Mapping Social Media Response to Crisis Events: A Spatial Deviation Analysis of Public Reaction on X and Tornado Paths in Oklahoma

Samiha Karim Subah
School of Civil Engineering & Environmental Science, University of Oklahoma
Samiha.karim.subah-1@ou.edu

Arif Mohaimin Sadri
School of Civil Engineering & Environmental Science, University of Oklahoma
sadri@ou.edu

ABSTRACT

Tornadoes pose a recurring threat in Oklahoma, United States, underscoring the need for timely information during emergencies. This study examines how geotagged tweets on X reflect actual tornado impacts in 2022. By analyzing spatial deviations between tweet locations and verified tornado paths using geoprocessing techniques and Moran's I, the study evaluates the spatial reliability of social media data for situational awareness and disaster response. The study introduces an interactive framework through which tornado intensity can be adjusted to dynamically refine the analysis. The framework visualizes social media disaster communication patterns and enables comparison of spatial patterns across tornado intensities (EF0–EF4). The results indicate that geotagged tweets cluster strongly near verified tornado paths for low- to moderate-intensity events, while higher-intensity tornadoes exhibit greater spatial deviation and reduced real-time reporting. The study presents an interactive geospatial framework that enhances understanding of public awareness and engagement, supporting the use of social media for real-time hazard monitoring and community response mapping.

Keywords

Social media, risk communication, disasters, social networks.

INTRODUCTION

Tornadoes represent one of the most destructive and unpredictable natural hazards in the United States, with states such as Oklahoma experiencing frequent and often severe events. These disasters pose significant challenges for emergency management due to their rapid onset, localized impacts, and limited warning times. As a result, timely dissemination of information and situational awareness are critical components of effective disaster preparedness, response, and recovery (Perry & Lindell, 2004). In recent years, social media platforms have emerged as influential tools for real-time communication during disasters, enabling individuals to share observations, warnings, and requests for assistance almost instantaneously (Sadri et al., 2017).

Among social media platforms, Twitter has been widely studied for its role in disaster communication due to its real-time nature, public accessibility, and the availability of geotagged content (Mojumder & Sadri, 2021). Previous research has demonstrated that Twitter data can provide valuable insights into public awareness, information diffusion, and collective responses during emergencies (Imran et al., 2013; Vieweg et al., 2010). Natural Language Processing (NLP) techniques have further enhanced the utility of social media data by enabling automated extraction of disaster-related information, sentiment analysis, topic detection, and identification of urgent needs from large volumes of unstructured text (Alam et al., 2021).

Despite its potential, the reliability of social media data, particularly the ones containing geographic locations remains a significant concern. User-generated locations may not accurately reflect the true spatial extent of an imminent risk, as such posts can originate from observers far from the actual event or be posted after relocation. Studies have shown that while social media activities often correlate with natural hazard occurrence, spatial inaccuracies and noise can limit its effectiveness for precise geographic analysis (Crooks, 2010). This uncertainty underscores the need for rigorous spatial validation of social media data against authoritative hazard datasets.

In the context of tornado events, relatively few studies (Coleman et al., 2024; González et al., 2019) have explicitly examined the spatial deviation between social media activity and verified tornado paths. Integrating social media data with official geospatial records such as tornado path datasets offers an opportunity to evaluate how well public communication reflects actual hazard impacts. By combining GIS-based proximity analysis, spatial statistics, and NLP-supported data processing, researchers can better understand where social media serves as a reliable proxy for on-the-ground conditions and where intended or unintended misinformation and spatial bias may arise.

The year 2022 experienced several notable tornado events across Oklahoma, reinforcing the state's vulnerability to severe convective storms. Tornado activity was particularly concentrated during the spring season, with significant outbreaks occurring in April and May. For example, on May 4, 2022, multiple tornadoes were reported across central Oklahoma, including areas near Seminole, Pottawatomie, and Oklahoma counties, causing structural damage and widespread disruptions (NOAA, 2022). Additional tornado events were observed in late April 2022, affecting regions such as McClain and Cleveland counties. These events align with the climatological peak of tornado season in Oklahoma (Oklahoma Climatological Survey, 2022). By anchoring tweet patterns to known hazard events, this study provides important temporal context for interpreting how public communication reflects real-world disaster occurrences and response dynamics.

