MTMD02
Operational Data Assimilation Techniques

Part I: Variational techniques (Ross Bannister)

1. Introduction
   (a) Inverse problems
   (b) Notation
   (c) History of data assimilation in meteorological operations and the data assimilation cycle
   (d) The scale/challenges of the operational problem
2. Variational techniques (VAR)
   (a) Euler-Lagrange equations
   (b) Error covariance matrices
   (c) Cost functions and simplifications for operational assimilation
   (d) Optimal interpolation and physical space analysis systems
3. A-priori information and the B-matrix
   (a) The null space of the observation operator and the importance of a-priori information
   (b) The role of the background error covariance matrix
   (c) Spatial aspects (inverse Laplacians, diffusion operators)
   (d) Multivariate aspects and balance
   (e) Control variable transforms and the implied B-matrix
   (f) Conditioning
4. Operational algorithms
5. Measuring the B-matrix
   (a) Analysis of innovations
   (b) NMC method
   (c) Monte-Carlo (ensemble) method
6. Hybrid (var/ensemble) formulations
   (a) Basic ideas
   (b) Incorporating a simple hybrid scheme in VAR
   (c) Incorporating a localized hybrid scheme in VAR
7. Data assimilation diagnostics
   (a) The Bennet-Talagrand theorem
   (b) Desrozier diagnostics

Further reading

- Bennett A.F., 2002, Inverse Modeling of the Ocean and Atmosphere (Euler-Lagrange equations and representers - sections 1.2, 1.3).
- Daley R., 1991, Atmospheric Data Analysis (historical aspects and basic ideas - chapters 1, 13).
- Kalnay E., 2003, Atmospheric Modeling, Data Assimilation and Predictability (basic aspects of data assimilation - chapter 5).
- Lewis J.M., Lakshmivarahan S., Dhall S.K., 2006, Dynamic data assimilation: a Least Squares Approach (applications - chapters 3,4, data assimilation algorithms - chapter 19).
- Schlatter T.W., 2000, Variational Assimilation of Meteorological Observations in the Lower Atmosphere: a Tutorial on How it Works, Journal of Atmospheric and Solar-Terrestrial Physics 62, pp. 1057-1070.
- Mathematics Aide Memoir handout.

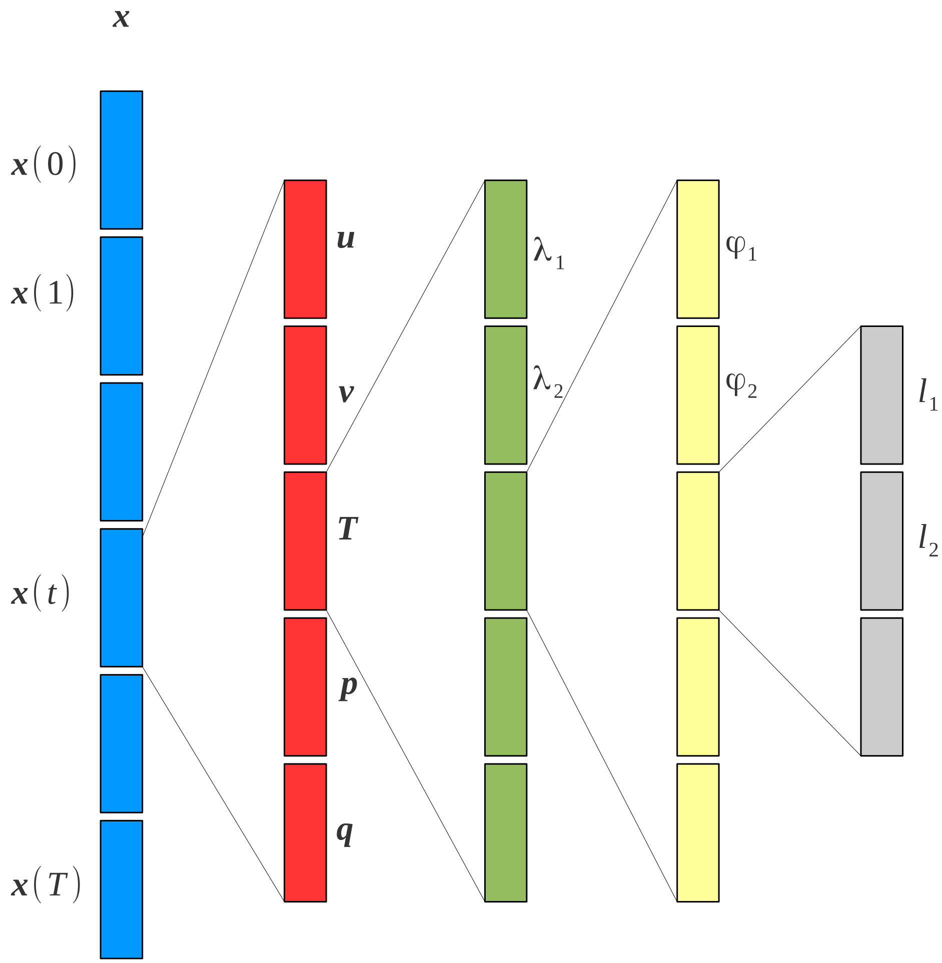
Note that page numbers on the slides and on the handouts do not always match.

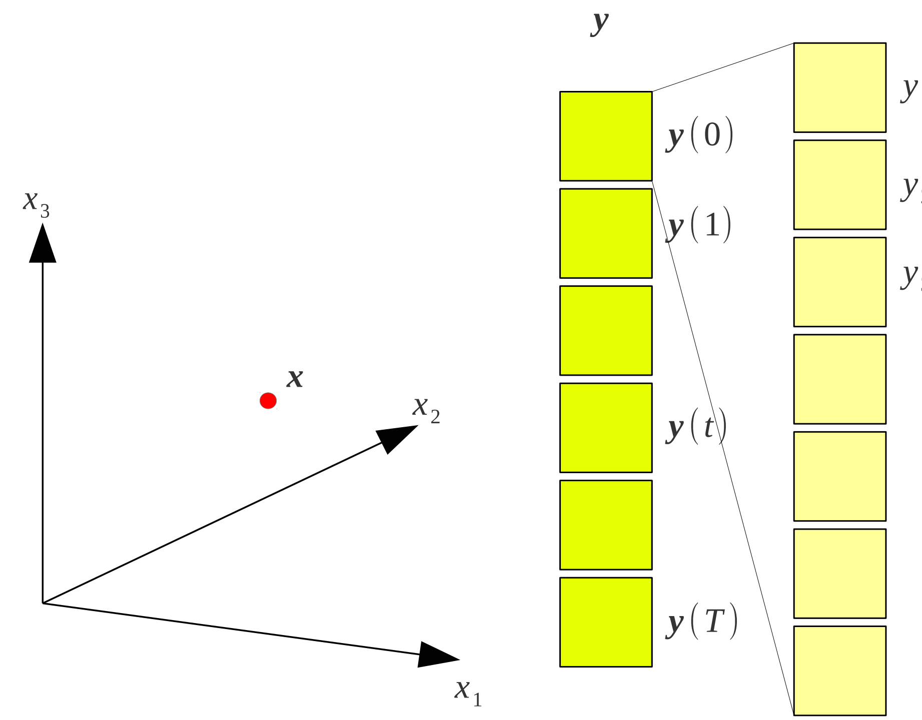
1. Introduction

1(a) Inverse problems

Field	Example inverse problem to be solved
Medical diagnosis	What is the 3-D structure of biological tissues from X-ray images (CAT scan)?
Seismology	Determination of subterranean properties from seismic data (e.g. porosity, hydrocarbon content)
Astrophysics	Determination of the internal structure of the Sun from surface observations
Astronomy	Orbit determination from observations
Astronautics	Landing a spacecraft safely on another planet
Parameter estimation	Determination of unknown model parameters
Atmospheric pollution	What is the source/sink field of an atmospheric pollutant?
Atmospheric retrievals	What is the vertical profile of atmospheric quantities from remotely sensed observations?
Weather forecasting	What are the initial conditions (e.g. u, v, T, p, q, cloud, SST, salinity) of an atmosphere or ocean forecast model that agrees with the latest observations?

1(b) Notation



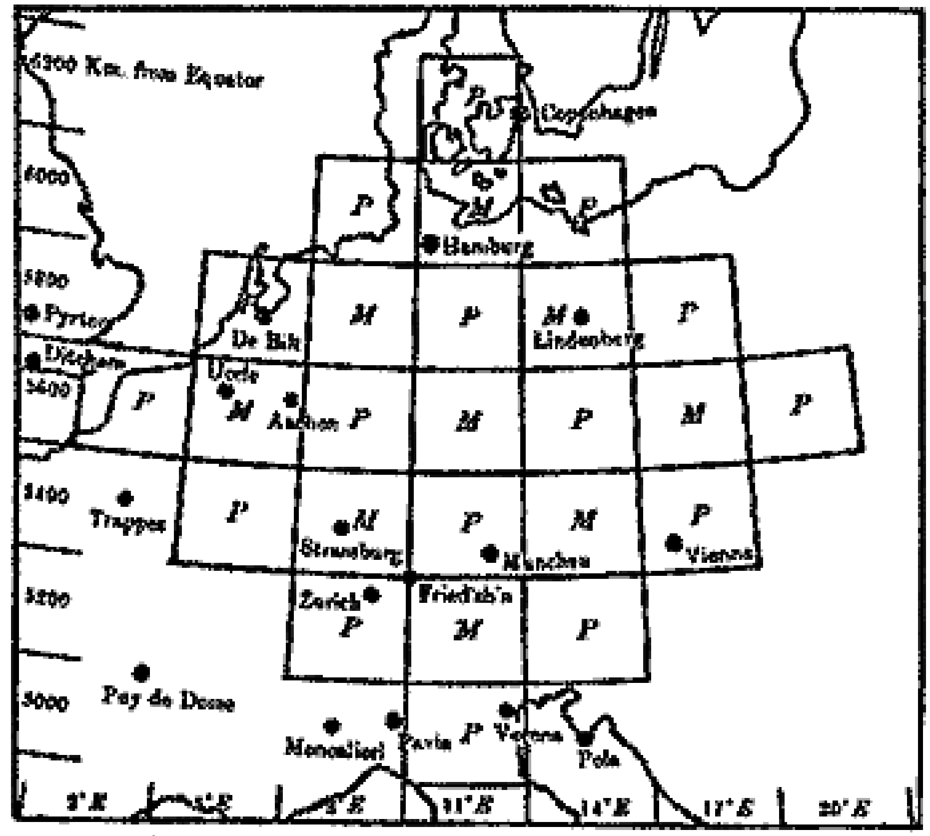


$\mathbf{x}_\mathrm{A}$ analysis state
$\mathbf{x}_\mathrm{B}$ background state
$\delta\mathbf{x}$ incremental state
Sometimes $\mathbf{x}$ and $\mathbf{y}$ are for only one time
$\mathbf{x}$-vectors have n elements in total
$\mathbf{y}$-vectors have p elements in total

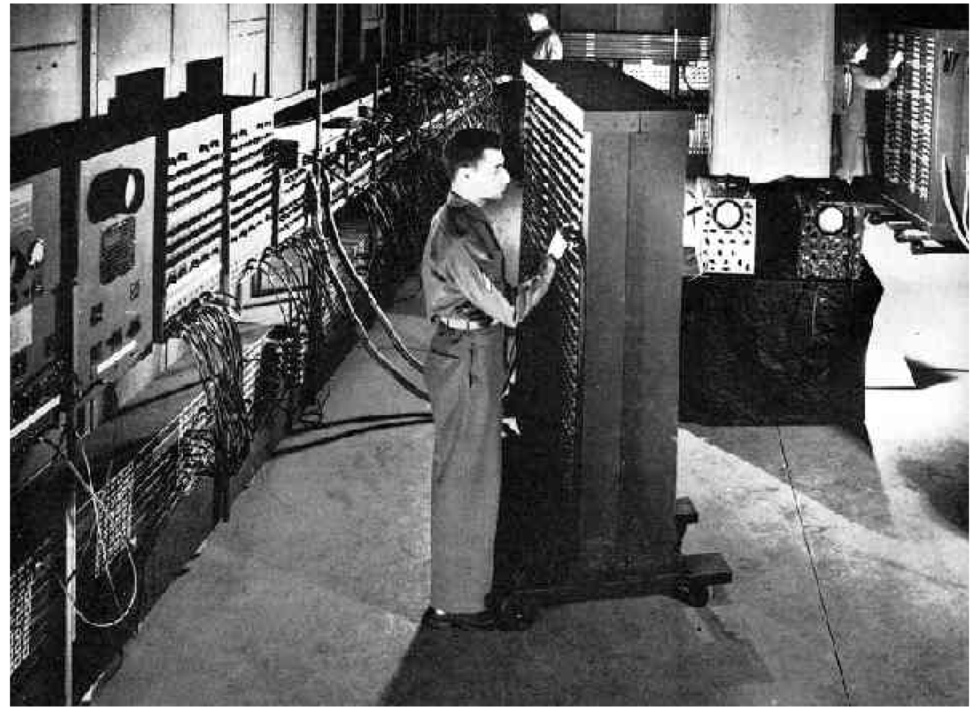
1(c) History of data assimilation in meteorological operations and the data assimilation cycle

Subjective 'data assimilation' 1910s, 1920s

- LF Richardson (1922) attempted a hind-cast (by hand!) for 20th May 1910.
- Primitive equation-based forecast model: resolution $\Delta\lambda = 3°$, $\Delta\phi = 1.8°$, 5 vertical levels.
- 'Data assimilation' was done for mass variables (T, p) separately from wind variables (u, v) (i.e. univariate) by interpolating observations subjectively.
- A disastrous forecast: $\Delta P/\Delta t \approx 145 \text{ hPa } /6 \text{ hours.}$
- Catastrophic growth rate not due to the model, but due to inadequate data assimilation – the mass and wind were out of balance.
- Bjerknes, 1911, described the analysis problem as, "The ultimate problem in Meteorology".



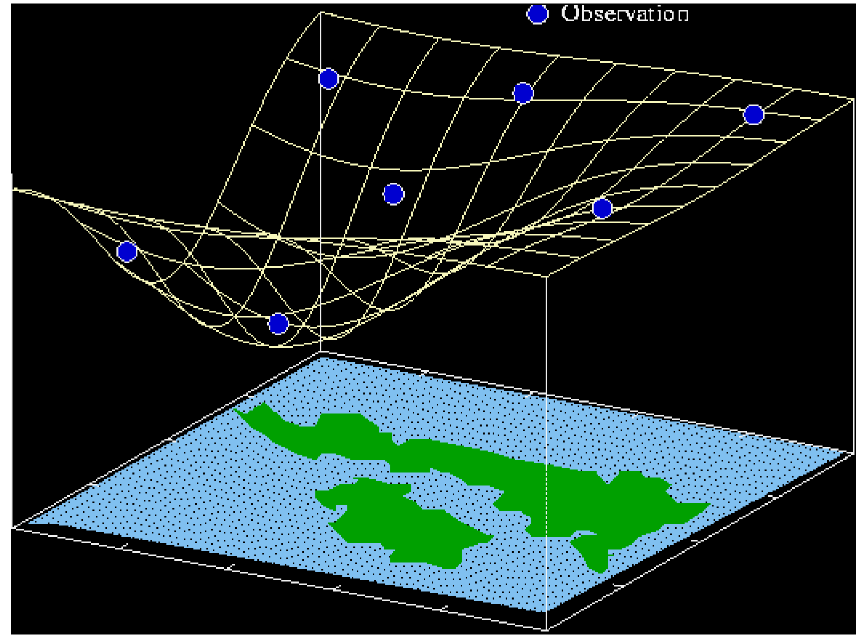
Successes in NWP, 1940s



- Success with filtered dynamical models containing balanced motion only (e.g. barotropic vorticity equation), even with subjective analysis.
- BVE is less accurate than the primitive equations, but is insensitive to imbalances in the initial conditions (there are no gravity waves in the BVE).
- ENIAC (Electronic Numerical Integrator and Computer).
- [We now use primitive equations for NWP, but with DA that inhibits imbalance.]

Beginnings of objective analysis: polynomial fitting, late 1940s

- Fit a polynomial expansion to observations.
- Made no account of observation accuracy.
- Different variables treated independently (univariate).
- Direct observations only.
- Unrealistic values in data voids.



Cressman analysis / method of successive corrections, 1950s, 1960s

- Use prior knowledge (a background state).
- Provides information in data voids.
- Prior knowledge can come from climatology or a previous forecast.
- Latter leads on to the 'data assimilation cycle'.

$$\mathbf{x}_i^0 = \text{first guess (background)}$$

$$\mathbf{x}_i^{n+1} = \mathbf{x}_i^n + \frac{\sum_{k=1}^{K_i^n} W_{ik}^n (y_k - \tilde{\mathbf{x}}_k^n)}{\sum_{k=1}^{K_i^n} W_{ik}^n + \epsilon_k^2}$$

- $\mathbf{x}_i^n$ estimate of field at grid point i after the nth iteration.
- $\tilde{\mathbf{x}}_k^n$ field value at grid location closest to observation k.
- W_{ik}^n weight of influence of observation k on grid point i (reduces with distance).
- K_i^n number of observations within distance R^n of grid point i.
- y_k kth observation value.
- ϵ_k controls the degree of influence of the observations on the analysis (diminishing influence as $\epsilon \to \infty$).

Nudging (Newtonian relaxation), 1970s - present

- Allows the analysis to be combined with the background state smoothly.
- Relies on an intermediate analysis, $\mathbf{x}_{\rm int}$ (e.g. from SCM).
- $\mathbf{x}_{\rm int}$ to be introduced over a timescale τ.
- Model equations:
$$\frac{\partial \mathbf{x}}{\partial t} = \mathbf{f}(\mathbf{x}),$$
- ... are modified to:
$$\frac{\partial \mathbf{x}}{\partial t} = \mathbf{f}(\mathbf{x}) - \frac{\mathbf{x} - \mathbf{x}_{\rm int}}{\tau}.$$

Example with a scalar (x) for a persistence model ($f(x) = 0$):

$$\begin{aligned} \frac{dx}{dt} &= -\frac{x - x_{\rm int}}{\tau}, \\ \Rightarrow x(t) &= x_{\rm int} + (x(0) - x_{\rm int}) \exp -\frac{t}{\tau}. \end{aligned}$$

Optimal interpolation, 1970s / 1980s

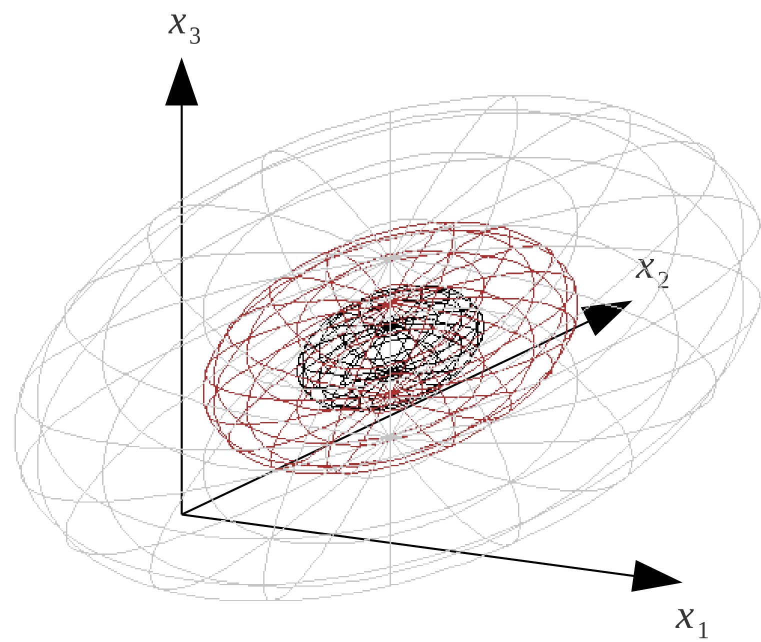
$$\mathbf{x}_{\mathrm{A}} = \mathbf{x}_{\mathrm{B}} + \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}})^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}_{\mathrm{B}}))$$

- A formal way of combining observations and models.
- Intimately related to method of least squares.
- Represents uncertainties of all information.
- Too expensive to solve for the global system (solve for patches and glue together for 'global' analysis).
- Need accurate estimates of $\mathbf{B}$ and $\mathbf{R}$ matrices.

- $\mathbf{x}_{\mathrm{A}}$ analysis state (posterior) $\in \mathbb{R}^n$.
- $\mathbf{x}_{\mathrm{B}}$ background state (prior) $\in \mathbb{R}^n$.
- $\mathbf{B}$ background error covariance matrix (accounts for uncertainty in $\mathbf{x}_{\mathrm{B}}$) $\in \mathbb{R}^{n\times n}$.
- $\mathbf{y}$ observation vector $\in \mathbb{R}^p$.
- $\mathbf{h}$ observation operator $\mathbb{R}^n \to \mathbb{R}^p$.
- $\mathbf{H}$ Jacobian of $\mathbf{h} \in \mathbb{R}^{p\times n}$.
- $\mathbf{R}$ observation error covariance matrix (accounts for uncertainty in $\mathbf{y}$) $\in \mathbb{R}^{p\times p}$.

Variational methods (VAR), 1990s / 2000s

- Broadly speaking (in the case of 3D-VAR) a way of solving the OI equations efficiently.
- Construct a cost functional, $J[\mathbf{x}]$ as the sum of squares of deviations from data.
- Analysis is defined as the $\mathbf{x}$ that minimizes $J[\mathbf{x}]$.
- $\mathbf{B}$ is not applied as an explicit matrix, but is instead modelled (see later).
- Efficient enough for a truly global analysis.
- Still need accurate estimates of $\mathbf{B}$ and $\mathbf{R}$ matrices. $\mathbf{B}$ is usually static.
- Variants: 1D-VAR / 3D-VAR / 4D-VAR / etc. (see later).



