Jul 1st, 12:00 AM

Modelling Lava Flow to Assess Hazard on Mount Etna (Italy). From Geological Data to a Preliminary Hazard Map

E. Bertino

M. L. Damiani

G. Groppelli

G. Norini

B. Aldighieri

See next page for additional authors

Follow this and additional works at: https://scholarsarchive.byu.edu/iemssconference

https://scholarsarchive.byu.edu/iemssconference/2006/all/345

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Presenter/Author Information
E. Bertino, M. L. Damiani, G. Groppelli, G. Norini, B. Aldighieri, S. Borgonovo, F. Comoglio, and G. Pasquaré
Modelling Lava Flow to Assess Hazard on Mount Etna (Italy). From Geological Data to a Preliminary Hazard Map

E. Bertino1, M.L. Damiani2,6, G. Groppelli3, G. Norini4,5, B. Aldighieri1, S. Borgonovo2, F. Comoglio4, G. Pasquare6

1 CS and ECE Departments, Purdue University, West Lafayette IN, USA
2 Dipartimento di Informatica e Comunicazione, Università degli Studi di Milano, via Comelico 39/41, 20135 Milano, Italy
3 Istituto per la Dinamica dei Processi Ambientali, sezione di Milano, CNR, via Mangiagalli 34, 20133 Milano, Italy
4 Dipartimento di Scienze della Terra, A. Desio, Università degli Studi di Milano, via Mangiagalli 34, 20133 Milano, Italy
5 Istituto Nazionale di Geofisica e Vulcanologia - Sezione di Catania, Piazza Roma 2, 95123 Catania, Italy
6 EPFL-IC-LBD, Lausanne, CH

Abstract: In this paper we present a systematic approach to the development of the lava flow hazard map for the Mount Etna, the most important active volcano in Europe. The basic idea is to determine the hazard zones by simulating the lava flows originated from a number of sample points localized in regions at high density of vents, called eruption zones. The key choice is the adoption of a probabilistic model for the simulation of the lava flow. In the paper we outline the characteristics of the model, called ELFM, and how it has been validated. On the other side, for the determination of the eruption zones, which likely contain the points of future lava flows emissions, we propose an approach based on data mining techniques.

Keywords: Volcanic hazard map; Probabilistic lava flow simulation model; Lava flow invasion assessment; Geological map; GIS

1. INTRODUCTION

For the development of land use and emergency plans, volcanic hazard maps represent a fundamental resource. Such maps specify the geographical areas that may be affected by dangerous volcanic phenomena, such as lava flows. Typically, a hazard map divides the volcanic region into a number of zones that are differently classified based on the probability of being damaged in a given period of time by a specific volcanic event. However, volcanic hazard maps are very complex to generate. Moreover, because they are typically based on a subjective interpretation of historical data, their reliability is difficult to estimate. In order to provide a more rigorous and systematic approach to the development of volcanic hazard maps, simulations of phenomena can be very helpful. For such purpose, we propose an approach for the evaluation of the lava flow invasion hazard and the construction of the hazard map for the Mount Etna (Italy), based on the integration of simulation techniques with the analysis of an accurate geological map of the volcano (Acireale Geological Map, in press) through data mining and spatial data management techniques. The Mount Etna, with an altitude of 3323 m, is a challenging case study since it is the most active volcano in Europe. The Etna volcanic activity started about 500,000 years ago (Gillot et al., 1994). Recent approaches based on the application of high-resolution stratigraphy allow recognition of hundreds lava flows emplaced on the Etna volcano flanks (Acireale Geological Map, in press; Branca et al., in press). These approaches are able to reconstruct a detailed history of the volcano and allow to recognise 4 main phases in Etna evolution (Branca et al., 2004 and references therein). The last phase started about 80,000 years ago when the construction began of the main stratovolcano, named Ellittico Volcano, which forms the bulk of the present edifice. About 15,000 years ago the Ellittico Volcano ended with a caldera...
collapse (Coltelli et al., 2000). The following Mongibello Volcano, which is the responsible for the last 15,000 years of the Etna’s activity, developed inside the Ellittico caldera. The lava flows belonging to the Mongibello Volcano (aged less than 15,000 years) crop out over about 85% of the Etnean edifice (Branca et al., 2004). During the history of the Mongibello Volcanoc there have been frequent eruptions not only in the past but even in recent times, mainly consisting of basalt lava flows (Azzaro and Neri, 1992; Behncke and Neri 2003). Further, several towns and important economic activities are located nearby and on the flanks of the volcano, so that a lava flow hazard map is definitely needed. Nevertheless the only hazard maps that have been realized so far for Mount Etna go back to twenty years ago and are based on incomplete knowledge of past eruptions and only on qualitative analysis (Guest and Murray, 1979; Duncan et al., 1981; Forgione et al., 1989; Chester et al., 2002).

