Estimating conditional probability of volcanic flows for forecasting event distribution and making evacuation decisions

E.R. Stefanescu[1], A.K. Patra [1], M.F. Sheridan [2], G. Cordoba [3]

[1]Mechanical and Aerospace Engineering

University at Buffalo

[2]Geology, University at Buffalo

[3]Universidad de Nariño, Colombia

Outline

1. Introduction
   - Volcanic hazard
   - Conditional probability framework
2. Methodology
   - Frequency analysis
   - Logistic regression
   - Bayes analysis
3. Conclusions

Galeras volcano

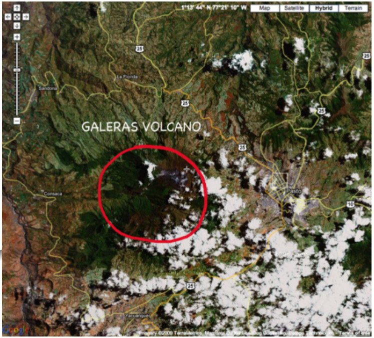
Galeras volcano is one of the most active volcanoes in the world. Nearly 400,000 people currently live near the volano.

- Deal with complex physics.
- Sparse observations - one cannot build a forecast based only on observations.
- To supplement the lack of observations we used a simulator /computer model - TITAN2D.

Questions

During a crisis

Should a village be evacuated and villagers moved/relocated to a different location?



The problem

- A critical action to reduce volcanic risk/hazard is the evacuation of people from threatened areas during a volcanic eruption.
- Decision making requires a comparison of hazards at different locations.
- The range of possible inputs (from experts) allows us to *frame the range of possible outcomes*.
- Range of inputs can be explored using the simulator by space filling sampling (Latin Hypercube Design) - the result is a relative **probability matrix**

The particular choice of method used in finding the relative probability can be based on the amount of information necessary in the evacuation decision and on the complexity of the analysis required in taking such decision.

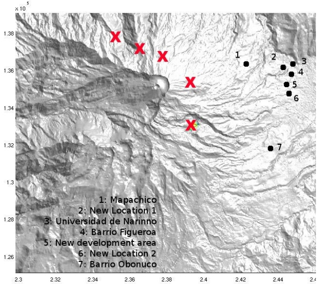


Figure: Galeras Volcano, Colombia - red X represents a possible vent location and black dot is a critical regions (village, infrastructure, city, university)

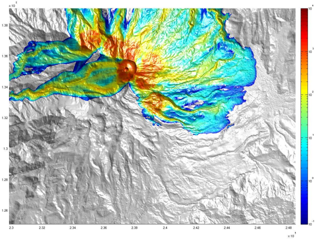


Figure: Maximum flow height over time - 4900 Titan2D runs

Frequency analysis - One possible starting zone

	Location 2			
Location 6		Nonflow	Flow	**Total**
	Nonflow	633	21	**654**
	Flow	11	35	**46**
	Total	**644**	**56**	**700**

Table: Location 2 and location 6

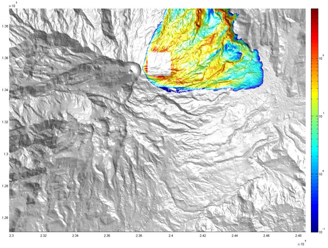


Figure: Maximum flow height over time (one starting zone) - 700 Titan2D

One possible starting zone - cont.

		Location 2	
		Nonflow	Flow
Location 6	Nonflow	0.452	0.015
	Flow	0.007	0.025

Table: Joint distribution

$$P_{location2} = \begin{pmatrix} Nonflow & Flow \\ 0.920 & 0.080 \end{pmatrix}$$

$$P_{location6} = \begin{pmatrix} Nonflow & Flow \\ 0.935 & 0.065 \end{pmatrix}$$

P(loc6 = Flow, loc2 = Flow) = 0.025

P(loc6 =Flow | loc2 =Flow) = 0.625

Seven possible starting zones

		Location 2		
		Nonflow	Flow	**Total**
Location 6	Nonflow	4687	83	**4770**
	Flow	30	100	**130**
	Total	**4717**	**183**	**4900**

Table: Location 2 and location 6

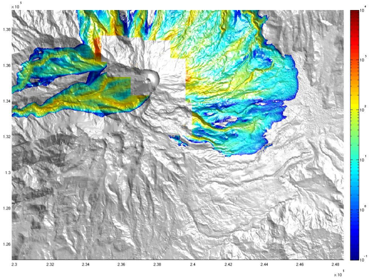


Figure: Maximum flow height over time (7 starting zones) - 4900 Titan2D

Frequency analysis - cont.

