Dear students,

Here is another set of data assimilation notes, perhaps a better set than my class notes, which are somewhat old.

These are obviously (if you look at them) from Prof. Wang's data assimilation class. I sat in her class for the first month a few years ago, and I have taken the liberty to piece together 3 different lectures from the first few weeks.

I am giving these to you in the hopes of providing further exposure to the ideas I tried to convey in a single lecture today.

For the exam, I expect you to understand the ideas presented in the first 21 pages, and then I would expect you to be able to derive, using some information from pg 87 in O. Talagrand's paper the final weights for each measurement.

Dr. Wicker

# Data assimilation basics

METR 6803

Advanced topics in data assimilation:

Ensemble Kalman filter techniques

Instructor
Dr. Xuguang Wang
School Of Meteorology, University of Oklahoma
xuguang.wang@ou.edu
Aug. 25, 2009

# Definition of data assimilation

- Data assimilation <-> state estimation

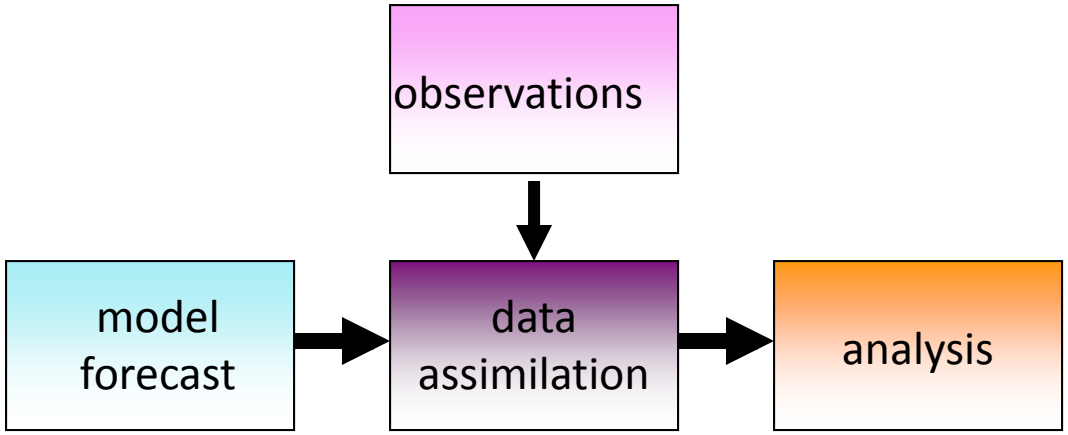
- Information sources to estimate the true state of the nature:

❑ Observations

❑ Physical laws (that govern the evolution and physical properties of the state, often embodied in computer model)

- DA is a process that fuses observations with physical laws (i.e. use all available information) to determine as accurately as possible what the true state is.

