Kickoff Meeting of Clemson Environmental Statistics Group

Whitney Huang

CLEMSON UNIVERSITY

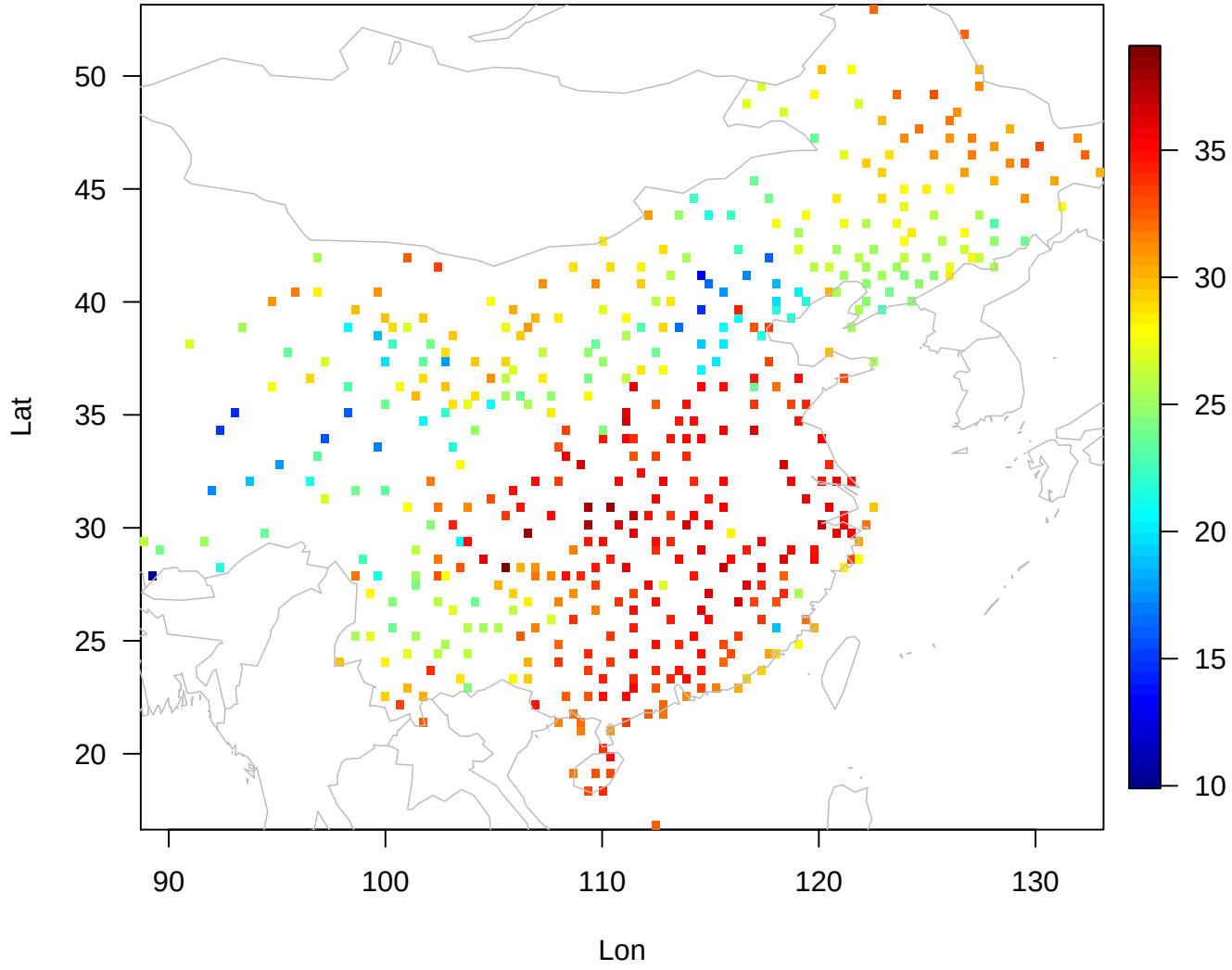
wkhuang@clemson.edu

Clemson ENVR Group, September 2, 2020

Possible Topics

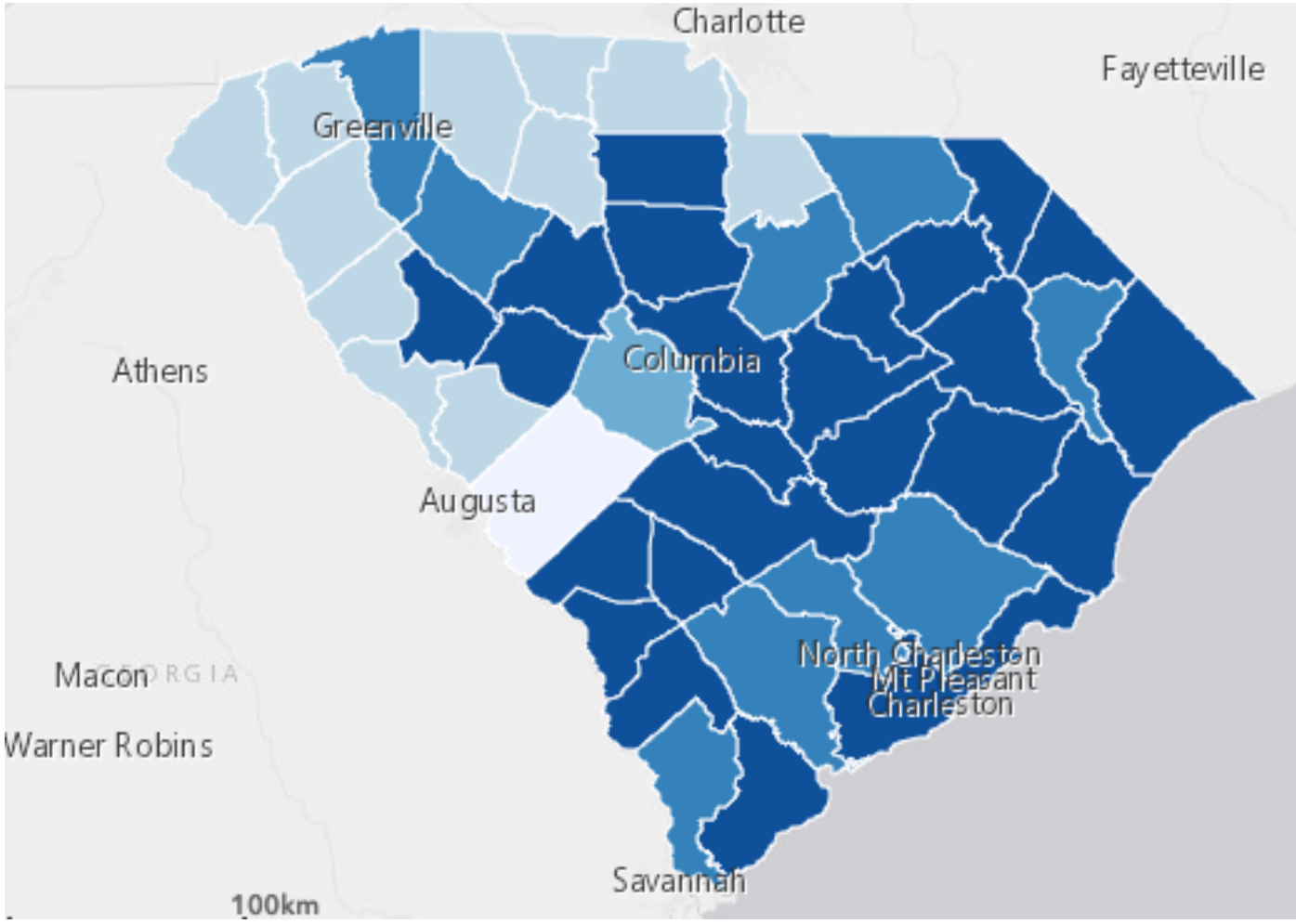
- Spatial interpolation with Gaussian processes
- Markov Random Fields (GMRF) for modeling discrete spatial variation
- Some selected topics in space-time data analysis
- Extreme value analysis
- Quantile regression
- Functional data analysis (FDA) and topological data analysis (TDA)
- Copula

