Intro to Data Assimilation (DA)

Chris Snyder

DA from 40K ft

Many atmospheric observations, from diverse instruments

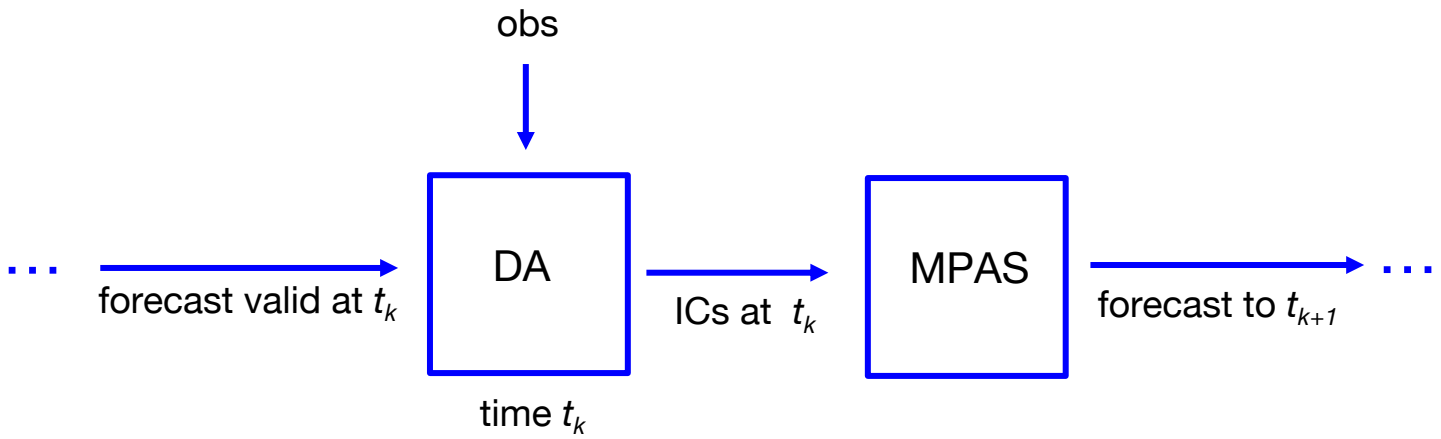
Large, complex numerical weather prediction (NWP) models

Want initial conditions for NWP models, based on available observations

Cycling DA

Short-range NWP is quite accurate

Both latest observations **and** latest forecast provide useful information



Scalar DA Example

Given: an observation and a forecast of some scalar, say *T* in this room.
How to combine these into an improved estimate of *T*?

Average, with weights chosen to minimize the expected squared error

$$T^a = kT^o + (1-k)T^b, \text{ with } k \text{ chosen to minimize } E((T^a - T)^2)$$

$$k = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}, \text{ where } \sigma_o^2 = E((T^o - T)^2), \quad \sigma_b^2 = E((T^b - T)^2)$$

"o" = observed, "b" = background (the forecast), "a" = analysis (the improved estimate)

Beyond the Scalar Example

NWP is trickier – as are other real applications

- Forecast models have many degrees of freedom (e.g., number of points X number of variables ~ 10^8 - 10^9)
- Many observations, and of diverse quantities (number of distinct observations ~ 10^7 - 10^8)
- Observed quantities differ from model quantities

DA Notation (don't sleep through this!)

x = "state"

Quantities to be estimated, represented in discrete basis and then concatenated into a column vector.
For MPAS-JEDI, consists of u v T q p_s (+ hydrometeors) at cell centers of MPAS mesh.

y = observations

All observations, concatenated into a column vector.

$\mathbf{y} = H(\mathbf{x}) + \boldsymbol{\varepsilon}$

Assumed relation of state and observations:
$H(.)$ = known observation or forward operator; predicts observations given state
$\boldsymbol{\varepsilon}$ = unknown, random observation error. Sum of all effects preventing perfect prediction of obs when state is known: instrument noise, influence of scales not included (represented) in state, imperfections in H.

