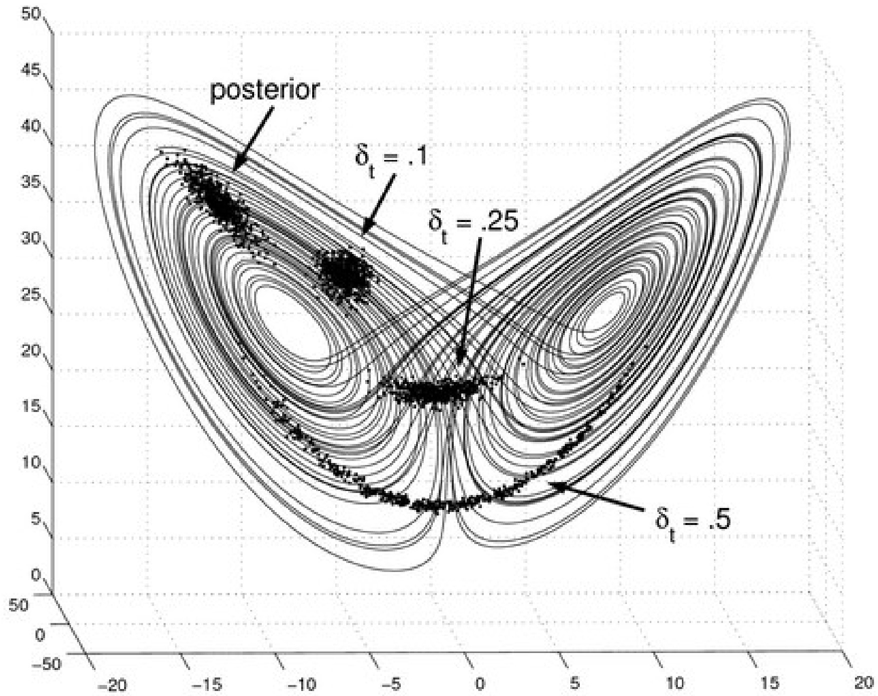
Non-Gaussianity in Data Assimilation (A Brief Tour)



Bengtsson et al., 2003: *J. Geophys. Res.*, **62**(D24), 8775-8785.

▷ Chris Snyder, National Center for Atmospheric Research

Preliminaries

Notation

- will follow Ide et al. (1997), generally
- $\mathbf{x}$ = atmospheric state written in terms of a finite, discrete basis, e.g. grid-point values or Fourier coefficients
- $\mathbf{y}$ = set of observations valid at a given time
- $\dim(\mathbf{x}) = N_x$, $\dim(\mathbf{y}) = N_y$

The Bayesian view

- true $\mathbf{x}$ can not be known, so consider $\mathbf{x}$ as random variable
- let subscripts indicate times, $\mathbf{x}_k = \mathbf{x}(t_k)$
- our goal is to calculate $p(\mathbf{x}_k|\mathbf{y}_1, \ldots, \mathbf{y}_l)$

Preliminaries (cont.)

Terminology

- "analysis" (pdf) is $p(\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_l)$ with $l=k$;
  i.e., pdf of state $\mathbf{x}_k$ conditioned on observations up to same time, t_k.
- "forecast" (pdf) is $p(\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_l)$ with $l=k-1$;
  i.e. pdf of state $\mathbf{x}_k$ conditioned on obs up to previous time, t_{k-1}.

Preliminaries (cont.)

Bayes rule

- definition of conditional pdf: $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x}, \mathbf{y})/p(\mathbf{y})$
- similarly, $p(\mathbf{y}|\mathbf{x}) = p(\mathbf{x}, \mathbf{y})/p(\mathbf{x})$
- thus,

$$p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})/p(\mathbf{y})$$

More terminology

- $p(\mathbf{x})$ is the "prior:" what we know about the state *before* the obs
- $p(\mathbf{x}|\mathbf{y})$ is the "posterior:" what we know *after* the observations
- $p(\mathbf{y}|\mathbf{x})$ is the "observation likelihood:" a conditional pdf for **y** that we treat as a function of **x**. Requires knowledge of the statistics of measurement and representativeness errors.

Sequential, Bayesian Assimilation

Suppose we have a forecast for $\mathbf{x}_k$ and new observations $\mathbf{y}_k$

- use Bayes rule to "update" (calculate analysis pdf)

  $$p(\mathbf{x}_k|\mathbf{y}_1,\dots,\mathbf{y}_k) = p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_1,\dots,\mathbf{y}_{k-1})/p(\mathbf{y}_k)$$

- sequential: $\mathbf{y}_k$ is needed for computation of $p(\mathbf{y}_k|\mathbf{x}_k)$, then discarded
- sequential form requires that $\mathbf{y}_k$ is conditionally independent of all previous observations given $\mathbf{x}_k$. In general (e.g. obs errors correlated in time), $p(\mathbf{y}_k|\mathbf{x}_k)$ and $p(\mathbf{y}_k)$ should be conditioned on $\mathbf{y}_1,\dots,\mathbf{y}_{k-1}$

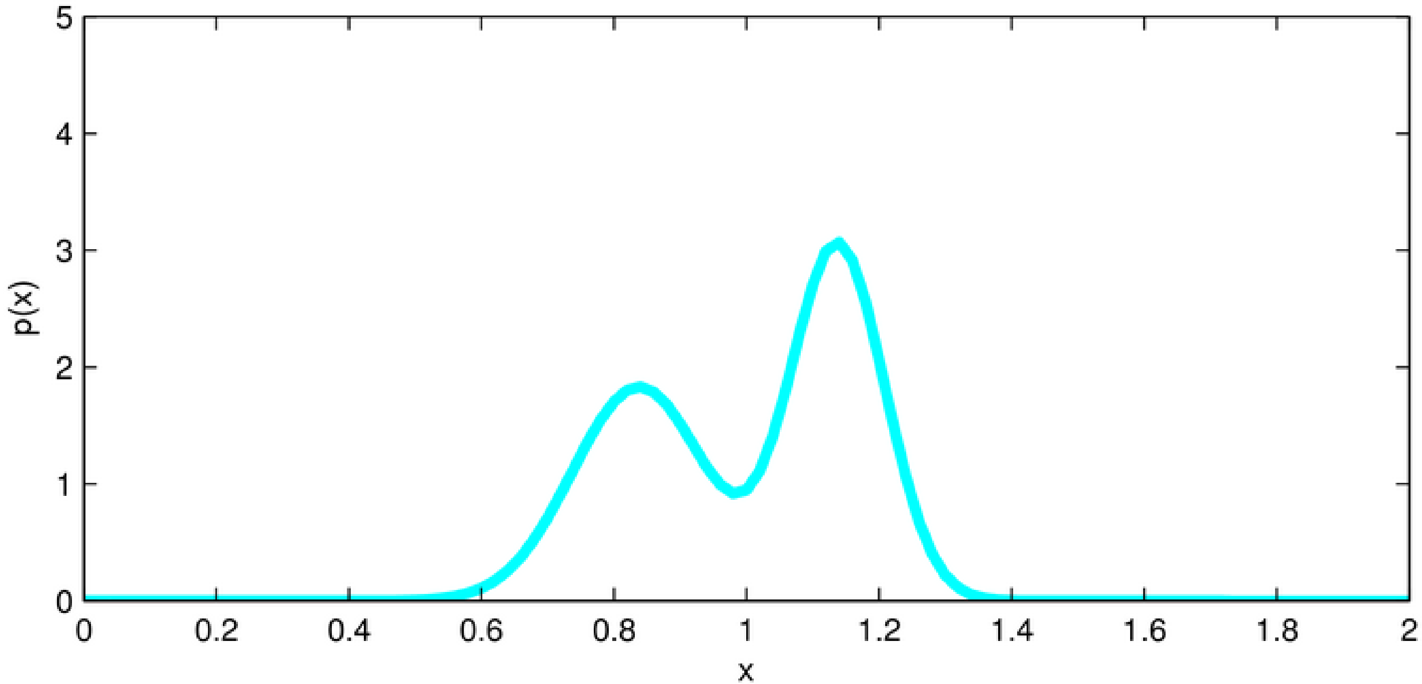
Simplified notation

- suppress reference to $\mathbf{y}_1,\dots,\mathbf{y}_{k-1}$ and to specific time t_k

