

WRFDA Overview

Hans Huang

Acknowledge:

NCAR/MMM/DAG

NCAR/RAL/JNT/DATC

USWRP, NSF-OPP, NCAR, AFWA, KMA, CWB, CAA, EUMETSAT, BMB, AirDat



Outline of Talk

- Introduction to data assimilation.
- Basics of modern data assimilation.
- Demonstration with a simple system.
- WRFDA overview.



Why data assimilation?

- Initial conditions
- Calibration and validation
- Observing system design, monitoring and assessment
- Reanalysis
- Better understanding:
 - Data assimilation methods
 - Model errors
 - Data errors
 - Physical process interactions
 - ...



Modern weather forecast (Bjerknes, 1904)

- A sufficiently accurate knowledge of the state of the atmosphere at the initial time
- A sufficiently accurate knowledge of the laws according to which one state of the atmosphere develops from another.

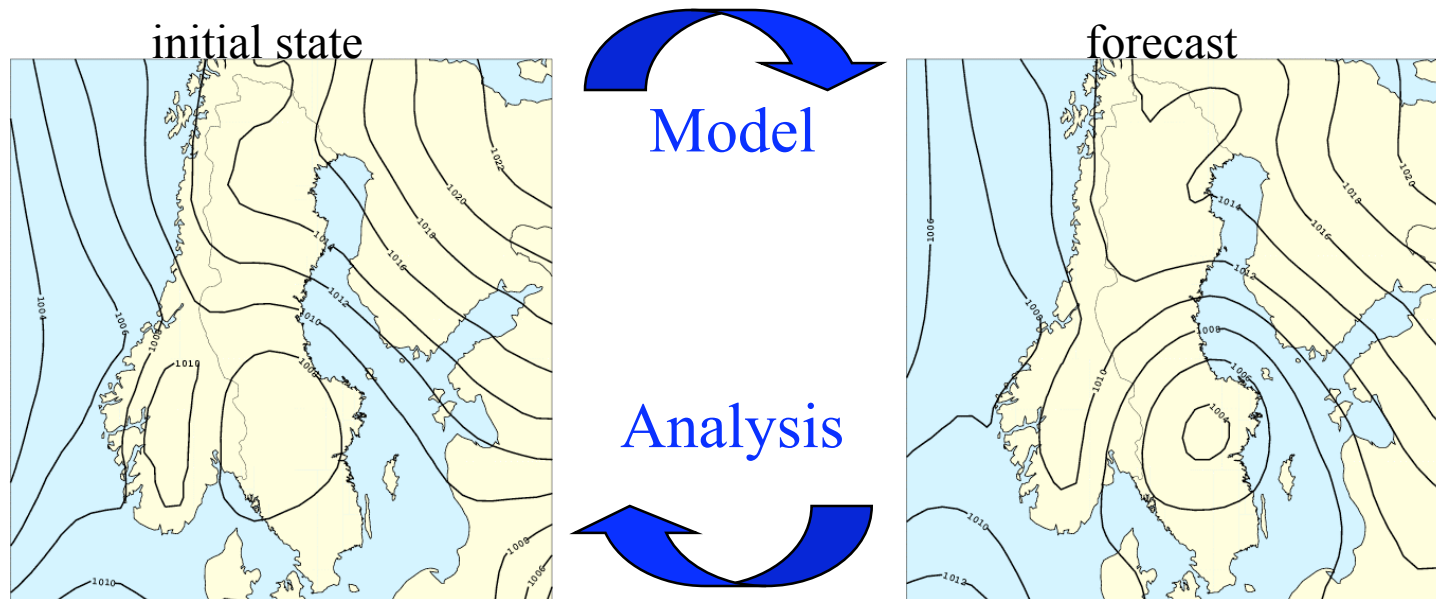


Vilhelm Bjerknes (1862–1951)

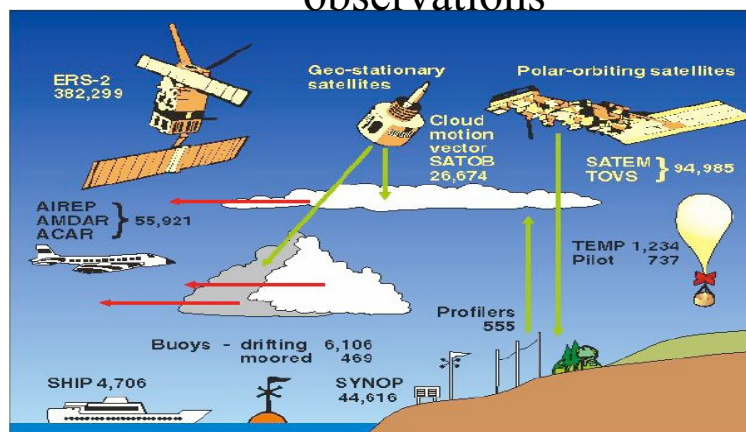
- **Analysis:** using observations and other information, we can specify the atmospheric state at a given initial time: “Today’s Weather”
- **Forecast:** using the equations, we can calculate how this state will change over time: “Tomorrow’s Weather”

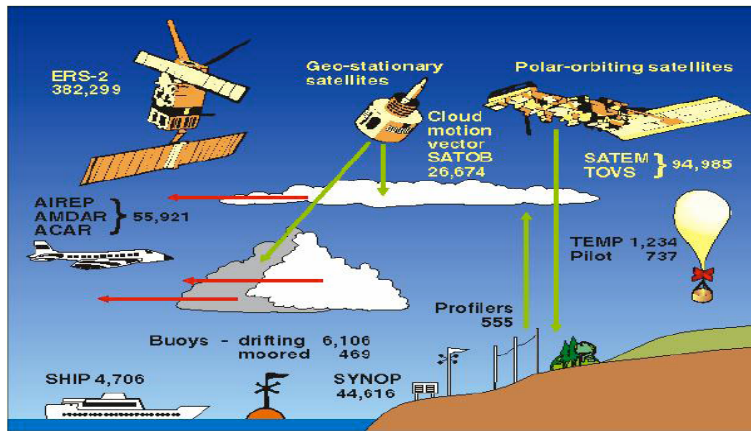
(Peter Lynch)





