Hazard Assessment for Pyroclastic Flows

Robert L Wolpert

Duke University
Department of Statistical Science
with
James O Berger (Duke),
E. Bruce Pitman, Eliza S. Calder & Abani K. Patra (U Buffalo),
M.J. Bayarri (U Valencia), & Elaine T. Spiller (Marquette U).

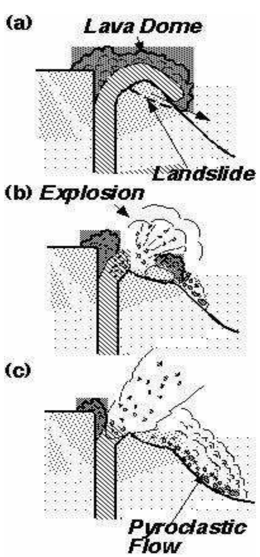
Games & Decisions in Reliability & Risk:
2011 May 21, Laggo Maggiore, IT

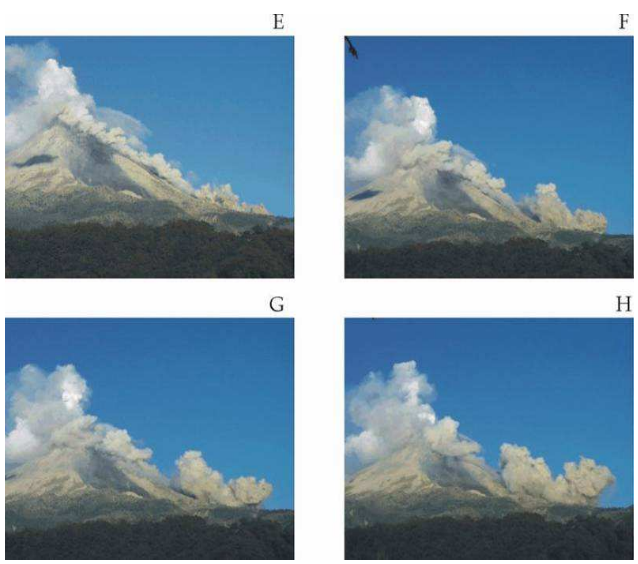
Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

The Soufrière Hills Volcano on Montserrat

Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

Pyroclastic Flows





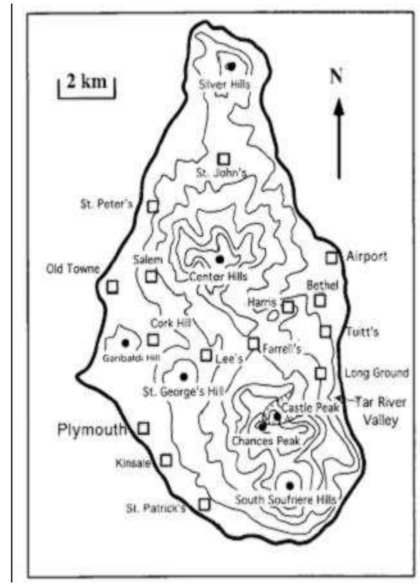
Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

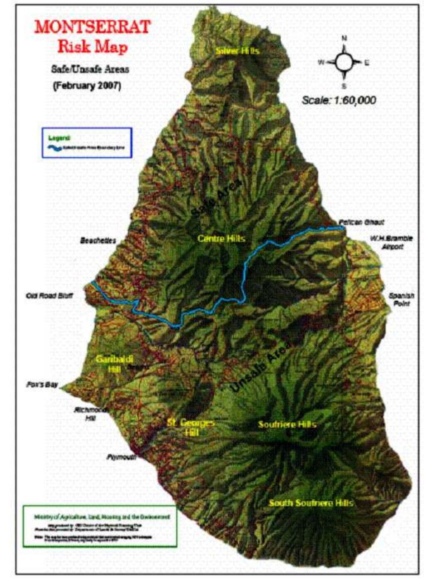
Pyroclastic Flows at Soufrière Hills Volcano

Where *is* Montserrat?

Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

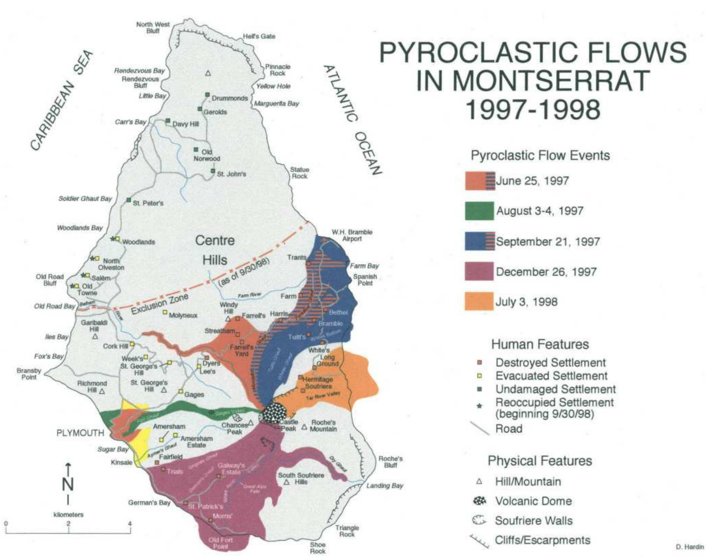
What's there?





Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

Montserrat PFs in One 13-month Period



Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

Plymouth, the former capital

Modelling Hazard

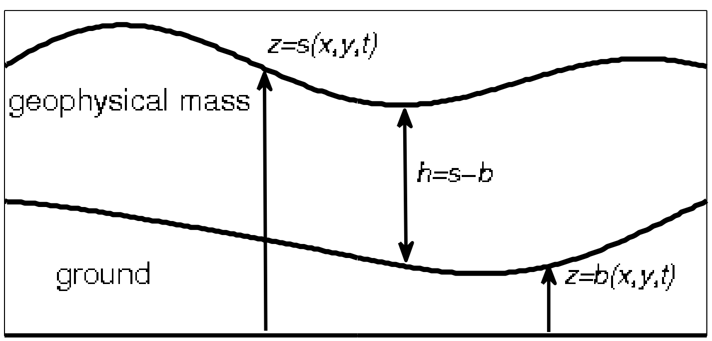
- Goal: quantify the hazard from Pyroclastic Flows (PFs)—
- For specific locations "x" on Montserrat, we would like to find the probability of a "catastrophic event" (inundation to $\geq$1m) within T years— for $T = 1$, $T = 5$, $T = 10$, $T = 20$ years
- Reflecting all that is uncertain, including:
  - How often will PFs of various volumes occur?
  - What initial direction will they go?
  - How will the flow evolve?
  - How are things changing, over time?
- This is *not* what is needed for short-term crisis and event management— here we consider only long-term hazard.

Pieces of the Puzzle

Our team (geologist/volcanologist, applied mathematicians, statisticians, stochastic processes) breaks the problem down into two parts:

- ▶ Try to model the volume, frequency, and initial direction of PFs at SHV, based on 15 years' data;
- ◀ Tools used: Probability theory, extreme value statistics.
- ▶ For specific volume V and initial direction θ (and other uncertain things like the basal friction angle ϕ), try to model the evolution of the PF, and max depth M_x at location x;
- ◀ Tools used: Titan2D PDE solver, emulator (see below).

Titan2D



- Use 'thin layer' modeling $\rightsquigarrow$ system of PDEs for flow depth and depth-averaged momenta
- Incorporates topographical data from DEM
- Consumes 20m–2hr of computing time (in parallel on 16 processors) for each run— say, about 1 hr.

Titan2D (cont)

TITAN2D (U Buffalo) computes solution to the PDE s with:

- Stochastic inputs whose randomness is the basis of the hazard uncertainty:

  $V =$ initial volume (initial flow magnitude, in m^3),
  $\theta =$ initial angle (initial flow direction, in radians).

- Deterministic inputs:

$\phi =$	basal friction	(deg)	important & very uncertain
$\psi =$	internal friction	(deg)	less important, ignored
$v_0 =$	initial velocity	(m/s)	less important, set to zero

- Output: flow height and depth-averaged velocity at each of thousands of grid points at every time step. We will focus on the maximum flow height at a few selected grid points.
- Each run takes about 1 hour on 16 processors

Hazard Assessment I: What's a Catastrophe?

- Let $M_x(z)$ be the computer model prediction with arbitrary input parameter vector $z \in \mathcal{Z}$ of whatever characterizes a catastrophe at $x \in \mathcal{X}$.

SHV: *$z = (V, \theta, \phi) \in \mathcal{Z} = (0, \infty) \times [0, 2\pi) \times (0, 90)$, $M_x(V, \theta, \phi)$ = max PF height at location x (downtown Plymouth, Bramble Airport, Dyers RV TP4) for PF of characteristics $z = (V, \theta, \phi)$.*

- Catastrophe occurs for z such that $M_x(z) \in \mathcal{Y}_C$.

SHV: *Catastrophe if $M_x(z) \geq 1\,m$ (Other options...)*

- Determine 'catastrophic region' $\mathcal{Z}_c$ in the input space:

$$\mathcal{Z}_c = \{z \in \mathcal{Z} : \; M_x(z) \in \mathcal{Y}_C\}$$

SHV: *$\mathcal{Z}_c = \{z : \; V > \Psi(\theta, \phi)\}$ where...*

Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

Which inputs are catastrophic at SHV?

By continuity and monotonicity,

$$\begin{aligned}\mathcal{Z}_c &= \{z \in \mathcal{Z} : \; M_x(z) \in \mathcal{Y}_C\} \\ &= \{(V, \theta, \phi) \in \mathcal{Z} \; : \; V > \Psi(\theta, \phi)\}\end{aligned}$$

for the *critical contour* Ψ, where

$$\Psi(\theta, \phi) \equiv \{\text{value of } V \text{ such that } M_x(V, \theta, \phi) = 1\,\text{m}\}$$

Now let's turn to the problem of evaluating Ψ.

Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

Emulation

To find Ψ, we'd like to evaluate $M_x(V_i, \theta_i, \phi_i)$ by running Titan2D for selected locations $x \in \mathcal{X}$ and at each of perhaps:

- ~ 100 Volumes V_i;
- ~ 100 Initiation Angles θ_j;
- ~ 100 Basal Friction Angles ϕ_k;

Which would entail maybe $100 \times 100 \times 100 = 1\,000\,000$ runs of Titan2D... but we don't have $1\,000\,000$ hours.
Our Solution:

Build an Emulator for our PDE flow model.

Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

Emulators

An Emulator is a (very fast):

- statistical model (based on Gaussian Processes) for our
- computer model (based on PDE solver) of the
- volcano.