		Location 2	
		Nonflow	Flow
Location 6	Nonflow	0.956	0.016
	Flow	0.006	0.020

Table: Joint distribution

$$P_{location2} = \begin{pmatrix} Nonflow & Flow \\ 0.851 & 0.149 \end{pmatrix}$$

$$P_{location6} = \begin{pmatrix} Nonflow & Flow \\ 0.973 & 0.026 \end{pmatrix}$$

P(loc6 = Flow, loc2 = Flow) = 0.020

P(loc6 =Flow | loc2 =Flow) = 0.546

Logistic regression

We consider:

$P(t=1\,|\,x)=f(x)$ and $P(t=0\,|\,x)=1-f(x), f\in[0,1]$

The logistic function is defined as:

$$f(x)=\sigma(x)=\frac{1}{1+exp(-x)} \tag{1}$$

Define: $P(t=1\,|\,x)=\sigma(b+x^T w)$, where b-bias parameter

The likelihood function can be written as:

$$P(D)=\prod_{i=1}^{N}P(t^i\,|\,x^i)=\prod_{i=1}^{N}P(t=1\,|\,x^i)^{t^i}(1-P(t=1\,|\,x^i)^{1-t^i} \tag{2}$$

Logistic regression - cont

Define L as the log of likelihood

$$L = logP(D\,|\,w,b) = \sum_{i=1}^{N}(t^i \log\sigma(b+w^Tx^i) + (1-t^i)\log(1-\sigma(b+w^Tx^i))) \tag{3}$$

Solve for w and b:

$$\nabla_w L = \sum_i^N (t^i - \sigma(b+w^Tx^i))x^i = 0$$

$$\nabla_b L = \sum_i^N (t^i - \sigma(b+w^Tx^i)) = 0$$

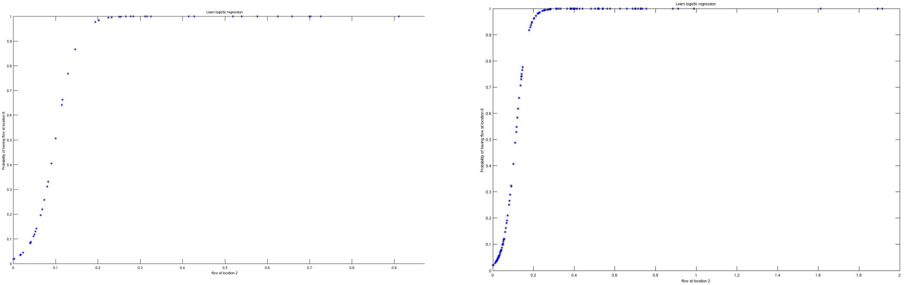


Figure: The probability of having flow at location 6 given flow at location a) one location b) seven locations

Bayes analysis

Define y - event of flow occurring ($y = 1$ - Flow, $y = 0$ - Nonflow) at a certain location.

Let ∂ be the probability that $y = 1 : P(y = 1) = \partial$.

Bayes' rule:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B)} \tag{4}$$

In our case we have:

$$\pi(\partial \mid data) = \frac{\pi(data \mid \partial)\pi(\partial)}{\pi(data)}, \tag{5}$$

where $\pi(data \mid \partial)$ - is the likelihood function.

Bayes analysis -cont.

The posterior distribution mean

$$E(\partial \,|\, data) = \frac{\alpha_0 + \sum_{i=1}^{N} y_i}{\alpha_0 + \sum_{i=1}^{N} + \beta_0 - N - \sum_{i=1}^{N} y_i} \tag{6}$$

$$\frac{\alpha_0 + \beta_0}{\alpha_0 + \beta_0 + N} \cdot \frac{\alpha_0}{\alpha_0 + \beta_0} + \frac{N}{\alpha_0 + \beta_0 + N} \cdot \bar{y}$$

where $\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$

The likelihood is binomial $\pi(data \,|\, \partial) = \partial^{\alpha_F} (1 - \partial)^{\alpha_N}$.

