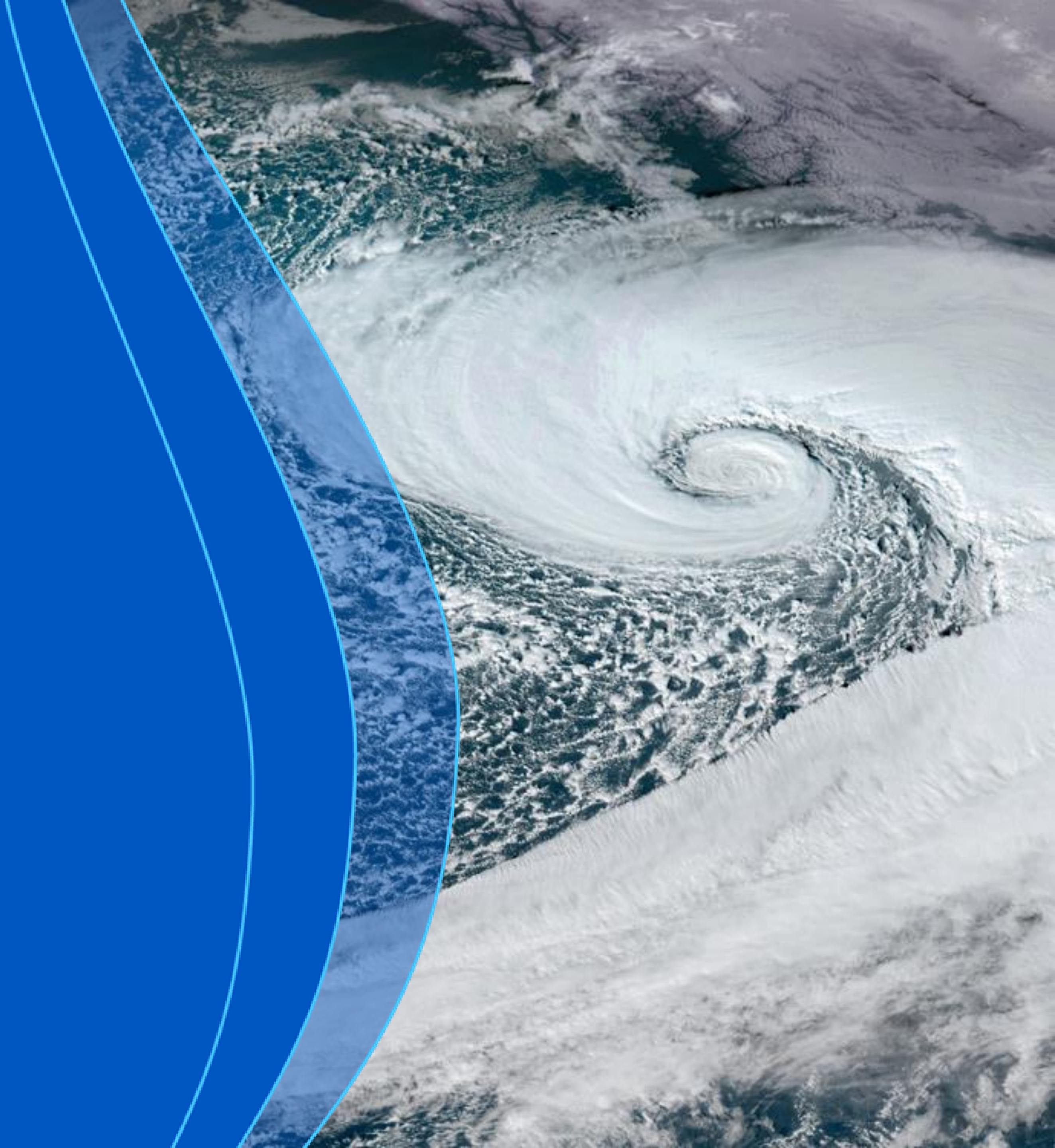
August 13, 2026

MPAS-JEDI 3D/4DEnVar

Materials prepared by Jake (Zhiquan) Liu, I-Han Chen, and Robert Nystrom

Presented by Junmei Ban



Overview

1. **Variational Cost Function**
2. Ensemble Error Covariance Matrix
3. Overview of 3DEnVar
4. Setting up a .yaml file for 3DEnVar
5. Overview of 4DEnVar
6. Setting up a .yaml file for 4DEnVar

The Problem to solve

We want to find the **analysis state** (x) that minimizing a cost function with an optimal fit to the background and observations.

