



**HAL**  
open science

# Locating volcano-seismic signals in the presence of rough topography: wave simulations on Arenal volcano, Costa Rica

Jean-Philippe Métaxian, G. S. O'Brien, C.J. Bean, Bernard Valette, M. Mora

## ► To cite this version:

Jean-Philippe Métaxian, G. S. O'Brien, C.J. Bean, Bernard Valette, M. Mora. Locating volcano-seismic signals in the presence of rough topography: wave simulations on Arenal volcano, Costa Rica. *Geophysical Journal International*, 2009, 179 (3), pp.1547 - 1557. 10.1111/j.1365-246X.2009.04364.x . insu-00498848

**HAL Id: insu-00498848**

**<https://insu.hal.science/insu-00498848>**

Submitted on 10 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Locating volcano-seismic signals in the presence of rough topography: wave simulations on Arenal volcano, Costa Rica

J. P. Métaxian,<sup>1</sup> G. S. O'Brien,<sup>2</sup> C. J. Bean,<sup>2</sup> B. Valette<sup>1</sup> and M. Mora<sup>3</sup>

<sup>1</sup>Laboratoire de Géophysique Interne et Tectonophysique, IRD:R157, CNRS, Université de Savoie, Le Bourget du lac, France.

E-mail: metaxian@univ-savoie.fr

<sup>2</sup>Seismology and Computational Rock Physics Laboratory, School of Geological Sciences, University College Dublin, Belfield, Dublin, Ireland

<sup>3</sup>Universidad de Costa Rica, Escuela Centroamericana de Geología, Costa Rica

Accepted 2009 August 12. Received 2009 August 6; in original form 2008 March 12

## SUMMARY

Quantifying the scattering effects of pronounced volcano topography on the seismic wavefield is an important component in locating and interpreting volcano seismic sources. In this study, we perform seismic wave simulations to quantify the scattering generated by a 3-D digital elevation map and 1-D velocity model of Arenal volcano, Costa Rica. Full waveform synthetic seismograms were generated using a 3-D elastic lattice method including complex topography. Several different simulations were performed where the source location, source type and topographic models were varied. Synthetic seismograms were calculated for 35 seismic arrays each one comprising nine stations. At each array, the slowness vector of wave propagation is estimated from the time delays between the sensors obtained using the cross-spectral method. Results show that the backazimuth estimated for some arrays, in particular those close to the source, deviate from the true source position suggesting strong topographic effects in these regions. The maximum of the probability density function, obtained by crossing the backazimuths of the remaining arrays, coincides exactly with the true source location. We also compare our synthetic seismograms with array results from a physical field study. The true and calculated location misfit depends largely on the topography, but also on the number of antennas, the distance from the source and the spatial resolution of the antennas. The results show that this kind of study could be undertaken prior to the installation of seismic arrays to select the sites that minimize the topographic effects leading to improved source locations.

**Key words:** Volcano seismology; Wave propagation; Volcano monitoring.

## 1 INTRODUCTION

Volcanoes can exhibit several different classes of seismic signals. These signals, particularly long-period (LP) events and tremor, have been directly related to magma movement or magma ascent. Therefore, source location of these emergent events is of fundamental importance to better understand the volcanic plumbing system. These signals can be complex in character, which presents several difficult challenges in extracting information about the location, source type and volcanic structure. The difficulty arises in volcanic settings because the source mechanism is generally poorly understood and poorly constrained. Near-field effects in the seismic wavefield often cannot be ignored, the wavelengths can be large relative to the source–receiver distance and hence the *P*- and *S*-wave phases can be intertwined. Also, the signals have emergent wave trains and may persist for days in the case of tremor so separate wave phases can not be distinguished and the wavefield can be significantly distorted by the complex topography and volcanic stratigraphy.

In this study, we focus on the role of topography and shallow velocity structure in scattering the seismic wavefield and the con-

sequent effect on our ability to accurately locate the source. As discussed above, the lack of clear body wave phase arrivals makes it extremely difficult to locate sources with classical hypocentre determination methods. Other methods must be used, such as particle motion analysis where the ground deformation vector is reconstructed from the three-component seismogram, dense array methods where the slowness vector is estimated (Almendros *et al.* 2001a,b, 2002; Métaxian *et al.* 2002; Saccorotti *et al.* 2003; La Rocca *et al.* 2004; Di Lieto *et al.* 2007a,b) or waveform inversion to quantify the source mechanism (Ohminato *et al.* 1998; Chouet 2003; Chouet *et al.* 2005; Nakano & Kumagai 2005; Lokmer *et al.* 2007; Bean *et al.* 2008). These methods use the whole or partial seismic waveform, not only arrival times. However, as briefly discussed, the wavefield is affected by the internal structure of the volcano, the complex free surface, characterized by a pronounced topography and the shallow location of seismic source generating surface waves. Neuberg & Pointer (2000) show that for broad-band waveforms, the angle of incidence as well as the backazimuth is affected by an inclined free surface. Numerical simulations of seismic wave propagation on Merapi (Indonesia), show that a complex, heterogeneous wavefield can

emerge due to the presence of steep topography and that strong surface waves are generated and dominate the wavefield at later times (Ripperger *et al.* 2003). Along with complex topography subsurface volcanic structures can clearly influence in the wavefield. Wegler & Luhr (2001) and Wegler (2003) pointed out the importance of scattering within the volcanic structure rather than scattering at the free surface. Parsiegla & Wegler (2008) modelled the effects of topography and volcanic stratigraphy reiterating the importance of scattering by internal structures. Almendros *et al.* (2001a,b) noted significant differences between the slowness vectors derived for a homogeneous velocity model including topography and a medium including both structure and topography of Kilauea. They used large antennas composed of 22–41 receivers and used the MUSIC algorithm (Schmidt 1986; Goldstein & Archuleta 1987) to estimate the slowness vectors. Along with complex topography subsurface volcanic structures can clearly influence in the wavefield.

In this paper, we address the problem of locating volcano-seismic signals by using seismic array methods in the presence of rough topography. Array methods designed for source location by estimating the slowness vector usually use large time windows ( $\sim 1$  s) compared to the whole duration of the signal. Thus, the processed part of the signal corresponding to the beginning may include different wave phases, *P*, *S* and surface waves. Waves can be intertwined depending on the source–receiver distance. The general idea of this paper is to estimate how the composition of the first part of the seismogram is affected by the topography and consequently how it can influence the source location accuracy. We have performed several numerical simulations of seismic wave propagation using a 3-D digital elevation map and 1-D velocity model of Arenal volcano (derived from Mora 2003), Costa Rica, where the source location, source type and topographic models were varied. Synthetic seismograms were calculated at 35 seismic arrays each one composed of nine stations. A polarization analysis was undertaken using the synthetics. We used array methods to estimate the slowness vector at several sites around the volcano including locations corresponding to a real field experiment. Finally, we compared our synthetic results with the results of the field experiment.

## 2 ARENAL VOLCANO

Arenal volcano is a small basaltic andesite stratovolcano 1100 m high, 1660 m above sea level located in northwest Costa Rica. It began to form 7000 BP by the superposition of lava fields, pyroclastic flows and epiclastic deposits. Rocks from the volcano basement are older than 20 000 yr and include tephra, tuffs, lahars and lava flows, all of them mildly to deeply weathered (Soto 1998; Alvarado *et al.* 2003). Arenal is located at the intersection of several active Holocene faults striking predominantly N–S, NW–SE and NNW–ESE. One of the main systems is the Danta fault, which is an active N–S normal fault located underneath the volcano (Alvarado 1989; Leandro & Alvarado 1999; Soto *et al.* 1999; Alvarado *et al.* 2003). Arenals edifice consists of a lava armour composed of at least five lava fields. At the lower parts of the flanks the geological structure mainly consists of a heterogeneous unconsolidated tephra (pyroclastic and epiclastic deposits) apron, (Borgia *et al.* 1988; Alvarado *et al.* 2003; Mora *et al.* 2006). According to Leandro & Alvarado (1999) this apron has *P*-wave velocities less than  $1.95 \text{ km s}^{-1}$  and is thicker on the western flank (25–140 m) than the eastern flank (20–50 m).

