RESEARCH ARTICLE

Laura Sandri · Warner Marzocchi · Lucia Zaccarelli

A new perspective in identifying the precursory patterns of eruptions

Abstract The complexity of the processes responsible for volcanic eruptions makes a theoretical approach to forecasting the evolution of volcanic unrest rather difficult. A feasible strategy for this purpose appears to be the identification of possible repetitive schemes (patterns) in the pre-eruptive unrest of volcanoes. Nevertheless, the limited availability and the heterogeneity of pre-eruptive data, and the objective difficulty in quantitatively recognizing complex pre-eruptive patterns, make this task very difficult. In this work we address this issue by using a pattern recognition approach applied to the seismicity recorded during 217 volcanic episodes of unrest around the world. In particular, we use two non-parametric algorithms that have proven to give satisfactory results in dealing with a small amount of data, even if not normally distributed and/or characterized by discrete or categorical values. The results show evidence of a longer period of instability in the unrest preceding an eruption, compared to isolated unrest. This might indicate, even if not necessarily, a difference in the energy of processes responsible for the two types of unrest. However, if the unrest is followed by an eruption, it seems that the seismic energy released during the unrest (parameterized by the duration of the swarm and the maximum magnitude recorded) is not indicative of the magnitude of the impending eruption. We also found that, in general, unrest followed by the largest explosive eruptions have a longer repose time than those related to moderate eruptions. This evidence supports the fact that the occurrence of a large eruption needs a sufficient amount of time after the last event in order to re-charge the feeding system and to achieve a closed-conduit regime so that a sufficiently large amount of gas can be accumulated.

L. Sandri (✉) · W. Marzocchi · L. Zaccarelli
INGV Bologna, Via Donato Creti 12, 40128 Bologna, Italy
e-mail: sandri@bo.ingv.it
Tel.: +39-051-4151422
Fax: +39-051-4151498

Introduction

Volcanic eruptions often have devastating effects. A basic step towards the mitigation of their consequences consists of forecasting the time evolution of pre-eruptive unrest at volcanoes. The present state of knowledge of the complex physical process responsible for the volcanic eruptions makes a theoretical approach to forecasting rather difficult. In this situation, the empirical identification of possible repetitive schemes (patterns) in the pre-eruptive unrest of volcanoes may represent a viable strategy to improve significantly our forecasting capability and our physical knowledge of the system. The potential identified patterns might not only indicate whether the unrest is evolving into an eruption but also provide an estimation of the energy associated with the volcanic eruption, for instance the Volcanic Explosivity Index (VEI).

Volcanic unrest has a complex nature, involving different interactive processes. An almost complete picture of the phenomena consists of a large variety of different signals, for example different seismicity variables, deformation, gas emission, and so on. A robust technique that aims to identify possible pre-eruptive patterns in volcanic unrest has to take into account all, or at least some, of these measurements simultaneously. This necessity has been recently discussed by Sparks (2003) as the key to successful forecasting, and it is the basic concept underlying many studies made by expert volcanologists in dealing with volcanic unrest (see, for instance, the volume by Newhall and Punongbayan 1996 and, in particular, Harlow et al. 1996; Voight et al. 1999; Hill et al. 2002). In these studies, to set some empirical rules mainly based on human experience, different parameters are actually taken into account simultaneously. Here, the human brain of the expert volcanologist works as a qualitative pattern recognition or

neural network code, identifying empirical multivariate patterns from past experience in order to establish some rules for future pre-eruptive unrest phases. However, even if the human brain is by far much more flexible and elastic than a computer in applying a pattern recognition or neural network code, it has some shortcomings that might limit the use of the human experience and favor the use of computer codes (e.g., Cammarata 1997). In particular, the human brain can hardly deal simultaneously with 3 or more variables. In other words, the higher the number of variables considered, the lower the capability of the human brain in detecting patterns (even simple ones). Also, the personal and subjective experience of a scientist is usually more difficult to extend to the rest of the volcanological community than rules obtained through a quantitative and reproducible approach. Moreover, the latter produces results that are definitely much easier to submit to rigorous scientific validation. Because of these reasons, it might also be useful to consider a computer-based approach, looking for strict objective and quantitative rules.

In fact, quantitative rules have been already sought in forecasting eruptive activity. These studies are very often based on the retrospective analysis of a single volcanic event (e.g., Shibata and Akita 2001; Gottsmann and Rymer 2002), or of events from a single volcano (e.g., Aki and Ferrazzini 2000; Londoño and Sudo 2002) that almost always take into account only one variable and not a multivariate dataset. The approach based on a single volcano, and above all, on a single eruptive episode, has an intrinsically strong limitation: the analysis, even though detailed, does not allow for the discrimination of the general pre-eruptive patterns from the peculiarities of the volcano or eruption considered. However, possible general pre-eruptive patterns are definitely the most important ones because they contain information useful for improving the knowledge of the physics of the erupting system. At the same time, their identification may furnish quantitative rules that can be profitably used to forecast the time evolution of the unrest in other volcanoes. In practice, in fact, we often have to cope with very dangerous volcanoes, for instance Mount Vesuvius, where no quantitative measurements relative to past pre-eruptive phases are available. In such cases, it becomes very difficult to understand when an unusual behavior of the volcano is really linked to an impending large eruption. Hence, the common experience acquired in other erupting explosive volcanoes (e.g., Shimozuru 1972; McNutt 1996) becomes the most relevant information.

The idea of this paper is to look for quantitative and complex (i.e., coming from a multivariate dataset) pre-eruptive patterns common to many different volcanic areas of the world. The main difficulties in reaching this goal are (1) the scarce availability and the incompleteness of pre-eruptive data, and (2) the ability of the methods used in objectively recognizing possible complex and quantitative pre-eruptive patterns.

As regards point (1), we have to collect a sufficient number of multivariate data coming from different volcanoes. In fact, some effort has been dedicated until now to collecting multivariate pre-eruptive data coming from different volcanoes into a single dataset. Remarkable examples are the catalogs compiled by Newhall and Dzurisin (1988) and Benoit and McNutt (1996), just to mention a few. The latter provides seismic data, in a time period of 10 years, relative to pre-eruptive phases on more than 100 volcanoes. In spite of the huge effort made by the authors, the collected data are rather rough, being very often categorical, strongly heterogeneous (for instance, the magnitude measurements) and in some cases qualitative.

In our opinion a definite improvement in this field can be achieved only through an international and coordinated effort, such as the WOVOdat project (www.wovo.org/wovodat.htm).

As regards point (2), the main difficulty is predominantly technical. A complication arises in the fact that the available data are usually few, often categorical, correlated, and their statistical distributions are seldom Gaussian (see Benoit and McNutt 1996). This precludes the use of all the parametric multivariate techniques successfully used in many other scientific fields (e.g., Fukunaga 1990). Note that, when few data are available, neural network codes cannot be used either, because they need a large amount of data to correctly perform the training, validating and testing steps (Tarassenko 1998). Furthermore, in this paper we are mainly interested in the physical meaning of the possible common patterns. With a neural network approach, we might be able to classify different types of unrest, but it would not be clear what the physical rules are that allow the network to discriminate between different types of unrest.

In this work, we provide a possible strategy of analysis that properly takes into account all of the issues discussed above. In particular, we apply two different non-parametric pattern recognition codes to search for common pre-eruptive patterns in a catalog containing data recorded during several volcanic unrest episodes around the world. The dataset consists of the Benoit and McNutt catalog (1996) plus all the available information concerning the seismic swarms related to the largest (VEI$\geq$4) explosive eruptions that occurred during the last century, and other episodes of volcanic unrest for which information was found in the literature. The main goal is to provide new insights concerning the following questions:

– Do the seismic unrest episodes occurring before volcanic eruptions have common patterns?
– Do these possible common patterns reflect the magnitude of the following eruption?
– Do these possible common patterns reflect the type of the following eruption or the initial state of the conduit (closed or open)?

