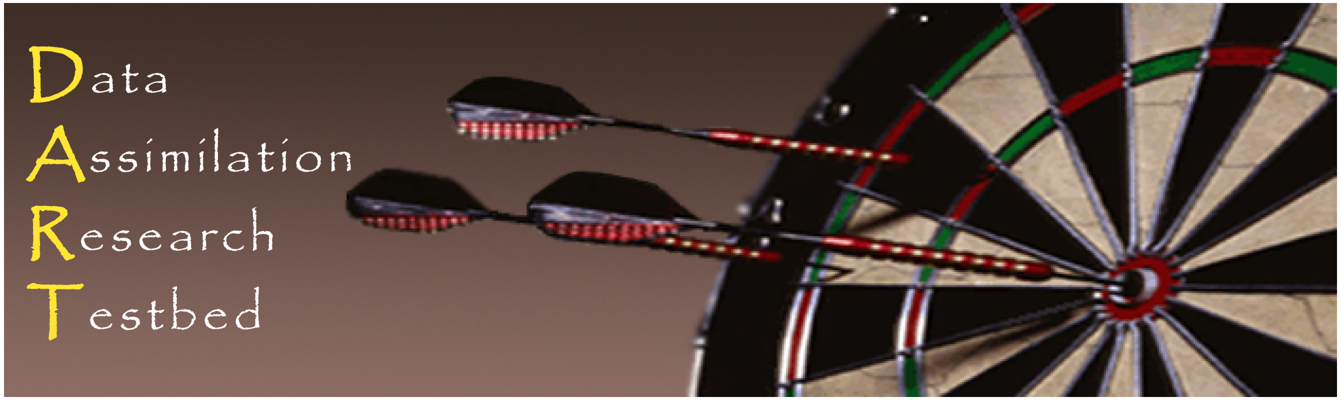
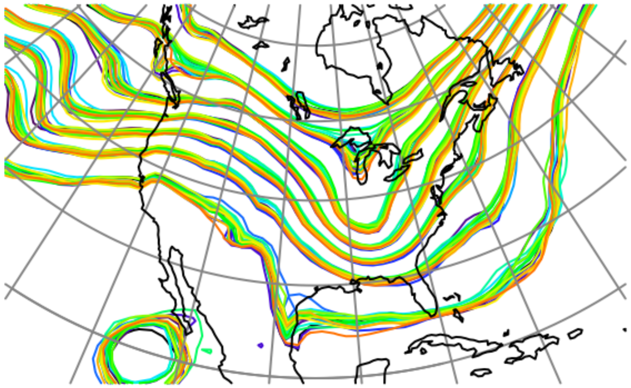


Verification of Ensemble Forecasts

Jeff Anderson, Tim Hoar
NCAR Data Assimilation Research Section



The National Center for Atmospheric Research is sponsored by the National Science Foundation.
Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

NCAR UCAR National Center for Atmospheric Research

The Prediction Problem for Geophysical Models

Dynamical system governed by (stochastic) Difference Equation:

$$dx_t = f(x_t, t) + G(x_t, t)\, d\beta_t, \qquad t \geq 0 \qquad (1)$$

Observations at discrete times:

$$y_k = h(x_k, t_k) + v_k; \qquad k = 1, 2, \ldots; \qquad t_{k+1} > t_k \geq t_0 \qquad (2)$$

Observational error white in time and Gaussian (nice, not essential).

$$v_k \rightarrow N(0, R_k) \qquad (3)$$

Complete history of observations is:

$$Y_\tau = \{ y_l; t_l \leq \tau \} \qquad (4)$$

<u>Goal: Find probability distribution for state at time t and subsequent:</u>

$p(x, t \mid Y_t)$ Analysis $\qquad$ $p(x, t^+ \mid Y_t)$ Forecast $\qquad$ (5)

What do we want from an ensemble (set) of analyses/forecasts?

- Random draw from forecast or analysis distribution. Normally just a simple random draw.

- Could try to draw using some ‘proposal’ density. Example, sample unlikely high-impact events more. Hard to do, not discussed further here.

Properties of Random Sample Ensemble

- Ensemble members indistinguishable from ‘truth’.
  If observations (and model) are unbiased.

- Observations are NOT drawn from same distribution.
  They also include an observational error.
  But they are all we know.
  Have to account for these differences when verifying.

Computing Sample Statistics from Ensembles

- Can compute sample statistics from ensemble:
  Mean,
  Variance / standard deviation,
  Higher order moments,
  Clusters,
  Probabilities of events, P(Rainfall > 1.0mm).

- Can compute statistics for any function.

- If observation is not in state, can compute an estimate for each ensemble (forward operator).

Common Ensemble Generation Mechanisms

- Empirical perturbations around a single mean: Mean usually from deterministic assimilation.
- Breeding: Grow perturbations around evolving mean.
- Singular vectors: Sample rapidly growing perturbations to an evolving mean.
- Ensemble data assimilation: Directly estimate forecast distribution.

Common Ensemble Generation Mechanisms: Weaknesses

- Empirical perturbations around a single mean:
  Usually assume gaussian structure,
  Empirical covariance estimates,
  No use of observations.

- Breeding:
  Empirical normalization size and period,
  No direct use of observations.

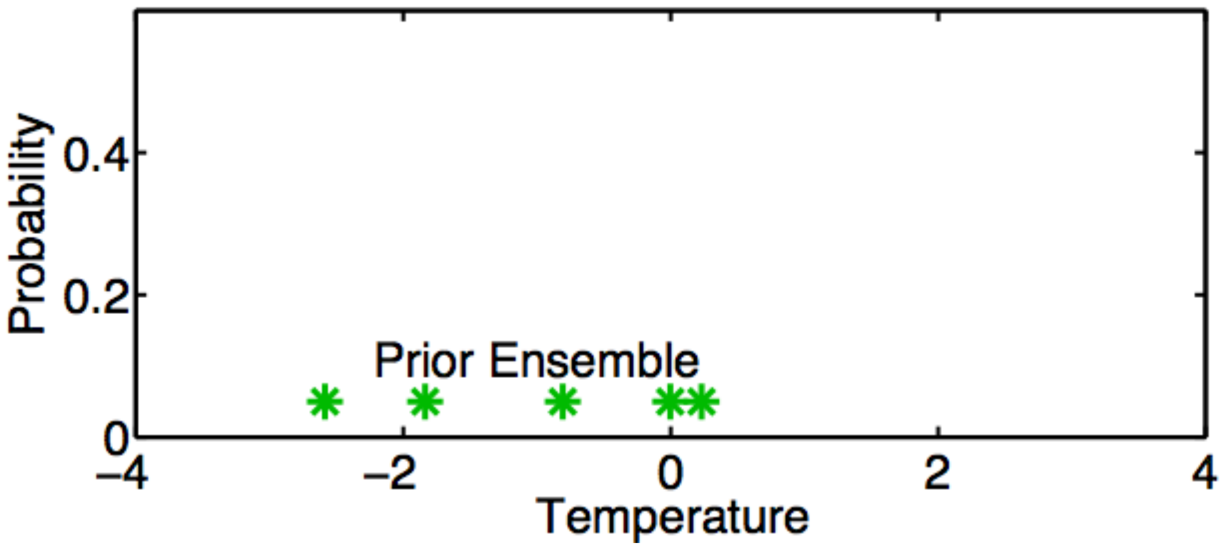
- Singular vectors:
  Implicit assumptions about projection of distribution,
  No direct use of observations.

Ensemble Data Assimilation Weaknesses

- Assumes all distributions are gaussian.
- Assumes forecast model is linear.
- Assumes observation forward operator is linear.

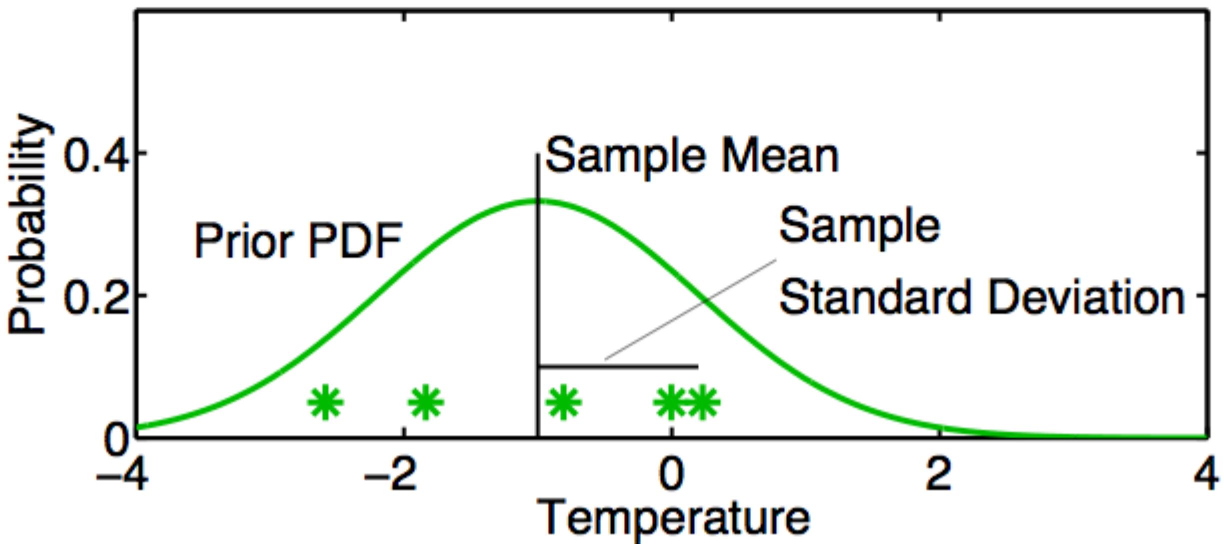
A One-Dimensional Ensemble Kalman Filter

Represent a prior pdf by a sample (ensemble) of N values:



A One-Dimensional Ensemble Kalman Filter

Represent a prior pdf by a sample (ensemble) of N values:



