Statistical Analysis of some Volcanologic Data Regarded as Series of Point Events

By R. A. REYMENT[1])

Summary – The case histories of some active volcanoes in various parts of the world are analyzed from the standpoint of their being observations of point events in a time continuum. The eruptive histories of the three Japanese volcanoes included show trend in the rate of occurrence of outbreaks. The possible existence of trend in rate of occurrence of events was found for certain Lower Cretaceous bentonites of Wyoming. The data investigated for Etna derive from a period of persistent activity and here also trend in the rate of occurrence of ejections could be identified. The remaining volcanoes studied do not display significant trend in the rate of occurrence of outbreaks over the time interval available. Various statistical tests indicate, that although some of the non-trend volcanoes may be fairly closely approximated as regards rate of occurrence of eruptions by the plausible Poisson model, none agree in all respects with the requirements of this process. The patterns of activity of volcanoes found to differ greatly from the Poisson model are complicated kinds of point processes, but owing to the shortness of the series available and their rather unsatisfactory accuracy, it is not possible to be explicit as to their precise nature. In order to elucidate some aspects of the analysis, a simulated series of outbreaks with exponentially distributed intervals between events was produced. The general scheme of analysis adopted has been firstly to test for trend; if trend in the rate of occurrence of events does not occur, the series have been tested for dependence. If there is no dependence between events, tests for agreement with a Poisson model have been carried out, with a negative conclusion leading to a test for agreement with some kind of renewal process. In order to provide a comparison with another type of natural phenomenon of a random nature, the earthquakes occurring in Fennoscandia over the period 1891 to 1950 were analyzed by the same methods. Perhaps surprisingly, the 322 shocks registered during this time (shocks $\geqq 3.0$ on the Gutenberg-Richter scale) show an indication of trend with a tendency for a decrease in the rate of occurrence of shocks. The eruption pattern of Mauna Loa is thought to be approximately a simple Poisson process. The patterns for Semeru, Bromo and Peak of Ternate seem to be reasonably consistent with a renewal process model, but appear to differ from a Poisson process. The Indonesian volcanoes have several features in common, among these a high coefficient of variation for the times between eruptions. It is tentatively suggested that this may be of some genetic significance. It is possible, that the Indonesian volcanoes erupt in accordance with a pattern approximating to some kind of stationary point process.

Introduction

Recently, WICKMAN [16–20][2]) has published statistical studies of the activity of some recent volcanoes. This material is important for the elucidation of the kind of problem represented by the bentonite data; for this reason, the most suitable case

---

[1]) Department of Historical Geology, Paleontologiska Institutet, Uppsala Universitet, Fack, Uppsala 1, Sweden.

[2]) Numbers in brackets refer to References, pages 76/77.

histories of volcanoes published in the Catalogue of the Active Volcanoes of the World were analyzed statistically to ascertain whether trends in the rate of occurrence of eruptions could be identified.

WICKMAN [16–20] was interested in determining whether stochastic models may be usefully applied to the description of volcanic activity; he concluded, that the repose patterns are amenable to this approach. The avenues of investigation adopted in this paper are largely different from WICKMAN's and we shall not be concerned with the development of models for the mechanics of eruption of the volcanoes studied, but rather with analyzing patterns of eruption. This approach is of an empirical nature, designed to represent the observable features of eruptive activity, but it does not include detailed technical knowledge about the presumed factors underlying the volcanologic processes. In connexion herewith, wide use is made of the results of LEWIS [12] and COX and LEWIS [4] with respect to the statistical analysis of series of point events.

Dr. PETER A. W. LEWIS, IBM Research Laboratories, Yorktown Heights, N.Y., U.S.A., has been of great advisory help and made a FORTRAN IV computer program available, prepared by himself and his associates, for carrying out many of the calculations involved in this study. This program was subsequently adapted for the Control Data 3600 at the University of Uppsala, and all computations have been checked on this machine as well. I am greatly indebted to Dr. LEWIS, Professor B. MATÉRN, Skogshögskolan, Stockholm, Professor D. R. COX, Birkbeck College, London, Professor MARKUS BÅTH and Dr. A. ÅGREN, Uppsala, and Professor FRANS E. WICKMAN, Pennsylvania State University, U.S.A., for careful criticism of the contents of the paper from which I have much benefited. I have also profited from discussions with Professor P. MORAN and Dr. D. VERE-JONES, Australian National University, Canberra. However, any faults and inaccuracies remaining are entirely my responsibility.

Origin of the data

The observations on the Lower Cretaceous bentonites were made by Dr. JOHN C. DAVIS while on the staff of the University of Wyoming. The data on Etna were obtained by WICKMAN and EL-HINNAWI [21] over two periods of observation. The observations on the three Japanese volcanoes, the Indonesian volcanoes and Mauna Loa were extracted from the published volumes of the Catalogue of the Active Volcanoes of the World, issued by the International Volcanological Association. The data for the artificial volcano were obtained by standard methods of computer simulation.

For the purposes of the present study, and the statistical methods employed, it was necessary to be stringent as regards the selection of data for the analyses; hence only a few of the case histories of the volcanoes listed in the above-mentioned catalog came near to being reasonably suitable and all are far from being without flaws (cf. WICKMAN [17], p. 136). Perhaps the most serious drawback is that due to the occurrence of periods of persistent activity; however, inexact records are a source of error.

General considerations

Many kinds of observations are known to occur randomly in time and one would indeed consider it strange if such were not so. Examples from everday life are: crashes of commercial aircraft, breakdowns in machinery, calls arriving at a telephone exchange outside of rush-hour periods, the appearance of mutations, and radioactive emission. Data of this kind may be studied in various ways. As regards volcanologic data, interest would tend to be directed towards analyzing the pattern of occurrence of outbreaks. The theory of 'point processes' has been developed in order to permit the analysis of individual occurrences of events, distinguished only by their positions in continuous, unidimensional time. It seems that volcanologic phenomena and earthquakes are best regarded as point processes as one may not always be basically interested in the number of eruptions or earthquakes that have taken place up to a particular time, but rather in the points in time at which the events happen.

It is clear that treating volcanic eruptions as a point event taking place at an instant of time is a vast oversimplification. However, data even less conformable to the point process model than that analyzed in this paper have proved to be amenable to the methods of statistical analysis here employed.

