APPENDIX A
Probability estimation at the nodes: general aspects

In this appendix we report the general features to estimate BET parameters. In particular, we discuss how different kinds of information (monitoring and non-monitoring) are included into the model (section A.1), the main features of the Bayesian inference to estimate the probability at nodes (section A.2), and the characteristics of the fuzzy approach to deal with monitoring measures (section A.3). More details and examples are given by Marzocchi et al. (2004, 2006a).

A1. Accounting for monitoring and non-monitoring information

Generally speaking, we have two broad classes of information that can be considered in EF: measurements from monitoring (dataset $\mathcal{M}$) and all the other kinds of data/information (dataset $\overline{\mathcal{M}}$) . This subdivision is mainly due to the fact that, usually, these two types of information have different weights in different states of a volcano. During an unrest, monitoring data may be the most relevant for EF purposes, while the same data do not carry relevant information about EF during a quiet period, apart from telling that the volcano is at rest. At the same time, it is obvious that monitoring data contain fundamental information that must be used to quantify mid- to short-term volcanic hazard. For these reasons, we introduce these kinds of information through two different functions. At the generic k-th node, the pdf of the j-th event ($[\theta_k^{(j)}]$, see notation used in section 2 and figure 1 of main text) is

$$[\theta_k^{(j)}] = \gamma_k [\theta_k^{(j)\{\mathcal{M}\}}] + (1-\gamma_k)[\theta_k^{(j)\{\overline{\mathcal{M}}\}}] \tag{1}$$

where γ_k is a variable in the interval $[0,1]$, $[\theta_k^{(j)\{\mathcal{M}\}}]$ and $[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]$ have the same meaning as $[\theta_k^{(j)}]$, but they are defined by using only monitoring information and all the other kinds of information (non-monitoring, hereinafter), respectively. In other words, $[\theta_k^{(j)}]$ is a linear combination of the probabilities based on these two types of knowledge, weighted with γ_k and $(1-\gamma_k)$ respectively.

It follows that, the unknowns of BET are the pdfs of the conditional probability of the selected events at the nodes. The evaluation of each one of these probabilities passes through the estimation of the three unknowns in equation 1, i.e., γ_k, $[\theta_k^{(j)\{\mathcal{M}\}}]$, and $[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]$. In particular:

1. The parameter γ_k sets the degree at which monitoring data (useful for short-term EF) control the posterior probabilities with respect to the non-monitoring part (useful for long-term EF); for the nodes where monitoring parameters are informative, γ_k is a function of the "state of unrest" η (see A.3.2 below), which, in turn, is a *fuzzy* parameter (Zadeh, 1965) in the interval $[0,1]$ that indicates the degree at which unrest is detected by the monitoring observations at $t = t_0$ (see below). In practice, through γ_k, BET switches dynamically from long-term (when the volcano is found to be at rest) to short-term (during unrest) probabilities.
2. $[\theta_k^{(j)\{\mathcal{M}\}}]$ is the monitoring part in equation 1, i.e., the leading term in short-term probability evaluation. It is determined through Bayes' rule, which combines estimated probabilities from monitoring measures at time t_0 and monitoring measurements from past episodes of unrest (if any). Here, the present and past monitoring measurements are transformed into probabilities by means of a physical model that depends on the node considered (see Appendix B).
3. $[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]$ is the non-monitoring part in equation 1, i.e., the leading term in long-term probability evaluation. It is determined through Bayes' rule, which combines estimated

probabilities from all our knowledge based on theoretical models and/or beliefs, and past data, i.e., past frequencies of occurrence.

The estimation of these three unknowns requires the use of two important technical concepts, namely the Bayesian inference (see section A.2) and the *fuzzy* approach (Zadeh, 1965; see section A.3). The Bayesian inference is necessary to merge together theoretical models/beliefs with data, and to deal with aleatory and epistemic uncertainties. The fuzzy approach is used to manage monitoring measures into the probability calculations and to define the state of unrest η (see subsection A.3.2). The state of unrest η is used to detect unrest from monitoring measures, and then to define γ_k. The specific estimation of γ_k, $[\theta_k^{(j)\{\mathcal{M}\}}]$, and $[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]$ for each node of BET (equation 1) is reported in Appendix B.

A.2 The Bayesian inference

Bayesian inference is the process of fitting a probability model to a set of data (i.e., past data from monitoring and other sources; y hereinafter), and summarizing the result by a probability distribution of the parameters of the model (i.e., θ_k). In order to make probability statements about θ_k given y, we must use a probabilistic model to link the variables, i.e., the Bayes' rule.

Each function $[\theta_k^{(j)\{\cdot\}}]$ of equation 1 is determined through Bayes theorem (see figure 4). Given a set of data y (monitoring measurements from past episodes of unrest in case of $[\theta_k^{(j)\{\mathcal{M}\}}]$, or past frequencies in case of $[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]$), we have:

$$[\theta_k^{(j)\{\cdot\}}] \equiv [{\theta_k}^{(j)\{\cdot\}}|y] = \frac{[\theta_k^{(j)\{\cdot\}}]_{prior}[y|\theta_k^{(j)\{\cdot\}}]}{[y]} \tag{2}$$

where

- $[\theta_k^{(j)\{\cdot\}}|y]$ is the posterior distribution, given the set of data y.
- $[\theta_k^{(j)\{\cdot\}}]_{prior}$ is the prior distribution containing all our knowledge based only on theoretical models and/or beliefs.
- $[y|\theta_k^{(j)\{\cdot\}}]$ is the sampling distribution (the so-called *likelihood* function), that is the probability distribution of observing the data y given a specific probability of occurrence of the j-th event ($\theta_k^{(j)\{\cdot\}}$) characterizing node k.
- The distribution $[y]$ is a normalizing factor accounting for the total probability of observing data y.

In case no data are available (y is empty), we rely on the prior distribution based on theoretical models and/or theoretical beliefs, if any:

$$[\theta_k^{(j)\{\cdot\}}] = [\theta_k^{(j)\{\cdot\}}]_{prior} \tag{3}$$

Bayes' rule is widely and commonly used in many classical and Bayesian probability calculations. Compared to the most classical applications where the probabilities of equation 2 are single values, here we use statistical distributions of the probabilities in order to account for the uncertainties on the probability estimates (epistemic and aleatory uncertainties; see Marzocchi et al., 2004).

The choice of the functional form of the prior distribution and of the likelihood function represents the core of the Bayesian inference, and it requires some physical and statistical assumptions on the process. Here, for the sake of example, we report the common strategy adopted

for the non-monitoring part of equation 1 that is used for all nodes. This example aims to clarify some concepts of Bayesian inference, and it introduces many conceptual details that will be used in Appendix B. As regards the monitoring part, it is not possible to establish a common procedure of Bayesian inference for each node. The detailed description of the strategy used for each node is reported in Appendix B.

A.2.1 *Prior distribution*

We model the prior distribution for the j-th event at the k- node ($[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]_{prior}$) with a Dirichlet distribution (Marzocchi et al., 2004; Gelman et al. 1995) that, for a generic multivariate random variable $\xi = (\xi^{(1)}, ...\xi^{(J)})$, reads

$$[\xi] = \mathrm{Di}_J(\alpha_1, ...\alpha_J) = \frac{\Gamma(\alpha_1 + ... + \alpha_J)}{\Gamma(\alpha_1)...\Gamma(\alpha_J)}[\xi^{(1)}]^{\alpha_1 - 1}...[\xi^{(J)}]^{\alpha_J - 1} \tag{4}$$

where $\Gamma(\cdot)$ is the Gamma function, $\alpha_j > 0$, $\xi^{(1)}, ..., \xi^{(J)} > 0$, $\sum_{j=1}^{J} \xi^{(j)} = 1$, and J is the number of possible mutually exclusive and exhaustive events, each parameterized by variable $\xi^{(j)}$. In our case, we have

$$[\theta_k^{\{\overline{\mathcal{M}}\}}]_{prior} \equiv \mathrm{Di}_{J_k}(\alpha_k^{(1)}, ..., \alpha_k^{(J_k)}) \tag{5}$$

where J_k is the number of possible mutually exclusive and exhaustive events at the k-th node, and $\alpha_k^{(1)}, ..., \alpha_k^{(J_k)}$ are the parameters of the distribution that depend on the node considered. The probability distribution of the j-th event at node k ($[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]_{prior}$) is the marginal distribution of $\theta_k^{(j)\{\overline{\mathcal{M}}\}}$ in equation 5.

