Data assimilation; comparison of 4D-Var and LETKF smoothers

Eugenia Kalnay
and many friends

University of Maryland
CSCAMM DAS13 June 2013

Contents

First part:

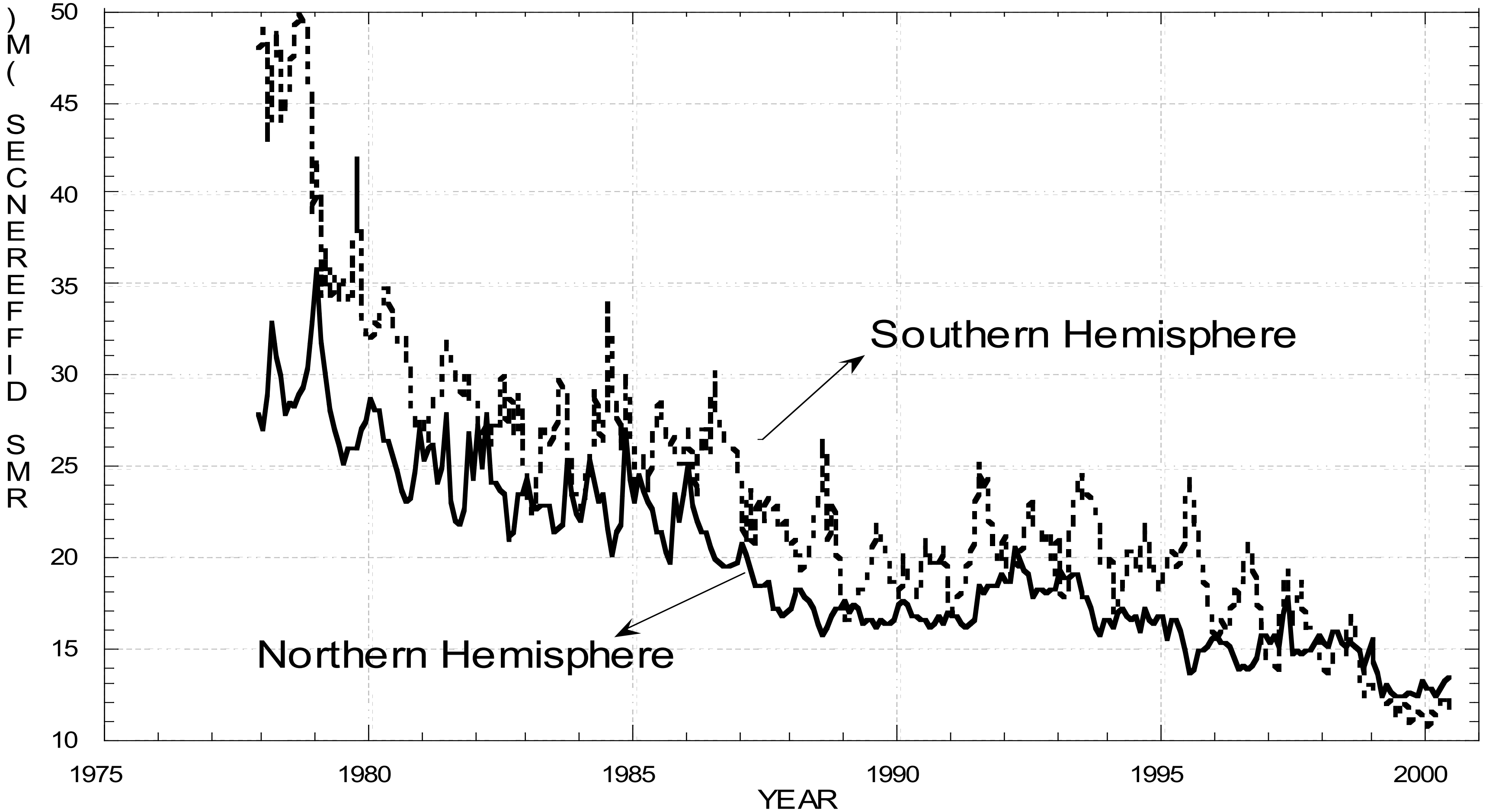
- Forecasting the weather - we are really getting better!
- Why: Better obs? Better models? Better data assimilation?
- Intro to data assim: a toy scalar example 1, we measure with two thermometers, and we want an accurate temperature.
- Another toy example 2, we measure radiance but we want an accurate temperature: we derive OI/KF, 3D-Var, 4D-Var and EnKF for the toy model.
- The equations for the huge real systems are the same as for the toy models.

Second Part: Compare 4D-Var and EnKF in a QG model

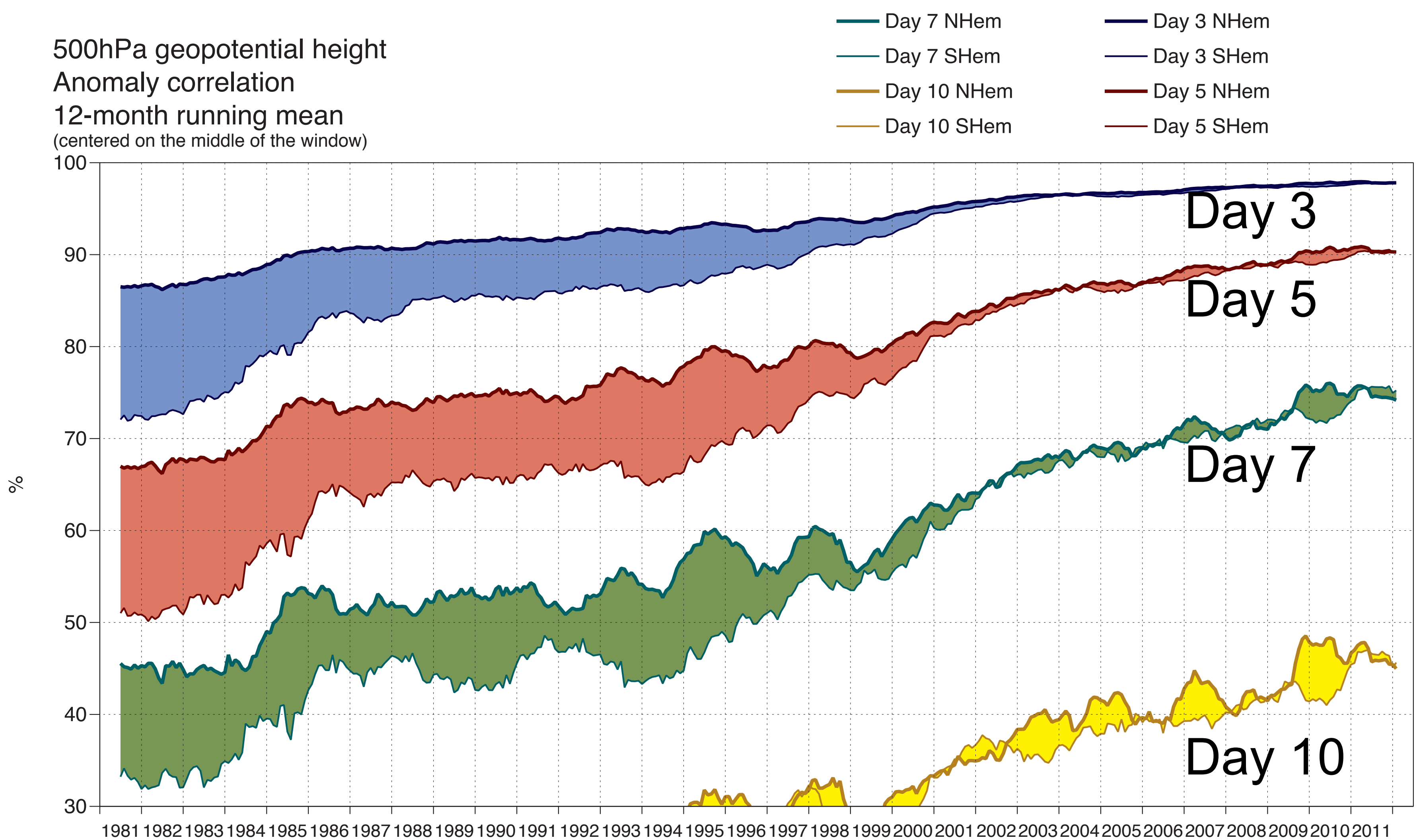
- 4D-Var increments evolve like Singular Vectors
- LETKF increments evolve like Lyapunov Vectors (~Bred Vs)
- Initial 4D-Var increments are norm dependent, not realistic