$$\boldsymbol{J}(x) = \boxed{\frac{1}{2}(x - x_b)^T \boldsymbol{B}^{-1}(x - x_b)} + \boxed{\frac{1}{2}(h(x) - y)^T \boldsymbol{R}^{-1}(h(x) - y)}$$

Distance to background

Distance to observations

$\boldsymbol{B}$: background error covariance matrix

$\boldsymbol{R}$: observation error covariance matrix

Incremental Cost Function in JEDI

Original-form:

$$J(x) = \frac{1}{2}(x - x_b)^T \boldsymbol{B}^{-1}(x - x_b) + \frac{1}{2}(h(x) - y)^T \boldsymbol{R}^{-1}(h(x) - y)$$

Incremental-form:

$$J(\delta x) = \frac{\mathbf{1}}{\mathbf{2}}(\delta x - \delta x_g)^T \boldsymbol{B}^{-1}(\delta x - \delta x_g) + \frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{H}\delta x - d)^T \boldsymbol{R}^{-1}(\boldsymbol{H}\delta x - d)$$

$\boldsymbol{\delta x = x - x_g}$ $\quad \delta x_g = x_b - x_g$ $\quad \boldsymbol{d = y - h(x_g)}$

The minimization deals with increments to a known reference state

- Cost function minimizes δx=x− x_g instead of the full state (x)
- Start from $x_g = x_b$ and $\delta x_g = 0$
- **After minimization->** $x_a = x_g + \delta x$

About dual-resolution

- Analysis increment and ensemble can be at a lower resolution than the background and analysis

$$J(\delta x) = \frac{1}{2}\left(\delta x - \delta x_g\right)^T \boldsymbol{B}^{-1}\left(\delta x - \delta x_g\right) + \frac{1}{2}(\boldsymbol{H}\delta x - d)^T \boldsymbol{R}^{-1}(\boldsymbol{H}\delta x - d)$$

$$\boldsymbol{\delta x = x - x_g} \qquad \delta x_g = x_b - x_g \qquad \boldsymbol{d = y - h(x_g)}$$

- Hofx : based on the high-resolution nonlinear background state
- inner-loop minimization : low-resolution
- After converge, the low-res analysis increment is interpolated up to high resolution, and added to the high-res background state, to obtain the final analysis.

Practical session 4.3: dual-resolution 3DEnVar test case at 120km-240km

Appropriately assign B and R is critical

We want to find the analysis state (x) that minimizing a cost function with **an optimal fit** to the background and observations.

Distance to background **Distance to observations**

$$J(\delta x) = \boxed{\frac{1}{2}(\delta x - \delta x_g)^T B^{-1} (\delta x - \delta x_g)} + \boxed{\frac{1}{2}(H\delta x - d)^T R^{-1} (H\delta x - d)}$$

The weighting between the two components is determined by B (background error) and R (observation error).

- A larger B means background is less accurate -> x will get closer to observation
- A larger R means observation is less accurate -> x will get closer to background

Two types of background error covariance (B)

$$J(\delta x) = \frac{1}{2}\left(\delta x - \delta x_g\right)^T B^{-1}\left(\delta x - \delta x_g\right) + \frac{1}{2}(H\delta x - d)^T R^{-1}(H\delta x - d$$

