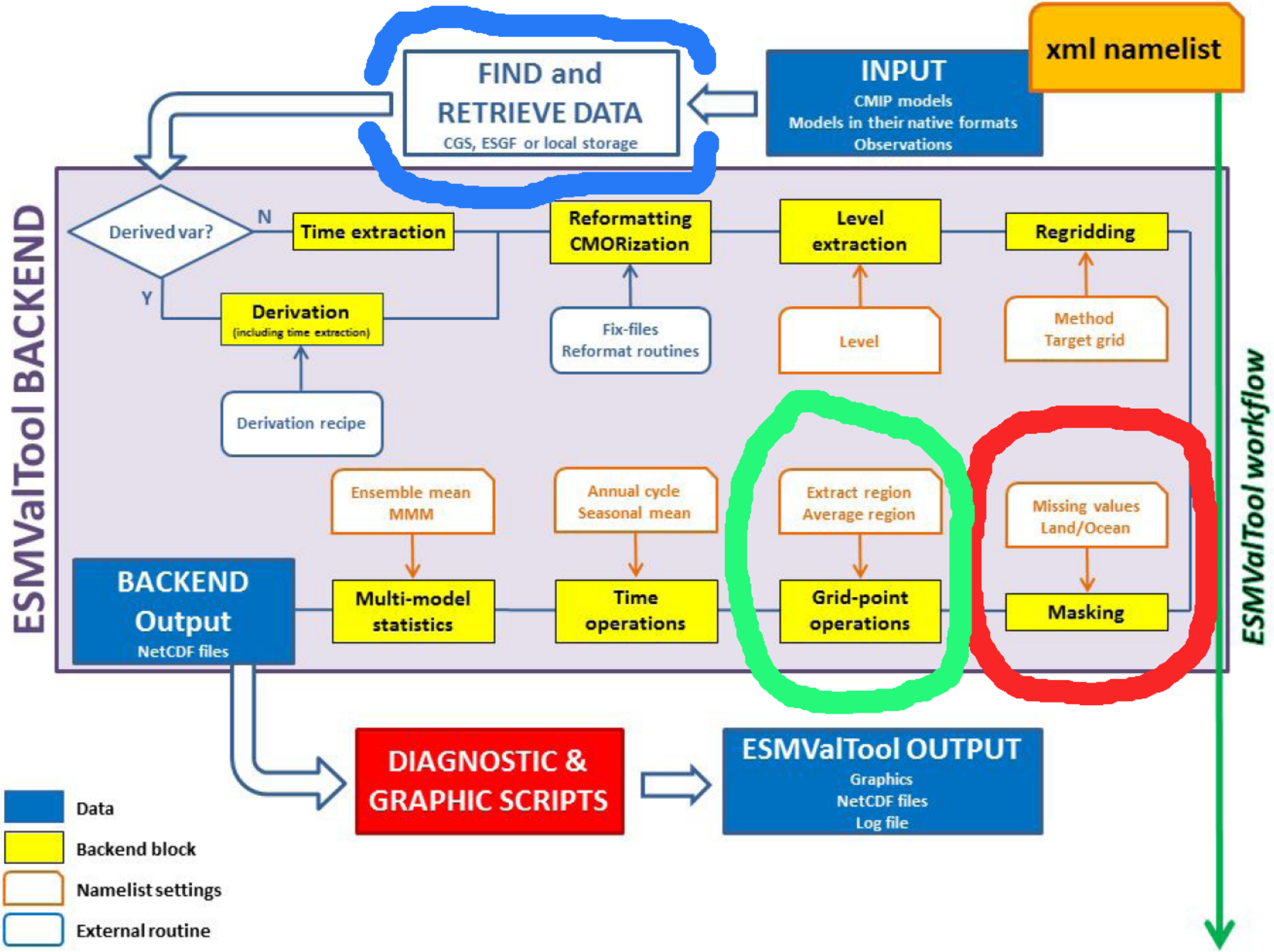


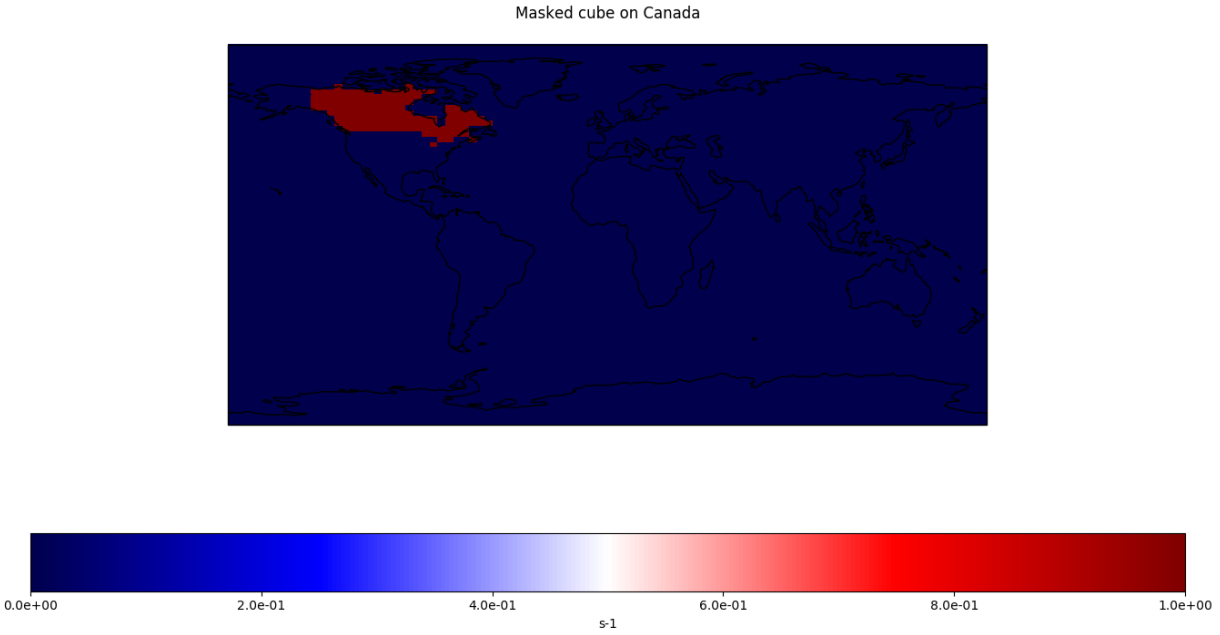
Figure 2. Schematic view of the revised ESMValTool backend.

Look up/download data

- synda_wrapper/get_data_synda.py
  https://github.com/valeriupredoi/synda_wrapper/blob/master/get_data_synda.py
- Looks for data specified either in a parameter file or as command line arguments – searches in a multitude of ESGF; tells you if a specific file already exists locally or if not downloads it;
- Very straightforward interface, with minimum user inputs;
- Comes with a checker script that checks the progress of your downloads.