- Example for strong constraint 4D-VAR:

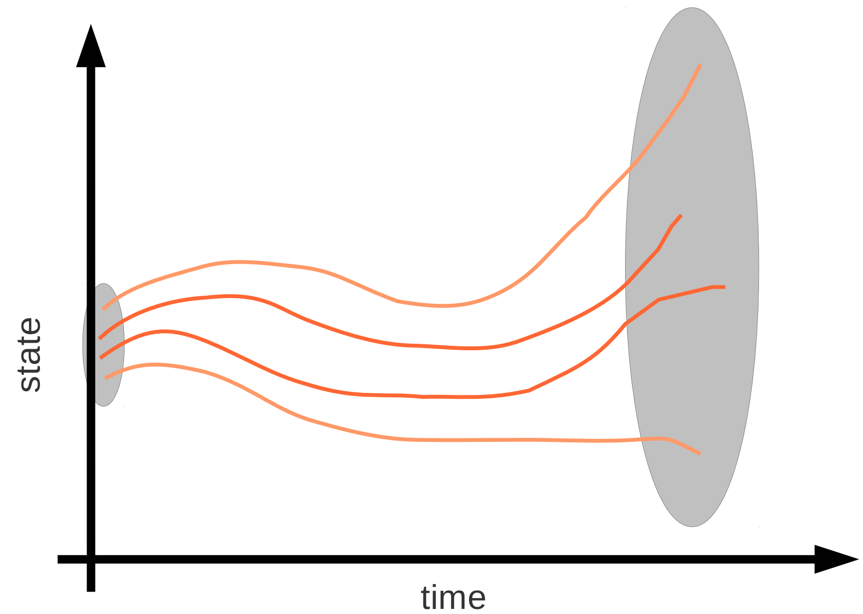
$$
\begin{aligned}
J(\mathbf{x}) \;=\; & \frac{1}{2}\left(\mathbf{x}-\mathbf{x}_{\mathrm{B}}\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}_{\mathrm{B}}\right) + \\
& \frac{1}{2}\sum_{t=0}^{T}\left(\mathbf{y}(t)-\mathbf{h}_t[\mathcal{M}_{t\leftarrow 0}(\mathbf{x})]\right)^{\mathrm{T}}\mathbf{R}_t^{-1}\times \\
& \qquad \left(\mathbf{y}(t)-\mathbf{h}_t[\mathcal{M}_{t\leftarrow 0}(\mathbf{x})]\right),
\end{aligned}
$$

where $\mathbf{x}$ is the state vector at $t=0$.

Part I of this course is mainly about variational methods.

Ensemble methods 2000s / 2010s

- The spread in an ensemble of N background forecasts has information about background uncertainty, member i $\mathbf{x}^{(i)}$.



- Flow-dependent background error covariances, $\mathbf{P}^{\mathrm{f}}$.
- Formulation starts with the OI equation ($\mathbf{B} \rightarrow \mathbf{P}^{\mathrm{f}}$), but for an ensemble of states.
- Does not need the $\mathbf{P}^{\mathrm{f}}$-matrix explicitly.
- Severe rank deficiency problems with $\mathbf{P}^{\mathrm{f}}$ due to undersampling (use, e.g., localization techniques to overcome).
- Deterministic (square-root) and non-deterministic (non-square-root) formulations exist - see part II of the course.

$$\mathbf{P}^{\mathrm{f}} \approx \mathbf{P}^{\mathrm{f}}_{(N)} = \frac{1}{N-1} \sum_{i=1}^{N} \left\langle \left(\mathbf{x}^{(i)} - \langle \mathbf{x} \rangle\right) \left(\mathbf{x}^{(i)} - \langle \mathbf{x} \rangle\right)^{\mathrm{T}} \right\rangle ,$$

$$\langle \mathbf{x} \rangle \approx \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}^{(i)}.$$

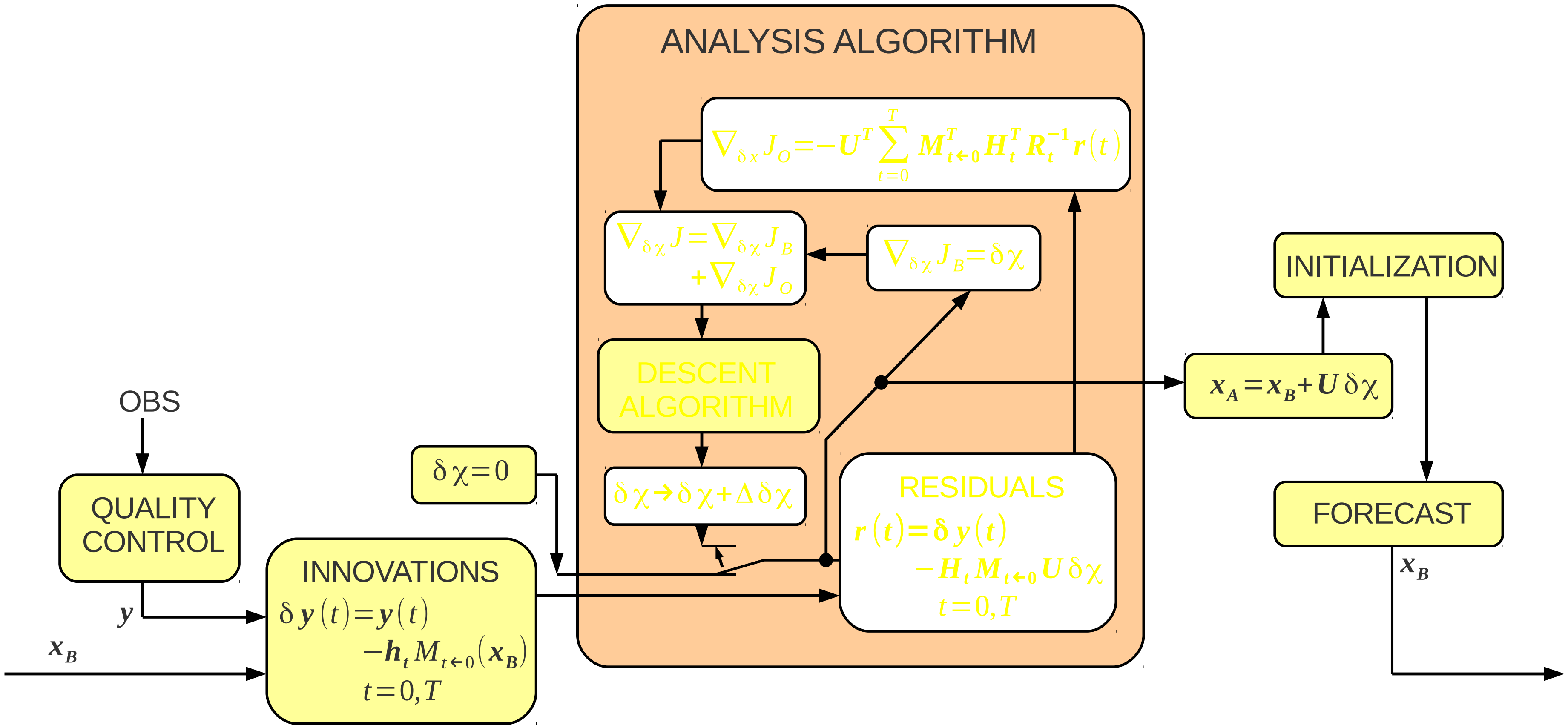
Hybrid methods 2010s

- Combine the robustness of the $\mathbf{B}$-matrix with the flow-dependence of the $\mathbf{P}^{\mathrm{f}}$-matrix.
- Most simple is the arithmetic average:

$$\mathbf{P}^{\mathrm{H}} = \alpha\mathbf{B} + (1-\alpha)\mathbf{P}^{\mathrm{f}}_{(N)}$$

- Solve a VAR-like problem but $\mathbf{B} \rightarrow \mathbf{P}^{\mathrm{H}}$.
- Still need localization methods.
- Other approaches exist too.
- Uses methods that avoid the need to hold large matrices explicitly.

The data assimilation cycle



2. Variational techniques

2(a) The Euler-Lagrange equations

This section teaches us formally about the variational solution of an inverse problem, backward (or adjoint) variables and the strong and weak constraint formulations. The method of representers, used to solve the Euler-Lagrange equations, is introduced.

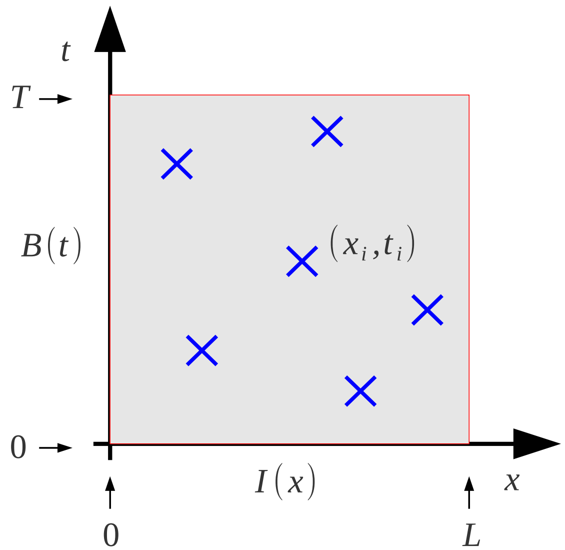
Statement of problem

What is the optimal state, $\phi(x, t)$ of the 1-D system whose dynamics are governed by

$$\frac{\partial \phi}{\partial t} + u\frac{\partial \phi}{\partial x} - F = e, \tag{1}$$

which lies close to some given observations, some initial conditions and some boundary conditions?The symbols have the following meanings:

- ϕ unknown tracer concentration,
- u known, but constant advection speed,
- F known source field, which may vary in time and
- e unknown model error, which may vary in time.



The (imperfectly known) information we have about the system is (see Fig.):

- $\phi(x,0) \approx I(x)$ imperfectly known initial conditions (i.c.s), $0 \le x \le L$,
- $\phi(0,t) \approx B(t)$ imperfectly known boundary conditions (b.c.s), $0 \le t \le T$, and
- y_m imperfect observation of the system (a direct observation of $\phi(x_m, t_m)$), $1 \le m \le p$.

The a-priori state $\phi_{\mathrm{B}}(x,t)$ satisfies the known bits of the problem (the specified i.c.s and b.c.s, and (1) with $e = 0$):

$$\frac{\partial \phi_{\mathrm{B}}}{\partial t} + u\frac{\partial \phi_{\mathrm{B}}}{\partial x} - F = 0, \qquad \phi_{\mathrm{B}}(x,0) = I(x), \qquad \phi_{\mathrm{B}}(0,t) = B(t). \tag{2}$$

The strong constraint formulation

The strong constraint formulation imposes the known parts of the system equations exactly (that is we assume that there is no model error, even though in reality there nearly always is). We still allow for imperfections in the other pieces of information though (i.c.s, b.c.s and observations). In order to find the optimal solution to the problem in this formulation, construct a functional $f[\phi]$:

$$f[\phi] \;=\; W_{\mathrm{ic}} \int_{x=0}^{L} dx\{\phi(x,0) - I(x)\}^2 + W_{\mathrm{bc}} \int_{t=0}^{T} dt\{\phi(0,t) - B(t)\}^2 +$$

$$W_{\rm ob}\sum_{i=1}^{p}\{\phi(x_i,t_i)-y_i\}^2. \tag{3}$$

We ask the question: what $\phi(x,t)$ makes $f[\phi]$ stationary, subject to the following model constraint:

$$g(x,t)=\frac{\partial\phi}{\partial t}+u\frac{\partial\phi}{\partial x}-F=0\,? \tag{4}$$

The constrained minimization problem introduces a new functional, J, which is the sum of f and all of the constraint terms. Each constraint term represents the constraint at a position and time and is represented as the product of $g(x,t)$ and the Lagrange multiplier, $2\lambda(x,t)$:

$$\begin{aligned} J[\phi,\lambda] &= f[\phi]+2\int_{x=0}^{L}dx\int_{t=0}^{T}dt\,\lambda(x,t)g(x,t),\\ &= W_{\rm ic}\int_{x=0}^{L}dx\{\phi(x,0)-I(x)\}^2+W_{\rm bc}\int_{t=0}^{T}dt\{\phi(0,t)-B(t)\}^2+\\ &\quad W_{\rm ob}\sum_{i=1}^{p}\{\phi(x_i,t_i)-y_i\}^2+2\int_{x=0}^{L}dx\int_{t=0}^{T}dt\,\lambda(x,t)\left(\frac{\partial\phi}{\partial t}+u\frac{\partial\phi}{\partial x}-F\right). \end{aligned} \tag{5}$$

This is standard variational calculus (see, e.g., the Aide Memoir handout). The problem now is to minimize (5) w.r.t. the fields $\phi(x,t)$ and $\lambda(x,t)$.

Variations of J (strong constraint formulation) Construct variations of J about some reference fields $\hat{\phi}$, $\hat{\lambda}$, i.e. $J[\hat{\phi}+\delta\phi,\hat{\lambda}+\delta\lambda]=J[\hat{\phi},\hat{\lambda}]+\delta J|_{\hat{\phi},\hat{\lambda}}$, where:

$$\delta J|_{\hat{\phi},\hat{\lambda}} = \int_{x=0}^{L}dx\int_{t=0}^{T}dt\,\frac{\partial J}{\partial\phi}|_{\hat{\phi},\hat{\lambda}}\delta\phi+\int_{x=0}^{L}dx\int_{t=0}^{T}dt\,\frac{\partial J}{\partial\lambda}|_{\hat{\phi},\hat{\lambda}}\delta\lambda+O(\delta\phi^2,\delta\lambda^2,\delta\phi\delta\lambda),$$

$$
\begin{aligned}
&= 2W_{\mathrm{ic}} \int\limits_{x=0}^{L} dx \{\hat{\phi}(x,0) - I(x)\}\delta\phi(x,0) + 2W_{\mathrm{bc}} \int\limits_{t=0}^{T} dt \{\hat{\phi}(0,t) - B(t)\}\delta\phi(0,t) + \\
&\quad 2W_{\mathrm{ob}} \sum_{i=1}^{p} \{\hat{\phi}(x_i,t_i) - y_i\}\delta\phi(x_i,t_i) + \\
(*) \qquad &\quad 2 \int\limits_{x=0}^{L} dx \int\limits_{t=0}^{T} dt\, \hat{\lambda}(x,t) \left(\frac{\partial \delta\phi}{\partial t} + u \frac{\partial \delta\phi}{\partial x} \right) + \\
&\quad 2 \int\limits_{x=0}^{L} dx \int\limits_{t=0}^{T} dt\, \delta\lambda(x,t) \left(\frac{\partial \hat{\phi}}{\partial t} + u \frac{\partial \hat{\phi}}{\partial x} - F \right) + O(\delta\phi^2, \delta\lambda^2, \delta\phi\delta\lambda). \qquad (6)
\end{aligned}
$$

We impose the following conditions on $\hat{\lambda}$: $\hat{\lambda}(x,T) = 0$, $\hat{\lambda}(L,t) = 0$.

Changing form (strong constraint formulation) We would like to separate terms to do with $\delta\phi$ at different positions and times. The line marked (*) in (6) is not in the required form as it involves *derivatives* of $\delta\phi$ in space and time. Use the integration by parts formula to rewrite. In generic form this is:

$$
\int\limits_{a}^{b} v \frac{du}{dx} dx = [uv]_a^b - \int\limits_{a}^{b} u \frac{dv}{dx} dx. \qquad (7)
$$

Using this to rewrite the first term in (*):

$$
\begin{aligned}
\int\limits_{t=0}^{T} dt\, \hat{\lambda}(x,t) \frac{\partial \delta\phi}{\partial t} &= [\delta\phi\hat{\lambda}]_0^T - \int\limits_{t=0}^{T} dt\, \delta\phi \frac{\partial \hat{\lambda}}{\partial t}, \\
&= \delta\phi(x,T)\hat{\lambda}(x,T) - \delta\phi(x,0)\hat{\lambda}(x,0) - \int\limits_{t=0}^{T} dt\, \delta\phi \frac{\partial \hat{\lambda}}{\partial t}, \qquad (8)
\end{aligned}
$$

and the second term in (*):

$$\begin{aligned}\int\limits_{x=0}^{L} dx\,\hat{\lambda}(x,t)u\frac{\partial\delta\phi}{\partial x} &= u[\delta\phi\hat{\lambda}]_0^L - \int\limits_{x=0}^{L} dx\,u\delta\phi\frac{\partial\hat{\lambda}}{\partial x},\\ &= u\delta\phi(L,t)\hat{\lambda}(L,t) - u\delta\phi(0,t)\hat{\lambda}(0,t) - \int\limits_{x=0}^{L} dx\,u\delta\phi\frac{\partial\hat{\lambda}}{\partial x}.\end{aligned} \tag{9}$$

Note also for the observation term:

$$\{\hat{\phi}(x_i,t_i) - y_i\}\delta\phi(x_i,t_i) = \int\limits_{x=0}^{L} dx \int\limits_{t=0}^{T} dt\,\{\hat{\phi}(x_i,t_i) - y_i\}\delta\phi(x,t)\delta(x-x_i)\delta(t-t_i). \tag{10}$$

Using (8), (9) and (10) in (11) and noting the conditions on $\hat{\lambda}$ given after (11):

$$\begin{aligned}\delta J|_{\hat{\phi},\hat{\lambda}} = {} & 2W_{\rm ic}\int\limits_{x=0}^{L} dx\{\hat{\phi}(x,0) - I(x)\}\delta\phi(x,0) + 2W_{\rm bc}\int\limits_{t=0}^{T} dt\{\hat{\phi}(0,t) - B(t)\}\delta\phi(0,t) + \\ & 2W_{\rm ob}\int\limits_{x=0}^{L} dx \int\limits_{t=0}^{T} dt \sum_{i=1}^{p}\{\hat{\phi}(x_i,t_i) - y_i\}\delta\phi(x,t)\delta(x-x_i)\delta(t-t_i) - \\ & 2\int\limits_{x=0}^{L} dx\,\hat{\lambda}(x,0)\delta\phi(x,0) - 2\int\limits_{x=0}^{L} dx \int\limits_{t=0}^{T} dt\,\frac{\partial\hat{\lambda}}{\partial t}\delta\phi(x,t) - \\ & 2\int\limits_{t=0}^{T} dt\,u\hat{\lambda}(0,t)\delta\phi(0,t) - 2\int\limits_{t=0}^{T} dt \int\limits_{x=0}^{L} dx\,u\frac{\partial\hat{\lambda}}{\partial x}\delta\phi(x,t) + \\ & 2\int\limits_{x=0}^{L} dx \int\limits_{t=0}^{T} dt\,\left(\frac{\partial\hat{\phi}}{\partial t} + u\frac{\partial\hat{\phi}}{\partial x} - F\right)\delta\lambda(x,t) + O(\delta\phi^2,\delta\lambda^2,\delta\phi\delta\lambda).\end{aligned} \tag{11}$$

The Euler-Lagrange equations for the strong constraint formulation Setting the linear part of (11) to zero, and using the conditions on $\hat{\lambda}$, gives Euler-Lagrange equations for the strong constraint:

$$
\begin{aligned}
\frac{\partial \hat{\phi}}{\partial t} + u\frac{\partial \hat{\phi}}{\partial x} - F &= 0, && (12)\\
W_{\mathrm{ic}}\{\hat{\phi}(x,0) - I(x)\} - \hat{\lambda}(x,0) &= 0, && (13)\\
W_{\mathrm{bc}}\{\hat{\phi}(0,t) - B(t)\} - u\hat{\lambda}(0,t) &= 0, && (14)\\
W_{\mathrm{ob}}\sum_{i=1}^{p}\{\hat{\phi}(x_i,t_i) - y_i\}\delta(x-x_i)\delta(t-t_i) - \left(\frac{\partial \hat{\lambda}}{\partial t} + u\frac{\partial \hat{\lambda}}{\partial x}\right) &= 0, && (15)\\
\hat{\lambda}(x,T) &= 0, && (16)\\
\hat{\lambda}(L,t) &= 0. && (17)
\end{aligned}
$$

(12) is known as the forward equation, and (13)/(14) are its initial/boundary conditions. (15) is known as the backward equation, and (16)/(17) are its conditions. The solution to these Euler-Lagrange equations for $\hat{\phi}$ solves the original problem that we posed in Section 1 (with the assumption of a perfect model).

The weak constraint formulation

The weak constraint formulation imposes the known parts of the system equations approximately (that is we acknowledge that there is unknown model error).

$$
\begin{aligned}
J[\phi] \;=\; & W_{\mathrm{ic}}\int_{x=0}^{L} dx\{\phi(x,0) - I(x)\}^2 + W_{\mathrm{bc}}\int_{t=0}^{T} dt\{\phi(0,t) - B(t)\}^2 + \\
& W_{\mathrm{ob}}\sum_{i=1}^{p}\{\phi(x_i,t_i) - y_i\}^2 + W_{\mathrm{e}}\int_{x=0}^{L} dx\int_{t=0}^{T} dt\left\{\frac{\partial \phi}{\partial t} + u\frac{\partial \phi}{\partial x} - F\right\}^2. \qquad (18)
\end{aligned}
$$

Variations of J (weak constraint formulation) Construct variations of J about some reference field $\hat{\phi}$, i.e. $J[\hat{\phi}+\delta\phi] = J[\hat{\phi}] + \delta J|_{\hat{\phi}}$, where:

$$\begin{aligned}
\delta J|_{\hat{\phi}} &= \int_{x=0}^{L} dx \int_{t=0}^{T} dt\, \frac{\partial J}{\partial \phi}|_{\hat{\phi}} \delta\phi + O(\delta\phi^2), \\
&= 2W_{\rm ic} \int_{x=0}^{L} dx \{\hat{\phi}(x,0) - I(x)\} \delta\phi(x,0) + 2W_{\rm bc} \int_{t=0}^{T} dt \{\hat{\phi}(0,t) - B(t)\} \delta\phi(0,t) + \\
&\quad 2W_{\rm ob} \int_{x=0}^{L} dx \int_{t=0}^{T} dt \sum_{i=1}^{p} \{\hat{\phi}(x_i,t_i) - y_i\} \delta\phi(x,t) \delta(x-x_i) \delta(t-t_i) + \\
(*) &\quad 2W_{\rm e} \int_{x=0}^{L} dx \int_{t=0}^{T} dt \left\{ \frac{\partial \hat{\phi}}{\partial t} + u \frac{\partial \hat{\phi}}{\partial x} - F \right\} \left\{ \frac{\partial \delta\phi}{\partial t} + u \frac{\partial \delta\phi}{\partial x} \right\} + O(\delta\phi^2). \qquad (19)
\end{aligned}$$

Define:

$$\hat{\mu}(x,t) = W_{\rm e} \left(\frac{\partial \hat{\phi}}{\partial t} + u \frac{\partial \hat{\phi}}{\partial x} - F \right), \qquad (20)$$

(like the $\lambda(x,t)$ in (5)) where $\hat{\mu}(x,T) = 0$, $\hat{\mu}(L,t) = 0$.

