Revealing Severe Weather Hotspots: Tornadoes, Hail, and Wind Events

Group 7: Elaine Wan, Erika Wan, Hongbo Zhao, Tianjia Li, Terence Wang

Table of contents

Introduction

Background - Tornado, hail, and strong winds:

- Major severe weather hazards within the U.S.
- Cause significant fatality, injuries, and property damage

Project objectives:

- Map spatial patterns of all three severe weather events
- Identify geographic hotspots using geospatial analysis
- Utilize parallel computing to speed up data processing

Research Questions:

- Where are severe weather events (e.g., tornado) most concentrated?
- Which regions or geographic hotspots exhibit the highest frequency of severe weather activity based on start/end latitude/longitude coordinates?
- Is there a statistically significant difference in the occurrence or severity of events between different seasons or times of day?

Data Source

We obtained our tornado, hail, wind dataset from the **Storm Prediction Center's "WCM" archive** (1950–2023) (https://www.spc.noaa.gov/wcm/#data) across **48 contiguous U.S. states**:

- Tornadoes: 71,398 events (1950-2023_torn.csv.zip)
- Hail: 395,904 events (1955-2023_hail.csv.zip)
- Wind: 516,701 events (1955-2023_wind.csv.zip)

Key Fields:

- Temporal fields: date (YYYY-MM-DD) and time (HHMM, local)
- Spatial fields: st (state code), sn (subdivision code), slat(latitude), slon(longitude)
- Meteorological details: mag (F-scale rating), length (miles), width (yards)
- Casualty counts: injuries, fatalities
- Loss: economic damages

Data Cleaning Process:

- Dropped all county-level summaries (sg= −9) to focus on state-level events
- Removed rows with invalid/missing geolocations

The **cleaned dataset** contains approximately **70653** tornado events—each with complete **temporal, spatial, and intensity information**—ready for our **temporal trend and spatial hotspot analyses**.

Data Cleaning and Computation on CHTC

- Developed an R script for **data cleaning and aggregation**.
- Filtered out **county-level entries** and kept **only single record** when multiple states were affected.
- Aggregated cleaned data at the **state** and **year-month** levels.
- Automated processing using a shell script (project.sh) and submission file (project.sub) to run in CHTC.
- **Ran 180 parallel jobs** on CHTC:
    a. Typical runtime: **~1 to 5 minutes** per job
    b. Memory requested: **1GB**
    c. Disk requested: **1GB**
- Collected and downloaded summary tables for further analysis.

Kernel Density Estimation (KDE)

Kernel Density Estimation (KDE) is a **non-parametric method** to estimate the **probability density function** of a random variable. In spatial analysis, KDE **transforms a set of discrete event locations** into **a continuous surface**, **revealing "hotspots"** where events cluster.

For geographic coordinates $\{(x_i, y_i)\}$, the estimate at (x, y) is

$$\hat{f}(x,y) = \frac{1}{n\,h_x\,h_y} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_x}, \frac{y - y_i}{h_y}\right).$$

With an isotropic Gaussian kernel ($h_x = h_y = h$), this simplifies to

$$\hat{f}(x,y) = \frac{1}{n\,h^2} \sum_{i=1}^{n} \frac{1}{2\pi} \exp\left(-\frac{(x-x_i)^2+(y-y_i)^2}{2h^2}\right).$$

Pros and Cons of KDE

Advantages

• Smooths point data into continuous "hotspots" for intuitive visuals.

• Makes no parametric assumptions about underlying distributions.

• Different hazard KDEs can be overlaid in layers for direct comparison.

• Can incorporate weights (e.g., intensity or casualties) if desired.

Disadvantages

• Highly sensitive to bandwidth choice: too large h over-smooths; too small h overfits.

• Boundary bias can underestimate densities near map edges without correction.