Spatial interpolation



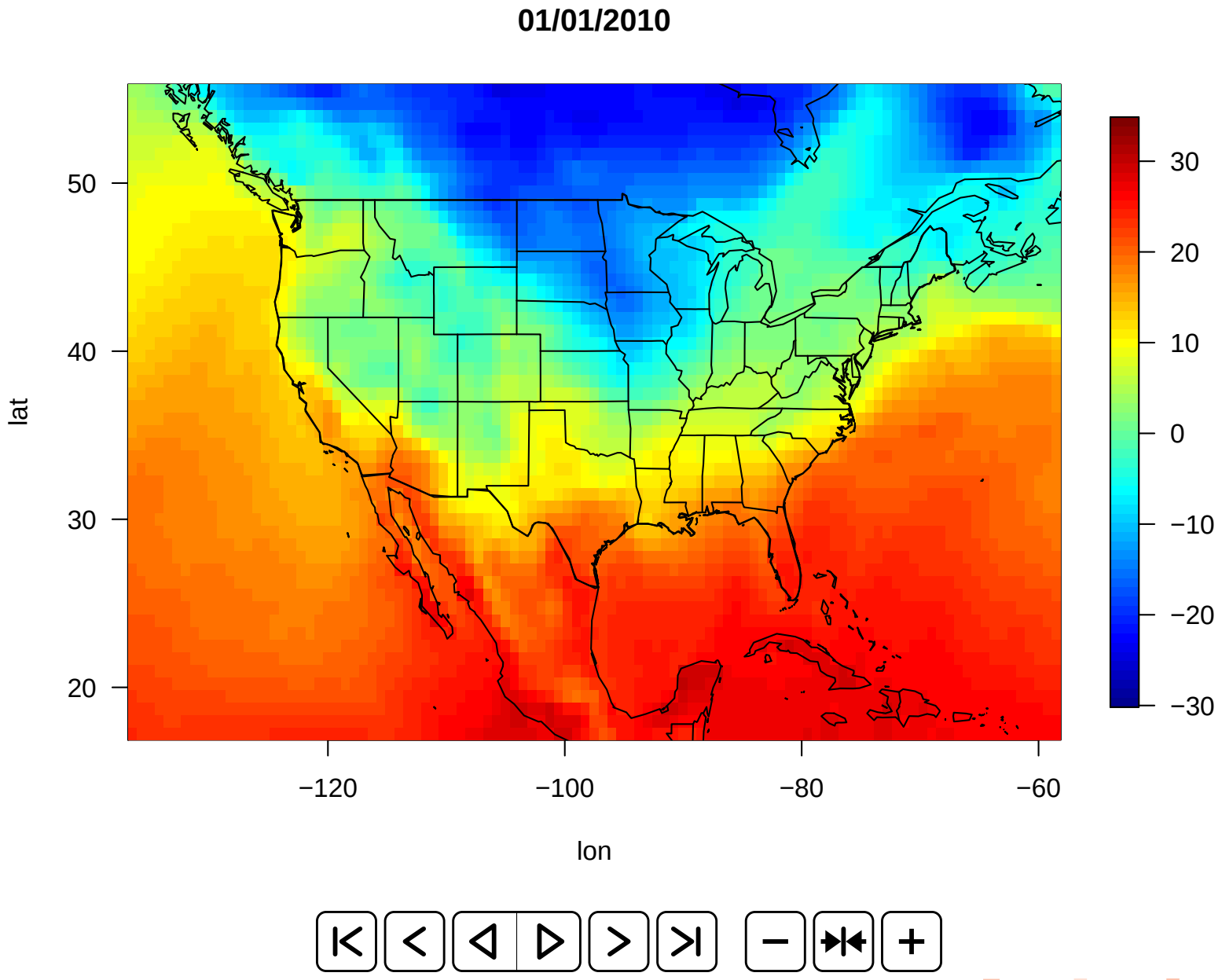
Goal: To interpolate the values in the spatial domain