Changing form (weak constraint formulation) Note that (*) of (21) is like (*) of (6) so change the form using the integration by parts formula (7), making (21) into:

$$\begin{aligned}
\delta J|_{\hat{\phi}} &= 2W_{\rm ic} \int_{x=0}^{L} dx \{\hat{\phi}(x,0) - I(x)\} \delta\phi(x,0) + 2W_{\rm bc} \int_{t=0}^{T} dt \{\hat{\phi}(0,t) - B(t)\} \delta\phi(0,t) + \\
&\quad 2W_{\rm ob} \int_{x=0}^{L} dx \int_{t=0}^{T} dt \sum_{i=1}^{p} \{\hat{\phi}(x_i,t_i) - y_i\} \delta\phi(x,t) \delta(x-x_i) \delta(t-t_i) -
\end{aligned}$$

$$2\int_{x=0}^{L} dx\, \hat{\mu}(x,0)\delta\phi(x,0) - 2\int_{x=0}^{L} dx \int_{t=0}^{T} dt\, \frac{\partial \hat{\mu}}{\partial t}\delta\phi(x,t) -$$
$$2\int_{t=0}^{T} dt\, u\hat{\mu}(0,t)\delta\phi(0,t) - 2\int_{t=0}^{T} dt \int_{x=0}^{L} dx\, u\frac{\partial \hat{\mu}}{\partial x}\delta\phi(x,t) + O(\delta\phi^2). \quad (21)$$

The Euler-Lagrange equations for the weak constraint formulation Setting the linear part of (21) to zero, using the model equation (1), and definition (21) gives Euler-Lagrange equations for the weak constraint:

$$\frac{\partial \hat{\phi}}{\partial t} + u\frac{\partial \hat{\phi}}{\partial x} - F = W_{\rm e}^{-1}\hat{\mu}, \quad (22)$$
$$W_{\rm ic}\{\hat{\phi}(x,0) - I(x)\} - \hat{\mu}(x,0) = 0, \quad (23)$$
$$W_{\rm bc}\{\hat{\phi}(0,t) - B(t)\} - u\hat{\mu}(0,t) = 0, \quad (24)$$
$$W_{\rm ob}\sum_{i=1}^{p}\{\hat{\phi}(x_i,t_i) - y_i\}\delta(x-x_i)\delta(t-t_i) - \left(\frac{\partial \hat{\mu}}{\partial t} + u\frac{\partial \hat{\mu}}{\partial x}\right) = 0, \quad (25)$$
$$\hat{\mu}(x,T) = 0, \quad (26)$$
$$\hat{\mu}(L,t) = 0. \quad (27)$$

Solving the weak-constraint Euler-Lagrange equations using the method of representers

The forward equation (22) is solved for $\hat{\phi}(x,t)$ 'upwards and to the right' (since the conditions for $\hat{\phi}$ are given for $x = 0$ and $t = 0$, see Fig.), and the backward equation (25) is solved for $\hat{\mu}(x,t)$ 'downwards and to the left' (since the conditions for $\hat{\phi}$ are given for $x = L$ and $t = T$, see Fig.).

Problem: In order to solve (22) for $\hat{\phi}(x,t)$, $\hat{\mu}(x,t)$ is needed, but in order to solve (25) for $\hat{\mu}(x,t)$, $\hat{\phi}(x,t)$ is needed! The set of Euler-Lagrange equations must be all solved together.

Recipe for the solution using the method of representers

1. Solve the background problem (2) for $\phi_{\rm B}(x,t)$. This is an exercise in solving partial differential equations (PDEs) analytically or numerically.

2. Define the p forward representer functions and the p backward representer functions (one each per observation) as:

$$\left.\begin{array}{ll}\text{Forward representer function} & r_i(x,t) \\ \text{Backward representer function} & \alpha_i(x,t)\end{array}\right\} 1 \leq i \leq p.$$

   The modified equations that these representers satisfy are based on the Euler-Lagrange equations, but have $\hat{\phi} \to r_i$, $\hat{\mu} \to \alpha_i$, $F = 0$, $I(x) = 0$, $B(t) = 0$ and replace the observations with a single impulse at the position and time of the ith observation $(W_{\rm ob} \sum_{i=1}^{p} \{\hat{\phi}(x_i,t_i) - y_i\} \delta(x - x_i)\delta(t - t_i) \to \delta(x - x_i)\delta(t - t_i))$.

$$\begin{aligned}
\frac{\partial r_i}{\partial t} + u\frac{\partial r_i}{\partial x} &= W_{\rm e}^{-1}\alpha_i, & (28)\\
W_{\rm ic} r_i(x,0) - \alpha_i(x,0) &= 0, & (29)\\
W_{\rm bc} r_i(0,t) - u\alpha_i(0,t) &= 0, & (30)\\
\delta(x - x_i)\delta(t - t_i) - \left(\frac{\partial \alpha_i}{\partial t} + u\frac{\partial \alpha_i}{\partial x}\right) &= 0, & (31)\\
\alpha_i(x,T) &= 0, & (32)\\
\alpha_i(L,t) &= 0. & (33)
\end{aligned}$$

3. Start with the backward representers. Solve (31), (32) and (33) for each i 'downwards and to the left' (again an exercise in solving PDEs). This gives the p backward representers, $\alpha_i(x,t)$. In the modified equations, the backward representers do not depend upon the forward representers, $r_i(x,t)$.

4. Now find the forward representers. Solve (28), (29) and (30) for each i 'upwards and to the right' (again an exercise in solving PDEs). This gives the p forward representers, $r_i(x,t)$, which can be found because the α_i are now known.

5. Look for a solution of $\hat{\phi}(x,t)$ (the field that we are really interested in) that is a linear combination of the forward representer functions:

$$\hat{\phi}(x,t) = \phi_{\rm B}(x,t) + \sum_{i=1}^{p} \beta_i r_i(x,t), \qquad (34)$$

where the β_i are the coefficients which are determined by insisting that $\hat{\phi}(x,t)$ satisfies the Euler-Lagrange equations.

6. To make (34) satisfy the Euler-Lagrange equations, act with $\partial/\partial t + u\partial/\partial x$ on (34), then use (22), (2) and (31):

$$\begin{aligned} \frac{\partial\hat{\phi}}{\partial t} + u\frac{\partial\hat{\phi}}{\partial x} &= \frac{\partial\phi_{\mathrm{B}}}{\partial t} + u\frac{\partial\phi_{\mathrm{B}}}{\partial x} + \sum_{i=1}^{p}\beta_i\left(\frac{\partial r_i}{\partial t} + u\frac{\partial r_i}{\partial x}\right), \\ \Rightarrow F + W_{\mathrm{e}}^{-1}\hat{\mu} &= F + \sum_{i=1}^{p}\beta_i W_{\mathrm{e}}^{-1}\alpha_i, \\ \Rightarrow \hat{\mu}(x,t) &= \sum_{i=1}^{p}\beta_i\alpha_i(x,t). \end{aligned} \tag{35}$$

7. Substitute (35) into (25), then use (34) and (31):

$$\begin{aligned} W_{\mathrm{ob}}\sum_{i=1}^{p}\{\hat{\phi}(x_i,t_i) - y_i\}\delta(x-x_i)\delta(t-t_i) &= \sum_{i=1}^{p}\beta_i\left(\frac{\partial\alpha_i}{\partial t} + u\frac{\partial\alpha_i}{\partial x}\right), \\ \Rightarrow W_{\mathrm{ob}}\sum_{i=1}^{p}\{\phi_{\mathrm{B}}(x_i,t_i) + & \\ \sum_{j=1}^{p}\beta_j r_j(x_i,t_i) - y_i\}\delta(x-x_i)\delta(t-t_i) &= \sum_{i=1}^{p}\beta_i\delta(x-x_i)\delta(t-t_i). \end{aligned} \tag{36}$$

8. Equate coefficients of impulses in (36):

$$\begin{aligned} W_{\mathrm{ob}}\left\{\phi_{\mathrm{B}}(x_i,t_i) + \sum_{j=1}^{p}\beta_j r_j(x_i,t_i) - y_i\right\} &= \beta_i, \\ \Rightarrow W_{\mathrm{ob}}\left\{\phi_{\mathrm{B}}(x_i,t_i) - y_i\right\} + \sum_{j=1}^{p}\left\{W_{\mathrm{ob}} r_j(x_i,t_i) - \delta_{ij}\right\}\beta_j &= 0, \end{aligned} \tag{37}$$

where δ_{ij} is the Kronecker delta-function. This is the equation that we have to solve for the β_i coefficients. Once these are known, the solution can be built using (34).

Finding the coefficients Equation (37) is the remaining equation to solve. We will use some linear algebra (vectors and matrices) to do this. This is a standard procedure in a wide range of numerical analysis problems. Let the vectors $\mathbf{y} \in \mathbb{R}^p$, $\boldsymbol{\beta} \in \mathbb{R}^p$ and $\boldsymbol{\phi}_{\mathrm{B}}^{\mathrm{ob}} \in \mathbb{R}^p$ (bold symbols) represent the following collections of information:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \qquad \boldsymbol{\phi}_{\mathrm{B}}^{\mathrm{ob}} = \begin{pmatrix} \phi_{\mathrm{B}}(x_1, t_1) \\ \phi_{\mathrm{B}}(x_2, t_2) \\ \vdots \\ \phi_{\mathrm{B}}(x_p, t_p) \end{pmatrix}.$$

These represent (respectively) the observations, the (as yet) unknown coefficients that we are trying to find and the background values at the observation positions and times. The equations represented by (37) $(1 \leq i \leq p)$ may be written in linear algebraic form:

$$W_{\mathrm{ob}} \left(\boldsymbol{\phi}_{\mathrm{B}}^{\mathrm{ob}} - \mathbf{y} \right) + \left(W_{\mathrm{ob}} \mathbf{P} - \mathbf{I} \right) \boldsymbol{\beta} = 0,$$

where $\mathbf{P} \in \mathbb{R}^{p \times p}$ is

$$\mathbf{P} = \begin{pmatrix} r_1(x_1, t_1) & r_2(x_1, t_1) & \cdots & r_p(x_1, t_1) \\ r_1(x_2, t_2) & r_2(x_2, t_2) & \cdots & r_p(x_2, t_2) \\ \vdots & \vdots & \ddots & \vdots \\ r_1(x_p, t_p) & r_2(x_p, t_p) & \cdots & r_p(x_p, t_p) \end{pmatrix},$$

and $\mathbf{I} \in \mathbb{R}^{p \times p}$ is the identity matrix. All of these vectors and matrices are known except for $\boldsymbol{\beta}$. Providing that the matrix $W_{\mathrm{ob}} \mathbf{P} - \mathbf{I}$ is full rank, then the solution is found to be

$$\boldsymbol{\beta} = W_{\mathrm{ob}} \left(W_{\mathrm{ob}} \mathbf{P} - \mathbf{I} \right)^{-1} \left(\mathbf{y} - \boldsymbol{\phi}_{\mathrm{B}}^{\mathrm{ob}} \right).$$

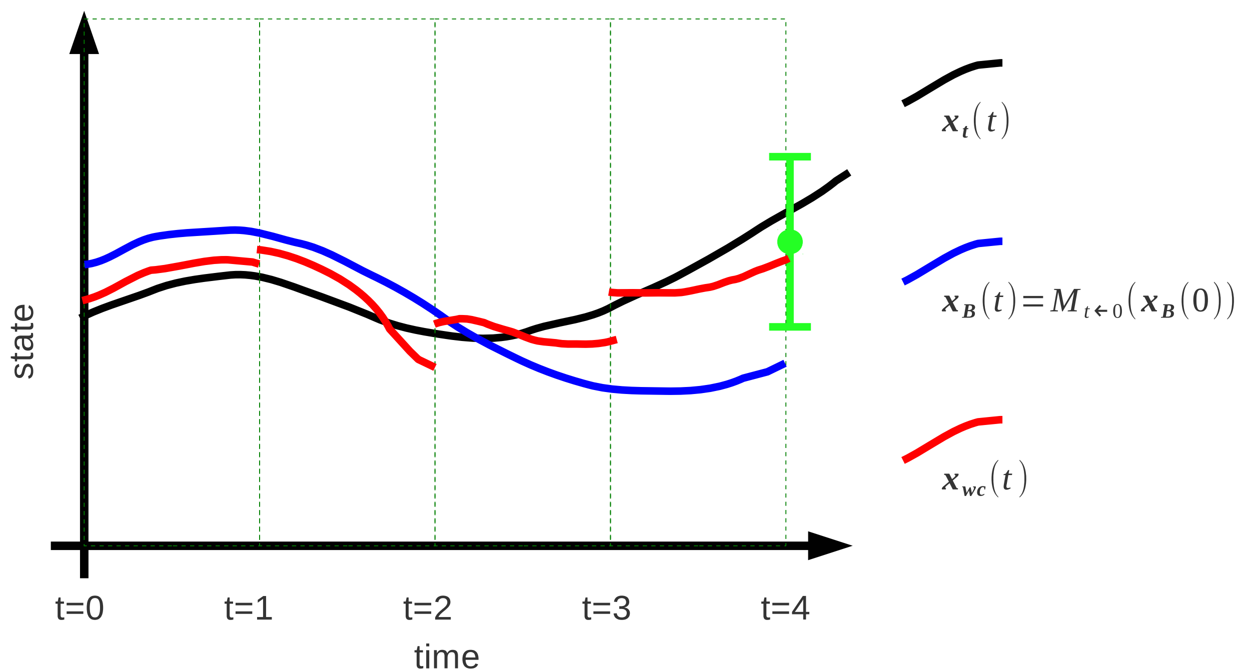
2(c) Cost functions and simplifications for operational assimilation

Weak constraint 4D-VAR

$$
\begin{aligned}
J_{\mathrm{wc}}[\mathbf{x}] \;=\; & \frac{1}{2}\left[\mathbf{x}(0)-\mathbf{x}_{\mathrm{B}}(0)\right]^{\mathrm{T}} \mathbf{P}^{\mathrm{f}-1}\left[\mathbf{x}(0)-\mathbf{x}_{\mathrm{B}}(0)\right]+ \\
& \frac{1}{2}\sum_{t=0}^{T}\left[\mathbf{y}(t)-\mathbf{h}_t\left(\mathbf{x}(t)\right)\right]^{\mathrm{T}} \mathbf{R}_t^{-1}\left[\mathbf{y}(t)-\mathbf{h}_t\left(\mathbf{x}(t)\right)\right]+ \\
& \frac{1}{2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\left[\mathbf{x}(t)-\boldsymbol{\mathcal{M}}_{t\leftarrow t-1}\left(\mathbf{x}(t-1)\right)\right]\left(\mathbf{Q}^{-1}\right)_{tt'}\left[\mathbf{x}(t')-\boldsymbol{\mathcal{M}}_{t'\leftarrow t'-1}\left(\mathbf{x}(t'-1)\right)\right].
\end{aligned}
$$

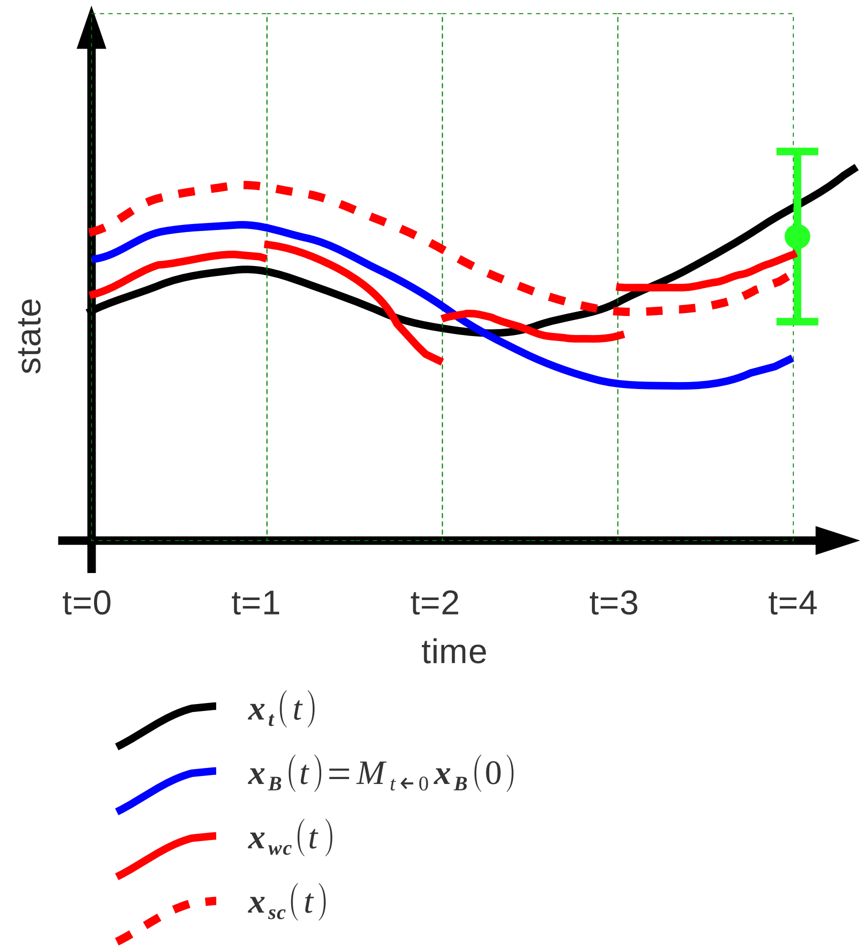
Here $\mathbf{x}$ is called the control variable and is the 4D state vector:

$$
\boldsymbol{x}=\begin{pmatrix} \boldsymbol{x}(0) \\ \boldsymbol{x}(1) \\ \vdots \\ \boldsymbol{x}(T) \end{pmatrix}
$$



Simplification 3: Assume that the numerical model is perfect (strong constraint 4D-VAR)

$$
\begin{aligned}
J_{\mathrm{sc}}[\mathbf{x}] &= \frac{1}{2}(\mathbf{x}-\mathbf{x}_{\mathrm{B}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_{\mathrm{B}}) + \\
&\quad \frac{1}{2}\sum_{t=0}^{T}\left[\mathbf{y}(t)-\mathbf{h}_t\left(\boldsymbol{\mathcal{M}}_{t\leftarrow 0}(\mathbf{x})\right)\right]^{\mathrm{T}}\mathbf{R}_t^{-1}\times \\
&\qquad\qquad \left[\mathbf{y}(t)-\mathbf{h}_t\left(\boldsymbol{M}_{t\leftarrow 0}(\mathbf{x})\right)\right], \\
\boldsymbol{\mathcal{M}}_{t\leftarrow 0}(\mathbf{x}) &= \begin{cases} \boldsymbol{\mathcal{M}}_{t\leftarrow t-1}\left(\cdots\boldsymbol{\mathcal{M}}_{2\leftarrow 1}\left(\boldsymbol{\mathcal{M}}_{1\leftarrow 0}(\mathbf{x})\right)\right) & t>0 \\ \mathbf{I} & t=0 \end{cases}.
\end{aligned}
$$



This is equivalent to making $\mathbf{Q}_{tt}\to 0$ in the weak constraint cost function.

