Assimilation of CYMS products by MET Norway

Per Dahlgren, Roger Randriamampianina

1 Introduction

MET Norway's (MET) contribution to the project was to assimilate SAR derived winds into the HARMONIE-AROME (H-A) Numerical Weather Prediction (NWP) system, [Bengtsson(2017)]. H-A is the NWP system used for operational weather predictions at MET. H-A is also used for creating reanalysis data sets, i.e. recreating past weather using a modern state of the art NWP and Data Assimilation (DA) system. Such a reanalysis data set is the Copernicus Arctic Regional ReAnalysis (CARRA) ([Yang(2020)]) in which the H-A system has been used over two domains (Figure 1) to reanalyze the atmospheric state between 1991 to 2022. The CARRA reanalysis system was constructed from an operational H-A system used at MET, called AROME-Arctic.

It was decided to use the CARRA system in the SAR wind assimilation described here. As MET's part of the project was quite small the work done can be seen as a *proof of concept*, i.e. we wanted to show that SAR winds can indeed be assimilated into the H-A system. Usually, the impact of an observation type is assessed by running the full NWP and DA system for at least one month with the observation type assimilated. For reference, a system with the full set of observation have to run again that without the observation type. Such a full assessment of an observation type is usually also preceded by a thorough process in which several aspects of quality control, thinning of data, tuning of observation errors are looked into. Such an impact study was not feasible under this project and instead we aimed at assimilating SAR winds in a case study.

2 The CARRA reanalysis system

The CARRA reanalysis system was built upon the HARMONIE version cy40. It uses a three dimensional variational (3D-Var) DA technique to analyze the atmospheric state. On top of the so called *conventional observations* (SYNOP, SHIP, DRIBU, TEMP, AIRCRAFT) a number of satellite data is also used from microwave sounders (AMSU-A, AMSU-B, MHS), infrared sounders (IASI), atmospheric motion vectors (winds derived from satellite data, AMV), Scatterometer winds and radio occultation bending angles from the Global Navigational Satellite System (GNSS).

CARRA is run with a three hour cycling, i.e. every third hour the atmospheric state is analyzed. From each analyzed state a 3-hour forecast model

Figure 1: Domains used in the CARRA reanalysis. The SAR wind assimilation was done using the eastern (NE) domain which is similar to the AROME-Arctic domain used for operational forecasting at MET Norway.

integration is made except at 00 and 12 UTC where a 30-hour integration is made.

3 Implementing SAR wind assimilation

MET received code updates used by Meteo-France for assimilating SAR winds in tropical cyclone cases [Duong(2021)]. Due to differences in model versions these updates had to be adapted and phased into the CARRA system.

Observations going into the H-A DA system have to go through the following steps:

- Adapt **BATOR** to read observation in its native format (usually BUFR, ASCII, NETCDF) and create an Observation Data Base (ODB) file. ODB is the format internally used by the H-A DA system.
- Adapt the **Screening** part of the H-A DA system. Here, observations are compared to the first-guess, or *background*, and quality control of the observation is performed. Here it is decided whether the observation is to be used in the next step or not.
- Adapt **Minim**, here the final analysis is calculated.

The SAR winds were provided in NETCDF format. In the H-A which CARRA is based on, the NETCDF interface was not very mature and a large part of the work was put into adapting BATOR to read the SAR winds and create an ODB.

In BATOR, the following parameters were read from the SAR wind NETCDF files:

- *sourceProduct*. From this string the satellite identifier was obtained.
- *measurementDate*. Date and time of SAR wind measurement
- *wind_speed*
- *wind_to_direction*
- *heterogeneity_mask*
- *mask_flag*
- *lat* and *lon*

BATOR converts the wind speed and direction to the u and v components of the wind which were written to the ODB and used in the DA. SAR winds were rejected if the *heterogeneity_mask* was not zero.