A Bayesian Perspective

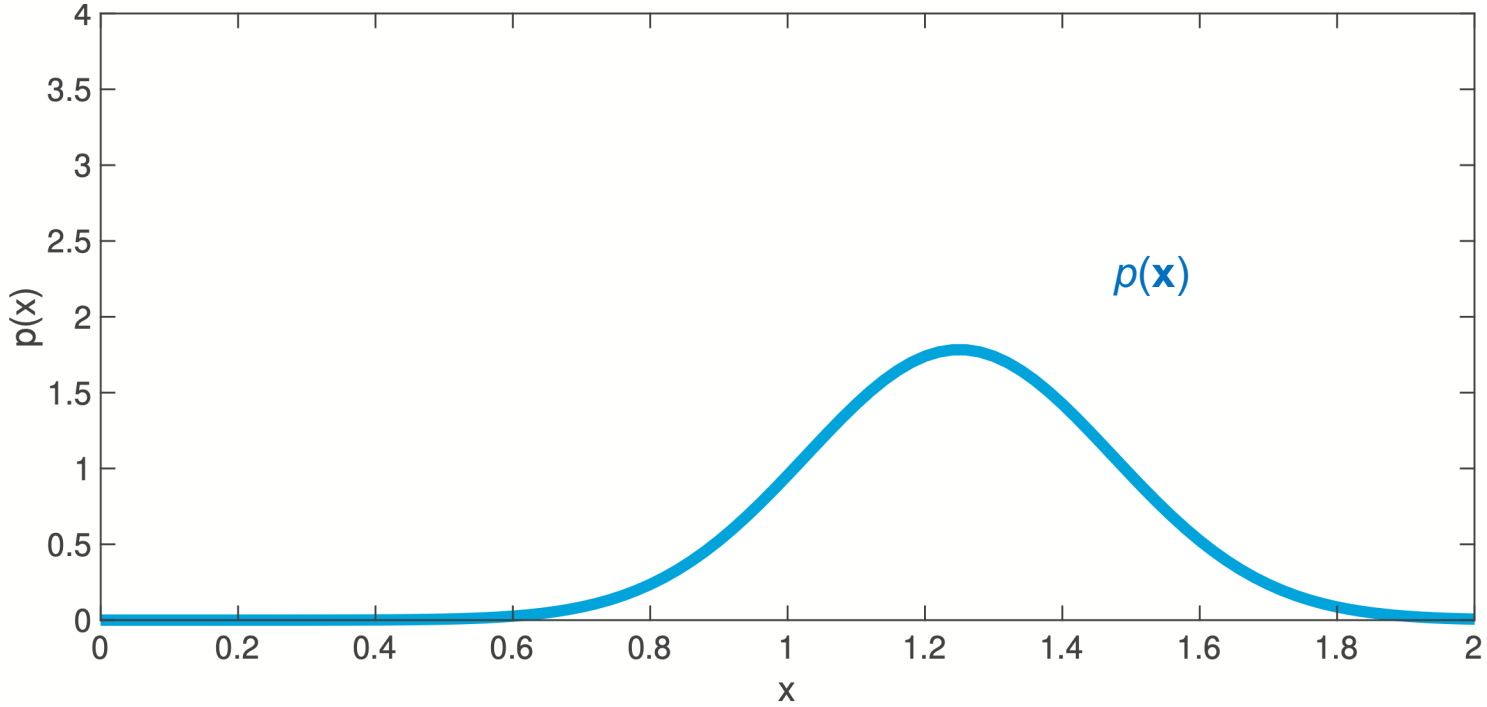
One view: seek “point” estimate $\mathbf{x}^a$ that minimizes expected squared error

