## Exercises - GEO4902-02 - Data assimilation

**The assignment is due XY , 2022**

***[Save google-collab as jupyter notebook, with your name in the filename, and upload on canvas]***

### Exercise 1 - Analyse the impact of “upper-air” data assimilation in forecasts

**Use:**

<https://github.com/franzihe/GEO4902_2021/blob/main/02/ModelAnalysis_and_background_upperair.ipynb>

The script gives you a framework to assess two different forecasts. Note, forecasts are available with initialization dates every 3 hours, i.e. we assimilate every three hours (rapid update cycling), but we do full forecasts only every 6 hours. You can thus analyse two different forecasts (i.e. which were initialized at different times) for a specific time.

1. Utilize the script to plot the impact of data assimilation on a point forecast. That means plot 2 or more forecasts with different initialization times in one figure. Use also at least 2 additional variables relevant for a weather forecast.
2. Utilize the script to plot a 2D-map of the impact of the data assimilation, this means the difference between the background xB and the analysis xA. In order to find the background, think about what goes into the data assimilation procedure from the previous forecast? Plot maps for the different variables in (1) and for different levels.

### Exercise 2 - Analyse the impact of “surface” data assimilation in forecasts

**Use:**

<https://github.com/franzihe/GEO4902_2021/blob/main/02/ModelAnalysis_and_background-SurfaceModel.ipynb>

It follows the same logic as Exercise 1 but this time for the (land-)surface model.

1. Utilize the script to plot the impact on point forecasts. That means plot 2 or more forecasts in one figure. Use also different variables.
2. Utilize the script to plot a 2D-map of the impact of the data assimilation, this means the difference between the background xB and the analysis xA. In order to find the background think about what goes into the data assimilation procedure from the previous forecast? Plot maps for different variables (levels are all surface (“screen”) levels).
3. Describe and interpret the difference between the result in Exercise 1

### 

### Exercise 3 - Data assimilation: Single observation experiment

**Use:**

<https://github.com/franzihe/GEO4902_2021/blob/main/02/SingleObservationExperiment.ipynb>

For (a) testing new observations, (b) analyzing their impact, and (c) understanding the impact of the data assimilation system we often perform so-called “*single-observation experiments*”. Only one single observation is included in the data assimilation system and we can compare the background and the analysis. In the following example, a radiosonde temperature observation at 500 hPa height has been assimilated.

The background xB and analysis xA include the prognostic variables of *temperature*, *wind*, and *specific humidity.* The data files are for the Norwegian area and distributed in 60 vertical levels. The python script for this exercise gives you the access points to this data and two plotting routines, in order to plot vertical sections and horizontal layers of the model data.

1. Find out where the radiosonde was assimilated. (level and geographical location)
2. Estimate the spatial impact this one observation has.
3. What is the impact on wind and humidity, and why&how could that be realized in the data assimilation system?

### Exercise 4 - 2D data assimilation example

**Use:**

<https://colab.research.google.com/github/maltemuellerm/GEO4902_2021/blob/main/02/Data_assimilation_2Dexample.ipynb>

This script (1) visualizes a 2D Multivariate Gaussian Distribution and (2) guides you through all steps of a data assimilation cycle by assuming a two-dimensional model background and two observations. (The observation operator is the identity operator.)

1. Run the entire script and understand the steps. Describe with your own words the last part, i.e. prior, observation, and posterior and how it relates to what we have talked about in the lectures.
2. What can you say about the Gaussian distribution of the background compared to the optimal solution? How can you interpret the differences between them?
3. Degrade the accuracy significantly of the background and/or observation (of one and both points). What do you observe concerning the optimal solution?

Reflection notes by students

## Reflection Note

First of all, the assignment was very interesting and fun to work with! Working with the exersice i continue to improve my coding skills, eventhough its a long way to go before i feel comfortable with the syntax. I learned that producing a weather forecast as accurate as possible, represent the "current state of the atmosphere" in the variables needed for a weather forecast, and put them on a spatial grid we need data assimilation. This is because we want to produce the best initial condition for the next or new weather forecast. Then we need to use the observations and the model information on the overlapping hours and analyse the deviation (errors) between them.

What was difficult - even if i know what to do, or at least understand the question, i always end up with spending tremendous time on index errors and attribute errors. tuples out of range, not included or missing attributes and i feel thats quite confusing because when it suddenly work im not quite sure what i did (ehhh, or how the error was connected to the metadata, or indexing within the variable.)

What was surprising - finding the location of the radiosonde was surprising since i did not see any change before layer 20-25.

What elements i will take with me - better understanding of the code. Also, ask for help a little bit sooner.

Where would you get more input - more input on data assimilation and e.g the colab git hub file for exersize 4 maybe more information about the calculation (more in depth on what is done.). I think working on these exersice and comparing with the solution is very good.

What did I learn?

- Got a better understanding of how the background, observation and analysis

state vector are used in weather models, and how they are used to get as accurate

forecasts as possible.

What was difficult?

- Struggled to get the vertical plots in Exercise 2 to run, but this became clear after

the last group session. The “.rename” function worked like a charm.

- I often struggle to read the data files from thredds. I usually have to get the url

directly from the server, and I still do not understand as to why… There should

not be any differences between these, and the ones provided in the already made

code examples.

- The theory behind the fourth exercise and the respective code were also

something that I used quite some time wrapping my head around, but it was a

great help to understand the general concept better.

What went well?

- After struggling with getting the different codes to run, I feel like I managed to

answer the respective questions and connect the dots in a way.

What was surprising?

- How intricate the relation between background, observation and analysis is,

meanwhile the principal idea is quite straight forward.

What elements will you take with you and use next time?

- A better general understanding of the subject, and further tips and tricks

regarding the coding of the coming exercises.

Where would you like to get some more input?

- More input around exercise 3 and 4.

Reflection note exercise 2

This exercise was difficult. The problem is not that I don’t understand the lectures, it is that without some introduction to xarrays and cartopy this is really hard to figure out. How was I supposed to know that I had to use argmax? How can I interpret it because I only saw some numbers and nothing useful.

Exercise 1 was well explained and nice to make. Everything was clear and it was a good example of how data assimilation works. Exercise 3 was a disaster. As I mentioned the script was not useful as we had to find some new functions. I struggled a lot and I think it is sad that I couldn’t do the exercise because of the programming because that is not what the course is about. It is also hard to ask questions through email and as an exchange student I also don’t have much time to go to the office to ask the questions. Exercise 4 was also fine. It’s a really hard and abstract concept so I tried to interpret the graphs but I’m looking forward to the lecture to get a more clear explanation from you.

So maybe next time I will start earlier so I have more time to maybe come by the office if I need help. Also for the fourth exercise I had to revise a bit my notes from the lecture and that helped so this is something I will take with me for the next exercise. I would like to have some more input in the script if we have to use some functions from xarray or cartopy because if you can’t find this you can’t make the exercise.

Reflection Note GEO4902 Exercise Number 2

In the assignment the focus was on data assimilation. This was also where I felt that I learned

the most. The theory feels quite abstract, so working on some data and also doing exercises in

only two dimensions, or one observation really helped with the understanding. I felt that once

one was able to connect the theory to the exercises they was more or less ok. The script for

the ex.1 and ex.3 was quite similar to those in ex.1, which shortened the time spent on coding.

The fourth exercise the script was new, but quite simple, so most of the time was spent on

trying to connect the different steps in the code to the theory. But, at the same time I felt that I

learned the most from this exercise together with ex.3, as one was able to remove the “noise”

from correlating variables and many different observations. Or in other words, use

simplifications and less observations, to easier see the different impacts of the steps in the

data assimilation.

For next time, I think I will mostly use the better understanding of the subject that I felt the

exercises gave me, together with the better understanding of unsing cartopy and netcdf. Since

these are still quite new, I feel that the learning curves are still quite steep, and it helps to be

forced to dig into them a bit and learn through struggling and errors.

For me I think just discussing the exercises will give me a better understanding of them in

general. Especially ex.4. I have tried to go through the theory and connect it to the exercise,

but I know that it for sure still is points and parts that I have not understood or thought about

yet.

Reflection Note

1) How you worked with the exercise.

At first I try to run all the cells to make sure they will run and I can get all the data. If I could not execute a cell in ex 1 I just tried to find a work around. Lots of time took me exploring data in ex. 3 (figuring out 500 hpa levels). I also needed to go through lecture notes to remind myself about data assimilation, B, H and R matrices. I spent some time reading about data assimilation from different sources.

2) What did you learn?

Hands-on approach on data assimilation, as well as MVGDF (multivariate gaussian distribution). So far I was only working with a normal distribution with one variable.

3) What was difficult?

Figuring out levels with different air pressure. Getting deeper to understand MVGDF, it should be obvious by drawing analogy from the 1d case, but I still got confused.

4) What went well?

Mostly 2 d plotting of temperatures.

5)What was surprising?

How small the difference between analysis and forecast was in case of single observation. I checked the values several times. I also expected argmax and max to return the same values. Probably I need to spend more time understanding the differences, or maybe I run it on different data.

6) What elements will you take with you and use next time?

Hands on experience on data assimilation process. Having more experience with differently organized forecasts data, MGDF.

7) Where would you like to get some more input?

Understanding levels in hydrostatic models, interpreting MGDF.