Since the random variable is a probability, the Dirichlet distribution is particularly suitable being unimodal and with domain $[0, 1]$ in each variate. The first two moments (mean and variance) of the distribution are

$$E[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]_{prior} \equiv \Theta_k^{(j)\{\overline{\mathcal{M}}\}} = \frac{\alpha_k^{(j)}}{\alpha_k^{(0)}} \tag{6}$$

$$\mathrm{V}[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]_{prior} = \frac{\alpha_k^{(j)}(\alpha_k^{(0)} - \alpha_k^{(j)})}{(\alpha_k^{(0)})^2(\alpha_k^{(0)} + 1)} \tag{7}$$

where $\alpha_k^{(0)} \equiv \sum_{j=1}^{J_k} \alpha_k^{(j)}$, and $E[\cdot]$ and $V[\cdot]$ are the expected value and the variance of the pdf in the square brackets, respectively.

The univariate case of equation 5 is called Beta distribution. It represents the Dirichlet distribution for two mutually exclusive events (e.g., magma or not). For a generic random variable ξ, Beta reads

$$[\xi] = \mathrm{Beta}(\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\xi^{\alpha - 1}(1 - \xi)^{\beta - 1} \tag{8}$$

where $\alpha, \beta > 0$, and a sufficient condition to have a finite pdf is $\alpha, \beta \geq 1$ (e.g., Gelman et al., 1995). Note that in the previous formula can be derived from equation 4 with $J = 2$, $\alpha_1 = \alpha$, $\alpha_2 = \beta$, $\xi^{(1)} = \xi$ and $\xi^{(2)} = 1 - \xi$. Consequently, also mean and variance for a Beta can be obtained by using the same substitutions in equations 6 and 7. Remarkably, the

marginal distribution of a specific $\theta_k^{(j)}$ (i.e., $[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]_{prior}$) in the more general case of the Dirichlet distribution ($J_k > 2$) is a Beta($\alpha_k^{(j)}, \alpha_k^{(0)} - \alpha_k^{(j)}$).

The average of those distributions ($\Theta_k^{(j)\{\cdot\}}$) represents an estimation of the aleatory uncertainty, i.e., the intrinsic (and unavoidable) random variability due to the complexity of the process. The dispersion around the central value (i.e., the variance), instead, mainly represents an estimation of the epistemic uncertainty, due to our limited knowledge of the process. In spite of the latter being neglected in past works, its estimation is very important for correct comparison between the probabilities of different hazards, and the confidence limits that are ascribed to them (cf. Gelman et al., 1995; Woo, 1999). Moreover, Marzocchi et al. (2004; 2006a) have shown that, by accounting for the epistemic uncertainties, also central values are significantly affected.

The Dirichlet (Beta) distribution has two important limit cases. The case of $\alpha_k^{(1)} = ... = \alpha_k^{(J)} = 1$ ($\alpha_k = \beta_k = 1$) represents the uniform distribution; in this case the mean and variance of the distribution at k-th node are

$$E[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]_{prior} \equiv \Theta_k^{(j)\{\overline{\mathcal{M}}\}} = \frac{1}{J_k} \tag{9}$$

$$V[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]_{prior} = \frac{J_k - 1}{J_k^2(J_k + 1)} \tag{10}$$

On the other hand, when $\alpha_k^{(j)} \to \infty$, $V[\theta_k^{(j)\{\overline{\mathcal{M}}\}}] \to 0$. This means that the distribution tends to a Dirac's δ distribution centered around the average of the j-th event.

Note that the choice of the Dirichlet (Beta) distribution is itself rather subjective, and in some case probabilities may be more appropriately characterized by other distributional forms. However, any possible bias introduced by this subjective choice is certainly less than by assuming an exact value for the probability, that is assuming a Dirac's δ distribution; actually, the latter is a much more subjective choice because it neglects the epistemic uncertainty. Further details on this choice can be found in Marzocchi et al. (2004).

The definition of the prior distribution consists of setting the Dirichlet (Beta) distribution by using a priori information. In general, theoretical models, a priori beliefs, and/or expert elicitation give estimation of the expected central value ($\Theta_k^{(j)\{\overline{\mathcal{M}}\}}$) of the prior distribution that represents the "best guess". The variance can be seen as a sort of "confidence degree" of our a priori information, i.e., an evaluation of the epistemic uncertainties. The confidence degree is set up writing the variance in terms of "equivalent number of data" ($\Lambda_k^{\{\overline{\mathcal{M}}\}}$)

$$\Lambda_k^{\{\overline{\mathcal{M}}\}} = \alpha_k^{(0)} - J_k + 1 \tag{11}$$

and then

$$V[\theta_k^{(j)\{\overline{\mathcal{M}}\}}]_{prior} = \frac{\Theta_k^{(j)\{\overline{\mathcal{M}}\}}(1 - \Theta_k^{(j)\{\overline{\mathcal{M}}\}})}{\Lambda_k^{\{\overline{\mathcal{M}}\}} + J_k} \tag{12}$$

The $\Lambda_k^{\{\overline{\mathcal{M}}\}}$ is a more friendly measure than V of the confidence on the prior distribution, or, in other terms, of the epistemic uncertainty. In general, the higher the $\Lambda_k^{\{\overline{\mathcal{M}}\}}$, 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), $\Lambda_k^{\{\overline{\mathcal{M}}\}}$ must be small, so that even a small number of past data can drastically modify the prior. The minimum possible $\Lambda_k^{\{\overline{\mathcal{M}}\}}$ is

1, and this number represent the maximum possible epistemic uncertainty. In the limit case of maximum ignorance ($\Lambda_k^{\{\overline{\mathcal{M}}\}} = 1$, that is, the reliability of our model is comparable to the information given by one datum) and a mean equal to 0.5 we obtain a prior uniform distribution. As $\Lambda_k^{\{\overline{\mathcal{M}}\}}$ increases, the Beta function becomes more and more spiked around the given mean. In the limit case discussed above of the Dirac's δ, this means that the epistemic uncertainty becomes negligible when we have a large amount of information.

In practice, the input parameters of the model to set up the prior distribution are usually $\Theta_k^{(j)\{\overline{\mathcal{M}}\}}$ and $\Lambda_k^{\{\overline{\mathcal{M}}\}}$. The parameters of the prior distribution ($\alpha_k^{(j)}, j = 1, ..., J_k$, see equation 5) are determined by resolving the system given by equations 6 and 11.