The present cycle of activity began with a lateral explosive eruption on the 1968 July 29 that started with a closed-conduit pelean

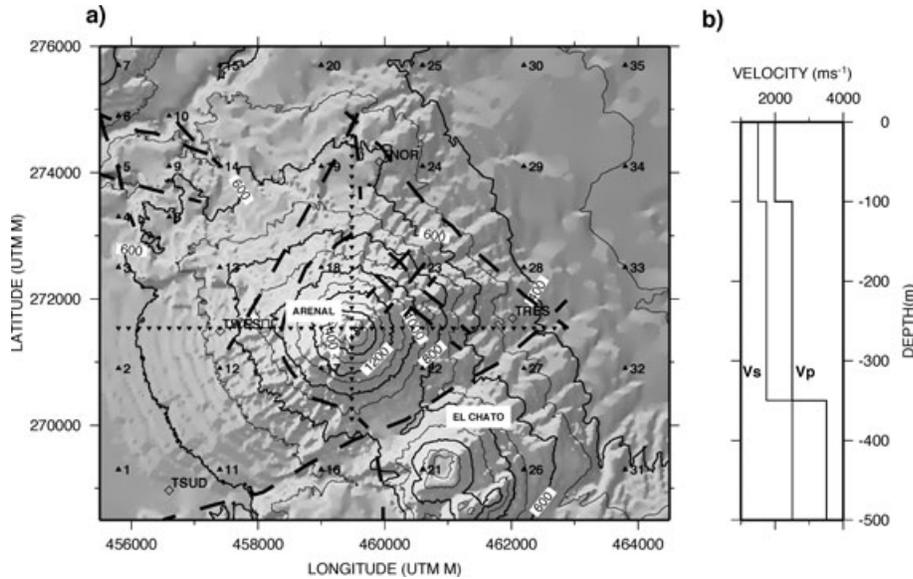
eruption and formed three new craters on the western flank. Those new craters together with the previous crater suggest an east–west fracture system. Arenals current activity is characterized by Strombolian eruptions, lava avalanches, and sporadic pyroclastic flows. The seismic activity is characterized by explosions accompanied by an acoustic phase (Alvarado & Barquero 1987; Garcès *et al.* 1998), LP events and various types of tremor (Lesage *et al.* 2006). A characteristic of Arenals activity is the similarity between the signal and the frequency content of the explosions and LP events. The energy is concentrated in several peaks between 1 and 3 Hz (Lesage *et al.* 2006). Several seismic studies have been conducted examining site effects, location of emergent events using seismic antennas techniques and tremor source analysis (Mora *et al.* 2001, 2006; Métaxian *et al.* 2002). The results of these studies indicate that possible topographic effects in the southwest flank may affect the source location of LP events and explosions (Fig. 9, Métaxian *et al.* 2002).

## 3 WAVEFIELD MODELLING

Full wavefield synthetic seismograms were generated using a 3-D elastic lattice method for the solution of the elastic wave equation including structural heterogeneity and complex topography. The 3-D elastic lattice method for the simulation of seismic waves consists of particles arranged on a cubic lattice, which interact through a central force term and a bond-bending force. Particle disturbances are followed through space by numerically solving their equations of motion. It can be shown mathematically that this approach is equivalent to the macroscopic continuum elastic wave equation in 2-D and 3-D (Monette *et al.* 1994; O'Brien & Bean 2004). Topography is included by removing particles above the free surface, (O'Brien & Bean 2004). The method is computational equivalent to a fourth order finite-difference method.

The numerical model comprises a volume of  $10.24 \text{ km} \times 8 \text{ km} \times 4.8 \text{ km}$  with a spatial resolution of 16 m. The digital elevation map is shown in Fig. 1 along with the locations of the synthetic arrays. Each of the 35 arrays is comprised of nine receivers arranged in a square with a distance between each receiver of 48 m. The array density is higher at the northwest of the volcano so as to study the effect of the valley crossing this area. We also included an east–west and a north–south profile across the crater along with four additional arrays corresponding to the field experiment sites of Métaxian *et al.* (2002). The layered velocity model used in this study (Fig. 1b) is derived from structural studies performed with two semi-circular arrays which were installed on the western and eastern flanks of the volcano (Mora 2003). Several days of noise recordings were analysed using the SPAC method (Métaxian *et al.* 1997). Results indicate higher velocities on the eastern flank, which corresponds to older and more compacted deposits. We adopted an average velocity model for the entire structure. The 1-D velocity model follows the topographic slope giving a constant depth for the two low velocity layers for the entire model. A Ricker wavelet source with a central frequency of 2 Hz was used in all simulations as the frequency content coincides with the main spectral content of the physically recorded events.

In total 12 simulations were performed which can be classified into two groups, the topographic simulations denoted by S1–S8 and the filtered topography models T1–T4. The parameters explored in the simulations are listed in Table 1. In the first group of simulations (S1–S8), the first three were performed with the layered velocity model shown in Fig. 1(b), using an explosive source located directly



**Figure 1.** (a) Map of Arenal volcano showing the locations of the 35 nine-receiver seismic antennas (triangles), the profiles (inverse triangles) and the triangular array (open diamonds) corresponding to the position of field experiment antennas. El Chato is an old inactive volcano located southeast of Arenal. (b) Stratified 1-D velocity model where zero is the free surface.

**Table 1.** Parameter space explored by the 12 different simulations.

Name	Type of the source	Altitude (m)	Topography	Radial ( $\mu\text{m}^{-1}$ )	Transverse ( $\mu\text{m}^{-1}$ )	Vertical ( $\mu\text{m}^{-1}$ )	$\sigma$ ( $^{\circ}$ )
S1	Explosion	1188	Real	45	18	38	3.7
S2	Explosion	868	Real	19	5	13	4.3
S3	Explosion	68	Real	12	3	9	4.6
S4	Explosion	1188	Real	21	4	16	5.4
S5	Vertical Pipe [2 2 1]	1188	Real	38	18	36	3.7
S6	Horizontal Crack [3 1 1]	1188	Real	51	36	48	5.3
S7	Horizontal Crack [1 3 1]	1188	Real	45	39	46	4.9
S8	Vertical Crack [1 1 3]	1188	Real	59	22	56	4.2
T1	Explosion	868	Smoothed	20	7	13	4.5
T2	Explosion	868	Smoothed	13	2	9	3.1
T3	Explosion	868	Smoothed	10	0.7	9	2.1
T4	Explosion	868	Flat	13	0.7	7	0.4

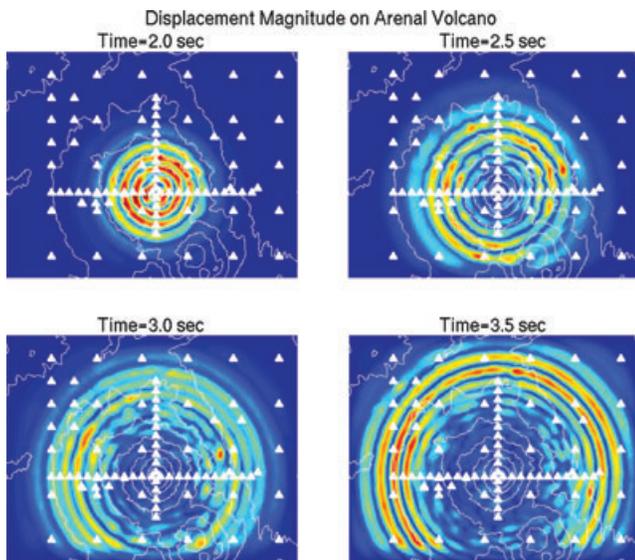
*Notes:* Maximum amplitude averaged over the entire synthetic signals for each simulation for each component is listed in the right-hand columns. The last column indicates the standard deviation  $\sigma$  of the difference between the backazimuth and the source azimuth for the 35 arrays.

beneath the summit crater at an altitude of 1188, 868 and 68 m for S1, S2 and S3, respectively. Simulation S4 was run in a homogeneous model ( $v_p = 3500 \text{ m s}^{-1}$  and  $v_s = 2020 \text{ m s}^{-1}$ ) with an explosive source located at an altitude of 1188 m. Simulations S5–S8 were performed with the 1-D velocity model and a source depth of 1188 m but with different source mechanisms. Four different sources were considered, a vertical pipe and two vertical cracks orientated in the north and east directions and a horizontal crack. From Aki & Richards (2002) the moment corresponds to [ $m_{xx} = 2 m_{yy} = 2 m_{zz} = 1$ ] with  $m_{xy} = m_{xz} = m_{yz} = 0$  for a vertical pipe and using the same notation [1 3 1], [3 1 1] and [1 1 3] with  $m_{xy} = m_{xz} = m_{yz} = 0$  for a vertical crack orientated in the north and east and a horizontal crack. These sources were considered as they represent the most common sources retrieved from moment tensor inversion of LP and very LP events recorded on active volcanoes. They also allow us to explore the effect of different radiation patterns on the source location. In the T1–T4 simulations the digital elevation map was filtered to smooth the topography from the original rough free surface to a smoother surface. The maximum amplitude of the topography is

1472 m and was decreased to 1152, 656, 240 and 0 m for models T1–T4, respectively. The source type and location along with the velocity model were fixed. This was done to quantify the effect of the scattering created by the topography.

