



Comment on se-2021-52

Javier Almendros (Referee)

Referee comment on "Multi-array analysis of volcano-seismic signals at Fogo and Brava, Cape Verde" by Carola Leva et al., Solid Earth Discuss.,
<https://doi.org/10.5194/se-2021-52-RC1>, 2021

GENERAL COMMENTS

This is an interesting paper dealing with the application of array techniques to the problem of earthquake location. It constitutes a very clear demonstration of the fact that when a distributed network around the seismogenic area is not available, seismic arrays are a feasible alternative. The authors deal with earthquakes that are distributed around Fogo and Brava islands. Any seismic network based on land stations will be ill-suited, due to the limited extent of the Cape Verde islands. The authors use time-domain beam-forming at three seismic arrays to estimate the back-azimuths to the source, and design a scheme based on probabilities to pinpoint the epicenter locations.

However, there are some aspects of the work that need a critical revision. I address some of them below.

EPICENTRAL LOCATION

The locations obtained by the multi-array approach are limited to the determination of the epicenters using the estimates of back-azimuth. But sometimes the authors seem to oversee this limitation. At several points in the paper (see detailed comments below) you seem to imply that your method is better than others because you do not need a velocity model, with all the uncertainties that it brings to the calculations. However this is not a real advantage, since you provide only epicentral locations. The determination of depth requires some knowledge or assumptions about the velocity structure.

There is some discussion about source depths, although they are heterogeneous and somewhat confusing. In general, depth estimates are not addressed systematically. Sometimes a velocity model is assumed, either a two-layer model (line 229) or the model of Vales et al. 2014 (line 203) which is more complex. At one instance the depth is fixed (line 230). Depths are obliterated at the beginning but they are brought back during the discussion. I suggest that, if possible, you should exploit fully the information provided by the arrays, not only the back-azimuths but also the apparent slowness results, and include depths as part of the results. Of course the depths would be biased by the choice of velocity model, but at least you would have an estimate.

EARTHQUAKES VS HYBRIDS

The classification of seismic signals in "earthquakes" and "hybrid events" is not clear. Figure 7 compares the seismogram and spectrogram of a volcano-tectonic earthquake and a hybrid event, both recorded at the AF array on Fogo. Why don't you think that the hybrid event is in reality a small volcano-tectonic earthquake located in Fogo? I don't see a low-frequency coda, or any hint of a band-limited frequency content. Looking at the spectrum, it is true that there is relatively more energy in the 0-10 Hz band. For the hybrid, the peaks in the 0-10 and 10-20 Hz bands have similar heights. While for the VT earthquake there is definitely more energy in the 10-20 Hz band. Still, this could be an effect of lower signal-to-noise ratio. The noise seems to have bursts of energy in the 0-10 Hz band (before, during and after the event), so the difference could be just an effect of the size of the earthquake.

Figure S2 shows a single volcano-tectonic earthquake recorded at the AF array on Fogo and the BR array on Brava. The right plot looks similar to Figure 7b, except for the difference in signal-to-noise ratio.

In any case, I suggest that you justify your classification, showing hybrid events that look clearly different from local earthquakes. This has no impact in the method that you are presenting, which is valid for any earthquake: volcano-tectonic, hybrid, or whatever. But it is important for the interpretation of the activity in the discussion.

ARRAY ABERRATIONS

I don't agree on the use of the term "aberration" referred to the array, e.g. line 64: "seismic arrays can exhibit systematic aberrations of backazimuth and slowness", line 186: "Effects along the ray path from the source to the array, such as heterogeneities, can result in a systematic aberration of the array".

The heterogeneity of the medium can have an effect on wave propagation, resulting in rays impinging at the array site with back-azimuths that do not point to the source. See for example the interesting results of Garcia-Yeguas et al. GJI 2011. However, this cannot be pinned to the array! The seismic arrays are working ok, providing the directions with which the wavefronts are propagating through the array, these estimates are not biased in any way. The wavefield distortions (or aberrations) are produced in the medium. Therefore, I'd rather talk about ray bending or back-azimuth deviations, instead of array aberrations.

METHOD

The description of the method, and particularly the uncertainty estimates, is a bit confusing. You take as uncertainty the standard deviation of the maximum energy obtained by varying the start and duration of the stacking window. How do you obtain the standard deviation of back-azimuth and apparent slowness? Are these the values reported in tables S1,S2? Standard deviations of back-azimuth in those tables are large (around 100 degrees), but in the figures the beams seem to be much narrower. The width of the back-azimuth wedges is related to the uncertainty of the solutions, but I don't fully understand this relationship.

Additionally, the width of the array response is not considered at all. Even if the energy stack in figure 3 provides a very narrow peak, you should consider the uncertainty produced by the finite array response. In figure S1 I see that the central peak has a diameter of about 0.1 s/km, this should be always added to any uncertainty estimate.

