Mathematical Geology, Vol. 23, No. 2, 1991

Nonhomogeneous Poisson Model for Volcanic Eruptions[1]

Chih-Hsiang Ho[2]

A simple Poisson process is more specifically known as a homogeneous Poisson process since the rate λ was assumed independent of time t. The homogeneous Poisson model generally gives a good fit to many volcanoes for forecasting volcanic eruptions. If eruptions occur according to a homogeneous Poisson process, the repose times between consecutive eruptions are independent exponential variables with mean $\theta = 1/\lambda$. The exponential distribution is applicable when the eruptions occur "at random" and are not due to aging, etc. It is interesting to note that a general population of volcanoes can be related to a nonhomogeneous Poisson process with intensity factor $\lambda(t)$. In this paper, specifically, we consider a more general Weibull distribution, WEI (θ, β), for volcanism. A Weibull process is appropriate for three types of volcanoes: increasing-eruption-rate ($\beta > 1$), decreasing-eruption-rate ($\beta < 1$), and constant-eruption-rate ($\beta = 1$). Statistical methods (parameter estimation, hypothesis testing, and prediction intervals) are provided to analyze the following five volcanoes: Aso, Etna, Kilauea, St. Helens, and Yake-Dake. We conclude that the generalized model can be considered a goodness-of-fit test for a simple exponential model (a homogeneous Poisson model), and is preferable for practical use for some nonhomogeneous Poisson volcanoes with monotonic eruptive rates.

KEY WORDS: prediction interval, volcanism, Weibull distribution.

INTRODUCTION

It is the complex, unpredictable interaction of the factors governing the behavior of a volcano that makes it doubtful whether the exact time of an outbreak of a volcano can be accurately predicted. Although every volcano has an individual repose-period pattern, there are, nevertheless, several general types of patterns (Wickman, 1966, 1976), which make long-term forecasting possible for volcanoes with simple extreme patterns. Wickman observed that, for some volcanoes, the eruption rates were independent of time. These volcanoes were called "simple Poisson volcanoes." Wickman also used a series of repose states characterized by increasingly larger but time-independent rate parameters to

---

[1]Manuscript received 5 December 1989; accepted 26 February 1990.

[2]Department of Mathematical Sciences, University of Nevada-Las Vegas, 4505 Maryland Parkway, Las Vegas, Nevada 89154.

0882-8121/91/0200-0167$06.50/1

describe the repose-period patterns of several volcanoes other than simple Poisson volcanoes. However, the models presented by Wickman were based on the assumption that the conditions of the volcanic activity mainly changed stepwise. This paper is concerned with the use of a time-dependent rate parameter—a continuous model.

Theoretically, the probability model for simple Poisson volcanoes is derived from the following assumptions:

- Volcanic eruptions in successive time periods of length t for each period are independent and should follow a Poisson distribution with a constant mean (average rate) $\mu = \lambda t$, where λ is the average eruptive rate in unit time and is assumed to be constant throughout the entire life of the volcanic activity.

If λ is assumed constant over t, the process is referred to as a *homogeneous Poisson process (HPP)*. Since λ is constant and the increments are independent, it turns out that one does not need to be concerned about the location of the observation time interval, and the model $X \sim \text{POI}(\mu)$ is applicable for any interval of length t, $[s, s + t]$, $\mu = \lambda t$. That is, regardless of the interval chosen, the variable remains Poisson with the appropriate mean. The Poisson process is an important model for the repose times of a volcano. In this terminology, the HPP assumptions imply that the time to first eruption is a random variable that follows the exponential distribution, and also that the time between eruptions is an independent exponential variable. The assumption of a constant eruptive rate λ suggests that the volcanism, which depends on the availability of magma and a functioning triggering mechanism, as well as on their mutual interaction, is relatively uniform and does not get "exhausted" by loss of gases or for other reasons. If the volcanism is waning or developing, the model should be generalized to allow λ to be, respectively, a decreasing or increasing function of t. More generally, one might want to allow the eruptive rate to be an arbitrary nonnegative function of t.

