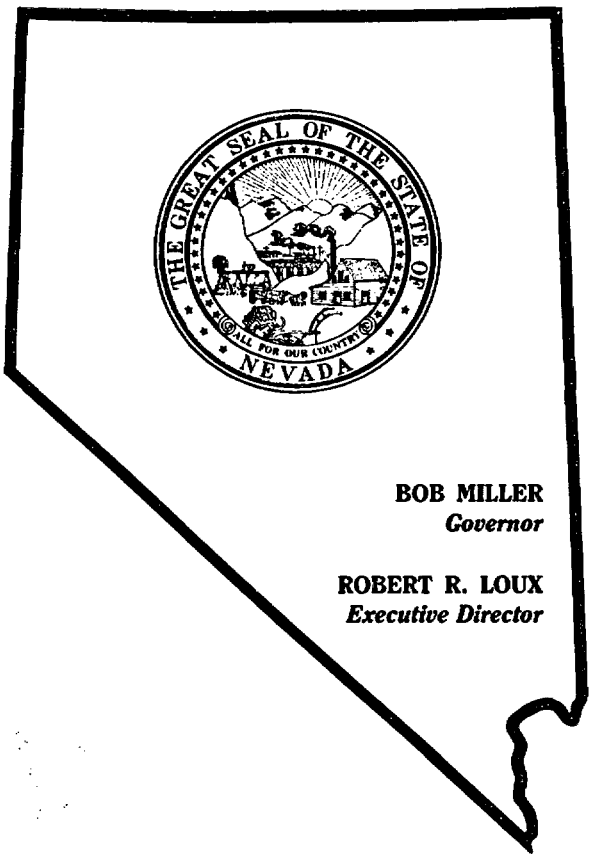
STATE OF NEVADA

AGENCY FOR NUCLEAR PROJECTS/ NUCLEAR WASTE PROJECT OFFICE



BOB MILLER
Governor

ROBERT R. LOUX
Executive Director

PRELIMINARY

John, FYI, K

THE MATHEMATICAL MODEL OF VOLCANISM AT YUCCA MOUNTAIN

Chih-Hsiang Ho

Department of Mathematical Sciences
University of Nevada, Las Vegas
Las Vegas, Nevada 89154

A Final Report Submitted to
the Nuclear Waste Project Office
State of Nevada

September 30, 1990

The Objectives of the Research

In the United States, spent fuel and high-level radioactive waste will be permanently disposed of in a geologic repository. Disposal of the spent fuel and high-level waste is scheduled to begin in the year 2010. The candidate site for the first U.S. geologic repository is located at Yucca Mountain, Nevada, approximately 100 miles, or about 160 kilometers, northwest of Las Vegas, Nevada. Comprehensive studies are underway on the potential host rock formation. These studies are called site characterizations.

An important element in assessing the suitability (or lack of suitability) of the Yucca Mountain site is an assessment of the potential for future volcanic activity. A potentially adverse condition with respect to volcanism is judged to be of concern at the Yucca Mountain site (Department of Energy, 1986) because of the presence of multiple basalt centers of Quaternary age.

Present understanding of eruptive mechanisms is not yet advanced enough to allow deterministic predictions of future activity. The only attempts at long-term forecasting have been made on statistical grounds, using historical records to examine eruption frequencies, types, patterns, risk, and probabilities. Reliable historical data make possible the construction of activity patterns for several volcanoes (Wickman, 1966, 1976; Klein, 1982, 1984; Mulargia et al., 1985; Mulargia et al., 1987; Condit et al., 1989). Unfortunately, there is no historical record of volcanism near Yucca Mountain, located within the Nevada Test Site (NTS). The volcanic record must therefore be developed by detailed field, geomorphic, and geochronologic studies.

Concern that future volcanism might disrupt the proposed Yucca Mountain repository site motivated the assessment of the volcanic risk to the Yucca Mountain area. Crowe and Carr (1980) calculate the probability of volcanic disruption of a repository at Yucca Mountain using a method developed largely by Crowe (1980). Crowe et al. (1982) refine the volcanic probability calculations for the Yucca Mountain area using a simple Poisson model. Although the simple Poisson model has proved successful in a wide range of situations, it might be inadequate to model the volcanism at NTS for the following reasons:

a) A simple Poisson model does not allow for the possibility of a waning (or developing) volcanic time trend, which is one of the major concerns in quantifying the volcanism at the Yucca Mountain region. It should be obvious that the chronological order in which the volcanic eruptions occur is an extremely important aspect of a historical eruptive data set. The simple Poisson model, however, assumes that the average recurrence rate (λ) is constant throughout the entire life of the volcanic activity.

A new development in volcanic studies is the possibility that the scoria cone of the Lathrop Wells volcanic center, one of the youngest volcanic centers in the Yucca Mountain region, is significantly younger than associated lava flows (Wells et al., 1988). Further studies have shown that at least three and possibly more of the seven Quaternary volcanic centers in the region exhibit polycyclic activity (Crowe et al., 1989 and Wells et al., 1990). It is therefore of interest to explore alternative model(s) derived from less restrictive model assumptions to see how the time trend could be evaluated.

b) Given the extremely limited nature of the geologic data at the Yucca Mountain area, the use of the most elementary statistical model (i.e., a simple Poisson model) is hard to justify. As has been mentioned earlier, there is a large and growing body of literature on probabilistic modeling for volcanism. Much of the debate in the literature is centered on the choice of distribution models (principally homogeneous Poisson versus nonhomogeneous Poisson models). There are several variations possible in goodness-of-fit testing. For the homogeneous Poisson 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 volcanic 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 general, relatively large samples are usually required to verify the validity of a specified model (at some probability level).

There is a developing literature on volcanic recurrence models that include plots of cumulative volume of volcanic events versus time (Shaw 1980, 1987; Wadge, 1982; Bacon, 1982). Crowe and Perry (1989) use the curve of cumulative magma volume plotted versus time to evaluate whether the volcanism at NTS indicates steady-state eruptive behavior in magma production. They then evaluate the slope of the curve as an estimate of the annual rate of magma production. I consider this approach questionable. First, a simple Poisson model requires a constant rate of occurrence, which is not the same as magma production. Also, the trend

of magma volume has not proved relevant to that of the frequency of the volcanic events (eruptions) at the NTS area. Second, the degree of erosional modification of volcanic landforms for the Yucca Mountain region should be studied to estimate volumes of missing volcanic deposits.

Crowe and Perry (1989) object that cone counts record only the recognition of a volcanic event, not its magnitude, and so they refine the parameter estimation by concentrating on the cumulative magma volume which is a continuous variable. Nonetheless, their model assumptions and development are still based on a discrete simple Poisson model, which treats each eruption equally in order to calculate the final probability. So far, the problems of model assumptions and parameter estimations have been treated only separately by Crowe et al. (1982) and Crowe and Perry (1989), despite the fact that the model assumptions and parameter estimation methods virtually always depend on each other in volcanic hazard and risk calculations.

For the reasons discussed, a formal structure, with conclusions depending on the model assumptions, needs to be developed to ensure that volcanic risk assessment is based on an adequate model and a complete volcanic record of the Yucca Mountain region.

Nonhomogeneous Poisson Models Developed

A Compound Poisson Model with a Gamma Compounding Density

The simple Poisson model generally gives a good fit to many volcanoes for volcanic eruption forecasting. Nonetheless, empirical evidence suggests that volcanic activity in successive equal time-period

tends to be more variable than a simple Poisson with constant eruptive rate. An alternative model is therefore examined in which eruptive rate (λ) for a given volcano or cluster(s) of volcanoes is described by a gamma distribution (prior) rather than treated as a constant value as in the assumptions of a simple Poisson model. Bayesian analysis is performed to link two distributions together to give the aggregate behavior of the volcanic activity. When the Poisson process is expanded to accommodate a gamma mixing distribution on λ, a consequence of this mixed (or compound) Poisson model is that the frequency distribution of eruptions in any given time-period of equal length follows the negative binomial distribution (NBD). Applications of the proposed model and comparisons between the generalized model and simple Poisson model are discussed based on the historical eruptive count data of volcanoes Mauna Loa (Hawaii) and Etna (Italy). Several relevant facts lead to the conclusion that the generalized model is preferable for practical use both in space and time. Moreover, this exceedingly flexible model has only two parameters which can be easily determined from the eruptive court data.

A Nonhomogeneous Poisson Process with Weibull Intensity

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 model, specifically, we consider a more general Weibull distribution, WEI(θ,β), for volcanism. A Weibull process is appropriate for three types of volcanoes: increasing-recurrence-rate (β>1), decreasing-recurrence-rate (β<1), and constant-recurrence-rate (β=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.

Statistical Estimation of Recurrence Rates of Volcanism near the Yucca Mountain Site

Investigations are currently underway to evaluate the impact of potentially adverse conditions (e.g., volcanism, faulting, seismicity) on the waste-isolation capability of the proposed nuclear waste repository at Yucca Mountain, Nevada, U.S.A. We discuss three estimating techniques for determining the recurrence rate of volcanic eruption (λ), an important parameter in the Poisson probability model. The first method is based on the number of events occurring over a certain observation period, the second is based on repose times, and the final is based on magma volume.

All of the statistical estimation methods considered for λ require knowing the value of E (total number of eruptions during the observation period). An accurate count of E is possible for volcanoes with a complete historical record. Identifying E, however, depends strictly on a clear understanding of eruptive processes and reliable dating techniques for the NTS region, since no historical record is available. Scientists differ in their opinions of volcanism at the NTS area. The following is the view of Crowe et al. (1983):

> Basalt centers are composed of multiple vents, each marked by a scoria cone. In this NTS region the cones are divided into two categories: large central cones, referred to as the main cones, and satellite cones. The average number of cones at a single center, based on cone counts of seven Quaternary basalt centers in the NTS region, is about 2 to 3 cones. Thus, field data suggest a general eruption pattern where the initial breakthrough of magma to the surface is marked by the development of an eruptive fissure with two or three loci of effusion. Each of these vents becomes the site of small scoria cones. As the eruption proceeds, activity shifts or concentrates at a single vent that becomes the site of the main scoria cone.

