The Philippine Statistician
1992, Vol. 41,
Pages 35-45

A STATISTICAL MODEL FOR RISK ASSESSMENT ON PHILIPPINE VOLCANIC HAZARD

by

Erniel B. Barrios
The Statistical Center
University of the Philippines Diliman

Abstract

Statistical Modelling techniques were used intensively to develop hazard probability models in the assessment of vulnerability of an area, and in valuation of elements at risk. The resulting models can be used for the assessment of the economic loss in an area due to an eventual eruption.

Introduction

Natural disasters can greatly affect the economy and can even result in loss of lives. The devastating impact of a volcanic eruption is not an exception. The damage brought about by an eruption is proportional to the extent of an activity. Lava flow for instance can result in total wreck.It is true that man cannot control natural phenomena like a volcanic eruption. The loss however, can be minimized by inculcating preparedness to the residents/concerned people in the hazard zone of the volcano. In this light, the project aims to determine the risks posed by volcanic hazards in an active volcano area. Due to absence of eruption data for other volcanoes, the study concentrated on Mayon volcano. The number of eruption with enough records is quite reasonable.

The following terminology will simplify comprehension of this paper.

Risk - the amount of losses that may result from the probable occurence of a destructive agent or hazard. It is determined by three elements or factors; the hazard, vulnerability to the hazard, and the value of the element at risk.

Hazard - a potentially destructive agent, process, or event whose direct interaction with the material environment could cause harm on man and his resource. (Note: this definition of hazard leads to the terminology of hazard probability as the probability of occurence of the hazard at a specific period of time.)

Vulnerability - the likelihood that an area will be hit by a particular hazard. It is a function of: 1) the characteristics and magnitude of the hazard; 2) site specific conditions/characteristics such as local and regional geologic conditions, distance form the hazard source. location with respect to major structural features, topography, slope, drainage, etc.; and 3) the presence of protective/mitigation and preparedness measures.

Elements at risk - the population that may be injured or killed, natural resources, buildings and civil engineering works, public services, utilities and infrastructures that may be damaged or destroyed, the economic activities that may be disrupted, and other elements that may adversely affected by a potentially harmful phenomenon.

1. Elements at risk

Baseline data on the elements at risk were generated from a survey. This survey provided information at the barangay level for various hazard zones. In assessing risk brought about by future eruptions, yield/count/value of elements at risk at that time will be needed. Thus forecasting models were constructed.

Identification

In the fourth quater of 1989, two-stage stratified sampling was used in identifying the elements at risk. The primary stage units (psu) were the barangays, and the secondary stage units (ssu) were the households. The population consists of 97 barangays which were stratified according to a composite degree of hazard (for the four hazards considered - ash, lava, lahar and pyroclastic). The barangays were classified as low, moderate, or high hazard zones. The sample psu's (barangays) were chosen using probability proportional to size sampling with number of households as size measure. Then, post-stratification according to various levels of hazard per type of hazard was done. SSU's were selected using systematic sampling.

People, properties, infrastructures, and crops and animals complete the composition of elements at risk. Among the dominant crops are palay, corn, tomato, cassava and camote. For the animals we have cattle, carabao, hog, chicken and duck.

Forecasting Models

The data collected in the survey reveals only the distribution of elements at risk to various hazard zones. This is actually the main reason why the survey was conducted. The data collected by the National Statistics Office (NSO) and the Bureau of Agricultural Statistics (BAS) do not provide information at this level. For long term planning, the economic values of elements at risk has to be predicted.

For long term forecasting, structural models are preferred over nonstructural models. In this case, exogenous variables are used to explain the behavior of the variables under study. For crops, area planted is the most stable exogenous variable. Hence, simple linear regression models of yield on area planted were constructed for the major crops listed in Section 1.1. The problem with these models is that the disaggregated data is not available. Palay and corn data are released at regional level only, for other crops, they are released at the provincial level. The areas that can be potentially afftected by eruption do not cover the entire province of Albay and hence, data disaggregation is a must.

Data disaggregation is an area which is not yet fully explored, thus literature is very limited. The procedure that will be adopted in this study has many limitations which will be pointed out later on.