  $$p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})/p(\mathbf{y})$$

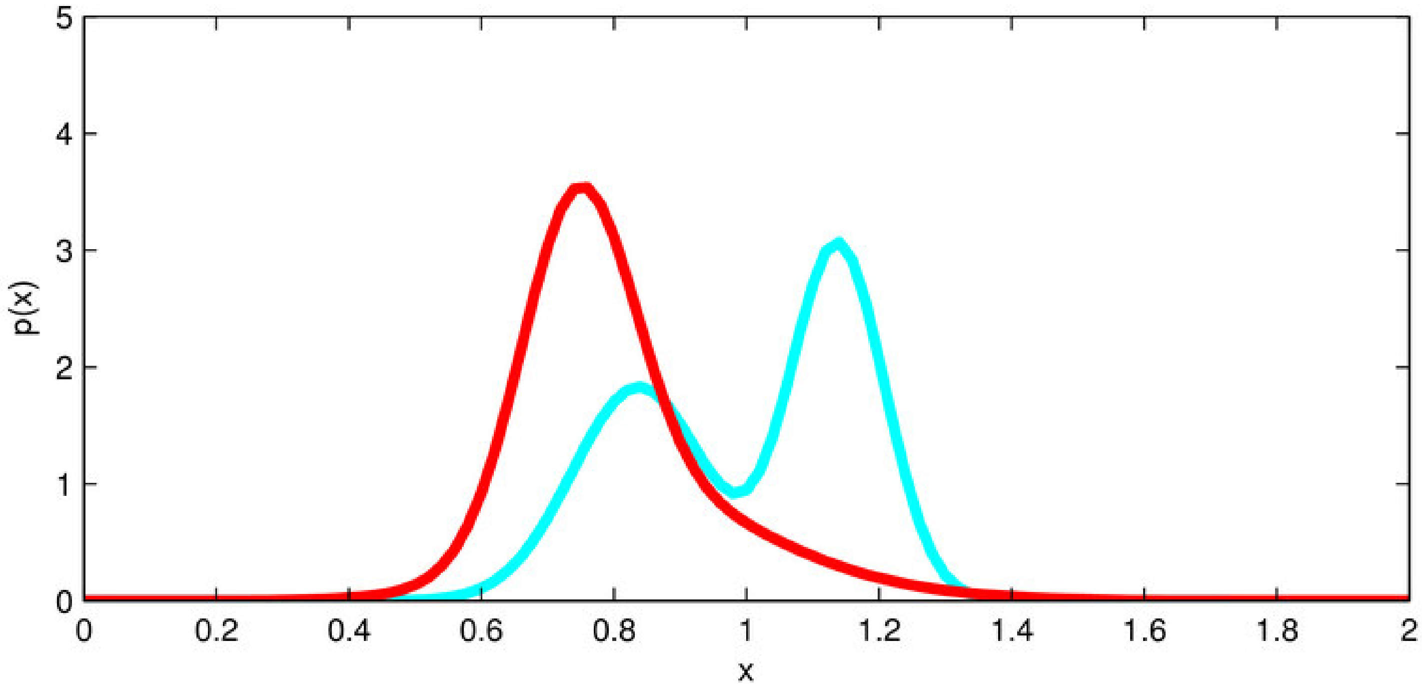
Bayesian Assimilation Illustrated

- forecast $p(x)$ for one-dimensional example: state x is a scalar



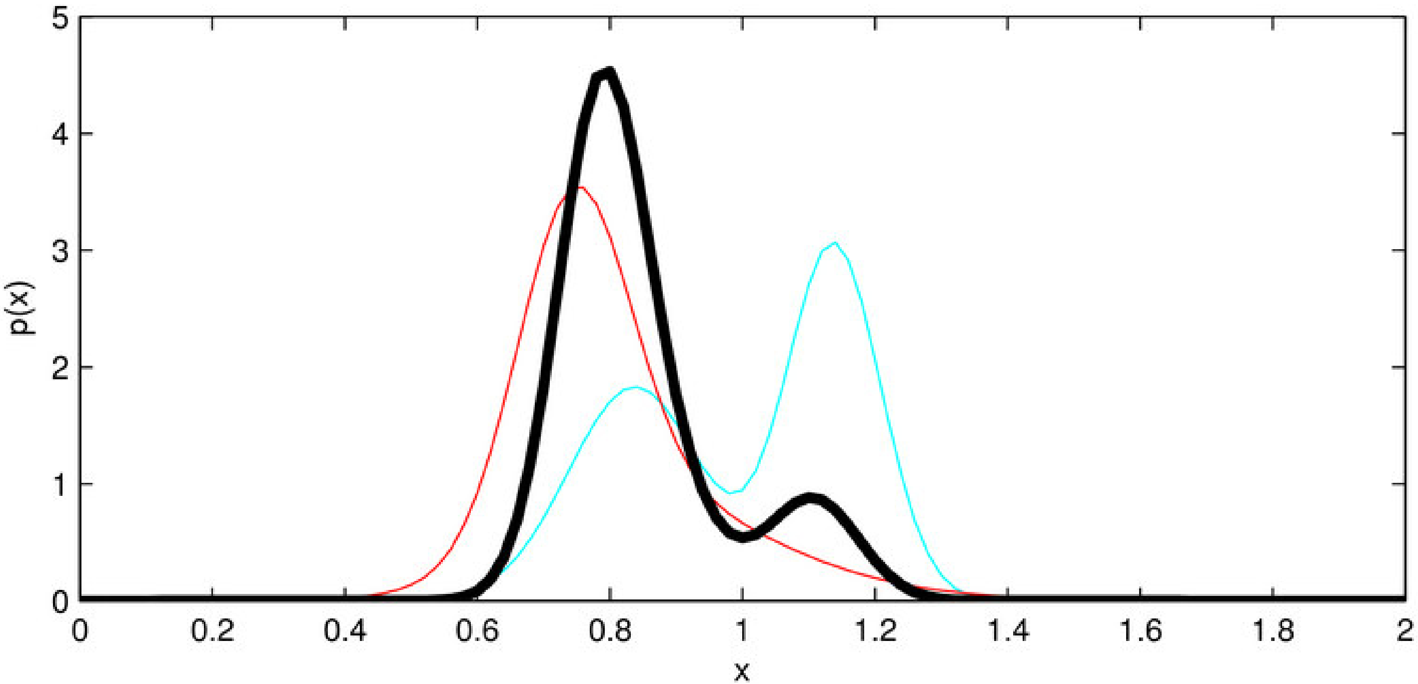
Bayesian Assimilation Illustrated

- observation likelihood $p(y|x)$ for $y = 0.8$



Bayesian Assimilation Illustrated

- analysis $p(x|y)$



Bayes Rule for Gaussians

"Linear, Gaussian case"

- linear observations with additive error, $\mathbf{y} = \mathbf{H}\mathbf{x} + \epsilon$
- prior/forecast $p(\mathbf{x})$ and pdf of ϵ are Gaussian

Consequences of linear, Gaussian case

- $\mathbf{y} = \mathbf{H}\mathbf{x} + \epsilon$ and $\epsilon \sim$ Gaussian $\Rightarrow p(\mathbf{y}|\mathbf{x})$ Gaussian
- analyis/posterior $p(\mathbf{x}|\mathbf{y})$ is product of Gaussians and so Gaussian too

Linear, Gaussian case (cont.)

Kalman filter = Bayes rule for linear, Gaussian case

- analysis equations:

$$\overline{\mathbf{x}}^a = (\mathbf{I} - \mathbf{KH})\overline{\mathbf{x}}^f + \mathbf{Ky}, \quad \mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f,$$

- Kalman gain

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{HP}^f \mathbf{H}^T + \mathbf{R})^{-1}$$

- notation: overbar indicates mean, superscript a (f) indicates analysis (forecast), $\mathbf{P}$ is state covariance, $\mathbf{R}$ is covariance of ϵ

Linear, Gaussian Case (cont.)

Kalman filter update

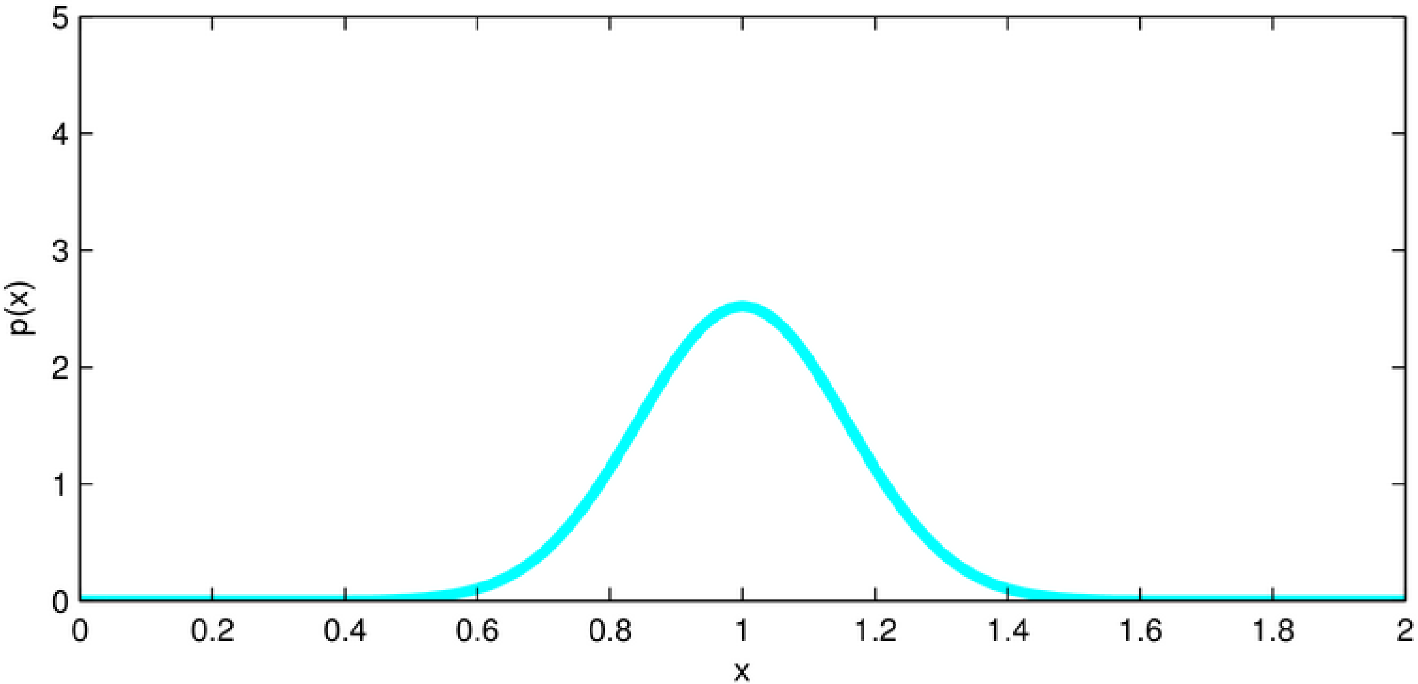
- $\overline{\mathbf{x}}^a = (\mathbf{I} - \mathbf{KH})\overline{\mathbf{x}}^f + \mathbf{Ky}, \quad \mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f$

Properties

- only need means and covariances:
  $\overline{\mathbf{x}}^f$ and $\mathbf{P}^f$ for prior, $\overline{\mathbf{x}}^a$ and $\mathbf{P}^a$ for posterior, $\mathbf{R}$ for ϵ
- $\overline{\mathbf{x}}^a$ depends linearly on $\overline{\mathbf{x}}^f$ and $\mathbf{y}$
- variance is smaller in analysis: since $\mathbf{KHP}^f$ is positive definite
  $$\mathbf{P}^a = \mathbf{P}^f - \mathbf{KHP}^f \Rightarrow \mathrm{tr}(\mathbf{P}^a) < \mathrm{tr}(\mathbf{P}^f)$$
- $\mathbf{P}^a$ does not depend on $\mathbf{y}$
- analysis is sensitive to outliers

