A Simple Pyroclastic Flow Model

Robert L Wolpert
October 23, 2006

1 Introduction

Page 14 of Pittman (2006) (reprinted as Figure (1) below) presents a frequency plot of the magnitudes of pyroclastic flows (PFs) of the Soufrière Hills Volcano on the Caribbean island of Montserrat. The log-log plot looks strikingly linear. In the hope of making predictions and inference about eruptions, let's build the simplest possible model for these data— later we can embellish (or abandon) it as needed.

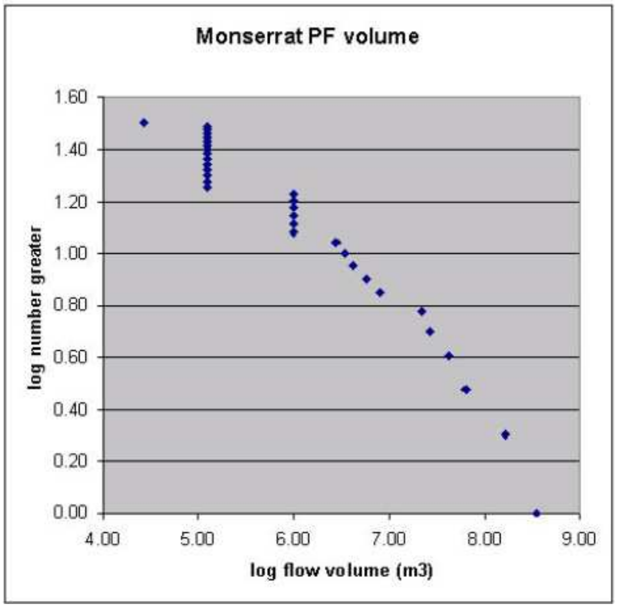

Figure 1: Reprint of figure from Pittman (2006, *p.* 14)

The data consist of a number J_ϵ of estimated flow volumes $\{v_j\}_{j \le J_\epsilon}$ (in m^3) observed at unspecified times $\{t_j\}_{j \le J_\epsilon}$. It is understood that these are all of the PFs exceeding some known threshold $\epsilon > 0$ (apparently $\epsilon = 10^4\,\text{m}^3$) observed during some time period of length T years (three years or so?).

2 A Simple Model

The simplest plausible model would accord the number J_ϵ of PFs exceeding threshold volume ϵ a Poisson distribution $J_\epsilon \sim \mathsf{Po}(\lambda_\epsilon T)$ with mean proportional to the time T, for some unknown rate $\lambda_\epsilon > 0$. Event times $\{t_j\}_{j\le J_\epsilon}$ and flow volumes $\{v_j\}_{j\le J_\epsilon}$ would be independent, with uniformly distributed event times $\{t_j\}_{j\le J_\epsilon} \overset{\text{iid}}{\sim} \mathsf{Un}(0,T)$ and with volumes $\{v_j\}_{j\le J_\epsilon} \overset{\text{iid}}{\sim} F_\epsilon(v)$ drawn from some (as yet) unspecified distribution with CDF $F_\epsilon(v) = \mathsf{P}[v_j \le v]$. Under such a model, for any intervals $A_i, B_i \subset \mathbb{R}_+$ the numbers $J_i \equiv J_\epsilon(A_i \times B_i)$ of flows of volumes $v \in A_i$ at times $t \in B_i$ would be independent Poisson-distributed random variables, for disjoint rectangles $\{A_i \times B_i\}_{i\le I}$, with means $\mathsf{E}[J_i] = F_\epsilon(A_i)|B_i|$ (the increment of F_ϵ over A_i times the length of B_i), so the expected number of PFs with volumes exceeding v in time T would satisfy

$$\mathsf{E}\Big[J_\epsilon\big((v,\infty) \times (0,T]\big)\Big] = \lambda_\epsilon T \bar{F}_\epsilon(v) \tag{1}$$

where $\bar{F}_\epsilon(v) = 1 - F_\epsilon(v) = \Pr[v_j > v]$ is the complementary CDF (cCDF). This will decrease linearly on a log-log scale for $v > \epsilon$ if and only if

$$\log(\lambda_\epsilon T) + \log \bar{F}_\epsilon(v) \quad = \quad c - \alpha \log v, \qquad v > \epsilon$$

