

₁ Digital Elevation Model (DEM) clustering for terrain
₂ modeling

₃ E. R. Stefanescu¹, A.K. Patra¹, and M. Bursik²

₄ ¹Department of Mechanical and Aerospace Engineering, University at
₅ Buffalo, Buffalo, NY 14260

₆ ²Department of Geology, University at Buffalo, Buffalo, NY 14260

₇ May 6, 2013

₈ **Abstract**

₉ We consider the problem of Digital Elevation Models (DEMs) segmentation in
₁₀ homogeneous regions, aiming for identification of plateaux, ridges, small drainages,
₁₁ straight front slopes, valleys, and crests, in order to create ensembles of DEMs. These
₁₂ are then used in a systematic hazard analysis, resulting in a complete and complex haz-
₁₃ ard maps. In the paper we explore and implement a method for segmentation using
₁₄ clustering, that is required / needed when we want to construct a sparse representation
₁₅ of the DEM. The method – spectral clustering, is extensively and successfully used in
₁₆ image segmentation. It is a complex method that accounts for the spatial correlation
₁₇ of the elevation points and has the advantage that it can be used for almost any ap-
₁₈ plication where relationships between topographic features and other components of

19 landscapes are to be assessed. Here, the method is adapted for the case in which each
20 data point has associated range of geomorphometric measures.

21 1 Introduction

22 Information about topography is necessary for landscape evaluation, erosion studies, hydrology
23 and geophysical modeling, natural hazard prevention, etc. The classic way to incorporate
24 relief units into a landscape assessment is to delineate them during field survey or using stereo
25 aerial photographs. This approach is relatively time-consuming and the results depend on
26 the subjective decision of the interpreter. Several methods for the creation of landform el-
27 ements using elevation-derived attributes are described in the literature. Commonly, these
28 techniques developed regions of homogeneity based on common attributes and then classified
29 those regions (or groups of regions) as elements. The most widely used techniques are: self
30 organizing map [Kohn et al., 1995], watershed segmentation [Najman and Schmitt, 1996],
31 support vector machine [Gunn, 1997], segmentation using heuristic rules and fuzzy logic
32 [Sonka et al., 1999], fuzzy K -means classification [Burrough et al., 2000] and object-based
33 image analysis [Carleer et al., 2005]. Many of these techniques have drawbacks, especially
34 when the method relies heavily on hydrological information and requires data-specific knowl-
35 edge; also the methods do not incorporate autocorrelation between the same attribute at
36 two locations.

37 Digital Elevation Models (DEMs) are digital representations of terrain, and consist of
38 an array of squared cells (data points/ pixels) with an elevation associated with each data
39 point. They can have different resolutions (5m, 30m, 90m, 120m, etc.) and can be obtained
40 from various methods (photogrammetry, radar interferometry, laser altimetry, etc.). Usually
41 the size of a DEM to be used in a particular application varies from tens to hundreds of

42 kilometers which can lead to thousands to millions of grid points.

43 The aim of this paper is to quantify the variation in the output of a computational
44 flow model for block and ash flows, when the model inputs, including the elevation values
45 represented in the DEM, are uncertain or given as a range of possible values. Integrating these
46 variations in the possible flows as a function of input uncertainties provides well-defined data
47 on the probability of hazard at specific locations, i.e., a hazard map [Dalbey et al., 2008]. In
48 particular, the focus here is on assessing the influence of DEM uncertainties and proposes an
49 improved method of generating ensembles of DEMs for probabilistic pyroclastic flow hazards
50 analysis as well as other applications in which DEM uncertainty could play a critical role in
51 decision making.

52 The particular problem that is addressed in the present contribution is that of the varia-
53 tion in data quality and error structure in different regions on a given DEM due to features
54 within the topography itself, or of sensors and their interaction with the topography. For ex-
55 ample, radar interferometry has difficulty in properly capturing topography in steep regions,
56 and especially in regions from which radar return is insignificant, such as slopes facing away
57 from the sensor. In these regions, errors are relatively high compared, e.g., to flat regions
58 of diffuse reflectance. In this context, we seek to implement a segmentation of a DEM of
59 Mammoth Mountain to create non-overlapping groupings of homogeneous regions. These
60 homogeneous regions consist not only of contiguous areas with similar slope or curvature,
61 but more precisely any noncontiguous areas having features that should result in their having
62 the same error structure.

63 Mammoth Mountain (Fig. 1) was chosen for this study due to the existence of appropriate
64 data sets and geologic setting. Mammoth Mountain is a large, geologically young, composite
65 dome volcano located on the southwestern rim of Long Valley Caldera, California [Bailey,
66 1989]. There are many active geophysical flow hazards occurring on Mammoth Mountain,

including snow avalanches, rock avalanches and debris flows. In addition, it is intersected by the Mono-Inyo Craters volcanic chain, which is the most active volcanic region in the southwestern U.S. If Mono-Inyo type activity occurs on Mammoth Mountain, then domes may form. These new domes would be growing atop a steep edifice, and therefore could become gravitationally unstable. Given that block and ash flows occurred at Mammoth Mountain during its older dome growth stage, there is reason to believe that renewed dome formation would result in block and ash flow activity. If this is so, then parts of Mammoth Lakes, CA, are at risk from block and ash flows. Our previous work on Mammoth Mountain (Stefanescu et al., 2012xx) was the testing of the hypothesis that different DEMs result in different model outputs of block and ash flow inundation. The present contribution thus in part represents an exploration of a refinement of the methodology.

