BET_VH: a probabilistic tool for long-term volcanic hazard assessment

Warner Marzocchi[1], Laura Sandri[2], Jacopo Selva[2]

1- Istituto Nazionale di Geofisica e Vulcanologia, Roma, Italy,

2- Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy

Submitted to Bulletin of Volcanology

Abstract

In this paper we illustrate a Bayesian Event Tree to estimate Volcanic Hazard (BET_VH). The procedure enables us to calculate the probability of any kind of long-term hazardous event for which we are interested, accounting for the intrinsic stochastic nature of volcanic eruptions and our limited knowledge regarding related processes. For the input, the code incorporates results from numerical models simulating the impact of hazardous volcanic phenomena on an area, and data from the eruptive history. For the output, the code provides a wide and exhaustive set of spatio-temporal probabilities of different events; these probabilities are estimated by means of a Bayesian approach that allows all uncertainties to be properly accounted for. The code is able to deal with many eruptive settings simultaneously, weighting each with its own probability of occurrence. In a companion paper, we give a detailed example of application of this tool to the Campi Flegrei caldera, in order to estimate the hazard from tephra fall.

1 Introduction

Volcanic hazard studies have a prominent impact on society and volcanology itself, being an area where the "rubber hits the road", that is, where science is applied to an important societal problem. Despite its importance, volcanic hazard assessment is still commonly presented in many different ways, ranging from maps of past deposits of the volcano to more quantitative probabilistic assessment (e.g., Scandone et al. 1993; Newhall and Hoblitt 2002; Marzocchi et al. 2004; Martin et al. 2004; Neri et al. 2008; Marti et al. 2008). The latter, being quantitative, has remarkable advantages: 1) it allows comparisons among different volcanoes and with other natural and non-natural hazards; 2) its reliability can be tested through statistical procedures; 3) it provides a basic component for rationale decision making (e.g. Marzocchi and Woo 2007, 2009; Woo 2008).

In order to distinguish unambiguously the quantitative approach from others more qualitative, it has been suggested to call it *Probabilistic Volcanic Hazard Assessment* (PVHA, hereinafter; see Marzocchi et al. 2007). The term "probabilistic" means that the extreme complexity, nonlinearities, limited knowledge, and the large number of degrees of freedom of a volcanic system make difficult, if not impossible, deterministic prediction of the evolution of volcanic processes (see, e.g., Marzocchi 1996; Sparks 2003). In other words, volcanic systems are stochastic and hazardous volcanic phenomena involve so many uncertainties that a probabilistic approach is needed.

That said, we note that full PVHA is still quite rare (Magill et al. 2006, Ho et al. 2006, Neri et al. 2008, Marti et al. 2008 are among the few remarkable exceptions). Most of the times hazard assessment represents, at best, a conditional probability of one specific hazard conditioned to the occurrence of one specific event (for instance, the most likely event, Cioni et al. 2003; Macedonio et al. 2008), or it is focused on one specific aspect of volcanic hazard, like the vent opening (Martin et al. 2004; Jaquet et al. 2008; Selva et al. this volume). In other cases, as mentioned above, hazard assessment merely consists of maps of volcanic deposits of past events. Despite all of them are valuable information that a reliable PVHA has to account for, we argue that a full PVHA is something more. For example, a full PVHA requires the assessment of the impact of hazardous phenomena associated to every possible "Eruptive Setting" (ES), and eventually the merging of all ESs, each of them weighted with its own probability of occurrence. Hereafter, with the terms "Eruptive Setting" we mean the occurrence of an eruption of a specific size or type from a specific vent.

Another basic feature of PVHA is that it has to account for all relevant sources of uncertainty. As a matter of fact, the great importance of PVHA is due to its practical implications for society. In this perspective, it is fundamental that PVHA is "accurate" (i.e., without significant biases), because a biased estimation would be useless in practice. On the other hand, PVHA may have a low "precision" (i.e., a large uncertainty) that would reflect our scarce knowledge of some physical processes involved, from the preparation of an impending eruption to the derived impact on the surrounding area.

Here, we present a probabilistic tool, named BET_VH (Bayesian Event Tree for Volcanic Hazard), to calculate and to visualize long-term PVHA, accounting for the features described above. In this paper, with the word "long-term" we refer to the time scale of the expected significant variations in volcanic processes. While during unrest the time variations occur in short time scales (from hours to few months), the changes expected during a quiet phase of the volcano are much longer. In input, BET_VH takes the output of different models (ES-based or not), and merges them with available data collected on the field. The result is a long-term probabilistic estimation of the hazard posed by different volcanic hazardous processes (e.g., either ash fall or pyroclastic density currents, or lava flows, or lahars, etc.), accounting for all possible sources of uncertainty in a Bayesian structure. BET_VH has an event tree structure (Newhall and Hoblitt 2002; Marzocchi et al. 2004; see figure 1, panel a), that is a tree graphical representation of events in which individual branches are alternative steps from a general prior event, state or condition, through increasingly specific subsequent events (intermediate outcomes) to final outcomes. In this way, the scheme shows all relevant possible outcomes of volcanic unrest at progressively higher degrees of detail. This structure allows different processes and relative uncertainties to be modeled and accounted for separately, merging output of models and field data (Marzocchi et al. 2008).

In the following sections, we describe the structure of the event tree for BET_VH, and the basic rules to estimate the probability distributions at each node. Finally, we show how a full PVHA is achieved, combining the probabilities of each node. A tutorial application of these concepts is outlined in the companion paper (Selva et al. this volume), where a PVHA for ash fall at Campi Flegrei Italy is reported.

2 The Bayesian Event Tree Scheme for long-term volcanic hazard

BET_VH scheme for long-term volcanic hazard is a natural evolution of the short-term eruption forecasting code BET_EF described in Marzocchi et al. (2008) and devoted to eruption forecasting purposes. Here, the event tree is expanded to consider also the probability of occurrence of the typical hazardous phenomena accompanying eruptions and impacting the territory, such as lava flows, pyroclastic density currents, ash fall, lahars, tsunami and so on. Its basic structure can be described as follows (Fig. 1a):

- Node 1-2-3: there is an eruption, or not, in the time interval $(t_0, t_0 + \tau]$, where t_0 is the present time, and τ is the forecasting time window. This node condenses the probabilities of nodes 1 (unrest), 2 (presence of magma given an unrest) and 3 (eruption given a magmatic unrest) in BET_EF by

Marzocchi et al. (2008) as regards the non-monitoring part.

- Node 4: the eruptive vent will open in a specific location, provided there is an eruption.

- Node 5: the eruption will be of a certain size or type, provided that there is an eruption in a given location.

- Node 6: a particular hazardous phenomenon will be generated, or not, given that an eruption of a specific size or type occurs. Several hazardous phenomena can be uploaded in BET_VH at a time; volcanic hazard due to different phenomena can be juxtaposed and compared.

- Node 7: a selected area around the volcano will be reached, provided the occurrence of a specific hazardous phenomenon produced by an eruption of given size or type and location.

- Node 8: a specific intensity threshold will be overcome, or not, provided that the area selected at node 7 is reached by a specific hazardous phenomenon generated by an eruption with a given size or type and location.

Note that since the definition of event tree is mainly driven by its practical utility, the branches at each node point to the whole set of different possible events, regardless of their probabilistic features. In other words, the events at each node need not be mutually exclusive (see upper part of figure 1). This makes the combination of nodes a little bit more complicated but it keeps a more logical and comprehensible structure (see Marzocchi et al. 2004, 2006, 2008 for more details).

