Analysis of tornadic activity in Texas, United States between 1950-2021

Joost Luijten, Tim Resing, Shuting Yu, Karolis Vilkevičius

Our Datasets

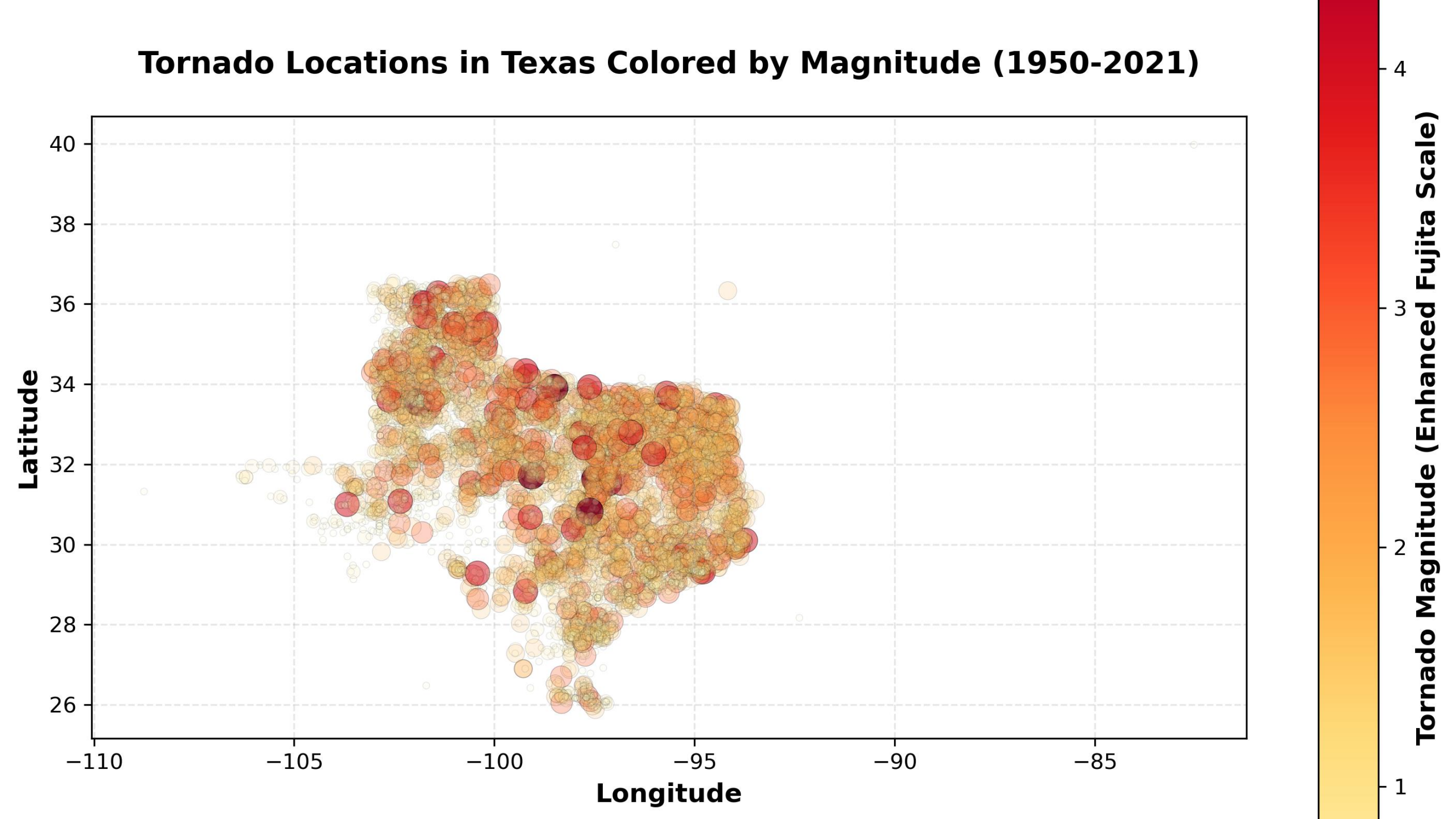
Historical US Tornado Data

- NOAA's Storm Prediction Center
- Historical occurrence spanning

  1950-2021
- Recorded impact (injury, fatality)
- Geospatial distribution
- Enhanced Fujita Scale (weakest EF0 -> strongest EF5)

ERA5 Global Climate

- Climate Data Store (EU)
- Diverse meteorological data
- Gridded


Tornado Locations in Texas Colored by Magnitude (1950-2021)
Latitude
Longitude
40
38
36
34
32
30
28
26
−110
−105
−100
−95
−90
−85
Tornado Magnitude (Enhanced Fujita Scale)
4
3
2
1

RQ 1.

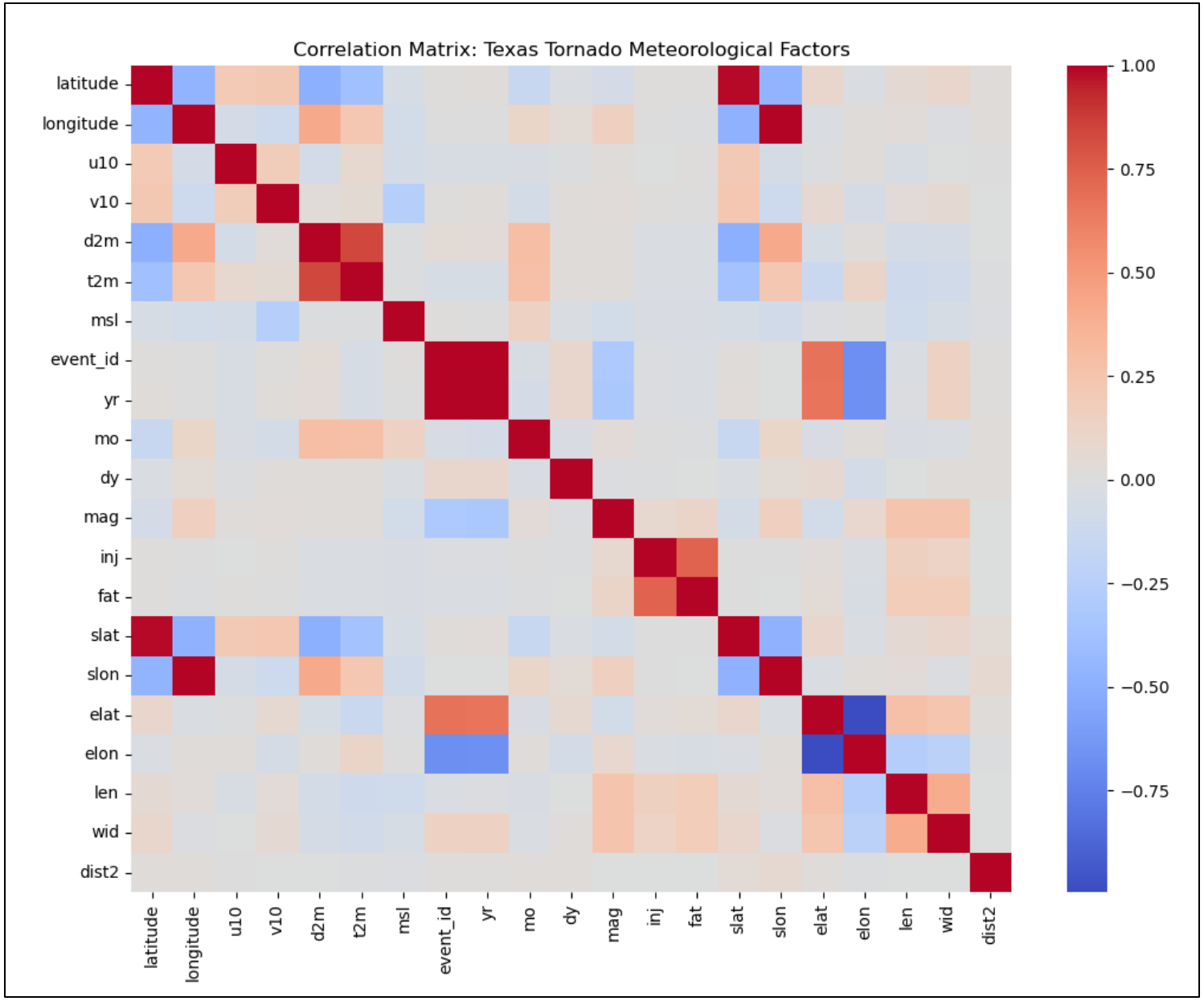
Do supercell storm indicative meteorological variables (cloud cover, precipitation, vertical wind / vertical shear, potential evaporation, etc.) have a statistically significant effect on tornado strength in Texas?

H0: *None of the meteorological variables have a statistically significant effect on tornado strength.*

H1: *At least one meteorological variable has a statistically significant effect on tornado strength.*

Correlation matrix I.

u10: east-west wind
v10: north-south wind
d2m: dewpoint at 2m
msl: mean sea level pressure



Correlation matrix II.

