Novel Empirical Models & Comparative Probabilistic Analysis of Interconnectedness of Volcano Eruption & Nearby Earthquakes of the Earth

Prithwish Ghosh[1], Debashis Chatterjee[2,*], Amlan Banerjee[3], and Shiladri Shekhar Das[4]

[1]Visva Bharati, Department of Statistics, Santiniketan, 731235, India
[2]Visva Bharati, Department of Statistics, Santiniketan, 731235, India
[3]Indian Statistical Institute, Geological Studies Unit, Kolkata, 700108, India
[4]Indian Statistical Institute, Geological Studies Unit, Kolkata, 700108, India
[*]debashis.chatterjee@visva-bharati.ac.in

ABSTRACT

This paper takes a probabilistic approach to validate novel empirical models and directional distributional similarities of nearby earthquake counts concerning a typical volcano with its eruption duration. We consider datasets on volcanic eruptions and earthquakes, mainly focusing on earthquakes within a 100-kilometer radius and within a three-year time frame of the volcano eruption. We propose empirical probabilistic models for the same; statistical model validation tests favor our proposed models. Moreover, we take a novel directional statistical approach to characterize the inter-connectedness and distributional similarities of volcanic eruptions and earthquakes near volcanoes, utilizing the directional nature of the datasets. We project and partition the volcanic eruption and earthquake data to assess its directional distribution. The analysis demonstrated that the data adhered to a Von Mises distribution and unsupervised equal partition revealed for both datasets, highlighting the interconnected nature of volcanic eruptions and earthquakes. We applied the Von Mises-Fisher distribution fit test; the analysis produced partition results that closely aligned with the partitions obtained through the 2D projection. This congruence emphasizes the robustness of our findings in a spherical context. Our proposed empirical models and conclusions on distributional similarities may provide insights into the underlying mechanisms connecting these geological phenomena.

1 Introduction

Accurately assessing volcanic hazards is the primary step in preventing disasters [1]. Most earthquakes emerge along the edges of tectonic plates, where most volcanoes are situated. However, at least theoretically, not all earthquakes can be related to volcanoes. The plates' interaction, not magma's movement, causes most earthquakes. On the other hand, the movement of magma also causes most earthquakes near a volcano[2,3]. The inside-formed magma wields pressure on the rocks until it decrypts the rock. Then, the magma splashes into the rupture and produces pressure again, making a small earthquake too weak to be felt but only can be detected and recorded by sensitive instruments. Deep volcanic tremor and magma ascent mechanism case study leading to an earthquake is addressed in many literatures like[4,5,6,7]. Once the volcano is open and magma flows through it, constant earthquake waves, called harmonic tremors, are recorded[2,7]. Earth Science-related literature on the relationship between volcano eruption and nearby occurring earthquakes is plenty. A review of how earthquakes trigger volcanic eruptions has been addressed in[7].[8] gathered observations concerning how mud volcanoes and various geological systems (including earthquakes, volcanoes, liquefaction, groundwater, and geysers) respond to seismic activity. The collected data of[8] reveals a distinct threshold of magnitude and distance that triggers these responses. The categorization developed in[7] demonstrates that most volcano types can be triggered by seismic activity, although they necessitate specific combinations of volcanic and seismic conditions. Triggering is improbable unless the volcanic system is in a state primed for eruption. Seismically-induced unrest is more prevalent, especially in connection with hydrothermal systems. Interactions between earthquakes and volcano activity are statistically addressed in [9], [2]. [10] addressed Changes over time in the characteristics of shallow volcano-tectonic earthquakes linked to the escalation of volcanic activity at the Kuchinoerabujima volcano, Japan. Researchers are actively developing statistical methods and diverse models within the time-space-magnitude parameter space of earthquakes and advances in pursuit of suitable stochastic modeling techniques that rely on the history of earthquake occurrences and pertinent geophysical information. This endeavor aims to describe and forecast earthquake activity accurately. These advancements aim to analyze seismic activity using regularly accumulated earthquake hypocenter catalogs, e.g., see[11], [12]. In particular, [12] reported a truncated exponential frequency-magnitude relationship observed in earthquake statistics. [13] addressed earthquake statistics and its importance. A statistical model is designed to characterize ground motion generated by earthquakes at local

and regional distances [14]. Recent model-based earth science research is adapting directional statistical tools. For instance, scaled von Mises–fisher distributions and regression models for paleomagnetic directional data have been proposed in [15]. The probabilistic assessment of volcanic hazards made in[16] serves to quantify volcanic hazards and elucidate uncertainties about the magnitude and potential consequences of volcanic activity. In particular, [16] presents an approach developed to estimate volcanic hazards related to tephra fallout and illustrates this approach with a tephra fallout hazard assessment for the city of Leon, Nicaragua, and the surrounding area.[17] proposed a Bayesian approach for the determination and parameterization of earthquake focal mechanisms. [18] demonstrated a numerical simulation of earthquake-induced excitation on tire-reinforced sand behind a retaining wall. [19] addressed elastoplastic models for pressure sources in a heterogeneous domain were developed to characterize, evaluate, and interpret observed deformation in volcanic regions and used the Finite Element Method (FEM) to simulate deformation in a three-dimensional domain, partitioned to incorporate volcano topography and the distribution of heterogeneous material properties. Based on novel measurements, statistical estimations, and models,[2] supported the speculation that a large earthquake can trigger subsequent volcanic eruptions over surprisingly long distances and time scales. However, the stress changes from usual earthquakes are typically smaller than those associated with solid-earth tides (about 0.001 MPa) and cannot directly influence a typical volcano eruption[2],[5],[4].[20] investigated the interaction between earthquakes and volcanic eruptions by analyzing a modern seismic data catalog and eruption records. The conclusion of[20] is that moderate earthquakes having moment magnitudes of 5 to 6 (Mw 5 to 6) are triggered within a 50 km horizontal distance from volcanoes for approximately 0.3 years after the initiation of an eruption and roughly 13% of volcanic eruptions are escorted by medium earthquakes. The likelihood of quakes every 0.1 years for 0.3 years after an explosion is about five times larger than regular. There are many natural cases, such as a recent example of earthquake-mud volcano triggering that surfaced when a mud volcano on the island of Bharatang, on the island of the Middle Andaman[21].

