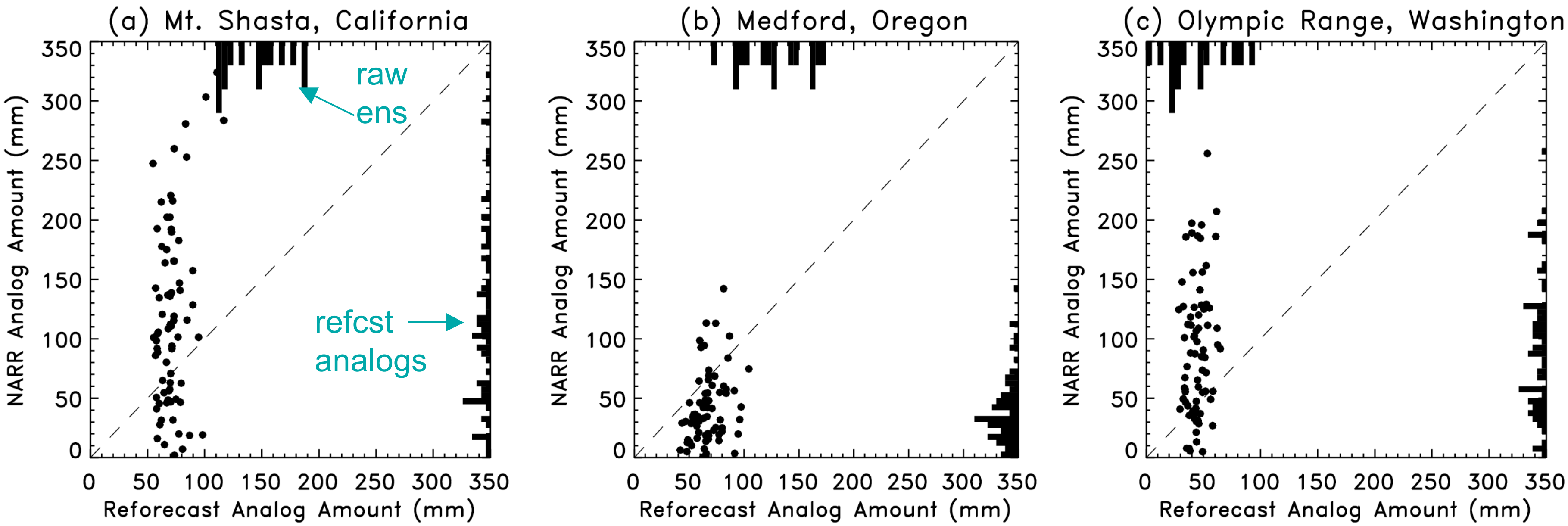
Filled + unfilled dots: general verification

Filled: comparison against CPC operational

Analog example:
Day 4-6 heavy precipitation in California, 0000 UTC 29 December 1996 - 0000 UTC 1 January 1997