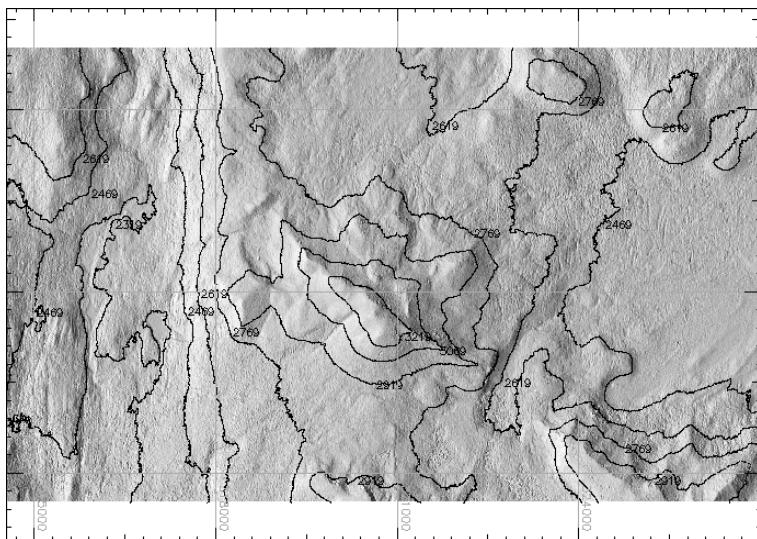


Figure 1: Hillshade plot of the Mammoth Mountain

78 **2 Methodology**

79 Segmentation is a broad term, covering a wide variety of problems and techniques. Segmen-
80 tation methods are based on characterizing the similarity between some data point or region
81 in relation to its local neighborhood, measured either spatially or categorically. A variety of
82 different methods have been proposed for image segmentation such as boundary-based seg-
83 mentation, region-based segmentation and pixel-labeling. One view of segmentation is that
84 we are trying to determine which components of the data set naturally “belong together”.
85 This is a problem known as *clustering*. Clustering is difficult to implement in practice for a
86 number of reasons. Real-life data may contain clusters of varying size and shape, whose num-
87 ber is unknown in advance. Noise and outliers can further complicate the task by connecting
88 separate clusters. Spectral clustering was first developed in the context of graph partitioning
89 problems [Donath and Hofmann, 1973], where the problem is to partition a weighted graph
90 into disjoint pieces, minimizing the sum of the weights of the edges linking the disjoint pieces.
91 In the graph the nodes represent the grid points and arcs represent affinities (“couplings”)
92 between nearby grid points. The final cluster assignments of the dataset can be achieved by
93 optimizing some clustering criteria defined on the graph. The criteria of some spectral clus-
94 tering methods are to optimize some cut value on an undirected graph, such as a normalized
95 cut [Shi and Malik, 2000], ratio cut [Hagen and Kahng, 1992], or min–max cut [Ding et al.,
96 2001].

97 To be able to perform the segmentation of the DEM in homogeneous regions we need to
98 specify a range of geomorphometric measures which can be extracted from the surface. We
99 define a *feature matrix* of DEM attributes, consisting of elevation and first and second deriva-
100 tives of elevation (slope, profile curvature and tangential curvature). Slope and curvature
101 are easily extracted from a DEM within a Geographical Information System (GIS).

102 In the next sections a basic methodology for generating ensembles of DEMs using segmen-

tation is presented, with emphasis on segmentation using spectral clustering. Subsequent sections summarize the TITAN2D flow simulation tool and its use in a systematic hazard analysis. The hazard analysis tool itself uses ensembles of TITAN2D simulations to construct statistical surrogate models of flow outcomes at different locations as a function of model inputs, such as flow volume and initial location. Sampling of these surrogates leads to the construction of effective hazard maps that reflect the range of uncertainty in the model inputs.

2.1 Spectral Clustering

A digital representation of a terrain surface is an approximation of reality and is often subject to significant error. The error is usually not known in terms of both magnitude and spatial distribution. There are, in fact large uncertainties associated with the construction of DEMs. DEM vendors generally provide users with a measure of vertical accuracy in the form of the root mean squared error (RMSE) statistic.

However, one key feature of DEM grid points, which are spatial data, is the autocorrelation of observations in space. Generally, spatial autocorrelation refers to the correlation between the same attribute at two locations. Observations in close spatial proximity tend to be more related than observations at larger distances or separation. Based on this assumption spectral clustering is performed to identify homogenous regions within a DEM.

Compared with traditional clustering algorithms, spectral clustering has some advantages: it can stably detect non-convex patterns and linearly non-separable clusters [Sakai and Imiya, 2009], and can obtain the globally optimal solutions in a continuous domain by eigendecomposition [Archip et al., 2005]. As a discriminative approach, they do not make assumptions about the global structure of data. Instead, local evidence on how likely two data points belong to the same class is first collected and a global decision is then made to

127 divide all data points into disjunct sets according to some criterion. Often, such a criterion
128 can be interpreted in an embedding framework, where the grouping relationships among data
129 points are preserved as much as possible in a lower-dimensional representation. What makes
130 spectral methods appealing is that their global-optima in the relaxed continuous domain
131 are obtained by eigendecomposition. However, to get a discrete solution from eigenvectors
132 often requires solving another clustering problem, albeit in a lower-dimensional space. That
133 is, eigenvectors are treated as geometrical coordinates of a point set. Unfortunately, when
134 the number of grid points (denoted as n) is large, spectral clustering encounters a quadratic
135 resource bottleneck in computing pairwise similarity between n nodes and storing that large
136 matrix. Moreover, the algorithm requires considerable computational time to find the small-
137 est k eigenvalues of a Laplacian matrix. Eigenvalues and eigenvectors are at the heart of
138 spectral clustering algorithms, and in spite of their importance, existing eigensolvers do not
139 scale well. Fast computation schemes for spectral clustering have been proposed by differ-
140 ent authors. They focus on the eigenvector computation of a graph Laplacian defined by
141 a matrix of data similarities. The Krylov subspace methods [Mahadevan, 2008], are itera-
142 tive algorithms for finding leading eigencomponents of a sparse matrix, while Dhillon et al.
143 [2007] assume the availability of the similarity matrix and propose a method that does not
144 use eigenvectors. Fowlkes et al. [2004] propose using the Nyström approximation to avoid
145 calculating the whole similarity matrix. In this paper we are using a method proposed by
146 Song et al. [2008], which parallelize spectral clustering on distributed computers to address
147 resource bottlenecks of both memory use and computation time.