Bayesian view: explicitly estimate probabilities of all outcomes

$$p(\mathbf{x}|\mathbf{y}^o) \propto p(\mathbf{x})p(\mathbf{y}^o|\mathbf{x})$$

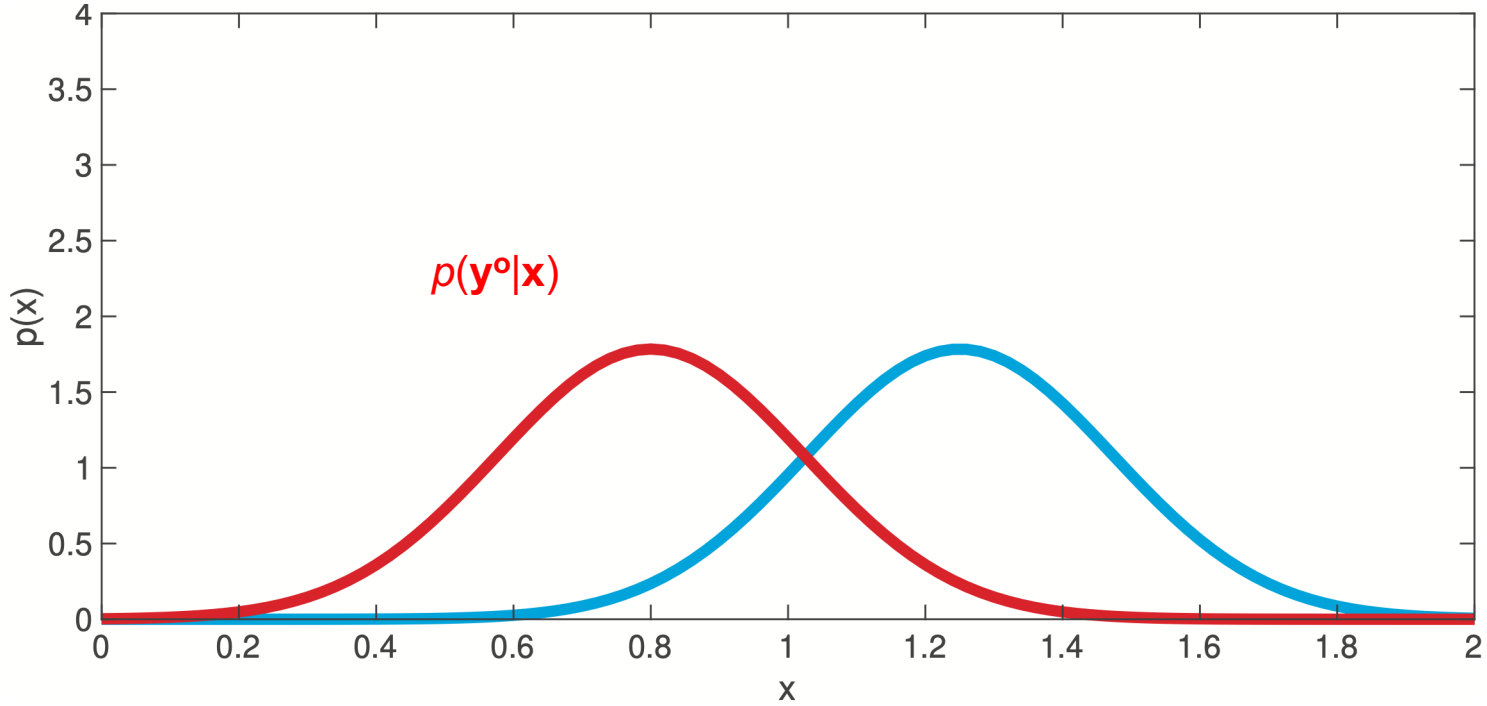
Bayes rule gives the conditional (or “posterior”) pdf in terms of prior pdf and the observation likelihood

Bayes Rule Illustrated



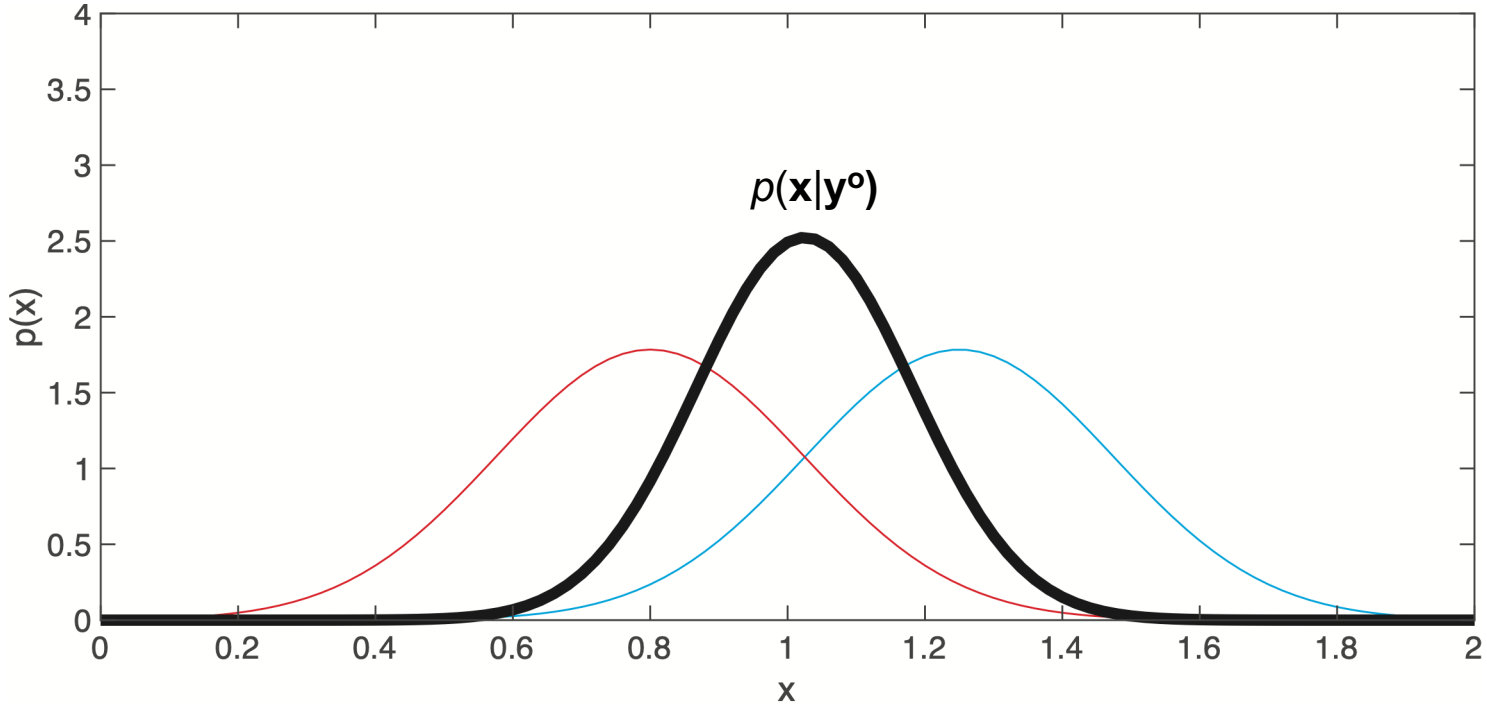
In this example, x has a Gaussian distribution with mean 1.25