1.1 Objective & Overview of this Paper

1. Holistically, the volcano, and nearby earthquake distribution fall under a directional statistical paradigm because of the inherent directional parameters such as latitude and longitude involved in the data. To our knowledge, little literature exists that proposes a novel regression model capturing the inter-dependencies of volcanoes and nearby earthquakes and uses directional statistical models to frame the research question on the inter-relationship of volcanos and nearby earthquakes and proceed toward answers. In this paper, we try to initialize to fill in the gap. In addition, we apply the Von mises-fisher clustering models presented in [22] on Volcano eruptions and their nearby earthquake occurrences and address the goodness-of-fit. We propose and validate a novel empirical model of volcanos and their nearby earthquakes. We perform a comparative statistical analysis of directional earthquakes and volcanic eruptions. We probabilistically searched for the geological directional relationship between volcanic eruptions and earthquakes by creating a secondary dataset that considers location and time factors. We propose empirical statistical models and explore directional distributional properties and similarities among the occurrence of volcano eruption and its nearby earthquake numbers and magnitude.

2. Here, we choose a well-known theory of dependencies of volcanoes with hassle related to tectonic-plate junctions. The same explanation applies to earthquakes. Hence, theoretically, there should be an abundance of observed earthquakes before a volcanic eruption and vice versa. In this paper, we propose empirical statistical models and wish to apply directional statistical tools to verify the same.

3. We propose empirical statistical models to express the interconnecting nature of volcanic eruptions and earthquakes. Data were collected on volcanic eruptions and earthquakes, focusing on events occurring within a 100-kilometer radius and within a one-year timeframe. By examining the count data for the occurrences of these natural phenomena, we intend to uncover and quantify the distributional similarities and employ directional statistical approaches.

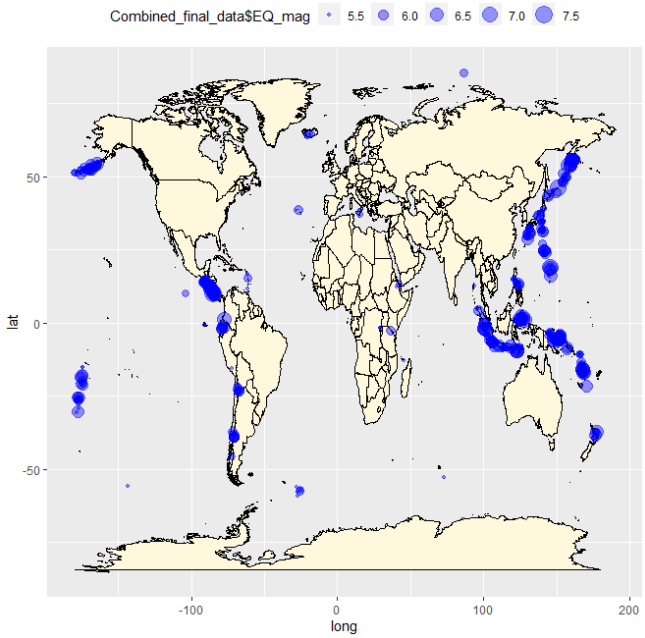
4. In addition, combined volcanic eruption and earthquake data were projected and partitioned to assess its distribution. Remarkably, the analysis demonstrated that the data adhered to a Von Mises distribution, highlighting again the interconnected nature of volcanic eruptions and earthquakes.

5. We applied the Von Mises-Fisher distribution fit test to delve deeper into the spherical domain. This analysis yielded partition results that closely aligned with the partitions obtained through the 2D projection. This congruence underscores the robustness of our findings in a spherical context.

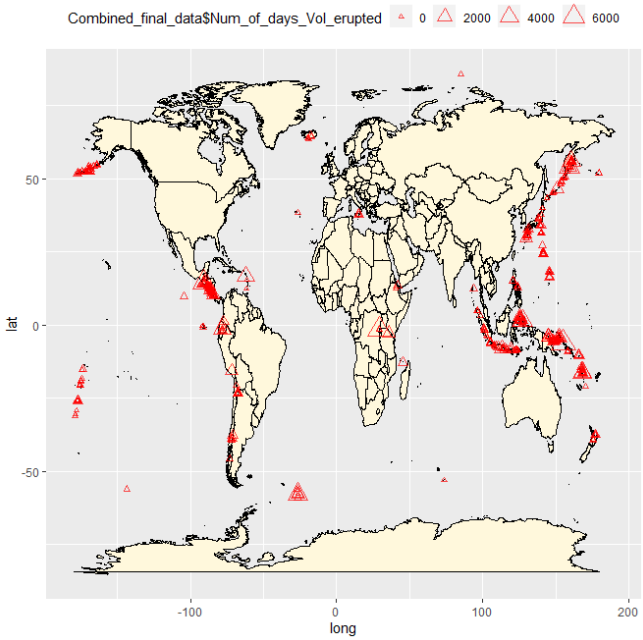
2 The Dataset

2.1 About the Earthquake and Volcano Irruption Dataset

The National Earthquake Information Center (NEIC) is tasked with identifying the precise locations and magnitudes of noteworthy earthquakes across the globe and sharing this information with scientists and scientific organizations. This database is estab-



(a) Plot of the Recorded Earth quack observed worldwide concerning the Earthquake Magnitude. Here, the point sizes will increase concerning the magnitude. The bigger magnitude means a bigger plot of that dataset.



(b) Plot of the Recorded Volcano Eruption all over the World Map concerning the number of days Volcano eruption happened. The larger the triangle is, the larger the Volcanic Eruption time

Figure 1. Plot of the Earthquake and Volcano all over the worldmap

lished through advanced national and global seismograph networks and collaborative international agreements. The NEIC earthquake dataset is collected from https://www.usgs.gov/programs/earthquake-hazards/earthquakes, given in the picture 1a and figure1b. The second dataset provides us with the volcano eruption; the volcanic dataset is collected from https://volcano.si.edu/, and the view of the volcano eruption is given in Figure 1b. Both data sets are uploaded to the Harvard dataverse https://doi.org/10.7910/DVN/SPLMMN.

2.2 About the novel secondary Data set Creation

1. First, we considered two datasets about the volcanic eruption and the earthquake information. Initially, The earthquake data set had 37331 data points, while the Volcano data set had 1193 data points.

2. Where we first sorted the dataset and considered the common time points for both the earthquake and volcanic eruption datasets.

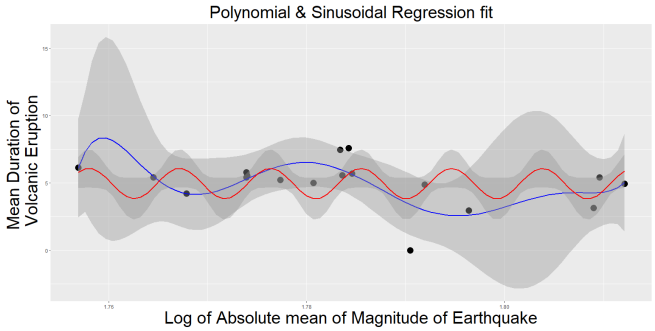
3. Then we took a hypothetical manual threshold of 100 KM around the volcanic eruption and sorted and found out how many earthquakes happened within that threshold.

