University at Buffalo
The State University of New York

Surrogate Models and Uncertainty Quantification for Hazard Map Construction

E. R. Stefanescu*, A. K. Patra*, K. Dalbey*, M. I. Bursik$, M. D. Jones^, E. B. Pitman#, M. Sheridan and E. Calder$

**Department of Mechanical and Aerospace Engineering,*
^Center for Computational Research,
#Department of Mathematics,
$Department of Geology, University at Buffalo, State University of New York
abani@buffalo.edu

Abstract

Computer models of hazardous phenomena, such as floods, hurricanes, and avalanches, are very expensive to run, and each run produces an enormous amount data. For example, a flood model output consists of water depth and velocity at every point in a large grid, at every instant of time. We describe the process of computing a hazard map due to a geophysical flow with uncertain model inputs. These include for instance digital elevation models (DEMs) to represent the terrain. The effect of the terrain on the output of the flow model is investigated by creating realizations of the DEMs using a stochastic method. We also present some effective computational strategies for constructing surrogate models for an ensemble of computer models.

Physics Model: TITAN

$$\frac{\partial h}{\partial t}+\frac{\partial hv_x}{\partial x}+\frac{\partial hv_y}{\partial y}=0$$

$$\frac{\partial hv_x}{\partial t}+\frac{\partial(hv_x^2+0.5\boxed{k_{ap}}g_z h^2)}{\partial x}+\frac{\partial hv_y v_x}{\partial y}=g_x h-\mathrm{sgn}(v_x)\left[g_z+\frac{1}{\kappa_x}v_x^2\right]h\tan(\boxed{\phi_{bed}})-\mathrm{sgn}\left(\frac{\partial v_x}{\partial y}\right)hk_{ap}\frac{\partial hg_z}{\partial y}\sin(\boxed{\phi_{int}})$$

$$\frac{\partial hv_y}{\partial t}+\frac{\partial hv_y v_x}{\partial x}+\frac{\partial(hv_y^2+0.5\boxed{k_{ap}}g_z h^2)}{\partial y}=g_y h-\mathrm{sgn}(v_y)\left[g_z+\frac{1}{\kappa_y}v_y^2\right]h\tan(\boxed{\phi_{bed}})-\mathrm{sgn}\left(\frac{\partial v_y}{\partial x}\right)hk_{ap}\frac{\partial hg_z}{\partial x}\sin(\boxed{\phi_{int}})$$

Modeled as a depth-averaged model of incompressible granular material resulting in a hyperbolic system of equations
- Parallel adaptive solution uses space filling curve based dynamic data management system
- Use Real topography from integrated Geographic Information Systems
- Code is open source and works on many platforms including PCs and large scale HPC platforms - code available from http://www.gmfg.buffalo.edu

DEM Uncertainty Using Two RV

- The Gaussian model is used to represent DEM uncertainty. Model assumes that the total error is the sum of a large number of random, additive effects.

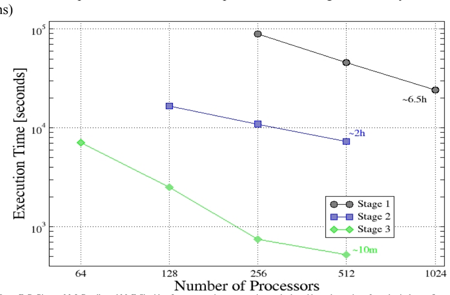
Generation of Ensemble DEMs

1.Error points were obtained as the difference between the 'true' elevation and given DEM dataset.
2.A geostatistical correlogram was employed to show the spatial autocorrelation of error - the correlogram of model was fitted to the error model correlogram by weighted least square estimation.
3.'Extract' the parameters that give the smallest difference between the error model correlogram and the random field.
4.Determine the probability distribution function for the stochastic simulation
5.A total of 64 equally probable potential elevation surfaces of the test area were created using this p.d.f.

$$R(\mathbf{u}) = m(\mathbf{u}) + m(m(T)) + (m(s2(T)).e)\, Z(\mathbf{u})$$

Workflow Strategy

Performance speedup of three stages of the hazard map workflow: Stage 1 is generation of direct simulation inputs, Stage 2 is emulator construction, and Stage 3 is emulator evaluation (only Stage 3 needs to be redone to produce a new hazard map based on the range covered by the initial direct simulations)