This study addresses this gap by investigating the spatial relationship between geotagged social media posts and verified tornado paths in Oklahoma during the year 2022. Using Python and the ArcPy module within ArcGIS Pro, the analysis applies a geoprocessing technique including spatial joins, buffering, clipping, proximity analysis, and spatial autocorrelation (Moran's I)(Anselin, 2010). to identify spatial patterns of public engagement and awareness. Additionally, the integration of Python geospatial libraries such as GeoPandas, PyProj, and Matplotlib supports automated data processing, geocoordinate transformation, and visualization. By incorporating an interactive framework that allows users to adjust tornado intensity scales Enhance Fujita (EF0–EF5), this study provides a flexible and reproducible approach for evaluating the spatial reliability of social media data in disaster contexts. The findings contribute to a better understanding of how social media can complement traditional hazard monitoring and support emergency management decision-making.

DATA COLLECTION AND DESCRIPTION

This study integrates three complementary geospatial datasets to systematically examine the spatial relationship between social media activity and verified tornado impacts across the state of Oklahoma. By combining user-generated geotagged social media posts with authoritative tornado path records and official administrative boundary datasets, the study enables a robust multi-scale spatial analysis. This integrated framework facilitates direct spatial comparison between public communication patterns and observed hazard extents, while also supporting aggregation and interpretation of results within established geographic units. Together, these datasets provide a comprehensive foundation for assessing the spatial reliability of social media data, identifying patterns of public awareness and engagement during tornado events, and contextualizing social media activity within verified disaster impact zones.

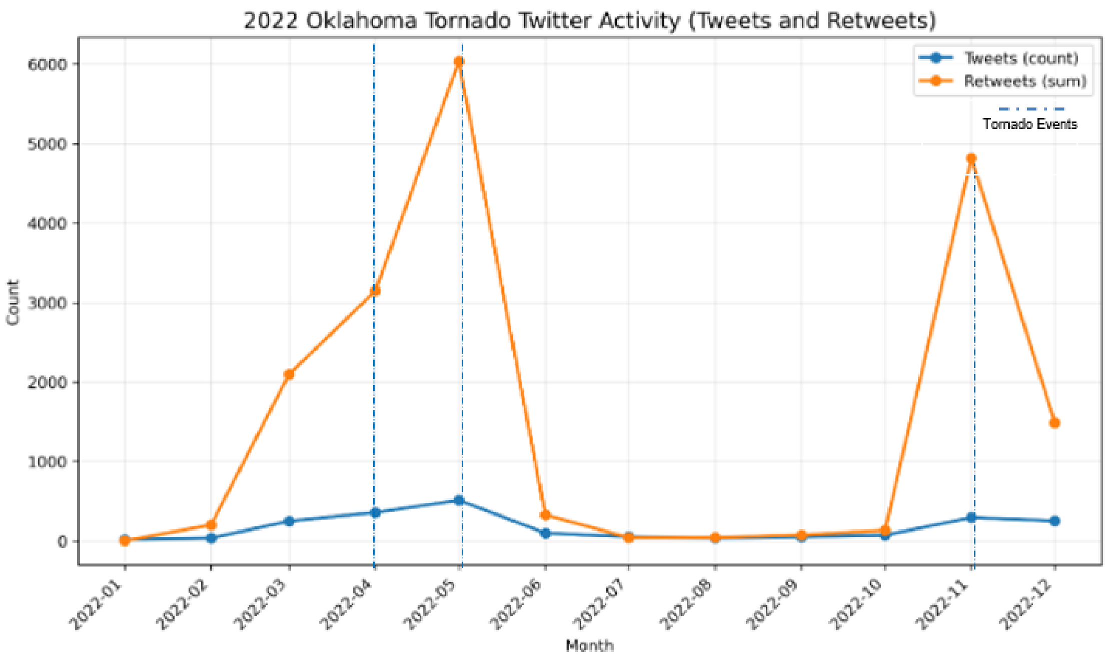


Figure 1. 2022 Oklahoma Tornado Tweet Activity (Tweets and Retweets) and Severe Tornado Events