Although the catalog used is certainly the best available, its intrinsic quality may still be too insufficient (few data, with missing measurements) for obtaining quantitative and useful rules to forecast volcanic events. At the same time, it is possible to achieve interesting

Table 1 This is a part of the dataset of seismic swarms we used in this work. In particular, all the largest post-1950 eruptions (VEI≥4) included are shown. A negative value for MXM, PRE and/or TRE means that a measurement is missing

Volcano name	VEI	Swarm date yyyy mm dd	DUR d	MXM	REP y	PRE 0=n, 1=y	TRE 0=n, 1=y	PHI
Bezymianny	5	1955 10 11	172	4.4	1000	0	1	7.52 E 0
Shiveluch	4	1964 11 2	10	4.9	14.6	1	1	9.07 E −1
Fernandina	4	1968 6 11	16	5.4	0.05	−9	1	7.85 E −3
Plosky Tolbachik	4	1975 6 27	9	0.5	0.5	−9	1	3.13 E 0
Augustine	4	1976 1 21	0.625	0.3	4.28	−9	−9	3.51 E 0
St. Helens	5	1980 3 20	59	5.0	123	0	1	1.36 E −2
Alaid	4	1981 4 26	22	3.5	8.5	−9	0	3.58 E 0
Pagan, north	4	1981 4 1	45	4.0	56	−9	1	3.61 E −3
El Chichon	5	1982 3 1	28	4.0	132	1	1	7.17 E −2
Galunggung	4	1982 4 4	0.260	−9.0	63.7	−9	1	3.14 E −3
Colo (Una Una)	4	1983 7 4	24	4.6	45	−9	1	1.23 E −1
Augustine	4	1986 2 10	45	2.1	8.7	1	−9	3.50 E 0
Kelut	4	1989 11 15	90	2.0	21.9	−9	1	8.15 E −3
Pinatubo	5	1991 4 5	67	4.3	675	0	1	5.37 E 0
Spurr	4	1991 8 15	180	1.7	38.1	−9	1	1.98 E +1
Rabaul	4	1994 9 18	1.125	5.1	50.7	1	−9	3.45 E 0

scientific insights for improving our knowledge of the physics of the eruptive processes. In any case, independent from the scientific results obtained here, an ambitious aim of this paper is to introduce a new quantitative perspective in approaching the eruption forecasting issue. As soon as a worldwide catalog of volcanic unrest of good quality is available, the strategy of analysis described here can provide a very powerful tool in the recognition of quantitative rules for forecasting the temporal evolution of unrest in volcanic areas.

The dataset

The bibliography about volcanic eruptions dates back to historical times, but only for the catastrophic events (such as Vesuvius, 79 A.D.). Furthermore, in these cases, the reports are purely qualitative: morphological descriptions, eruptive products and descriptive temporal evolution. This information is not particularly useful for a quantitative approach to eruption forecasting. Thanks to the evolution in instrumentation, in the last few decades quantitative investigations have begun together with geophysical data reports. It is now possible to find a great number of different data relative to volcanic unrest in the form of seismological records, deformation measures, temperature or magnetic field variation detections, and so on. Among these, seismological information are the most available and reliable, mainly because of the great diffusion of seismometers compared to that of other instrumentation. Furthermore, among all the precursors of an eruption, volcanic earthquakes almost always characterize periods of volcanic unrest. For these reasons, we concentrated our study on the seismic data relative to the episodes of unrest which occurred in the last 50 years.

The dataset that we collected and analyzed consists of measurements relative to 217 seismic swarms in volcanic areas (see Table 1). For each swarm, we collected as many measurements as possible that are potentially related to the occurrence of a volcanic eruption and/or to the estimation of its VEI, in case an eruption occurs. Concerning the seismicity that characterizes the unrest followed by either low explosive events or not followed by eruptive activity (we call the latter isolated unrest), we mainly referred to Benoit and McNutt (1996). These two types of unrest are quite frequent, and in that catalog we were able to find a sufficient number of cases to be analyzed.

On the other hand, due to the rarity of this type of the event, many more difficulties were encountered concerning the seismicity that characterizes the unrest before the most explosive eruptions (VEI≥4). Because of this, further research was necessary.

For the VEI≥4 events from 1950 to 1994, we primarily consulted the *Volcanoes of the World* (Simkin and Siebert 1994). We complete the list up to 2001 through the personal communications of Lee Siebert.

One problem was where to find information about the seismicity, characterizing the unrest that preceded the VEI≥4 eruptions. We started by systematically seeking articles about these eruptive events since 1950. Some of the articles are hardly available (not being published in an easily accessible magazine, e.g., Taylor 1957). Some others, not being related to seismicity, were not useful for our purposes (e.g., Buell and Stoiber 1976). However, all the papers provided a bibliography containing a further list of references. As regards the last events of this century, we found the most interesting information on web sites.

For every useful article found, we had to be careful in giving the right meaning to the information since the author's interpretation can greatly influence the data. In order to obtain continuity among the different types of eruptions, acknowledging that they are subjective, we interpreted the data according to the definitions of seismic swarm and its duration used by Benoit and McNutt (1996) in their database.

To summarize, we consulted the database of Benoit and McNutt (1996), the catalog of Simkin and Siebert (1994), the Bulletin of Volcanic Eruptions (1963–90), and

the available literature on the large eruptions (VEI$\geq$4) of the last century, especially in its second half (Gorshkov 1959; Gorshkov and Dubik 1970; Simkin and Howard 1970; Zobin 1971; Reeder et al. 1977; Faberov 1983; Fedotov et al. 1983; Gorel'chik et al. 1983; Zobin 1983; Decker and Decker 1981; Jensen et al. 1983; Tokarev 1985; Swanson and Kienle 1988; Smithsonian Institution's Global Volcanism Network 1990; Miller and McGimsey 1998; Paolo Papale, personal communication, 2002; http://www.volcano.und.nodak.edu; http://www.volcano.si.edu; http://www.vulcan.wr.usgs.gov).

For each swarm, we found measurements of the following variables:

– The duration (DUR) of the swarm (in days)
– The repose time (REP) associated to the swarm, i.e., the time (in years) elapsed between the end of the last eruption and the beginning of the swarm
– The maximum magnitude (MXM) recorded in the swarm
– A binary indicator (PRE) of the occurrence of a previous swarm (0=no, 1=yes)
– A binary indicator (TRE) of the occurrence of volcanic tremor (0=no, 1=yes)
– The ϕ function value (PHI). Considering the k-th swarm occurring in a certain volcanic area, PHI ($\phi^{(k)}$) is a perturbation function (Marzocchi 2002) that mimics the stress induced on this volcanic system by all the large remote earthquakes that occurred in the 35 years preceding the k-th swarm. In particular:

$$\phi^{(k)} = \sum_{j=1}^{N} M_{0j}\omega(d_{jk}) \tag{1}$$

where N is the number of earthquakes that occurred in the 35 years preceding the onset time of the k-th swarm, M_{0j} is the seismic moment of the j-th earthquake and $\omega(d_{jk})$ is a weight function dependent on the relative distance between the location of the k-th swarm and the epicenter of the j-th earthquake (see Fig. 3 in Marzocchi 2002). The seismic data are taken from the catalog of Pacheco and Sykes (1992) for the period 1900–1989, and from the CMT Harvard catalog (Dziewonsky et al. 1981; Dziewonsky and Woodhouse 1983) for recent years. The earthquakes considered are the events with $M_s \geq 7$ and depth ≤ 70 km.

While DUR, PHI and REP have been retrieved for almost the totality of the catalog, PRE, TRE and MXM retrieving has been much more difficult (see Fig. 1). The magnitude measurements are in different scales for every country, and have been assumed to be consistent. For the parameters TRE and PRE, a 1 value simply means that some information regarding the feature has been reported, as in Benoit and McNutt (1996). If a report states that, for example, "TRE measurements have been conducted," the TRE feature is set to 1 in Benoit and McNutt (1996), which we have also done in our dataset, regardless of the occurrence of TRE. A 0 value means that a negative result on the occurrence of TRE or PRE was reported.
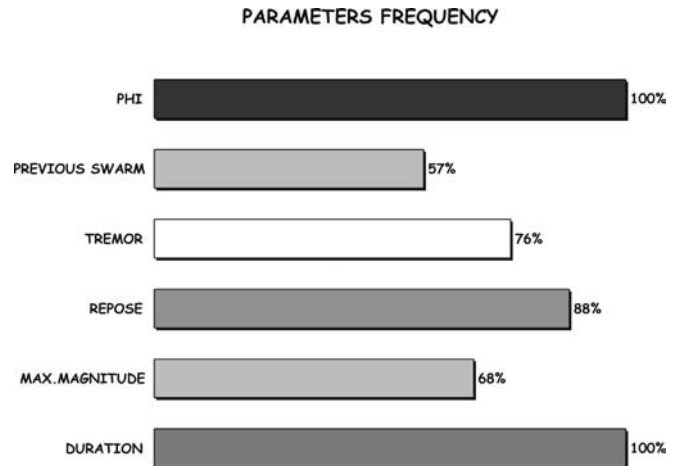

Fig. 1 Relative frequencies of the features in the dataset

The measurements which could not be retrieved are set to a number standing for a missing value that will not be used in the analysis. As a final remark, we emphasize that the resulting catalog, even though it is still rough and needs further improvement, is certainly the largest one available at present.

