Computational Strategies in Uncertainty Quantification for Hazard Mapping

Regis Rutarindwa
Marquette University

COMPUTATIONAL STRATEGIES IN UNCERTAINTY QUANTIFICATION FOR HAZARD MAPPING

by

Regis Rutarindwa

A Dissertation submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

Milwaukee, Wisconsin

December 2017

ABSTRACT
COMPUTATIONAL STRATEGIES IN UNCERTAINTY QUANTIFICATION FOR HAZARD MAPPING

Regis Rutarindwa

Marquette University, 2017

There are many hazards associated with volcanic activities. Amongst them are Pyroclastic flows; a mixture of rock fragments, debris and hot gases that flow down the slope of actives volcanoes at high velocities. These flows have proven to be devastating, and at the same time more than 500 millions people in the world live within potential exposure to such a hazard. A few approaches have been used to try to mitigate the impact of volcanic hazard in general. These include remote sensing technology and developing hazard maps – a graphic representation of safe and risky zones for a given volcanic area.

In this dissertation, we develop a workflow for fast creation of accurate hazard maps. We apply this workflow on the case of the Long Valley volcanic region in northern California (USA). We have also made a couple of contributions that, while pertinent to the problem at hand, also have merit in a wide range of applications. First, we develop a Hierarchical Bayesian model that combines data on Pyroclastic flow behavior from various volcanic sites into a "global" dataset and reduces predictive uncertainty at volcanoes with sparse data. Of particular interest to us is the uncertainty in key input variables for computer simulations of Pyroclastic flows. Secondly, we develop a learning algorithm for experimental resource allocation in the case where multiple objectives need to be achieved simultaneously. This algorithm allows us to compute probability of hazard for multiple locations at the same time, and vastly reduce the time it takes to create hazard maps. These two contributions form the basis of a tool for geo-scientists to rapidly assess risk spatially at a moment notice, and provide hazard maps that can be used as a teaching tool for communities at risk.

ACKNOWLEDGMENTS

Regis Rutarindwa

I would like to thank, from the bottom of my heart, my parents and my siblings for the help and sacrifices they made to help me through this journey. I owe special thanks to my good friends Meghan – for all the care packages and upbeat messages – and Nathalie – for helping me cross the finish line. I would not have been able to accomplish this without the much-needed distraction and support from my partners in crime (Renato & Fidele), all my friends (too many to list here), and my graduate school colleagues Casey, Kawsar, Adib, Prachi.

I would like to extend my sincere appreciation to Dr. Jim Berger, Dr. Robert Wolpert, Dr. Abani Patra, Dr. Bruce Pitman, Dr. Marcus Bursik, Dr. Greg Valentine and the rest of my research group for all the valuable help, comments and insights on my work. I would also like to thank Dr Stephen Merrill, Dr Medhi Maadooliat and Dr Ting Lin for serving on my doctoral committee and helping me to polish my dissertation. This work would have been hard to realize without the support of NSF grants SES 1521855, DMS 1228265, and EAR 1331353.

Last but not least, my adviser and mentor, Dr Elaine Spiller – Thank you, Thank you, Thank you …

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

There are many hazards associated with volcanic activity; these include lava flows, tephra/ash fall, volcanic gases, lahars [1] , volcanic landslides, and pyroclastic flows (See Figure 1.2). According to a historical review of volcanic event reports by Doocy et al., these hazards globally resulted in 91,789 deaths, 14,068 injuries and 4.72 million people affected from 1900 to 2009. On top of this, more than 500 millions people live within potential exposure to a volcanic hazard, with this number set to increase as world's population increases [40]. These catastrophic events also lead to significant economic loss and invaluable infrastructure damages.

In trying to mitigate the impact of these hazards, several approaches are used in concert. These include continuously monitoring high-risk volcanic areas using a wide variety of sensors and satellite images, and training programs to educate and raise the awareness of the public, policy-makers, and emergency planners [74]. Along the lines of educational programs is developing hazard zoning maps that provide a visual representation of safe and risky areas in a region of interest. Such tools have been successfully used for pyroclastic flows at Soufrière Hills Volcano [8, 9, 110], lahars at Mount St. Helen [72] and Mount Rainier [59], and lava flows at Mount Etna [47] and Mount Vesuvius [105]. Figure 1.1 shows one example of such hazard map for the case of pyroclastic flows originating from Soufrière Hills Volcano on the island of Montserrat (UK) [9].

In this work, we focus on pyroclastic flows. These are volcanic avalanches made up of a mixture of ash, rock fragments, boulders, and gas. They can be as hot as 1,500°F, and move at speeds upwards of 150 miles per hour [78]. Pyroclastic flows can

[1] Lahars: Wet mixture of water, rock fragments and debris that flows down slopes of active volcanoes.

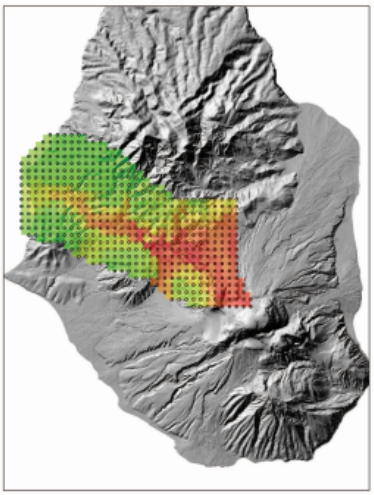


Figure 1.1: Probabilistic hazard map of the Belham valley North-West of Soufrière Hills Volcano . The color represents probability of inundation ranging from 0 (green) to 1 (red)

be quite massive, in the order of 10^6 to 10^{10}m^3 of material (rock, ash, boulders, etc.). Small flows can travel long distances depending on their compositions and the topography of an area, and larger flows are known to leave deposits as thick as 200 meters deep. There are many ways for pyroclastic flows to form: collapse of an eruption column leading to hot gases and volcanic ash moving down the flank of a volcano [109], "boiling over" from an explosive eruption [116], and the collapse of a growing unstable lava dome [76]. Regardless of how they are formed, they can be very deadly. In fact, pyroclastic flows alone accounted for roughly 75% of all volcanic hazard related deaths between 1900 and 2009 [40].

1.1 Dissertation Contribution

In this dissertation, we devise a strategy for fast and flexible generation of hazard zoning maps, and create probabilistic hazard maps of the Long Valley volcanic region in California. In part, we achieve this by combining a high-fidelity physical model of pyroclastic flows, TITAN2D , with Gaussian-Process based emulators, GaSP, to estimate the probability of inundation of locations *conditioned* on an event of a particular volume happening.

1.2 Dissertation Organization

The following work is split in three chapters each with a separate objective but all leading to the goal described in the previous section.

- **Chapter II:** Hierarchical Bayesian Analysis of Pyroclastic Density Current Mobility Metrics

  In this chapter, we develop a hierarchical Bayesian model to combine the mobility data (basal friction angle & volume) of several volcanoes and improve our knowledge of the relationship between a pyroclastic flow volume and its associated basal friction angle. This allows us to accurately constrain input parameters for TITAN2D simulations and quantify the uncertainty associated with those parameters. We apply this technique on data from Volcán de Colima, Mount Merapi, Soufrière Hills, Mount Unzen all of which have significant amount of data and drastically improve predictive capability at Mount Semeru which has limited data.

- **Chapter III:** Probabilistic Inundation Maps of the Long Valley Volcanic Region

  In this chapter, we create probabilistic hazard maps of the Long Valley volcanic region of California. We use the hierarchical Bayesian model developed in the previous chapter to determine the input distribution the volume and basal friction angle parameter for TITAN2D runs. In the framework we develop here, computing probabilities of inundation becomes a post-processing step of TITAN2D simulation output. The novelty of this strategy is the fact that it allows for easy plug-in of different aleatoric variability models of locations and volume. This enables us explore various different scenarios without making any additional runs of TITAN2D .

- **Chapter IV:** Multi-Objective Adaptive Design of Experiments For Hazard Mapping

In this chapter, we develop an algorithm that allows us to cleverly combine TITAN2D simulations runs we already have and vastly reduce the number of simulations needed to get better estimates of inundation contours – the key to fast probability calculations in Chapter III – for multiple locations. The breakthrough of this method lies in devising a strategy for resource allocations, in this case TITAN2D runs, when having to satisfy multiple objectives, in this case contour estimates.

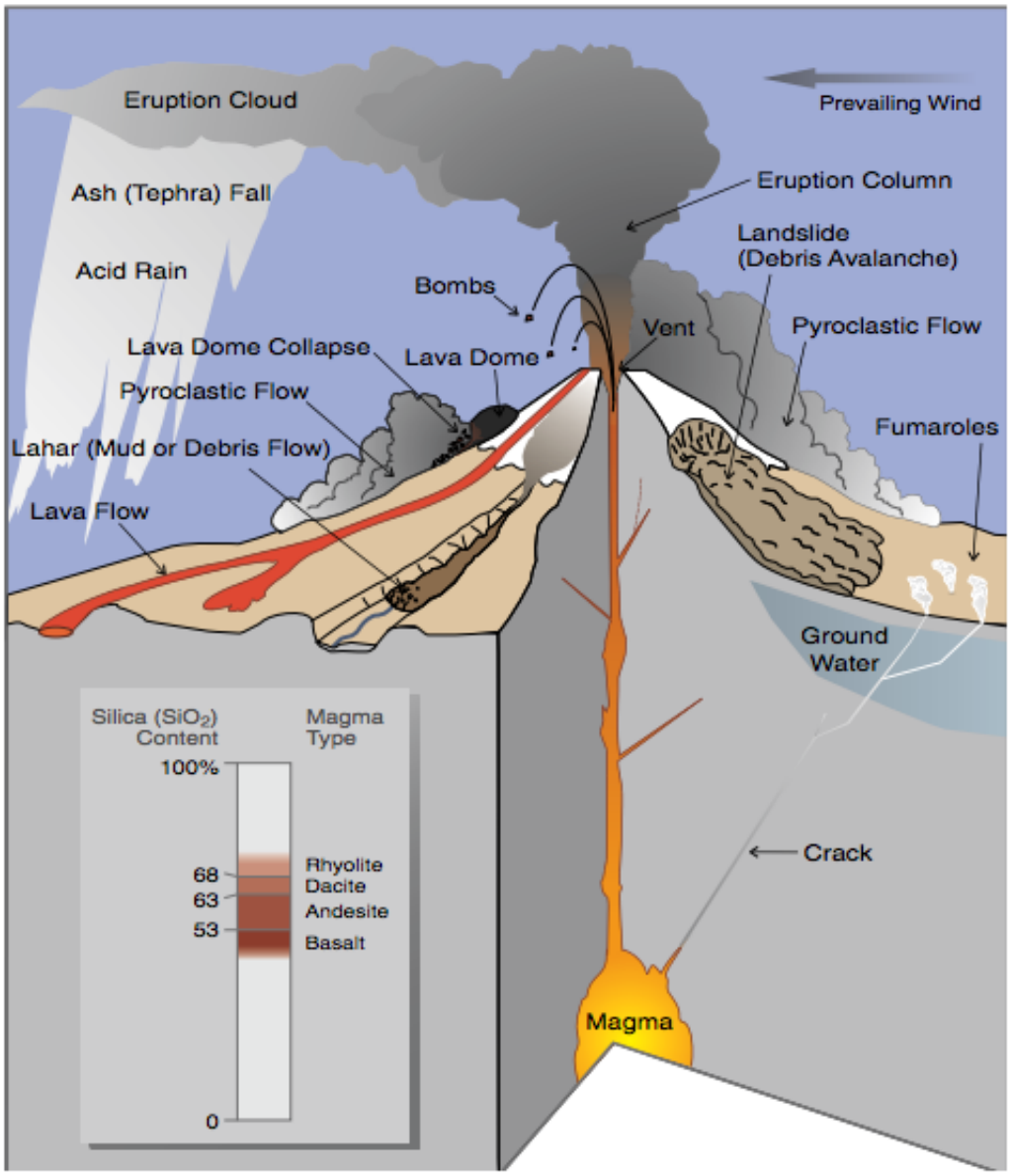


Figure 1.2: Illustration of the various hazard caused by volcanism. Lahars and landslides can occur even without an eruption

[2]Graphics courtesy of U.S. Geological Survey. https://pubs.usgs.gov/fs/fs002-97/fs002-97.pdf

Chapter 2

Hierarchical Bayesian Analysis of Pyroclastic Density Current Mobility Metrics

2.1 Introduction

A number of projects have been undertaken to create databases of volcanic data for quantitative volcanic hazard assessment (QVHA). Examples of these include the Global Volcanism Program database [45], WOVOdat [121], Geologic Survey of Japan (GSJ) Quaternary and Active volcanoes databases [44], LaMEVE [35], DomeHaz [83, 86], and FlowDat [84, 85]. As individual datasets provide limited information on their own, new avenues are explored to make use of the global record of volcanic data. The projects mentioned above enables us to exploit the notion that combining volcanic datasets via statistical methods, for comparative studies or further research, is vital to understanding volcanic phenomena.

Despite these efforts, scientists still deal with the issue of data sparsity at some volcanic sites. This can be caused by either past flows not being adequately exposed due to vegetation or human settlements nearby [35, 17, 64, 125], or the remoteness of deposits and how it impacts data collection [125]. In addition to this, it has been observed that small effusive flows are frequently poorly recorded [39, 43, 107, 35, 17], and large flows from newly active volcanoes are often misunderstood because of the lack of data from these previously dormant volcanoes [2, 124].

Broadly speaking, there are two "extreme" ways to deal with data sparsity. First, one can assume that a volcanic process is similar for every site, and thus use global volcanic data for analysis. Alternatively, it can be assumed that a volcanic phenomenon is different at every sites and one uses only the data available for each

site for analysis. However, a more sensible approach is to assume that while a physical phenomenon has the same characteristics across all volcanoes, it can differ from site to site by some *slight* difference that we can try quantify with a probability distribution. Hierarchical Bayesian models allow us to *borrow* information from various different sites, draw insights from the aggregate data and improve predictive capabilities at individual sites with insufficient data [101].

In this chapter, we apply this last approach on mobility metrics of various volcanoes to improve our understanding of pyroclastic flow hazards. Mobility metrics play a crucial role in hazard mitigation and forecast of possible flow paths. Specifically, they are commonly used in empirical models such as PFz [30, 126], or computational flow models such as TITAN2D [87]. Normally, linear regression of these mobility metrics such as the Heim coefficient (height drop/ runoff length, H/L of a pyroclastic flow) are used as direct input [30, 126], or proxy input for computational models ($e.g.$ basal friction angle [1] for the case of TITAN2D). With this approach, models have proven successful at replicating historical flows [28, 77, 106, 24, 85]. However, the dearth of data at some volcanic sites renders the task of forward modeling with approach complicated.

We develop a systematic method to extract as much information from the limited data available on mobility data (Heim coefficient) while quantifying the uncertainty associated with each related input. We develop a hierarchical bayesian model to pool mobility metrics data of pyroclastic flows from different volcanoes and improve our characterization of the relationship between basal friction and volume at volcanoes with sparse data. The objective of this study is to be able to quantify the uncertainty associated with a flow's basal friction angle and volumes, which are used as inputs for TITAN2D . We used the global record of mobility metrics data from the FlowData mass flow database [84, 85] (See Appendix B).

[1] Basal Friction Angle = $\tan^{-1} H/L$

First, it is useful to consider some background to the problem of assessing mobility of pyroclastic density currents (PDC) and how their mobility metrics are used with, and subsequently propagated through, flow modeling, is presented in Section 2.2. The hierarchical Bayesian analysis of the compiled data is presented in 2.3, and the results discussed in Section 2.4.

2.2 Mobility metrics for mass flows

2.2.1 Frictional vs. resisting shear stress models

The most widely used mobility metric for concentrated mass flows of (e.g. volcanic and non-volcanic debris avalanches, dome- and column-collapse PDC) is the Heim coefficient [50], commonly denoted as H/L, where H is the vertical fall height traversed by a flow and L is the runout length. H/L is equivalent to the coefficient of friction following a Mohr-Coulomb friction model, in which shear stress at the initiation of failure is proportional to the normal stress.

According to Mohr-Coulomb friction models, the mass or volume V, of the flow should be irrelevant to mobility, and the coefficient of friction should be a function of material properties. Numerous studies of real deposits, however, have shown a linear inverse relationship between $\log(V)$ of a mass flow (of any type) and $\log(H/L)$ [50, 98, 97, 55], with large volume flows demonstrably being more mobile than small volume flows. It is clear then that capturing this behavior is essential for accurate hazard forecasting of PDC.

An alternative to the frictional model approach is that of the constant resisting shear stress models. In these models, the mobility of mass flows is described by a constant resisting shear stress (CRS), or yield strength, and planimetric area, A_p, is related to $V^{2/3}$ via scaling arguments [56, 59, 37, 23]. This model indicates a relationship between inundated area and resisting shear stress, suggesting a yield stress rheology [63, 34, 46].

Both these metrics (H/L vs. V and A_p vs. $V^{2/3}$) have been applied to PDC mobility with success [112, 79, 42, 102, 12, 41, 49, 23, 31, 119, 27] and have become standard mobility metrics with which to compare and contrast PDC behavior.

2.2.2 Mobility metrics for flow modeling

Many empirical flow inundation models are based directly on measurements of H/L vs V or A_p vs. $V^{2/3}$. This concept has been applied at a variety of volcanoes [105, 123, 54, 1, 103] and also forms the basis for the FLOW2D and FLOW3D computer models [67, 38, 104, 53] which base shear resistance on basal friction (taken directly from H/L), viscosity, and turbulence.

H/L also informs computational flow models that use a Coulomb friction law, including TITAN2D [87], which have built upon the work of [96], who used Coulomb friction laws in conjunction with depth-averaged equations for mass and momentum (See equation 3.6). The Heim coefficient can therefore provide a guideline for choosing appropriate basal friction input angles for different flow volumes for TITAN2D [82, 85, 28, 29].

LAHARZ and PFZ use semiempirical equations for planimetric area ($A_p = cV^{2/3}$) and cross-sectional area ($A_{xs} = CV^{2/3}$) to predict lahar [59], debris flow, rock avalanche [46] and PDC [126] inundation using empirically derived coefficients (c and C) from a variety of mass flow deposits worldwide. These relationships also form the basis of flow models using constant shear stress instead of constant friction [62].

With increasing application of these respective flow modeling approaches, it is now timely and appropriate to undertake more considered approaches to understanding and quantifying the uncertainty related to the use of mobility metrics as model inputs. This work has been driven by our specific interest in constraining the basal friction input parameter required by TITAN2D when undertaking ensemble runs for generating probabilistic hazards maps [8, 110, 9], by using the H/L-volume mobility

relationships for block-and-ash flows from the FlowDat database. The application of the method developed can, however, be applied widely.

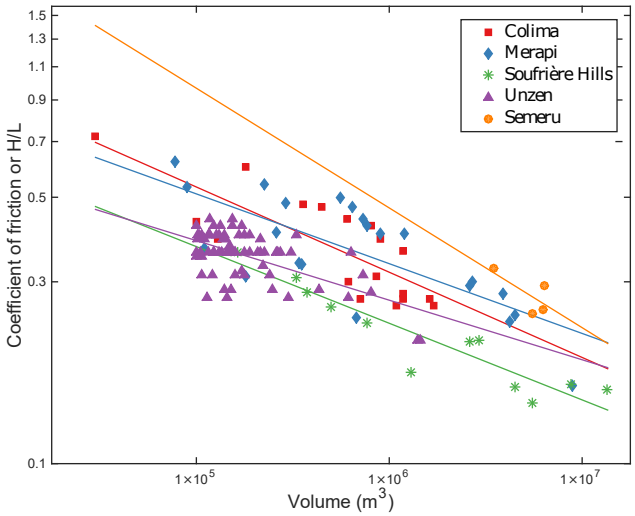


Figure 2.1: Data from all volcanoes considered for each of the two countries relationships along with their respective linear regression lines. Colima, Merapi, SHV, and Unzen volcanoes have plentiful data, while the data set for Semeru is sparse

2.3 Statistical analyses

Herein, we present a method using hierarchical Bayesian modeling to leverage the global record of mobility metrics for PDCs, which can aid in cases where data for a particular volcano is sparse. We use the FlowDat database of mass flow mobility metrics [84, 85], which is current through 2014. From FlowDat [84], 4 volcanoes were selected with plentiful (14 to 80 flows) H/L and volume data for dome-collapse PDC: Colima (Mexico) [94, 93, 95], Merapi (Indonesia)[14, 15, 99, 27, 26, 65], Soufrière Hills Volcano (Montserrat) [23, 31, 48, 66, 70, 32], and Unzen (Japan) [80, 115] (Figure 2.1). Semeru volcano (Indonesia) was also included due to the scarcity of the data available [117]. These flows are all dense, concentrated dome-collapse PDC (block and ash flows), for which it is reasonable to assume broadly similar flow behavior. Error was rarely reported by the sources of the data, but is shown as error

bars where available. However, the error bars were often smaller than the markers themselves.

For the frictional model of mobility (H/L vs. V), the strong linear relationship between the logarithm of PDC volume and the logarithm of the coefficient of friction (discussed in section 2) suggests the use of a linear model, such as a regression model

$$y = \alpha + \beta x + \epsilon, \qquad \epsilon \overset{iid}{\sim} N(0, \sigma^2)$$

where x is the log-volume[2]), y is the log-coefficient of friction (H/L), α and β are the intercept and slope of the regression line, and ϵ is random error. Graphically, this model would correspond to fitting a straight line through all of the data $\boldsymbol{y}$ in Figure 2.1 which minimizes the errors between estimated and observed values. This approach corresponds to one end-member option, that is, to assume that the relationship between the coefficient of friction and flow volume for block-and-ash flows is constant at every volcano, and thus use information from all the volcanoes to fit a regression.

