# Introduction to Atmospheric Data Assimilation: Basic Concepts and Methodologies

Xiaolei Zou
Florida State University
Email: zou@met.fsu.edu
Phone: 850-644-6025

**FORMOSAT-3/COSMIC Science Summer Camp in Taiwan**

*30 May - 3 June 2005*

# Outline

- Needs --- Why is data analysis/assimilation needed?
- Concerns --- What are concerns of data assimilation?
- Methods --- How is data analysis/assimilation done?
- Details --- How important are details of data assim. in practical applications?
- Summary
- Important areas of DA research

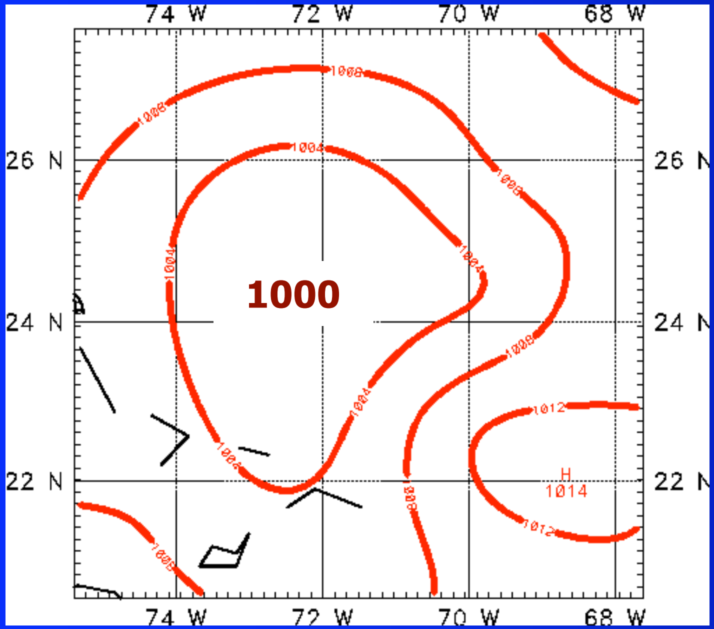
# Why is data analysis/assimilation needed?

**Analysis --- the best estimate of the true state of a physical system at a given time and specified resolution.**

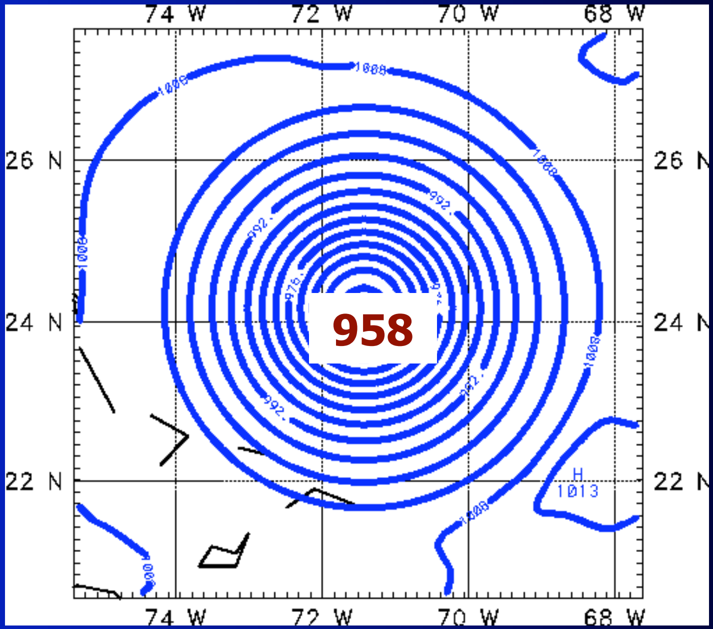
1. Produce analysis of an *incompletely sampled* atmospheric state variable
2. Produce analysis of an *unobserved* atmospheric state variable from observations which are dynamically and/or physically related to the analysis variable
3. Produce analysis of an *overly sampled* atmospheric state variable

