# Abstract

Daily precipitation data is critical for monitoring and responding to climate hazards like floods and droughts. Daily precipitation data could also help agriculture businesses better manage and utilize water resources. Climate Hazards Center (CHC) is a global leader in generating precipitation dataset based on satellites and on ground stations. The Climate Hazards Center InfraRed Precipitation with Stations (CHIRPS) made by CHC is a gridded dataset that covers almost the entire globe at spatial resolution 5km x 5km. CHIRPS is widely used for monitoring climate hazards and understanding long term precipitation patterns. The project goal is to compare CHIRPS 3.0 (newest version of CHIRPS) to other independent precipitation datasets and quantify the daily performance of CHIRPS dataset with statistical methods relative to other datasets.

# Background

**Our Sponsor** - Shrad Shukla at [Climate Hazard Center (CHC)](https://www.chc.ucsb.edu/)

* The Climate Hazards Center is an alliance of multidisciplinary scientists and food security analysts utilizing climate and crop models, satellite-based earth observations, and socioeconomic data sets to predict and monitor droughts and food shortages among the world's most vulnerable populations. Through partnerships with USAID, USGS, and FEWS NET, the CHC provides early warning to save lives and secure livelihoods.

**Gridded dataset**

* a spatial data format that consists of a matrix of cells, or pixels, organized into rows and columns where each cell contains a value for each grid point across a two dimensional surface.
* In our case, data is formatted with longitude and latitude coordinates with time in date and precipitation in millimeters.

[**CHIRPS 2.0 and 3.0**](https://www.chc.ucsb.edu/data/chirps)

* Quasi-globe, daily, high resolution (5 x 5 km) precipitation data
* Used for early drought warnings and environmental monitoring
* Constructed using satellite images and ground station data

**PRISM**

* Similar to CHIRPS but only covers the US and has higher resolution (4 x 4 km)

**Xarray**

* Python package for dealing with gridded dataset
* Specialized for geographic data (gridded data with coordinates and/or dates)
* Used in project for slicing, interpolating, and regridding the gridded datasets for comparison and for deriving 1d time series data from the spatial data

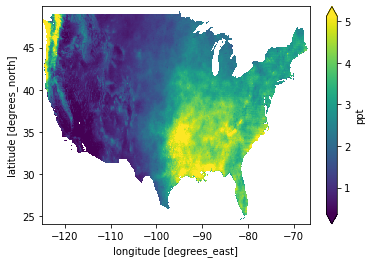
**Regridding/interpolating data**

* Since two datasets had different resolutions and different coordinates for each pixel of the data, we had to find a way to alter one of the datasets to match the other.
* We first sliced the data using the longitude and latitude boundaries.
* We used linear interpolation to create a continuous surface of precipitation values and generated a discrete, gridded dataset that matches the other dataset’s coordinates and resolution.
* We then masked one of the datasets using the value from another dataset(where the precipitation is nonnegative). This makes sure that the two datasets contain the same area.

# Datasets

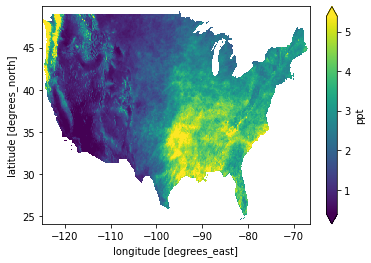
[CHIRPS 2.0 and 3.0](https://www.chc.ucsb.edu/data/chirps)

* Quasi-globe, daily, high resolution (5 x 5 km) precipitation data
* Used for early drought warnings and environmental monitoring
* Constructed using satellite images and ground weather station data
* Example Visualization for 2015 data (yearly average)



PRISM

* Similar to CHIRPS but only covers the US and has higher resolution (4 x 4 km)
* Constructed using ground weather stations
* Example Visualization for 2015 data (yearly average)



