Variational Data Assimilation

Background and Methods[1]

Quanling Deng

YMSC, Tsinghua University

Dec. 4, 2025

[1] Based on lecture slides by Ross Bannister & Amos Lawless (U. of Reading)

Real global NWP systems and their DA schemes

Centre & model	Role	Operational DA scheme(s)
ECMWF – IFS	Global medium-range NWP and seasonal prediction	Multi-resolution **4DVar** with tangent-linear physics; hybrid with ensemble (EDA/EnKF) to provide flow-dependent B
Met Office – Unified Model (Global)	Global NWP and ensembles	**Hybrid 4DVar** (climatological B + MOGREPS ensemble B); development of hybrid **4DEnVar** for global and UK models
NOAA/NCEP – GFS (FV3 core)	Global NWP and ensembles (GDAS/GFS)	**GSI**-based hybrid **4DEnVar** (EnVar with ensemble covariances); moving toward JEDI-based hybrid systems
JMA – Global Spectral Model	Global NWP and typhoon forecasting	**4DVar** with hybrid **LETKF/4DVar** (ensemble covariances from GEPS LETKF)
NASA GMAO – GEOS	Global NWP and atmospheric reanalyses	Hybrid **4DEnVar** atmospheric DA system for GEOS (variational + ensemble)

Notes: All of these systems are strongly influenced by incremental 4DVar ideas, but increasingly blend variational and ensemble DA to obtain flow-dependent error covariances.

Real global climate / Earth system models and DA usage

Climate models based on operational NWP systems

- **EC-Earth (EC-Earth consortium)**
  Atmosphere uses ECMWF IFS (seasonal prediction configuration); coupled with NEMO ocean and LIM sea ice. For seasonal/decadal prediction, initial conditions are typically taken from the IFS-based DA system (4DVar + ensemble hybrid) used for ECMWF forecasts.
- **HadGEM3 / UKESM1 (Met Office + UK partners)**
  Global coupled climate and Earth-system models (HadGEM3-GC3.1 is the physical core of UKESM1). Initialised climate predictions and reforecasts are driven by the Met Office global DA system (hybrid 4DVar / 4DEnVar) used for NWP, then integrated freely for climate time-scales.

Earth-system DA frameworks for coupled models

- **CESM (NCAR) + DART**
  CESM is coupled to the **Data Assimilation Research Testbed (DART)**, which provides ensemble adjustment Kalman filter (EAKF) and EnKF schemes for atmosphere and ocean, often in a weakly-coupled configuration for seasonal to decadal prediction.
- **MIROC (JAMSTEC / Univ. Tokyo / NIES)**
  MIROC climate models (e.g. MIROC3m, MIROC4h, MIROC5, MIROC6) use **ocean data assimilation** (e.g. hydrographic anomalies with incremental analysis update) to initialise decadal climate predictions, followed by free coupled integrations.

In most long CMIP-style climate projections, the GCM/ESM runs without DA; DA is crucial mainly for reanalyses and initialised prediction (seasonal–decadal).

Bayes' Theorem

$$p(x \mid y) = \frac{p(x)\, p(y \mid x)}{p(y)}$$

Interpretation

$$\underbrace{p(x \mid y)}_{\text{posterior distribution}} = \frac{\underbrace{p(x)}_{\text{prior distribution}} \times \underbrace{p(y \mid x)}_{\text{likelihood}}}{\underbrace{p(y)}_{\text{normalizing constant}}}.$$

- **Prior distribution**: PDF of the state *before* observations are considered (e.g. PDF of a model forecast).
- **Likelihood**: PDF of observations given that the state is x.
- **Posterior**: PDF of the state *after* the observations have been considered.

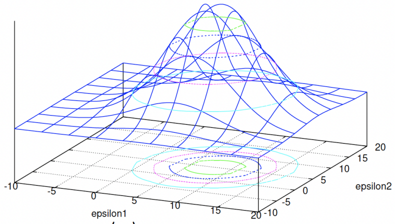
The Gaussian assumption

- PDFs are often described by Gaussians (normal distributions).
- Gaussian PDFs are described by a mean and covariance only.

