NOAA Earth System
Research Laboratory

Statistical Post-Processing of Ensemble Weather Predictions

Tom Hamill

NOAA Earth System Research Lab

tom.hamill@noaa.gov

Statistical post-processing, holistic view

- Would like probability distribution ϕ of true state T (or samples thereof) given all available information, including today's ensemble forecast

$$\phi\left(T \,\middle|\, \mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}\right)$$

- $\mathbf{z}$ might be past observations and/or forecasts, output from other modeling systems, your intuition, etc.
- ϕ might be a field rather than a scalar

Statistical post-processing, in terms of specific rationales.

- **Rationale 1**: Infer large-sample probabilities from small ensemble, assuming ensemble is calibrated (truth and ensemble drawn from same distribution).
- **Rationale 2**: Remove biases, increase forecast reliability while preserving as much sharpness as possible. Guided by discrepancies between past observations and forecasts.
- **Rationale 3**: Predict probabilities of a variable that may not be forecast by the model using variables that are forecast by the model, e.g., tornado probability from CAPE, shear.
- Combinations of these three, or perhaps others.

Rationale 1:

Infer large-sample probabilities from small ensemble, assuming ensemble is calibrated

Perfect model, but finite ensemble size → unreliable probabilities

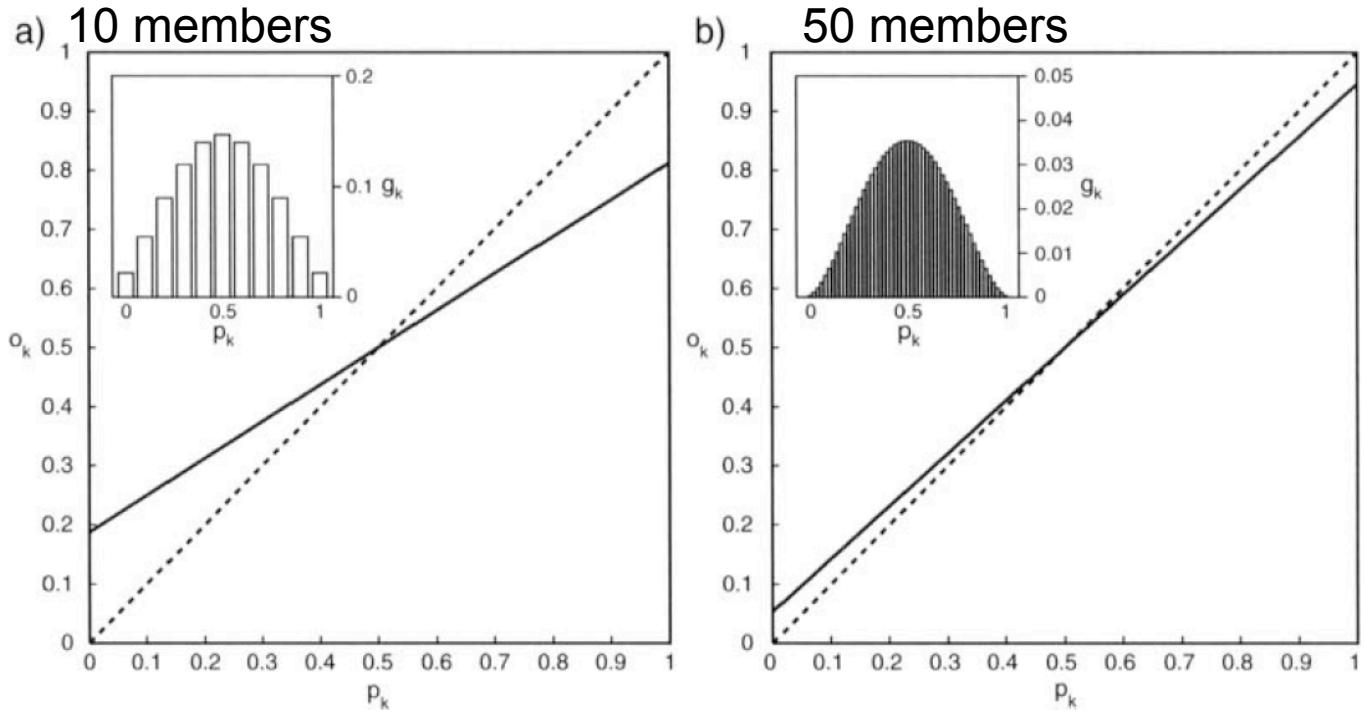
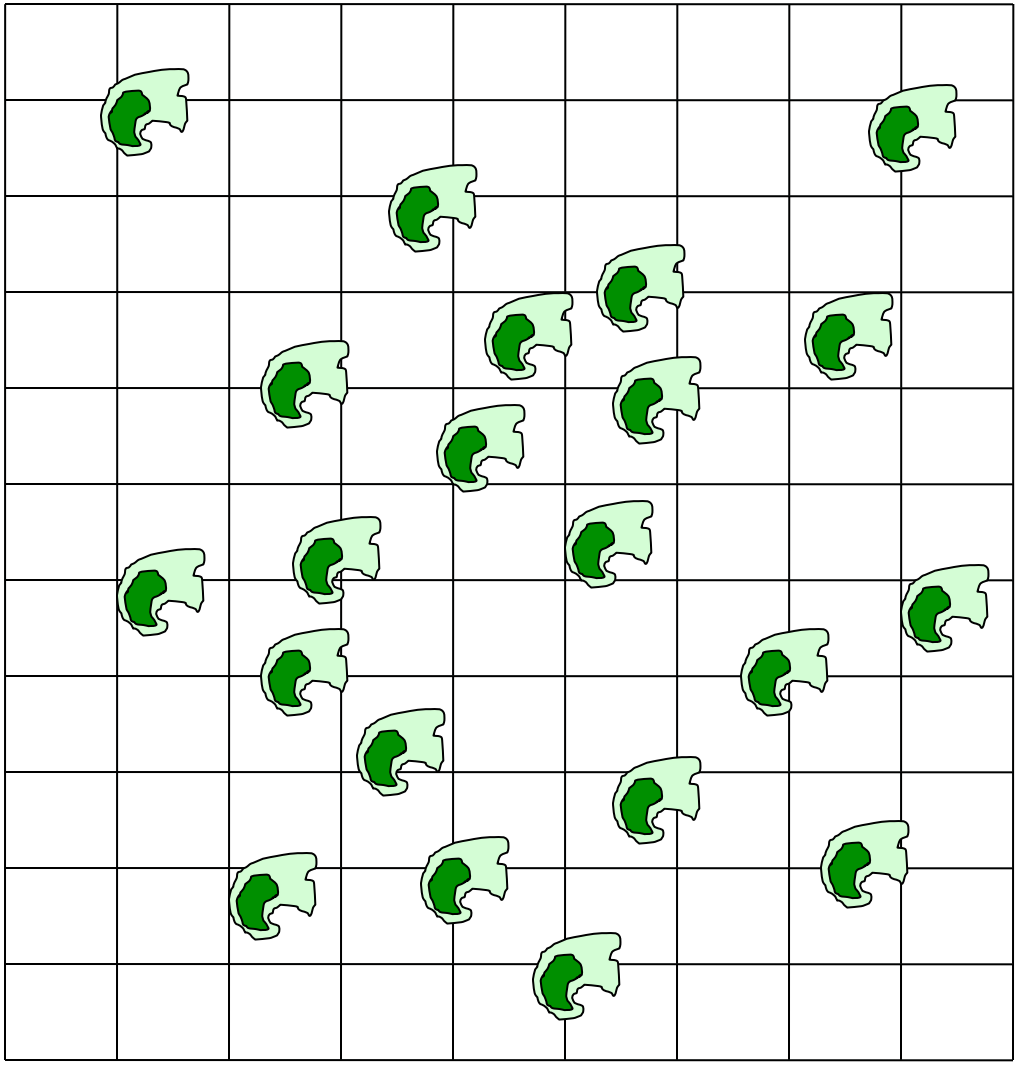


Figure 3. Reliability diagrams for theoretical ensemble forecasts for (a) a 10-member ensemble prediction system (EPS) and (b) a 50-member EPS. Distribution of underlying forecast probabilities is completely reliable and specified by a beta distribution with $r = s = 3$. See text for details and explanation of symbols.

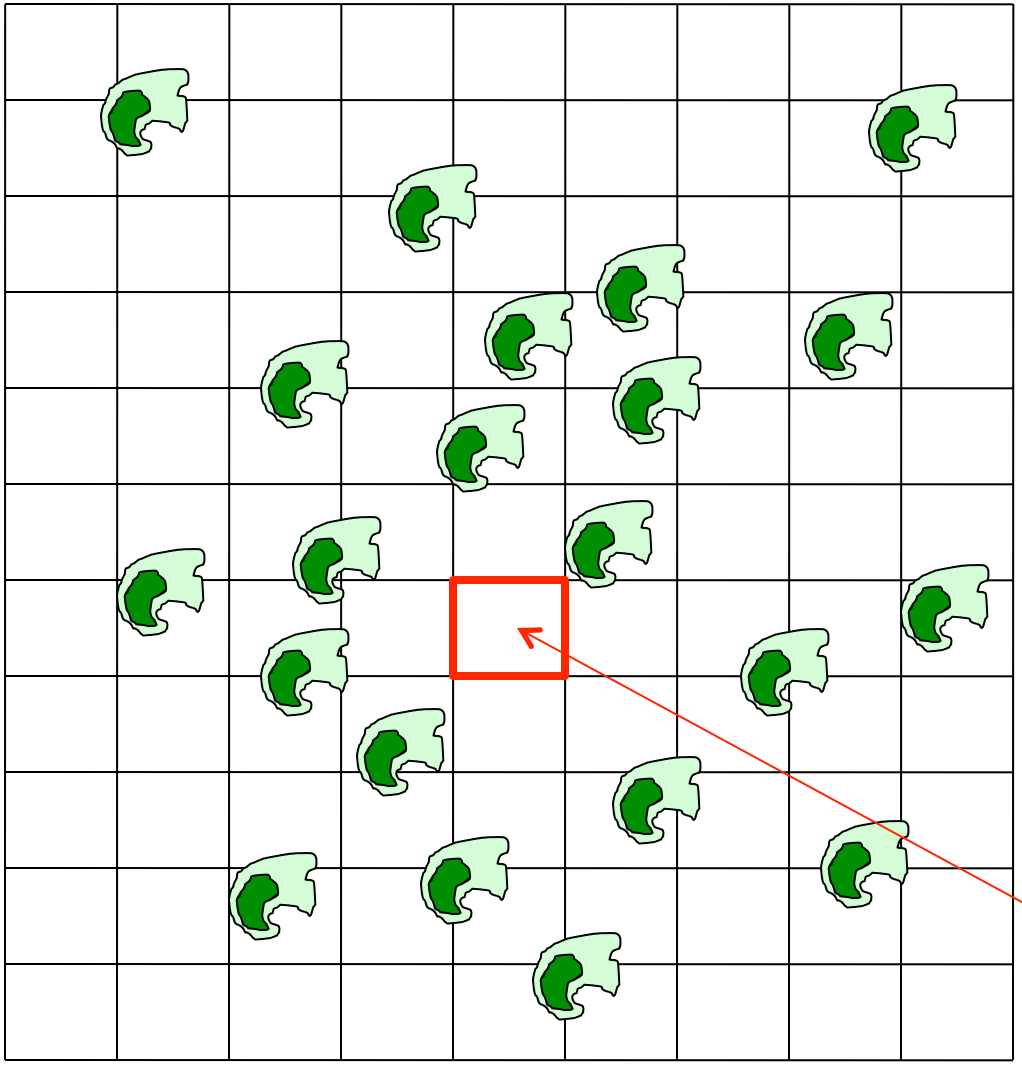
Probabilities directly estimated from an ensemble system where truth and ensemble members drawn from the same distribution. With only a few members, you may not produce reliable probabilities.

ref: Richardson 2001 QJRMS



Imagine: your 20-member storm-scale ensemble (which is a calibrated system, truth consistent with a random draw from ensemble)

Your job: estimate reliable probabilities on the grid.



Imagine: your 20-member storm-scale ensemble (which is a calibrated system, truth consistent with a random draw from ensemble)

Your job: estimate reliable probabilities on the grid.

Zero probability for this cell? Yes if you use ensemble relative frequency

There are **many** ways to estimate probabilities you'd get with a large sample, such as with kernel density estimation.

Kernel density estimation – Wikipedia, the free encyclopedia

http://en.wikipedia.org/wiki/Kernel_density_estimation

kernel density estimation tricube

Home Bookmarks AMS MTS TinyURL ACUF ESRL Library NOAA Directory NCARPeople NOAA Web Cal PSD Wiki PSD Intranet TC Tracker HFIP Stamps Comcast PSD Activities Google Calendar GFS EnKF

New features Log in / create account

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate

Interaction
About Wikipedia
Community portal
Recent changes
Contact Wikipedia
Help

Toolbox

Print/export

Languages
Deutsch
Français
日本語
中文

Article Discussion Read Edit View history Search

Kernel density estimation

From Wikipedia, the free encyclopedia

It has been suggested that *Multivariate kernel density estimation* be merged into this article or section. (Discuss)

In statistics, **kernel density estimation** is a non-parametric way of estimating the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample. It is also known as the **Parzen-Rosenblatt window** method, named after Emanuel Parzen and Murray Rosenblatt, who are usually credited with independently creating it in its current form,[1][2] though it appears that this latter nomenclature is more common amongst certain specialists e.g. signal processing, econometrics. Within statistical and mathematical sciences, kernel density estimation is more prevalent.

Contents [hide]

1 Definition
1.1 Relation to the characteristic function density estimator
2 Bandwidth selection
3 Statistical implementation
3.1 Example in R
3.2 Example in Matlab/octave
4 See also
5 External links
6 References

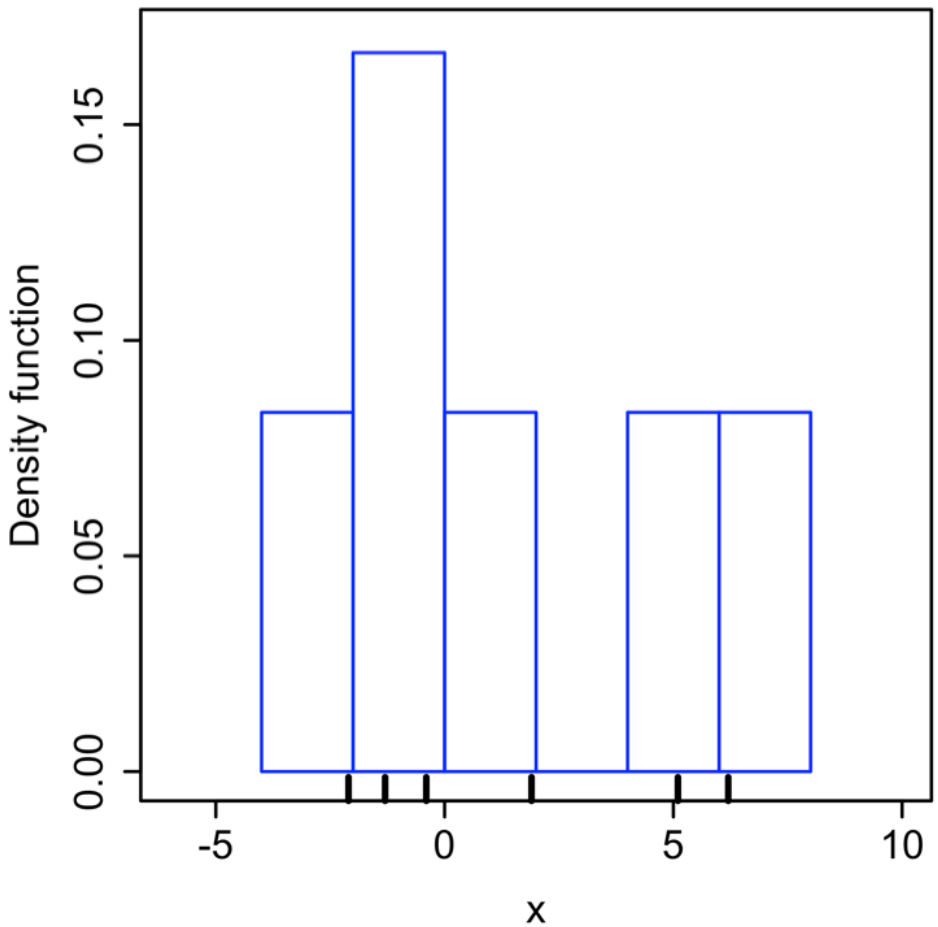
Definition [edit]

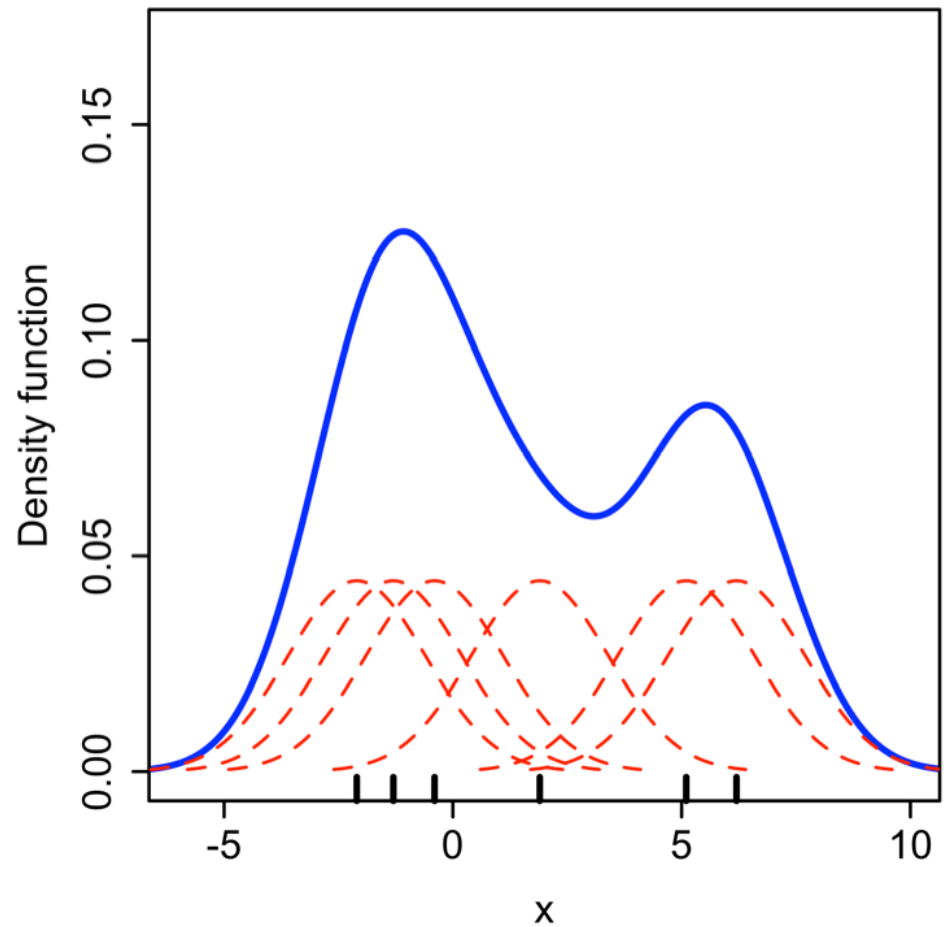
Let $(x_1, x_2, \ldots, x_n)$ be an iid sample drawn from some distribution with an unknown density f. We are interested in estimating the shape of this function f. Its **kernel density estimator** is

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) \quad = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),$$

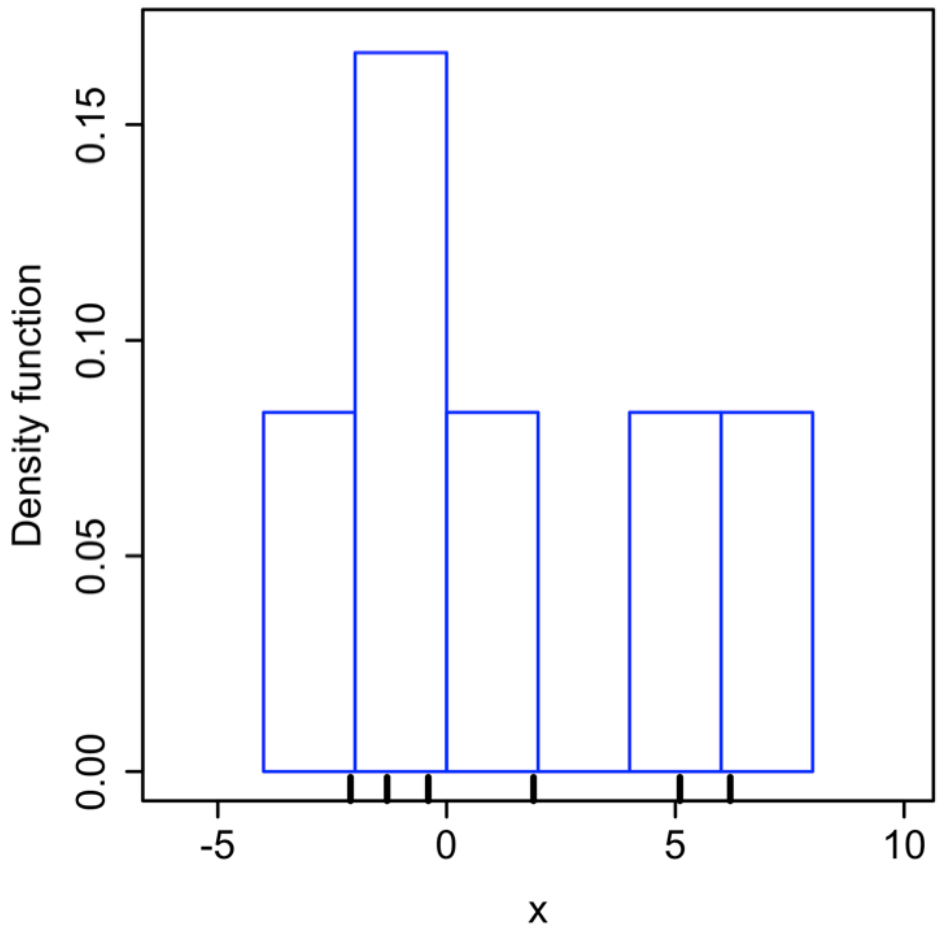
where $K(\cdot)$ is the kernel — a symmetric but not necessarily positive function that integrates to one — and $h > 0$ is a smoothing parameter called the **bandwidth**. A kernel with subscript h is called the *scaled kernel* and defined as $K_h(x) = 1/h\, K(x/h)$. Intuitively one wants to choose h as small as the data allows, however there is always a tradeoff between the bias of the estimator and its variance; more on the choice of bandwidth later. A range of kernel functions are commonly used: uniform, triangular, biweight, triweight, Epanechnikov, normal, and others. The Epanechnikov kernel is optimal in a minimum variance sense,[3] though the loss of efficiency is small for the kernels listed previously[4], and due to its convenient mathematical properties, the normal kernel is often used $K(x) = \phi(x)$, where ϕ is the standard normal density function.

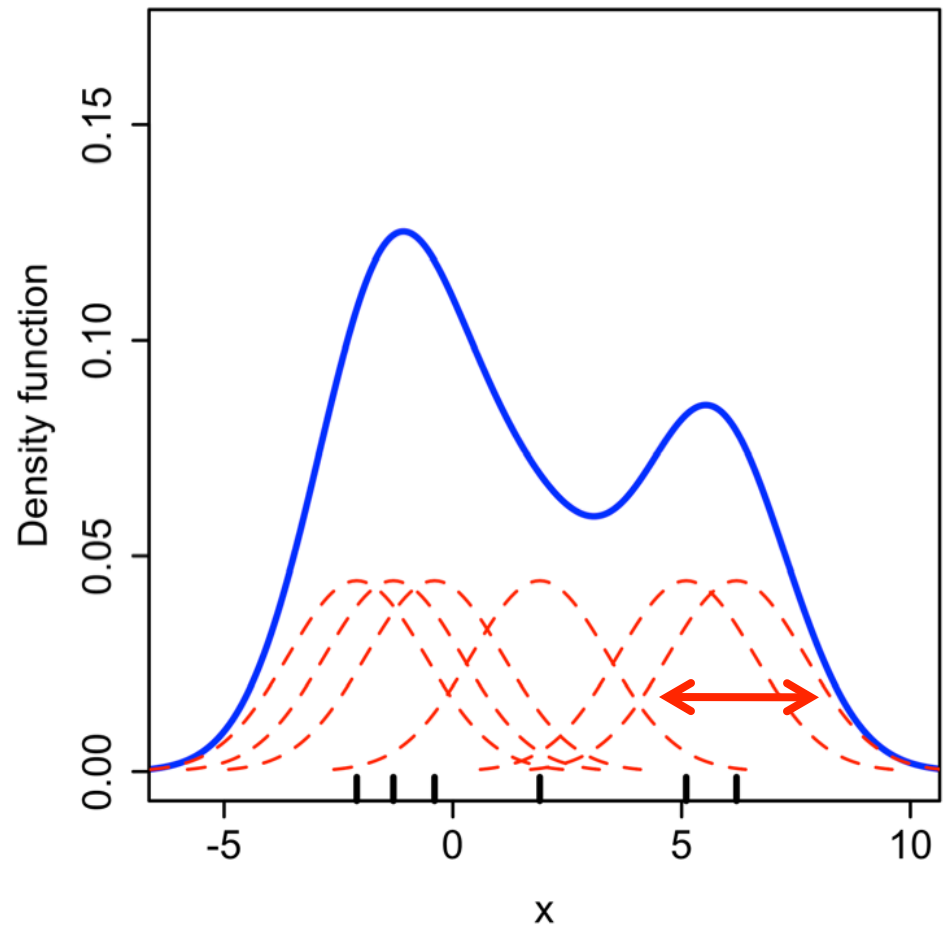
Density estimate from histogram and from kernel density estimation.