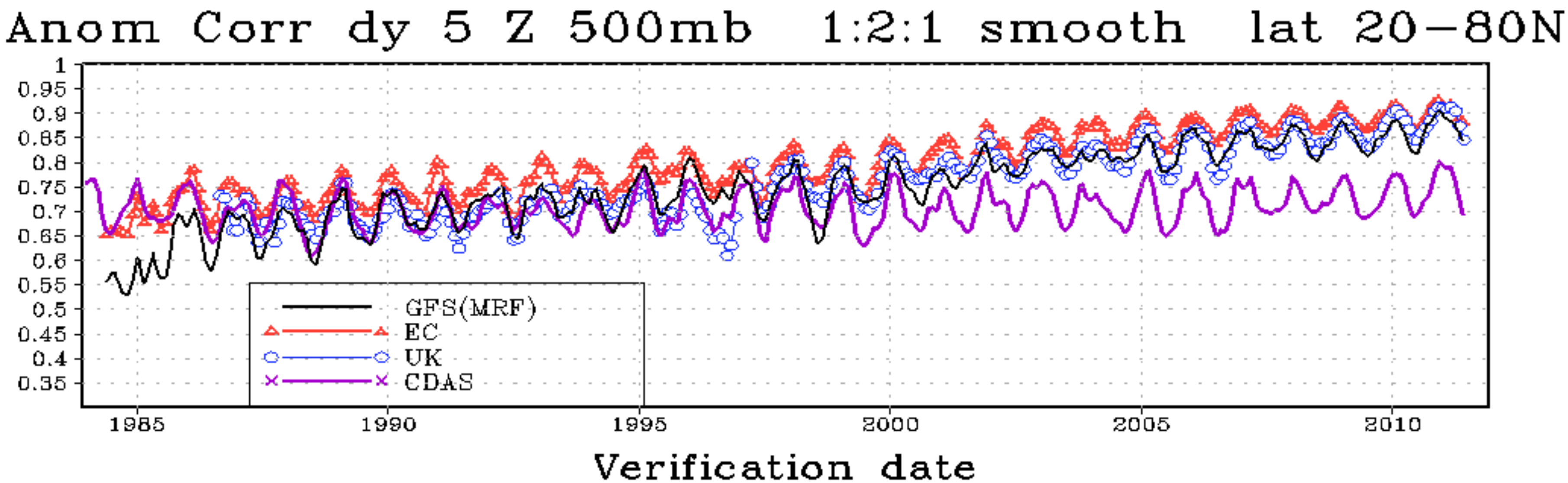
NCEP observational increments

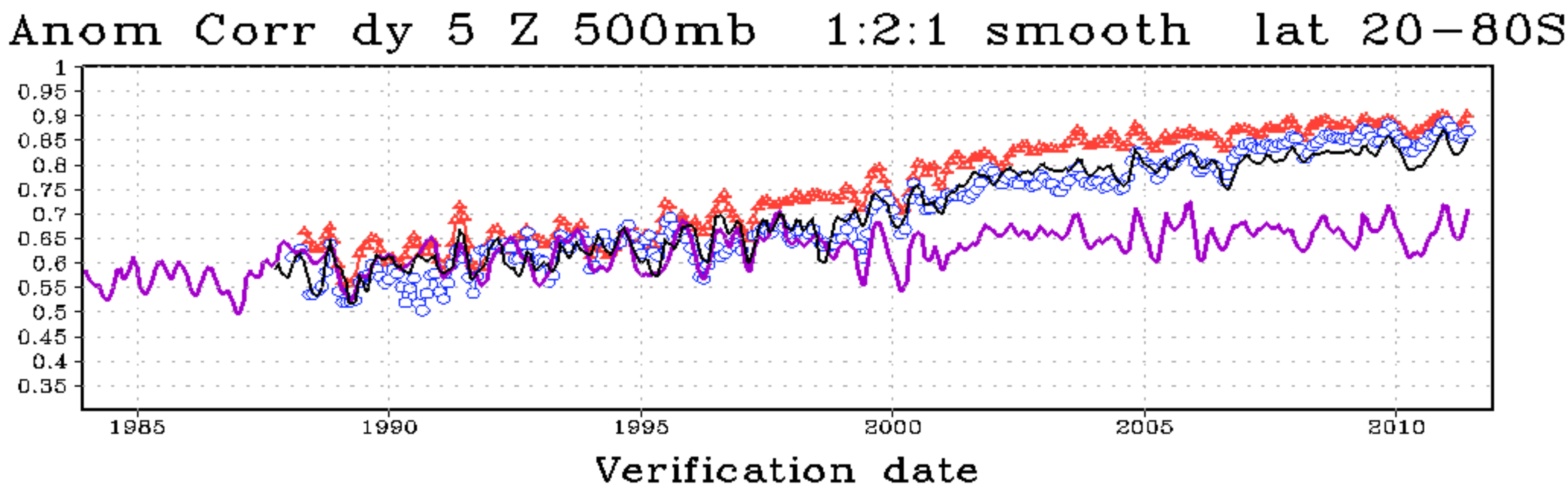
500MB RMS FITS TO RAWINSONDES
6 HR FORECASTS



Comparisons of Northern and Southern Hemispheres

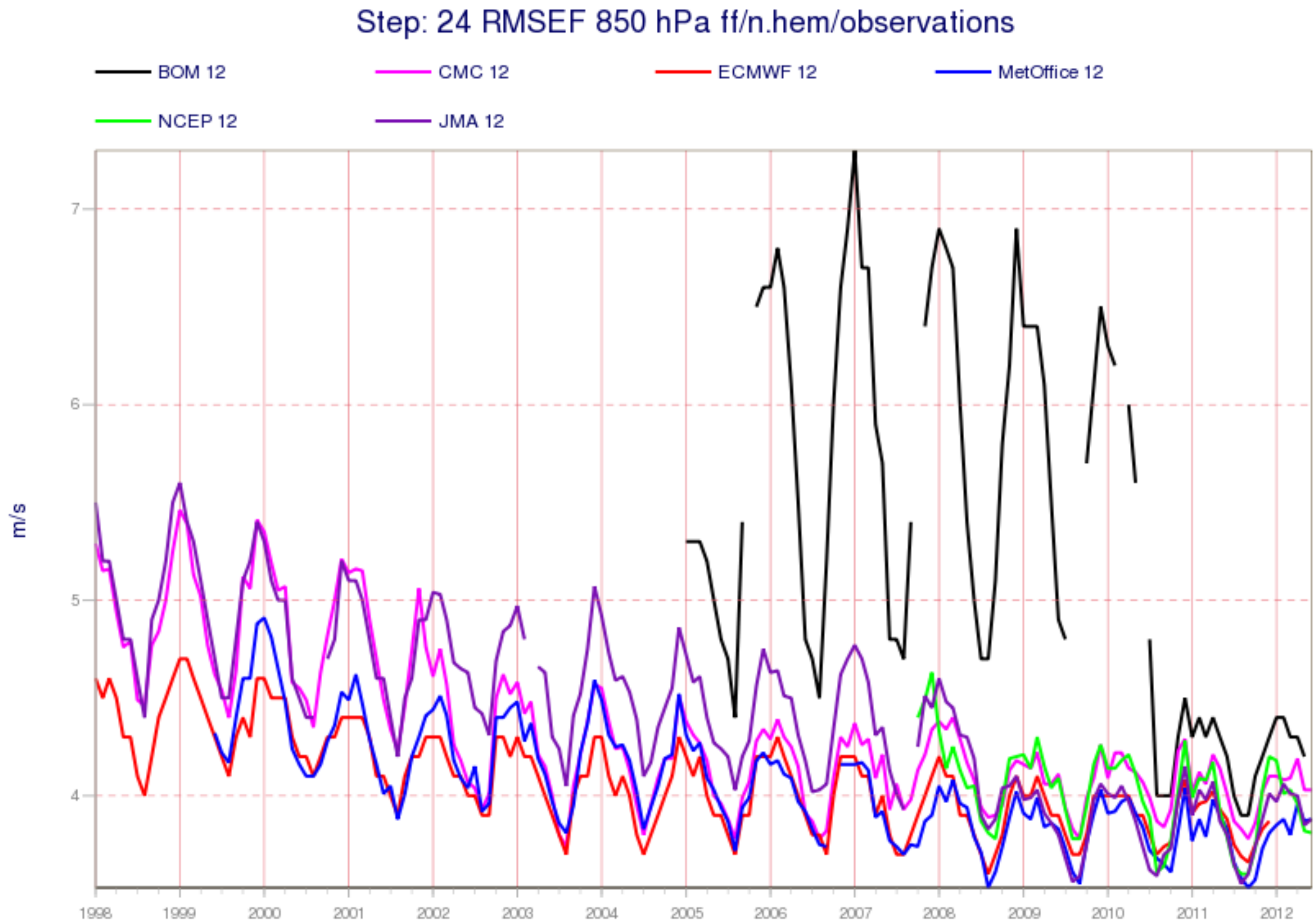



Anom Corr dy 5 Z 500mb 1:2:1 smooth lat 20-80N
GFS(MRF)
EC
UK
CDAS
Verification date

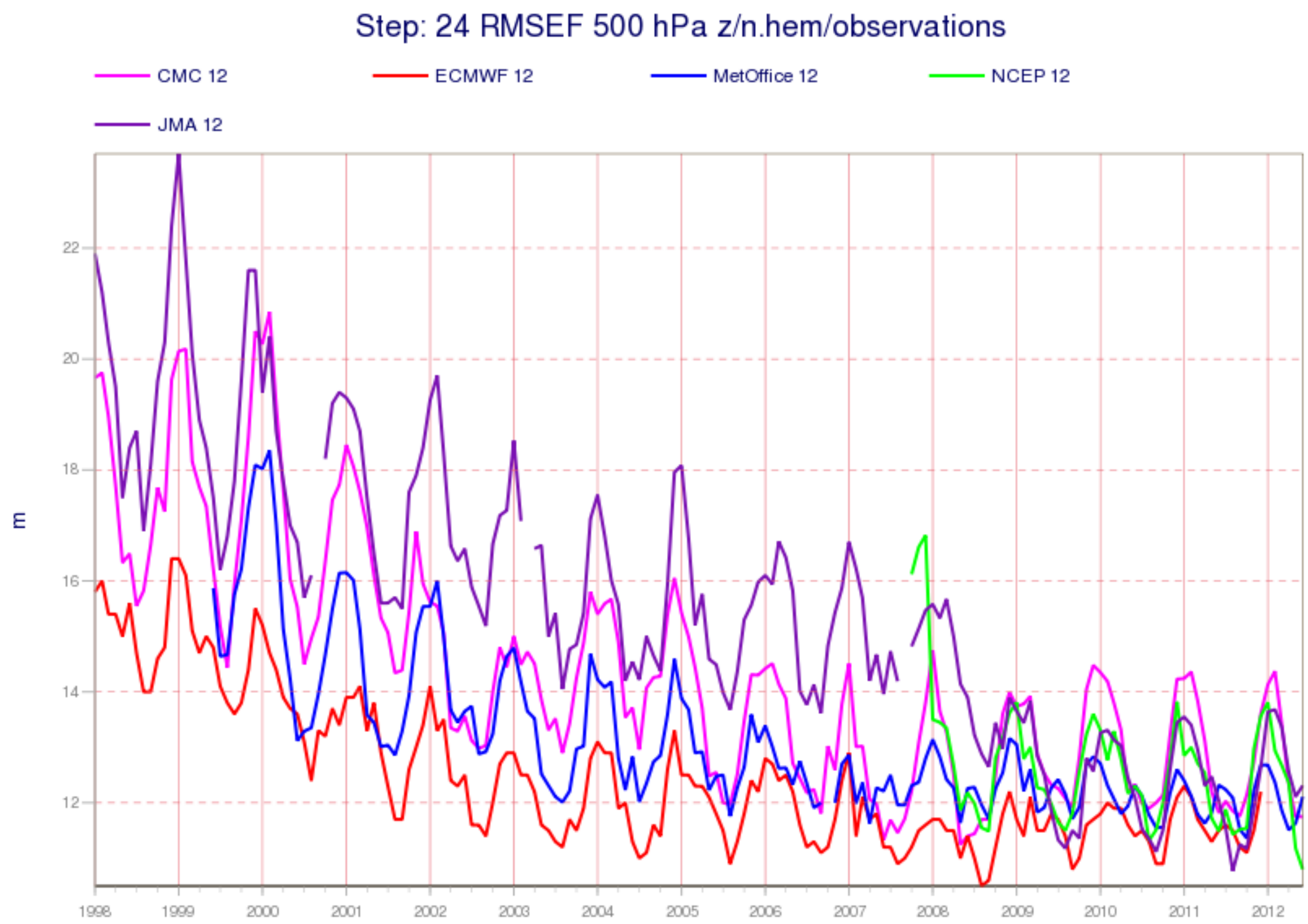

Anom Corr dy 5 Z 500mb 1:2:1 smooth lat 20-80S
Verification date

Comparisons verifying forecasts against observations

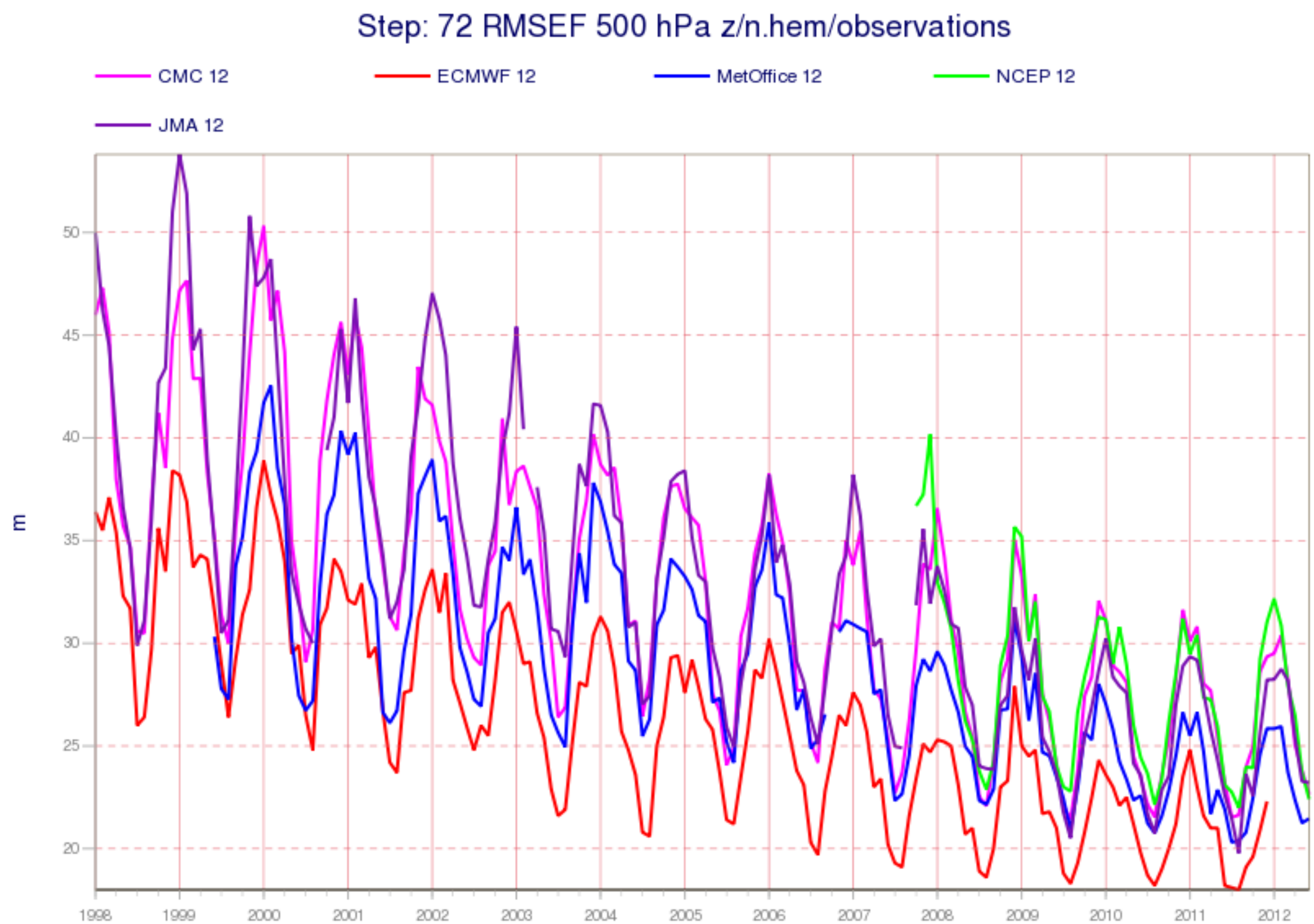
1-day forecasts, 850hPa, NH, verification of wind



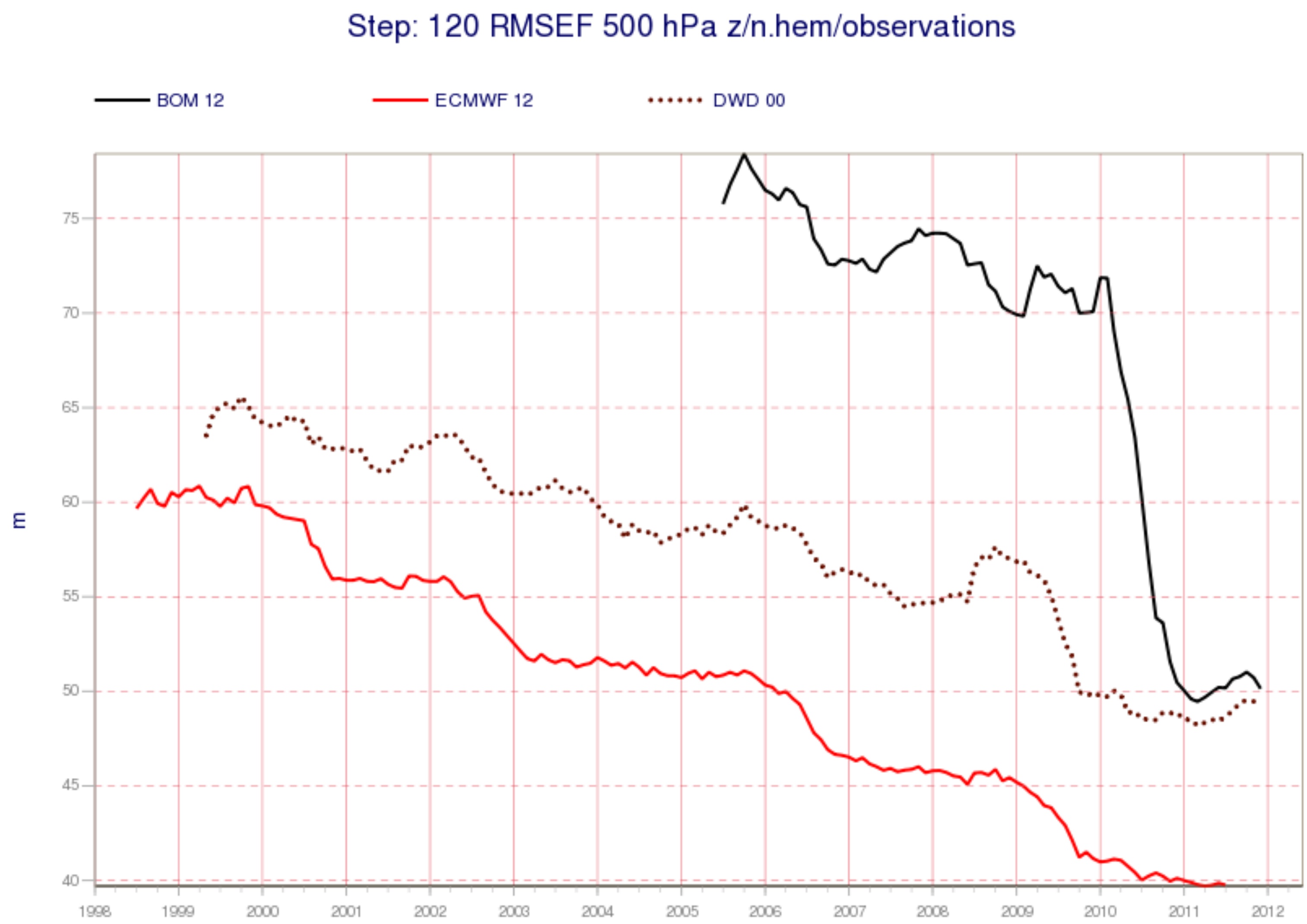
1-day forecast 500hPa Z, NH



3-day forecast, 500hPa, NH against observations



5-day forecast, 500hPa, NH, 12 month average



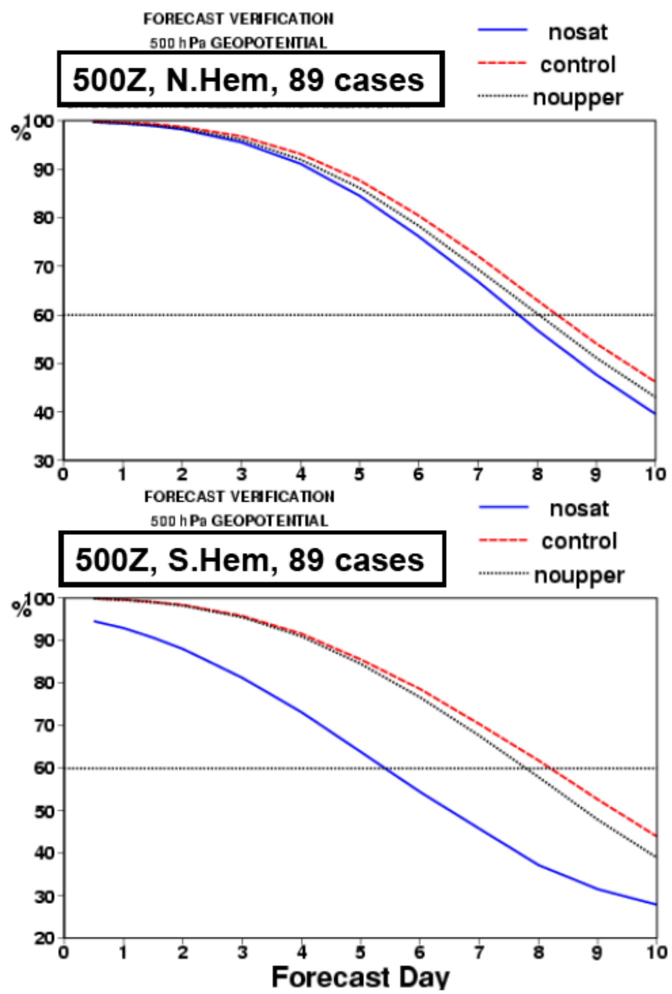
Satellite radiances are essential in the SH

Observing System Experiments (ECMWF - G. Kelly et al.)