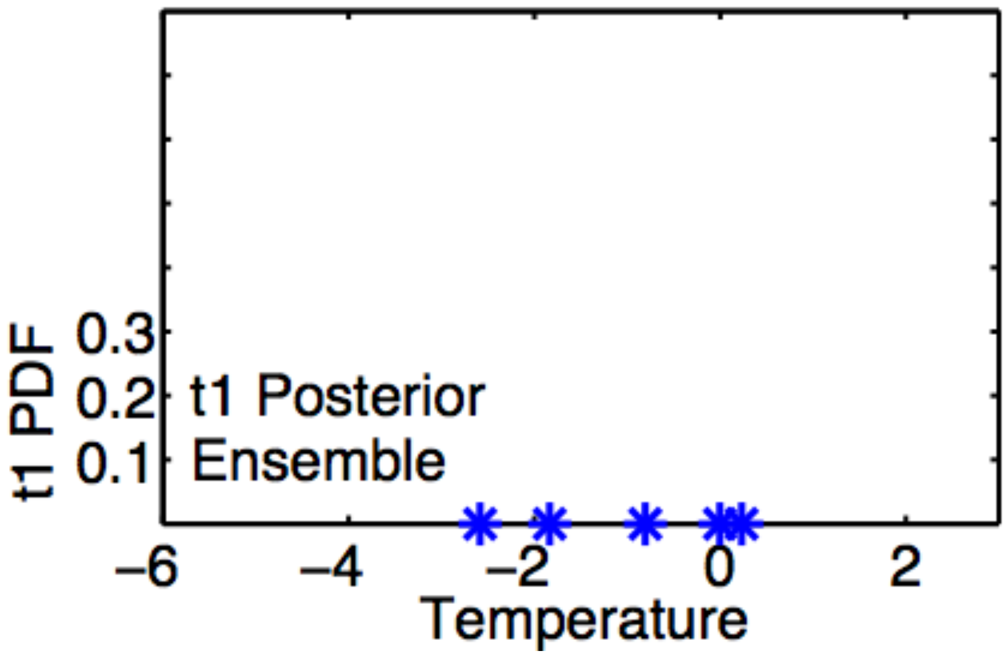
Use sample mean $\overline{T} = \sum_{n=1}^{N} T_n / N$

and sample standard deviation $\sigma_T = \sqrt{\sum_{n=1}^{N} \left(T_n - \overline{T}\right)^2 / (N-1)}$

to determine a corresponding continuous distribution $Normal\left(\overline{T}, \sigma_T\right)$

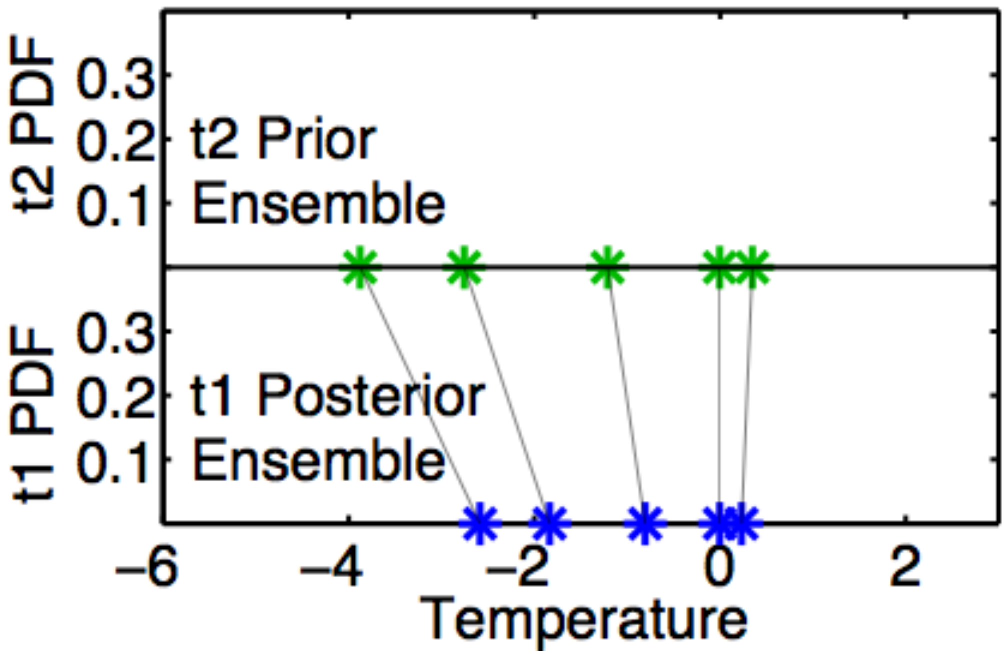
A One-Dimensional Ensemble Kalman Filter: Model Advance

If posterior ensemble at time t_1 is $T_{1,n}$, $n = 1, \ldots, N$



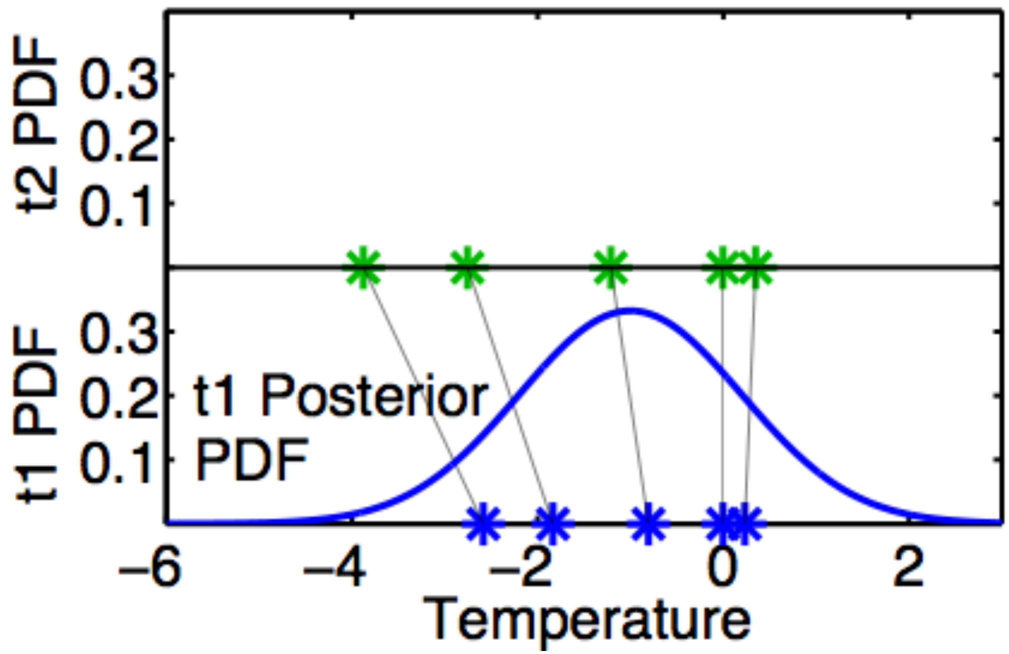
A One-Dimensional Ensemble Kalman Filter: Model Advance

If posterior ensemble at time t_1 is $T_{1,n}$, $n = 1, \ldots, N$,
advance each member to time t_2 with model, $T_{2,n} = L(T_{1,n})$ $n = 1, \ldots, N$.



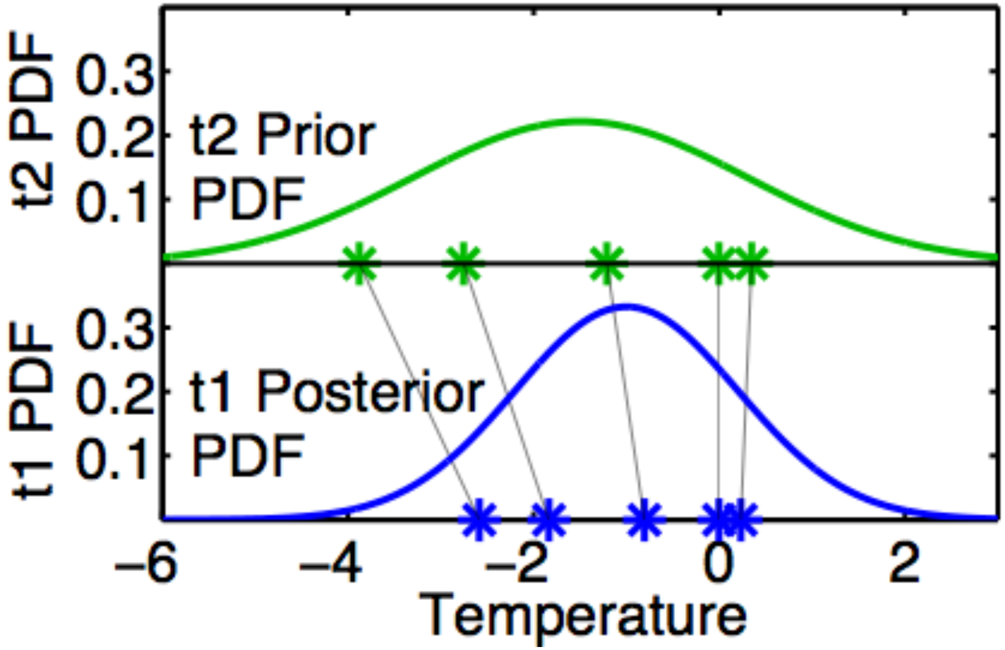
A One-Dimensional Ensemble Kalman Filter: Model Advance