A.2.2 *Likelihood function*

Using Bayes' rule with a chosen probability model means that the past data at node k (y_k) affect the posterior inference (see equation 2) only through the function $[y_k|\theta_k^{\{\overline{\mathcal{M}}\}}]_{prior}$ that is called likelihood function. In the present case, we choose a multinomial distribution that for a multivariate random variable $\xi = (\xi^{(1)}, ..., \xi^{(J)})$ reads

$$[y|\xi] = \mathrm{Mu}_J(y^{(1)}, ..., y^{(J)}; \xi) = \begin{pmatrix} \sum_{j=1}^{J} y^{(j)} \\ y^{(1)} ... y^{(J)} \end{pmatrix} (\xi^{(1)})^{y^{(1)}} ... (\xi^{(J)})^{y^{(J)}} \tag{13}$$

where $y = (y^{(1)}, ..., y^{(J)})$ is the data vector, and the elements $y^{(j)}$ are the number of successes (occurrence) relative to the event labeled j with probability $\xi^{(j)}$. The sum $\sum_{j=1}^{J} y^{(j)}$ represents the total number of data n. In our case, we have

$$[y_k|\theta_k^{\{\overline{\mathcal{M}}\}}] \equiv \mathrm{Mu}_{J_k}(y_k^{(1)}, ... y_k^{(J_k)}; \theta_k^{\{\overline{\mathcal{M}}\}}) \tag{14}$$

where J_k is the number of possible mutually exclusive and exhaustive events at the k-th node. The likelihood function of the j-th event at node k given y_k ($[y_k|\theta_k^{(j)\{\overline{\mathcal{M}}\}}]$) is the marginal distribution of $\theta_k^{(j)\{\overline{\mathcal{M}}\}}$ in equation 14. Note that this distribution assumes that any data of the set y_k is independent from the others. This assumption seems reasonable in the present case.

In analogy with the previous section, in case $J = 2$ a multinomial distribution becomes a binomial distribution; let y_k be the variable that counts the number of occurrence of an event in a dataset of n_k data (e.g., at node 2, the number of magmatic intrusions y_2 out of the number of recorded unrest episodes n_2), then

$$[y_k|\theta_k^{\{\overline{\mathcal{M}}\}}] \equiv \mathrm{Bin}(y_k, n_k; \theta_k) = \begin{pmatrix} n_k \\ y_k \end{pmatrix} \theta_k^{y_k} (1 - \theta_k)^{n_k - y_k} \tag{15}$$

which can be obtained from equation 13 with the substitutions $J = 2$, $\xi^{(1)} = \theta_k$, $\xi^{(2)} = 1 - \theta_k$, $y^{(1)} = y_k$, and $y^{(2)} = n_k - y_k$. Also in this univariate case, the assumption is that the n_k data are independent from each other.

A.2.3 *Posterior distribution*

The choice of the Dirichlet and Multinomial (Beta and Binomial in the univariate case) distributions simplifies the computation, because they are conjugate distributions (Gelman et al., 1995), i.e., a Dirichlet multiplied by a Multinomial is still a Dirichlet. Through Bayes theorem

and adopting the results of the conjugate families, we obtain the following posterior distribution for $\theta_k^{\{\overline{\mathcal{M}}\}}$:

$$[\theta_k^{\{\overline{\mathcal{M}}\}}] \equiv [\theta_k^{\{\overline{\mathcal{M}}\}}|y_k] = \mathrm{Di}_{J_k}(\alpha_k^{(j)} + y_k^{(j)}; j = 1, ..., J_k) \tag{16}$$

where $y_k^{(j)}$ have the same meaning as in equation 14, and $\alpha_k^{(j)}$ as in equation 5.

In analogy, as regards the univariate case, through Bayes theorem and adopting the results of the conjugate families, we obtain the following posterior distribution for θ_k:

$$[\theta_k^{\{\overline{\mathcal{M}}\}}] \equiv [\theta_k^{\{\overline{\mathcal{M}}\}}|y_k] = \mathrm{Beta}(\alpha_k + y_k, \beta_k + n_k - y_k) \tag{17}$$

Note that from equations 16 or 17, it is possible to see that the relative weight between prior and likelihood to form the posterior is controlled by the ratio between the equivalent number of data $\Lambda_k^{\{\overline{\mathcal{M}}\}}$ of the prior and the total number of past data n_k of the likelihood, since $\Lambda_k^{\{\overline{\mathcal{M}}\}}$ is function of the sum of $\alpha_k^{(j)}$ (see equation 11) and n_k is the sum of $y_k^{(j)}$ (see subsections A.2.1 and A.2.2).

A.3 From monitoring measures to probabilities

In the common practice, volcanologists interpret monitoring measures in order to have indications about the actual state of the volcano and hints about its possible evolution. This is logically done by evaluating the following two distinct sentences (see figure 5):

- STEP 1: *The measure x is anomalous*
- STEP 2: *Anomalous values of x are indicative of one specific state of the volcano*

In this paper, STEP 1 is evaluated by means of the fuzzy approach, defining the *degree of anomaly* of the monitoring observations; in STEP 2 we define the procedure to calculate the probabilities from the degree of anomaly of the monitoring measurements. In the following we report the general features of the fuzzy approach applied to our case, and the definition of the state of unrest η. The details of the STEP 2 for each node are reported in Appendix B.

A.3.1 *Fuzzy approach for monitoring measures*

The introduction of the fuzzy approach (Zadeh, 1965) in BET helps to manage the monitoring measures. A detailed explanation of fuzzy set theory is beyond the aims of the present paper. Here, we only explain how the fuzzy approach enters into BET.

After having selected the relevant monitoring parameters for each node, we have to define when a monitoring measure can be considered "anomalous". In Marzocchi et al. (2004) we have introduced a single threshold for each monitoring parameter that distinguish two possible states of the monitoring measures, "anomalous" or "not anomalous".

The use of one threshold may be too rough, because of several reasons. At first, a system can move to an anomalous state gradually, rather than overcoming one specific value of a monitoring variable. Secondly, since the definition of a threshold is strongly subjective, and that such a threshold may strongly affect the results, a more "fuzzy" definition seems to be more appropriate. For these reasons, we introduce the fuzzy set theory that quantitatively emulates the expert-like flexible judgment of the anomalous state of a monitoring parameter. In fact, the proposition in STEP 1 stands for the membership of the present measure of the parameter $x_k^{(i)}$ to the set of anomalous values of the parameter. In a fuzzy perspective, this membership is not only TRUE or FALSE as in Boolean logic, but can be true by some degree. As a consequence, the proposition

in STEP 1 can be partly true and partly wrong, to some extent. In other words, BET associates a real value between 0 and 1 to the logical statement in STEP 1, representing the *degree of truth* of the statement, i.e., the *degree of anomaly* $z_k^{(i)}$ of a given measure of $x_k^{(i)}$ (see figure 6). In practice, for each monitoring parameter $x_k^{(i)}$, a function called *membership function* $\mu(x_k^{(i)})$ is defined; $\mu(x_k^{(i)})$ associates to each measurable value of $x_k^{(i)}$ its relative *degree of anomaly*

$$z_k^{(i)} = \mu(x_k^{(i)}) \tag{18}$$

Note that, for the sake of simplicity, in equation 18 we introduce a single notation $\mu(\cdot)$ to indicate a generic membership function, even though different parameters $x_k^{(i)}$ have different $\mu(\cdot)$.

In BET, membership functions have two possible fixed shapes:

$$\mu(x) = \begin{cases} 1 & \text{if } x \geq t_u \\ \frac{1}{2}\{\sin[\pi(\frac{x-t_l}{t_u-t_l}) - \frac{\pi}{2}] + 1\} & \text{if } t_l < x < t_u \quad \text{for increasing trend} \\ 0 & \text{if } x \leq t_l \end{cases} \tag{19}$$

$$\mu(x) = \begin{cases} 0 & \text{if } x \geq t_u \\ \frac{1}{2}\{\sin[\pi(\frac{x-t_l}{t_u-t_l}) + \frac{\pi}{2}] + 1\} & \text{if } t_l < x < t_u \quad \text{for decreasing trend} \\ 1 & \text{if } x \leq t_l \end{cases} \tag{20}$$

where t_l and t_u are the lower and upper thresholds of the parameter. The membership function $\mu(x)$ is controlled by three parameters: t_l, t_u, and the kind of trend, i.e., increasing or decreasing. All of them are input of the model that can be set up either by the elicitation of a pool of experts, or by analysis of past measures, or both. The latter factor, i.e., the trend, accounts for the fact that measures may be anomalous either for large (increasing trend, as in figure 6) or for small (decreasing trend) values. Finally, note that the Boolean case is also included, when upper and lower thresholds collapse to the same value (dotted line in figure 6), i.e., $t_l = t_u$.

