Probabilistic modelling for volcanic eruptions

Peter K. Anderson
Dulcie Kaivika

Abstract

A study of the mathematics involved in the modelling of volcanic eruptions is presented with a view to a future analysis of historical volcanic eruption data from the Bismarck Volcanic Arc to the north of PNG. The Poisson distribution is considered initially but strict criteria of IID and a fixed rate of events λ limit its applicability in the present context. The more realistic Weibull model is then considered which allows λ to vary as a function of time thereby allowing greater flexibility for modelling a wider range of volcanic types where eruption rates might not be constant over time. Finally, a demonstration is provided of the use of a Weibull plot method to graphically determine the parameters θ and β of the Weibull model for a given set of eruption data.

Keywords: Probability, distribution, models, deterministic, probabilistic, Poisson, Weibull, likelihood, MLE.

Introduction

Risk assessment and mitigation strategies of volcanic hazards are an ongoing need in PNG to avoid major economic loss, environmental damage and widespread mortality. Reliable assessment should lead to more accurate forecasting using deterministic or probabilistic modelling (Marzocchi & Bebbington, 2012, Varley et al. 2006).

The former involves accurate measuring of seismic events, such as ground-level elevation from GPS measurements and gas pressures in magma chambers. The assumption is that if enough data can be gathered to completely specify a deterministic model, the laws of physics can be applied, eruption events can be forecast for any future time (Varley, 2006) and evacuations planned. However, eruptive mechanisms of volcanoes are not sufficiently well understood to allow deterministic forecasting to be made with confidence (Jones et al., 1999).

Probabilistic modelling, which will be the focus of this paper and which can be used concurrently with deterministic modelling, recognizes the existence of random causative factors which can be estimated from historical eruption records provided these are complete, accurate and extensive (Damaschke et al., 2018). By analyzing historical data, hidden patterns can be discovered by a process pioneered by Wickman (1966a &b, cited in Marzocchi & Bebbington, 2012) using statistical modelling.

This paper will examine several statistical models which could be used to model volcanic eruptions using historical data and make probability estimates to date possible future eruptions. Mathematics for modelling of volcanic eruptions with the Poisson distribution of discrete random events will be considered initially. This will be shown to have limited application in the present context because of the strict requirements of Independent Identically Distributed (IID) data and a fixed rate of occurrence of events λ. It will be noted that a related continuous exponential distribution models inter-event times under the same assumptions. From here, it will also be noted that the exponential distribution is a special case of the Weibull distribution. This latter distribution incorporates a time variable t and allows λ (now = $\lambda(t)$) to vary as a function of time. Thus, the Weibull model will be shown to allow greater flexibility for modelling a wider range of volcanic types. Finally, a demonstration will be provided of the use of a Weibull plot to graphically determine the parameters θ and β of the Weibull model.

Statistical Models

A statistical model enables probabilistic forecasting using one of several well-known continuous or discrete probability distributions provided certain fixed points or parameters can be estimated from historical records of eruptions.

Probability Distributions depend on data from IID events. Certain fixed points called parameters need to be determined to allow predictions to be made with estimated probabilities. Here we consider the method of Maximum Likelihood Estimation (MLE) used to determine necessary distribution parameters.

Maximum Likelihood Estimate

Here we write: $P(X \mid p)$ as the probability of obtaining a given vector of data, X, under a particular distribution. However, we first need an estimate of parameter p which can be written in terms of a likelihood function: $L(p \mid X)$ which is the value of p which will make a set of historical data vector X most likely. Thus, we seek the value of p which maximises the likelihood L thereby estimating the parameter from the data (Law, 2013, p 330).

Given an observed set of discrete data, $x_1, x_2, .., x_n$, for example (continuous distributions can be treated analogously), and an unknown parameter θ, the Likelihood function becomes the joint probability function:

$$L(\theta) = P_\theta(x_1).\ P_\theta(x_2) \ldots.\ P_\theta(x_n) \qquad (1)$$

where P_θ is the probability mass function for this distribution.