Disease Mapping

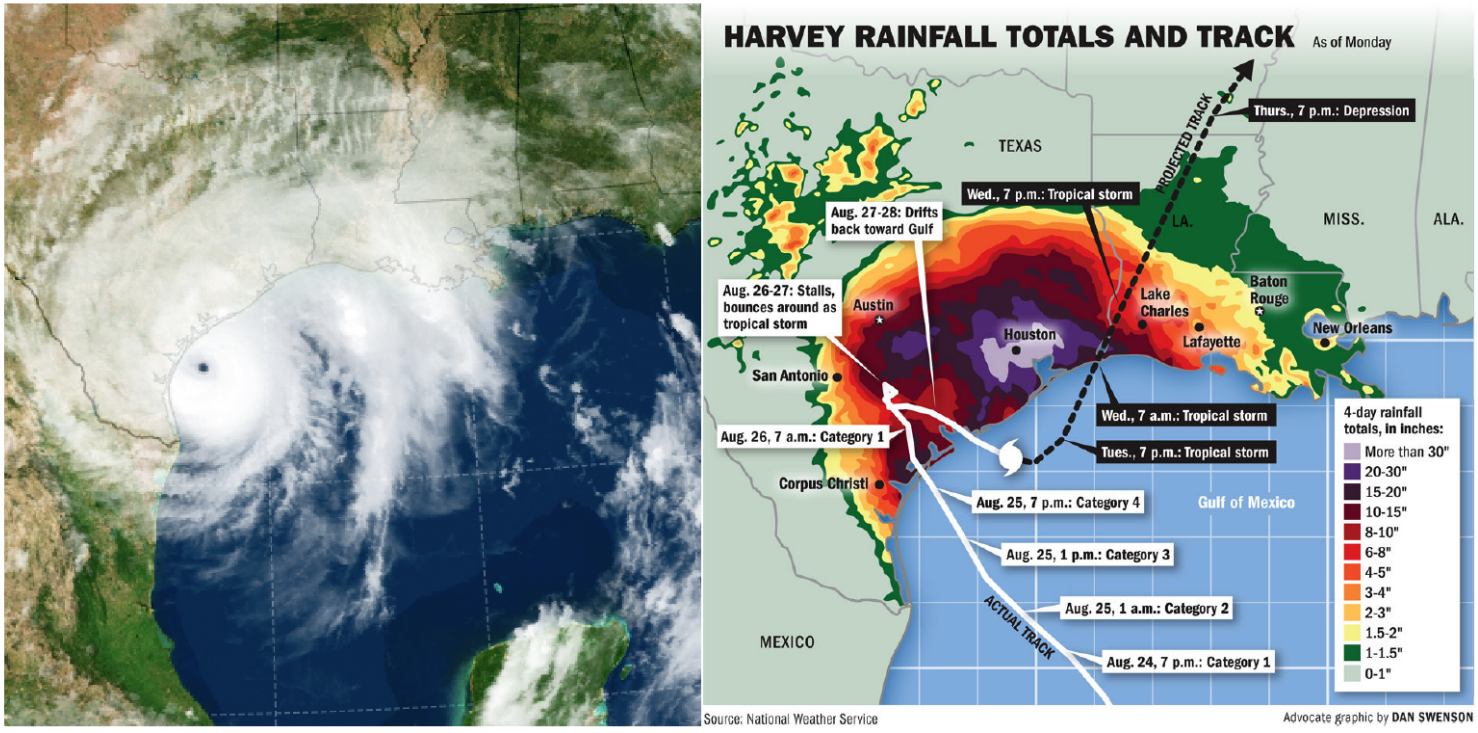


Goal: To estimate local disease risk

Modeling Space-Time Process



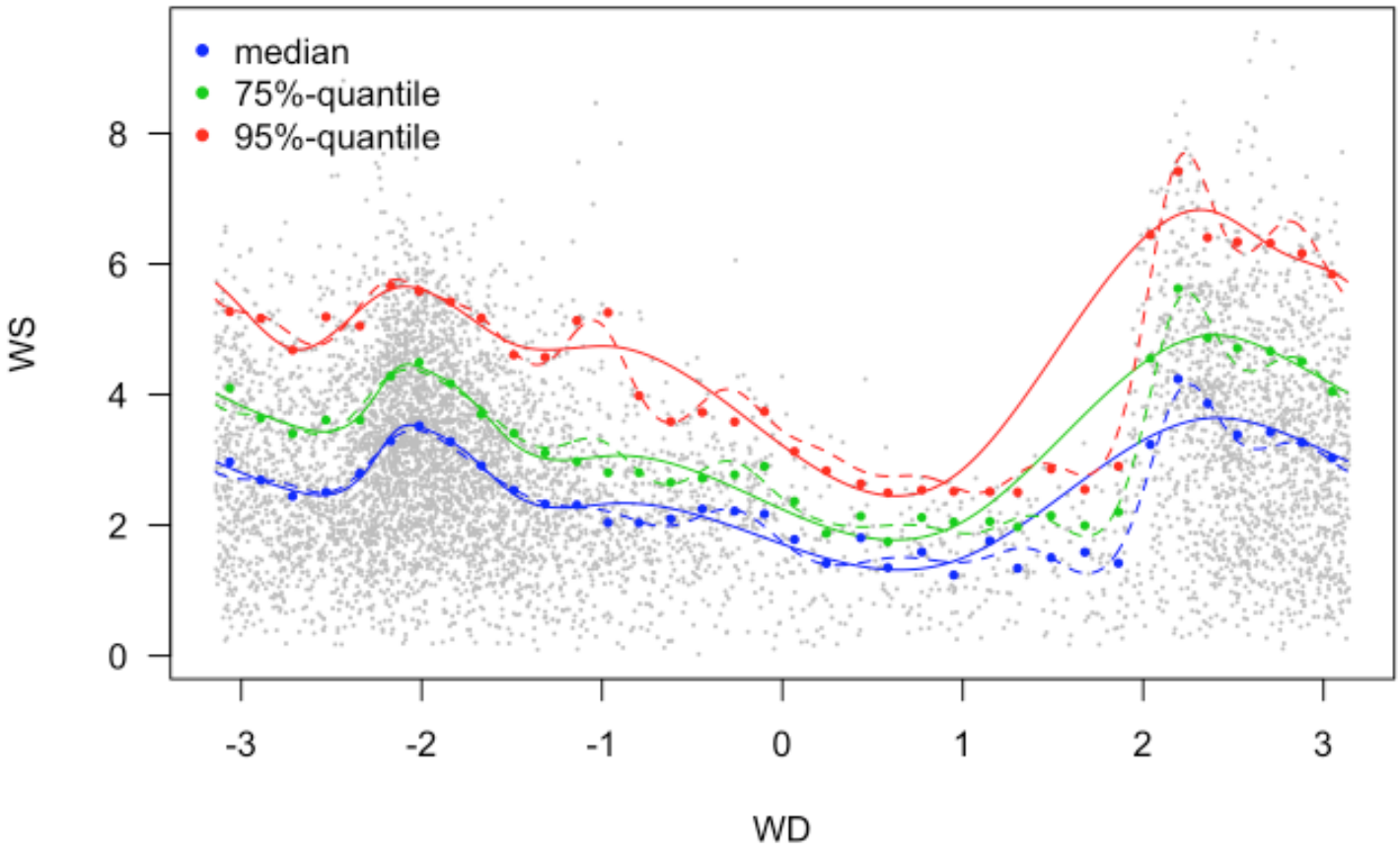