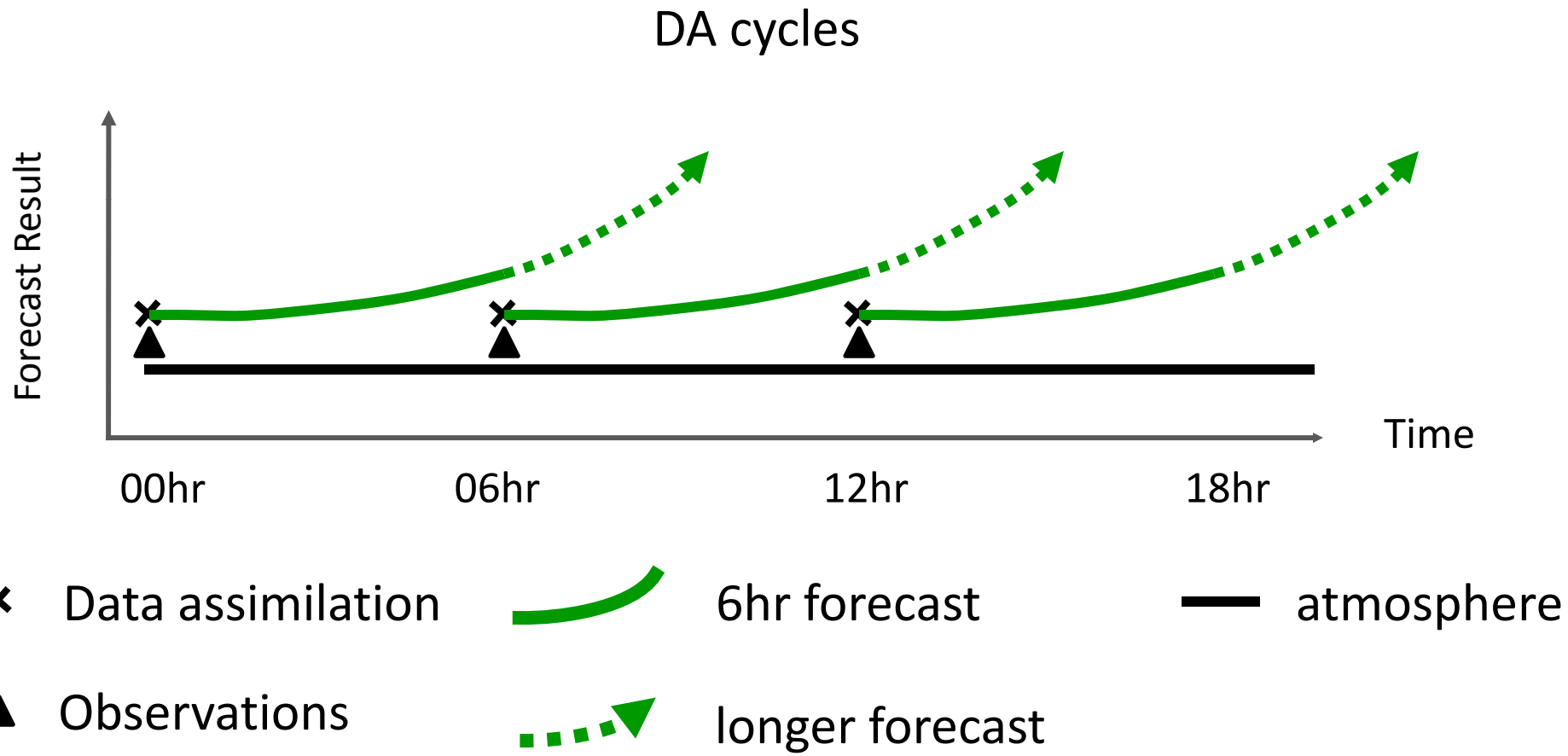
# Definition of data assimilation



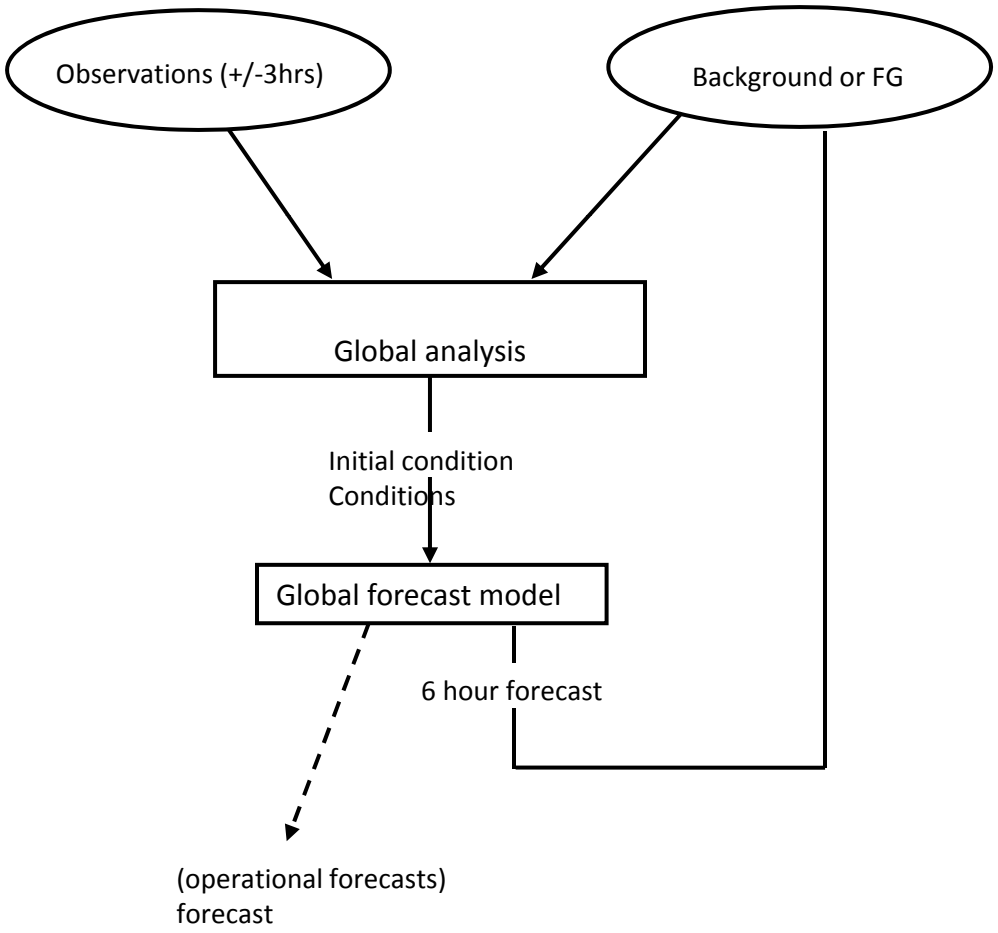

• For a commonly used DA framework, DA is a statistical, objective analysis process that “optimally” combines observations with short-range numerical model forecasts.

