Introduction to Data Assimilation

Saroja Polavarapu
Meteorological Service of Canada
University of Toronto

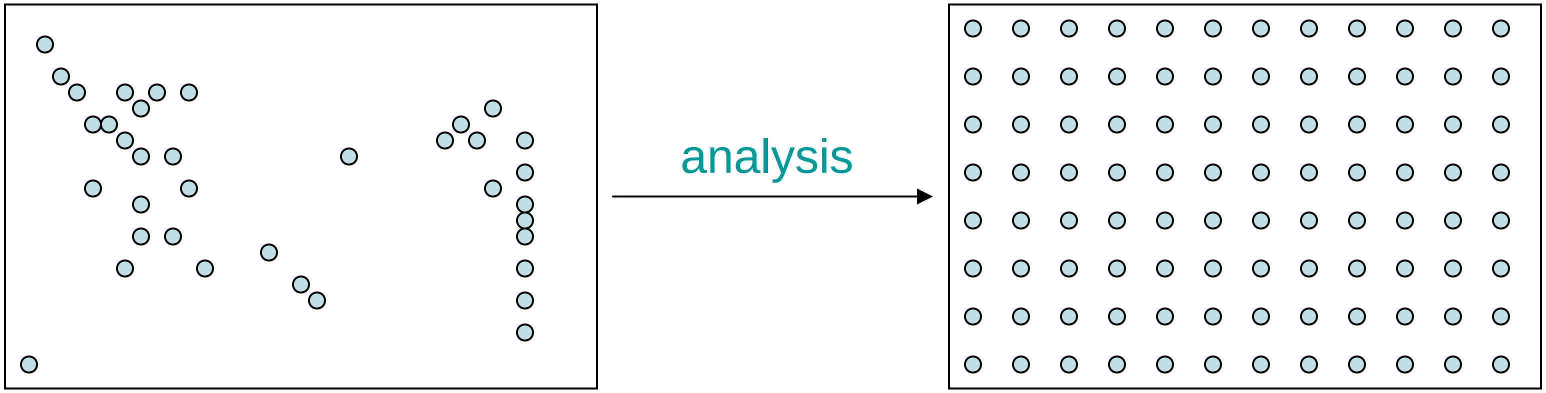
GCC Summer School, Banff. May 22-28, 2004

Outline of lectures

- General idea
- Numerical weather prediction context
- Simple scalar example
- Two observations on a 1D grid
- Optimal Interpolation
- Basic estimation theory
- 3D-Variational Assimilation (3Dvar)
- Covariance modelling

Atmospheric Data Analysis

Goal: To produce a regular, physically consistent, four-dimensional representation of the state of the atmosphere from a heterogeneous array of in-situ and remote instruments which sample imperfectly and irregularly in space and time. (Daley, 1991)



- Approach: Combine information from past observations, brought forward in time by a model, with information from new observations, using
  - statistical information on model and observation errors
  - the physics captured in the model
- Observation errors
  - Instrument, calibration, coding, telecommunication errors
- Model errors
  - “representativeness”, numerical truncation, incorrect or missing physical processes

Analysis = Interpolation + Filtering

Why do people do data assimilation?

1. To obtain an initial state for launching NWP forecasts
2. To make consistent estimates of the atmospheric state

   for diagnostic studies.

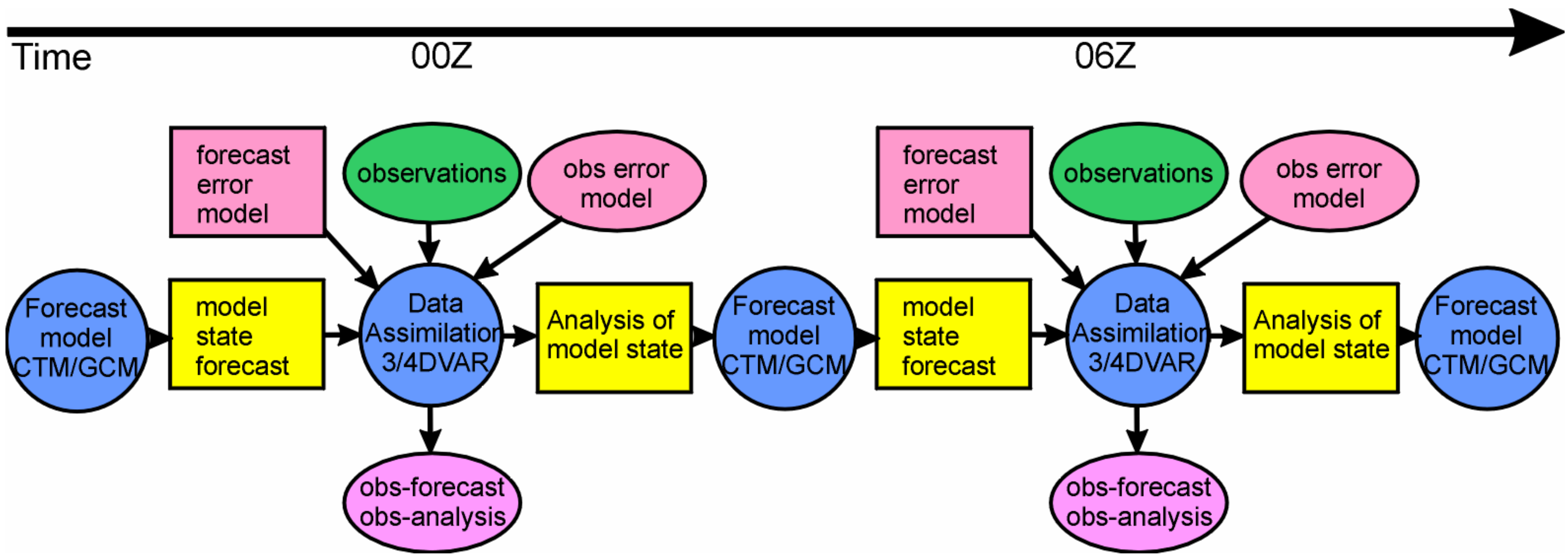
   - reanalyses (eg. ERA-15, ERA-40, NCEP, etc.)
3. For an increasingly wide range of applications (e.g. atmospheric chemistry)
4. To challenge models with data and vice versa
   - UKMO analyses during UARS (1991-5) period

Producing a Numerical Weather Forecast

1. Observation
   - Collect, receive, format and process the data
   - quality control the data
2. Analysis
   - Use data to obtain a spatial representation of the atmosphere
3. Initialization
   - Filter noise from analysis
4. Forecast
   - Integrate initial state in time with full PE model and parameterized physical processes

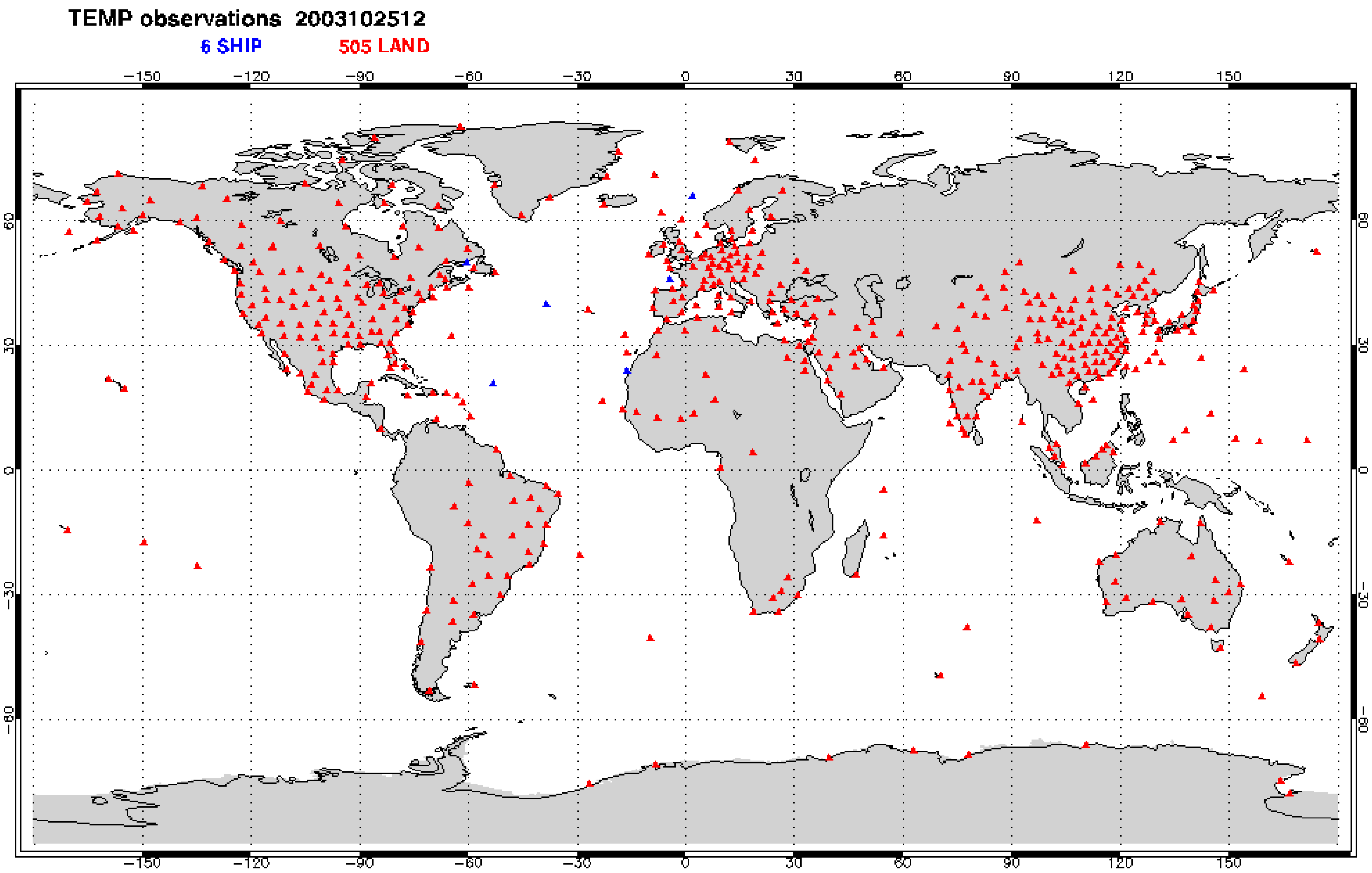
Data Assimilation

Data Assimilation Cycles

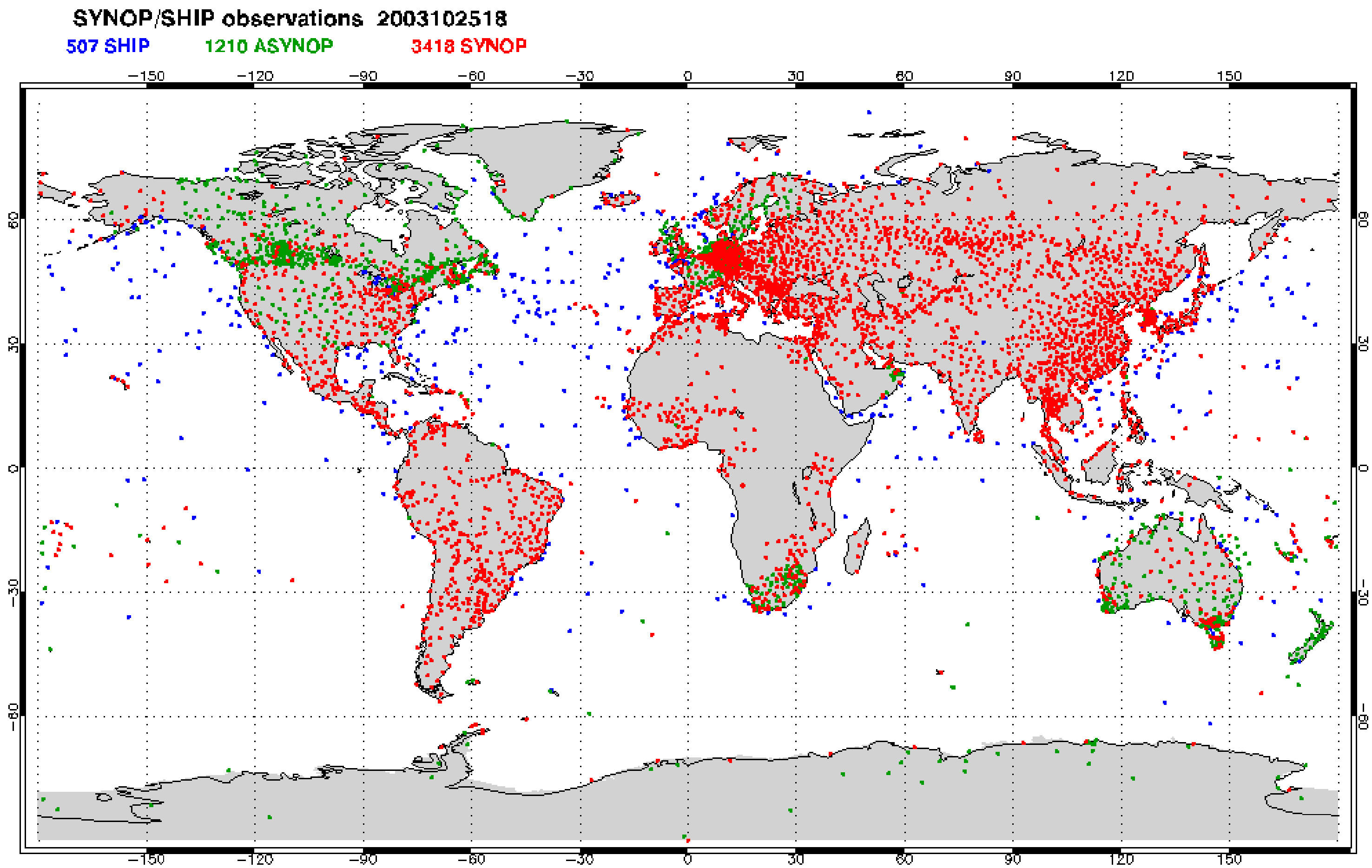


Observations currently in use at CMC

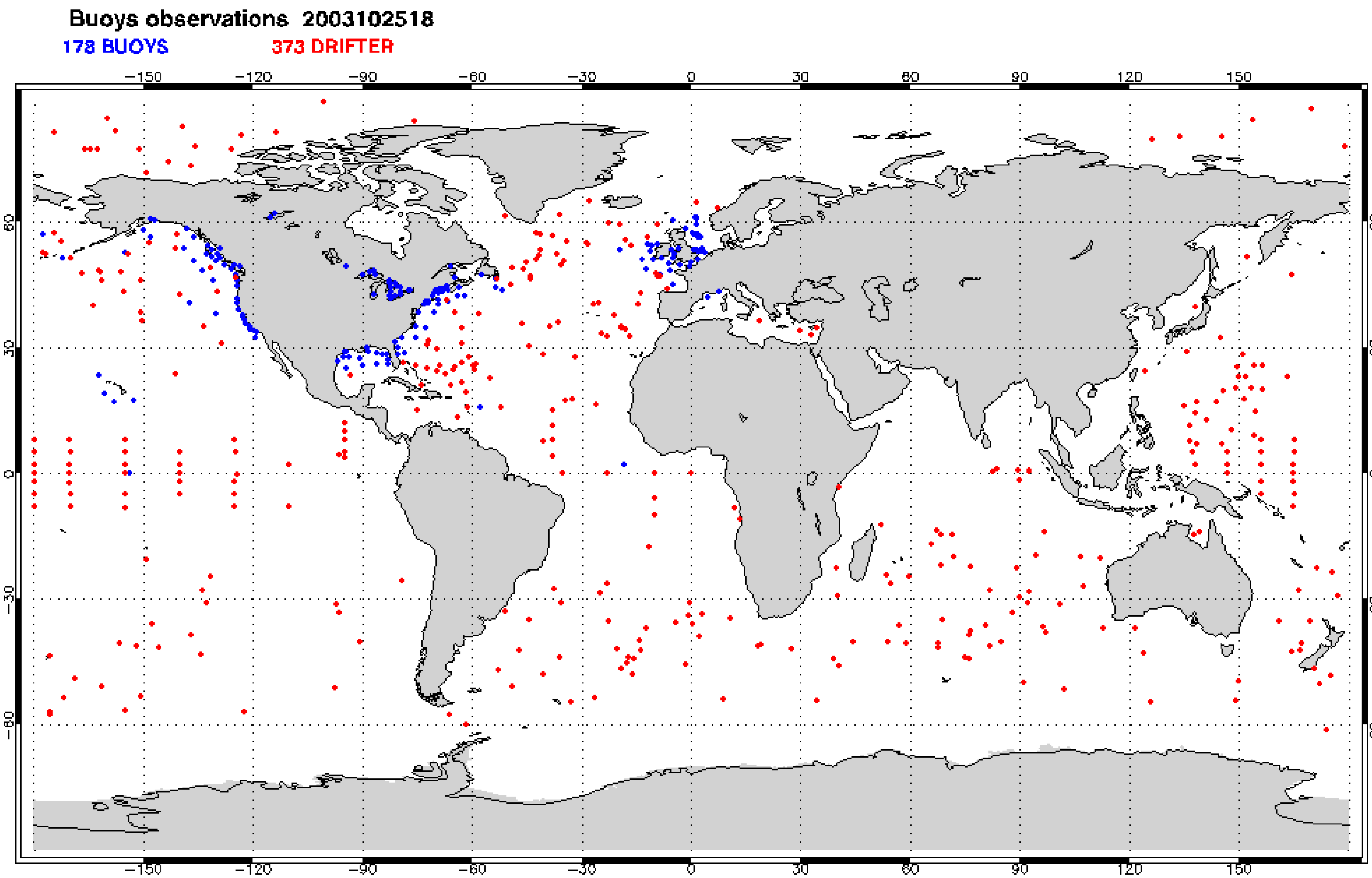
Radiosonde observations used in GDAS (12Z)