Pattern recognition analysis

Pattern recognition (PR) is a set of very powerful multivariate analysis techniques allowing, in principle, the identification of possible repetitive schemes or patterns among the objects belonging to distinct classes. While usual data analysis takes into account only one variable of the process at a time, PR methods are able to extract information from any possible combination (linear or not) of variables that are suspected to have an influence on the process. Moreover, PR methods do not need the construction of a theoretical model, but are usually based on a basic and sole hypothesis, i.e., the assumption that the phenomenon under study is governed by a finite number of complex, but repetitive patterns of the variables.

For these appealing properties, we believe that PR might also be a very promising tool in earth science. Until now, the only few remarkable efforts in this direction are CN and M8 algorithms (Keilis-Borok et al. 1988; Keilis-Borok and Kossobokov 1990), and applications to volcanology (Mulargia et al. 1991,; Vinciguerra et al. 2001). Most of these algorithms, including CN and M8, are based on a different type of PR analysis: the so-called logical PR. This type of PR analysis requires the arbitrary choice (by the user) of several parameters influencing the behavior of the algorithm. Because of this, the risk of overfitting the data increases drastically. Furthermore, any systematical evaluation of how the values chosen for the parameters influence the performance of these algorithms has not as yet been conducted. For these reasons, in this study we prefer using different algorithms based on a different approach: the so-called statistical PR. The

algorithms belonging to this category do not need the selection of parameter values by the user.

From a technical point of view, the main goal of PR methods is to classify objects. Every object is represented by an array of qualitative or quantitative variables. The procedure of analysis consists of three different steps: the learning phase, the voting phase, and the control experiments. In the learning phase, a set of known and classified objects is analyzed in order to recognize all the possible patterns that characterize each class, i.e., the combinations of variables that allow for the discrimination of objects belonging to different classes. This step turns out to be very useful also from a theoretical point of view, since it allows for the recognition, among all the suspected variables, of those that really play an important role in the process under study. In the voting phase, the patterns identified during the learning are used to classify new objects, whose class is unknown to the algorithm. Finally, the control experiments allow one to check the stability of the results by repeating the learning and the voting phases with different values of the algorithm's parameters.

In the present study, the main goal of the analysis is to recognize the prominent characteristic of the seismic swarms preceding a volcanic eruption and to find possible relationships with the VEI of the impending eruption. Due to the very limited amount of data available, in this paper we will perform only the learning phase, and attempt to recognize, as a first step, all the possible patterns in our dataset. In spite of the impossibility of testing the results on independent data (voting), we have used some empirical strategies to check the presence of possible overfit in the results.

Before performing the learning phase, we first had to:

1. Define the objects to be analyzed and the classes involved in the problem, and
2. Select the statistical PR algorithm that is most suitable to the problem we are dealing with

We shall explain these two steps more accurately in the following.

Definition of the objects and of the classes

The objects of the analysis are the seismic swarms. Any object is represented by a vector that contains all the measurements (the features) that we can associate to the object. Due to the large differences between the maximum and minimum measurements in the catalog for DUR, REP and PHI, we decided to use the logarithm of these features. Thus, each vector has the following components: Log (DUR), Log (REP), MXM, PRE, TRE and Log (PHI).

Each vector has then a further component: it is the VEI associated to the eruption (if any) following the swarm described by the vector. If the swarm has not been followed by an eruption, a fictitious VEI is associated to it equal to -1. In this paper, the attribution of an object to its class depends on the VEI of the subsequent eruption (if any). Since we had eight different values for the VEI (-1, 0, 1, 2, 3, 4, 5, 6) in our catalog, in principle we had eight different classes of objects. For simplicity, the VEIs of the swarms were be grouped in order to reduce the problem to a two-class problem, i.e., class 1 versus class 2. We kept at least one unit of VEI between the lowest VEI of the upper class and the highest VEI of the lower class. For example, in order to find patterns that distinguish a swarm preceding a small eruption from one preceding a large eruption, we considered as class 1, all the swarms with VEI$\geq$4, and as class 2, all the swarms with 0$\leq$VEI$\leq$2. The VEI=3 events were excluded to emphasize the distinction between the classes. In this way we avoided more safely, with no loss of generality, any kind of overlapping between the classes. Note that we are interested in the most general features distinguishing the two classes.

The complete list of the various analyses (class 1 vs. class 2) performed is provided in the following.

Selection of the most suitable statistical PR algorithms

In this paper, we will try to identify repetitive patterns between two distinct categories of objects. Many statistical PR 2-class algorithms, both parametric (e.g., maximum likelihood estimation, see Duda and Hart 1973) and non-parametric (e.g., binary decision tree, Fisher's analysis, K-nearest neighbors, linear or quadratic discriminant analysis, see Rounds 1980; Duda and Hart 1973; Fukunaga 1991), have been successfully used in other scientific fields such as engineering, biology, economy, medicine. In these disciplines, the available datasets are large and continuous, and the variables are normally distributed.

Our dataset, as well as most of the datasets in earth sciences, do not have these "nice" features. In particular, it is composed of a small amount of data, some of the variables (if not all) are not normally distributed (e.g., the duration of the swarm, and the occurrence of previous swarms and tremor), and some might be also correlated (e.g., the duration and the maximum magnitude). Moreover, some of the variables we have collected in the catalog are probably completely irrelevant to the eruptive process. Indeed, we compiled our catalog by taking the largest possible number of potentially relevant variables available for each seismic swarm, because we did not know which (if any) of these variables are important for the subsequent occurrence of a volcanic eruption, or for the determination of the VEI of that eruption.

As a result, we needed to use a statistical PR algorithm that could perform satisfactorily on small datasets and is characterized by continuous and discrete or categorical variables that are perhaps correlated. Possibly, we are including in the analysis some variables which do not affect the eruption occurrence or its VEI, thus it was necessary to make use of a statistical PR algorithm that was able to extract those variables having a predominant influence on the processes related to volcanic unrest. According to these considerations, in this work we used

two statistical PR 2-class algorithms that we had previously simulated using synthetic data and that had proved capable of recognizing patterns satisfactorily on small datasets, also with correlated and/or discrete (also categorical) data, and identifying the variables having a predominant role in the process (Sandri and Marzocchi 2003). These two non-parametric algorithms are called binary decision tree (BDT; Rounds 1980; Mulargia et al. 1992) and Fisher discriminant analysis (FIS; e.g., Duda and Hart 1973). The use of both algorithms, based on very different approaches, allowed us to check if the results that we obtained are due to the type of algorithm used. Although the risk of overfit can be excluded only by voting a set of independent data, the stability of the results obtained by these two different algorithms is indirect evidence that the risk of overfit is reduced.

Algorithm BDT was originally designed for hierarchically ordered data, but it has also exhibited very good performance on different data. It builds up a decisional tree where, at each level, a threshold value for a certain variable determines which branch has to be followed. BDT provides automatically the subset of variables playing an important role in the process.

Algorithm FIS is a non-parametric method because, although it assumes that the boundary between the classes is a hyperplane, it does not make any a priori assumption on the distribution of the data. It is a type of linear discriminant analysis, in which the original data are projected along a direction maximizing the ratio of the dispersion between the two classes to the dispersion inside each class. This algorithm, according to Fukunaga (1990), is here applied through a so-called branch-and-bound technique in order to identify the relevant features of the process. The feature selection performance by the branch-and-bound technique has been previously tested on synthetic data as well.

For a more complete definition of the algorithms and of the branch-and-bound technique, see Appendixes A, B and C.