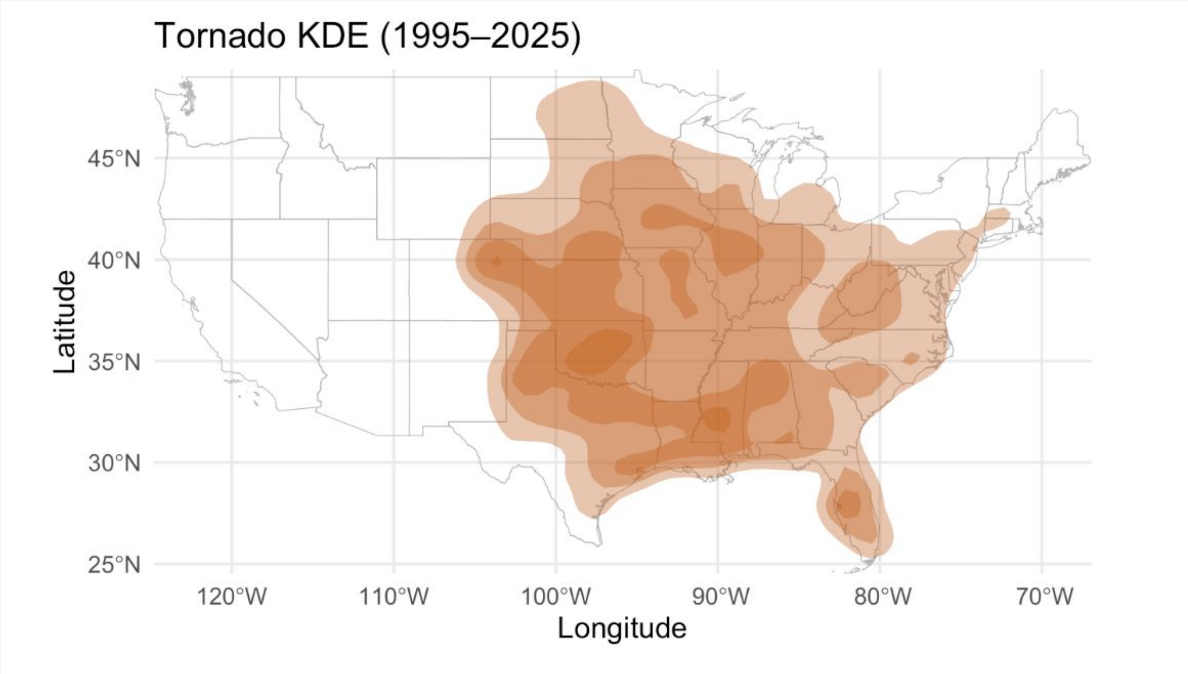
• Computationally intensive for large datasets or high grid resolution.

• Lacks formal statistical inference; results are exploratory and descriptive.