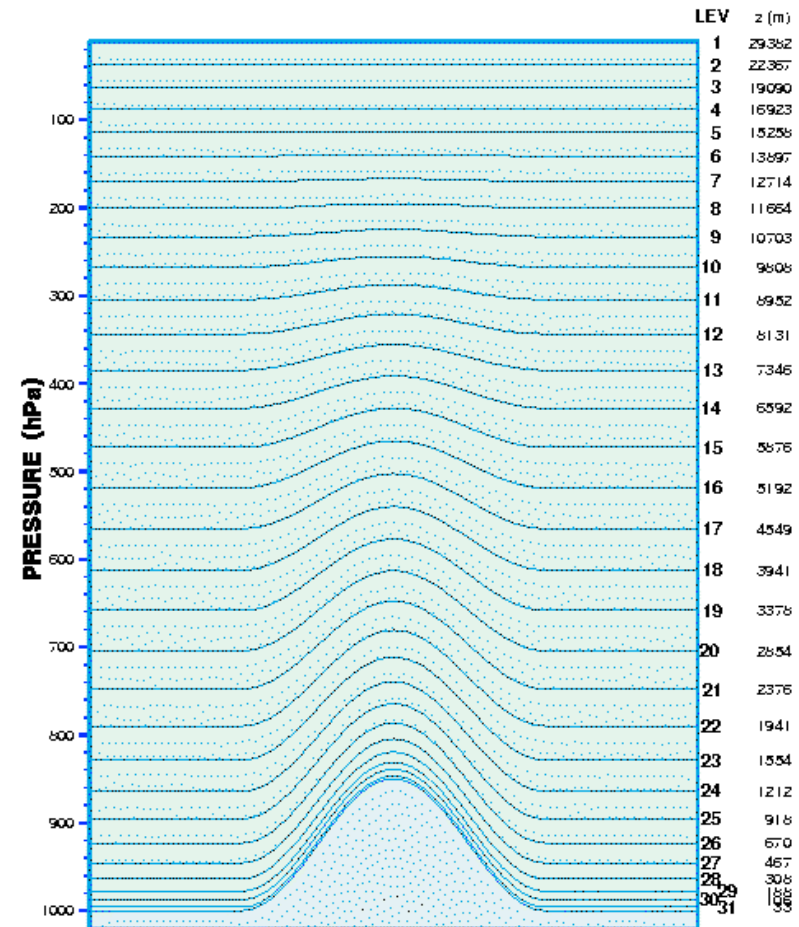
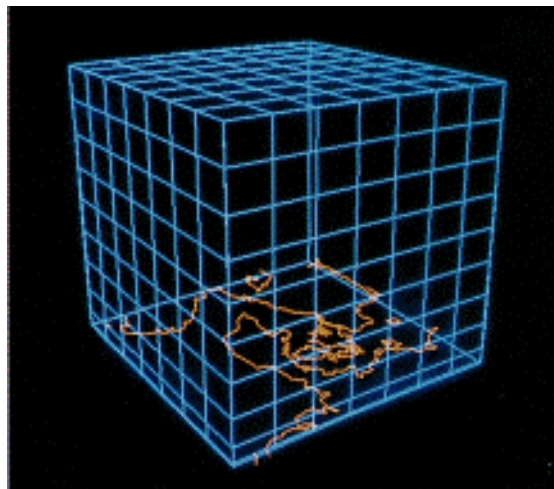
observations





Observations
 $y^0, \sim 10^5 - 10^6$

Model state
 $x, \sim 10^7$



Vertical resolution of the DMI-HIRLAM system



Assimilation methods

- Empirical methods
 - Successive Correction Method (SCM)
 - Nudging
 - Physical Initialisation (PI), Latent Heat Nudging (LHN)
- Statistical methods
 - Optimal Interpolation (OI)
 - 3-Dimensional VARIational data assimilation (3DVAR)
 - 4-Dimensional VARIational data assimilation (4DVAR)
- Advanced methods
 - Extended Kalman Filter (EKF)
 - Ensemble Kalman Filter (EnFK)



Outline of Talk

- Introduction to data assimilation.
- **Basics of modern data assimilation.**
- Demonstration with a simple system.
- WRFDA overview.



The analysis problem for a given time

Consider a scalar x .

The background (normally a short-range forecast):

$$x^b = x^t + b.$$

The observation:

$$x^r = x^t + r.$$

The error statistics are assumed to be known:

$\langle b \rangle = 0$, mean error (unbiased),

$\langle r \rangle = 0$, mean error (unbiased),

$\langle b^2 \rangle = B$, background error variance,

$\langle r^2 \rangle = R$, observation error variance,

$\langle br \rangle = 0$, no correlation between b and r ,

where $\langle \cdot \rangle$ is ensemble average.



BLUE: the Best Liner Unbiased Estimate

The analysis: $x^a = x^t + a$.

Search for the best estimate: $x^a = \alpha x^b + \beta x^r$

Substitute the definitions, we have:

$$\alpha + \beta = 1.$$

$$\langle a \rangle = 0$$

The variance:

$$A = \langle a^2 \rangle = B - 2\beta B + \beta^2 (B + R)$$

$$\text{To determine } \beta: \frac{dA}{d\beta} = -2B + 2\beta(B + R) = 0$$

$$\text{we have } \beta = \frac{B}{B + R}$$

$$\text{The analysis: } x^a = x^b + \frac{B}{B + R} (x^r - x^b)$$

$$\text{The analysis error variance: } A^{-1} = B^{-1} + R^{-1}$$



The analysis: $x^a = x^b + \frac{B}{B+R}(x^r - x^b)$

$$0 < \frac{B}{B+R} < 1$$

The analysis value should be between background and observation.

$$\lim_{B \rightarrow 0} x^a = x^b$$

If B is too small, observations are less useful.

$$\lim_{R \rightarrow 0} x^a = x^r$$

If R can be tuned, analysis can fit observations as close as one wants!

The analysis error variance: $A^{-1} = B^{-1} + R^{-1}$

$A < B$ Statistically, analyses are better than background.

$A < R$ Statistically, analyses are better than observations!