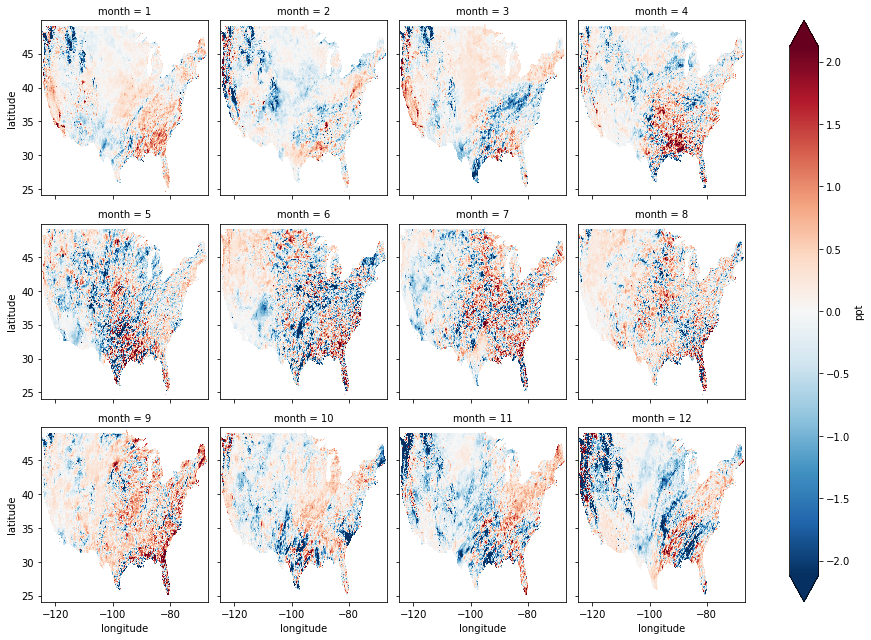
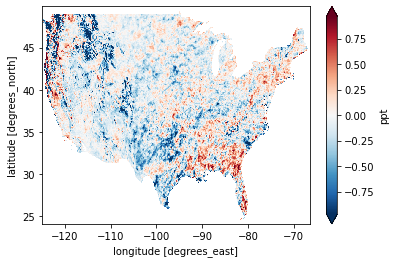
# Objectives

* Investigate how daily rainfall intensity and number of rainy days in CHIRPS 2.0 and CHIRPS 3.0 dataset compare with other widely used global precipitation dataset
* Examine the improvements in CHIRPS 3.0 daily precipitation relative to the version 2.0

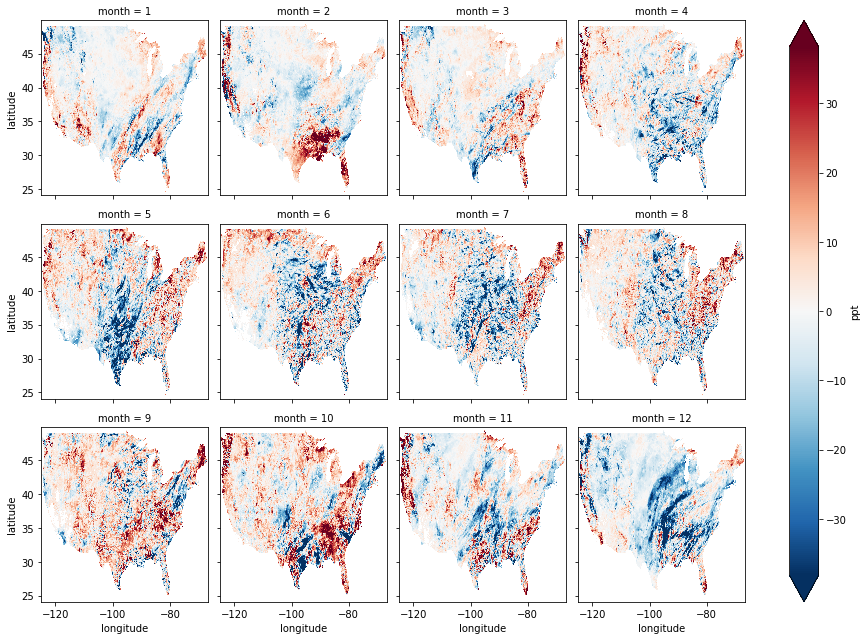
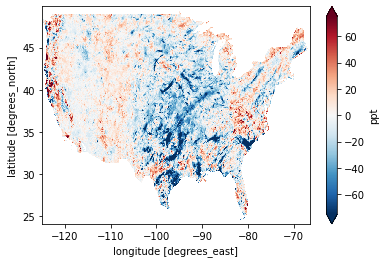
# Current/planned work

* Construct a data processing and analysis pipeline that takes in two yearly data of CHIRPS and other dataset and outputs all the visualizations and time series metrics to see the differences between two datasets
  + Current matrix:

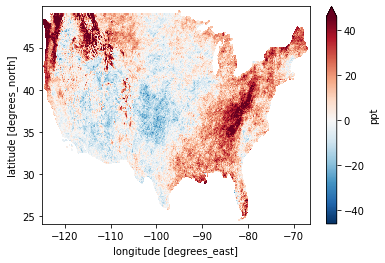
Differences in yearly average precipitation and monthly average precipitation



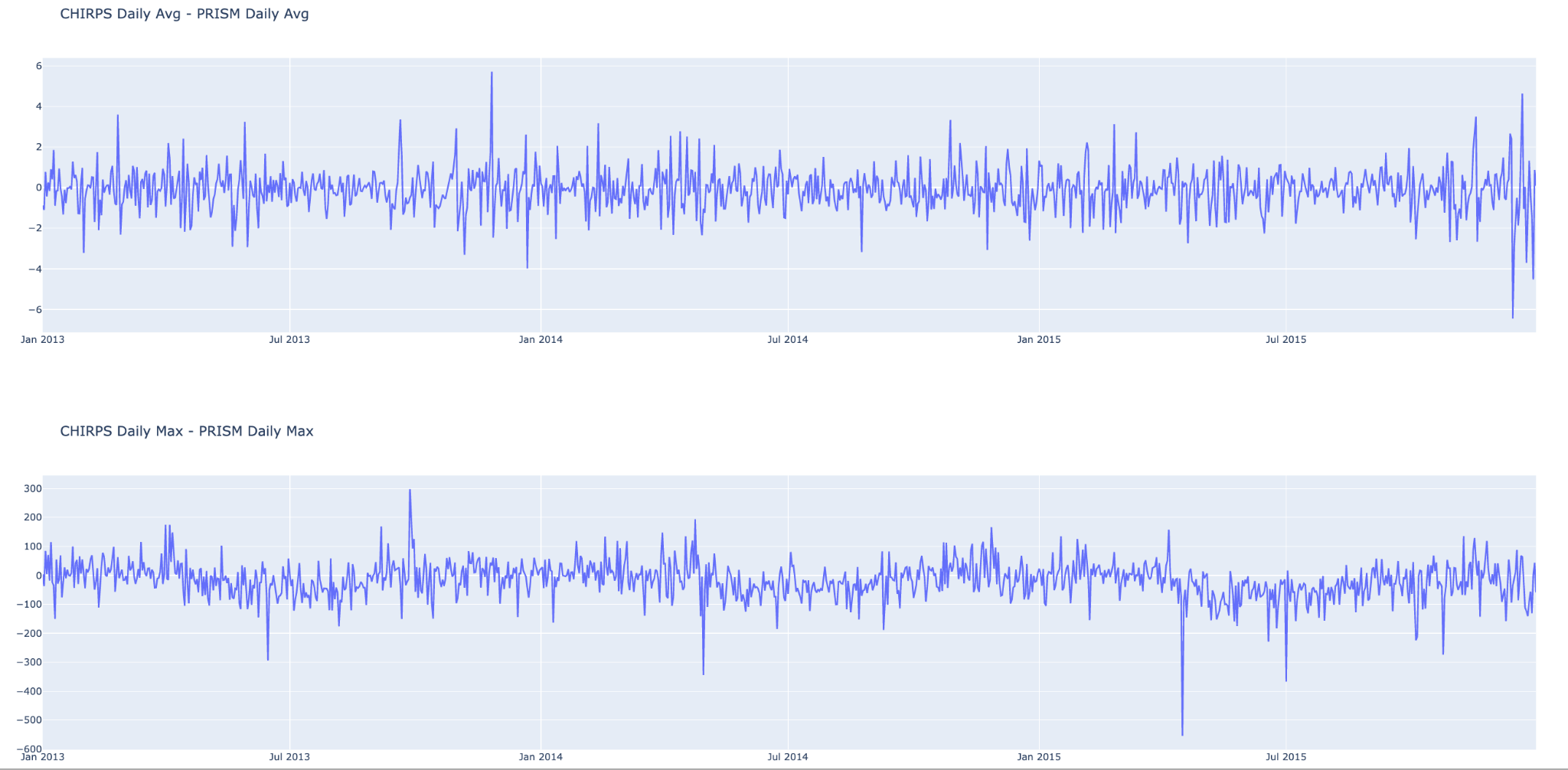
* + Differences in yearly maximum precipitation and monthly maximum precipitation