Extreme Value Analysis



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

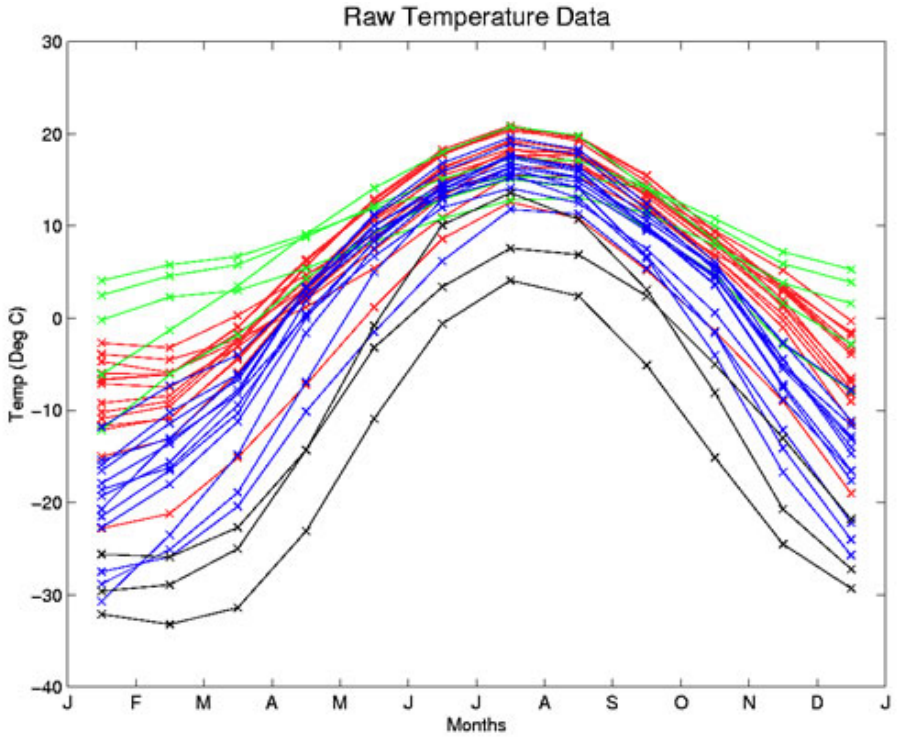
Goal: To estimate the magnitude of extreme event (e.g., 100-year flood)