for some constants $c \in \mathbb{R}$ and $\alpha > 0$, *i.e.*,

$$\bar{F}_\epsilon(v) \quad \propto \quad v^{-\alpha}, \qquad v > \epsilon.$$

The solution with $\bar{F}_\epsilon(\epsilon) = 1$ is the Pareto $V \sim \mathsf{Pa}(\alpha, \epsilon)$, with CDF $F_\epsilon(v) = 1 - (\epsilon/v)^\alpha$ for $v > \epsilon$ and density function

$$f_\epsilon(v) = \alpha \epsilon^\alpha v^{-\alpha-1}\, \mathbf{1}_{\{v>\epsilon\}}.$$

From Equation (1), the expected number of flows of volumes exceeding any fixed number $v \ge \epsilon$ in time T is now $\lambda_\epsilon T \bar{F}_\epsilon(v) = \lambda_\epsilon T (\epsilon/v)^\alpha$. Since this cannot depend on ϵ, the quantity $\lambda \equiv \lambda_\epsilon \epsilon^\alpha$ must be constant and we can rewrite this as

$$\mathsf{E}\Big[J_\epsilon\big((v,\infty) \times (0,T]\big)\Big] = \lambda T v^{-\alpha}, \qquad v \ge \epsilon \tag{2}$$

as a function of the two parameters $\alpha, \lambda \in \mathbb{R}_+$.

2.1 Likelihood

The likelihood function upon observing $\{(v_j, t_j)\}_{j \le J_\epsilon}$ is thus

$$\begin{aligned} L(\alpha, \lambda_\epsilon) &= \frac{(\lambda_\epsilon T)^{J_\epsilon}}{J_\epsilon!} e^{-\lambda_\epsilon T} \prod_{j \le J_\epsilon} \left[\alpha \epsilon^\alpha v_j{}^{-\alpha-1} \mathbf{1}_{\{v_j > \epsilon\}} T^{-1} \mathbf{1}_{\{0 < t_j \le T\}} \right] \\ &\propto (\lambda_\epsilon \, \alpha)^{J_\epsilon} \exp\Big[-\lambda_\epsilon T - \alpha \sum_{j \le J_\epsilon} \log(v_j/\epsilon) \Big], \qquad \text{or} \qquad (3) \\ L(\alpha, \lambda) &\propto (\lambda \, \alpha)^{J_\epsilon} \exp\Big[-\lambda T \epsilon^{-\alpha} - \alpha \sum_{j \le J_\epsilon} \log v_j \Big] \qquad (4) \end{aligned}$$

for $\epsilon < \min\{v_j\}$ and $0 < \min\{t_j\} \le \max\{t_j\} \le T$. Maximum likelihood estimators (MLEs) are easy to compute, and depend on the data only through the sufficient statistics J_ϵ and $S_\epsilon \equiv \sum \log(v_j/\epsilon)$:

$$\hat{\alpha} = \frac{J_\epsilon}{S_\epsilon} \qquad \hat{\lambda}_\epsilon = \frac{J_\epsilon}{T} \qquad \hat{\lambda} = \frac{J_\epsilon \, \epsilon^{\hat{\alpha}}}{T}.$$

More interesting is that Equation (3) is proportional to a product of gamma densities in λ_ϵ and α; with independent gamma (or reference) prior distributions on $\alpha \sim \mathsf{Ga}(a_\alpha, b_\alpha)$ and $\lambda_\epsilon \sim \mathsf{Ga}(a_{\lambda_\epsilon}, b_{\lambda_\epsilon})$, the posterior distributions are again independent gammas:

$$\alpha|_{\text{ data}} \sim \mathsf{Ga}\Big(a_\alpha + J_\epsilon, \, b_\alpha + S_\epsilon\Big) \quad \lambda_\epsilon|_{\text{ data}} \sim \mathsf{Ga}\big(a_{\lambda_\epsilon} + J_\epsilon, \, b_{\lambda_\epsilon} + T\big)$$