Differentiating (1) to maximise $L(\theta)$ can be simplified by taking natural logs of both sides of (1) to obtain a sum of terms which can be more readily differentiated. Thus, we differentiate:

$$ln(L(\theta)) = ln(P_\theta(x_1)) + ln(P_\theta(x_2)) + \ldots + ln(P_\theta(x_n)), \quad (2)$$

set the derivative to 0 and solve for θ, as MLE estimate. $ln(L(\theta))$ will have the same maximum position as $L(\theta)$ since the log function is strictly increasing with $L(\theta)$ as $L(\theta)$ increases.

Volcanic eruption models

The important parameter for volcanic eruptions is λ, the rate of eruptions over time which is used in the Poisson distribution of random events, and related distributions. This rate may be a constant over time (λ) as required for a Poisson Distribution or it may be a function of time ($\lambda = \lambda(t)$). Timing can be based on absolute time of event occurrence (onset of eruption) or it can be inter-event, or repose times as will be further discussed.

Homogeneous Poisson Process

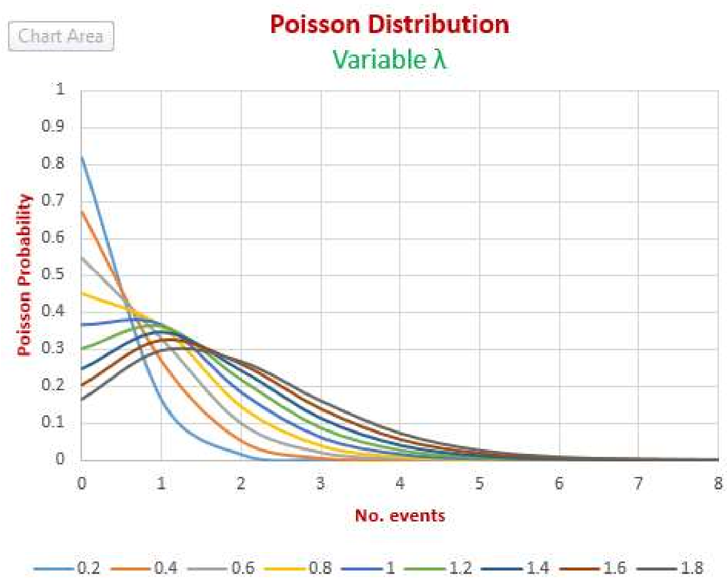

Figure 1: A series of Poisson curves with emphasis on small values of λ such as can be expected in this analysis. Initially, the number of time intervals with no events exceeds the number with one event until λ = 1 (dark blue line) when the situation reverses. The maximum turning point on each distribution moves slowly to the right.

We consider initially Homogeneous Poisson events where λ is constant over time with discrete probability mass function (Law, 2013, P312):

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \text{ x} = 0,1,2,.. \tag{3}$$

which gives the probability of x events (here eruptions, Figure 1), assumed a small number in the given time frame (say 1 year), where the parameter λ is estimated from sampled data using the MLE method as explained above and as demonstrated below (Jones et al., 1999, P35).

For data samples $x_1, x_2, \ldots\ldots, x_n$ of an assumed Poisson distribution, we consider the joint probability mass function (1) as the likelihood function:

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{xi}}{x!i} .$$

We maximize the corresponding natural logarithm function as

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

which has derivative

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

$=$ 0, for a turning point

giving estimated λ as the sample mean

$$\hat{\lambda} = \sum_{i=1}^{n}\frac{x_i}{n} = \bar{x} . \tag{4}$$

It can be further shown that the second derivative is negative implying that the turning point is a maximum. Thus, the constant Poisson parameter λ for homogeneous Poisson events can be estimated from the mean number of events occurring in a given time interval. This parameter can also be understood as characterizing the intensity of eruptions.