1. **Static B**

   -> from statistic (e.g. a month of model forecasts), does not vary with time

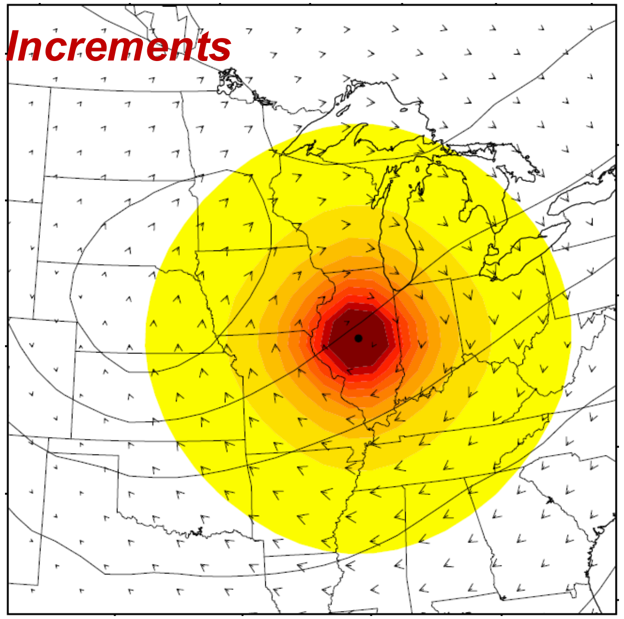
2. **Ensemble B**

   -> flow-dependent, reflect the background error in different time

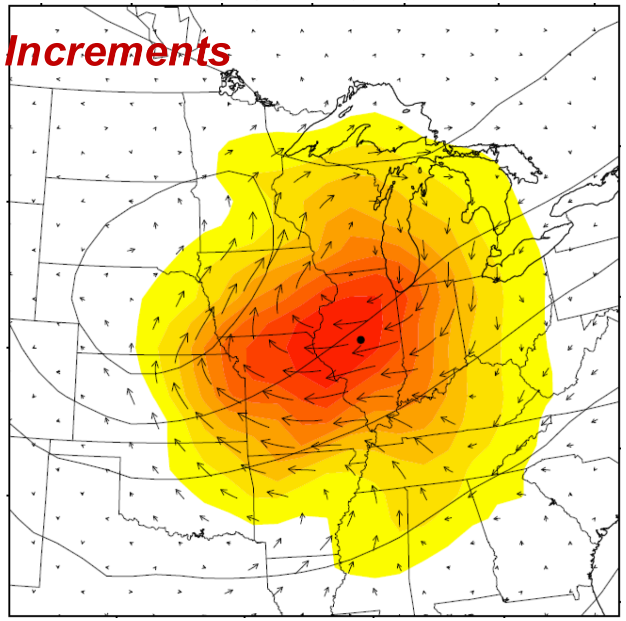
Example to show the B effect (Single observation tests)

Assimilate a single Temperature observation

Static B

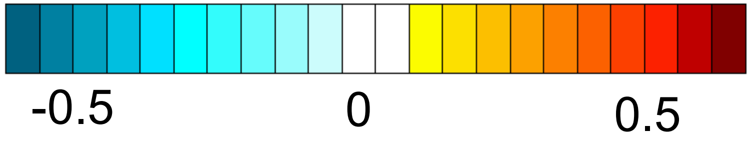


Ensemble B



Ensemble $\boldsymbol{B}$:

- Errors of the day are sampled
- flow-dependent update



Increments of temperature (shaded) and horizontal winds (vector)

Overview

1. Variational Cost Function
2. **Ensemble Error Covariance Matrix**
3. Overview of 3DEnVar
4. Setting up a .yaml file for 3DEnVar
5. Overview of 4DEnVar
6. Setting up a .yaml file for 4DEnVar

Derive B matrix from an ensemble of forecasts

ensemble size

State variable of ensemble mean

$$\boldsymbol{B_e} = \frac{1}{n-1}\sum_{i=1}^{\boldsymbol{n}}(\boldsymbol{x_i} - \boldsymbol{\bar{x}})(x_i - \bar{x})^T$$

State variable of each ensemble member

$$\boldsymbol{B_e} = \frac{1}{n-1}\sum_{i=1}^{\boldsymbol{n}}(\delta x_{\boldsymbol{i}})(\delta x_{\boldsymbol{i}})^T$$