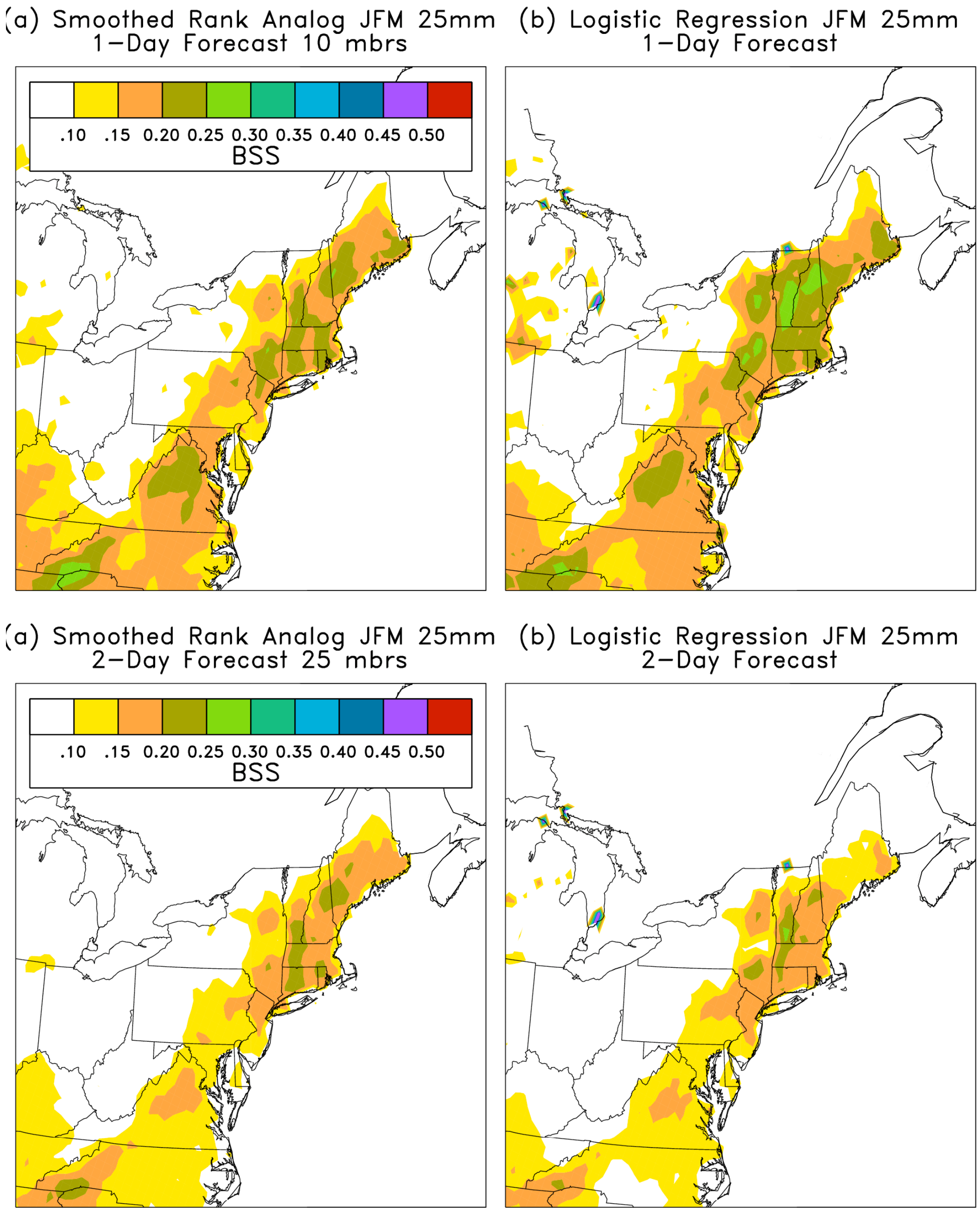
Bias, spread, and downscaling corrections in analog technique



Can't find any other reforecast analogs with precip as heavy. But introduce large scatter by taking associated observed analogs.

