

Using Statistical and Computer Models to Quantify Volcanic Hazards

M. J. Bayarri James O. Berger Eliza S. Calder Keith Dalbey
Simon Lunagomez Abani K. Patra E. Bruce Pitman
Elaine T. Spiller Robert L. Wolpert *

July 14, 2009

Abstract

Risk assessment of rare natural hazards— such as large volcanic block and ash or pyroclastic flows— is addressed. Assessment is approached through a combination of computer modeling, statistical modeling, and extreme-event probability computation. A computer model of the natural hazard is used to provide the needed extrapolation to unseen parts of the hazard space. Statistical modeling of the available data is needed to determine the initializing distribution for exercising the computer model. In dealing with rare events, direct simulations involving the computer model are prohibitively expensive. Solution instead requires a combination of adaptive design of computer model approximations (emulators) and rare event simulation. The techniques that are developed for risk assessment are illustrated on a test-bed example involving volcanic flow.

Keywords and phrases: Bayesian analysis; Catastrophic events; Emulators; Extreme events; Inverse Problems.

1 INTRODUCTION

The focus of this paper is on assessing the risk of extreme natural hazards, such as volcanic pyroclastic flows large enough to devastate a nearby city. The scenario we focus upon is that in which the potentially catastrophic events $\{E_i\}$ occur with some frequency, for example, hurricanes, tsunamis, floods, earthquakes, forest fires, volcano eruptions, etc., but in which truly catastrophic consequences at any specific time and place are rare. We then wish to determine the probability that at least one catastrophic event C will occur in the next T years— for instance, the probability of a pyroclastic flow event in the next T years that results in a significant amount of debris reaching a designated location (e.g., the center of a town, the location of a hospital or airport, etc.) and hence causing catastrophic damage.

*Affiliations: Bayarri — Universitat de València; Berger, Lunagomez and Wolpert — Duke University; Calder, Dalbey, Patra and Pitman — University of Buffalo; Spiller — Marquette University.

This focus on rare events typically precludes the use of purely data-based approaches to the problem; there will not be enough data concerning the rare events to provide a sound statistical answer. It is common to employ expert opinion to help assess the risk of extreme events, but these phenomena are so complex that expert opinion can be quite uncertain.

An increasingly popular more formal approach to the problem is to construct a deterministic computer implementation of a mathematical model for the phenomena that produce the catastrophic events. This computer model can be run under a range of natural conditions that are possible, but haven't yet been observed, in order to aid in risk assessment. As one example of the use of such computer models in evaluating hurricane risk, see Iman et al. (2005).

In this paper we focus on the particular problem of assessing the risk from volcanic activity, employing a computer model— TITAN2D— developed for modeling the process of volcanic flow. Given a digital elevation map specifying the topography of a volcano and the surrounding area, and the values of input parameters including the initial volume of erupted material and the direction at which the discharged mass begins to move, TITAN2D can predict the thickness of a volcano pyroclastic flow at any location (such as the center of a given town) at any time during the flow (Patra et al. 2005). If the flow thickness is large enough (say, one meter deep) at any time during the simulation, we will call that flow a catastrophic event. See Section 2 for a brief introduction to this computer model and its capabilities for predicting volcanic flow.

While the development of a mathematical model of a rare phenomenon is one fundamental step in formal risk assessment, there are two other key advances that are needed. The first is to determine the input distributions needed to exercise the computer model; e.g., what are the volumes of volcanic flow that can be expected from a given volcano? In determining such input distributions, availability of data concerning the geophysical process being modeled is crucial, as is the development of suitable statistical models of the data. In Section 4 we propose a method for determining such input distributions, based on scarce and challenging data of the type that is typical for volcanic flows.

The final challenge is that of computing the probability of a catastrophic event, since such events are typically very rare. Direct simulation of catastrophic events— utilizing the input distribution and the computer model— is generally hopeless, because of the expense (time needed) to run the computer model. To surmount this final challenge we utilize adaptively designed emulators (approximations) for the computer model, to identify the *threshold* inputs that define a catastrophic event. This process is described in Section 3. It is then possible to compute the probability of the catastrophic event using more standard computational techniques, as demonstrated in Section 5.

The test site used in the paper and the computations is the Soufrière Hills Volcano (often abbreviated *SHV*) on the island of Montserrat. While we focus on this volcanic application, the methodology we propose is immediately applicable to other dome-forming volcanoes (Mount St. Helens, Merapi, and others); indeed, there is considerable interest in understanding such lava dome eruptions (Voight et al. 1999). Moreover, the general approach taken is, in principle, applicable to any hazard situation for which there is an available computer model of the natural process and there exists a reasonable amount of data (but not necessarily of extreme events). The recent development and emerging accessibility of new remote sensing techniques is leading to rapidly growing classes of problems suitable for

analysis using the methods described here.

2 THE COMPUTER MODEL

2.1 The Geophysical/Mathematical/Computer Model

Geophysical mass flows include pyroclastic flows, rockfalls and granular avalanches. These flows often arise consequent to volcanic activity, and they pose serious challenges to mathematical modeling and risk assessment. In particular, they activate physics across six orders of magnitude, from centimeter-sized particles to kilometer-sized runouts, and depend on the terrain and geography over which they run. These mass flows may contain 10^4 – 10^9m^3 or more of solids material, and travel at tens of meters per second, leaving deposits that may be tens of meters deep and several kilometers long. Needless to say, these flows can have potentially devastating effects.

The enormous span of scales presents significant challenges to those developing models of the flow process. Although models of rolling and sliding grains traversing a pile of otherwise stationary grains have gained popularity, these models have not been able to provide quantifiable predictions in systems of engineering interest. Instead, research characterizes geophysical mass flows via ‘thin layer’ modeling— that is, by recognizing the long runout of these flows compared to their thickness, and scaling accordingly. The resulting mathematical system, akin to the shallow water equations for fluids, consists of a system of partial differential equations for the flow depth and the depth-averaged momenta (Savage and Hutter 1989; Iverson 1997; Patra et al. 2005). These models provide reasonable comparisons with experimental findings, using terrain data and only a few input modeling parameters.

We use a depth-averaged model for dry geophysical mass flows. Depth averaging circumvents the difficult problem of computing the free surface of the flowing mass, allowing the computation of flows that otherwise would be prohibitively expensive. However averaging also blurs any internal structure, such as stratification, that might develop within a flow. As always, the physical assumptions under which a mathematical model is derived must be respected as one applies the model equations.

TITAN2D is a simulation tool developed at the University of Buffalo for computing solutions to this mathematical model. Validation of the computer code and verification of the results have been performed and reported in the literature. For example, simulation results from TITAN2D have been tested against field measurements of historical mass flows, and give a good description of important flow features (Sheridan et al. 2005).

A principal feature of TITAN2D is the incorporation of topographical data from GIS (Geographic Information System) sources into the simulation and grid structure. Two other key inputs are $\boldsymbol{x} = (x_1, x_2)$, where x_1 is the size of the initial flow and x_2 the direction in which material first begins to flow (an angle measured from a reference direction). The output— the flow height and a depth-averaged velocity, at every grid point at every timestep— is a complete description of the mass flow over realistic terrain.

TITAN2D is still evolving, particularly as concerns the two constitutive inputs— internal friction angle and basal friction angle— that describe the frictional dissipation of the flowing material. The values of these inputs that were used in the analysis in this paper were 35 degrees and 21 degrees, respectively, values determined from laboratory experiments of

flowing material. It is well known, however, that large mass flows exhibit an effective friction angle that is much smaller than laboratory measured values (Dade and Huppert 1998; Collins and Melosh 2003). Hence, there is an ongoing effort to obtain geophysical data that can be used to better calibrate the friction angles used in TITAN2D. For example, the uncertainty of friction angles has been examined in the context of hazard maps in Dalbey et al. (2008). Acknowledgement of uncertainty in these two inputs also requires a new and considerably more complex analysis, involving new design, emulators, computer experiment, physical data and probability modeling. These improvements are still ongoing work and will be reported elsewhere. Of course, allowing friction angles to depend on flow volumes will likely have a significant effect on the final risk assessments, but the model-emulator methodology developed in this paper is not altered by this future change in the computer model itself. Indeed, computer models are continually undergoing such improvements, and it is important to have risk assessment methodology compatible with a sequence of improving computer models.

2.2 The test bed: Soufrière Hills Volcano

Montserrat is a small island, about 10 km. wide and 16 km. long, in the British West Indies. The Soufrière Hills volcano on Montserrat has been erupting since July 1995 and continues to date. There had been seismic activity for about 3 years prior to the 1995 eruption, and other spells of activity during the twentieth century, but the last volcanic eruption is estimated to be about 300–350 years ago. The recent eruption consists of the extrusion of a sequence of lava domes, piles of viscous magma that, in essence, form steep bulbous plugs on top of volcanic vents. They grow slowly (about $2\text{--}8\text{m}^3/\text{s}$) and can host internal pressures as great as 10 MPa. Lava dome eruptions are notorious for suddenly transforming from a benignly effusive state to a violently explosive one. Sudden removal of these ‘plugs’, either by vertically-driven explosions, or by sudden collapse and spontaneous disintegration as the piles grow and become unstable, can have devastating consequences. These dome collapse events spawn one of the major hazards in volcanology—devastating pyroclastic density currents, which move down the flanks of the volcano at speeds of 60 m/s or more (Calder et al. 1999).

Now in its fourteenth year, the Soufrière Hills eruption is one of the longest-lived dome forming eruptions known. Importantly for us, the Soufrière Hills volcano is one of only a very small number of eruptions that have been intensely monitored. The scientific insights gained during the course of this eruption are providing much of the basis for recent progress in volcanology (e.g., Voight et al. 1999).