In the following section, we describe how the probability distribution at each node is assigned. We will try to make use of the same symbols and terminology already published in Marzocchi et al. (2008) and in the Electronic Supplementary Material of that paper.

3 Estimating the probability distribution at each node

BET_VH focusses on long-term PVHA only. Because of this choice, the information to be used is related to geological and/or physical models, and past data. We do not account for monitoring measures. The latter, in fact, are related to short-term variations in the state of the volcano. This is one of the basic differences with BET_EF by Marzocchi et al. (2008), and it implies a simpler formalism for the probability computation at the nodes of the event tree in common with BET_EF. On the other hand, BET_VH has a more complex structure than BET_EF, because it accounts also for the impact of hazardous phenomena on the territory; this leads to a more complex dependence on the selected path, i.e., on the events selected at the previous nodes. In order to keep the notation as simple as possible, hereinafter we set specific indexes for specific selected outcomes at the different nodes: i indicates the vent location, j the eruption size or type, p the hazardous phenomenon, k the area, and s the threshold related to the phenomenon p. Despite the description of the following nodes should be self consistent, we recommend to refer to Marzocchi et al. (2008) for more detailed description of the statistical distributions used (Beta and Dirichlet distributions).

3.1 Node 1-2-3

This node has two possible outcomes: *eruption* or *no eruption*, in the time interval $(t_0, t_0 + \tau]$. A suitable time window for long-term hazard purposes is $\tau = 1$ year. The long-term probability of eruption is estimated similarly to Marzocchi et al. (2008) for the non-monitoring probability. Here BET_VH condenses the first three nodes of that tree in one. In particular, BET_VH assigns a prior Beta distribution for the two possible outcomes:

$$[\theta_{1-2-3}]_{prior} = \text{Beta}(\alpha_1, \beta_1) \tag{1}$$

where the parameters α_1 and β_1 are determined from the average Θ_1 (i.e., the best guess for the long-term probability of eruption) and from the equivalent number of data, Λ_1, by inverting the system of equations (see Marzocchi et al. 2008 for more details):

$$\begin{cases} \Theta_1 = \frac{\alpha_1}{\alpha_1+\beta_1} \\ \Lambda_1 = \alpha_1 + \beta_1 - 1 \end{cases} \tag{2}$$

While Θ_1 comes from models and/or *a priori* considerations, Λ_1 is set according to the degree of confidence that the user puts on the estimate of the best guess. The parameter Λ_i (here the index i is relative to the node considered) is a friendly measure of the confidence on the prior distribution, or, in other terms, of the epistemic uncertainty. In general, the higher the Λ_i, the larger our confidence on the reliability of the model, so that the number of past data needed to modify significantly the prior must be larger. On the contrary, if we believe that the prior is poorly informative (i.e., our a priori information is very scarce), Λ_i must be small, so that even a small number of past data can drastically modify the prior (for more information see Marzocchi et al. 2008).

In case a catalog of past eruptions is available, BET_VH updates the prior distribution into the posterior through:

$$[\theta_{1-2-3}] = [\theta_{1-2-3}|y_{1-2-3}] = \mathrm{Beta}(\alpha_1 + y_{1-2-3}, \beta_1 + n_{1-2-3} - y_{1-2-3}) \qquad (3)$$

where y_{1-2-3} represents the number of past eruptions having occurred, and n_{1-2-3} represents the total number of not overlapping time windows of length τ (for example, years) available in the catalog, starting without unrest or eruption (see Electronic Supplementary Material in Marzocchi et al. 2008).

3.2 Node 4

At this node, BET_VH estimates the spatial probability of vent opening, given that an eruption occurs. The basic assumption here is that only one vent at a time will erupt. Because of this, this node has I_4 possible and mutually exclusive outcomes, corresponding to the number of possible vent locations defined by the user.

The probability estimation at this node is identical to the non-monitoring case in Marzocchi et al. (2008). In practical terms, BET_VH assigns a prior Dirichlet distribution for the I_4 vent locations:

$$[\theta_4]_{prior} = \mathrm{Di}_{I_4}(\alpha_4^{(1)}, ...\alpha_4^{(I_4)}) \qquad (4)$$

where each $\alpha_4^{(i)}$ (i=1,...,I_4) is determined (as in Marzocchi et al. 2008) on the basis of $\Theta_4^{(i)}$ (i.e., the expected value, or best guess, of the long-term probability of eruption in vent location i) and of Λ_4 (i.e., the equivalent number of data), set similarly to node 1-2-3. The parameters $\alpha_4^{(i)}$ are obtained by inverting the system of equations (see Marzocchi et al. 2008 for more details):

$$\begin{cases} \Theta_4^{(1)} = \frac{\alpha_4^{(1)}}{\sum_{i=1}^{I_4} \alpha_4^{(i)}} \\ ... \\ \Theta_4^{(I_4)} = \frac{\alpha_4^{(I_4)}}{\sum_{i=1}^{I_4} \alpha_4^{(i)}} \\ \Lambda_4 = \sum_{i=1}^{I_4} \alpha_4^{(i)} - I_4 + 1. \end{cases} \tag{5}$$

In case a catalog of past data is available, BET_VH updates the prior distribution into the posterior through:

$$[\theta_4] = [\theta_4 | y_4] = \mathrm{Di}_{I_4}(\alpha_4^{(1)} + y_4^{(1)}, ..., \alpha_4^{(I_4)} + y_4^{(I_4)}) \tag{6}$$

where $y_4^{(i)}$ represents the number of past eruptions occurred in vent location i.

3.3 Node 5

Here, we examine the probabilities related to the size or type of the eruption. The magnitude can be represented either by the type of the eruption (explosive, effusive, phreatomagmatic, and so on), or by the size (e.g., VEI), or by groups of types or sizes that for the user's purposes are considered homogeneous (e.g., VEI $\geq$4). In the following, we use the term *size–class*, meaning anyone of these parametrizations. Thus, the number of possible outcomes at this node is J_5, corresponding to the number of possible size–classes defined.

At this node we estimate the probability of a specific size–class of the eruption, given its occurrence and given that the vent opens in a specific location. Here, we propose a substantial improvement over Marzocchi et al. (2008), because in BET_VH (and in the version 2.0 of BET_EF that can be downloaded from the website: http://www.bo.ingv.it/bet), we allow the probability of each size–class to vary as a function of the vent location. In this way it is possible to account for potential differences among vent locations. For example, for a certain volcano we might want to assign a higher probability of a hydromagmatic type of eruption if the vent is located under shallow water than on land. Similarly, with this improvement we can take into account flank instability in case the vent opens in a radial sector of a central, steep flank volcano rather than in a central crater. Note that J_5 does not depend on vent location, since only the probabilities of the specific size–classes do.

The computation of the prior and posterior distributions is similar to the one at node 4, except for the fact that each vent location has its own distribution that might be different from those relative to other locations. If the user defines J_5 size–classes, BET_VH assigns a prior Dirichlet distribution for the size–classes at the i-th vent location:

$$[\theta_{5;i}]_{prior} = \mathrm{Di}_{J_5}(\alpha_{5;i}^{(1)}, ...\alpha_{5;i}^{(J_5)}) \tag{7}$$

where each $\alpha_{5,i}^{(j)}$ (j=1,...J_5) is determined (as for node 4, eq. 5) on the basis of $\Theta_{5;i}^{(j)}$ (the best guess of the long-term probability of an eruption having the j-th size–class in the i-th location), and $\Lambda_{5;i}$ (the equivalent number of data for the i-th location). These parameters are set similarly to node 1-2-3. Note that the parameter $\Lambda_{5;i}$ can differ from vent locations, since we may be more confident on the best guess related to a vent location than to another.