Outcome & Conclusion

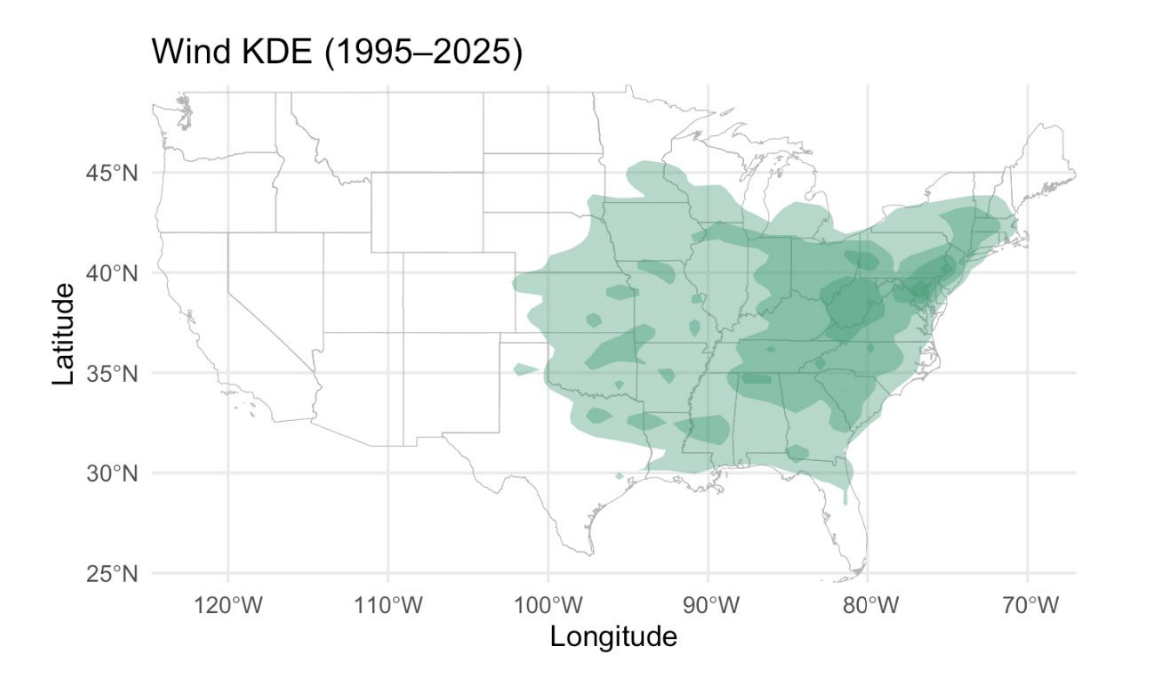
1. **Spatial hotpots:**

   Tornadoes events cluster in the central and Lower Mississippi Valley, with shifting contours northward since the 1990s.



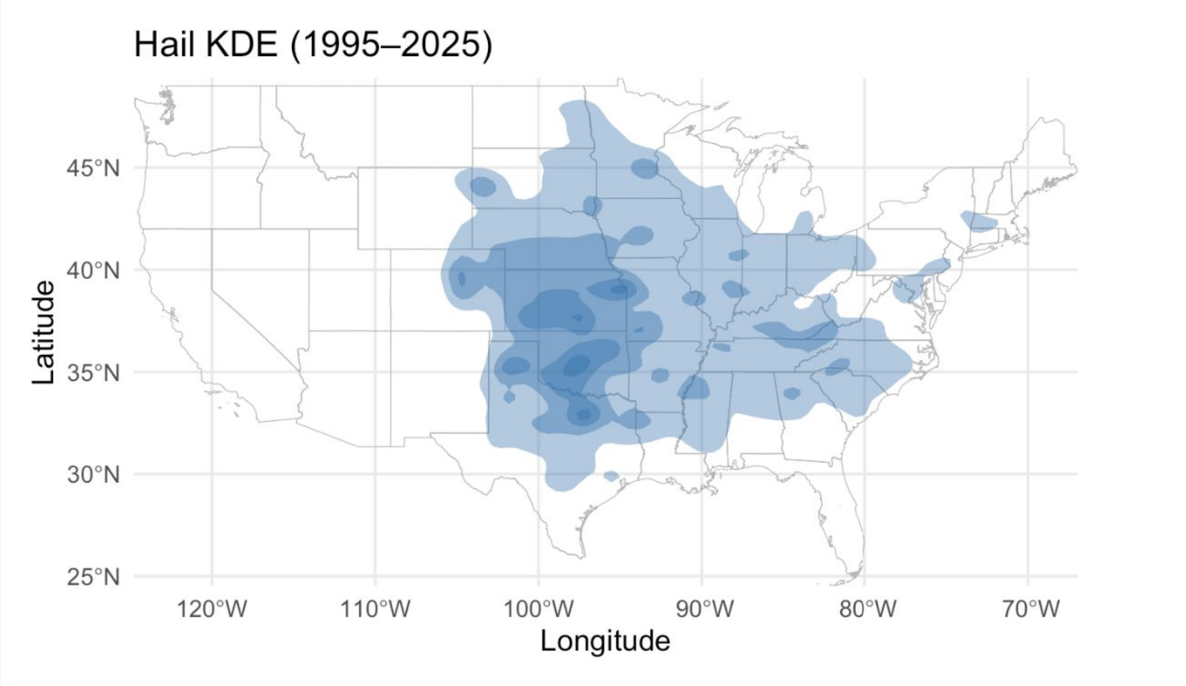
1. **Spatial hotpots:**

   High-wind events cluster in the central and eastern Plains, with shifting contours northward since the 1990s.



