Bollettino di Geofisica Teorica ed Applicata Vol. 54, n. 1, pp. 39-52; March 2013
DOI 10.4430/bgta0077

Detecting dynamical regimes by Self-Organizing Map (SOM) analysis: an example from the March 2006 phreatic eruption at Raoul Island, New Zealand Kermadec Arc

R. Carniel[1], L. Barbui[1] and A.D. Jolly[2]

[1] *DICA, Università di Udine, Italy*
[2] *GNS Science, Taupo, New Zealand*

(Received: May 17, 2012; accepted: July 5, 2012)

ABSTRACT We propose a technique to improve the analysis of volcanic seismic data and highlight possible dynamical or precursory regimes, by using an efficient class of artificial neural network, the Self-Organizing Maps (SOMs). SOMs allow an automatic pattern recognition, as independent as possible from any *a priori* knowledge. In the training phase, volcanic tremor spectra are randomly presented to the network in a competitive iterative process. Spectra are then projected, ordered by time, onto the map. Every spectrum will take up a node on the map and their time evolution on the map can highlight the existence of different regimes and the transitions between them. We show a practical application on data recorded at Raoul Island during the period around the March 2006 phreatic eruption which reveals both a diurnal anthropogenic signal and the post-eruption system excitation.

Key words: Volcanic eruptions precursors, dynamical analysis, Self-Organizing Maps, spectral analysis, 2006 Raoul eruption.

1. Introduction

Since volcanoes can devastate and modify the scenery of wide areas, an eruption is likely to have a significant impact on population and anthropic activities. Volcanoes can remain in a quiescent state for a long period and the volcanic unrest is often prolonged over a period of months to years making the identification of precursory patterns of the uttermost importance.

The core of the procedure is an efficient post-processing technique that uses a particular kind of artificial neural network called Self-Organizing Maps [SOMs: see Kohonen (1982)]. A SOM is composed of a single layer of neurons (arranged in the nodes of the grid). Using a competitive learning algorithm, the data are mapped onto the 2-dimensional grid, trying to preserve topological relations: patterns close in the *n*-dimensional input space should be mapped to nodes close on the map output space and the results can be represented as 2-dimensional feature maps. The SOM capability in data clustering allows the analysis of multidimensional, non-linear and highly noisy geophysical data (Klose, 2006). SOM has been employed to improve the H/V spectral ratio method [or HVSR technique or Nakamura's method (Nakamura, 1989, 2000)] to evaluate, in a cheap and relatively easy way, the fundamental frequency of a given site from its frequency spectra (Carniel *et al.*, 2009).

The basic idea is that if the volcanic system condition is related to some observed parametres

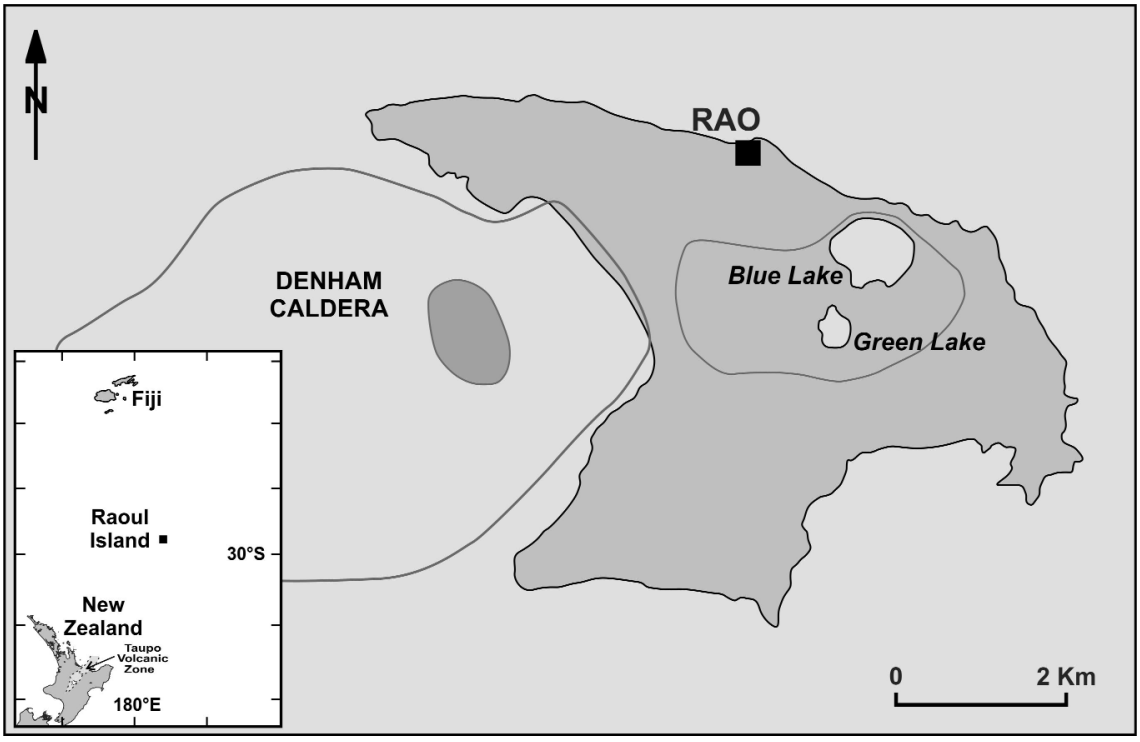


Fig. 1 - Location map of Raoul Island, showing location of seismic station RAO and major volcanic features.

(e.g., the frequency content of the acquired seismic signal) and if the SOM process has trained a well organized map (e.g., every possible cluster onto the map has at least one distinctive feature with respect to the other clusters and clusters with similar features are topologically close on the map), the projection of the data (e.g., frequency spectra) onto the map, ordered by time, could detect possible modifications of the volcanic system condition.