In case a catalog of past data is available, BET_VH updates the prior distribution for the i-th vent location into the posterior through:

$$[\theta_{5;i}] = [\theta_{5;i}|y_{5;i}] = \mathrm{Di}_{J_5}(\alpha_{5;i}^{(1)} + y_{5;i}^{(1)}, ...\alpha_{5;i}^{(J_5)} + y_{5;i}^{(J_5)}) \tag{8}$$

where $y_{5;i}^{(j)}$ represents the number of past eruptions of size–class j occurred in vent location i.

3.4 Node 6

For a specific hazardous phenomenon p, this node has two possible outcomes: *generation* or *no generation* of the phenomenon from an eruption of size–class j. The generation of a specific phenomenon depends only on the size–class, and not on the vent location. The prior distribution is set, similarly to node 1-2-3, as

$$[\theta^{(p)}_{6;j}]_{prior} = \mathrm{Beta}(\alpha^{(p)}_{6;j}, \beta^{(p)}_{6;j}) \tag{9}$$

where the parameters $\alpha^{(p)}_{6;j}$ and $\beta^{(p)}_{6;j}$ are determined (as for node 1-2-3, eq. 2) on the basis of $\Theta^{(p)}_{6;j}$ (i.e., the best guess for the long-term probability of generation of the phenomenon from a size–class j eruption) and of $\Lambda^{(p)}_{6;j}$ (i.e., the equivalent number of data). The parameters $\Theta^{(p)}_{6;j}$ and $\Lambda^{(p)}_{6;j}$ are set similarly to node 1-2-3. Note that the apex p on θ is necessary because we have different (and independent) distributions depending on the phenomenon considered.

In case a catalog of past data is available, BET_VH updates the prior distribution into the posterior through:

$$[\theta^{(p)}_{6;j}] = [\theta^{(p)}_{6;j}|y^{(p)}_{6;j}] = \mathrm{Beta}(\alpha^{(p)}_{6;j} + y^{(p)}_{6;j}, \beta^{(p)}_{6;j} + n_{6;j} - y^{(p)}_{6;j}) \tag{10}$$

where $y^{(p)}_{6;j}$ represents the number of past eruptions of size–class j that generated the phenomenon p, and $n_{6;j}$ represents the total number of past eruptions of size–class j available in the catalog.

3.5 Nodes 7 and 8

These two nodes have a similar structure. First, the surrounding of the volcano is divided into a number (K_7) of areas (not necessarily equal and equally spaced). For each area k, at node 7 we have two possible outcomes: *area k is reached* or *area k is not reached* by the hazardous phenomenon p generated by an eruption of size–class j and location i. At node 8, the two outcomes are: *the selected threshold s is overcome* or *the selected threshold s is not overcome*, considering an area k reached by the hazardous phenomenon p, generated from an eruption of size–class j and location i. Both probabilities are assumed to be homogeneous all over the area k.

Each prior distribution is respectively set as

$$[\theta^{(k)}_{7;i,j,p}]_{prior} = \mathrm{Beta}(\alpha^{(k)}_{7;i,j,p}, \beta^{(k)}_{7;i,j,p}) \tag{11}$$

and

$$[\theta^{(s)}_{8;i,j,p,k}]_{prior} = \mathrm{Beta}(\alpha^{(s)}_{8;i,j,p,k}, \beta^{(s)}_{8;i,j,p,k}) \tag{12}$$

where the parameters α_* and β_* are determined (as for node 1-2-3, eq. 2) by the correspondent Θ_* (i.e., the best guess) and Λ_* (i.e., the equivalent number of data) as for the previous nodes. The use of the Beta distribution for these nodes implies that each area is independent from the others; this assumption requires some additional comments. Let us consider node 7; obviously, the probability that a specific area will be hit by an hazardous phenomenon is somehow correlated to the probability on adjacent cells. This sort of dependency is accounted for, in the prior distribution, by the parameters Θ (the best guesses) for every area that are set by a model. In other words, a reliable model will produce probability estimations that are strongly spatially correlated (see the companion paper Selva et al. this volume), and therefore also the Θ values for adjacent areas will be correlated. The characteristics of such a correlation are set by the physics of the model used. The use of a Beta distribution for each area implies some sort of independence also among adjacent areas. In particular, the average probability of hitting an area will be very similar to the average probability of hitting an adjacent area, but each single probability value is sampled from the Beta distribution independently from the sampled values in adjacent areas. The rationale of this choice is mostly technical, because it greatly simplifies the computations of the code; nevertheless, this choice does not introduce any bias in the final assessment, because the BET_VH code never combines the probabilities of different areas.

In case a catalog of past data is available, BET_VH updates the prior distributions into the posterior respectively through:

$$
\begin{aligned}
&[\theta^{(k)}_{7;i,j,p}] = [\theta^{(k)}_{7;i,j,p}|y^{(k)}_{7;i,j,p}] = \\
&\quad = \mathrm{Beta}(\alpha^{(k)}_{7;i,j,p} + y^{(k)}_{7;i,j,p}, \beta^{(k)}_{7;i,j,p} + n_{7;i,j,p} - y^{(k)}_{7;i,j,p})
\end{aligned} \tag{13}
$$

and

$$
\begin{aligned}
&[\theta^{(s)}_{8;i,j,p,k}] = [\theta^{(s)}_{8;i,j,p,k}|y^{(s)}_{8;i,j,p,k}] = \\
&\quad = \mathrm{Beta}(\alpha^{(s)}_{8;i,j,p,k} + y^{(s)}_{8;i,j,p,k}, \beta^{s)}_{8;i,j,p,k} + n_{8;i,j,p,k} - y^{(s)}_{8;i,j,p,k})
\end{aligned} \tag{14}
$$

where y_* represents the number of successes (reaching area k for node 7, overcoming threshold s for node 8) and n_* represents the total number of past data (past eruptions of size–class j and vent location i generating the hazardous phenomenon p for node 7, and number of such eruptions in which it reached area k for node 8).

4 Estimating PVHA

In the previous chapters we have described the general features of the probability distributions at each node. Their combination allows a full and complete PVHA to be determined. In order to accomplish that, we have to still explore in detail three issues: 1) how model output can be used to set prior distributions for the nodes described before; 2) how to combine the conditional probability at each node to get absolute probabilities; 3) how to account for different ESs. The following three subsections are devoted to describe these three issues. The last subsection reports the kind of outputs provided by the code.

4.1 Numerical models to define prior distribution

As we have seen so far, the setup of prior probability distributions is mainly based on models, through the best guess Θ_* and the number of equivalent data Λ_* that represents a sort of confidence on our best guess (see Marzocchi et al. 2008). For the prior distribution of nodes 1 to 5, we refer to the estimation of the non-monitoring part in Marzocchi et al. (2008). For nodes 6, 7 and 8, we can use results from numerical models that are available for most of the hazardous phenomena related to volcanic eruptions (e.g., Favalli et al. 2005, for lava flows; Neri et al. 2007 for pyroclastic density currents; Pfeiffer et al. 2005 and Costa et al. 2006 for ash dispersion). For a more detailed discussion about this point see Selva et al. (this volume). Here, we highlight the basic philosophy behind models' usage in PVHA.

A single realization of a model very rarely represents a reliable forecast of the future activity. Several factors, acting simultaneously or separately, are responsible for this lack of determinism:

1. intrinsic stochasticity of the process (the so-called *aleatory* uncertainty);

2. *epistemic* uncertainty in the model parameters and in the boundary conditions at the time and during the eruption (e.g., for an ash fall model, the uncertainties in wind conditions);

3. epistemic uncertainty in the (volcanological) input parameters (e.g., for an

ash fall model, the uncertainties in the relevant eruption parameters given a specific eruption size–class);

4. any model is always a simplification of the reality, leading to unavoidable uncertainties into the forecasting.