NoSAT= no satellite radiances or winds

Control= like operations

NoUpper=no radiosondes, no pilot winds, no wind profilers



Intro. to data assim: toy example 1 summary

A forecast b and an observation o optimally combined (analysis):

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

If the statistics of the errors are exact, and if the coefficients are optimal, then the "precision" of the analysis (defined as the inverse of the variance) is the sum of the precisions of the measurements.

Second toy example of data assimilation including remote sensing.

The importance of these toy examples is that the equations are identical to those obtained with big models and many obs.

Intro. to remote sensing and data assimilation: toy example 2

- Assume we have an object, like a stone in space
- We want to estimate its temperature T (^oK) accurately but we measure the radiance y (W/m^2) that it emits. We have an *obs. model,* e.g.: $y = h(T) \sim \sigma T^4$

Intro. to remote sensing and data assimilation: toy example 2

- Assume we have an object, a stone in space
- We want to estimate its temperature T (^{o}K) accurately but we can only measure the radiance y (W/m^2) that it emits. We have an *obs. model, e.g.*: $y = h(T) \sim \sigma T^4$
- We also have a *forecast model* for the temperature

$$T(t_{i+1}) = m\left[T(t_i)\right];$$

$$\text{e.g., } T(t_{i+1}) = T(t_i) + \Delta t\left[\text{SW heating+LW cooling}\right]$$

Intro. to remote sensing and data assimilation: toy example 2

- Assume we have an object, a stone in space
- We want to estimate its temperature *T* (°K) accurately but we measure the radiance *y* (W/m^2) that it emits. We have an *obs. model, e.g.*: $y = h(T) \sim \sigma T^4$

- We also have a *forecast model* for the temperature

$$T(t_{i+1}) = m\left[T(t_i)\right];$$

$$\text{e.g.,}\ T(t_{i+1}) = T(t_i) + \Delta t\left[\text{SW heating+LW cooling}\right]$$

- We will derive the data assim eqs (KF and Var) for this toy system (easy to understand!)

Intro. to remote sensing and data assimilation: toy example 2

- Assume we have an object, a stone in space
- We want to estimate its temperature *T* (°K) accurately but we measure the radiance *y* (W/m^2) that it emits. We have an *obs. model,* e.g.: $y = h(T) \sim \sigma T^4$

- We also have a *forecast model* for the temperature

$$T(t_{i+1}) = m\left[T(t_i)\right];$$

$$\text{e.g.}, T(t_{i+1}) = T(t_i) + \Delta t\left[\text{SW heating+LW cooling}\right]$$

- We will derive the data assim eqs (OI/KF and Var) for this toy system (easy to understand!)
- Will compare the toy and the real huge vector/matrix equations: they are exactly the same!

Toy temperature data assimilation, measure radiance

We have a forecast T_b (prior) and a radiance obs $y_o = h(T_t) + \varepsilon_0$

The new information (or innovation) is the observational increment:

$$y_o - h(T_b)$$

Toy temperature data assimilation, measure radiance

We have a forecast T_b (prior) and a radiance obs $y_o = h(T_t) + \varepsilon_0$

The new information (or innovation) is the observational increment:

$$y_o - h(T_b)$$

The final formula is very similar to that in toy model 1:

$$T_a = T_b + w(y_o - h(T_b))$$

with the optimal weight $w = \sigma_b^2 H(\sigma_o^2 + H\sigma_b^2 H)^{-1}$

Recall that $T_a = T_b + w(y_o - h(T_b)) = T_b + w(\varepsilon_o - H\varepsilon_b)$

So that, subtracting the truth, $\varepsilon_a = \varepsilon_b + w(\varepsilon_o - H\varepsilon_b)$

Toy temperature data assimilation, measure radiance

Summary for Optimal Interpolation/Kalman Filter (sequential):

$$T_a = T_b + w(y_o - h(T_b))$$ analysis

with $$w = \sigma_b^2 H(\sigma_o^2 + \sigma_b^2 H^2)^{-1}$$ optimal weight

The analysis error is obtained from squaring $\varepsilon_a = \varepsilon_b + w\left[\varepsilon_o - H\varepsilon_b\right]$

$$\sigma_a^2 = \overline{\varepsilon_a^2} = (1 - wH)\sigma_b^2 = \frac{\sigma_o^2}{\sigma_o^2 + \sigma_b^2 H^2}\sigma_b^2$$

It can also be written as

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

analysis precision=
forecast precision + observation precision

Toy temperature data assimilation, variational approach

We have a forecast T_b and a radiance obs $y_o = h(T_t) + \varepsilon_0$

Innovation: $y_o - h(T_b)$

From a 3D-Var point of view, we want to find a T_a that minimizes the cost function J:

$$J(T_a) = \frac{(T_a - T_b)^2}{2\sigma_b^2} + \frac{(h(T_a) - y_o)^2}{2\sigma_o^2}$$

Toy temperature data assimilation, variational approach

We have a forecast T_b and a radiance obs $y_o = h(T_t) + \varepsilon_0$

Innovation: $y_o - h(T_b)$

From a 3D-Var point of view, we want to find a T_a that minimizes the cost function J:

$$J(T_a) = \frac{(T_a - T_b)^2}{2\sigma_b^2} + \frac{(h(T_a) - y_o)^2}{2\sigma_o^2}$$

This analysis temperature T_a is closest to both the forecast T_b and the observation y_o and maximizes the likelihood of $T_a \sim T_{truth}$ given the information we have.

It is easier to find the analysis increment $T_a - T_b$ that minimizes the cost function J

Toy temperature data assimilation, variational approach

We have a forecast T_b and a radiance obs $y_o = h(T_t) + \varepsilon_0$

Innovation: $y_o - h(T_b)$

From a 3D-Var point of view, we want to find a T_a that minimizes the cost function J:

$$J(T_a) = \frac{(T_a - T_b)^2}{2\sigma_b^2} + \frac{(h(T_a) - y_o)^2}{2\sigma_o^2}$$

The cost function is derived from a maximum likelihood analysis:

Toy temperature data assimilation, variational approach

We have a forecast T_b and a radiance obs $y_o = h(T_t) + \varepsilon_0$

Innovation: $y_o - h(T_b)$

From a 3D-Var point of view, we want to find a T_a that minimizes the cost function *J:*

$$J(T_a) = \frac{(T_a - T_b)^2}{2\sigma_b^2} + \frac{(h(T_a) - y_o)^2}{2\sigma_o^2}$$

Likelihood of T_{truth} given T_b: $\frac{1}{\sqrt{2\pi}\sigma_b} \exp\left[-\frac{(T_{truth} - T_b)^2}{2\sigma_b^2} \right]$

Toy temperature data assimilation, variational approach

We have a forecast T_b and a radiance obs $y_o = h(T_t) + \varepsilon_0$

Innovation: $y_o - h(T_b)$

From a 3D-Var point of view, we want to find a T_a that minimizes the cost function J:

$$J(T_a) = \frac{(T_a - T_b)^2}{2\sigma_b^2} + \frac{(h(T_a) - y_o)^2}{2\sigma_o^2}$$

Likelihood of T_{truth} given T_b: $\frac{1}{\sqrt{2\pi}\sigma_b} \exp\left[\frac{\left(T_{truth} - T_b\right)^2}{2\sigma_b^2} \right]$

Likelihood of $h(T_{truth})$ given y_o: $\frac{1}{\sqrt{2\pi}\sigma_o} \exp\left[-\frac{\left(h(T_{truth}) - y_o\right)^2}{2\sigma_o^2} \right]$

Toy temperature data assimilation, variational approach

We have a forecast T_b and a radiance obs $y_o = h(T_t) + \varepsilon_0$

Innovation: $y_o - h(T_b)$

From a 3D-Var point of view, we want to find a T_a that minimizes the cost function J:

$$2J_{\min} = \frac{(T_a - T_b)^2}{\sigma_b^2} + \frac{(h(T_a) - y_o)^2}{\sigma_o^2}$$

Likelihood of T_{truth} given T_b: $\frac{1}{\sqrt{2\pi}\sigma_b}\exp\left[-\frac{\left(T_{truth} - T_b\right)^2}{2\sigma_b^2}\right]$

Likelihood of $h(T_{truth})$ given y_o: $\frac{1}{\sqrt{2\pi}\sigma_o}\exp\left[-\frac{\left(h(T_{truth}) - y_o\right)^2}{2\sigma_o^2}\right]$

Joint likelihood of T_{truth}:

$$\frac{1}{\sqrt{2\pi}\sigma_b}\exp\left[-\frac{\left(T_{truth} - T_b\right)^2}{2\sigma_b^2} - \frac{\left(h(T_{truth}) - y_o\right)^2}{2\sigma_o^2}\right]$$

Minimizing the cost function maximizes the likelihood of the estimate of truth

Toy temperature data assimilation, variational approach

Again, we have a forecast T_b and a radiance obs $y_o = h(T_t) + \varepsilon_0$

Innovation: $y_o - h(T_b)$

We want to find $(T_a - T_b)$ that minimizes the cost function J. This maximizes the likelihood of $T_a \sim T_{truth}$ given both T_b and y_o

$$2J_{\min} = \frac{(T_a - T_b)^2}{\sigma_b^2} + \frac{(h(T_a) - y_o)^2}{\sigma_o^2}$$

Toy temperature data assimilation, variational approach

We have a forecast T_b and a radiance obs $y_o = h(T_t) + \varepsilon_0$

Innovation: $y_o - h(T_b)$

We want to find $(T_a - T_b)$ that minimizes the cost function J. This maximizes the likelihood of $T_a \sim T_{truth}$ given both T_b and y_o

$$J(T_a) = \frac{(T_a - T_b)^2}{2\sigma_b^2} + \frac{(h(T_a) - y_o)^2}{2\sigma_o^2}$$

To find the minimum we use an incremental approach: find $T_a - T_b$:

$$h(T_a) - y_o = h(T_b) - y_o + H(T_a - T_b)$$

So that from $\partial J / \partial (T_a - T_b) = 0$ we get

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

Toy temperature data assimilation, variational approach

We have a forecast T_b and a radiance obs $y_o = h(T_t) + \varepsilon_0$

Innovation: $y_o - h(T_b)$

From a 3D-Var point of view, we want to find $(T_a - T_b)$ that minimizes the cost function J.

$$J(T_a) = \frac{(T_a - T_b)^2}{2\sigma_b^2} + \frac{(h(T_a) - y_o)^2}{2\sigma_o^2}$$

To find the minimum we use an incremental approach: find $T_a - T_b$:

$$h(T_a) - y_o = h(T_b) - y_o + H(T_a - T_b)$$

So that from $\partial J / \partial(T_a - T_b) = 0$ we get

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

or $$T_a = T_b + w\left(y_o - h(T_b)\right)$$ where now

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

Toy temperature data assimilation, variational approach

We have a forecast T_b and a radiance obs $y_o = h(T_t) + \varepsilon_0$

Innovation: $y_o - h(T_b)$

3D-Var: T_a minimizes the distance to both the background and the observations

$$2J_{\min} = \frac{(T_a - T_b)^2}{\sigma_b^2} + \frac{(h(T_a) - y_o)^2}{\sigma_o^2}$$

3D-Var solution $T_a = T_b + w(y_o - h(T_b))$ with $w = \left(\sigma_b^{-2} + H\sigma_o^{-2}H\right)^{-1} H\sigma_o^{-2} = \sigma_a^2 H\sigma_o^{-2}$

Toy temperature data assimilation, variational approach

We have a forecast T_b and a radiance obs $y_o = h(T_t) + \varepsilon_0$

Innovation: $y_o - h(T_b)$

3D-Var: T_a minimizes the distance to both the background and the observations

$$2J_{\min} = \frac{(T_a - T_b)^2}{\sigma_b^2} + \frac{(h(T_a) - y_o)^2}{\sigma_o^2}$$

3D-Var solution $T_a = T_b + w(y_o - h(T_b))$ with $w_{3D-Var} = \left(\sigma_b^{-2} + H\sigma_o^{-2}H\right)^{-1} H\sigma_o^{-2} = \sigma_a^2 H\sigma_o^{-2}$

This variational solution is the **same** as the one obtained before with Kalman filter (a sequential approach, like Optimal Interpolation, Lorenc 86)):

KF/OI solution $T_a = T_b + w(y_o - h(T_b))$ with $w_{OI} = \sigma_b^2 H(\sigma_o^2 + \sigma_b^2 H^2)^{-1}$