It can be noted that Eqn. (4) does not carry any information regarding the distribution of x_i*'s*, whereas this information is available and could potentially carry important information for model selection, checking the adequacy of model selection and parameter estimation (Jones et al., 1999, P35). To use this extra available information, the parameter λ for equation (5) can be estimated by maximizing the likelihood (joint probability) function:

$$L(\lambda; t_1, t_2 .. t_n)$$

where t_i is the time between events (see worked example, Jones et al P.36). These inter-event or reposal times are also used to model non-homogeneous Poisson events for which mean rates of events are defined with:

$$\lambda = \lambda(t), \text{ a function of time.}$$

As previously noted, a Poisson distribution of occurrences of eruptions implies an exponential distribution for inter-arrival or repose times t between eruptions (Ho, 91, Law, 2013, p313) as given by:

$$f(t) = \lambda e^{-\lambda t} \text{ for } t, \lambda > 0 \tag{5}$$

where $\lambda = 1/\theta$ and θ is the mean of inter-event times. Thus, we can also write:

$$f(t) = 1/\theta \; e^{-t/\theta} \tag{6}$$

which is a special case of the Weibull distribution:

$$f(t) = \beta\theta^{-\beta} t^{\beta-1} e^{-(t/\theta)^{\beta}} \tag{7}$$

when $\beta = 1$. We can then, by analogy with (6) write $\lambda(t)$ for the Weibull distribution as the coefficient of the exponential term:

$$\lambda(t) = (\beta/\theta)(t/\theta)^{\beta-1} \tag{8}$$

where the variable t can model change in the rate of eruption with time as will be discussed in the next section where the non-homogeneous Poisson distributions are introduced.

Non-homogeneous Poisson events

Here we consider the Weibull density function defined as

$$f(t) = \beta\theta^{-\beta}t^{\beta-1}e^{-t^{\beta}}$$

where θ is a scale parameter and *β* shape parameter. The introduction of the time variable *t,* typically a time to fail, but here time to an eruption, enables modelling where eruption rates vary with *t* and so $\lambda = \lambda(t)$ as will be demonstrated and as commonly observed.

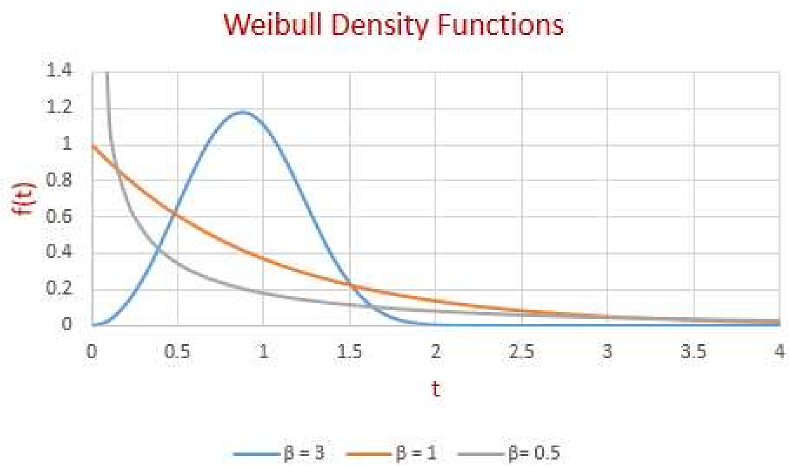

Figure 2: Weibull density function ($f(t) = \beta\theta^{-\beta}t^{\beta-1}e^{-t^{\beta}}$) is presented for $\theta = 1$ and three distinguishing values of *β.* For $\beta = 1$ and $f(t) = 1/e^t$ we have the special case of the exponential distribution which models inter-event times for a Poisson distribution for events with $\lambda(t) = \lambda$, and so constant with time. For $\beta > 1$ (here 3) we have now event rates decreasing more rapidly with time. For $\beta < 1$ (here 0.5) and event rates initially increase from zero to a maximum and then decrease as shown.

The Weibull function can be considered for 3 distinguishing values of the shape parameter each with $\theta = 1$ (Figure 2). For $\beta = 1$ and so $f(t) = 1/e^t$ we have the special case of the exponential distribution which models inter-event times for a Poisson distribution where $\lambda(t) = \lambda$ and where average event rates are constant over time. For $\beta > 1$ (here 3) we have $f(t) = 3t/e^{t^3}$ now decreasing more rapidly with time. For $\beta < 1$ (here 0.5) and f(t)=

$0.5 \big/ \left(e^{t^{0.5}} t^{0.5}\right)$, f(t) is now initially increasing from zero to a maximum and then decreasing as shown with the eventual overriding of the effect of the factor $t^{0.5}$.