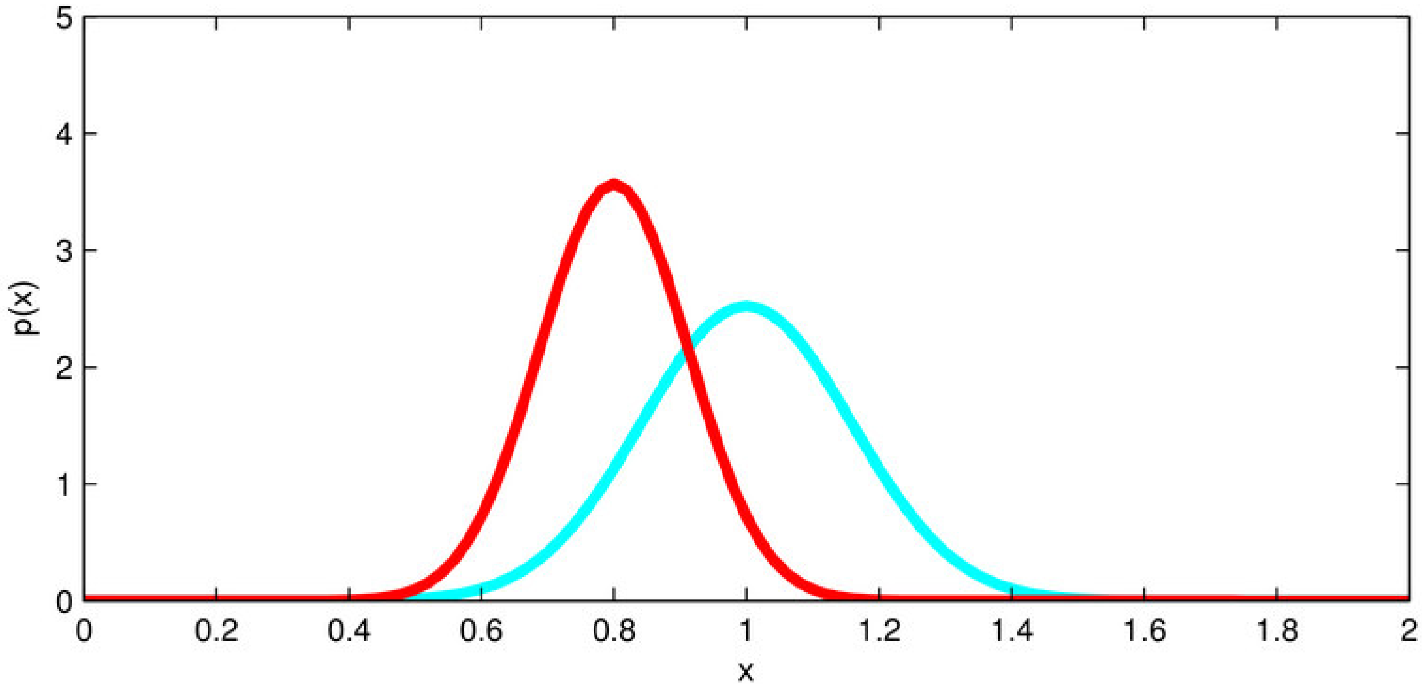
Gaussians Illustrated

▷ prior $p(x)$



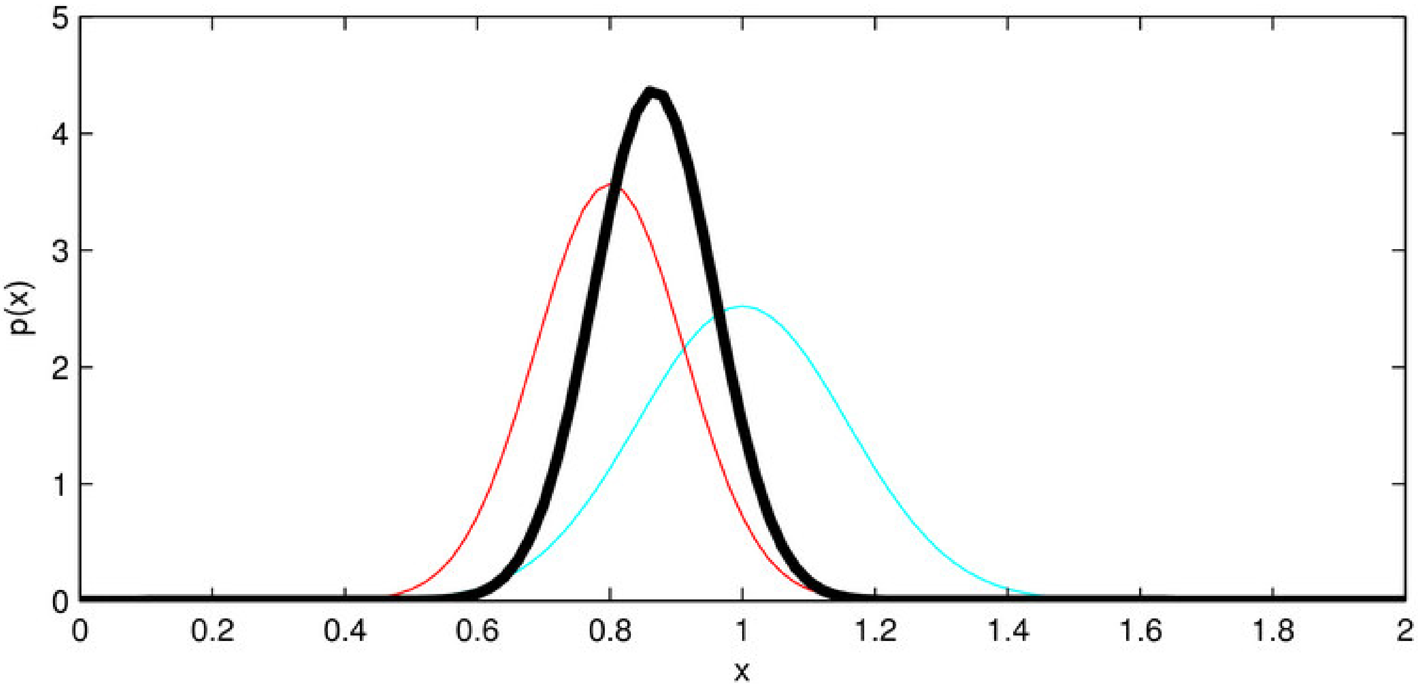
Gaussians Illustrated

▷ observations likelihood $p(y|x)$



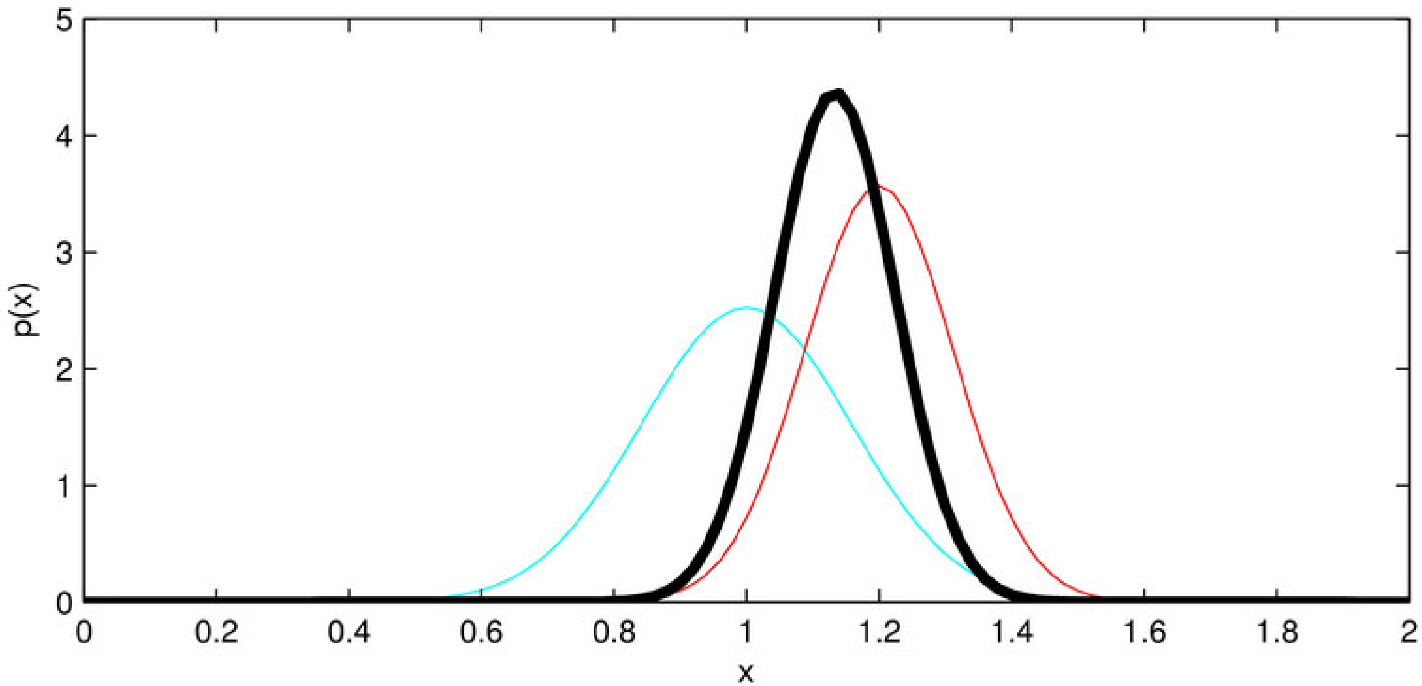
Gaussians Illustrated

- posterior $p(x|y)$



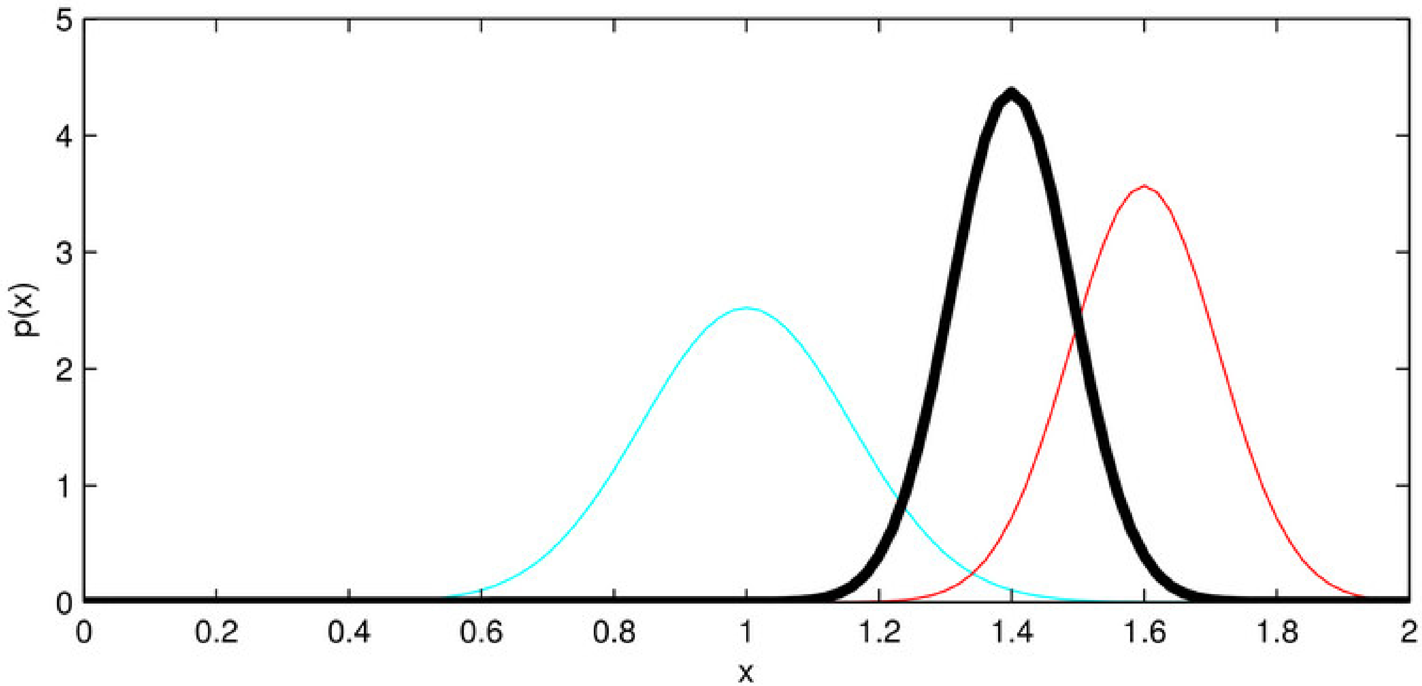
Gaussians Illustrated

- analysis variance is independent of y: $y = 1.2$



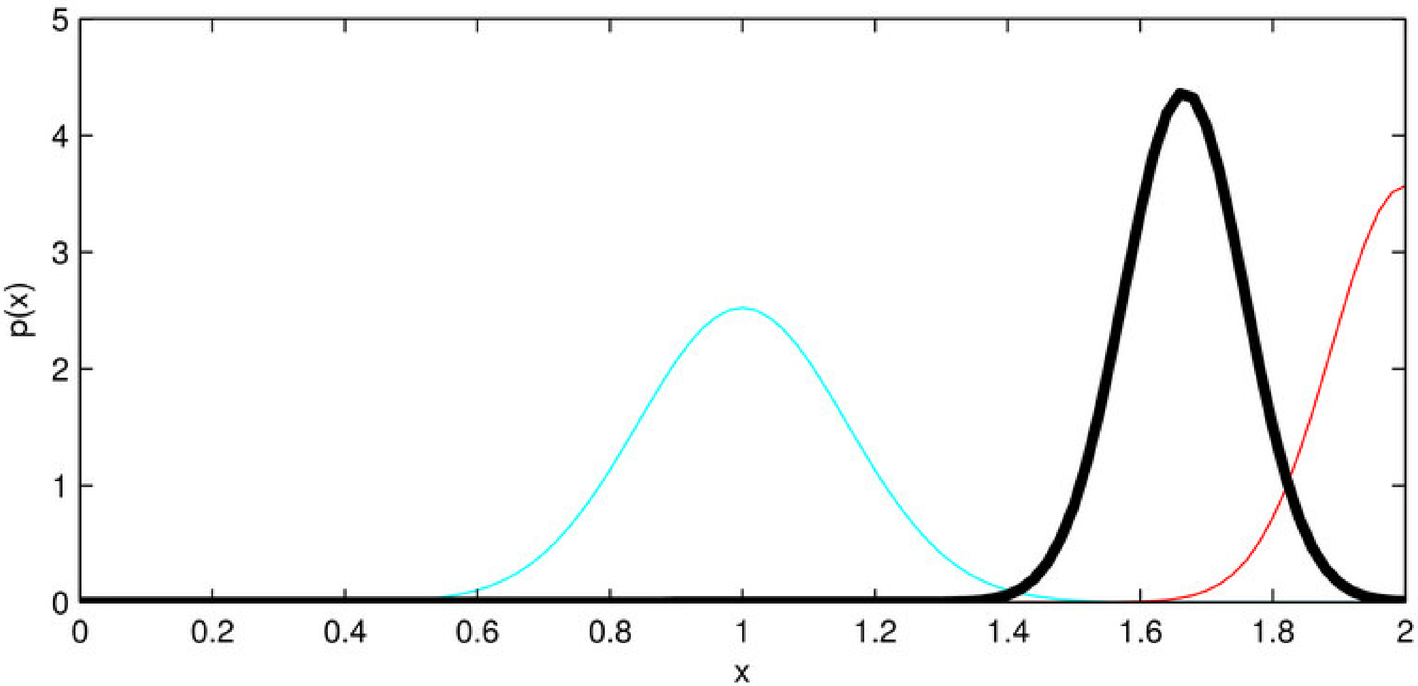
Gaussians Illustrated

