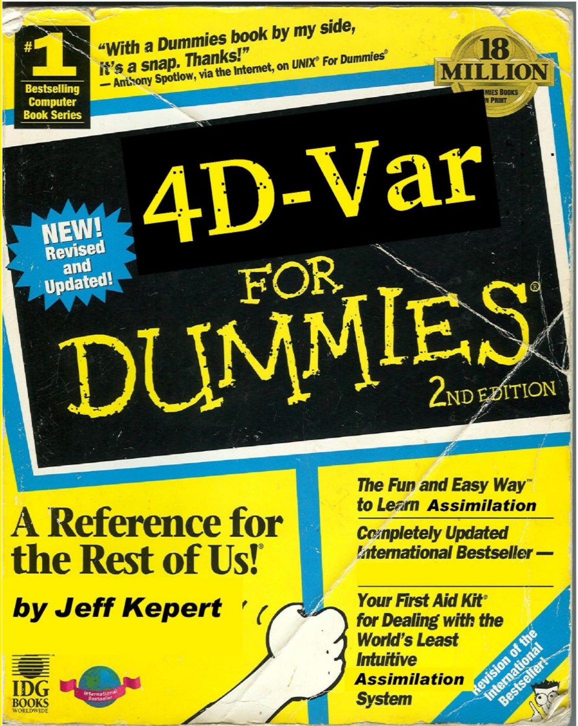
4D-Var for Dummies

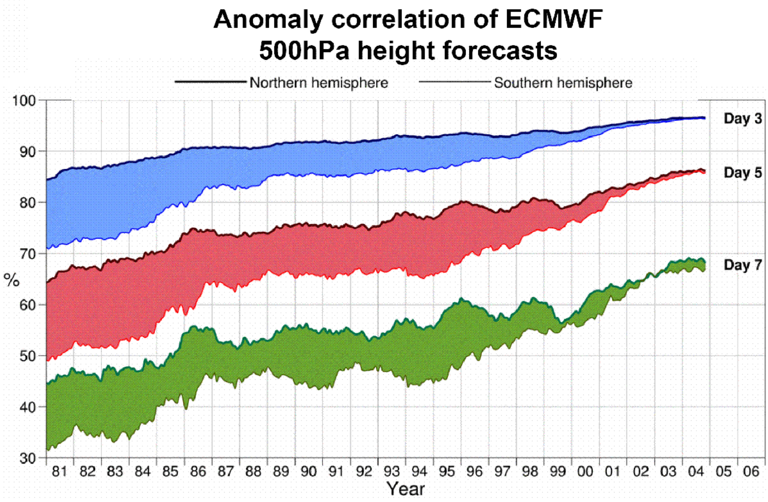
Jeff Kepert

Centre for Australian Weather and Climate Research
A partnership between the Australian Bureau of Meteorology and CSIRO

8th Adjoint Workshop, Pennsylvania, May 17-22 2009


#1
Bestselling Computer Book Series
"With a Dummies book by my side, It's a snap. Thanks!"
— Anthony Spotlow, via the Internet, on UNIX® For Dummies®
18 MILLION
4D-Var
NEW! Revised and Updated!
FOR DUMMIES®
2ND EDITION
The Fun and Easy Way™ to Learn Assimilation
Completely Updated International Bestseller —
Your First Aid Kit® for Dealing with the World's Least Intuitive Assimilation System
A Reference for the Rest of Us!®
by Jeff Kepert
IDG BOOKS WORLDWIDE
International Bestseller
Revision of the International Bestseller!

- 1854: Meteorological Dept of the British Board of Trade created
- “...in a few years .... we might know in this metropolis the condition of the weather 24 hours beforehand.” (M.J.Ball MP, House of Commons, 30 June 1854.)
- Response from House: “Laughter”



Anomaly correlation of ECMWF
500hPa height forecasts
Northern hemisphere
Southern hemisphere
Day 3
Day 5
Day 7
%
100
90
80
70
60
50
40
30
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 00 01 02 03 04 05 06
Year

Why Data Assimilation is Important

- Numerical Weather Prediction (NWP) is (largely) an initial value problem.
  - Has contributed to enormous forecast improvements
  - Extracts the maximum value from expensive observations
- Accurate analyses are necessary for getting the most from field programs.
- Reanalyses of past data using modern methods are an essential resource for climate research.

