Communications in Applied Analysis **xx** (200x), no. N, xxx–xxx

MATHEMATICAL AND STATISTICAL ANALYSIS OF TEPHRA FALLOUT

R. D. WOOTEN[1] AND C. P. TSOKOS[2]

[1]Department of Mathematics & Statistics
University of South Florida
Tampa, FL 33620 USA
E-mail: rwooten@cas.usf.edu

[2]Department of Mathematics & Statistics
University of South Florida
Tampa, FL 33620 USA
E-mail: profcpt@cas.usf.edu

ABSTRACT. In this study, the bivariate probability distributions of volcanic explosivity index as well as the tephra fallout as measured at Cerro Negro are considered and the skewness of the distribution is considered empirically and the non-skewed bivariate Gaussian probability distribution is compared to the skewed Gaussian distribution.

Keywords. Bivariate Probability Distribution

1. INTRODUCTION

In the present study the dispersion of ash fall from a volcanic event in two parts. First, we consider the empirical probability of a given Volcanic Explosivity Index (VEI); that is, the associated proportion of the volcanic eruptions at Cerro Negro which are a given VEI: 1, 2 or 3; see [2].

Second, we will perform parametric inferential analysis on the mass of tephra measured at 80 sites around where ash fall and was presented in [1] and [4]. If there were no external forces other than gravity and all particles were perfect in shape (round), we would expect the dispersion to be bivariate Gaussian (normal) probability distribution to characterize the key variable, but with the rotation of the earth and the resulting wind shear, the distribution is skewed; see [3]. Therefore, four variations of the standard bivariate normal are being considered in the present study. The fit of these probability distributions were compared using χ^2 and R^2 to determine the best-fit probability distribution and percent of empirical distribution explained by the statistical model when best characterizes the behavior of the subject phenomenon.

Received September 10, 2007

Establishing the probability distribution of the subject variable (mass in cubic meters) enables us to estimate the amount of mass that is likely to land in a given location. This is extremely important in urban development as well as for strategic planning and risk analysis.

In our present study we will address the following questions:

1. *Identify the volcano addressed and why.*
2. *What is the probable VEI in a volcanic event?*
3. *What is the probability distribution of tephra, combined and by grain size?*
4. *What is the best-fit bivariate probability distribution?*

2. CERRO NEGRO, NICARAGUA

The volcano of interest in this study is Cerro Negro, Nicaragua. Located at 12.5°N and 86.7°W, this volcano has an elevation of 2214 feet (675 meters) and a summit of 2388 feet (728 meters). Since its birth in 1850, there have been approximately 24 eruptions; the last eruption was in 1999. At 155 years, this is the youngest of Central America's volcanoes in the Maribios volcanic chain.

There are many uncertain data from the dates of eruptions to the magnitude of the eruptions. Searching Cerro Negro, Nicaragua, there are many sites, which offer information on volcanoes. The Global Volcanism Program, maintained by the Smithsonian Institution, has posted information on the duration of eruptions, the volcano explosivity index (VEI), column height, the tephra fallout, the lava volume and the source area, see Table 1. Additional information was gleaned from other sources, see [2].

Table 1 gives the eruptions data for Cerro Negro including the year the volcanic event occurred, the approximate duration of the event, the cumulative volume (cubic meters), the approximate fall volume for the given event (cubic meters), the tephra (cubic meters), lava volume (cubic meters), column height (meters) and area effect.

3. ANALYSIS OF VOLCANIC EXPLOSIVITY INDEX OF CERRO NEGRO, NIGERIA

Consider the VEI for the volcanic eruptions at Cerro Negro, Nigeria. There are several capricious data sources, see Figures 1 and 2. Figure 1 shows VEI of 0, 1 or 2 (Smithsonian), whereas Figure 2 shows VEI of 0, 1, 2, 2.5 and 3; see [1]. One source has a VEI of zero when there is little appreciable tephra fallout, no column height and little or no lava flow, while other sources have VEI of 1 for the same eruption. Such as the eruption in 1995; one source states there was an eruption which lasted 79 days but shows a VEI of 0, and for that same year an eruption which lasted only 13 days and expelled a significant amount of volume with a VEI of 1. Whereas, according to

the Smithsonian and historical data, the eruption that occurred in 1995 shows a VEI of 2.

While the two sources, (Smithsonian) and [2], are different, they are highly correlated with an estimate of the correlation coefficient, $R^2 = 71.8\%$. Compare the VEI from the two sources, see Table 2. The second source has several zeros whereas the first source only has record of one or higher. That is, the first source states all eruptions have an index of at least one. If we consider the proportions associated with the various levels of VEI, we see that the second source indicates that 35% of all eruptions are insignificant with a VEI of 0, and a VEI of 1 or 2 is equally likely at 26%, but that a VEI of 3 is likely to occur 13% of the time. More realistically, the first source indicates that a VEI of two is most likely at 61% of the eruptions, whereas a VEI of three is the second most likely magnitude of eruption at 30% and a VEI of one occurs the remaining 9% of the time. Note, both sources indicate that there has never been an eruption with a VEI of four or five; this powerful eruption has not occurred at Cerro Negro, yet.

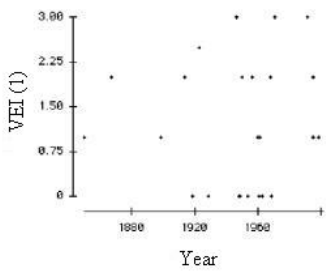


FIGURE 1. Line graph of VEI over the years (first source).

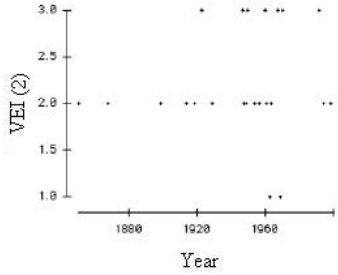


FIGURE 2. Line graph of VEI over the years (second source).

Under the assumption, the first source is more realistic and we will proceed with the analysis of the remaining variables; namely, the direction of the deposit, and the thickness of the deposit.

4. TREND ANALYSIS

Consider the cumulative volumes in cubic meters of volcanic fallout. Throughout time, volcanic eruptions of magnitude 3 are commonly followed by eruptions of magnitude two or one. Let $x_i(t) = n\left(VEI_i(t)\right)$ be the cumulative frequencies for each of the three main magnitudes, $i = 1, 2, 3$, shown in Figure 3, are the overall cumulative frequencies defined by $n(t) = \sum_{i=1}^{3} x_i(t)$. Then the probability of an eruption of a given magnitude is

$$p_i(t) = \frac{x_i(t)}{n(t)},$$

for $i = 1, 2, 3$ as illustrated in Figure 3.

Figure 4 shows the convergence of the percentage of given VEI over time; that is, approximately 9% of volcanic eruptions at Cerro Negro have VEI of 1, 61% of volcanic eruptions at Cerro Negro have VEI of 2, and approximately 30% of volcanic eruptions at Cerro Negro have VEI of 3.

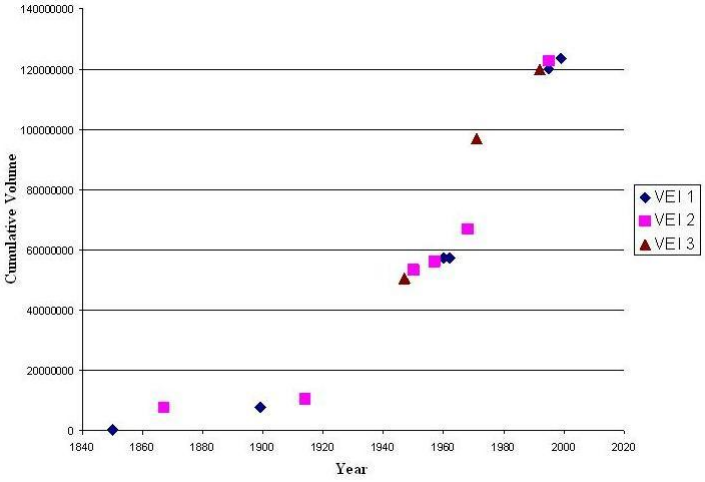


FIGURE 3. Line graph of cumulative volume (m^3) since the birth of Cerro Negro.