These issues have a different impact depending on the model that we are considering. The compelling necessity to include all sources of uncertainties requires to explore the whole range of possible variations in the relevant parameters and conditions. In practice, this need calls for the use of a model able to run thousands of times in a reasonable CPU time, accounting for many different sets of initial/boundary conditions and model parameters realizations. Usually, faster models are simpler than more complex models. Anyway, we argue that errors due to the use of a simple model (see point 4 above) are often smaller than errors introduced by uncertainties in model parameters and initial and/or boundary conditions (points 1, 2, and 3; see also Grezio et al. 2010).

If different models can be used, BET_VH may use the output of each one of them recursively. In the companion paper (Selva et al. this volume), we propose a general scheme to introduce models' results from a large number of runs and from more than one model, presenting also a practical application for tephra fall hazard estimation around the Campi Flegrei caldera.

4.2 Combination of the probabilities to obtain PVHA

The probability distributions at each node are conditional on the selection of a well defined path (see section 2). In general, PVHA, as well as the evaluation of probability of each event for which we may be interested, requires their combination. With BET_VH it is possible to compute the probability associated to a single ES (an eruption of size–class and location specified) or to a combination of possible ESs. In the latter case, at nodes 4 and 5 we can select more than one branch at the same time (i.e., a set $\mathcal{J}$ of J possible eruption size–classes, and/or a set $\mathcal{I}$ containing I possible vent locations; see also figure 1, (panel b) for a snapshot of the main window of BET_VH code, where it is possible to see that more than one location and/or size–class can be selected). This feature of the code is quite remarkable because these new combined probabilities are usually very important for practical purposes and for a full PVHA (Selva et al. this volume). In the following, we will carefully describe how to obtain meaningful probabilities for ESs combinations.

4.2.1 Absolute probability

In order to compute the absolute probability Φ of an outcome at node m, we multiply the probabilities along all the selected path, from node 1-2-3 to the selected outcome at node m (see Marzocchi et al. 2008). For example, the long-term absolute probability $[\Phi_a]$ of an area k being impacted by a specific hazardous phenomenon p overcoming the selected threshold s, generated by an eruption of size–class j occurring in location i (i.e., a single ES), in any time window of duration τ in the future, is

$$[\Phi_a] = [\theta_{1-2-3}][\theta_4^{(i)}][\theta_{5;i}^{(j)}][\theta_{6;j}^{(p)}][\theta_{7;i,j,p}^{(k)}][\theta_{8;i,j,p,k}^{(s)}] \tag{15}$$

while the long-term absolute probability $[\Phi_b]$ of an area k being impacted by the same phenomenon p overcoming the selected threshold s, associated to a combination of possible ESs (see above) in any time window of duration τ in the future, is

$$[\Phi_b] = [\theta_{1-2-3}] \sum_{i \in \mathcal{I}} \left([\theta_4^{(i)}] \sum_{j \in \mathcal{J}} [\theta_{5;i}^{(j)}][\theta_{6;j}^{(p)}][\theta_{7;i,j,p}^{(k)}][\theta_{8;i,j,p,k}^{(s)}] \right) \tag{16}$$

Note that the latter estimation is particularly important in practical applications because the risk in a specific area depends on the phenomenon reaching that area, not the specific characteristic of the eruption that creates the phenomenon.

4.2.2 Conditional probability

Beyond the absolute probabilities, in many practical applications some conditional probabilities are particularly important and useful, like the conditional probabilities that can be obtained by the combination of different ESs (see Selva et al. this volume). To this purpose, the code BET_VH gives the possibility to average, with proper weights, the conditional probabilities for different size–classes and/or locations.

Here, we describe how BET_VH calculates these new conditional probabilities ($[\phi]$) starting from the conditional probability distributions at each node. In order to distinguish clearly these new conditional probabilities from the conditional probabilities of the nodes, we decide to indicate them with the symbol ϕ instead of θ. Let us consider all the possibilities offered by the code. For node 5 (size–classes), the probability $[\theta_{5;i}^{(j)}]$ is by definition the conditional probability of an eruption of size–class j, given the occurrence of an eruption at a specific location i. If a set $\mathcal{I}$ of I possible vent locations is considered, then the conditional probability $[\phi_a]$ of an eruption of size–class j occurring in a location belonging to the set $\mathcal{I}$, given an eruption occurs, becomes

$$[\phi_a] = \frac{\sum_{i \in \mathcal{I}} \left([\theta_4^{(i)}][\theta_{5;i}^{(j)}] \right)}{\sum_{i \in \mathcal{I}} [\theta_4^{(i)}]} \tag{17}$$

If a set of size–classes is selected at the same time, the resulting conditional probability is the algebraic sum of the conditional probability of the single size–classes.

For nodes 6, 7 and 8, the conditional probabilities are always conditioned to the ES chosen (single or combined). We define the conditional ES probability $[\phi_{ES}]$ (given an eruption occurs) as:

$$[\phi_{ES}] \equiv [\theta_{\{i\in\mathcal{I},j\in\mathcal{J}\}}] = \sum_{i\in\mathcal{I}}\left([\theta_4^{(i)}]\sum_{j\in\mathcal{J}}[\theta_{5;i}^{(j)}]\right) \tag{18}$$

The conditional ES probability $[\phi_{ES}]$ represents the probability of occurrence of the selected single or combined ES, given an eruption occurs. This probability is particularly important, because it gives the "weight " to every possible ES selected; this is necessary either to compare or to merge them (see Selva et al. this volume, for more details).

With this definition, we have

- at node 6, the conditional probability $[\phi_b]$ of a specific phenomenon p, given the occurrence of an eruption within the selected combination of ESs is

$$[\phi_b] = \frac{\sum_{i\in\mathcal{I}}\left([\theta_4^{(i)}]\sum_{j\in\mathcal{J}}[\theta_{5,i}^{(j)}][\theta_{6,j}^{(p)}]\right)}{\phi_{ES}} \tag{19}$$

- at node 7, the conditional probability $[\phi_c]$ of a specific area k being reached by the phenomenon p, given the occurrence of an eruption within the selected combination of ESs is

$$[\phi_c] = \frac{\sum_{i\in\mathcal{I}}\left([\theta_4^{(i)}]\sum_{j\in\mathcal{J}}[\theta_{5,i}^{(j)}][\theta_{6,j}^{(p)}][\theta_{7,i,j,p}^{(k)}]\right)}{\phi_{ES}} \tag{20}$$

- at node 8, the conditional probability $[\phi_d]$ of overcoming the selected threshold s in a specific area k reached by the phenomenon p, given the occurrence of an eruption within the selected combination of ESs is

$$[\phi_d] = \frac{\sum_{i\in\mathcal{I}}\left([\theta_4^{(i)}]\sum_{j\in\mathcal{J}}[\theta_{5,i}^{(j)}][\theta_{6,j}^{(p)}][\theta_{7,i,j,p}^{(k)}][\theta_{8,i,j,p,k}^{(s)}]\right)}{\phi_{ES}} \tag{21}$$

In analogy to what said for node 5, such new combined conditional probabilities (eqs. 19, 20 and 21) represent conditional probabilities averaged for every possible ES; in other words, the conditional probabilities for each ES are summed up weighting them with the ES probability. In case of a single ES (i-th vent location, j-th size–class), $\mathcal{I} = \{i\}$ and $\mathcal{J} = \{j\}$.