Simplification 4: Incremental data assimilation

$$\mathbf{x}(t) = \mathbf{x}_k^{\mathrm{ref}}(t) + \delta\mathbf{x}(t).$$

Linearizing the forecast model:

$$\begin{aligned}
\mathbf{x}(t) &= \boldsymbol{\mathcal{M}}_{t\leftarrow t-1}(\mathbf{x}(t-1)), \\
\mathbf{x}_k^{\mathrm{ref}}(t) + \delta\mathbf{x}(t) &= \boldsymbol{\mathcal{M}}_{t\leftarrow t-1}\left(\mathbf{x}_k^{\mathrm{ref}}(t-1) + \delta\mathbf{x}(t-1)\right), \\
&\simeq \boldsymbol{\mathcal{M}}_{t\leftarrow t-1}\left(\mathbf{x}_k^{\mathrm{ref}}(t-1)\right) + \mathbf{M}_{t\leftarrow t-1}\delta\mathbf{x}(t-1), \\
\delta\mathbf{x}(t) &= \mathbf{M}_{t\leftarrow t-1}\delta\mathbf{x}(t-1),
\end{aligned}$$

where the reference state

$$\begin{aligned}
\mathbf{x}_k^{\mathrm{ref}}(t) &\equiv \boldsymbol{\mathcal{M}}_{t\leftarrow t-1}\left(\mathbf{x}_k^{\mathrm{ref}}(t-1)\right), \\
\text{and } \mathbf{M}_{t\leftarrow t-1} &\equiv \left.\frac{\partial \boldsymbol{\mathcal{M}}_{t\leftarrow t-1}\left(\mathbf{x}(t-1)\right)}{\partial \mathbf{x}(t-1)}\right|_{\mathbf{x}_k^{\mathrm{ref}}} \in \mathbb{R}^{n\times n},
\end{aligned}$$

with matrix elements

$$\{\mathbf{M}_{t\leftarrow t-1}\}_{ij} = \left.\frac{\partial\left\{\boldsymbol{\mathcal{M}}_{t\leftarrow t-1}\left(\mathbf{x}(t-1)\right)\right\}_i}{\partial\left\{\mathbf{x}(t-1)\right\}_j}\right|_{\mathbf{x}_k^{\mathrm{ref}}}.$$

Linearizing the observation operator:

$$\begin{aligned}
\mathbf{y}^{\mathrm{mo}}(t) &= \mathbf{h}_t(\mathbf{x}(t)), \\
&= \mathbf{h}_t\left(\mathbf{x}_k^{\mathrm{ref}}(t) + \delta\mathbf{x}(t)\right), \\
&\simeq \mathbf{h}_t\left(\mathbf{x}_k^{\mathrm{ref}}(t)\right) + \mathbf{H}_t\delta\mathbf{x}(t), \\
&\simeq \mathbf{y}_{\mathrm{ref},k}^{\mathrm{mo}}(t) + \mathbf{H}_t\delta\mathbf{x}(t), \\
\delta\mathbf{y}^{\mathrm{mo}}(t) &= \mathbf{H}_t\delta\mathbf{x}(t),
\end{aligned}$$

$$\begin{aligned}
\text{where } \delta\mathbf{y}^{\mathrm{mo}}(t) &\equiv \mathbf{y}^{\mathrm{mo}}(t) - \mathbf{y}_{\mathrm{ref},k}^{\mathrm{mo}}(t), \\
\text{and } \mathbf{H}_t &\equiv \left.\frac{\partial \mathbf{h}_t(\mathbf{x}(t))}{\partial \mathbf{x}(t)}\right|_{\mathbf{x}_k^{\mathrm{ref}}} \in \mathbb{R}^{p\times n},
\end{aligned}$$

with matrix elements

$$\{\mathbf{H}_t\}_{ij} = \left.\frac{\partial\left\{\mathbf{h}_t(\mathbf{x}(t))\right\}_i}{\partial\left\{\mathbf{x}(t)\right\}_j}\right|_{\mathbf{x}_k^{\mathrm{ref}}}.$$

By writing the background as a perturbation with respect to the reference state, $\mathbf{x}_{\mathrm{B}}(t) \equiv \mathbf{x}_k^{\mathrm{ref}}(t) + \delta\mathbf{x}_{\mathrm{B}}(t)$, and defining $\delta\mathbf{y}(t) \equiv \mathbf{y}(t) - \mathbf{h}_t\left(\mathcal{M}_{t\leftarrow 0}(\mathbf{x}_k^{\mathrm{ref}})\right)$, the strong constraint cost function becomes:

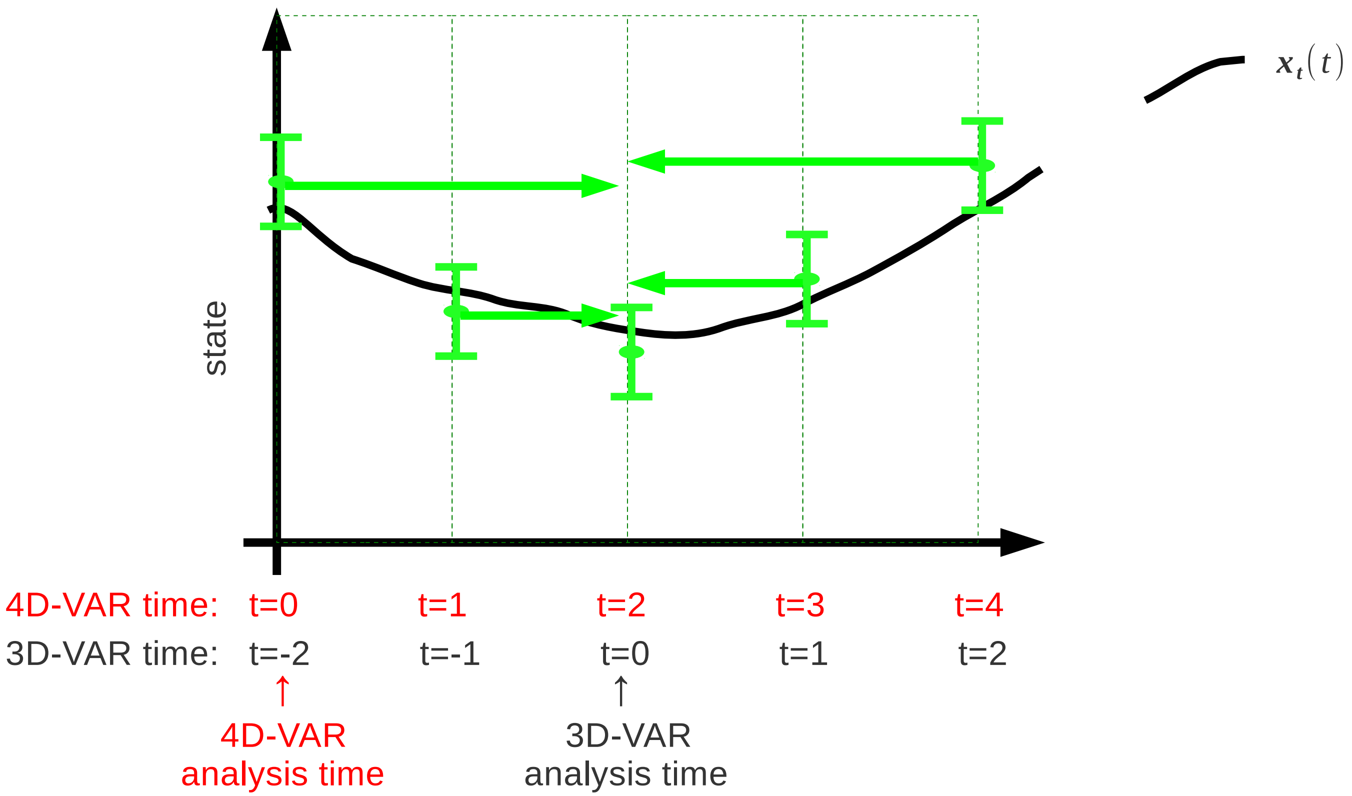
$$J_{\text{4Dinc}}[\delta\mathbf{x}] = \frac{1}{2}(\delta\mathbf{x} - \delta\mathbf{x}_{\text{B}})^{\text{T}}\mathbf{B}^{-1}(\delta\mathbf{x} - \delta\mathbf{x}_{\text{B}}) + \frac{1}{2}\sum_{t=0}^{T}\left[\delta\mathbf{y}(t) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\delta\mathbf{x}\right]^{\text{T}}\mathbf{R}_t^{-1}\left[\delta\mathbf{y}(t) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\delta\mathbf{x}\right].$$

- The control variable is $\delta\mathbf{x} = \delta\mathbf{x}(0)$ in this incremental formulation.
- Later we will call $\delta\mathbf{y}(t) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\delta\mathbf{x}$ the residual vector, $\mathbf{r}(t)$.
- $J_{\text{4Dinc}}[\delta\mathbf{x}]$ is exactly quadratic and so is easier to minimize than $J_{\text{4D}}[\delta\mathbf{x}]$.
- If the value of $\delta\mathbf{x}$ that minimizes this is $\delta\mathbf{x}_{\text{A}}$ ('inner loop'), then the analysis is

$$\mathbf{x}_{\text{A}} = \mathbf{x}_k^{\text{ref}} + \delta\mathbf{x}_{\text{A}}.$$

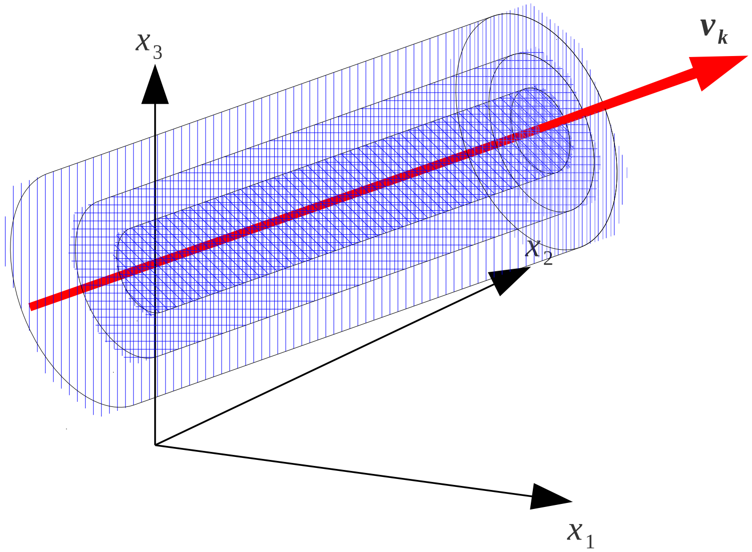
- Set $\mathbf{x}_{k+1}^{\text{ref}}(t) = \mathbf{x}_{\text{A}}$ and repeat ('outer loop').

Simplification 6: 3D-VAR



3. A-priori information and the B-matrix

3(a) The null space of the observation operator and the importance of a-priori information

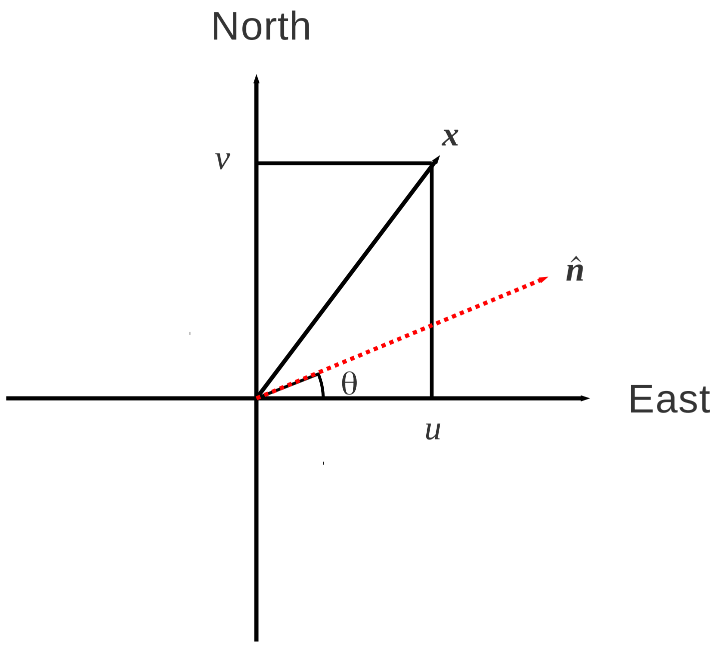


Physical example of an observation operator and null space

Let

$$\begin{aligned} \mathbf{x} &= \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \text{uniform zonal wind} \ {}^{-}\leftrightarrow^{+} \\ \text{uniform meridional wind} \ \updownarrow^{+}_{-} \end{pmatrix}, \\ \mathbf{y} &= \text{measurement of wind component in a direction } \theta \text{ from E}, \\ \sigma_{\mathbf{y}}^2 &= \text{Error variance of measurement.} \end{aligned}$$

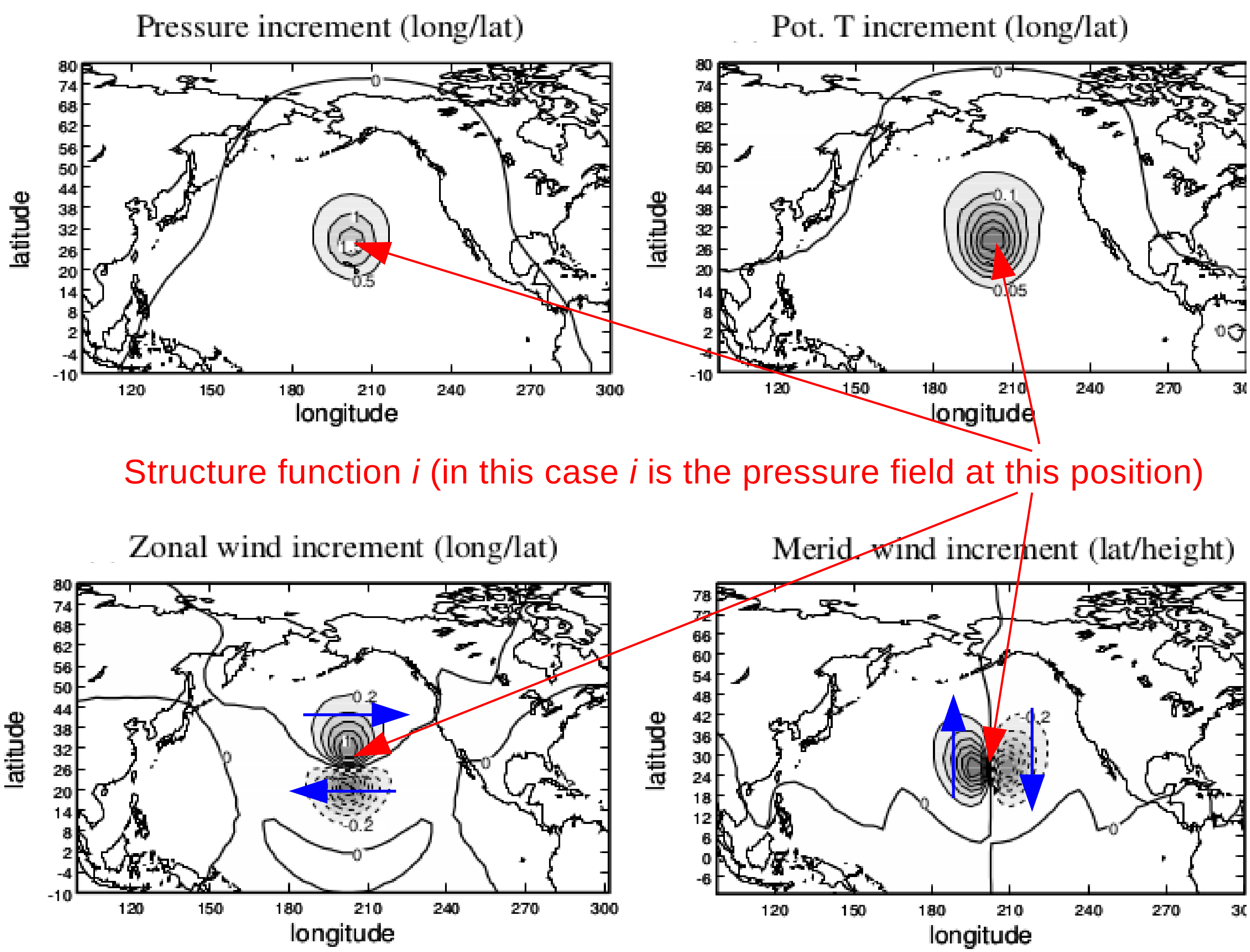
This measurement is given e.g. by a Doppler radar instrument.



$\hat{\mathbf{u}}$ is the unit vector in the line of sight of the radar beam,

$$\hat{\mathbf{u}} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}.$$

3(b) The role of the background error covariance matrix



In this case the wind part of the structure function is in geostrophic balance with the pressure

3(c) Spatial aspects (inverse Laplacians, diffusion operators)

Inverse Laplacians

Consider the following form of a $\mathbf{B}$-matrix for a single field (univariate):

$$\begin{aligned}
\mathbf{COR} &= \gamma \left(\mathbf{I} + \frac{l^4}{2} (\nabla^2)^2 \right)^{-1}, \\
\mathbf{B} = \mathbf{COV} &= \mathbf{\Sigma} \gamma \left(\mathbf{I} + \frac{l^4}{2} (\nabla^2)^2 \right)^{-1} \mathbf{\Sigma}, \\
\mathbf{B}^{-1} &= \gamma^{-1} \mathbf{\Sigma}^{-1} \left(\mathbf{I} + \frac{l^4}{2} (\nabla^2)^2 \right) \mathbf{\Sigma}^{-1},
\end{aligned}$$

(where l is the (chosen) correlation length-scale). What is the result of acting with $\mathbf{COR}$ on the arbitrary function $f(x)$ in 1-D?

$$\text{Let } g(x) = \mathbf{COR}\{f(x)\} = \gamma \left(1 + \frac{l^4}{2} \frac{d^4}{dx^4} \right)^{-1} f(x).$$

This can be easily solved in Fourier space:

$$f(x) = \frac{1}{\sqrt{2\pi}} \int dk\, \bar{f}(k) e^{ikx} \qquad g(x) = \frac{1}{\sqrt{2\pi}} \int dk\, \bar{g}(k) e^{ikx},$$

$$f(x) = \gamma^{-1} \left(1 + \frac{l^4}{2} \frac{d^4}{dx^4} \right) g(x),$$

$$\begin{aligned}
\int dk\, \bar{f}(k) e^{ikx} &= \gamma^{-1} \left(1 + \frac{l^4}{2} \frac{d^4}{dx^4} \right) \int dk\, \bar{g}(k) e^{ikx}, \\
&= \int dk\, \bar{g}(k) \gamma^{-1} \left(1 + \frac{l^4 k^4}{2} \right) e^{ikx}.
\end{aligned}$$

Multiply each side by $e^{-ik'x}$, integrate over x, and use orthogonality of complex exponentials:

$$\bar{g}(k) = \gamma \left(1 + \frac{l^4 k^4}{2} \right)^{-1} \bar{f}(k).$$

Inverse Fourier transform this to get the result in x-space:

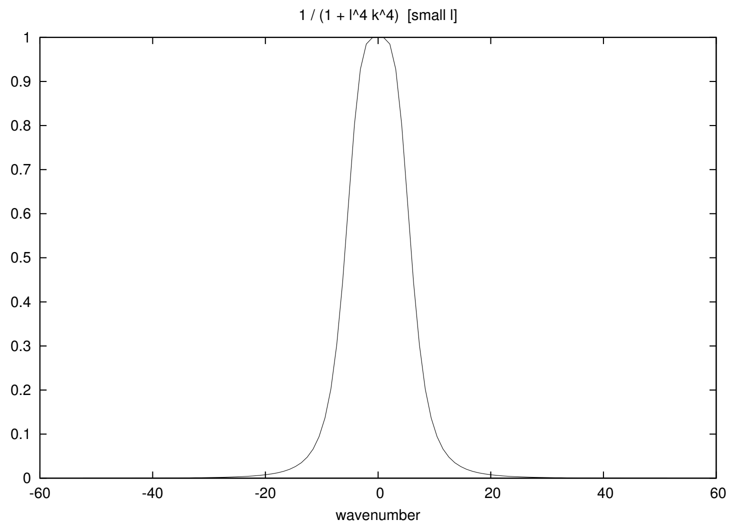
$$\begin{aligned}
g(x) &= \text{I.F.T.} \left\{ \gamma \left(1 + \frac{l^4 k^4}{2} \right)^{-1} \bar{f}(k) \right\}, \\
&= \text{I.F.T.} \left\{ \bar{c}(k) \bar{f}(k) \right\}, \\
&= \frac{1}{2\pi} \int dx'\, c(x - x') f(x'),
\end{aligned}$$

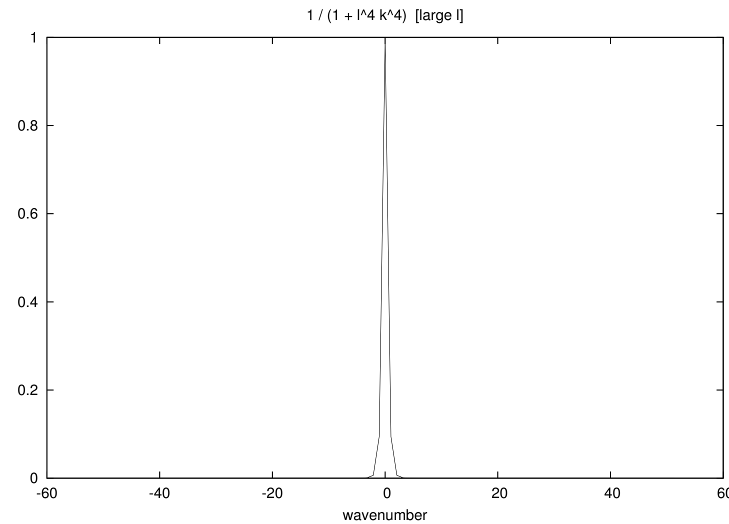
by the convolution theorem of Fourier transforms. $c(x)$ is the inverse Fourier transform of $\gamma/(1 + l^4 k^4/2)$.

SMALL LENGTHSCALE

LARGE LENGTHSCALE

SPECTRAL SPACE

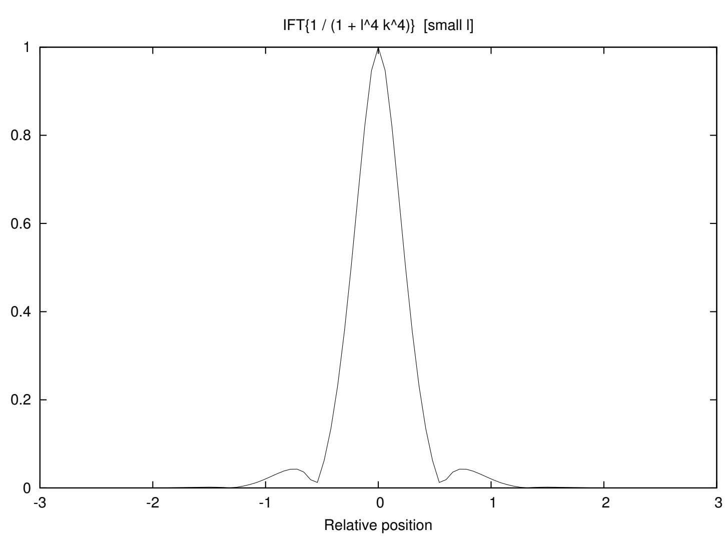


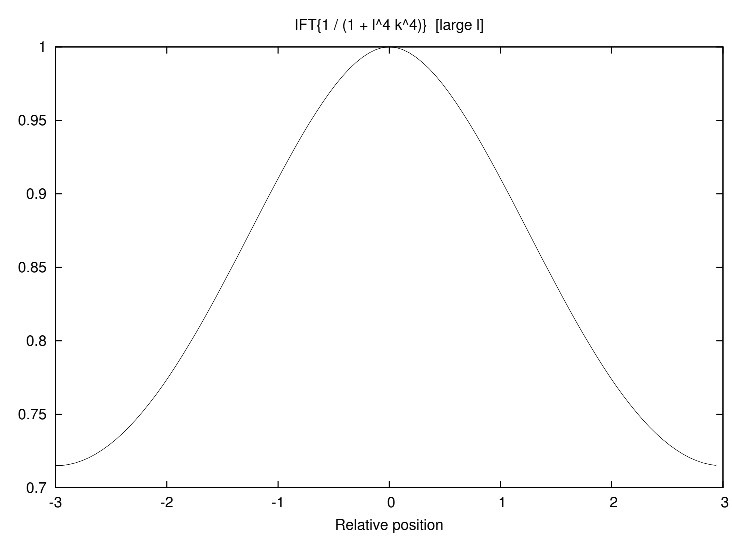


Filter away very large wavenumbers only
(keep all but smallest scales)

Filter away all except very small wavenumbers
(keep large−scales)

REAL SPACE





Diffusion operators

Consider the following diffusion equation for integration from $t = 0$ to T:

$$\frac{\partial g(x,t)}{\partial t} - \kappa \frac{\partial^2 g(x,t)}{\partial x^2} = 0,$$

κ : diffusion co-efficient, initial condition $g(x,0) = f(x)$.