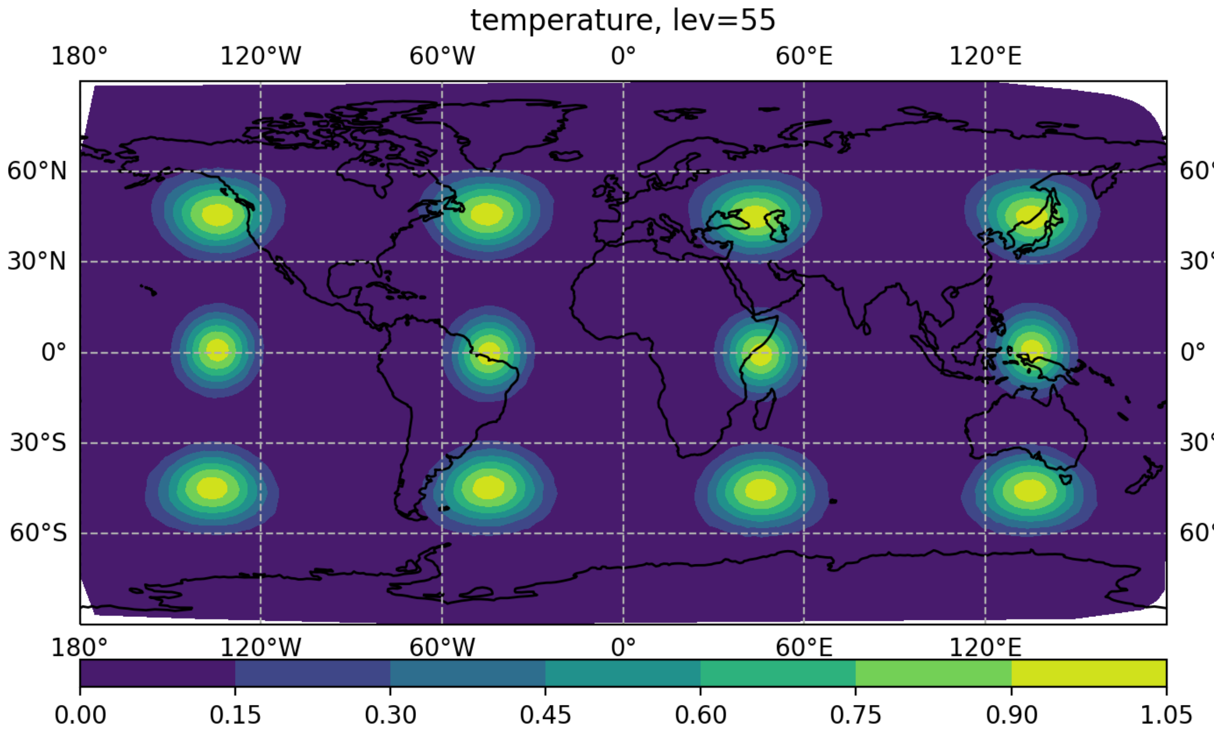
ensemble perturbation

- Sample covariance matrix
- The ensemble mean provides an estimation of the truth
- The perturbations from the mean estimate the uncertainty, which is used to model background-error covariance matrix.
- Ensemble spread is the standard deviation from ensemble mean.

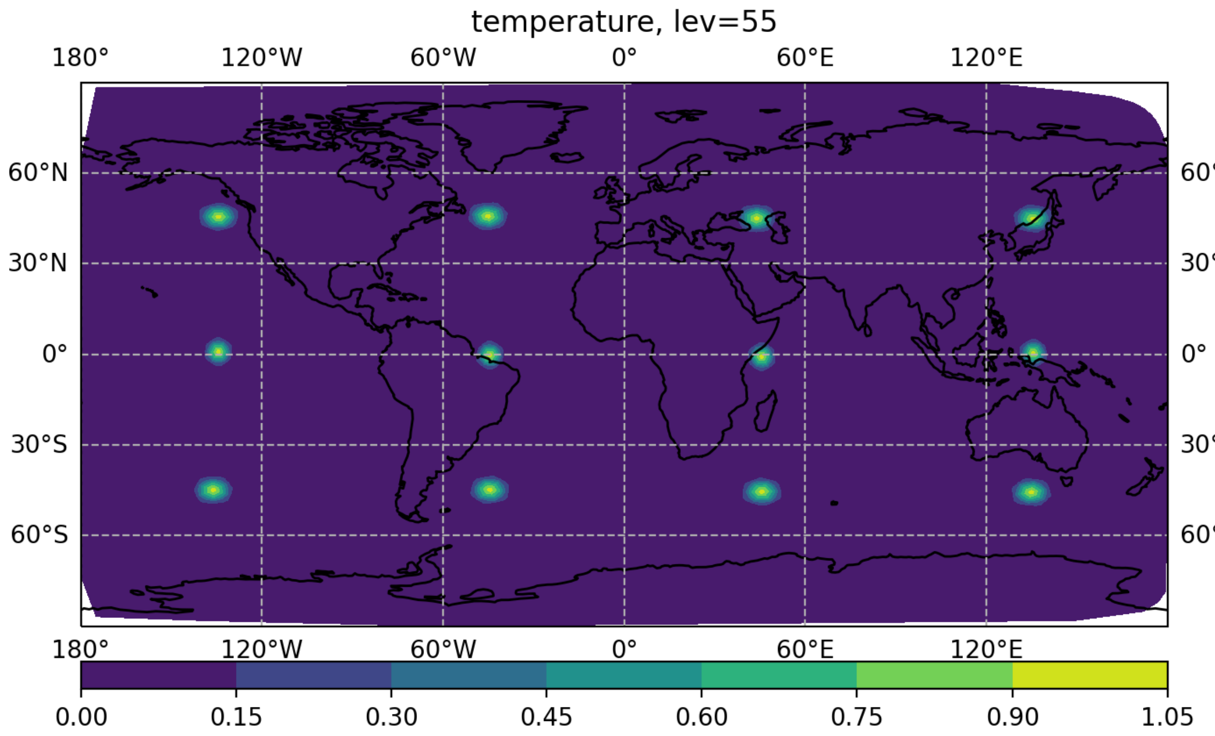
Localization of ensemble B matix: $B = L \circ B_e$