• Model forecast is also termed as first guess, or background

# Definition of data assimilation

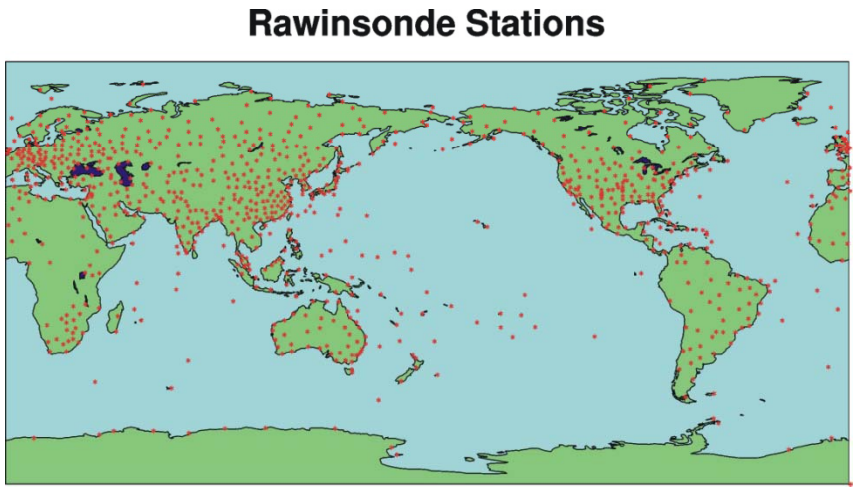


# DA cycles: another view

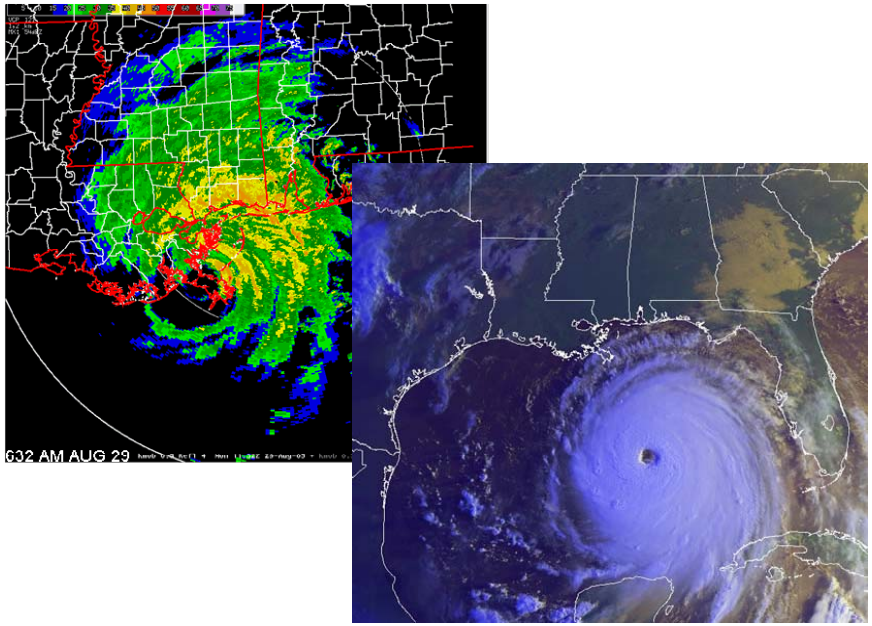


# Why need to combine with model forecast?

- ✓ Need estimate of the state on **regular** model grids and variables (wind, temperature, moisture, etc.).
- ✓ But, **Non-Uniform, Indirect** observations.
- ✓ Model helps to interpolate irregular and indirect obs. to model grid and variables.