The diffusion equation can be integrated analytically in Fourier space. For wavenumber k:

$$\frac{\partial \bar{g}(k,t)}{\partial t} + \kappa k^2 \bar{g}(k,t), \qquad \bar{g}(k,0) = \bar{f}(k).$$

Integrate from $t = 0$ to T:

$$\begin{aligned} \int_{t=0}^{T} d \ln \bar{g}(k,t) + \kappa k^2 \int_{t=0}^{T} dt &= 0, \\ \ln \bar{g}(k,T) - \ln \bar{g}(k,0) + \kappa k^2 T &= 0, \end{aligned}$$

$$\bar{g}(k,T) = \bar{f}(k) \exp(-\kappa k^2 T).$$

To find the solution in real space, inverse Fourier transform the above. The right hand side is a product of functions in Fourier space, so use the convolution theorem again:

$$g(x,T) = \frac{1}{2\pi} \int dx' \, f(x') \, c(x - x').$$

$c(x)$ is here the inverse Fourier transform of $\exp(-\kappa k^2 T)$, which is $\sqrt{\pi/\kappa T} \exp(-x^2/4\kappa T)$ (a Gaussian function with length-scale $\sqrt{2\kappa T}$). The solution is thus:

$$g(x,T) = \frac{1}{\sqrt{4\pi\kappa T}} \int dx' \, f(x') \exp(-(x - x')^2/4\kappa T).$$

Note the correspondence between the convolution and action with a homogeneous covariance matrix (as in the previous section on inverse Laplacians), which means that the structure functions have the form:

$$\frac{1}{\sqrt{4\pi\kappa T}} \exp(-(x - x')^2/4\kappa T).$$

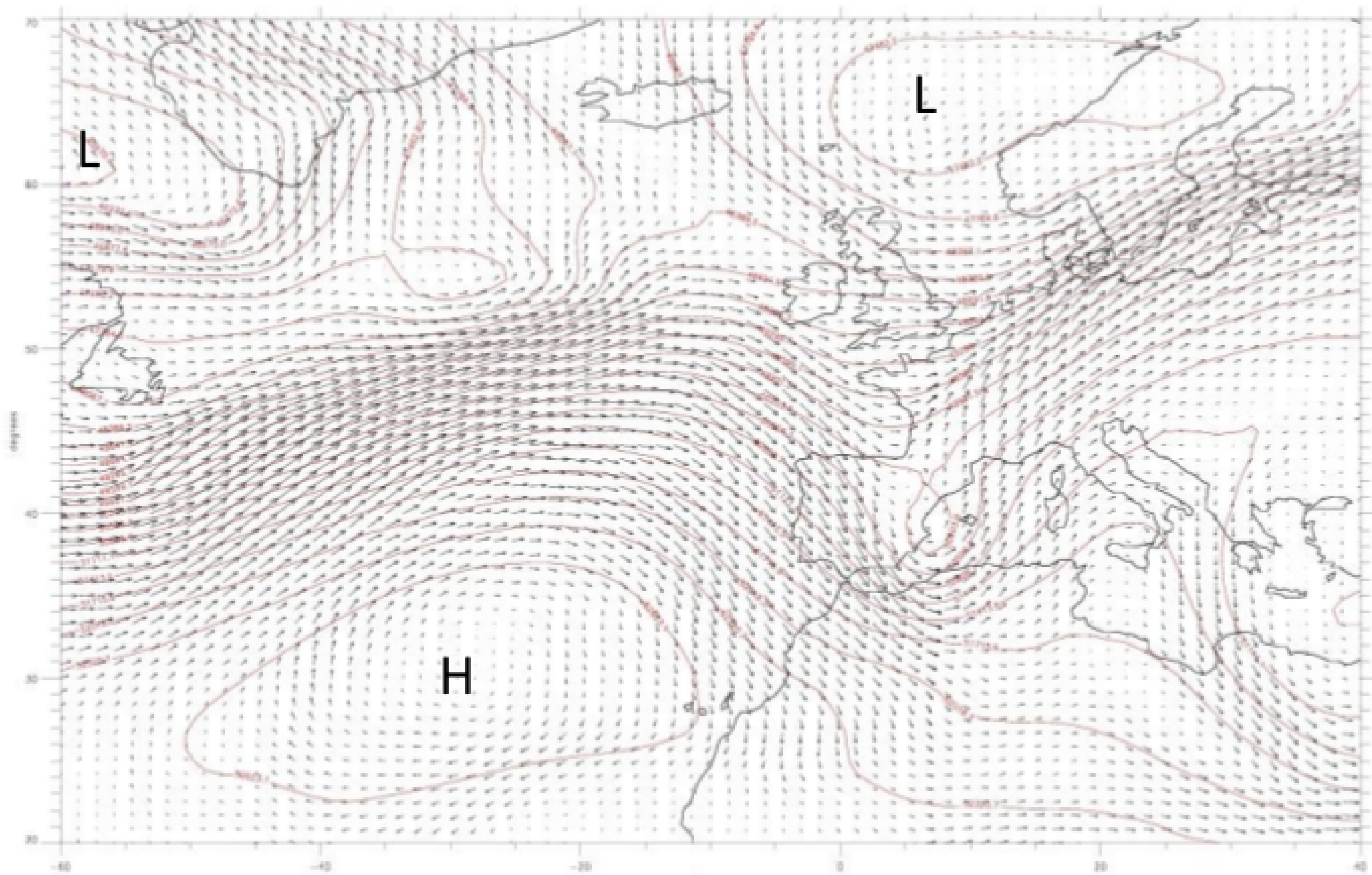
3(d) Multivariate aspects and balance

Example with perfect geostrophic balance

For flows with small Rossby number, $\mathrm{Ro} = U/fL \ll 1$, the momentum equations approximate to the following diagnostic equations:

$$v = \frac{1}{f\rho}\frac{\partial p}{\partial x}, \quad u = -\frac{1}{f\rho}\frac{\partial p}{\partial y},$$

(this is geostrophic balance).



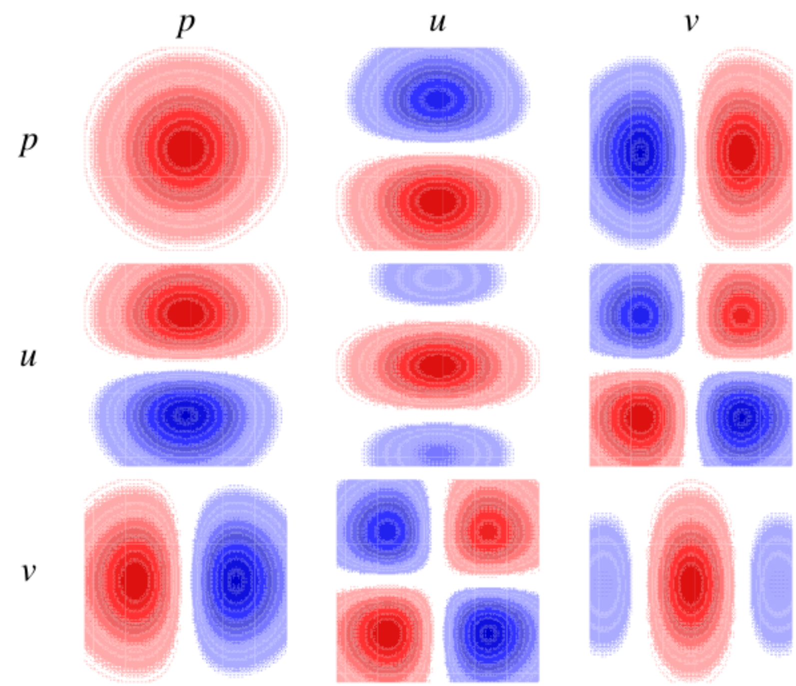
Now derive the multivariate error covariances between positions i and j:

$$
\begin{aligned}
p-p \text{ covs: } \langle \delta p_i \delta p_j \rangle &= \sigma_p^2 \mu_{ij} \text{ (by definition)},\\
p-u \text{ covs: } \langle \delta p_i \delta u_j \rangle &= -\frac{1}{f\rho}\left\langle \delta p_i \frac{\partial \delta p_j}{\partial y_j}\right\rangle = -\frac{1}{f\rho}\frac{\partial}{\partial y_j}\langle \delta p_i \delta p_j \rangle = -\frac{\sigma_p^2}{f\rho}\frac{\partial \mu_{ij}}{\partial y_j},\\
p-v \text{ covs: } \langle \delta p_i \delta v_j \rangle &= \frac{1}{f\rho}\left\langle \delta p_i \frac{\partial \delta p_j}{\partial x_j}\right\rangle = \frac{1}{f\rho}\frac{\partial}{\partial x_j}\langle \delta p_i \delta p_j \rangle = \frac{\sigma_p^2}{f\rho}\frac{\partial \mu_{ij}}{\partial x_j},\\
u-p \text{ covs: } \langle \delta u_i \delta p_j \rangle &= -\frac{1}{f\rho}\left\langle \frac{\partial \delta p_i}{\partial y_i}\delta p_j \right\rangle = -\frac{1}{f\rho}\frac{\partial}{\partial y_i}\langle \delta p_i \delta p_j \rangle = -\frac{\sigma_p^2}{f\rho}\frac{\partial \mu_{ij}}{\partial y_i},\\
u-u \text{ covs: } \langle \delta u_i \delta u_j \rangle &= \frac{1}{f^2\rho^2}\left\langle \frac{\partial \delta p_i}{\partial y_i}\frac{\partial \delta p_j}{\partial y_j}\right\rangle = \frac{1}{f^2\rho^2}\frac{\partial^2}{\partial y_i \partial y_j}\langle \delta p_i \delta p_j \rangle = \frac{\sigma_p^2}{f^2\rho^2}\frac{\partial^2 \mu_{ij}}{\partial y_i \partial y_j},\\
u-v \text{ covs: } \langle \delta u_i \delta v_j \rangle &= -\frac{1}{f^2\rho^2}\left\langle \frac{\partial \delta p_i}{\partial y_i}\frac{\partial \delta p_j}{\partial x_j}\right\rangle = -\frac{1}{f^2\rho^2}\frac{\partial^2}{\partial y_i \partial x_j}\langle \delta p_i \delta p_j \rangle = -\frac{\sigma_p^2}{f^2\rho^2}\frac{\partial^2 \mu_{ij}}{\partial y_i \partial x_j},\\
v-p \text{ covs: } \langle \delta v_i \delta p_j \rangle &= \frac{1}{f\rho}\left\langle \frac{\partial \delta p_i}{\partial x_i}\delta p_j \right\rangle = \frac{1}{f\rho}\frac{\partial}{\partial x_i}\langle \delta p_i \delta p_j \rangle = \frac{\sigma_p^2}{f\rho}\frac{\partial \mu_{ij}}{\partial x_i},\\
v-u \text{ covs: } \langle \delta v_i \delta u_j \rangle &= -\frac{1}{f^2\rho^2}\left\langle \frac{\partial \delta p_i}{\partial x_i}\frac{\partial \delta p_j}{\partial y_j}\right\rangle = -\frac{1}{f^2\rho^2}\frac{\partial^2}{\partial x_i \partial y_j}\langle \delta p_i \delta p_j \rangle = -\frac{\sigma_p^2}{f^2\rho^2}\frac{\partial^2 \mu_{ij}}{\partial x_i \partial y_j},\\
v-v \text{ covs: } \langle \delta v_i \delta v_j \rangle &= \frac{1}{f^2\rho^2}\left\langle \frac{\partial \delta p_i}{\partial x_i}\frac{\partial \delta p_j}{\partial x_j}\right\rangle = \frac{1}{f^2\rho^2}\frac{\partial^2}{\partial x_i \partial x_j}\langle \delta p_i \delta p_j \rangle = \frac{\sigma_p^2}{f^2\rho^2}\frac{\partial^2 \mu_{ij}}{\partial x_i \partial x_j}.
\end{aligned}
$$

Note the following first and second derivatives of μ:

$$
\begin{aligned}
\frac{\partial \mu_{ij}}{\partial x_i} &= -\mu_{ij}\frac{(x_i - x_j)}{L^2} \\
\frac{\partial \mu_{ij}}{\partial x_j} &= \mu_{ij}\frac{(x_i - x_j)}{L^2}, \\
\frac{\partial \mu_{ij}}{\partial y_i} &= -\mu_{ij}\frac{(y_i - y_j)}{L^2}, \\
\frac{\partial \mu_{ij}}{\partial y_j} &= \mu_{ij}\frac{(y_i - y_j)}{L^2}, \\
\frac{\partial^2 \mu_{ij}}{\partial x_i \partial x_j} &= \frac{\mu_{ij}}{L^2}\left(1 - \frac{(x_i - x_j)^2}{L^2}\right), \\
\frac{\partial^2 \mu_{ij}}{\partial y_i \partial y_j} &= \frac{\mu_{ij}}{L^2}\left(1 - \frac{(y_i - y_j)^2}{L^2}\right), \\
\frac{\partial^2 \mu_{ij}}{\partial y_i \partial x_j} &= -\mu_{ij}\frac{(x_i - x_j)(y_i - y_j)}{L^4}, \\
\frac{\partial^2 \mu_{ij}}{\partial x_i \partial y_j} &= -\mu_{ij}\frac{(x_i - x_j)(y_i - y_j)}{L^4}.
\end{aligned}
$$

Example structure functions giving the output field (p, u or v down the side) associated with a point in the centre of the domain (either of p, u or v along the top). Red is positive, blue is negative.



3(e) Control variable transforms and the implied B-matrix

Solving a variational problem using CVTs involves the following steps:

- Assume that we know the CVT, $\mathbf{U}$, and its adjoint and that they are practical to apply.
- Minimize $J[\delta\boldsymbol{\chi}]$ with respect to varying $\delta\boldsymbol{\chi}$. The cost function is:

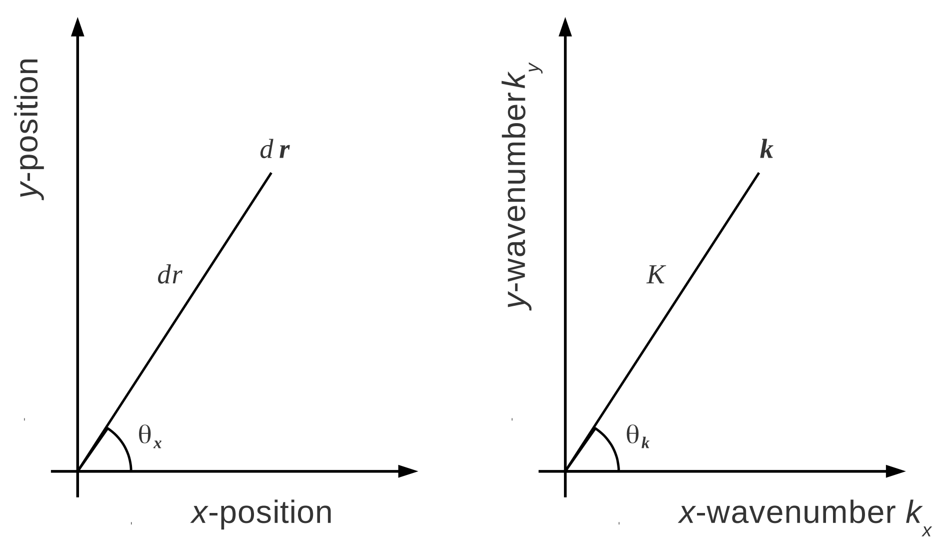
$$
\begin{aligned}
J[\delta\boldsymbol{\chi}] &= \frac{1}{2}\delta\boldsymbol{\chi}^{\mathrm{T}}\delta\boldsymbol{\chi} + \\
&\quad \frac{1}{2}\sum_{t=0}^{T}\left[\delta\mathbf{y}(t) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}\delta\boldsymbol{\chi}\right]^{\mathrm{T}}\mathbf{R}_t^{-1}\left[\delta\mathbf{y}(t) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}\delta\boldsymbol{\chi}\right].
\end{aligned}
$$

- The analysis increment in control variable space that minimizes the above is $\delta\boldsymbol{\chi}_{\mathrm{A}}$.
- The analysis in model space is $\mathbf{x}_{\mathrm{A}} = \mathbf{x}_{\mathrm{B}} + \mathbf{U}\delta\boldsymbol{\chi}_{\mathrm{A}}$.
- This is equivalent to minimizing the original cost function $J[\delta\mathbf{x}]$ with the implied background error covariance matrix $\mathbf{B}_{\mathrm{imp}} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$.

Example of the CVT method to model horizontal background error covariances (e.g. for pressure, p)

See Fig. for definitions of angles and lengths in real and Fourier spaces. Note the following:

$$
\begin{aligned}
\Delta\mathbf{r} &= \begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix} = \Delta r\begin{pmatrix}\cos\theta_x \\ \sin\theta_x\end{pmatrix}, \\
\mathbf{k} &= \begin{pmatrix}k_x \\ k_y\end{pmatrix} = K\begin{pmatrix}\cos\theta_k \\ \sin\theta_k\end{pmatrix}, \\
d\mathbf{k} &= K\,dK\,d\theta_k.
\end{aligned}
$$



What is the saving of this CVT method of modelling B compared to an explicit matrix method?

- No. of grid points: $n_x \times n_y$.
- No. of pieces of information in $\delta\mathbf{x}$: $3 \times n_x \times n_y$.
- No. of pieces of information in $\delta\boldsymbol{\chi}$: $n_x \times n_y$.
- No. of independent elements in explicit $\mathbf{B}$: $\sim \frac{1}{2}(3 \times n_x \times n_y)^2 \sim \frac{9}{2}n_x^4$ (assuming $n_x \sim n_y$).
- No. of pieces of information needed for CVT: $\sim$ No. of total wavenumbers needed to know $\lambda_p(K) \sim \sqrt{2}n_x$.

If $n_x = 1000$, then

- No. of independent elements in explicit $\mathbf{B}$: $\sim 5 \times 10^{12}$.
- No. of pieces of information needed for CVT: ~ 1500.

Operational CVTs

- The Met Office use a similar approach in its operational 4D-VAR and 3DFGAT systems. Geostrophic balance (imposed weakly) and hydrostatic balance are used. The spatial component includes a similar approach as shown above (spectral space) for the horizontal structure of background error covariances, and vertical modes (empirical orthogonal functions) for the vertical structure. *Lorenc A.C., Ballard S.P., Bell R.S., Ingleby N.B., Andrews P.L.F., Barker D.M., Bray J.R., Clayton A.M., Dalby T., Li D., Payne T.J., Saunders F.W., The Met Office global 3-dimensional variational data assimilation scheme, Q.J.R.Meteor.Soc. 126 pp.2991-3012 (2000).*
- The ECMWF use similar balance relationships, but use a spatial component that makes use of wavelets. *Fisher M., Andersson E., Developments in 4d-Var and Kalman filtering, ECMWF Research Report No. 347 pp.36 (2001).*
- The diffusion operator approach is used in ocean data assimilation systems. *Weaver A.T., Deltel C., Machu E., Ricci S., Daget N., A multivariate balance operator for variational ocean data assimilation, Q.J.R.Meteor.Soc. 131 pp.3605-3626 (2005).*

3(f) Conditioning of the variational problem

The rate of convergence of the variational problem is affected strongly by the *conditioning* of the variational problem. Consider the case when $\delta\mathbf{x}$ is the control variable. A Taylor expansion of $J(\mathbf{x})$ with respect to perturbations $\delta\mathbf{x}$ about $\mathbf{x}$ is:

$$\begin{array}{ccccccc} J(\mathbf{x}+\delta\mathbf{x}) & = & J(\mathbf{x}) & + & \left.\dfrac{\partial J}{\partial \delta\mathbf{x}}\right|_{\mathbf{x}} \delta\mathbf{x} & + & \dfrac{1}{2}\delta\mathbf{x}^{\mathrm{T}} \left.\dfrac{\partial^2 J}{\partial \delta\mathbf{x}^2}\right|_{\mathbf{x}} \delta\mathbf{x}. \\ & & & & \text{gradient vector} & & \text{Hessian matrix} \\ (1\times 1) & & (1\times 1) & & (1\times n)(n\times 1) & & (1\times n)(n\times n)(n\times 1) \end{array}$$

The Hessian matrix is an $n \times n$ matrix that describes all possible second derivatives of J with respect to the control variable elements:

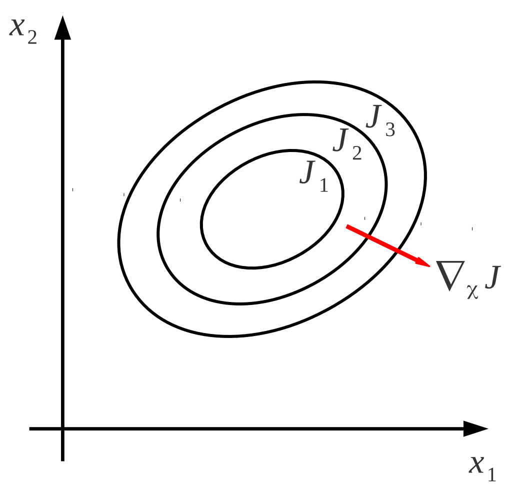
$$\frac{\partial^2 J}{\partial \delta \mathbf{x}^2} = \begin{pmatrix} \frac{\partial^2 J}{\partial x_1^2} & \frac{\partial^2 J}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 J}{\partial x_1 \partial x_n} \\ \frac{\partial^2 J}{\partial x_2 \partial x_1} & \frac{\partial^2 J}{\partial x_2^2} & \cdots & \frac{\partial^2 J}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J}{\partial x_n \partial x_1} & \frac{\partial^2 J}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 J}{\partial x_n^2} \end{pmatrix},$$

and describes the eccentricity and orientation of the ellipsoids that describe surfaces of constant J in phase space. In particular, the condition number is important:

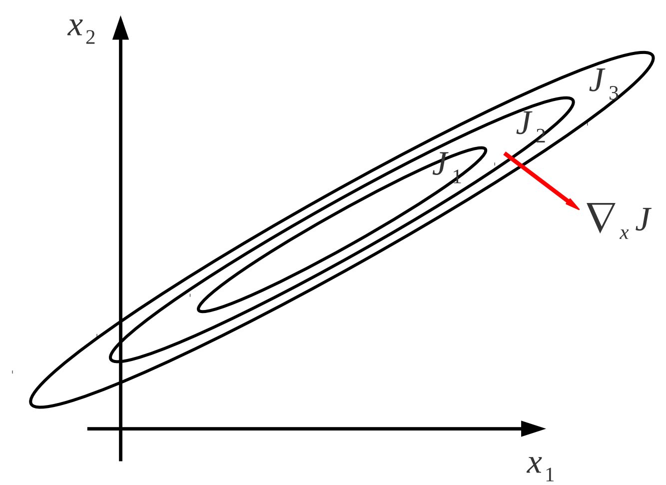
$$\kappa = \text{condition number} = \frac{\text{maximum eigenvalue of the Hessian}}{\text{minimum eigenvalue of the Hessian}}.$$

- If $\kappa \approx 1$, then the variational problem is well conditioned and it will be possible for the solution to be found to a high accuracy.
- If $\kappa \gg 1$, then the variational problem will converge slowly and it is hard for the solution to be found to a high accuracy.