Raoul Island (Fig. 1) provides a useful case study of this approach because volcanic activity was monitored by a single seismic sensor and activity at the volcano includes a range of volcanic activity including large magmatic eruptions (Healy *et al.*, 1965; Lloyd and Nathan, 1981; Worthington *et al.*, 1999; Smith *et al.*, 2010), and small phreatic events (Christenson *et al.*, 2007). In addition, a small semi-permanent population of New Zealand Department of Conservation (DoC) staff is present. Hence, improvements in near-real time event discrimination may have positive benefits to the local population. On March 17, 2006 (08:21 NZST) [March 16, 2006 (20.21 UT)], a small phreatic eruption occurred. The event was preceded by a swarm of small earthquakes located 10-20 km from the seismic sensor (Christenson *et al.*, 2007) on March 12, decaying to background seismicity by March 16 NZST. The eruption seismic signal (Fig. 2) was composed of several pulses and had a duration of 8 minutes (Christenson *et al.*, 2007). The spectrogram of the day is shown in Fig. 3. On March 16 (20.21 UT), the phreatic eruption was recorded on the local seismograph. Unfortunately, the eruption took the life of DoC ranger Mark Kearney, who was in the volcanic crater at the time of the eruption. The goal of this paper is to examine retrospectively the activity of this event and determine if improvements in seismic data

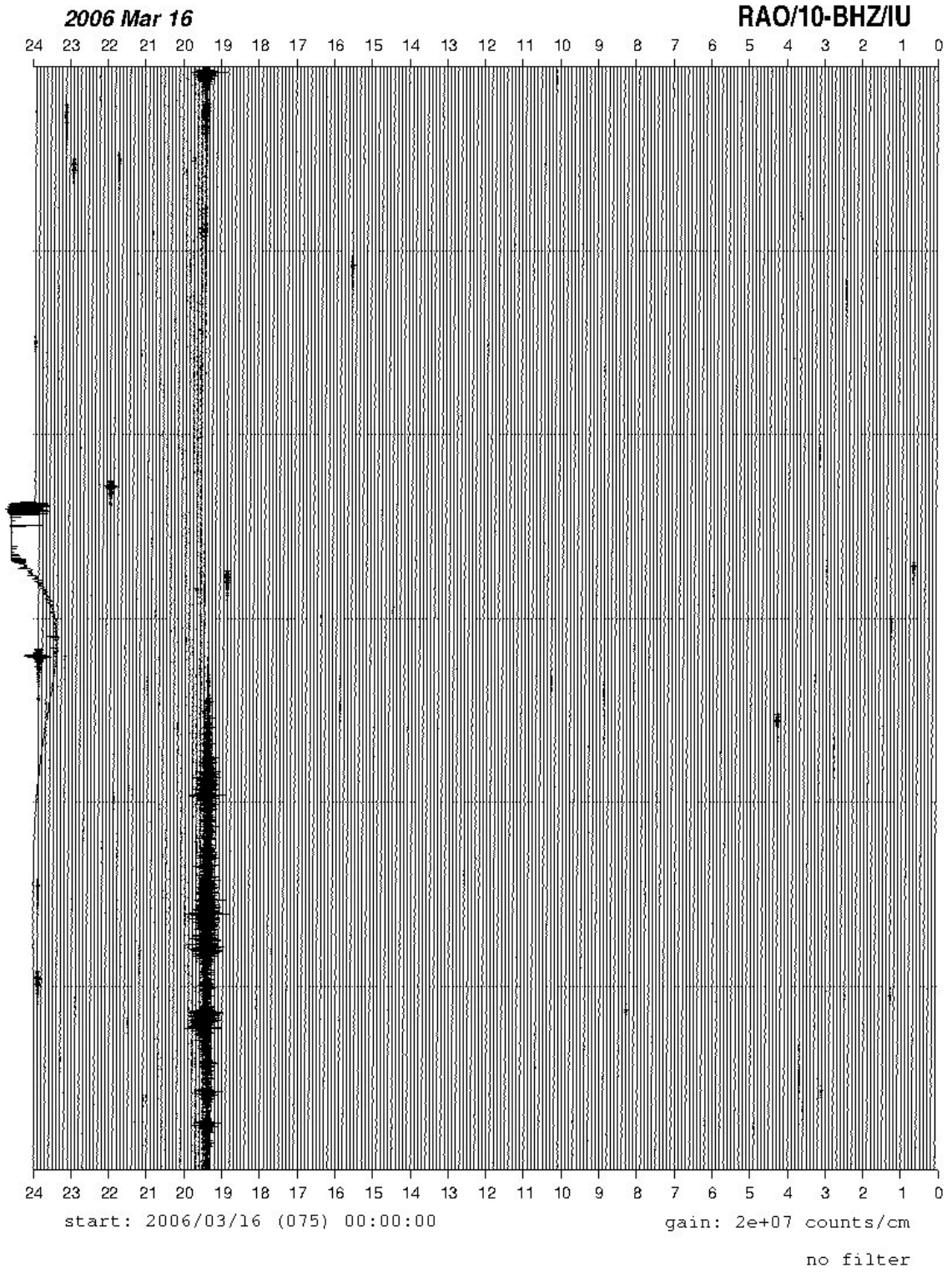


Fig. 2 - Original Helicorder recording acquired at seismic station RAO on Raoul Island, showing the pulsating eruption seismic signal on March 16, 2006 UT. Each line represents 6 minutes of data. The earthquake seen at the very end of the Helicorder recording is a tectonic event.