Sequential and frequent collapses of the Soufrière Hills lava dome occur as a natural consequence of growth by extrusion of lava, and range from conventional rockfalls, through small pyroclastic flows, to major dome-collapse events involving (to date) up to $2.1 \times 10^8\text{m}^3$ of material (Herd et al. 2005, Fig. 1). Since 1995, the eruption has generated tens of thousands of individual rockfalls, and of the order of 300 pyroclastic flows with runouts exceeding 2 km. On 47 occasions, these pyroclastic flows were associated with large collapse events removing in excess of 10^6m^3 of the dome. More detail about these data can be found in Section 4.

Despite efforts towards understanding the physical mechanisms behind dome collapse at Montserrat, the propensity for lava domes to collapse with little or no apparent warning

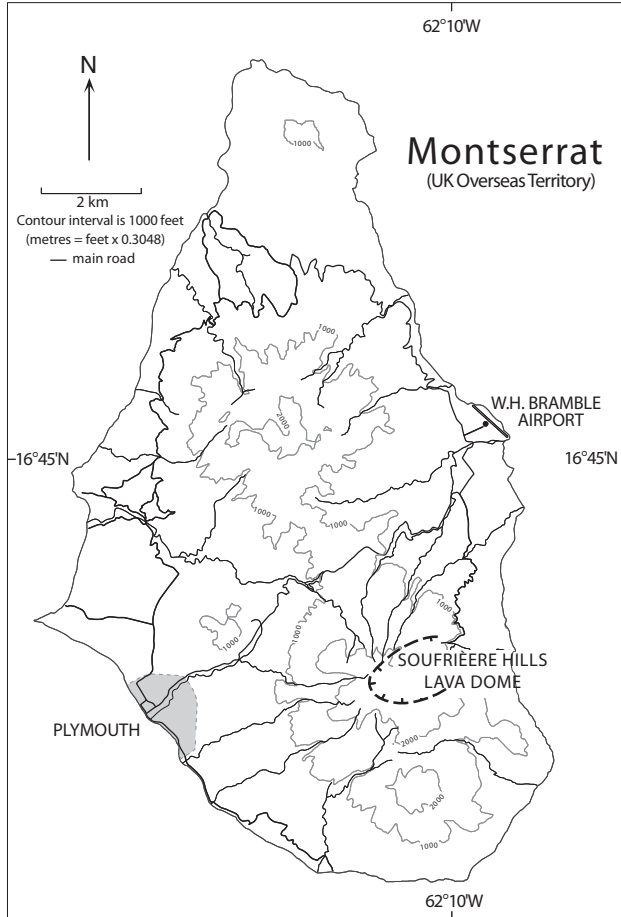


Figure 1: The island of Montserrat. Adapted with permission from (Calder et al. 2002).

remains a serious issue.

3 RISK ASSESSMENT AND EMULATING THE COMPUTER MODEL

3.1 Introduction

We consider problems in which there is a scalar quantity that determines whether or not a particular flow event is “catastrophic.” In this paper, for example, we consider a specified location (e.g., the center of town, or a hospital), and monitor the maximum flow depth at that location throughout the entire flow event as the quantity of interest, to see whether this maximum flow depth exceeds some specified benchmark value. The multivariate situation and the extension to hazard maps of entire areas will be treated elsewhere.

For inputs $\mathbf{x} \in \mathcal{X}$ to the computer model, denote the computer model prediction of the characteristic of interest by $y^M(\mathbf{x})$, and denote by \mathcal{Y}_C the critical set of values of that characteristic regarded as catastrophic; thus input \mathbf{x} results in a catastrophic event if $y^M(\mathbf{x}) \in \mathcal{Y}_C$. Since we are focusing on univariate critical outputs, the critical set \mathcal{Y}_C would typically be an interval $[y_{\min}, y_{\max}]$ or (most often) $[y_{\min}, \infty)$. For the SHV in Montserrat, $\mathbf{x} = (x_1, x_2) =$

(volume, initialization angle) of the volcanic eruption, and $y^M(\mathbf{x}) =$ the maximum height at the center of the target area of a pyroclastic flow from an eruption with characteristics \mathbf{x} . A catastrophic event occurs (by definition for this paper) if \mathbf{x} is such that $y^M(\mathbf{x}) \geq 1\text{m}$.

Note that, although the computer models considered here are deterministic, the model input \mathbf{x} (e.g., flow volume and initial flow angle) will be random, with a distribution that will be determined in Section 4; this will determine the *probability* of a catastrophic event in some specified interval of time.

The naïve approach to determination of a risk probability is simply to generate \mathbf{x} from its input distribution and observe the fraction of time that the resulting $y^M(\mathbf{x}) \in \mathcal{Y}_C$. This is not feasible for typical scenarios involving catastrophic events, both because the events are typically very rare, and because the computer models are extremely expensive to run. For instance, each run of TITAN2D takes approximately one hour on a single processor.

The obvious alternative is to identify the critical set of inputs $\mathcal{X}_C = \{\mathbf{x} \in \mathcal{X} : y^M(\mathbf{x}) \in \mathcal{Y}_C\}$, and then attempt to compute the probability of \mathcal{X}_C under the input distribution. Alas, even the precise determination of \mathcal{X}_C in problems such as this is typically infeasible, as it would require numerically solving for a surface (the boundary of \mathcal{X}_C) using the computer model, which is prohibitively expensive.

A way forward is to develop an *emulator* approximation to the computer model, and use the emulator to find an approximation to \mathcal{X}_C . An emulator is a statistical predictor of the model output (which we call the “simulator”) at untried values of the inputs. Typically, emulator predictions pass through all points at which the simulator has been exercised, and interpolate in between while providing an estimate of the error incurred. Here we use a Gaussian process (GaSP) response-surface approximation to the computer model, following on work by Sacks et al. (1989), Currin et al. (1991) and others. For complex $y^M(\mathbf{x})$ and/or \mathcal{X}_C , developing an emulator near the boundary of \mathcal{X}_C , while simultaneously trying to determine \mathcal{X}_C , will require an iterative or adaptive process.

For the test-bed problem, it is convenient to switch to a more mnemonic notation: we denote the flow volume x_1 by V and the initial flow angle x_2 by φ . Thus $(V, \varphi) \in \mathcal{X} = (0, \infty) \times [0, 2\pi)$, and we desire to find the critical contour in this space that separates catastrophic events from benign events. This contour, which we will call Ψ , is most conveniently represented by finding, for each angle $\varphi \in [0, 2\pi)$, the minimum volume V that causes catastrophic damage (all larger volumes will cause even worse damage). Thus we will write $\Psi = \Psi(\varphi) = \inf\{V : y^M(V, \varphi) \geq 1\text{m}\}$. Also for the test-bed, we focus on risk at two locations: Plymouth city and Bramble Airport.

3.2 Design Points

To begin the search for Ψ we ran 256 flow simulations at design points in a large region $\mathcal{X} = [10^5\text{m}^3, 10^{9.5}\text{m}^3] \times [0, 2\pi)$ of the input space. These 256 design points were chosen according to a Latin hypercube design and are plotted in Figure 2. Latin hypercubes are space-filling designs which have proven very successful for all-purpose designs of computer experiment runs, since they require relatively few design points per input to “fill” the design space (see, for example, Sacks et al. 1989; Tang 1993; Koehler and Owen 1996; Santner et al. 2003, Chap. 5).

Initially the emulator is based on a subset of these design points, which is then augmented

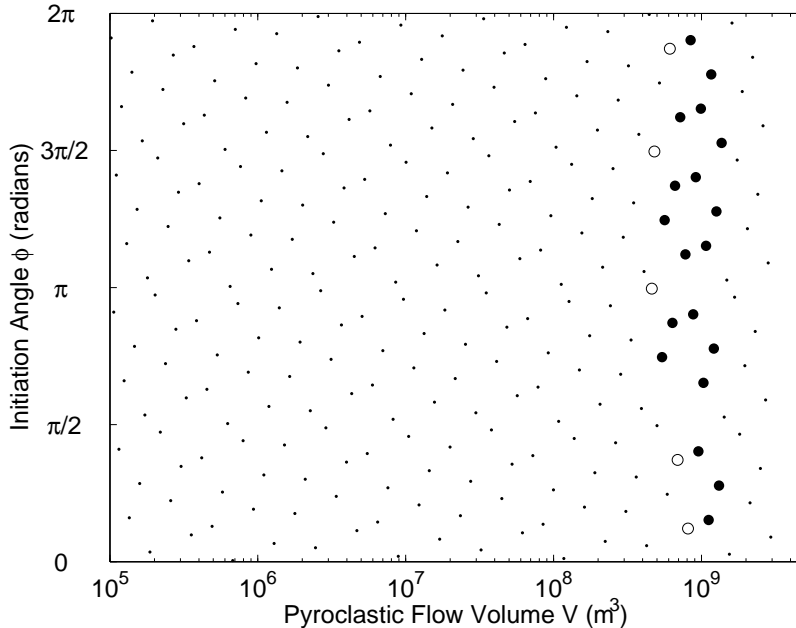


Figure 2: A representation of the design points covering the space \mathcal{X} . Dots represent all of the (V, φ) pairs used as initial conditions in the full simulation runs. Large dots identify points used to fit GaSP emulator: filled large dots are those that achieved $y^M(V, \varphi) > 0$ at Plymouth, open large dots are those for which $y^M(V, \varphi) = 0$ at Plymouth.

in an adaptive way (described in Section 3.5) to improve the estimated frontier Ψ . First we separate the design points corresponding to zero height at the location of interest from those corresponding to $y^M(V, \varphi) > 0$, i.e., some flow activity. We further reduce this set by eliminating very large volume cases, which could lead to difficulty in fitting the emulator. (We made this choice location-specific: for Plymouth, $V > 10^{8.6}\text{m}^3$ and, for the Bramble Airport, $V > 10^{8.8}\text{m}^3$). Next, to ensure that the emulated max-height surface goes to zero near the actual curve that separates \mathcal{X} into flow and no-flow events, we also include some of the largest volumes giving zero height at the specific location. Specifically, we divide the initiation angle domain into M equal-width subdomains, and within each, pick the largest volume design point such that $y^M(V, \varphi) = 0$. Furthermore to ensure that the emulated height surface is (approximately) a periodic function of initiation angle, we periodically extend design points so the resulting initiation angle domain for the emulated surface is $[-\pi/2, 5\pi/2]$.