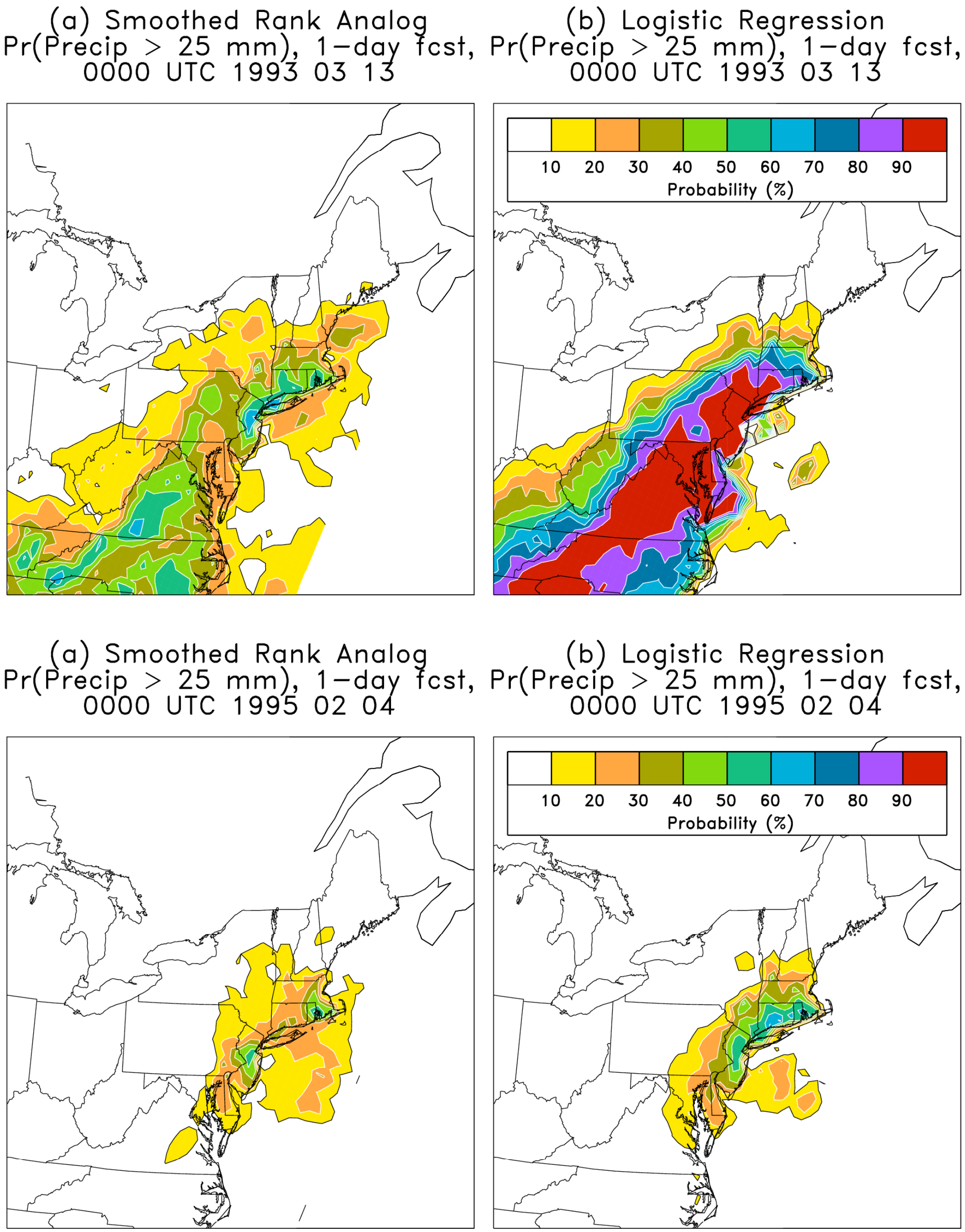
Again, few close reforecast analogs. But observed data recognizes overforecast bias.

Here there are close reforecast analogs. Observed data introduces spread, increases amount.



We compare here the smoothed rank analog approach to the logistic regression approach for wintertime (JFM) data over the northeast USA. The focus is specifically on the 25-mm threshold, i.e., the quality of forecasting heavy-precipitation events.

First, notice that maps of the overall precipitation forecast skill are relatively similar, here for day-1 and day-2 forecasts. The logistic regression appears to be slightly more skillful over New England on day 1.



Next, consider some individual storms and their forecasts. For record-setting events like 1993's "Storm of the Century", logistic regression "extrapolates the regression" and produces much higher probabilities. 10-member rank analog techniques produced much lower probabilities, since most if not all reforecast analogs that were selected inevitably had lower forecast (and presumably analyzed) precipitation amounts.

Finding analogs, cont'd: horizontal weighting

5x5 arrays of shear, CAPE, CIN weighted by distance from center grid point; controlled by e-folding distance

- E-folding of 7.5 grid points

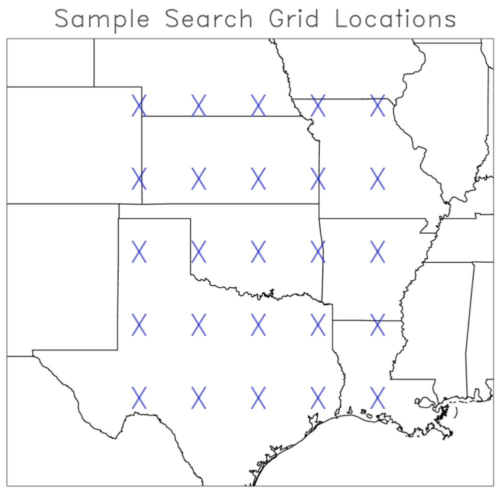
0.867	0.915	0.931	0.915	0.867
0.915	0.965	0.982	0.965	0.915
0.931	0.982	1.000	0.982	0.931
0.915	0.965	0.982	0.965	0.915
0.867	0.915	0.931	0.915	0.867