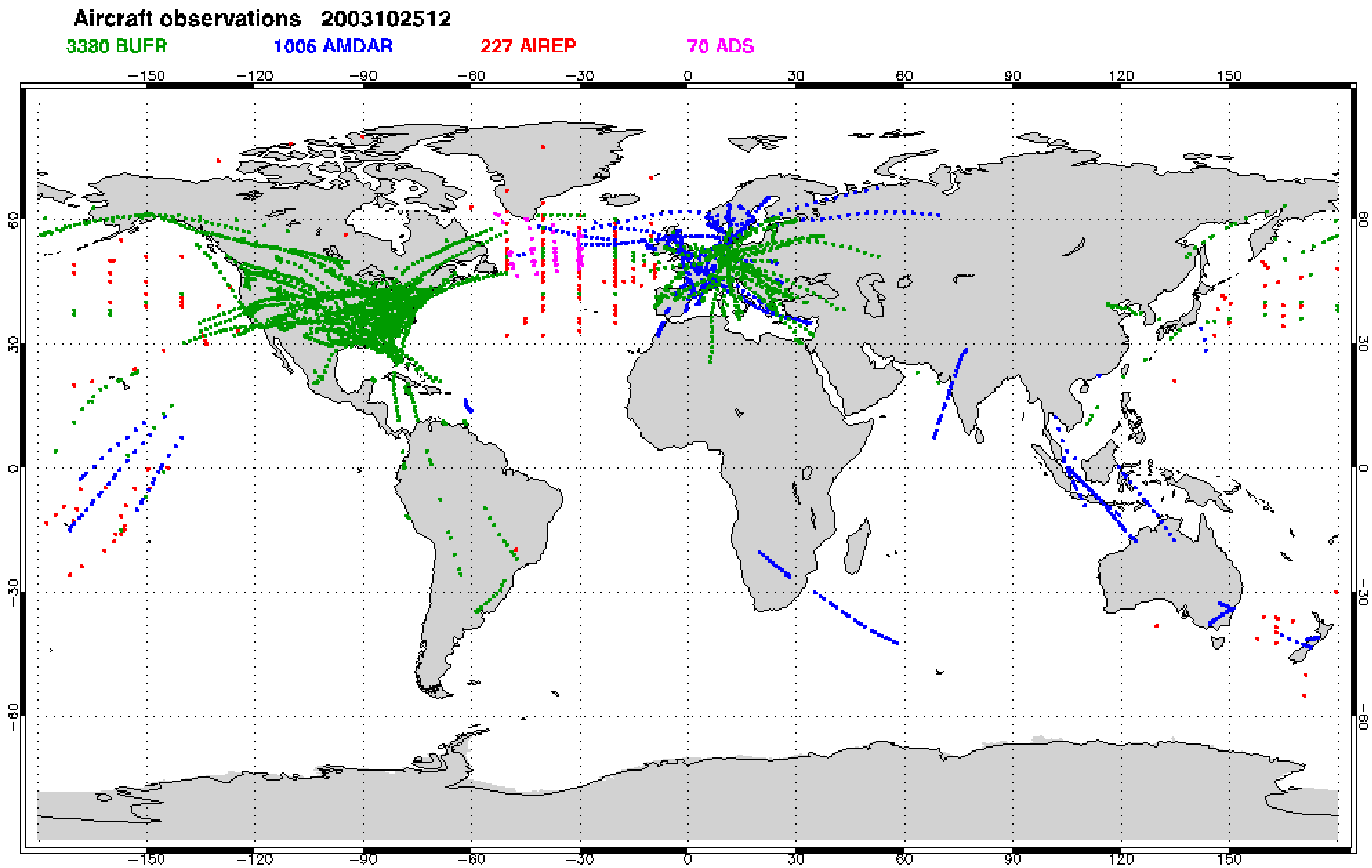
SYNOP and SHIP observations used in GDAS



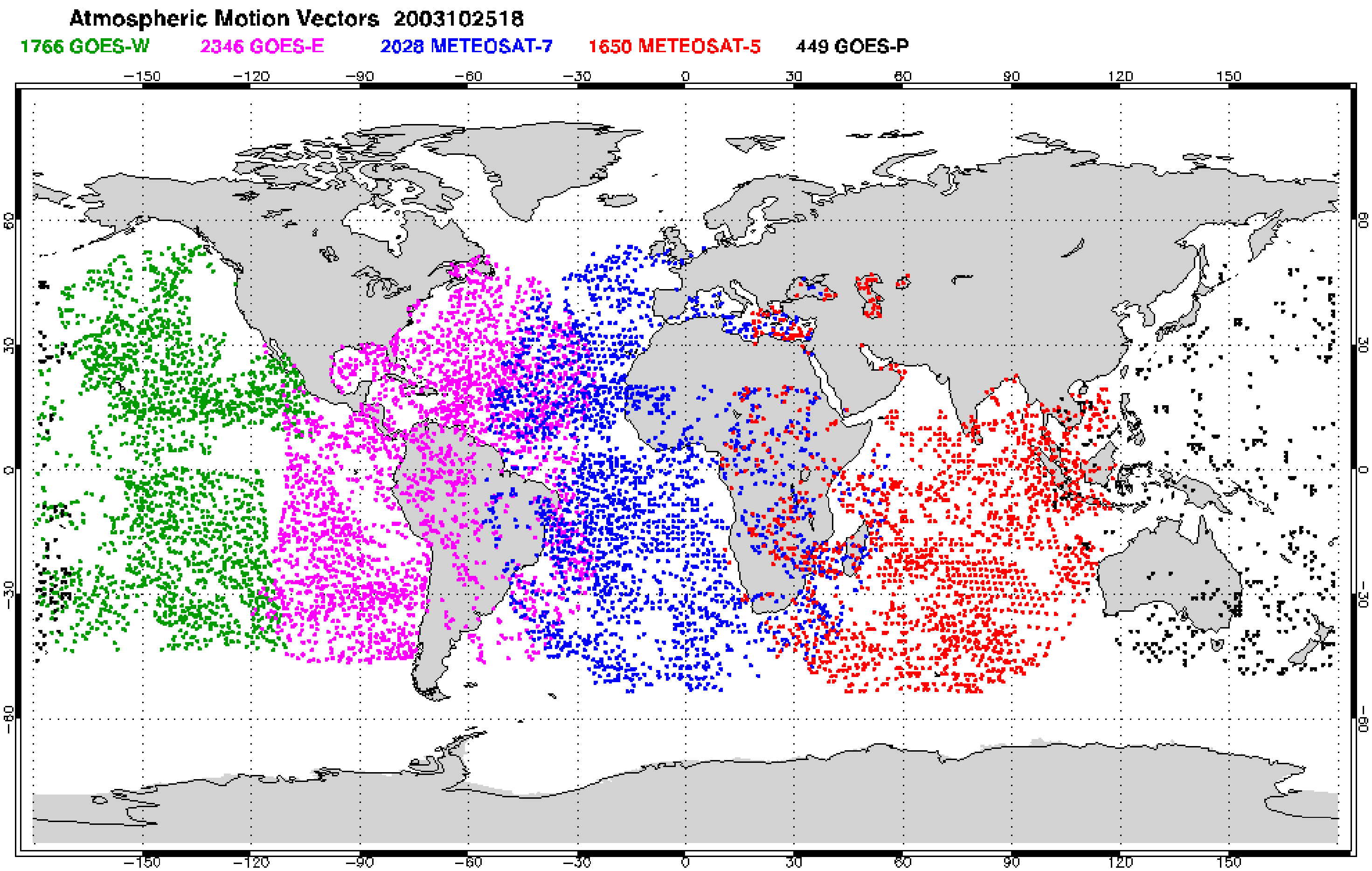
Buoy observations used in GDAS



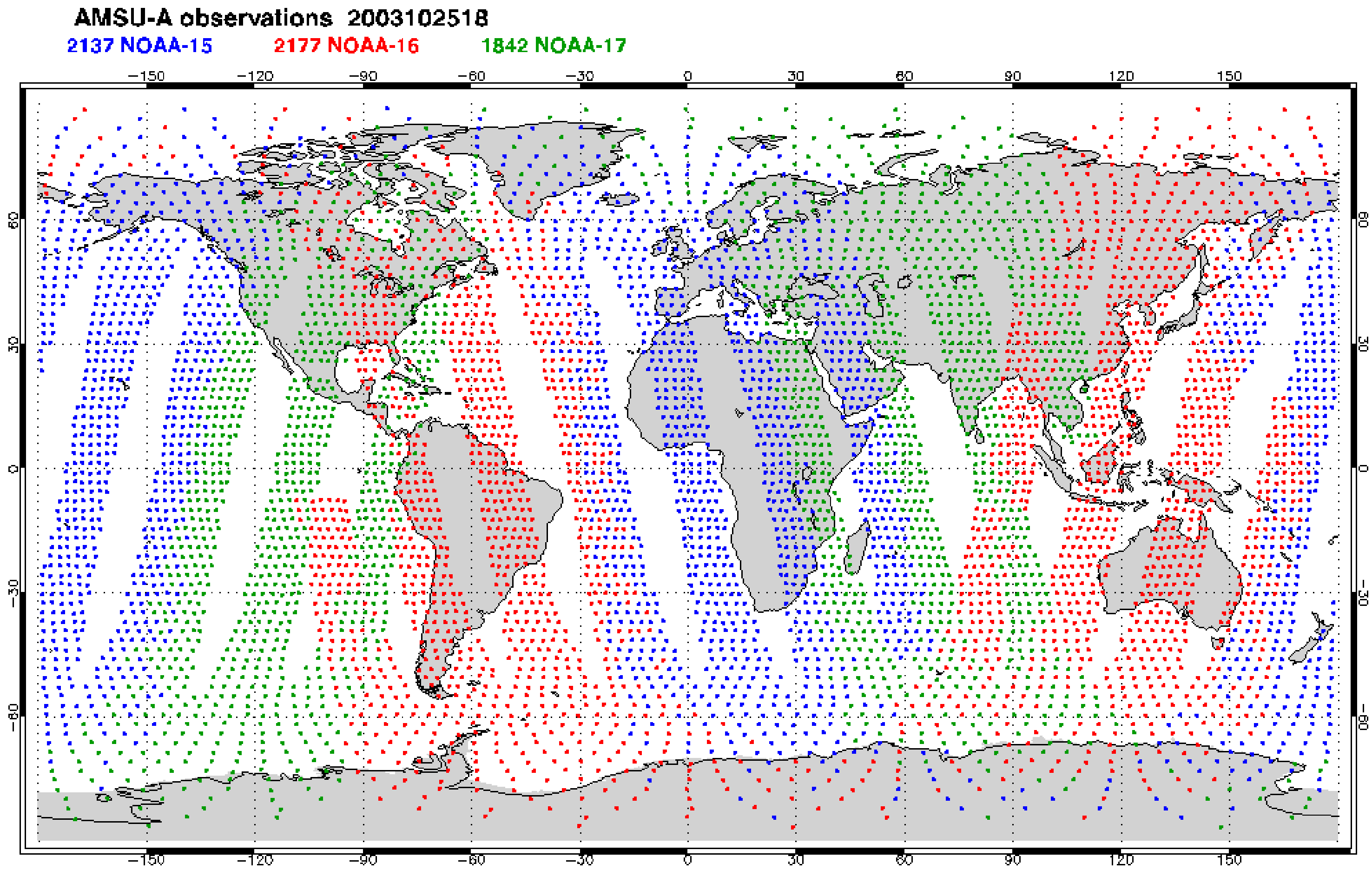
Aircraft observations used in GDAS



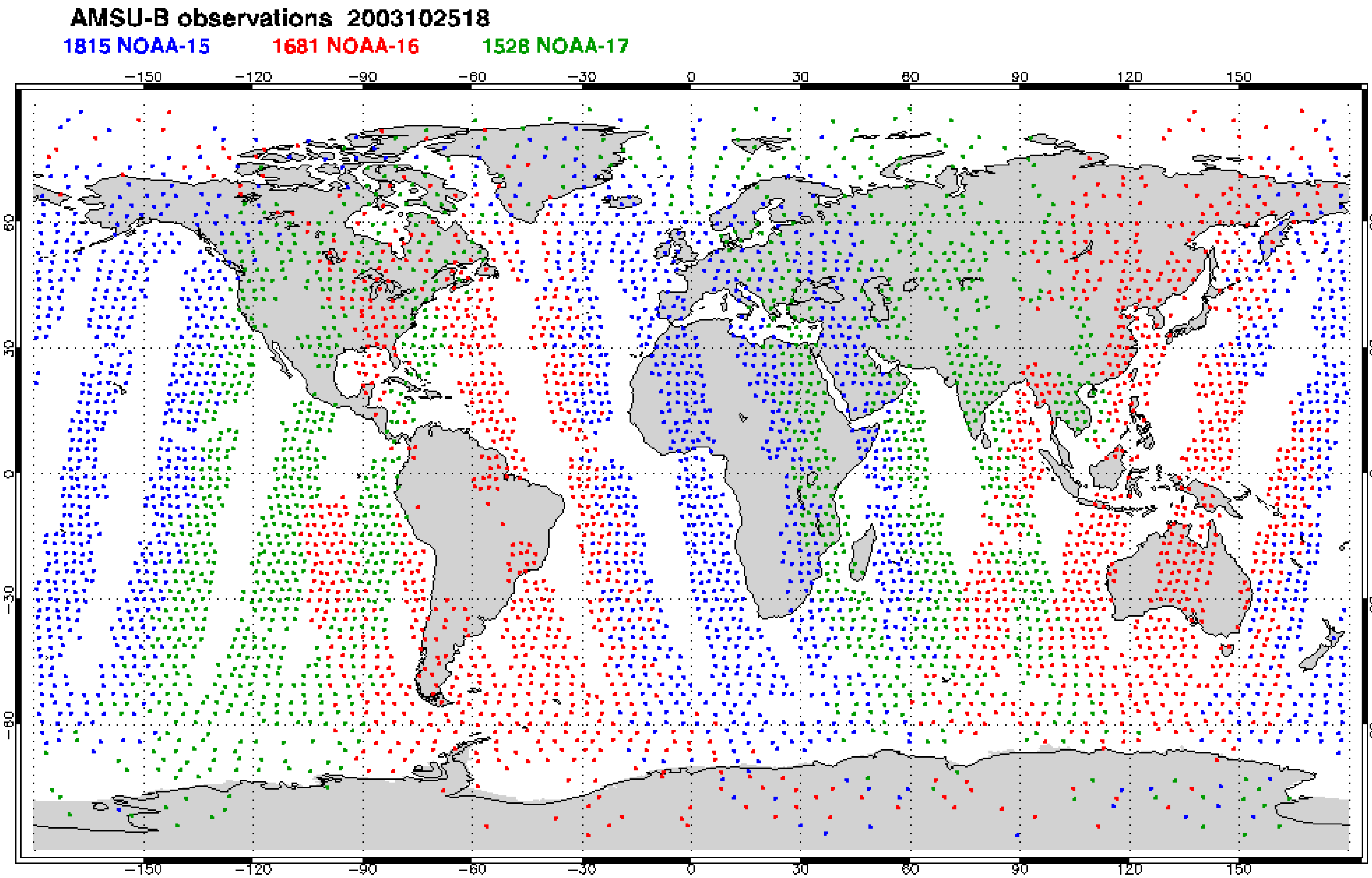
Cloud motion wind observations used in GDAS



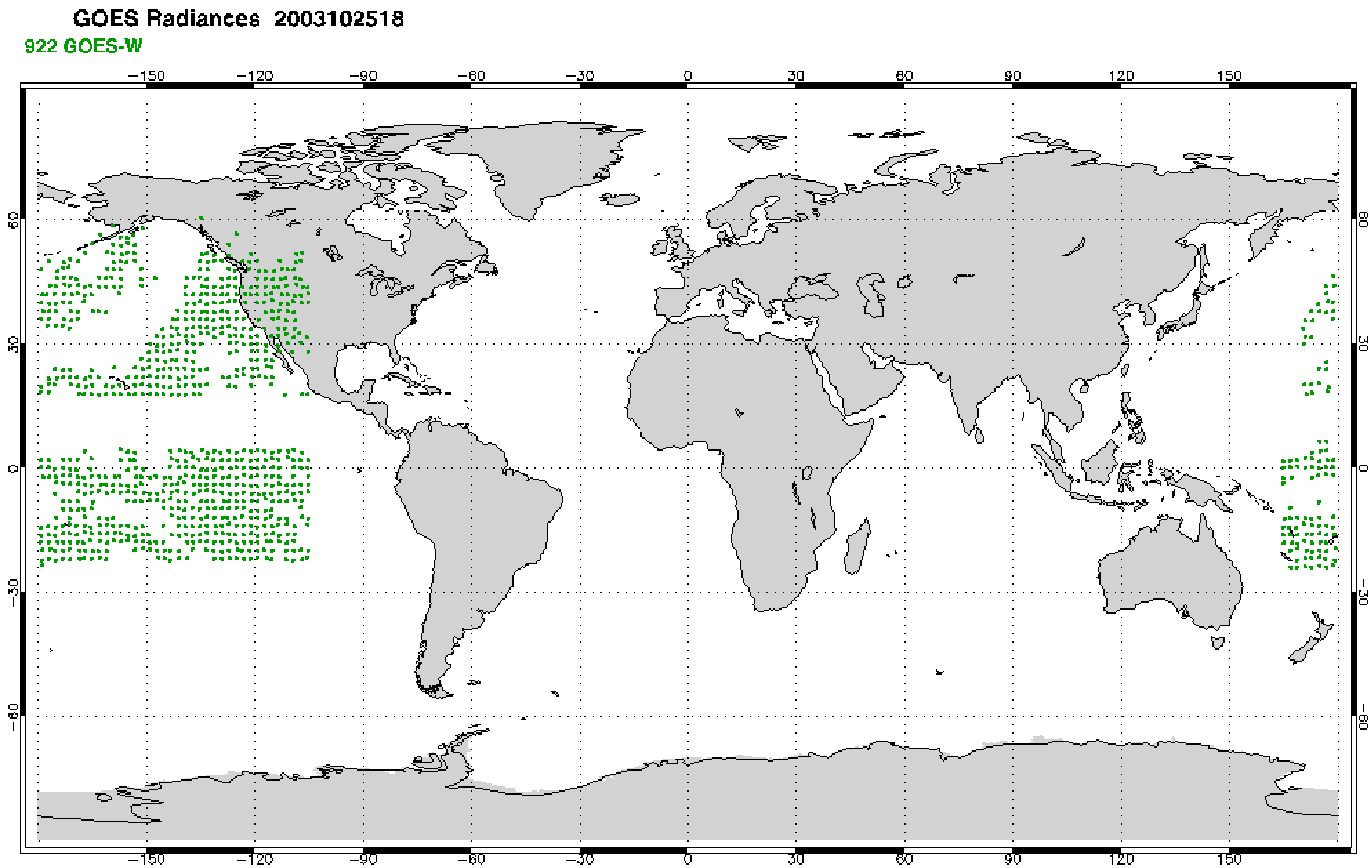
AMSU-A observations used in GDAS



AMSU-B observations used in GDAS



GOES radiances used in GDAS



Underdeterminacy

X = state vector

Model	Lat x long x lev x variables
CMC global oper.	400x200x28x4 =9x10^6
CMC meso-global	800x400x58x4 =7x10^7

Z = observation vector

Data	Reports x items x levels
sondes	500x5x15
AMSU	6000x20
SM, ships, buoys	4000x5
aircraft	4000x3
Sat. winds	8000x2
TOTAL	2x10^5

$$\frac{N_X}{N_Z} \approx 45$$

- Cannot do X=f(Y), must do Y=f(X)
- Problem is underdetermined, always will be
- Need more information: prior knowledge, time evolution, nonlinear coupling

A Scalar Example with a single observation

The analysis equation is:

$$x^a = x^b + W(x^{obs} - x^b). \tag{3}$$

Subtract the truth from both sides:

$$x^a - x^t = x^b - x^t + W(x^{obs} - x^t - x^b + x^t)$$

Analysis error	$\epsilon^a = x^a - x^t$
Background error	$\epsilon^b = x^b - x^t$
Observation error	$\epsilon^{obs} = x^{obs} - x^t$

The analysis equation in terms of errors is:

$$\epsilon^a = \epsilon^b + W(\epsilon^{obs} - \epsilon^b) \tag{4}$$

A Scalar Example with a single observation

Take an ensemble average:

$$< \epsilon^a >=< \epsilon^b > +\mathrm{W}(< \epsilon^{\mathrm{obs}} > - < \epsilon^b >).$$

If $< \epsilon^b >=< \epsilon^{\mathrm{obs}} >= 0$, then $< \epsilon^a >= 0$.