The above description indicates that a main scoria cone is the final stage of a single eruption, and a single eruption could have several small vents to accompany the main cone. However, the possibility of polygenetic (and polycyclic) volcanism at all the volcanic centers needs

to be evaluated. λ would be underestimated if nearby vents have distinguishable ages. We, therefore, estimate E as follows:

Let I denote the number of volcanic centers under investigation, and let J_i be the number of main cones in the ith volcanic center, where i=1, ..., I. The proposed estimate of E is:

$$\hat{E} = \sum_{i=1}^{I} \sum_{j=1}^{J_I} (m_{ij} + e_{ij}), \qquad (1)$$

where m_{ij} = number of multiple, time-separate eruptions of the jth main cone in the ith volcanic center,

and e_{ij} = number of vents that are separate in space and time (with distinguishable age measurements) from the jth main cone in the ith volcanic center.

The rationale for this estimate is that significant information has emerged that some of the volcanic centers are polycyclic volcanoes (e.g., Lathrop Wells center (Wells et al., 1990)). This estimation for parameter E (total number of eruptions) given by Equation 1 takes into account such a possibility for the NTS area.

Future Work

Crowe and Perry (1989, Figure 1) divide the Cenozoic volcanism of the Yucca Mountain region into three episodes that include 1) an older episode of large volume basaltic volcanism (12 to 8.5 Ma) that coincides in time with the termination of silicic volcanic activity; 2) the formation of five clusters of small volume basalt scoria cones and lava flows (9 to 6.5 Ma), all located north and east of the Yucca Mountain site; and 3) the formation of three clusters of small volume

basalt centers (3.7 to .01 Ma), all located south and west of the Yucca Mountain site. The two youngest episodes form northwest-trending zones that parallel the trend of structures in the Spotted Range-Mine Mountain section of the Walker Lane belt. Crowe and Perry (1989) suggest a southwest migration of basaltic volcanism in the Yucca Mountain area based on this structural parallelism, a pattern that may reflect an earlier southwest migration of silicic volcanism in the Great Basin. Smith et el. (1990) also provide a different point of view of the migration trends of volcanism in the Yucca Mountain region.

The important questions that we attempt to answer for assessment of volcanic risk for the Yucca Mountain site include:

1. What is the overall time trend of the volcanic activity?
2. Can the next episode be predicated?
3. What does the youngest episode (3.7 to 0.01 Ma) tell us? Is the time trend increasing?
4. Can the probabilities of future eruptions and site disruption be predicted?

REFERENCES

Bacon, C.R., 1982. Time-Predictable Bimodal Volcanism in the Coso Range, California. Geology, 10: 65-69.

Crowe, B.M., 1980. Disruptive Event Analysis: Volcanism and Igneous Intrusion. Battelle Pacific Northwest Laboratory Rep PNL - 2882.

Crowe, B.M. and Carr, W.J., 1980. Preliminary Assessment of the Risk of Volcanism at a Proposed Nuclear Waste Repository in the Southern Great Basin. US Geologic Survey Open-File Report 80-375, 15p.

Crowe, B.M., Harrington, C., Turrin, B., Champion, D., Wells, S., Perry, F., McFadden, L. and Renault, C., 1989. Volcanic Hazard Studies for the Yucca Mountain Project. Waste Management 89, Volume 1 - High Level Waste and General Interest, pp. 485-491.

Crowe, B.M., Johnson, M.E. and Beckman, R.J., 1982. Calculation of the Probability of Volcanic Disruption of a High-Level Radioactive Waste Repository within Southern Nevada, U.S.A. Radioactive Waste Management, 3: 167-190.

Crowe, B.M. and Perry, F.V., 1989. Volcanic Probability Calculations for the Yucca Mountain site: Estimation of Volcanic Rates, Focus '89, Nuclear Waste Isolation in the Unsaturated Zone, September, 1989, hosted by the Nevada section of the American Nuclear Society, pp. 326-334.

Department of Energy, 1986. Final Environmental Assessment: Yucca Mountain site, Nevada Research and Development, Nevada: Washington, D.C., DOE/RW-0073.

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

Klein, F.W., 1984. Eruption Forecasting at Kilauea Volcano, Hawaii. Journal of Geophysical Research, 89: 3059-3073.

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

Mulargia, G., Gasperini, P. and Tinti, S., 1987. Identifying Different Regimes in Eruptive Activity: an Application to Etna Volcano. Journal of Volcanology and Geothermal Research, 34: 89-106.

Shaw, H.R., 1980. The Fracture Mechanisms of Magma Transport from the Mantle to the Surface, In: R.B. Hargraves (editor), Physics of Magmatic Transport. Princeton University Press, pp. 201-264.

Shaw, H.R., 1987. Uniqueness of Volcanic Systems. In: R.W. Decker, T.L. Wright, and P.H. Staufer (editors), Volcanism in Hawaii. U.S. Geologic Survey Professional Paper 1350.

Smith, E.I., Feuerbach, D.L., Naumann, T.R. and Faulds, J.E., 1990. The Area of Most Recent Volcanism Near Yucca Mountain, Nevada: Implications for Volcanic Risk Assessment. Proceedings of the International Topical Meeting, High-Level Radioactive Waste Management: American Nuclear Society/American Society of Civil Engineers, Vol. 1: pp. 81-90.

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

Wadge, G., 1982. Steady State Volcanism: Evidence from Eruption Histories of Polygenetic Volcanoes. Journal of Geophysical Research, 87: 4035-4049.

Wells, S.G., McFadden, L.D., Renault, C.E., Turrin, B.D. and Crowe, B.M., 1988. A Geomorphic Assessment of Quaternary Volcanism in the Yucca Mountain Area, Nevada Test Site, Southern Nevada. Proceedings of the Geological Society of America, Cordilleran Section, Las Vegas, Nevada, March, Vol. 20, p. 242.

Wells, S.G., McFadden, L.D., Renault, C.E. and Crowe, B.M., 1990. Geomorphic Assessment of Late Quaternary Volcanism in the Yucca Mountain area, Southern Nevada: Implications for the Proposed High-level Radioactive Waste Repository. Geology, 18: 549-553.

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

Wickman, F.E., 1976. Markov Models of Repose-Period Patterns of Volcanoes. In: D.F. Merriam (Editor), Random Processes in Geology. Springer, New York, N.Y., pp. 135-161.

APPENDIX

Copies of Professional Papers

A. "Bayesian Analysis of Volcanic Eruptions," accepted for publication in Journal of Volcanology and Geothermal Research

B. "Nonhomogeneous Poisson Model for Volcanic Eruptions," accepted for publication in Mathematical Geology

C. "Eruptive Probability Calculation for the Yucca Mountain Site, U.S.A.: Statistical Estimation of Recurrence Rates," with E.I. Smith et al., revised and resubmitted to Bulletin of Volcanology

I am pleased to inform you that your article

Date 12 April, 1990

"Bayesian analysis of volcanic eruptions"

has been accepted for publication in

Journal of Volcanology & Geothermal Research

Dr. C-H. Ho
Department of Mathematical Sciences
University of Nevada
4505 Maryland Parkway
Las Vegas, Nevada 89154
U.S.A.

Formal written transfer of copyright on each article from the author(s) to the publisher is needed. This transfer enables the publisher to provide for the widest possible dissemination of the article through activities such as: distribution of reprints; authorization of reprints, translation, or photocopies by others; production of microfilm editions; and authorization of indexing and abstracting services in print and data base formats. Without this transfer of copyright, such activities by the publisher and the corresponding spread of information are limited.

We therefore ask that you complete and return the attached form promptly.
Your cooperation is essential and appreciated.

Editor

ELSEVIER SCIENCE PUBLISHERS B.V.
P.O. BOX 330, 1000 AH AMSTERDAM, THE NETHERLANDS

BAYESIAN ANALYSIS OF VOLCANIC ERUPTIONS

Chih-Hsiang Ho

Department of Mathematical Sciences
University of Nevada, Las Vegas, Nevada
4505 Maryland Parkway
Las Vegas, Nevada 89154 (U.S.A.)

ABSTRACT

The simple Poisson model generally gives a good fit to many volcanoes for volcanic eruption forecasting. Nonetheless, empirical evidence suggests that volcanic activity in successive equal time-periods tends to be more variable than a simple Poisson with constant eruptive rate. An alternative model is therefore examined in which eruptive rate (λ) for a given volcano or cluster(s) of volcanoes is described by a gamma distribution (prior) rather than treated as a constant value as in the assumptions of a simple Poisson model. Bayesian analysis is performed to link two distributions together to give the aggregate behavior of the volcanic activity. When the Poisson process is expanded to accommodate a gamma mixing distribution on λ, a consequence of this mixed (or compound) Poisson model is that the frequency distribution of eruptions in any given time-period of equal length follows the negative binomial distribution (NBD). Applications of the proposed model and comparisons between the generalized model and simple Poisson model are discussed based on the historical eruptive count data of volcanoes Mauna Loa (Hawaii) and Etna (Italy). Several relevant facts lead to the conclusion that the generalized model is preferable for practical use both in space and time.