NONHOMOGENEOUS POISSON PROCESS

If we replace the constant λ with a function of t, denoted by $\lambda(t)$, then another type of Poisson process can be derived, known as a *nonhomogeneous Poisson process (NHPP)*. If $X(t)$ denotes the number of occurrences in a specified interval $[0, t]$ for an NHPP, then it can be shown that (Parzen, 1962, p. 138)

$$X(t) \sim \text{POI}(\mu(t))$$

where

$$\mu(t) = \int_0^t \lambda(s)\, ds$$

The cumulative distribution function for the time to first occurrence, t_1, now becomes

$$F_1(t) = 1 - \exp[-\mu(t)]$$

An important choice for a nonhomogeneous intensity function is

$$\lambda(t) = (\beta/\theta)(t/\theta)^{\beta-1}$$

which gives

$$\mu(t) = (t/\theta)^{\beta}$$

In this case, the time to first occurrence follows a Weibull distribution, WEI (θ, β). This intensity parameter is an increasing function of t if $\beta > 1$, and a decreasing function of t if $\beta < 1$. Of course, the Weibull process is a generalization of the exponential case ($\beta = 1$, which assumes a no-memory property), so it is useful for situations which entail waning, growth, etc. For example, the birth process (new volcanoes) and the death process (extinction) of volcanoes are included also. In a Weibull process, the time to first occurrence, say T_1, follows a Weibull distribution WEI (θ, β). The time to second occurrence or the time between occurrences does not follow a Weibull distribution. This is in contrast to the exponential case in which the times between occurrences are also exponentially distributed. Thus, in the exponential case, the data could have come from either times between occurrences of a single Poisson process or from repeated observations on the time to first occurrence of several Poisson processes. (Or the data could be from variables not interpreted in terms of a Poisson process.) Thus, if Weibull data are to be interpreted in terms of the Weibull process, it must be remembered that the data represent repeated observations on the time to first occurrence of a Weibull process. As in the exponential case, the successive times of occurrences from a single Weibull process are of main interest, and some statistical results in this framework are discussed in the next section.

ANALYSIS FOR THE WEIBULL PROCESS

Bain (1978) discusses inference procedures of the Weibull process, and also gives additional references. Suppose we assume that the successive volcanic eruptions of a specific volcano follow a single Weibull process. Let t_1, . . . , t_n be the first n successive times of eruptions of a volcano. These times are measured from the beginning of the observation period (cumulative length of time over which the eruptions occur), so $t_1 \leq t_2 \leq \ldots \leq t_n$. The following theoretical results (for proof see Bain, 1978, Ch. 4) are useful for volcanic eruptive studies:

1) The maximum likelihood estimators for β and θ are

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n-1} \ln (t_n/t_i)}$$

and

$$\hat{\theta} = \frac{t_n}{n^{1/\hat{\beta}}}$$

2. A size α test of H_0: $\beta = \beta_0$ against H_A: $\beta \neq \beta_0$ is to reject H_0 if $2n\beta_0/\hat{\beta} < \chi^2_{\alpha/2}(2n - 2)$ or $2n\beta_0/\hat{\beta} > \chi^2_{1-\alpha/2}(2n - 2)$, where $\chi^2_{\alpha/2}(2n - 2)$ is the $100\alpha/2$ percentile of a chi-square distribution with $2n - 2$ degrees of freedom.
3. An r level prediction interval for t_{n+1} is $L \leq t_{n+1} \leq U$, where

$$U = t_n \exp \left\{ \left[\left(\frac{1 - r}{2} \right)^{-1/(n-1)} - 1 \right] \Big/ \hat{\beta} \right\}$$

and

$$L = t_n \exp \left\{ \left[\left(\frac{1 + r}{2} \right)^{-1/(n-1)} - 1 \right] \Big/ \hat{\beta} \right\}$$

First, the parameters estimated from Eqs. (1) and (2) provide us with a quantitative model to characterize the volcanic activity, which is the first step toward forecasting of future eruptions. Second, suppose we wish to decide whether an exponential distribution seems appropriate for the data collected or whether the more general Weibull distribution seems required. This suggests a test of H_0: $\beta = 1$ against H_A: $\beta \neq 1$. Result 2 indicates that a chi-square test is appropriate. And third, consider a single Weibull process for the volcanic eruptive times, and suppose that successive eruptive times $t_1, \ldots, t_n$ have been recorded. Perhaps the most natural question concerns when the next eruption will occur. This suggests that a prediction interval for t_{n+1} would be quite useful and meaningful in this framework. A prediction interval is a confidence interval for a future observation. Result 3 serves this purpose.