For one variable (1D): $x \sim \mathcal{N}(\langle x \rangle, \sigma^2)$,

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \langle x \rangle)^2}{2\sigma^2} \right).$$

For n variables (nD): $x \sim \mathcal{N}(\langle x \rangle, C)$,



$$\varepsilon = \mathbf{x} - \langle \mathbf{x} \rangle$$

$$P(x) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} \exp\left[-\frac{1}{2} (x - \langle x \rangle)^{\mathrm{T}} C^{-1} (x - \langle x \rangle) \right].$$

Meaning of x and y

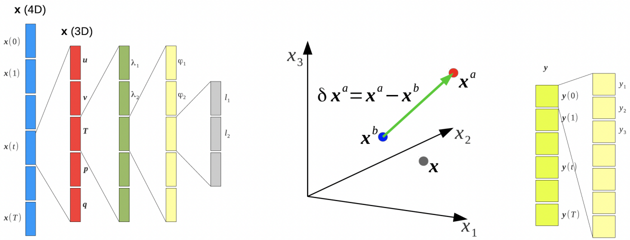


Figure: Illustration of states, increments, and observation vectors.

- x^a: analysis; x^b: background state;
- y: observations; δx: increment (perturbation).
- $y^m = \mathcal{H}(x)$: model (or simulated) observations.
- $\mathcal{H}(x)$ is the observation operator / forward model.
- Sometimes x and y are for only one time (3DVar).
- x-vectors have n elements; y-vectors have p elements.

Back to the Gaussian assumption

Prior with mean x^b and covariance B:

$$P(x) = \frac{1}{\sqrt{(2\pi)^n \det(B)}} \exp\left[-\frac{1}{2}(x - x^b)^{\mathrm{T}} B^{-1}(x - x^b)\right].$$

Likelihood with mean $\mathcal{H}(x)$ and covariance R:

$$P(y \mid x) = \frac{1}{\sqrt{(2\pi)^p \det(R)}} \exp\left[-\frac{1}{2}(y - \mathcal{H}(x))^{\mathrm{T}} R^{-1}(y - \mathcal{H}(x))\right].$$

Posterior:

$$p(x \mid y) = \frac{p(x)\, p(y \mid x)}{p(y)} \propto \exp\left[-\frac{1}{2}\left\{(x - x^b)^{\mathrm{T}} B^{-1}(x - x^b) + (y - \mathcal{H}(x))^{\mathrm{T}} R^{-1}(y - \mathcal{H}(x))\right\}\right].$$

Variational DA – the idea

In variational (Var) methods, we seek a solution that maximizes the posterior probability $p(x \mid y)$ (maximum-a-posteriori).

- This is the most likely state, given the observations and the background: the **analysis** x^a.
- Maximizing $p(x \mid y)$ is equivalent to minimizing $-\ln p(x \mid y)$, which we denote by $J(x)$ (a least-squares problem).

$$p(x \mid y) = C \exp\left[-\frac{1}{2} \left\{ (x - x^b)^{\mathrm{T}} B^{-1} (x - x^b) + (y - \mathcal{H}(x))^{\mathrm{T}} R^{-1} (y - \mathcal{H}(x)) \right\} \right],$$

so

$$\begin{aligned} J(x) &= -\ln C + \frac{1}{2}(x - x^b)^{\mathrm{T}} B^{-1} (x - x^b) + \frac{1}{2}(y - \mathcal{H}(x))^{\mathrm{T}} R^{-1} (y - \mathcal{H}(x)) \\ &= \text{constant (ignored)} + J_b(x) + J_o(x), \end{aligned}$$

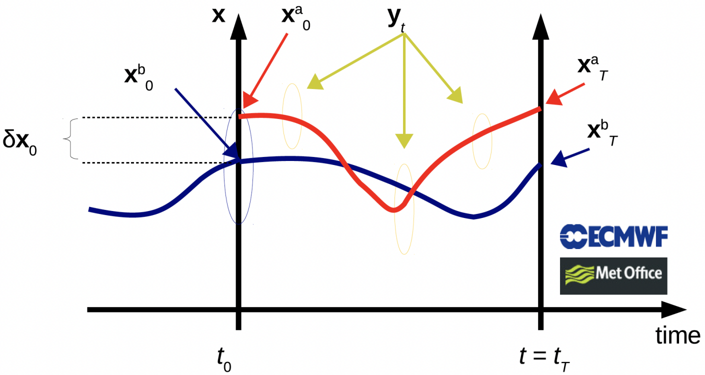
where J_b is the background term and J_o is the observation term.