The prior is beta distribution: $\pi(\partial) < \partial^{\beta_F - 1} (1 - \partial)^{\beta_N - 1} = B(\beta_F, \beta_N)$

The posterior is beta distribution: $\pi(\partial \,|\, data) = B(\beta_F + \alpha_F, \beta_N + \alpha_N)$.

Results one zone

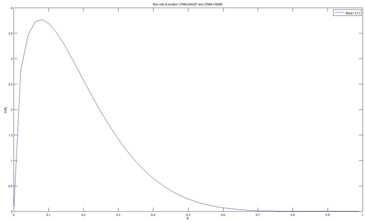


Figure: Probability density function (pdf) - Beta distribution

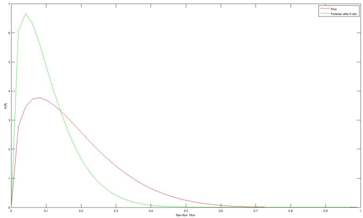


Figure: Prior and Posterior PDF after 6 observations

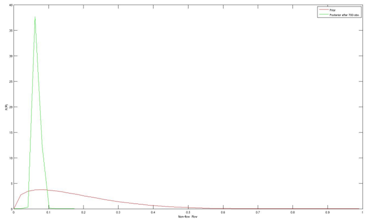


Figure: Prior and Posterior pdf after 4900 observations

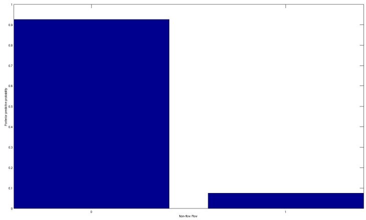


Figure: Posterior predictive probability

Results seven zones

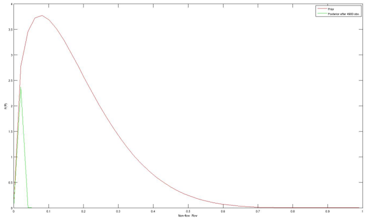


Figure: Prior and Posterior pdf after 4900 observations - 7 zones

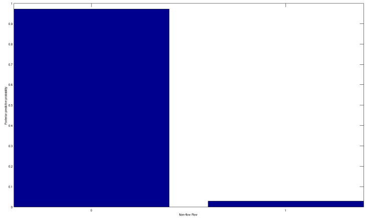
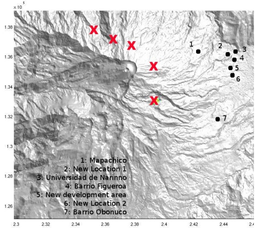
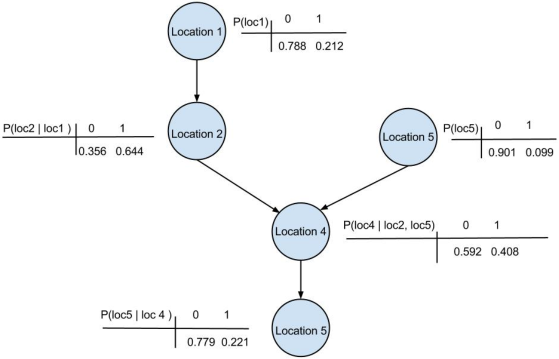


Figure: Posterior predictive probability

Graphical Models - Bayesian Network

In a probabilistic graphical model each node represents a random variable and the edges express probabilistic relationship between variables





Directed Graphical Model

Product terms are conditional distributions of each node conditioned on variables corresponding to parents of that node in the graph:

$$p(X) = \prod_{k=1}^{K} p(x_k \mid pa_k) \tag{7}$$

where $x = (x_1, \ldots, x_K)$ and pa_k denotes set of parents of a_k

By inspection, joint ditribution of the "flow graph" is given by:

$$P(loc3, loc4, loc5, loc2, loc1) = P(loc3 \mid loc4) \cdot P(loc4 \mid loc2, loc5) \cdot \tag{8}$$

$$P(loc5) \cdot P(loc2 \mid loc1) \cdot P(loc1) = 0.001$$

Conclusions

- Developed a framework for forecasting event distribution and making evacuation decision.
- Space filling sampling design used to explore the input parameter space.
- The choice of method used for finding relative probability is based on the amount of information necessary in the evacuation decision.
- This work was supported by ....