Alternatively, one could fit separate regression lines for each of the J volcanoes, namely

$$y_j = \alpha_j + \beta_j x + \epsilon, \qquad \epsilon \overset{iid}{\sim} N(0, \sigma_j^2)\,,$$

based on the data $\boldsymbol{y}_j$ from volcano j alone. The result of separate regression fits is shown in Figure 2.1. This approach represents the alternative end-member option, that is, to assume that the relationship between the coefficient of friction and volume at different volcanoes is unrelated and thus use only the information from a given volcano to fit a regression.

The hierarchical analysis presented in the next section could prove useful for *any* linear relationship suggested by transformations of volcanic datasets – the frictional relationships for dome-collapse PDCs used here is just one example.

[2]Actually $x = \log_{10}(\text{volume}/10^{5.5})$. This x-origin then corresponds volume of $10^{5.5}\text{m}^3$, roughly where the slope and intercept are least correlated.

2.3.1 Hierarchical Bayesian model

In situations such as this, where it is unclear whether to fit an overall regression or separate regressions, it has become common statistical practice to use the hierarchical or multilevel approach, which is a happy medium between these end-member alternatives. Hierarchical modeling is carried out via Bayesian analysis, wherein a *prior* probability distribution is chosen to describe knowledge about the unknown model parameters (here the various regression parameters); this distribution will then be updated by the data to form *posterior* probability distributions of the unknown model parameters.

The version of hierarchical modeling that we utilize here links together the separate regressions by assuming that the regression line slopes arose from the common normal distribution (part of the prior distribution)

$$\beta_j \quad \text{are i.i.d.} \quad N(\mu, \tau^2)\,,$$

with unknown hyper-mean (the mean of the prior distribution) μ and hyper-variance (the variance of the prior distribution) τ^2. Note that, if $\tau^2 = 0$, then all the β_j would be equal, so we would be back to the case of a single regression. At the other extreme, as $\tau^2 \to \infty$, this model would yield the same answers as the separate regression models. The performance of the hierarchical model, in situations such as this, is typically better than that of either of the two extremes.

Initially we will presume that little is known about μ and τ^2 (a vague prior distribution will be used for these parameters), but we will learn about them from the data through their posterior distribution and they, in turn, will affect the posterior distribution of the β_j.

If data were plentiful at each volcano, there would be little need (but also no harm) in employing the hierarchical model, as the effect of the posterior distribution of

μ and τ^2 on the β_j would then be minimal. When data is sparse for one or more volcanoes, however, the gains with the hierarchical approach can be considerable. For instance, from the left panel of Figure 2.1 it can be seen that there are only four data points from Semeru for a very narrow range of PDC volumes, and attempting to fit a separate regression to just four points will lead to a very uncertain result. In contrast, the hierarchical modeling approach allows for 'borrowing strength' from the other volcanoes in estimating Semeru's regression line slope (because of the assumption that all slopes arose from a common normal distribution), and will be seen to result in much tighter credible intervals for the regression line for Semeru.

To complete the specification of the hierarchical model, we need to also choose 'prior' distributions for the other unknown parameters in the model. Whereas the regression coefficients from Figure 2.1 appear quite related, the intercepts, α_j, seem considerably more variable. We could utilize a hierarchical model for the intercepts but, since there will be little gain, we instead employ an objective constant prior distribution $\pi^O(\alpha_1, \ldots, \alpha_J) = 1$; although this objective prior does not induce any sharing of intercept information across volcanoes, the changes in the slope parameters through their hierarchical modeling will influence the intercepts.

In developing prior distributions for the regression variances σ_j^2, it is important to consider that the PDC data represented in Figure 2.1 contain data from both highly channelized and unchannelized (unconfined) flows, which experience different frictional forces and exhibit different mobilities [85, 27, 114]. Modeling by Stinton [114] using TITAN2D, showed that flows confined in synthetic channels had longer runouts, higher velocities, and shorter travel times than flows simulated over synthetic unconfined terrain.

This difference is also apparent in the data. Indeed, Table 2.1 gives the results of separate regressions at the five volcanoes, and the mean square residual (MSR) are very similar for the three volcanoes with dominantly channelized flows and much

smaller than the MSR for the volcanoes with dominantly unchannelized flow deposits. The higher MSR for unchannelized flows or those that inundate multiple channels makes intuitive sense, as these flows travel over extremely varied topography, with greater variation in slope and surface roughness than flows which travel down channels.

Volcano	**Lin. Reg. Slope**	**Lin. Reg. Intercept**	**MSR** ($\times 10^{-4}$)
Colima*	-0.224	-0.386	66.5
Merapi*	-0.183	-0.384	95.2
SHV	-0.201	-0.531	24.8
Unzen	-0.156	-0.493	26.3
Semeru	-0.314	-0.172	24.3

Table 2.1: Linear regression parameters and MSR for each volcano for H/L vs. V relationship. The (*) marks volcanoes with flows which are generally **unchannelized**

Therefore, it would be natural to have a separate variance for the channelized and the unchannelized flow data. We, thus, assign Merapi and Colima a common variance σ_C^2 and the other volcanoes common variance σ_U^2, with the two variances being unknown.

To complete the Bayesian model, prior distributions are needed for σ_C^2 and σ_U^2 and for the hyperparameters μ and τ^2 from the hierarchical prior. For these parameters we utilize a standard objective prior – called the *reference prior* – $\pi^R(\mu, \sigma_\beta^2, \sigma_C^2, \sigma_U^2)$. The reference prior is chosen so as to minimize the influence of the prior distribution on the analysis, i.e., to ensure that the posterior distribution of the model parameters only reflects what the data has to say.

With this, one now simply applies Bayes theorem to obtain the distribution of all unknowns parameters, given all the data $\boldsymbol{y}$

$$\pi(\alpha_1,\ldots,\alpha_J,\beta_1,\ldots,\beta_J,\mu,\tau^2,\sigma_C^2,\sigma_U^2 \mid \boldsymbol{y}) \propto \prod_{j=1}^{J} f(\boldsymbol{y}_j \mid \alpha_j,\beta_j,\sigma_j^2) \\ \times\pi^O(\alpha_1,\ldots,\alpha_J)\pi^R(\mu,\sigma_\beta^2,\sigma_C^2,\sigma_U^2)\prod_{j=1}^{J} N(\beta_j \mid \mu,\tau^2)\,, \tag{2.1}$$

where $f(\boldsymbol{y}_j \mid \alpha_j,\beta_j,\sigma_j^2)$ is the likelihood arising from the data at volcano j and the σ_j^2 are either the channelized or unchannelized variance.

2.3.2 Analysis

There are no closed form analytical expressions for estimates of unknown parameters or for credible intervals, but there is a relatively straightforward Markov Chain Monte Carlo (MCMC) method – described below – for drawing samples from the posterior distribution in (2.1). From this set of samples,

$$\{(\alpha_1^i,\ldots,\alpha_J^i,\beta_1^i,\ldots,\beta_J^i,\mu^i,(\tau^2)^i,(\sigma_C^2)^i,(\sigma_U^2)^i), \quad i=1,\ldots,m\}\,,$$

all desired inferences can be performed.

The typical parameter estimate would be the posterior mean, computed as the average of all of the samples; enough samples are typically chosen ($m = 10^6$ was used in the computations herein) that the numerical error in this computation is negligible. Similarly a 95% credible interval, for example, would be formed as the interval containing the central 95% of the ordered sample. Even more informatively, the entire posterior distribution of a parameter could be approximated by simply making a histogram of the sample values. These histograms are illustrated in Figure 2.2 and 2.3

The technical details of the MCMC method used in the hierarchical Bayesian analysis are given herein. First, some notation: write the design matrix for the j^{th} regression (i.e., the intercept constant 1 and the transformed volume input values) as

(with n_j being the number of observations for Volcano j)

$$\boldsymbol{X}_j = \begin{pmatrix} 1 & x_{j1} \\ 1 & x_{j2} \\ \vdots & \vdots \\ 1 & x_{jn_j} \end{pmatrix},$$

and define (recalling that the σ_j^2 are σ_C^2 or σ_U^2 for the channelized and unchannelized volcanoes respectively)

$$\bar{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ji} \qquad \lambda_j = \frac{\tau^2}{\sigma_j^2} \qquad v_j = v_j(\sigma_j^2, \tau^2) = d_j + \tau^2$$

$$S_j = \sum_{i=1}^{n_j}(x_{ji} - \bar{x}_j)^2 \qquad d_j = \frac{\sigma_j^2}{S_j} \qquad \boldsymbol{v} = (v_1, \ldots, v_J)$$

$$\begin{pmatrix} \hat{\alpha}_j \\ \hat{\beta}_j \end{pmatrix} = (\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1}\boldsymbol{X}_j'\boldsymbol{y}_j \qquad n = \sum_{j=1}^{J} n_j \qquad \hat{\mu}(\boldsymbol{v}) = \frac{\sum_{j=1}^{J} \hat{\beta}_{j2}/v_j}{\sum_{j=1}^{J} 1/v_j}$$

From [11], the objective reference prior for the parameters $(\mu, \tau^2, \sigma_U^2, \sigma_C^2)$ is

$$\pi(\mu, \tau^2, \sigma_U^2, \sigma_C^2) \propto \left(\frac{1}{\sigma_U^2\sigma_C^2}\right)\sqrt{\sum_{j=1}^{J}\frac{1}{v_j^2(\sigma_j^2, \tau^2)}}\,.$$

Then a Gibbs sampler [25] can be constructed as follows, to draw samples from the posterior distribution in (2.1).

Step 1. Draw the β_j, given σ_j^2, μ and τ^2, from the following distribution:

$$N\left(\hat{\beta}_j - \frac{(\hat{\beta}_j - \mu)}{1 + \lambda_j S_j}, \quad \frac{\sigma_j^2\lambda_j}{1 + \lambda_j S_j}\right).$$

This is the marginal posterior distribution of β_j, given σ_j^2, μ and τ^2 (i.e., α_j has been integrated out). Note that we could have also integrated out μ, but that should not be necessary because below we generate μ from its marginal

posterior distribution with the βs integrated out.

Step 2. Draw the α_j, given σ_j^2 and β_j, from the $N\left(\hat{\alpha}_j - \bar{x}_j(\beta_j - \hat{\beta}_j),\ \ \sigma_j^2/n_j\right)$ distribution. This is the conditional posterior distribution of α_j, given σ_j^2 and β_j. (It happens to not depend on τ^2 or μ.)

Step 3A. Propose a value of σ_U^2, given the $\{\beta_j\}, j = 1, 2$, by drawing a random variable from the inverse gamma distribution with shape parameter $\alpha_U = (n_1 + n_2)/2$ and rate parameter

$$\beta_U = \frac{1}{2}\sum_{j=1}^{2}(y_{ji} - [\alpha_j + x_{ji}\beta_j])^2\,.$$

Draw a uniform random variable U on $(0, 1)$ and accept the proposed σ_U^2 if

$$U < \frac{\sqrt{\sum_{j=1}^{J} 1/v_j^2(\sigma_j^2, \tau^2)}}{\sqrt{\sum_{j=1}^{2} 1/v_j^2(0, \tau^2) + \sum_{j=3}^{5} 1/v_j^2(\sigma_j^2, \tau^2)}}\,;$$

else discard σ_U^2 and propose a new σ_U^2, repeating as necessary until a σ_U^2 is accepted. This arises from the standard accept-reject algorithm because the numerator above, which is the unnormalized ratio of the target posterior distribution and the inverse gamma proposal distribution, is maximized at $\sigma_U^2 = 0$.

Step 3B. Propose a value of σ_C^2, given the $\{\beta_j\}, j = 3, 4, 5$, by drawing a random variable from the inverse gamma distribution with shape parameter $\alpha_C = (n_3 + n_4 + n_5)/2$ and rate parameter

$$\beta_C = \frac{1}{2}\sum_{j=3}^{5}(y_{ji} - [\alpha_j + x_{ji}\beta_j])^2\,.$$

Draw a uniform random variable U on $(0, 1)$ and accept σ_C^2 if

$$U < \frac{\sqrt{\sum_{j=1}^{J} 1/v_j^2(\sigma_j^2, \tau^2)}}{\sqrt{\sum_{j=1}^{2} 1/v_j^2(\sigma_j^2, \tau^2) + \sum_{j=3}^{5} 1/v_j^2(0, \tau^2)}};$$

else discard σ_C^2 and draw a new σ_C^2, repeating as necessary until a σ_C^2 is accepted. The rationale is as in Step 3A.

These steps yield draws from the conditional posterior distributions of σ_U^2 and σ_C^2, given the $\{\alpha_j, \beta_j\}$, and do not depend on the other parameters.

Step 4. Draw μ, given the σ_j^2 and τ^2, from the following distribution:

$$N\left(\hat{\mu}(\boldsymbol{v}), \quad \frac{1}{\sum_{j=1}^{J} 1/v_j}\right).$$

This is the marginal posterior distribution of μ, given the σ_j^2 and τ^2, i.e., all the βs have been integrated out.

Step 5. Generate τ^2, given μ, the $\{\beta_j\}$ and the σ_j^2, by the following accept-reject algorithm:

- Generate τ^2 from the inverse gamma distribution with shape parameter $\alpha = (J-2)/2$ and rate parameter $\beta = \frac{1}{2}\sum_{j=1}^{J}(\beta_j - \mu)^2$.
- Draw a uniform random variable U on $(0, 1)$ and accept τ^2 if

$$U < \frac{\sqrt{\sum_{j=1}^{J} 1/v_j^2(\sigma_j^2, \tau^2)}}{\sqrt{\sum_{j=1}^{J} 1/v_j^2(\sigma_j^2, 0)}};$$

  else discard τ^2 and draw a new τ^2, repeating as necessary until a τ^2 is accepted.

This algorithm follows from noting that the likelihood for τ^2, given all the other

parameters, is proportional to the given inverse gamma distribution. The posterior distribution of τ^2, given all the other parameters, is then proportional to this likelihood times the prior; a sample is then drawn from this posterior using accept/reject with the likelihood as the proposal distribution.

To *view* samples from the posterior and assess that the MCMC algorithm is behaving properly, we consider histograms and trace plots, respectively. Trace plots illustrate the entire sequence of samples from the posterior distribution, or chain, (after the first few thousand are discarded) with the value of the random variable plotted on the vertical axis vs. the sequence index. The reader unfamiliar with MCMC sampling should note that a well-mixing algorithm should not get "stuck" at one value for many samples, should not have too many vertical outliers, and should not have a discernible periodic envelope. Note, the samples (and trace plots) have been "thinned" keeping every fifth sample from the MCMC sequence.

Of particular interest are slope parameters for each volcano, β_j, illustrated for the frictional model in Figure 2.2. Histograms of slope parameter samples for each volcano give reassurance that we are sampling around a common slope, near -0.2. Spread in each individual histogram reflects the uncertainty of the slope parameter for each volcano. Of course, wider histograms indicate more uncertainty.

Samples for *any* of the unknown parameters described by the posterior distribution can be visualized in this manner. For example, one might be interested in estimating the inferential variance parameters for the two flow categorizations, channelized vs. unchannelized. Descriptive illustrations of these samples are presented in Figure 2.3. Of course, which unknown parameters are of particular interest is dependent on the scientific questions at hand for a given problem.

Next, we turn to use of the regression samples to draw the desired geophysical conclusions.

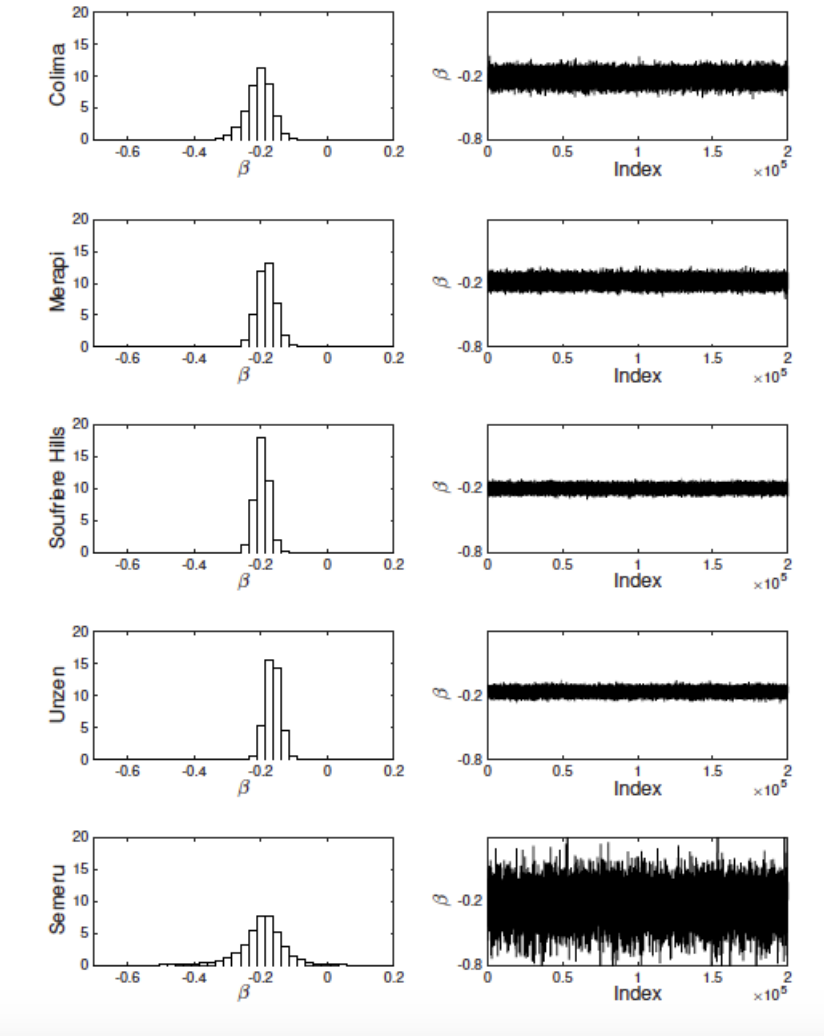


Figure 2.2: Left: Normalized histograms of sampled slopes for the frictional model for each of the five volcanoes considered. Right: corresponding trace plots from MCMC samples

2.4 Geophysical Results and Discussion

The coefficient of friction - volume relationship can be studied in several ways from the posterior sample of parameters. Thus, for volcano j, we have a sample $\{(\alpha_j^i, \beta_j^i),\ \ i = 1, \ldots, m\}$ of the intercepts and slopes. We thus immediately have a sample from the posterior distribution of all regressions lines, illustrated in Figure 2.4.

Samples of regression lines are useful for the computation of inundation

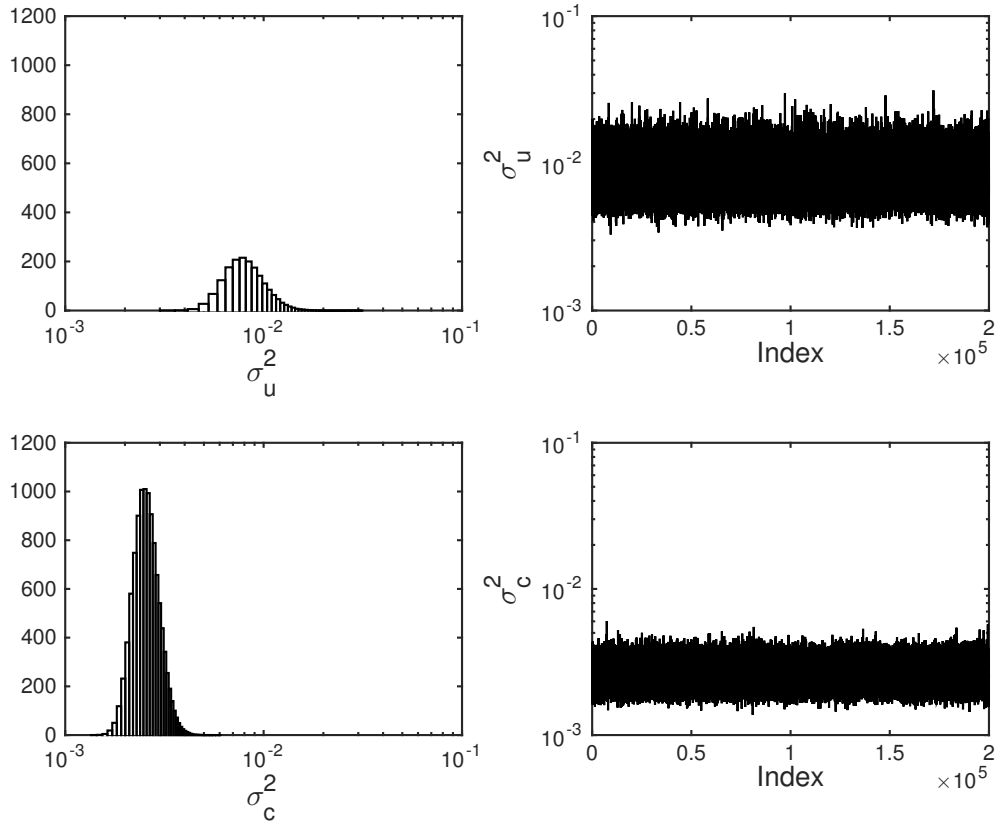


Figure 2.3: Left: Normalized histograms of the inferential variances, σ_u^2 (unchannelized, top) and σ_c^2 (channelized, bottom), for linear regression model applied to the frictional relationship, plotted on a log scale. Right: corresponding trace plots from MCMC samples

probabilities from PDCs; for example, where it is necessary to consider different possible mobilities for flows over a range of volumes. Samples from regression lines can be used directly for empirical models such as the energy line/cone method or for estimating the basal friction input parameter for a geophysical model like TITAN2D [8, 85, 110] or the constant resisting shear stress input parameter in VolcFlow [85]. Furthermore, using regression samples generated by this method allows one to account for uncertainty in probabilistic assessments of PDC inundation.