For the development of a more accurate hazard map of Mount Etna, we propose an approach based on four main steps:

1) Development of a simulation model of lava flows to determine the extent of lava flows originated from emission points. This includes the improvement and testing of a probabilistic lava flow simulation model (ELFM, Damiani et al., 2006), previously developed for volcanoes at Canary Islands (Felpeto et al., 2001), based on a high accuracy digital elevation model (DEM) to obtain the morphological constraint of the lava flow simulation and the most probable lava flow path (Pareschi et al., 1999). ELFM has been extensively tested through a validation procedure based on the comparison of simulation results with historical lava flows.

2) Organization of an accurate and comprehensive spatial database containing topographical and geological data including data about past eruptions and analyses of the geological data, which include the extension of the most recent lava flows, lava flow length, vents location, and vents age (Groppelli and Norini, 2005).

3) Application of clustering techniques to group past vents based on their location, age and characteristics of the generated lava flows in order to outline eruption zones, that is, the zones in which the probability of eruption is higher.

4) Validation of the preliminary hazard map through comparison with the recent volcanic evolution. The idea is to combine the above three steps as follows: first the eruption zones are determined based on historical data; then lava flow simulations are generated from sample emission points located in eruption zones, so that the areas which might be covered by lava flows are outlined; finally, the whole procedure is tested through a validation procedure based on the comparison of simulation results with historical lava flows.

In this paper we focus in particular on the techniques developed for the simulation of the lava flows (ELFM), its validation procedure and the clustering of emission points. We present the basic characteristics of the ELFM system integrated with a high quality geological database. Such system is based on a simulation model which extends the stochastic model developed by Araña et al. (2000) and Felpeto et al. (2001). A key aspect of the method is that it does not require many different types of data and yet it provides an accurate modelling for the estimation of the lava flow hazard. With respect to the clustering of emission points, we have adopted a relatively recent clustering technique which has its roots in data mining. In this paper we discuss the motivations for selecting this algorithm and, because the work is still in progress, some preliminary results and open issues.

The remainder of the paper is structured as follows. In the next section we present the lava flow simulation system, in particular the underlying simulation model and the validation process which has been applied to estimate the reliability of the simulation model. Then we discuss the challenges concerning the application of data-mining based clustering techniques for the identification of the eruption zones. Open issues and plans for future work are presented in the concluding section.
2. SIMULATION OF THE LAVA FLOWS – ELFM

The lava flow simulation system is the core component of the process proposed for the construction of the hazard map. The key choice has been the adoption of a probabilistic model for lava flow simulation. In particular we have adopted and then extended the model developed by Araña et al. (2000) and Felpeto et al. (2001) (hereinafter called Felpeto’s model) and so far only applied to the evaluation of the hazard of the Tenerife and Lanzarote volcanoes (Canary Islands). In what follows, we summarize the basic characteristics of the model. For the details we refer the reader to Damiani et al. (2006).

2.1 An overview of the lava flow simulation model

The basic parameters of the model include: the vent (emission point), which we assume consisting of a point with coordinates \((x,y)\); and the DEM, which provides the grid representation of the topographic height. Given a DEM and an emission point, the algorithm iteratively computes a number of paths for the lava flow. Each path represents a sequence of grid cells which are determined based primarily on the topography of the terrain near the emission point. Then the model assigns each point of the region a probability to be invaded by the lava flow. The maximum length of each single path is specified as an input parameter and the number of paths for the lava flow alike.