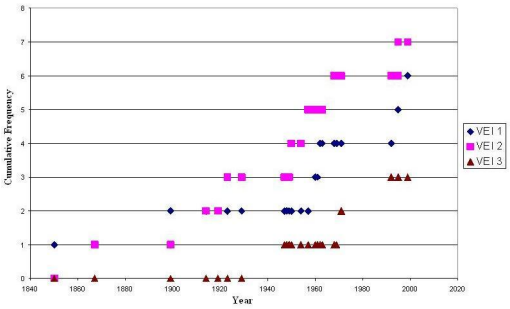


FIGURE 4. Line graph of cumulative frequencies.

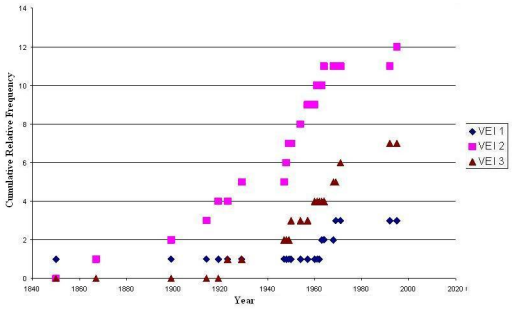


FIGURE 5. Line graph of percentages over time (probabilities).

We see that there are two large gaps in the line graphs given by Figures 4 and 5. The first gap appears between 1867 and 1899 (32 years), and the second gap between 1971 and 1992 (21 years); but on the average, there is an eruption every 6.2 years. In addition, the first few eruptions after this twenty-plus year lull were one of magnitude 3 and then two eruptions of magnitude 2 over a seven year period. In fact, over half of the differences indicate an eruption every 3 years. However, there was a time that this volcano lay dormant for three decades. This study will not address the probability that there is an event in a given year, but the conditional hazard that relates to the probabilities that a given event has a specified magnitude and where, relative to the main vent of the volcano, the probability that the tephra will fall in a specified direction.

5. DIRECTION OF DEPOSITION: CONIC SECTIONS

Data collected at Cerro Negro by the University of South Florida's Geology department, the mass of tephra by grain size that can be used to analyze the probability distribution of tephra fallout. With the main vent of the volcano set as the origin, consider the eight sections (45° each) enumerated counter-clockwise off true east. These eight conical regions defined finding the angle off due east and then divided into equal sectors

$$s = \text{int}\left(\frac{\theta}{45^\circ} + 1\right).$$

The majority of the tephra fallout is in the sixth sector as shown in Figure 6. As given in Table ??, 63% of the mass falls to the south-southwest of the volcano. Also, all mass falls south of the main vent and very little (11%) falls southeast. The differences between the locations where the data is collected and where tephra falls should be minimal; that is, the data is assumed to be gathered at random, (systematically selected) to represent all locations where tephra falls. Figure 6 illustrates that the probability distribution of the given data is not best characterized by the symmetric bivariate normal probability distribution.

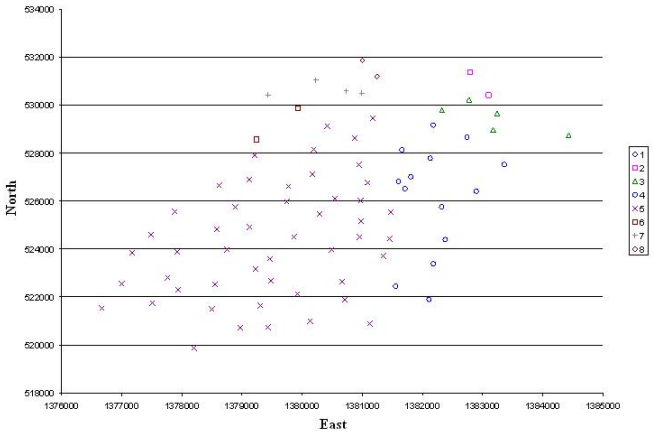


FIGURE 6. Scatter plot of Tephra Fallout by conical section.

Further refine this directional partition into twenty-four conical sectors shown by Figure 7 where the cones defined finding the angle off due east and then divided into equal sectors, $s = \text{int}\left(\frac{\theta}{15^\circ} + 1\right)$. This refined partition shows that while 3.8% of the samples are taken near the main vent of the volcano, 20.7% of the mass falls in this direction. The more refined the sectors, the more normal the distribution appears as shown in Table 4.

This analysis shows that the data is not Gaussian; the dispersion of the tephra is not symmetrical with respect to the center of the main vent. This analysis also shows that even if the standard (non-correlated) bivariate normal is assumed, the data

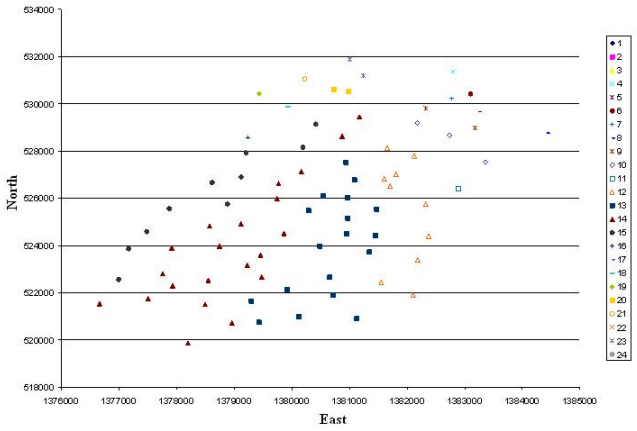


FIGURE 7. Scatter plot of Tephra Fallout by conical section.

must either be rotated to a primary and secondary axis or use the general (correlated) bivariate normal.

6. RADIAL ANALYSIS

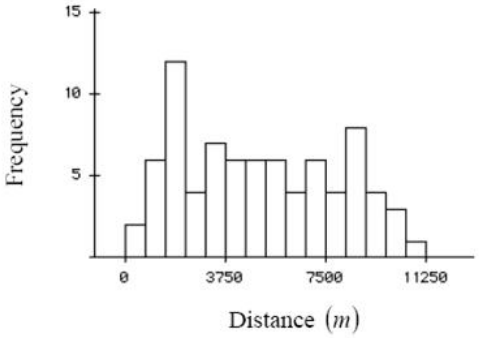
Let the location of the main vent be the center of our volcanic eruption. Assuming the converted latitude and longitude denoted in meters (northing and easting, Universal in a Transverse Mercator coordinate), and then we can compute the distance from this center marker as well as the angle. That is, the distance

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$$

between (x_i, y_i) is the location of the i^{th} sample and (x_c, y_c) is the location of the main vent or center and the angle $\theta_i = \tan^{-1}\left(\frac{y_i - y_c}{x_i - x_c}\right)$ is off due east. Then we can best analyze the distance and angle independently.

Consider the histogram of the distances sampled as shown by Figure 8 along with the basic statistics that describe the data. More samples were taken closer to the main vent, and fewer were taken more than $10,000$ meters from the main vent, but all in all the number of samples are uniform. Percentages for distance (with thirty contours) are given in Table 5, but more interesting, the mass measured at these various distances illustrated by Figure 9. Also in the table that follows, in addition to the distance contour, it shows the count, present by distance, sample mean of the mass present and the percent of mass at each distance. The basic descriptive statistics are also given by the accompanying table.

Furthermore, consider the distribution of the angle θ_i, samples appear to be normally distributed as illustrated by Figure 10 with a mean of $194.5°$ off due east with a standard deviation of $47.1°$; however, this is simply the sampling distribution. However, as Figure 9 illustrates the distribution of mass is also centered about this



Statistic	Estimation
Count	79
Mean	5025.153
Median	4666.93
Std Dev	2916.963
Variance	8508671
Range	10936.3
Min	0
Max	10936.3
IQR	5397.227
25th%	2246.24
75th%	7643.467

FIGURE 8. Histogram of the estimated distance from the center (main vent) including descriptive statistics.