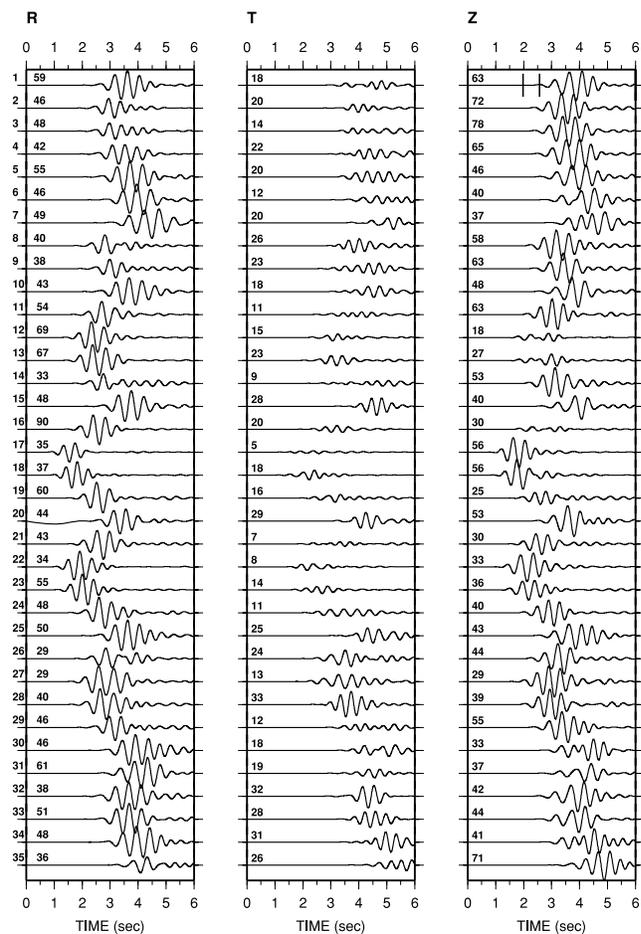
#### 4 SIMULATION RESULTS

The vertical component of the seismic wavefield moving across the surface of the volcano is displayed in Fig. 2 for four different times for simulation S1. As the source is explosive, the wavefield moves across the surface in a clear circular pattern. After some time the scattering effect of the layered velocity model and volcano topography perturb the circular wavefield. This is seen in the distortion of the wave front after 2.5 s in the southeast quadrant and also in the northwest quadrant at 4 s. In the final two temporal snapshots the complexity of the wavefield is clearly seen behind the main wave front. This wavefield consists of reflected and refracted  $P$  waves,  $S$  waves generated by  $P$  to  $S$  conversions, trapped waves inside the



**Figure 2.** Each subplot shows the normalized displacement magnitude on the surface of Arenal for different time periods for simulation S1. The topography of Arenal is contoured and the location of the seismic antennas are also shown. As the source is explosive, the circular spreading of the wavefield is clearly visible. As the wave front propagates across the surface the effect of the topography perturbs the wavefield and the wavefield becomes irregular.

layered structure and surface waves. The seismograms at the central receiver of each array are displayed in Fig. 3. The radial and transverse components were calculated by rotating the north and east component in the direction of the source. This is possible as we know the exact source location. The traces are normalized to the highest amplitude of the three components. The signals are complex and not easily interpretable; the initial *P* wave (explosive source and ballistic wave) is followed by two or more distinct phases. The general features show a dominant *P* wave onset for the closest sites on the vertical and radial components and several phases on the three components for more distant sites. There is clearly less energy on the transverse component as the explosive source generates no *S* waves. The *S* waves are created by *P* to *S* conversions. For distant stations after the initial *P* wave, different phases clearly dominate the signal and are probably surface waves. In Table 1, the average amplitude for each component calculated over the entire synthetic signals for each simulation is shown. As expected the values decrease when the source is deeper, S1–S3. For an explosive source with a stratified or a homogeneous velocity (S1–S4) the energy is always higher on the radial component than the vertical component. The energy on the transverse component is two to four times lower than on the radial component. For a perfectly horizontal free surface with horizontal layers and using an explosive source, the transverse energy should be zero. Therefore, the transverse component represents the scattering produced by the stratified velocity model and topography. For the non-isotropic sources (S5–S8) a similar pattern is observed except the relative ratios of the components differ. This is easily explained in terms of the generation of *S* waves in the source and the different seismic radiation patterns. For an explosive source with a stratified medium and a progressive smoothed topography (T1–T4), the energy on the transverse component decreases strongly and tends zero for T3 and T4. This confirms that the energy on this component is generated by scattering from the topography. Clearly, the seismic signals from all the simulations are complex in nature and difficult to interpret even though we have access to the

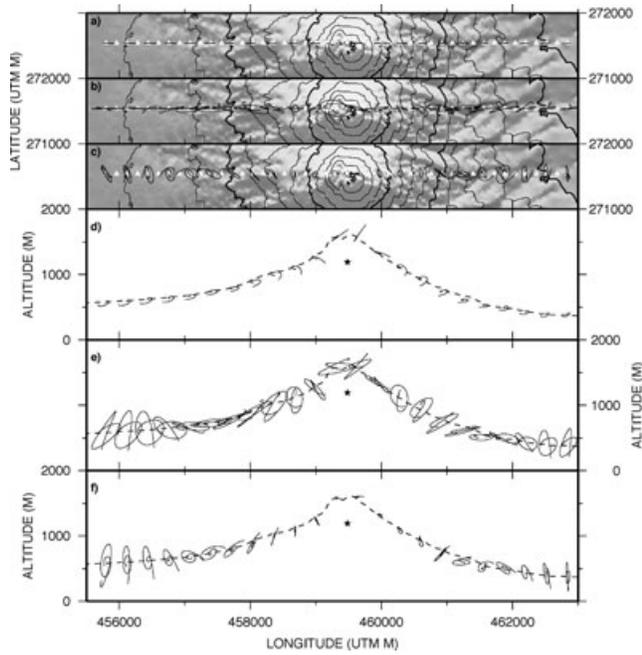


**Figure 3.** The radial, transverse and vertical synthetic seismograms calculated at the central receiver of every array (labelled on the far left-hand side) are displayed. The source is explosive (simulation S1). The location of each array is shown in Fig. 1. Traces are normalized to the maximum of all three components. The maximum amplitude of each component is shown on the left-hand side of the trace.

full 3-D wavefield. To help quantify the effect of the topography we perform standard polarization wavefield analysis on the synthetic data along with locating the synthetic events using a slowness vector method.

## 5 POLARIZATION ANALYSIS

The particle motion was plotted for three different time windows (defined as a function of the hypocentral distance) to explore the beginning, central and last part of the signal from simulation S1. Projections of the motion are displayed for the horizontal components and in the vertical-radial plane for the west–east profile (Fig. 4). The first time window includes the first onset for all the simulations. The particle motion is dominated by *P*-wave arrivals aligned with the source direction on the horizontal plane. The alignment is less clear in the vertical plane particularly for the stations nearest to the source where the *P* and *S* waves are intertwined. In the second time window a transverse and vertical motion appears for several sites probably related to the arrival of *S* waves. Elliptical motion in the vertical plane probably represents the arrival of surface waves. At intermediate distances, between 457 000 and 458 000 (UTM m) and 461 000 and 462 000 (UTM m) there is less

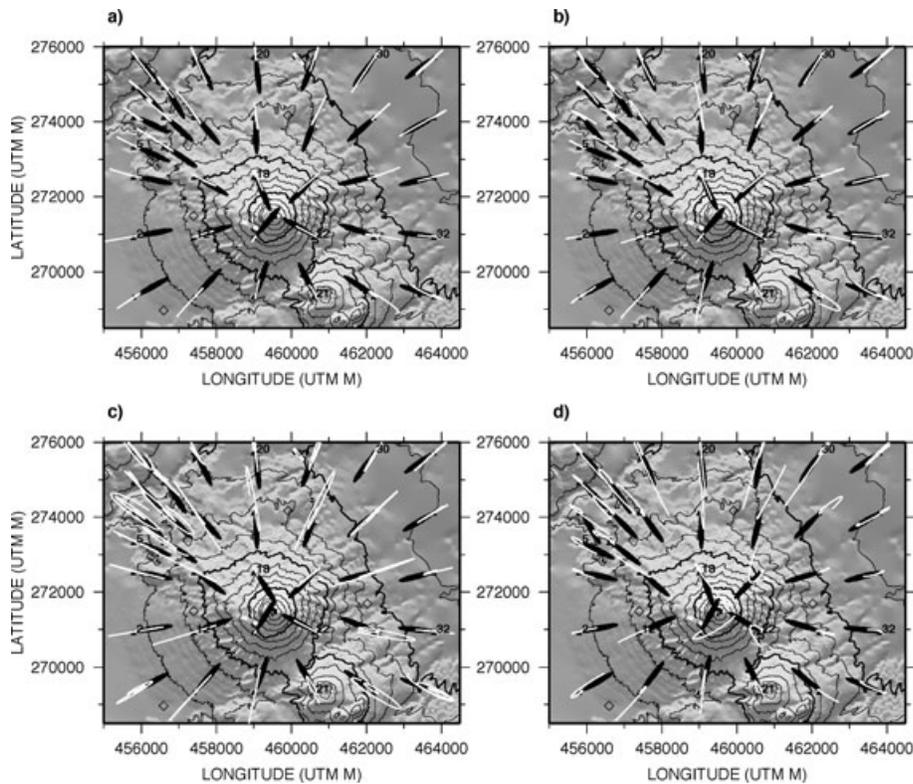


**Figure 4.** Projections of the particle motions in the horizontal plane and the vertical-radial plane for three different time-windows. (a) and (d) [0 1.15 s], (b) and (e) [1.15 1.95 s], (c) and (f) [1.95 2.75 s]. Triangles indicate the positions of the stations; the open star indicates the source position. The amplitudes are given in arbitrary units and are represented with the same scale.