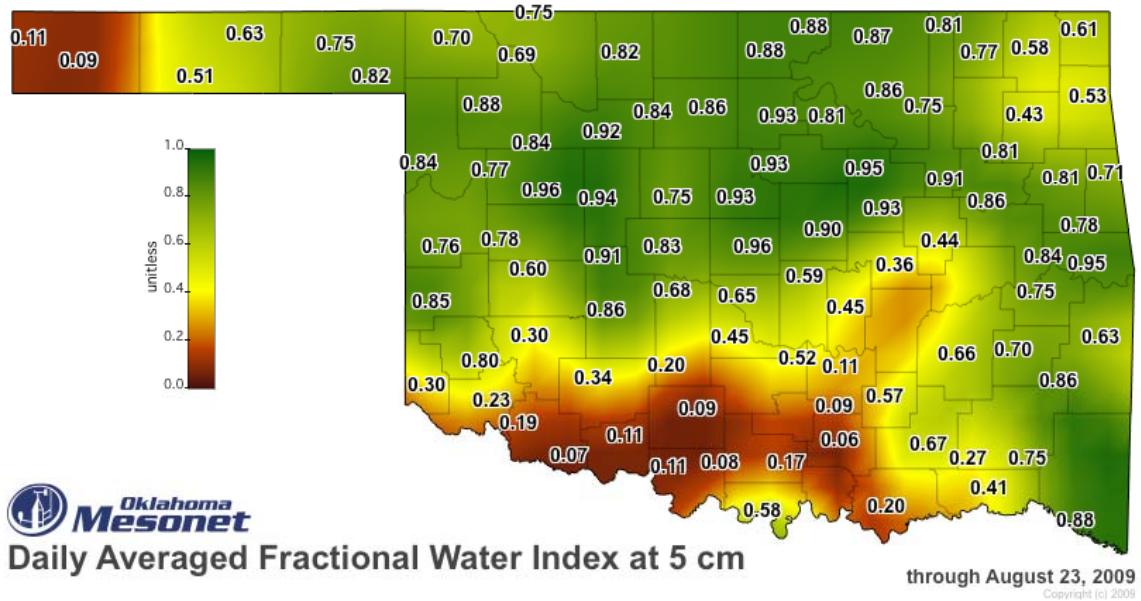



• Irregularly distributed observation network




• Indirectly observed variables: radar reflectivity, satellite radiance

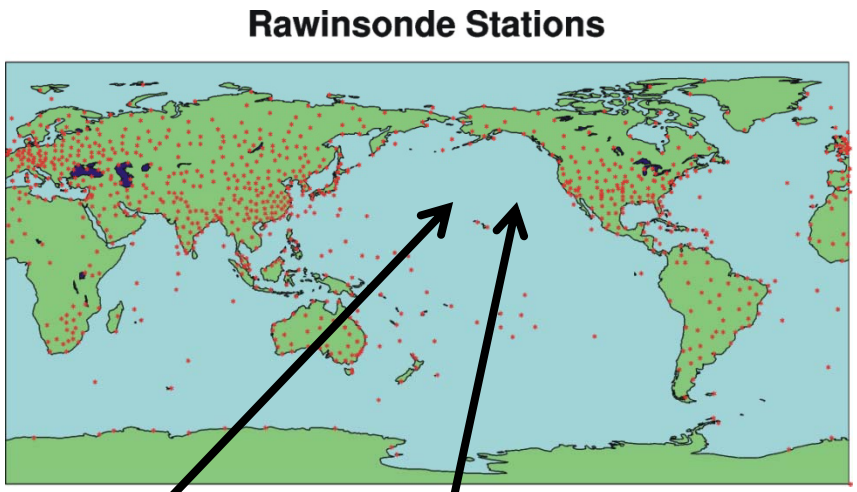
# Other examples of non-uniform observation network



# Why need to combine with model forecast?

✓ Model can provide **additional** information.

• Observations are often insufficient to determine all the unknowns of the state (esp. high dimensional system).




• **First guess/ background** is introduced to solve this problem. It is our best estimate of the state prior to the use of current observations.

• Provide info. in poorly observed regions

• Propagate obs. info. from densely observed to poorly observed region

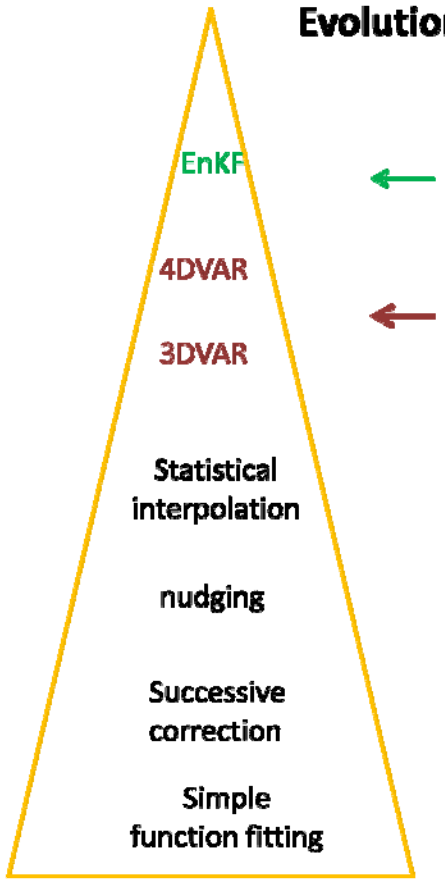
• Through DA cycles, the model organizes and propagates forward the information from previous observations. In other words, information from previous time is preserved.

# A brief review of Statistical Interpolation, 3DVAR and 4DVAR

METR 6803
Advanced topics in data assimilation:
Ensemble Kalman filter techniques

Instructor
Dr. Xuguang Wang
xuguang.wang@ou.edu
Sep. 10, 2009

## Evolution and current status of data assimilation




Topic of this class; being actively explored in many fields; start to be operational in some Meteorology centers; easy to implement

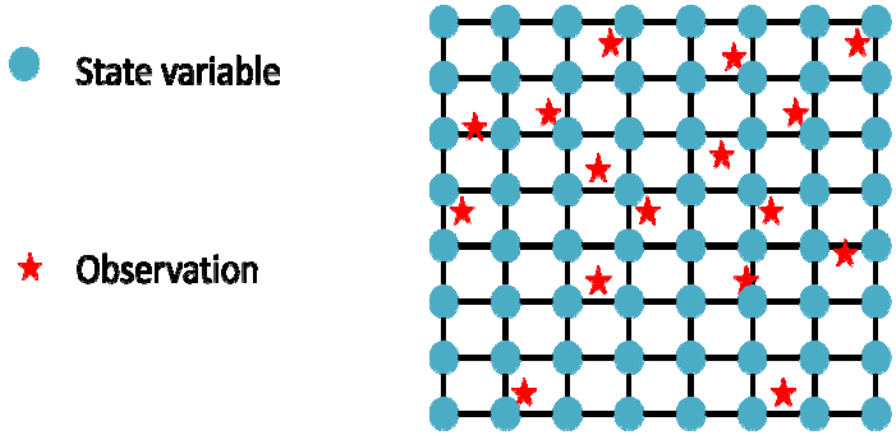
Variational methods are popular methods in operational centers; 4DVAR state of the art

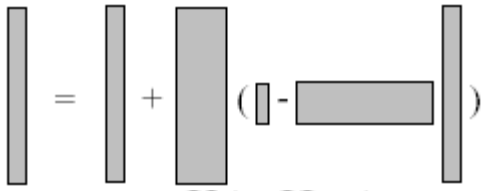
# Statistical Interpolation

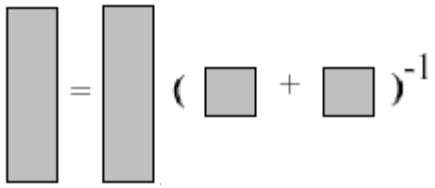
- The analysis is given by

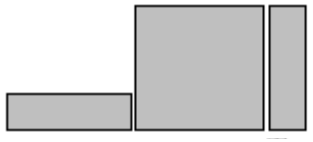
$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{W}\left(\mathbf{y} - H\left(\mathbf{x}^{\mathrm{b}}\right)\right) = \mathbf{x}^{\mathrm{b}} + \mathbf{W}\mathbf{d}$$