In this way, we are left with N design points (35 and 42, respectively, for Plymouth and Bramble Airport), many fewer than the initial 256, that we use to fit the emulator of the max-height flow at a specific location over a reduced region $\mathcal{X}_{\mathcal{R}}$ of the input space around the critical frontier. We denote this initial set of points by \mathcal{D} . It should be emphasized that the goal of this reduction to N design points was to obtain a more accurate emulator in the region where the critical contour appears to be. That emulators often perform better when localized was demonstrated in Gramacy et al. (2004).

3.3 GaSP Emulator

Let \mathbf{y} denote the N -dimensional vector of simulator runs $y^M(\mathbf{x})$ for $\mathbf{x} = (V, \varphi) \in \mathcal{D}$. We can fit an emulator (GaSP response surface) to \mathbf{y} . Indeed, Figure 3 shows the mean surfaces of the emulated max-height GaSP's (developed below) for the two sites under consideration, as a function of the volume and angle inputs, along with dots indicating the actual heights at the design-points in \mathcal{D} that were obtained from the computer model runs.

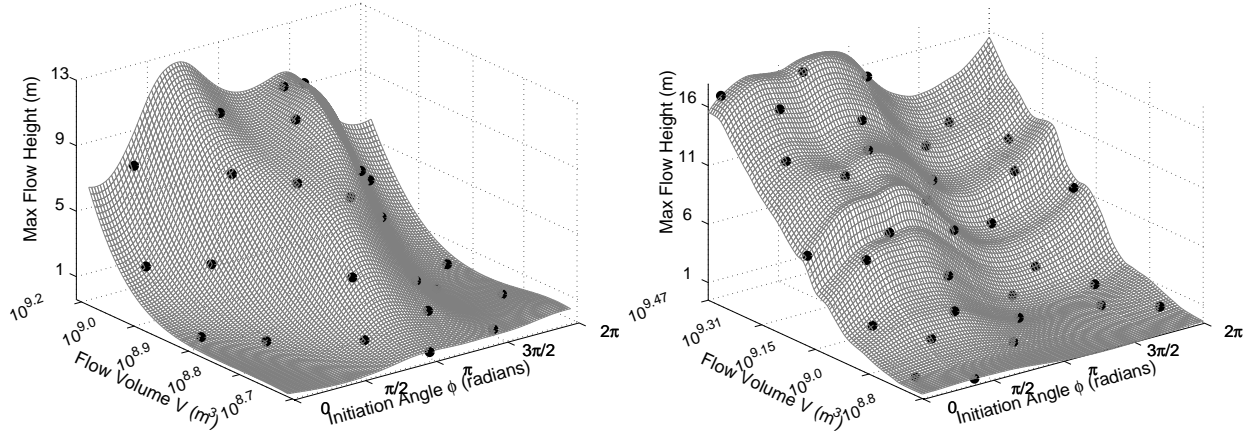


Figure 3: Left: Plymouth; Right: Bramble Airport. Max-height surfaces are the mean of the GaSP emulators. Dark points represent the max-height simulation output at design points.

Gaussian emulators may be developed not only for $\mathbf{y} = y^M(\mathbf{x})$, but also for $g(\mathbf{y})$ (which we denote by \mathbf{y}^g) for any strictly monotone function g , with the emulated values then transformed back to the original units. A judicious choice of transformation function g may lead to a problem where Gaussian interpolation does a better job of interpolating model outputs, leading to more efficient and more accurate emulation. In our ongoing example, we consider two functions: the identity $g(\mathbf{y}) = \mathbf{y}$ (no transformations needed) and $g(\mathbf{y}) = \log(\mathbf{y} + \mathbf{1})$. This latter transformation was chosen because it “spreads out” small positive values of \mathbf{y} , de-emphasizes extremely large values of \mathbf{y} , and is well defined at $\mathbf{y} = 0$.

We must specify a GaSP prior distribution for the (unknown) function $g(y^M(V, \varphi))$ over the region of interest $(V, \varphi) \in \mathcal{X}_{\mathcal{R}}$. Specifically, we take

$$g(y^M(V, \varphi)) = \beta + mV + z(V, \varphi), \quad (1)$$

where $\beta + mV$ constitutes a linear trend (mean) function and where $z(V, \varphi)$ is a zero-mean spatial Gaussian process, with covariance specified below. Note, it is reasonable to assume that y^M and hence $g \circ y^M$ increase monotonically in the volume V , but not in the angle φ ; thus there is no reason to include φ in the trend. The Gaussian process $z(V, \varphi)$ has $\mathbf{E}[z(V, \varphi)] = 0$ and $\mathbf{Var}[z(V, \varphi)] = \sigma^2$. A periodic extension of the standard product exponential correlation structure, when evaluated at the design points $\mathbf{x}_i = (V_i, \varphi_i) \in \mathcal{D}$, would result in the correlation matrix $\mathbf{R} = [R_{ij}]$ (an $N \times N$ matrix) given for $1 \leq i, j \leq N$

by

$$\begin{aligned}
R_{ij} &= R[(V_i, \varphi_i), (V_j, \varphi_j)] \\
&= \exp(-\theta_V |V_i - V_j|^{\alpha_V}) \frac{1}{c(\theta_\varphi, \alpha_\varphi)} \sum_{k=-\infty}^{\infty} \exp(-\theta_\varphi |\varphi_i - \varphi_j + 2\pi k|^{\alpha_\varphi}) \quad (2)
\end{aligned}$$

for range parameters θ and smoothness parameters α for each “direction” in parameter space \mathcal{X} , where $c(\theta, \alpha) = 1 + 2 \sum_{k=1}^{\infty} e^{-\theta|2\pi k|^\alpha}$ is the appropriate normalizing constant. For computational reasons we employ a one-term approximation to this infinite sum

$$\approx \exp(-\theta_V |V_i - V_j|^{\alpha_V}) \exp(-\theta_\varphi |\varphi_i - \varphi_j|^{\alpha_\varphi}) \quad (3)$$

and extend the data periodically to the interval $[-\pi/2, 5\pi/2]$ to minimize edge effects. This approximation will be excellent for sufficiently large values of θ_φ ; we are exploring alternatives in ongoing work. The likelihood for this model is given by

$$\begin{aligned}
p(\mathbf{y}^g \mid \theta_V, \theta_\varphi, \alpha_V, \alpha_\varphi, \sigma^2, \beta, m) &= \\
&= \frac{1}{2\pi\sigma^2 |\mathbf{R}|^{1/2}} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{y}^g - \beta \mathbf{1} - m \mathbf{V})^T \mathbf{R}^{-1} (\mathbf{y}^g - \beta \mathbf{1} - m \mathbf{V}) \right], \quad (4)
\end{aligned}$$

where $\mathbf{1}$ is the vector of ones in \mathbb{R}^N and $\mathbf{V} = (V_1, \dots, V_N)$ with $(V_i, \varphi_i) \in \mathcal{D}$.

A popular, easy emulator is the GaSP resulting from plugging in maximum likelihood estimates into the conditional posterior distribution of the function given the data \mathbf{y}^g (and the parameters). The mean and variance are then given by the usual kriging expressions. This strategy, while convenient, typically results in underestimation of the variance for the predicted simulator at untried inputs. In practice this underestimation is usually negligible (see Bayarri et al. (2007) for discussion), but the number N of design points we are considering is modest, and so we incorporate the most crucial uncertainty (namely, that of β, m and σ) via a fully Bayesian analysis, while retaining the MLE’s for the parameters determining the correlation structure $(\theta_V, \theta_\varphi, \alpha_V, \alpha_\varphi)$ (a full Bayesian analysis would be prohibitively expensive computationally). We use the GaSP software of Welch (2003) to compute these MLE’s, and from now on $(\theta_V, \theta_\varphi, \alpha_V, \alpha_\varphi)$ are taken to be known and equal to these estimates, and regard the approximate correlation function R of Equation 3 as known.