- The emulator can predict (in seconds) what Titan2D would return (in hours), *with* an estimate of its accuracy;
- Based on Gaussian Stochastic Process (GaSP) Model. We:
  - Pick a few hundred LHC "design points" (V_j, θ_j, ϕ_j, ...);
  - Run Titan2D at each of them to find Output $M_x(V_j, \theta_j, \phi_j)$;
  - Train the GaSP to return a statistical estimate

$$\mathsf{E}\big[M_x(V, \theta, \phi) \mid \{M_x(V_j, \theta_j, \phi_j)\}\big]$$

of model output M for site x at untried points (V, θ, ϕ).

Why?

Our immediate goal is to find a "threshhold function" for each location of concern x in Montserrat:

$$\Psi(\theta,\phi) = \text{Smallest volume } V \text{ that would inundate } x \text{ if flow begins in direction } \theta \text{ with friction angle } \phi$$

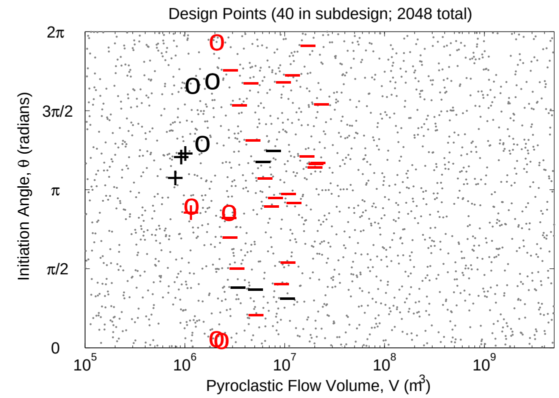
for each possible direction θ (0–360 in degrees, or 0–2π radians, with 0= due East and $\pi/2$= due North) and basal friction angle ϕ.
Simplification: Take $\phi_i \equiv \hat{\phi}(V_i)$ (empirical; see below).
Then we can quantify the hazard at x for T years as

$$\Pr\left\{V_i \geq \Psi(\theta_i) \text{ for some PF } (V_i,\theta_i) \text{ in time } (t_0, t_0+T]\right\}$$

which in turn we study with the probability models.

Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

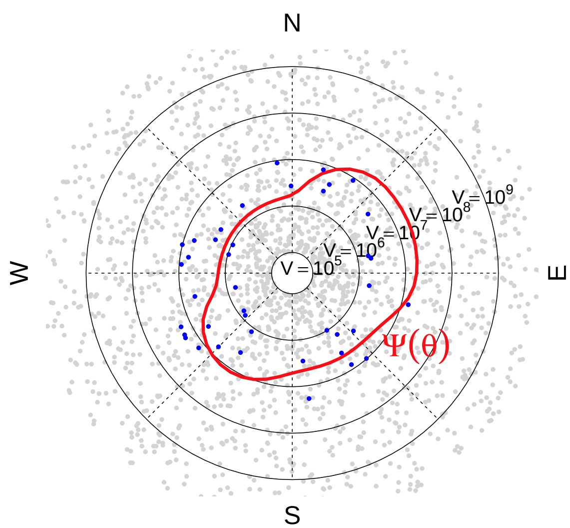
Design points V, θ



2048 Design Points total; 40 used for determining Ψ at x=
Level 4 Trigger Point: Dyers River Valley (head of Belham valley) .
Black $\iff M_x(V,\theta,\phi)=0$, Red $\iff M_x(V,\theta,\phi)>0$.

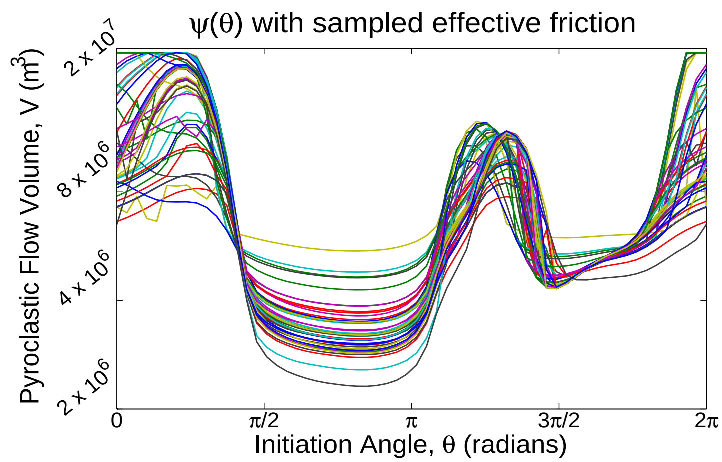
Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

Polar View of Design, Subdesign, & Ψ



Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

50 Simulations of $\Psi(\theta)$ for Level 4 Trigger Point, Dyers RV



Some Details about our Emulator...

Emulators are very fast *approximations* for (some of) the outputs $M_x(z)$ of (slow) computer models. They are used for many purposes (design, optimization, inference, sensitivity analyses, ...) for expensive computer models.

- Begin with a Maxi-Min Latin Hypercube statistical design to select some number N of design points z_i in the large region $\mathcal{Z} = [10^5, 10^{9.5}]\mathrm{m}^3 \times [0, 2\pi)\mathrm{rad} \times [5, 25)\mathrm{deg}$.
- Run the slow computer model $M_x(z_i)$ at these N preliminary points.
- For fitting an emulator to find $\mathcal{Z}_c$, keep only design points $\mathcal{D}$ in a region 'close' to the boundary $\partial\mathcal{Z}_c$:

Too many details? Skip ahead 5 or 6 frames...