148 **2.1.1 Approach**

For a given data set $P = \{p_1, \dots, p_n \in R^d\}$, spectral clustering finds a set of data clusters,
 $\{C_1, \dots, C_k \subset P\}$, on the basis of spectral analysis of a similarity graph. Spectral clustering

builds a weighted graph $G(V, E)$, where V represents vertices and E , edges. We represent each elevation point as a node in the graph G and the links between the adjacent data points will form the edges of the graph. Spectral clustering partitions data points into groups such that the members of a group are similar to each other and dissimilar to data points outside of the group. Given data points, an affinity matrix can be represented by a weighted adjacency matrix $W \in \mathcal{R}^{n \times n}$, where w_{ij} is a measure of the similarity between grid point i and grid point j . The affinity matrix is used to preserve the local structure of the patterns. It expresses the degree of similarity between points, and it must have the following properties: i) non-negative; ii) symmetric; iii) invertible. We have chosen the heat kernel for calculating the affinity matrix, as:

$$\mathbf{W}_{ij} = \begin{cases} \exp \frac{-\|F(i)-F(j)\|}{\sigma_F^2} * \exp \frac{-\|x(i)-x(j)\|}{\sigma_x^2}, & \text{if } \|x(i) - x(j)\| \leq r \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where $F(i)$ represents the DEM feature vector for node i , and $x(i)$ represents the coordinate location of i^{th} node. σ_F and σ_x are positive tuning parameter that controls the decay of the affinity [Tung et al., 2010]. The graph partitioning can be interpreted as a minimization problem of an objective function. Common objective functions are the ratio cut (Rcut), normalized cut (Ncut) and min–max cut (Mcut) expressed as:

$$Rcut(C_1, \dots, C_k) = \sum_{l=1}^k \frac{(C_l, P \setminus C_l)}{\text{card } C_l} \quad (2)$$

$$Ncut(C_1, \dots, C_k) = \sum_{l=1}^k \frac{(C_l, P \setminus C_l)}{\text{cut}(C_l, P)} \quad (3)$$

and

$$Mcut(C_1, \dots, C_k) = \sum_{l=1}^k \frac{(C_l, P \setminus C_l)}{\text{cut}(C_l, C_l)} \quad (4)$$

Here, $\text{cut}(X, Y)$ is the sum of the edge weights between $\forall p \in X$ and $\forall p \in Y$, $P \setminus C_l$ is the complement of $C_l \subset P$, and $\text{card } C_l$ denotes the number of points in C_l . The degree d_i of

node i is the sum of all edge weights incident on x_i :

$$d_i = \sum_{j=1}^n w_{ij} \quad (5)$$

Let h_l be an n -dimensional vector indicating the members of the cluster C_l by its binary components. The minimization problem of any objective function in Eqs. ??, ?? and ?? can be rewritten as a trace minimization problem under a constraint on a matrix $H = [h_1 \dots h_k]$:

$$\min_H \text{tr}(H^\top LH) \text{ subject to } H^\top LH = I. \quad (6)$$

The matrix N is equal to I , D or W for Rcut, Ncut or Mcut, respectively [Shi and Malik, 2000, Ng et al., 2002]. The spectral clustering algorithms were derived from the minimization problem in Eq. 6 by relaxing the binary constraint on h_l . The relaxed trace minimization for $H \in \mathcal{R}^{n \times k}$ is the generalized eigenvalue problem [Von Luxburg, 2007]:

$$LH = NH\Lambda \quad (7)$$

The eigenvectors for Ncut and Mcut are identical due to this relaxation, while for Ncut, Eq.7 can be converted into a normal eigenvalue problem:

$$SZ = Z\Delta \quad (8)$$

where

$$S = S_{sym} = D^{-1/2}WD^{-1/2}, \quad Z = D^{1/2}H \text{ and } \Delta = I - \Lambda \quad (9)$$

or

$$S = S_{rw} = D^{-1}W, \quad Z = H \text{ and } \Delta = -\Lambda \quad (10)$$

The above method leads to the normalized graph Laplacians defined as:

$$L_{sym} = I - D^{-1/2}WD^{-1/2} \text{ and } L_{rw} = I - D^{-1}W \quad (11)$$

¹⁴⁹ L_{sum} is a symmetric matrix, while L_{rw} it is closely related to a random walk. A more
¹⁵⁰ detailed description for normalized graph Laplacian can be found in Chung [1997]. The data
¹⁵¹ clustering by graph-cut boils down to the eigenvalue decomposition problem of S for finding
¹⁵² the cluster indicators h_1, \dots, h_k . These eigenvectors induce an embedding of the data points
¹⁵³ in a low-dimensional subspace wherein a partitioning based on the normalized cut (NCut)
¹⁵⁴ can be used. The solution of the minimization problem can be obtained from the Fiedler
¹⁵⁵ eigenvector [Yu and Shi, 2003]. The steps involved in DEM segmentation using spectral
clustering are summarized in Figure 2.