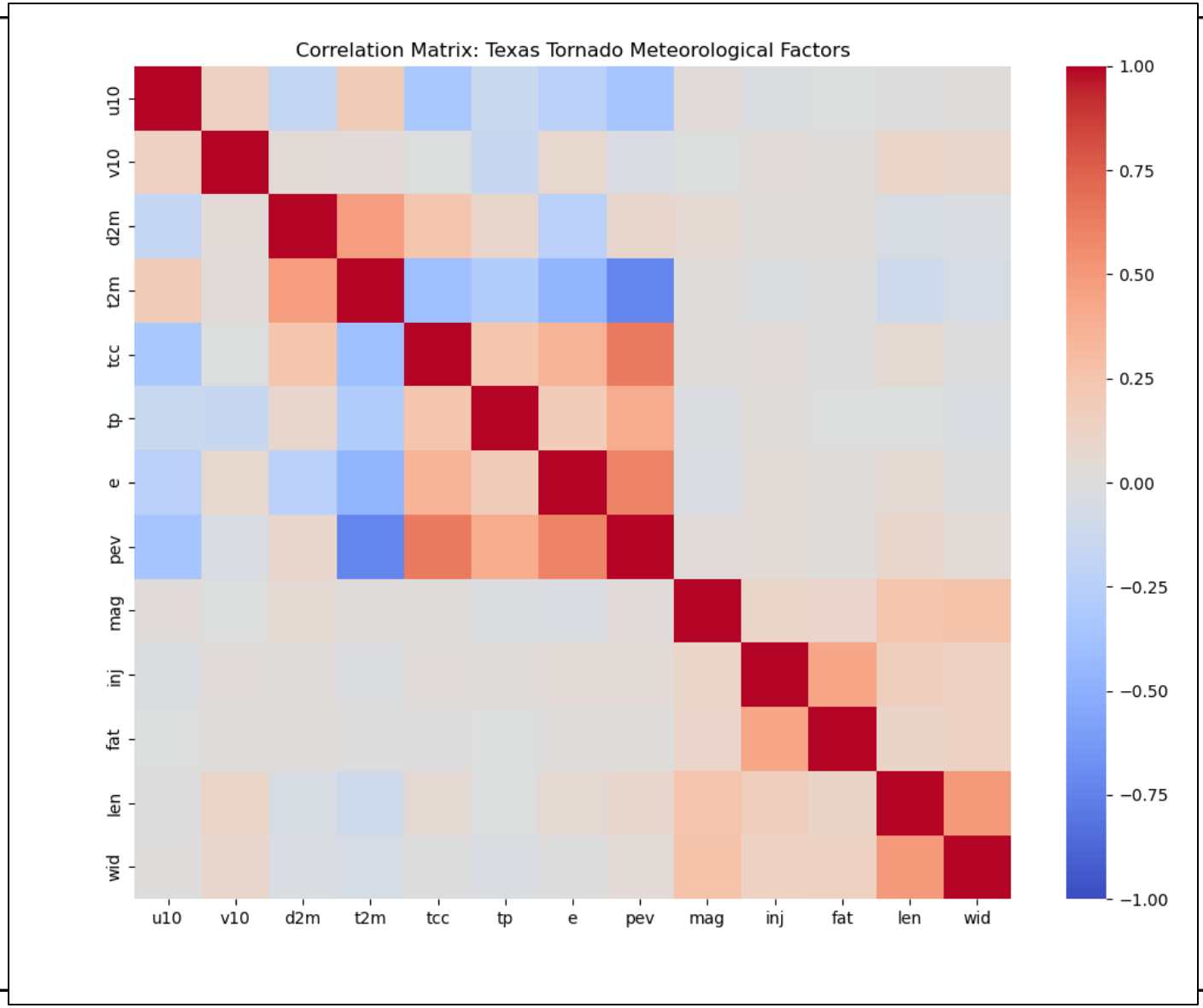
tcc: total cloud coverage

tp: total precipitation

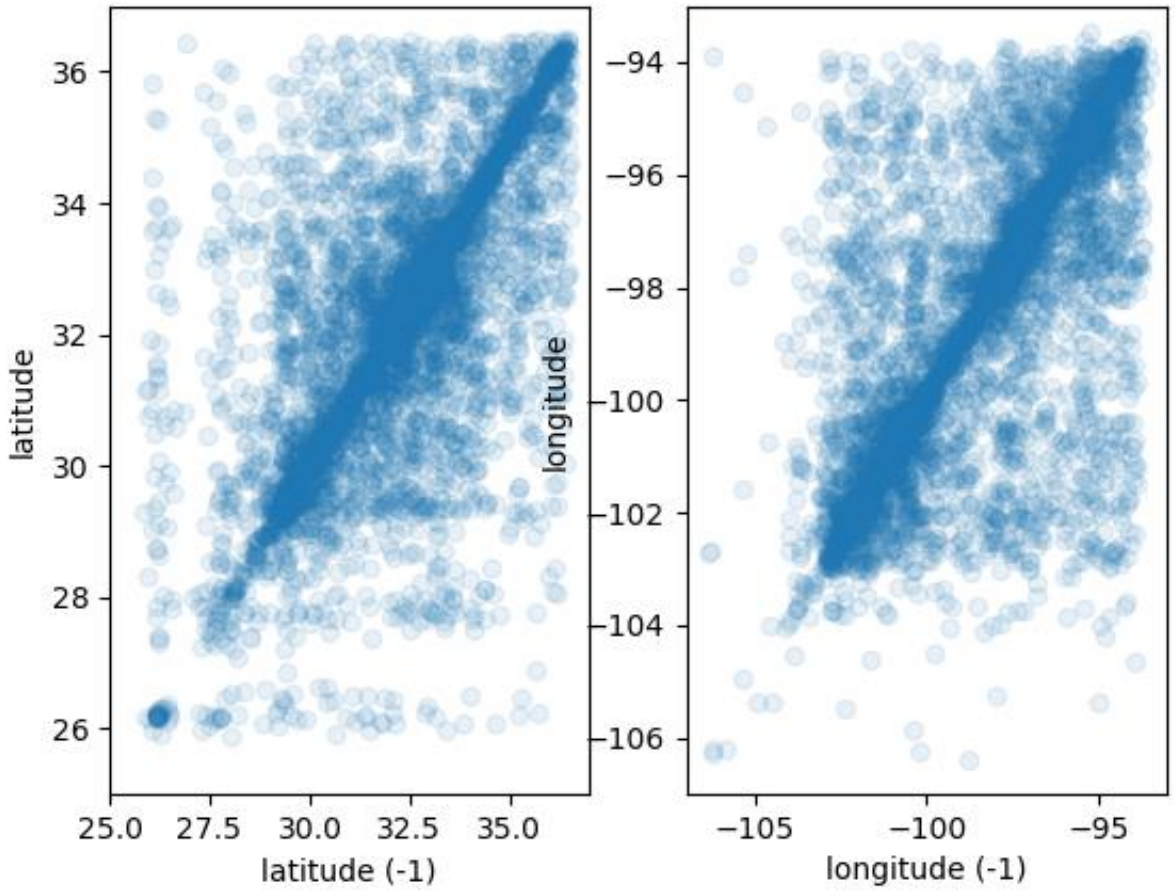
e: evaporation

pev: potential evaporation

FAIL TO REJECT THE NULL HYPOTHESIS!



RQ 2a.

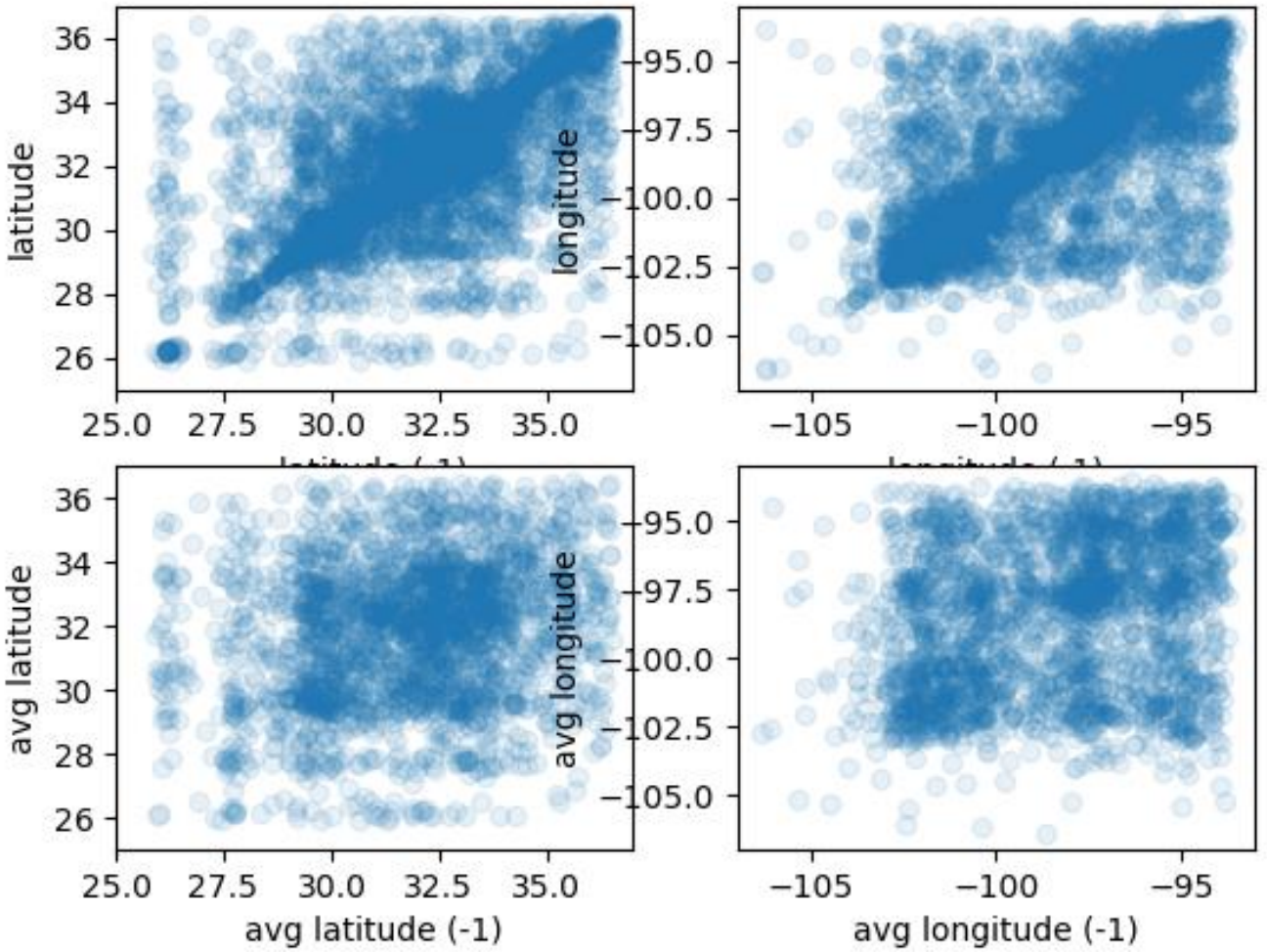


Time-series I

H0: The occurance of tornado's behaves like a time-series.

H1: The occurance of tornado's does not behave like a time series.