EMPIRICAL EXAMPLES

The eruption records (adopted from *Volcanoes of the World*, Simkin et al., 1981) of the following four volcanoes are studied for monotonic trends: Aso, St. Helens, Kilauea, and Yake-Dake. The time series of occurrence of flank eruptions of Etna is found to follow a homogeneous Poisson process (Mulargia et al., 1985). For comparison purposes, the data set of Etna in the paper of Mulargia et al. (1985) was added and assumed to be from a Weibull process. Several simplifying assumptions must be made in treating eruptions as events in time. Although the onset date of an eruption is generally well-defined by the time when lava first breaks the surface, the duration is harder to determine because of such problems as slowly cooling flows or lava lakes and the gradual decline of activity. We adopt the same definition for repose time as defined by Klein (1982). We, therefore, ignore eruption duration; instead, we take the onset date (based on year only) as most physically meaningful, and measure repose times from one onset date to the next. Thus, our definition of "repose time" differs from the classic one (a noneruptive period). This procedure seems justified because most eruption durations are much shorter than typical repose intervals (Klein, 1982). Each data set of a Weibull process consists of the cumulative length of time (measured in years) over which the eruptions occur. Based on the above definition of repose times, the eruptive dates may be directly transformed into a sequence of the intervals (repose times) between eruption for further analyses if an exponential distribution seems appropriate for the data set. Results of the statistical analyses are summarized in Table 1. Based on $\hat{\beta}$, the data suggest a waning or decrease in the eruptive rates through time for both St. Helens and Yake-Dake. All other volcanoes show the opposite trend ($\beta > 1$). The results of the significance tests are interesting. The p values indicate that the hypothesis of an exponential distribution is rejected at a 0.05 significance level ($p < 0.05$) for all five volcanoes, although Etna and Yake-Dake show moderate evidence against $H_0(\beta = 1)$. A verbal definition of p value is the chance of getting a departure from H_0 as or more extreme than that observed, calculated assuming H_0 to be true. Using Kilauea eruptions during the period 1884–1961, Wickman (1966) found nonrandom behavior, revealed as a lowered eruption rate after a repose of 2 years. A different result was obtained by Klein (1982), who used eruptions during the period 1918–1979. The result of an HPP for Kilauea obtained by Klein (1982), however, is not substantiated by the present approach of a general Weibull model, which uses eruptions over a longer period (1800–1979). For Kilauea, the number of reposes in each of the

Table 1. Summary of Empirical Results

Name of volcano	Aso	Etna	St. Helens	Kilauea	Yake-Dake
Analysis period	1800–1980	1605–1978	1831–1980	1800–1979	1910–1962
Number of eruptions (n)	66	48	9	69	21
$\begin{pmatrix}\hat{\beta}\\ \hat{\theta}\end{pmatrix}$	$\begin{pmatrix}1.7931\\ 17.3997\end{pmatrix}$	$\begin{pmatrix}1.432\\ 24.996\end{pmatrix}$	$\begin{pmatrix}0.4927\\ 1.7236\end{pmatrix}$	$\begin{pmatrix}2.2358\\ 26.9405\end{pmatrix}$	$\begin{pmatrix}0.6369\\ 0.4366\end{pmatrix}$
Test statistic $2n/\hat{\beta}$ for H_0: $\beta = 1$ vs. H_A: $\beta \neq 1$ (p value)	73.62 (0.000)	67.04 (0.032)	36.53 (0.005)	61.72 (0.000)	65.94 (0.012)
90% prediction interval for t_{n+1}	$\begin{pmatrix}1980\\ 1985\end{pmatrix}$	$\begin{pmatrix}1978\\ 1996\end{pmatrix}$	$\begin{pmatrix}1982\\ 2206\end{pmatrix}$	$\begin{pmatrix}1979\\ 1983\end{pmatrix}$	$\begin{pmatrix}1962\\ 1977\end{pmatrix}$

periods 1918–1924, 1925–1959, and 1960–1979 were 6, 11, and 28, respectively (Klein, 1982, Table 3), suggesting an increase in volcanic activity in the observed periods. There are several variations possible in goodness-of-fit testing. For the HPP model, the chi-square goodness-of-fit test (e.g., see Steel and Torrie, 1980, p. 529), based on count data, is often not reliable because of low degrees of freedom or low expected cell counts for some historical eruptive data. The Kolmogorov–Smirnov test (e.g., see Steel and Torrie, 1980, p. 535) is considered more reliable and is based on the repose times between eruptions, but does not take into account the relative positions of repose times. In other words, any random permutations of the same data set of repose times yield the same result if the Kolmogorov–Smirnov test was applied. In contrast, Eq. (1) is sensitive to the locations, numbers, and relative sizes (to t_n) of the ordered t_i's. If early sparse t_i's were accompanied later by dense t_j's toward t_n, then $\hat{\beta}$ would be large, showing an increasing rate of eruption through time, and vice versa. For this reason, a test of the eruptive rate of Etna indicates that it is increasing, but only slightly.

