NOAA Earth System
Research Laboratory

A kernel-density based ensemble filter applicable (?) to high-dimensional systems

Tom Hamill
NOAA Earth System Research Lab
Physical Sciences Division

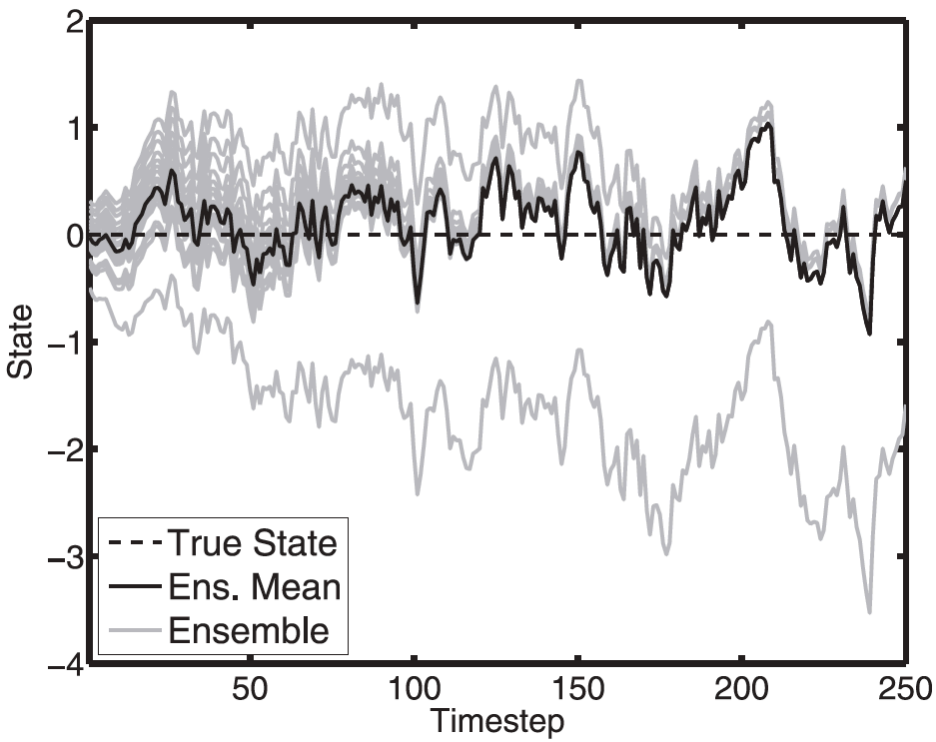
Presentation at 91st AMS Annual, Seattle, WA, 15th IOAS-AOLS

“Research is the process of going up alleys to see if they are blind.”

Marston Bates

What to do when in ensemble filters when prior is obviously non-Gaussian?

Cycled EnKF can create non-Gaussian states, especially when (a) ensemble size is large, and (b) when the forecast model contains significant nonlinearity. See Lawson and Hansen, MWR, 2004; Mitchell and Houtekamer, MWR, 2009 and Anderson, MWR, 2010.



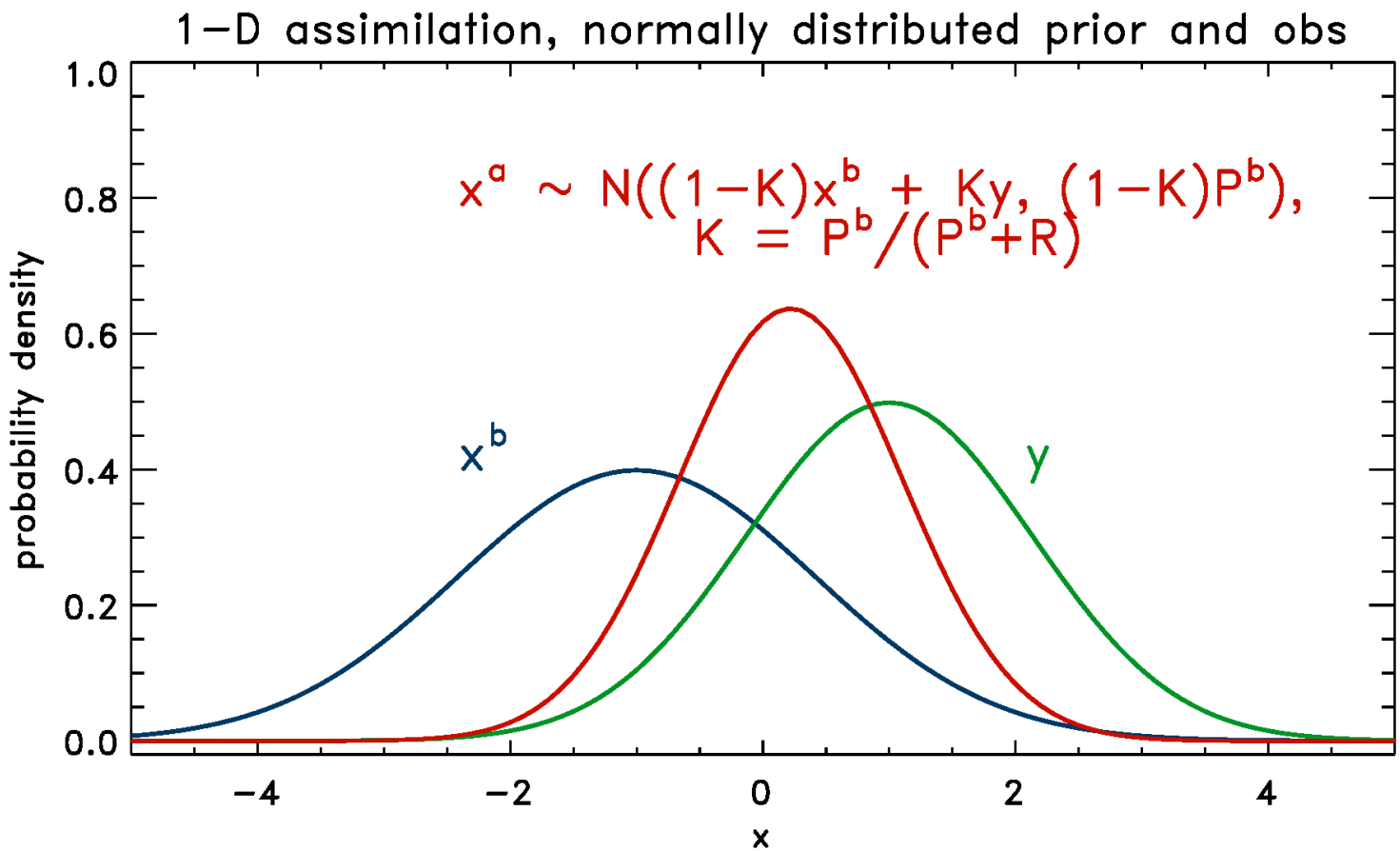
Example: data assimilation with EAKF in simple dynamical system where