Same as advancing continuous pdf at time t_1 …

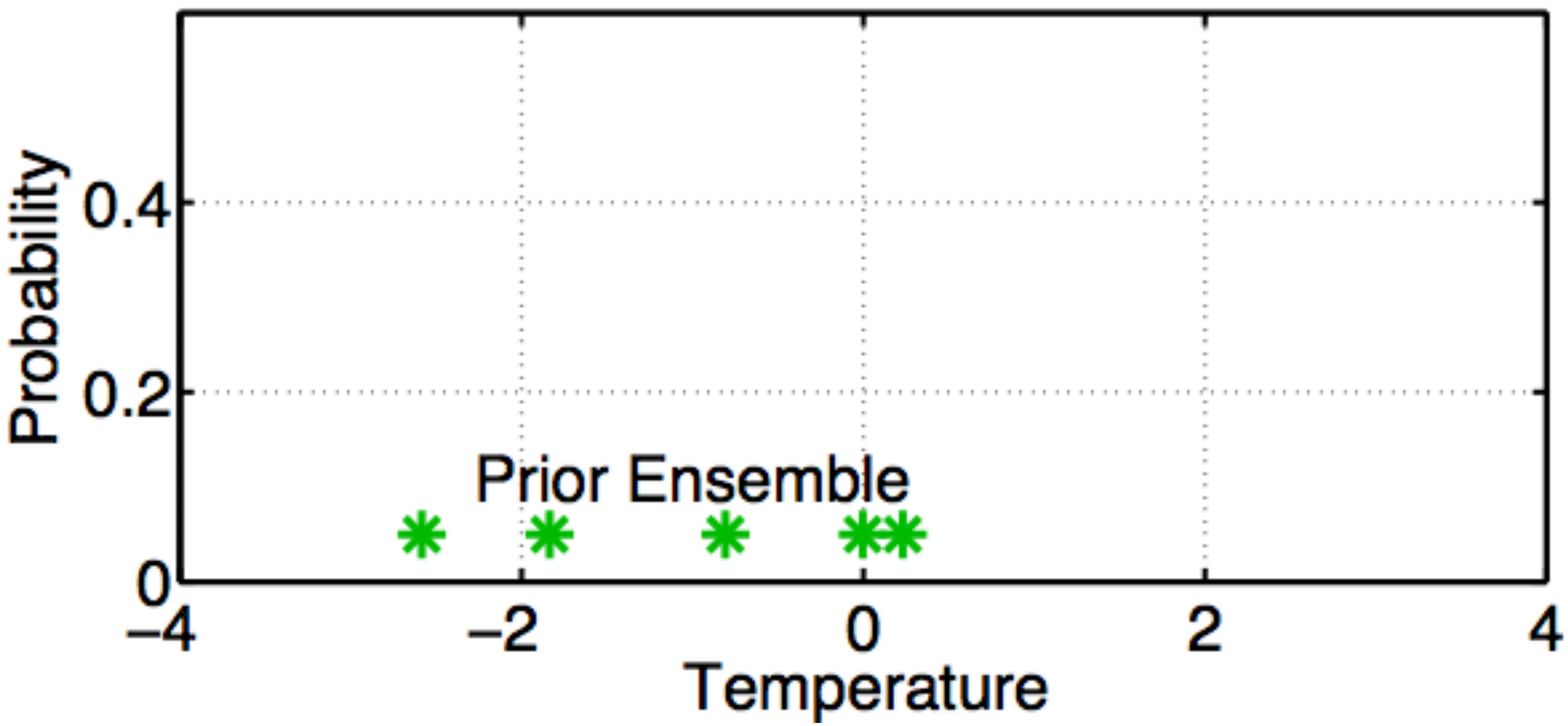


A One-Dimensional Ensemble Kalman Filter: Model Advance

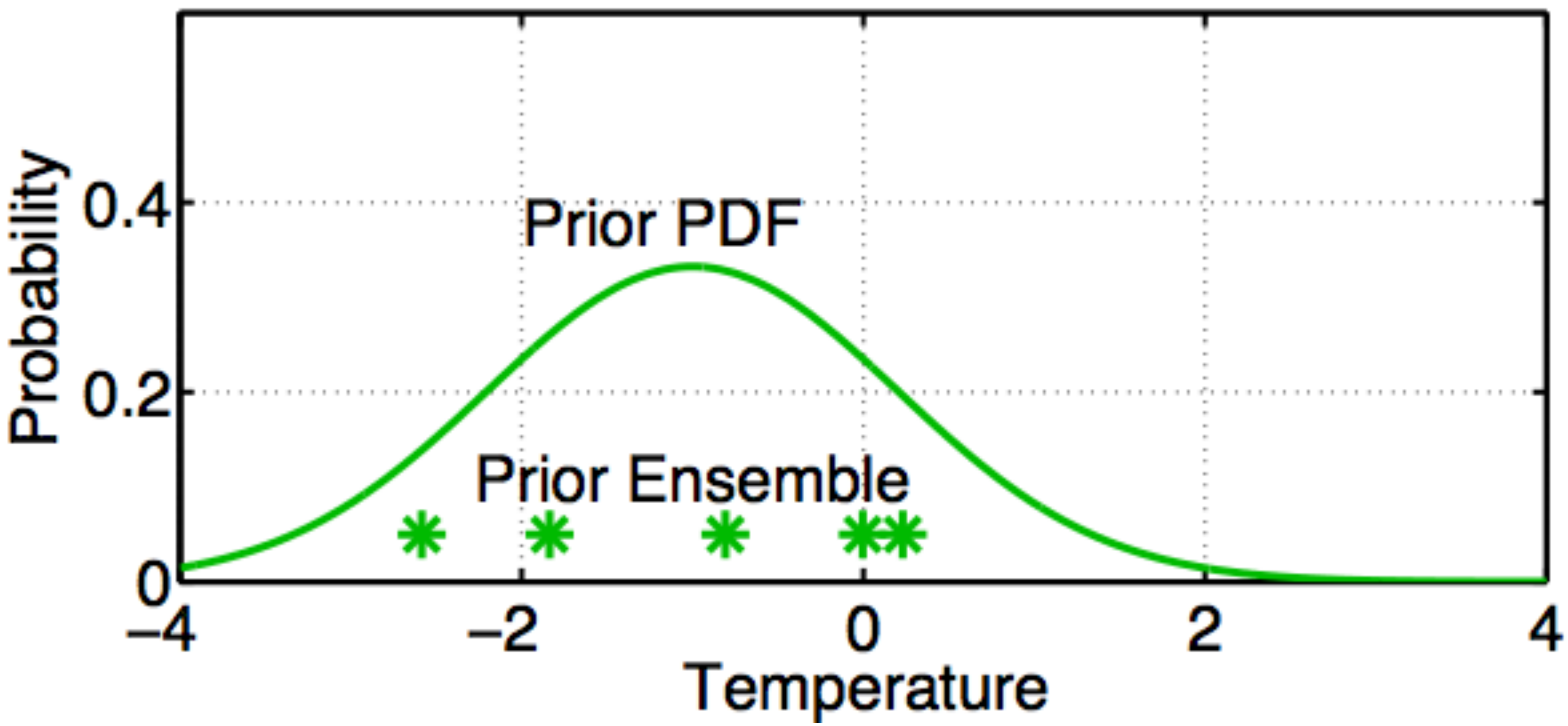
Same as advancing continuous pdf at time t_1 to time t_2 with model L.



