The University of Arizona
ATMOSPHERIC SCIENCES
Ave Arellano

DATA ASSIMILATION AND INFORMATION

Fundamental Perspectives of DA

ASP AQ Colloquium, Boulder, CO

August 2, 2016

Data Assimilation (DA) Perspectives:

– "interpolating fields for subsequent use as initial data in a model integration" (Bennett, 2002)

– "statistical combination of observations and short-range forecasts" (Kalnay, 2003)

– "using all the available information, to define as accurate as possible the state" (Talagrand, 1997)

– "incorporating data into the law" (Lewis et al., 2006)

“process by which observations are incorporated into a computer model of a real system” (Wikipedia)

For our purposes, Data Assimilation will be viewed as a method of combining information (whether embodied in observations or models).

Statistical Perspective

Fusing data (observations) with prior knowledge (e.g., physical laws, model output) to get an estimate of the true state of the physical system.

Sources of Information

- **observations**
  (these are measurements of the system)
- **models**
  (understanding of the spatio-temporal evolution of the system)

- physical constraints (moisture must be > 0)
- climatology

from a point of view of information, models and observations are not distinct; it is the mechanism of obtaining this information that is distinct

WHILE EXTREMELY USEFUL, THESE OBSERVATIONS ARE:

1. MOSTLY INDIRECT MEASUREMENTS OF THE STATE OF THE ENVIRONMENT
2. HAVE ASSOCIATED ERRORS
3. INCOMPLETE (DISCRETE)
4. IRREGULAR SAMPLES

$$y_k^o = h(x_k^t) + e_k, \qquad e_k \sim N(0, (\sigma^o)^2)$$

The true state at time k x_k^t is related to the observation through $h(\;)$.

errors: random (precision), systematic (bias), representativeness

Models on the other hand are imperfect. While they typically encapsulate our current understanding of the system, they

1. are incomplete (and discrete) representation of the system
2. have associated errors

$$\boldsymbol{x}^t(t_{k+1}) = \boldsymbol{M}_k[\boldsymbol{x}^t(t_k)] + \eta_k, \qquad \eta_k \sim N(0, q^2)$$

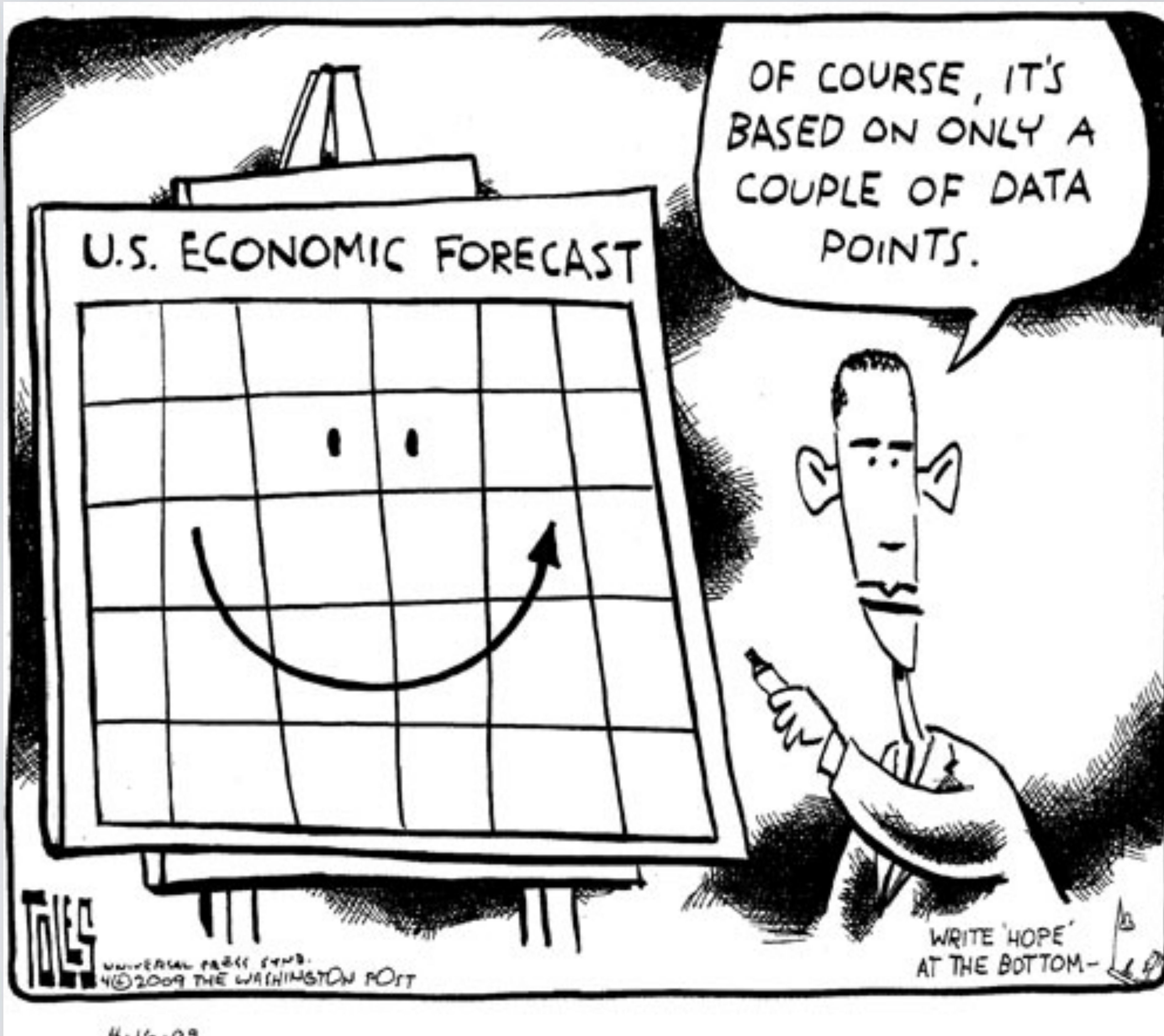
The evolution of the state is typically described as PDEs. e.g.,

$$\frac{\partial \rho_i}{\partial t} = \left[\frac{\partial \rho_i}{\partial t}\right]_{adv} + \left[\frac{\partial \rho_i}{\partial t}\right]_{mix} + \left[\frac{\partial \rho_i}{\partial t}\right]_{conv} + \left[\frac{\partial \rho_i}{\partial t}\right]_{scav} + \left[\frac{\partial \rho_i}{\partial t}\right]_{chem} + \left[\frac{\partial \rho_i}{\partial t}\right]_{em} + \left[\frac{\partial \rho_i}{\partial t}\right]_{dep}$$

Eq. 4.10 of Brasseur and Jacob, 2016

How do we produce a best estimate of the state of the Earth System given incomplete observations and imperfect models?

How do we ensure an observationally-constrained estimate that is at the same time consistent with our model understanding?


U.S. ECONOMIC FORECAST
OF COURSE, IT'S BASED ON ONLY A COUPLE OF DATA POINTS.
TOLES
UNIVERSAL PRESS SYND.
4 © 2009 THE WASHINGTON POST
WRITE 'HOPE' AT THE BOTTOM-
4-16-09

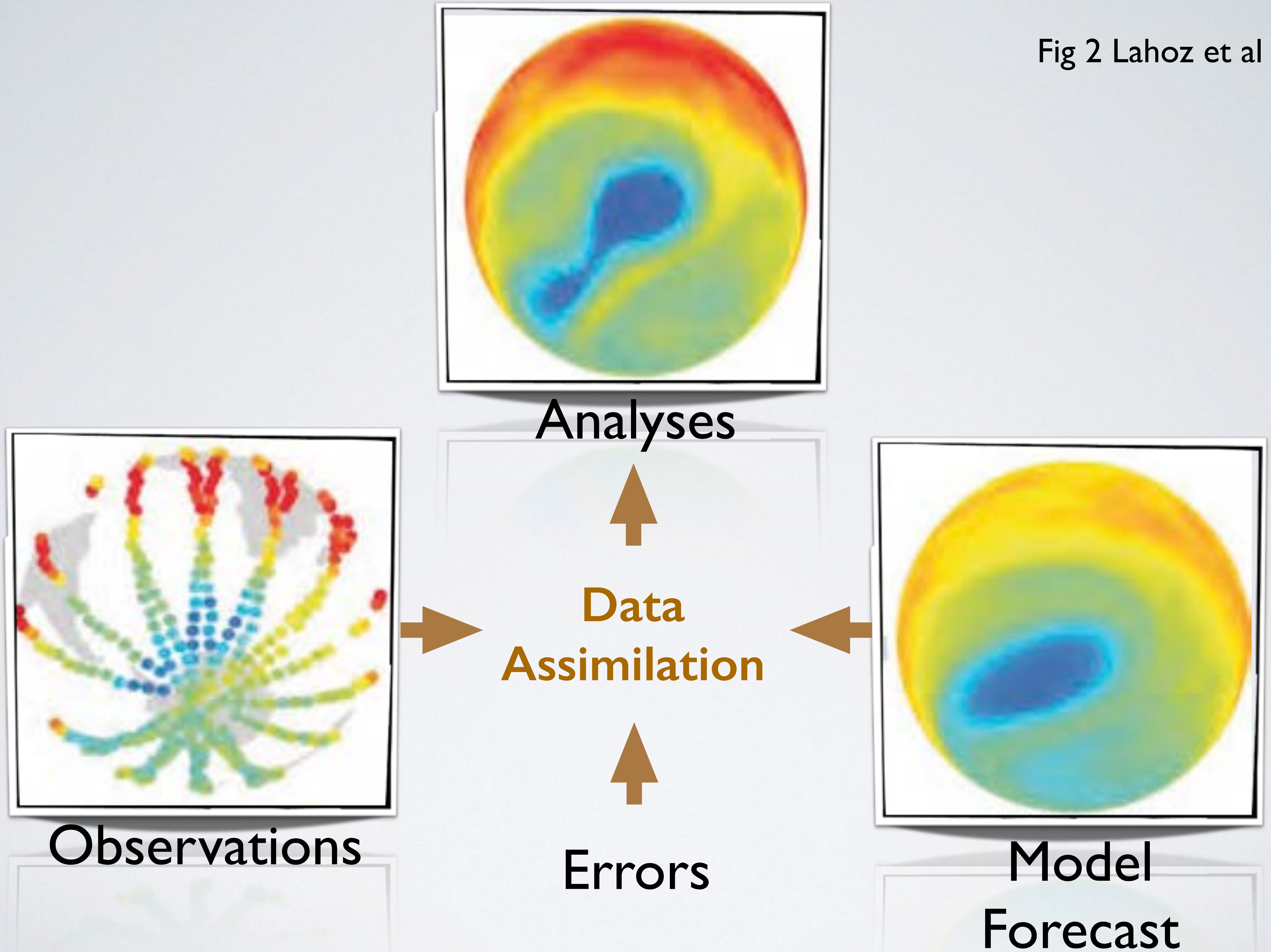
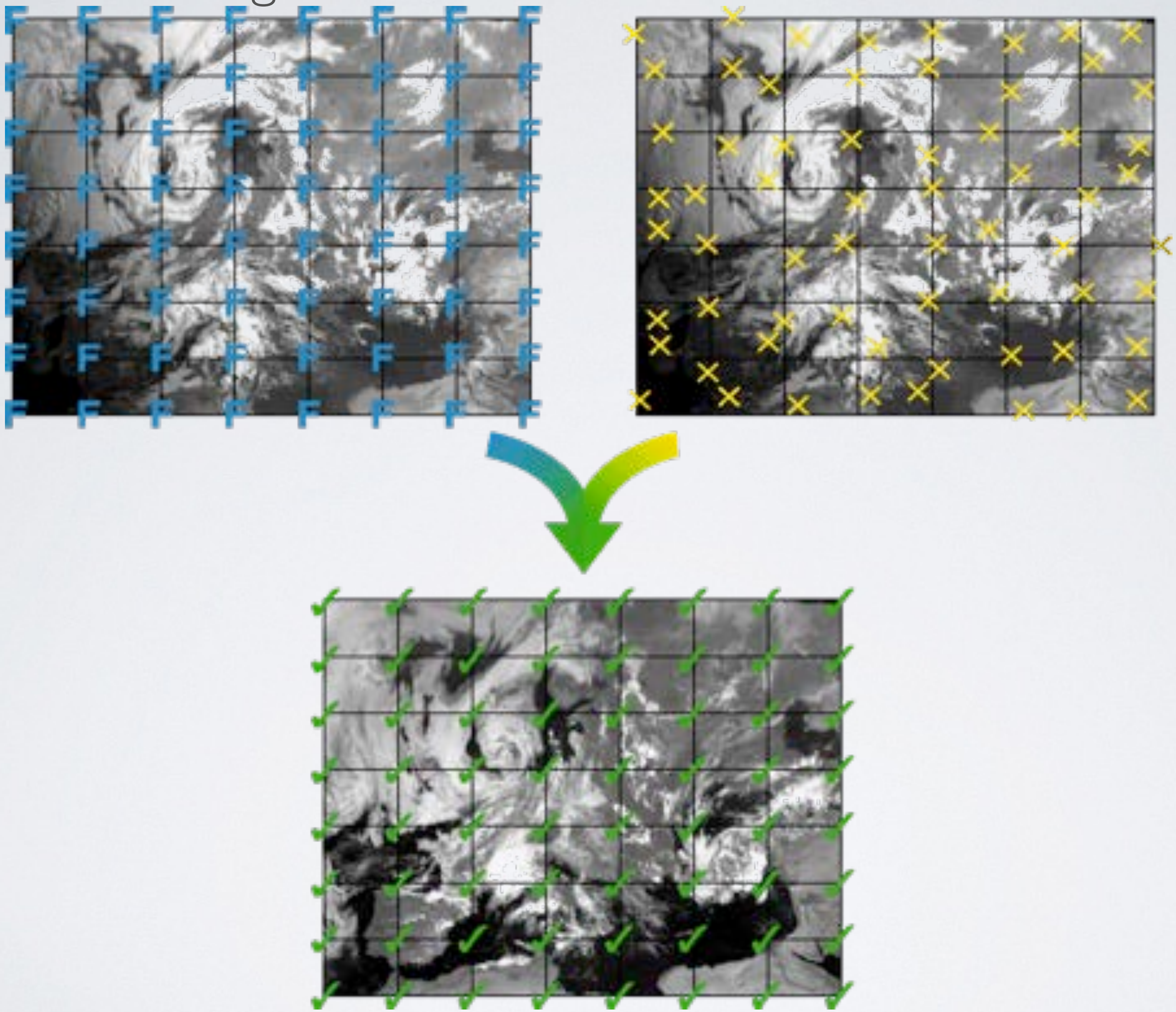


Fig 2 Lahoz et al 2010

short-range forecast

measurements



corrected representation of the atmosphere

How do we combine information?

Toy Example:

We have two measurements of temperature in this room (T_1 and T_2).

What would be our best estimate of the temperature (state) of this room?

Mean, RMS, Variance/Covariance, and Correlation

Supposed we have 2 data sets containing the values $x_1, x_2, x_3, \ldots x_n$ and $y_1, y, y_3, \ldots y_n$.

The mean, $\bar{x}$, is defined as a measure of central tendency (expected value of a random variable $\boldsymbol{x}$ or $E(\boldsymbol{x})$ or $\langle \boldsymbol{x} \rangle$, i.e.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The variance, σ_x^2 is defined as a measure of spread (expected value of squared deviation from the mean, or $E([\boldsymbol{x} - \bar{x}]^2)$, i.e.

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$var(\boldsymbol{x}) = E([\boldsymbol{x} - \bar{x}]^2) = cov(\boldsymbol{x}, \boldsymbol{x}) = \sigma_{x,x} \text{ or}$$

$$var(\boldsymbol{x}) = E(\boldsymbol{x}^2) - [E(\boldsymbol{x})]^2$$

The standard deviation σ_x is the square root of the variance, σ_x^2.

The covariance $\sigma_{x,y}$ is a measure of how $\boldsymbol{x}$ and $\boldsymbol{y}$ change together, i.e.

$$\sigma_{x,y} = \frac{1}{n}\sum_{i=1}^{n}[x_i - \bar{x}][y_i - \bar{y}]$$

$$cov(\boldsymbol{x},\boldsymbol{y}) = E([x_i - \bar{x}][y_i - \bar{y}]) = \langle[x_i - \bar{x}][y_i - \bar{y}]\rangle$$

$$cov(\boldsymbol{x},\boldsymbol{y}) = E(\boldsymbol{x}\,\boldsymbol{y}) - E(\boldsymbol{x})E(\boldsymbol{y}) = cov(\boldsymbol{y},\boldsymbol{x})$$

The correlation is defined as a measure of linear relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$, i.e.

$$\rho_{x,y} = \frac{\sigma_{x,y}}{\sqrt{\sigma_{x,x}}\sqrt{\sigma_{y,y}}} = \frac{\sigma_{x,y}}{\sigma_x\sigma_y}$$

For uncorrelated $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$cov(\boldsymbol{x},\boldsymbol{y}) = E(\boldsymbol{x}\,\boldsymbol{y}) - E(\boldsymbol{x})E(\boldsymbol{y}) = E(\boldsymbol{x})E(\boldsymbol{y}) - E(\boldsymbol{x})E(\boldsymbol{y}) = 0.$$

The variance of $\boldsymbol{x} + \boldsymbol{y}$ is defined as

$$var(\boldsymbol{x}+\boldsymbol{y}) = var(\boldsymbol{x}) + var(\boldsymbol{y}) + 2\,cov(\boldsymbol{x},\boldsymbol{y})$$

The root mean square of $\boldsymbol{x}$ is defined as a measure of the magnitude of $\boldsymbol{x}$ (or quadratic mean), i.e.

$$x_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2} = \sqrt{E(\boldsymbol{x}^2)} = \sqrt{\langle \boldsymbol{x}^2 \rangle}$$

What is the relation between x_{RMS} and σ_x?

$$\sigma_x = \sqrt{var(\boldsymbol{x})} = [E(\boldsymbol{x^2}) - [E(\boldsymbol{x})]^{\boldsymbol{2}}]^{1/2}$$

If $E(\boldsymbol{x}) = 0$, then

$$\sigma_x = x_{RMS}$$

Toy Example:

We want to measure the temperature in this room, and we have two thermometers that measure temperature with errors:

$$T_1 = T_t + e_1$$

$$T_2 = T_t + e_2$$

where T_t is the true value (which, like the errors, we never exactly know in reality).

We assume that the errors are random and unbiased and normally distributed: i.e.

$$E(e_1) = E(e_2) = 0$$

where $E(\ )$ is the "expectation". We also know the variances of these errors: i.e.

$$E(e_1^2) = \sigma_1^2 \quad \text{and} \quad E(e_2^2) = \sigma_2^2$$

Assume that the errors of the two measurements are uncorrelated:

$$E(e_1, e_2) = 0$$

Question: How can we estimate the true temperature in an objective (feasible) way?

Solution:

Estimate T_t as a linear combination of two pieces of information:

$$T_a = a_1 T_1 + a_2 T_2 \qquad (1)$$

Since $E(e_1) = E(e_2) = 0$, it follows that $E(e_a) = 0$ and that the 'analysis' T_a should be unbiased, i.e. $E(T_a) = T_t, E(T_1) = T_t$ and $E(T_2) = T_t$

$$E(T_a) = a_1 E(T_1) + a_2 E(T_2)$$

$$T_t = a_1 T_t + a_2 T_t$$

$$1 = a_1 + a_2 \qquad (2)$$

T_a will be the best estimate of T_t if coefficients minimize mean square error (Least Square Method). Since it is unbiased, minimizing the mean square error is the same as minimizing the error variance (Minimum Variance Method).

$$\sigma_a^2 = E[(T_a - T_t)^2] = E\left[\left(a_1(T_1 - T_t) + (1 - a_1)(T_2 - T_t)\right)^2\right] \qquad (3)$$

Expand Eq. (3), taking into account assumptions on correlation and bias. We then find a_1, a_2, σ_a^2 and T_a by first minimizing Eq. 3 (take the derivative with respect to a_1, equate to zero and solve for a_1).

We find:

$$a_1 = \frac{\sigma_2^2}{(\sigma_1^2 + \sigma_2^2)} = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

$$a_2 = \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)} = \frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

The weights of the observations are proportional to the precision/accuracy of the measurements (define here as the inverse of the variances of the observation errors). In other words, the weights depend on the relative accuracies of the observed estimates.

We also find:

$$T_a = \frac{\frac{T_1}{\sigma_1^2} + \frac{T_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} = \frac{\sigma_2^2}{(\sigma_1^2 + \sigma_2^2)} T_1 + \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)} T_2$$

$$\sigma_a^2 = \frac{\sigma_1^2 \sigma_2^2}{(\sigma_1^2 + \sigma_2^2)} = \sigma_2^2 (1 - a_1)$$

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

The error variance associated with the combined information is generally lower than the error associated with any of the 2 pieces of information being combined. And that at worse, it is equal to the minimum of the errors of the individual pieces of information but never larger.

If the error of one piece of information is infinitely large, the information from this piece of information being combined becomes vanishing small.

In the end, the analysis is the weighted average of the relative accuracies.

How about an analysis from an observation and a model (guess) information

Rewrite analysis equation in terms of a first guess (prior/forecast/background) information T_b and observation T_o

$$T_a = a_1 T_o + (1 - a_1) T_b$$

We find the same solution:

$$a_1 = \frac{\sigma_b^2}{(\sigma_o^2 + \sigma_b^2)} = \frac{\frac{1}{\sigma_o^2}}{\frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}}$$

$$T_a = \frac{\frac{T_o}{\sigma_o^2} + \frac{T_b}{\sigma_b^2}}{\frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}}$$

$$\sigma_a^2 = \frac{\sigma_o^2 \sigma_b^2}{(\sigma_o^2 + \sigma_b^2)} = \sigma_b^2 (1 - a_1)$$

In the end, the analysis is the weighted average of the relative accuracies of the observations and the model (first guess).

Our basic analysis equation:

$T_a = T_b + W(T_o - T_b)$ **analysis state estimate**

$\sigma_a^2 = \sigma_b^2(1 - W) = W\sigma_o^2 = \left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}\right)^{-1}$ **analysis error estimate**

If we let $\alpha = \frac{\sigma_o^2}{\sigma_b^2}$

$W = \frac{\frac{\sigma_b^2}{\alpha}}{\frac{\sigma_b^2}{\alpha} + \sigma_o^2} = \frac{1}{1+\alpha}$ **weights**

If $\sigma_o^2 \ll \sigma_b^2$, then $\alpha \approx 0, W \approx 1$ and $T_a \approx T_o, \sigma_a^2 \approx \sigma_o^2$ **obs highly certain**

If $\sigma_o^2 \gg \sigma_b^2$, then $\alpha \gg 1, W \approx 0$ and $T_a \approx T_b, \sigma_a^2 \approx \sigma_b^2$ **model highly certain**

If $\sigma_o^2 = \sigma_b^2$, then $\alpha = 1, W = 1/2$ and $T_a = 1/2(T_b + T_o), \sigma_a^2 = \frac{1}{2}\sigma_b^2 = \frac{1}{2}\sigma_o^2$ **average**

Our best (optimal) estimate (or analysis) of the state (temperature, in our toy example) is a linear combination of two pieces of information (model and observation). The weight applied to each information is associated with its relative accuracy. Our analysis is optimal since the corresponding error variance is minimum (i.e., it has the least mean square error).

Variational (cost function) Approach

We can also obtain the same best estimate of T_t by minimizing a function of the temperature defined as the sum of the square of the distance (or misfit) of the estimate T to the model and observations, weighted by their associated error precisions/accuracies:

$$J = \frac{1}{2}\left[\frac{(T - T_o)^2}{\sigma_o^2} + \frac{(T - T_b)^2}{\sigma_b^2}\right]$$

The squared deviation of T from either the model or observation is weighted in inverse proportion of the variance of the error on the model (or observation). Minimization of the 'cost' function J therefore imposes that T must fit either observation to within its own accuracy. This leads to an estimate $T = T_a$ given in the previous example using the method of weighted least squares.