$x_{t+1}=x_t + 0.05\,(x_t + \alpha x_t\,|x_t|)$

where $\alpha=0.2$, n=20, obs~ N(0,1)

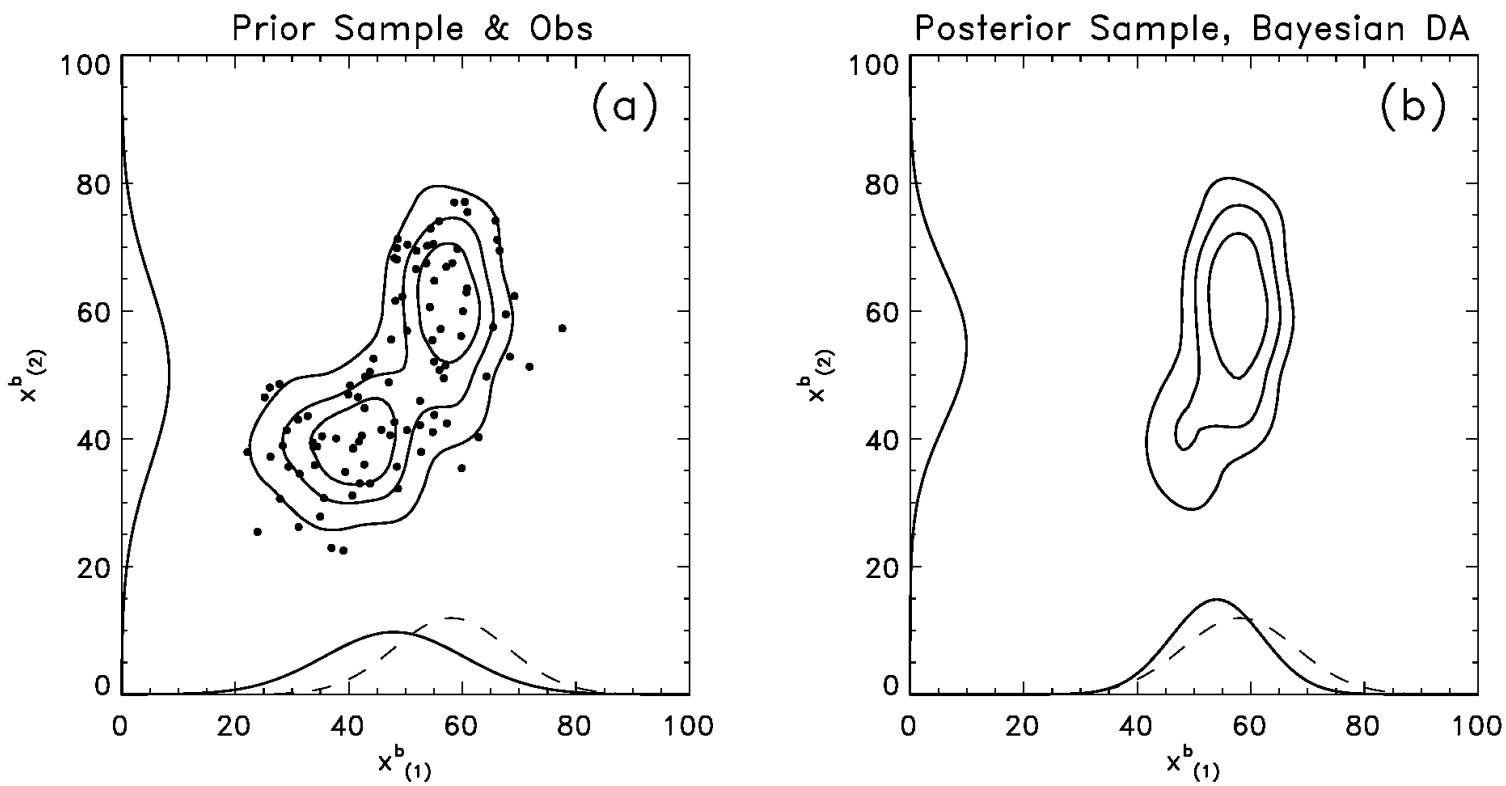
from Jeff Anderson, MWR, 2010.

Gaussian distributions easy.



