Introduction to aggfly

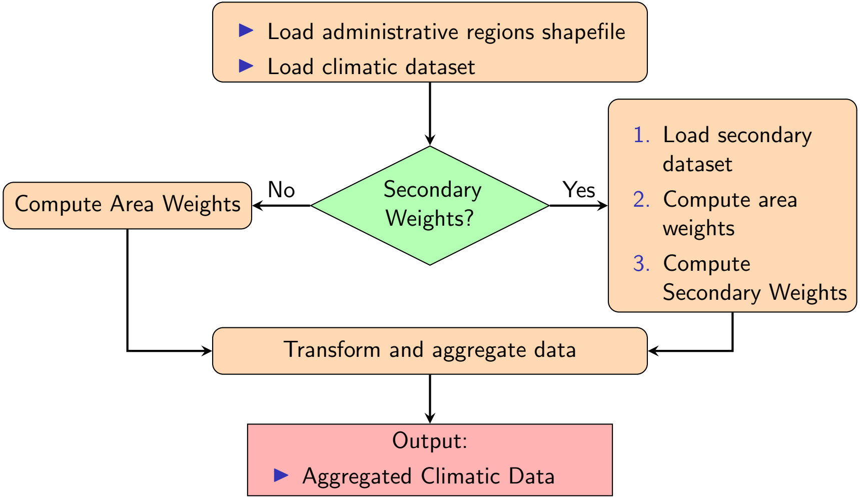
Your Name

June 12, 2024

Overview of the aggfly package

- Python package designed for the **temporal** and **spatial** aggregation of gridded climate data
  - **Temporal**: Nonlinear aggregations of weather data across time, e.g., degree days, daily polynomials
  - **Spatial**: Weighted aggregation of gridded data to administrative boundaries accounting for local exposure, e.g., humans, crops
- Efficiently **automates compute and memory-intensive geospatial operations**, a common barrier to entry for new researchers
- Different functional forms allow for **wide variety of weather specifications**
  - Currently: average, sum, min, max, degree days, bins, polynomials
- Useful for studying the effects of weather and climate on human health, agriculture, economic growth, etc.

Workflow for Aggregating Climatic Data

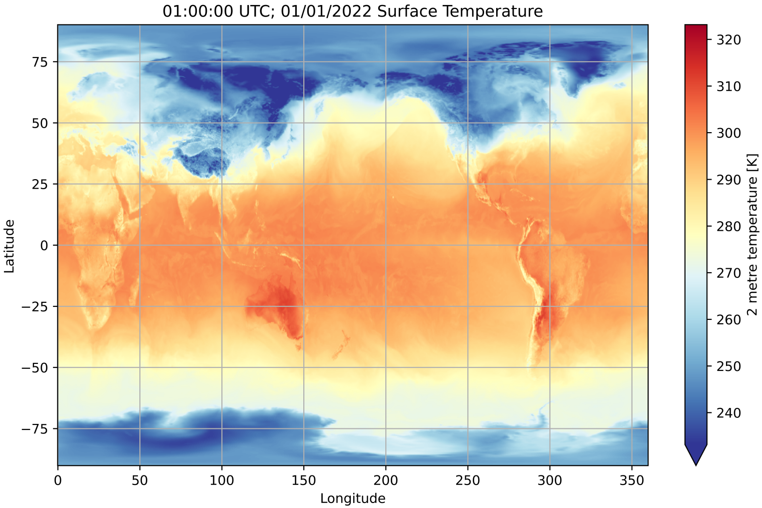


Input Datasets

1. **Climatic Dataset**: Gridded dataset with weather observations for a given longitude, latitude, and point in time
2. **Shapefile**: Geospatial data representing boundaries of the target administrative regions
3. **(Optional) Secondary Weights Dataset**: Raster dataset providing high resolution exposure, e.g., population or crop coverage, used to weight climate grid cells