Given the model

$$y = a + bx + e.$$

If each of the variables y and x are multiplied by a constant c, the model becomes

$$cy = ca + bcx + ce \quad \text{or}$$
$$y^* = a^* + bx^* + e^*.$$

Thus, the parameter estimates for the diaggregated level can be obtained by adjusting the estimate of the intercept from the aggregated data. The drawback of this approach is that it has implied the assumption that the proportion of the disaggregated level relative to the total is always the same over time. However, this can be adjusted as soon as the data becomes available.

For Palay:

The given data are totals for the whole Bicol region. Least squares estimation results

Yield = 70265.1622 + 1.6425 Area (R-sqr = 87.22%)

The yield is expressed in metric tons while area is in hectares. Based on the 1981 Census of Agriculture, Albay shares approximately 20% of the palay area from the total of Bicol region. Furthermore, approximately 50% of the total farm area in Albay can be affected by Mayon's eruption. Taking these figures into account, we can use the model

Yield = 7026.5162 + 1.6245 Area

in forecasting yield of palay in affected areas of Albay given the area planted to the crop. To use this forecasting in smaller domains, the intercept should be adjusted further.

For Corn:

The data on the toals for the whole Bicol region yields the model

Yield = -19620 + 0.9176 Area (R-sqr = 92.73%).

The share of corn area to the whole of Bicol region is approximately 15%. Adjusting the intercept using similar procedure as in the case of palay,

Yield = - 1471.5 + 0.9176 Area.

The model above can also be used in forecasting yield of corn for the total area which can be affected by the eruption.

For Tomato:

The data for Albay results

Yield = -28967630.37 + 19575.99 Area (R-sqr = 91.94%)

and after an adjustment on the intercept

Yield = -14483815.19 + 19575.99 Area.

For Cassava:

The model given the data for Albay is

Yield = 44752.67 + 8759.703 Area (R-sqr = 48.81%)

and adjusting the intercept

Yield = 22376.335 + 8759.703 Area.

For Camote:

The model is

Yield = -21199983.83 + 13278.89 Area (R-sqr = 96.1%)

and the adjusted model is

$$\text{Yield} = -10599991.92 + 13278.89 \text{ Area}.$$

For poultry and livestock head count, smoothing was used. Single exponential smoothing was used for each of the five animals. The data given were all totals for the whole of Albay. The individual models are:

For Cattle:

$$\text{Forecast} = 0.2\ Y + 0.8\ \text{Forecast}(-1)$$

where Y is the most recent head count value, Forecast is the forecast of the head count in the next period, Forecast(1) is the forecast of the most recent head count made in the previous period. We can use 13109.273 as initial value of Forecast(-1)

For Carabao:

$$\text{Forecast} = 0.2\ Y + 0.8\ \text{Forecast}(-1)$$

with 47829.042 as initial value of Forecast(-1).

For Chicken:

$$\text{Forecast} = 0.2\ Y + 0.8\ \text{Forecast}(-1)$$

with 526778.79 as initial value of Forecast(-1).

For Duck:

$$\text{Forecast} = 0.2\ Y + 0.8\ \text{Forecast}(-1)$$

with 57802.13 as initial value of Forecast(-1).

For Hog:

$$\text{Forecast} = 0.2\ Y + 0.8\ \text{Forecast}(-1)$$

with 102180.85 as initial value of Forecast(-1).

2. Hazard Probability Models

To assess the potential hazards in the surroundings of the volcano, we need to assess the probability of eruption at various points in time. The first model that we will construct is a probability model for volcanic eruption, regardless of the type and the associated hazard. The next step will be the determination of the probability of occurrence of a specific hazard in a given type of eruption.

Probability of Eruption

The repose period (in years) or the distance between two consecutive eruptions will be used in construction a probabilty a model for eruption. The data on 44 eruptions (includes minor ash emissions) were considered. A total of 43 repose periods can be determined. The values range from 1 to 150 years. The repose period of 150 years was deleted from analysis since it is clearly an outlying value (see Table

1). The cumulative probability (at x years of repose) given in Table 1 gives the probability that will erupt in not more than x years. Thus, the curve relating cumulative probabililty to repose period will give us the desired probability model.