A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation

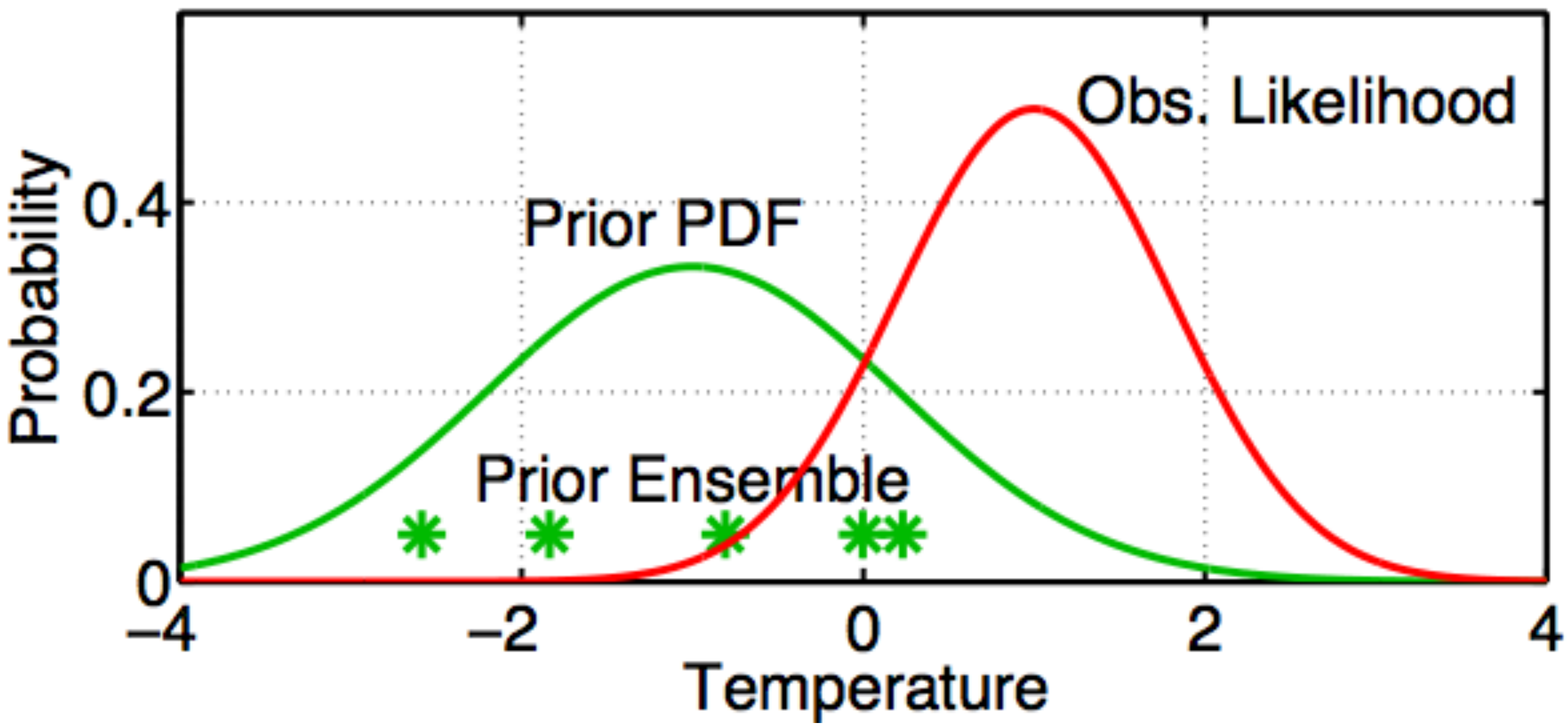


A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



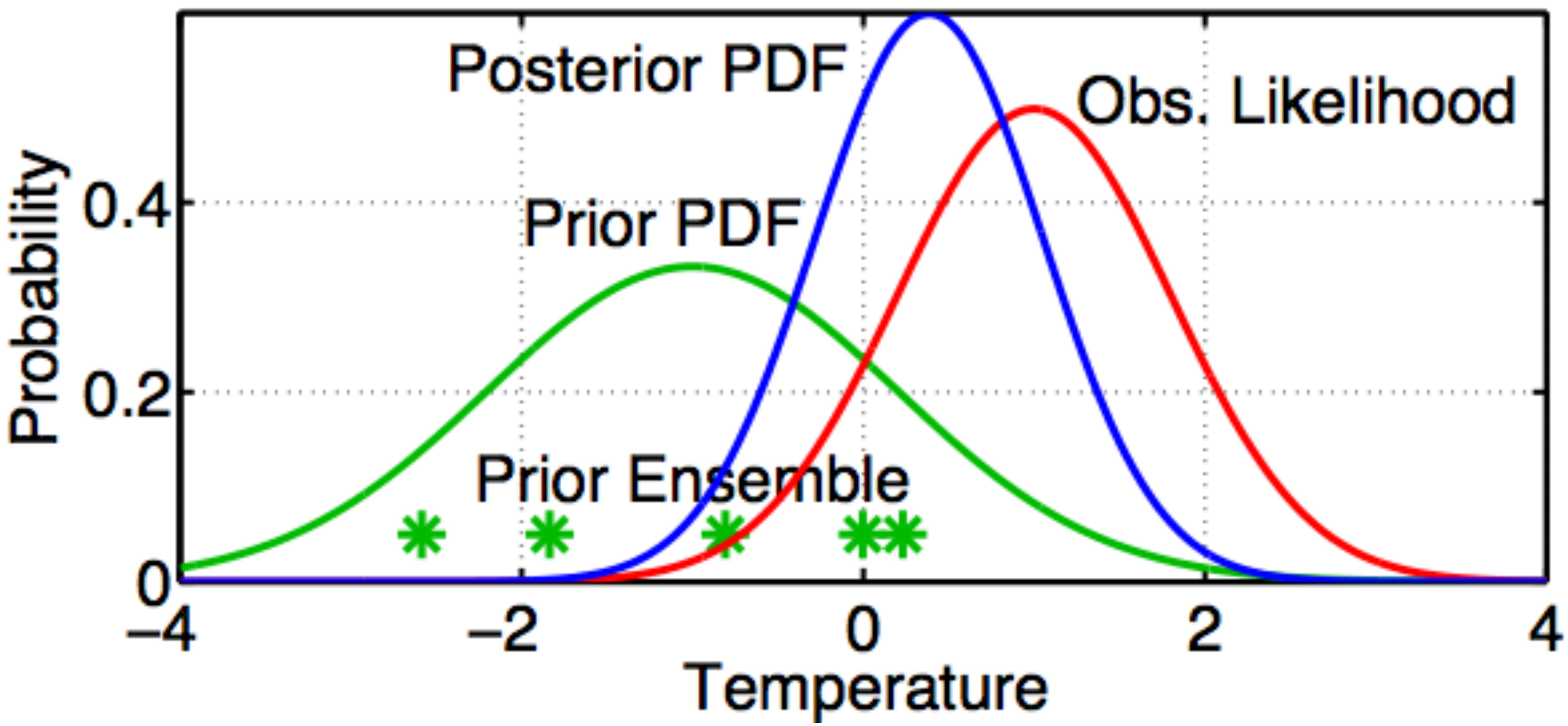
Fit a Gaussian to the sample.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



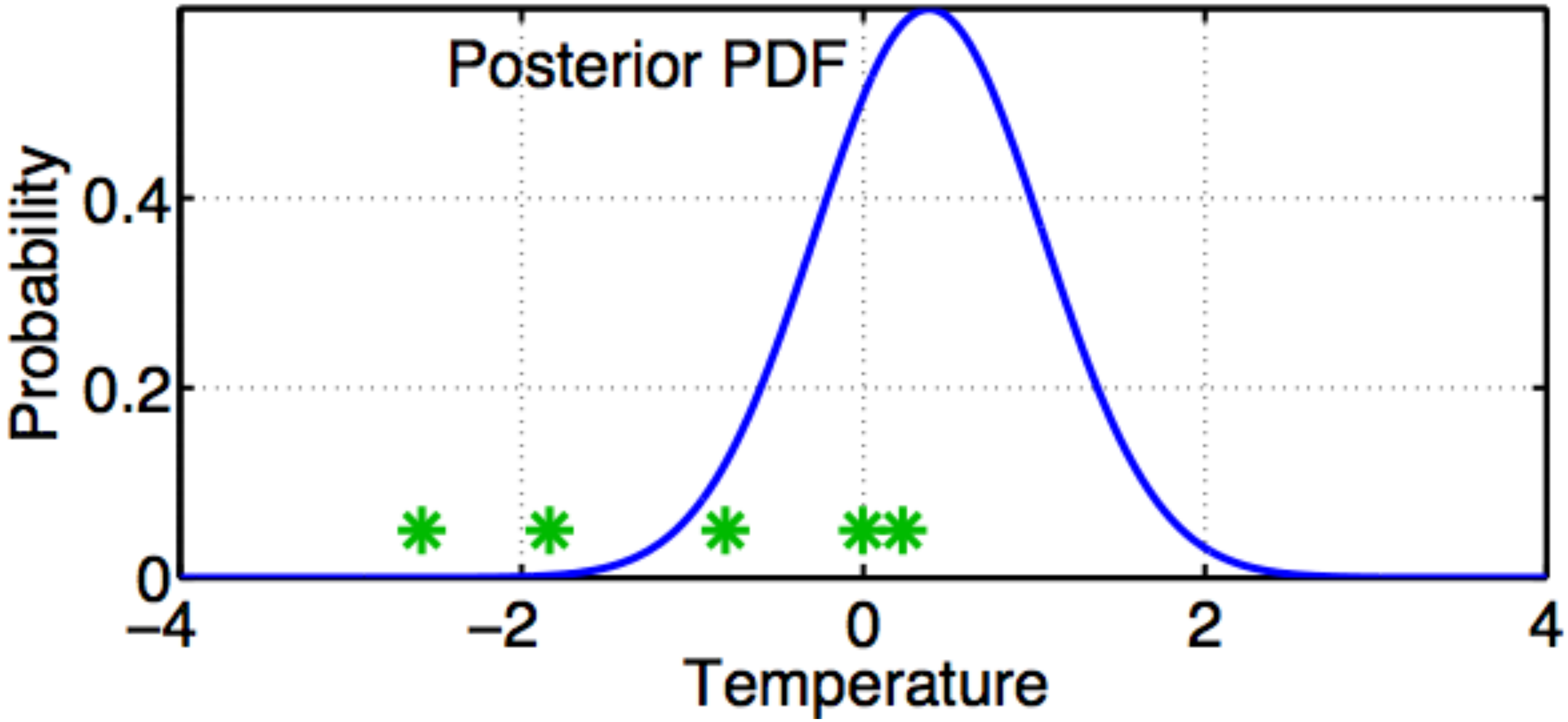
Get the observation likelihood.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



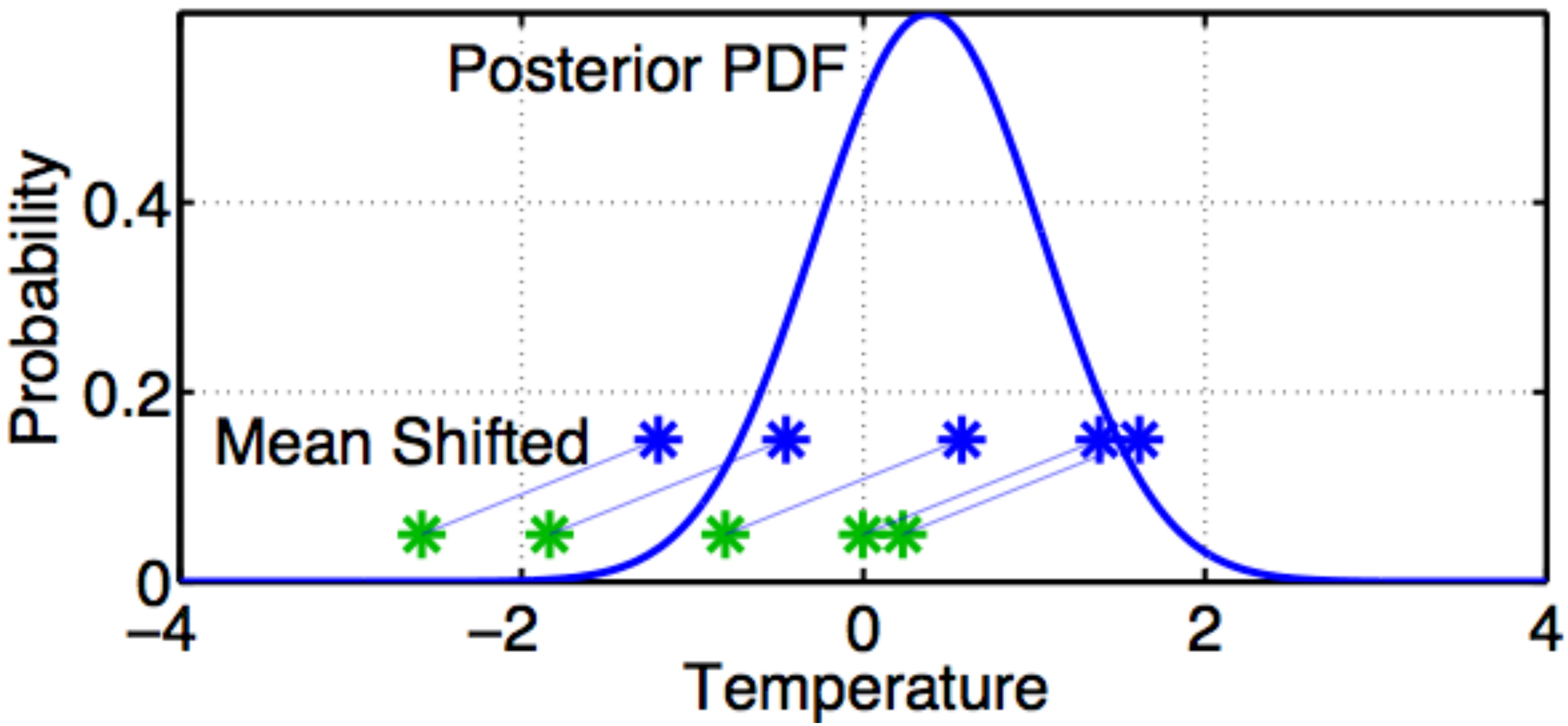
Compute the continuous posterior PDF.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