The single path is computed stepwise: at each step, starting form the emission point, the lava propagates from the current cell to one of the adjacent cells. The next cell in the sequence is selected by first assigning a slope-dependent probability to the eight neighbouring cells and then randomly selecting one of the candidate cells for which that value is positive.

The major extensions we have introduced in the model are as follows: a) We consider the DEM to be dynamic. This allows us to account of the natural fact that the height of the terrain in a point increases as the lava flow covers that point. Therefore, in our model the altitude of a point is increased of a factor \(\delta\), representing the lava thickness, as the point is covered by the simulated lava flow. b) The second extension is to account of the fact that the thickness of lava flow may vary depending on the physical properties of the lava flow and the topography. In order to address this requirement, we have introduced an additional parameter specifying how the lava flow height would vary. The value of such a parameter is a mathematical function; in the simplest case, it is a constant function.

3. VALIDATION PROCESS

In the development of a simulation model a crucial phase is the validation of the model which is fundamental to estimate the reliability of simulations. The validation strategy we propose is based on a comparison of the simulated lava flow against eruptive events occurred in the past. To that extent, we have selected some lava flow samples (Fig. 1), as their coverage can be precisely outlined and also because of the topographic and geological characteristics of the lava flows. Finally the simulations of lava flows originated from the corresponding emission points have been generated and compared with historical data (Fig. 2). We have applied a quantitative analysis to evaluate the fitting between the historical lava flow and the simulated flow. We identify three different areas, shown in Fig. 3. We use a simple algorithm to evaluate each area, based on the ratio among each area and the grand total of A, B and C (Table 1). Particularly, the area C represents the historical lava flow area which is not covered by the simulation flow. This value outlines the under-evaluated area and in our simulation is always less than 10%, and mostly less than 5%. At the opposite the area B is the overestimated area in respect to the real case, but this area is covered by a few lava flow paths, so the probability results are in any case very low. As it can be observed (Fig. 2), there is a geometric mismatch between the simulated lava flow and the historical boundaries. However we claim that such discrepancy can be tolerated because the purpose of the simulations is not that of predicting the path of a single lava flow but rather of shaping the hazard zones of the volcano. In addition the areas (C of Fig. 3 and table 1) not covered by the simulated lava flow with respect to the historical flow are very limited and so it means the under-evaluation can be acceptable for a lava flow hazard assessment. Moreover it is important to remark that these simulations are generated using minimal information about the eruptions and do not include for example any data about effusion rate, eruption duration and the type of lava flow field. Ultimately the model is able to work well although some volcanic parameters are not known and that is extremely important for the development of the hazard map. Another important result of this activity has been the specification of possible values for the parameters of the simulation model.
complete data set of emission points, which are currently available for the volcano. Based on the consideration that behavioural patterns extracted from the records of past eruptions may be useful to predict the future behaviour of the volcano, we assume that the regions at high density of emission points (i.e. eruption zones) are those which contain the probable origins of future lava flows. To identify the possible eruption zones, we are currently investigating the suitability of a particular class of

4. CLUSTER ANALYSIS

A major step of the overall methodology proposed for the construction of the volcanic hazard map is the identification of the regions presenting a high density of emission points. For that purpose, we consider the emission points of the lava flows that occurred in the last 4 centuries, precisely since year 1669 (Branca & Del Carlo, 2005). This set consists of 134 points and constitutes the most reliable and

Figure 2. Simulated lava flows against occurred lava flows. For the legend see fig. 3 and table 1; for the location, see fig. 1. A and B) 1991-93 lava flow; C and D) 2001 lava flow.
Table 1. Data entry used to simulate the 1991-93 and 2001 lava flows. A, B and C refer to Fig. 3 and represent the parameter used to validate the simulation model. Note that C is the area not covered by our simulation software in respect to the real lava flow.