4. Then, for the time points, we consider another threshold which tells us, considering the previous threshold, how many earthquakes happen within the time point 365 days or one year.

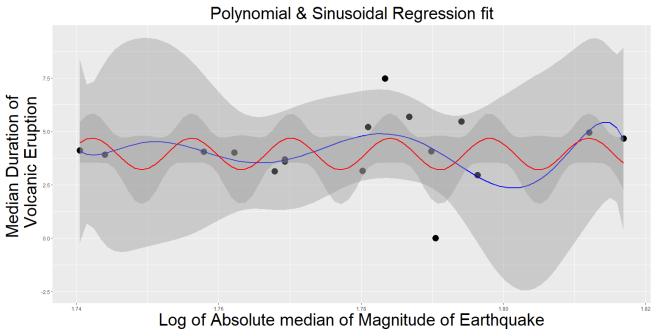
5. Considering those thresholds, we get the earthquake and volcanic dataset, which occurs within the 100 km radius and within the time point of 1 year. This gave us the final combined dataset of 478 data points and 16 parameters, giving us information about the volcano Eruption and Earthquake.

6. We hypothesize that earthquakes can trigger volcanic eruptions, and volcanic eruptions can also trigger earthquakes. So we partitioned the data into two parts: one is the volcano that triggered the earthquake, and another one is the earthquake triggered by the volcano, which led to two datasets, one with 215 data points and 16 parameters and another one with 263 data points with 16 parameters.

7. The pictorial illustration of the secondary dataset creation is given in the figure 3b as an example, where We exhibited it for one exemplar Volcanic Eruption (Rincon De Vieja).



(a) Plot of the absolute mean magnitude of earthquake vs mean period of volcano eruption. We used the absolute mean of the Earthquake's Magnitude nearby vs the mean of the Volcano's Duration for the observed Earthquake count near the Volcano. The threshold distance is 100 km and three years. We used the polynomial regression with 7 degrees to plot(for 5 degree polynomial regression fit, R^2 is 0.341).



(b) Plot of the median of the magnitude of earthquake vs median of the period of volcano eruption. We used the Median of the Earthquake's Magnitude nearby vs the Median of the Volcano's Duration. The threshold distance is 100 km and three years. The multiple R^2 is 0.1597. We used the polynomial regression with 7 degrees.

Figure 2. (a) Plot of absolute mean earthquake magnitude vs mean volcano eruption period, using polynomial regression with 7 degrees (for 5 degrees, R^2 is 0.341), considering earthquakes within 100 km and three years of the volcano. (b) Plot of earthquake magnitude median vs volcano eruption period median, using polynomial regression with 7 degrees (multiple R^2 = 0.1597), considering earthquakes within 100 km and three years of the volcano.

3 Methodologies

3.1 Novel Empirical Models on Earthquake Count & Volcano Eruption Time

A simple visualization of the logarithm of volcanic eruption time vs. nearby (100 km range) earthquake count over the past three years (figure 3a) reveals a notable pattern, suggesting a noteworthy probabilistic theoretical association. The following observations are apparent:

1. The Mean duration of the logarithm of volcano eruption time seems sinusoidally increasing with the number of earthquakes observed before eruption starts.

2. The precision in determining the duration of volcano eruption increases as the number of earthquakes observed increases.

We define two random variables: the volcanic eruption time (days) is denoted by D, and the nearby (100 km range) earthquake magnitudes over the past three years as M_{i,N_i} (N_i denote the corresponding earthquake count) for the i the volcano. If we have N many volcanoes in our dataset, we denote $\mathbf{D} = \{D_1, D_2, \cdots, D_N\}$. For each D_i, suppose we have $\mathbf{M}_i := \{M_{i,1}, M_{i,2}, \cdots, M_{i,N_i}\}$ many earthquakes count over the past three years (100 km range), which we denote as N_i. We use notations:

1. $A(\log M_i) := \dfrac{1}{N_i} \displaystyle\sum_{k=1}^{N_i} \log M_{i,k}$ (log mean),

2. $B(\log M_i) := median\{\log \mathbf{M}_{i,1}, \log \mathbf{M}_{i,2}, \cdots, \log \mathbf{M}_{i,N_i}\}$ (log median).

Model 1 (1st empirical Model on Volcano eruption Time) *Consider the conditional random variable* $[\log D_i | \boldsymbol{M}_i]$. *We hypothesize*

$$E(\log D_i | \boldsymbol{M}_i, N_i) = p(A(\boldsymbol{M}_i)),$$
$$[\log D_i | \boldsymbol{M}_i, N_i] \sim N\left(p(A(\boldsymbol{M}_i)), \frac{\sigma^2}{N_i^{1.35}}\right),$$

and similarly,

$$E(\log D_i | \boldsymbol{M}_i, N_i) = p(B(\boldsymbol{M}_i)),$$
$$[\log D_i | \boldsymbol{M}_i, N_i] \sim N\left(p(B(\boldsymbol{M}_i)), \frac{\sigma^2}{N_i^{1.35}}\right),$$

Where $p(\cdot)$ is a polynomial of the unknown degree to be determined using the goodness of fit, $N(\cdot)$ denotes normal distribution, and σ is an unknown model parameter to be estimated using MLE (Maximum likelihood estimation).

Model 2 (2nd empirical Model on Volcano eruption Time) *This model is the same as model 1, except the mean is hypothesized to be a sinusoidal function. Consider the conditional random variable $[\log D_i|M_{i,N_i}]$. We hypothesize*

$$E(\log D_i|\boldsymbol{M}_i) = a\sin(b(A(\boldsymbol{M}_i)+c))+d,$$
$$[\log D_i|\boldsymbol{M}_i,N_i] \sim N\left(p(a\sin(b(A(\boldsymbol{M}_i)+c))+d, \frac{\sigma^2}{N_i^{1.35}}\right),$$

and similarly,

$$E(\log D_i|\boldsymbol{M}_i) = a\sin(b(B(\boldsymbol{M}_i)+c))+d,$$
$$[\log D_i|\boldsymbol{M}_i,N_i] \sim N\left(p(a\sin(b(B(\boldsymbol{M}_i)+c))+d, \frac{\sigma^2}{N_i^{1.35}}\right),$$

where $a,b,c,d,\alpha,\beta,\sigma$ are unknown model parameters. We estimated using MLE (Maximum likelihood estimation).