INTRODUCTION

The evaluation of eruptive probabilities for a given volcano or a volcanic center remains an open problem in the definition of volcanic risk. There are many unknown areas with respect to

geologic understanding of volcanic activity, despite the fact that there are well recognized means of gathering data (field mapping, determinations of the eruptive history of basaltic centers, petrology, geochemistry, geochronology including magnetic polarity determinations, tectonic setting, and geophysical studies) that are well advanced. Present understanding of eruptive mechanisms is not yet advanced enough to allow deterministic predictions of future activity to be put forward. The only attempts at long term forecasting have been made on statistical grounds, using historical records to examine eruption frequencies, types, patterns, risk, and probabilities (Wickman 1966, 1976; Crowe et al., 1982; Klein 1982; Mulargia et al., 1985; Condit et al., 1989). The application of statistical methods to volcanic eruptions was put onto a sound analytical footing by Wickman (1966, 1976), in a series of papers which discussed the applicability of the methods and the evaluation of eruptive rates for a number of volcanoes. Wickman observed that, for some volcanoes, the eruption rates were independent of time. These volcanoes were called "simple Poissonian volcanoes". Theoretically, the probability model for simple Poissonian 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) λt, where λ is the average eruptive rate in unit time and is assumed to be constant throughout the entire life of the volcanic activity. λ is called the parameter of a Poisson model.

Relevant standard statistical methods (Runs test, Kolmogorov - Smirnov test, etc.) have been surveyed by Davis (1986, Ch. 4) and Tazieff (1983, Ch. 6), and have been widely applied to closely examine the above assumptions for volcanoes such as Etna (Mulargia et al., 1985), Kilauea, and Mauna Loa (Klein, 1982). Although the simple Poisson model has proved useful in some situations, assumptions of a simple Poisson process are rather restrictive for longer observation time and, consequently, fail to cover a more general population of volcanoes. It is therefore of interest to explore suitable models derived from different assumptions for volcanoes other than simple Poissonian volcanoes.

In this article we are concerned with the assumption that the eruptive rate in unit time, λ, should be constant. One well-known characteristic of the Poisson distribution is that the mean and variance have the same value. In some cases these properties may not be valid, and a more flexible model is required.

GAMMA PRIOR AND BAYESIAN ANALYSIS

In searching for a more flexible model, we consider, at least conceptually, a large volcanic center (belt) with cluster(s) of volcanoes in which, for any given volcano, the number of volcanic eruptions in [0,t] follows a Poisson process with rate $\mu = \lambda t$; however, λ may be different from volcano to volcano. In particular, for this conceptually large population of volcanoes, we assume that λ is a continuous random variable that follows a probability density function $g(\lambda)$. There is one more scenario that suggests this generalization, which is also suitable for a single

volcano. It arises from the fact that, although eruptions are caused by specific physical events or processes, there might be many causal factors with random influences on the sequence of eruptions. As a result, parameter λ is a random variable. Note that in the context of a Bayesian analysis of a Poisson model, the density $g(\lambda)$ corresponds to a prior density for the parameter, and the mathematics involved in the Appendix is essentially equivalent to that involved in the associated Bayesian analysis. The differences between the suggested scenarios for this generalization and the Bayesian approach to the same problem depend on the philosophy for introducing a density function for the parameter and the interpretation of the results. In a Bayesian analysis the prior density reflects a degree of belief about the value of the parameter or some previous information about the value of the parameter (Interested readers are referred to Berger, 1988, Ch. 4).

We now turn to the modification of the Poisson scheme which is derived by supposing that the average eruptive rates in successive time periods of equal length are not the same. Without loss of generality, each observation period will be treated as unit time, i.e. $t=1$. One naturally commences with the assumption that $g(\lambda)$ has a normal distribution. But the assumption of a normal distribution cannot be justified since values of λ may range from zero far in the positive direction so that $g(\lambda)$ should be skew. The distribution of λ we adopt is the gamma distribution. This is plausible, since such a distribution is fairly flexible (having two

adjustable parameters) and the right shape (i.e. a continuous distribution for non-negative values that is reversed-J-shaped or hump-backed and always positively skewed). In the Bayesian analysis, the gamma distribution is a prior from the conjugate family for the Poisson population (DeGroot, 1970, p. 164). Klein (1984) has also used the gamma distribution as one of the smoothing curves to fit the empirical data of the eruption forecasting parameters.

NEGATIVE BINOMIAL PREDICTORS

When the Poisson process is expanded to accommodate a gamma mixing distribution on λ, an immediate consequence of this mixed (or compound) Poisson model is that the frequency distribution of eruptions in any given time-period of equal length follows the negative binomial distribution (abbreviated NBD). (See Appendix.) The probability of x eruptions is:

$$P(x) = \frac{\Gamma(r+x)}{\Gamma(r)x!} \left[\frac{a}{a+1}\right]^r \left[\frac{1}{a+1}\right]^x, \quad x = 0, 1, 2, \ldots, \qquad (1)$$

where r and a are, respectively, the shape and scale parameters of the gamma distribution (mean = r/a, variance = r/a^2). If the future observation time interval increased by a factor of t then a in the right-hand side of (1) is replaced by at. Equation (1) is a form of negative binomial distribution, and it is referred to as a compound Poisson distribution with a gamma compounding density. Thus, the NBD represents a generalization of the Poisson distribution, and it converges to the Poisson distribution when the gamma prior density becomes degenerate at a constant (Feller, 1968, p. 281). The negative binomial model, therefore, can be used as an

alternative to the Poisson model in analyzing volcanic eruption count data, particularly when the variance and the mean cannot be assumed equal.

THE POSTERIOR MEAN OF λ

Bayesian analysis is performed by combining the prior information ($g(\lambda)$) and the sample information (x) into what is called the posterior distribution of λ given x, from which all decisions and inferences are made. The posterior distribution of λ given x (or posterior for short) is denoted $g(\lambda|x)$. The name "posterior distribution" is indicative of the role of $g(\lambda|x)$. Just as the prior distribution reflects the value of λ prior to experimentation, so $g(\lambda|x)$ reflects the updated beliefs about λ after (posterior to) observing the sample x. In other words, the posterior distribution combines the prior beliefs about λ with the information about λ contained in the sample, x, to give a composite picture of the final beliefs about λ. The eruptive rate λ plays an important role in the evaluation of volcanic risk and volcanic forecast. The mean of this up-dated posterior distribution is derived in the Appendix as

$$E\,[\lambda|x] = \frac{r + n\bar{x}}{a + n}, \qquad (2)$$

where n is the total number of observation periods with equal length and $\bar{x}$ is the sample average. Again, if the future observation period increased by a factor of t then the right-hand side of (2) is multiplied by t.

PARAMETER ESTIMATION AND MODEL ADEQUACY CHECKING

General sampling theory of the NBD has been discussed by Anscombe (1950), who also gives earlier references. The mean and variance for the negative binomial variable in (1) are given by

$$EX = r/a,$$

and

$$Var(x) = r(a + 1)/a^2.$$

In spite of the fact that maximum likelihood estimates for the NBD were given (implicitly) by Anscombe (1950), the two common techniques for estimation of (r,a) (required to evaluate (1)) have been the method of moments and fitting the mean and zeros (Anscombe, 1950). The second method is more attractive in terms of efficiency. The technique is to equate the observed proportion of zero counts P_0 to its expected value, i.e. to write

$$\left[\frac{\hat{a}}{\hat{a} + 1}\right]^{\hat{a}\,\bar{x}} = P_0, \tag{3}$$

and

$$\hat{r} = \hat{a}\,\bar{x}, \tag{4}$$

where $\bar{x}$ is the observed sample mean. The first equation is solved for $\hat{a}$ by iteration. Once $\hat{a}$ and $\hat{r}$ are estimated, the probability of future eruptions can be calculated based on (1). If $P_0 = 0$, then the method of moments is a natural substitute. Method of moment estimates for the NBD are:

$$\hat{a} = \bar{x}/(s^2 - \bar{x}),$$

$$\hat{r} = \hat{a}\,\bar{x},$$

where $\bar{x}$ and s^2 are the observed mean and variance, respectively.

The goodness-of-fit of the estimated distribution can be tested by comparing the observed and theoretical frequencies using the chi-square goodness-of-fit test. A slightly more laborious test, for large samples, is to compare the observed variance with the theoretical value $r(\alpha+1)/\alpha^2$ as proposed by Anscombe, 1950.

EMPIRICAL EXAMPLES

The data set used here consists of historical eruptions of volcanoes Etna (Table 1) and Mauna Loa (Table 2).

Volcano Etna

A complete record exists for the eruptive activity of Etna. The time series of occurrence of flank eruptions in the period 1600-1980 is found to follow a simple Poisson process (Mulargia et al., 1985). A similar result was found by Vickman (1976), although on a less reliable data set. The purpose of the following analysis is not to reexamine the previous results, but to investigate the sensitivity of the proposed NBD model to a simple Poissonian volcano based on eruptive count data. For comparison purpose, we adopt the same observation time intervals as described in Mulargia et al., 1985. The year 1605 was taken as a starting point and the period 1605-1980 was divided into 25 fifteen-year intervals. The frequency distribution of this set of data yields sample variance $s^2 = 2.826$ and sample mean $\bar{x} = 1.920$. These results suggest that the mean and variance of X may not be equal. However, additional distribution results are needed in order to indicate whether this magnitude of difference between s^2 and $\bar{x}$ reflects a true difference or is due to random variation. It has been shown in the literature

(e.g., Steel and Torrie, 1980, p. 530-531) for this case that approximately

$$\frac{(n-1)\ S^2}{\bar{X}} \sim \chi^2(n-1)$$

when X does follow a Poisson model (it may yield different conclusions than other methods). In our problem $(n-1)s^2/\bar{x}$ = 24(2.826)/1.920 = 35.325. Now $P[\chi^2(24) \geq 35.325] \doteq 0.06$, thus the observed ratio is larger than would be likely. However, there is about a 6% (p-value) chance of getting such a result when the Poisson model is valid. Therefore, the hypothesis of a simple Poisson model can not be rejected at 0.05 significance level, which is consistent with the results in the previous studies (Mulargia et al., 1985, and Vickman, 1976). We then use the generalized model (NBD) to fit this data set. Equations 3 and 4 yield $\hat{a} = 1.079 \times 10^8$ and $\hat{r} = 2.072 \times 10^8$. The estimated mean and variance of the gamma prior distribution are:

$$\hat{\mu}_\lambda = \hat{r}/\hat{a} = 1.920 = \bar{x}$$

and

$$\hat{\sigma}^2_\lambda = \hat{r}/\hat{a}^2 = 1.78 \times 10^{-8} \doteq 0.$$

The near zero variance of the gamma prior indicates that the gamma prior density degenerates at $\bar{x} = 1.920$, which is the same estimate for the eruptive rate as if a simple Poisson model were used. The probability of x eruptions based on equation (1) can then be approximated by the Poisson probability distribution function:

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!},\ x = 0,\ 1,\ \ldots$$

Thus, the NBD does represent a generalization of the Poisson distribution. Further analyses for this simple Poisson model were not attempted to avoid duplicating similar work already in the literature.