# Hurricane Bonnie (1998) at 12 UTC, 23 August 1998

**TPC observed parameters:** $P_c$ = 958 hPa, $R_{max}$=25 km, $V_{max}$=100kt, $R_{34kt}$=255 km.

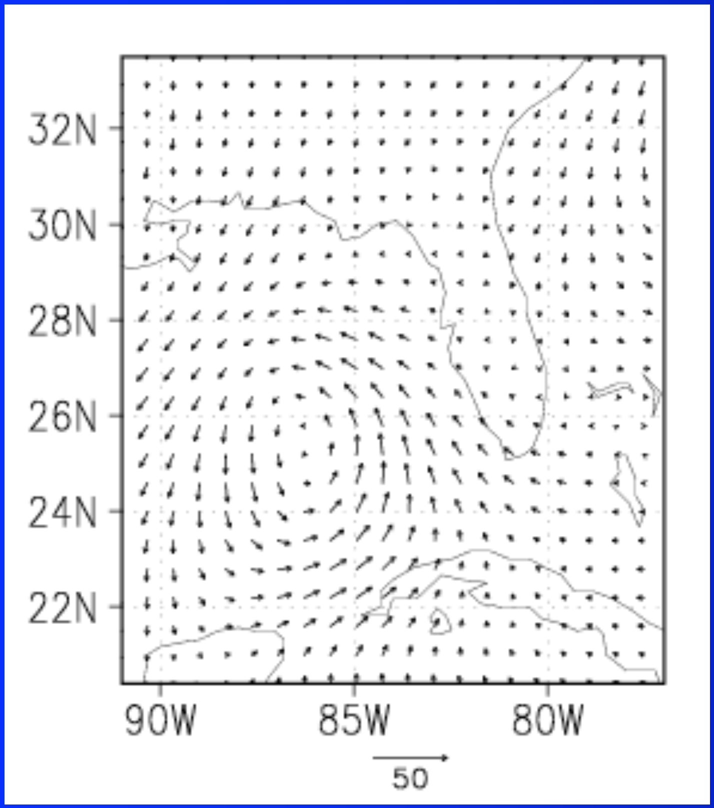



**Large-scale analysis**

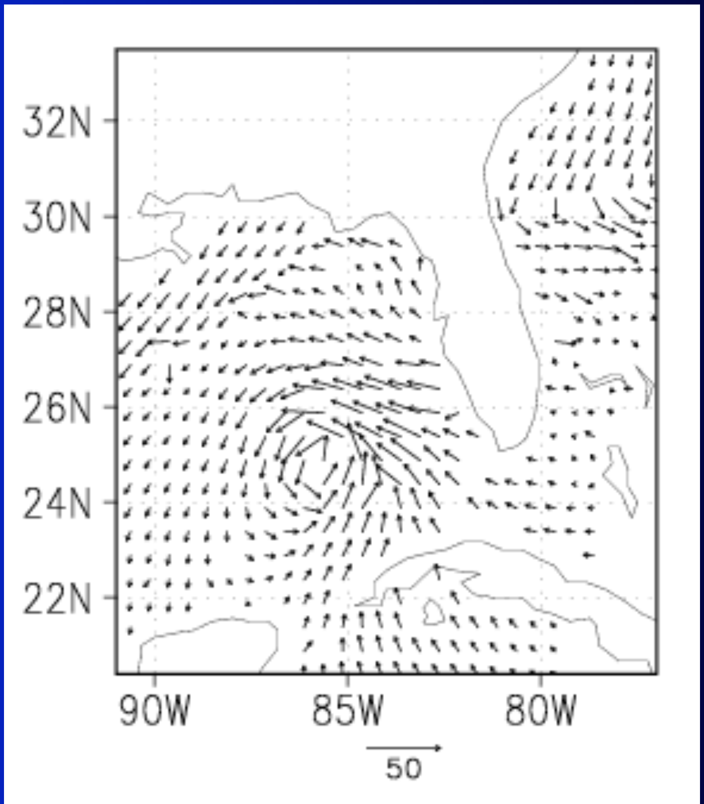



**After bogus data initialization**

# Surface Wind of Hurricane Gordon at 00 UTC, 17 Sept. 2000

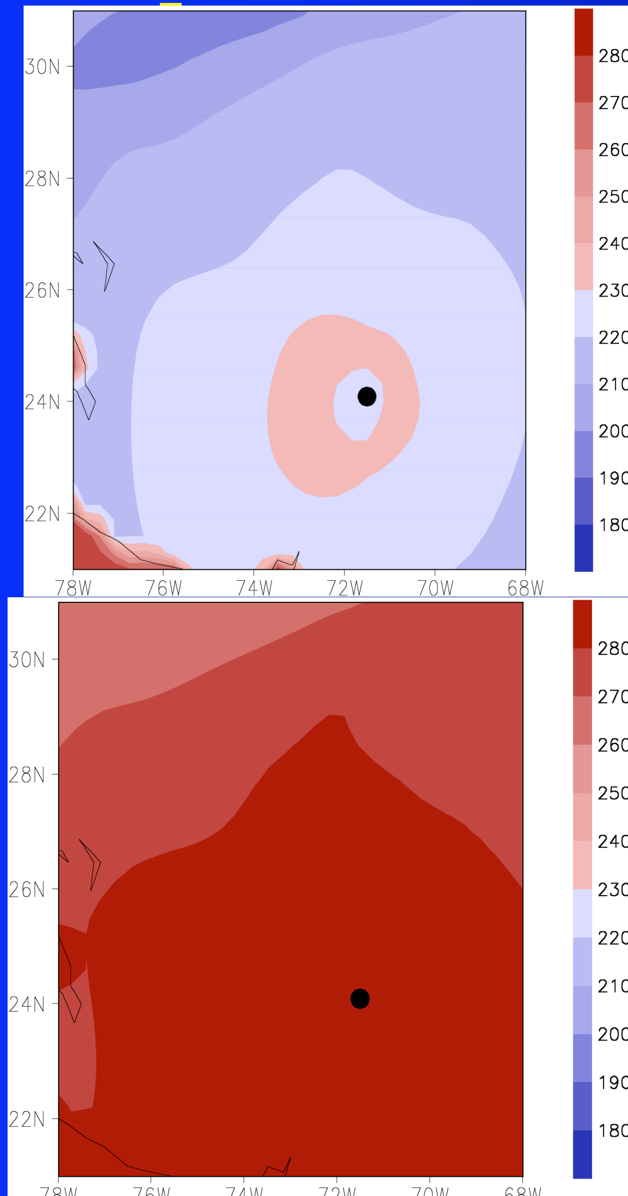



NCEP large-scale analysis

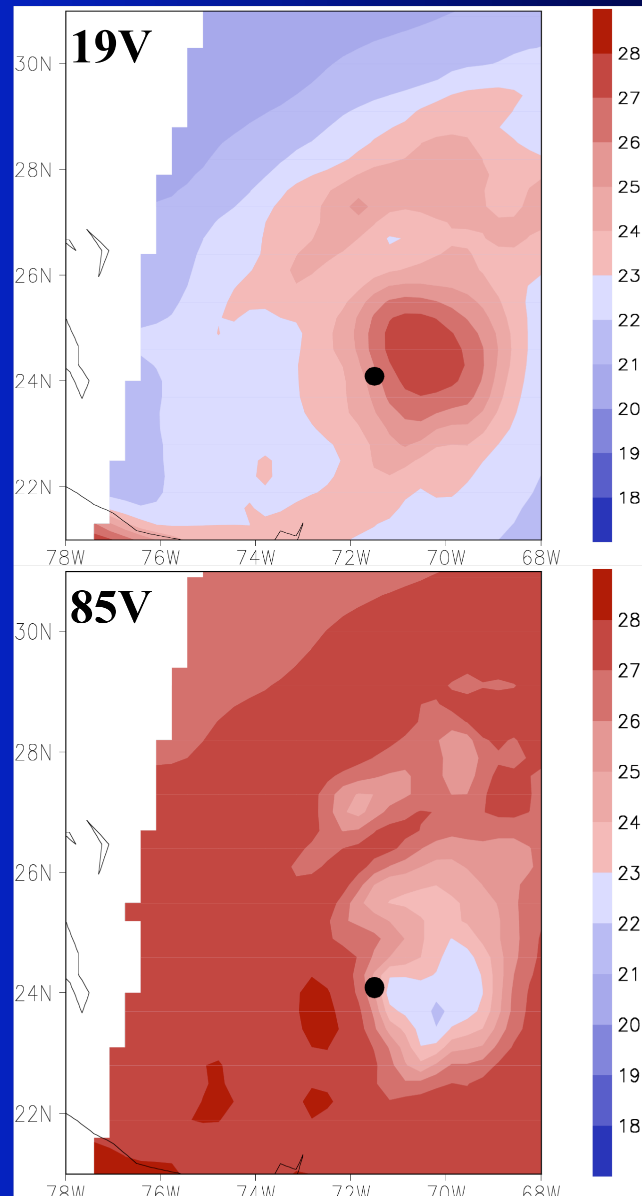



QuikSCAT observations

# SSM/I $T_b$s of Hurricane Bonnie at 1200 UTC 08/23/1998

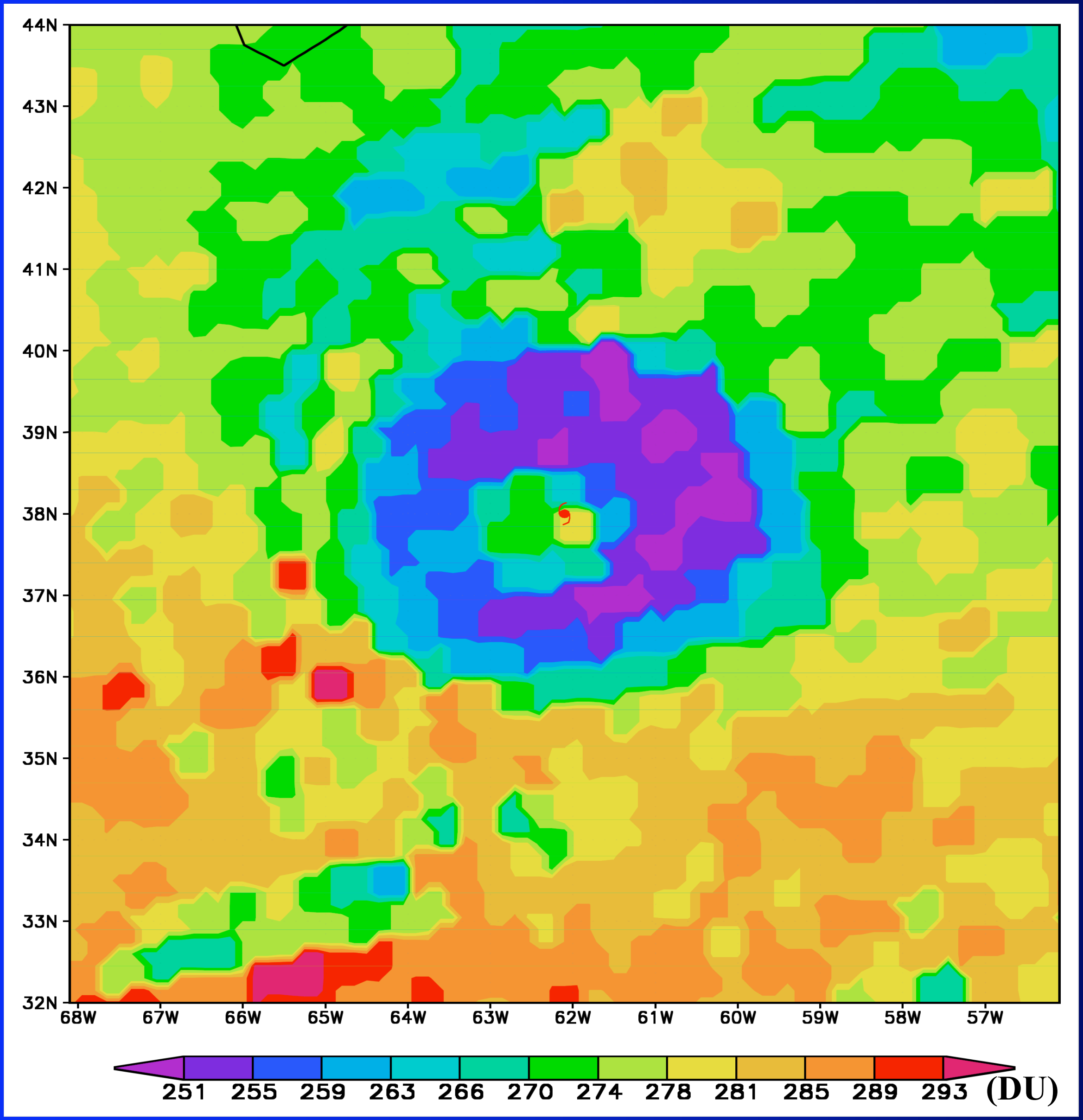



Simulation based on large-scale analysis




SSM/I observations

# Level-2 raw EP/TOMS ozone distribution of Hurricane Erin at 15 UTC 12 September 2001



# A simple function fitting interpolation example

**Function fitting is probably the simplest interpolation method used to produce analysis of incompletely sampled atmospheric state variables.**

*General procedures:*

1. **The analysis variable is expressed in terms of a chosen set of expansion functions**
2. **Coefficients of the expansion are determined by either requiring the analyzed values equal to observed values at observational locations (an exact fit to observations) or through a least-square fit between the analysis and observations within a chosen analysis domain**
3. **The analysis variable is actually a continuous function of its spatial coordinates and hence values of the analyzed variable can be calculated at any specified resolution.**

## *Problem:*

Assume there are two (K=2) zonal wind (u) observations: $u_1^{obs}$ and $u_2^{obs}$ at two spatial locations $x=x_1$ and $x=x_2$ within an interval of $[x_a, x_b]$, where $x_a < x_1 < x_2 < x_b$. Find the analysis of u at any point within the interval $[x_a, x_b]$ through a polynomial function fitting procedure.

### Exact Linear Fit

$$u^a(x) = a_0 + a_1 x$$

$$u^a(x_1) = a_0 + a_1 x_1 \equiv u_1^{obs}$$
$$u^a(x_2) = a_0 + a_1 x_2 \equiv u_2^{obs}$$

$$u^a(x) = \underbrace{\frac{x_2 - x}{x_2 - x_1}}_{W_1} u_1^{obs} + \underbrace{\frac{x - x_1}{x_2 - x_1}}_{W_2} u_2^{obs}$$

**A posteriori weights**

## *Problem:*

Assume there are two (K=2) zonal wind (u) observations: $u_1^{obs}$ and $u_2^{obs}$ at two spatial locations $x=x_1$ and $x=x_2$ within an interval of $[x_a, x_b]$, where $x_a < x_1 < x_2 < x_b$. Find the analysis of u at any point within the interval $[x_a, x_b]$ through a polynomial function fitting procedure.

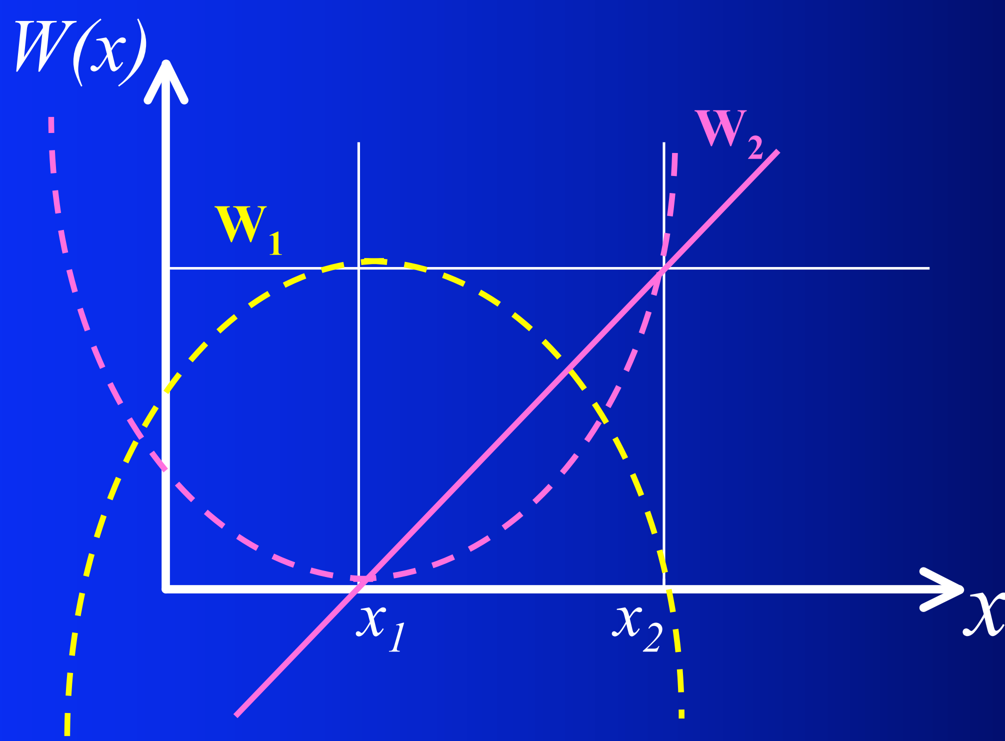
### Exact Quadratic Fit

$$u^a(x) = a_0 + a_1 x^2$$

$$u^a(x_1) = a_0 + a_1 x_1^2 \equiv u_1^{obs}$$
$$u^a(x_2) = a_0 + a_1 x_2^2 \equiv u_2^{obs}$$

$$u^a(x) = \underbrace{\frac{x_2^2 - x^2}{x_2^2 - x_1^2}}_{W_1} u_1^{obs} + \underbrace{\frac{x^2 - x_1^2}{x_2^2 - x_1^2}}_{W_2} u_2^{obs}$$

# *A posteriori* weights




$$u^a(x) = \underbrace{\frac{x_2 - x}{x_2 - x_1}}_{W_1} u_1^{obs} + \underbrace{\frac{x - x_1}{x_2 - x_1}}_{W_2} u_2^{obs}$$

$$u^a(x) = \underbrace{\frac{x_2^2 - x^2}{x_2^2 - x_1^2}}_{W_1} u_1^{obs} + \underbrace{\frac{x^2 - x_1^2}{x_2^2 - x_1^2}}_{W_2} u_2^{obs}$$

## Linear fitting

$$u^a(x) = a_0 + a_1 x$$

$$u^a(x) = \underbrace{\frac{x_2 - x}{x_2 - x_1}}_{W_1} u_1^{obs} + \underbrace{\frac{x - x_1}{x_2 - x_1}}_{W_2} u_2^{obs}$$

## Quadratic fitting

$$u^a(x) = a_0 + a_1 x^2$$

$$u^a(x) = \underbrace{\frac{x_2^2 - x^2}{x_2^2 - x_1^2}}_{W_1} u_1^{obs} + \underbrace{\frac{x^2 - x_1^2}{x_2^2 - x_1^2}}_{W_2} u_2^{obs}$$

## *Some elementary concepts:*

- **Analysis is a weighted sum of observations.**
- **Analysis at the observation location takes observations at that point.**
- **Weightings depend on distances between observation location and analysis point.**
- **The sum of all the weighting coefficients is equal to 1.**
- **Differences between analyses from different function fitting are rather small within two observed stations but could be very large outside the observed area.**

## *Consider observation error:*

Assume there are two (K=2) zonal wind (u) observations: $u_1^{obs}$ and $u_2^{obs}$ at two spatial locations $x=x_1$ and $x=x_2$ within an interval of $[x_a, x_b]$, where $x_a < x_1 < x_2 < x_b$. The observational error variances are known to be $\sigma_k^2$.