Remark 1 *In our empirical models 1 and 2. The factor $\frac{\sigma^2}{N_i^{1.35}}$ is empirical (based on observation). Had it been replaced by an unknown parameter, 1.35 is a close approximation to the maximum likelihood estimate (MLE) of that parameter.*

Remark 2 *We have done goodness-of-fit for polynomial regression. We obtained the best degree for polynomial $p(M_i)$ is 7 (0.341 is the multiple R^2 Value) and 7 (0.1597 is the multiple R^2 Value) for the Absolute mean and Median of volcanic eruption duration, respectively.*

Model 3 *Here, we hypothesize the joint distribution of the logarithm of the mean of the Earthquake's magnitude nearby and the i th Volcano's Duration D_i, $(i \in \{1,2,\cdots,N\}$ follows a bivariate normal distribution. The threshold distance is 100 km and three years. For convenience of notation, we may denote the sample points with (x_i,y_i) defined as (mean of Earthquake Magnitude, Volcano Duration).*

$$[\log D_i, \log A(M_i)] \sim \phi(\log D_i, \log A(M_i)), \tag{1}$$

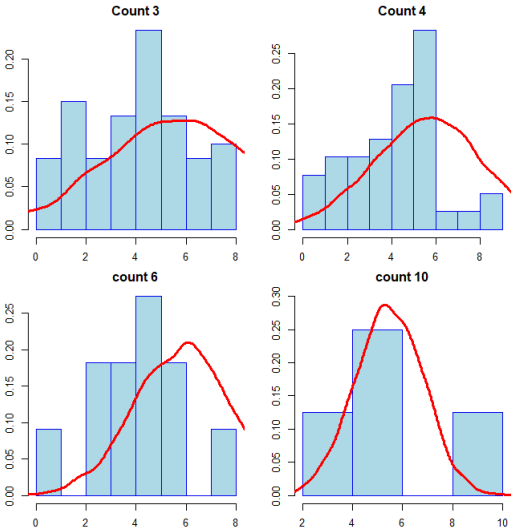
where bivariate normal density $\phi(x,y) = \frac{1}{2\pi|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)$. Here $\mu,\Sigma = \begin{bmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2\end{bmatrix}$ are unknown parameters.

Remark 3 *We obtained the estimated variance is $\sigma_x = 0.0985, \sigma_y = 3.75$, and the correlation $\rho = 0.00901$. The two parameters are taken with logarithmic transformation. Fitting Bivariate normal distribution (see [23]). We used the Anderson Darling and Cramer Von Mises test[24] for the goodness of Fit, yielding results in our hypothesis. For Anderson- Darling test [25], which gives us the p-value = $9.999e^{-5}$ and for the Cramer Von Mises test gives us the p-value = $9.999e^{-5}$.*

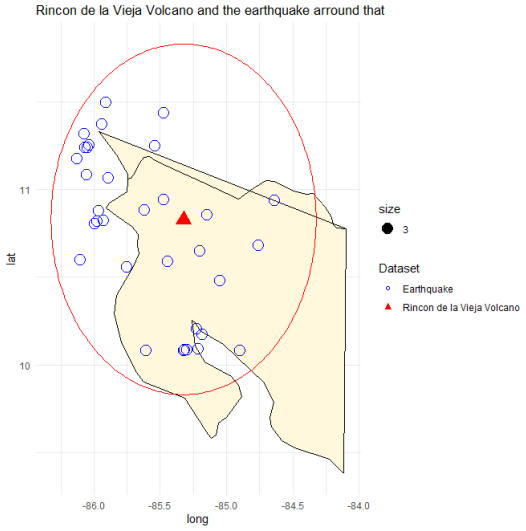
The estimated mean vector and covariance matrix can then be used to characterize the fitted bivariate normal distribution.

Remark 4 *We apply the KS test (Kolmogorov–Smirnov test, see [26]) for model validation of our proposed empirical model 1. The p-values of the count are in the table 1, which asserts our model 1 assumption satisfies all the cases. Note that Model 2 is very similar to model 1 (same variance assumption).*

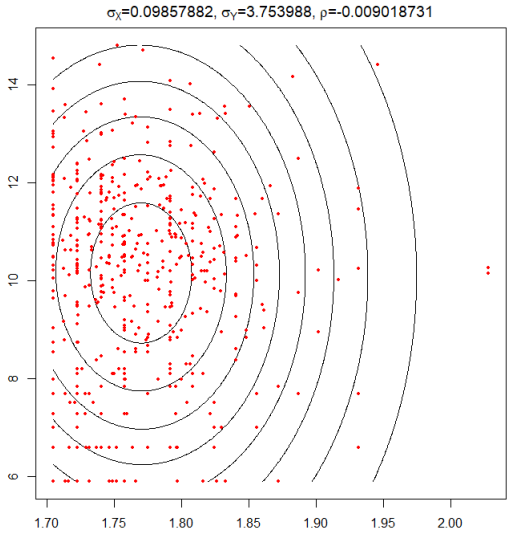
Figure 3a is the density Plot of the Duration of volcanic eruption concerning the count of Earthquakes observed within a 100 km radius and three years of the volcano. Here, we considered the earthquake count of 3,4,6 and 10. Here, the red line indicates the sample drawn from normal distribution from the proposed model given in the model 1. The figure 2a shows the fitted regression line for the 5-degree regression model where we used the absolute mean of the Earthquake's Magnitude nearby vs. the mean of the Volcano's Duration for the observed Earthquake count near the Volcano. The threshold distance is 100 km and three years, and figure 2b shows the regression model of a 10-degree polynomial where we used the Median of the Earthquake's Magnitude nearby vs the Median of the Volcano's Duration. The threshold distance is 100 km and three years. The multiple R^2 is 0.3649. The detailed table for all polynomial regressions for both, starting from 1 degree to 10 degrees, is given in the supplementary materials. Figure 3c shows the Plot of the Count dataset where we used the arithmetic mean of the Magnitude of the Earthquake vs the Duration of the Volcano and overlayed the Contour plot of Bivariate Normal Distribution.