Best Linear Unbiased Estimate (BLUE)

- Observations y_1 and y_2 of a true state x_t:

$$y_1 = x_t + \epsilon_1 \qquad\qquad y_2 = x_t + \epsilon_2$$

- The statistical properties of the errors are known:

$$\langle \epsilon_1 \rangle = 0 \qquad \langle \epsilon_1^2 \rangle = \sigma_1^2 \qquad \langle \epsilon_1 \epsilon_2 \rangle = 0$$
$$\langle \epsilon_2 \rangle = 0 \qquad \langle \epsilon_2^2 \rangle = \sigma_2^2$$

- Estimate x_a of x_t as a linear combination of the observations such that $\langle x_a \rangle = x_t$ (unbiased) and $\sigma_a^2 = \langle (x_a - x_t)^2 \rangle$ is minimised (best).
- Then

$$x_t = \frac{\sigma_2^2 y_1 + \sigma_1^2 y_2}{\sigma_1^2 + \sigma_2^2}$$

Best Linear Unbiased Estimate (cont'd)

- Same estimate found by minimising

$$J(x_a) = \frac{(x_a - y_1)^2}{\sigma_1^2} + \frac{(x_a - y_2)^2}{\sigma_2^2}$$

- Minimising J is the same as maximising $\exp(-J/2)$
- Hence for Gaussian errors the BLUE is the maximum likelihood (or optimal) estimate.
- For many pieces of data $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$,

$$J(x_a) = (x_a - \mathbf{y})^T \mathbf{P}^{-1} (x_a - \mathbf{y})$$

where $\mathbf{P}$ is the error covariance matrix of $\mathbf{y}$.

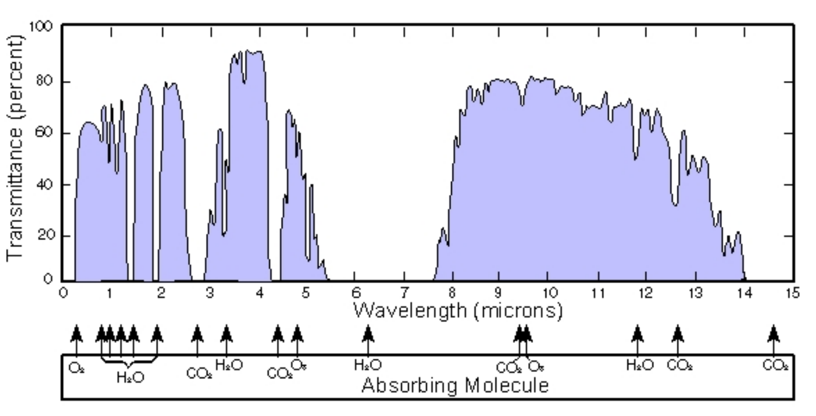
Assimilation: The Big BLUE 

Assimilation combines a short-term numerical forecast with some observations:

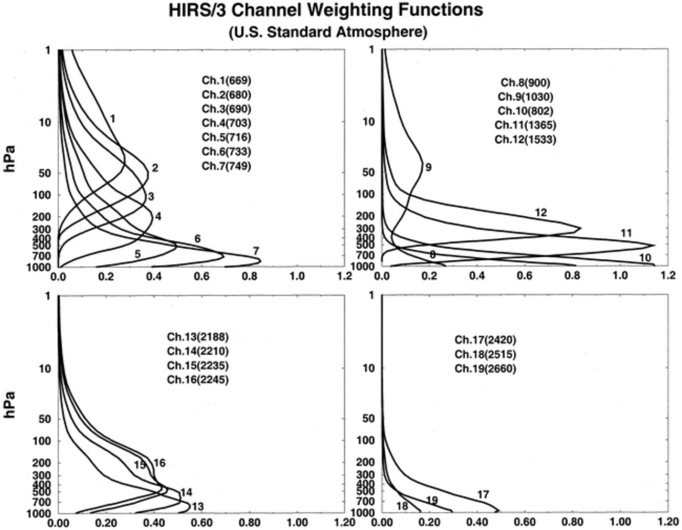
$$J(\mathbf{x}_a) = (\mathbf{x}_a - \mathbf{x}_f)^T \mathbf{B}^{-1} (\mathbf{x}_a - \mathbf{x}_f) + (\mathcal{H}(\mathbf{x}_a) - \mathbf{y})^T \mathbf{R}^{-1} (\mathcal{H}(\mathbf{x}_a) - \mathbf{y})$$

