JEDI-LETKF Experiments

DARC/NCEO data assimilation training course 2025

This practical uses the Object-Oriented Prediction System (OOPS) repository within the Joint Effort for Data assimilation Integration (JEDI) framework. OOPS implements most data assimilation algorithms and includes two toy models for experimentation.

We use a two-level quasi-geostrophic (QG) model to explore the performance of the Local Ensemble Transform Kalman Filter (LETKF) under various configurations. We will examine how factors such as ensemble size, localization radius and inflation factor, as well as observation distributions and types affect the resulting analysis.

This practical will give us a better understanding of

- *How background error statistics depend on ensemble size*
- *How observation networks affect the analysis accuracy*
- *How to tune the LETKF system to reduce analysis error*

Step-by-step instructions are provided below.

Step 1: Set up the experiment by running the following commands (after logging to RACC)

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

bash setup_letkf_exps.sh

cd letkf_exps
```

Step 2: Run the LETKF

```
bash run_letkf.sh
```

This produces three NetCDF files in the "output" folder.

Alternatively, we can submit the job to a compute node for faster execution:

```
sbatch run_letkf.sh
```

Step 3: Load the Anaconda environment for Python

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

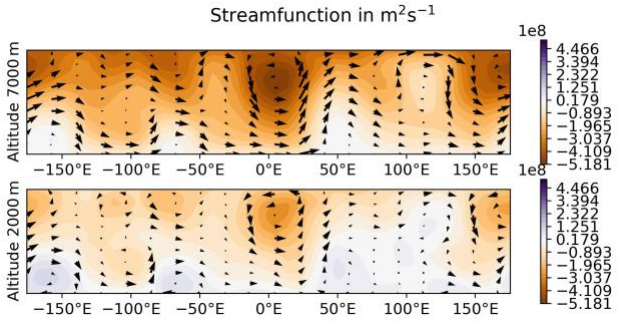
Step 4: Plot the results:

```
bash draw_letkf.sh
```

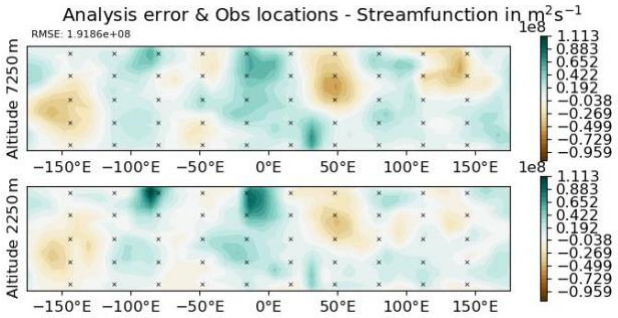
This generates eight JPG files:

- Four plots of the true fields: streamfunction (x), potential vorticity (q), eastward wind (u) and northward wind (v) of the two-level QG model.
- Four plots of the corresponding analysis error fields.

For example, the figure below shows the true streamfunction field, with the arrows indicating the wind vectors.



In addition, the figure 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.



Questions to discuss:

1. *How do background error statistics vary with different ensemble sizes?*
2. *How does the spatial distribution of observations influence the analysis field?*
3. *How is the analysis variable (streamfunction) related to different observation types via the observation operator? Which type contributes most effectively to improving the analysis?*
4. *How does changing the localization radius affect the analysis error? Why is localization a critical component of ensemble Kalman filters?*
5. *How does the analysis respond to changes in the inflation factor?*

At the end of the practical, you are welcome to share your configuration that produced the best analysis.

Hints:

a) Explore background error covariances

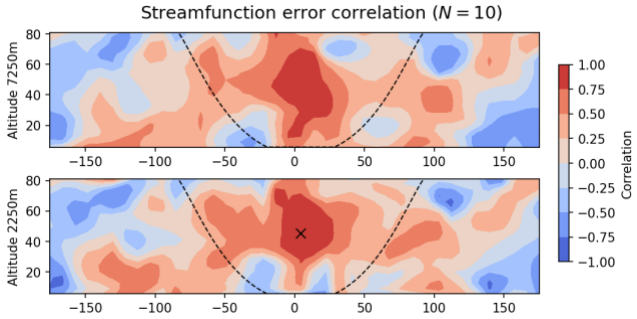
To visualize the correlation between a selected point and all grid points in physical space, run

python plot_ensemble_covariance.py x <ensemble_size> <radius_m>

Replace <ensemble_size> with an integer between 2 and 100.

Replace <radius_m> with the localization radius in metres, e.g., 5e6.

This generates "correlation_fields.png" (for an example see below). The selected point is marked with '×', and the dashed line outlines the localisation domain.



b) Use different observation files

Three observation files are pre-generated:

- Regularly distributed observations over the entire globe
- Regularly distributed observations on the southern half of the domain
- Regularly distributed observations on the northern half of the domain

To change the observation file:

- Edit line 27 in "letkf.yaml"
- Edit line 8 in "draw_letkf.sh"

Re-run Steps 2 and 4.

c) Select observation types

There are three observation types:

- Streamfunction
- Eastward and northward winds
- Wind speed

To exclude one type from the assimilation, comment out its corresponding block in "letkf.yaml" (add # at the beginning of the line). For example, to exclude wind speed observations (WSpeed), comment out lines 51-63.

Re-run Steps 2 and 4.

d) Tune LETKF parameters

Paramter	**Edit "letkf.yaml"**	**Notes**
Ensemble size	Line 17 (nmembers:)	Range: 2-100
Localization radius	Line 37 (lengthscale:)	Try e.g., 5e6 → 3e6 (metres)
Inflation factor	Line 75 (mult:)	Multiplicative factor

Re-run Steps 2 and 4.