- analysis variance is independent of y: $y = 1.6$



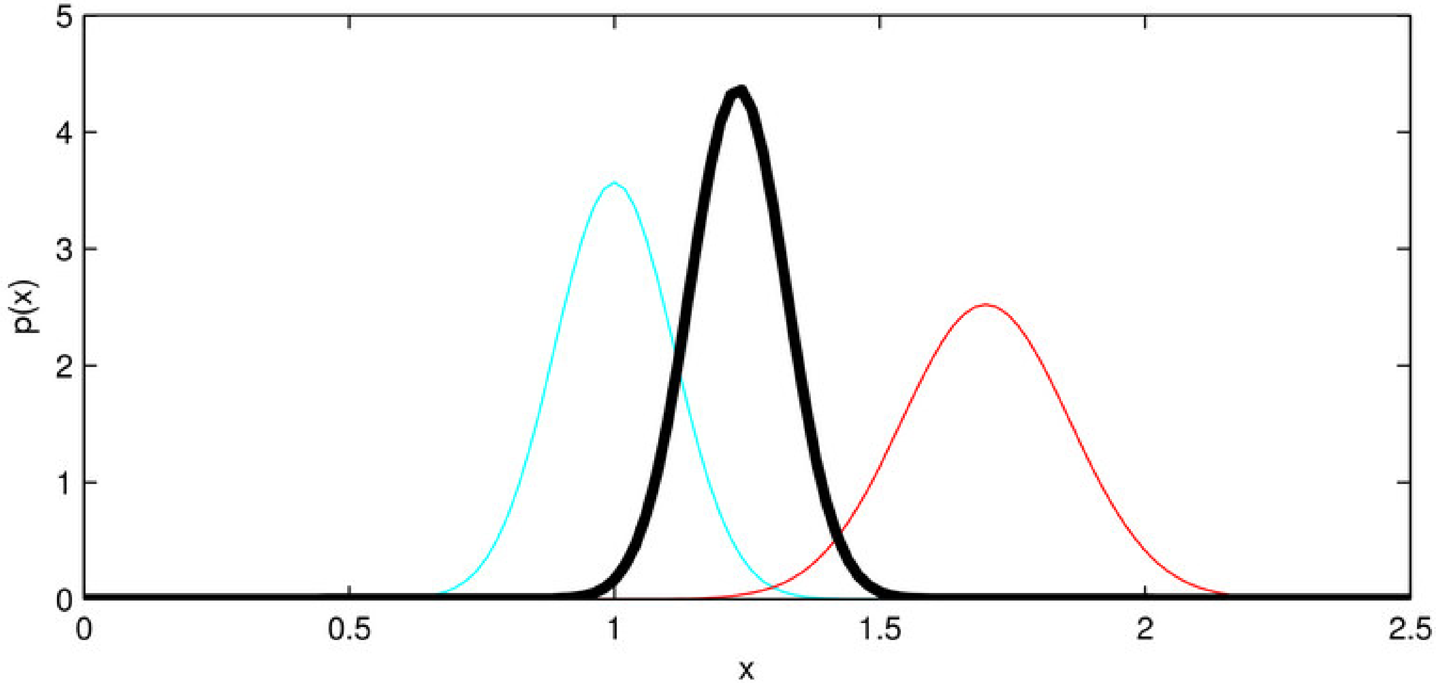
Gaussians Illustrated

- analysis variance is independent of y: $y = 2.0$



Gaussians and Outliers

- forecast mean and observation differ by "6σ:" $\bar{x}^f = 1$, $y = 1.7$



- Analysis mean has *very* low probability under both prior and likelihood
- If observation errors are assumed Gaussian but in fact are not (e.g. occasional large errors), then analysis will be strongly degraded