Figure 2.5 gives, for each volcano, a posterior summary consisting of the regression line corresponding to the posterior median values of the sample regression lines (the solid red line) – this would be the natural estimated regression line from the

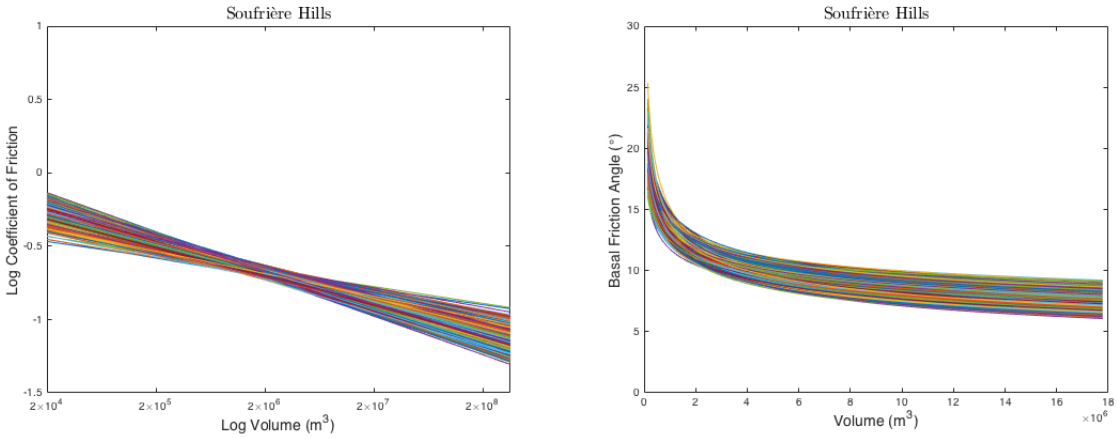


Figure 2.4: Both figures represent samples from the hierarchical linear regression model of the frictional relationship applied to the data, but show the same sample curves on different scales. Left: coefficient of friction and volume each on a log scale (which the linear model was fit to), Right: Basal friction angle (calculated as $\arctan$ of the coefficient of friction) versus volume on a linear scale ($1 \times 10^6 \mathrm{m}^3$)

Bayesian analysis. 95% credible intervals (the dashed red lines) are also shown and are obtained, at each volume value V, by taking the central 95% interval of values of $\alpha_j^i + \beta_j^i \log_{10}(V)$, over the posterior samples.

For comparison, the confidence intervals on the regression function from classical individual regressions are also given in Figure 2.5, with the solid black line being the standard estimated regression function and the dashed black lines being the standard 95% confidence intervals. As expected, for the volcanoes with abundant data, there is not much difference between the hierarchical model regression summaries and the classical regressions. But, for Semeru, which had only four data points all of which are closely clustered in volume, the differences found would affect the results of a probabilistic analysis, with the hierarchical approach providing tighter uncertainty estimates.

This type of approach is broadly applicable to other types of mass flows (debris avalanches, lahars, or column-collapse PDCs, for example) or other types of data entirely (ash-dispersion metrics, for example), but it is important that the datasets selected describe phenomena that are similar. This work focused only pyroclastic

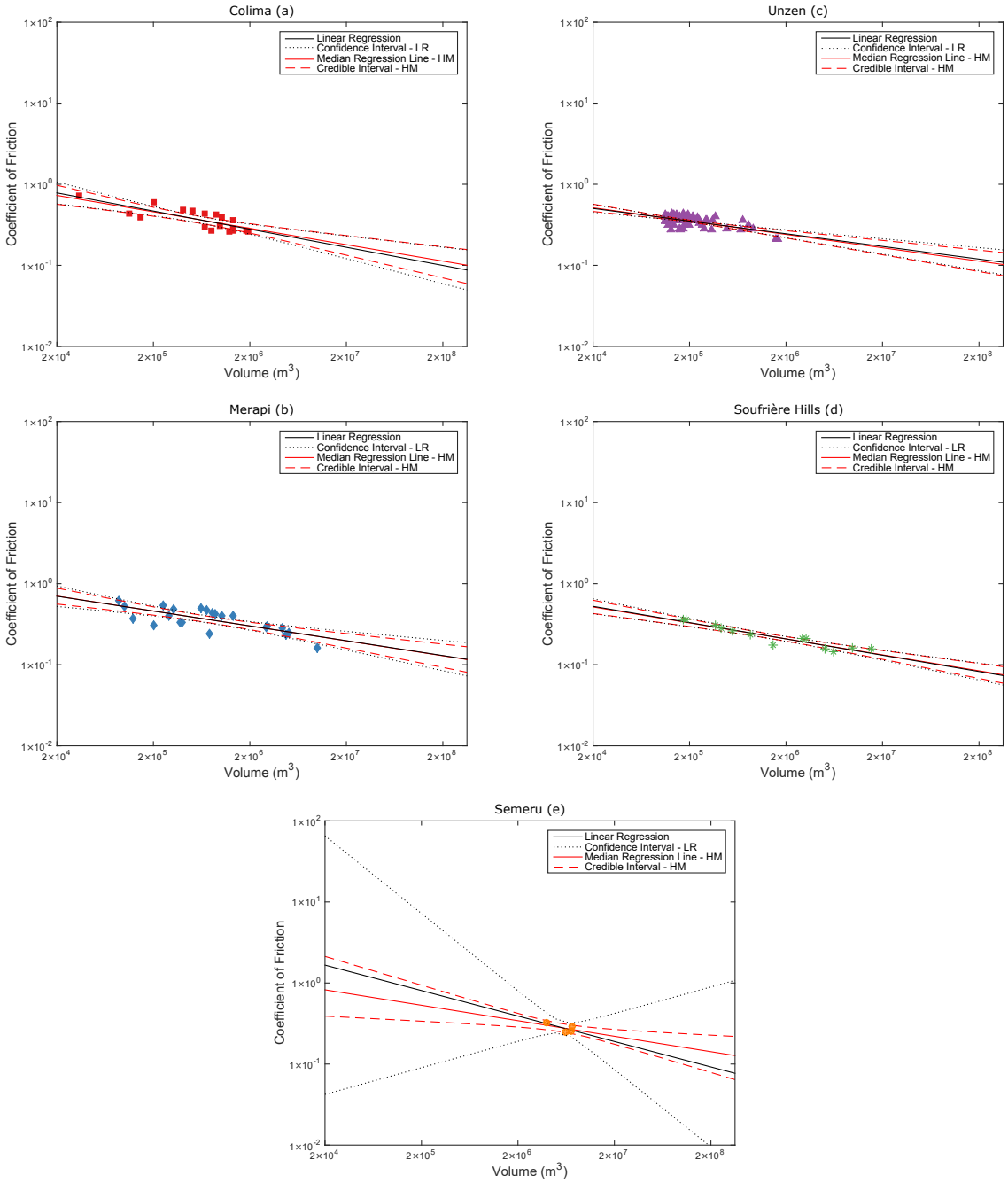


Figure 2.5: Comparison of the 95% confidence intervals (black dotted line) on the regression line for each individual volcano (black solid lines) and credible intervals (red dotted line) obtained from the hierarchical model (red solid line) as applied to the coefficient of friction vs volume relationship. PDCs in a & b were considered unchannelized; and c & d were considered channelized in this analysis. PDCs from Semeru (e) were also considered channelized, but with only four data points

density currents which are considered to have broadly similar emplacement dynamics; and accounted for dissimilarities (i.e. differences in channelization) by allowing for different variances. However, the more similar the phenomena at different volcanoes, the better the method is able to reduce uncertainty.

2.5 Conclusions

Understanding the past behavior of a particular volcano is the foundation upon which assessments of potential future hazards are based. However, complete and robust datasets are very rare, and really only exist for a handful of very well-studied volcanoes. Additionally, newly active volcanoes may produce hazards with poorly constrained characteristics due to limited data. This problem can be handled by: 1) using only the data from a particular volcano (which may be sparse, and thus introduces large uncertainties into hazard assessments); or 2) using the global record of volcanoes (which may ignore or downplay any particularities of the volcano in question). The hierarchical Bayesian method presented herein allows one to achieve a happy medium between these two approaches by "borrowing strength" from the global record of PDC behavior and thus greatly reducing the uncertainty for volcanoes with sparse data.

The modeling results obtained herein are used in the next chapter as guidelines to create experimental designs for large batches of computer experiments. The experiments will in turn be used to produce hazard maps of the Long Valley Volcanic region.

Chapter 3

Probabilistic Inundation Maps of the Long Valley Volcanic Region

3.1 Introduction

The Long Valley volcanic region is located in Northern California, east of the Sierra Nevadas. It comprises the Long Valley caldera, a volcanic depression 17 kilometers wide and 900 meters deep, and the Mono-Inyo craters, a series of more than thirty volcanic craters aligned from the western end of the caldera to Mono Lake at the northern end [108].

The volcanism in the area started more than 3 Ma ago, with the single largest event being the creation of the caldera $\sim$760 Ka ago from the eruption of 600 km^3 of magma [51] [1]. More recent activities include the eruptions of the Mono-Inyo craters $\sim$60 Ka [108, 22], the eruption of 1 km^3 of lava from the Mono craters 1325 – 1350 AD [108], and the eruption that created the Paoha Island in Mono Lake around 1700 AD [51].

In 1978, a M 5.8 earthquake beneath Wheeler Crest, southeast of the caldera, started a period of unrest characterized by earthquakes, ground uplift, and volcanic gas emissions at Mammoth Mountain that is still going on to this day [52]. This extended period of volcanic activities may be a precursor of a much larger future eruption with devastating impact on the nearby surrounding but also on a wider scale. In immediate danger are the town of Mammoth Lakes with 8000 inhabitants (along with its popular winter ski resort), the Casa Diablo power plant, and part of the aqueduct connecting the city of Los Angeles to the Lake Crowley reservoir. As a result of this, the U.S. Geological Survey has installed a network of sensors to monitor the area and has

[1] Ma (Mega-annum): $1,000,000$ years
Ka (Kilo-annum): $1,000$ years

described its threat potential as *Very High* [118].

In this chapter, we focus on the hazard of pyroclastic density currents, PDC or simply pyroclastic flows, in the Long Valley volcanic region. These flows can be devastating as they expose surrounding areas to extreme temperatures, ash falls and volcanic rock avalanches. We aim to quantify the probability of the hazard associated with these flows, and develop a tool to visually represent our results. Although pyroclastic flows give rise to many different hazard, we focus on inundation by the flowing mass. We describe a location as being inundated if there is more than a critical height of material, h_{crit}, present at that location at any time during the flow event. Our intent is that this tool, which can be used to differentiate between safe and risky areas, be used as a learning and experimental tool for policymakers and emergency planners to quickly assess the hazard threat posed by various catastrophic scenarios.

A rudimentary approach to create such a tool would be to determine the limit of previously observed flows and use that as a boundary between safe and risky areas [105]. However, this proves to be a complicated task because distal deposits are often hard to identify due to many factors such as human activities and vegetation [35, 17]. Also, activities are well described by heavy-tailed probability models. Thus, the assumption that *the future will look similar to the past* can seriously underestimate the threat of hazard.

We propose a simulation-based approach. We use a combination of physical model of pyroclastic flow (TITAN2D) and statistical emulators (GaSP) to develop a tool that takes into account the uncertainty of important input parameters, while at the same being computationally cheap to run. The hierarchical model we developed in the previous chapter helps us accomplish this by combining different sources of data to provide a better, more accurate, description of the uncertainty in the basal friction and volume input parameters that are needed for TITAN2D simulations. Gaussian Process Response Surface models (GaSP) are used as cheap surrogates of complex models. We

use GaSPs to explore the output of TITAN2D at untried inputs. The tool we develop allows for fast and easy plug-in of different scenario models and computation of probability of inundation *conditioned* on flow event's volume for all locations in the Long Valley region. Illustrative hazard maps are presented.

The chapter has the following organization. Section 3.2 gives an overview of the computational tools TITAN2D and GaSP models. Section 3.3 presents our datasets and the details of our algorithm. Finally, Section 3.4 presents hazard maps of the Long Valley volcanic region and a brief discussion of our findings.

3.2 Computational Tools

Various mathematical tools were used to generate these maps. First, we will present a brief description of these and how they fit together in the next 2 sections.

3.2.1 TITAN2D

TITAN2D is the primary physical model that we use to simulate the flow of pyroclastic density current. It is based on previous work by Savage & Hutter [96], Pitman et at. [89], Iverson et al. [57, 58], and Bursik et al. [21]. It solves a system of equations describing a depth-average model for granular flow governed by Mohr-Coulomb friction interactions using a second order Godunov solver with an adaptive grid. This system of equations is as follow:

$$\frac{\partial}{\partial t}\begin{pmatrix} h \\ hv^x \\ hv^y \end{pmatrix} + \frac{\partial}{\partial x}\begin{pmatrix} hv^x \\ hv^{x^2} + \frac{1}{2}\eta g^z h^2 \\ hv^x v^y \end{pmatrix} + \frac{\partial}{\partial y}\begin{pmatrix} hv^y \\ hv^x v^y \\ h^{v^{y^2}} + \frac{1}{2}\eta g^z h^2 \end{pmatrix} = \tag{3.1}$$

$$\begin{pmatrix} 0 \\ hg^x - \left(\frac{v^x}{\sqrt{v^{x^2}+v^{y^2}}}\right) h \tan\delta_{\text{bed}}[g^z + (v^x)^2 \frac{db}{dx}] - sgn(\frac{\partial}{\partial y}v^x)\frac{\partial}{\partial y}(\frac{\eta}{2}\sin(\delta_{int})h^2 g^z) \\ hg^y - \left(\frac{v^y}{\sqrt{v^{x^2}+v^{y^2}}}\right) h \tan\delta_{\text{bed}}[g^z + (v^y)^2 \frac{db}{dy}] - sgn(\frac{\partial}{\partial x}v^y)\frac{\partial}{\partial x}(\frac{\eta}{2}\sin(\delta_{int})h^2 g^z) \\ \end{pmatrix}$$

where $h(x, y, t)$ is the depth of the flow at location (x, y) and time t. v^x, v^y, hv^x, hv^y are flow velocities and depth-averaged momenta. g^z is the gravitational acceleration, and δ_{int} and δ_{bed} are internal and basal friction angle, respectively [110], and η is a rheological parameter.

TITAN2D takes in a wide array of user input, including flow-related variables (internal and basal friction angle, volume, granular pile geometry, initial flow velocity, etc) and location or scenario-based variables such as a digital elevation map, vent location and initial flow direction. The main mechanism to slow down a flow is encoded in the internal and basal friction angle variables. However, it has been observed that TITAN2D simulations are most sensitive to the basal friction angle compared to the internal friction angle. It is for this reason that we developed the Hierarchical Bayesian model of Chapter II, so as to reduce the uncertainty in our knowledge of the basal friction angle, and thus improve the accuracy of TITAN2D simulations.

It outputs a time series of flow depth and velocity at each location in the elevation map. An example of this is illustrated in Figure 3.1, where we plot the maximum flow depth at each location over time for case of the Long Valley region.

TITAN2D uses the message parsing interface MPI and can be run on clusters of computer to speed up processing and handle large amount of data. However, one TITAN2D simulation for an average flow for 1 hour of simulated time can take around 30 minutes of compute time on 4 processors. This computational cost is the main drive behind our use of statistical emulators to speed up computation.

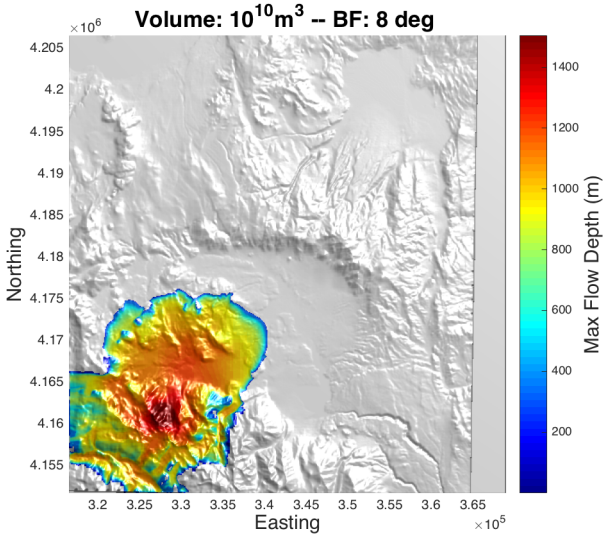


Figure 3.1: Estimated maximum flow depth of a pyroclastic flow in Long Valley for an event of volume 10^{10}m^3 and basal friction of 8 degrees

3.2.2 Gaussian Process Response Surface (GaSP)

Gaussian Process Response Surface models GaSP (also knowns as emulators, meta-models, surrogate models) are statistical tools commonly used to model the output of physical models that are expensive to exercise, such as TITAN2D [92, 61, 127, 33]. With a few input/output pairs from a complex simulator, a GaSP model can replicate the behavior of the complex model at untried inputs, and do so much faster than the simulator. For our purposes, the simulator is TITAN2D [2], a physical model of block-ash flow that is widely used to model pyroclastic flow paths. Inputs are vent locations (Northing, Easting), flow volume, and basal friction angle,

[2]For more on TITAN2D see Section 3.2.1

and the output is the maximum flow depth at a location.

A GaSP model works as follow. Let $\mathbf{X} = [x_1, x_2, \ldots, x_n]^T$ be the set of inputs for a simulator, and $\mathbf{Y} = [y_1, y_2, \ldots, y_n]^T$ be the corresponding response. $\mathbf{X}$ is an $n \times d$ matrix called the *design* and $\mathbf{Y}$ is a n-dimensional vector. A GaSP emulator models the response $\mathbf{Y}$ as if it is a realization from a n-dimensional normal distribution $\mathcal{Y} \sim N(\mu, \sigma_z^2 R)$. In other words,

$$y(x_i) = \mu(x_i) + z(x_i) \qquad \forall x_i \in X_D \tag{3.2}$$

where $\mu = f(\cdot\,; \theta)$ is the mean value of Y, which can be some pre-defined function (constant, linear/quadratic regression) of the input X_D. $z(x)$ is a Gaussian process with zero mean $E[z(x)] = 0$ and covariance structure given by $cov(z(x_i, x_j)) = \sigma_z^2 R_{ij}$, with R_{ij} being a $n \times n$ correlation matrix. This covariance kernel can be defined any number of ways depending on the stationarity of the simulator response Y. We use the Matérn-5/2 kernel (See equation 3.3) with separate length-scale parameter σ_m and a standard deviation σ_f [91].

$$R(x_i, x_j) = \left(1 + \sqrt{5}\beta + \frac{5}{3} r\beta^2\right) \exp\left(-\sqrt{5}\beta\right) \tag{3.3}$$

where

$$\beta = \sqrt{\sum_{m=1}^{d} \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}}$$

and m denotes the m^{th} element of an input vector x_i. Any covariance kernel in the Matérn-ν family of kernels can be re-written as a product of a Squared Exponential (SE) function – the most widely used kernel – and a polynomial of order p, where $\nu = p + 1/2$. Unlike the *infinitely-differentiable* SE kernel, the Matérn-5/2 kernel is only twice differentiable, which makes it more suitable for modeling non-linear processes. Stein [111] recommends using Matérn kernels, as many natural physical

processes do not behave as smoothly – a primary reason for using SE kernels. Under normal assumptions, the likelihood of a GaSP model can be written as

$$L(\sigma_m, \mu, \sigma_f) = \frac{1}{(2\pi\sigma_z^2)^{n/2}|\boldsymbol{R}|^{1/2}} \exp\left[-\frac{1}{2\sigma_z^2}(\boldsymbol{Y} - \mathbf{1_n}\mu)^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{Y} - \mathbf{1_n}\mu)\right] \quad (3.4)$$

and the best linear unbiased predictor at an untried input x^* has mean and variance

$$\hat{y}(x^*) = \hat{\mu} + \boldsymbol{r}^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{Y} - \mathbf{1_n}\hat{\mu}) \quad (3.5)$$

$$s^2(x^*) = \hat{\sigma}_z^2\left[1 - \boldsymbol{r}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{r} + \frac{(1 - \mathbf{1_n}\boldsymbol{R}^{-1}\boldsymbol{r})^2}{\mathbf{1}_{\boldsymbol{n}}^{\mathrm{T}}\boldsymbol{R}^{-1}\mathbf{1_n}}\right] \quad (3.6)$$

where, $\boldsymbol{r} = (r_1(x^*), \ldots, r_n(x^*))^{\mathrm{T}}$ with $r_i(x^*) = \boldsymbol{R}(x^*, x_i)$ from equation. 3.3. In practice, parameters associated with μ, e.g intercept and slope if μ is linear, are substituted with their maximum likelihood estimates, likewise with σ_f and all σ_m [92].

A few key aspects of GaSP models ought to be noted. First, the mean prediction at a design point $x_i \in X_D$ is equal to its corresponding simulator response y_i, i.e $\hat{y}(x^*) = y_i$. Accordingly, $s^2(x_i) = 0$ for all input points in our experimental design X_D. The predictive variance $s^2(x^*)$ tends to be larger in regions of the input space that are less represented in the experimental design X_D. For this reason, X_D is usually set up in a space-filling manner, such as a Latin Hypercube design, so as to reduce regions of large variance in the response surface.

3.3 Methodology

3.3.1 Dataset

First, a log of the historical record of eruptions in the area gives us a basis for a location model to be used in computing the probability of inundation (Figure 3.2 and Table A.1 in Appendix A). Secondly, we used the results from the hierarchical model from Chapter II for Mount Unzen, in place for the basal friction and volume required to make TITAN2D runs (Figures 3.3 and 3.5).