(a) Plot of Volcanic Eruption Duration(Logarithmic Transformation) for counts of Earthquake happened within 100 km and three years of the volcano. Here, we considered the earthquake count of 3,4,6 and 10. Here, the red line indicates the sample drawn from normal distribution from the proposed model given in the model 1. We statistically verify that our model 1 assumption is valid for all the cases (table 1). see remark 4 for detailed explanation



(b) This plot illustrates one instance of creating the dataset. We have plotted the location (latitude, longitude) of 1 typical volcano (Rincon De Vieja), indicated by the Red Triangle, and the earthquake that happened near its 100 km (denoted with the blue circles) that happened in the coastal area of U.S



(c) Plot of the Arithmetic mean magnitude of earthquake vs period of volcano eruption. We used the Arithmetic mean of the Earthquake's Magnitude nearby vs the Volcano's Duration. The threshold distance is 100 km and three years. Here (x, y) is defined as (Arithmetic mean of Earthquake Magnitude, Volcano Duration). Here the estimated variance is $\sigma_x = 0.0985$, and $\sigma_y = 3.75$ and the covariance ρ is 0.00901. The two parameters are taken with logarithmic transformation. Fitting Bivariate normal distribution, we used the Anderson- Darling test, which gives us the p-value = $9.999e^{-5}$ and the Cramer Von Mises test gives us the same p-value = $9.999e^{-5}$.

Figure 3. (a) Log-transformed eruption duration vs earthquake counts for validating Model 1. (b) Location of Rincon De Vieja volcano (Red Triangle) and nearby earthquakes (Blue Circles) within 100 km in the U.S. coastal area. (c) Earthquake magnitude vs volcano duration (100 km, three years). Bivariate normal distribution fitted, p-value = $9.999e^{-5}$ (tests).

3.2 A spherical Mixture model based Statistical Approach

Formally, a mixture model corresponds to the mixture distribution representing the probability distribution of observations in the overall population. The Gaussian mixture model is commonly extended to fit a vector of unknown parameters (here, Von Mises Fisher Distribution)[27].

we define the density for directional parametric mixture distribution $p(x|\theta)$, where $\theta = (\theta_1, \theta_2, \cdots, \theta_K)$ represent appropriate parameter set, as follows.

$$p(\theta) = \sum_{i=1}^{K} \phi_i F(x|\theta_i). \tag{2}$$

- K = number of mixture components
- N = number of observations
- $\theta_{i=1\ldots K}$ = parameter of distribution of observation associated with component i
- $\phi_{i=1\ldots K}$ = mixture weight, i.e., the prior probability of a particular component i
- Φ = K-dimensional vector set $\phi_{1\cdots K}$; sum to 1
- $z_{i=1\ldots N}$ = component of observation i
- $x_{i=1\ldots N}$ = observation i
- $F(x|\theta)$ = probability distribution of an observation, parametrized on θ
- $z_{i=1\ldots N}$ = Categorical(ϕ)
- $x_{i=1\ldots N}|z_{i=1\ldots N} = F(x_{i=1\ldots N}|\theta_{z_i})$

In our paper, the i th vector component is characterized by $F(x|\theta_i)$ as either von Mises-Fisher distribution with weights ϕ_i, parameter θ_i as means μ_i and concentration matrices κ_i, or as circular uniform.[28]
For fitting the mixture Von Mises Fisher distribution to both datasets, we take the help of R package **movMF**[29].

4 Results

4.1 Directional Statistical Results

We refer to[30] and[31] for essential definitions related to directional statistics. Moreover, we state some essential definitions and properties of certain directional statistical distributions and tests in the Supplementary file (see section 6. Circular data analysis techniques are employed to study the characteristics of observed Earthquake locations, which inherently exhibit cyclical or angular properties. In earthquake scenarios, these locations are often measured in degrees along a circular scale, mirroring the Earth's surface.

Count	Pvalues	Count	Pvalues	Count	PValue
1	2.2e-16	6	0.1438	11	0.5502
2	1.039e-06	7	0.001103	12	0.2962
3	0.1697	8	0.1058	13	0.1798
4	0.5475	9	0.05988	14	0.8619
5	0.1123	10	0.05257	15	0.0774
16	0.2186	17	0.2322	18	0.9745

Table 1. Kolmogorov–Smirnov-test for every count with respect to the model 1, from where the men and variances are estimated

Parameter	Test Statistics	Critical value
Longitude	0.237	0.081
Latitude	0.0454	0.09

(a) Watson Test Result for Dataset of Earth Quack dataset for 0.01 level of Significance both of location parameters, where Longitude follows Von Mises but Latitude does not.

Parameter	Test Statistics	Critical value
Longitude	0.3133	0.081
Latitude	0.1395	0.09

(b) Watson Test Result for Dataset of Volcano Eruption dataset for 0.01 level of Significance both of location parameters of the dataset does not follow a von Mises distribution

Table 2. (a) Significant deviation for longitude (Von Mises), not for latitude in Earthquake dataset at 0.01 significance. (b) Significant deviation for longitude (Von Mises), not for latitude in Volcano dataset at 0.01 significance.

Test statistics	Critical Value	Output of null Hypothesis
0.318	0.268	Reject Null Hypothesis

(a) Homogeneity test to get if both the location parameters of the Earth quack dataset follow the same distribution or not

Test statistics	Critical Value	Output of null Hypothesis
0.6433	0.268	Reject Null Hypothesis

(b) Test of Homogeneity to get if both the location parameters of the Volcano Eruption dataset follow the same distribution or not

Table 3. (a) Homogeneity test for assessing whether the location parameters of two earthquake datasets follow the same distribution. (b) Homogeneity test for assessing whether the location parameters of two Volcano datasets follow the same distribution.

4.2 Results on Projected Circular Distributions of earthquake & Volcano on the Earth

Observe that the spherical plot of the raw data (figure1a and 1b) indicates clusters and the need for partition. Indeed, for unpartitioned whole data, the table 2asays the test rejects the null hypothesis for the Watson Test for Dataset of Earth Quake for 0.01 level of Significance both of location parameters, so overall unpartitioned locations of the earthquake as a whole does not follow the Von Mises distribution. Indeed, we show that both data indicate the same number of partitions under unsupervised learning. The results of Homogeneity for the whole dataset are mentioned in table 3a, which tells us that if not partitioned, the entire dataset of the two location parameters rejects the null hypothesis.

Remark 5 *The tables 2a and table 3a indicate that the location parameter as a whole does not follow the proposed circular distribution (Von Mises), and the spherical plot of the raw data plot of figure1a and 1b indicates clusters and need for partition. Hence, the immediate logical conclusion is partitioning the whole dataset and testing for spherical distributional similarity.*

4.3 A comparison between Kent distribution and Von Mises distribution

In this section, we used the limiting null distribution of Kent's (1982) statistic (see[32]) to test whether a sample comes from the Fisher distribution (Von Mises Fisher distribution) when K, the concentration parameter, goes to inf. A modification is suggested, the limiting null distribution of which is χ^2_2 when either κ or n, the sample size, goes to inf. Tests of Fisher based on the eigenvalue of the sample cross-product matrix are also considered. Numerical examples are presented.[33]

We tested the Hypothesis test for von Mises-Fisher distribution over Kent distribution for earthquake and volcano datasets. The null hypothesis is whether a von Mises-Fisher distribution fits the data well, whereas the alternative is that the Kent distribution is more suitable. The details of the hypothesis testing with the p-value are given in the table 4.

