NOAA Earth System Research Laboratory

How we make ensemble forecasts, and how we verify them

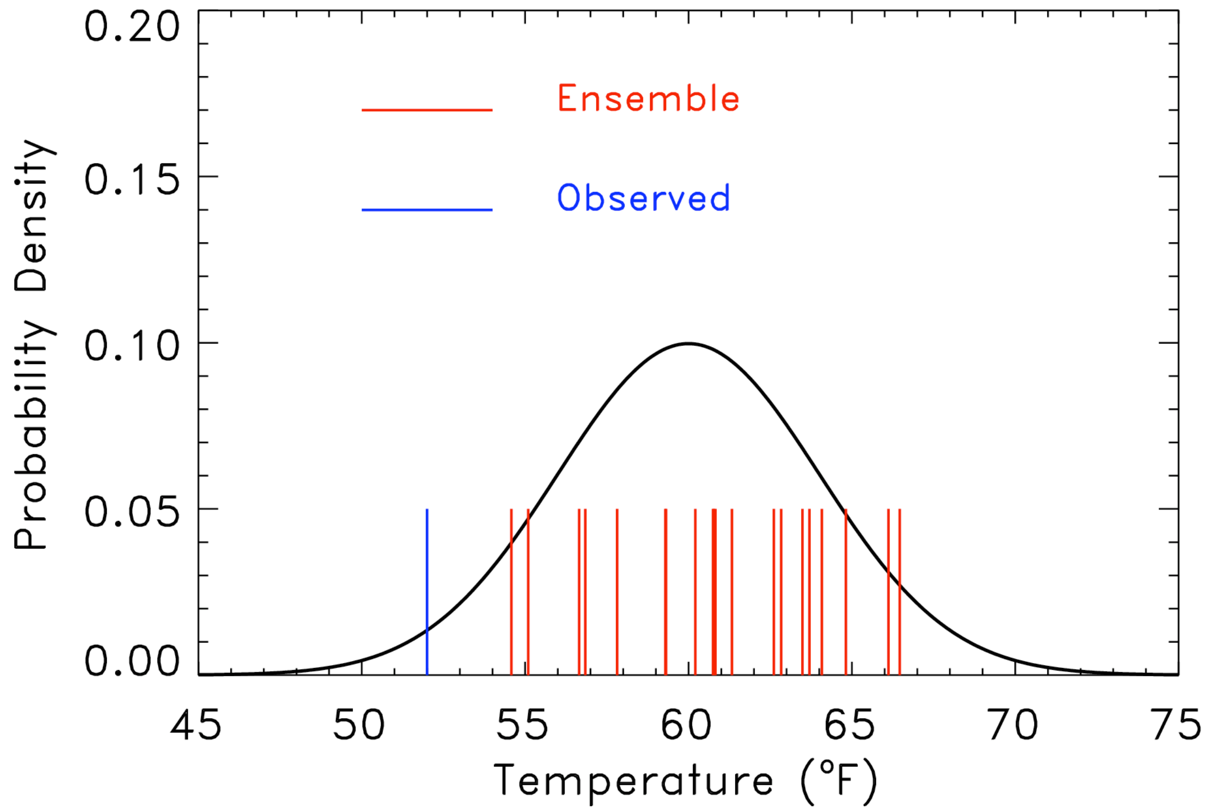
Tom Hamill

NOAA Earth System Research Lab,

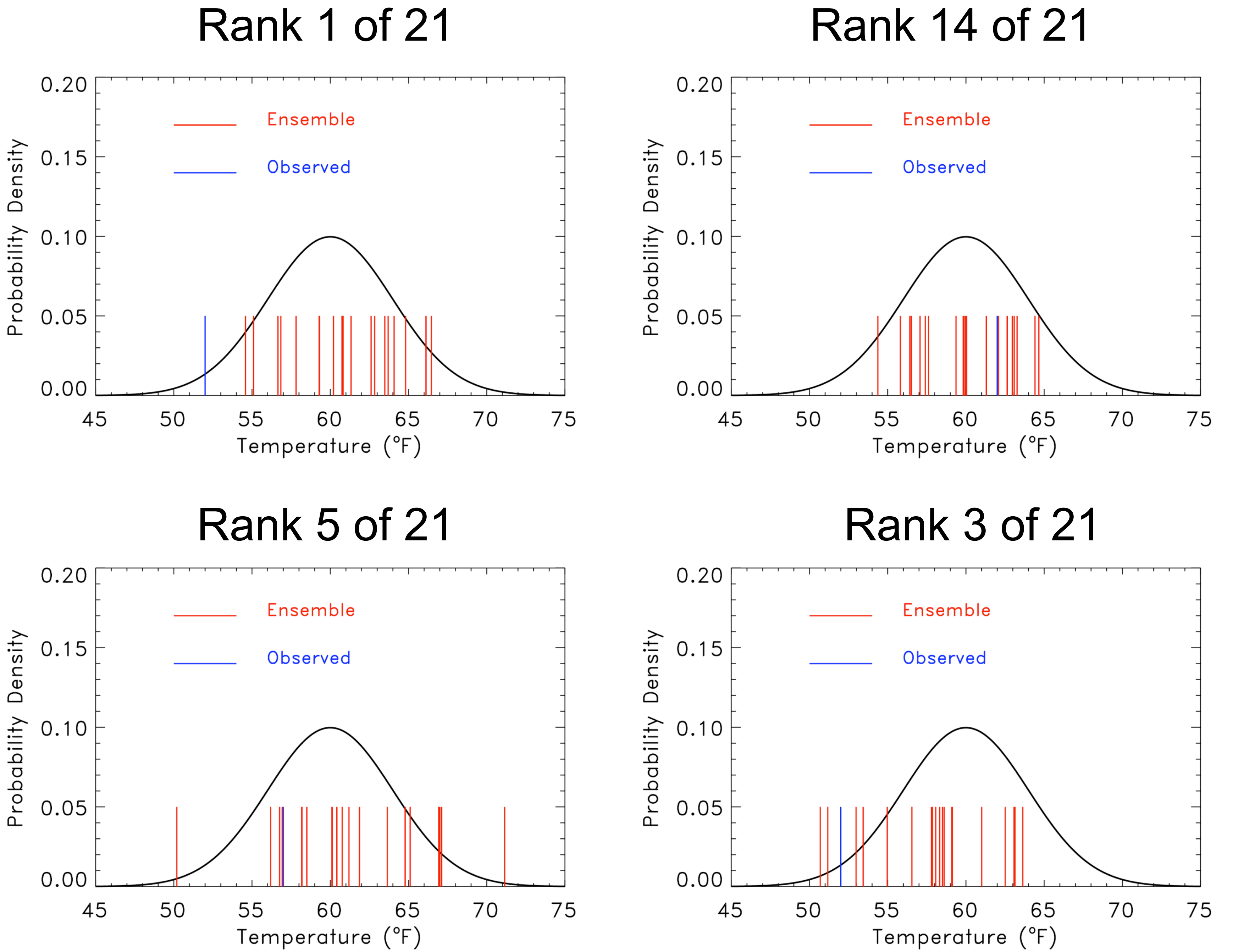
Physical Sciences Division, Boulder, CO

tom.hamill@noaa.gov

What constitutes a “good” ensemble forecast?

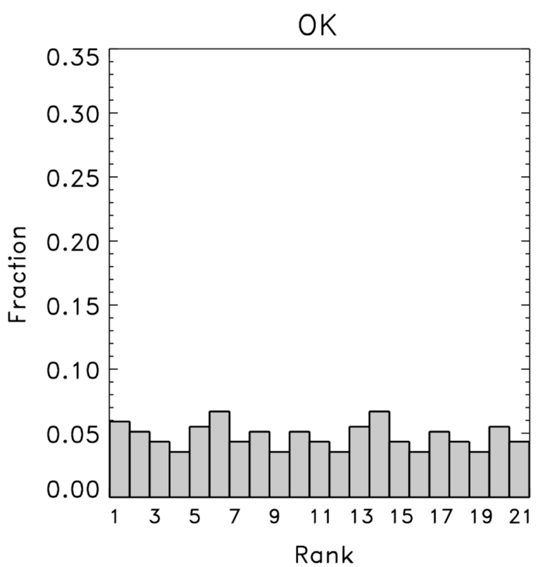


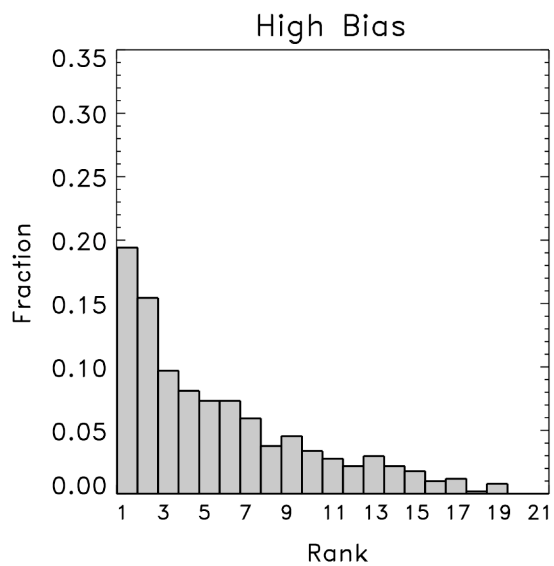
Here, the observed is outside of the range of the ensemble, which was sampled from the pdf shown. Is this a sign of a poor ensemble forecast?

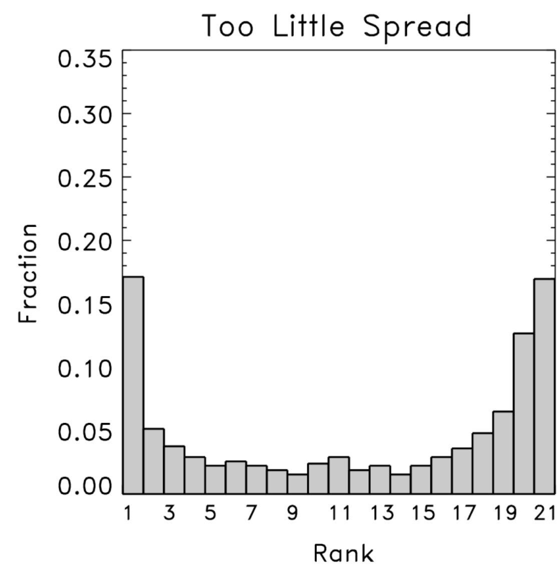

Rank 1 of 21
Rank 14 of 21
Rank 5 of 21
Rank 3 of 21
Ensemble
Observed
Probability Density
Temperature (°F)
0.00
0.05
0.10
0.15
0.20
45
50
55
60
65
70
75

One way of evaluating ensembles: "rank histograms"

We need lots of samples from many situations to evaluate the characteristics of the ensemble.





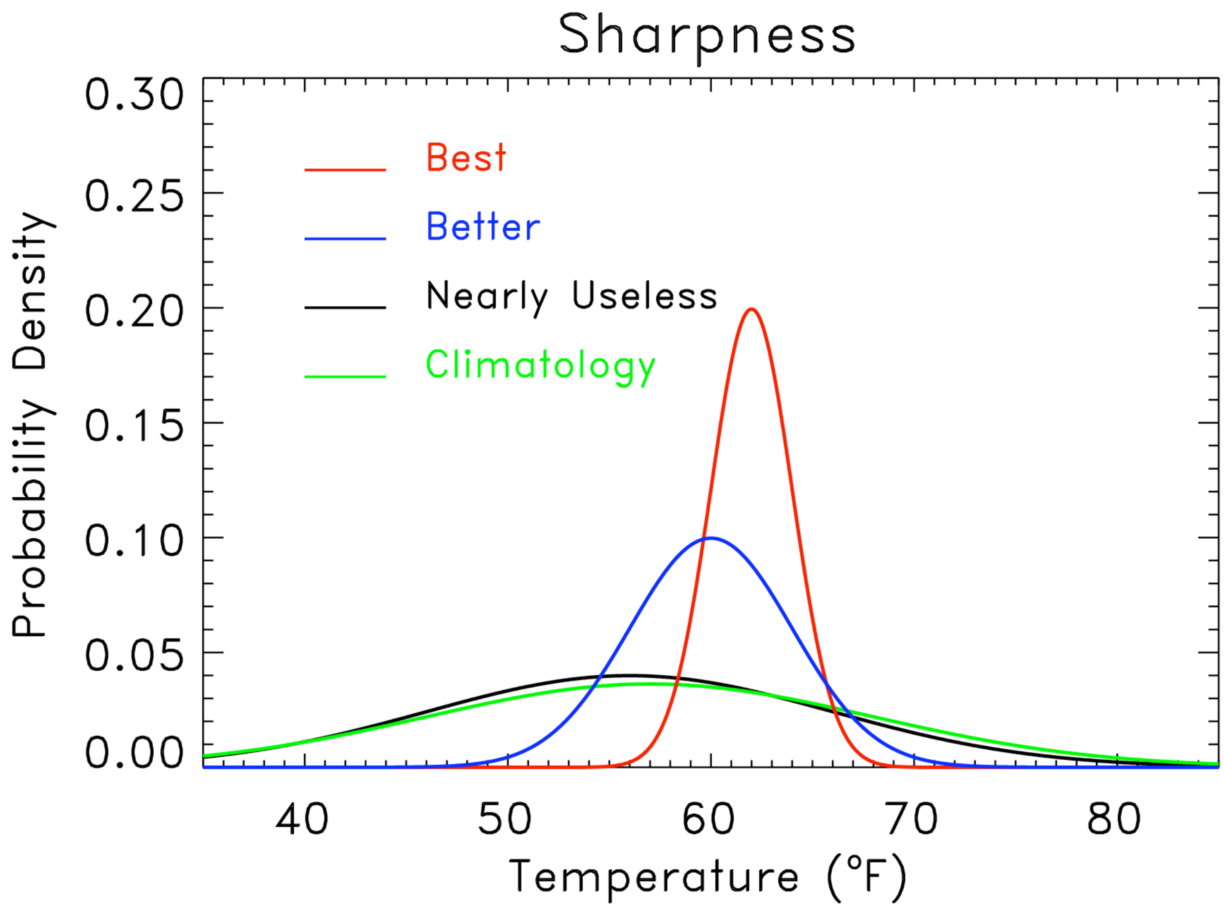


Happens when observed is indistinguishable from any other member of the ensemble. Ensemble is "reliable"

Happens when observed too commonly is lower than the ensemble members.

Happens when there are either some low and some high biases, or when the ensemble doesn't spread out enough.