- $\mathbf{x}_a$ is the analysis
- $\mathbf{x}_f$ the short-term forecast
- $\mathbf{y}$ are the observations
- $\mathcal{H}$ produces the analysis estimate of the observed values
- $\mathbf{R}$ is the observation error covariance
- $\mathbf{B}$ is the forecast error covariance

The Atmospheric Infrared Transmission Spectrum



HIRS (High resolution InfraRed Sounder) Channel Weights



Finding the minimum of J

$$J(\mathbf{x}_a) = (\mathbf{x}_a - \mathbf{x}_f)^T \mathbf{B}^{-1} (\mathbf{x}_a - \mathbf{x}_f) + (\mathcal{H}(\mathbf{x}_a) - \mathbf{y})^T \mathbf{R}^{-1} (\mathcal{H}(\mathbf{x}_a) - \mathbf{y})$$

Solve directly $\nabla J = 0$.

- Have to manipulate big matrices
- Nonlinear $\mathcal{H}$ is very difficult (satellite radiances)

Iterative minimisation (a.k.a. variational assimilation).

- Finds full 3-D structure of the atmosphere (3D-Var)
- Other observations and background helps constrain the poorly-conditioned and underdetermined inversion of the satellite radiances

Minimising J

$$J(\mathbf{x}_a) = (\mathbf{x}_a - \mathbf{x}_f)^T \mathbf{B}^{-1} (\mathbf{x}_a - \mathbf{x}_f) + (\mathcal{H}(\mathbf{x}_a) - \mathbf{y})^T \mathbf{R}^{-1} (\mathcal{H}(\mathbf{x}_a) - \mathbf{y})$$

To minimise J, we need the gradient:

$$\nabla J(\mathbf{x}_a) = 2\mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_f) + 2\mathbf{H}^T \mathbf{R}^{-1} (\mathcal{H}(\mathbf{x}_a) - \mathbf{y})$$

$\mathbf{H} = \left[\frac{\partial \mathcal{H}_i}{\partial \mathbf{x}_{a,j}} \right]$ is the Jacobian of $\mathcal{H}$ (a.k.a. the tangent linear)
$\mathbf{H}^T$ is the adjoint of $\mathcal{H}$

The Importance of **B**

$$J(\mathbf{x}_a) = (\mathbf{x}_a - \mathbf{x}_f)^T \mathbf{B}^{-1} (\mathbf{x}_a - \mathbf{x}_f) + (\mathcal{H}(\mathbf{x}_a) - \mathbf{y})^T \mathbf{R}^{-1} (\mathcal{H}(\mathbf{x}_a) - \mathbf{y})$$

B is important:

- Conditioning and speed of convergence
- Getting the statistics right
- Describing atmospheric balance
- Spatial scale of analysis

B in the model variables fails miserably:

- Rank deficient
- Too large to store, let alone operate on

B the best you can B

Representing **B** typically involves:

- *Transform to less-correlated variables.*
  - $(u, v) \Longrightarrow (\psi, \chi)$
  - $u = -\partial\psi/\partial y + \partial\chi/\partial x$, $v = \partial\psi/\partial x + \partial\chi/\partial y$
  - Replace mass field by unbalanced mass: $\phi_{\text{unbal}} = \phi - \phi_{\text{bal}}(\psi)$
- *Transform to spectral space.*
- *Rescale.*

These make **B** diagonal $\Longrightarrow$ good conditioning and computational efficiency.

- *Truncate the small scales.* Forecast error spectrum is red, with little power at small scales. So truncate **B**.

Incremental Formulation

Replace

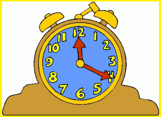
$$J(\mathbf{x}_a) = (\mathbf{x}_a - \mathbf{x}_f)^T \mathbf{B}^{-1} (\mathbf{x}_a - \mathbf{x}_f) + (\mathcal{H}(\mathbf{x}_a) - \mathbf{y})^T \mathbf{R}^{-1} (\mathcal{H}(\mathbf{x}_a) - \mathbf{y})$$

by

$$J(\delta\mathbf{x}) = \delta\mathbf{x}^T \mathbf{B}^{-1} \delta\mathbf{x} + (\mathcal{H}(\mathbf{x}_f) + \mathbf{H}\delta\mathbf{x} - \mathbf{y})^T \mathbf{R}^{-1} (\mathcal{H}(\mathbf{x}_f) + \mathbf{H}\delta\mathbf{x} - \mathbf{y})$$

where $\delta\mathbf{x} = \mathbf{x}_a - \mathbf{x}_f$ and **H** is the Jacobian of $\mathcal{H}$ (tangent linear).

