Algorithms (2):

Static B, 3DVar, and Hybrid-EnVar

Presented by BJ Jung

NSF NCAR/MMM

MPAS-A and MPAS-JEDI Tutorial
University of St Andrews, UK, June 23-27, 2025

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

Overview

1. 3DVar and static B
2. B designed and how it operates
3. B training and tuning
4. Yaml configuration for 3DVar
5. Yaml configuration for Hybrid-3DEnVar

3DVar and Static Background Error Covariance

$$\mathrm{J(x)} = \underbrace{\frac{1}{2}(\mathrm{x}-\mathrm{x_b})^{\mathrm{T}}\mathbf{B}^{-\mathbf{1}}(\mathrm{x}-\mathrm{x_b})}_{\mathbf{J_b}} + \underbrace{\frac{1}{2}(\mathrm{h(x)}-\mathrm{y})^{\mathrm{T}}\mathbf{R}^{-\mathbf{1}}(\mathrm{h(x)}-\mathrm{y})}_{\mathbf{J_o}}$$

- 3DVar uses $\mathbf{B} = \mathbf{B_s}$.
- $\mathbf{B_s}$ has a **climatological** characteristics.
- 3DVar is computationally efficient.
- 3DVar is simple in terms of workflow.
- The hybrid B can be benefit from $\mathbf{B_s}$.

B designed and how it operates (1/6)

$$\mathbf{B} = < \mathbf{x_b} - \mathbf{x_t}, \mathbf{x_b} - \mathbf{x_t} >$$

- Role of $\mathbf{B}$
  - weights the importance of $\mathbf{x_b}$ for a given y and $\mathbf{R}$.
  - spreads the observed information in the vertical and horizontal direction.
  - spreads the observed information to other variables and imposes the balance properties.
- It is difficult to know $\mathbf{B}$ exactly.
  - We don't know the "true" state, $\mathbf{x_t}$.
  - The dimension of B is too large.
- Thus, $\mathbf{B}$ is modeled in a practical way.

See Bannister (2008a,b) for more detailed review on B.

B designed and how it operates (2/6)

- MPAS-JEDI's $\mathbf{B}_s$ is designed following that of the GSI* (Wu et al., 2002), except the univariate spatial correlation.

$$\mathbf{B}=\mathbf{K}_1\mathbf{K}_2\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}^T\mathbf{K}_2^T\mathbf{K}_1^T$$

- It is constructed as a series of linear variable changes ($\mathbf{K}_1$, $\mathbf{K}_2$ and their adjoint; ^{T}) around the univariate covariance ($\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}^T$).

- Note that…
  - **B** operates to the MPAS-JEDI's analysis variable ($\delta\mathbf{x}$), a set of $\{\delta u,\ \delta v,\ \delta T,\ \delta Q,\ \delta p_s\}$ @ cell center.
  - $\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}^T$ is represented with $\{\delta\psi,\ \delta\chi_u,\ \delta T_u,\ \delta Q,\ \delta p_{s,u}\}$.

**Gridpoint Statistical Interpolation*

$B = K_1 K_2 \Sigma C \Sigma^T K_2^T K_1^T$ (3/6)

- K_1 is a linear variable change
  from stream function ($\delta\psi$) and velocity potential ($\delta\chi$)
  to zonal (δu) and meridional (δv) winds

$$\begin{bmatrix} \delta u \\ \delta v \\ \delta T \\ \delta Q \\ \delta p_s \end{bmatrix} = \begin{bmatrix} -\partial_y & -\partial_x & 0 & 0 & 0 \\ \partial_x & -\partial_y & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{bmatrix} \begin{bmatrix} \delta\psi \\ \delta\chi \\ \delta T \\ \delta Q \\ \delta p_s \end{bmatrix}$$

- K_1 directly operates on the native MPAS mesh (code).
- K_1^T is a corresponding adjoint operator (code).

$B = K_1 K_2 \Sigma C \Sigma^T K_2^T K_1^T$ (4/6)

- **K_2** applies the vertical cross-variable correlation via linear regression
  : throuth BUMP Vertical BALance (VBAL) operator

  - $\delta\chi = \delta\chi_b + \delta\chi_u = \boldsymbol{L}\delta\psi + \delta\chi_u$
  - $\delta T = \delta T_b + \delta T_u = \boldsymbol{M}\delta\psi + \delta T_u$
  - $\delta p_s = \delta p_{s,b} + \delta p_{s,u} = \boldsymbol{N}\delta\psi + \delta\chi_u$

$$\begin{bmatrix} \delta\psi \\ \delta\chi \\ \delta T \\ \delta Q \\ \delta p_s \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{L} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{M} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{N} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \delta\psi \\ \delta\chi_u \\ \delta T_u \\ \delta Q \\ \delta p_{s,u} \end{bmatrix}$$

  – $\delta\psi$ is a predictor for the *balanced* part of $\delta\chi$, δT , and δp_s.
  – Full matrix for **M** & **N**, diagonal matrix for **L**

Following Derber and Bouttier (1999)

- **K_2^T** is a corresponding adjoint operator.