Most common data assimilation updates are well behaved when prior and observation are normally distributed, as shown here, with an analytical solution for this 1-D problem.

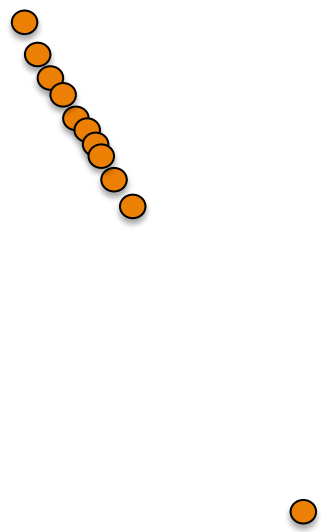
Estimating prior as fully non-Gaussian may work when state dimension is small.



Here ensemble is used to estimate probability density in 2-D, which is then Bayesian updated to an observation in one component. Approaches such as these are computationally and scientifically impractical for large dimensional systems, e.g., "curse of dimensionality."

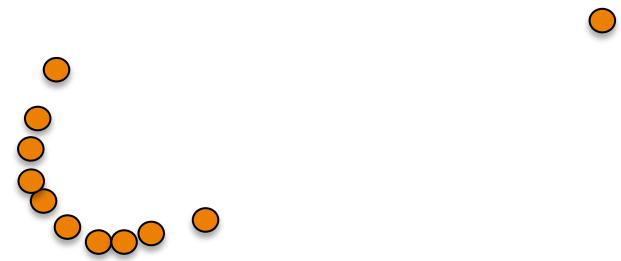
from Hamill, chapter 6 in "Predictability of Weather and Climate"

Flavors of non-Gaussianity



(1) Ensemble non-Gaussian, but very high correlation between state components, i.e, effectively non-Gaussian only in 1 direction.

(potentially solveable, & intended focus of this talk)



(2) Ensemble non-Gaussian, curved attractor

(well beyond my current capabilities & available computational resources)

Two insights of Jeff Anderson relevant to non-Gaussian data assimilation (of the first type)

(1) Serial ensemble filters can split the update into two steps (Anderson, MWR, 2003) :

(a) Update prior at observation location to the observation

(b) Regress increments to the rest of the state to the updated observation prior.

Insight 2: only in updating the *observation prior*, relax the Gaussianity assumption.

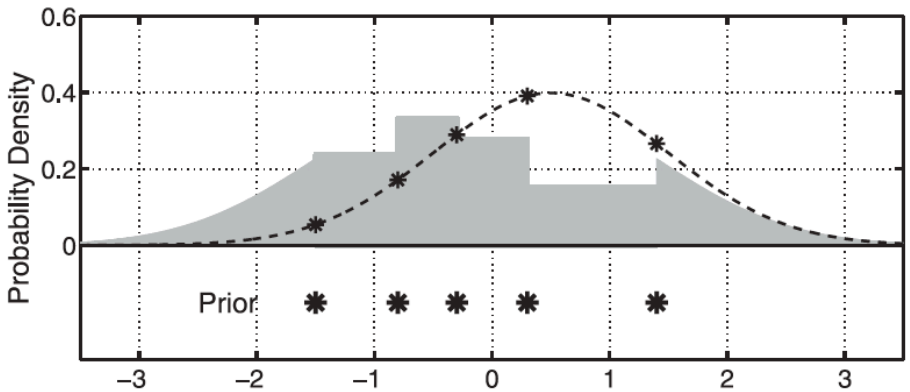


FIG. 6. Schematic of first phase of rank histogram filter algorithm. The locations of five prior ensemble members are indicated by large asterisks at the bottom. The continuous approximation to the prior probability density is indicated by the four shaded boxes and the shaded portions of Gaussians on the tails. The continuous likelihood is the dashed line with the values at the ensemble members marked by small asterisks.