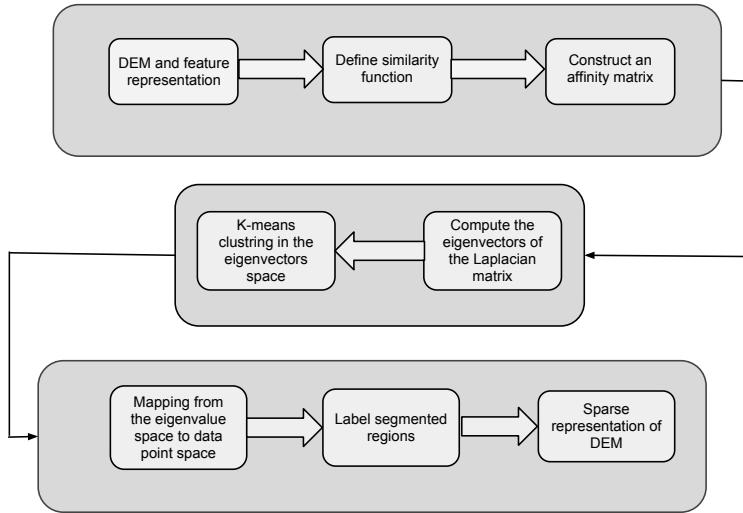


Figure 2: Spectral clustering for DEM segmentation workflow

¹⁵⁶

¹⁵⁷ 2.1.2 Parallel implementation

¹⁵⁸ When the number of data points (n) is large, the computational complexity of spectral
¹⁵⁹ decompositions can reach $O(n^3)$ (W dense). If the affinity matrix W is define as in Equation
¹⁶⁰ 1 then its construction takes $\mathcal{O}(n^2d)$ flops, which can be also computationally intensive if the

161 data cardinality n or dimensionality d is large. Chen et al. [2011] investigate approaches for
162 large-scale spectral clustering and propose a parallel implementation, which is also adopted
163 in this paper. The strategy to address the computational and memory difficulties involves
164 distributing n data instances onto p distributed machine node. On each node, parallel
165 spectral clustering (PSC) computes the similarities between local data and the whole set in
166 a way that uses minimal disk I/O. Then PSC stores the eigenvector matrix on distributed
167 nodes to reduce per-node memory use. Together with parallel eigensolver and K -means,
168 PSC archives good speedup with large data sets.

169 3 DEM ensembles

When propagating uncertainty in DEMs through a geophysical system, stochastic methods are considered to be an effective way to estimate the probability density function of outputs by addressing uncertainties present in initial conditions and in model approximations. In previous work Stefanescu et al. [2012a] presented a basic methodology for generating ensembles of DEMs representative of the true DEM. Here, we are extended the methodology at the cluster level, by making the assumption that each homogenous regions has its own error model which leads to different random fields. The random fields are used in creating multiple equally likely representations of an actual terrain surface, following the approach suggested by Ehlschlaeger and Goodchild [1994]. A normal distribution (mean of 0.0 and variance of 1.0) of maps or realizations is computed to reproduce the spatial autocorrelation encountered in the original error surface, filtered using a Gaussian convolution filter, with kernel sizes derived from autocorrelation analysis of the original error surfaces. The random field function derives its spatial dependence from the use of a distance based decay filter

function. The following equation is used to generate the random field:

$$Z(\mathcal{U}) = \frac{\sum_v w_{u,v} \epsilon_v}{\sqrt{\sum_v w_{u,v}^2}}, \quad u \in \mathcal{U}, v \in \mathcal{V} \quad (12)$$

$$w_{u,v} = \begin{cases} 1 & : d_{u,v} \leq F \\ \left(1 - \frac{d_{u,v}-F}{D-F}\right)^E & F < d_{u,v} \leq D, u \in \mathcal{U}, v \in \mathcal{V} \\ 0 & : d_{u,v} > D \end{cases} \quad (13)$$

where \mathcal{V} is the set of points potentially influencing points in a given area, \mathcal{U} , $w_{u,v}$ is the spatial autocorrelative effect between points $u \in \mathcal{U}$ and $v \in \mathcal{V}$, ϵ_v is a Gaussian random variable with a mean of 0 and variance of 1, $d_{u,v}$ is the distance between u and v , D is the minimum distance of spatial independence, E is the distance decay exponent, and F the distance at which errors are completely correlated.

A set of random fields are created for each homogenous region/ cluster and are calibrated to the spatial variation of the field being simulated using a correlogram function. This is done by fitting the correlogram and choosing the best descriptive parameters of the random field (the minimum distance of spatial independence, the correlated distance decay exponent and the filter parameter) in a weighted least-square estimator. After running hundreds of tests with multiple combinations of D , E and F , the best random field was found by fitting the error map characteristics such that the sum of least squares difference between an error field's correlogram and the target correlogram is minimized.

Each error realization was added to the “true” DEM indicated as $m(\mathcal{U})$, to generate equally probable realizations of the topography for the error structure of a DEM under consideration:

$$R(\mathcal{U}) = m(\mathcal{U}) + m(m(\mathcal{T})) + (m(s^2(\mathcal{T})) \cdot \epsilon) \cdot Z(\mathcal{U}) \quad (14)$$

where $R(\mathcal{U})$ is a realization of the elevation dataset $m(\mathcal{U})$, \mathcal{T} is a group of sets of spatially uncorrelated sample points in $m(\mathcal{U})$, and ϵ is a Gaussian random variable with mean 0.0

185 and variance 1.0. $m(m(\mathcal{T}))$ and variance $m(s^2(\mathcal{T}))$ is mean and variance, respectively, of all
 186 sets in \mathcal{T} . $Z(\mathcal{U})$ specifies the random field as defined in Equation 12 for each homogenous
 187 region.