with parameters $a^*_\alpha \equiv a_\alpha + J_\epsilon$, $b^*_\alpha \equiv b_\alpha + S_\epsilon$, $a^*_{\lambda_\epsilon} \equiv a_{\lambda_\epsilon} + J_\epsilon$, and $b^*_{\lambda_\epsilon} \equiv b_{\lambda_\epsilon} + T$. Conditional on α, the implicit prior distribution of the rescaled $\lambda \equiv \lambda_\epsilon \epsilon^\alpha$ would be $\lambda \sim \mathsf{Ga}(a_\lambda, b_\lambda)$ with $a_\lambda \equiv a_{\lambda_\epsilon}$ and $b_\lambda \equiv b_{\lambda_\epsilon} \epsilon^{-\alpha}$, with posterior distribution

$$\lambda \mid_{\text{ data},\alpha} \sim \mathsf{Ga}\big(a_{\lambda_\epsilon} + J_\epsilon, \, (b_{\lambda_\epsilon} + T)\epsilon^{-\alpha}\big) = \mathsf{Ga}\big(a_\lambda + J_\epsilon, \, b_\lambda + T\epsilon^{-\alpha}\big)$$

The parameters α and λ won't be independent under the posterior, but we can still exploit the simple form of the likelihood given in Equation (4).

2.2 Predictions

Fix some number $v > 0$ and time $t > 0$. Under the model of Section (2), the probability that the Poisson-distributed number J^*_v of PFs of magnitudes

$v_i > v\,\mathrm{m}^3$ during a future interval of length t yrs is exactly *zero* would be simply $\exp(-\lambda_\epsilon t(\epsilon/v)^\alpha)$; conversely, the probability of at least one catastrophic eruption of volume $V > v$ in the next t years would be

$$\mathsf{P}[\text{At least one PF} > v \text{ in } t \text{ yrs} \mid \alpha, \lambda_\epsilon] = 1 - e^{-\lambda_\epsilon t(\epsilon/v)^\alpha}, \qquad v \ge \epsilon$$

as a function of the parameters α and λ_ϵ. Typically these parameters are unknown, but under the independent conjugate prior distributions of Section (2.1), we can compute the posterior probability

$$\mathsf{P}[\text{At least one PF} > v \text{ in } t \text{ yrs} \mid \text{data}] = \iint_{\mathbb{R}_+^2} \left(1 - e^{-\lambda_\epsilon t\,(\epsilon/v)^\alpha}\right) \pi^*(d\alpha, d\lambda_\epsilon)$$

for the posterior distribution $\pi^*(d\alpha, d\lambda_\epsilon)$ with density

$$\pi^*(\alpha, \lambda_\epsilon) \quad = \quad \frac{\alpha^{a_\alpha^*-1}(b_\alpha^*)^{a_\alpha^*}}{\Gamma(a_\alpha^*)} e^{-b_\alpha^*\alpha} \times \frac{{\lambda_\epsilon}^{a_{\lambda_\epsilon}^*-1}(b_{\lambda_\epsilon}^*)^{a_{\lambda_\epsilon}^*}}{\Gamma(a_{\lambda_\epsilon}^*)} e^{-b_{\lambda_\epsilon}^*\lambda_\epsilon}, \qquad \alpha, \lambda_\epsilon > 0.$$

The integral with respect to $d\lambda_\epsilon$ can be evaluated in closed form, leaving the one-dimensional integral with respect to $\pi^*(d\alpha) = \mathsf{Ga}(a_\alpha^*, b_\alpha^*)$:

$$\mathsf{P}\Big[\text{PF} > v \text{ in} \le t \text{ yrs} \mid \text{data}\Big] = 1 - \int_{\mathbb{R}_+} \left[1 + (t/b_{\lambda_\epsilon}^*)(\epsilon/v)^\alpha\right]^{-a_{\lambda_\epsilon}^*} \pi^*(d\alpha). \quad (5)$$

The reference case ($a_\alpha = b_\alpha = a_{\lambda_\epsilon} = b_{\lambda_\epsilon} = 0$) leads to:

$$\mathsf{P}\Big[\text{PF} > v \text{ in} \le t \text{ yrs} \mid \text{data}\Big] = 1 - \int_{\mathbb{R}_+} \frac{\alpha^{J_\epsilon - 1} e^{-\alpha S_\epsilon} {S_\epsilon}^{J_\epsilon}}{\left[1 + (t/T)(\epsilon/v)^\alpha\right]^{J_\epsilon} \Gamma(J_\epsilon)}\, d\alpha$$

where $S_\epsilon \equiv \sum \log(v_j/\epsilon)$. Figure (2) shows a plot of this function, for the simulated data used in Figure (3) below, for $t = 10$ and $t = 100$; horizontal lines help identify selected quantiles of the distribution (10%, 25%, 50%, 75%, 90%). The largest observed flow in the three-year simulation was $1.42 \times 10^8\,\mathrm{m}^3$, but the model predicts an even chance of a flow exceeding $V > 3.55 \times 10^{11}$ in a decade and $V > 10^{15}$ in a century.