energy in the vertical plane and more in the transverse direction. This could be interpreted as SH or Love waves. Finally, the third time-window shows that for the distant stations a clear elliptical motion is seen and can be interpreted as Rayleigh waves. As with the east–west profile, the north–south profile has a clear linear polarization of the *P*-wave arrivals at all the receivers. The particle motions for the later part of the signal are also less easily interpretable and show more complex patterns. The particle motion was analysed for the other simulations. No clear differences can be observed between S1, S2 and S3, save for changes in the vertical motion orientation that is related to the source depth. For simulations S5–S8, the first time window shows *P*-wave arrivals perfectly aligned with the source direction on the horizontal plane. It is less clear and interpretable in the vertical plane for the closest and intermediate receivers for the first and second time windows. For more distant receivers, we observe clear elliptical motion as for the explosive sources. Particle motion obtained at the central station of each array for the first part of the signal is shown in Fig. 5 for simulations S1, S3, S4 and S7. As with the profiles, a linear polarization is clearly observable for the first time window and is aligned in most cases towards approximately the source location.

### 6 SLOWNESS VECTOR ANALYSIS

To evaluate the topographic effects on the wave propagation, we aim to compare an estimated value of the slowness vector with the source direction at each array. We applied the slowness vector analysis method described in Métaxian *et al.* (2002). We assume that the wave field is composed of plane waves, is non dispersive, only



**Figure 5.** Horizontal particle motion of each array for the first part of the signal is shown as white lines for simulations S1, S3, S4 and S7 (plots a, b, c and d, respectively). A linear polarization is observable and is aligned in most cases to approximately the source location. The PDF of the backazimuth represented as rose diagrams with 1 increments for each of the 35 arrays for the same simulations is also plotted. Some weak differences of a few degrees are observed between various antennas for these four simulations.

one wave propagates across the array at the same time or, at least, one is dominant and the medium is laterally homogeneous underneath the array. These assumptions may not hold for our numerical simulations but are usually applied to real data. We will show that these assumptions provide a reliable estimate of the source location. The first step consists of measuring time delays between the receivers by using the cross-spectral method (Poupinet *et al.* 1985; Got *et al.* 1990, 1994). The number of delays is  $N(N - 1)/2$ , where  $N$  is the number of receivers. The uncertainties in the time lags are estimated and can be less than half the sampling spacing. These measurements are carried out on sliding time windows that move along the entire signal. Time-series, either of slowness vectors or of backazimuths, apparent velocities and incident angle, are obtained as a second step by solving the relation

$$\tau_{i,j} = \mathbf{s} \bullet \mathbf{r}_{i,j}, \quad (1)$$

where the dot denotes the scalar product,  $\tau_{i,j}$  is the time delay between receivers  $i$  and  $j$ ,  $\mathbf{s} = -(s \sin \theta, s \cos \theta, 1/s \tan \phi)$  is the slowness vector,  $\theta$  is the backazimuth, measured clockwise from the north,  $s$  the apparent slowness ( $s = 1/v_{\text{app}}$ , where  $v_{\text{app}}$  is the apparent velocity),  $\phi$  the incident angle and  $\mathbf{r}_{i,j}$  is the relative position vector (Rost & Thomas 2002). If the  $\mathbf{r}_{i,j}$  vectors are coplanar only the components of eq. (1), defined in the receivers plane, can be obtained. The orthogonal components are completely unknown. Therefore, in this case only the backazimuth and the apparent slowness can be calculated. The expression of the slowness vector can then be simplified and expressed with two parameters as  $\mathbf{s} = -(s \sin \theta, s \cos \theta)$ . If the receivers are not in the same plane, (which is the normal situation on a volcano) the incident angle and the local slowness under the array can be obtained. The components of the slowness vector are estimated by linear inversion of the  $N(N - 1)/2$  equations. The problem is overdetermined for  $N \geq 3$  in the case of two parameters (two independent measures of the time delay) and overdetermined for  $N \geq 4$  in the case of three parameters (at least three independent measures of the time delay). We used eq. (43) of Tarantola & Valette (1982) to estimate the parameters and eq. (44) for the calculation of the *a posteriori* covariance. This method estimates the slowness vector only for the dominant wavefield crossing the array. Therefore, the results are usually only interpretable for the beginning of the signal, dominated by  $P$  and  $S$  waves. For synthetic data, the known velocity model and access to the full wavefield solution allows us to interpret the results for the latter part of the seismogram. However, only values calculated from the first part of the signal, including the  $P$  wave were considered as a reliable indication of the backazimuth, the apparent slowness and the incident angle. To validate the method of using the slowness vector estimation to quantify topographic effects, we first processed data from simulation T4, which has a perfectly horizontal free surface. The difference between the backazimuth calculated from the slowness vector method and the true backazimuth is negligible. The slowness vector was estimated for the nine receivers in each of our synthetic arrays (36 delays). To determine whether the incident angle can be estimated, we defined an index of coplanarity  $\zeta$  ranging between 0 and 1 by

$$\begin{aligned} \zeta^3 = & \frac{\|(\mathbf{v}_3 \bullet \mathbf{v}_1) \mathbf{v}_2 - (\mathbf{v}_3 \bullet \mathbf{v}_2) \mathbf{v}_1\|^2}{\|v_3\|^2(\|v_1\|^2\|v_2\|^2 - (\mathbf{v}_1 \bullet \mathbf{v}_2)^2)} \\ & \times \frac{\|(\mathbf{v}_2 \bullet \mathbf{v}_1) \mathbf{v}_3 - (\mathbf{v}_2 \bullet \mathbf{v}_3) \mathbf{v}_1\|^2}{\|v_2\|^2(\|v_1\|^2\|v_3\|^2 - (\mathbf{v}_1 \bullet \mathbf{v}_3)^2)} \\ & \times \frac{\|(\mathbf{v}_1 \bullet \mathbf{v}_2) \mathbf{v}_3 - (\mathbf{v}_1 \bullet \mathbf{v}_3) \mathbf{v}_2\|^2}{\|v_1\|^2(\|v_2\|^2\|v_3\|^2 - (\mathbf{v}_2 \bullet \mathbf{v}_3)^2)}, \end{aligned} \quad (2)$$

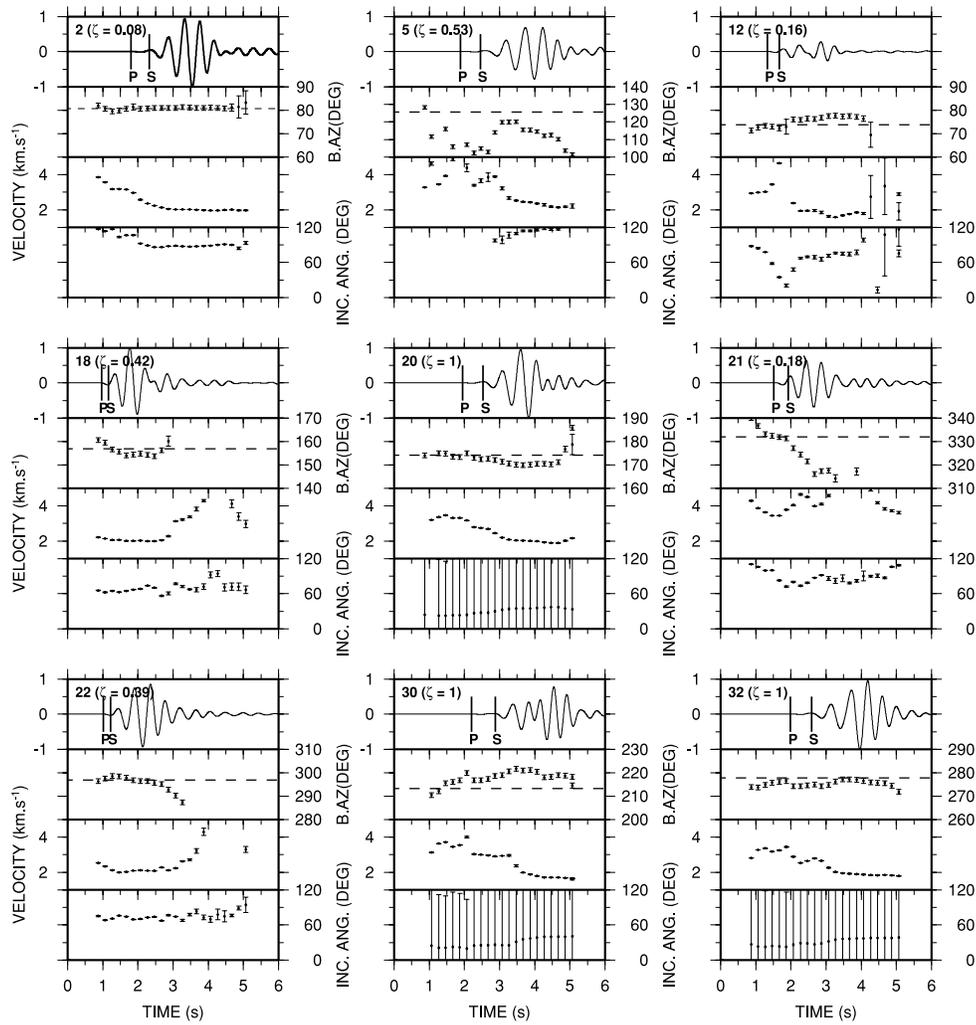
where for each spatial coordinate  $i$  ( $i = 1, 2, 3$ ), the vector  $\mathbf{v}_i$  is composed of the difference in this coordinate for all the pairs of sensors, and where the dot denotes the scalar product. Thus, for an array of  $N$  sensors,  $\mathbf{v}_i$  has  $N(N - 1)/2$  components. The procedure leading to eq. (2) is detailed in the appendix. Note that  $\zeta$  equal to 1 corresponds to a perfectly planar array. Calculating this index for all the arrays, we found that 10 of the arrays are perfectly planar (1, 3, 4, 7, 8, 20, 30, 32, 33, 34).