A reasonable model to select for the 'average volcano' is some form of point process, such as the Poisson Process. A generalization of the Poisson Process is the renewal process. The intervals between successive events in a renewal process are independently and identically distributed; if this distribution is of the exponential type, we have the special case of the Poisson process.

A few remarks as regards *notation*. The time-to-events will be denoted as T_i, and the times-between-events as X_i.

$$0 < T_1 < T_2 < T_3 < \cdots .$$

The letter T is used to denote a fixed length of time of observation on a series during which n events are observed.

The Poisson Process

In the present connexion, it will be taken, that an event may have occurred immediately prior to the point $T=0$, and that there is no chance that two or more events can occur simultaneously. If a series of observed events occurs at random, the Poisson model requires the assumption, that the chance of a new event taking place in any interval is independent of both the previous states of the system and the present state. The chance of a new eruption in an observed series of eruptions during the short interval ΔT is $\lambda\,\Delta T + 0(\Delta T)$, where λ is some constant characterizing the density of volcanic activity, and $0(\Delta T)$ is a remainder term. The probability distribution of a Poisson process $P_n(T)$ is:

$$P_n(T) = e^{-\lambda T}(\lambda\, T)^n/n! \quad (n = 0, 1, 2, \dots). \tag{1}$$

$P_n(T)$ is thus the probability that in an arbitrary interval of length T, n events will occur in a Poisson process with intensity λ.

The null hypothesis one would be testing for volcanoes is that the eruptions take place randomly at a constant rate (they are said to be homogeneous in time). The Poisson distribution has an additive property; this means, that if $A_1, \ldots, A_m$ are mutually independent and follow Poisson distributions with means $\mu_1, \ldots, \mu_m$ $(\mu = \lambda T)$, then the sum of the A's has a Poisson distribution with mean $\Sigma\, \mu_i$. This fact is utilized in the analysis of the Etna observations, where there are a few minor gaps.

If trends are not found to occur in the series of eruptions being studied, it may not be unreasonable to assume, that the series is stationary in the statistical sense, thence the marginal distributions of the times between events are identical. It is also interesting to ascertain whether significant serial correlation exists between the successive times between events. If the correlations found are not significant, it may be taken that the times between successive eruptions are independent and identically distributed.

The Wyoming bentonite data

As already noted, one series of observations made by DAVIS for a sequence of Lower Cretaceous bentonites in Wyoming gave an indication of trend in the rate of occurrence of events, thus suggesting the possibility of non-stationarity in the data. Thus, instead of the rate parameter for a Poisson process model for these data being constant over time, the following functional form may be used to describe the rate of occurrence of events:

$$\lambda(T) = e^{\alpha + \beta T}. \tag{2}$$

The test procedure used here to ascertain the existence of this trend in the data is that presented in COX and LEWIS ([4], p. 47) in the following terms:

$$U = \frac{\sum_{i=1}^{n} T_i/n - \frac{1}{2} T}{T \sqrt{(1/12\, n)}}, \tag{3}$$

where the series is observed for the interval $(0, T)$ and $(T_1, \ldots, T_n)$ are the times at which the events occur and n is the number of events. If n is known, and for $\beta = 0$, $\sum_{i=1}^{n} T_i/n$ is distributed as the sum of n independent rectangular random variables and the distribution of the standardized random variable (3) approaches the standardized normal form rapidly as n increases. The test based on (3) is an optimum test against the trend as represented by (2) (cf. COX and LEWIS [4], p. 47).

The Wyoming material gave for (3) the significant value of $U = 3.8$, which being positive, indicates that $\beta > 0$ in (2). One conclusion that might be ventured from this result is that some Lower Cretaceous volcano in the Wyoming area was, during the period represented by the bentonite deposits, possibly showing an increase in the rate of occurrence of eruptions. Another possibility is that the effect is spurious and, in fact, Dr. DAVIS has been able to produce a trend-free sequence of events by means of dip adjustments.

Volcanoes showing trend in the rate of occurrence of eruptions over time

The graphs of cumulative number of events against time for three volcanoes are shown in Figs. 1–3. These figures demonstrate that there may be pronounced trend in the data for the Japanese volcanoes Kirisima, Asama and Aso, particularly in the last section of each set of measurements for Asama and Kirisima, while the trending in the rate of occurrence of eruptions for Aso occurs over a longer interval of time.

