SMSS First Year Graduate Student Seminar

Whitney Huang

CLEMSON UNIVERSITY

wkhuang@clemson.edu

February 28, 2023

Agenda

About Myself

My Research

Some General Advice

Who am I?

- **Fourth year** Assistant Professor of Applied Statistics and Data Science
- Born in Laramie, Wyoming, grew up in Taiwan

- Got a B.S. in Mechanical Engineering, switched to Statistics in graduate school
- Got a Ph.D. in Statistics, 2017 at Purdue; did a SAMSI/CANSSI Postdoc before moving to Clemson

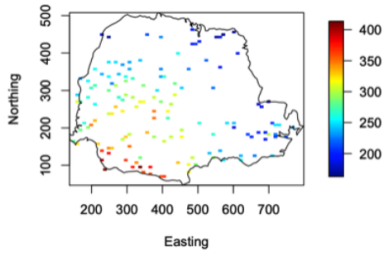

samsi
NSF Duke NCSU UNC

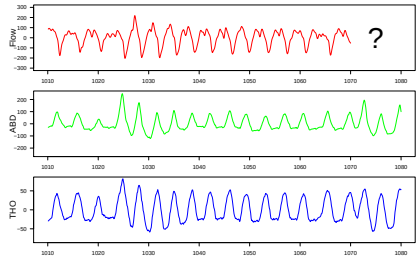


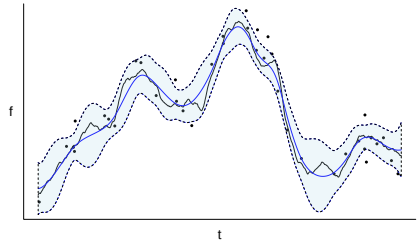
University of Victoria

Overview of My Research

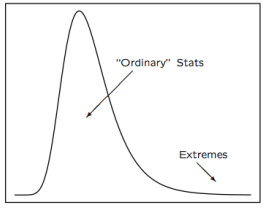
- ## Spatio-Temporal Statistics

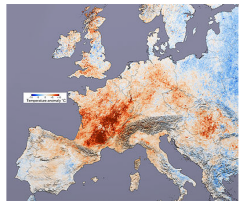


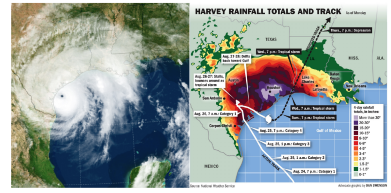




- ## Extreme Value Analysis



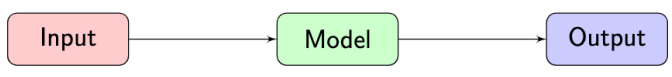


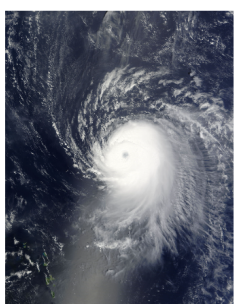


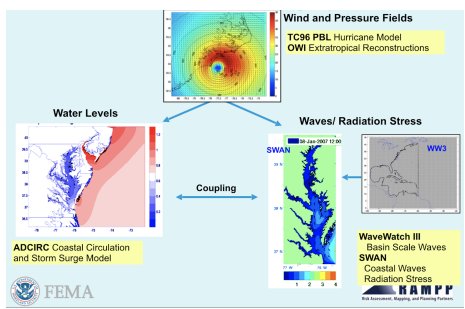
Source: NASA (Left); National Weather Service (Right)

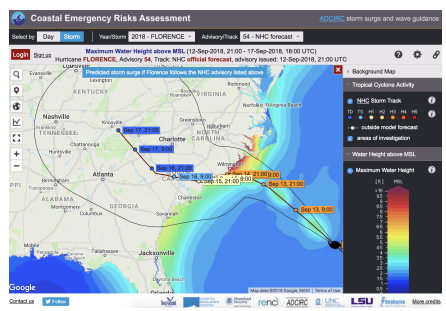
- ## Surrogate Modeling of Computer Experiments

$\boldsymbol{x} \in \mathcal{X}$ $\quad \eta : \mathcal{X} \mapsto \mathcal{Y}$ $\quad y = \eta(\boldsymbol{x})$

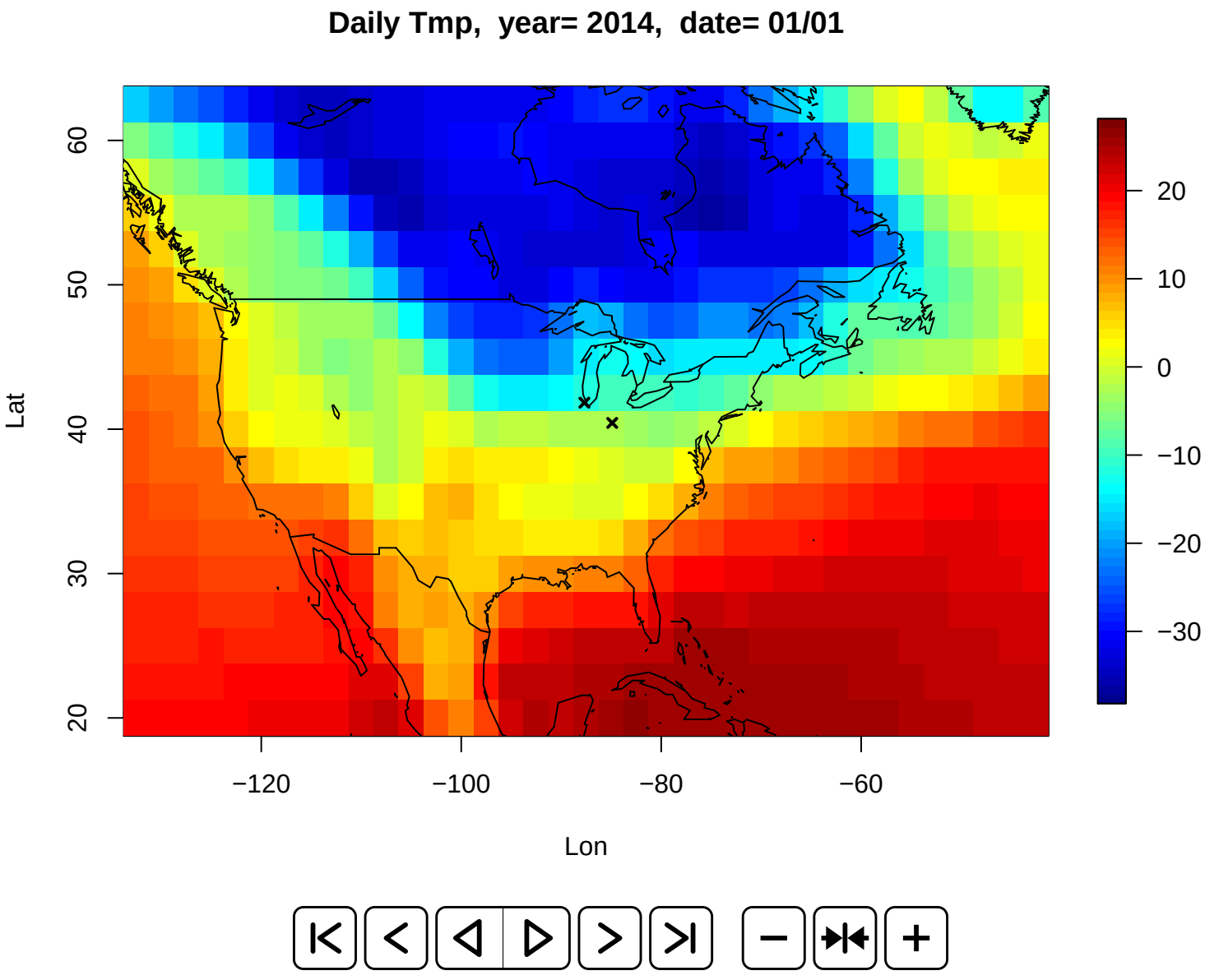






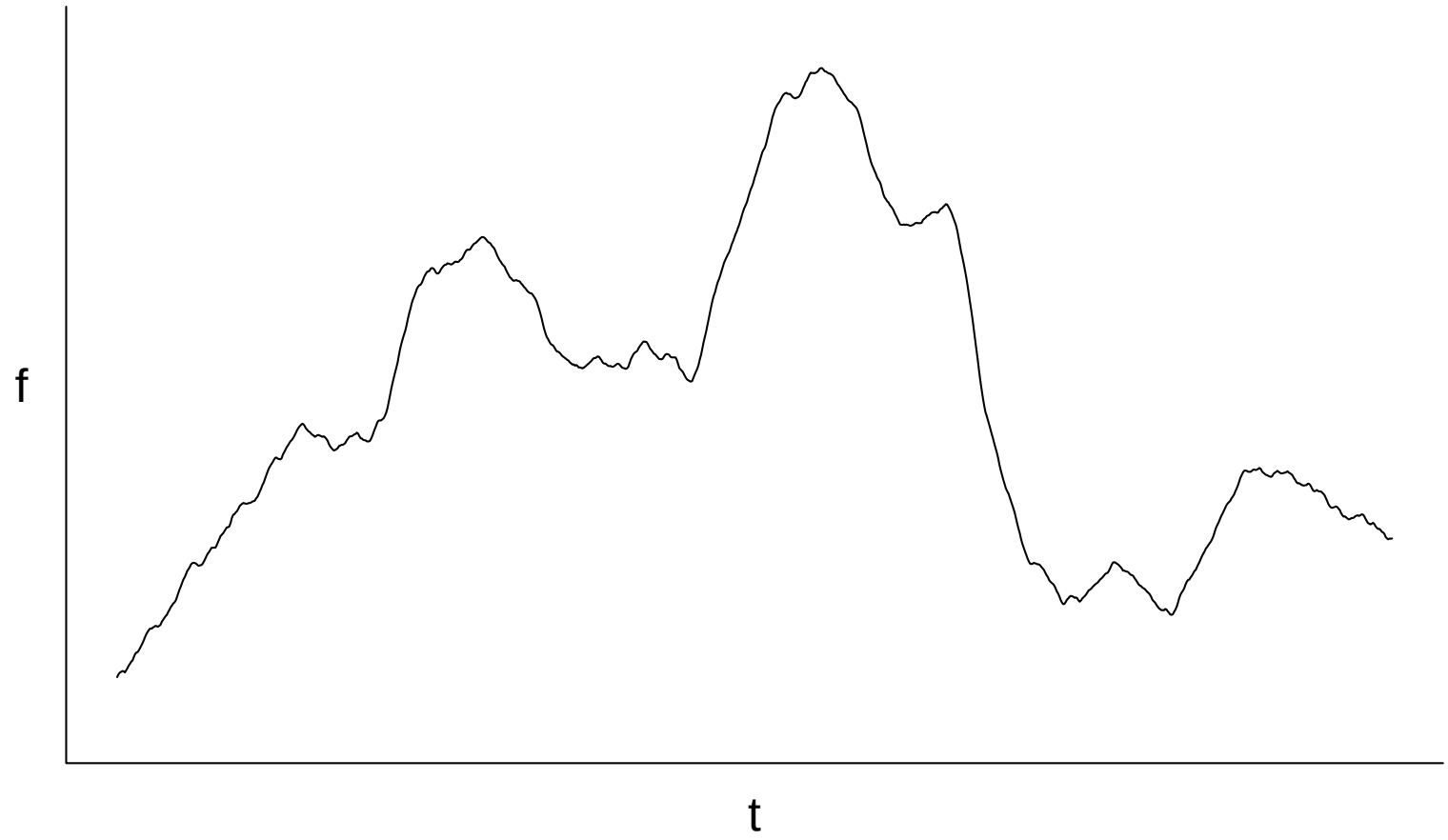


Spatio-Temporal Data



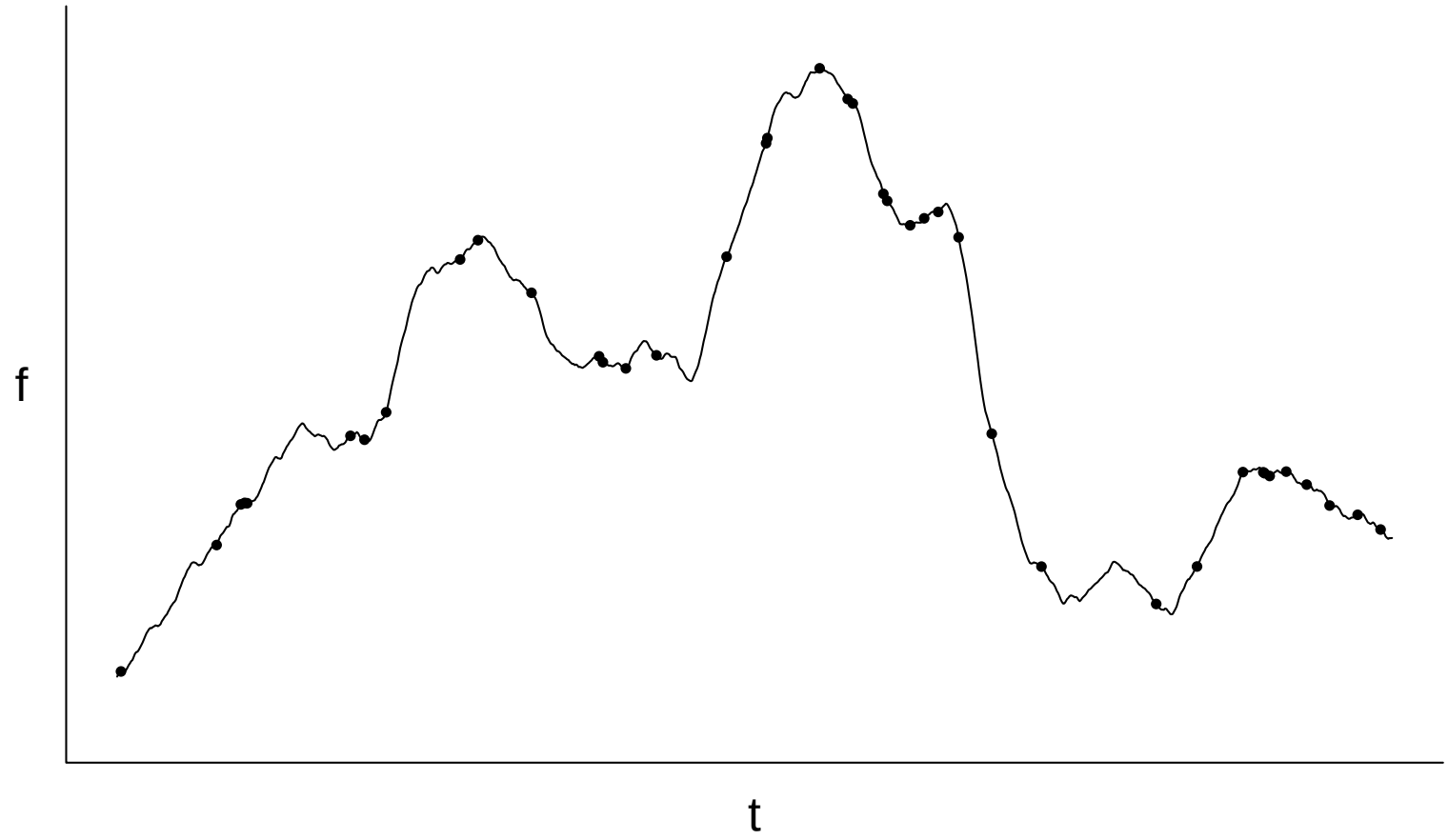
Gaussian Processes

Function Estimation



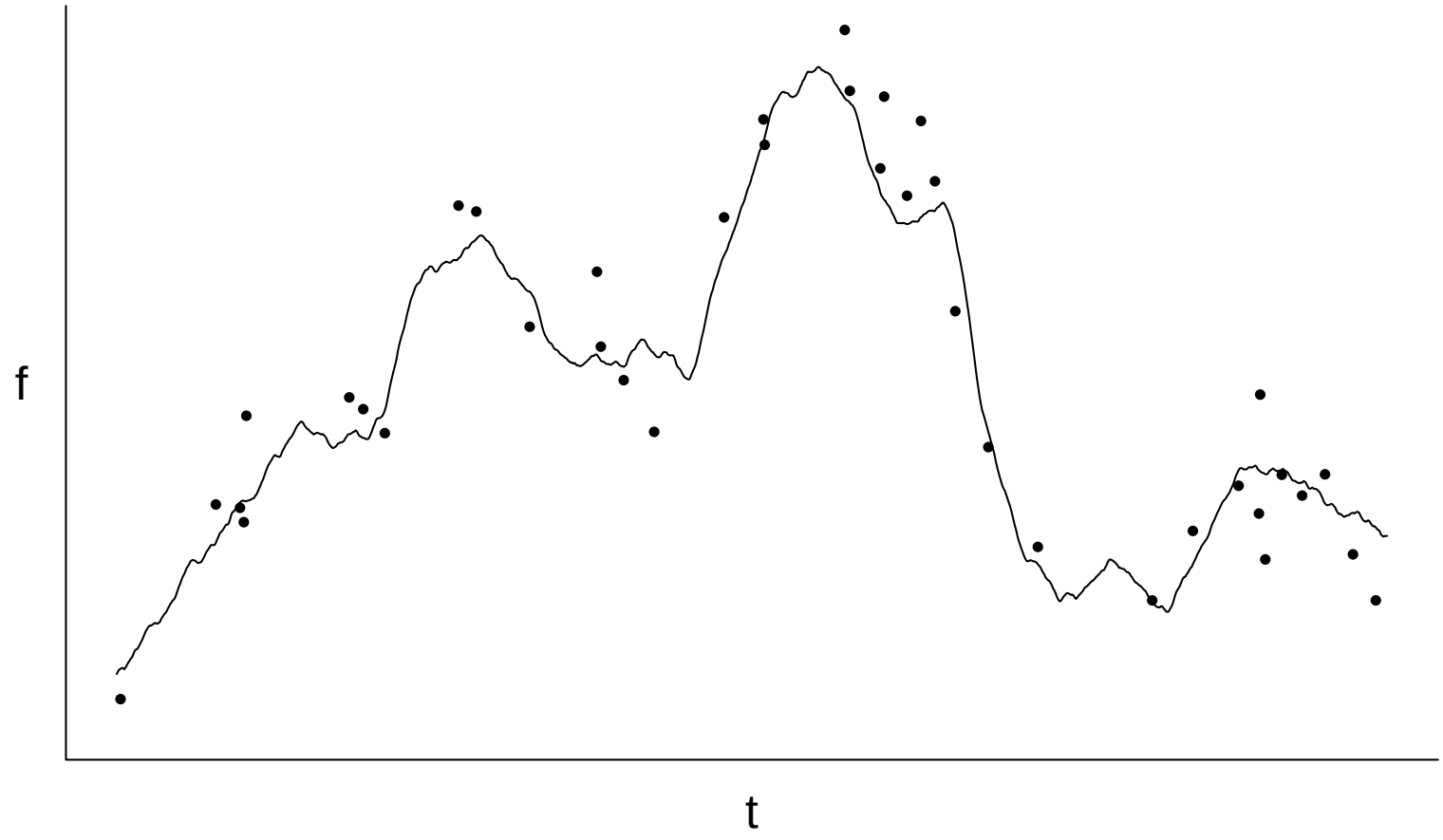
Consider a function $f(t), t \in \mathcal{T}$

Function Estimation



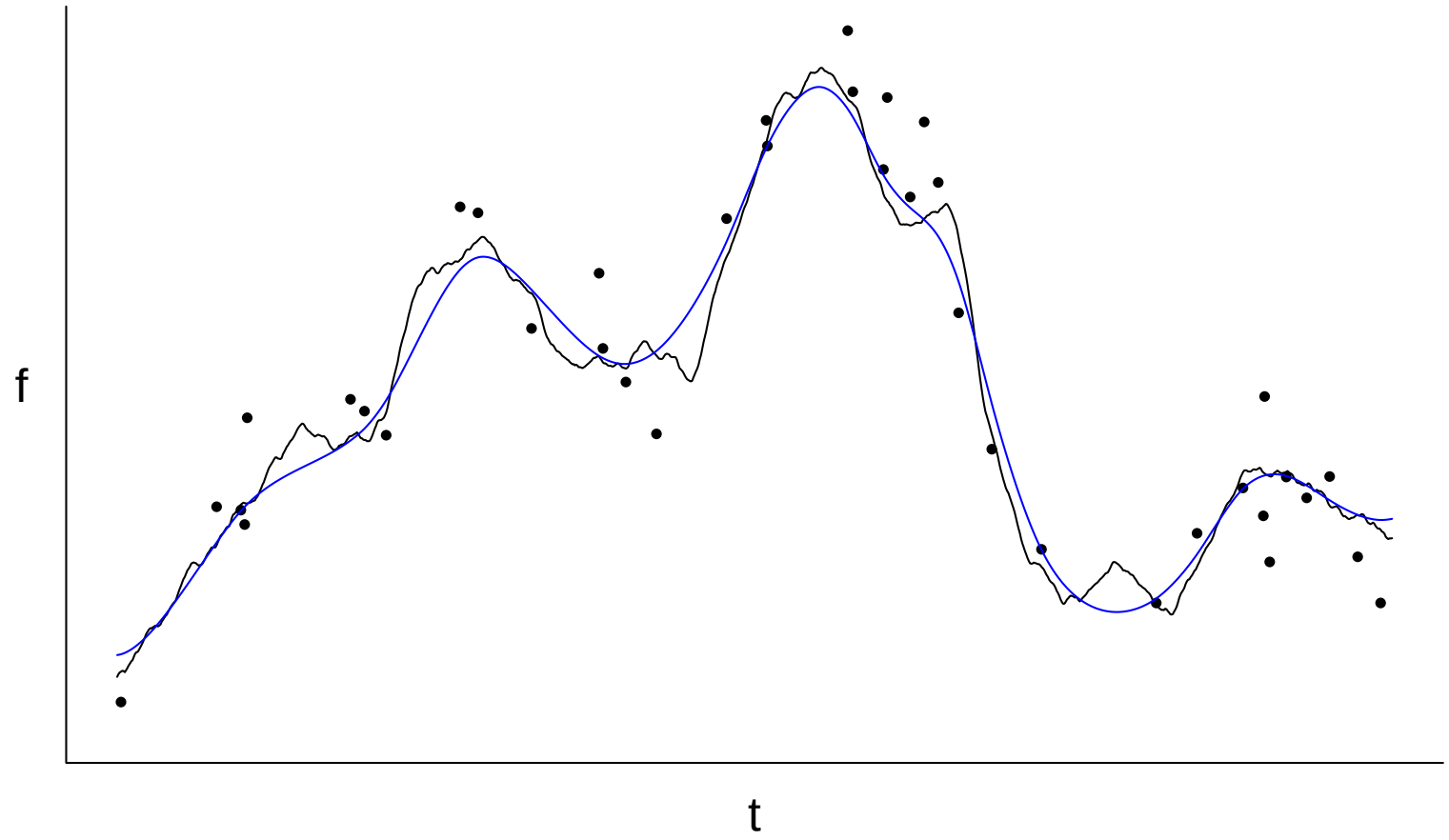
Consider $f(t)$, observed incompletely $\{f(t_i)\}_{i=1}^{n}$

Function Estimation



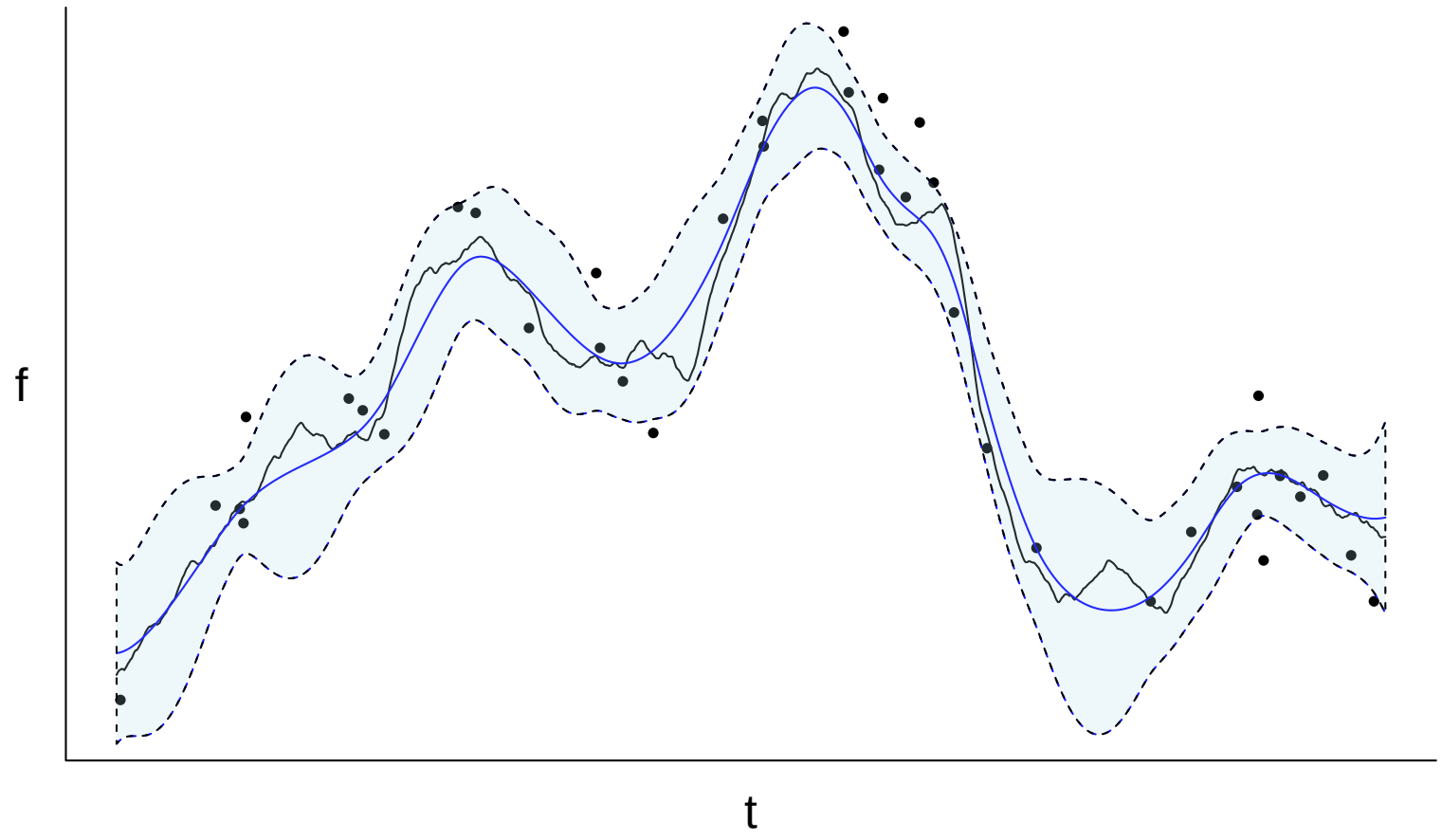
Consider $f(t)$, observed incompletely $\{f(t_i)\}_{i=1}^n$, and with noise $\{\varepsilon_i\}_{i=1}^n$

Gaussian Processes (GPs): Function Estimators



Main idea: exploit the inter-point correlation to estimate $f(t)$

GPs: Probabilistic Function Estimators



GP provides an "optimal" estimate $\hat{f}(t)$ along with "localized" uncertainty quantification ("error bars")

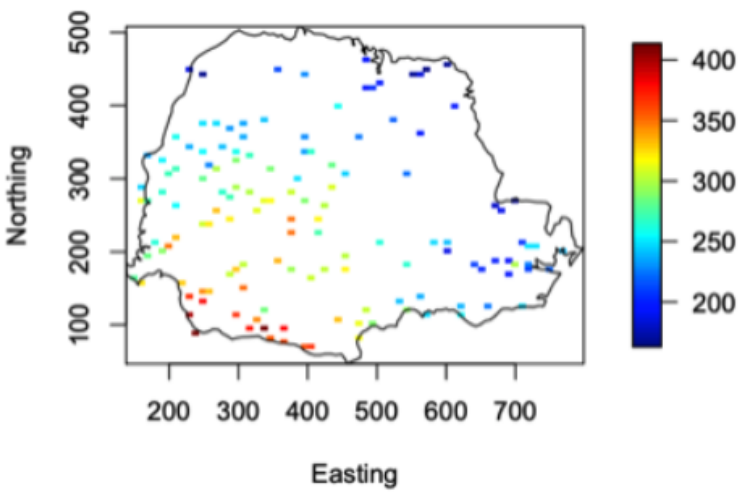
Spatial Interpolation via GP Model

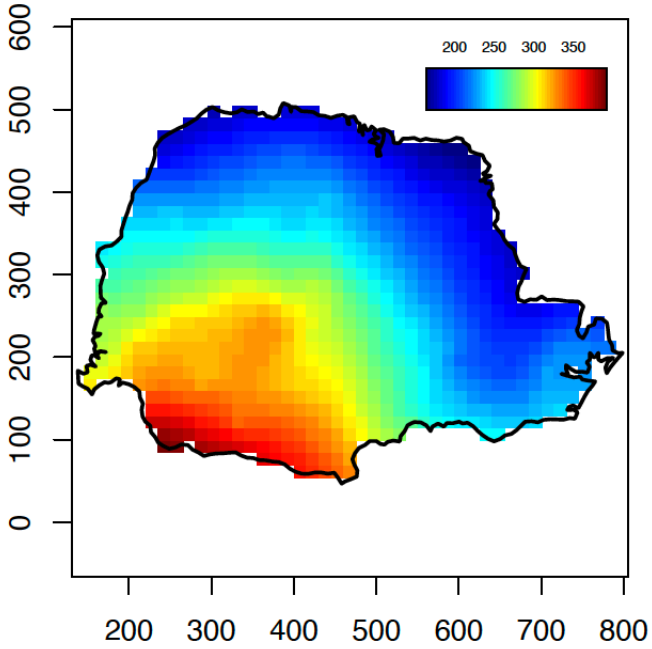
The "generic" spatial model

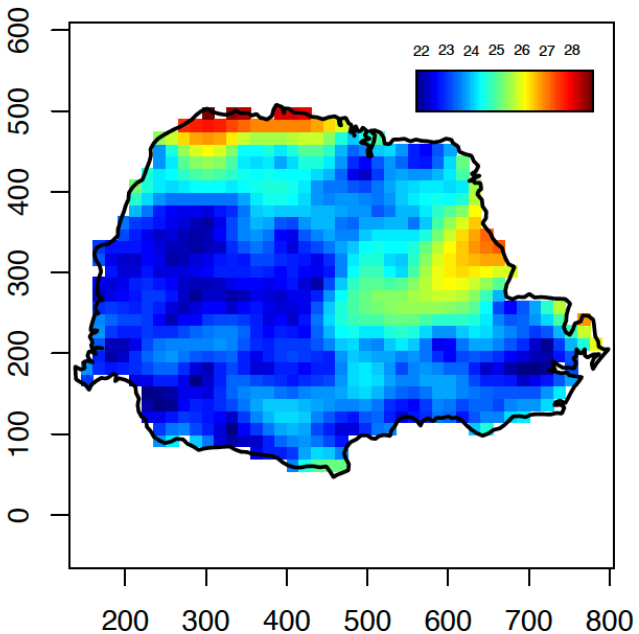
$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \eta(\boldsymbol{s}),\ \boldsymbol{s} \in \mathcal{S},$$

where

- $\mu(\boldsymbol{s})$ is the mean function, e.g., $\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}$
- $\eta(\boldsymbol{s})$ is a GP with mean zero and covaraince function $c_{\boldsymbol{\theta}}(h)$

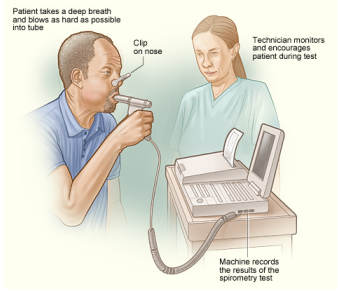


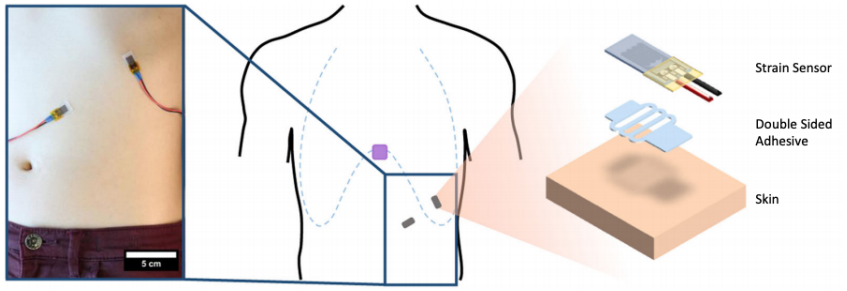


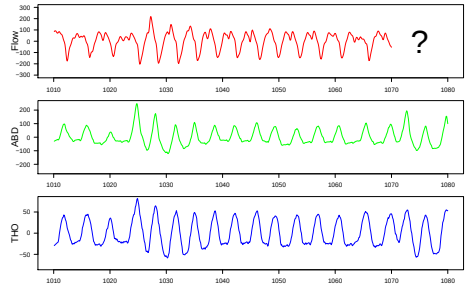


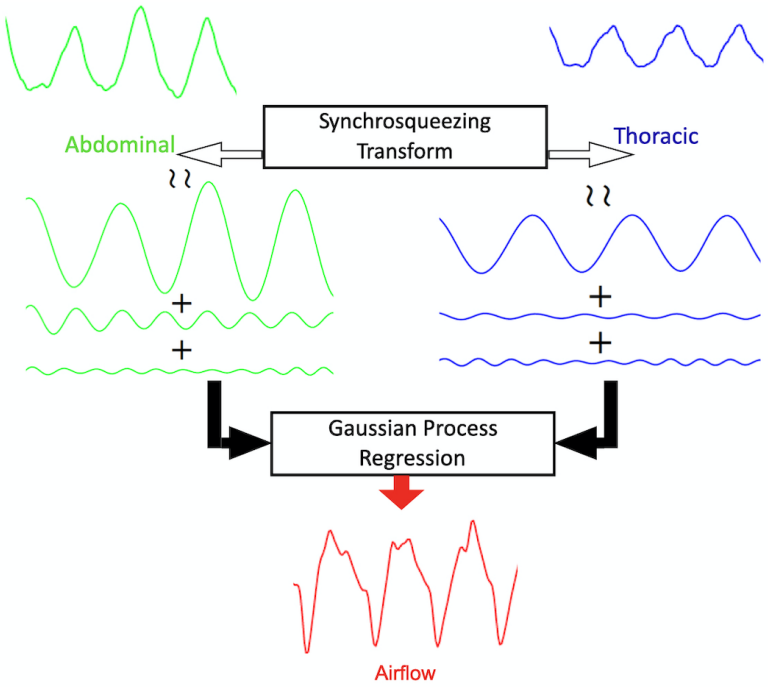
High-Frequency Physiological Waveform Signal Analysis

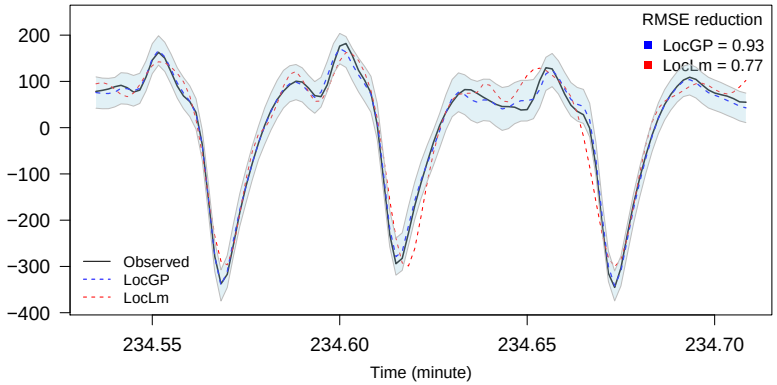
[joint with Chung (Eli Lilly), Wang (Clemson), Mandel (UPenn), and Wu (Duke)]

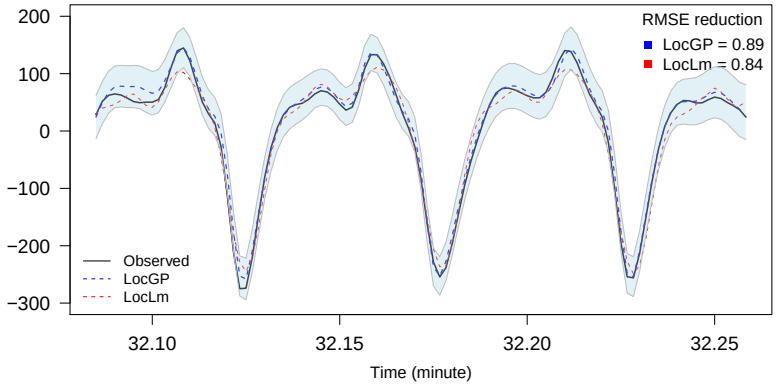




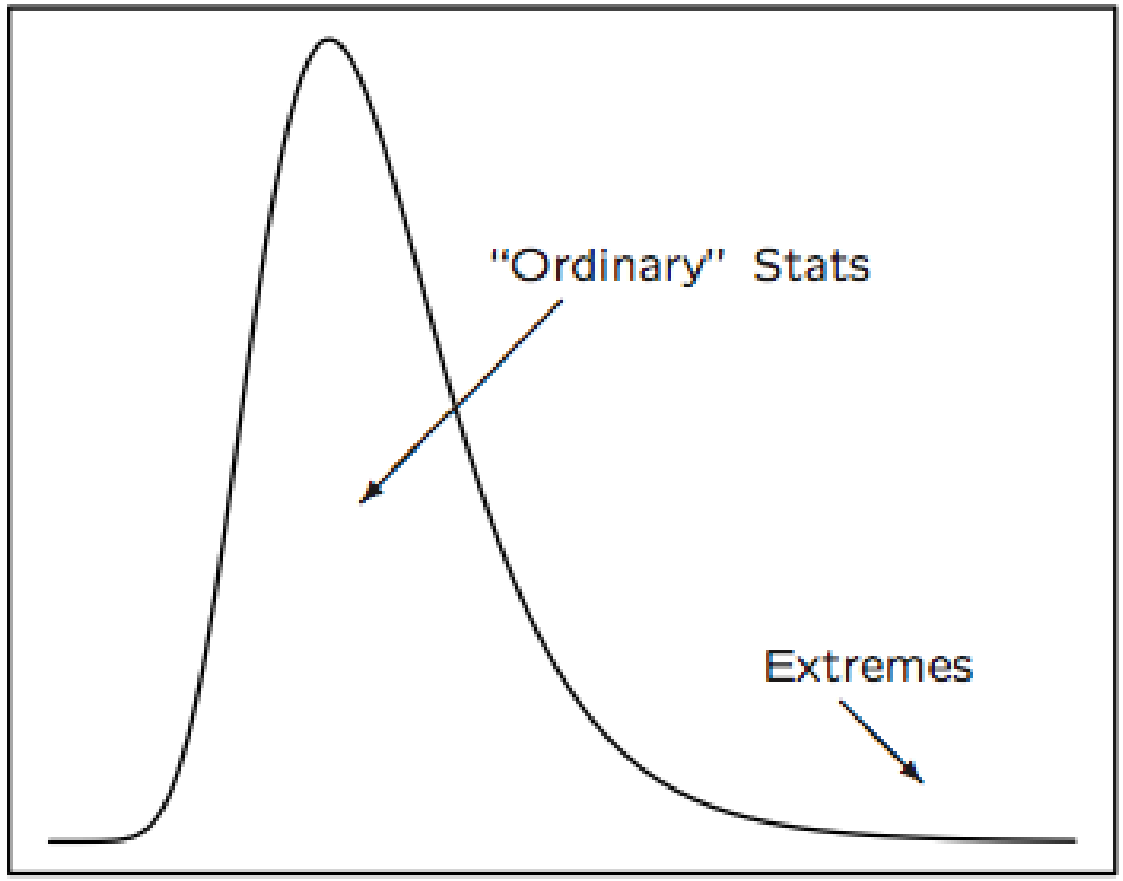






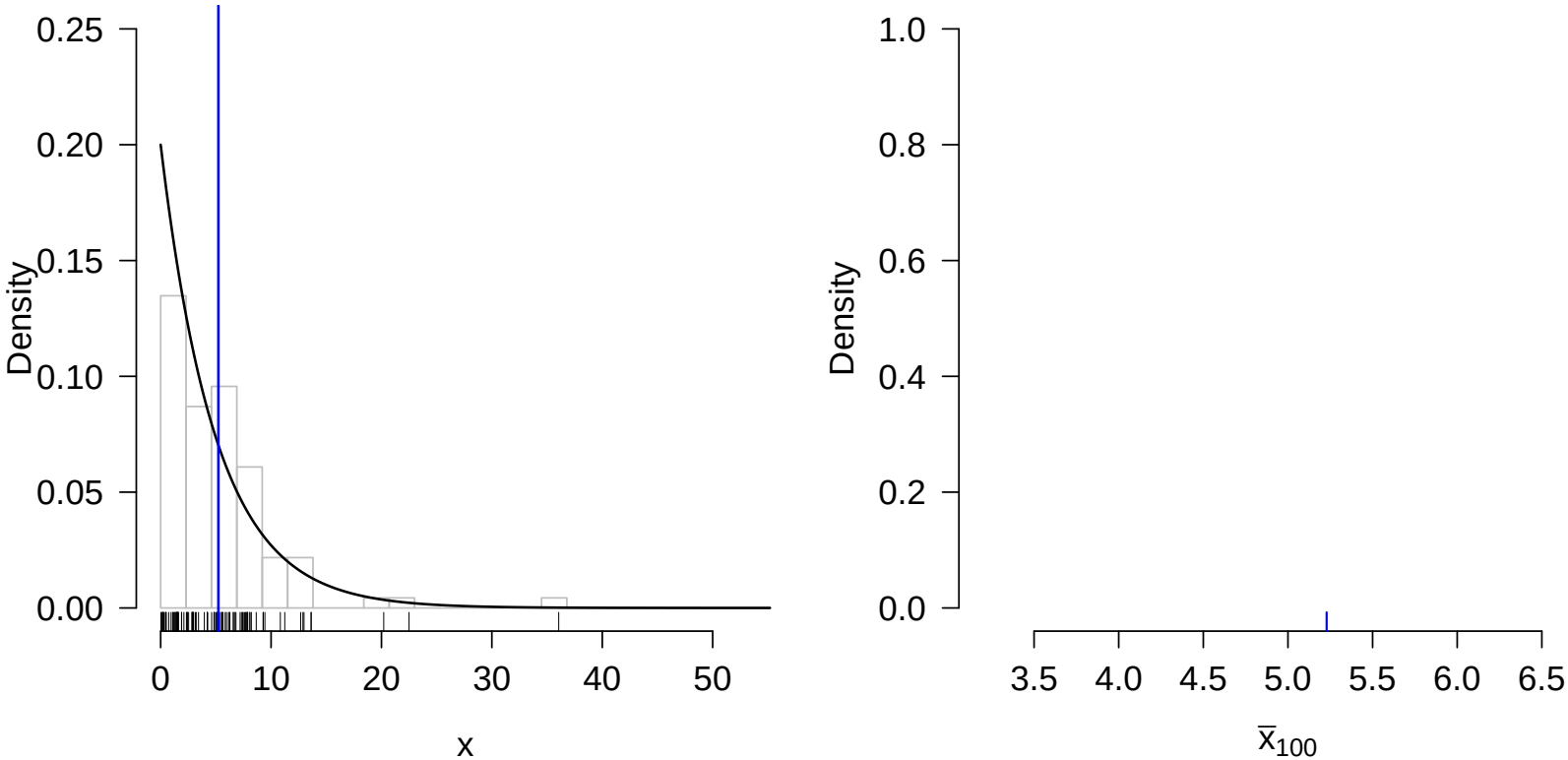


Extreme Value Analysis



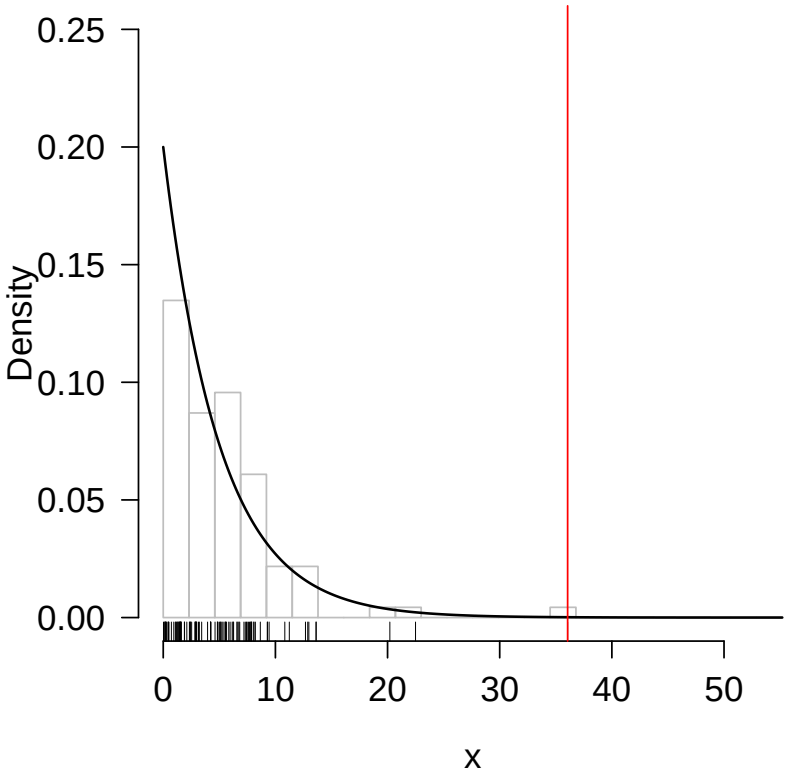
Central Limit Theorem Demonstration

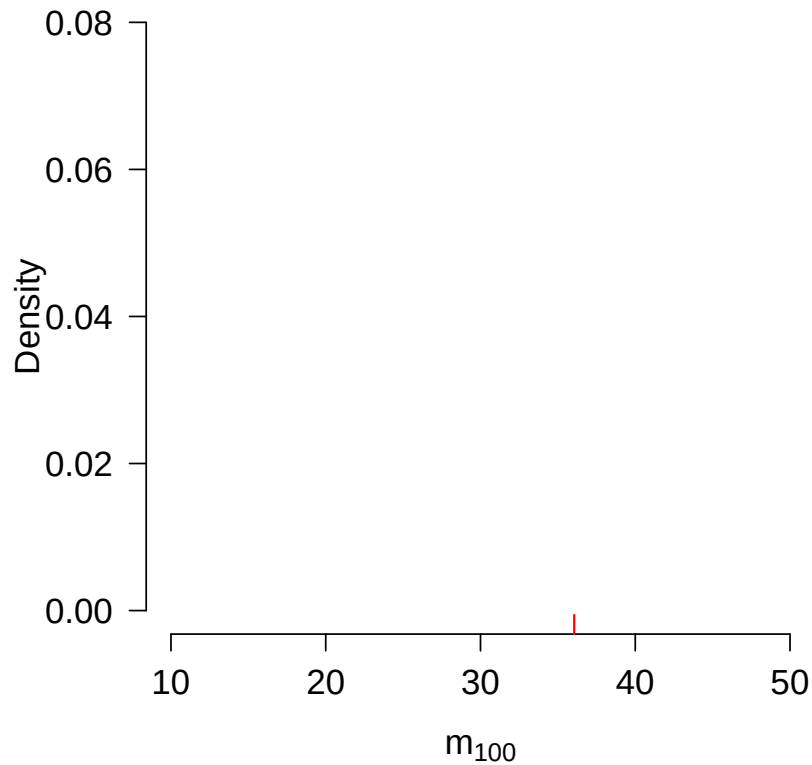
1. Generate 100 random numbers $(n = 100)$ from an Exponential distribution
2. Compute the sample mean of these 100 random numbers
3. Repeat this process 120 times

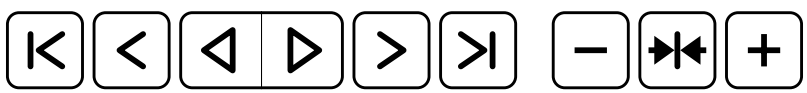


Demo: Distribution of the Sample Maximum

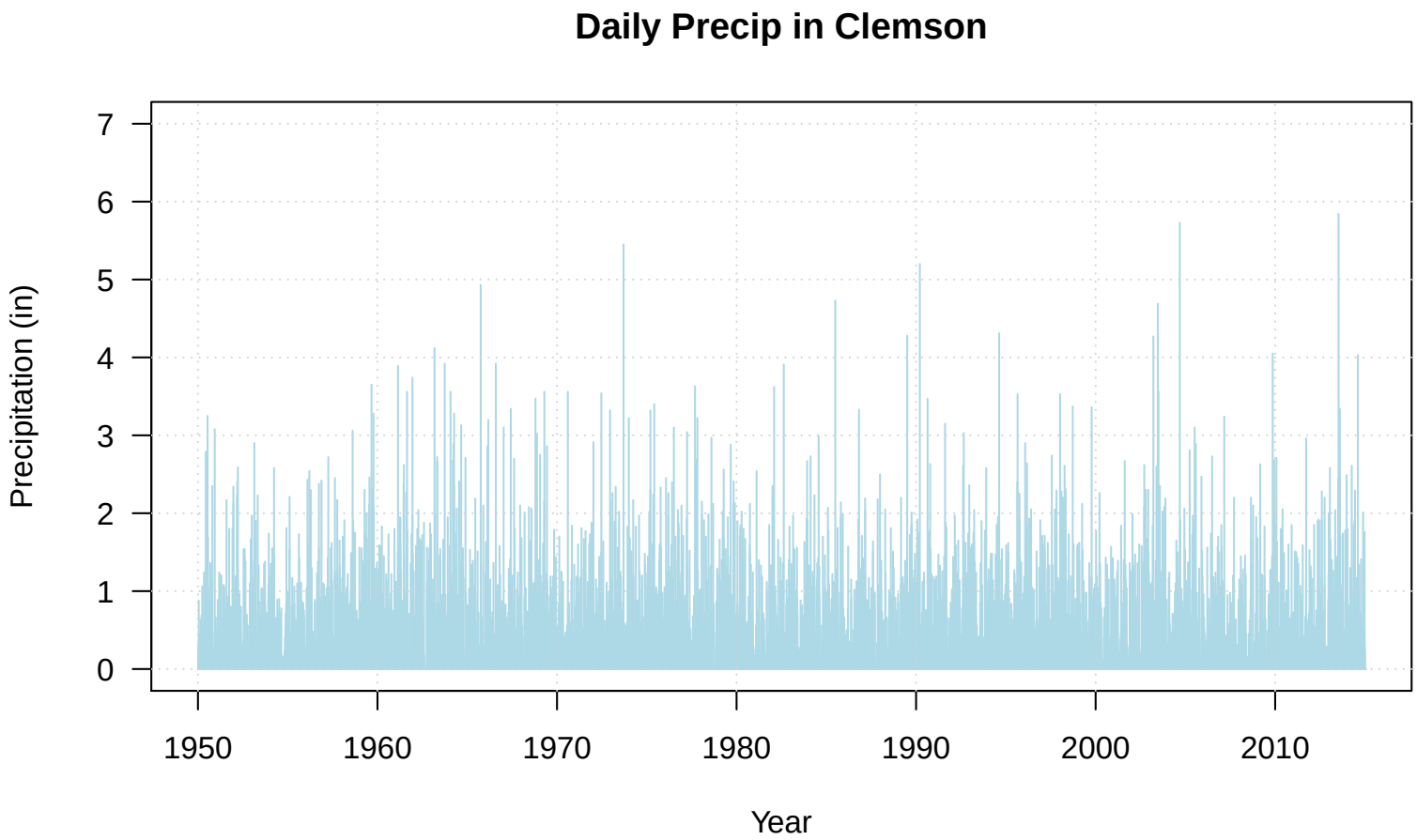
1. Generate 100 random numbers $(n = 100)$ from an Exponential distribution
2. Compute the sample maximum of these 100 random numbers
3. Repeat this process 120 times







Estimating Clemson[1] Daily Precipitation[2] Extremes



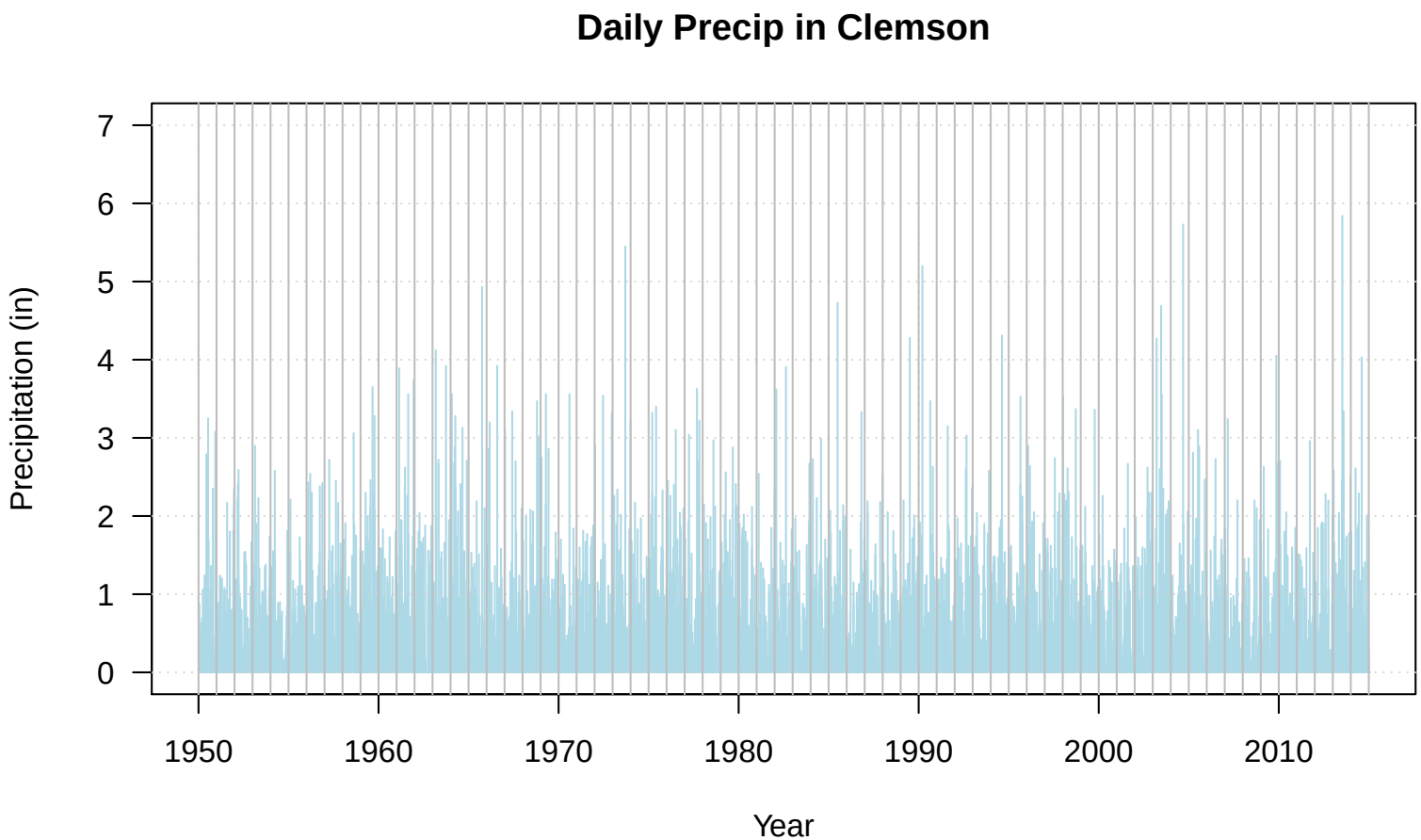
We may want to estimate, say the 50-year return level

[1] Near LaMaster dairy center

[2] Data Source: U.S. Historical Climatology Network (USHCN)

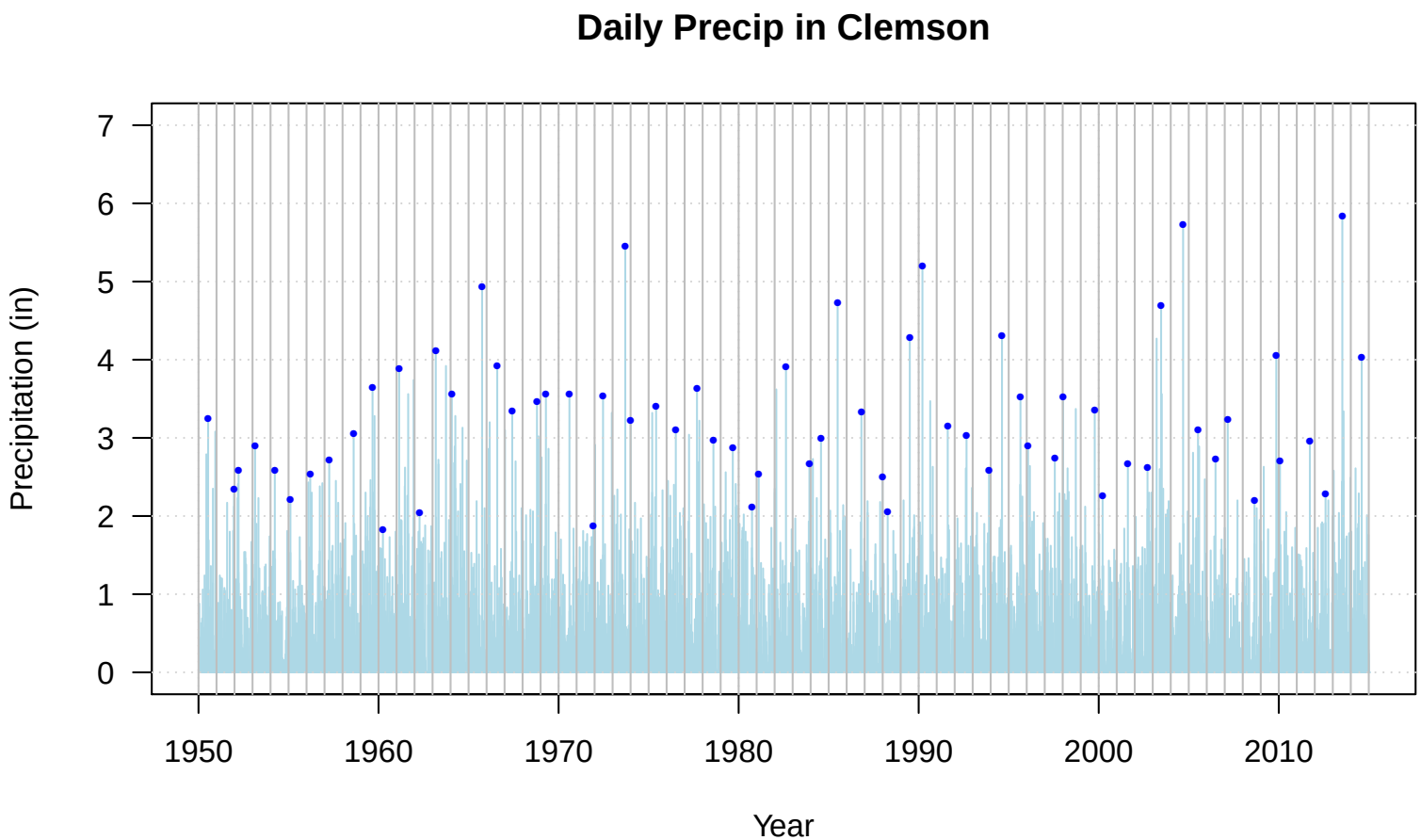
Block Maxima Method (Gumbel 1958)

1. Determine the block size and extract the block maxima



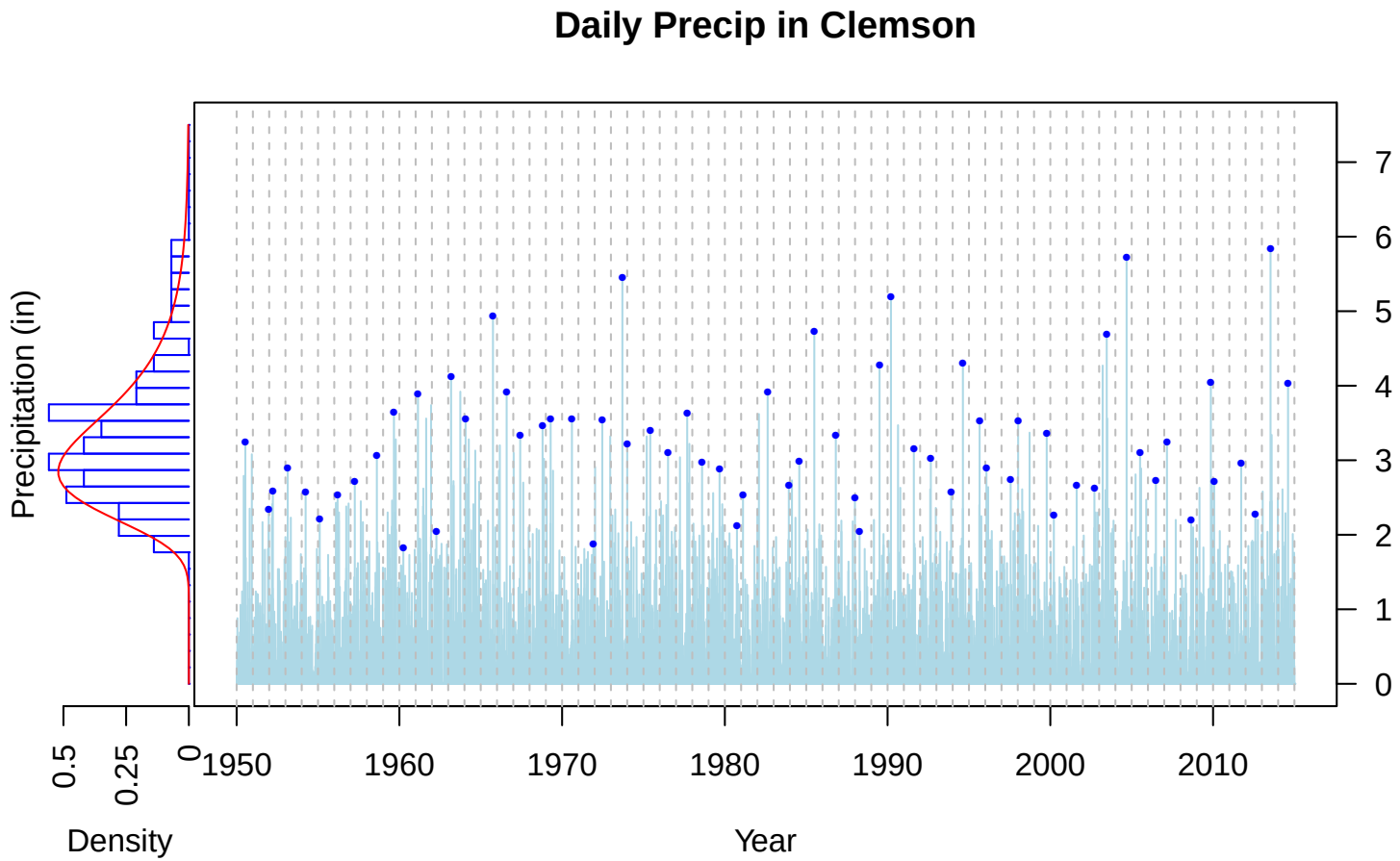
Block Maxima Method (Gumbel 1958)

1. Determine the block size and extract the block maxima



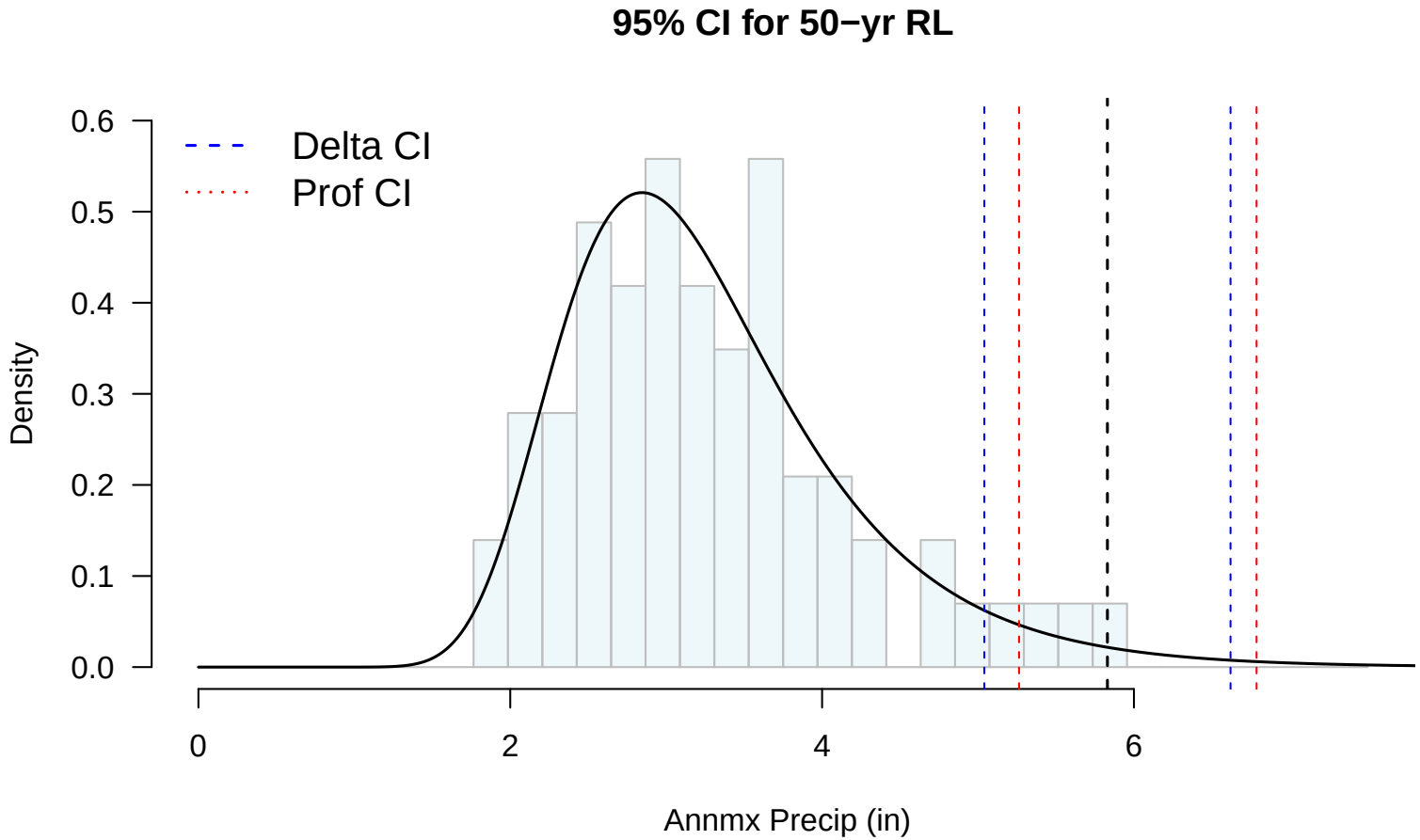
Block Maxima Method (Gumbel 1958)

2. Fit the Generalized extreme value (GEV) distribution to the maxima and assess the fit



Block Maxima Method (Gumbel 1958)

3. Perform inference for return levels or exceedance probabilitie

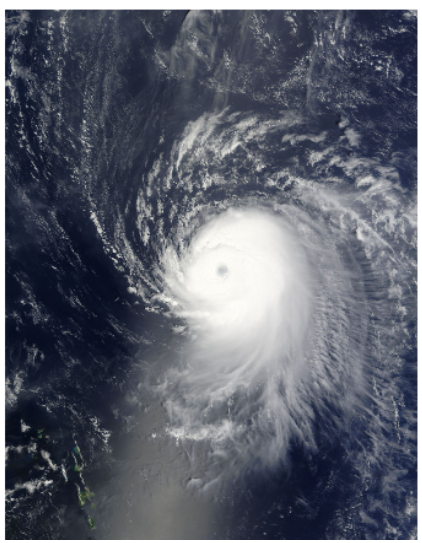


Point estimate: 5.68.

Interval estimate: $[4.63, 6.72]$ and $[4.98, 7.47]$

Some Research Questions

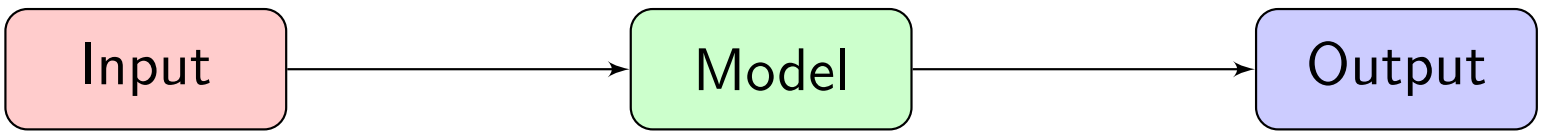
- How extremes vary in space? How extremes may change in future climate conditions?
- How to model extremes when the process of interest involves several variables (i.e., compound extremes)?
- How to combine data from different sources to infer extremes?
- How to leverage physical knowledge to better model extremes?



Estimating Extreme Surges: Physical-Statistical Approach

- **Goal**: Estimate high quantile of storm surge in coastal region
- Analysis based on historical storm surge observations alone can be unreliable due to limited data in space and time
- We take a physical-statistical approach:

$\boldsymbol{x} \in \mathcal{X}$ $\quad \eta : \mathcal{X} \mapsto \mathcal{Y}$ $\quad y = \eta(\boldsymbol{x})$



TC Characteristics

- Records are more complete than surge levels
- Input to simulate storm surge levels

Task: Estimating $f(\boldsymbol{x})$

Computer Model

- Simulate high fidelity surge response
- computationally extensive

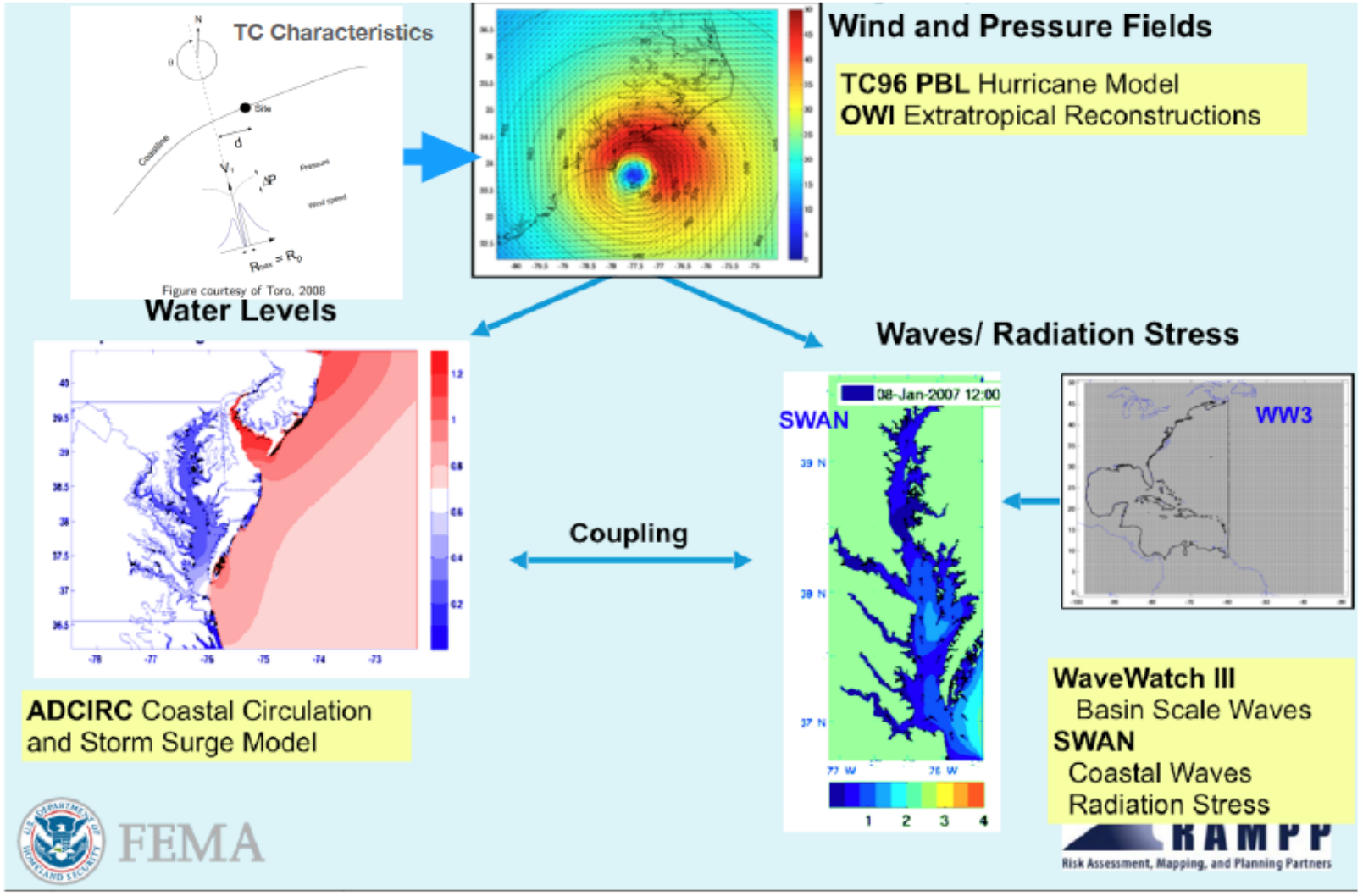
Task: Estimating $\eta(\boldsymbol{x})$

Surge Level

- simulate synthetic storms
- generate surge response for risk analysis

Task: Estimating y_r

Computer Model Linking Input to Output $\eta : x \mapsto y$

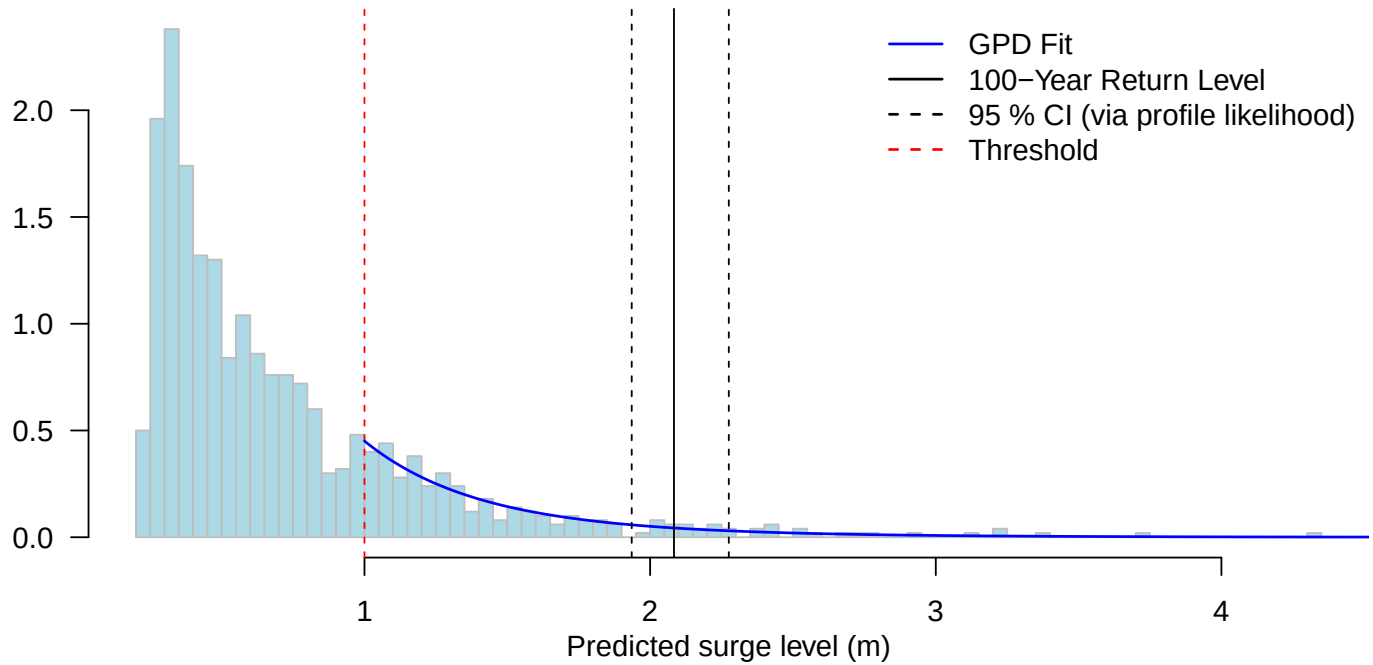


Courtesy of Gangai (Dewberry) & Danforth (FEMA)

Estimating Extreme Surges: Extreme Value Analysis[3]

We employed the peaks-over-threshold method [Davison and Smith, 1990] to estimate the r-year return levels

- Assuming upper tail follow a generalized Pareto distribution (GPD)
- Using profile likelihood method to construct confidence interval (CI), which gives asymmetric interval



[3] There will be a 3-day NSF workshop on Extreme Value Analysis this summer organized by myself and Dr. Russell

My Research Group Members

- Eva Murphy, PhD, "Estimating Spatio-Temporal Variation of Wind Speed and Wind Direction"
- Kanon Kamronnaher, PhD, "Estimating Financial and Environmental Risk"
- Jiyun Huang, PhD (joint with Dr. Russell), "Intensity-Duration-Frequency Curve Estimation"
- Katherine Kreuser, PhD, "Uncertainty Quantification for Environmental Extremes"
- Andrew Bellucco, MS, "*Estimating Financial Risk*" (joint with Dr. Gallagher, Graduated Dec. 2019. Current position: Recommendation Analytics at Credit Karma, Charlotte, NC
- Emily Tidwell, MS, "*A Combined Physical-Statistical Approach for Estimating Storm Surge Risk*" Graduated May 2021. Current position: Dynetics, Huntsville, AL
- Adam Diaz, MS, "Contributions to Statistical Analysis of Wildfires" Graduated May 2022. Current position: Principal Research Statistician at Northern California Institute for Research and Education, San Francisco, CA

Reading/Working Groups

Currently ruining a reading group on GP/EVA where we meet every Friday 11:15am - 12:15pm at Watt Center 323

Past reading groups:

- Spatio-Temporal Statistics: joint with Drs. Russell, Ma, Ranganathan, and Li

  Link:
  https://whitneyhuang83.github.io/ENVR/ClemsonENVR.html

- Uncertainty Quantification (UQ): joint with Drs. Brown and Zhang, Ma

  Link:
  https://whitneyhuang83.github.io/UQ/ClemsonUQ.html

Some General Advice

Being a Graduate Student

- "How to Succeed in Graduate School" by Marie desJardins
  http://www.ai.sri.com/~marie/papers/advice-summary.html
- "Notes on the PhD Degree" by Douglas Comer
  https://www.cs.purdue.edu/homes/dec/essay.phd.html
- "Writing and Presenting your Thesis or Dissertation" by S. Joseph Levine
  http://www.learnerassociates.net/dissthes/

Some Useful Computing Skills

- R (https://www.r-project.org), RStudio (https://www.rstudio.com), and R markdown (http://rmarkdown.rstudio.com)
- LaTeXand Overleaf (https://www.overleaf.com/project)
- Python and jupyter (https://jupyter.org/)
- GitHub

Attend Seminars!

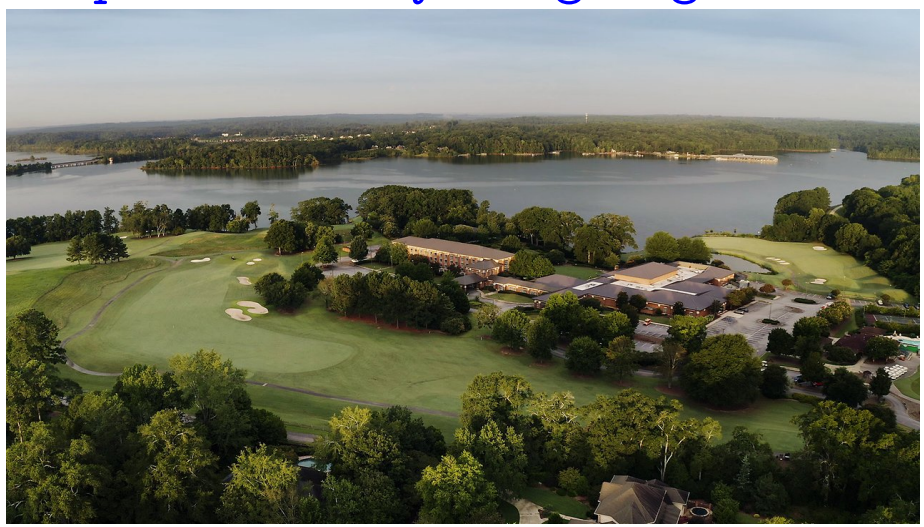
- Graduate Student Seminars
  http://siam.people.clemson.edu/gss/schedule.php
- Research Seminars
- School Colloquia
- Many One World Seminar Series

Attend Workshops!

- NSF-CBMS Regional Research Conferences in the Mathematical Sciences
  https://www.cbmsweb.org/regional-conferences/

- National Science Foundation (NSF) funded mathematical sciences institutes https://mathinstitutes.org/

- Interdisciplinary Workshop on Weather and Climate Extremes
  https://whitneyhuang83.github.io/WCE2023



Attend Conferences!

- Joint Mathematics Meetings (JMM): January
- Society for Industrial and Applied Mathematics (SIAM) Annual Meeting: July
- Institute for Operations Research and the Management Sciences (INFORMS) Annual Meeting: October or November
- Joint Statistical Meetings (JSM): Late July or early August
- Neural Information Processing Systems (NIPS) Conference: December

How to React Me?

- **Websites** 🖱: https://whitneyhuang83.github.io/
- **Email** ✉: wkhuang@clemson.edu
- **Office**: O-221 Martin Hall

Go Tigers!