The scatter plot of empirical cumulative probability (ecp) versus repose period is given in Figure 1. The ecp increases sharply at shorter repose periods and tend to level off after repose period of 7 years. The Michaelis-Menten model given by the function

$$F(x)=\frac{\phi_1 x}{\phi_2+x}$$

best describes the ecp.

The nonlinear least squares procedure give the following estimates:

$$\hat{F}(x)\frac{1.05059x}{2.257235+x}$$

When $\hat{F}(x)$ exceeds 1, we truncate it to 1. To assess the fit of the model, we compute the predicted cumulative probability and compare it to actual values (see Table 2). From (2), the probability that will erupt within 10 years after the most recent eruption is 0.8751.

Probability of a Hazard

In this project we consider four volcanic hazards namely: ashfall, pyroclastic flow, lava flow and lahars. Furthermore, we consider only primary lahars or lahars which occure only during eruptions. Lahar

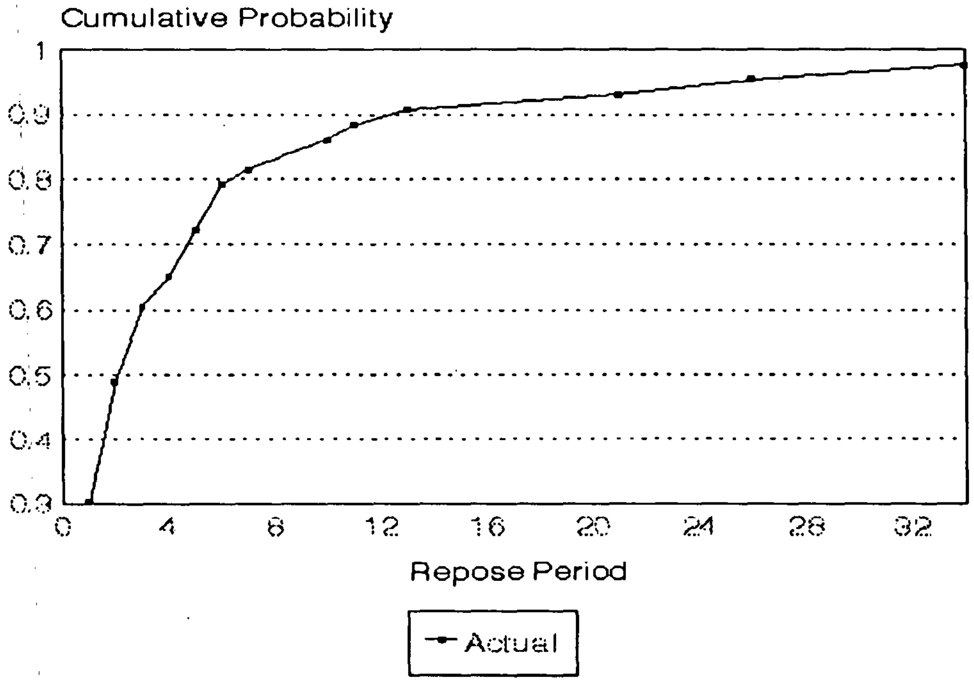


Fig. 1 Cumulative Probability Distribution

Table 1

Repose Period (in years)	Frequency	Cumulative Frequency	Cumulative Probability
1	13	13	0.3023
2	8	21	0.4884
3	5	26	0.6047
4	2	28	0.6512
5	3	31	0.7209
6	3	34	0.7907
7	1	35	0.8140
10	2	37	0.8605
11	1	38	0.8837
13	1	39	0.9070
21	1	40	0.9302
26	1	41	0.9535
34	1	42	0.9767
150	1	43	1.0000

Table 2

Repose Period (in years)	Cumulative Probability	Predicted
1	0.3023	0.3225
2	0.4884	0.4936
3	0.6047	0.5995
4	0.6512	0.6716
5	0.7209	0.7238
6	0.7907	0.7634
7	0.8140	0.7944
10	0.8605	0.8571
11	0.8837	0.8717
13	0.9070	0.8952
21	0.9302	0.9486
26	0.9535	0.9667
34	0.9767	0.9852
150	1.0000	1.0000

is a volcanic hazardwhich may occur even during a volcano's repose period. The occurence of hazards will depend in part on the type of eruption, e.g., the likelihood of a pyroclastic is very high during a vulcanian eruption and low during strombolian eruption. The following are the frequency of occurence of each hazard and the corresponding empirical probabilities:

Type	**Frequency**	**Probability**
Ashfall	24	1.000
Pyroclastic Flow	18	0.750
Lava Flow	15	0.625
Lahar	14	0.583

3. Vulnerability models

Vulnerabililty is defined as the likelihood that an area will be hit by a particular hazard. It is a function of: 1) the characteristics and magnitude of the hazard; 2) site specific conditions/characteristics such as local and regional geologic conditions, distance from the hazard source, location with respect to major structural features, topograhpy, slope, drainage, etc.; 3) the presence or absence of protective/mitigation and preparedness structures. The vulnerability models were based on: 1) geologic maps; 2) stratigraphic maps; 3) hazard zonation maps; 4) land use maps; and 5) historical accounts.

Lava Flow

In constructing the probability model for lava flows, we first determine the probability it will reach a certain distance from the crater. The likelihood that an area will be hit by the hazard was computed based on geologic maps and stratigraphic columns. Note that a lava flow reaching a certain distance also affected the areas it had passed. For example, a lava flow that had travelled 5 km. for the crater had also affected the areas 4, 3, 2 and 1 km. from the crater. Using nonlinear least squares, the likelihood that an area will be hit by the hazard is

$$D(x) = \frac{1}{0.9647 + 0.0004 \cdot \exp(1.7765 \cdot x)}$$

where x is the distance of the area from the crater.

Lava flows are usually channel-confined or they follow existing topographic depressions. To determine the effect of topography, we calculated the proportion of cumulative channel area with other areas within 1 sq. km. The result is shown in Figure 2. The vulnerability model of an area is then

$$V(x) = D(x) * T(x)$$

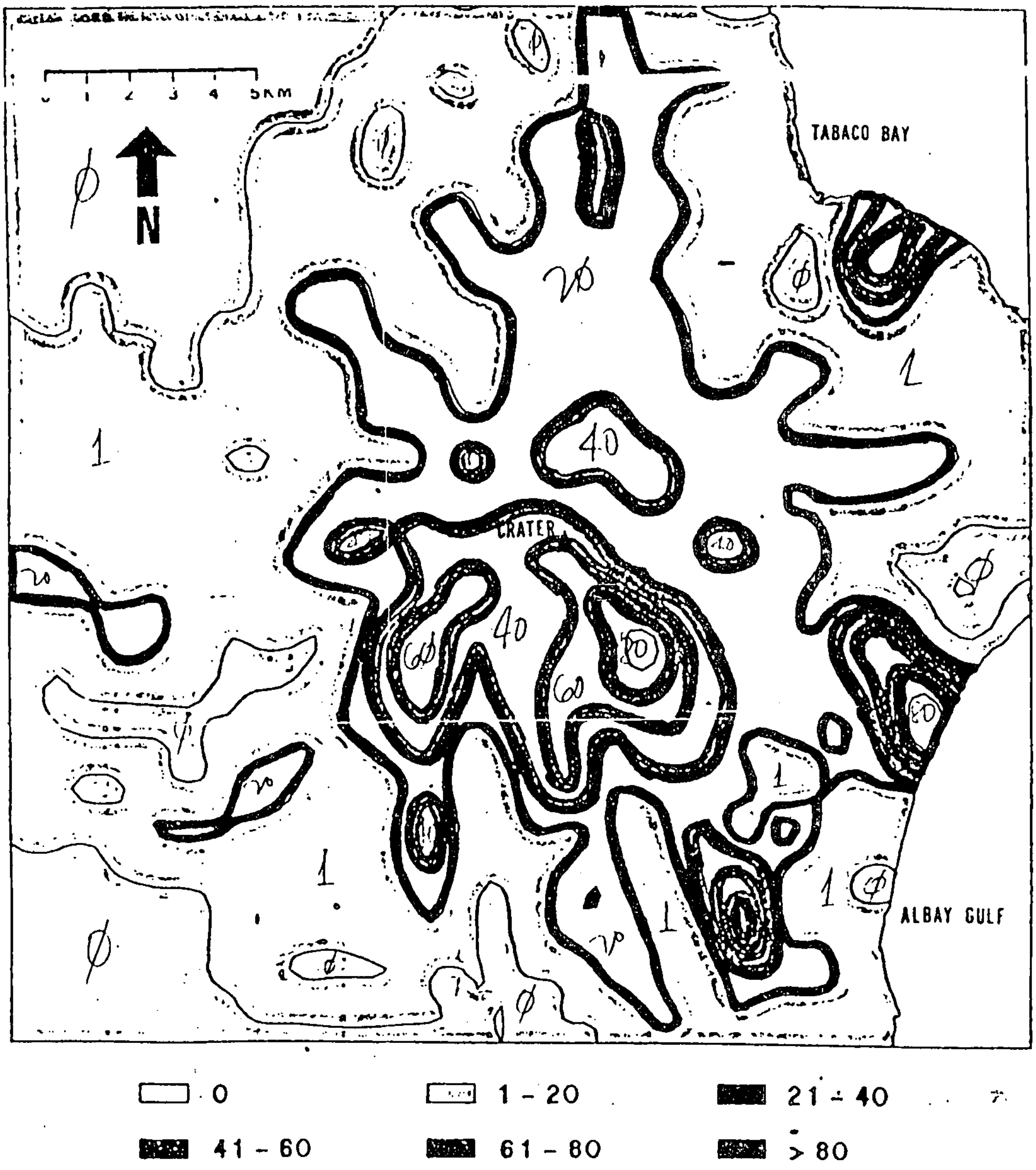
where $T(x)$ is the topography factor given in Figure 2.