- E-folding of 4.5 grid points

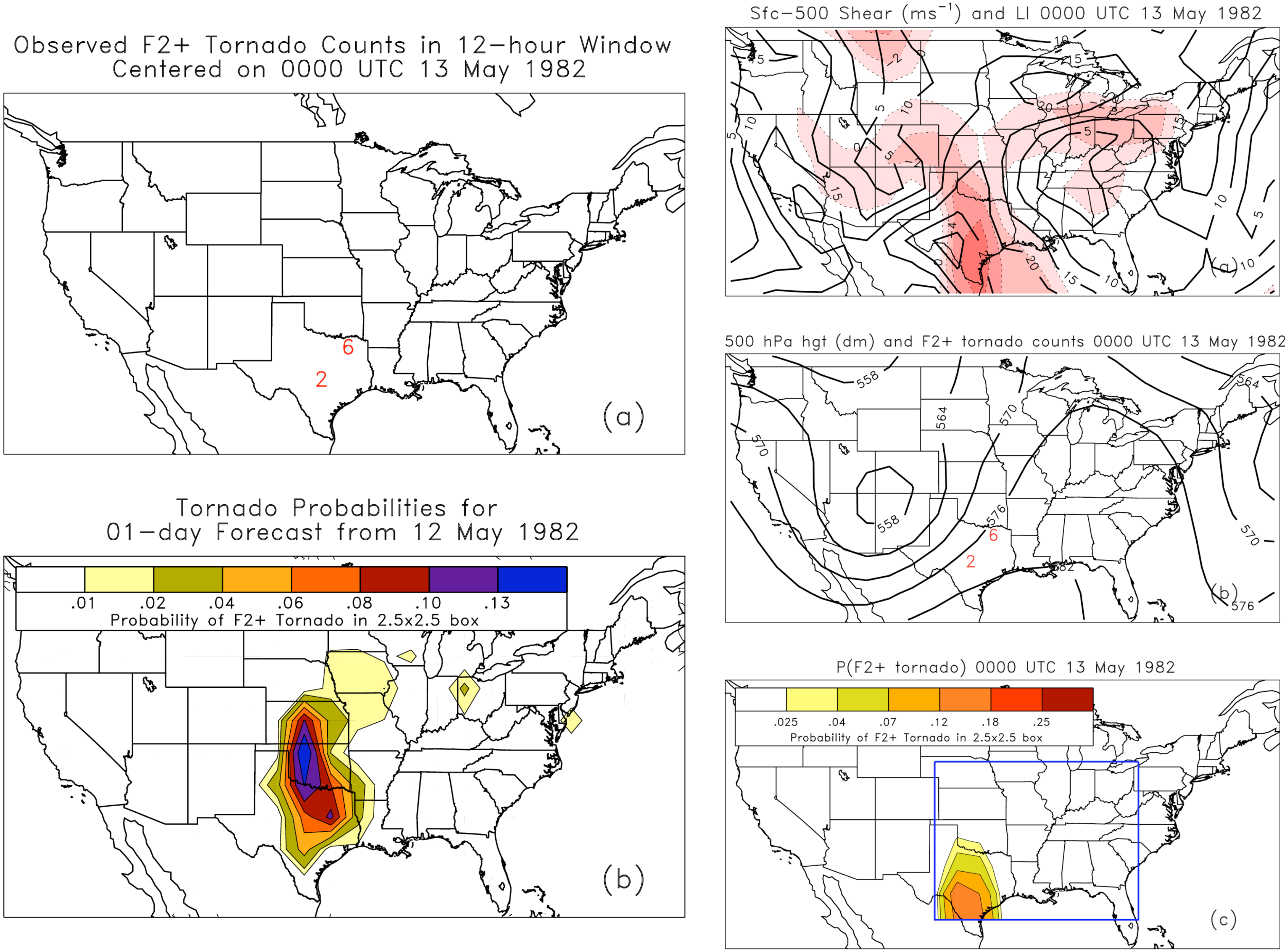
0.674	0.781	0.821	0.781	0.674
0.781	0.906	0.952	0.906	0.781
0.821	0.952	1.000	0.952	0.821
0.781	0.906	0.952	0.906	0.781
0.674	0.781	0.821	0.781	0.674