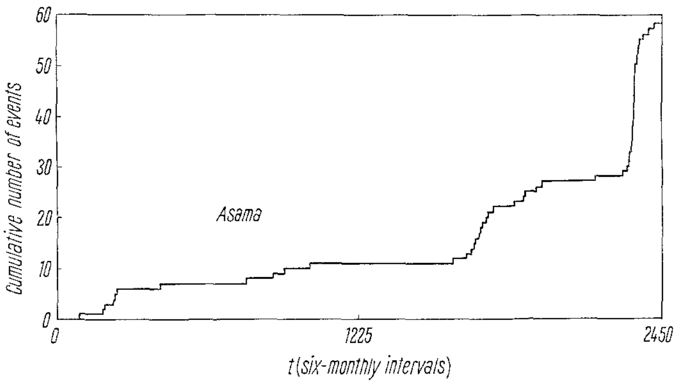

Figure 1
Asama, Japan. Cumulative plot of number of eruptions

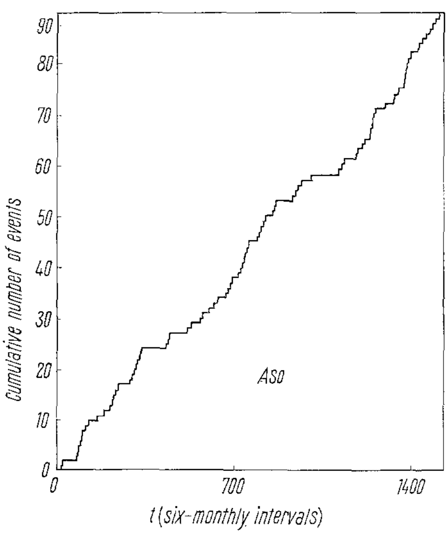

Figure 2
Aso, Japan. Cumulative plot of number of eruptions

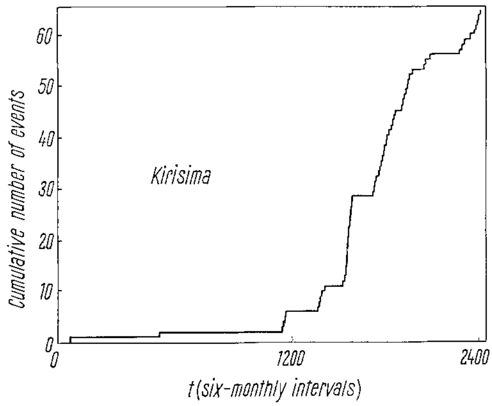

Figure 3
Kirisima, Japan. Cumulative plot of number of eruptions

The observations for Etna are particularly interesting. These were made by WICKMAN and EL-HINNAWI [21] during one of the periods of persistent activity of this volcano in April 1960 and constitute records of the ejections during two nights (here referred to as Etna I, respectively, Etna II). WICKMAN and EL-HINNAWI considered the pattern of ejection to conform with a Poisson process model (*op. cit.* p. 372). The results of the calculations (see Table 1) indicate a more likely model to be that of (2). The coefficient of variation for the intervals between events of Etna I is $V=0.74$ for 160 observations; for a Poisson process, $V=1$. The 145 observations for Etna II have a coefficient of variation of $V=0.76$, which does not differ significantly from Etna I. These two series of observations disclose the presence of some kind of trend in the rate of occurrence of the ejections over relatively short periods

Table 1
Tests for trend

Observations on	n	U	Time Units
Wyoming bentonite	29	3.80*	unspecified
Asama	64	9.42*	half year
Aso	108	3.43*	half year
Kirisima	59	6.74*	half year
Etna I	145	3.87*	seconds
Etna II	160	2.87*	seconds
Mauna Loa	36	−1.14	months
Semeru	53	0.30	half year
Bromo	32	−1.25	months
Peak of Ternate	40	−1.93	months
Simulated volcano	89	0.0036	unspecified
Fennoscandian earthquakes	322	−3.20*	days

The asterisks denote that the value is significant at at least the 5%-level.

during a phase of persistent activity. There is a basic difference in the nature of the data for Etna, deriving as they do from persistent activity, and the other volcanoes considered in this paper and it is not suggested that both types of activity showing trend in the rate of occurrence of events can be directly homologized.

The graphs of cumulative number of events against time for both Etna series do not show obvious trending and this is reflected in the values obtained for U.

For each of the volcanoes displaying trend, the bentonite sample, the volcanoes not showing trend and the simulated volcano, the number of events n, the time units, and U, the test statistic are given in Table 1. The result for Peak of Ternate is near the 5% level for trend and it cannot be ruled out, that longer series might have disclosed the presence of trend in the rate of occurrence of eruptions. The simulated volcano with exponentially distributed times between events gives a very low value of U. The data displaying trend in the rate of occurrence of outbreaks will not be considered further in this analysis.

The Japanese volcanoes and Etna yield positive values of U, indicating $\beta > 0$, and thus suggesting an increase in the rate of occurrence of events. The data on the Fennoscandian earthquakes were obtained from BÅTH [3]. The relatively high value of $U = -3.2$ indicates that there is a significant trend in the rate of occurrence of events in Fennoscandia over the period involved (magnitude greater than 3 on the Gutenberg-Richter scale). The negative value may indicate a trend towards a decreasing rate of occurrence of events.

It is clearly also important to look for cyclic trends in data of the kind treated here and the appropriate methods are given in COX and LEWIS ([4], p. 243). In the current study, the series are too short to allow a meaningful examination for such trends.

Volcanoes not showing significant trend in the rate of occurrence of eruptions

The following analysis rests on the assumption of stationarity. The sets of data are analysed, firstly, to see whether they accord with the properties of a renewal process; if a positive result is obtained, the data are tested to ascertain whether the Poisson model supplies a reasonably adequate description. If the Poisson model is clearly inadequate, it is sometimes possible to suggest some special renewal model, etc., although this will mostly require much longer series than are available in volcanology. The test for independence of successive intervals is based on the serial correlation coefficients. The plots of the cumulative numbers of events for Semeru, Bromo, Peak of Ternate and Mauna Loa do not give any indication of trend.

Tests for renewal processes

Rejection of the renewal hypothesis implies rejection of the hypothesis of a Poisson process of a series of events. Some estimates of quantities which characterize the second order, joint properties of the intervals between eruptions will now be reviewed. The serial coefficients of correlation may be represented as:

$$\left.\begin{aligned} \varrho_j &= \frac{E\{(X_i - E(X)), (X_{i+j} - E(X))\}}{\operatorname{Var}(X)} \\ &= \frac{\operatorname{Cov}(X_i, X_{i+j})}{\operatorname{Var}(X)} \quad (j = \cdots, -1, 0, 1, \ldots). \end{aligned}\right\} \tag{4}$$

This is for the intervals between events $(X_1, X_2, \ldots, X_n)$. In the above equations, $E(X)$ denotes the mean of the X's.

The serial correlation coefficients are the Fourier coefficients of the second quantity used in this analysis, notably, the *Spectral Density Function*, which may be expressed in the following form (Lewis [13], p. 213):

$$f_+(\omega) = \frac{1}{\pi} \{1 + 2 \sum_{j=1}^{\infty} \varrho_j \cos(j\,\omega)\} \quad 0 \leqq \omega \leqq \pi. \tag{5}$$

A renewal process will have a constant spectrum of intervals $f_+(\omega) = 1/\pi$. Where sample sizes are much less than 100, and the marginal distributions are highly skewed, the usefulness of the serial correlations is limited. Under suitable conditions, the ϱ_j may be considered under the null hypothesis $\varrho_j = 0; j = 1, 2, \ldots$ (see Cox and Lewis [4], p. 91, for further details).