Toy temperature data assimilation, variational approach

We have a forecast T_b and a radiance obs $y_o = h(T_t) + \varepsilon_0$

Innovation: $y_o - h(T_b)$

3D-Var: T_a minimizes the distance to both the background and the observations

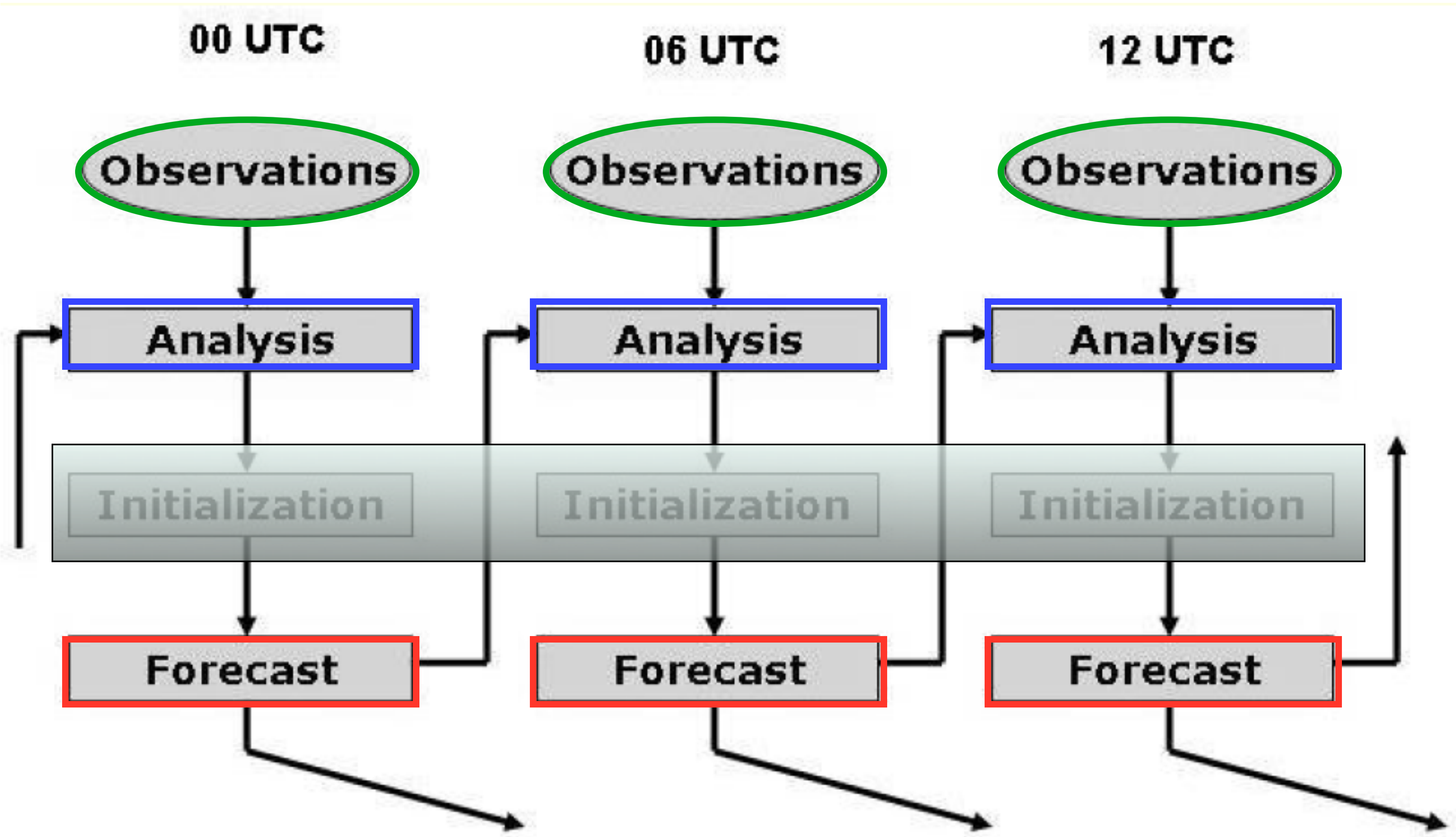
$$2J_{\min} = \frac{(T_a - T_b)^2}{\sigma_b^2} + \frac{(h(T_a) - y_o)^2}{\sigma_o^2}$$

3D-Var solution $T_a = T_b + w(y_o - h(T_b))$ with $w_{3D-Var} = \left(\sigma_b^{-2} + H\sigma_o^{-2}H\right)^{-1} H\sigma_o^{-2} = \sigma_a^2 H\sigma_o^{-2}$

This variational solution is the **<u>same</u>** as the one obtained before with Kalman filter (a sequential approach, like Optimal Interpolation, Lorenc 86)):

KF/OI solution $T_a = T_b + w(y_o - h(T_b))$ with $w_{OI} = \sigma_b^2 H(\sigma_o^2 + \sigma_b^2 H^2)^{-1}$

Show that the 3d-Var and the OI/KF weights are the same: both methods find the same optimal solution!



Typical 6-hour analysis cycle.

Forecast phase, followed by Analysis phase

Toy temperature analysis cycle (Kalman Filter)

Forecasting phase, from t_i to t_{i+1}: $$T_b(t_{i+1}) = m\left[T_a(t_i)\right]$$

Forecast error: $$\varepsilon_b(t_{i+1}) = T_b(t_{i+1}) - T_t(t_{i+1}) =$$
$$m\left[T_a(t_i)\right] - m\left[T_t(t_i)\right] + \varepsilon_m(t_{i+1}) = M\varepsilon_a(t_i) + \varepsilon_m(t_{i+1})$$

So that we can predict the forecast error variance

$$\sigma_b^2(t_{i+1}) = M^2\sigma_a^2(t_i) + Q_i; \quad Q_i = \overline{\varepsilon_m^2(t_{i+1})}$$

(The forecast error variance comes from the analysis and model errors)

Toy temperature analysis cycle (Kalman Filter)

Forecasting phase, from t_i to t_{i+1}: $$T_b(t_{i+1}) = m\left[T_a(t_i)\right]$$

Forecast error: $$\varepsilon_b(t_{i+1}) = T_b(t_{i+1}) - T_t(t_{i+1}) =$$
$$m\left[T_a(t_i)\right] - m\left[T_t(t_i)\right] + \varepsilon_m(t_{i+1}) = M\varepsilon_a(t_i) + \varepsilon_m(t_{i+1})$$

So that we can predict the forecast error variance

$$\sigma_b^2(t_{i+1}) = M^2\sigma_a^2(t_i) + Q_i; \quad Q_i = \overline{\varepsilon_m^2(t_{i+1})}$$

(The forecast error variance comes from the analysis and model errors)

Now we can compute the optimal weight (KF or Var, whichever form is more convenient, since they are equivalent):

$$w = \sigma_b^2 H(\sigma_o^2 + H\sigma_b^2 H)^{-1} = \left(\sigma_b^{-2} + H\sigma_o^{-2}H\right)^{-1} H\sigma_o^{-2}$$

Toy temperature analysis cycle (Kalman Filter)

Analysis phase: we use the new observation $y_o(t_{i+1})$

compute the new observational increment $y_o(t_{i+1}) - h\left(T_b(t_{i+1})\right)$

and the new analysis:

$$T_a(t_{i+1}) = T_b(t_{i+1}) + w_{i+1}\left[y_o(t_{i+1}) - h\left(T_b(t_{i+1})\right)\right]$$

We also need the compute the new analysis error variance:

from $\sigma_a^{-2} = \sigma_b^{-2} + H\sigma_o^{-2}H$

we get $\sigma_a^2(t_{i+1}) = \left(\dfrac{\sigma_o^2\sigma_b^2}{\sigma_o^2 + H^2\sigma_b^2}\right)_{i+1} = (1 - w_{i+1}H)\sigma_{b\,i+1}^2 < \sigma_{b\,i+1}^2$

now we can advance to the next cycle $t_{i+2}, t_{i+3}, \ldots$

Summary of toy Analysis Cycle (for a scalar)

$$T_b(t_{i+1}) = m\left[T_a(t_i)\right] \qquad \sigma_b^2(t_{i+1}) = M^2\left[\sigma_a^2(t_i)\right] \qquad M = \partial m / \partial T$$

Interpretation…

"We use the model to forecast T_b and to update the forecast error variance from t_i to t_{i+1} "

Summary of toy Analysis Cycle (for a scalar)

$$T_b(t_{i+1}) = m\left[T_a(t_i)\right] \qquad \sigma_b^2(t_{i+1}) = M^2\left[\sigma_a^2(t_i)\right] \qquad M = \partial m / \partial T$$

"We use the model to forecast T_b and to update the forecast error variance from t_i to t_{i+1} "

At t_{i+1} $\quad T_a = T_b + w\left[y_o - h\left(T_b\right)\right]$

"The analysis is obtained by adding to the background the innovation (difference between the observation and the first guess) multiplied by the optimal weight:

Summary of toy Analysis Cycle (for a scalar)

$$T_b(t_{i+1}) = m\left[T_a(t_i)\right] \qquad \sigma_b^2(t_{i+1}) = M^2\left[\sigma_a^2(t_i)\right] \qquad M = \partial m / \partial T$$

"We use the model to forecast T_b and to update the forecast error variance from t_i to t_{i+1} "

At t_{i+1} $\quad T_a = T_b + w\left[y_o - h\left(T_b\right)\right]$

"The analysis is obtained by adding to the background the innovation (difference between the observation and the first guess) multiplied by the optimal weight:

$$w = \sigma_b^2 H(\sigma_o^2 + H\sigma_b^2 H)^{-1}$$

"The optimal weight is the background error variance divided by the sum of the observation and the background error variance. $H = \partial h / \partial T$ ensures that the magnitudes and units are correct."

Summary of toy Analysis Cycle (for a scalar)

$$w = \sigma_b^2 H (\sigma_o^2 + H \sigma_b^2 H)^{-1}$$

"The optimal weight is the background error variance divided by the sum of the observation and the background error variance. $H = \partial h / \partial T$ ensures that the magnitudes and units are correct."

Note that the larger the background error variance, the larger the correction to the first guess.

Summary of toy Analysis Cycle (for a scalar)

The analysis error variance is given by

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

"The analysis error variance is reduced from the background error by a factor (1 - scaled optimal weight)"

Summary of toy system equations (cont.)

The analysis error variance is given by

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

"The analysis error variance is reduced from the background error by a factor (1 - scaled optimal weight)"

This can also be written as

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

"The analysis precision is given by the sum of the background and observation precisions"

Equations for toy and real huge systems

These statements are important because they hold true for data assimilation systems in very large multidimensional problems (e.g., NWP).

Instead of model, analysis and observational scalars, we have 3-dimensional vectors of sizes of the order of 10^7-10^9

Equations for toy and real huge systems

These statements are important because they hold true for data assimilation systems in very large multidimensional problems (e.g., NWP).

Instead of model, analysis and observational scalars, we have 3-dimensional vectors of sizes of the order of 10^7-10^8

We have to replace scalars (obs, fcasts, analyses) by vectors

$$T_b \rightarrow \mathbf{x}_b; \quad T_a \rightarrow \mathbf{x}_a; \quad y_o \rightarrow \mathbf{y}_o;$$

and their error variances by error covariances:

$$\sigma_b^2 \rightarrow \mathbf{B}; \quad \sigma_a^2 \rightarrow \mathbf{A}; \quad \sigma_o^2 \rightarrow \mathbf{R};$$

Equations for toy and real huge systems

These statements are important because they hold true for data assimilation systems in very large multidimensional problems (e.g., NWP).

Instead of model, analysis and observational scalars, we have 3-dimensional vectors of sizes of the order of 10^7-10^8

We have to replace scalars (obs, forecasts) by vectors

$$T_b \rightarrow \mathbf{x}_b; \quad T_a \rightarrow \mathbf{x}_a; \quad y_o \rightarrow \mathbf{y}_o;$$

and their error variances by error covariances:

$$\sigma_b^2 \rightarrow \mathbf{B}; \quad \sigma_a^2 \rightarrow \mathbf{A}; \quad \sigma_o^2 \rightarrow \mathbf{R};$$

Interpretation of the NWP system of equations

"We use the model to forecast from t_i to t_{i+1} "

$$\mathbf{x}_b(t_{i+1}) = M\left[\mathbf{x}_a(t_i)\right]$$

At t_{i+1} $\quad \mathbf{x}_a = \mathbf{x}_b + \mathbf{K}\left[\mathbf{y}_o - H\left(\mathbf{x}_b\right)\right]$

Interpretation of the NWP system of equations

"We use the model to forecast from t_i to t_{i+1} "

$$\mathbf{x}_b(t_{i+1}) = M\left[\mathbf{x}_a(t_i)\right]$$

At t_{i+1} $\quad \mathbf{x}_a = \mathbf{x}_b + \mathbf{K}\left[\mathbf{y}_o - H\left(\mathbf{x}_b\right)\right]$

"The analysis is obtained by adding to the background the innovation (difference between the observation and the first guess) multiplied by the optimal Kalman gain (weight) matrix"

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

Interpretation of the NWP system of equations

"We use the model to forecast from t_i to t_{i+1} "

$$\mathbf{x}_b(t_{i+1}) = M\left[\mathbf{x}_a(t_i)\right]$$

At t_{i+1} $\quad \mathbf{x}_a = \mathbf{x}_b + \mathbf{K}\left[\mathbf{y}_o - H\left(\mathbf{x}_b\right)\right]$

"The analysis is obtained by adding to the background the innovation (difference between the observation and the first guess) multiplied by the optimal Kalman gain (weight) matrix"

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

"The optimal weight is the background error covariance divided by the sum of the observation and the background error covariance. $\mathbf{H} = \partial H / \partial \mathbf{x}$ ensures that the magnitudes and units are correct. The larger the background error covariance, the larger the correction to the first guess."

Interpretation of the NWP system of equations

Forecast phase:

"We use the model to forecast from t_i to t_{i+1} "

$$\mathbf{x}_b(t_{i+1}) = M\left[\mathbf{x}_a(t_i)\right]$$

Interpretation of the NWP system of equations

Forecast phase:

"We use the model to forecast from t_i to t_{i+1} "

$$\mathbf{x}_b(t_{i+1}) = M\left[\mathbf{x}_a(t_i)\right]$$

"We use the linear tangent model and its adjoint to forecast **B**"

$$\mathbf{B}(t_{i+1}) = \mathbf{M}\left[\mathbf{A}(t_i)\right]\mathbf{M}^T$$

Interpretation of the NWP system of equations

Forecast phase:

"We use the model to forecast from t_i to t_{i+1} "

$$\mathbf{x}_b(t_{i+1}) = M\left[\mathbf{x}_a(t_i)\right]$$

"We use the linear tangent model and its adjoint to forecast **B**"

$$\mathbf{B}(t_{i+1}) = \mathbf{M}\left[\mathbf{A}(t_i)\right]\mathbf{M}^T$$

"However, this step is so <u>horrendously</u> expensive that it makes Kalman Filter <u>completely unfeasible</u>".

