TOWARDS HOMOGENIZATION OF OPERA RADAR DATA

Report from RC LACE stay
Ljubljana, 25. 03. 2018 – 12. 04. 2018
Máté Mester, OMSZ, *mester.m@met.hu*
in cooperation with Benedikt Strajnar, Peter Smerkol, ARSO

I. INTRODUCTION

Following the development of the assimilation methodology at Meteo France, radar data assimilation (DA) has been a long-term plan for RC LACE countries for many years, yet the speed of progress in this field of numerical weather prediction (NWP) has been far from desirable. In order to change this state, an agreement was made this year to coordinate radar DA efforts more efficiently. Part of the agreement was to use radar datasets provided by OPERA (Operational Programme for the Exchange of Weather Radar Information), which would prove to be a common interface for exchanging radar observations within the scope of the interconnected model domains. Due to the release of new cycle versions for the AROME model and tendencies in the Meteo France-related NWP communities it was decided that the primary file format for radar DA should be HDF5, which has already been used by most countries as most radar factory-built softwares generate HDF files from the measurements. OPERA collects these nationally generated HDF files and after some post processing and quality control they are available in a theoretically standardized format and structure on the OPERA Data Hub. This standardized, common structure and content, especially in the case of radar metadata should be one of the main advantages of using OPERA HDF5 files. However, in practice standardization is not fully present, which generates a problem right before any quality control and assimilation task. Following the proposal for further work by Alena Trojakova, a 3-week stay was dedicated to overcome this issue by making a list of the contents of each file and creating a namelist-like database for radar metadata. This report summarizes the results so far and findings of the stay.

II. THE STANDARD STRUCTURE OF THE OPERA-ODIM HDF5 FILE

HDF is an abbreviation for Hierarchical Data Format, which is an information model that attempts to store many types of different data (in our case: individual scans, images, products, metadata) in a structure similar to directories, files and links on a hard-drive. Metadata is referred to as “attributes” while the actual data is referred to as “datasets”. These attributes are organized together in so-called “groups”. One can distinguish between 3 levels in each HDF file, and each level can be referred to as groups. Root group can be found at the top level, which includes datasets that act both as containers for actual data and subgroups for storing attributes. The lowest group can be found in each dataset in the form of data and quality subgroups that also have attributes on the lowest level. For our further work, we are only interested in attributes. Again, these are present at all 3 levels. This way, a standardized ODIM HDF5 file should have *“how”*, *“what”*, *“where”* attributes at root level and at the level of each dataset, plus *“how”* and *“what”* attached to each data and quality group within each dataset.
Although this structure is similar to one of the standards advised by OPERA (Fig. 1), it is not followed by many countries, which makes it difficult to create a tool to handle every radar and use the data for NWP purposes (as mentioned previously by Martin Ridal). This gives reason for the work mentioned in the next part.

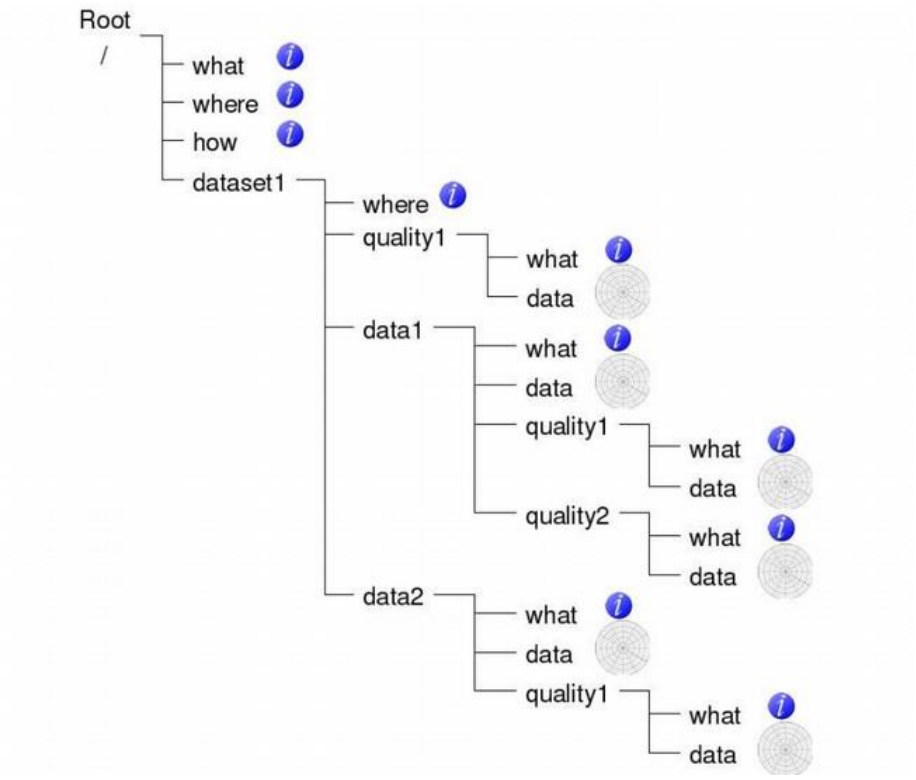

Figure 1: Data information model for a polar scan containing 2 parameters and associated quality metrics, advised by OPERA. What we work with is similar to this.

