Hazard analysis using large scale computer simulations and combined physical/statistical models

E.R. Stefanescu[1], A.K. Patra [1], E.B. Pitman [1], M. Bursik [1], M. Jones [1], K. Dalbey [2]

[1]Geophysical Mass Flow Group
University at Buffalo
[2]Sandia National Laboratories

Outline

Geophysical flows

Debris flows

Pyroclastic flows

Questions when studying hazardous natural flows

- *Is a particular village likely to be affected by flows in a given time frame?*
- *Should we construct a road along this valley or along a different one?*
- *Should a village be evacuated and villagers moved to a different location?*

Hazard maps

- “A volcanic hazard refers to the probability that a given area will be affected by a potentially destructive volcanic process" – d’Albe, 1979
- Hazard maps portray the impact of harmful future volcanic events

Some problems

- We have to deal with complex physics
- Poorly characterized observation data, poorly known extents of units, and imprecise volumes of deposits
- Uncertainties in flow characteristics and model parameters complicate the construction of accurate hazard maps
- Changes in the topography represent a major uncertainty for future flows

The geophysical mass flow model TITAN2D

1. Depth-averaged model of incompressible granular material
2. Combines numerical simulations of the flow with digital data of the natural terrain
3. The code runs in parallel –using MPI (message passing interface)
4. The algorithm uses local adaptive mesh refinement and dynamic load balancing for parallel processing

$$
\begin{aligned}
\textit{Mass conservation}: \quad & \frac{\partial h}{\partial t} + \frac{\partial (V_x \cdot h)}{\partial x} + \frac{\partial (V_y \cdot h)}{\partial y} = 0 \qquad (1) \\
\textit{Momentum conservation}: \quad & \frac{\partial h V_x}{\partial t} + \frac{\partial (V_x \cdot h V_x + 0.5 k_{ap} g_z h^2)}{\partial x} + \frac{\partial (V_y \cdot h V_x)}{\partial y} = S_x \\
& \frac{\partial h V_y}{\partial t} + \frac{\partial (V_x \cdot h V_y)}{\partial x} + \frac{\partial (V_y \cdot h V_y + 0.5 k_{ap} g_z h^2)}{\partial y} = S_y
\end{aligned}
$$

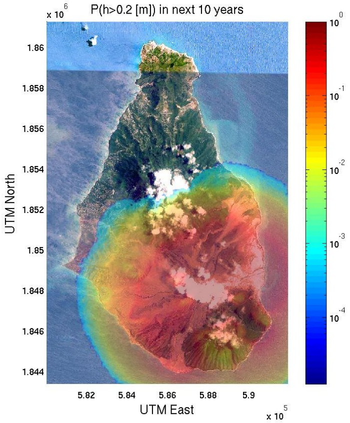
The geophysical mass flow model TITAN2D

Inputs to TITAN2D are initial volume, initiation location, basal and internal friction angles and DEM

$$k_{ap} = 2\frac{1 \mp \sqrt{1 - cos^2(\phi_{int})(1+\tan^2(\phi_{bed}))}}{\cos^2} - 1 \tag{2}$$

$$S_x = g_x h - \frac{V_x}{\sqrt{V_x^2 + V_y^2}} \max\left(g_z + \frac{V_x^2}{r_x}, 0\right) h \tan((\phi_{bed}) - \\ -sgn\left(\frac{\partial V_x}{\partial y}\right) h k_{ap} \frac{\partial(g_x h)}{\partial y} \sin((\phi_{int})$$

Input uncertainty



Computational issues

- To obtain three-digit accuracy in the expected value would require $O(10^6)$ simulations
- One million 20 minute simulations running non-stop in parallel on 64 processors will take 217 days to complete

Hazard map construction

- Straightforward way to account for uncertain inputs and stochastic forcing is a Monte Carlo approach
- A simulator with uncertain inputs is a stochastic process
- An emulator is a statistical model of a stochastic process that can be built from multiple sources of different fidelity data
- To construct a **Ga**usian **S**tochastic **P**rocess (GASP) emulator, the covariance structure of the Gaussian must be assumed and parameters determined by Bayesian methodology
- A Bayes Linear Model (BLM) emulator is a least square fit plus a GASP error model that maps inputs to outputs and interpolates the "simulated" data (Dalbey et. al. '08,'09')