3D-Var

The analysis is obtained by minimizing the cost function J , defined as :

$$J = \frac{1}{2} \left(x - x^b \right)^T B^{-1} \left(x - x^b \right) + \frac{1}{2} \left(x - x^r \right)^T R^{-1} \left(x - x^r \right)$$

The gradient of J with respect to x :

$$J' = B^{-1} \left(x - x^b \right) + R^{-1} \left(x - x^r \right)$$

At the minimum, $J'=0$, we have:

$$x^a = x^b + \frac{B}{B+R} \left(x^r - x^b \right)$$

the same as BLUE.



Sequential data assimilation (I)

True states : ..., x_{i-1}^t , x_i^t , x_{i+1}^t , ...

Observations : ..., x_{i-1}^r , x_i^r , x_{i+1}^r , ...

Forecasts : ..., x_{i-1}^f , x_i^f , x_{i+1}^f , ...

Analyses : ..., x_{i-1}^a , x_i^a , x_{i+1}^a , ...



Sequential data assimilation (II)

Forecast model:

$$x_{i+1}^t = M(x_i^t) + q_i$$

Where q_i is the model error.

As q_i is unknown and x_i^a is the best estimate of x_i^t the forecast model usually takes the form:

$$x_{i+1}^f = M(x_i^a)$$

OI (and 3DVAR):

$$x_i^a = x_i^f + \frac{B}{B+R} \left(x_i^r - x_i^f \right)$$



Sequential data assimilation (III)

4DVAR analysis is obtained by minimizing the cost function J , defined as:

$$J(x_i) = \frac{1}{2} \left(x_i - x_i^f \right)^T B^{-1} \left(x_i - x_i^f \right) + \frac{1}{2} \sum_{k=0}^K \left[M_{k-1}(x_i) - x_{i+k}^r \right]^T R^{-1} \left[M_{k-1}(x_i) - x_{i+k}^r \right]$$

where, K is the assimilation window and

$$M_{-1}(x_i) = x_i$$

$$M_0(x_i) = M(x_i)$$

$$M_{k-1}(x_i) = \underbrace{M(M(\dots M(x_i)\dots))}_k$$



Sequential data assimilation (IV)

4DVAR (continue)

The gradient of J with respect to x :

$$J' = B^{-1} \left(x_i - x_i^f \right) + \sum_{k=0}^K \prod_{j=0}^{k-1} \mathbf{M}_{i+j}^T R^{-1} \left[M_{k-1}(x_i) - x_{i+k}^r \right]$$

where, $\mathbf{M}_{i+j}^T$ is the adjoint model of M at time step $i+j$.



Sequential data assimilation (V)

Extended Kalman Filters:

True states: $x_{i+1}^t = M(x_i^t) + q_i$

Model states: $x_{i+1}^f = M(x_i^a)$

Forecast error: $x_{i+1}^f - x_{i+1}^t = M(x_i^a) - M(x_i^t) - q_i \approx M_i(x_i^a - x_i^t) - q_i$

Forecast error covariance matrix:

$$\begin{aligned} P_{i+1}^f &= \left\langle \left(x_{i+1}^f - x_{i+1}^t \right) \left(x_{i+1}^f - x_{i+1}^t \right)^T \right\rangle \\ &\approx M_i \left\langle \left(x_i^a - x_i^t \right) \left(x_i^a - x_i^t \right)^T \right\rangle M_i^T + \left\langle q_i q_i^T \right\rangle \\ &= M_i P_i^a M_i^T + Q_i \end{aligned}$$



Sequential data assimilation (VI)

Extended Kalman Filters (continue):

For the analysis step:

$$K_i = P_i^f \left(P_i^f + R \right)^{-1}$$

$$x_i^a = x_i^f + K_i \left(x_i^r - x_i^f \right)$$

$$P_i^a = \left(I - K_i \right) P_i^f$$

For the forecast step:

$$x_{i+1}^f = M \left(x_i^a \right)$$

$$P_{i+1}^f = M_i P_i^a M_i^T + Q_i$$

(Ensemble KF use ensembles
to calculate P^f)



Sequential data assimilation (VII)

From scalar to vector:

Number of grid points $\approx 10^7$:

$$x \rightarrow \mathbf{x}$$

$$x^b \rightarrow \mathbf{x}^b$$

Dimension of $\mathbf{B}, \mathbf{P} \approx 10^7 \times 10^7$

Number of observations, 10^6 :

$$x^r \rightarrow \mathbf{y}^o$$

$$x - x^r \rightarrow H(\mathbf{x}) - \mathbf{y}^b$$

Dimension of $\mathbf{R} \approx 10^6 \times 10^6$



Sequential data assimilation (VIII)

OI:

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{B}\mathbf{H}^T \left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R} \right)^{-1} \left[\mathbf{y}^o - H \left(\mathbf{x}_i^f \right) \right]$$

$$\mathbf{x}_{i+1}^f = M \left(\mathbf{x}_i^a \right)$$

4D-Var:

$$J(\mathbf{x}_i) = \frac{1}{2} \left(\mathbf{x}_i - \mathbf{x}_i^f \right)^T \mathbf{B}^{-1} \left(\mathbf{x}_i - \mathbf{x}_i^f \right)$$

$$+ \frac{1}{2} \sum_{k=0}^k \left[H \left(M_{k-1}(\mathbf{x}_i) \right) - \mathbf{y}_{i+k}^o \right]^T \mathbf{R}^{-1} \left[H \left(M_{k-1}(\mathbf{x}_i) \right) - \mathbf{y}_{i+k}^o \right]$$

$$J' = \mathbf{B}^{-1} \left(\mathbf{x}_i - \mathbf{x}_i^f \right) + \sum_{k=0}^K \prod_{j=0}^{k-1} \mathbf{M}_{i+j}^T \mathbf{H}^T \mathbf{R}^{-1} \left[H \left[M_{k-1}(\mathbf{x}_i) \right] - \mathbf{y}_{i+k}^o \right]$$



Sequential data assimilation (IX)

The Extended Kalman Filter:

For the analysis step i :