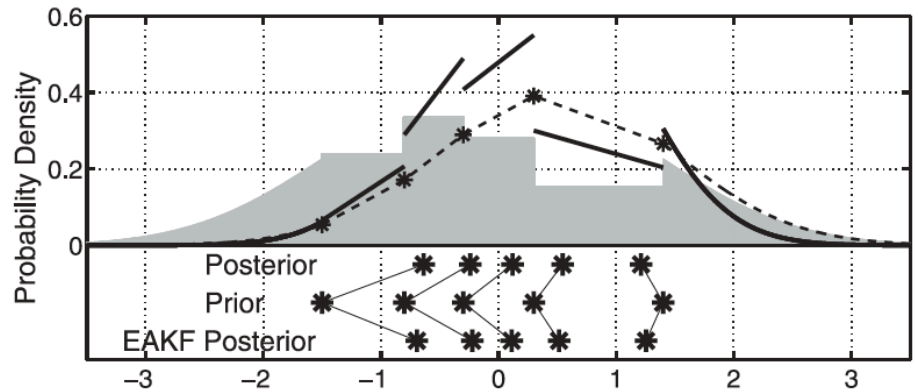
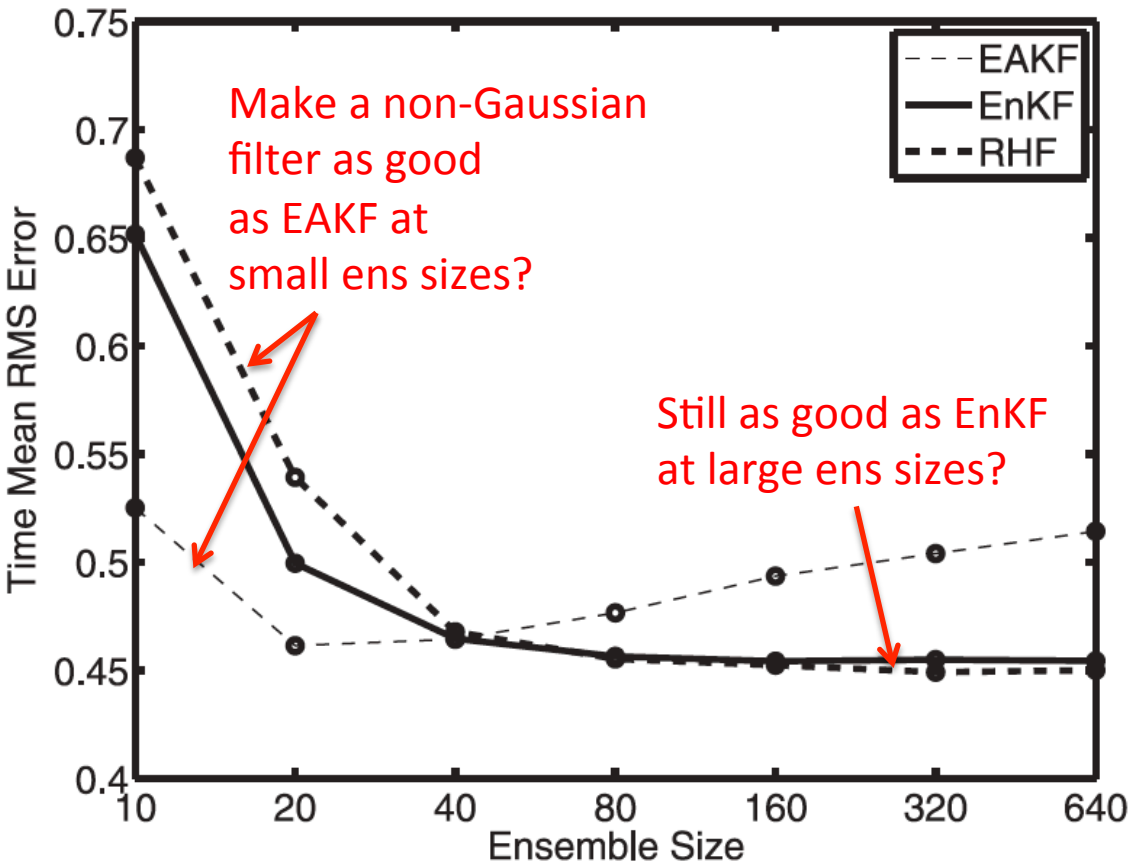


FIG. 7. Schematic of the rank histogram filter algorithm. The prior and posterior ensembles along with the posterior from an ensemble adjustment Kalman filter are marked by asterisks at the bottom. The continuous approximation to the prior probability density is shaded. The dashed line is a piecewise linear interior approximation to the likelihood. The continuous posterior probability distribution is the thick solid line.

"Rank Histogram Filter" -- **a probability mass of 1 / (n+1) is assigned between each ensemble member observation prior**. Given the observation likelihood (dashed), piecewise construct a product of the prior and likelihood. For i^{th} sorted observation prior, determine the value associated with the posterior CDF at i / (n+1). This replaces step (1) of Anderson (2003). From Anderson, MWR, (2010).

Anderson's tests with Lorenz '96 model



F=8.0; fixed localization "half width" of 12 grid points.

Observations

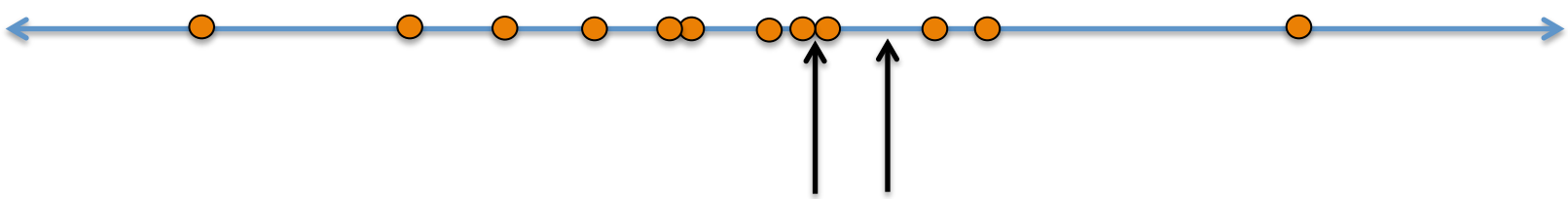
$$y_j = (x_j + x_{j+1})/2 + \text{Normal}(0,4), \quad j = 1, \ldots, 40.$$

Adaptive spatially varying inflation following Anderson, Tellus, 2009.