Fig. 6 shows the slowness vector, backazimuth, velocity and incident angle for simulation S1 obtained for several arrays distributed around the volcano at different distances from the source, (see Fig. 1 for the array locations). We used a 1.3-s window sliding along the recordings with a 0.2 s increment. We observe (1) the back azimuth corresponding to the first part of the signal corresponds well in most cases with the source azimuth (backazimuth calculated between the north direction and the segment made by the centre of the array and the source position). The difference is lower than  $10^\circ$ , except for array 5. For the majority of the arrays, the wavefield is probably dominated by the  $P$  waves. For arrays 5 and 18 the backazimuth differs from the true azimuth. Array 5 is located on the flank of a valley, which causes the misfit. Array 18 is close to the source and  $P$  and  $S$  waves are probably mixed. (2) In most cases, the velocity is higher at the beginning of the signal with values in agreement with the  $P$ -wave velocity for the deeper layer and decreases rapidly to values consistent with  $S$ -waves velocities or the  $P$ -wave velocity in the shallow layers. The later value of the velocity could also correspond to surface waves. For arrays 18 and 22, which are near the summit, the velocity is lower at the beginning of the signal and is consistent with  $S$ -waves velocities in the shallower layers. As  $P$  and  $S$  waves are probably mixed, it seems that  $S$  wave dominate the signal. (3) Incident angles cannot be determined for arrays 20, 30, 32 as all the receivers are situated in the same plane. Incident angles have high values for some far-field arrays (2, 5), which can be interpreted as wave propagation parallel to the surface. For the receivers close to the source (12, 18, 21, 22), low values of the incident angle are observed, which can be interpreted as direct wave propagation from the source. Nevertheless, incident angles estimated at these four arrays for the three different source depths do not show any significant variations. These results show the efficiency of the method for determining backazimuths and differentiating between different wave types when they are separated by enough time (at least one or more time window increments). Determining any reliable information from the incident angle is more problematic.

The slowness vector has been calculated at each of the 35 arrays for each simulation, yielding a time-series of backazimuths, apparent velocities and in some cases the incident angles. In the following, we focus on the backazimuth. Following the procedure described in Métaxian *et al.* (2002), a probability density function (PDF) of the backazimuth has been obtained for each array  $k$  by representing the backazimuth  $\alpha$  as a Gaussian variable with mean  $\theta_k(t)$ , standard deviation  $\sigma_k(t)$  and PDF

$$\rho_1^k(\alpha, t) = \frac{1}{\sqrt{2\pi} \sigma_k(t) \operatorname{erf} \left[ \frac{\pi}{\sqrt{2} \sigma_k(t)} \right]} \exp \left\{ -\frac{[\alpha - \theta_k(t)]^2}{2 \sigma_k(t)^2} \right\}. \quad (3)$$

We obtain a family of PDFs corresponding to each backazimuth value for a given array. The PDF are not independent from each other in the probabilistic sense. Each one is representative of the arrival of the seismic phase at a given time. Thus, we define a new probabilistic function, which is the average of the preceding one for a time window including the first arrival waves. As the time window (1.3 s) is large compared to the entire signal, the time



**Figure 6.** Backazimuth, velocity, apparent velocity (for arrays 20, 30 and 32) and incident angle estimated at different sites situated around the volcano. The location of the sites, given by the number in the top left-hand side, are shown in Fig. 1. The coplanar index ( $\zeta$ ) is indicated in the upper window at the right-hand side of array number. The dashed line in the Backazimuth subplot shows the true azimuth. The incident angle is shown up to 120 as the Arenal slope can reach 30. Vertical bars represent *P* and *S* wave arrivals obtained by ray tracing.

window includes the *P* and *S* waves. For the arrays nearest to the source, the average is performed over two or three values of the PDF of the backazimuth. For the most distal arrays, it is done over five to six values. A PDF is obtained for the backazimuth, the apparent velocity and the incident angle.

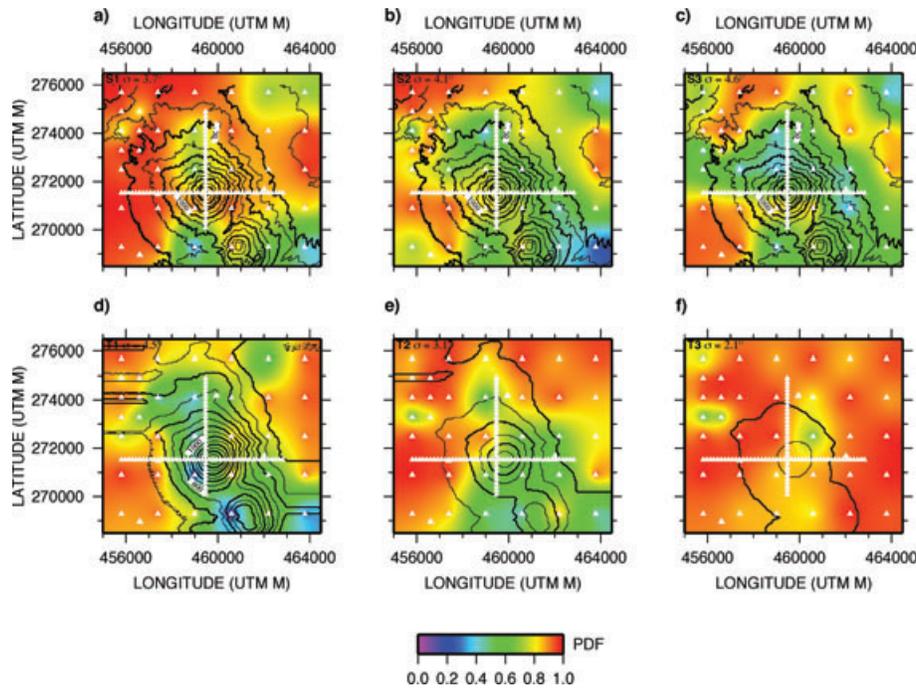
Fig. 5 displays the PDF of the backazimuth represented as rose diagrams for each of the 35 arrays, for simulation S1, S3, S4 and S7. Some weak differences of a few degrees are observed between various antennas for these four simulations. This is particularly evident for the antennas close to the source. To compare these differences over the 35 antennas, the standard deviation  $\sigma$  of the difference between the backazimuth and the source azimuth for the 35 arrays is listed in Table 1. Taking into account the topography, for different types of sources and different depths,  $\sigma$  varies between 4.1° and 5.4°. For an explosive source,  $\sigma$  increases slightly with depth, from 3.7° to 4.6° for S1–S3. This might be explained by the method of estimating the slowness vector. The uncertainty could be higher for rays arriving with lower incident angles.  $\sigma$  is higher for the other simulations with a maximum deviation of 5.4°. The same depth was used for these different simulations. This result could be a consequence of a different response of the topography to

different radiation patterns or wave paths. Also, for the non-isotropic source mechanisms *P* and *S* waves are generated which increases the intertwining of different phases leading to higher errors in the estimation of the slowness vector. Smoothing the topography T1–T4 reduces significantly the standard deviation. Without topography,  $\sigma$  is almost null, which clearly reveals the effect of the topography.