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

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

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

For the forecast step, from i to $i+1$:

$$\mathbf{x}_{i+1}^f = M \left(\mathbf{x}_i^a \right)$$

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

(Ensemble KF use ensembles
to calculate P^a and P^f)



Sequential data assimilation (X)

EKF

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

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

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

$$\mathbf{x}_{i+1}^f = M \left(\mathbf{x}_i^a \right)$$

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

(Ensemble KF use ensembles to calculate P^a and P^f)

$(\mathbf{B} = \mathbf{P}_{i+1}^f)$ EKF $\rightarrow$ OI or VAR

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

OI or VAR

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

$$\mathbf{x}_{i+1}^f = M \left(\mathbf{x}_i^a \right)$$



Issues on data assimilation

- Observations $\mathbf{y}^o$
- Observation operator H
- Observation errors $\mathbf{R}$
- Background $\mathbf{x}^b$
- Size of $\mathbf{B}$: statistical model and tuning
- Minimization algorithm (Quasi–Newton; Conjugate Gradient; ...)
- $\mathbf{M}$ and $\mathbf{M}^T$: development and validity
- Size of $\mathbf{P}^f$ and $\mathbf{P}^a$: simplifications
- Model errors $\mathbf{Q}$



H - Observation operator

H maps variables from “model space” to “observation space”

$$\mathbf{x} \longrightarrow \mathbf{y}$$

- Interpolations from model grids to observation locations
- Extrapolations using PBL schemes
- Time integration using full NWP models
- Transformations of model variables (u , v , T , q , p_s , etc.) to “indirect” observations (e.g. satellite radiance, radar radial winds, etc.)
 - Simple relations like PW, radial wind, refractivity, ...
 - Radar reflectivity $Z = Z(T, IWC, LWC, RWC, SWC)$
 - Radiative transfer models $L(\nu) \approx \int_0^\infty B(\nu, T(z)) \left[\frac{dTR(\nu)}{dz} \right] dz$
 - Precipitation using simple or complex models
 - ...

!!! Need H , $\mathbf{H}$ and $\mathbf{H}^T$, not $\mathbf{H}^{-1}$!!!



Outline of Talk

- Introduction to data assimilation.
- Basics of modern data assimilation.
- **Demonstration with a simple system.**
- WRFDA overview.



The Lorenz 1964 equation

Nonlinear equation (NL):

$$x_{i+1} = ax_i - x_i^2 = M(x_i)$$

Depending on a , three types of solution are found; steady state; limited cycle; chaotic.

Tangent Linear equations (TL):

$$x_{i+1}^{tl} = \left(a - 2x_i^{bs} \right) x_i^{tl} = M_i x_i^{tl}$$

(linearized around basic state x_i^{bs})

Adjoint equation (AD):

$$x_i^{ad} = \left(a - 2x_i^{bs} \right) x_{i+1}^{ad} = M_i^T x_{i+1}^{ad}$$

Note here for this simple case we have

$$M_i = M_i^T = \left(a - 2x_i^{bs} \right)$$



Issues on data assimilation for the system based on the Lorenz 64 equation

- Observation operator $H = H = H^T = 1$
- Estimate of the true states - generated (with model error!)
- Observation errors $x^o - x^t = \sigma_o G$
- Size of B is 1, but $B = \sigma_b^2$ still needs attention
- Model errors Q: $q_i = \sigma_m G$ When "true" states are generated; but we assume $Q = B/4$
- Size of P^f or P^a is 1.
- M and M^T : no effort in development but their validity is still a major problem
(model, assimilation window length, ...)
- Gaussian statistics



Too simple?

Try: Lorenz63 model

Try: 1-D advection equation

Try: 2-D shallow water equation

...

Try: ARW!!!