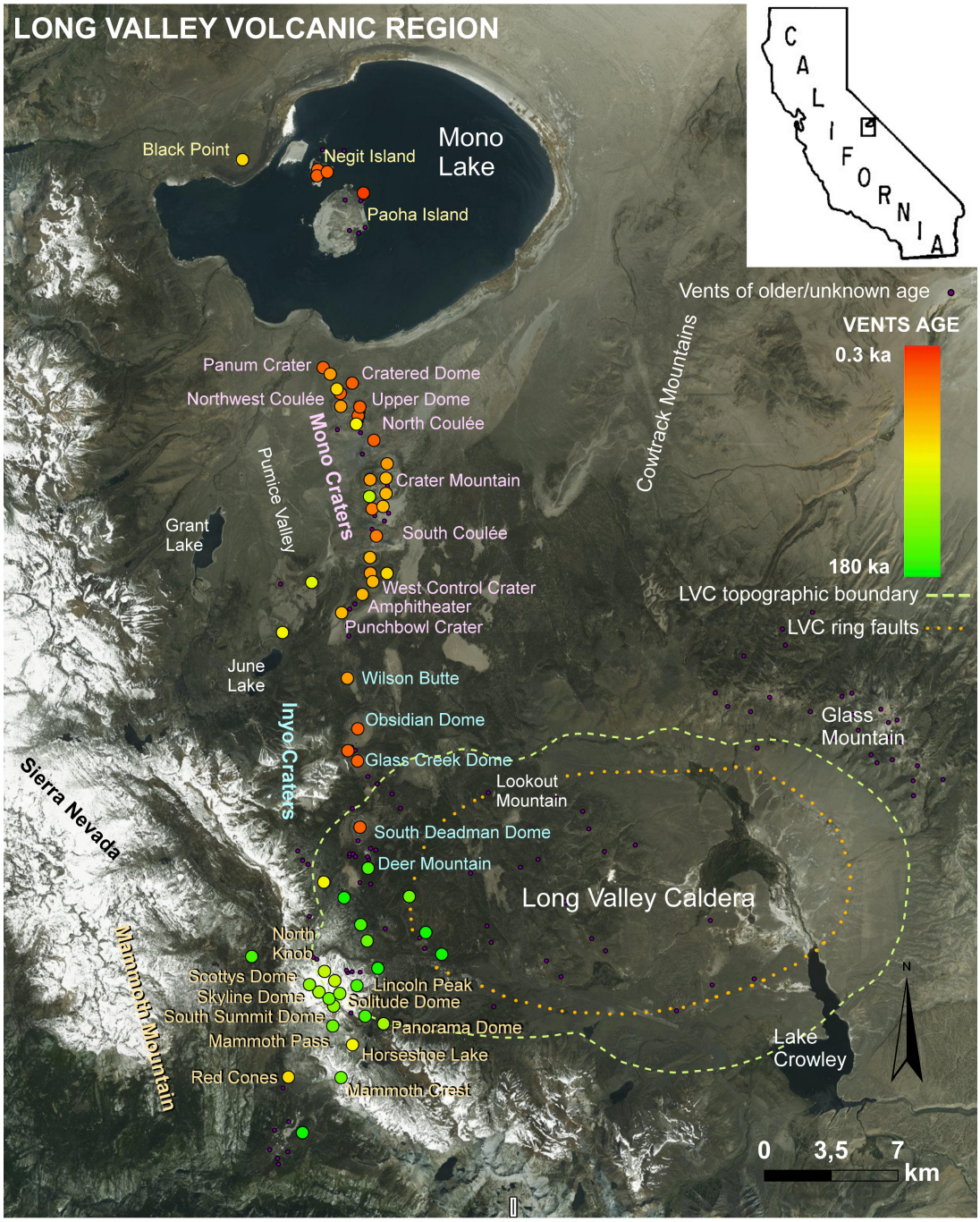


Figure 3.2: Historical records of the eruptions in the Mono-Inyo craters and Long Valley Caldera region in the last 180 ka. Colors represent the age of eruptions. The different colored labels represents different spatial groups of vents. The border of the Long Valley Caldera is also displayed as from [4] (Picture taken from Andrea et al. [13])

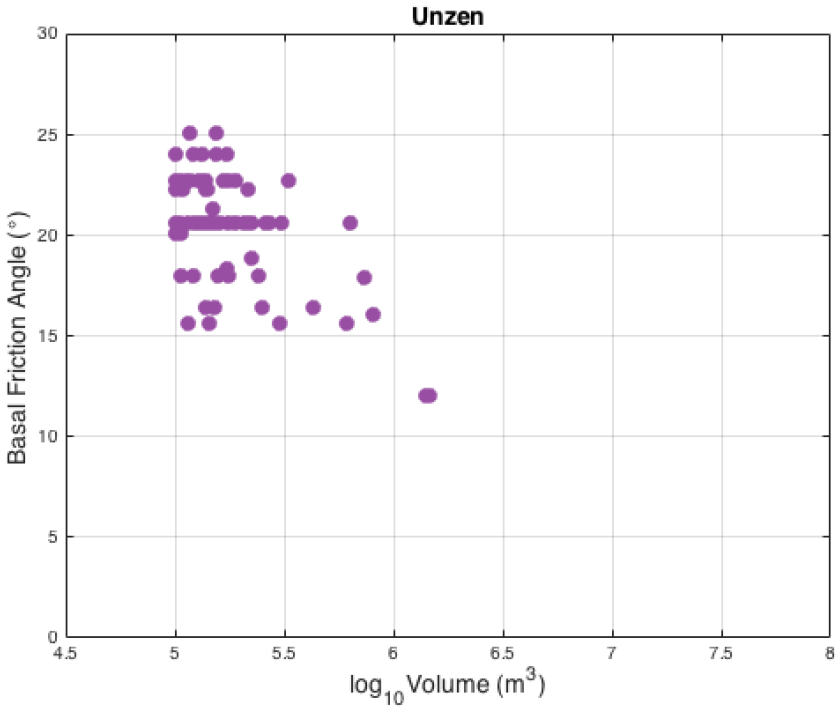


Figure 3.3: Basal Friction and Volume data from Mount Unzen (Japan)

3.3.2 Experimental Design

As explained in section 3.2.1, the initial set of TITAN2D runs requires a 4-dimensional experimental design with the following features: Northing, Easting, basal friction, and volume. This is achieved using a Latin Hypercube sampling scheme to generate 6000 design points that span the design space described below.

The Easting-Northing design is restricted to a smaller section of the digital elevation map (DEM) that correspond to more recents events (see Figure 3.4). This allowed us to reduce the total number of required TITAN2D simulations. Furthermore, it allows us to explore more historically relevant regions of the design space.

From personal communication with experts [20], it was established that Mount Unzen is good stand-in for the Long Valley region. Hence, the basal friction-volume portion of the experimental design is derived from the BF-Volume relationship at Mount Unzen we obtain from the Hierarchical model presented in the previous

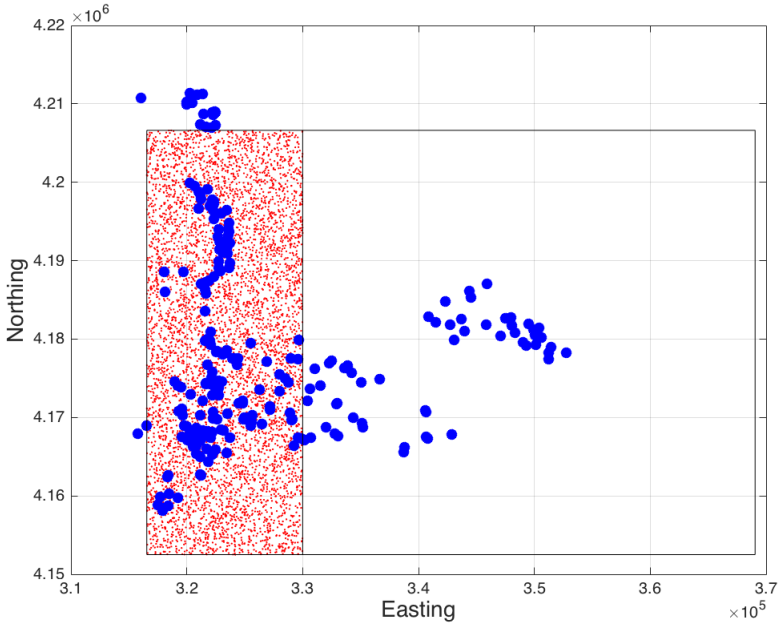


Figure 3.4: Juxtaposition of the entire DEM (larger rectangle) and the smaller area of interest our experimental design is focused on (smaller rectangle). The blue dots represent the vent location data, and the red dots represent the (Northing, Easting) vent design points

chapter. This relationship is then extended up to 10^{10}km^3, to include the largest events to be expected in the Long Valley region. Figure 3.5 illustrates this.

3.3.3 Algorithm

Using the experimental design described in the previous section, 6000 separate TITAN2D simulations were run in parallel on a distributed system, using 4 cores per node and up to 100 compute nodes. Each one of these simulations is initialized with one 4-dimensional design point, for example E = 320,000, N = 4,180,000, $\log_{10} V = 6$, and BF = $15°$. Subsequently, these simulations yielded 6000 records of maximum pile heights at all locations on the D.E.M of the Long Valley area (see Figure 3.2). It is to be noted that, most scenarios do not reach all locations, i.e many locations will have a pile height of 0 meters.

In order to create a map of probabilities of inundation, we compute these

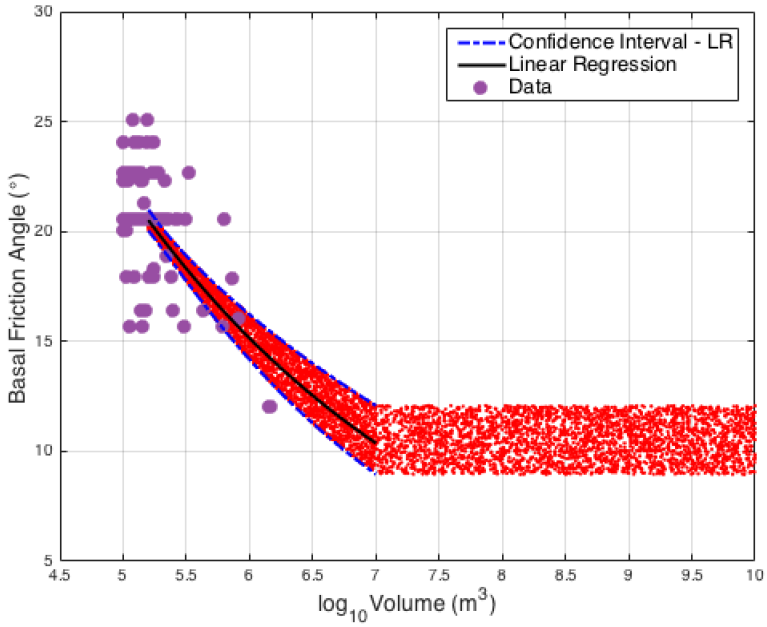


Figure 3.5: Comparison of the Basal friction – Volume subset of the experimental design we used (red dots), and the BF-Volume relationship from the Bayesian hierarchical model for the case of Mt Unzen (purple dots and 90^{th} percentile lines). The results from the hierarchical model were extended horizontally from $\log_{10} V = 7$ to $\log_{10} V = 10$, to accommodate the much larger volumes expected to be observed while staying physically accurate

probabilities for thousands of individual locations and make a map with this information. The following steps outlines our algorithm to compute a probability of inundation profile curve for *each* of these individual locations.

For each location L_i

Step 1: *Get a subset of design points relevant for* L_i

Since we are interested in estimating a critical flow height contour (h_{crit} = 1 meter), we chose, out of the initial 6000 points, those scenarios that produced a flow height $0.2 \text{ m} \leq H \leq 50$ m. Design points that yielded flows higher than 50 meters were deemed to not have enough information concerning our pre-defined critical height. Likewise, flow heights of less than 0.2 meters were set to 0, as this is just above the modeling error of TITAN2D for shallow flows. On top of this, we complemented this sub-design with a n *zeros* design points (events of $H = 0$ meter) that are closest to L_i

in euclidian distance. So, for location L_i situated at (E_l, N_l), we added to the sub-design $n = 20$ new design points with the smallest

$$d(x_i, L_i) = \left[(E_i^x - E^l)^2 + (N_i^x - N^l)^2\right]^{1/2}$$

where x_i is a *zero* design point with location (E_i^x, N_i^x) and (E^l, N^l) corresponds to location L_i

Step 2: *Fit a GaSP to sub-design*

A version of the BLUP described in section 3.2.2 is fit to the sub-design after it had been normalized to the hypercube $[0, 1] \times [0, 1] \times [0, 1]$, to reduce the effect of the difference in magnitudes of the different variables. Similarly, we scaled the output flow height by $\log(h + 1)$, as this emphasizes small values of flow heights, and de-emphasizes the magnitude of very high height values and at the same time does not affect the sign of the height values (must always be positive). The following Matérn-5/2 covariance function is used

$$k(x_i, x_j) = \sigma_f^2 \left(1 + \frac{\sqrt{5r}}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right) \exp\left(-\frac{5r}{\sigma_l}\right)$$

where $r = \sqrt{(x_i - x_j)^T (x_i - x_j)}$

Step 3: *Evaluate GaSP and impose monoticity constraint*

Once the GaSP is fitted, we evaluate it on a dense grid of 75 points along each our 3 input dimensions (Easting, Northing and Volume)[3], producing mean predictions $\hat{y}(x)$ and their standard deviation $\sigma(x)$. With this evaluation, the contour h_{crit} can be easily visualized, as it is the level surface corresponding to $\hat{y}(x) = 1$ m (see Figure 3.6). This level surface separates regions of the design space that lead to inundation ($h > 1$ m) and those that do not.

A monotonic constraint is also imposed along the volume axis to force the pile

[3]From here on, the basal friction is omitted from our design as it is dependent on volume

height at a location L_i to be monotonically increasing as the volume of an event is increasing. While this assumption is realistic, the mean predictions $\hat{y}(x)$ from the GaSP do not always follow it, hence the need for a correction. The pooled adjacent violator algorithm, PAV, presented in [6], [69] is used to impose this constraint (see Figure 3.7).

Briefly, this algorithm for a 1-D problem is as follow:

Given a sequence of inputs $x_0, x_1, \ldots, x_i, x_{i+1}, \ldots, x_n$, and their corresponding outputs $y(x_0), y(x_1), \ldots, y(x_i), y(x_{i+1}), \ldots, y(x_n)$

Step a: Iterate in increasing order of x_i, until the first pair $y(x_i), y(x_{i+1})$ that violates the monoticity is found, i.e $y(x_i) > y(x_{i+1})$. Replace both values with their average $(y(x_i) + y(x_{i+1})/2$

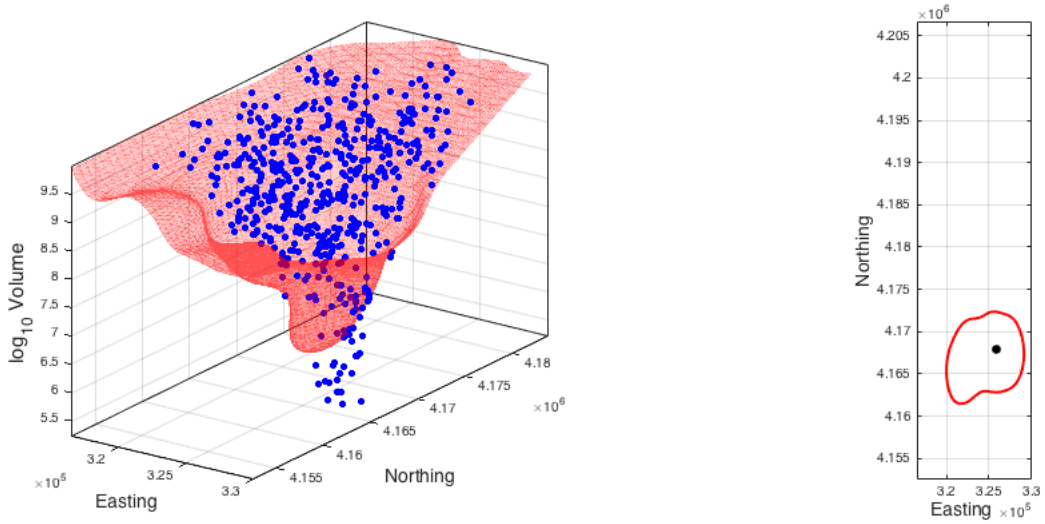
Step b: Check $y(x_{i-1})$ and $y(x_i)$. If $y(x_{i-1}) > y(x_i)$, replace $[y(x_{i-1}), y(x_i), y(x_{i+1})]$ with their average. Continue this correction going to the left until the monoticity is respected again. Then, proceed to the right again starting from $y(x_{i+1})$.

Step 4: *Compute Probability of Inundation*

In order to provide an estimate of the probability of inundation of a given location that takes into account all the random inputs, one would have to have input models; for the vent locations (Northing & Easting) and the flow volume. Given these, a probability of inundation $P(I)$ of a location L_i can be computed using Baye's theorem

$$P_{L_i}(\mathbf{I}) = \int\int\int P(\mathbf{I}|\text{E,N,V}) \times P(\text{E,N}) \times P(\text{V})\, dE\, dN\, dV \tag{3.7}$$

However, in the absence of a probabilistic model of volume $P(V)$, we use models of vent locations and computed probabilities of inundation *conditioned* on a flow volume. The two models of vent location we use are, a simple uniform



(a) (b)

Figure 3.6: **(a)** Example of a h_{crit} level surface (in red) computed from the evaluation of a GaSP fitted over a dense grid. The blue dots are the sub-design used as support for the GaSP. **(b)** Top view of a slice of the level surface shown on the left at Volume $= 10^8 \text{m}^3$. Scenario events whose location fall within this contour lead to a flow height of 1 meter or more at the Town of Mammoth Lake (represented by the black dot)

distribution and a Gaussian mixture model based on the relevant historical vent locations (see Figure 3.8). The Gaussian mixture model (GMM) is fitted using 3 distinct means and full covariance matrices.

These two distributions are then sampled and used in a Monte Carlo scheme to compute conditional probabilities of inundation using the following equation which follows naturally from equation (3.7) (see Figure 3.9)

$$P_{L_i}(\mathbf{I}\,|V \geq 10^v m^3) = \int_{E\times N} \mathbf{1}_{h>1\text{m}}\, p(E,N)\; dE\, dN$$

where $v \in [6, 10]$, $E \in [316550, 330000]$, $N \in [4152500, 4206620]$, and $\mathbf{1}_{h>1m}$ is an indicator function that is 1 if $h > 1$m, and 0 otherwise. In other words, for one geographical location, we get a function $P(\mathbf{I}|V)$ of the flow volume, which we can illustrate by the probability profile curve in Figure 3.10

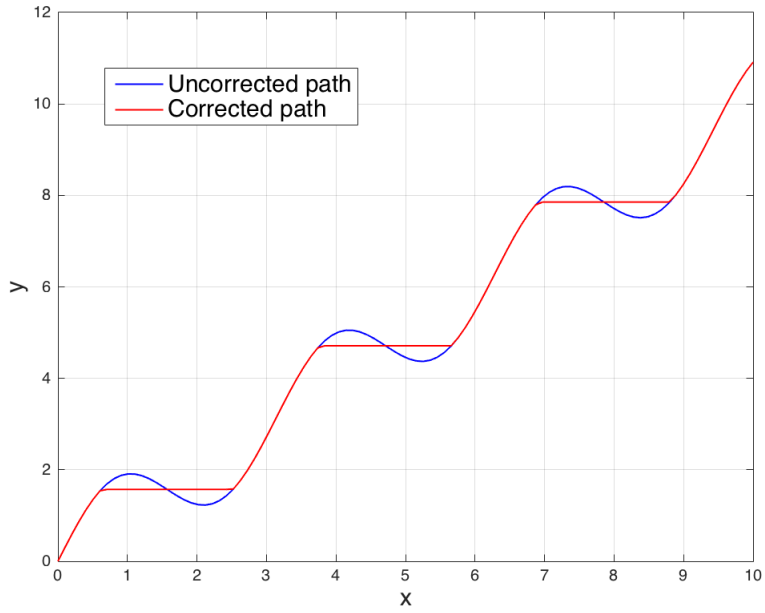


Figure 3.7: Illustration of the pooled adjacent violators algorithm applied on the function $y = \sin(2x) + x$ in the range $[0, 10]$. The corrected path (in red) agrees with the original function (in blue) in places where the monotonicity is respected, and runs a moving average in places where it does not

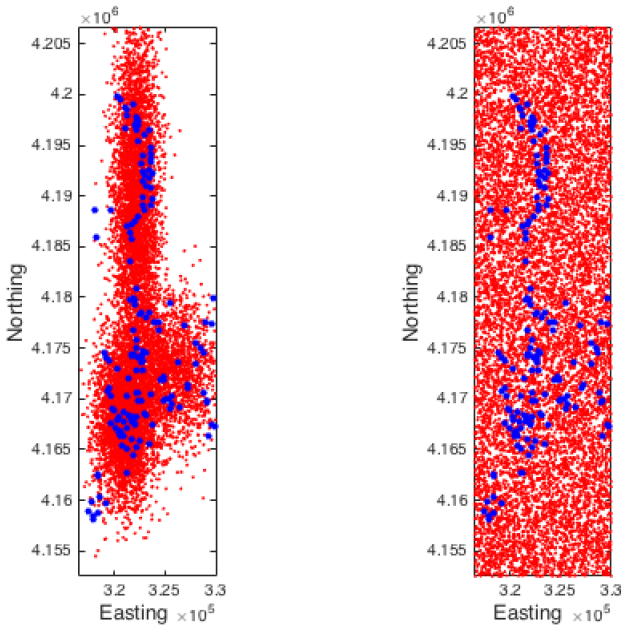


Figure 3.8: On the left, samples (in red) from a Gaussian mixture model fit over historical vents locations (in blue). On the right, samples from a uniform distribution of vent locations over our region of interest of the D.E.M