As an example, let $x_k^{(i)}$ be a monitoring parameter at the k-th node. Based on our knowledge of the volcano background for that variable, we define its relative membership function $\mu(x_k^{(i)})$. In practice we select between increasing or decreasing trend (equation 19 and 20, respectively), and we fix t_l and t_u. In figure 6 we show a possible $\mu(x_k^{(i)})$ (solid line) and, for comparison, the corresponding Boolean case (dashed line). The difference should be clear: for example, in the fuzzy case, a measured value $x_k^{(i)} = 5$ has a $\mu(5) = 0.3455$, meaning that such value is indicative of some degree of anomaly in the monitored variable. However, in the Boolean case, such value is considered completely not anomalous. Therefore, this fuzzy procedure also implies that small changes in monitoring measures do not lead to jumps in probability (as for measures close to the threshold in a Boolean approach), if this is not wanted.

To sum up, there are two main advantages in using a fuzzy logic approach in BET: (i) the variations in probabilities are smooth for small changes of the monitoring measures; (ii) rather than a sharp threshold, there is a wider range of possible values to be defined in order to identify probability changes. The latter advantage helps in defining threshold ranges supported by a wider expert consensus.

A.3.2 *State of unrest* η

One of the key points in the approach of BET is to define the degree at which the volcano is found in unrest at the time $t = t_0$. As mentioned above, for some nodes the parameter γ_k (see section A.1) is a function of the state of unrest η. Even though each γ_k is computed from η, the

form of this relationship is different from node to node, and depends only on the information content of the monitoring at the node in case of unrest.

The state of unrest η is defined through the monitoring parameters and their relative membership function $\mu(\cdot)$ at node 1. Generally, there are several possible monitoring parameters indicative of unrest in a volcanic system, for instance, seismic activity, ground deformation, geochemical signatures, and so on. At a specific volcano, once the monitored variables indicative of unrest have been selected, they are processed in a fuzzy perspective.

We define the degree of unrest of the volcano at the time of monitoring t_0 (see STEP 2 in section A.3) as:

$$\eta = 1 - \prod_{i=1}^{L_1} (1 - z_i^{(1)}) \tag{21}$$

where L_1 is the number of monitored parameters selected at node 1, and the $z_i^{(1)}$ are the degrees of anomaly of measures at time $t = t_0$ (see equation 18).

This choice is "conservative", because as soon as one monitored variable approaches the truth of its proposition (see STEP 1), the volcano is considered certainly in unrest ($\eta = 1$). In other words, this means that the recognition of one anomalous parameter is enough to identify a state of unrest.

APPENDIX B
Probability estimation at the nodes: specific calculations

In this section, we report the details for the conditional probability assessment for each node. Specifically, we describe the explicit formulation for the three unknowns in equation 1. More details and examples are given by Marzocchi et al. (2004, 2006a).

B.1 Node 1

The first node has two possible outcomes: *presence of unrest* and *not presence of unrest* in a time interval $(t_0, t_0 + \tau]$.

B.1.1 *Parameter* γ_1

At this node we set $\gamma_1 = \eta$. This means that we consider the monitoring measures fully informative during an unrest, and the past frequency of unrest is not longer relevant.

B.1.2 *The posterior distribution* $[\theta_1^{\{\mathcal{M}\}}]$

The distribution $[\theta_1^{\{\mathcal{M}\}}]$ is the pdf of the posterior probability to have an unrest in the time interval $(t_0, t_0+\tau]$ using the monitoring data when the volcano is in unrest at time t_0, i.e., $\eta = 1$. As mentioned before, in this case we do not use the Bayes' rule, but we identically set

$$[\theta_1^{\{\mathcal{M}\}}] = \delta(\theta_1^{\{\mathcal{M}\}} - 1) \tag{22}$$

where $\delta(\cdot)$ indicate the Dirac function. This means that if the volcano is in unrest at time t_0, the unrest is assumed to be also in the time interval $(t_0, t_0 + \tau]$. We use a Dirac distribution rather than a Beta distribution as for the subsequent nodes, because the concept of "unrest" is purely subjective, and what we can realistically assign is a sort of "degree of unrest" (i.e., the parameter η). The Beta distribution is more suitable when a "real" value of θ exists and we want to estimate it; in this case, it is necessary to estimate also the uncertainty about our estimation that can be considered by using a Beta distribution.

B.1.3 *The posterior distribution* $[\theta_1^{\{\overline{\mathcal{M}}\}}]$

We define the following prior distribution (see equation 8) for θ_1:

$$[\theta_1^{\{\overline{\mathcal{M}}\}}]_{prior} = \mathrm{Beta}(\alpha_1, \beta_1) \tag{23}$$

The parameters α_1, β_1 are determined by

$$\alpha_1 = \Theta_1^{\{\overline{\mathcal{M}}\}}(\Lambda_1^{\{\overline{\mathcal{M}}\}} + 1) \tag{24}$$

$$\beta_1 = (\Lambda_1^{\{\overline{\mathcal{M}}\}} + 1) - \alpha_1 \tag{25}$$

As previously discussed, $\Theta_1^{\{\overline{\mathcal{M}}\}}$ is the central value inferred by *a priori* models and/or of the theoretical beliefs, while $\Lambda_1^{\{\overline{\mathcal{M}}\}}$ controls the confidence at which $\Theta_1^{\{\overline{\mathcal{M}}\}}$ is considered a reliable estimate. Both these parameters are input of the BET model.

In writing the likelihood for node 1, the two possible outcomes can be treated as "success" and "failure", respectively, in a binomial model only under specific conditions. Specifically, if we define $n^\star$ as the total number of non-overlapping time windows of the available catalog, we obtain a sequence of 0 (no unrest) and 1 (unrest) that could be autocorrelated if the duration of unrest is larger than the time window considered (τ). In order to remove this autocorrelation and therefore to use the binomial model, we consider n_1 that is the total number of non-overlapping time windows investigated that begin in a state of no unrest. In this scheme, y_1 is the variable that counts the number of (non-overlapping) time windows containing onset of unrest that occur in a set of n_1 inspections. In other words, each unrest counts as one, regardless its duration. Following equation 15, we set

$$[y_1|\theta_1^{\{\overline{\mathcal{M}}\}}] = \mathrm{Bin}(y_1, n_1; \theta_1) \tag{26}$$

From equation 17, we obtain the following posterior distribution for $[\theta_1^{\{\overline{\mathcal{M}}\}}]$:

$$[\theta_1^{\{\overline{\mathcal{M}}\}}] \equiv [\theta_1^{\{\overline{\mathcal{M}}\}}|y_1] = \mathrm{Beta}(\alpha_1 + y_1, \beta_1 + n_1 - y_1), \tag{27}$$

B.2 Node 2

Node 2 has two possible outcomes: success is *magmatic unrest* and failure is *not magmatic unrest*, given that unrest has occurred.

B.2.1 *Parameter* γ_2

At this node $\gamma_2 = \eta$. This means that we consider the monitoring measures fully informative during an unrest, i.e., the past frequency of magmatic unrest is no longer relevant. Note that η (and then γ_2) is computed from the monitoring measures at node 1.

B.2.2 *The posterior distribution* $[\theta_2^{\{\mathcal{M}\}}]$

At first, we need to select the monitored variables indicative of magmatic origin for the unrest. It is worth remarking that a monitored variable can be also a combination of different measures (i.e., a variable can be a simultaneous increase in seismicity and an inflation). For each monitored variable selected $x_2^{(i)} (i = 1, ..., L_2$, where L_2 is the number of selected monitored variables), it is necessary to define the membership function (i.e., lower and upper thresholds, and trend).

At each monitoring variable $x_2^{(i)}$, we can associate a weight $\omega_2^{(i)}$ ($\omega_2^{(\cdot)} \geq 1$) that ranks the relative importance of each parameter compared to the others. For example, if $\omega_2^{(1)} = 1$ and $\omega_2^{(2)} = 2$, this means that the second parameter is considered more informative in identifying the presence of magma.