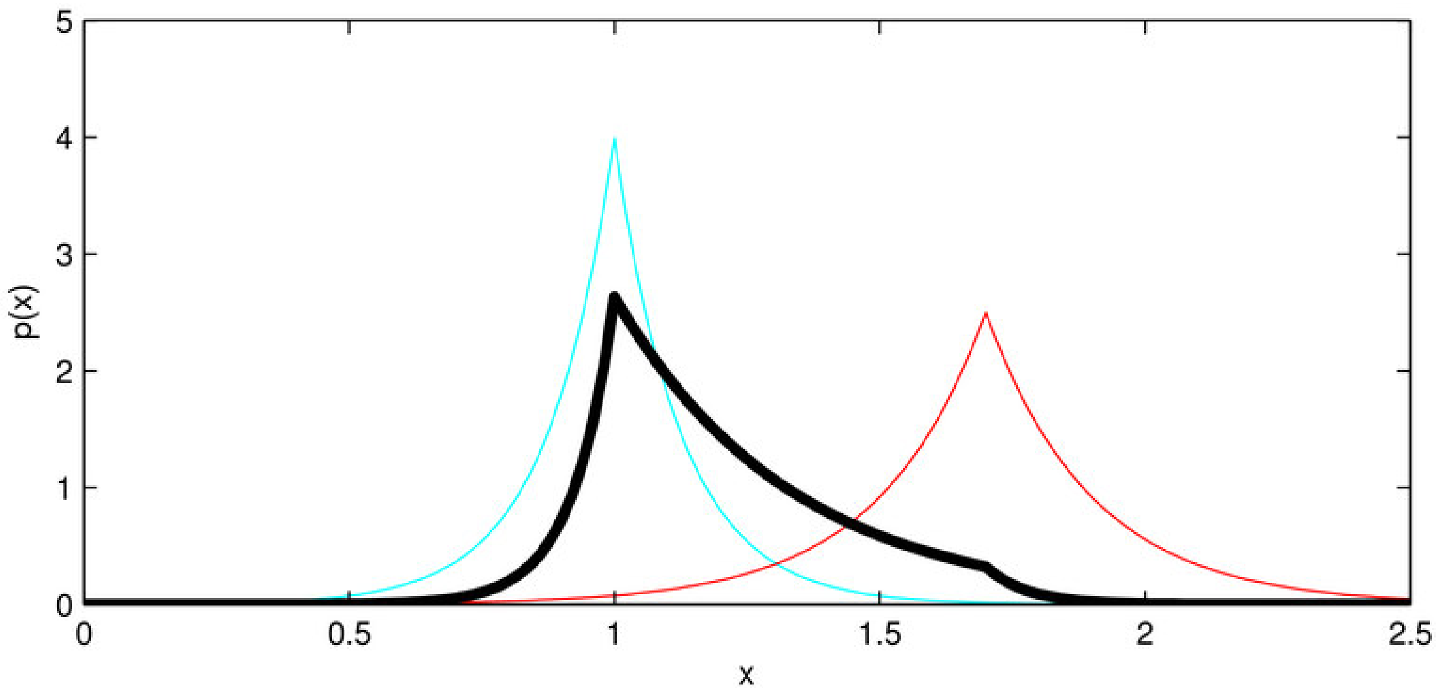
Non-Gaussian Effects

Results for general pdfs may be qualitatively different

- e.g., differences between mean and mode (most likely state)
- analysis mean depends nonlinearly on observations
- analysis variance depends on value of observations
- analysis variance can be larger than that of forecast
- pdfs with longer tails are less sensitive to outliers

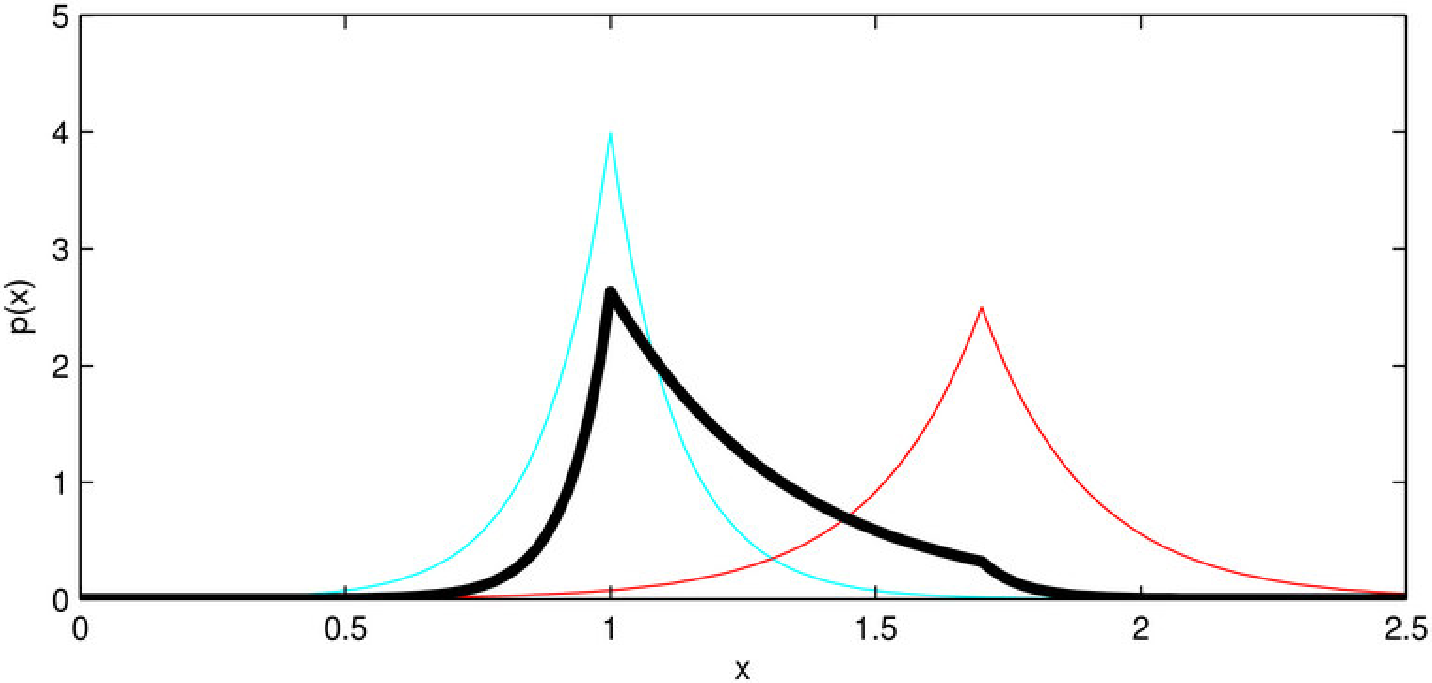
Non-Gaussian Effects (cont.)

- suppose $p(x)$ and $p(y|x)$ are exponential pdfs
- analysis variance depends on y: $\mathrm{var}(x|y) = 0.23^2$ for $y = 1.7$



Non-Gaussian Effects (cont.)

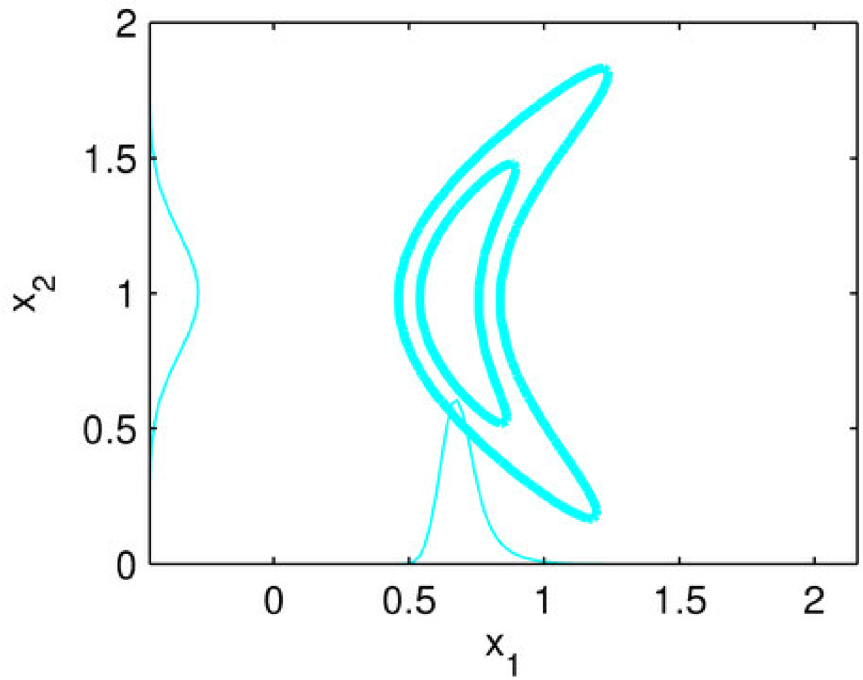
- suppose $p(x)$ and $p(y|x)$ are exponential pdfs
- analysis variance depends on y: $\mathrm{var}(x|y) = 0.23^2$ for $y = 1.7$



- analysis variance larger than forecast variance (0.18^2)
- analysis pdf is close to forecast pdf, despite outlying observation

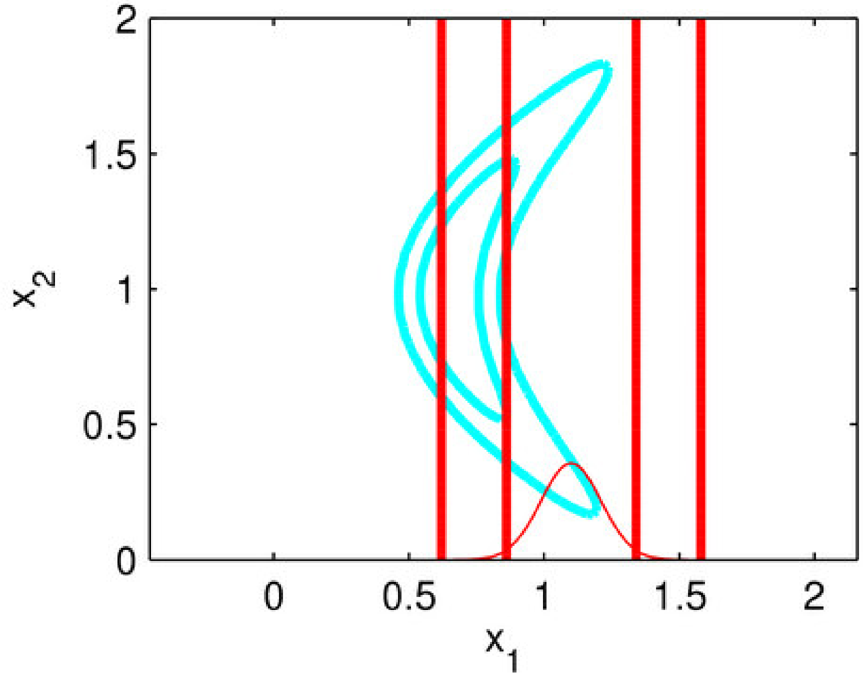
Non-Gaussian Effects (cont.)