Volcano Mauna Loa

Klein (1982) has performed statistical test for randomness on the repose times between eruptions for Mauna Loa using eruptions during the period 1832-1975. Except for small phases of summit activity, data for Mauna Loa eruptions are complete since 1832, owing to visibility of the mountain from several coastal towns. Our attempt to analyze longer periods of time leads us to adopt the historical eruptions recorded in the *Volcanoes of the World* (Simkin et al., 1981). According to their record, the first eruption started in 1750. After a period of total inactivity (or loss of data) of 82 years, in 1832 Mauna Loa renewed its activity. The reliability of the historical eruptive counts is not our concern here. On the contrary, we intentionally include this possibly unreliable eruption to perturb a potential simple Poisson process in contrast to the previous analysis for Etna. Based on Table 2, the year 1750 was taken as a starting point and the period 1750-1979 was divided into 46 five-year intervals. The frequency distribution provides sample variance $s^2 = 1.169$ and sample mean $\bar{x} = 0.826$. The chi-square test statistic (= 63.69) for the observed ratio yields p-value = 0.03, which rejects the hypothesis of a simple Poisson model at 0.05 significance level and concludes that a NBD model is indicated for further analyses. The results are summarized in

Table 3. Generally speaking, the fit is adequate except at $x = 3$ where the observed count is zero. Note that the observed and theoretical frequencies at $x=0$ are always identical since the parameter estimates were obtained by fitting the sample mean and observed proportion of zero eruptions. A chi-square goodness-of-fit test was performed to check the adequacy of the fit of the negative binomial distribution. The test statistic is 8.69 which leads to p-value $\doteq 0.07$, and indicates that the model is acceptable at 0.05 significance level. The estimated theoretical variance is 1.031 ($= \hat{r}\ (\hat{a} + 1)/\hat{a}^2$), which is also close to (relative to the theoretical value (=0.826) from a simple Poisson model) the sample variance $s^2 = 1.169$. Moreover, the posterior mean is estimated as 0.826, which coincides with the sample mean, $\bar{x}$. In other words, the updated belief about the future long run average eruptive rate, from the Bayesian point of view, after observing the sample x, is the sample average. The value 0.826 is based on five-year time interval. It is equivalent to 0.1652 ($year^{-1}$), or 1.38×10^{-2} ($month^{-1}$), or 3.77×10^{-5} (day^{-1}). Indeed, the eruptive rate which has been postulated as a random variable need not be observable in any sense. It is only necessary to suppose that in any one period of time eruptive rates behave as if they were a random sample from a gamma distribution.

CONCLUSIONS

Volcanic activity is governed by the complex interaction of several geological, geophysical and geochemical factors. Because of this complexity even with the present knowledge eruptions cannot

theoretically be predicted. Several circumstances make it difficult to work out more realistic statistical models yielding results in 100 percent quantitative agreement with the observed data. First of all, reliable historical records of volcanic activity are available for only a few volcanoes and, when this favorable circumstance occurs, the records are generally either short or inhomogeneous. In spite of these difficulties and limitations, the statistical model proposed in this work concludes the following relevant facts:

a) The whole analyses are strictly based on the eruptive count data, which are more reliable and consistent than measurements based on repose times, magmatic volume or other geologic variables. For the simple Poisson model, the chi-square goodness-of-fit test is often not reliable due to low degree of freedom or low expected cell counts. The Kolmogorov-Smirnov test or other survival curves, is then applied based on the repose times between eruptions, which is expected to be exponentially distributed. However, analyses based on repose times are limited to a description of its general pattern of activity, rather than the prediction of the outbreaks of a volcano. For the purpose of prediction, one has to convert the exponential distribution model to the Poisson model, which requires additional estimation of the eruptive rate based on count data. Often these two models produce inconsistent estimates, despite the fact that they should be identical theoretically.

b) The assumptions of the proposed NBD model are less restrictive and thus more applicable for a variety of situations. Moreover, this exceedingly flexible model has only two parameters which can be easily determined from the available data. The estimated posterior mean for the future eruptive rate is the observed sample average, which is intuitively acceptable and simple. Individual probability of x number of eruptions can be predicted based on the aggregate behavior of the volcanic activity. The applicability of the NBD model extends to cluster(s) of volcanoes and longer observation time. One does not need to trim the data and strive for the homogeneity of the processes. In several parts of the world, socio-economic conditions have resulted in the growth of large human settlements near active volcanic regions (e.g., Italy, Japan, U.S.A., Indonesia and Central America). Such is the high rate of industrial and agricultural development in these areas that it is becoming increasingly urgent to be able to evaluate the risk to which they are exposed. The record of global volcanism, however, is fragmentary and scattered. Many volcanoes were remote from human habitation, and the historic record, in view of the long intervals between many eruptions, is often too short to aid in predicting future volcanic activity. Predictions are needed, though, and there is always a chance that if we exclude any long repose time(s) (or interval(s) suspected as loss of data) we are discarding real information. We need solutions to this problem: protection against discarding good

data. The flexible model proposed in this paper would be helpful in this regard.

c) General sampling theory of the negative binomial distribution is well documented. The parameters estimation method adopted for the volcanic eruption modelling is not new. However, the most promising development for the NBD model, perhaps, is that the observed and theoretical counts for zero eruptions are exactly identical (except for some rounding errors). The probability distribution of volcanic count data decreases rapidly for high count values as in most of the cases, where P(0)(=P(no eruptive events)) sometimes is the largest frequency and plays an important role in decision making processes such as selection of a high-level nuclear waste repository site (e.g., Yucca Mountain, Nevada, U.S.A.). This model reflects 100 percent accuracy for fitting the theoretical value of P(0) to the observed data.

d) Finally, the estimated parameter $\hat{\alpha}$ in the NBD model could also serve as a diagnostic index for a simple Poisson process. A huge $\hat{\alpha}$ automatically sends a signal of a degenerated gamma prior, indicating a simple Poisson model for future volcanic forecasting. This is a valuable unique distinction of the generalized NBD model.

There are many unknown areas with respect to geologic understanding of volcanic activity. One possible improvement would be to reconfirm all of the crucial assumptions from statistical evidence using data, which are the only basis we have for making necessary plans, calculations and decisions. 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, where data are rare and expensive (e.g., dating of volcanism, cone counts). For the reasons discussed, the gain of Bayesian analysis, letting λ be random, is obtained with virtually no loss in predictive ability both in space and time.

ACKNOWLEDGEMENTS

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

APPENDIX

The NBD Model.

Treating the average eruptive rate as a random variable λ means that the probability distribution function $f(x;\lambda)$ is actually a conditional probability, the condition being that λ is in state λ. Thus, when using a probability distribution for λ, it is more appropriate to use the notation $f(x|\lambda)$ for the data X. From the conditional distribution of X and the given prior distribution for λ, we can calculate the joint distribution of (X,λ):

$$f(x,\lambda) = f(x|\lambda)g(\lambda),$$

and the corresponding marginal or absolute distribution of X, with probability

$$P(x) = E_g[f(x,\lambda)] = \int f(x|\lambda)g(\lambda)d\lambda. \qquad (A1)$$

For the case as described in the text, assuming that λ follows a

gamma distribution, GAM(r,a), then

$$g(\lambda) = \frac{a^r}{\Gamma(r)} \lambda^{r-1} e^{-a\lambda}, \ \lambda > 0; \ r,a > 0$$

and

$$f(x|\lambda) = e^{-\lambda}\lambda^x/x!, \ x = 0, 1, \ldots .$$

Thus, from equation (A1), the marginal probability distribution for the number of eruptions in unit time interval is given by

$$P(x) = \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!} \frac{a^r}{\Gamma(r)} \lambda^{r-1} e^{-a\lambda} d\lambda$$

$$= \frac{\Gamma\ (r + x)}{\Gamma\ (r)\ x!} \left[\frac{a}{a+1}\right]^r \left[\frac{1}{a+1}\right]^x, \ x = 0, 1, 2, \ldots,$$

as indicated in the text's expression (1). The interest in the marginal distribution centers around the fact that, if X has the conditional density $f(x|\lambda)$ and λ actually is random with density $g(\lambda)$, then $P(x)$ is the probability according to which X will actually occur. For this reason $P(x)$ is sometimes called the predictive distribution for X, since it describes what one would "predict" that X would be.