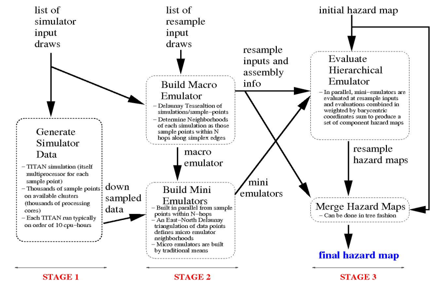
References
- K. Dalbey, A. K. Patra, E. B. Pitman, M. I. Bursik, and M. F. Sheridan. Input uncertainty propagation methods and hazard mapping of geophysical mass flows. J. Geophys. Res, 113, 2008..
- M. J. Bayarri, J. O. Berger, E. S. Calder, K. Dalbey, S. Lunagomez, A. K. Patra, E. B. Pitman, E. T. Spiller, and R. L. Wolpert. Using statistical and computer models to quantify volcanic hazards. submitted to Technometrics Jan. 25, 2008.
- A. K. Patra, A. C. Bauer, C. C. Nichita, E. B. Pitman, M. F. Sheridan, M. I. Bursik, B. Rupp, A. Webber, A. J. Stinton, L. M. Namikawa, and C. S. Renschler. Parallel adaptive numerical simulation of dry avalanches over natural terrain. Journal of Volcanology and Geothermal Research, 139:1–21, 2005.
- C.R. Ehlschlaeger, A.M. Shortridge and M.F. Goodchild. Visualizing spatial data uncertainty using animation. Computers & Geoscience 23(4), pp.387-395, 1997
- E.R Stefanescu, M.Bursik, K.Dalbey, M.Jones, A.K.Patra and E.B.Pitman. DEM uncertainty and hazard analysis using a geophysical flow model, 2010 International Congress on Environmental Modelling and Software. Proceedings, Ottawa, 2010.

Approach

Stage 1: Evaluate an ensemble of several hundred to several thousand multiprocessor landslide simulations, dynamically assigning simulations to processors as they become available to continually **use the entire pool** of processors efficiently.

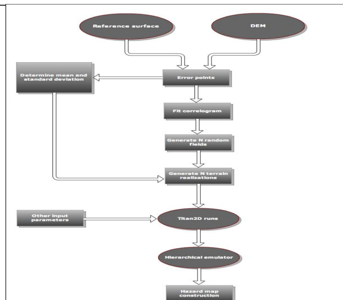
Stage 2: Create a multi-level hierarchical emulator (a statistical model) from the output of the ensemble of simulations. Its **hierarchical** nature allows the emulator's components to be constructed (and evaluated) **concurrently.** Emulator acts as a fast surrogate of the simulator.

Stage 3: Use the emulator through importance sampled Monte Carlo to compute **a map of the probability that a hazard criterion will be met at hundreds of thousands (or more) of locations.**



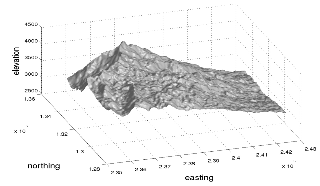
Input Uncertainty

- In a model, uncertainty is a measure of the lack of knowledge about input data, model and inherent variability.
- Inputs to TITAN2D are **initial volume, pile aspect ratio, initiation location, basal and internal friction angles and Digital Elevation Model (DEM)** – 8 random variables (RV)
- **Characterization of uncertainty in each of these and propagation through TITAN can be used to estimate hazard.**
- Many approaches to propagating uncertainty in such PDE based models such as Monte Carlo, Polynomial Chaos, Response surface.
- Sampling driven **by strategy to minimize cost followed by** computation of statistics
- Naive approaches to DEM uncertainty can lead to the use of thousands if not millions of random variables.
- Need a method that yields O(10) random variables – see next for 2 RV process.

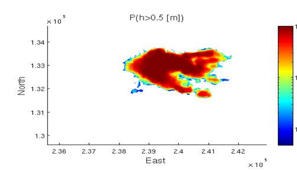


Case Study: Mammoth Mountain, CA

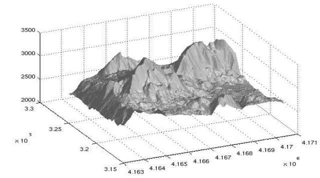
- Hazard maps at Mammoth mountain, CA, using 4 and 8 random variables as input.



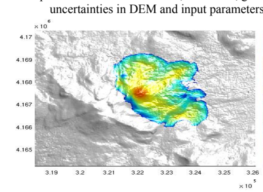
Aster DEM in easting, northing and elevation coordinates; Galeras Volcano, Colombia



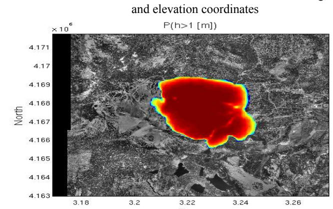
Probability that a flow will exceed 0.5 m in depth as a function of position on Galeras Volcano, Columbia, given the uncertainties in DEM and input parameters

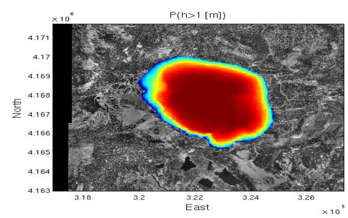


The Mammoth TOPSAR 30m DEM terrain surface in easting, northing and elevation coordinates



Sample simulation result on flow at Mammoth





Hazard maps for Mammoth Mountain computed using 64 multi-processor TITAN simulations and 10^{15} resamples of hierarchical emulator. A) Input random variables are volume, basal and internal friction angles and DEM (using one RV) B) Input random variables are volume, initial pile aspect ratio, starting location, basal and internal friction angles and DEM (2 random variables.)