- $\mathcal{H}(\mathbf{x}_a)$ becomes $\mathcal{H}(\mathbf{x}_f) + \mathbf{H}\delta\mathbf{x}$
- Computational efficiency since $\delta\mathbf{x}$ now at reduced resolution of **B**, **H** maybe cheaper to compute than $\mathcal{H}$, true quadratic form.



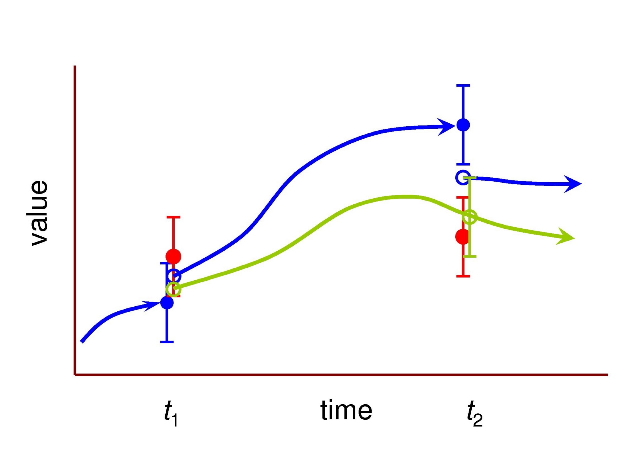
A Matter of Time

So far all data assumed to be at the analysis time.

- Assimilate e.g. four times a day.
- All data in 6-hour window assumed to occur at the middle of that window.
- Introduces some errors $\Longrightarrow$ weather systems move and develop!
- Reduce errors by assimilating more frequently, but that has its own problems.

A better way is to introduce the time dimension into the assimilation, 4-dimensional variational assimilation (4D-Var).

Observations at two times



Red: Observations. Blue: 3D-Var. Green: 4D-Var.

4D-Var

Add a term for the later time:

$$J(\mathbf{x}_a) = \ldots + (\mathcal{H}_2(\mathcal{M}(\mathbf{x}_a)) - \mathbf{y}_2)^T \mathbf{R}_2^{-1} (\mathcal{H}_2(\mathcal{M}(\mathbf{x}_a)) - \mathbf{y}_2)$$

- $\mathcal{M}$ is the model forecast from t_1 to t_2
- Subscripts 2 refer to the time t_2.

The gradient becomes

$$\nabla J(\mathbf{x}_a) = \ldots + 2\mathbf{M}^T \mathbf{H}_2^T \mathbf{R}_2^{-1} (\mathcal{H}_2(\mathcal{M}(\mathbf{x}_a)) - \mathbf{y}_2)$$

- $\mathbf{H}_2^T$ is the adjoint of the Jacobian of $\mathcal{H}$, takes information about the observation-analysis misfit from radiance space to analysis space
- $\mathbf{M}^T$ is the adjoint of the Jacobian of $\mathcal{M}$ and propagates this gradient information back in time from t_2 to t_1.

4D-Var

Minimising

$$\begin{aligned} J(\mathbf{x}_a) &= (\mathbf{x}_a - \mathbf{x}_f)^T \mathbf{B}^{-1} (\mathbf{x}_a - \mathbf{x}_f) \\ &+ (\mathcal{H}(\mathbf{x}_a) - \mathbf{y})^T \mathbf{R}^{-1} (\mathcal{H}(\mathbf{x}_a) - \mathbf{y}) \\ &+ (\mathcal{H}_2(\mathcal{M}(\mathbf{x}_a)) - \mathbf{y}_2)^T \mathbf{R}_2^{-1} (\mathcal{H}_2(\mathcal{M}(\mathbf{x}_a)) - \mathbf{y}_2) \end{aligned}$$

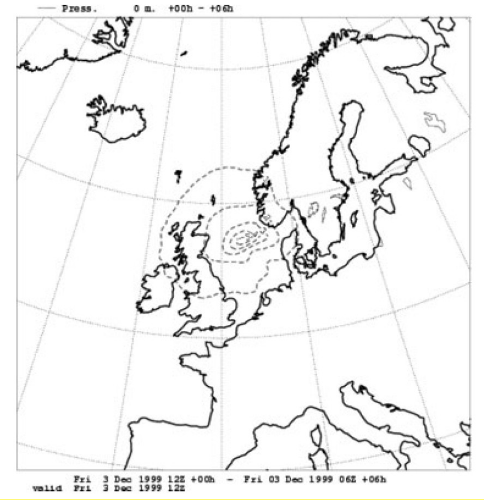
gives an analysis $\mathbf{x}_a$ at time t_1 that