Posterior Mean $E[\lambda|x]$

Assume $X = (x_1, \ldots, x_n)$ is a sample from a Poisson distribution with mean λ, and assume $\lambda \sim$ GAM (r,a), then λ and X have joint distribution

$$f(x,\lambda) = f(x|\lambda)g(\lambda)$$

$$= \frac{e^{-n\lambda}\ \lambda^{n\bar{x}}}{\Pi_{i=1}^{n}[x_i!]} \cdot \frac{\lambda^{r-1}e^{-\alpha\lambda}\ \alpha^{r}I_{(0,\infty)}(\lambda)}{\Gamma(r)}$$

$$= \frac{e^{-\lambda(n+\alpha)}\ \lambda^{(n\bar{x}+r-1)}\ \alpha^{r}\ I_{(0,\infty)}(\lambda)}{\Gamma(r)\ \Pi_{i=1}^{n}[x_i!]}$$

The factor involving λ in this last expression is recognizable as belonging to a GAM $(n\bar{x} + r,\ n + \alpha)$ distribution. This must then be $g(\lambda|x)$. Since this posterior is a gamma distribution, it has mean

$$E[\lambda|x] = \frac{r + n\ \bar{x}}{\alpha + n},$$

as indicated in the text's expression (2).

TABLE 1

Years of flank eruptions of the volcano Etna for the period 1605-1980

1610	1766	1874	1928
1614	1780	1879	1942
1634	1792	1879	1947
1646	1792	1883	1949
1651	1802	1886	1950
1669	1809	1892	1971
1689	1811	1908	1971
1702	1819	1910	1974
1755	1832	1911	1974
1759	1843	1918	1978
1763	1852	1923	1978
1763	1865	1928	1978

TABLE 2

Years of eruptions of the volcano Mauna Loa for the period 1750-1979

1750	1870	1880	1926
1832	1871	1887	1933
1843	1872	1892	1935
1849	1873	1896	1940
1851	1873	1899	1942
1852	1875	1903	1949
1855	1875	1907	1950
1859	1876	1914	1975
1865	1877	1916	
1868	1880	1919	

TABLE 3

Summary of the fit of a negative binomial distribution

(volcano Mauna Loa)

NUMBER OF ERUPTIONS	FREQUENCIES	
	OBSERVED	THEORETICAL
0	22	22.0
1	15	14.6
2	7	6.3
3	0	2.2
4	1	0.2
5	1	0.2
6+	0	0.0
$\bar{x} = 0.826$, $P_0 = 0.478$, $\hat{a} = 4.023$, $\hat{r} = 3.322$		

REFERENCES

Anscombe, F.J., 1950. Sampling Theory of the Negative Binomial and Logarithmic Distributions. Biometrika, 37: 358-382.

Berger, J.O., 1988. Statistical Decision Theory and Bayesian Analysis (2nd ed.). Springer, New York, N.Y.

Condit, C.D., Crumpler, L.S., Aubele, J.C. and Elston W.E., 1989. Patterns of Volcanism along the Southern Margin of the Colorado Plateau: The Springerville Field. Journal of Geophysical Research, 94: 7975-7986.

Crowe, B.M., Johnson, M.E. and Beckman, R.J., 1982. Calculation of the Probability of Volcanic Disruption of a High-Level Radioactive Waste Repository within Southern Nevada, U.S.A. Radioactive Waste Management, 3: 167-190.

Davis, J.C., 1986. Statistics and Data Analysis in Geology (2nd ed.). Wiley, New York.

DeGroot, M.H., 1970. Optimal Statistical Decisions. McGraw-Hill, New York.

Feller, W., 1968. An Introduction to Probability Theory and Its Applications (3rd ed.). Wiley, New York.

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

Klein, F.W., 1984. Eruption Forecasting at Kilauea Volcano, Hawaii. Journal of Geophysical Research, 89: 3059-3073.

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

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

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

Tazieff, H. and Sabroux J.-C. (Editors), 1983. Forecasting Volcanic Events. Elsevier, New York.

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

Wickman, F.E., 1976. Markov models of repose-period patterns of volcanoes. In: D.F. Merriam (Editor), Random Processes in Geology. Springer, New York, N.Y., pp. 135-161.

TRANSFER OF COPYRIGHT

To be signed by the author or, in the case of multiple authorship, by at least one of the authors who agrees to inform the others, or, in the case of a work made for hire, by the employer. The signed statement must be received by the Editor before the manuscript can be accepted for publication.

Copyright to the article by Chih-Hsiang Ho

submitted under the title Nonhomogeneous Poisson Model for Volcanic Eruptions

is hereby transferred to the International Association for Mathematical Geology effective if and when the article is accepted for publication in MATHEMATICAL GEOLOGY

MATHEMATICAL GEOLOGY

For authors who are employees of the United States Government, the transfer of copyright is understood to be effective only to the extent that such copyright is transferable.

The authors explicitly reserve the following rights:

All proprietary rights other than copyright, such as patent rights.

The right to use all or part of this article in future works of their own, such as lectures, reviews, textbooks, or reprint books.

The right to make copies for the authors' own teaching use.

The right to use figures and tables in future publications, provided explicit acknowledgment is made of their initial appearance in this journal.

Signature	Signature
CHIH-HSIANG HO	
Name (print)	Name (print)
Title, if not Author	Title, if not Author
2/16/90	
Date	Date

NONHOMOGENEOUS POISSON MODEL FOR VOLCANIC ERUPTIONS

Chih-Hsiang Ho

Department of Mathematical Sciences
University of Nevada, Las Vegas, Nevada
4505 Maryland Parkway
Las Vegas, Nevada 89154 (U.S.A.)

ABSTRACT

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, $WEI(\theta,\beta)$, 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.

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 exactly 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 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 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 $X \sim POI(\mu)$ 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 a NHPP, then it can be shown that (Parzen, 1962, p. 138)

$$X(t) \sim POI(\mu(t)),$$

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

Veibull process, it must be remembered that the data represent repeated observations on the time to first occurrence of a Veibull process. As in the exponential case, the successive times of occurrences from a single Veibull process are of main interest, and some statistical results in this framework are discussed in the next section.

ANALYSIS FOR THE VEIBULL PROCESS

Bain (1978) discusses inference procedures of the Veibull process, and also give additional references. Suppose we assume that the successive volcanic eruptions of a specific volcano follow a single Veibull process. Let $t_1, \ldots, t_n$ 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

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

and

$$\hat{\theta} = \frac{t_n}{n^{1/\hat{\beta}}} . \tag{2}$$

2) A size α test of H_0: $\beta = \beta_0$ against H_A: $\beta \neq \beta_0$ is to reject H_0 if $2n\beta o/\hat{\beta} < \chi^2_{\alpha/2}(2n-2)$ 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) A r level prediction interval for t_{n+1} is $L \leq t_{n+1} \leq U$, where

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

and

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

First, the parameters estimated from Equations 1 and 2 provide us with a quantitative model to characterize the volcanic activity, which is the first step towards 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 trend: 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 the cumulative length of time (measured in years) over which the eruptions occur. Based on the above definition of repose times, this form of 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 0.05 significance level (p-value <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 non-random behavior, revealed as a lowered eruption rate after a repose of two years. A different result was obtained by Klein (1982), who used eruptions during the period 1918-1979. The result of a 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 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 HHP 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 Kolmogorov-Smirnov test was applied. In contrast, Equation 1 is sensitive to the locations, numbers, and relative sizes (to t_n) of the ordered t_i's. If an 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-run frequency interpretation of probability: To say that an event A has probability .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 include 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 VEI(θ, β), and then decide whether some subset of this family such as VEI(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 VEI(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.

ACKNOWLEDGEMENTS

I benefited considerably from discussions with geologists E.I. Smith, D. Feuerbach and T. Naumann of the Center of Volcanic and Tectonic Studies (CVTT) at UNLV. The 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.

Klein, F.W., 1982. Patterns of Historical Eruptions at Hawaii Volcanoes. Journal of Volcanology and Geothermal Research, 12: 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, 23: 263-272.

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

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

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

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

Wickman, F.E., 1976. Markov models of repose-period patterns of Volcanoes. In: D.F. Merriam (Editor), Random Processes in Geology. Springer, New York, N.Y., pp. 135-161.

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
$\begin{bmatrix} \hat{\beta} \\ \hat{\theta} \end{bmatrix}$	$\begin{bmatrix} 1.7931 \\ 17.3997 \end{bmatrix}$	$\begin{bmatrix} 1.432 \\ 24.996 \end{bmatrix}$	$\begin{bmatrix} 0.4927 \\ 1.7236 \end{bmatrix}$	$\begin{bmatrix} 2.2358 \\ 26.9405 \end{bmatrix}$	$\begin{bmatrix} 0.6369 \\ 0.4366 \end{bmatrix}$
Test statistic $2n/\hat{\beta}$ for H_0: $\beta=1$ vs. H_A: $\beta\neq1$ (p-value)	73.62 (0.000)	67.04 (0.032)	36.53 (0.005)	61.72 (0.000)	65.94 (0.012)
a 90% prediction interval for t_{n+1}	$\begin{bmatrix} 1980 \\ 1985 \end{bmatrix}$	$\begin{bmatrix} 1978 \\ 1996 \end{bmatrix}$	$\begin{bmatrix} 1982 \\ 2206 \end{bmatrix}$	$\begin{bmatrix} 1979 \\ 1983 \end{bmatrix}$	$\begin{bmatrix} 1962 \\ 1977 \end{bmatrix}$

ERUPTIVE PROBABILITY CALCULATION FOR THE YUCCA MOUNTAIN SITE, U.S.A.: STATISTICAL ESTIMATION OF RECURRENCE RATES (*)

Chih-Hsiang Ho[1], Eugene I. Smith[2], Daniel L. Feuerbach[2], Terry R. Naumann[2]

1. Department of Mathematical Sciences
University of Nevada, Las Vegas
Las Vegas, Nevada, 89154 (U.S.A.)

2. Center for Volcanic and Tectonic Studies
Department of Geoscience
University of Nevada, Las Vegas
Las Vegas, Nevada 89154 (U.S.A.)