Time-series II

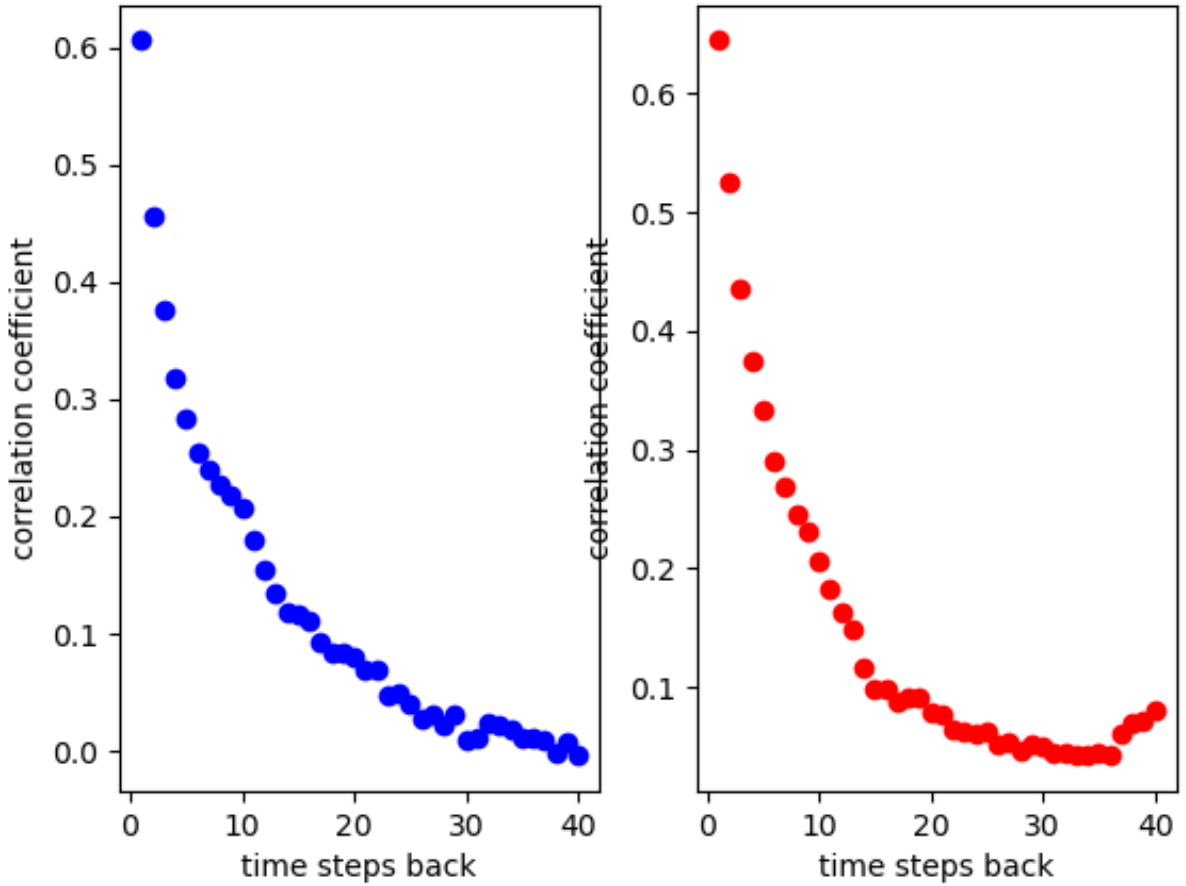


Lat : r = 0.209 and p = 3.4 e-24

Lon : r = 0.238 and p = 4.8 e-31

Time-series III

Cross correlation



RQ 2b.

Temporal Trends

- **Has the (tornado variable) shifted over time?**
- Importance?
- Linear Regression

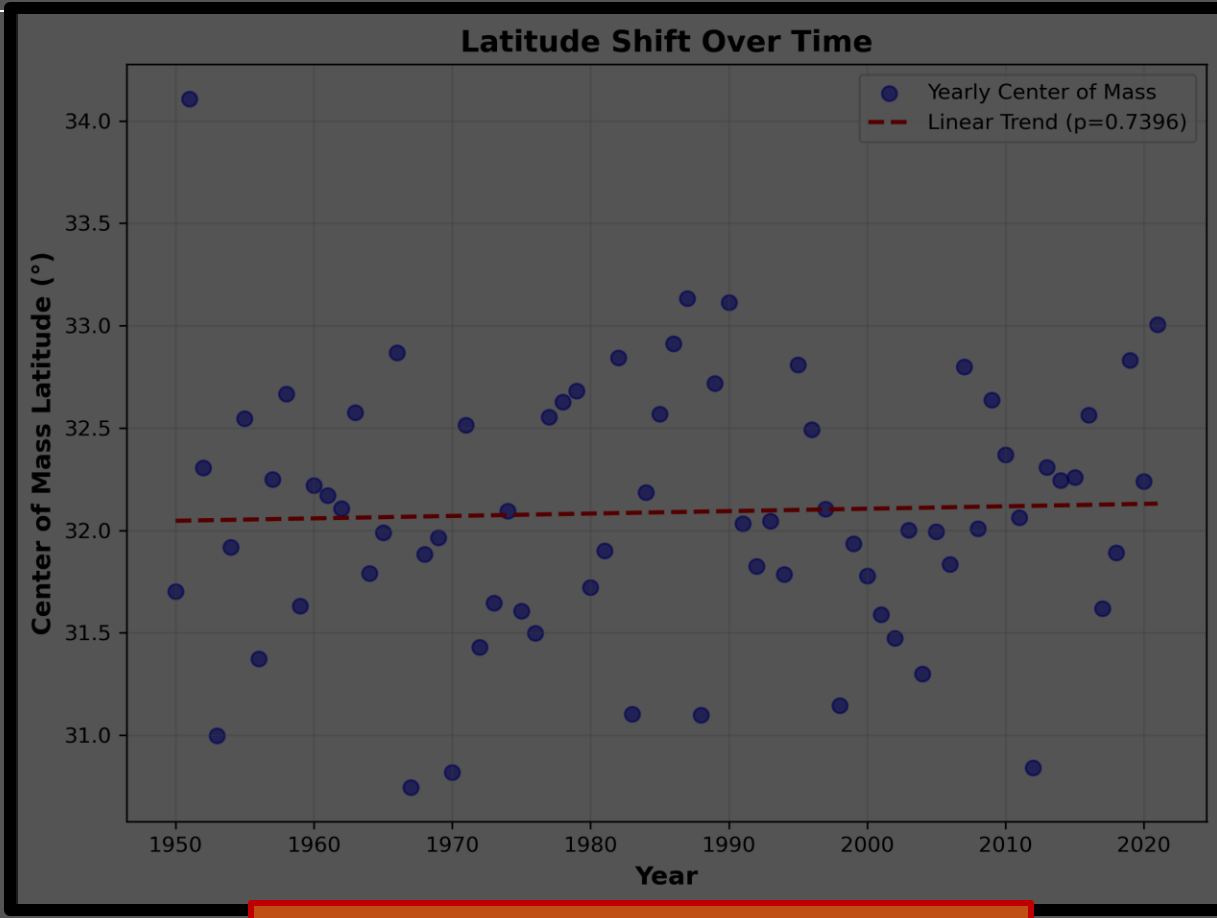
Latitude / Longitude

Path Width / Length

Magnitude

H0: **(variable)** has not shifted over time.

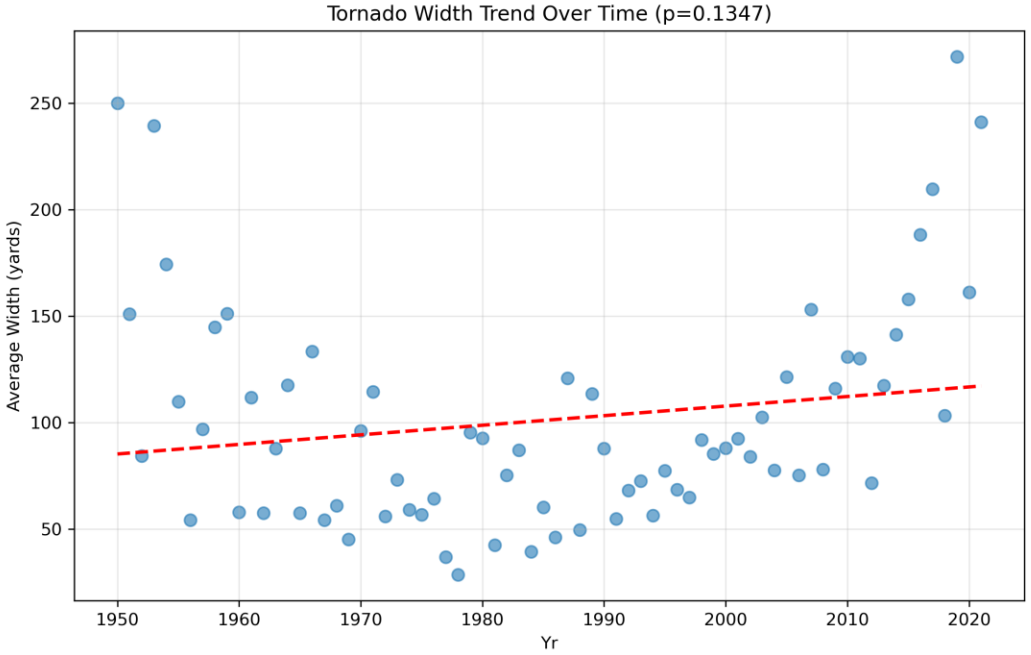
Temporal trends I.

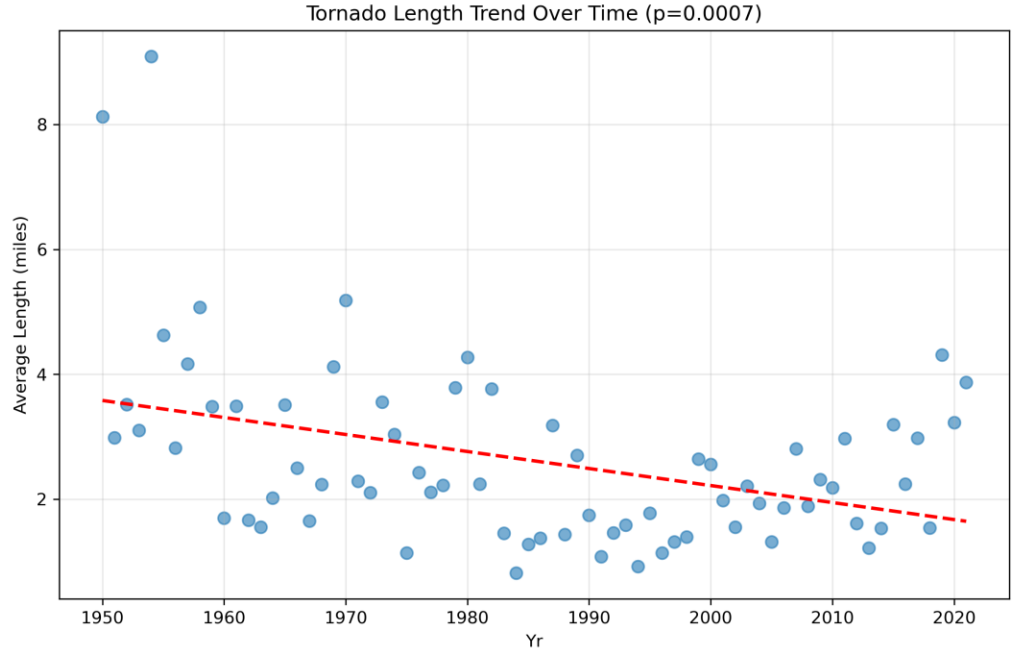


fail to reject the null hypothesis.

fail to reject the null hypothesis.