<table>
<thead>
<tr>
<th>FLOW</th>
<th>PIXEL WIDTH</th>
<th>FLOW HEIGHT</th>
<th>FLOW HEIGHT VARIATION</th>
<th>ITERATIONS</th>
<th>FLOW LENGTH</th>
<th>THRESHOLD (%)</th>
<th>A (%)</th>
<th>B (%)</th>
<th>C (%)</th>
<th>FIGURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991-93</td>
<td>10</td>
<td>10</td>
<td>LOG</td>
<td>1000</td>
<td>4500</td>
<td>0</td>
<td>55.9</td>
<td>38.4</td>
<td>2.7</td>
<td>Fig. 2A</td>
</tr>
<tr>
<td>1991-93</td>
<td>10</td>
<td>10</td>
<td>LOG</td>
<td>1000</td>
<td>4500</td>
<td>0.5</td>
<td>60.0</td>
<td>15.9</td>
<td>4.50</td>
<td>Fig. 2B</td>
</tr>
<tr>
<td>1991-93</td>
<td>25</td>
<td>10</td>
<td>LOG</td>
<td>1000</td>
<td>1000</td>
<td>0</td>
<td>66.10</td>
<td>32.86</td>
<td>0.23</td>
<td>Fig. 2C</td>
</tr>
<tr>
<td>2001</td>
<td>10</td>
<td>10</td>
<td>LIN</td>
<td>1000</td>
<td>2100</td>
<td>0</td>
<td>39.5</td>
<td>59.7</td>
<td>1.8</td>
<td>Fig. 2D</td>
</tr>
<tr>
<td>2001</td>
<td>10</td>
<td>10</td>
<td>LIN</td>
<td>1000</td>
<td>2100</td>
<td>0.5</td>
<td>44.2</td>
<td>48.3</td>
<td>9.5</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>25</td>
<td>10</td>
<td>LIN</td>
<td>1000</td>
<td>500</td>
<td>0</td>
<td>37.9</td>
<td>61.9</td>
<td>0.2</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Parameter used to validate our software and simulation model: we consider the percentage of three different areas (A, B and C) to evaluate the reliability of our simulation model (Felpeto’s model). The results are described in Table 1.

Clustering methods, the *density-based clustering techniques*. Clustering is the process of grouping data into classes or clusters so that objects within a cluster have high similarity with respect to each other, but are very dissimilar with respect to objects in other clusters (Han & Kamber, 2001). In particular, density-based algorithms regard clusters as dense regions of objects in the data space that are separated by regions of low density, which represent noise. Density-based algorithms have notable characteristics for the grouping of emission points:
- They are able to discover clusters of arbitrary and irregular shapes. It is likely in fact that the clusters of emission points do not have a regular shape.
- They are insensitive to noise; thus clusters are not affected by noise observations.
- No a-priori knowledge about the data or the number of clusters to be created is required. This is important in our application, since it is not possible to specify the number of clusters of emission points.

To carry out this study we have used the clustering algorithm called OPTICS (Ankerst & al., 1999). This algorithm requires the specification of the following two parameters:
- The minimum number \( N \) of objects contained in each cluster;
- The distance \( \text{eps} \) representing the radius of the circle (\( \text{eps-neighborhood} \)) containing at least \( N \) objects.