Bayes Rule Illustrated

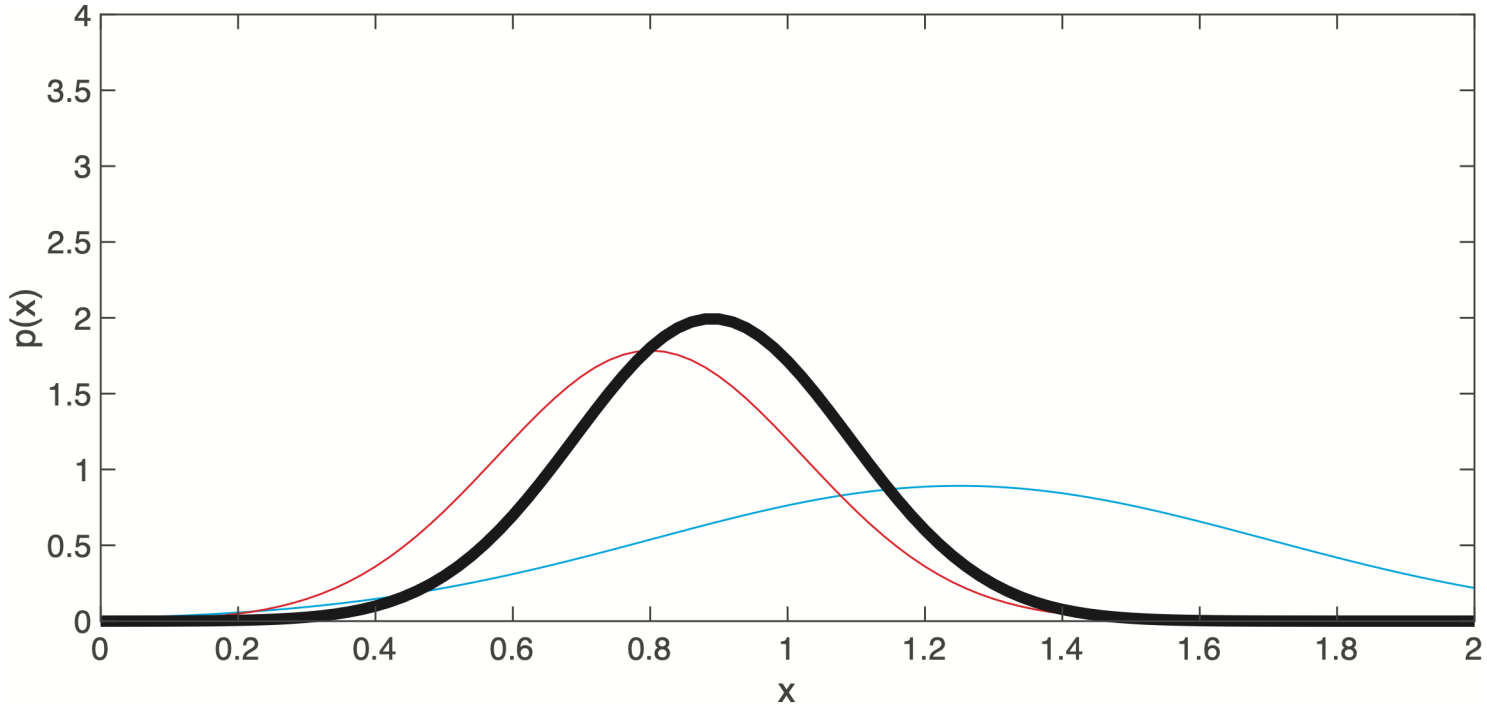


In this example, $y^o = 0.8$ and ε is Gaussian with mean zero

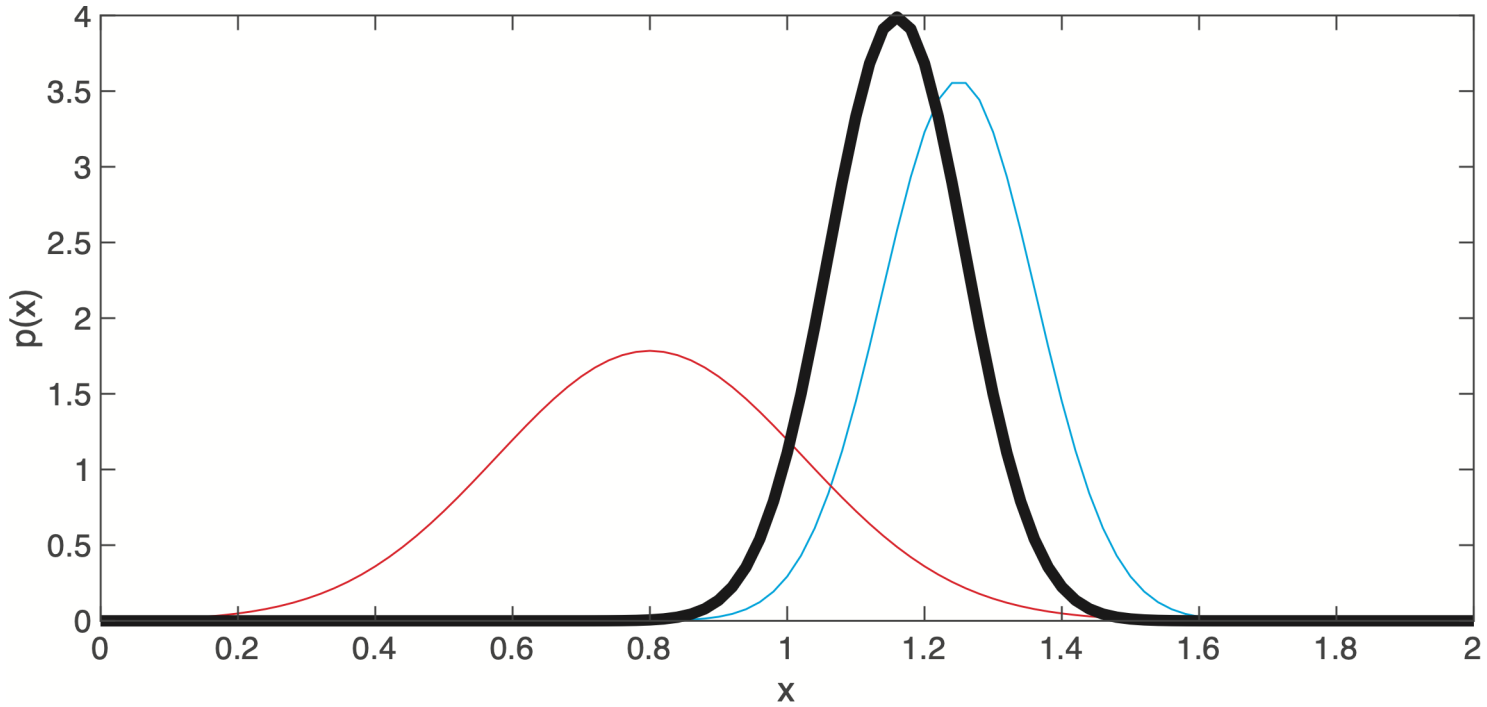
Bayes Rule Illustrated



Bayes Rule Illustrated



Bayes Rule Illustrated



Beyond the Scalar Example

Let's suppose that

- $\mathbf{x}$ is Gaussian, with mean $\mathbf{x}^b$ and covariance $\mathbf{B}$
- $\boldsymbol{\varepsilon}$ is Gaussian, with mean $\mathbf{0}$ and covariance $\mathbf{R}$
- H is linear, i.e. $H(\mathbf{x}) = \mathbf{H}\mathbf{x}$

"$\mathbf{x}$ is Gaussian" = pdf of $\mathbf{x}$ is Gaussian, i.e. $p(\mathbf{x}) = c \exp[-(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) /2]$

Beyond the Scalar Example

Let's suppose that

- $\mathbf{x}$ is Gaussian, with mean $\mathbf{x}^b$ and covariance $\mathbf{B}$
- $\boldsymbol{\varepsilon}$ is Gaussian, with mean $\mathbf{0}$ and covariance $\mathbf{R}$
- H is linear, i.e. $H(\mathbf{x}) = \mathbf{H}\mathbf{x}$