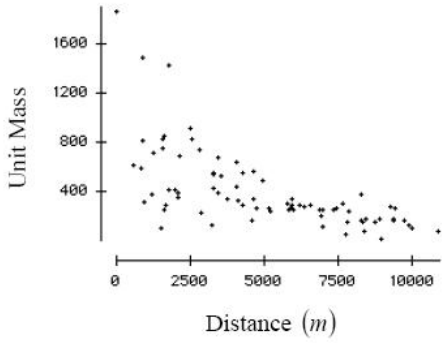
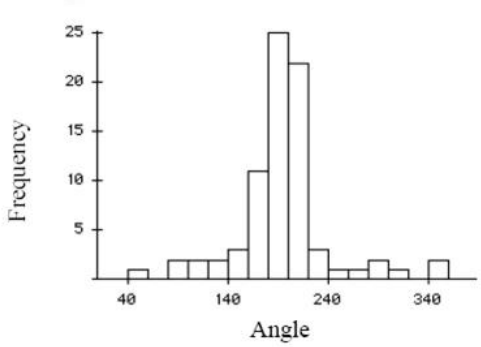


FIGURE 9. Scatter plot of mass kg/m^2 and distance from the center (main vent).

angle as well. The basic descriptive statistics are also given by the accompanying table.



Statistic	Estimation
Count	78
Mean	194.533
Median	194.744
Std Dev	47.117
Variance	2220.012
Range	292.932
Min	49.472
Max	342.404
IQR	30.18
25th%	178.896
75th%	209.076

FIGURE 10. Histogram of the estimated angles off the horizon including descriptive statistics.

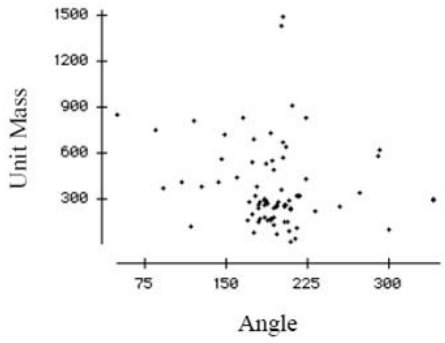


FIGURE 11. Scatter plot of mass kg/m^2 and angle off due east.

Further analysis of the mass by angle indicates that the distribution of the angle of fallout is not normally distributed. The normal plot and box plot given by Figures 12 and 13, respectively, indicate that the data is more uniformly distributed near the central angle determined by the rotation of the earth and the direction of the wind near the main vent. All other directions are outliers as illustrated in the box plot given by Figure 13, where an outlier is any point which falls further than three sample standard deviations from the mean. These angles of trajectory would be uniformly distributed over all 360°. This is due to the fact that without the external forces (and assuming perfectly spherical and uniform particle size) the dispersion of the ash fall would be bivariate normal (Gaussian).

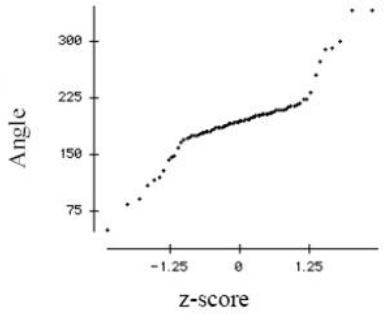


FIGURE 12. Normal probability plot for direction of fallout.

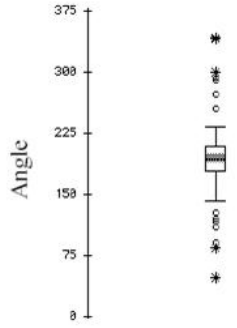


FIGURE 13. Box plot for direction of tephra fallout.

7. PARTICLE SIZE PROBABILITY DISTRIBUTION

These eruptions produced an ash-rich column extending 2 kilometers. Here we see the smallest particles of ash falling away from the mushrooming column as well as vibrator dust whereas the majority of the particles form a more liquidous buoyant state. In general, Strombolian eruptions are characterized by the sporadic explosion or spewing forth basaltic lava from a single vent or crater. Each event is caused by the release of volcanic gases, and they typically occur periodically — sometimes with an

appearance pattern and others more randomly. The lava fragments generally consist of partially molten volcanic bombs that become rounded as they fly through the air.

These particles were gathered and sifted into sixteen different particle sizes, ϕ, $\phi = -\log_2 d$ where d is the particle diameter measured in millimeters. The ranges in diameters are as listed in Table 6.

Consider when the mass is plotted first versus diameter as shown by Figure 14 and then versus ϕ shown by Figure 15. Furthermore, consider the probability distribution of the diameter and ϕ using mass as the frequency.

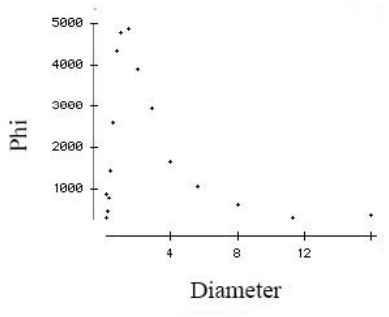


FIGURE 14. Scatter plot of mass (kg/m^2) by diameter d.

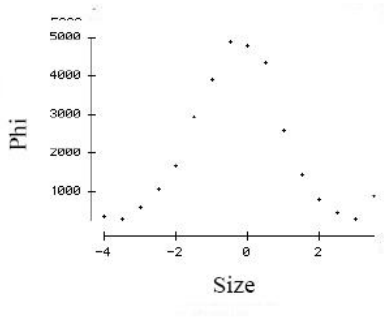


FIGURE 15. Scatter plot of mass (kg/m^2) by size $\phi = -\log_2 d$.

Phi does demonstrate a more normal probability distribution, but does not compensate for the distributions in the tail. The best-fit probability distribution is the Log-Normal probability distribution, see Table 7 below; however, many volcanologists use the normal probability distribution which will give misleading results.

Here the empirical distribution is computed by the mass of the various particle sizes;

$$P(j = \phi) = \frac{M(j)}{\sum_{i \in \Phi} M(i)},$$

where Φ is the set of particle sizes defined by $\phi = -\log_2 d$ with d being the diameter size of the particle in millimeters, $d > 0$. Note: there is loss of mass when converting the percentage particle size at each given location back to comparable mass unit. The manipulated data is accurate up to $\pm 2\%$ of the actual recorded percent mass.

The normal (Gaussian) probability distribution of the size phi is given by

$$f_\Phi(\phi) = \frac{1}{\sigma_\phi \sqrt{2\pi}} \exp\left(-\frac{(\phi - \mu_\phi)^2}{2\sigma_\phi^2}\right), \tag{7.1}$$

where μ_ϕ is the expected value (true mean) of the size ϕ and σ_ϕ is the associated standard deviation, where the recorded mass is the frequency. It may be necessary to include a separate weighing system to break the mass into frequency or count of number of particles of a given size in a given mass. To consider this interpretation — given the number of particles, the probability that a given particle is of a given size

— would require estimations on the mass of particles of a given diameter size. Here we have the cumulative probability distribution of ϕ given by

$$F_\Phi(\phi) = 1 - \exp\left\{-\frac{(\phi - \mu_\phi)^2}{2\sigma_\phi^2}\right\},$$

and that of d given by

$$\begin{aligned} F_D(d) &= P\{D \leq d\} = P\{\ln D < \ln d\} \\ &= P\{\Phi < \ln d\} = F_\Phi(\ln d) \\ &= 1 - \exp\left\{-\frac{(\ln d - \mu_\phi)^2}{2\sigma_\phi^2}\right\}. \end{aligned}$$

We can simplify this cumulative probability distribution if d is given by

$$F_D(d) = \frac{1}{d\sigma_\phi^2} \exp\left\{-\frac{(\ln d - \mu_\phi)^2}{2\sigma_\phi^2}\right\}. \tag{7.2}$$

Furthermore, note that we can write

$$f_D(d) = \frac{f_\Phi(\ln d)}{d},$$

which is the probability distribution function of the mass by particle size as measured by the diameter of tephra.