Square (4) and take an ensemble average:

$$< (\epsilon^a)^2 >=< (\epsilon^b)^2 > +\mathrm{W}^2 < (\epsilon^{\mathrm{obs}} - \epsilon^b)^2 > +2\mathrm{W} < \epsilon^b(\epsilon^{\mathrm{obs}} - \epsilon^b) > .$$

Minimize $< (\epsilon^a)^2 >$ with respect to W assuming $< \epsilon^b \epsilon^{\mathrm{obs}} >= 0$:

$$d < (\epsilon^a)^2 > /d\mathrm{W} = 2\mathrm{W} < (\epsilon^{\mathrm{obs}})^2 + (\epsilon^b)^2 > -2 < (\epsilon^b)^2 >= 0$$

Let

$$(\sigma^{\mathrm{obs}})^2 =< (\epsilon^{\mathrm{obs}})^2 >, \quad (\sigma^b)^2 =< (\epsilon^b)^2, \quad (\sigma^a)^2 =< (\epsilon^a)^2 > .$$

so that

$$\boxed{\mathrm{W} = \frac{(\sigma^b)^2}{(\sigma^b)^2 + (\sigma^{\mathrm{obs}})^2}.} \tag{5}$$

A Scalar Example with a single observation

Note that $0 \leq \mathrm{W} \leq 1$. If $(\sigma^{\mathrm{obs}})^2 = 0$ and W=1. If $(\sigma^b)^2 = 0$ and W=0.
With this choice of weight,

$$(\sigma^a)^{-2} = (\sigma^b)^{-2} + (\sigma^{\mathrm{obs}})^{-2}. \tag{6}$$

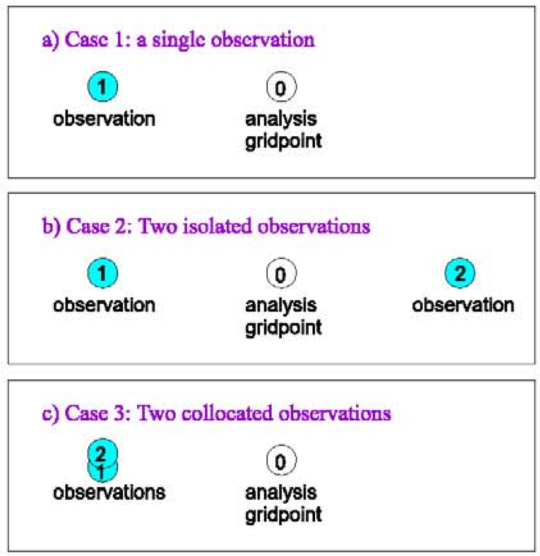
The analysis equation may be written as:

$$\mathrm{x}^a = \mathrm{x}^b + W(\mathrm{x}^{\mathrm{obs}} - \mathrm{x}^b), \quad \mathrm{W} = \frac{1}{1+\alpha}.$$

where $\alpha = (\sigma^{\mathrm{obs}})^2/(\sigma^b)^2$.

- If $(\sigma^{\mathrm{obs}})^2 << (\sigma^b)^2$, then $\alpha = 0$, W=1 and $\mathrm{x}^a = \mathrm{x}^{\mathrm{obs}}$, $\sigma^a = \sigma^{\mathrm{obs}}$.
- If $(\sigma^{\mathrm{obs}})^2 >> (\sigma^b)^2$, then $\alpha >> 1$, W=0 and $\mathrm{x}^a = \mathrm{x}^b$, $\sigma^a = \sigma^b$.
- If $(\sigma^{\mathrm{obs}})^2 = (\sigma^b)^2$, then W=1/2 and $\mathrm{x}^a = 0.5(\mathrm{x}^b + \mathrm{x}^{\mathrm{obs}})$, $(\sigma^a)^2 = (\sigma^b)^2/2 = (\sigma^{\mathrm{obs}})^2/2$.

Two observations on a 1-D grid



Now consider the case of an analysis grid point influenced by two observations. Both obs are of the same type and so have the same error variance of $(\sigma^r)^2$. Also, the background error variance at both obs stations is assumed the same:

$$< (\epsilon_1^b)^2 > = < (\epsilon_2^b)^2 > = (\sigma^b)^2,$$

The obs error is assumed horizontally uncorrelated, i.e.

$$< (\epsilon_1^r)(\epsilon_2^r) >= 0.$$

The obs and background errors are uncorrelated:

$$< (\epsilon_c^b)(\epsilon_d^r) >= 0$$

where $c, d \in \{0, 1, 2\}$. The analysis equation is

$$\mathsf{x}_0^a = \mathsf{x}_0^b + w_1(\mathsf{x}_1^r - \mathsf{x}_1^b) + w_2(\mathsf{x}_2^r - \mathsf{x}_2^b). \quad (10)$$

To determine the weights applied to each observation according to a minimum variance principle, first form the analysis error variance from (10) and apply the expectation operator.

$$< (\epsilon_0^a)^2 >= (\sigma^b)^2 + (w_1^2 + w_2^2)[(\sigma^r)^2 + (\sigma^b)^2]$$
$$-2w_1\rho_{10}(\sigma^b)^2 - 2w_2\rho_{20}(\sigma^b)^2 + 2w_1w_2\rho_{12}(\sigma^b)^2 \quad (11)$$

where we have defined

$$< \epsilon_0^b\epsilon_1^b >= \rho_{10}(\sigma^b)^2, \quad < \epsilon_0^b\epsilon_2^b >= \rho_{20}(\sigma^b)^2, \quad < \epsilon_1^b\epsilon_2^b >= \rho_{12}(\sigma^b)^2$$

and where all terms involving correlations of observation and background errors have been dropped.

Now minimize $< (\epsilon_0^a)^2 >$ w.r.t. w_1 and w_2:

$$\begin{aligned} w_1(1+\alpha) + w_2\rho_{12} &= \rho_{10} \\ w_1\rho_{12} + w_2(1+\alpha) &= \rho_{20} \end{aligned}$$

where

$$\alpha = (\sigma^r)^2/(\sigma^b)^2.$$

Solving for w_1 and w_2 yields:

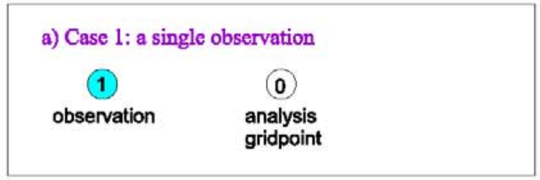
$$w_1 = \frac{\rho_{10}(1+\alpha) - \rho_{12}\rho_{20}}{(1+\alpha)^2 - \rho_{12}^2} \tag{12}$$

$$w_2 = \frac{\rho_{20}(1+\alpha) - \rho_{12}\rho_{10}}{(1+\alpha)^2 - \rho_{12}^2}. \tag{13}$$

With these optimal weights, (11) becomes

$$< (\epsilon_0^a)^2 > = (\sigma^b)^2 \left\{ 1 - \frac{(1+\alpha)(\rho_{10}^2 + \rho_{20}^2) - 2\rho_{10}\rho_{20}\rho_{12}}{(1+\alpha)^2 - \rho_{12}^2} \right\}. \tag{14}$$

Case One: A single observation



What is the analysis at gridpoint 0 if only the observation at gridpoint 1 is available? In this case, the analysis equation, (10) reduces to

$$x_0^a = x_0^b + w_1(x_1^r - x_1^b),$$

the weight, (12) becomes

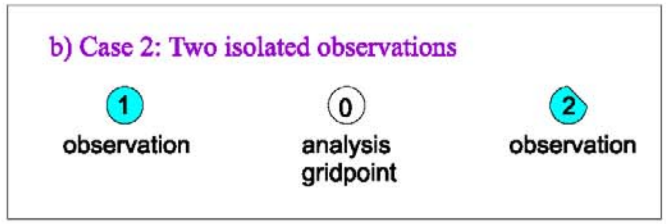
$$w_1 = \frac{\rho_{10}}{1+\alpha} \tag{15}$$

and the analysis error variance in (14) becomes

$$< (\epsilon_0^a)^2 >= (\sigma^b)^2 \left\{ 1 - \frac{\rho_{10}^2}{1+\alpha} \right\}. \tag{16}$$

The weight given to an observation depends on the distance between it and the analysis grid point and the way the background error correlation varies with distance.

Case Two: Two isolated observations



Now the obs are located on either side of the analysis grid point. Assume that $\rho_{12} \approx 0$, and that $\rho_{10} = \rho_{20} = \rho$. In this case, (12) and (13) reduce to

$$w_1 = w_2 \approx \frac{\rho}{1+\alpha}, \tag{17}$$

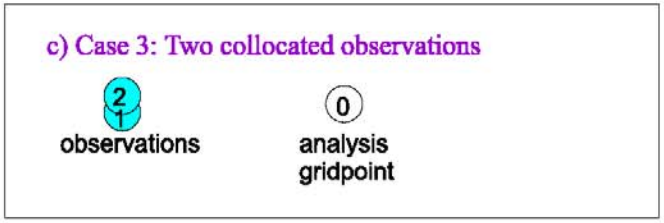
and (14) reduces to

$$< (\epsilon_0^a)^2 >= (\sigma^b)^2 \left\{ 1 - \frac{2\rho^2}{1+\alpha} \right\}. \tag{18}$$

Comparing (16) and (18) reveals that having 2 observations results in a lower analysis error than having only 1 observation.

Two obs are better than one.

Case Three: Two collocated observations



What if, instead of being located on either side of the analysis gridpoint, the two observations are collocated? In this case, $\rho_{12} = 1$ and $\rho_{10} = \rho_{20} = \rho$ so that

$$w_1 = w_2 = \frac{\rho}{2 + \alpha}, \tag{19}$$

and

$$< (\epsilon_0^a)^2 >= (\sigma^b)^2 \left\{ 1 - \frac{2\rho^2}{2 + \alpha} \right\}. \tag{20}$$

The weight given to the collocated observations is less than that for the isolated observations. The analysis error is also smaller when there are isolated observations. Why? More information is obtained for *independent* observations. Two collocated observations do not provide independent information so they each contribute less than if they had been independent.

Two isolated obs are better than two collocated ones.

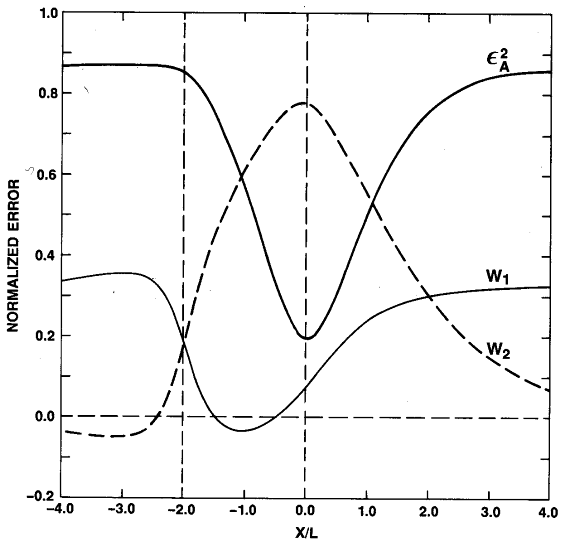
Case Four: Two observations on a 1-D network

If you have two observations, where is the best place to put them? To find out, hold one the the observations fixed at x/L=-2. The analysis gridpoint is at x/L=0. The second obs's location will vary with x from -∞ to +∞. Daley's Fig. 4.7 plots the normalized analysis error variance, $< (\epsilon_0^a)^2 > /(\sigma^b)^2$ for $\alpha_1 = \alpha_2 = 0.25$ and $\rho_{10} = 0.406$. To determine ρ_{20}, and ρ_{12}, the covariance model was used:

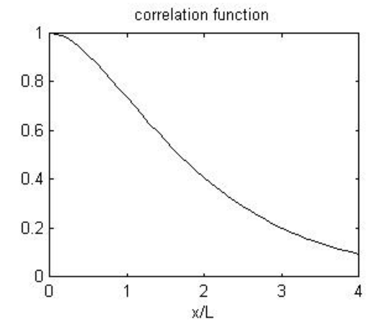
$$\rho^b(\Delta x) = \left(1 + \frac{|\Delta x|}{L}\right) \exp\left(-\frac{|\Delta x|}{L}\right).$$

$$w_1 = \frac{\rho_{10}(1+\alpha) - \rho_{12}\rho_{20}}{(1+\alpha)^2 - \rho_{12}^2}, \qquad w_2 = \frac{\rho_{20}(1+\alpha) - \rho_{12}\rho_{10}}{(1+\alpha)^2 - \rho_{12}^2}.$$

Obs 1 analysis



Daley (1991)



Optimal Interpolation

Consider a model state vector,

$$\mathbf{x} = (\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n)^\top.$$

The background, $\mathbf{x}^b$, and analysis, $\mathbf{x}^a$ are both on this grid and are n-vectors. For an obs network of m measurements, define the obs vector as

$$\mathbf{z} = (\mathsf{z}_1, \mathsf{z}_2, \ldots, \mathsf{z}_m)^\top.$$

Since the obs are not necessarily at analysis grid points, we need a spatial interpolation from the observation locations to the model grid. Let's introduce H, the *forward model*, which maps the model state to the obs variables and locations. H is nonlinear in general. Our analysis equation is then

n x m

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{z} - H(\mathbf{x}^b)). \tag{21}$$

n x 1 n x 1 m x 1 m x 1

$\mathbf{W}$ was renamed $\mathbf{K}$.

The stochastic measurement equation is:

$$\mathbf{z} = \mathbf{z}^t + \boldsymbol{\nu} \tag{22}$$

$\boldsymbol{\nu}$ is the measurement error. $\mathbf{z}^t$ is the "true" atmospheric quantity being sensed.

To allow an imperfect forward model operator we write

$$\begin{aligned} \mathbf{z} &= H(\mathbf{x}^t) + \boxed{\mathbf{z}^t - H(\mathbf{x}^t) + \boldsymbol{\nu}} \\ &= H(\mathbf{x}^t) + \mathbf{v} \end{aligned} \tag{23}$$

where

representativeness　　measurement

$$\mathbf{v} = \underbrace{[\mathbf{z}^t - H(\mathbf{x}^t)]}_{\text{representativeness}} + \boldsymbol{\nu}. \tag{24}$$

The term is square brackets is called the *representativeness* error and reflects the fact that our forward model, H, is not perfect. Recall that H includes a mapping of model variables to observed variables and a spatial interpolation from the model grid (or state) to the observed locations. The sum of the measurement and representativeness errors form the *observation* error, $\mathbf{v}$. The observation error bias is given by

$$< \mathbf{v} >= \bar{\mathbf{v}}$$

and the observation error covariance matrix is:

$$\mathbf{R} =< (\mathbf{v} - \bar{\mathbf{v}})(\mathbf{v} - \bar{\mathbf{v}})^{\top} > .$$

Our errors can then be defined as before, with a new addition:

$$\begin{array}{lrcl} n \times 1 & \mathbf{e}^a & = & \mathbf{x}^a - \mathbf{x}^t \\ n \times 1 & \mathbf{e}^b & = & \mathbf{x}^b - \mathbf{x}^t \\ m \times 1 & \mathbf{v} & = & \mathbf{z} - H(\mathbf{x}^t). \end{array}$$

The analysis equation in terms of errors is then

$$\begin{aligned} \mathbf{e}^a &= \mathbf{e}^b + \mathbf{K}[\mathbf{z} - H(\mathbf{x}^t) + H(\mathbf{x}^t) - H(\mathbf{x}^b)] \\ &= \mathbf{e}^b + \mathbf{K}[\mathbf{v} + H(\mathbf{x}^b + \mathbf{x}^t - \mathbf{x}^b) - H(\mathbf{x}^b)] \\ &\approx \mathbf{e}^b + \mathbf{K}[\mathbf{v} + H(\mathbf{x}^b) + \mathbf{H}(\mathbf{x}^t - \mathbf{x}^b) - H(\mathbf{x}^b)] \\ &= \mathbf{e}^b + \mathbf{K}[\mathbf{v} - \mathbf{H}(\mathbf{e}^b)]. \end{aligned} \tag{25}$$

The *Tangent Linear Forward Model* operator is defined as

$$\mathbf{H} = \left.\frac{dH}{d\mathbf{x}}\right|_{\mathbf{x}^b}. \tag{26}$$

$\mathbf{H}$ is the derivative of the forward model operator with respect to the model state vector and evaluated at the model background state. Thus we have performed a linearization of the nonlinear observation operator around the background state, implicitly assuming that the truth is not too far from the background.