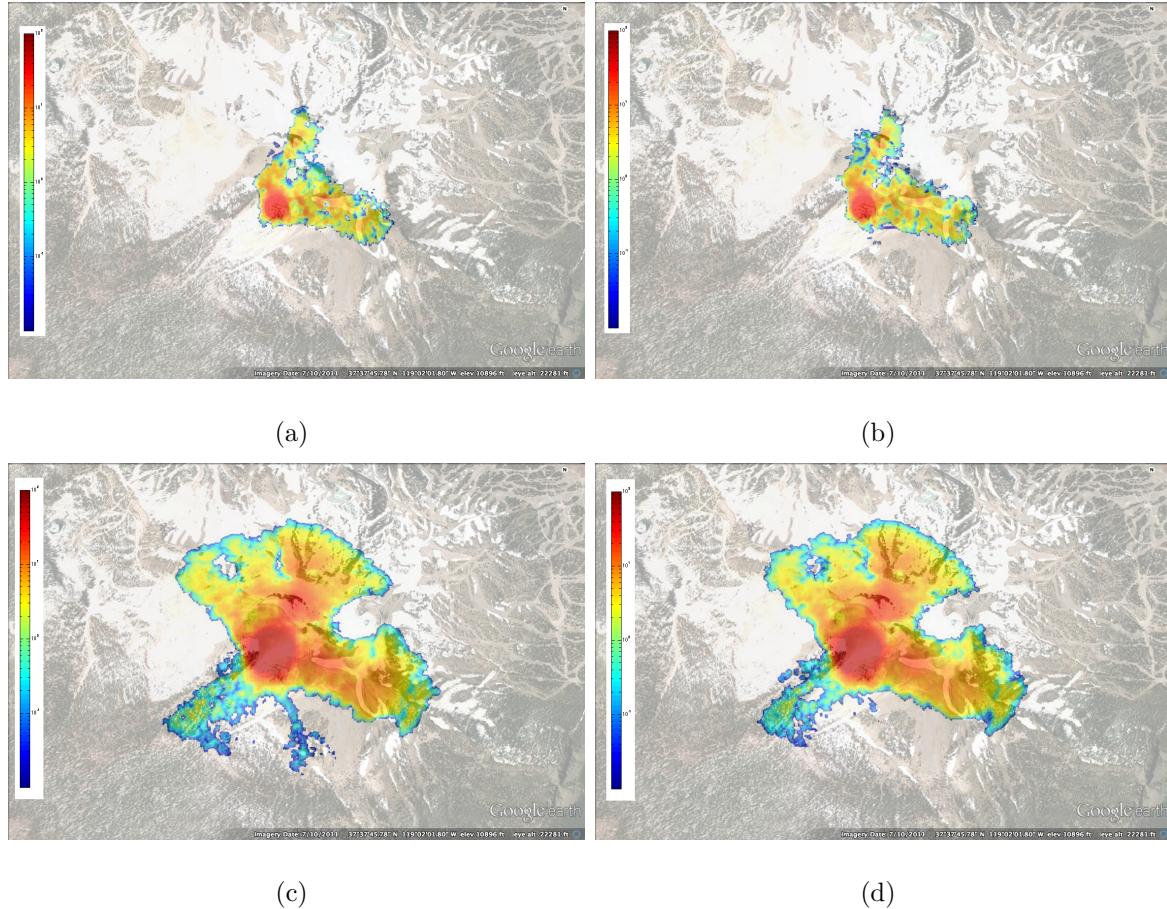


Figure 3: Flow maps for two selected parameters values when DEM was created using a) no
 cluster b) cluster

188 4 Flow simulator

189 The threat of avalanches, block-and-ash flows, and mud-flows at volcanoes is a major global
 190 problem. These are complex phenomena and involve physics at multiple spatial and temporal

191 scales. Developing satisfactory models and computational simulations of potential debris
192 flows on natural terrains and integrating them with appropriate geographical information
193 is difficult but extremely necessary. In recent years, many significant advances have been
194 made in the modeling and simulation of such flows by taking advantage of new models of
195 the physics, new stable and accurate solution schemes for nonlinear hyperbolic systems, and
196 the wide availability of high-performance computers.

197 Hazard maps are one of the most important instruments [Sheridan et al., 2010] for rep-
198 resenting areas of potential inundation by dangerous phenomena at volcanoes. One major
199 problem with using model to construct hazard maps is the strong dependence on the outcome
200 on the choice of model parameters, such as bed friction, initial volumes and terrain.

201 4.1 TITAN2D

202 TITAN2D [Patra et al., 2005, Sheridan et al., 2005] was developed for modeling dry geo-
203 physical granular flows, such as debris avalanches and block and ash flows. The TITAN2D
204 code combines numerical simulations of a natural granular flow with digital terrain data.
205 It is based on a depth-averaged model for an incompressible granular materia governed by
206 Coulomb-type friction interactions [Savage and Hutter, 1989]. The governing equations are
207 obtained by applying conservation laws to the incompressible continuum, providing appro-
208 priate constitutive modeling assumptions, and then taking advantage of the shallowness of
209 the flows (flows are much longer and wider than they are deep) to obtain simpler depth-
210 averaged representations [Bursik et al., 2005]. The motion of the material is considered to
211 be gravitationally driven and resisted by both internal and bed friction. The stress boundary
212 conditions are: no stress at the upper free-surface and a Coulomb-like friction law imposed
213 at the interface between the material and the basal surface.

214 A principal feature of TITAN2D is the incorporation of topographical data from geo-

215 graphic information system (GIS) sources into the simulation and grid structure. Topo-
216 graphic data are included in the simulation through a preprocessing routine in which the
217 digital elevation data are imported. TITAN2D performs flow simulations on a DEM of a
218 desired region, the simulation accuracy being highly dependent on the level of the DEM
219 resolution and quality [Stefanescu et al., 2012b].