The basic idea underlying the algorithm is to insert in the same cluster the objects which are at a distance less than \( \text{eps} \) from some other object of the cluster. For each cluster an initial seed is found. The objects which do not belong to any cluster are classified as *noise*. The outcome of the algorithm is a set of clusters in which each cluster is identified by a unique identifier \( \text{Id} \). Members of clusters can then be distinguished based on such value assigned to each point. The question that remains to address is how to define the values of the parameters \( N \) and \( \text{eps} \). Intuitively, a low value for \( N \) means that clusters may be rather rarefied, while a high value of \( \text{eps} \) means that the cluster may consist of points which are quite far from each other. To address this issue, we have analyzed the behaviour of the system for different values of \( N \) and \( \text{eps} \). With respect to the value of \( \text{eps} \), on the basis of the known distance between vents, we have concluded that reasonable values for \( \text{eps} \) are: \( \text{eps}=1 \) km and \( \text{eps}=1.5 \) km. With respect to the value of \( N \), it can be noticed that the percentage of noise, when the value of \( N \) changes, varies more significantly than when \( \text{eps} \) changes. We interpret this fact as follows: the clusters obtained when \( N \) is set to high values are those which are more “selective” in the sense that group points which are very closely linked. The clusters obtained with a low value of noise are instead those grouping points which are more loosely linked. To maintain, however, the noise into a reasonable percentage, we have assigned \( N \) the following values: \( N=5 \) and \( N=10 \). It can be also noticed that the number of clusters varies significantly for varying values of \( \text{eps} \). For example, with \( \text{eps}=1 \) km, we obtain five clusters. Conversely with \( \text{eps}=1 \) km, the number of clusters is less or equal to two. This is the natural consequence of the fact that a greater value of \( \text{eps} \) results into spatially wider aggregations of points. By varying the two parameters \( N \) and \( \text{eps} \) we thus obtain clusters at different levels of detail.

A related issue is how to define the boundaries of clusters. Since the outcome of the clustering process is a set of points, a relevant question is how to determine the spatial extent of each cluster. A possible approach is to draw such a boundary by connecting the nodes which are on the border of the cluster. The shortcoming is that since the spatial extent of the cluster is meant to contain also future emission points, its shape should not necessarily coincide with the area delimited by past emission points.
points. We have thus adopted a different approach. The idea is to determine for each point of the cluster a buffer area surrounding such point. The buffer denotes an area of uncertainty which accounts of future emission points close to the point. Then we obtain the cluster areas by geometrically merging the generated buffers. Fig. 4 reports the clusters which have been generated by combining the following values: \( N=10, N=5, \) \( \epsilon =1.5 \text{ km} \), \( \epsilon =1 \text{ km} \). The radius of buffers is set to 600 metres. The corresponding percentages of noise and the number of clusters are reported in Table 2.

**Figure 4.** Boundaries of clusters obtained assuming: A) \( N=10, \epsilon =1 \text{ km} \); B) \( N=5, \epsilon =1 \text{ km} \); C) \( N=10, \epsilon =1.5 \text{ km} \); D) \( N=5, \epsilon =1.5 \text{ km} \). The points in each cluster are represented in different colours.
5. CONCLUSIONS

In the paper we have presented the general methodology proposed for the development of the volcanic hazard map for Mount Etna. The core component is the ELFM lava flow simulation model. We have also discussed on-going research on data mining techniques for clustering emission points. A viable approach is represented by the use of density-based algorithms. The experiments we have done so far are based on the OPTICS algorithm. OPTICS is able to identify clusters of arbitrary form and is not affected by noise. However, such algorithm has two major drawbacks. The first is that the outcome is affected by the order according to which the points are considered. In practice it means that some points may be not included in the proper cluster depending on the ordering of input points. Whereas this aspect may be not relevant in the case of clusters consisting of many points, it becomes important when, like in our case, the number of points is relatively low, because the percentages of points which are not properly classified may be significant. The second drawback is that, in order to compute the clusters for different values of \( N \), we need to run the algorithm several times. To improve efficiency it would thus be useful to compute in a single run clusters for a range of values of \( N \). To address these issues we plan to extend the experimental activity to consider improved clustering strategies. In parallel, we plan to address the issue concerning the generation of multiple lava flow simulations from a set of sample emission points and achieve a possibly quantitative estimation of the hazard probability.

6. REFERENCES

Ankerst, M., M. Breunig, M. Kriegel, and J. Sander, OPTICS: Ordering Points To Identify the Clustering Structure, SIGMOD Conference, 49-60, 1999.


<table>
<thead>
<tr>
<th>N</th>
<th>Eps (km)</th>
<th>% noise</th>
<th>N_clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.5</td>
<td>32</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>41</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>1.5</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>28</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 2. Percentage of noise and number of clusters in our four experiments.


