Feature Based Error model construction for Digital Elevation Models and Hazard Analysis

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

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

Outline

Geophysical flows

Debris flows

Pyroclastic 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

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

The geophysical mass flow model TITAN2D

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

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

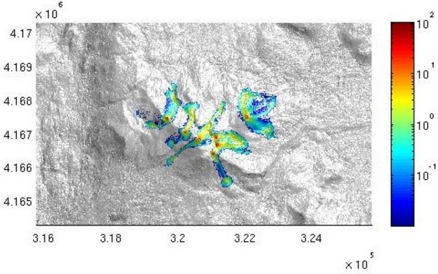
$$S_x = g_x h - \frac{V_x}{\sqrt{V_x^2 + V_y^2}} \max\left(g_z + \frac{V_x^2}{r_x}, 0\right) h \tan((\phi_{bed}) -$$

$$-sgn\left(\frac{\partial V_x}{\partial y}\right) h k_{ap} \frac{\partial(g_x h)}{\partial y} \sin((\phi_{int})$$

Volume - Pareto distribution (Bayari et.al. ’09), initial location, basal and internal friction - Uniform distribution (Dalbey et.al. ’08)

Previous work

For moderate and smaller scale flows it is important to capture the terrain's features in order to get an accurate footprint of the flow (Stefanescu et.al '12).



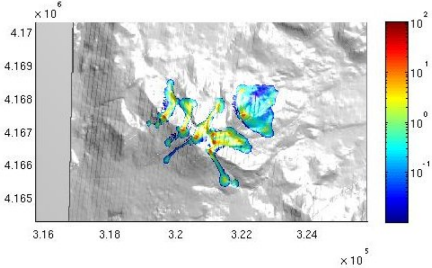


Figure: Fow maps for Mammoth Mountain (a) Topsar 5m (b) NED 30m

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

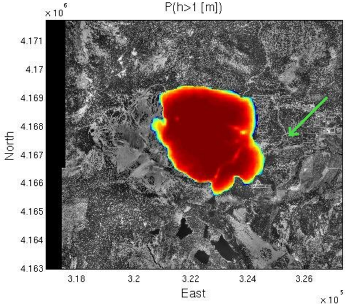
$$R(\mathbf{u}) = m(\mathbf{u}) + m(m(T)) + (m(s^2(T)) \cdot \epsilon) \cdot Z(\mathbf{u}) \tag{2}$$

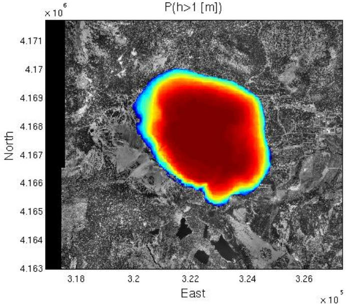
where R(**u**) is a possible DEM and Z(**u**) is a random field which captures the autocorrelative effect between points

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')

Hazard map

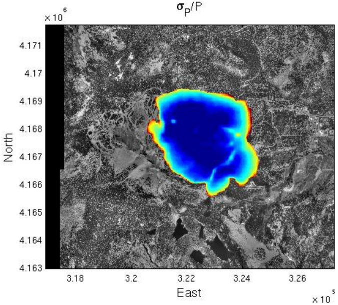




(a) (b)

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



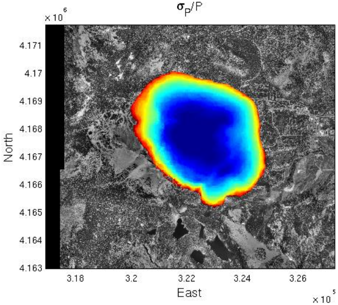
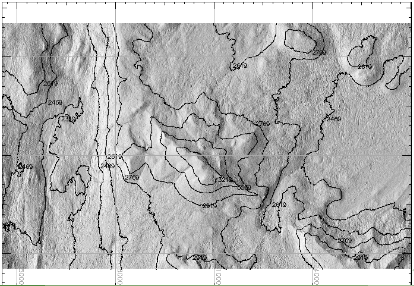


Figure: Standard deviation in the estimate that the flow will exceed 1 m in depth. Estimation error is concentrated at flow margins. (a) 4 uncertain parameters (b) 8 uncertain parameters

DEM segmentation

- We consider the problem of DEM segmentation in homogeneous regions, aiming for identification of plateaux, ridges, small drainages, straight front slopes, valleys, and crests.
- It is required / needed when we want to construct a sparse representation of the DEM.
- We define a *feature matrix* of DEM attributes, consisting of elevation and first and second derivatives of elevation (slope, profile curvature and tangential curvature).