Hierarchical Emulator

$$s(\underline{x}) = \mu(\underline{x}) + \epsilon(\underline{x}) \quad (3)$$

$$Cov(\hat{\underline{\epsilon}}(\underline{x}), \hat{\underline{\epsilon}}(\underline{y_i})) = \sigma^2 r_i(\underline{x}) \quad (4)$$

$$r_i(\underline{x}) = r(\underline{x} - \underline{y_i}) = exp(-\textstyle\sum_{i_{in}}^{N_{in}} \partial_{i_{in}} (x_{i_{in}} - y_{i,i_{in}})^p) \quad (5)$$

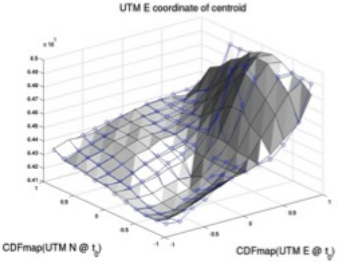
$$R_{i,j} = r_i(\underline{y_j}) = r_j(\underline{y_i}) \quad (6)$$

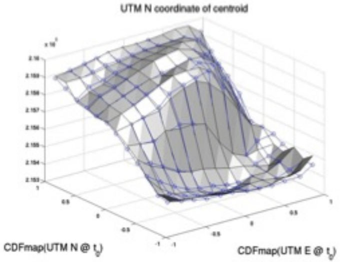
s - represents the simulator output, $\underline{x}$ is an arbitrary input

$$E(s_{BLM}(\underline{x})|\underline{s_y}) = \underline{g}(\underline{x})^T \underline{\beta} + \underline{r}(\underline{x})^T \underline{\underline{R}}^{-1} \underline{\epsilon} \quad (7)$$

$$Var(s_{BLM}(\underline{x})|\underline{s_y}) = \sigma^2 (1 - \underline{r}(\underline{x})^T \underline{\underline{R}}^{-1} \underline{r}(\underline{x})) \quad (8)$$

g - are the least squares basis functions, β are their coefficients, $\epsilon(x)$ is "Gaussian" model of the error, p=2,1 (absolute val.)





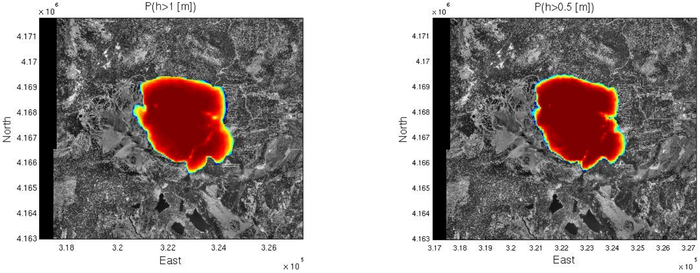
Strategies for R^{-1}

- Some solutions come from geostatistics – *moving window* (*local neighborhood kriging*) (Davis '84)
- *Projected process kriging* - the covariance matrix is updated with new observations while the rank remains unchanged (Ingram '07)
- Replace single global emulator by ensemble of m emulators of size $n \longrightarrow$ cost reduces form $O(N^3)$ to $O(mn^3)$ – assemble local emulators into global emulator using partition of unity
- Use a properly localized covariance function (Wood'95, Gneiting '97,'99)
- Wood defined the sufficient conditions to allow truncation of valid covariance function defined on $(-\infty, \infty)$ to $[-K, K], K \geq 0$ (Gaussian form (p=2) cannot be truncated!)
- Need to maintain the positivity of Fourier coefficients of chosen function

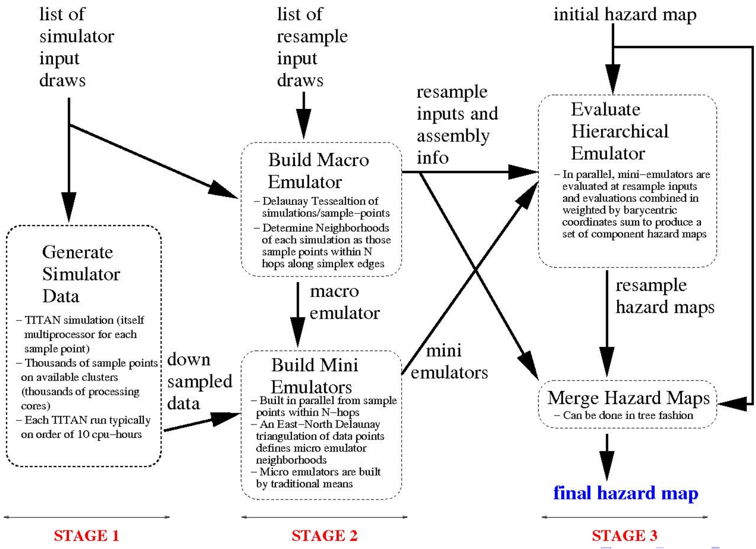
Strategies for R^{-1}

- Gneiting '97, '99 introduced truncated power law function – "virtually identical to the Gaussian" – we use "this" and also examine p=1 which meets the Wood criteria for truncation
- Can use Moore-Penrose inverse (if needed) for local emulator construction