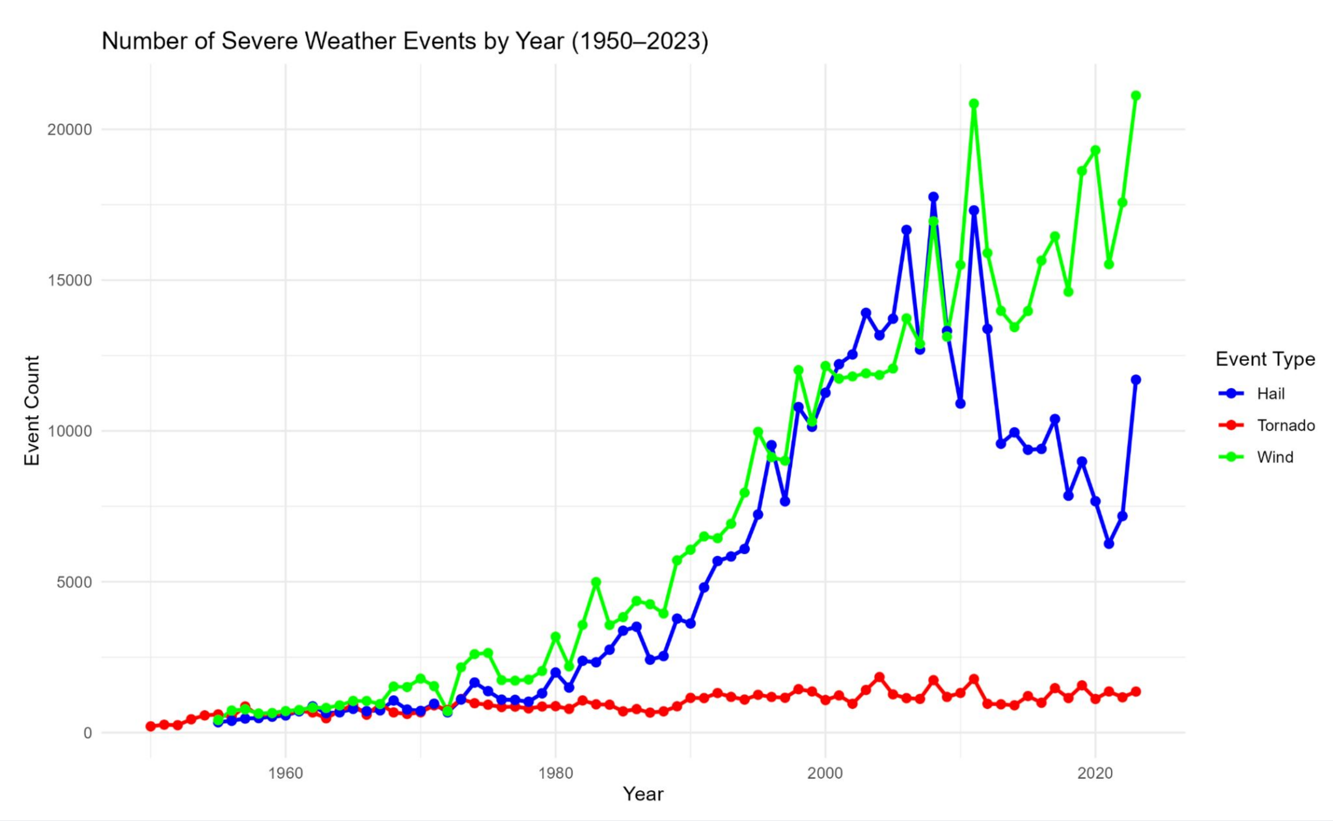
1. **Spatial hotpots:**

   Hail shows stronger concentration in the central Midwest and Front Range regions.