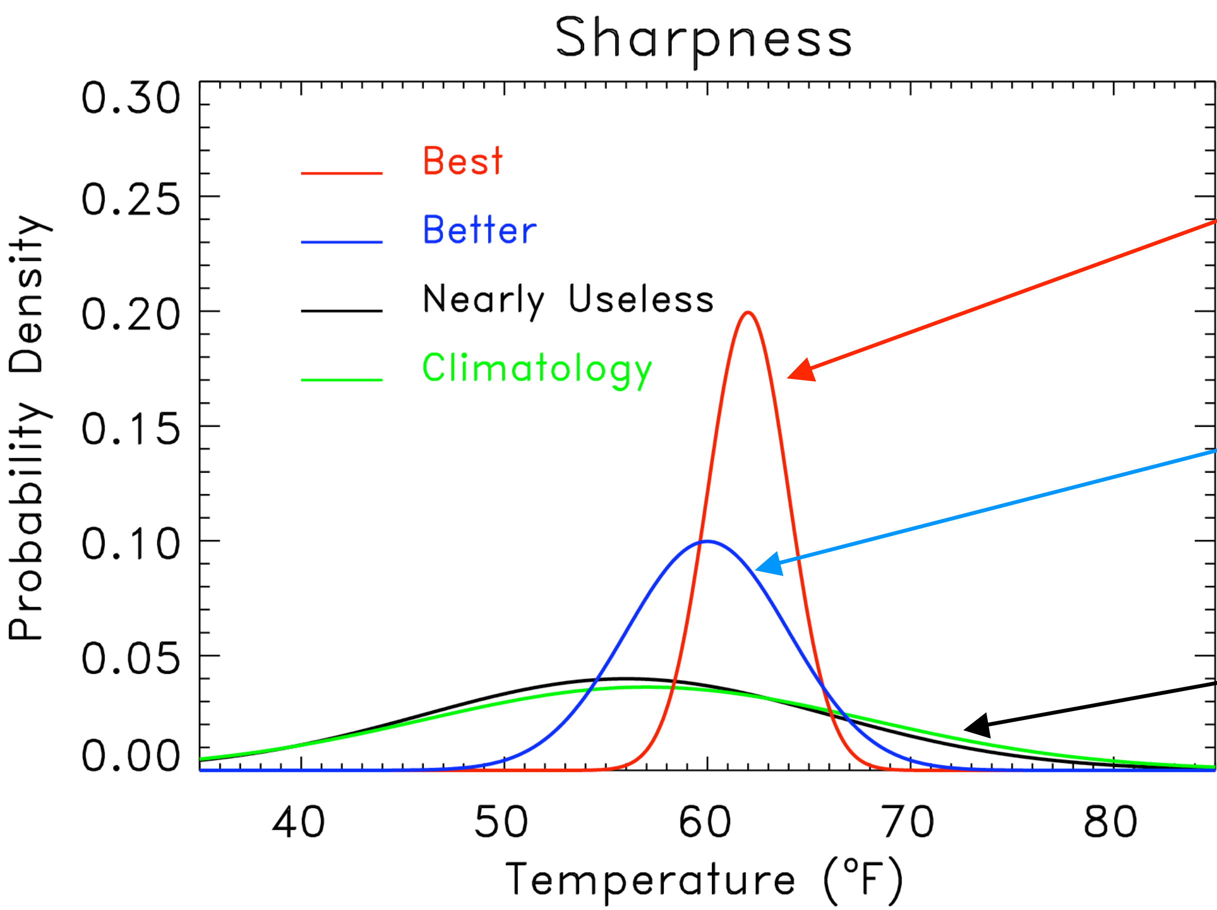
Rank histograms tell us about reliability - but what else is important?



"Sharpness" measures the specificity of the probabilistic forecast. Given two reliable forecast systems, the one producing the sharper forecasts is preferable.

"Spread-skill" relationships are important, too.

Small-spread ensemble forecasts should have less ensemble-mean error than large-spread forecasts.



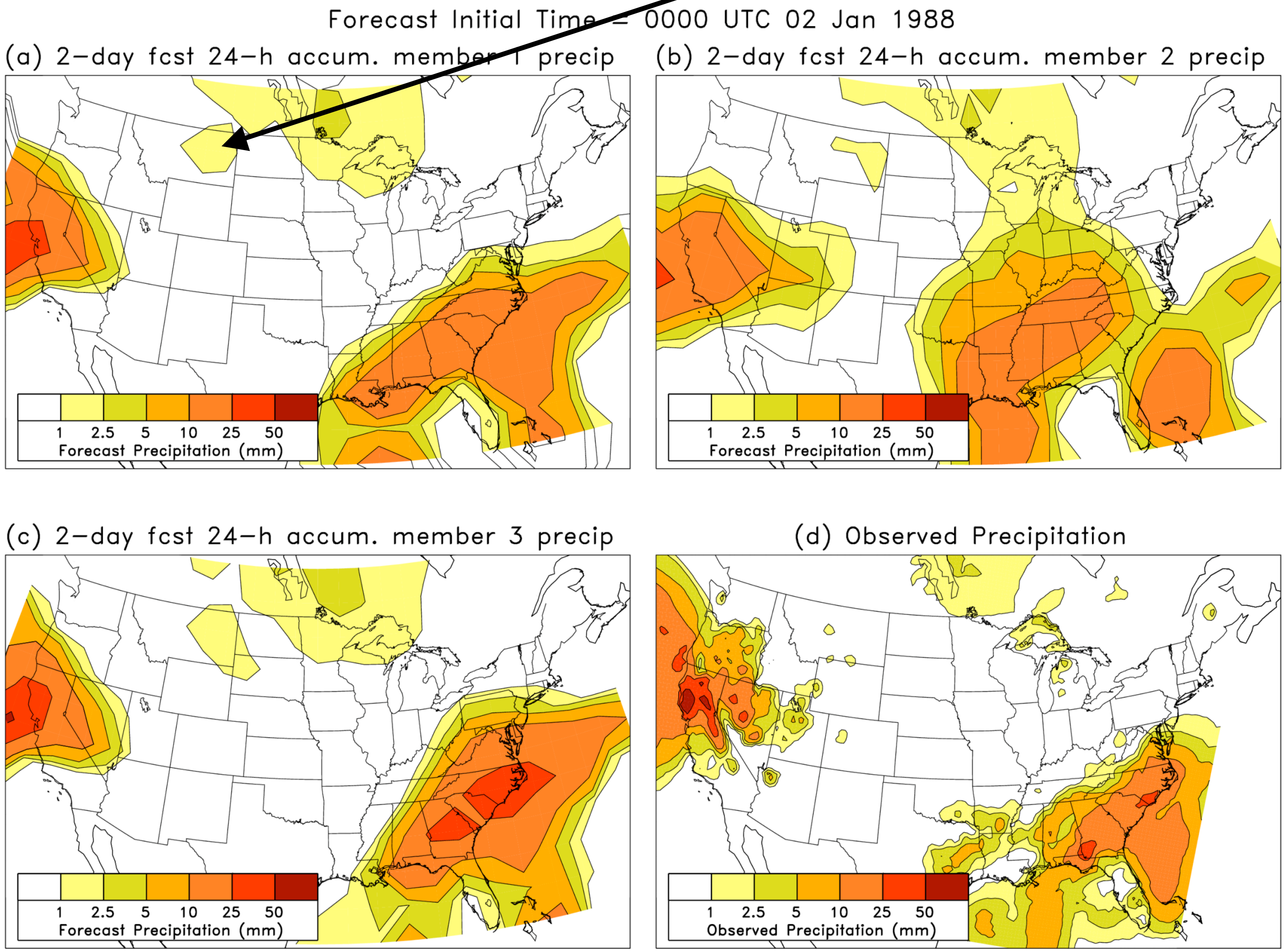
ensemble-mean error from a sample of this pdf should be low.

ensemble-mean error should be moderate

ensemble-mean error should be large

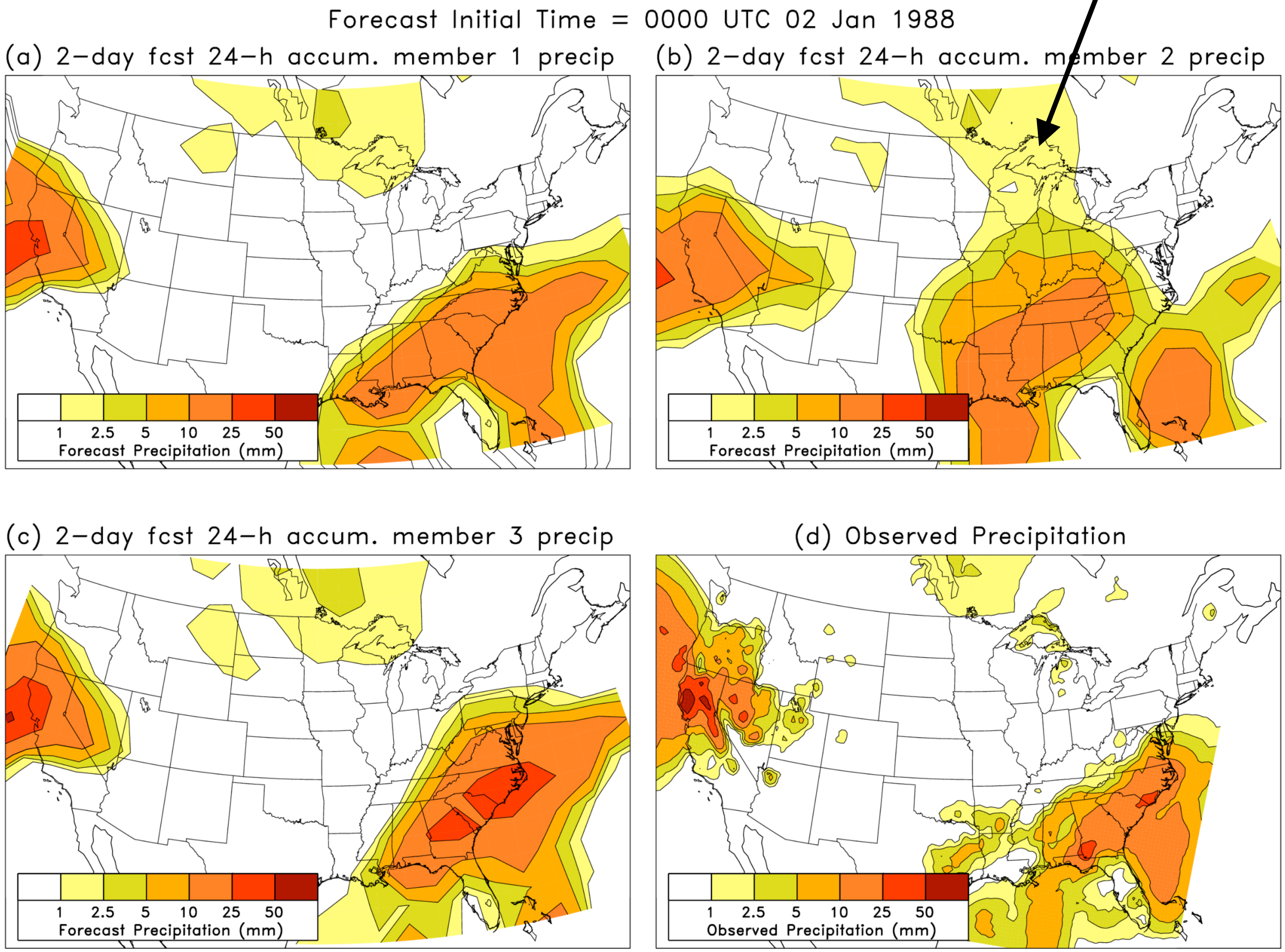
An example of overpopulating the extreme ranks of the histogram

ensemble members too similar to each other.