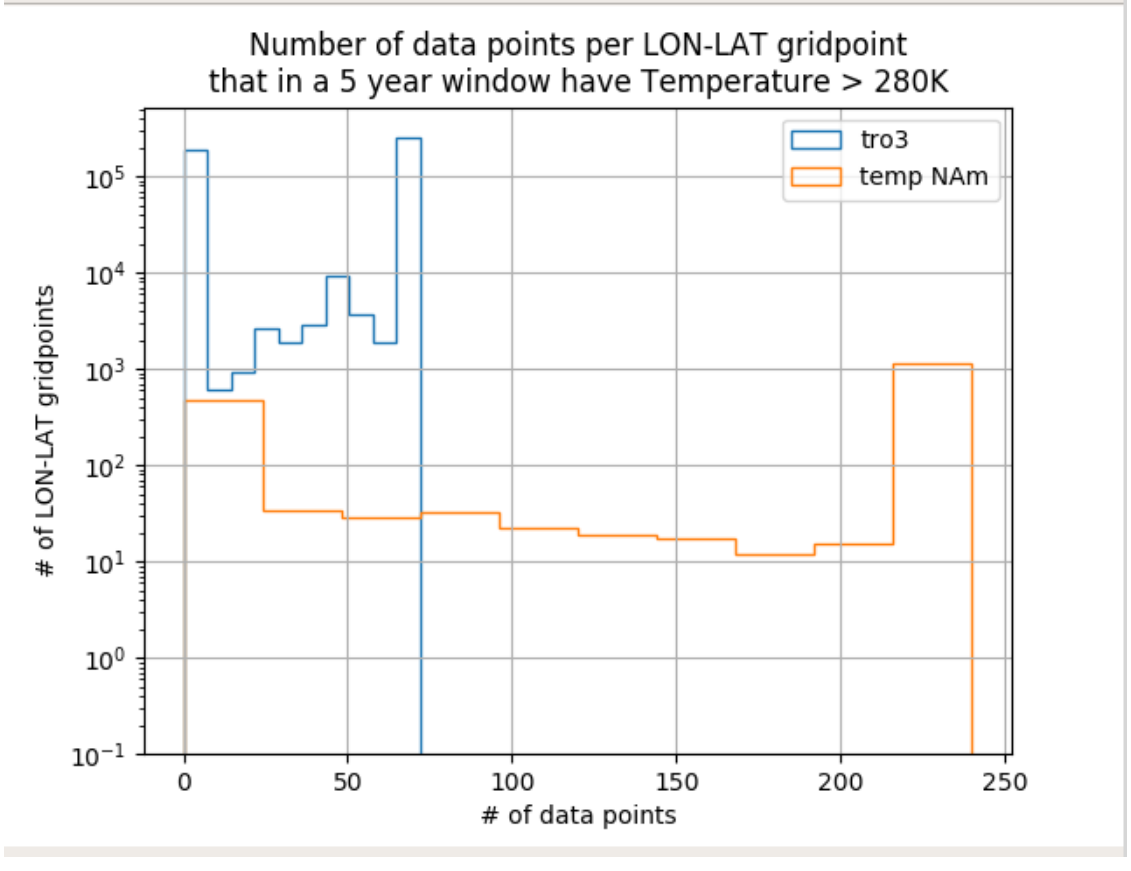
Masking + Region Extract

- https://github.com/ESMValGroup/ESMValTool/blob/REFACTORING_mask/backend/masking/mask_suite.py
- Simple LON-LAT box extraction; simple thresholding on a value masking
- Multi-point region extraction e.g. land/ocean (can be applied to other /backend/ tasks *ie* Region Extraction):