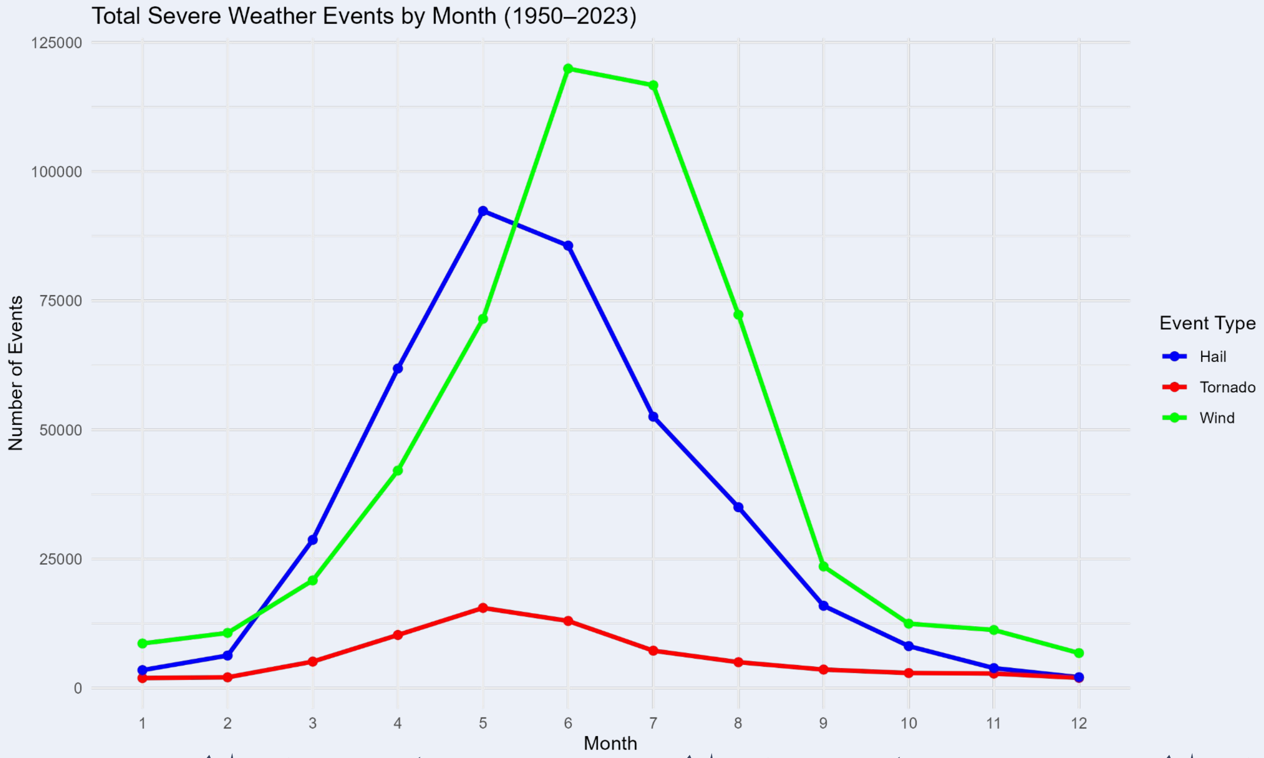
2. Temporal trends:

A clear, statistically significant upward trend in **high-wind** reports over 1950–2023, likely reflecting both true increases in severe weather and enhanced detection.