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







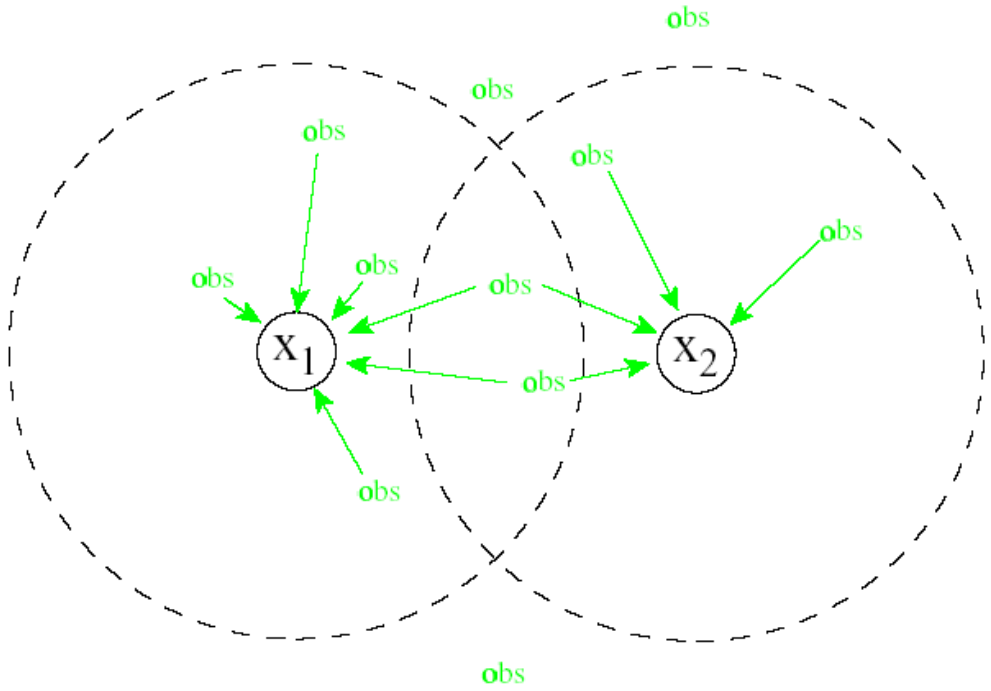
- The actual implementation requires simplifications in the computation of the weight **W**.

- The equation for $\mathbf{x}^a$ can be regarded as a list of scalar analysis equations, one per model variable.

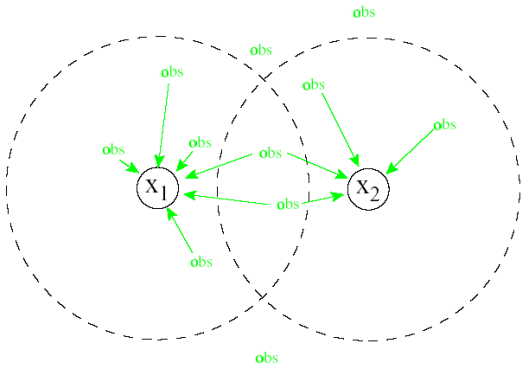
- For each model variable the analysis increment is given by the corresponding row of **W** times the innovation.

- The fundamental hypothesis in the typically implementation of statistical interpolation is: *For each model variable, only a few observations are important in determining the analysis increment*.

- Based on this assumption, the problem of matrix product and inversion is reduced by including only a smaller number of observations for the analysis at a given grid point.



- The actual implementation can be as follows:




1) For each model variable $x_i$, select a small number of $p_i$ observations using empirical selection criteria.

2) Form the corresponding innovation vector

3) Form the $p_i$ background error covariances between the model variable $x_i$ and the model state interpolated at the $p_i$ observation points (i.e. the relevant $p_i$ coefficients of the $i$-th row of $\mathbf{P}^b\mathbf{H}^T$), and

4) Form the $p_i \times p_i$ background and observation error covariance submatrices formed by the selected observations ($\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^T + \mathbf{R}$).

5) Invert the matrix in 4) for selected observations.

6) Multiply it by the $i$-th line of $\mathbf{P}^b\mathbf{H}^T$ to get the necessary row of $\mathbf{W}$.

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{W}\left(\mathbf{y} - H\left(\mathbf{x}^{\mathrm{b}}\right)\right) = \mathbf{x}^{\mathrm{b}} + \mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}\left(\mathbf{y} - H\left(\mathbf{x}^{\mathrm{b}}\right)\right)$$