For earthquakes, apparent slowness is smaller at BR. This should imply a larger uncertainty in back-azimuth, which is not clear in the plots.

Why do you use the time-domain beam-forming method? There are other options that may provide improved estimates of the back-azimuth and apparent slowness. Have you tested any other array method?

How are the results from the three arrays combined to provide the locations? (as in figure 6). It looks like you simply add the "energy" grids obtained for the arrays. However, if you interpret these maps in terms of probabilities of a grid node being the epicenter location, it make more sense multiplying them. After all this is an "and" operation, you want to determine if the position of a given node complies with the back-azimuths at array AF AND array CG AND array BR.

NETWORK LOCATIONS

I'd like to have some information about the quality of the network locations. They are used as ground truth in the comparison with the array locations. But the station coverage is poor and I assume that the network locations may have large uncertainties as well, even if you restrict them to the best cases. Perhaps you can show examples with error ellipses (in the supplementary information?).

OTHER COMMENTS

- line 54: the method of localization "has the advantage of being independent of velocity models". This is a bit misleading because you are just providing an epicentral location. If you try to locate the earthquake (including depth) you do need a velocity model.
- line 55: "The velocity structure is often very complex in a volcanic regime", I think that the term "regime" is not adequate here. Perhaps "environment", "setting", or "medium". (The same in line 242).
- lines 57-62: this description of the method is too detailed for the introduction, it could be later in the method section.
- section 2, the arrays are made of different types of seismic stations, short period and broadband. I assume you have taken good care of this and removed the instrument response, in order to make the waveforms comparable?
- line 73: "the arrays were designed for events with frequencies between 5 and 10 Hz". What do you mean? More than frequency of events, what's important in array design is the wavenumber, that includes the effect of frequency and apparent slowness.
- line 83: "absolute slowness", what do you mean with "absolute"?
- figures 3,4,5: one figure for one array may be enough to show how the method works.
- line 103: grid size of 124x124. If the grid search extends from -0.3 to 0.3 s/km, the grid spacing is about 0.005 s/km. This is unnecessarily small, given the shape of the array response.
- line 104: "the resulting energy stack is shown in Fig 3b". What do you mean by "energy stack"? You delay and sum the signals in the time domain, thus obtaining a seismogram representing the beam for a given apparent slowness and back-azimuth (e.g. green line in figure 3c). But this is still a time series, how do you get a unique value to be plotted in figure 3b? Is this the maximum amplitude of the beam? Is it the rms?
- line 106: the formula $BAZ=90-\text{atan}(s_x/s_y)$ does not follow the usual convention that $BAZ=0$ for north direction. When $s_x=0$ we obtain $BAZ=90$. This formula should be

corrected, or else explain your convention for azimuths.

- line 109: "the slowness is related to the angle of incidence by $\sin(i)=s*vc$, with the mean crustal velocity vc ". I have two criticisms. First, this s in the formula is APPARENT slowness (or ray parameter). Slowness (without "apparent") is just the inverse of velocity, and therefore it is not related to the incidence angle. Second, this formula is valid at points along the seismic ray, with local values of i , s_{ap} , and velocity. I don't think that the use of a "mean crustal velocity" is adequate in this context. What are i and s_{ap} then? If you are thinking of the arrival at the array, i is the incidence angle at the surface, and s_{ap} the apparent slowness measured by the array. But then v should be the velocity of the shallow layer, not an average of the crust.

- figure 6: there is a light-blue beam going SSW from the CG array, what is it? (the same happens in figures 12 and 13). The error bars seem underestimated. In the color scale of this plot, 90% of the maximum seems to correspond to the transition from yellow to orange. But in the map, the error bars do not include that transition, I'd say that the error area should include the whole yellow patch, e.g. ± 0.03 degrees approx from the epicenter.

- figure 6, caption: "the location of maximum energy estimated" is unclear. You are locating the epicenter of an earthquake. The colors indicate the directions in which the beamforming provides maximum energy, but in what sense the overlapping area has more energy? Also the sentence "The beams correspond to the beams" should be rephrased.

- line 156: "discrepancy of the amount of detected and with the multi-array analysis located earthquakes", weird sentence, rephrase?

- line 164: "on Brava the dominant frequencies of the same event are lower", why do you mention this, and even show figure S2? Since the earthquakes are closer to Brava than Fogo, is this an indication of a site effect?

- line 166: "ray turning point", that is true for rays propagating downward from the source. For an upward propagating ray there is no turning point.