Pyroclastic Flow

As in lava flows, we follow the same procedure for pyroclastic flows since both are flowage phenomena although pyroclastic flows have a higher velocity and greater energy. Inputs for $D(x)$ were also taken from geologic maps and stratigraphic columns. The likelihood is thus

$$D(x) = \frac{1}{0.9876 + 0.0029 \cdot \exp(1.0324 \cdot x)}$$

Figure 2


0 1 2 3 4 5KM
N
TABACO BAY
CRATER
ALBAY GULF
0
1
20
40
60
80
0
1 - 20
21 - 40
41 - 60
61 - 80
> 80

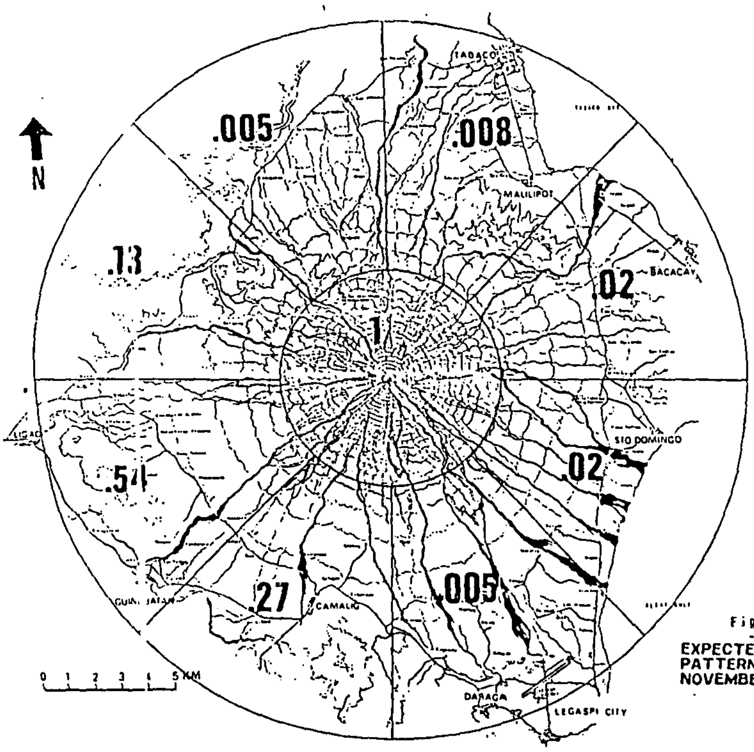


Figure 3

EXPECTED ASH DISPERSAL PATTERN FROM NOVEMBER TO MAY

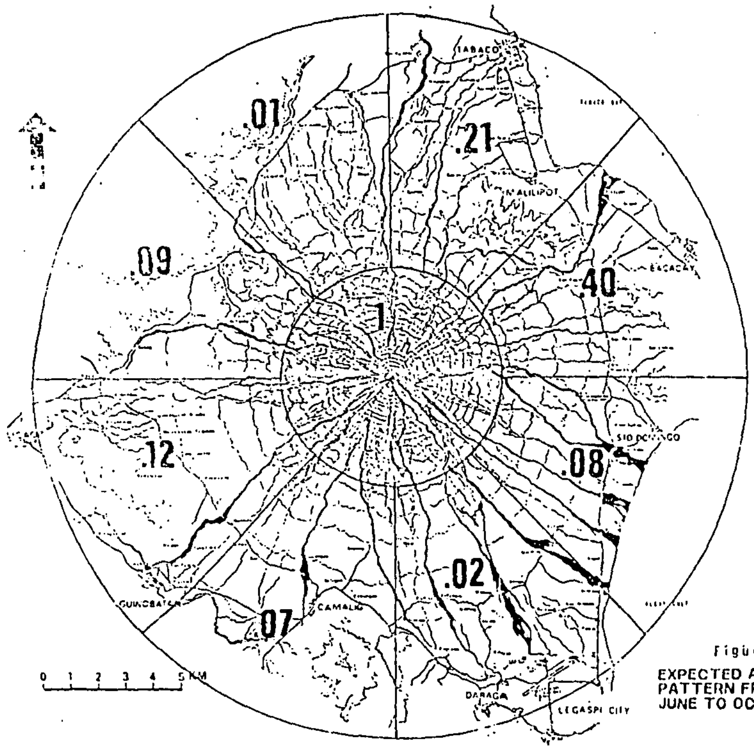


Figure 4

EXPECTED ASH DISPERSAL PATTERN FROM JUNE TO OCTOBER

And the vulnerability model is

$$V(x) = D(x) * T(x).$$

Lahars

Lahar is a volcanic hazard that occurs even during a volcano's repose period. Here we deal only with primary or eruption-related lahars. Most of the lahars originated at the 750 m. elevation or about 3.5 km. from the crater. Based on geologic maps, stratigraphic columns, hazard zonation map and recent field surveys, the probability of occurrence of lahar give the distance from the crater x is

$$D(x) = \frac{1}{1.6233 + 408234.8126 \cdot \exp(-2.12121 \cdot x)}$$

Since lahar is also a flowage phenomena, the vulnerability model is

$$V(x) = D(x) * T(x)$$

Ash fall

In the construction of the probability model for ashfall, we considered only areas that will be covered by at least 20 cm. ashfall deposit. Based on the hazard zonation map, the likelihood model is