8. STATISTICAL MODELING OF TEPHRA FALLOUT

Consider the three variables — mass, distance and angle — associated with tephra fallout. Let the mass of the tephra at a given location be denoted by m, then we can consider the linear statistical model given by

$$m = \beta_0 + \beta_1 d + \beta_2 \theta + \varepsilon, \tag{8.1}$$

where the β_i's are the weights that drive the estimate of the subject response and ε is the random error.

Statistically, we find that only the distance away from the main vent is significant as contributing variables with p-value < 0.0001 explaining 36.5% of the variation in the amount of mass recorded at a given location. The direction in which the mass of tephra found is dependent on the wind, but does not significantly contribute to the dispersion of the mass. Thus, an acceptable estimate of the statistical model given by equation (7.2) is,

$$\hat{m} = 810.368 - 0.0613542d. \tag{8.2}$$

That is, the response variable only depends on distance from main vent. Also, turning these roles around, consider the distances of the mass by particle size m_{ϕ_i} is given by

$$d = \beta_0 + \sum_{i=1}^{16} \beta_i m_{\phi_i} + \varepsilon, \tag{8.3}$$

where the β_i's are the weights that drive the estimate of the subject response and ε is the random error.

The developed statistical model explains 52.7% of the variation in the distance, but no one particle size was found to be significant. This model is important when considering that most advection equations assume that the location where tephra is expected to fall depends on particle size. However, the present study shows that the location, at least in terms of distance, does not depend on particle size.

Bivariate Distribution. Consider the mass m over the northing distance y and easting distance x shown by Figure 16. We see that the majority of the mass falls near the main vent and depends on the direction the wind is blown from the volcano. This wind direction can be measured. Then the major and minor axis can be rotated from the north and east points at the required angle, after which a simple non-correlated bivariate Normal distribution can be used; however, there are many contributing factors and the wind is not the only determining factor.

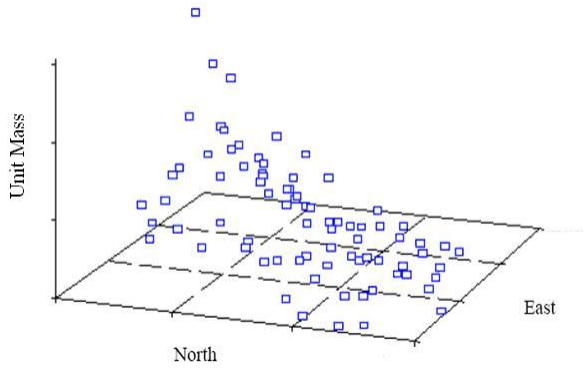


FIGURE 16. Scatter plot of mass (kg/m^2).

Consider the standard correlated bivariate normal probability distribution given by

$$f(x, y|\mu_x, \sigma_x, \mu_y, \sigma_y, \rho_{xy}) = K \exp\left\{ \frac{1}{-2(1+\rho_{xy}^2)} \left[z_x^2 - 2\rho_{xy} z_x z_y + z_y^2 \right] \right\},$$

where

$$z_x = \frac{x - \mu_x}{\sigma_x}, \quad z_y = \frac{x - \mu_y}{\sigma_y}, \quad \rho_{xy} \in (-1, 1)$$

and

$$K = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}, \text{ for } -\infty < x < \infty, \ -\infty < y < \infty.$$

Notice the function estimates given by the above probability distribution are on a different scale. This is simply due to the fact that the true empirical probability is

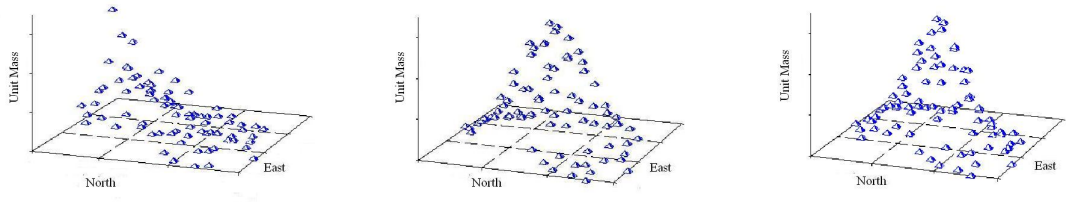
$$P(x,y) = \frac{\gamma m(x,y)}{\sum_{\Omega} m(x,y)},$$

where γ is the percent of the mass collected; that is, if the total fallout mass expelled by the volcano is m and

$$\gamma = \frac{\sum_{\Omega} m(x,y)}{m}$$

is the percent of the total mass measured in the sample.

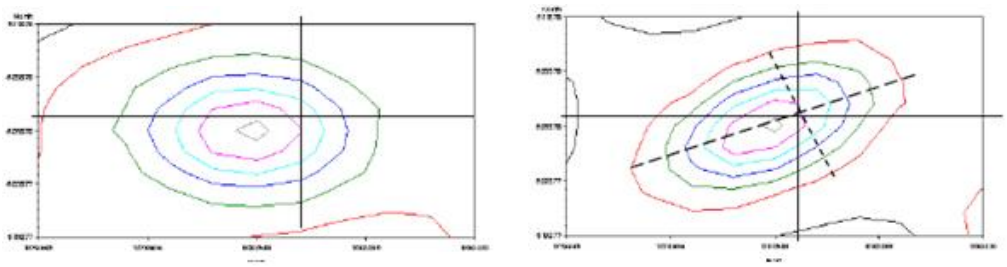
However, even with different scales, we see as shown by Figure 17(A), the distribution of the data collected is skewed toward the volcano's main vent; whereas, as shown by Figures 17(B) and 17(C), the non-correlated and the correlated bivariate Gaussian are symmetrical.



(A) Empirical (B) $\rho_{xy} = 0.0$ (C) $\rho_{xy} = 0.504$ (dependent)

FIGURE 17. Empirical probability distribution for the northern and eastern coordinates.

When comparing the empirical probability distribution with that of the general non-correlated bivariate Gaussian probability distribution and the correlated bivariate Gaussian probability distributions, neither account for the skewness of the data's distribution toward the volcano's main vent. Compare the contour plots for the non-correlated bivariate Gaussian probability distribution and the correlated bivariate Gaussian probability distribution as shown by Figures 18(A) and 18(B), respectively.



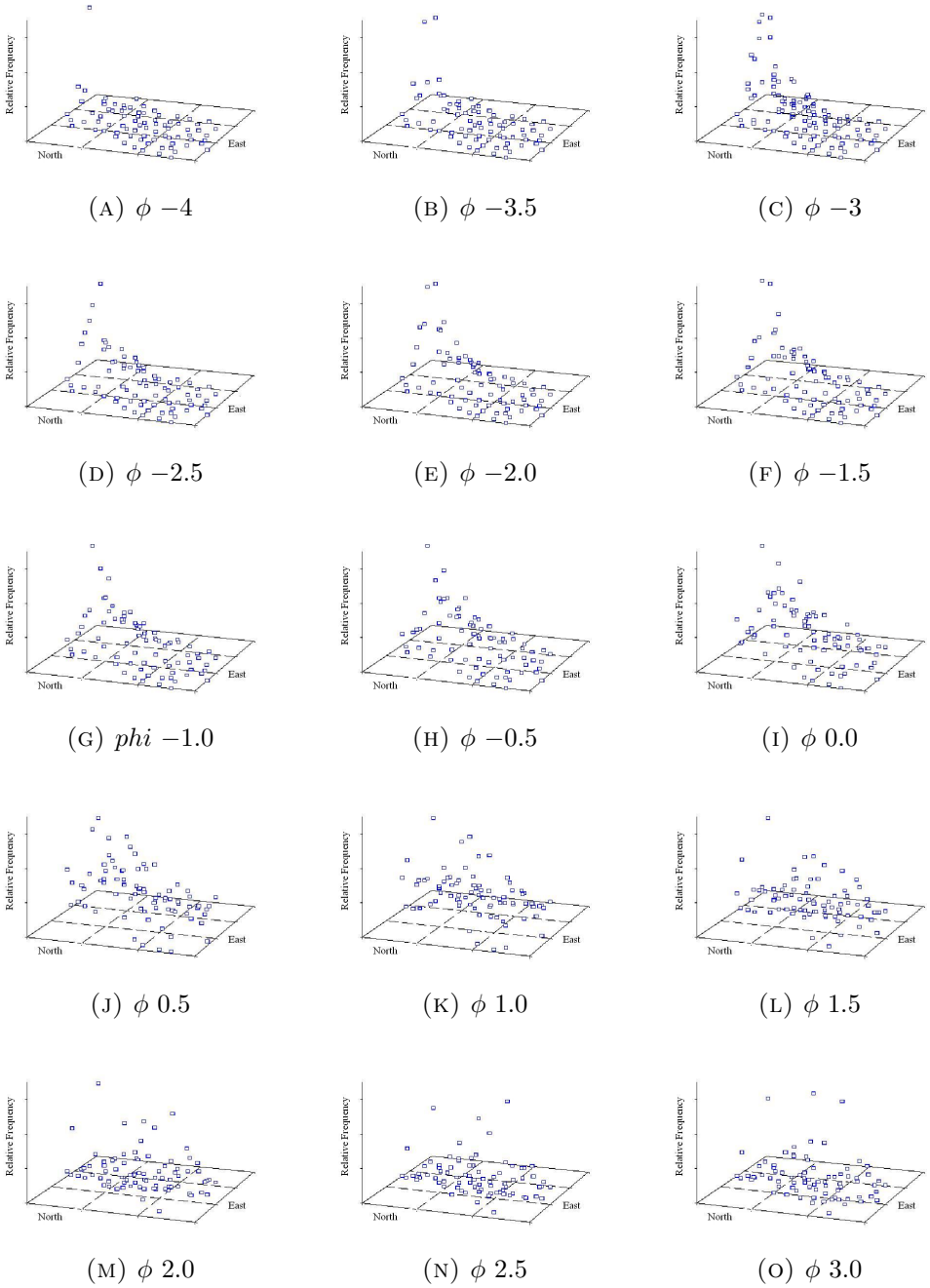
(A) $\rho_{xy} = 0.0$ (B) $\rho_{xy} = 0.53$