4.3 PVHA output maps

As we have seen, BET_VH code allows any conditional or absolute probability of interest to be computed. Some probability maps have usually a practical immediate usefulness. Because of this, BET_VH is able to display them in gif or GoogleEarth format, or to export them in a raster file that can be uploaded and visualized in a GIS program. Such probability maps are:

- Spatial probability of vent opening (conditional on eruption occurrence), which is related to the same probability of the susceptibility map proposed by Felpeto et al. (2007); see figure 2 for a snapshot example from the code.
- Absolute spatial probability of a specific size–class or a specific phenomenon (in the code, it is called *Absolute map*); this map shows, for each area, the absolute probability of a specific eruption size–class, or the absolute occurrence probability of a specific phenomenon; see figure 3 for a snapshot example from the code.
- ES conditional probability (conditional on eruption occurrence): this map (called *Sizes map* in the code) displays, for each area (or group of areas), the probability of a specific eruption size–class, weighted with the spatial probability of vent opening, conditional on eruption occurrence in that area

(or group of areas). In practice, if one class is selected, the map displays $[\phi_a]$ (see equation 17); if more size–classes are selected, the map shows the sum of $[\phi_a]$ for all size–classes belonging to $j \in \mathcal{J}$. Figure 4 displays a snapshot example of the code.

- Absolute or conditional probability of the phenomenon impacting the territory, in terms of "reaching" proximal or distal areas (node 7; $[\phi_c]$), or in terms of overcoming a specific threshold (node 8; $[\phi_d]$) in proximal or distal areas; this is called *Outcome map* in the code; see figure 5 for a snapshot example from the code.

The probabilities reported in each map are represented as distributions. The average value represents the "best guess" of such a probability; the dispersion around it gives the uncertainty about this guess.

5 Discussion and Final Remarks

This paper aims at introducing a new statistical tool, named BET_VH to calculate and visualize the long-term probabilistic volcanic hazard assessment (PVHA). BET_VH has some paramount features:

- it provides PVHA in a user-friendly and transparent way (it is not a black box);

- it estimates almost all probabilities useful for hazard and risk applications. In particular, it calculates and visualizes ES maps, as well as weighted combinations of all possible ESs;

- it is based on the Bayesian approach; this enables us to take into account different sources of information, such as models output, field data, and relevant geological and historical information. Moreover, the Bayesian approach allows aleatory and epistemic uncertainties to be estimated and visualized;
- the outputs are provided in different formats (maps in GoogleEarth, GIS, gif formats), in order to make easy the use of the results.

In the companion paper (Selva et al. this volume), an extensive application of BET_VH to ash hall hazard at Campi Flegrei, Italy, is reported.

6 Acknowledgements

Laura Sandri wishes to thank Alex Garcia-Aristizabal for his help with the figures of this paper. We also thank Thea Hincks whose suggestions help us to improve the manuscript.

7 References

Cioni R, Longo A, Macedonio G, Santacroce R, Sbrana A, Sulpizio R, Andronico D (2003) Assessing pyroclastic fall hazard through field data and numerical simulations: Example from Vesuvius. J Geophys Res 108:2063. doi:10.1029/2001JB000642

Costa A, Macedonio G, Folch A (2006) A three-dimensional Eulerian model for transport and deposition of volcanic ashes. Earth Planet Sci Lett 241:634-647. doi:10.1016/j.epsl.2005.11.019

Favalli M, Pareschi MT, Neri A, Isola I (2005) Forecasting lava flow paths by a stochastic approach. Geophys Res Lett 32:3. doi:10.1029/2004GL021718

Felpeto A, Marti J, Ortiz R (2007) Automatic GIS-based system for volcanic hazard assessment. J Volcanol Geotherm Res 166:106-116. doi:10.1016/j.jvolgeores.2007.07.008

Grezio A, Marzocchi W, Sandri L, Gasparini P (2010) A Bayesian procedure for probabilistic tsunami hazard assessment. Nat Haz (in press). doi:10.1007/s11069-009-9418-8

Ho C-H, Smith EI, Keenan DL (2006) Hazard area and probability of volcanic disruption of the proposed high-level radioactive waste repository at Yucca Mountain, Nevada, USA. Bull Volcanol 69:117123. doi:10.1007/s00445-006-0058-5.

Jaquet O, Connor C, Connor L (2008) Probabilistic methodology for long-term assessment of volcanic hazards. Nucl Technol 163:180-189.

Macedonio G, Costa A, Folch A (2008) Ash fallout scenarios at Vesuvius: Numerical simulations and implications for hazard assessment. J Volcanol Geoth Res (in press). doi: 10.1016/j.jvolgeores.2008.08.014

Magill CR, Hurst AW, Hunter LJ, Blong RJ (2006) Probabilistic tephra fall simulation for the Auckland Region, New Zealand. J Volcanol Geoth Res 153:370-386

Marti J, Aspinall WR, Sobradelo R, Felpeto A, Geyer A, Ortiz R, Baxter P, Cole PD, Pacheco J, Blanco MJ, Lopez C (2008) A long-term volcanic hazard event tree for Teide-Pico Viejo stratovolcanoes (Tenerife, Canary Islands). J Volcanol Geoth Res 178:543-552. doi: 10.1016/j.jvolgeores.2008.09.023

Martin AJ, Umeda K, Connor CB, Weller JN, Zhao D, Takahashi M (2004) Modeling long-term volcanic hazards through Bayesian inference: An example from the Tohoku volcanic arc, Japan. J Geophys Res 109:B10208. doi:10.1029/2004JB003201

Marzocchi W (1996) Chaos and stochasticity in volcanic eruptions: the case of Mt. Etna and Vesuvius. J Volcan and Geoth Res 70:205-212.

Marzocchi W, Sandri L, Gasparini P, Newhall C, Boschi E (2004) Quantifying probabilities of volcanic events: the example of volcanic hazard at Mt Vesuvius. J Geophys Res 109:B11201. doi:10.1029/2004JB003155

Marzocchi W, Sandri L, Furlan C (2006) A quantitative model for volcanic hazard assessment. In: Mader HM, Coles SG, Connor CB, Connor LJ (eds) Statistics in Volcanology. IAVCEI publications, pp. 31-37.

Marzocchi W, Woo G (2007) Probabilistic eruption forecasting and the call for an evacuation. Geophys Res Lett 34:L22310. doi:10.1029/2007GL031922

Marzocchi W, Neri A, Newhall CG, Papale P (2007) Probabilistic volcanic hazard and risk assessment. EOS Trans AGU 88:318.

Marzocchi W, Sandri L, Selva J (2008) BET_EF: a probabilistic tool for long- and short-term eruption forecasting. Bull Volcanol 70:623-632. doi:10.1007/s00445-007-0157-y

Marzocchi W, Woo G (2009) Principle of volcanic risk metrics: theory and the case study of Mt Vesuvius and Campi Flegrei (Italy). J Geophys Res 114:B03213. doi:10.1029/2008JB005908.

Neri A, Esposti Ongaro T, Menconi G, De'Michieli Vitturi M, Cavazzoni C, Erbacci G, Baxter PJ (2007) 4D simulation of explosive eruption dynamics at Vesuvius. Geophys Res Lett 34:L04309. doi:10.1029/2006GL028597

Neri A, Aspinall WP, Cioni R, Bertagnini A, Baxter PJ, Zuccaro G, Andronico D, Barsotti S, Cole PD, Esposti Ongaro T, Hincks TK, Macedonio G, Papale P, Rosi M, Santacroce R, Woo G (2008) Developing an event tree for probabilistic hazard and risk assessment at Vesuvius. J Volcanol Geoth Res 178:397-415. doi:10.1016/j.jvolgeores.2008.05.014

Newhall CG, Hoblitt RP (2002) Constructing event trees for volcanic crises. Bull Volcanol 64: 3-20. doi:10.1007/s004450100173

Pfeiffer T, Costa A, Macedonio G (2005) A model for the numerical simulation of tephra fall deposits. J Volcanol Geotherm Res 140:273-294. doi:10.1016/j.jvolgeores.2004.09.001

Scandone R, Arganese G, Galdi F (1993) The evaluation of volcanic risk in the

Vesuvian area. J Volcanol Geoth Res 58:263-271.