The computations are significantly more involved for independent prior distributions on α and λ than for those on α and λ_ϵ.

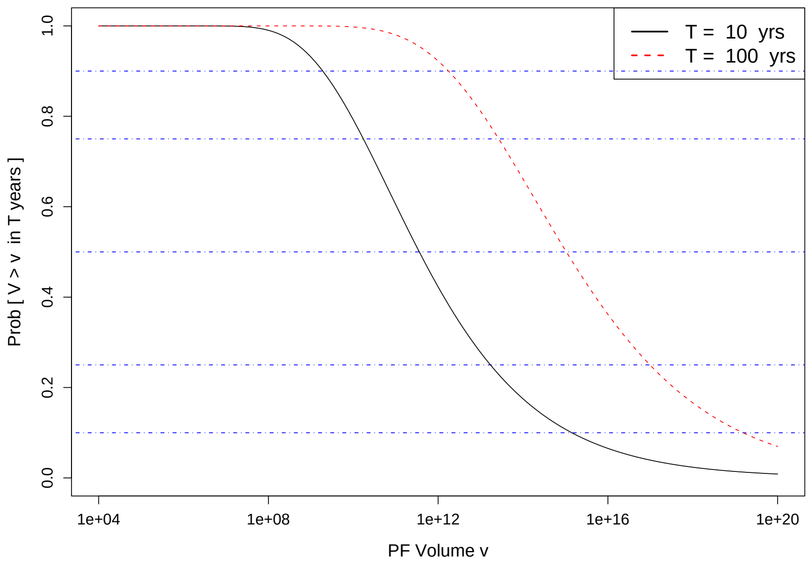

Figure 2: Posterior expected probability of PF volume exceeding v in t years, for $10^4 \le v \le 10^{20}$ and $t = 10$ (solid line) and $t = 100$ (dashed line).

2.3 Checking the Model

Under the model of Section (2), the points

$$\{(x_j, y_j) \equiv \big((\epsilon/v_j)^\alpha, t_j/T\big)\}_{j\le J_\epsilon}$$

and, with $t_0 \equiv 0$,

$$\{(x_j, z_j) \equiv \big((\epsilon/v_j)^\alpha, e^{-\lambda_\epsilon(t_j-t_{j-1})}\}_{j\le J_\epsilon}$$

would each be J_ϵ independent draws from the unit square. We might be able to detect departures from that distribution, which might suggest ways to alter and improve the model. The times may be *less* irregular than *i.i.d.* uniform draws, for example, if there is a “refractory period” as in neural spike trains; they may be *more* irregular than *i.i.d.* uniform draws if there are “aftershocks” as in earthquakes. Flow volumes may tend to be larger after relatively long inter-event times and smaller after shorter intervals, perhaps. All these departures can be modeled, once we discover them.

3 Relation to α-Stable

The threshold $\epsilon > 0$ was set arbitrarily. What happens if we change ϵ? Or take the limit as $\epsilon \to 0$? Under the model of Section (2) the cumulative PF volume over the interval $(0, t]$

$$X_t^\epsilon \equiv \sum \{v_j \mid v_j > \epsilon,\ t_j \le t\}$$

has an infinitely-divisible (ID) distribution— it is, in fact, a stationary independent increment (SII) or "Lévy" process with Lévy measure