FIGURE 18. Non-correlated and correlated bivariate Gaussian distribution.

In the non-correlated bivariate, the wind shear effect is not present; assuming that the directions are independent. A common practice to compensate for this is to use the initially defined advection diffusion equations under the assumption that each layer in the atmosphere moves collectively and falls in normally distributed piles when considered by grain size. That is, the assumption is that the fallout is symmetrical to a center point and not skewed toward the main vent. This "kernel"-like approach need not directly account for the wind shear, but moreover does not account for the skewness of the fallout toward the main vent of the volcano.

Furthermore, while the correlated bivariate Gaussian does address the issue of orientation without any intermediate transformations, it does not address the skewness of the data toward the main vent as illustrated by Figure 19(A). Therefore, either we need to sift (profile) the data according to particle size in an attempt to justify the general bivariate Gaussian probability distribution function or test the goodness-of-fit of a skewed probability distribution such as some form continued within the generalized extreme value distribution (GEVD); (the Weibull, the Gumbel, the Frechet or the Pareto) or the skewed Normal distribution.

Furthermore, the categorization of the tephra fallout by particle size defined by $\phi = -\ln(d)$, where d is the diameter of the tephra, is normalized by the scale and homogenized by the variance. For the majority of the particle sizes, the distribution is skewed toward the main vent of the volcano, but as the particle's size becomes smaller, indicated by the large values of ϕ, the distribution loses its center of concentration and becomes more variegated. Therefore, it is possible for the larger particle sizes to be combined and characterized by the same parametric distribution, whereas non-parametric techniques may need to be used for the smaller particle size.



(A) ϕ −4 (B) ϕ −3.5 (C) ϕ −3

(D) ϕ −2.5 (E) ϕ −2.0 (F) ϕ −1.5

(G) *phi* −1.0 (H) ϕ −0.5 (I) ϕ 0.0

(J) ϕ 0.5 (K) ϕ 1.0 (L) ϕ 1.5

(M) ϕ 2.0 (N) ϕ 2.5 (O) ϕ 3.0

FIGURE 19. Tephra fallout by particle size.

The general empirical distributions for the majority of the larger particle sizes are very similar in skewness and dispersion to the distribution of the combined information. The smaller particle size, the more uniform the distribution, yet the dispersed is in the same general region as the large particle sizes. Hence, the overall correlated bivariate probability distribution can be applied to the total mass and not by particle size to approximate the best probabilistic behavior of the subject phenomenon (location of fallout) using one of the three forms of the bivariate normal distribution:

the rotated non-correlated normal probability distribution, the non-rotated correlated normal probability distribution and the rotated (independent) skew normal probability distribution.

For the data gathered at Cerro Negro, the correlation coefficient between the northing and easting distances is $\hat{\rho}_{xy} = 0.508$ using data location in listed form and the estimate of the correlation coefficient $\hat{\rho}_{xy} = 0.5378$, using the associated mass as weights.

The sample mean north coordinate is $\hat{\mu}_x = 525944$ ($\hat{\mu}_x = 527329$) and sample mean east coordinate is $\hat{\mu}_y = 1380440$ ($\hat{\mu}_y = 1380896$), with sample standard deviations of $\hat{\sigma}_x = 3222$ ($\hat{\sigma}_x = 2751.3$) and $\hat{\sigma}_y = 1720$ ($\hat{\sigma}_y = 1416.5$). Compare the north-south and the east-west cross-section of the empirical data versus the correlated bivariate normal distribution as shown by Figures 20–23; these figures illustrate that there the standard bivariate normal is not the best-fit distribution. There is skewness in the data which is not simulated by a symmetric normal bivariate distribution. This appears as a lean in the data as illustrated by Figure 20. However, the empirical probability distribution shown in Figure 22 is better characterized by the normal probability distribution shown in Figure 23.

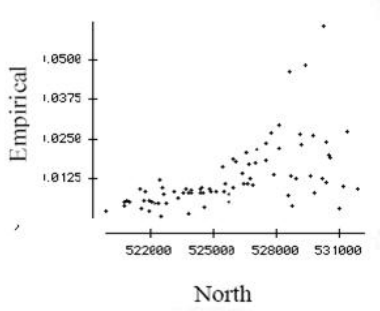


FIGURE 20. Scatter plot for the empirical probability distribution and distance north.

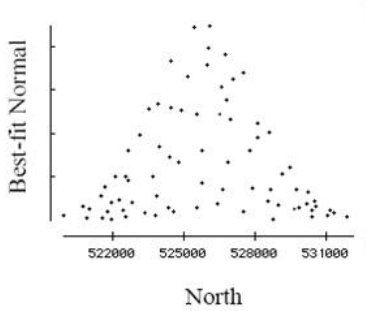


FIGURE 21. Scatter plot for the correlated bivariate normal distribution and distance north.

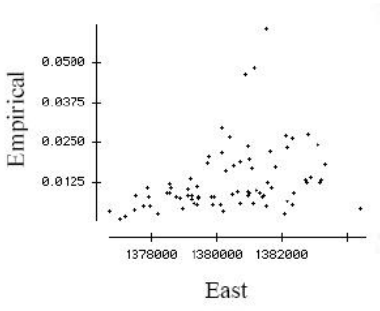


FIGURE 22. Scatter plot for the empirical probability distribution and distance east.

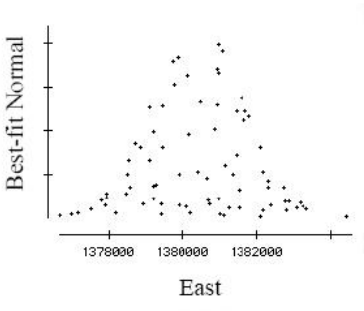


FIGURE 23. Scatter plot for the correlated bivariate normal distribution and distance east.

Furthermore, consider the three-dimensional plot of the empirical probability distribution and the correlated bivariate normal distribution with respect to the map of distance north by distance east shown by Figures 20 and 21, respectively. This further illustrates the contour plots shown by Figures 22 and 23, respectively. In these contour plots, the lines represent μ_x (true horizontal mean) and μ_y (true vertical mean) and the main vent is indicated by the dashed line.