Modelling λ (t)

Two approaches for modelling λ (t) have been used in the literature. The first (Ho, 1991) uses the Weibull Cumulative Distribution Function (CDF), (Law, 2013, p290) to model λ (t) with *β, θ* and t where t_i belongs to a set of repeated observations of time from the first occurrence. Thus for $t_1, t_2 \ldots t_n$, we require $t_1 \le t_2 \le \ldots t_n$.

For use with the Weibull distribution, we have already derived (8) λ(t) as:

$$\lambda\,(t) = \left(\beta \big/ \theta\right)\left(t \big/ \theta\right)^{\beta-1}$$

from which we can now derive:

$$\mu(t) = \left(t \big/ \theta\right)^{\beta} \qquad (9)$$

as the predicted numbers of events in time *t* by introducing the parameters β and θ as shown by the following integration:

$$\begin{aligned} \mu(t) &= \int_0^t \lambda(t)dt \\ &= \int_0^t \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1} dt \\ &= \frac{\beta}{\theta^{\beta}}\int_0^t t^{\beta-1}dt \\ &= \left[\frac{\beta}{\theta^{\beta}}\frac{t^{\beta}}{\beta}\right]_0^t = \frac{\beta}{\theta^{\beta}}\frac{t^{\beta}}{\beta} = \left(\frac{t}{\theta}\right)^{\beta} \end{aligned} \qquad (10)$$

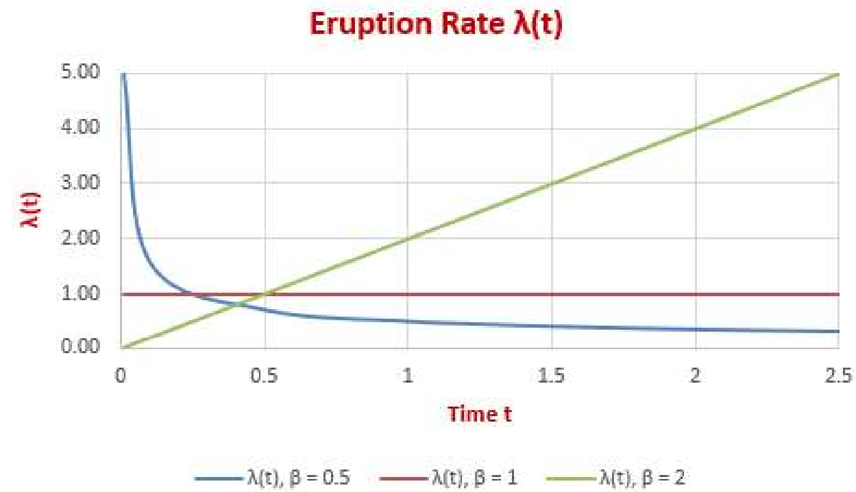

Figure 3: Eruption rates λ are a function of time with the Weibull model for θ = 1 and distinguishing values of θ. For β > 1, rates (λ(t)) increase with time, for β < 1, rates decrease with time and for β = 1, rate is constant. Thus, the Weibull distribution enables the modelling of varying rates.

If $\beta = 1$, $\lambda(t) = \lambda = 1/\theta$ and so is constant with time giving an exponential distribution for inter-event times which would also imply a Poisson distribution for events. For β > 1 we have the general Weibull model with inter-event times with $\lambda = \lambda(t)$ initially increasing with time. β < 1 gives a Weibull distribution for inter-event times which with $\lambda = \lambda(t)$ decreasing with time. Thus, by identifying θ and β for a given set of historical data it should be possible to predict future events in the historical series.

MLE Indicators of parameters

The MLE indicators for parameters θ and β can be shown (Ho, p170) to be given by:

$$\hat{\beta} = n \Big/ \sum_{i=1}^{n-1} \ln\left(t_n / t_i \right) \quad \text{and} \quad \hat{\theta} = t_n \Big/ n^{1/\beta} . \qquad (11)$$

Relation to Poisson Distribution

It can be noted here that this will simplify to the Homogeneous Poisson Distribution if $\beta = 1$ when $\lambda(t) = \lambda = 1/\theta$.