To form the analysis error covariance, multiply (25) by the transpose of itself and apply the expectation operator:

$$\mathbf{P}^a = \mathbf{P}^b + \mathbf{K}(\mathbf{R} + \mathbf{H}\mathbf{P}^b\mathbf{H}^\top)\mathbf{K}^\top - \mathbf{K}\mathbf{H}\mathbf{P}^b - \mathbf{P}^b\mathbf{H}^\top\mathbf{K}^\top. \tag{27}$$

We now minimize the analysis error variance or trace of $\mathbf{P}^a$ with respect to the weight $\mathbf{K}$. Thus

$$0 = \frac{d\mathrm{Tr}(\mathbf{P}^a)}{d\mathbf{K}} = 2\mathbf{K}(\mathbf{R} + \mathbf{H}\mathbf{P}^b\mathbf{H}^\top) - 2\mathbf{P}^b\mathbf{H}^\top \tag{28}$$

or, on solving for $\mathbf{K}$:

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^\top(\mathbf{H}\mathbf{P}^b\mathbf{H}^\top + \mathbf{R})^{-1}. \tag{29}$$

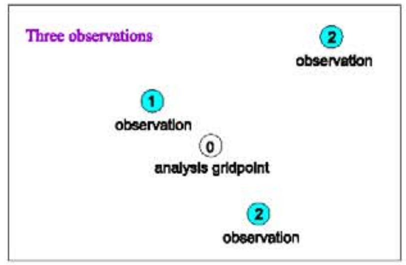
This is the choice of weight that gives the minimum variance of the estimate. Substituting (29) into (27) reveals the analysis error covariance for this optimal weight:

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b. \tag{30}$$

Summary of the OI algorithm:

$$\begin{aligned}
\mathbf{x}^a &= \mathbf{x}^b + \mathbf{K}[\mathbf{z} - H(\mathbf{x}^b)] \\
\mathbf{K} &= \mathbf{P}^b\mathbf{H}^\top(\mathbf{H}\mathbf{P}^b\mathbf{H}^\top + \mathbf{R})^{-1} \\
\mathbf{P}^a &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b
\end{aligned}$$

Example: 3 observations



$$\mathsf{x}_0^a = \mathsf{x}_0^b + \begin{bmatrix} K_1 & K_2 & K_3 \end{bmatrix} \begin{bmatrix} < z_1 - H(\mathbf{x}^b) > \\ < z_2 - H(\mathbf{x}^b) > \\ < z_3 - H(\mathbf{x}^b) > \end{bmatrix}. \quad (31)$$

$$\boxed{\mathbf{K}(\mathbf{H}\mathbf{P}^b\mathbf{H}^\top + \mathbf{R}) = \mathbf{P}^b\mathbf{H}^\top}$$

If the same instrument is used for each measurement and the obs error is uncorrelated in space, $\mathbf{R} = (\sigma^r)^2\mathbf{I}$.

$$\mathbf{H}\mathbf{P}^b\mathbf{H}^\top = \mathbf{H} < \mathbf{e}^b(\mathbf{e}^b)^\top > \mathbf{H}^\top = < (\mathbf{H}\mathbf{e}^b)(\mathbf{H}\mathbf{e}^b)^\top >$$

$$\mathbf{P}^b\mathbf{H}^\top = < \mathbf{e}^b(\mathbf{e}^b)^\top > \mathbf{H}^\top = < (\mathbf{e}^b)(\mathbf{H}\mathbf{e}^b)^\top >$$

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix}^\top \left\{ \begin{bmatrix} < \epsilon_1^b\epsilon_1^b > & < \epsilon_1^b\epsilon_2^b > & < \epsilon_1^b\epsilon_3^b > \\ < \epsilon_2^b\epsilon_1^b > & < \epsilon_2^b\epsilon_2^b > & < \epsilon_2^b\epsilon_3^b > \\ < \epsilon_3^b\epsilon_1^b > & < \epsilon_3^b\epsilon_2^b > & < \epsilon_3^b\epsilon_3^b > \end{bmatrix} + \mathbf{I}(\sigma^r)^2 \right\} = \begin{bmatrix} < \epsilon_1^b\epsilon_0^b > \\ < \epsilon_2^b\epsilon_0^b > \\ < \epsilon_3^b\epsilon_0^b > \end{bmatrix}$$

Optimal Interpolation in practice

$$
\begin{aligned}
\mathbf{x}^a &= \mathbf{x}^b + \mathbf{K}[\mathbf{z} - H(\mathbf{x}^b)] \\
\mathbf{K} &= \mathbf{P}^b\mathbf{H}^\top(\mathbf{H}\mathbf{P}^b\mathbf{H}^\top + \mathbf{R})^{-1} \\
\mathbf{P}^a &= \mathbf{K}\mathbf{R}\mathbf{K}^\top + (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b(\mathbf{I} - \mathbf{K}\mathbf{H})^\top
\end{aligned}
$$

OI for NWP is run at every synoptic hour (00, 06, 12, 18Z), each day. Observations are binned into 6 hr intervals centered on the analysis time. The same $\mathbf{P}^b$ is used every analysis. Thus we are assuming stationary statistics.

To solve for the weights we need to invert an $m \times m$ matrix. For m=10^5, this matrix inversion is too expensive.

1. Assume the generalized interpolation, H, is linear. Then H(**x**) = **H x**.
2. $\mathbf{P}^b$ is continuous. $\mathbf{H}\mathbf{P}^b\mathbf{H}^\top$ can then be evaluated at observation sites as an $m \times m$ matrix without ever needing to know $\mathbf{P}^b$ on the model grid, which is $n \times n$. Linear dynamical constraints can then be applied through modelling of $\mathbf{P}^b$.
3. Data selection is used so that the analysis equation is solved n times. Each equation is then solved for a scalar x^a. By limiting the number of observations that influence a given analysis point to p ($<$100), we can further reduce the size of $\mathbf{H}\mathbf{P}^b\mathbf{H}^\top$ to $p \times p$. Thus the inversion of an $m \times m$ matrix has been replaced by n inversions of $p \times p$ matrices.

Table 5.1 Characteristics of operational numerical analysis schemes.

Organization or country	Present operational analysis methods	Analysis area	Analysis forecast cycle	Plans
Australia	. Successive correction method . Variational blending techniques	. S Hemisph . Regional	12 hours 6 hours	
Canada	. Multivariate 3-dimensional statistical interpolation	. N Hemisph . Regional	6 hours (3 hours for the surface)	
France	. Successive correction method, windfield and massfield balance through first guess fields . Multivariate 3-dimensional statistical interpolation	. N Hemisph . Regional	6 hours	
F.R.G.	. Successive correction method. Upper-air analyses are built up, level by level, from the surface . Variational height/wind adjustment	. N Hemisph	12 hours (6 hours for the surface) Climatology only as preliminary fields	Multivariate statistical interpolation is being developed
Japan	. Successive correction method. . Height field analyses are corrected by wind analyses	. N Hemisph . Regional	12 hours	Multivariate statistical interpolation is tested
Sweden	. Uni-variate 3-dimensional statistical interpolation . Variational height/wind adjustment	. N Hemisph . Regional	12 hours 3 hours	A multivariate scheme is being tested
U.K.	. Hemispheric orthogonal polynomial method . Uni-variate statistical interpolation (repeated insertion of data)	. Global	6 hours	Multivariate schemes considered
U.S.A.	. Spectral 3-dimensional analysis . Multivariate 3-dimensional statistical interpolation	. Global . Global	6 hours	
U.S.S.R.	. 2-dimensional statistical interpolation	. N Hemisph	12 hours	
E.C.M.W.F.	. Multivariate 3-dimensional statistical interpolation	. Global	6 hours	

Gustafsson (1981)

OI was the standard assimilation method at weather centres from the early 1970’s to the early 1990’s.

Canada was the first to implement a multivariate OI scheme.

Filtering Properties

The OI analysis equation is:

$$\mathbf{x}^a = \mathbf{x}^b + \overbrace{\mathbf{P}\mathbf{H}^\top(\mathbf{H}\mathbf{P}\mathbf{H}^\top + \mathbf{R})^{-1}}^{\mathbf{K}}[\mathbf{z} - \mathbf{H}(\mathbf{x}^b)] \qquad (32)$$

For simplicity, $\mathbf{P} = \mathbf{P}^b$. Assume $\mathbf{H} = \mathbf{I}$. Then define $\mathbf{y} = [\mathbf{z} - \mathbf{H}(\mathbf{x}^b)]$. Then (32) can be written as:

$$\mathbf{x}^a - \mathbf{x}^b = \mathbf{d} = \mathbf{P}(\mathbf{P} + \mathbf{R})^{-1}\mathbf{y}$$

or

$$\boxed{\mathbf{d} = \mathbf{A}\mathbf{y}}$$

where

$$\mathbf{A} = \mathbf{P}(\mathbf{P} + \mathbf{R})^{-1} = (\mathbf{I} + \mathbf{R}\mathbf{P}^{-1})^{-1}.$$

To simplify $\mathbf{A}$ further let $\mathbf{P} = (\sigma^b)^2\mathbf{C}$. Also assume that all observations are from the same type of instrument so $\mathbf{R} = (\sigma^r)^2\mathbf{I}$. Finally, define the eigenvalues and eigenvectors of $\mathbf{C}$ as λ and $\mathbf{e}$, i.e.,

$$\mathbf{C}\mathbf{e} = \lambda\mathbf{e} \qquad \mathbf{C}^{-1}\mathbf{e} = \frac{1}{\lambda}\mathbf{e}.$$

Then,

$$\mathbf{P}^{-1}\mathbf{e} = (\sigma^b)^{-2}\mathbf{C}^{-1}\mathbf{e} = \frac{1}{\lambda(\sigma^b)^2}\mathbf{e}$$

Finally, we see that,

$$\mathbf{Ae} = (\mathbf{I} + \mathbf{RP}^{-1})^{-1}\mathbf{e} = \frac{1}{1 + \frac{\alpha}{\lambda}}\mathbf{e}, \qquad \alpha = \frac{(\sigma^r)^2}{(\sigma^b)^2}.$$

If the observation increment can be written as a superposition of eigenvectors of $\mathbf{C}$, i.e.,

$$y = \sum_{i=1}^{N} c_i \mathbf{e}_i$$

then

$$\mathbf{d} = \mathbf{Ay} = \sum_{i=1}^{N} c_i \mathbf{Ae}_i = \sum_{i=1}^{N} c_i \left(\frac{1}{1 + \frac{\alpha}{\lambda_i}} \right) \mathbf{e}_i$$

Large eigenvalues (large scales): If $\lambda_i \gg \alpha$ then $1/(1 + \alpha/\lambda_i) \to 1$

Small eigenvalues (small scales): If $\lambda_i \ll \alpha$ then $1/(1 + \alpha/\lambda_i) \to 0$.

The spectral structure of the correlation matrix for background errors determines the filtering properties of the analysis.

If the background error correlation function has most energy at large (small) scales, the OI will act as a low (high)-pass filter.

Estimation theory

The general problem of data assimilation is this: given a set of observations and a model of some physical parameters, what does knowledge of the observations tell us about the model state?

a posteriori p.d.f. →

$$p_{x|z}(\mathbf{x}|\mathbf{z}) = \frac{p_{xz}(\mathbf{x},\mathbf{z})}{p_z(\mathbf{z})} = \frac{p_{z|x}(\mathbf{z}|\mathbf{x})p_x(\mathbf{x})}{p_z(\mathbf{z})}. \quad (33)$$

where $p_z(\mathbf{z}) \neq 0$. How do we choose an *estimator*, $\hat{\mathbf{x}}(\mathbf{z})$, based on $p_{x|z}(\mathbf{x}|\mathbf{z})$?

Let's try to minimize the risk $\mathcal{J}$ or expected cost function, J:

$$\begin{aligned}\mathcal{J}(\hat{\mathbf{x}}) &= E(J(\tilde{\mathbf{x}})) = \int_{-\infty}^{\infty} J(\tilde{\mathbf{x}})p_x(\mathbf{x})d\mathbf{x} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J(\tilde{\mathbf{x}})p_{xz}(\mathbf{x},\mathbf{z})d\mathbf{z}d\mathbf{x} \quad (34)\end{aligned}$$

where

$$\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}.$$

The quadratic cost function is

$$J(\tilde{\mathbf{x}}) = |\tilde{\mathbf{x}}|_S^2 = \tilde{\mathbf{x}}^\mathsf{T}\mathbf{S}\tilde{\mathbf{x}} \tag{35}$$

where $\mathbf{S}$ is a non-negative definite, symmetric matrix. The Minimum Variance estimator yields the conditional mean:

$$\boxed{\hat{\mathbf{x}}_{\mathsf{MV}} = E(\mathbf{x}|\mathbf{z}).} \tag{36}$$

The uniform cost function is

$$J(\tilde{\mathbf{x}}) = \begin{cases} 0 & |\tilde{\mathbf{x}}| < \epsilon \\ \frac{1}{2\epsilon} & |\tilde{\mathbf{x}}| \geq \epsilon \end{cases} \tag{37}$$

MAP estimation provides the maximum or "mode" of the *a posteriori* p.d.f.

$$\boxed{\frac{\partial}{\partial \mathbf{x}} p_{x|z}(\mathbf{x}|\mathbf{z})\Big|_{\mathbf{x}=\hat{\mathbf{x}}_{\mathsf{MAP}}} = 0.} \tag{38}$$

The absolute error cost function is

$$J(\tilde{\mathbf{x}}) = |\tilde{\mathbf{x}}| = |\mathbf{x} - \hat{\mathbf{x}}|. \tag{39}$$

The estimator with the minimum absolute error is the "median" of the *a posteriori* p.d.f.

Example: Estimation of a constant vector

Consider the measurement equation,

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}. \tag{40}$$

Assume $\mathbf{x}$ is N($\boldsymbol{\mu}$,$\mathbf{P}$), $\mathbf{v}$ is N($\mathbf{0}$,$\mathbf{R}$) and $\mathbf{x}$ and $\mathbf{v}$ are independent.

For a given $\mathbf{z}$, we want to know

$$p_{x|z}(\mathbf{x}|\mathbf{z}) = \frac{p_{zx}(\mathbf{z},\mathbf{x})}{p_z(\mathbf{z})} = \frac{p_{z|x}(\mathbf{z}|\mathbf{x})p_x(\mathbf{x})}{p_z(\mathbf{z})}. \tag{41}$$

$$\begin{aligned} p_{z,x}(\mathbf{z},\mathbf{x}) \;&=\; \frac{1}{(2\pi)^{n/2}|\mathbf{P}|^{1/2}}\frac{1}{(2\pi)^{m/2}|\mathbf{R}|^{1/2}} \\ &\times\; \exp\left\{-\frac{1}{2}(\mathbf{z}-\mathbf{H}\mathbf{x})^\top\mathbf{R}^{-1}(\mathbf{z}-\mathbf{H}\mathbf{x}) - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{P}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}. \end{aligned}$$

$$p_z(\mathbf{z}) = \frac{1}{(2\pi)^{m/2}|\mathbf{H}\mathbf{P}\mathbf{H}^\top + \mathbf{R}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{z}-\mathbf{H}\boldsymbol{\mu})^\top(\mathbf{H}\mathbf{P}\mathbf{H}^\top + \mathbf{R})^{-1}(\mathbf{z}-\mathbf{H}\boldsymbol{\mu})\right\}$$

After lots of algebra, we can show that

$$p_{x|z}(\mathbf{x}|\mathbf{z}) = \frac{|\mathbf{H}\mathbf{P}\mathbf{H}^\top + \mathbf{R}|^{1/2}}{(2\pi)^{m/2}|\mathbf{P}|^{1/2}|\mathbf{R}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\hat{\mathbf{x}})^\top \mathbf{P}_x^{-1}(\mathbf{x}-\hat{\mathbf{x}})\right\} \tag{43}$$

where

$$\hat{\mathbf{x}} = \mathbf{P}_x(\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{z} + \mathbf{P}^{-1}\boldsymbol{\mu}) \tag{44}$$

and

$$\mathbf{P}_x^{-1} = \mathbf{P}^{-1} + \mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H}. \tag{45}$$

Minimum variance estimator: OI

For our example, the MV estimator is the the mean of (42):

$$\widehat{\mathbf{x}}_{\mathsf{MV}} = E(\mathbf{x}|\mathbf{z}) = (\mathbf{P}^{-1} + \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H})^{-1}(\mathbf{H}^\top \mathbf{R}^{-1}\mathbf{z} + \mathbf{P}^{-1}\boldsymbol{\mu}). \tag{45}$$

Using the Sherman-Morrison-Woodbury formula

$$(\mathbf{P}^{-1} + \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\top \mathbf{R}^{-1} = \mathbf{P}\mathbf{H}^\top(\mathbf{H}\mathbf{P}\mathbf{H}^\top + \mathbf{R})^{-1} \tag{46}$$

and adding and subtracting $\boldsymbol{\mu}$ from the right side gives

$$\begin{aligned} \widehat{\mathbf{x}}_{\mathsf{MV}} &= \boldsymbol{\mu} + \mathbf{K}(\mathbf{z} - \mathbf{H}\boldsymbol{\mu}) \\ &= \mathbf{x}^b + \mathbf{K}(\mathbf{z} - \mathbf{H}\mathbf{x}^b) \end{aligned} \tag{47}$$

$$\mathbf{K} = \mathbf{P}\mathbf{H}^\top(\mathbf{H}\mathbf{P}\mathbf{H}^\top + \mathbf{R})^{-1} \tag{48}$$

MAP estimator: 3D-Var

For MAP estimation, it is sufficient to maximimize the numerator of

$$p_{x|z}(\mathbf{x}|\mathbf{z}) = \frac{p_{xz}(\mathbf{x},\mathbf{z})}{p_z(\mathbf{z})} = \frac{p_{z|x}(\mathbf{z}|\mathbf{x})p_x(\mathbf{x})}{p_z(\mathbf{z})}$$

since the $p_z(\mathbf{z})$ is not a function of $\mathbf{x}$. In our example, assuming Gaussian statistics, the MAP estimator could also be obtained by minimizing:

$$J_{\text{MAP}}(\mathbf{x}) \;=\; (\mathbf{z}-\mathbf{H}\mathbf{x})^{\top}\mathbf{R}^{-1}(\mathbf{z}-\mathbf{H}\mathbf{x}) + (\mathbf{x}-\boldsymbol{\mu})^{\top}\mathbf{P}^{-1}(\mathbf{x}-\boldsymbol{\mu}). \tag{49}$$

This is the 3DVAR cost function. In exercise 3, the solution is

$$\begin{aligned} \hat{\mathbf{x}}_{\text{MV}} &= \boldsymbol{\mu} + \mathbf{K}(\mathbf{z}-\mathbf{H}\boldsymbol{\mu}). \\ \mathbf{K} &= \mathbf{P}\mathbf{H}^{\top}(\mathbf{H}\mathbf{P}\mathbf{H}^{\top}+\mathbf{R})^{-1} \end{aligned} \tag{50}$$

For linear observation operators, and Gaussian background and observation errors, OI and 3DVAR are equivalent (theoretically).