$B = K_1 K_2 \Sigma C \Sigma^T K_2^T K_1^T$ (5/6)

- **$\Sigma C \Sigma^T$** represents the univariate covariance for $\{\delta\psi, \delta\chi_u, \delta T_u, \delta Q, \delta p_{s,u}\}$. These variables do not have a cross variable covariance anymore.

- **Σ** (and **Σ^T**) is a diagonal matrix with error standard deviation
: through either BUMP VARiance (VAR) operator or SABER StdDev operator

- **C** is a block diagonal matrix. Each blocks represents the spatial correlation for each variables for $\{\delta\psi, \delta\chi_u, \delta T_u, \delta Q, \delta p_{s,u}\}$.
  – through BUMP Normalized Interpolated Convolution on an Adaptive Subgrid (NICAS; Ménétrier, 2020) operator
  – **BUMP NICAS directly operates on the MPAS's unstructured mesh.**

$B=K_1K_2\Sigma C\Sigma^T K_2^T K_1^T$ (6/6)

- Even with a single variable, the dimension for spatial correlation is still large.
- BUMP NICAS applies the spatial correlation with *a subset of full grid* ($\mathbf{C^s}$), rather than with a full grid ($\mathbf{C}$). The correlation on the subgrid is interpolated ($\mathbf{S}$) to the full grid with a normalization factorn ($\mathbf{N}$).

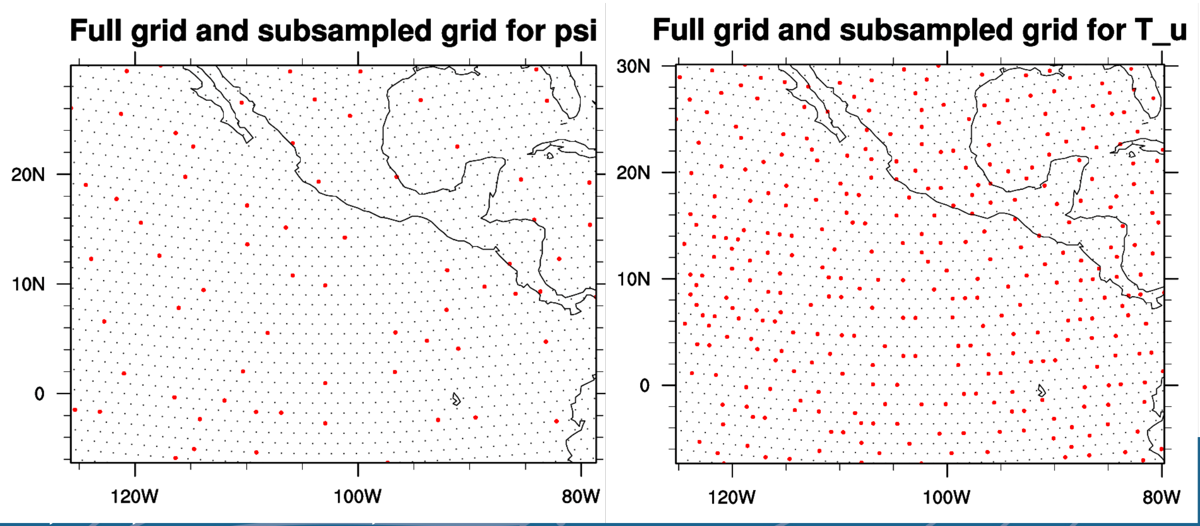
$$\mathbf{C} \cong \tilde{\mathbf{C}} = \mathbf{N}\mathbf{S}\mathbf{C^s}\mathbf{S^T}\mathbf{N^T}$$

$\mathbf{C} \rightarrow \mathbb{R}^{n\times n}$; $\mathbf{C^s} \rightarrow \mathbb{R}^{n_s\times n_s}$ with $n_s \ll n$

$\mathbf{N}$: diagonal matrix for normalization
(to ensure the diagonal component of C equals "1")

$\mathbf{S} = \mathbf{S^v}\mathbf{S^h}$: Interpolation from sub-sampled to full mesh

- The correlation function follows the shape of Gaspari-Cohn (1999)'s 5th order functions, for the given lengths.



$B = K_1 K_2 \Sigma C \Sigma^T K_2^T K_1^T$

- In previous slides, we learned how B is designed and each operator works.
- Some of the operators requires the *known parameters*.

- $\mathbf{K_2}$ needs *the regression coefficients* (i.e., **L**, **M**, and **N**).
- *The error standard deviation* constructs **Σ**.
- *The horizontal and vertical correlation lengths* are required for **C**.

- How can we estimate (or train or calibrate) these parameters ?
  - From *some* samples : proxy for background error
  - Need to apply the inverse of each operator

(Like GEN_BE in WRFDA or GSI)

B training and tuning (1/6)

- We can estimate the required static B parameters from samples of forecast differences with two different lead times, which is called "*NMC method*".
- In the hands-on practice, we will take the 48 hour- and 24 hour- forecast differences as a proxy of background error.

More in the hands-on practice

- Alternatively, we can use the ensemble forecast as a proxy of background error, which is called "*ensemble method*".

- After a proxy of background error (some form of perturbations) is determined, we can process the following procedures to estimate the static B parameters.