Selva J, Costa A, Marzocchi W, Sandri L (2010) BET_VH: Exploring the influence of natural uncertainties on long-term hazard from tephra fallout at Campi Flegrei (Italy). Bull Volcanol (this volume).

Sparks RSJ (2003) Forecasting volcanic eruptions. Earth Planet Sci Lett 210:1-15.

Woo G (2008) Probabilistic criteria for volcano evacuation decisions. Nat Haz 45:87-97.

Figure captions

Figure 1: Panel a: General scheme of event tree for BET_VH. The selected path (black solid arrows connecting blue labels) is related to the probability of an eruption in location i, of size–class j (see text for more explanations), generating tephra that is able to reach area k and overcome the threshold set for tephra accumulation; Panel b: a snapshot from the main selection window of BET_VH code for an imaginary volcano called Monte Donato. Here the selection is related to the probability of an eruption in any location, of size–class 2 or greater, generating tephra that is able to reach proximal or distal areas and overcome the threshold of 5 cm.

Figure 2: A snapshot from the BET_VH code applied to Monte Donato, showing the conditional probability of vent opening in every possible vent location (conditioned to the occurrence of an eruption). The figure reports the average of the probability distribution. The table below the figure reports the 10th, 50th (median), and 90th percentiles. The term "vent loc" means vent location, and "perc" means percentile.

Figure 3: A snapshot from BET_VH code applied to Monte Donato, showing the absolute probability of an eruption of size–class (see text for more explanations) 2 or greater, in every possible vent location. The figure reports the average of the probability distribution. The table below the figure reports also the 10th, 50th (median), and 90th percentiles. The term "vent loc" means vent location, and "perc" means percentile.

Figure 4: A snapshot from BET_VH code applied to Monte Donato, showing the conditional probability of an eruption of size–class 2 or greater in every

possible vent location, weighted by the probability of vent opening in each location, given the occurrence of an eruption. Note the difference in palette scale compared to figure 3. The figure reports the average of the probability distribution. The table below the figure reports the 10th, 50th (median), and 90th percentiles. The term "vent loc" means vent location, and "perc" means percentile.

Figure 5: A snapshot from the BET_VH code applied to Monte Donato, showing the outcome map. Here, we display the absolute probability related to the selection of figure 1, panel b, i.e., the absolute probability of overcoming 5 cm of tephra due to an eruption in any possible vent location with size–class (see text for more details) 2 or greater. The probability is computed for different areas around the imaginary volcano. The centers of the areas are marked by white squares. The description of each area can be retrieved by clicking the relative white square as shown in the figure. Note the different spatial scale compared to figures 2, 3 and 4.