# Three Dimensional Variational method (3DVAR)

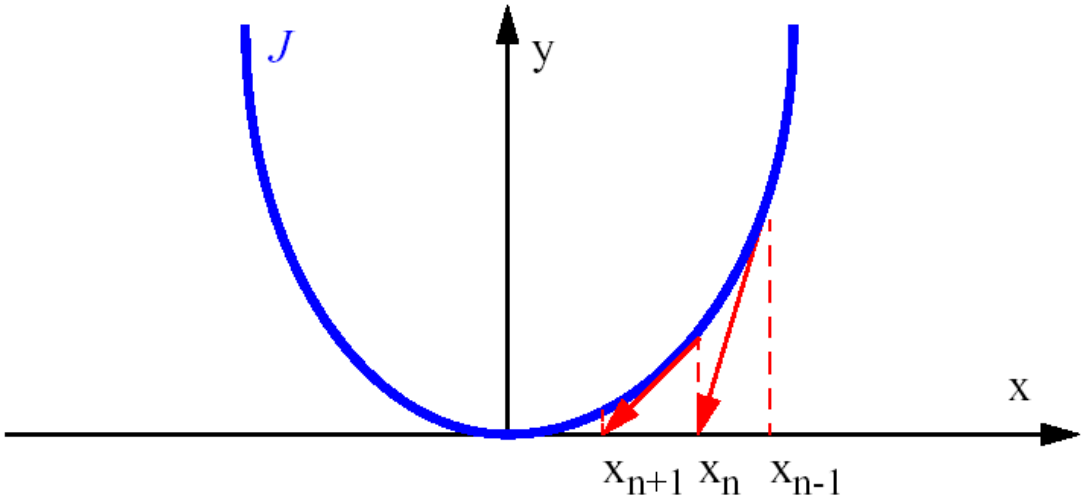
We look for the x that minimizes

$$\boldsymbol{J}(\mathbf{x}) = \frac{1}{2}\left[\left(\mathbf{y} - H(\mathbf{x})\right)^T (\mathbf{R})^{-1}\left(\mathbf{y} - H(\mathbf{x})\right) + \left(\mathbf{x}^{\mathrm{b}} - \mathbf{x}\right)^T \left(\mathbf{P}^{\mathrm{b}}\right)^{-1}\left(\mathbf{x}^{\mathrm{b}} - \mathbf{x}\right)\right]$$

Explicit solution never used!

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \left(\left(\mathbf{P}^{\mathrm{b}}\right)^{-1} + \mathbf{H}^T(\mathbf{R})^{-1}\mathbf{H}\right)^{-1} \mathbf{H}^T(\mathbf{R})^{-1}\left(\mathbf{y} - H\left(\mathbf{x}^{\mathrm{b}}\right)\right)$$

Instead, the minimum is found by numerical minimization, which is usually an iterative procedure.




Linear model: *J* quadratic

**Typical iterative procedure for numerically minimizing *J:***

1. Starting from background as the first guess, compute the cost function *J*

2. Compute the gradient of cost function *J* with respect to the analysis variable **x**

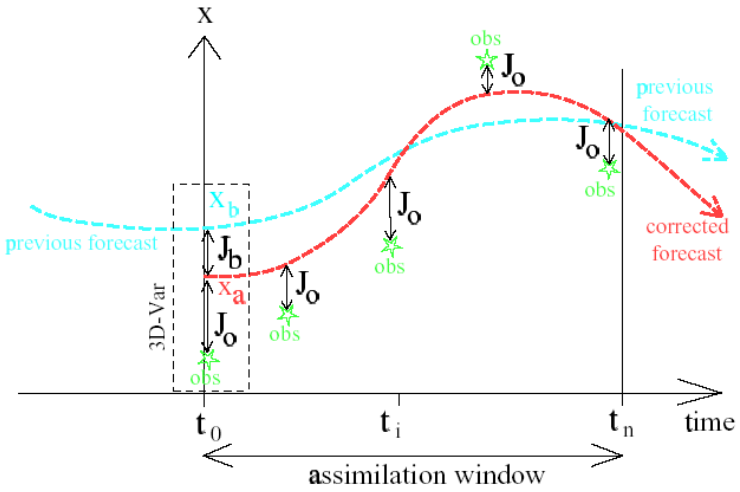
3. Call certain optimization subroutine (using e.g., the conjugate gradient method), passing into the subroutine the cost function and gradient vector, and determine the amount of correction to ***x***

4. Check if the optimal solution has been found by computing the norm of the gradients or the value of $J$ itself to see if either is less than a prescribed tolerance. If not, go back to step 1, with the updated value $\mathbf{x}^{n+1}$, and repeat the steps until convergence is reached. The solution obtained at convergence is the optimal solution that minimizes $J$.

# Four Dimensional Variational method (4DVAR)

4DVAR is actually a direct generalization of 3DVAR to handle observations that are distributed in time. The cost function is the same, provided that the observation operators are generalized to include a forecast model that will allow a comparison between the model state and the observations at the appropriate time.

4DVAR seeks the *initial condition* such that the forecast best fits the observations within the assimilation interval.



Cost function for 4DVAR

Let $\mathbf{x}^f(t_i) = M_{i-1}\left[\mathbf{x}^a(t_{i-1})\right]$ represent the (nonlinear) model forecast that advances from the previous analysis time $t_{i-1}$ to the current $t_i$.