Geotagged X Dataset (Point Feature)

The X dataset used in this study consists of geotagged tweets (also known as tweets) collected during the year 2022. The monthly distributions of these posts are shown in **Figure *1***. For this study, approximately 2100 tweets were analyzed. The dataset was collected using the Application Programming Interface (API) with authorized access through an API token. The API token was generated by registering a developer application on the Twitter Developer Platform, which provides secure authentication credentials required to access publicly available X data. Using the API, tweets were queried for the year 2022 and filtered based on both geographic location and content relevance. Geographic filtering was applied to retrieve tweets originating from within or near the state of Oklahoma, ensuring that only tweets relevant to the study area were included. Content-based filtering was performed using tornado-related keywords such as “tornado” and “storm” to isolate tweets associated with severe weather events.

The API also returned structured metadata responses containing both spatial and non-spatial attributes. From these responses, relevant fields were extracted and stored in a structured format. These fields include a unique user identifier, timestamp of the tweet, self-reported user location (state), follower and following counts, tweet text, and geographic coordinates in the form of latitude and longitude. Only tweets containing valid geographic coordinates were retained to ensure spatial usability. This dataset serves as a proxy for real-time public awareness and engagement during severe weather events, while also introducing uncertainty due to the voluntary and user-generated nature of geotagged information.

Tornado Path Dataset (Polyline Feature)

The tornado path dataset represents verified tornado tracks across the state of Oklahoma and was obtained from the NOAA Storm Events Database((NOAA), 2022).The dataset includes tornado events recorded between 2020 and 2024, with each tornado path represented as a polyline feature that captures the spatial trajectory of the tornado from initial touchdown to dissipation. Each record contains detailed geospatial and temporal attributes, including georeferenced start and end locations (latitude and longitude), event date and time, path length, and tornado intensity.

Tornado intensity is classified using the Enhanced Fujita (EF) Scale, which ranges from EF0 to EF4 and estimates tornado strength based on observed damage to structures and vegetation. Lower EF-scale values correspond to weaker tornadoes that cause limited damage, while higher EF-scale values indicate more intense tornadoes with severe and widespread impacts. The inclusion of the EF-scale attribute enables stratified spatial analysis by tornado intensity, allowing for comparison of social media activity patterns and spatial deviation metrics across different levels of tornado severity.

Oklahoma County Boundary Dataset (Polygon Feature)

The Oklahoma county boundary dataset provides the official administrative boundaries for all counties within the state of Oklahoma and serves as an essential spatial framework for aggregating and visualizing analysis results. This polygon dataset was obtained from the (Bureau, 2022) which offers authoritative and standardized geographic boundary data for use in spatial analysis. Each polygon feature represents a single county and includes identifying attributes such as county name and Federal Information Processing Standards (FIPS) codes.

This dataset is used to summarize and compare spatial metrics derived from social media activity and tornado path analysis at the county level. Specifically, county boundaries support the aggregation of tweet counts, spatial deviation statistics, and social vulnerability indicators within administrative units that are meaningful for planning and decision-making. The dataset also serves as the base geometry for rasterization and spatial statistical analyses, enabling consistent visualization of regional patterns across Oklahoma. All county polygons are projected into a common coordinate reference system consistent with the tornado path and tweet datasets to ensure spatial accuracy in distance and area calculations. By providing a stable administrative reference, the Oklahoma county boundary dataset facilitates the interpretation of spatial patterns of public awareness, engagement, and vulnerability in relation to verified tornado impacts.

METHODS

This study employed a comprehensive geospatial workflow for data analysis and visualization. The methodology integrates social media data, verified tornado path data, and spatial analysis techniques to examine the relationship between geotagged tweets and tornado impacts in Oklahoma during the year 2022. All geoprocessing tasks were conducted within a geodatabase which ensured consistent handling of feature classes generated throughout the workflow. The tornado path dataset contained verified tornado tracks for multiple years. To align the hazard data temporally with the X dataset, tornado paths were filtered to retain only events from 2022. To associate each geotagged tweet with the nearest tornado path, a spatial join was performed. Tweets served as the target features, while tornado paths were used as the join features. To examine how social media activity varies by tornado intensity, tweet counts were summarized by EF scale. This enabled efficient counting of tweets associated with each EF category. Next, spatial deviation between tweet locations and tornado paths was analyzed. Distances were grouped by EF scale and distribution patterns are presented through boxplots to compare proximity across tornado intensities. Additionally, a histogram was created to display the overall distribution of tweet distances from tornado paths. Distances were interpreted in feet, corresponding to the linear units of the projected coordinate system used in the analysis. To support modularity and reusability, advanced spatial analyses procedures were encapsulated in a custom Python programming module. To examine spatial patterns of tweet activity within proximity zones around tornado paths, buffering techniques were used i.e. generating multiple buffer rings around the selected tornado paths. The buffers define concentric proximity-based zones around each tornado path to represent graduated levels of potential impact, capturing the decay of tornado influence with increasing distance from the path. These buffers enable classification of tweet locations based on their proximity to tornado paths and support comparative analysis of engagement intensity across distance bands. All buffer rings are dissolved to produce unified proximity zones for each EF scale. We used Global Moran's I function to evaluate whether the spatial distribution of tweet-to-tornado distances exhibits statistically significant clustering, dispersion, or randomness. This method applies inverse-distance weighting and Euclidean distance measurement to assess spatial dependence across tweet locations. The output includes the Moran's I index and associated statistical significance values, which help identify spatial clustering patterns of social media activity relative to tornado paths.

RESULTS

Figure 2 illustrates the number of geotagged tweets associated with tornado events of varying intensities (EF0–EF4) in Oklahoma during 2022. A significant difference in tweet volume is observed across EF-scale categories, indicating that social media activity is not evenly distributed by tornado intensity. EF1 tornadoes account for the largest share of tweets by a substantial margin, with tweet counts far exceeding those associated with all other EF categories. EF0 tornadoes show the second-highest tweet count, reflecting their high frequency but limited impact. Although EF0 tornadoes are the weakest category. EF2 tornadoes are associated with very few tweets. EF3 and EF4 tornadoes, despite being more destructive, show relatively low tweet counts compared to EF0–EF1 events.

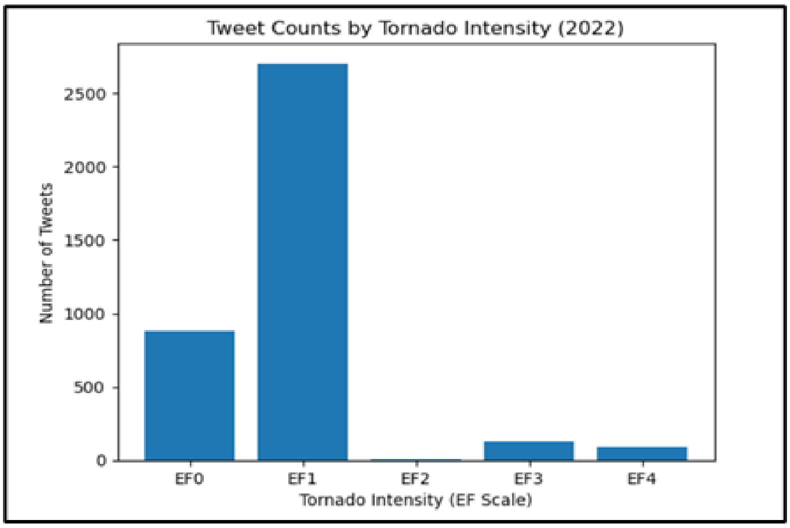


Figure 2. Tweet Counts by Tornado Intensity

Figure 3 illustrates the distribution of distances between geotagged tweet locations and the nearest verified tornado path across different EF scale categories. Distances are measured in feet and represent the spatial deviation between user-reported locations and actual tornado trajectories. Tweets associated with EF0 tornadoes are generally located very close to tornado paths, as indicated by a low median distance and a relatively narrow interquartile range. While most tweets fall within close proximity, several outliers are present. EF1 tornadoes show a slightly higher median distance and greater variability compared to EF0 events. The presence of several extreme outliers indicates that some tweets are posted at substantial distances from tornado paths. Tweets associated with EF2 tornadoes exhibit a moderate median distance and a tighter distribution compared to EF1. For EF3 tornadoes, tweet distances show a moderate spread with a median similar to EF2 but with some higher-distance outliers. EF4 tornadoes exhibit the highest median distances and the widest interquartile range among all EF categories.

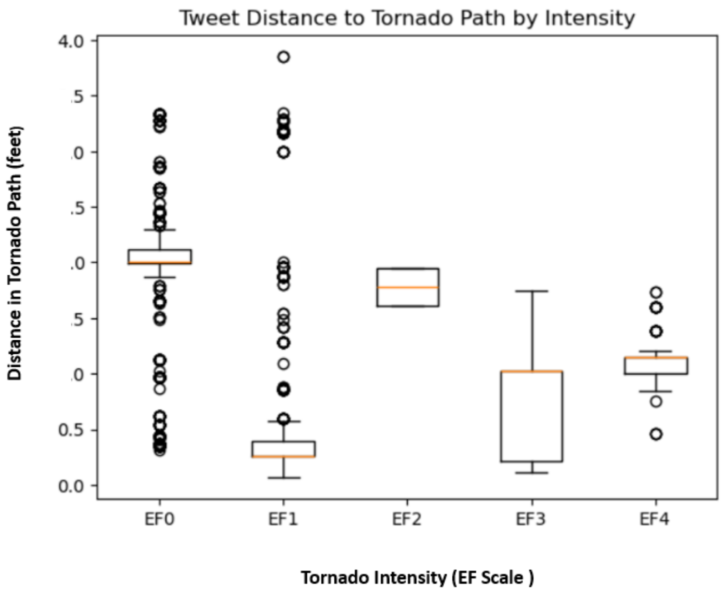
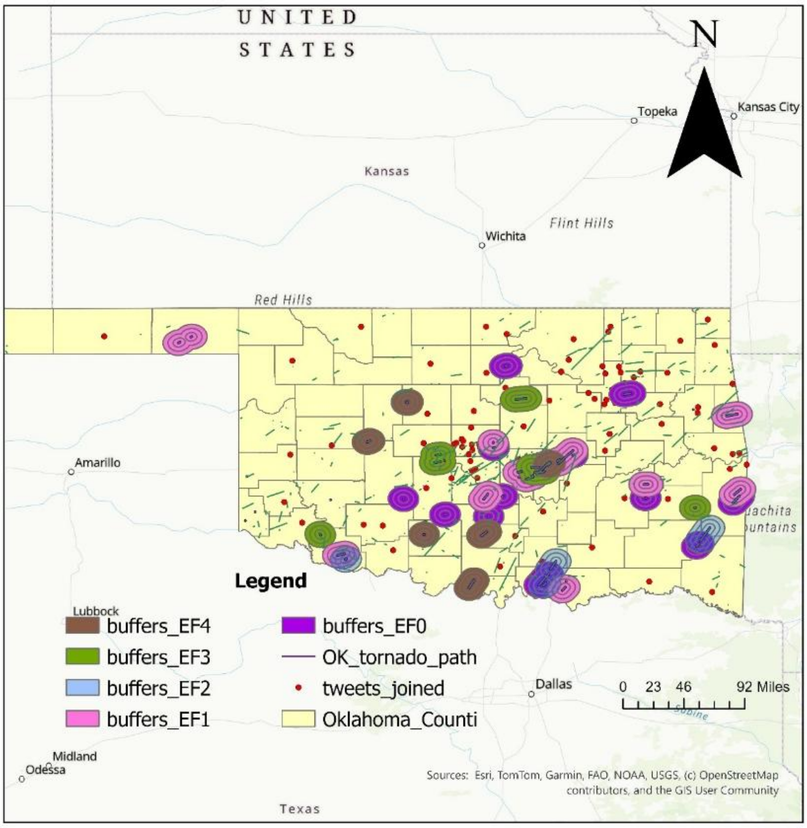
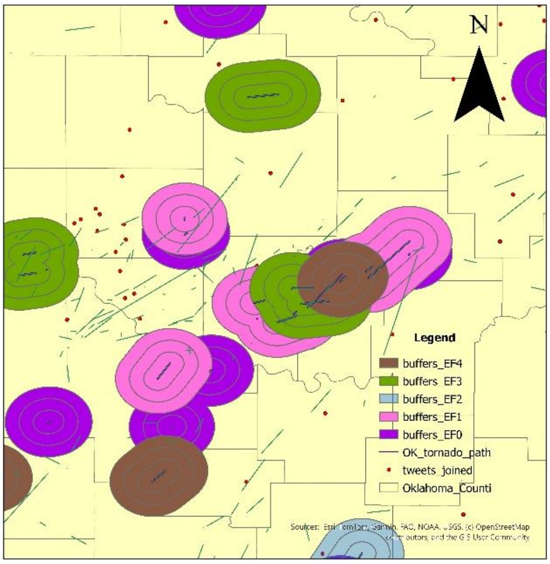


Figure 3. Tweet Distance to Tornado Path by Intensity

Distinct spatial patterns are evident when comparing tornado paths and buffer zones across different Enhanced Fujita (EF) scale categories in Oklahoma (Figure 4) Lower-intensity tornadoes (EF0 and EF1) are the most widespread spatially and occur across much of the state. Their buffers are smaller and more localized, and while they are numerous, they tend to be associated with fewer clustered tweet locations.



(a)



(b)

Figure 4. Spatial Analysis of Tornado Path and Tweet Locations in Oklahoma (Year 2022):
(a) Statewide distribution of tornado paths with multi-distance buffer zones (EF1–EF4) and geotagged tweets across Oklahoma counties, (b) Zoomed-in view of a selected region highlighting the spatial overlap between tornado path buffers (EF1–EF4) and nearby tweet locations.

Moderate-intensity tornadoes (EF2) display more spatial clustering than EF0–EF1 events, particularly in central and eastern Oklahoma. Buffers associated with EF2 tornadoes intersect with noticeably higher concentrations of tweet locations, indicating an increase in public engagement as tornado impacts become more severe. These events appear to affect larger areas and generate broader awareness compared to weaker tornadoes. Higher-intensity tornadoes (EF3 and EF4) show the most pronounced spatial patterns. Their buffer zones are larger and more clearly defined, and they are strongly associated with dense clusters of geotagged tweets. EF3–EF4 tornadoes are primarily concentrated in central and southeastern Oklahoma, where repeated and intense tornado activity occurred during 2022. The inset map highlights a localized area in central Oklahoma (e.g., Pottawatomie–Seminole region), providing a detailed view of overlapping tornado buffers and dense tweet clusters. The arrow connects the statewide overview to the zoomed-in inset, emphasizing how social media activity intensifies near verified tornado tracks.

Global spatial autocorrelation analysis revealed consistently high Moran's I values (approximately 0.994) for tweet-to-tornado distance across all Enhanced Fujita (EF) scale categories. These values indicate an extremely strong positive spatial autocorrelation, demonstrating that tweet locations with similar distances to tornado paths are highly clustered in space rather than randomly distributed.