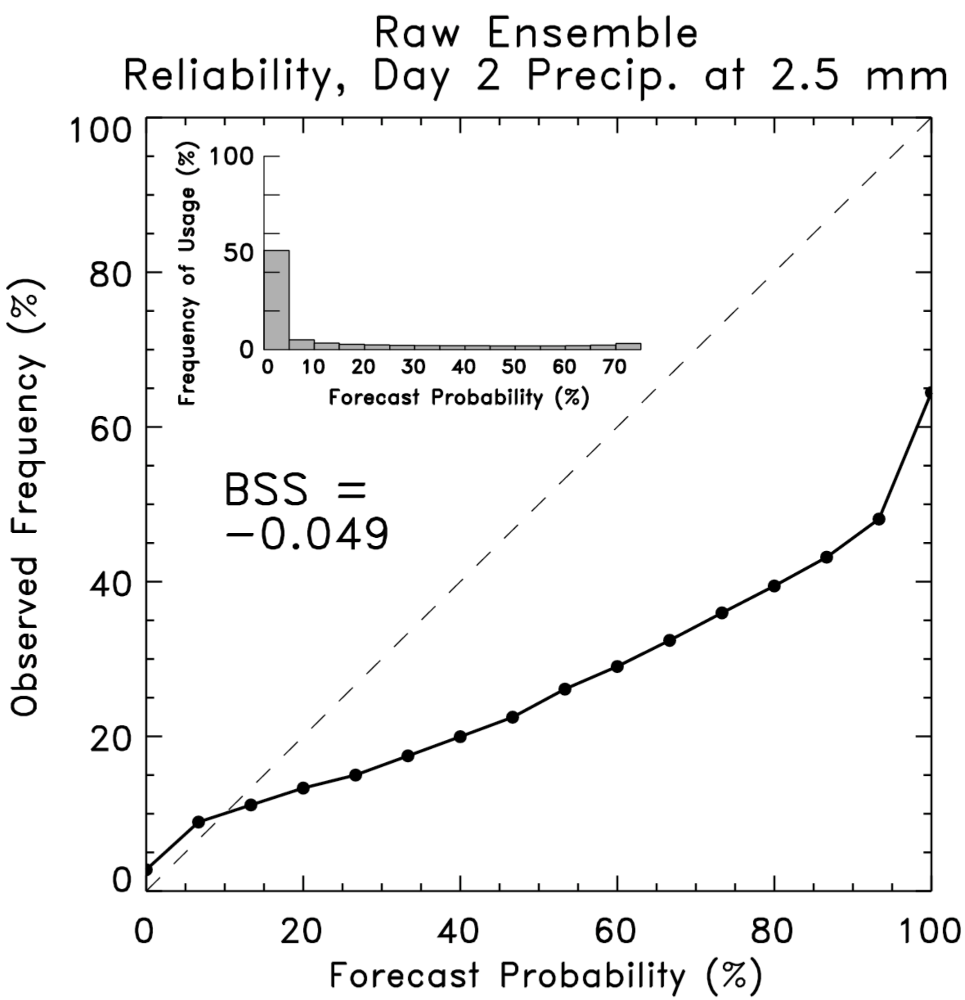


Seen another way, this could be considered a model bias …

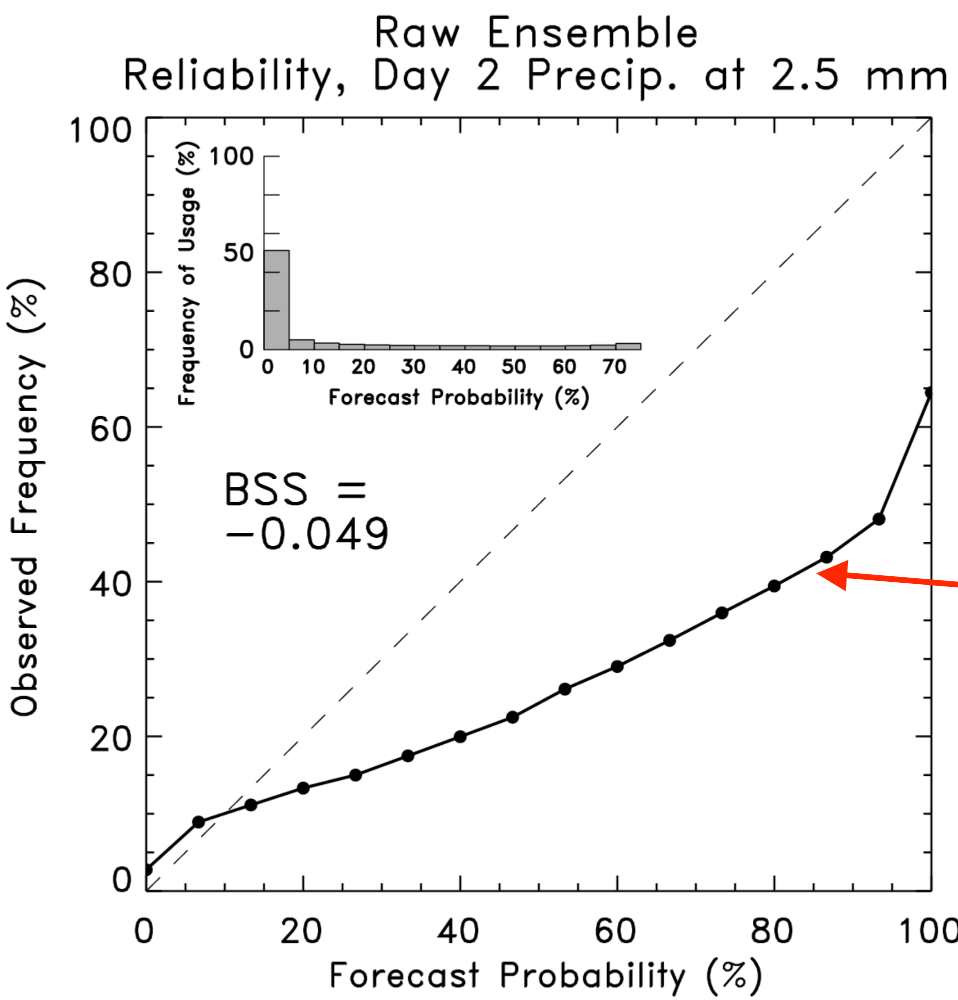
drizzle over-forecast;
too much yellow in forecast plots



“Reliability Diagrams”

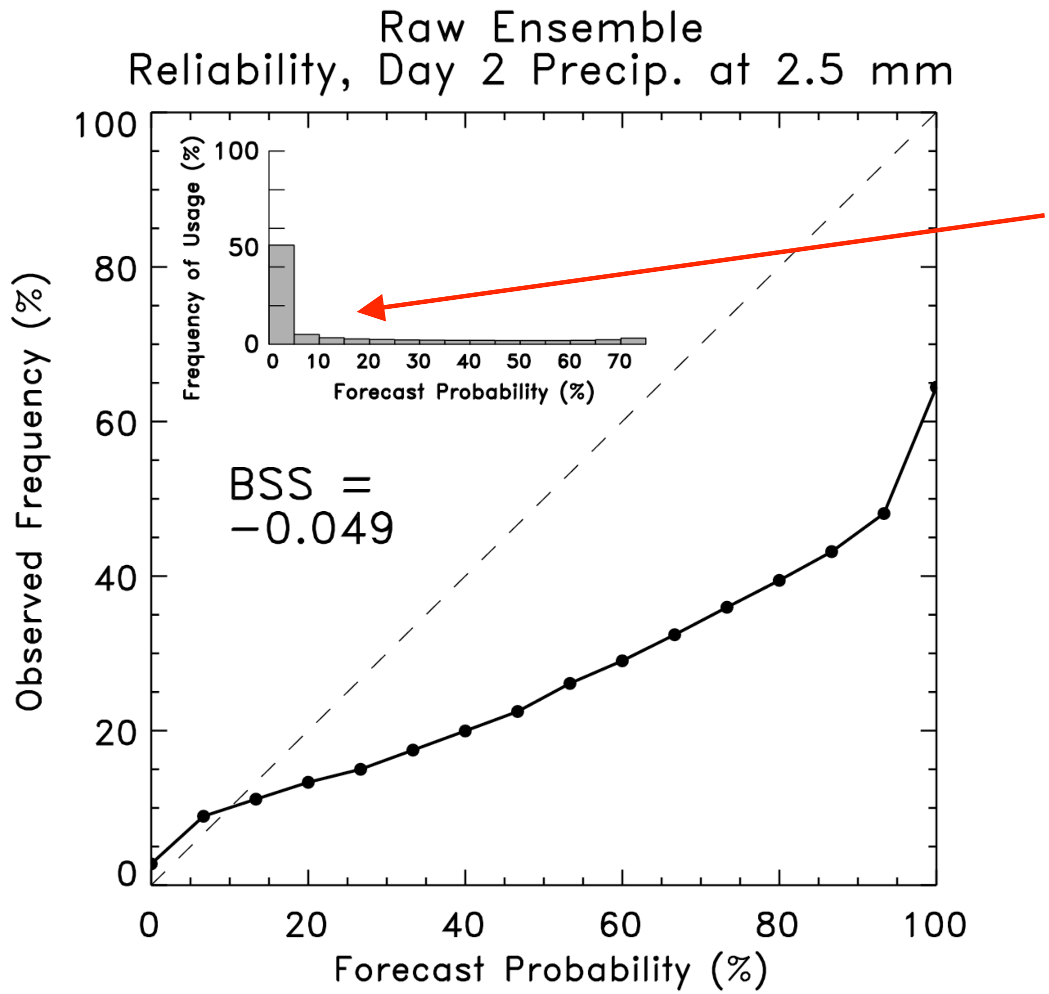


“Reliability Diagrams”



curve tells you what the observed frequency was each time you forecast a given probability. This curve ought to lie along $y = x$ line.

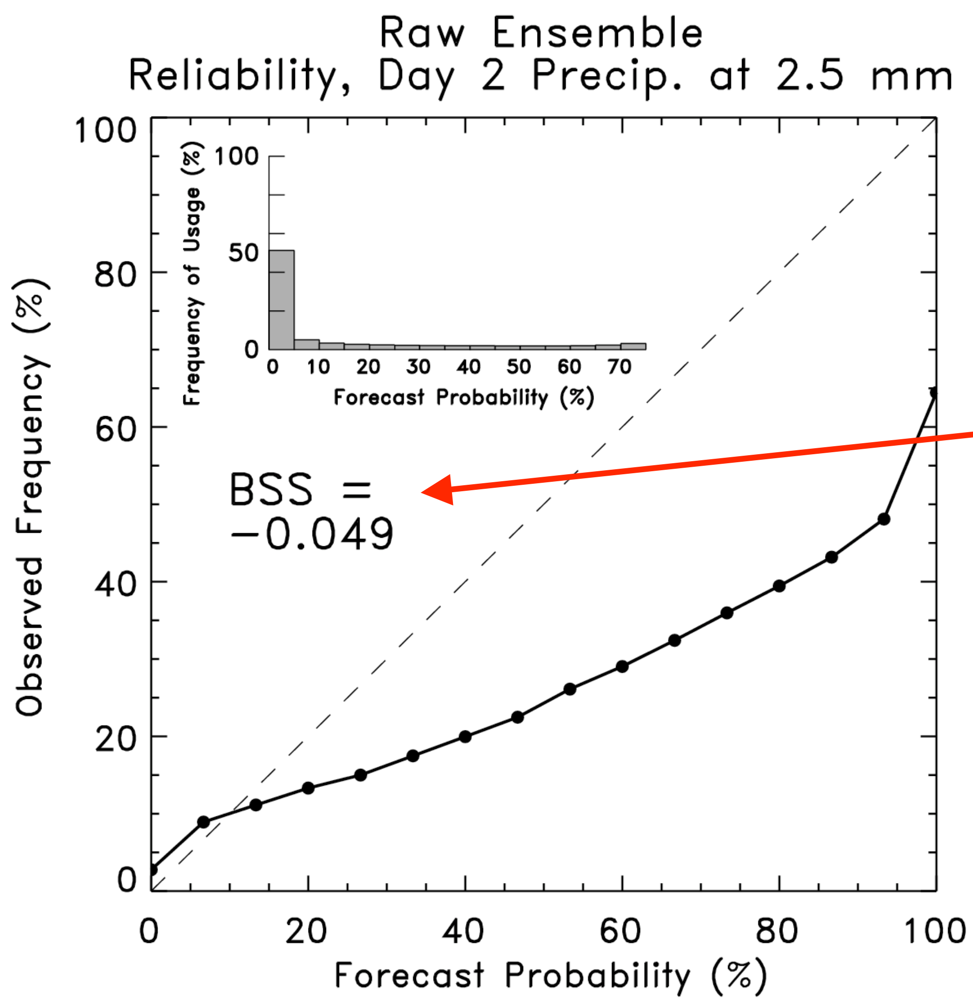
"Reliability Diagrams"



inset histogram tells you how frequently each probability was issued.

Perfectly sharp: frequency of usage populates only 0% and 100%.

"Reliability Diagrams"



BSS = Brier Skill Score

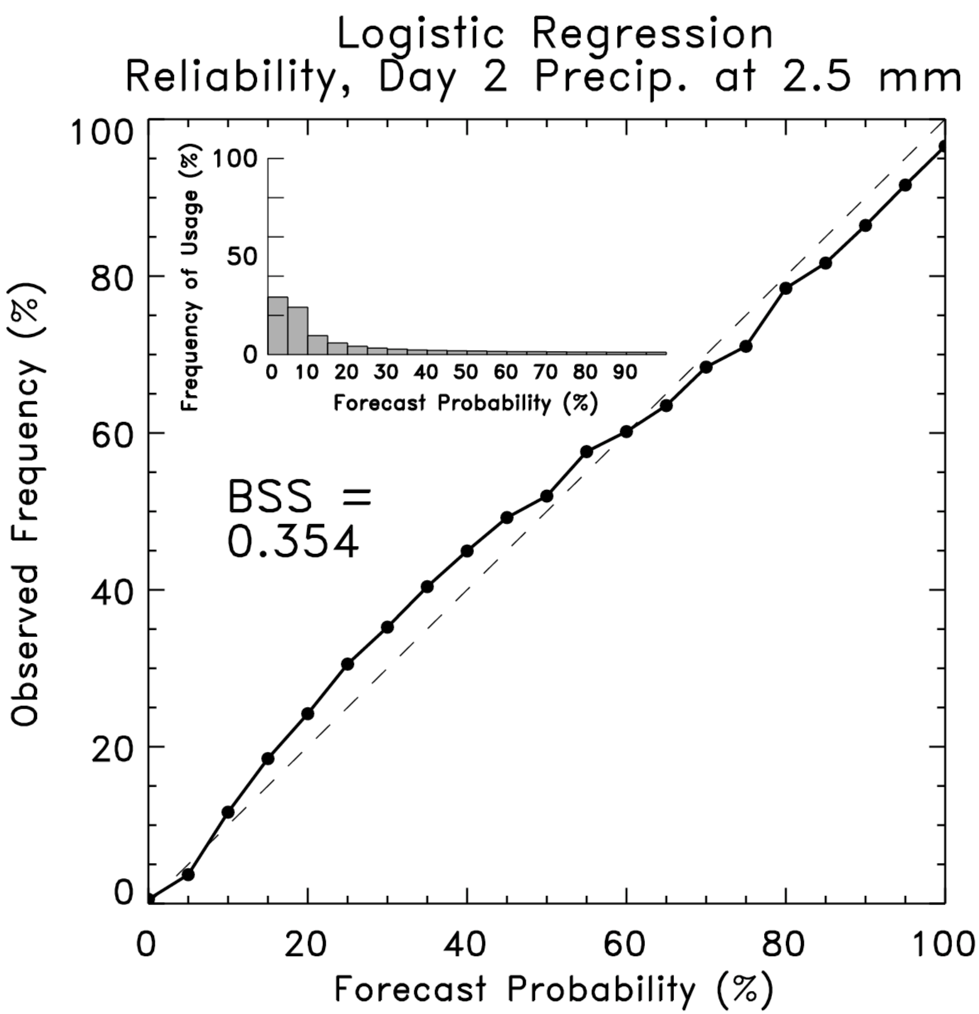
$$\boxed{BSS = \frac{BS(CLimo) - BS(Forecast)}{BS(CLimo)}}$$

BS(•) measures the Brier Score, which you can think of as the squared error of a probabilistic forecast.

Perfect: *BSS* = 1.0

Climatology: *BSS* = 0.0

Reliability after post-processing



Statistical correction of forecasts using a long, stable set of prior forecasts from the same model (like in MOS).

More in Maj. Tony Eckel's talk.