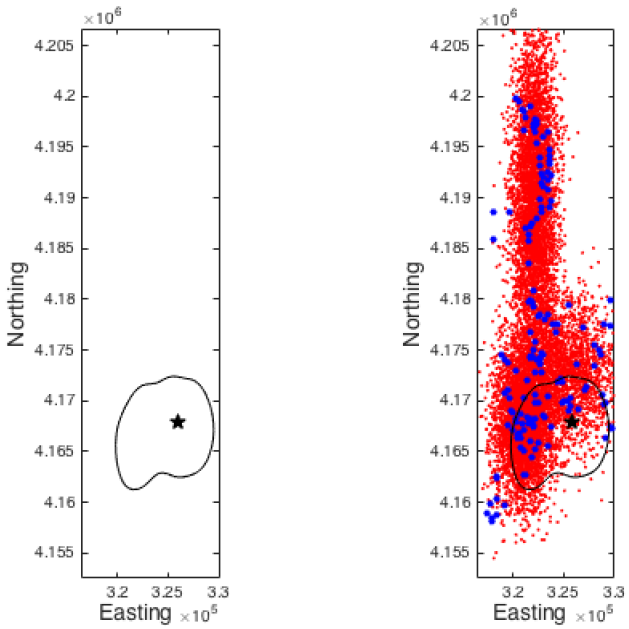


Figure 3.9: Illustration of the Monte Carlo method used to compute conditional probabilities of inundation for the town of Mammoth Lakes (black star) using a Gaussian mixture model (GMM) of vent locations. First, the 1-meter level surface is sliced at $V = 10^8\, m^3$, which yields the 2-D contour line seen on the left. Then, $P_{L_i}(\mathbf{I}\,|V \geq 10^8)$ is approximated by the ratio of samples from the GMM (in red) that fall within this contour over the total number of samples (right)

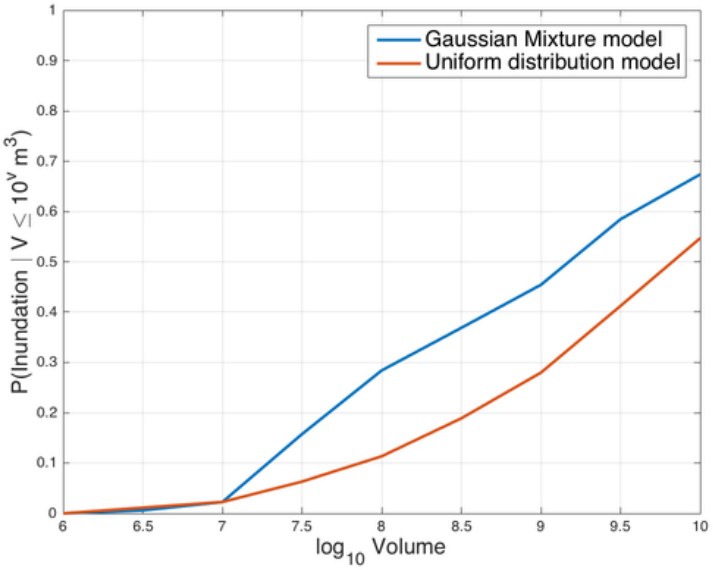
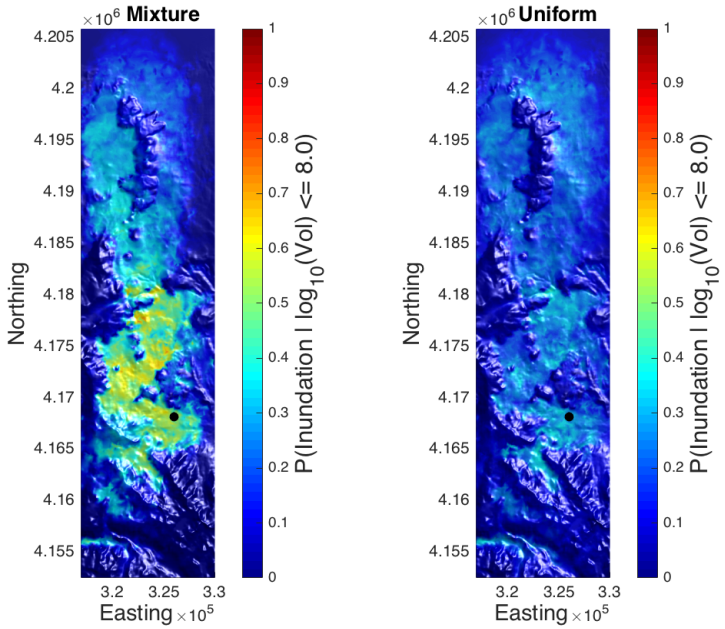


Figure 3.10: Probability profile curves for the town of Mammoth Lakes (CA), under two different vent location models

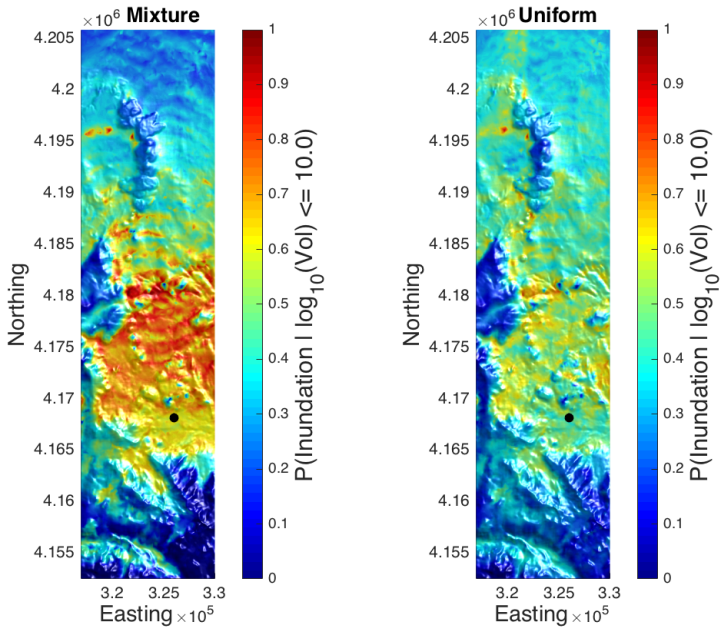
3.4 Results and Discussion

The algorithm detailed in the previous section is run for 10,000 different map locations, yielding $10,000$ probability profile curves similar to Figure 3.10. Then, we interpolated these values to a much finer grid on the map for higher resolution map of inundations, which we then combine into one pictogram, i.e the hazard map (Figure 3.10)

While both vent location models lead to results that show a strong correlation between probability of inundation and altitude, they differ on a couple of points. Since the GMM describes the vent locations more closely, its results tend yield lower probability estimates far away from the Mono-Inyo craters compared to the the uniform distribution. Moreover, we note that at volumes 10^8m^3 or larger, a large portion of the DEM has a relatively high probability of being inundated by more than 1 meter of pyroclastic flow (40% or higher).



(a)



(b)

Figure 3.11: Maps of probability of inundation of the LGV region conditioned on a flow volume being 10^8m^3 or smaller (top row), and 10^{10}m^3 (bottom row), using a uniform distribution (right column) and a Gaussian mixture model (left column) fitted over recorded vent locations. The black dot shows the location of the town of Mammoth Lakes. Computing the color of each pixels is equivalent to repeating the algorithm described in Section 3.3.3, Figures 3.9 and 3.10

Chapter 4

Multi-Objective Adaptive Design for Estimation of Multiple Contours

4.1 Introduction

In the previous chapter, we showed how a very large number of computationally expensive runs of TITAN2D are required to yield a probabilistic hazard map of an entire region. This large number of simulations, based on a input DoE[1] henceworth refered to as *global design*, is required to sample the aleatoric uncertainty of volcanic systems, i.e, the intrinsic randomness of their process. Usually, this global design would be chosen using some variant of the Latin Hypercube sampling technique [75], or a space-filling design [60].

However, these generic schemes tend to not be very useful for contour or threshold estimation, which is our goal, as contours constitute a very small section of a design space and a primary objective of these schemes is to explore as much of the design space as possible. We aim to develop a strategy that reduces the number of simulations needed by targeting a contour of interest. *On top of this, the issue of having a useful DoE is compounded by the fact that for each physical location, only a subset of the global design proves relevant for that particular location and this relevant subset varies as we change location*. Specifically, this is done because for most locations the majority of simulations lead to either an estimated pyroclastic flow height of 0 meters (no flow) or a flow so large (500 meters or more), as to not give any information in the search for a critical height contour h_{crit}. So, while this location-specific *sub-design* aims at providing a suitable support for a Gaussian

[1] Design of Experiment

emulator from which to compute a contour, it also means that, at each location, our predictions are based on a limited number of design points. We need a multi-objective oriented strategy for choosing a global design which provides a useful sub-design at multiple locations.

One way to solve this problem is to start with a relatively small initial design DoE and adaptively add design points (or batches of design points) to improve the accuracy of an emulator in the vicinity of the contour(s) level of interest. In the same spirit as Jones' Efficient Global Optimization algorithm (EGO) [61], Ranjan et al. [90] and Picheny et al. [88] proposed algorithms that balance space exploration with reducing emulator's variance around a level of interest. Shan et al. [100] proposed a rough-set based approach to estimating areas less or equal to a given contour. Viana et al. [122] used an ensemble of surrogate models, where each model suggested one new design point to add, and iteratively added batches of new points at each step of their algorithm.

This problem is also found in the engineering reliability literature where instead of estimating a contour level, the focus is on trying to find a limit state function that separates safe and failure regions of design space. Kuczera et al. [68] suggested a mixture of meta-models and clustering algorithm to estimate transition zones and from that, non-linear state function. Basudhar et al. [7] and Arenbeck et al. [3] both used support vector machine as meta-models, and then used clustering and importance sampling, respectively, to separate failure regions from safe ones.

In this chapter, we are concerned with improving our estimates of inundation contours for locations in the Mono-Inyo region of California. We develop a new methodology which builds on the work of Picheny et al. [88]. Our methodology sequentially selects batches of new design points at which simulations are to be run in order to improve estimate of inundation contours for *more than one location at a time*. Unlike the previous works summarized above, the novelty of this method lies in

devising a strategy for resource allocations, in this case TITAN2D runs, when having to satisfy multiple objectives, in this case contour estimates. The algorithm we develop allows us to cleverly combine experimental results we already have and vastly reduce the number of simulations needed to get good estimates of inundation contours for multiple locations. Ultimately, this can be used in concert with the hazard mapping method from Chapter III to speed up studies of new hazard threats.

The organization of this chapter is as follow. Section 4.2 presents our algorithm, both from a theoretical and computational perspective. Then, we present applications of our algorithm on a pedagogical example, Section 4.3, and on the hazard mapping tool of the previous chapter, Section 4.4.

4.2 Methods

Traditionally, DoE algorithms such as Latin Hypercube sampling [75], or space-filling design [60] choose the entire design of experiment before any computer simulation is made (experiment). However, algorithms that iteratively choose design points are more flexible. Primarily, this is because they allow for objective-oriented constraints, which discourages unnecessary experiments. Examples of such objectives include finding a global extrema [61], a contour level [90, 88, 100], or a limit-state function [68, 7, 3].

Objective-oriented constraints are defined in terms of a utility function that assesses how much experimental improvement would be achieved if a candidate design point were to be added to the design of a particular experiment. Candidates with high utility value are then iteratively added to the DoE by a sequential algorithm. In our case, the utility of a design point is generally defined as some weighted average of its closeness to a contour and the variance of an emulator at that point.

A general step-by-step procedure to improve contour estimation for more than one location at a time look works as follows. First, we select an initial small DoE for all locations being considered and feed them to TITAN2D as inputs. Then, we

compute utility functions for each locations being considered using predictions from emulators fit to the TITAN2D output for each of those locations. Next, we devise a multi-objective optimization strategy to choose optimal design points that maximize all utilities simultaneously. These points are then added to each location's sub-design and simulations are run at those new design points. This process is repeated until the Mean Squared Error (MSE) of each GaSP emulators near contours has decreased to a satisfactory level.

4.2.1 Modified Integrated Mean Square Error criterion (I.M.S.E)

Picheny *et al.* [88] proposed this criterion as a specialized version of the Integrated Mean Square Error criterion of Sacks et al. [92]. Both criteria use a GaSP with predictive mean and variance $\hat{y}(x)$ and $s_k^2(x)$ for the basis of their computations. The original criterion is defined in equation 4.1 below

$$\text{IMSE} = \int_X \text{MSE}[\hat{y}(x)]\phi(x)dx \tag{4.1}$$

$$\text{IMSE}_T = \int_X s^2(x)W(x)dx \tag{4.2}$$

where $\phi(x)$ and $W(x)$ are both weight factors,
$W(x) = \frac{1}{\sqrt{2\pi(\sigma_\epsilon^2 + s^2(x))}} \exp\left[-1/2\frac{(\hat{y}(x)-T)^2}{\sigma_\epsilon^2+s_K^2(x)}\right]$, $\hat{y}(x)$ and $s^2(x)$ the mean and variance of an GaSP, and σ_ϵ a constant defining a range around the level of interest T. It measures the average MSE of a meta-model over the whole input space.

According to [88], the modified IMSE criterion (equation 4.2) provides a measure of the uncertainty of the GaSP ($s_k^2(x)$) weighted by the closeness of $\hat{y}(x)$ to the contour of interest T. In other words, both points with a predictive response $\hat{y}(x)$ close to the contour and points with a high variance $s_k^2(x)$ in the vicinity of the contour are weighted higher in this scheme. The algorithm we designed samples the input space proportionally to the IMSE, and updates the DoE with the new found design points. In doing so, it refines the sampling of the input space near the contour, and

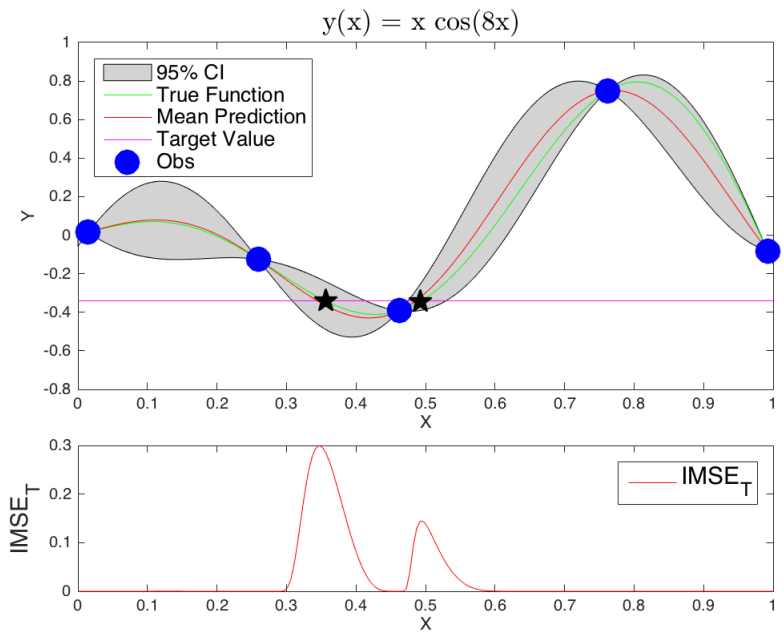


Figure 4.1: Illustration of the IMSE criterion of Picheny et al. [88]. The 'contour' (target value) of the 1-D function is set at $a = -0.34$. The bottom figure shows how more importance is given to regions near the target value represented by the black stars

improves emulators' approximations of the inundation contour.

As example of this criterion, Figure 4.1 below illustrates the IMSE for a simple 1-D function. Regions of the input space that are close to the contour of interest have higher values of IMSE criterion. Hence, if we were to run our algorithm for this example, the next points to be added to the DoE are most likely to come from these regions.

4.2.2 Aggregation and Optimization

After IMSE criteria (utility functions) have been set up for multiple contours, comes the problem of devising a strategy to find "compromise" design points that improves contour estimation for all contours under consideration – each corresponding to a different map location of interest.

We solved this resource allocation problem using the Nash arbitration scheme [81], which suggests that the solution to such a problem be the maximum of the

product of all criteria. Essentially, this approach turns a constrained high-dimensional optimization problem into a manageable 1-D optimization problem. It is a special case of the *product of powers* composite utility function of Bridgman [16], equation4.3, where the weights w_i of each weights are all set to 1.

$$U_C = \prod_{i=1}^{k} [U_i(x)]^{w_i} \tag{4.3}$$

where $U_i(x)$ is the utility function of the ith objective (in our case the IMSE for one particular contour), and w_i is the corresponding weight associated with its importance. However, criteria can have different magnitudes. So, a normalization process is first required to avoid some criteria from dominating others. For our purposes, we rescaled our vector-shaped criteria to have a norm L_2 of 1.

Once we setup the composite utility function equation 4.3, we seek new design points that maximize it with the added constraint that we maximize the minimum distance between those new design points and the DoE we already have at hand. This constraint avoids clustering of the design points and provides an appropriate support for GaSP emulators.

For our work, we optimize equation 4.3 using a search method over a fine grid in the input space $[0, 1] \times [0, 1] \times [0, 1]$. First, we compute each $U_i(x)$ over the grid, and evaluate the composite criteria U_C. Then, we select design points whose U_C value are above the 99^{th} percentile. From these, we iteratively pick a new design point (up to a certain number n of design points) in a manner that is space-filling vis-a-vis of all the current DoE. We use the euclidian distance to compute the distance between design points. Figure 4.2 illustrates this algorithm.

It is to be noted that, while it may seem like we are computing many more simulation input/output and choosing from them new design points, we are not doing so. The composite criteria function of equation 4.3 is based on equation 4.2, which is

itself based on evaluations of a cheap GaSP. So, a grid search optimization algorithm is relatively fast. In cases where such technique is not feasible, we recommend a multi-start optimization algorithm, as U_C can be highly non-linear.

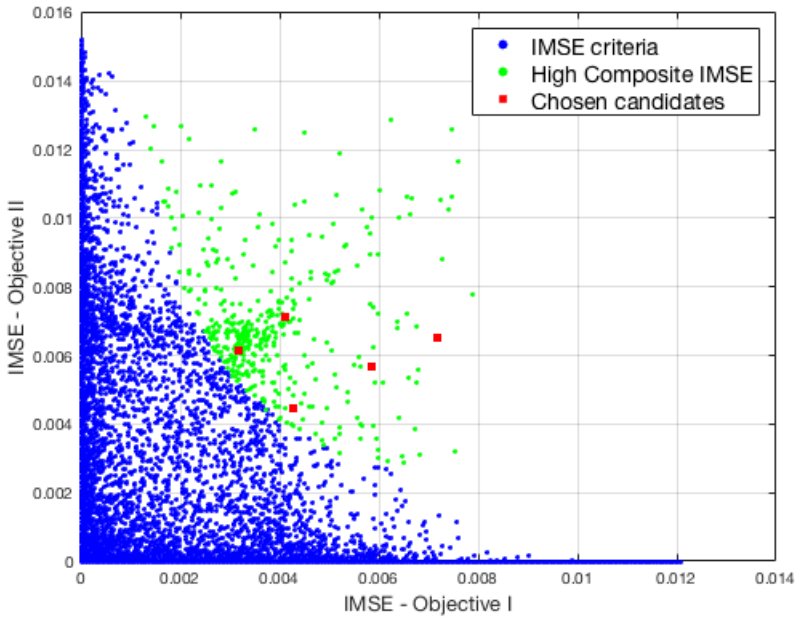


Figure 4.2: Illustration of one optimization step of the composite IMSE function of equation 4.3 in the criteria space. Objective I and II represent two different contours. The green dots are design points in the 99^{th} percentile of the Composite IMSE U_C, and the red dots represents points that we decided to add to our DoE

4.3 Numerical Example

Consider the following 'family' of functions

$$F_2(x, y; a) = \exp\left[\frac{-(x-0.5)^2}{0.02}\right] \times \exp\left[\frac{-(y-0.4)^2}{0.02}\right] + \exp\left[\frac{-(x-x_a)^2}{0.045}\right] \times \exp\left[\frac{-(y-y_a)^2}{0.045}\right] \tag{4.4}$$

where $x_a = 0.4\cos(\frac{\pi}{4}a) - 0.7\sin(\frac{\pi}{4}a)$ and $y_a = 0.4\sin(\frac{\pi}{4}a) - 0.7\cos(\frac{\pi}{4}a)$ and a is a free parameter in the range $[0, 1]$. Each of these functions is defined on a range that include the value $T = 0.6$ which we will consider to be our contour level of interest

(see Figure 4.3). We apply our multi-objective sequential algorithm (summarized in Table B.4) to the problem of improving GaSP model predictions of the contours $T = 0.6$ of any two of these functions. Several different sets of experiments are made to illustrate the advantage of our algorithm. One instance of such experiment is set up as follow.

With a contour level of $T = 0.6$, the parameter σ_ϵ, which determines the width of a neighborhood around the contour, is set at $\sigma_\epsilon = 0.1$ (roughly 5% of the output range of the functions). This has the effect of focusing our search window around the contour and reduces input space exploration. An initial DoE of size $n = 10$ for each objective is chosen, and batches of 5 design points are added after each iteration, up to 6 iterations. Each experiment is run 15 times with different random seeds to minimize the effect of the initial DoE on our results. Table 4.1 summarizes this experimental setup.

Experimental setup	
Initial DoE size	10
Batch size	5
Number of steps	6
Grid size	100×100
Values of free parameters a	(0.1, 0.6)

Table 4.1: Summary of the experimental setup of one run of our algorithm. In this case, we are trying to estimate the contours corresponding to $a = 0.1$ and $a = 0.6$

Furthermore, we consider three metrics of 'goodness' to assess the performance of the algorithm. The first metric measures the average euclidean distance between the j^{th} contour C_j and the true contour C_T [90].

$$M_1 = \frac{1}{|C_j|} \sum_{x \in C_j} d(x, C_T) \tag{4.5}$$

where $d(x, C_T) = \min\{\|x - y\|_2 : y \in C_T\}$. The second metric measures the

maximum distance between contour C_j and the true contour C_T [90] and is given by

$$M_2 = \max\{d(x, C_T) : x \in C_j\} \tag{4.6}$$

Lastly, we measure the average GaSP variance $s_k^2(x)$ in the neighborhood of the contour. That is, we have

$$M_3 = \frac{\sum_x \boldsymbol{I}(x) s^2(x)}{\sum_x \boldsymbol{I}(x)} \tag{4.7}$$

where $\boldsymbol{I}(x) = \begin{cases} 1, & \text{if } |\hat{y}(x) - T| \leq \sigma_\epsilon \\ 0, & \text{otherwise} \end{cases}$

1.	Set up an initial DoE X_i for each objective i and perform experiments to get corresponding responses Y_i
2.	Create a dense grid of candidates design points X_D
3.	Fit a GaSP emulator to the data $\{X_i, Y_i\}$ of each objective
4.	Compute the IMSE criterion $U_i(x_k)$ of each $x_k \in X_D$, for each objective i using equation 4.2
5.	Combine all criteria using the composite utility function U_C of equation 4.3
6.	Identify k design points $X_k = \{x_1, x_2, \ldots, x_k\} \in X_D$ that maximize U_C in a space-filling manner
7.	Perform experiments to obtain responses Y_k for the new k design points for each objectives
8.	Update DoE $X_i = \{X_i, X_k\}$, and $Y_i = \{Y_i, Y_k\}$ and $X_D = X_D \setminus X_k$
9.	Repeat 3-8 until experimental resources are exhausted.