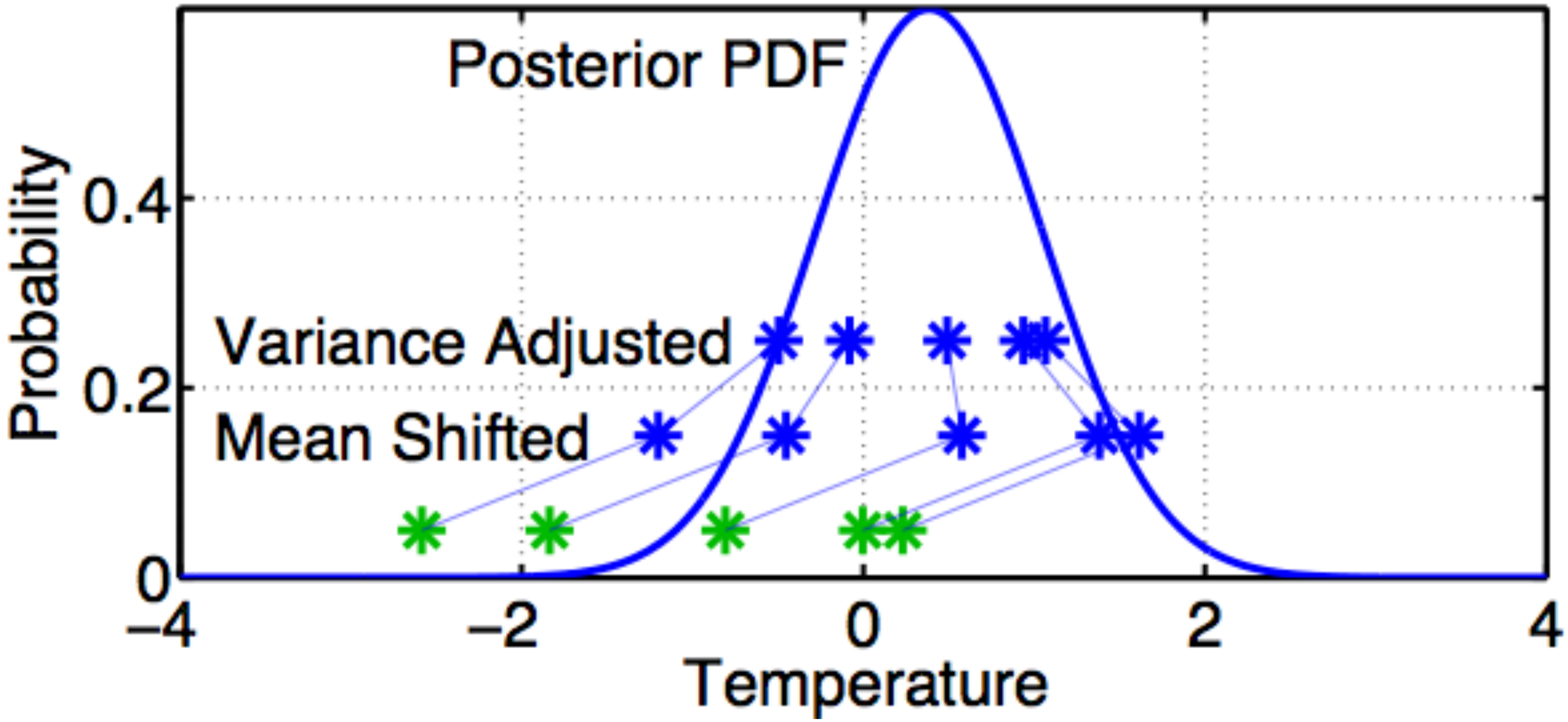
Use a deterministic algorithm to 'adjust' the ensemble.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



First, 'shift' the ensemble to have the exact mean of the posterior.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



First, 'shift' the ensemble to have the exact mean of the posterior.

Second, linearly contract to have the exact variance of the posterior.

Sample statistics are identical to Kalman filter.

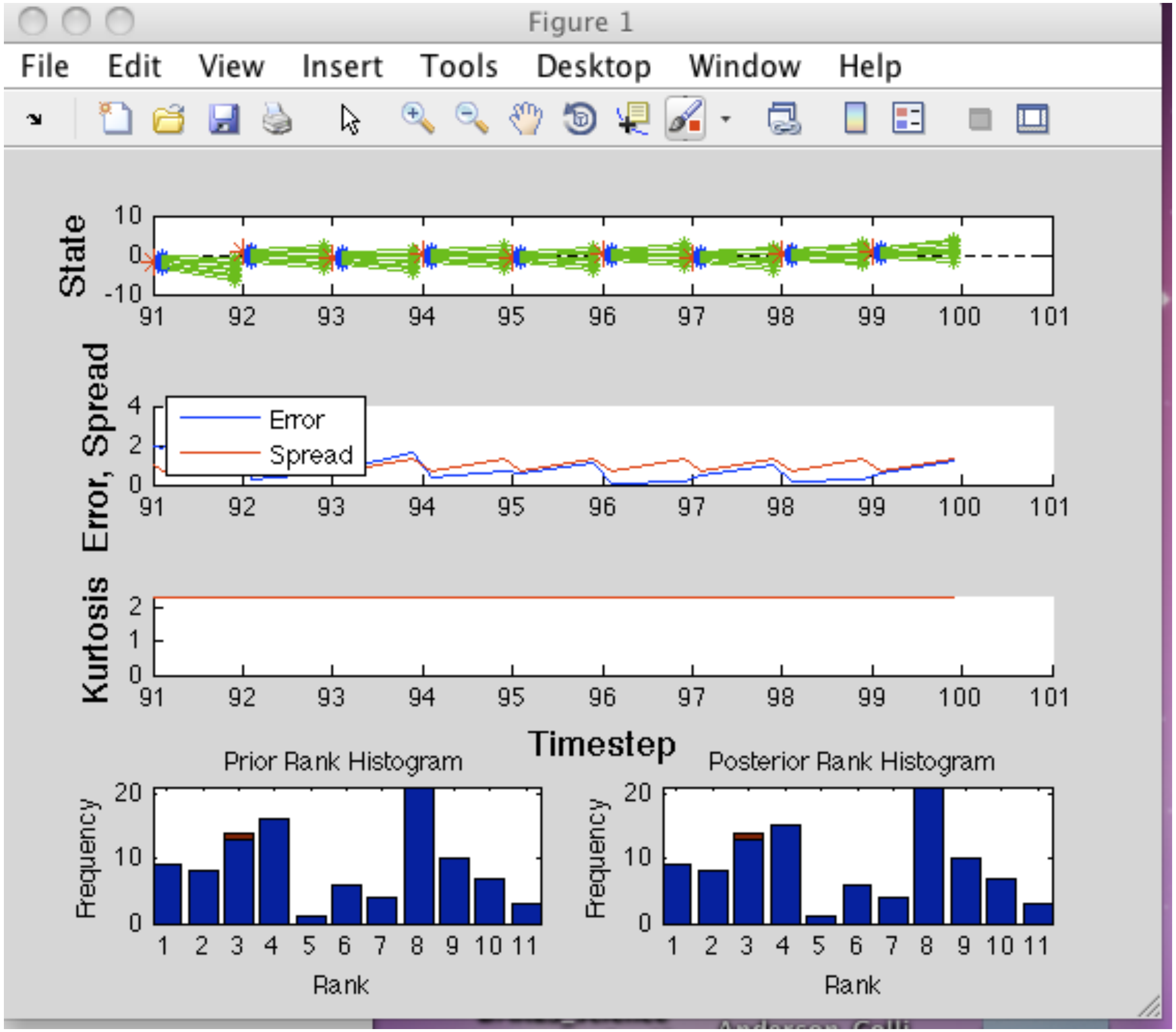
Ensemble Kalman Filter Weaknesses

- (Ensemble) KF optimal for linear model, gaussian likelihood.
- In KF, only mean and variance have meaning.

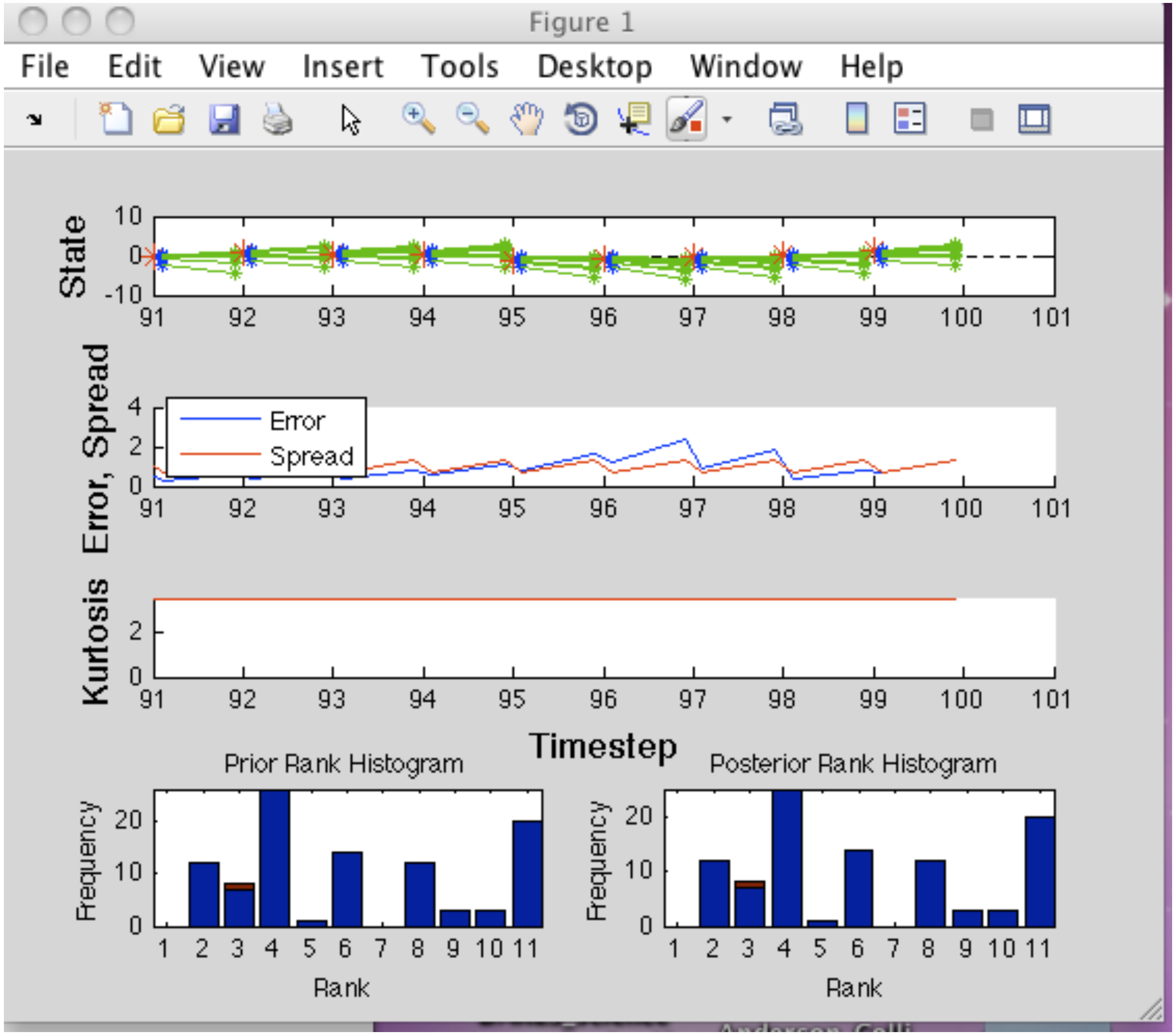
- Ensemble allows computation of many other statistics.
- What do they mean? Not entirely clear.