Low condition number



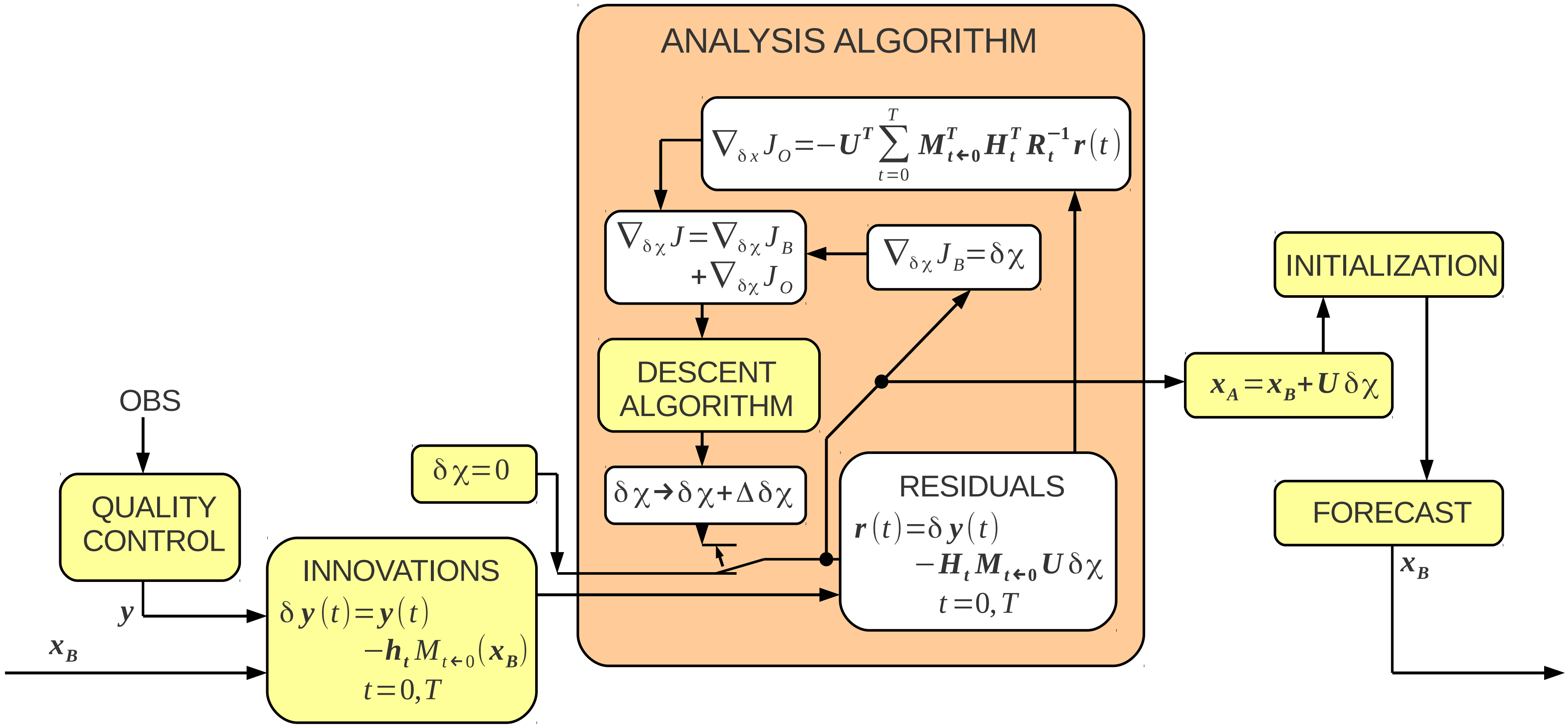
High condition number



The following table compares weak constraint 4D-VAR with $\delta\boldsymbol{\chi}$ and $\delta\mathbf{x}$ as the control variable.

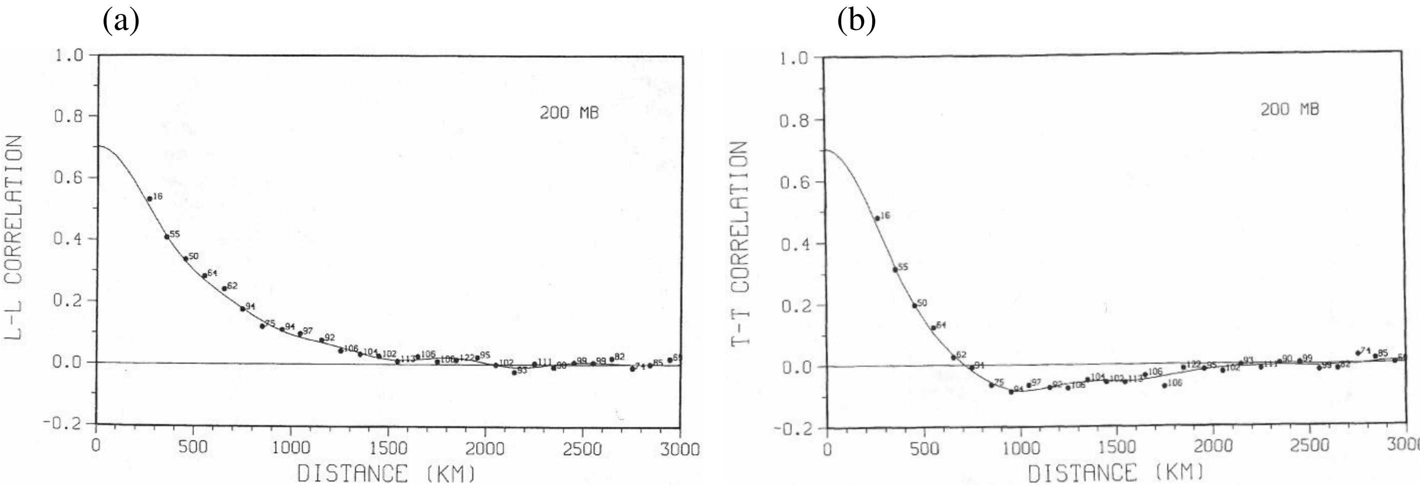
	$\delta\boldsymbol{\chi}$	$\delta\mathbf{x}$
Hessian	$\mathbf{I}+\sum_{t=0}^{T}\mathbf{U}^{\mathrm{T}}\mathbf{M}_{t\leftarrow 0}^{\mathrm{T}}\mathbf{H}_t^{\mathrm{T}}\mathbf{R}_t^{-\mathrm{T}}\mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}$	$\mathbf{B}^{-1}+\sum_{t=0}^{T}\mathbf{M}_{t\leftarrow 0}^{\mathrm{T}}\mathbf{H}_t^{\mathrm{T}}\mathbf{R}_t^{-\mathrm{T}}\mathbf{H}_t\mathbf{M}_{t\leftarrow 0}$
min eigenvalue	$\lambda_{\min}^{\boldsymbol{\chi}} \gtrsim 1$	$\lambda_{\min}^{\mathbf{x}} \geq 0$
max eigenvalue	$\lambda_{\max}^{\boldsymbol{\chi}}$	$\lambda_{\max}^{\mathbf{x}} \gg 1$ in practice
condition No.	$\lambda_{\max}^{\boldsymbol{\chi}}/1 \sim \lambda_{\max}^{\boldsymbol{\chi}}$	$\lambda_{\max}^{\mathbf{x}}/0^{+} \to \infty$

4. Operational algorithms



5. Measuring the B-matrix

5(a) Analysis of innovations

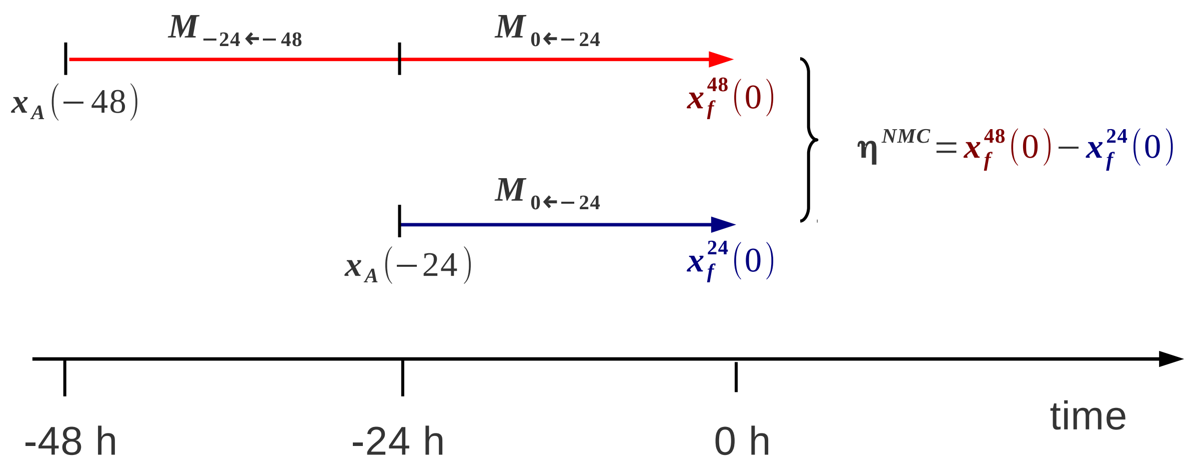


- The H+L method was popular in the 1980s and 1990s.
- It replies on a huge number of direct (in-situ) observations.
- Not useful in practice to probe flow dependence of $\mathbf{B}$, or $\mathbf{B}$ in unobserved regions.
- *Hollingsworth A., Lonnberg P., The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, Tellus 38A pp.111-136 (1986). Lonnberg P., Hollingsworth A., The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors, Tellus 38A pp.137-161 (1986).*

5(b) The NMC method

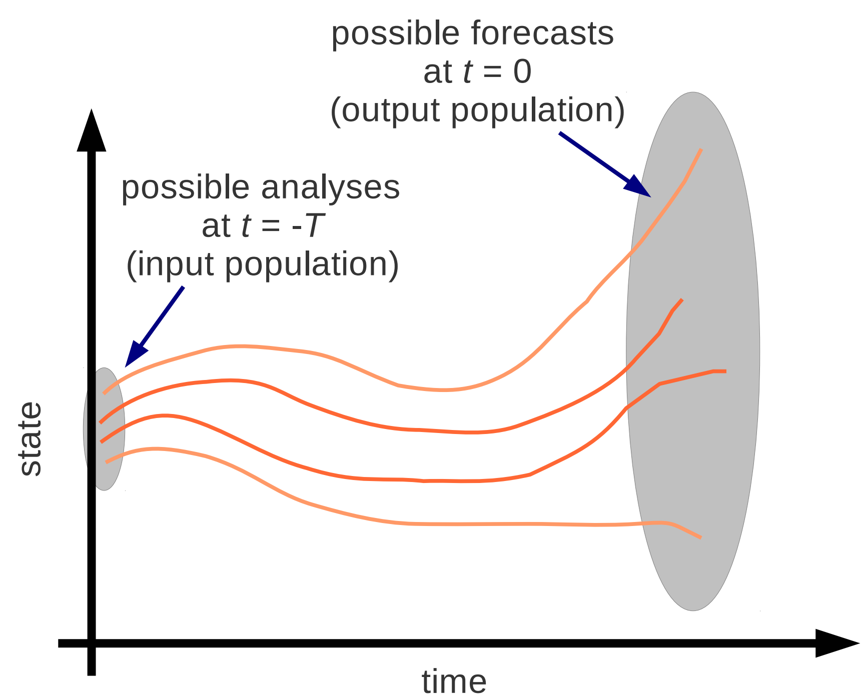
Propose a proxy for forecast error:

$$\boldsymbol{\eta}^{\mathrm{NMC}} \approx \mathbf{x}_{\mathrm{f}}^{48}(0) - \mathbf{x}_{\mathrm{f}}^{24}(0).$$



5(c) Monte-Carlo (ensemble) method

Generate an ensemble that ideally simulates all known sources of forecast error.



For the ith ensemble member ($1 \leq i \leq N$):

$$\mathbf{x}^{(i)}(t+\delta t) = \boldsymbol{\mathcal{M}}_{t+\delta t \leftarrow t}\left(\mathbf{x}^{(i)}(t)\right) + \mathbf{e}^{(i)}(t),$$

integrated from $t=-T$ to $t=0$. The following sources of error are considered:

- Initial condition error, $\delta\mathbf{x}_{\mathrm{A}}^{(i)}(-T)$, e.g.:

  $$\mathbf{x}^{(i)}(-T) = \mathbf{x}_{\mathrm{A}}(-T) + \delta\mathbf{x}_{\mathrm{A}}^{(i)}(-T),$$

  where

  $$\frac{1}{N-1}\sum_{i=1}^{N}\delta\mathbf{x}_{\mathrm{A}}^{(i)}(-T)\delta\mathbf{x}_{\mathrm{A}}^{(i)\mathrm{T}}(-T) \approx \mathbf{P}_{\mathrm{A}}(-T).$$

  All errors inherited from previous DA cycles are represented as initial condition errors.

- Model error, the integrated effect of $\mathbf{e}^{(i)}(t)$. The model error is unknown, but can be included stochastically during the integration of the model. Practical methods of implicitly approximating model error include:
  - Multi-model/multi-physics methods (these use different models, different parameterizations or different parameter values of the parameterizations for each ensemble methods to approximate the effect of $\mathbf{e}^{(i)}(t)$).
  - Stochastic kinetic energy backscatter (SKEB) methods (forecast models do not represent the energy well at scales close to the grid-scale - leading to significant model errors; SKEB injects kinetic energy into the model to try to make up for this).
  - Stochastically perturbed tendencies (SPT) (tendencies from the - imperfect - parametrization schemes are scaled and added as possible model errors).

- Other errors (e.g. boundary condition perturbations for limited area models, perturbations to the unknown forcings of the model).

6. Hybrid (var/ensemble) formations

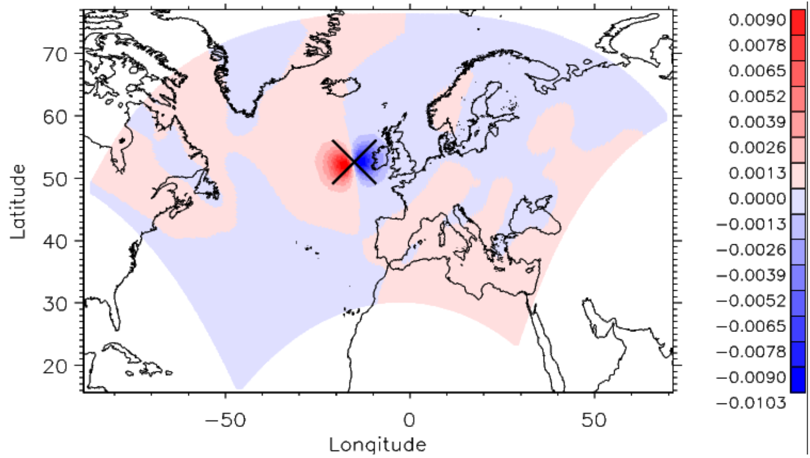
6(a) Basic ideas

Let us consider the pros and cons of variational data assimilation and ensemble data assimilation (such as the ensemble Kalman filter discussed in part II of this course).

	VARIATIONAL DATA ASSIMILATION	ENSEMBLE KALMAN FILTER	
1. Efficiency	Good	Good	
2. Data voids	Reverts to the background state, $\mathbf{x}_\mathrm{B}$	Reverts to the background state, $\mathbf{x}_\mathrm{B}$	
3. Processing	Continuous (within assimilation window)	Intermittent	
4. Scaling for parallel computing	Limits to scaling	No limits to scaling	
5. Errors in inputs	Allows for errors in $\mathbf{x}_\mathrm{B}$ and $\mathbf{y}$	Allows for errors in $\mathbf{x}_\mathrm{B}$ and $\mathbf{y}$	
6. Errors in model	Accounted for in WC 4D-VAR	Accounted for	
7. Indirect observations	Yes	Yes	
8. Balance and smoothness of analysis	Yes	No, unless N is sufficiently large	*
9. Flow dependent background error covariance matrix	No, $\mathbf{P}^\mathrm{f}$ is approximated by $\mathbf{B}$	Yes, $\mathbf{P}^\mathrm{f}$ is approximated by $\mathbf{P}^\mathrm{f}_{(N)}$	*
10. Rank of background error covariance matrix	Full rank	$\mathrm{rank} \leq N$	*

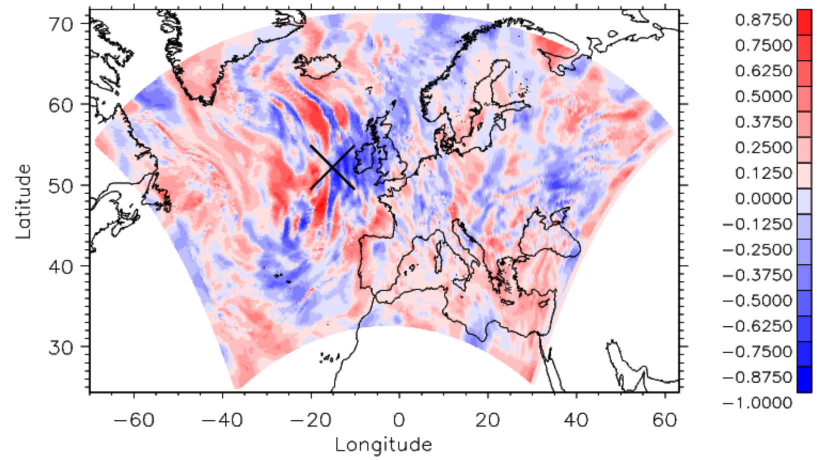
* These issues are related. The aim of hybrid data assimilation is to combine VAR with an ensemble to get the best bits of each approach.

Variational assimilation structure function

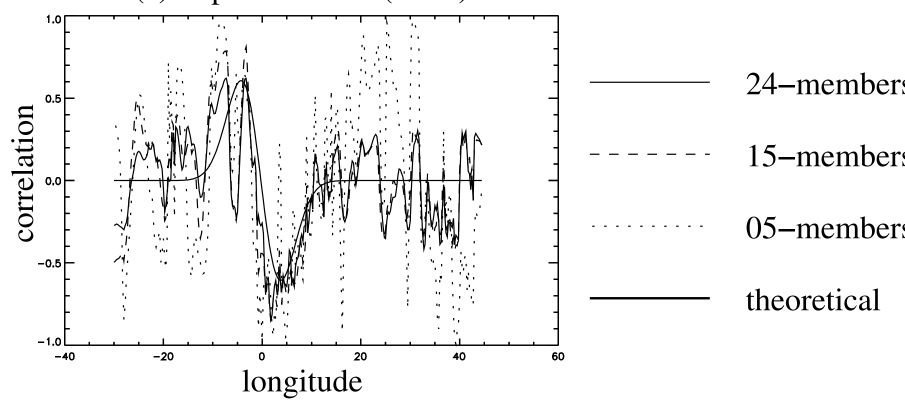


- Full rank, but not flow dependent.

Ensemble-derived structure function ($N = 24$)



(c) v–p correlation (NAE)



- Flow dependent, but rank deficient.

In the hybrid solution, we solve a VAR-like problem but $\mathbf{B} \to \mathbf{P}^{\mathrm{H}}$:

$$\mathbf{P}^{\mathrm{H}} = \alpha \mathbf{B} + (1-\alpha)\mathbf{P}^{\mathrm{f}}_{(N)}, \text{ where } 0 \le \alpha \le 1.$$

6(b) Incorporating a simple hybrid scheme in VAR

In order to use $\mathbf{P}^{\mathrm{H}} = \alpha\mathbf{B} + (1-\alpha)\mathbf{P}^{\mathrm{f}}_{(N)}$ in variational assimilation, $\mathbf{P}^{\mathrm{H}}$ needs to be made compatible with the control variable transform (CVT).