1. Seek an estimate $\mathbf{x}^a$ that is a linear function of $\mathbf{x}^b$ and $\mathbf{y}^o$, the realization of obs we receive, and which minimizes the expected squared error

$$\mathbf{x}^a = \mathbf{K}\mathbf{y}^o + (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{x}^b = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)$$

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

Beyond the Scalar Example

Let's suppose that

- $\mathbf{x}$ is Gaussian, with mean $\mathbf{x}^b$ and covariance $\mathbf{B}$
- $\boldsymbol{\varepsilon}$ is Gaussian, with mean $\mathbf{0}$ and covariance $\mathbf{R}$
- H is linear, i.e. $H(\mathbf{x}) = \mathbf{H}\mathbf{x}$

2. Find the most likely $\mathbf{x}^a$, i.e. that maximizes $p(\mathbf{x} \mid \mathbf{y}^o)$

$$p(\mathbf{x}|\mathbf{y}^o) = c\exp\left(-\left[\frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + \frac{1}{2}(\mathbf{y}^o-\mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y}^o-\mathbf{H}\mathbf{x})\right]\right)$$

$$\mathbf{x}^a = \arg\max p = \arg\min J(\mathbf{x}), \text{ where } J(\mathbf{x}) = [\ldots]$$

Beyond the Scalar Example

Let's suppose that

- $\mathbf{x}$ is Gaussian, with mean $\mathbf{x}^b$ and covariance $\mathbf{B}$
- $\boldsymbol{\varepsilon}$ is Gaussian, with mean $\mathbf{0}$ and covariance $\mathbf{R}$
- H is linear, i.e. $H(\mathbf{x}) = \mathbf{H}\mathbf{x}$

Though it's not immediately obvious, (1) and (2) give same $\mathbf{x}^a$

Each approach also comes with (equivalent) equations for updated covariance, $\text{cov}(\mathbf{x} \mid \mathbf{y}^o)$

Importance of Covariances

Covariance measures whether two quantities vary together, independently, or in opposition

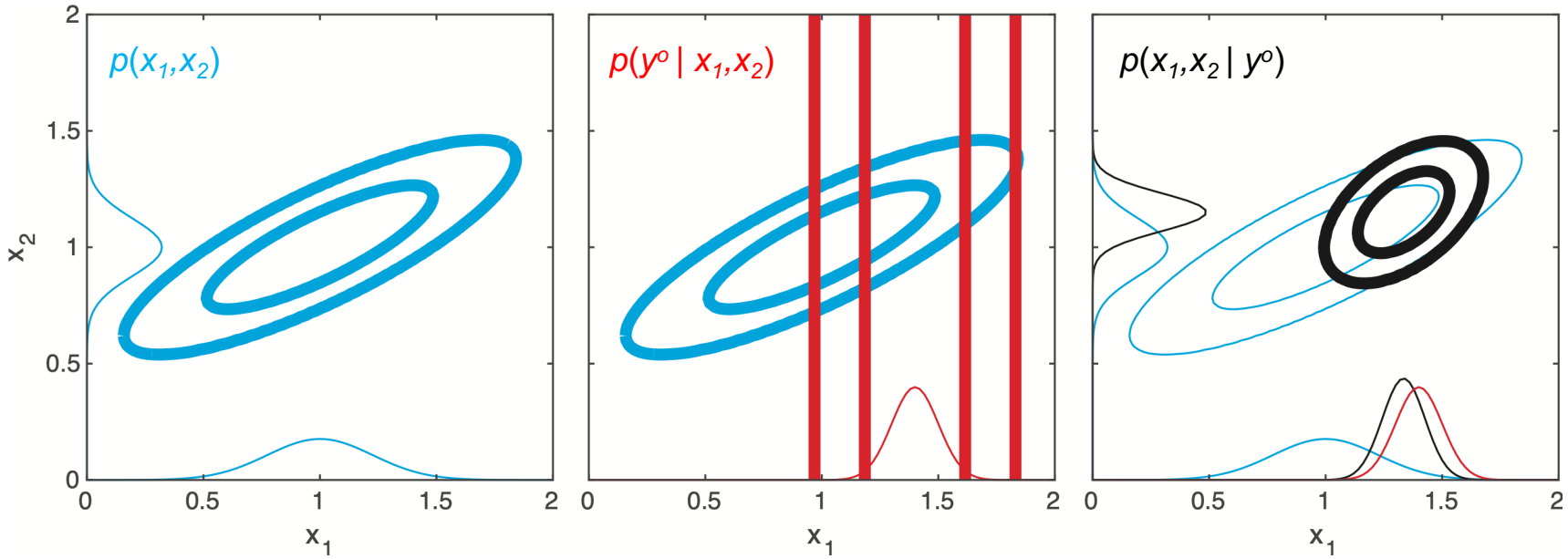
$$\mathbf{B} = E\left((\mathbf{x} - \mathbf{x}^b)(\mathbf{x} - \mathbf{x}^b)^T\right), \qquad [\mathbf{B}]_{ij} = E\left((x_i - x_i^b)(x_j - x_j^b)\right)$$

Why do covariances appear everywhere in DA?