Solution:

$$J = \frac{1}{2}\left[\frac{T^2 - 2TT_o + T_o{}^2}{\sigma_o^2} + \frac{T^2 - 2TT_o + T_o{}^2}{\sigma_b^2}\right]$$

$$\frac{\partial J}{\partial T} = \frac{T}{\sigma_o^2} - \frac{T_o}{\sigma_o^2} + \frac{T}{\sigma_b^2} - \frac{T_b}{\sigma_b^2} = 0$$

$$T\left(\frac{\sigma_o^2 + \sigma_b^2}{\sigma_o^2\sigma_b^2}\right) = \frac{\sigma_b^2 T_o + \sigma_o^2 T_b}{\sigma_o^2\sigma_b^2}$$

$$T = T_a = \frac{\sigma_b^2 T_o + \sigma_o^2 T_b}{(\sigma_o^2 + \sigma_b^2)} = \frac{\sigma_b^2}{(\sigma_o^2 + \sigma_b^2)}T_o + \frac{\sigma_o^2}{(\sigma_o^2 + \sigma_b^2)}T_b$$

$$T = T_a = a_1 T_o + (1 - a_1)T_b = T_b + W(T_o - T_b)$$

$$\frac{1}{\sigma_a^2} = \frac{\partial^2 J}{\partial T^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}$$

$$\sigma_a^2 = \left(\frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}\right)^{-1} = \frac{\sigma_o^2\sigma_b^2}{(\sigma_o^2 + \sigma_b^2)} = \sigma_b^2(1 - W) = W\sigma_o^2$$

Now, consider 3 pieces of information:

How do we combine the 3 pieces of information of temperature in this room to find our best estimate of temperature?

$$T_a = T_b + W_1\left(T_{o,1} - T_b\right) + W_2\left(T_{o,2} - T_b\right)$$

least squares

or

$$J = \frac{1}{2}\left[\frac{\left(T - T_{o,1}\right)^2}{\sigma_{o,1}^2} + \frac{\left(T - T_{o,2}\right)^2}{\sigma_{o,1}^2} + \frac{(T - T_b)^2}{\sigma_b^2}\right]$$

variational

or

$$T_{a,1} = T_b + \frac{\sigma_b^2}{\left(\sigma_{o,1}^2 + \sigma_b^2\right)}\left(T_{o,1} - T_b\right)$$

sequential filter

$$\sigma_{a,1}^2 = \left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_{o,1}^2}\right)^{-1}$$

$$T_a = T_{a,1} + \frac{\sigma_b^2}{\left(\sigma_{o,2}^2 + \sigma_b^2\right)}\left(T_{o,2} - T_{a,1}\right)$$

$$= T_b + \frac{\sigma_b^2}{\left(\sigma_{o,1}^2 + \sigma_b^2\right)}\left(T_{o,1} - T_b\right) + \frac{\sigma_b^2}{\left(\sigma_{o,2}^2 + \sigma_b^2\right)}\left(T_{o,2} - \left(T_b + \frac{\sigma_b^2}{\left(\sigma_{o,1}^2 + \sigma_b^2\right)}\left(T_{o,1} - T_b\right)\right)\right)$$

$$T_a = T_b + \frac{\sigma_b^2}{\left(\sigma_{o,1}^2 + \sigma_b^2\right)}\left(T_{o,1} - T_b\right) + \frac{\sigma_b^2}{\left(\sigma_{o,2}^2 + \sigma_b^2\right)}\left(T_{o,2} - T_b\right) - \frac{\sigma_b^2}{\left(\sigma_{o,2}^2 + \sigma_b^2\right)}\frac{\sigma_b^2}{\left(\sigma_{o,1}^2 + \sigma_b^2\right)}\left(T_{o,1} - T_b\right)$$

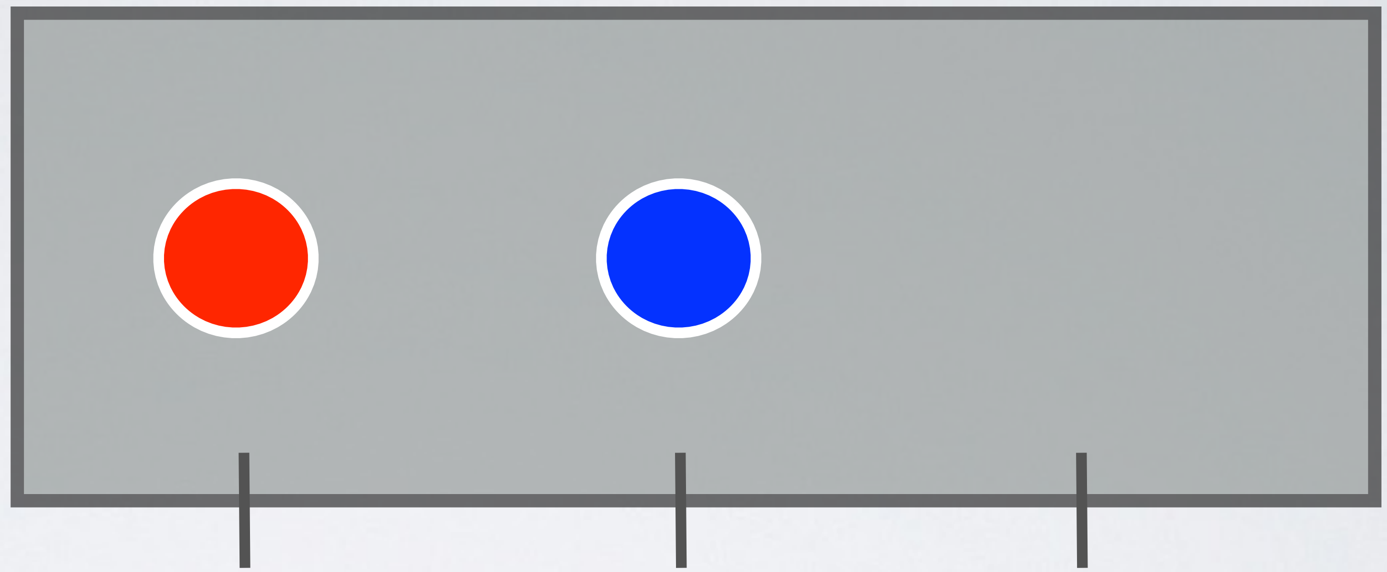
$$T_a = \sigma_a^2\left(\frac{T_b}{\sigma_b^2} + \frac{T_{o,1}}{\sigma_{o,1}^2} + \frac{T_{o,2}}{\sigma_{o,2}^2}\right)$$

$$\sigma_a^2 = \left(\left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_{o,1}^2}\right) + \frac{1}{\sigma_{o,2}^2}\right)^{-1}$$

OK, that was easy. What if we have 2 pieces of information that are not in the same location?

Guess and Observations at different location

Case 1: 1 observation & 2 model guess



Consider the case where the observation ($T_{o,1}$) is located at different location (say, the other room). We want to find our best estimate of the temperature in this room given our first guess (model) of the temperature ($T_{B,0}$ $T_{B,1}$) in both rooms.

Given:

Observation of room temperature at grid 1 and its error characteristics,

$$T_{o,1} = T_{t,1} + e_{o,1} \text{ where } E\left(e_{o,1}\right) = 0, E\left(e_{o,1}^2\right) = \sigma_{o,1}^2$$

as well as, a first guess of the room temperature at grid 0 and 1 and their error characteristics,

$$T_{b,0} = T_{t,0} + e_{b,0} \text{ where } E\left(e_{b,0}\right) = 0, E\left(e_{b,0}^2\right) = \sigma_{b,0}^2$$

$$T_{b,1} = T_{t,1} + e_{b,1} \text{ where } E\left(e_{b,1}\right) = 0, E\left(e_{b,1}^2\right) = \sigma_{b,1}^2$$

where

$$\sigma_{b,0}^2 = \sigma_{b,1}^2 \text{ and } E\left(e_{b,0}, e_{b,1}\right) = \rho_{0,1}\sigma_b^2$$

Assume that the error in observation is uncorrelated with the errors in our first guess, i.e.,

$$E\left(e_{o,1}, e_{b,0}\right) = \sigma_{\{o,1\}\{b,0\}} = 0$$

$$E\left(e_{o,1}, e_{b,1}\right) = \sigma_{\{o,1\}\{b,1\}} = 0$$

Solution:

Using our analysis expression $T_{a,0} = T_{b,0} + W\left(T_{o,1} - T_{b,1}\right)$, we can subtract from this equation the true temperature T_t to formulate the analysis error equation.

$$\left(T_{a,0} - T_{t,0}\right) = \left(T_{b,0} - T_{t,0}\right) + W\left(\left[T_{o,1} - T_{t,1}\right] - \left[T_{b,1} - T_{t,1}\right]\right)$$

such that,

$$\left(e_{a,0}\right) = \left(e_{b,0}\right) + W\left(\left[e_{o,1}\right] - \left[e_{b,1}\right]\right)$$

We find W by finding the least square error of the analysis assuming that the model, observation, and analysis are unbiased.

First, form an expression of the square error of the analysis.

$$\left(e_{a,0}\right)^2 = \left(e_{b,0}\right)^2 + 2\left(e_{b,0}\right) W\left(\left[e_{o,1}\right] - \left[e_{b,1}\right]\right) + W^2\left(\left[e_{o,1}\right] - \left[e_{b,1}\right]\right)^2$$

Second, take the ensemble average (expected value) of the square error of the analysis

$$E\left(e_{a,0}^2\right) = E\left(e_{b,0}^2\right) + 2WE\left(e_{b,0}\left[e_{o,1} - e_{b,1}\right]\right) + W^2 E\left(\left[e_{o,1} - e_{b,1}\right]^2\right)$$

$$\sigma_{a,0}^2 = \sigma_{b,0}^2 + 2W\,\sigma_{\{o,1\}\{b,0\}} - 2W\,\rho_{0,1}\sigma_b^2 + W^2\sigma_{o,1}^2 - 2W^2\sigma_{\{o,1\}\{b,1\}} + W^2\sigma_{b,1}^2$$

Since the error in the observation is not correlated with the model,

$$\sigma_{a,0}^2 = \sigma_{b,0}^2 + 0 - 2W\,\rho_{0,1}\sigma_b^2 + W^2\sigma_{o,1}^2 - 0 + W^2\sigma_{b,1}^2$$

Third, find the derivative of $\sigma_{a,0}^2$ with respect to W and equate to zero, i.e.,

$$\frac{d\sigma_{a,0}^2}{dW} = -2\rho_{0,1}\sigma_b^2 + 2W\sigma_{o,1}^2 + 2W\sigma_{b,1}^2 = 0$$

$$2W\left(\sigma_{o,1}^2 + \sigma_{b,1}^2\right) = 2\rho_{0,1}\sigma_b^2$$

$$W = \frac{\rho_{0,1}\sigma_b^2}{\left(\sigma_{o,1}^2 + \sigma_{b,1}^2\right)} = \frac{\rho_{0,1}}{1+\alpha} \quad \text{where } \alpha = \frac{\sigma_{o,1}^2}{\sigma_{b,1}^2}$$

and so

$$T_{a,0} = T_{b,0} + \frac{\rho_{0,1}\sigma_b^2}{\left(\sigma_{o,1}^2 + \sigma_{b,1}^2\right)}\left(T_{o,1} - T_{b,1}\right)$$

or

$$T_{a,0} = T_{b,0} + \frac{\rho_{0,1}}{1+\alpha}\left(T_{o,1} - T_{b,1}\right)$$

The analysis mean square error is:

$$\sigma_{a,0}^2 = \sigma_{b,0}^2 - 2\left(\frac{\rho_{0,1}\sigma_b^2}{\left(\sigma_{o,1}^2 + \sigma_{b,1}^2\right)}\right)\rho_{0,1}\sigma_b^2 + \left(\frac{\rho_{0,1}\sigma_b^2}{\left(\sigma_{o,1}^2 + \sigma_{b,1}^2\right)}\right)^2\sigma_{o,1}^2 + \left(\frac{\rho_{0,1}\sigma_b^2}{\left(\sigma_{o,1}^2 + \sigma_{b,1}^2\right)}\right)^2\sigma_{b,1}^2$$

to simplify let $a = \sigma_{a,0}^2$, $b = \sigma_{b,0}^2 = \sigma_{b,1}^2 = \sigma_b^2$, $\rho = \rho_{0,1}$, and $c = \sigma_{o,1}^2$

$$a = b - \frac{2\rho^2 b^2}{(b+c)} + \frac{\rho^2 b^2 b}{(b+c)^2} + \frac{\rho^2 b^2 c}{(b+c)^2}$$

$$a = \frac{(b+c)(b+c)b - (b+c)2\rho^2 b^2 + (b+c)\rho^2 b^2}{(b+c)(b+c)}$$

$$a = \frac{(b+c)b - 2\rho^2 b^2 + \rho^2 b^2}{(b+c)} = \frac{b\big((b+c) - \rho^2 b\big)}{(b+c)}$$

$$\sigma_{a,0}^2 = \sigma_b^2 \frac{\left(\sigma_b^2 + \sigma_{o,1}^2 - {\rho_{0,1}}^2 \sigma_b^2\right)}{\left(\sigma_b^2 + \sigma_{o,1}^2\right)} = \sigma_b^2 \left(1 - \frac{{\rho_{0,1}}^2 \sigma_b^2}{\left(\sigma_b^2 + \sigma_{o,1}^2\right)}\right)$$

$$\sigma_{a,0}^2 = \sigma_b^2\left(1 - \rho_{0,1} W\right) = \sigma_b^2 \left(1 - \frac{{\rho_{0,1}}^2}{1+\alpha}\right)$$

In summary,

$$T_{a,0} = T_{b,0} + \frac{\rho_{0,1}\sigma_b^2}{\left(\sigma_{o,1}^2 + \sigma_{b,1}^2\right)}\left(T_{o,1} - T_{b,1}\right) = T_{b,0} + \frac{\rho_{0,1}}{1+\alpha}\left(T_{o,1} - T_{b,1}\right)$$

analysis state estimate

$$\sigma_{a,0}^2 = \sigma_b^2\left(1 - \rho_{0,1} W\right) = \sigma_b^2 \left(1 - \frac{{\rho_{0,1}}^2}{1+\alpha}\right)$$

analysis state error estimate

Summary

1 obs, 1 guess (collocated)

$$T_a = T_b + \frac{1}{1+\alpha}(T_o - T_b)$$

$$\sigma_a^2 = \sigma_b^2\left(1 - \frac{1}{1+\alpha}\right) = \frac{\sigma_b^2\sigma_o^2}{(\sigma_o^2 + \sigma_b^2)} = \left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}\right)^{-1}$$

1 obs, 2 guesses at different locations

$$T_{a,0} = T_{b,0} + \frac{\rho_{0,1}}{1+\alpha}(T_{o,1} - T_{b,1})$$

$$\sigma_{a,0}^2 = \sigma_b^2\left(1 - \frac{{\rho_{0,1}}^2}{1+\alpha}\right) = \sigma_b^2\left(1 - \frac{{\rho_{0,1}}^2\sigma_b^2}{(\sigma_b^2 + \sigma_{o,1}^2)}\right)$$

2 obs, 3 guesses at different locations

$$T_{a,0} = T_{b,0} + \frac{\rho_{0,1}(1+\alpha) - \rho_{0,2}\,\rho_{1,2}}{(1+\alpha)^2 - {\rho_{1,2}}^2}(T_{o,1} - T_{b,0}) + \frac{\rho_{0,2}(1+\alpha) - \rho_{0,1}\,\rho_{1,2}}{(1+\alpha)^2 - {\rho_{1,2}}^2}(T_{o,2} - T_{b,2})$$

$$\sigma_{a,0}^2 = \sigma_b^2\left(1 - \frac{(1+\alpha)\left([\rho_{0,1}]^2 + [\rho_{0,2}]^2\right) - 2\,\rho_{0,1}\,\rho_{0,2}\,\rho_{1,2}}{(1+\alpha)^2 - (\rho_{1,2})^2}\right)$$

Indirect Measurements (Use of Observation Operator)

We have an object, a stone in space. We want to estimate its temperature T_a (in Kelvin units) accurately but we measure the radiance flux density, y (in Watts/m^2) that it emits. We have an observation model $y = h(T_t)$, i.e., $y = \sigma {T_t}^4$, where $h(\quad)$ is a non-linear forward model (observation) operator that includes in some cases transformation and grid interpolation.

We have the following expressions for the measurement process and estimation:

$$y = h(T_t) + e_o$$

$$T_b = T_t + e_b$$

$$T_a = T_t + e_a$$

$$T_a = T_b + K\big(y - h(T_b)\big)$$

assuming $E(e_o) = E(e_b) = E(e_a) = 0,\ E(e_o, e_b) = 0, E(e_o^2) = \ \sigma_o^2$ and $E(e_b^2) = \ \sigma_b^2$.

Problem: Estimate T_a and $E(e_a^2) = \ \sigma_a^2$.

Note: From Taylor Series,

$$h(T_t) = \ h(T_b) + \left.\frac{dh(T_t)}{dT_t}\right|_{T_b} (T_t - T_b)$$

$$H = \left.\frac{dh(T_t)}{dT_t}\right|_{T_b}$$

H is the derivative of the forward model operator with respect the model state and evaluated at the model first guess (background state). Here, we have performed a linearization of the nonlinear operator around the background state, implicitly assuming that the truth is not too far from the background.

After estimating T_a and σ_a^2, consider a simpler linear case (i.e. $y = hT_t$).

Solution:

Our analysis (unbiased) is a linear combination of our first guess (model information) and our measurement (observed information):

$$T_a = T_b + K\big(y - h(T_b)\big) \qquad (1)$$

To estimate T_a we find the 'weights' K such that the mean square error of T_a is minimum (least squares), i.e.

1) set the expression for mean square error

$$\sigma_a^2 = E[(T_a - T_t)^2] \qquad (2)$$

2) take its derivative and equate to zero

$$\frac{d\sigma_a^2}{dK} = 0$$

3) solve for K

Expand Eq. 2 by first substituting Eq. 1 to T_a in Eq. 2

$$\sigma_a^2 = E[(T_a - T_t)^2] = E\left[\left(T_b + K\left(y - h(T_b)\right) - T_t\right)^2\right] \tag{3}$$

We know that,

$T_b = T_t + e_b$ and $y = h(T_t) + e_o$ such that

$h(T_b) = h(T_t) + h(e_b)$

Substituting these to Eq. 3

$$\sigma_a^2 = E[(T_a - T_t)^2] = E\left[\left(T_b + K\left(h(T_t) + e_o - \left(h(T_t) + h(e_b)\right)\right) - T_t\right)^2\right]$$

We then linearize $h(T_t)$ at T_b

$$h(e_b) = h(T_b) - h(T_t) = h(T_b) - \left(h(T_b) + H(T_t - T_b)\right) = He_b$$

$$\sigma_a^2 = E[(T_a - T_t)^2] = E\left[\left(T_b + K\left(h(T_t) + e_o - (h(T_t) + He_b)\right) - T_t\right)^2\right]$$

$$\sigma_a^2 = E[(T_a - T_t)^2] = E\left[\left(e_b + K(e_o - He_b)\right)^2\right]$$

$$\sigma_a^2 = E[(T_a - T_t)^2] = E\left[\left(e_b + K(e_o - He_b)\right)^2\right]$$

Taking its derivative and assuming $E(e_o^2) = \sigma_o^2$ and $E(e_b^2) = \sigma_b^2$

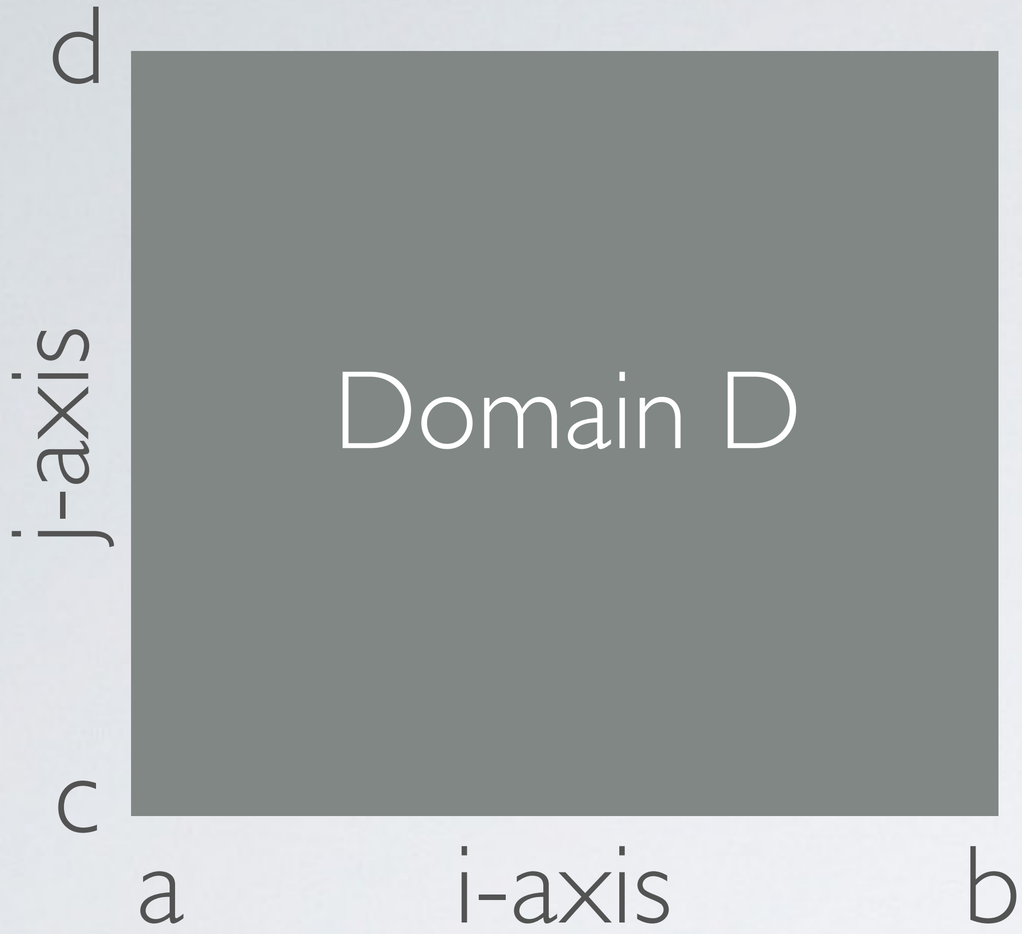
$$\frac{d\sigma_a^2}{dK} = -2H\sigma_b^2 + 2\mathrm{K}\sigma_o^2 + 2KH^2\sigma_b^2 = 0$$

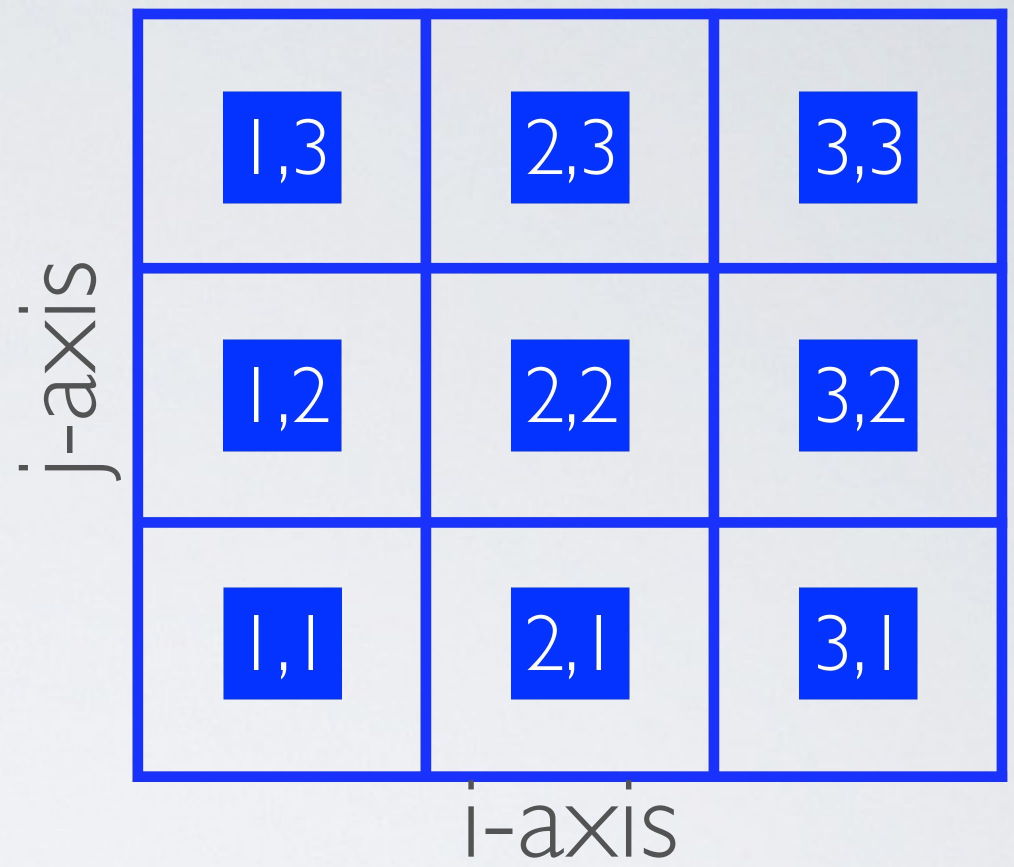
$$K = \frac{H\sigma_b^2}{\sigma_b^2 + H^2\sigma_b^2}$$

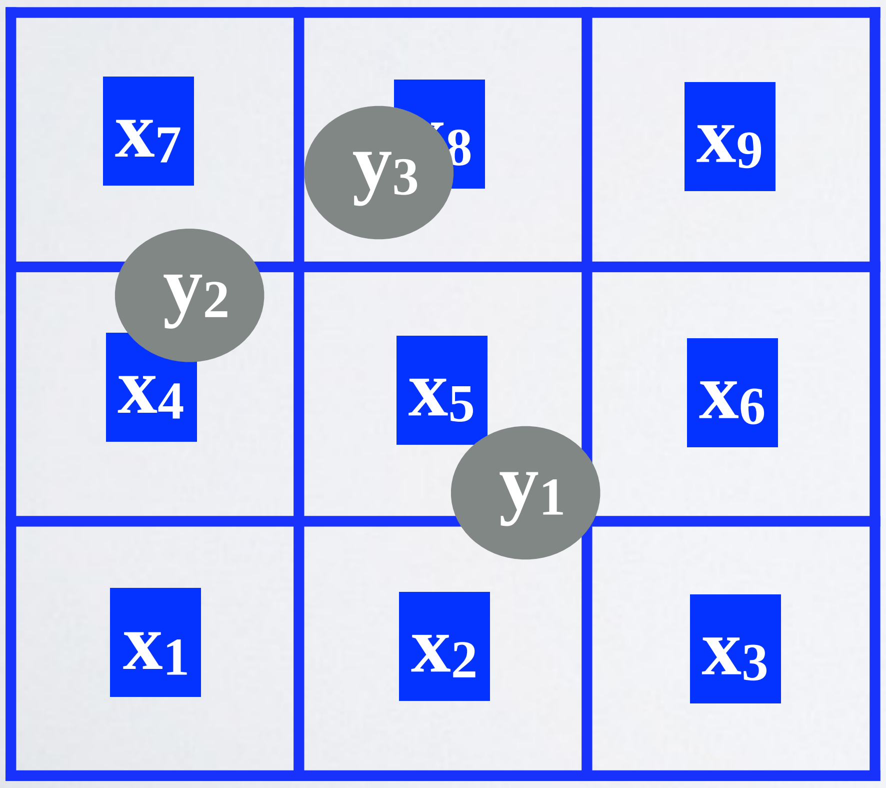
$$T_a = T_b + \frac{H\sigma_b^2}{\sigma_b^2 + H^2\sigma_b^2}\left(y - h(T_b)\right)$$