Next, we can calculate a spatial probability density function of the difference between the backazimuths estimated with the synthetics and the source azimuths by considering a straight line propagation path between the source and the array. Both estimated backazimuths and the source azimuths are measured with the centre of the arrays as the reference. The spatial coverage, one array every kilometre, should reveal the influence of the main topographical features. The PDF is expressed by

$$\alpha^k(x, y) = \frac{1}{\cosh\left[\frac{(\alpha_s - \alpha_c)}{\sigma_k}\right]}, \tag{4}$$

where  $\alpha_s$  and  $\alpha_c$  are, respectively, the source azimuth and the calculated backazimuth,  $\sigma_k$  controls the variability of the discrepancy between the backazimuth and the source azimuth. Its value has been



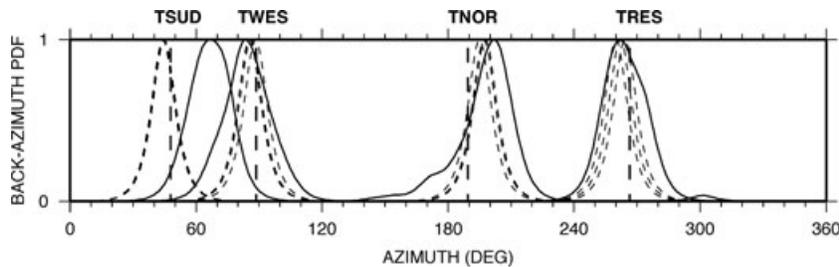
**Figure 7.** Spatial probability density function of the difference between the true and calculated backazimuth at each antenna with (using eq. 3) for simulations S1–S3 (a) to (c) and for simulations T1–T3 (d) to (f).

fixed to  $4^\circ$ . Fig. 7 displays the PDF of the difference between the source azimuths and backazimuths. For simulations S1–S3, (a–c) this function has lower values indicating a more pronounced difference for more pronounced topographic areas, such as the Arenal cone itself or El Chato structure. A semi-circular structure corresponding to a cliff is also responsible for low PDF values at the northwest of the volcano. For a progressively smoothed topography (Figs 7d–f), the PDF has globally higher values. There is a good correlation between the presence of the topography and the difference between the true and calculated backazimuths. The case for the perfectly flat horizontal surface is not shown, as the PDF value is almost 1 for the entire surface. This result suggests that topographic effects could explain most of the differences observed between the calculated and the theoretical slowness vector.

### 7 COMPARISON WITH REAL DATA

Finally, we compared the backazimuths calculated for real data recorded at the four triangular arrays consisting of three stations shown in Fig. 1 and the synthetic data recorded at the same location. A set of 20 volcanic explosions were selected. The PDF of backazimuth was calculated for each event and each antenna fol-

lowing the same procedure described above for synthetic data with eqs (3) and (4). This function is estimated for a time-series of backazimuths including the beginning of the signal. Considering that explosions occurred approximately in the same area and consequently the waves are affected in the same manner by the structure and the topography, we averaged the PDFs at each antenna. The resultant averaged PDFs are displayed in Fig. 8 where they are compared to functions estimated from simulations S1 to S3 calculated at the same four antennas. As observed with antennas composed of nine receivers, the difference between the true and calculated backazimuths is less than  $10^\circ$  for the simulations. The maximum of the PDF for the different simulations and the maximum of the average PDF for volcanic explosions differ by  $1^\circ$  to  $3^\circ$  and  $4^\circ$  for TWES,  $+7^\circ$  to  $+8^\circ$  and  $+14^\circ$  for TNOR,  $-3^\circ$  to  $0^\circ$  and  $4^\circ$  for TRES,  $-5^\circ$  to  $1^\circ$  and  $+20^\circ$  for TSUD. For the 3 antennas, TWES, TNOR and TRES, the PDF average value for the explosions fits quite well with the PDF obtained with the 3 different simulations. Topographic effects could thus explain part of the difference between source and observed backazimuths. The maximum of the averaged PDF for explosions recorded at TSUD differs by up to  $20^\circ$  from the source azimuth as well as from the backazimuth obtained with the simulations with a value close to  $N70^\circ$ . This direction corresponds



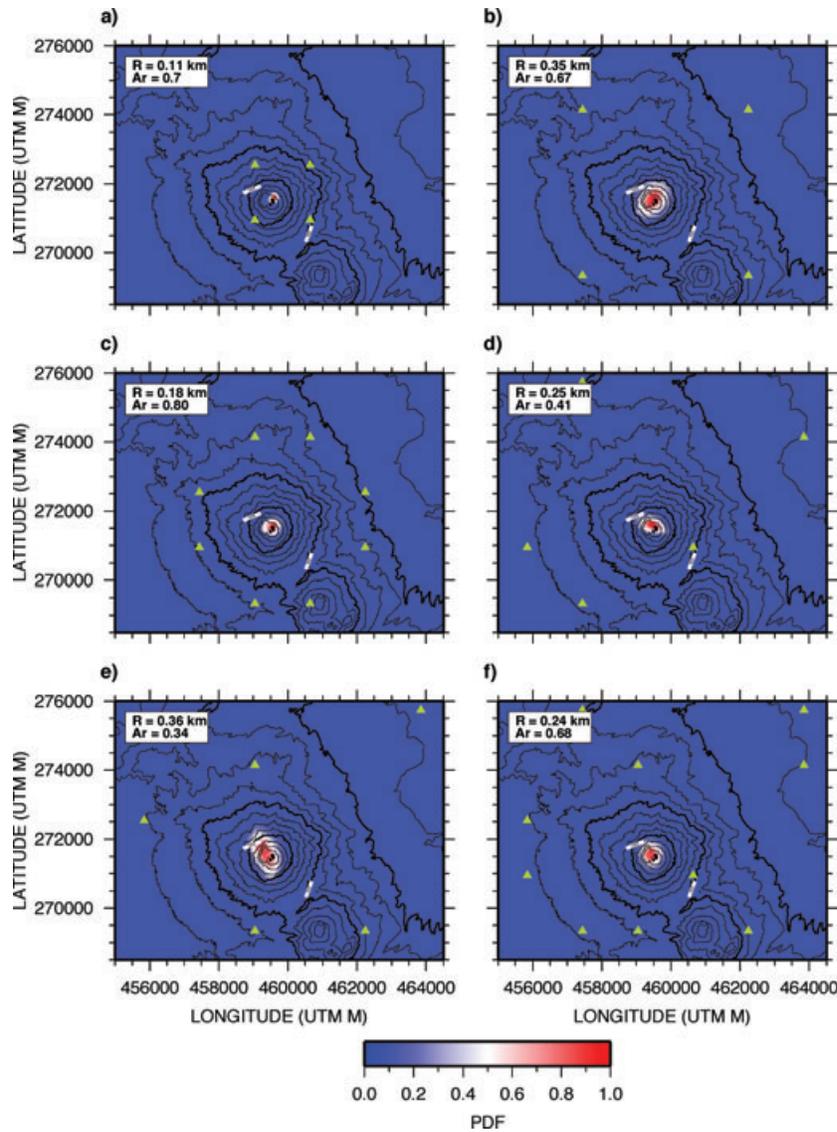
**Figure 8.** Probability density function of the calculated backazimuth for simulations S1–S3 (dashed lines) and averaged probability density function for 20 explosions (black lines) displayed for each triangular antenna, TSUD, TWES, TNOR and TRES. The vertical dashed lines indicate the true backazimuth of the synthetic events.

to a fault identified by (Mora *et al.* 2006) by seismic refraction experiments oriented SW–NE (Fig. 1). Moreover, this fault is associated with a depression or a graben situated between Arenal and El Chato and including Arenal lake situated SE from Arenal. Part of the real wavefield is possibly trapped in this depression and dominating the wavefield recorded by TSUD array. In this case, the difference between real and simulated data can't be explained by topographical effects. This difference is likely explained by an incompatibility between the real and the stratified medium adopted in this study and highlights the need for good near surface structural models.