### **Least-square fit**

$$u^a(x) = a_0 + a_1 x$$

$$J(a_1, a_2) = \sum_{k=1}^{K} \sigma_k^{-2} \left( a_0 + a_1 x_k - u_k^{obs} \right)^2$$

$$u^a(x) = \underbrace{\left[ b_{11}\sigma_1^{-2} + b_{12}\sigma_1^{-2}(x_1 - x_i) \right]}_{W_1} u_1^{obs} + \underbrace{\left[ b_{11}\sigma_2^{-2} + b_{12}\sigma_2^{-2}(x_2 - x_i) \right]}_{W_2} u_2^{obs}$$

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \equiv \begin{pmatrix} \sigma_1^{-2} + \sigma_2^{-2} & \sigma_1^{-2}(x_1 - x_i) + \sigma_2^{-2}(x_2 - x_i) \\ \sigma_1^{-2}(x_1 - x_i) + \sigma_2^{-2}(x_2 - x_i) & \sigma_1^{-2}(x_1 - x_i)^2 + \sigma_2^{-2}(x_2 - x_i)^2 \end{pmatrix}^{-1}$$

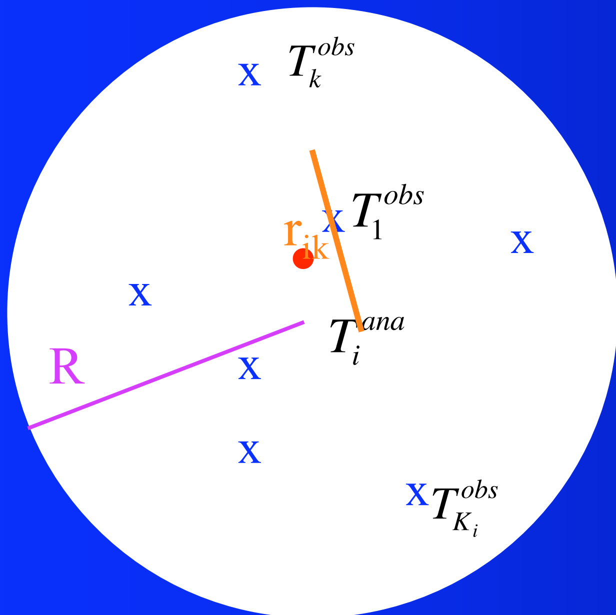
# What does data assimilation do?

**Produce an analysis which combines information in background field, time distributed observations and a dynamic model.**

***Important (ideal) considerations:***

- **Observations are fitted to within (presumed) observation error.**
- **Background information is included.**
- **Observations are enough to over-determine the problem.**
- **Background, observational and model errors are accounted for.**
- **Appropriate dynamic and physical constraints are incorporated.**
- **Noises are suppressed.**
- **Analysis error statistics are known.**

# Successive Corrections (SCs)




*Construct K+1 estimates:*

| | **K+1 estimates** | **variances** |
|---|---|---|
| **obs** | $T_1^c = T_i^b + \left(T_1^{obs} - T_1^b\right)$ | $\sigma_{c1}^2 = \frac{\sigma_o^2}{w(r_{i1})}$ |
| | $T_2^c = T_i^b + \left(T_2^{obs} - T_2^b\right)$ | $\sigma_{c2}^2 = \frac{\sigma_o^2}{w(r_{i2})}$ |
| | … | … |
| | $T_{K_i}^c = T_i^b + \left(T_{K_i}^{obs} - T_{K_i}^b\right)$ | $\sigma_{cK_i}^2 = \frac{\sigma_o^2}{w(r_{iK_i})}$ |
| **bg** | $T^b$ | $\sigma_b^2$ |

**Minimum variance estimate:**

$$T_i^{ana} = \frac{\sigma_b^{-2} T_i^b + \sum_{k=1}^{K_i} \sigma_{ck}^{-2} T_k^c}{\sigma_b^{-2} + \sum_{k=1}^{K_i} \sigma_{ck}^{-2}} = T_i^b + \sum_{k=1}^{K_i} \underbrace{\frac{\sigma_{ok}^{-2} w(r_{ik})}{\sigma_b^{-2} + \sum_{k=1}^{K} \sigma_{ok}^{-2} w(r_{ik})}}_{W_{ki}} \left(T_k^{obs} - T_k^b\right)$$

# Weighting function in SCs

$$T_k^c = T_i^b + \left(T_k^{obs} - T_k^b\right) \quad \Leftrightarrow \quad \sigma_{ck}^2 = \frac{\sigma_o^2}{w(r_{ki})}$$

Weighting functions for observation increments are specified *a priori* as a monotonically decreasing function of the distance between an observation station and an analysis point:

$$w(r) = \begin{cases} \dfrac{R^2 - r^2}{R^2 + r^2} & \text{(Cressman)} \\ e^{-\frac{4r^2}{R^2}} & \text{(Barnes)} \end{cases}$$

The successive corrections method is a local scheme. The analysis is carried out point by point and only observations that lie within the radius of influence (R) of the analysis grid are allowed to influence the analysis.

# Remarks on SCs

$$T_i^{ana} = T_i^b + \sum_{k=1}^{K_i} \underbrace{\frac{\sigma_{ok}^{-2} w(r_{ik})}{\sigma_b^{-2} + \sum_{k=1}^{K} \sigma_{ok}^{-2} w(r_{ik})}}_{W_{ki}} \left(T_k^{obs} - T_k^b\right)$$

*Advantages of SCs over function fitting:*

- A background field is introduced into analysis procedure.
- Observation increments are analyzed to produce analysis increments.
- Weighting functions for observation increments are specified *a priori*.

*Assumptions:*

- Background errors are unbiased, uncorrelated and homogeneous.
- Observation errors are unbiased and uncorrelated.
- Observation errors are not correlated with background errors.
- The $K_i$ estimates, $T_k^c$ (k=1,2,…,K), and their error variances are crudely constructed.

# Optimal Interpolation (OI)

## SCs:

$$T_i^{ana} = T_i^b + \sum_{k=1}^{K_i} \underbrace{\frac{\sigma_{ok}^{-2} w(r_{ik})}{\sigma_b^{-2} + \sum_{k=1}^{K} \sigma_{ok}^{-2} w(r_{ki})}}_{W_{ki}} \left(T_k^{obs} - T_k^b\right) \equiv T_i^b + \sum_{k=1}^{K_i} W_{ki}\left(T_k^{obs} - T_k^b\right)$$

Empirically given in SCs

## OI:

$$x_i^{ana} = x_i^b + \sum_{k=1}^{K_i} W_{ki}\left(x_k^{obs} - x_k^b\right) \equiv x_i^b + \mathbf{w}_i^T\left(\mathbf{x}^{obs} - \mathbf{x}^b\right), \quad \mathbf{w}_i = \begin{pmatrix} W_{1i} \\ W_{2i} \\ \vdots \\ W_{K_i i} \end{pmatrix}$$

$$\mathbf{w}_i = \left(\mathbf{B} + \mathbf{O}\right)^{-1} \mathbf{b}_i$$

$$\sigma_{ai}^2 = \sigma_{bi}^2 - \mathbf{b}_i^T \left(\mathbf{B} + \mathbf{O}\right)^{-1} \mathbf{b}_i \quad \left(\le \overline{\left(x_i - x_i^{true}\right)\left(x_i - x_i^{true}\right)}\right)$$

**Weighting is chosen to give the smallest analysis error variance.**

# Covariances Involved in OI

**Background error covariance matrix:**

$$\mathbf{B}=\begin{pmatrix} \overline{(x_1^b-x_1^t)(x_1^b-x_1^t)} & \overline{(x_1^b-x_1^t)(x_2^b-x_2^t)} & \cdots & \overline{(x_1^b-x_1^t)(x_{K_i}^b-x_{K_i}^t)} \\ \overline{(x_2^b-x_2^t)(x_1^b-x_1^t)} & \overline{(x_2^b-x_2^t)(x_2^b-x_2^t)} & \cdots & \overline{(x_2^b-x_2^t)(x_{K_i}^b-x_{K_i}^t)} \\ \vdots & \vdots & \cdots & \vdots \\ \overline{(x_{K_i}^b-x_{K_i}^t)(x_1^b-x_1^t)} & \overline{(x_{K_i}^b-x_{K_i}^t)(x_2^b-x_2^t)} & \cdots & \overline{(x_{K_i}^b-x_{K_i}^t)(x_{K_i}^b-x_{K_i}^t)} \end{pmatrix}_{K_i\times K_i}$$

**Observation error covariance matrix:**

$$\mathbf{O}=\begin{pmatrix} \overline{(x_1^{obs}-x_1^t)(x_1^{obs}-x_1^t)} & \overline{(x_1^{obs}-x_1^t)(x_2^{obs}-x_2^t)} & \cdots & \overline{(x_1^{obs}-x_1^t)(x_{K_i}^{obs}-x_{K_i}^t)} \\ \overline{(x_2^{obs}-x_2^t)(x_1^{obs}-x_1^t)} & \overline{(x_2^{obs}-x_2^t)(x_2^{obs}-x_2^t)} & \cdots & \overline{(x_2^{obs}-x_2^t)(x_{K_i}^{obs}-x_{K_i}^t)} \\ \vdots & \vdots & \cdots & \vdots \\ \overline{(x_{K_i}^{obs}-x_{K_i}^t)(x_1^{obs}-x_1^t)} & \overline{(x_{K_i}^{obs}-x_{K_i}^t)(x_2^{obs}-x_2^t)} & \cdots & \overline{(x_{K_i}^{obs}-x_{K_i}^t)(x_{K_i}^{obs}-x_{K_i}^t)} \end{pmatrix}_{K_i\times K_i}$$

**Background error covariance vector:**

$$\mathbf{b}_i=\begin{pmatrix} \overline{(x_1^b-x_1^t)(x_i^b-x_i^t)} \\ \overline{(x_2^b-x_2^t)(x_i^b-x_i^t)} \\ \vdots \\ \overline{(x_{K_i}^b-x_{K_i}^t)(x_i^b-x_i^t)} \end{pmatrix}_{K_i\times 1}$$

**About OI:**

$$x_i^{ana} = x_i^b + \overbrace{\underbrace{\mathbf{b}_i^T}_{\text{interpolation}} \underbrace{(\mathbf{B}+\mathbf{O})^{-1}}_{\text{weighting}}}^{\mathbf{w}_i^T} (\mathbf{x}^{obs} - \mathbf{x}^b)$$

$(\mathbf{B}+\mathbf{O})^{-1}(\mathbf{x}^{obs} - \mathbf{x}^b)$:

- Observation increments are weighted by the inverse of the sum of background and observation error covariance matrices.
- Observations that are less accurate or are located over areas where the background field is less accurate are given smaller weights.
- This term does not depend on the position of an analysis grid.