L is a correlation matrix, to reduce spurious correlation due to limited ensemble size

larger localization scales



Smaller localization scale



Influence a a small area

EnVar uses a pure ensemble B to updates a deterministic forecast

In hybrid methods, B can be a weighted sum of static B ($\boldsymbol{B_s}$) and ensemble B ($\boldsymbol{B_e}$).

$$\boldsymbol{B} = \beta_s \boldsymbol{B_s} + \beta_e \boldsymbol{B_e}$$

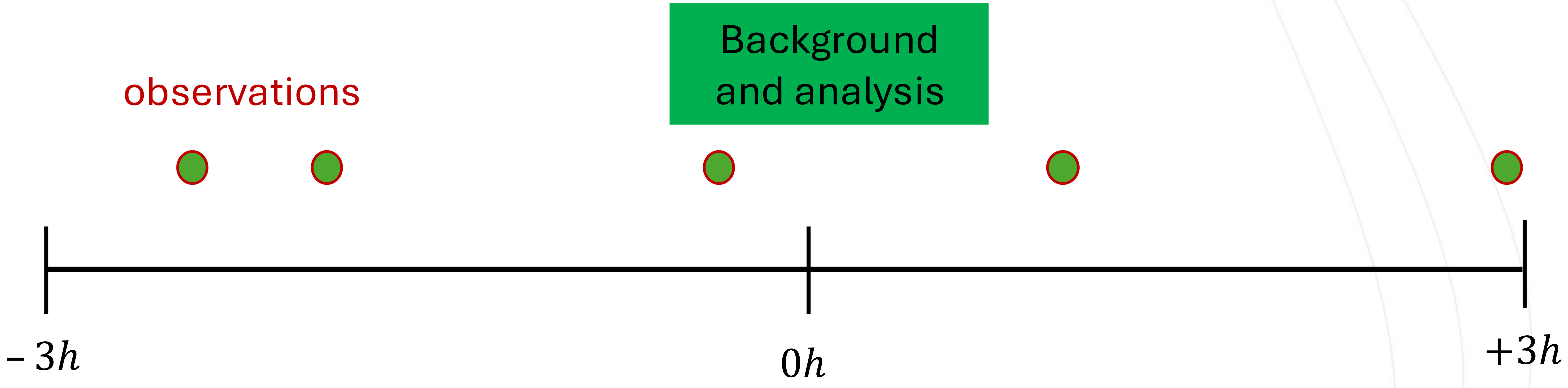
$$\underbrace{\beta_s}_{=0} + \underbrace{\beta_e}_{=1} = 1$$

pure ensemble B

Overview

1. Variational Cost Function
2. Ensemble Error Covariance Matrix
3. **Overview of 3DEnVar**
4. Setting up a .yaml file for 3DEnVar
5. Overview of 4DEnVar
6. Setting up a .yaml file for 4DEnVar

3DEnVar using a 6h assimilation window



- All observations in 3DEnVar are assumed to be valid at the same time as the background.
- Innovation $(\boldsymbol{Hx} - \boldsymbol{y_o})$ is calculated relative to background at that single time

Overview

1. Variational Cost Function
2. Ensemble Error Covariance Matrix
3. Overview of 3DEnVar
4. **Setting up a .yaml file for 3DEnVar**
5. Overview of 4DEnVar
6. Setting up a .yaml file for 4DEnVar

Configure the analysis time for 3DEnvar

```
_member config: &memberConfig
  date: &analysisdate '2018-04-15T00:00:00Z'
  state variables: &incvars
  - air_temperature
  - water_vapor_mixing_ratio_wrt_moist_air
  - eastward_wind
  - northward_wind
  - air_pressure_at_surface
  stream name: ensemble
cost function:
  cost type: 3D-Var
  time window:
    begin: '2018-04-14T21:00:00Z'
    length: PT6H
  geometry:
    nml_file: "./Data/480km/namelist.atmosphere_2018041500"
    streams_file: "./Data/480km/streams.atmosphere"
    deallocate non-da fields: true
  analysis variables: *incvars
  background:
    state variables: [air_temperature, water_vapor_mixing_ratio_wrt_moist_air...]
    filename: "./Data/480km/bg/restart.2018-04-15_00.00.00.nc"
    date: *analysisdate
```

analysis time (center of window)