Temporal trends II.

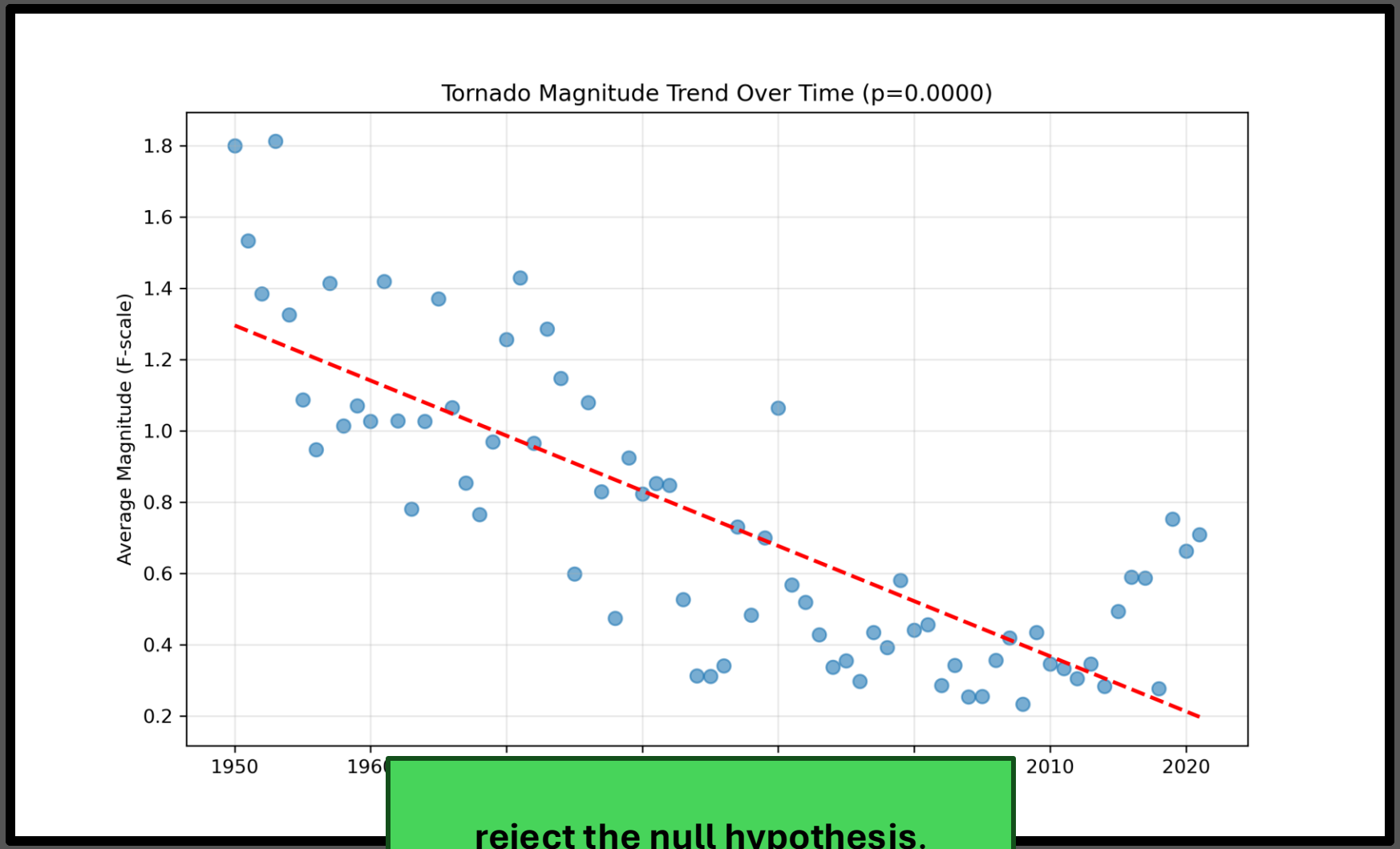




fail to reject the null hypothesis.

reject the null hypothesis.

Temporal trends III.



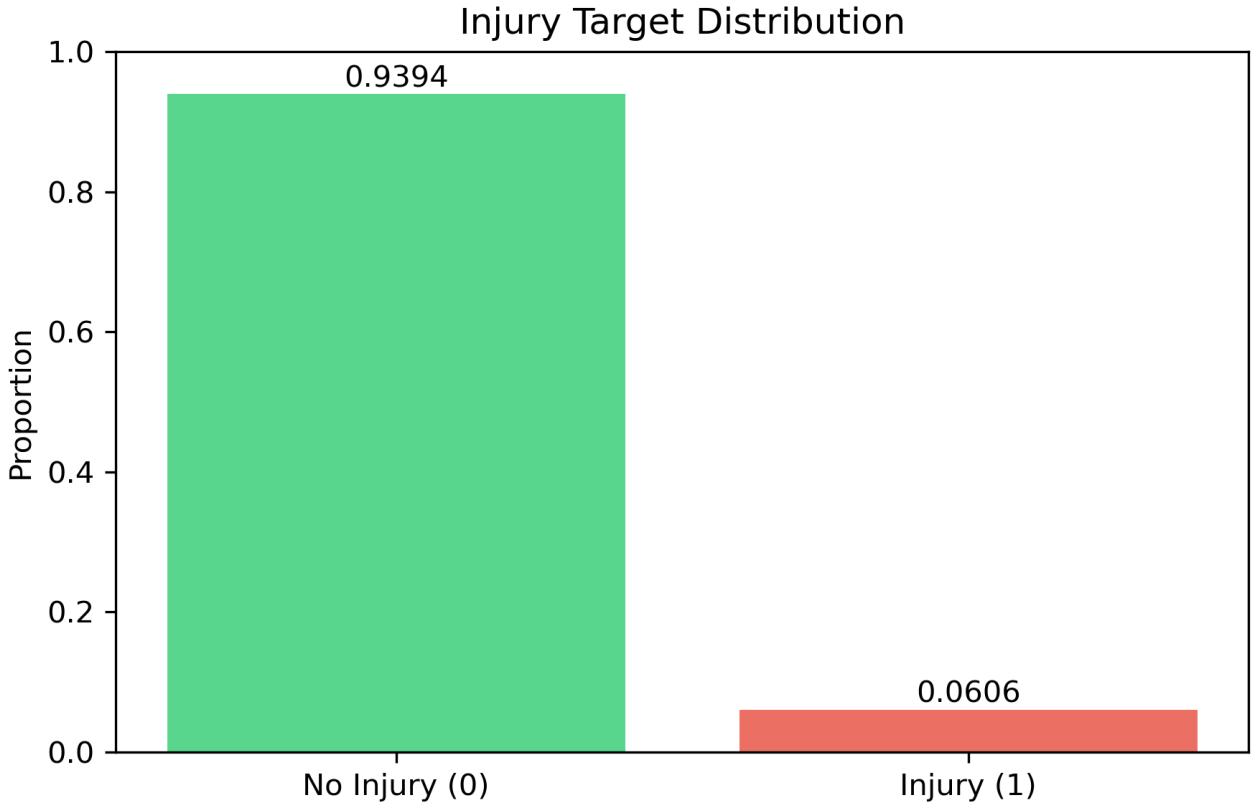
reject the null hypothesis.

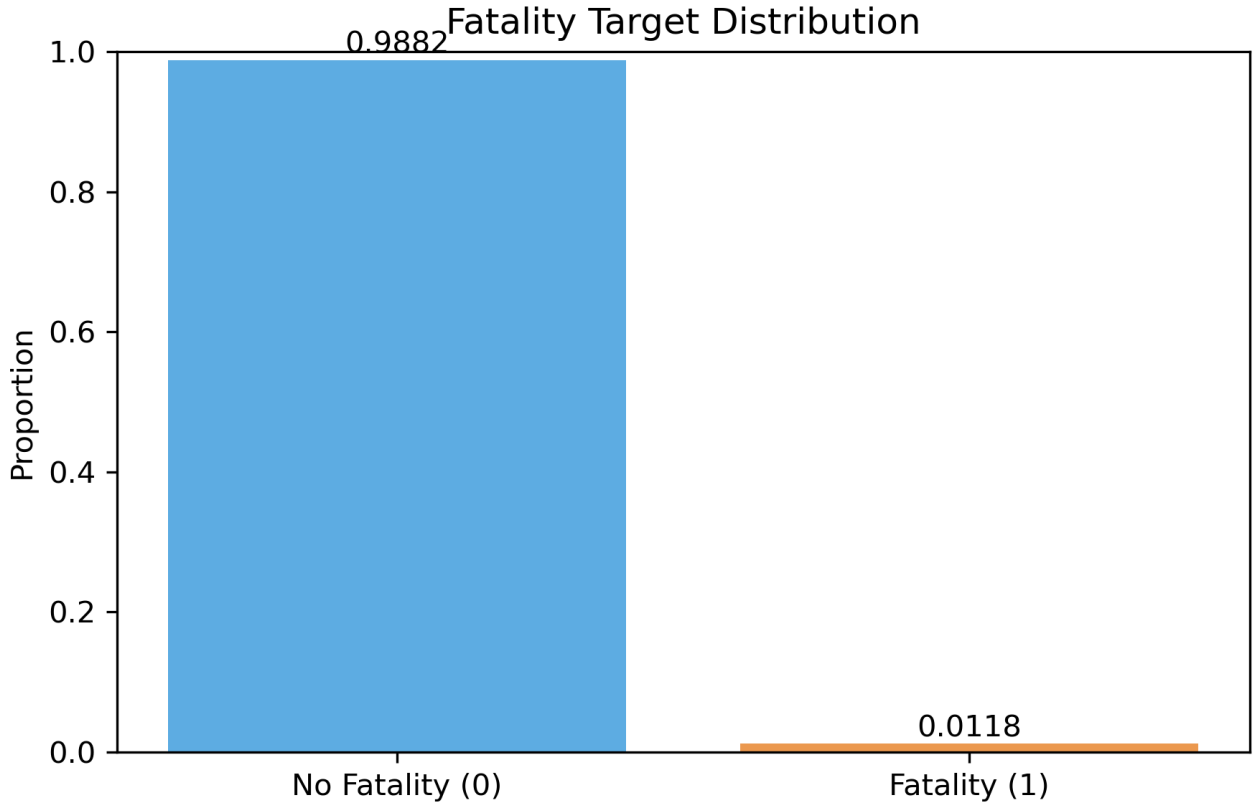
RQ 3.