Gaussian Process Emulators in the Region $\mathcal{Z}_c$

- Since we are interested in regions where the flow is small (1m), fit an emulator to $\tilde{M}_x(z) \equiv \log(1 + M_x(z))$. Let $\tilde{\mathbf{y}}$ be the transformed vector of computer model runs $\tilde{M}_x(z)$ for $z \in \mathcal{D}$.
- Model the unknown $\tilde{M}(z)$ as a Gaussian process

$$\tilde{M}(z) = \beta + mV + Z(V, \theta, \phi)$$

(note that we expect a monotonic trend in V, but not θ; ϕ discussed later), where $Z(V, \theta, \phi)$ is a stationary GP with

- Mean 0, Variance σ_z^2;
- Product exponential correlation: *i.e.*, $(\forall z_i = (V_i, \theta_i, \phi_i) \in \mathcal{D})$, the correlation matrix $\mathbf{R}$ is:

$$R_{ij} = \exp\left\{ -\left|\frac{V_i - V_j}{\rho_V}\right|^{\alpha_V} - \left|\frac{\theta_i - \theta_j}{\rho_\theta}\right|^{\alpha_\theta} - \left|\frac{\phi_i - \phi_j}{\rho_\phi}\right|^{\alpha_\phi} \right\}$$

with range parameters $\rho_\bullet$, smoothness parameters $\alpha_\bullet$.

Handling the Unknown Hyperparameters $\boldsymbol{\theta} = (\rho_V, \rho_\theta, \rho_\phi, \alpha_V, \alpha_\theta, \alpha_\phi, \sigma_z^2, \beta, m)$

- Deal with the crucial parameters (σ_z^2, β, m) via a fully Bayesian analysis (here an extension of Kriging) using objective priors: $\pi(\beta) \propto 1$, $\pi(m) \propto 1$, and $\pi(\sigma_z^2) \propto 1/\sigma_z^2$;
- Compute the marginal posterior mode, $\hat{\boldsymbol{\xi}}$, of $\boldsymbol{\xi} = (\rho_V, \rho_\theta, \rho_\phi, \alpha_V, \alpha_\theta, \alpha_\phi)$ using the above priors and the reference prior for $\boldsymbol{\xi}$; then $\hat{\mathbf{R}} \equiv \mathbf{R}(\hat{\boldsymbol{\xi}})$ is completely specified (big simplification— no matrix decomp inside MCMC loop).
  - A fully Bayesian analysis, accounting for uncertainty in $\hat{\boldsymbol{\xi}}$, is difficult and rarely affects the final answer significantly because of confounding of variables.

The Posterior Mode of $\boldsymbol{\xi} = (\rho_V, \rho_\theta, \rho_\phi, \alpha_V, \alpha_\theta, \alpha_\phi)$

- MLE fitting of $\boldsymbol{\xi}$ has enormous problems; we've given up on it.
- A big improvement is finding the *marginal* MLE of $\boldsymbol{\xi}$ from the marginal likelihood for $\boldsymbol{\xi}$, available by integrating lh wrt the objective prior $\pi(\beta, m, \sigma_z^2) = 1/\sigma_z^2$. The expression is:

$$L(\boldsymbol{\xi}) \propto |\mathbf{R}(\boldsymbol{\xi})|^{-\frac{1}{2}} \left|\mathbf{X}'\mathbf{R}(\boldsymbol{\xi})^{-1}\mathbf{X}\right|^{-\frac{1}{2}} (S^2(\boldsymbol{\xi}))^{-(\frac{n-q}{2})}, \text{ where}$$

  - $\mathbf{X} = [\mathbf{1}, \mathbf{V}]$ is the design matrix for the linear parameters, *i.e.*, $\mathbf{1}$ is the column vector of ones and $\mathbf{V}$ is the vector of volumes $\{V_i\}$ in the data set, and $\boldsymbol{\mu} = (\beta, m)$ (of dimension $q = 2$);
  - $S^2(\boldsymbol{\xi}) = [\tilde{\mathbf{y}} - \mathbf{X}\hat{\boldsymbol{\mu}}]'\mathbf{R}(\boldsymbol{\xi})^{-1}[\tilde{\mathbf{y}} - \mathbf{X}\hat{\boldsymbol{\mu}}]$;
  - $\hat{\boldsymbol{\mu}} = [\mathbf{X}'\mathbf{R}(\boldsymbol{\xi})^{-1}\mathbf{X}]^{-1}\mathbf{R}(\boldsymbol{\xi})^{-1}\tilde{\mathbf{y}}$.

- An even bigger improvement arises by finding the posterior mode from $L(\boldsymbol{\xi})\pi^R(\boldsymbol{\xi})$, where $\pi^R(\boldsymbol{\xi})$ is the reference prior for $\boldsymbol{\xi}$ (Paulo, 2005 AoS). Note that it is computationally expensive to work with the reference posterior in an MCMC loop, but using it for a single maximization to determine the posterior mode is cheap.
  The reference prior for $\boldsymbol{\xi}$ is $\pi^R(\boldsymbol{\xi}) \propto |I^\star(\boldsymbol{\xi})|^{1/2}$, where