~test/testinput/3denvar_bumploc.yaml

Start of assimilation window

length of assimilation window

background at analysis time

Configure the ensemble B

```yaml
background error:
  covariance model: ensemble
  localization:
    localization method: SABER
    saber central block:
      saber block name: BUMP_NICAS
      active variables: *incvars
      read:
        io:
          files prefix: Data/bump/mpas_parametersbump_loc
        drivers:
          multivariate strategy: duplicated
          read local nicas: true
  members:
  - filename: Data/480km/bg/ensemble/mem01/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig
  - filename: Data/480km/bg/ensemble/mem02/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig
  - filename: Data/480km/bg/ensemble/mem03/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig
  - filename: Data/480km/bg/ensemble/mem04/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig
  - filename: Data/480km/bg/ensemble/mem05/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig
```

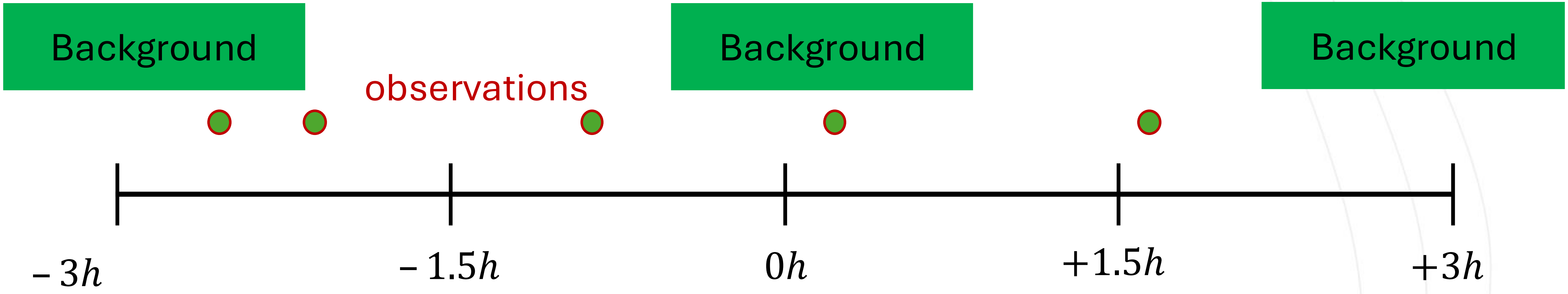
set ensemble B for 3DEnVar

Specifying members used to compute ensemble B

Overview

1. Variational Cost Function
2. Ensemble Error Covariance Matrix
3. Overview of 3DEnVar
4. Setting up a .yaml file for 3DEnVar
5. **Overview of 4DEnVar**
6. Setting up a .yaml file for 4DEnVar