4.4 Results for Two-dimensional partition of the dataset

The table 5 contains the partition of the combined earthquake and volcano dataset, which gives us seven partitions which are following Vonmises distribution, considering we took the projection of the location parameter and Individually tested the distributions of that dataset with the help of Watson Test.

4.5 Results regarding the fit of Mixture of Fisher Von Mises Distribution Model

4.6 partitions for the volcano part dataset:

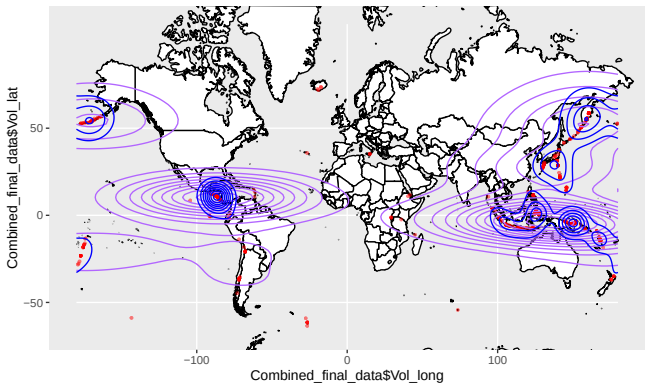
We started with 20 of the partitions for the volcano part dataset; the details are in table 6a; obtaining an optimal 14 partitions can provide a mixture of Fisher Von Mises Distribution for the location parameter(Longitude and Latitude together). We checked 20 values of k from 1 to 20, where the optimal number of partitions is 14. We used the Volcano Part dataset to get this table. We get that this will give us the weights(ϕ) or α values as 0.014515901, 0.078723287, 0.037609859, 0.004643843, 0.351158355, 0.01908429, 0.177467094, 0.062865949, 0.036460001, 0.041481650, 0.071710780, 0.070833349, 0.020899085, 0.012546552. This is done for the Volcano part of the dataset.

Data part	test	Bootstrap p-value
Earthquake	195.979	0.583
Volcano	194.7918	0.5270

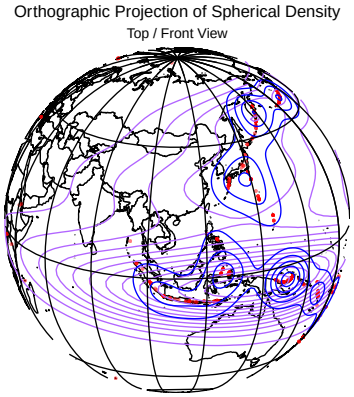
Table 4. Hypothesis test for von Mises-Fisher distribution over Kent distribution for earthquake dataset where the p-value is 0.583 and the null hypothesis is whether a von Mises-Fisher distribution fits the data well, where the alternative is that Kent distribution is more suitable and similar for volcano dataset where the p-value of this dataset is 0.5270 .

EQ Latitude	EQ Longitude	Vol Latitude	Vol Longitude	Distribution
-56.254 to 55.918	-177.64 to 177.26	-56.300 to 56.056	-177.19 to 177.18	Von Mises
-37.341 to 85.644	-177.16 to 177.33	-37.520 to 85.608	-177.19 to 177.18	Von Mises
-39.744 to 56.170	-177.98 to 177.52	-39.420 to 56.056	-177.19 to 177.18	Von Mises
-59.307 to 64.681	-179.95 to 176.17	-59.017 to 64.633	-179.03 to 179.58	Von Mises
-30.199 to 55.918	-178.09 to 168.46	-30.543 to 56.056	-178.56 to 168.37	Von Mises
-57.679 to 64.782	-178.36 to 177.33	-57.80 to 64.42	-177.17 to 177.18	Von Mises
-57.241 to 63.965	-90.91 to 158.93	-57.800 to 63.983	-90.60 to 158.20	Von Mises

Table 5. This table contains the partition of the combined earthquake and volcano dataset, which gives us seven partitions that follow Vonmises distribution considering we took the projection of the location parameter and Individually tested the distributions of that dataset with the help of Watson Test.

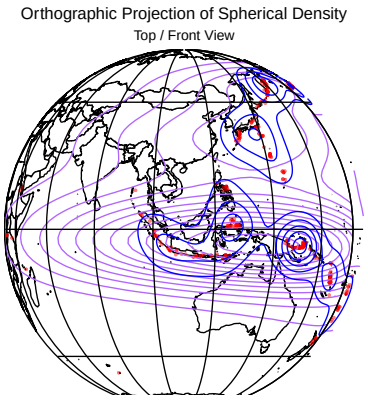


(a) Plot of the volcano part of the dataset all over the two-dimensional world map with the densities.

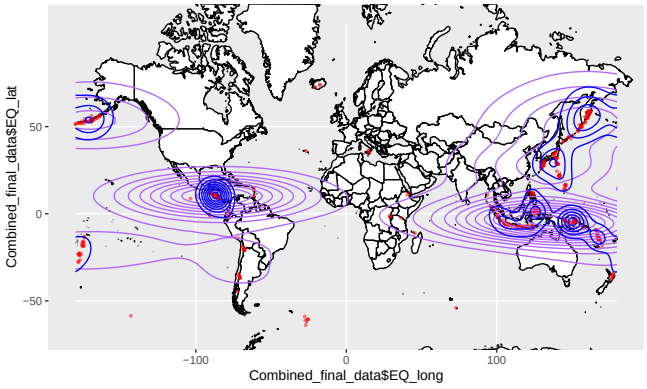


(b) Plot of the volcano part of the dataset all over the three-dimensional world map with the density.

Figure 4. Density Plot of Volcano in the 3D and 2D Plane



(a) Plot of the earthquake part of the dataset all over the spherical world map with the densities.



(b) Plot of the earthquake part of the dataset all over the two-dimensional world map with the densities.