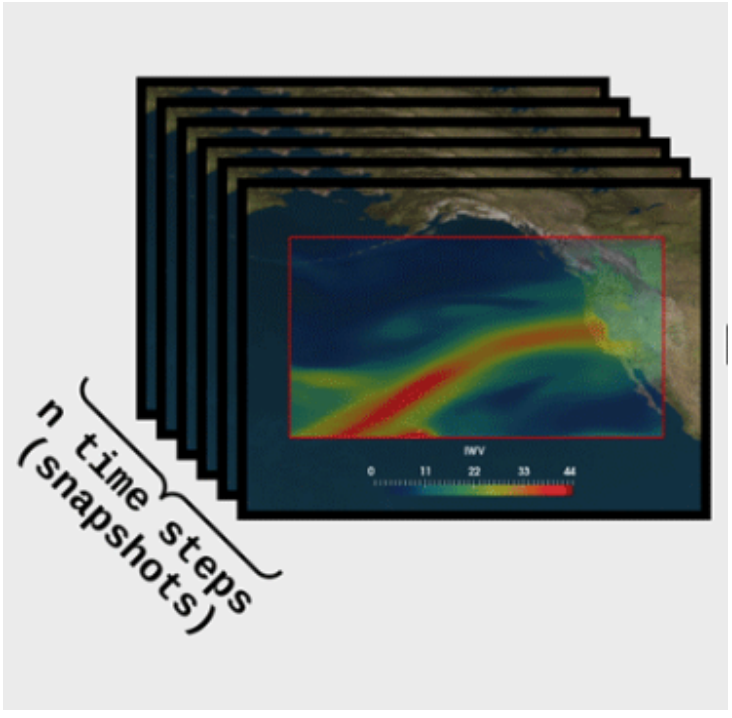
Quantile regression



Goal: To estimate conditional quantile curves

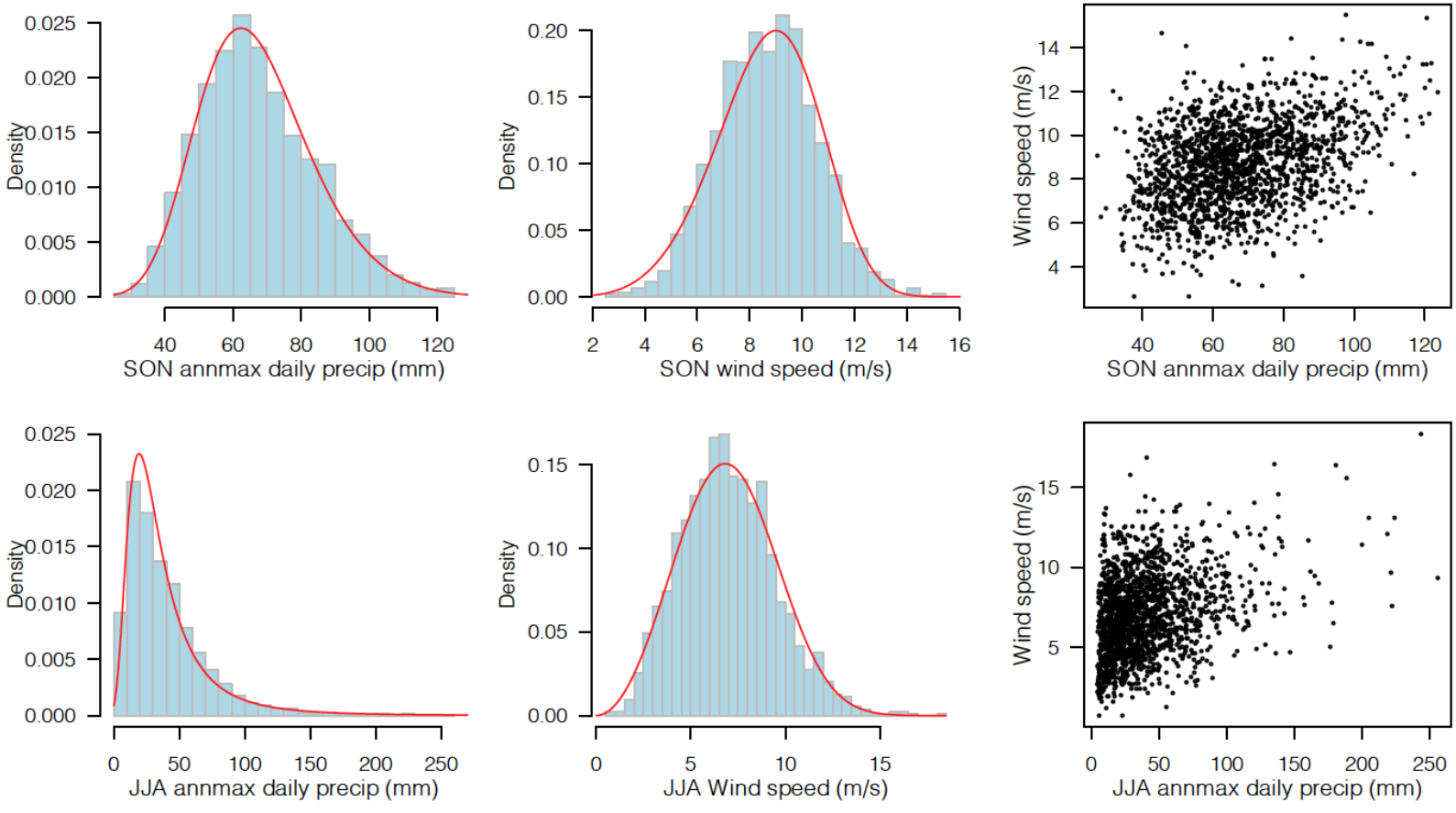
Functional and topological data analysis (FDA/TDA)





Source: `http://ego.psych.mcgill.ca/misc/fda/ex-weather-a1.html`
(Left), "Topological data analysis and machine learning for recognizing atmospheric river patterns in large climate datasets", Muszynski et al 2019
Right

Dependence Modeling with Copulas



Goal: To describe the dependence between random variables

Related NSF-CBMS Conferences

- Environmental Statistics, 2001, Richard L.Smith (UNC)
- Statistical Climatology, 2012, Doug Nychka (NCAR, now at Colorado School of Mines)
- Topological Data Analysis: Topology, Geometry, and Statistics, 2016, Sayan Mukherjee (Duke)
- Elastic Functional and Shape Data Analysis, 2018, Anuj Srivastava (FSU)
- Topological Methods in Machine Learning and Artificial Intelligence, 2019, Gunnar Carlsson (Stanford)

Environmental Data

- Observations
- Computer Model Output
- Reanalysis

ENVR Data Challenge

Daily values of maximum temperature (MAXT), minimum temperature (MINT), and precipitation (PRCP) over the continental United States of America (CONUS)

- WRF: 4 km, 2000 Oct. - 2013 Sept. $[1359 \times 1015 \times 4748]$ for each variable
- ERA-Interim: $\sim$ 80 km, 1979 - 2017. $[105 \times 52 \times 14245]$ for each variable
- CESM-LENS: 1 degree, 1920-2005, 40 (42) ensemble members. $[63 \times 41 \times 31390]$ for each ensemble member.

Other Datasets

- United States Historical Climatology Network
  `http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html`
- CPC: Daily precipitation, 0.5 degree, 1979-2020 `https://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/`
- IBTrACS: Tropical cyclone best track data
  `https://www.ncdc.noaa.gov/ibtracs/`
- NOAA tide gauge data
  `https://www.ngdc.noaa.gov/hazard/tide.shtml`
- Cabauw wind tower data
  `http://www.cesar-database.nl/Welcome.do`
- Community Multiscale Air Quality Modeling System (CMAQ) model output

Wildfire Data