$$\nu(dv) = \alpha\lambda\, v^{-\alpha-1}\mathbf{1}_{\{v>\epsilon\}}\, dv.$$

For $\alpha \ge 1$ there is no nontrivial convergence as ϵ shrinks— the tiny PFs accumulate without bound, so $X_t^\epsilon \to \infty$ almost-surely as $\epsilon \to 0$. For $0 < \alpha < 1$, however, there is an interesting limit as $\epsilon \to 0$: the distribution of $\{X_t^\epsilon\}$ will converge to the fully-skewed pure-jump α-stable distribution $X_t \sim \mathsf{St}(\alpha, 1, \gamma t, \delta)$ with rate constant γ and drift constant δ that depend[1] only on α and λ. Moreover, the observed sequence $\{(v_j, t_j)\}_{j \le J_\epsilon}$ is exactly the collection of jumps $\{\Delta X_t \equiv [X_t - X_{t-}]\}$ and jump times of magnitudes $\Delta X_t > \epsilon$ of the fully skewed α-stable process X_t. Note that the flow volumes do not have finite means or variances, and the central limit theorem does not apply; their cumulative sums converge to stable distributions, not Gaussians, as $\epsilon \to 0$ or as $t \to \infty$.

3.1 A Simulation

A crude estimate from Figure (1) suggests that the slope is about -0.3, suggesting the plausibility of this model with $\epsilon \approx 10^4$, $\alpha \approx 0.30$, and that $J_\epsilon \approx 1 + 10^{1.50} \approx 33$, so $\lambda_\epsilon T \approx 33$. Figure (3) shows the analogous survival plot from a simulation of the jumps of size $v > \epsilon = 10^4$ of the skewed α-stable process $X_t \sim \mathsf{St}(\alpha, 1, \gamma t, \delta)$ with $\alpha = 0.30$, $\beta = 1$ (*i.e.*, fully skewed), γ chosen to attain $\mathsf{E}[J_\epsilon] = 33$, and $\delta = \gamma \tan(\pi\alpha/2)$ for a pure-jump process. Repeated runs of the simulation give some idea of how variable are the flow distributions (and α estimates).

References

Pittman, E. B. (2006) Granular avalanche modeling. Overheads from 2006-10-02 presentation in SAMSI Granular Materials working group.

---

[1] The dependence is given by: $\gamma = \pi\lambda/\{2\Gamma(\alpha)\sin(\pi\alpha/2)\}$ and $\delta = \gamma\tan(\pi\alpha/2)$; the latter balances exactly the usual stable-process "compensation" for a pure-jump process.

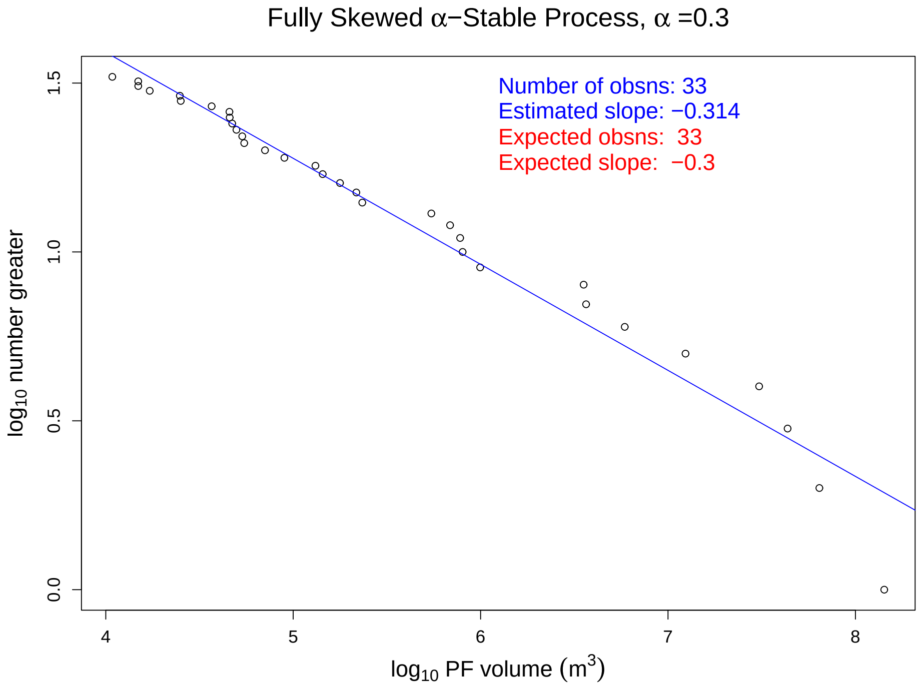

Figure 3: Simulated PF volumes from skewed α-stable process.