Results of the analysis and discussion

We performed three different 2-class analyses with different goals. In particular they are:

1. *VEI≥1 vs. VEI=−1* where class 1 is represented by all the swarms followed by a volcanic eruption (VEI≥1)

and class 2 by all the isolated swarms (VEI=−1). This analysis was done in order to recognize the general differences between the swarms preceding a volcanic eruption and the isolated swarms.
2. *VEI≥4 vs. VEI=−1* where class 1 is represented by all the swarms followed by a strongly explosive eruption (VEI≥4) and class 2 by all the isolated swarms (VEI=−1). This analysis was done in order to recognize the differences between the swarms preceding a strong explosive volcanic eruption and the isolated swarms.
3. *VEI≥4 vs. 0≤VEI≤2* where class 1 is represented by all the swarms followed by a strong explosive eruption (VEI≥4) and class 2 by all the swarms followed by moderate eruptions (0≤VEI≤2). This analysis was done in order to recognize the differences between the swarms preceding a strongly explosive volcanic eruption and the swarms preceding small or moderate eruptions.

In each analysis, we used only "complete" objects, i.e., the objects having no missing values for the features considered in the analysis. We start by considering all the six features. Due to the missing measurements, the analysis that considers all the six features was carried out on a low number of objects (see Tables 2, 3, 4). In order to perform the analysis on a higher number of objects, and to check the stability of the results obtained on different learning datasets, we performed two additional learning phases that concern a smaller number of features. In particular, we repeated the statistical PR analysis concerning (1) DUR, REP, MXM and PHI, and (2) DUR, REP and PHI (see Tables 2, 3, 4). The choice of the features in (1) and (2) was due to their more common reporting (allowing for a larger number of complete objects) and to their importance in the process as suggested by the analysis carried out on all the six features (see below).

Since we were interested in the recognition of possible patterns in our swarms dataset, in each analysis we used all of the available complete objects for the learning phase. This allowed us to make use of as much data as possible to define the patterns in the data.

VEI≥1 vs. VEI=−1

As shown in Table 2, both algorithms recognize the DUR as the predominant variable for the discrimination

Table 2 VEI≥1 (class 1) vs. VEI=−1 (class 2). In the first column, the features used in the learning phase are shown; in the second column, the number of available objects with no missing values for any of the features used is shown for each class; in the two last columns, the features recognized as relevant are reported (third column, for BDT algorithm, fourth column for FIS algorithm). An empty box means that the algorithm does not recognize any pattern

Features analyzed	Number of complete objects	Relevant features identified by BDT	Relevant features identified by FIS
DUR, MXM, REP, PRE, TRE, PHI	Class 1: 22; Class 2: 13	DUR	DUR, REP
DUR, MXM, REP, PHI	Class 1: 46; Class 2: 85	DUR	DUR
DUR, REP, PHI	Class 1: 66; Class 2: 121	DUR	

Table 3 VEI≥4 (class 1) vs. VEI=−1 (class 2). In the first column, the features used in the learning phase are shown; in the second column, the number of available objects with no missing values for any of the features used is shown for each class; in the two last columns, the features recognized as relevant are reported (third column for BDT algorithm, fourth column for FIS algorithm)

Features analyzed	Number of complete objects	Relevant features identified by BDT	Relevant features identified by FIS
DUR, MXM, REP, PRE, TRE, PHI	Class 1: 5; Class 2: 13	DUR	DUR, PHI
DUR, MXM, REP, PHI	Class 1: 17; Class 2: 85	MXM	DUR
DUR, REP, PHI	Class 1: 18; Class 2: 121	REP, DUR	REP

Table 4 VEI≥4 (class 1) vs. 0≤VEI≤2 (class 2). In the first column, the features used in the learning phase are shown; in the second column the number of available objects with no missing values for any of the features used is shown for each class; in the two last columns, the features recognized as relevant are reported (third column for BDT algorithm, fourth column for FIS algorithm). An empty box means that the algorithm does not recognize any pattern

Features analyzed	Number of complete objects	Relevant features identified by BDT	Relevant features identified by FIS
DUR, MXM, REP, PRE, TRE, PHI	Class 1: 5; Class 2: 12	REP	
DUR, MXM, REP, PHI	Class 1: 17; Class 2: 17		REP
DUR, REP, PHI	Class 1: 18; Class 2: 34	REP	REP, PHI

between class 1 and class 2. In particular, swarms preceding a volcanic eruption are generally longer than isolated swarms. This agrees with the results obtained by Benoit and McNutt (1996). Due to the very limited amount of data available, the parameters of the pattern (i.e., the thresholds in DUR by which the algorithms classify a swarm as pre-eruptive or isolated) have a very large uncertainty. However, just to give an idea of the magnitude, the thresholds are in the order of a few days (a week). For example, Figs. 2 and 3 show the case in which four features (DUR, MXM, REP and PHI) are considered in the analysis for BDT and FIS algorithms, respectively. The BDT plot (Fig. 2) is very intuitive. The FIS plot (Fig. 3) instead needs a little explanation. In Fig. 3, we plotted the frequency of the learning objects (class 1 and class 2 separately) as a function of the pattern found, i.e., the combination of relevant variables identified. In this case, it is only Log (DUR). The data shown are standardized (mean and variance values are given in the figure caption). Should a new object have to be voted, we should first standardize it, then project it along Fisher's criterion line (in this case it is simply the standardized Log (DUR) axis). Then, the new object to be voted will be attributed to class 1 (precursory swarm) if it falls to the right of the decision boundary (i.e., if its standardized Log (DUR) is larger than 0.075), otherwise it will be attributed to class 2.

We interpreted the longer pre-eruptive swarm duration as an indication of a prolonged instability during pre-eruptive unrest. We did not observe any significant difference in the magnitude of the earthquakes among the episodes of unrest belonging to the two classes. Because of this, if we assume that the seismic rate among the episodes of unrest is comparable, we might interpret the prolonged duration of precursory unrest as an indication of higher energy involved.
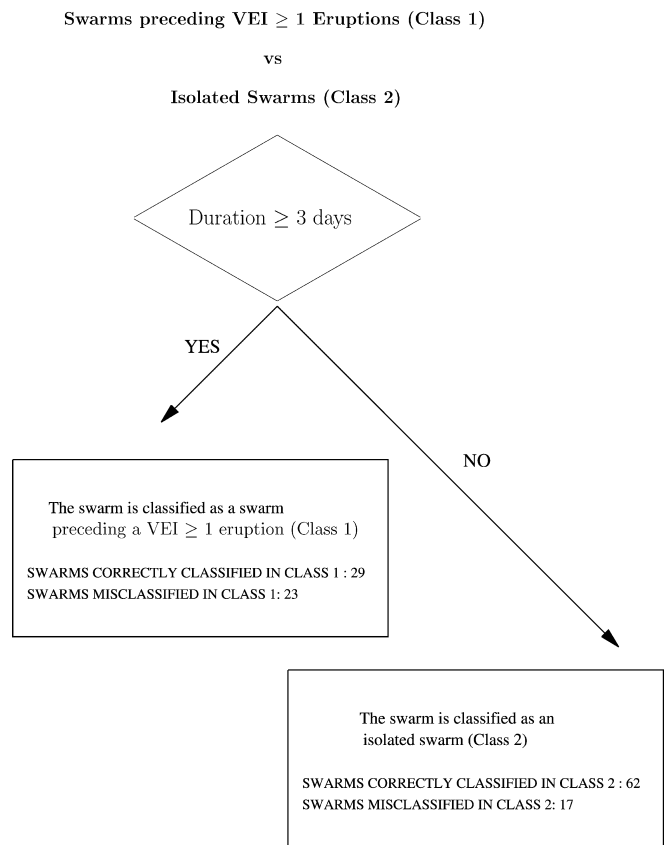