$$I^\star(\boldsymbol{\xi}) = \begin{pmatrix} (n-q) & \mathrm{tr}\mathbf{W}_1 & \mathrm{tr}\mathbf{W}_2 & \cdots & \mathrm{tr}\mathbf{W}_p \\ & \mathrm{tr}\mathbf{W}_1^2 & \mathrm{tr}\mathbf{W}_1\mathbf{W}_2 & \cdots & \mathrm{tr}\mathbf{W}_1\mathbf{W}_p \\ & & \ddots & \cdots & \vdots \\ & & & & \mathrm{tr}\mathbf{W}_p^2 \end{pmatrix}$$

$$\mathbf{W}_k = \left(\frac{\partial}{\partial \xi_k}\right) \mathbf{R}(\boldsymbol{\xi})^{-1} \left\{ \mathbf{I}_n - \mathbf{X}\left[\mathbf{X}'\mathbf{R}\boldsymbol{\xi}\right]^{-1}\mathbf{X} \right\}^{-1} \mathbf{X}'\mathbf{R}(\boldsymbol{\xi})^{-1},$$

with $q = 2$ the dimension of $\boldsymbol{\mu}$ and $p = 3$ the dimension of $\boldsymbol{\xi}$.

The posterior distribution of (σ_z^2, β, m), conditional on $\widetilde{\mathbf{y}}$ and $\hat{\boldsymbol{\xi}}$, yields the final emulator (in transformed space) at input z^*:

$$\tilde{M}_x(z^*) \mid \widetilde{\mathbf{y}},\ \hat{\boldsymbol{\xi}} \ \sim\ t(\,y^*(z^*),\ s^2(z^*),\ N-2)\,,$$

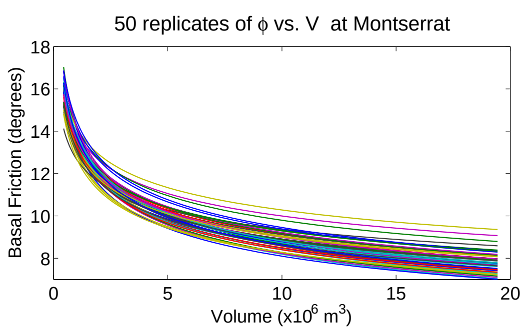
noncentral t-distribution with $N-2$ degrees of freedom and location & scale parameters:

$$y^*(z^*) = \mathbf{r}^T\mathbf{R}^{-1}\widetilde{\mathbf{y}} + \frac{\mathbf{1}^T\mathbf{R}^{-1}\widetilde{\mathbf{y}}}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}}[1 - \mathbf{r}^T\mathbf{R}^{-1}\mathbf{1}] + \frac{\widetilde{\mathbf{V}}^T\mathbf{R}^{-1}\widetilde{\mathbf{y}}}{\widetilde{\mathbf{V}}^T\mathbf{R}^{-1}\widetilde{\mathbf{V}}}[\widetilde{V}^* - \mathbf{r}^T\mathbf{R}^{-1}\widetilde{\mathbf{V}}]$$

$$s^2(z^*) = \left[(1 - \mathbf{r}^T\mathbf{R}^{-1}\mathbf{r}) + \frac{(1 - \mathbf{r}^T\mathbf{R}^{-1}\mathbf{1})^2}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}} + \frac{(\widetilde{V}^* - \mathbf{r}^T\mathbf{R}^{-1}\widetilde{\mathbf{V}})^2}{\widetilde{\mathbf{V}}^T\mathbf{R}^{-1}\widetilde{\mathbf{V}}}\right] \times \frac{1}{N-2}\left[\widetilde{\mathbf{y}}^T\mathbf{R}^{-1}\widetilde{\mathbf{y}} - \frac{(\mathbf{1}^T\mathbf{R}^{-1}\widetilde{\mathbf{y}})^2}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}} - \frac{(\widetilde{\mathbf{V}}^T\mathbf{R}^{-1}\widetilde{\mathbf{y}})^2}{\widetilde{\mathbf{V}}^T\mathbf{R}^{-1}\widetilde{\mathbf{V}}}\right],$$

where $\widetilde{V}_i = V_i - V_R$, $\widetilde{V}^* = V^* - V_R$, $V_R = \mathbf{1}^T\mathbf{R}^{-1}\mathbf{V}/\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}$, and $\mathbf{r}^T = (R(z^*, z_1), \ldots, R(z^*, z_N))$. Tedious, but tractable.

Theoretically the Basal Friction ϕ *should* be constant, but empirically it is highly dependent on the Volume V:



Currently we simply estimate the function $\phi(V)$, and replace ϕ in the emulator by this function. The emulator thus becomes only a function of (V, θ). We are now moving to a full 3-dimensional $\mathcal{Z}_c$.

With this short-cut, $\Psi(z)$ depends on only one quantity: θ.

Response Surfaces

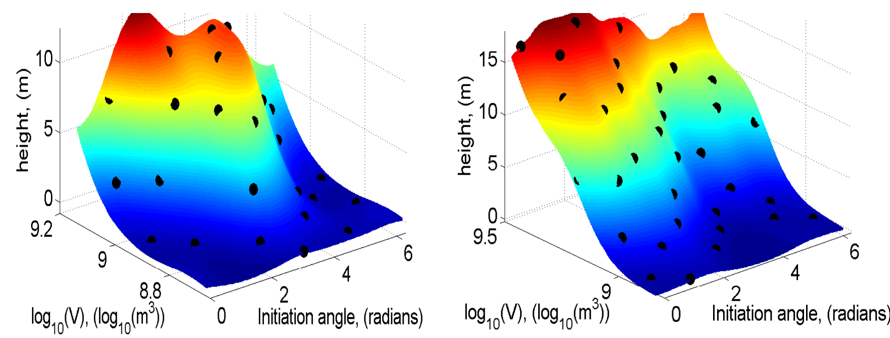


Figure: Median of the emulator, transformed back to the original space.
Left: Plymouth, Right: Bramble Airport.
Black points: max-height simulation outputs at subdesign points $\mathcal{D}_c$.