4DEnVar



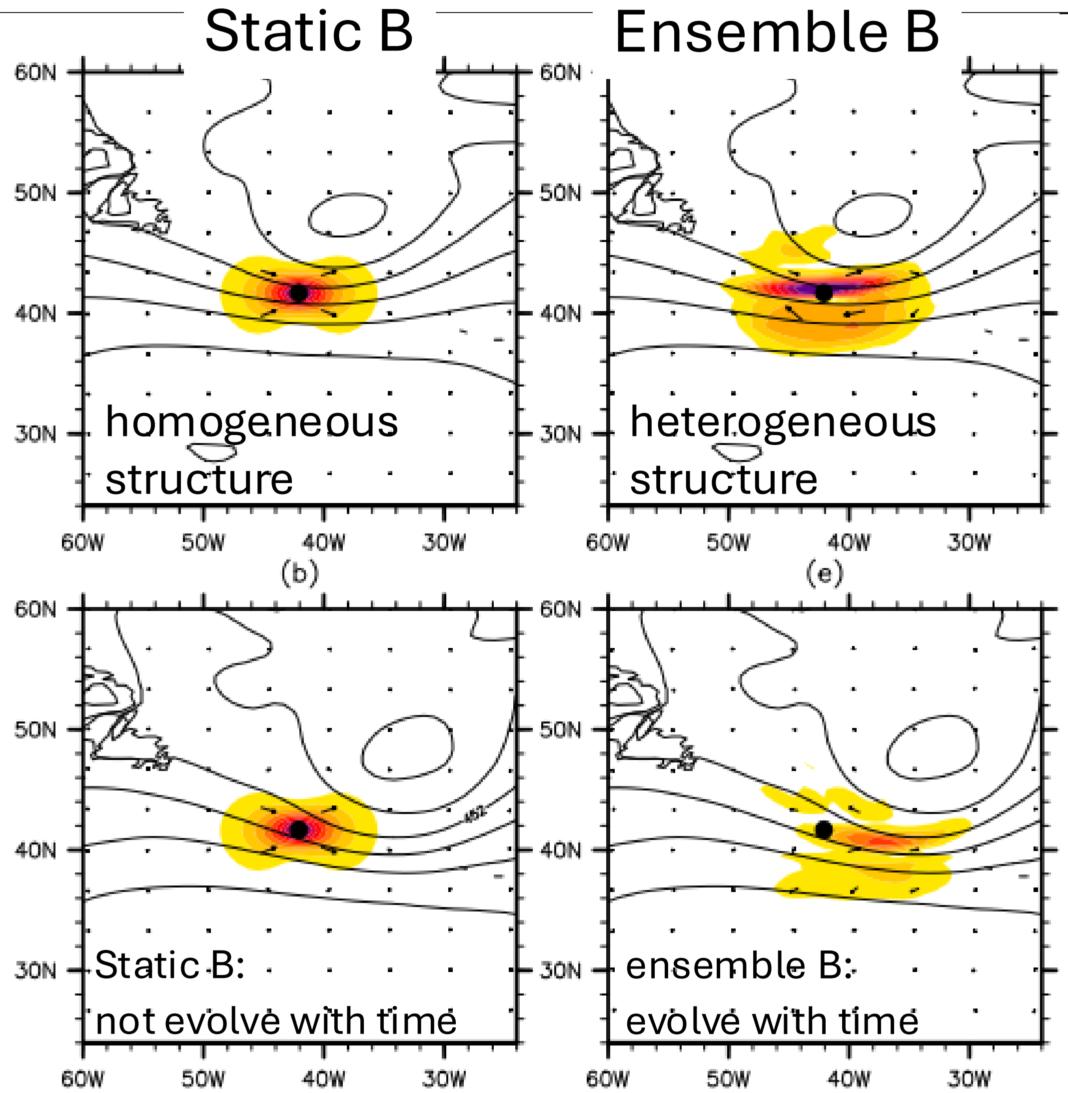
- All observations in 4DEnVar are binned within a smaller subwindow and innovations ($\boldsymbol{Hx} - \boldsymbol{y}_o$) are calculated relative to background valid at that time.
- Ensemble needed at the center of each subwindow ($\boldsymbol{K}$ ensemble required).

The 4D ensemble $\boldsymbol{B}$ is used to propagate the innovation

Analysis increment

Start of window

end of window



Lorenc et al. (2015)

increments have moved downstream by the end of the assimilation window

Overview

1. Variational Cost Function
2. Ensemble Error Covariance Matrix
3. Overview of 3DEnVar
4. Setting up a .yaml file for 3DEnVar
5. Overview of 4DEnVar
6. **Setting up a .yaml file for 4DEnVar**

Configure the analysis times for 4DEnvar

```yaml
- _member config 1: &memberConfig1
    date: &date1 '2018-04-14T21:00:00Z'
    state variables: &incvars
    - air_temperature
    - water_vapor_mixing_ratio_wrt_moist_air
    - eastward_wind
    - northward_wind
    - air_pressure_at_surface
    stream name: ensemble
  _member config 2: &memberConfig2
    <<: *memberConfig1
    date: &date2 '2018-04-15T00:00:00Z'
  _member config 3: &memberConfig3
    <<: *memberConfig1
    date: &date3 '2018-04-15T03:00:00Z'
  cost function:
    cost type: 4D-Ens-Var
    time window:
      begin: '2018-04-14T21:00:00Z'
      length: PT6H
    subwindow: PT3H
```

subwindow1

subwindow2

subwindow3

test/testinput/4denvar_bumploc.yaml

Background needed for each subwindow

```
background:
  states:
  - state variables: &stvars
    filename: "./Data/480km/bg/restart.2018-04-14_21.00.00.nc"
    date: *date1
  - state variables: *stvars
    filename: "./Data/480km/bg/restart.2018-04-15_00.00.00.nc"
    date: *date2
  - state variables: *stvars
    filename: "./Data/480km/bg/restart.2018-04-15_03.00.00.nc"
    date: *date3
```

[air_temperature, water_vapor_mixing_ratio_wrt_moist_air, eastward_wind, northward_wind, air_pressure_at_surface, air_potential_temperature, dry_air_density, u, water_vapor_mixing_ratio_wrt_dry_air, air_pressure, landmask, seaice_fraction, snowc …]

bg (subwindow 1)

bg (subwindow 2)

bg (subwindow 3)