We use independent objective prior distributions for β, m and σ with density functions $\pi(\beta) \propto 1$, $\pi(m) \propto 1$, and $\pi(\sigma) \propto \sigma^{-1}$. Combined with the likelihood (4) this produces the joint posterior distribution of (β, m, σ) . This posterior is used to find the predictive distribution of the simulator output at any untried input $\mathbf{x}^* = (V^*, \varphi^*)$ (see, for example, Santner et al. 2003, Section 4.1.3). The expression for the predictive distribution is considerably simplified if one first standardizes the flow volumes V , by defining $\tilde{V}_i \equiv V_i - V_R$ and $\tilde{V}^* \equiv V^* - V_R$, where $V_R = \mathbf{1}^T \mathbf{R}^{-1} \mathbf{V} / \mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}$. Then, defining $\tilde{\mathbf{V}}^T \equiv (\tilde{V}_1, \dots, \tilde{V}_N)$ and $\mathbf{r}^T \equiv (R(\mathbf{x}^*, \mathbf{x}_1), \dots, R(\mathbf{x}^*, \mathbf{x}_N))$ for $\mathbf{x}_i \in \mathcal{D}$, the predictive distribution of $y^M(\mathbf{x}^*)$ is

$$y^M(\mathbf{x}^*) \mid \mathbf{y}^g \sim t(y^*(\mathbf{x}^*), s^2(\mathbf{x}^*); N-2),$$

the noncentral Student t distribution with $N-2$ degrees of freedom and parameters

$$\begin{aligned}
y^*(\mathbf{x}^*) &= \mathbf{r}^T \mathbf{R}^{-1} \mathbf{y}^g + \frac{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}^g}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}} (1 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{1}) + \frac{\tilde{\mathbf{V}}^T \mathbf{R}^{-1} \mathbf{y}^g}{\tilde{\mathbf{V}}^T \mathbf{R}^{-1} \tilde{\mathbf{V}}} (\tilde{\mathbf{V}}^* - \mathbf{r}^T \mathbf{R}^{-1} \tilde{\mathbf{V}}) \\
s^2(\mathbf{x}^*) &= \left[(1 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r}) + \frac{(1 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{1})^2}{(\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1})} + \frac{(\tilde{\mathbf{V}}^* - \mathbf{r}^T \mathbf{R}^{-1} \tilde{\mathbf{V}})^2}{(\tilde{\mathbf{V}}^T \mathbf{R}^{-1} \tilde{\mathbf{V}})} \right] \\
&\quad \times \frac{1}{N-2} \left[(\mathbf{y}^g)^T \mathbf{R}^{-1} \mathbf{y}^g - \frac{(\mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}^g)^2}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}} - \frac{(\tilde{\mathbf{V}}^T \mathbf{R}^{-1} \mathbf{y}^g)^2}{\tilde{\mathbf{V}}^T \mathbf{R}^{-1} \tilde{\mathbf{V}}} \right].
\end{aligned}$$

Note that the computations are the same level of difficulty as those for the plug-in GaSP, so there is no computational cost in accounting for the uncertainty in β , m and σ .

This is our emulator for the identity transformation. Under the transformation $g(y^M(\mathbf{x})) = \log(y^M(\mathbf{x}) + 1)$, posterior moments for the inverse transformation do not exist, so we resort to quantiles instead and use the median to predict $y^M(\mathbf{x}^*)$ and appropriate quantiles to provide credible bands (note that it suffices to find the quantiles in the emulator for the log transformation and then transform them back to the original space). In our analyses, we found that the GaSPs in the transformed scale were much more stable than those developed directly in the original space, so from now on, unless explicitly said, we limit ourselves to using emulators derived in the log-transformed space.

3.4 Catastrophic Event Contours

We can approximate the catastrophic event contour Ψ by determining the appropriate contour numerically from the emulated GaSP surface. For Plymouth and Bramble Airport, $\Psi(\varphi)$ is shown in Figure 4 for the emulator obtained by fitting a GaSP to $\log(y^M(\cdot) + 1)$ and transforming back.

Since the emulators are only approximations to the computer model, there will be error in the estimation of Ψ . This is reflected in the 90% credible bands given in Figure 4 (found simply by transforming back to meters the 5% and 95% quantiles of the posterior predictive distribution of the Bayesian emulator for the log transformation).

Figure 4 suggests that there is considerable uncertainty for both the Plymouth and the Bramble Airport contours. Reducing this uncertainty requires additional design points and computer model runs.

3.5 Adapting the Design

The critical frontier—the 1m maximum-flow contour—is close to the 0m flow frontier, and (of course) there are no negative flow heights. Gaussian emulators can have difficulty near such hard boundaries. It is thus particularly important to improve the emulator near this 0m flow boundary. Figure 2 indicates (as large open dots) design points where there is no pyroclastic flow at the target location, and nearby points (as large filled dots) where the maximum flow is positive. By adding new design points to fill in the larger ‘gaps’ between these two classes of points, we may be able to improve the precision with which we estimate

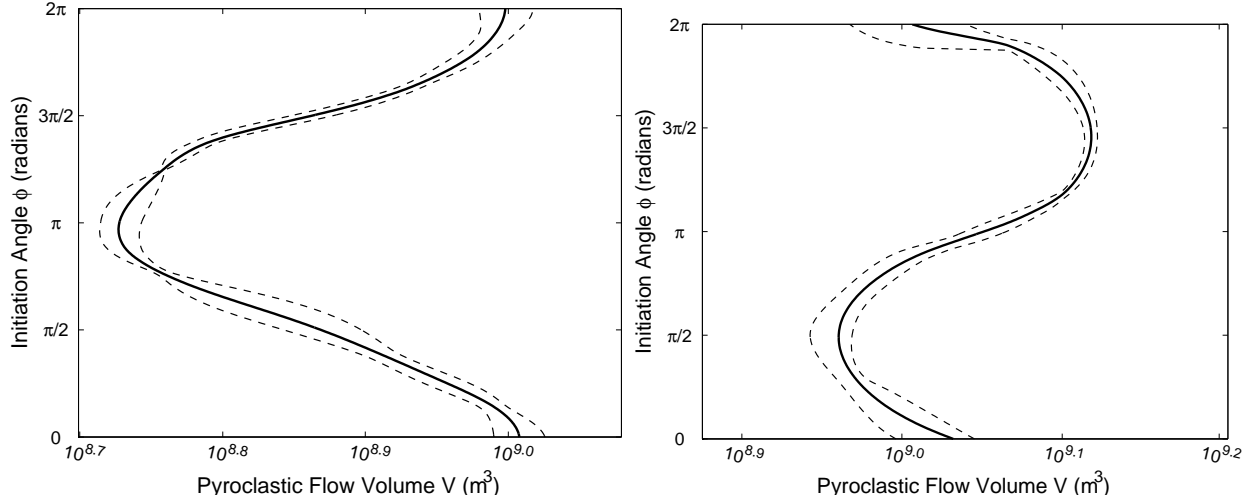


Figure 4: Median estimates (solid curves) and 90% credible bands (dashed curves) of frontier $\Psi(\phi)$, based on Gaussian process fit to log-transformed simulation output, at Plymouth (on left) and at Bramble Airport (on right).

Ψ . In addition, Figure 4 indicates (with the 90% interval) the uncertainty in the estimated Ψ curve at various input angles, and it is natural to choose additional design points at locations of large uncertainty.

Additional computer model runs were obtainable in batches of roughly 10 runs, so these two considerations were informally combined to arrive at a choice of 9 additional design points for computer model runs, as a preliminary investigation of the improvement that would be obtained. This augmented design is given in Figure 5 at Plymouth.

Surprisingly, this augmented design turned out to be sufficiently accurate for the risk assessment goals of this study (see below), so additional adaptation was not required. It is clear, however, that, for the larger problem of constructing overall risk maps, this process of adaptive design will need to be automated.

New emulators were then fit to the augmented design region. The new critical frontier for Plymouth is shown in Figure 6, along with the 90% credible bands for the resulting frontier (as dashed lines); for contrast, the 90% credible bands from the original design are also shown (as dotted lines). While the critical frontier did not change markedly, the uncertainty decreased substantially. The results for Bramble Airport with the new design points were similar, and so were omitted.

Systematic methods for progressive refinement of models in highly uncertain regions, long used in optimization (see Booker et al. 1999, for example), also offer promise for use in adaptive emulators. Other approaches to adaptive designs can be found in Gramacy et al. (2004), Ranjan et al. (2008), and Williams et al. (2000).

As partial validation for the accuracy of the emulator we conducted a leave-one-out cross-validation study, and found that 65 of 71 points were within the nominal 90% prediction intervals.

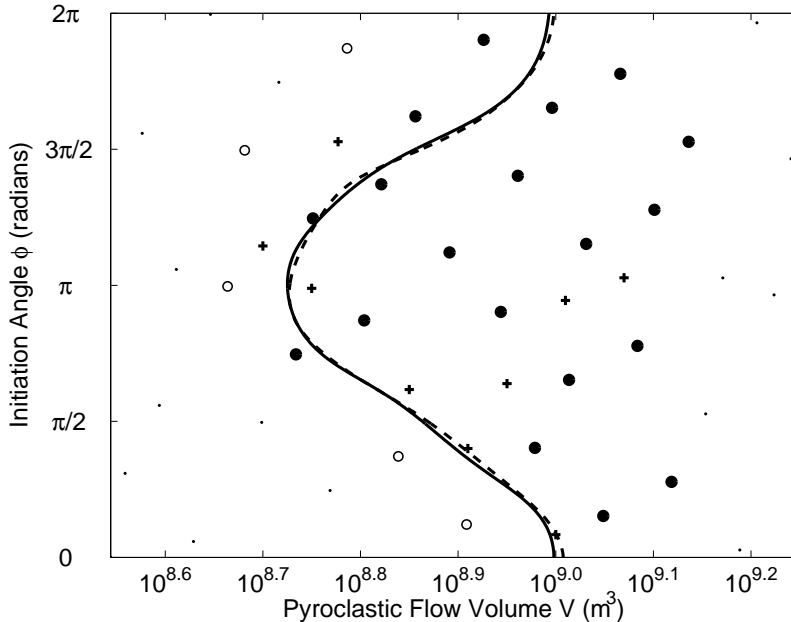


Figure 5: Adapted subdesign used to fit GaSP max height surface at Plymouth. Plus marks represent the additional points added to the design; large filled and open dots are original design points that achieved $y^M(V, \varphi) > 0$ and $y^M(V, \varphi) = 0$ at Plymouth, respectively; small dots are design points not used to fit Ψ for Plymouth. Dashed curve is Ψ obtained from the GaSP max-height surface fit to the original design and solid curve is Ψ fit to the updated design.

4 INPUT MODELING: THE FREQUENCY AND SEVERITY OF EXTREME EVENTS

One key to successful risk assessment using computer models is determination of a suitable input distribution for exercise of the computer model. This is particularly challenging when the goal is to study extreme events. In this section we propose a type of modeling of the needed input distribution that is particularly suited for this enterprise.