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_b^2} + \frac{H^2}{\sigma_o^2}$$

$$\sigma_a^2 = \sigma_b^2(1 - KH)$$







Unknown States

$$\mathbf{x} = (x_1, x_2, .. x_l, .. x_n)^T$$

$$l=(j-1)n_i+i$$

Background States

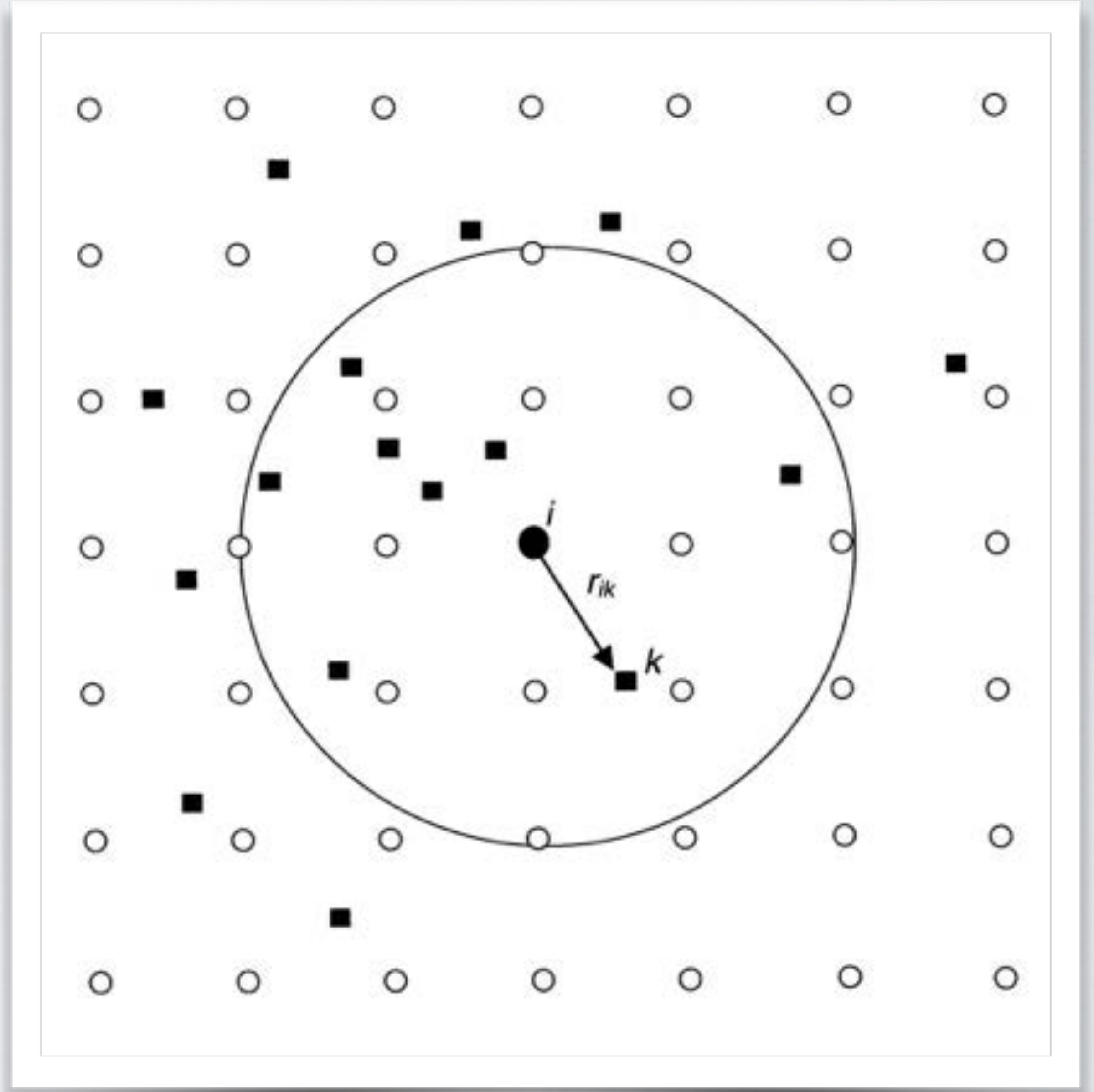
$$\mathbf{x_b} = (x_1^b, x_2^b, .. x_l^b, .. x_n^b)^T$$

Observations

$$\mathbf{y_o} = (y_1^o, y_2^o, \ldots\ y_m^o)^T$$

Observations are in general, different from the modeled state variables by: a) being located in different points and b) possibly being indirect measures of the modeled state variables.

$$\mathbf{y} = \mathbf{Hx} + \mathbf{v}$$



$\mathbf{H}$ is the forward observational operator that converts the background field into 'first guess of the observations'. $\mathbf{H}$ can be nonlinear (or just an interpolation operator). The innovation (obs increment) is $\mathbf{d} = \mathbf{y}_o - \mathbf{y}_b = \mathbf{y}_o - \mathbf{H}\mathbf{x}_b$

Let $\mathbf{d} = \mathbf{y}_o - \mathbf{y}_b = \mathbf{y}_o - \mathbf{H}\mathbf{x}_b$ and $\hat{\boldsymbol{e}} = \hat{\mathbf{x}} - \mathbf{x}$

Similar to our previous examples, we find a weight matrix W such that our estimate minimizes the mean square error

$$\mathbf{x} - \mathbf{x}_b = \boldsymbol{W}(\mathbf{y}_o - \mathbf{H}\mathbf{x}_b) - \hat{\mathbf{e}} = \boldsymbol{W}\mathbf{d} - \hat{\mathbf{e}}$$

An error covariance matrix is obtained by multiplying a vector error $\boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$ by its transpose $\boldsymbol{e}^T = [e_1 \quad e_2 \quad \cdots \quad e_n]$

and averaging over many cases to obtain the expected value:

$$\mathbf{P} = \mathrm{E}(\boldsymbol{e}\boldsymbol{e}^T) = \overline{\boldsymbol{e}\boldsymbol{e}^T} = \begin{bmatrix} \overline{e_1 e_1} & \overline{e_1 e_2} & \cdots & \overline{e_1 e_n} \\ \overline{e_2 e_1} & \overline{e_2 e_2} & \cdots & \overline{e_2 e_n} \\ \vdots & \vdots & & \vdots \\ \overline{e_n e_1} & \overline{e_n e_2} & \cdots & \overline{e_n e_n} \end{bmatrix}$$

This matrix is symmetric and positive definite. The diagonal elements are the variances of the vector error components

$$\overline{e_\iota e_\iota} = \sigma_i^2$$

If we normalize the covariance matrix, dividing each component by the product of the standard deviations: $\overline{e_i e_j}/_{\sigma_i \sigma_j} = corr(e_i, e_j) = \rho_{ij}$

we obtain a correlation matrix:

$$\mathbf{C} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & & \vdots \\ \rho_{1n} & \rho_{2n} & \cdots & 1 \end{bmatrix}$$

and if $\mathbf{D} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$

is the diagonal matrix of the variances, then we can write

$$\mathbf{P} = \mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2}$$

Statistical Assumptions

$$e_b(i,j) = x_b(i,j) - x(i,j)$$
$$\hat{e}(i,j) = \hat{x}(i,j) - x(i,j)$$

$$e_{oi} = y_o(r_i) - y(r_i) = y_o(r_i) - Hx(r_i)$$

We do not know the truth x, thus we do not know the errors of the available background and observations. But we can make a number of assumptions about their statistical properties. The background and observations are assumed to be unbiased.

$$E\{e_b(i,j)\} = E\{x_b(i,j)\} - E\{x(i,j)\} = 0$$
$$E\{e_o(r_i)\} = E\{y_o(r_i)\} - E\{y(r_i)\} = 0$$

If the forecasts (background) and the observations are biased, in principle we can and should correct the bias before proceeding. The bias can also be estimated as part of the analysis (Dee and Da Silva (1998).

Statistical Assumptions

$$\hat{\mathbf{P}} = \mathbf{P}_{\mathbf{x}} = E\{\hat{\mathbf{e}}\hat{\mathbf{e}}^T\}, \qquad \mathbf{P}_{\boldsymbol{b}} = \mathbf{B} = E\{\mathbf{e}_{\mathbf{b}}\mathbf{e}_{\mathbf{b}}{}^{\boldsymbol{T}}\}, \quad \mathbf{P}_{\boldsymbol{o}} = \mathbf{R} = E\{\mathbf{e}_{\mathbf{o}}\mathbf{e}_{\mathbf{o}}{}^{\boldsymbol{T}}\}$$

$$E\{\mathbf{e}_{\mathbf{o}}\mathbf{e}_{\mathbf{b}}{}^{\boldsymbol{T}}\} = 0$$

The nonlinear observation operator $\boldsymbol{H}$ that transforms model state variables into observed variables can be linearized as:

$$H(\mathbf{x} + \delta\mathbf{x}) = H(\mathbf{x}) + \mathbf{H}\delta\mathbf{x}$$

where $\mathbf{H}$ is a pxn matrix denoting the linear observation operator with elements $h_{i,j} = {\partial H_i}/{\partial x_j}$

We also assume that the background (usually a model forecast) is a good approximation of the truth, so that the analysis and the observations are equal to the background values plus small increments. That is,

$$\mathbf{d} = \mathbf{y}_{\boldsymbol{o}} - \mathbf{y}_{\boldsymbol{b}} = \mathbf{y}_{\boldsymbol{o}} - H\big(\mathbf{x} + (\mathbf{x}_{\boldsymbol{b}} - \mathbf{x})\big)$$
$$\mathbf{d} = \mathbf{y}_{\boldsymbol{o}} - H(\mathbf{x}) - \mathbf{H}(\mathbf{x}_{\boldsymbol{b}} - \mathbf{x}) = \mathbf{e}_{\mathbf{o}} - \mathbf{H}\mathbf{e}_{\mathbf{b}}$$

The solution to this problem (least squares) is:

$$\hat{\mathbf{x}} = \mathbf{x}_{\boldsymbol{b}} + \boldsymbol{W}(\mathbf{y}_{\boldsymbol{o}} - \mathbf{H}\mathbf{x}_{\boldsymbol{b}})$$

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

$$\hat{\mathbf{P}} = (\mathbf{I}_n - \boldsymbol{W}\mathbf{H})\mathbf{B}$$

Recall:

Toy Example 1:

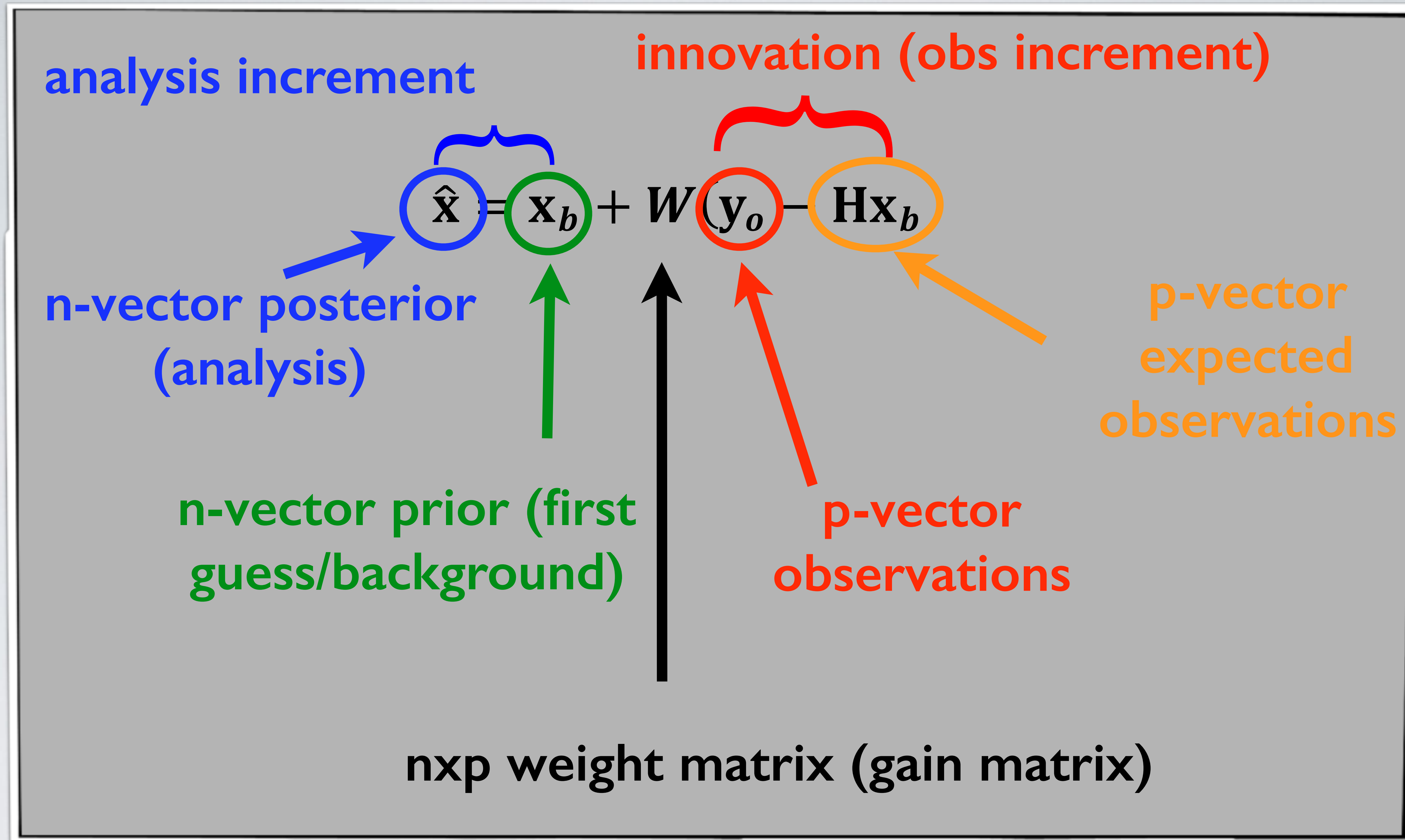
$$T_a = T_b + a_1(T_o - T_b)$$

$$a_1 = \frac{\sigma_b^2}{(\sigma_b^2 + \sigma_o^2)}$$

Toy Example 2:

$$T_a = T_b + K\big(y - h(T_b)\big)$$

$$K = \frac{H\sigma_b^2}{H^2\sigma_b^2 + \sigma_o^2} = \frac{(H\sigma_b)\sigma_b}{(H\sigma_b)^2 + \sigma_o^2}$$


analysis increment
innovation (obs increment)
$\hat{\mathbf{x}} = \mathbf{x}_b + W(\mathbf{y}_o - \mathbf{H}\mathbf{x}_b)$
n-vector posterior (analysis)
p-vector expected observations
n-vector prior (first guess/background)
p-vector observations
nxp weight matrix (gain matrix)

$$\hat{\mathbf{x}} = \mathbf{x}_b + W(\mathbf{y}_o - \mathbf{H}\mathbf{x}_b)$$

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

nxp error covariance between background and expected observation

pxp observation error covariance

pxp expected observation error covariance

Think Toy Example:

$$\hat{\mathbf{x}} = \mathbf{x}_b + W(\mathbf{y}_o - \mathbf{H}\mathbf{x}_b)$$

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

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

regression (map to model space) — ratio of model error to total error (shrink in obs space)

i.e.,

$$T_{a,0} = T_{b,0} + \frac{\rho_{0,1}\sigma_b^2}{(\sigma_{o,1}^2 + \sigma_{b,1}^2)}(T_{o,1} - T_{b,1})$$

$$W = \left(\rho_{0,1}\frac{\sigma_{b,0}}{\sigma_{b,1}}\right)\left(\frac{\sigma_{b,1}^2}{\sigma_{o,1}^2 + \sigma_{b,1}^2}\right) = \left(\frac{\sigma_{0,1}^2}{\sigma_{b,1}^2}\right)\left(\frac{\sigma_{b,1}^2}{\sigma_{o,1}^2 + \sigma_{b,1}^2}\right), \qquad \frac{\sigma_{b,0}}{\sigma_{b,1}} = 1$$

$$\frac{\partial \rho_i}{\partial t} = \left[\frac{\partial \rho_i}{\partial t}\right]_{adv} + \left[\frac{\partial \rho_i}{\partial t}\right]_{mix} + \left[\frac{\partial \rho_i}{\partial t}\right]_{conv} + \left[\frac{\partial \rho_i}{\partial t}\right]_{scav} + \left[\frac{\partial \rho_i}{\partial t}\right]_{chem} + \left[\frac{\partial \rho_i}{\partial t}\right]_{em} + \left[\frac{\partial \rho_i}{\partial t}\right]_{dep}$$

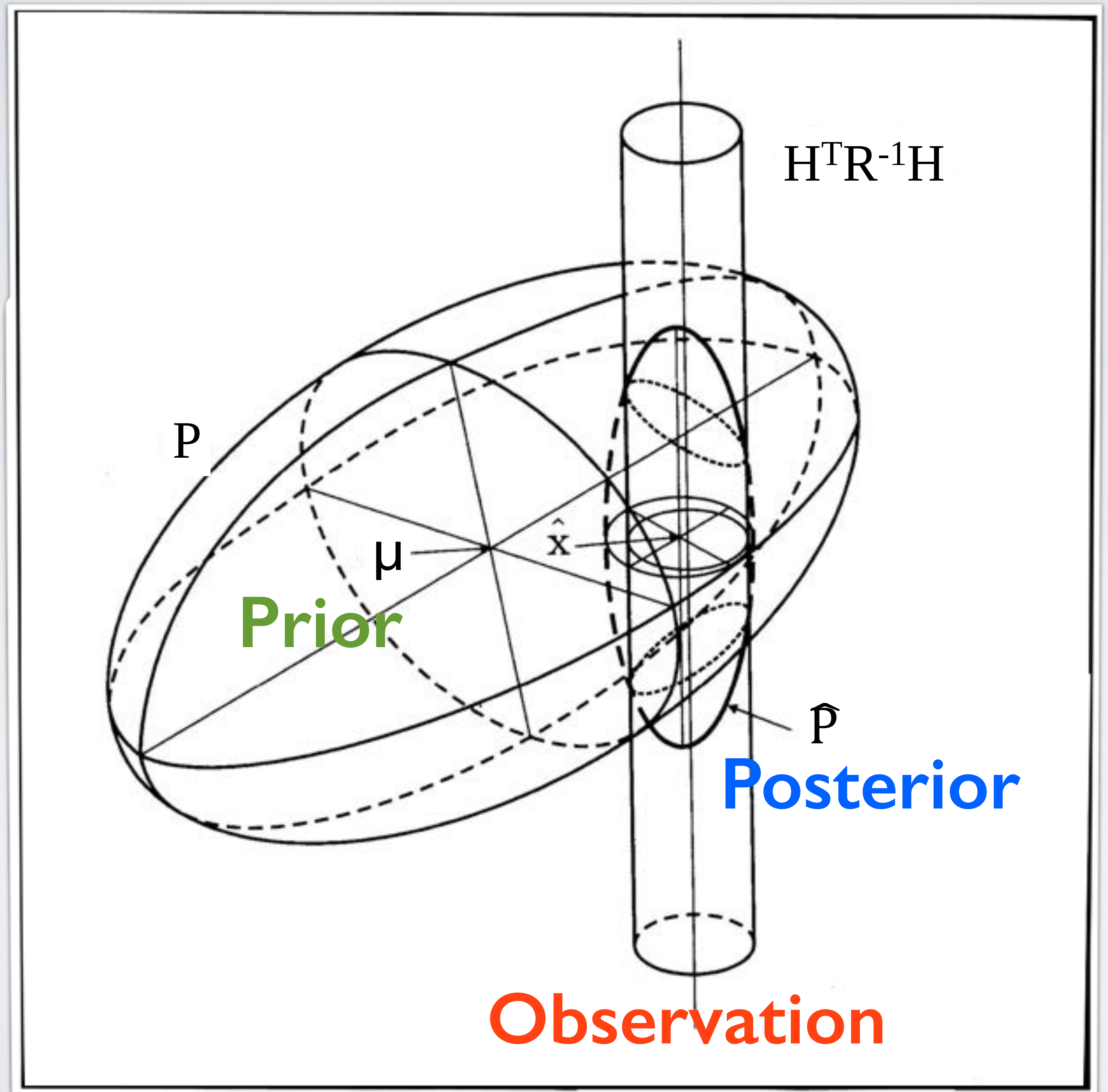
Eq. 4.10 of Brasseur and Jacob, 2016

General Problem: Given a set of observations and a model of some physical parameters, what does knowledge of the observations tell us about the model state?

Let $\mathbf{x}$ be n-vector of model state and $\mathbf{y}$ be p-vector of observations. The information we want to know is given by the conditional pdf, $p(\mathbf{x}|\mathbf{y})$.

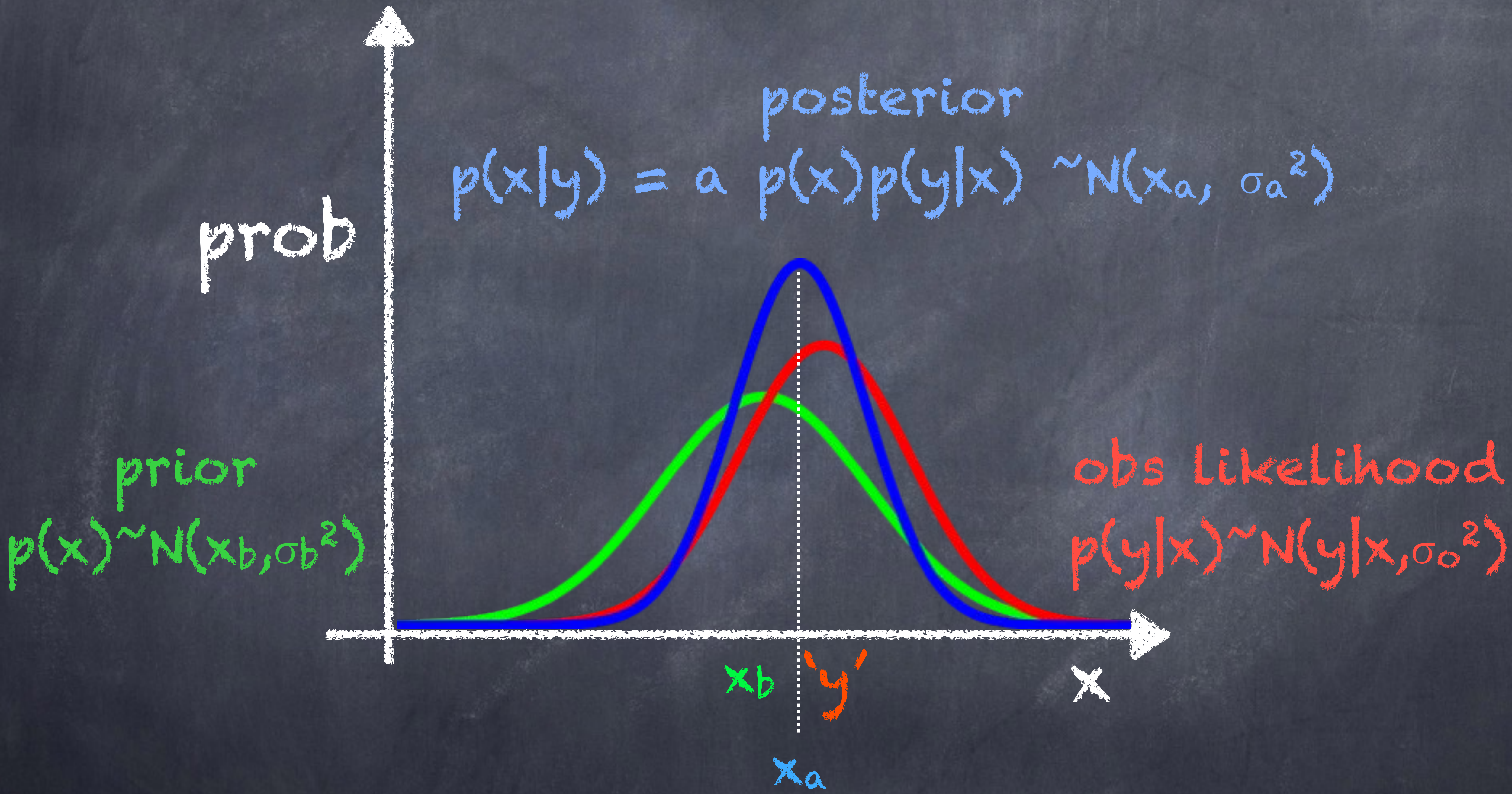
$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}$$

In practice, it is difficult to obtain $p(\mathbf{x}|\mathbf{y})$. Typically, we find attributes of $p(\mathbf{x}|\mathbf{y})$ which can help us estimate $\mathbf{x}$.