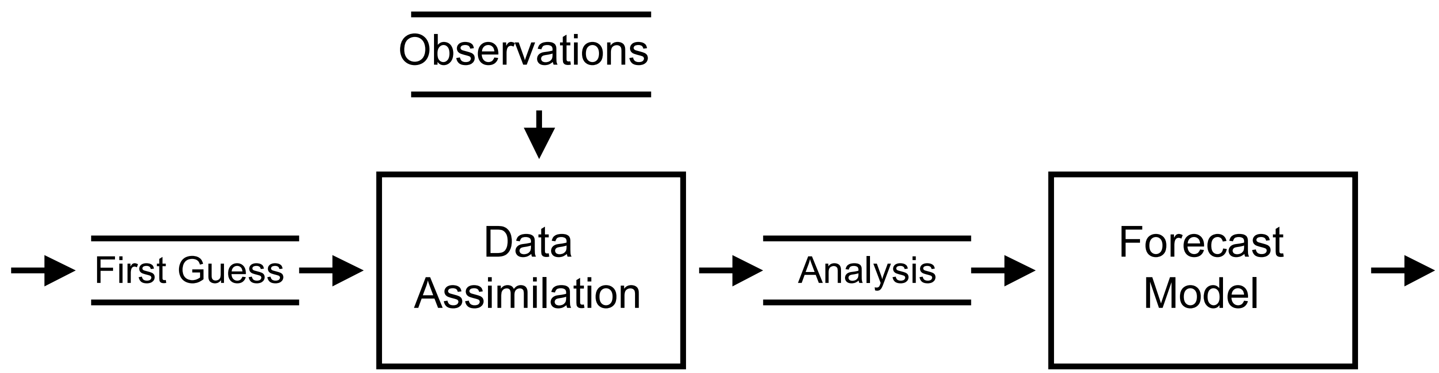
Ensemble forecast system problems and cures

- Problems:
  - Biases.
  - Spread deficiencies
  - Lack of spread / skill relationship.
  - Etc.
- Cures
  - Design a better ensemble forecast system (this talk)
    - Improved methods of generating initial conditions
    - Methods for dealing with model errors
  - Post-process the ensemble forecasts to ameliorate these errors (Maj. Tony Eckel's talk)

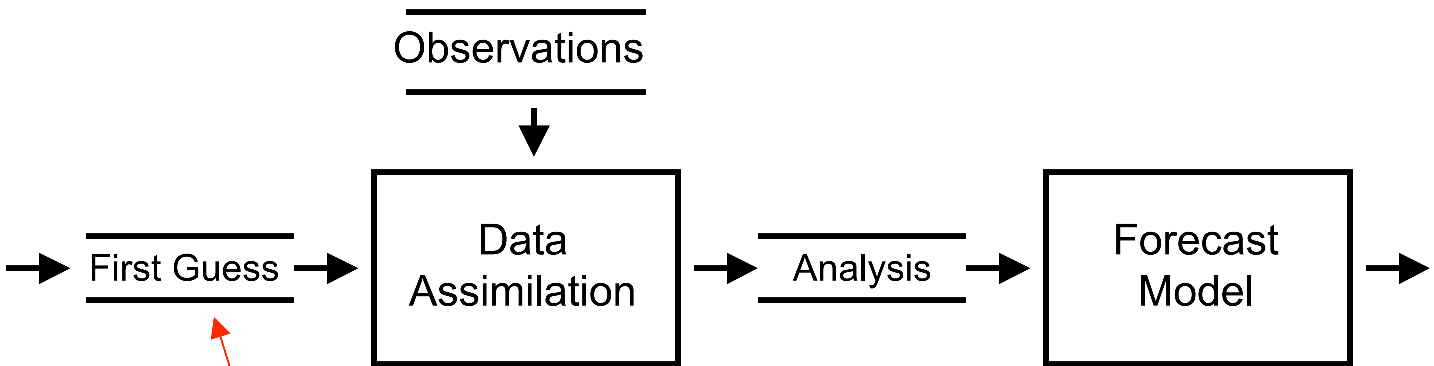
Ensemble forecast system problems and cures

- Problems:
  - Biases.
  - Spread deficiencies
  - Lack of spread / skill relationship.
  - Etc.
- Cures
  - Design a better ensemble forecast system (this talk)
    - Improved methods of generating initial conditions
    - Methods for dealing with model errors
  - Post-process the ensemble forecasts to ameliorate these errors (Maj. Tony Eckel's talk)

Understanding analysis error characteristics



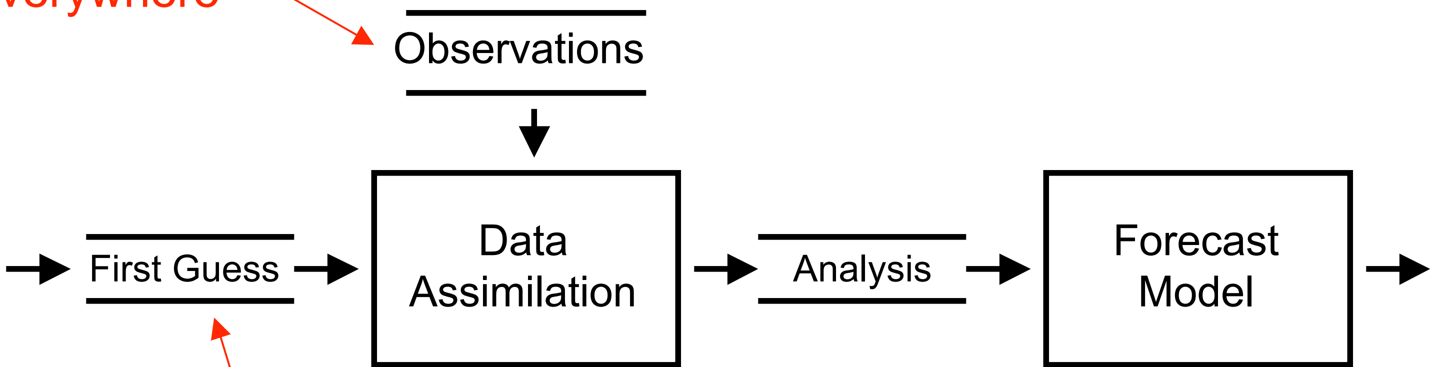
Understanding analysis error characteristics



this will inevitably have some errors, else why assimilate new observations?

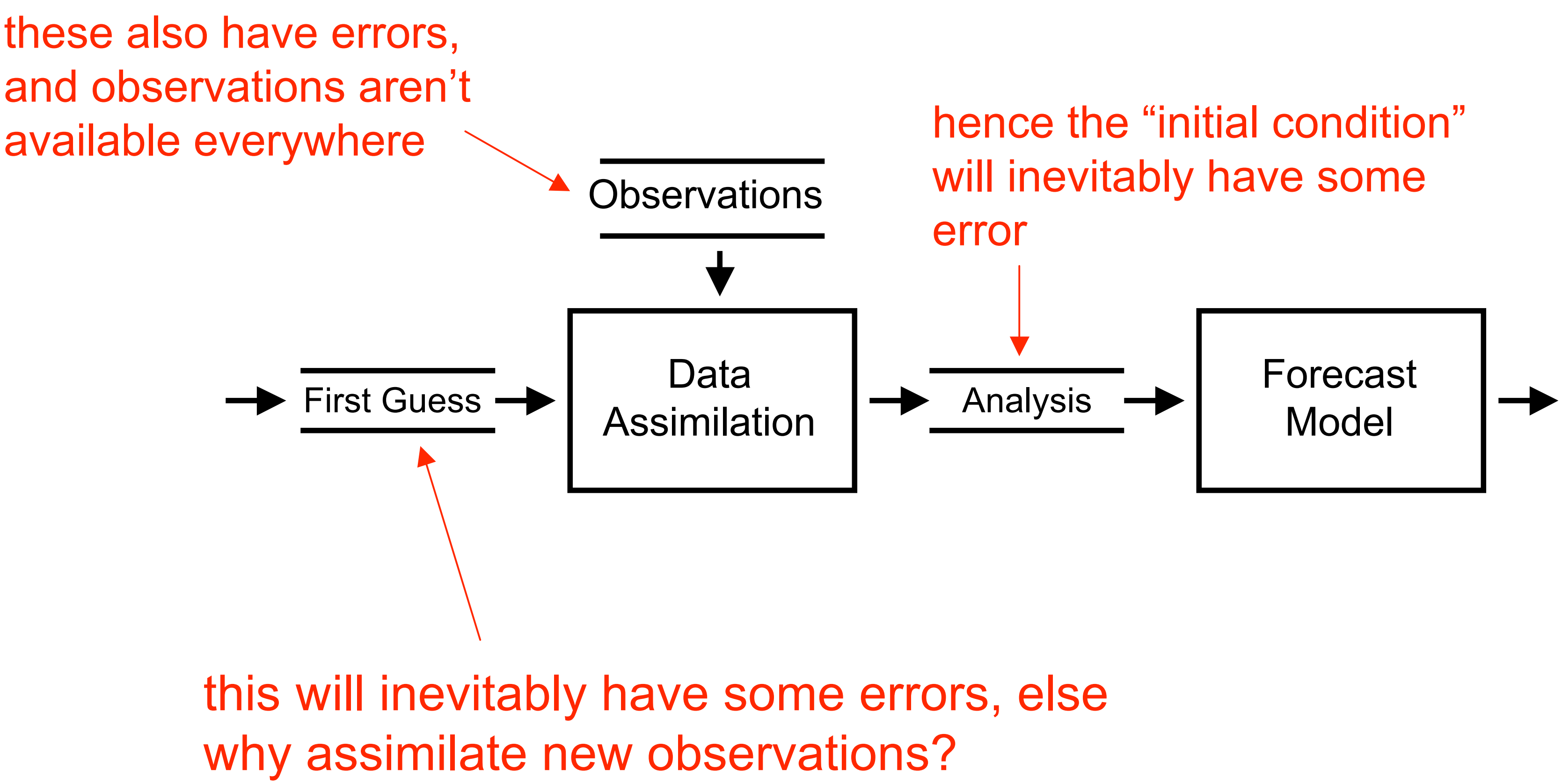
Understanding analysis error characteristics

these also have errors,
and observations aren't
available everywhere

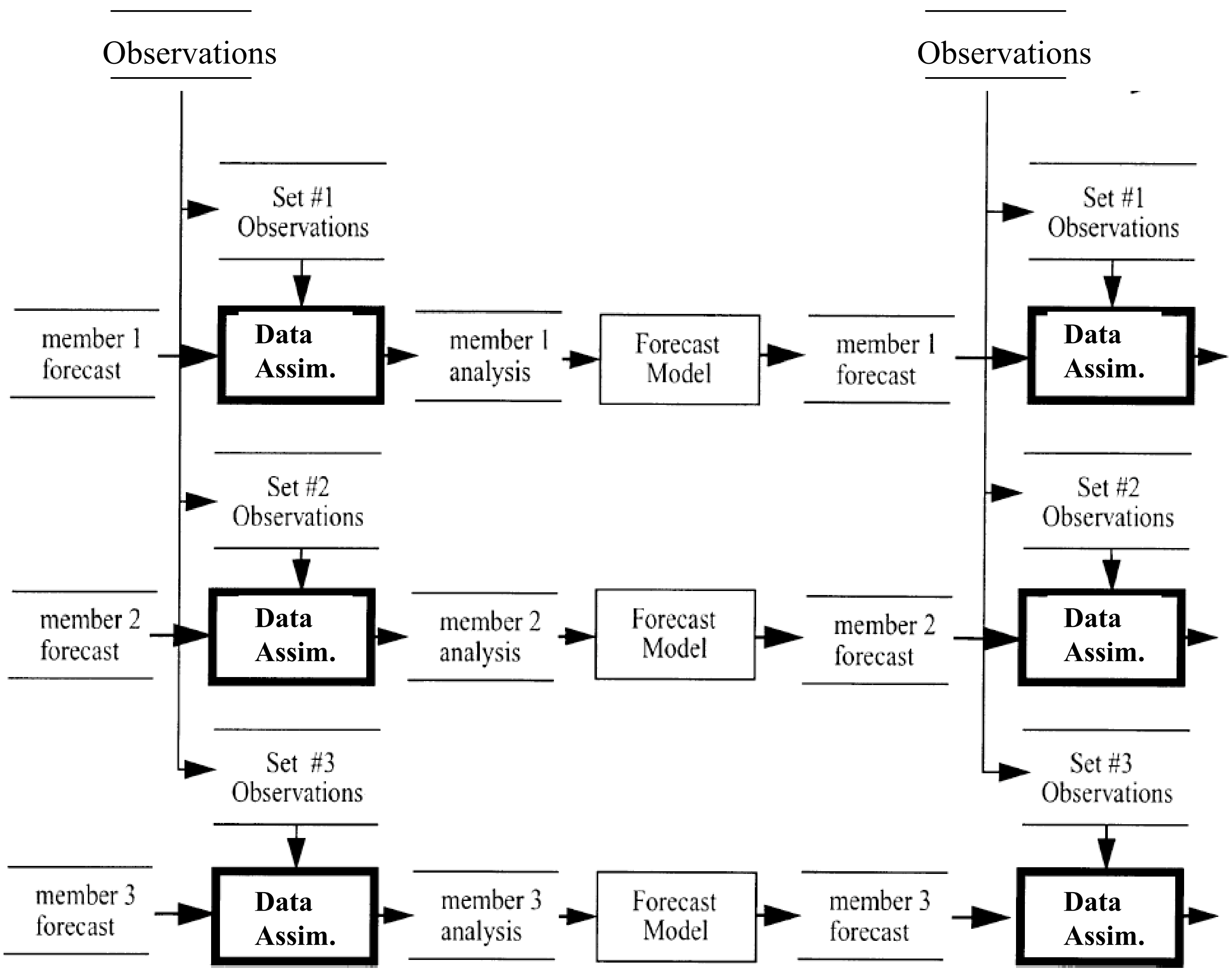


this will inevitably have some errors, else
why assimilate new observations?

Understanding analysis error characteristics



An experiment to examine the characteristics of analysis error pdfs



Use a process like this, with multiple forecasts and multiple realizations of observations, to examine characteristics of analysis errors.

Analysis error slide(s)

- Placeholder: insert graphics showing things like:
  - analysis errors larger in data voids.
  - larger scales pretty well analyzed, noisy at smaller scales.
  - analysis errors will be larger in a region of a growing cyclone.

The Breeding Technique (NCEP)

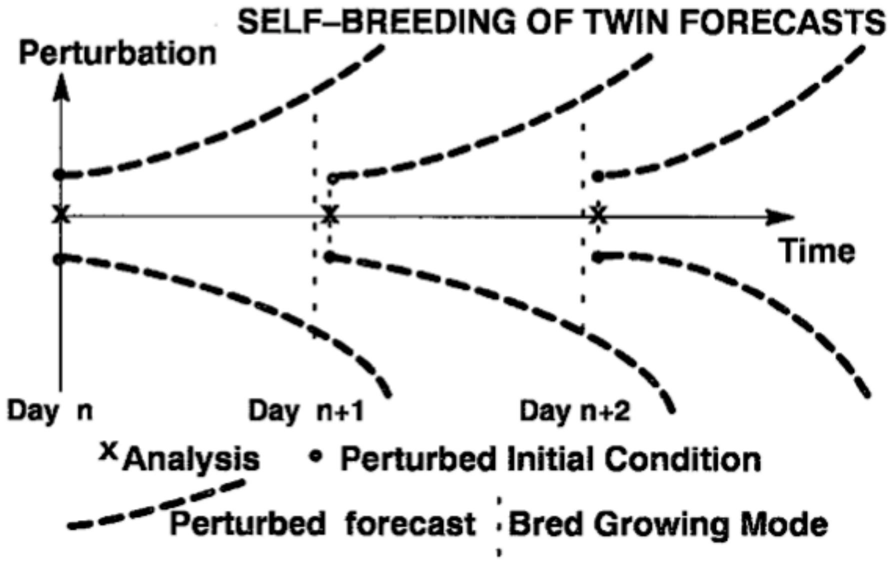


FIG. 5. Schematic of a self-contained breeding pair of ensemble forecasts.