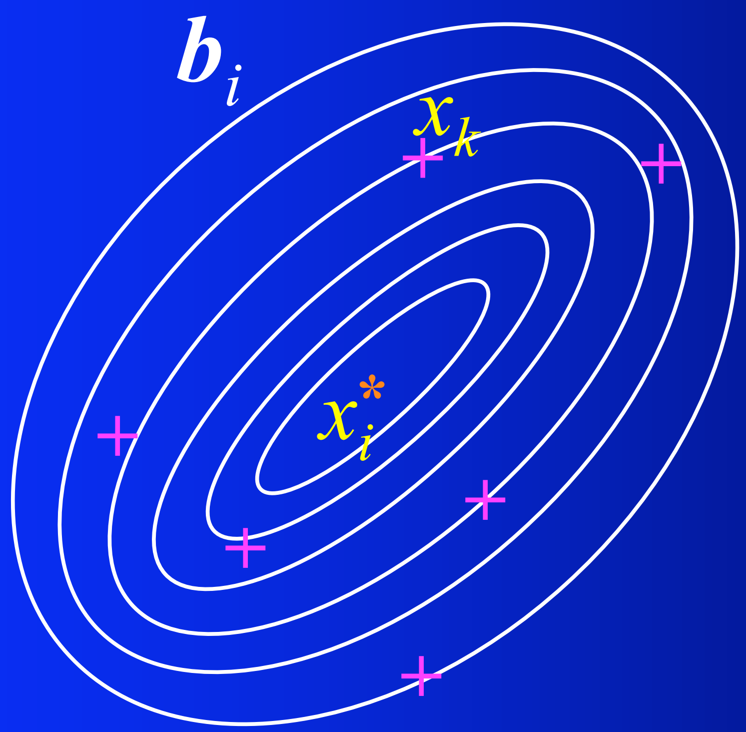
**About OI:**

$$x_i^{ana} = x_i^b + \underbrace{\mathbf{b}_i^T (\mathbf{B} + \mathbf{O})^{-1}}_{\mathbf{w}_i^T} \left(\mathbf{x}^{obs} - \mathbf{x}^b\right)$$

$\mathbf{b}_i^T (\mathbf{B} + \mathbf{O})^{-1} \left(\mathbf{x}^{obs} - \mathbf{x}^b\right)$:

- **Information in observation increments is spread out to the analysis grid based on $b_i$ (the spatial structure of background error covariance).**
- **An observation at a location for which the background error covariance between this location and the analysis point is larger is given a larger weight and thus has a larger impact on the analysis.**

*A larger covariance could imply higher correlation.*




**The OI interpolation strategy based on background error covariance is physically sound and usually produces a better analysis than the function fitting methods where interpolation is determined by the structure of arbitrarily chosen basis functions or SCs in which interpolation is done by an empirically specified weighting function.**

# Remarks on OI

***Common features compared to SCs:***

- A background field is introduced into analysis procedure.
- Observation increments are analyzed to produce analysis increments.
- A data selection procedure is involved which determines the total number of observations that will influence the analysis at a given grid.

***Advantage of OI over SCs:***

- Analysis produced by OI is more accurate than that of SCs.

***Assumption:***

- Observation errors are not correlated with background errors.

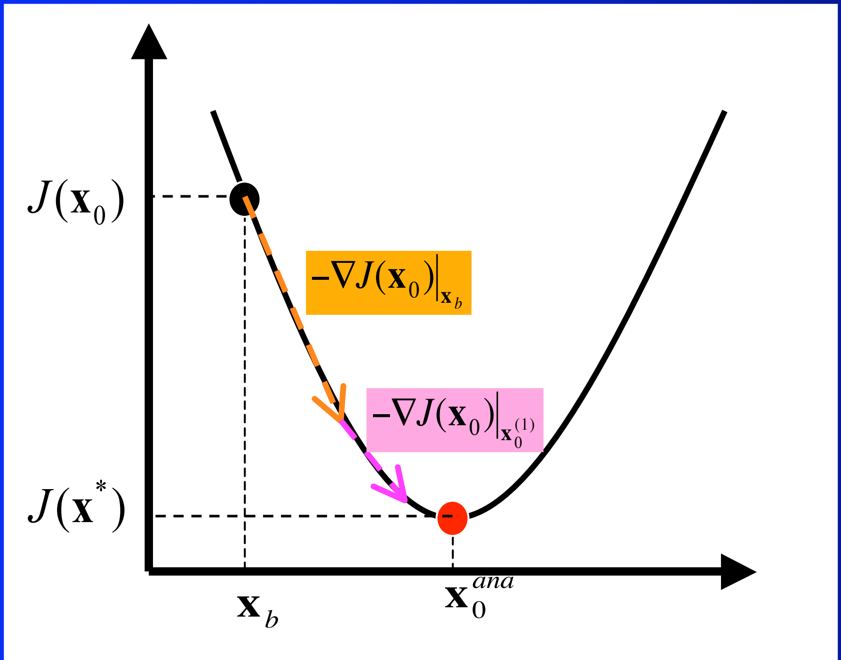
***Challenges:***

- The error covariances in **B**, **O** and $\mathbf{b}_i$ must be estimated.
- The matrix **B**+**O** of order $K_i x K_i$ must be inverted to produce analysis at every grid.

# 3D-Var (1)

**The following scalar cost function is minimized:**

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(H(\mathbf{x}) - \mathbf{y}^{obs})^T (\mathbf{O} + \mathbf{F})^{-1}(H(\mathbf{x}) - \mathbf{y}^{obs})$$



# 3D-Var (2)

**Minimization of J requires the gradient value of J:**

$$\nabla J(\mathbf{x}_0) = \mathbf{H}^T(\mathbf{O}+\mathbf{F})^{-1}(H(\mathbf{x}_0)-\mathbf{y}^{obs}) + \mathbf{B}^{-1}(\mathbf{x}_0-\mathbf{x}^b)$$

$$\mathbf{H} = \left.\frac{\partial H(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}$$ **is the tangent linear model**

$\mathbf{H}^T$ **is called the adjoint model**

$$\mathbf{H}^{\mathrm{T}}$$

$$(\mathbf{O}+\mathbf{F})^{-1}(H(\mathbf{x}_0)-\mathbf{y}^{obs})$$

$$\nabla J(\mathbf{x}_0)$$

$$\mathbf{B}^{-1}(\mathbf{x}_0-\mathbf{x}^b)$$

# Statistical Equivalence of 3D-Var Solution (1)

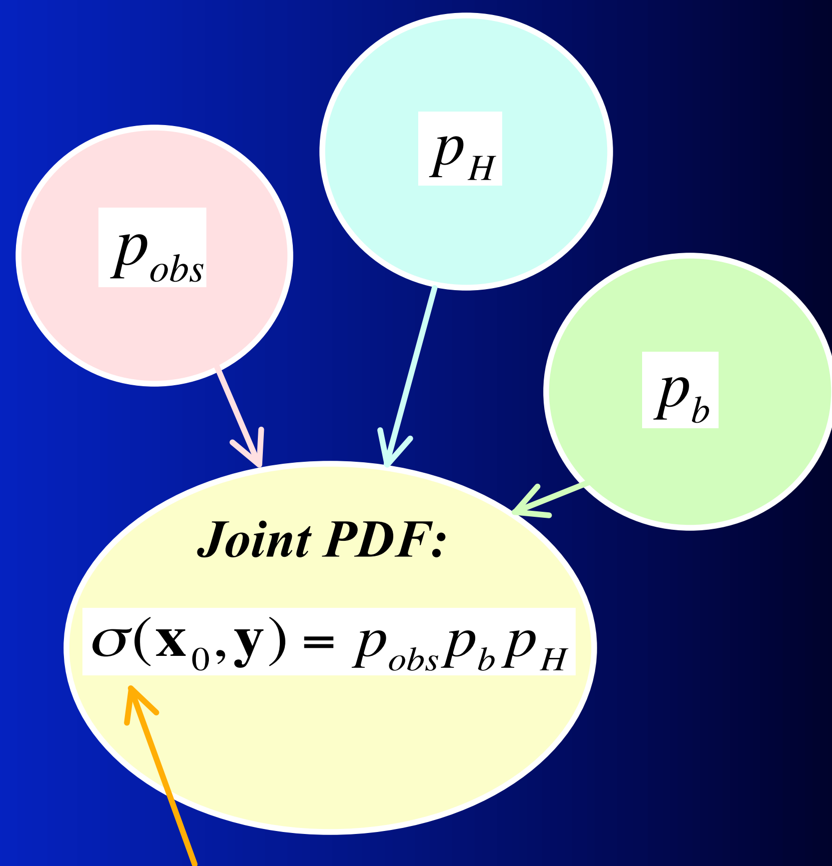
All background fields, models and observations are approximate.