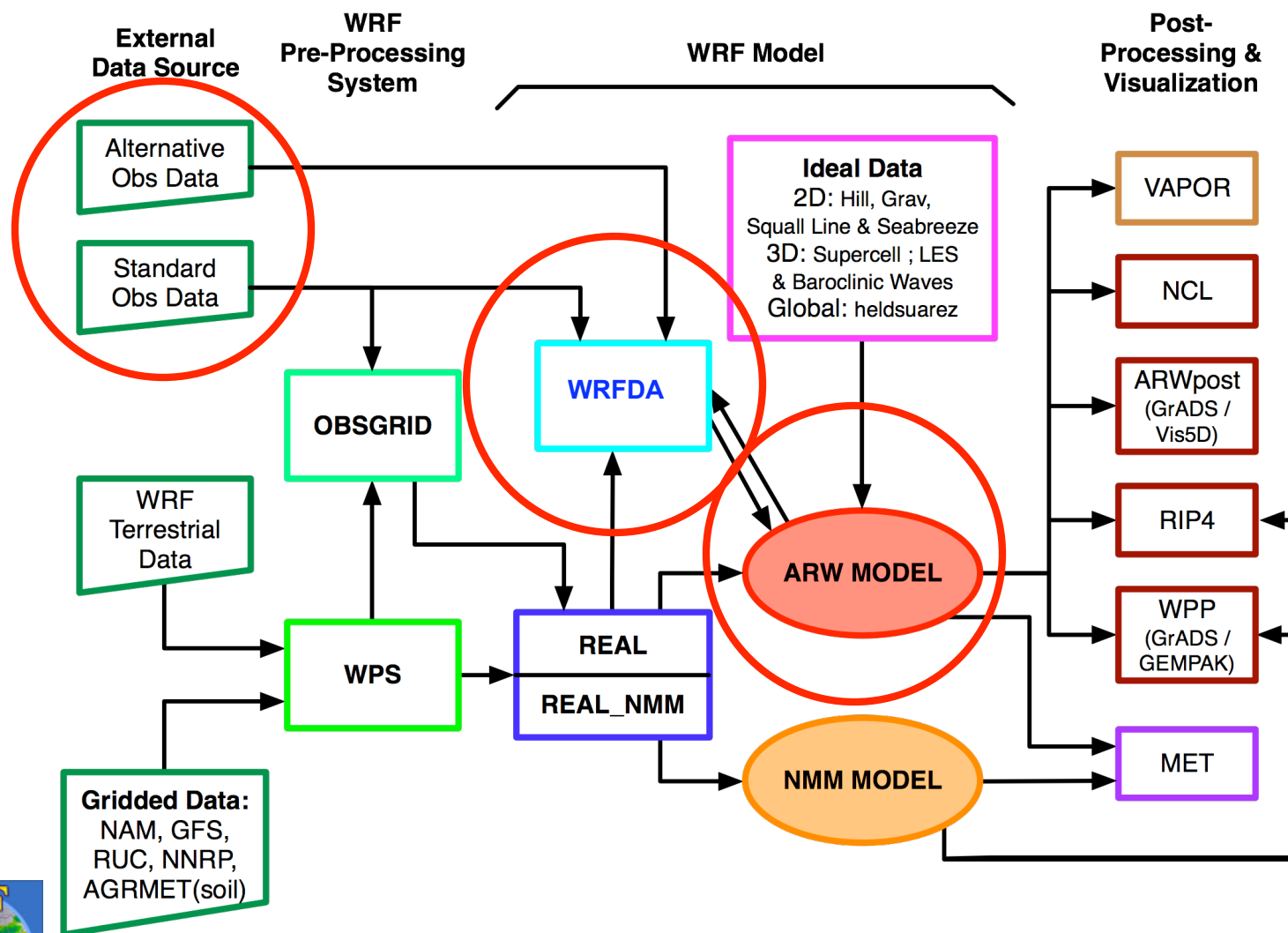
Outline of Talk

- Introduction to data assimilation.
- Basics of modern data assimilation.
- Demonstration with a simple system.
- WRFDA overview...

WRFDA is a **D**ata **A**ssimilation system built within the **WRF** software framework, used for application in both research and operational environments....

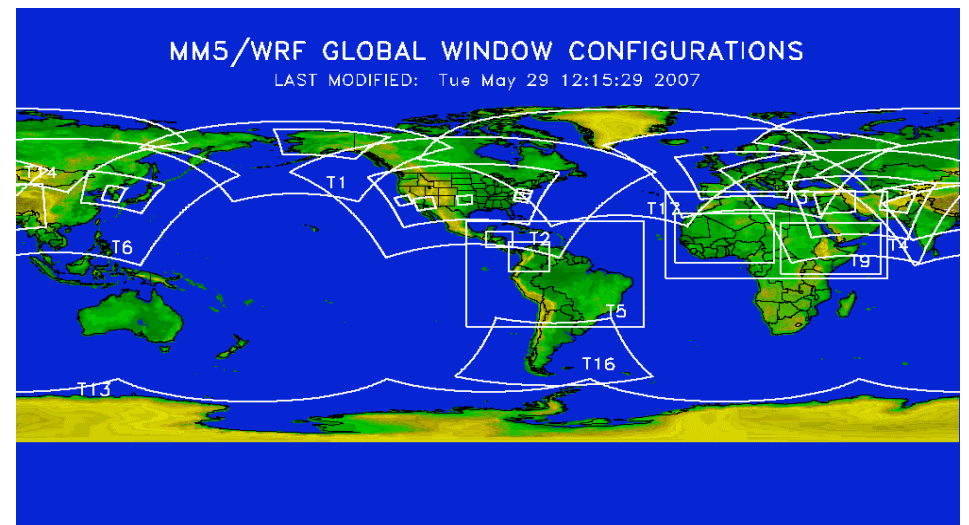
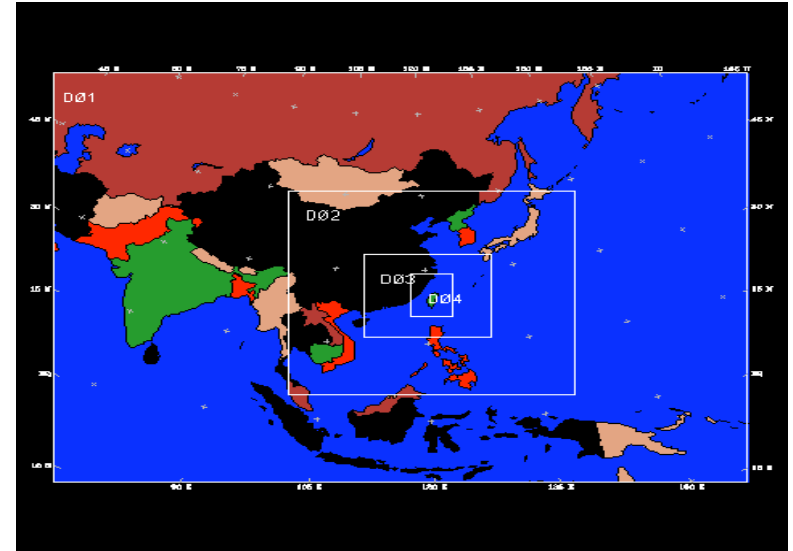