- Example: Kurtosis. Completely constrained by initial ensemble.
  It is problem specific whether this is even defined!

Ensemble Kalman Filter Weaknesses



Ensemble Kalman Filter Weaknesses



Use Caution with Any Ensemble Statistics

- No a priori reason to believe most statistics.
- Ensemble mean may be approximately correct.
- Variance may be correct from ensemble DA.
- Anything else, who knows (a priori).

Calibration and Validation

- No a priori reason to believe most statistics.
- Need to calibrate and validate (verify).
- No more about calibration here, but it's essential.
- Will only validate against observations.

(Weather services sometimes do weird other stuff).

Caution: Verifying Analyses

- Best practice is to verify forecasts.
  Observations at forecast times are independent.

- Verifying analyses can be tricky.
  Assimilated observations are not independent of analysis.
  Don't expect them to have same relationship to ensemble.
  Can verify analyses against withheld (not used) observations.

- Assimilated observations be used to verify analyses with ensemble DA.
  Requires sophisticated processing, rarely used.

Verifying Ensemble Forecasts: Ensemble Mean

- Need observation forward operator to compare to observations.
- Can compute RMSE or MAE of ensemble mean averaged over sets of observations.

  (Sets can span time and/or space).
- Generally want RMSE and MAE to be smaller.

Verifying Ensemble Forecasts: Ensemble Variance

- Sample variance is estimate of squared error of ensemble mean from truth.

  Sample standard deviation is estimate of error.

  Sometimes called spread / skill relation.

Verifying Ensemble Forecasts: Ensemble Variance

- But, only know observation, not truth.
- Have estimate of observation error variance (from somewhere).
- Estimate of squared error from observation is sum of ensemble and observation error variances.
- Square root of variance sum is sometimes called total spread.
- Are RMSE and total spread the same for set of observations?
  If not, ensemble isn't what we'd hoped.

Verifying Ensemble Forecasts: Other Statistics

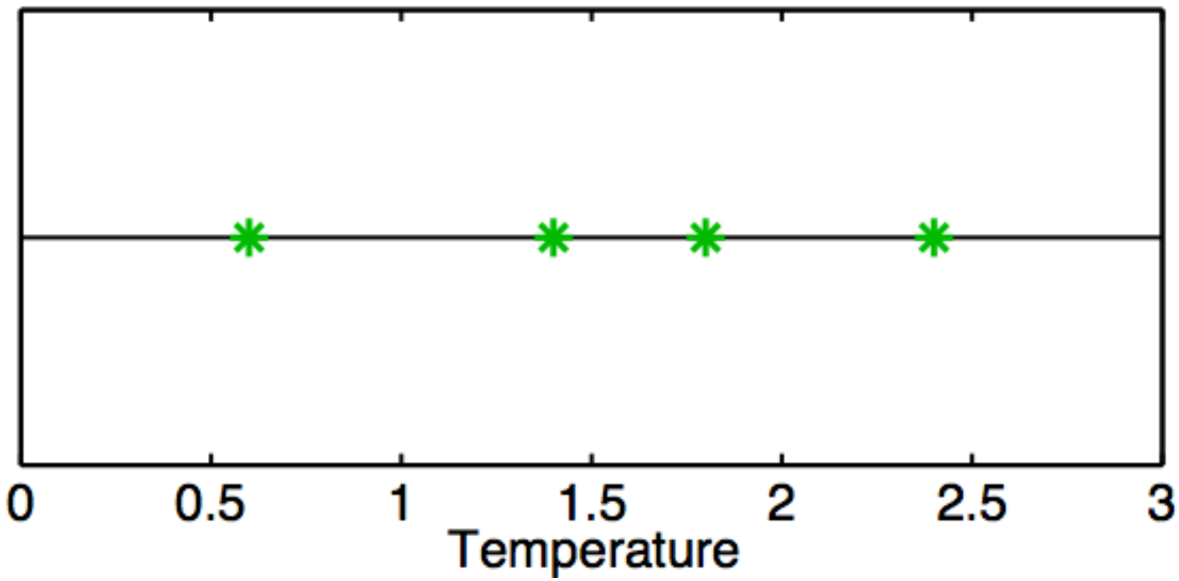
- Can compute many sample statistics from ensemble.
- When verifying, remember that observations aren't the truth.
- Example: What's probability that precipitation exceeds 1mm?
- First add random draw from observation error to ensemble members.
- Then, see what fraction of these are greater than 1mm.

The Rank Histogram

A non-parametric tool to evaluate the entire ensemble distribution.

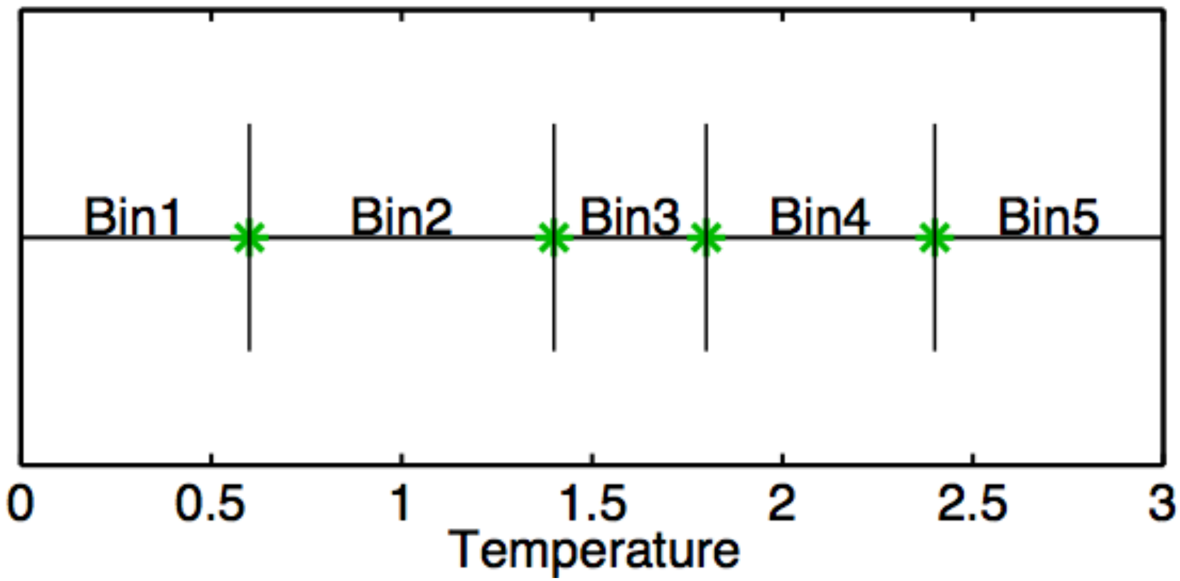
The Rank Histogram

Draw 5 values from a real-valued distribution.
Call the first 4 'ensemble members'.



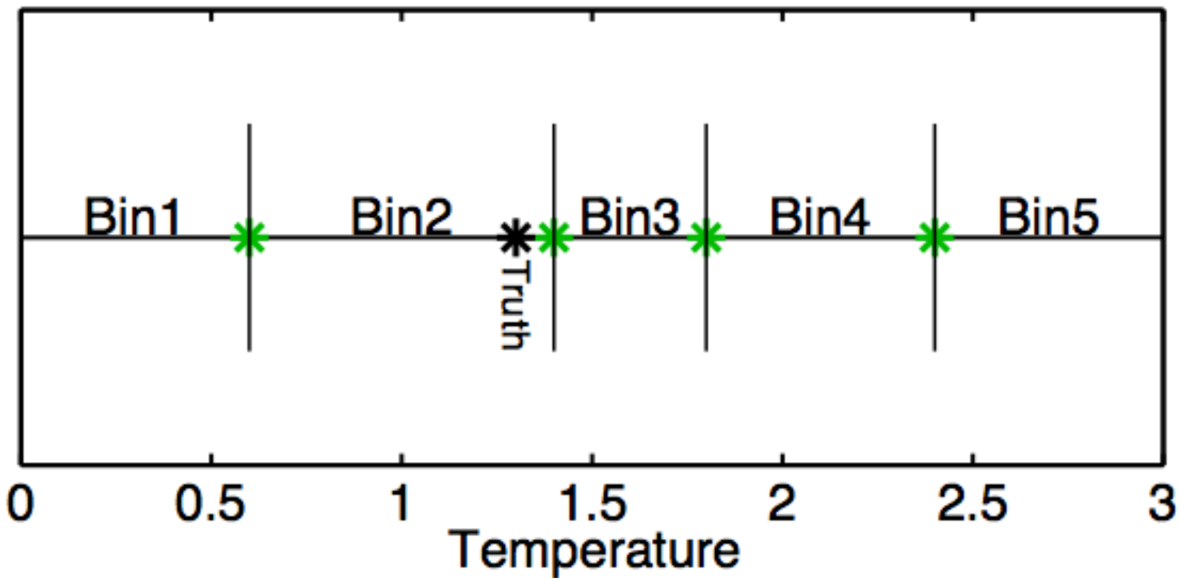
The Rank Histogram

These partition the real line into 5 bins.