Four-dimensional Var (4DVar)

Aim

To find the *best* estimate of the true state of the system (analysis), consistent with

- the observations, the background (prior),
- and the system dynamics.



Towards a 4DVar cost function

Consider the observation operator in the 4D case:

$$\mathcal{H}(x) = \mathcal{H}\begin{pmatrix} x_0 \\ \vdots \\ x_T \end{pmatrix} = \begin{pmatrix} \mathcal{H}_0(x_0) \\ \vdots \\ \mathcal{H}_T(x_T) \end{pmatrix}.$$

Assume that R is block diagonal:

$$R = \begin{pmatrix} R_0 & 0 & \cdots & 0 \\ 0 & R_1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & R_T \end{pmatrix}.$$

Then the observation part of the cost function is

$$\begin{aligned} J_o &= \frac{1}{2}(y - \mathcal{H}(x))^{\mathrm{T}} R^{-1}(y - \mathcal{H}(x)) \\ &= \frac{1}{2}\sum_{i=0}^{T}\big(y_i - \mathcal{H}_i(x_i)\big)^{\mathrm{T}} R_i^{-1}\big(y_i - \mathcal{H}_i(x_i)\big), \end{aligned}$$

subject to the model constraint $x_{i+1} = \mathcal{M}_i(x_i)$, where $\mathcal{M}_i$ is the model.

The 4DVar cost function ("full 4DVar")

The total cost function can be written as

$$\begin{aligned} J(x) &= \frac{1}{2}(x_0 - x_0^b)^{\mathrm{T}} B_0^{-1} (x_0 - x_0^b) + \frac{1}{2}(y - \mathcal{H}(x))^{\mathrm{T}} R^{-1} (y - \mathcal{H}(x)) \\ &= \frac{1}{2}(x_0 - x_0^b)^{\mathrm{T}} B_0^{-1} (x_0 - x_0^b) + \frac{1}{2}\sum_{i=0}^{T} \big(y_i - \mathcal{H}_i(x_i)\big)^{\mathrm{T}} R_i^{-1} \big(y_i - \mathcal{H}_i(x_i)\big), \end{aligned}$$

subject to the constraint

$$x_{i+1} = \mathcal{M}_i(x_i).$$

Here

- x_0^b: a-priori (background) state at t_0,
- y_i: observations at time t_i,
- $\mathcal{H}_i(x_i)$: observation operator at t_i,
- B_0: background error covariance matrix at t_0,
- R_i: observation error covariance matrix at t_i.

How to minimize this cost function?

We minimize $J(x)$ iteratively.

Generic iteration:

$$x_0^{k+1} = x_0^k + \alpha \nabla J(x_0^k),$$

where α is a step length and the gradient

$$\nabla J(x_0) = \begin{pmatrix} \partial J / \partial (x_0)_1 \\ \vdots \\ \partial J / \partial (x_0)_n \end{pmatrix}.$$

- $-\nabla J$ points in the direction of steepest descent.
- Methods:
  - Steepest descent (inefficient),
  - Conjugate gradient (more efficient),
  - and others.

The gradient of the cost function (with respect to $x(t_0)$)

Two equivalent approaches:

1. Minimize $J(x_0)$ with respect to x_0 with

$$x_i = \mathcal{M}_{i-1}\big(\mathcal{M}_{i-2}(\cdots \mathcal{M}_0(x_0)\cdots)\big).$$

2. Minimize $J(x_0, \ldots, x_T)$ with respect to $x_0, \ldots, x_T$ subject to the constraint

$$x_{i+1} - \mathcal{M}_i(x_i) = 0.$$

   Introduce a Lagrangian

$$\mathcal{L}(x, \lambda) = J(x) + \sum_{i=0}^{T-1} \lambda_{i+1}^{\mathrm{T}}\big(x_{i+1} - \mathcal{M}_i(x_i)\big),$$

   where λ_i are Lagrange multipliers.