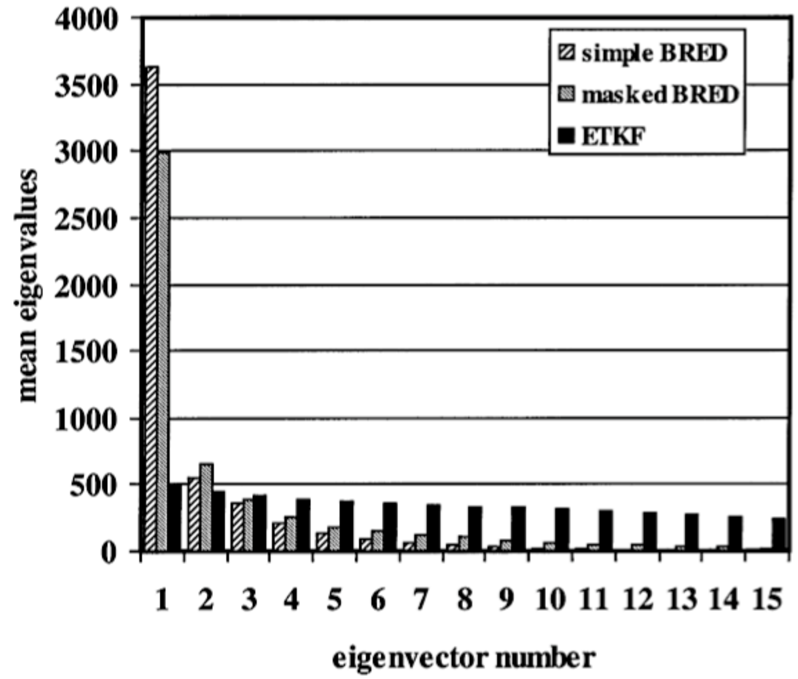


FIG. 5. Seasonal mean spectra of eigenvalues of ensemble-based 12-h forecast covariance matrices normalized by observation error covariance in observation sites for 16-member ETKF, simple breeding and masked breeding ensembles.

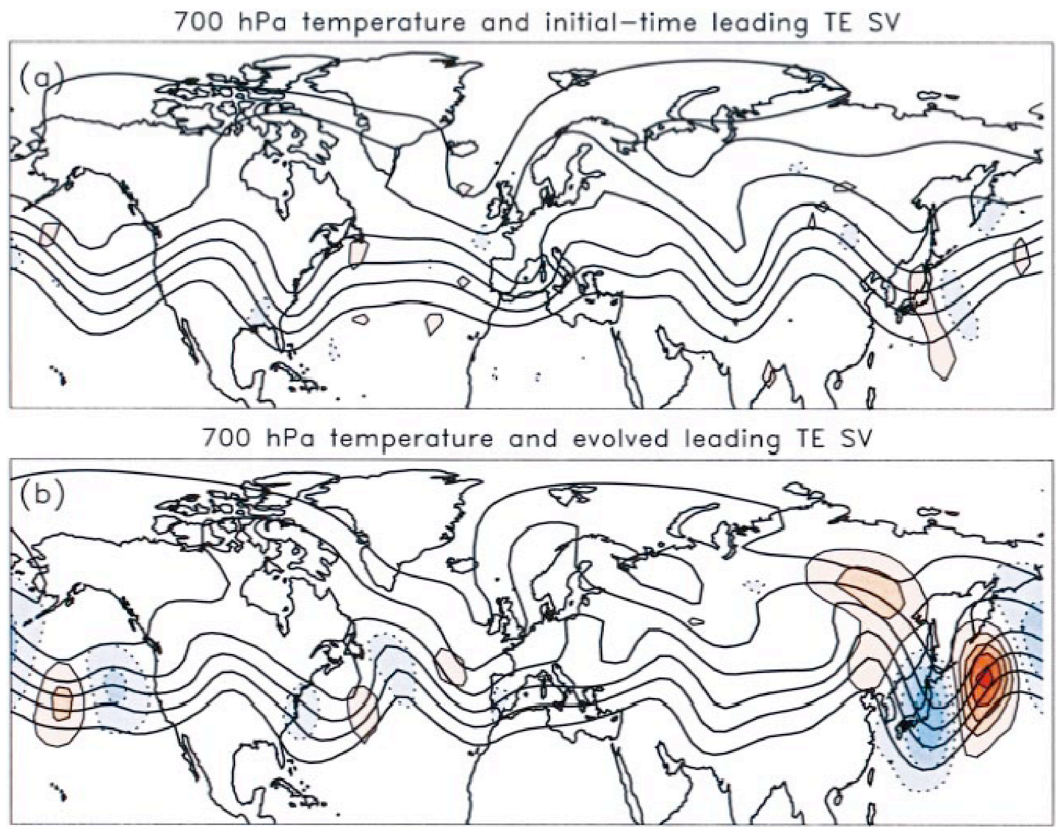
Breeding takes a pair of forecast perturbations and periodically rescales them down and adds them to the new analysis state. Perturbations only grossly reflect analysis errors.

Wang and Bishop showed that a larger ensemble formed from independent pairs of bred members will be comprised of pairs that have almost identical perturbation structures

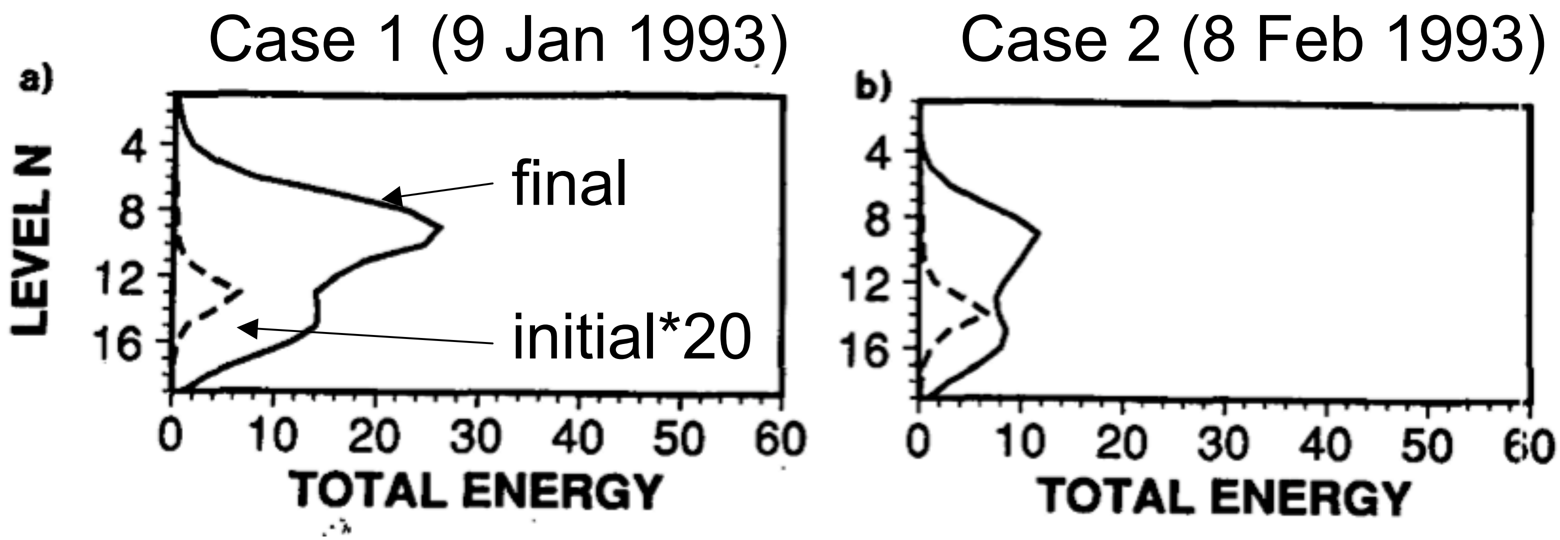
Singular Vectors (ECMWF)

Singular vectors (SVs) here indicate the field of perturbations that are expected to grow the most rapidly in a short-range forecast.
The SV structure depends upon a choice of norm.

The small perturbation in the top panel grows into the much larger perturbation in the bottom panel.



Singular vectors, continued

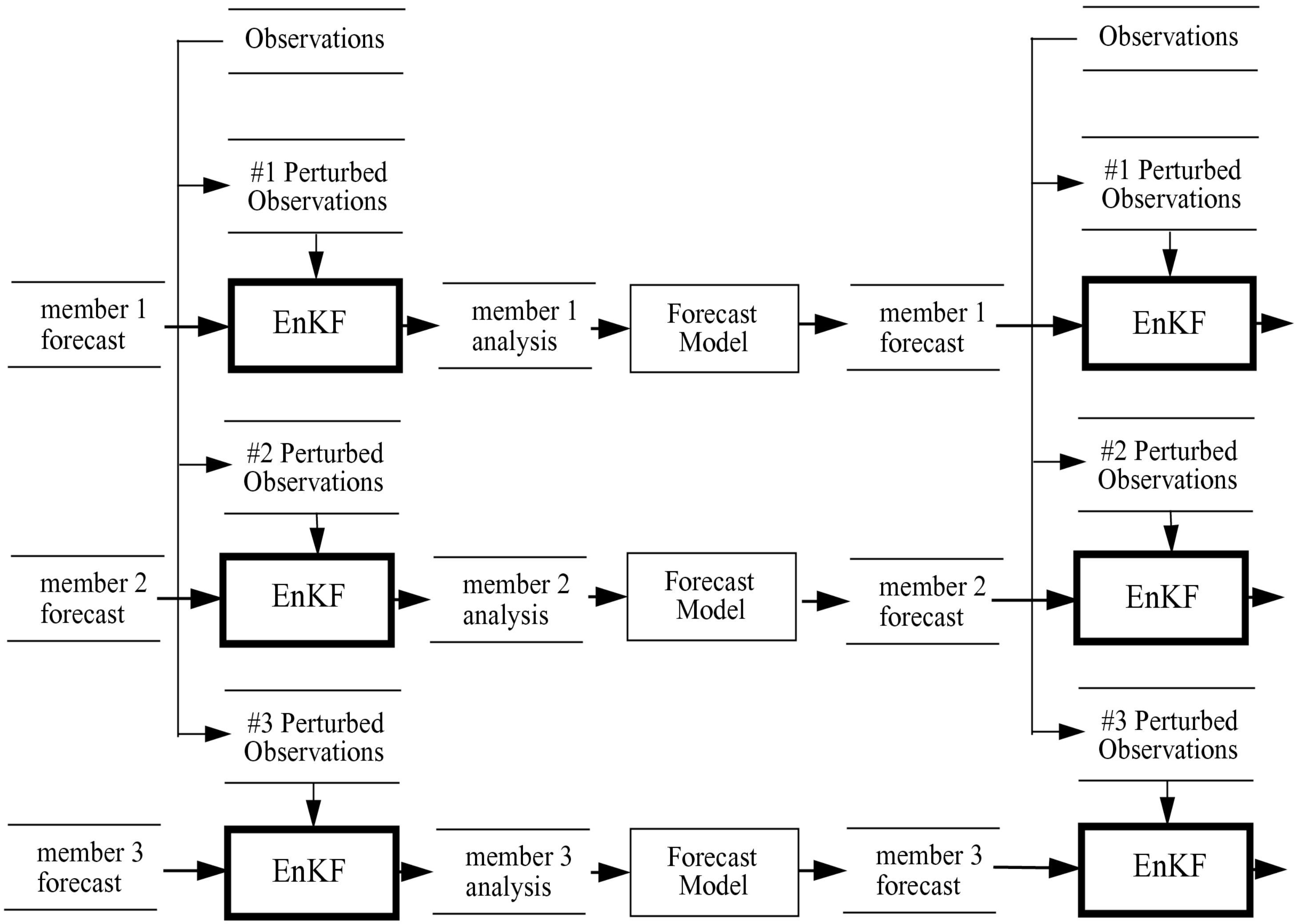


SVs tend to have their initial perturbations in the mid-troposphere, little amplitude near surface or tropopause. If they're meant to sample analysis errors, are analysis errors really near-perfect at these levels?

As SVs evolve, they grow to have amplitudes aloft and near the surface.

Expect SV ensemble forecast spread to be unrepresentative in the early hours of the forecast (e.g., spread of EFs too small near the surface).