Assume the observations distributed within a time interval $(t_0, t_n)$ will be used. The cost function includes a term measuring the distance to the background *at the beginning of the interval*, and a summation over time of the cost function for each observational increment computed with respect to the model integrated to the time of the observation:

$$J\left(\mathbf{x}(t_0)\right) = \frac{1}{2}\left[\begin{array}{l} \sum_{\mathrm{i=0}}^{\mathrm{M}} \left(\mathbf{y}_i - H(\mathbf{x}_i)\right)^{\mathrm{T}} (\mathbf{R}_i)^{-1} \left(\mathbf{y}_i - H(\mathbf{x}_i)\right) \quad + \\ \left(\mathbf{x}^{\mathrm{b}}(t_0) - \mathbf{x}(t_0)\right)^T \left(\mathbf{P}_0^{\mathrm{b}}\right)^{-1} \left(\mathbf{x}^{\mathrm{b}}(t_0) - \mathbf{x}(t_0)\right) \end{array}\right]$$

where M is the number of observational vectors $\mathbf{y_i}$ distributed over time.

The control variable (the variable with respect to which the cost function is minimized) is the *initial* state of the model at the beginning of the time interval, $\mathbf{x}(t_0)$, whereas the analysis at the end of the interval is given by the *model integration* from the solution $\mathbf{x}(t_n) = M_0\left[\mathbf{x}(t_0)\right]$.

In other words, 4DVAR seeks the initial condition such that the forecast best fits the observations within the assimilation interval. The 4DVAR tries to use all observations in the assimilation time interval as well as possible.

## Numerical minimization of 4DVAR cost function

- Similar to 3DVAR except the gradient calculation involves the calculation of tangent linear version of the model M and its adjoint (the transpose of the tangent linear model).

- A tangent linear model, TLM, is obtained by linearizing the model about the nonlinear trajectory of the model between $t_{i-1}$ and $t_i$, so that if we introduce a perturbation in the initial conditions, the final perturbation is given by

$$\mathbf{x}(t_i) + \delta\mathbf{x}(t_i) = M_{i-1}\left[\mathbf{x}(t_{i-1}) + \delta\mathbf{x}(t_{i-1})\right] = M_{i-1}\left[\mathbf{x}(t_{i-1})\right] + \mathbf{L}_{i-1}\delta\mathbf{x}(t_{i-1}) + O(|\,\delta\mathbf{x}\,|^2)$$

The linear tangent model $\mathbf{L}_{i-1}$ is a matrix that transforms an small initial perturbation at $t_{i-1}$ to the final perturbation at $t_i$.

The TLM equation is then

$$\delta\mathbf{x}(t_i) = \mathbf{L}_{i-1}\delta\mathbf{x}(t_{i-1}) .$$

# Least Square Method

METR 6803
Advanced topics in data assimilation:
Ensemble Kalman filter techniques

Instructor
Dr. Xuguang Wang
xuguang.wang@ou.edu
Sep. 1, 2009

# 1D Example

Data assimilation aims to optimally combine the background forecast/first guess with observations.

Given two independent measures of a true scalar value $x^t$:

$$x_1 = x^t + \varepsilon_1,$$
$$x_2 = x^t + \varepsilon_2.$$

Assume a) the errors are unbiased, i.e.,

$$E(\varepsilon_1) = 0,$$
$$E(\varepsilon_2) = 0,$$

b) the variances of the measurement errors are

$$E(\varepsilon_1^2) = \sigma_1^2,$$
$$E(\varepsilon_2^2) = \sigma_2^2,$$

c) the errors of the two measures are not correlated, i.e.,

$$E(\varepsilon_1 \varepsilon_2) = 0.$$

Now let's estimate $x^t$ from a linear combination of the two measures,

$$x^a = (1-w)\cdot x_1 + w \cdot x_2 = x_1 + w \cdot (x_2 - x_1)$$

$x^a$ will be the best estimate of the true scalar if the weight chosen minimizes the mean square error of $x^a$.

Derive yourself!

We obtain

$$w = \frac{\sigma_1^{\,2}}{\sigma_1^{\,2} + \sigma_2^{\,2}}$$

Best weight

$$x_a = \frac{\sigma_2^{\,2}}{\sigma_1^{\,2} + \sigma_2^{\,2}} \cdot x_1 + \frac{\sigma_1^{\,2}}{\sigma_1^{\,2} + \sigma_2^{\,2}} \cdot x_2$$

$$= x_1 + \frac{\sigma_1^{\,2}}{\sigma_1^{\,2} + \sigma_2^{\,2}} \cdot (x_2 - x_1)$$

More accurate measure given more weight

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

If optimal weight obtained, error variance of the analysis is smaller than that of the two measures.

# What about high dimensional system?

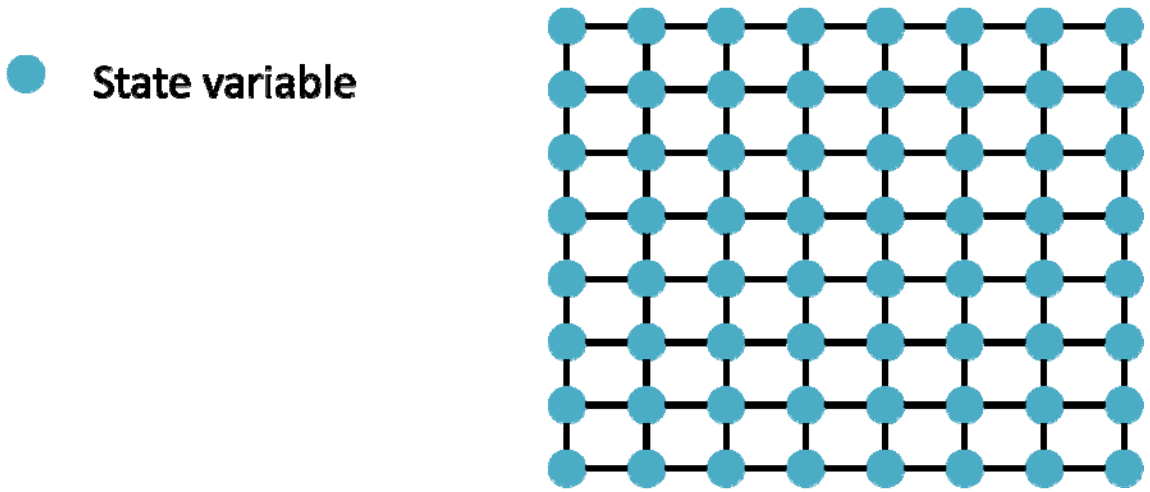
**The DA problem is now defined as finding**