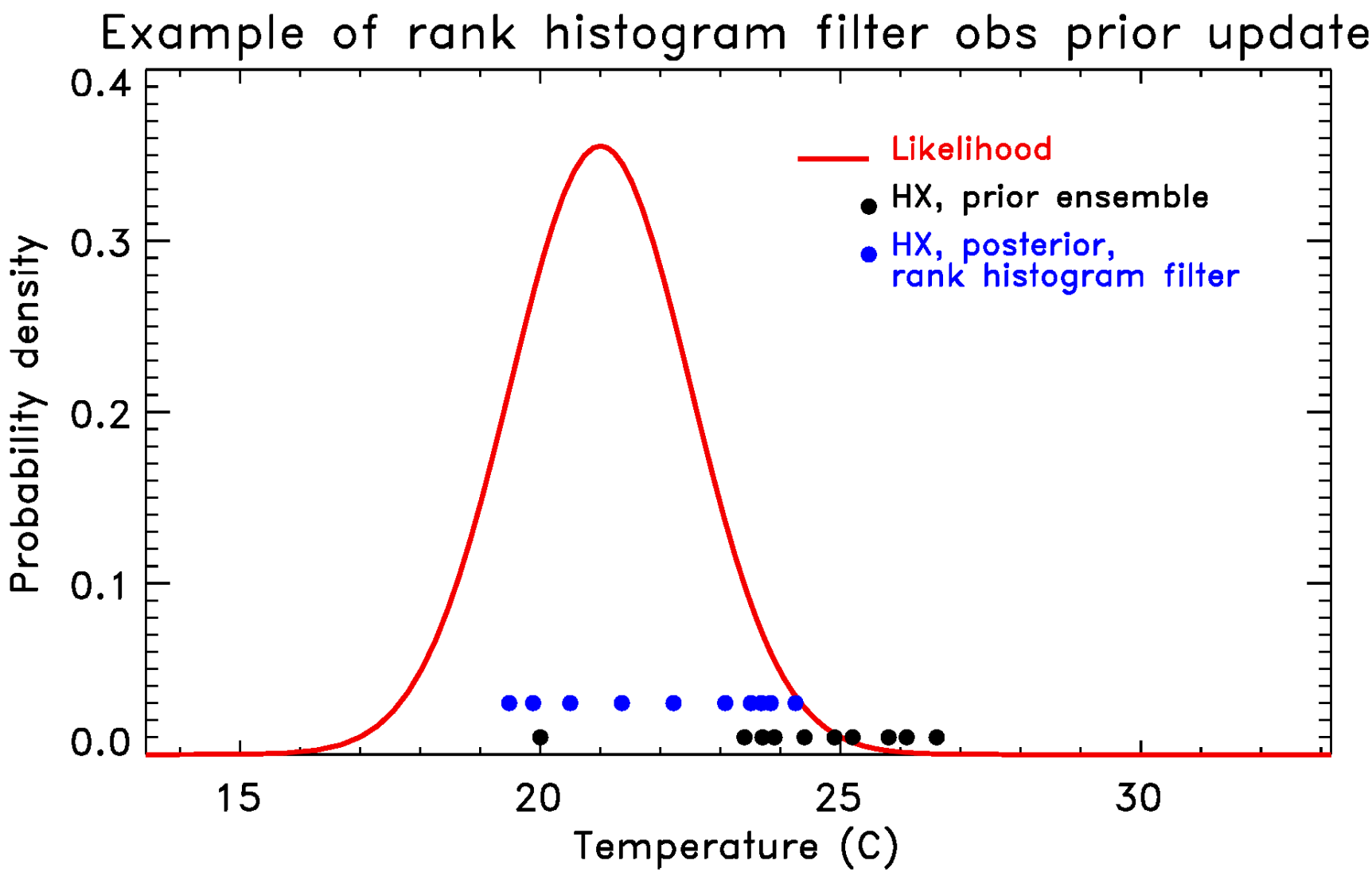
Note poor performance of both RHF and EnKF when ensemble size is 10 (see Whitaker and Hamill MWR 2002), poor performance of EAKF when ensemble size is large (Lawson and Hansen, MWR 2003).

Possible errors introduced with rank histogram filter (RHF)

- Ensembles may, due to sampling variability, have large or small deviations between members. Are there consequences of having 1/(n+1) probability between each?



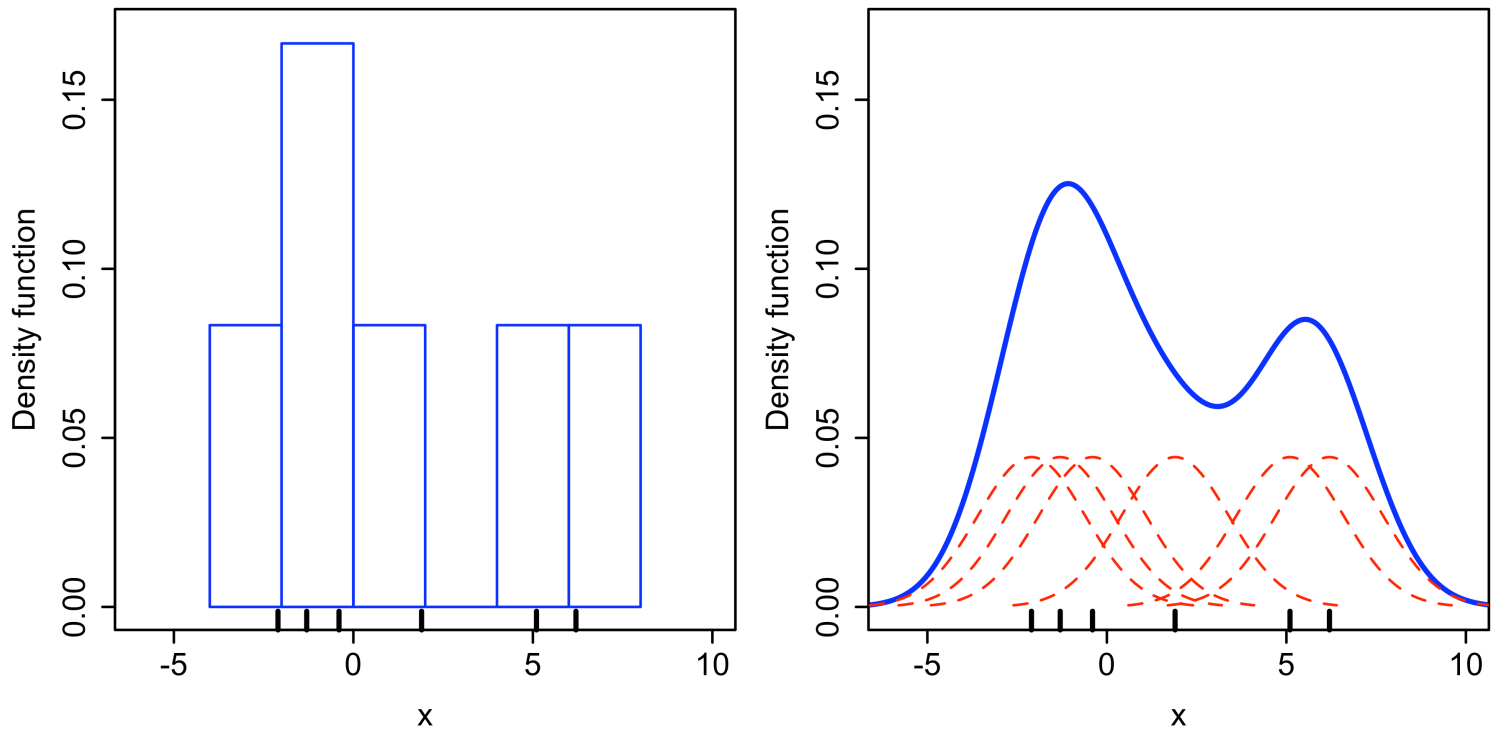
same 1/(n+1) probability mass between both?