The ensemble Kalman filter: a schematic



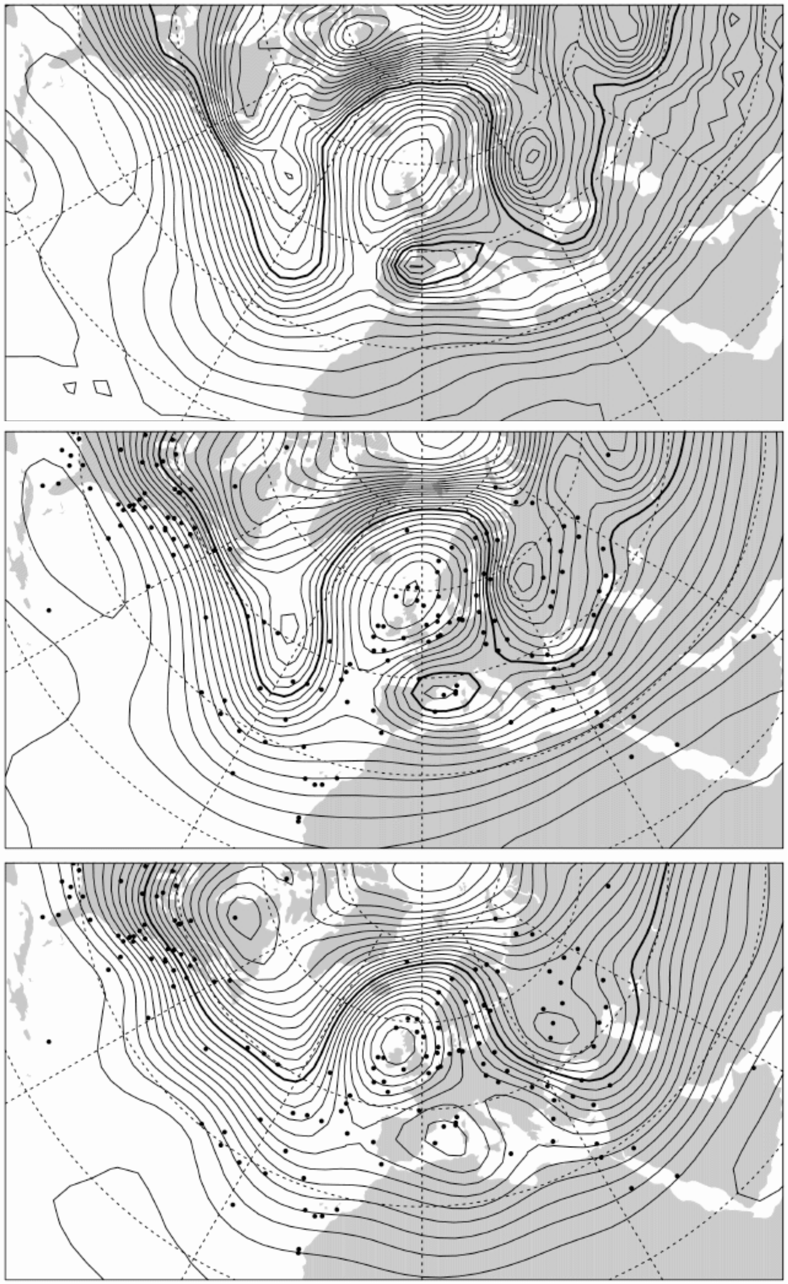
(This schematic is a bit of an inappropriate simplification, for EnKF uses every member to estimate background-error covariances)

500 hPa height analyses assimilating only SfcP observations

Full NCEP-NCAR Reanalysis (3D-Var)
(120,000+ obs)

Ensemble Filter
(214 surface pressure obs)

Optimal Interpolation
(214 surface pressure obs)

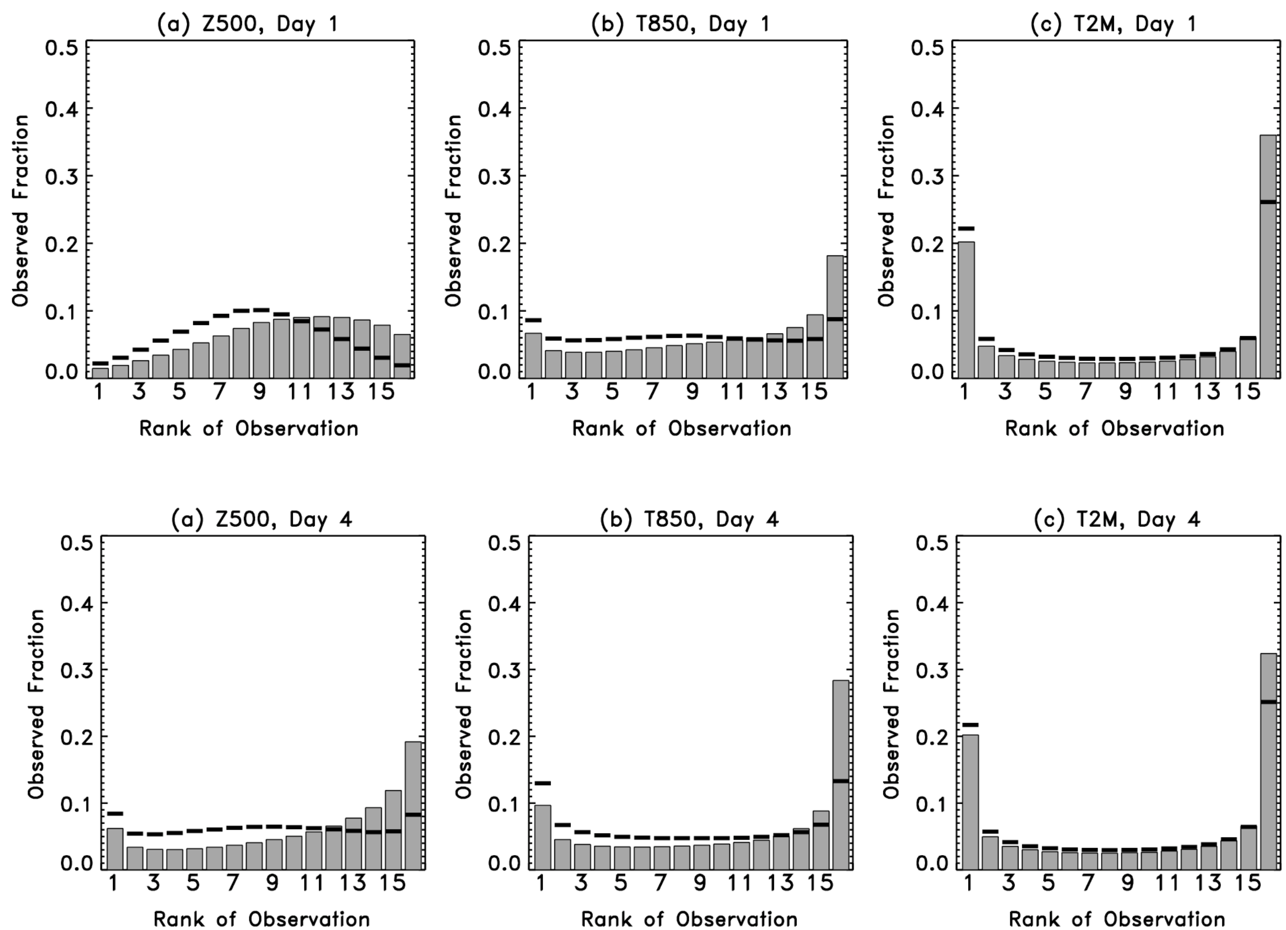


Black dots show pressure ob locations

RMS = 39.8 m

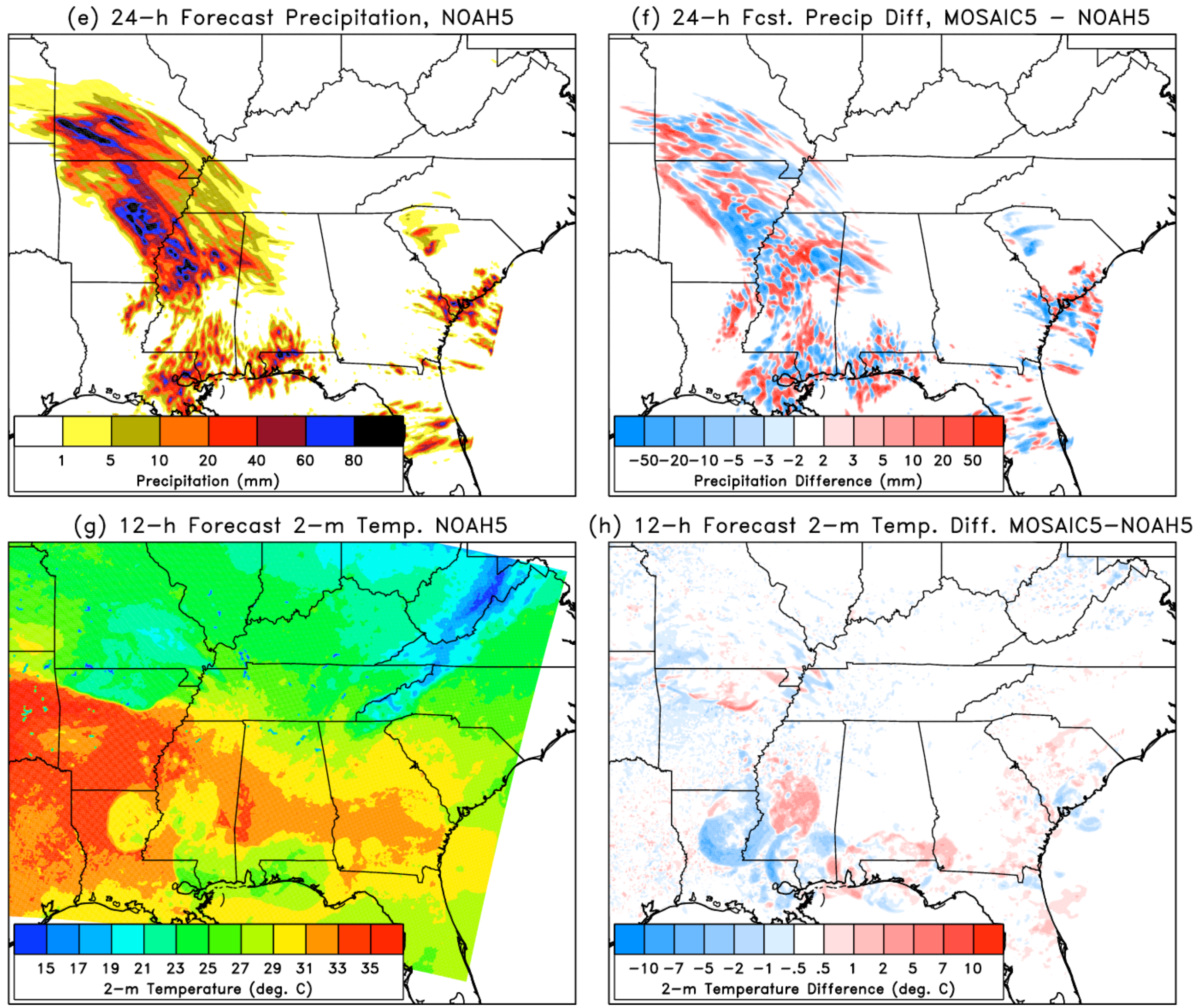
RMS = 82.4 m

Rank histograms of Z_{500}, T_{850}, T_{2m}



Notice the U-shaped rank histograms for 2-m temperature. What makes this worse near the surface?

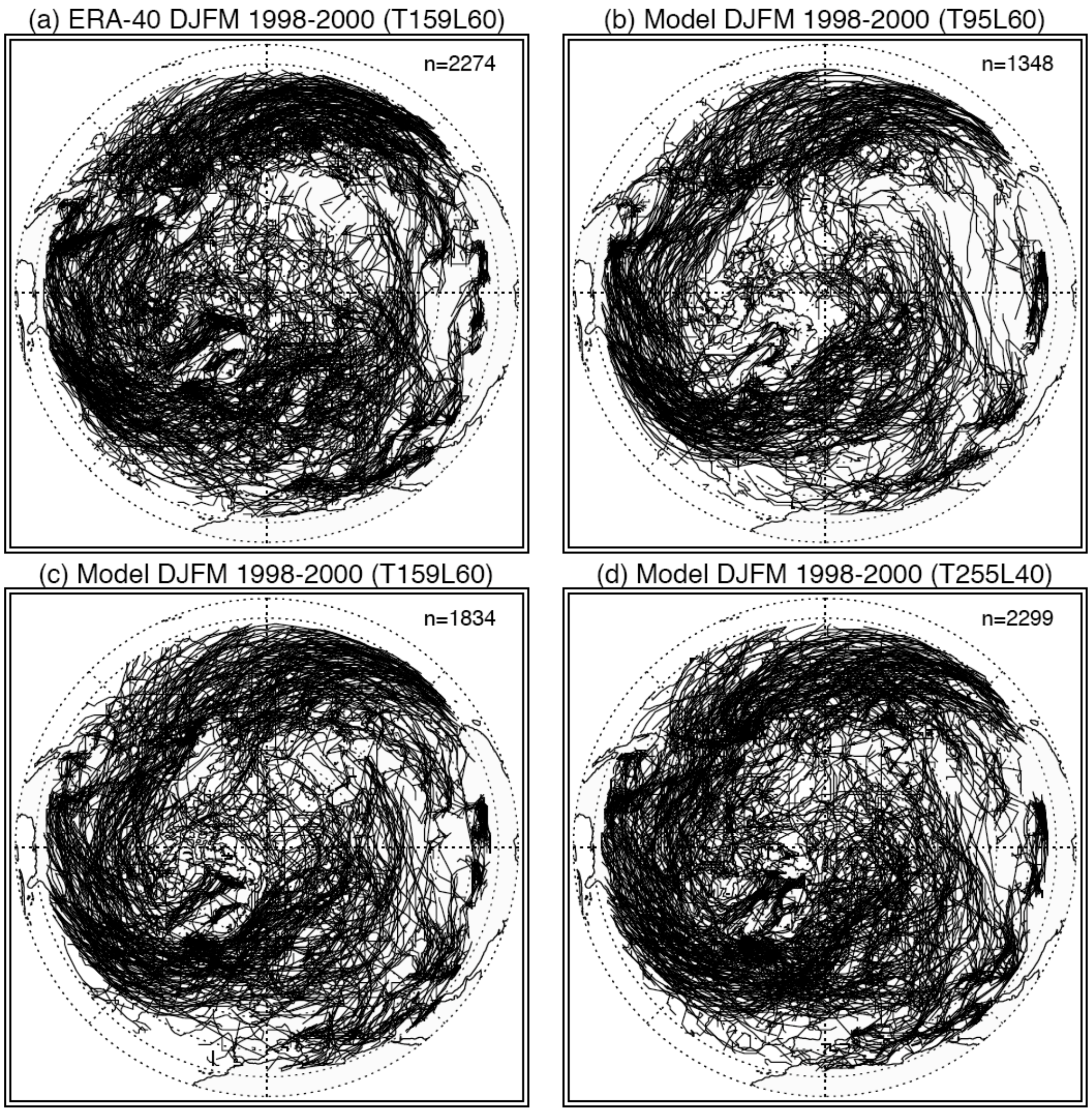
Perturb the land surface in EFs?