Spectral Clustering

Weighted Graph Representation

Given a DEM a weighted graph is set up **G(V,E)**, where each pixel of the DEM is a node in the graph G and the links between the adjacent elevation points form the edges of the graph (Ng et. al. '01, von Luxburg '06

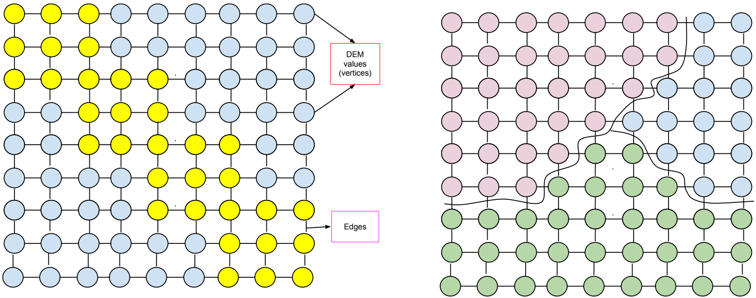


Figure: Graph based representation of DEM values

Graph Laplacian

The Affinity Matrix

$$D_{ij} = \begin{cases} \exp \frac{-\|F(i)-F(j)\|}{\sigma_F^2} * exp \frac{-\|x(i)-x(j)\|}{\sigma_x^2}, & if \parallel x(i) - x(j) \parallel \leq r \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

The graph Laplacian matrix is defined as:

$$L = D - W, \tag{4}$$

where W is the weighted adjacency matrix. L satisfies the following properties:

- For every $f \in \mathcal{R}^n$ we have :

$$f' L f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} (f_i - f_j)^2. \tag{5}$$

- L is symmetric and positive semi-definite.

Spectral Clustering

Results

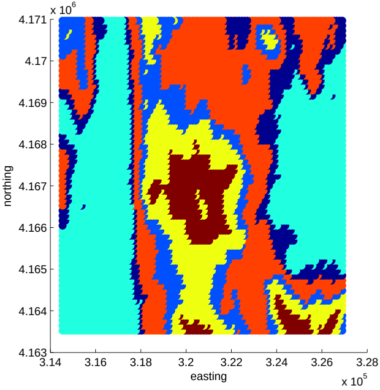


Figure: a) $\sigma_F = 0.8$ and $\sigma_x = 0.8$ b) $\sigma_F = 0.8$ and $\sigma_x = 1$

Clustering with Gaussian Mixtures

To define a Gaussian mixture model with $K > 1$ components in $\mathcal{R}^D$ for $D \geq 1$, let x_n be the observation of the nth data point of a DEM. The density function $p(x_n)$ is given by:

$$p(x_n) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \tag{6}$$

where $\pi_1, \pi_2 \cdots \pi_k$ are the mixing coefficients and μ_k, Σ_k are the Gaussian distribution's parameters for each kth component. The mixing coefficients satisfy the following conditions:

$$0 \leq \pi_k \leq 1 \ , \ \sum_{k=1}^{K} \pi_k = 1 \tag{7}$$

For the Gaussian mixtures, each component density is a Gaussian probability with μ_k and covariance Σ_k:

$$p(x_n) = \mathcal{N}(x_n|\mu_k, \Sigma_k) = \frac{1}{2\pi^{\frac{D}{2}}} \frac{1}{\det(\Sigma_k)^{\frac{1}{2}}} \exp\{-\frac{(x_n - \mu_k)^T (x_n - \mu_k)}{2\Sigma_k}\} \tag{8}$$