- **Knowledge of the error probability distribution function (PDF) enables the optimal combination of inaccurate information.**
- **There are not sufficient amount of observations and forecasts to quantify accurately these error PDFs.**
- **An approximation is made: PDFs are modeled (assumed) by multi-dimensional Gaussian distributions, which can be described by their mean and covariance.**
- **The DA problem of optimally combining new observations with a background field (a prior estimate of the atmospheric state) becomes tractable under such an approximation.**

# Statistical Equivalence of 3D-Var Solution (1)

**Write the PDFs for all three sources of information as**

| *Available information* | PDF |
|---|---|
| $\mathbf{y}^{obs}$ | $p_{obs}(\mathbf{y} \mid \mathbf{y}^{obs})$ |
| $\mathbf{x}_b$ | $p_b(\mathbf{x}_0 \mid \mathbf{x}^b)$ |
| $H(\mathbf{x}_0)$ | $p_H(\mathbf{y} \mid H(\mathbf{x}_0))$ |




**PDF of the *a posteriori* state of information**

# Statistical Equivalence of 3D-Var Solution (2)

**The marginal PDF of the *a posteriori* state of information:**

$$\sigma(\mathbf{x}_0) = \int \sigma(\mathbf{x}_0, \mathbf{y}) d\mathbf{y}$$

$$= p_b(\mathbf{x}_0 \mid \mathbf{x}^b) \int p_{obs}(\mathbf{y} \mid \mathbf{y}^{obs}) p_H(\mathbf{y} \mid H(\mathbf{x}_0)) d\mathbf{y}$$

**is the PDF of the a posteriori state of information in model space.**

***(The Bayes theorem)***

**Data assimilation only derives some features of this *a posteriori* PDF, such as the maximum likelihood estimate (analysis) and the covariance matrix (analysis error covariance) .**

# Statistical Equivalence of 3D-Var Solution (3)

**The PDFs for the observed value $y^{obs}$, the background value $x_b$, and the forward model $y=H(x_0)$ are all Gaussian:**

$$p_{obs}(\mathbf{y}^{obs},\mathbf{y}) = C_1 \exp\left(-\frac{1}{2}\left(\mathbf{y}-\mathbf{y}^{obs}\right)^T \mathbf{O}^{-1}\left(\mathbf{y}-\mathbf{y}^{obs}\right)\right)$$

$$p_b(\mathbf{x}_b,\mathbf{x}_0) = C_2 \exp\left(-\frac{1}{2}\left(\mathbf{x}_0-\mathbf{x}_b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}_0-\mathbf{x}_b\right)\right)$$

$$p_H(\mathbf{y},\mathbf{x}_0) = C_3 \exp\left(-\frac{1}{2}\left(\mathbf{y}-H(\mathbf{x}_0)\right)^T \mathbf{F}^{-1}\left(\mathbf{y}-H(\mathbf{x}_0)\right)\right)$$

$\Downarrow$ $\Downarrow$ $\Downarrow$

$$\sigma(\mathbf{x}_0) = C \exp\left(-\frac{1}{2}\left(\underbrace{\left(\mathbf{x}_0-\mathbf{x}_b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}_0-\mathbf{x}_b\right) + \left(H(\mathbf{x}_0)-\mathbf{y}^{obs}\right)^T (\mathbf{O}+\mathbf{F})^{-1}\left(H(\mathbf{x}_0)-\mathbf{y}^{obs}\right)}_{J(\mathbf{x}_0)}\right)\right)$$

**Maximize** $\sigma(\mathbf{x}_0)$ $\Leftrightarrow$ **Minimize** $J(\mathbf{x}_0)$

# Statistical Equivalence of 3D-Var Solution (4)

**Maximize** $\sigma(\mathbf{x}_0)$ $\Leftrightarrow$ **Minimize** $J(\mathbf{x}_0)$

*Maximum likelihood estimate* *3D-Var problem*

**Therefore, 3D-Var solves a general inverse problem using maximum likelihood estimate under the assumptions that all errors are Gaussian.**

# 3D-Var with a Linear Forward model

$$\mathbf{y} = \mathbf{H}\mathbf{x}_0$$

**The *a posteriori* PDF is Gaussian, with the following mean and covariance matrix:**

$$\bar{\mathbf{x}}_0 = \mathbf{x}_b + \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{y}^{obs} - \mathbf{H}\mathbf{x}_0\right) \quad \textit{(Analysis space form)}$$

$$= \mathbf{x}_b + \mathbf{B}\mathbf{H}^T\left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^{obs} - \mathbf{H}\mathbf{x}_0\right) \quad \text{(Observation space form)}$$

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

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

## 3D-Var with Linear Approximation for the Forward Model

$$\mathbf{y} = H(\mathbf{x}_0) \approx H(\mathbf{x}_b) + \mathbf{H}\left(\mathbf{x}_0 - \mathbf{x}_b\right)$$

if $\mathbf{x}_b$ is not too far from $\mathbf{x}_a$

**The *a posteriori* PDF is approximately Gaussian, with the following mean and covariance matrix:**

$$\mathbf{x}_a \equiv \bar{\mathbf{x}}_0 = \mathbf{x}_b + \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{y}^{obs} - H(\mathbf{x}_b)\right)$$

$$= \mathbf{x}_b + \mathbf{B}\mathbf{H}^T\left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^{obs} - H(\mathbf{x}_b)\right)$$

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

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

$$\mathbf{A}^{-1} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1} \equiv \nabla^2 J$$

# Information Content (1)

**What does** $\mathbf{A}^{-1} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1} \equiv \nabla^2 J$ **imply?**

**A** --- analysis error covariance matrix

||**A**|| --- some norm of **A**

As ||**A**|| decreases the error decreases and ||$\mathbf{A}^{-1}$|| increases.

When the error is small, the information content is large, the value of ||$\mathbf{A}^{-1}$|| is large.

$\mathbf{A}^{-1}$ is referred to as an information content matrix.

# Information Content (2)

**What does** $\mathbf{A}^{-1} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1} \equiv \nabla^2 J$ **imply?**

$\mathbf{R}$ is positive definite

$\mathbf{B}$ is positive definite

$\Longrightarrow$ $\mathbf{A}$ is positive definite

$$\left\|\mathbf{A}^{-1}\right\| \geq \left\|\mathbf{B}^{-1}\right\|$$
$$\left\|\mathbf{A}^{-1}\right\| \geq \left\|R^{-1}\right\|$$

**The information content of the 3D-Var analysis is greater than the information content in either the background or the observations.**

# About the Two Formulations of 3D-Var

## 1. 3D-Var in analysis space:

$$\mathbf{x}_a - \mathbf{x}_b = \Big(\underbrace{\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1}}_{\mathbf{Q}_A}\Big)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{y}^{obs} - H(\mathbf{x}_b)\right)$$

- Advantageous when there are many observations and few gridpoints
- Constraints $\mathbf{g}(\mathbf{x}_a)=0$ (geostrophy, balance eq., and suppress of fast gravity modes) can be imposed weakly or strongly.

## 2. 3D-Var in observation space:

$$\mathbf{x}_a - \mathbf{x}_b = \mathbf{B}\mathbf{H}^T\Big(\underbrace{\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}}_{\mathbf{Q}_O}\Big)^{-1}\left(\mathbf{y}^{obs} - H(\mathbf{x}_b)\right)$$

- Advantageous when there are many gridpoints and few observations.
- Only weak constraints can be added by treating the constraints as extra observations.

# Similarity between OI and 3D-Var

*Background error covariance between observation locations*

**OI:**

$$x_i^{ana} = x_i^b + \mathbf{b}_i^T (\mathbf{B} + \mathbf{O})^{-1} (\mathbf{x}^{obs} - \mathbf{x}^b)$$

*Background error covariance between grid locations observation locations*

**3D-Var:**

$$\mathbf{x}^{ana} = \mathbf{x}^b + \mathbf{B}\mathbf{H}^T (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} (\mathbf{y}^{obs} - H(\mathbf{x}^b))$$

*Observation Space form*

Suppose $\mathbf{y}^{obs} \sim \mathbf{x}$, *H* is simply spatial interpolation ($H$=$\mathbf{H}$, $\mathbf{R}$=$\mathbf{O}$), then

$$\mathbf{B}\mathbf{H}^T \leftrightarrow \mathbf{b}_i^T \text{ (in OI)}$$
$$\mathbf{H}\mathbf{B}\mathbf{H}^T \leftrightarrow \mathbf{B} \text{ (in OI)}$$
$$H(\mathbf{x}^b) \leftrightarrow \mathbf{x}^b \text{ (in OI)}$$

$$\Longrightarrow \quad \text{OI} \approx \text{3D-Var}$$

# Advantages of 3D-Var over OI

- Observations that are not related directly to analysis variables can be more easily assimilated in 3D-Var using a set of general forward models

- All observations could influence the analysis at every gridpoint. A priori data selection is not required. The 3D-Var analysis fields are smoother than OI analysis.

- Constraints $\mathbf{g}(\mathbf{x}_a)=0$ (geostrophy, balance eq., and suppress of fast gravity modes) can be imposed straightforwardly.

# 4D-Var and Representer Algorithms

**4D-Var is the 4-D extension of the 3D-Var in analysis space.**

**Representer method is the 4-D extension of the 3D-Var in observation space.**

In 3D-Var, $H$ includes only the observation operator.

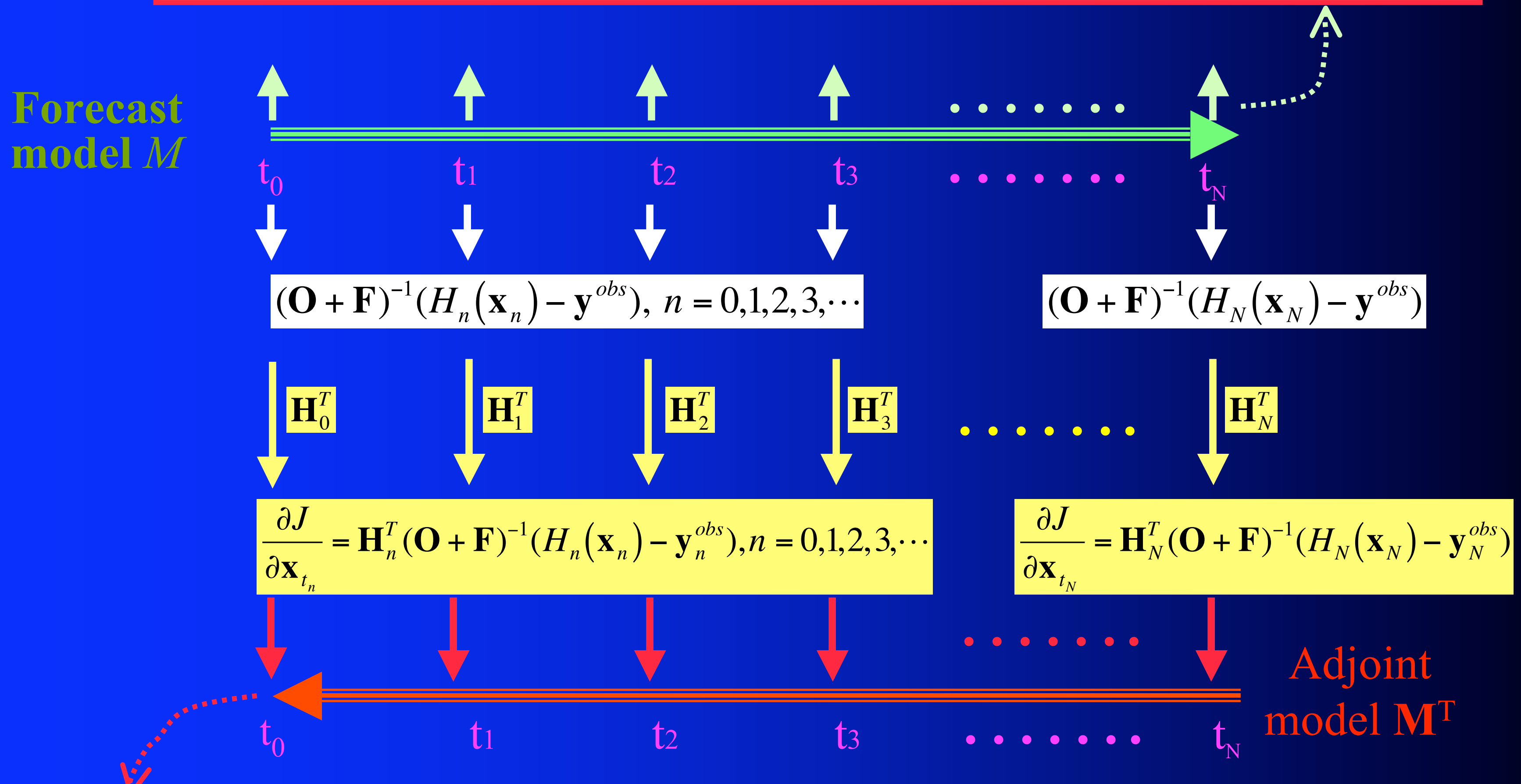
In 4D-Var or representer method, $H$ includes both the forecast model and the observation operator.