Recall from 3(e), $\mathbf{B}$ is modelled by minimizing the cost function with respect to a control variable $\delta\boldsymbol{\chi}$:

$$\begin{aligned} J[\delta\boldsymbol{\chi}] &= \frac{1}{2}\delta\boldsymbol{\chi}^{\mathrm{T}}\delta\boldsymbol{\chi} + \\ &\quad \frac{1}{2}\sum_{t=0}^{T}\left[\delta\mathbf{y}(t) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}\delta\boldsymbol{\chi}\right]^{\mathrm{T}}\mathbf{R}_t^{-1} \times \\ &\qquad \left[\delta\mathbf{y}(t) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}\delta\boldsymbol{\chi}\right], \\ \text{where } \delta\mathbf{x} &= \mathbf{U}\delta\boldsymbol{\chi}, \\ \text{and } \left\langle\delta\boldsymbol{\chi}\delta\boldsymbol{\chi}^{\mathrm{T}}\right\rangle &= \mathbf{I}, \end{aligned}$$

and the implied background error covariance matrix is:

$$\mathbf{B}_{\mathrm{imp}} = \mathbf{U}\mathbf{U}^{\mathrm{T}}.$$

Now consider the following cost function and modification to the control variable and its CVT:

$$\begin{aligned} J^{\mathrm{H}}[\delta\boldsymbol{\chi}^{\mathrm{H}}] &= \frac{1}{2}\delta\boldsymbol{\chi}_{\mathrm{var}}^{\mathrm{T}}\delta\boldsymbol{\chi}_{\mathrm{var}} + \frac{1}{2}\delta\boldsymbol{\chi}_{\mathrm{ens}}^{\mathrm{T}}\delta\boldsymbol{\chi}_{\mathrm{ens}} + \\ &\quad \frac{1}{2}\sum_{t=0}^{T}\left[\delta\mathbf{y}(t) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}^{\mathrm{H}}\delta\boldsymbol{\chi}^{\mathrm{H}}\right]^{\mathrm{T}}\mathbf{R}_t^{-1} \times \\ &\qquad \left[\delta\mathbf{y}(t) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}^{\mathrm{H}}\delta\boldsymbol{\chi}^{\mathrm{H}}\right], \\ \text{where } \delta\mathbf{x} &= \mathbf{U}^{\mathrm{H}}\delta\boldsymbol{\chi}^{\mathrm{H}}, \\ \text{and } \left\langle\delta\boldsymbol{\chi}^{\mathrm{H}}\delta\boldsymbol{\chi}^{\mathrm{HT}}\right\rangle &= \mathbf{I}, \\ \text{but now } \delta\boldsymbol{\chi}^{\mathrm{H}} &= \begin{pmatrix} \delta\boldsymbol{\chi}_{\mathrm{var}} \\ \delta\boldsymbol{\chi}_{\mathrm{ens}} \end{pmatrix}, \qquad \delta\boldsymbol{\chi}_{\mathrm{var}} \in \mathbb{R}^n,\ \delta\boldsymbol{\chi}_{\mathrm{ens}} \in \mathbb{R}^N, \\ \text{and } \mathbf{U}^{\mathrm{H}} &= \begin{pmatrix} \sqrt{\alpha}\mathbf{U} & \sqrt{\frac{1-\alpha}{N-1}}\mathbf{X} \end{pmatrix}. \end{aligned}$$

What is the implied background error covariance matrix of this scheme?

$$\begin{aligned} \mathbf{B}^{\mathrm{H}}_{\mathrm{imp}} &= \left\langle\delta\mathbf{x}\delta\mathbf{x}^{\mathrm{T}}\right\rangle = \mathbf{U}^{\mathrm{H}}\left\langle\delta\boldsymbol{\chi}^{\mathrm{H}}\delta\boldsymbol{\chi}^{\mathrm{HT}}\right\rangle\mathbf{U}^{\mathrm{HT}} = \mathbf{U}^{\mathrm{H}}\mathbf{U}^{\mathrm{HT}}, \\ &= \begin{pmatrix} \sqrt{\alpha}\mathbf{U} & \sqrt{\frac{1-\alpha}{N-1}}\mathbf{X} \end{pmatrix}\begin{pmatrix} \sqrt{\alpha}\mathbf{U}^{\mathrm{T}} \\ \sqrt{\frac{1-\alpha}{N-1}}\mathbf{X}^{\mathrm{T}} \end{pmatrix} = \alpha\mathbf{U}\mathbf{U}^{\mathrm{T}} + \frac{1-\alpha}{N-1}\mathbf{X}\mathbf{X}^{\mathrm{T}}, \\ &= \alpha\mathbf{B} + (1-\alpha)\mathbf{P}^{\mathrm{f}}_{(N)}. \end{aligned}$$

The first term contains $\mathbf{U}\mathbf{U}^{\mathrm{T}}$, which is the implied background error covariance matrix from the pure variational scheme, and the second term contains $\mathbf{X}\mathbf{X}^{\mathrm{T}}/(N-1)$, which is the ensemble-derived background error covariance matrix (we used this notation in section 3(b), and in problem 11).

6(c) Incorporating a localized hybrid scheme in VAR

The ensemble contribution to the hybrid covariance is noisy when N is small. How can we mitigate this noise?

- A statistical result tells us that the error in the sample correlation between two variables x and y has expectation $(1 - \mathrm{cor}^2(x, y))/\sqrt{N-1}$.
- For a given N, sampling errors are expected to be largest when the correlations are close to zero.
- Correlations are expected to be smaller at larger separations.
- 'Localization' artificially reduces covariances between variables separated by large distances.

Let $x = \boldsymbol{\eta}_{\mathrm{B}}(\mathbf{r}_1)$ and $y = \boldsymbol{\eta}_{\mathrm{B}}(\mathbf{r}_2)$. The raw covariance between x and y is:

$$\mathbf{P}^{\mathrm{f}}_{(N)}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{N-1} \sum_{i=1}^{N} \boldsymbol{\eta}^{(i)}_{\mathrm{B}}(\mathbf{r}_1) \boldsymbol{\eta}^{(i)}_{\mathrm{B}}(\mathbf{r}_2).$$

For the covariance actually used in the hybrid scheme, we wish to multiply this by a moderation function that decreases with separation between $\mathbf{r}_1$ and $\mathbf{r}_2$: $\boldsymbol{\Omega}(\mathbf{r}_1, \mathbf{r}_2) =$ prescribed function of $|\mathbf{r}_1 - \mathbf{r}_2|$, $0 \leq \boldsymbol{\Omega}(\mathbf{r}_1, \mathbf{r}_2) \leq 1$. The covariance used is then:

$$\mathbf{P}^{\mathrm{f,l}}_{(N)}(\mathbf{r}_1, \mathbf{r}_2) = \mathbf{P}^{\mathrm{f}}_{(N)}(\mathbf{r}_1, \mathbf{r}_2) \boldsymbol{\Omega}(\mathbf{r}_1, \mathbf{r}_2).$$

This is for a particular matrix element. For the whole covariance matrix, introduce the Schur product of matrices:

$$\mathbf{P}^{\mathrm{f,l}}_{(N)} = \mathbf{P}^{\mathrm{f}}_{(N)} \circ \boldsymbol{\Omega}, \qquad \boldsymbol{\Omega} \in \mathbb{R}^{n \times n}.$$

How do we incorporate this into the CVT?

This section is provided for information only. In outline:

- We know that $\mathbf{P}^{\mathrm{f}}_{(N)} = \frac{1}{N-1} \mathbf{X}\mathbf{X}^{\mathrm{T}}$, $\mathbf{P}^{\mathrm{f}}_{(N)} \in \mathbb{R}^{n \times n}$, $\mathbf{X} \in \mathbb{R}^{n \times N}$.
- Now suppose that we can decompose $\boldsymbol{\Omega}$ in terms of M members in $\mathbf{Y}$: $\boldsymbol{\Omega} = \frac{1}{M-1} \mathbf{Y}\mathbf{Y}^{\mathrm{T}}$, $\boldsymbol{\Omega} \in \mathbb{R}^{n \times n}$, $\mathbf{Y} \in \mathbb{R}^{n \times M}$.

- Then the localized background error covariance matrix is:

$$\begin{aligned}\mathbf{P}^{\mathrm{f,l}}_{(N)} &= \mathbf{P}^{\mathrm{f}}_{(N)} \circ \boldsymbol{\Omega}, \\ &= \left(\frac{1}{N-1}\mathbf{X}\mathbf{X}^{\mathrm{T}}\right) \circ \left(\frac{1}{M-1}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\right), \\ &= \frac{1}{(M-1)(N-1)}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right) \circ \left(\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\right).\end{aligned}$$

- It is possible to construct a new matrix $\mathbf{X}_\Omega$ such that $\mathbf{P}^{\mathrm{f,l}}_{(N)} = \frac{1}{(N-1)(M-1)}\mathbf{X}_\Omega\mathbf{X}_\Omega^{\mathrm{T}}$, $\mathbf{X}_\Omega \in \mathbb{R}^{n\times NM}$.
- This new matrix has the form:

$$\mathbf{X}_\Omega = \left(\begin{array}{cccccccccc} \uparrow & \uparrow & & \uparrow & \uparrow & & \uparrow & & & \uparrow \\ \boldsymbol{\eta}_{\mathrm{B}}^{(1)} \circ \mathbf{y}^{(1)} & \boldsymbol{\eta}_{\mathrm{B}}^{(1)} \circ \mathbf{y}^{(2)} & \cdots & \boldsymbol{\eta}_{\mathrm{B}}^{(1)} \circ \mathbf{y}^{(M)} & \boldsymbol{\eta}_{\mathrm{B}}^{(2)} \circ \mathbf{y}^{(1)} & \cdots & \boldsymbol{\eta}_{\mathrm{B}}^{(2)} \circ \mathbf{y}^{(M)} & \cdots & \cdots & \boldsymbol{\eta}_{\mathrm{B}}^{(N)} \circ \mathbf{y}^{(M)} \\ \downarrow & \downarrow & & \downarrow & \downarrow & & \downarrow & & & \downarrow \end{array} \right),$$

  where $\boldsymbol{\eta}_{\mathrm{B}}^{(i)}$ is the ith column of $\mathbf{X}$ and $\mathbf{y}^{(j)}$ is the jth column of $\mathbf{X}_\Omega$. There are other compact ways to write this matrix: Buehner M., Ensemble derived stationary and flow dependent background error covariances: Evaluation in a quasi-operational NWP setting, Q.J.R.Meteor.Soc. 131 pp.1013-1043 (2005).

- The localized hybrid scheme is then the same as the unlocalized one, but with
  - the N-element part of the control vector $\delta\boldsymbol{\chi}^{\mathrm{H}}$ relaced with an NM-element control vector, and
  - $\sqrt{\frac{1-\alpha}{N-1}}\mathbf{X}$ in the CVT replaced with $\sqrt{\frac{1-\alpha}{(N-1)(M-1)}}\mathbf{X}_\Omega$.

N.B. There are other ways of representing a hyrid system in terms of control variables: Lorenc A.C., The potential of the ensemble Kalman filter for NWP - a comparison with 4d-Var, Q.J.R.Meteor.Soc. 129 pp.3183-3203 (2003).

7. Data assimilation diagnostics

- What can go wrong with a data assimilation scheme? For a strong constraint 4D-VAR, e.g.:

- Incorrect error covariance matrices.
- Non-Gaissian or biased errors in the background or the observations.
- Errors in $\boldsymbol{\mathcal{M}}$, $\mathbf{h}$, $\mathbf{M}$ or $\mathbf{H}$.
- Strong non-linearities in $\boldsymbol{\mathcal{M}}$ or $\mathbf{h}$.
- Variational procedure not converged to the minimum.
- Background and observation errors are correlated.

- How can we assess if a given data assimilation scheme is sub-optimal? E.g. for variational data assimilation:
  - Bennett-Talagrand diagnostic.
  - Desrozier's diagnostics.

7(a) The Bennett-Talagrand theorem[1]

Twice the cost function value at the minimum (i.e. at the analysis) for an optimal assimilation system is a random variable that obeys χ^2 statistics and therefore has a particular expectation value[2]. Statistics tells us that the expectation value of a χ^2 distribution that results from a fit of ν degrees of freedom to q pieces of data is $\mathcal{E}(2J_{\min}) = q - \nu$. The data assimilation problem tries to fit $\nu = n$ pieces of information to $q = n + p$ pieces of information (the background state and the observations). Then, $\mathcal{E}(2J_{\min}) = n + p - n = p$. Therefore the expected value of $J_{\min}$ is

$$\mathcal{E}(J_{\min}) = \frac{p}{2}.$$

If a given assimilation run does not give a value of $J_{\min}$ close to this value then it is an indication that something is wrong with the data assimilation. This can also be proved directly for the data assimilation problem (the Bennet-Talagrand theorem).

[1]Based on notes by T. Payne, Met Office

[2]For any one assimilation, there will be one value of the cost function at the minimum, so what do we mean by the "expected value of the cost function at the minimum"? Imagine doing a very large number of assimilations of the same situation, but each with slightly different backgrounds and observations (where perturbations are consistent with the background and observations error covariance matrices). This is like doing different data assimilation runs in parallel universes. The expected value of the cost function at the minimum is the average of these experiments.

This section is provided for information only. Assume a data assimilation system that is optimal (e.g. all error covariance matrices are correctly specified). Then

$$\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}} = \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_{\mathrm{b}}) \text{ where } \mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}} \left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\right). \tag{38}$$

We wish to evaluate the expected value of the cost function at its minimum, $\mathbf{x} = \mathbf{x}_{\mathrm{a}}$. This expected value is written $\mathcal{E}[J(\mathbf{x}_{\mathrm{a}})]$ and the cost function at the analysis is (given a specific background state and set of observations)

$$\begin{aligned} J(\mathbf{x}_{\mathrm{a}}) &= J_{\mathrm{b}}(\mathbf{x}_{\mathrm{a}}) + J_{\mathrm{o}}(\mathbf{x}_{\mathrm{a}}), & (39)\\ \text{where } J_{\mathrm{b}}(\mathbf{x}_{\mathrm{a}}) &= \frac{1}{2}\left(\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}^{-1} \left(\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}}\right), & (40)\\ \text{and } J_{\mathrm{o}}(\mathbf{x}_{\mathrm{a}}) &= \frac{1}{2}\left(\mathbf{y} - \mathbf{H}\mathbf{x}_{\mathrm{a}}\right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\mathbf{y} - \mathbf{H}\mathbf{x}_{\mathrm{a}}\right). & (41) \end{aligned}$$

The analysis, background and observation errors are (again given a specific background state and set of observations)

$$\varepsilon_{\mathrm{a}} = \mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{t}}, \quad \varepsilon_{\mathrm{b}} = \mathbf{x}_{\mathrm{b}} - \mathbf{x}_{\mathrm{t}}, \quad \varepsilon_{\mathrm{o}} = \mathbf{y} - \mathbf{H}\mathbf{x}_{\mathrm{t}}. \tag{42}$$

The analysis error can be developed as follows using (38) and (42):

$$\begin{aligned} \varepsilon_{\mathrm{a}} &= \mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}} + \varepsilon_{\mathrm{b}} = \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}_{\mathrm{b}}\right) + \varepsilon_{\mathrm{b}},\\ &= \mathbf{K}\left(\mathbf{y} - \mathbf{H}(\mathbf{x}_{\mathrm{b}} - \mathbf{x}_{\mathrm{t}}) - \mathbf{H}\mathbf{x}_{\mathrm{t}}\right) + \varepsilon_{\mathrm{b}},\\ &= \mathbf{K}\left(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{b}}\right) + \varepsilon_{\mathrm{b}} = \left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)\varepsilon_{\mathrm{b}} + \mathbf{K}\varepsilon_{\mathrm{o}}. \end{aligned} \tag{43}$$

Equations (40) and (41) are inner products. To evaluate them, the following identity is useful

$$\mathbf{u}^{\mathrm{T}}\mathbf{C}\mathbf{v} = \sum_{i,j} u_i C_{ij} v_j = \mathrm{tr}\left(\mathbf{C}\mathbf{v}\mathbf{u}^{\mathrm{T}}\right). \tag{44}$$

The background term

The expectation of the background term (40) is, using (44):

$$\begin{aligned} \mathcal{E}[J_{\mathrm{b}}(\mathbf{x}_{\mathrm{a}})] &= \frac{1}{2}\mathcal{E}\left[\mathrm{tr}\left(\mathbf{B}^{-1}(\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}})(\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}})^{\mathrm{T}}\right)\right] = \frac{1}{2}\mathrm{tr}\left(\mathbf{B}^{-1}\mathcal{E}\left[(\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}})(\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}})^{\mathrm{T}}\right]\right),\\ &= \frac{1}{2}\mathrm{tr}\left(\mathbf{B}^{-1}\mathcal{E}\left[(\varepsilon_{\mathrm{a}} - \varepsilon_{\mathrm{b}})(\varepsilon_{\mathrm{a}} - \varepsilon_{\mathrm{b}})^{\mathrm{T}}\right]\right), \end{aligned} \tag{45}$$

where (42) have been used for the last line. Part of the last line is the expression $\mathcal{E}[(\varepsilon_\mathrm{a}-\varepsilon_\mathrm{b})(\varepsilon_\mathrm{a}-\varepsilon_\mathrm{b})^\mathrm{T}]$ which may be developed using (43)

$$\begin{aligned}
\mathcal{E}\left[(\varepsilon_\mathrm{a}-\varepsilon_\mathrm{b})(\varepsilon_\mathrm{a}-\varepsilon_\mathrm{b})^\mathrm{T}\right] &= \mathcal{E}\left[\varepsilon_\mathrm{a}\varepsilon_\mathrm{a}^\mathrm{T}+\varepsilon_\mathrm{b}\varepsilon_\mathrm{b}^\mathrm{T}-\varepsilon_\mathrm{a}\varepsilon_\mathrm{b}^\mathrm{T}-\varepsilon_\mathrm{b}\varepsilon_\mathrm{a}^\mathrm{T}\right], \\
&= (\mathbf{I}-\mathbf{KH})\mathcal{E}\left[\varepsilon_\mathrm{b}\varepsilon_\mathrm{b}^\mathrm{T}\right](\mathbf{I}-\mathbf{KH})^\mathrm{T}+\mathbf{K}\mathcal{E}\left[\varepsilon_\mathrm{o}\varepsilon_\mathrm{o}^\mathrm{T}\right]\mathbf{K}^\mathrm{T}+\mathcal{E}\left[\varepsilon_\mathrm{b}\varepsilon_\mathrm{b}^\mathrm{T}\right] \\
&\quad -(\mathbf{I}-\mathbf{KH})\mathcal{E}\left[\varepsilon_\mathrm{b}\varepsilon_\mathrm{b}^\mathrm{T}\right]-\mathcal{E}\left[\varepsilon_\mathrm{b}\varepsilon_\mathrm{b}^\mathrm{T}\right](\mathbf{I}-\mathbf{KH})^\mathrm{T}, \\
&= (\mathbf{I}-\mathbf{KH})\mathbf{B}(\mathbf{I}-\mathbf{KH})^\mathrm{T}+\mathbf{KRK}^\mathrm{T}+\mathbf{B}-(\mathbf{I}-\mathbf{KH})\mathbf{B}-\mathbf{B}(\mathbf{I}-\mathbf{KH})^\mathrm{T}, \\
&= \mathbf{B}+\mathbf{KHB}(\mathbf{KH})^\mathrm{T}-\mathbf{B}(\mathbf{KH})^\mathrm{T}-\mathbf{KHB}+\mathbf{KRK}^\mathrm{T}+\mathbf{B}-\mathbf{B} \\
&\quad +\mathbf{KHB}-\mathbf{B}+\mathbf{B}(\mathbf{KH})^\mathrm{T}, \\
&= \mathbf{KHB}(\mathbf{KH})^\mathrm{T}+\mathbf{KRK}^\mathrm{T}.
\end{aligned}$$

These steps assume that background and observation errors are mutually uncorrelated. Using the definition of $\mathbf{K}$ (38) turns the above into:

$$\begin{aligned}
\mathcal{E}\left[(\varepsilon_\mathrm{a}-\varepsilon_\mathrm{b})(\varepsilon_\mathrm{a}-\varepsilon_\mathrm{b})^\mathrm{T}\right] &= \mathbf{BH}^\mathrm{T}(\mathbf{HBH}^\mathrm{T}+\mathbf{R})^{-1}\mathbf{HB}(\mathbf{BH}^\mathrm{T}(\mathbf{HBH}^\mathrm{T}+\mathbf{R})^{-1}\mathbf{H})^\mathrm{T} \\
&\quad +\mathbf{BH}^\mathrm{T}(\mathbf{HBH}^\mathrm{T}+\mathbf{R})^{-1}\mathbf{R}(\mathbf{BH}^\mathrm{T}(\mathbf{HBH}^\mathrm{T}+\mathbf{R})^{-1})^\mathrm{T}, \\
&= \mathbf{BH}^\mathrm{T}(\mathbf{HBH}^\mathrm{T}+\mathbf{R})^{-1}\mathbf{HBH}^\mathrm{T}(\mathbf{HBH}^\mathrm{T}+\mathbf{R})^{-1}\mathbf{HB} \\
&\quad +\mathbf{BH}^\mathrm{T}(\mathbf{HBH}^\mathrm{T}+\mathbf{R})^{-1}\mathbf{R}(\mathbf{HBH}^\mathrm{T}+\mathbf{R})^{-1}\mathbf{HB}, \\
&= \mathbf{BH}^\mathrm{T}(\mathbf{HBH}^\mathrm{T}+\mathbf{R})^{-1}\mathbf{HB}=\mathbf{KHB}.
\end{aligned}$$

Inserting this into (45) gives

$$\mathcal{E}[J_\mathrm{b}(\mathbf{x}_\mathrm{a})] = \frac{1}{2}\mathrm{tr}(\mathbf{B}^{-1}\mathbf{KHB}). \tag{46}$$

Note the following identity, which holds for matrices $\mathbf{E}$ and $\mathbf{F}$, where $\mathbf{E}$ is $r\times s$ and $\mathbf{F}$ is $s\times r$

$$\mathrm{tr}(\mathbf{EF})=\sum_{j=1}^{r}\sum_{i=1}^{s}E_{ji}F_{ij}=\sum_{i=1}^{s}\sum_{j=1}^{r}F_{ij}E_{ji}=\mathrm{tr}(\mathbf{FE}), \tag{47}$$

i.e., the order of the operators inside the trace can be reversed. Applying this to (46) gives

$$E[J_\mathrm{b}(\mathbf{x}_\mathrm{a})]=\frac{1}{2}\mathrm{tr}(\mathbf{KH}). \tag{48}$$