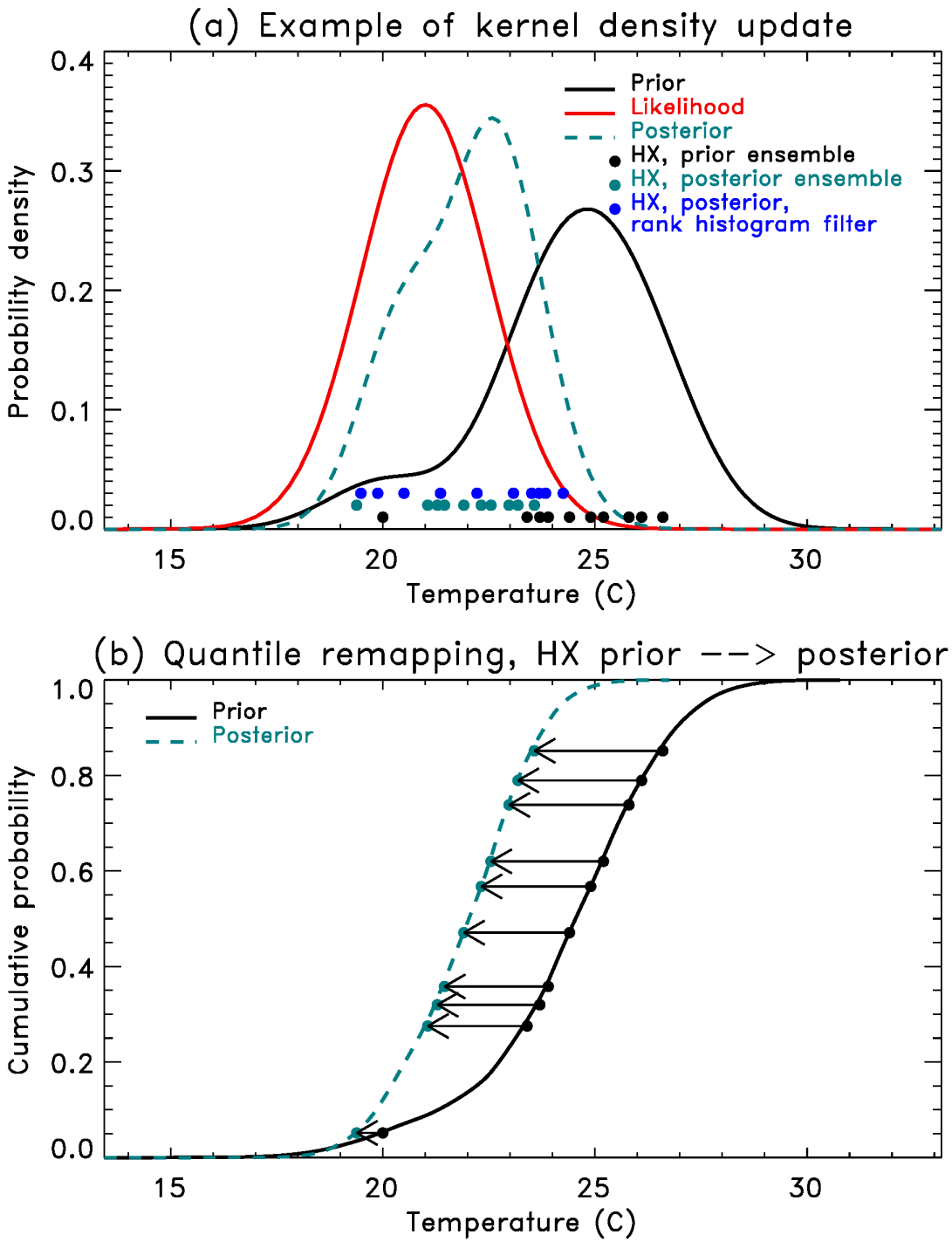
Surprisingly, the gap between the 1st and 2nd ensemble member in the prior is obliterated with the rank histogram filter.

Kernel density approach?

- Model observation prior with kernel density estimation techniques.



source: Wikipedia, “kernel density estimation.”