Each approach leads to the **adjoint method**:

- an efficient means of computing the gradient;
- uses the linearized and adjoint operators of $\mathcal{M}_i$ and $\mathcal{H}_i$.

The adjoint method

An equivalent gradient formula is

$$\begin{aligned}\nabla J &\equiv \nabla J(x_0)\\ &= \nabla J_b(x_0) + \nabla J_o(x_0)\\ &= B_0^{-1}(x_0 - x_0^b) - \sum_{i=0}^{T} M_0^{\mathrm{T}} \cdots M_{i-1}^{\mathrm{T}} H_i^{\mathrm{T}} R_i^{-1}\big(y_i - \mathcal{H}_i(x_i)\big),\end{aligned}$$

where

$$M_i = \frac{\partial \mathcal{M}_i(x_i)}{\partial x_i}, \qquad H_i = \frac{\partial \mathcal{H}_i(x_i)}{\partial x_i}.$$

Alternatively, define the adjoint variables λ_i:

$$\begin{aligned}\lambda_{T+1} &= 0,\\ \lambda_i &= H_i^{\mathrm{T}} R_i^{-1}\big(y_i - \mathcal{H}_i(x_i)\big) + M_i^{\mathrm{T}}\lambda_{i+1},\end{aligned}$$

so that

$$\lambda_0 = \nabla J_o \quad \text{and} \quad \nabla J = \nabla J_b + \nabla J_o = B_0^{-1}(x_0 - x_0^b) + \lambda_0.$$

Simplifications and complications

The full 4DVar method is *expensive* and *difficult* to solve.

- The model $\mathcal{M}_i$ is nonlinear.
- Observation operators $\mathcal{H}_i$ can be nonlinear.

Cost function shape

- Linear $H \Rightarrow$ quadratic cost function – easier to minimize: $J_o \sim \frac{1}{2}\frac{(y-ax)^2}{\sigma_o^2}$.
- Nonlinear $H \Rightarrow$ non-quadratic cost function – harder to minimize: $J_o \sim \frac{1}{2}\frac{(y-f(x))^2}{\sigma_o^2}$.

- Later, we will recognize that models are *wrong*!

Look for simplifications:

- Incremental 4DVar (linearized)
- 3DVar/3D-FGAT (imperfect)

Complications:

- Weak-constraint 4DVar

Incremental 4DVar (1)

Define a **reference trajectory** x_i^R:

$$x_{i+1}^R = \mathcal{M}_i\left(x_i^R\right).$$

Write

$$x_i = x_i^R + \delta x_i, \qquad x_0^b = x_0^R + \delta x_0^b.$$

The model equation $x_{i+1} = \mathcal{M}_i(x_i)$ becomes, after linearization about x_i^R,

$$x_{i+1}^R + \delta x_{i+1} \approx \mathcal{M}_i(x_i^R) + M_i \delta x_i,$$

so that $\delta x_{i+1} \approx M_i \delta x_i$.

For the observations:

$$y_i^m = \mathcal{H}_i(x_i),$$

linearizing about x_i^R gives

$$y_i^{m,R} + \delta y_i^m \approx \mathcal{H}_i(x_i^R) + H_i \delta x_i,$$

so

$$\delta y_i^m \approx H_i \delta x_i.$$

Incremental 4DVar (2)

The incremental 4DVar cost function is

$$J(\delta x_0) = \frac{1}{2}(\delta x_0 - \delta x_0^b)^{\mathrm{T}} B_0^{-1} (\delta x_0 - \delta x_0^b) + \frac{1}{2}\sum_{i=0}^{T} \big(y_i - \mathcal{H}_i(x_i^R) - H_i \delta x_i\big)^{\mathrm{T}} R_i^{-1} \big(y_i - \mathcal{H}_i(x_i^R) - H_i \delta x_i\big),$$

with $\delta x_i \approx M_{i-1} M_{i-2} \cdots M_0 \, \delta x_0$.