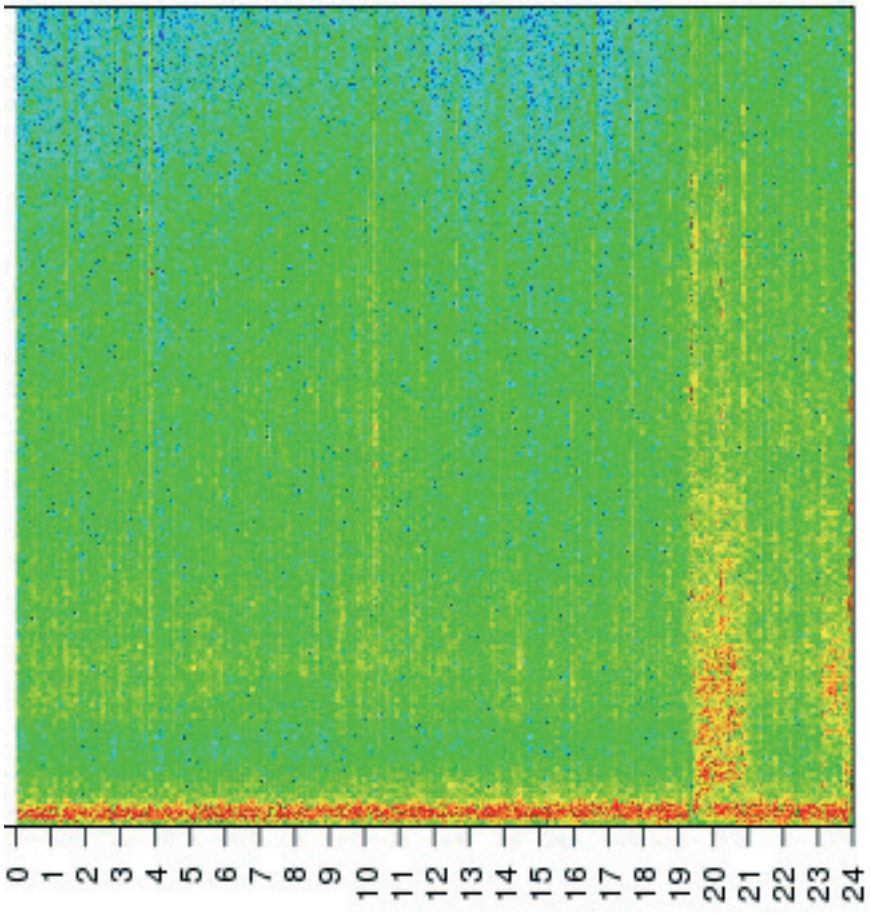


Fig. 3 - Spectrogram of the seismic data recorded at seismic station RAO on Raoul Island on March 16, 2006 UT.

processing and interpretation can be made. In this paper, we propose a technique that focuses on the volcanic tremor and background noise analysis without any *a priori* knowledge of the system. However, the methodology is easily extendable to other kinds of datasets.

2. Materials and methods

A SOM is an artificial neural network that carries out unsupervised competitive learning. Artificial neurons are arranged on a low-dimensional grid and each neuron is described by a n-dimensional vector w_i (called code vector), where n is the dimensionality of the input data space. Each neuron is connected to the neighbouring neurons, determining usually a rectangular or hexagonal organization of the map. Each neuron of the network is completely connected to all the input space data and the network represents a feed-forward structure with only one computational layer.

When an input vector (e.g., a frequency spectrum) is presented to the network, it causes a localized region or "bubble" of activity. Position and nature of this region usually changes with the input vector and during the learning process. All the neurons of the network should be

exposed to a sufficient number of input vectors to ensure that the self-organization process is suitable.

There are four essential processes involved in the formation of the SOM, listed below.

1. Initialization of code vectors. Since topology and dimensions of the map have been fixed, it is necessary to initialize the values of the neurons. In the application described in this paper a random initialization, choosing the code vectors among the available spectra, has been used.

2. Competition process. For each input vector, the neurons in the map compute their respective value of a discriminant function. The single neuron with the largest value of the discriminant function is declared the winner of the competition (also called best matching unit, BMU). Let n denote the dimension of the input space and the generic input vector be denoted by $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,n}]$. Let the code vector of neuron j be denoted by $\mathbf{w}_j = [w_{j,1}, w_{j,2}, \ldots, w_{j,n}]^T$ with j=1, 2, … l, where l is the total number of neurons in the map. To find the best match of the input vector x_i with the code vectors $\mathbf{w}$, a criterion is used based on maximizing the dot product $w_j \cdot x_i$, mathematically equivalent to minimizing the Euclidean distance between the vectors:

$$\left\| \mathbf{x}_i - \mathbf{w}_j \right\| = \sqrt{\sum_{m=1}^{n} \left[x_{i,m} - w_{j,m} \right]^2}. \tag{1}$$

If we call $\mathbf{w}_{c(x)}$ the BMU neuron for the input vector $\mathbf{x}_i$, the position in the map of the $\mathbf{w}_{c(x)}$ neuron determines the centre of the topological neighbourhood $h_{c(x),j}$. It will be verified:

$$\left\| \mathbf{x}_i - \mathbf{w}_{c(x)} \right\| = \min_j \left\{ \left\| \mathbf{x}_i - \mathbf{w}_j \right\| \right\}. \tag{2}$$

The effect of the competition process is that the continuous input space X of activation patterns is projected onto the discrete output space M. The Euclidean distance is not the only possible similarity measure, we then carry out some experiments with the correlation coefficients.

3. Cooperation process. The winning neuron $\mathbf{w}_{c(x)}$ determines the spatial location of a topological neighbourhood of excited neurons on the map. The neighbourhood function determines how strongly the neurons are connected to each other. It must be uni-modal with the lateral distance $d_{c(x),j}=\|r_c-r_j\|$ between the BMU neuron $\mathbf{w}_{c(x)}$ and the generic neuron $\mathbf{w}_j$ (r_c and r_j determine the position of the two neurons on the map). A typical choice of $h_{c(x),j}$ is the Gaussian function:

$$h_{c(x),j}(t) = \exp\left(-\frac{d^2_{c(x),j}}{2\sigma(t)^2} \right). \tag{3}$$

The neighbourhood function's value also depends on discrete time t that identifies the iteration number: at every step t the whole dataset will be processed by the network examining each input vector in random order. The parametre $\sigma(t)$ defines the effective width of the neighbourhood

function. During the learning process the neighbourhood radius $\sigma(t)$ must be reduced monotonically with the regression step t. A popular choice for the effective width of the neighbourhood function is the exponential decay:

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{T}\right) \tag{4}$$

where σ_0 is the value of σ at the beginning of the SOM algorithm and T is a time constant.

4. Adaptive process. It enables modification of the code vectors of excited neurons to increase the individual values of the discriminant function in relation to the input vector. The adjustment takes place in a way that reinforces the answer of the BMU neuron for similar input patterns, that allow the map training process and the partitioning of the map in clusters. By using discrete-time formalism, the adaptive process usually takes place according to the following model:

$$\mathbf{w}_j(t+1) = \mathbf{w}_j(t) + \alpha(t) \cdot h_{c(x),j}(t) \cdot \left[\mathbf{x}_i(t) - \mathbf{w}_j(t)\right] \tag{5}$$

where $0 \leq \alpha(t) \leq 1$ is the learning-rate factor, a monotonically function which decreases with the regression step t. Two suitable functions are indicated below:

$$\alpha(t) = \alpha_0 \cdot \exp\left(-\frac{t}{\tau}\right) \tag{6}$$

$$\alpha(t) = \alpha_0 \cdot \frac{\tau}{\tau + t} \tag{7}$$

where t is, as usual, the iteration number and τ is a new time constant.