Once we have selected the set of parameters and their weights we have to translate these measures into probabilities. At first (STEP 1 in section A.3), we use the fuzzy approach to estimate the degrees of anomaly $z_2^{(i)}$ at $t = t_0$ for all the monitoring parameters at node 2 through equation 18 (where the $\mu(\cdot)$ is the membership function relative to the i-th parameter at node $k = 2$). Then we calculate (Marzocchi et al., 2004)

$$Z_2 = \sum_{i=1}^{L_2} w_2^{(i)} z_2^{(i)} \tag{28}$$

Equation (28) defines Z_2 as a linear combination of the weighted membership functions, that represent the degree of anomaly of the monitored parameters as measured at time t_0. In practice, Z_2 provides an index of the "magmatic unrest"; higher Z_2 implies higher probability to have a magmatic unrest.

Now, we have to translate this information in terms of probability (see STEP 2 in section B.3). Basically, we use a functional relationship between Z_2 and $\Theta_2^{\{\mathcal{M}\}}$ (that is the average of a Beta distribution), and we set $\Lambda_2^{\{\mathcal{M}\}} = 1$, i.e., the maximum variance allowed. We use an exponential relationship like

$$\Theta_2^{\{\mathcal{M}\}} = 1 - ae^{-bZ_2} \tag{29}$$

where a and b are positive parameters that will be set by means of Bayesian inference (see below). Before describing that, we want to highlight the implications of the choice of the exponential function given by equation 29. First, this relationship is monotonically increasing so that the larger Z_2, the larger $\Theta_2^{\{\mathcal{M}\}}$. Second, such kind of relationship implies that the largest increase in probability mean occurs when one of the monitoring variables shows some degree of anomaly. As more monitored variables become anomalous, the probability mean keeps on rising, but slower.

We set the prior distribution for the parameter $\theta_2^{\{\mathcal{M}\}}$ as the marginal distribution of $\theta_2^{\{\mathcal{M}\}}$ by integrating over a and b the joint prior distribution of $\theta_2^{\{\mathcal{M}\}}$, a, and b:

$$\begin{aligned}
[\theta_2^{\{\mathcal{M}\}}]_{prior} &= \int_{0.5}^{1} da \int_{0}^{2} db\ [\theta_2^{\{\mathcal{M}\}}, a, b]_{prior} \\
&= \int_{0.5}^{1} da \int_{0}^{2} db\ [\theta_2^{\{\mathcal{M}\}}|a, b][a, b]_{prior} \\
&= \int_{0.5}^{1} da \int_{0}^{2} db\ \mathrm{Beta}(1 - ae^{-bZ_2}, ae^{-bZ_2}) \cdot 2 \cdot 0.5
\end{aligned} \tag{30}$$

where a and b are assumed independent and their prior distribution is a uniform distribution, i.e., $[a, b]_{prior} = [a]_{prior}[b]_{prior} = 2 \cdot 0.5$. The domain of the parameters a and b is chosen to have a reliable prior distribution. For instance, the domain of a could be [0,1], because this set of values guarantees that $[\theta_2^{\{\mathcal{M}\}}]_{prior}$ (equation 30) is still a probability. Nevertheless, we prefer to restrict the domain in the interval $[0.5, 1]$ in order to account for the "damping effect" that we have if $Z_2 = 0$ [see *Marzocchi et al.*, 2004]. Let us suppose, for example, that we are in clear unrest, but the present monitoring at node 2 gives $Z_2 = 0$. In this case, the monitoring parameters stand for a not magmatic unrest. However, if the domain for a is [0,1] (average is 0.5), this leads to $\Theta_2^{\{\mathcal{M}\}} = 0.5$ (see equation 29). Instead, by restricting the domain for a, $\Theta_2^{\{\mathcal{M}\}}$ is significantly damped (at 0.25), because the average of a is now 0.75 (cf. Marzocchi et al., 2004).

The posterior distribution of the values of a and b in equation 29 is based on Bayesian inference. At node 2, we assume that past monitoring data can modify the prior distribution for $\theta_2^{\{\mathcal{M}\}}$ (equation 30) only by refining the values of a and b. In Bayesian terms, given $n_2^{\{\mathcal{M}\}}$ past monitored episodes of unrest, each with an observed Z_{2_i} (as defined in equation 28) and with outcome y_{2_i} (where y_{2_i} is 1 if the unrest turned out to be magmatic, or 0 on the opposite), the joint posterior distribution of $\theta_2^{\{\mathcal{M}\}}$, a and b is:

$$\begin{aligned}
[\theta_2^{\{\mathcal{M}\}}, a, b|Z_2, \{Z_{2_i}\}, \{y_{2_i}\}] &\propto [\theta_2^{\{\mathcal{M}\}}|a, b, Z_2, \{Z_{2_i}\}, \{y_{2_i}\}][a, b|Z_2, \{Z_{2_i}\}, \{y_{2_i}\}] \\
&\propto [\theta_2^{\{\mathcal{M}\}}|a, b, Z_2][a, b|\{Z_{2_i}\}, \{y_{2_i}\}]
\end{aligned} \tag{31}$$

where the sign $\propto$ is because the term on the right side is not normalized to 1, and $\{Z_{2_i}\}$ and $\{y_{2_i}\}$ are the sets of Z_{2_i} from past monitored unrest episodes, and of their outcomes y_{2_i}, respectively, for $i = 1, ...n_2^{\{\mathcal{M}\}}$. The simplification of the right member comes from the fact that $[\theta_k^{(M)}|a, b, Z_2, \{Z_{2_i}\}, \{y_{2_i}\}]$ is actually independent on $\{Z_{2_i}\}$ and $\{y_{2_i}\}$ (it does only through a and b), and that $[a, b|Z_2, \{Z_{2_i}\}, \{y_{2_i}\}]$ is actually independent on Z_2.

The last term of equation 31, $[a, b|\{Z_{2_i}\}, \{y_{2_i}\}]$, can be further factorized by means of Bayes' rule:

$$[a, b|\{Z_{2_i}\}, \{y_{2_i}\}] \propto [a, b]_{prior}[\{y_{2_i}\}|a, b, \{Z_{2_i}\}] \tag{32}$$

Under the assumption of independent past data, the likelihood $[\{y_{2_i}\}|a, b, \{Z_{2_i}\}]$ is given by the product

$$[\{y_{2_i}\}|a, b, \{Z_{2_i}\}] = \prod_{j=1}^{n_2^{\{\mathcal{M}\}}} (1 - a\exp(-bZ_{2_i}))^{y_{2_i}} (a\exp(-bZ_{2_i}))^{1-y_{2_i}} \tag{33}$$

This likelihood resembles the so-called dose-response relationship based on the multinomial scheme (Gelman et al., 1995). It means that the most likely values of a and b are those maximizing the probability of having observed those monitoring data (yielding the computed $\{Z_{2_i}\}$) in past episodes of unrest with the outcomes given by $\{y_{2_i}\}$.

Finally, the posterior distribution of $\theta_2^{\{\mathcal{M}\}}$ is given by

$$\begin{aligned}
[\theta_2^{\{\mathcal{M}\}}] &= \int_{0.5}^{1} da \int_{0}^{2} db\ [\theta_2^{\{\mathcal{M}\}}, a, b|Z_2, \{Z_{2_i}\}, \{y_{2_i}\}] \\
&\propto \int_{0.5}^{1} da \int_{0}^{2} db\ \mathrm{Beta}(1 - ae^{-bZ_2}, ae^{-bZ_2}) \cdot 2 \cdot 0.5 \\
&\quad \cdot \prod_{j=1}^{n_2^{\{\mathcal{M}\}}} (1 - a\exp(-bZ_{2_i}))^{y_{2_i}} (a\exp(-bZ_{2_i}))^{1-y_{2_i}}
\end{aligned} \tag{34}$$

In practice, in order to determine the values for a and b, we start by random sampling 1000 pairs of (a, b) from their posterior bivariate joint distribution given by equation 32 with likelihood as in equation 33, and prior $[a, b]$ distribution as a bivariate uniform on the domains [0.5,1] for a and [0,2] for b. Then each pair is used in equation 30 to build up 1000 different Beta distributions. From each of them, a single random sampling is performed. The obtained 1000 sampling represent the posterior $[\theta_2^{\{\mathcal{M}\}}]$ searched.