- **Inner loop**: iterations to find

  $$\delta x_0^a = \arg\min_{\delta x_0} J(\delta x_0)$$

  (use adjoint method).
- **Outer loop**: update the reference: $x_0^R \leftarrow x_0^R + \delta x_0^a$ and recompute trajectory.

- Inner loop is exactly quadratic (e.g. has a unique minimum).
- Inner loop can be simplified (e.g. lower resolution, simplified physics).

Simplification 1: 3D-FGAT

3D-FGAT = Three-dimensional variational data assimilation with the *first guess* (the reference trajectory x_i^R) computed at the appropriate time.

Simplification:

$$M_i \to I, \quad \text{i.e.} \quad \delta x_i = M_{i-1} \cdots M_0 \delta x_0 \to \delta x_0.$$

The cost function becomes

$$\begin{aligned} J_{\text{3DFGAT}}(\delta x_0) = & \frac{1}{2}(\delta x_0 - \delta x_0^b)^{\mathrm{T}} B_0^{-1} (\delta x_0 - \delta x_0^b) \\ & + \frac{1}{2} \sum_{i=0}^{T} \big(y_i - \mathcal{H}_i(x_i^R) - H_i \delta x_0\big)^{\mathrm{T}} R_i^{-1} \big(y_i - \mathcal{H}_i(x_i^R) - H_i \delta x_0\big). \end{aligned}$$

Simplification 2: 3DVar

3DVar has no time dependence within the assimilation window.

- Not used (these days "3DVar" usually means 3D-FGAT).

The 3DVar cost function (dropping the time index) is

$$J_{3\mathrm{DVar}}(\delta x_0) = \frac{1}{2}(\delta x_0 - \delta x_0^b)^{\mathrm{T}} B_0^{-1}(\delta x_0 - \delta x_0^b) + \frac{1}{2}\sum_{i=0}^{T}\Big(y_i - \mathcal{H}_i(x_0^R) - H_i \delta x_0\Big)^{\mathrm{T}} R_i^{-1}(\cdot).$$

3DVar is *not* an approx. All obs. are at $t = 0$. For $x_0^R = x_0^b$:

$$J_{3\mathrm{DVar}}(\delta x_0) = \frac{1}{2}\delta x_0^{\mathrm{T}} B_0^{-1}\delta x_0 + \frac{1}{2}\Big(y_0 - \mathcal{H}_0(x_0^b) - H_0\delta x_0\Big)^{\mathrm{T}} R_0^{-1}(\cdot).$$

3DVar and the Kalman Filter

Setting $\nabla J_{\mathrm{3DVar}} = 0$ gives

$$B_0^{-1}\delta x_0 - H_0^{\mathrm{T}} R_0^{-1}\big(y_0 - \mathcal{H}_0(x_0^b) - H_0\delta x_0\big) = 0.$$

Solving for δx_0 leads to

$$x_0^a = x_0^b + \delta x_0 = x_0^b + \big(B_0^{-1} + H_0^{\mathrm{T}} R_0^{-1} H_0\big)^{-1} H_0^{\mathrm{T}} R_0^{-1}\big(y_0 - \mathcal{H}_0(x_0^b)\big).$$

This is equivalent to the Kalman filter analysis:

$$x_0^a = x_0^b + B_0 H_0^{\mathrm{T}}\big(R_0 + H_0 B_0 H_0^{\mathrm{T}}\big)^{-1}\big(y_0 - \mathcal{H}_0(x_0^b)\big).$$

Properties of 4DVar

- Observations are treated at the *correct time*.
- Use of dynamics means that more information can be obtained from observations.
- The covariance B_0 is implicitly evolved:

$$B_i = (M_{i-1} \cdots M_0)\, B_0\, (M_{i-1} \cdots M_0)^{\mathrm{T}}.$$

- In practice, development of linear and adjoint models is complex.
- M_i, H_i, M_i , H_i , M_i^{T}, and H_i^{T} are usually implemented as subroutines, so the “matrices” are not in explicit matrix form.

Model error

- Standard 4DVar assumes the model is *perfect*.
- This can lead to sub-optimalities.
- Weak-constraint 4DVar relaxes this assumption.