220 Inputs to the code are the size and location of the initial volume, the internal and bed
221 friction and the DEM. Dalbey et al. [2008] presented several methods for characterizing the
222 effect of input data uncertainty on model output – except DEM, where uncertainty associated
223 with spatial parameters like terrain elevation were not well understood. The output – the
224 flow height at every grid point at every timestep – is a complete description of the mass flow
225 over realistic terrain.

226 We define the stochastic input as $\Omega = (UTM_E, UTM_N, \mathbf{V}, \theta_r)^\top$. UTM_E and UTM_N
227 are the East and North, respectively coordinates (UTM values) of the location of possible
228 vents. These are considered to be uniform distributed around 321095 E and 4166433 N,
229 within 400m radius. \mathbf{V} is the initial volume of material, uniform distributed between 10^5
230 m^3 and $10^{7.5} m^3$. θ_r includes the DEM uncertainty as described in Section 3. We ran 1024
231 flow simulations at design points in the region of the input space. These 1024 design points
232 were chosen according to a Latin hypercube design, which is a space-filling design. This has
233 been proven very successful for all-purpose designs of computer experiment runs since they
234 require relative few design points per input to “fill” the design space [Sacks et al., 1989].

235 **5 Results**

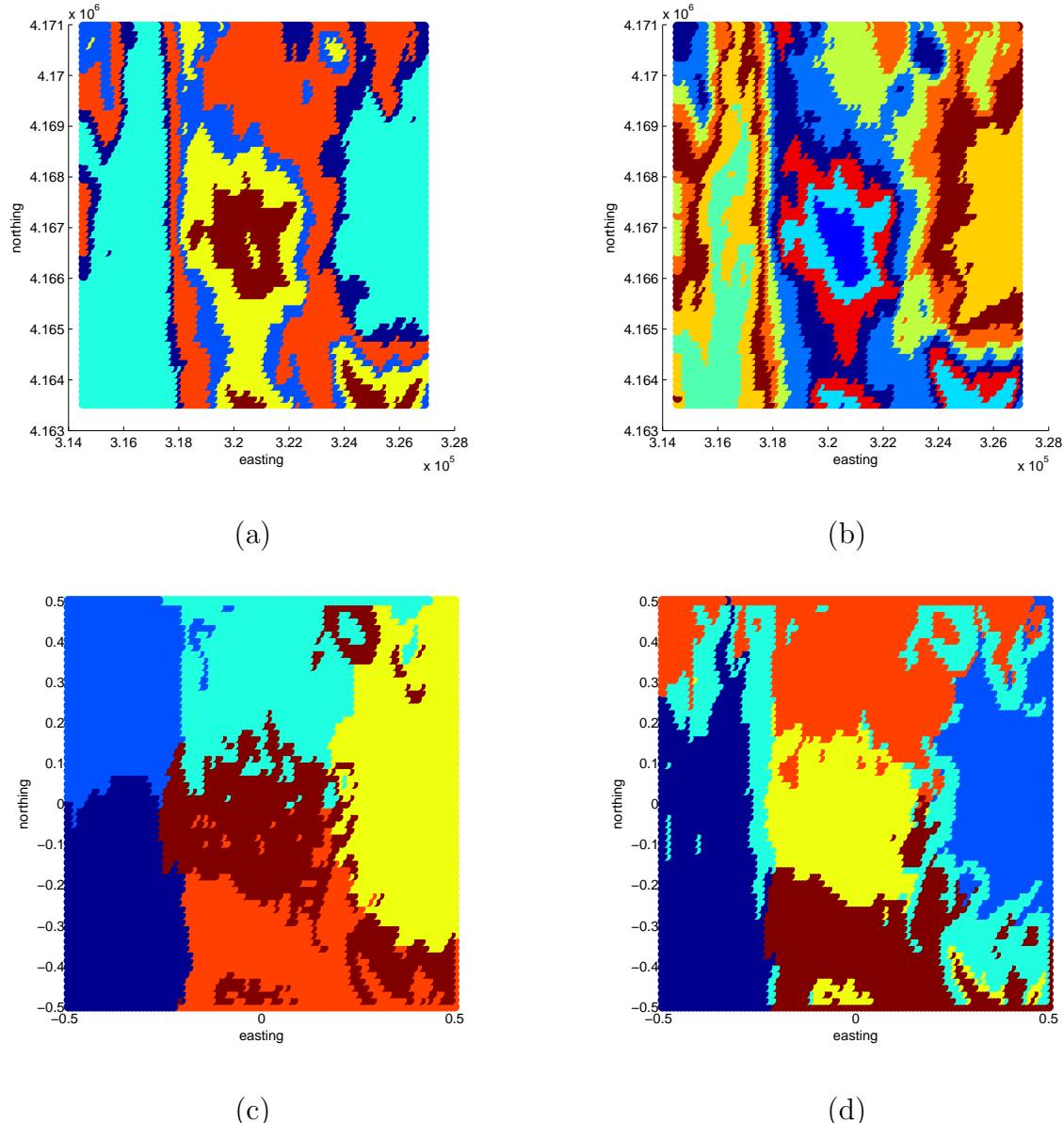


Figure 4: Spectral Clustering for a) $K=6$ – one feature b) $K=20$ – one feature c) $K=6$ where $\sigma_F = 0.8$ and $\sigma_x = 1$ – 4 feature d) $K=6$ $\sigma_F = 0.8$ and $\sigma_x = 0.8$ – 4 feature