B.2.3 *The posterior distribution* $[\theta_2^{\{\overline{\mathcal{M}}\}}]$

The formulation of $[\theta_2^{\{\overline{\mathcal{M}}\}}]$ is identical to the one of $[\theta_1^{\{\overline{\mathcal{M}}\}}]$. The symbols (obviously with index k relative to node 2 instead of 1) have a different physical meaning. In particular: $\Theta_2^{\{\overline{\mathcal{M}}\}}$ is the expected value of the probability for the unrest being magmatic, given there is unrest, provided by theoretical models and beliefs; $\Lambda_2^{\{\overline{\mathcal{M}}\}}$ is the number of equivalent data that we assign to our a priori model; y_2 is the number of observed magmatic unrest episodes at the volcano in the past; n_2 is the total number of observed unrest episodes at the volcano in the past, whose source process (magmatic or not) is known, i.e., $n_2 \leq y_1$.

Therefore, the posterior $[\theta_2^{\{\overline{\mathcal{M}}\}}]$ is:

$$[\theta_2^{\{\overline{\mathcal{M}}\}}] \equiv [\theta_2^{\{\overline{\mathcal{M}}\}}|y_2, y_1] = \text{Beta}(\alpha_2 + y_2, \beta_2 + y_1 - y_2), \tag{35}$$

B.3 Node 3

Node 3 has two possible outcomes: success is *occurrence of eruption* and *not occurrence of eruption* in a time interval $(t_0, t_0 + \tau]$, given magmatic unrest.

B.3.1 *Parameter* γ_3

At this node $\gamma_3 = \eta$. See considerations made at node 2 for implications of this choice.

B.3.2 *The posterior distribution* $[\theta_3^{\{\mathcal{M}\}}]$

The treatment of the monitored variables to set up the probability distribution for the monitoring part at node 3 is exactly the same of node 2, except that the monitored variables to be selected $(x_i^{(3)})$ should be indicative of magma erupting. At this node, the posterior distribution of $\theta_3^{\{\mathcal{M}\}}$ is given by

$$\begin{aligned}
[\theta_3^{\{\mathcal{M}\}}] &= \int_{0.5}^{1} da \int_{0}^{2} db\ [\theta_3^{\{\mathcal{M}\}}, a, b | Z_2, \{Z_{3_i}\}, \{y_{3_i}\}] \\
&\propto \int_{0.5}^{1} da \int_{0}^{2} db\ \text{Beta}(1 - ae^{-bZ_3}, ae^{-bZ_3}) \cdot 2 \cdot 0.5 \\
&\quad \cdot \prod_{j=1}^{n_3^{\{\mathcal{M}\}}} (1 - a\exp(-bZ_{3_i}))^{y_{3_i}} (a\exp(-bZ_{3_i}))^{1-y_{3_i}}
\end{aligned} \tag{36}$$

where a and b have posterior distributions different from the ones determined at node 2, $n_3^{\{\mathcal{M}\}}$ is the number of past monitored episodes of magmatic unrest, $\{Z_{3_i}\}(i = 1, ..., n_3^{\{\mathcal{M}\}})$ is the set of Z_{3_i} coming from past monitored magmatic unrest (see equation 28 with index equal to 3), and $\{y_{3_i}\}(i = 1, ..., n_3^{\{\mathcal{M}\}})$ is the set of outcomes of the past monitored magmatic unrest (0 for no eruption, 1 for eruption).

B.3.3 *The posterior distribution* $[\theta_3^{\{\overline{\mathcal{M}}\}}]$

The formulation of $[\theta_3^{\{\overline{\mathcal{M}}\}}]$ is identical to the one of $[\theta_1^{\{\overline{\mathcal{M}}\}}]$. The symbols (obviously with index k relative to node 3 instead of 1) have a different physical meaning. In particular: $\Theta_3^{\{\overline{\mathcal{M}}\}}$ is the expected value of the probability of eruption, given there is a magmatic unrest, provided by theoretical models and beliefs; $\Lambda_3^{\{\overline{\mathcal{M}}\}}$ is the number of equivalent data that we assign to our a priori model; y_3 is the number of observed eruptions at the volcano in the past; n_3 is the total number of observed magmatic unrest at the volcano in the past, whose outcome (eruption or not) is known, i.e., $n_3 \leq y_2$.

Therefore, the posterior $[\theta_3^{\{\overline{\mathcal{M}}\}}]$ is:

$$[\theta_3^{\{\overline{\mathcal{M}}\}}] \equiv [\theta_3^{\{\overline{\mathcal{M}}\}}|y_3, y_2, y_1] = \text{Beta}(\alpha_3 + y_3, \beta_3 + y_2 - y_3), \tag{37}$$

B.4 Node 4

Node 4 has J_4 possible outcomes, each one related to a specific location of the eruption.

B.4.1 *Parameter* γ_4

At node 4, the monitoring information never fully controls the probability $[\theta_4^{(j)}]$. In fact, several phenomena, i.e., migration of seismic events within the volcanic setting, may lead to false localizations of the vent location (e.g. Ozawa et al., 2004). Therefore, here we decide to give the same weight to monitoring and non-monitoring part, even in case of clear unrest ($\eta = 1$); accordingly, we set

$$\gamma_4 = \text{Min}(\eta,\, 0.5) \tag{38}$$

This means that during an unrest the probability of a specific vent opening accounts for both monitoring and non-monitoring information.

B.4.2 *The posterior distribution* $[\theta_4^{(j)\{\mathcal{M}\}}]$

For this node BET does not require a specific monitoring. In theory, each monitoring measurements already chosen from all previous nodes may be localized, depending on the measurement network, and the specific physical meaning of the parameter. When it is possible to localize relevant monitored measurements, we calculate the fraction of the measured phenomenon occurring in the specific vent location. Note that the sum of these fractions on all the vent locations must give 1 for every monitored parameter.

All localized monitoring measures are then combined to form the monitoring probability distributions. In this case the prior distribution is

$$[\theta_4^{\{\mathcal{M}\}}]_{prior} = \text{Di}_{J_4}(\alpha_4^{(j)} = \frac{J_4}{L_4}\sum_{n=1}^{L_4} f_n^{(j)}; j = 1, 2, ..., J_4) \tag{39}$$

where $f_n^{(j)}$ is the fraction of the n-th localized monitoring measure (among L_4) in the j-th vent location, and $\text{Di}_{J_4}(\cdot)$ is a Dirichlet distribution (see equation 5).

For the j-th vent location, the posterior distribution for the monitoring part is the marginal distribution of $\theta_4^{(j)\{\mathcal{M}\}}$ in equation 39 (i.e., $\text{Beta}(\alpha_4^{(j)}, J_4 - \alpha_4^{(j)})$). Therefore, the mean of the j-th vent location is

$$E[\theta_4^{(j)\{\mathcal{M}\}}] \equiv \Theta_4^{(j)\{\mathcal{M}\}} = \frac{\sum_{n=1}^{L_4} f_n^{(j)}}{L_4} \tag{40}$$

In practice, at each location, the mean of the distribution is the average of the fractions of all localized monitoring measures at that vent location. The variance of the probability distribution is taken as the maximum allowed, i.e., $\Lambda_4^{\{\mathcal{M}\}} = 1$.