$$\text{3D}-\text{Var} \xrightarrow{\text{4D extension}} \text{4D}-\text{Var}$$

$$H(\mathbf{x}_0) \xleftrightarrow{\text{replaces}} H\left(M(\mathbf{x}_0)\right)$$

# A Schematic Illustration of 4D-Var

$$J(\mathbf{x}_0)=\frac{1}{2}(\mathbf{x}_0-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}_0-\mathbf{x}^b)+\frac{1}{2}\sum_{n=0}^{N}(H_n(M_n(\mathbf{x}_0))-\mathbf{y}^{obs})^T(\mathbf{O}+\mathbf{F})^{-1}(H_n(M_n(\mathbf{x}_0))-\mathbf{y}^{obs})$$




$$(\mathbf{O}+\mathbf{F})^{-1}(H_n(\mathbf{x}_n)-\mathbf{y}^{obs}),\ n=0,1,2,3,\cdots$$

$$(\mathbf{O}+\mathbf{F})^{-1}(H_N(\mathbf{x}_N)-\mathbf{y}^{obs})$$

$$\frac{\partial J}{\partial \mathbf{x}_{t_n}}=\mathbf{H}_n^T(\mathbf{O}+\mathbf{F})^{-1}(H_n(\mathbf{x}_n)-\mathbf{y}_n^{obs}),n=0,1,2,3,\cdots$$

$$\frac{\partial J}{\partial \mathbf{x}_{t_N}}=\mathbf{H}_N^T(\mathbf{O}+\mathbf{F})^{-1}(H_N(\mathbf{x}_N)-\mathbf{y}_N^{obs})$$

$$\nabla J(\mathbf{x}_0)=\mathbf{B}^{-1}(\mathbf{x}_0-\mathbf{x}^b)+\sum_{n=0}^{N}\mathbf{M}_n^T\mathbf{H}_n^T(\mathbf{O}+\mathbf{F})^{-1}(H_n(\mathbf{x}_n)-\mathbf{y}^{obs})$$

# Incremental 4D-Var

## 4D-Var

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{n=0}^{N}\left(\mathbf{y} - \mathbf{y}^{obs}\right)^T \mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{y}^{obs}\right)$$

$$\mathbf{y} = H_n\left(M_n(\mathbf{x}_0)\right)$$

## Incremental 4D-Var

$$J(\delta\mathbf{x}_0) = \frac{1}{2}\left(\delta\mathbf{x}_0\right)^T \mathbf{B}^{-1}\delta\mathbf{x}_0 + \frac{1}{2}\sum_{n=0}^{N}\left(\tilde{\mathbf{y}} - \mathbf{y}^{obs}\right)^T \mathbf{R}^{-1}\left(\tilde{\mathbf{y}} - \mathbf{y}^{obs}\right)$$

$$\tilde{\mathbf{y}} = H_n\left(M_n(\mathbf{x}^b)\right) + \mathbf{H}_n\delta\mathbf{x}_n,\ \delta\mathbf{x}_n = \mathbf{M}_{n-1}\delta\mathbf{x}_{n-1} + \boldsymbol{\eta}_{n-1}$$

*About incremental 4D-Var:*

- Introduced as a cost saving method for operational implementation of 4D-Var
- Justified as a filter of scales and processes not well forecasted by NWP models

# Advantages of 4D-Var

*3D-Var advantages are retained in 4D-Var.*

- Indirect observations can be easily assimilated in 4D-Var.
- All observations could influence the analysis at every gridpoint. A priori data selection is not required.
- Constraints $g(x_a)=0$ can be imposed straightforwardly.

- Allows implementing a 4D covariance model.
- Effective use of the synergistic information in sequential observations (such as tracer field).

# Kalman Filter (1)

***Introducing the following notation:***

$\left\{\mathbf{x}_n^f\right\}$ --- A sequence of state vectors satisfying a forward-time-stepping linear model:

$$\mathbf{x}_n^f = \mathbf{M}(t_n, t_{n-1})\mathbf{x}_{n-1}^{ana} + \boldsymbol{\varepsilon}_n^f, \quad n = 1,2,\cdots$$

$\left\{\mathbf{y}_n^{obs}\right\}$ --- A sequence of observation vectors satisfying the following linear measurement equation:

$$\mathbf{y}_n^{obs} = \mathbf{H}_n \mathbf{x}_n^{true} + \boldsymbol{\varepsilon}_n^{obs}, \quad n = 1,2,\cdots$$

$\left\{\mathbf{x}_n^{true}\right\}$ --- A sequence of the true state vectors

$\mathbf{Q}_n = \overline{\boldsymbol{\varepsilon}_n^f \left(\boldsymbol{\varepsilon}_n^f\right)^T}$ --- The time-evolving forecast error covariance matrix

$\mathbf{R}_n = \overline{\boldsymbol{\varepsilon}_n^{obs} \left(\boldsymbol{\varepsilon}_n^{obs}\right)^T}$ --- The observation error covariance matrix

# Kalman Filter (2)

In KF, DA is carried out at every time step of a forward model integration. At the $n^{th}$ time step,

*(previous cycle)*

$$\left.\begin{array}{c}\mathbf{x}^{ana}_{n-1}, \mathbf{P}^{f}_{n-1}, \mathbf{A}_{n-1} \\ + \\ \mathbf{y}^{obs}_{n}, \mathbf{H}_{n}, \mathbf{R}_{n}, \mathbf{Q}_{n}\end{array}\right\} \longrightarrow \left.\begin{array}{c}\mathbf{x}^{ana}_{n}, \mathbf{P}^{f}_{n}, \mathbf{A}_{n} \\ + \\ \mathbf{y}^{obs}_{n+1}, \mathbf{H}_{n+1}, \mathbf{R}_{n+1}, \mathbf{Q}_{n+1}\end{array}\right\} \dashrightarrow$$

*(new information)*

*KF*

$$\mathbf{x}^{f}_{n} = \mathbf{M}(t_n, t_{n-1})\mathbf{x}^{ana}_{n-1}$$

$$\mathbf{P}^{f}_{n} = \mathbf{M}(t_n, t_{n-1})\mathbf{A}_{n-1}\mathbf{M}^{T}(t_n, t_{n-1}) + \mathbf{Q}_{n-1}$$

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

$$\mathbf{x}^{ana}_{n} = \mathbf{x}^{f}_{n} + \mathbf{K}_{n}\left(\mathbf{y}^{obs}_{n} - \mathbf{H}_{n}\mathbf{x}^{f}_{n}\right)$$

$$\mathbf{A}_{n} = \mathbf{P}^{f}_{n} - \mathbf{P}^{f}_{n}\mathbf{H}^{T}_{n}\left(\mathbf{H}_{n}\mathbf{P}^{f}_{n}\mathbf{H}^{T}_{n} + \mathbf{R}_{n}\right)^{-1}\mathbf{H}_{n}\mathbf{P}^{f}_{n}$$

# Kalman Filter (3)

Restricted to linear models

$$\mathbf{x}_n^f = \mathbf{M}(t_n, t_{n-1})\mathbf{x}_{n-1}^{ana}$$

*Too expensive!*

$$\mathbf{P}_n^f = \mathbf{M}(t_n, t_{n-1})\mathbf{A}_{n-1}\mathbf{M}^T(t_n, t_{n-1}) + \mathbf{Q}_{n-1}$$

Very poorly known

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

$$\mathbf{x}_n^{ana} = \mathbf{x}_n^f + \mathbf{K}_n\left(\mathbf{y}_n^{obs} - \mathbf{H}_n \mathbf{x}_n^f\right)$$

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

The gain matrix $\mathbf{K}_n$ is chosen to produce an analysis (linear unbiased estimate) with minimum analysis error variance under the assumption that $\boldsymbol{\varepsilon}^f$ and $\boldsymbol{\varepsilon}^{obs}$ are both Gaussian white-noise sequences.

# Assumptions Used Kalman Filter

1. The forecast model is linear and model error consists of a Gaussian white-noise sequences

$$\mathbf{x}_n^f = \mathbf{M}(t_n, t_{n-1})\mathbf{x}_{n-1}^{ana} + \boldsymbol{\varepsilon}_n^f, \quad \overline{\boldsymbol{\varepsilon}_m^f \left(\boldsymbol{\varepsilon}_n^f\right)^T} = \mathbf{Q}_n \delta_{mn}$$

2. The observation operator is linear and observation error consists of a Gaussian white-noise sequences

$$\mathbf{y}_n^{obs} = \mathbf{H}_n \mathbf{x}_n^{true} + \boldsymbol{\varepsilon}_n^{obs}, \quad \overline{\boldsymbol{\varepsilon}_m^{obs} \left(\boldsymbol{\varepsilon}_n^{obs}\right)^T} = \mathbf{R}_n \delta_{mn}$$

3. The KF analysis has the minimum error variance of all linear unbiased estimate.

$$\sigma_n^{ana}(\mathbf{K}_n) = \min_{\mathbf{W}_n} \sigma_n(\mathbf{W}_n)$$

$$\sigma_n^{ana}(\mathbf{K}_n) = \overline{\left(\mathbf{x}_n^f + \mathbf{K}_n\left(\mathbf{y}_n^{obs} - \mathbf{H}_n \mathbf{x}_n^f\right) - \mathbf{x}_n^{true}\right)^T \left(\mathbf{x}_n^f + \mathbf{K}_n\left(\mathbf{y}_n^{obs} - \mathbf{H}_n \mathbf{x}_n^f\right) - \mathbf{x}_n^{true}\right)}$$

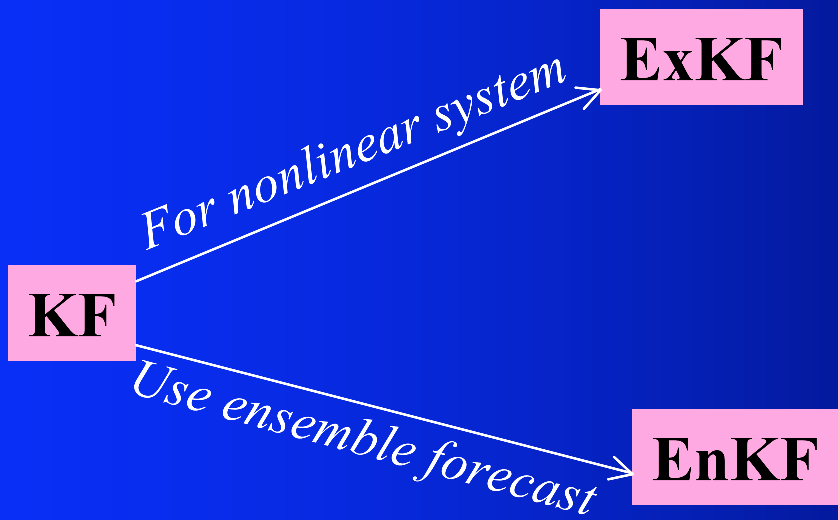
$$\sigma_n(\mathbf{W}_n) = \overline{\left(\mathbf{x}_n^f + \mathbf{W}_n\left(\mathbf{y}_n^{obs} - \mathbf{H}_n \mathbf{x}_n^f\right) - \mathbf{x}_n^{true}\right)^T \left(\mathbf{x}_n^f + \mathbf{W}_n\left(\mathbf{y}_n^{obs} - \mathbf{H}_n \mathbf{x}_n^f\right) - \mathbf{x}_n^{true}\right)}$$