Fig. 2 Pattern recognition results for BDT relative to the first experiment, in which class 1 is given by swarms preceding VEI≥1 eruptions and class 2 by isolated unrest when the features DUR, MXM, REP and PHI are considered. The classification was performed on the basis of the DUR and of the threshold value indicated. In total, 91 swarms are correctly classified and 40 are not
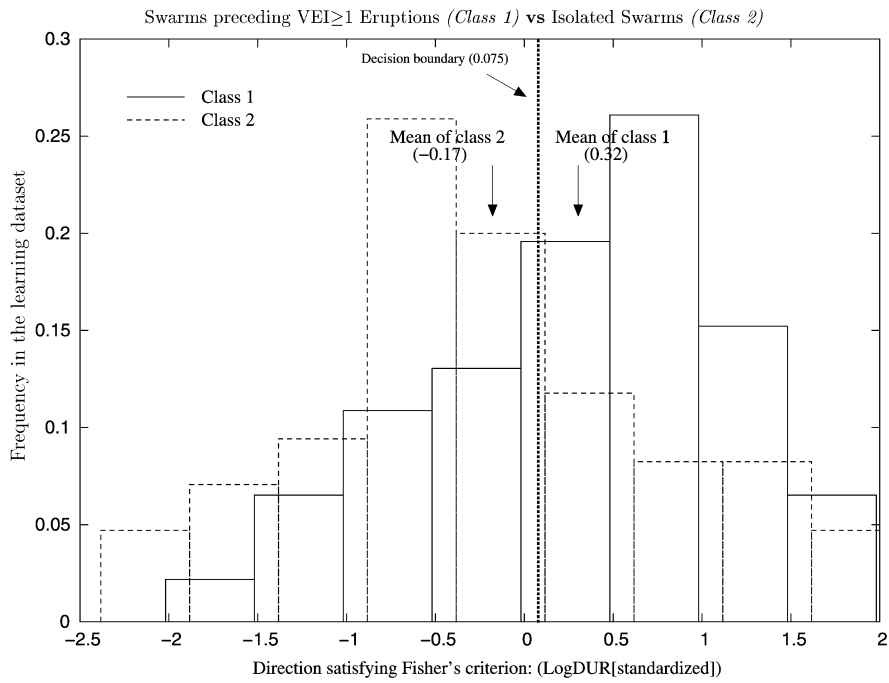
Swarms preceding VEI≥1 Eruptions *(Class 1)* **vs** Isolated Swarms *(Class 2)*

Fig. 3 Pattern recognition results for FIS relative to the first experiment in which class 1 is given by swarms preceding VEI≥1 eruptions and class 2 by isolated unrest when the features DUR, MXM, REP and PHI are considered. The classification was performed on the basis of DUR, which is the only relevant variable identified by FIS algorithm in this case. The frequency of the learning objects is given as a function of this variable. The mean of Log (DUR) for the whole set of learning data is 2.24, and the variance is 4.5. The standardized voting objects should be attributed to class 1 if they are larger than 0.075; otherwise, class 2. Out of our learning swarms, 90 swarms are correctly classified and 41 are not
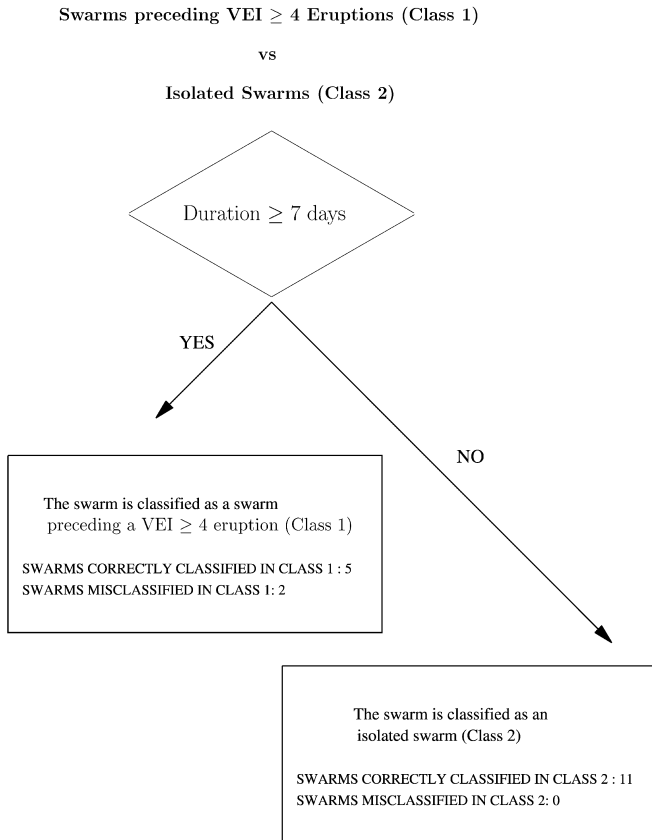
VEI≥4 vs. VEI=−1

As shown in Table 3, here we find again that the predominant variable is DUR for both algorithms. A swarm preceding a large explosive eruption is generally longer than an isolated swarm. The same considerations regarding the parameters of the pattern made above apply here. Figures 4 and 5 are shown as examples for the BDT and FIS, respectively, when all six features are considered in the analysis. In this particular case, FIS recognized two relevant features (the second one is PHI). Should we need to vote a new swarm, we should first standardize its Log (DUR) and Log (PHI) according to the means and variances given in the caption of Fig. 5. Then, we should project it along Fisher's criterion line, which is a linear combination of the relevant variables identified (given in the *x*-axis of Fig. 5). Finally, the object will be attributed to class 1 if its standardized and projected value is larger than 0.40, otherwise it will be attributed to class 2. Even if FIS recognizes two relevant variables, the largest part of the discriminating capability in Fig. 5 is given by DUR

Fig. 4 Same as in Fig. 2 but relative to the case in which class 1 is given by swarms preceding VEI≥4 eruptions and class 2 by isolated unrest when all the features are considered. The classification was performed on the basis of the DUR and of the threshold value indicated. In total, 16 swarms are correctly classified and 2 are not
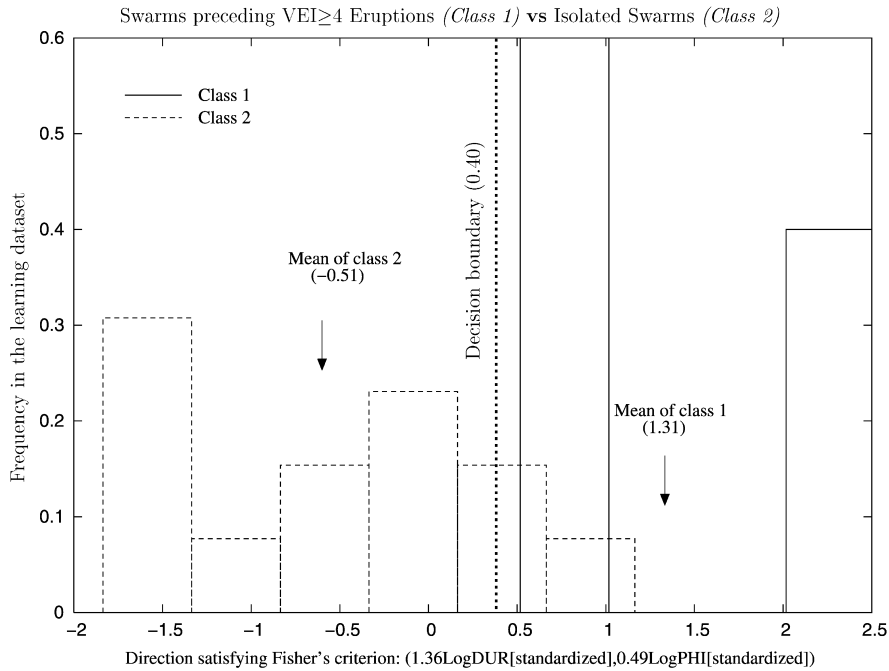
Fig. 5 Same as in Fig. 3, but relative to the case in which class 1 is given by swarms preceding VEI≥4 eruptions and class 2 by isolated unrest when all the features are considered. The classification was performed on the basis of DUR and PHI, which are the relevant variables identified by FIS algorithm in this case. The mean of Log (DUR) for the whole set of learning data is 1.63, and the variance is 5.2. The mean of Log (PHI) for the whole set of learning data is −0.74, and the variance is 7.2. The standardized voting objects should be projected along the Fisher's criterion line, i.e., their DUR and PHI measurements should be linearly combined as $x=1.36$ Log (DUR)+0.49 Log (PHI). Then, they should be attributed to class 1 if their projected values are larger than 0.40; otherwise, class 2. In total, 17 swarms are correctly classified and 1 is not
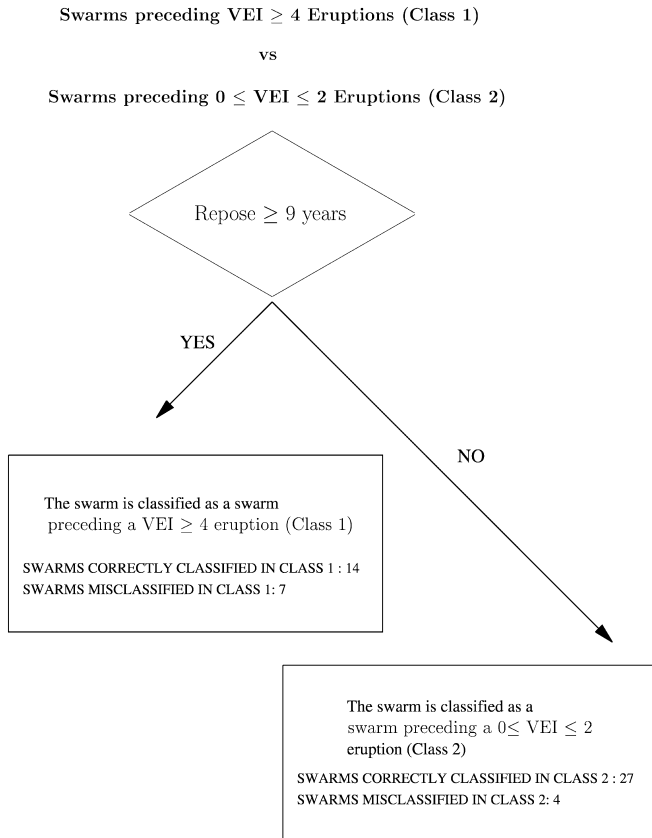
(see the much larger coefficient for Log (DUR), compared to the one for Log (PHI), in Fisher's criterion line).