Example of observation prior update with "kernel density filter" (KDF)

If prior ensemble is significantly non-Gaussian (as determined through Kolmogorov-Smirnov test) then replace standard ensemble filter's obs prior update with KDF update. Model the observation prior (black lines, panel a) with kernel density estimation. Construct posterior as product of prior and likelihood (panel a). Generate posterior members by finding the ensemble members' quantiles in the prior cdf. Posterior members are the values associated with those quantiles in posterior cdf (panel b).

When to apply KDF

- It's much more expensive than EnKF, EnSRF, EnAF for updating obs prior.
- My choice: use only when observation prior is statistically significantly non-Gaussian (Kolmogorov-Smirnov test, α=0.05)
- Departures from Gaussianity more commonly determined to be significant with large ensemble sizes than with small.

Testing in Lorenz '96 model

- Classic L96 model, 40 variables, perfect model, F=8.0, dt = 0.05, cycle for 10 years after spinup.
- Test EnKF, EnSRF, EnAF/KDF, EnAF/RHF over range of localization length scales, ensemble sizes.
- Covariance inflation : $1 + 1./\text{nanals}^{0.96}$

Lorenz '96, perfect model

(obs σ=5.0, 6 h between obs, obs every grid point)

Size	EnKF RMS error	fraction fail K-S test	EnSRF RMS error	KDF RMS error	RHF RMS error	fraction fail K-S test
10	1.524	0.0	1.480	1.480	1.480	0.0
20	1.316	0.0001	1.310	1.285	1.304	0.0007
40	1.226	0.0002	1.214	1.206	1.205	0.005
80	1.165	0.004	1.169	1.142	1.139	0.04
160	1.133	0.03	1.157	1.117	1.126	0.15
320	1.114	0.13	1.156	1.122	1.164	0.31
640	1.102	0.31	1.159	1.180	1.185	0.62

(1) RHF/EAKF and KDF/EAKF not employed at smallest ensemble sizes since departures from Gaussianity not detected. Same error as EnSRF.

(2) Surprisingly, benefit of RHF/EAKF & KDF/EAKF over EnSRF is at moderate ensemble sizes, with degradation at large ensemble sizes.

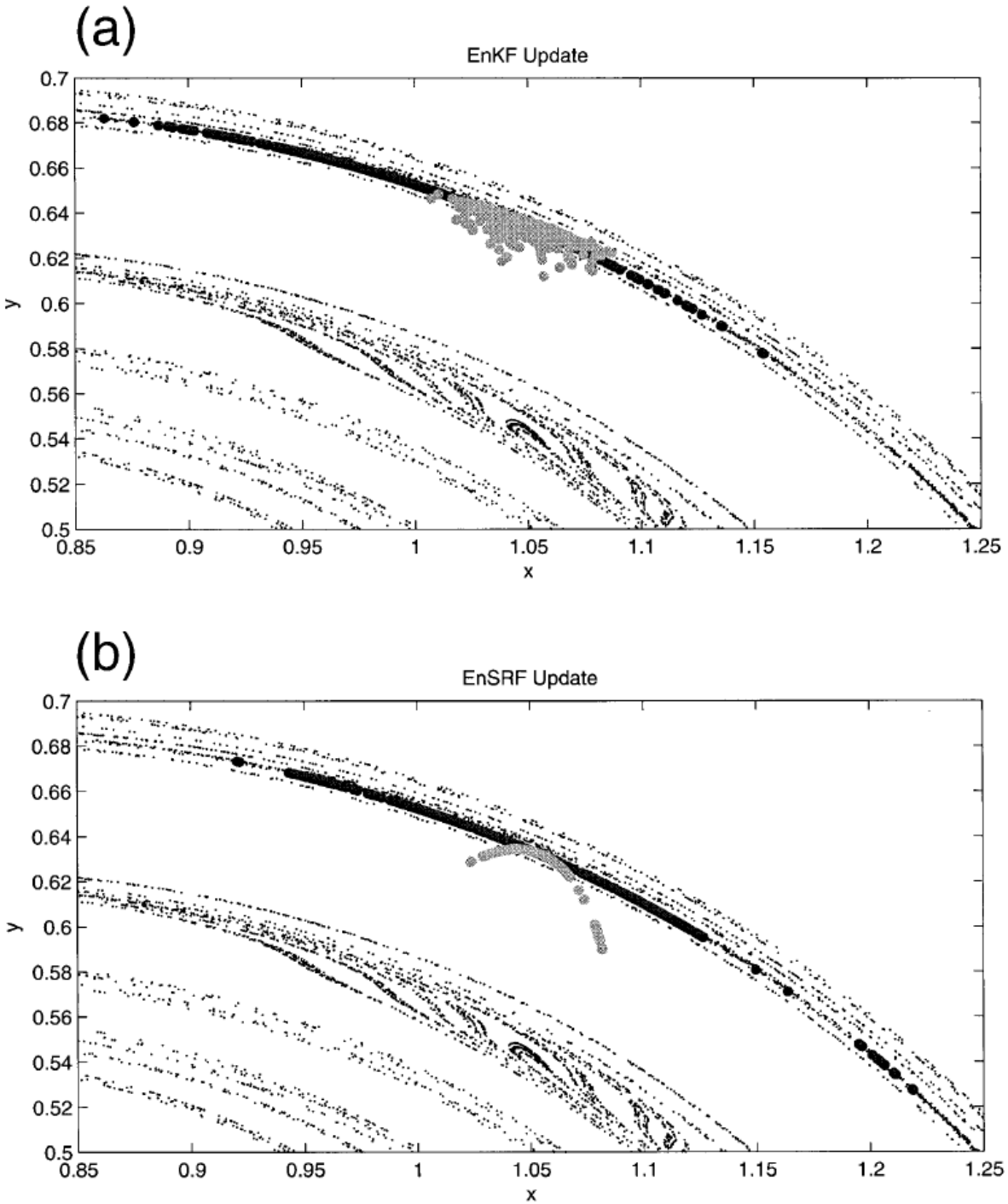
Lorenz '96, perfect model

(obs σ=5.0, 6 h between obs, obs every grid point)