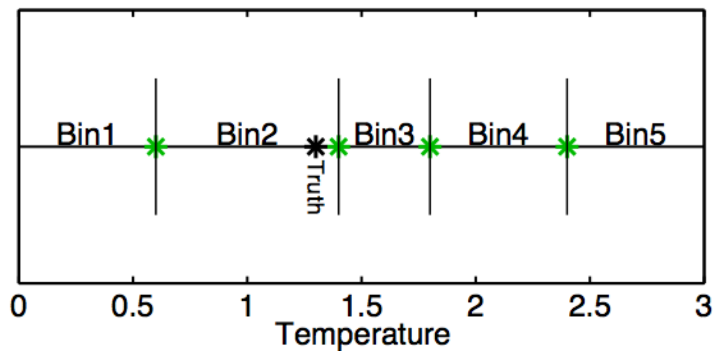
The Rank Histogram

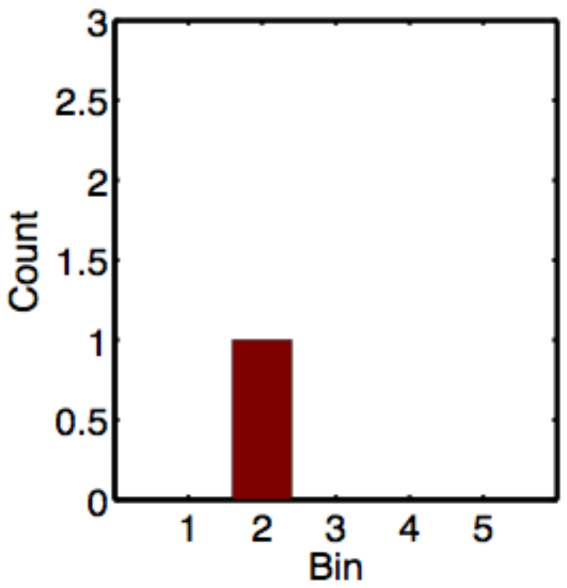
Call the 5th draw the ‘truth’.
1/5 chance that this is in any given bin.



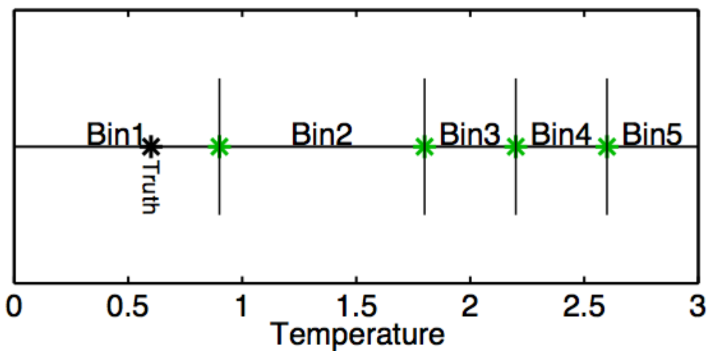
The Rank Histogram

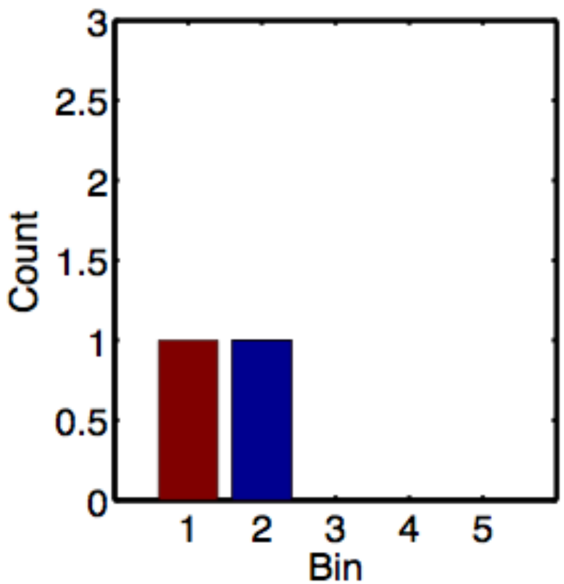
Rank histogram shows the frequency of the truth in each bin over many assimilations.





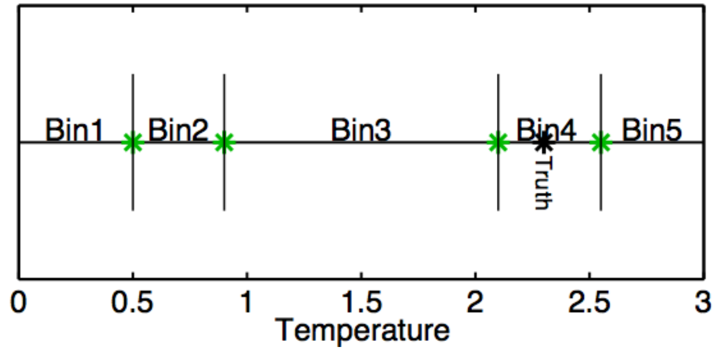
Rank histogram shows the frequency of the truth in each bin over many assimilations.

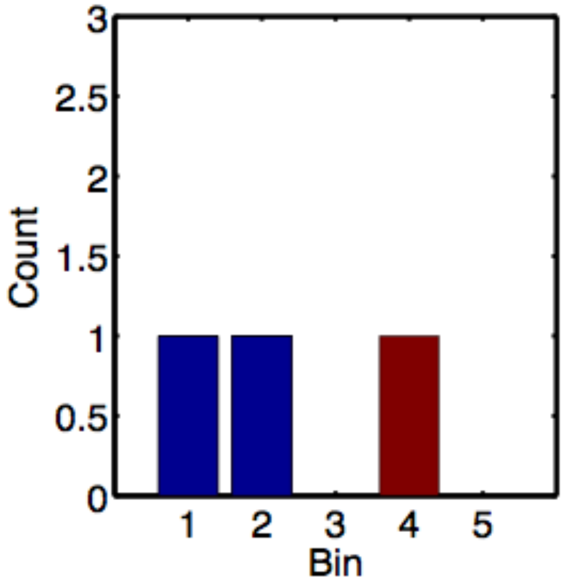




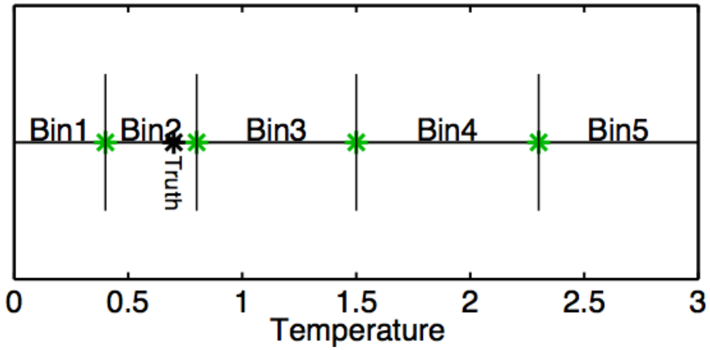
The Rank Histogram

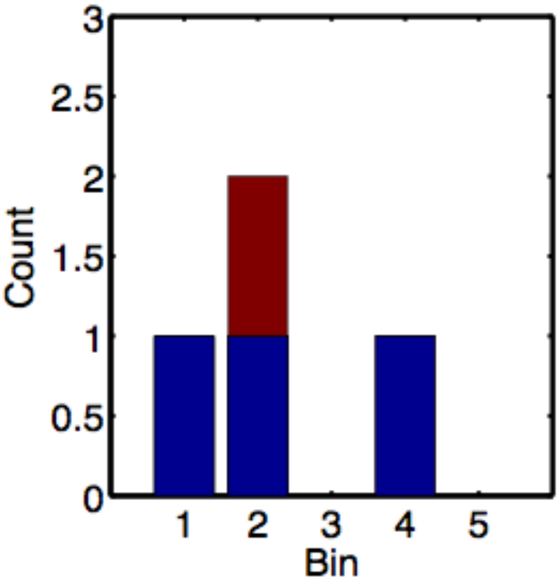
Rank histogram shows the frequency of the truth in each bin over many assimilations.





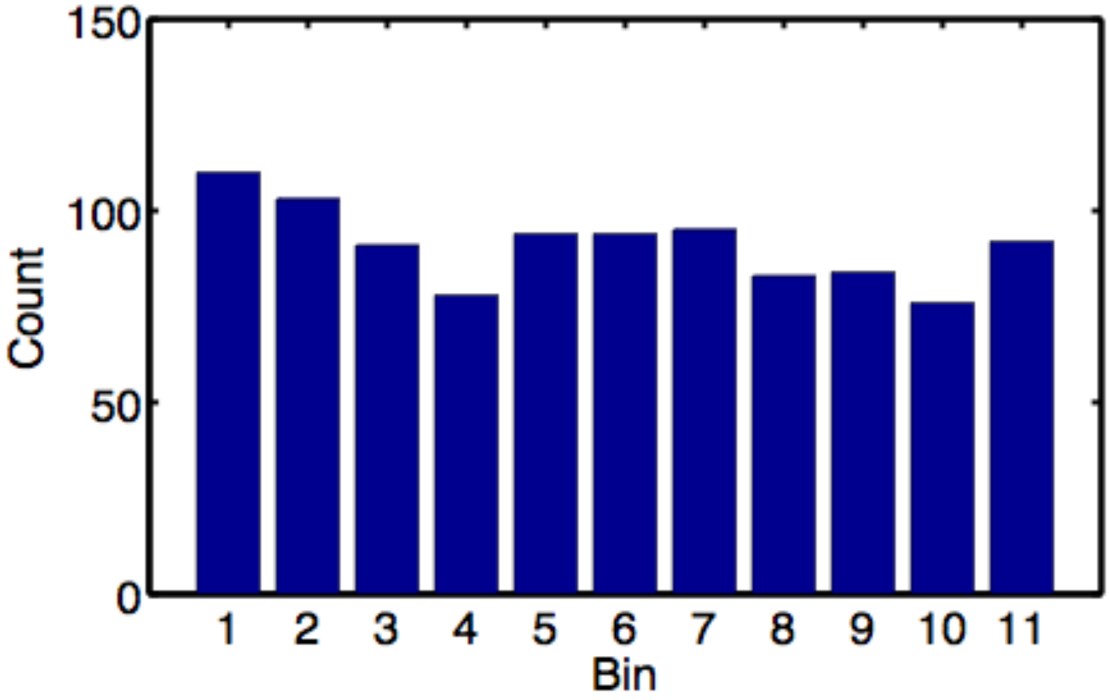
Rank histogram shows the frequency of the truth in each bin over many assimilations.