There are two control mechanisms on the unsupervised self-organization algorithm: the adaptive learning-rate parameter $\alpha(t)$ and the neighbourhood function $h_{c(x),j}(t)$. Since $\alpha(t)$ is a slowly decreasing function, the updating of the code vectors decreases with the regression step t. The second mechanism shrinks the kernel neighbourhood of each BMU neuron gradually over time. A large neighbourhood will help to achieve a stable convergence of the map. Beginning with a large neighbourhood and then gradually reducing it to a very small neighbourhood, the SOM achieves both ordering and convergence properties. In this work the effective width of the neighbourhood $\sigma(t)$ starts with a value equal to the dimension of the greater map side. Calling I the total number of iterations, when $t=I/2$ the value of $\sigma(t)$ is null, so during the remaining $I/2$ iterations, only the code vector of the BMU will be modified. The number of iterations I is fixed *a priori*, at each iteration t the network processes the whole dataset examining each input vector in random order.

3. Results

In this work the SOM algorithm has been applied to the eruption tremor frequency spectra at Raoul Island (station RAO) located in the southern Pacific Kermadec Islands (Fig. 1). The seismic data is sampled at 100 Hz, and has been split in time windows of 1024 points (corresponding to a time window of 10.24 s). Time windows have a 50% of overlap. Fast Fourier Transform (FFT) has been applied to each time window to calculate the corresponding spectrum in a frequency range from 0.1 to 20 Hz.

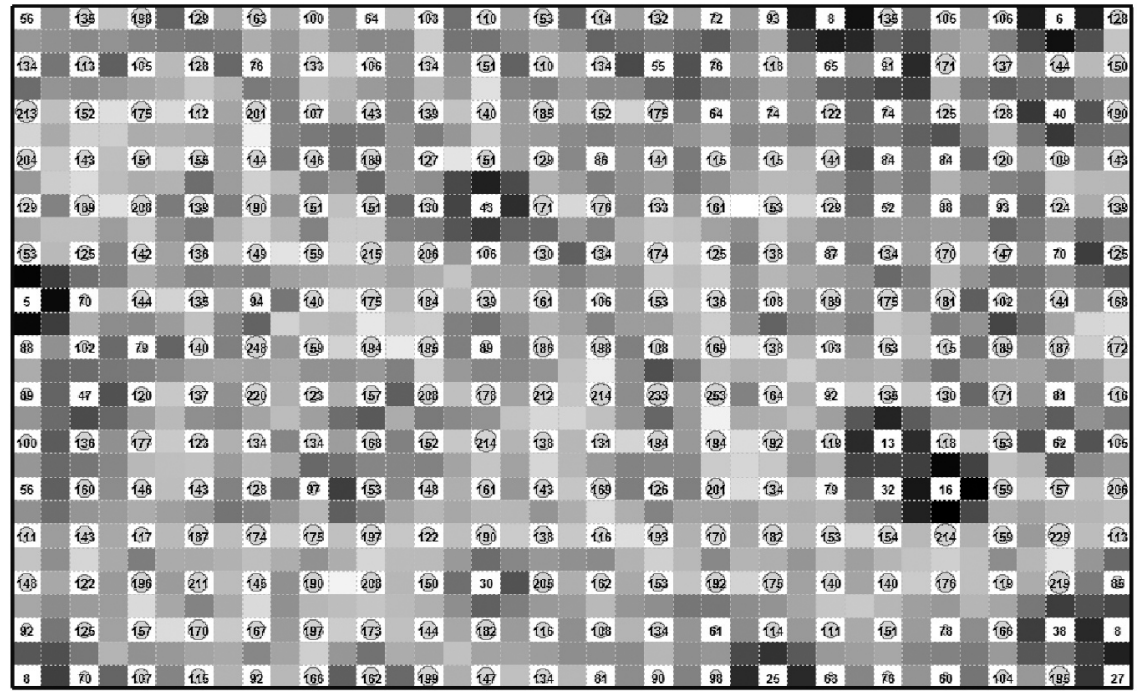
Each spectrum is an input vector for the SOM analysis, the process is split in two phases: the training phase and the data projection phase. During the training phase a large part of the whole data set is presented to the network a certain number of times. Excluded from the training data set are the spectra corresponding to the volcanic eruptive events, because the SOM trained map must be able to recognize both the hypothetical dynamical regimes as well as every singular event such as a volcanic explosion. During the training phase the whole training data set, composed of about ten thousand spectra, is presented to the network 100 times. At each iteration t, the network processes the training data set examining each input vector in random order. As the number of iterations t increases, the map topological order is improved. Obviously, this is a finite process and a large number of iterations causes an increase in the computation time, so, after many tests, a good compromise value for t has been found at 100 iterations. Two opposed needs compete to choose the map dimensions: a larger map increases the number of clusters that could arise; moreover a larger number of neurons generally increases the separation between these clusters and so the final resolution of the map. On the other hand, a too large map decreases the SOM capability to provide an overview of the data set structure.

At the end of the training phase the results of SOM analysis are shown using the U-matrix [(Ultsch, 2003), Fig. 4]: map nodes correspond to white cells, each node contains one neuron, the number and the circle dimension in the cell indicates how many input vectors have been mapped into the node. Cells which do not contain a number use a gray scale to show the similarity of the nearby neurons code vectors: light gray indicates a strong similarity and vice versa. In this way the light areas on the map indicate possible clusters consisting of more than one node.