Weak-constraint 4DVar

Modify the evolution equation to include model error:

$$x_{i+1} = \mathcal{M}_i(x_i) + \eta_i,$$

where $\eta_i \sim \mathcal{N}(0, Q_i)$.

State formulation of WC4DVar:

$$J_{\text{wc}}(x_0, \ldots, x_T) = J_b + J_o + \frac{1}{2}\sum_{i=0}^{T-1}\big(x_{i+1} - \mathcal{M}_i(x_i)\big)^{\mathrm{T}} Q_i^{-1}\big(x_{i+1} - \mathcal{M}_i(x_i)\big).$$

Error formulation of WC4DVar:

$$J_{\text{wc}}(x_0, \eta_0, \ldots, \eta_{T-1}) = J_b + J_o + \frac{1}{2}\sum_{i=0}^{T-1}\eta_i^{\mathrm{T}} Q_i^{-1}\eta_i.$$

View it as an extra penalty term.

Implementation of weak-constraint 4DVar

- The vector to be determined (the **control vector**) increases from size n in 4DVar to $n + n(T-1)$ in WC4DVar.
- The model error covariance matrices Q_i need to be estimated.
  - This is a non-trivial problem: *How?*
- The state formulation (determine $x_0, \ldots, x_T$) and the error formulation (determine $x_0, \eta_0, \ldots, \eta_{T-1}$) are mathematically equivalent, but can behave differently in practice.
- There is an incremental form of WC4DVar, analogous to incremental 4DVar.

Summary of 4DVar

- The variational method forms the basis of many operational weather and ocean forecasting systems (e.g. ECMWF, Met Office, Météo-France, etc.).
- It allows complicated observation operators to be used (e.g. assimilation of satellite data).
- It has been very successful.
- Incremental (quasi-linear) versions are usually implemented.
- It requires specification of
  - B_0, the background error covariance matrix,
  - R_i, the observation error covariance matrices.
- 4DVar requires the development of linear and adjoint models – *not* a simple task!
- Weak-constraint formulations additionally require specification of Q_i.

Some challenges ahead

Methods generally assume that error covariance matrices are correctly known.

- Representing B_0:
  - Better models of B_0,
  - Flow dependency (e.g. Ensemble-Var or hybrid methods).
- Representing R_i:
  - Allowing for observation error covariances.
- Representing Q_i.
- Numerical conditioning of the problem.
- Application to more complicated systems, e.g.:
  - high-resolution models,
  - coupled atmosphere–ocean DA,
  - chemical data assimilation.
- Variational bias correction.
- Moist processes, including clouds.
- Effective use on massively parallel computer architectures.

Selected references

Original application of 4DVar: Talagrand O., Courtier P., Variational assimilation of meteorological observations with the adjoint vorticity equation I: Theory, *Q. J. R. Meteorol. Soc.* **113**, 1311–1328 (1987).

Excellent tutorial on Var: Schlatter T. W., Variational assimilation of meteorological observations in the lower atmosphere: A tutorial on how it works, *J. Atmos. Sol. Terr. Phys.* **62**, 1057–1070 (2000).

Incremental 4DVar: Courtier P., Thépaut J.-N., Hollingsworth A., A strategy for operational implementation of 4D-Var using an incremental approach, *Q. J. R. Meteorol. Soc.* **120**, 1367–1387 (1994).

High-resolution application of 4DVar: Park S. K., Zupanski D., Four-dimensional variational data assimilation for mesoscale and storm scale applications, *Meteorol. Atmos. Phys.* **82**, 173–208 (2003).

Selected references (cont.)

Met Office 4DVar: Rawlins F., Ballard S. P., Bovis K. J., Clayton A. M., Li D., Inverarity G. W., Lorenc A. C., Payne T. J., The Met Office global four-dimensional variational data assimilation scheme, *Q. J. R. Meteorol. Soc.* **133**, 347–362 (2007).

Weak-constraint 4DVar: Trémolet Y., Model-error estimation in 4D-Var, *Q. J. R. Meteorol. Soc.* **133**, 1267–1280 (2007).