Figure 5. Density Plot of Earthquake in the 3D and 2D Plane

K Values	BIC	K Values	BIC
1	-54.06857	11	-1804.32167
2	-1114.07092	12	-1894.01874
3	-1268.51948	13	-1828.33372
4	-1424.21191	14	-1903.50719
5	-1467.47741	15	-1844.48016
6	-1507.52284	16	-1787.84740
7	-1730.72499	17	-1719.25160
8	-1748.07882	18	-1755.34958
9	-1799.561584	19	-1740.73555
10	-1815.63021	20	-1690.66637

(a) Table containing the Bayesian Information Criteria Score for Each K value. From that, 14 partitions can provide a mixture of Fisher Von Mises Distribution for the location parameter(Longitude and Latitude together). We checked 20 values of k from 1 to 20, where the optimal number of partitions is 14. We used the Volcano Part dataset to get this table. We get that this will give us the weights(ϕ) or α values as 0.014515901, 0.078723287, 0.037609859, 0.004643843, 0.351158355, 0.01908429, 0.177467094, 0.062865949, 0.036460001, 0.041481650, 0.071710780, 0.070833349, 0.020899085, 0.012546552. This is done for the Volcano part of the dataset.

K Values	BIC	K Values	BIC
1	-49.87195	11	-1808.27896
2	-1112.30709	12	-1796.88277
3	-1261.87566	13	-1802.01574
4	-1417.94000	14	-1802.40150
5	-1505.86819	15	-1782.84385
6	-1501.74819	16	-1774.68752
7	-1731.04052	17	-1767.89547
8	-1758.54639	18	-1756.11820
9	-1778.91169	19	-1717.64136
10	-1789.60595	20	-1768.21031

(b) Table containing the Bayesian Information Criteria Score for Each K value. From that, 14 partitions can provide a mixture of Fisher Von Mises Distribution for the location parameter(Longitude and Latitude together). We checked 20 values of k from 1 to 20, and the optimal number of partitions is 14. We used the earthquake part of the combined dataset to get this table. We get that this will give us the weights(ϕ) or α values as 0.06643275, 0.02084306, 0.18108015, 0.06747089, 0.01665956, 0.00550244, 0.10448460, 0.01455958, 0.03741776, 0.01246680, 0.35071396, 0.01846428, 0.02948653, 0.07441763. This is done for the Volcano part of the dataset.

Table 6. BIC table for both Volcano and Earthquake after fitting vonMises Fishers distribution

4.7 Partitions for the earthquake part dataset:

We started with 20 of the partitions for the earthquake part of the combined dataset, and the details are given in the table 6b. We checked values of k from 1 to 20, and the optimal number of partitions is 14, providing a mixture of Fisher Von Mises Distribution for the location parameter(Longitude and Latitude together). We used the earthquake part of the combined dataset to get this table. We get that this will give us the weights(ϕ) or α values as 0.06643275, 0.02084306, 0.18108015, 0.06747089, 0.01665956, 0.00550244, 0.10448460, 0.01455958, 0.03741776, 0.01246680, 0.35071396, 0.01846428, 0.02948653, 0.07441763.

Remark 6 *We can observe that Earthquakes and volcanoes have the same 14 partitions as the optimal number of mixtures in the von Mises Fishers mixture model. In the table 5, we got seven partitions that follow Von Mises distribution where we took spherical to circular projection, and in the spherical paradigm, it is 14 (Multiple of 7)*

5 Discussions

The primary findings of this work can be summarized as follows:

1. This study identified distributional similarities between volcanic eruptions and earthquakes, supporting the geological explanations and relationships between these two natural phenomena.

2. The dataset we constructed by collecting earthquake data within a 100 km radius of volcano eruption locations within one year Count data for the number of earthquakes and volcano eruptions were used to assess the frequency of these events. We propose empirical models and statistically validate the models based on that dataset. Although our model is empirical (observation-based), it may provide a more subtle perspective on the potential interactions between volcanoes and earthquakes.

3. The study utilized 2D projection partitioning techniques, which resulted in seven partitions for the combined volcano and earthquake data. This partitioning suggests that distinct clusters or groups of events exhibit similar characteristics within the dataset. All partitions follow Vonmises distribution for both parts (volcano and earthquake), and the optimal number of partitions is 7.

4. The Von Mises Fisher distribution fit test was applied for the spherical analysis. This test indicated that the optimal number of partitions for volcanic eruptions was 14, while for earthquakes, it was 14. These results align with the 2D projection partitioning findings, further supporting the existence of distinct patterns within the data.

6 Conclusion & Future Work

This work may provide a stepping stone for a directional statistical model and insights into the geological relationship between volcanic eruptions and earthquakes. In the future, an advanced combination of directional statistical spatial and temporal analysis, count data examination, and Bayesian statistical modeling may reveal more intriguing patterns and similarities and warrant further investigation. Understanding the underlying dynamics of volcano-earthquake interactions using Bayesian directional statistical tools is another future work that will have important implications for hazard assessment and disaster preparedness in regions prone to these events, leading to a deeper understanding of the complex geophysical processes at play.

Author contributions statement

D.C. designed, conceptualized, and developed the research and synthesized interdisciplinary statistical methodologies and models. P.G. collected and prepared the datasets, wrote codes for various modified datasets, and performed code-based analysis (mainly using R and Python). A.B. and S.D. supervised the alignments of relevant Earth science-related theories and methodologies. P.G., D.C., A.B., and S.D. wrote and reviewed the manuscript.

acknowledgments

P.G. and D.C. are thankful to Visva Bharati University, Santiniketan; A.B. and S.D. are thankful to the Indian Statistical Institute, Kolkata, for the facilities provided to the researchers.

Data Availability

The secondary data we have created for the sake of the analysis in this paper, has been attached as a supplementary.
The primary data used for this article can be accessed through the links provided below. The primary data is from[34] https://www.usgs.gov/programs/earthquake-hazards/earthquakes[35] and volcano[36] https://volcano.si.edu/[37].
The first data set, which contains information about the Earthquake Dataset, can also be found at the following[38]. The second data set provides valuable Earth Quack Direction information and is available at Earth Quake Direction dataset https://doi.org/10.7910/DVN/SPLMMN. All the datasets used in this paper are combined, and we have uploaded them to the Harvard Dataverse (weblink: https://doi.org/10.7910/DVN/SPLMMN). We have also uploaded the same in GitHub, weblink https://github.com/Prithwish-ghosh/Earth-Quack.

Code Availability

We have uploaded the most recent version of all the codes written for the analysis of this paper at the following Github link: https://github.com/Prithwish-ghosh/Earth-Quack.

Additional Information

The authors declare that they have no competing interests. No funding was available for this research from our institutes or organizations.