B training and tuning (2/6)

- K_1 does not have parameters. It is basically some spatial derivative operator. However, we still need to apply its inverse to the training samples to estimate the other static B parameters.
- K_1^{-1} (converting u and v to ψ and χ) is **not trivial** on the unstructured grid.
- So we…
    1) Interpolate $\{u, v\}$ on the unstructured grid to regular lat/lon grid,
    2) Calculate $\{\psi, \chi\}$ on the regular lat/lon grid through "uv2sfvpf" function of **NCL**,
    3) Interpolate $\{\psi, \chi\}$ on the regular lat/lon to the MPAS unstructured grid.
- For convenience, "temperature" and "specific humidity" are also pre-calculated.
- So the **perturbation fields** $\{\delta\psi, \delta\chi, \delta T, \delta Q, \delta p_s\}$ can be obtained by subtracting the 24 hour forecast from 48 hour forecast at the same valid time.

More in the hands-on practice

B training and tuning (3/6)

- The parameters in K_2 can be estimated with BUMP **V**ertical **BAL**ance (VBAL) driver.
- The regression coefficients ($\boldsymbol{L}$, $\boldsymbol{M}$, and $\boldsymbol{N}$) can be obtained by multiplying **inverse of auto-covariance matrix** $< \boldsymbol{\delta\psi}, \boldsymbol{\delta\psi} >$ to the cross-covariance matrices $< \delta\psi, \delta\chi >$, $< \delta\psi, \delta T >$, and $< \delta\psi, \delta p_s >$.
- Usually, the auto- and cross-covariance matrices are averaged over some latitude bands.
- Due to the large condition number of auto-covariance matrix, a pseudo-inverse is used with an Eigenvalue Decomposition.
- This step will also apply K_2^{-1} to the training samples and write out the ***unbalanced*** **samples {$\boldsymbol{\delta\psi}$, $\boldsymbol{\delta\chi_u}$, $\boldsymbol{\delta T_u}$, $\boldsymbol{\delta Q}$, $\boldsymbol{\delta p_{s,u}}$},** which are used for the next step.

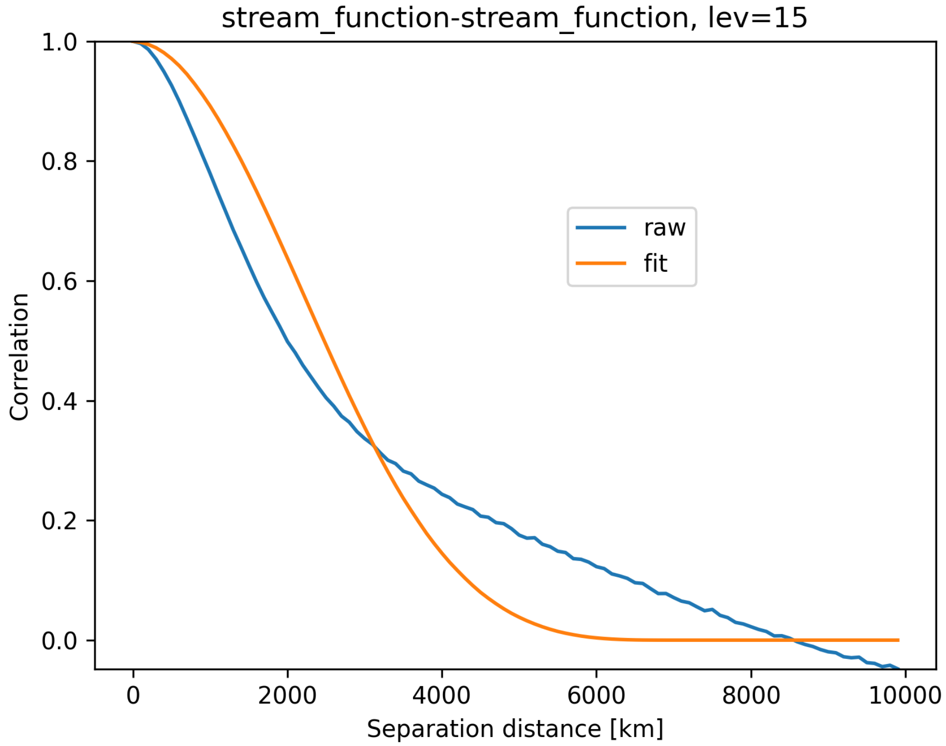
B training and tuning (4/6)

- Σ is estimated from unbalanced samples with BUMP **VAR**iance (VAR) driver.

- The horizontal-, vertical- correlation lengths for the unbalanced variables are estimated from samples with BUMP **H**ybrid **Diag**nostics (HDIAG) driver.
  - It defines the low-resolution grid for computation and the diagnostic point is randomly selected for different horizontal separation distances.
  - For different horizontal separation distances, the raw sample correlations are calculated.
  - The vertical correlation is also calculated at each computation grid, between each level and the neighboring levels.
  - These raw sample correlations are averaged over all computation grid or locally.
  - Finally, HDIAG fits the Gaspari-Cohn's function for each averaged correlation curves.