- is close to the background $\mathbf{x}_f$ at t_1
- is close to the observations $\mathbf{y}$ at t_1
- initialises a (linearised) forecast that is close to the observations $\mathbf{y}_2$ at time t_2

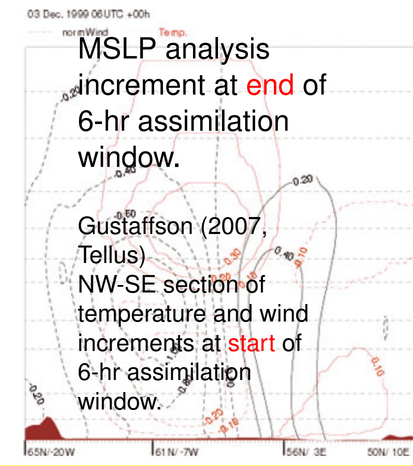
Adding additional time levels is straightforward, as is the incremental formulation (exercise).

4D-Var analysis of a single pressure observation

One pressure observation at centre of low, 5hPa below background, at end of 6-hr assimilation window.



MSLP analysis increment at end of 6-hr assimilation window.



Gustaffson (2007, Tellus)
NW-SE section of temperature and wind increments at start of 6-hr assimilation window.

In practice ...

This is not a small problem!

- Atmospheric model has $O(10^6 \text{ to } 10^7)$ variables
- Millions of observations per day
- Limited time available under operational constraints

The model has several hundred thousand lines of code, 4D-Var requires

- operations by the Jacobian of the model
- operations by the adjoint of the Jacobian

Good results require accurately estimating the necessary statistics ($\mathbf{R}$ and $\mathbf{B}$) and careful quality control of the observations.

Extensions

Multiple “Outer Loops”

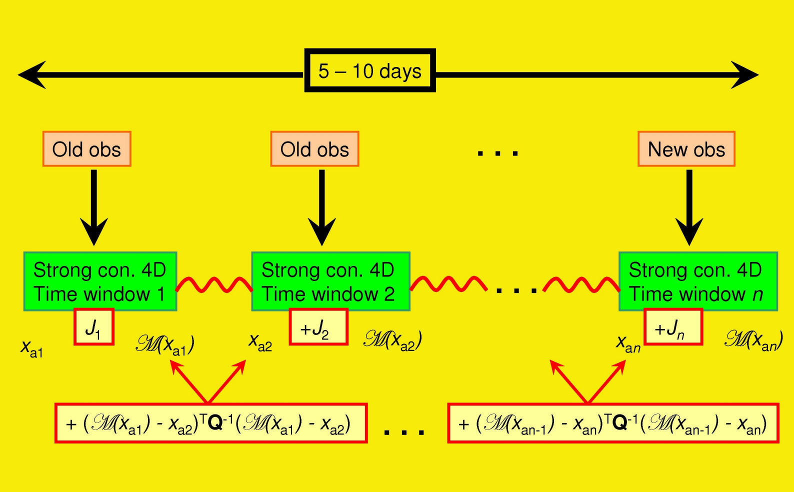
- *Problem:* Accuracy is limited by the linearisations of $\mathcal{H}$ and $\mathcal{M}$.
- *Solution:* Update the nonlinear forecast (outer loop) several times during the minimisation of the $J(\delta\mathbf{x})$ (inner loop).

Multi-incremental 4D-Var

- *Problem:* Balancing speed of convergence against need to resolve small scales.
- *Solution:* Begin minimising with $\delta\mathbf{x}$ at low resolution, and increase resolution after each iteration of the outer loop.

Extensions: Weak Constraint 4D-Var

- Doesn't assume that the model is perfect
- Allows a longer window.



Summary

Var is better than direct solution (a.k.a. Optimum Interpolation) because:

- Can handle lots of observations
- Can better cope with nonlinear observation operator $\mathcal{H}$
- Solves for the whole domain at once

4D-Var is better than 3D-Var because:

- Uses observations at the correct time
- Calculates analysis at the correct time
- Implicitly generates flow-dependent **B**

Special issues of QJRMS and JMetSocJapan from WMO DA workshops.

Kepert, J.D., 2007: Maths at work in meteorology. *Gazette of the Australian Mathematical Society,* **34**, 150 – 155. Available from http://www.austms.org.au/Publ/Gazette/