DISCUSSIONS

Figure 1 shows monthly tweet and retweet activity related to tornado events in Oklahoma during 2022. A seasonal pattern is observed, with tweet activity increasing during spring months (March–May) and peaking in May, which corresponds to the peak tornado season in Oklahoma. This indicates that social media activity varies across seasons and aligns with periods of increased tornado occurrence. The higher volume of retweets compared to original tweets reflects information amplification within the platform. A secondary increase in activity is observed in November, which may be associated with late-season storm events or other factors. In contrast, minimal activity during winter months corresponds to periods of lower tornado occurrence. While the figure presents aggregated monthly trends, it illustrates temporal variation in social media activity in relation to seasonal hazard patterns.

The results from tweet counts by tornado intensity indicate that EF1 events are associated with higher volumes of tweets compared to other categories. This pattern may reflect their frequency and spatial distribution. EF0 tornadoes also generate noticeable levels of social media activity. In contrast, fewer tweets are associated with EF2 and higher-intensity tornadoes, which may be related to their lower frequency or spatial occurrence. These observations indicate that tweet volume varies across intensity categories and may be influenced by both hazard characteristics and population distribution. Overall, tweet frequency appears to be associated with tornado occurrence patterns and exposure rather than intensity alone.

The observed spatial patterns indicate that tweet locations exhibit increasing deviation from tornado paths as tornado intensity increases. For lower-intensity events (EF0), tweets are generally located closer to tornado paths. As intensity increases to EF1 and EF2, greater variability in tweet distances is observed. For higher-intensity tornadoes (EF3–EF4), tweets are more widely distributed relative to tornado paths. These patterns indicate differences in spatial distribution of tweet locations across intensity levels. While factors such as evacuation, infrastructure disruption, or user behavior may influence these patterns, such mechanisms cannot be directly confirmed within this study and are therefore not explicitly inferred.

The spatial analysis also shows differences in clustering patterns across tornado intensity levels. Lower-intensity tornadoes (EF0–EF1) are more widely distributed and associated with dispersed tweet activity. In contrast, moderate- to high-intensity events (EF2–EF4) exhibit stronger clustering of tweets in certain regions. This reflects variation between individual-level spatial dispersion and aggregate-level clustering of activity. These patterns indicate that tweet distributions vary both in proximity to tornado paths and in regional concentration.

The global spatial autocorrelation analysis indicates that geotagged tweet locations exhibit strong spatial dependence across all EF categories. The similarity of Moran's I values across EF0 to EF4 tornadoes suggests that spatial structure in tweet locations is not strongly differentiated by tornado intensity. Instead, the observed spatial dependence likely reflects broader geographic patterns, such as population distribution and regional communication behavior. These findings indicate that while tweet locations may deviate from tornado paths, the overall spatial distribution remains structured.

CONCLUSION

This study investigated the spatial relationship between geotagged Twitter activity and verified tornado paths in Oklahoma during 2022 to assess the spatial reliability of social media data in disaster contexts. By integrating geotagged tweets, authoritative tornado path data, and administrative boundaries, the analysis demonstrated that social media activity exhibits meaningful spatial alignment with tornado impacts, particularly for lower- to moderate-intensity events. Distance analysis and buffer-based proximity assessments showed that tweets related to EF0 and EF1 tornadoes were generally located closer to tornado paths, reflecting more precise real-time reporting during less severe events. As tornado intensity increased, greater spatial deviation between tweet locations and tornado paths was observed. Higher-intensity tornadoes (EF3–EF4) were associated with larger median distances and increased spatial dispersion of tweet locations, likely due to evacuation behavior, infrastructure damage, and safety-driven communication patterns. Despite this increased deviation, strong tornado events still generated identifiable clusters of social media activity, indicating heightened public awareness and engagement even when real-time reporting from impact zones was limited.