Modelling inter-event times

A second approach (Bebbington & Lai, 1996) considers a renewal process and uses a Weibull Distribution (again with parameters β and θ) to model inter-occurrence or reposal times which are presumed to be independent and identically distributed. This distribution is given by:

$$F(t) = P(T \leq t) = 1 - \ell^{(t/\theta)^{\beta}} \quad (12)$$

to model time to eruption or repose time (i.e. failure time or 1 – time to at least one event).
Here, however, we will consider:

$$R(t) = P(T > t) = 1 - F(t) = \ell^{-(t/\theta)^{\beta}} \quad (13)$$

to model survival time (R(t)) or time when no event occurs (i.e. 1 – failure time). A Weibull plot, explained as follows, can be used to estimate β and θ graphically. We write:

$$1/R(t) = \ell^{(t/\theta)^{\beta}} \text{ and so } \ln 1/R(t) = \beta \ln (t/\theta). \quad (14)$$

This equation can be linearized by taking natural logarithms of both sides again and giving:

$$\ln\left(\ln 1/R(t)\right) = \beta \ln t + \beta \ln \left(1/\theta\right). \quad (15)$$

A Weibull plot involves regressing $\ln\left(\ln 1/R(t)\right)$ against ln. The slope of the regression line will estimate β and the *y* intercept, β ln(1/θ), can be used to calculate θ using the estimated β.

Weibull Plot demonstrated

Weibull Analysis is typically used as a method of modelling failure data to measure long term performance of manufactured devices, but in the present context to measure time to eruption of a potential volcanic source. The data that can be measured is typically a time to failure, but for a Weibull plot, the time must be converted into a measure of unreliability which here will be *1 – MR* where MR is known as the Median Rank given by Bernard's

Approximation (http://reliawiki.org/index.php/Parameter_Estimation):

$$MR = (i - 0.3)/(N + 0.4) \qquad (16)$$

where i is the absolute rank order of failure and N is the total number of events (failures). The MR is a measure of the proportion (cumulative %) of failures (See Table 1), for each time of failure. The measure of unreliability required for plotting is then given by:

$$R(t) = 1 - MR \qquad (17)$$

and this will be plotted against time (Figure 4) to determine the parameters β and θ in (15).

Table 1: Portion of observed and calculated data for the Karkar Island volcano for a Weibull plot showing the use of formulas as described in the text for Weibull plot (16) and for the regression (15). Median Rank is a proportional rank order used as a measure of unreliability. Data Source:RVO.

	A	B	C	D	E	F	G
4						Regression	
5		Weibull Plot					
6		i-values	X-age to failure sorted	Median Rank		ln(ln(1/Rt)	ln(t)
7		1	0.67	=(B7-0.3)/(B$22+0.4)		=LN(LN(1/(1-D7)))	=LN(C7)
8		2	0.8	=(B8-0.3)/(B$22+0.4)		=LN(LN(1/(1-D8)))	=LN(C8)
9		3	0.8	=(B9-0.3)/(B$22+0.4)		=LN(LN(1/(1-D9)))	=LN(C9)
10		4	0.8	=(B10-0.3)/(B$22+0.4)		=LN(LN(1/(1-D10)))	=LN(C10)
18							
19		12	29	=(B19-0.3)/(B$22+0.4)		=LN(LN(1/(1-D19)))	=LN(C19)
20		13	55	=(B20-0.3)/(B$22+0.4)		=LN(LN(1/(1-D20)))	=LN(C20)
21		14	67	=(B21-0.3)/(B$22+0.4)		=LN(LN(1/(1-D21)))	=LN(C21)
22		15	187	=(B22-0.3)/(B$22+0.4)		=LN(LN(1/(1-D22)))	=LN(C22)

Table 2: Portion of observed and calculated data produced by the formulas shown in Table 1 for a Weibull plot as described in the text.