Configure the ensemble B

```yaml
background error:
  covariance model: ensemble
  localization:
    localization method: SABER
    saber central block:
      saber block name: BUMP_NICAS
      active variables: *incvars
      read:
        io:
          files prefix: Data/bump/mpas_parametersbump_loc
        drivers:
          multivariate strategy: duplicated
          read local nicas: true
```

set ensemble B for 4DEnVar

Ensemble files needed for each subwindow

```yaml
members:
- states:
  - filename: Data/480km/bg/ensemble/mem01/x1.2562.init.2018-04-14_21.00.00.nc
    <<: *memberConfig1
  - filename: Data/480km/bg/ensemble/mem01/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig2
  - filename: Data/480km/bg/ensemble/mem01/x1.2562.init.2018-04-15_03.00.00.nc
    <<: *memberConfig3
- states:
  - filename: Data/480km/bg/ensemble/mem02/x1.2562.init.2018-04-14_21.00.00.nc
    <<: *memberConfig1
  - filename: Data/480km/bg/ensemble/mem02/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig2
  - filename: Data/480km/bg/ensemble/mem02/x1.2562.init.2018-04-15_03.00.00.nc
    <<: *memberConfig3
- states:
  - filename: Data/480km/bg/ensemble/mem03/x1.2562.init.2018-04-14_21.00.00.nc
    <<: *memberConfig1
  - filename: Data/480km/bg/ensemble/mem03/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig2
  - filename: Data/480km/bg/ensemble/mem03/x1.2562.init.2018-04-15_03.00.00.nc
    <<: *memberConfig3
- states:
  - filename: Data/480km/bg/ensemble/mem04/x1.2562.init.2018-04-14_21.00.00.nc
    <<: *memberConfig1
  - filename: Data/480km/bg/ensemble/mem04/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig2
  - filename: Data/480km/bg/ensemble/mem04/x1.2562.init.2018-04-15_03.00.00.nc
    <<: *memberConfig3
- states:
  - filename: Data/480km/bg/ensemble/mem05/x1.2562.init.2018-04-14_21.00.00.nc
    <<: *memberConfig1
  - filename: Data/480km/bg/ensemble/mem05/x1.2562.init.2018-04-15_00.00.00.nc
    <<: *memberConfig2
  - filename: Data/480km/bg/ensemble/mem05/x1.2562.init.2018-04-15_03.00.00.nc
    <<: *memberConfig3
```

A simpler way to configure ensemble input

```yaml
members from template:
  template:
    states:
    - state variables: *incvars
      date: 2018-04-14T21:00:00Z
      filename: ../ensemble/2018041418/%iMember%/mpasout.2018-04-14_21.00.00.nc
    - state variables: *incvars
      date: 2018-04-15T00:00:00Z
      filename: ../ensemble/2018041418/%iMember%/mpasout.2018-04-15_00.00.00.nc
    - state variables: *incvars
      date: 2018-04-15T03:00:00Z
      filename: ../ensemble/2018041418/%iMember%/mpasout.2018-04-15_03.00.00.nc
  pattern: %iMember%
  start: 1
  zero padding: 2
  nmembers: 20
```

See 4denvar.yaml in practical session 4.4

Nodes/Cores setting

3DEnVar

```
#PBS -l select=1:ncpus=64:mpiprocs=64:mem=235GB
```

4DEnVar with 3 sub-windows

```
#PBS -l select=3:ncpus=64:mpiprocs=64:mem=235GB
```

References

- Liu, Z., and Coauthors, 2022: Data assimilation for the Model for Prediction Across Scales - Atmosphere with the Joint Effort for Data assimilation Integration (JEDI-MPAS 1.0.0): EnVar implementation and evaluation. *Geosci. Model Dev.*, **15**, 7859–7878, https://doi.org/10.5194/gmd-15-7859-2022.

- Lorenc, A. C., N. E. Bowler, A. M. Clayton, S. R. Pring, and D. Fairbairn, 2015: Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Mon. Weather Rev.*, **143**, 212–229, https://doi.org/10.1175/MWR-D-14-00195.1.

- Nystrom, R. G., Snyder, C., Liu, Z., Jung, B. J., Ban, J., & Banos, I. H. (2025). A hybrid four-dimensional variational data assimilation system for the Model for Prediction Across Scales (MPAS-atmosphere): Leveraging the Joint Effort for Data assimilation Integration (JEDI). *Journal of Advances in Modeling Earth Systems*, 17, e2025MS005183. https://doi.org/10.1029/2025MS005183