JEDI experiments for DARC/NCEO DA training course 2025

The Joint Effort for Data assimilation Integration (JEDI) is the unified data assimilation (DA) framework collaboratively developed by multiple partners under the Joint Center for Satellite Data Assimilation (JCSDA). The Met Office has adopted the JEDI code framework to develop its next-generation observation processing and assimilation system. JEDI is structured around a collection of modular repositories, each representing a key component of a DA system. These components are integrated and executed according to user-defined YAML configuration files, which specify how the building blocks are assembled and interact (see below for an illustration).

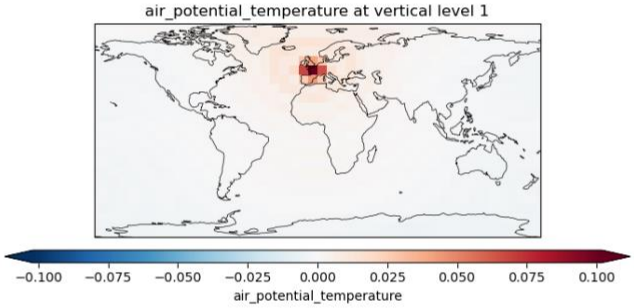
In this course, we will get a taste of how JEDI works through two practical experiments. The first experiment uses the System Agnostic Background Error Representation (SABER) repository to compute the background error covariance matrix (commonly known as the **B** matrix). In this experiment, we will see how the **B** matrix spread observation information into the model space. The second experiment involves the Object-Oriented Prediction System (OOPS) repository, which implements most of DA algorithms. Here, we will focus on one popular DA algorithm: the Local Ensemble Transform Kalman Filter (LETKF). We will adjust key parameters, including ensemble size, localization radius and inflation factor, to see how they affect the performance of the LETKF. More details of these experiments are given below.

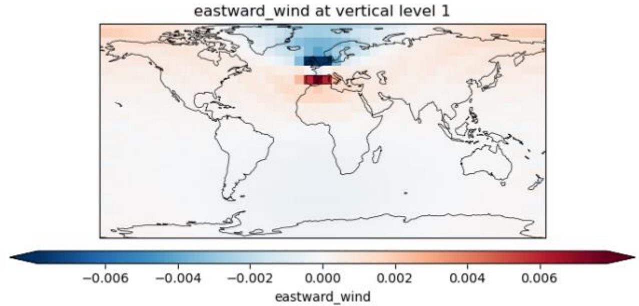
The **B** matrix experiment

We use a static **B** matrix similar to that used in the Met Office system. To explore how observation information is distributed by the **B** matrix, we apply an impulse to a specific grid point for a chosen model variable. We then look at

1. how this impulse introduces increments to other grid points for the same variable, and
2. how it introduces changes in other model variables due to the physical balance among atmospheric quantities.

For example, we can apply an impulse at longitude 0.0, latitude 50.0 and model level 1 in the air potential temperature field. The resulting increments in the air potential temperature (left) and eastward wind (right) fields are shown below.

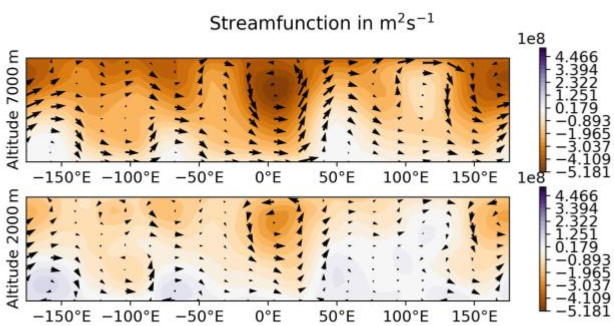




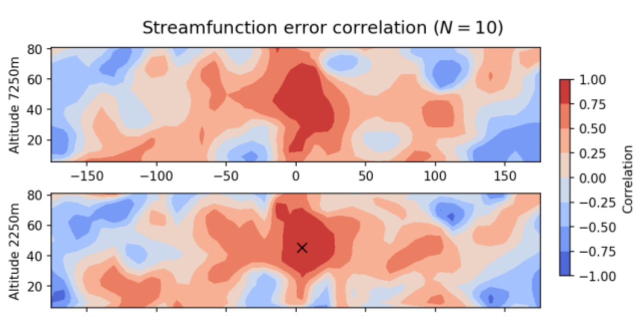
We may also explore the interactions between eastward and northward winds, as well as between air density and air potential temperature. This experiment provides flexibility to explore a wide range of model variables.