- Offprint request to: C.-H. Ho
- (*) Paper financed by the Nuclear Waste Project Office, State of Nevada, U.S.A.
- Tel. (702) 739-3494, Telex (702) 6204-8164, FAX (702) 739-3850

ABSTRACT

Investigations are currently underway to evaluate the impact of potentially adverse conditions (e.g., volcanism, faulting, seismicity) on the waste-isolation capability of the proposed nuclear waste repository at Yucca Mountain, Nevada, U.S.A. This paper is the first in a series that will examine the probability of disruption of the Yucca Mountain site by volcanic eruption. In it, we discuss three estimating techniques for determining the recurrence rate of volcanic eruption (λ), an important parameter in the Poisson probability model. The first method is based on the number of events occurring over a certain observation period, the second is based on repose times, and the final is based on magma volume. All three require knowledge of the total number of eruptions in the Yucca Mountain area during the observation period (E). Following this discussion we then propose an estimate of E which takes into account the possibility of polygenetic (and polycyclic) volcanism at all the volcanic centers near the Yucca Mountain site.

INTRODUCTION

The Yucca Mountain region is located within the Great Basin portion of the Basin and Range physiographic province, a large area of the western United States characterized by alternating linear mountain ranges and alluvial valleys. Crowe and Perry (1989, Figure 1) divide the Cenozoic volcanism of the Yucca Mountain region into three episodes that include 1) an older episode of large volume basaltic volcanism (12 to 8.5 Ma) that coincides in time with the termination of silicic

volcanic activity; 2) the formation of five clusters of small volume basalt scoria cones and lava flows (9 to 6.5 Ma), all located north and east of the Yucca Mountain site; and 3) the formation of three clusters (seven volcanic centers) of small volume basalt centers (3.7 to .01 Ma), all located south and west of the Yucca Mountain site. The two youngest episodes form northwest-trending zones that parallel the trend of structures in the Spotted Range-Mine Mountain section of the Walker Lane belt. Crowe and Perry (1989) suggest a southwest migration of basaltic volcanism in the Yucca Mountain area based on this structural parallelism, a pattern that may reflect an earlier southwest migration of silicic volcanism in the Great Basin. Smith et al. (1990) also provide a different point of view of the migration trends of volcanism in the Yucca Mountain region.

Concern that future volcanism might disrupt the proposed Yucca Mountain repository site motivated the assessment of the volcanic risk to the Yucca Mountain area, located within the Nevada Test Site (NTS). Crowe and Carr (1980) calculate the probability of volcanic disruption of a repository at Yucca Mountain, Nevada using a method developed largely by Crowe (1980). Crowe et al. (1982) refine the volcanic probability calculations for the Yucca Mountain area using the following mathematical model:

$$\Pr[\text{ disruptive event before time } t\ ] = 1 - \exp^{-\lambda t p},$$

where λ is the recurrence rate of volcanic events and p is the probability of a repository disruption, given an event. The parameter p is estimated as a/A, where a is the area of the repository and A is some minimal area that encloses the repository and the area of the

volcanic events. Crowe et al. (1982) develop a computer program to find either the minimum area circle or minimum area ellipse (defined as A) that contains the volcanic centers of interest and the repository site. A is defined to accommodate tectonic controls for the localization of volcanic centers and to constrain λ to be uniform within the area of either the circle or ellipse. The rate of volcanic activity is calculated by determination of the annual rate of magma production for the NTS region and by cone counts using refined age data. Resulting probability values using the refined mathematical model are calculated for periods of 1 year and 100,000 years. Two procedures (explained below) are used for the rate calculations (Crowe et al., 1982). As calculated by Crowe et al. (1982), the annual probability of volcanic disruption of a waste repository located at Yucca Mountain falls in the range of 4.7×10^{-8} to 3.3×10^{-10}.

ISSUES THAT ARISE IN CONNECTION WITH THE WORK OF CROWE et al. (1982)

Although the procedure outlined in Crowe et al. (1982) represents a more formal approach to this problem than ever attempted previously, flaws exist. The method must be modified because the existing data base is inadequate to reasonably constrain λ. Despite the fact that there are well-recognized means of gathering data in the NTS region (field mapping; determinations of the eruptive history of basaltic centers; petrology; geochemistry; geochronology, including magnetic polarity determinations; tectonic setting; and geophysical studies) many considerations are still unknown, e.g., age of volcanism and vent counts.

Present understanding of eruptive mechanisms is not yet advanced enough to allow deterministic predictions of future activity. The only attempts at long-term forecasting have been made on statistical grounds, using historical records to examine eruption frequencies, types, patterns, risk, and probabilities. Reliable historical data make possible the construction of activity patterns for several volcanoes (Wickman, 1966, 1976; Klein, 1982, 1984; Mulargia et al., 1985; Condit et al., 1989; Ho, 1990ab). Unfortunately, detailed geologic mapping of volcanic centers is in its infancy in the Yucca Mountain area. A formal structure, with conclusions depending on the model assumptions, needs to be developed to evaluate volcanic risk for NTS.

This paper investigates important parameters required to calculate the probability of site disruption and provides estimates for the unknown parameter(s) that are meaningful both statistically and geologically, taking into account the limited availability of precise ages in the NTS region.

The Poisson Model

The application of statistical methods to volcanic eruptions is put onto a sound analytical footing by Wickman (1966, 1976) in a series of papers that discuss the applicability of the methods and the evaluation of recurrence rates for a number of volcanoes. Wickman observes that, for some volcanoes, the recurrence rates are independent of time. Volcanoes of this type are called "Simple Poissonian Volcanoes". Theoretically, the probability formula (Crowe et al., 1982) is derived for this type of volcanic activity from the following assumptions:

1. 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) λt, i.e., a simple Poissonian volcano.
2. Every eruption has the same probability of repository disruption p. That is, there is no heterogeneity with respect to disruptiveness.
3. The disruptiveness of the eruptions are independent of one another.

This very brief description is purely mathematical and has no direct interpretation in geologic terms. However, the Poisson model has proved successful in a wide range of situations and does have a number of mathematical implications that in turn can have a practical interpretation. Of course, exploring alternative models derived from different assumptions based on detailed geologic data and statistical analyses would be valuable, as well. Since a simple Poisson model serves as a basis for future refined model(s), we do not question the assumptions in this article. Therefore, the following statistical development uses the same assumptions as described above.

Probability Formula

The probability model of Crowe et al. (1982) is based on the following relationship:

$$\Pr[\text{disruptive event before time } t] = 1 - \exp^{-\lambda t p}.$$

The power series expansion for $\exp^{-\lambda t p}$ (Ellis and Gulick, 1986, p. 545)

is:

$$\exp^{-\lambda tp} = \sum_{k=0}^{\infty} \frac{(-\lambda tp)^k}{k!} = 1 - \lambda tp + \frac{(\lambda tp)^2}{2!} - \frac{(\lambda tp)^3}{3!} + \ldots$$

Therefore, the final probability calculation can be simplified as:

$$\Pr[\text{ disruptive event before time } t\ ]$$

$$= \lambda tp - \frac{(\lambda tp)^2}{2!} + \frac{(\lambda tp)^3}{3!} + \ldots$$

$$\doteq \lambda tp, \text{ for small } \lambda \text{ and } p \text{ relative to } t.$$

The approximation is reasonable and is true for virtually all of the calculations in Crowe et al. (1982) since all of their estimated values of both λ ($<10^{-5}$) and p ($<10^{-3}$) are small. Therefore, the accuracy of estimating the unknown parameters λ and p directly influences the significance of values for the probability of repository disruption.

Rates of Volcanic Activity

The Poisson processes are used to describe a wide variety of phenomena that share certain characteristics and phenomena in which some "happening" takes place sporadically over a period of time in a manner that is commonly thought of as "at random". We will refer to a "happening" as an event. If events in a Poisson process occur at a mean rate of λ per unit time (1 yr, 10^5 yr, etc.), then the expected number (long-run average) of occurrences in an interval of time in t units is λt. In quantifying volcanism at Yucca Mountain, we define a volcanic eruption as an event. Therefore, the collection of data is directly or indirectly based on the number of eruptions.

Crowe et al. (1982) try three methods to calculate λ in their probability calculations. These are: 1) evaluation of intervals of volcanic activity for evidence of periodicity, 2) counts of volcanic events in Quaternary time, and 3) evaluation of the ratio of magma production rate and mean magma volume. Based on method 1, they conclude that the data suggest no distinct patterns or periodicity of basaltic volcanism in late Cenozoic time. Therefore, the data are insufficient to analyze interval patterns and thus cannot be used to calculate future rates of volcanic activity (Crowe et al., 1982). We believe, however, that according to the Poisson model assumptions, intervals of volcanic activity should follow an exponential distribution and thus λ can be estimated statistically. We shall demonstrate such statistical sampling and estimation techniques in the following section.

Based on method 2, Crowe et al. (1982) calculate λ as N/T where N is the number of scoria cones and T is the period of time represented by the age of the cones or some other specified time period. Thus λ is the average number of eruptions per unit time. In their calculation of N/T, they define no statistical sampling technique that is associated with the assumed model. Moreover, they do not provide evidence that counting cones is equivalent to counting eruptive events. Crowe et al. (1989) and Wells et al. (1990) now recognize and classify the Lathrop Wells volcanic center as a polycyclic volcano and hence that some cones may have erupted more than once. Note that the Lathrop Wells volcanic center is only twelve miles away from the proposed Yucca Mountain repository site. We shall introduce a statistical estimation procedure to interpret the estimator.