236 6 Conclusions

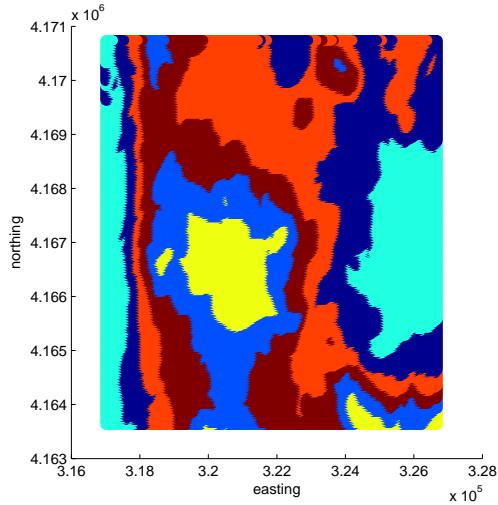


Figure 5: Parallel spectral clustering

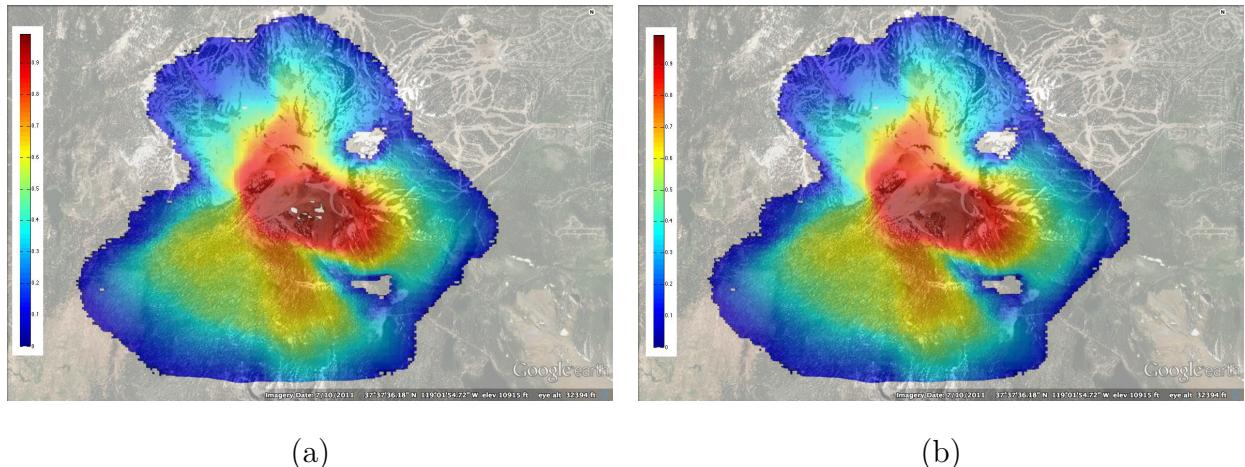


Figure 6: Hazard map a) no cluster b)clusters

²³⁷ **Acknowledgment**

²³⁸ **References**

- ²³⁹ N. Archip, P. Rohling, H. Tahmasebpour, and S.K. Warfield. Spectral clustering algorithms
²⁴⁰ for ultrasound image segmentation. *in: MICCAI 2005, Lecture notes in Computer Science*,
²⁴¹ 3750:862–869, 2005.
- ²⁴² R.A. Bailey. *Geologic Map of Long Valley Caldera, Mono-Inyo Craters Volcanic Chain and*
²⁴³ *Vicinity, Eastern California*. Department of the Interior, Reston, VA (US), 1989.
- ²⁴⁴ P.A. Burrough, P.F.M. van Gaans, and R.A. McMillan. High-resolution landform classifica-
²⁴⁵ tion using fuzzy k-means. *Fuzzy Sets and Systems*, 113:37–52, 2000.
- ²⁴⁶ M. Bursik, A. Patra, E. B. Pitman, C. Nichita, J. L. Macias, R. Saucedo, and O. Girina.
²⁴⁷ Advances in studies of dense volcanic granular flows. *Reports on Progress in Physics*, 68:
²⁴⁸ 271–301, 2005. doi:10.1088/0034-4885/68/2/R01.
- ²⁴⁹ A.P. Carleer, O. Debeir, and E. Wolff. Assessment of very high spatial resolution satellite
²⁵⁰ image segmentations. *Photogrammetric Engineering and Remote Sensing*, pages 1285–
²⁵¹ 1294, 2005.
- ²⁵² W.Y. Chen, Y. Song, H. Bai, C.J. Lin, and E.Y. Chang. Parallel spectral clustering in
²⁵³ distributed systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,
²⁵⁴ 33 (3):568–586, 2011.
- ²⁵⁵ F.R.K Chung. Spectral graph theory. *CBMS Regional Conference Series in Mathematics*,
²⁵⁶ *American Mathematical Society*, 92, 1997.

- 257 K. Dalbey. Predictive simulation and model based hazard maps of geophysical mass flows.
258 *PhD thesis, Department of Mechanical and Aerospace Engineering, University at Buffalo,*
259 2009.
- 260 K. Dalbey, A.K. Patra, E.B. Pitman, M.I. Bursik, and M.F. Sheridan. Input uncertainty
261 propagation methods and hazard mapping of geophysical mass flow. *Journal of Geophysical*
262 *Research*, 113:5203–5219, 2008. doi:10.1029/2006JB004471.
- 263 I.S. Dhillon, P. Guan, and B. Kulis. Weighted graph cuts without eigenvectors: A multilevel
264 approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29 (11):
265 1944–1957, 2007.
- 266 C.H.Q Ding, X. He, M. Gu, and H.D. Simon. A minmax cut algorithm for graph partitioning
267 and data clustering. *Proceedings of the IEEE International Conference on Data Mining*
268 (*ICDM*), pages 107–114, 2001.
- 269 W.E. Donath and A.J. Hofmann. Lower bounds for the partitioning of graphs. *IBM Journal*
270 *of Research and Development*, 17:420–425, 1973.
- 271 C. Ehlschlaeger and M.F. Goodchild. Uncertainty in spatial data: Defining, visualizing, and
272 managing data errors. In *Proceedings GIS/LIS'94*, pages 246–253, 1994.
- 273 C. Fowlkes, F. Belongie, F. Chung, and J. Malik. Spectral grouping using the nyström
274 method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26 (2):214–
275 225, 2004.
- 276 M. Goldstein. Bayes linear methods I adjusting beliefs: Concepts and properties. *Part 1 of*
277 *3 of online tutorial, website:*, 2007.