Again, we interpreted this result as an indication of the prolonged instability during unrest that precedes large explosive eruptions, compared to isolated unrest. The same considerations for the above section apply here.

VEI≥4 vs. 0≤VEI≤2

As shown in Table 4, in this case, there is no evidence of magnitude or duration difference in these two seismic swarm types, which suggests that the intrinsic characteristics of the seismic swarm may not be indicative of the eruption magnitude. This result agrees with statements from Newhall and Hoblitt (2002). Here, the only (or the most) relevant feature identified by both algorithms is REP (see Table 4). Generally, the swarms corresponding to the most explosive eruptions have a longer repose time than those related to moderate eruptions (Simkin and Siebert 1994; Newhall and Hoblitt 2002), as shown in Figs. 6 and 7 (for the case in which DUR, REP and PHI

Fig. 6 Same as in Fig. 2 but relative to the case in which class 1 is given by swarms preceding VEI≥4, class 2 by swarms preceding 0≤VEI≤2 eruptions when the features DUR, REP and PHI are considered. The classification was performed on the basis of REP and of the threshold value indicated. In total, 41 swarms are correctly classified and 11 are not
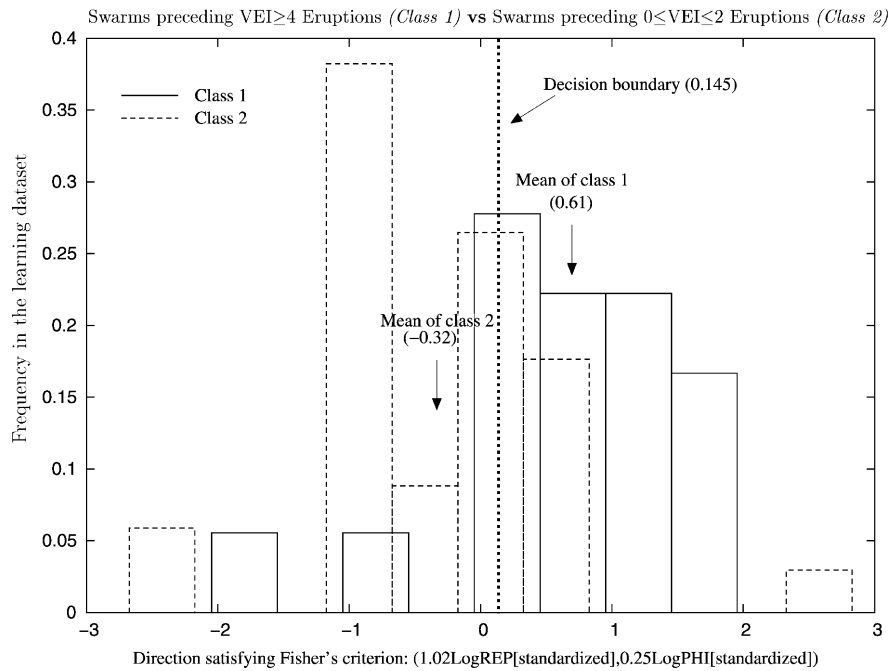
Fig. 7 Same as in Fig. 3 but relative to the case in which class 1 is given by swarms preceding VEI≥4, class 2 by swarms preceding 0≤VEI≤2 eruptions when the features DUR, REP and PHI are considered. The classification was performed on the basis of REP and PHI, which are the relevant variables identified by FIS algorithm in this case. The mean of Log (REP) for the whole set of learning data is 1.77, and the variance is 7.2. The mean of Log (PHI) for the whole set of learning data is −1.76, and the variance is 10.5. The standardized voting objects should be projected along the Fisher's criterion line, i.e., their DUR and PHI measurements should be linearly combined as x=1.02 Log (REP)+0.25 Log (PHI). Then, they should be attributed to class 1 if their projected values are larger than 0.145; otherwise to class 2. In total, 41 swarms are correctly classified and 11 are not

are considered in the analysis for BDT and FIS, respectively). In this case, FIS recognizes both REP and PHI as relevant variables. Should a new object need to be voted, we should first standardize its Log (REP) and Log (PHI) according to the mean and variance given in the caption of Fig. 7. Then, we should project it along Fisher's criterion line, which is a linear combination of the relevant variables identified (given in the x-axis of Fig. 7). Finally, the object will be attributed to class 1 if its standardized and projected value is larger than 0.145, otherwise, class 2. Even if FIS recognizes two relevant variables, the largest part of the discriminating capability in Fig. 7 is given by REP (see the much larger coefficient for Log (REP), compared to the one for Log (PHI), in Fisher's criterion line).

A long repose time might indicate that the volcano system had sufficient time, since the last eruption, to recharge the system and to achieve the closed-conduit regime. In this way the volcano can accumulate a sufficient amount of gas to give a large explosive eruption. Actually, according to Newhall and Decker (2002), most large eruptions are preceded by long repose times, but most long repose times are not followed by large explosive eruptions.

As in the two previous subsections, the parameters of the pattern have large uncertainties. However, the repose time typical for unrest followed by large explosive eruptions is of a magnitude of 10 years or longer.

Concluding remarks

The main goal of this paper is to identify common pre-eruptive patterns in worldwide volcanic unrest. For this purpose we applied non-parametric pattern recognition codes to a catalog of seismic data relative to seismic swarms recorded in volcanic areas. The use of two algorithms based on very different "philosophies" allows for a checking of the stability of the results and a reduction of the risk of overfitting. We used seismic data because they were the easiest to retrieve and because seismic information is of prominent importance in characterizing unrest in volcanic areas.

The results obtained in this study are quantitative patterns distinguishing different types of volcanic unrest. However, the still poor quality of the dataset used does not allow us to use these quantitative patterns as profitable and satisfactory rules for eruption forecasting. In particular, the limited amount of data produces large uncertainties concerning the parameters of each pattern found, and does not allow us to evaluate the performance of the patterns, i.e., the percent of missed events and false alarms concerning an independent dataset.

In any case, the results reported here provide interesting insights into the physics of the pre-eruptive processes. In particular, there is evidence of a prolonged instability in pre-eruptive periods of unrest, compared to the isolated ones, both in consideration of only large explosive

eruptions (VEI $\geq$4) and all the eruptions with VEI$\geq$1. In considering that no significant difference is found in the maximum magnitude recorded in these two types of swarms, a longer seismic unrest might be interpreted as an indication of an energetic difference in the processes responsible for pre-eruptive and isolated swarms. On the contrary, no significant magnitude or duration difference is found between unrest episodes preceding large explosive eruptions (VEI$\geq$4) and moderate eruptions (0$\leq$VEI$\leq$2), which suggests that the energy released during precursory unrest is not a good indicator of the VEI of the impending eruption. This also may indicate that the magnitude of the eruption (i.e., the VEI) can be mostly due to random factors such as that for other complex systems like earthquakes, landslides, and so on (Bak et al. 1988). Here, although less evident, the only pattern found, compared to ones that precede small to moderate events, is based on a longer time of repose preceding the unrest occurring before the largest eruptions. The correlation to a longer repose for a large eruption might be linked to the time needed to re-charge the feeding system and to reach the state of a closed-conduit volcano. In this way, the volcano can accumulate a sufficiently large amount of gas to be able to give a large explosive eruption.