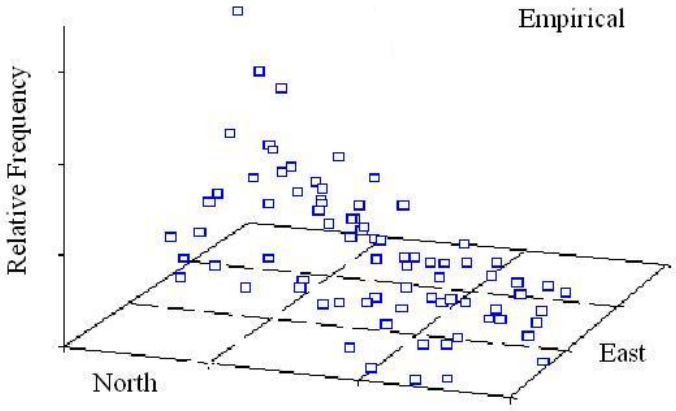


FIGURE 24. Three-dimensional scatter plot of the empirical probability distribution over the underlying distance north and distance east.

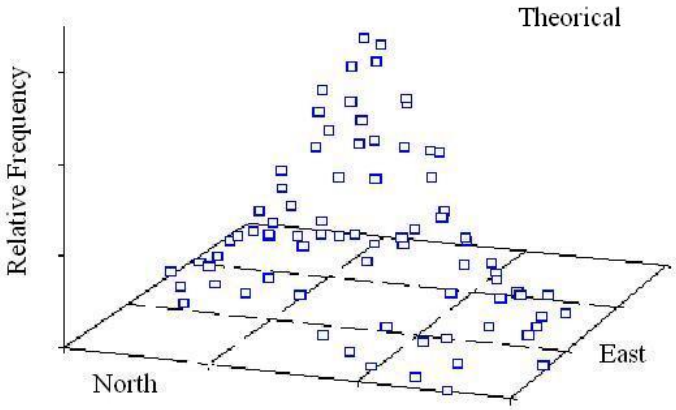


FIGURE 25. Three-dimensional scatter plot of the correlated bivariate normal distribution over the underlying distance north and distance east.

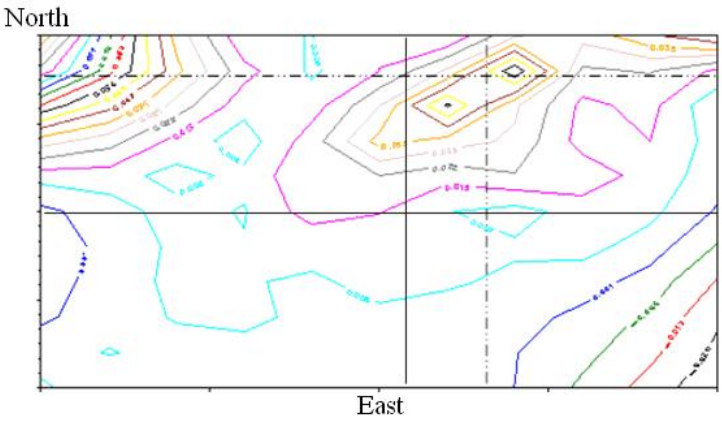


FIGURE 26. Three-dimensional scatter plot of the empirical probability distribution over the underlying distance north and distance east.

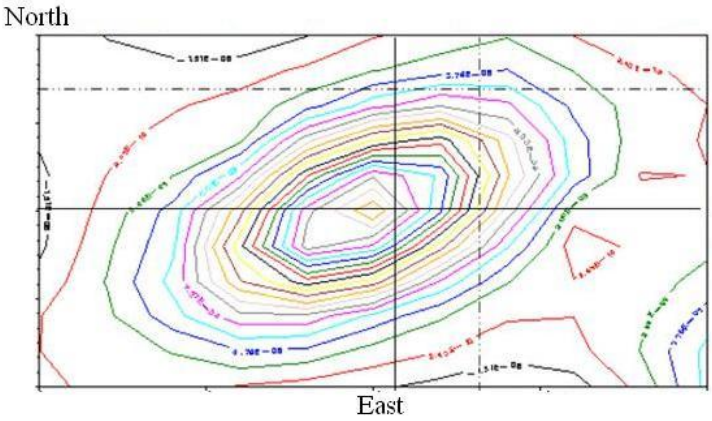


FIGURE 27. Three-dimensional scatter plot of the correlated bivariate normal distribution over the underlying distance north and distance east.

Comparison of Four Forms of the Bivariate Normal Probability Distribution. Recall that we defined the distance $d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$ between (x_i, y_i) is the location of the i^{th} sample and (x_c, y_c) is the location of the main vent or center and the angle $\theta_i = \tan^{-1}\left(\frac{y_i - y_c}{x_i - x_c}\right)$, off due east. We can therefore transform the northern and eastern directions into a centralized Cartesian plane by considering the transformed data $x_i := x_i - x_c$ and $y_i := y_i - y_c$. Then in *cis* notation, we have $r_i = \sqrt{x_i^2 + y_i^2} = d_i$ which corresponding radians are given by θ_i. To rotate this data to a primary and secondary axis where there is minimum or no correlation, we can rotate the data by a given angle α by defining $x_i' = r_i \cos(\theta_i + \alpha)$ and $y_i' = r_i \sin(\theta_i + \alpha)$. We can estimate the necessary value of α by considering the regressed slope between x' and y' set equal to zero. Accurate to the first decimal we have $\hat{\alpha} = 18.8°$ which shows an associated slope of $m = 0.000357$ and simple correlation of $r = 0.000801$. Scatter plots of these two coordinate systems are as shown by Figures and .

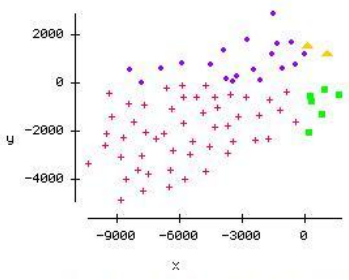


FIGURE 28. Scatter plot of original data.

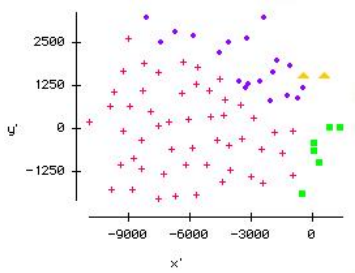


FIGURE 29. Scatter plot of the rotated data.

However, even with this rotation the mass is skewed and the majority of the tephra fall more in one direction than in any other, namely the direction in which the wind is blowing.

To compare these statistical models to the empirical distribution, consider the unitized probability distribution. Recall, the sample sets consist of $\Omega = \{(x_i, y_i) : i = 1, 2, \ldots, 80\}$ and the empirical probability at the given location is defined in terms of the mass at that location $p_i = P(x_i, y_i) = \dfrac{M(x_i, y_i)}{\sum\limits_{(x,y)\in\Omega} M(x, y)}$. Then, we can compute the best-fit correlated bivariate normal probability distribution for the transformed data $f(x_i, y_i)$ at the various locations for which we have data and then define the associated probability as $\hat{f}(x_i, y_i) = \dfrac{f(x_i, y_i)}{\sum\limits_{(x,y)\in\Omega} f(x_i, y_i)}$. Similarly, we can compute the best-fit non-correlated bivariate normal distribution for the transformed and rotated data $g(x_i', y_i')$ as defined by $\hat{g}(x_i', y_i') = \dfrac{g(x_i', y_i')}{\sum\limits_{(x,y)\in\Omega} g(x_i', y_i')}$. Then we can determine which distribution yields the best-fit.

Since x' and y' are independent, then $f_{\Omega'}(x', y') = f_{X'}(x') f_{Y'}(y')$, where $\Omega' = \{(x', y') \,|\, (x, y) \in \Omega\}$, $X' = \{x' \mid x \in X\}$ and $Y' = \{y' \mid y \in Y\}$. In this rotated coordinate system, the distribution of the minor axis is normally distributed; that is, $y' \sim N\left(\mu_{y'}, \sigma_{y'}^2\right)$. However, the distribution of the major axis is skewed; that is the variable x' is better described by the skewed distribution given by equation 6 where $g(x)$ is the standard normal probability distribution function with parameters $\mu_{x'}$ and $\sigma_{x'}^2$. The skewed probability distribution function is given by

$$f_{X'}(x') = 2g(x')\, G(\lambda x'), \tag{8.4}$$

where G is the cumulative normal probability distribution function, λ is the skewing factor and x' is the rotated coordinate defined above.