- 278 S.R. Gunn. Support vector machines for classification and regression. *Technical report image*
279 *speech and intelligent group Southampton University*, 1997.
- 280 L. Hagen and A. Kahng. New spectral methods for ratio cut partitioning and clustering.
281 *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 11 (9):
282 1074–1085, 1992.
- 283 J. Kohn, M.S. Suk, and S.M. Bhandarkar. A multilayer self organizing feature map for range
284 image segmentation. *Neural Networks*, 8 (1), 1995.
- 285 S. Mahadevan. Fast spectral learning using Lanczos eigenspace projections. *in AAAI*, pages
286 1472–1475, 2008.
- 287 L. Najman and M. Schmitt. Geodesic saliency of watershed contours and hierarchical seg-
288 mentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):
289 1163–1173, 1996.
- 290 A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: analysis and an algorithm. *In*
291 *T. Dietterich, S. Becker, and Z. Ghahramani (Eds.), Advances in Neural Information*
292 *Processing Systems*, 14:849–856, 2002.
- 293 A. Patra, M.I. Bursik, J. Dehn, M. Jones, M. Pavolonis, E. B. Pitman, T. Singh, P. Singla,
294 E.R. Stefanescu, S. Pouget, and P. Webley. Challenges in Developing DDDAS Based
295 Methodology for Volcanic Ash Hazard Analysis Efect of Numerical Weather Prediction
296 Variability and Parameter Estimation. In *submitted to International Conference on Com-*
297 *putational Science*, 2013.
- 298 A.K. Patra, A.C. Bauer, C. Nichita, E.B. Pitman, M. F. Sheridan, M. Bursik, B. Rupp,
299 A. Webber, L. Namikawa, and C. Renschler. Parallel adaptive numerical simulation of dry

- 300 avalanches over natural terrain. *Journal of Volcanology and Geothermal Research*, 139:
301 1–21, 2005.
- 302 J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn. Design and analysis of computer
303 experiments. *Statistical Science*, 4(4):409–435, 1989.
- 304 T. Sakai and A. Imiya. Fast spectral clustering with random projection and sampling.
305 *Machine Learning and Data Mining in Pattern Recognition*, pages 372–384, 2009.
- 306 S.B. Savage and K. Hutter. The motion of a finite mass of granular material down a rough
307 incline. *Journal of Fluid Mechanics*, 199:177–215, 1989.
- 308 M.F. Sheridan, A.J. Stinton, A. Patra, E.B. Pitman, A. Bauer, and C.C. Nichita. Evaluating
309 TITAN2D mass-flow model using the 1963 Little Tahoma Peak avalanches, Mount Rainier,
310 Washington. *Journal of Volcanology and Geothermal Research*, 139:275–308, 2005.
- 311 M.F. Sheridan, A.K. Patra, K. Dalbey, and B Hubbard. Probabilistic digital hazard maps for
312 avalanches and massive pyroclastic flows using TITAN2D. *in Grippelli G. and Viereck-*
313 *Goette, L., editors, Stratigraphy and Geology of Volcanic Areas , Geological Society of*
314 *America Special Paper*, 464:281–291, 2010.
- 315 J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern*
316 *Analysis and Machine Intelligence*, 22 (8):888–905, 2000.
- 317 Y. Song, W.Y. Chen, H. Bai, C.J. Lin, and E.Y. Chang. Parallel spectral clustering. *in:*
318 *Daelemans, W., Goethals, B., Morik, K. (eds.) ECML PKDD 2008, Part II*, 5215:374–389,
319 2008.
- 320 M. Sonka, V. Hlavac, and R. Boyle. Image processing, analysis and machine vision. *PWS*,
321 1999.

- 322 E. R. Stefanescu, M. Bursik, G. Cordoba, K. Dalbey, M. D. Jones, A. K. Patra, D. C. Pieri,
323 E. B. Pitman, and M. F. Sheridan. Digital elevation model uncertainty and hazard analysis
324 using a geophysical flow model. *Proceedings of the Royal Society A: Mathematical, Physical*
325 and *Engineering Science*, 468(2142):1543–1563, 2012a. doi: 10.1098/rspa.2011.0711. URL
326 <http://rspa.royalsocietypublishing.org/content/468/2142/1543.abstract>.
- 327 E.R. Stefanescu, M. Bursik, and A.K. Patra. Effect of digital elevation model on mohr-
328 coulomb geophysical flow model output. *Natural Hazards*, 62(2):635–656, 2012b. doi:
329 10.1007/s11069-012-0103-y. URL <http://dx.doi.org/10.1007/s11069-012-0103-y>.
- 330 F. Tung, A. Wong, and D.A. Clausi. Enabling scalable spectral clustering for image segmen-
331 tation. *Pattern Recognition*, 43:4069–4076, 2010.
- 332 U. Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17 (4):395–416,
333 2007.
- 334 S.X. Yu and J. Shi. Multiclass spectral clustering. In: *Proc. Internat. Conf. on Computer*
335 *Vision*, pages 313–319, 2003.