Do the selected meteorological variables affect tornado impact measures (occurrence of injuries/fatalities) in Texas?

H0: *The meteorological variables have no effect on tornado impact measures.*

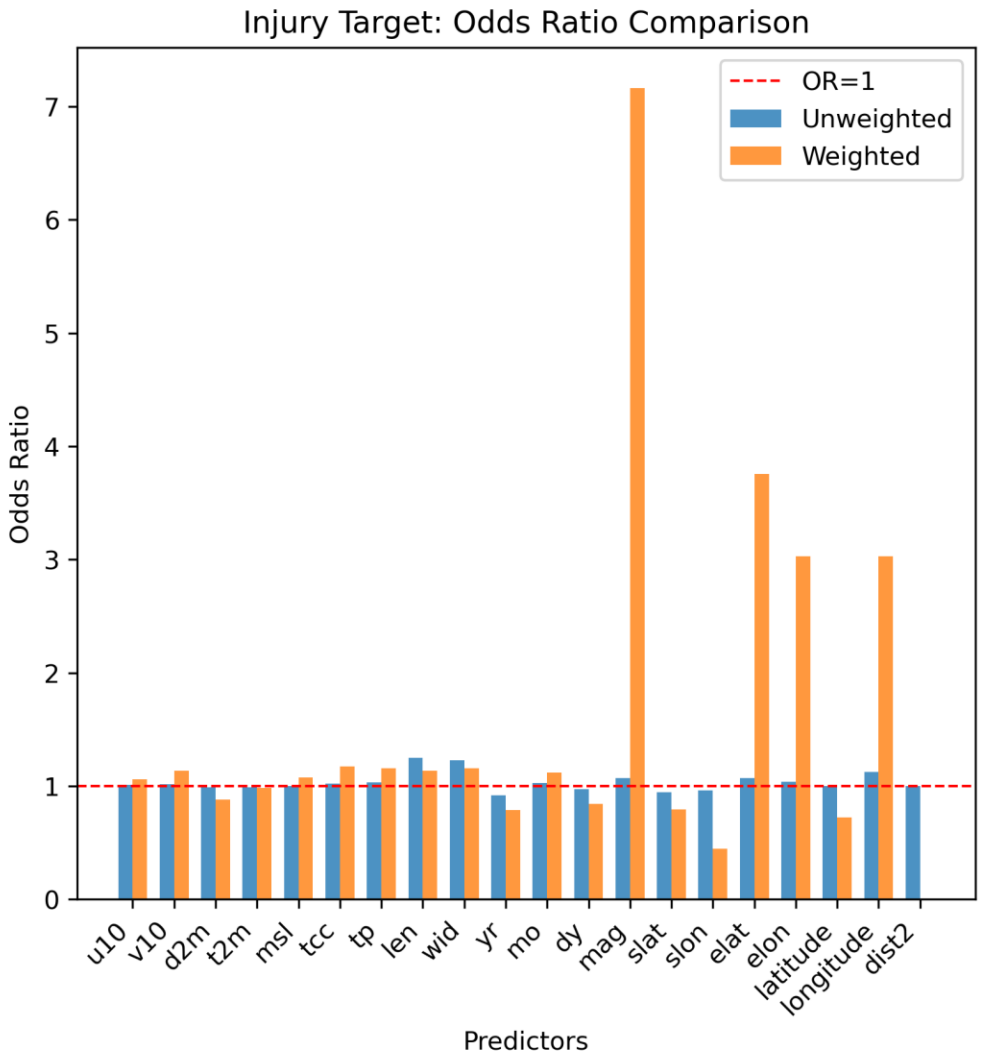
H1: *At least one meteorological variable has a statistically significant effect on tornado strength.*