Since the classical “three types” theorem of Fisher and Tippett (1928) and the influential book of Gumbel (1958) it has been common to use the generalized extreme value (GEV) distribution to model extreme events, such as the maximum volume of the PFs during some time interval. This approach of approximating the probability distribution of the maximum of independent identically-distributed random variables is well-suited to finding approximate probability distributions of the most extreme occurrences of nearly-stationary processes in fixed periods of time, when the stationary distribution is unknown, but is wasteful and inefficient for problems (such as ours) in which evidence is available about the uncertain *rates* of occurrence of events over a wide range of extremity. Instead we employ the “peaks over threshold” approach introduced by Todorović and Zelenhasić (1970) (see Coles (2001, Chap. 4) for a modern treatment), based on point-process methods modeling the joint distribution of the times and extremities of events over fixed time intervals. This lets us

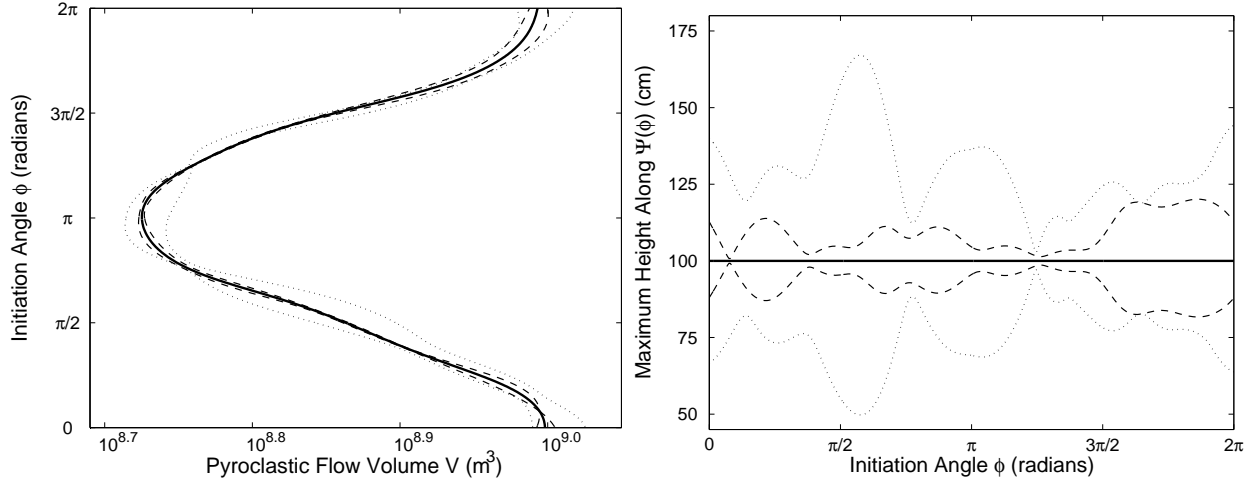


Figure 6: Left: $\Psi(\varphi)$ (solid) with 90% credible bands for original (dotted) and updated (dashed) designs. Right: 90% credible bands along $(\Psi(\varphi), \varphi)$ for GaSP max-height surface fit to original design (dotted) and updated design (dashed).

exploit more of the information contained in the rates of past smaller flows to predict those of future larger ones.

4.1 The Proposed Pareto-Poisson Model

Figure 7 shows an empirical plot of the number of pyroclastic flows exceeding volume $V_j \geq v$ vs. v during the period from March, 1996 through July, 2008 for the Soufrière Hills Volcano, on a log-log scale, for large volumes $v \geq \epsilon$ (here $\epsilon = 5 \cdot 10^4 \text{ m}^3$). Three principal sources of information were used to assemble this inventory: (1) the record of daily field observations made by the Montserrat Volcano Observatory (MVO) scientists and reported in daily reports or weekly reports (<http://www.mvo.ms/>), supplemented with (2) published papers describing specific events at Soufrière Hills Volcano, and (3) internal open-file reports of the MVO. The vertical banding is an artifact, suggesting some rounding-off of smaller reported PF volumes.

The approximately-linear fall-off on a log-log scale suggests that the probability distribution of flow volumes satisfies the linear equation

$$\log \mathbb{P}[V \geq v \mid V \geq \epsilon] \approx -\alpha \log(v) + c \quad (5)$$

for some constants $\alpha > 0$ and $c \in \mathbb{R}$, and hence the distribution of the $\{V_j\}$ is approximately the Pareto, with

$$\mathbb{P}[V \geq v] \approx (v/\epsilon)^{-\alpha}, \quad v \geq \epsilon, \quad (6)$$

just as expected for a peaks-over-threshold model in the domain of the Fréchet extreme value distribution (Coles 2001, Sec. §4.2). With the estimated value of $\alpha \approx 0.64 < 1$ (see Section 4.3), the PF volumes will not have finite means or variances, and their sums and averages will not obey the central limit theorem. The total accumulated flow over any fixed

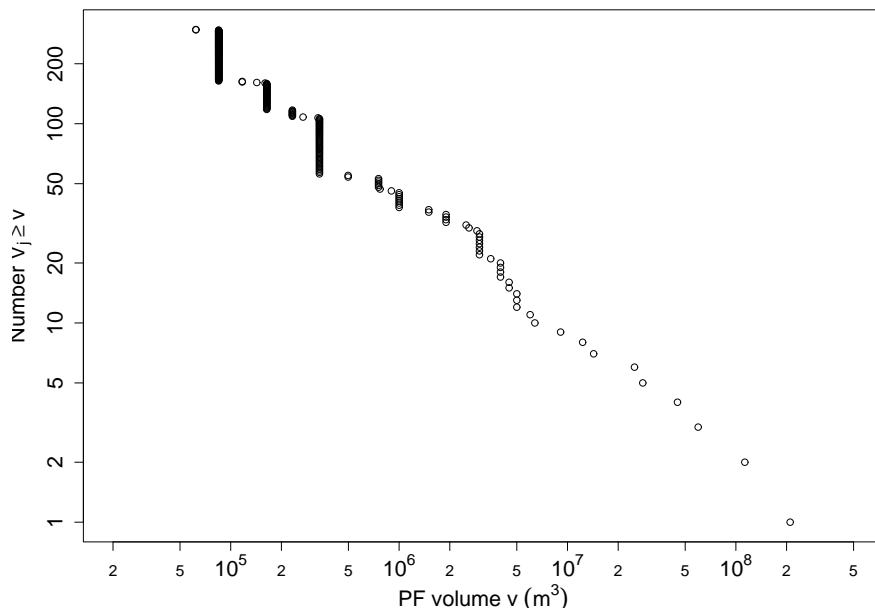


Figure 7: Frequency-*vs.*-magnitude plot for pyroclastic flows at Soufrière Hills Volcano.

time interval will have approximately an α -Stable distribution, not a normal distribution (see Section 4.2).

We model the PFs whose volume exceeds some fixed threshold ϵ as a marked Poisson process of initial volumes and initiation angle pairs $\{(V_j, \varphi_j)\}$ at times $\tau_j > 0$. With little evidence of nonstationarity over this time period, we take the Poisson rate to be some constant λ_ϵ ; we also assume independence of the Pareto-distributed volumes $\{V_j\}$. This model cannot be exactly correct for pyroclastic flows, of course, for many reasons— eruption rates will evolve over time; large PFs will deplete dome volumes, precluding a subsequent large PF within a short time interval; large PFs in one direction change the dome geometry, influencing the initiation angle for subsequent PFs, for example. Nevertheless the empirical evidence against these simplifying modeling assumptions is very weak, so we adopt them provisionally. In Section 4.5 we present more specific criticisms of the present model, introduce some approaches for model validation, and discuss possible modifications of this model to meet its anticipated shortcomings.

4.2 Relation to α -Stable Distributions

Under this model the total accumulated volume

$$X_t^\epsilon := \sum_{j=1}^{J_\epsilon(t)} V_j$$

of the $J_\epsilon(t) \sim \text{Po}(\lambda_\epsilon t)$ PFs exceeding volume ϵ at times $\tau_j \leq t$ has an infinitely-divisible (ID) distribution— it is, in fact, a stationary independent increment (SII) or Lévy process with

Lévy measure

$$\nu_\epsilon(du) = \alpha \lambda_\epsilon \epsilon^\alpha u^{-\alpha-1} \mathbf{1}_{\{u>\epsilon\}} du. \quad (7)$$

For $0 < \alpha < 1$, as in our data, the distribution of $\{X_t^\epsilon\}$ is identical to that of the jumps exceeding ϵ of the fully-skewed α -Stable process $X_t \sim \text{St}(\alpha, 1, \gamma t, 0)$ with shape parameter α and with rate constant $\gamma = \pi \lambda / \{2\Gamma(\alpha) \sin(\pi\alpha/2)\}$ depending only on α and on

$$\lambda \equiv \lambda_\epsilon \epsilon^\alpha. \quad (8)$$

We now turn to the problem of estimating the parameters α and λ from data.

4.3 The Likelihood Function

The problem of estimating the parameter α for α -Stable distributions and the closely-related one of estimating the tails of probability distributions (such as the α -Stable and Pareto) are long-studied and notoriously difficult (DuMouchel 1983; Smith 1987). For our Bayesian approach, growing out of Wolpert and Ickstadt (2004) and Wolpert and Taqqu (2005), we begin with the likelihood function, upon observing $\{(V_j, \tau_j) : V_j > \epsilon, 0 < \tau_j \leq T\}_{j \leq J_\epsilon}$,

$$\begin{aligned} L(\alpha, \lambda) &\propto (\alpha \lambda)^{J_\epsilon} \exp \left[-\lambda T \epsilon^{-\alpha} - \alpha \sum_{j \leq J_\epsilon} \log V_j \right] \\ &= (\alpha \lambda \epsilon^{-\alpha})^{J_\epsilon} e^{-\lambda T \epsilon^{-\alpha} - \alpha S_\epsilon}. \end{aligned} \quad (9)$$

Maximum likelihood estimators (MLEs) are easy to compute, and depend on the data only through the sufficient statistics J_ϵ and $S_\epsilon := \sum_{j \leq J_\epsilon} \log(V_j/\epsilon)$:

$$\hat{\alpha} = J_\epsilon / S_\epsilon \quad \hat{\lambda} = J_\epsilon \epsilon^{\hat{\alpha}} / T.$$

These will be used below to construct an importance function for evaluating a key integral needed for our estimates of risk.