Importance of Covariances

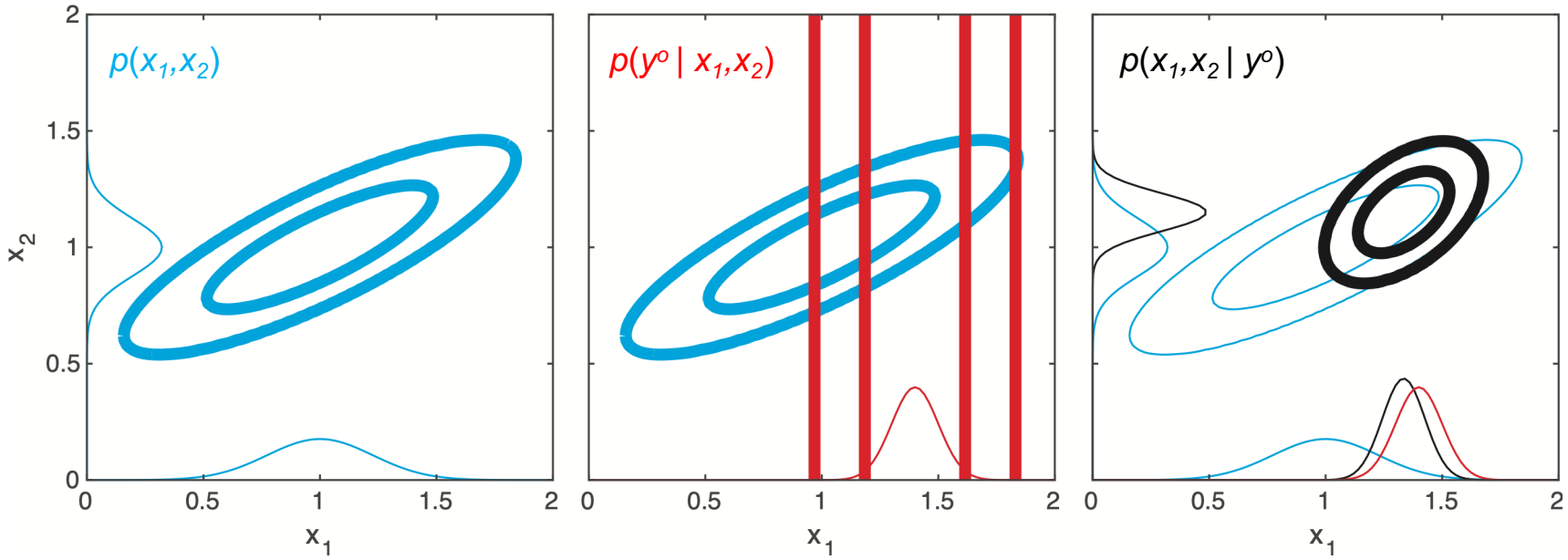
Consider a two-dimensional example, $\mathbf{x} = (x_1, x_2)$ and $y = x_1 + \boldsymbol{\varepsilon}$

Importance of Covariances



In this example, $y^o = 1.4$

Importance of Covariances



⇒ *non-zero covariances allow DA to update unobserved variables*

Modeling Covariances (I)

The bad news is that **B**, **R** are huge

- If **x** has dimension N_x, then **B** has dimension N_x by N_x … too large even to store on existing computers

1. Represent **B** implicitly, as sequence of sparse/cheap operations and their adjoints

$$\mathbf{x} = \mathbf{L}_2\mathbf{L}_1\mathbf{u} \text{ with } \mathrm{cov}(\mathbf{u}) \text{ diagonal}, \quad \mathbf{B} = \mathrm{cov}(\mathbf{x}) = \mathbf{L}_2\mathbf{L}_1\mathrm{cov}(\mathbf{u})\mathbf{L}_1^T\mathbf{L}_2^T$$

2. Assume **R** is diagonal or block diagonal
   - Independence of errors for different observations or observation types

Modeling Covariances (II)