Finally, how can we give meaning to "90% prediction interval"? The answer lies in recalling the long-term frequency interpretation of probability: To say that an event A has probability 0.90 is to say that, if an experiment in which an event A is possible is performed over and over again, in the long run A will occur 90% of the time. That is, the procedure (Result 3) outlined for obtaining a 90% prediction interval succeeds 90% of the time in producing an interval that includes the next future eruption. For example, a 90% prediction interval for the true value of t_{n+1} for St. Helens is (1982, 2206). This interval is obviously quite wide, reflecting substantial variability in cumulative eruptive times (4, 11, 16, 17, 18, 22, 23, 26, 149) and a small sample size ($n = 9$). It seems that we have no choice but to form our notion of governing laws on the basis of data and to act accordingly. This is particularly true in volcanic studies, in which there are many unknown areas with respect to geologic understanding of volcanism.

CONCLUSIONS

In volcanic eruption modeling, geological considerations may suggest a certain distribution, but it is also important to have statistical techniques available to aid in selecting an appropriate model. One difficulty is that with a small sample size of historical eruptive data, several different distributions may appear acceptable, yet tail probabilities from these distributions may vary considerably. Thus, relatively large samples are usually required to verify the validity of a specified model (at some probability level); however, even with smaller sample sizes it may be possible to eliminate some models from consideration. There are several variations possible in goodness-of-fit testing. Our approach demonstrated in this paper is to consider a general family of distributions such as WEI (θ, β), and then decide whether some subset of this family, such as WEI (θ, 1) = EXP (θ), is valid. Thus, in this case, the test of H_0: $\beta = 1$ may be considered a goodness-of-fit test, and several examples of this type have already been discussed. It is desirable to arrive at the simplest model which can properly describe the volcanic activity. Of course, if the simple model is not correct, then poorer results may be achieved than if a more general model is used. The preceding type of example, of course, incorporates the assumption that the original model, in this case WEI (θ, β), is at least general enough. This assumption may have been settled earlier, either with previous data showing a developing (or waning) trend, or using some other geologic knowledge. If a valid model is assumed, then predictions of future eruptions should be useful and reliable.

ACKNOWLEDGMENTS

I benefited considerably from discussions with geologists E. I. Smith, D. Feuerbach, and T. Naumann of the Center of Volcanic and Tectonic Studies (CVTS) at UNLV. This work was supported through the Nevada Nuclear Waste Project office.

REFERENCES

Bain, L. J., 1978, Statistical Analysis of Reliability and Life-Testing Models: Marcel Dekker, New York, 450 p.

Klein, F. W., 1982, Patterns of Historical Eruptions at Hawaii Volcanoes: Journal of Volcanology and Geothermal Research, v. 12, p. 1–35.

Mulargia, F., Tinti, S., and Boschi, E., 1985, A Statistical Analysis of Flank Eruptions on Etna Volcano: Journal of Volcanology and Geothermal Research, v. 23, p. 263-272.

Parzen, E., 1962, Stochastic Processes: Holden-Day, San Francisco, 324 p.

Simkin, T., Seibert, L., McClelland, L., Bridge, D., Newhall, C., and Latter, J. H., 1981, Volcanoes of the World: Smithsonian Institution and Hutchinson Ross, Stroudsburg, Pennsylvania, 232 p.

Steel, R. G. D., and Torrie, J. H., 1980, Principles and Procedures of Statistics (2nd ed): McGraw-Hill, New York, 633 p.

Wickman, F. E., 1966, Repose-Period Patterns of Volcanoes: Ark. Mineral. Geol., v. 4, p. 291–367.

Wickman, F. E., 1976, Markov Models of Repose-Period Patterns of Volcanoes: *in* D. F. Merriam (Ed.), Random Processes in Geology: Springer, New York, p. 135–161.