4.4 Objective Bayesian Inference

Let $\pi(\alpha, \lambda, \varphi)$ denote an appropriate prior density function for the Pareto and Stable shape parameter α , rate parameter λ from Equation (8), and initiation angle φ . We will base inferences on the posterior density

$$\pi^*(\alpha, \lambda, \varphi) = \pi(\alpha, \lambda, \varphi \mid \text{data}) \propto L(\alpha, \lambda) \pi(\alpha, \lambda, \varphi);$$

note that the likelihood does not depend on the initiation angle φ . Little expert prior information was available concerning α , λ , and φ , so objective prior distributions were used.

The obvious objective prior distribution for the angle φ is the uniform distribution on $[0, 2\pi)$. This was deemed reasonable by the geologists in the project, as was the assumption that φ may be viewed as independent of the other parameters. Thus we turn to objective choices for $\pi(\alpha, \lambda)$.

A popular objective prior is the invariant Jeffreys-rule prior (Jeffreys 1961, pg. 181), given by

$$\pi_J(\alpha, \lambda) \propto |I(\alpha, \lambda)|^{1/2},$$

where $I(\alpha, \lambda)$ is the expected (Fisher) information matrix and where $|\cdot|$ denotes matrix determinant. The Fisher information matrix for the Pareto-Poisson model of Section 4.1 is available in closed form:

$$I(\alpha, \lambda) = T \epsilon^{-\alpha} \begin{bmatrix} \lambda(\alpha^{-2} + \log^2 \epsilon) & -\log \epsilon \\ -\log \epsilon & \lambda^{-1} \end{bmatrix}$$

and so the Jeffreys-rule prior density is

$$\pi_J(\alpha, \lambda) \propto \alpha^{-1} \epsilon^{-\alpha} \mathbf{1}_{\{\alpha > 0, \lambda > 0\}}.$$

It is generally recognized, however, that the Jeffreys-rule prior can behave badly in multiparameter problems (see, e.g., Bernardo and Smith 1994, §5.4); a more promising choice is the one-at-a-time reference prior distribution of Berger and Bernardo (1992). This reference prior requires the specification of a “parameter of interest”—unfortunately, in this problem, the quantity of greatest interest is the risk probability, a complicated function of α and λ . Instead we construct two reference priors, based on declaring first α and then λ to be the parameter of interest. Each is computed with an easy application of the algorithm in Berger and Bernardo (1992), restricting (α, λ) to compact sets of the form $[1/k, k] \times [1/l, l]$ which grow to \mathbb{R}_+^2 as $k \rightarrow \infty$ and $l \rightarrow \infty$, leading to

$$\pi_{R\alpha}(\alpha, \lambda) \propto \lambda^{-1/2} \alpha^{-1} \epsilon^{-\alpha/2} \mathbf{1}_{\{\alpha > 0, \lambda > 0\}}$$

and

$$\pi_{R\lambda}(\alpha, \lambda) \propto \lambda^{-1/2} [\alpha^{-2} + (\log \epsilon)^2]^{1/2} \epsilon^{-\alpha/2} \mathbf{1}_{\{\alpha > 0, \lambda > 0\}}.$$

Each of these three objective prior candidates is of the form

$$\pi(\alpha, \lambda) = \lambda^{a-1} \alpha^{-1} g(\alpha) \mathbf{1}_{\{\alpha > 0, \lambda > 0\}} \quad (10)$$

for some number $a \geq 0$ and some bounded function $g(\alpha)$. We will use each of them in the calculations of Section 5.2.

4.5 Checking the Statistical Model

Surely some of the assumptions presented in Section (4.1) that underlie this model are false; for example,

1. Dome collapse generation mechanisms are clearly linked to the non-stationary process of lava extrusion. Lava extrusion rates typically vary between 0–20 m³/s and entail fluctuations that occur on time-scales of days to months (Voight et al. 1999). Over short time-scales, frequency-magnitude models that assume stationarity would certainly be incorrect. For multi-year periods such as those we consider here, however, the assumption of stationarity for time-averaged extrusion rates is not unreasonable.

2. Some collapses, especially those of the small tail end member, are not strictly independent events. Rockfalls can trigger other rockfalls. However, the larger dome collapse events we consider can reasonably be considered as independent.

It is useful to examine the data to assess how well the model fits these data. Under the proposed model, the points

$$\{(x_j, y_j) := ((\epsilon/V_j)^\alpha, \tau_j/t)\}_{j \leq J_\epsilon}$$

and, with $\tau_0 := 0$,

$$\{(x_j, z_j) := ((\epsilon/V_j)^\alpha, e^{-\lambda \epsilon^{-\alpha}(\tau_j - \tau_{j-1})})\}_{j \leq J_\epsilon}$$

would each be J_ϵ independent draws from the unit square. Any departures we detect from that uniform distribution will 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 (in which case we may wish to replace our Poisson event process with an inhibitory process such as the Strauss or area interaction processes); they may be *more* irregular than *i.i.d.* uniform draws if there are aftershocks as in earthquakes (in which case a Poisson cluster process or a Cox process may be more appropriate). Flow volumes may tend to be larger after relatively long inter-event times and smaller after shorter intervals, perhaps, suggesting a Markov sequence rather than independent random variables for flow volumes. All these departures can be modeled, once they are discovered. Preliminary investigation suggests that both point-clouds appear sufficiently uniform; in later investigations we will explore this issue more deeply.

5 RISK ASSESSMENTS AT KEY LOCATIONS

We are finally in a position to compute the probability of a catastrophic event in the next t years for two key locations: Plymouth and Bramble Airport.

5.1 Probability of a Catastrophic Event

Fix some time $t > 0$, and consider the contour $\Psi : [0, 2\pi) \rightarrow \mathbb{R}_+$ from Section 3.1. Under the model of Section 4.1, the number of PFs in a future time interval of length t years whose volume V_i and initiation angle φ_i satisfy $V_i > \Psi(\varphi_i)$ (i.e., the number of catastrophic PFs in t years) will have a Poisson probability distribution with conditional expectation

$$\mathbb{E}(\# \text{ catastrophic PFs in } t \text{ yrs} \mid \alpha, \lambda) = \frac{t \lambda}{2\pi} \int_0^{2\pi} \Psi(\varphi)^{-\alpha} d\varphi, \quad (11)$$

for given values of the parameters α and λ , so the probability of a catastrophic event is

$$\mathbb{P}(\text{ At least one catastrophic PF in } t \text{ yrs} \mid \alpha, \lambda) = 1 - \exp \left[-\frac{t \lambda}{2\pi} \int_0^{2\pi} \Psi(\varphi)^{-\alpha} d\varphi \right].$$

There is a delicate point here that needs to be taken into account in risk assessment of rare events. Given the angle φ (and α and λ),

$$\mathbb{P}(\text{ At least one catastrophic PF in } t \text{ yrs} \mid \alpha, \lambda, \varphi) = 1 - \exp \left[-t \lambda \Psi(\varphi)^{-\alpha} \right].$$

We also know the distribution of φ (uniform on $[0, 2\pi)$), so it is tempting to view the above expression as the conditional risk probability given φ (and α and λ), and then integrate over the known distribution of φ . This is wrong because the relevant Poisson mean is given in (11), i.e., the integration over φ must take place in the Poisson mean.

Although the parameter vector (α, λ) is typically unknown, the maximum likelihood estimator of this probability is easy to compute:

$$\begin{aligned}\hat{P}(t) &= \text{MLE of } \mathbf{P}[\text{At least one PF} > \Psi(\varphi) \text{ in } t \text{ yrs} \mid \alpha, \lambda] \\ &= 1 - \exp \left[-\frac{t \hat{\lambda}}{2\pi} \int_0^{2\pi} \Psi(\varphi)^{-\hat{\alpha}} d\varphi \right].\end{aligned}\tag{12}$$

The MLE distorts the risk estimate by ignoring uncertainty in the estimation of α and λ ; a better choice is to compute the posterior probability of catastrophe in t years, using the likelihood function of Equation (9) and a prior density of the form in Equation (10) from Section 4.4:

$$\begin{aligned}P(t) &= \mathbf{P}[\text{At least one PF} > \Psi(\varphi) \text{ in } t \text{ yrs} \mid \text{data}] \\ &= 1 - \iint_{\mathbb{R}_+^2} \exp \left[-\frac{t \lambda}{2\pi} \int_0^{2\pi} \Psi(\varphi)^{-\alpha} d\varphi \right] \pi^*(\alpha, \lambda) d\alpha d\lambda\end{aligned}\tag{13}$$

for the posterior density $\pi^*(\alpha, \lambda) = Z^{-1} L(\alpha, \lambda) \lambda^{a-1} \alpha^{-1} g(\alpha) \mathbf{1}_{\{\alpha > 0, \lambda > 0\}}$, with normalizing constant $Z := \iint_{\mathbb{R}_+^2} L(\alpha, \lambda) \lambda^{a-1} \alpha^{-1} g(\alpha) d\alpha d\lambda$. The λ integral in Equation (13) is available in closed form, leaving:

$$\begin{aligned}P(t) &= 1 - Z^{-1} \iint_{\mathbb{R}_+^2} \exp \left[-\lambda t \epsilon^{-\alpha} I_\epsilon(\alpha) \right] (\alpha \lambda \epsilon^{-\alpha})^{J_\epsilon} e^{-\lambda T \epsilon^{-\alpha} - \alpha S_\epsilon} \lambda^{a-1} \alpha^{-1} g(\alpha) d\alpha d\lambda \\ &= 1 - Z^{-1} \iint_{\mathbb{R}_+^2} \lambda^{J_\epsilon + a - 1} \exp \left[-\lambda \epsilon^{-\alpha} (T + t I_\epsilon(\alpha)) \right] e^{-\alpha S_\epsilon} (\alpha \epsilon^{-\alpha})^{J_\epsilon} \pi(\alpha) d\alpha d\lambda \\ &= 1 - \tilde{Z}^{-1} \int_{\mathbb{R}_+} \left[1 + (t/T) I_\epsilon(\alpha) \right]^{-J_\epsilon - a} \alpha^{J_\epsilon - 1} e^{-\alpha [S_\epsilon - a \log \epsilon]} g(\alpha) d\alpha\end{aligned}\tag{14}$$

where

$$I_\epsilon(\alpha) := \frac{1}{2\pi} \int_0^{2\pi} [\Psi(\varphi)/\epsilon]^{-\alpha} d\varphi$$

and where

$$\tilde{Z} = \frac{Z T^{J_\epsilon + a}}{\Gamma(J_\epsilon + a)} = \int_{\mathbb{R}_+} \alpha^{J_\epsilon - 1} e^{-\alpha [S_\epsilon - a \log \epsilon]} g(\alpha) d\alpha.$$