$\mathbf{x}^a$ - an optimum analysis of a field of model variables with length $n$

given

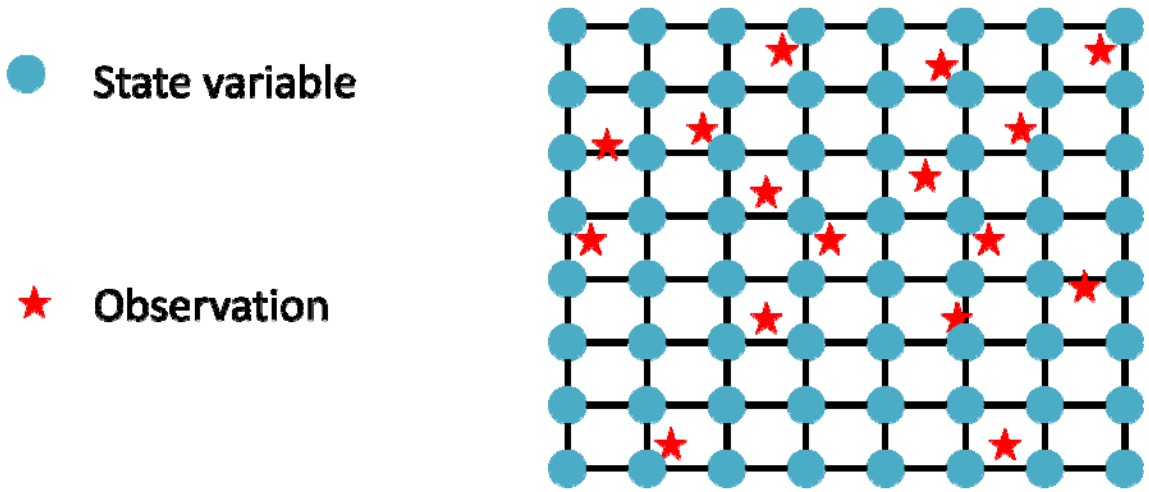
$\mathbf{x}^b$ - a background field available at grid points with length $n$, and
$\mathbf{y}$ - a set of $p$ observations available at irregularly spaced points

$$\mathbf{x}^a = \begin{pmatrix} x_1^a \\ \vdots \\ x_n^a \end{pmatrix} \qquad \mathbf{x}^b = \begin{pmatrix} x_1^b \\ \vdots \\ x_n^b \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}$$




The unknown analysis and the known background can be 2D fields of a single variable like the temperature analysis, or the 3D field of the initial conditions for all the model prognostic variables.

These model variables are ordered by grid point and by variable, forming a single vector of length $n$, the product of the number of points times the number of variables. The (unknown) "truth" $\mathbf{x}^t$, discretized at the model points, is also a vector of length $n$.




A different variable **y** is for the observations.

The observed variables are, in general, different from the model variables by

a)being located in different points, and

b)by possibly being *indirect* measures of the model variables.

# Deriving optimal analysis and analysis error covariance in high dimension

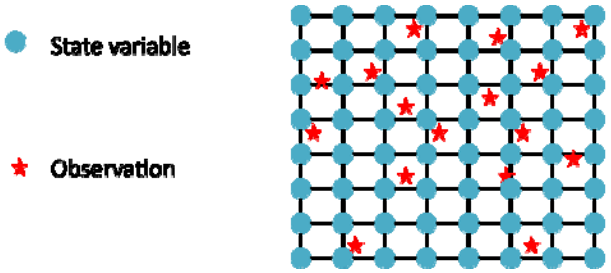
**Definition**
The optimal analysis is equal to the background plus the innovation weighted by optimal weights which are determined so as to minimize the analysis error variance.

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{W}\left(\mathbf{y} - H\left(\mathbf{x}^{\mathrm{b}}\right)\right) = \mathbf{x}^{\mathrm{b}} + \mathbf{W}\mathbf{d}$$

- The weights are given by a matrix of dimension $(n \times p)$, denoted as $\mathbf{W}$.

- $H$ is the forward observational operator that converts the background field into "observed first guesses." $H$ can be nonlinear (e.g., the radiative transfer equations that go from temperature and moisture vertical profiles to the satellite observed radiances) or linear.

radiance = $H$(temperature, moisture, etc)




observation location = $H$(model grid)

- The vector $\mathbf{d}$, also of length $p$, is called the "innovation" or "observational increments" vector:

$$\mathbf{d} = \mathbf{y} - H\left(\mathbf{x}^{\mathrm{b}}\right)$$

which is defined as the difference between the observation and the background mapped to the observational point via forward operator $H$.