B training and tuning (5/6)

- Additional tunings are applied.
- Reducing the error STD for all variables by a factor of 1/3
  - To scale the 24 hour difference in the NMC-type perturbations as 6 hour (typical DA cycle)
- Reducing the diagnosed horizontal lengths for $\delta\psi$ and $\delta\chi_u$ by a factor of ½
  - To match the wind variance near the small separations better
  - The implied wind variance is proportional to the second derivatives of $\delta\psi$ and $\delta\chi_u$.



Sample correlation structure
Fitted structure to Gaspari-Cohn's 5th order function

B training and tuning (6/6)

- For a given horizontal and vertical correlation lengths, NICAS driver will pre-generate the correlation values on the subgrid, the interpolation weights and indices between subgrid and full grid, and the normalization factors.

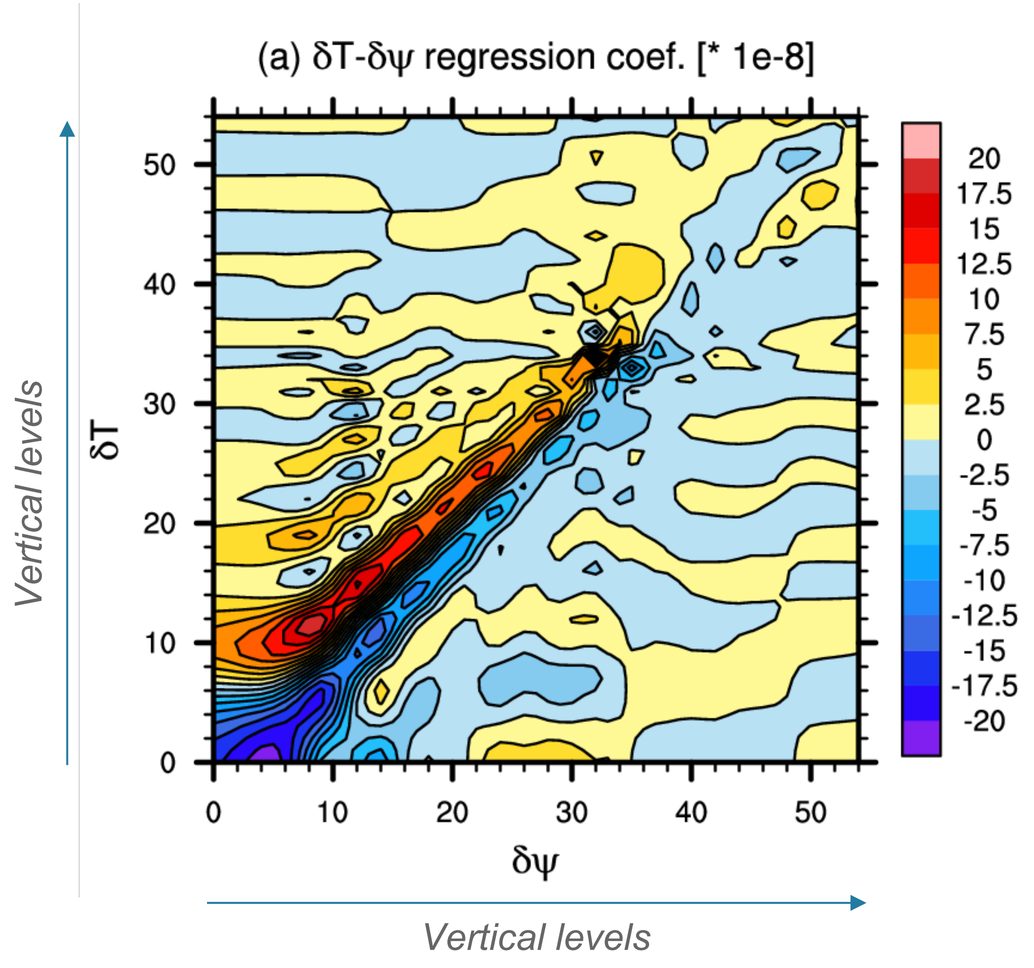
- The subgrid will be determined by a given horizontal and vertical correlation lengths and some yaml keys.

```
nicas:
  resolution: 8
  max horizontal grid size: 15000  #default
```

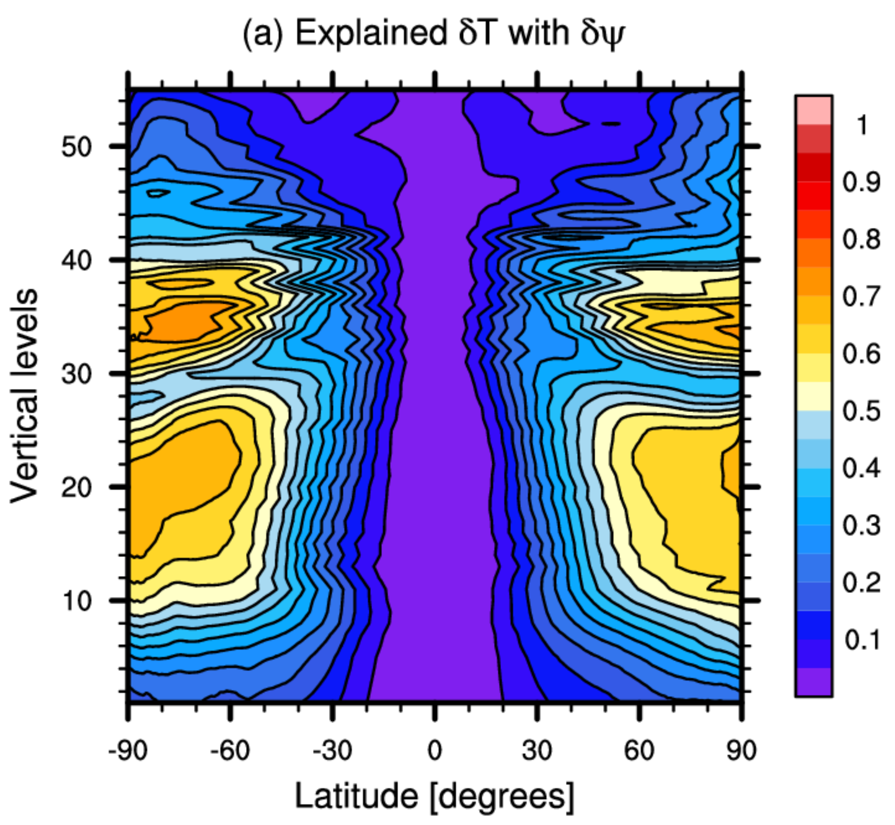
number of grid point to represent a single correlation function (resolution)

number of maximum horizontal grid to define the "subgrid" (max horizontal grid size)

Regression coefficients $\boldsymbol{M}$ (part of $\mathbf{K_2}$) at 34.8° N latitude

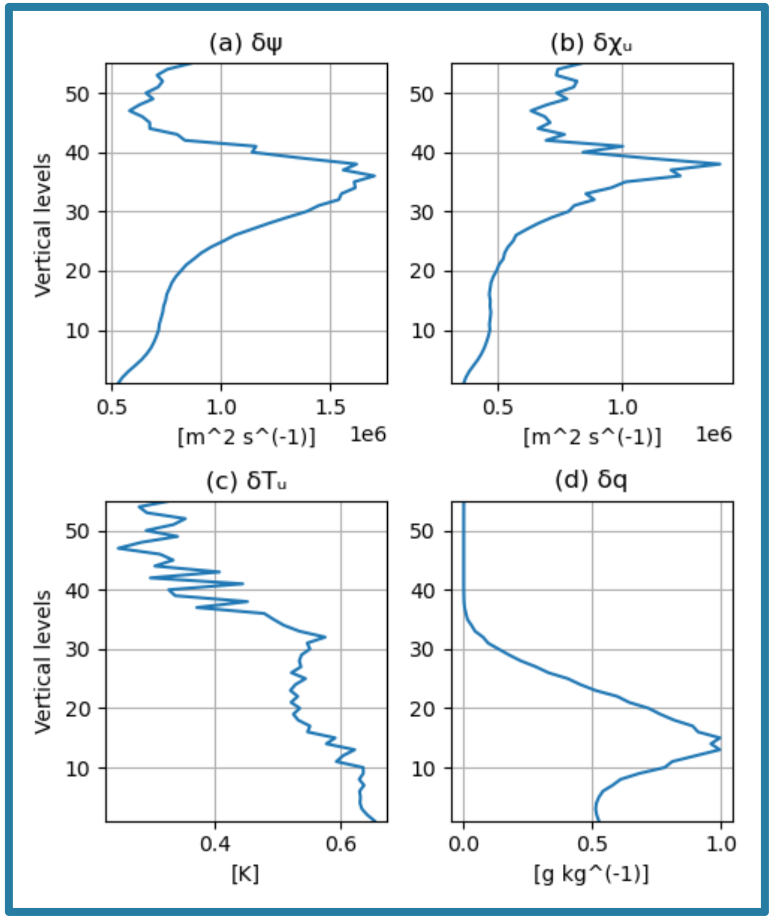


Ratio of balanced variance to total variance

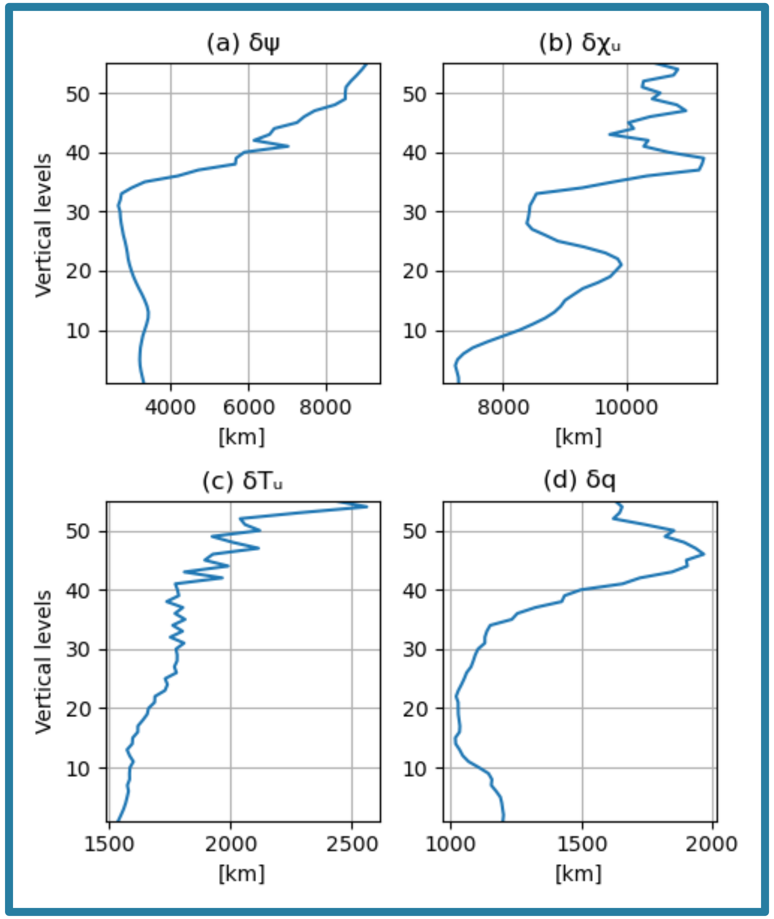


$$\frac{\langle \delta T_b, \delta T \rangle}{\langle \delta T, \delta T \rangle}$$

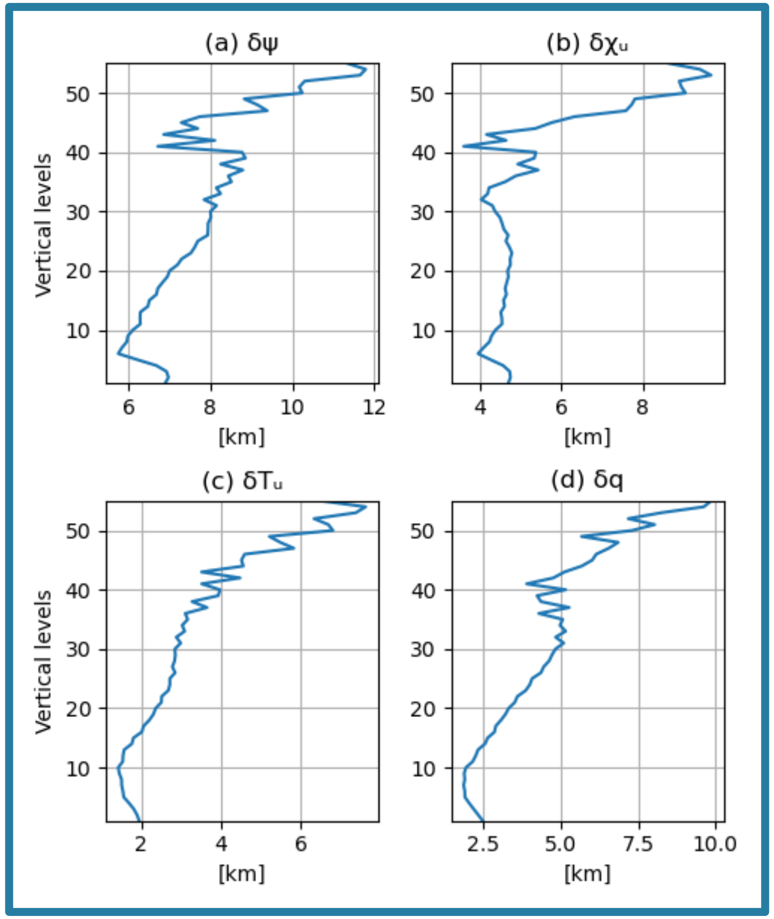
Error standard dev. Σ



Horizontal corr. lengths



Vertical corr. lengths

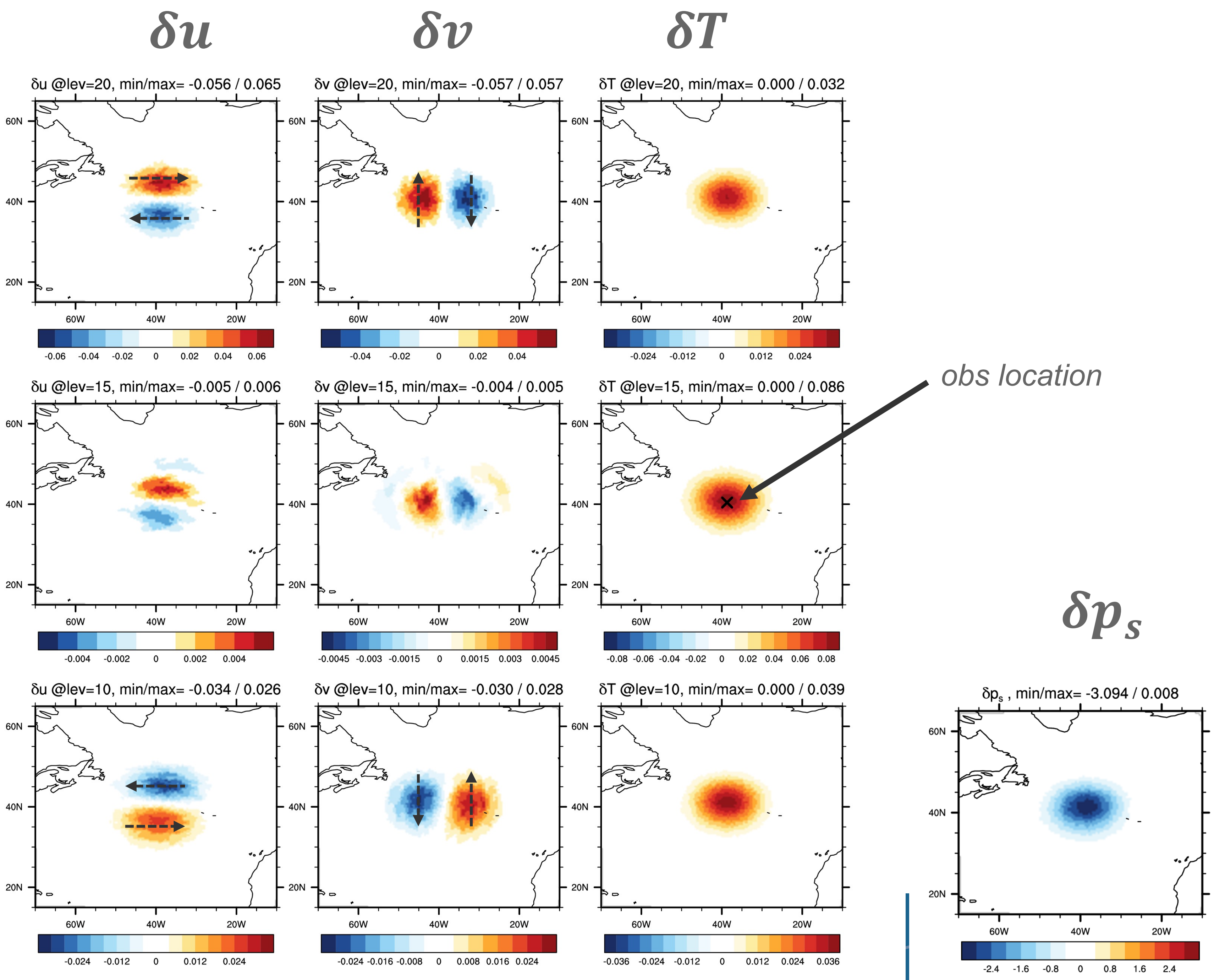


Single T obs test

Model level = 20

Model level = 15

Model level = 10



Previous slides describe ‘multivariate’ B.
MPAS-JEDI can easily do ‘univariate’ B, in that case:

$$\mathbf{B} = \mathbf{\Sigma} \mathbf{C} \mathbf{\Sigma}^{\mathrm{T}}$$

i.e., no cross-variable correlation between analysis variables of $\{\delta u, \delta v, \delta T, \delta Q, \delta p_s\}$

This should work both for **global** and **regional** MPAS meshes.

With a focus on "`cost function.background error`" section

Yaml configuration for 3DVar (1/6)

```
cost function:
  cost type: 3D-Var
  window begin: 2018-04-14T21:00:00Z
  window length: PT6H

  analysis variables: &incvars
[spechum,surface_pressure,temperature,uReconstructMeridional,uReconstructZonal]

  background:
    state variables:
[spechum,surface_pressure,temperature,uReconstructMeridional,uReconstructZonal,theta,rho,u,qv,pressure,landmask,xice,snowc,skintemp,ivgtyp,isltyp,snowh,vegfra,u10,v10,lai,smois,tslb,pressure_p]
    filename: ./bg.2018-04-15_00.00.00.nc
    date: &analysisDate 2018-04-15T00:00:00Z
```

{ 3D-Var, 4D-Ens-Var, 3D-FGAT } (cost type)

The variables that are used for the minimization algorithm. (analysis variables)

Background state usually includes more that just "analysis variables" (state variables)

Background state is read from this file. (filename)

Analysis time is at the center of assimilation window. (date)

Yaml configuration for 3DVar (2/6)

$$B = K_1 K_2 \Sigma C \Sigma^T K_2^T K_1^T$$

```
cost function:
  ...
  background error:
    covariance model: SABER
    saber central block:
C       saber block name: BUMP_NICAS
        ... more config ...
    saber outer blocks:
Σ     - saber block name: StdDev
        ... more config ...
K2    - saber block name: BUMP_VerticalBalance
        ... more config ...
K1  linear variable change:
      linear variable change name: Control2Analysis
      ... more config ...
```

In order of forward (TL) operations

Yaml configuration for 3DVar (3/6)

$$B = K_1 K_2 \Sigma C \Sigma^T K_2^T K_1^T$$

```
background error:
  covariance model: SABER
  saber central block:
    saber block name: BUMP_NICAS
    active variables: &ctlvars
[stream_function,velocity_potential,temperature,spechum,surface_pressure]
    read:
      io:
        data directory: ./BUMP_files/bump_nicas
        files prefix: bumpcov_nicas
      drivers:
        multivariate strategy: univariate
        read local nicas: true
```

"univariate" strategy is used for covariance C.f. "duplicated" or "crossed" strategy is used when NICAS is used for localization

Read the pre-generated NICAS files. Here, "local" nicas means one NICAS file per one processor. C.f. "read global nicas: true" reads single NICAS file, which contains the information over all domain.