Inner and outer loops: Lawless A. S., Gratton S., Nichols N. K., *QJRMS* (2005); Gratton S., Lawless A. S., Nichols N. K., *SIAM J. Optim.* (2007).

Survey of variational and hybrid methods: Bannister R. N., A review of operational methods of variational and ensemble-variational data assimilation, *Q. J. R. Meteorol. Soc.* **143**, 607–633 (2017).

Complementary slides

Making variational DA work – control variable transforms

B_0 is an $n \times n$ matrix.

- In operational problems, B_0 is too large to store, let alone invert.
- It is an *unknowable* matrix.

We can model the essential features of B_0 with a change of variable (a **control variable transform**):

$$\delta x = U\,\delta v.$$

- Hypothesis: the problem is much simpler when posed in terms of δv rather than δx.

Idea

- Rewrite the cost function in terms of δv.
- Minimize w.r.t. δv (with a trivial B-matrix).
- Convert back to δx and iterate.

This is equivalent to solving the original problem w.r.t. δx with $B_0 = UU^{\mathrm{T}}$

Control variable transforms (CVTs) in more detail

Illustrate in the simplest case: 3DVar with $x_0^R = x_0^b$ (drop the time index).

The 3DVar cost function is

$$J_{3\mathrm{DVar}}(\delta x) = \frac{1}{2}\delta x^{\mathrm{T}} B^{-1} \delta x + \frac{1}{2}\big(y - \mathcal{H}(x^b) - H\delta x\big)^{\mathrm{T}} R^{-1} \big(y - \mathcal{H}(x^b) - H\delta x\big).$$

Make the change of variable $\delta x = U\delta v$ and assume that the B-matrix of errors in the δv representation is identity.
Denote background covariance by

$$B = \left\langle \delta x \delta x^{\mathrm{T}} \right\rangle_b = U \left\langle \delta v \delta v^{\mathrm{T}} \right\rangle_b U^{\mathrm{T}} = UIU^{\mathrm{T}} = UU^{\mathrm{T}}.$$

CVTs in more detail (cont.)

Substitute $\delta x = U\delta v$ into J_{3DVar} to get a cost function in terms of δv:

$$J_{\text{3DVar}}(\delta v) = \frac{1}{2}\delta v^{\mathrm{T}}\delta v + \frac{1}{2}\big(y - \mathcal{H}(x^b) - HU\delta v\big)^{\mathrm{T}} R^{-1}\big(y - \mathcal{H}(x^b) - HU\delta v\big).$$

The gradient w.r.t. δv is

$$\nabla_{\delta v} J_{\text{3DVar}} = \delta v - U^{\mathrm{T}} H^{\mathrm{T}} R^{-1}\big(y - \mathcal{H}(x^b) - HU\delta v\big).$$

The analysis is then

$$x^a = x^b + U \arg\min_{\delta v} J_{\text{3DVar}}(\delta v).$$

Estimating a B-matrix

Let

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

The background error covariance matrix is

$$B = \left\langle (x^b - x^t)(x^b - x^t)^{\mathrm{T}} \right\rangle_b.$$

In component form

$$B = \begin{pmatrix} \langle (x_1^b - x_1^t)^2 \rangle_b & \cdots & \langle (x_1^b - x_1^t)(x_n^b - x_n^t) \rangle_b \\ \vdots & \ddots & \vdots \\ \langle (x_n^b - x_n^t)(x_1^b - x_1^t) \rangle_b & \cdots & \langle (x_n^b - x_n^t)^2 \rangle_b \end{pmatrix}.$$

Here $\langle \cdot \rangle_b$ denotes an average over a population of possible backgrounds.

Problem: The truth x^t appearing in the definition of the error $\varepsilon = x^b - x^t$ is *unknowable*, so we need a proxy.

Approaches to estimating a B-matrix (1)

"Canadian quick" method

Assume

$$x^b - x^t \sim \frac{x^b(t+T) - x^b(t)}{\sqrt{2}}.$$

- Take the population from one long time run of the model.
- Use differences of forecasts separated by a fixed time lag T.
- Reference: Polavarapu et al. (2005).