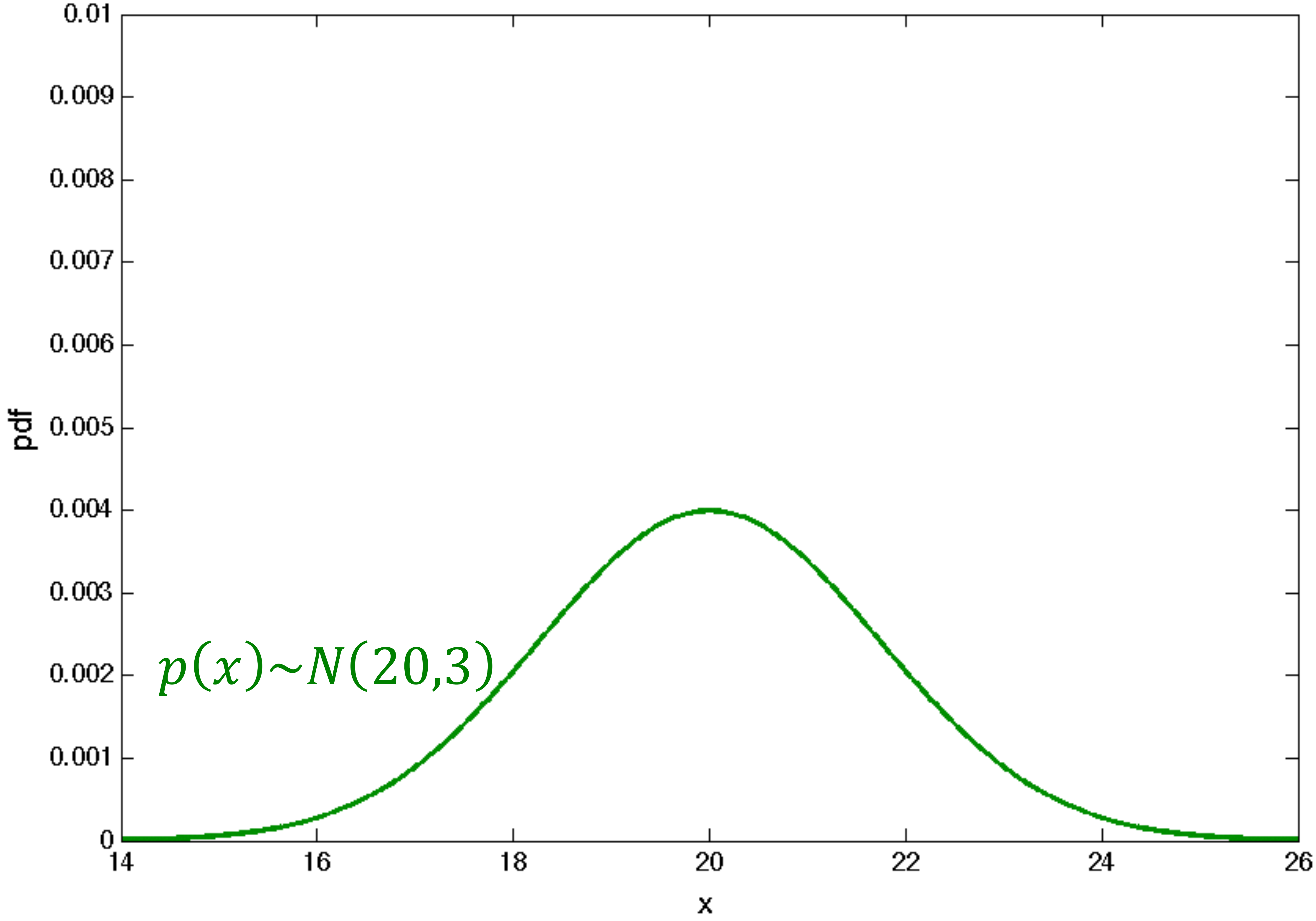


H^T R^-1 H
P
μ
x̂
Prior
P̂
Posterior
Observation

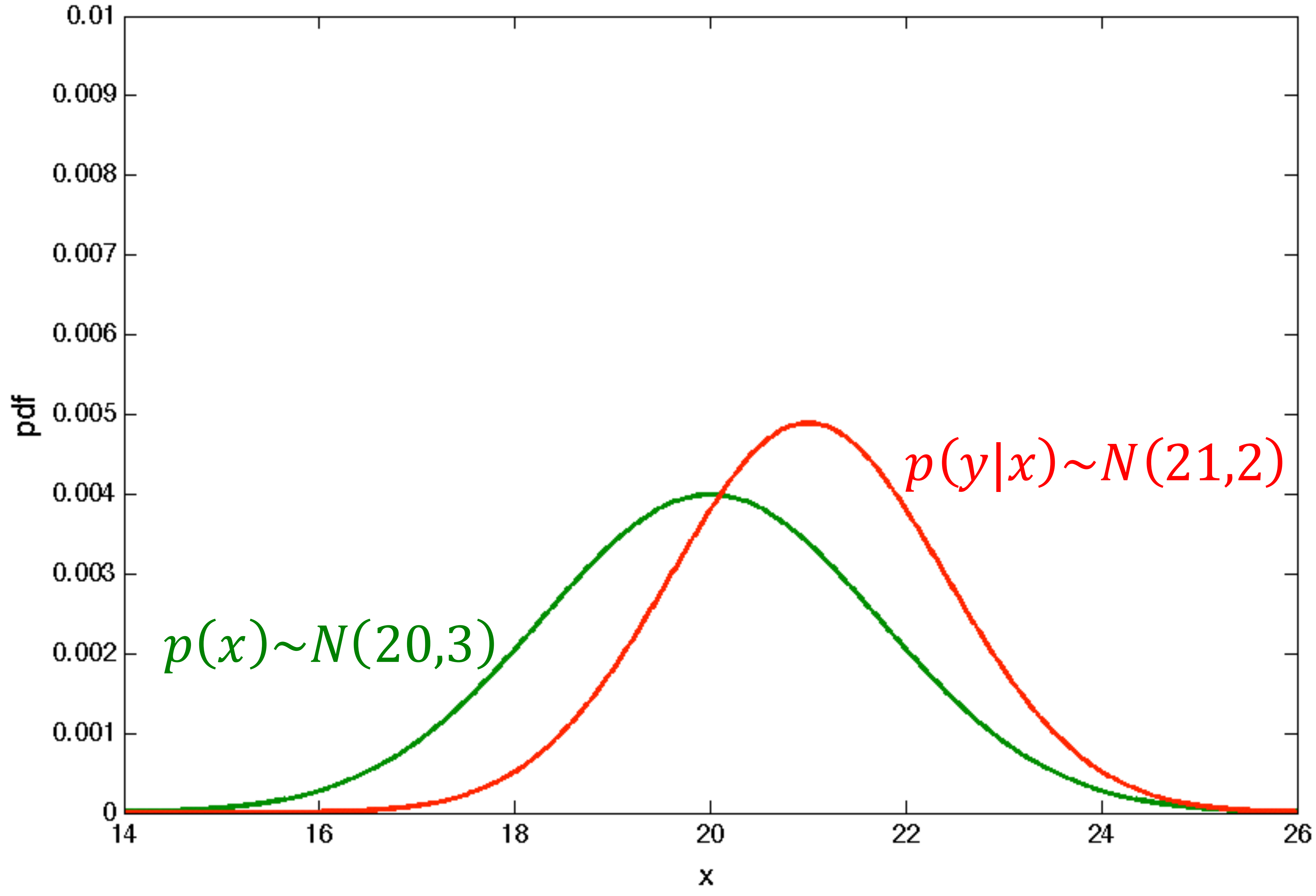
bayesian inference

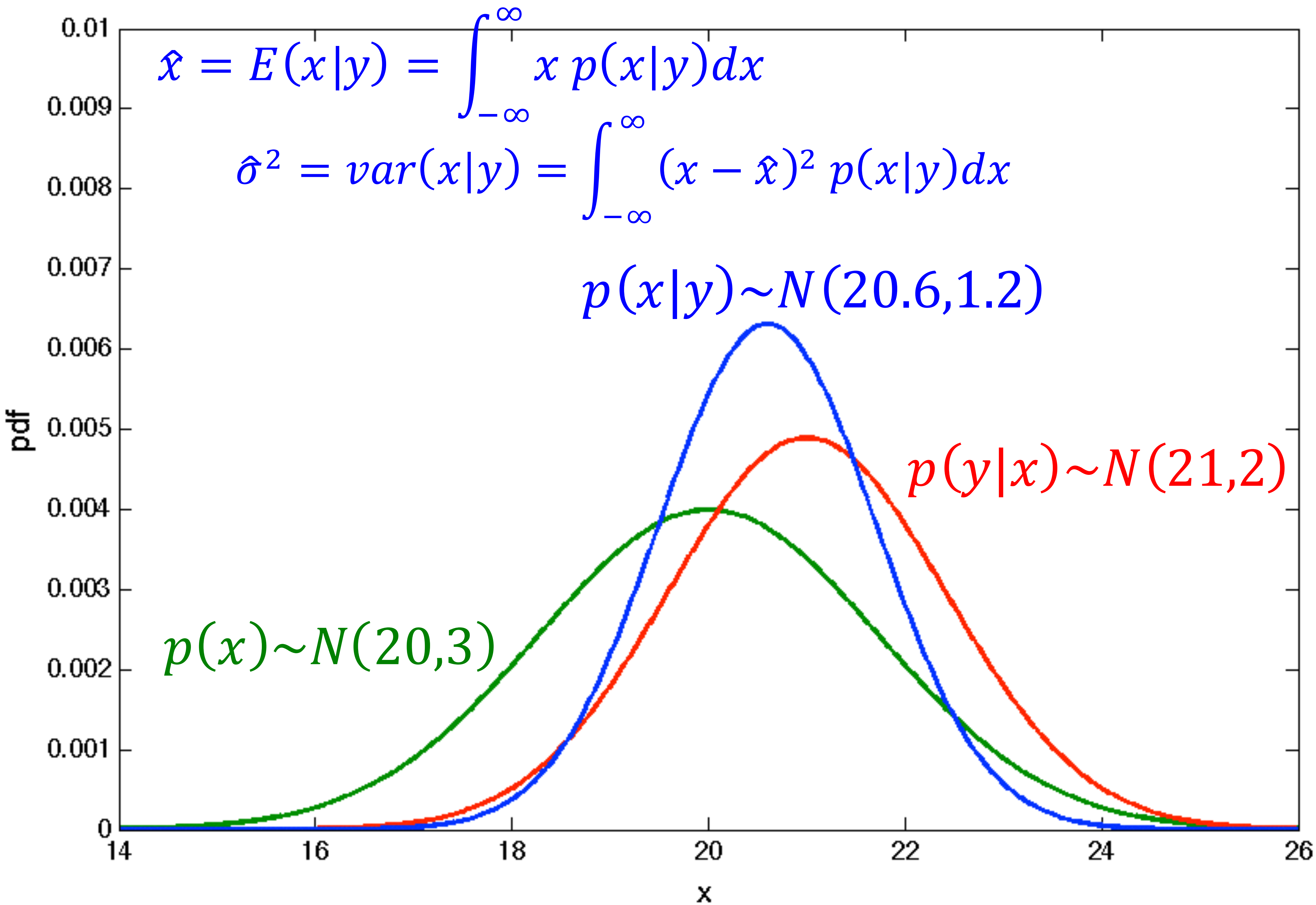


maximize $p(x|y)$ to find x_a



0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
pdf
14
16
18
20
22
24
26
x
$p(x) \sim N(20,3)$


0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
pdf
14
16
18
20
22
24
26
x
$p(y|x) \sim N(21,2)$
$p(x) \sim N(20,3)$


$\hat{x} = E(x|y) = \int_{-\infty}^{\infty} x\, p(x|y)dx$
$\hat{\sigma}^2 = var(x|y) = \int_{-\infty}^{\infty} (x-\hat{x})^2\, p(x|y)dx$
$p(x|y) \sim N(20.6, 1.2)$
$p(y|x) \sim N(21,2)$
$p(x) \sim N(20,3)$
pdf
x
0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
14
16
18
20
22
24
26

Let $\mathbf{x}$ be n-vector of model state and $\mathbf{y}$ be p-vector of observations with error $\mathbf{v}$. We assume that:

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{P})$$

$$\mathbf{y} = \mathbf{Hx} + \boldsymbol{v}$$

$$\mathbf{v} \sim N(\mathbf{0}, \boldsymbol{R})$$

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

$$p(\mathbf{x}) = \frac{1}{2\pi^{n/2}|\boldsymbol{P}|^{1/2}} exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{P}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{2\pi^{m/2}|\boldsymbol{R}|^{1/2}} exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{Hx})^T \boldsymbol{R}^{-1}(\mathbf{y}-\mathbf{Hx})\right)$$

$$p(\mathbf{y}) = \frac{1}{2\pi^{m/2}|\mathbf{H}\boldsymbol{P}\mathbf{H}^{\boldsymbol{T}} + \boldsymbol{R}|^{1/2}} exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{H}\boldsymbol{\mu})^T (\mathbf{H}\boldsymbol{P}\mathbf{H}^{\boldsymbol{T}} + \boldsymbol{R})^{-1}(\mathbf{y}-\mathbf{H}\boldsymbol{\mu})\right)$$

$$\mathrm{p}(\mathbf{x}|\mathbf{y}) = \frac{\mathrm{p}(\mathbf{y}|\mathbf{x})\mathrm{p}(\mathbf{x})}{\mathrm{p}(\mathbf{y})} = \frac{N(\mathbf{Hx}, \boldsymbol{R})N(\boldsymbol{\mu}, \boldsymbol{P})}{N(\mathbf{H}\boldsymbol{\mu}, \mathbf{H}\boldsymbol{P}\mathbf{H}^{T} + \boldsymbol{R})} = N(\hat{\mathbf{x}}, \boldsymbol{P}_{\tilde{x}})$$

$$p(\mathbf{x}|\mathbf{y}) = \frac{|\mathbf{H}\boldsymbol{P}\mathbf{H}^{T} + \boldsymbol{R}|^{1/2}}{2\pi^{n/2}|\boldsymbol{P}|^{1/2}|\boldsymbol{R}|^{1/2}} exp\left(-\frac{1}{2}J\right)$$

where

$$J = (\mathbf{x} - \boldsymbol{\mu})^{T}\boldsymbol{P}^{-1}(\mathbf{x} - \boldsymbol{\mu}) + (\mathbf{y} - \mathbf{Hx})^{T}\boldsymbol{R}^{-1}(\mathbf{y} - \mathbf{Hx}) - (\mathbf{y} - \mathbf{H}\boldsymbol{\mu})^{T}(\mathbf{H}\boldsymbol{P}\mathbf{H}^{T} + \boldsymbol{R})^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu})$$

we also know,

$$p(\mathbf{x}|\mathbf{y}) = \frac{|\mathbf{H}\boldsymbol{P}\mathbf{H}^{T} + \boldsymbol{R}|^{1/2}}{2\pi^{n/2}|\boldsymbol{P}|^{1/2}|\boldsymbol{R}|^{1/2}} exp\left(-\frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^{T}\boldsymbol{P}_{\tilde{x}}^{-1}(\mathbf{x} - \hat{\mathbf{x}})\right)$$

And so using completing squares, we arrive at a solution of our estimator:

$$\hat{\mathbf{x}} = E(\mathbf{x}|\mathbf{y}) = \boldsymbol{P}_{\tilde{\mathbf{x}}}(\mathbf{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\mathbf{y} + \boldsymbol{P}^{-1}\boldsymbol{\mu})$$

$$\boldsymbol{P}_{\tilde{\mathbf{x}}} = E[(\mathbf{x} - \hat{\mathbf{x}})^{2}] = (\boldsymbol{P}^{-1} + \mathbf{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\mathbf{H})^{-1}$$

Note the similarity of these estimates to our estimates using the least squares approach

$$\hat{\mathbf{x}} = E(\mathbf{x}|\mathbf{y}) = \boldsymbol{P}_{\tilde{\mathbf{x}}}(\mathbf{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\mathbf{y} + \boldsymbol{P}^{-1}\boldsymbol{\mu})$$

$$\boldsymbol{P}_{\tilde{\mathbf{x}}} = E[(\mathbf{x} - \hat{\mathbf{x}})^2] = (\boldsymbol{P}^{-1} + \mathbf{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\mathbf{H})^{-1}$$

Toy Example 1:

$$T_a = \sigma_a^2\left(\frac{1}{\sigma_o^2}T_o + \frac{1}{\sigma_b^2}T_b\right), \quad \sigma_a^2 = \left(\frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}\right)^{-1}$$

Toy Example 2:

$$T_a = \sigma_a^2\left(\frac{H}{\sigma_o^2}T_o + \frac{1}{\sigma_b^2}T_b\right), \quad \sigma_a^2 = \left(\frac{H^2}{\sigma_o^2} + \frac{1}{\sigma_b^2}\right)^{-1}$$

This is also similar to:

$$\hat{\mathbf{x}} = E(\mathbf{x}|\mathbf{y}) = \boldsymbol{\mu} + \mathbf{P}\mathbf{H}^{\mathbf{T}}(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathbf{T}} + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}) = \boldsymbol{\mu} + \mathbf{K}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu})$$

$$\mathbf{P}_{\tilde{\mathbf{x}}} = E\left[\left(\mathbf{x} - E(\mathbf{x}|\mathbf{y})\right)^{\mathbf{2}}\right] = \mathbf{P} - \mathbf{P}\mathbf{H}^{\mathbf{T}}(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathbf{T}} + \mathbf{R})^{-1}\mathbf{H}\mathbf{P} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}$$

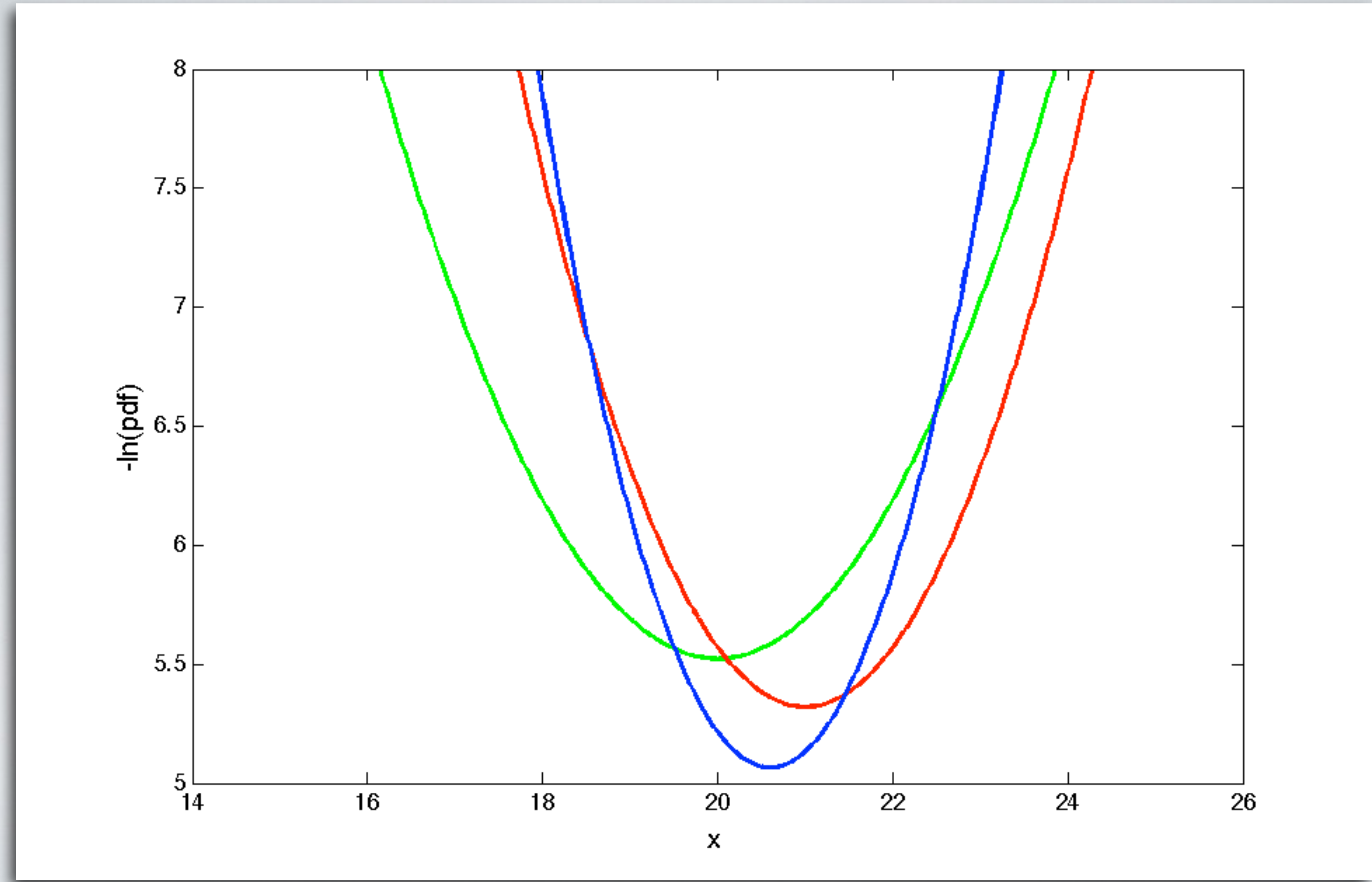
Toy Example 1:

$$T_a = T_b + W(T_o - T_b), \quad \sigma_a^2 = (1 - W)\sigma_b^2, \quad W = \frac{\sigma_b^2}{(\sigma_o^2 + \sigma_b^2)}$$

Toy Example 2:

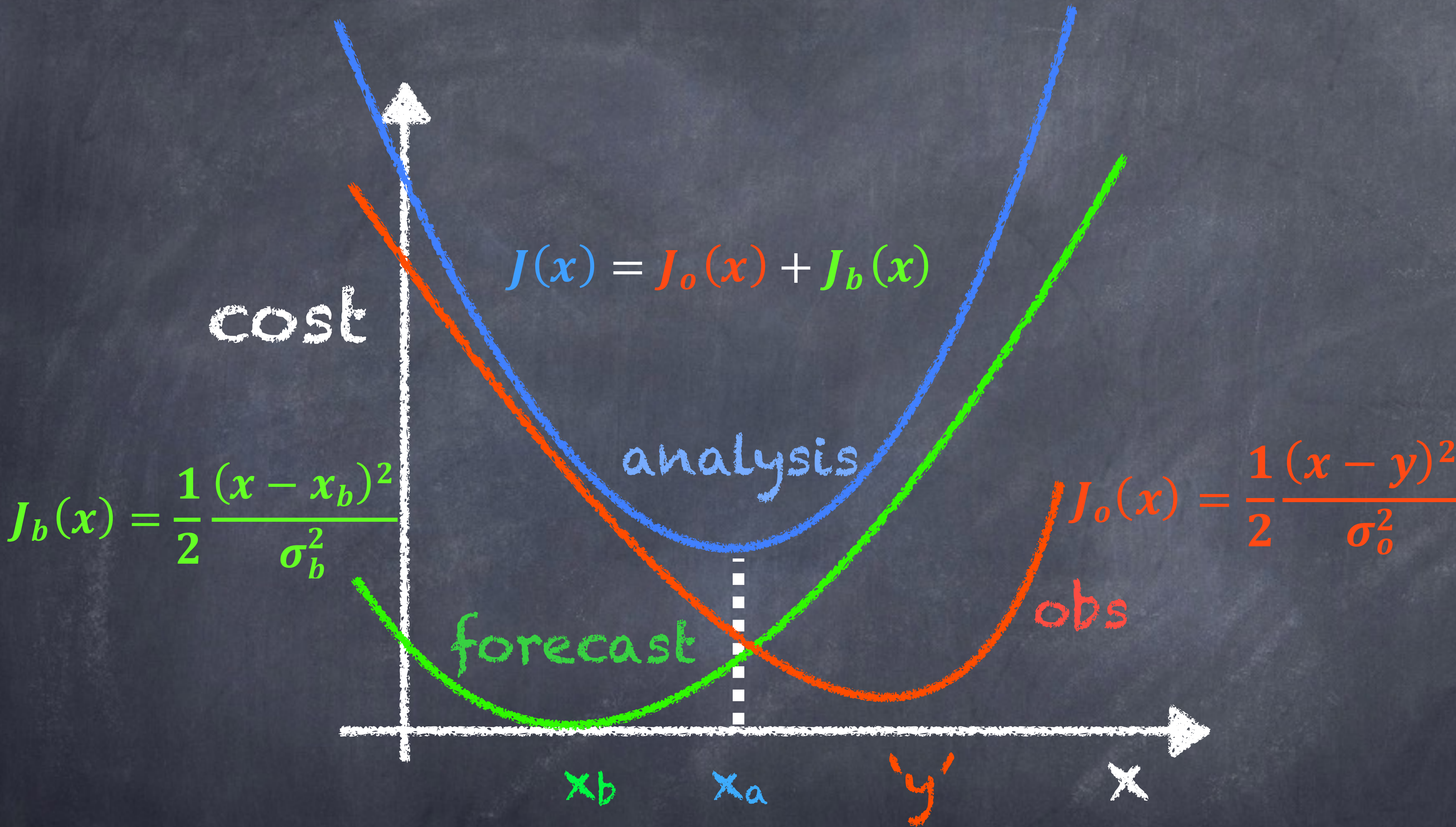
$$T_a = T_b + K(T_o - h(T_b)), \quad \sigma_a^2 = (1 - KH)\sigma_b^2 \quad K = \frac{H\sigma_b^2}{(\sigma_o^2 + H^2\sigma_b^2)}$$

equivalent to finding the minimum of -ln{p(*x*|*y*)}



$$J = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{P}^{-1} (\mathbf{x} - \boldsymbol{\mu}) + (\mathbf{y} - \mathbf{H}\mathbf{x})^T \boldsymbol{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}) - (\mathbf{y} - \mathbf{H}\boldsymbol{\mu})^T (\mathbf{H}\boldsymbol{P}\mathbf{H}^{\boldsymbol{T}} + \boldsymbol{R})^{-1} (\mathbf{y} - \mathbf{H}\boldsymbol{\mu})$$

variational method



minimize J(x) to find x_a

This is similar to what we did earlier (variational approach):

$$\mathbf{J} = (\mathbf{x}-\boldsymbol{\mu})^{\mathbf{T}}\mathbf{P}^{-1}(\mathbf{x}-\boldsymbol{\mu}) + (\mathbf{y}-\mathbf{H}\mathbf{x})^{\mathbf{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x}) + (\mathbf{y}-\mathbf{H}\boldsymbol{\mu})^{\mathbf{T}}(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathbf{T}}+\mathbf{R})^{-1}(\mathbf{y}-\mathbf{H}\boldsymbol{\mu})$$

We find our estimate by minimizing the cost function and equate to zero

$$\frac{\partial \mathbf{J}}{\partial \mathbf{x}} = \mathbf{P}^{-1}(\mathbf{x}-\boldsymbol{\mu}) - \mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x}) = \mathbf{0}$$

$$(\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})\,\mathbf{x} - (\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{y} + \mathbf{P}^{-1}\boldsymbol{\mu}) = \mathbf{0}$$

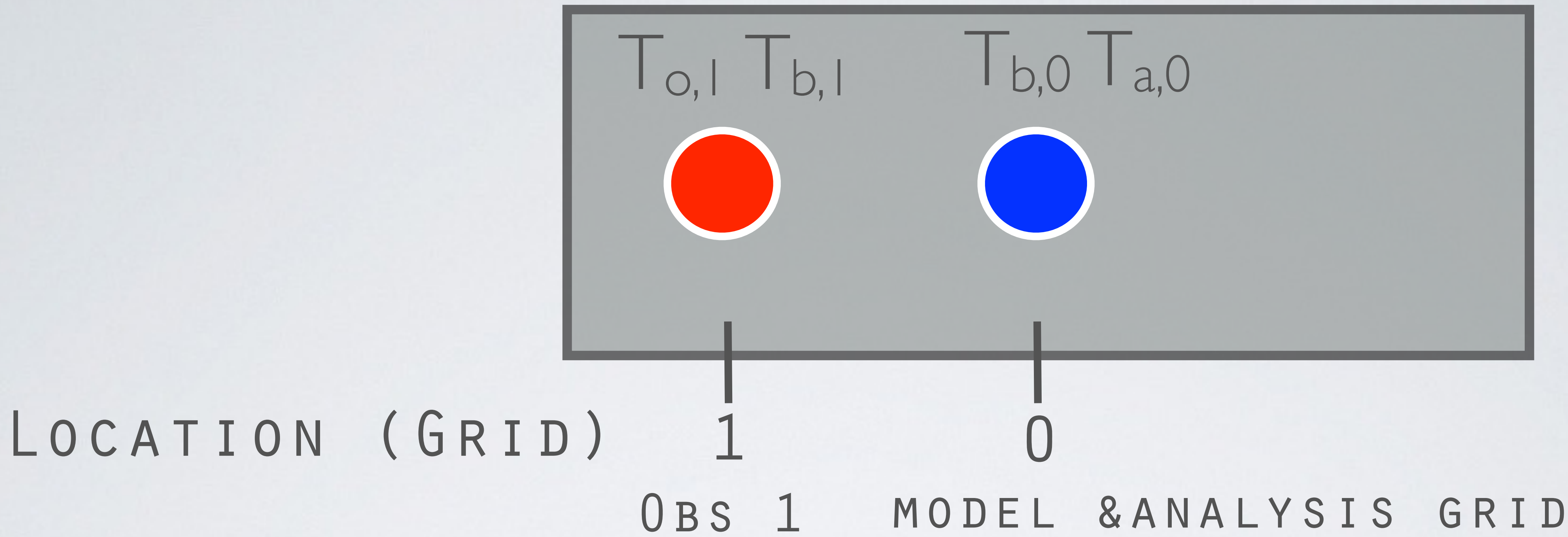
$$\hat{\mathbf{x}} = (\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})^{-1}(\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{y} + \mathbf{P}^{-1}\boldsymbol{\mu})$$

We can also approximate the error covariance of the estimate by taking the Hessian or second partial derivative:

$$\frac{\partial^2 \mathbf{J}}{\partial \mathbf{x}^2} = \mathbf{P}^{-1} + \mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{H}$$

$$\mathbf{P}_{\tilde{\mathbf{x}}} = (\mathbf{P}^{-1} + \mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}$$

This is also similar to our one obs + 2 model guesses example:



$$[T_{o,1}] = [0 \quad 1]\begin{bmatrix} T_{t,0} \\ T_{t,1} \end{bmatrix} + [e_{o,1}] \quad \text{or} \quad [y_1] = [0 \quad 1]\begin{bmatrix} x_0 \\ x_1 \end{bmatrix} + [\nu_1]$$

$$\begin{bmatrix} T_{b,0} \\ T_{b,1} \end{bmatrix} = \begin{bmatrix} T_{t,0} \\ T_{t,1} \end{bmatrix} + \begin{bmatrix} e_{b,0} \\ e_{b,1} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix} + \begin{bmatrix} \tau_0 \\ \tau_1 \end{bmatrix}$$

more generally,

$$\mathbf{y} = \mathbf{Hx} + \boldsymbol{\nu}, \quad \mathbf{R} = E(\boldsymbol{\nu}\boldsymbol{\nu}^T) = [\sigma_o{}^2]$$