Each observation in the H-A DA system is assigned an *observation type (obstype)* and *observation code type (codetype)*. The *obstype* places the observation into a group of observations, such as e.g. AMV, Radiance, Scatterometer. If we use the *obstype=Scatterometer* example, then *Codetype* tells which type of Scatterometer product the observation is, e.g. ASCAT, ERS. The SAR winds were used with *obstype=Scatterometer*, i.e. the SAR winds are treated in the parts of the code that deals with Scatterometer observations but assigned with its own *codetype*. What separates this SAR wind product from other Scatterometer data is that it provides one u and v component and not the ambiguous counter parts. To make this work in the H-A DA system the handling of ambiguous wind components were disabled which disabled the use of other Scatterometer observations alongside the SAR winds. Due to time constraints there was no time to make sure that the Screening performed a proper first-guess check in which the first-guess departure, the distance between the observation and the NWP model background, is checked. The observation is rejected if the first-guess departure is too large. The SAR winds were thinned and were subjected to the blacklist.

Thinning is performed in the Screening and is applied to data with high spatial density. Such data often have horizontal error correlations which we do not know how to properly handle at this point. Instead the data is thinned to make sure each data point is far enough apart so that no horizontal error correlations will be present. Here the SAR winds were thinned to 50km.

The Screening also utilize a blacklist file for quality control purposes. In the Blacklist file one can specify on a detailed level which data to use and which data to reject. If e.g. a certain channel from a certain instrument is know to be bad for a certain date, then this can be specified in the blacklist file and the erroneous channel can be rejected for these dates. It is possible to provide the first guess departure to the blacklist. The SAR winds were rejected, or blacklisted, if the first guess departure was bigger than 10m/s. This type of first-guess check in the blacklist is less sophisticated than the proper first-guess check made inside the Screening (mentioned two sections above). The first-guess check in the Screening also takes into account the observation and background error. To do the first-guess check in the blacklist is a crude and quick remedy to the fact that no proper first-guess check was done.

In summary, the SAR winds were rejected in BATOR if the *heterogeneity_mask* was not zero. The Screening thinned the data to 50km and the blacklist rejected the SAR winds if the first-guess departure was bigger than 10m/s.

4 Case study

We selected a case with an intense mesoscale development that hits the north coast of Norway where several observations reported wind speeds over 25m/s. The storm developed over sea and well within the CARRA domain, see Figure 2. It was very useful to use the CARRA reanalysis system for this study because whatever case we select the CARRA reanalysis data would provide *warm start files*. This means that the first guess for the first analysis can be taken from the CARRA reanalysis data set, i.e. we do not need to start a couple of days earlier to spin up the system. The H-A DA system uses a variational bias correction technique (VARBC) to correct for biases in satellite radiances. VARBC coefficients usually need to be spun up for about two weeks before being useful. In this case, spun up VARBC coefficients could also be taken from the CARRA reanalysis, i.e. we could run this case study with the full observation data set. That being said, due to reasons explained in the previous section we could not use other Scatterometer data alongside the SAR winds.

The case study was started on 00UTC on 8 Dec 2016 (using the first guess coming from the CARRA reanalysis) and run with 3 hour cycling (an analysis being done every third hour) until 00UTC on 9 Dec 2016. At 00 and 12UTC a +30h forecast integration was done, at other times a 3h integration was done to provide a first guess for the next DA cycle. A baseline, or reference, was run which assimilate the full observing system (described in first paragraph of section 2) minus Scatterometers. Then the experiment was run which is identical to the reference except that SAR winds were assimilated when available; at 03, 06 and 15UTC. Figure 3 show the coverage of the SAR winds assimilated, the effect of the thinning is clearly visible and the data gaps are due to the rejection in BATOR of data where the *heterogeneity_mask* is not zero. The difference between the analysis and the first-guess for wind speed at 10m (analysis increments) for the experiment run will show the effect of the SAR winds on the analysis as there are very few other wind measurements in the area, (Figure 4). Increments are both positive and negative (red and blue) which is good, but they are also rather large, up to 9 m/s at 15UTC. Typical 10m wind speed