WRFDA in WRF Modeling System



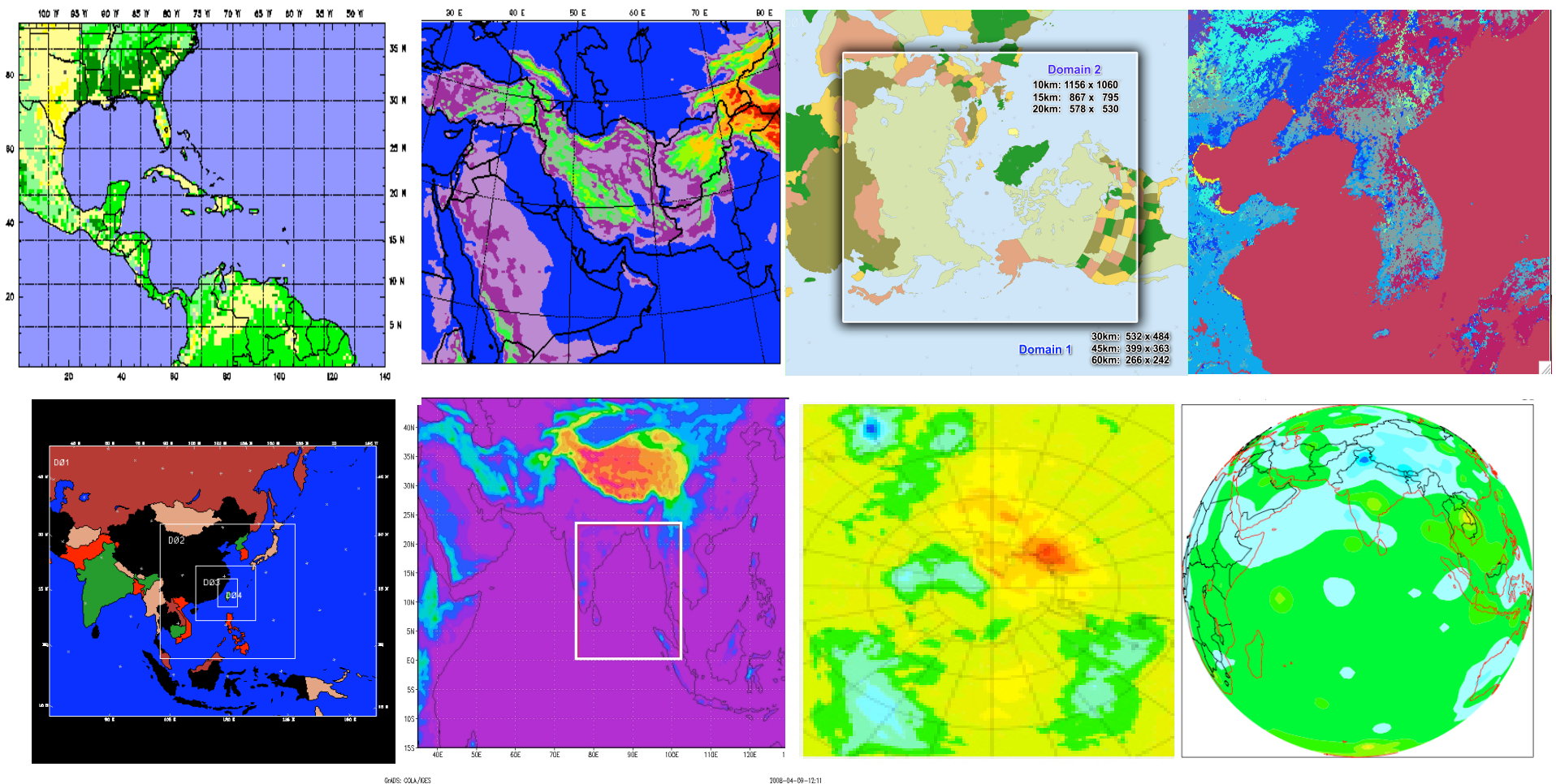
WRFDA

- **Goal:** Community WRF DA system for
 - regional/global,
 - research/operations, and
 - deterministic/probabilistic applications.
- **Techniques:**
 - 3D-Var
 - 4D-Var (regional)
 - Ensemble DA,
 - Hybrid Variational/Ensemble DA.
- **Model:** WRF (ARW, NMM, Global)
- **Support:**
 - NCAR/ESSL/MMM/DAG
 - NCAR/RAL/JNT/DATC
- **Observations:** Conv.+Sat.+Radar
(+bogus)



The WRFDA Program

- NCAR staff (DAG,DATC): 20FTE, ~10 projects.
- Non-NCAR collaborators (AFWA, KMA, CWB, BMB, etc): ~10FTE.
- Community users: ~30 (more in 10,000 general WRF downloads?).



WRFDA Observations

- **In-Situ:**

- Surface (SYNOP, METAR, SHIP, BUOY).
- Upper air (TEMP, PIBAL, AIREP, ACARS, TAMDAR).

- **Remotely sensed retrievals:**

- Atmospheric Motion Vectors (geo/polar).
- SATEM thickness.
- Ground-based GPS Total Precipitable Water/Zenith Total Delay.
- SSM/I oceanic surface wind speed and TPW.
- Scatterometer oceanic surface winds.
- Wind Profiler.
- Radar radial velocities and reflectivities.
- Satellite temperature/humidity/thickness profiles.
- GPS refractivity (e.g. COSMIC).

- **Radiative Transfer (RTTOV or CRTM):**

- HIRS from NOAA-16, NOAA-17, NOAA-18, METOP-2
- AMSU-A from NOAA-15, NOAA-16, NOAA-18, EOS-Aqua, METOP-2
- AMSU-B from NOAA-15, NOAA-16, NOAA-17
- MHS from NOAA-18, METOP-2
- AIRS from EOS-Aqua
- SSMIS from DMSP-16

- **Bogus:**

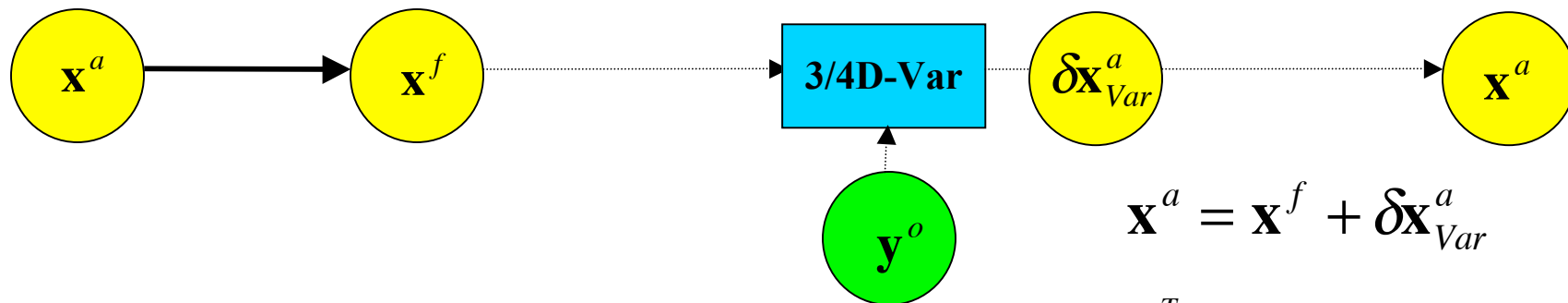
- TC bogus.

- Global bogus.