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\tau}, \quad \mathbf{P} = E(\boldsymbol{\tau}\boldsymbol{\tau}^T) = \begin{bmatrix} \sigma_{b,0}^2 & \rho_{0,1}\sigma_{b,0}\sigma_{b,1} \\ \rho_{1,0}\sigma_{b,1}\sigma_{b,0} & \sigma_{b,1}^2 \end{bmatrix} = \sigma_b^2 \begin{bmatrix} 1 & \rho_{0,1} \\ \rho_{0,1} & 1 \end{bmatrix}$$

Our estimates are given as:

$$\hat{\mathbf{x}} = \boldsymbol{\mu} + \mathbf{P}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu})$$

$$\begin{bmatrix} \hat{x}_0 \\ \hat{x}_1 \end{bmatrix} = \begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix} + \mathbf{P}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\left(y_1 - \begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix}\right)$$

where

$$\mathbf{P}\mathbf{H}^{\mathrm{T}} = \sigma_b^2 \begin{bmatrix} 1 & \rho_{0,1} \\ \rho_{0,1} & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \sigma_b^2 \begin{bmatrix} \rho_{0,1} \\ 1 \end{bmatrix}$$

$$\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R} = \begin{bmatrix} 0 & 1 \end{bmatrix}\sigma_b^2 \begin{bmatrix} \rho_{0,1} \\ 1 \end{bmatrix} + \sigma_o^2 = \sigma_b^2 + \sigma_o^2$$

$$\mathbf{P}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1} = \sigma_b^2 \begin{bmatrix} \rho_{0,1} \\ 1 \end{bmatrix}(\sigma_b^2 + \sigma_o^2)^{-1}$$

and so

$$\begin{bmatrix} \hat{x}_0 \\ \hat{x}_1 \end{bmatrix} = \begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix} + \sigma_b^2 \begin{bmatrix} \rho_{0,1} \\ 1 \end{bmatrix}(\sigma_b^2 + \sigma_o^2)^{-1}(y_1 - \mu_1)$$

$$\hat{x}_0 = \mu_0 + \frac{\rho_{0,1}\sigma_b^2}{(\sigma_b^2 + \sigma_o^2)}(y_1 - \mu_1) \quad \text{and} \quad \hat{x}_1 = \mu_1 + \frac{\sigma_b^2}{(\sigma_b^2 + \sigma_o^2)}(y_1 - \mu_1)$$

$$T_{a,0} = T_{b,0} + \frac{\rho_{0,1}\sigma_b^2}{(\sigma_b^2 + \sigma_o^2)}(T_{o,1} - T_{b,1}) \quad \text{and} \quad T_{a,1} = T_{b,1} + \frac{\sigma_b^2}{(\sigma_b^2 + \sigma_o^2)}(T_{o,1} - T_{b,1})$$

Again, this is similar to:

$$\hat{\mathbf{x}} = E(\mathbf{x}|\mathbf{y}) = \boldsymbol{\mu} + (\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-\mathbf{1}}\mathbf{H} + \mathbf{P}^{-\mathbf{1}})^{-\mathbf{1}}\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-\mathbf{1}}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}) = \boldsymbol{\mu} + \mathbf{K}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu})$$

Toy Example 1:

$$T_a = T_b + \left(\frac{\frac{1}{\sigma_o^2}}{\left(\frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}\right)} \right)(T_o - T_b)$$

Toy Example 2:

$$T_a = T_b + \left(\frac{\frac{H}{\sigma_o^2}}{\left(\frac{H^2}{\sigma_o^2} + \frac{1}{\sigma_b^2}\right)} \right)(T_o - h(T_b))$$

The corresponding error covariance of our estimates is given as:

$$\mathbf{P}_{\tilde{\mathbf{x}}} = \mathbf{P} - \mathbf{P}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{H}\mathbf{P} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}$$

$$\mathbf{P}_{\tilde{\mathbf{x}}} = \begin{bmatrix} \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \end{bmatrix} = \sigma_b^2 \begin{bmatrix} 1 & \rho_{0,1} \\ \rho_{0,1} & 1 \end{bmatrix} - \sigma_b^2 \begin{bmatrix} \rho_{0,1} \\ 1 \end{bmatrix} (\sigma_b^2 + \sigma_o^2)^{-1} [0 \quad 1] \sigma_b^2 \begin{bmatrix} 1 & \rho_{0,1} \\ \rho_{0,1} & 1 \end{bmatrix}$$

$$\mathbf{P}_{\tilde{\mathbf{x}}} = \begin{bmatrix} \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \end{bmatrix} = \sigma_b^2 \begin{bmatrix} 1 & \rho_{0,1} \\ \rho_{0,1} & 1 \end{bmatrix} - \sigma_b^2 (\sigma_b^2 + \sigma_o^2)^{-1} \sigma_b^2 \begin{bmatrix} \rho_{0,1} \\ 1 \end{bmatrix} [\rho_{0,1} \quad 1]$$

$$\mathbf{P}_{\tilde{\mathbf{x}}} = \begin{bmatrix} \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \end{bmatrix} = \sigma_b^2 \left(\begin{bmatrix} 1 & \rho_{0,1} \\ \rho_{0,1} & 1 \end{bmatrix} - \frac{\sigma_b^2}{(\sigma_b^2 + \sigma_o^2)} \begin{bmatrix} {\rho_{0,1}}^2 & \rho_{0,1} \\ \rho_{0,1} & 1 \end{bmatrix} \right)$$

And so the error variance of our estimates are:

$$\hat{\sigma}_0^2 = \sigma_b^2 \left(1 - \frac{{\rho_{0,1}}^2 \sigma_b^2}{(\sigma_b^2 + \sigma_o^2)} \right) \quad \text{and} \quad \hat{\sigma}_1^2 = \sigma_b^2 \left(1 - \frac{\sigma_b^2}{(\sigma_b^2 + \sigma_o^2)} \right)$$

$$\sigma_{a,0}^2 = \sigma_b^2 \left(1 - \frac{{\rho_{0,1}}^2 \sigma_b^2}{(\sigma_b^2 + \sigma_o^2)} \right) \quad \text{and} \quad \sigma_{a,1}^2 = \sigma_b^2 \left(1 - \frac{\sigma_b^2}{(\sigma_b^2 + \sigma_o^2)} \right)$$

Let’s do a Gallery Walk.

Bayesed and Confused

- Group yourselves into 1 or more (need 6 groups)
- Groups 1 to 3 will work on the left side (4 to 6 on the right side)
- Your task: Give your best description (or better yet identify) what is the picture in the poster all about. write it down (1-3 minutes)
- Indicate the level of uncertainty of your description/ identification) by annotating with stars

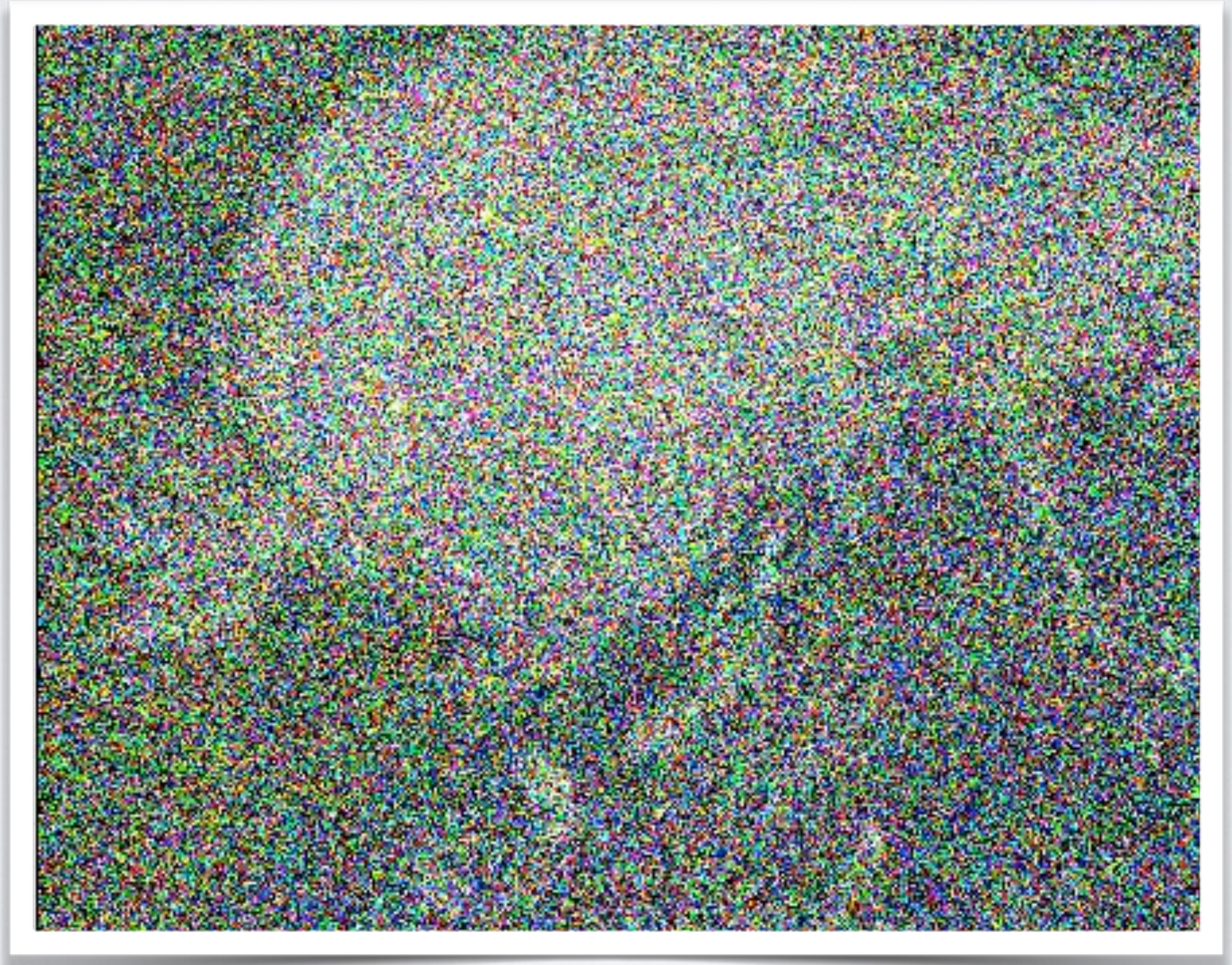
  1 star = Have no idea

  2 star = Hmm, looks familiar

  3 star = Gotcha
- Go to another poster and do the same (but now taking into account the added information from the previous group)
- After you have gone through all 3 posters, assign a reporter from your group. He/She will report your description/ identification of the picture.

(2) Incomplete Guess, Noisy (large errors) & Complete Observations

____________ captured this stunning visible image of ___________ *at 8:32 a.m. EDT, just 28 minutes before Irene's landfall* in New York City.

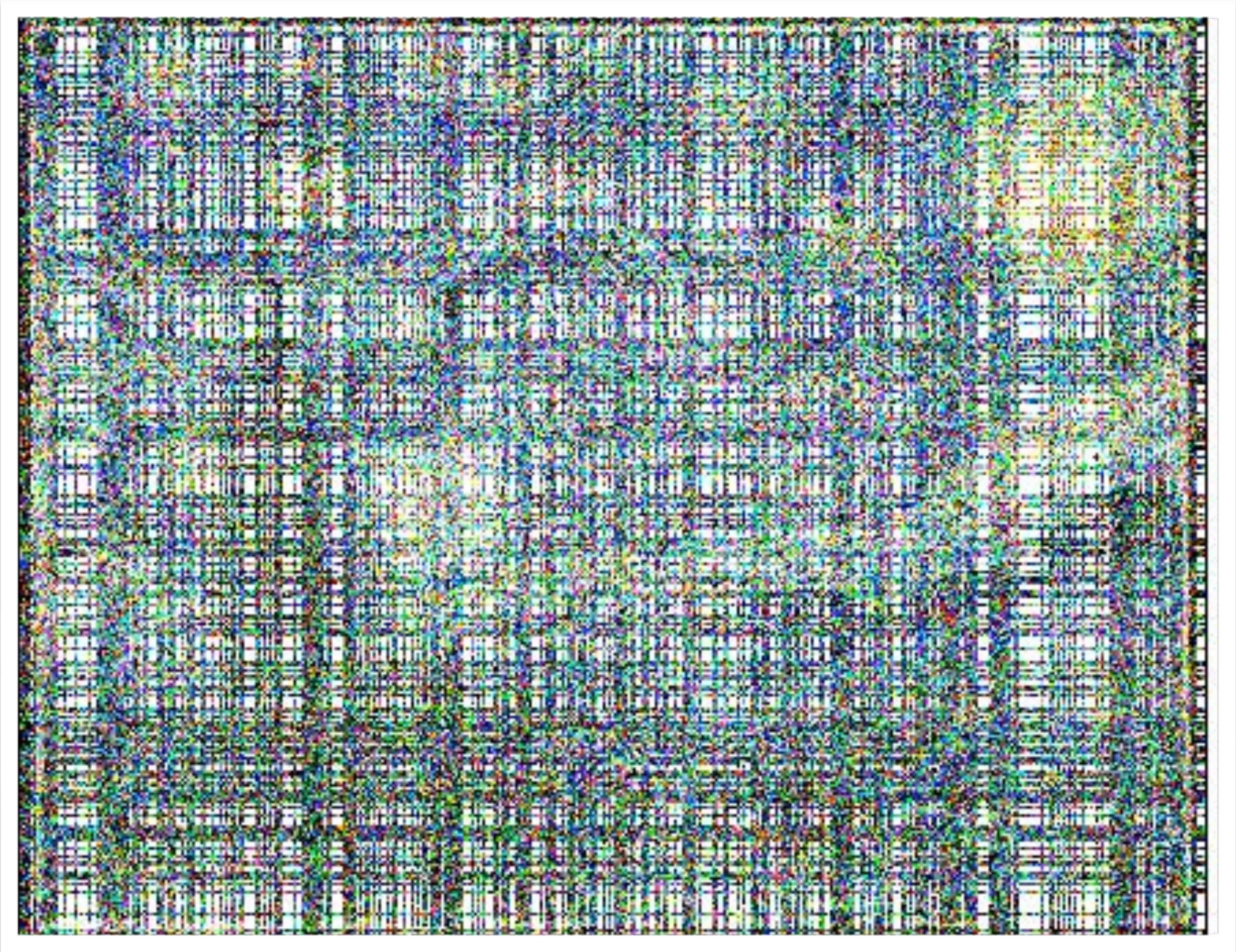


(2) Incomplete Guess, Noisy (large errors) & Complete Observations

The GOES-13 satellite captured this stunning visible image of **Hurricane Irene** *at 8:32 a.m. EDT, just 28 minutes before Irene*'s *landfall* in New York City.

http://www.nasa.gov/mission_pages/hurricanes/archives/2011/h2011_Irene.html

(3) Somewhat 'Accurate' & 'Complete' Guess, Noisy (large errors) & Somewhat Few Observations



(3) Somewhat 'Accurate' & 'Complete' Guess, Noisy (large errors) & Somewhat Few Observations

<u>The Starry Night</u> vibrates with rockets of burning yellow while planets gyrate like cartwheels. The hills quake and heave, yet the cosmic gold fireworks that swirl against the blue sky are somehow restful.

http://www.ibiblio.org/wm/paint/auth/gogh/starry-night/

(4) Wrong Guess, Noisy (low errors) & Few Observations

Nadal cruises to straight-set win at US Open

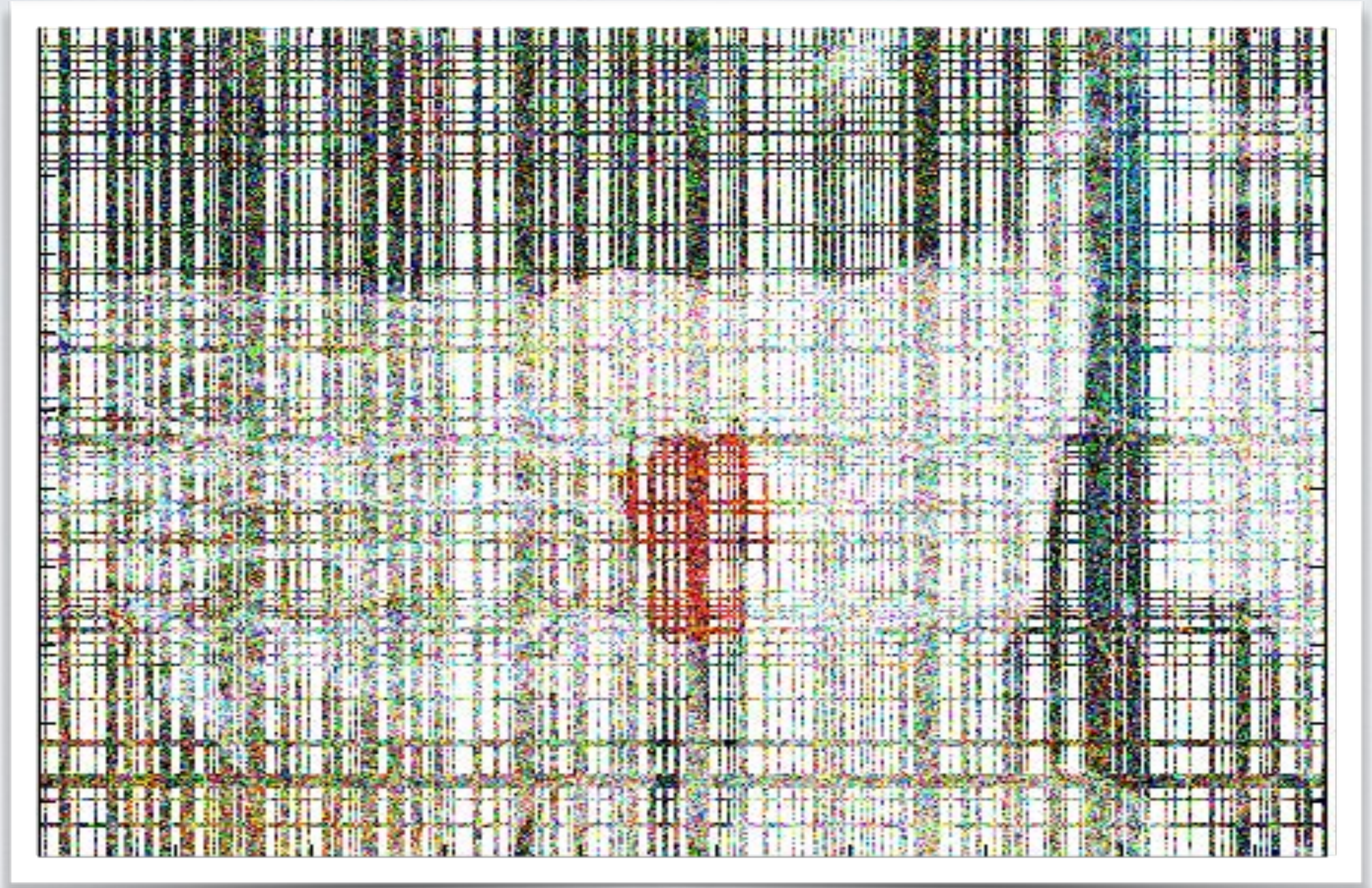
(4) Wrong Guess, Noisy (low errors) & Few Observations

Djokovic cruises to straight-set win at US Open

http://www.boston.com/sports/other-sports/tennis/2013/08/27/djokovic-cruises-straight-set-win-open/1hDa8MfxY2UOv2rATqI0XK/story.html

(5) Incomplete Guess, Noisy (low errors) & Few Observations

Images of the _____ flood. A woman near __________ Creek.

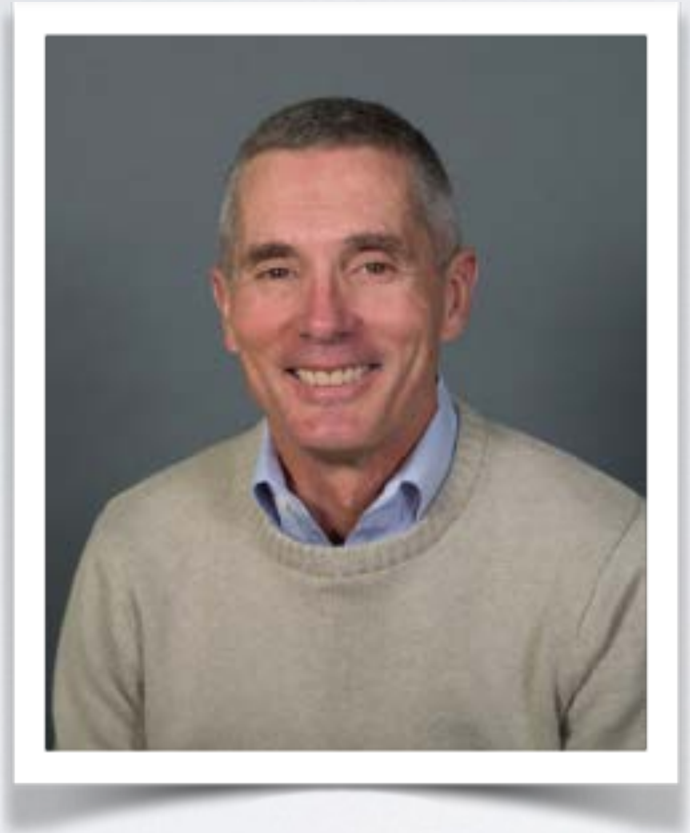


(5) Incomplete Guess, Noisy (low errors) & Few Observations

Images of the **Colorado** flood. A woman near __**Boulder**_ Creek.

(1) No Guess, Noisy (large errors) & Few Observations

(1) No Guess, Noisy (large errors) & Few Observations

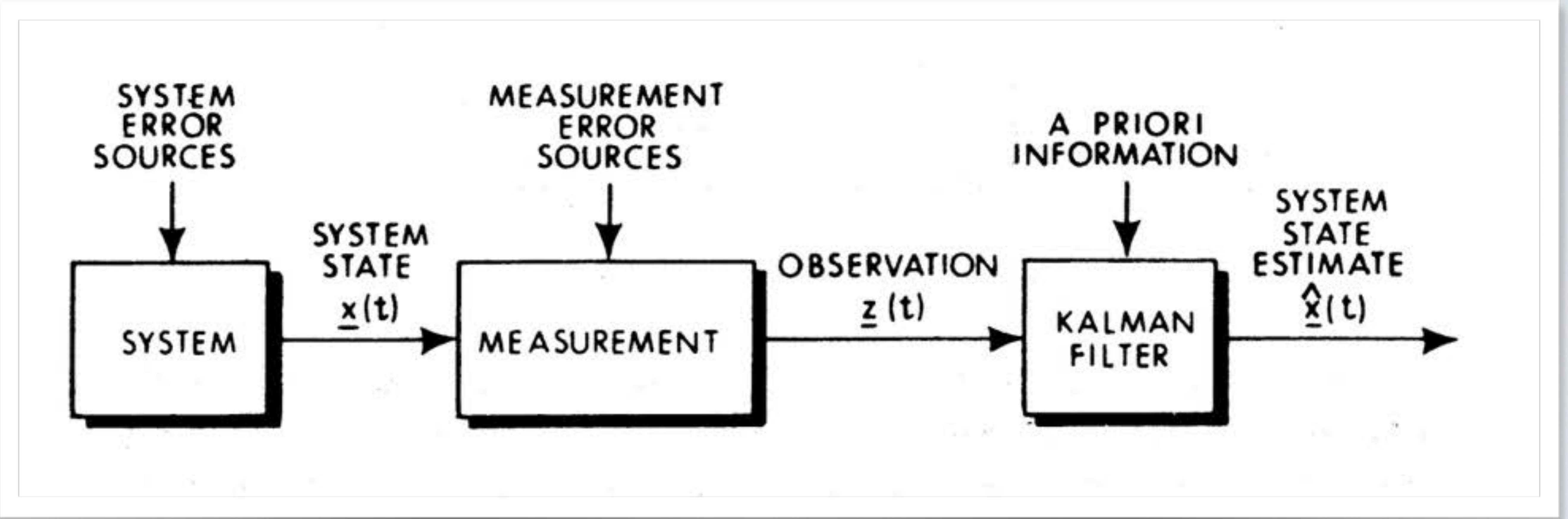


Arthur Mizzi
Project Scientist

http://www.acom.ucar.edu/cgi-bin/acd/pictureBoard.py

Ingredients of a Kalman Filter

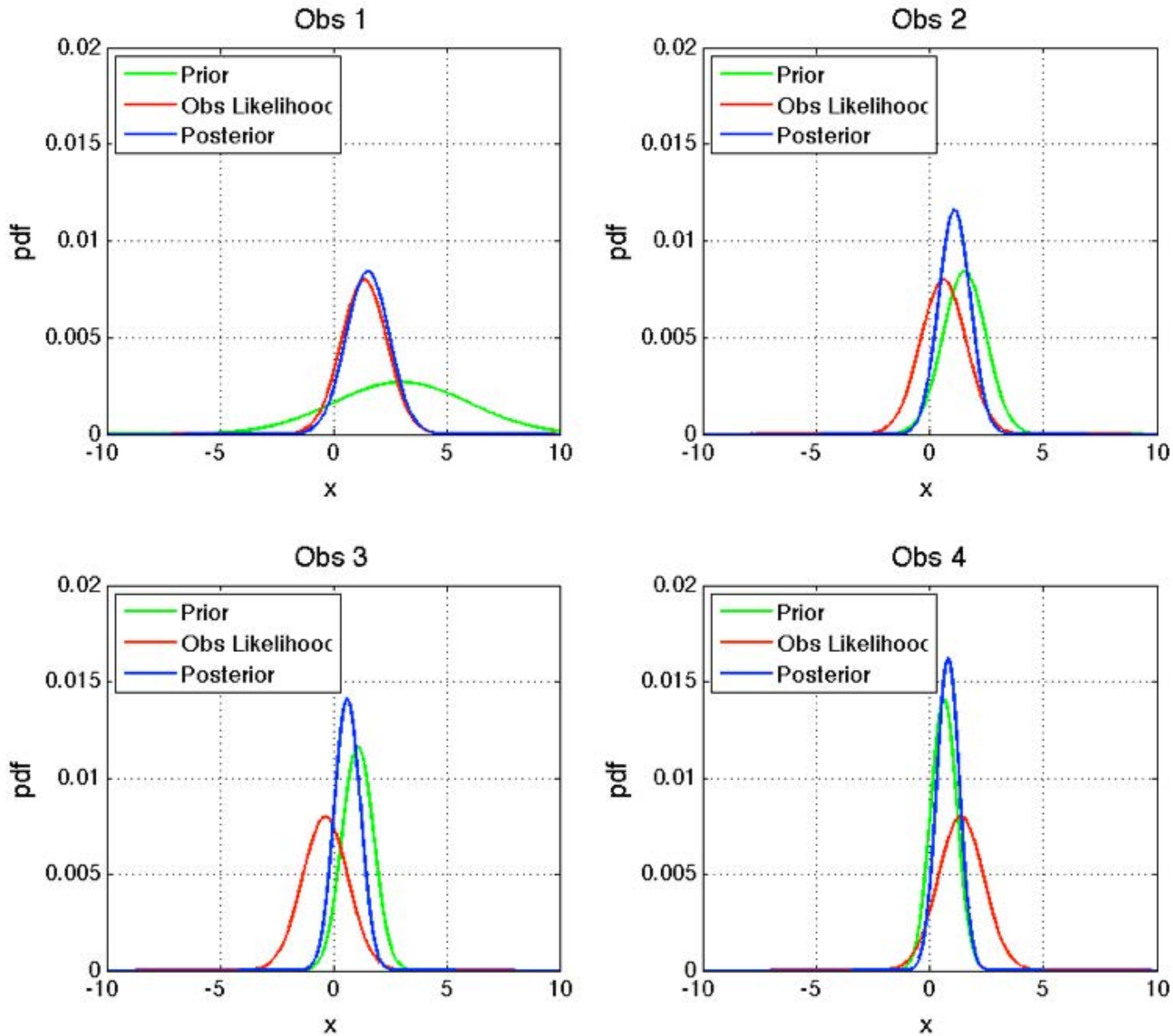
- **A discrete process model**
  - change in state over time
  - linear difference equation
- **A discrete measurement model**
  - relationship between state and measurement
  - linear function
- **Noise Characteristics**
  - process noise
  - measurement noise

