Background Error Covariance Modelling

Data Assimilation Training Course

Marcin Chrust

European Centre for Medium-Range Weather Forecasts

March 2026

Acknowledgements: Elias Hólm

Outline

Foundations

Definitions

State vector: $\mathbf{x}$ contains all analysed model variables at all grid points and levels

Background and analysis:

- Background (first guess): $\mathbf{x}_b$
- Analysis: $\mathbf{x}_a$
- True state (unknown): $\mathbf{x}_t$
- Background error: $\boldsymbol{\epsilon}^b = \mathbf{x}_b - \mathbf{x}_t$
- Observation error: $\boldsymbol{\epsilon}^o = \mathbf{y}^o - \mathcal{H}(\mathbf{x}_t)$

Error covariances:

$$\mathbf{B} = \mathbb{E}[\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^b)^T], \quad \mathbf{R} = \mathbb{E}[\boldsymbol{\epsilon}^o(\boldsymbol{\epsilon}^o)^T]$$

Key point: $\mathbf{B}$ and $\mathbf{R}$ define how information is weighted and spread

Observation-Error Assumptions and Implications

Common assumptions:

- Uncorrelated observation errors (diagonal $\mathbf{R}$)
- Errors unbiased and independent of background errors
- Representativeness error included in $\mathbf{R}$

Meaning: $\mathbb{E}[\boldsymbol{\epsilon}^o] = 0$ and $\mathbb{E}[\boldsymbol{\epsilon}^o(\boldsymbol{\epsilon}^b)^T] = 0$

Reality check:

- Many instruments have spatial/temporal correlations
- Mis-specified $\mathbf{R}$ distorts diagnosed $\mathbf{B}$
- Correlated $\mathbf{R}$ reduces effective information content

Practical approach: Use inflation, thinning, or explicit $\mathbf{R}$ correlations where feasible

Importance of Background Covariances

The $\mathbf{B}$ matrix is crucial to the performance of analysis systems!

Simple Example: Single observation at gridpoint i

$$\mathbf{H} = [0, \ldots, 0, 1, 0, \ldots, 0]$$

Gradient of 3D-Var cost function (at optimum):

$$\nabla_{\mathbf{x}} J = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}^o)$$

For linear $\mathbf{H}$, the analysis increment is:

$$\boxed{\mathbf{x}_a - \mathbf{x}_b = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}_b)}$$

For a single observation, the increment is proportional to a column of $\mathbf{B}$ scaled by the innovation and total error variance.

Roles of the Background Error Covariance Matrix $\mathbf{B}$ – Part 1

The $\mathbf{B}$ matrix is critical for:

1. **Spreading Information**
   - Observations at single gridpoint affect neighbouring points
   - Horizontal correlations determine spread length
   - Vertical correlations determine depth of influence
   - Without $\mathbf{B}$: observations only affect single gridpoint
2. **Providing Statistically Consistent Increments**
   - Ensures increments respect spatial/vertical correlations
   - Neighbouring gridpoints receive compatible adjustments
   - Different model levels receive physically consistent changes

Roles of the Background Error Covariance Matrix $\mathbf{B}$ – Part 2

3. **Ensuring Dynamical Consistency**

    - Temperature observations produce vorticity/divergence changes
    - Wind observations generate balanced temperature adjustments
    - Inter-variable correlations encoded in $\mathbf{B}$
    - Respects balance relationships in atmosphere (geostrophy, hydrostatic balance)

Key Insight: $\mathbf{B}$ transforms single observations into physically consistent increments across the model state

What Does a Column of B Look Like?

Single-Observation Experiment – the key diagnostic tool:

Assimilate one observation (e.g., T at 500 hPa) and examine the increment:
$\delta\mathbf{x} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(y^o - H\mathbf{x}_b)$

What you see:

- **Horizontal:** Smooth, localized spread around obs (correlation length scale)
- **Vertical:** Increment extends above/below obs level (vertical correlation)
- **Multivariate:** T obs produces $\mathbf{u}, \mathbf{v}$ increments via balance (and vice versa)

What goes wrong with a bad B:

- *Too-wide correlations:* smooths out real features
- *Too-narrow correlations:* noisy, data-void gaps
- *Missing balance:* T obs produces no wind increment $\rightarrow$ dynamically inconsistent