For a renewal process, $\operatorname{Var}(\tilde{\varrho}_j) \sim 1/(n-j)$, where $\tilde{\varrho}_j$ is the estimate of the serial correlation. The test of the hypothesis $\varrho_j = 0$ is made by using $\tilde{\varrho}_j \sqrt{(n-j)}$, which has, approximately a unit normal distribution if n is large and a renewal process is assumed. Although this test using the serial correlation coefficient is useful and readily accessible, it does suffer from certain shortcomings (Cox and Lewis [4], p. 165). The first ten serial correlation coefficients for Mauna Loa, the three Indonesian volcanoes, and the simulated volcano are shown in Table 2. It should be understood, that the interpretation of higher order serial correlation coefficients in a situation such as the one here studied is fraught with difficulty and the significant values occurring in Table 2 could easily have arisen by pure chance, remembering the 5% confidence level being used although there does seem to be a tendency towards positive values. In view of the short sequences involved and the fact that the series of events for the simulated volcano also yields a significant second serial correlation coefficient, all volcanoes will be tentatively accepted as giving rise to series of eruptions which accord roughly with a renewal model.

Tests for Poisson Processes

The first three sample moments of the intervals between events are given in Table 3 for the volcanoes considered in this section. Particular significance attaches to the coefficient of variation, which offers a rough means of assessing departures from the exponential distribution. A simple Poisson process will have a coefficient of variation

Table 2

Estimated serial correlation coefficients

i	1	2	3	4	5	6	7	8	9	10
$\tilde{\varrho}_j$										
$\tilde{\varrho}_j \sqrt{(n-j)}$										
Mauna Loa	0.104	0.309	−0.023	0.301	0.042	0.364*	−0.023	0.104	−0.208	0.085
	0.617	1.802	−0.132	1.700	0.233	1.991	−0.122	0.549	−1.082	0.435
Semeru	0.002	−0.057	−0.015	0.028	−0.040	−0.028	0.024	0.058	0.016	−0.057
	0.017	−0.405	−0.103	0.196	−0.278	−0.194	0.166	0.389	0.103	−0.371
Bromo	0.067	−0.194	−0.188	0.128	−0.093	−0.312	−0.341	0.151	0.445*	0.304
	0.370	−1.061	−1.011	0.678	−0.483	−1.590	−1.707	0.740	2.136	1.424
Peak of Ternate	0.226	0.086	0.093	−0.106	−0.183	−0.196	−0.059	−0.114	0.040	0.055
	1.412	0.529	0.568	−0.635	−1.083	−1.144	−0.337	−0.647	0.221	0.302
Simulated volcano	0.062	0.230*	−0.129	−0.030	−0.132	−0.047	−0.135	0.051	−0.049	0.129
	0.580	2.150	−1.195	−0.274	−1.211	−0.424	−1.224	0.461	−0.440	1.147

Asterisks denote values significant at at least the 5% level.

for times between events of around unity (cf. the results for the simulated volcano in Table 3). Mauna Loa and the simulated volcano have coefficients of variation around one or less than one and therefore accord with what would be expected for a Poisson process. The three Indonesian volcanoes have higher coefficients of variation for times between events, although Bromo and Peak of Ternate lie around the upper limit of what might be acceptable for the sample sizes in question. Nevertheless, the fact that the three volcanoes with a high coefficient of variation all are located within the same geographic area may have a special significance.

Table 3

Sample moments for the intervals between events

	Mauna Loa	Bromo	Ternate	Semeru	Simulated Volcano
Mean	41.42	23.63	26.53	5.09	113.60
Variance	434.42	882.11	1110.05	85.63	11376.99
Standard Deviation	37.87	29.70	33.32	9.25	106.66
Coefficient of Variation	0.91	1.26	1.26	1.82	0.94
Measure of Skewness	1.47	1.67	2.26	5.81	1.53

Tests for Poisson processes

Some special distribution-free tests for Poisson processes have been made use of. These are, the one-sided Kolmogorov-Smirnov statistics, the two-sided Kolmogorov-Smirnov statistic, the Anderson-Darling statistic. If the series has been observed for a fixed time T and n events occur in $(0, T)$, the test considered here on the normalized times to events for the Poisson process, $y_i = T_i/T$ $(i=1, 2, \ldots, n)$, is called the uniform conditional test for a Poisson process.

$F_n(y)$ denotes the empirical distribution function of the observations y_i.

$$F_n(y) = \frac{\text{number of } y_i \leqq y}{n},$$

where $0 \leqq y \leqq 1$. The one-sided Kolmogorov-Smirnov statistic is defined as:

$$\left.\begin{aligned} D_n^+ &= \sqrt{n} \sup_{0 \leqq y \leqq 1} \{F_n(y) - y\} = \sqrt{n} \max_{1 \leqq i \leqq n} \left\{\frac{i}{n} - y_i\right\} \\ D_n^- &= \sqrt{n} \sup_{0 \leqq y \leqq 1} \{y - F_n(y)\} = \sqrt{n} \max_{1 \leqq i \leqq n} \left\{y_i - \frac{(i-1)}{n}\right\}. \end{aligned}\right\} \quad (6)$$

The two-sided Kolmogorov-Smirnov statistic is defined as:

$$D_n = \sqrt{n} \sup_{0 \leqq y \leqq 1} |F_n(y) - y| = \max\{D_n^+, D_n^-\} \quad (7)$$

The Anderson-Darling statistic is defined as:

$$\left.\begin{aligned} W_n^2 &= n \int_0^1 \left[\frac{F_n(y) - y}{\sqrt{\{y(1-y)\}}} \right]^2 dy\,, \\ &= -n - \frac{1}{n} \sum_{i=1}^{n} \{(2\,i-1) \log_e y_i + (2(n-i)+1) \log_e (1-y_i)\} \end{aligned}\right\} \quad (8)$$

The tests above in their original form have certain drawbacks (LEWIS [13], p. 210), but they may be considerably bettered with respect to power by means of a transformation. If the intervals between events are denoted by $X_1, X_2, \ldots, X_n$, and X_{n+1} is the interval between the last observed event and the end of the observation period, the intervals ordered by magnitude give the order statistics (LEWIS [13], p. 211):

$$0 < X_1 \leqq X_2 \leqq \cdots \leqq X_n \leqq X_{n+1}\,.$$

The quantities:

$$W_i = \frac{X_1}{T} + \frac{X_2}{T} + \cdots + \frac{(n+2-i)\,X_i}{T} \qquad (i = 1, \ldots, n)$$

under the null Poisson hypothesis have the same distributional properties as the y's previously considered.