# Approximate KFs

**Kalman Filter (KF)**

**Extended Kalman Filter (ExKF)**

**Ensemble Kalman Filter (EnKF)**



# KF ---> ExKF

*KF*

$$\mathbf{x}_n^f = \mathbf{M}(t_n, t_{n-1})\mathbf{x}_{n-1}^{ana}$$

$$\mathbf{P}_n^f = \mathbf{M}(t_n, t_{n-1})\mathbf{A}_{n-1}\mathbf{M}^T(t_n, t_{n-1}) + \mathbf{Q}_{n-1}$$

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

$$\mathbf{x}_n^{ana} = \mathbf{x}_n^f + \mathbf{K}_n\left(\mathbf{y}_n^{obs} - \mathbf{H}_n \mathbf{x}_n^f\right)$$

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

*ExKF*

$$\mathbf{x}_n^f = M(t_n, t_{n-1})\left(\mathbf{x}_{n-1}^{ana}\right)$$

$$\mathbf{P}_n^f = \mathbf{M}(t_n, t_{n-1})\mathbf{A}_{n-1}\mathbf{M}^T(t_n, t_{n-1}) + \mathbf{Q}_{n-1}$$

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

$$\mathbf{x}_n^{ana} = \mathbf{x}_n^f + \mathbf{K}_n\left(\mathbf{y}_n^{obs} - H_n\left(\mathbf{x}_n^f\right)\right)$$

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

# KF ---> EnKF

Use ensemble forecasts to approximately calculate the forecast error covariance $\mathbf{B}_n$ required in the gain matrix $\mathbf{K}_n$.

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

$$\mathbf{H}_n \mathbf{P}_n^f \mathbf{H}_n^T = \frac{1}{M-1} \sum_{m=1}^{M} \left( H_n\left(\mathbf{x}_n^f(m)\right) - H_n\left(\overline{\mathbf{x}_n^f(m)}\right) \right) \left( H_n\left(\mathbf{x}_n^f(m)\right) - H_n\left(\overline{\mathbf{x}_n^f(m)}\right) \right)^T$$

$$\mathbf{P}_n^f \mathbf{H}_n^T = \frac{1}{M-1} \sum_{m=1}^{M} \left( \mathbf{x}_n^f(m) - \overline{\mathbf{x}_n^f(m)} \right) \left( H_n\left(\mathbf{x}_n^f(m)\right) - H_n\left(\overline{\mathbf{x}_n^f(m)}\right) \right)^T$$

$$\mathbf{x}_n^a(m) = \mathbf{x}_n^f(m) + \mathbf{K}_n \left( \mathbf{y}_n^{obs} + \mathbf{v}_n(m) - H_n\left(\mathbf{x}_n^f(m)\right) \right)$$

# Kalman Filter and 3D-Var

**3D-Var:**

$$\mathbf{x}_0^{ana} = \mathbf{x}_b + \mathbf{B}\mathbf{H}^T\left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^{obs} - \mathbf{H}\mathbf{x}_0\right)$$

**KF:**

$$\mathbf{x}_n^{ana} = \mathbf{x}_n^f + \underbrace{\mathbf{P}_n^f\mathbf{H}_n^T\left(\mathbf{H}_n\mathbf{P}_n^f\mathbf{H}_n^T + \mathbf{R}_n\right)^{-1}}_{\mathbf{K}_n}\left(\mathbf{y}_n^{obs} - \mathbf{H}_n\mathbf{x}_n^f\right)$$

$$\mathbf{P}_n^f = \mathbf{M}(t_n, t_{n-1})\mathbf{A}_{n-1}\mathbf{M}^T(t_n, t_{n-1}) + \mathbf{Q}_{n-1}$$

**The estimate of analysis is similar in KF and 3D-Var if $x^b = x^f$. But in 3D-Var**

- Analysis is done at a longer time interval (6 h).
- The background error covariance is not updated at every estimate in a DA cycle.
- Model error is not considered.

# Kalman Filter and 4D-Var

Incremental 4D-Var ~ Extended Kalman Filter

- Incremental 4D-Var can be viewed as a practical implementation of the ExKF for a finite time window.
- The ExKF is equivalent to 4D-Var at the end of the finite 4D-Var time window.

# Practical Implementation

**Practical implementation of any data assimilation algorithm requires numerous assumptions, approximation and decisions to be made on**

- **Error characteristics of background, observations, and models**
- **Assimilation variables (“raw” observations)**
- **Analysis variables**
- **Statistical and/or dynamical balance at the scales of interest**
- **Incremental approach (3D-var and 4D-Var)**
- **Multi-processor run**

It is extremely important to keep in mind all the assumptions and approximations made in developing the data assimilation system when interpreting results and incorporating new constraints and observations.

# Forecast Covariance **B** (1)

$$\mathbf{B} = \overline{\left(\mathbf{x}^{b} - \mathbf{x}^{true}\right)\left(\mathbf{x}^{b} - \mathbf{x}^{true}\right)^{T}}$$

- **Quantifying likely errors in forecasts for users.**
- **Determining the weights given to observations.**

*How to obtain* **B** *?*

# Covariance Statistics (2)

***Estimating forecast error covariance has always been a challenging problem.***

$$\mathbf{B}_n = \mathbf{M}(t_n, t_{n-1})\mathbf{A}_{n-1}\mathbf{M}^T(t_n, t_{n-1}) + \mathbf{Q}_{n-1}$$

$$\mathbf{A}_{n-1} = \mathbf{B}_{n-1} - \mathbf{B}_{n-1}\mathbf{H}_{n-1}^T\left(\mathbf{H}_{n-1}\mathbf{B}_{n-1}\mathbf{H}_{n-1}^T + \mathbf{R}_{n-1}\right)^{-1}\mathbf{H}_{n-1}\mathbf{B}_{n-1}$$

1. The matrices $\mathbf{B}_n$ and $\mathbf{A}_n$ are too large to evaluate.
2. The input matrices $\mathbf{Q}_n$ are poorly known.
3. There are not enough actual forecasts and validation data.

**Only some structures of $\mathbf{B}_n$ and $\mathbf{A}_n$ are deduced from actual forecasts and observations based on known atmospheric dynamics and physics.**

*It is important to know which properties of the covariances to retain.*

# Covariance Statistics (3)

- **Deviations of the background from radiosondes**

$$B_{kl} \propto \frac{\overline{\left(x_k^{obs} - x_k^b\right)\left(x_l^{obs} - x_l^b\right)}}{\sqrt{\overline{\left(x_k^{obs} - x_k^b\right)^2}}\sqrt{\overline{\left(x_l^{obs} - x_l^b\right)^2}}}$$

- **Differences between lagged forecasts**

$$B \propto \overline{\left(\mathbf{x}^{48\text{-h forecast}} - \mathbf{x}^{24\text{-h forecast}}\right)\left(\mathbf{x}^{48\text{-h forecast}} - \mathbf{x}^{24\text{-h forecast}}\right)^T}$$

- **Ensemble forecasts**

$$\mathbf{B} \propto \overline{\left(\mathbf{x}^{\text{ensemble forecast}} - \overline{\mathbf{x}^{\text{ensemble forecast}}}\right)\left(\mathbf{x}^{\text{ensemble forecast}} - \overline{\mathbf{x}^{\text{ensemble forecast}}}\right)^T}$$

**A variable transform is used to diagonalize B so that $B^{-1}$ can be evaluated.**

# Determining Covariance Statistics Based on Deviations of the Background from Radiosondes

*Under the assumptions:*

- *Background errors time-invariant, homogeneous and isentropic*
- *Observation errors spatially uncorrelated*
- *Background and observation errors uncorrelated*

*then*

$$B_{kl} = \begin{cases} \sigma_b^2 & k = l \\ \sigma_b^2 \rho_b(r_{kl}), & k \neq l \end{cases} \qquad O_{kl} = \begin{cases} \sigma_o^2, & k = l \\ 0 & k \neq l \end{cases}$$

$$Q(r) \approx \frac{\overline{\left(x_k^{obs} - x_k^b\right)\left(x_l^{obs} - x_l^b\right)}}{\sqrt{\overline{\left(x_k^{obs} - x_k^b\right)^2}}\sqrt{\overline{\left(x_l^{obs} - x_l^b\right)^2}}} = \frac{\sigma_b^2 \rho_b(r)}{\sigma_b^2 + \sigma_{obs}^2}$$

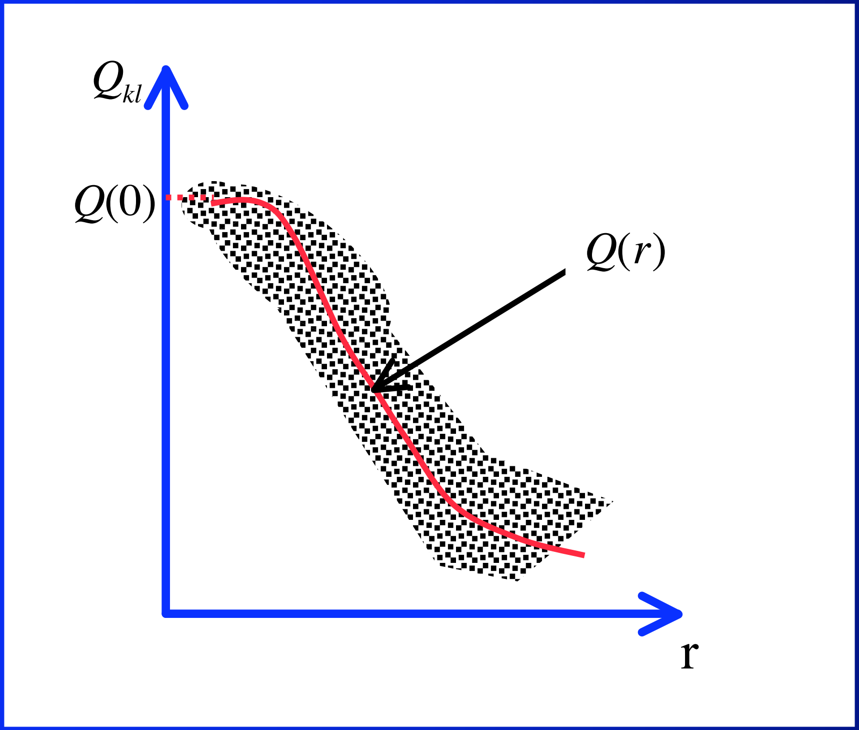
$$Q(0) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_{obs}^2}$$

$$\overline{\left(x_k^{obs} - x_k^b\right)^2} = \sigma_b^2 + \sigma_{obs}^2 \qquad \Longrightarrow \quad \sigma_b^2,\ \sigma_{obs}^2$$

$$\rho_b(r) = \frac{\sigma_b^2 + \sigma_{obs}^2}{\sigma_b^2} Q(r)$$

# Extracting Homogeneous and Isentropic Components in the Covariance between Background and Radiosondes

$$Q_{kl} = \frac{\overline{\left(x_k^{obs} - x_k^{b}\right)\left(x_l^{obs} - x_l^{b}\right)}}{\sqrt{\left(x_k^{obs} - x_k^{b}\right)^2}\sqrt{\left(x_l^{obs} - x_l^{b}\right)^2}} \overset{?}{\Rightarrow} Q(r)$$



# Keeping the Known Dynamic Constraint in Forecast Covariance (1)

In the extratropics, the true and forecast atmospheres are approximately in geostrophic balance. The errors should have the same property.