Takeaway: Single-obs tests directly reveal the spatial and multivariate structure of $\mathbf{B}$

Main Issues in Covariance Modelling

Two fundamental problems:

Problem 1: Unknown True State

Cannot directly diagnose background errors (true state unknown). Must use indirect methods:

- Innovation statistics from observations
- Forecast differences (NMC method)
- Ensemble of analyses (EDA method)

Problem 2: Enormous Dimensionality

$\mathbf{B}$ matrix is $\sim 10^8 \times 10^8$:

- Impossible to store or compute explicitly
- Must model through structured decomposition: $\mathbf{B} = \mathbf{L}\mathbf{L}^T$ and implement in operator form.

Diagnosing Background Error Statistics

Diagnosing Background Error Statistics

Three Classic Methods to Estimate $\mathbf{B}$:

1. **Innovation Statistics (Hollingsworth-Lönnberg, 1986)** Use obs − background: partitions background from observation error, but only in observation space and biased towards data-dense regions
2. **NMC Forecast Differences (Parrish & Derber, 1992)** 24h vs 48h forecasts that verify at the same time: inexpensive, global coverage, but static
3. **Ensemble of Analyses (Fisher, 2003)** Perturb all inputs, differences have background error statistics, flow-dependent, but expensive and feedback risk

Method 1: Innovation Statistics (Hollingsworth-Lönnberg)

Key Idea: Innovation contains both errors:

$$d_i = y_i - \mathcal{H}_i(\mathbf{x}_b)$$

Decomposition: Assuming obs errors uncorrelated:

$$\mathrm{Var}(d_i) = \mathrm{Var}(\epsilon_i^b) + \mathrm{Var}(\epsilon_i^o)$$

For separated obs:

$$\mathrm{Cov}(d_i, d_k) \approx \mathrm{Cov}(\epsilon_i^b, \epsilon_k^b)$$

Advantages:

- Direct real data
- Observation space

Limitations:

- Regional only
- Data-dense bias

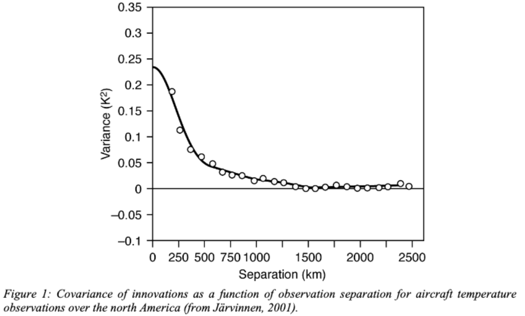


Figure 1: Covariance of innovations vs observation separation (aircraft T, N. America)

Result: Covariance vs separation reveals spatial structure of background error

Method 2: Forecast Differences (NMC Method)

Key Logic: Subtract two forecasts *valid at the same time* to eliminate the unknown truth

Step 1: Both forecasts verify at time t_k:

$$\mathbf{x}_{24h}(t_k) = \mathbf{x}_t(t_k) + \boldsymbol{\epsilon}_{24h}$$
$$\mathbf{x}_{48h}(t_k) = \mathbf{x}_t(t_k) + \boldsymbol{\epsilon}_{48h}$$

Step 2: Difference removes truth:

$$\mathbf{d}_k = \mathbf{x}_{24h}(t_k) - \mathbf{x}_{48h}(t_k) = \boldsymbol{\epsilon}_{24h} - \boldsymbol{\epsilon}_{48h}$$

Step 3: Covariance from time-averaging (after rescaling):

$$\mathbf{B}_{NMC} \propto \overline{\mathbf{d}_k \, \mathbf{d}_k^T}$$

Key assumption: 24h forecast errors approximate *cycling* background errors

Reality: Overestimates large-scale variance (48h errors are larger), but *correlation structure* is reasonable

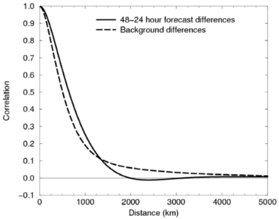


Figure 5: Background error correlation for 500hPa geopotential estimated using the NMC method (solid line) and the analysis ensemble method (dashed line).