Catastrophic event contours Ψ (response surface slices at ht. 1 m)

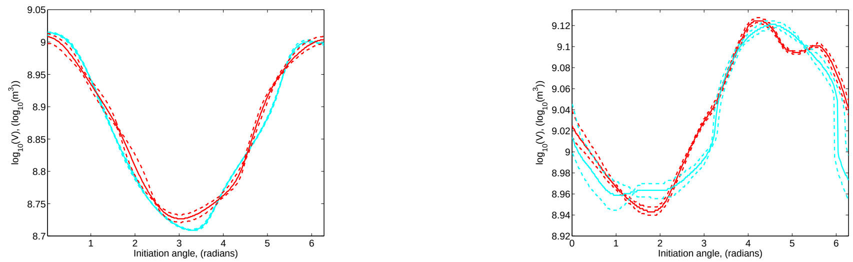


Figure: Left: Plymouth, Right: Bramble Airport.
Solid Cyan: linear max-height. Solid Red: log transformation.
Dashed: 75% pointwise confidence bands for Ψ.

Adapting the design

- We added new design points near the boundary $\partial\mathcal{Z}_c$ of the critical region where:
  - contours $\Psi(z)$ pass between design points z_i with $M_x(z_i) = 0$ and z_j with $M_x(z_j) \gg 1$; or
  - the confidence bands for $\Psi(z)$ are widest.
- The computer model was re-run at these new design points.
- The emulator was then re-fit and and critical contour Ψ was re-computed.
- Median contours Ψ did not change much, but confidence bands for Ψ were much narrower, so it was judged that no further adaptation was needed.

(For other adaptive designs see R.B. Gramacy *et al.* (ICML, 2004), P. Ranjan *et al.* (Technometrics, 2008), B.J. Williams *et al.* (Stat Sin, 2000))

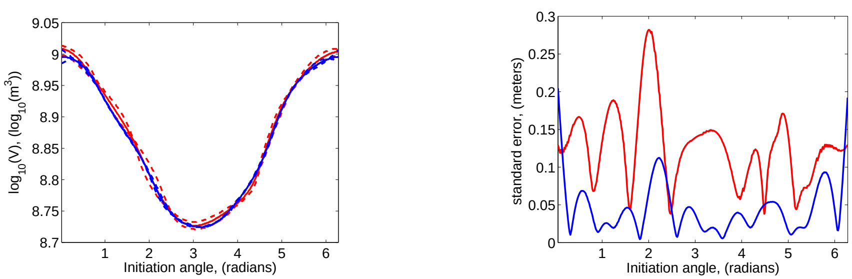


Figure: Left: $\Psi(\theta)$ (solid) with implicit standard error curves (dashed), Right: standard error; Red: original design. Blue: updated design.

The critical region $\mathcal{Z}_c$ of inputs (V, θ) producing a catastrophic event is the region above the critical contour Ψ.

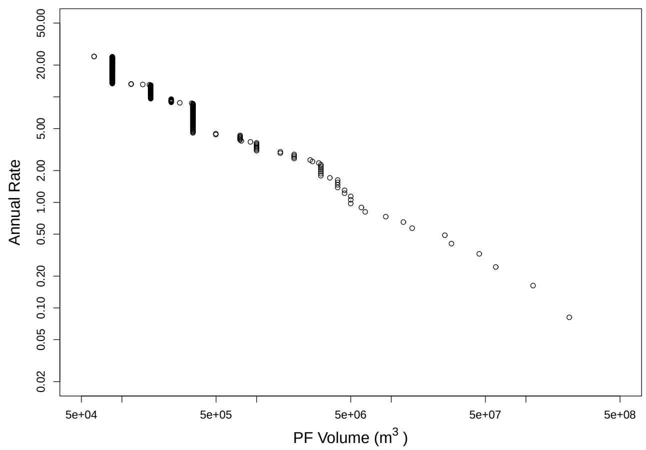
Hazard Assessment II: Probability of Catastrophe

For us a PF is *catastrophic* if its volume V exceeds an uncertain threshold $\Psi(\theta)$ that depends on the initiation angle θ.
For a hazard summary we wish to report, for each $T > 0$,

$$\Pr[\text{ Catastrophe at } x \text{ within } T \text{ Years }] = \Pr[\{(V_i, \theta_i, \tau_i)\} : \; V_i > \Psi(\theta_i), \; \tau_i \le T \,]$$

SO, we need a joint model for points $\{(V_i, \theta_i, \tau_i)\} \subset \mathbb{R}^3$.
Let's do it in that order: first Volumes, then Angles, then Times.

PF Volume *vs.* Frequency



Seems kind of linear, on log-log scale...