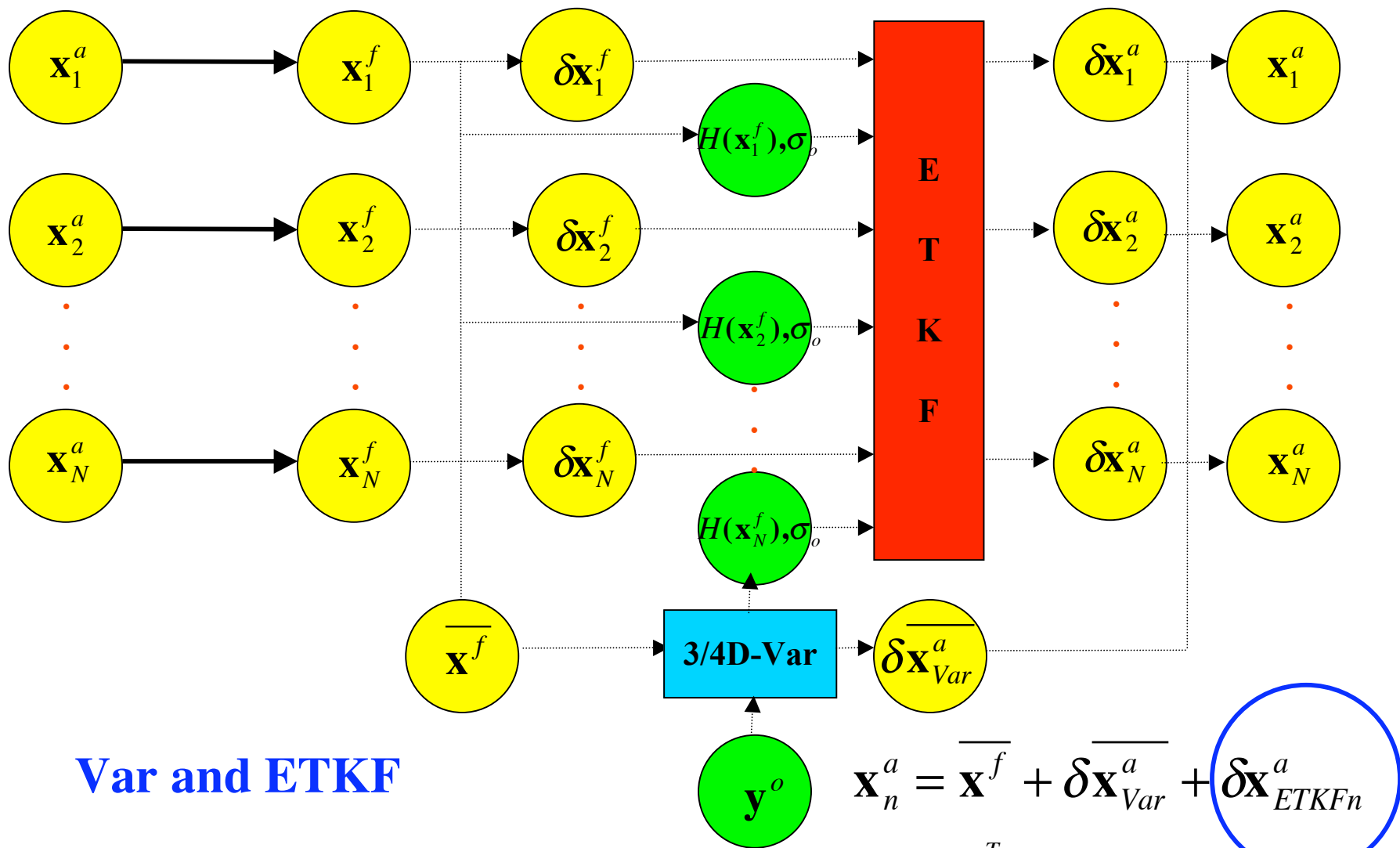


WRFDA

3/4D-Var

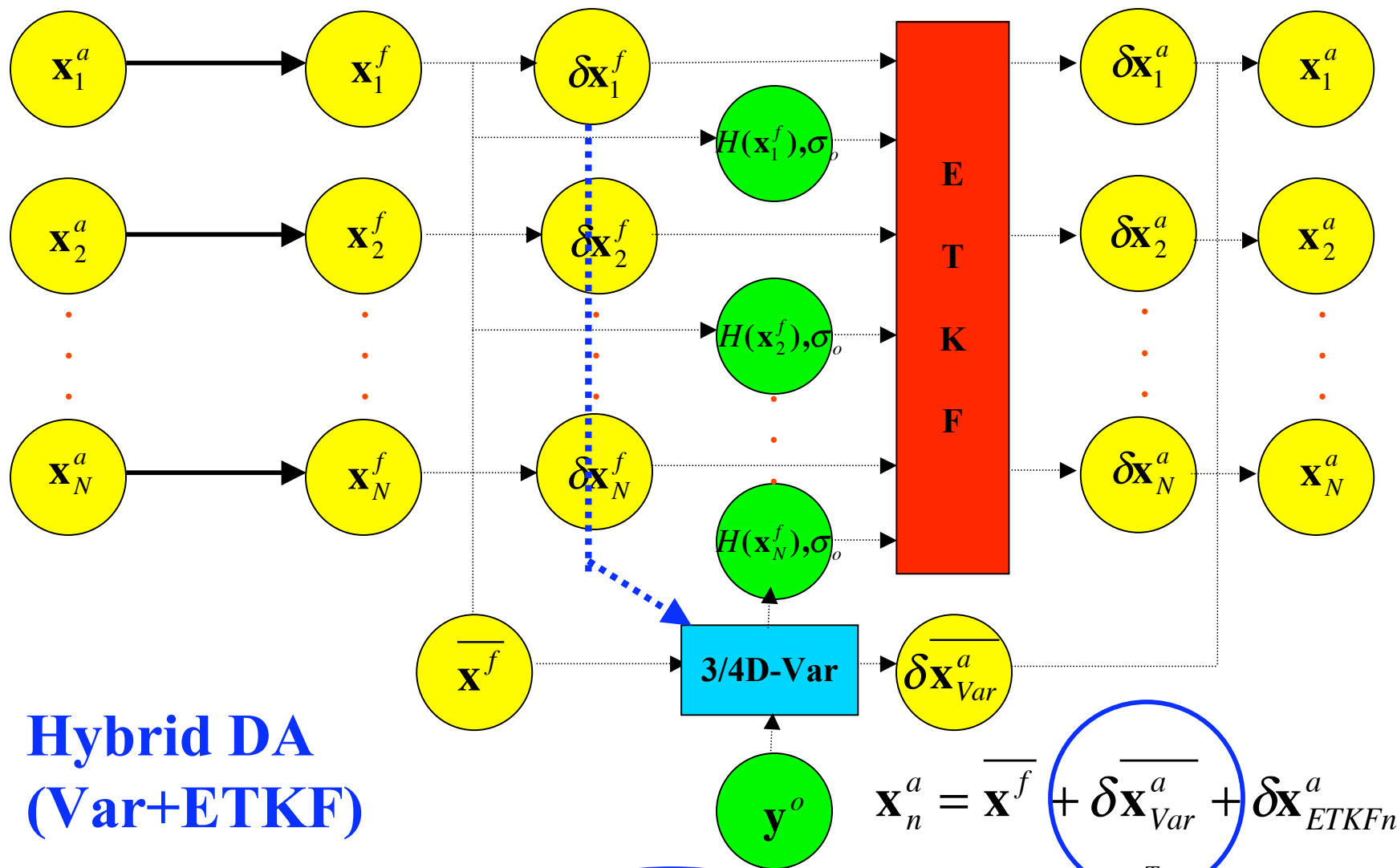


$$J = \frac{1}{2} \delta \mathbf{x}_0^T \mathbf{B}_o^{-1} \delta \mathbf{x}_0 + \frac{1}{2} \sum_{i=0}^n \left[\mathbf{H}_i \delta \mathbf{x}(t_i) - \mathbf{d}_i \right]^T \mathbf{R}_i^{-1} \left[\mathbf{H}_i \delta \mathbf{x}(t_i) - \mathbf{d}_i \right]$$



$$J = \frac{1}{2} \delta \mathbf{x}_0^T \mathbf{B}_o^{-1} \delta \mathbf{x}_0 + \frac{1}{2} \sum_{i=0}^n \left[\mathbf{H}_i \delta \mathbf{x}(t_i) - \mathbf{d}_i \right]^T \mathbf{R}_i^{-1} \left[\mathbf{H}_i \delta \mathbf{x}(t_i) - \mathbf{d}_i \right]$$





$$J = \frac{W_b}{2} \delta \mathbf{x}_0^T \mathbf{B}_o^{-1} \delta \mathbf{x}_0 + \frac{W_\alpha}{2} \mathbf{a}^T \mathbf{A}^{-1} \mathbf{a} + \frac{1}{2} \sum_{i=0}^n \left[\mathbf{H}_i \delta \mathbf{x}(t_i) - \mathbf{d}_i \right]^T \mathbf{R}_i^{-1} \left[\mathbf{H}_i \delta \mathbf{x}(t_i) - \mathbf{d}_i \right]$$

www.mmm.ucar.edu/wrf/users/wrfda

The screenshot shows a web browser window titled "WRFDA Model Users Site". The address bar displays the URL "http://www.mmm.ucar.edu/wrf/users/wrfda/". The browser's "Most Visited" bar shows links to Gmail, Google Calendar, and the WRFDA Model Users Site. The website has a green header with the text "WRFDA USERS PAGE" and a background image of a weather map. Below the header is a navigation menu with links: Home, Analysis System, User Support, Download, Doc / Pub, Links, and Users Forum. A search bar is located on the right side of the page. The main content area is titled "WRF Data Assimilation System Users Page" and contains a welcome message and a description of the WRFDA system. A sidebar on the left contains links to "wrf-model.org", "Public Domain Notice", and "Contact WRF Support". A right sidebar titled "ANNOUNCEMENTS" lists recent updates, including WRF Tutorials, WRF Version 3.1 Release Information, and WRF Version 3.0.1.1 Release information.