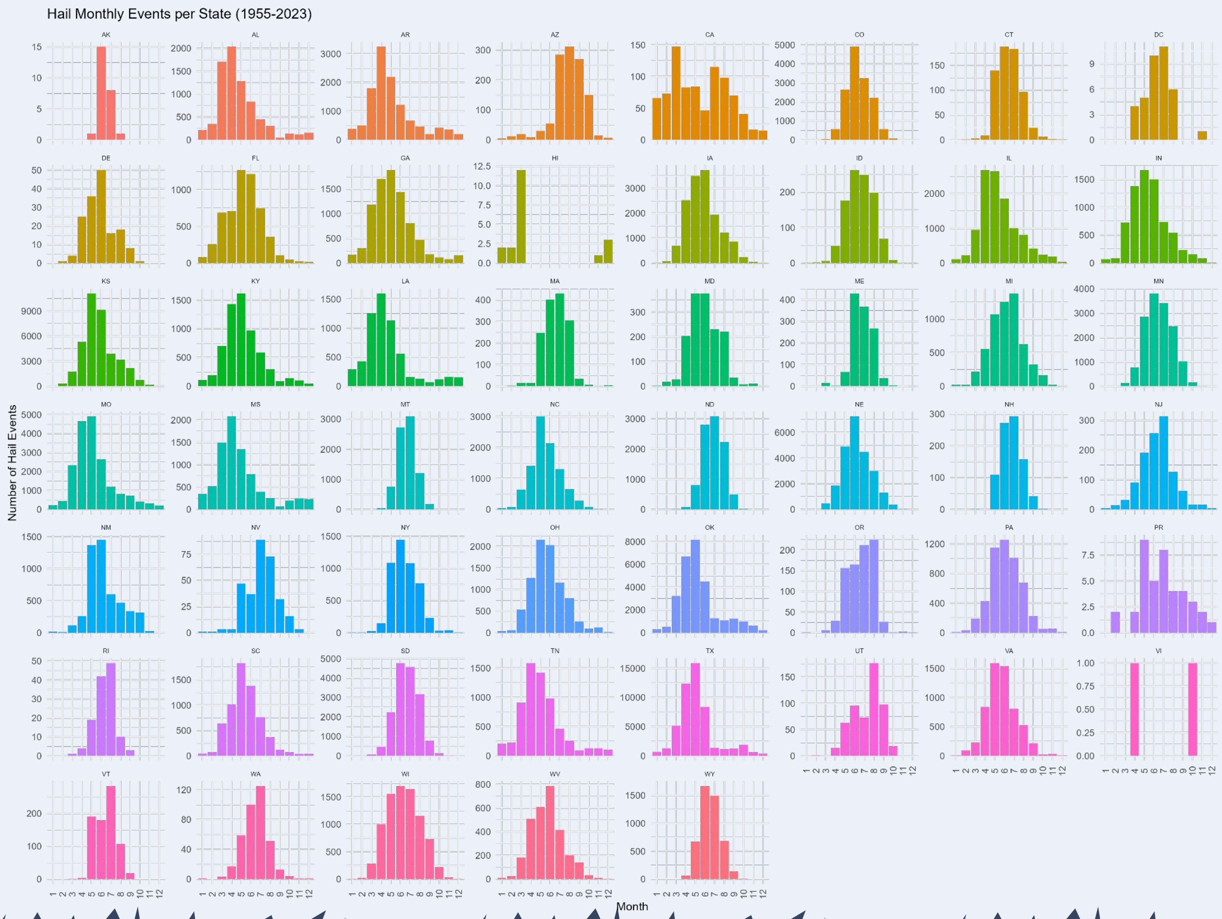


2. Temporal trends:

Two-thirds of all events occur between May and August.