In practice, approximations are made (e.g. data selection in OI or doing only a few descent steps in 3DVAR). Thus, the details of the implementation will determine the performance of our DA system.

In OI, we could solve the matrix equation variationally to get PSAS:

$$\begin{aligned} \mathbf{y} &= (\mathbf{H}\mathbf{P}\mathbf{H}^{\top}+\mathbf{R})^{-1}(\mathbf{z}-\mathbf{H}\boldsymbol{\mu}) \\ \hat{\mathbf{x}}_{\text{MV}} &= \boldsymbol{\mu} + \mathbf{P}\mathbf{H}^{\top}\mathbf{y}. \end{aligned} \tag{51}$$

Data Selection

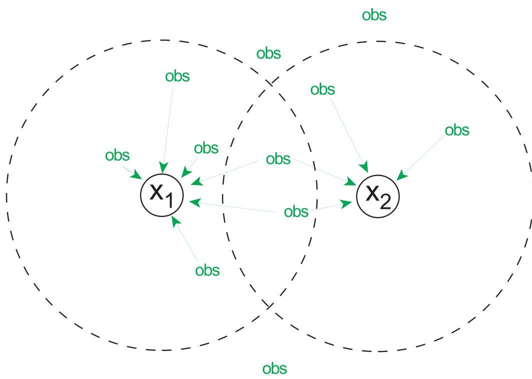


Figure 9. One OI data selection strategy is to assume that each analysis point is only sensitive to observations located in a small vicinity. Therefore, the observations used to perform the analysis at two neighbouring points x_1 or x_2 may be different, so that the analysis field will generally not be continuous in space. The cost of the analysis increases with the size of the selection domains.

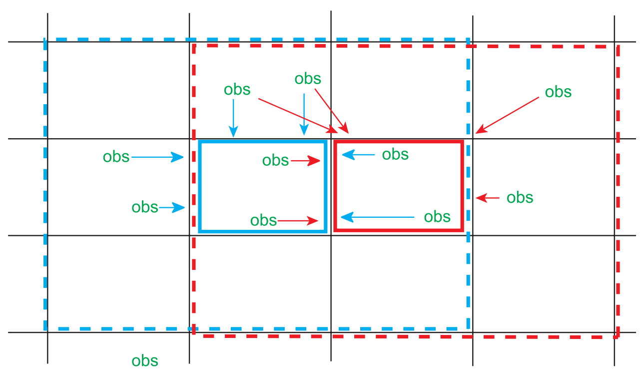
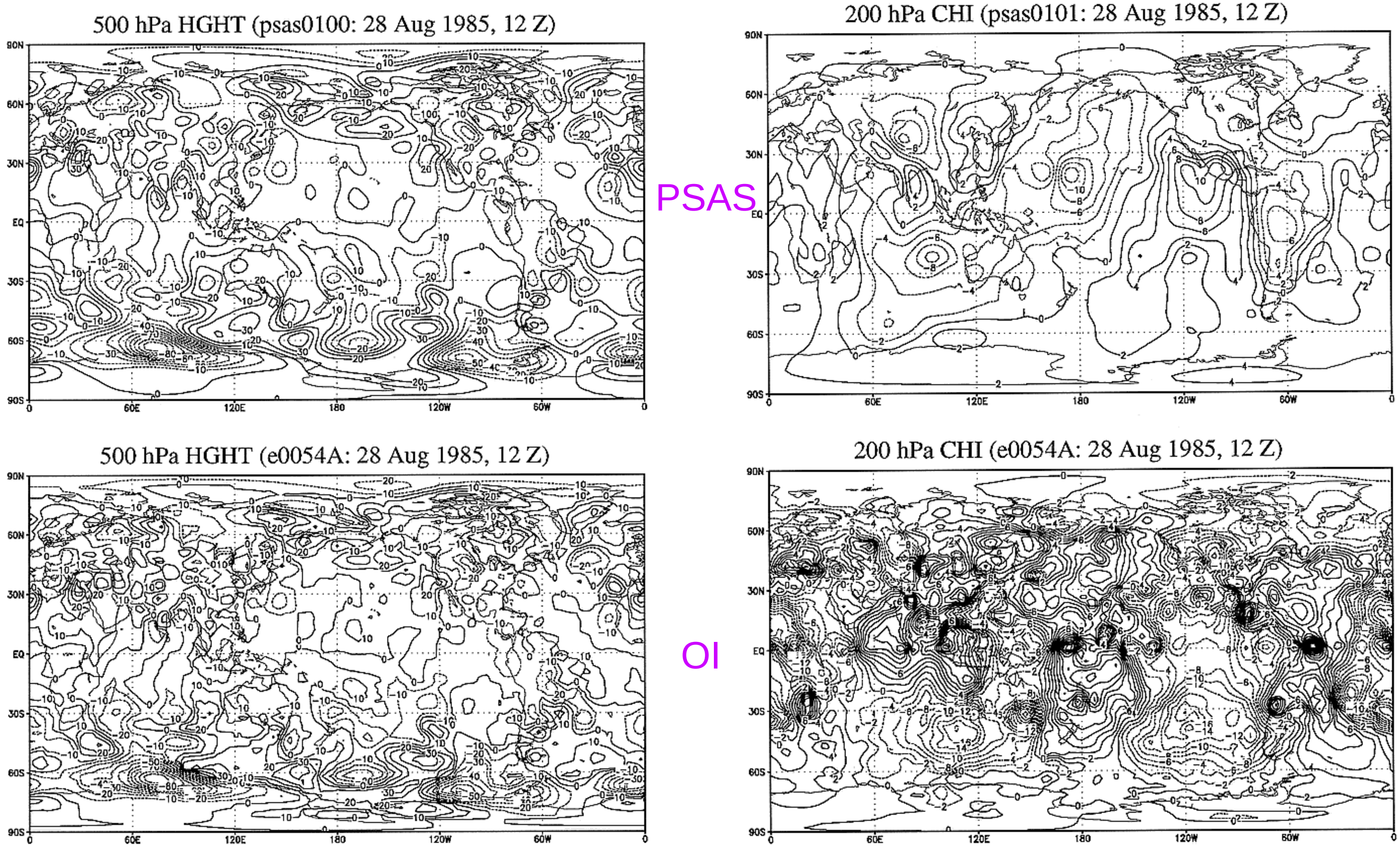


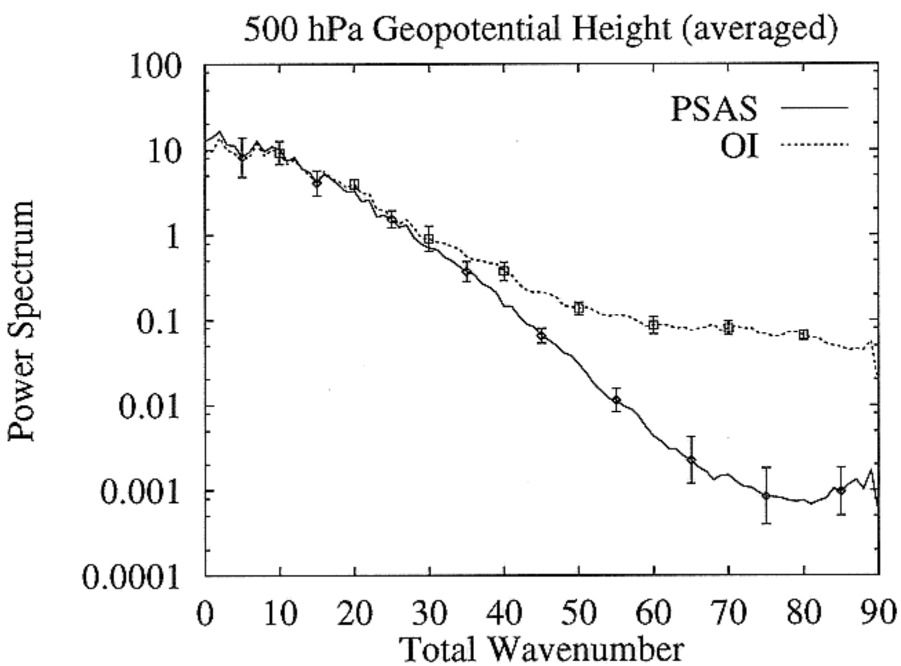
Figure 10. A slightly more sophisticated and more expensive OI data selection is to use, for all the points in an analysis box (black rectangle), all observations located in a bigger selection box (dashed rectangle), so that most of the observations selected in two neighbouring analysis boxes are identical.

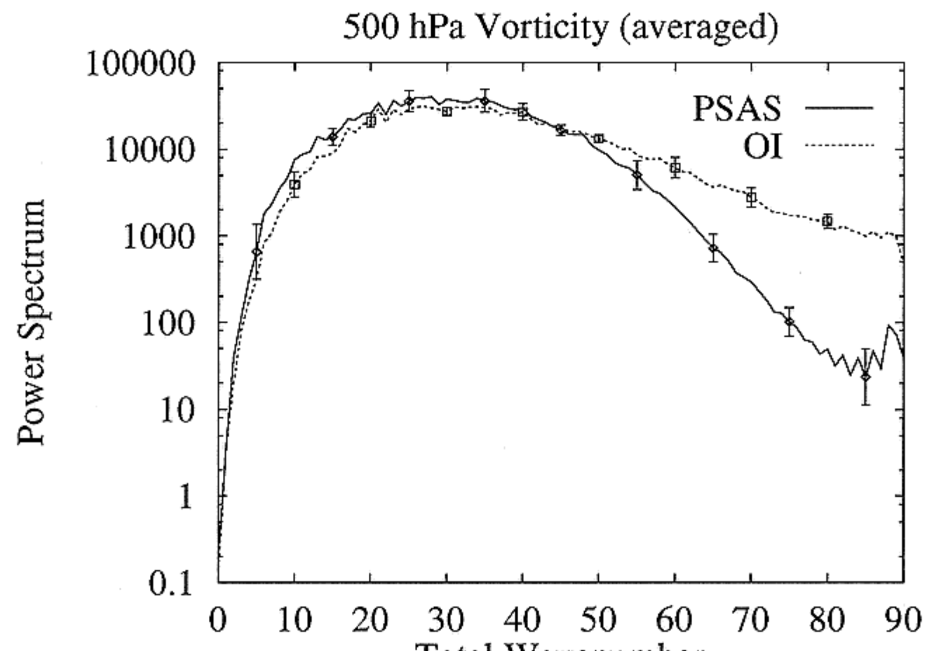
The effect of data selection

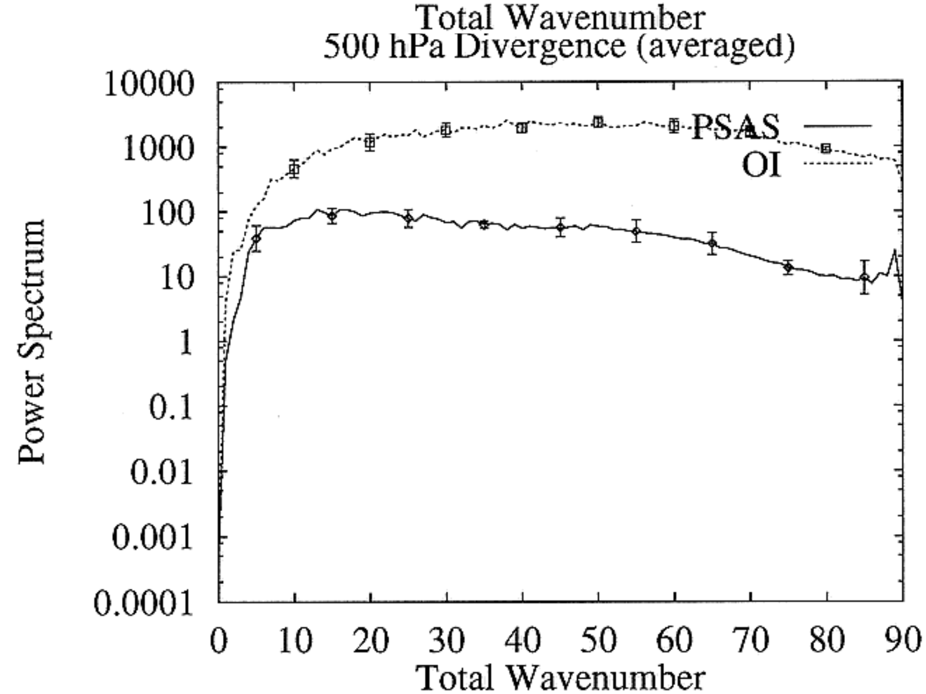


Cohn et al. (1996)

The effect of data selection







Cohn et al. (1996)

Advantages of 3D-var

$$J(\mathbf{x}) = (\mathbf{x}-\mathbf{x}^{\mathbf{b}})^T \mathbf{B}^{-\mathbf{1}}(\mathbf{x}-\mathbf{x}^{\mathbf{b}}) + (\mathbf{z}-H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{z}-H(\mathbf{x}))$$

1. Obs and model variables can be nonlinearly related.
   - H(X), H, H^T need to be calculated for each obs type
   - No separate inversion of data needed – can directly assimilate radiances
   - Flexible choice of model variables, e.g. spectral coefficents
2. No data selection is needed.

With covariances in spectral space, longer correlation lengths scales are permitted in the stratosphere

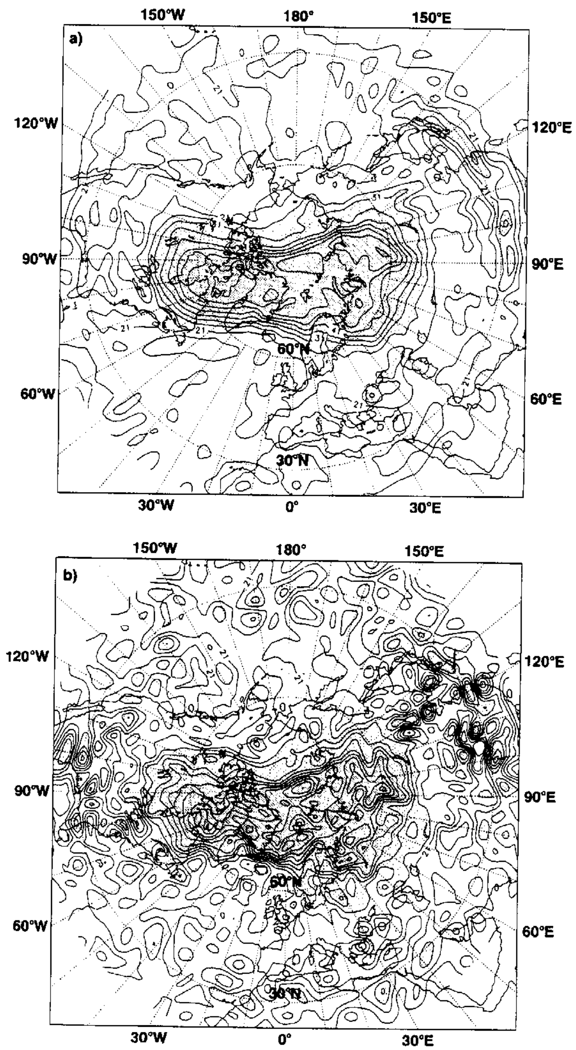


Figure 10. Potential vorticity on the 475 K isentropic surface in northern mid- to high latitudes at 12 UTC 29 January 1996: (a) 3D-Var; (b) OI.