Size	EnKF RMS error	fraction fail K-S test	EnSRF RMS error	KDF RMS error	RHF RMS error	fraction fail K-S test
10	1.524	0.0	1.480	1.480	1.480	0.0
20	1.316	0.0001	1.310	1.285	1.304	0.0007
40	1.226	0.0002	1.214	1.206	1.205	0.005
80	1.165	0.004	1.169	1.142	1.139	0.04
160	1.133	0.03	1.157	1.117	1.126	0.15
320	1.114	0.13	1.156	1.122	1.164	0.31
640	1.102	0.31	1.159	1.180	1.185	0.62

EnKF detects non-Gaussianity less frequently than deterministic filters. For typical 50-100 members, non-Gaussian conditions virtually never diagnosed.

Lawson & Hansen's insight using Ikeda Map



black dots = prior ensemble
grey dots = posterior ensemble

EnKF & effect of perturbed observations creates more normally distributed posterior, while EnSRF compresses but keeps curved prior shape.

(EnAF/KDF or EnAF/RHF would do something similar).

Lawson & Hansen, MWR, 2004.

Tests of KDF in global primitive equation model

- Tested in 2-level dry global PE model recently used in Hamill and Whitaker (2011 MWR). Same uniform observation network. Imperfect model assumption.
- Bottom line: effectively no change from using EnSRF data assimilation to using EnSRF/KDF data assimilation.
  - Reason is that KDF is virtually never invoked, since only invoked when prior is significantly non-Gaussian. EnSRF consistently used.
- Admittedly, it's still a toy model w/o moisture, other complicating aspects.

Why is non-Gaussianity actually more rare than one might think?

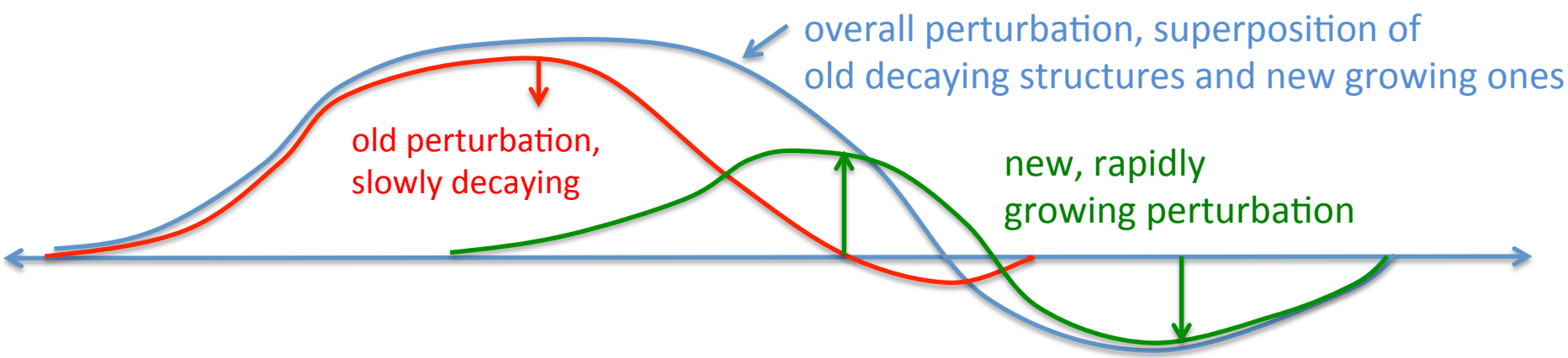
- (1) In practical NWP applications, we often account for model error or sampling variability through additive noise with Gaussian properties.

$$\mathbf{P}^b = \mathbf{M}\mathbf{P}^a\mathbf{M}^T + \mathbf{Q}$$

$$\mathbf{x}^b = \mathbf{M}\mathbf{x}^a + \mathbf{e}, \qquad \mathbf{e} \sim N(\mathbf{0},\mathbf{Q})$$

Why is non-Gaussianity actually more rare than one might think?

- (2) Ensemble forecast members in more complex models may project onto several modes, some growing, some decaying. This may randomize the perturbation structures.



Conclusions

- Many previous researchers have dashed themselves to pieces trying to surf the reef that is non-Gaussian data assimilation (me too).
- Some oldies but goodies, like perturbed-obs EnKF & deterministic filters with noise to account for model and sampling error, continue to be hard to beat for NWP applications.
- Possible increased relevance of non-Gaussian techniques as ensemble filters begin to deal with moisture-related variables.

KDF technical detail.

- Q: How does one determine the kernel type and width used?
- A: Gaussian kernel used, but may not be a critical detail. Optimal width is a function of ensemble size, smaller for large sizes. The optimal width estimated by repeatedly:
  - creating an ensemble drawn from standard normal.
  - choosing kernel width, creating pdf estimate.
  - evaluating the integrated square error (ISE) of the estimated to the analytical cdf.
  - finally, for a given ensemble size, choose the width "w" that has the lowest average integrated square error.
  - width that's actually used for an ensemble with a spread of s is s*w.