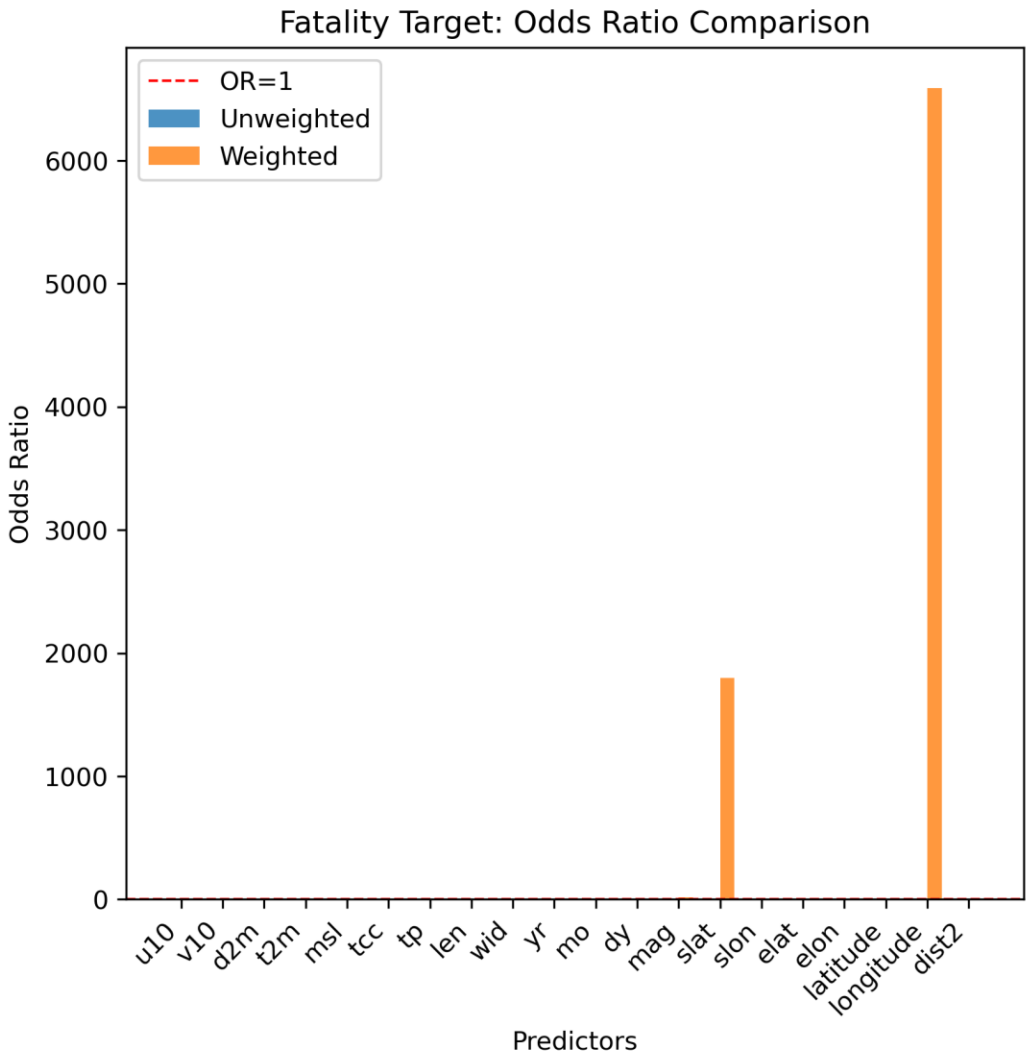




Predicting Fatality & Injury

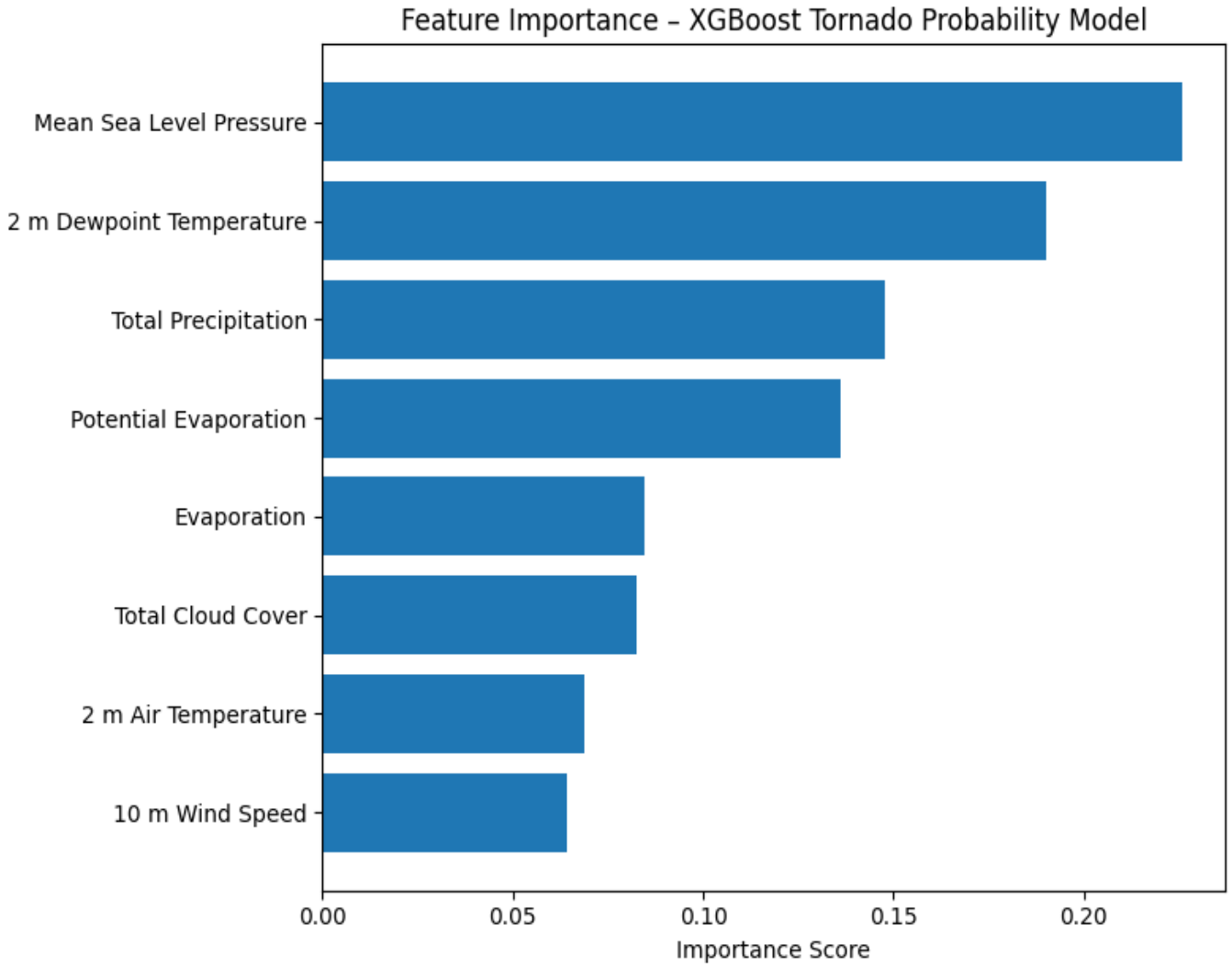
Prodicting Fatality & Injury

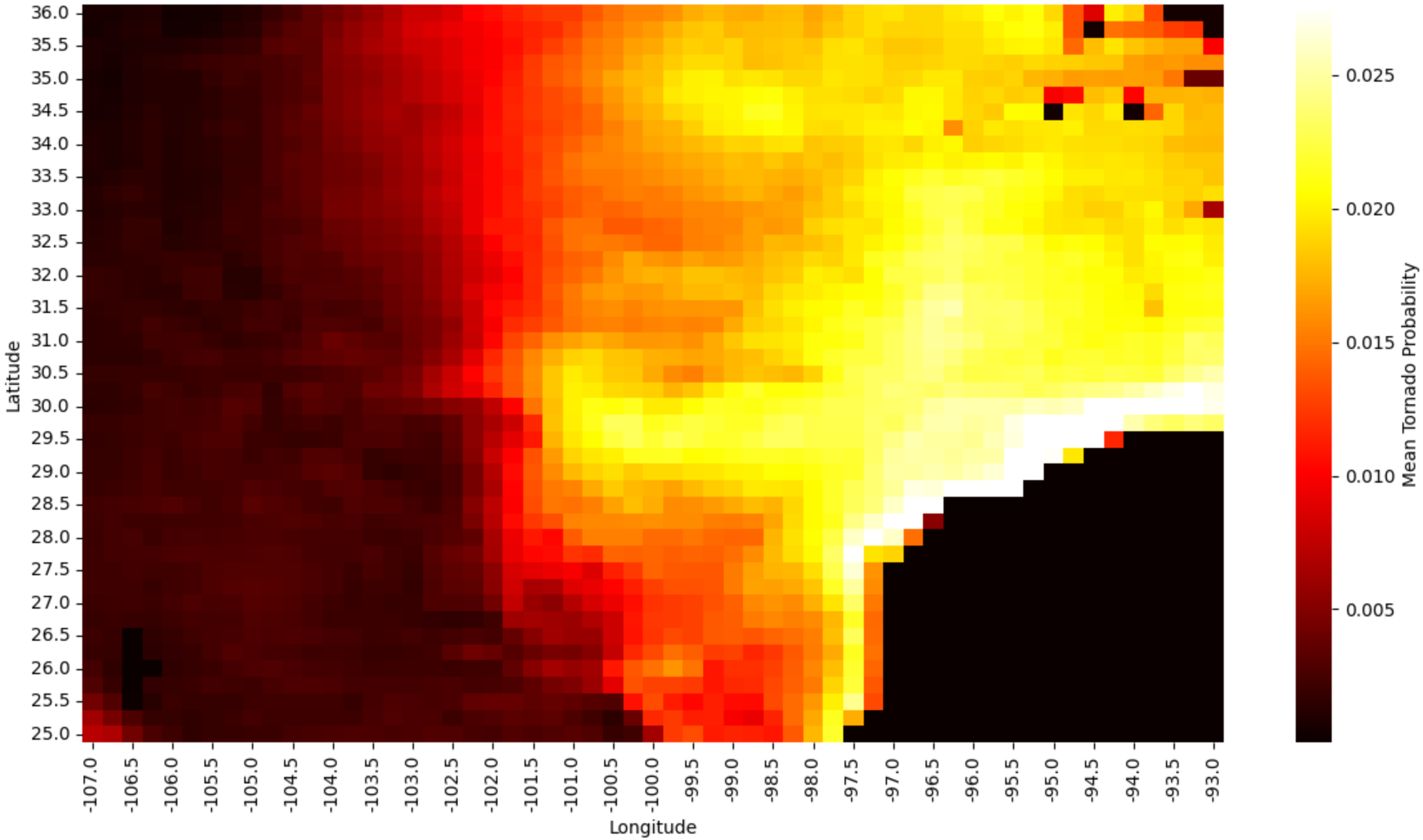




Tornado Forecasting

- Task: Create a model which predicts when tornado occurs based on the weather data
- Binary classification problem:
- 1 if tornado occurs
- 0 if no tornado
- Solution: XGBoost model (gradient-boosted decision tree model)
- Model evaluation: ROC-AUC = 0.939



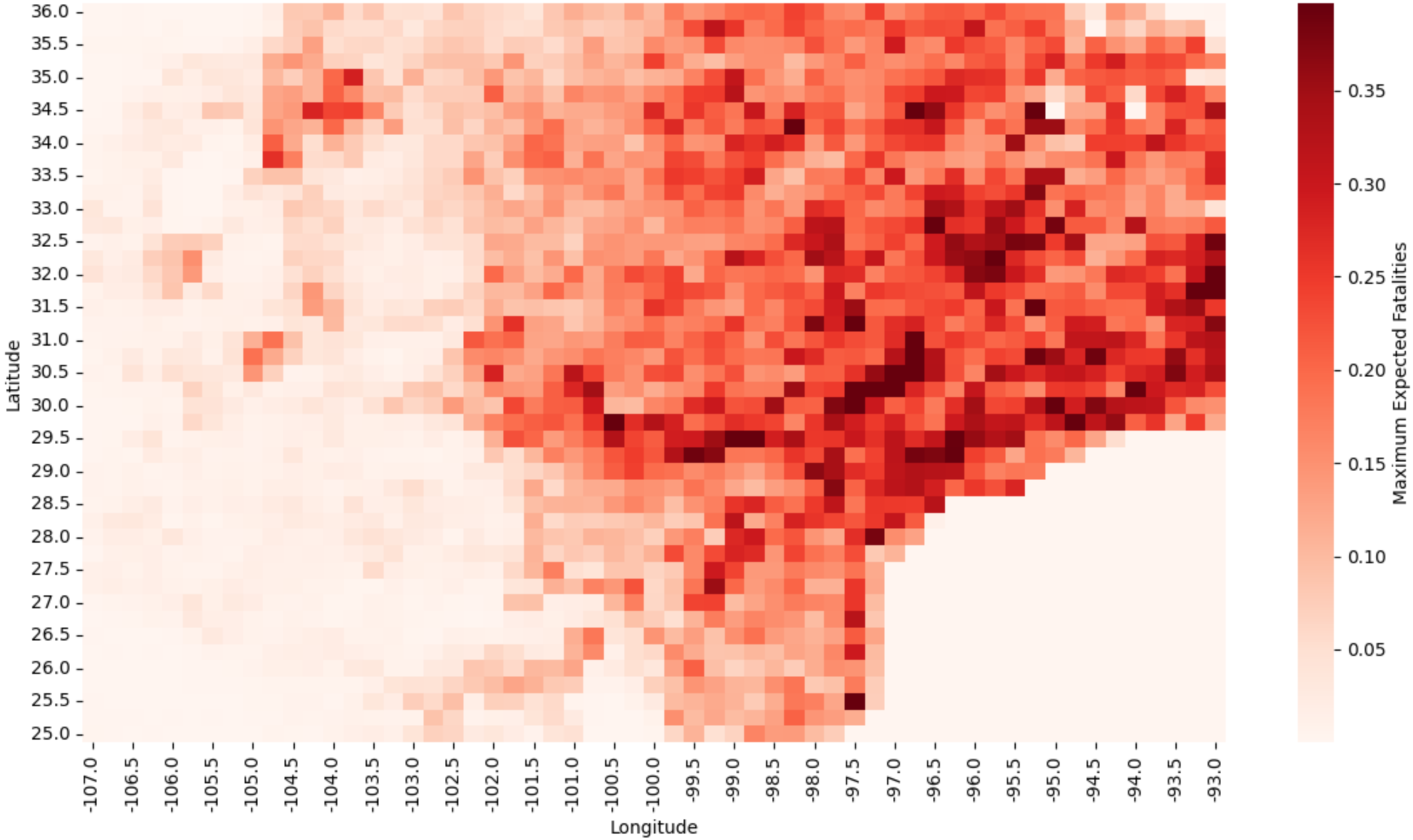

Heatmap of Tornado Probability (2015–2020)
Latitude
36.0
35.5
35.0
34.5
34.0
33.5
33.0
32.5
32.0
31.5
31.0
30.5
30.0
29.5
29.0
28.5
28.0
27.5
27.0
26.5
26.0
25.5
25.0
Longitude
-107.0
-106.5
-106.0
-105.5
-105.0
-104.5
-104.0
-103.5
-103.0
-102.5
-102.0
-101.5
-101.0
-100.5
-100.0
-99.5
-99.0
-98.5
-98.0
-97.5
-97.0
-96.5
-96.0
-95.5
-95.0
-94.5
-94.0
-93.5
-93.0
Mean Tornado Probability
0.025
0.020
0.015
0.010
0.005

