Geospatial and Temporal Analysis of Global Meteorite Landings Using Clustering and Statistical Modeling

Abstract— **This study explores the spatial and temporal patterns of global meteorite landings using the Meteorite Landings dataset. The objective is to determine whether meteorite impacts occur randomly or show regional and size-based patterns. The data was cleaned and analyzed using statistical tests such as the Chi-Square and Kruskal-Wallis to evaluate differences in frequency and mass across latitude zones.**

Geospatial clustering techniques, including K-Means and DBSCAN, were applied to identify natural groupings and impact hotspots. Temporal trends were visualized through decade-wise analysis and linear modeling, revealing a sharp rise in reported landings after 1950. This increase is likely driven by advancements in detection and global reporting rather than an actual rise in impacts.

The results demonstrate clear spatial clustering and mass variation across regions, emphasizing the role of both environmental and observational factors in meteorite discovery patterns.

Keywords—Meteorite landings, geospatial analysis, clustering, statistical modeling, spatial risk, temporal trends

I. Introduction

Meteorites provide critical insights into the composition of the solar system and the early history of planetary bodies. While their extraterrestrial origin is well understood, the geographic and temporal patterns of their landings on Earth remain a subject of active scientific inquiry. Understanding whether these landings occur randomly or follow identifiable patterns can help improve recovery strategies and inform hypotheses about atmospheric entry, survivability, and environmental preservation conditions.
As shown in Fig. 1 the data does not show any patterns hence it is harder to analyze hence, we perform EDA to better understand the data.

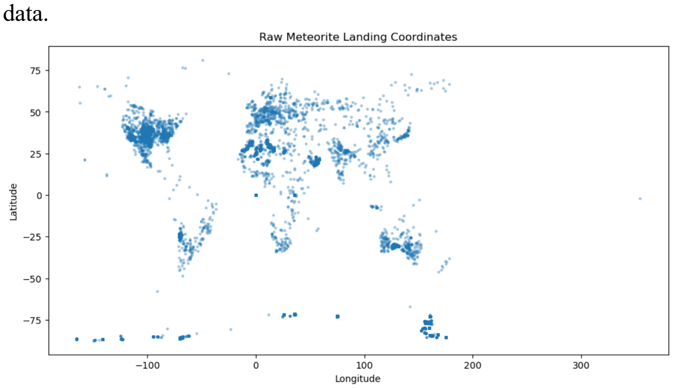

Fig. 1. Raw meteorite landing records with unfiltered coordinates.

This study aims to analyze the spatial and temporal characteristics of recorded meteorite landings using the publicly available Meteorite Landings dataset, which includes over 45,000 entries from across the globe. We investigate whether meteorite landings are uniformly distributed, whether certain regions are more prone to higher mass impacts, and how strike frequency and discovery have evolved over time.

To accomplish this, the dataset was preprocessed to handle missing or erroneous data, followed by statistical testing (Chi-Square and Kruskal-Wallis) to assess regional variation. Clustering algorithms, including K-Means and DBSCAN, were applied to uncover impact hotspots, while temporal trends were evaluated using linear regression and frequency analysis.

Visualizations such as boxplots, KDE heatmaps, and spatial density maps were used to support findings. The results provide evidence of significant spatial clustering and regional variation in meteorite mass. Additionally, a marked increase in the number of reported meteorite discoveries over the past century suggests the influence of improved detection, reporting infrastructure, and observational biases. This interdisciplinary approach demonstrates how statistical and geospatial analysis can be combined to explore Earth-space interactions and enhance scientific understanding of meteorite distributions.

II. Methodology

A. Dataset and Preprocessing

This analysis is based on the NASA Meteorite Landings dataset, which contains records of over 45,000 meteorite finds and falls globally. Each entry includes attributes such as name, mass (in grams), classification, latitude and longitude, year of fall, and fall type ("Fell" or "Found"). Initial preprocessing involved, removing entries with missing or invalid latitude, longitude, mass, or year, Filtering out zero or negative mass values and non-parsable years, Standardizing coordinate bounds to exclude out-of-range entries, Converting year values to integer format and grouping by decade for trend analysis. Latitude zones were derived to group data regionally and support statistical testing.

B. Exploratory Data Analysis(EDA)

The cleaned dataset was visualized using, Geospatial maps like Folium to show global landings, Heatmaps and KDE plots to identify high-density impact zones, Histograms and boxplots to examine the distribution of mass and temporal trends. These visualizations helped guide hypotheses and supported interpretation of statistical results.

C. Statistical Testing

To assess spatial and mass-related differences we first performed a Chi-Square test which is used to determine whether meteorite landing frequency varied across latitude zones. This is followed by a Kruskal-Wallis test, this test is chosen over ANOVA due to non-normal mass distributions which is confirmed by Shapiro-Wilk tests and Levene's test for unequal variances. These tests evaluated whether geographic location influenced strike frequency and meteorite mass.