Since at node 4 the monitored measures and their localizations are dynamically chosen, the type and number of localized monitored parameters may change in time during the unrest,

depending on the availability of data. This prevents the possibility of setting up the form of the likelihood distribution, thus the posterior equals the prior:

$$[\theta_4^{\{\mathcal{M}\}}] = [\theta_4^{\{\mathcal{M}\}}]_{prior} \tag{41}$$

B.4.3 *The posterior distribution* $[\theta_4^{(j)\{\overline{\mathcal{M}}\}}]$

We define the following prior distribution:

$$[\theta_4^{\{\overline{\mathcal{M}}\}}]_{prior} = \mathrm{Di}_{J_4}(\alpha_4^{(j)}; j = 1, 2, ..., J_4) \tag{42}$$

that is a Dirichlet distribution (see equation 5) with J_4 parameters $\alpha_4^{(j)}$ set as

$$\alpha_4^{(j)} = \Theta_4^{(j)\{\overline{\mathcal{M}}\}}(\Lambda_4^{\{\overline{\mathcal{M}}\}} + J_4 - 1) \tag{43}$$

where $\Theta_4^{(j))\{\overline{\mathcal{M}}\}}$ and $\Lambda_4^{\{\overline{\mathcal{M}}\}}$ are parameters that depend on the expected value and variance of the theoretical models and/or beliefs. Note that the $\alpha_4^{(j)}$ parameters here are different from the ones used in the monitoring part.

For the likelihood, we use a multinomial distribution with J_4 possible outcomes (one for each vent location). As seen above (see equation 16), through Bayes theorem and adopting the results of the conjugate families, we obtain the following posterior distribution

$$[\theta_4^{\{\overline{\mathcal{M}}\}}] \equiv [\theta_4^{\{\overline{\mathcal{M}}\}}|y_4, y_3, y_2, y_1] = \mathrm{Di}_{J_4}(\alpha_4^{(j)} + y_4^{(j)}; j = 1, 2, ..., J_4) \tag{44}$$

where $y_4^{(j)}$ is the number of eruptions observed at location j in the past, and $\sum_{j=1}^{J_4} y_4^{(j)}$ is the total number of eruptions at the volcano with known localization. Those events must represent a complete eruption catalog of the period considered.

For the j-th vent location, the posterior distribution for the non-monitoring part is the marginal distribution of $\theta_4^{(j)\{\overline{\mathcal{M}}\}}$ in equation 44, i.e., $\mathrm{Beta}(\alpha_4^{(j)}, (\sum_{i=1}^{J_4} \alpha_4^{(i)}) - \alpha_4^{(j)})$.

B.5 Node 5

Node 5 has J_5 possible outcomes, each one related to a specific style of eruption, which is grouped in J_5 classes generically defined "sizes/types". The definition of sizes/types is generic since it may depend on the goal of the BET application, and on the characteristics of the target volcano. In most of the cases, such a size can be estimated through the VEI (Newhall and Self, 1982), but it is not the only possible choice; for instance, we can use different eruptive styles like "effusive" and "explosive".

In order to avoid a cumbersome notation, here we assume that the size distribution is independent from the vent location. This means that we assume that in all possible vent locations the probability estimations of sizes are identical. This assumption can be dropped without introducing any new conceptual complication.

B.5.1 *Parameter* γ_5

Even though there are some promising studies (for example degassing in volcanic areas inhibiting explosive activity; Newhall, 2003), until now there is no reliable precursor of the eruption size (e.g., Sandri et al., 2004). Because of this, no monitoring variable at the moment

could be selected to improve the probability estimation at this node. Therefore we set $\gamma_5 = 0$ and we do not compute $[\theta_5^{(j)\{\mathcal{M}\}}]$. However, when more sound monitoring precursors are found, BET can be implemented in order to account for it.

B.5.2 *The posterior distribution* $[\theta_5^{(j)\{\overline{\mathcal{M}}\}}]$

The formulation of $[\theta_5^{(j)\{\overline{\mathcal{M}}\}}]$ is identical to the one of $[\theta_4^{(j)\{\overline{\mathcal{M}}\}}]$. The symbols (obviously with index k relative to node 5 instead of 4) have a different physical meaning. In particular: $\Theta_5^{(j)\{\overline{\mathcal{M}}\}}$ is the expected value of the probability of an eruption of j-th size, given there is an eruption, provided by theoretical models and beliefs; $\Lambda_5^{\{\overline{\mathcal{M}}\}}$ is the equivalent number of data that defines the variance of the distribution; $y_5^{(j)}$ is the number of observed eruptions of the j-th size at the volcano in the past; $n_5 = \sum_{j=1}^{J_5} y_5^{(j)}$ is the total number of observed eruptions at the volcano in the past with known size. Those events must represent a complete eruption catalog of the period considered.

The posterior $[\theta_5^{(j)\{\overline{\mathcal{M}}\}}]$ is:

$$[\theta_5^{\{\overline{\mathcal{M}}\}}] \equiv [\theta_5^{\{\overline{\mathcal{M}}\}} | y_5, y_3, y_2, y_1] = \mathrm{Di}_{J_5}(\alpha_5^{(j)} + y_5^{(j)}; j = 1, ..., J_5) \tag{45}$$

For the j-th size class, the posterior distribution for the non-monitoring part is the marginal distribution of $\theta_5^{\{\overline{\mathcal{M}}\}}$ in equation 45, i.e., $\mathrm{Beta}(\alpha_5^{(j)}, (\sum_{i=1}^{J_5} \alpha_5^{(i)}) - \alpha_5^{(j)})$.