$$D(x) = \frac{1}{1 + \exp(--17.671 + 1.8225 \cdot x)}$$

The dispersal of ash and therefore the likely to be afftected will depend highly on the prevailing wind direction. From November to May, the winds come from the NE while from June to October the prevailing wind direction is form SW. Thus, the calculation of vulnerability from ashfall will depend on which month the assessment is targeted. Projected ash dispersal was based on near surface wind direction data from 1903 to 1980. The vulnerability to ashfall may then be calculated from

$$V(x) = D(x) * W(x)$$

where $W(x)$ is the wind factor given in Figures 3 and 4.

4. Risk Estimation

The risk of an element is the product of its economic value and the likelihood that it will be damaged. The probability that it will be damaged is dependent on the occurrence of an eurption, the type of associated hazard and the vulnerabilty of an area where this element is located. Thus, to compute the risk of an element, we use Bayes' theorem repeatedly. The risk function is given by

Risk = P(eruption) * P(hazard/eruption)
* P(vulnerability/eruption and hazard)
* economic value of element at risk.

The value of this function for all elements in a barangay are added to derive the average loss for that place in the event of an eruption.

References

Bates, D. M. and D. G. Watts. **Nonlinear Regression Analysis and Its Applications**. John Wiley and Sons, 1988.

Philippine Journal of Volcanology. Special Issue on Mayon Volcano, PHIVOLCS