The values obtained for the foregoing tests and for the transformed variables are shown in Table 4. An asterisk above the value of a test statistic in this table indicates that the Poisson hypothesis is rejected at the 0.05 level. The data on the Japanese volcanoes, Etna, and the Fennoscandian earthquakes have not been included, as the indication of trend already obtained constitutes rejection of a Poisson hypothesis. The significance decisions for the statistics D_n^-, D_n^+ and D_n are of necessity very approximate, owing to the small sample sizes involved.

Table 4

Tests for Poisson Processes

	Untransformed Data					Transformed Data			
Volcano	n	D_n^+	D_n^-	D_n	W_n^2	D_n^+	D_n^-	D_n	W_n^2
Mauna Loa	36		0.62	1.10	1.83	0.32		0.51	0.28
Semeru	53		0.92	1.44*	2.26*	1.95*		2.72*	9.78*
Bromo	32		0.29	0.85	1.26		0.31	1.79*	3.77*
Ternate	40		0.23	1.45*	2.59*		0.76	1.30	2.70*
Artificial volcano	89	0.61		0.77	0.72		0.46	0.53	0.34

An asterisk above the value of the test statistic indicates rejection of the Poisson hypothesis at a level less than 0.05.

Mauna Loa appears to fall in line with a Poisson hypothesis to a remarkable degree for both the transformed and untransformed data and these results are compatible with those obtained for the artificial volcano. Where the coefficient of variation for times between events is greater than unity, the transformed data will usually give higher values than the untransformed data, although this has not occurred for Peak of Ternate. Summing up, the results yielded by the tests employed in this section might indicate that only Mauna Loa erupts in accordance with a Poissonian type pattern. The other four volcanoes studied here could follow some other pattern within the group of renewal processes.

Logarithmic empirical survivor function

The graphs of the log empirical survivor function for four volcanoes and the simulated volcano are shown in Figs. 4–8. The graph of this function offers a rough

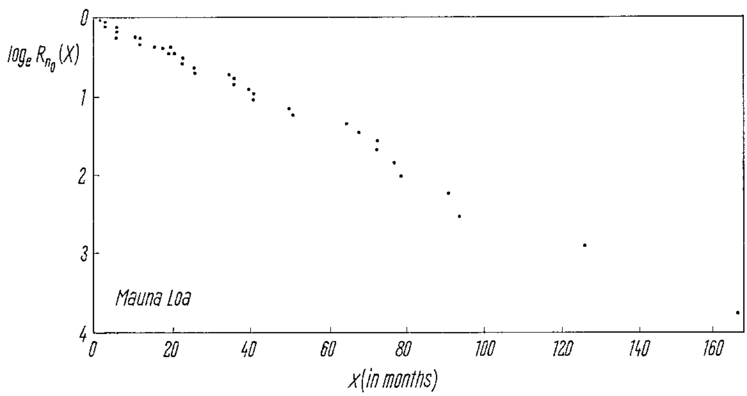

Figure 4
Mauna Loa. Log empirical survivor function for intervals between eruptions

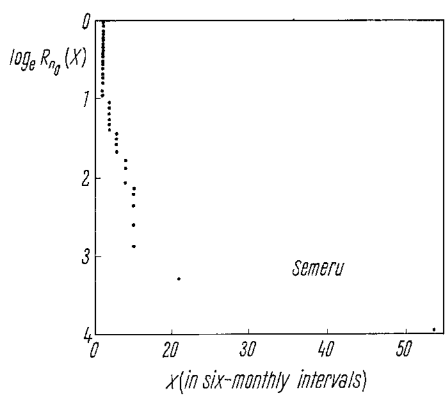

Figure 5
Semeru. Log empirical survivor function for intervals between eruptions

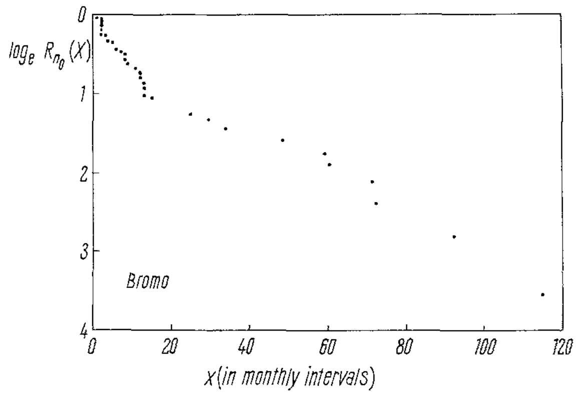

Figure 6

Bromo. Log empirical survivor function for intervals between eruptions

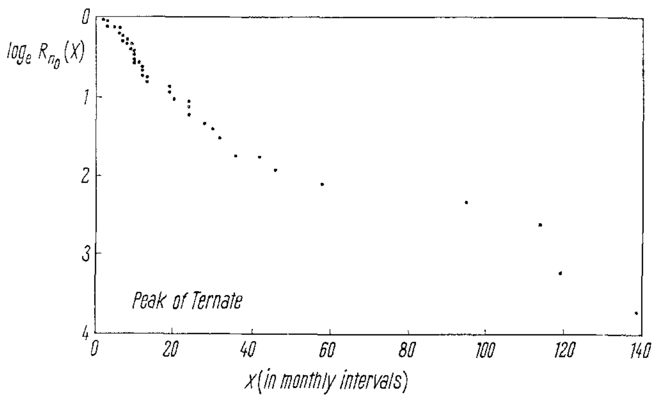

Figure 7

Peak of Ternate. Log empirical survivor function for intervals between eruptions

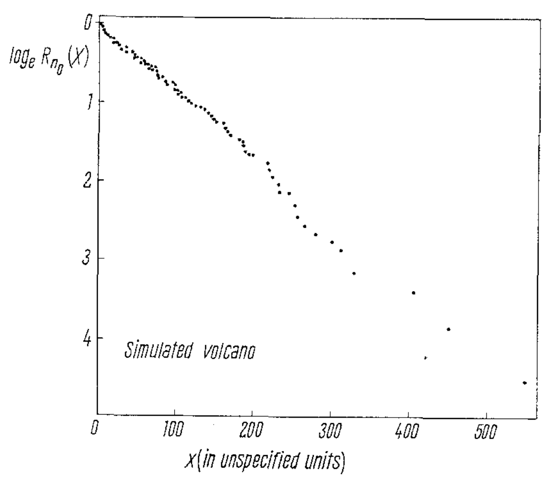

Figure 8

Simulated Volcano. Log empirical survivor function for intervals between simulated eruptions check on the feasability of a Poisson hypothesis for a series of events. If the tail of the plot has a constant slope, this suggests agreement of the times between happenings with a Poisson hypothesis. The n observed times between events X_i may be ordered to give the observed-order-statistics:

$$0 < X_1 \leqq X_2 \leqq \cdots \leqq X_n .$$

The empirical survivor function may be defined as

$$R_n(X) = \text{Proportion of intervals longer than } X ,$$

which is plotted against X. Here, n is the number of intervals in the sample. Where the distribution is more or less exponential in nature, the plot of the natural logarithm of the empirical survivor function, instead of the foregoing plot, is useful, as departures from linearity will indicate divergencies from the exponential distribution. The values used in the graphs in this paper, plotted against X_i, have been obtained from:

$$\log_e \left\{ 1 - \frac{i}{n+1} \right\} .$$

A monotone non-increasing hazard is suggestive of the logarithm of the empirical survivor function being concave (for a discussion of this terminology, see Cox and Lewis [4]).

The monotone non-decreasing hazard corresponds to a convex log survivor function; it is related to a coefficient of variation for the times between events of less than unity. The graph for Mauna Loa (Fig. 4) suggests the possibility of a linear tail for the points and consequently an approximately exponential distribution. The general shape of the plot is quite close to that for the points of the simulated volcano (Fig. 8). The graph of the log empirical survivor function for Semeru (Fig. 5) seems to differ considerably from the linear form for approximately exponentially distributed intervals and there is thus a reasonable possibility that the volcano differs from a Poisson type process; the coefficient of variation is also relatively high. The graph for Bromo has an overall form which differs from that of the simulated volcano, although the tail might be considered roughly linear. The coefficient of variation for the times between events for this volcano is greater than unity. The form displayed by the plot of the log empirical survivor function for Peak of Ternate (Fig. 7) is generally similar to that for Bromo (Fig. 6), but the tendency to form an upward concavity in the tail, only suggested by the data for Bromo, is more pronounced for Ternate and leaves some doubt about the linearity of the tail. It may be significant, that both of these volcanoes display a comparable pattern of eruption. The concavity may be interpreted as meaning that the longer the time from an eruption, the lesser is the risk for a new eruption. Owing to the small size of the samples, these interpretations must needs be very approximate.

Periodogram analysis

For testing the independence of the sequence of intervals $\{X_i\}$, it is useful to compute the unsmoothed estimate of the power spectral density, $\text{Var}(X) f_+(\omega)$. This is often referred to as the *Periodogram* when plotted. It is defined by the formula (cf. Cox and Lewis [4], p. 98):

$$I_p = \frac{A_n^2(\omega_p) + B_n^2(\omega_p)}{4\pi} \tag{9}$$

where,

$$A_n(\omega_p) = \sqrt{\left(\frac{2}{n}\right)} \sum_{s=1}^{n} X_s \cos(s\,\omega_p),$$

and,

$$B_n(\omega_p) = \sqrt{\left(\frac{2}{n}\right)} \sum_{s=1}^{n} X_s \sin(s\,\omega_p).$$

Here, ω_p is of the form $\omega_p = 2\pi p/n$ $(p = 1, 2, \ldots, n/2)$. I_p is known as the periodogram estimate for the power spectrum, when plotted against p or ω. The spectral transformation reduces the test for renewal processes to a test for a Poisson process. Deviating estimates of the power spectrum will show trends of some sort in their values rather than in the distributional form. The test for independence amounts to testing whether the I_p for successive values of p accord with a Poisson process. The results of the analysis for trend and independence of the periodogram calculations are shown in Table 5. The statistics used are those denoted by formulas (6), (7) and (8) (cf. Lewis [13], p. 214). The values for Mauna Loa, Semeru, Bromo and Peak of Ternate suggest support for the renewal hypothesis, although it is necessary to bear in mind, that all sample sizes are small for this kind of analysis.

There seems to be a reasonable amount of evidence for the independence of the sequences of intervals between eruptions for Mauna Loa, Semeru, Bromo and Peak of Ternate.

Table 5

Tests for trend and independence for I_p (the periodogram)

Volcano	U	D_n^+	D_n	W_n^2	Remarks
Mauna Loa	1.24	0.36	0.71	0.84	Does not reject renewal hypothesis
Semeru	0.56	0.29	0.35	0.23	Does not reject renewal hypothesis
Bromo	1.15	0.43	0.74	0.86	Does not reject renewal hypothesis
Peak of Ternate	1.76	0.10	1.01	1.16	Does not reject renewal hypothesis
Simulated volcano	0.77	0.76	1.04	1.55	Does not reject renewal hypothesis

N.B. The values of D_n^- and D_n coincide so that the former have been left out of the table.

Spectral density of the counting process

A useful second order property of the counting process N_t (cf. Cox and Lewis [4]) is the spectrum thereof, designated $g_+(\omega)$, the Fourier transform of a covariance density, which may be expressed in the following form:

$$g_+(\omega) = \frac{m}{\pi} + \frac{1}{\pi}\int_0^\infty \{e^{-i\omega\tau} + e^{i\omega\tau}\}\,\gamma_+(\tau)\,d\tau \quad \omega \geqq 0 \tag{10}$$

where $\gamma_+(\tau)$ is the covariance density. This function has certain relatively simple properties which make it useful in the analysis of the kind of statistical problem here considered. The estimate of (10) used in the computer program is (Lewis [13], p. 219):

$$\tilde{g}_+(\omega) = \frac{1}{\pi}\left(\frac{n}{T} + \frac{2}{T}\sum_{s=1}^{n-1}\sum_{j=1}^{n-s}\cos\{\omega(T_{(s+j)} - T_{(j)})\}\right) \tag{11}$$

where T is the length of the period and n is the number of eruptions. For a Poisson process with parameter λ,

$$g_+(\omega) = \frac{m}{\pi} = \frac{\lambda}{\pi} \quad \omega \geqq 0\,. \tag{12}$$

Here, m denotes the mean rate of occurrence of events. The function discussed here may be interpreted as a so-called *power spectrum*. The results in this paper are expressed in terms of the normalized spectrum, that is, as $\pi\, g_+(\omega)/m$. This has the value one for all ω for a Poisson process.