Linear log-log plots of Magnitude *vs.* Frequency are a hallmark of the Pareto probability distribution $\text{Pa}(\alpha, \epsilon)$,

$$\text{P}[V > v] = (v/\epsilon)^{-\alpha}, \qquad v > \epsilon.$$

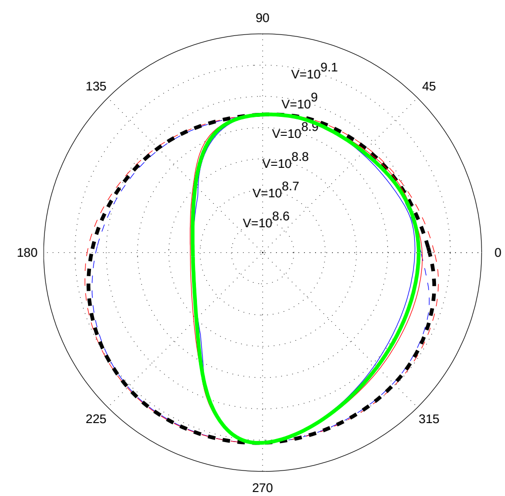
Which is kind of bad news.

The Pareto Distribution

The negative slope seems to be about $\alpha \approx 0.64$ or so.
The Pareto distribution with $\alpha < 1$ has:

- Heavy tails;
- Infinite mean $\text{E}[V] = \infty$, infinite variance $\text{E}[V^2] = \infty$;
- No Central Limit Theorem for sums (skewed α-Stable);
- Significant chance that, in the future, we will see volumes larger than any we have seen in the past. Like $V > 10^9 \text{m}^3$.
- The Pareto comes up all the time in the Peaks over Threshhold (PoT) approach to the Statistics of Extreme Events— related to Fisher/Tippett Three Types Theorem.

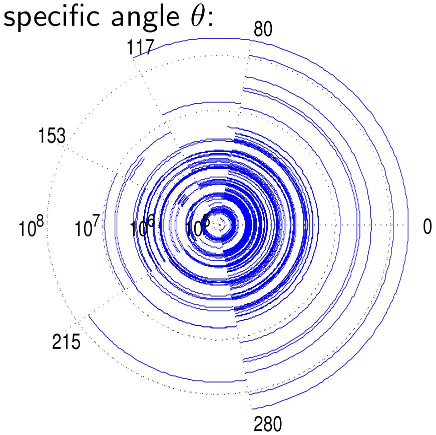
How big is bad?



Kinda depends on which direction it goes...

PF Initiation Angles

Our data on angles is quite vague— we only know which of 7 or 8 *valleys* were reached by a given PF, from which we can infer a *sector* but not a specific angle θ:



Any nonuniformity for θ? Any dependence of θ on Volume V?

Angle/Volume Data (cont)

We need a *joint* density function for V and θ. Without much evidence against independence, we use product pdf:

$$V, \theta \sim \alpha\epsilon^{\alpha} V^{-\alpha-1} \mathbf{1}_{\{V>\epsilon\}} \pi_\kappa(\theta)$$

where $\pi_\kappa(\theta)$ is the pdf for the von Mises vM(μ, κ) distribution,

$$\pi_\kappa(\theta) = \frac{e^{\kappa \cos(\theta-\mu)}}{2\pi I_0(\kappa)},$$

centered at $\theta \approx \mu$ close to zero (East) with concentration κ that might depend on V if the data support that.

PF Times

If we observe λ PFs per year of volume $V > \epsilon$, then what is the probability that such a PF will occur in the next 24 hours?
For *short-term* predictions, it may be important to note this can depend on many things, such as:

- How high is the dome just now?
- Any seismic activity suggesting dome growth and instability?
- Has it rained recently?
- How long since last PF?

But, for *long-term* predictions all these factors average out and we assume that:

- PF occurrences in disjoint time intervals are *independent*
- PF rates are constant over time, neither rising nor falling.

PF Times (cont)

Under those assumptions (which we will revisit), the number of PFs in any time interval has a Poisson Distribution, with mean proportional to its length.
For an interval of length $\Delta t = 1/365$, a single day, the expected number of PFs is $\lambda\Delta t$ and so the *probability* of

$$\text{P[At least one PF during time } \Delta t \text{]} = 1 - \exp(-\lambda\Delta t)$$

Or about $1 - e^{-22/365} \approx 5.8\%$ for one day with the SHV data from 1995–2010.

A Summary of our Stochastic Model:

The data suggest a (provisional) model in which:

1. PF Volumes are iid from a Pareto $V \sim \mathsf{Pa}(\alpha, \epsilon)$ distribution for some shape parameter $\alpha(\approx 0.63)$ and minimum flow ϵ $(\approx 5 \cdot 10^4\ \mathrm{m}^3)$; and
2. PF Initiation Angles have a von Mises $\theta \sim \mathsf{vM}(\mu, \kappa)$ distribution on $[0, 2\pi)$ with $\mu \approx 0$ and $\kappa \approx 0.4$; and
3. PF Arrival Times follow a stationary Poisson process at some rate λ (≈ 22)/yr.

These have the beautiful and simplifying consequence that the number of PFs in *any* region of three-dimensional $(V \times \theta \times \tau)$ space has a Poisson probability distribution— so, we can evaluate:

Hazard

P[Catastrophy within T Years]

$$\begin{aligned} &= 1 - \mathsf{P}[Y_T = 0] \qquad \text{(where } Y_T \text{ is the } \textit{number} \text{ of catastrophes)} \\ &= 1 - \exp\big(-\mathsf{E}Y_T\big) \\ &= 1 - \exp\left(-\lambda T\, \epsilon^{\alpha} \int_0^{2\pi} \Psi(\theta)^{-\alpha}\, \pi_\kappa(d\theta)\right) \end{aligned}$$

Which we can compute pretty easily on a computer.