The observation term

The expectation of the observation term (41) is, using (44):

$$\begin{aligned}\mathcal{E}[J_o(\mathbf{x}_a)] &= \frac{1}{2}\mathcal{E}\left[\mathrm{tr}\left(\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x}_a)(\mathbf{y}-\mathbf{H}\mathbf{x}_a)^{\mathrm{T}}\right)\right], \\ &= \frac{1}{2}\mathrm{tr}\left(\mathbf{R}^{-1}\mathcal{E}\left[(\mathbf{y}-\mathbf{H}\mathbf{x}_a)(\mathbf{y}-\mathbf{H}\mathbf{x}_a)^{\mathrm{T}}\right]\right), \\ &= \frac{1}{2}\mathrm{tr}\left(\mathbf{R}^{-1}\mathcal{E}\left[(\varepsilon_o-\mathbf{H}\varepsilon_a)(\varepsilon_o-\mathbf{H}\varepsilon_a)^{\mathrm{T}}\right]\right). \end{aligned} \tag{49}$$

where (42) have been used for the last line. Part of the last line is the expression $\mathcal{E}[(\varepsilon_o-\mathbf{H}\varepsilon_a)(\varepsilon_o-\mathbf{H}\varepsilon_a)^{\mathrm{T}}]$ which may be developed using (43):

$$\begin{aligned}\mathcal{E}[(\varepsilon_o-\mathbf{H}\varepsilon_a)(\varepsilon_o-\mathbf{H}\varepsilon_a)^{\mathrm{T}}] &= \mathbf{H}\mathcal{E}[\varepsilon_a\varepsilon_a^{\mathrm{T}}]\mathbf{H}^{\mathrm{T}}+\mathcal{E}[\varepsilon_o\varepsilon_o^{\mathrm{T}}]-\mathbf{H}\mathcal{E}[\varepsilon_a\varepsilon_o^{\mathrm{T}}]-\mathcal{E}[\varepsilon_o\varepsilon_a^{\mathrm{T}}]\mathbf{H}^{\mathrm{T}}, \\ &= \mathbf{H}\{(\mathbf{I}-\mathbf{K}\mathbf{H})\mathcal{E}[\varepsilon_b\varepsilon_b^{\mathrm{T}}](\mathbf{I}-\mathbf{K}\mathbf{H})^{\mathrm{T}}+\mathbf{K}\mathcal{E}[\varepsilon_o\varepsilon_o^{\mathrm{T}}]\mathbf{K}^{\mathrm{T}}\}\mathbf{H}^{\mathrm{T}} \\ &\quad +\mathcal{E}[\varepsilon_o\varepsilon_o^{\mathrm{T}}]-\mathbf{H}\mathbf{K}\mathcal{E}[\varepsilon_o\varepsilon_o^{\mathrm{T}}]-\mathcal{E}[\varepsilon_o\varepsilon_o^{\mathrm{T}}]\mathbf{K}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}, \\ &= \mathbf{H}\{(\mathbf{I}-\mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I}-\mathbf{K}\mathbf{H})^{\mathrm{T}}+\mathbf{K}\mathbf{R}\mathbf{K}^{\mathrm{T}}\}\mathbf{H}^{\mathrm{T}}+\mathbf{R}-\mathbf{H}\mathbf{K}\mathbf{R}-\mathbf{R}\mathbf{K}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}, \\ &= \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}+\mathbf{H}\mathbf{K}\mathbf{H}\mathbf{B}(\mathbf{K}\mathbf{H})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}-\mathbf{H}\mathbf{B}(\mathbf{K}\mathbf{H})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}-\mathbf{H}\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} \\ &\quad +\mathbf{H}\mathbf{K}\mathbf{R}\mathbf{K}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}+\mathbf{R}-\mathbf{H}\mathbf{K}\mathbf{R}-\mathbf{R}\mathbf{K}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}. \end{aligned}$$

These steps assume that background and observation errors are mutually uncorrelated. Using the definition of $\mathbf{K}$ (38):

$$\begin{aligned}\mathcal{E}[(\varepsilon_o-\mathbf{H}\varepsilon_a)(\varepsilon_o-\mathbf{H}\varepsilon_a)^{\mathrm{T}}] &= \mathbf{HBH}^{\mathrm{T}}+\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} \\ &\quad -\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}}-\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} \\ &\quad +\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{R}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}}+\mathbf{R} \\ &\quad -\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{R}-\mathbf{R}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}}. \end{aligned}$$

Merging the 2nd and 5th terms leads to

$$\begin{aligned}\mathcal{E}[(\varepsilon_o-\mathbf{H}\varepsilon_a)(\varepsilon_o-\mathbf{H}\varepsilon_a)^{\mathrm{T}}] &= \mathbf{HBH}^{\mathrm{T}}+\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} \\ &\quad -\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}}-\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} \\ &\quad +\mathbf{R}-\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{R}-\mathbf{R}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}}. \end{aligned}$$

Further simplifications can be made by merging the 3rd and 6th terms and the 4th and 7th terms

$$
\begin{aligned}
\mathcal{E}[(\varepsilon_{\rm o} - \mathbf{H}\varepsilon_{\rm a})(\varepsilon_{\rm o} - \mathbf{H}\varepsilon_{\rm a})^{\rm T}] &= \mathbf{HBH}^{\rm T} + \mathbf{HBH}^{\rm T}(\mathbf{HBH}^{\rm T} + \mathbf{R})^{-1}\mathbf{HBH}^{\rm T} - \mathbf{HBH}^{\rm T} - \mathbf{HBH}^{\rm T} + \mathbf{R}, \\
&= \mathbf{HBH}^{\rm T}\left\{(\mathbf{HBH}^{\rm T} + \mathbf{R})^{-1}\mathbf{HBH}^{\rm T} - \mathbf{I}\right\} + \mathbf{R}. \qquad (50)
\end{aligned}
$$

Consider the term inside the curly brackets in the above:

$$
\begin{aligned}
(\mathbf{HBH}^{\rm T} + \mathbf{R})^{-1}\mathbf{HBH}^{\rm T} - \mathbf{I} &= (\mathbf{HBH}^{\rm T} + \mathbf{R})^{-1}\mathbf{HBH}^{\rm T} - (\mathbf{HBH}^{\rm T} + \mathbf{R})^{-1}(\mathbf{HBH}^{\rm T} + \mathbf{R}), \\
&= (\mathbf{HBH}^{\rm T} + \mathbf{R})^{-1}[\mathbf{HBH}^{\rm T} - \mathbf{HBH}^{\rm T} - \mathbf{R}], \\
&= -(\mathbf{HBH}^{\rm T} + \mathbf{R})^{-1}\mathbf{R}.
\end{aligned}
$$

Using this to rewrite (50):

$$
\mathcal{E}[(\varepsilon_{\rm o} - \mathbf{H}\varepsilon_{\rm a})(\varepsilon_{\rm o} - \mathbf{H}\varepsilon_{\rm a})^{\rm T}] = -\mathbf{HBH}^{\rm T}(\mathbf{HBH}^{\rm T} + \mathbf{R})^{-1}\mathbf{R} + \mathbf{R},
$$

and then substituting this into (49) and then using (47) gives

$$
\begin{aligned}
\mathcal{E}[J_{\rm o}(\mathbf{x}_{\rm a})] &= \frac{1}{2}\mathrm{tr}\left(\mathbf{R}^{-1}[-\mathbf{HBH}^{\rm T}(\mathbf{HBH}^{\rm T} + \mathbf{R})^{-1}\mathbf{R} + \mathbf{R}]\right), \\
&= \frac{1}{2}\mathrm{tr}\left(-\mathbf{HBH}^{\rm T}(\mathbf{HBH}^{\rm T} + \mathbf{R})^{-1} + \mathbf{I}\right), \\
&= \frac{1}{2}\mathrm{tr}\left(-\mathbf{HK} + \mathbf{I}\right) = \frac{1}{2}\left(-\mathrm{tr}(\mathbf{HK}) + p\right) = \frac{1}{2}\left(-\mathrm{tr}(\mathbf{KH}) + p\right), \qquad (51)
\end{aligned}
$$

where p is the number of observations.

The sum of the background and observation terms

The sum of the background and observation terms is (using (39), (48) and (51)):

$$
\mathcal{E}(J(\mathbf{x}_{\rm a})) = \mathcal{E}(J_{\rm b}(\mathbf{x}_{\rm a})) + \mathcal{E}(J_{\rm o}(\mathbf{x}_{\rm a})) = \frac{1}{2}\left(\mathrm{tr}(\mathbf{KH}) - \mathrm{tr}(\mathbf{KH}) + p\right) = \frac{p}{2}.
$$

This is a very involved derivation, but leads to the very simple result that the expectation of the minimum of the cost function has value equal to half the number of observations. Some people have called this the Bennett-Talagrand theorem. If the value of the cost function at the minimum does not have this value in practice then this is an indication that the error characteristics of the data assimilation do not match those of the actual data, or other things are wrong with the set-up like the forward operator, $\mathbf{H}$. Note that this result applies to systems that are Gaussian and linear.

Desrozier's Diagnostics[3]

Desrozier diagnostics use the following quantities calculated for a data assimilation run (all in observation space):

- Innovations (observation minus background): $\mathbf{d}_{\mathrm{b}}^{\mathrm{o}} = \mathbf{y} - \mathbf{H}\mathbf{x}_{\mathrm{b}}$.
- Analysis increment (analysis minus background):$\mathbf{d}_{\mathrm{b}}^{\mathrm{a}} = \mathbf{H}\delta\mathbf{x}_{\mathrm{a}}$.
- Residuals (observation minus analysis): $\mathbf{d}_{\mathrm{a}}^{\mathrm{o}} = \mathbf{y} - \mathbf{H}\mathbf{x}_{\mathrm{a}}$.

The covariances of these quantities reveals the consistency (or inconsistency) of the data assimilation. E.g. for 3D-VAR:

Covariance	Actual result (sub-optimal)	Result if optimal
$\mathcal{E}\{\mathbf{d}_{\mathrm{b}}^{\mathrm{o}}\mathbf{d}_{\mathrm{b}}^{\mathrm{o}\,\mathrm{T}}\}$	$\mathbf{R}+\mathbf{HBH}^{\mathrm{T}}$	$\mathbf{R}+\mathbf{HBH}^{\mathrm{T}}$
$\mathcal{E}\{\mathbf{d}_{\mathrm{b}}^{\mathrm{a}}\mathbf{d}_{\mathrm{b}}^{\mathrm{o}\,\mathrm{T}}\}$	$\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}+\hat{\mathbf{R}})^{-1}(\mathbf{R}+\mathbf{HBH}^{\mathrm{T}})$	$\mathbf{HBH}^{\mathrm{T}}$
$\mathcal{E}\{\mathbf{d}_{\mathrm{a}}^{\mathrm{o}}\mathbf{d}_{\mathrm{b}}^{\mathrm{o}\,\mathrm{T}}\}$	$(\mathbf{I}-\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}+\hat{\mathbf{R}})^{-1})(\mathbf{R}+\mathbf{HBH}^{\mathrm{T}})$	$\mathbf{R}$
$\mathcal{E}\{\mathbf{d}_{\mathrm{b}}^{\mathrm{a}}\mathbf{d}_{\mathrm{a}}^{\mathrm{o}\,\mathrm{T}}\}$	$\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}+\hat{\mathbf{R}})^{-1}(\mathbf{R}+\mathbf{HBH}^{\mathrm{T}})(\mathbf{I}-\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}+\hat{\mathbf{R}})^{-1})^{\mathrm{T}}$	$\mathbf{HAH}^{\mathrm{T}}$

Here $\mathbf{B}$ and $\mathbf{R}$ are the true background and observation error covariances matrices, and $\hat{\mathbf{B}}$ and $\hat{\mathbf{R}}$ are the ones assumed for the data assimilation. $\mathbf{H}$ is assumed perfect.

This section is provided for information only. Proofs of these results are as follows. Consider a sub-optimal variational data assimilation scheme where the specified statistics (indicated with hats) may have been given incorrectly. Consider the following analysis increment that result:

$$\delta\mathbf{x}_{\mathrm{a}} = \mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}} = \hat{\mathbf{K}}\mathbf{d}_{\mathrm{b}}^{\mathrm{o}},$$

where the Kalman gain used in the assimilation is

$$\hat{\mathbf{K}} = \hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}+\hat{\mathbf{R}})^{-1}, \tag{52}$$

and $\mathbf{d}_{\mathrm{b}}^{\mathrm{o}}$ is the innovation vector (observation minus background - see below). $\hat{\mathbf{B}}$ and $\hat{\mathbf{R}}$ are the (potentially incorrect) background and observation error covariance matrices that are actually specified in the data assimilation (and $\hat{\mathbf{K}}$ is the Kalman gain that

[3] Desroziers G., Berre L., Chapnik B., Poli P., 2005, Diagnostics of observation, background and analysis-error statistics in observation space. Q.J.R. Meteorol. Soc. 131, 3385-3396.

follows). $\mathbf{B}$ and $\mathbf{R}$ (without the hats) are the correct background and observation error covariance matrices and $\mathbf{K}$ is the correct Kalman gain (38) that follows. We now examine various 'difference statistics' in observation space.

O-B, A-B, O-A expressions

The 'observation minus background' difference in observation space is:

$$\mathbf{d}_{\mathrm{b}}^{\mathrm{o}} = \mathbf{y} - \mathbf{H}\mathbf{x}_{\mathrm{b}} \approx \varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{b}}, \tag{53}$$

where ε_{o} is the observation error, and ε_{b} is the background error as in (42). We now express other important differences in terms of the innovations. The 'analysis minus background' difference in observation space is:

$$\mathbf{d}_{\mathrm{b}}^{\mathrm{a}} = \mathbf{H}\delta\mathbf{x}_{\mathrm{a}} = \mathbf{H}\hat{\mathbf{K}}\mathbf{d}_{\mathrm{b}}^{\mathrm{o}}, \tag{54}$$

and the 'observation minus analysis' difference in observation space is:

$$\begin{aligned} \mathbf{d}_{\mathrm{a}}^{\mathrm{o}} &= \mathbf{y} - \mathbf{H}\mathbf{x}_{\mathrm{a}} = \mathbf{y} - \mathbf{H}(\mathbf{x}_{\mathrm{b}} + \delta\mathbf{x}_{\mathrm{a}}) \\ &= \mathbf{d}_{\mathrm{b}}^{\mathrm{o}} - \mathbf{H}\hat{\mathbf{K}}\mathbf{d}_{\mathrm{b}}^{\mathrm{o}} = (\mathbf{I} - \mathbf{H}\hat{\mathbf{K}})\mathbf{d}_{\mathrm{b}}^{\mathrm{o}}. \end{aligned} \tag{55}$$

The vector $\mathbf{d}_{\mathrm{b}}^{\mathrm{o}}$ is otherwise known as the 'innovation vector' and the vector $\mathbf{d}_{\mathrm{a}}^{\mathrm{o}}$ is otherwise known as the 'residual vector'. The key thing is that these vectors are measurable directly from an existing data assimilation system. We will now use their equivalents in the above to see what we can learn about the system.

Measured statistics

Now we have these expressions, let us look at their covariance statistics.

O-B \ O-B statistics Assuming that background and observation errors are uncorrelated, the covariance matrix between $\mathbf{d}_{\mathrm{b}}^{\mathrm{o}}$ and $\mathbf{d}_{\mathrm{b}}^{\mathrm{o}}$ are[4]:

[4] The $\mathcal{E}$ operator performs an average over a population of realizations of the assimilation system (as though we had access to results from parallel universes). In practice though we do not have access to parallel universes so instead the average is taken between pairs of different observations that have (say) similar separations.

$$\begin{aligned}\mathcal{E}\{\mathbf{d}^{\mathrm{o}}_{\mathrm{b}}\mathbf{d}^{\mathrm{o}\mathrm{T}}_{\mathrm{b}}\} &= \mathcal{E}\{(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{b}})(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{b}})^{\mathrm{T}}\},\\ &= \mathcal{E}\{\varepsilon_{\mathrm{o}}\varepsilon_{\mathrm{o}}^{\mathrm{T}}\} - \mathcal{E}\{\varepsilon_{\mathrm{o}}\varepsilon_{\mathrm{b}}^{\mathrm{T}}\}\mathbf{H}^{\mathrm{T}} - \mathbf{H}\mathcal{E}\{\varepsilon_{\mathrm{b}}\varepsilon_{\mathrm{o}}^{\mathrm{T}}\} + \mathbf{H}\mathcal{E}\{\varepsilon_{\mathrm{b}}\varepsilon_{\mathrm{b}}^{\mathrm{T}}\}\mathbf{H}^{\mathrm{T}},\\ &= \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}.\end{aligned} \tag{56}$$

A-B \ O-B statistics Using (54), (56) and (52), the covariance matrix between $\mathbf{d}^{\mathrm{a}}_{\mathrm{b}}$ and $\mathbf{d}^{\mathrm{o}}_{\mathrm{b}}$ are:

$$\begin{aligned}\mathcal{E}\{\mathbf{d}^{\mathrm{a}}_{\mathrm{b}}\mathbf{d}^{\mathrm{o}\mathrm{T}}_{\mathrm{b}}\} &= \mathbf{H}\hat{\mathbf{K}}\mathcal{E}\{\mathbf{d}^{\mathrm{o}}_{\mathrm{b}}\mathbf{d}^{\mathrm{o}\mathrm{T}}_{\mathrm{b}}\} = \mathbf{H}\hat{\mathbf{K}}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}),\\ &= \mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}} + \hat{\mathbf{R}})^{-1}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}).\end{aligned}$$

If $\hat{\mathbf{B}} = \mathbf{B}$ and $\hat{\mathbf{R}} = \mathbf{R}$ then this becomes

$$\mathcal{E}\{\mathbf{d}^{\mathrm{a}}_{\mathrm{b}}\mathbf{d}^{\mathrm{o}\mathrm{T}}_{\mathrm{b}}\} = \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}) = \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}. \tag{57}$$

O-A \ O-B statistics Using (55), (56) and (52), the covariance matrix between $\mathbf{d}^{\mathrm{o}}_{\mathrm{a}}$ and $\mathbf{d}^{\mathrm{o}}_{\mathrm{b}}$ are:

$$\begin{aligned}\mathcal{E}\{\mathbf{d}^{\mathrm{o}}_{\mathrm{a}}\mathbf{d}^{\mathrm{o}\mathrm{T}}_{\mathrm{b}}\} &= (\mathbf{I} - \mathbf{H}\hat{\mathbf{K}})\mathcal{E}\{\mathbf{d}^{\mathrm{o}}_{\mathrm{b}}\mathbf{d}^{\mathrm{o}\mathrm{T}}_{\mathrm{b}}\} = (\mathbf{I} - \mathbf{H}\hat{\mathbf{K}})(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}),\\ &= (\mathbf{I} - \mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}} + \hat{\mathbf{R}})^{-1})(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}).\end{aligned}$$

If $\hat{\mathbf{B}} = \mathbf{B}$ and $\hat{\mathbf{R}} = \mathbf{R}$ then this becomes:

$$\mathcal{E}\{\mathbf{d}^{\mathrm{o}}_{\mathrm{a}}\mathbf{d}^{\mathrm{o}\mathrm{T}}_{\mathrm{b}}\} = (\mathbf{I} - \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1})(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}) = \mathbf{R}. \tag{58}$$

A-B \ O-A statistics Using (54), (55), (56) and (52), the covariance matrix between $\mathbf{d}^{\mathrm{a}}_{\mathrm{b}}$ and $\mathbf{d}^{\mathrm{o}}_{\mathrm{a}}$ are:

$$\begin{aligned}\mathcal{E}\{\mathbf{d}^{\mathrm{a}}_{\mathrm{b}}\mathbf{d}^{\mathrm{o}\mathrm{T}}_{\mathrm{a}}\} &= \mathbf{H}\hat{\mathbf{K}}\mathcal{E}\{\mathbf{d}^{\mathrm{o}}_{\mathrm{b}}\mathbf{d}^{\mathrm{o}\mathrm{T}}_{\mathrm{b}}\}(\mathbf{I} - \mathbf{H}\hat{\mathbf{K}})^{\mathrm{T}} = \mathbf{H}\hat{\mathbf{K}}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}})(\mathbf{I} - \mathbf{H}\hat{\mathbf{K}})^{\mathrm{T}},\\ &= \mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}} + \hat{\mathbf{R}})^{-1}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}})(\mathbf{I} - \mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^{\mathrm{T}} + \hat{\mathbf{R}})^{-1})^{\mathrm{T}}.\end{aligned}$$

If $\hat{\mathbf{B}} = \mathbf{B}$ and $\hat{\mathbf{R}} = \mathbf{R}$ then this becomes:

$$\mathcal{E}\{\mathbf{d}^{\mathrm{a}}_{\mathrm{b}}\mathbf{d}^{\mathrm{o}\mathrm{T}}_{\mathrm{a}}\} = \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{I} - \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1})^{\mathrm{T}}.$$

By writing $\mathbf{I} = (\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}$ then the above becomes:

$$\begin{aligned}
\mathcal{E}\{\mathbf{d}_{\mathrm{b}}^{\mathrm{a}}\mathbf{d}_{\mathrm{a}}^{\mathrm{oT}}\} &= \mathbf{HBH}^{\mathrm{T}}((\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1} - \mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1})^{\mathrm{T}}, \\
&= \mathbf{HBH}^{\mathrm{T}}([\mathbf{HBH}^{\mathrm{T}} + \mathbf{R} - \mathbf{HBH}^{\mathrm{T}}](\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1})^{\mathrm{T}}, \\
&= \mathbf{HBH}^{\mathrm{T}}(\mathbf{R}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1})^{\mathrm{T}}, \\
&= \mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{R}.
\end{aligned}$$

Note that the inverse Hessian has the form $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$ and the Sherman-Morrison-Woodbury formula in terms of $\mathbf{A}^{-1}$ is $\mathbf{A}^{-1}\mathbf{BH}^{\mathrm{T}} = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{R} + \mathbf{HBH}^{\mathrm{T}})$. This makes the above into:

$$\mathcal{E}\{\mathbf{d}_{\mathrm{b}}^{\mathrm{a}}\mathbf{d}_{\mathrm{a}}^{\mathrm{oT}}\} = \mathbf{HAH}^{\mathrm{T}}, \tag{59}$$

which is the analysis error covariance matrix in observation space.

These results are important because they allow the error statistics to be checked. If (57), (58) or (59) are not satisfied then the assumptions that $\hat{\mathbf{B}} = \mathbf{B}$ and $\hat{\mathbf{R}} = \mathbf{R}$ may not be correct. Even in this case, these equations can help us to improve the error statistics in the ways discussed in the Desroziers et al. paper.