Climatology Calculation Support in the GeoCAT Ecosystem

Heather R. Craker[1,2] and Alea Kootz[2]
[1] Purdue University, [2] National Center for Atmospheric Research

NCL to GeoCAT

Two main groups of functionality in the NCAR Command Language (NCL) need to be added to the GeoCAT Python ecosystem.

Visualization
- Example gallery (GeoCAT-examples)
- Wrapper functions for plotting (GeoCAT-viz)

Computation (GeoCAT-comp)
- Handling data from different atmospheric models
- Interpolating/transforming data
- Climatology calculations (means, anomalies)

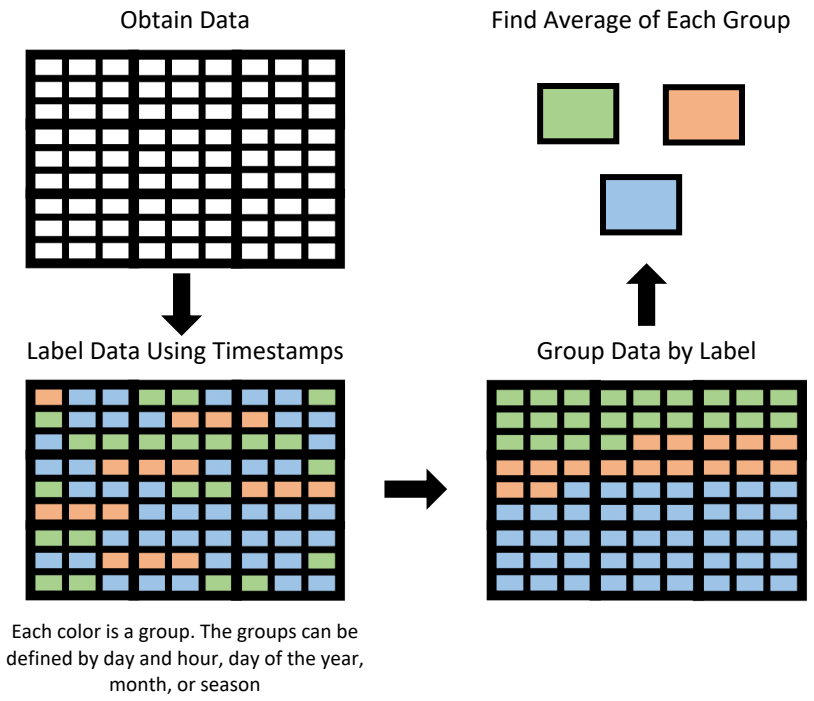
This project focuses on adding computational support for doing climatological calculations, specifically climate averages.

Climatological Calculations

What is a Climatological Average?
- Climatological averages are long term averages aggregated over multiple years
- Data must be grouped by a time period before the average is taken (i.e. by month)
  - Grouping must disregard years
  - All data in any January (regardless of year) is a part of one group

Grouping by datetimes
- Grouping sequential datetimes can be easily done by slicing an xarray.Dataset
- It is harder to group data that isn't sequential
  - Months aren't all the same length
  - Leap years add variability
  - Non-standard calendars add complexity
- xarray.DateTimeAccessors and the groupby() function make this easier

Processing Data



Fundamental Packages

NumPy

xarray

Challenges

- Supporting non-standard calendars for different weather model outputs
  - Data where there are never leap years
  - Data recorded in the Julian calendar
  - Data where months are all 30 days (including February)
- Doing a weighted averages when going from monthly averages to seasonal averages

`climatology_average()`

Calculates long term hourly, daily, monthly, or seasonal averages across all years in the data

Inputs:
- `dset`: The data as an xarray.DataArray or xarray.Dataset
- `freq`: A string representing the frequency of the calculated climatological averages
  - 'hour', 'day', 'month', 'season'
- `time_dim`: The name of the time dimension
  - Optional, will be inferred if not given

Output:
- Climatological averages for each period

Future Work

- Address user feedback now that geocat-comp 2021.7.1 is released
- Add functionality to change how time bounds are handled
- Implement NCL climate anomaly functions along with other computational routines
- Create any new tools that users need

Acknowledgements
Thank you to Alea Kootz and Orhan Eroglu for their mentorship and help with learning how to design software. Special thanks to Anderson Banihirwe, Deepak Cherian, and Max Grover for their help with development and user feedback.

GeoCAT Website
geocat.ucar.edu

GeoCAT-comp Repository