Despite providing valuable insights into the spatial relationship between geotagged social media activity and verified tornado paths, this study has several limitations. Firstly, the analysis relies on aggregated temporal data at a monthly scale, which does not capture fine-grained temporal dynamics of tornado events that occur over short durations (e.g., hours or minutes), limiting the ability to assess real-time communication patterns before, during, and after individual events. Secondly, geotagged Twitter data represent only a small and self-selected subset of users, as location sharing is optional and not uniformly adopted across populations, potentially introducing spatial and demographic bias. Thirdly, tweet locations may not accurately correspond to the location of the tornado event itself, as users often post from safe locations, after relocation, or while sharing second-hand information, which contributes to spatial deviation. Fourthly, the spatial analyses rely on planar distance calculations and buffer thresholds that may oversimplify complex human movement and communication behaviors during disasters. Additionally, the consistently high Moran's I values indicate strong spatial dependence, which, while informative, may be influenced by underlying population distribution and road networks rather than tornado impacts alone. Furthermore, reproducibility of the study may be constrained by platform-specific API access and data availability, as social media data collection processes and policies can change over time. Finally, the analysis is limited to a single year (2022) and one geographic region, restricting the generalizability of the findings to other time periods or hazard types. Future studies could address these limitations by incorporating multi-year datasets, additional social media platforms, and finer-scale temporal analysis to better capture dynamic patterns of public response during extreme weather events.

The application of spatial statistical methods, including proximity analysis and spatial autocorrelation, further highlighted non-random clustering patterns in tweet locations, reinforcing the value of social media as a complementary data source for hazard monitoring. Overall, the findings suggest that while geotagged Twitter data should be used cautiously due to spatial uncertainty, it can provide valuable insights into public awareness, engagement, and response during tornado events. This integrated geospatial framework offers a reproducible approach for enhancing situational awareness and supporting emergency management efforts by combining user-generated data with authoritative hazard records.

ACKNOWLEDGMENTS

The research presented in this paper was supported by the National Science Foundation under Grant No. CAREER-2339100. Any opinions, findings, conclusions, and recommendations expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

(NOAA), N. O. a. A. A. (2022). *Storm Events Database*.

Alam, F., Sajjad, H., Imran, M., & Ofli, F. (2021). CrisisBench: Benchmarking Crisis-related Social Media Datasets for Humanitarian Information Processing. *Proceedings of the International AAAI Conference on Web and Social Media*, *15*, 923–932. https://doi.org/10.1609/icwsm.v15i1.18115

Anselin, L. (2010). Local Indicators of Spatial Association—ISA. *Geographical Analysis*, *27*, 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Bureau, U. S. C. (2022). *TIGER/Line Shapefiles: Oklahoma County Boundaries*.

Coleman, T., Thompson, R., & Forbes, G. (2024). A Comprehensive Analysis of the Spatial and Seasonal Shifts in Tornado Activity in the United States. *Journal of Applied Meteorology and Climatology*, *63*, 717–730. https://doi.org/10.1175/JAMC-D-23-0143.1

Crooks, A. T. (2010). Constructing and implementing an agent-based model of residential segregation through vector GIS. *International Journal of Geographical Information Science*, *24*(5), 661–675.

González, J., Hahn, U., & Mateu, J. (2019). Analysis of Tornado Reports Through Replicated Spatiotemporal Point Patterns. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *69*. https://doi.org/10.1111/rssc.12375

Imran, M., Elbassuoni, S. M., Castillo, C., Diaz, F., & Meier, P. (2013). Extracting information nuggets from disaster-related messages in social media. *Proc. of ISCRAM, Baden-Baden, Germany*.

Mojumder, M. N. H., & Sadri, A. M. (2021). Exploring the Properties and Growth of Student Interaction

Networks on Twitter: Insights on STEM Learning and Engagement. 2021 ASEE Annual Conference & Exposition.

Perry, R., & Lindell, M. (2004). Preparedness for Emergency Response: Guidelines for the Emergency Planning Process. *Disasters*, *27*, 336–350. https://doi.org/10.1111/j.0361-3666.2003.00237.x

Sadri, A. M., Hasan, S., & Ukkusuri, S. V. (2017). Joint Inference of User Community and Interest Patterns in Social Interaction Networks. *arXiv preprint arXiv:1704.01706*.

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events: what twitter may contribute to situational awareness. Proceedings of the SIGCHI conference on human factors in computing systems.