- E-folding of 1.5 grid points

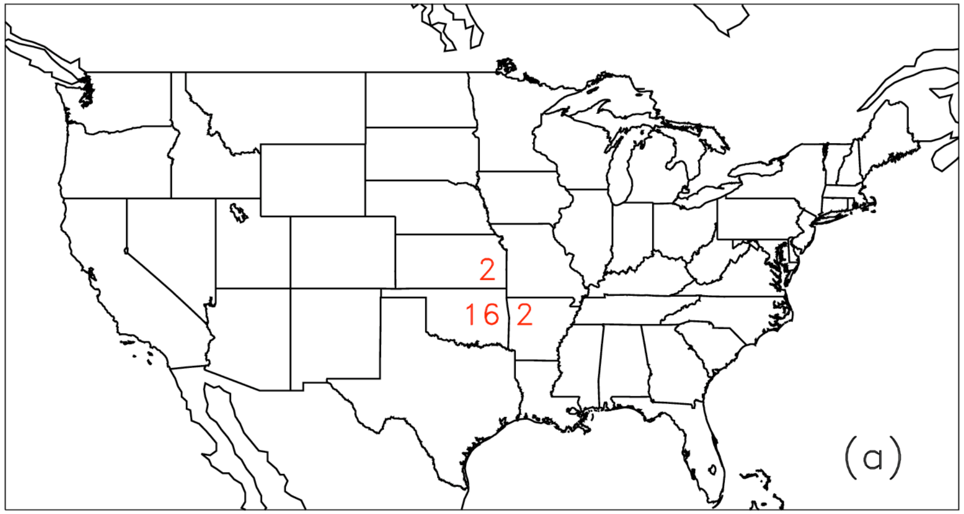
0.029	0.108	0.169	0.108	0.029
0.108	0.411	0.641	0.411	0.108
0.169	0.641	1.000	0.641	0.169
0.108	0.411	0.641	0.411	0.108
0.029	0.108	0.169	0.108	0.029



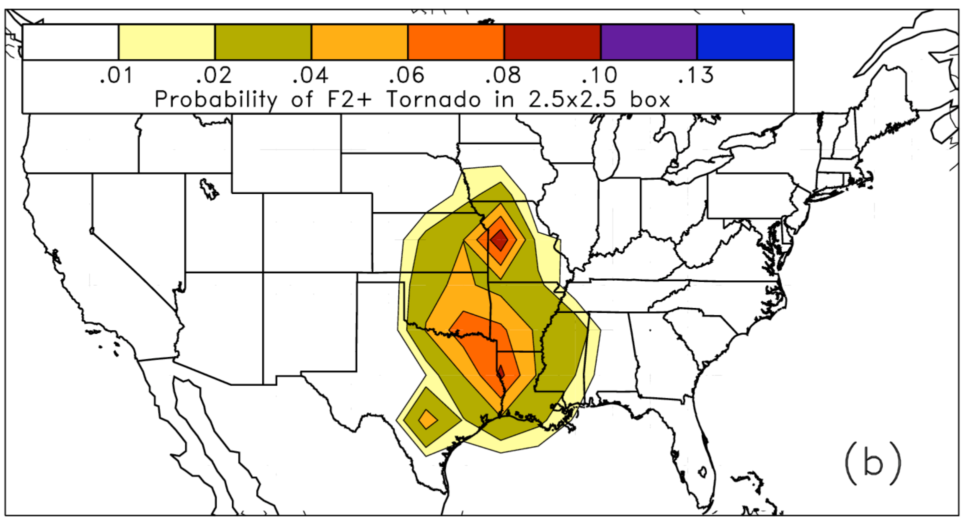
(Comparing skill using many e-folding distances this will indicate if the local information is of primary importance, or the larger-scale pattern.)


Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 13 May 1982
6
2
(a)
Tornado Probabilities for
01-day Forecast from 12 May 1982
.01 .02 .04 .06 .08 .10 .13
Probability of F2+ Tornado in 2.5x2.5 box
(b)
Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 13 May 1982
(a)
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 13 May 1982
558
564
570
576
6
2
(b)
P(F2+ tornado) 0000 UTC 13 May 1982
.025 .04 .07 .12 .18 .25
Probability of F2+ Tornado in 2.5x2.5 box
(c)

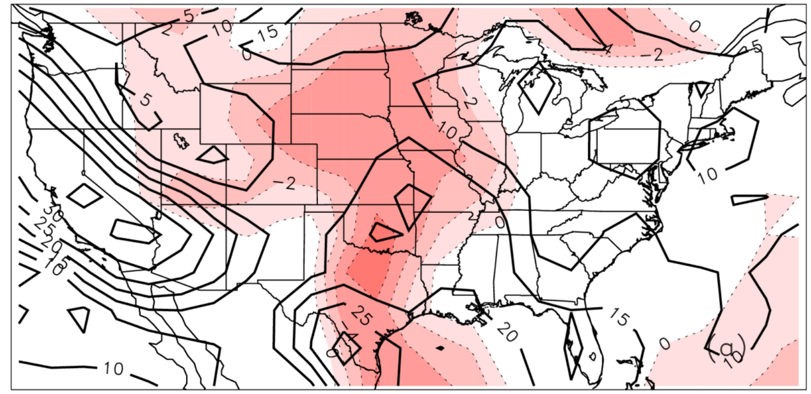
Observed F2+ Tornado Counts in 12–hour Window
Centered on 0000 UTC 04 May 1999



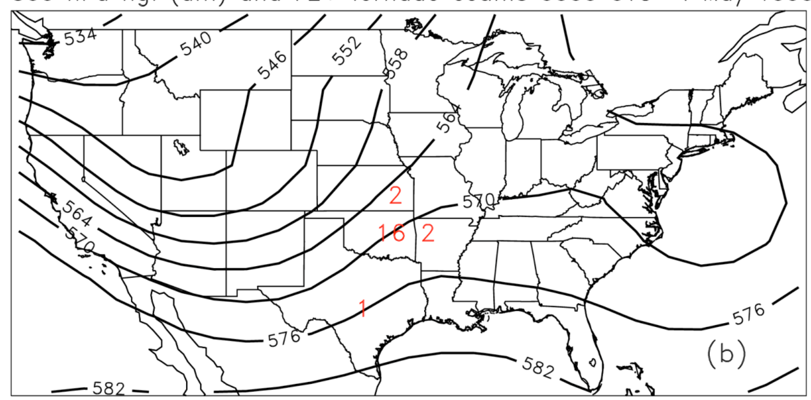
Tornado Probabilities for
01–day Forecast from 03 May 1999



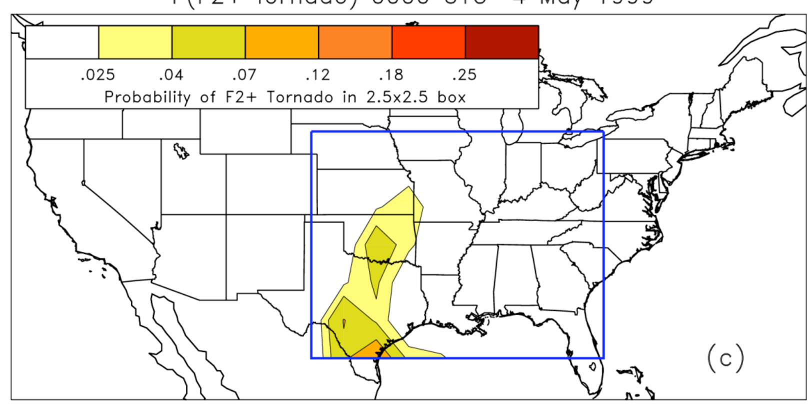
Sfc–500 Shear (ms^{-1}) and LI 0000 UTC 4 May 1999

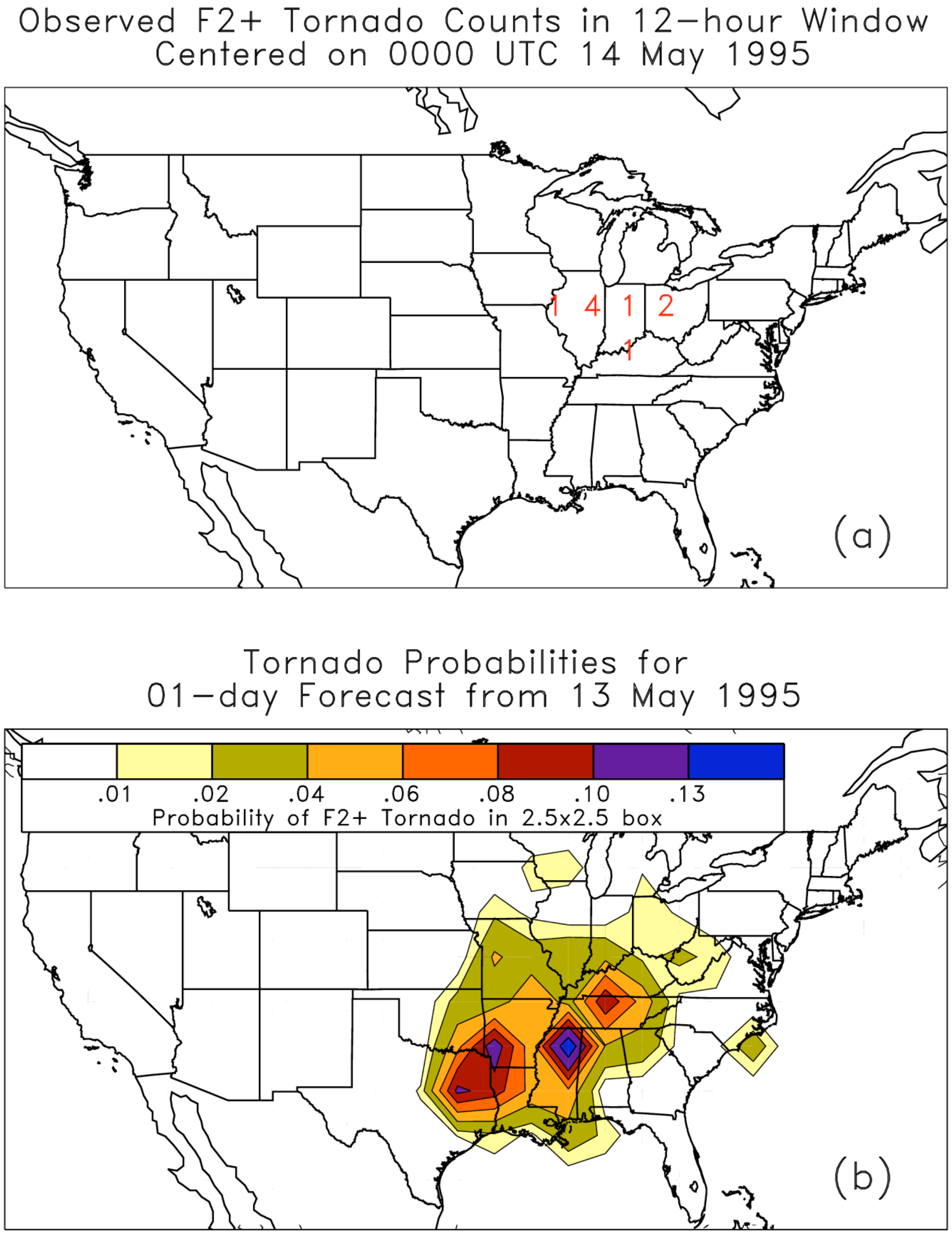


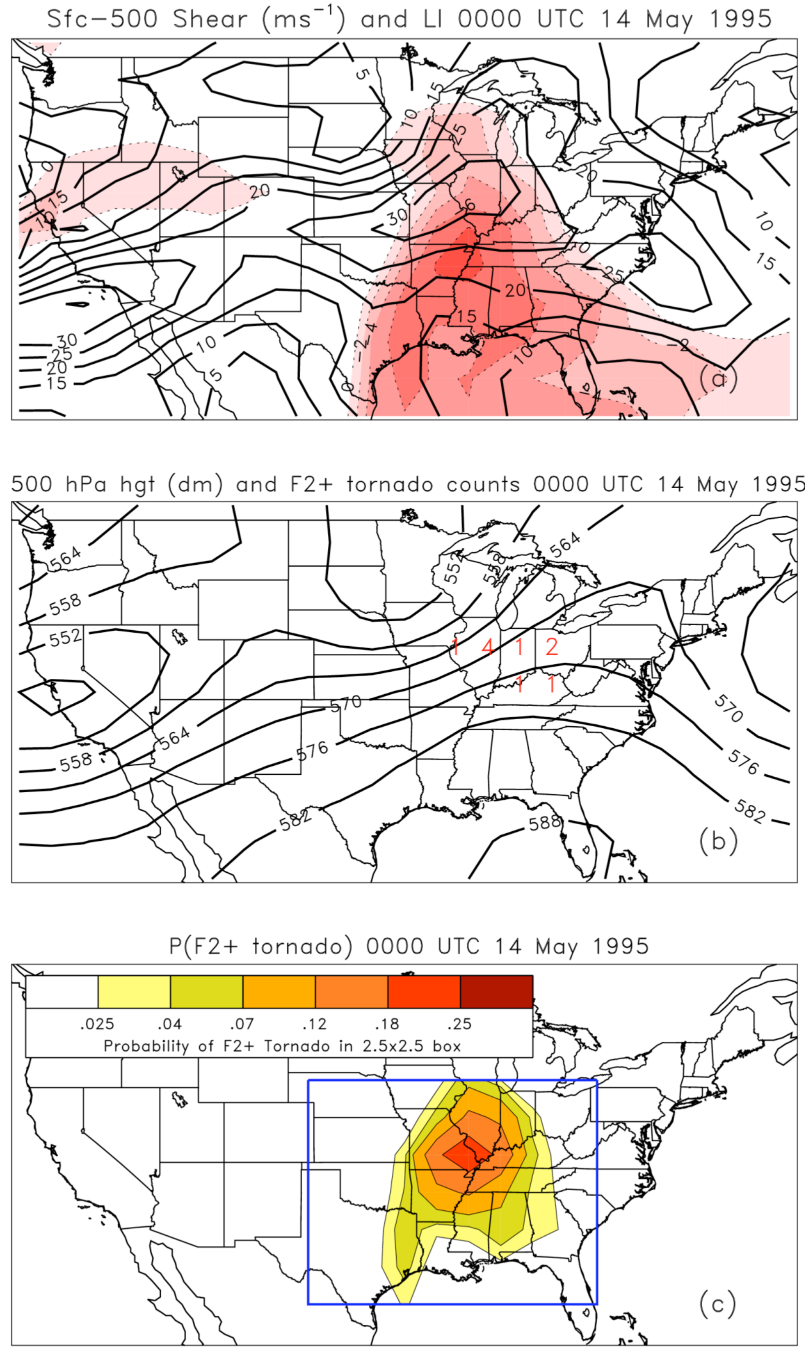
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 4 May 1999



P(F2+ tornado) 0000 UTC 4 May 1999




Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 14 May 1995
1 4 1 2
1
(a)
Tornado Probabilities for
01-day Forecast from 13 May 1995
.01 .02 .04 .06 .08 .10 .13
Probability of F2+ Tornado in 2.5x2.5 box
(b)


Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 14 May 1995
(a)
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 14 May 1995
564
558
552
558
564
570
576
582
588
1 4 1 2
1 1
(b)
P(F2+ tornado) 0000 UTC 14 May 1995
.025 .04 .07 .12 .18 .25
Probability of F2+ Tornado in 2.5x2.5 box
(c)

Making probabilistic forecasts from analogs

- **Method 1**: Use raw relative frequency of observed tornado occurrence in *n* analogs

$$P(T) = \frac{1}{n}\sum_{i=1}^{n} T_i$$

T_i = 1 if F2+ occurred,
T_i = 0 if no F2+ occurred

- **Method 2**: Use *weighted* relative frequency of observed tornado occurrence in *n* analogs

$$P(T) = \frac{1}{n}\sum_{i=1}^{n} w_i\, T_i$$