The plots of values for the Simulated Volcano, Mauna Loa, Semeru, Bromo and Peak of Ternate are shown in Figs. 9–13. The values for Semeru fluctuate around 1 and those for the Peak of Ternate around 1.5. The spectra for Mauna Loa (Fig. 10) and Semeru (Fig. 11) show a certain element of non-randomness. The former displays a surfeit of high values at about $(\omega\, T)/2\pi = 18$, 36, 54 and the latter has a perfect alternation of low and high values.

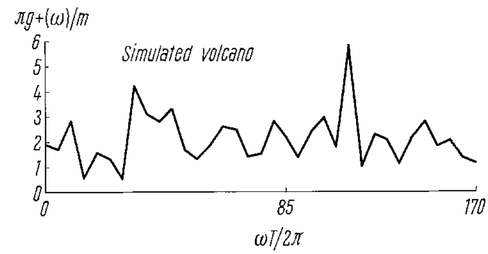

Figure 9
Simulated Volcano. Estimate of the normalized spectrum of counts

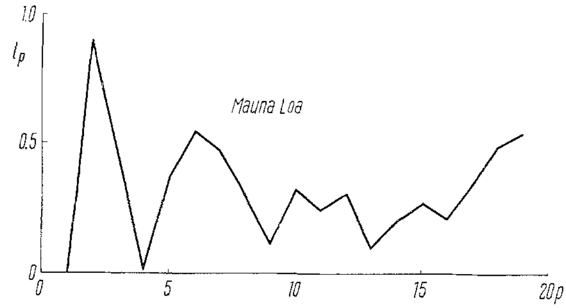

Figure 10
Mauna Loa. Estimate of the normalized spectrum of counts

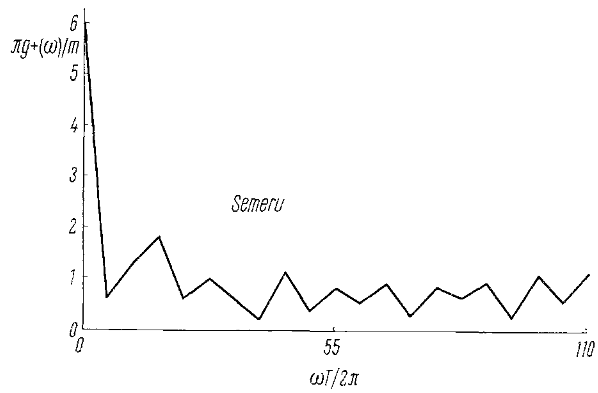

Figure 11
Semeru. Estimate of the normalized spectrum of counts

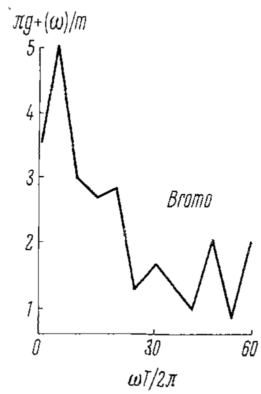

Figure 12
Bromo. Estimate of the normalized spectrum of counts

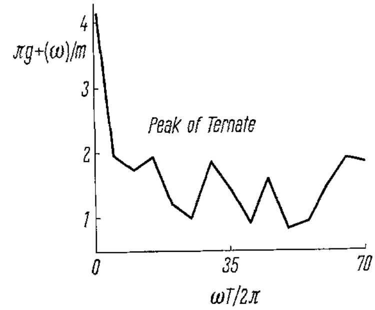

Figure 13
Peak of Ternate. Estimate of the normalized spectrum of counts

Summary for the Trend Volcanoes

The three Japanese volcanoes studied show significant trend in the rate of occurrence of eruptions. The data for Etna, which derive from a period of persistent activity, also display significant trend in the rate of occurrence of ejections. The 5% level of significance used to decide the level of significant trend may seem rather arbitrary, with Peak of Ternate barely excluded from the trend volcanoes, although the volcanoes showing significant trend lie well above this level of significance.

Thus, an important point arising from the results of this section is that Recent volcanoes do show trends in the rate of occurrence not only of eruptions, but also the rate of occurrence of ejections in connexion with persistent activity. Hence, the series represented by the data for the Wyoming bentonites is not necessarily an 'unreal' one in the sense that a volcano could not produce such a result.

Statistical methods have no magic powers and the level of accuracy of the results obtained can be no better than the data put into them. It is clear, that a change in the procedure of recording volcanologic activity could easily lead to the introduction of false trend into the records and, as already noted, none of the eruptive histories recorded in the International Catalogue of Active Volcanoes is particularly satisfactory.

It may therefore be reassuring to note, that the direct observations by WICKMAN and EL-HINNAWI [21] for Etna show trend as do also the earthquake data of BÅTH [3], to use an example based on another kind of crustal activity. As regards recording the times between events for volcanoes showing phases of persistent activity, it would seem, that the most satisfactory way of going about this is to take the time from the cessation of a particular period of persistent activity to the beginning of the next following eruption. Volcanoes for which the phases of persistent activity are relatively long in relation to the times between events are hardly amenable to the methods of analysis here considered.

The graph of the cumulative number of events is an invaluable aid in picking up trend, but it is inadequate if used on its own.

Summary for Mauna Loa

According to the information in Part III (p. 16) of the Catalog of the Active Volcanoes of the World, this volcano erupts almost exclusively by the extrusion of fluid basaltic magma. The test (3) did not disclose the presence of trend in the observations for this volcano, hence, the series could be treated as being a stationary one. The coefficient of variation is slightly less than unity, which is what could be expected for a Poisson process. Specific tests for Poisson processes do not show Mauna Loa to deviate from the Poisson hypothesis. The analysis of the periodogram did not suggest any deviations from the simple renewal model. The plot of the log empirical survivor function is similar in general shape to that for the Simulated Volcano, which has exponentially distributed times between events. The analysis using the spectral density function of the counting process also agrees with the Simulated Volcano. It would therefore seem a not entirely unreasonable assumption if a Poisson process model were applied as an approximation to the eruptive pattern of Mauna Loa for the period covered by the observations used in the present analysis.