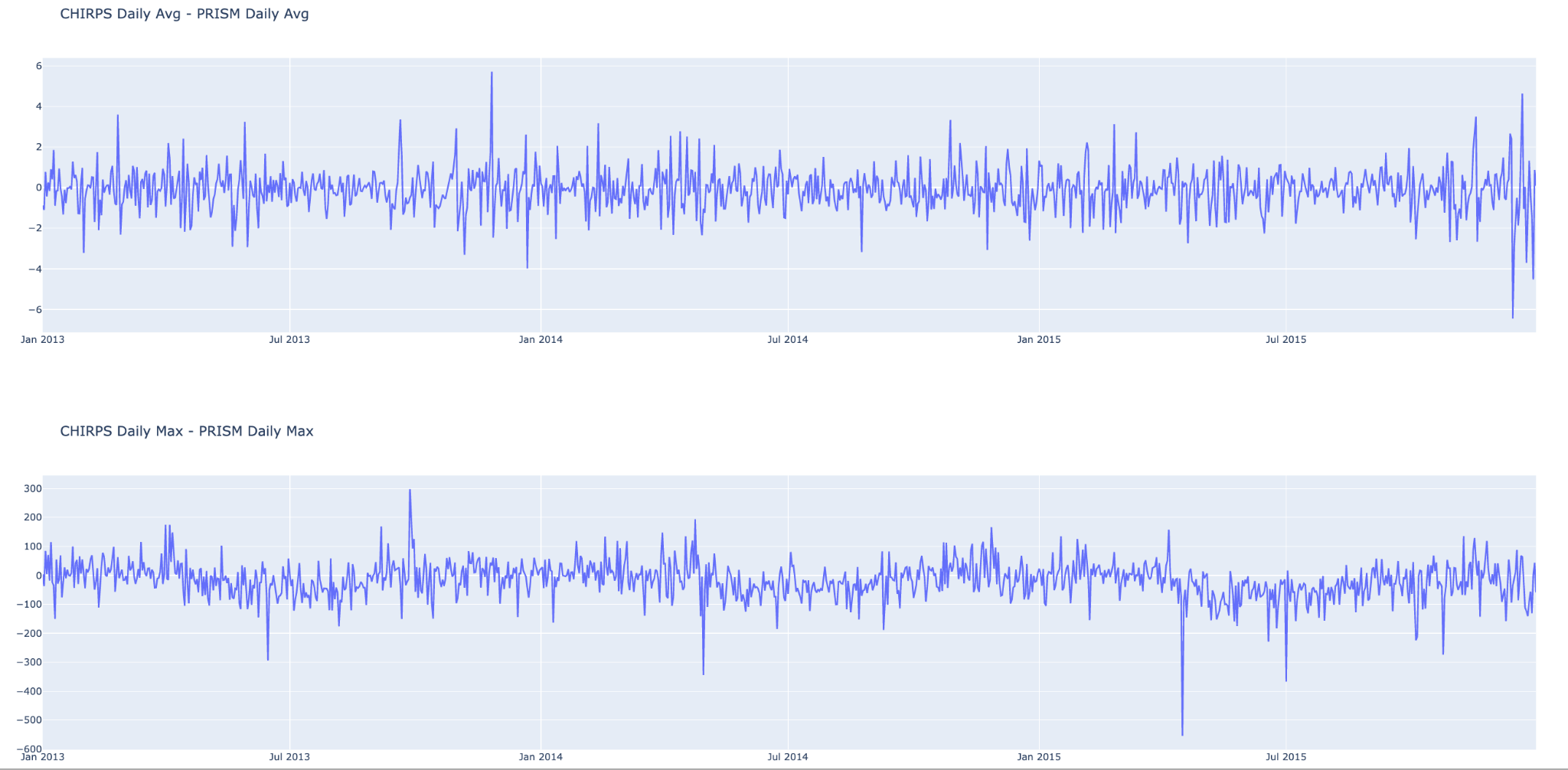
Differences in the number of days for difference precipitation thresholds in a year/each month (i.e. First specify a range of precipitation; then count the number of days that have precipitations in that range)



Time series of the average daily differences in precipitation



Time series of the maximum daily difference in precipitation



* Key findings thus far (after comparing with the PRISM dataset)

CHIRPS contains LOTS OF dry data (precipitation = 0mm)

Generally, CHIRPS detects more rainfall in eastern area compared to PRISM

In 2-5 mm precipitation range, PRISM contains more data, especially in eastern area

* Run the analysis on 30 years worth of data and document the results and evaluation of CHIRPS 3.0 compared to other precipitation datasets.