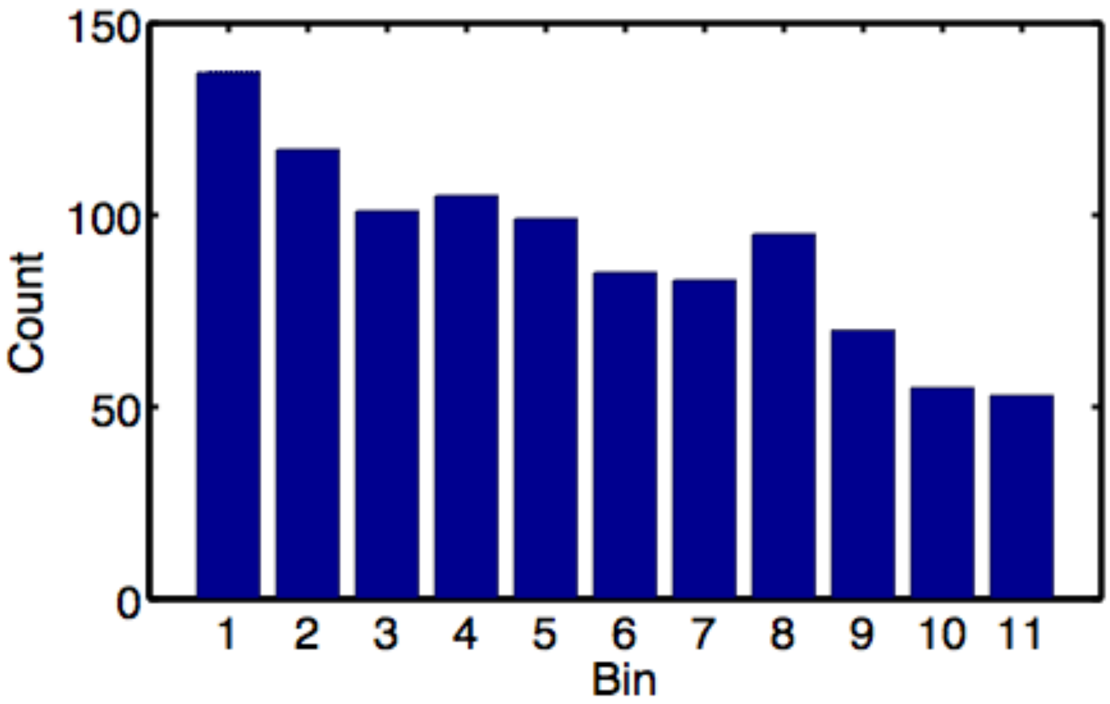
The Rank Histogram

- Rank histograms for good ensembles should be uniform (caveat sampling noise).
- Want truth to look like random draw from



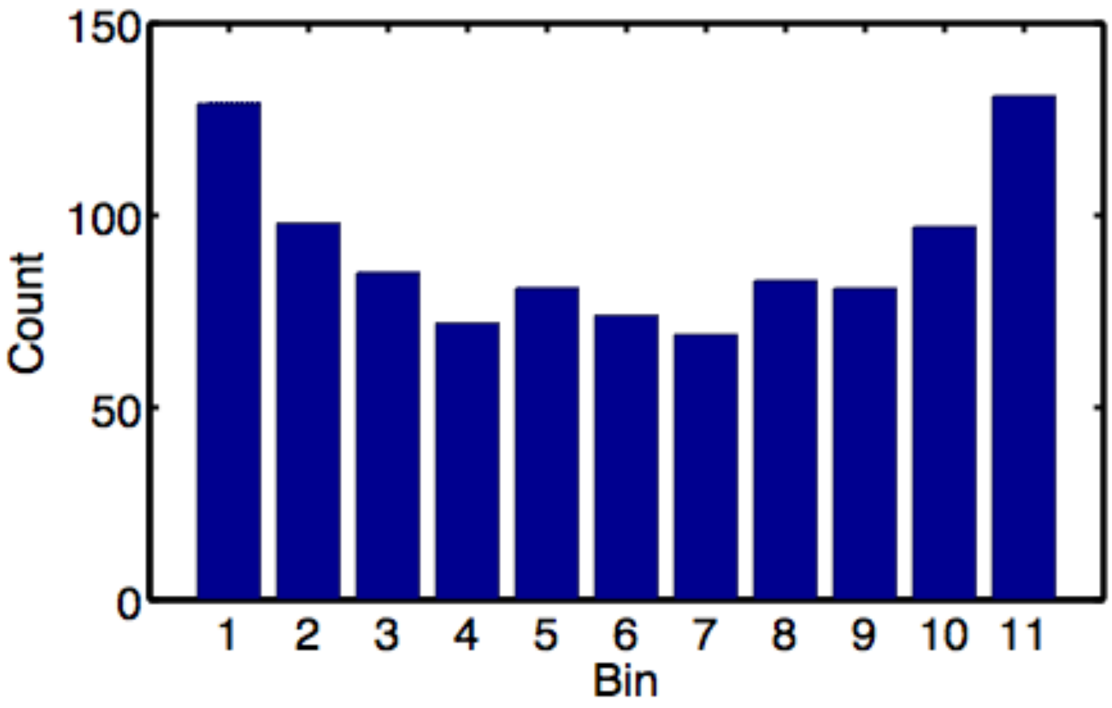
The Rank Histogram

A biased ensemble leads to skewed histograms.



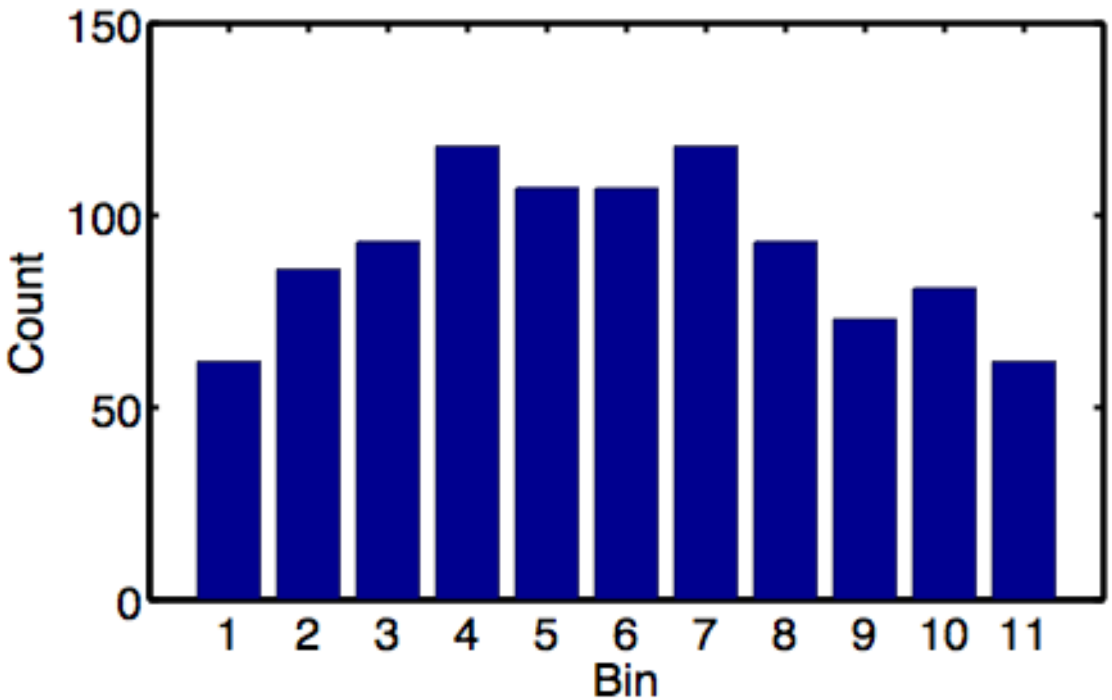
The Rank Histogram

An ensemble with too little spread gives a u-shape.
This is the most common behavior for geophysics.



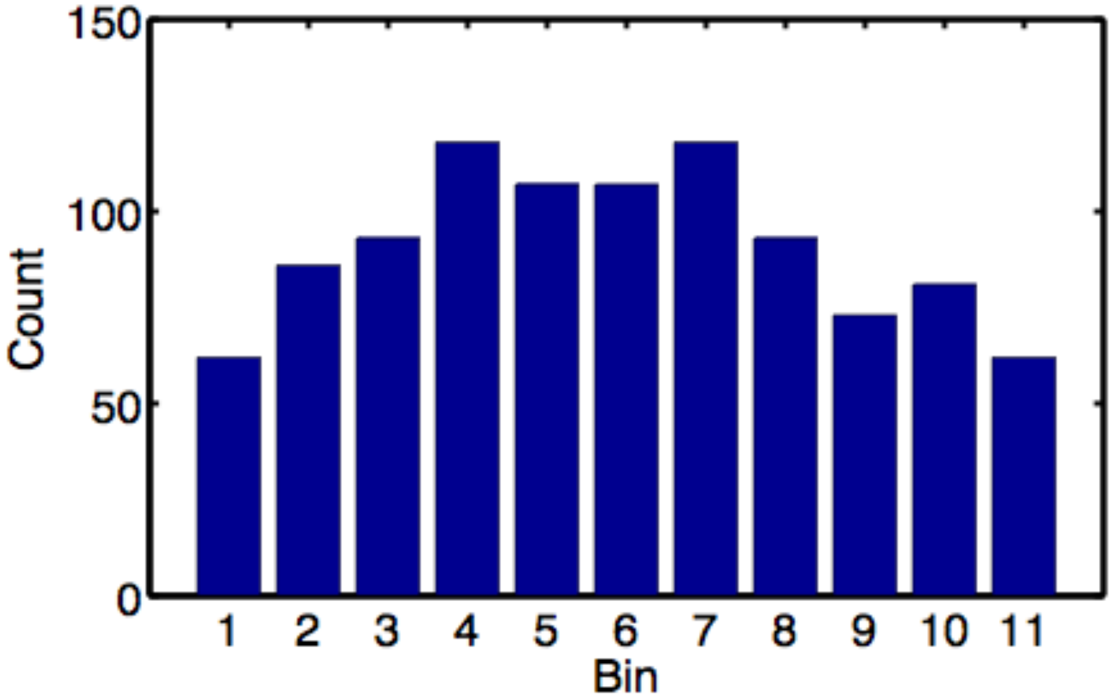
The Rank Histogram

An ensemble with too much spread is peaked in the center.



The Rank Histogram

- Observations are not the truth.
- Must add random draw from observation error distribution to each ensemble member first.



Multi-model Ensemble Verification

- Check statistics for each model independently.
- May want to weight ensembles when computing mean, variance, other statistics.
- For instance, weight by inverse of time mean model error variance.

Are Two Ensembles Different?

- A different but interesting problem.

- Use a variety of standard statistical tools.
  T-test on points, or regions, or time means, or …
  Kolmogorov-Smirnov test on full ensembles at points.

- Need to do field significance when statistics are computed at points.