▷ $p(x_1, x_2)$ for 2D state (x_1, x_2); thin lines indicate marginal pdfs



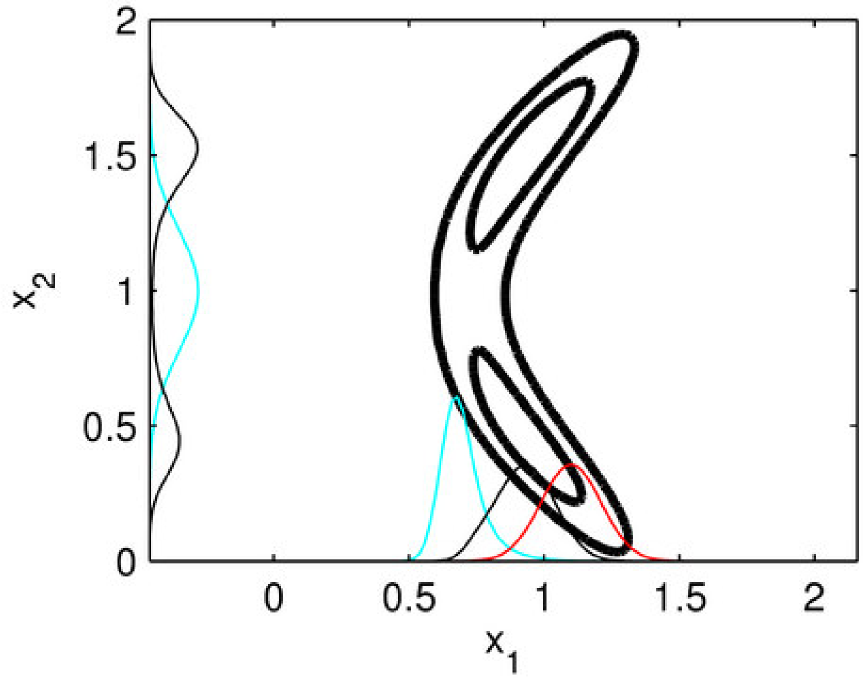
Non-Gaussian Effects (cont.)

- observation $y = x_1 + \epsilon = 1.1$
- $p(y|x_1, x_2)$ does not depend on x_2



Non-Gaussian Effects (cont.)

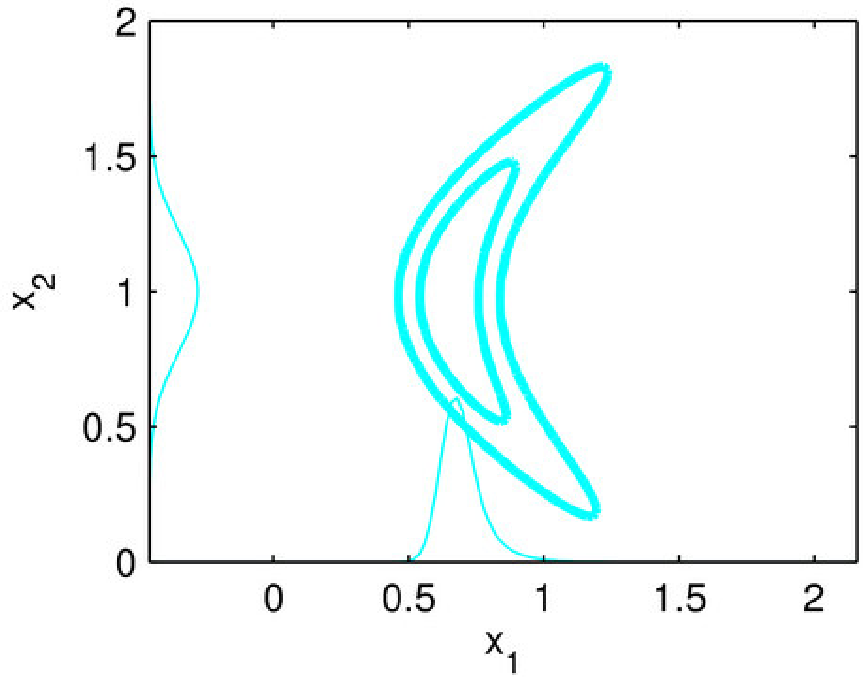
- $p(x_1, x_2|y)$
- marginal variances increase, marginal for x_2 becomes bimodal



Dealing with non-Gaussianity

Direct calculation of Bayesian update

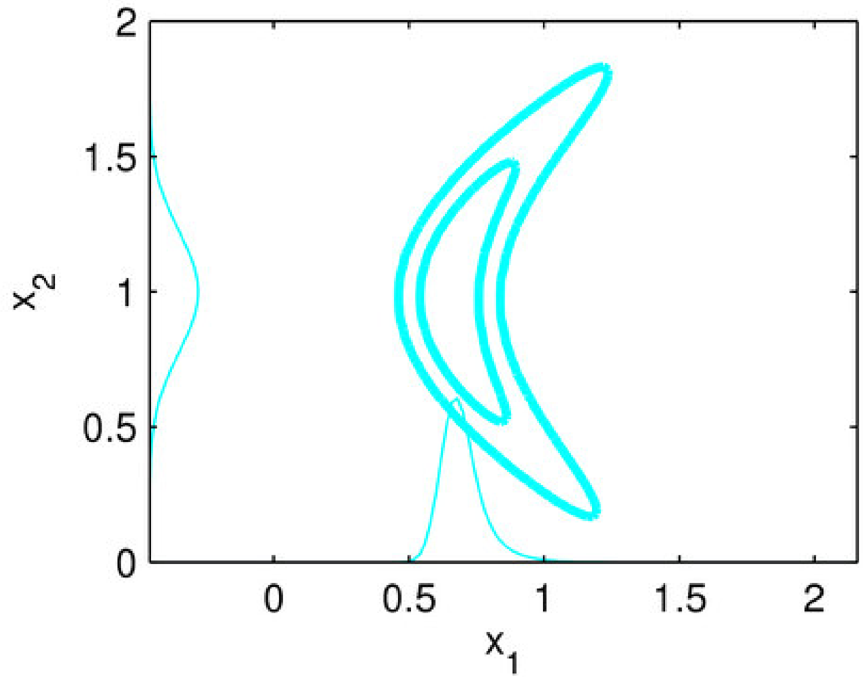
- in principle, could represent required pdfs on discrete grid, then perform multiplication directly
- no approximations, other than those required in specifying observation operators and errors and in evolving $p(\mathbf{x})$ from analysis to forecast times.



Dealing with non-Gaussianity

Direct calculation is difficult when dimension is large

- recall that $p(\mathbf{x})$ is a function in $N_x = \dim(\mathbf{x})$ variables
- thus, gridded representation of $p(\mathbf{x})$ requires number of grid points that scales as $\exp(N_x)$ ... computationally intractable
- e.g. if $\dim(\mathbf{x}) = 100$ and we allow 10 grid points for each of the variables $x_1, \ldots, x_{100}$, then we need 10^{100} points (!)



Dealing with non-Gaussianity (cont.)

Maximum likelihood estimation

- calculate the posterior mode, $\mathbf{x}$ s.t. $p(\mathbf{x}|\mathbf{y})$ is maximum, rather than entire posterior pdf
- equivalently, minimize $-\log\left(p(\mathbf{x}|\mathbf{y})\right)$ ... as in 4DVar
- does not provide $p(\mathbf{x}|\mathbf{y})$; also requires models for $p(\mathbf{x})$

Dealing with non-Gaussianity (cont.)

Particle filter (PF)

- Monte-Carlo approach: start from ensemble $\{\mathbf{x}_i^f, i = 1, \ldots, N_e\}$ that is assumed to be random draw from $p(\mathbf{x})$
- approximate prior pdf as sum of point masses,

$$p(\mathbf{x}) \approx N_e^{-1} \sum_{i=1}^{N_e} \delta(\mathbf{x} - \mathbf{x}_i^f)$$

- Bayes $\Rightarrow$

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}) \sum_{i=1}^{N_e} \delta(\mathbf{x} - \mathbf{x}_i^f) = \sum_{i=1}^{N_e} p(\mathbf{y}|\mathbf{x}_i^f)\delta(\mathbf{x} - \mathbf{x}_i^f)$$

- thus, posterior pdf approximated by weighted sum of point masses

$$p(\mathbf{x}|\mathbf{y}) \approx \sum_{i=1}^{N_e} w_i \delta(\mathbf{x} - \mathbf{x}_i^f), \quad \text{with} \quad w_i = \frac{p(\mathbf{y}|\mathbf{x}_i^f)}{\sum_{j=1}^{N_e} p(\mathbf{y}|\mathbf{x}_j^f)}$$

Dealing with non-Gaussianity (cont.)

Asymptotically convergent to Bayes rule

- PF yields an exact implementation of Bayes' rule as $N_e \to \infty$; no approximations other than finite ensemble size

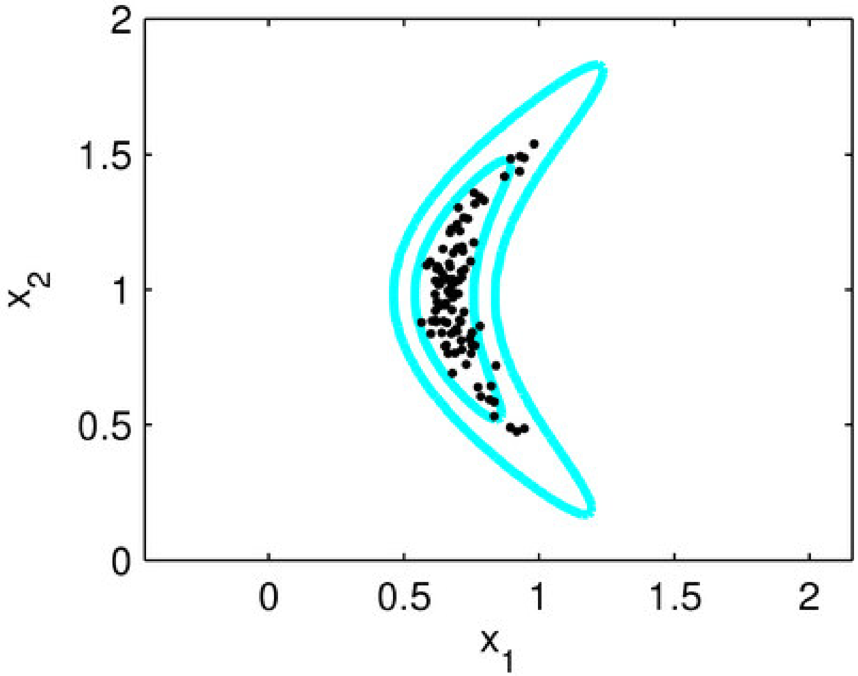
Exceedingly simple

- main calculation is $p(\mathbf{y}|\mathbf{x}_i^f)$ for $i = 1, \ldots, N_e$.