Only for some prior distributions can the normalizing constant \tilde{Z} be computed explicitly (for the Jeffreys-rule prior, for example, it is $\tilde{Z} = \Gamma(J_\epsilon) S_\epsilon^{-J_\epsilon}$ and the posterior distribution of α is $\alpha \sim \text{Ga}(J_\epsilon, S_\epsilon)$), but the computational method presented in Section 5.2 does not require \tilde{Z} .

5.2 Numerical Computation

To compute the probabilities in Equation (14) we use importance sampling with an *importance function* $f_I(\alpha)$ —so that, for any prior of the form of Equation (10) we approximate

the risk probability of Equation (14) as

$$P(t) \cong 1 - \frac{\sum_i w_i \left[1 + \frac{t}{NT} \sum_{n=1}^N (\Psi_n/\epsilon)^{-\alpha_i} \right]^{-J_\epsilon - a}}{\sum_i w_i} \quad (15)$$

where the $\{\alpha_i\}$ are draws from the distribution with density function $f_I(\alpha)$, and where $\Psi_n \equiv \Psi(2\pi n/N)$ for $1 \leq n \leq N$ for some large integer $N \in \mathbb{N}$. The weights are given by

$$w_i \equiv \alpha_i^{J_\epsilon - 1} e^{-\alpha_i [S_\epsilon - a \log \epsilon]} g(\alpha_i) \mathbf{1}_{\{\alpha_i > 0\}} / f_I(\alpha_i);$$

note that in this form of importance sampling it is unnecessary to know the normalizing constant \tilde{Z} . For efficiency, the same importance sample is used to determine the risk probability simultaneously for all times t .

For the Jeffreys-rule prior distribution we use the Gamma $\text{Ga}(J_\epsilon, S_\epsilon)$ density as an importance function, leading to constant weights w_i (i.e., these are exact draws from the posterior distribution) and a simple bound of $1/(4M)$ for the Monte Carlo sampling variance of an estimate of $P(t)$ based on M samples, so a sample of at least $M \geq 16588$ will ensure a 99% error bound of no more than ± 0.001 . We used $M = 1,000,000$, consuming under ten minutes of computing time. For the one-at-a-time reference priors we use the Student $t_\nu(\hat{\mu}, s)$ distribution

$$f_I(\alpha) \propto (1 + (\alpha - \hat{\mu})^2 / 3s^2)^{-2}$$

with $\nu = 3$ degrees of freedom, centered at the MLE $\hat{\mu} = 0.64$, with scale $s = \sqrt{(\hat{i}^{-1})_{\alpha\alpha}} = 0.03714$, where \hat{i} denotes the observed information matrix evaluated at the MLE $(\hat{\alpha}, \hat{\lambda})$. This choice gives bounded weights $\{w_i\}$, leading to efficient sampling and negligible Monte Carlo sampling error with $M = 250,000$ draws.

One final source of uncertainty that needs to be taken into account is the uncertainty in the critical contour Ψ , as displayed in Figure 6. This uncertainty could be treated as above, with draws of Ψ from its posterior distribution, but we instead chose to present 90% upper and lower credible bands for $P(t)$, based on the 90% upper and lower credible bands for Ψ from Figure 6. We took this approach because uncertainty in Ψ is not inherent to the problem, but rather is “removable” in the sense that additional runs of TITAN2D in the construction of the critical contours could reduce it arbitrarily. For our problem, this removable uncertainty was quite small, rendering additional designed TITAN2D runs unnecessary.

The risk predictions resulting from this computation are presented in Figure 8. As expected, the predicted risks at Plymouth are higher than at Bramble Airport and, as suggested above, the uncertainty arising from uncertainty in Ψ was quite small. The three reference prior distributions all gave nearly identical answers in each case, so no effort was made to distinguish the curves for the different priors. This is a sign that the data (through the likelihood function) are far more influential than the prior; one would expect a greater difference in problems with less extensive data sets.

It is worth emphasizing that the two major data-modeling assumptions used in the computation of the catastrophe probabilities were conservative. The stationarity assumption was conservative because it ignores the well-known feature of vulcanism that periods of volcanic

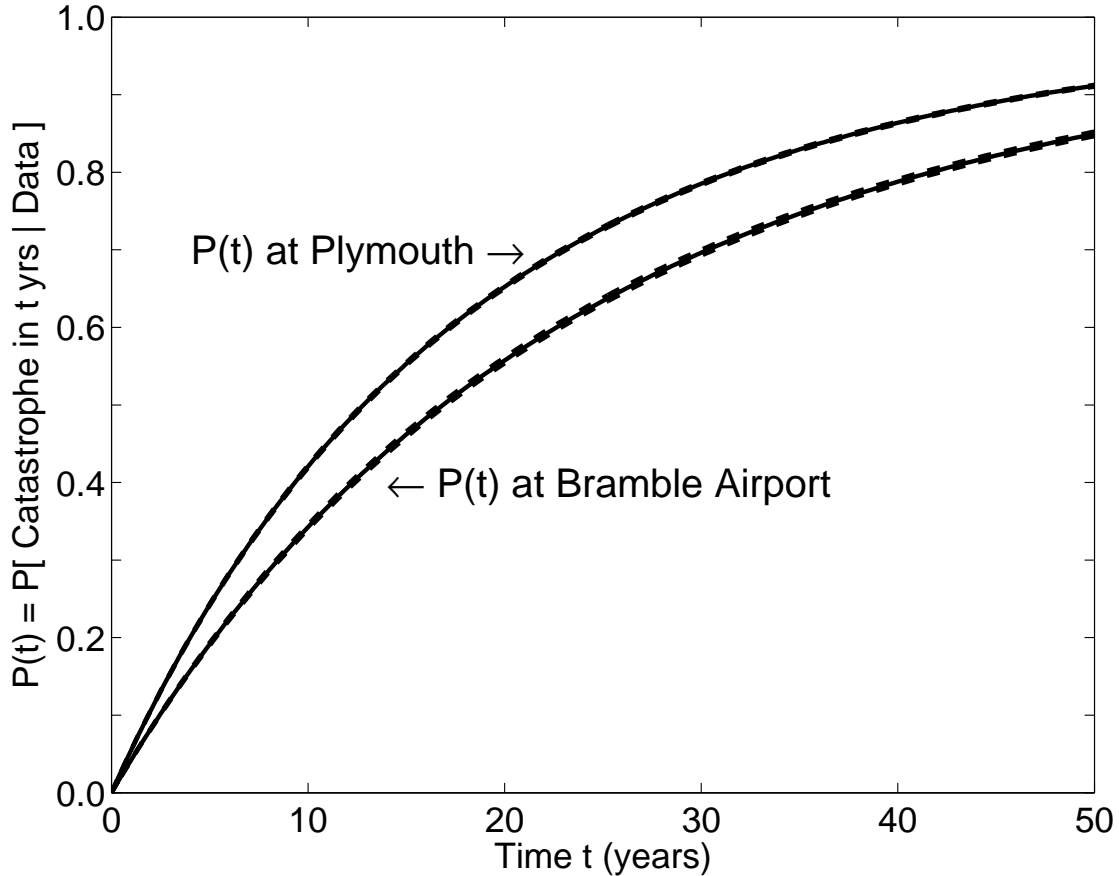


Figure 8: Posterior probability of catastrophe $P(t)$ at Plymouth (upper curves) and Bramble Airport (lower curves) within t years, for $0 \leq t \leq 50$. Solid lines indicate posterior medians, dashed curves indicate pointwise 90% credible bands.

activity seldom last more than one or two decades, separated by (often quite long) periods of relative inactivity; our risk predictions do not reflect the possibility that the activity at Montserrat might “turn off.” Unfortunately the data are quite sparse in this regard and improving the stationarity assumption is thus difficult. Also, the Pareto tails assume that possible flow volumes are unbounded, while in fact there are clear physical limits to the magnitude of flow that could possibly occur. This is not likely to affect significantly the probability of “moderate” catastrophes. Also, the reliability of these predictions rely on the adequacy of the computer model to predict reality. Although some satisfactory checks have been performed, full validation of the computer model with extremely rare data on catastrophic events is very problematic.