WRFDA Model Users Site

http://www.mmm.ucar.edu/wrf/users/wrfda/

Most Visited

Gmail - Inbox - hans.xy.huang... Google Calendar WRFDA Model Users Site

WRFDA USERS PAGE

Home Analysis System User Support Download Doc / Pub Links Users Forum

Search

wrf-model.org

Public Domain Notice

Contact WRF Support

WRF Data Assimilation System Users Page

Welcome to the users home page for the Weather Research and Forecasting (WRF) model data assimilation system (WRFDA). The WRFDA system is in the public domain and is freely available for community use. It is designed to be a flexible, state-of-the-art atmospheric data assimilation system that is portable and efficient on available parallel computing platforms. WRFDA is suitable for use in a broad range of applications across scales ranging from kilometers of regional mesoscale to thousands of kilometers of global scales.

The Mesoscale and Microscale Meteorology Division of NCAR is currently maintaining and supporting a subset of the overall WRF code (Version 3) that includes:

ANNOUNCEMENTS

[WRF Tutorials](#) - January 26 - February 5, 2009, Boulder, Colorado.

[WRF Version 3.1 Release Information](#)

[WRF Version 3.0.1.1 Release:](#)
August 22, 2008

[WRF Var Version 3.0.1.1 Release:](#)
August 29, 2008

New 'Known Problems' posts for V3 [WRF](#) (1/6/09) and [WPS](#) (8/4/08)

The 9th WRF Users' Workshop was held June 23 - 27, 2008 in Boulder, Colorado. [Workshop Presentations](#) is now online.

Done

Issues on data assimilation

(covered by this tutorial)

- Observations $\mathbf{y}^o$
- Observation errors $\mathbf{R}$
- Observation operator H
- Background $\mathbf{x}^b$
- Size of $\mathbf{B}$: statistical model and tuning
- Minimization algorithm (Quasi–Newton; Conjugate Gradient; ...)
- $\mathbf{M}$ and $\mathbf{M}^T$: development and validity
- Size of $\mathbf{P}^f$ and $\mathbf{P}^a$: simplifications
- Model errors $\mathbf{Q}$