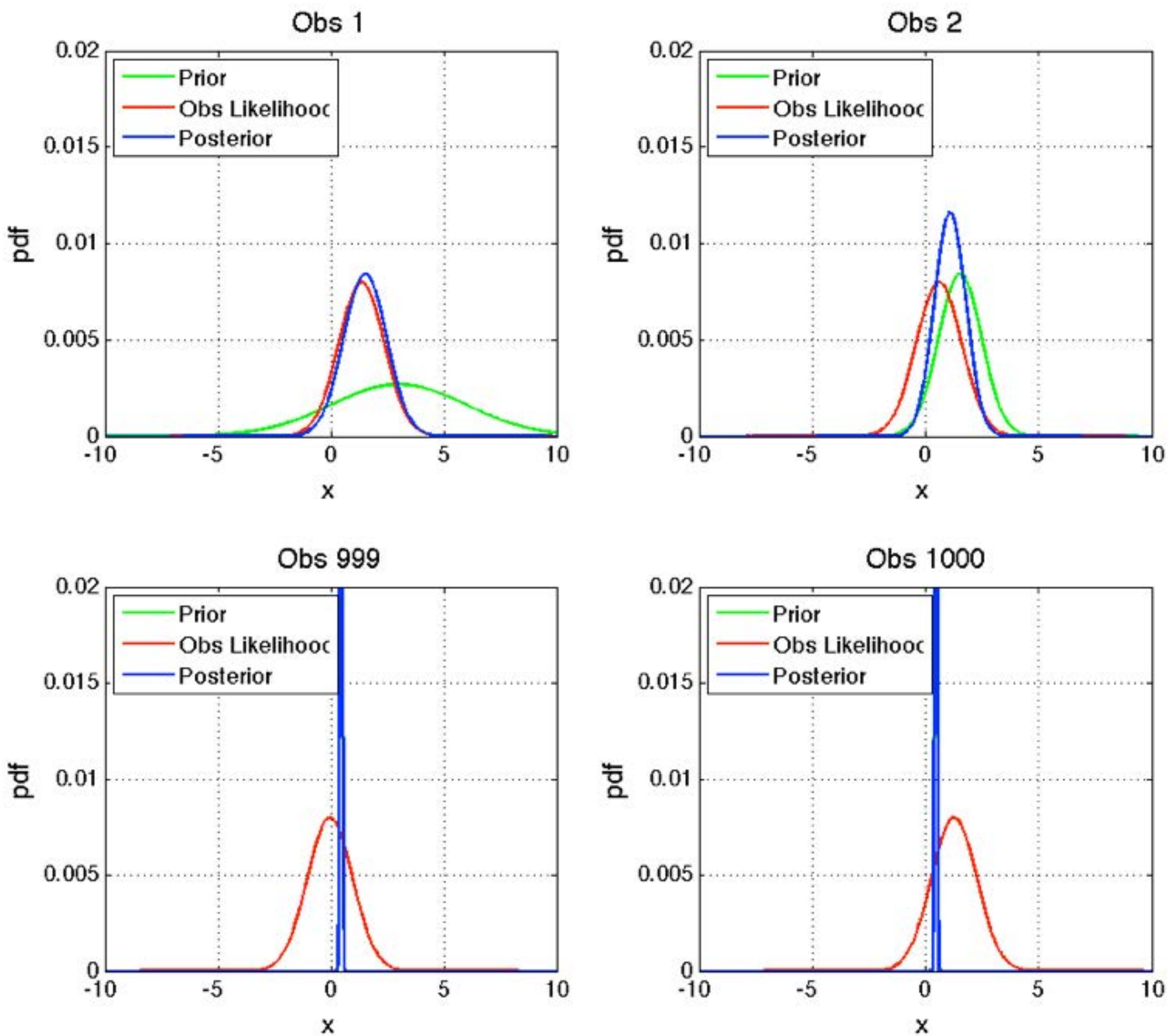


Gelb (1974)

$$\frac{\partial \rho_i}{\partial t} = \left[\frac{\partial \rho_i}{\partial t}\right]_{adv} + \left[\frac{\partial \rho_i}{\partial t}\right]_{mix} + \left[\frac{\partial \rho_i}{\partial t}\right]_{conv} + \left[\frac{\partial \rho_i}{\partial t}\right]_{scav} + \left[\frac{\partial \rho_i}{\partial t}\right]_{chem} + \left[\frac{\partial \rho_i}{\partial t}\right]_{em} + \left[\frac{\partial \rho_i}{\partial t}\right]_{dep}$$

Eq. 4.10 of Brasseur and Jacob, 2016


Obs 1
Prior
Obs Likelihooc
Posterior
pdf
x
Obs 2
Prior
Obs Likelihooc
Posterior
pdf
x
Obs 3
Prior
Obs Likelihooc
Posterior
pdf
x
Obs 4
Prior
Obs Likelihooc
Posterior
pdf
x



Obs 1
Obs 2
Obs 999
Obs 1000
Prior
Obs Likelihooc
Posterior
pdf
x
0.02
0.015
0.01
0.005
0
-10
-5
0
5
10

Kalman Filter in a Nutshell

Initial Guess

$$\mathbf{x}_0^f,\ \mathbf{P}_0^f$$

2. Compute Kalman Gain

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

1. Advance in Time

$$\mathbf{x}_{k+1}^f = \mathbf{M}_k \mathbf{x}_k^a$$

$$\mathbf{P}_{k+1}^f = \mathbf{M}_k \mathbf{P}_k^a \mathbf{M}_k^T + \mathbf{Q}_k$$

3. State Update

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k \left(\mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^f \right)$$

4. Error Covariance Update

$$\mathbf{P}_k^a = \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right) \mathbf{P}_k^f$$

Recall: Kalman Filter in a Nutshell

Initial Guess

$$\mathbf{x}_0^f,\ \mathbf{P}_0^f$$

2. Compute Kalman Gain

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

PREDICT

1. Advance in Time

$$\mathbf{x}_{k+1}^f = \mathbf{M}_k \mathbf{x}_k^a$$

$$\mathbf{P}_{k+1}^f = \mathbf{M}_k \mathbf{P}_k^a \mathbf{M}_k^T + \mathbf{Q}_k$$

CORRECT

3. State Update

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k \left(\mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^f\right)$$

4. Error Covariance Update

$$\mathbf{P}_k^a = \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right)\mathbf{P}_k^f$$

Information Filter in a Nutshell

Initial Guess

$$\mathbf{x}_0^f,\ \mathbf{P}_0^f$$

1. Advance in Time

$$\mathbf{x}_{k+1}^f = \mathbf{M}_k \mathbf{x}_k^a$$

$$\mathbf{P}_{k+1}^f = \mathbf{M}_k \mathbf{P}_k^a \mathbf{M}_k^T + \mathbf{Q}_k$$

2. Compute Kalman Gain

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

3. State Update

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k \left(\mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^f \right)$$

4. Error Covariance Update

$$\left(\mathbf{P}_k^a\right)^{-1} = \left(\mathbf{P}_k^f\right)^{-1} + \mathbf{H}_k^T \left(\mathbf{R}_k\right)^{-1} \mathbf{H}_k$$

Variational Data Assimilation

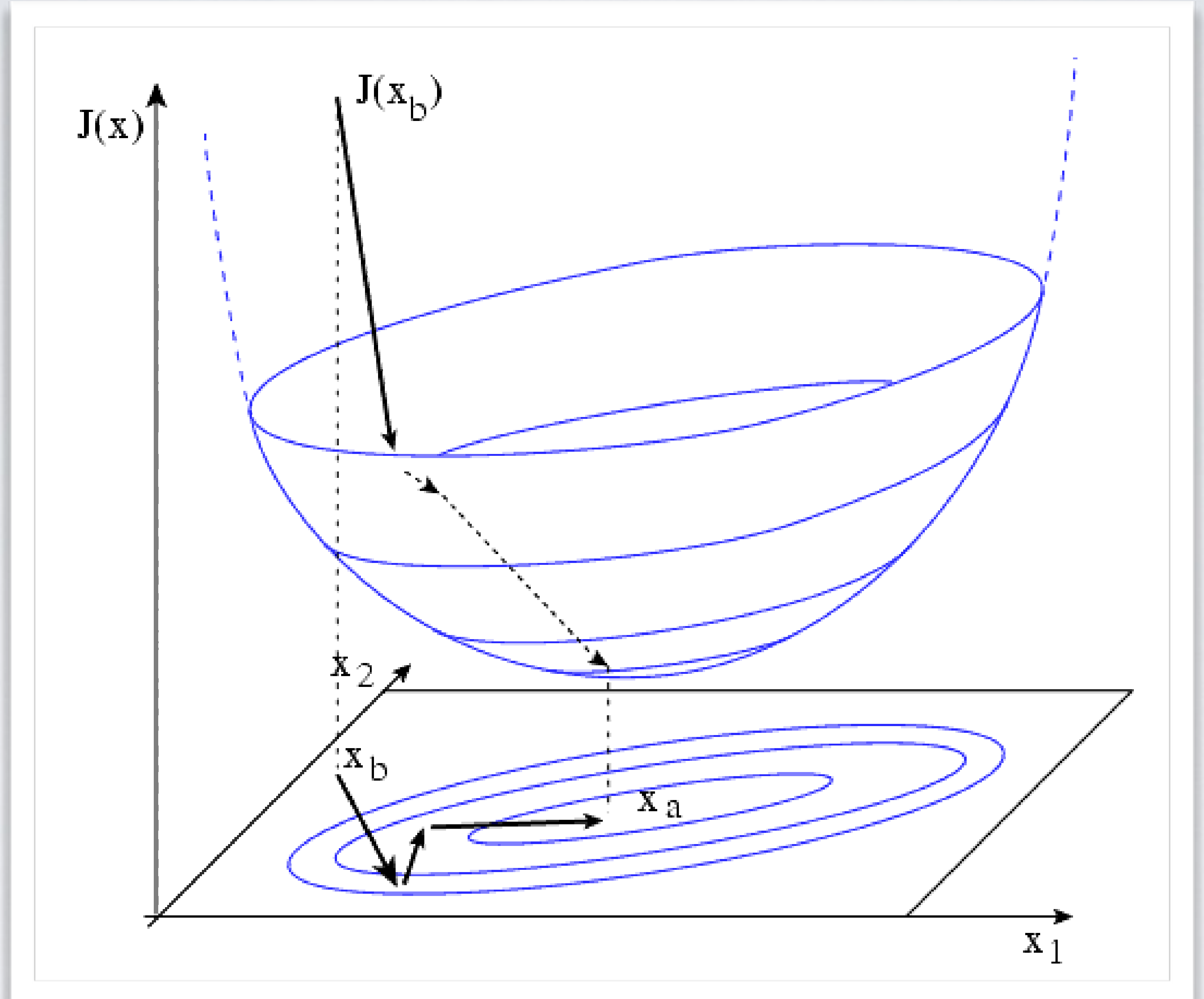
A class of assimilation algorithms in which the field to be estimated are explicitly determined as minimizers of a scalar function, called, objective or cost function, that measure the misfit to the available data.

We can construct an objective function of the form:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T(\mathbf{P}^b)^{-1}(\mathbf{x}-\mathbf{x}^b) + \frac{1}{2}(H(\mathbf{x})-\mathbf{y}^o)^T(\mathbf{R})^{-1}(H(\mathbf{x})-\mathbf{y}^o) = J_b + J_o$$

which measure the deviation of our state from the prior (background) information and the deviation from the observation. Our estimate of the state, $\mathbf{x}^a$ can be derived by minimizing the cost function, $\nabla_{\mathbf{x}} J(\mathbf{x}^a) = 0$

Graphically for n=2, the geometry of the minimization of the cost function term for the background state is:



The minimization works by performing several line-searches to move the control variable to areas where the cost-function is smaller, usually by looking at the local slope (the gradient) of the cost-function.

The objective function :

$$J(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right)^T \left(\mathbf{P}_0^b\right)^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^b\right) + \frac{1}{2}\sum_{k=0}^{K} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)^T \, (\mathbf{R}_k)^{-1} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)$$

Minimization of the cost function will define the initial condition of the model solution that fits the data most closely. Following Sasaki (1970), this is called strong constraint four-dimensional variational assimilation (**4D-Var**).

If we consider the model error, we have the following objective function to minimize:

$$J(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right)^T \left(\mathbf{P}_0^b\right)^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^b\right) + \frac{1}{2}\sum_{k=0}^{K} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)^T \, (\mathbf{R}_k)^{-1} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)$$
$$+ \frac{1}{2}\sum_{k=0}^{K-1} (\mathbf{x}_{k+1} - M_k(\mathbf{x}_k))^T \, (\mathbf{Q}_k)^{-1} (\mathbf{x}_{k+1} - M_k(\mathbf{x}_k))$$

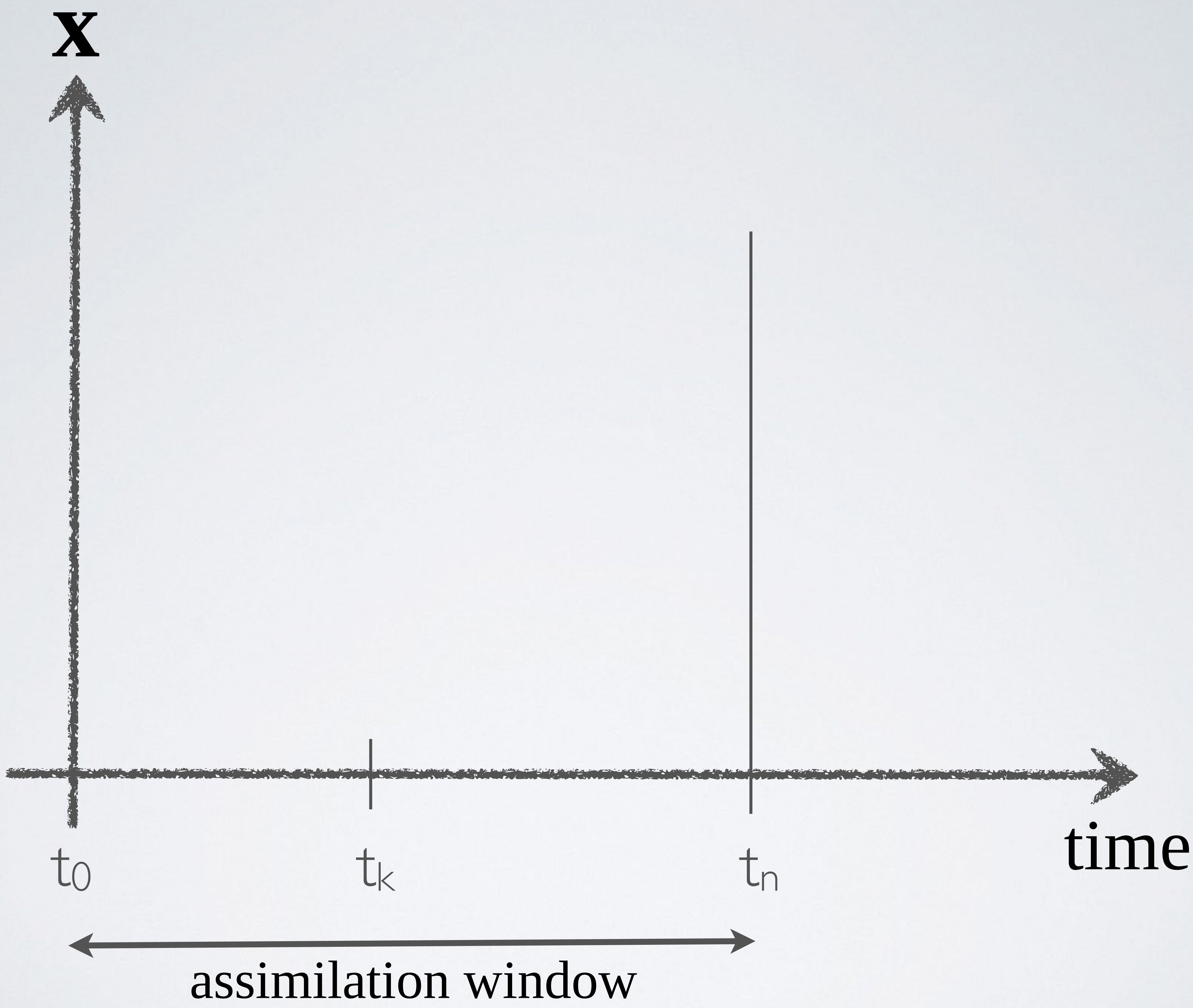
Minimizing this cost function where the model equations are present as noisy data to be fitted by the analysed fields like any other data is called weak constraint **4D-Var**.

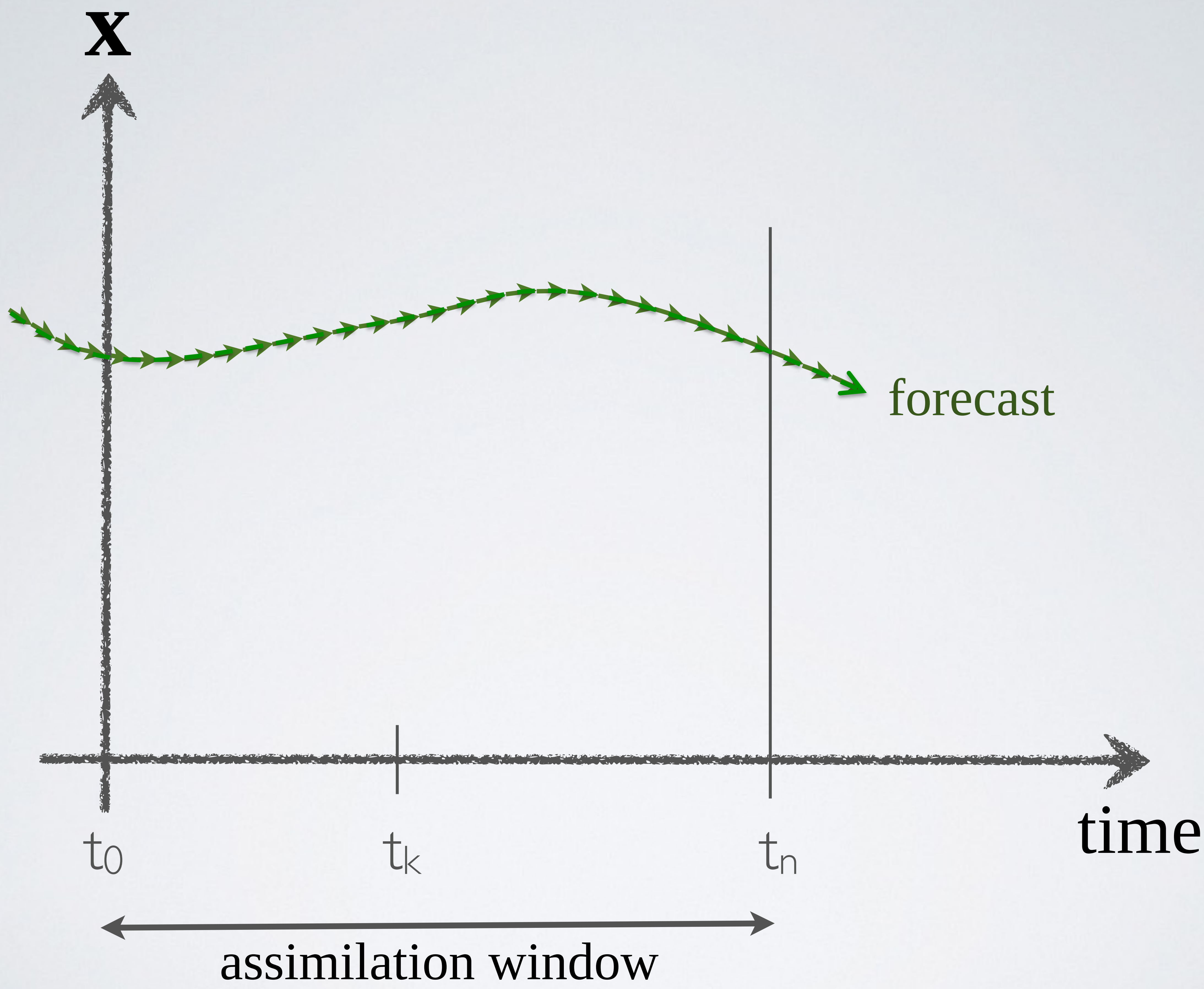
4D-Var minimizes the misfit between a temporal sequence of model states and the observations that are available over a given assimilation window. In contrast to Kalman filter (and to sequential algorithms), it propagates the information contained in the data both forward and backward in time.

The general idea behind 4D-Var is to find the initial conditions which lead to the best fit to observations which are spread over a time interval. The notion of 'best' is defined by a scalar cost function.

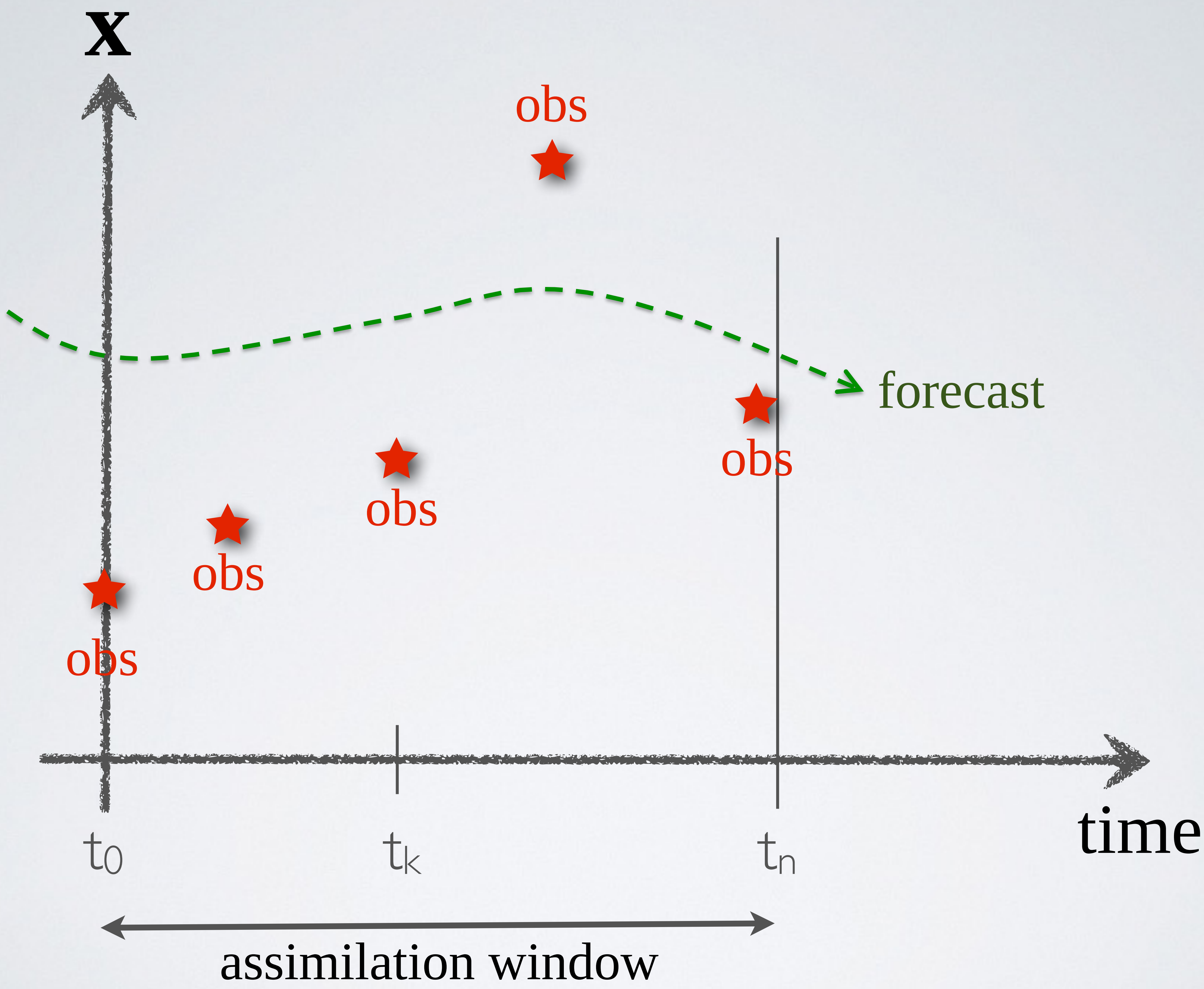
Find the initial state which produces a model trajectory (when integrated in time using the forecast model) that 'best' fits the observations.