Table 4.2: Outline of a multi-objective sequential DoE algorithm. A GaSP model is used as meta-model and design points are added in batches of size k

The plots in Figure 4.4, we see significant reduction in the average GaSP variance near the contour and distance between approximated contours and true contours, which shows that our algorithm is apt at finding design points that are useful for both objectives *simultaneously*. This same behavior is observed for any combinations of 2 or more functions we pick from equation 4.4.

On the other hand, if we choose to only approximate one contour (primary objective) and observe how another *near-by* contour (secondary objective) is approximated *by proxy*, we get the same behavior as in the prior case but with some dissimilarities. This is illustrated in Figure 4.5. In that case, our algorithm is able to approximate both contours but more so the principal objective than for the secondary objective. This matches one's intuition, since in this case our search algorithm focuses on the primary contour and the "goodness" of the approximation of the secondary contour varies depending on how close the two objectives are.

Next, we analyze the relationship between the improvement our algorithm achieves for two objectives (contours of function from equation 4.4) versus how close/similar those objectives are to each other. We use the radial basis function kernel (RBF) in equation 4.8 to quantify the similarity of two functions.

$$K(y, y') = \exp(-\gamma * \|y - y'\|_2) \tag{4.8}$$

where the free parameter $\gamma = 0.05$. This measure ranges from 0 for dissimilar objectives to 1 for identical objectives.

For example, if we are looking to trade-off between two objectives C_1 and C_2, we run our algorithm as usual and then compute the improvement made in, say, metric M_1 for C_2 as the ratio M_1^i/M_1^0, where M_1^i and M_1^0 are the final and initial value of M_1 respectively. This ratio is greater than 1 if our algorithm reduced the average distance metric M_1 between approximated contour C_1^i and the true contour C_1, and less than 1 otherwise. We then record that improvement versus the RBF similarity measure of the function generating C_1 and C_2.

Figure 4.6 shows that there's a directly proportional relationship between how similar objectives are and how well our algorithm performs. Indeed, the more similar objectives are, the more likely our algorithm will make great improvements toward

both of them. This observation makes sense as it is intuitively harder to find a good compromise between conflicting objectives than it is for those that are more aligned.

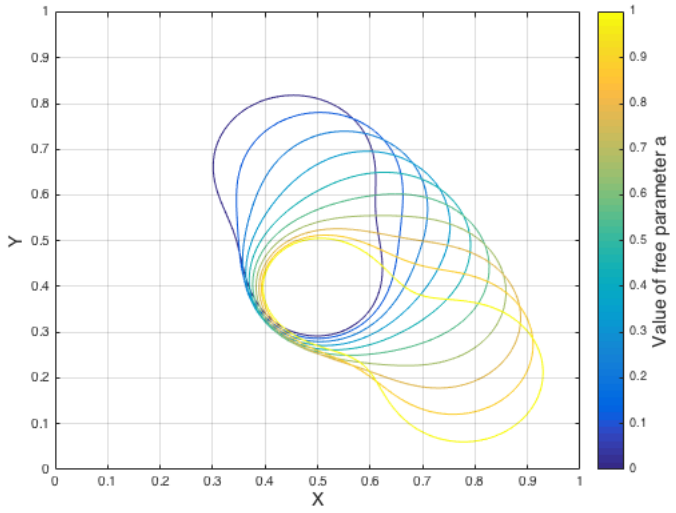


Figure 4.3: Contour level $T = 0.6$ of 10 different functions from the family of functions in equation 4.4 with $a = 0.1, 0.2 \ldots, 1$. Each of the contours is considered as an objective in our algorithm

4.4 Inundation Contours

We apply our algorithm on the problem of estimating inundation contours of locations in the Long Valley volcanic region of California (see Figure 3.6). This problem is similar to the pedagogical example in the previous section, only extended to three dimensions (Easting, Northing, and Volume) and now the $h_{\text{crit}} = 1\text{m}$ inundation contour is a level surface.

First, we create a Latin Hypercube DoE with 600 points spread out in the input space and run a TITAN2D simulation for each of those design points (event scenarios). This constitutes our *global design*. Next, for each location in consideration, we choose sets of 50 points each – *sub-designs* – that are most relevant for each location. These sub-designs are made up of a combinations of design points that lead to a flow height $0.2\,\text{m} \leq H \leq 100\text{m}$ (*non-zeros*) and design points that lead to a flow $H < 0.2\,\text{m}$

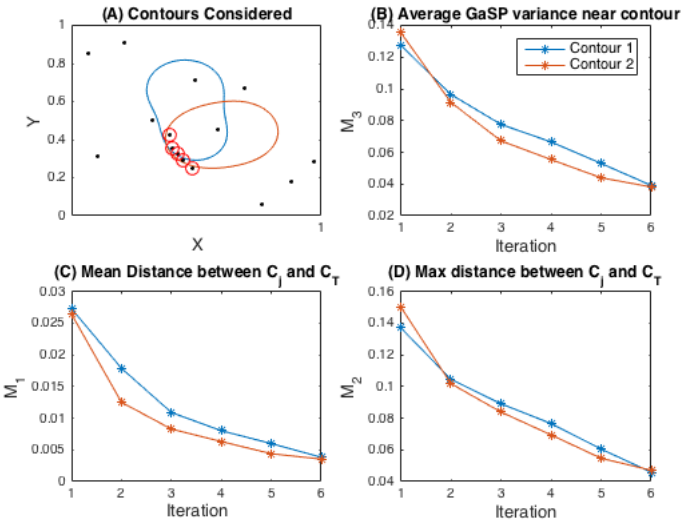


Figure 4.4: Plots illustrating the performance of our algorithm in simultaneously estimating 2 different contours (A) of two functions from equation 4.4 using a GaSP model and a IMSE criterion to select design points that are relevant for both objectives. In (A), the black dots represent the initial DoE, while the red circles mark the new design points added after the first iteration of the algorithm

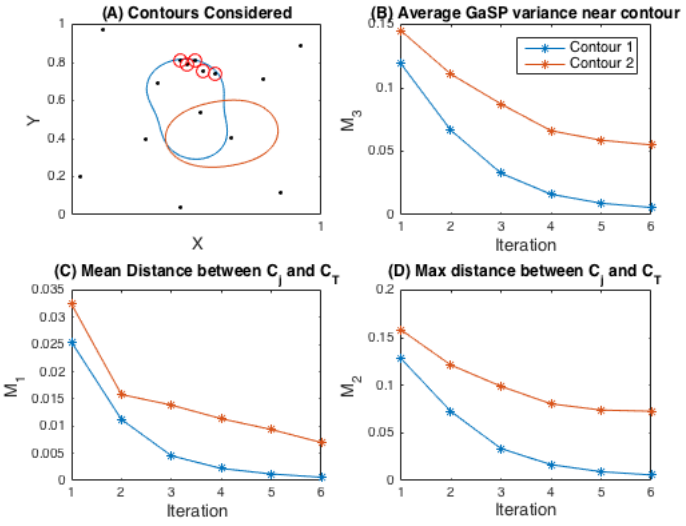


Figure 4.5: Plots illustrating how an objective (red contour) is affected by our algorithm estimating another objective *near-by* (the blue contour). The two objectives (A) are contours of two functions from equation 4.4. In (A), the black dots represent the initial DoE, while the red circles mark the new design points added after the first iteration of the algorithm

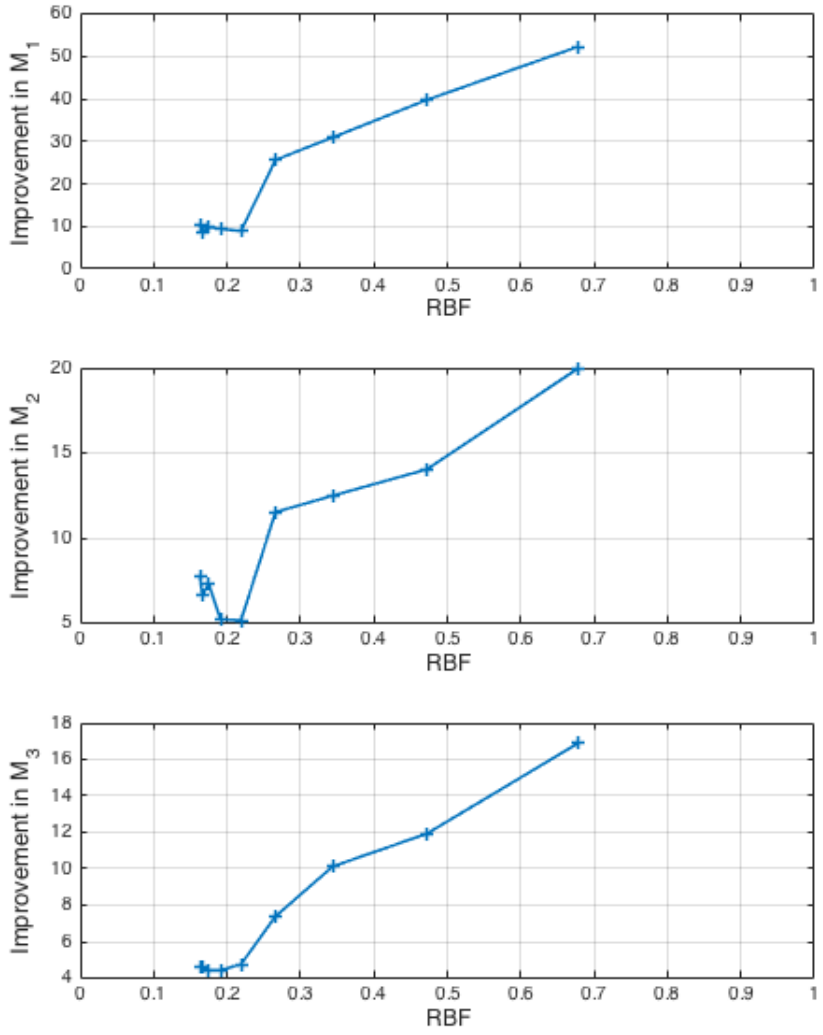


Figure 4.6: Relationship between the improvement our algorithm makes in trading-off between two objectives versus the similarity between those objectives. The more similar objectives (higher RBF) the more our algorithm is apt at finding a compromise. For example, after running our algorithm for 6 iterations, we achieve an improvement in M1 of about 51 for objectives that are as similar as RBF $= 0.68$ (top graph). This means that our algorithm reduced the average euclidian distance to those objectives by about 50 times

(*zeros*) that we think are relevant for a location – See Section 3.3.3 for a brief description of why we classify design points as such. These *zeros* design points are chosen in such a way to maximize the minimum distance between them and the

non-zeros design points already in place. This avoids clusters of design points and gives a suitable support for GaSP emulators. For each location, we fit a GaSP emulator over the *sub-design*, which we then evaluate over a dense grid, and extract the $h_{\text{crit}} = 1\text{m}$ contour. Figure 4.7 illustrates an example of one such *sub-design*. The IMSE is calculated using equation 4.2, with $\sigma_\epsilon = 0.1$, and at each iteration of our algorithm we add 5 new design points, up to 10 iterations.

Since in this case the true contour is unknown, we use a modified version of our metrics of 'goodness'. We replace C_T with C_{j-1} in equation (4.5) and (4.6). Now, the metric M_1 measures the average distance between two *successive* contours C_{j-1} and C_j. The same applies to M_2. We now interpret our metrics as measuring the stability of our estimates of contours. In other words, as these metrics go to 0, this indicates that our approximations are becoming more and more similar to one another, hinting that we may be getting closer to the true solution.

Figure 4.8 shows the results we obtain from refining our estimates of the probability of inundation of two locations: the center of the Town of Mammoth Lakes (CA), and another location 100 meters away. As with the previous example, we can see that our algorithm is apt at reducing our metrics for multiple contours.

However, like in Figure 4.6, it is not trivial to correlate the similarity between two objectives and how well our algorithm balances them. We suspect that is made harder by the fact that the obvious high correlation between the topography and pile height adds an extra spatial-orientation aspect to the problem that standard GaSP emulation is blind to. In other words, two locations that are close together as-the-bird-fly may be really hard to optimize for simultaneously due to the difference in altitude of those two locations. This happens very often in the mountainous region of the Long Valley Caldera.

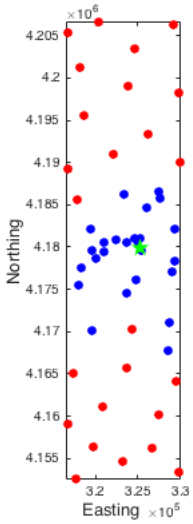


Figure 4.7: Example of a *sub-design* DoE for one location, marked in green, in the Long Valley Region. Blue dots are design points that lead to flows height between 20 cm and 100 meters. Red dots are design points that lead to flows below 20 cm. The red dots were chosen in a space-filling manner to cover the rest of the input space

4.5 Results & Discussion

In this chapter, we presented an iterative design of experiment algorithm that incorporates objective-specific information from various objectives and finds the right compromise Design of Experiment scheme suitable for all objectives. In particular, we look at the case where an objective is defined as a contour level from a high-dimensional function (or black-box model). This algorithm uses Gaussian Response Surface (GaSP) models to emulate the function in consideration and a constrained optimization strategy to choose new design points in such a way as to improve estimation of contours. We also discovered a relationship between objectives that can be combined together and how well our algorithm would do in trading-off between them (See Figure 4.6 and 4.8). In short, the more objectives agree with each other, the easier and the more apt our algorithm is at finding a trade-off between them.

This algorithm is used in estimating inundation contours for locations in the Long Valley volcanic region of California. In this problem, the black-box model is the

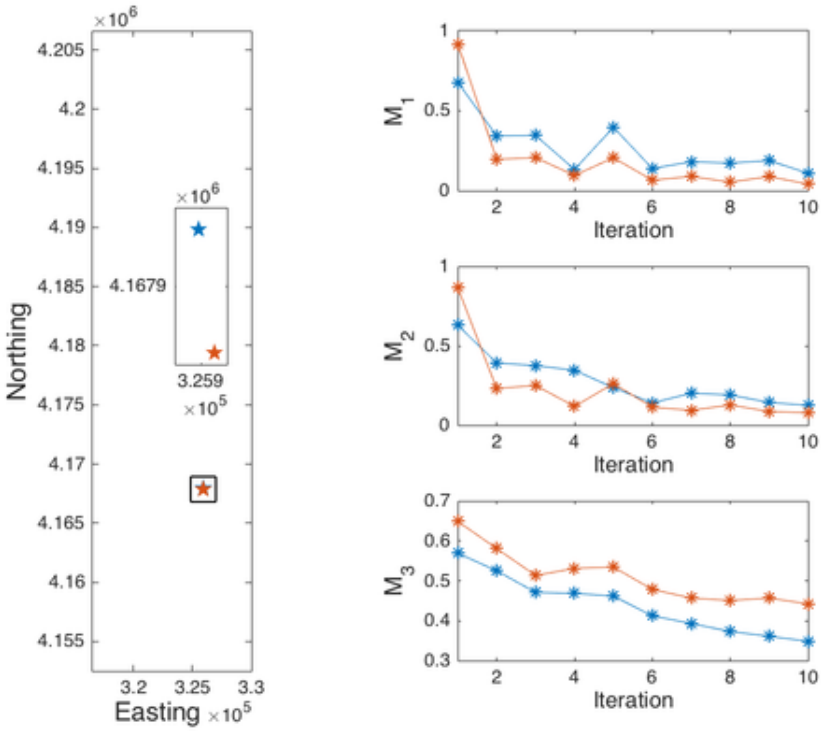


Figure 4.8: Results of our multi-objective sequential algorithm to enhance our approximation of inundation contours for two different location simultaneously (marked with two stars on the left picture). Plots of convergence metrics (on the right side) show that the algorithm reduces the variance of a GaSP emulator near the contour (M_3), and that our estimated contours are slowly converging toward a stable solution

computationally intensive geophysical moodel TITAN2D . *Our algorithm allows us to improve estimate of inundation contours for more than one location at a time using fewer runs of TITAN2D* .

It is to be noted that, our approach is a general strategy for creating a Design of Experiments that is targeted for multiple response-specific goals. As such, it can be used in any situation where one (or more) utility function(s) can be defined and a user has to satisfy all of them. Also, our use of Gaussian-based Response surface is specific to this case and be replaced by any problem specific meta-model.

Chapter 5

Conclusion

In the second chapter, we develop a model that is very useful for either volcanic sites with a dearth of data, or any newly active sites that becomes of interest. This model draws from the similarity in mobility phenomena across all volcanic sites, makes use of the global dataset available about mobility metrics and proposes a characterization of the mobility metric relationship for new locations with very little data. We tested this model on basal friction and volume for 5 volcanic sites, one of which had very little data compared to the others.

In the third chapter, we use the model developed in the previous chapter to quantify the uncertainty in the basal friction-volume relationship within TITAN2D simulations of pyroclastic flows and use GaSP models to allows for fast computation of probability of inundation. These probabilities are then illustrated in hazard maps. The process we propose allows for a flexible plug-in of any scenarios model and a fast computation of probability of inundations for large areas.

In the fourth chapter, we develop a strategy for even faster generation of hazard maps. This is achieved by recognizing the similarities in flow deposits of different locations and devising a clustering metric and algorithm to improve the accuracy of inundation contour computations for multiple locations at the same time. This both speeds up map creation and reduces the number of expensive computer model runs (TITAN2D) needed for hazard assessment.

Risk analysis and uncertainty quantification in volcanology is a complex field of study. There are many factors that contributes to this. These include; the complexity of the physical phenomena, the paucity of data available, the difficulty of collecting data, the statistical analysis – and identification – of epistemic and aleatoric

uncertainty, the inherent inability to provide conventional confidence bounds on forecasts, the communication challenge between scientists and policymakers/public, the human lives and financial interests at stake, and many more.

It is an immense task to develop a process that tackles all of these obstacles at once, as they span multiple disciplines and require the collaboration of a wide range of concerned parties; from the public to policymakers to scientists. This has been the objective of the research group that I am a member of, and the research grants that have been funding my work [1]. Ultimately, the goal of this dissertation is to provide a *fast* and *flexible* software tool for policymakers to quickly assess hazard threats based on various scenarios and make well-informed decision. We foresee our methodology being used in a model selection process or advisory role. We achieve this by developing new statistical forecasting approach along with software to reduce and quantify uncertainty, and allow for fast and accurate assessment of hazard threats.

There are a couple of avenues to follow up on. First, we might explore more specialized covariance kernel functions for GaSP models to handle the non-linearity and discontinuities from TITAN2D output. While the Matérn 5/2 is appropriate for modeling non-linear systems, TITAN2D represents a unique challenge as uncertainty in digital elevation maps, which drive flow movement, affect TITAN2D results.

Secondly, while developing our multi-objective strategy in chapter IV, we noted that more study ought to be done in the selection of new design points as non-linearity and topography can make it very hard to identify similar location on a digital elevation map. Sometimes locations close together as the crow flies can have very distinct TITAN2D output, which leads to drastically different inundation contours. This is caused by either the non-linearity of TITAN output or topography orientation. A spatially-aware flavor of GaSP models that takes into account a region's topography would be useful in these cases.

[1] See Acknowledgement section

Bibliography

[1] I. Alberico, L. Lirer, P. Petrosino, and R. Scandone. A methodology for the evaluation of long-term volcanic risk from pyroclastic flows in Campi Flegrei (Italy). *Journal of Volcanology and Geothermal Research*, 116(1–2):63–78, 2002.

[2] F. Alfano, C. Bonadonna, A. Volentik, C. Connor, S. Watt, D. Pyle, and L. Connor. Tephra stratigraphy and eruptive volume of the May, 2008, Chaiten eruption, Chile. *Bulletin of Volcanology*, 73(5):613–630, 2011.

[3] H. Arenbeck, S. Missoum, A. Basudhar, and P. Nikravesh. Reliability-based optimal design and tolerancing for multibody systems using explicit design space decomposition. *Journal of Mechanical Design*, 132(2):021010, 2010.

[4] R. Bailey. Geologic map of long valley caldera, mono-inyo craters volcanic chain and vicinity, eastern california, scale 1: 62,500, 11 pp. *US Geol. Surv. Geophys. Invest. Map I?1993*, 1989.

[5] R. Bailey, G. Dalrymple, and M. Lanphere. Volcanism, structure, and geochronology of long valley caldera, mono county, california. *Journal of Geophysical Research*, 81(5):725–744, 1976.

[6] R. Barlow, D. Bartholomew, J. Bremner, and H. Brunk. *Statistical inference under order restrictions: The theory and application of isotonic regression*. Wiley New York, 1972.

[7] A. Basudhar and S. Missoum. An improved adaptive sampling scheme for the construction of explicit boundaries. *Structural and Multidisciplinary Optimization*, 42(4):517–529, 2010.

[8] M. Bayarri, J. Berger, E. Calder, K. Dalbey, S. Lunagomez, A. Patra, E. Pitman, E. Spiller, and R. Wolpert. Using statistical and computer models to quantify volcanic hazards. *Technometrics*, 51(4):402–413, 2009.

[9] M. Bayarri, J. Berger, E. Calder, A. Patra, E. Pitman, E. Spiller, and R. Wolpert. Probabilistic quantification of hazards: a methodology using small ensembles of physics-based simulations and statistical surrogates. *International Journal for Uncertainty Quantification*, 5(4):297–325, 2015.