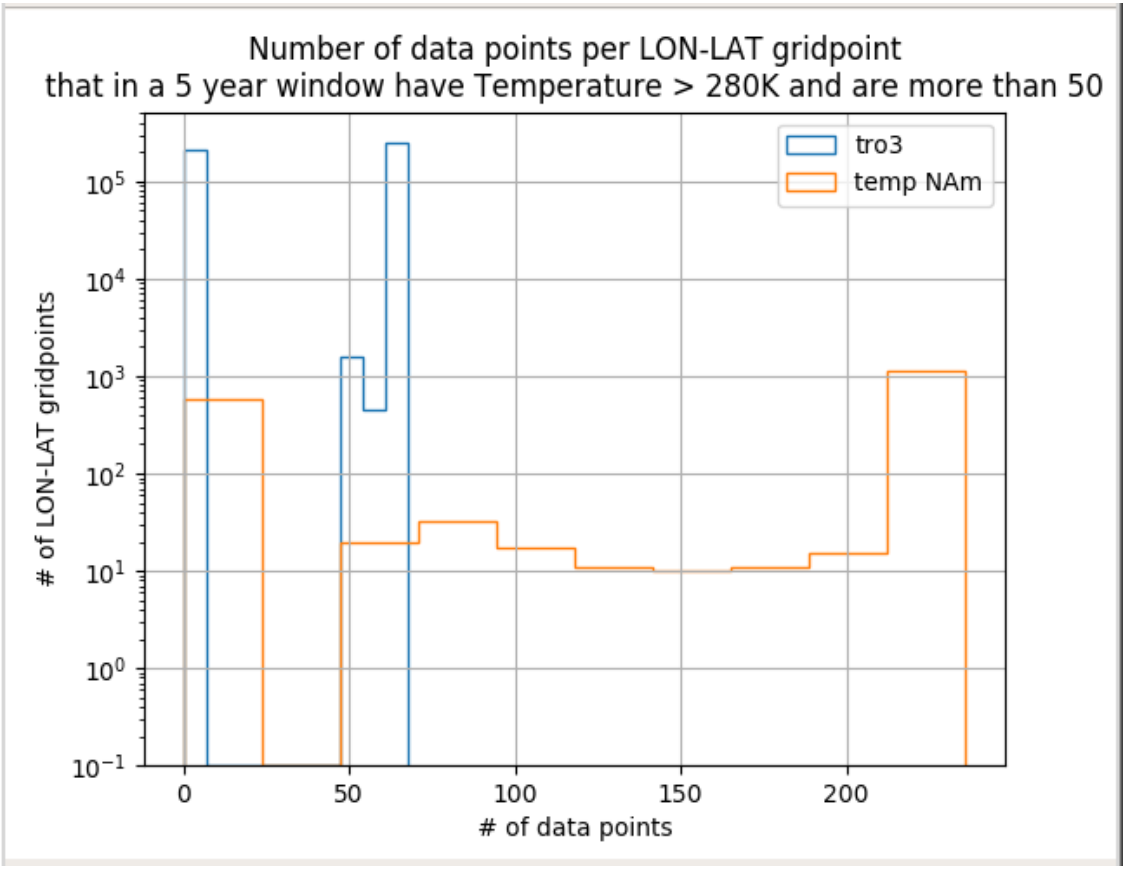
Masking – Missing Data

- Dealing with missing data: say we have two model data files that (after regridding to a common grid) each have missing data: first, we chunk the temporal data in **windows** (or blocks) (optionally apply a threshold on the variable)