1	0.67	0.045455	-3.067872615	-0.400477567
2	0.8	0.11039	-2.145823454	-0.223143551
3	0.8	0.175325	-1.646280772	-0.223143551
4	0.8	0.24026	-1.29178935	-0.223143551
12	29	0.75974	0.354897648	3.36729583
13	55	0.824675	0.554526136	4.007333185
14	67	0.88961	0.79015558	4.204692619
15	187	0.954545	1.128508398	5.231108617

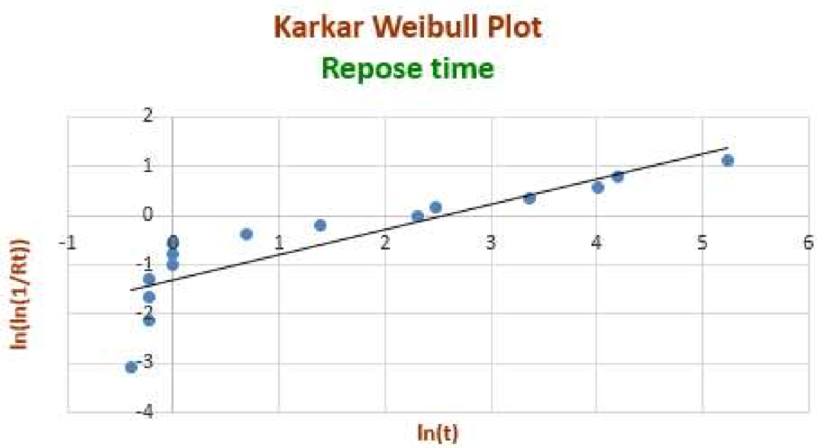

Figure 4: Weibull plot using sample data from the Karkar Island (Tables 1 and 2) volcano for purposes of illustration. The trend line models the linear relation (15) with the slope being used to determine β and the intercept to determine θ using β.

Summary & Conclusion

Mathematics for modelling of volcanic eruptions was presented starting with the Poisson distribution of random events. This has a limited application on the present context because of the strict required criteria of IID and a fixed rate of events λ. It was noted that an equivalent continuous exponential distribution models inter-event times under the same assumptions. However, it was also noted that the exponential distribution is a special case of the Weibull distribution. This distribution incorporates a time variable *t* and allows λ (= $\lambda(t)$) to vary as a function of time. Thus, the Weibull model, by incorporating the variable t allows greater flexibility for modelling a wider range of volcanic types. Finally, a demonstration was provided of the use of a Weibull plot to graphically determine the parameters θ and β of the Weibull model.

Glossary

CDF Cumulative Distribution Function (F(x))
PDF Probability Density Function (f(x)) for continuous distributions. Probability mass function is equivalent to a PDF for discrete distributions.
θ Weibull scale parameter theta (cf. β in Law, 2013)
β Weibull shape parameter beta (cf. α in Law, 2013)
λ Poisson Distribution parameter measuring event rate assumed constant over time
IID Independent Identically Distributed
MLE Maximum Likelihood Estimate
MR Median Rank
RVO Rabaul Volcanic Observatory
Poisson distribution: Discrete probability distribution used here to predict numbers of events in a given interval of time when random events are occurring at a constant rate, given by the parameter λ.
Exponential distribution: Continuous probability distribution with parameter θ (= 1/λ) used here to model inter-event times when events follow a Poisson distribution. This is a special case of the Weibull distribution when θ = 1.
Weibull distribution: Continuous distribution used here to model volcanic activity (time to failure of a potential volcanic source). Parameters are β (shape parameter) and θ (scale parameter). This can be understood as a generalisation of the exponential distribution when λ is allowed to vary as a function of time t (λ(t)).
Parameter notation: β and θ in the literature quoted here occur respectively as α and β in Law, 2013.

References

Bebbington, M.S., & Lai, C.D., (1996A). On nonhomogeneous models for volcanic eruptions. Mathematical Geology, Vol 28. No. 5. 585 – 600

Bebbington, M.S., & Lai, C.D., (1996B)., Statistical analysis of New Zealand volcanic occurrence data. Journal of Volcanology and Geothermal Research, 74: 101 – 110.

Cronin S.J., Bebbington, M., & Lai, Chin Diew, (2001). A probabilistic assessment of eruption recurrence on Taveuni Volcano, Fiji. *Bulletin of Volcanology,* 63: 274 – 288. DOI 10. 1007/s004450 100144.