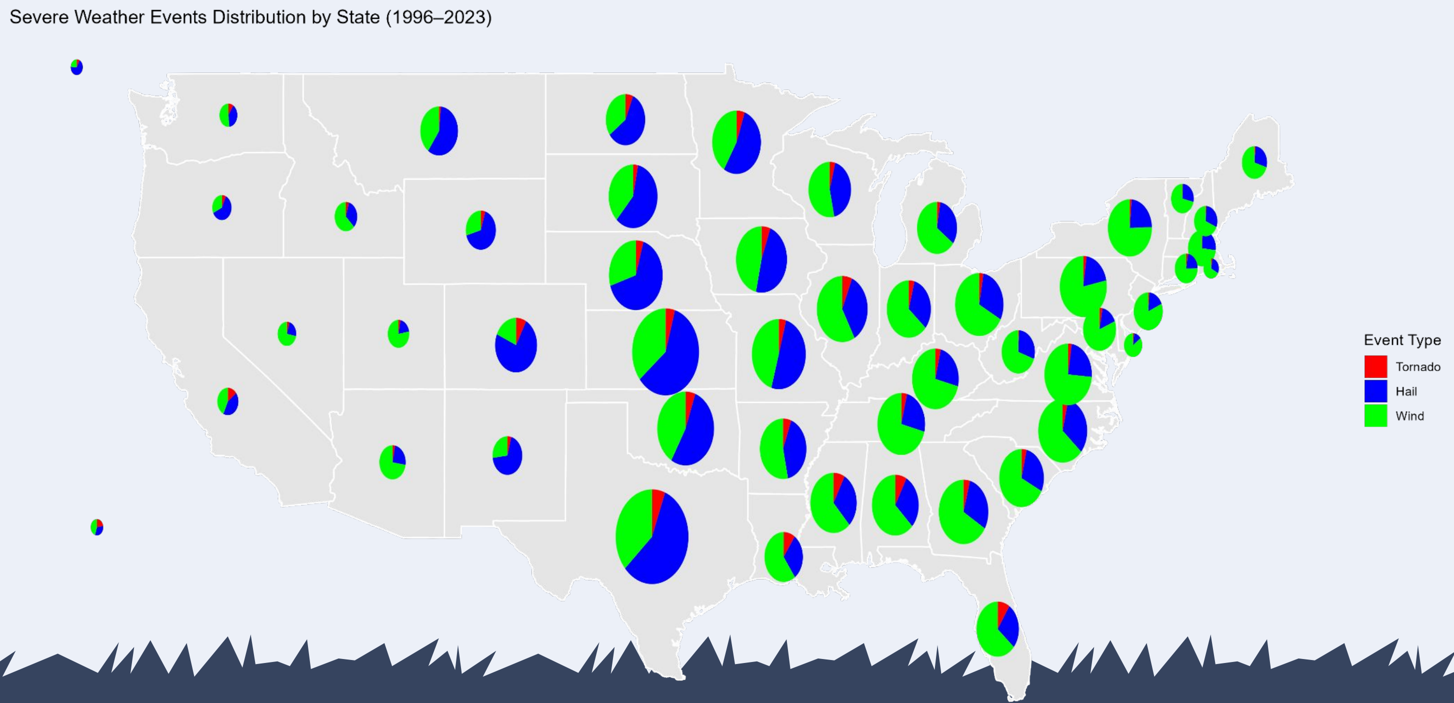
Wind Monthly Events per State (1955-2023)


Hail Monthly Events per State (1955-2023)
Number of Hail Events
Month
AK
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
PR
RI
SC
SD
TN
TX
UT
VA
VI
VT
WA
WI
WV
WY

Tornado Monthly Events per State (1950-2023)
AK
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
PR
RI
SC
SD
TN
TX
UT
VA
VI
VT
WA
WI
WV
WY
Number of Tornado Events
Month

3. The Comparative Total Impact or Losses Between States

- **Event Distribution** (Within States):
  - Each pie chart breaks down the percentage of Tornadoes, Hail, and Wind events for a specific state.