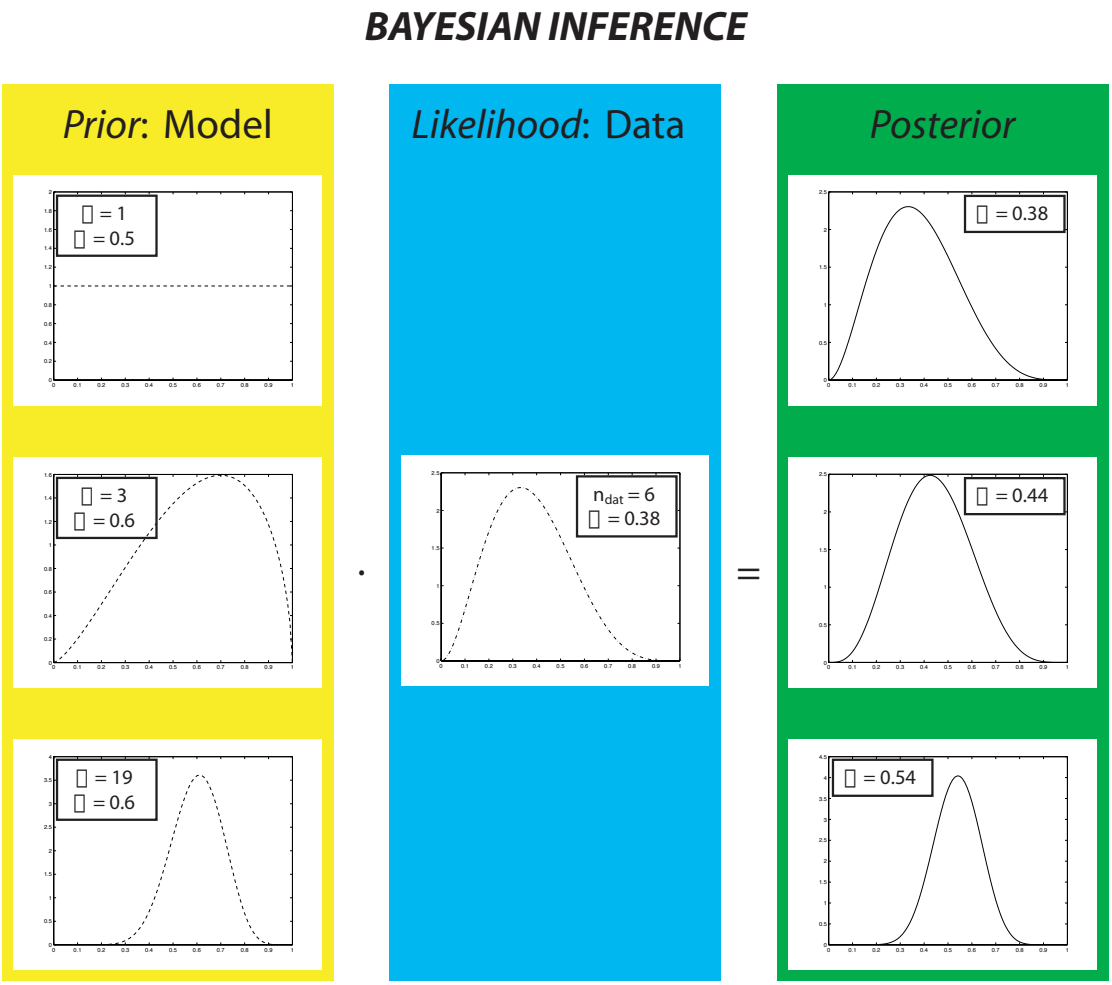

Figure 4: Scheme of Bayesian inference. Models and theoretical beliefs form the prior pdf (three different examples in the left panel) that is reshaped according to past data (middle panel) to give the posterior pdf (right panel). In the left panel (prior) we report three different pdfs with same average (Θ) but different "equivalent number of data" (Λ). High values of Λ correspond to more spiked distributions, and to a higher weight compared to past data in order to influence the corresponding posterior.

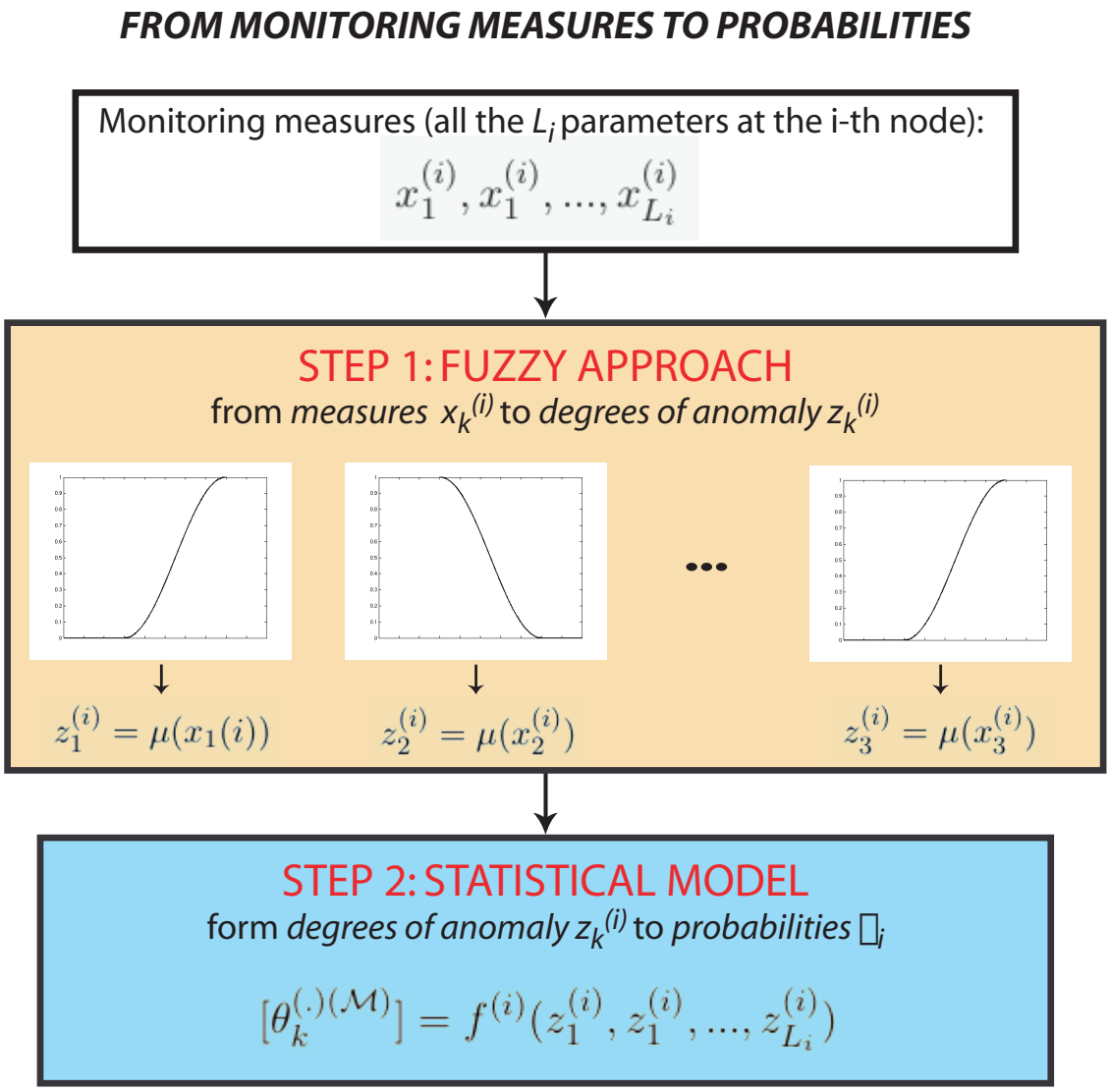

Figure 5: Monitoring measures are processed to compute the monitoring part of probability distributions. At first (STEP 1) the degrees of anomaly of each measure is computed through a fuzzy approach; then (STEP 2) degrees of anomaly are used to form the probability distribution.

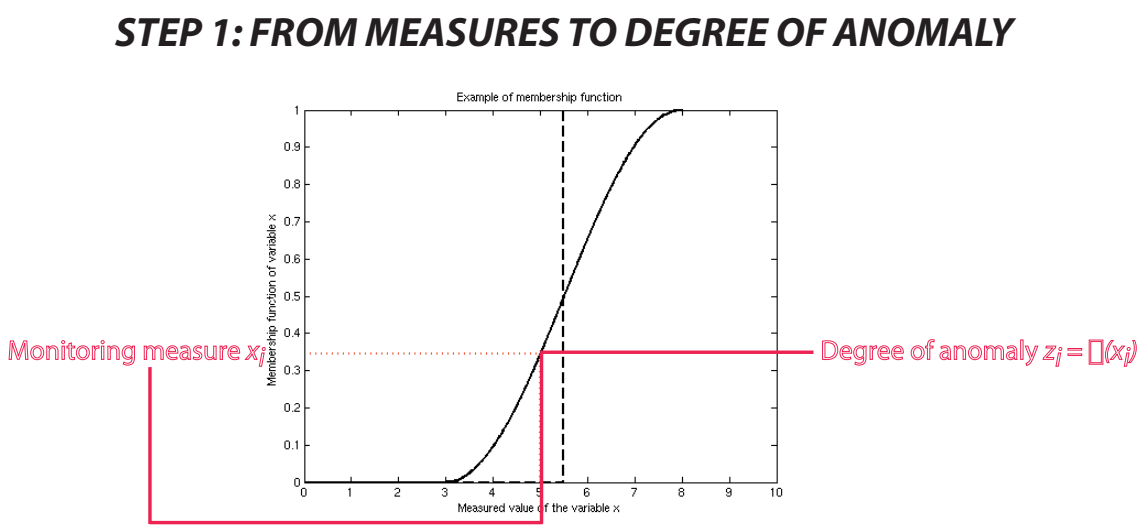

Figure 6: The degree of anomaly of each measure is computed in a fuzzy perspective, so that measure is not strictly either anomalous or not anomalous (Boolean case, dashed line), but it might be anomalous by some degree (Fuzzy case, solid line).