Andersson et al. (1998)

With flexibility of choice of obs, can assimilate many new types of obs such as scatterometer

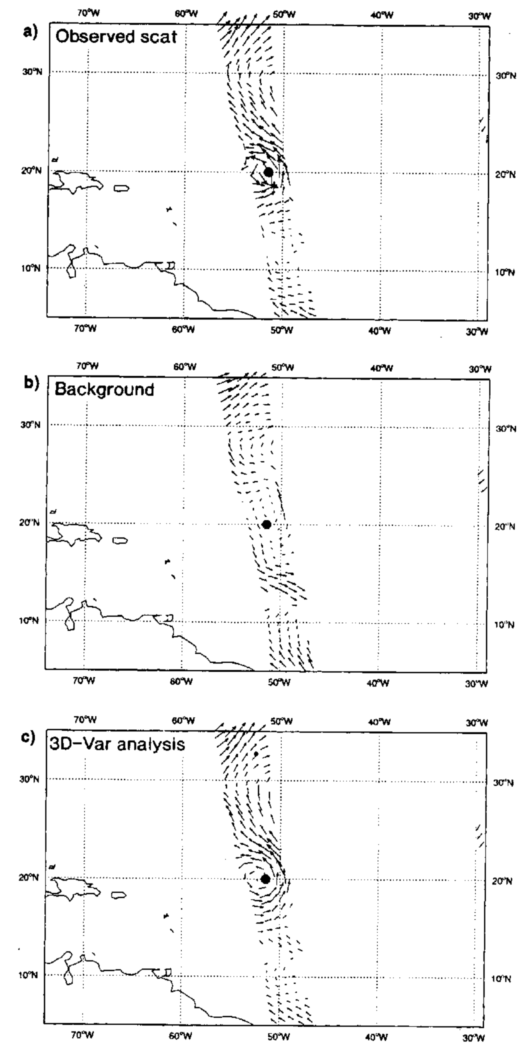
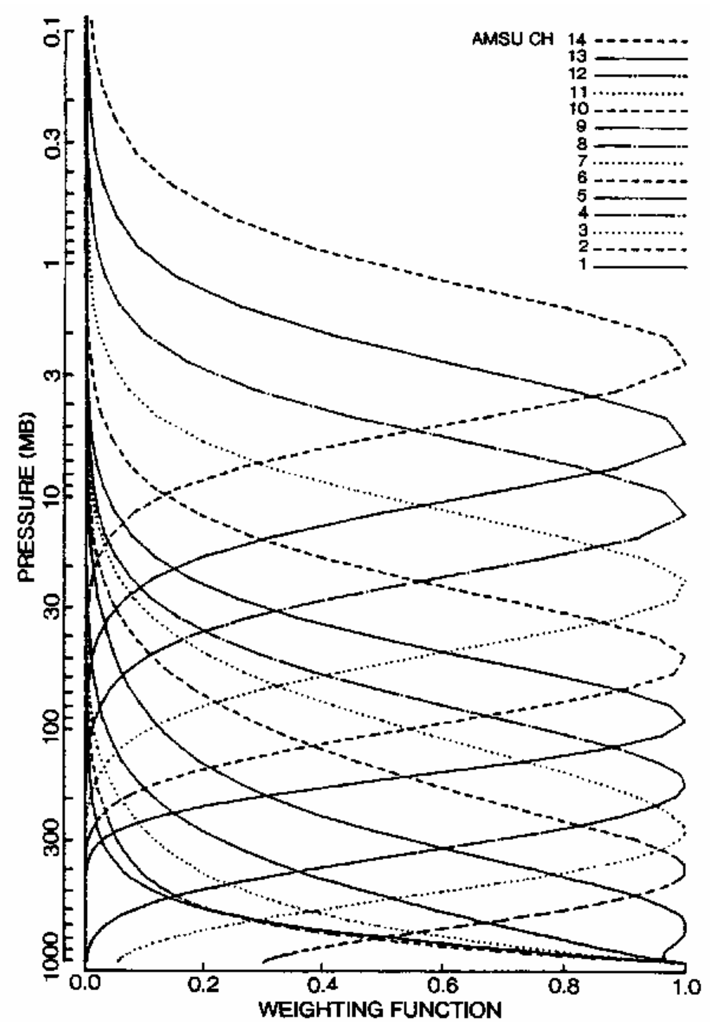


Figure 11. Winds beneath an orbit which passes over tropical cyclone *Karen* located at 20°N, 52°W (large dot) on 31 August 1995: (a) observed by scatterometer; (b) background (six-hour) forecast valid for the same time; (c) 3D-Var analysis. (b) and (c) are interpo;ated to the positions of the scatterometer observations.

Andersson et al. (1998)

To assimilate radiances directly, H includes an instrument-specific radiative transfer model

$$J(\mathbf{x}) = (\mathbf{x}-\mathbf{x}^{\mathbf{b}})^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^{\mathbf{b}}) + (\mathbf{z}-H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{z}-H(\mathbf{x}))$$



14
13
12
11
10
9
8
7
6
5

Normalized AMSU weighting functions

Impact of Direct Assimilation of Radiances

Anomaly = difference between forecast and climatolgy
Anomaly correlation – pattern correlation between forecast anomalies and verifying analyses

1974 – improved NESDIS VTPR Retrievals
1978 – TOVS retrievals

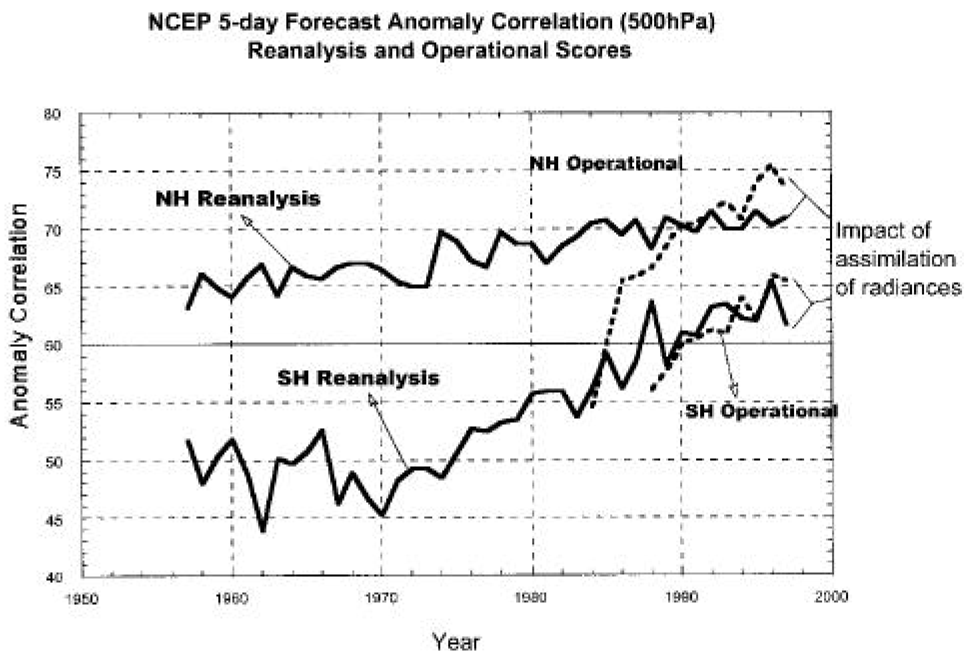


FIG. 6. Comparison of operational and reanalysis 5-day forecast anomaly correlations for the NH and the SW. The large improvement in operational forecasts observed in 1996–97 is due, to a large extent, to the direct assimilation of TOVS radiances (data courtesy of R. Kistler).

Kalnay et al. (1998)

Weather centers using 3D-var operationally

Center	Region	Started	Opera-tional	Ref.
NCEP	U.S.A.		June 1991	Parrish& Derber (1991)
ECMWF	Europe	1987	Jan. 1996	Courtier et al. (1997)
CMC	Canada	1993	June 1997	Gauthier et al. (1998)
Met Office	U.K.		Mar. 1999	Lorenc et al. (2000)
DAO	NASA		1997	Cohn et al. (1997)
NRL	US Navy		2000?	Daley& Barker (2001)

Summary

- Data assimilation combines information of observations and models and their errors to get a best estimate of atmospheric state (or other parameters)
- For Gaussian errors, 3D-var and OI are equivalent in theory, but different in practice
- 3D-var allows easy extension for nonlinearly related obs and model variables. Also allows more flexibility in choice of analysis variables.
- 3D-var does not require data selection so analyses are in better balance.
- Improvement of 3D-var over OI is not statistically significant for same obs. Systematic improvement of 3DVAR over OI in stratosphere and S. Hemisphere. Scores continue to improve as more obs types are added.

Covariance Modelling

1. Innovations method
2. NMC-method
3. Ensemble method

The role of the forecast error covariance matrix in analysis

model space (nx1) observation space (mx1)

$$\mathbf{x}_a = \mathbf{x}_b + K\,[\mathbf{z} - H(\mathbf{x}_b)]$$

(mxm)

$$\mathbf{K} = \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}$$

For a single observation, at gridpoint i

$$\mathbf{K} = \mathbf{B}_i /(\sigma_b{}^2+\sigma_r{}^2) = c\ \mathbf{B}_i$$

The forecast error covariance matrix determines the spatial influence of the observations.

Background error covariance matrix

$$\mathbf{P}^{\mathbf{b}} = \left\langle \left(\mathbf{x}^{\mathbf{b}} - \mathbf{x}^{\mathbf{t}}\right)\left(\mathbf{x}^{\mathbf{b}} - \mathbf{x}^{\mathbf{t}}\right)^{T} \right\rangle$$

- If x is 10^7, P^b is 10^7 x 10^7.
- With 10^5 obs, cannot estimate P^b.
- Need to model P^b.
- The fewer the parameters in the model, the easier to estimate them, but less likely the model is to be valid

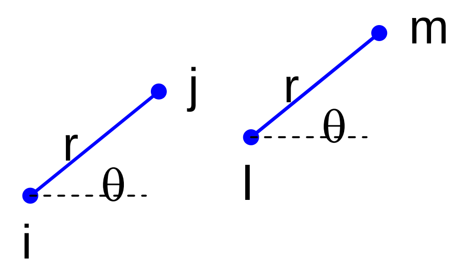
1. Innovations method

- Historically used for Optimal Interpolation (e.g. Hollingsworth and Lonnberg 1986, Lonnberg and Hollingsworth 1986, Mitchell et al. 1990)
- Typical assumptions:
  - separability of horizontal and vertical correlations

$$\mathbf{C}^b(x_i, y_i, z_i, x_j, y_j, z_j) = \mathbf{C}_H^b(x_i, y_i, x_j, y_j)\mathbf{C}_V^b(z_i, z_j)$$

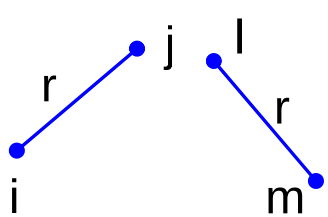
  - Homogeneity

$$\mathbf{C}_H^b(x_i, y_i, x_j, y_j) = \mathbf{C}_x^b(r_i, \theta_j)$$



  - Isotropy

$$\mathbf{C}_H^b(x_i, y_i, x_j, y_j) = \mathbf{C}_x^b(r_i)$$



Innovations method of computing B Matrix

If H is linear then $\mathbf{z} - H(\mathbf{x}^b) = \mathbf{v} - \mathbf{H}\mathbf{e}^b$.

Instrument+ representativeness (→ $\mathbf{v}$)

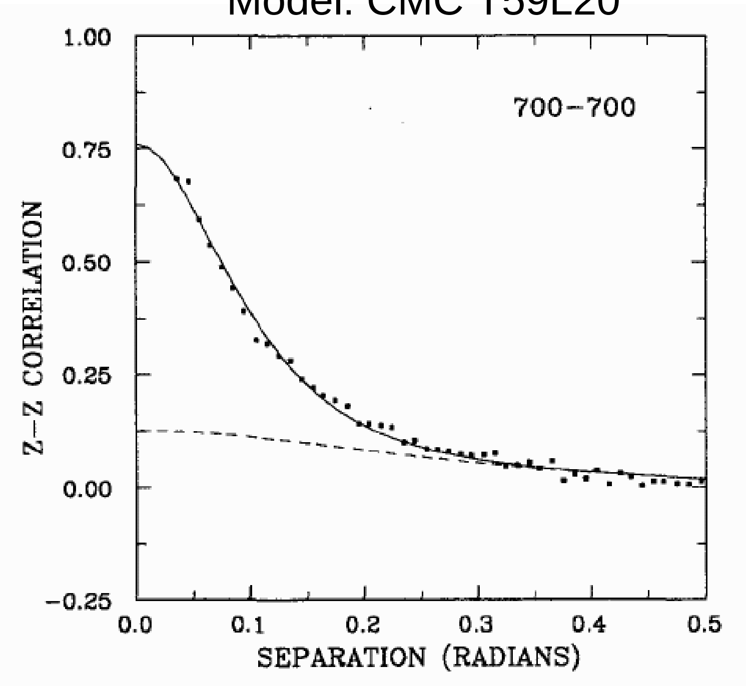
Background error (→ $\mathbf{e}^b$)

$$
\begin{aligned}
< (\mathbf{v} - \mathbf{H}\mathbf{e}^b)(\mathbf{v} - \mathbf{H}\mathbf{e}^b)^\top > &= < (\mathbf{v})(\mathbf{v})^\top > + \mathbf{H} < \mathbf{e}^b {\mathbf{e}^b}^\top > \mathbf{H}^\top \\
&- \mathbf{H} < \mathbf{e}^b(\mathbf{v})^\top > - < \mathbf{v}(\mathbf{e}^b)^\top > \mathbf{H}^\top
\end{aligned}
$$

Choose obs s.t. these terms =0

$$
\begin{aligned}
&= \mathbf{R} + \mathbf{H}\mathbf{P}^b\mathbf{H}^\top \\
&= (\sigma^r)^2\mathbf{I} + \mathbf{H}\mathbf{P}^b\mathbf{H}^\top.
\end{aligned}
$$

Dec. 15/87-Mar. 15/88 radiosonde data.
Model: CMC T59L20



Mitchell et al. (1990)

Assume homogeneous, isotropic correlation model. Choose a continuous function ρ(r) which has only a few parameters such as L, correlation length scale. Plot all innovations as a function of distance only and fit the function to the data.

Obs and Forecast error variances

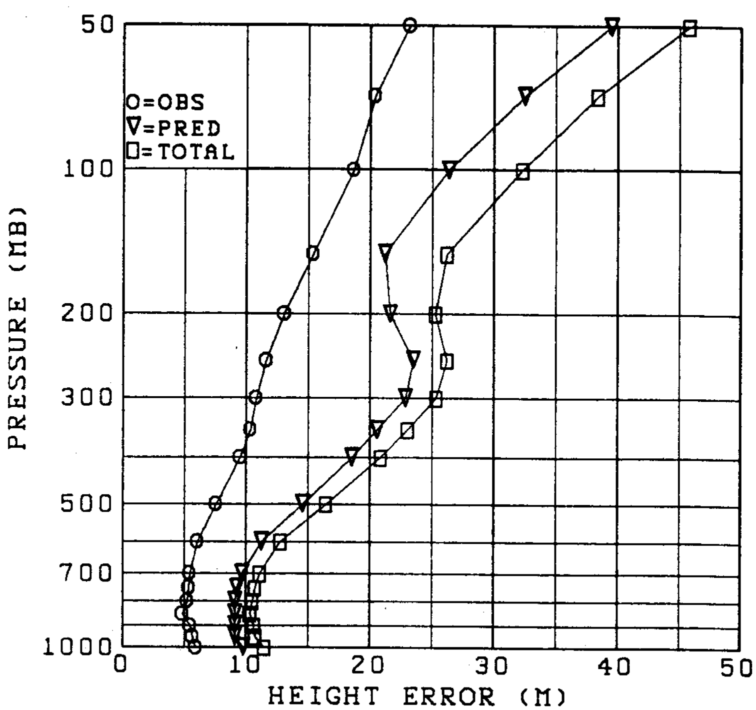


FIG. 8. Vertical profile of the observed height residual (m) (i.e., total perceived forecast error) denoted TOTAL, and the corresponding profiles of prediction and observation error.

Mitchell et al. (1990)

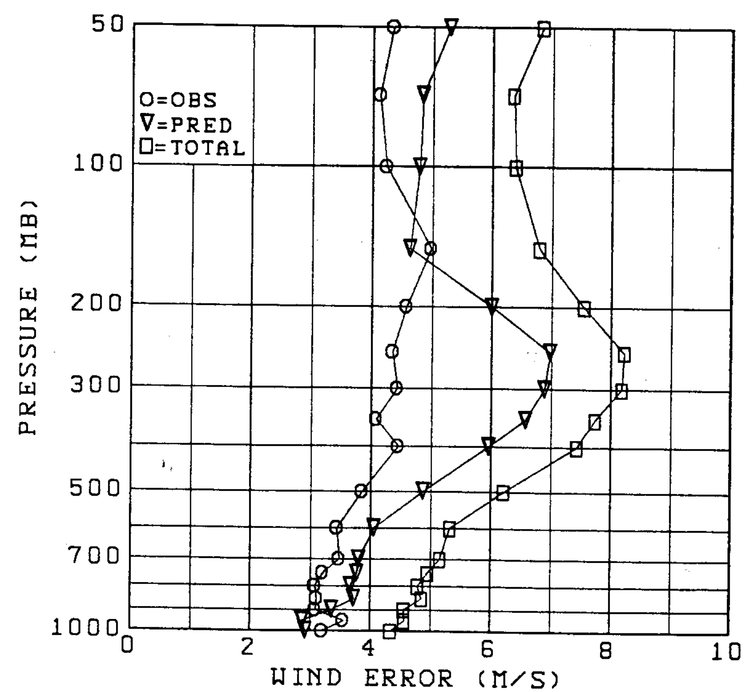


FIG. 5. Vertical profile of the observed wind residual (meters per second) (i.e., total perceived forecast error) denoted TOTAL, and the corresponding profiles of prediction and observation error.

Mitchell et al. (1990)

Vertical correlations of forecast error

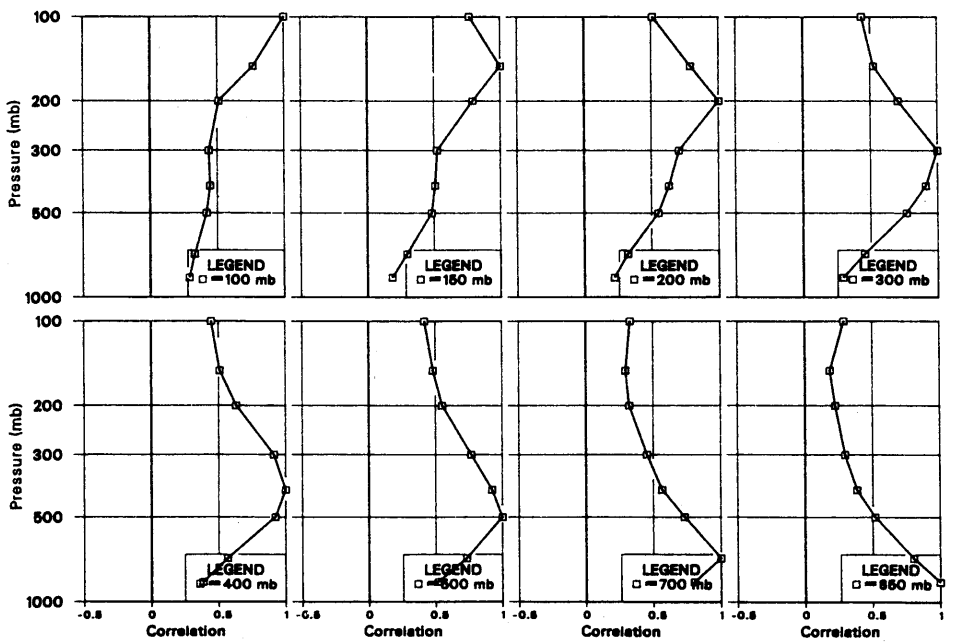


Fig. 10. Height prediction error vertical correlations for a selected set of standard levels, indicated in the legend of each frame of the plot.