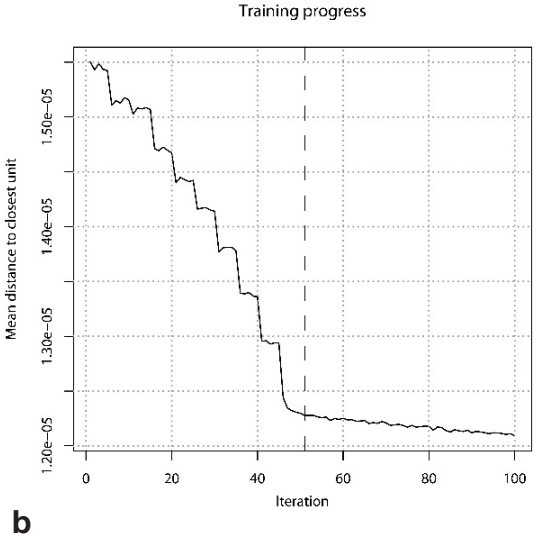
The data projection phase is conceived to analyse the temporal evolution of the volcanic system condition, monitoring the frequency content by the SOM pattern recognition capability. Each spectrum will take up a position on the map related with its frequency content and map topology. Since the code vectors are not modified, the map does not change its topology during the projection phase.

At each step, a group of spectra, equivalent to a fixed time length window of the acquired signal, gets projected onto the map. In order to keep the map unchanged during the projection phase, the neighbourhood function value $h_{c(x),j}(t)$ and/or the learning-rate factor $\alpha(t)$ have to be fixed at the null value. In this work the length of the time windows is fixed at about 10 minutes.

For each group of spectra projected onto the trained map, the Euclidean distance between each input vector and the correspondent BMU code vector has been measured. If this error is low for each input vector of the whole data set, then the map is well organized, and it is possible to explore the whole input space. Moreover, the input vectors with a larger error value could be singular data, potentially related to unusual conditions of the volcanic system, which cannot find a suitable strong similarity to a code vector on the map.



a



b

Fig. 4 - The SOM obtained at the end of a training phase using the Raoul seismic data (a). The graph (b) shows, for each iteration t of the SOM process, the mean error calculated as the mean Euclidean distance between input vector and the corresponding BMU code vector. The dashed line shows the iteration from which neighborhood radius σ becomes null.

One of the greatest issues of this in-time analysis is the large number of maps that have to be analysed and the consequent difficulty to have an overall view of the data set. This is achieved by summarising three key variables over time. These are, for each group of spectra projected: the position of the centre of the distribution, the position of the neuron with the largest number of input vectors and the moment of inertia of the distribution. The moment of inertia can be

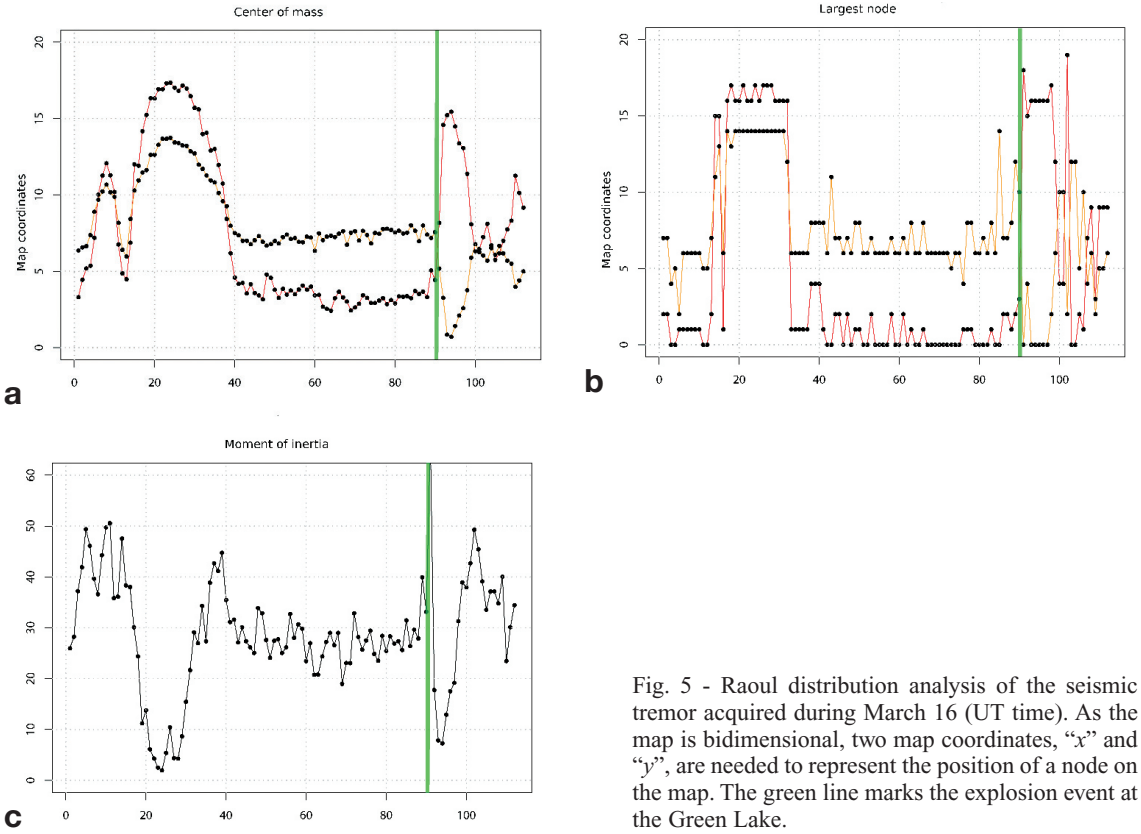


Fig. 5 - Raoul distribution analysis of the seismic tremor acquired during March 16 (UT time). As the map is bidimensional, two map coordinates, "*x*" and "*y*", are needed to represent the position of a node on the map. The green line marks the explosion event at the Green Lake.

calculated as:

$$M = \frac{1}{N}\sum_{i=1}^{N} N_{c_i} \cdot \sqrt{\left(x_{c_i} - x_{br}\right)^2 + \left(y_{c_i} - y_{br}\right)^2} \qquad (8)$$

where N is the total number of input vectors projected onto the map at the iteration t, N_{ci} is the number of input vectors projected onto the cell i of the map (with coordinates x_{ci} and y_{ci}) and the coordinates x_{br} and y_{br} define the position of the centre of the distribution. The moment of inertia assumes higher values when the vector projections fall farther from the centre of the distribution. On the contrary, a low value of the moment of inertia indicates a set of vectors very concentrated close to the centre of the distribution. This is physically analogous to the difference between an object whose mass is spread across a wide volume and one where the mass is mostly concentrated close to the baricentre.