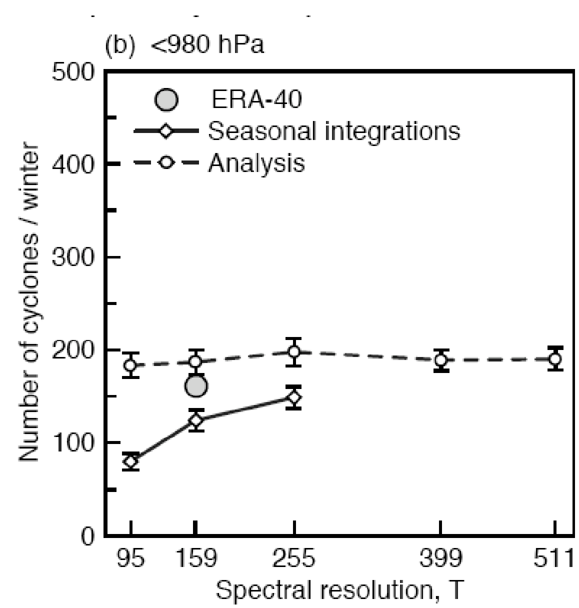
Problems and cures

- Problems:
  - Biases.
  - Spread deficiencies
  - Lack of spread / skill relationship.
  - Etc.
- Cures
  - Design a better ensemble forecast system (this talk)
    - Improved methods of generating initial conditions
    - Methods for dealing with model errors
  - Post-process the ensemble forecasts to ameliorate these errors (Maj. Tony Eckel's talk)

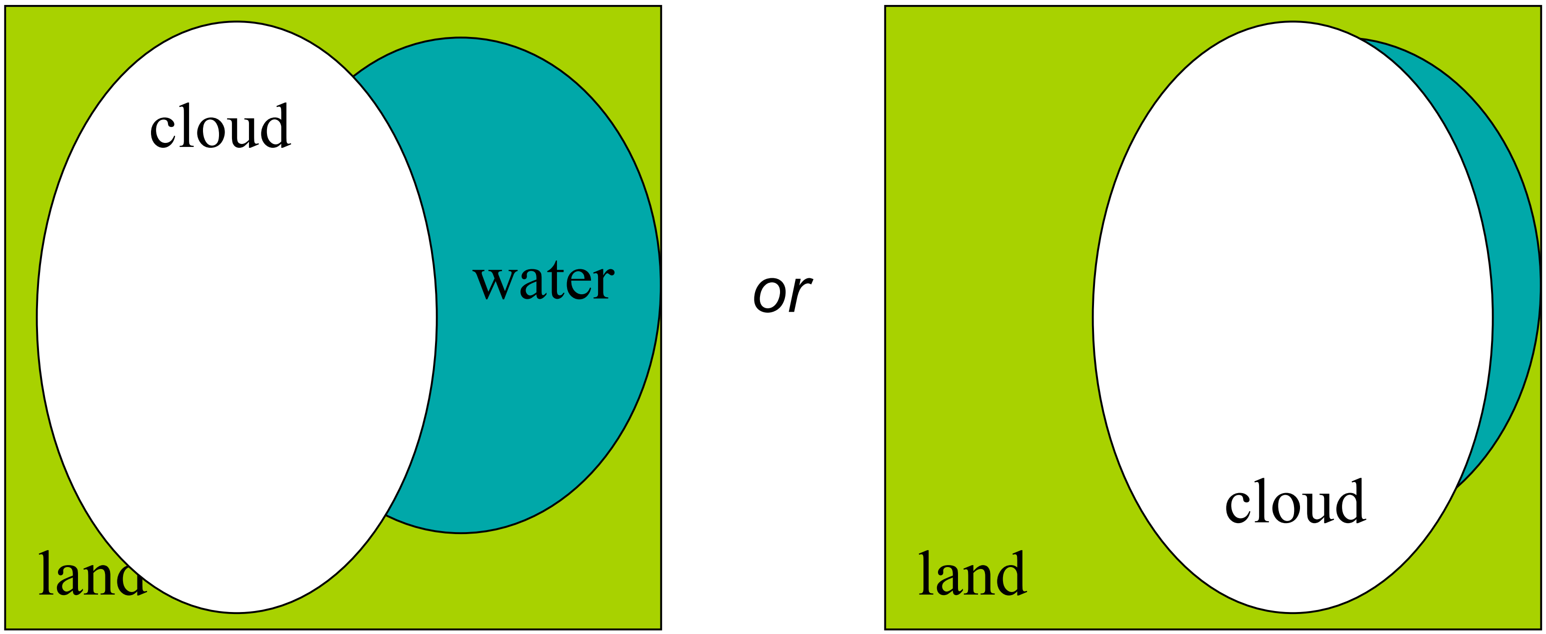
Many problems with forecast models that cause systematic bias. Example: cyclone frequency and model resolution



Fewer cyclones in low-resolution ECMWF models than in higher-res. models.



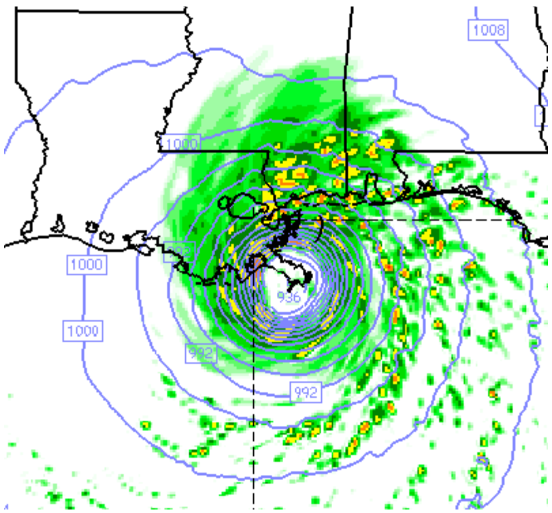
Forecast error from random model uncertainties



Envision a grid box comprised of part land, part water. It is 40 % covered by cloud. The amount of heating from the ground will depend on where the cloud is positioned inside the grid box. The unknowable sub-gridscale detail thus contributes an element of randomness to the forcing at the scale of the grid box. This may contribute to a lack of spread in ensemble forecasts.

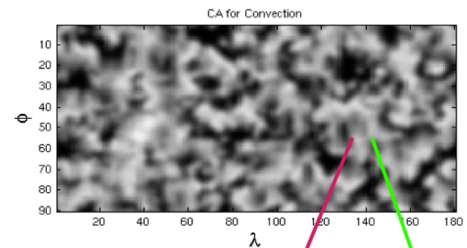
Dealing with model errors

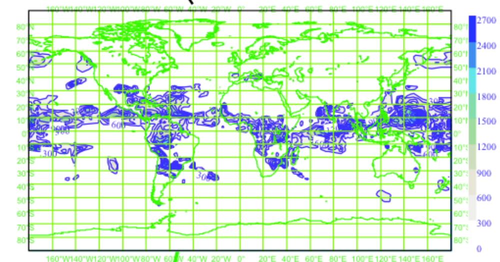
(1) Better models (4-km, 60-h WRF for Katrina)



(2) Introduce stochastic element into model

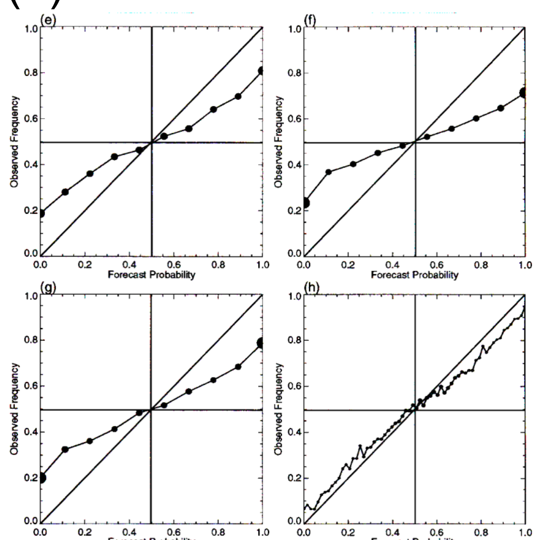
$$\Delta\psi = f(\Psi(\lambda, \varphi, t), z) * CAPE\ (\lambda, \phi, t)$$



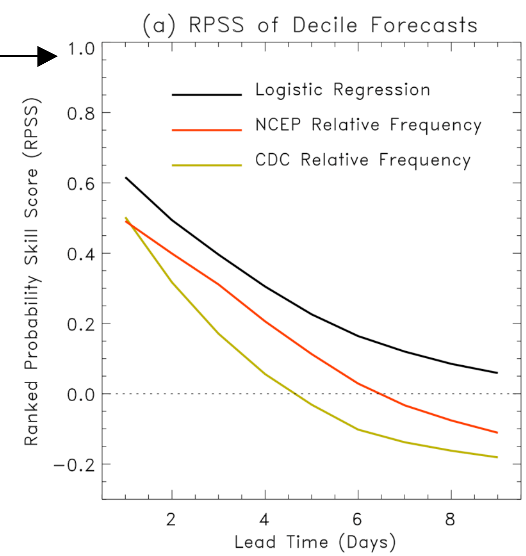


Non-local, quasi-random, state-dependent

(3) Multi-model ensembles



(4) Calibration →



References and notes

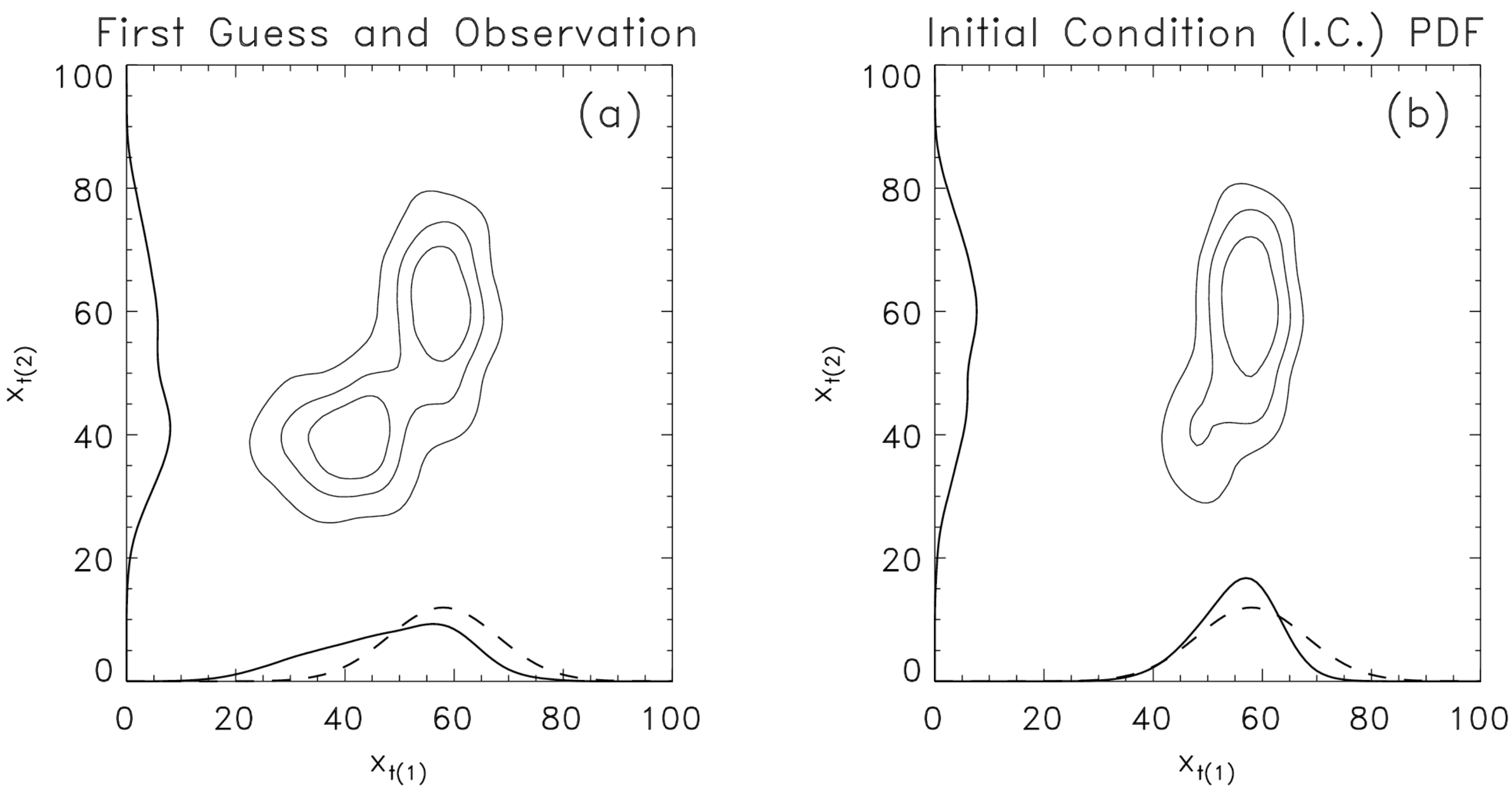
Ensemble forecast verification:
(1) Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 2006.
(2) Beth Ebert's verification web page, tinyurl.com/y97c74 (references and links to many papers and sites).

Generating ensemble forecast initial conditions:
(1) Buizza, R., and T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434-1456.
(2) Barkmeijer, J. M. Van Gijzen, and F. Bouttier, 1998: Singular vectors and estimates of the analysis-error covariance metric. *Quart. J. Royal Meteor. Soc.*, **124**, 1697-1713.
(3) Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation. Chapter 6 of *Predictability of Weather and Climate*, Cambridge Press, 124-156.
(4) Toth, Z. and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

Model errors:
(1) WRF forecast from http://wrf-model.org/plots/realtime_hurricane.php (try 2005-08-27, and rain mixing ratio, 60-h forecast)
(2) Stochastic element picture from Judith Berner's presentation at the ECMWF workshop on the representation of sub-grid processes using stochastic-dynamic models. http://www.ecmwf.int/newsevents/meetings/workshops/2005/Sub-grid_Processes/Presentations.html .
(3) Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. *Tellus*, **57A**, 219-233. (multi-model ensemble picture)
(4) Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2005: Reforecasts, an important data set for improving weather predictions, *Bull. Amer. Meteor. Soc.*, in press. Available at http://www.cdc.noaa.gov/people/tom.hamill/reforecast_bams4.pdf