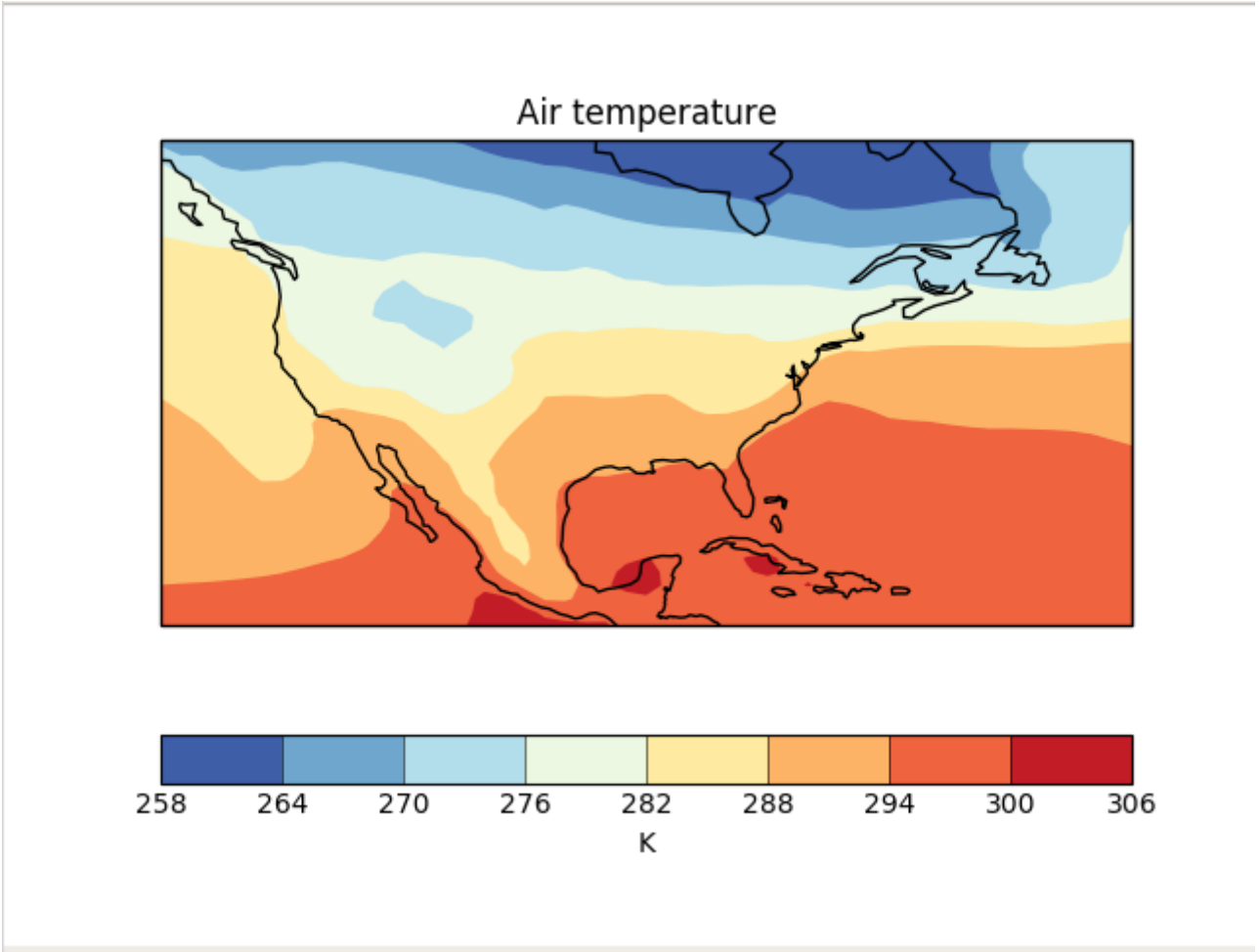


Masking – Missing Data

- Dealing with missing data: afterwards, we compute an optimal global thresold for the population of missing values in each window - ``global'' because we'll apply it in each cell, and build a common missing data mask