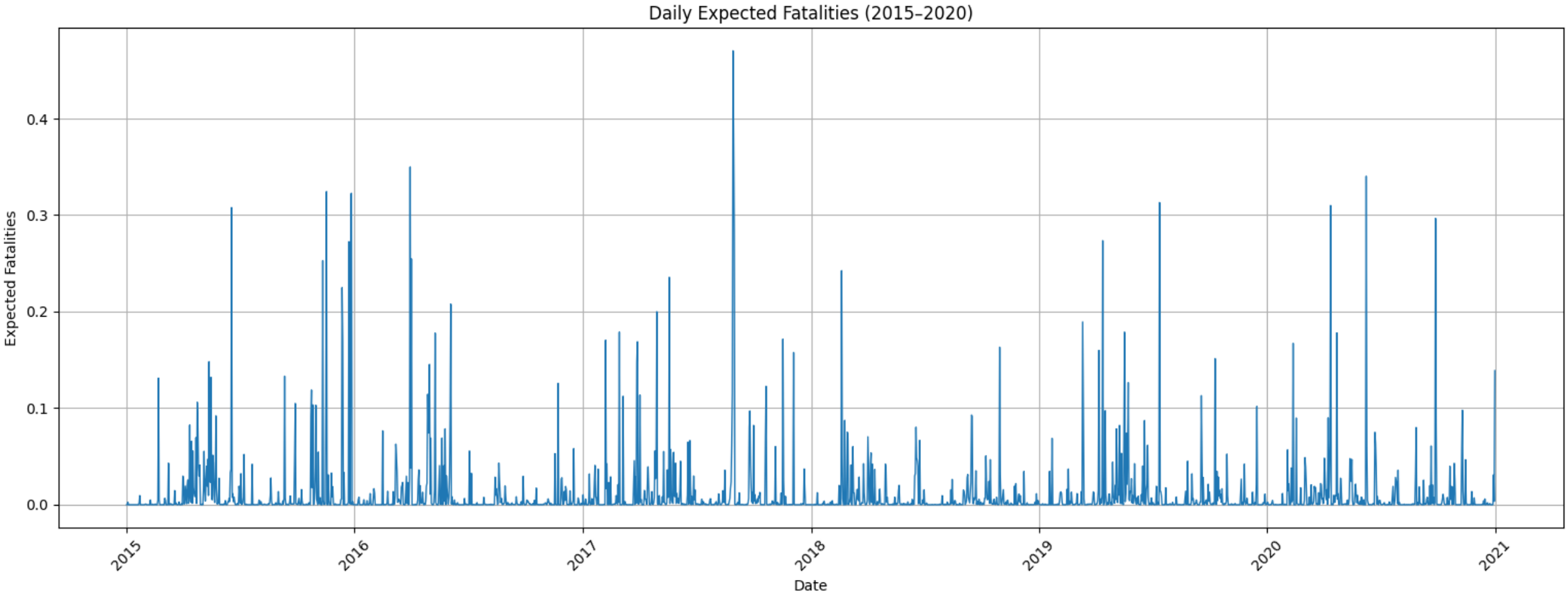
Expected Tornado Fatalities: A Probabilistic Risk Model

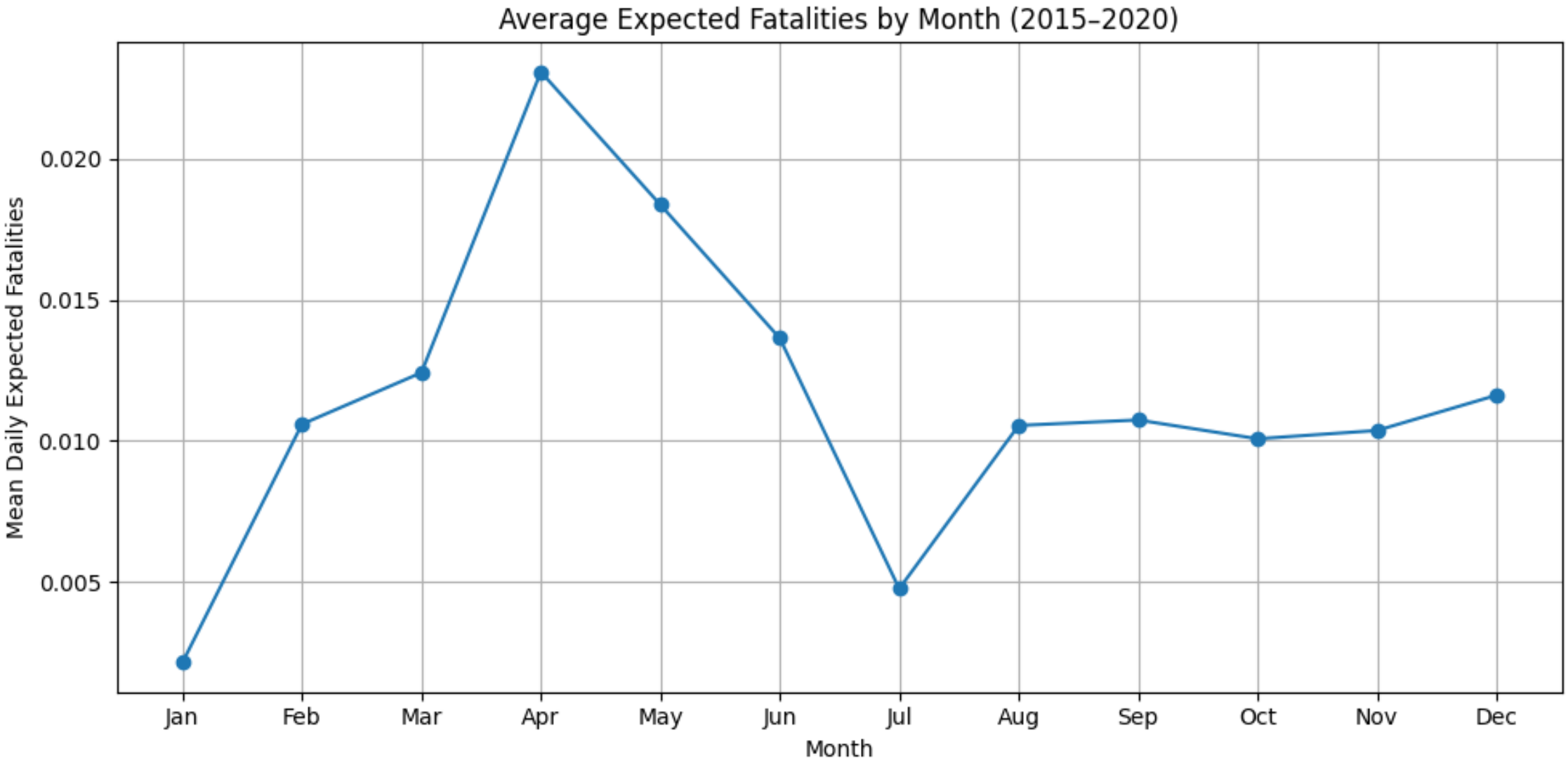
Expected fatalities = P(Tornado) ×
P(Fatal | Tornado) × E(Fatalities | Fatal Tornado)

- P(Fatal | Tornado):
- Given that a tornado occurred, will it cause at least one fatality?
  → From a Random Forest classifier, trained only on tornado events.
  → ROC-AUC ≈ 0.54
- E(Fatalities | Fatal Tornado):
- If a tornado is fatal, how many people are expected to die?
  → From Random Forest Regressor, trained on fatal tornadoes.
  → Mean Absolute Error (MAE) ≈ 2.87

Year	Real Fatalities	Predicted Fatalities
2015	17	5.65
2016	1	3.53
2017	4	5.35
2018	0	2.79
2019	2	3.96
2020	4	3.93

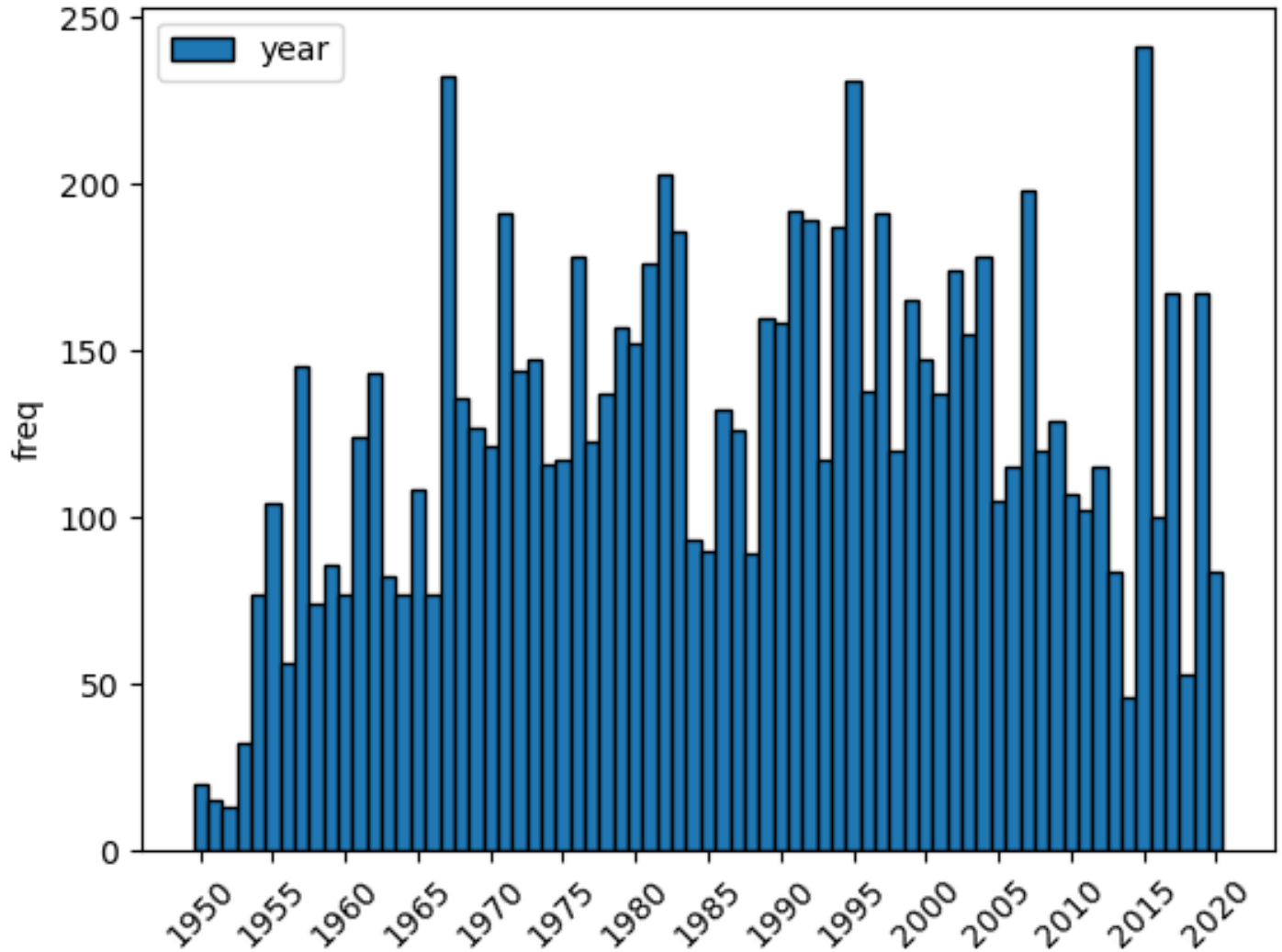

Maximum Expected Tornado Fatality Risk per Grid Cell (2015–2020)
Latitude
36.0
35.5
35.0
34.5
34.0
33.5
33.0
32.5
32.0
31.5
31.0
30.5
30.0
29.5
29.0
28.5
28.0
27.5
27.0
26.5
26.0
25.5
25.0
Longitude
-107.0
-106.5
-106.0
-105.5
-105.0
-104.5
-104.0
-103.5
-103.0
-102.5
-102.0
-101.5
-101.0
-100.5
-100.0
-99.5
-99.0
-98.5
-98.0
-97.5
-97.0
-96.5
-96.0
-95.5
-95.0
-94.5
-94.0
-93.5
-93.0
Maximum Expected Fatalities
0.35
0.30
0.25
0.20
0.15
0.10
0.05

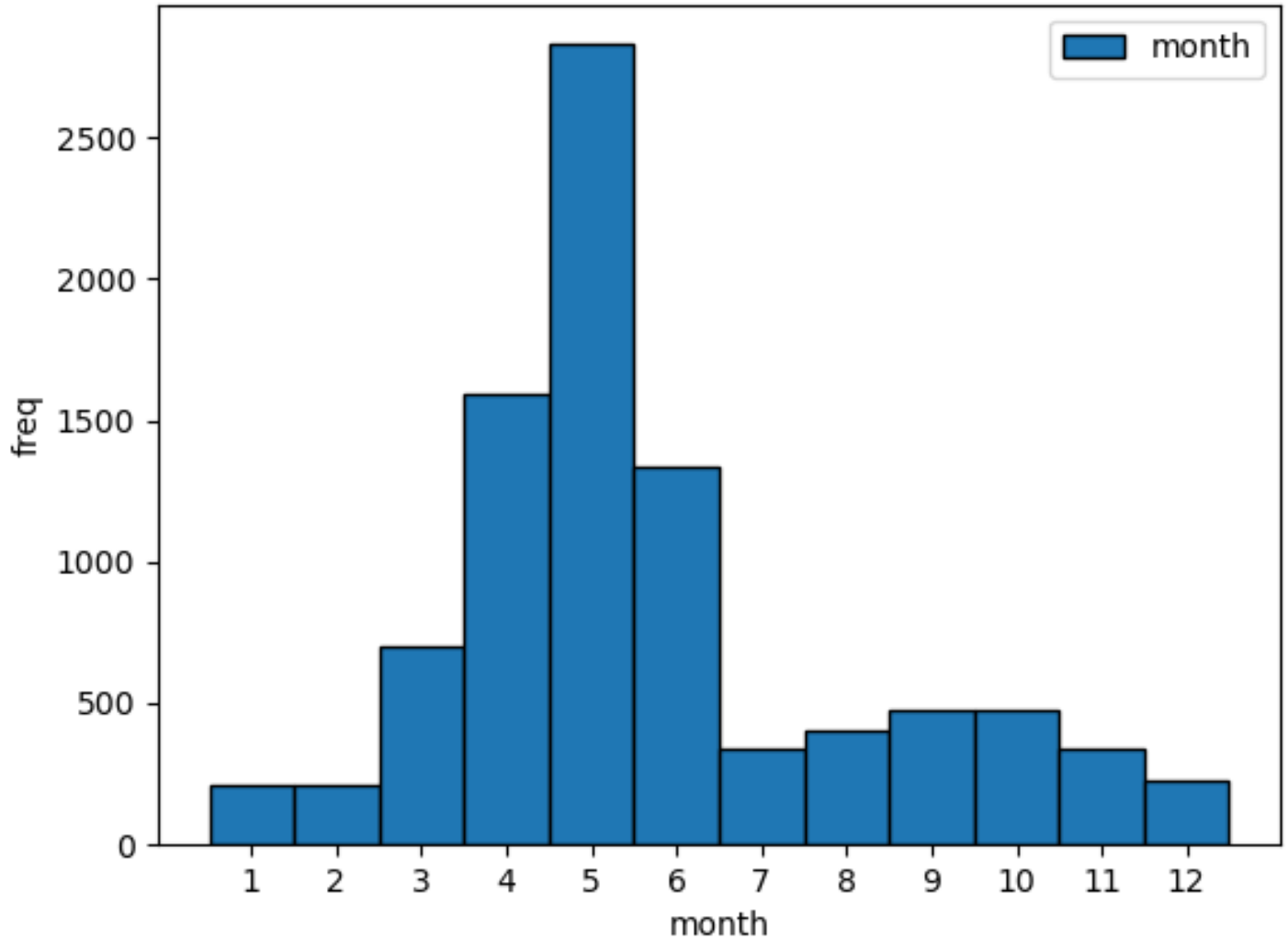

Daily Expected Fatalities (2015–2020)
Expected Fatalities
0.0
0.1
0.2
0.3
0.4
2015
2016
2017
2018
2019
2020
2021
Date

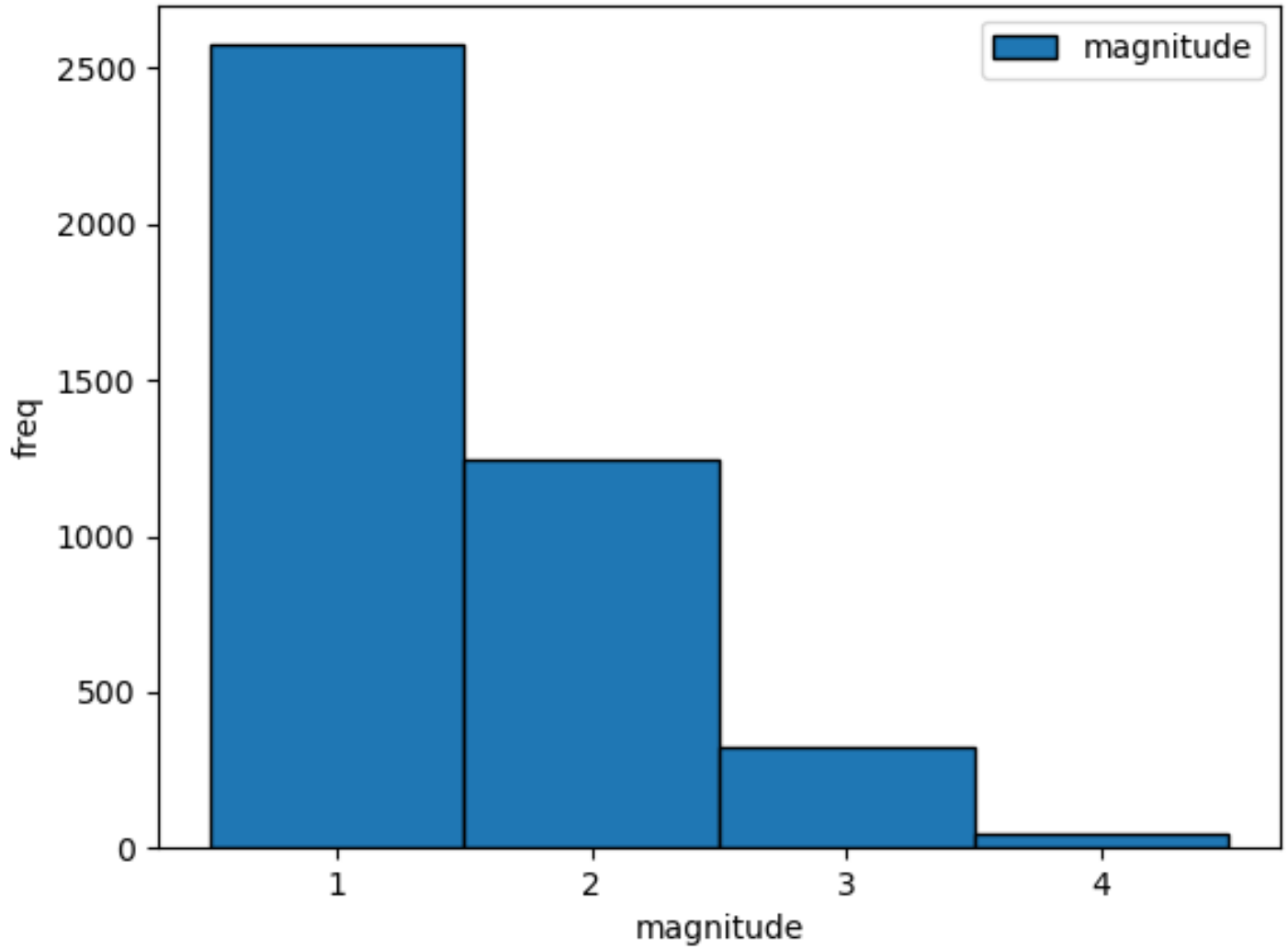

Average Expected Fatalities by Month (2015–2020)
Mean Daily Expected Fatalities
0.020
0.015
0.010
0.005
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

Final Thoughts and conclusion

- No significant correlations
- Data does not behave like a time-series
- Magnitude and path length show a trend over time
- Injury/fatality impact are difficult to predict
- Article that states climatological trend of tornado pattern moving south-east
- We can make well predictions based on weather data.


year
freq
250
200
150
100
50
0
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020


month
2500
2000
1500
1000
500
0
freq
1
2
3
4
5
6
7
8
9
10
11
12
month


magnitude
2500
2000
1500
1000
500
0
freq
1
2
3
4
magnitude