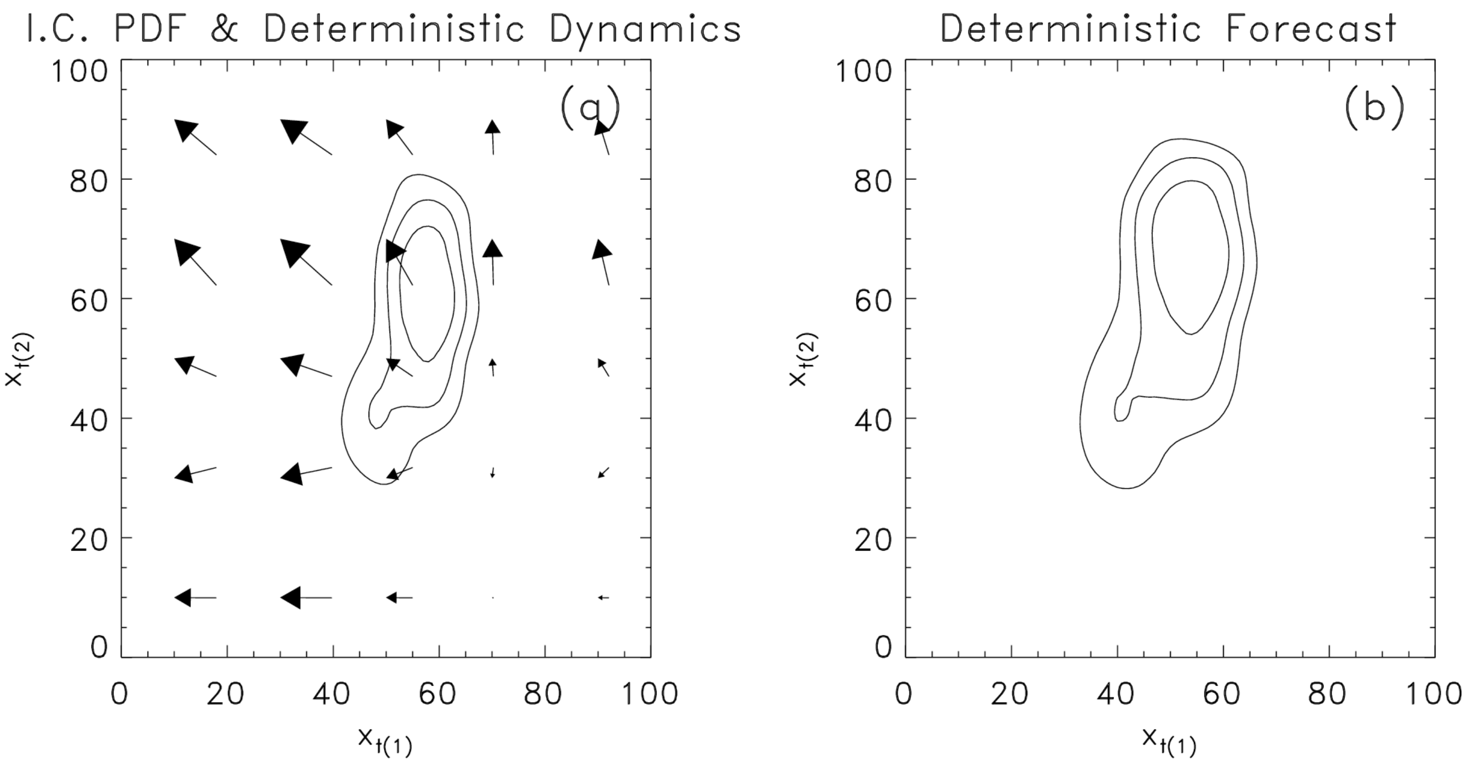
Backup slides

Probabilistic NWP: data assimilation



Envision a prior estimate of a two-dimensional model state (solid lines) assimilating a new observation (dashed line), which measures only 1st component of the state.

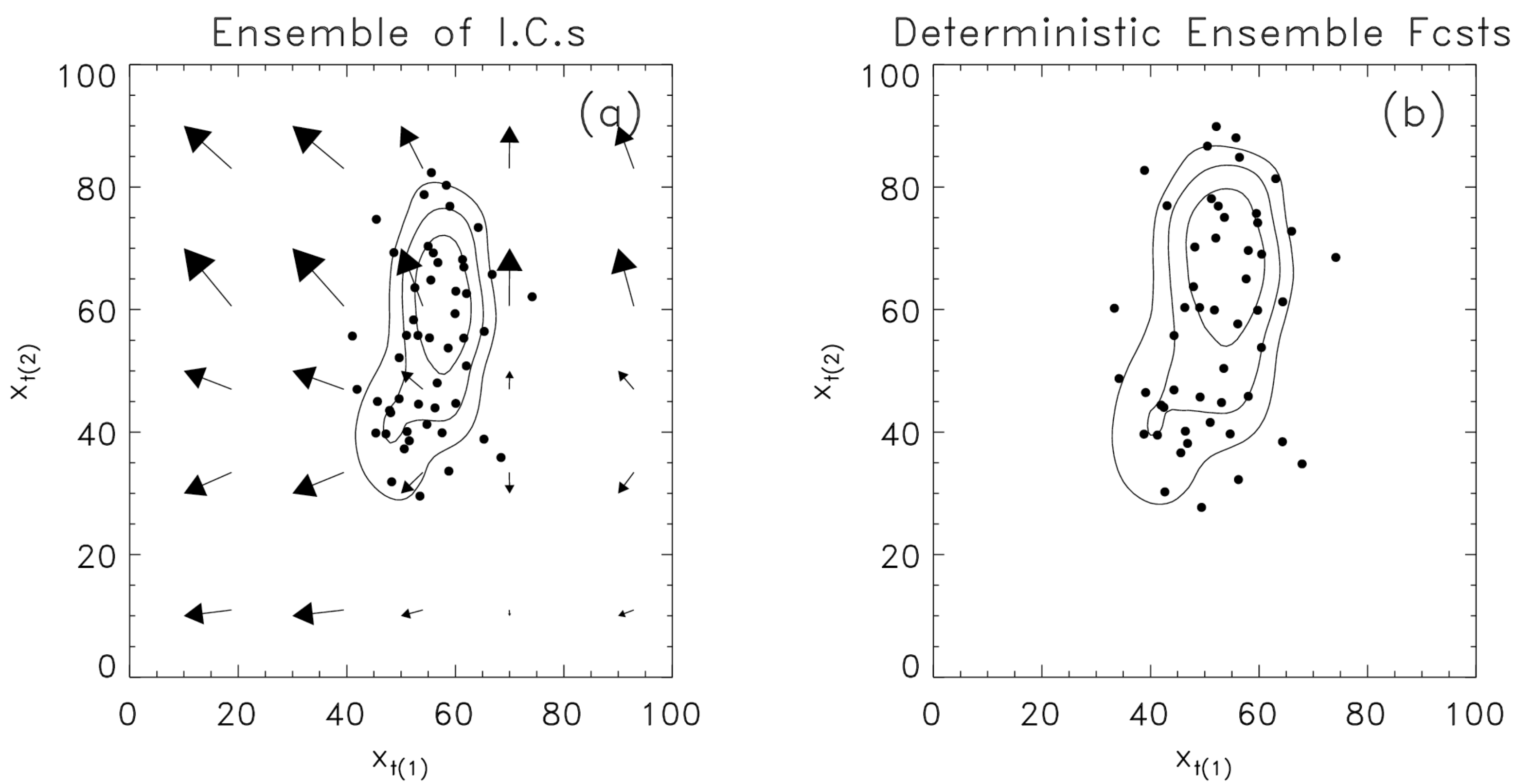
Probabilistic NWP: forecasting pdf with perfect model



Arrows indicate state-dependent deterministic forecast dynamics, which tend to smear out probabilities with increasing time.

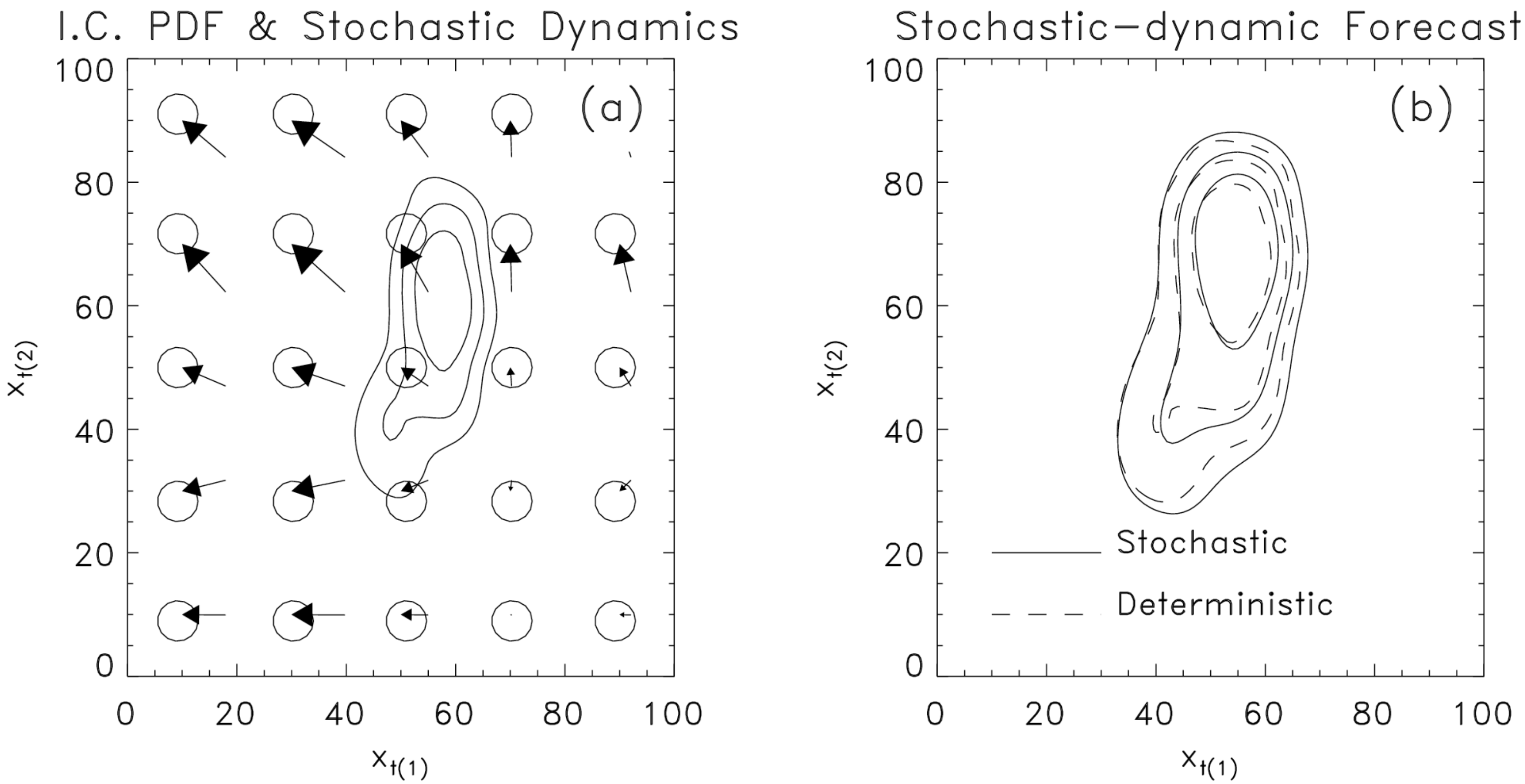
(actually forecasting the whole pdf in this manner is prohibitively expensive with high-dimensional NWP models)

Probabilistic NWP: approximating with deterministic ensemble forecasts



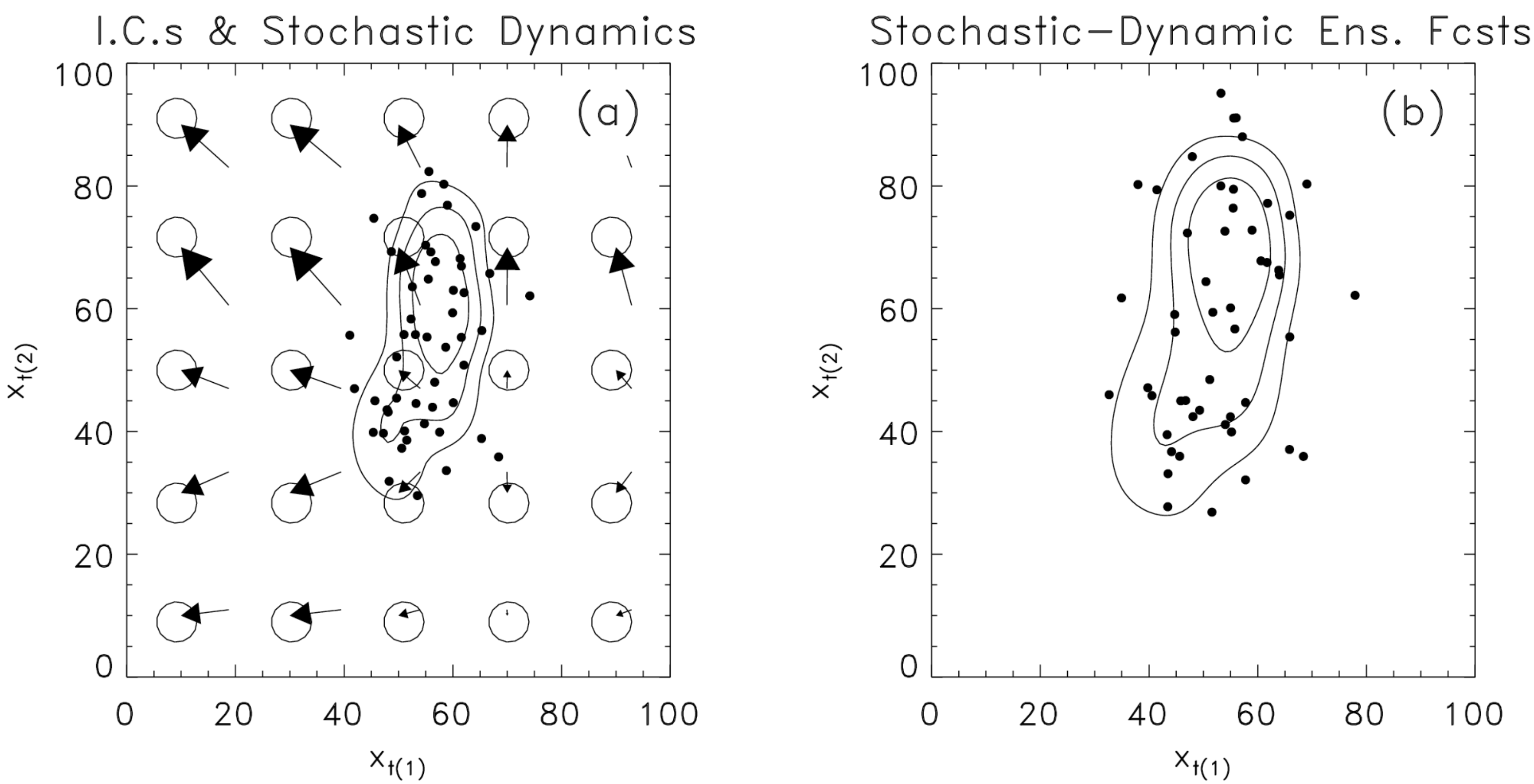
Generate samples from initial pdf. Propagate each sample forward in time using deterministic forecast model dynamics.

Probabilistic NWP: forecasting pdf with imperfect model



In reality, it’s more appropriate to think of forecast model as part deterministic (arrows) plus part stochastic (circles). Stochastic part may include state-dependent bias correction that shifts pdf and addition of random error that diffuses pdf.

Probabilistic NWP: approximating with stochastic ensemble forecasts



Generate samples from initial pdf. Propagate each sample forward in time using "stochastic-dynamic" forecast model.