Note that there is NO distinction with "total" and "unbalanced" variable names in the YAMLs or files.

Yaml configuration for 3DVar (4/6)

$$B = K_1 K_2 \Sigma C \Sigma^T K_2^T K_1^T$$

```
background error:
  covariance model: SABER
  saber central block:
    saber block name: BUMP_NICAS
    ... more config ...
  saber outer blocks:
  - saber block name: StdDev
    read:
      model file:
        filename: ./BUMP_files/stddev/mpas.stddev_0p33.2018-04-15_00.00.00.nc
        date: *analysisDate
        stream name: control
```

Read the standard deviation values from MPAS file format.

Yaml configuration for 3DVar (5/6)

$$B = K_1 \mathbf{K_2} \Sigma C \Sigma^T \mathbf{K_2^T} K_1^T$$

```
- saber block name: BUMP_VerticalBalance
  read:
    io:
      data directory: ./BUMP_files/bump_vertical_balance
      files prefix: bumpcov_vbal
    drivers:
      read local sampling: true
      read vertical balance: true
    vertical balance:
      vbal:
      - balanced variable: velocity_potential
        unbalanced variable: stream_function
        diagonal regression: true
      - balanced variable: temperature
        unbalanced variable: stream_function
      - balanced variable: surface_pressure
        unbalanced variable: stream_function
```

Read the pre-generated "local" VBAL & sampling files.

These describe the designed balance (regression) relationship.

Yaml configuration for 3DVar (6/6)

$$B = K_1 K_2 \Sigma C \Sigma^T K_2^T K_1^T$$

```
background error:
  covariance model: SABER
  saber central block:
    saber block name: BUMP_NICAS
    ... more config ...
  saber outer blocks:
  - saber block name: StdDev
    ... more config ...
  - saber block name: BUMP_VerticalBalance
    ... more config ...
  linear variable change:
    linear variable change name: Control2Analysis
    input variables: *ctlvars
    output variables: *incvars
```

Yaml configuration for Hybrid-3DEnVar (1/2)

- 3DVar uses

```
cost function:
  background error:
    covariance model: SABER
    ... more config ...
```

- 3DEnVar uses

```
cost function:
  background error:
    covariance model: ensemble
    ... more config ...
```

- We can configure the hybrid covariance as a linear combination of other Bs !

$$\mathbf{B}_{\mathbf{hybrid}} = \alpha \mathbf{B}_{\mathbf{static}} + \beta \mathbf{B}_{\mathbf{ensemble}}$$

(Hamill and Snyder, 2000)

Yaml configuration for Hybrid-3DEnVar (2/2)

- We can configure the hybrid covariance as a linear combination of other Bs !

```
background error:
   covariance model: hybrid
   components:
   - weight:
       value: 0.5
     covariance:
       covariance model: SABER
       ... more config ...
   - weight:
       value: 0.5
     covariance:
       covariance model: ensemble
       ... more config ...
```

$$\mathbf{B}_{\mathbf{hybrid}} = \alpha \mathbf{B}_{\mathbf{static}} + \beta \mathbf{B}_{\mathbf{ensemble}}$$

The weights can be a simple scalar value or they can be read from a file, which varies spatially.

References

Bannister, R.N. (2008a), A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. Q.J.R. Meteorol. Soc., 134: 1951-1970. https://doi.org/10.1002/qj.339

Bannister, R.N. (2008b), 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. https://doi.org/10.1002/qj.340

Derber, J. and Bouttier, F.: A reformulation of the background error covariance in the ECMWF global data assimilation system, Tellus A: Dynamic Meteorology and Oceanography, 51, 195–221, https://doi.org/10.3402/tellusa.v51i2.12316 , 1999.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Quarterly Journal of the Royal Meteorological Society, 125, 723–757, https://doi.org/https://doi.org/10.1002/qj.49712555417 , 1999.

Hamill, T. M. and Snyder, C.: A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme, Monthly Weather Review, 128, 2905 – 2919, https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2 , 2000.

Jung, B.-J., Ménétrier, B., Snyder, C., Liu, Z., Guerrette, J. J., Ban, J., Baños, I. H., Yu, Y. G., and Skamarock, W. C.: Three-dimensional variational assimilation with a multivariate background error covariance for the Model for Prediction Across Scales–Atmosphere with the Joint Effort for data Assimilation Integration (JEDI-MPAS 2.0.0-beta), Geosci. Model Dev., 17, 3879–3895, https://doi.org/10.5194/gmd-17-3879-2024 , 2024.

Ménétrier, B.: Normalized Interpolated Convolution from an Adaptive Subgrid documentation, https://github.com/benjaminmenetrier/nicas_doc/blob/master/nicas_doc.pdf , 2020.

Wu, W.-S., Purser, R. J., and Parrish, D. F.: Three-dimensional variational analysis with spatially inhomogeneous covariances, Monthly Weather Review, 130, 2905–2916, https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2 , 2002.