Covariances in **B** also vary in space and time

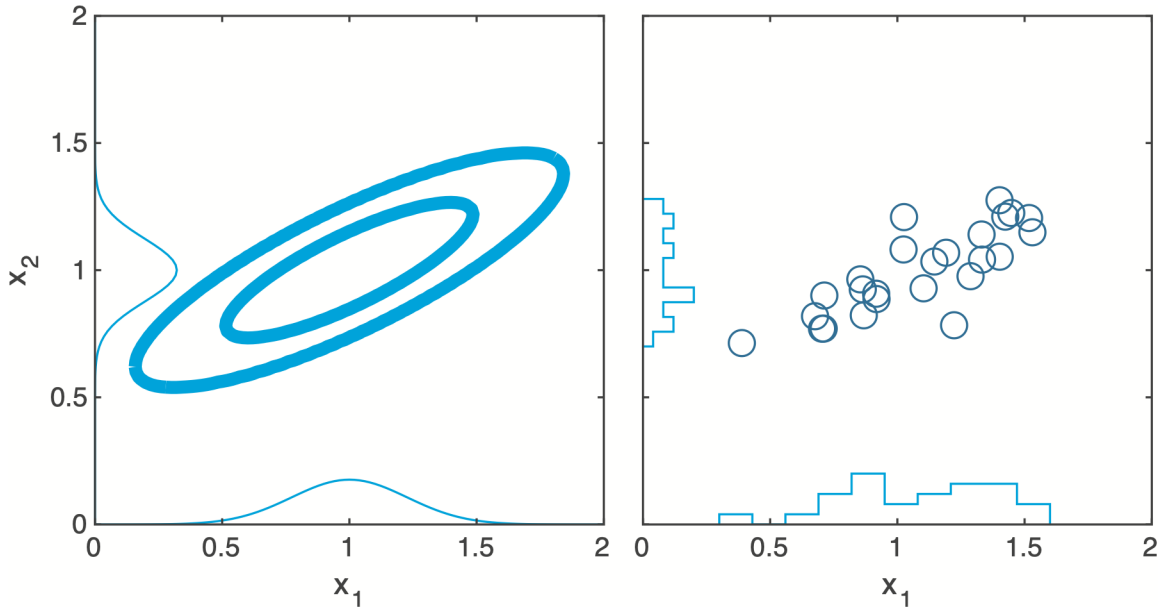
- Response of forecast errors to variation in observing network and in growth of forecast errors
- The latter depends on details of atmospheric flow (strong jet? conditionally unstable? etc)

How to capture flow dependence of **B**?

Ensemble Techniques

First, sample randomly from $p(\mathbf{x})$

- Typical algorithm: Generate ensemble of forecasts from ensemble of analyes; use ensemble DA to produce ensemble of analyses given ensemble of forecasts; cycle



Ensemble Techniques

Second, use sample covariance as approximation to required covariances

$$\mathbf{B}_e = (N_e - 1)^{-1} \sum_{i=1}^{N_e} (\mathbf{x}^i - \hat{\mathbf{x}})(\mathbf{x}^i - \hat{\mathbf{x}})^T, \text{ where } \hat{\mathbf{x}} \text{ is the ensemble mean}$$

Advantages of using $\mathbf{B}_e$ instead of $\mathbf{B}$

- Drastically reduces memory needs: $O(N_e N_x)$ rather than $O(N_x^2)$
- Errors in each element are $O(1/\text{sqrt } N_e)$

"e" = estimated from ensemble

Localization of $\mathbf{B}_e$

Disadvantages of using $\mathbf{B}_e$

- Most eigenvalues are zero; nonzero eigenvalues are much too large
- Results in most of obs information being ignored and analysis ensemble having too little spread

For most applications, “localization” is essential.

- Assume correlations decrease with spatial separation, then enforce in ensemble estimate of **B**

For DA schemes discussed later today, implement localization via

$$[\mathbf{L} \circ \mathbf{B}_e]_{ij} = [\mathbf{L}]_{ij}[\mathbf{B}_e]_{ij}$$

- **L** is correlation matrix whose entries decay with spatial separations; $\mathbf{L} \circ \mathbf{B}_e$ then reflects fact that, at sufficient separations, covariances are small
- Ameliorates disadvantages of raw $\mathbf{B}_e$
- Other schemes (e.g. LETKF) implement other forms of localization

Observations (again)

Observations of course are crucial to DA and to skill of subsequent forecasts

Key steps, which are only touched on in this tutorial:

- Acquisition
- Processing, including variable changes, thinning and superobing
- Specification of **R**, the observation-error covariance
- Bias correction, especially for satellite radiances
- Quality control to remove or decrease weight of observations that may be contaminated by gross errors