Thus, we have four probability distributions to consider: the transformed non-correlated bivariate normal (TNCN), the transformed correlated bivariate normal (TCN), the transformed rotated non-correlated bivariate distribution (TRNCN) and

finally, the (independent) skewed transformed rotated non-correlated bivariate normal (STRNCN).

To determine which of the four probability distributions best-fits the data, we proceed to determine which of the four probability distribution minimizes the measure of $\chi^2 = \sum_i \frac{(f(x_i, y_i) - p_i)^2}{p_i}$ and $\chi^2_{\text{adj}} = \sum_i \frac{\left(\hat{f}(x_i, y_i) - p_i\right)^2}{p_i}$. We can improve the fit of the distribution by letting $\hat{\lambda} = 5.23$; this skewness yields a significant decrease in the χ^2. According to this statistic, the transformed correlated bivariate normal is the best-fit. Let χ^2_0 represent the minimum measured χ^2. Then the larger such statistics are given as a multiple of this minimum statistic, see Table 8. Furthermore, with the adjusted statistics, χ^2_{adj} we have that the first two probability distributions (statistical models), the TNCN and the TCN are very similar, but the STRNCN shows a vast improvement over all the remaining statistical models. Therefore, we can also consider the correlation between p_i and f_i for the given probability distribution function, respectively. We compare these probability estimates given by Figures –.

This again indicates that the STRNCN explains more variation in the empirical probability distribution than the other statistical models with $R^2 = 39.3\%$; however, when we consider the simple correlation between p_i and $\hat{f}_i$ for the given probability distribution function, respectively — the TRNCN is the most explanatory. Note: all data is considered in its transposed form with graph given by Figures – in the coordinate system (x, y).

Figures 30–33, are on different scales; where the empirical probability in Figure 30 sum to one, the estimated probabilities in Figures 34 are based on the bivariate distribution, which is only a small portion of the total probability as stated previously.

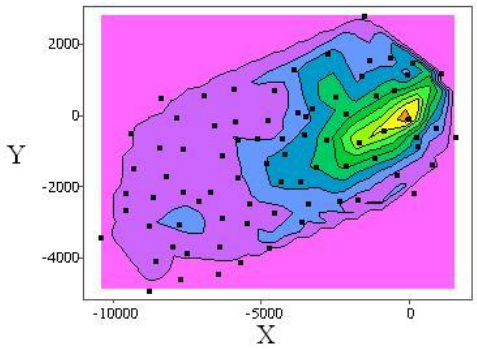


FIGURE 30. Contour plot of empirical probability distribution.

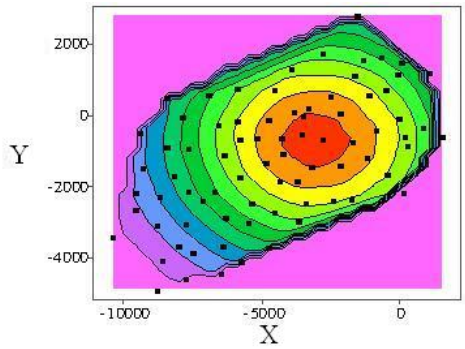


FIGURE 31. Contour plot for the estimated non-correlated bivariate normal distribution.

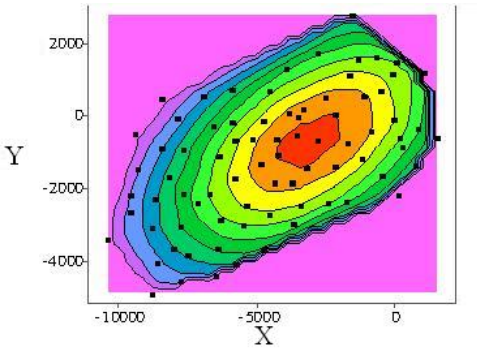


FIGURE 32. Contour plot for the estimated correlated bivariate normal distribution.

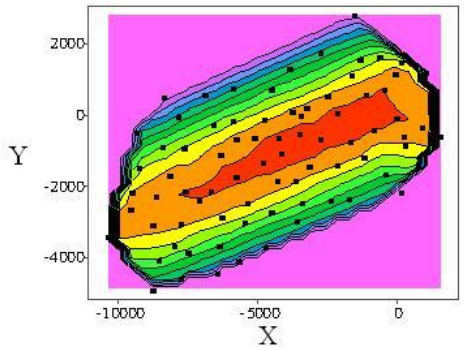


FIGURE 33. Contour plot for the estimated rotated non-correlated bivariate normal distribution.

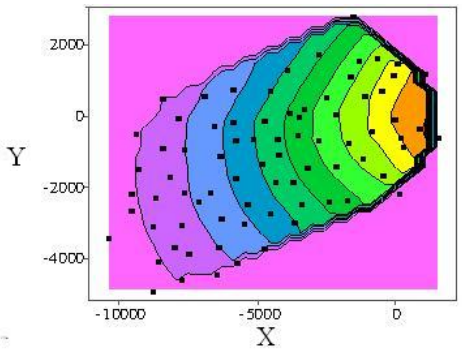


FIGURE 34. Contour plot for the estimated rotated non-correlated skewed bivariate normal distribution (assuming independence).

9. USEFULNESS

This type of analysis is extremely important on both a global scale as well as on a local scale. Global impacts are not just the possibility of ash in the area, but also the placement of such things as nuclear power plants and biohazard facilities. On a local scale, there is urban planning for economic growth and, more importantly, evacuation planning in case of a volcanic eruption. Unfortunately, this is not an exact science. In 2001, volcanologists stated with 95% confidence, that Cerro Negro would erupt again before 2005, but this has not yet come into fruition and it is over a year past this forecast [2]. Does this mean we are over due, and how does this additional time affect the probable magnitude (VEI) of the next volcanic eruption? Consider the trend over time.

10. CONCLUSION

The volcano data analyzed in this study was obtained from Cerro Negro, Nicaragua. We have 80 sample sites from which the tephra was measured in terms of mass and

sieved down into mass by particle size. The most probable VEI for Cerro Negro is 2. In terms of grain size, the probability distributions are the same for larger particle size; the skewed bivariate normal. Whereas the smaller the particle size the more uniform the distribution. The majority of the particles, however, are large enough to consider these masses combined.

To determine the bivariate probability distribution which best characterizes the subject response (location of ash fall), four probability distributions are tested for goodness-of-fit: the transformed non-correlated bivariate normal probability distribution (TNCN), the transformed correlated bivariate normal probability (TCN), the transformed rotated non-correlated bivariate normal probability distribution (TRNCN) and finally, the (independent) skewed transformed rotated non-correlated bivariate normal probability distribution (STRNCN). While with both the rotated non-correlated bivariate normal probability distribution and the correlated bivariate normal probability distribution, the fit is extremely tight, when measuring the goodness-of-fit using the chi-square statistics, χ^2 indicates that the correlated bivariate normal probability distribution best characterizes the distribution of tephra. However, R^2 indicates that the distribution fit to the data by first rotating the data to a primary axis and the non-correlated bivariate normal probability distribution used. Thus, neither of these probability distributions accounts for the skewness in the data. According the third way to determine the best-fit, χ^2_{adj} indicates that the skewed transposed (transformed by center to the main vent of the volcano), rotated (based on removing the correlation between northern and eastern direction) non-corrected (made to be independent) bivariate normal probability distribution best characterizes the behavior of the subject phenomenon. The TCN explains an estimated 35.2% to 54.6% of the variation in the empirical distribution; these values correspond to the correlation coefficient R^2, first between f_i and p_i, and second between $\hat{f}_i$ and p_i. The TRNCN explains an estimated 17% to 58.9% of the variation in the empirical distribution and the STRNCN explains an estimated 39.3% to 45.1% of the variation in the empirical distribution.

