Documentation of the Data Structure

A Documentation from Jessica Ahring
j.ahring@fz-juelich.de

Forschungszentrum Jülich
JSC
ESDE
DeepRain
Jülich
February 27, 2020

This is a documentation about the data structure of the files we get from the DWD, how this changed through the preprocessing steps and what data structure we import to rasdaman at the end.

Beginning: DWD

filename: cdeYYYYMMDD.FF.mEE.grb2
location: /mnt/nfs/grb/YYYY/MM/DD

This data files include the following data

- variables (temperature, relative humidity, ...)
- time
  - values: actual times (8 per file)
  - refTime = midnight of day of model run
- longitude
- latitude
- modellevel (for some variables)

The name holds the following data:

- date (of the model runs)
- forecast hour
- ensemble member

First Preprocessing Step: '01-prepro.py'

In this script the data files where split per variable, such that every remaining data file just holds data for one variable. The files are stored in different directories. Also the files are converted to netCDF.

filename: preproc-cdeYYYYMMDD.FF.mEE.nc
location: /mnt/rasdaman/DeepRain/playground/preproc/VAR/

This data files include the following data

- values for the one variable (temperature, relative humidity, ...)
- time
  - values: actual times (8 per file)

  - refTime = midnight of day of model run
- longitude
- latitude
- modellevel (for some variables)

The name holds the following data:

- date (of the model runs)
- forecast hour
- ensemble member

Second Preprocessing Step: '02-split.py'

Here the data files where split per time value. So every file gets split into eight different files and the RefTime is adjusted such that processing in the next step is easier (and makes sense).

filename: time:YYYYMMDD-HH.FF.mEE.nc
location: /mnt/rasdaman/DeepRain/playground/preproc/VAR/

This data files include the following data

- values for the one variable (temperature, relative humidity, ...)
- time
  - value: actual time
  - refTime = start of that model run (start time)
- longitude
- latitude
- modellevel (for some variables)

The name holds the following data:

- date (of the model run)
- hour (of the model run)
- forecast hour
- ensemble member

Third "'Preprocessing"' Step: '03-import.py'

The data files corresponding to one model run where merged together.

filename: processed:YYYYMMDDHH.mEE.nc
location: /mnt/rasdaman/DeepRain/playground/preproc/VAR/

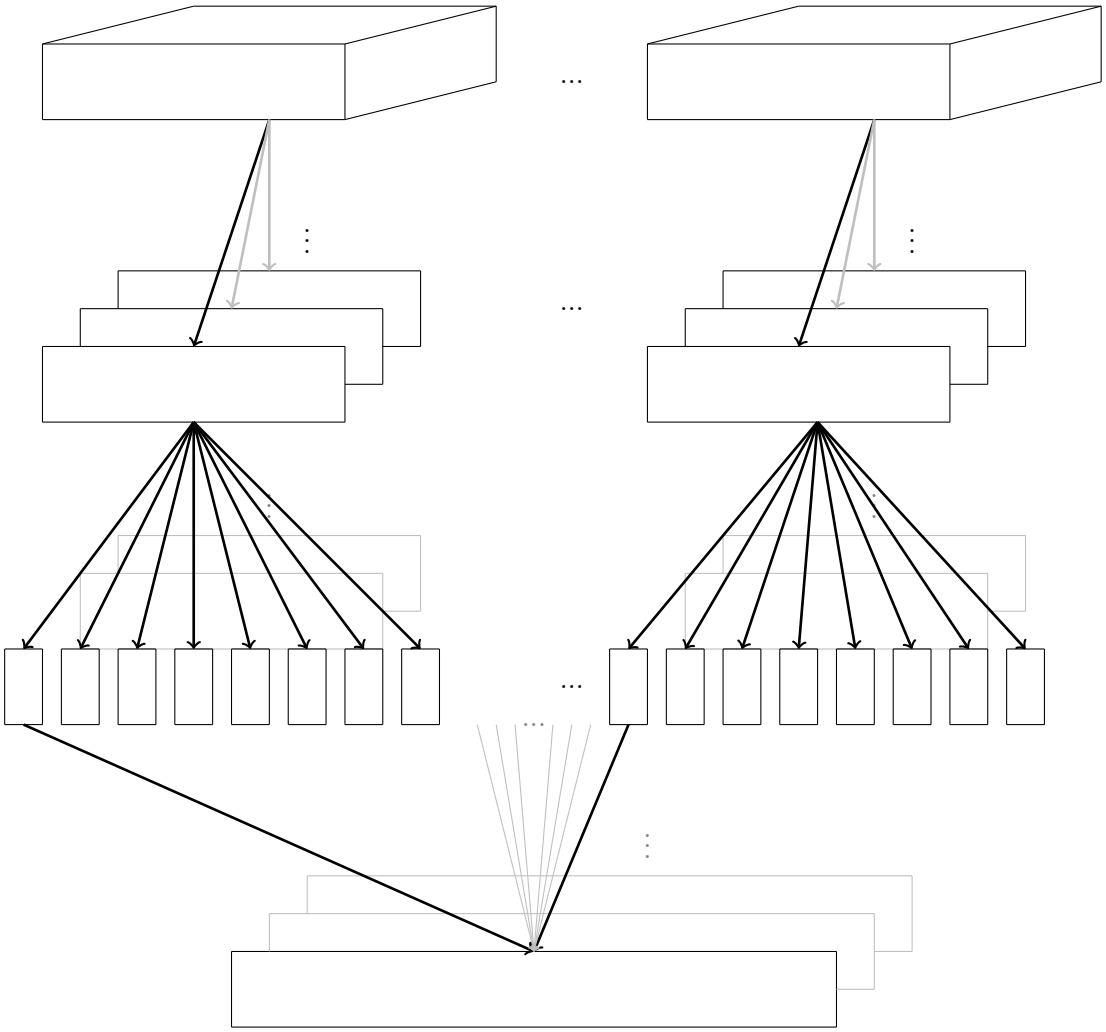
This data files include the following data

- values for the one variable (temperature, relative humidity, ...)
- time
  - value: actual times (25 (00-24))
  - refTime = start of that model run (start time)
- longitude
- latitude
- modellevel (for some variables)

The name holds the following data:

- date (of the model run)
- hour (of the model run)
- ensemble member

The following picture should illustrate the splitting and merging:



In the first line we see the data coming from DWD and containing all information. By using "01-prepro.py" we split every datafile into its variables (relative humidity, temperature, ...). This files can be seen in the second line. After that "02-split.py" splits this datafiles into its 8 timestamps (actual times) that correspond to the different model runs stating at one day. The last script "03-import.py" merges the corresponding timestemps together, such that we get 20 files per model run, where each file contains all forecast hours (00-24). This are 20 files, since we have 20 ensemble member that are not possible to merge in one file.

Since some of the values change through the preprocessing steps the following illustrate what happens. "01-prepro.py" only splits the data files in its variables and the values of the variables, the longitude, latitude and modellevel stay untouched, so this is not illustrated to make it more simple.