Damaschke, M., Cronin, S. J., & Bebbington, M. S., (2018). A volcanic event forecasting model for multiple tephra records, demonstrated on Mt. Taranki, New Zealand. *Bulletin of Volcanology,* 80:9. https://doi.org/10.1007/s0045-017-1184-y

Davies, H. L., (2012). The geology of New Guinea – Cordilleran Margin of the Australian continent. *Episodes,* Vol. 35, No. 1

Dzierma, Y., & Wehrmawn, H., (2010). Eruption time series statistically examined: Probabilities of future eruptions at Villarrica and Lima volcanoes, Southern Volcanic Zone, Chile. *Journal of Volcanology and Geothermal Research,* 193: 82 – 92.

Ho Chih–Hsiang, (1991). Nonhomogeneous Poisson model for volcanic eruptions. *Mathematical Geology,* Vol 23, No.2: 167 – 173.

Ho Chih–Hsiang, (1996). Volcanic time – trend analysis. *Journal of Volcanology and Geothermal Research,* 74: 171 – 177.

Jones, G., Chestr, D.K., & Shooshtarian, F., (1999). Statistical analysis of the frequency of eruptions at Farnes Volcano, Sao Miguel, Azores. *Journal of Volcanology and Geothermal Research*, 92: 31 – 38.

Law, A.M., (2013). *Simulation Modelling and analysis*. (5th Ed.) VS: McGraw Hill.

Marzocchi, W., & Bebbington, M, S., (2012). Probabilistic eruption forecasting at short and long time scales. *Bulletin of Volcanology,* 74: 1777 - 1805. DOI 10.1007/s00445 – 012 – 0633 – x.

Turner, M.B., Cronin, S.J., & Bebbington, M.S., (2008). Developing Probabilistic eruption forecasts for dormant volcanoes: A case study from Mt Tarasabi New Zealand. *Bulletin of Volcanology,* 70: 507 – 515. DOI 10. 1007/s00445 – 007 – 0151 – 4.

Varley, N., Johnson, J., Reiz, M., & Marlen, K. (2006). Applying statistical analysis to understanding the dynamics of volcanic explosions. DOI: https//dot.org/ 10. 1144/ AVCE 1001. 6.

Weibull Plot. htps://www.qualitydigest.com/magazine/1999/ jan/article/using-microsoft-excel-weibull-analysis.html

Wickman, F.E., (1996a). Repose – period patterns of volcanoes. I. Volcanic eruptions regarded as random phenomena. *Arch. Miner. Geol.,* 4: 291 – 367.

Wickman, F.E., (1996b). Repose – period patterns of volcanoes. V. General discussion and a tentative stochastic model. *Arch. Miner. Geol.,* 4: 351 – 367. Cited in Marzocchi & Bebbington 2012.

Suppose $\ln(L(p))$ is maximal for $p=p*$, but it does *not* maximize $L(p)$. Then, there exists $p{\sim}\neq p*$ such that $L(p{\sim})>L(p*)$. Because the logarithm is increasing, this also means that $\ln(L(p{\sim}))>\ln(L(p*))$. This is a contradiction with $\ln(L(p*))$ being maximal.

Benard's Approximation for Median Ranks

Another quick, and less accurate, approximation of the median ranks is also given by:

$$MR = \frac{j - 0.3}{N + 0.4}$$

This approximation of the median ranks is also known as *Benard's approximation*.

http://reliawiki.org/index.php/Parameter_Estimation

Acknowledgements

This paper originated in a final year student research project conducted by Dulcie which modelled volcanic eruptions in the Bismarck Volcanic arc to the North of PNG using the Poisson distribution. Limitations of a strictly constant rate of eruption events led to a wider investigation of possible mathematical models allowing eruption rates to vary with time as has been recorded in this paper. The authors wish to thank Mr Cyril Sarsorsuo for reviewing this paper.

Authors

Peter K. Anderson PhD
Head, Department of
Information Systems
Divine Word University
Email: panderson@dwu.ac.pg

Ms Dulcie Kaivika
Data Officer
Divine Word University
Email: dkaivika@dwu.ac.pg