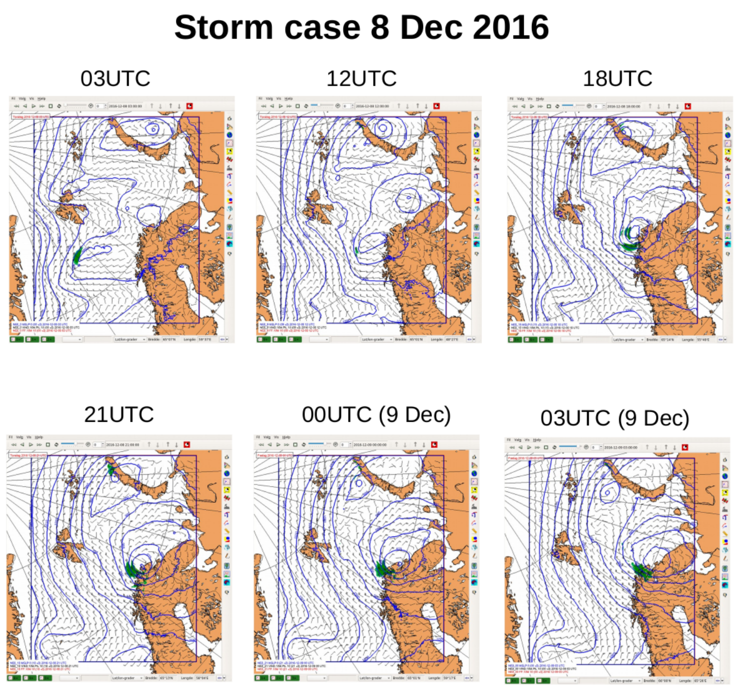

Figure 2: Storm case development on the 8 Dec 2016. Green color shade show where the wind speed is greater than 20 m/s. Blue solid lines show the mean sea level pressure in hPa units with 5hPa intervals. Wind arrows, or barbs, are in black. Maps shows 3h forecasts, i.e. the map labeled 03UTC is the 3h forecast from the 00 UTC analysis.

increments are ±3 m/s. In the DA process a cost function is minimized with an iterative procedure. The cost function has two terms; J_b which measure the distance of the analyzed model state $\mathbf{x}$ to the first-guess (or background) $\mathbf{x_b}$ scaled by the background error covariance matrix $\mathbf{B}$ and J_o which measure the distance of the observations $\mathbf{y}$ to the analyzed state scaled by the observation error matrix $\mathbf{R}$. H is an operator that projects the model state to the observation space. $\mathbf{x}$ is the state we want to determine in the minimization. Usually the minimization start with $\mathbf{x} = \mathbf{x_b}$ which means $J_b = 0$ and J_o is large. As the minimization progresses J_b will grow while J_o gets smaller.

$$J = (\mathbf{x} - \mathbf{x_b})^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x_b}) + (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x}))$$

Figure 5 shows the J_o values for the assimilation cycles where SAR winds were used from the reference and experiment runs. We can see that the J_o values are larger (red lines) when SAR winds are used but it minimizes smoothly in all three cases. At 03 and 06UTC it takes more iterations to reach a minimum when SAR winds are assimilated while at 15UTC it takes a few iterations less.

To investigate how the assimilation of SAR winds affected the CARRA system's representation of the storm a group of surface observations (Figure 6), along the coast of Norway were selected. These stations were in the area where the storm hits and reported high wind over 25 m/s. The stations reported every hour while the CARRA system produce an analysis every third hour. In order to compare the CARRA system to observations we constructed a time series that is a mix of analyses and forecasts, e.g. at 00UTC on 8 Dec 2016 the analyzed 10m wind speed was used while at 01 and 02UTC the 1 and 2h forecast from the 00 analysis was used. After 00UTC on 9 Dec 2016 we have only forecast values originating from the 00UTC analysis.

In Figure 7 such time series have been constructed for each of the eight selected stations. For most stations there is a clear spike with high wind speeds as the storm hits which is between 18UTC on 8 Dec 2016 to 03UTC on 9 Dec 2016. If we focus on how well the NWP system reproduces the high wind speeds, the impact of SAR wind assimilation is rather mixed. For station 01042 the reference (REF) and SAR wind experiment (EXP) produce the same wind speed but EXP has it bit too early. For 01086 EXP is better while for 01055 EXP produce slightly higher wind speeds but a bit too late. For 01083 EXP is worse than REF while for 01074 they are similar. For 01078 REF is better while for 01088 both REF and EXP produce the high wind spike too late, REF gives higher winds which is better. For 01043 REF and EXP gives similar wind speeds but EXP has it at the correct time.