In order to carry out the SOM training process, the seismic tremor acquired during the March 16, 17, 18 and 19 (UT time) by the seismometre positioned on Raoul Island has been split in windows of 1024 points, for each window the frequency spectra has been calculated and then fed to the SOM. The dimensions of the map have been fixed at 20´15 (equivalent to a map with 300

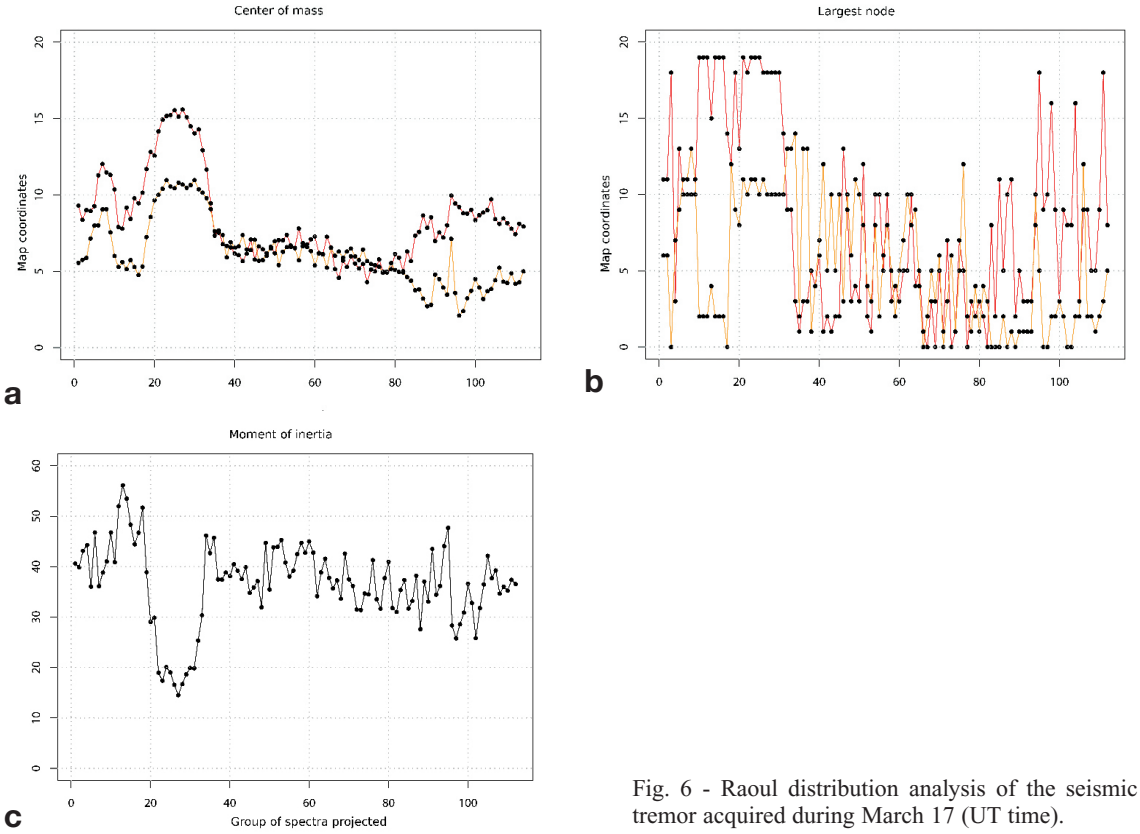


Fig. 6 - Raoul distribution analysis of the seismic tremor acquired during March 17 (UT time).

neurons).

At each step of the projection phase, a group of 120 spectra (corresponding to a window length of about 13 minutes) is projected onto the map. Each point (or pair of points) in Figs. 5, 6, 7, and 8 show the position of the centre of mass, the position of the neuron with the largest number of input vectors and the value of the moment of inertia for one group of spectra. It is important to underline that the seismic tremor data acquired during March 16 has not been used in order to train the map.

It has been verified that the error defined as the Euclidean distance between each input vector and its BMU onto the map always stays low, this assures that the SOM training produced a well organized map, able to explore the input space data used during the training as well the data that did not aid the map topology organization. Only after the explosion at the Green Lake the mean error shows a slight increase that lasts about one hour and a half (Fig. 5).

Considering the graphs of the day of the explosion and the three days after (Figs. 5, 6, 7, and 8), it can be noticed that the trend on the position of the centre of mass, the position of the greatest node and the value of the moment of inertia for the groups from approximately the 20th to the 40th is considerably different to the mean trend. This means that the spectra of those groups have been projected into a restricted (looking at the moment of inertia graph) area onto the map, implying a strong coherency of such spectra, which show almost the same shape within each of