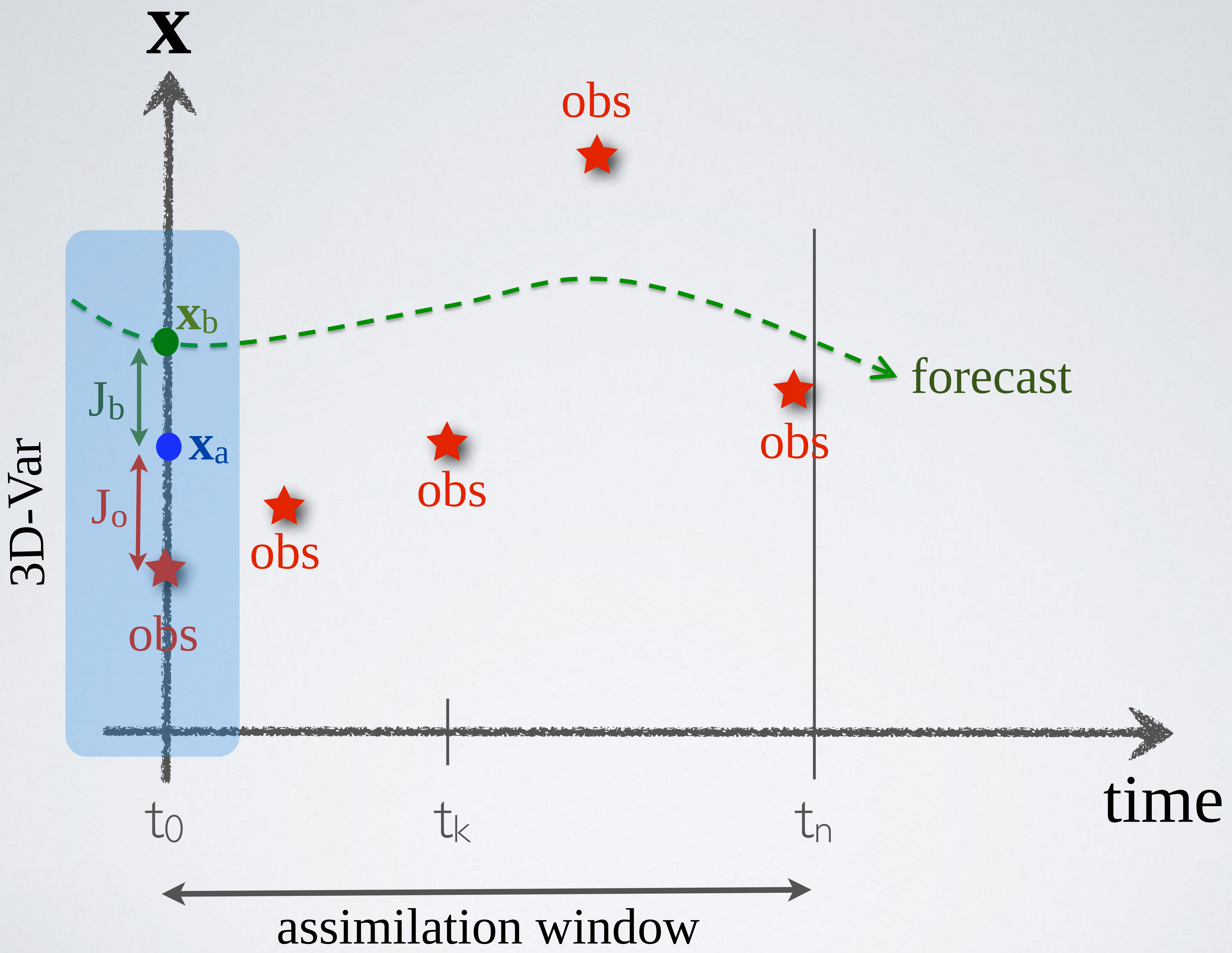
$$J(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right)^T \left(\mathbf{P}_0^b\right)^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^b\right) + \frac{1}{2}\sum_{k=0}^{K} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)^T\, (\mathbf{R}_k)^{-1} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)$$


x
time
t0
tk
tn
assimilation window

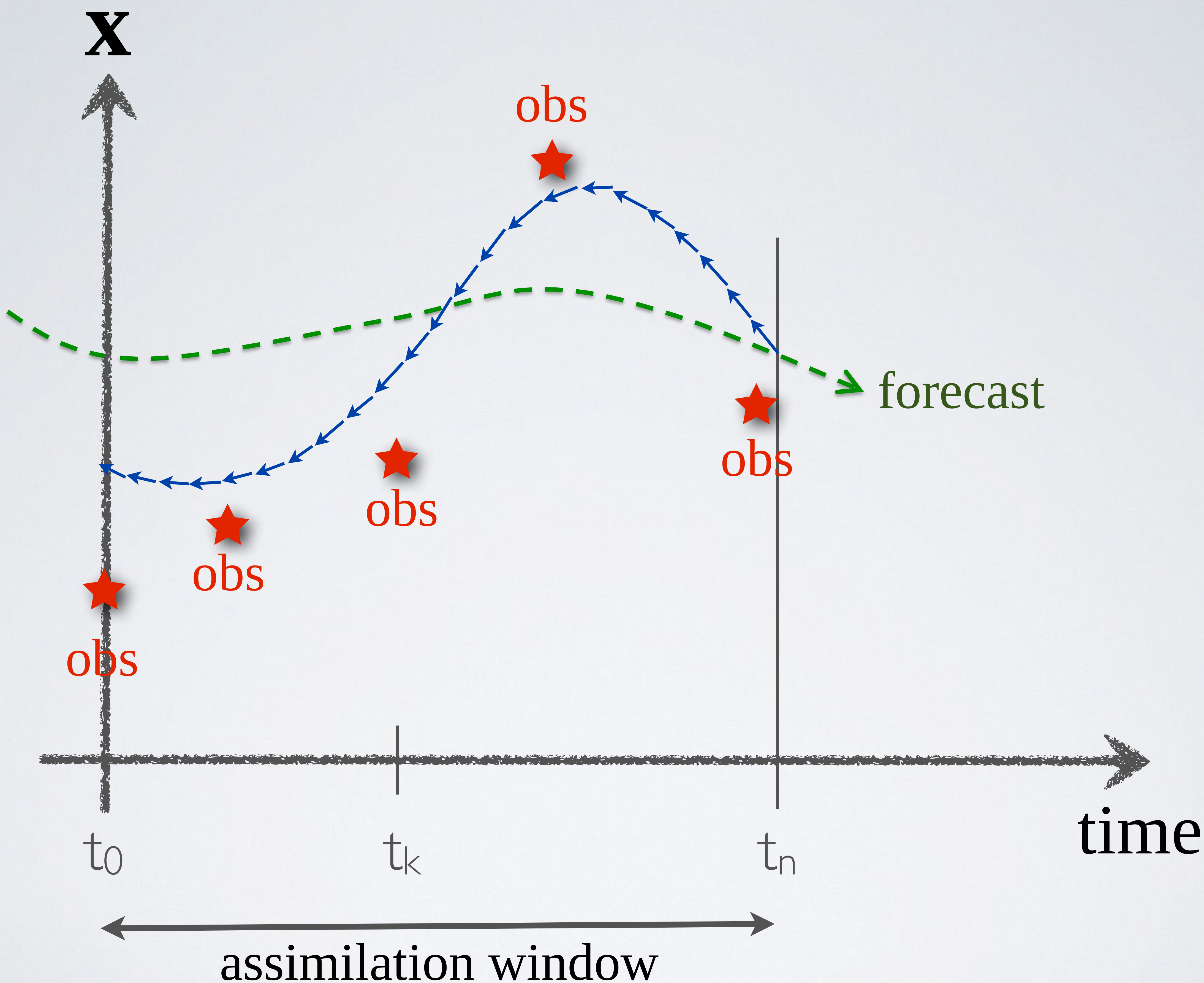


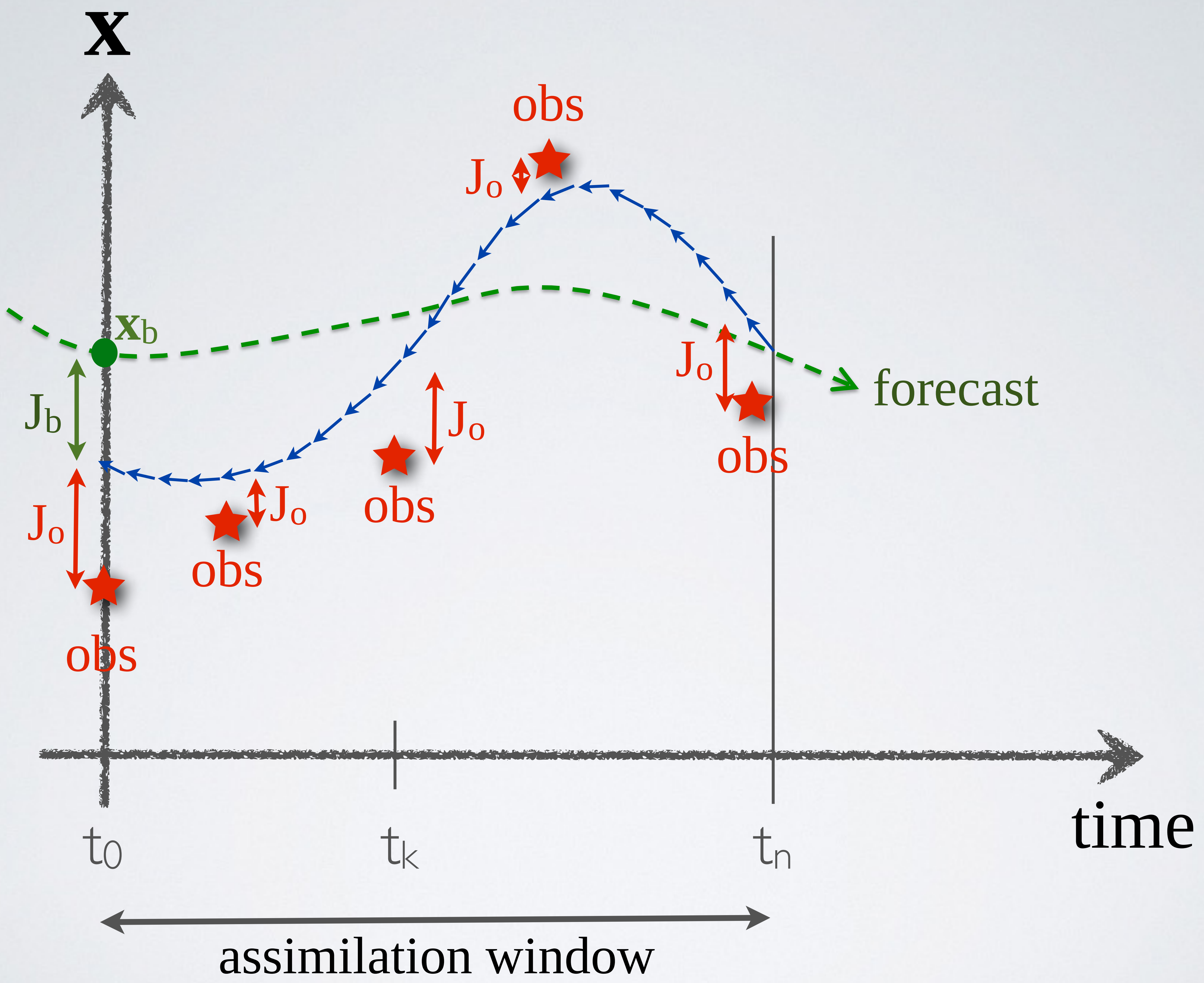
x
forecast
time
t0
tk
tn
assimilation window

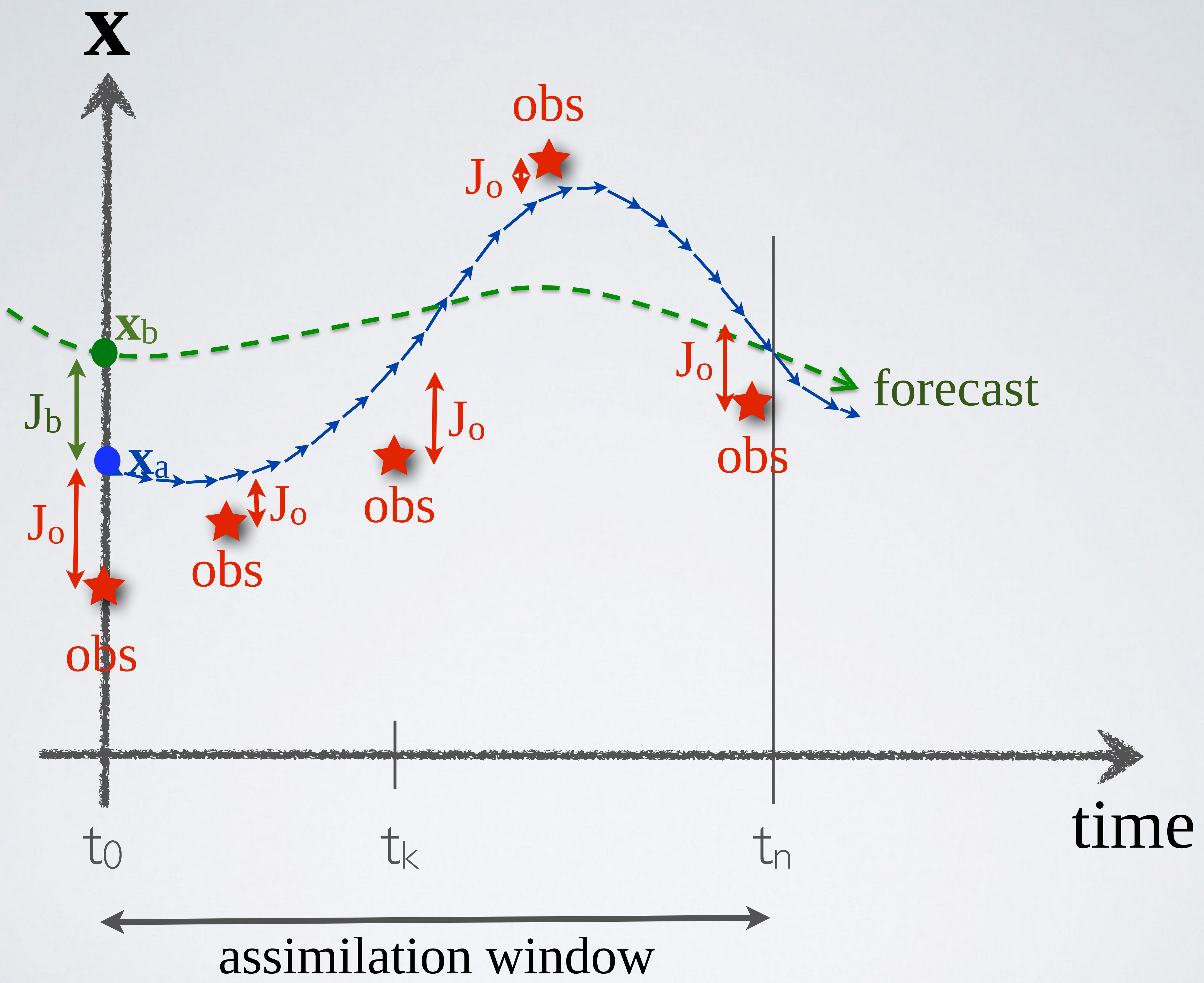

x
obs
forecast
obs
obs
obs
obs
time
t0
tk
tn
assimilation window

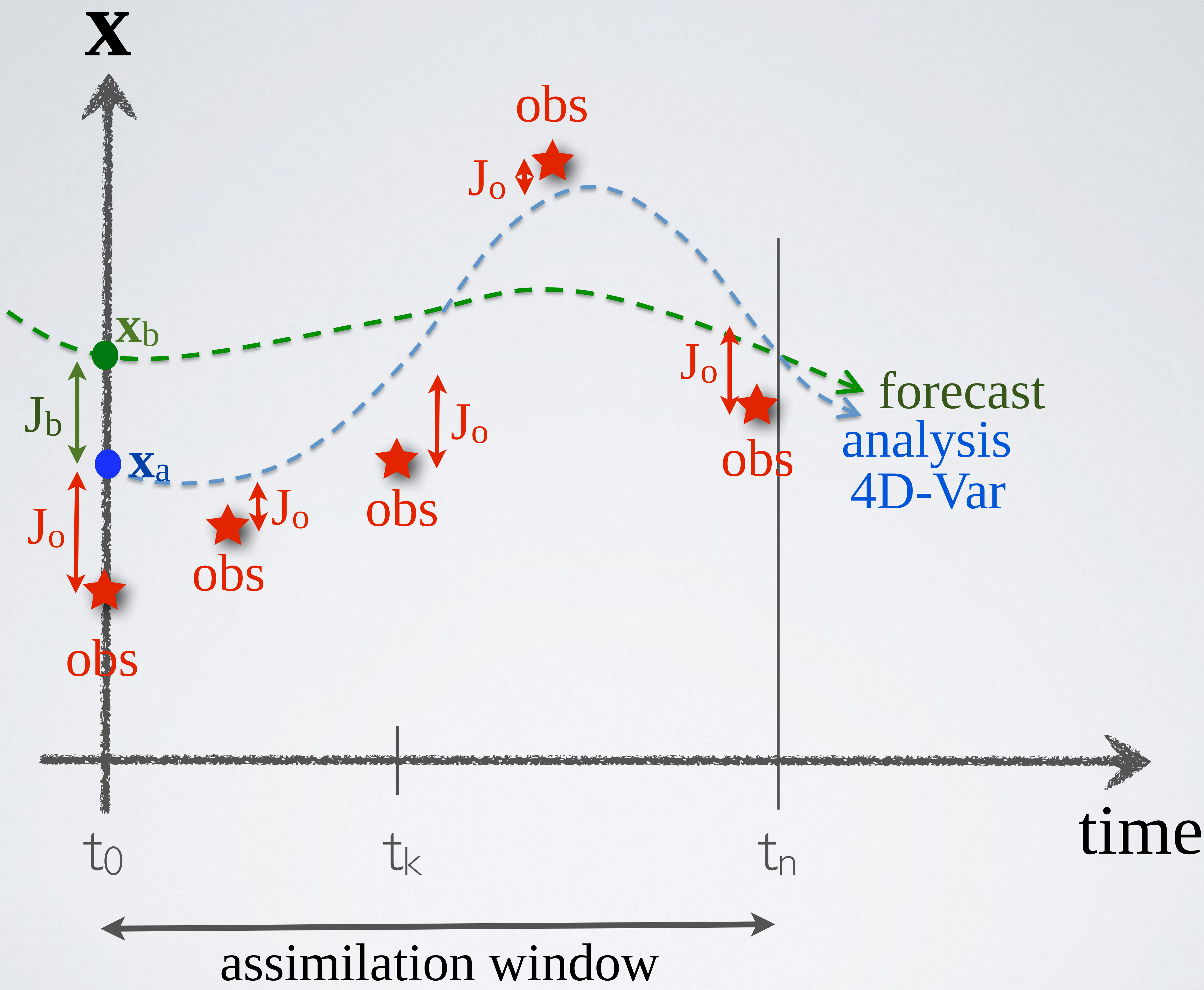


x
obs
forecast
x_b
J_b
x_a
3D-Var
J_o
obs
obs
obs
obs
time
t_0
t_k
t_n
assimilation window

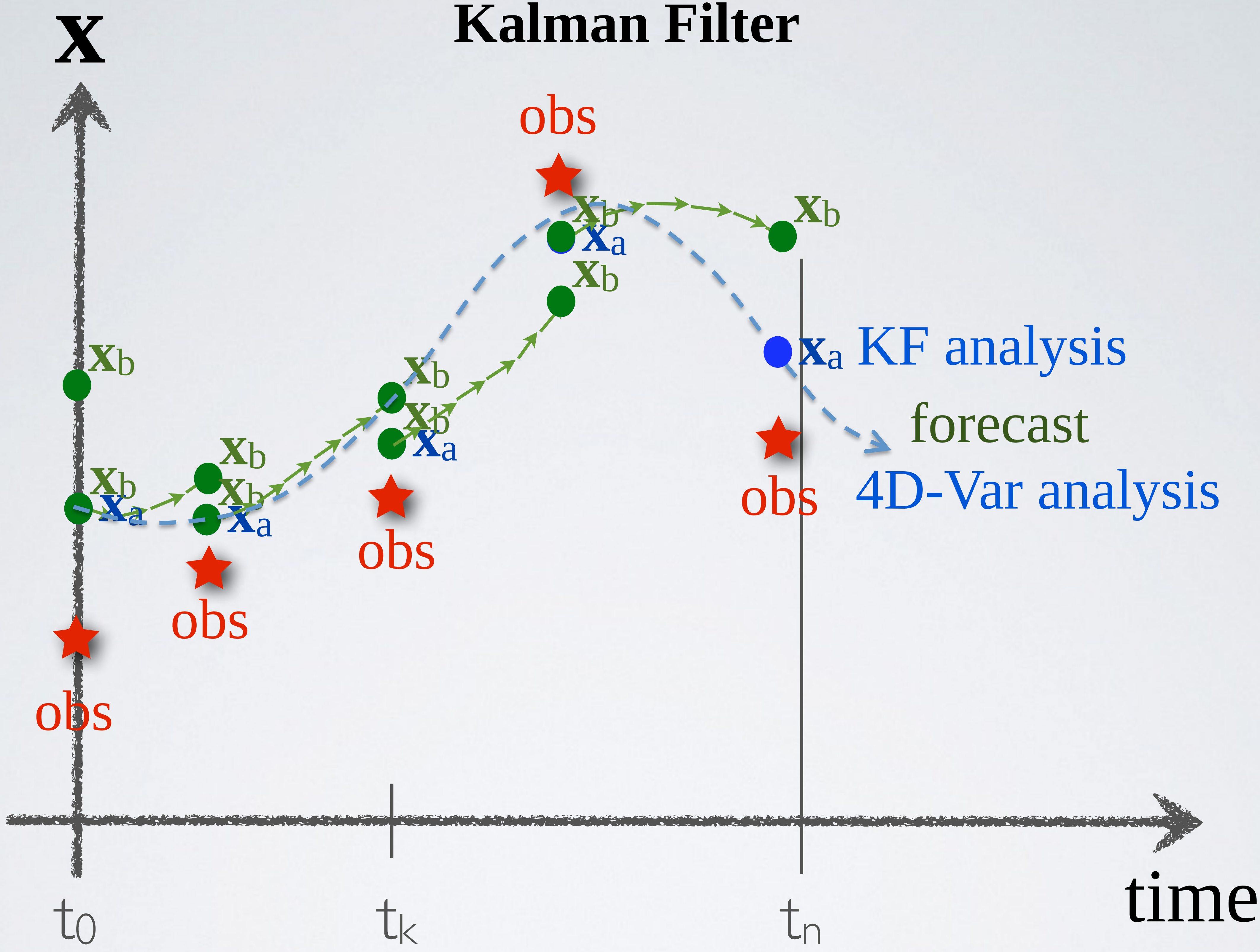

x
obs
obs
obs
obs
obs
forecast
time
t0
tk
tn
assimilation window


x
obs
J_o
x_b
J_o
forecast
J_b
J_o
obs
J_o
obs
J_o
obs
obs
time
t_0
t_k
t_n
assimilation window


x
obs
J_o
x_b
J_o
forecast
J_b
J_o
x_a
obs
J_o
obs
J_o
obs
obs
time
t_0
t_k
t_n
assimilation window


x
obs
Jo
xb
Jb
xa
Jo
obs
Jo
obs
Jo
obs
Jo
obs
forecast
analysis
4D-Var
obs
t0
tk
tn
time
assimilation window

Kalman Filter

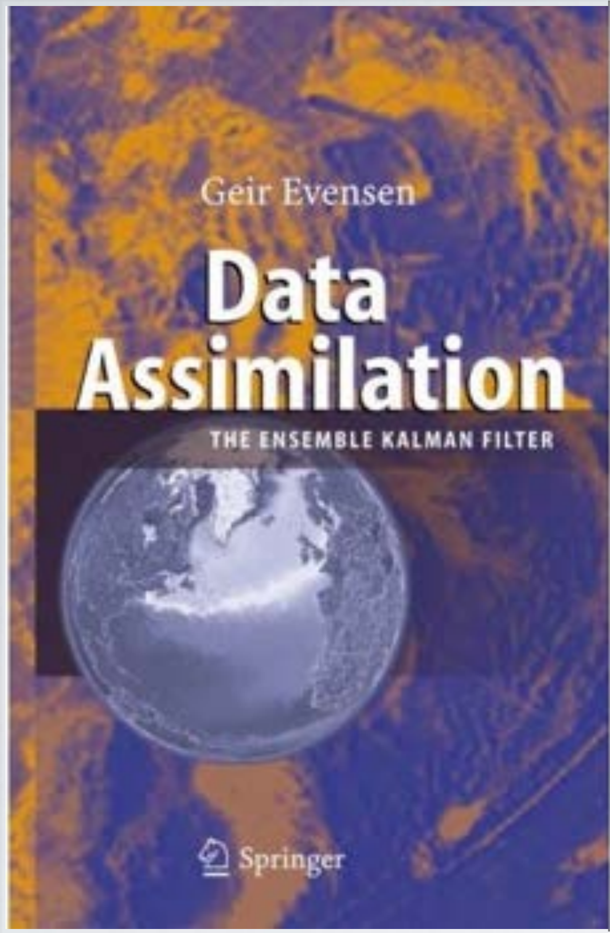


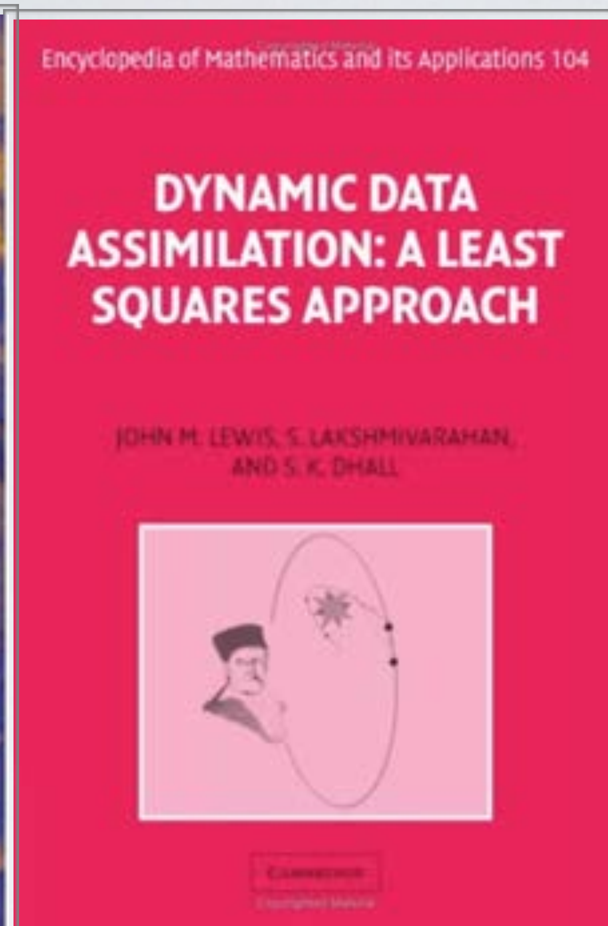
My Own Final Thoughts

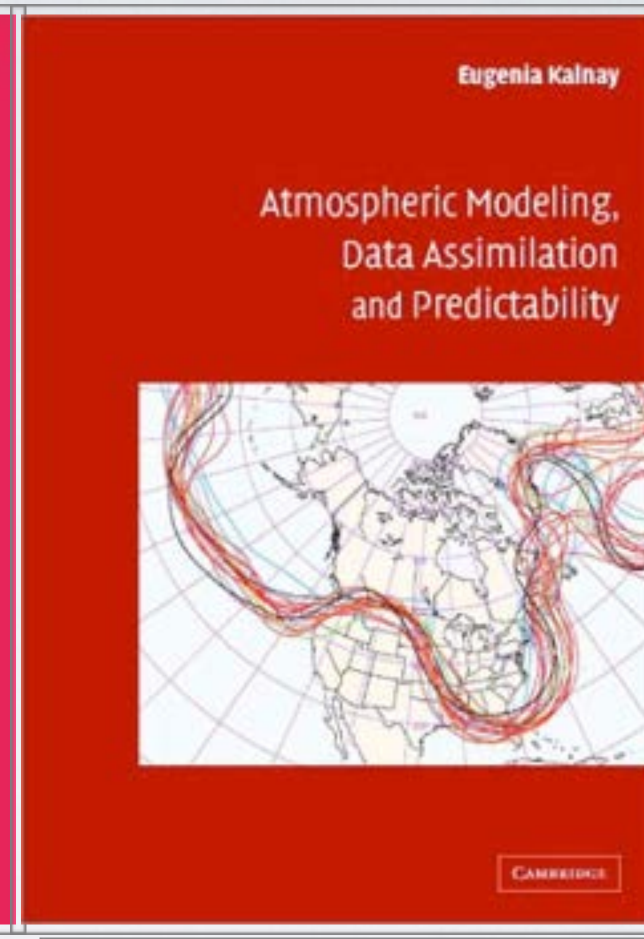
A data assimilation person:

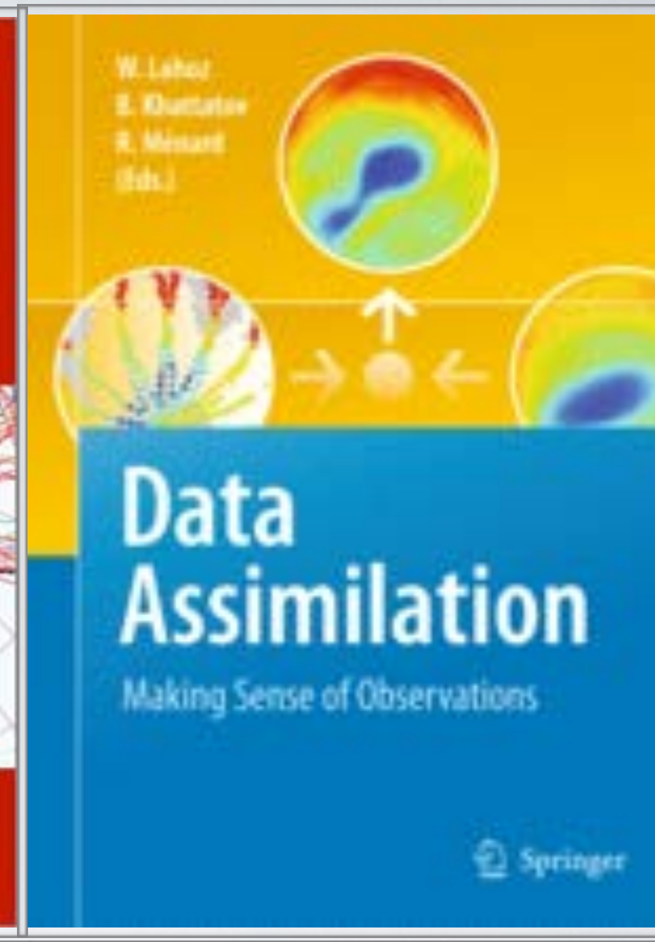
- is a HPC (computing) 'hog'
- 'abuses' the data
- thinks the model is wrong but 'blames' the data anyway
- has the 'constitutional' right to change the model and/or data but is very conservative about change
- must do everything right — the devil is in the details
- pretends to know the truth

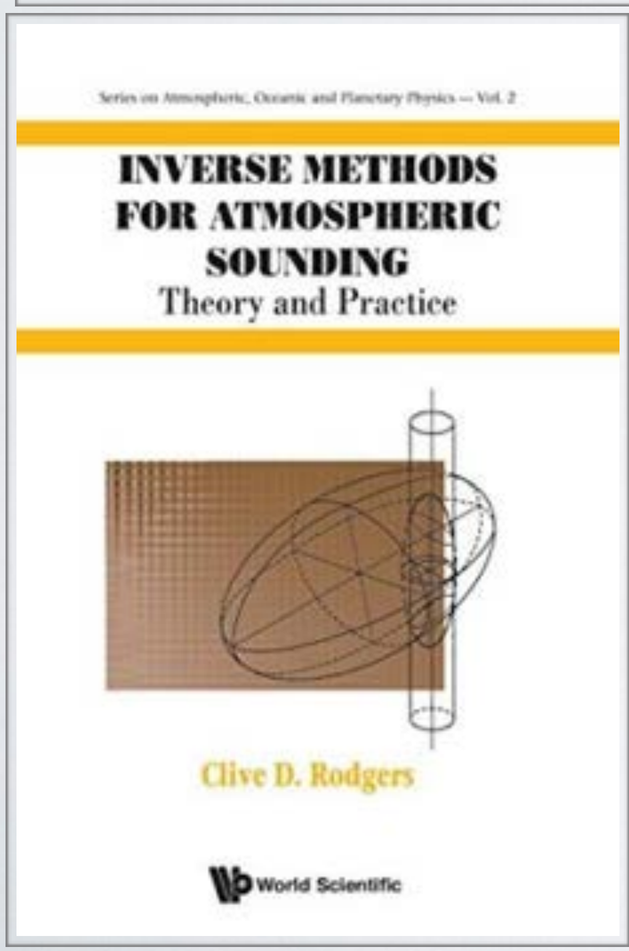
Some References

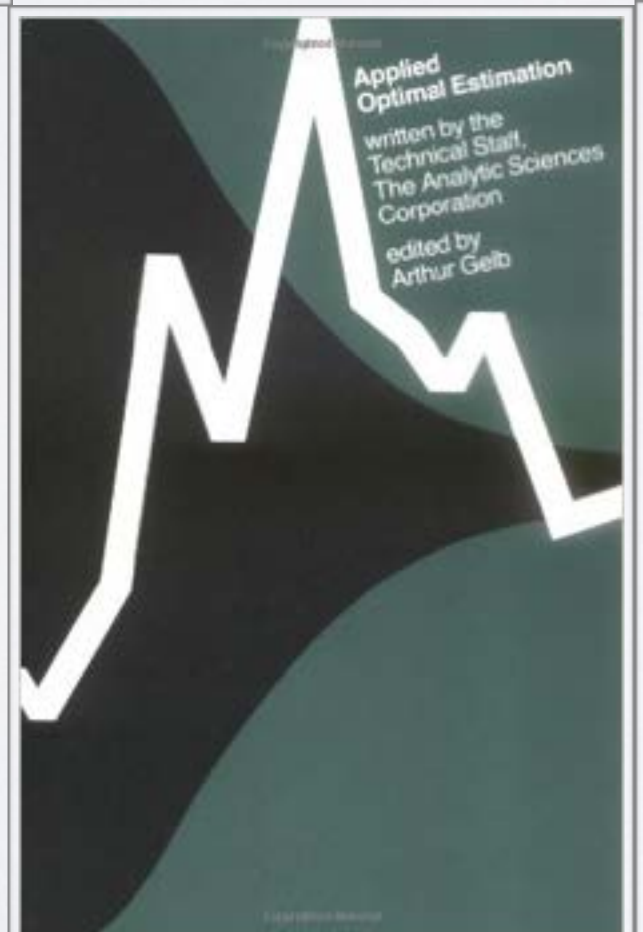


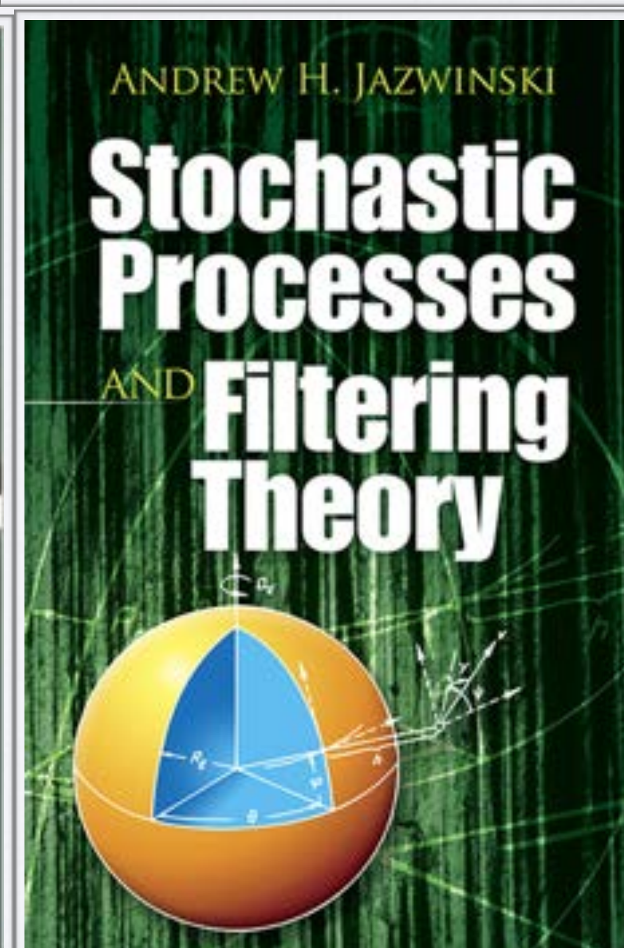


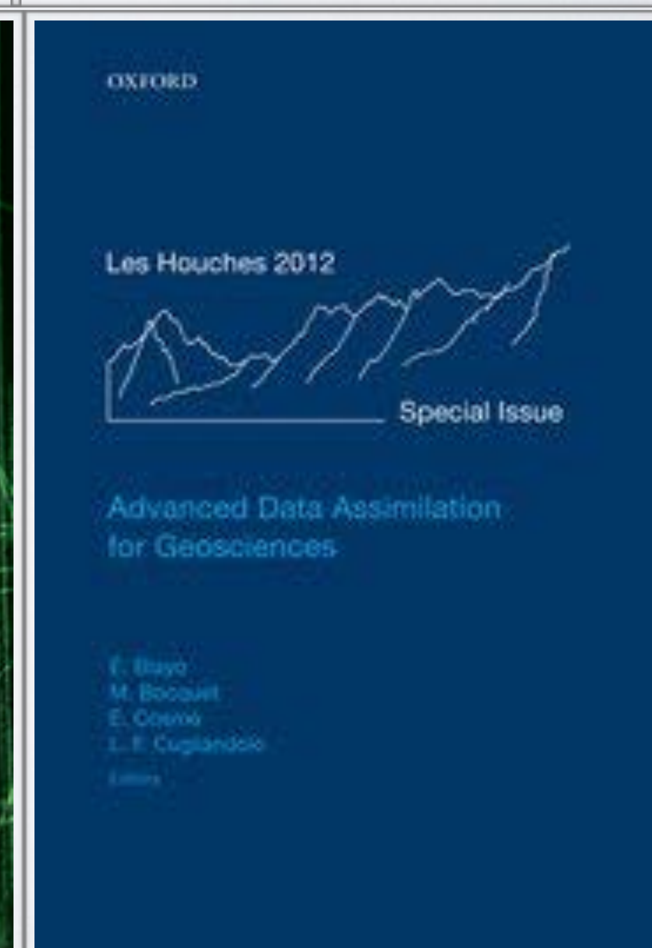












"Ave, may i go home? I cant assimilate any more data today."

http://www.condenaststore.com

EXTRA SLIDES

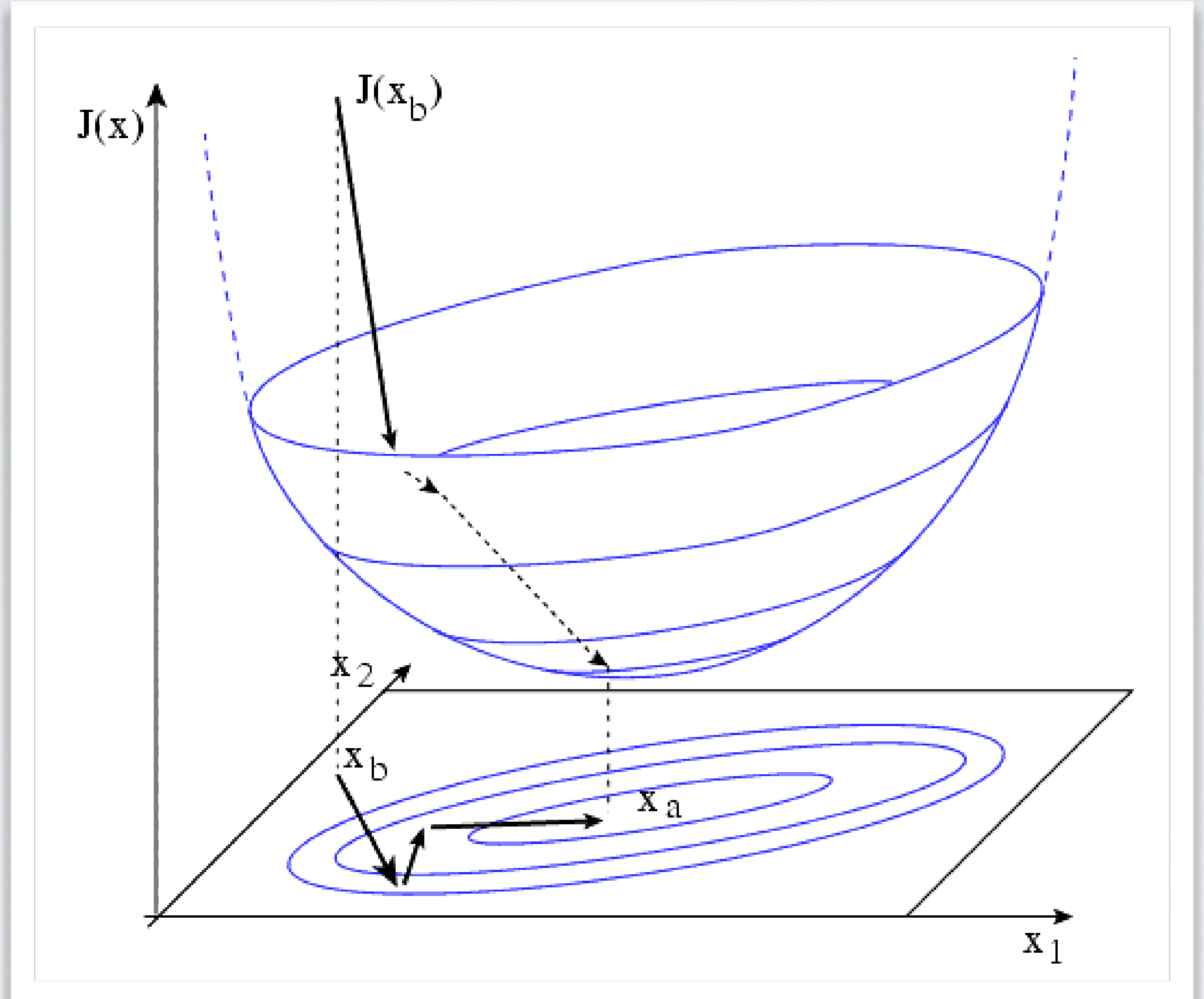
4D-Var Implementation

Given a scalar cost function

$$J(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right)^T \left(\mathbf{P}_0^b\right)^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^b\right) + \frac{1}{2}\sum_{k=0}^{K} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)^T\,(\mathbf{R}_k)^{-1}(H_k(\mathbf{x}_k) - \mathbf{y}_k^o)$$

we want to find an estimate of $\mathbf{x}_0$ that minimizes the cost function.

Graphically for n=2, the geometry of the minimization of the cost function term for the background state is:



The minimization works by performing several line-searches to move the control variable to areas where the cost-function is smaller, usually by looking at the local slope (the gradient) of the cost-function.

4D-Var Implementation

4D-Var can be seen to be an iterative algorithm. For iteration, i, we will:

1. Run the nonlinear model with initial conditions, $\mathbf{x}_0^i$ from t_0 to t_K
2. Compute the cost, $J(\mathbf{x}_0^i)$.
3. Compute the gradient with respect to the initial state, $\nabla_{\mathbf{x}_0^i} J$ to find out the direction of steepest descent.
4. Choose the descent direction, d^i based on the direction of steepest descent, and choose a step size, α^i.
5. Modify the initial state: $x_0^{i+1} = x_0^i - \alpha^i d^i$

The iteration is continued until the minimum of the cost function is found.

4D-Var Implementation

1. Run the nonlinear model with initial conditions, $\mathbf{x}_0^i$ from t_0 to t_K

$$\mathbf{x}_{k+1} = M_k(\mathbf{x}_k)$$

Typically, we have a nonlinear model which is written as a set of N nonlinear coupled ODEs

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \mathbf{F} = \begin{bmatrix} F_1 \\ \vdots \\ F_N \end{bmatrix}$$

Once we choose a time-difference scheme, it becomes a set of nonlinear coupled difference equations (e.g. Crank-Nicholson)

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \Delta t\mathbf{F}\left(\frac{\mathbf{x}^k + \mathbf{x}^{k+1}}{2}\right)$$

A numerical solution starting from an initial time can be readily obtained by integrating the model numerically between the initial time and a final time ('running the model'). This gives us a nonlinear model solution that depends only on the initial conditions:

$$\mathbf{x}(t) = \boldsymbol{M}[\mathbf{x}(t_0)]$$

4D-Var Implementation

This gives us a nonlinear model solution that depends only on the initial conditions: $\mathbf{x}(t) = \boldsymbol{M}[\mathbf{x}(t_0)]$

where $\boldsymbol{M}$ is the time integration of the numerical scheme from the initial condition to time t.

A small perturbation $\delta\mathbf{x}(t)$ can be added to $\mathbf{x}(t)$ such that:

$$\boldsymbol{M}[\mathbf{x}(t_0) + \delta\mathbf{x}(t_0)] = \boldsymbol{M}[\mathbf{x}(t_0)] + \frac{\partial \boldsymbol{M}}{\partial \mathbf{x}}\delta\mathbf{x}(t_0) + O[\delta\mathbf{x}(t_0)^2]$$

$$\boldsymbol{M}[\mathbf{x}(t_0) + \delta\mathbf{x}(t_0)] = \mathbf{x}(t) + \delta\mathbf{x}(t) + O[\delta\mathbf{x}(t_0)^2]$$

At any given time, the linear evolution of the small perturbation $\delta\mathbf{x}(t)$ will be given by:

$$\frac{d\delta\mathbf{x}}{dt} = \frac{\partial \boldsymbol{F}}{\partial \mathbf{x}}\delta\mathbf{x}$$

TLM in differential form

Its solution between the initial time to final time can be obtained by integrating the TLM in time:

$$\delta\mathbf{x}(t) = \mathbf{M}(t_0, t)\,\delta\mathbf{x}(t_0)$$

where $\mathbf{M}(t_0, t) = \frac{\partial \boldsymbol{M}}{\partial \mathbf{x}}$ is known as the resolvent or propagator of the TLM

It propagates an initial perturbation at time t_0 into the final perturbation at time t.

4D-Var Implementation

TLM: $$\mathbf{M}(t_0, t) = \frac{\partial \boldsymbol{M}}{\partial \mathbf{x}}$$

Because it is linearized over the flow from t_0 to t, it depends on the basic trajectory $\mathbf{x}(t)$ (the solution of the nonlinear model) but it does not depend on the perturbations $\delta\mathbf{x}(t)$.

The adjoint of an operator $\mathbf{M}$ is defined by the property

$$\langle \mathbf{x}, \mathbf{M}\mathbf{y} \rangle = \langle \mathbf{M}^T \mathbf{x}, \mathbf{y} \rangle$$

In the case of a model with real variables, the **adjoint of the tangent linear model** $\mathbf{M}(t_0, t)$ is simply the **transpose of the tangent linear model**.

4D-Var Implementation

$$\mathbf{M}(t_0, t) = \frac{\partial \boldsymbol{M}}{\partial \mathbf{x}}$$

In the case of a model with real variables, the adjoint of the tangent linear model $\mathbf{M}(t_0, t)$ is simply the transpose of the tangent linear model.

Now assume that we separate the interval (t_0, t) into two successive time intervals, say: $t_0 < t_1 < t$

$$\mathbf{M}(t_0, t) = \mathbf{M}(t_1, t)\mathbf{M}(t_0, t_1)$$

Since the adjoint of the tangent linear model is the transpose of the TLM, the property of the transpose of a product is also valid:

$$\mathbf{M}^T(t_0, t) = \mathbf{M}^T(t_0, t_1)\mathbf{M}^T(t_1, t)$$

This shows that the TLM can be cast as a product of TLM matrices corresponding to short integrations. This also shows that the adjoint of the model can also be separated the same way but they are executed backwards in time starting from the last time step and ending with the first time step.

4D-Var Implementation

Why do we need the adjoint?

Looking back

4D-Var can be seen to be an iterative algorithm. For iteration, i, we will:

1. Run the nonlinear model with initial conditions, $\mathbf{x}_0^i$ from t_0 to t_K

2. Compute the cost, $J\left(\mathbf{x}_0^i\right)$.

$$J\left(\mathbf{x}_0^i\right) = \frac{1}{2}\left(\mathbf{x}_0^i - \mathbf{x}_0^b\right)^T \left(\mathbf{P}_0^b\right)^{-1}\left(\mathbf{x}_0^i - \mathbf{x}_0^b\right) + \frac{1}{2}\sum_{k=0}^{K}(H_k(\mathbf{x}_k) - \mathbf{y}_k^o)^T\,(\mathbf{R}_k)^{-1}(H_k(\mathbf{x}_k) - \mathbf{y}_k^o)$$

Looking back

4D-Var can be seen to be an iterative algorithm. For iteration, i, we will:

1. Run the nonlinear model with initial conditions, $\mathbf{x}_0^i$ from t_0 to t_K

2. Compute the cost, $J\left(\mathbf{x}_\mathbf{0}^{\boldsymbol{i}}\right)$.

$$J\left(\mathbf{x}_\mathbf{0}^{\boldsymbol{i}}\right) = \frac{1}{2}\left(\mathbf{x}_\mathbf{0}^{\boldsymbol{i}} - \mathbf{x}_0^b\right)^T \left(\mathbf{P}_0^b\right)^{-1} \left(\mathbf{x}_\mathbf{0}^{\boldsymbol{i}} - \mathbf{x}_0^b\right) + \frac{1}{2}\sum_{k=0}^{K} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)^T \, (\mathbf{R}_k)^{-1} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)$$

3. Compute the gradient with respect to the initial state, $\nabla J\left(\boldsymbol{x}_0^i\right) = \lambda_0$ to find out the direction of steepest descent(using the adjoint)

$$\lambda_K = \mathbf{H}_K^T (\mathbf{R}_K)^{-1} (H_K(\mathbf{x}_K) - \mathbf{y}_K^o)$$

$$\lambda_k = \mathbf{M}_k^T \lambda_{k+1} + \mathbf{H}_k^T (\mathbf{R}_k)^{-1} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)$$

$$\lambda_0 = \mathbf{M}_0^T \lambda_1 + \mathbf{H}_0^T (\mathbf{R}_0)^{-1} (H_0(\mathbf{x}_0) - \mathbf{y}_0^o) + \left(\mathbf{P}_0^b\right)^{-1} \left(\mathbf{x}_\mathbf{0}^{\boldsymbol{i}} - \mathbf{x}_0^b\right)$$

4D-Var Implementation

4D-Var can be seen to be an iterative algorithm. For iteration, i, we will:

1. Run the nonlinear model with initial conditions, $\mathbf{x}_0^i$ from t_0 to t_K
2. Compute the cost, $J(\mathbf{x}_0^i)$.
3. Compute the gradient with respect to the initial state, $\nabla J(\boldsymbol{x}_0^i) = \lambda_0$ to find out the direction of steepest descent (using the adjoint).
4. Choose the descent direction, d^i based on the direction of steepest descent, and choose a step size, α^i.

For our case, we can use Newton's method (or quasi-Newton's method:

$$\alpha^i = 1, \qquad d^i = \nabla^{-2} J(\mathbf{x}_0^i) \nabla_{\mathbf{x}_0^i} J$$

Steepest descent is the product of inverse Hessian and the gradient of the cost function:

$$\alpha^i = 1, \qquad d^i = \nabla^{-2} J(\mathbf{x}_0^i) \nabla_{\mathbf{x}_0^i} J$$

Or,

$$d^i = \nabla^{-2} J(\boldsymbol{x}_0^i) \;\; \lambda_0$$

The inverse hessian, $\nabla^{-2} J(\boldsymbol{x}_0^i)$ is typically approximated for non-scalar system by simply perturbing the gradient and take the finite difference between perturbed gradient and unperturbed gradient. The Hessian will be:

$$\nabla^2 J(\boldsymbol{x}_0^i) = \frac{1}{\delta}[\nabla J(\boldsymbol{x}_0^i + \delta I) - \nabla J(\boldsymbol{x}_0^i)]$$

where δ is a perturbation constant.

4D-Var Implementation

4D-Var can be seen to be an iterative algorithm. For iteration, i, we will:

1. Run the nonlinear model with initial conditions, $\mathbf{x}_0^i$ from t_0 to t_K
2. Compute the cost, $J(\mathbf{x}_0^i)$.
3. Compute the gradient with respect to the initial state, $\nabla J(\boldsymbol{x}_0^i) = \lambda_0$ to find out the direction of steepest descent (using the adjoint).
4. Choose the descent direction, d^i based on the direction of steepest descent (use Newton's method to find inverse Hessian, $\nabla^{-2} J(\boldsymbol{x}_0^i)$).
5. Modify the initial state: $x_0^{i+1} = x_0^i - \alpha^i d^i$

$$\boldsymbol{x}_0^{i+1} = \boldsymbol{x}_0^i - \nabla^{-2} J(\boldsymbol{x}_0^i)\ \lambda_0$$

when using Newton's method: iteration, $i = 1,$ & $\alpha^i = 1$.

4D-Var Implementation

4D-Var can be seen to be an iterative algorithm. For iteration, i, we will:

1. Run the nonlinear model with initial conditions, $\mathbf{x}_0^i$ from t_0 to t_K
2. Compute the cost, $J(\mathbf{x}_0^i)$.
3. Compute the gradient with respect to the initial state, $\nabla J(\boldsymbol{x}_0^i) = \lambda_0$ to find out the direction of steepest descent (using the adjoint).
4. Choose the descent direction, $\boldsymbol{d}^i$ based on the direction of steepest descent (use Newton's method to find inverse Hessian, $\nabla^{-2} J(\boldsymbol{x}_0^i)$).
5. Modify the initial state: $\boldsymbol{x}_0^{i+1} = \boldsymbol{x}_0^i - \nabla^{-2} J(\boldsymbol{x}_0^i)\ \lambda_0$
6. Calculate the analysis for by running the nonlinear model with updated initial conditions.