[10] L. Benson, B. Linsley, J. Smoot, S. Mensing, S. Lund, S. Stine, and A. Sarna-Wojcicki. Influence of the pacific decadal oscillation on the climate of thesierra nevada, california and nevada. *Quaternary Research*, 59(2):151–159, 2003.

[11] J. Berger and J. Bernardo. Reference priors in a variance components problem. In *Bayesian Analysis in Statistics and Econometrics*, volume 75 of *Lecture Notes in Statistics*, pages 177–194. Springer, New York, NY, 1992.

[12] J. Berger and A. Limke. Two-dimensional kinematic and rheological modeling of the 1912 pyroclastic flow, Katmai, Alaska. *Bulletin of Volcanology*, 50(3):148–160, 1988.

[13] A. Bevilacqua, M. Bursik, P. A.K., and P. E.B. Bayesian construction of a long-term vent opening probability map in the long valley volcanic region (ca, usa). in press, 2017.

[14] G. Boudon, G. Camus, A. Gourgaud, and J. Lajoie. The 1984 nuée-ardente deposits of Merapi volcano, Central Java, Indonesia: stratigraphy, textural characteristics, and transport mechanisms. *Bulletin of Volcanology*, 55(5):327–342, 1993.

[15] J. Bourdier and E. Abdurachman. Decoupling of small-volume pyroclastic flows and related hazards at Merapi volcano, Indonesia. *Bulletin of Volcanology*, 63(5):309–325, 2001.

[16] P. Bridgman. *Dimensional analysis*. Yale University Press, 1922.

[17] S. Brown, H. Crosweller, R. Sparks, E. Cottrell, N. Deligne, N. Guerrero, L. Hobbs, K. Kiyosugi, S. Loughlin, L. Siebert, and S. Takarada. Characterisation of the Quaternary eruption record: analysis of the Large Magnitude Explosive Volcanic Eruptions (LaMEVE) database. *Journal of Applied Volcanology*, 3:22, 2014.

[18] B. Browne, M. Bursik, J. Deming, M. Louros, A. Martos, and S. Stine. Eruption chronology and petrologic reconstruction of the ca. 8500 yr bp eruption of red cones, southern inyo chain, california. *Geological Society of America Bulletin*, page B30070. 1, 2010.

[19] S. Burkett. *Geomorphic mapping and petrography of mammoth mountain, California*. ProQuest, 2007.

[20] M. Bursik. Personal Communication, 2015.

[21] M. Bursik, A. Patra, E. Pitman, C. Nichita, J. Macias, R. Saucedo, and O. Girina. Advances in studies of dense volcanic granular flows. *Reports on Progress in Physics*, 68(2):271, 2005.

[22] M. Bursik and K. Sieh. . *Digital database of the Holocene tephras of the Mono-Inyo Craters, California*, 2013.

[23] E. Calder, P. Cole, W. Dade, T. Druitt, R. Hoblitt, H. Huppert, L. Ritchie, R. Sparks, and S. Young. Mobility of pyroclastic flows and surges at the Soufrière Hills Volcano, Montserrat. *Geophysical Research Letters*, 26(5):537–540, 1999.

[24] L. Capra, V. Manea, M. Manea, and G. Norini. The importance of digital elevation model resolution on granular flow simulations: a test case for Colima volcano using TITAN2D computational routine. *Natural Hazards*, 59(2):665–680, 2011.

[25] G. Casella and E. George. Explaining the Gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.

[26] S. Charbonnier, A. Germa, C. Connor, R. Gertisser, J. JKomorowski, K. Preece, F. Lavigne, T. Dixon, and L. Connor. Evaluation of the impact of the 2010 pyroclastic density currents at Merapi volcano from high-resolution satellite imagery, field investigation and numerical simulations. *Journal of Volcanology and Geothermal Research*, 261:295–315, 2013.

[27] S. Charbonnier and R. Gertisser. Deposit architecture and dynamics of the 2006 block-and-ash flows of Merapi Volcano, Java, Indonesia. *Sedimentology*, 58(6):1573–1612, 2011.

[28] S. Charbonnier and R. Gertisser. Evaluation of geophysical mass flow models using the 2006 block-and-ash flows of merapi volcano, java, indonesia: Towards a short-term hazard assessment tool. *Journal of Volcanology and Geothermal Research*, 231–232:87–108, 2012.

[29] S. Charbonnier, J. Palma, and S. Ogburn. Application of 'shallow-water' numerical models for hazard assessment of volcanic flows: the case of TITAN2D and Turrialba volcano (Costa Rica). *Revista Geologica de America Central*, pages 107–128, 2015.

[30] M. C.M. and M. F.S. Computer-assisted mapping of pyroclastic surges. *Science*, 217(4560):637–640, 1982.

[31] P. Cole, E. Calder, R. Sparks, A. Clarke, T. Druitt, S. Young, R. Herd, C. Harford, and G. Norton. Deposits from dome-collapse and fountain-collapse pyroclastic flows at Soufrière Hills Volcano, Montserrat. In T. H. Druitt and B. P. Kokelaar, editors, *The eruption of Soufrière Hills Volcano, Montserrat, from 1995 to 1999*, volume 21, pages 231–262. Geological Society of London, Memoirs, 2002.

[32] P. Cole, P. Smith, A. Stinton, H. Odbert, M. Bernstein, J. Komorowski, and R. Stewart. Vulcanian explosions at Soufriere Hills Volcano, Montserrat between 2008 and 2010. In *The eruption of Soufrière Hills Volcano, Montserrat, from 2000 to 2010*, volume 39, pages 93–111. Geological Society of London, Memoirs, 2014.

[33] S. Conti, J. Gosling, J. Oakley, and A. O'hagan. Gaussian process emulation of dynamic computer codes. *Biometrika*, page asp028, 2009.

[34] G. Crosta, S. Cucchiaro, and P. Frattini. Validation of semi-empirical relationships for the definition of debris-flow behavior in granular materials. In D. Rickenmann and C.-L. Chen, editors, *Proceedings of the Third International Conference on Debris-Flow Hazards Mitigation: Mechanics, Prediction and Assessment, Davos, Switzerland*, pages 821–831. Millpress, Rotterdam, NL, 2003.

[35] H. Crosweller, B. Arora, S. Brown, E. Cottrell, N. Deligne, N. Guerrero, L. Hobbs, K. Kiyosugi, S. Loughlin, J. Lowndes, M. Nayembil, L. Siebert, R. Sparks, S. Takarada, and E. Venzke. Global database on large magnitude explosive volcanic eruptions (LaMEVE). *Journal of Applied Volcanology*, 1(4):13, 2012.

[36] J. Crowley, B. Schoene, and S. Bowring. U-pb dating of zircon in the bishop tuff at the millennial scale. *Geology*, 35(12):1123–1126, 2007.

[37] W. Dade and H. Huppert. Long-runout rockfalls. *Geology*, 26(5):803–806, 1998.

[38] A. Del Pozzo, A. Lillian, M. Sheridan, D. Barrera, L. Hubp, and V. Lorenzo. Potential hazards from Colima volcano, Mexico. *Geofísica Internacional*, 34(4):363–376, 1995.

[39] N. Deligne, S. Coles, and R. Sparks. Recurrence rates of large explosive volcanic eruptions. *Journal of Geophysical Research*, 115(B6), 2010.

[40] S. Doocy, A. Daniels, S. Dooling, and Y. Gorokhovich. The human impact of volcanoes: a historical review of events 1900-2009 and systematic literature review. *PLoS currents*, 5, 2013.

[41] R. Fisher and H. Schmincke. *Pyroclasic Rocks*. Springer-Verlag, New York, NY, 1984.

[42] P. Francis and M. Baker. Mobility of pyroclastic flows. *Nature*, 270:164–165, 1977.

[43] C. Furlan. Extreme value methods for modeling historical series of large volcanic magnitudes. *Statistical Modelling*, 10(2):113–132, 2010.

[44] Geological Survey of Japan and the National Institute of Advanced Industrial Science and Technology (AIST). Catalog of eruptive events during the last 10,000 years in japan, version 2.2, 2013.

[45] Global Volcanism Program. Volcanoes of the World, v. 4.4.0, 2013.

[46] J. Griswold and R. Iverson. Mobility and statistics and automated hazard mapping for debris flows and rock avalanches, 2008. U. S. Geological Survey Scientific Investigations Report 2007–5276.

[47] J. Guest and J. Murray. An analysis of hazard from mount etna volcano. *Journal of the Geological Society*, 136(3):347–354, 1979.

[48] V. Hards, M. Strutt, S. De Angelis, G. Ryan, T. Christopher, T. Syers, and V. Bass. Report to the Scientific Advisory Committee on activity at Soufrière Hills Volcano Montserrat: April 2008. *Montserrat Volcano Observatory Open File Report*, OFR 09-01, 2008.

[49] J. Hayashi and S. Self. A comparison of pyroclastic flow and debris avalanche mobility. *Journal of Geophysical Research B: Solid Earth*, 97(B6):9063–9071, 1992.

[50] A. Heim. Bergsturz und Menschenleben. *Zürich Vierteljahrsschrift*, 77:218, 1932. [English translation by N. A. Skermer, 1989. Landslides and human lives: Vancouver, B.C., BiTech Publishers, 195p].

[51] W. Hildreth. Volcanological perspectives on long valley, mammoth mountain, and mono craters: several contiguous but discrete systems. *Journal of Volcanology and Geothermal Research*, 136(34):169 – 198, 2004.

[52] D. Hill. Unrest in long valley caldera, california, 1978–2004. *Geological Society, London, Special Publications*, 269(1):1–24, 2006.

[53] D. Hooper and G. Mattioli. Kinematic modeling of pyroclastic flows produced by gravitational dome collapse at Soufrière Hills Volcano, Montserrat. *Natural Hazards*, 23(1):65–86, 2001.

[54] A. Höskuldsson and J. Cantagrel. Volcanic hazards in the surroundings of Pico de Orizaba, eastern Mexico. *Natural Hazards*, 10(3):197–219, 1994.

[55] K. Hsü. Catastrophic debris streams (sturtzstroms) generated by rockfalls. *Geological Society of America Bulletin*, 86(1):129–140, 1975.

[56] O. Hungr. Mobility of rock avalanches. *Report of the National Research Institute for Earth Science and Disaster Prevention, Japan*, 46:11–20, 1990.

[57] R. Iverson. The physics of debris flows. *Reviews of geophysics*, 35(3):245–296, 1997.

[58] R. Iverson and R. Denlinger. Flow of variably fluidized granular masses across three-dimensional terrain: 1. coulomb mixture theory. *Journal of Geophysical Research: Solid Earth*, 106(B1):537–552, 2001.

[59] R. Iverson, S. Schilling, and J. Vallance. Objective delineation of lahar-inundation hazard zones. *Geological Society of America Bulletin*, 110(8):972–984, 1998.

[60] M. Johnson, L. Moore, and D. Ylvisaker. Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2):131 – 148, 1990.

[61] D. Jones, M. Schonlau, and W. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, DEC 1998.

[62] K. Kelfoun and T. Druitt. Numerical modelling of the emplacement of the 7500 bp socompa rock avalanche, chile. *Journal of Geophysical Research B: Solid Earth*, 110(B12):B12202, 2005.

[63] C. Kilburn and S. Sorenson. Runout lengths of sturztroms; the control of initial conditions and of fragment dynamics. *Journal of Geophysical Research B: Solid Earth*, 103(B8):17877–17884, 1998.

[64] K. Kiyosugi, C. Connor, R. Sparks, H. Crosweller, S. Brown, L. Siebert, T. Wang, and S. Takarada. How many explosive eruptions are missing from the geologic record? Analysis of the quaternary record of large magnitude explosive eruptions in Japan. *Journal of Applied Volcanology*, 4(1):17, 2015.

[65] J. Komorowski, S. Jenkins, P. Baxter, A. Picquout, F. Lavigne, S. Charbonnier, R. Gertisser, K. Preece, N. Cholik, A. Budi-Santoso, and Surono. Paroxysmal dome explosion during the Merapi 2010 eruption: Processes and facies relationships of associated high-energy pyroclastic density currents. *Journal of Volcanology and Geothermal Research*, 261:260–294, 2013.

[66] J. Komorowski, Y. Legendre, T. Christopher, M. Bernstein, R. Stewart, E. Joseph, N. Fournier, L. Chardot, A. Finizola, G. Wadge, R. Syers, C. Williams, and V. Bass. Insights into processes and deposits of hazardous vulcanian explosions at Soufrière Hills Volcano during 2008 and 2009 (Montserrat, West Indies). *Geophysical Research Letters*, 37(19):1–6, 2010.

[67] T. Kover and M. Sheridan. Numerical models for pyroclastic flows of mt. unzen, japan (abstract), 1993. Geological Society of America, Abstracts with programs 25, 268.

[68] R. Kuczera and Z. Mourelatos. On estimating the reliability of multiple failure region problems using approximate metamodels. *Journal of Mechanical Design*, 131(12):121003, 2009.

[69] L. Lin and D. Dunson. Bayesian monotone regression using gaussian process projection. *Biometrika*, page ast063, 2014.

[70] S. Loughlin, R. Luckett, and G. Ryan. An overview of lava dome evolution, dome collapse and cyclicity at Soufriere Hills Volcano, Montserrat, 2005-2007. *Geophysical Research Letters*, 37:1–6, 2010.

[71] G. Mahood, J. Ring, and M. McWilliams. Contemporaneous mafic and silicic eruptions during the past 160 ka at long valley caldera, californai: implications of new 40ar/39ar eruption ages for current volcanic hazards. *EOS, Transactions, American Geophysical Union*, 81(48):F1321, 2000.

[72] M. Malin and M. Sheridan. Computer-assisted mapping of pyroclastic surges. *Science*, 217(4560):637–640, 1982.

[73] M. Marcaida, M. Mangan, J. Vazquez, M. Bursik, and M. Lidzbarski. Geochemical fingerprinting of wilson creek formation tephra layers (mono basin, california) using titanomagnetite compositions. *Journal of Volcanology and Geothermal Research*, 273:1–14, 2014.

[74] W. McGuire. Volcanic hazards and their mitigation. *Geological Society, London, Engineering Geology Special Publications*, 15(1):79–95, 1998.

[75] M. McKay, R. Beckman, and W. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, May 1979.

[76] R. Mellors, R. Waitt, and D. Swanson. Generation of pyroclastic flows and surges by hot-rock avalanches from the dome of mount st. helens volcano, usa. *Bulletin of Volcanology*, 50(1):14–25, 1988.

[77] H. Murcia, M. Sheridan, J. Macías, and G. Cortés. TITAN2D simulations of pyroclastic flows at Cerro Machín Volcano, Colombia: Hazard implications. *Journal of South American Earth Sciences*, 29(2):161–170, 2010.

[78] B. Myers, S. Brantley, and J. Hendley II. *What are Volcano Hazards?* U.S. Geological Survey, Vancouver, WA, 002–97 edition, March 2008.

[79] I. Nairn and S. Self. Explosive eruptions and pyroclastic avalanches from Ngauruhoe in February 1975. *Journal of Volcanology and Geothermal Research*, 3(1–2):39–60, 1978.

[80] S. Nakada, H. Shimizu, and K. Ohta. Overview of the 1990–1995 eruption at unzen volcano. *Journal of Volcanology and Geothermal Research*, 89(1):1–22, 1999.

[81] J. Nash Jr. The bargaining problem. *Econometrica: Journal of the Econometric Society*, pages 155–162, 1950.

[82] S. Ogburn. Potential hazards at Soufrière Hills Volcano, Montserrat: Northwards-directed dome-collapses and major explosive eruptions. Master's thesis, State University of New York at Buffalo, Buffalo, NY, 2008.

[83] S. Ogburn. DomeHaz: Dome-forming eruptions database, 2012. On Vhub at `https://vhub.org/groups/domedatabase`.

[84] S. Ogburn. FlowDat: Mass flow database, 2012. On Vhub at `https://vhub.org/groups/massflowdatabase`.

[85] S. Ogburn. *Reconciling field observations of pyroclastic density currents with conceptual and computational analogs using a GIS and a newly developed global database*. PhD thesis, State University of New York at Buffalo, Buffalo, NY, 2014.

[86] S. Ogburn, S. Loughlin, and E. Calder. The association of lava dome growth with major explosive activity (vei 4): Domehaz, a global dataset. *Bulletin of Volcanology*, 77:1–17, 2015.

[87] A. Patra, A. Bauer, C. Nichita, E. Pitman, M. Sheridan, and M. Bursik. Parallel adaptive numerical simulation of dry avalanches over natural terrain. *Journal of Volcanology and Geothermal Research*, 139(1–2):1–21, 2005.

[88] V. Picheny, D. Ginsbourger, O. Roustant, R. Haftka, and N. Kim. Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design*, 132(7):071008, JUL 2010.

[89] E. Pitman, C. Nichita, A. Patra, A. Bauer, M. Sheridan, and M. Bursik. Computing granular avalanches and landslides. *Physics of fluids*, 15(12):3638–3646, 2003.

[90] P. Ranjan, D. Bingham, and G. Michailidis. Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4):527–541, 11/01; 2015/03 2008.

[91] C. Rasmussen. *Gaussian processes for machine learning*. Citeseer, 2006.

[92] J. Sacks, S. Schiller, and W. Welch. Designs for computer experiments. *Technometrics*, 31(1):41–47, 1989.

[93] R. Saucedo, J. Macias, and M. Bursik. Pyroclastic flow deposits of the 1991 eruption of Volcán de Colima, Mexico. *Bulletin of Volcanology*, 66(4):291–306, 2004.

[94] R. Saucedo, J. Macias, M. Bursik, J. Mora, J. Gavilanes, and A. Cortes. Emplacement of pyroclastic flows during the 1998-1999 eruption of Volcán de Colima, México. *Journal of Volcanology and Geothermal Research*, 117:129–153, 2002.

[95] R. Saucedo, J. Macias, J. Gavilanes, J. Arce, J. Komorowski, J. Gardner, and G. Valdez-Moreno. Eyewitness, stratigraphy, chemistry, and eruptive dynamics of the 1913 Plinian eruption of Volcán de Colima, México. *Journal of Volcanology and Geothermal Research*, 191(3-4):149–166, 2010.

[96] S. Savage and K. Hutter. The motion of a finite mass of granular material down a rough incline. *Journal of Fluid Mechanics*, 199:177–215, 1989.

[97] A. Scheidegger. On the prediction of the release and velocity of catastrophic rockfalls. *Rock Mechanics*, 5(4):231–236, 1973.

[98] A. Scheller. Beitrag zum Bewegungsverhalten grosser Berstürze. *Eclogae Geologicae Helvetiae*, 64:195–202, 1971.

[99] L. Schwarzkopf, H. Schmincke, and S. Cronin. A conceptual model for block-and-ash flow basal avalanche transport and deposition, based on deposit architecture of 1998 and 1994 Merapi flows. *Journal of Volcanology and Geothermal Research*, 139(1-2):117–134, 2005.

[100] S. Shan and G. Wang. Space exploration and global optimization for computationally intensive design problems: a rough set based approach. *Structural and Multidisciplinary Optimization*, 28(6):427–441, 2004.

[101] T. Sheldrake. Long-term forecasting of eruption hazards: A hierarchical approach to merge analogous eruptive histories. *Journal of Volcanology and Geothermal Research*, 286:15–23, 2014.

[102] M. Sheridan. Emplacement of pyroclastic flows— a review. In C. E. Chapin and W. E. Elston, editors, *Ash-Flow Tuffs*, pages 125–136. Geological Society of America, Special Paper 180, 1979.

[103] M. Sheridan, B. Hubbard, G. Carrasco-Núñez, and C. Siebe. Pyroclastic flow hazard at Volcán Citlaltépetl. *Natural Hazards*, 33(2):209–221, 2004.

[104] M. Sheridan and J. Macias. Estimation of risk probability for gravity-driven pyroclastic flows at Volcán Colima, México. *Journal of Volcanology and Geothermal Research*, 66(1–4):251–256, 1995.

[105] M. Sheridan and M. Malin. Application of computer-assisted mapping to volcanic hazard evaluation of surge eruptions: Vulcano, Lipari, and Vesuvius. *Journal of Volcanology and Geothermal Research*, 17(1–4):187–202, 1983.

[106] M. Sheridan, A. Patra, K. Dalbey, and B. Hubbard. Probabilistic digital hazard maps for avalanches and massive pyroclastic flows using TITAN2D. In *Stratigraphy and geology of volcanic areas*, pages 281–291. Geological Society of Amer, 2010.

[107] L. Siebert, T. Simkin, and P. Kimberly. *Volcanoes of the World*. Smithsonian Institution, Washington D.C and University of Califronia Press, Berkeley, 2010.

[108] K. Sieh and M. Bursik. Most recent eruption of the mono craters, eastern central california. *Journal of Geophysical Research: Solid Earth*, 91(B12):12539–12571, 1986.

[109] R. Sparks, L. Wilson, and G. Hulme. Theoretical modeling of the generation, movement, and emplacement of pyroclastic flows by column collapse. *Journal of Geophysical Research: Solid Earth*, 83(B4):1727–1739, 1978.

[110] E. Spiller, M. Bayarri, J. Berger, E. Calder, A. Patra, E. Pitman, and R. Wolpert. Automating emulator construction for geophysical hazard maps. *SIAM/ASA Journal on Uncertainty Quantification*, pages 126–152, 01/01; 2014/03 2014.

[111] M. Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.

[112] R. Stephen and J. Sparks. Grain size variations in ignimbrites and implications for the transport of pyroclastic flows. *Sedimentology*, 23(2):147–188, 1976.

[113] S. Stine. Late holocene fluctuations of mono lake, eastern california. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 78(3-4):333–381, 1990.

[114] A. Stinton. *Effects of changes in valley geomorphology on the behavior of volcanic mass-flows*. PhD thesis, State University of New York at Buffalo, Buffalo, NY, 2014.

[115] S. Takarada, 2008. (personal communication).