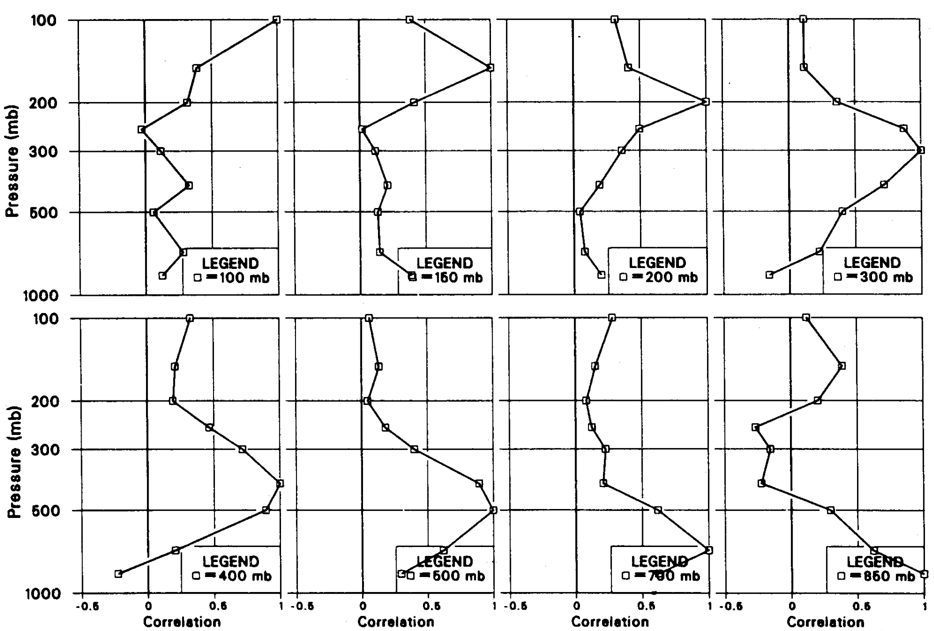


Fig. 16. Non-divergent wind prediction error vertical correlations for a selected set of standard levels, indicated in th legend of each frame of the plot. The plots correspond to particular columns of the vertical correlation matrix fo non-divergent wind.

Height

Lonnberg and Hollingsworth (1986)

Non-divergent wind

Hollingsworth and Lonnberg (1986)

Multivariate analyses

For K observation locations, define:

$$\mathbf{x}^\top = [p_1, u_1, v_1, p_2, u_2, v_2, \ldots, p_K, u_K, v_K].$$

$$\mathbf{P}^b = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1K} \\ b_{21} & b_{22} & \cdots & b_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ b_{K1} & b_{K2} & \cdots & b_{KK} \end{bmatrix}, b_{ij} = \begin{bmatrix} C_{pp}(\mathbf{r}_j, \mathbf{r}_k) & C_{pu}(\mathbf{r}_j, \mathbf{r}_k) & C_{pv}(\mathbf{r}_j, \mathbf{r}_k) \\ C_{up}(\mathbf{r}_j, \mathbf{r}_k) & C_{uu}(\mathbf{r}_j, \mathbf{r}_k) & C_{uv}(\mathbf{r}_j, \mathbf{r}_k) \\ C_{vp}(\mathbf{r}_j, \mathbf{r}_k) & C_{vu}(\mathbf{r}_j, \mathbf{r}_k) & C_{vv}(\mathbf{r}_j, \mathbf{r}_k) \end{bmatrix}$$

If $\mathbf{r}_i = (x_i, y_i)$, $\mathbf{r}_j = (x_j, y_j)$ and $u_j = u(x_j, y_j)$, $v_j = v(x_j, y_j)$ then

$$C_{pu}(\mathbf{r}_i, \mathbf{r}_j) = < p_i, u_j > .$$

Now introduce a linear relationship, e.g.

$$fu = -\frac{\partial \phi}{\partial y}, \qquad fv = \frac{\partial \phi}{\partial x}.$$

Then

$$C_{pu}(x_i, y_i, x_j, y_j) = < p_i, u_j > = -\frac{1}{f} < \phi_i \frac{\partial \phi_j}{\partial y_j} > = -\frac{1}{f} \frac{\partial}{\partial y_j} < \phi_i \phi_j > .$$

Assume homogeneity. Define,

$$\tilde{\mathbf{r}}^2 = \tilde{x}^2 + \tilde{y}^2, \qquad \tilde{x} = x_i - x_j, \qquad \tilde{y} = y_i - y_j.$$

Then,

$$\frac{\partial}{\partial y_i} = \frac{\partial}{\partial \tilde{y}}, \qquad \frac{\partial}{\partial y_j} = -\frac{\partial}{\partial \tilde{y}}$$

and similarly for x derivatives. Then

$$
\begin{aligned}
C_{pp}(\mathbf{r}_i, \mathbf{r}_j) &= <\phi_i \phi_j> \\
C_{pu}(\mathbf{r}_i, \mathbf{r}_j) &= -C_{up}(\mathbf{r}_i, \mathbf{r}_j) = \frac{1}{f}\frac{\partial}{\partial \tilde{y}} <\phi_i \phi_j> \\
C_{pv}(\mathbf{r}_i, \mathbf{r}_j) &= -C_{vp}(\mathbf{r}_i, \mathbf{r}_j) = -\frac{1}{f}\frac{\partial}{\partial \tilde{x}} <\phi_i \phi_j> \\
C_{uu}(\mathbf{r}_i, \mathbf{r}_j) &= -\frac{1}{f^2}\frac{\partial^2}{\partial \tilde{y}^2} <\phi_i \phi_j> \\
C_{vv}(\mathbf{r}_i, \mathbf{r}_j) &= -\frac{1}{f^2}\frac{\partial^2}{\partial \tilde{x}^2} <\phi_i \phi_j> \\
C_{uv}(\mathbf{r}_i, \mathbf{r}_j) &= C_{vu}(\mathbf{r}_i, \mathbf{r}_j) = \frac{1}{f^2}\frac{\partial^2}{\partial \tilde{x} \partial \tilde{y}} <\phi_i \phi_j>
\end{aligned}
$$

Multivariate correlations

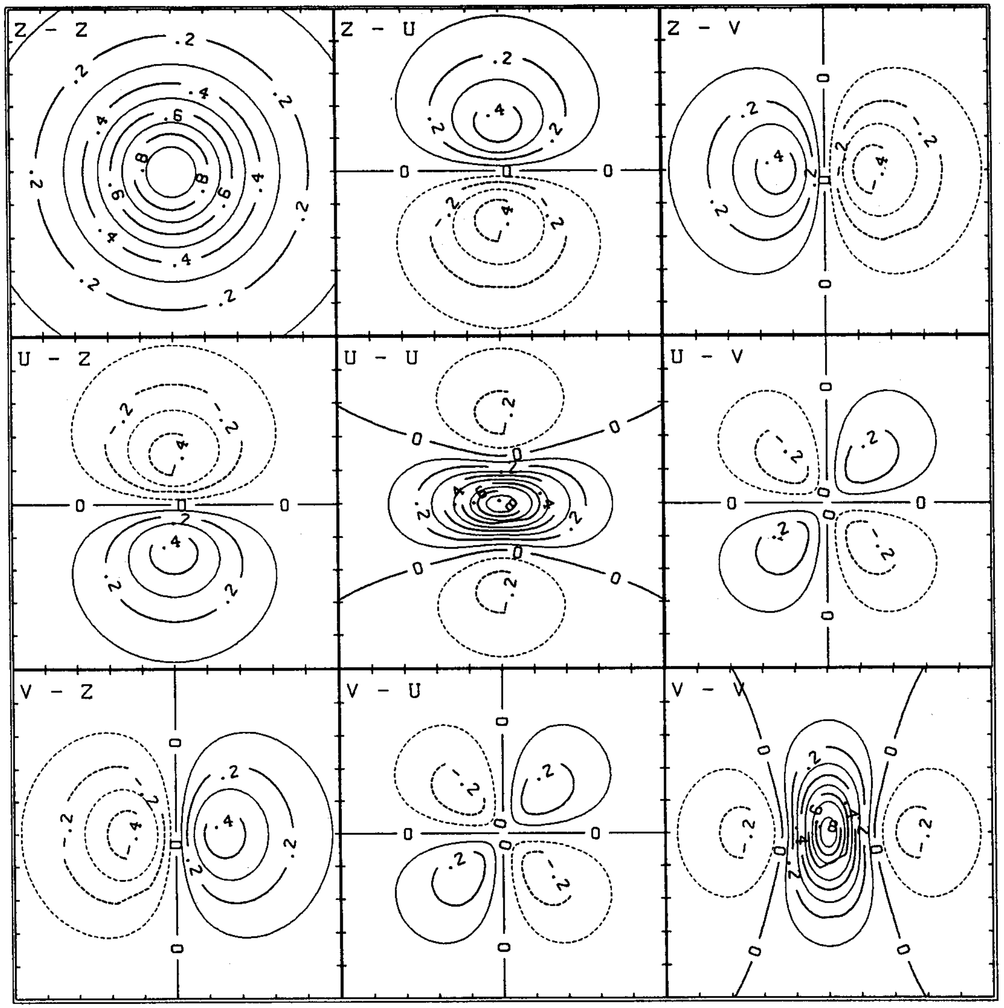


FIG. 17. The complete set of horizontal prediction error correlations for the variables z, u and v based on (7a) with $N = 3$ and $\alpha = 0.2$. The scale parameter, c, is that obtained at 500 mb. Tic marks along margins are 250 km apart.

Mitchell et al. (1990)

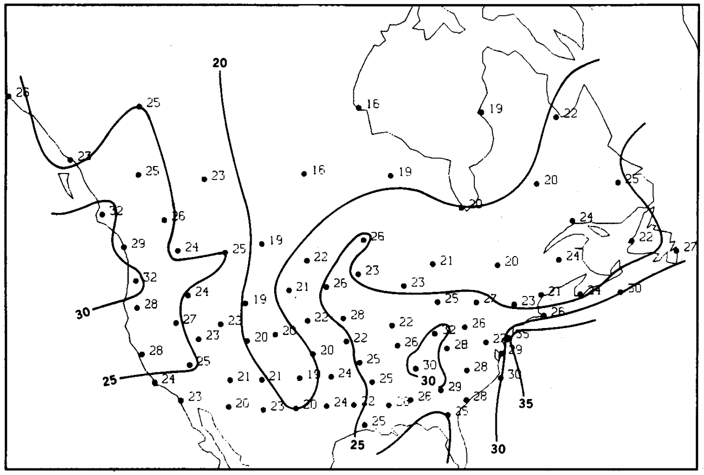


Figure 4.13 The standard deviation of the 250 mb geopotential background (forecast) error over North America (m). (After Lönnberg and Hollingsworth 1986)

If covariances are homogeneous, variances are independent of space

Covariances are not homogeneous

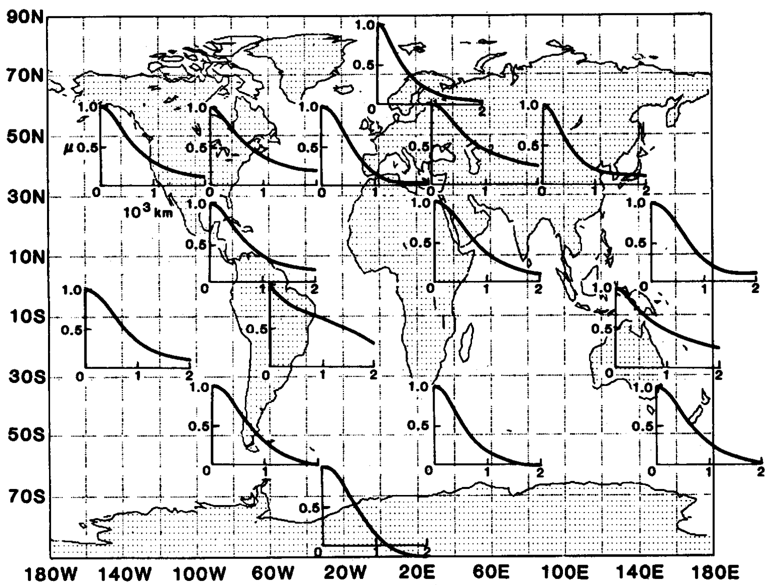


Figure 4.5 Isotropic component of 500 mb geopotential background (forecast) error correlation in different parts of the globe. (After Baker et al. *Mon. Wea. Rev.* **115**: 272, 1987. The American Meteorological Society.)

If correlations are homogeneous, correlation lengths are independent of location

Correlations are not homogeneous

Correlations are not isotropic

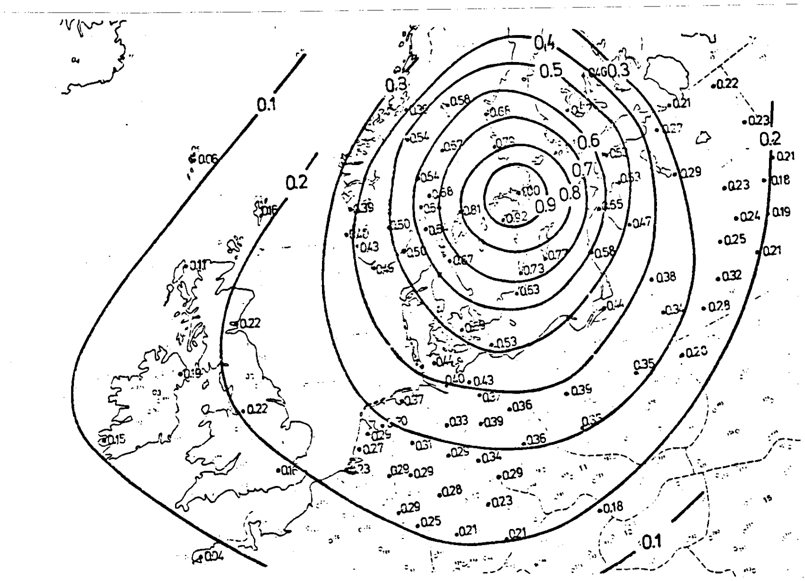


Figure 5.1.1.4.1 Auto-correlation of errors in 12h numerical forecasts of surface pressure in a reference station (Stockholm) and other stations.

Gustafsson (1981)

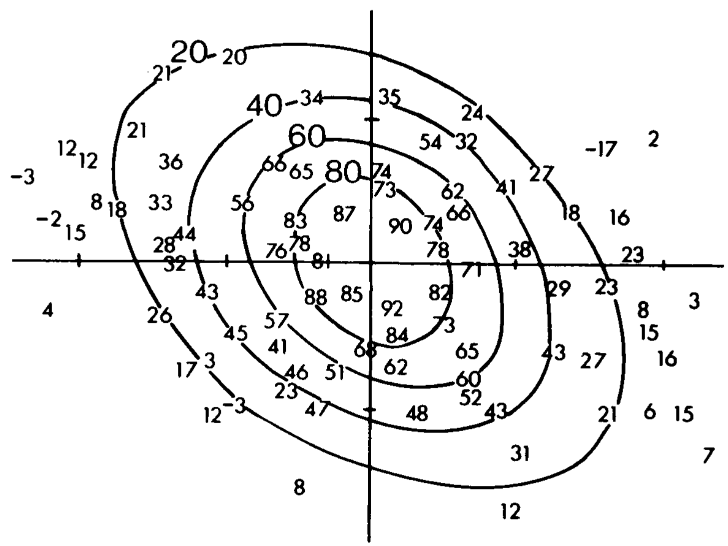


Figure 4.2 Observed-minus-background (climatology) correlation for the 500 mb geopotential field over Australia. All correlations are with respect to the observation station at the origin. (After Seaman, *Aus. Met. Mag.* **30**, 133, 1982. AGPS Canberra, reproduced by permission of Commonwealth of Australia copyright.)

Daley (1991)

Are correlations separable?

If so, correlation length should be
Independent of height.