As a final consideration, we want to stress that the quality and the practical usefulness (eruption forecasting) of the results can be dramatically improved by using this kind of technique on large worldwide datasets of volcanic unrest such as the one proposed in the WOVOdat project.

Appendix A: Binary decision tree

This method was developed by Rounds (1980) and, slightly modified, successfully applied to volcanic data by Mulargia et al. (1992). It can be used only in the 2-class problem and it was originally designed for hierarchically ordered datasets, even though tests on synthetic data have shown very good behavior also on different types of datasets.

Once the data have been collected, and objects and classes have been defined, BDT integrates feature selection and binary decision tree according to the following steps:

1. The fixing of a level α for the decision rule. This level represents the risk we accept of a wrong attribution at each step. We used α=0.1.
2. The computation of the cumulative distribution in both classes for each feature taken one at a time, and the identification of the feature and the relative threshold value for which the statistical difference between the cumulative of the two classes is the largest. This means that the significance level of this statistical difference must be (a) lower than the level α and (b) lower than the significance level of the statistical difference calculated for any other feature. The feature (if any) for which both the (a) and (b) conditions are satisfied is the first-order feature, often called the "root" of the pattern. On the basis of the root feature and its threshold value, each object is assigned to either one of two subsets formed respectively by data with a value of the root feature lower/higher than the threshold.
3. The identification of the second-order features and their thresholds for which the statistical difference again satisfies the (a) and (b) conditions. These features, which are at most two (i.e., one for each subset), are found by reanalyzing all the features in the two subsets separately, as in step 2.
4. The repeating of step 3 for each second-order feature in order to identify progressively higher orders, as long as it is possible to find a feature for which the cumulatives in the two classes are statistically different at a significance level lower than α. The progressive branching of the tree gives all the possible patterns. The procedure automatically terminates when no further branching is possible at the given level α.

Steps 2–4 are performed by means of non-parametric Kolmogorov-Smirnov two-sample statistics (Hollander and Wolfe 1973). Note that the use of an a priori fixed level α reduces the possibility of obtaining overfitting patterns.

Appendix B: Fisher's discriminant analysis

This method (see e.g., Duda and Hart 1973) is based on the reduction of the n-dimensional space of the objects (where n is the number of variables describing the objects, i.e., the dimension of the vectors) to an L-1 dimensional space (where L is the number of classes). In our 2-class problem (L=2), Fisher's method simply projects the objects onto a line. The basic idea, called Fisher criterion, is to project the objects onto the direction that maximizes the ratio of the dispersion between the classes to the dispersion within the classes. More rigorously, suppose we have N objects $\mathbf{x}$, each represented by a vector consisting of n components x_k (k=1,...n). Of these, N_1 belong to class 1 and N_2 to class 2. We linearly combine the components of $\mathbf{x}$, i.e., the x_k (k=1,...n), in order to obtain a one-dimensional vector $\mathbf{y}$=(y):

$$y = w_k^T x_k \tag{2}$$

where w_k are the elements of an n-dimensional vector that projects $\mathbf{x}$ onto $\mathbf{y}$. In this way, we obtain N objects $\mathbf{y}$=(y) spread over the two classes.

The unknown in Eq.(2) is the projector, i.e., the vector $\mathbf{w}$. As mentioned above, we would like to choose the projection for which the ratio of the dispersion between the classes to the dispersion within the classes is

maximum. In order to do this, first we need to define some quantities.

We define $\mathbf{m}_i$ as the average vector of class i:

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{\mathbf{x} \in \text{class}\,i} \mathbf{x} \tag{3}$$

We also define $\mathbf{m}$ as the average of all the $\mathbf{x}$:

$$\mathbf{m} = \frac{1}{N} \sum_{\mathbf{x}} \mathbf{x} \tag{4}$$

Thus, the dispersion matrix within the class i is given by:

$$S_i = \sum_{\mathbf{x} \in \text{class}\,i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \tag{5}$$

and the dispersion within all of the classes is

$$S_w = \sum_{i=1}^{2} S_i \tag{6}$$

The total dispersion matrix is given by

$$S_T = \sum_{\mathbf{x}} (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T \tag{7}$$

It follows that

$$
\begin{aligned}
S_T &= \sum_{i=1}^{2} \sum_{\mathbf{x} \in \text{class}\,i} (\mathbf{x} - \mathbf{m}_i + \mathbf{m}_i - \mathbf{m})(\mathbf{x} - \mathbf{m}_i + \mathbf{m}_i - \mathbf{m})^T \\
&= \sum_{i=1}^{2} \sum_{\mathbf{x} \in \text{class}\,i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \\
&\quad + \sum_{i=1}^{2} \sum_{\mathbf{x} \in \text{class}\,i} (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \\
&= S_w + \sum_{i=1}^{2} N_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T
\end{aligned}
\tag{8}
$$

The second addendum of the right side term of Eq. (8) is a dispersion matrix S_b that gives an idea of the dispersion between the partial means $\mathbf{m}_i$ over the different classes and the total mean $\mathbf{m}$:

$$S_b = \sum_{i=1}^{2} N_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \tag{9}$$

In order to achieve the vector $\mathbf{w}^*$ that maximizes the ratio of the S_b to the S_w, we need to project these matrixes onto the $\mathbf{y}$ space and compute the $\mathbf{w}^*$ such that:

$$\frac{|w^{*T} S_b w^*|}{|w^{*T} S_w w^*|} = \max_{w} \frac{|w^T S_b w|}{|w^T S_w w|} \tag{10}$$

Once the maximization has been carried out, Fisher's analysis projects the $\mathbf{x}$ vectors onto the $\mathbf{y}$ space, which is a line. Then, each object $\mathbf{y}$ is assigned to the class i whose mean $\mathbf{m}_i$, projected onto the same line, is closest to $\mathbf{y}$.

Appendix C: Branch-and-bound technique

This technique (see e.g., Fukunaga 1990) allows us to select the subset of relevant features among those available. In fact, given n features for each object, apart from few statistical PR algorithms (e.g., BDT) that automatically provide the subset of features by which the classification is carried out (named optimal subset), most of the statistical PR algorithms just perform the pattern recognition and the classification of the objects, but do not explicitly provide the optimal subset. The basic concept in the selection of the optimal subset of features is to find, among all possible subset of the n features, the one leading to the lowest classification error and consisting of the smallest number of features. In such a situation, we are confident that we are considering all of the important variables (otherwise the classification error would not be the lowest) and we are excluding the irrelevant ones (otherwise the number of features in the optimal subset would not be the smallest).

A simple, but very time consuming way to find such an optimal subset consists of exploring the performance of the chosen statistical PR algorithm on all the possible subsets of the n features. This becomes prohibitive as n increases, since we have to explore $\sum_{k=1}^{n} \binom{n}{k}$ subsets. In order to avoid the application of the chosen statistical PR algorithm to all the possible subsets of features, the branch-and-bound technique was developed. This technique is applied iteratively n times; at each iteration k ($k=1,...n$), it allows for the identification of the suboptimal subset consisting of k features by applying the statistical PR algorithm only to the "most promising" subsets of k features. The suboptimal subset is then the one consisting of k features and leading to the lowest classification error.

The branch-and-bound method relies on a basic assumption; i.e., it assumes that the noise introduced by irrelevant features does not deteriorate the signal given by the relevant features. In a previous study we have tested the validity of this assumption for algorithms BDT and FIS. Based on this assumption, when a certain subset of k features does not produce a good discrimination rule, the branch-and-bound method assumes that any other subset of $k+l$ ($l=1,...n-k$) features containing those k features will not be the optimal one. In this way, a considerable portion of all the possible subsets is discarded a priori thus saving computation time and effort.