[116] G. Taylor, G. Australia Bureau of Mineral Resources, and Geophysics. *The 1951 eruption of Mount Lamington, Papua*. Canberra: Australian Government Publishing Service, 2nd ed edition, 1983.

[117] J. Thouret, F. Lavigne, H. Suwa, B. Sukatja, and Surono. Volcanic hazards at Mount Semeru, East Java (Indonesia), with emphasis on lahars. *Bulletin of Volcanology*, 70(2):221–244, 2007.

[118] US Geological Survey. Long valley caldera, May 2017.

[119] J. Vallance, K. Bull, and M. Coombs. Pyroclastic flows, lahars and mixed avalanches generated during the 2006 eruption of Augustine Volcano. In *The 2006 eruption of Augustine Volcano, Alaska: Professional Paper 1769*, chapter 10, pages 219–267. US Geological Survey, Menlo Park, CA, 2010.

[120] J. Vazquez and M. Lidzbarski. High-resolution tephrochronology of the wilson creek formation (mono lake, california) and laschamp event using 238 u-230 th sims dating of accessory mineral rims. *Earth and Planetary Science Letters*, 357:54–67, 2012.

[121] D. Venezky and C. Newhall. WOVOdat design document: the schema, table descriptions, and create table statements for the database of worldwide volcanic unrest (WOVOdat Version 1.0). *Geological Survey Open File Report 2007-1117*, page 184, 2007.

[122] F. Viana, R. Haftka, and L. Watson. Sequential sampling for contour estimation with concurrent function evaluations. *Structural and Multidisciplinary Optimization*, 45(4):615–618, APR 2012.

[123] G. Wadge and M. Isaacs. Mapping the volcanic hazards from the Soufrière Hills Volcano,Montserrat, West Indies, using an image processor. *Journal of the Geological Society, London*, 145(4):541–551, 1988.

[124] S. Watt, D. Pyle, and T. Mather. Evidence of mid- to late-Holocene explosive rhyolitic eruptions from Chaiten Volcano, Chile. *Andean Geology*, 40(2):216–226, 2013.

[125] P. Whelley, C. Newhall, and K. Bradley. The frequency of explosive volcanic eruptions in Southeast Asia. *Bulletin of Volcanology*, 77(1):1, 2015.

[126] C. Widiwijayanti, B. Voight, D. Hidayat, and S. Schilling. Objective rapid delineation of areas at risk from block-and-ash pyroclastic flows and surges. *Bulletin of Volcanology*, 71(6):687–703, 2008.

[127] J. William, J. Robert, S. Jerome, H. P.W., T. J.M., and M. D.M. Screening, predicting, and computer experiments. *Technometrics*, 34(1):15–25, 1992.

[128] S. Wood. Chronology of late pleistocene and holocene volcanics, long valley and mono basin geothermal areas, eastern california. *Chronology of Late Pleistocene and Holocene Volcanics, Long Valley and Mono Basin Geothermal Areas, Eastern California*, 1983.

[129] S. Zimmerman, S. Hemming, D. Kent, and S. Searle. Revised chronology for late pleistocene mono lake sediments based on paleo-intensity correlation to the global reference curve. *Earth and Planetary Science Letters*, 252(1):94–106, 2006.

Appendix A

Vent Location Dataset

Age (in years)	Northing	Easting	Age (in years)	Northing	Easting
215	4207338.98	321139.72	89000	4166281	320516
612	4180931.99	322095.45	91000	4170644	322239
612	4175778.42	322217.02	91000	4166665	320453
612	4199888.73	320281.11	93000	4170234	321145
612	4199071.61	321810.41	94000	4169947	324984
612	4197362.53	322115.8	95000	4167441	330730
612	4197823.21	322210.28	95000	4170250	319655
612	4196478.97	323426.6	95000	4170811	319278
612	4173782.23	322722.57	98000	4162718	321172
612	4180931.99	322095.45	103000	4168943	325551
612	4175778.42	322217.02	104000	4179940	329682
612	4199888.73	320281.11	106000	4171817	324472
612	4199888.73	320281.11	107000	4177151	326926
612	4199888.73	320281.11	110000	4171829	324833
612	4196478.97	323426.6	110000	4169741	324873
612	4196478.97	323426.6	111000	4165278	322213
612	4179254.58	322086.47	113000	4174088	322324
612	4180931.99	322095.45	118000	4167675	321647
612	4175778.42	322217.02	123000	4168323	321398
612	4175778.42	322217.02	126000	4167964	320293
702	4168092	321661	126000	4167161	330139
1329	4190881.56	323513.7	145000	4175487	327997
1574	4197823	322210	145000	4168477	322979
1660.5	4183588.98	321557	149000	4171436	327187
1660.5	4183588.98	321557	151000	4171051	327178
6000	4177151	326926	164000	4174962	328459
6000	4176755.63	321812.26	165000	4173537	326299
8490	4162569.97	318364.49	171000	4169741	324873
21680	4189650.19	323739.6	192000	4181015	343956
24000	4177498	324397	210000	4173537	326299
29000	4179494	325524	284000	4169841	322254
32000	4177529	324095	324000	4167314	340785
37000	4178534	323437	349000	4170668	340685
38000	4178356	322532	468000	4176884	332260
39800	4193120.39	322710.5	509000	4176297	333574
38700	4193120.39	322710.5	634000	4174471	328736

40800	4193120.39	322710.5	658000	4176193	331020
44000	4177529	324095	660000	4172116	330428
62000	4167439	329611	664000	4171746	332912
63000	4167558	321978	670000	4167671	333052
65000	4167964	320293	673000	4177441	329630
67000	4170644	322239	675000	4169630	329071
76000	4167896	315803	730000	4168762	335186
79000	4168637	320847	767100	4165611	338700
82000	4164986	321146	767100	4165611	338700
82000	4166665	320453			
83000	4167409	323735			
89000	4167275	329849			

Table A.1: Historical Record of Eruptions of the Mono-Inyo Chain and Long Valley Caldera [129, 128, 120, 113, 108, 71, 73, 36, 22, 19, 18, 10, 5]

Appendix B

Hierarchical Bayesian Model Implementation

```
# Implementation of a Hierarchical Bayesian
# for coeficient of friction vs volume relationship

# Author: Regis Rutarindwa

from numpy import *
import time
import matplotlib.pyplot as plt
from scipy import stats

# load data
clm = recfromtxt('Clm_Vol_CoefFriction.txt');
mrp = recfromtxt('Mrp_Vol_CoefFriction.txt');
shv = recfromtxt('Shv_Vol_CoefFriction.txt');
unz = recfromtxt('Unz_Vol_CoefFriction.txt');
smr = recfromtxt('Smr_Vol_CoefFriction.txt');
alldata = [clm,mrp,shv,unz,smr]

name = ['Clm','Mrp', 'SHV','Unz','Smr']
nvolc = len(alldata)

mid = 10**5.5
polys = zeros([nvolc,2])
njs = zeros([nvolc,1])
xj = zeros([nvolc,1])
Sj_lst = zeros([nvolc,1])

# Data manipulation: remember col1 = y, col2 = x
for i in range(nvolc):
    alldata[i][:,1] = alldata[i][:,1] * 10**6
    alldata[i][:,1] = alldata[i][:,1]/mid
    alldata[i] = log10(alldata[i])

    polys[i,:] = polyfit(alldata[i][:,1], alldata[i][:,0],1)
    njs[i] = shape(alldata[i])[0]
    xj[i] = mean(alldata[i][:,1])
    Sj_lst[i] = sum((alldata[i][:,1] - xj[i])**2)

Sj_lst = squeeze(Sj_lst)
```

```
njs = squeeze(njs)
xj = squeeze(xj)


# sample size
nsamples = 10**3
nflows = sum(njs)

# Allocating spaces for arrays
thetas1= zeros([nvolc,nsamples])
thetas2= zeros([nvolc,nsamples])
mus = zeros([nsamples,1])
st2_lst = zeros([nsamples,1])
sj2_lst = zeros([nvolc,nsamples])
lams = zeros([nvolc,nsamples])

# Initializing first values of matrices
t1 = polys[:,1]
t2 = polys[:,0]
thetas1[:,0] = squeeze(t1)
thetas2[:,0] = squeeze(t2)
mus[0] = mean(t2)
sigj2init = map(lambda x: var(x[:,1],ddof=1),alldata)
sj2_lst[:,0] = sigj2init
st2_lst[0] = var(t2,ddof=1)


Sj_lst1 = Sj_lst[:2]
Sj_lst2 = Sj_lst[2:]

vj_func = lambda s_j2,s_t2,Sj_part: (1.0*s_j2/Sj_part)+\
          s_t2

mu_hat_func = lambda t2j, s_j2,s_t2,Sj_part: \
                    sum(t2j/vj_func(s_j2,s_t2,Sj_part))/ \
                    sum(1/vj_func(s_j2,s_t2,Sj_part))

tic = time.time()
q = nsamples/10
for k in range(0,nsamples-1):

    # progress report
    if (k%q == 0):
        print "%3.1f %% done"%(k*10/q)
```

```
84
85      ########################################################
86
87      # STEP 1: Draw th2
88      lamtemp = st2_lst[k]/sj2_lst[:,k]
89      m1 = t2 - (t2 - mus[k])/(1 + lamtemp*Sj_lst)
90      s21 = (sj2_lst[:,k]*lamtemp)/(1 + lamtemp*Sj_lst)
91      thetas2[:,k+1] = squeeze(random.randn(nvolc,1)) *\
92                       sqrt(s21) + m1
93
94      ########################################################
95
96      # STEP 2: Draw th1
97      m = t1 - xj*(thetas2[:,k+1]-t2) #use new t2
98      s22 = sj2_lst[:,k]/njs
99      thetas1[:,k+1] = squeeze(random.randn(nvolc,1)) *\
100                      sqrt(s22) + m
101
102     ########################################################
103
104     # STEP 3A: Draw sig2_u
105     cu = 0
106     found_u = 0
107     while (found_u != 1):
108         btmp = 0
109         for i in range(2):
110             xji = alldata[i][:,1]
111             yji = alldata[i][:,0]
112             btmp = btmp + sum((yji-(thetas1[i,k+1] + \
113                                     xji*thetas2[i,k+1]))**2)
114
115         betau = 0.5*btmp
116         s2_u = 1/random.gamma(sum(njs[0:2])/2, 1.0/betau)
117
118         n1 = sum(1/vj_func(sj2_lst[:,k],\
119                            st2_lst[k],Sj_lst)**2)
120         b1 = sum(1/vj_func(0,st2_lst[k],Sj_lst1)**2)
121         b2 = sum(1/vj_func(sj2_lst[2:,k],\
122                            st2_lst[k],Sj_lst2)**2)
123
124         u_u = random.rand()
125         condu = sqrt(n1)/sqrt(b1 + b2)
126
127         if (u_u < condu):
```

```
            found_u = 1
        cu = cu + 1

    ########################################################

    # STEP 3B: Draw sig2_c
    cc = 0
    found_c = 0
    while (found_c != 1):
        btmp = 0
        for i in range(2,nvolc): #i=3:nvolc
            xji = alldata[i][:,1]
            yji = alldata[i][:,0]
            btmp = btmp + sum((yji-(thetas1[i,k+1]+\
                                    xji*thetas2[i,k+1]))**2)

        betac = 0.5*btmp
        s2_c = 1/random.gamma(sum(njs[2:])/2, 1.0/betac)

        n1 = sum(1/vj_func(sj2_lst[:,k],\
                           st2_lst[k],Sj_lst)**2)
        b1 = sum(1/vj_func(sj2_lst[:2,k],\
                           st2_lst[k],Sj_lst1)**2)
        b2 = sum(1/vj_func(0,st2_lst[k],Sj_lst2)**2)

        u_c = random.rand()
        condc = sqrt(n1)/sqrt(b1 + b2)

        if (u_c < condc):
            found_c = 1

        cc = cc + 1

    for i in range(2):
        sj2_lst[i,k+1] = s2_u
    for i in range(2,5):
        sj2_lst[i,k+1] = s2_c

    ########################################################

    # STEP 4: Draw mu
    vs = vj_func(sj2_lst[:,k+1],st2_lst[k],Sj_lst)
    ms = mu_hat_func(thetas2[:,k+1],sj2_lst[:,k+1],\
                     st2_lst[k],Sj_lst)
```

```
172     mus[k+1] = random.randn(1)*sqrt(1/sum(1/vs)) + ms
173
174     ########################################################
175
176     # STEP 5: Generate sig_t2
177     beta = 0.5*sum((thetas2[:,k+1]-mus[k+1])**2)
178     sig_t2_cand = 1/random.gamma((nvolc-2)/2, 1.0/beta)
179     u_t = random.rand()
180     condt = sqrt(sum(1/vj_func(sj2_lst[:,k+1],\
181                                sig_t2_cand,Sj_lst)**2))/ \
182             sqrt(sum(1/vj_func(sj2_lst[:,k+1],0,Sj_lst)**2))
183
184     ct = 0;
185
186     while ((u_t < condt) == 0):
187         beta = 0.5*sum((thetas2[:,k+1]-mus[k+1])**2)
188         sig_t2_cand = 1/random.gamma((nvolc-2)/2, 1.0/beta)
189
190         u_t = random.rand()
191         condt = sqrt(sum(1/vj_func(sj2_lst[:,k+1],\
192                                    sig_t2_cand,\
193                                    Sj_lst)**2))/ \
194                 sqrt(sum(1/vj_func(sj2_lst[:,k+1],0,\
195                                    Sj_lst)**2))
196
197         ct = ct + 1
198
199     st2_lst[k+1] = sig_t2_cand
200
201
202
203 toc = time.time()
204 print
205 print '%10.3f secs elapsed'%(toc-tic)
206
207
208 burn = nsamples*1.0/100
209 thetas1 = thetas1[:,burn:]
210 thetas2 = thetas2[:,burn:]
211 mus = mus[burn:]
212 st2_lst = st2_lst[burn:]
213 sj2_lst = sj2_lst[:,burn:]
214
215
```

```
216
217 plt.figure(1)
218 plt.clf()
219 for i in range(nvolc):
220     plt.hist(thetas1[i,:], bins=100, \
221              histtype='stepfilled', normed=True,\
222             alpha=0.5, label=name[i])
223 plt.title('Theta1')
224 plt.legend()
225 plt.show()
226
227
228 plt.figure(2)
229 plt.clf()
230 for i in range(nvolc):
231     plt.hist(thetas2[i,:], bins=100, \
232              histtype='stepfilled', normed=True,\
233             alpha=0.5, label=name[i])
234 plt.title('Theta2')
235 plt.legend()
236 plt.show()
237
238
239 ##############################################
240
241 n = 10;
242
243 plt.figure(3)
244 plt.clf()
245 figind = 1
246
247 for j in range(nvolc):#j = 1 ######### volcano index
248
249     plt.subplot(3,2,figind)
250     figind += 1
251     x = alldata[j][:,1]
252     y = alldata[j][:,0]
253     vv = linspace(-1.5, 2.5, n);
254
255     p975 = zeros([n,1]);
256     p25 = zeros([n,1]);
257     p50 = zeros([n,1]);
258     fvv = zeros([n,1]);
259
```

```
260     for i in range(n):
261         cf = thetas2[j,:] * vv[i] + thetas1[j,:]
262         p25[i] = percentile(cf,2.5)
263         p50[i] = percentile(cf,50)
264         p975[i] = percentile(cf,97.5)
265 
266     nn = len(x)
267     p = polyfit(x,y,1)
268     sxx = sum((x - mean(x))**2)
269     yihat = polyval(p,x)
270     seps2 = sum((y - yihat)**2)/(nn-2)
271 
272     topy = zeros([n,1])
273     boty = zeros([n,1])
274 
275     for i in range(n):
276         val = stats.t.ppf(0.975,nn) * sqrt(seps2) * \
277                 sqrt((1/nn) + ((vv[i] - mean(x))**2/sxx));
278         topy[i] = polyval(p,vv[i]) + val;
279         boty[i] = polyval(p,vv[i]) - val;
280 
281 
282 
283 
284     plt.plot(x,y,'bo',ms=5,label='data')
285     plt.plot(vv,p25,'r--', label='HLM - CI')
286     plt.plot(vv,p975,'r--')
287     plt.plot(vv,p50,'r-',label='HLM - Mean')
288     plt.plot(vv,topy, 'k:', label='LR - CI')
289     plt.plot(vv,boty, 'k:')
290     plt.plot(vv,polyval(p,vv), 'k-',label='LR')
291     plt.xlim([-1.5,2.5])
292     plt.ylim([-3,3])
293     plt.title(name[j])
294 
295 
296 plt.show()
```

0.3900000000000	0.1300000000000
0.4300000000000	0.1000000000000
0.2700000000000	1.6200000000000
0.2600000000000	1.7100000000000
0.3600000000000	1.1880000000000
0.2600000000000	1.0800000000000
0.3100000000000	0.8640000000000
0.2700000000000	0.7110000000000
0.3000000000000	0.6120000000000
0.2800000000000	1.1850000000000
0.2700000000000	1.1816500000000
0.7200000000000	0.0300000000000
0.6000000000000	0.1800000000000
0.4400000000000	0.6000000000000
0.4200000000000	0.8000000000000
0.4700000000000	0.4500000000000
0.4800000000000	0.3600000000000
0.3900000000000	0.9000000000000

Table B.1: File *Clm_Vol_coefFriction.txt* with mobility metric for Volcán de Colima

0.2461538461538	4.5000000000000
0.2923076923077	2.6000000000000
0.4054054054054	0.2600000000000
0.4833333333333	0.2900000000000
0.3666666666667	0.1100000000000
0.3103448275862	0.1800000000000
0.3333333333333	0.3500000000000
0.4000000000000	1.1914010000000
0.3000000000000	2.6983110000000
0.4000000000000	0.9004950000000
0.4200000000000	0.7684380000000
0.4400000000000	0.7313370000000
0.4700000000000	0.6395460000000
0.5000000000000	0.5597460000000
0.5300000000000	0.0901080000000
0.5400000000000	0.2258180000000
0.6200000000000	0.0772460000000
0.2794701986755	3.8900000000000
0.2349272349272	4.2100000000000
0.2410423452769	0.6800000000000
0.1598765432099	8.8700000000000
0.3361344537815	0.3400000000000

Table B.2: File *Mrp_Vol_coefFriction.txt* with mobility metric for Mt Merapi

0.29260000000000	6.40000000000000
0.25443478260870	6.30000000000000
0.24660869565217	5.50000000000000
0.32511111111111	3.50000000000000

Table B.3: File *Smr_Vol_coefFriction.txt* with mobility metric for Mt Semeru

0.2805555555556	0.8080000000000
0.3200000000000	0.1720000000000
0.2107142857143	1.4100000000000
0.2107692307692	1.4520000000000
0.4201183431953	0.1225000000000
0.2861111111111	0.1517000000000
0.4382022471910	0.1539000000000
0.4382022471910	0.1179000000000
0.2861111111111	0.2496000000000
0.3896713615023	0.1410000000000
0.4201183431953	0.1731000000000

0.3595041322314	0.1434000000000
0.3966480446927	0.1673000000000
0.3595041322314	0.6345000000000
0.3595041322314	0.2258000000000
0.3896713615023	0.1377000000000
0.2734584450402	0.6107000000000
0.3896713615023	0.1011000000000
0.3966480446927	0.1171000000000
0.3896713615023	0.1094000000000
0.3595041322314	0.1002000000000
0.3966480446927	0.1903000000000
0.3595041322314	0.1466000000000
0.3966480446927	0.1060000000000
0.3966480446927	0.1035000000000
0.3595041322314	0.1356000000000
0.3595041322314	0.1480000000000
0.3966480446927	0.1277000000000
0.3595041322314	0.1236000000000
0.3595041322314	0.1238000000000
0.3502109704641	0.1059000000000
0.3595041322314	0.2611000000000
0.3137254901961	0.1764000000000
0.3595041322314	0.1516000000000
0.4201183431953	0.1004000000000
0.3595041322314	0.2653000000000
0.3595041322314	0.1152000000000
0.3595041322314	0.1227000000000
0.3595041322314	0.3090000000000
0.2734584450402	0.3006000000000
0.3966480446927	0.1147000000000
0.3896713615023	0.2156000000000
0.2734584450402	0.1431000000000
0.4201183431953	0.1322000000000
0.3595041322314	0.1597000000000
0.3966480446927	0.3305000000000
0.3595041322314	0.1313000000000
0.3595041322314	0.2084000000000
0.3966480446927	0.1763000000000
0.3595041322314	0.2155000000000
0.3595041322314	0.1491000000000
0.4201183431953	0.1535000000000
0.3595041322314	0.1772000000000
0.2861111111111	0.1392000000000
0.3595041322314	0.1583000000000

0.3966480446927	0.1395000000000
0.3966480446927	0.1164000000000
0.3595041322314	0.1890000000000
0.3723849372385	0.1482000000000
0.3595041322314	0.2709000000000
0.3502109704641	0.1049000000000
0.2861111111111	0.4298000000000
0.3299319727891	0.2224000000000
0.3595041322314	0.1051000000000
0.3137254901961	0.1217000000000
0.3137254901961	0.2419000000000
0.3137254901961	0.1576000000000
0.3502109704641	0.1001000000000
0.3966480446927	0.1012000000000
0.3966480446927	0.1312000000000
0.2734584450402	0.1139000000000
0.3137254901961	0.1058000000000
0.3125000000000	0.7300000000000

Table B.5: File *Unz_Vol_coefFriction.txt* with mobility metric for Mt Unzen

0.362849210970	0.152000000000
0.308474576271	0.331000000000
0.208718563625	2.600000000000
0.358781810597	0.163000000000
0.211366838891	2.900000000000
0.282152230971	0.375000000000
0.233886320370	0.766000000000
0.143710870803	5.538000000000
0.162103746398	8.750000000000
0.156916254477	13.563000000000
0.158307087476	4.500000000000
0.258020668576	0.500000000000
0.173046378653	1.300000000000

Table B.4: File *Shv_Vol_coefFriction.txt* with mobility metric for Soufrière Hill volcano