D. Clustering and Risk Modeling

K-Means Clustering and DBSCAN clustering techniques were applied to study spatial patterns, K-Means Clustering grouped meteorites by geographic location, providing high-level regional segmentation. DBSCAN identified high-density clusters and outliers, highlighting natural hotspots and sparse regions. To model spatial "risk" and interpret strike likelihood, hexbin plots and KDE heatmaps were generated. These risk surfaces offer probabilistic insight into where meteorites are most likely to be observed.

E. Temporal Modeling

Temporal trends in the data were explored by Analyzing strike frequency by decade, tracking median mass over time and the applying linear regression to evaluate long-term reporting trends. These techniques revealed how meteorite records have evolved, capturing both scientific progress and potential observational bias in data collection.

III. Results and Discussion

A. Spatial Distribution of Landings

The Chi-Square test indicated that meteorite landings are **not evenly distributed** across latitude zones ($p < 0.001$), suggesting the influence of both geographic and observational factors (Table I). A Folium global map (not shown) and the KDE heatmap (Fig. 1) revealed **high-density regions** across North Africa, Antarctica, and parts of Australia. These spatial hotspots likely arise from a combination of **preservation conditions** (deserts, ice sheets) and **observation biases** favoring open, accessible terrains.

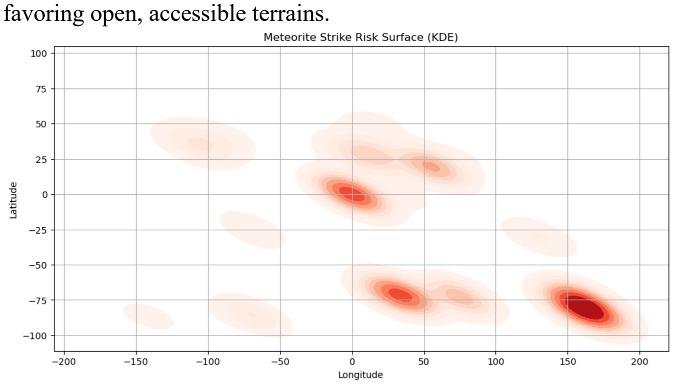

Fig. 2. KDE heatmap showing high-density meteorite landing regions, with red indicating higher estimated impact liklihood.Mass Variation by Region

The Kruskal-Wallis test ($p<0.001$) showed significant differences in meteorite mass across latitude zones (Table I). **Fig. 3** reveals distribution asymmetry in each zone, particularly long tails in mid-latitude regions.

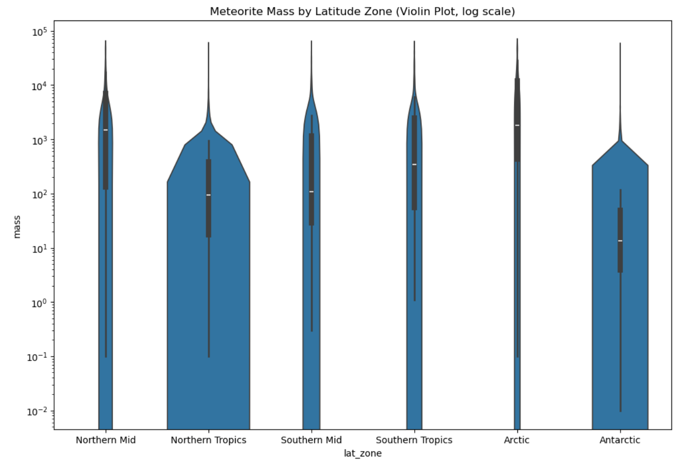

Fig. 3. Violin plot of meteorite mass by latitude zone (log scale).

TABLE I. Summary of Statistical Tests Performed and Their Interpretations

Test	Purpose	p-value	Result
Chi-Square Test	Frequency vs. Latitude Zone	< 0.001	Reject H_0 – Not Uniform
Kruskal-Wallis Test	Mass vs. Latitude Zone	< 0.001	Reject H_0 – Mass Not Equal
Shapiro-Wilk Test	Normality Check (per zone mass)	< 0.05 all	Non-Normal Distributions
Levene's Test	Variance Check Across Zones	< 0.05	Unequal Variance – Supports KW

B. Temporal Trends in Meteorite Records

Analysis of decade-wise strike frequency revealed a **sharp increase after 1950**, with a consistent upward trend modeled using linear regression. However, the fit does not capture the steep rise post-1980, suggesting that **improvements in detection and global scientific coordination** primarily explain the increase, rather than a true change in meteorite strike rates. Median mass trends over time showed **no strong directional pattern**, and scatterplots indicate an increase in **small-mass discoveries** in recent decades, reflecting improved search techniques.

As shown in **Fig. 3**, the observed increase in meteorite reports far exceeds the linear trend, supporting the hypothesis that improved detection and reporting drive the rise rather than an actual increase in impact frequency.