Widely applied, and effective, in low-dim'l systems

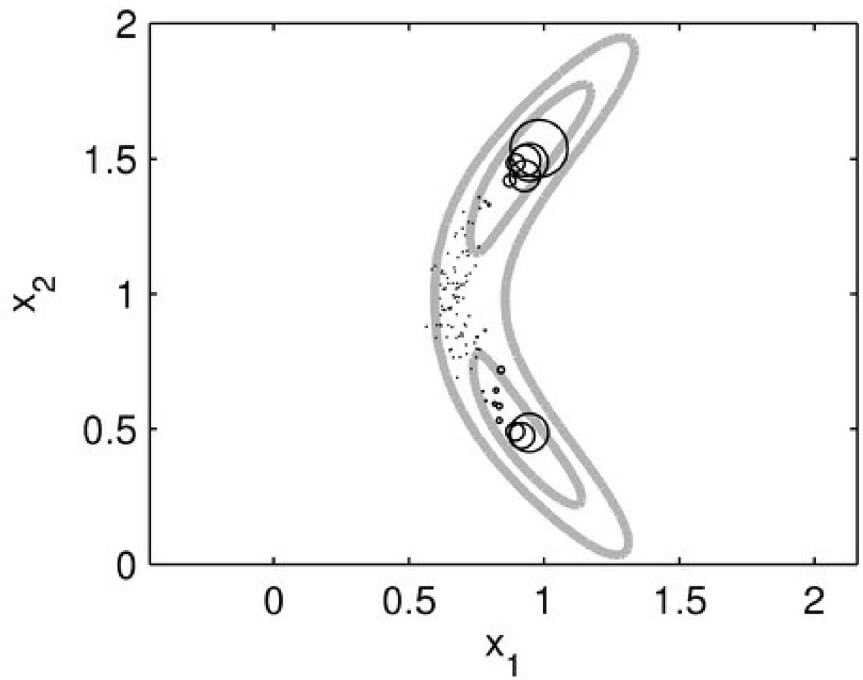
PF Illustrated

- $p(\mathbf{x})$, as before, and prior ensemble



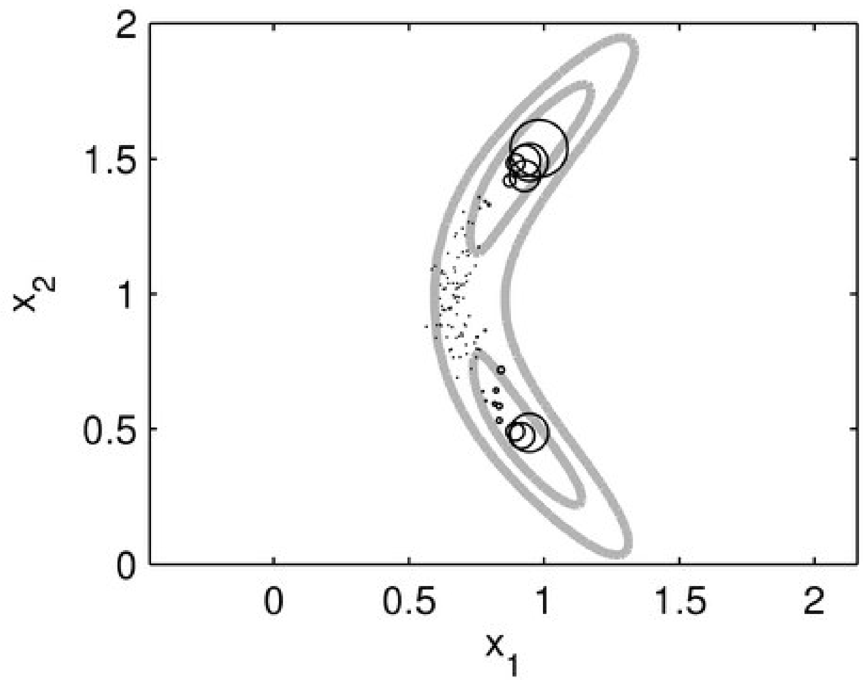
PF Illustrated

▷ $p(\mathbf{x}|\mathbf{y})$ and "weighted" ensemble (size $\propto$ weight)



PF Illustrated

- $p(\mathbf{x}|\mathbf{y})$ and "weighted" ensemble (size $\propto$ weight)



- weighted ensemble captures bimodality

Refinements of PF

Many members recieve very small weights

- resampling: need to "refresh" ensemble; members with small weights are dropped, while additional members are added near members with large weights
- importance sampling: draw original ensemble from another distribution that incorporates additional information, for example from latest observations

Problems arise for high-dimensional systems

- strong tendency for $\max w_i \to 1$

Dealing with non-Gaussianity (cont.)

Best linear unbiased estimator (BLUE)

- ask for the linear estimator (analysis) that has minimum expected squared error
- to fix ideas, consider the scalar case, but can generalize to multivariate
- given: $y = x + \epsilon$ and a prior or forecast estimate $\hat{x}^f = x + \epsilon^f$

BLUE

Linear estimator

- estimator $\hat{x}$ depends linearly on y and $\hat{x}^f$,

$$\hat{x} = ay + b\hat{x}^f$$

Unbiased

- want $E(\hat{x} - x) = 0$ if $E(\epsilon) = E(\epsilon^f) = 0$
- since $\hat{x} - x = (a + b - 1)x + a\epsilon + b\epsilon^f$, must have

$$a + b = 1$$

- note $\hat{x}^f$ must be the prior mean of x if $E(\epsilon^f) = 0$

BLUE (cont.)

"Best" = minimum expected squared error

- expected squared error of $\hat{x}$ given by

$$E\left((\hat{x}-x)^2\right) = a^2\sigma_o^2 + (1-a)^2\sigma_f^2 + 2E(\epsilon\epsilon^f)$$

- take $E(\epsilon\epsilon^f) = 0$ for simplicity; minimizing w.r.t. a gives

$$a = \sigma_f^2/(\sigma_o^2+\sigma_f^2), \quad b = \sigma_o^2/(\sigma_o^2+\sigma_f^2)$$

- back substitution yields $E\left((\hat{x}-x)^2\right)$

Estimator involves only mean and covariances

- equivalent to Kalman filter in linear, Gaussian case
- *but*, no assumption of Gaussianity of ϵ and ϵ^f; BLUE properties hold for arbitrary pdfs

Bayesian View of the BLUE

Begin with $p(\mathbf{x})$ and observations $\mathbf{y}$

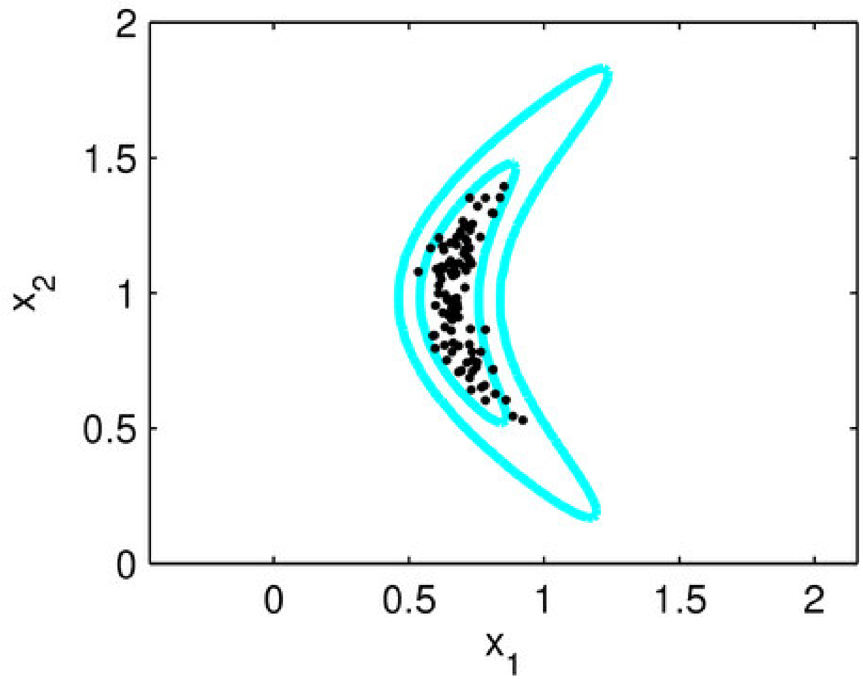
BLUE defines linear (affine) transformation of $\mathbf{x}$

- i.e., a new random variable $\mathbf{x}^a = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}$
- $\mathbf{x}^a$ has known mean and covariance matrix given by BLUE formulas
- $\mathbf{x}^a$ need not be Gaussian
- in linear, Gaussian case, $\mathbf{x}^a$ has pdf $p(\mathbf{x}|\mathbf{y})$