5 Outlook

This short study shows that we can assimilate SAR winds and that the DA system can handle the information as the cost function minimize smoothly. A number of quality control problems in the Screening were not looked into due to time constraints. The small impact in the storm case is actually quite

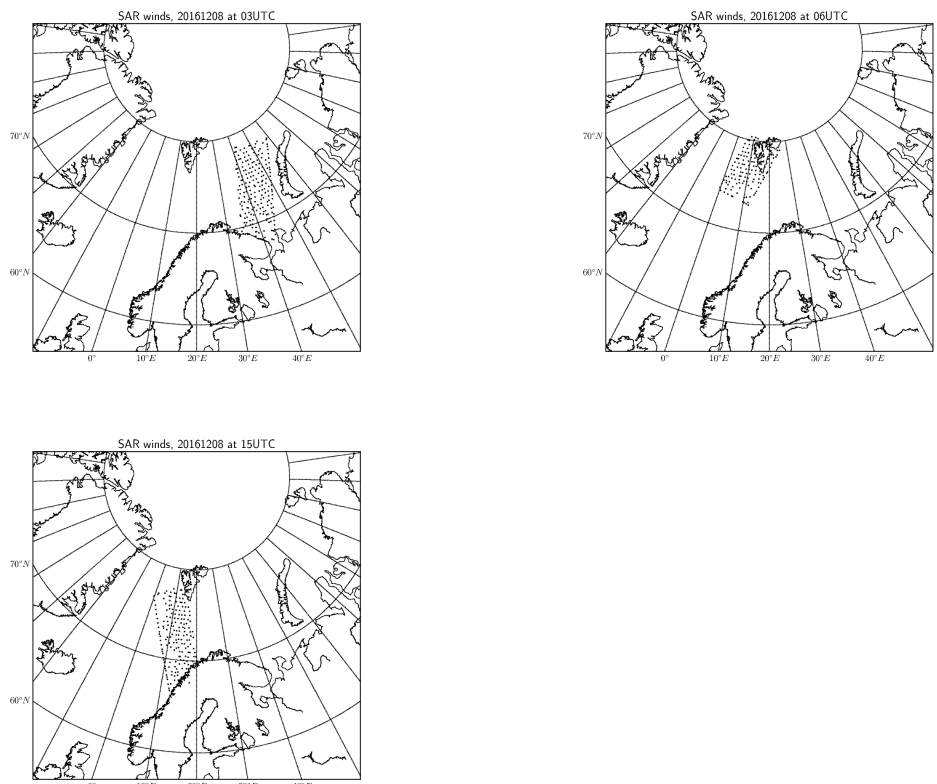

Figure 3: Coverage of SAR winds used in data assimilation.