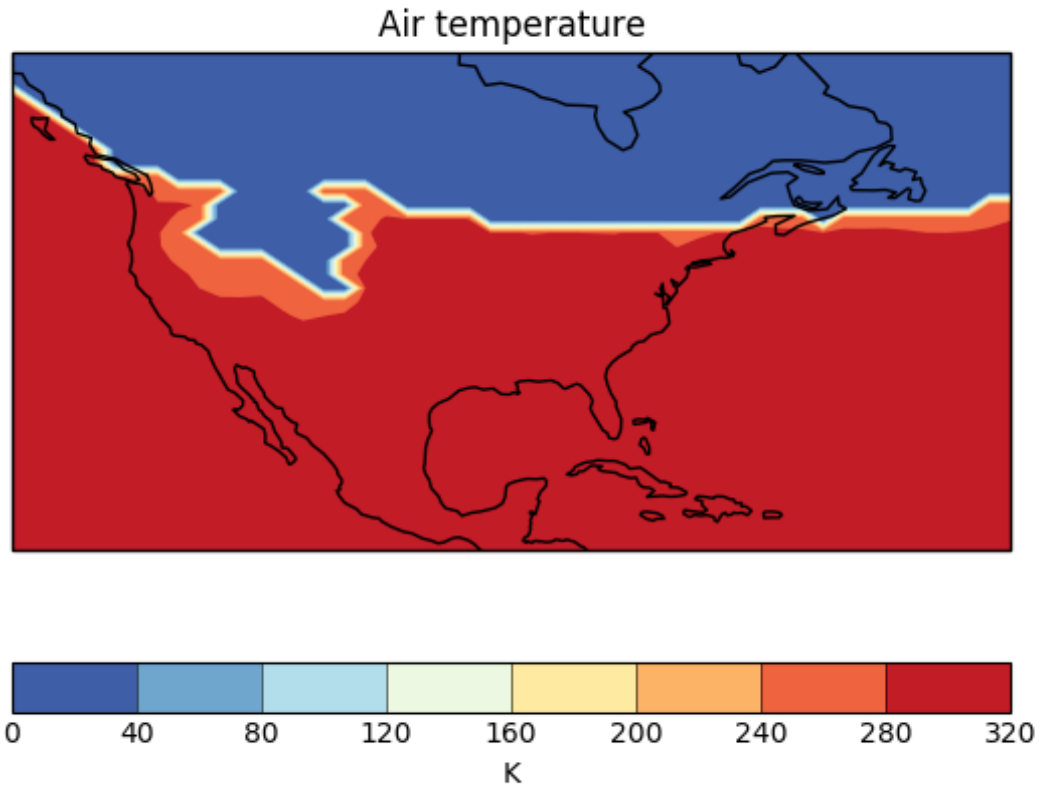


Masking – Missing Data



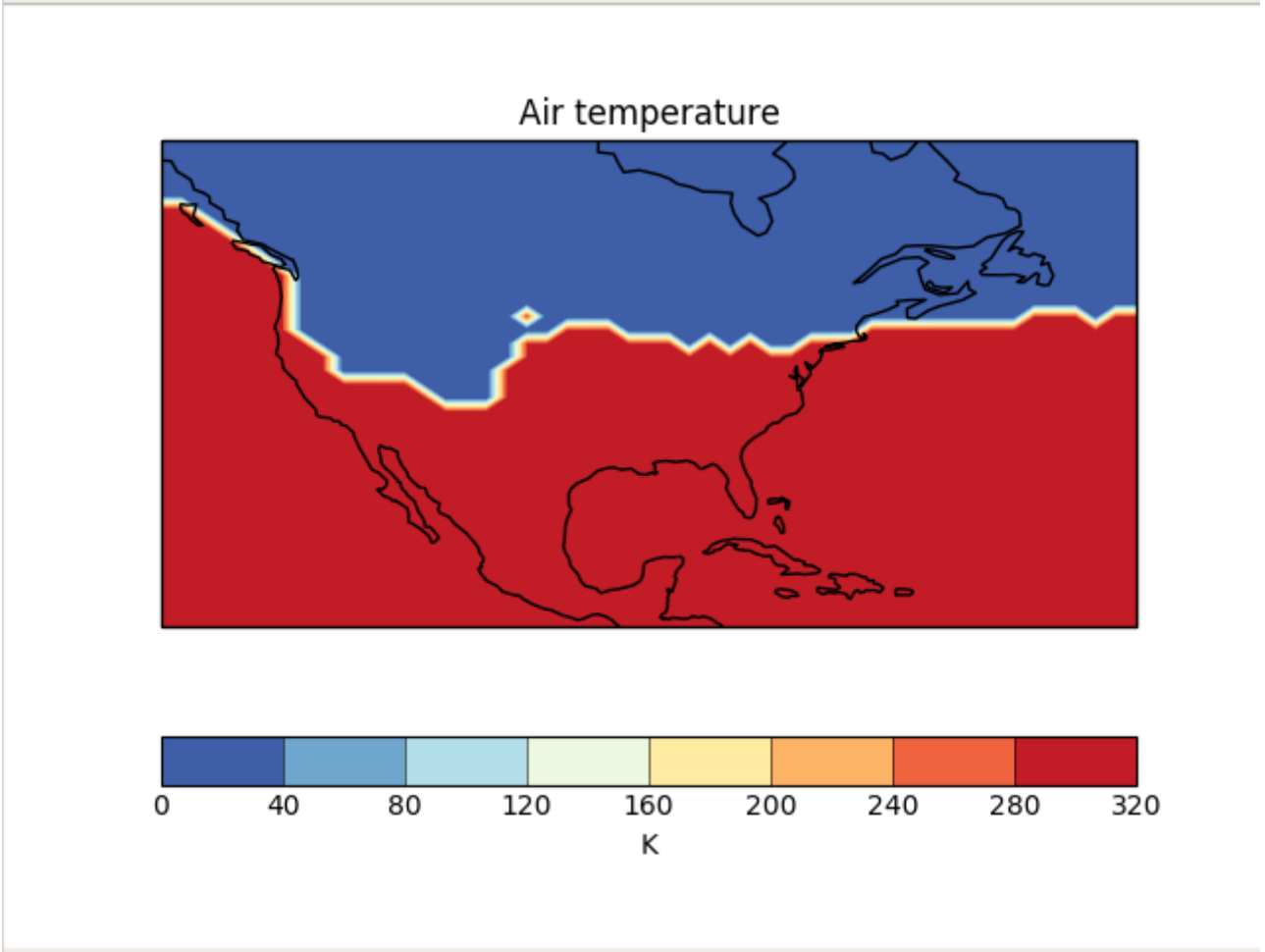
Unmasked – original data

Masking – Missing Data



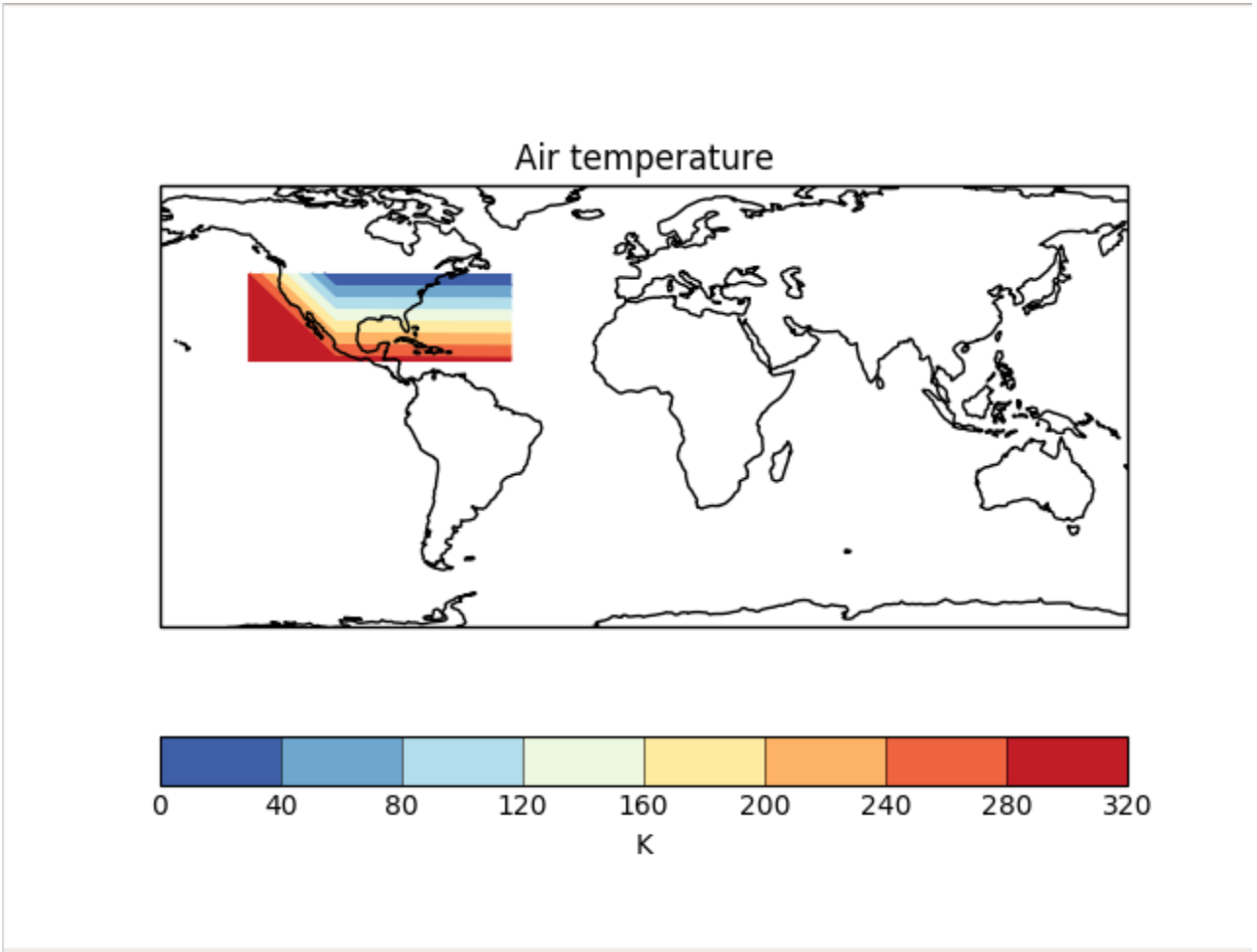
Masked: 5-years window, threshold T>280K, more than 50 time-points per window and LON-LAT gridpoint

Masking – Missing Data



Masked: 5-years window, threshold T>280K, more than 200 time-points per window and LON-LAT gridpoint

Masking – Missing Data



Masked: 5-years window, threshold T>280K, more than 50 time-points per window and LON-LAT gridpoint, REGRIDDED to a global grid