EnKF is Monte-Carlo implementation of BLUE in joint state-obs space

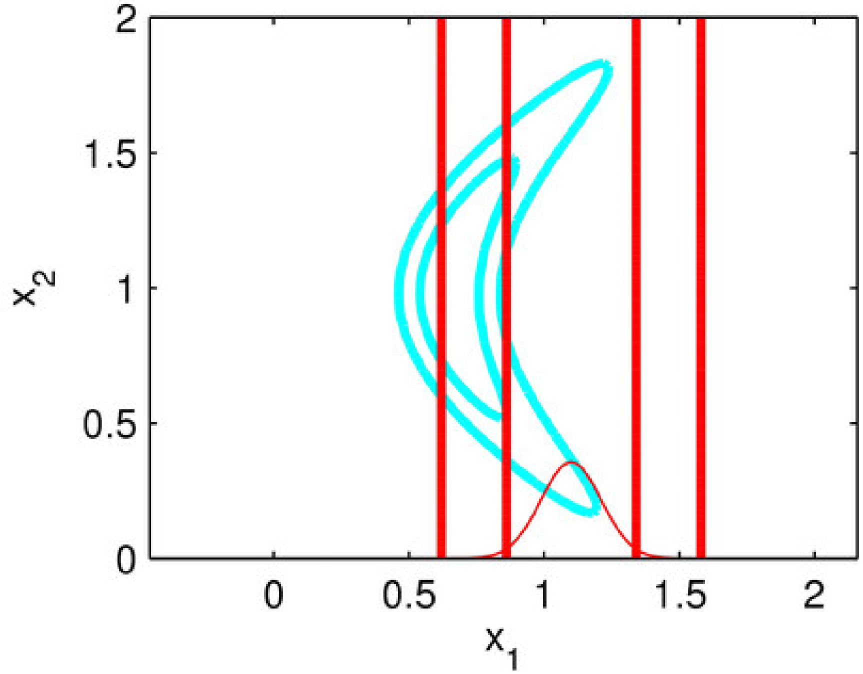
BLUE/EnKF Illustrated

- ▷ $p(\mathbf{x})$ and ensemble



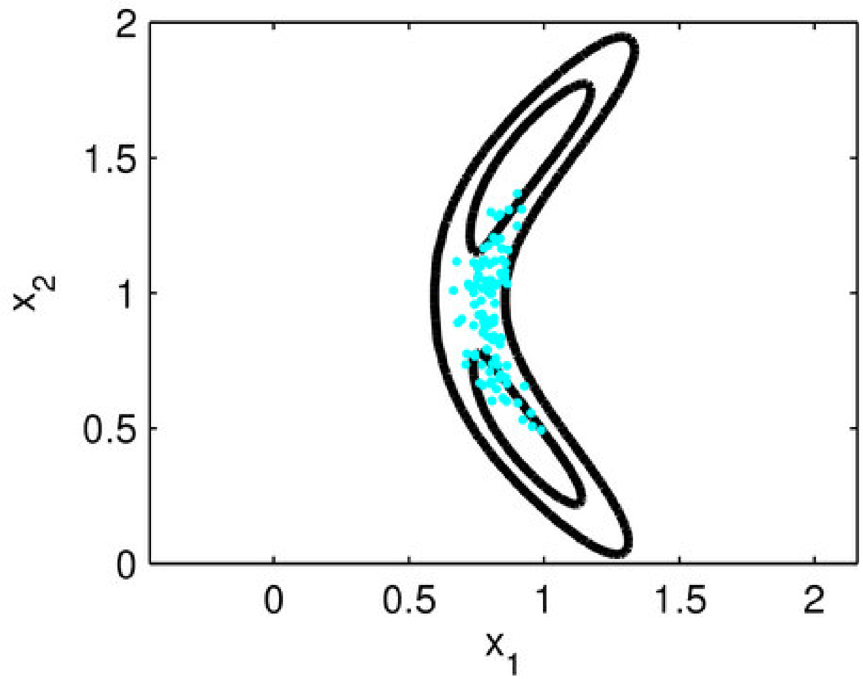
BLUE/EnKF Illustrated

- $p(y|\mathbf{x})$ for $y = 1.1$



BLUE/EnKF Illustrated

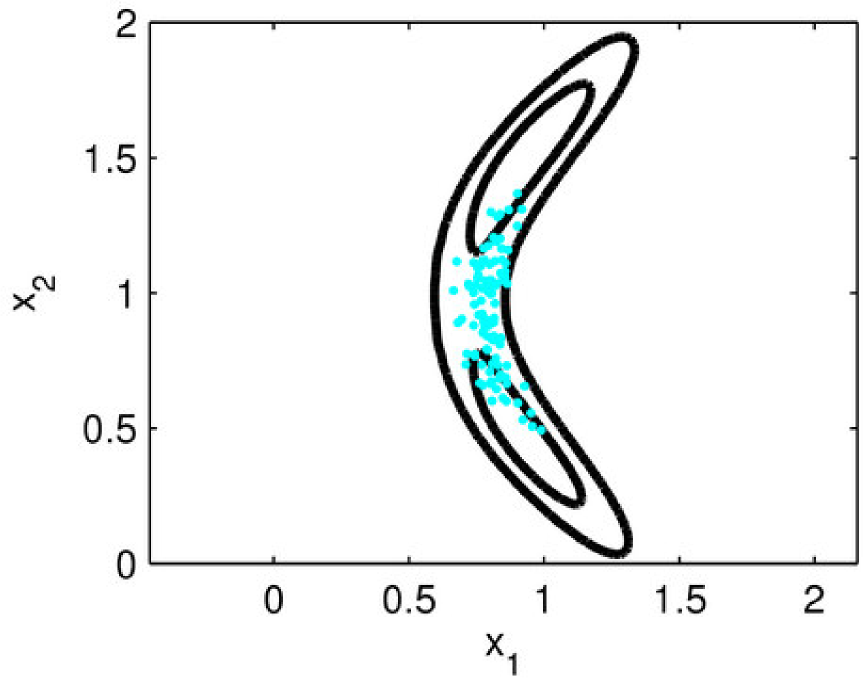
- $p(\mathbf{x}|y)$ from Bayes rule and analysis ensemble from BLUE/EnKF



- sample retains non-Gaussian curvature but does not capture bimodality

BLUE/EnKF Illustrated

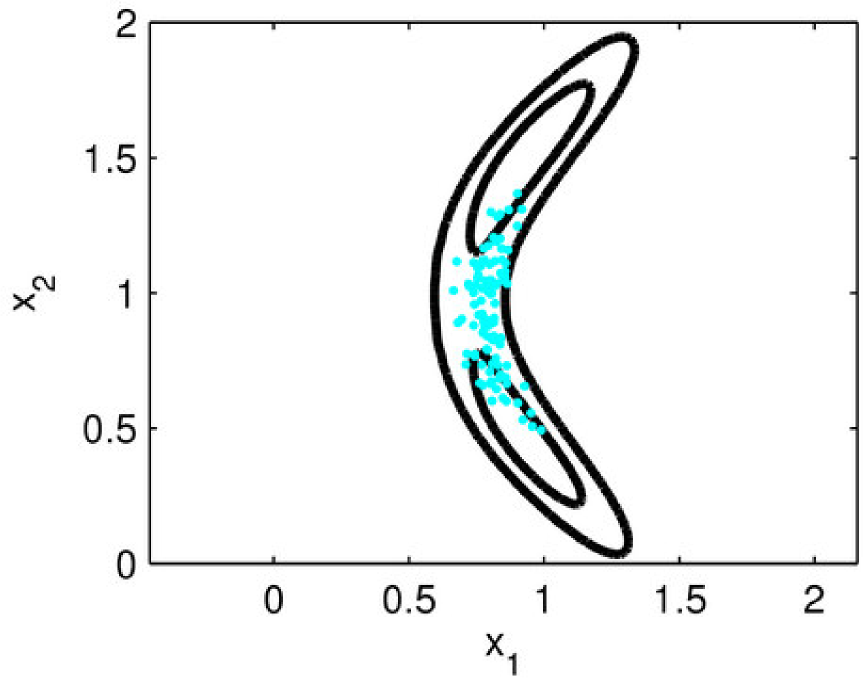
- $p(\mathbf{x}|y)$ from Bayes rule and analysis ensemble from BLUE/EnKF



- sample retains non-Gaussian curvature but does not capture bimodality

BLUE/EnKF Illustrated

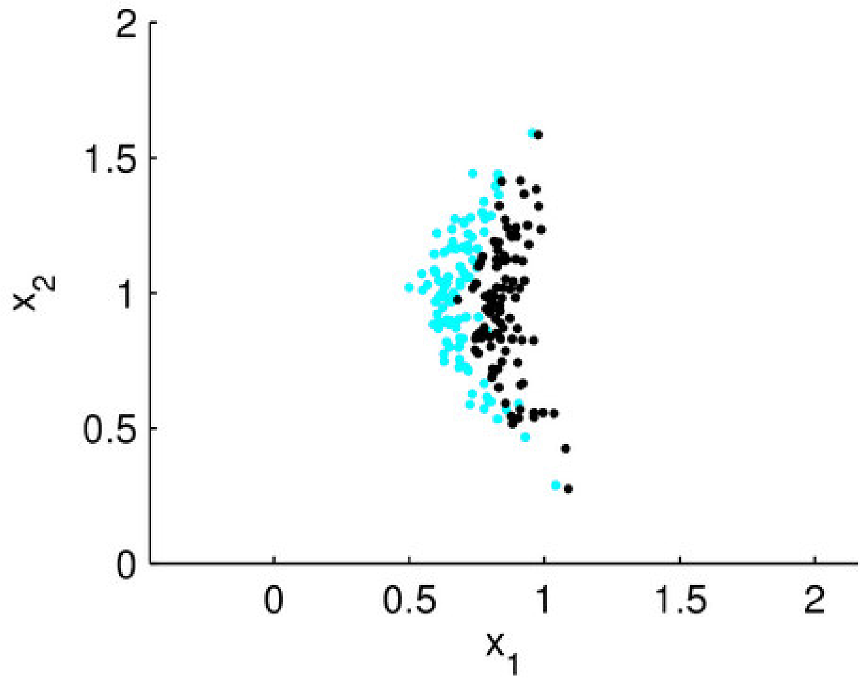
- $p(\mathbf{x}|y)$ from Bayes rule and analysis ensemble from BLUE/EnKF



- sample retains non-Gaussian curvature but does not capture bimodality

BLUE/EnKF Illustrated

- prior (blue) and analysis (black) ensembles from BLUE/EnKF



- transformation by BLUE shifts ensemble toward observation; little contraction of variance in analysis ensemble in this case

Closing Thoughts

General treatment of non-Gaussian effects is hard

- direct calculations are overwhelmingly expensive
- particle filters also problematic, except for low-dimensional systems

Linear or approximately Gaussian approaches often work well

- other issues, such as model error and flow-dependence of covariances, more important?

Non-Gaussian effects significant in some applications

- tailored treatments based on specific assumptions about form of non-Gaussianity
- e.g., variational quality control