Accommodating uncertainty in λ, κ, α, *etc.* is easy in Objective Bayesian statistics— we use Reference Prior distributions, and simulation-based methods (MCMC) to evaluate the necessary integrals.

Posterior distribution of (α, λ)

For a given minimum volume $\epsilon > 0$ and period $(0, t]$, the sufficient statistics and Likelihood Function are

$$\begin{aligned} J &= \text{Number of PF's in } (0, t], \text{ and} \\ S &= \sum \log(V_j), \text{ the log-product of their volumes} \\ L(\alpha, \lambda) &\propto (\lambda\, \alpha)^J \exp\left\{-\lambda t\, \epsilon^{-\alpha} - \alpha\, S\right\} \end{aligned}$$

Objective Priors:

- Jeffreys prior is $\pi_J(\alpha, \lambda) \propto |I(\alpha, \lambda)|^{1/2} \propto \alpha^{-1}\epsilon^{-\alpha}$;
- Reference priors:
  - α of interest: $\pi_{R1}(\alpha, \lambda) \propto \lambda^{-1/2}\alpha^{-1}\epsilon^{-\alpha/2}$
  - λ of interest: $\pi_{R2}(\alpha, \lambda) \propto \lambda^{-1/2}[\alpha^{-2} + (\log\epsilon)^2]^{1/2}\epsilon^{-\alpha/2}$, which is also Jeffreys' independent prior.

Posterior: $\pi(\alpha, \lambda \mid \text{data}) \propto L(\alpha, \lambda)\, \pi(\alpha, \lambda)$, quite tractable.

Computing the probability of catastrophe

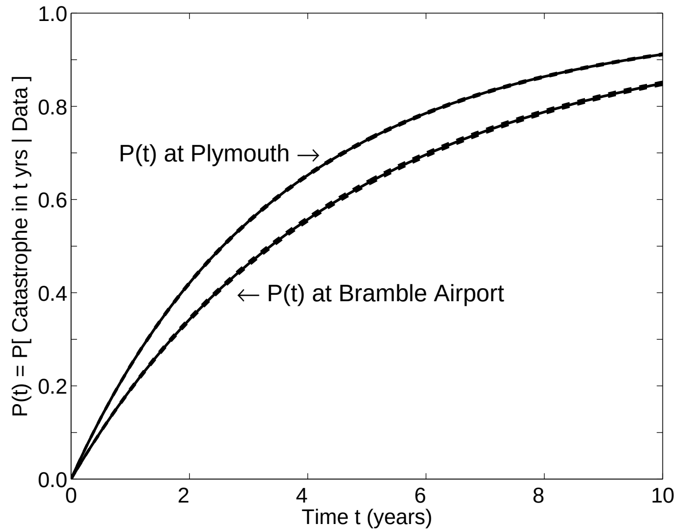
To compute Pr(at least one catastrophic event in T *years* | data) for a range of T, an importance sampling estimate is

$$P(T) \cong 1 - \frac{\sum_i \exp\left[-\lambda_i\, T\, \epsilon^{\alpha}\, \widehat{\Psi}(\alpha_i)\right] \frac{\pi^*(\alpha_i, \lambda_i)}{g(\alpha_i, \lambda_i)}}{\sum_i \frac{\pi^*(\alpha_i, \lambda_i)}{g(\alpha_i, \lambda_i)}},$$

where

- $\widehat{\Psi}(\alpha)$ is an MC estimate of $\int_0^{2\pi} \Psi(\theta)^{-\alpha}\, \pi_\kappa(d\theta)$ based on draws $\theta_i \sim \mathsf{vM}(\mu, \kappa)$ (or use quadrature);
- $\pi^*(\alpha, \lambda)$ is the un-normalized posterior density;
- $\{(\alpha_i, \lambda_i)\}$ are iid draws from the importance sampling density $g(\alpha, \lambda) = t_2(\alpha, \lambda \mid \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}, 3)$, with d.f. 3, mean $\widehat{\boldsymbol{\mu}}^t = (\hat{\alpha}, \hat{\lambda})$, and scale $\widehat{\boldsymbol{\Sigma}}$ = inverse of observed Information Matrix.

Hazard Over Time at Plymouth & Bramble



How are things at Bramble Airport?
Not so good... here's the runway...

Discussion

We have argued that:

- Hazard assessment of catastrophic events (in the absence of lots of extreme data) requires
  - Mathematical computer modeling to support extrapolation beyond the range of the data;
  - Statistical modeling of available (possibly not extreme) data to determine input distributions;
  - Statistical development of emulators of the computer model to determine critical event contours.
- Major sources of uncertainty can be combined and incorporated with Objective Bayesian analysis.

Ongoing Work & Extensions:

- Extend the methodology to create entire Hazard Maps (*i.e.*, find hazard for all locations x simultaneously).
- Go beyond stationarity with
  - Change-point model for intensity λ_t;
  - Model (heavy-tailed) duration of activity;
  - Model caldera evolution (μ, κ for vM).
- Reflect uncertainty and change in DEM s.

A Collaborative Effort...

Introduction Hazard Assessment I: Defining a Catastrophe Hazard Assessment II: Probability of a Catastrophe Discussion

Thanks— to Organizers and Collaborators!

For more, see: www.stat.duke.edu/∼rlw/ or www.mvo.ms