NMC (solid) vs Ensemble (dashed), 500 hPa geopotential

Method 3: Ensemble of Analyses (Fisher, 2003)

Concept: Run the analysis system multiple times with independent perturbations

Cycle: Perturb observations, surface fluxes, model physics $\rightarrow$ analyze & forecast $\rightarrow$ repeat

Background differences after spin-up sample BG error statistics:

$$\mathbf{d}^{(i,j)} = \mathbf{x}_b^{(i)} - \mathbf{x}_b^{(j)}$$

Advantages:

- Direct error sample (not a proxy)
- Diagnoses the real assimilation system
- Naturally flow-dependent

Challenges:

- Needs good inputs: obs error stats, surface flux perturbations, model physics perturbations
- Feedback loops: perturbations affect analyses, which affect perturbations
- Expensive: requires multiple analyses and forecasts per cycle

EDA: Forecast vs. Analysis Error Correlations

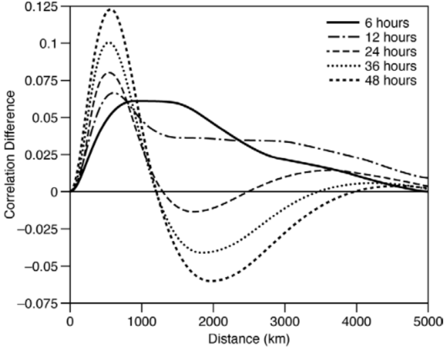


Figure 4: The difference between forecast correlation and analysis correlation calculated using the analysis ensemble method for 500hPa geopotential, and for a selection of forecast ranges.

Figure 4 (Fisher, 2003): Difference between forecast and analysis error correlation vs separation distance, for various forecast lengths (500 hPa geopotential).

Key observations:

- Forecast errors are more broadly correlated than analysis errors
- The difference grows with forecast range
- At long range, correlations approach the NMC-method picture
- Short-range EDA differences give the best estimate of true BG error correlations

Vertical Correlations: NMC vs Ensemble Methods

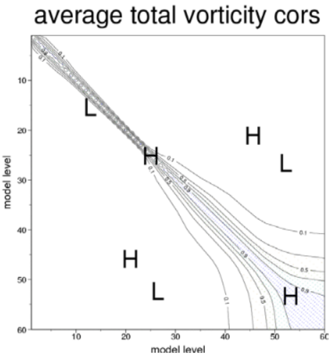


Figure 6: Mean vertical correlation for vorticity as a function of model level, calculated using the analysis-ensemble method.

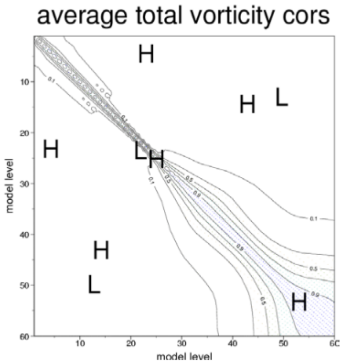


Figure 7: Mean vertical correlations for vorticity calculated using the NMC method.

Ensemble Method:

- More realistic vertical structure
- Captures flow-dependent variations

NMC Method:

- Broader vertical correlations
- Static (climatological)

Comparison of Methods for Estimating B

Method	Pros	Cons
Innovation Statistics	Direct data Observation space	Limited coverage Data-sparse bias
NMC Forecast Differences	Global statistics All levels Inexpensive	Analysis + BG mix Poor sparse regions Static (not flow-dep.)
Ensemble Analyses	Direct sampling Actual system Flow-dependent	Requires inputs Feedback risk Expensive

Key Insight: Each method has complementary strengths and weaknesses

Current Practice: Most centres use **multiple methods** for robustness

Covariance Formulations

J_b Formulation: The Control Variable

Key Challenge: Cannot explicitly store $\mathbf{B}$ matrix

Solution: Define $\mathbf{B}$ implicitly through control variable transformation

Change of variables:

$$\boldsymbol{\chi} = \mathbf{L}^{-1}(\mathbf{x} - \mathbf{x}_b) \quad \text{where} \quad \mathbf{B} = \mathbf{L}\mathbf{L}^T$$

Cost function in control space:

$$J(\boldsymbol{\chi}) = \frac{1}{2}\boldsymbol{\chi}^T\boldsymbol{\chi} + \frac{1}{2}\sum_{i=0}^{N} \|\mathbf{y}_i - \mathcal{H}_i(\mathbf{M}_i(\mathbf{x}_b + \mathbf{L}\boldsymbol{\chi}))\|^2_{\mathbf{R}^{-1}}$$

Advantages:

- $\mathbf{L}$ can be factorized into steps adding specific correlations
- Each step handles different physical aspect (balance, spectral, spatial)
- Computationally efficient (avoids explicit $\mathbf{B}^{-1}$ matrix)
- **Preconditioning:** Hessian $\approx \mathbf{I} + \ldots$ in control space (vs. $\mathbf{B}^{-1} + \ldots$ in model space) $\Rightarrow$ faster minimization convergence

The Balance Operator $\mathbf{K}$: Why and How

Problem: A wind observation should also adjust temperature (and vice versa). Without balance, each variable is corrected independently $\rightarrow$ dynamically inconsistent analysis.

Solution: Factor $\mathbf{L}$ as $\quad \mathbf{L} = \mathbf{K}\,\mathbf{B}_u^{1/2}$

Control variables (what the minimizer adjusts):

- ψ – streamfunction (vorticity)
- χ_u – unbalanced divergence
- T_u – unbalanced temperature
- p_s^u – unbalanced surface pressure
- q – specific humidity

Superscript u = "unbalanced´ ' = what remains after removing the part explained by ψ

What $\mathbf{K}$ adds (the balanced parts):

1. $\psi \xrightarrow{\text{NL balance}} \Phi_{bal}$ (geopotential from vorticity)
2. $\Phi_{bal} \xrightarrow{\text{hydrostatic}} T_{bal}$ (temperature from geopotential)
3. $\psi \xrightarrow{\text{regression}} p_{s,bal}$ (surface pressure)
4. $\psi \xrightarrow{\text{QG } \omega \text{ eqn}} \chi_{bal}$ (balanced divergence)

Total field = balanced + unbalanced:

$T = T_{bal}(\psi) + T_u$

Result: A single ψ increment propagates to Φ, T, p_s, χ via physical balances

The Balance Operator $\mathbf{K}$: Flow-Dependence

Key point: $\mathbf{K}$ is *not* a fixed matrix – it depends on the current atmospheric state

Linear (geostrophic) balance:

$$f\,\zeta \approx \nabla^2\Phi$$

- Good approximation at mid-latitudes, large scales
- *Independent* of background state

Nonlinear balance (used in IFS):

- Includes wind shear and curvature terms
- *Linearized about the background* $\mathbf{x}_b$
- $\Rightarrow$ $\mathbf{K}$ changes every cycle

Why does this matter?

- Near a jet: strong curvature $\rightarrow$ large ageostrophic component $\rightarrow$ balance operator accounts for this
- In the tropics: geostrophic balance is weak $\rightarrow$ $\mathbf{K}$ contributes less balanced mass
- After a cyclone deepens: balance relationships tighten $\rightarrow$ stronger cross-variable coupling

Practical effect: The multivariate structure of the increment adapts to the current weather – a source of flow-dependence even without ensembles

Fundamental Decomposition: Standard Deviations and Correlations

Key Design Principle: Separate $\mathbf{B}$ into magnitude and structure

For each control variable i: $\mathbf{B}_i = \mathbf{\Sigma}_i \, \mathbf{C}_i \, \mathbf{\Sigma}_i$ where $\mathbf{\Sigma}_i = \mathrm{diag}(\sigma_{b,1}, \sigma_{b,2}, \ldots)$ and $\mathbf{C}_i$ is the correlation matrix

Standard deviations $\mathbf{\Sigma}$:

- Control the *amplitude* of increments
- Vary geographically (larger in baroclinic zones)
- Can be updated daily from EDA

Correlations $\mathbf{C}$:

- Control the *shape* of increments
- Horizontal correlations $\rightarrow$ length scales
- Vertical correlations $\rightarrow$ depth of influence
- Can vary with wavenumber, location, flow

Why separate? Independent modelling, calibration, and flow-dependent updating of each

Spectral B Model (Derber & Bouttier, 1999)

Formulation: $\mathbf{L} = \mathbf{K}\mathbf{B}_u^{1/2}$ so $\mathbf{B} = \mathbf{K}\mathbf{B}_u\mathbf{K}^T$ where

$$\mathbf{B}_{u,i} = \boldsymbol{\Sigma}_i \, \mathbf{C}_i \, \boldsymbol{\Sigma}_i$$

$\boldsymbol{\sigma}_i$ = standard deviations; $\mathbf{C}_i$ = correlation; $\mathbf{K}$ = balance operator

Spectral Space: $\mathbf{C}_n$ block-diagonal (one block per total wavenumber n)

Each block: $\mathbf{C}_n^{(k)} \in \mathbb{R}^{N_{lev} \times N_{lev}}$ (full vertical correlation)

Key Feature: Full resolution of the variation of vertical correlations with horizontal scale (wavenumber)

- No variation of vertical correlations with location (physically unrealistic)
- Wavenumber-dependent vertical correlations (physically realistic)

Wavelet $\mathbb{B}$ Model (Fisher, 2003) – Part 1

Improvement over Spectral: Allow both spatial and spectral variation

Wavelet Basis: Band-limited, spatially localized functions $\mathbf{C} = \sum_j \mathbf{C}_j^{(wavelets)}$

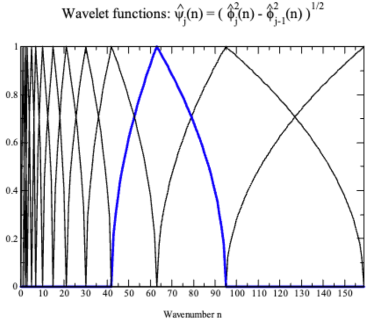


Figure 9: Spectral coefficients of the wavelet functions defined by equation (10).

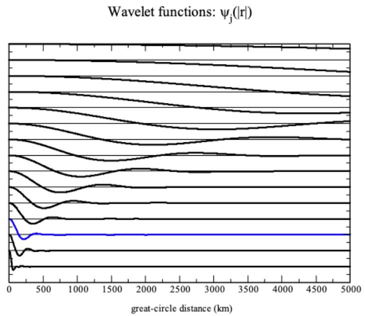


Figure 10: Functions of great-circle distance corresponding to the functions shown in Figure 9.

Spectral band-limited functions with finite support in wavenumber space

Spatial localization increases at higher wavenumbers (narrower peaks)

Wavelet B Model (Fisher, 2003) – Part 2

Key idea: Each wavelet band j has its own *location-dependent* vertical covariance $\mathbf{C}_j(\lambda, \phi)$

The total correlation is a sum over bands: $\mathbf{C} = \sum_j \mathbf{\Psi}_j \, \mathbf{C}_j \, \mathbf{\Psi}_j^T$, where $\mathbf{\Psi}_j$ are wavelet basis functions

Advantages:

- Variation in both wavenumber AND location

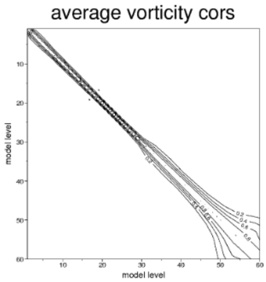


Figure 11: The effective wavenumber-averaged vertical correlation matrix for vorticity over north America implied by the wavelet J_b covariance model.

Shallow boundary layer, deep troposphere/strat. coupling

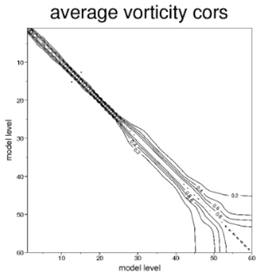


Figure 12: As Figure 11, but for a point in the equatorial Pacific ocean.

Deeper boundary layer, different tropopause signature

Horizontal Inhomogeneity in Wavelet Model

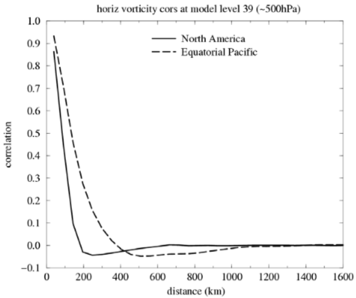


Figure 13: Effective horizontal structure functions for vorticity for a point over north America and a point in the equatorial Pacific.

Figure 13: Horizontal structure functions showing location-dependent length scales

Key Finding: Correlation length scales vary with location

- **North America:** Shorter length scales (continental effect)
- **Equatorial Pacific:** Longer length scales (ocean effect)

Physical Mechanism: Wavelets allow $\mathbf{C}_j(\lambda, \phi)$ to vary with location

Flow-Dependent Extension:

- Daily update from EDA
- Captures weather regime variations
- More realistic error statistics

Diffusion Operators

Motivation: Spectral/wavelet problematic for irregular domains (e.g., ocean models)

Idea: Use diffusion equation to generate correlations

One-dimensional diffusion:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

Solution at time T:

$$u(x,T) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi T}} \exp\left(-\frac{(x-x')^2}{4T}\right) u(x',0)dx'$$

$=$ *convolution with Gaussian of width* $\sqrt{T}$

Advantages:

- Handles irregular boundaries
- Tensor coefficients allow anisotropy
- Efficient numerical implementation

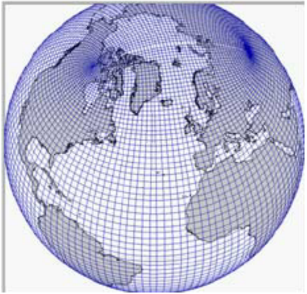


Figure: Irregular grid used in ocean models (ORCA grid)

Digital Filters

Alternative to Diffusion: Digital filters (Purser et al., 2003)

Recursive filter: forward pass $v_i = \alpha\, v_{i-1} + (1-\alpha)\, u_i$, then backward pass; the cascade approximates a Gaussian convolution

Explicit (non-recursive) filter: repeated symmetric 3-point smoothing: $\tilde{u}(x) = \beta\, u(x{+}1) + (1{-}2\beta)\, u(x) + \beta\, u(x{-}1)$; multiple passes also converge to a Gaussian

2D Implementation: Factorization of 2D Gaussian

$$\exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) = \exp\left(-\frac{x^2}{2\sigma_x^2}\right)\exp\left(-\frac{y^2}{2\sigma_y^2}\right)$$

Apply 1D filter in x-direction, then y-direction

Advantages:

- Computationally very efficient
- Easy to implement with sparse operations
- Allows spatially varying filter coefficients (inhomogeneity)

Disadvantages:

- Determining filter coefficients from data is difficult
- Requires careful treatment of boundaries and poles

Comparison of B Formulations

Property	Spectral	Wavelet	Diffusion/Filter
Irregular domains	**No**	Yes	Yes
Inhomogeneous	No	Yes	Yes
Anisotropic	No	Partially	Yes
Flow-dependent	Difficult	Easy (EDA)	Possible
Computation	Very efficient	Efficient	Balanced
Complexity	Low	Medium	High

Current Practice:

- **Atmospheric model**: Wavelet formulation with flow-dependence from EDA
- **Ocean models**: Diffusion operators (irregular domains) with flow-dependence from EDA
- **Research**: Flow-dependent ensembles + wavelets

Advanced Topics

Variances of the Day (from EDA)

Operational concept: Update $\boldsymbol{\Sigma}_b$ daily using the Ensemble of Data Assimilations

How it works:

1. Run N_e perturbed analyses/forecasts (EDA)
2. Compute ensemble spread at each grid point:

$$\sigma_b(\mathbf{r}, t) = \sqrt{\frac{1}{N_e - 1} \sum_{m=1}^{N_e} (x_b^{(m)} - \bar{x}_b)^2}$$

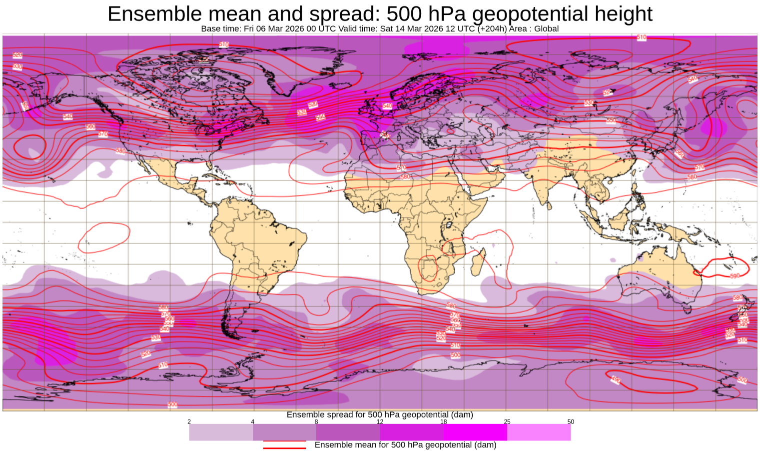
3. Re-scale the static $\boldsymbol{\Sigma}$ in J_b

Physical effect:

- Larger σ_b where forecast is uncertain (e.g. near fronts, cyclones)
- Smaller σ_b in quiescent regions
- Analysis trusts observations more in high-spread areas

Impact: Situation-dependent weighting between background and observations – one of the most impactful flow-dependent $\mathbf{B}$ features in operations

EDA Spread and Ensemble Mean: 500 hPa Geopotential



Ensemble spread (shading) highlights regions of high forecast uncertainty; ensemble mean (contours) shows the large-scale flow. Spread is largest near baroclinic zones, fronts, and developing cyclones.

$\mathbf{B}$ in 4D-Var: Implicit Flow-Dependence

Key insight: In 4D-Var, even a *static* $\mathbf{B}$ produces flow-dependent increments

3D-Var: increment $= \mathbf{B}\mathbf{H}^T(\ldots)^{-1}\mathbf{d}$

- Increment structure is fixed by $\mathbf{B}$
- Isotropic if $\mathbf{B}$ is isotropic
- No time dimension

4D-Var: uses $\mathbf{M}\mathbf{B}\mathbf{M}^T$ implicitly

- TL model $\mathbf{M}$ propagates $\mathbf{B}$ along trajectory
- Increments stretched along flow (e.g., along jet)
- Effectively anisotropic and flow-dependent

Practical consequence:

- 4D-Var extracts more information from the same observations than 3D-Var
- Longer assimilation windows $\Rightarrow$ stronger implicit flow-dependence
- Combining with *explicit* flow-dependence (EDA, hybrid) gives further gains

Hybrid B and Localization

Hybrid formulation (static + ensemble):

$$\mathbf{B}_{hyb} = (1-\alpha)\,\mathbf{B}_{static} + \alpha\,\mathbf{B}_{ens}$$

Ensemble covariance:

$$\mathbf{B}_{ens} = \frac{1}{N_e - 1}\sum_{m=1}^{N_e}(\mathbf{x}_b^{(m)} - \overline{\mathbf{x}}_b)(\mathbf{x}_b^{(m)} - \overline{\mathbf{x}}_b)^T$$

Perturbations from EDA: perturbed observations, stochastic physics, surface flux perturbations, and IC/BC perturbations

Why hybrid?

- Retains flow dependence from ensembles
- Stabilizes sampling noise with static climatology
- Widely used in hybrid EnVar and 4D-EnVar

Localization to reduce sampling error:

$$\mathbf{B}_{ens}^{loc} = \mathbf{L} \circ \mathbf{B}_{ens}$$

where $\mathbf{L}$ is a compactly supported correlation matrix

Localization and Scale-Dependent Localization

Localization (EnKF/EnVar):

- Suppresses spurious long-range correlations from finite ensembles
- Implemented via Schur product: $\mathbf{B}_{ens}^{loc} = \mathbf{L} \circ \mathbf{B}_{ens}$
- $\mathbf{L}$ often Gaspari–Cohn with compact support

Scale-dependent localization:

- Different correlation length scales for different spatial scales or variables
- Example: shorter localization for small scales, longer for synoptic scales
- Implemented with multi-scale transforms (spectral or wavelet) or adaptive $\mathbf{L}(k)$

Benefit: Preserves physically meaningful large-scale correlations while reducing noise at small scales

Model Error and Weak-Constraint 4D-Var

Strong-constraint 4D-Var: model is perfect ($\mathbf{x}_{k+1} = \mathcal{M}_k(\mathbf{x}_k)$)

Weak-constraint 4D-Var: allow model error $\boldsymbol{\eta}_k$

$$\mathbf{x}_{k+1} = \mathcal{M}_k(\mathbf{x}_k) + \boldsymbol{\eta}_k, \quad \boldsymbol{\eta}_k \sim \mathcal{N}(0, \mathbf{Q})$$

Impact on $\mathbf{B}$:

- Model error inflates background uncertainty
- $\mathbf{Q}$ interacts with $\mathbf{B}$ and affects flow dependence
- Ignoring model error can overfit observations and mis-shape increments

Desroziers Diagnostics (Desroziers et al., 2005)

Idea: If $\mathbf{B}$ and $\mathbf{R}$ are correctly specified, innovation and residual statistics must be mutually consistent

Define: $\mathbf{d}^b = \mathbf{y}^o - \mathcal{H}(\mathbf{x}_b)$ (innovation), $\mathbf{d}^a = \mathbf{y}^o - \mathcal{H}(\mathbf{x}_a)$ (analysis residual)

Three consistency relations:

$$\mathbb{E}[\mathbf{d}^b(\mathbf{d}^b)^T] = \mathbf{HBH}^T + \mathbf{R}$$

$$\mathbb{E}[\mathbf{d}^a(\mathbf{d}^b)^T] = \mathbf{R} \qquad \rightarrow \text{diagnose } \mathbf{R}$$

$$\mathbb{E}[(\mathbf{d}^b - \mathbf{d}^a)(\mathbf{d}^b)^T] = \mathbf{HBH}^T \qquad \rightarrow \text{diagnose } \mathbf{B}$$

Practical use:

- Compare diagnosed σ_b, σ_o against prescribed values
- Iteratively tune $\mathbf{B}$ and $\mathbf{R}$ towards consistency
- Monitor per obs type, channel, region

Caveat: Relations are exact only when the analysis is *optimal*; in practice the diagnostic is approximate and should not be iterated blindly

Further Validation and Diagnostics of B

Complementary diagnostics:

- Single-observation experiments: check spatial structure of increments
- Innovation whitening: reduced correlation in $\mathbf{R}^{-1/2}d$
- Degrees of freedom for signal (DFS): effective information extracted from observations

Verification indicators:

- Fit-to-observations vs forecast skill (bias–variance tradeoff)
- Consistency across channels, variables, and regimes
- Sensitivity to thinning, localization, and inflation

Summary

Summary: Background Error Covariance Modelling

1. $\mathbf{B}$ **matrix is crucial** for spreading information and ensuring consistency
2. **Three estimation methods:** Innovation statistics, Forecast differences (NMC), Ensemble analyses
3. **Control variable decomposition:** $\mathbf{B} = \mathbf{L}\mathbf{L}^T$ with $\mathbf{L} = \mathbf{K}\mathbf{B}_u^{1/2}$
4. **Three formulation approaches:**
   - Spectral: efficient for global regular grids
   - Wavelets: handles inhomogeneity and flow-dependence
   - Diffusion/Filters: flexible for irregular domains
5. **No single perfect solution** – choose based on model geometry and constraints
6. **Current trend:** Flow-dependent $\mathbf{B}$ from EDA (Ensemble Data Assimilation)

Thank You!

Questions & Discussion

Marcin Chrust
marcin.chrust@ecmwf.int

European Centre for Medium-Range Weather Forecasts

References and Further Reading

- Bannister, R.N., 2008: A review of forecast-error covariance statistics in atmospheric variational data assimilation. *Q.J.R. Meteorol. Soc.*, **134**, 1951–1970.
- Bouttier, F. & Courtier, P., 2002: Data Assimilation Concepts and Methods. *ECMWF Meteorological Training Course Lecture Notes*.
- Derber, J. & Bouttier, F., 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus A*, **51**, 195–221.
- Desroziers, G., Berre, L., Chapnik, B. & Poli, P., 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Q.J.R. Meteorol. Soc.*, **131**, 3385–3396.
- Fisher, M., 2003: Background error covariance modelling. *Proc. ECMWF Seminar on Recent Developments in Data Assimilation*, 45–63.
- Hollingsworth, A. & Lönnberg, P., 1986: The statistical structure of short-range forecast errors. *Tellus A*, **38**, 111–136.
- Parrish, D.F. & Derber, J.C., 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Isaksen, L., et al., 2010: Ensemble of data assimilations at ECMWF. *ECMWF Tech. Memo.*, **636**.