Given observations of Φ and v at K stations, the analysis of F at the i$^{th}$ grid:

$$\Phi_i^{ana} - \Phi_i^b = \mathbf{W}_{\Phi\Phi}^T\left(\mathbf{\Phi}^{ana} - \mathbf{\Phi}^b\right) + \mathbf{W}_{\Phi v}^T\left(\mathbf{v}^{ana} - \mathbf{v}^b\right)$$

where

$$\left(\mathbf{B}_{\Phi\Phi} + \mathbf{O}_{\Phi\Phi}\right)\mathbf{W}_{\Phi\Phi} + \left(\mathbf{B}_{\Phi v} + \mathbf{O}_{\Phi v}\right)\mathbf{W}_{\Phi v} = B_{\Phi\Phi}(\vec{r}_i)$$

$$\left(\mathbf{B}_{v\Phi} + \mathbf{O}_{v\Phi}\right)\mathbf{W}_{\Phi\Phi} + \left(\mathbf{B}_{vv} + \mathbf{O}_{vv}\right)\mathbf{W}_{\Phi v} = B_{v\Phi}(\vec{r}_i)$$

*If the property of geostrophic balance is retained in the covariance, the evaluation of* **B** *requires less data and the analysis increments also satisfy the geotrophic constraint.*

# Keeping the Known Dynamic Constraint in Forecast Covariance (2)

In the extratropics, the true and forecast atmospheres are approximately in geostrophic balance. The errors should have the same property.

$$u = -\frac{1}{f}\frac{\partial \Phi}{\partial y}, \; v = \frac{1}{f}\frac{\partial \Phi}{\partial x}$$

$$u = -\frac{\partial \psi}{\partial y} + \frac{\partial \chi}{\partial x}, \; v = \frac{\partial \psi}{\partial x} + \frac{\partial \chi}{\partial y}$$

→

$$\Phi = f\psi, \; \chi = 0$$

↓

$$fB_{\Phi\psi} = B_{\Phi\Phi}, \; B_{\Phi\psi} = 0$$

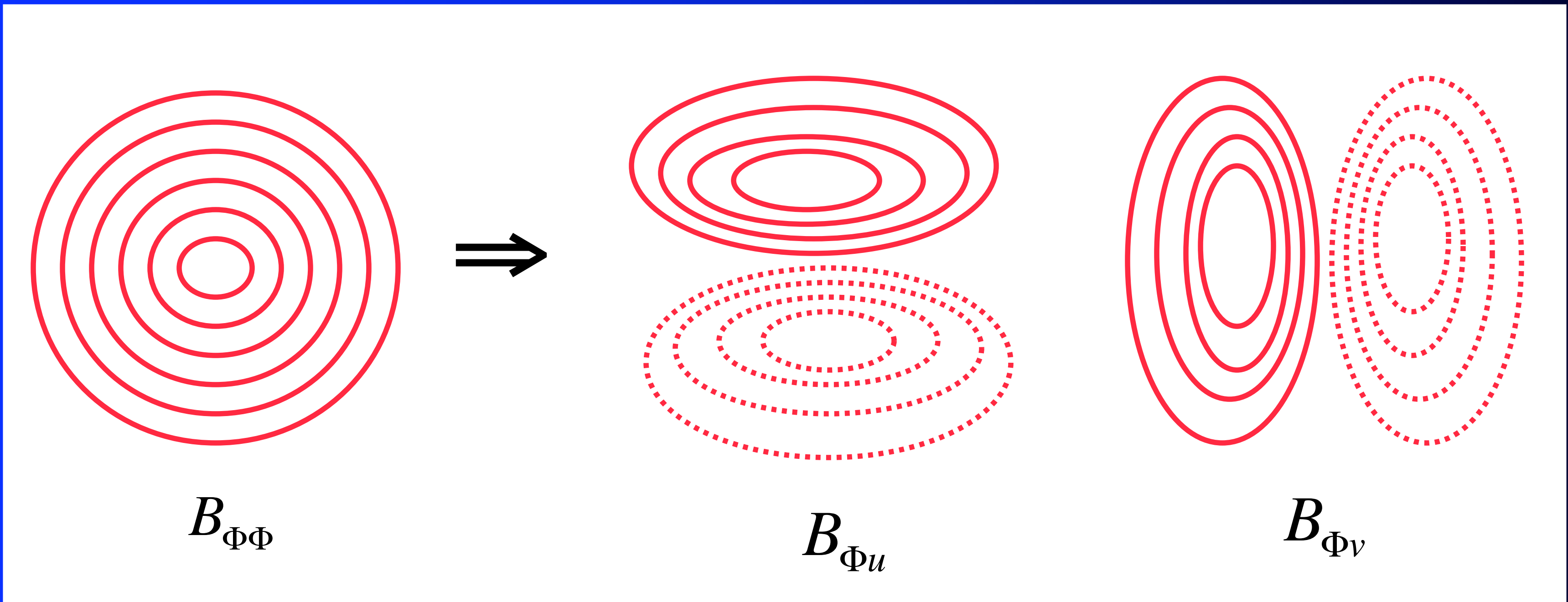
↓

$$B_{\Phi u} = -\sin\phi \frac{1}{f}\frac{\partial}{\partial r} B_{\Phi\Phi}$$

$$B_{\Phi v} = \cos\phi \frac{1}{f}\frac{\partial}{\partial r} B_{\Phi\Phi}$$

*When the properties of geostrophic and hydrostatic balances are retained in the covariance, other components, such as gravity waves, are automatically eliminated.*

# A Multi-Variable Correlation Example




$$B_{\Phi u} = -\sin\phi \frac{1}{f}\frac{\partial}{\partial r} B_{\Phi\Phi}$$

$$B_{\Phi v} = \cos\phi \frac{1}{f}\frac{\partial}{\partial r} B_{\Phi\Phi}$$

# Forecast Error Covariance Model in 3D-Var

**In 3D-Var, a variable transform is used to diagonalize B so that $B^{-1}$ can be evaluated.**

**Original variables x** → **Transformed variables z**

$$\mathbf{x}_0 - \mathbf{x}^b = \mathbf{Lz},\ \mathbf{B}(\mathbf{x}) = \mathbf{LB}(\mathbf{z})\mathbf{L}^T$$

Based on dynamical, physical and mathematical arguments, such as

- Separate balanced and unbalanced variables
- Vertical transform (e.g., EOF)
- Horizontal transform (e.g., spectral transform)

Assume that errors in different components of the transformed variable **z** are uncorrelated, then **B**(**z**) is diagonal (but **B**(**x**) is not). Then the transformed 3D-Var problem becomes

$$J(\mathbf{z}) = \frac{1}{2}\mathbf{z}^T\mathbf{B}^{-1}(\mathbf{z})\mathbf{z} + \frac{1}{2}(\mathbf{y}(\mathbf{z}) - \mathbf{y}^{obs})^T(\mathbf{O}+\mathbf{F})^{-1}(\mathbf{y}(\mathbf{z}) - \mathbf{y}^{obs}),\ \mathbf{y}(\mathbf{z}) = H(\mathbf{x}^b) + \mathbf{HLz}$$

# Implication of Incremental 4D-Var on Covariance Model

$$\delta\mathbf{x}_n = \mathbf{M}_{n-1}\delta\mathbf{x}_{n-1} + \boldsymbol{\eta}_{n-1}$$

$\boldsymbol{\eta}_{n-1}$ Estimate of the stochastic errors at the (n-1)th time step

$\Downarrow$

$$\delta\mathbf{x}_N = \mathbf{M}_{N-1}\mathbf{M}_{N-2}\cdots\mathbf{M}_0\delta\mathbf{x}_0 + \mathbf{M}_{N-1}\mathbf{M}_{N-2}\cdots\mathbf{M}_1\boldsymbol{\eta}_0 + \mathbf{M}_{N-1}\mathbf{M}_{N-2}\cdots\mathbf{M}_2\boldsymbol{\eta}_1 + \cdots + \boldsymbol{\eta}_{N-1}$$

$$= \begin{pmatrix} \mathbf{M}_{N-1}\mathbf{M}_{N-2}\cdots\mathbf{M}_0 & & & \\ & \mathbf{M}_{N-1}\mathbf{M}_{N-2}\cdots\mathbf{M}_1 & & \\ & & \ddots & \\ & & & \mathbf{I} \end{pmatrix} \begin{pmatrix} \delta\mathbf{x}_0 \\ \boldsymbol{\eta}_0 \\ \vdots \\ \boldsymbol{\eta}_{N-1} \end{pmatrix} \equiv L \begin{pmatrix} \delta\mathbf{x}_0 \\ \boldsymbol{\eta}_0 \\ \vdots \\ \boldsymbol{\eta}_{N-1} \end{pmatrix}$$

$\Downarrow$

covariance

$$\mathbf{B}(\mathbf{x}_N) = \mathbf{L} \begin{pmatrix} \mathbf{B}(\mathbf{x}_0) & & & \\ & \mathbf{Q}_0 & & \\ & & \ddots & \\ & & & \mathbf{Q}_{N-1} \end{pmatrix} \mathbf{L}^T$$

The linear model ($\mathbf{M}_n$) extends the covariance relationships to the time dimension.

# Summary (1)

- Function Fitting
- Successive Corrections
- Optimal Interpolation
- 3D-Var
- 4D-Var and Incremental 4D-Var
- KF, ExKF and EnKF

**Analysis of the following form**

$$\mathbf{x}^{ana} = \mathbf{x}^{b} + \mathbf{W}\left(\mathbf{y}^{obs} - \mathbf{y}\right) \qquad *$$

**is derived exactly, approximately or implicitly. Different methods differ in how the *a posteriori* weight W is evaluated and at what time Interval is eq. * implemented.**

# Summary (2)

Atmospheric data assimilation is a process of incorporating various observed information into a NWP model to produce the "best" description of the atmospheric state at desired resolutions in a statistically "optimal" way.

Atmospheric data assimilation is more than an inverse problem in statistics. Physical understanding of what are observed and what structures are we looking for is essential. Knowledge of the computational constraint is also important.

# Areas of Future DA Research

- Assimilation of new observations (GPS RO, …)
- Scale-dependent, weather-dependent background error statistics
- Model errors
- Mesoscale and storm scale balances
- Initialization of ensemble
- New concepts of variable phase correction, analysis of hydrometeor quantities, hurricane initialization, cloud and precipitation, ……
- Coupled DA, climate DA
- Education