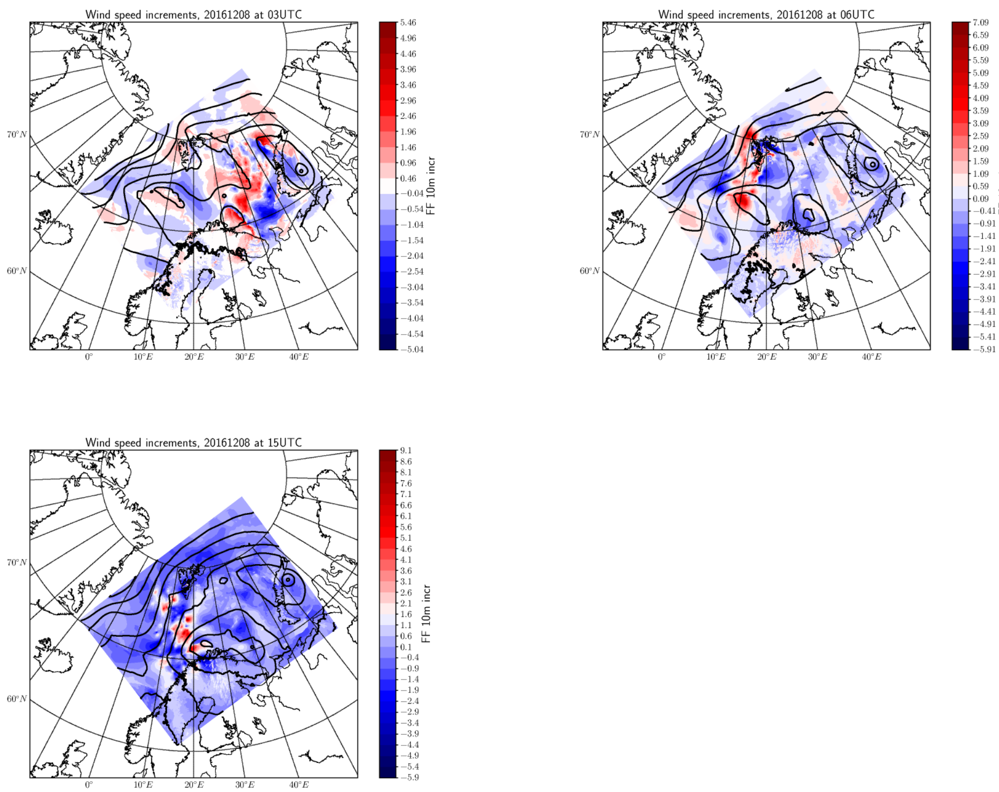

Figure 4: Wind speed at 10m increments, i.e. difference between analysis and first guess.

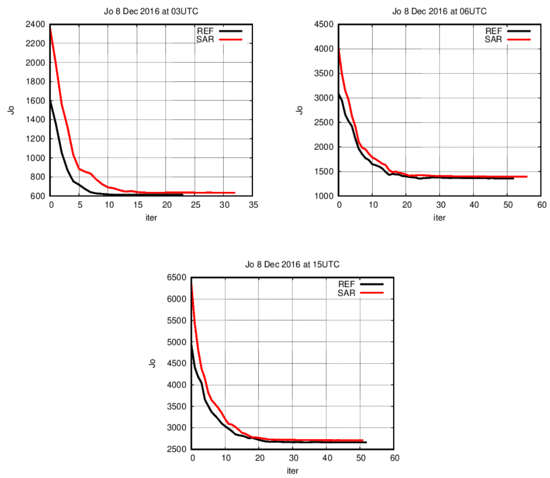

Figure 5: J_o cost function values for the three analysis cycles where SAR winds were used. Red lines are J_o values for SAR wind experiment and black is the reference.

normal even when established observation types are studied, especially when the reference performance is quite good which was the case here, i.e. the CARRA system performed well in representing this storm. With that in mind it is a good indicator that the SAR wind assimilation only gave minor impact. We outline below some points to be further considered:

Data selection. We received some feedback from the data providers that there is more to do when selecting data from the NETCDF file. E.g. that the *wind streak direction* might be better to use for wind direction in some cases.

Aggregating NETCDF files. If SAR data is available from different satellite passages withing the assimilation window then they would be in different NETCDF files. There is already a solution to this! In later HARMONIE-AROME versions this can be handled by the system in the BATOR script.

Quality control. There should be more time spent on developing a more mature handling of SAR winds especially in the Screening e.g. making sure the first guess check is done and tuned properly. Also, making sure that SAR winds can be used together with other scatterometer data such as ASCAT.

Use of ambiguous wind components. Overall, we think that since other scatterometer products are assimilated with the ambiguous wind components that DA methodology is more mature. It would therefore be better if the ambiguities were kept in the data provided to NWP. Then a SAR DA methodology with ambiguities can and should be developed.

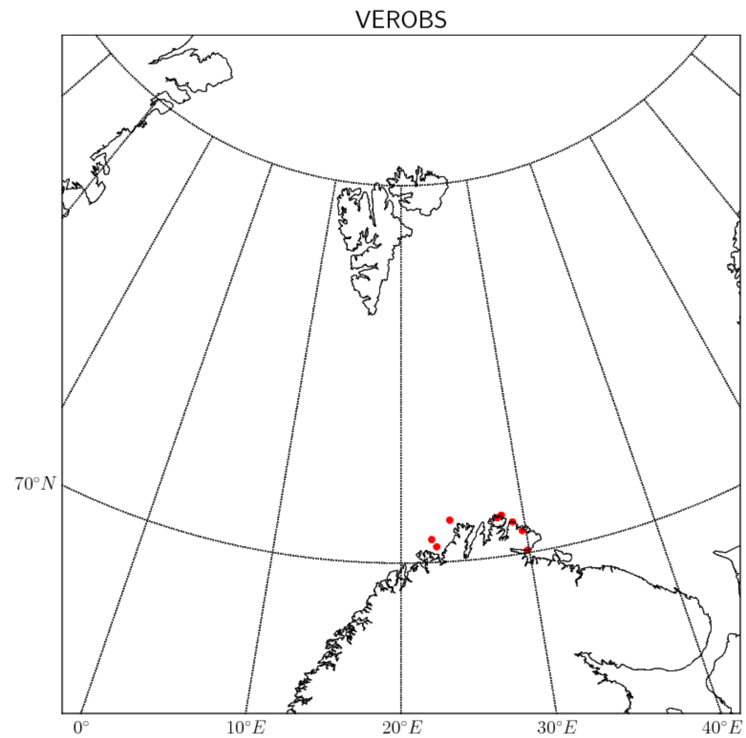

Figure 6: Observations used for assessing the impact of assimilating SAR winds the CARRA system.

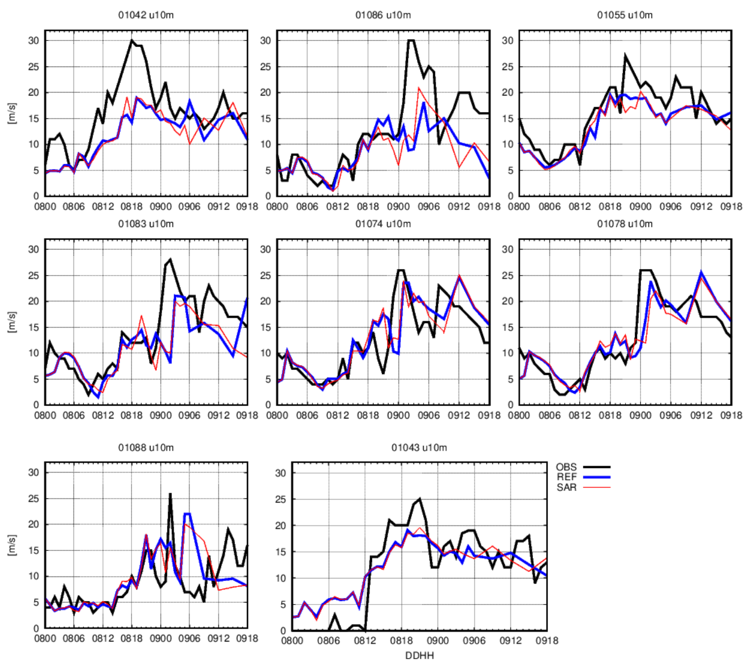

Figure 7: Assessment of impact from assimilating SAR winds. x axis show the date and time. Black lines are observed values, blue is the reference run and red is the experiment where SAR winds have been assimilated.

References

[Bengtsson(2017)] Bengtsson, L.; Andrae, U.; Aspelien, T.; Batrak, Y.; Calvo, J.; de Rooy, W.; Gleeson, E.; Hansen-Sass, B.; Homleid, M.; Hortal, M.; Ivarsson, K.; Lenderink, G.; Niemelä, S.; Nielsen, K. P.; Onvlee, J.; Rontu, L.; Samuelsson, P.; Muñoz, D. S.; Subias, A.; Tijm, S.; Toll, V.; Yang, X.; Køltzow, M. Ø. The HARMONIE–AROME Model Configuration in the ALADIN–HIRLAM NWP System. *Monthly Weather Review* **2017**, *145(5)*, 1919–1935.
https://doi.org/10.1175/MWR-D-16-0417.1

[Duong(2021)] Duong, Q.-P.; Langlade, S.; Payan, C.; Husson, R.; Mouche, A.; Malardel, S. C-Band SAR Winds for Tropical Cyclone Monitoring and Forecast in the South-West Indian Ocean. *Atmosphere* **2021**, 12, 576.
https://doi.org/10.3390/atmos12050576

[Yang(2020)] Yang X.; Schyberg H.; Palmason B.; Bojarova J.; Box J.; Pagh Nielsen K.; Amstrup B.; Peralta C.; Høyer J.; Nielsen Englyst P.; Homleid M. Køltzow M. A. Ø.; Randriamampianina R.; Dahlgren P.; Støylen E.; Valkonen T.; Thorsteinsson S.; Kornich H.; Lindskog M.; Mankoff K. C3S Arctic regional reanalysis - Full system documentation. **2020**. Available online:
https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-carra-single-levels?tab=doc