### 8 DISCUSSIONS AND CONCLUSIONS

Topographic effects were investigated at Arenal volcano by using full wavefield synthetics for several receivers distributed all around the volcano. The receivers were arranged so that we could perform some array analysis. Several different simulations were performed where the source location, source type, velocity model and digital

elevation model were varied. The numerical output of the wavefield on the surface, 3-D temporal snapshots of the entire wavefield and a visual inspection combined with polarization analysis indicates that the wavefield is comprised of an initial onset of *P* waves followed by a mixture of scattered *P*, *S* and surface waves. Due to the complex topography and layered velocity model, this complex wavefield is to be expected. The slowness vector was estimated for the different simulations for every seismic antenna. Simulations show that most of the ray deviations observed at the antennas can be explained mainly by the topography. Without topography, deviations are close to zero. The ray deviations are spatially different, higher for more elevated or steeper areas or for structures with a specific or marked shape. The ray deviation also varies with the source depth and with the source mechanism. By comparing the backazimuth obtained with real explosions and with simulations it is noticeable that the results differ from the source azimuth in the same manner except for one antenna. The difference between the source azimuth for real data and synthetics can have different origins: (1) the low number of



**Figure 9.** Probability density function of the source position for (a) and (b) four antennas with the same azimuthally distribution but at different distances from the source, (c) eight antennas at intermediate distances, (d) five antennas situated on sites with low topographic effects, (e) five antennas situated on sites with strong topographic effects, (f) sum of the antennas of (d) and (e). The values of the mean quadratic (*R*) value and the aspect ratio (*Ar*) of the PDF are written in the upper left-hand side inset of each map.

receivers, the estimation of the slowness vector is less constrained with three receivers and the result can be strongly affected if one of the receivers is subject to a local site effect, for example, (2) the lack of a more realistic structure model. (Almendros *et al.* 2001a,b) show that the bias due to the cumulated effects of the structure and the topography can be up to  $10^\circ$  on Kilauea volcano. Therefore, without a realistic structural model, one may expect errors greater than  $10^\circ$  and (3) source depth variations can account for several degrees as seen with simulations S1–S3. Using such a coarse approximation to the subsurface velocity structure our horizontal locations fit our true synthetic locations surprisingly well.

Close to the source, direct waves and reflected or scattered waves on the free surface arrive nearly at the same time. Analysing the first part of the signal, we see that the backazimuth is influenced by the propagation of different waves. We obtain an averaged value of the back azimuth, which can point to the source if the *P* wave is dominating or can be influenced by other waves arriving just after if they have an equivalent amount of energy. The estimated backazimuth for the beginning of the signal can therefore be considered as a response of the structure and the medium to the wave propagation. In our case, the structure appears to be the dominate influence on the wave propagation. This response can be taken into account for locating real events.

The type of study carried out here could be useful in advance of deploying a field experiment to identify potential areas that can generate substantial distortions in the wavefield using numerical simulations. Fig. 9 compares the influence of the number and spatial distribution of antennas on our ability to recover the source location. In Figs 9(a) and (b), we compare the location PDF calculated using four antennas equally azimuthally distributed, but at different distances from the source. The mean quadratic radius is much smaller when the antennas are close to the source. In Fig. 9(c), we considered eight antennas at an intermediate distances with a good azimuthally distribution. The mean quadratic radius is almost as small as for (a) and the aspect ratio is higher. When antennas cannot be set up near to the supposed source location, a larger number of sites can provide similar results. In Figs 9(d) and (e), we considered antennas located in weak and strong topographic zones, respectively, (see Fig. 7 for spatial topographic effects). The azimuthal distribution of sites is comparable in both cases. The mean quadratic radius is clearly smaller and the aspect ratio is higher for weaker topographic zones. When combining the sites considered in (d) and (e), the location quality is greatly improved. As for (c), it is shown that a greater number of antennas, even if part of them are situated in sites with evident topographic effects, improve the location quality. The lack of true structural models is of course a very significant limitation as was shown by (Almendros *et al.* 2001b) and as observed with the analysis of signals recorded at the TSUD antenna. However, the backazimuth deviation for real data are not that different from the deviation observed with the synthetics for the three other antennas. Therefore, topographic effects could be enough to explain deviations in these cases. Synthetics can also be used to produce a slowness vector model for the entire 3-D edifice (Almendros *et al.* 2001b). We suggest a procedure to locate emergent volcanic earthquakes by using seismic antenna method: (1) detect the areas with strong topographic effects by using synthetics analysis before performing the field experiment, (2) if there is no well constrained structural model, compare the slowness vector estimated for synthetics and real data to select the seismic antennas affected mainly by topographic effects and (3) estimate a slowness vector model to correct these effects. The application of this methodology to Arenal volcano will be presented in a future paper.

## ACKNOWLEDGMENTS

The GSOB and CJB wish to acknowledge that this study was carried out in part by the Cosmogrid Ireland project funded under the Irish Government Programme for Research in Third Level Institutions, and by the 6th framework EU project VOLUME. The authors wish to acknowledge the SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support.

## REFERENCES

- Aki, K. & Richards, P.G., 2002. *Quantitative Seismology*, 2nd edn., University Science Books Sausalito, CA
- Almendros, J., Chouet, B. & Dawson, P., 2001a. Spatial extent of a hydrothermal system at Kilauea volcano, Hawaii, determined from array analyses of shallow long-period seismicity. Part I: method, *J. geophys. Res.*, **106**, 13 565–13 580.
- Almendros, J., Chouet, B. & Dawson, P., 2001b. Spatial extent of a hydrothermal system at Kilauea volcano, Hawaii, determined from array analyses of shallow long-period seismicity. Part II: results, *J. geophys. Res.*, **106**, 13 581–13 597.
- Almendros, J., Chouet, B., Dawson, P. & Huber, C., 2002. Mapping the sources of the seismic wave field at Kilauea volcano, Hawaii, using data recorded on multiple seismic antennas, *Bull. seism. Soc. Am.*, **92**, 2333–2351.
- Alvarado, G., 1989. Consideraciones neotectónicas recientes en los alrededores de la Laguna de Arenal, Costa Rica, *Boletín del Observatorio Sismológico y Vulcanológico de Arenal y Miravalles*, **2**, 6–21.
- Alvarado, G. & Barquero, R., 1987. Las se-ales sísmicas del volcán Arenal (Costa Rica) y su relación con las fases eruptivas (1968–1986), *Ciencia y Tecnología*, **11**, 15–35.
- Alvarado, G., Carboni, S., Cordero, M., Avilés, E., Valverde, M. & Leandro, C., 2003. Estabilidad del cono y comportamiento de la fundación del edificio volcánico del Arenal (Costa Rica), *Boletín del Observatorio Sismológico y Vulcanológico de Arenal y Miravalles*, **14**, 21–73.
- Bean, C., Lokmer, I. & O'Brien, G., 2008. Influence of near-surface volcanic structure on long-period seismic signals and on moment tensor inversions: simulated examples from Mount Etna, *J. geophys. Res.*, **113**, B08308, doi:10.1029/2007JB005468.
- Borgia, A., Poore, C., Carr, M.J., Melson, W.G. & Alvarado, G., 1988. Structural, stratigraphic, and petrological aspects of the Arenal-Chato volcanic system, Costa Rica: evolution of a young stratovolcanic complex, *Bull. Volcanol.*, **50**, 86–105.
- Chouet, B., 2003. Volcano seismology, *Pure appl. Geophys.*, **160**, 739–788.
- Chouet, B., Dawson, P. & Arciniega-Ceballos, A., 2005. Source mechanism of Vulcanian degassing at Popocatepetl Volcano, Mexico, determined from waveform inversions of very long period signals, *J. geophys. Res.*, **110**, B07301, doi:10.1029/2004JB003524.
- Di Lieto, B., Saccorotti, G., Zuccarello, L., La Rocca, M. & Scarpa, R., 2007a. *Continuous Tracking of Volcanic Tremor at Mount Etna*, Blackwell Synergy, Italy.
- Di Lieto, B., Saccorotti, G., Zuccarello, L., La Rocca, M. & Scarpa, R., 2007b. Continuous tracking of volcanic tremor at Mount Etna, Italy, *Geophys. J. Int.*, **169**, 699–705.
- Garcès, M.A., Hagerty, M.T. & Schwartz, S.Y., 1998. Magma acoustics and time-varying melt properties at Arenal volcano, Costa Rica, *Geophys. Res. Lett.*, **25**, 2293–2296.
- Goldstein, P. & Archuleta, R.J., 1987. Array analysis of seismic signals, *Geophys. Res. Lett.*, **14**, 13–16.
- Got, J.-L., Poupinet, G. & Fréchet, J., 1990. Changes in source and site effect compared to Coda Q-1 temporal variations using microearthquake doublets in California, *Pure appl. Geophys.*, **134**, 195–228.
- Got, J.L., Fréchet, J. & Klein, F.W., 1994. Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea, *J. geophys. Res.*, **99**, 15 375–15 386.

La Rocca, M., Saccorotti, G., Del Pezzo, E. & Ibanez, J., 2004. Probabilistic source location of explosion quakes at Stromboli volcano estimated with double array data, *J. Volc. Geotherm. Res.*, **131**, 123–142.

Leandro, C. & Alvarado, G., 1999. Estudio Geológico-Géofísico de una sección oriental y occidental en el volcán Arenal, *Boletín del Observatorio Sismológico y Vulcanológico de Arenal y Miravalles*, **20–21**, 48–58.

Lesage, P., Mora, M., Alvarado, G., Pacheco, J.F. & Métaxian, J.-P., 2006. Complex behavior and source model of the volcanic tremor at Arenal volcano, Costa Rica, *J. Volc. Geotherm. Res.*, **157**, 49–59.

Lokmer, I., Bean, C.J., Saccorotti, G. & Patané, D., 2007. Moment-tensor inversion of LP events recorded on Etna in 2004 using constraints obtained from wave simulation tests, *Geophys. Res. Lett.*, **34**, L22316, doi:22310.21029/22007GL031902.