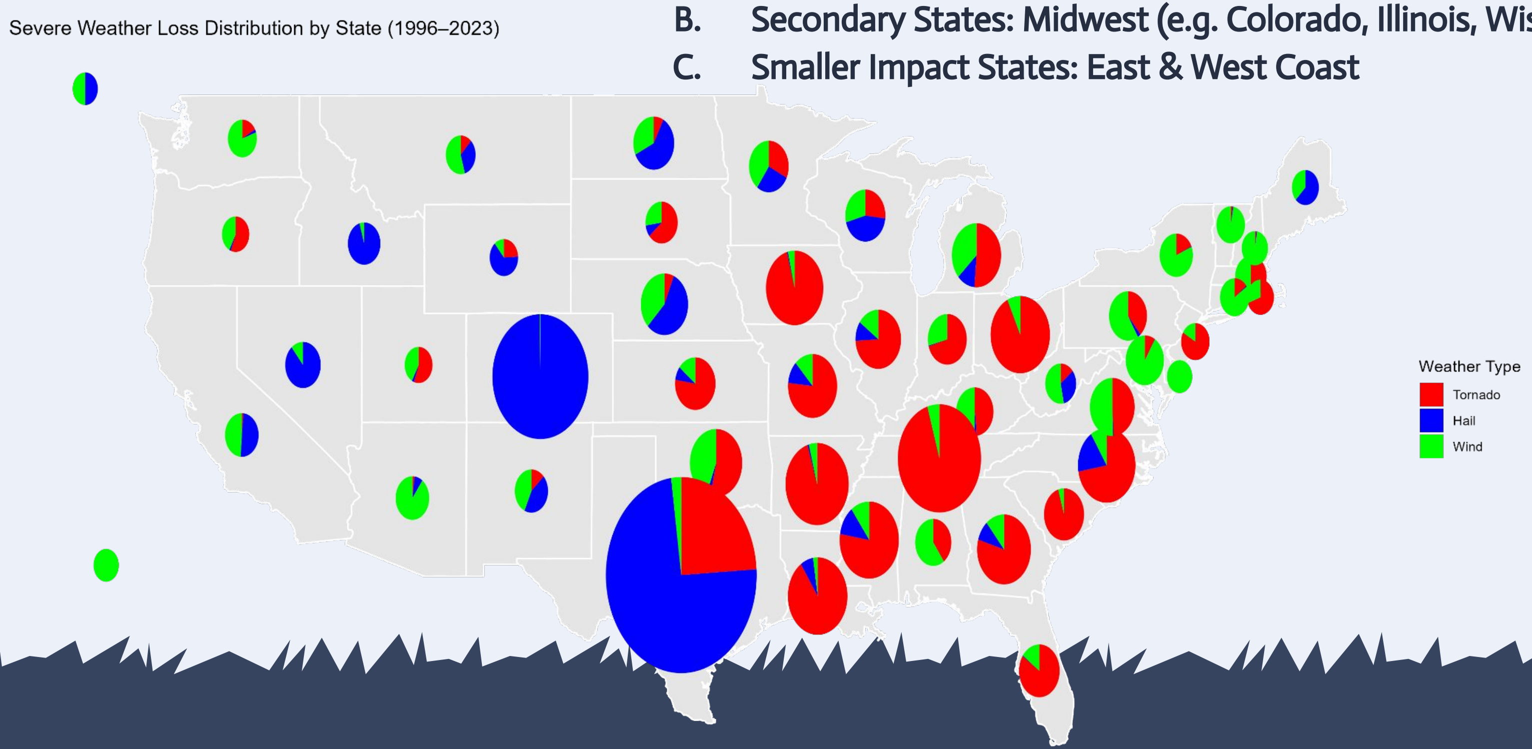


3. The Comparative Total Impact or Losses Between States (cont'd)

Comparative Impact (Between States):

The size of each pie chart reflects the relative severity or total losses experienced by each state.

A. **Dominant States: Central Region (e.g. Texas, Oklahoma, Florida)**
B. **Secondary States: Midwest (e.g. Colorado, Illinois, Wisconsin)**
C. **Smaller Impact States: East & West Coast**



Future Work

Future work could include:

- **Adaptive KDE** with location-dependent bandwidths to refine hotspot boundaries.
- **Spatio-temporal modeling** (e.g., space–time kriging or Bayesian hierarchical models) to quantify year-to-year shifts and predict emerging hotspots under climate change scenarios.
- **Integration of environmental covariates** (CAPE, shear, soil moisture) to link physical drivers with event clustering.
- **Higher-resolution county or grid-cell analyses** using expanded computational resources to capture local risk variations.
- **Event intensity weighting** in density estimates, so that more destructive events contribute proportionally to the hotspot surfaces.

Resources

Storm Prediction Center. (n.d.). *Severe weather event database: Tornadoes, hail, and damaging winds (1950–2023)*. NOAA National Weather Service. https://www.spc.noaa.gov/wcm/#data

Thanks!