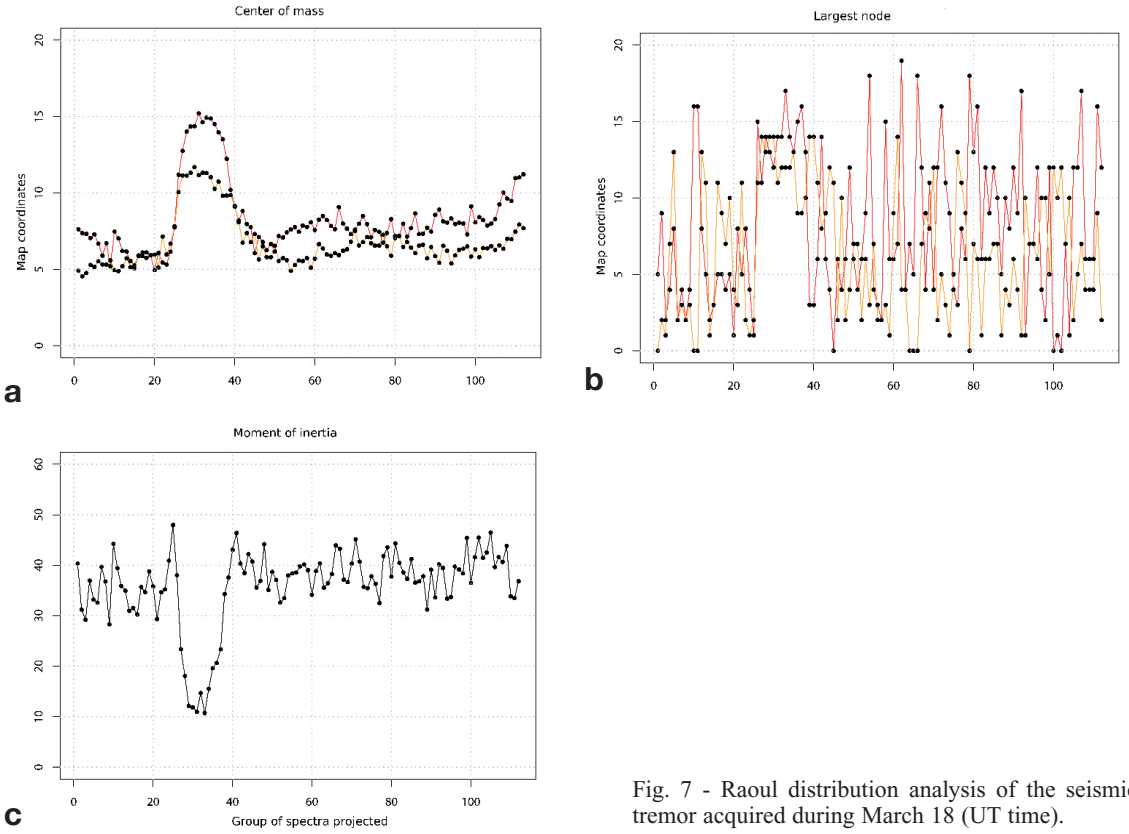


Fig. 7 - Raoul distribution analysis of the seismic tremor acquired during March 18 (UT time).

those groups. Moreover, the area of the map involved is different from the "usual" areas, due to a different mean frequency content. This excursion occurred each day analysed and at the same time of the day, hence we interpret the excursion as resulting from an anthropogenic source. We note however that the island was evacuated on the day following the eruption. We surmise that the diurnal cycle for groups 20-40 represents some relict cultural noise effect, such as an automated scheduled activity, that persisted even after the human population departed. A natural source for the diurnal cycle is considered unlikely. Moreover, during the one hour and half after the explosion the data has been arranged into an area of the map that was previously unfilled. We interpret this transition as a result of the explosion and post explosion system excitation.

4. Discussion

The time evolution of summary parametres of the SOM analysis, such as the centre of mass, the position of the neuron with the largest number of input vectors and the moment of inertia of the data distribution can provide information about the existence of relatively stable regimes and about the transitions between them by either slow or abrupt volcanic processes as expressed by transitions through the clusters on the map.

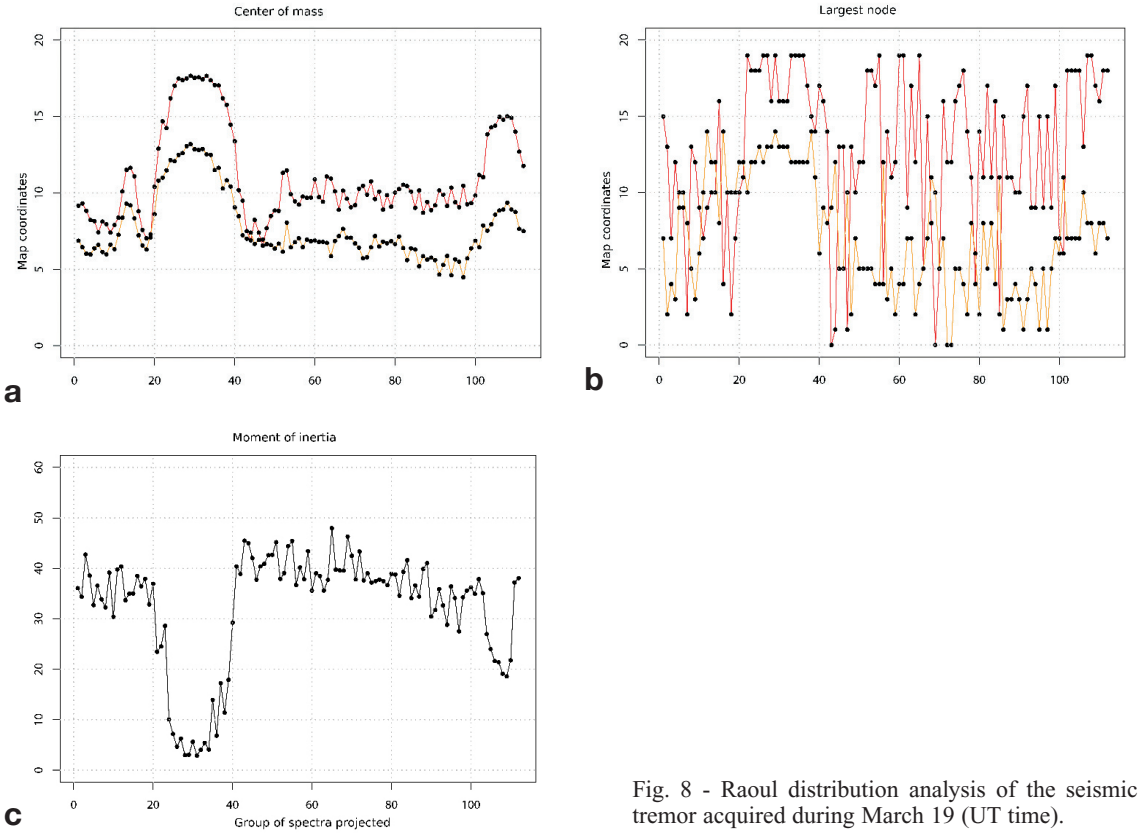


Fig. 8 - Raoul distribution analysis of the seismic tremor acquired during March 19 (UT time).