The LETKF experiment

In this experiment, we use a two-level quasi-geostrophic toy model. The figure below shows the true streamfunction field at the assimilation time, with the arrows indicating the wind vectors. In addition, we may also see the wind field derived from the streamfunction.

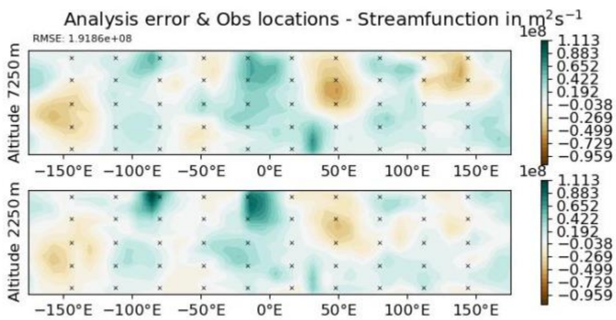


The LETKF is a widely used type of ensemble Kalman filters. It uses an ensemble estimate of the **B** matrix. As a first step, we examine the structure of this ensemble covariance using different ensemble sizes. The figure below shows the correlation between each model grid point and a selected point (marked with '×'). In this experiment, the ensemble size can vary from 2 to 100.



In addition to ensemble size, the localization radius and inflation factor also affect the performance of LETKF. In this experiment, we will tune the LETKF system using these three parameters to evaluate their impact and identify configurations that give the most accurate analysis. Furthermore, three synthetic observation files are provided: 1) observations regularly distributed across the entire globe, 2) observations regularly distributed on the southern half of the domain, and 3) observations regularly distributed on the northern half of the domain. There are three observation types: streamfunction, wind and wind speed.

The plot below shows the difference between the analysis and the truth fields, with observation locations marked by '×'. The Root Mean Square Error (RMSE) for the analysis is indicated at top-left of the upper panel.



The commands required to run these experiments are provided on the following pages.

How to Run JEDI Experiments

Set Up Working Directories

Copy batch script to your home directory:

```
cp /storage/research/nceo/DA-training-course/docs/set_up_exps.sh .
```

Run the script:

```
bash set_up_exps.sh
```

Two folders (letkf_exps and saber_exps) should be created.

Experiments with Background Error Covariances (SABER)

Navigate to the saber_exps directory:

```
cd saber_exps
```

Edit “dirac_spectralb_gauss_vader_1.yaml”: Select a model variable at line 162 (as guided by in-file comments). The default variable is “air_potential_temperature”.

Run the experiment:

```
bash run_saber.sh
```

This will generate a NetCDF file named “output.nc”.

Load anaconda for Python:

```
module load anaconda/2023.09-0/met-env
```

Plot results (replace <variable> and <level> with allowed options):

```
python plot_saber.py output.nc <variable> <level>
```

View allowed options of <variable> and <level>:

```
python plot_saber.py -h
```

Example:

```
python plot_saber.py output.nc eastward_wind 1
```

This will generate a PNG file named “figure_ eastward_wind.png”. The plot can be viewed by

display figure_ eastward_wind.png

LETKF Experiments

Navigate to the letkf_exps directory:

cd letkf_exps

Plot background error covariances (replace <variable> and <ensemble_size> with allowed options):

python plot_ensemble_covariance.py <variable> <ensemble_size>

View allowed options of <variable> and <ensemble_size>:

python plot_ensemble_covariance.py -h

Example:

python plot_ensemble_covariance.py x 5

This will generate a PNG file named "correlation_fields.png".

Edit "letkf.yaml" to change:

- Number of ensemble members (edit "nmembers:" at line 17)
- Localization radius (edit "lengthscale:" at lines 37, 50 and 63)
- Covariance inflation factor (edit "mult:" at line 75)
- Observation files (edit "obsfile:" at line 27)

Additionally, specific observation types can be excluded from the assimilation by commenting out their corresponding blocks in the "letkf.yaml" file. For example, to exclude wind speed observations (WSpeed), comment out lines 51-63.

Run the experiment:

bash run_letkf.sh

This will generate three NetCDF files in output folder.

Plot (need to load anaconda):

bash draw_letkf.sh

We need to specify which observation file to use for plotting by opening "draw_letkf.sh" and selecting one of the three available observation files. This will generate eight JPG files: four showing the true fields of the variables q, u, v, and x, respectively; and four showing the corresponding analysis error fields for those variables.