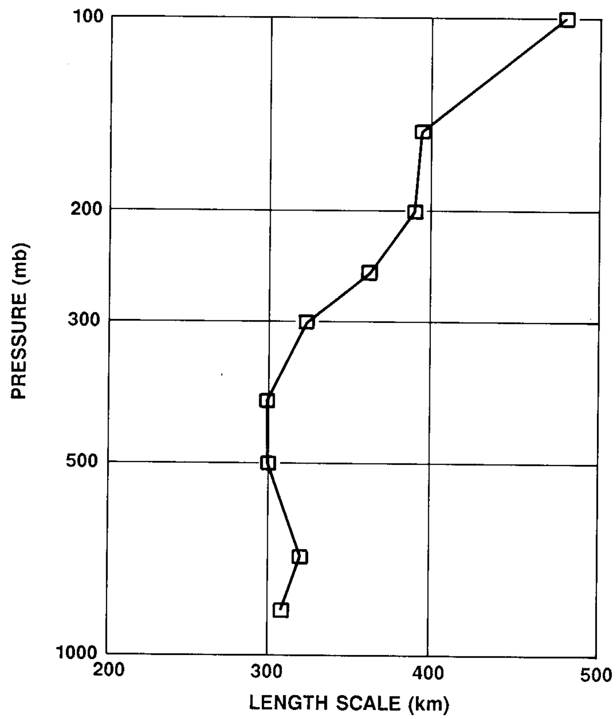


Figure 4.12 The characteristic scale of the geopotential background (forecast) error correlation for the North American radiosonde network as a function of pressure. (After Lönnberg and Hollingsworth 1986)

Lonnberg and Hollingsworth (1986)

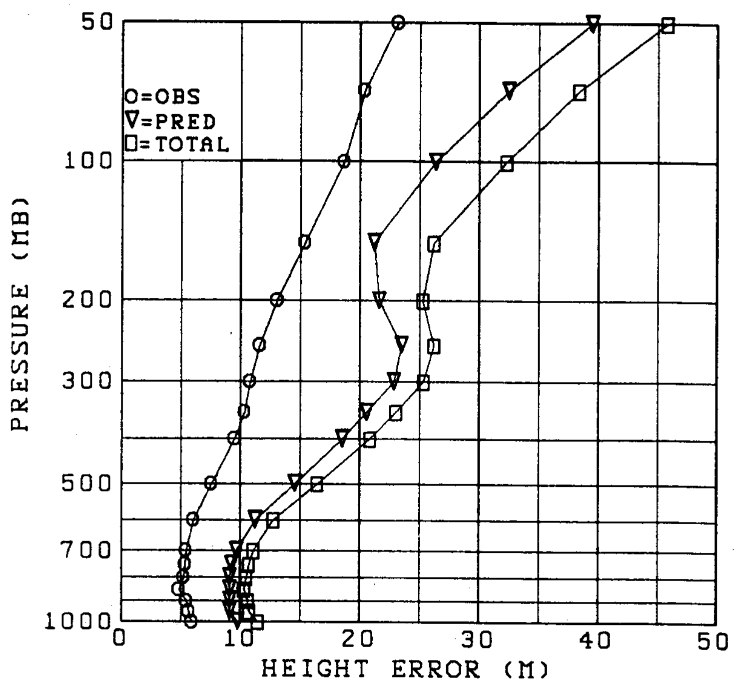


FIG. 8. Vertical profile of the observed height residual (m) (i.e., total perceived forecast error) denoted TOTAL, and the corresponding profiles of prediction and observation error.

Mitchell et al. (1990)

Covariance modelling assumptions:

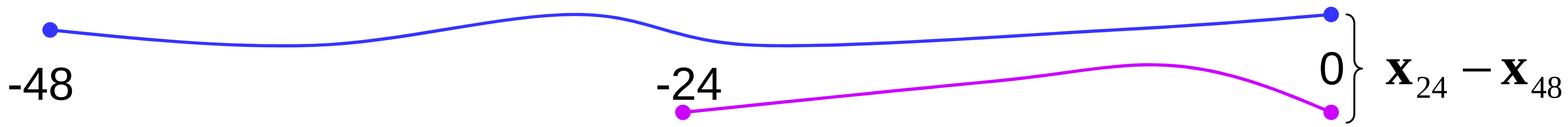
1. No correlations between background and obs errors
2. No horizontal correlation of obs errors
3. Homogeneous, isotropic horizontal background error correlations
4. Separability of vertical and horizontal background error correlations

None of our assumptions are really correct. Therefore Optimal Interpolation is not optimal so it is often called Statistical Interpolation.

2. NMC-method

- Need global statistics
- N. American radiosonde network is only 4000 km in extent defining only up to wavenumber 10. Vertical and horizontal resolution is too coarse.
- A posteriori justification: compare resulting statistics with those obtained using other methods

The NMC-method



- Compares 24-h and 48-h forecasts valid at same time
- Provides global, multivariate corr. with full vertical and spectral resolution
- Not used for variances
- Assumes forecast differences approximate forecast error

$$\mathbf{x}_{24} - \mathbf{x}_{48} \approx \mathbf{x}_{0} - \mathbf{x}_{6}\ ?$$

Why 24 – 48 ?

- 24-h start forecast avoids "spin-up" problems
- 24-h period is short enough to claim similarity with 0-6 h forecast error. Final difference is scaled by an empirical factor
- 24-h period long enough that the forecasts are dissimilar despite lack of data to update the starting analysis
- 0-6 h forecast differences reflect assumptions made in OI background error covariances

NMC-method usage

Center	Region	Reference
NCEP	U.S.A.	Parrish & Derber 1991
ECMWF	Europe	Rabier et al.1998
CDC	Canada	Gauthier et al. 1999
Met Office	U.K.	Ingleby et al. 1996
BMRC	Australia	Steinle et al. 1995
Meteo-Fr.	France	Desroziers et al. 1995

Properties of the NMC-method Bouttier (1994)

- For linear H, no model error, 6-h forecast difference, can compare NMC P calc. to what Kalman Filter suggests.
- NMC-method breaks down if there is no data between launch of 2 forecasts. With no data P is under-estimated
- For dense, good quality hor. uncorr. obs, P is over-estimated
- For obs at every gridpoint, where obs and bkgd error variances are equal, the NMC-method P estimate is equivalent to that from the KF.

A posteriori justification:
compare NMC results to innovation-method results

Horizontal correlation length scale

NMC

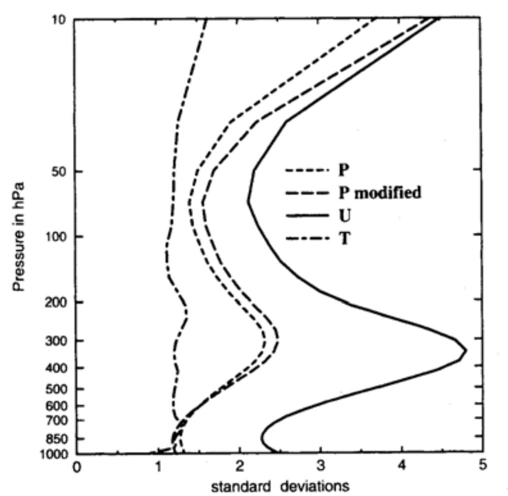


Figure 1. Horizontally averaged standard deviations as a function of pressure (hPa) for the mass variable P (dashed line) ($100\ m^2s^{-2}$), mass variable modified as explained in section 4 of the text (long-dashed line) ($100\ m^2s^{-2}$), wind components (solid line) ($m\ s^{-1}$) and temperature (dash-dotted line) (K).

Rabier et al. (1998)

Innovations

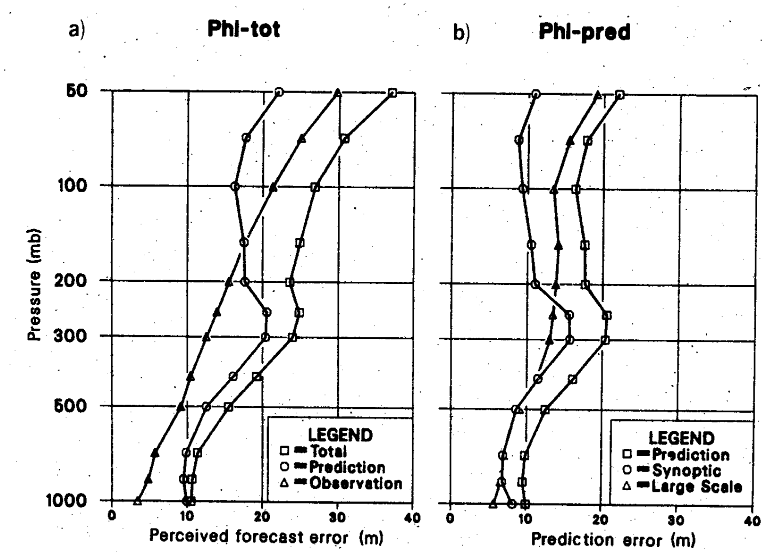


Fig. 2. (a) Vertical profiles of the total, or perceived, forecast error of height, together with the contributions to this error from prediction error, and the observation error. The unit is metres. (b) Vertical profiles of the prediction error (copied from 2a) and of the contributions of the synoptic-scale and large-scale components to the prediction error. The sum of the squares of the components gives the square of the prediction error.

Hollingsworth and Lonnberg (1986)

Different vertical correlation lengths for different wavenumbers

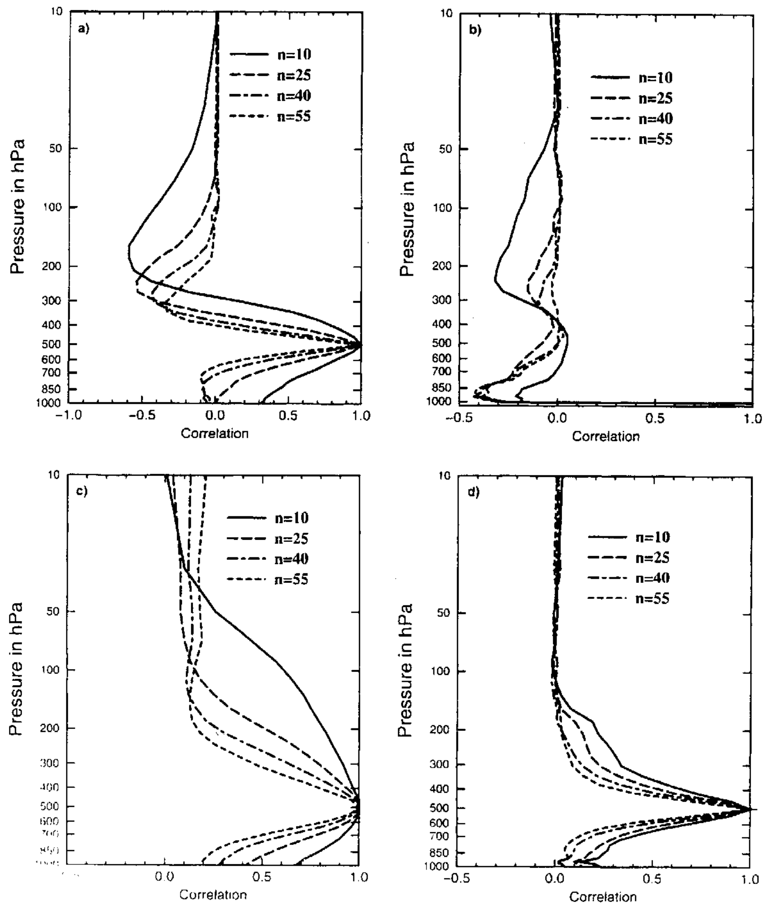


Figure 9. Vertical correlations for level 18 (approximately 500 hPa) as a function of pressure (hPa) for selected horizontal wave-numbers: (a) temperature T; (b) cross-correlation between surface pressure (represented by the lowest point on the curve) and temperature; (c) the mass variable P; (d) specific humidity Q. Wave numbers 10, 25, 40 and 55 are shown by solid, long-dashed, dash-dotted and dashed lines respectively.

Rabier et al. (1998)

Different horizontal correlation lengths for different vertical levels

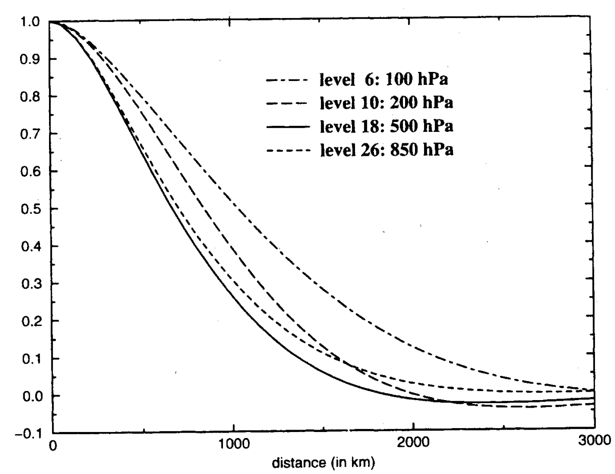


Figure 4. Horizontal autocorrelations as a function of horizontal distance for the mass variable P and selected model-levels. Level 6 (approximately 100 hPa) is denoted by a dash-dotted line, level 10 (approximately 200 hPa) by a long-dashed line, level 18 (approximately 500 hPa) a solid line and level 26 (approximately 850 hPa) a dashed line.

Rabier et al. (1998)

3. Ensemble method of calc. B matrix

Buehner (2004), Fisher (1999)

1. Run a control forecast, and a perturbed forecast:

$$\begin{aligned} \mathbf{x}^f &= M(\mathbf{x}^a) \\ \tilde{\mathbf{x}}^f &= M(\tilde{\mathbf{x}}^a) + \mathbf{w} \end{aligned}$$

   where **w** is $\mathcal{N}(\mathbf{0}, \mathbf{Q})$ and **Q**=s*__B__. s is a scaling factor computed every 6 hrs and is a function of latitude band (NH, SH or tropics) and vertical level.

2. Run a control and a perturbed analysis:

$$\begin{aligned} \mathbf{x}^a &= \mathbf{x}^f + \mathbf{K}(\mathbf{z} - H(\mathbf{x}^f)) \\ \tilde{\mathbf{x}}^a &= \tilde{\mathbf{x}}^f + \mathbf{K}(\mathbf{z} + \mathbf{v} - H(\tilde{\mathbf{x}}^f)) \end{aligned}$$

   where **v** is $\mathcal{N}(\mathbf{0}, \mathbf{R})$.

3. Repeat steps 1 and 2 every cycle, for one month or so.

4. Compute the forecast error:

$$\mathbf{e}_6^f = \boldsymbol{x}_6^f - \tilde{\boldsymbol{x}}_6^f.$$

By definition,

$$\mathbf{P}_6^f =< (\mathbf{e}_6^f - < \mathbf{e}_6^f >)(\mathbf{e}_6^f - < \mathbf{e}_6^f >)^\mathsf{T} > .$$

Replace $<>$ with time averages, [ ] to get

$$\mathbf{B} \approx [(\mathbf{e}_6^f - [\mathbf{e}_6^f])(\mathbf{e}_6^f - [\mathbf{e}_6^f])^\mathsf{T}].$$

- Fisher (1999) uses stochastic physics to represent model error instead of **Q**.
- Actual EnsKF uses $\mathbf{P}^f$ from ensembles in step 2, while Buehner (2004) uses **B**.

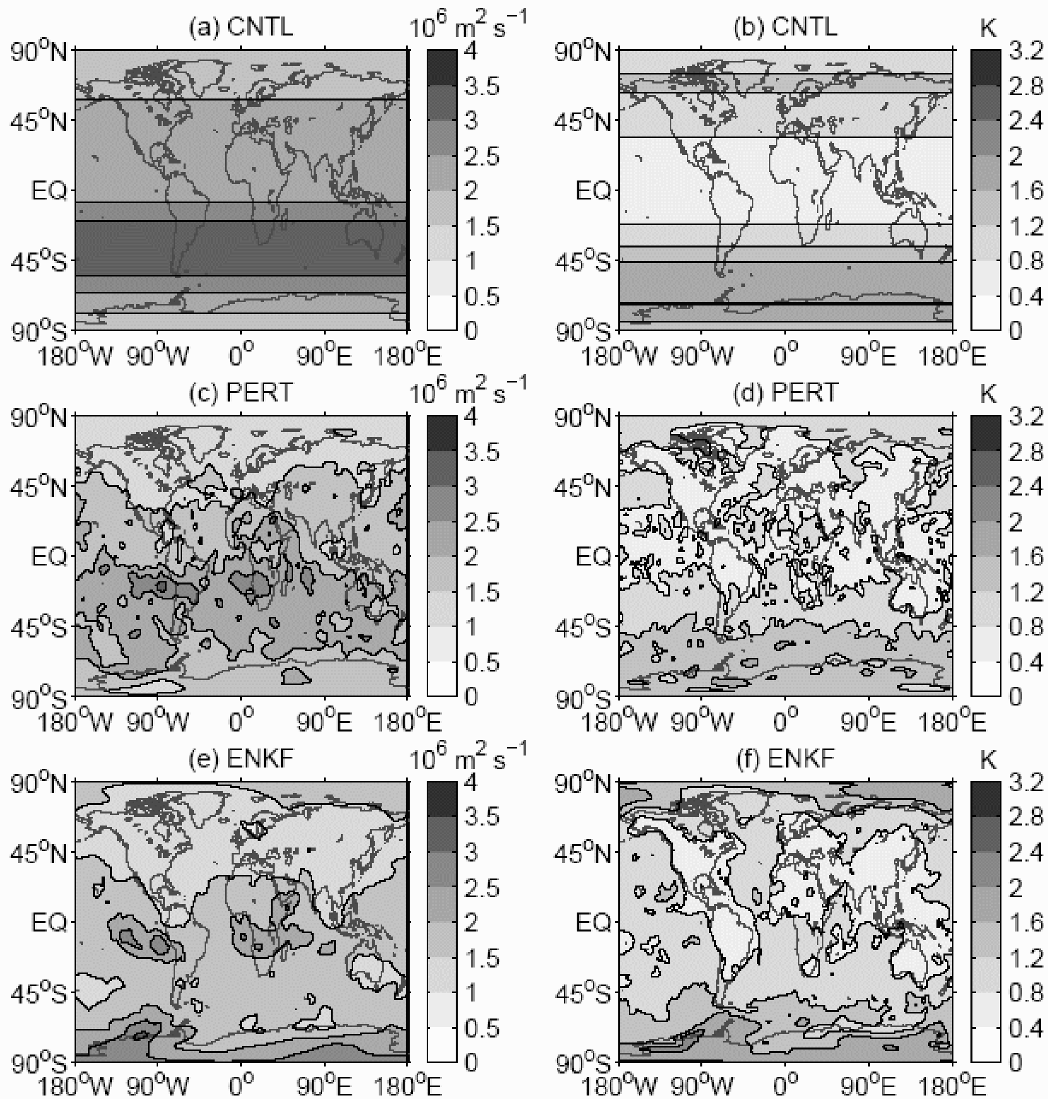


Figure 3. Estimated background error std dev of (a) streamfunction near 250 hPa ($\eta = 0.258$) and (b) temperature near 500 hPa ($\eta = 0.516$) computed using the "NMC method". Similarly, panels (c) and (d) show the same results estimated from the perturbed 3D-Var experiments and panels (e) and (f) show the results estimated by temporally averaging the background ensemble spread variances from the EnKF.

Buehner (2004, submitted)

Some final words…

Fifteen years ago, data assimilation was a minor and often neglected sub-discipline of numerical weather prediction. The situation is very different today. Data assimilation is now felt to be important for all climate and environmental monitoring and estimating the ocean state. There have been great advances in both modelling and instrumentation for a variety of atmospheric phenomena and variables, and data assimilation provides the bridge between them….

(Roger Daley, 1997)