Supplimentary

The supplementary materials are mentioned in the Journal and the Harvard Data Verse https://doi.org/10.7910/DVN/SPLMMN

References

1. Mendoza-Rosas, A. & De la Cruz-Reyna, S. Hazard estimates for el chichón volcano, chiapas, méxico: a statistical approach for complex eruptive histories. *Nat. Hazards Earth Syst. Sci.* **10**, 1159–1170 (2010).
2. Hill, D. P., Pollitz, F. & Newhall, C. Earthquake–volcano interactions. *Phys. Today* **55**, 41–47 (2002).
3. Soosalu, H. *et al.* Lower-crustal earthquakes caused by magma movement beneath askja volcano on the north iceland rift. *Bull. Volcanol.* **72**, 55–62 (2010).
4. Aki, K. & Koyanagi, R. Deep volcanic tremor and magma ascent mechanism under kilauea, hawaii. *J. Geophys. Res. Solid Earth* **86**, 7095–7109 (1981).

5. Aki, K. Earthquake mechanism. *Tectonophysics* **13**, 423–446 (1972).

6. Anderson, O. L. The role of magma vapors in volcanic tremors and rapid eruptions. *Bull. Volcanol.* **41**, 341–353 (1978).

7. Seropian, G., Kennedy, B. M., Walter, T. R., Ichihara, M. & Jolly, A. D. A review framework of how earthquakes trigger volcanic eruptions. *Nat. Commun.* **12**, 1004 (2021).

8. Manga, M., Brumm, M. & Rudolph, M. L. Earthquake triggering of mud volcanoes. *Mar. Petroleum Geol.* **26**, 1785–1798 (2009).

9. Lemarchand, N. & Grasso, J.-R. Interactions between earthquakes and volcano activity. *Geophys. Res. Lett.* **34** (2007).

10. Triastuty, H., Iguchi, M. & Tameguri, T. Temporal change of characteristics of shallow volcano-tectonic earthquakes associated with increase in volcanic activity at kuchinoerabujima volcano, japan. *J. volcanology geothermal research* **187**, 1–12 (2009).

11. Ogata, Y. Statistics of earthquake activity: Models and methods for earthquake predictability studies. *Annu. Rev. Earth Planet. Sci.* **45**, 497–527 (2017).

12. Cosentino, P., Ficarra, V. & Luzio, D. Truncated exponential frequency-magnitude relationship in earthquake statistics. *Bull. Seismol. Soc. Am.* **67**, 1615–1623 (1977).

13. Michael, A. J. Fundamental questions of earthquake statistics, source behavior, and the estimation of earthquake probabilities from possible foreshocks. *Bull. Seismol. Soc. Am.* **102**, 2547–2562 (2012).

14. Ou, G.-B. & Herrmann, R. B. A statistical model for ground motion produced by earthquakes at local and regional distances. *Bull. Seismol. Soc. Am.* **80**, 1397–1417 (1990).

15. Scealy, J. & Wood, A. T. Scaled von mises–fisher distributions and regression models for paleomagnetic directional data. *J. Am. Stat. Assoc.* (2019).

16. Connor, C. B., Hill, B. E., Winfrey, B., Franklin, N. M. & Femina, P. C. L. Estimation of volcanic hazards from tephra fallout. *Nat. Hazards Rev.* **2**, 33–42 (2001).

17. Walsh, D., Arnold, R. & Townend, J. A bayesian approach to determining and parametrizing earthquake focal mechanisms. *Geophys. J. Int.* **176**, 235–255 (2009).

18. Lazizi, A., Trouzine, H., Asroun, A. & Belabdelouhab, F. Numerical simulation of tire reinforced sand behind retaining wall under earthquake excitation. *Eng. Technol. & Appl. Sci. Res.* **4**, 605–611 (2014).

19. Currenti, G., Bonaccorso, A., Del Negro, C., Scandura, D. & Boschi, E. Elasto-plastic modeling of volcano ground deformation. *Earth Planet. Sci. Lett.* **296**, 311–318 (2010).

20. Nishimura, T. Interaction between moderate earthquakes and volcanic eruptions: Analyses of global data catalog. *Geophys. Res. Lett.* **45**, 8199–8204 (2018).

21. Mellors, R., Kilb, D., Aliyev, A., Gasanov, A. & Yetirmishli, G. Correlations between earthquakes and large mud volcano eruptions. *J. Geophys. Res. Solid Earth* **112** (2007).

22. Gopal, S. & Yang, Y. Von mises-fisher clustering models. In *International Conference on Machine Learning*, 154–162 (PMLR, 2014).

23. Korkmaz, S., Göksülük, D. & Zararsiz, G. Mvn: An r package for assessing multivariate normality. *R JOURNAL* **6** (2014).

24. Koziol, J. A. A class of invariant procedures for assessing multivariate normality. *Biometrika* **69**, 423–427 (1982).

25. Henze, N. & Zirkler, B. A class of invariant consistent tests for multivariate normality. *Commun. statistics-Theory Methods* **19**, 3595–3617 (1990).

26. Berger, V. W. & Zhou, Y. Kolmogorov–smirnov test: Overview. *Wiley statsref: Stat. reference online* (2014).

27. Wood, A. T. Simulation of the von mises fisher distribution. *Commun. statistics-simulation computation* **23**, 157–164 (1994).

28. Mammasis, K., Stewart, R. W. & Thompson, J. S. Spatial fading correlation model using mixtures of von mises fisher distributions. *IEEE Transactions on Wirel. Commun.* **8**, 2046–2055 (2009).

29. Hornik, K. & Grün, B. movmf: An r package for fitting mixtures of von mises-fisher distributions. *J. Stat. Softw.* **58**, 1–31 (2014).

30. Jammalamadaka, S. R. & SenGupta, A. *Topics in circular statistics*, vol. 5 (world scientific, 2001).

31. A SenGupta, S. R. J. *Topics in Circular Statistics*. Volume 5 (World Scientific Publishing Co. Pte. Ltd, 2001).

32. Kent, J. T. The fisher-bingham distribution on the sphere. *J. Royal Stat. Soc. Ser. B (Methodological)* **44**, 71–80 (1982).

33. Rivest, L.-P. Modified kent's statistics for testing goodness of fit for the fisher distribution in small concentrated samples. *Stat. & probability letters* **4**, 1–4 (1986).

34. Primary dataset usgs. https://www.usgs.gov/programs/earthquake-hazards/earthquakes. Accessed: 2023-10-30.

35. usgs. Usgs (2023). https://www.usgs.gov/programs/earthquake-hazards/earthquakes [Accessed: (August 2023)].

36. Volcano dataset. https://volcano.si.edu/. Accessed: 2023-10-30.

37. of Natural History Global Volcanism Program, S. I. N. M. Smithsonian institution national museum of natural history global volcanism program (2023). https://volcano.si.edu/ [Accessed: (August 2023)].

38. Primary dataset usgs kaggle version. https://www.kaggle.com/datasets/usgs/earthquake-database. Accessed: 2023-10-30.