Similar volcanic regime changes have been noticed for Ambrym Volcano (Vanuatu Islands) (Carniel *et al.*, 2003). In the seismic tremor acquired there during a one month period, two regimes were observed, characterized by durations of a few days, that differ in terms of frequency content and more generally in terms of volcanic activity with different levels of tremor energy, degassing processes and small explosions at the lava lake (Carniel *et al.*, 2003). A similar alternation of two main regimes has been observed at the Stromboli Volcano (Sicily, Italy) with time scales going from minutes to weeks (Carniel and Iacop, 1996; Ripepe *et al.*, 2002). The Erta Ale lava lake (Ethiopia, Africa) also showed a similar alternation of volcanic activity regimes (Harris *et al.*, 2005; Jones *et al.*, 2006).

At Raoul Island, we observe a transition in spectral characteristics at the onset of the eruptive activity into a spectral regime that lasted for approximately 1.5 hours. Christenson *et al.* (2007) surmised that the proximal cause of the eruption was due to the failure of a shallow hydrothermal seal which became pressurised by gas released from a deeper magmatic carapace. The carapace itself is surmised to have failed due to a swarm of hybrid and volcano-tectonic earthquakes (Lahr *et al.*, 1994) occurring on March 12, which released gas from magma and caused pressurisation beneath the hydrothermal seal. If this model is correct, then the failure of the hydrothermal seal was instantaneous and included no precursors seen in either the observed spectra or the SOM analysis. However, the SOM highlights the post-failure system excitation which is possibly due

to the re-equilibration of the hydrothermal system.

5. Conclusions

The SOM method provides a new capability to explore the input data space applying an unsupervised pattern recognition algorithm, and allows us to recognize outliers as samples that have been projected onto unusual areas of the map. The analysis of the time evolution of a volcanic system through the SOM methodology offers then the possibility to highlight data patterns possibly related to a precursory activity of an upcoming volcanic crisis. At Raoul Island, the SOM method allows recognition of diurnal pattern in seismic data that may be anthropogenic in nature, and are not readily apparent in visual spectrogram analysis. The SOM did not reveal precursors to the eruption in Raoul Island seismicity, but it highlighted the post-failure system excitation.

Acknowledgements. This work was funded by project PRIN 2007PTRC4C_002 at the University of Udine (RC and LB) and New Zealand MSI funding (AJ). The EQC funded GeoNet Project provided data used in the project. Discussions with Steve Sherburn and the comments of two anonymous reviewers substantially improved the quality of the paper.

REFERENCES

Carniel R. and Iacop F.; 1996: *Spectral precursors of paroxysmal phases of Stromboli.* Ann. Geof., **39,** 327-345.

Carniel R., Barbui L. and Malisan P.; 2009: *Improvement of HVSR technique by self-organizing map (SOM) analysis.* Soil Dyn. Earthquake Eng., **29**, 1097-1101.

Carniel R., Di Cecca M. and Rouland D.; 2003: *Ambrym, Vanuatu (July - August 2000): spectral and dynamical transitions on the hours-to-days timescale.* J. Volcanol. Geotherm. Res., **128**, 1-13.

Christenson B., Werner C., Reyes A.G., Sherburn S., Scott B.J., Miller C., Rosenburg M.J., Hurst A.W. and Britten K.A.; 2007: *Hazards from hydrothermally sealed volcanic conduits.* Eos, Trans. Am. Geophys. Un., **88**, 53-55.

Harris A., Carniel R. and Jones J.; 2005: *Identification of variable convective regimes at Erta Ale lava lake.* J. Volcanol. Geotherm. Res., **142**, 207-223.

Healy J., Lloyd E.F., Banwell C.J. and Adams R.D.; 1965: *Volcanic eruption on Raoul Island, November 1964.* Nature, **205**, 743-745.

Jones J., Carniel R., Harris A.J.L. and Malone S.; 2006: *Seismic characteristics of variable convection at Erta Ale lava lake, Ethiopia.* J. Volcanol. Geotherm. Res., **153**, 64-79.

Klose C.D.; 2006: *Self-organizing maps for geoscientific data analysis: geological interpretation of multidimensional geophysical data.* Comput. Geosci., **10**, 265-277.

Kohonen T.; 1982: *Self-organised formation of topologically correct feature map.* Biol. Cybern., **43**, 56-69.

Lahr J.C., Chouet B.A., Stephens C.D., Power J.A. and Page R.A.; 1994: *Earthquake classification, location, and error analysis in a volcanic environment: implications for the magmatic system of the 1989-1990 eruptions at Redoubt Volcano, Alaska.* J. Volcanol. Geotherm. Res., **62**, 137-151.

Lloyd E.F. and Nathan S.; 1981: *Geology and tephrochronology of Raoul Island, Kermadec Group, New Zealand.* New

Zealand Geol. Surv., Bull. 95, 105 pp.

Nakamura Y.; 1989: *A method for dynamic characteristic estimation of subsurface using microtremor on the ground surface.* QR Railway Tech. Res. Inst., **30**, 25-33.

Nakamura Y.; 2000: *Clear identification of fundamental idea of Nakamura's technique and its application.* In: Proc. 12th World Conf. Earthquake Eng., Auckland, New Zealand, paper 2656.

Ripepe M., Harris A.J.L. and Carniel R.; 2002: *Thermal, seismic and infrasonic evidences of variable degassing rates at Stromboli Volcano.* J. Volcanol. Geotherm. Res., **118**, 285-297.

Smith I., Stewart R.B., Price R.C. and Worthington T.J.; 2010: *Are arc-type rocks the products of magma crystallisation? Observations from a simple oceanic arc volcano: Raoul Island, Kermadec Arc, SW Pacific.* J. Volcanol. Geotherm. Res., **190**, 219-234.

Ultsch A.; 2003: *U*-matrix: a tool to visualize clusters in high dimensional data.* DataBionics Res. Lab., Dept. Computer Sci., Univ. Marburg, Germany, Tech. Report 36, 12 pp.

Worthington T.J., Gregory M.R. and Bondarenko V.; 1999: *The Denham Caldera on Raoul Volcano: dacitic volcanism in the Tonga-Kermadec Arc.* J. Volcanol. Geotherm. Res., 90, 29-48.

Corresponding author: Roberto Carniel
Laboratorio di misure e trattamento dei segnali, DICA, Università di Udine
Via delle Scienze 206, 33100 Udine, Italy
Phone: +39 0432 558749; fax: +39 0432 558700; e-mail: roberto.carniel@uniud.it