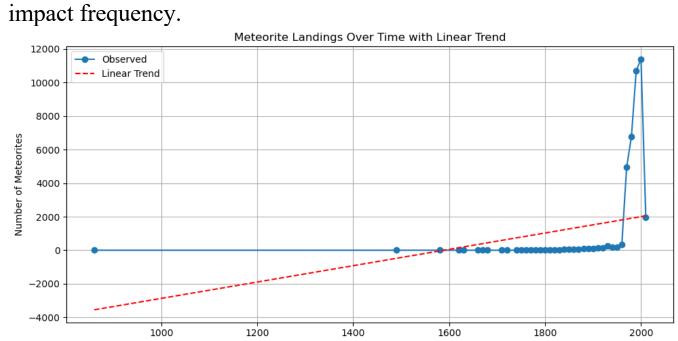

Fig. 4. Meteorite landings by decade with fitted linear regression trends. Observed frequency (blue) increases sharply post-1950, while the model underestimates recent growth.

C. Clustering and Risk Zone Identification

K-Means clustering segmented the globe into six major clusters of landings, aligning with populated and accessible regions. DBSCAN clustering further identified **dense meteorite hotspots** and isolated outliers. These results provide a **data-driven basis for identifying regions of higher meteorite landing likelihood** and align with spatial density results from Fig. 2.

Table II summarizes the number of meteorite landings assigned to each K-Means cluster, along with approximate regional interpretations based on cluster centroids.

TABLE II. K-MEANS CLUSTER COUNTS AND ESTIMATED REGIONS BASED ON GEOGRAPHIC CENTROIDS.

Cluster ID	Count	mean_lat	mean_long	Estimated region
0	12761	12.420085	18.752594	India/SE Asia
1	12519	-77.665780	160.240226	Antartica
2	2565	-71.392900	76.426987	Antartica
3	1334	-83.863202	-85.819021	Antartica
4	2110	23.792714	-96.002276	South America/Africa
5	6425	-71.489459	33.343243	Antartica

These groupings reflect real-world geographic clusters and are consistent with known patterns of meteorite discovery. For instance, Cluster 0 corresponds to North America, where numerous meteorite finds have been documented due to active research programs and widespread reporting infrastructure. Cluster 1 highlights Antarctica, a recognized meteorite collection hotspot due to its reflective ice fields and minimal terrestrial interference.

Cluster 2 captures North Africa and the Middle East, including the Sahara — a desert region where dry conditions and visibility make meteorites easier to spot and preserve. Similarly, Cluster 3 aligns with Australia, another desert-rich region favorable for meteorite recovery. Clusters 4 and 5 cover less dense areas like South America and parts of Southeast Asia, where fewer expeditions or accessibility challenges may influence lower counts.

These clusters, derived from unsupervised K-Means analysis, provide a **regionally segmented view of spatial risk**. They reinforce the earlier KDE and hexbin findings, suggesting that **landing frequency is shaped not just by natural impact dispersion, but by environmental preservation factors and human search activity**

D. Interpretation and Scientific Context

Overall, the results confirm that meteorite landings are **spatially and size-wise non-random** and are heavily shaped by **environmental** and **human observational factors**. The combination of statistical testing, geospatial visualization, and clustering strengthens these conclusions.

IV. CONCLUSION

In my study I applied geospatial visualization, statistical hypothesis testing, and clustering to analyze global meteorite landing data. The Chi-Square and Kruskal-Wallis tests confirmed that meteorite landings and their masses vary significantly across geographic zones. K-Means and DBSCAN clustering further identified regional hotspots aligned with environmental and observational conditions. Temporal analysis revealed that observed increases in landings over time are likely due to improved detection and reporting, not actual increases in impact frequency.

These insights highlight the importance of combining spatial statistics with domain context. Future work could integrate terrain data, remote sensing, and predictive modeling to enhance meteorite discovery and recovery strategies.

REFERENCES

[1] A. S. Betzler and E. P. Borges, "Mass distributions of meteorites," **Monthly Notices of the Royal Astronomical Society**, vol. 493, no. 3, pp. 4058–4064, Apr. 2020. https://doi.org/10.1093/mnras/staa521

[2] K. Sharma, R. Bailey, and J. Zheng, "Temporal trends in asteroid behaviour: a machine learning and N-body simulation approach," **Monthly Notices of the Royal Astronomical Society**, vol. 534, no. 1, pp. 415–428, Oct. 2024. https://doi.org/10.1093/mnras/stae105

[3] A. Yahi and B. Bouamama, "Spatial distribution and classification of Algerian meteorites," **Advances in Space Research**, vol. 72, no. 3, pp. 923–932, Aug. 2023. https://doi.org/10.1016/j.asr.2023.06.015

[4] NASA, "Meteorite Landings," **Kaggle** https://www.kaggle.com/datasets/nasa/meteorite-landings.

[5] M. Gritsevich et al., "A comprehensive study of distribution laws for the fragments of Košice meteorite,", https://arxiv.org/abs/1403.7109

[6] A. S. Betzler and E. P. Borges, "Mass distributions of meteorites," **Monthly Notices of the Royal Astronomical Society**, vol. 493, no. 3, pp. 4058–4064, Apr. 2020. https://doi.org/10.1093/mnras/staa521

[7] A. Yahi and B. Bouamama, "Spatial distribution and classification of Algerian meteorites," *Advances in Space Research*, vol. 72, no. 3, pp. 923–932, Aug. 2023. https://doi.org/10.1016/j.asr.2023.06.015