

Some ideas for emulators and sampling rare events in the pyroclastic flow problem — a working document

SAMSI granular flow group
Elaine Spiller

March 17, 2007

Our ultimate goal is to draw a probability map around a volcano where contours represent the probability of a flow over some critical value, say 2m, over a substantial length of time, say 100 years. We are specifically interested in mapping regions of very low $P(h_{flow} > h_{crit})$ to determine safe locations for municipal development (e.g. hospitals, schools). Our main tool in this mapping is the TITAN software which is a computer model of pyroclastic flows.

This problem raises some significant statistical challenges

1. We seek to predict the probability of events for which we hope to never have data, i.e. catastrophic events.
2. Data collection for large volume flows is a bit spotty due to obvious safety concerns, and data for frequent yet small volume flows is “pre-binned”. However, Robert and Simon have been making some progress on characterizing a distribution for the flow volume.
3. The TITAN software is moderately expensive (~ 1 hour per sample). This rules out any hope of using TITAN output directly as samples in a Monte-Carlo (MC) simulation of rare events, even if utilizing variance reduction techniques (importance sampling, etc).
4. There are several issues in designing emulators of the TITAN model for this system.
 - a) Zeros – the output from one run will typically have no flow over a significant portion of the land area we wish to map.
 - b) Designing a “spatial” emulator, i.e., one that can interpolate output of interest (max height, say) at untested locations around the volcano is challenging both due to the size of the design problem and the zeros mentioned previously.

1 Proposal for a “simple” emulator/sampling problem: Calculate $P(h_{max} > h_{crit})$ at a single spatial location of interest

First lets consider the model inputs:

1. Initial mass location (stochastic) — East and North coordinates, each uniformly distributed
2. Flow volume (stochastic) — Volume of flow, Pareto distributed with parameters determined from flow data by Robert and Simon.
3. Initial flow velocity (stochastic?) — Starting velocity of flow. Currently taken to be zero.
4. Friction — two sorts (?) one material dependent – friction between interacting grains of material, one friction of material interacting with the ground – depends on rainfall.
5. any others?

First problem

A reasonable goal for our first problem is to find the probability that the maximum flow height, h_{max} , at one specific interesting location, $x_{target} = (E_{target}, N_{target})$, exceeds some critical flow height, h_{crit} . First, let's consider fixed friction coefficients and zero initial velocity (these will be issues to address later). Then we are looking for the *most likely* combination(s) of volumes and initial flow coordinates that generate a flow such that $h_{max}(x_{target}) \geq h_{crit}(x_{target})$.

Naive idea

Ideally, we would sample volumes from Robert and Simon's distribution, varying the initial coordinates uniformly (? or according to some known distribution for a given volcano) and use these values to initialize TITAN runs for a Monte-Carlo (MC) simulation and compute the probability, $P(h_{max}(x_{target}) \geq h_{crit}(x_{target}))$, directly.

Problems with this idea

1. It would take an unreasonable number of TITAN runs to find this probability with decent error bounds, even if the probability was fairly large.
2. Presumably, many TITAN runs in this scenario will give us no flow at the site of interest – obviously a waste of resources.
3. If $P(h_{max}(x_{target}) \geq h_{crit}(x_{target}))$ is small, we may only directly sample a few volumes large enough to get a flow with h_{max} near h_{crit} and those flows may be “wasted” on initial coordinates that, say, send the flow down the other side of the mountain.

Slightly less naive idea: importance sampling an emulator

Emulating the surface that we wish to sample is attractive because (obviously) emulator evaluations are very cheap. Designing the emulator for this use probably won't be trivial. Working on the assumption that we are looking for the smallest volume flow to cause a h_{max} near h_{crit} at x_{target} , we can restrict the domain of the initial flow coordinates to be some interval centered around $\theta = 0$ (consider east and north coordinates as an angle) where $\theta = 0$ “lines up” with the x_{target} . How large to make the range of angles will probably be problem specific (and informed by the geography, river valleys etc). We will also want the design space to cover a range of volumes that give us a range of $h_{max}(x_{target})$ around $h_{crit}(x_{target})$, say $h_{max} \in (\epsilon, 2h_{max}]$. Presumably this will take some tinkering to figure out an appropriate design domain. Once we have a surface

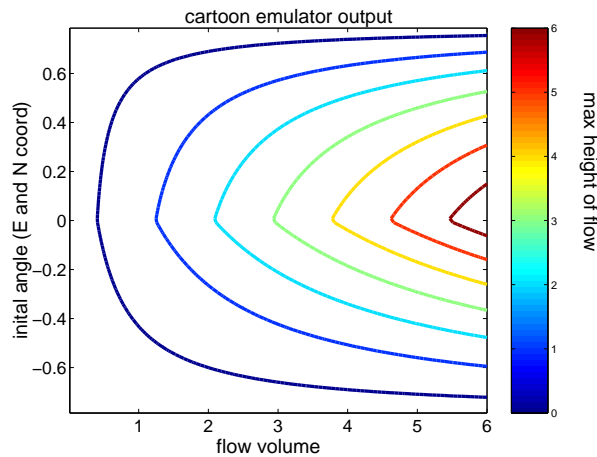


Figure 1: Cartoon of surface we wish to emulate and sample.

established, we can use importance sampling to focus the samples in the region of the design domain.

How to proceed initially?

1. We could work backwards. That is, choose an x_{target} , a range of angles, and a range of volumes, then based on the h_{max} surface generated, choose an h_{crit} . (Maybe some runs already exist that might be suitable to use?) Clearly this is not the way to proceed for a realistic problem, but it seems like easier test of concept problem.
2. It would be nice to have a truth mesh of the initial design domain to get a feel for what this surface looks like. This might illuminate how to proceed somewhat, i.e., give us a better sense of what the design domain should look like, suggest the sort of regression functions to consider.
3. Other thoughts?