"Ensemble Kalman Filter solves this problem by estimating **B** using an ensemble of forecasts."

Summary of NWP equations (cont.)

The analysis error covariance is given by

$$\mathbf{A} = \left(\mathbf{I} - \mathbf{KH}\right)\mathbf{B}$$

"The analysis covariance is reduced from the background covariance by a factor (**I** - scaled optimal gain)"

Summary of NWP equations (cont.)

The analysis error covariance is given by

$$\mathbf{A} = \left(\mathbf{I} - \mathbf{KH}\right)\mathbf{B}$$

"The analysis covariance is reduced from the background covariance by a factor (**I** - scaled optimal gain)"

This can also be written as

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^{T}\mathbf{R}^{-1}\mathbf{H}$$

"The analysis precision is given by the sum of the background and observation precisions"

Summary of NWP equations (cont.)

The analysis error covariance is given by

$$\mathbf{A} = \left(\mathbf{I} - \mathbf{KH}\right)\mathbf{B}$$

"The analysis covariance is reduced from the background covariance by a factor (**I** - scaled optimal gain)"

This can also be written as

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^{T}\mathbf{R}^{-1}\mathbf{H}$$

"The analysis precision is given by the sum of the background and observation precisions"

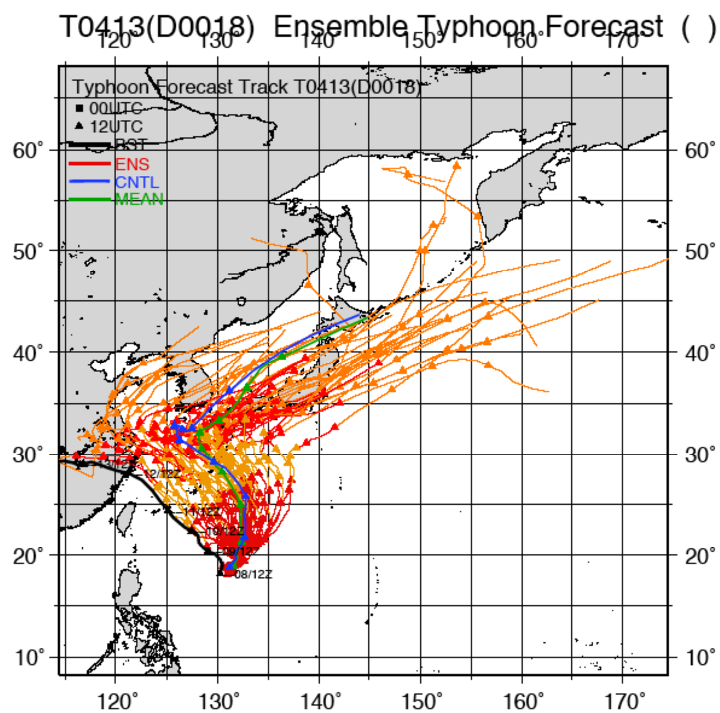
$$\mathbf{K} = \mathbf{B}\mathbf{H}^{T}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{T})^{-1} = (\mathbf{B}^{-1} + \mathbf{H}^{T}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{T}\mathbf{R}^{-1}$$

"The variational approach and the sequential approach are solving the same problem, with the same **K**, but only KF (or EnKF) provide an estimate of the analysis error covariance"

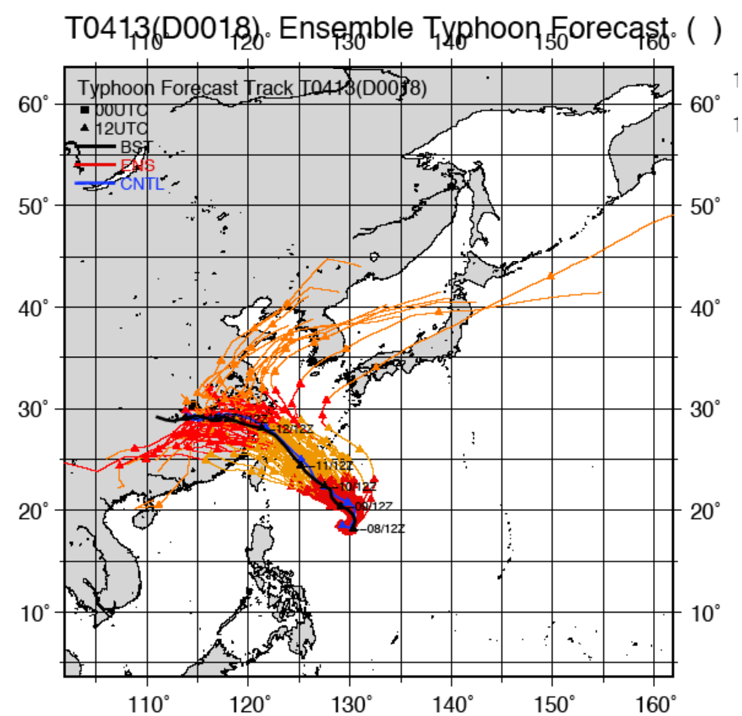
Comparison of 4-D Var and LETKF at JMA

18th typhoon in 2004, IC 12Z 8 August 2004

T. Miyoshi and Y. Sato



operational



LETKF

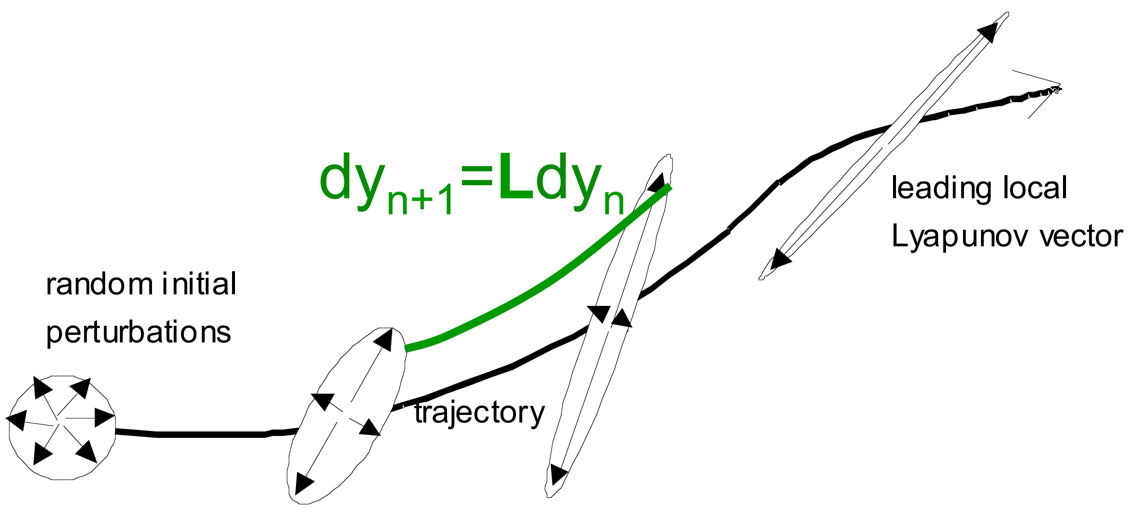
2^{nd} part: Comparison of 4D-Var/SV and LETKF/BVs

Lorenz (1965) introduced (without using their current names) all the concepts of: Tangent linear model, Adjoint model, Singular vectors, and Lyapunov vectors for a low order atmospheric model, and their consequences for ensemble forecasting.

He also introduced the concept of "errors of the day": predictability is not constant: It depends on the stability of the evolving atmospheric flow (the basic trajectory or reference state).

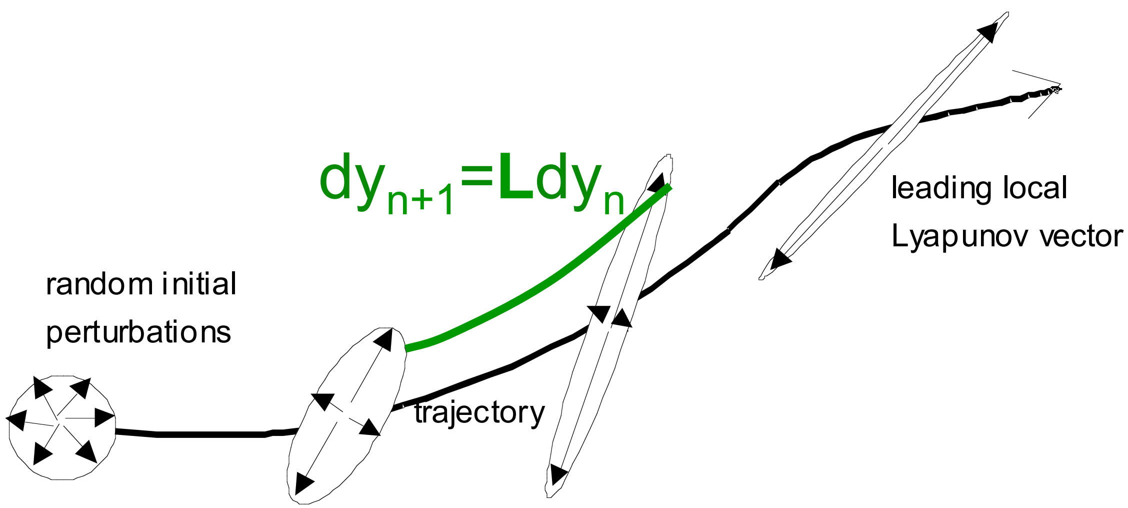
When there is an instability, all perturbations converge towards the fastest growing perturbation (leading Lyapunov Vector). The LLV is computed applying the linear tangent model **L** on each perturbation of the nonlinear trajectory

Fig. 6.7: Schematic of how all perturbations will converge towards the leading Local Lyapunov Vector



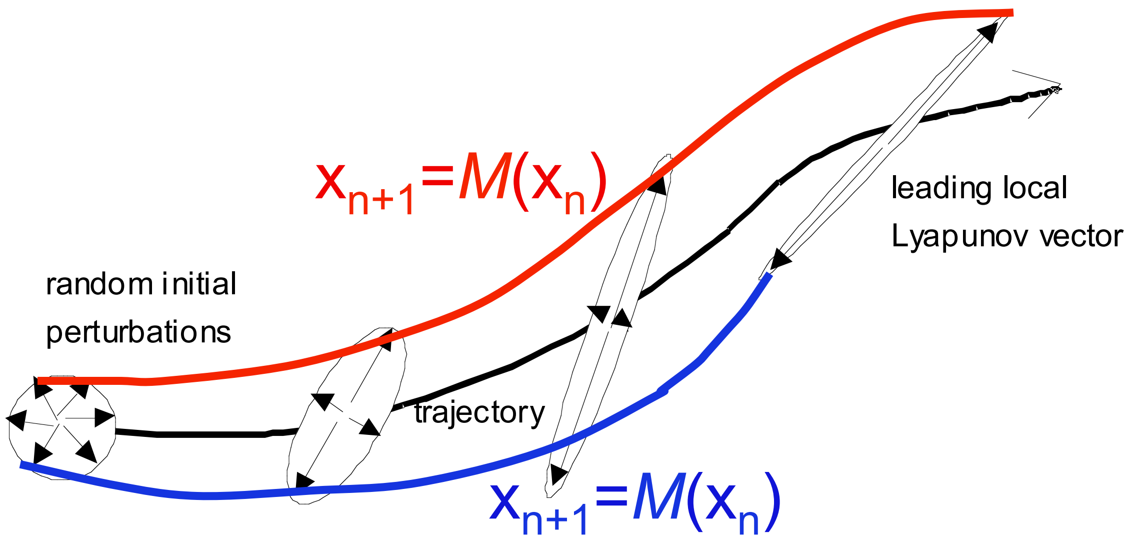
When there is an instability, all perturbations converge towards the fastest growing perturbation (leading Lyapunov Vector). The LLV is computed applying the linear tangent model **L** on each perturbation of the nonlinear trajectory

Fig. 6.7: Schematic of how all perturbations will converge towards the leading Local Lyapunov Vector



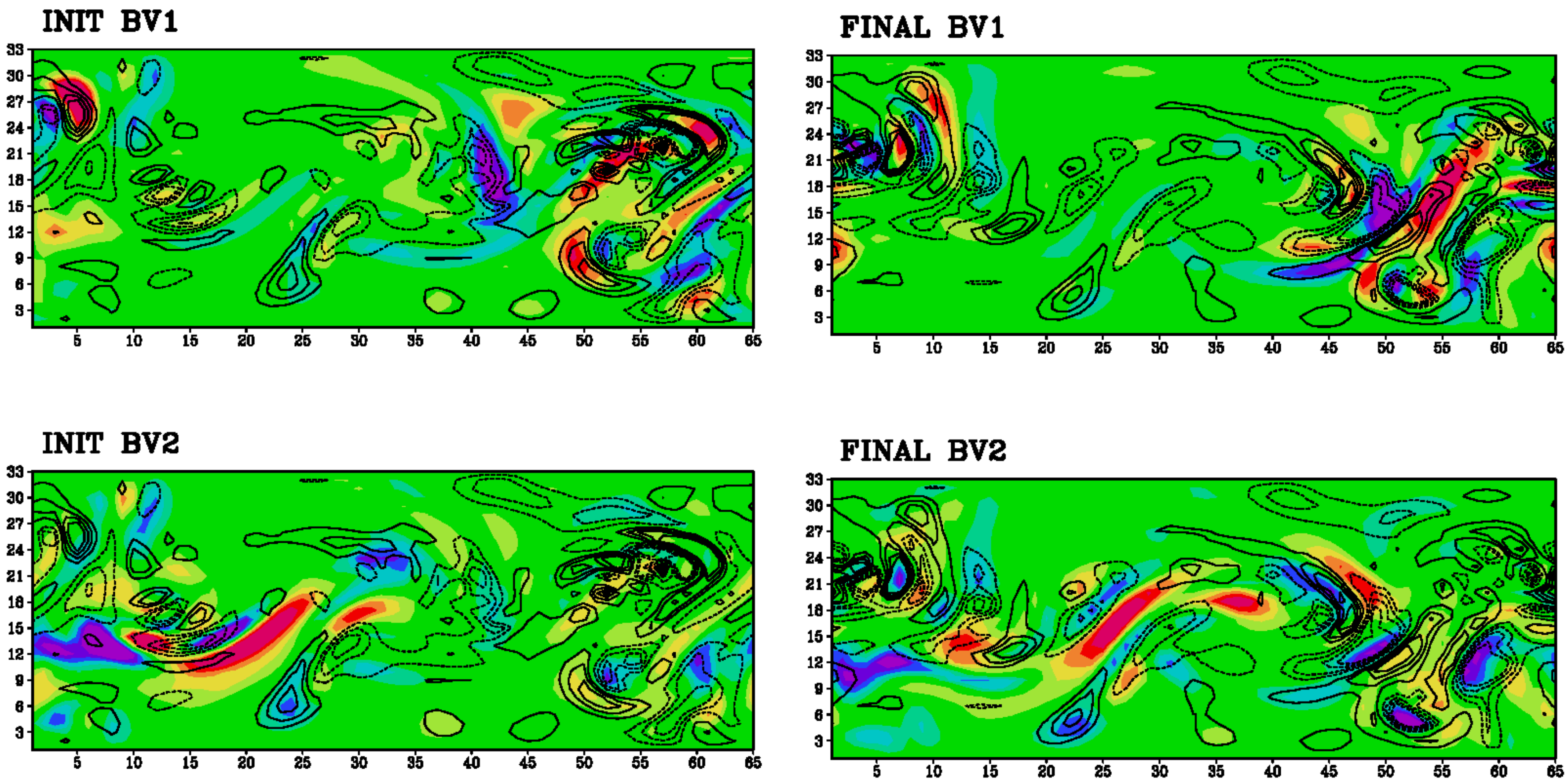
Bred Vectors: nonlinear generalizations of Lyapunov vectors, finite amplitude, finite time