Based on method 3, Crowe et al. (1982) determine the rate of magma production for the NTS region by fitting a linear regression line to a data set of four points collected from four volcanic centers. Each value thus represents magmatic volume of a single eruption at a corresponding volcanic center. The mean magma volume for 4 m.y. is calculated by taking the average of these four values. The ratio (rate/mean) is then calculated as an estimate for the annual recurrence rate λ. Similarly, the annual recurrence rate for Quaternary time is obtained using only the two Quaternary data points. We consider this approach questionable, since a simple Poisson model requires a constant rate of occurrence, which is not the same as magma production. We shall show that such calculations based on magma volume duplicate those of method 2, if the rate of magma volume is constant. Moreover, we shall also point out that, in this case, they apparently assume only four (two) eruptions in the NTS region during the last 4 m.y. (Quaternary time). This apparent assumption explains why their final probabilities based on magma volume are consistently smaller than those based on cone counts (Crowe et al., 1982, Tables IV and V).

The rate of volcanic eruption, λ, is a critical parameter for the probability calculation. We shall now examine various statistical methods for calculating λ, how the geologic record of volcanism in the Yucca Mountain can be used to estimate values of λ, and the limitations in calculating λ.

ESTIMATION BASED ON POISSON COUNT DATA

In dealing with distributions, repeating a random experiment several times to obtain information about the unknown parameter(s) is useful. The collection of resulting observations, denoted $x_1, x_2, \ldots, x_n$, is

a sample from the associated distribution. Often these observations are collected so that they are independent of each other. That is, one observation must not influence the others. If this type of independence exists, it follows that $x_1, x_2, \ldots, x_n$ are observations of a random sample of size n. The distribution from which the sample arises is the population. The observed sample values, $x_1, x_2, \ldots, x_n$, are used to determine information about the unknown population (or distribution).

Assuming that $x_1, x_2, \ldots, x_n$ represent a random sample from a Poisson population with parameter λ, the likelihood function is

$$L(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda) = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}.$$

Many good statistical procedures employ values for the population parameters that "best" explain the observed data. One meaning of "best" is to select the parameter values that maximize the likelihood function. This technique is called "maximum likelihood estimation," and the maximizing parameter values are called "maximum likelihood estimates," also denoted MLE or $\hat{\lambda}$. Note that any value of λ that maximizes $L(\lambda)$ will also maximize the log-likelihood, $\ln L(\lambda)$. Thus, for computational convenience, the alternate form of the maximum likelihood equation,

$$\frac{d}{d\lambda} \ln L(\lambda) = 0$$

will often be used, and the log-likelihood for a random sample from a Poisson distribution is:

$$\ln L(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \ln\lambda - \ln\left(\prod_{i=1}^{n} x_i!\right).$$

The maximum likelihood equation is:

$$\frac{d}{d\lambda} \ln L(\lambda) = -n + \sum_{i=1}^{n} \frac{x_i}{\lambda} = 0,$$

which has the solution $\hat{\lambda} = \sum_{i=1}^{n} \frac{x_i}{n} = \bar{x}$. This is indeed a maximum because the second derivative

$$\frac{d^2}{d\lambda^2} \ln L(\lambda) = -\sum_{i=1}^{n} \frac{x_i}{\lambda^2},$$

is negative when evaluated at $\bar{x}$.

Let us demonstrate this estimation technique. Let X denote the number of volcanic eruptions for a 10^5-year period for the NTS region from this assumed Poisson process. Then X follows a Poisson distribution with average recurrence rate μ , with $\mu = \lambda t = 10^5\lambda$. If we wish to estimate λ for the Quaternary using the Poisson count data for the NTS region, the successive number of eruptions from the last eighteen consecutive intervals of length 10^5 years ($18 \times 10^5 = 1.8 \times 10^6 \doteq$ Quaternary period) must be estimated. The number of observed eruptions per interval are denoted as $x_1, x_2, \ldots, x_{18}$. Thus, these eighteen values represent a sample of size eighteen from a Poisson random variable with average recurrence rate μ. Estimating the mean of the Poisson variable from these count data gives

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{18} x_i}{18},$$

and $$\hat{\lambda} = \frac{\hat{\mu}}{10^5} = \frac{\sum_{i=1}^{18} x_i}{1.8 \times 10^6}.$$

This shows that the estimated annual recurrence rate, $\hat{\lambda}$, is the average number of eruptions during the observation period. Based on this estimation technique, $\hat{\lambda}$ can be defined as

$$\hat{\lambda} = \frac{E}{T}, \tag{1}$$

where E = total number of eruptions during the observation period,

and T = observation period.

Note that for the estimation of λ in this model, an individual observation x_i is not required. However, the distribution of x_i's can provide information for model selection, for model-adequacy checking, and for parameter estimations in general.

ESTIMATION BASED ON REPOSE TIMES

With any Poisson process there is an associated sequence of continuous waiting times for successive occurrences. If events occur according to a Poisson process with parameter λ, then the waiting time until the first occurrence, T_1, follows an exponential distribution, $T_1 \sim \exp(\theta)$ with $\theta = 1/\lambda$. Furthermore, the waiting times between consecutive occurrences are independent exponential variables with the same mean time between occurrences, $1/\lambda$ (Parzen, 1962, p. 135). Several simplifying assumptions must be made in treating eruptions as events in time. Although the onset date of an eruption is generally well-defined as 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 used by Klein (1982). Therefore, we ignore eruption duration, we choose the onset date as the most physically meaningful parameter, and we measure repose times from one onset date to the next. Thus, our definition of "repose time" differs from the classic one as a noneruptive period. This procedure seems justified because most eruption durations are much shorter than typical repose intervals. The mean time between two events (eruptions), θ, is inversely related to the volcanic recurrence rate λ. Assumptions of the Poisson process are rather restrictive, but at least a very tractable and easily analyzed model can be proposed. The maximum likelihood estimator for $\theta = \bar{t}$ (Hogg and Tanis, 1988, p. 336) is:

$$\hat{\theta} = \bar{t} = \frac{\sum_{i=1}^{m} t_i}{m}, \text{ and } \hat{\lambda} = \frac{1}{\hat{\theta}} = \frac{1}{\bar{t}},$$

where $t_1, \ldots, t_m$ represent values of a random sample of size m from an exponential population with parameter θ. For the NTS region, the exact values of t_i's (repose times) are difficult to obtain because the precise date of each eruption is not known. However, based on the definition of repose times we can calculate:

$\sum_{i=1}^{m} t_i$ = time between the first and last eruptions during the observation period,

and

m = total number of repose times = E - 1, which gives

$$\hat{\lambda} = \frac{E - 1}{T_0 - T_y}, \tag{2}$$

where

E = Total number of eruptions between T_0 and T_y, inclusive,

with

T_0 = age of the oldest eruption,

T_y = age of the youngest eruption.

Note that the numerical values of E in both Equations 1 and 2 are identical for the same observation period of length T. In practice, however, the observation period for the exponential model (Equation 2) must be trimmed to a period between T_0 and T_y, inclusive, to reflect that exactly m (= E - 1) repose times (t_1 through t_m) are obtained. Theoretically, the two estimates obtained for λ (Equations 1 and 2) should be consistent, but not identical.

ESTIMATION BASED ON MAGMA VOLUME

Let V be the total volume of basaltic magma erupted at the surface in the NTS region during the observation period of length T. From Equation 1, we obtain

$$\hat{\lambda} = E/T = EV/TV = (V/T)/(V/E) = r/\bar{v} \tag{3}$$

where

$r = V/T$, the annual rate of magma production,

and

$\bar{v} = V/E$, the mean volume of magma during the observation period of length T.

Equation 3 is valid, but it also requires an accurate estimate of E for $\bar{v}$. If E (or r) is underestimated, so is λ. The most efficient way to

calculate r is V/T. Crowe and Perry (1989) present a refined method to calculate r. They evaluate r as the slope of the curve of cumulative magma volume plotted versus time. It is essentially identical to V/T, assuming a constant rate of magma volume (an assumption that Crowe et al. (1982) and Crowe and Perry (1989) have been striving to prove). In this case, the degree of erosional modification of volcanic landforms should be studied to estimate volumes of missing volcanic deposits. The overall error, which is multiplicative, is compounded in the values of Crowe and Perry (1989) for r. Moreover, E must be estimated when calculating $\bar{v}$, the mean volume of magma. Therefore, we see no economy in Equation 3 and consider it to duplicate Equation 1. We derive Equation 3 merely to demonstrate that the estimation procedures used by Crowe et al. (1982) and Crowe and Perry (1989) are flawed and therefore must be modified.

ESTIMATION OF E

All of the statistical estimation methods considered for λ (Equations 1-3) require knowing the value of E (total number of eruptions during the observation period). An accurate count of E is possible for volcanoes with a complete historical record. Identifying E, however, depends strictly on a clear understanding of eruptive processes and reliable dating techniques for the NTS region, since no historical record is available. Scientists differ in their opinions of volcanism at the NTS area. The following is the view of Crowe et al. (1983):

Basalt centers are composed of multiple vents, each marked by a scoria cone. In the NTS region the cones are divided into two categories: large central cones, referred to as the main cones, and satellite cones. The average number of cones at a single center, based on cone counts of seven Quaternary basalt centers in the NTS region, is about 2 to 3 cones. Thus, field data suggest a general eruption pattern where the initial breakthrough of magma to the surface is marked by the development of an eruptive fissure with two or three loci of effusion. Each of these vents becomes the site of small scoria cones. As the eruption proceeds, activity shifts or concentrates at a single vent that becomes the site of the main scoria cone.

The above description indicates that a main scoria cone is the final stage of a single eruption, and a single eruption could have several small vents to accompany the main cone. However, the possibility of polygenetic (and polycyclic) volcanism at all the volcanic centers needs to be evaluated. λ would be underestimated if nearby vents have distinguishable ages. We, therefore, estimate E as follows:

Let I denote the number of volcanic centers under investigation, and let J_i be the number of main cones in the ith volcanic center, where i=1, ..., I. The proposed estimate of E is:

$$\hat{E} = \sum_{i=1}^{I} \sum_{j=1}^{J_I} (m_{ij} + e_{ij}), \tag{4}$$

where m_{ij} = number of multiple, time-separate

eruptions of the jth main cone in the ith volcanic center,

and e_{ij} = number of vents that are separate in space and time (with distinguishable age measurements) from the jth main cone in the ith volcanic center.