References

Aki K, Ferrazzini V (2000) Seismic monitoring and modeling of an active volcano for prediction. J Geophys Res 105(B7)16617–16640

Bak P, Tang C, Wiesenfeld K (1988) Self-organized criticality. Phys Rev Am 38:364–374

Benoit JR, McNutt SR (1996) Global volcanic earthquake swarm database 1979–1989. USGS Open-File Report 1996, No 69, US Department of the Interior, Washington, DC

Buell RE, Stoiber M (1976) Eruption of Volcan Fuego: October 14th, 1974. Bull Volcanol 38:861–870

Cammarata S (1997) Reti neuronali, 2nd edn. Etaslibri, Milano, pp 1–291

Decker R, Decker B (1981) The eruptions of Mount St Helens, vol 244, No 3. Scientific American, New York, pp 68–80

Duda RO, Hart PE (1973) Pattern classification and scene analysis. Wiley, New York, pp 1–482

Dziewonsky AM, Woodhouse JH (1983) An experiment in systematic study of global seismicity: centroid-moment tensor solutions for 201 moderate and large earthquakes of 1981. J Geophys Res 88:3247–3271

Dziewonsky AM, Chou TA, Woodhouse JH (1981) Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. J Geophys Res 86:2825–2852

Faberov AI (1983) Activity of the Plosky Tolbachik volcano June–July 1975. In: Fedotov SA, Markhinin YEK (eds) The great Tolbachik fissure eruption: geological and geophysical data 1975–1976. Cambridge University Press, Cambridge, pp 36–40

Fedotov SA, Gorel'chik VI, Stepanov VV, Garbuzova VT (1983) The development of the great Tolbachik fissure eruption in 1975 from seismological data. In: Fedotov SA, Markhinin YEK (eds) The great Tolbachik fissure eruption: geological and geophysical data1975–1976. Cambridge University Press, Cambridge, pp 189–203

Fukunaga K (1990) Statistical pattern recognition, 2nd edn. Academic Press, San Diego, pp 1–591

Gorshkov GS (1959) Gigantic eruption of the volcano Bezymianny. Bull Vulcanol 20:77–109

Gorshkov GS, Dubik YM (1970) Gigantic directed blast at Sheveluch volcano (Kamchatka). Bull Vulcanol 34:261–288

Gorel'chik VI, Stepanov VV, Khanzutin VP (1983) Volcanic tremor during the great Tolbachik fissure eruption of 1975. In: Fedotov SA, Markhinin YEK (eds) The great Tolbachik fissure eruption: geological and geophysical data 1975–1976. Cambridge University Press, Cambridge, pp 204–212

Gottsmann J, Rymer H (2002) Deflation during caldera unrest: constraints on subsurface processes and hazard prediction from gravity-height data. Bull Volcanol 64:338–348

Harlow DH, Power JA, Laguerta EP, Ambubuyog G, White RA, Hoblitt RP (1996) Precursory seismicity and forecasting of the June 15, 1991, eruption of Mount Pinatubo. In: Newhall CG, Punongbayan RS (eds) Fire and mud: eruptions and lahars of Mount Pinatubo, Philippines. Philippines Institute of Volcanology and Seismology, Quezon City, University of Washington Press, Seattle, pp 285–308

Hill DP, Dzurisin D, Ellsworth WL, Endo ET, Galloway DL, Gerlach TM, Johnston MJS, Langbein J, McGee KA, Miller CD, Oppenheimer D, Sorey ML (2002) Response plan for volcano hazards in the Long Valley Caldera and Mono Craters region, California. USGS Bull 2185:58

Hollander M, Wolfe DA (1973) Non-parametric statistical methods. Wiley, New York, 503 pp

Jensen JH, De La Cruz-Reina S, Singh SK, Medina-Martinez F, Gutierrez-Martinez C (1983) Actividad sismica relacionada con las erupciones del volcan Chichonal en marzo y abril de 1982, Chiapas. In: El volcan Chichonal, UNAM, Instituto de Geologia, pp 36–48

Keilis-Borok VI, Knopoff L, Rotwain IM, Allen CR (1988) Intermediate-term prediction of occurrence times of strong earthquakes. Nature 335:690–694

Keilis-Borok VI, Kossobokov V (1990) Premonitory activation of earthquake flow: algorithm M8. Phys Earth Planet Int 61:73–83

Londoño JM, Sudo Y (2002) A warning model based on temporal changes of coda Q for volcanic activity at Nevado Del Ruiz Volcano, Colombia. Bull Volcanol 64:303–315

Marzocchi W (2002) Remote seismic influence on large explosive eruptions. J Geophys Res 107:B1 DOI 10.1029/2001JB000307

Mc Nutt, SR (1996) Seismic monitoring and eruption forecasting of volcanoes: a review of the state-of-the-art and case histories. In: Scarpa R, Tilling R (eds) Monitoring and mitigation of volcano hazards, Springer, Berlin Heidelberg New York, 99–146

Miller TP, McGimsey RG (1998) Catalog of the historically active volcanoes of Alaska. Department of the Interior USGS Open-File Report 1998 No 582

Mulargia F, Gasperini P, Marzocchi W (1991) Pattern recognition applied to volcanic activity: identification of the precursory patterns to Etna recent flank eruptions and periods of rest. J Volcanol Geotherm Res 45:187–196

Mulargia F, Marzocchi W, Gasperini P (1992) Statistical identification of physical patterns which accompany eruptive activity on Mount Etna, Sicily. J Volcanol Geotherm Res 53:289–296

Newhall CG, Dzurisin D (1988) Historical unrest at large calderas of the world. USGS Bull 1855:1108

Newhall CG, Punongbayan RS (eds) (1996) Fire and mud: eruptions and lahars of Mount Pinatubo, Philippines. Philippines Institute of Volcanology and Seismology, Quezon City, University of Washington Press, Seattle, pp 1–1126

Newhall CG, Decker R (2002) Can the VEI of an eruption be forecast? In: Proc IAVCEI Int. Congr., Martinique, 12–16 May 2002

Newhall CG, Hoblitt RP (2002) Constructing event trees for volcanic crises. Bull Volcanol 64:3–20

Pacheco JF, Sykes LR (1992) Seismic moment catalog of large shallow earthquakes, 1900–1989. Bull Seismol Soc Am 82:1306–1349

Reeder JW, Lahr JC, Thomas J, Conens S, Blackford M (1977) Seismological aspects of the recent eruption of Augustine volcano. EOS Transact 58:12

Rounds EM (1980) A combined nonparametric approach to feature selection and binary decision tree design. Pattern Recogn 12:313–317

Sandri L, Marzocchi W (2003) Testing the performance of some nonparametric pattern recognition algorithms in realistic cases. Pattern Recogn (in press)

Shibata T, Akita F (2001) Precursory changes in well water level prior to the March, 2000 eruption of Usu volcano, Japan. Geophys Res Lett 28(9):1799–1802

Shimozuru D (1972) A seismological approach to the prediction of volcanic eruptions. In: The surveillance and prediction of volcanic activity, vol 8. UNESCO Earth Sci Monograph, Paris, pp 19–45

Simkin T, Howard KA (1970) Caldera collapse in the Galapagos Islands, 1968. Science 169(3944):428–437

Simkin T, Siebert L (1994) Volcanoes of the world, 2nd edn. Geoscience Press, Tucson, Arizona, pp 1–349

Smithsonian Institution's Global Volcanism Network (1990) Summary of recent volcanic activity. Bull Volcanol 52:407

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

Swanson SE, Kienle J (1988) The 1986 eruption of Mt. St. Augustine: field test of a hazard evaluation. J Geophys Res 93:4500–4520

Tarassenko L (1998) A Guide to neural computing applications. Wiley, New York, pp 1–139

Taylor GA (1957) The 1951 eruption of Lamington, Papua Commonwealth of Australia. Bureau of Mineral Resources. Geol Geophys Bull 38:117

Tokarev PI (1985) The prediction of large explosions of andesitic volcanoes. J Geodyn 3:219–244

Vinciguerra S, Latora V, Bicciato S, Kamimura RT (2001) Identifying and discriminating seismic patterns leading flank eruptions at Mt. Etna volcano during 1981–1996. J Volcanol Geotherm Res 106:211–228

Voight B, Sparks RSJ, Miller AD, Stewart RC, Hoblitt RP, Clarke A, Ewart J, Aspinall WP, Baptie B, Druitt TH, Herd RA, Jackson P, Lockhart AB, Loughlin SC, Lynch L, McMahon J, Norton GE, Robertson R, Watson IM, Young SR (1999) Magma flow instability and cyclic activity at Soufriere Hills Volcano, Montserrat, BWI. Science 283:1138–1142

Volcanological Society of Japan (1960–1993) Bull volc eruptions, vols 1–30. Published in Bull Volcanol since 1986

Zobin VM (1971) Mechanism of volcanic earthquake of the Sheveluch volcano, Kamchatka. Bull Vulcanol 35(1):225–229

Zobin VM (1983) The focal mechanism and dynamic parameters of volcanic earthquakes preceding the great Tolbachik fissure eruption of 1975. In: Fedotov SA, Markhinin YEK (eds) The great Tolbachik fissure eruption: geological and geophysical data 1975–1976. Cambridge University Press, Cambridge pp 243–256