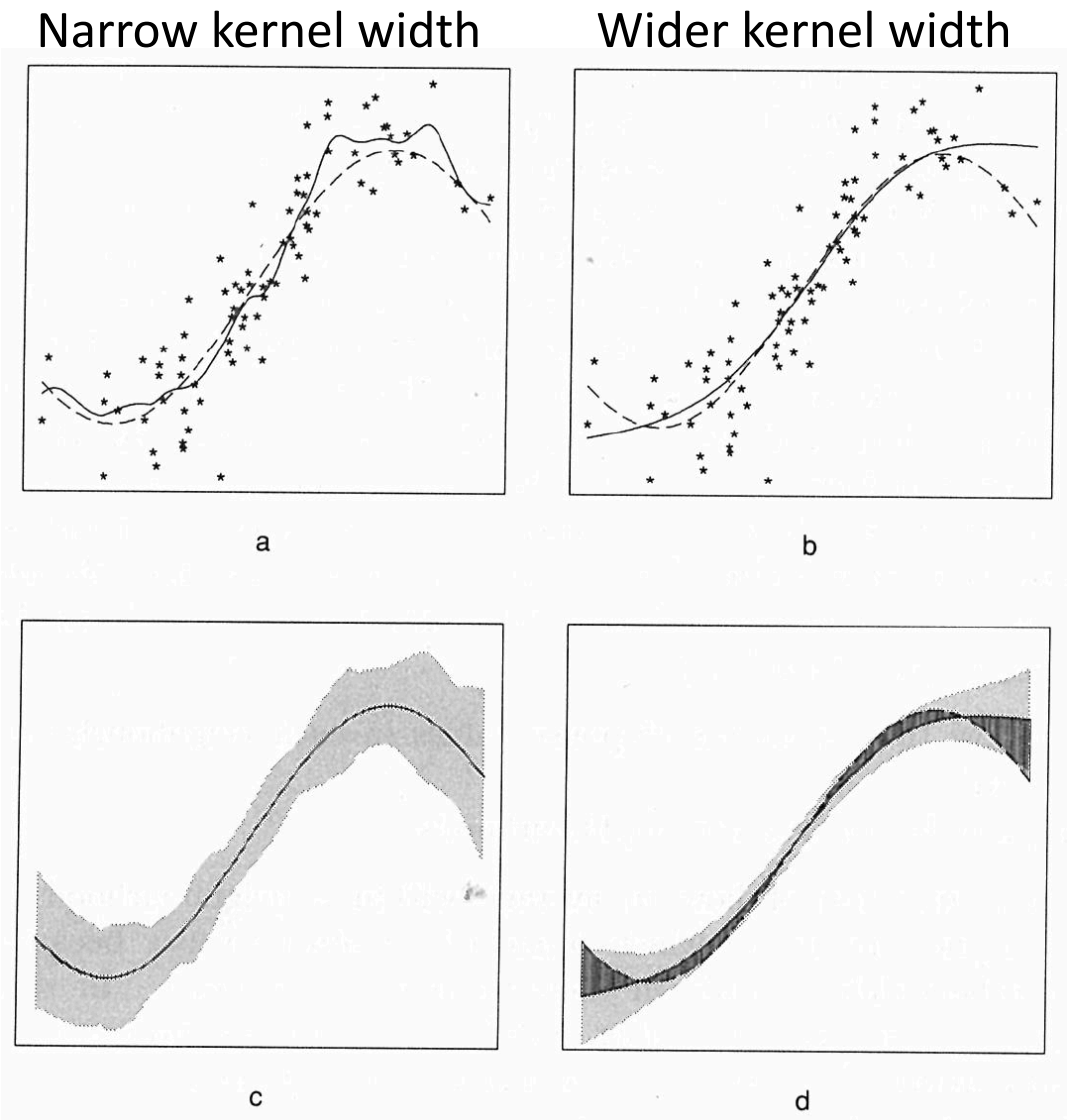
Density estimate from histogram and from kernel density estimation.





Key question: how do you set the "kernel width?"

“Bias-variance tradeoff”



- Example: 80 samples from $Y = \sin(X) + e$, where $X \sim N(0,1)$, $e \sim N(0,1/9)$.
- Large grey area shows twice the standardized pointwise errors, reflect the variance under repeated samplings of 80 pairs.
- Darker grey region shows bias, difference between generating curve and average value of the estimated function.

Ref: Hastie and Tibshirani, “*Generalized Additive Models*” (an excellent book)

More on "bias-variance tradeoff"

- Same dilemma in the decision of how many predictors to use in a linear regression.
- Need to select some criteria to decide the optimal tradeoff. Perhaps to minimize mean squared error (MSE)?

$$E\left\{\hat{f}_w(x_i) - f(x_i)\right\}^2 = \mathrm{var}\left\{\hat{f}_w(x_i)\right\} + \left[E\hat{f}_w(x_i) - f(x_i)\right]^2$$

estimated value at data point x_i

actual value at data point x_i

variance of estimate value, perhaps by "jackknife"

squared bias of estimated value

Ref: ibid.

2-D example of inferring large-sample pdf from small-sized ensemble

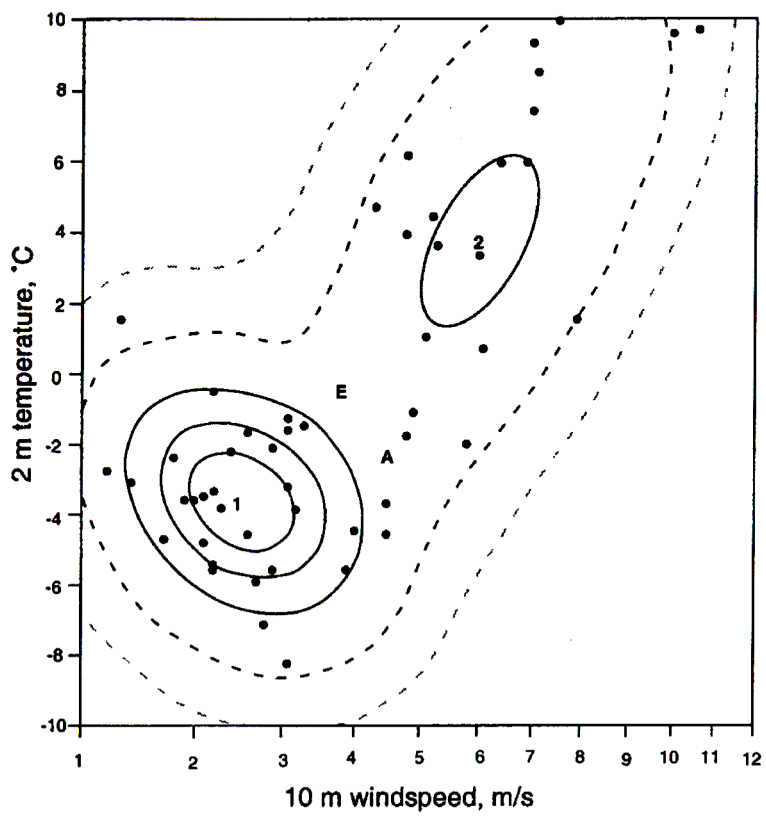


Figure 2. Example ensemble distribution with fitted Gaussian mixture, jointly for the temperature and windspeed forecast at 12 UTC 10 January 1997 at Manchester, made at the 180 h lead time. Dots indicate individual forecasts made by the 51 ensemble members, with the ensemble mean located at 'E'. The two bivariate Gaussian densities $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are centred at '1' and '2', respectively, and the smooth lines indicate level curves of their mixture $f_{\text{mix}}(\mathbf{x})$, formed with $\alpha = 0.57$ (see text). Contour interval is 0.05, and the thick and thin dashed lines are for 0.01 and 0.001, respectively. Subsequent verifying analysis is 'A'.

Wilks (*QJRMS*, **128**, p 2821) explored fitting parametric distributions, or mixtures thereof, to ECMWF forecasts in perfect-model context. Power-transformed non - Gaussian variables prior to fitting.

Inferring large-sample pdf from small-sized ensemble

kernels of probability density around each sample, pdf constructed from sum of these

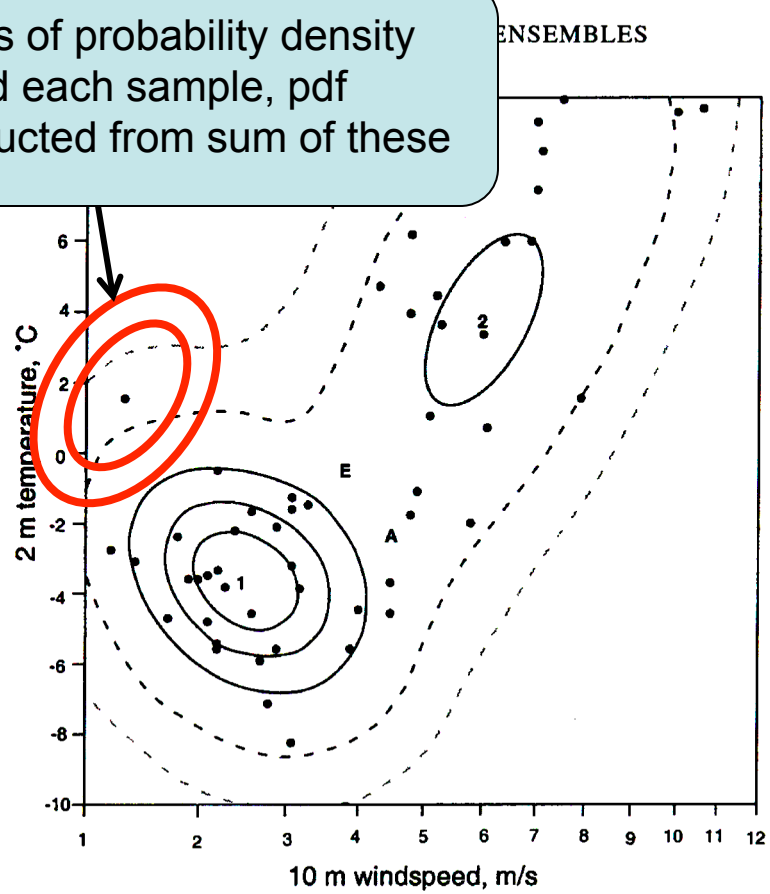
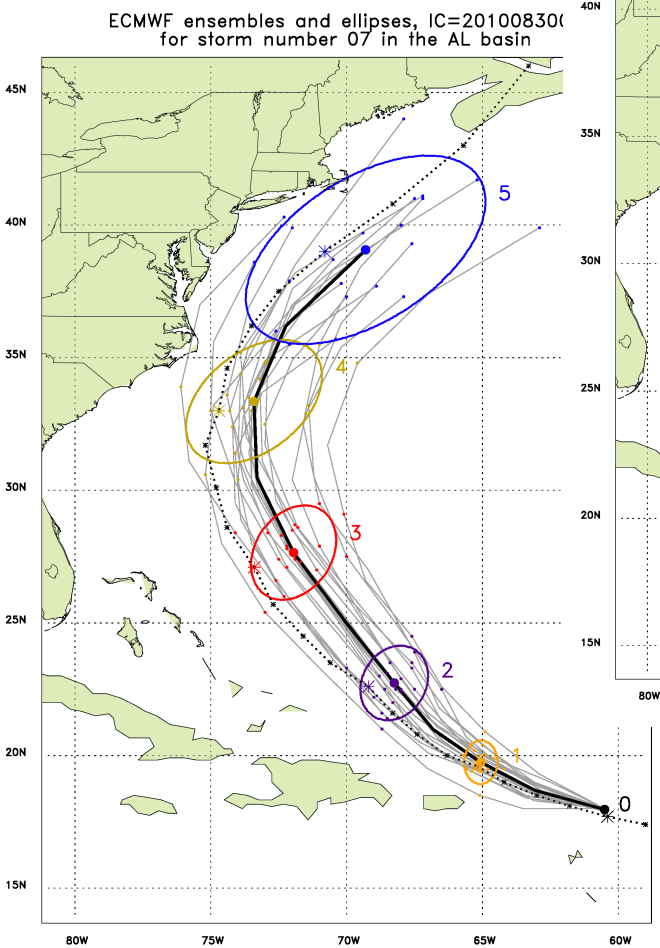
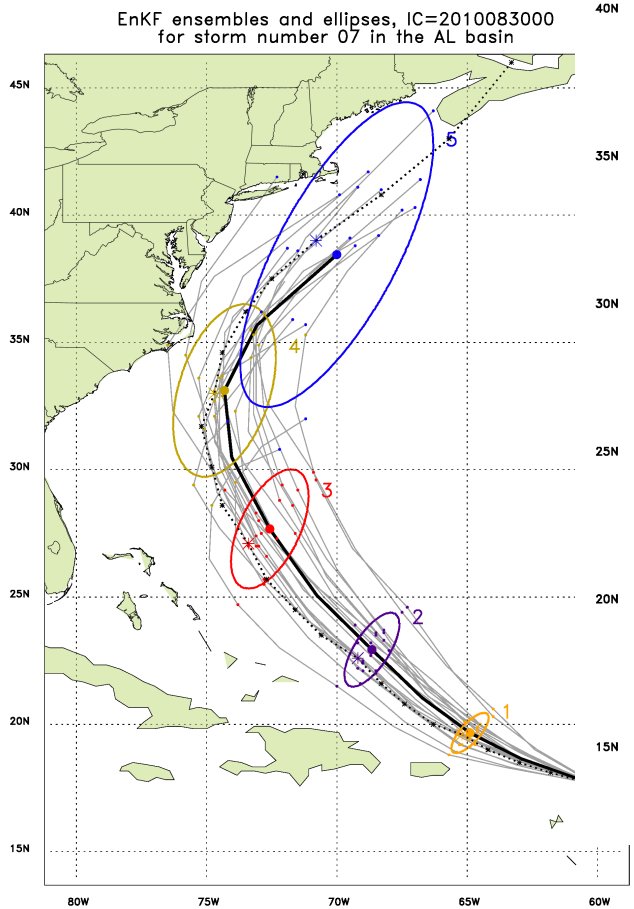


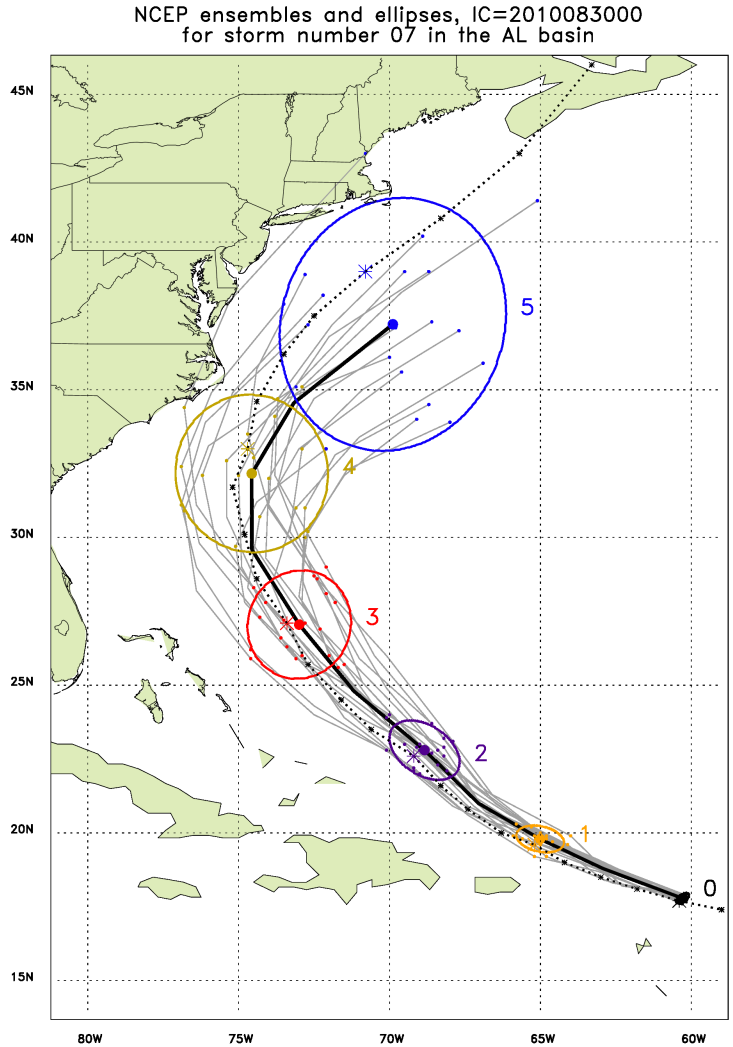
Figure 2. Example ensemble distribution with fitted Gaussian mixture, jointly for the temperature and wind-speed forecast at 12 UTC 10 January 1997 at Manchester, made at the 180 h lead time. Dots indicate individual forecasts made by the 51 ensemble members, with the ensemble mean located at 'E'. The two bivariate Gaussian densities $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are centred at '1' and '2', respectively, and the smooth lines indicate level curves of their mixture $f_{\text{mix}}(\mathbf{x})$, formed with $\alpha = 0.57$ (see text). Contour interval is 0.05, and the thick and thin dashed lines are for 0.01 and 0.001, respectively. Subsequent verifying analysis is 'A'.

Wilks (*QJRMS*, **128**, p 2821) explored fitting parametric distributions, or mixtures thereof, to ECMWF forecasts in perfect-model context. Power-transformed non - Gaussian variables prior to fitting.

Earl, 2010/08/30/00Z







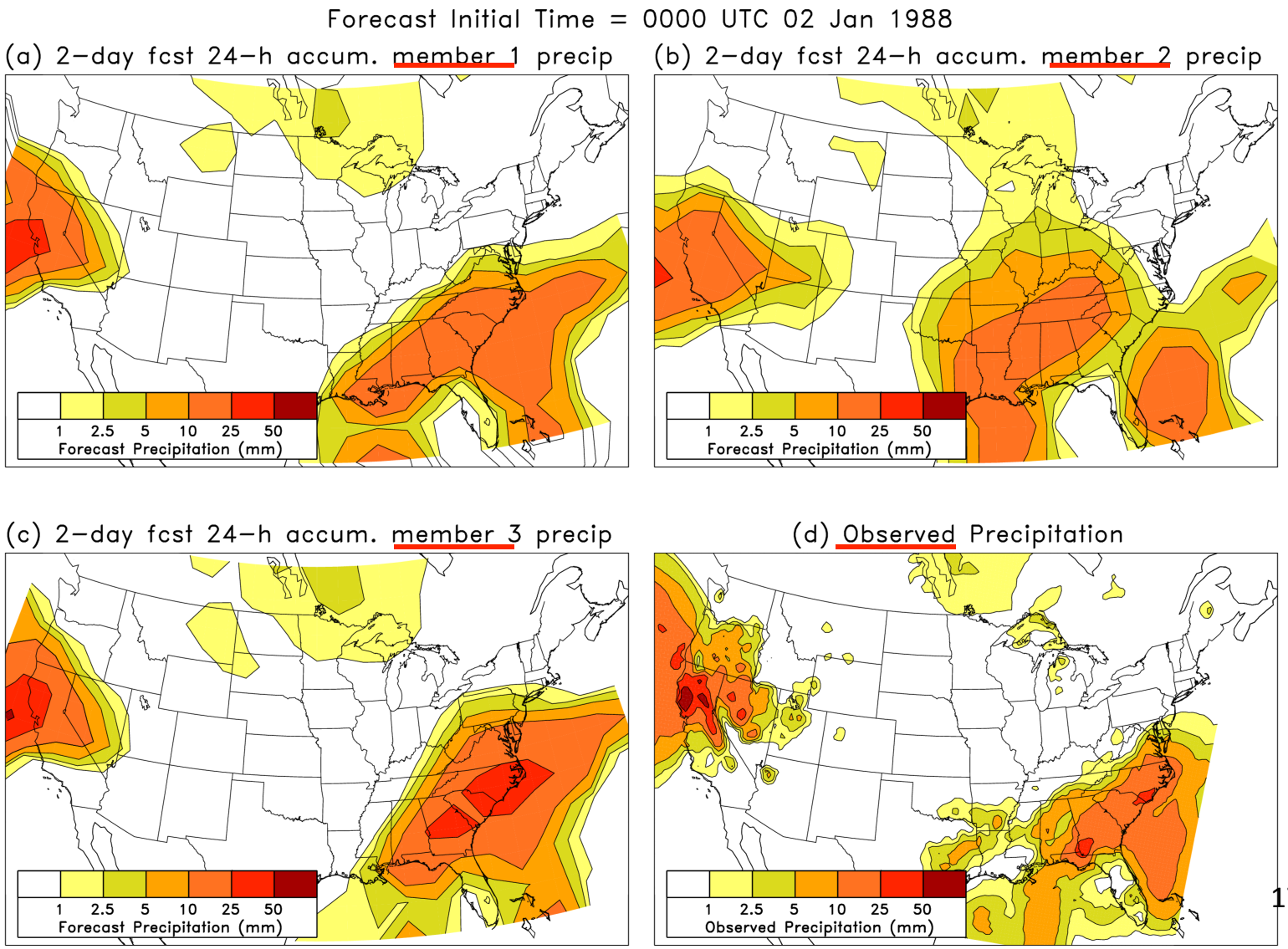
ellipses from a bivariate normal distribution fit to the ensemble data. Contours enclose 90% of the fitted probability.

Rationale 2:

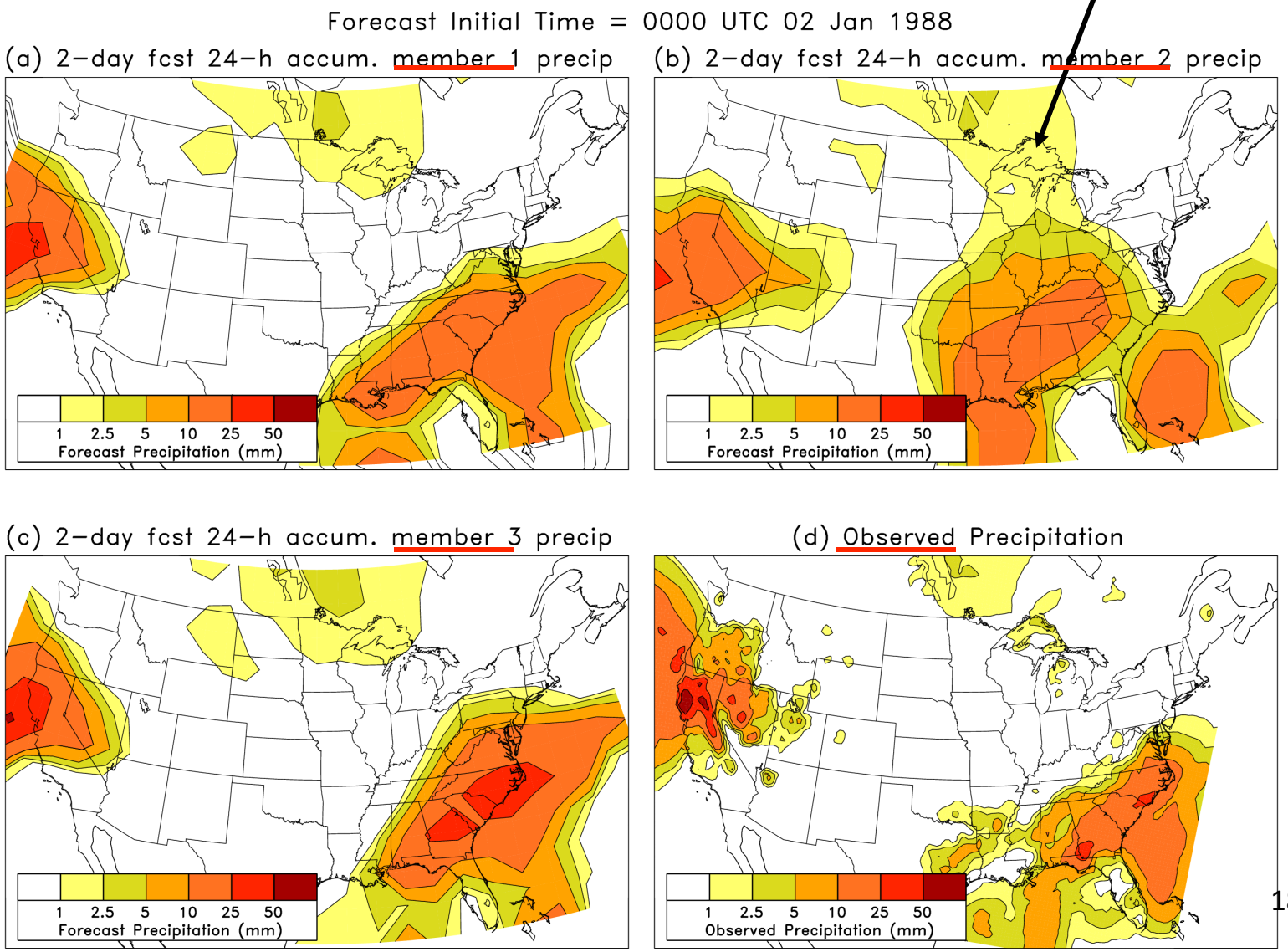
Remove biases, increase forecast reliability while preserving as much sharpness as possible. Guided by discrepancies between past observations and forecasts.

[let's call this "post-processing" now]

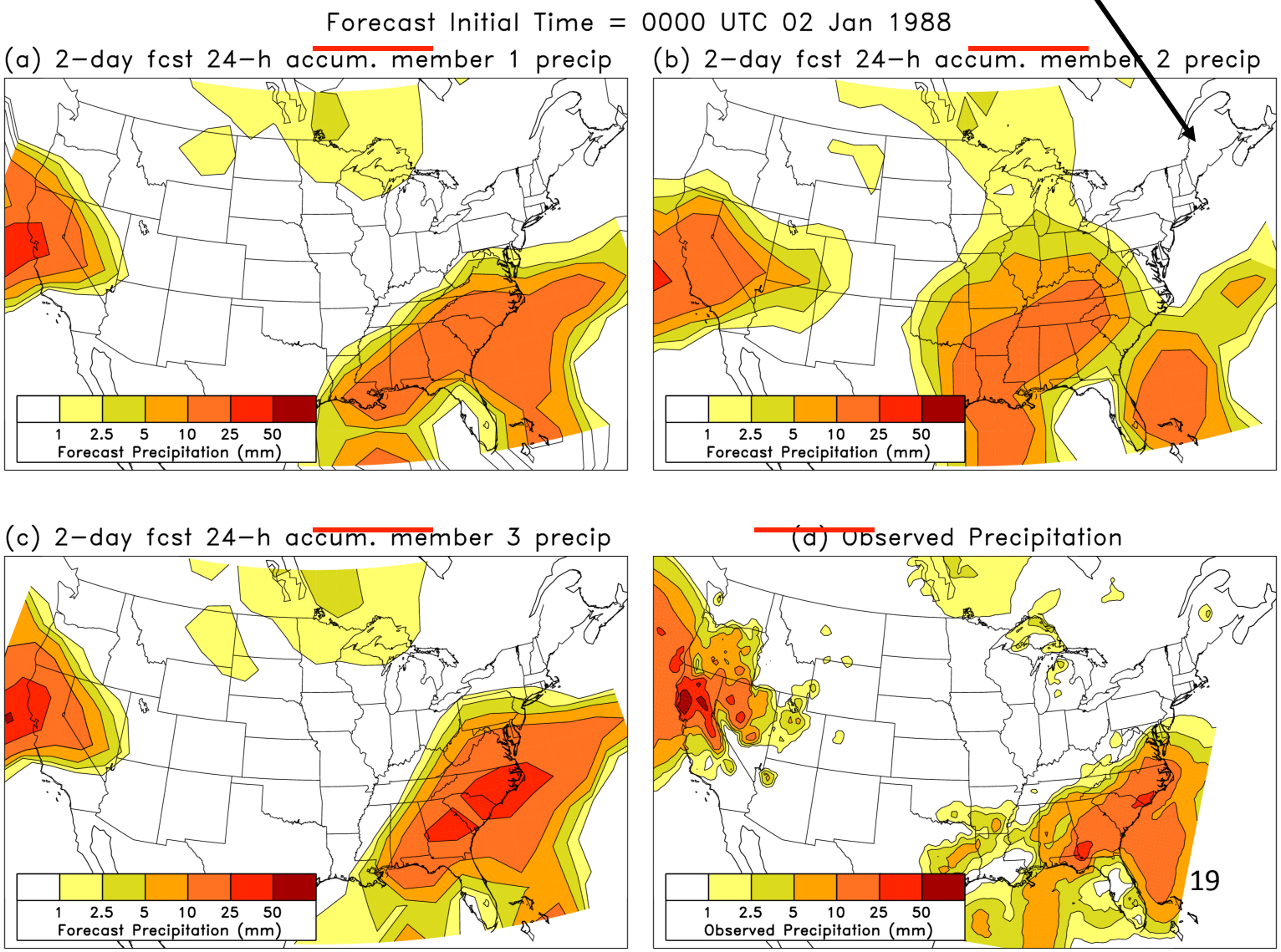
Ensemble-base probabilistic forecasts: problems we'd like to correct through post-processing



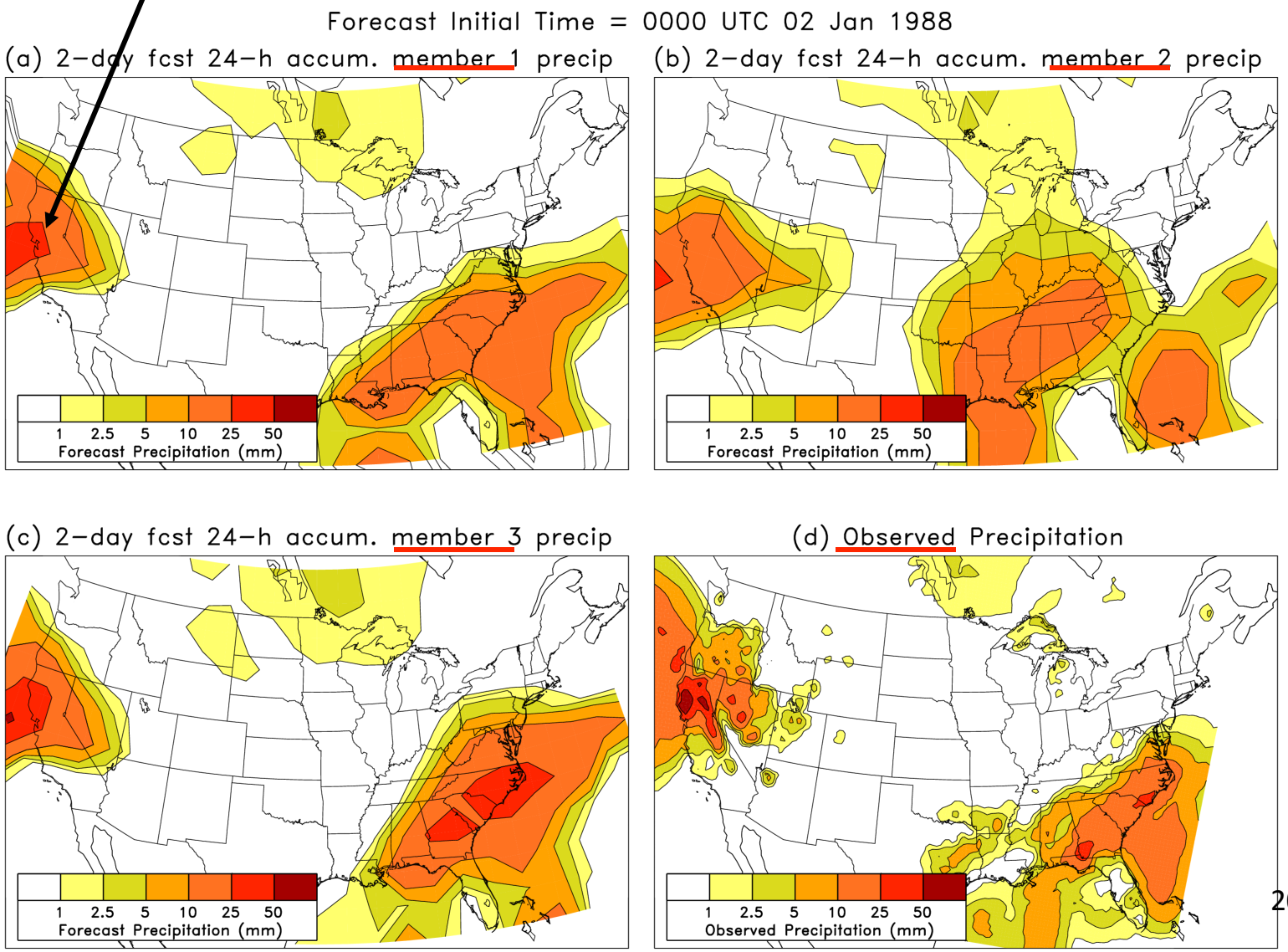
(1) possible conditional bias
(drizzle over-forecast)



(2) ensemble members too similar to each other (or ensemble size too small).

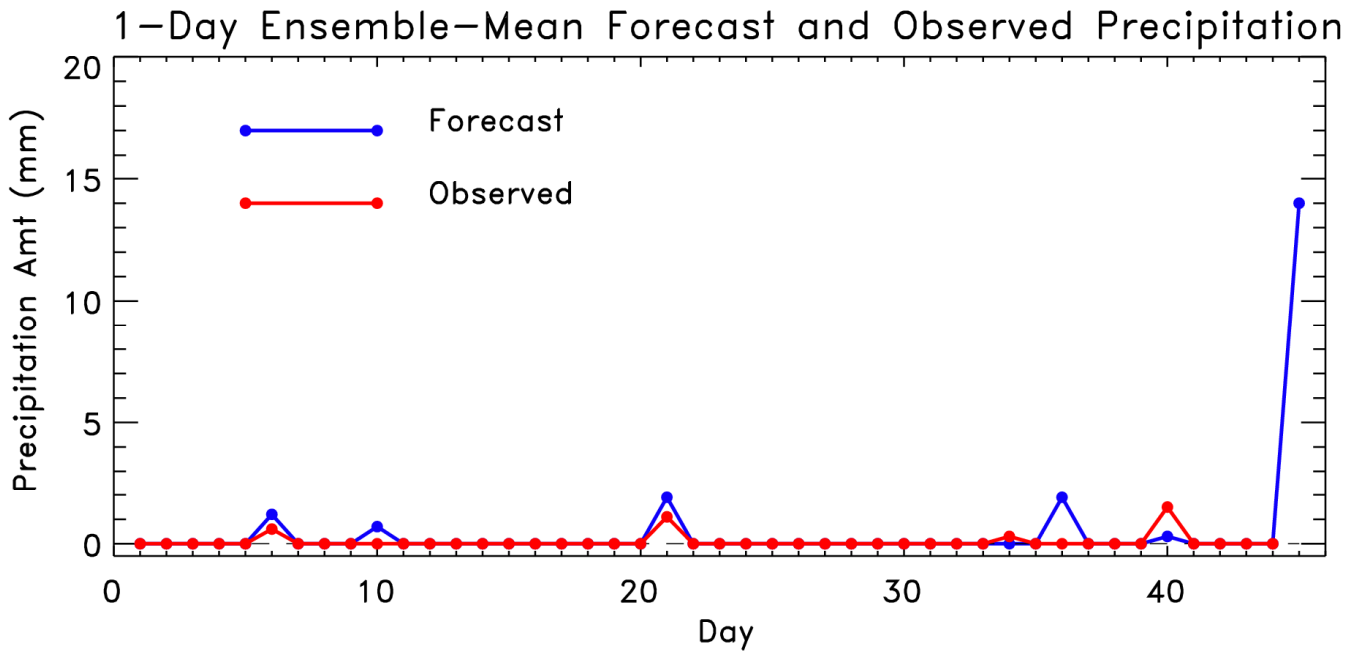


(3) Ensembles are too smooth, not capturing intense local precipitation due to orographic forcing. *Downscaling* needed.



Post-processing of PQPF & rare events: importance of sample size

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.



(this has motivated our exploration of reforecasts)

Key post-processing questions

- Is there a best technique, or best for this particular forecast problem?
- How much training data (past forecasts & obs/ analyses) do you have / need?
  - Generally, more needed to do good job with rare events, longer-lead forecasts where chaotic noise is large.
- Tradeoff between the two:
  - Lots more algorithm development involved in trying to get a good result with a short training data set, but less computations required.

Disadvantages to post-processing

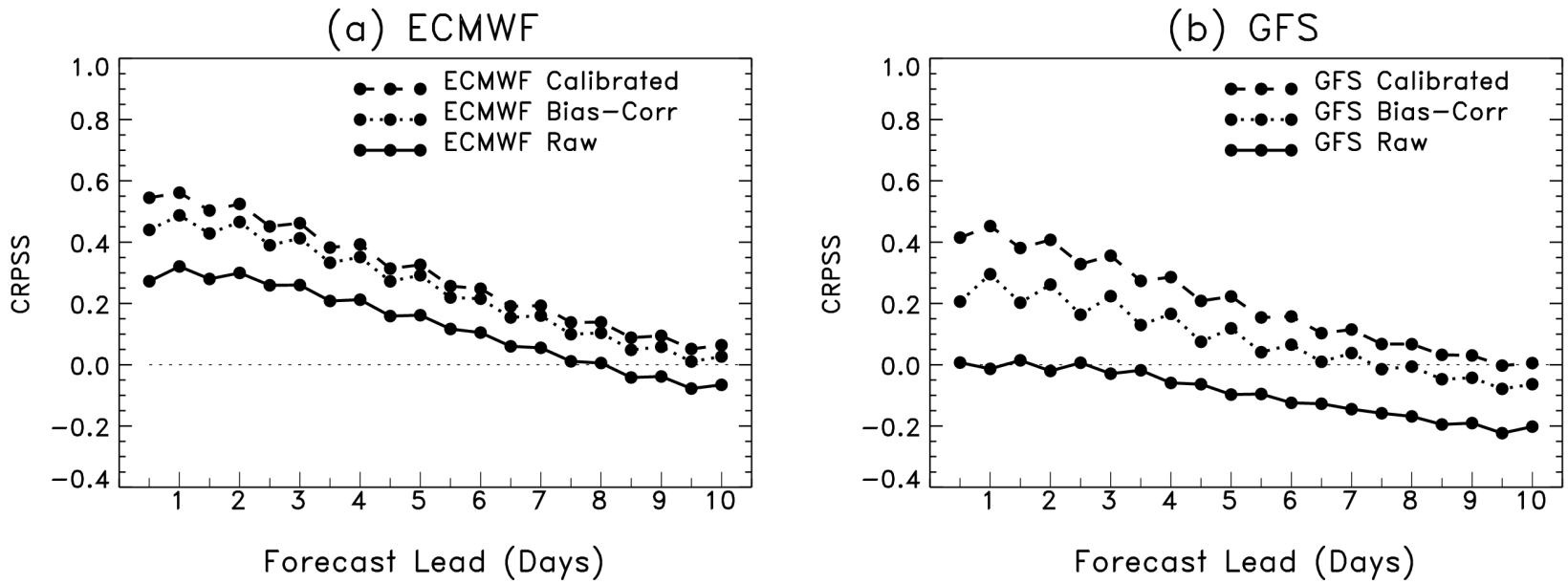
- Post-processing won't correct the underlying problem. Prefer to achieve unbiased, reliable forecasts by doing numerical modeling correctly in the first place.
- No one general approach that works best for all applications.
- Corrections may be model-specific; the calibrations for NCEP v 2.0 may not be useful for ECMWF, or even NCEP v 3.0.
- Simple techniques applied to date mostly focus on point-by-point post-processing, may lose spatial correlation information of original ensemble.
- Could constrain model development. Calibration ideally based on long database of prior forecasts (reforecasts) from same model. Upgrading model good for improving raw forecasts, may be bad for skill of post-processed forecasts.
- Users beware: Several calibration techniques that have been recently proposed may only work properly in certain circumstances. Can't use naively.

A quick survey of common post-processing techniques

- Simpler methods
  - Gross bias correction
  - Kalman-inspired filters
  - CDF-based bias corrections
  - Linear regression
- Some more complex methods
  - Logistic regression
  - Analog approach
  - Bayesian model averaging (MDL's EKDMOS very similar)
  - Bayesian processor of forecasts
  - Non-homogeneous Gaussian regression
  - Rank histogram-based calibration

Gross bias correction

- Given sample of past forecasts $x_1, \ldots, x_n$ and observations $y_1, \ldots, y_n$, gross bias correction is simply $\overline{y} - \overline{x}$



In surface-temperature calibration experiments with NCEP's GFS and ECMWF, simple gross bias correction achieved a large percentage of the improvement that was achieved through more sophisticated, bias+spread correction.

Ref: Hagedorn et al., MWR, 2008, in press.

Gross bias correction

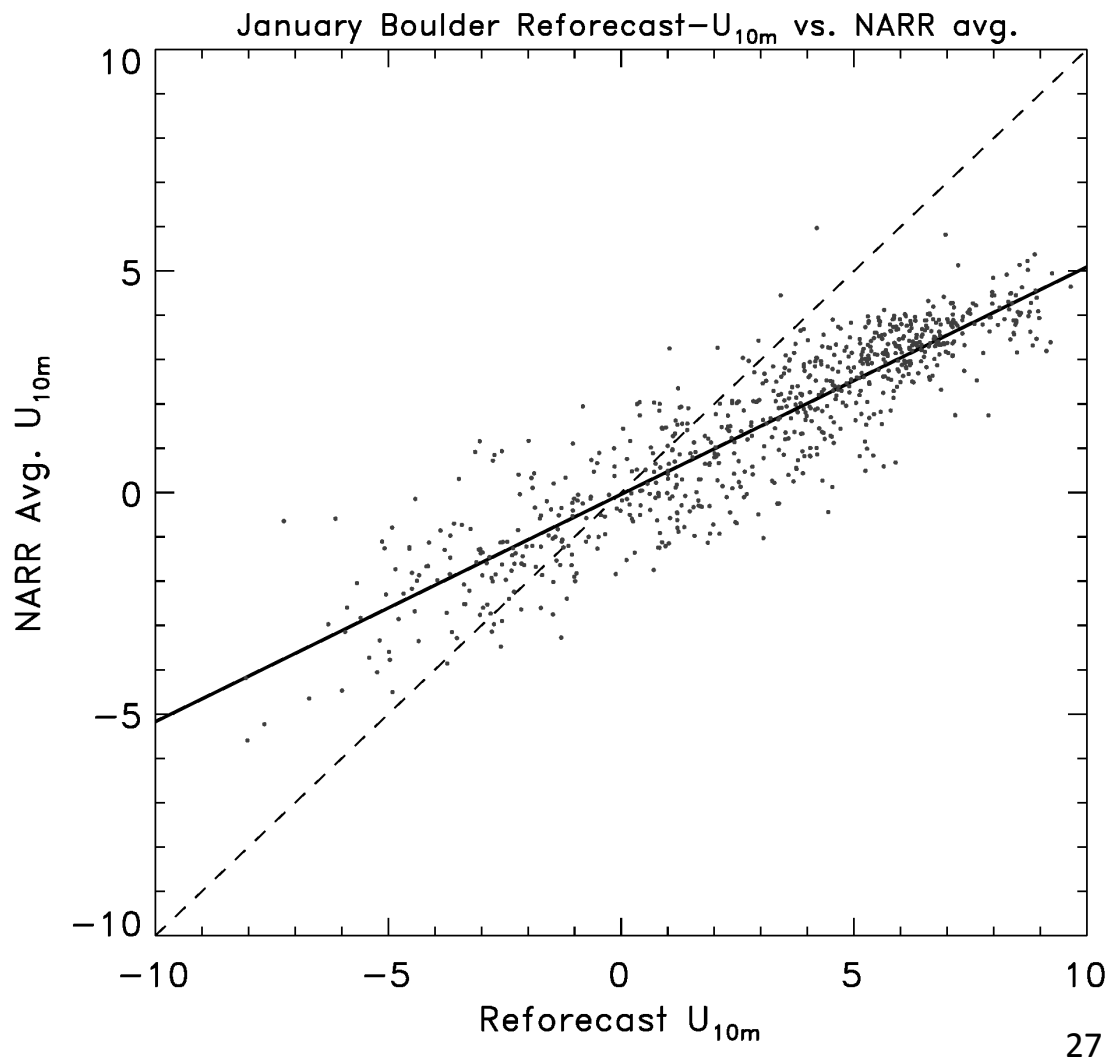
- Effectively, the implied statistical model is the following:

$$Y_i = \beta + X_i + \varepsilon_i$$

  – assumes normality of errors; uncorrelated errors, error not *state dependent* (next slide).

State-dependent errors

For this 10-m wind, the bias is conditional, depends on the forecast amount. Linear regression (discussed later) a much better choice.



"Kalman-inspired" filter

Today's forecast bias estimate

Yesterday's bias estimate

Yesterday's observed bias

$$\hat{b}_t^f = \hat{b}_{t-1}^f + K_t\left(\varepsilon_t - \hat{b}_{t-1}^f\right)$$

Kalman gain: weighting applied to residual. Larger K_t, more weight to recent data, and vice versa.

Pro:

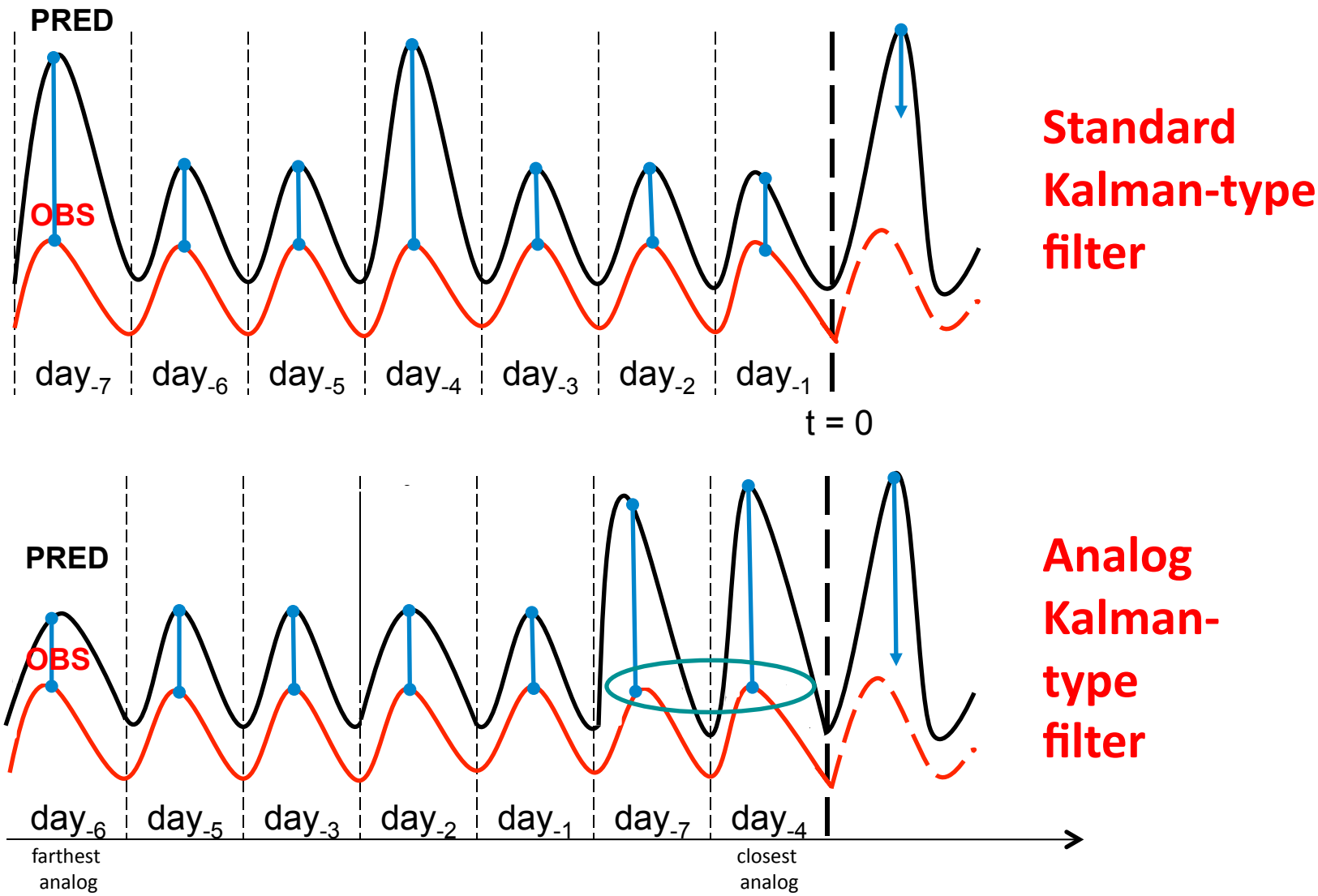
- memory in system, amount tunable through K_t
- adaptive

Con:

-assuming there is state-dependent bias, takes time to adapt after regime change and change of state.

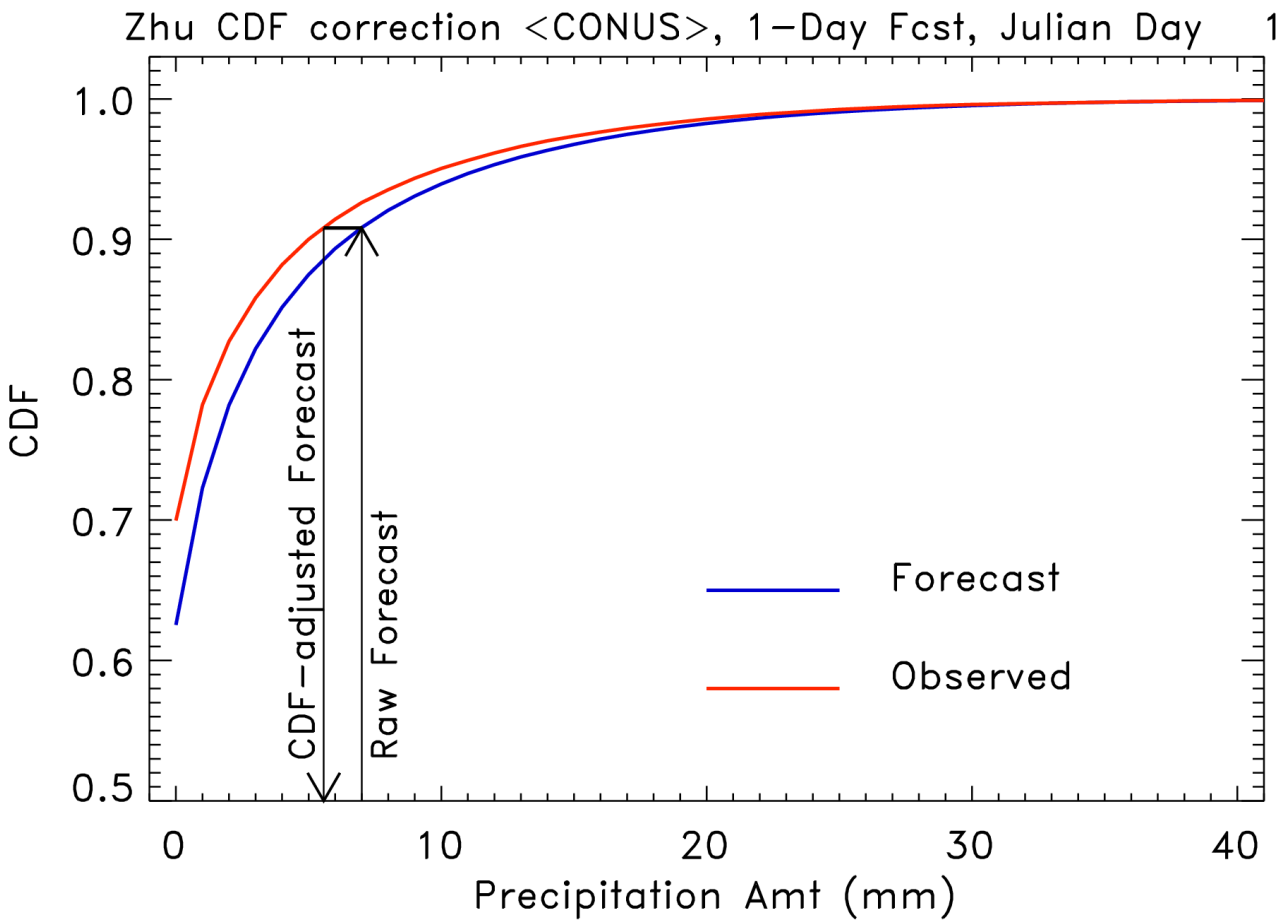
Ref: Cheng and Steenburg, conferences.dri.edu/WxPrediction/Weather12/Cheng_Steenburgh.ppt

An alternative "analog" formulation



c/o Luca delle Monache, NCAR/RAL, MWR, submitted.

CDF-based bias corrections



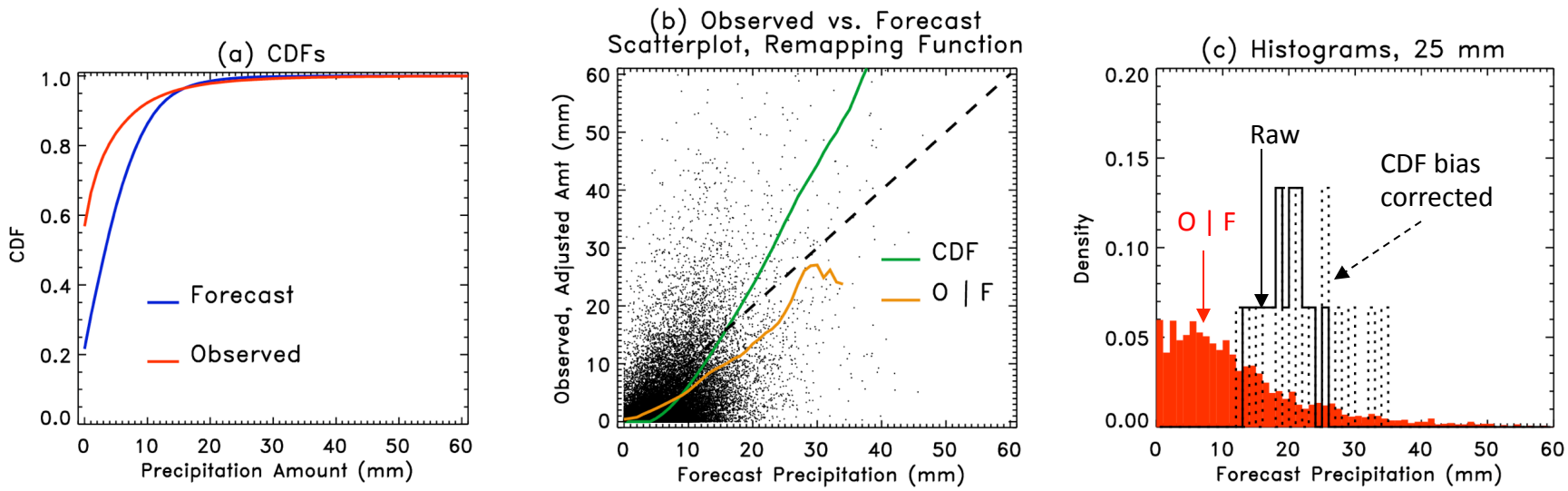
Use difference in CDFs to correct each ensemble member's forecast. In example shown, raw 7-mm forecast corrected to ~5.6 mm forecast.

NOTE: bias only, not spread correction or downscaling.

Ref: Zhu and Toth, 2005 AMS Annual Conf., and many others

CDF corrections: example of problem

1-day forecasts in Northern Mississippi (US), mid-August.
Consider a forecast precipitation of 25 mm.

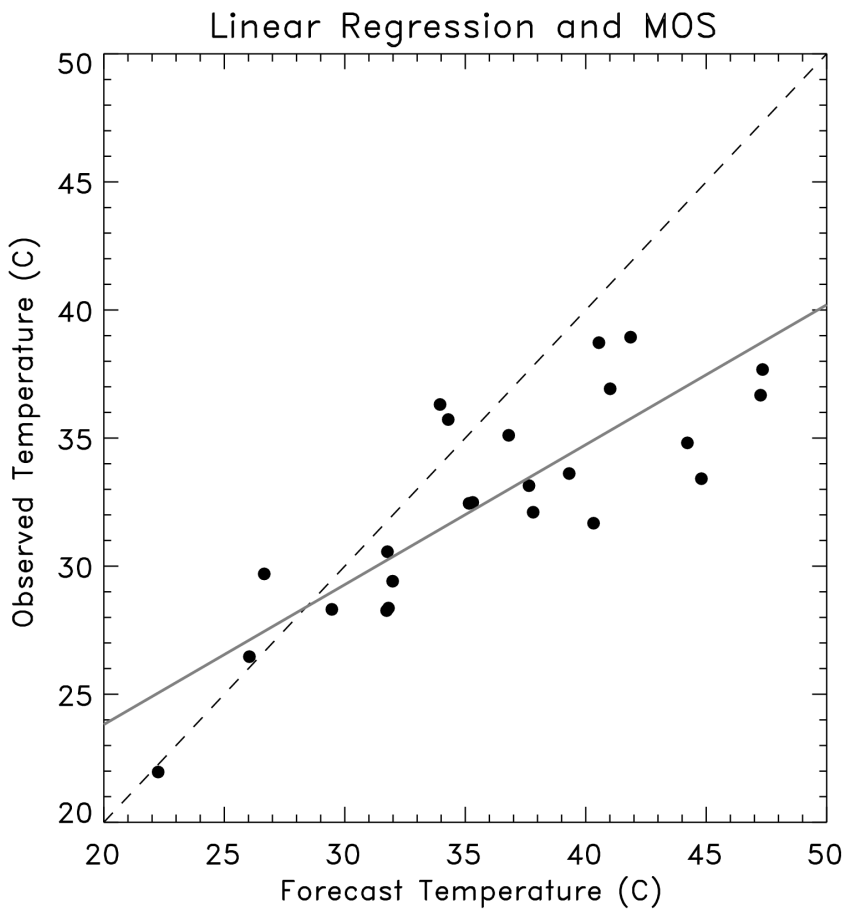


CDF-based corrections at high amounts suggest further increasing precipitation amount forecast. O|F indicates decrease.

At root of problem is assumption that Corr (F,O) ≅ 1.0

Linear regression

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \ldots + \beta_N X_{i,2} + \varepsilon_i$$



Corrects for state-dependent bias; when no predictive skill of forecast, regresses to observed sample climatology.

Diagnostics include statistics on error, so can infer (largely non-state dependent) pdf.

Ref: any applied statistics textbook

When is linear regression approach useful?

- Some assumptions:
  - Normality of errors.
  - Linear relation between predictors and predictand.
  - Homoscedasticity, error variance doesn't depend on state x.
  - Errors are uncorrelated between samples.

Ref: any applied stats text, e.g., Neter, Wasserman, and Kutner, Applied Linear Statistical Models.

When is linear regression approach useful?

- Some assumptions:
  - Normality of errors.
  - Linear relation between predictors and predictand.
  - Homoscedasticity, error variance doesn't depend on state x.
  - Errors are uncorrelated between samples.

Well, there goes using this for precipitation!

Ref: any applied stats text, e.g., Neter, Wasserman, and Kutner, *Applied Linear Statistical Models*.

When is linear regression approach useful?

- Some assumptions:
  - Normality of errors.
  - Linear relation between predictors and predictand.
  - Homoscedasticity, error variance doesn't depend on state x.
  - Errors are uncorrelated between samples.

Problematic for weather, if samples every day.
"Serial correlation," smaller "effective sample size."
But can deal with this problem.

Ref: any applied stats text, e.g., Neter, Wasserman, and Kutner, *Applied Linear Statistical Models*.

Linear regression – big assumption!

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \ldots + \beta_N X_{i,2} + \varepsilon_i$$

unknown, with error

known, assumed no error

In our practice, the Y's typically have some small error (obs) and the X's have larger error (forecast model state).

Practically, the method works well enough to gloss over what error ε_i represents, but there is a whole branch of statistics (regression with "errors in variables") that deals with this more formally. This incorrect assumption applies to most of the rest of the methods discussed, too.

Ref: Casella and Berger, *Statistical Inference*, 1990. Also Vannitsem & Nicolis, Feb 2008 MWR (though I think their analysis is unfairly critical of MOS)

Model Output Statistics ("MOS")

most elements based on multiple linear regression

```
KBID   GFS MOS GUIDANCE    2/16/2005  1800 UTC
 DT /FEB  17                   /FEB  18                 /FEB  19
 HR   00 03 06 09 12 15 18 21 00 03 06 09 12 15 18 21 00 03 06 12 18
 N/X              32          40          25          35       19
 TMP  42 39 36 33 32 36 38 37 35 33 30 28 27 30 32 31 28 25 23 19 27
 DPT  34 29 26 22 19 18 17 17 17 17 17 15 14 13 11  8  7  6  5  2  4
 CLD  OV FW CL CL SC BK BK BK BK BK BK BK SC BK BK BK BK FW CL CL CL
 WDR  26 30 32 32 32 31 29 28 30 32 31 31 31 31 30 29 31 32 33 33 27
 WSP  12 12 12 11 08 08 09 08 09 09 10 10 10 12 13 13 15 16 15 09 08
 P06        17     0     0     0     4     0    10     6     8  0  0
 P12              17           0          10          17        8
 Q06         0     0     0     0     0     0     0     0     0  0  0
 Q12              0           0           0           0        0
 T06      0/ 2  0/ 0  1/ 0  1/ 2  0/ 1  0/ 1  1/ 0  0/ 1  0/ 0  0/ 0
 T12                 1/ 0        1/ 2        1/ 1        0/ 1  0/ 0
 POZ   0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
 POS  13 47 70 84 91100 96100100100100 92100 98100100100 94 92100100
 TYP   R  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S
 SNW                                       0                       0
 CIG   7  8  8  8  8  8  8  8  8  7  7  7  8  7  7  7  8  8  8  8  8
 VIS   7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
 OBV   N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N
```

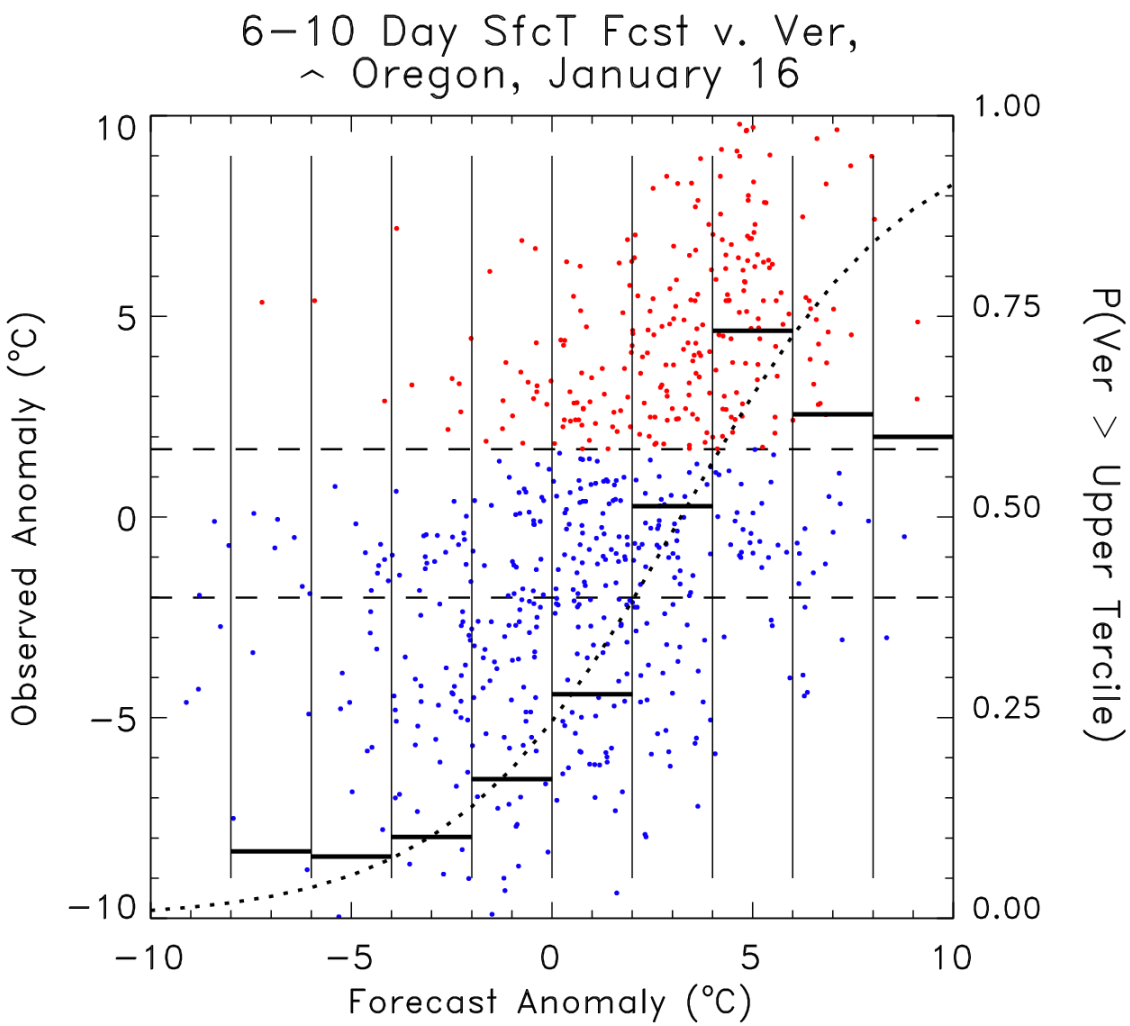
US: Statistical corrections to operational US NWS models, some fixed (NGM), some not (Eta, GFS). Refs: http://www.nws.noaa.gov/mdl/synop/index.htm, Carter et al., *WAF*, **4**, p 401, Glahn and Lowry, *JAM*, **11**, p 1580. **Canadian** models discussed in Wilson and Vallee, *WAF*, **17**, p. 206, and *WAF*, **18**, p 288. **Britain**: Met Office uses "updateable MOS" much like perfect prog.

Logistic regression

- Useful for making probabilistic forecasts for some binary event, e.g, precip above threshold.
- 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 | *x*) as a continuous function of *x*.
- Logistic Regression: $$P = 1 - \frac{1}{1 + exp\ (\beta_0 + \beta_1 x)}$$

Ref: any applied statistics text.

Logistic regression using a long data set of observed and forecast anomalies



Seeking to predict probability of warmer than normal conditions (upper tercile of observed). Using reforecasts (a later talk), 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.

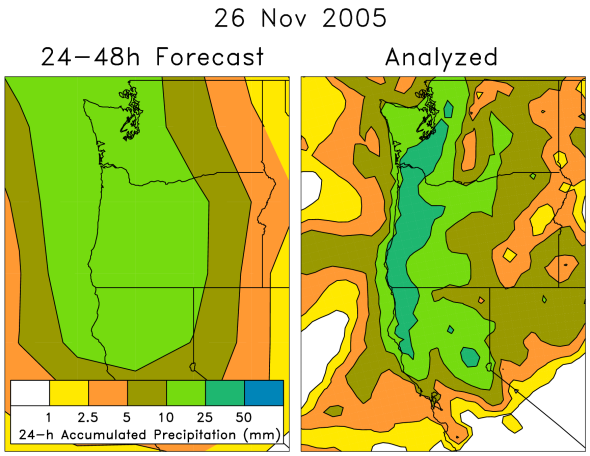
Dashed lines: tercile boundaries
Red points: samples above upper tercile
Blue points: samples below upper tercile
Solid bars: probabilities by bin count
Dotted line: logistic regression curve

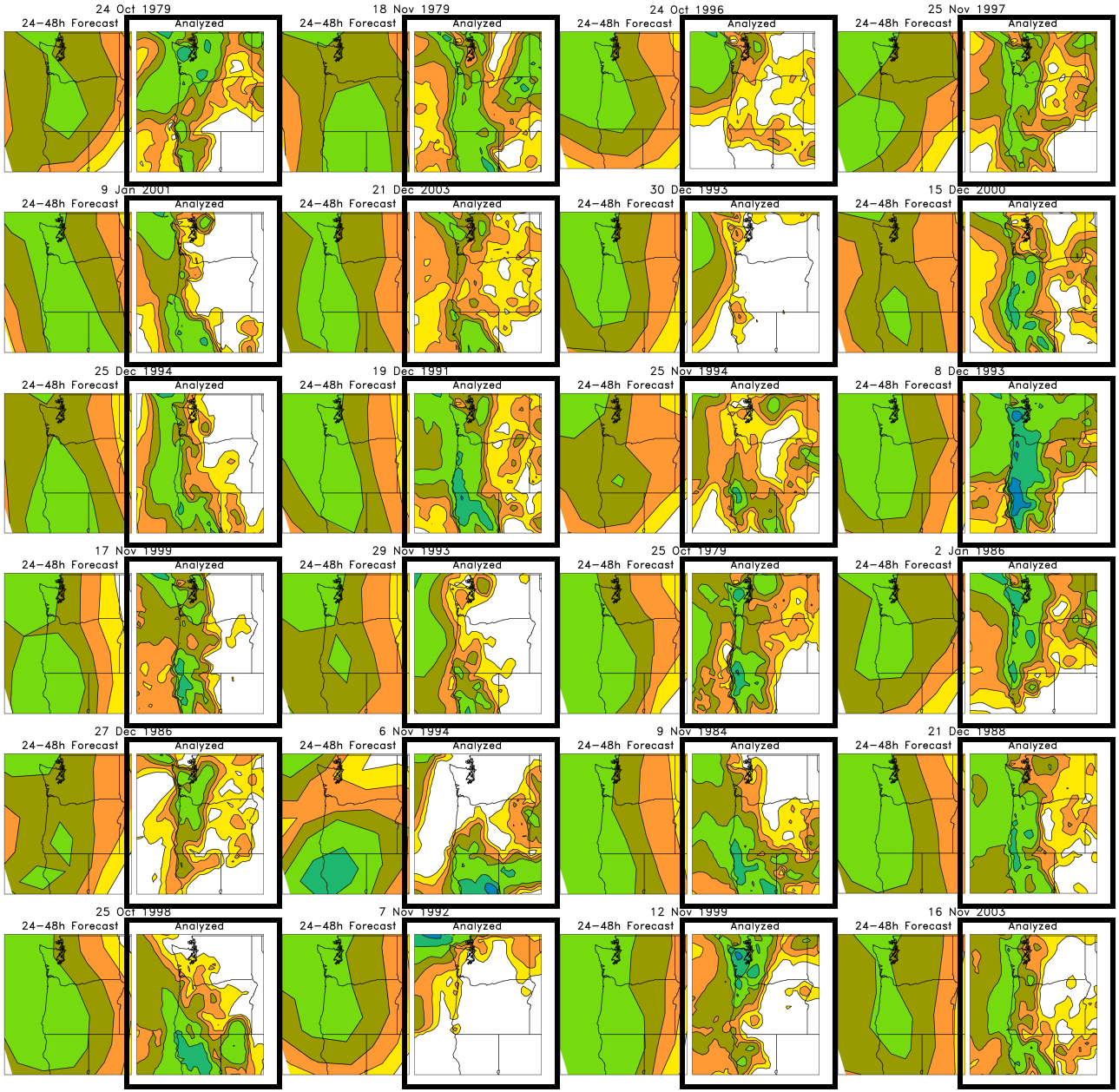
Ref: Hamill et al. MWR, June 2004

Logistic regression drawbacks

- Doesn't generate full pdf (though see Wilks, 2009, *Met Apps*, p. 361).
- With ensembles, what do you use as predictors? Ens. mean? Spread? Every member?
- Iterative technique, can be slower.
- Better have training set with distribution of 1's and 0's, otherwise software will croak.

Analog technique using reforecasts

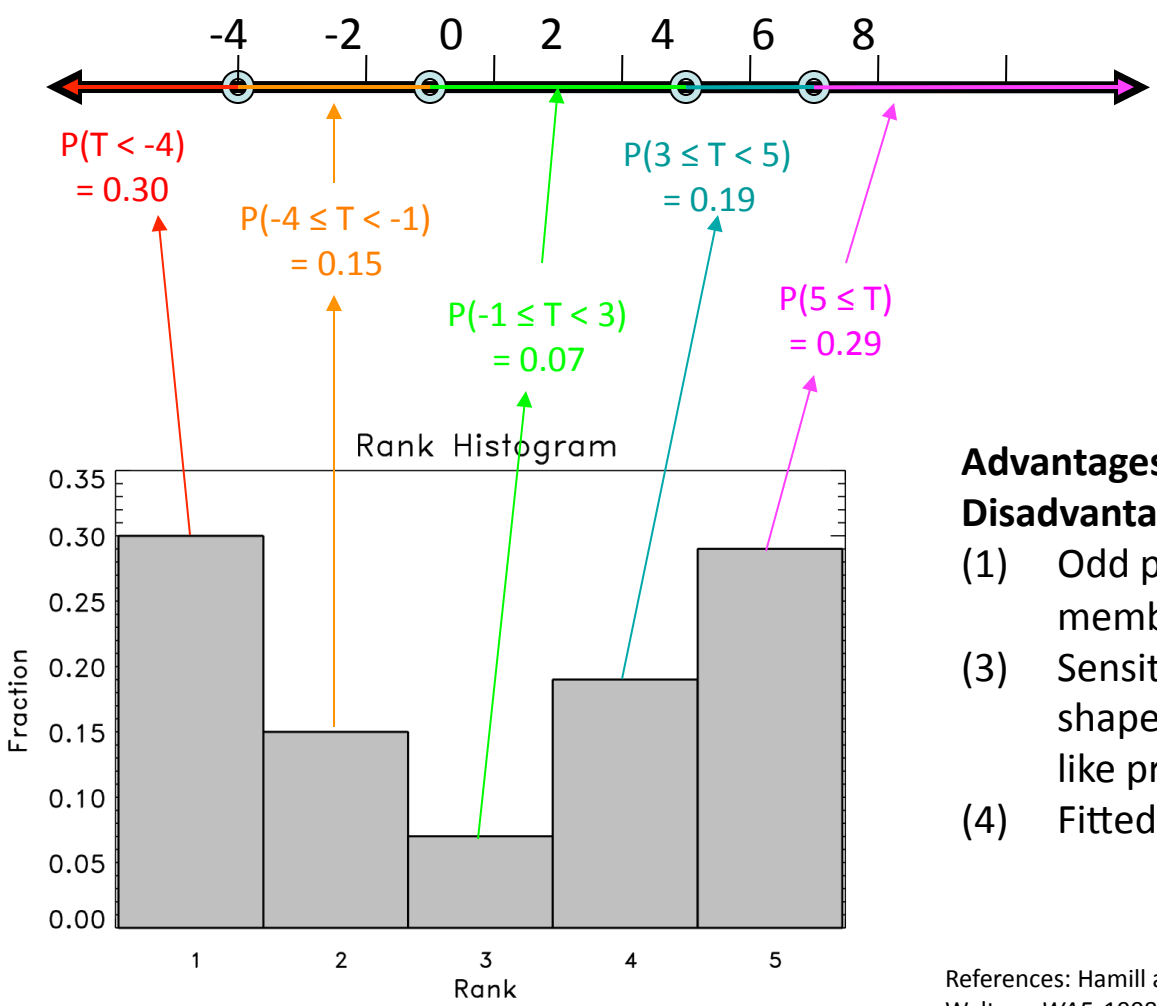




On the left are old forecasts similar to today's ensemble-mean forecast. The data on the right, the analyzed precipitation conditional upon the forecast, can be used to statistically adjust and downscale the forecast.

Analog approaches like this may be particularly useful for hydrologic ensemble applications, where an ensemble of realizations is needed.

Rank histogram technique for ensemble calibration



NCEP MRF precipitation forecasts, from Eckel and Walters, 1998

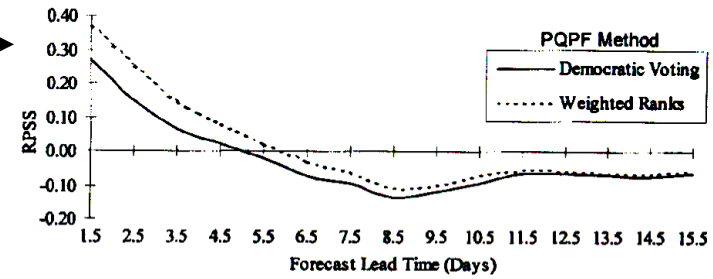


FIG. 10. Ranked probability skill score (RPSS) results for all forecast lead times.

Advantages: Demonstrated skill gain

Disadvantages:

(1) Odd pdfs, especially when two ensemble members close in value.

(3) Sensitive to shape of rank histogram, and shape of histogram may vary with aspects like precip amount --> sample size issues.

(4) Fitted parametric distributions as skillful

References: Hamill and Colucci (*MWR*, 1997, 1998; Eckel and Walters, *WAF*, 1998; used at Met Office)

Bayesian model averaging (BMA)

$$p(y \mid f_1, \ldots, f_K) = \sum_{k=1}^{K} w_k \; g_k(y \mid f_k)$$

Weighted sum of kernels centered around individual, bias-corrected forecasts.

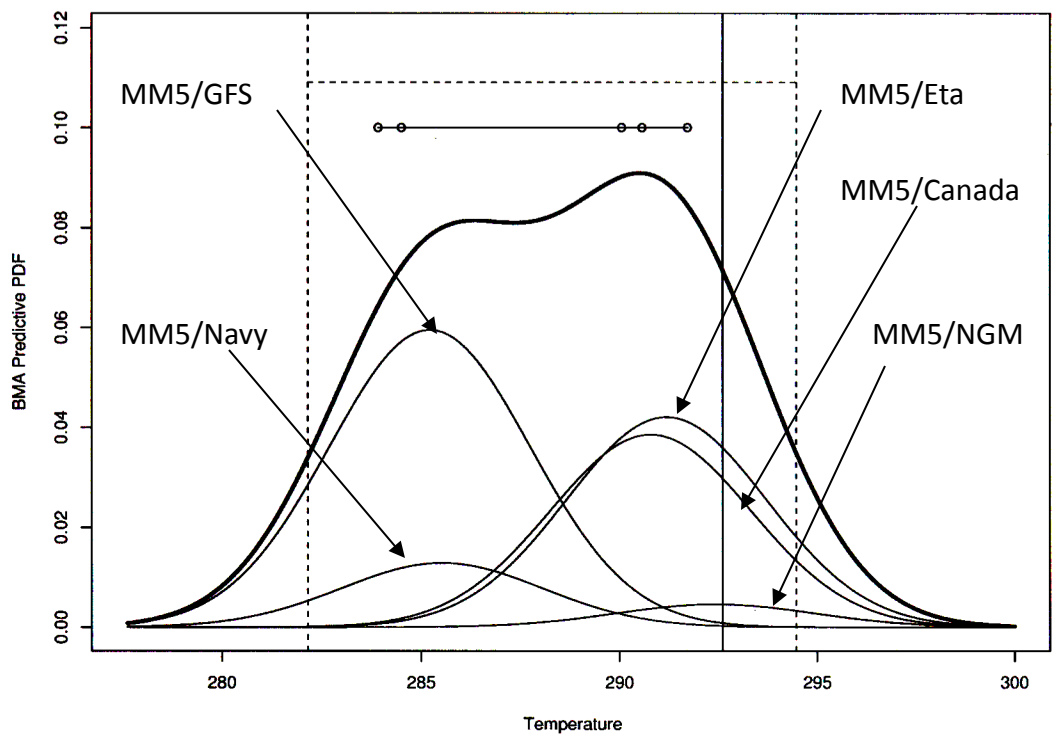


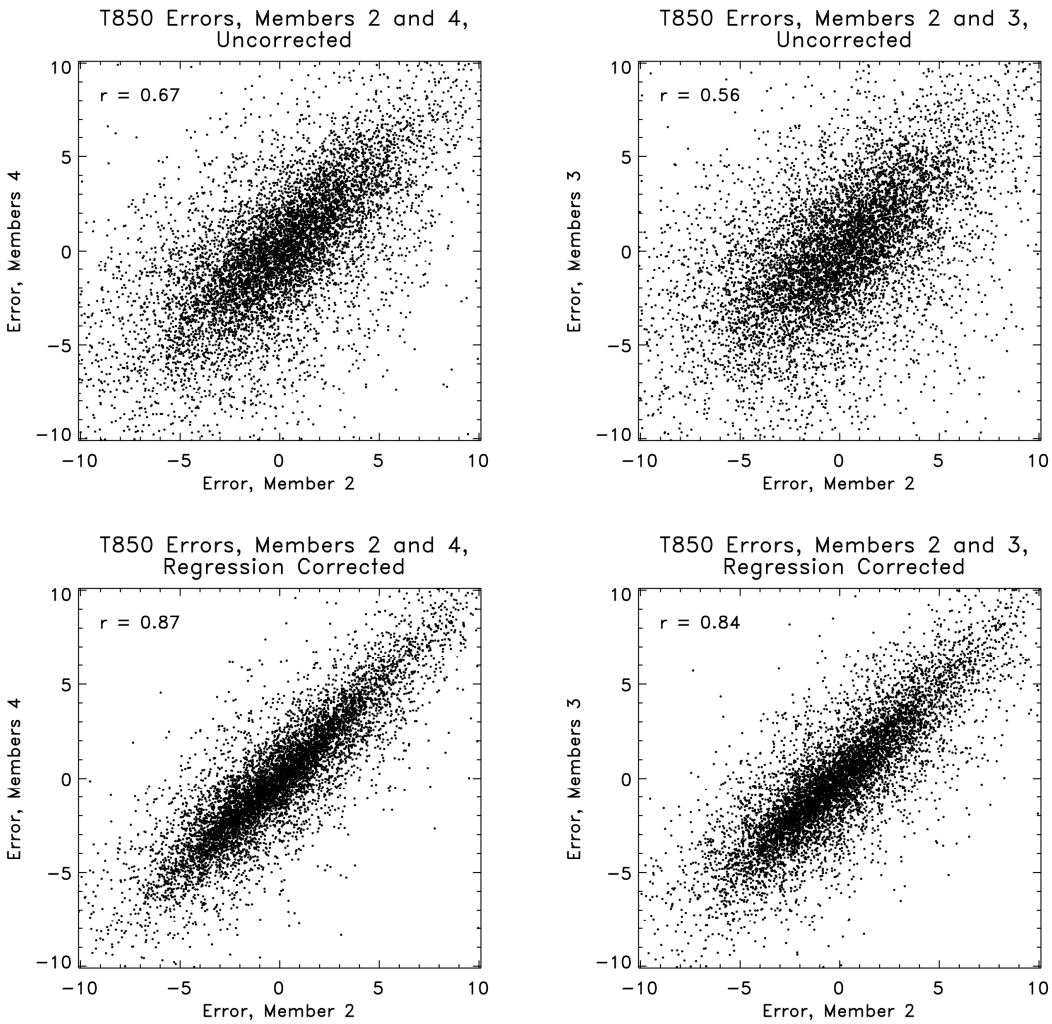
Figure 3: BMA predictive PDF (thick curve) and its five components (thin curves) for the 48-hour surface temperature forecast at Packwood, Wash., initialized at 0000 UTC on June 12, 2000. Also shown are the ensemble member forecasts and range (solid horizontal line and bullets), the BMA 90% prediction interval (dotted lines), and the verifying observation (solid vertical line).

Advantages: Theoretically appealing. No parameterized distribution assumed, weights applied proportional to their independent information (in concept).

Disadvantages: When trained with small sample, BMA radically de-weighted some members due to "overfitting" See Hamill, *MWR*, Dec. 2007.

Ref: Raftery et al., MWR, 2005. Wilson et al., MWR, 2007

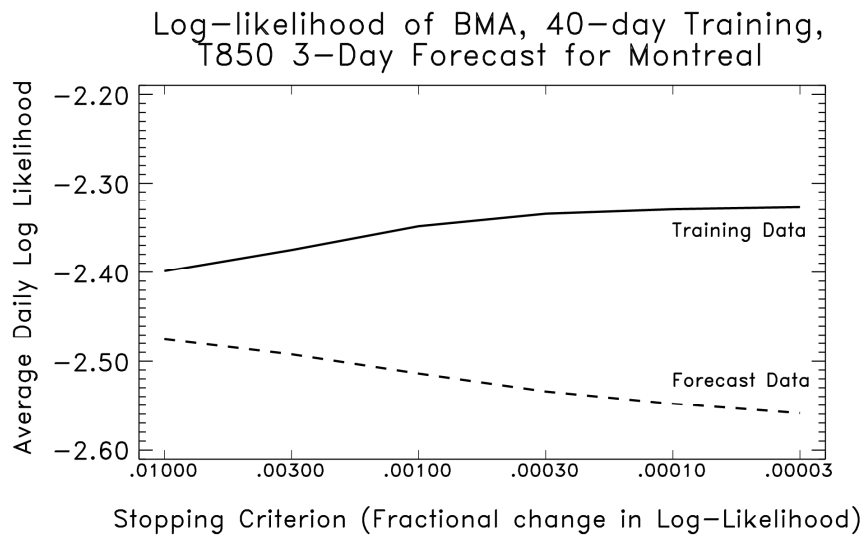
Why BMA's unequal weights?
(1) regression correction accentuates error correlations.

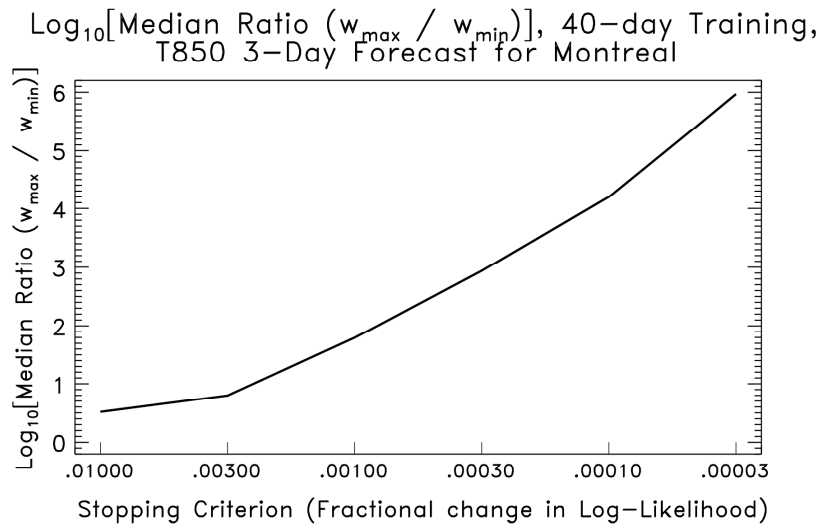


Ref: Hamill, *MWR*, Dec. 2007

Why BMA's unequal weights?
(2) E-M overfits with little training data

An "estimation-minimization" (E-M) algorithm is used to determine the weights applied to ensemble members. If two forecasts have highly co-linear errors, E-M will weight one very highly, the other very little.

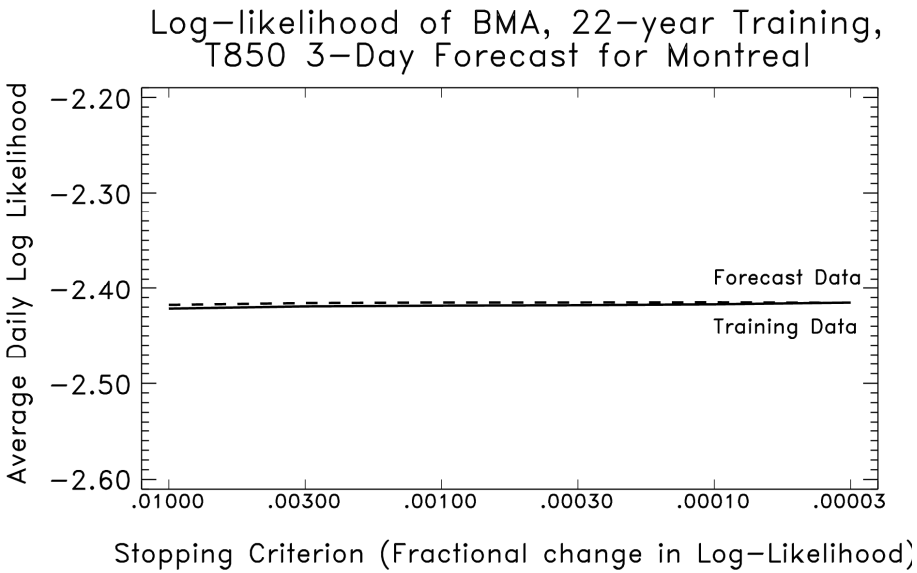


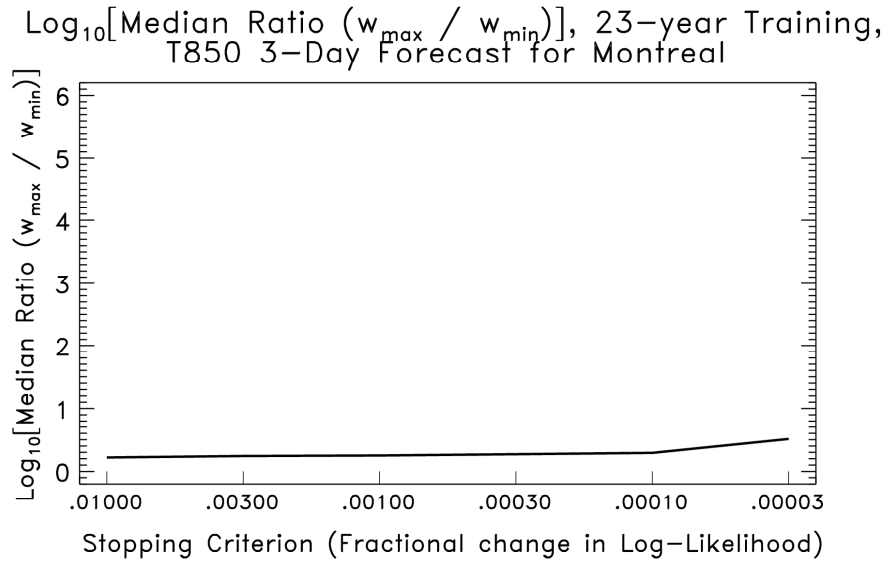


E-M is an iterative technique, and we can measure the accuracy of the fit to the data through the log-likelihood. Something odd happens here; as the E-M convergence criteria is tightened, the fit of the algorithm to independent data gets worse.

This plots the ratio of the weights of the highest-weighted member to the lowest-weighted member. As the convergence criterion is tightened, the method increasingly weights a few select members and de-weights others.

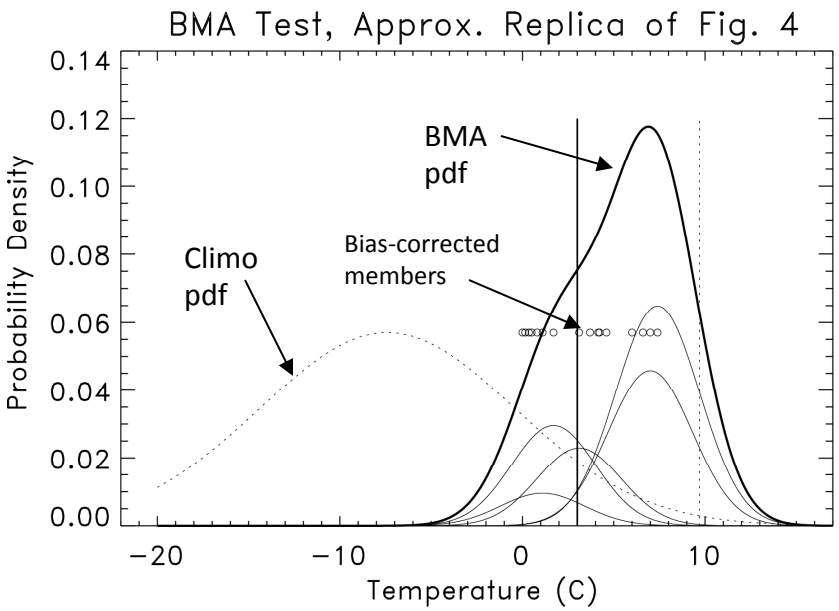
(BMA overfitting not a problem with 2+ decades training data)

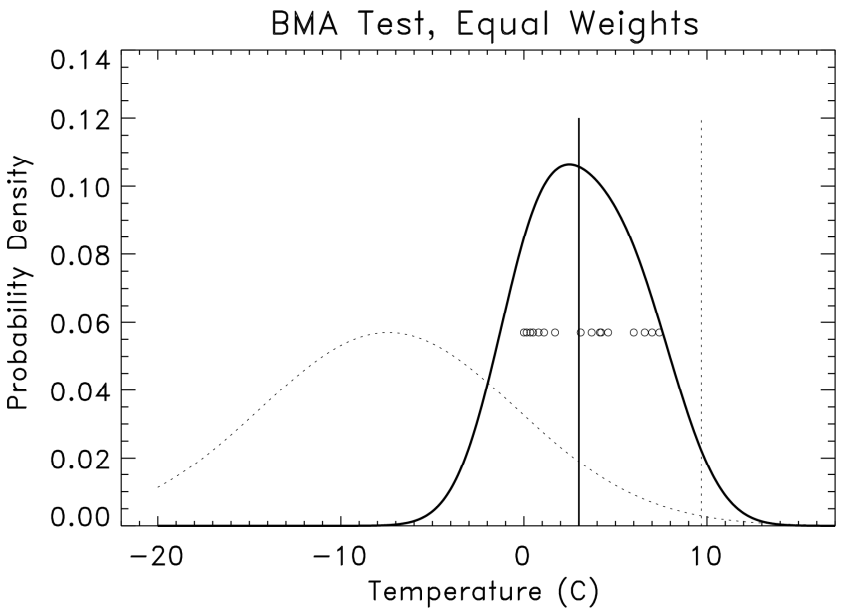




With reforecast data set, we can train with a very large amount of data. When we do so, the weights applied to individual members are much more equal. This indicates that the unequal weighting previously is incorrect.

BMA's problem: an example





Here's a test of BMA in the winter season for a grid point near Montreal. BMA ends up highly weighting the warmest members (inappropriately so), thus producing a very high probability of a warm forecast.

Ref: Laurie Wilson, Met Service Canada, personal correspondence. See also upcoming Bishop et al. MWR manuscript.

"Bayesian Processor of Forecasts"

- Two key ideas:
  - (1) Bayes' Rule: leverage prior non-NWP information, whether from climatology, persistence, whatever. Update with NWP information

$$\phi\left(o\middle|f\right) \propto \phi\left(f\middle|o\right) g\left(o\right)$$

  - (2) If data non-normally distributed, transform data to space where normally distributed before performing regression analysis.

Ref: Krzysztofowicz and Evans, WAF, April 2008

Non-homogeneous Gaussian regression

- **Reference**: Gneiting et al., *MWR*, **133**, p. 1098
- **Predictors**: ensemble mean and ensemble spread
- **Output**: mean, spread of calibrated Gaussian distribution

$$f^{CAL}\left(\overline{\mathbf{x}}, \sigma\right) \sim N\left(a + b\overline{\mathbf{x}}, c + d\sigma\right)$$

- **Advantage**: leverages possible spread/skill relationship appropriately. Large spread/skill relationship, c ≈ 0.0, *d* ≈1.0. Small, *d* ≈ 0.0
- **Disadvantage**: iterative method, slow…no reason to bother (relative to using simple linear regression) if there's little or no spread/skill relationship.

Is there a "best" calibration technique?

Using Lorenz '96 toy model, direct model output (DMO), rank histogram technique, MOS applied to each member, dressing, logistic regression, non-homogeneous Gaussian regression (NGR), "forecast assimilation", and Bayesian model averaging (with perturbed members assigned equal weights) were compared. Comparisons generally favored logistic regression and NGR, though differences were not dramatic, and results may not generalize to other forecast problems such as ones with non-Gaussian errors.

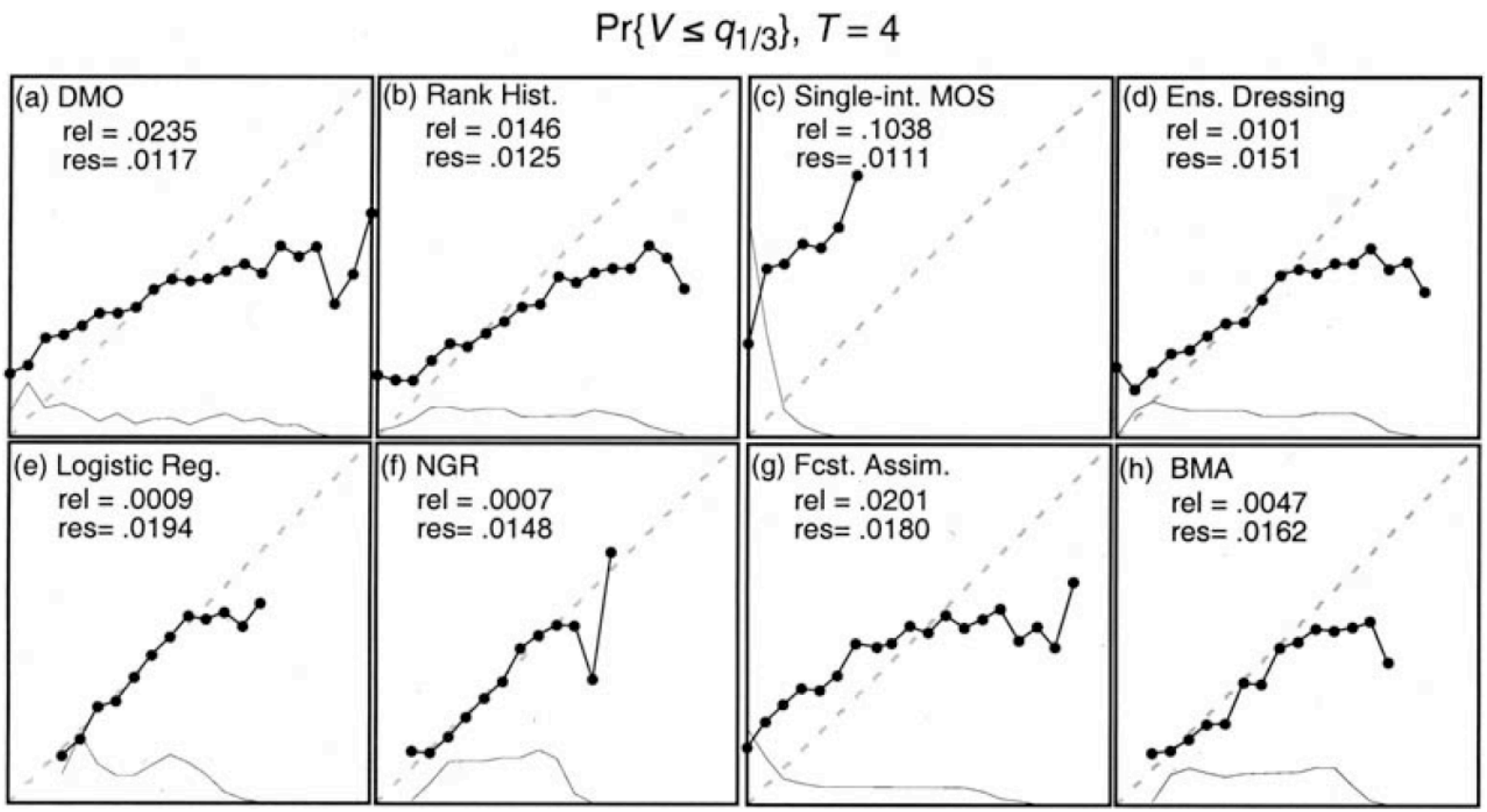


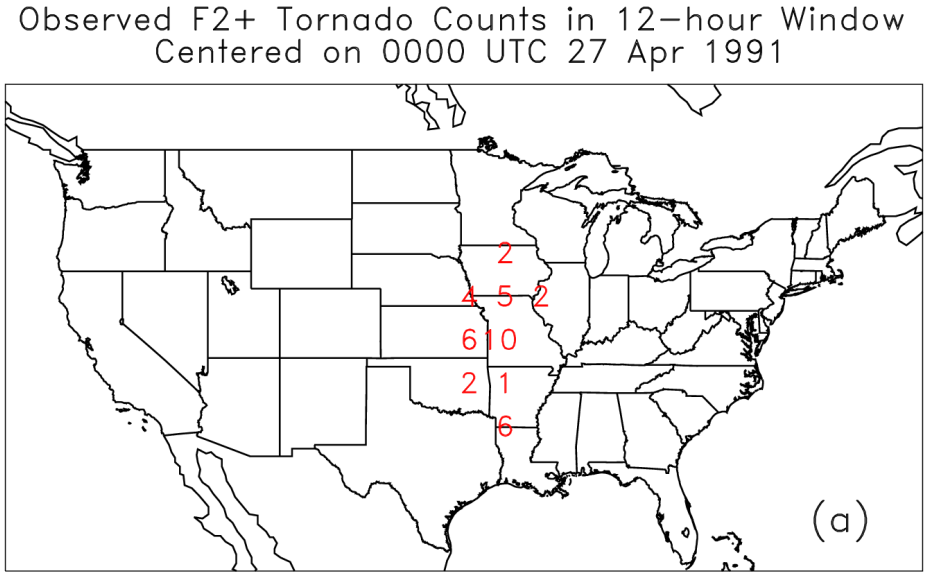
Figure 8. *As Figure 5, for $Pr\{V \le q_{1/3}\}$ at lead time $T = 4$.*

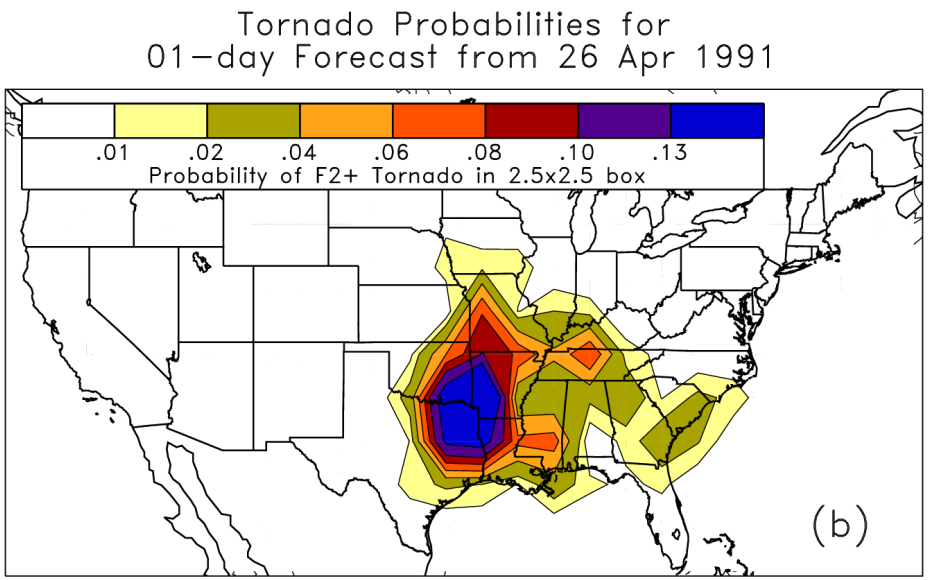
Ref: Wilks, *Met. Apps*, 2006, **13**, p. 243

Rationale 3 for post-processing:

Predict probabilities of a variable that may not be forecast by the model using variables that are forecast by the model, e.g., tornado probability from CAPE, CIN, shear.

Tornado probabilities from analogs

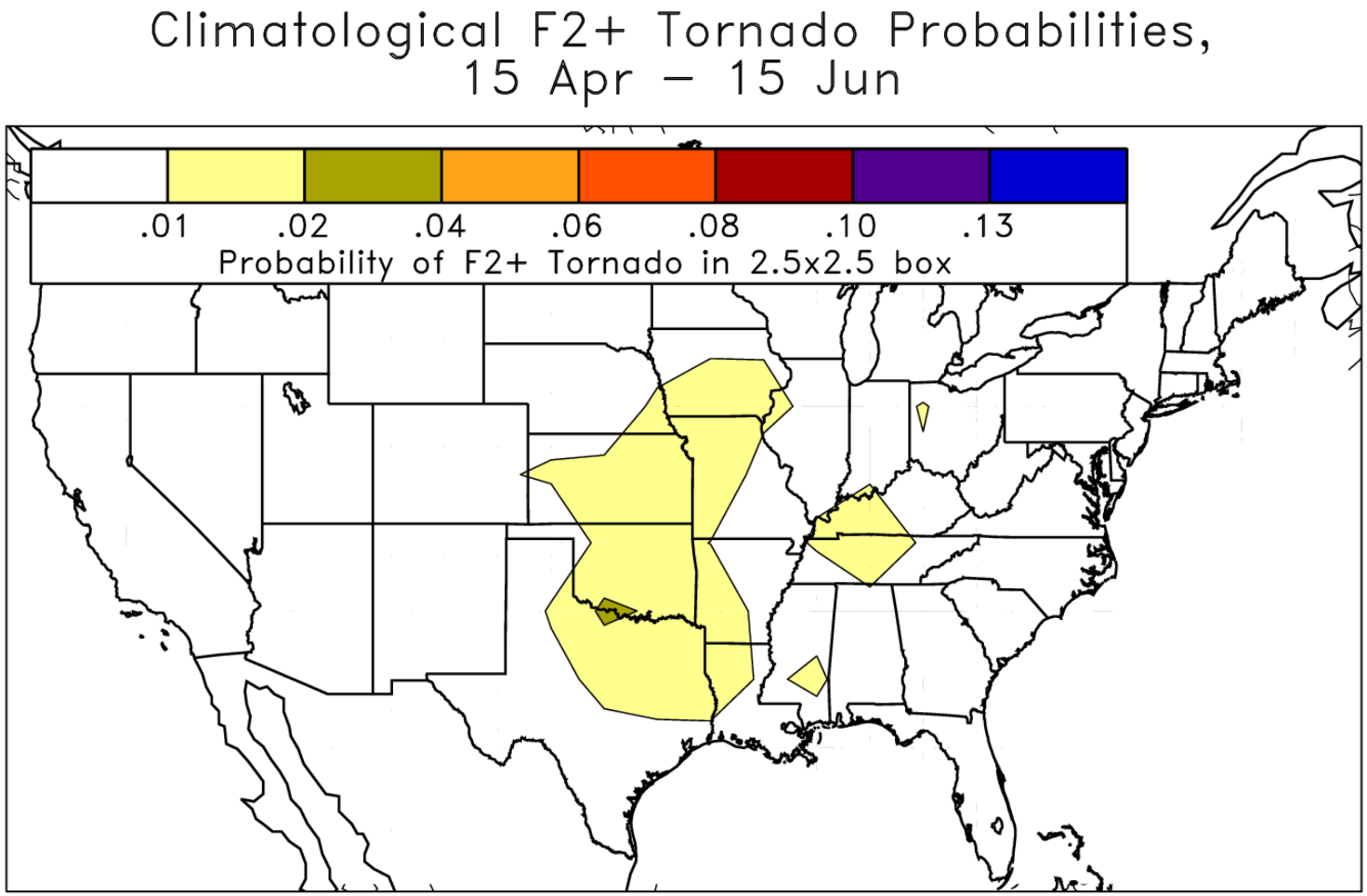




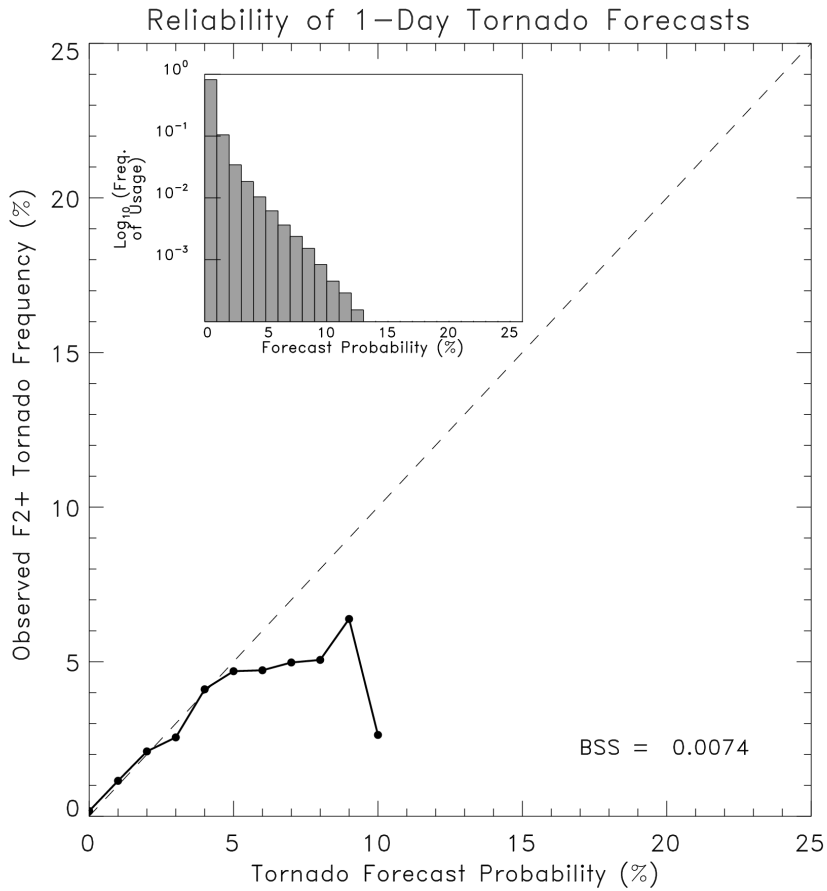
For each grid point, in a region around the point I compared that day's forecast ensemble-mean shear, CAPE, and CIN to the values of shear, CAPE, CIN from past days at a similar time of the year in the reforecast data set. This provided a set of dates. The probability of F2+ tornadoes for that box was determined by counting the number of analog days with tornadoes in that box divided by the number of analog days (50).

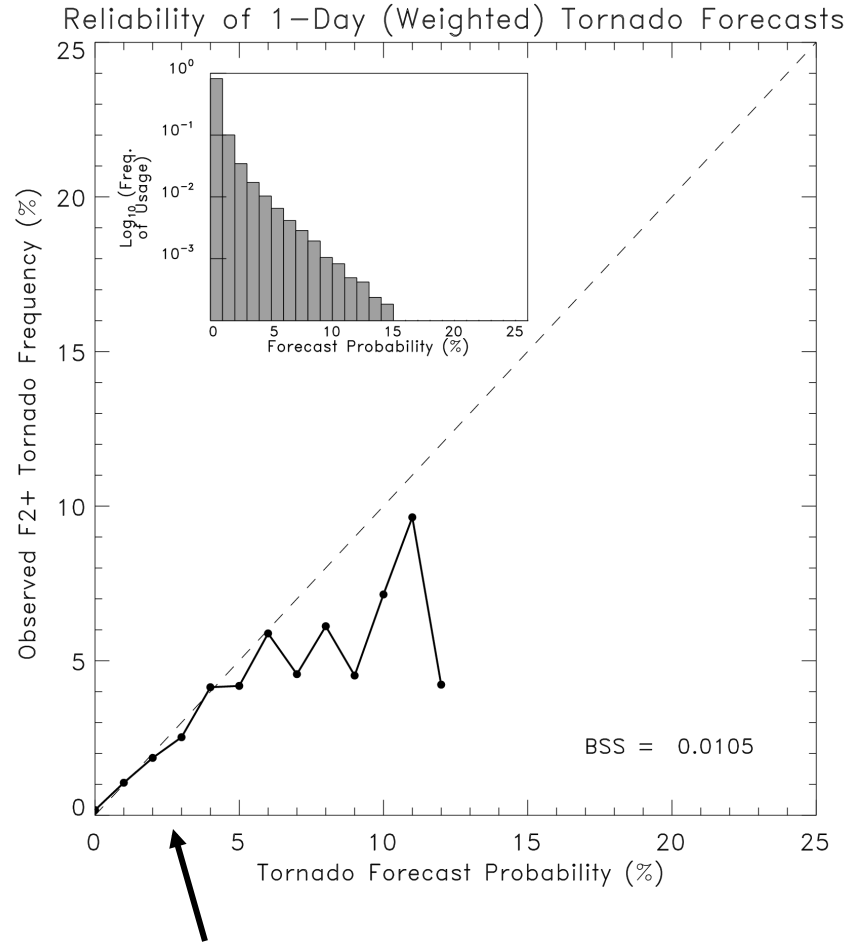
(details in supplementary slides after conclusion).

Climatology of F2+ tornadoes



Reliability and skill of tornado analog probabilities





Note: low BSSes inevitable when verifying rare events.

here, closer analogs are weighted more than further ones

Tradeoffs between reforecast sample size and statistical sophistication

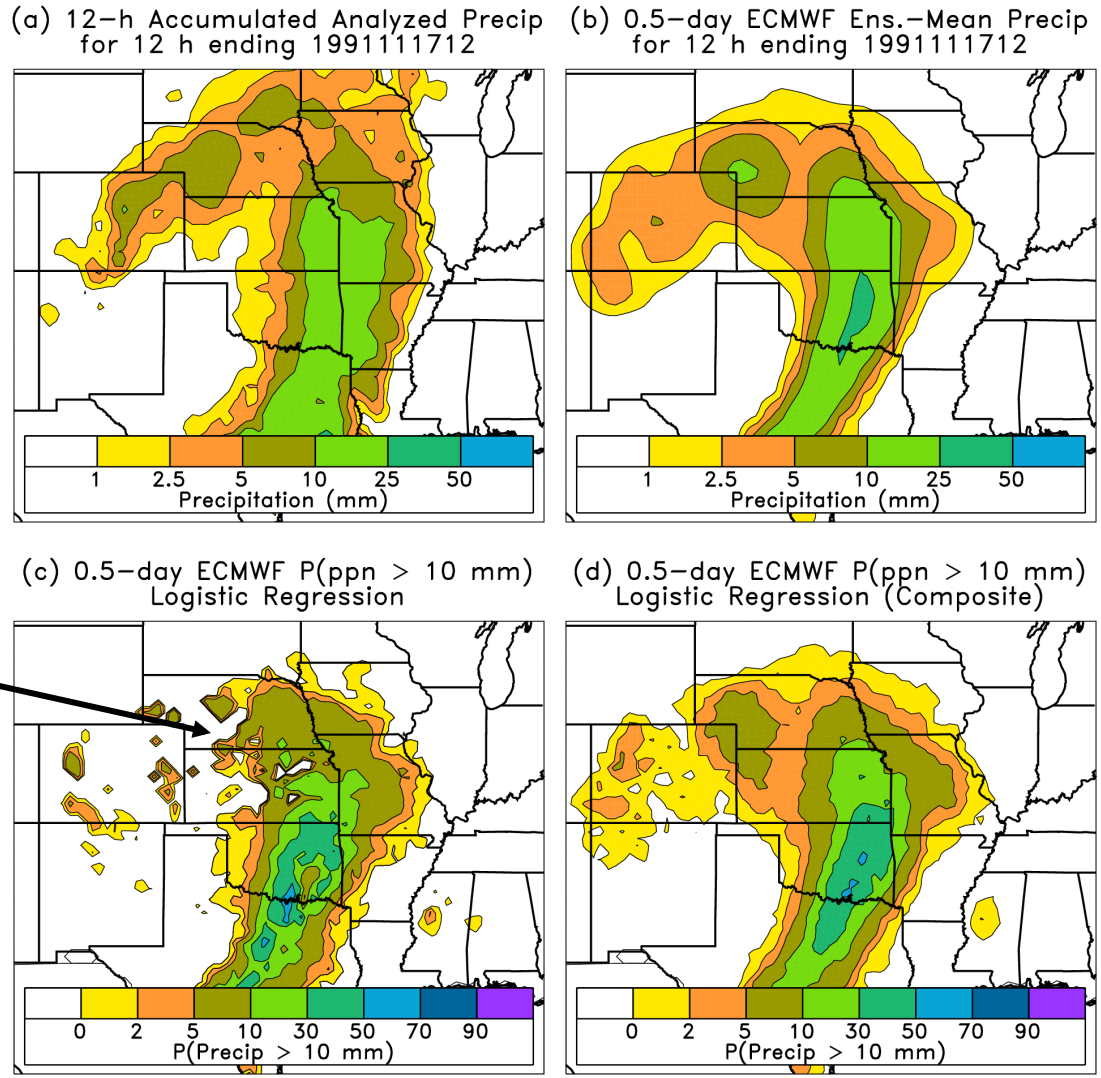
- Large reforecast: almost any reasonable post-processing approach gives some benefit.
- Smaller training data set: you have to find ways such as "compositing" to increase the sample size.

Example: precipitation calibration with ECMWF reforecasts

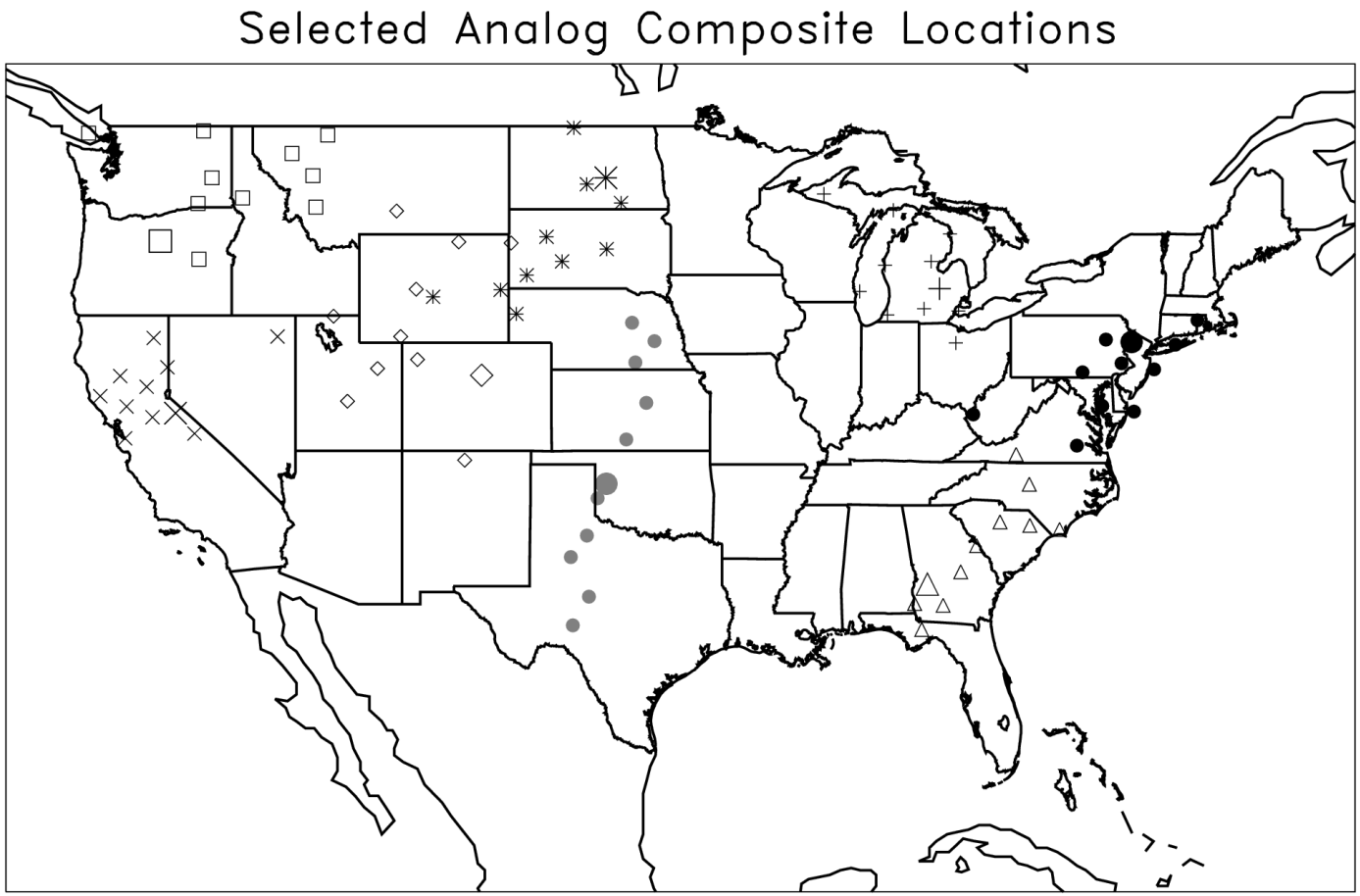
- NARR CONUS 12-hourly data used for training, verification. ~32 km grid spacing
- Logistic regression
- Use all of their 20 year * 1x/week Sep-Dec training data, except for data from the year being forecast.

Problem: patchy probabilities when grid point X trained with only grid point X's forecasts / obs

Even 20 years of weekly forecast data (260 samples after cross-validation) is not enough for stable regression coefficients, especially at higher precipitation thresholds.



Tested method: add in training data at other grid points that have similar analyzed climatologies



Big symbol: grid point where we do regression

Small symbols: analog locations with similar climatologies

When is it proper to use training data at location B to supplement regression analysis at location A?

(1) When location B's errors are independent of location A's errors.

(2) When observed CDF at A and B are very similar.

(3) When forecast CDF at A and B are very similar.

(4) When corr(forecast, observed) at A and B are similar.

When is it proper to use training data at location B to supplement regression analysis at location A?

(1) When location B's errors are independent of location A's errors.

Make sure location A is not too close to location B

(2) When observed CDF at A and B are very similar.

(3) When forecast CDF at A and B are very similar.

(4) When corr(forecast, observed) at A and B are similar.

When is it proper to use training data at location B to supplement regression analysis at location A?

(1) When location B's errors are independent of location A's errors.

(2) When observed CDF at A and B are very similar.

Need lots of samples. Luckily, ~28 year NARR provides them.

(3) When forecast CDF at A and B are very similar.

(4) When corr(forecast, observed) at A and B are similar.

When is it proper to use training data at location B to supplement regression analysis at location A?

(1) When location B's errors are independent of location A's errors.

(2) When observed CDF at A and B are very similar

(3) When forecast CDF at A and B are very similar.

(4) When corr(forecast, observed) at A and B are similar.

Judging this would be tough with ECMWF forecasts. Only 14 weeks*20 years, not a large sample for non-normally distributed data. Can be fooled by rare events.

When is it proper to use training data at location B to supplement regression analysis at location A?

(1) When location B's errors are independent of location A's errors.

(2) When observed CDF at A and B are very similar

(3) When forecast CDF at A and B are very similar.

(4) When corr(forecast, observed) at A and B are similar.

Tricky to compute in dry regions, where overwhelming bulk of the samples are zero's.

Conclusions #1: post-processing can be very beneficial if you have the following:

- Enough training data (past forecasts with the same model as you are running operationally, plus quality obs/analyses)
  - more data needed for rare events, long-lead forecasts.
- A reasonable statistical correction method.

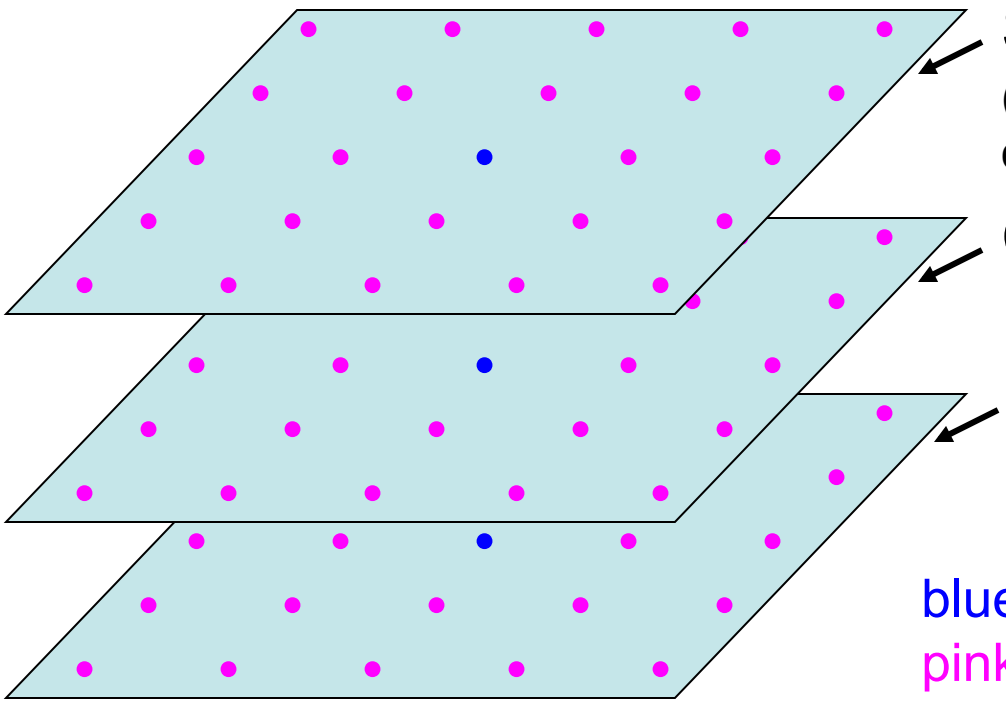
Poorly post-processed guidance is worse than nothing at all; misleading, another thing the forecaster must compensate for.

Conclusions #2: What's the best post-processing technique?

- Enlarging training sample size has a bigger effect for many variables than changing the post-processing technique.
- Preferred techniques may vary from user to user. Hydrologists want bias-corrected and downscaled members, others want smooth pdfs.
- If possible, KISS (Keep it simple, stupid). Increasing focus on complicated techniques. Often parametric distributions, linear techniques work just fine.
- We are just learning how to begin to do joint post-processing in multiple dimensions; no idea yet what's best.

Technique for finding tornado forecast analogs

For a given grid point, match today's *ensemble mean* fields with past forecast fields. Find *n* closest analog dates.



blue dot: point to find analogs for
pink dots: points to match up current forecasts with old ones.
Note: can vary weights horizontally

Result:

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

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

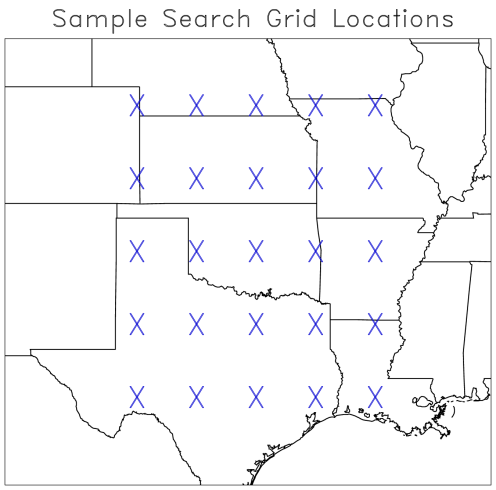
  0.867 0.915 0.931 0.915 0.867
  0.915 0.965 0.982 0.965 0.915
  0.931 0.982 1.000 0.982 0.931
  0.915 0.965 0.982 0.965 0.915
  0.867 0.915 0.931 0.915 0.867

- 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.)

Making probabilistic forecasts from analogs

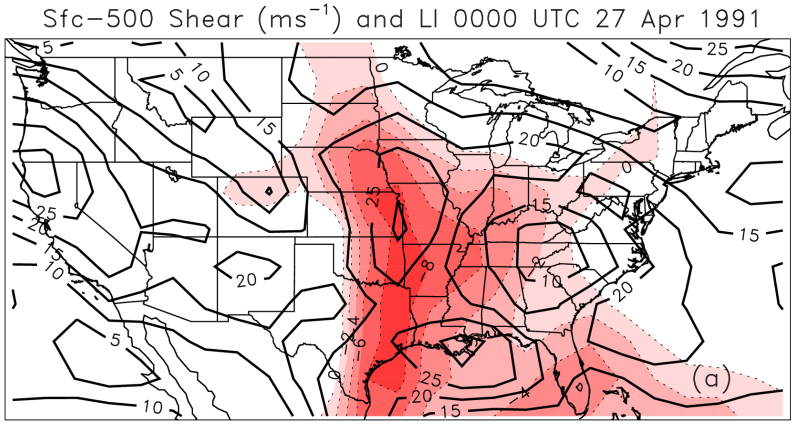
- **Method 1**: Use raw relative frequency of observed tornado occurrence in *n* analogs

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

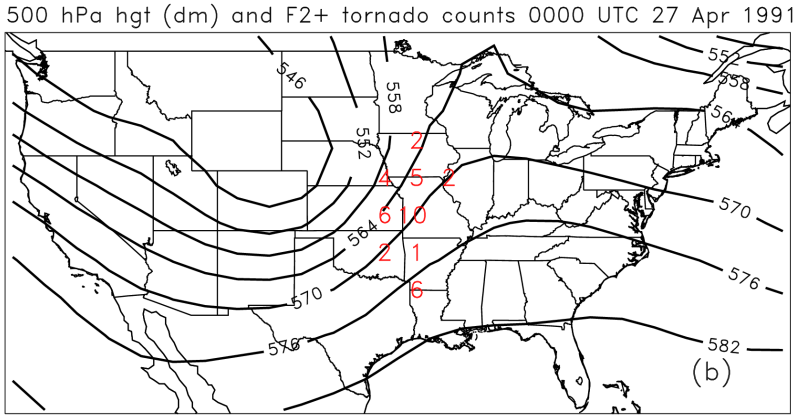
T_i = 1 if F2+ occurred,
T_i = 0 if no F2+ occurred

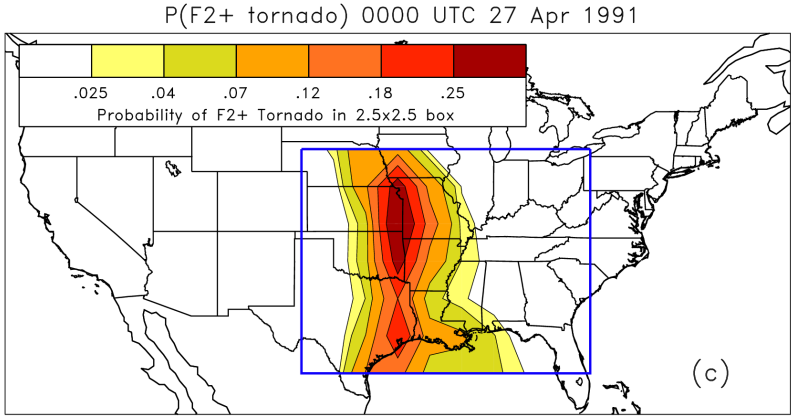
- **Method 2**: Use *weighted* relative frequency of observed tornado occurrence in *n* analogs

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



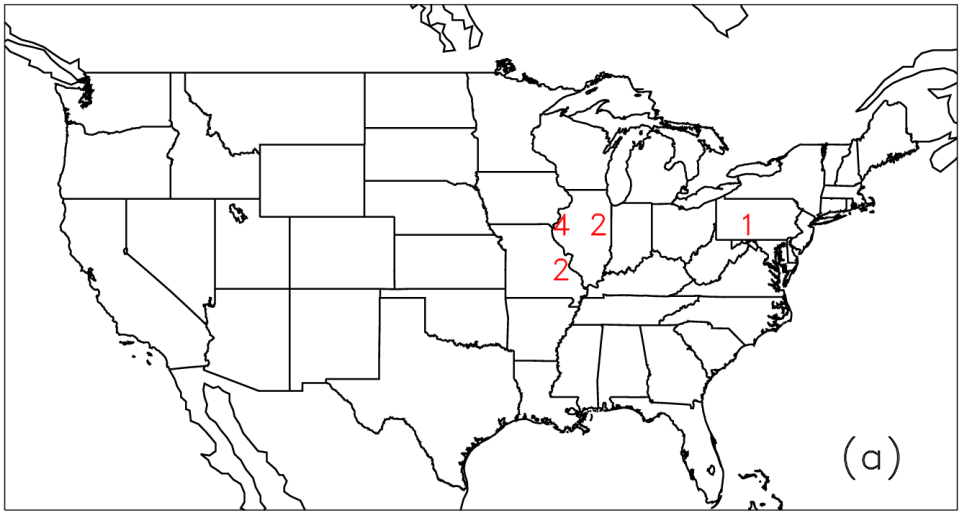
Observed
shear and LI
from NCEP-NCAR
reanalysis



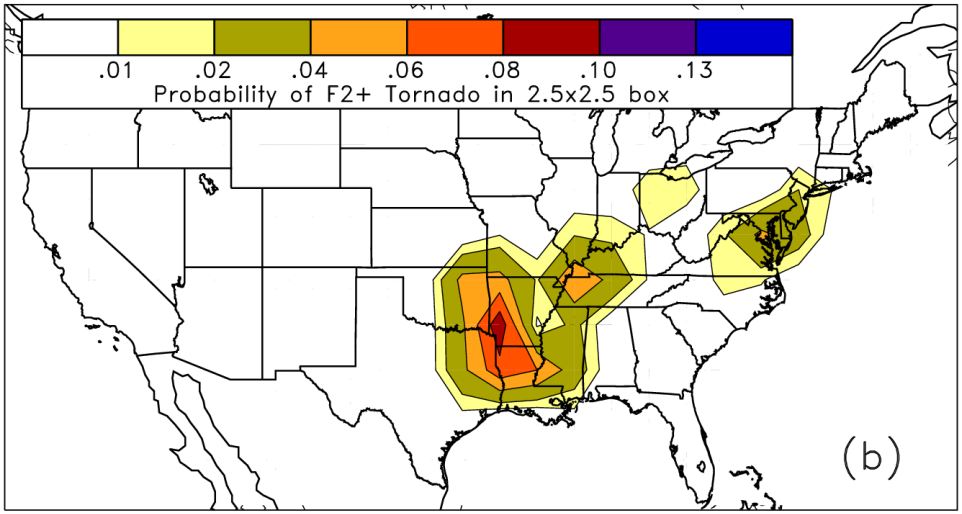


Nowcast tornado
probabilities based
on analyzed shear, LI
(see upcoming BAMS article
on May ‘03 outbreak)

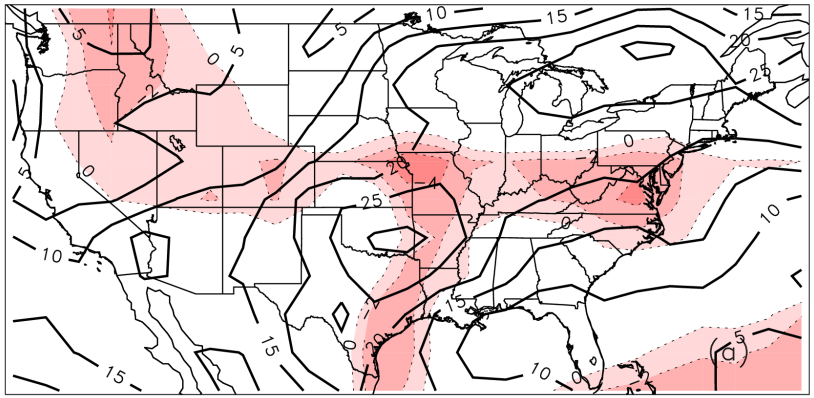
Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 13 May 1980



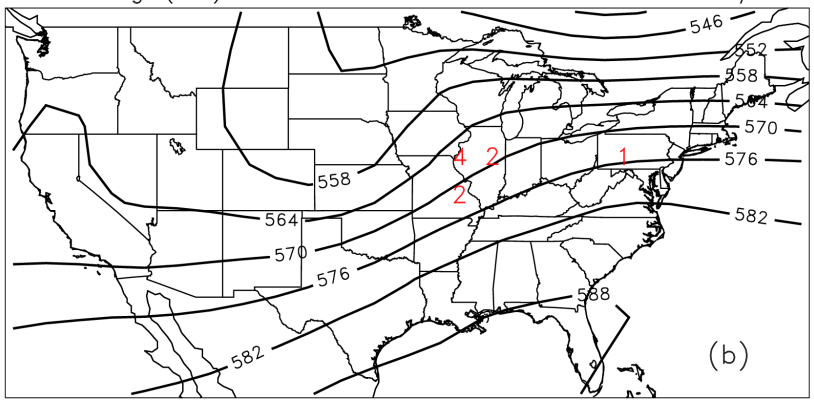
Tornado Probabilities for
01-day Forecast from 12 May 1980



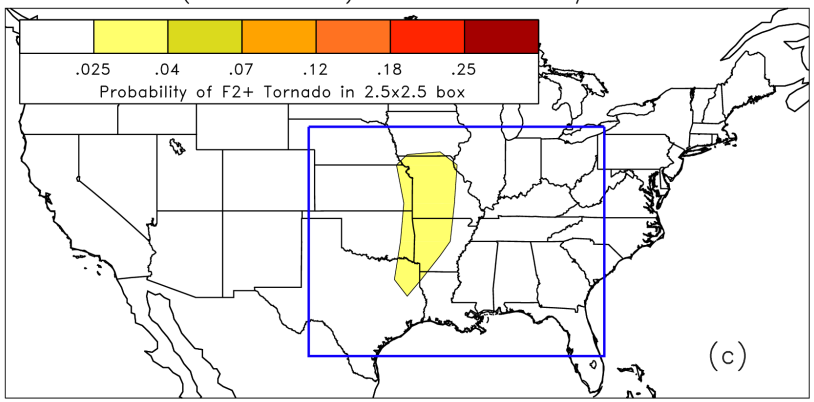
Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 13 May 1980



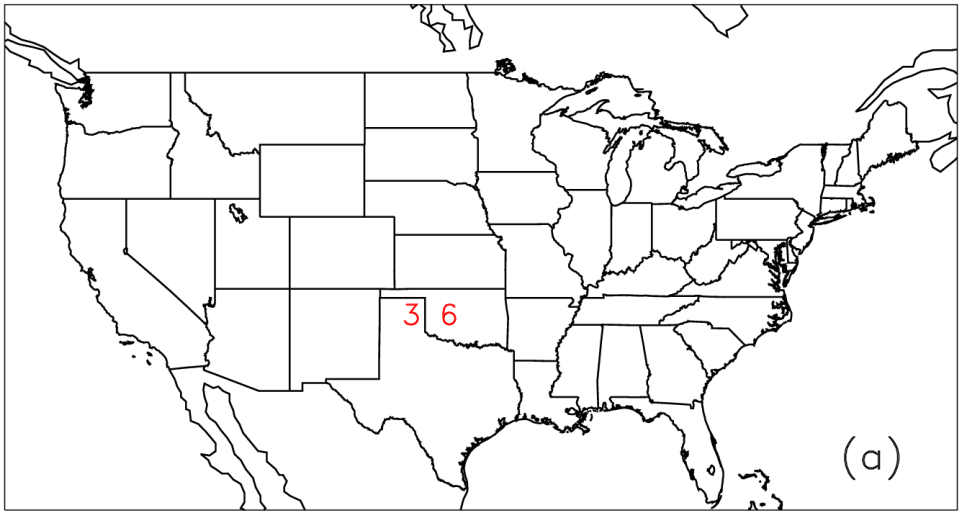
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 13 May 1980



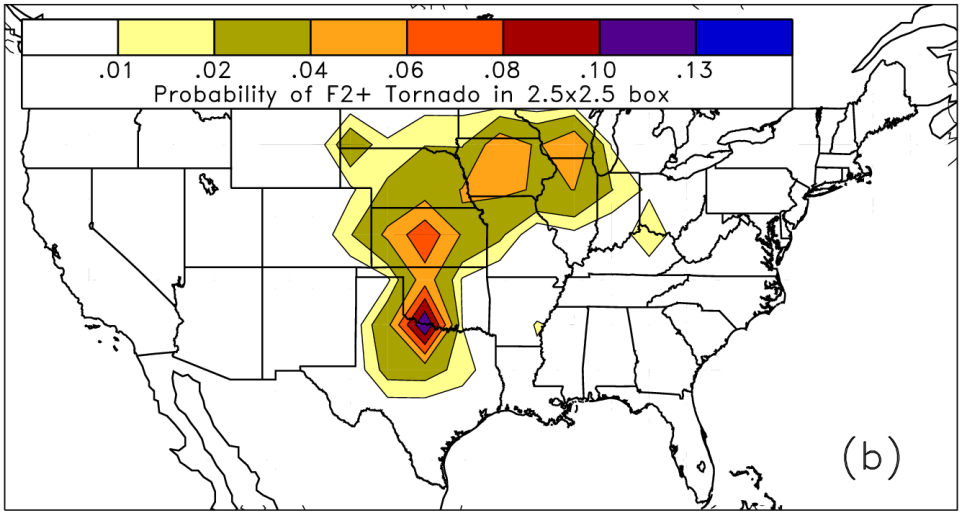
P(F2+ tornado) 0000 UTC 13 May 1980



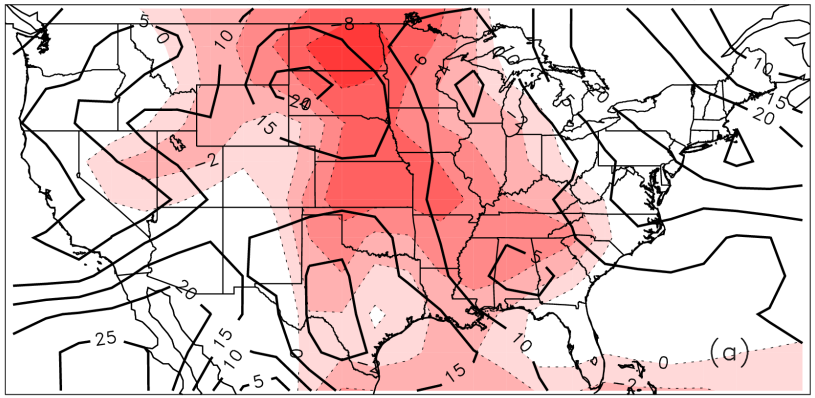
Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 29 May 1980



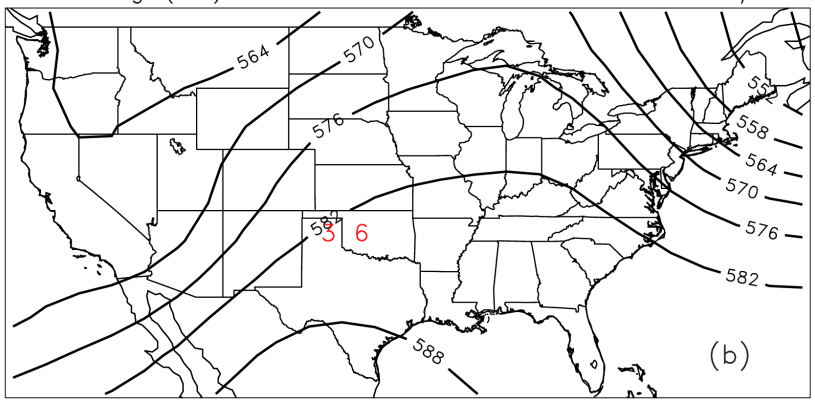
Tornado Probabilities for
01-day Forecast from 28 May 1980



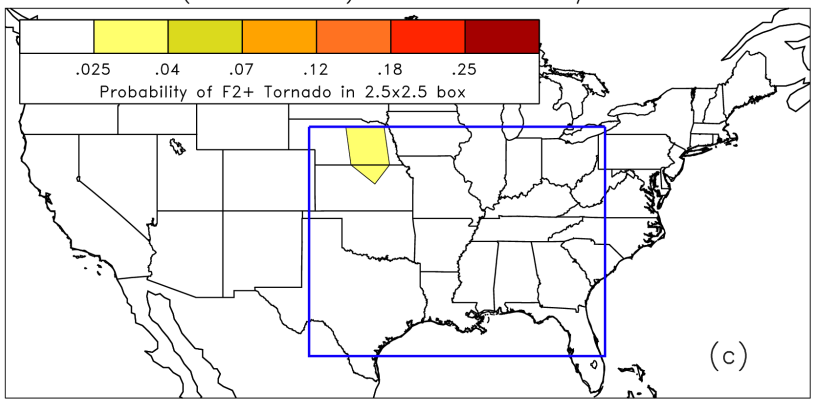
Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 29 May 1980



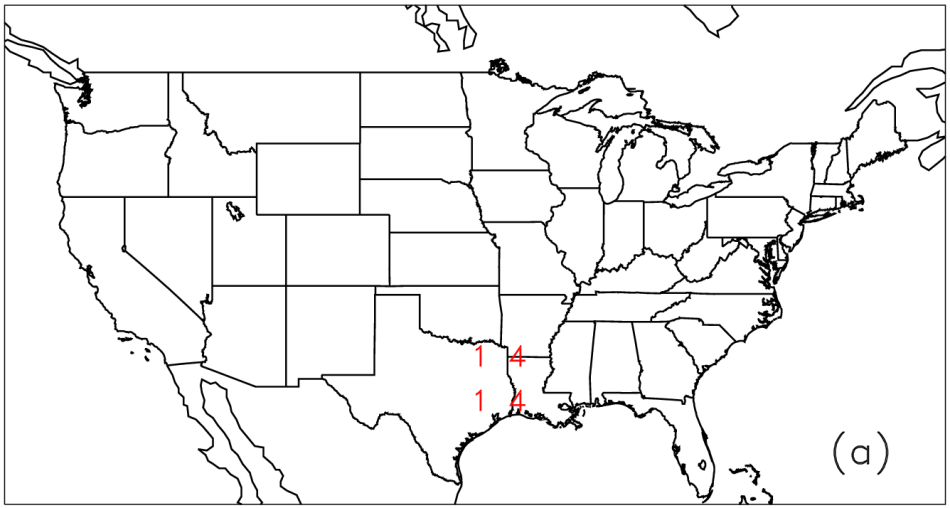
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 29 May 1980



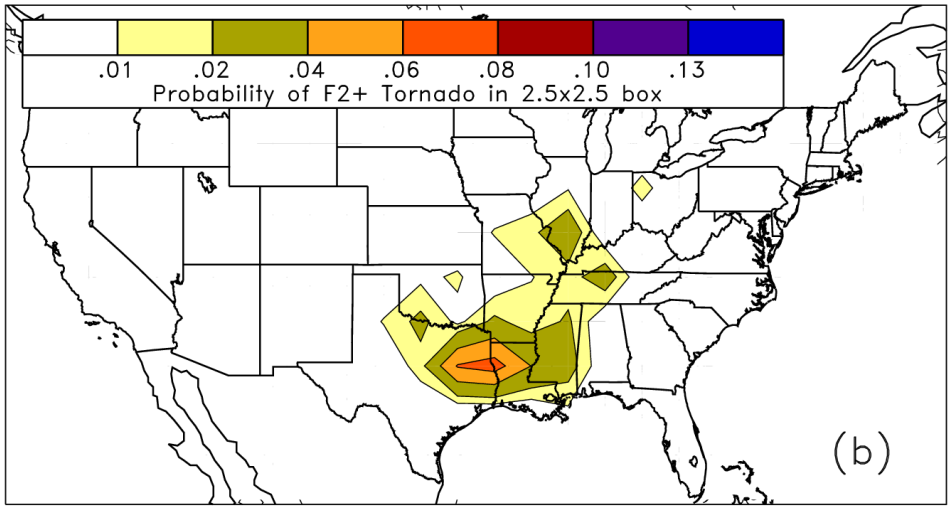
P(F2+ tornado) 0000 UTC 29 May 1980



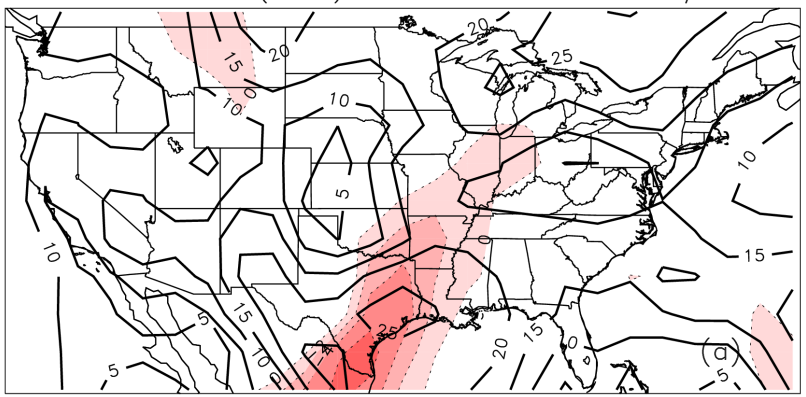
Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 10 May 1981



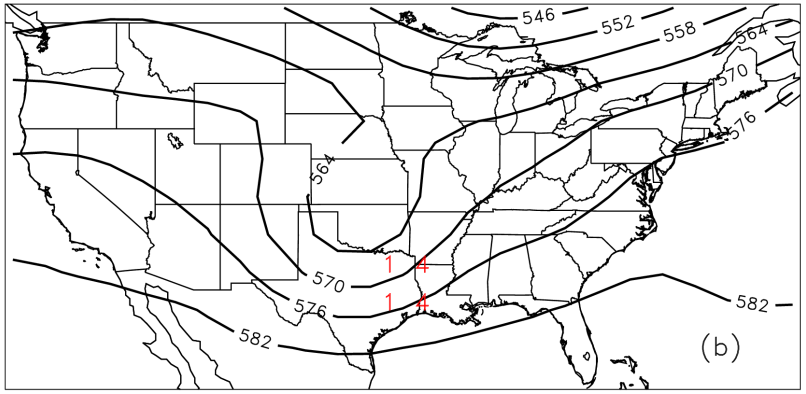
Tornado Probabilities for
01-day Forecast from 09 May 1981



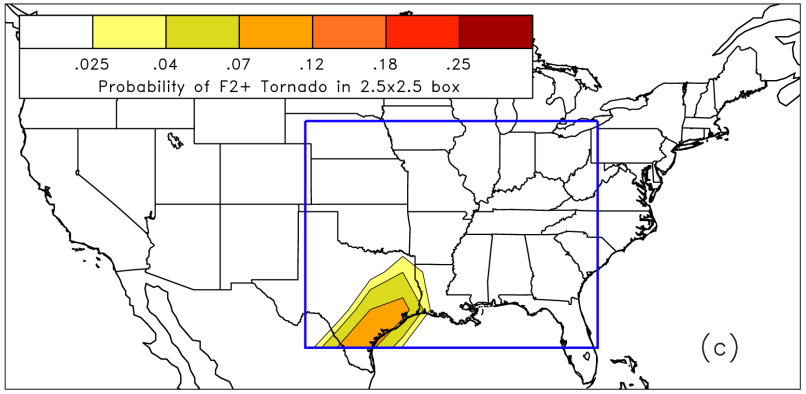
Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 10 May 1981



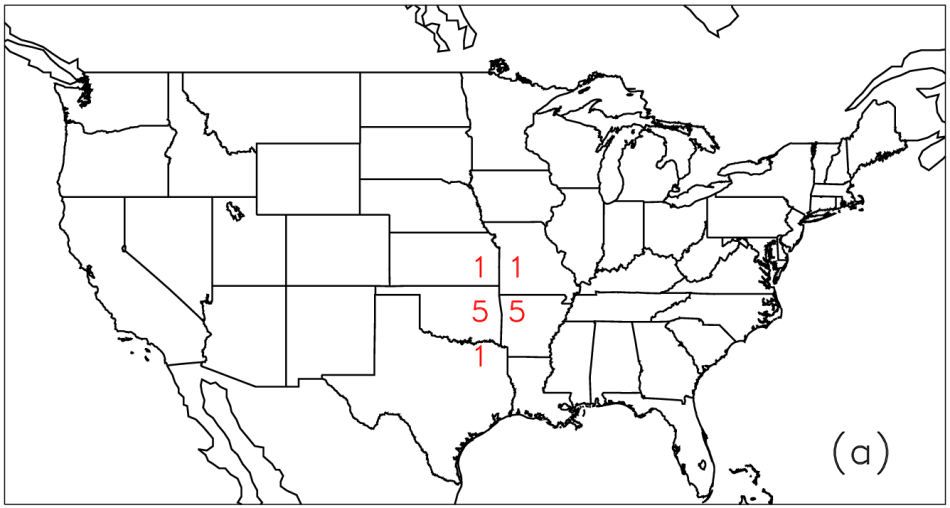
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 10 May 1981



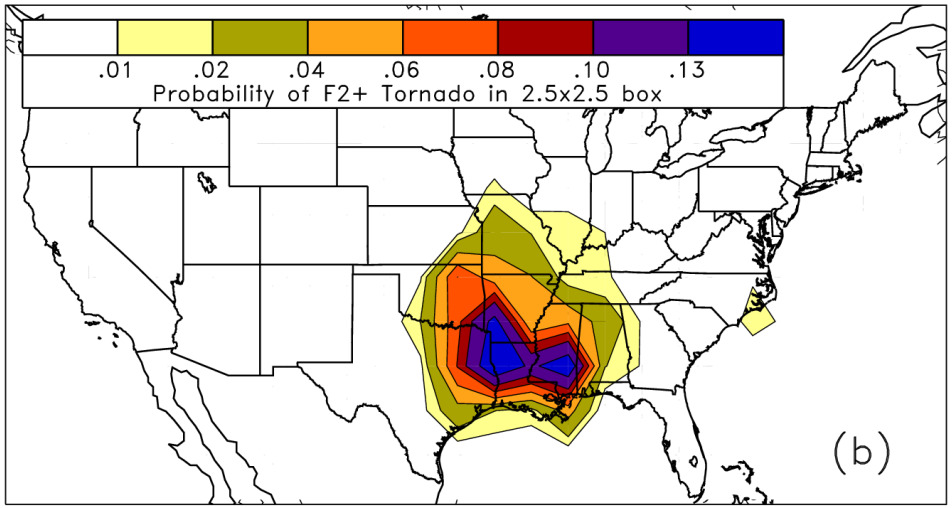
P(F2+ tornado) 0000 UTC 10 May 1981



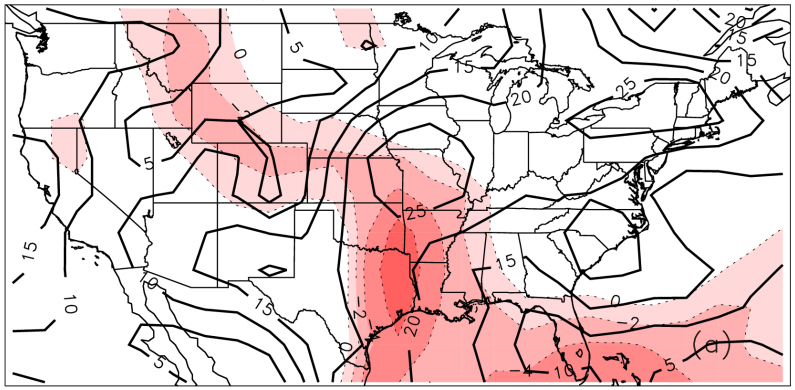
Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 18 May 1981



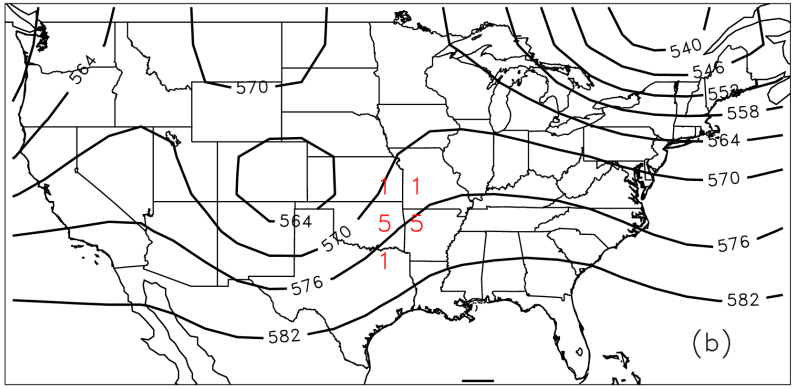
Tornado Probabilities for
01-day Forecast from 17 May 1981



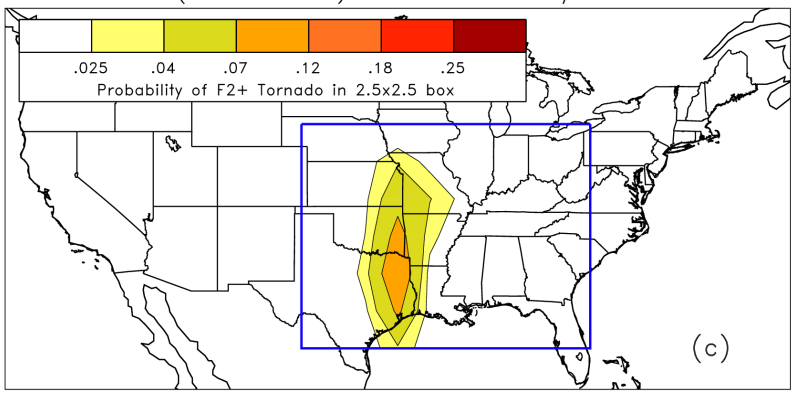
Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 18 May 1981



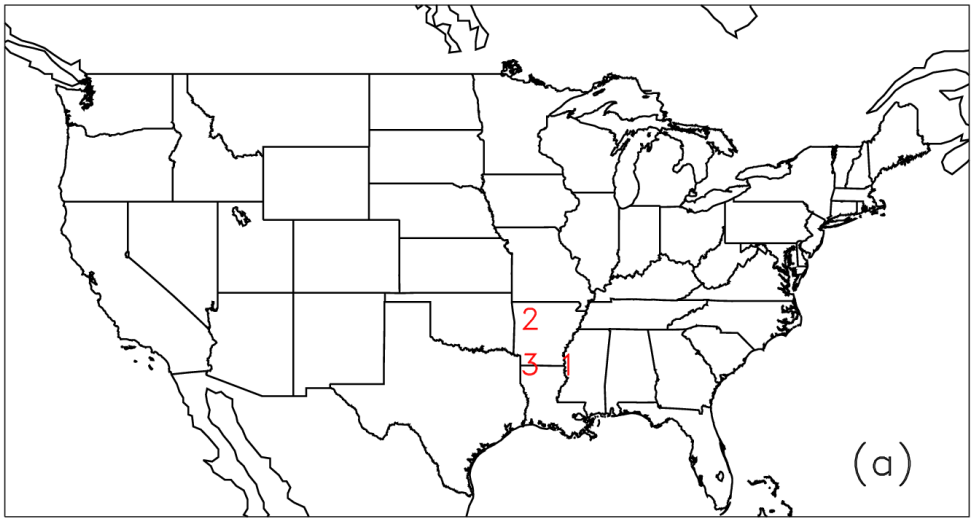
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 18 May 1981



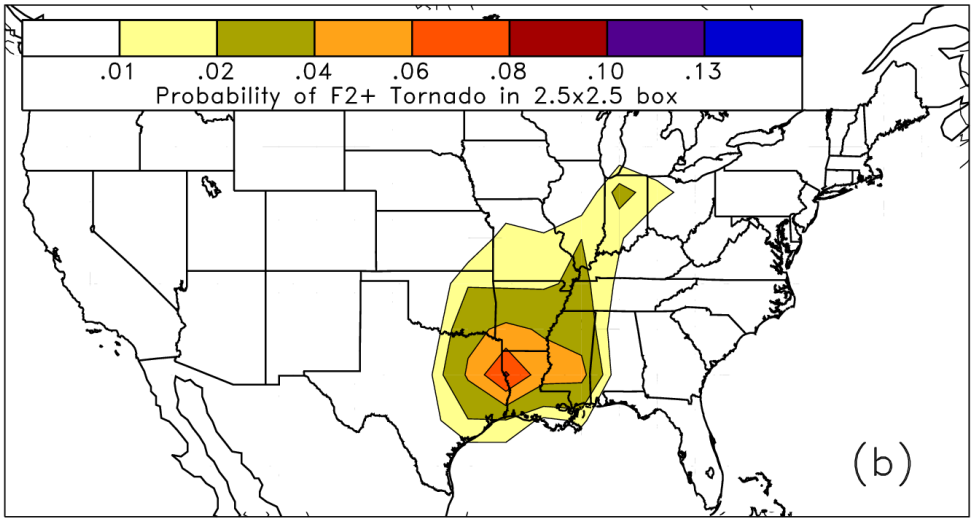
P(F2+ tornado) 0000 UTC 18 May 1981



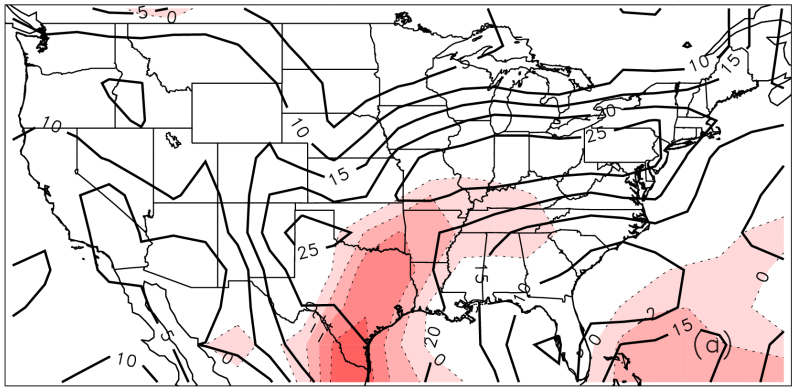
Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 14 May 1981



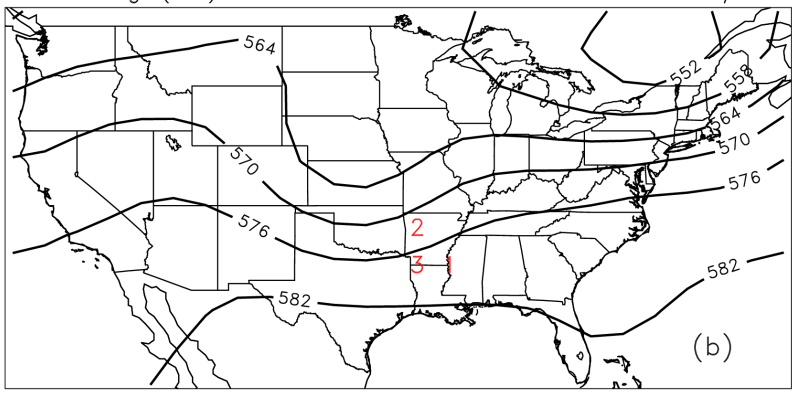
Tornado Probabilities for
01-day Forecast from 13 May 1981



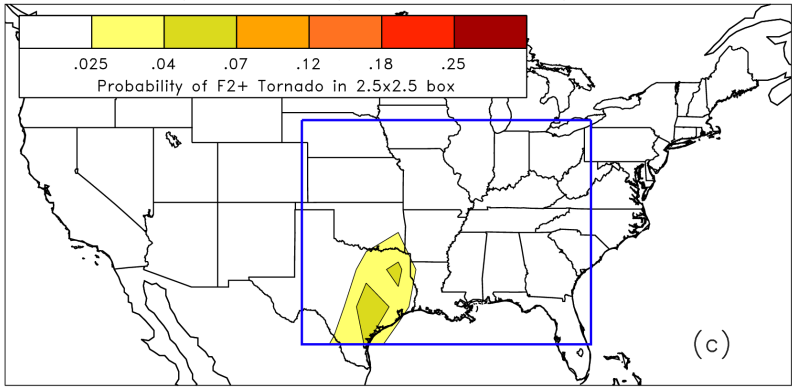
Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 14 May 1981



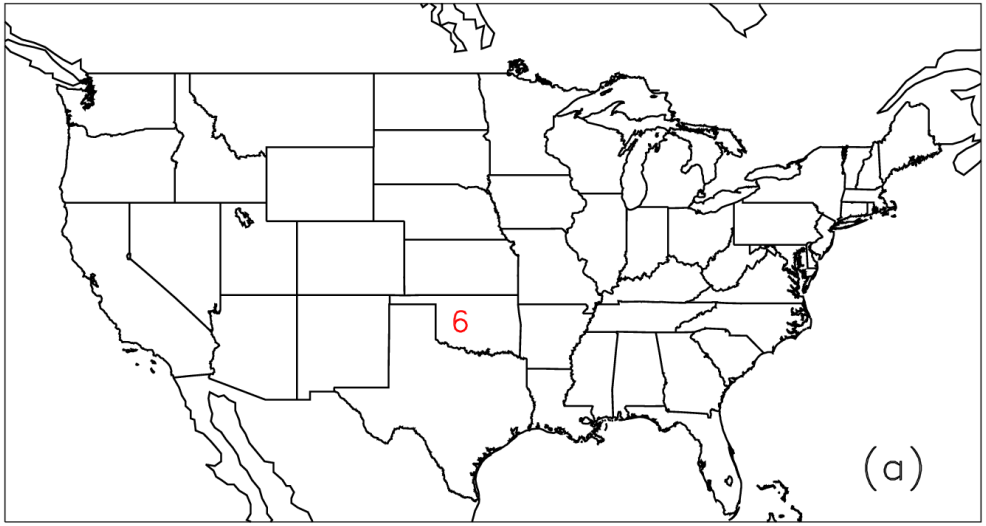
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 14 May 1981



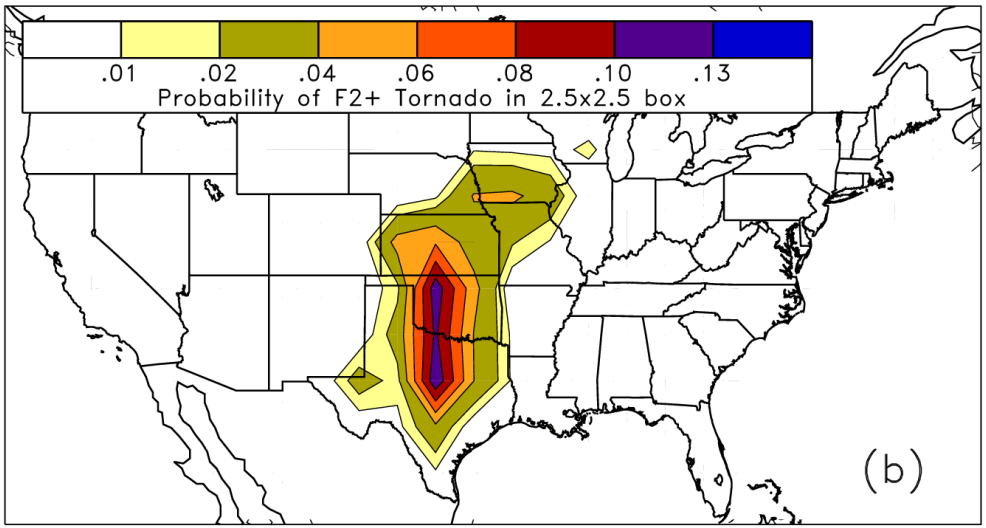
P(F2+ tornado) 0000 UTC 14 May 1981



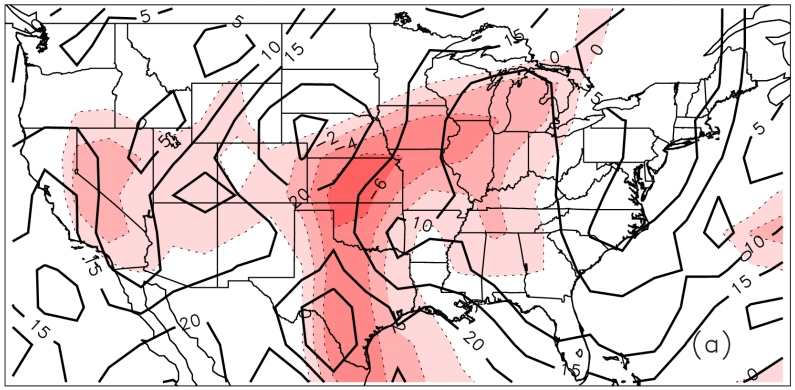
Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 12 May 1982



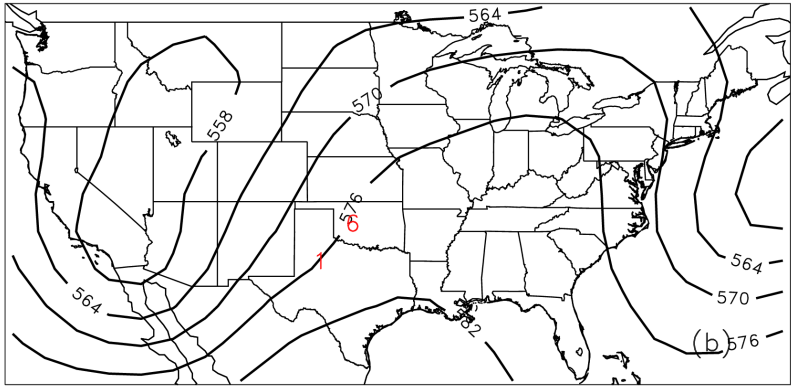
Tornado Probabilities for
01-day Forecast from 11 May 1982



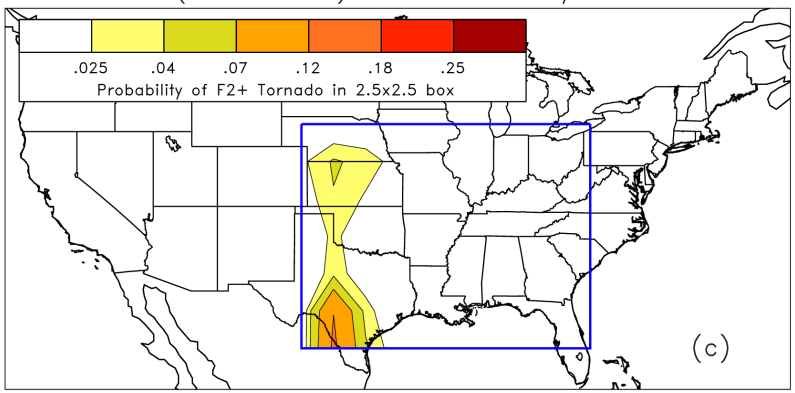
Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 12 May 1982



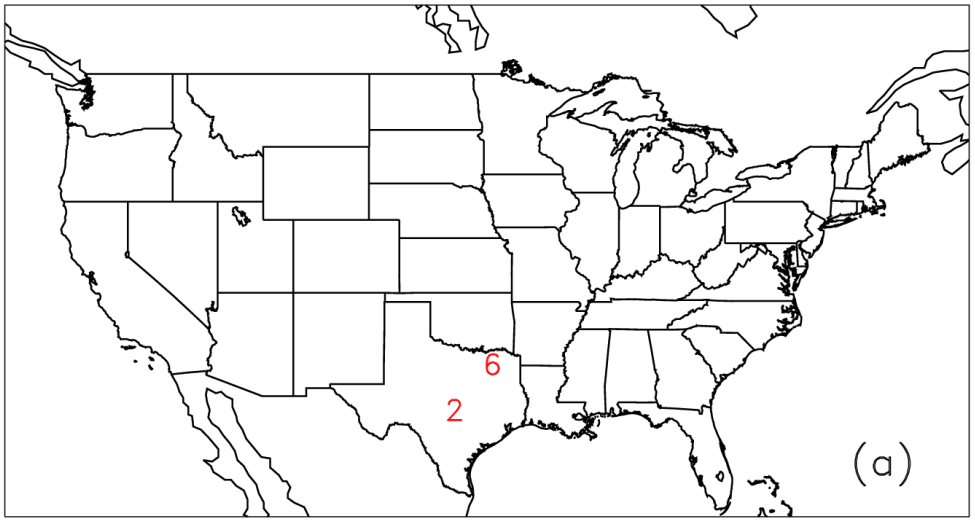
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 12 May 1982



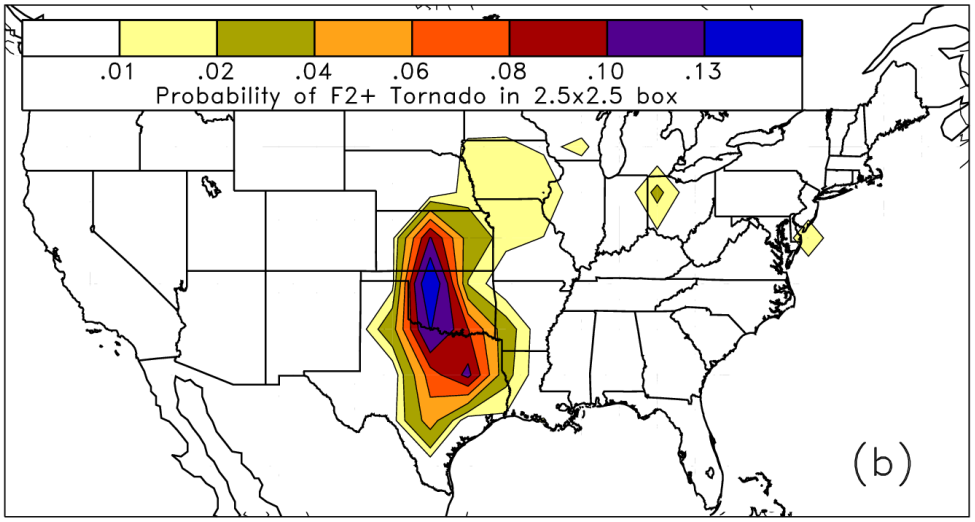
P(F2+ tornado) 0000 UTC 12 May 1982



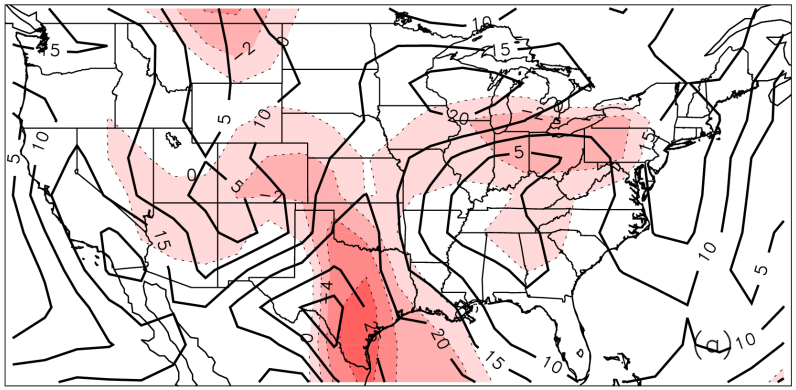
Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 13 May 1982



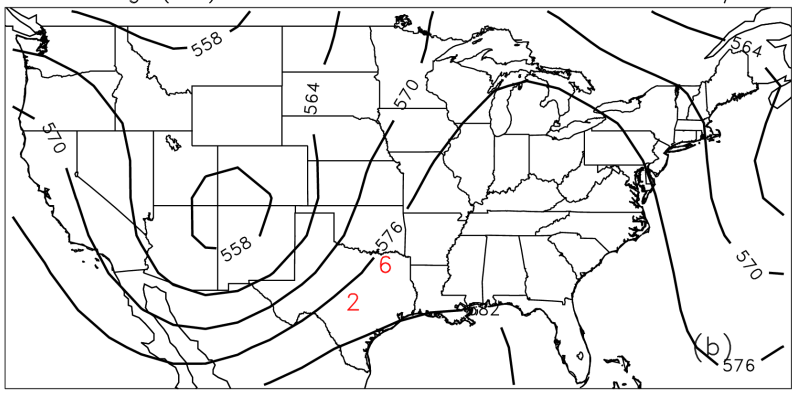
Tornado Probabilities for
01-day Forecast from 12 May 1982



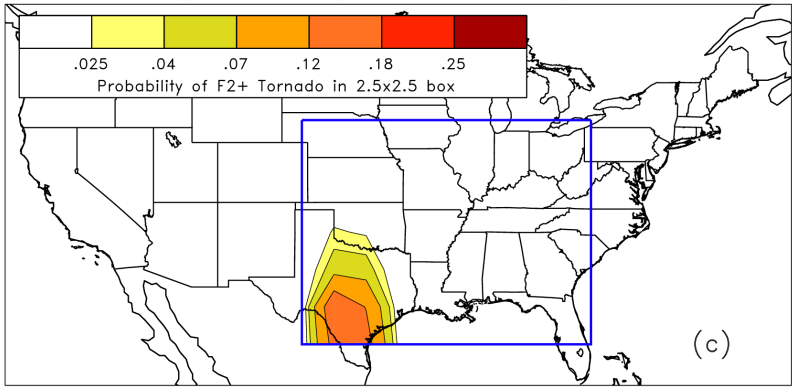
Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 13 May 1982



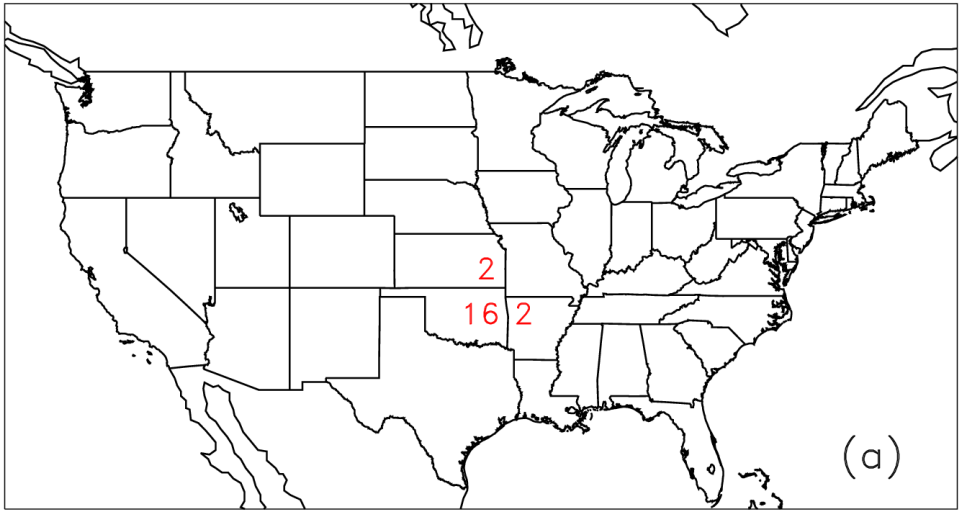
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 13 May 1982



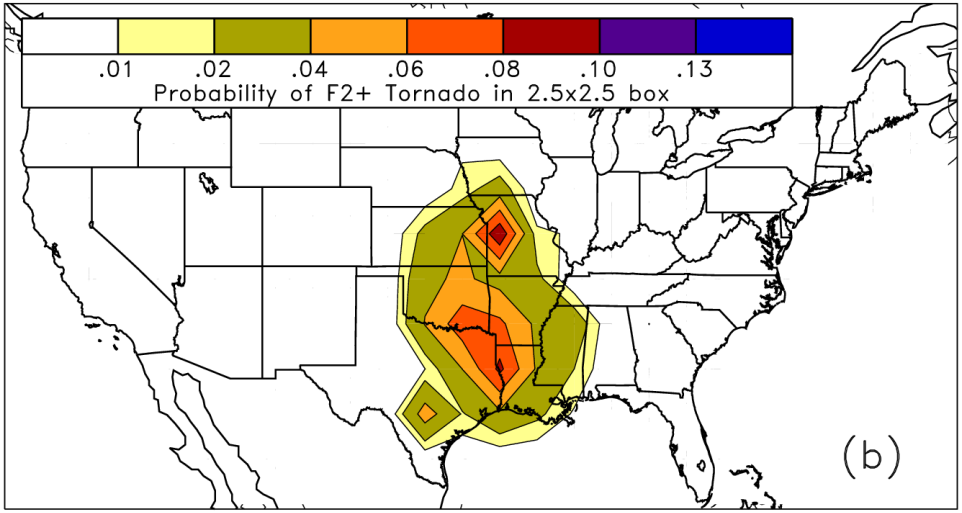
P(F2+ tornado) 0000 UTC 13 May 1982



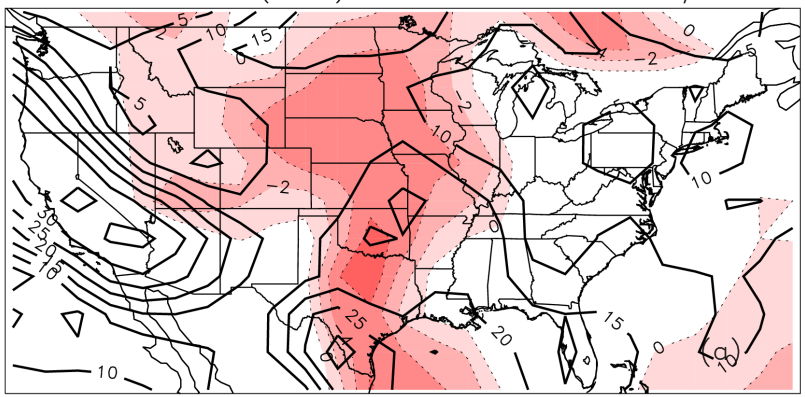
Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 04 May 1999



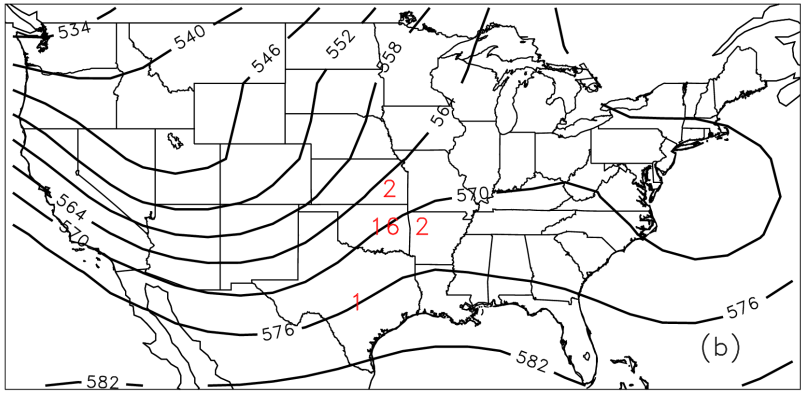
Tornado Probabilities for
01-day Forecast from 03 May 1999



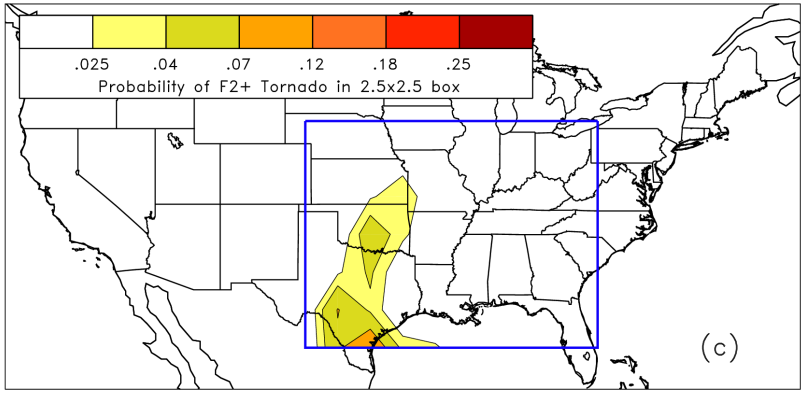
Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 4 May 1999



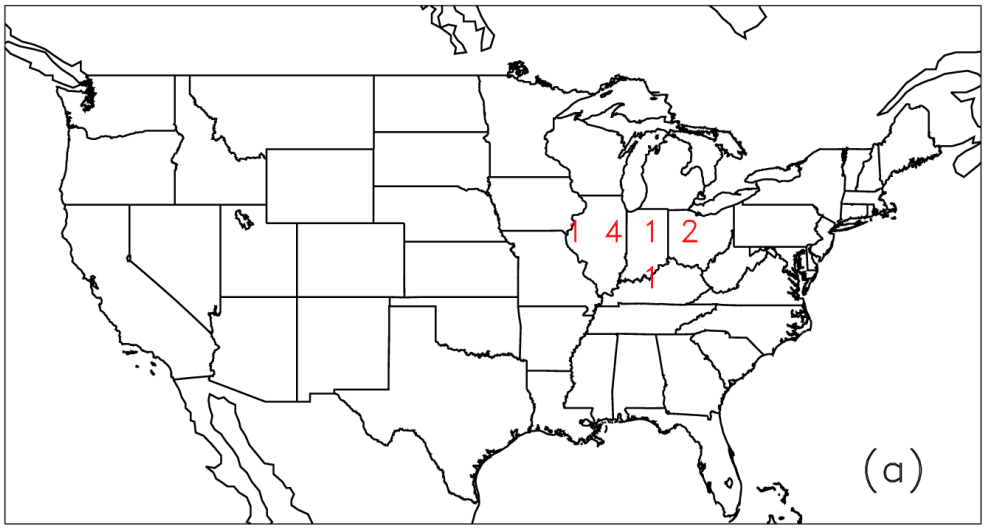
500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 4 May 1999



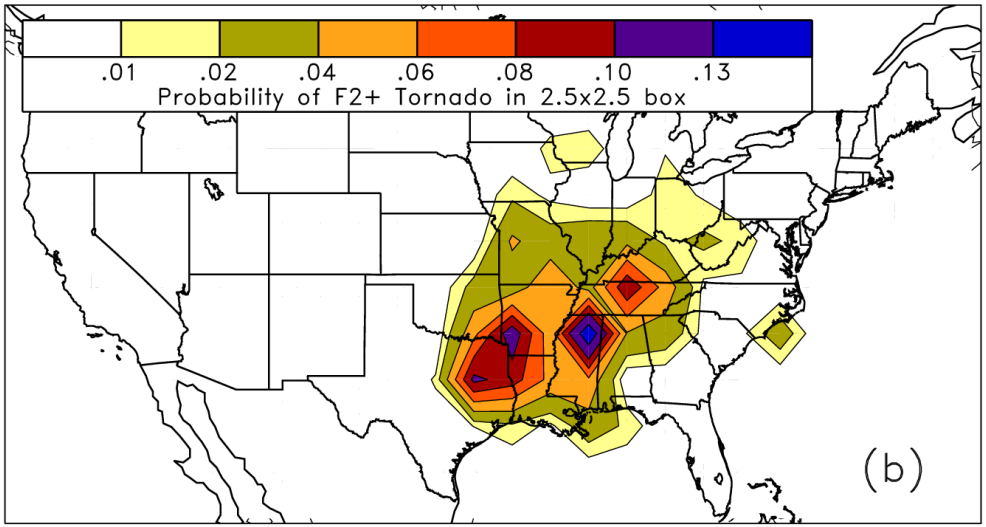
P(F2+ tornado) 0000 UTC 4 May 1999



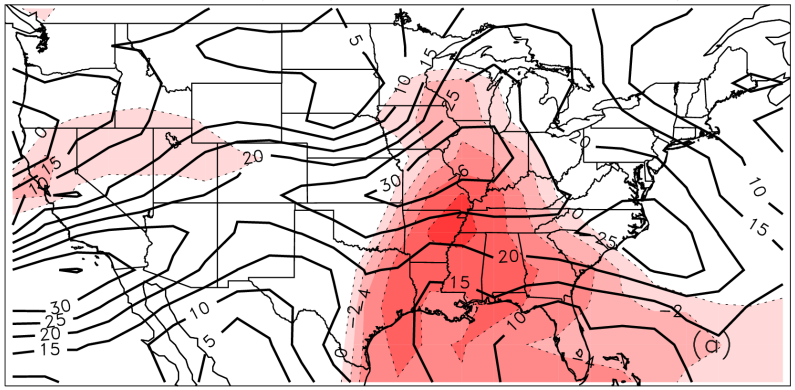
Observed F2+ Tornado Counts in 12-hour Window
Centered on 0000 UTC 14 May 1995



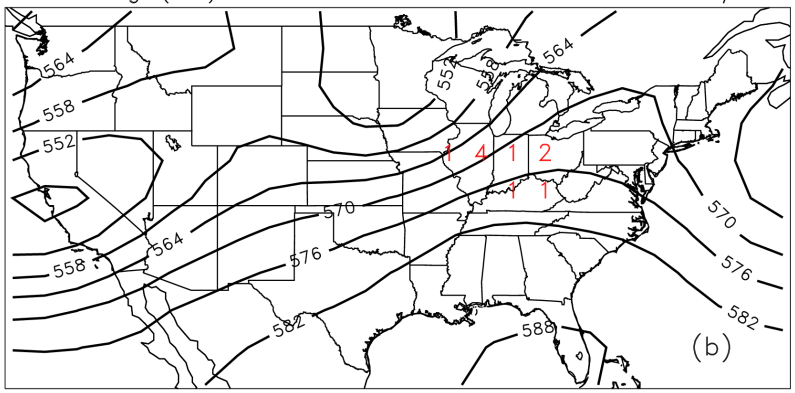
Tornado Probabilities for
01-day Forecast from 13 May 1995



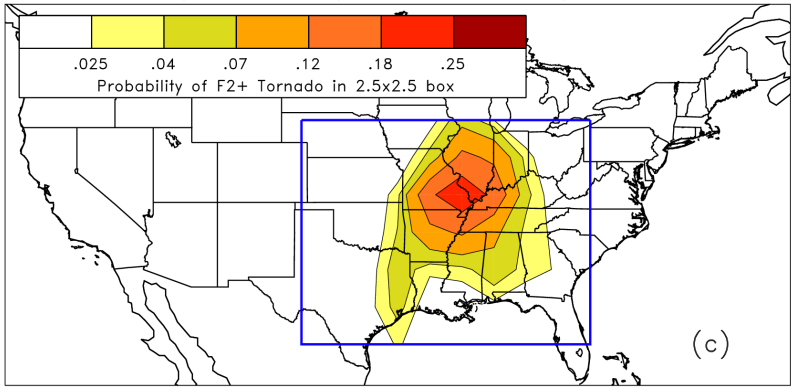
Sfc-500 Shear (ms^{-1}) and LI 0000 UTC 14 May 1995



500 hPa hgt (dm) and F2+ tornado counts 0000 UTC 14 May 1995

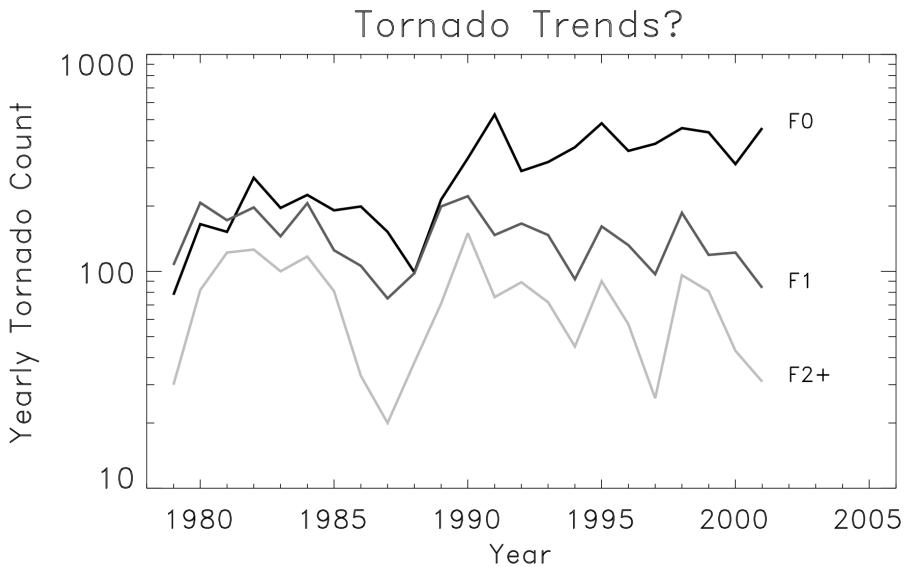


P(F2+ tornado) 0000 UTC 14 May 1995



Other results

- Better to use large e-folding distance, e.g., nearly equal weights for 5x5 box of grid points (not shown).
- CIN useful as predictor.
- *n*=100 much better than *n*=50.
- sfc-650 shear not as good 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+