Hazard map superimposed on an aerial photo of Mammoth Mountain



Hierarchical Emulator



DEM uncertainty- spatial autocorrelation

- Certain types of terrain are more suited to the creation of accurate DEMs
- DEM uncertainty is correlated to the feature of the terrain

- The spatial variability and spatial autocorrelation were identified as significant feature of DEMs (Hunter and Goodchild '97)
- There is strong relationship between elevation error and terrain, and that is influenced by the spatial autocorrelation of the error (as the difference of two realizations) (Wechsler et. al. '06, Hebeler and Purves '08)

The error model

The stochastic model

- The Gaussian model is commonly used as an error model (Ehlschlaeger and Shortridge '96)
- The model assumes that the total error is the sum of a large number of random, additive effects

$$R(\mathbf{u}) = m(\mathbf{u}) + m(m(T)) + (m(s^2(T)) \cdot \epsilon) \cdot Z(\mathbf{u}) \tag{9}$$

where R(**u**) is a possible DEM and Z(**u**) is a random field which captures the autocorrelative effect between points

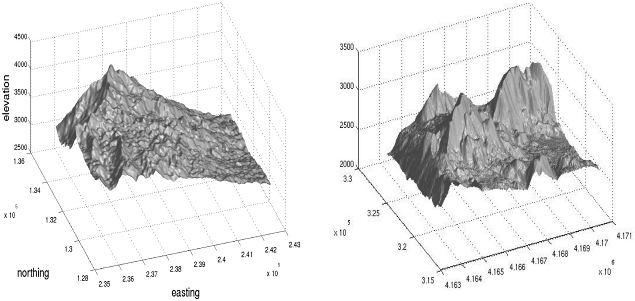
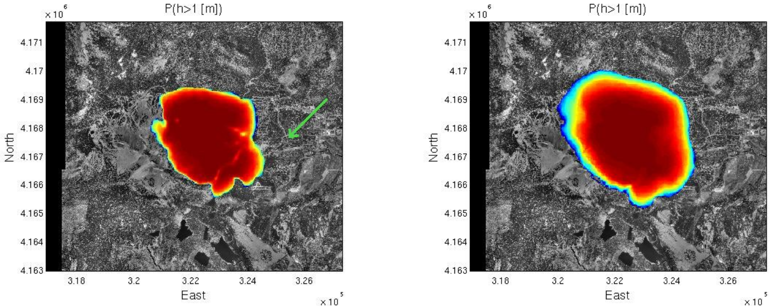


Figure: (a) Galeras Volcano (b) Mammoth Mountain

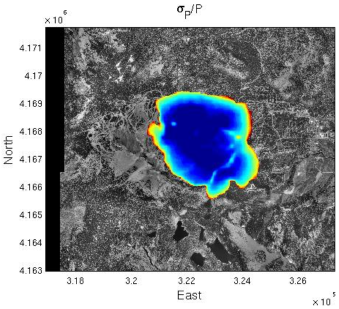
Hazard map



(a) (b)

Figure: Probability that a flow will exceed 1 m in depth as a function of position on Mammoth Mountain, CA, given the uncertainties in DEM and input parameters (a) 4 uncertain parameters (east and north location, basal friction, height)– the arrow indicates the center of the town (b) 8 uncertain parameters (including DEM)

Hazard map



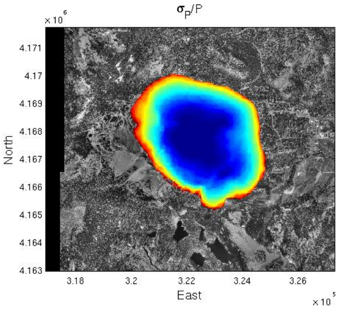


Figure: Standard deviation in the estimate that the flow will exceed 1 m in depth. Estimation error is concentrated at flow margins. (a) 4 uncertain parameters (b) 8 uncertain parameters

Conclusions

- Developed effective computation methods for conducting hazard analysis using modest ensemble of simulations
- Hierarchical emulators based on approximate localized covariance
- DEM representation of the terrain has errors which are uncertain. These errors impact the flow modeling and hence the hazard computation
- Uncertainties in the DEM can be quantified and the impact on hazard analysis can be studied
- This work was supported by NASA grant NNX08AF75G.