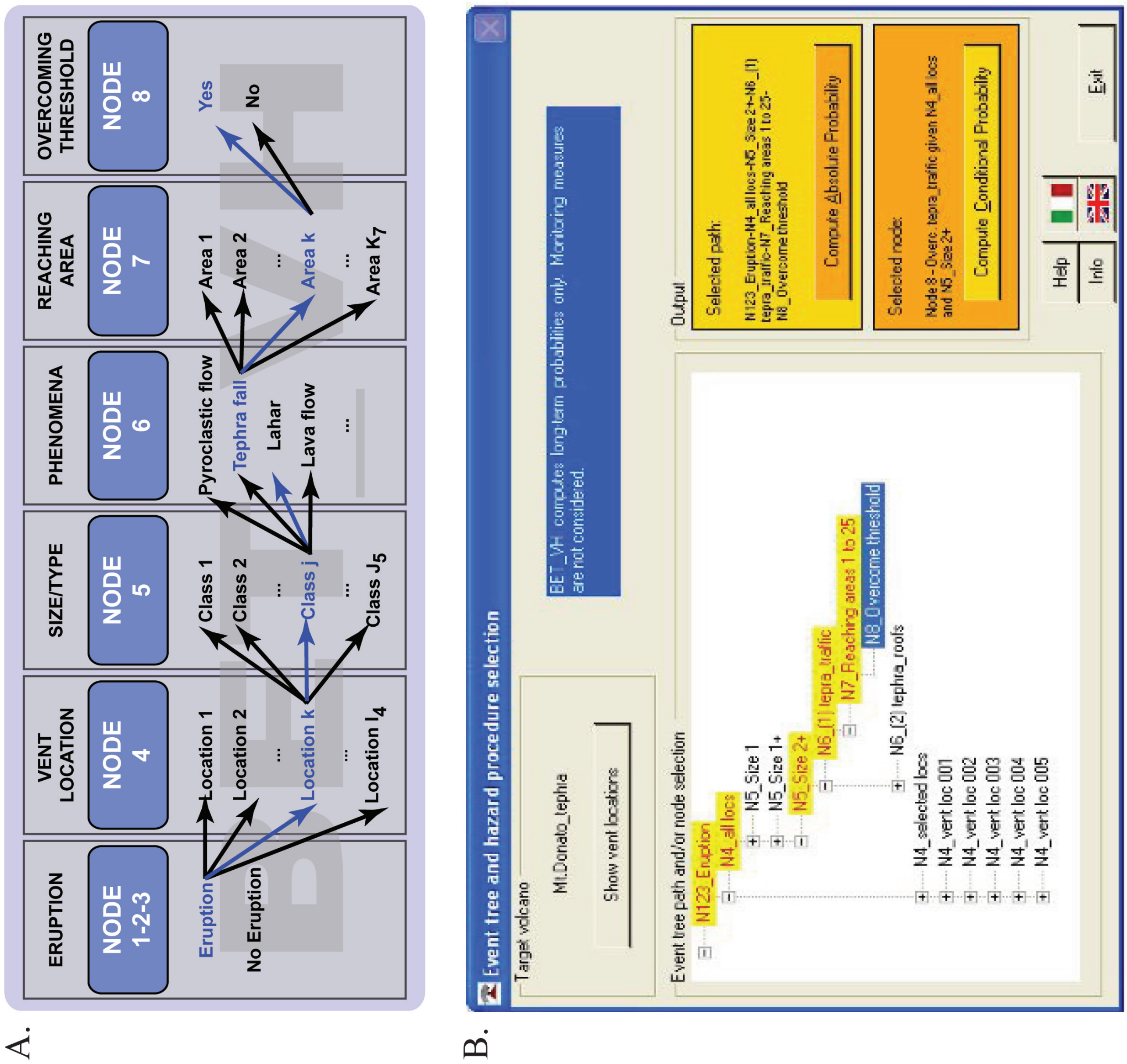
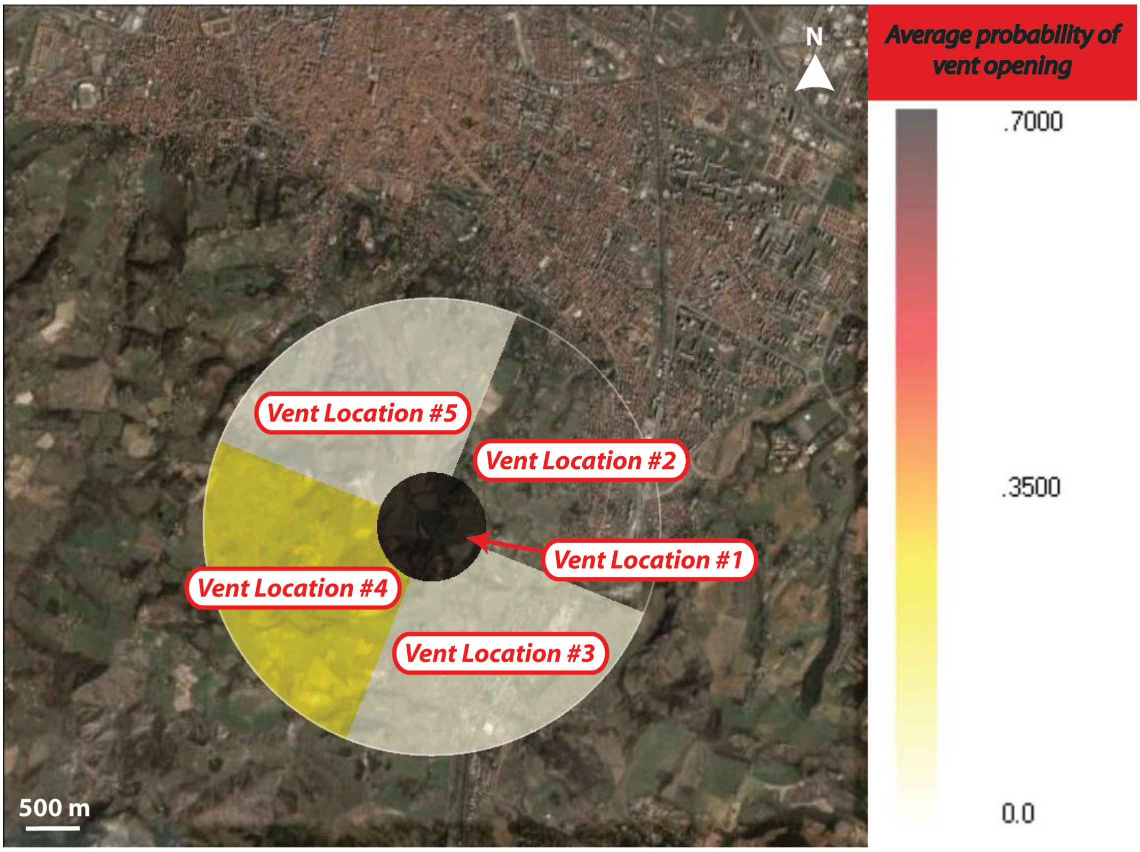
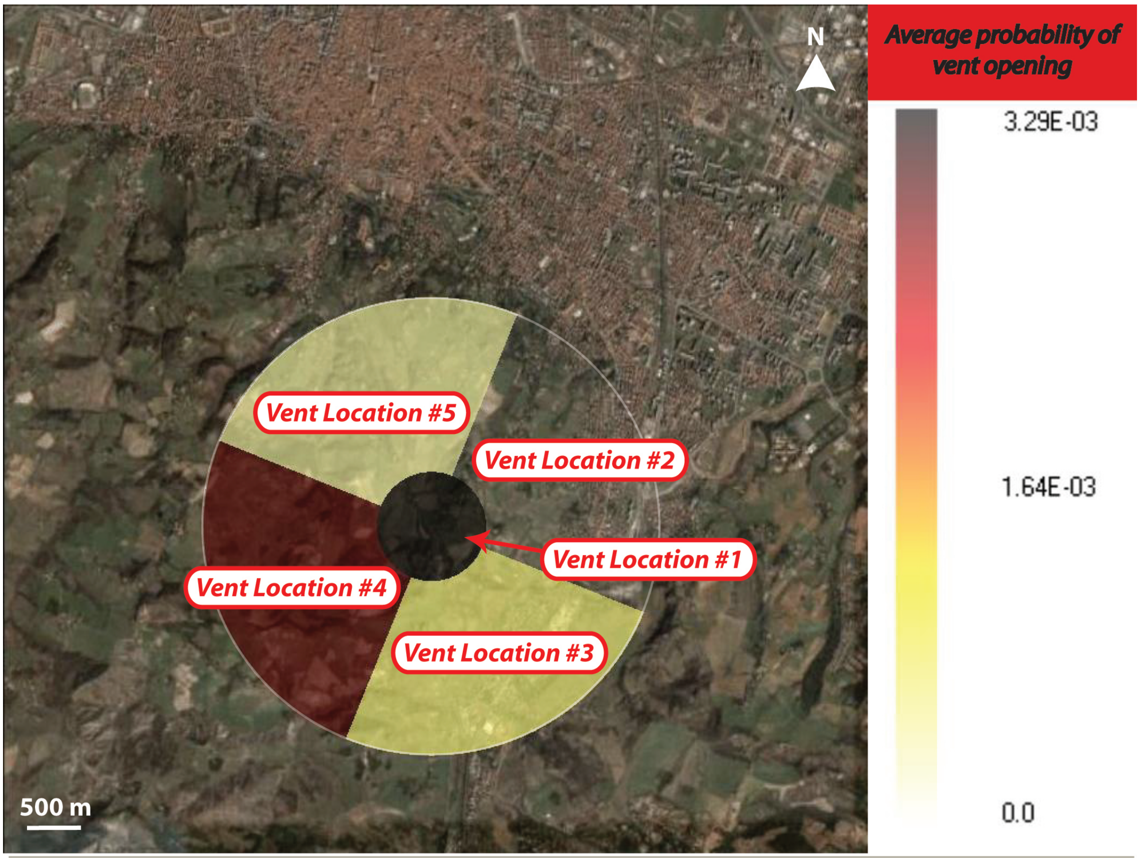

A.
ERUPTION
NODE 1-2-3
Eruption
No Eruption
VENT LOCATION
NODE 4
Location 1
Location 2
Location k
SIZE/TYPE
NODE 5
Class 1
Class 2
Class j
PHENOMENA
NODE 6
Pyroclastic flow
Tephra fall
Lahar
Lava flow
REACHING AREA
NODE 7
Area 1
Area 2
Area k
OVERCOMING THRESHOLD
NODE 8
Yes
No
B.
Event tree and hazard procedure selection
Target volcano
Mt.Donato_tephra
Show vent locations
Event tree path and/or node selection
BET_VH computes long term probabilities only. Monitoring measures are not considered.
Output
Selected path:
Compute Absolute Probability
Selected node:
Compute Conditional Probability
Help
Info
Exit

Figure 1:



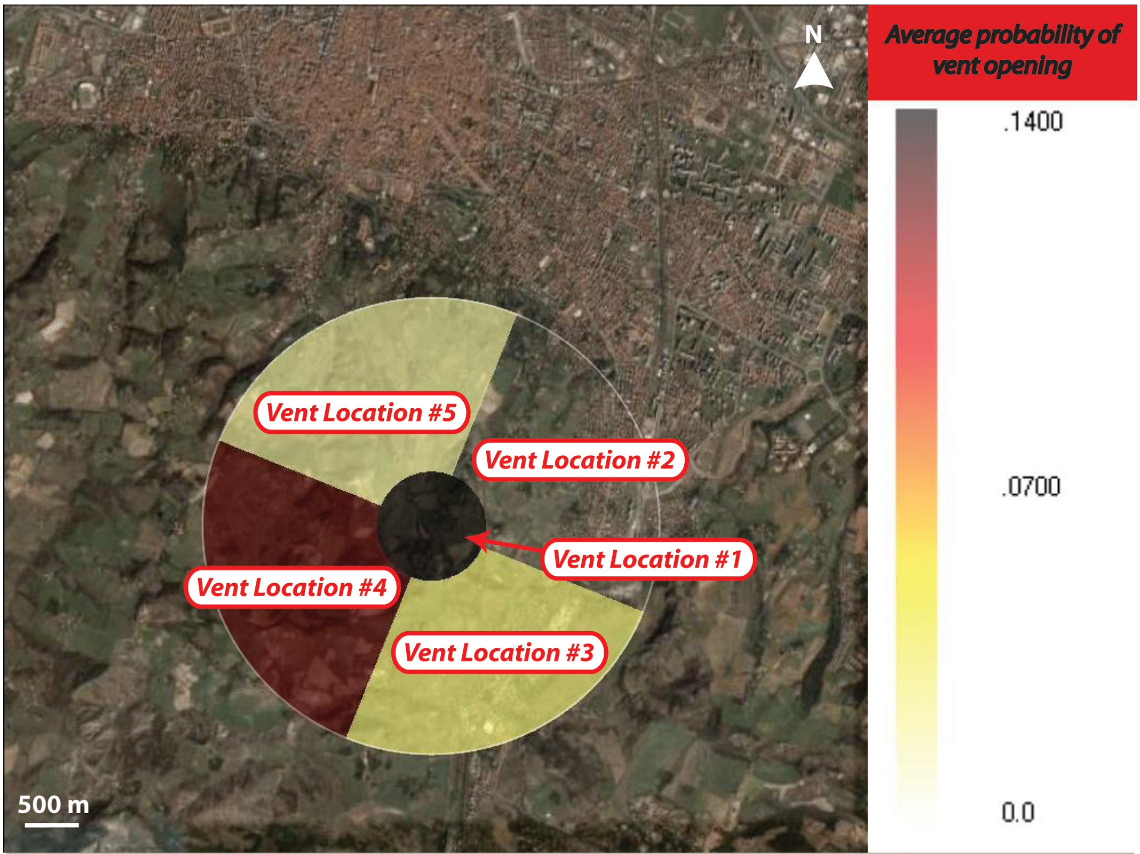
	Average	10th perc	50th perc	90th perc
Vent loc #1	7.00E-01	5.99E-01	7.02E-01	7.99E-01
Vent loc #2	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Vent loc #3	3.00E-02	2.58E-03	2.05E-02	6.82E-02
Vent loc #4	2.50E-01	1.62E-01	2.51E-01	3.46E-01
Vent loc #5	2.00E-02	9.18E-04	1.12E-02	4.02E-02

Figure 2:



	Average	10th perc	50th perc	90th perc
Vent loc #1	3.29E-03	7.36E-04	2.65E-03	6.84E-03
Vent loc #2	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Vent loc #3	5.64E-04	3.91E-05	3.13E-04	1.35E-03
Vent loc #4	2.94E-03	7.53E-04	2.44E-03	5.84E-03
Vent loc #5	3.76E-04	1.22E-05	1.68E-04	8.29E-04

Figure 3:



	Average	10th perc	50th perc	90th perc
Vent loc #1	1.40E-01	4.73E-02	1.25E-01	2.58E-01
Vent loc #2	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Vent loc #3	2.40E-02	2.10E-03	1.63E-02	5.13E-02
Vent loc #4	1.25E-01	4.88E-02	1.23E-01	2.17E-01
Vent loc #5	1.60E-02	6.61E-04	8.92E-03	3.27E-02

Figure 4:

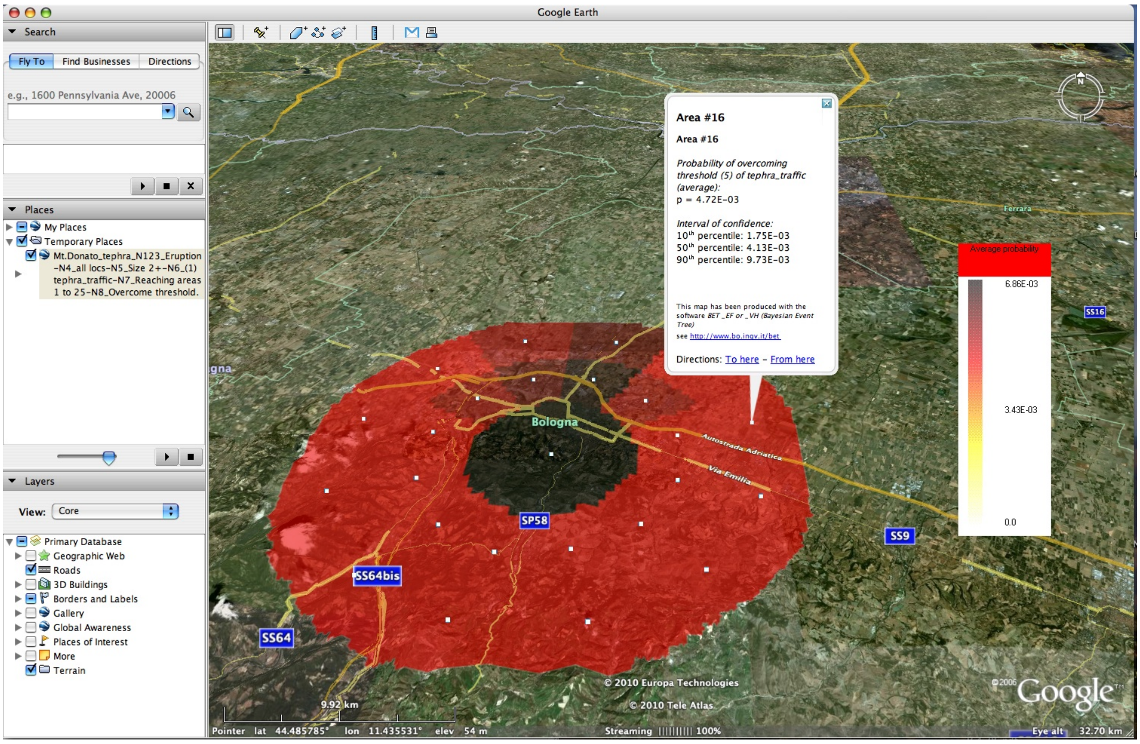

Google Earth
Search
Fly To
Find Businesses
Directions
e.g., 1600 Pennsylvania Ave, 20006
Places
My Places
Temporary Places
Mt.Donato_tephra_N123_Eruption-N4_all locs-N5_Size 2+-N6_(1) tephra_traffic-N7_Reaching areas 1 to 25-N8_Overcome threshold.
Layers
View:
Core
Primary Database
Geographic Web
Roads
3D Buildings
Borders and Labels
Gallery
Global Awareness
Places of Interest
More
Terrain
Area #16
Area #16
Probability of overcoming threshold (5) of tephra_traffic (average):
p = 4.72E-03
Interval of confidence:
10th percentile: 1.75E-03
50th percentile: 4.13E-03
90th percentile: 9.73E-03
This map has been produced with the software BET_EF or _VH (Bayesian Event Tree)
see http://www.bo.ingv.it/bet
Directions: To here - From here
Average probability
6.86E-03
3.43E-03
0.0
Bologna
Autostrada Adriatica
Via Emilia
SP58
SS64bis
SS64
SS9
SS16
Ferrara
© 2010 Europa Technologies
© 2010 Tele Atlas
Google
9.92 km
Pointer lat 44.485785° lon 11.435531° elev 54 m
Streaming 100%
Eye alt 32.70 km

Figure 5: