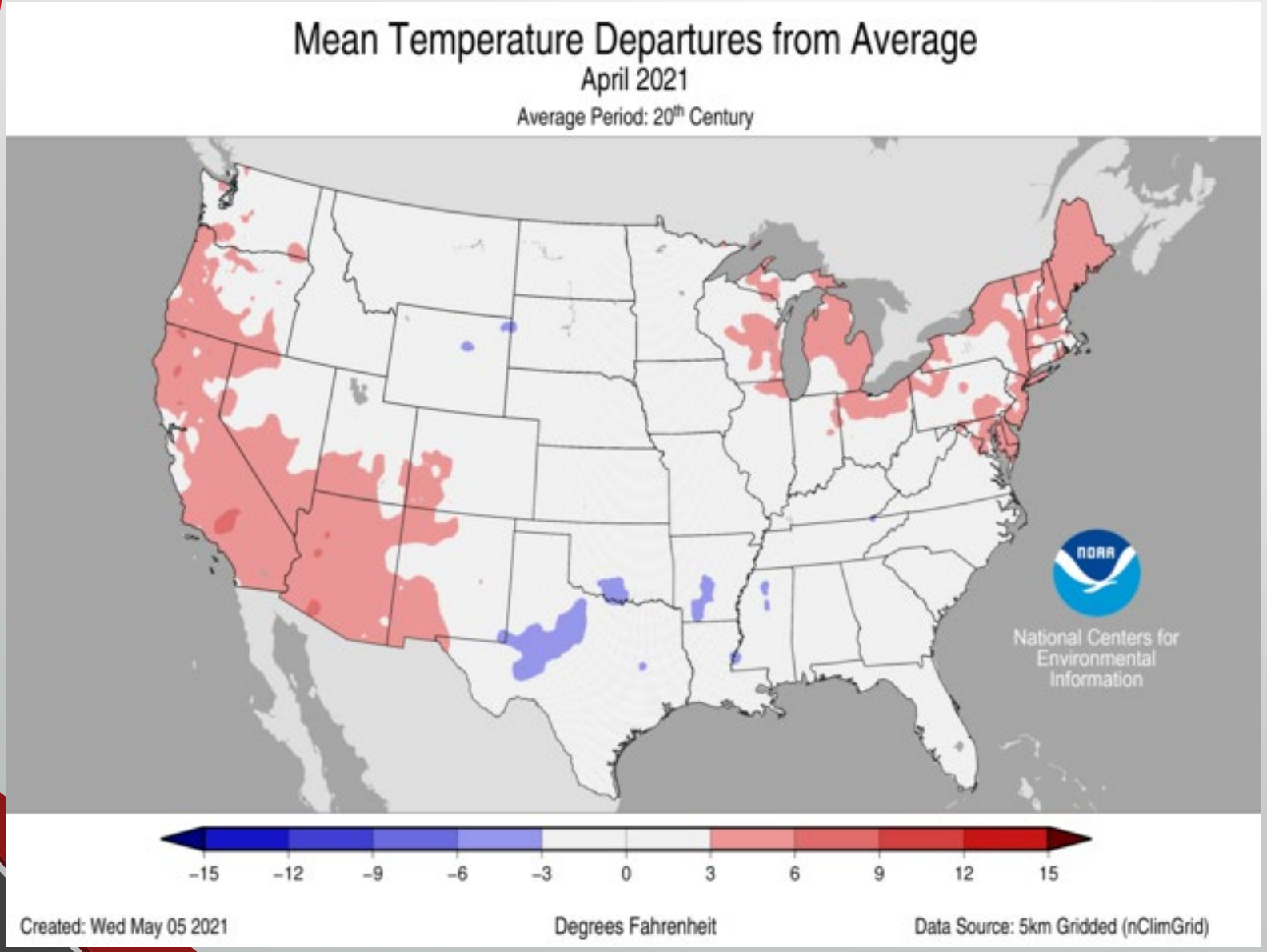
June 17, 2021

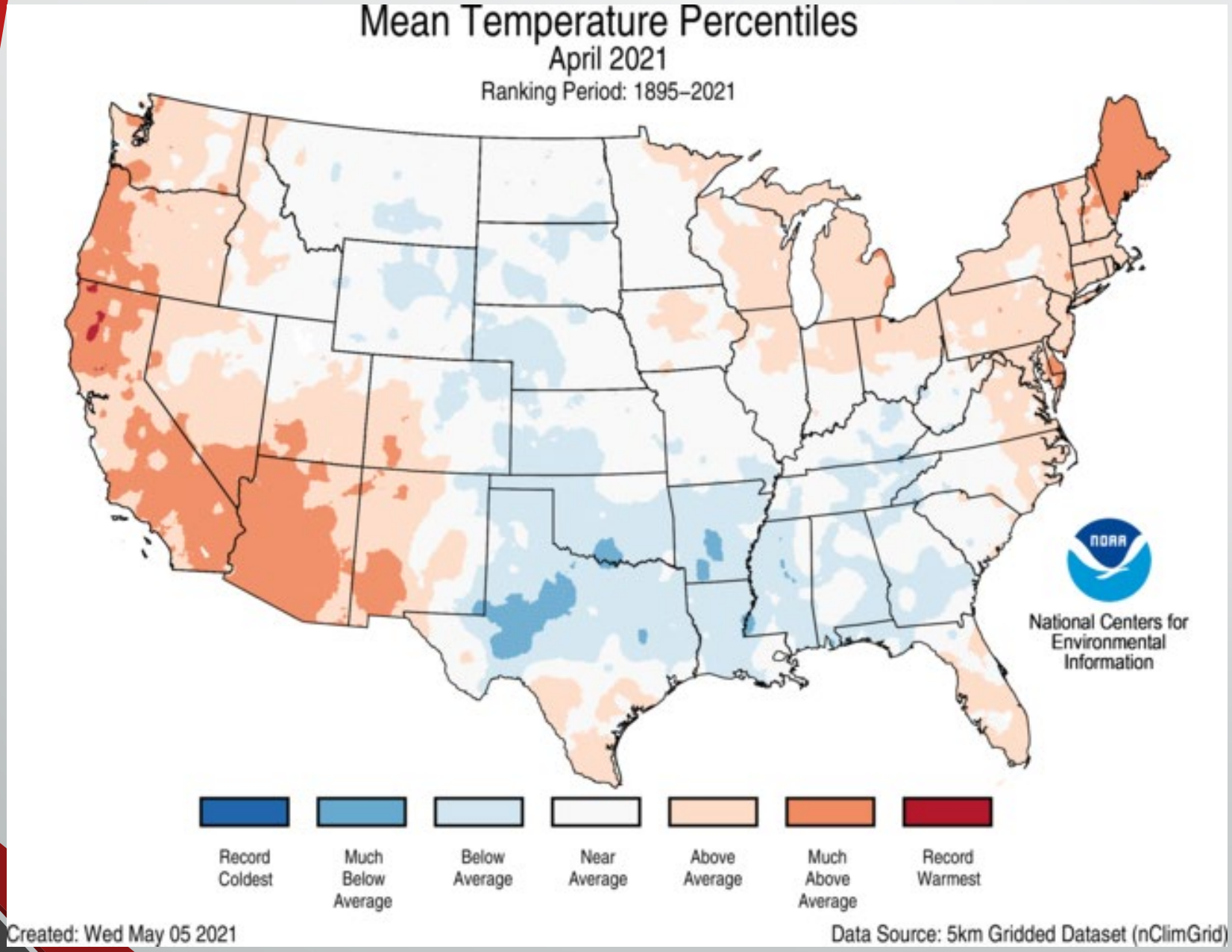
2. Exploratory Univariate Data Analysis
a. Examining time series of data
b. Data Distributions and Histograms
c. Central value, spread, and symmetry
d. Transforming Data
e. Exploratory Univariate Analysis of Fluid Velocity
f. Check Your Understanding

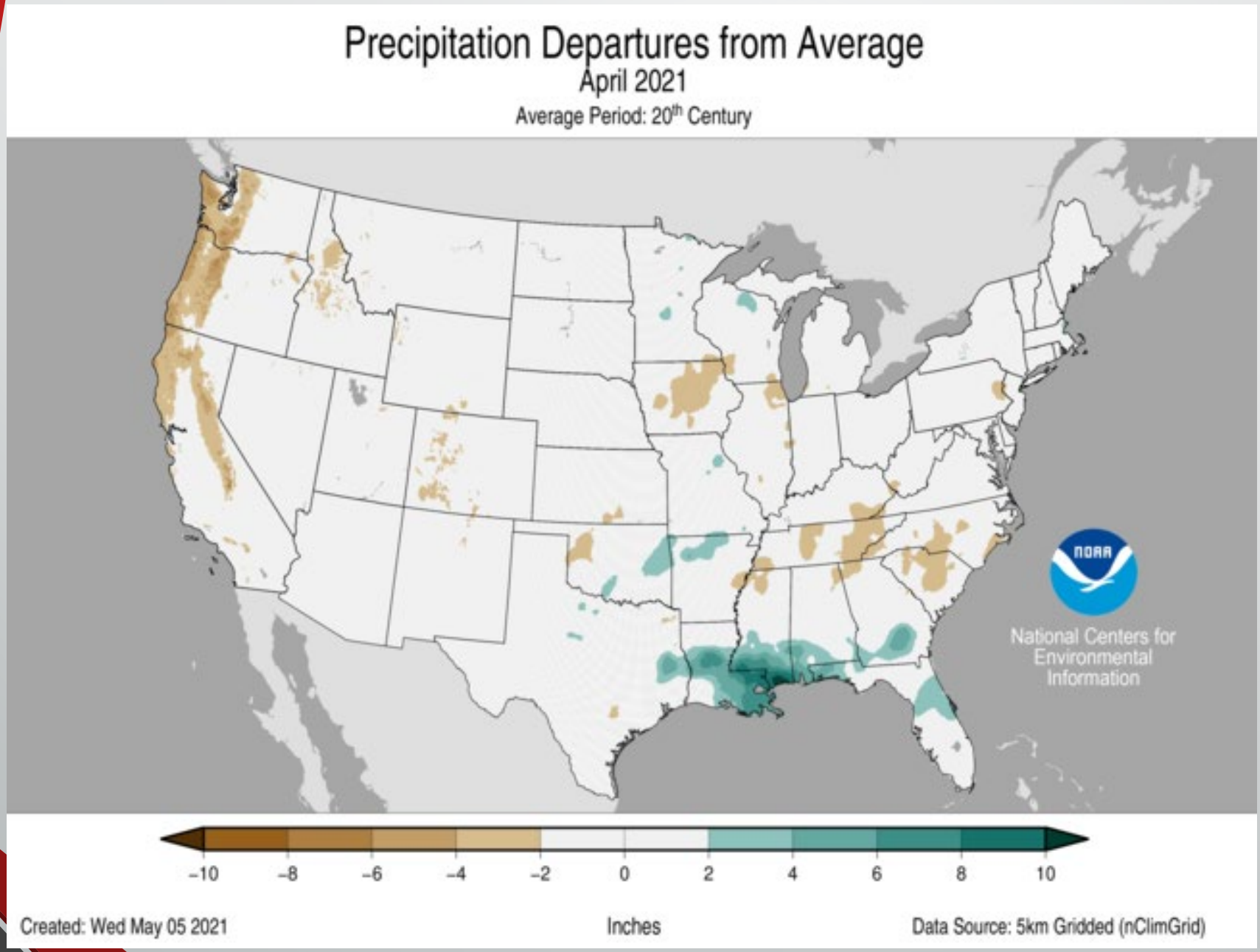
U.S. 2020 Billion-Dollar Weather and Climate Disasters

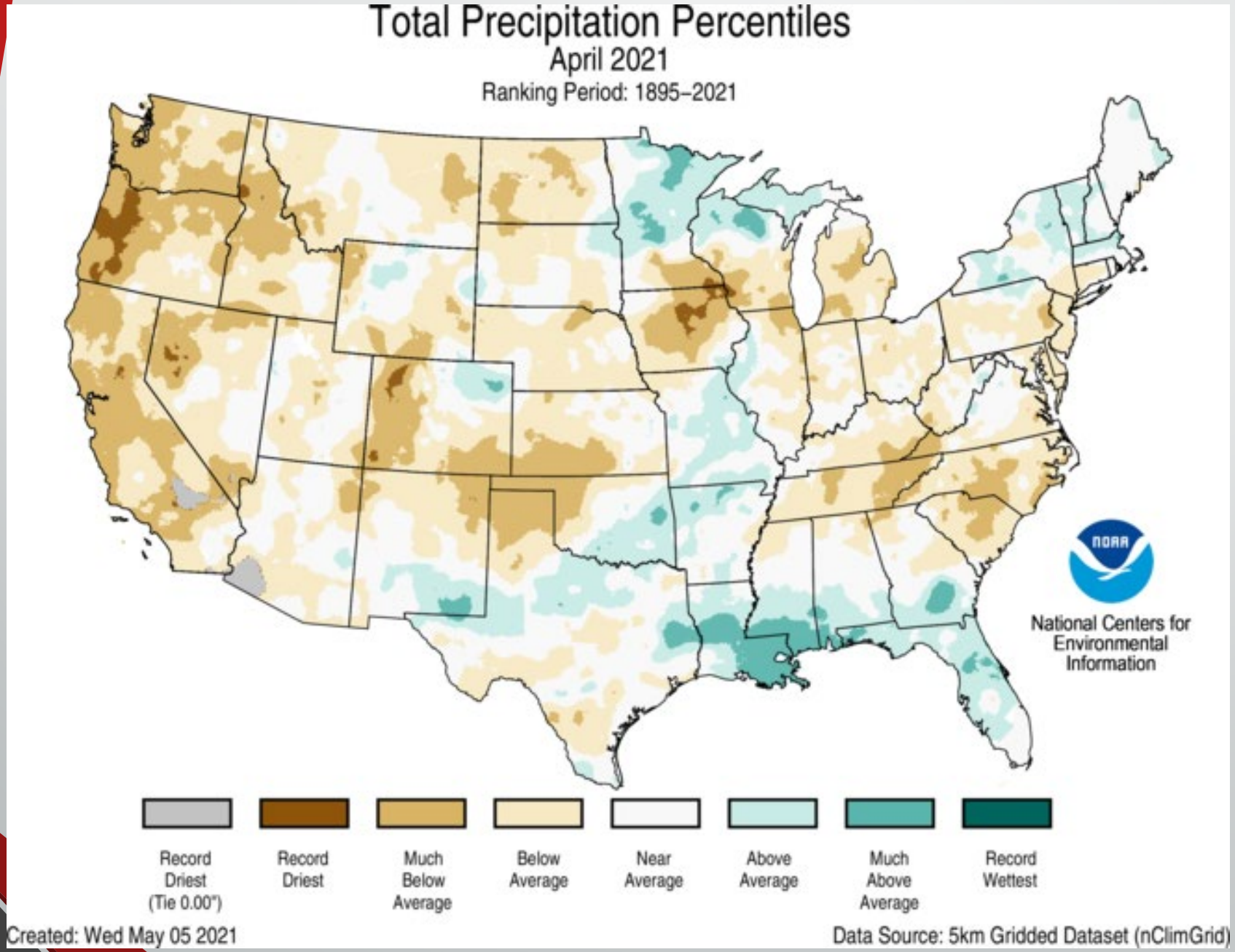
*This map denotes the approximate location for each of the **16 separate billion-dollar weather and climate disasters** that impacted the United States **from January–September 2020**.*

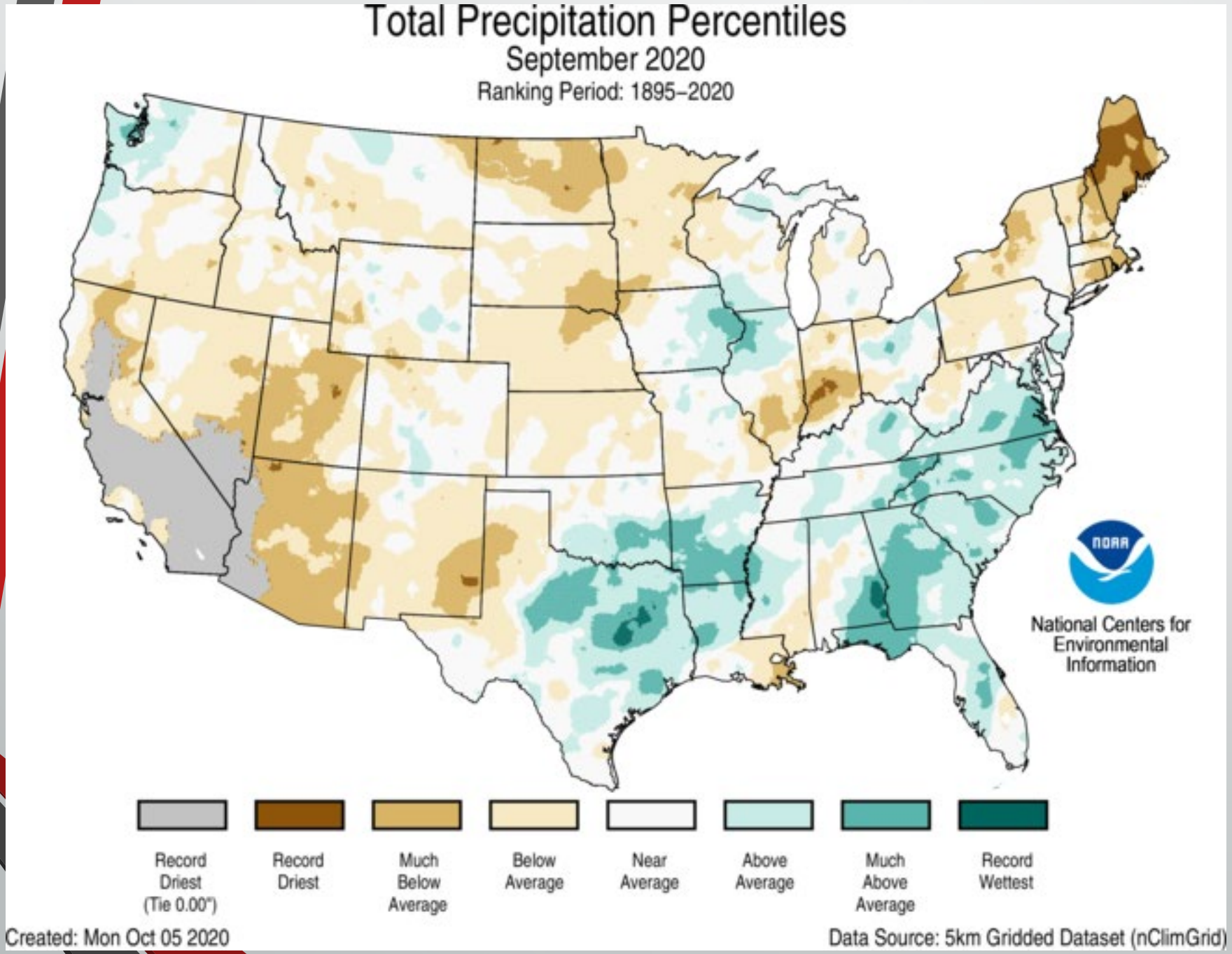
https://www.ncdc.noaa.gov/temp-and-precip/us-maps/1/202104#us-maps-select




Mean Temperature Percentiles
April 2021
Ranking Period: 1895–2021
National Centers for Environmental Information
Record Coldest
Much Below Average
Below Average
Near Average
Above Average
Much Above Average
Record Warmest
Created: Wed May 05 2021
Data Source: 5km Gridded Dataset (nClimGrid)


Precipitation Departures from Average
April 2021
Average Period: 20th Century
National Centers for Environmental Information
−10
−8
−6
−4
−2
0
2
4
6
8
10
Created: Wed May 05 2021
Inches
Data Source: 5km Gridded (nClimGrid)


Total Precipitation Percentiles
April 2021
Ranking Period: 1895–2021
National Centers for Environmental Information
Record Driest (Tie 0.00")
Record Driest
Much Below Average
Below Average
Near Average
Above Average
Much Above Average
Record Wettest
Created: Wed May 05 2021
Data Source: 5km Gridded Dataset (nClimGrid)


Total Precipitation Percentiles
September 2020
Ranking Period: 1895–2020
National Centers for Environmental Information
Record Driest (Tie 0.00")
Record Driest
Much Below Average
Below Average
Near Average
Above Average
Much Above Average
Record Wettest
Created: Mon Oct 05 2020
Data Source: 5km Gridded Dataset (nClimGrid)

June 17, 2021

Use opendemand to run chapter_2e.ipynb

Did it run without any errors? If not, try to debug. Send a message in Teams if you can't figure it out

Exploratory Empirical Analyses

OBJECTIVE IS TO REDUCE THE COMPLEXITY (DIMENSIONALITY) WITHIN A LARGE DATA SET

WHAT IS A VALUE COMMONLY OBSERVED?

HOW MUCH VARIABILITY IS THERE AMONG ALL THE VALUES?

WHAT ARE EXTREME CASES THAT HAVE BEEN OBSERVED?

What we want…

Metrics that are:

- ***robust***- which means not overly sensitive to the characteristics of the entire sample of data values. In other words, we want the measure to perform reasonably well no matter how the data values are distributed.
- ***resistant-*** not unduly influenced by outliers in the sample. For example, the range is not resistant to outliers.

c. Central Value, Spread, and Symmetry

- **central value** (central tendency or typical value)
- **spread** (variation or dispersion about the central or typical value)
- **symmetry** (degree to which the values tend to be larger or smaller than the central value)

Metrics

Central tendency

- Median
- Mode
- Mean
- Trimmed mean

Spread

- Range
- Interquartile Range
- Median absolute deviation
- Standard deviation
- Variance

Mode, Mean, Trimmed Mean

- Mode- most common values
- Mean- not robust or reliant- sensitive to outliers

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Trimmed mean- robust and reliant but throwing away data: k values on either end of the distribution

$$\overline{X_\alpha} = \frac{1}{n-2k}\sum_{i=k+n}^{n-k} x_i$$

Spread

- Range- very sensitive to outliers
- IQR- $75^{th} - 25^{th}$ percentiles (ignores 50% of the data)
- MAD = median $| x_i - q_{.5} |$- robust and reliant
- **standard deviation**, s, is the most commonly used measure of spread, but not resistant to outliers or robust

(2.c.3)

Standard Deviation/Variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Standard deviation, s, is not resistant to outliers or robust

Variance, s^2, an **unbiased estimate of the population variance**

Why n-1 rather than n???

Symmetry

- Skewness- a measure of how the values are symmetric or asymmetric

$$\gamma = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}$$

Transforming Data:
Anomalies: departure from long-term mean

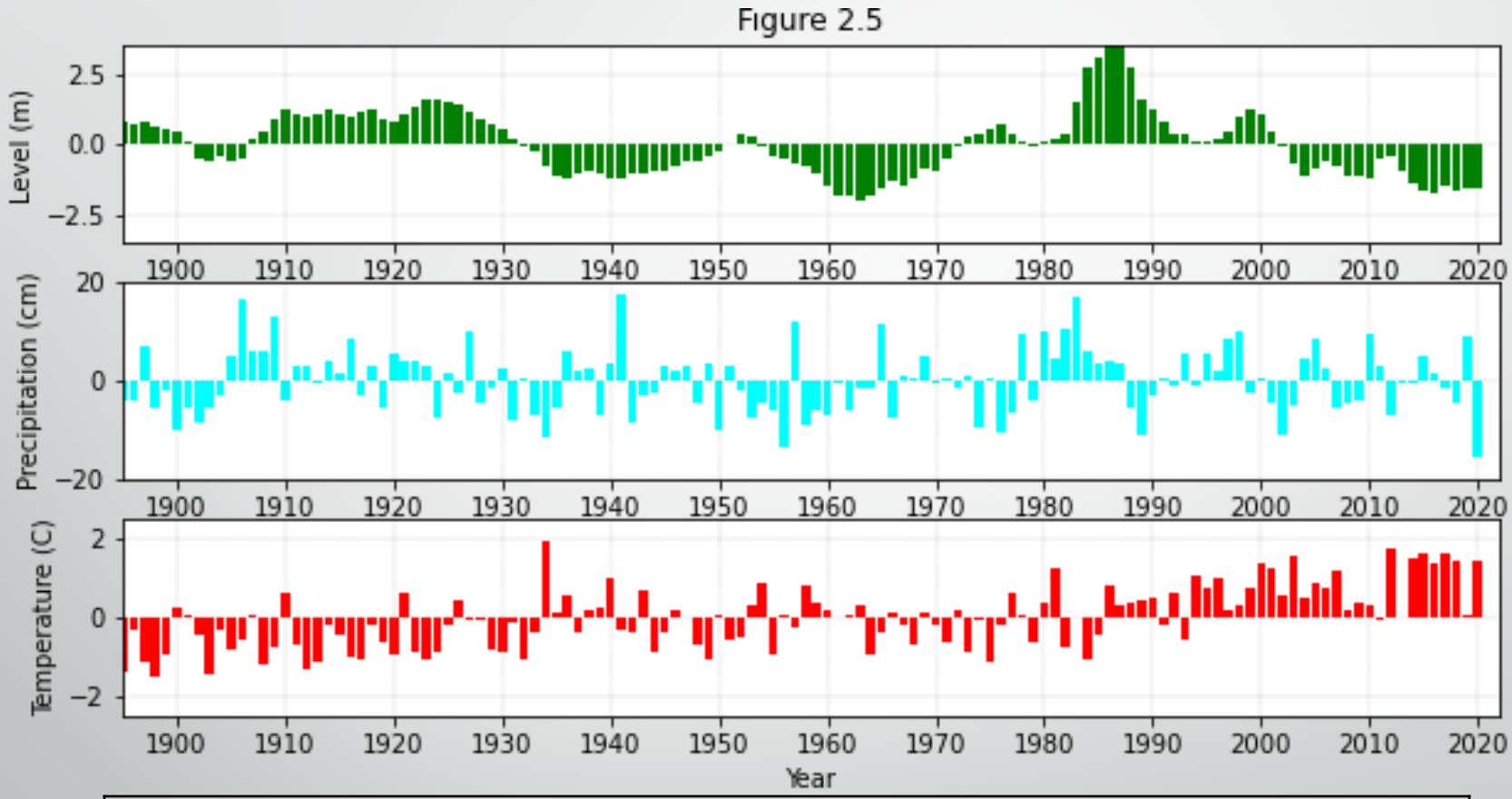


Figure 2.5. Top. Departure (m) of lake level from 126 year sample mean. Middle. Departure (cm) of Utah precipitation from sample mean. Bottom. Departure (^{o}C) of Utah temperature from sample mean.

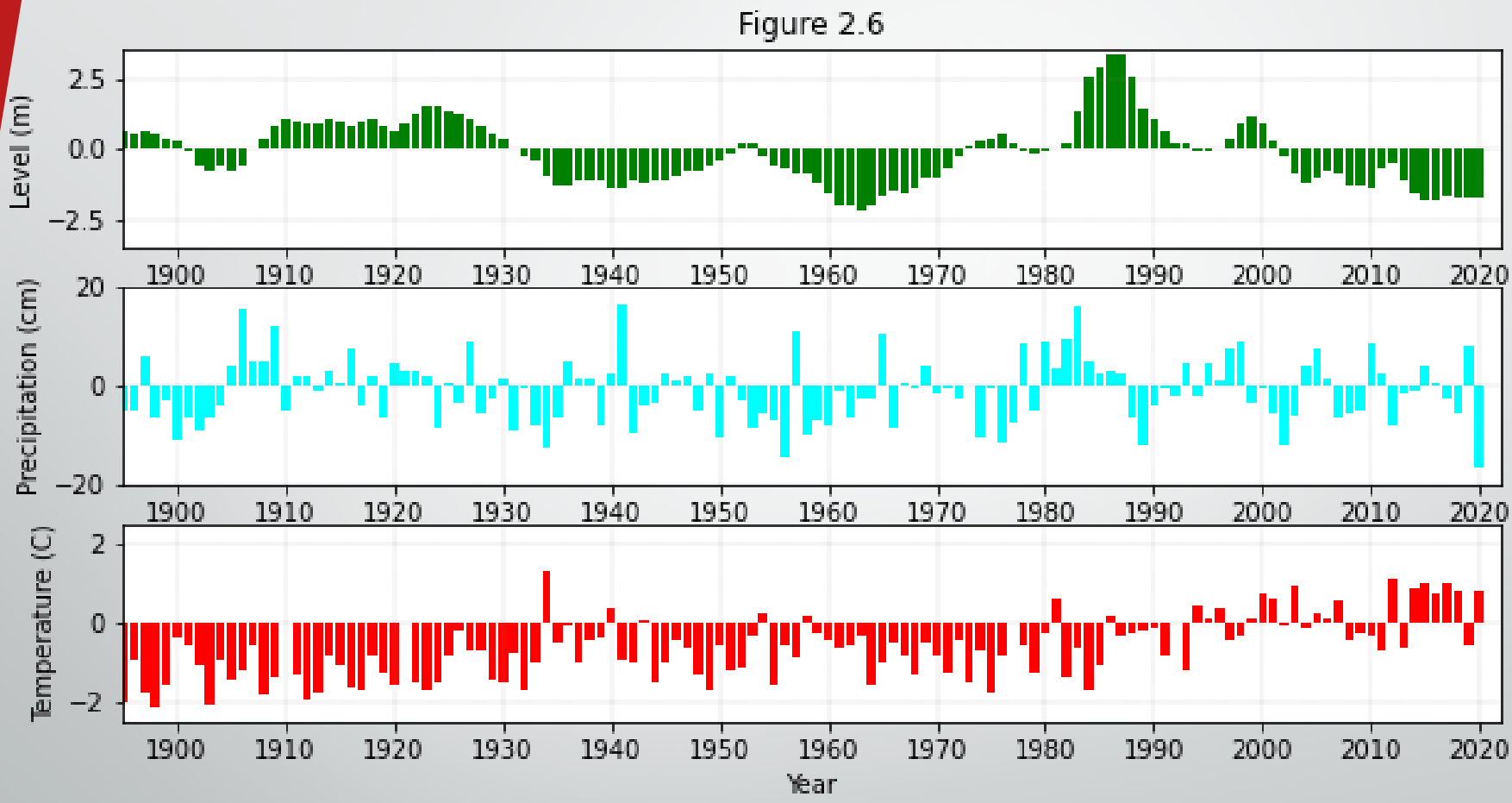


Figure 2.6. Anomalies relative to 1991-2020 climate normal.

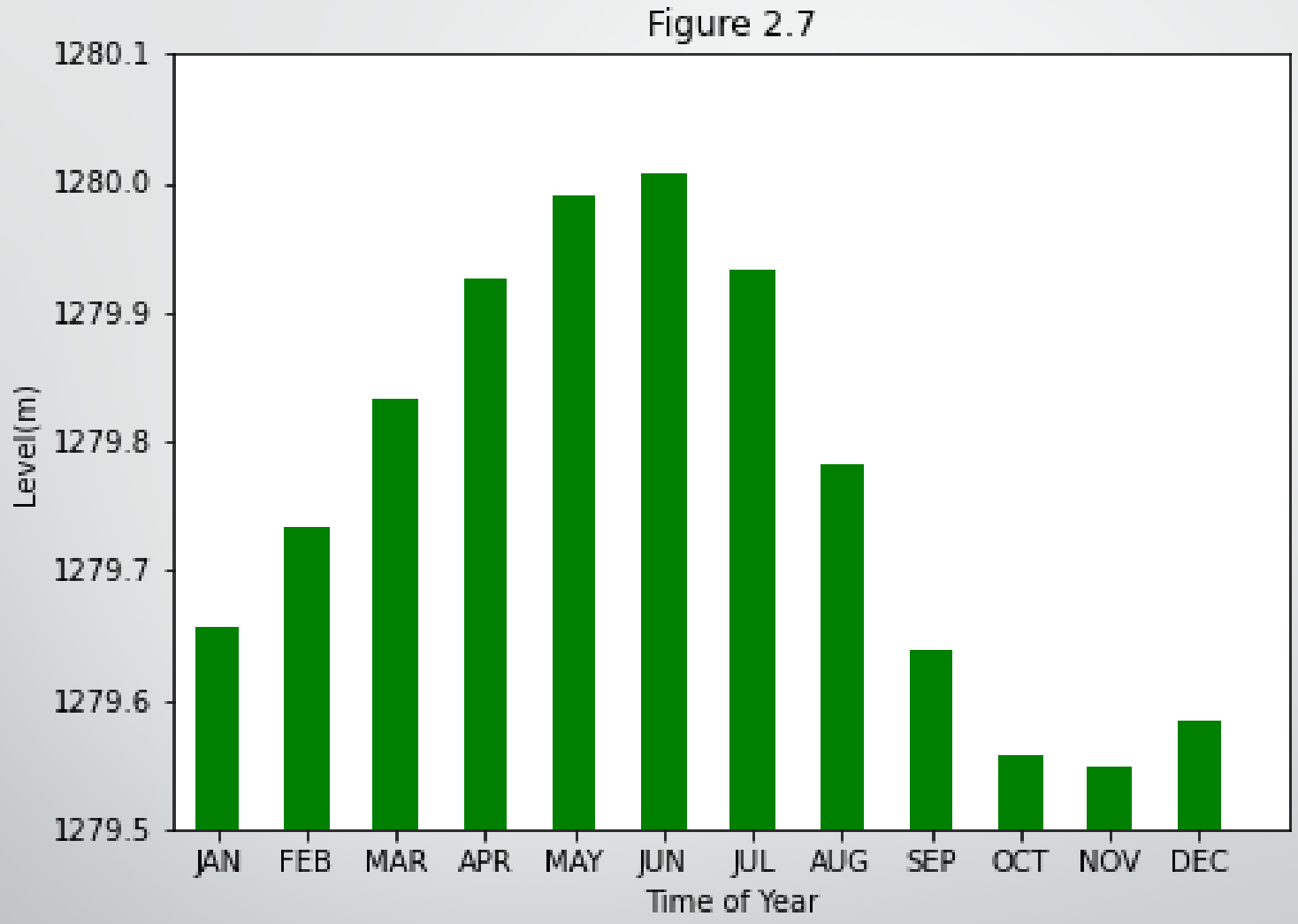


Figure 2.7. Lake level (m) averaged separately for each calendar month over the 118 years.

Transforming data

- Removing climatological seasonal cycle
- Computing standardized (non-dimensional) anomalies

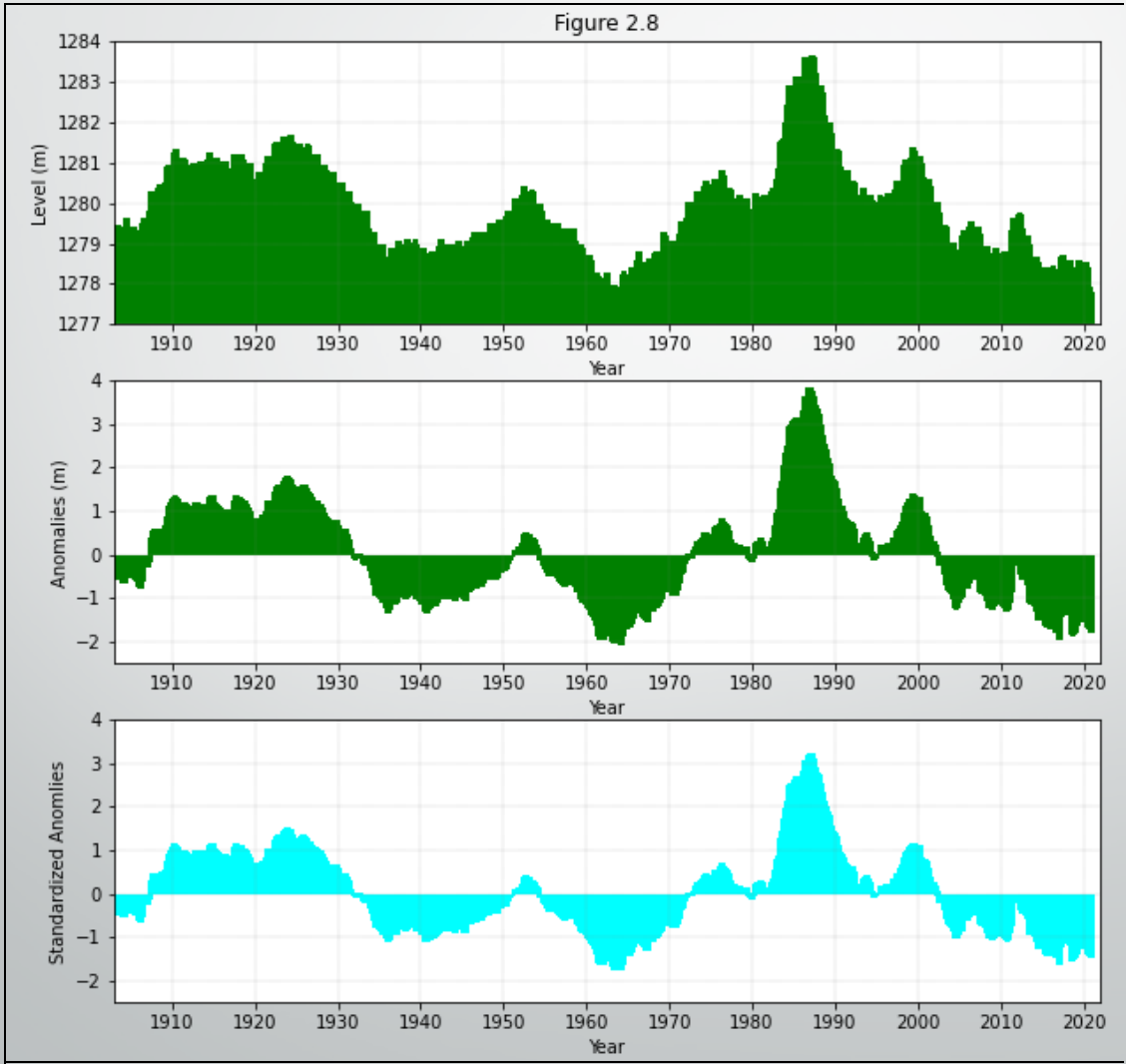


Figure 2.8. Top panel. Monthly lake level (m). Middle panel. Departures from monthly means (m). Bottom panel. Standardized anomalies (nondimensional).

CDF of Monthly Standardized Anomalies

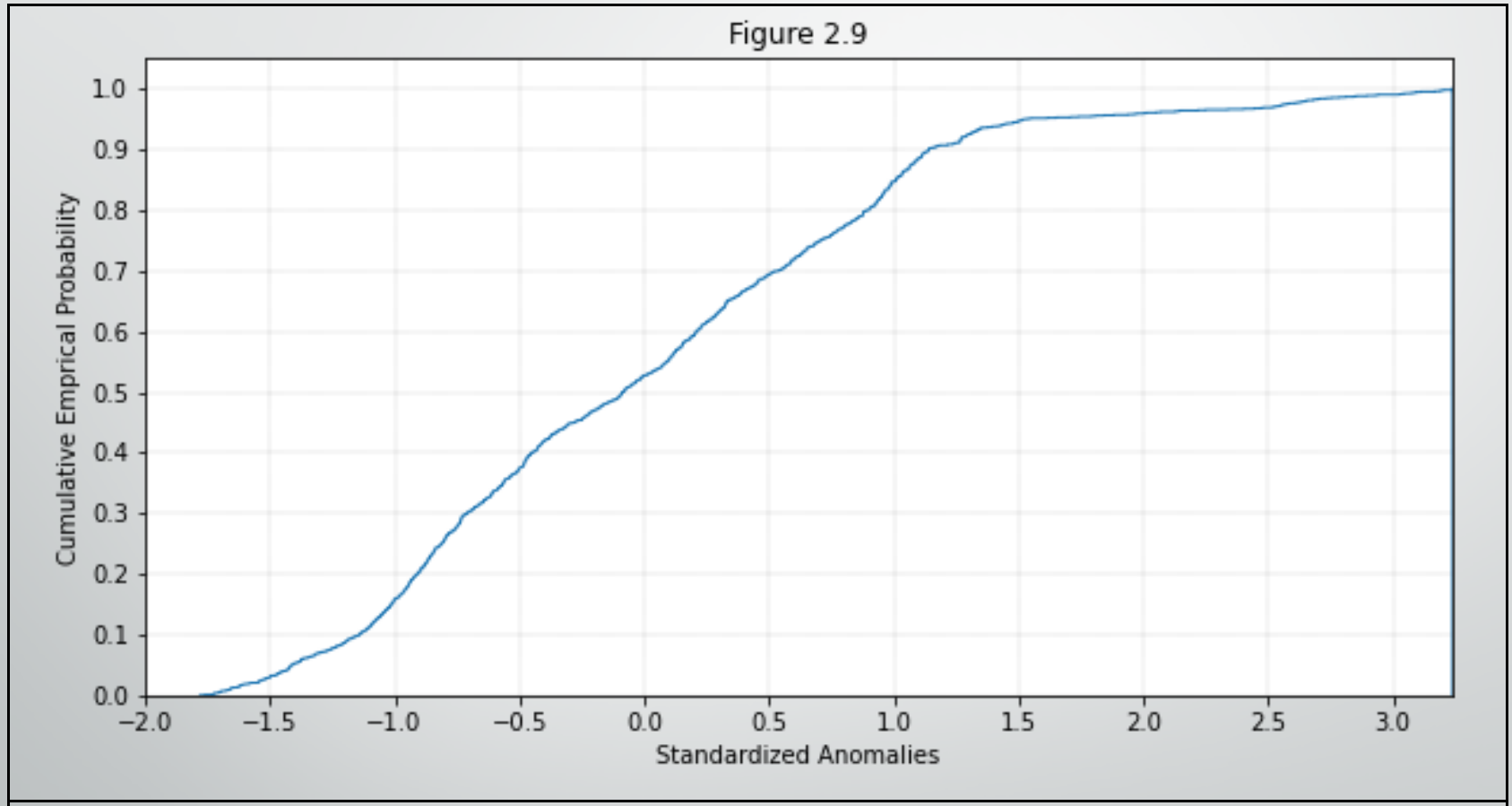


Figure 2.9. Cumulative frequency distribution for monthly values of lake level.

Filter characteristics

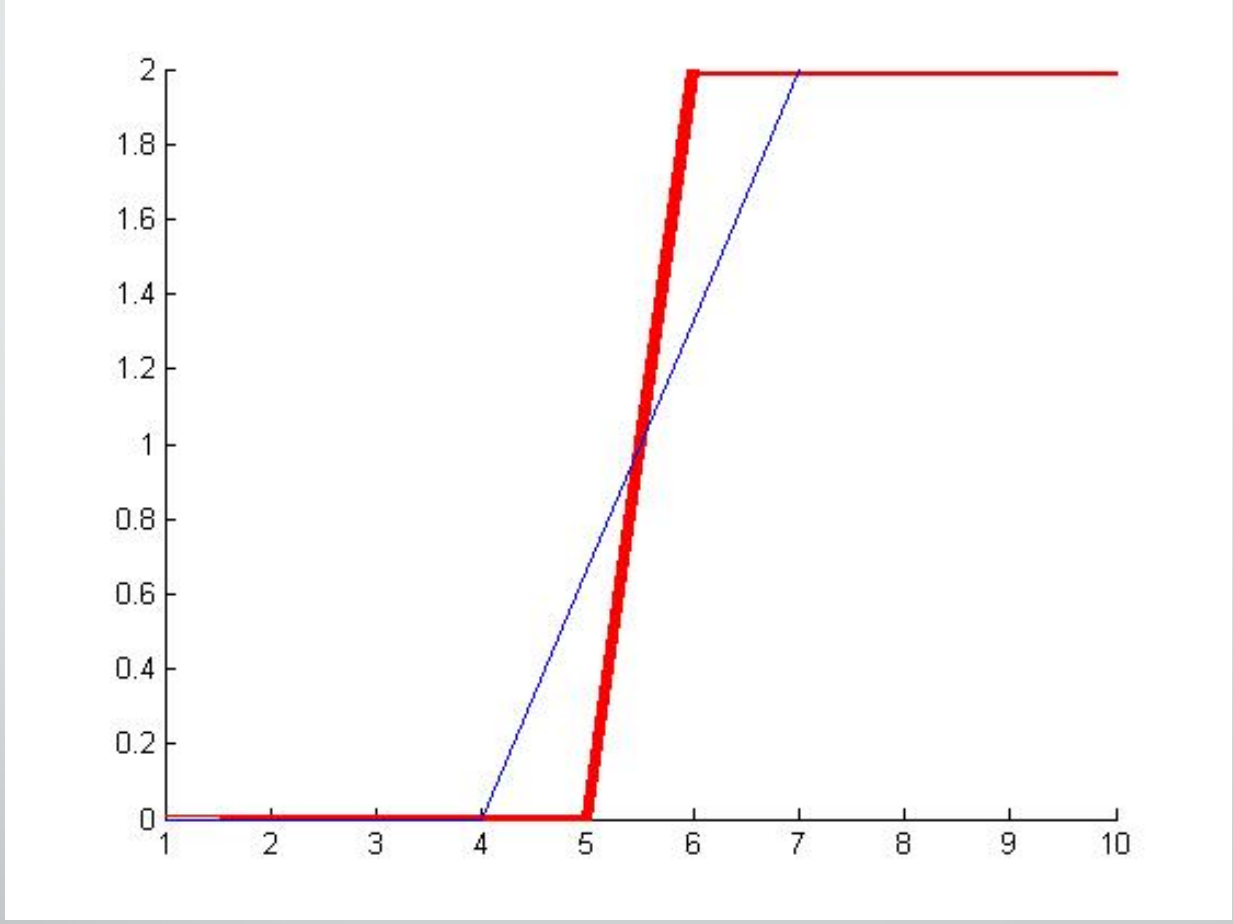


Figure 2.10. Example of original and median smoother (red line) and running mean smoother (blue line).

Transforming Data

- Low pass filter: keep slow variations, remove fast ones

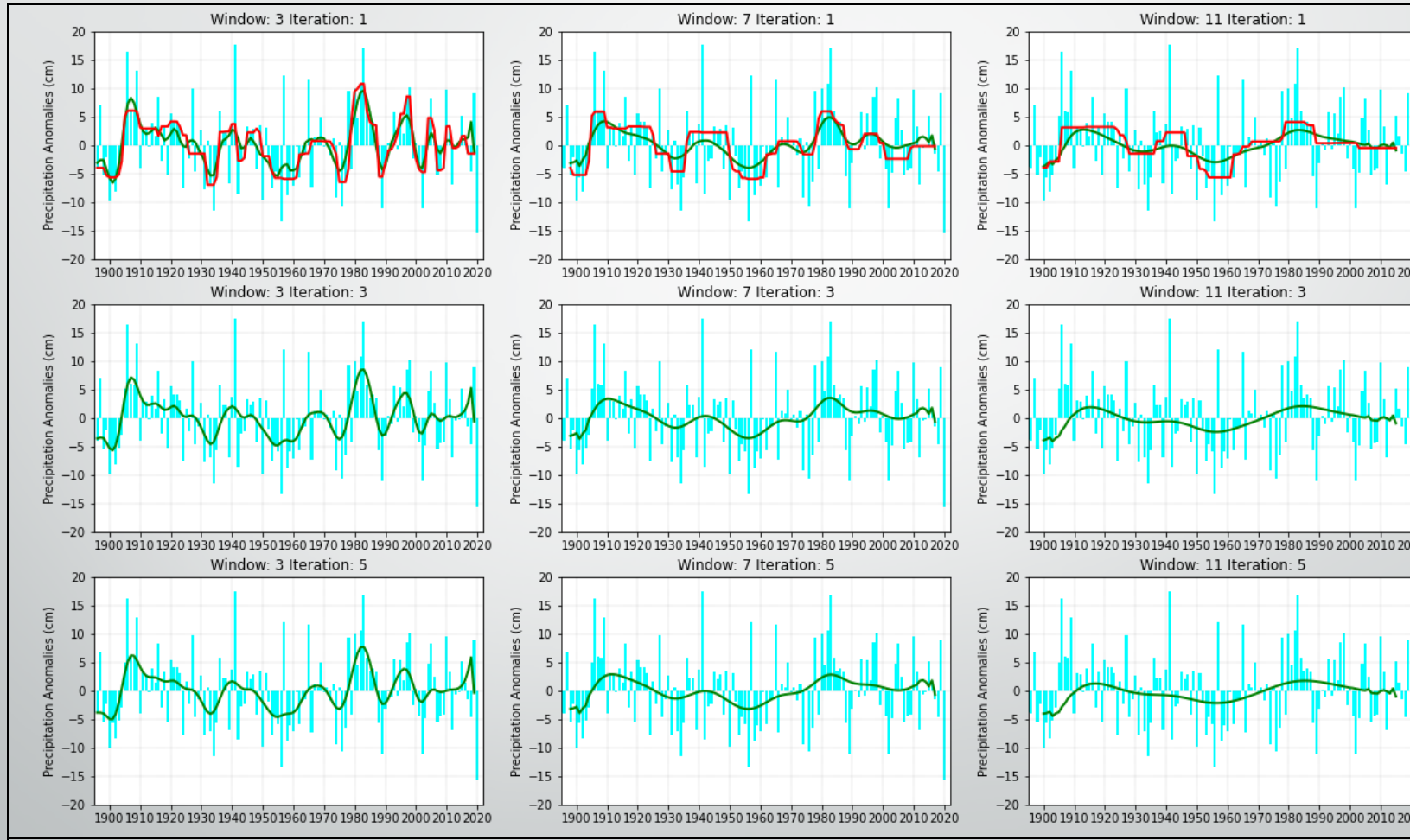
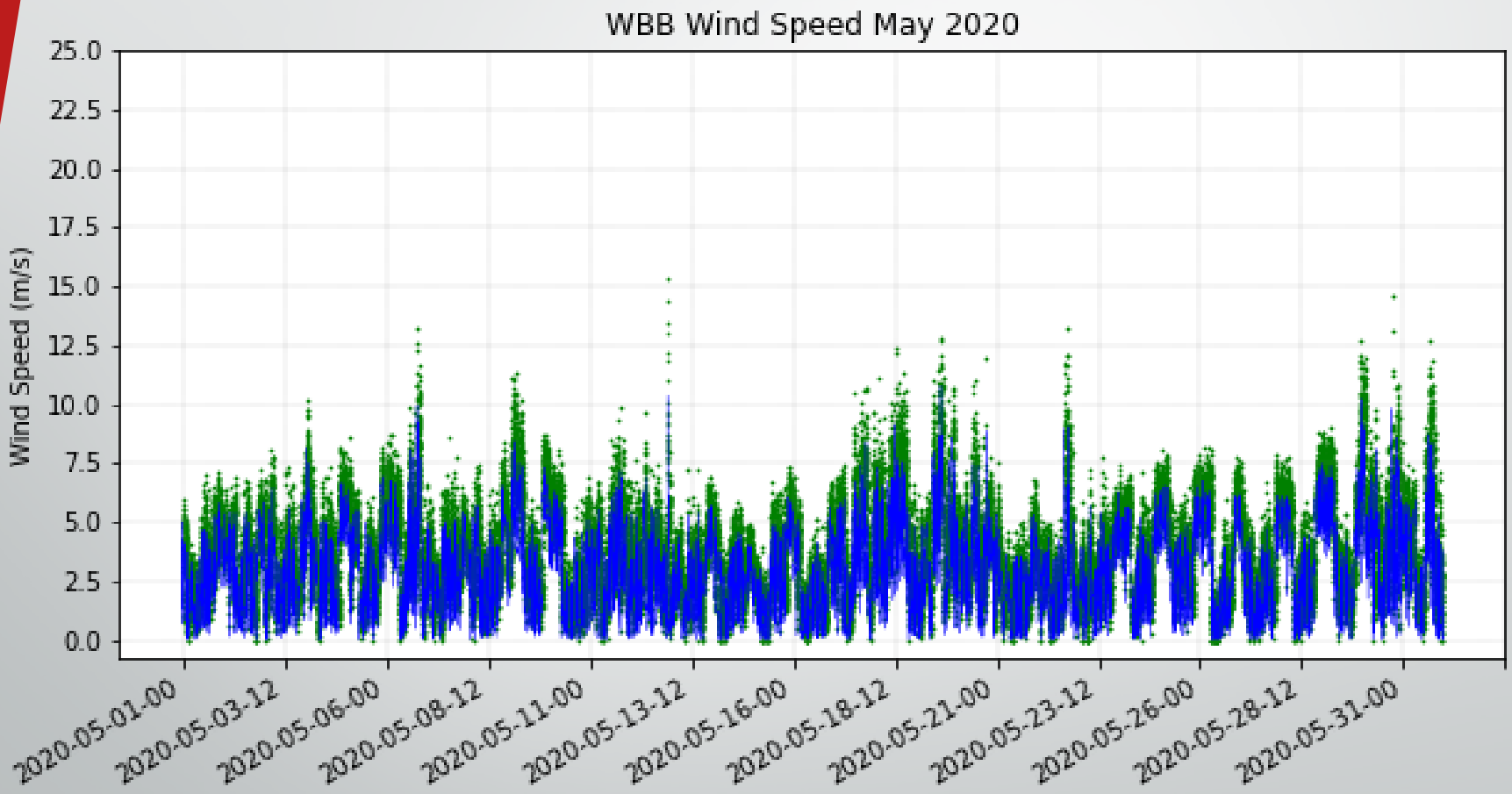


Figure 2.11. Utah annual precipitation anomalies (cyan bars) and the data smoothed using running mean smoother (green lines) using different windows (sample sizes; columns) and iterations (rows). The red lines in the top row indicate running median smoothers using different sample sizes (there would be no difference is iterated multiple times).

Run chapter_2e_2021.ipynb
Any errors?



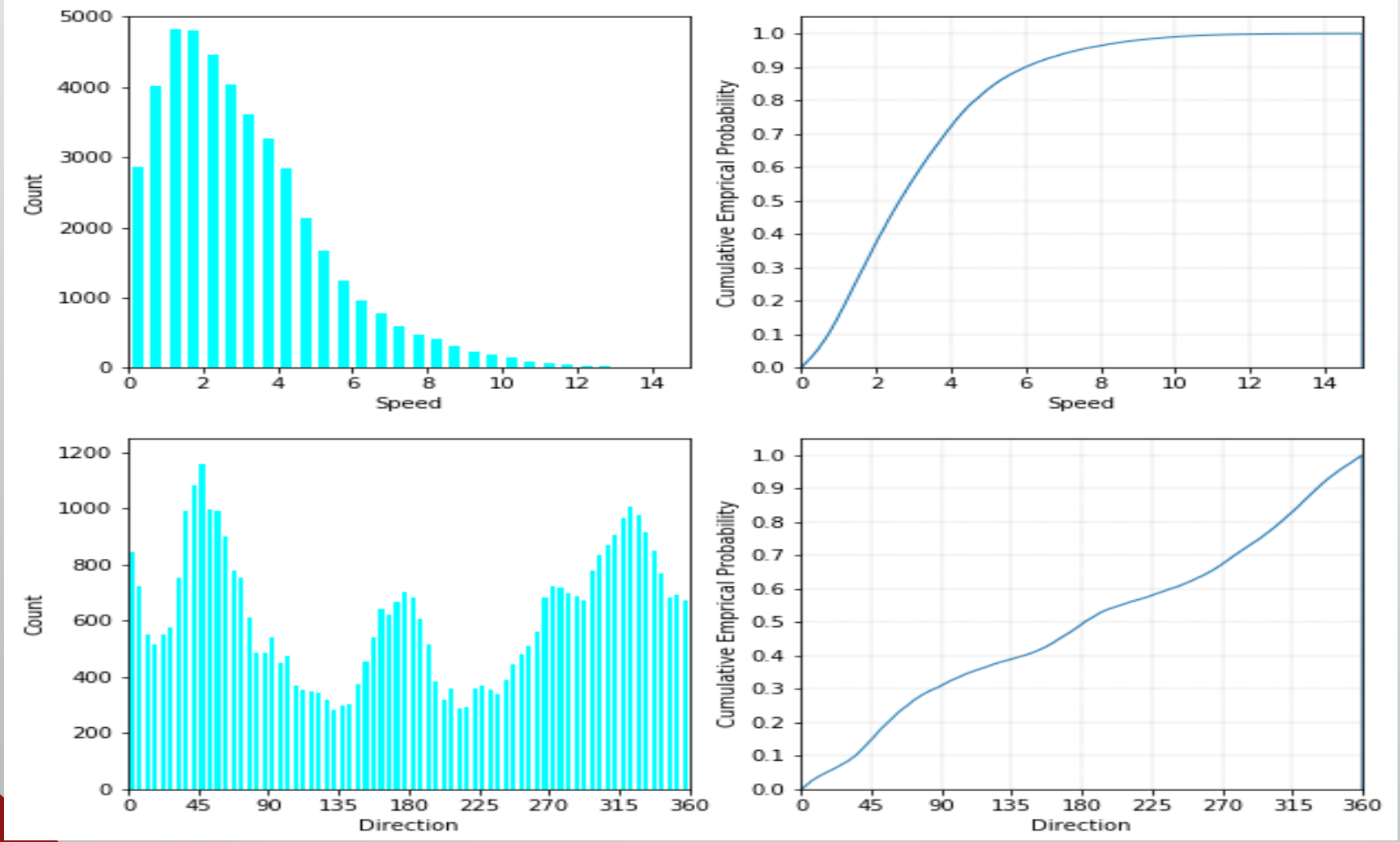
Basic Statistical Methods for Fluid Flow

- horizontal speed V and direction (θ)
- θ is the direction from which the wind blows: north wind is 0; east wind is 90; south wind is 180; west wind is 270
- horizontal Cartesian components,
  - zonal *u* (east-west with *u* positive when fluid motion is from west to east)
  - meridional *v* (north-south with *v* positive when fluid motion is from south to north)
- $\vec{V} = u\hat{\imath} + v\hat{\jmath}$ and $\vec{V} = \left|\vec{V}\right| \hat{t}$
  - $V = \left|\vec{V}\right| = \sqrt{u^2 + v^2}$
  - $\theta = 180 + tan^{-1}\, u\,/v$

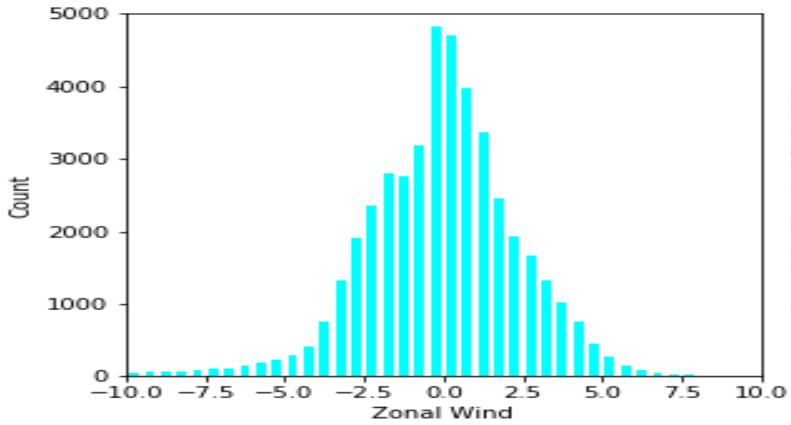
Basic Statistical Methods for Fluid Flow

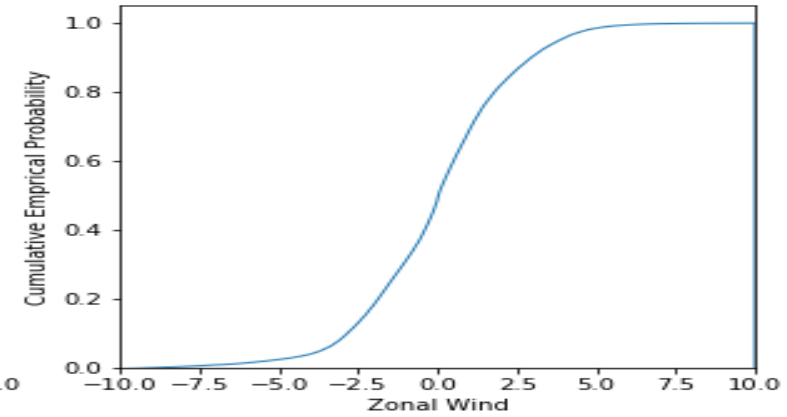
- Horizontaal fluid motion can be described as:
  - speed $\left|\vec{V}\right|$ and direction (θ)
  - Cartesian components, zonal *u* (east-west with *u* positive when fluid motion is from west to east) and meridional *v* (north-south with *v* positive when fluid motion is from south to north
-

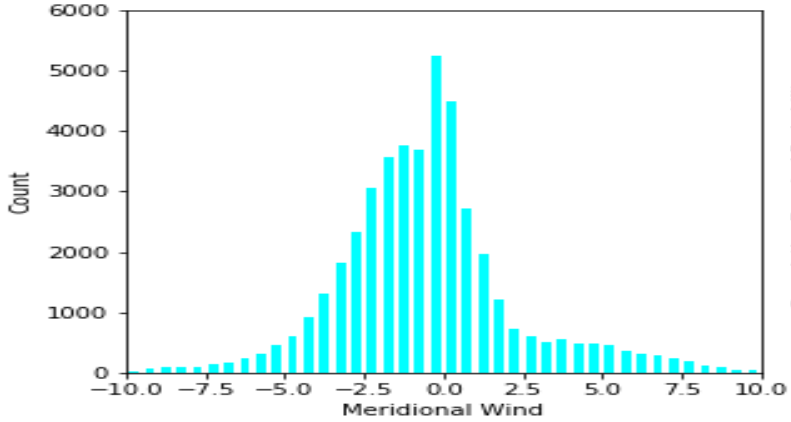
Histograms and cumulative frequency distributions of wind speed and direction

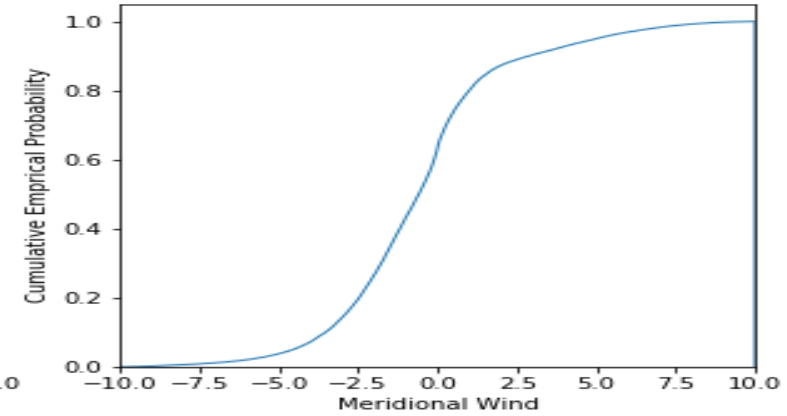


Histograms and cumulative frequency distributions of zonal and meridional wind components

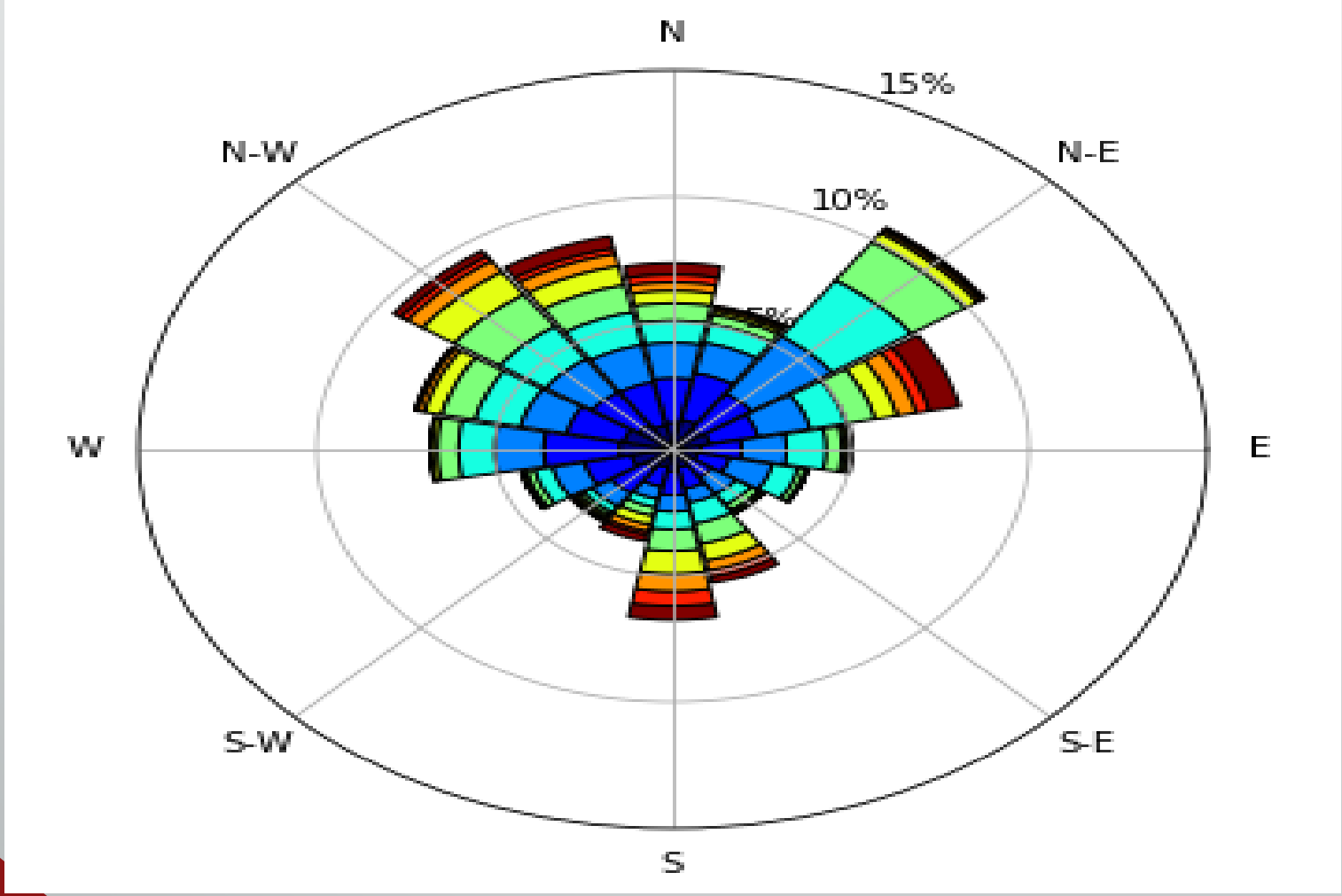






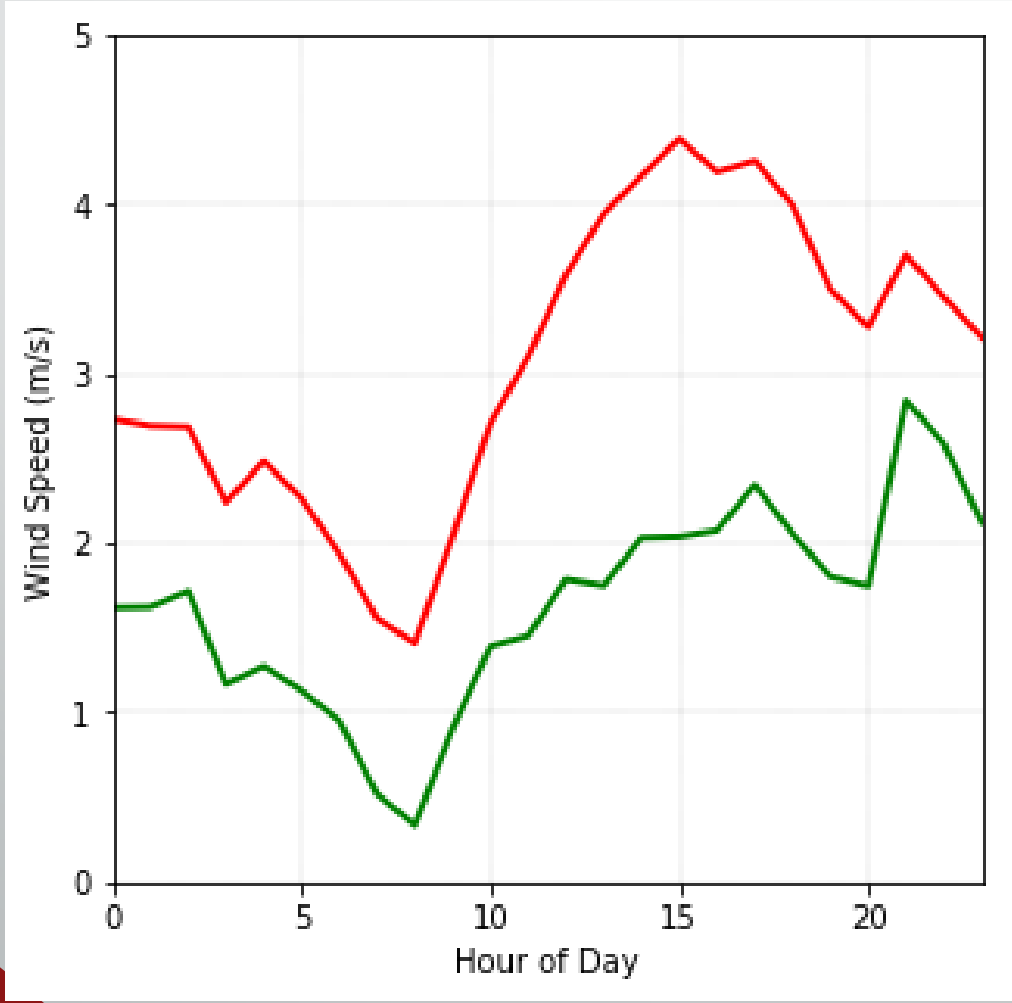


Wind Rose: Counting Speed in Wind Direction Bins



http://home.chpc.utah.edu/~u0553130/Brian_Blaylock/cgi-bin/roses.cgi

Hourly mean wind speed (red line) and resultant wind speed (green line)



What you should be doing

- Maqke sure you can run the chapter 2 python notebooks
- Read all of Chapter 2 notes
- Complete Check Your Understanding #2