Clustering with Gaussian Mixtures

The inherent steps of the EM for the Gaussian mixture approach can be summarized as:

- Initialize the means μ_k, covariance Σ_k, mixing coefficients π_k and the log likelihood.
- **E step** evaluate the responsibilities using the current parameters values:

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} \tag{9}$$

- **M step** Re-estimate the parameters using the current responsibilities:

$$\mu_k^{new} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_{nk}) x_n \tag{10}$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_{nk})(x_n - \mu_k^{new})(x_n - \mu_k^n ew)^T \tag{11}$$

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} \gamma(z_{nk}) \tag{12}$$

Clustering with Gaussian Mixtures

Results

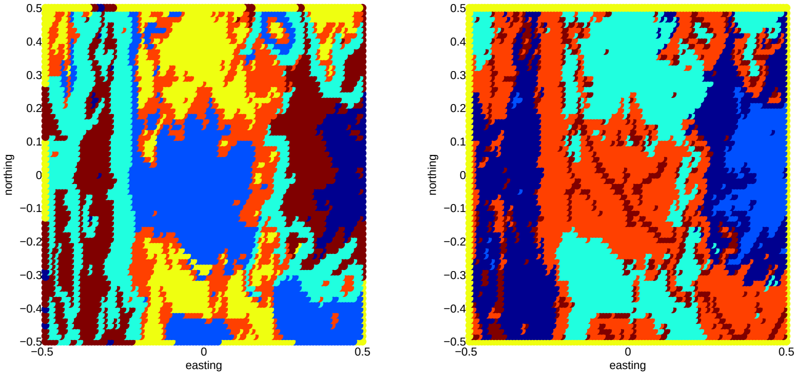


Figure: a) K=6 using 2 features (elevation and slope) b) K=6 using 4 feature (elevation, slope, transversal, and profile curvature)

Future work

Adaptive Spatial Gaussian Mixture Model

Propose a way of incorporating spatial relationships. Therefore, in the calculation of the density function, the data point x_n will be influenced by its neighbors

$$p(x_n) = \frac{1}{2\pi^{\frac{D}{2}}} \frac{1}{\det(\Sigma_k)^{\frac{1}{2}}} \times [\exp\{-[\frac{\eta_n^k (x_n - \mu_k)^T (x_n - \mu_k)}{2\Sigma_k} + \frac{\eta_n^k}{8} \sum_{X_l \in X_{x_n}} \frac{(x_n - \mu_k)^T (x_n - \mu_k)}{2\Sigma_k}]\}] \tag{14}$$

where η_n is the parameter that controls the neighbors influence and X_{x_n} is the subset of neighborhood data points of x_n in a 3×3 window. η_n is calculated using the following formula:

Future work

Adaptive Spatial Gaussian Mixture Model

$$\eta_i^k = df_{std}^k(i)/x_{std}(i) \tag{15}$$

where

$$df_{std}^k(i) = (\frac{1}{9}[\sum_{X_l \in X_{x_n}} \{(df_{x_n}^k - \mu)^2\} + (df_{x_i}^k - \mu)^2])^{1/2} \tag{16}$$

μ is the mean value of *df* in the 3×3 window:

$$df_{x_n}^k = (\frac{(x_n - \mu_k)^T (x_n - \mu_k)}{2\Sigma_k} \tag{17}$$

In order to eliminate the unbalanced effect on the weighting functions between smooth and sharp edges, the *df* is divided by the standard deviation of all the data points in the 3×3 window

$$x_{std}(i) = \{(\frac{1}{9}\sum_{X_l \in X_{x_n}} (x_l - \hat{x})^2 + (x_n - \hat{x})^2)\}^{1/2} \tag{18}$$

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.

E.R.Stefanescu, M.Bursik, G.Cordoba, K.Dalbey, M.D.Jones, A.K.Patra, E.B.Pitman and M.F. Sheridan, Digital elevation model uncertainty and hazard analysis using a geophisical flow model, Proc.R.Soc. A, 2012.