The rationale for this estimate is that significant information has emerged that some of the volcanic centers are polycyclic volcanoes (e.g., Lathrop Wells center (Wells et al., 1990)). This estimation for parameter E (total number of eruptions) given by Equation 4 takes into account such a possibility for the NTS area. Studies are in progress to attempt to evaluate the values of m_{ij}'s and e_{ij}'s for the Quaternary volcanic centers of the Yucca Mountain region using K-Ar age determinations.

EMPIRICAL EXAMPLE

The following calculations demonstrate the applications of Equations 1 through 3 using the historical eruptions of Mt. Etna, since at this preliminary stage of our work data for the Yucca Mountain region are incomplete. A complete record exists for the eruptive activity of Etna between 1600 and the present. The time series of occurrence of flank eruptions in the period 1600-1980 follow a simple Poisson process (see Mulargia et al., 1985, and Ho, 1990ab). We provide this example to show consistency among the three estimation techniques for determining the recurrence rate.

Using the data for Etna, we take the year 1605 as a starting point

and then divide the period 1605-1980 into 25 fifteen-year intervals. Consistent with the notations used in the previous sections, Table 1 yields:

$n = 25$, $t = 15$, $T = 375$, $E = 48$, $V = 2830 \times 10^6 m^3$ (we treat the missing volumes in 1974 and 1983 as negligible).

Therefore, based on Equation 1,

$$\hat{\lambda} = E/T = 48/375 = 0.128.$$

Based on Equation 2,

$$\hat{\lambda} = (E-1)/(T_0-T_y) = (48-1)/(1978-1610) = 0.1277.$$

Based on Equation 3,

$$\hat{\lambda} = r/\bar{v} = (2830 \times 10^6/375)/(2830 \times 10^6/48) = 0.128.$$

As expected, two of the estimated rates are identical, while the third is also correct to two significant figures.

DISCUSSION AND CONCLUSION

The statistical estimation of eruptive rates λ requires a reliable count of distinguishable vents. This approach is based on the geologic record of volcanism at the NTS region. The methods of the approach are supported by sound statistical sampling theory. Crowe and Perry (1989), however, object that vent counts record only the recognition of a volcanic event, not its magnitude, and so they refine the parameter estimation by concentrating on the cumulative magma volume, which is a continuous variable. Nonetheless, their model assumptions and development are still based on a discrete simple Poisson model, which treats each eruption equally in order to calculate the final probability. Crowe and Perry (1989) also state:

Vent counts depend on the item scale of observation and may be insensitive to variations in rates of volcanic activity. Volcanic activity has been intermittent in the Yucca Mountain region during the Pliocene and Quaternary periods. Short periods of eruptive activity are bounded by long periods of inactivity. Because of this, differing observation periods for determining vent counts can lead to variations in rates of volcanic activity. For example, high rates of volcanic activity are obtained by narrowing the observation period to intervals of eruptive activity. Conversely, vanishingly small values of λ are obtained by restricting the observation period of intervals of inactive volcanism. The use of volcanic vent counts can therefore lead to an arbitrary biasing of calculated volcanic rates.

Their criticism is perfectly understandable based on the following facts:

a) Their recommended method for estimating λ is to construct a curve of cumulative magma volume versus time, which is also affected by the counts of vents ($\hat{E}$) in the observation period (T). Their ignorance of the critical factor E in Equation 3 leads them to believe that estimation based on magma volume is the most acceptable method (Crowe and Perry, 1989); this questionable belief, in turn, handicaps their estimates for $\bar{v}$ and thus for λ.

b) All of the published results that demonstrate statistical sampling techniques for volcanic activity require a representative sample and a sufficiently large sample size to calculate a reliable long-run average with precision at a desired level (flipping a coin twice does not tell the whole story of the fairness of the coin).

c) Their recognition of the fact that short periods of eruptive activity are bounded by long periods of inactivity at NTS indicates that their choice of a simple Poisson model should be adequately checked based on more detailed geologic data. So far, the problems of model assumptions and parameter estimations have been treated only separately by Crowe et al. (1982) and Crowe and Perry (1989), despite the fact that the model assumptions and parameter estimation methods virtually always depend on each other in volcanic hazard and risk calculations.

Yucca Mountain is remote from human habitation. There is no historical record of volcanism near Yucca Mountain. Therefore, the volcanic record must be developed by detailed field, geomorphic, and geochronologic studies. Precise ages are critical for volcanic rate calculations, but traditional K-Ar dating commonly has a large error in the age range recorded by the volcanoes near Yucca Mountain (1.1 Ma to 20 Ka). Until more precise techniques are developed, there will be uncertainties with regard to the age and duration of volcanism. Since predictions are needed, one possible improvement would be to reconfirm all of the crucial assumptions using data that are the only basis we

have for making necessary plans, calculations, and model selections. 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, where data are rare and expensive (≐ $300 - $600 per age of a vent at Yucca Mountain). Our efforts for future studies will be devoted to considerably more detailed data collection and statistical modelling. At this preliminary stage of our work, all we can conclude is that the probabilistic results of Crowe et al. (1982) are based on idealized model assumptions, a premature data base, and inadequate estimates of the required parameters. For the reasons discussed, we think that Crowe et al. underestimate the risk of volcanism at the proposed Yucca Mountain repository site.

ACKNOWLEDGEMENTS

We wish to thank Dr. Donald A. Swanson for his useful and thoughtful suggestions to this work.

REFERENCES

Condit CD, Crumpler LS, Aubele JC, Elston WE (1989) Patterns of volcanism along the southern margin of the Colorado plateau: the Springerville Field. J Geophys Res 94: 1975-1986

Crowe BM (1980) Disruptive event analysis: Volcanism and igneous intrusion. Battelle Pacific Northwest Laboratory Rep PNL - 2882

Crowe BM, Carr WJ (1980) Preliminary assessment of the risk of volcanism at a proposed nuclear waste repository in the Southern Great Basin. US Geol Survey Open-File Rep 80-375, 15p

Crowe BM, Johnson ME, Beckman RJ (1982) Calculation of the probability of volcanic disruption of a high-level radioactive waste repository within Southern Nevada, USA. Radioactive Waste Management 3: 167-190

Crowe BM, Self S, Vaniman D, Amos R, Perry F (1983) Aspects of potential magmatic disruption of a high-level radioactive waste repository in Southern Nevada. J Geol 91: 259-276

Crowe BM, Harrington C, Turrin B, Champion D, Wells S, Perry F, McFadden L, Renault C (1989) Volcanic hazard studies for the Yucca Mountain Project. Waste Management 89, Volume 1 - High Level Waste and General Interest, 485-491

Crowe BM, Perry FV (1989) Volcanic probability calculations for the Yucca Mountain site:' Estimation of volcanic rates. Focus '89, Nuclear waste isolation in the unsaturated zone, September, 1989, hosted by the Nevada section of the American Nuclear Society, pp. 326-334

Ellis R, Gulick D, (1986) Calculus (3rd Ed.). Harcourt Brace Jovanovich, New York.

Ho CH (in press) Nonhomogeneous Poisson model for volcanic Eruptions. Math Geol

Ho CH (in press) Bayesian analysis of volcanic eruptions. J Volcanol Geotherm Res

Hogg RV, Tanis EA (1988) Probability and statistical inference (3rd Ed.). Macmillan, New York.

Klein FW (1982) Patterns of historical eruptions at Hawaii volcanoes. J volcanol Geotherm Res 12: 1-35

Klein FW, (1984) Eruption forecasting at Kilauea volcano, Hawaii. J Geophys Res 89: 3059-3073

Mulargia F, Tinti S, Boschi E (1985) A statistical analysis of flank eruptions on Etna volcano. J Volcanol Geotherm Res 23: 263-272

Parzen E (1962) Stochastic processes. Holden-Day, San Francisco

Smith EI, Feuerbach DL, Naumann TR, Faulds JE (1990) The area of most recent volcansim near Yucca Mountain, Nevada: Implications for volcanic risk assessment. Proceedings of the International Topical Meeting, High-Level Radioactive Waste Management: American Nuclear Society/American Society of Civil Engineers, Vol. 1: pp.81-90

Wells SG, McFadden LD, Renault CE, Crowe BM (1990) Geomorphic assessment of late Quaternary volcanism in the Yucca Mountain area, southern Nevada: Implications for the proposed high-level radioactive waste repository. Geology, 18: 549-553

Wickman FE (1966) Repose-period patterns of volcanoes. Ark. Mineral. Geol 4: 291-367

Wickman FE (1976) Markov models of repose-period patterns of volcanoes. In: Merriam DF (Editor) Random processes in geology. Springer, New York, N.Y., pp. 135-161

TABLE 1

Date and volume of erupted lava for Etna flank eruptions in the period 1605 - 1980 (Source: Mulargia et al., 1985, Table 1)

Date	Erupted lava Volume ($10^6 m^3$)	Date	Erupted lava Volume ($10^6 m^3$)	Date	Erupted lava Volume ($10^6 m^3$)
1610	29	1802	7	1911	40
1614	424	1809	19	1918	1
1634	101	1811	23	1923	50
1646	184	1819	33	1928	8
1651	426	1832	51	1928	25
1669	435	1843	53	1942	2
1689	16	1852	118	1947	5
1702	11	1865	46	1949	4
1755	5	1874	1	1950	55
1759	67	1879	1	1971	3
1763	16	1879	40	1971	30
1763	28	1883	--	1974	--
1766	115	1886	59	1974	4
1780	16	1892	105	1978	27
1792	12	1908	3	1978	4
1792	79	1910	38	1978	11