III. PROBLEM, 3-STEP WORKPLAN, THE JOB DONE SO FAR, METHODS

A 3-step approach was advised to create the repository of attributes by Peter Smerkol. The 3 points are as follows:

1. Before creating the actual “database”, it is necessary to check the structure of each file (each radarsite at a given timestamp). Even at this point all datasets, and inside datasets, all data and quality groups have to be taken into account and listed. Each dataset should represent a scan of a given elevation angle according to the OPERA standard, and each data group should correspond to a measured quantity. OPERA applies a 4-step quality control to every file, which means 4 quality groups are supposed to be present within one dataset. Nevertheless, all attributes have to be listed on all levels as these will be the elements of the database. If something is missing according to the official structure, or there are one or multiple extra groups or attributes, it should be recorded. Consistency is also an important matter to be cared for. In this case, it should be examined if one dataset actually represents one elevation angle and if there are multiple measured or calculated radar moments and corresponding data groups. It was mentioned before that there should be 4 quality groups, but to keep things clear this also needs to be checked. Consistency check also affects the content of attributes as those on the same level should have the same content in OPERA files (to clarify by using an example: if */dataset1/data1/what* has attributes such as *gain*, *offset*, *nodata*, etc. then */dataset1/data2/what* and */dataset2/data1/what* and so forth are supposed to have the same attributes too). This should be done to the following attributes:

   /datasetX/how

/datasetX/what
/datasetX/where
/datasetX/dataY/how
/datasetX/dataY/what
/datasetX/qualityY/how
/datasetX/qualityY/what

2. At this step, the union of attributes of the same levels should be created for all radarsites (or a chosen number of radars). In other words, we create a repository of the attributes at the same levels. For example, by reading */datasetX/how* from radar A, we get attributes such as *NEZ, NI*. By also reading */datasetX/how* from radar B, we get attributes *lowprf* and *radconstH*. The union of these will then contain *NEZ, NI, lowprf and radcontH*. This way we will have the maximum variety of attribute options present at a given level (coming back to our example, at */datasetX/how* level). After this step, attribute repositories will be available for "*how*", "*what*" and "*where*" attributes at all 3 levels (root, dataset, data/quality levels).
3. Finally, we have to check back the attributes in the case of each radar by using the repositories created previously. The main reason for this is to determine which attributes are present for which radar. This is important because in later stages, after determining exactly which metadata is required for the assimilation, it is desirable to know where to find them exactly. The details of this step and the output are still not determined.

IV. PROBLEMS AND FINDINGS THAT ARE STILL NOT HANDLED

During our work, it was realized that other problems also need some attention even at this early preprocessing stage. Such issues include as follows:

- In the case of some countries like Czech Republic, datasets with different polarisation (vertical or horizontal) belong to separate datasets in the HDF5 files. In other words, one full scan for one elevation angle can be obtained from two datasets. This may not affect all metadata, but should be cared for.
- In some cases, the same scan may be present in two consequent HDF5 files (reported by Benedikt Strajnar). The reason for this may be the concatenating algorithm applied by OPERA. One HDF5 file consists of measurements for a 15-minute period. When there is a scan which is on the border of two 15-minute periods (in other words, the scan is "interrupted" by the concatenating), the algorithm may put it into both files according to the time stamp of the measurement. This implies that both the measurement time and the HDF5 file production time should be taken into consideration when using the files.
- According to previous reports, missing data about radar constants posed a major problem because it was requested by Bator code up to cy40. In most cases, it is present as metadata. If not, it can be calculated from other metadata (OPERA suggests a function for this). However, in some cases there is not enough metadata included to do so. Referring back to discussions with remote sensing colleagues, this is caused by the fact that many countries send corrected reflectivity to OPERA, and correction is carried out by using the radar constant. So radar constant is not included as metadata, it is present as "part" of the actual dataset in these cases.

V. CONCLUSION

Further work needs to be carried out in terms of finishing our 3-step workplan. It is also desirable to ensure that the repository of metadata is created automatically, and that it updates itself regularly to keep up with the changes in the radar system (for example a new radar is

added to the OPERA database or there is a change in the measurement protocol).

At its current state, the script only reads HDF files. In a later stage, it might be possible to rearrange the structure of files based on the information of the metadata repository. This way HDF files with a standardized structure would be available for assimilation, and preprocessing would have a clear output for use apart from the metadata database.

Finally, I would like to thank Benedikt Strajnar, Anja Fettich and Peter Smerkol for their help and support. The stay was financed by the LACE DA project.

REFERENCES:

Trojáková A, Strajnar B, Smerkol P. 2018. Radar data pre-processing. Report from RC LACE stay in Ljubljana, CHMI, ARSO
http://www.rclace.eu/File/Data_Assimilation/2018/rep_Trojakova_RadarPreProc_Ljubljana_201803.pdf

Michelson D, Lewandowski R, Szewczykowski M, Beekhuis H, Haase G. 2014. EUMETNET OPERA weather radar information model for implementation with the HDF5 file format. Version 2.2. SMHI, IMGW-PIB, KNMI
http://eumetnet.eu/wp-content/uploads/2017/01/OPERA_hdf_description_2014.pdf

Ridal M, Dahlbom M. 2017. Assimilation of multinational radar reflectivity data in a mesoscale model: a proof of concept. Journal of the American Meteorological Society, Boston, MA.
https://journals.ametsoc.org/doi/full/10.1175/JAMC-D-16-0247.1