In conclusion, while some improvements in the statistical modeling are possible, we are quite confident that these would not significantly alter the predictions in Figure 8. Note, however, that both Plymouth and Bramble Airport experienced catastrophic pyroclastic flows during the first two years of the eruptions at Soufrière Hills Volcano, which suggests that the risk predictions in Figure 8 are actually too low. This is indeed likely true, for the

reasons indicated at the end of Section 2.2, relating to the need for better choices of friction coefficients for TITAN2D. Studies are currently being conducted to obtain improved friction coefficients, and this should result in improved risk predictions.

6 SUMMARY AND DISCUSSION

Risk assessment of catastrophic events— in the absence of sufficient data relating to the events— requires a delicate interplay of mathematical computer modeling and statistical analysis. Mathematical modeling is needed for the necessary extrapolation beyond the range of the data. Statistical modeling of the available (not necessarily extreme) data is crucial for determining the distribution of inputs that will drive the computer model. More subtly, statistical techniques for emulating computer models are needed when (as usual) the computer model is expensive to exercise, in order to determine the critical event contours needed for computation of the probability of catastrophic events.

In risk assessment, incorporation of all major sources of uncertainty into the analysis is important. The Bayesian approach we adopted is ideally suited to this task. In particular, we were able to incorporate not only the dominant uncertainty, the randomness in the volcanic process itself, but also uncertainty in the model for the input distribution (i.e., uncertainty about the values of model parameters α and λ).

Another potential source of uncertainty in problems such as this is error in the computer model. The computational mesh in TITAN2D is refined adaptively during the computation in order to maintain a high degree of accuracy, so numerical instability or convergence problems do not contribute significant variability or error. Other elements of the computer model, however, such as the topographic inputs and the (already mentioned) treatment of friction, may potentially introduce additional uncertainty or bias in the model output. Unfortunately, the data needed to investigate model validity are extremely sparse, as are techniques for assessing validity for complicated surface output such as that of TITAN2D. We hope to develop and present (elsewhere) an analysis of the bias and uncertainty in the output of TITAN2D, but that is quite beyond the scope of this paper. Note that any such bias analysis could, in principle, be incorporated directly into the probabilistic risk assessments considered here through Bayesian methods.

Acknowledgments

This research was supported in part by the National Science Foundation (GRANTS ITR 0121254, EAR 0087665, DMS-0103265, DMS-0635449, DMS-0757549, DMS-0757367, and DMS-0757527) and by the Spanish Ministry of Education and Science (Grant MTM2007-61554). The research began under the auspices and support of the Statistical and Applied Mathematical Sciences Institute (SAMSI) 2006-7 research program on the Development, Assessment and Utilization of Complex Computer Models.

References

- Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J. C., Lin, C.-H., and Tu, J. (2007), “A Framework for Validation of Computer Models,” *Technometrics*, 49, 138–154.
- Berger, J. O. and Bernardo, J. M. (1992), “On the development of the Reference Prior Method,” in *Bayesian Statistics 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, UK: Oxford Univ. Press, pp. 35–49.
- Bernardo, J. M. and Smith, A. F. M. (1994), *Bayesian Theory*, Wiley Series in Probability and Statistics, John Wiley & Sons.
- Booker, A. J., Dennis, J., John E., Frank, P. D., Serafini, D. B., Torczon, V., and Trosset, M. W. (1999), “A Rigorous Framework for Optimization of Expensive Functions by Surrogates,” *Structural Optimization*, 17, 1–13.
- Calder, E. S., Cole, P. D., Dade, W. B., Druitt, T. H., Hoblitt, R. P., Huppert, H. E., Ritchie, L. J., Sparks, R. S. J., and Young, S. R. (1999), “Mobility of pyroclastic flows and surges at the Soufrière Hills Volcano, Montserrat,” *Geophysical Research Letters*, 26, 537–540.
- Calder, E. S., Lockett, R., Sparks, S. J., and Voight, B. (2002), “Mechanisms of lava dome instability and generation of rockfalls and pyroclastic flows at Soufrière Hills Volcano, Montserrat,” in *The Eruption of Soufrière Hills Volcano, Montserrat from 1995 to 1999*, eds. T. H. Druitt and B. P. Kokelaar, number 21 in Geological Society Memoirs, London, UK: London Geological Society, pp. 173–190.
- Coles, S. G. (2001), *An Introduction to Statistical Modeling of Extreme Values*, New York, NY: Springer-Verlag.
- Collins, G. S. and Melosh, H. J. (2003), “Acoustic Fluidization and the Extraordinary Mobility of Sturzstroms,” *Journal of Geophysical Research*, 108, 2473.
- Currin, C., Mitchell, T. J., Morris, M. D., and Ylvisaker, D. (1991), “Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments,” *J. Am. Stat. Assoc.*, 86, 953–963.
- Dade, W. B. and Huppert, H. E. (1998), “Long-runout rockfalls,” *Geology*, 26, 803–806.
- Dalbey, K., Patra, A. K., Pitman, E. B., Bursik, M. I., and Sheridan, M. F. (2008), “Input Uncertainty Propagation Methods and Hazard Mapping of Geophysical Mass Flows,” *Journal of Geophysical Research*, 113, 1–16, doi:doi:10.1029/2006JB004471.
- DuMouchel, W. H. (1983), “Estimating the stable index α in order to measure tail thickness,” *Ann. Stat.*, 11, 1019–1036.
- Fisher, R. A. and Tippett, L. H. C. (1928), “Limiting forms of the frequency distributions of the largest or smallest member of a sample,” *Proc. Cambridge Philos. Soc.*, 24, 180–190.

- Gramacy, R. B., Lee, H. K. H., and MacReady, W. G. (2004), “Parameter Space Exploration With Gaussian Process Trees,” in *ICML-2004: Proceedings of the 21st International Conference on Machine Learning*, eds. R. Greiner and D. Schuurmans, ACM Press, pp. 353–360.
- Gumbel, E. J. (1958), *Statistics of Extremes*, New York, NY: Columbia University Press.
- Herd, R. A., Edmonds, M., and Bass, V. A. (2005), “Catastrophic lava dome failure at Soufrière Hills Volcano, Montserrat, 12-13 July 2003,” *Journal of Volcanology and Geothermal Research*, 148, 234–252.
- Iman, R. L., Johnson, M. E., and Watson, J., Charles C. (2005), “Uncertainty Analysis for Computer Model Projections of Hurricane Losses,” *Risk Analysis*, 25, 1299–1312.
- Iverson, R. M. (1997), “The physics of debris flows,” *Reviews of Geophysics*, 35, 245–296.
- Jeffreys, H. (1961), *Theory of Probability*, Oxford, UK: Oxford Univ. Press.
- Koehler, J. R. and Owen, A. B. (1996), “Computer Experiments,” in *Handbook of Statistics*, eds. S. Ghosh and C. R. Rao, Amsterdam, NL: Elsevier Science, volume 13, pp. 261–308.
- Patra, A. K., Bauer, A. C., Nichita, C. C., Pitman, E. B., Sheridan, M. F., and Bursik, M. I. (2005), “Parallel Adaptive Numerical Simulation of Dry Avalanches over Natural Terrain,” *Journal of Volcanology and Geothermal Research*, 139, 1–21.
- Ranjan, P., Bingham, D. R., and Michailidis, G. (2008), “Sequential Experiment Design for Contour Estimation from Complex Computer Codes,” *Technometrics*, 50, 527–541.
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), “Design and Analysis of Computer Experiments,” *Stat. Sci.*, 4, 409–435.
- Santner, T. J., Williams, B. J., and Notz, W. (2003), *The Design and Analysis of Computer Experiments*, Springer Series in Statistics, New York, NY: Springer-Verlag.
- Savage, S. B. and Hutter, K. (1989), “The motion of a finite mass of granular material down a rough incline,” *Journal of Fluid Mechanics*, 199, 177–215.
- Sheridan, M. F., Stinton, A. J., Patra, A. K., Pitman, E. B., Bauer, A. C., and Nichita, C. C. (2005), “Evaluating TITAN2D mass-flow model using the 1963 Little Tahoma Peak avalanches, Mount Rainier, Washington,” *Journal of Volcanology and Geothermal Research*, 139, 89–102.
- Smith, R. L. (1987), “Estimating tails of probability distributions,” *Ann. Stat.*, 15, 1174–107.
- Tang, B. (1993), “Orthogonal Array-Based Latin Hypercubes,” *J. Am. Stat. Assoc.*, 88, 1392–1397.
- Todorović, P. and Zelenhasić, E. (1970), “A stochastic model for flood analysis,” *Water Resources Research*, 6, 1641–1648.

- Voight, B., Sparks, R. S. J., Miller, A. D., Stewart, R. C., Hoblitt, R. P., Clarke, A. B., Ewart, J., Aspinall, W. P., Baptie, B., Calder, E. S., Cole, P. D., Druitt, T. H., Hartford, C., Herd, R. A., Jackson, P., Lejeune, A.-M., Lockhart, A. B., Loughlin, S. C., Luckett, R., Lynch, L., Norton, G. E., Robertson, R., Watson, I. M., Watts, R. B., and Young, S. R. (1999), “Magma Flow Instability and Cyclic Activity at Soufrière Hills Volcano, Montserrat, British West Indies,” *Science*, 283, 1138–1142.
- Welch, W. J. (2003), “GaSP (Gaussian Stochastic Process),” Personal communication.
- Williams, B. J., Santner, T. J., and Notz, W. I. (2000), “Sequential design of computer experiments to minimize integrated responses functions,” *Stat. Sinica*, 10, 1133–1152.
- Wolpert, R. L. and Ickstadt, K. (2004), “Reflecting Uncertainty in Inverse Problems: A Bayesian Solution using Lévy Processes,” *Inverse Problems*, 20, 1759–1771.
- Wolpert, R. L. and Taqqu, M. S. (2005), “Fractional Ornstein-Uhlenbeck Lévy Processes and the Telecom Process: Upstairs and Downstairs,” *Signal Processing*, 85, 1523–1545.