Summary for Semeru

The eruptive activity displayed by this volcano is the extrusion of mainly ash and bombs, but also some lava, as reported in Part I, p. 142 of the Catalog of Active Volcanoes of the World. The coefficient of variation for Semeru is considerably greater than unity, which gives a rough suggestion of deviation from the Poisson type process. Specific tests for Poisson processes support this indication. The serial correlation coefficients are consistent with the theoretical value of zero required for the renewal process model and this is supported by the results of the periodogram analysis and the spectral density function. The graph of the log empirical survivor function deviates from the form of a monotone non-decreasing hazard. It is tentatively concluded, that the pattern of eruption displayed by this volcano approximates to some kind of renewal process. A more definite conclusion is hardly possible, owing to phases of persistent activity which are not fully accounted for in the existing record of eruptive activity for this volcano.

Summary for Bromo

Bromo has only been reported to extrude ash, lapilli and lava bombs and not lava, according to the Catalog of Active Volcanoes of the World (Pt. I, p. 146). The relatively high coefficient of variation found for the times between events gives a rough indication of deviation from a Poisson process; it is of interest to note that this has the same value as that for the Peak of Ternate. Specific tests for Poisson processes (accounted for in Table 4) reject this hypothesis. The tail of the log empirical survivor function for this volcano is roughly linear, but the form deviates from linearity. The periodogram analysis supports a renewal process model and it is suggested that the eruptive pattern of this volcano may represent some kind of renewal process.

Summary for Peak of Ternate

None of the serial correlation coefficients in Table II for this volcano differ significantly from zero, which may mean that the times between eruptions are reasonably consistent with a renewal process. The relatively high coefficient of variation suggests the series to deviate from the Poisson process model and specific tests reject the Poisson hypothesis. The shape of the graph of the log empirical survivor function is fairly close to that of Bromo, but it would be stretching the point if the tail were claimed to be linear. The analysis of the periodogram calculations supports a renewal hypothesis. Inadequately recorded persistent activity puts a more detailed analysis of the history of this volcano out of the question.

Final remarks

The methods of statistical analysis employed in this study have been largely concerned with the concept of a Poisson process. An important feature to be recognized concerns the identification of the existence of trend in the data. Methods of analysis appropriate for stationary series free of trend have been used and some of the methods of spectral analysis and serial correlation. Tests have been made for renewal processes in which the periods between successive events are independently and identically distributed. It was considered adequate for the present purposes to identify trend in the rate of occurrence; if an analysis of the nature of the trend were desired, approximate regression techniques could be applied with advantage.

It should be quite clear, that a volcanic eruption cannot be the outcome of a single factor and that several causes must interplay to bring about an outbreak. This makes it therefore natural to consider some sort of generalized renewal process as a likely near approximation. If a Poisson process does not fit the record of eruptive activity documented for a volcano, some other kind of renewal process might be considered, such as a simple renewal process, a simple semi-markov process or a superposition of point processes. The latter has a number of properties which make it attractive as a rough model for some kind of non-Poisson volcano.

References

[1] N. T. J. Baily, *The Elements of Stochastic Processes with Applications to the Natural Sciences* (Wiley and Sons, 1964).
[2] M. S. Bartlett, *Introduction to Stochastic Processes* (Cambridge University Press, 1955).
[3] M. Bath, *An Earthquake Catalogue for Fennoscandia for the Years 1891–1950*, Årsbok 50, S.G.U. [Ser. C] *545* (1956), 52 pp.
[4] D. R. Cox and P. A. W. Lewis, *The Statistical Analysis of Series of Events* (Methuens Monographs on Applied Probability and Statistics, London 1966).

[5] D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes* (Methuen, London 1966).
[6] H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, 1946).
[7] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I (1957) and Vol. II (1966) (Wiley and Son).
[8] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods* (Methuen, London 1964).
[9] E. J. Hannan, *Time Series Analysis* (Methuen, London 1960).
[10] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 3: *Design and Analysis and Time Series* (Griffin, London 1966).
[11] H. Kuno, *Japan, Taiwan and Marianas*, Part XI of the *International Volcanological Association Catalogue of the Active Volcanoes of the World* (Naples 1962).
[12] P. A. W. Lewis, *A Branching Poisson Model for the Analysis of Computer Failure Patterns*, J. Roy. Stat. Soc. [B] (1964), 398–456.
[13] P. A. W. Lewis, *A computer program for the statistical analysis of series of events*, IBM Systems Journal *4* (1967), 202–225.
[14] G. A. MacDonald, *Hawaian Islands*, Part III of the *International Volcanological Association Catalogue of the Active Volcanoes of the World* (Naples 1955).
[15] M. Neumann van Padang, *Indonesia*, Part I of the *International Volcanological Association Catalogue of the Active Volcanoes of the World* (Naples 1951).
[16] F. E. Wickman, *Repose Period Patterns of Volcanoes*, I: *Volcanic Eruptions Regarded as Random Phenomena*, Arkiv för Mineralogi och Geologi *4* (1966), 291–301.
[17] F. E. Wickman, *Repose Period Patterns of Volcanoes*, II: *Eruption Histories of some East Indian Volcanoes*, Ibid. *4* (1966), 303–317.
[18] F. E. Wickman, *Repose Period Patterns of Volcanoes*, III: *Eruption Histories of some Japanese Volcanoes*, Ibid. *4* (1966), 319–335.
[19] F. E. Wickman, *Repose Period Patterns of Volcanoes*, IV: *Eruption Histories of some Selected Volcanoes*, Ibid. *4* (1966), 337–350.
[20] F. E. Wickman, *Repose Period Patterns of Volcanoes*, V: *General Discussion and a Tentative Stochastic Model*, Ibid. *4* (1966), 351–367.
[21] W. E. Wickman and E. E. El-Hinnawi, *The Time Distribution of Lava Fragment Ejection from Volcanoes*, Ibid. *3* (1963), 363–383.

(Received 21st February 1968)