Fig. 6.7: Schematic of how all perturbations will converge towards the leading Local Lyapunov Vector



Two initial and final BV (24hr)

contours: 3D-Var forecast errors, colors: BVs



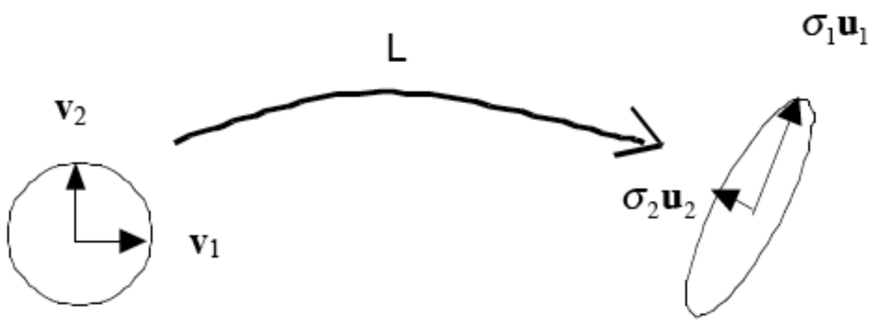
The BV (colors) have shapes similar to the forecast errors (contours)

SV: Apply the linear tangent model forward in time to a ball of size 1

$$\mathbf{L}\mathbf{v}_i = \sigma_i \mathbf{u}_i$$

$\mathbf{v}_i$ are the initial singular vectors
$\mathbf{u}_i$ are the final singular vectors
σ_i are the singular values

Fig. 6.3: Schematic of the application of the TLM to a sphere of perturbations of size 1 for a given interval (t_0,t_1).



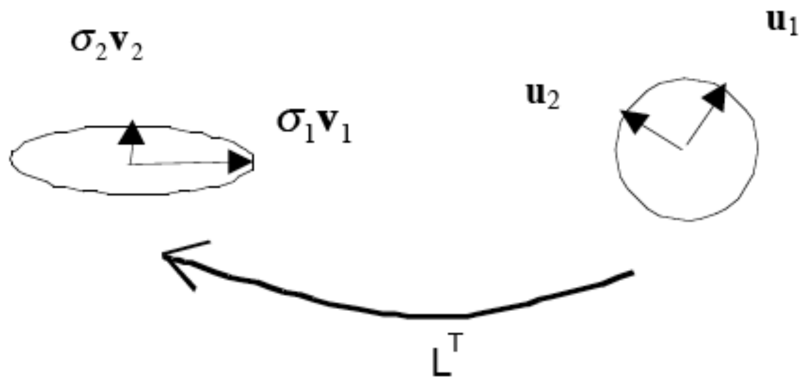
- The ball becomes an ellipsoid, with each final SV $\boldsymbol{u}_i$ multiplied by the corresponding singular value σ_i .
- Both the initial and final SVs are orthogonal.

If we apply the adjoint model backwards in time

$$\mathbf{L}^T\mathbf{u}_i = \sigma_i \mathbf{v}_i$$

$\mathbf{v}_i$ are the initial singular vectors
$\mathbf{u}_i$ are the final singular vectors
σ_i are the singular values

Fig. 6.4: Schematic of the application of the adjoint of the TLM to a sphere of perturbations of size 1 at the final time.



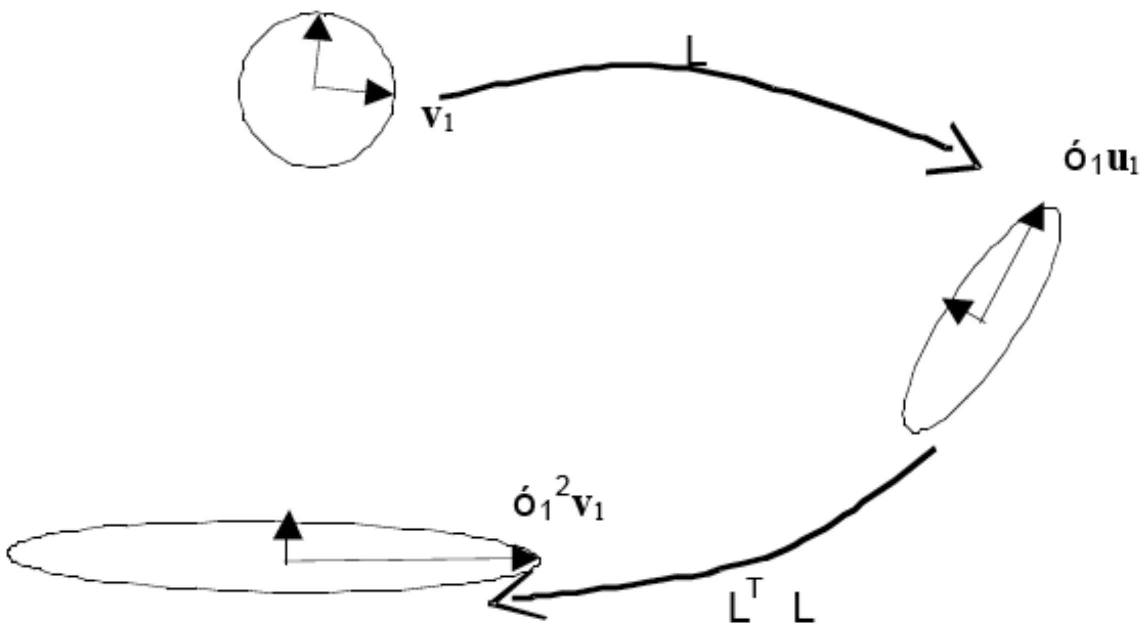
- The final SVs get transformed into initial SVs, and are also multiplied by the corresponding singular value σ_i .

Apply both the linear and the adjoint models

$$\mathbf{L}^T\mathbf{L}\mathbf{v}_i = \sigma_i\mathbf{L}^T\mathbf{u}_i = \sigma_i^2\mathbf{v}_i$$

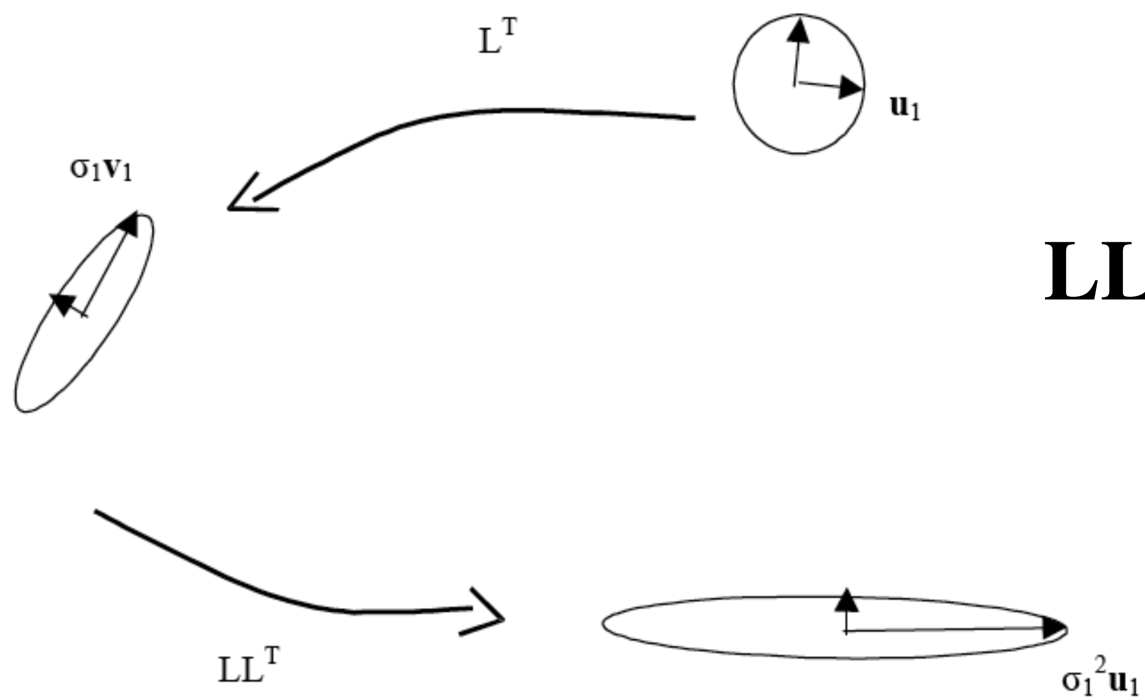
So that $\mathbf{v}_i$ are the eigenvectors of $\mathbf{L}^T\mathbf{L}$ and σ_i^2 are its eigenvalues (singular values)

Fig. 6.5: Schematic of the application of the TLM forward in time followed by the adjoint of the TLM to a sphere of perturbations of size 1 at the initial time.



Conversely, apply the adjoint model first and then the TLM

Fig. 6.6: Schematic of the application of the adjoint of the TLM backward in time followed by the TLM forward to a sphere of perturbations of size 1 at the final time.



$$\mathbf{L}\mathbf{L}^T\mathbf{u}_i = \sigma_i^2\mathbf{u}_i$$

More generally,

A perturbation is advanced from t_n to t_{n+1} $\quad \mathbf{y}_{n+1} = \mathbf{L}\mathbf{y}_n$

Find the final size with a final norm **P**:

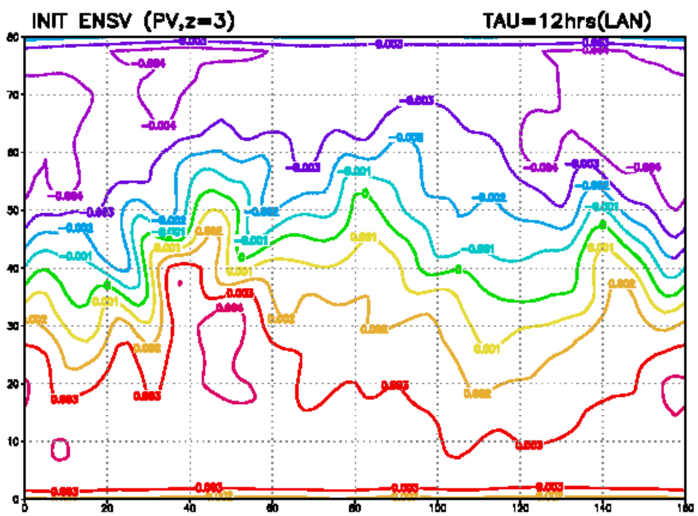
$$\left\|\mathbf{y}_{n+1}\right\|^2 = (\mathbf{P}\mathbf{y}_{n+1})^T(\mathbf{P}\mathbf{y}_{n+1}) = \mathbf{y}_n^T\mathbf{L}^T\mathbf{P}^T\mathbf{P}\mathbf{L}\mathbf{y}_n$$

This is subject to the constraint that all the initial perturbations being of size 1 (with some norm **W** that measures the *initial* size): $\quad \mathbf{y}_n^T\mathbf{W}^T\mathbf{W}\mathbf{y}_n = 1$

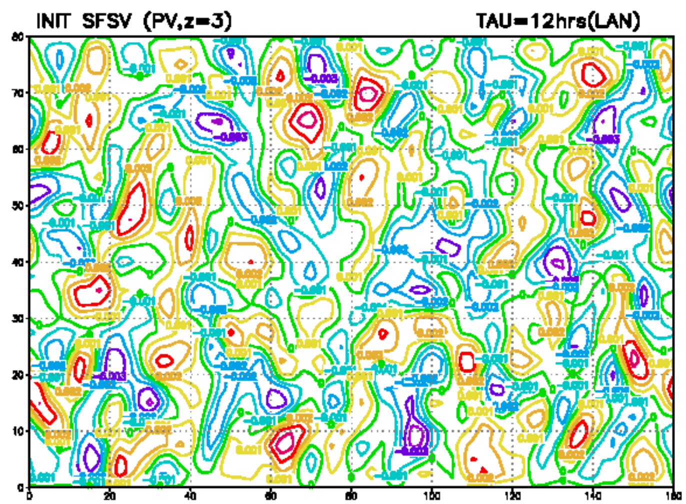
The **initial** leading SVs depend strongly on the initial norm **W** and on the optimization period $T = t_{n+1} - t_n$

QG model: Singular vectors using either enstrophy/streamfunction initial norms (12hr)