Métaxian, J.-P., Lesage, P. & Dorel, J., 1997. The permanent tremor of Masaya volcano, Nicaragua: wave field analysis and source location, *J. geophys. Res.*, **102**, 22 529–22 545.

Métaxian, J.-P., Lesage, P. & Valette, B., 2002. Locating sources of volcanic tremor and emergent events by seismic triangulation: application to Arenal volcano, Costa Rica, *J. geophys. Res.*, **107**, 2243.

Minster, J.B., Jordan, T.H., Molnar, P. & Haines, E., 1974. Numerical modelling of instantaneous plate tectonics, *Geophys. J. R. astr. Soc.*, **36**, 541–576.

Monette, L., Anderson, M.P., Wagner, H.D. & Mueller, R.R., 1994. The Young's modulus of silica beads/epoxy composites: experiments and simulations, *J. Appl. Phys.*, **75**, 1442–1455.

Mora, M., 2003. Etude de la structure superficielle et de l'activité sismique du volcan Arenal, Costa Rica, *PhD thesis*, Université de Savoie, France.

Mora, M., Lesage, P., Dorel, J., Bard, P.-Y., Métaxian, J.-P., Alvarado, G.E. & Leandro, C., 2001. Detection of seismic site effects by using H/V spectral ratios at Arenal volcano (Cost Rica), *Geophys. Res. Lett.*, **28**, 2991–2994, doi:12310.11029/12005GL022666.

Mora, M., Lesage, P., Valette, B., Alvarado, G., Leandro, C., Métaxian, J.-P. & Dorel, J., 2006. Shallow velocity structure and seismic site effects at Arenal volcano, Costa Rica, *J. Volc. Geotherm. Res.*, **152**, 121–139.

Nakano, M. & Kumagai, H., 2005. Waveform inversion of volcano-seismic signals assuming possible source geometries, *Geophys. Res. Lett.*, **32**, L12302, doi:12310.11029/12005GL022666.

Neuberg, J. & Pointer, T., 2000. Effects of volcano topography on seismic broad-band waveforms, *Geophys. J. Int.*, **143**, 239–248.

O'Brien, G.S. & Bean, C.J., 2004. A 3D discrete numerical elastic lattice method for seismic wave propagation in heterogeneous media with topography, *Geophys. Res. Lett.*, **31**, doi:10.1029/2004GL020069.

Ohminato, T., Chouet, B.A., Dawson, P. & Kedar, S., 1998. Waveform inversion of very long period impulsive signals associated with magmatic injection beneath Kilauea volcano, Hawaii, *J. geophys. Res.*, **103**, 23 839–23 862.

Parsiegla, N. & Wegler, U., 2008. Modelling of seismic energy transport at volcanoes with real topography and complex propagation medium, *J. Volc. Geotherm. Res.*, **171**, 229–236.

Poupinet, G., Fréchet, J., Ellsworth, W.J., Fremont, M.J. & Glangeaud, F., 1985. Doublet analysis: improved accuracy for earthquake prediction studies, *Earth. Predict. Res.*, **3**, 147–159.

Ripperger, J., Igel, H. & Wassermann, J., 2003. Seismic wave simulation in the presence of real volcano topography, *J. Volc. Geotherm. Res.*, **128**, 31–44.

Rost, S. & Thomas, C., 2002. Array seismology: methods and applications, *Rev. Geophysics*, **40**, 1008, doi:10.1029/2000RG000100.

Saccorotti, G., Chouet, B. & Dawson, P., 2003. Shallow velocity models at Kilauea Volcano, Hawaii, determined from array analysis of tremor, *Geophys. J. Int.*, **152**, 633–648.

Schmidt, R.O., 1986. Multiple emitter location and signal parameter estimation, *IEEE Trans. Antennas Propag. AP*, **34**, 276–280.

Soto, G., 1998. Ceniza eruptada por el volcán Arenal, 1992-1997, *Boletín del Observatorio Sismológico y Vulcanológico de Arenal y Miravalles*, **10**, 14–24.

Soto, G.J., López, D.L., Fernández, J.F. & Alvarado, G.E., 1999. Caracterización geoquímica de las aguas termales del volcán Arenal (Costa Rica) dentro de su marco geovulcanológico., *Boletín del Observatorio Sismológico y Vulcanológico de Arenal y Miravalles*, **11**, 120.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys. Space Phys.*, **20**, 219–232.

Wegler, U., 2003. Analysis of multiple scattering at Vesuvius volcano, Italy, using data of the TomoVes active seismic experiment, *J. Volc. Geotherm. Res.*, **128**, 45–63.

Wegler, U. & Luhr, B.-G., 2001. Scattering behaviour at Merapi volcano (Java) revealed from an active seismic experiment, *Geophys. J. Int.*, **145**, 579–592.

## APPENDIX A: INDEX OF COPLANARITY OF A SET OF VECTORS

Let us first consider an overdetermined inverse problem, that is, a data space  $\mathbf{D}$  a model space  $\mathbf{M}$  and an injective linear operator  $G$  mapping  $\mathbf{M}$  into  $\mathbf{D}$ . The range  $R(G)$  of  $G$  is the linear set of data that are exactly explained by the theory. This suggests to evaluate the consistency of a given data vector  $d$  with the theory by considering the angle between  $d$  and the linear manifold  $R(G)$ . More precisely we can define the index of consistency  $i(d)$  of a given vector  $d$  in  $\mathbf{D}$  as

$$i(d) = \frac{P d \bullet d}{d \bullet d} = \frac{(P d \bullet d)^2}{(d \bullet d)(P d \bullet d)}, \quad (\text{A1})$$

where  $P$  is the orthogonal projector onto  $R(G)$  in  $\mathbf{D}$ , and  $\bullet$  stands for the scalar product in  $\mathbf{D}$ . Clearly,  $i(d)$  is the square of the cosines of the angle between  $d$  and  $R(G)$  and thus ranges between 0 and 1. Denoting by a star the adjoint operator, we can prove that the expression of  $P$  is

$$P = G(G^*G)^{-1}G^* \quad (\text{A2})$$

by remarking that  $G^*G$  is one to one since  $G$  is injective, and that:  $P^2 = P$ ,  $P^* = P$ ,  $P G = G$ . Note also that this index can be related to the concept of importance of data introduced by Minster *et al.* (1974).

Let us now turn to the definition of the index of coplanarity of a set of  $m$  vectors  $(v_1, \dots, v_i, \dots, v_m)$  in  $\mathbf{R}^3$ . First, define the three vectors  $v_\alpha$  ( $\alpha = 1, 2, 3$ ) in  $\mathbf{R}^m$  that, respectively, consists of one of the three components of the  $m$  vectors:  $v_\alpha = (v_1^\alpha, \dots, v_i^\alpha, \dots, v_m^\alpha)$ . Then, consider the inverse problem which, given vectors  $v_\alpha$ , corresponds to the following mapping from  $\mathbf{R}^2$  into  $\mathbf{R}^m$ :

$$\forall (p, q) \in \mathbf{R}^2, \mathbf{G}(p, q) = p v_1 + q v_2 \in \mathbf{R}^m. \quad (\text{A3})$$

Clearly, the mapping corresponds to the coplanarity of the  $m$  vectors  $(v_1, \dots, v_i, \dots, v_m)$  in  $\mathbf{R}^m$ , that is,  $v_3$  belongs to  $R(G)$  if and only if the  $m$  vectors  $(v_1, \dots, v_i, \dots, v_m)$  are coplanar in  $\mathbf{R}^3$ . A first step in the definition of the index of coplanarity of the  $m$  vectors is to use the index of consistency of  $v_3$  for this inverse problem. Using expressions (A1) and (A3), an easy calculation yields

$$i(d) = \frac{\|(v_3 \bullet v_1) v_2 - (v_3 \bullet v_2) v_1\|^2}{\|v_3\|^2 [\|v_1\|^2 \|v_2\|^2 - (v_1 \bullet v_2)^2]}. \quad (\text{A4})$$

By symmetry, we finally define the index of coplanarity of the set of vectors as

$$\zeta^3 = \frac{\|(v_3 \bullet v_1) v_2 - (v_3 \bullet v_2) v_1\|^2}{\|v_3\|^2 [\|v_1\|^2 \|v_2\|^2 - (v_1 \bullet v_2)^2]} \times \frac{\|(v_2 \bullet v_1) v_3 - (v_2 \bullet v_3) v_1\|^2}{\|v_2\|^2 [\|v_1\|^2 \|v_3\|^2 - (v_1 \bullet v_3)^2]} \times \frac{\|(v_1 \bullet v_2) v_3 - (v_1 \bullet v_3) v_2\|^2}{\|v_1\|^2 [\|v_2\|^2 \|v_3\|^2 - (v_2 \bullet v_3)^2]} \quad (\text{A5})$$

which yields expression (2).