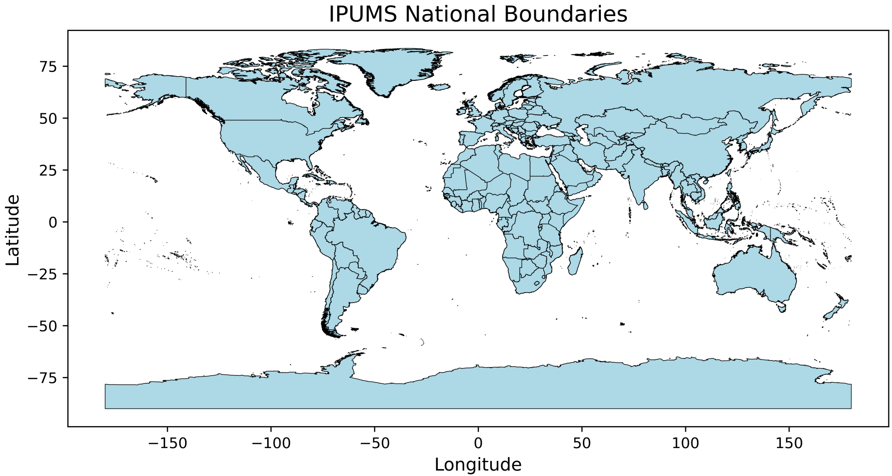
Input Dataset 1: Climatic dataset

- Global hourly temperatures in 2022 from ERA5

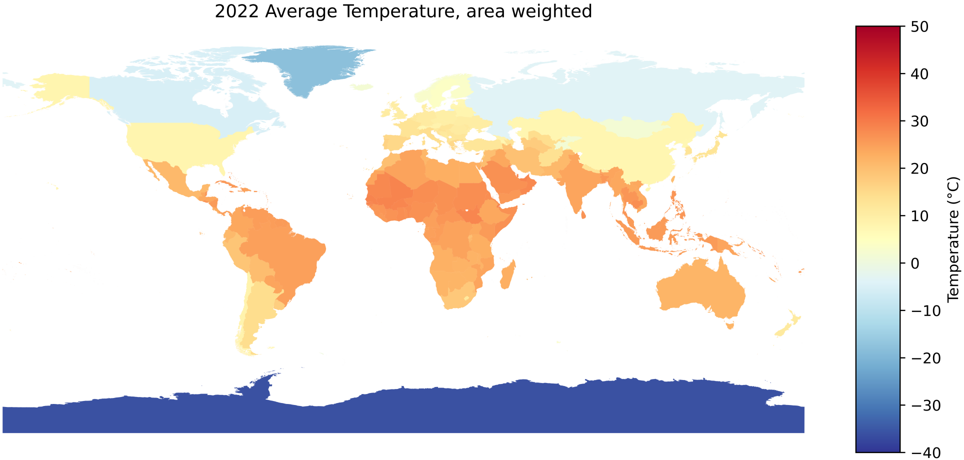


Administrative Regions Shapefile

- IPUMS World national boundaries



Output usage:



Data: ERA5 from the Copernicus Climate Data Store

Types of Weights

- **Area Weights**:
  - Default weights based on the area of an administrative region that overlaps each grid cell.
  - Reduces measurement error for climate datasets with large grid cells or regions with irregular borders.
- **Secondary Weights**:
  - Optional weights used in the spatial aggregation of climate data to target exposure based on a relevant variable, e.g., population or crop cover
  - Useful for studies focusing on human health, agricultural productivity, etc., where high resolution data on exposure are available
  - Population weights example: cells with higher population contribute more to the computation of the average U.S. temperature.

Current Output Variables:

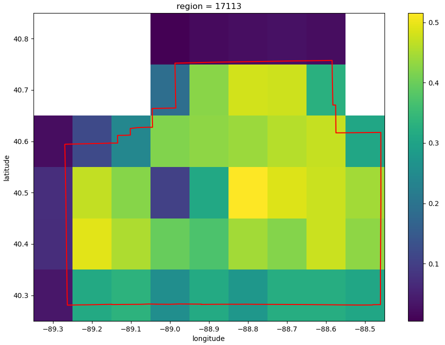
With **aggfly**, you can construct and combine various aggregated climatic variables, such as:

- **Average**: eg., mean temperature over specified time periods
- **Sum**: eg., sum of precipitation over specified time periods
- **Minimum**: eg., daily minimum temperatures
- **Maximum**: eg., daily maximum temperatures
- **Polynomials**: e.g., the sum of the polynomials of daily temperature up to the n^{th} degree
- **Degree Days**: e.g., temperature degree days - the sum of daily temperatures exceeding a base threshold, useful for agricultural studies
- **Bins**: e.g., temperature bins - the number of days in which the average temperature was in each 5 degrees Celsius interval in each year

Example: Aggregating crop-weighted climate data to the county level

Step 1. Assign crop-area weights to each grid cell

1. Average cropland mask to climate data resolution to get cropped share c_i for each grid cell i
2. Calculate overlapping area shares for each grid cell i and target region r, a_{ir}
3. Multiply cell-level cropped share with overlapping area $w_{ir} = c_i \times a_{ir}$

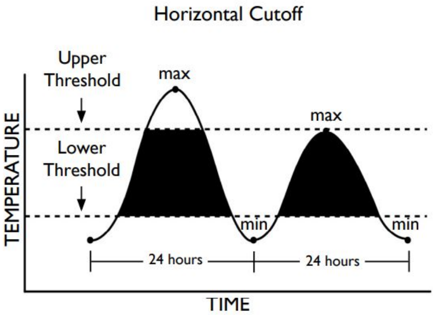


Example: Aggregating crop-weighted climate data to the county level

Step 1. Assign crop-area weights to each grid cell

Step 2. Calculate daily grid cell-level nonlinear weather terms

- Temperature: degree days, time exposed to each 1C bin in each day
- Precipitation: Quadratic function of daily total rainfall



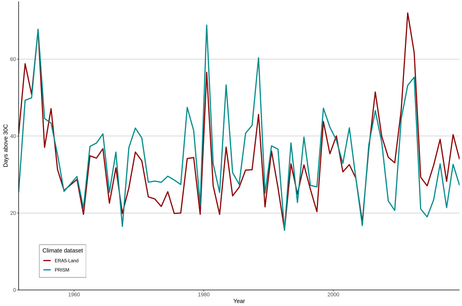
Example: Aggregating crop-weighted climate data to the county level

Step 1. Assign crop-area weights to each grid cell

Step 2. Calculate daily grid cell-level nonlinear weather terms

Step 3. Sum daily grid cell-level weather observations over the growing season

- Uniform March-August growing season definition



Example: Aggregating crop-weighted climate data to the county level

Step 1. Assign crop-area weights to each grid cell

Step 2. Calculate daily grid cell-level nonlinear weather terms

Step 3. Sum daily grid cell-level weather observations over the growing season

Step 4. Calculate crop-area weighted average of weather observations for each county

- $h_{rt} = \frac{1}{N} \sum_{i=1}^{N} w_{ir} h_{it}$