Approaches to estimating a B-matrix (2)

Analysis of innovations

Choose a pair of *direct/independent* observation locations separated by Δr:

$$\begin{pmatrix} y_r - x_r^b \\ y_{r+\Delta r} - x_{r+\Delta r}^b \end{pmatrix} = \begin{pmatrix} y_r - x_r^t \\ y_{r+\Delta r} - x_{r+\Delta r}^t \end{pmatrix} - \begin{pmatrix} x_r^b - x_r^t \\ x_{r+\Delta r}^b - x_{r+\Delta r}^t \end{pmatrix} = \begin{pmatrix} \varepsilon_r^y - \varepsilon_r^b \\ \varepsilon_{r+\Delta r}^y - \varepsilon_{r+\Delta r}^b \end{pmatrix}$$

Taking the expectation gives

$$\left\langle (\varepsilon_r^y - \varepsilon_r^b)(\varepsilon_{r+\Delta r}^y - \varepsilon_{r+\Delta r}^b) \right\rangle = \langle \varepsilon_r^y \varepsilon_{r+\Delta r}^y \rangle + \langle \varepsilon_r^b \varepsilon_{r+\Delta r}^b \rangle = \sigma_o^2 \delta_{\Delta r,0} + \sigma_b^2 \operatorname{cor}_b(\Delta r).$$

Above assumes observation and background errors are uncorrelated, as are errors between observations at different locations.

- Take the population from many pairs with the same Δr.
- References: Rutherford (1972), Hollingsworth and Lönnberg (1986), Järvinen (2001).

Approaches to estimating a B-matrix (3)

National Meteorological Center (NMC) method

Choose pairs of lagged forecasts valid at the same time, e.g.:

$$x^b - x^t \sim \frac{x^b_{48}(t) - x^b_{24}(t)}{\sqrt{2}},$$

where subscripts indicate forecast lead-times (48 h, 24 h).

- Take the population from such differences at many times.
- References: Parrish and Derber (1992), Berre et al. (2006).

Approaches to estimating a B-matrix (4)

Ensemble method

If you have an ensemble that is correctly spread, then

$$x^b - x^t \sim x^{b,(i)} - \langle x^b \rangle$$

or

$$x^b - x^t \sim \frac{x^{b,(i)} - x^{b,(j)}}{\sqrt{2}}.$$

- Take the population from ensemble members and over many times.
- References: Houtekamer et al. (1996), Buehner (2005), Bonavita et al. (2015).

References (control variable transforms and B-matrix methods)

Control variable transforms:
Bannister R. N., A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics, *Q. J. R. Meteorol. Soc.* **134**, 1971–1996 (2008).

Canadian "quick" method:
Polavarapu S., Ren S., Rochon Y., Sankey D., Ek N., Koshyk J., Tarasick D., Data assimilation with the Canadian middle atmosphere model, *Atmos.-Ocean* **43**, 77–100 (2005).

Analysis of innovations:
Hollingsworth A., Lönnberg P., The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, *Tellus* **38A**, 111–136 (1986).

Järvinen H., Temporal evolution of innovation and residual statistics in the ECMWF variational data assimilation systems, *Tellus* **53A**, 333–347 (2001).

References (cont.)

NMC method:
Parrish D. F., Derber J. C., The National Meteorological Center's spectral statistical interpolation analysis system, *Mon. Wea. Rev.* **120**, 1747–1763 (1992).

NMC method:
Berre L., Ştefănescu S. E., Pereira M. B., The representation of the analysis effect in three error simulation techniques, *Tellus* **58A**, 196–209 (2006).

Ensemble method:
Houtekamer P. L., Lefaivre L., Derome J., Ritchie H., Mitchell H. L., A system simulation approach to ensemble prediction, *Mon. Wea. Rev.* **124**, 1225–1242 (1996).

Buehner M., Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operational NWP setting, *Q. J. R. Meteorol. Soc.* **131**, 1013–1043 (2005).

Bonavita M., Holm E., Isaksen L., Fisher M., The evolution of the ECMWF hybrid data assimilation system, *Q. J. R. Meteorol. Soc.* (2015).