Initial SV with enstrophy norm

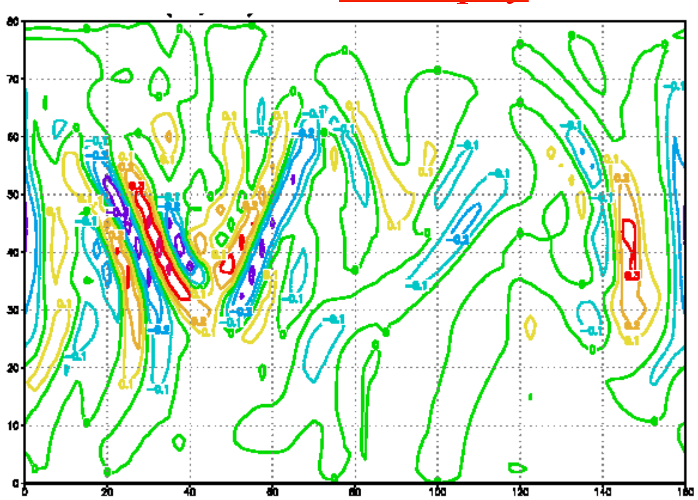


Initial SV with streamfunction norm

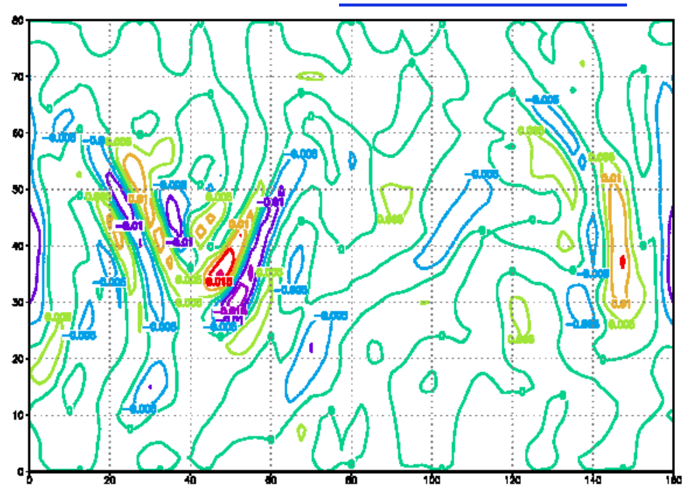


Initial SVs are very sensitive to the norm

Final SV with enstrophy norm



Final SV with streamfunction norm

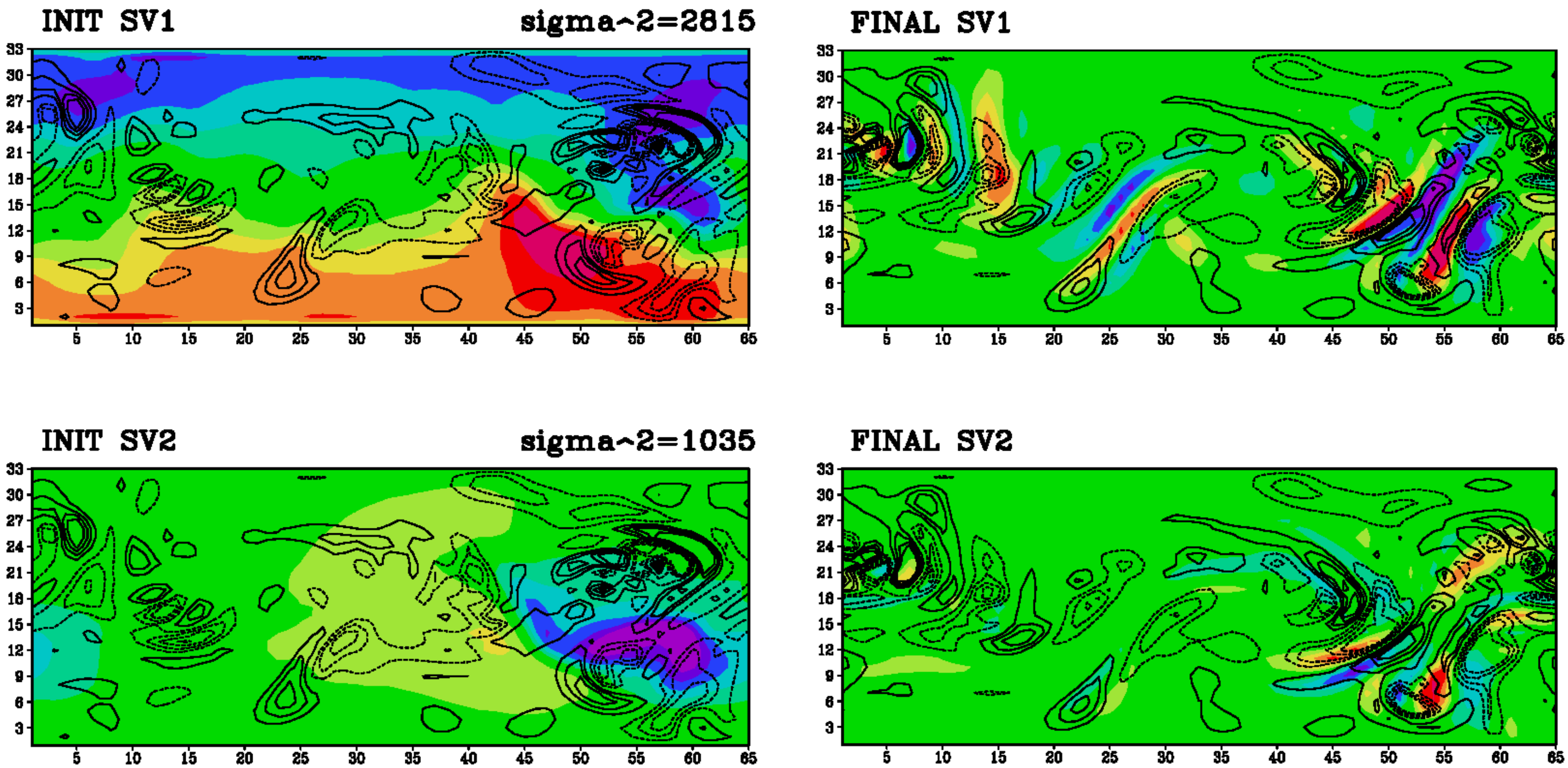


Final SVs look like bred vectors (or Lyapunov vectors)

(Shu-Chih Yang)

Two initial and final SV (24hr, vorticity^2 norm)

contours: 3D-Var forecast errors, colors: SVs



With an enstrophy norm, the initial SVs have large scales, by the end of the"optimization" interval, the final SVs look like BVs (and LVs)

How to compute nonlinear, tangent linear and adjoint codes:

Lorenz (1963) third equation: $\boxed{\dot{x}_3 = x_1 x_2 - b x_3}$

Nonlinear model, forward in time
$x_3(t + \Delta t) = x_3(t) + [x_1(t) x_2(t) - b x_3(t)]\Delta t$

M

The nonlinear model is used directly by BVs and by EnKF

Example of nonlinear, tangent linear and adjoint codes:

Lorenz (1963) third equation: $\boxed{\dot{x}_3 = x_1 x_2 - b x_3}$

Nonlinear model, forward in time

$$x_3(t+\Delta t) = x_3(t) + [x_1(t)x_2(t) - bx_3(t)]\Delta t$$

M

Tangent linear model, forward in time

$$\delta x_3(t+\Delta t) = \delta x_3(t) + [x_2(t)\delta x_1(t) + x_1(t)\delta x_2(t) - b\delta x_3(t)]\Delta t$$

L

The TLM is needed to construct the adjoint model L^T. Linearizing processes like convection is very hard!

Example of nonlinear, tangent linear and adjoint codes:

Lorenz (1963) third equation: $\boxed{\dot{x}_3 = x_1 x_2 - b x_3}$

Nonlinear model, forward in time — M

$$x_3(t+\Delta t) = x_3(t) + [x_1(t)x_2(t) - bx_3(t)]\Delta t$$

Tangent linear model, forward in time — L

$$\delta x_3(t+\Delta t) = \delta x_3(t) + [x_2(t)\delta x_1(t) + x_1(t)\delta x_2(t) - b\delta x_3(t)]\Delta t$$

In the adjoint model the above line becomes — L^T

$$\delta x_3^*(t) = \delta x_3^*(t) + (1 - b\Delta t)\delta x_3^*(t+\Delta t)$$
$$\delta x_2^*(t) = \delta x_2^*(t) + (x_1(t)\Delta t)\delta x_3^*(t+\Delta t)$$
$$\delta x_1^*(t) = \delta x_1^*(t) + (x_2(t)\Delta t)\delta x_3^*(t+\Delta t)$$
$$\delta x_3^*(t+\Delta t) = 0$$

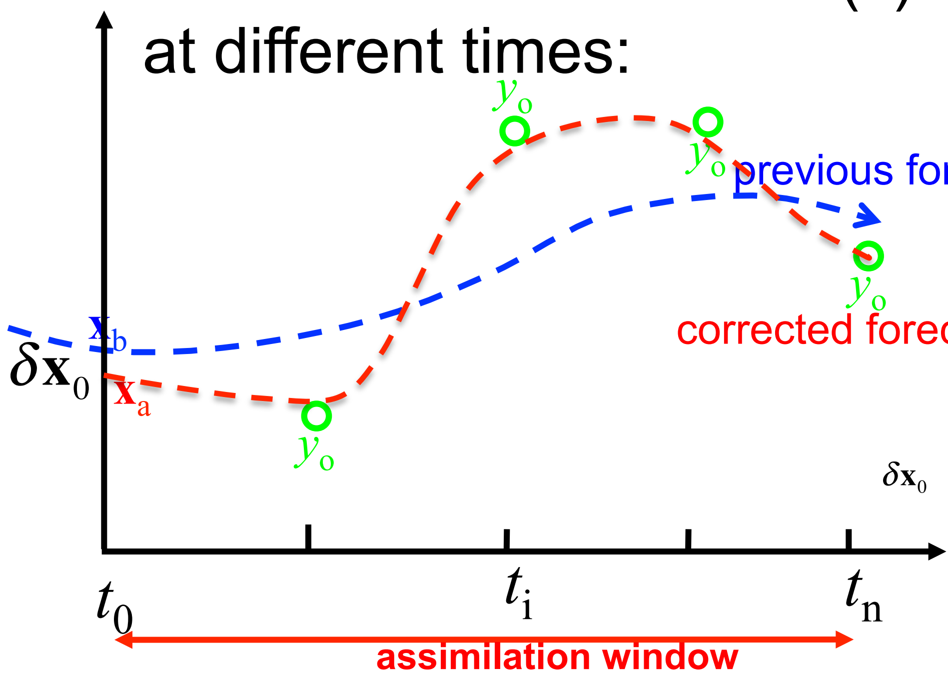
backward in time

SVs summary and extra properties

- To obtain the SVs we need the TLM and the ADJ models.
- The leading SVs are obtained by the Lanczos algorithm.
- One can define an initial and a final norm (size), this gives flexibility (and arbitrariness, Ahlquist, 2000).
- The leading initial SV is the vector that will grow fastest (starting with a very small initial norm and ending with the largest final norm).
- The leading SVs grow initially faster than the Lyapunov vectors, but at the end of the period, they look like LVs (and bred vectors always look like LVs).
- The initial SVs are *very* sensitive to the norm used. The final SVs look like LVs~BVs.

4D-Var

The 3D-Var cost function $J(\mathbf{x})$ is generalized to include observa
at different times:



Find the smallest initial perturbation such that its forecast **best fits the observations within the assimilation interval**

Minimize the 4D-Var cost function for the initial perturbation:

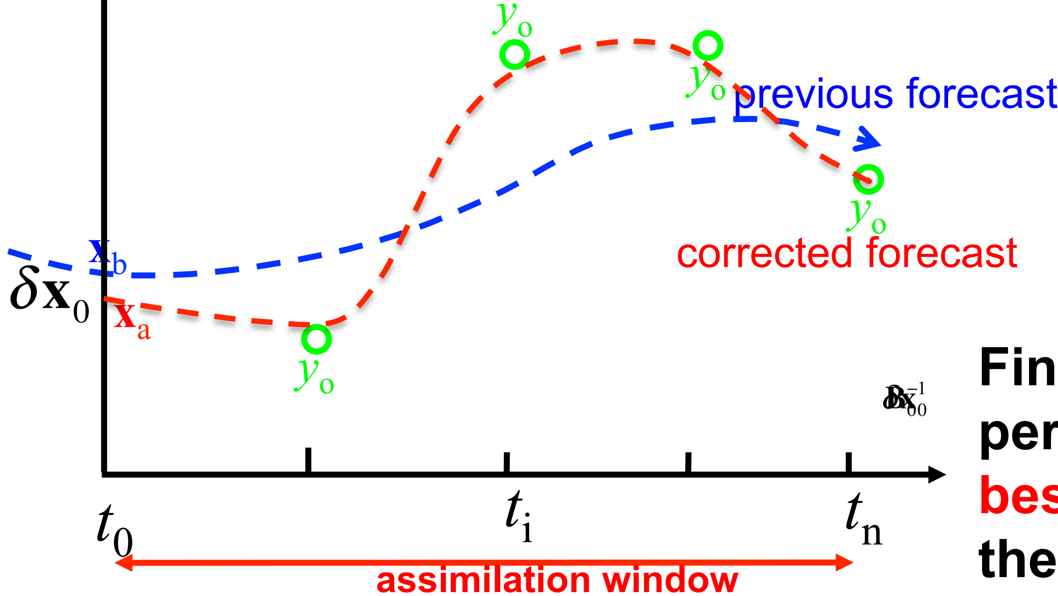
$$J(\delta\mathbf{x}_0) = \frac{1}{2}\delta\mathbf{x}_0^T\mathbf{B}_0^{-1}\delta\mathbf{x}_0 + \frac{1}{2}\sum_{i=1}^{N}\left[\left(\mathbf{H}_i\mathbf{L}(t_0,t_i)\delta\mathbf{x}_0 - \mathbf{d}_i\right)^T\mathbf{R}_i^{-1}\left(\mathbf{H}_i\mathbf{L}(t_0,t_i)\delta\mathbf{x}_0 - \mathbf{d}_i\right)\right]$$

We are looking for the **smallest initial perturbation** that will **grow close to canceling the observational increments**

$$\mathbf{d}_i = \mathbf{y}_i^o - H_i(\mathbf{x}(t_i))$$

4D-Var

The 3D-Var cost function $J(\mathbf{x})$ is generalized to include observa
at different times:



Find the smallest initial perturbation such that its forecast **best fits the observations within the assimilation interval**

Minimize the 4D-Var cost function for the initial perturbation:

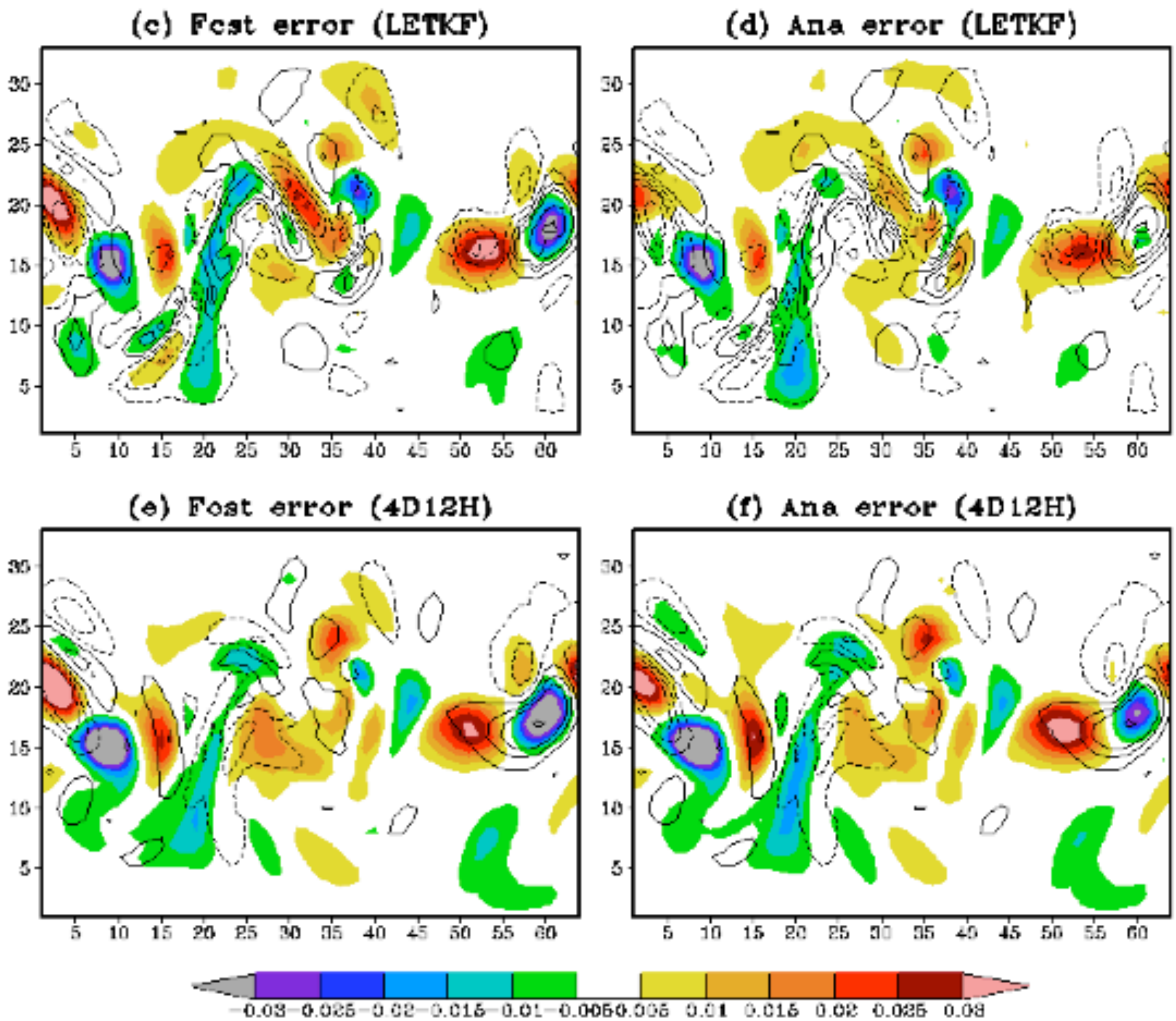
$$J(\delta\mathbf{x}_0) = \frac{1}{2}\delta\mathbf{x}_0^T\mathbf{B}_0^{-1}\delta\mathbf{x}_0 + \frac{1}{2}\sum_{i=1}^{N}\left[\left(\mathbf{H}_i\mathbf{L}(t_0,t_i)\delta\mathbf{x}_0 - \mathbf{d}_i\right)^T\mathbf{R}_i^{-1}\left(\mathbf{H}_i\mathbf{L}(t_0,t_i)\delta\mathbf{x}_0 - \mathbf{d}_i\right)\right]$$

It is evident that the solution to this variational problem will be **dominated by the leading singular vectors** with initial norm $\mathbf{B}_0^{-1}$

Analyses and forecasts at the end of a window

Colors: Forecast errors (left), Analysis errors (right)

Contours: Analysis errors



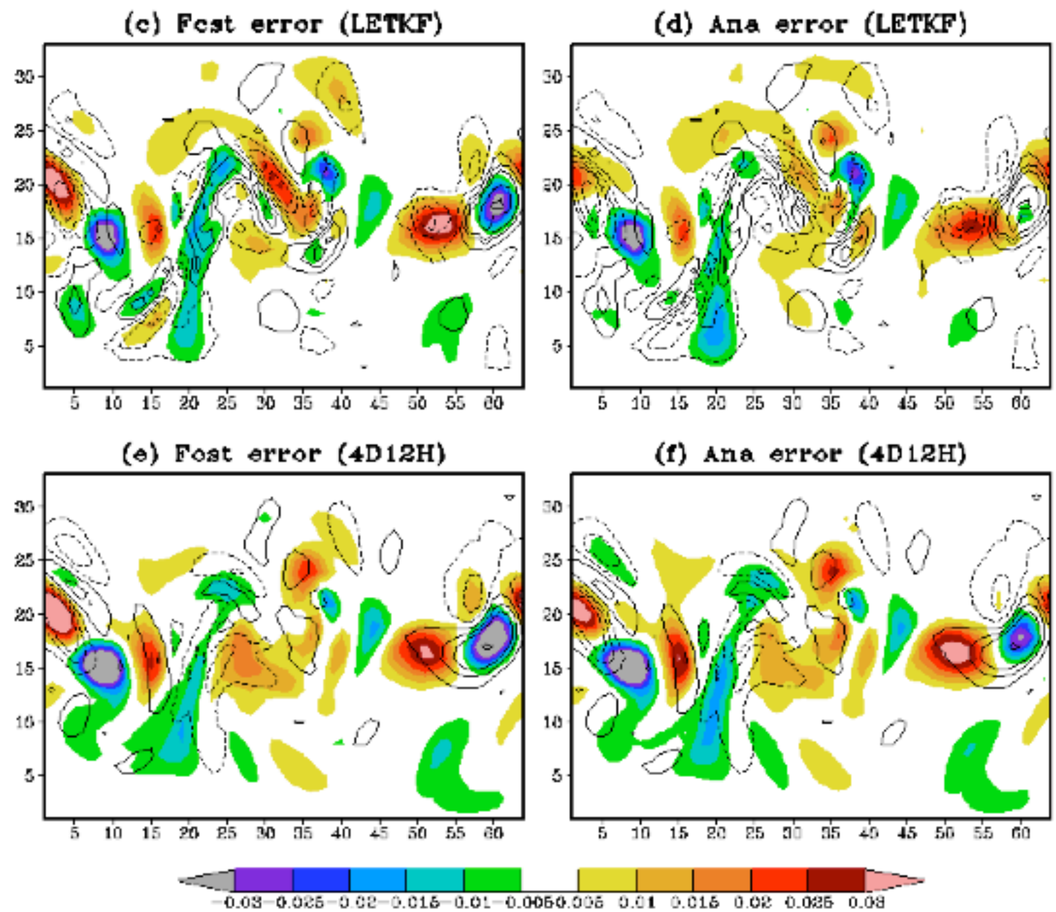
LETKF

4D-Var-12hr

At the end of the assimilation window, the 4D-Var and LETKF corrections are clearly very similar.

What about at the beginning of the assimilation window?

4D-Var is already a smoother, we know the initial corrections.
We can use the “no-cost” LETKF smoother to obtain the

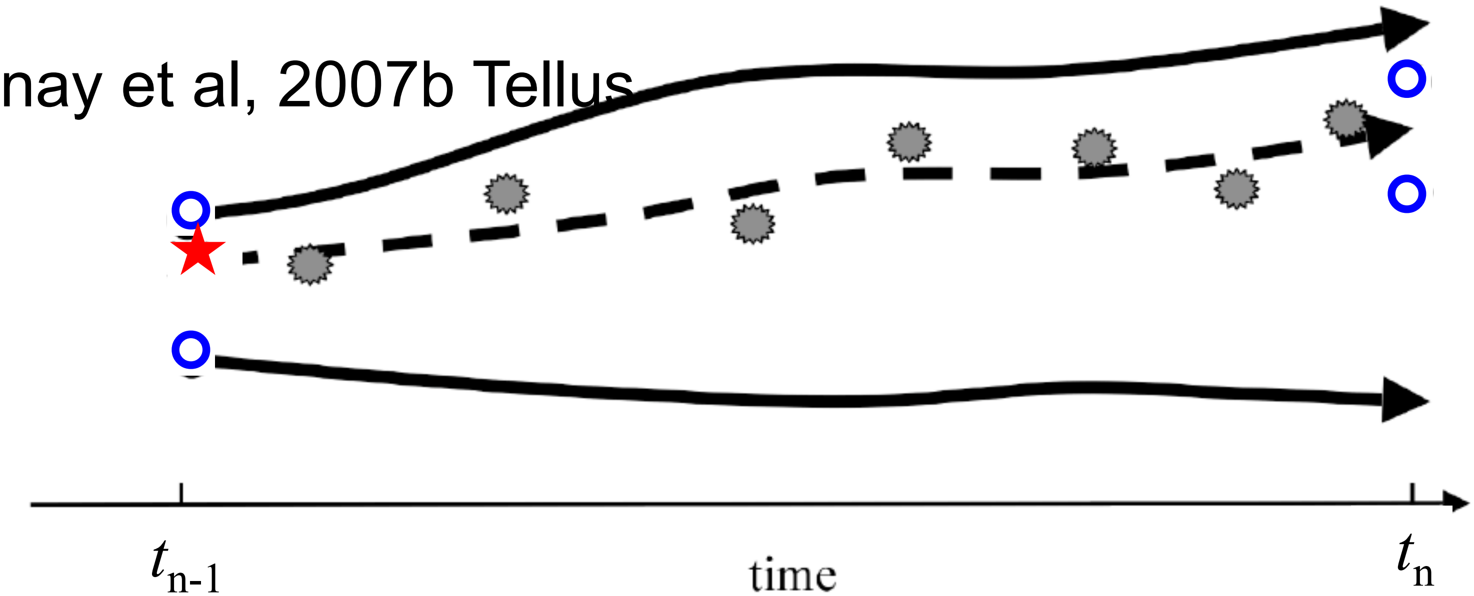


LETKF

4D-Var-12hr

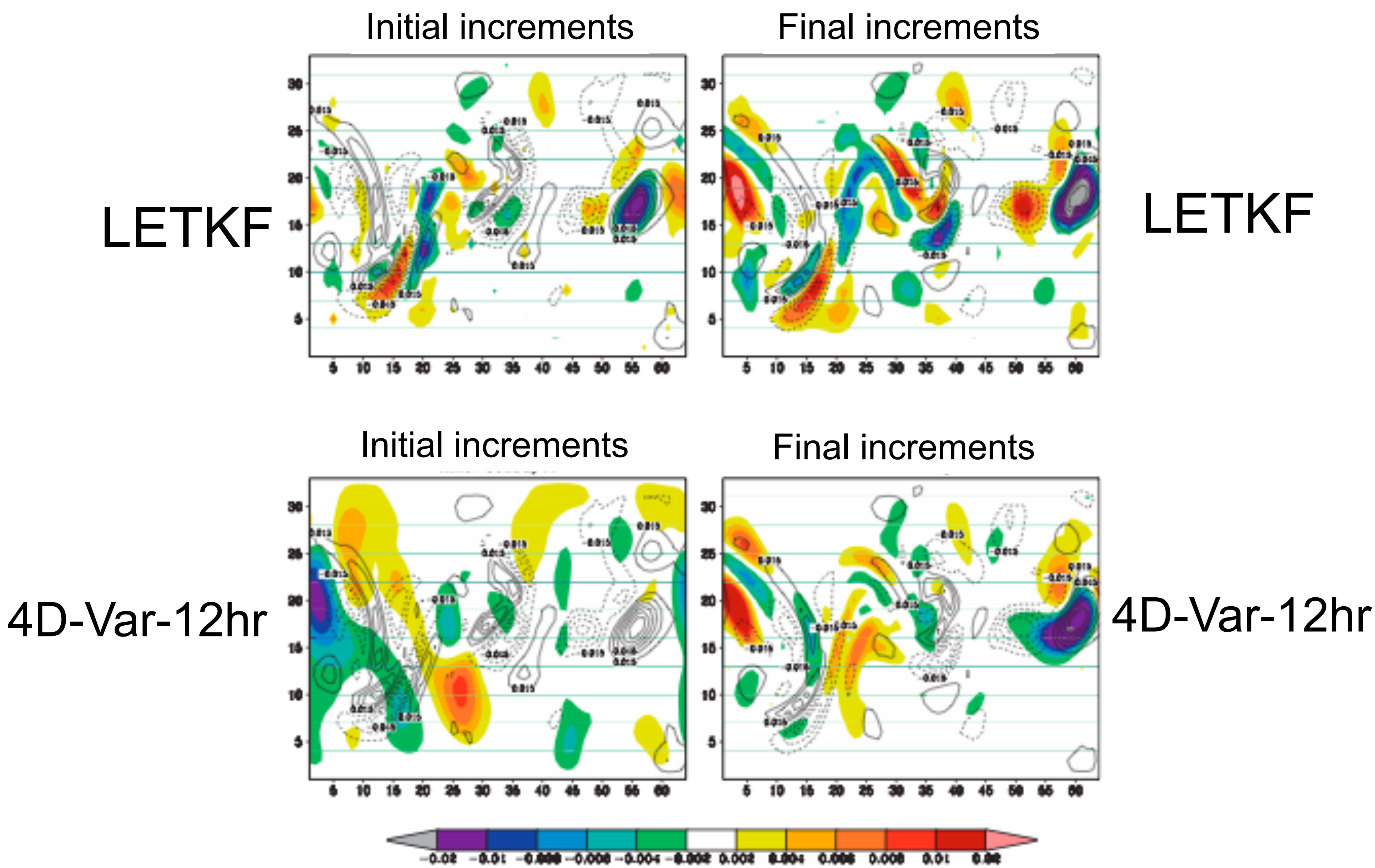
No cost smoother for the LETKF

Kalnay et al, 2007b Tellus



The optimal ETKF weights are obtained at the end of the window, but they are valid for the whole window. We can estimate the 4D-LETKF at any time, simply by applying the weights at that time.

Initial and final analysis corrections (colors), with one BV (contours)



Summary

- Bred Vectors, like leading Lyapunov vectors are norm-independent.
- Initial Singular Vectors depend on the norm.
- 4D-Var is a smoother: it provides an analysis throughout the assimilation window.
- We can define a “No-cost” smoother for the LETKF.
- Applications: Outer Loop and “Running in Place”.
- Comparisons: 4D-Var and LETKF better than 3D-Var.
- Analysis corrections in 3D-Var: missing errors of the day
- Analysis corrections in 4D-Var and LETKF are very similar at the end of the assimilation window.
- Analysis corrections at the beginning of the assimilation window look like bred vectors for the LETKF and like norm-dependent leading singular vectors for 4D-Var.

References

Kalnay et al., Tellus, 2007a (pros and cons of 4D-Var and EnKF)

Kalnay et al., Tellus, 2007b (no cost smoother)

Yang, Carrassi, Corazza, Miyoshi, Kalnay, MWR (2009) (comparison of 3D-Var, 4D-Var and EnKF, no cost smoother)