Knowing the bivariate probability distribution which best characterizes the behavior of the subject phenomenon (location of ash fall) is extremely important to urban planning, strategic planning and risk analysis.

ACKNOWLEDGEMENTS

We wish to thank Chuck Connor, the Chair of the Geology Department at the University of South Florida, for his data and valuable suggestions.

REFERENCES

1. C. B. Connor and B. E. Hill, *Three non-homogeneous Poisson models for the probability of basaltic volcanism: Application to the Yucca Mountain region, Nevada*, Journal of Geophysical Research **100** (1995), no. B6, 10,107–10,125.
2. C. B. Connor, B. E. Hill, B. Winfrey, N. M. Franklin, and P. C. La Femina, *Estimation of volcanic hazards from tephra fallout*, Natural Hazards Review **2** (2001), 33–42.
3. M. G. Genton (ed.), *Skew-elliptical distributions and their applications: A journey beyond normality*, Chapman and Hall/CRC, Boca Raton, FL, 2004.
4. R. Vogel, *The impact of natural disaster on urban and economic structures*, The Review of Radical Political Economics **30** (1998), no. 3, 114–122.

TABLE 1. Eruption data for Cerro Negro.

Year	Duration	VEI	Cum. (T) Volume m^3	Fall Volume m^3	Tephra Volume m^3	Lava Volume m^3	Column Height m	Area
1850	10 to 44	1 to 2	6.0E+06	4.3E+05	6.5E+05	5.4E+06		Form.
1867	16	2	1.0E+07	7.4E+06	8.6E+06		3.0E+03	NE/SW
1899	7 to 8	1 to 2	1.1E+07		1.7E+06			
1914	6 to 6	2	1.2E+07	2.8E+06	2.8E+06			
1919	10	0 to 2	1.2E+07					
1923	49	2 to 3	3.9E+07	1.7E+07	3.6E+07	1.0E+07	2.0E+03	Summit N R
1929	19	0 to 2	3.9E+07			1.0E+05		
1947	13 to 24	3 to 3	5.1E+07	2.3E+07	3.1E+07	3.8E+06	6.0E+03	Summit N F
1948	1	0 to 2	5.1E+07					
1949	1	0 to 2	5.1E+07					
1950	26	2 to 3	6.8E+07	2.8E+06	3.8E+07	1.0E+05	1.5E+04	
1954	1	0 to 2	6.8E+07					
1957	20	2	7.4E+07	2.8E+06	2.8E+06	4.5E+06	2.0E+03	Summit E F
1960	89	1 to 3	9.5E+07	1.1E+06	3.4E+07	5.2E+06	1.0E+03	Summit S F
1961	1	0 to 2	9.5E+07					NE Flank
1962	2	1	9.6E+07					
1963	1	0 to 1	9.6E+07					
1964	0	2	9.6E+07					
1968	48	2 to 3	1.2E+08	9.7E+06	2.7E+07	6.9E+06	2.0E+03	Summit S F
1969	10	0 to 1	1.2E+08					
1971	10.6 to 11	3	1.4E+08	3.0E+07	5.8E+07		5.0E+03	Summit E F
1992	3.6 to 5	3	1.5E+08	2.3E+07	2.6E+07		5.0E+03	
1995	79 to 191	1 to 2	1.5E+08		5.8E+06	3.7E+06	2.3E+03	
1995	13 to 15	2	1.6E+08	2.8E+06				
1999	2 to 3	1 to 2	1.6E+08	8.4E+05	1.0E+06	6.0E+05	1.0E+03	S. Flank

TABLE 2. Frequencies and proportions for the two data sources.

VEI	Frequency (1)	Frequency (2)	Probability (1)	Probability (2)
0	0	8	0%	35%
1	2	6	9%	26%
2	14	6	61%	26%
3	7	3	30%	13%
4	0	0	0%	0%
5	0	0	0%	0%

TABLE 3. Data by direction (eight sectors).

Sector	Freq.	Probability
1	0	0%
2	2	3%
3	5	6%
4	14	18%
5	49	63%
6	2	3%
7	4	5%
8	2	3%

TABLE 4. Data by direction (twenty-four sectors).

Sector	Count	Probability Cone	Mean Mass	Probability Mass in Cone
4	1	1.282	850.883	12%
6	1	1.282	753.294	11%
7	1	1.282	373.504	5%
8	2	2.564	267.685	4%
9	2	2.564	601.575	9%
10	3	3.846	568.754	8%
11	1	1.282	441.885	6%
12	10	12.821	362.741	5%
13	18	23.077	303.451	4%
14	21	26.923	394.468	6%
15	10	12.821	354.771	5%
16	1	1.282	215.712	3%
18	1	1.282	246.921	3%
19	1	1.282	343.031	5%
20	2	2.564	604.44	9%
21	1	1.282	97.82	1%
23	2	2.564	294.861	4%

TABLE 5. Mass by distance.

Distance Contour	Frequency	Relative Frequency	Total Mass	Percent Mass by Distance
1	1	1.266	1872.39	6%
2	1	1.266	620.231	2%
3	4	5.063	3214.428	10%
4	2	2.532	1098.318	4%
5	8	10.127	4914.92	16%
6	4	5.063	1838.584	6%
7	2	2.532	1755.53	6%
8	2	2.532	953.39	3%
9	4	5.063	1654.296	5%
10	3	3.797	1598.424	5%
11	1	1.266	329.566	1%
12	5	6.329	2241.005	7%
13	4	5.063	1307.428	4%
14	2	2.532	753.998	2%
15	1	1.266	225.219	1%
16	3	3.797	793.821	3%
17	4	5.063	1136.98	4%
18	2	2.532	540.326	2%
19	4	5.063	792.1	3%
20	1	1.266	238.067	1%
21	2	2.532	538.862	2%
22	3	3.797	420.882	1%
23	5	6.329	913.04	3%
24	1	1.266	142.721	0%
25	2	2.532	184.05	1%
26	4	5.063	848.004	3%
27	2	2.532	281.058	1%
28	1	1.266	89.403	0%
30	1	1.266	62.848	0%

TABLE 6. Mass by diameter and size.

Diameter d mm	Size $\phi = -\log_2 d$	Mass $^{kg}/_{km^2}$	Diameter d mm	Size $\phi = -\log_2 d$	Mass $^{kg}/_{km^2}$
16.00	-4	325.4	1.00	0	4792.37
11.31	-3.5	264.48	0.71	0.5	4370.21
8.00	-3	572.21	0.50	1	2621.48
5.66	-2.5	1079.65	0.35	1.5	1463.36
4.00	-2	1662.49	0.25	2	777.45
2.83	-1.5	2955.97	0.18	2.5	428.41
2.00	-1	3921	0.13	3	276.76
1.41	-0.5	4905.6	<0.09	>3	872.66

TABLE 7. Test for best-fit distribution.

Test: Diameter (Phi)	Normal	Lognormal	Exponential	Weibull
Kolmogorov–Smirnov	< 0.010	< 0.001	< 0.001	< 0.010 (< 0.001)
Cramer Von Miser	< 0.005	< 0.001	<0.001	< 0.005 (< 0.001)
Anderson Darling	< 0.005	< 0.001	< 0.001	< 0.005 (< 0.001)

TABLE 8. Descriptive statistics and regressed slope and correlation coefficient.

Statistic (x, y)	TNCN		TCN	
μ	-2964.3	-637.82	-2964.29	-637.82
σ^2	2751.46	1416.3	2751.46	1416.27
ρ	0		0.5378	
R^2	28.90% (52.9%)		35.20% (54.6%)	
χ^2	$1.3\chi_0^2$		χ_0^2	
χ^2_{adj}	0.413		0.422	
Statistic (x', y')	TRNCN		STRNCN	
μ	-3196.3	396.23	-3196.34	396.23
σ^2	2760.22	1137.5	2760.22	1137.48
ρ	0.000		0.000	
R^2	17% (58.9%)		39.30% (45.1%)	
χ^2	$11.9\chi_0^2$		$16.2\chi_0^2$	
χ^2_{adj}	0.607		0.265	