- line 180: you superimpose the beams for all hybrids, why would you do that? Are the beams weighted in some way? Why didn't you do this for earthquakes as well? The white boundary in figure 10b leaves out the largest hybrid events, e.g. the northernmost, magnitude 1.5 event. Do you assume that the sources must be the same? What is the interpretation of this figure 10b?

- figure 8: the cumulative number of hybrids seems unrelated to the daily number of hybrids. The curve grows when there are no events (e.g. during April or June 2017), and the peak with 5 events in a day is not coincident with any increase in slope. I think this may be due to building this curve just joining points, when it should have a stair-like aspect (you should have right angles instead of the diagonal). The difference is not significant when there are many earthquakes, as if figure 8a. But it pops out when the number of earthquakes is small, as in figure 8b.

- figure 9: very small events ($M \sim 0-0.5$) near array AF and in areas near Brava, how can you locate these? Do you get good results for these earthquakes even at distances of 30 km from the array? This could be very interesting and emphasize the power of seismic arrays, but of course you need very fine noise conditions to perform the analysis, specially since beam-forming is carried out in the time domain.

- figure 10 has error bars in the locations of the hybrid events, but there were no error

bars in figure 9.

- section 4.3 is partly related to the method, in fact you refer here to section 3.2. Wouldn't the part on array parameters fit better in the method section?

- around line 195: it is not clear if you are talking about the start of the stacking window or the length of the stacking window.

- line 206: $rms < 0.25$ s?

- line 207: "(theoretical) back-azimuth". I don't think you should call "theoretical" to the back-azimuth obtained from the network location. The network location constitutes just another estimate of the epicenter location. And given the sparse network distribution, there is no reason to think that the network location will be better than the array location. So I would pick another word, perhaps "reference" back-azimuth?

- line 207: "slowness values", you mean apparent slowness? For this you do need a velocity model, I assume you use the same velocity model from Vales et al. (2014) ?

- figure 11: this is difficult to understand, since you represent back-azimuth and apparent slowness independently. The red dots are located in the coordinates (baz1,s1) given by the array analysis. But the green dots are not located at the coordinates (baz2,s2) given by the network locations. Instead, they are located at (baz2,s1) in the left panels and (baz1,s2) in the right panels. Why splitting them like this? These green points do not represent real estimates, and I find this figure very confusing, I don't know what you want to emphasize with this kind of plot. Have you considered x,y plots instead of polar plots?

- line 211: "yields back-azimuths pointing too far to the south by about 7 degrees", this is difficult to see in figure 11 because all points overlap in a tight cluster.

- figure 12: these examples may serve as a test to understand if your selection of uncertainty levels is adequate. Ideally, the uncertainty region in panels a and c should contain the solution shown in panels b and d. For example, if you choose a level of 0.7 to define the uncertainty region, the solution of panel b would be contained within the uncertainty region of panel a. It is important to understand that even if you use the maximum as best solution (red dots) the source could be anywhere within the uncertainty region (e.g. the orange-yellow areas, sum of energy above 0.7).

- line 249: "cannot establish a link between them". I agree, I think that even bringing up the subject is a bit far-reached.

- line 260: "from February to March and from June to September". Looking at figure S6 I see that there were more than 10 events in June and from September to November. There is a minimum in August. This does not correspond with your description, there is no increment in the hybrids in the period "June to September".

- line 266: "With the multi-array analysis it is not possible to estimate the depth of the events". This is not true in general. You can estimate the depth as well, if you exploit the information from apparent slowness and not only back-azimuth, as demonstrated for example by Almendros et al. 2001 and La Rocca et al. 2004. If you want to keep this sentence, you should specify that you are referring to your own implementation of multi-array analysis, focused on epicenter location.

- line 281: "Low rupture velocities and strong path effects result in the long low-frequency coda", regarding this discussion see also the paper by Bean et al. 2014 in Nature Geoscience, who discuss the same issue for long-period events.

- line 283: "we conclude". It is not clear to me the reason of this whole discussion about the origin of hybrids. It is not based on your results, and I don't see what is the consequence regarding the array analysis. Perhaps you could explain a little more, is there any evidence in the locations, or the array results, pointing to a fluid-related origin? Is there any evidence pointing to the contrary?
- line 288: "partly strong", what do you mean?
- line 289: "slowness variations", compared to what?
- line 302: "Being independent of any velocity model and able to locate events", as commented somewhere above, this is misleading. The multi-array method proposed does not use a velocity model BUT is able to identify just epicenters. You should make this clear at all instances, claiming otherwise is misleading. The location of seismic events (hipocentral location, complete with depth) requires a velocity model.
- line 313: parallel beams imply over-estimated distance? What do you mean exactly? To determine if the distance is over-estimated, you must know the distance. And this is precisely what you are trying to estimate.
- line 314: "closer to the expected location". What is the expected location?
- line 326: "This application allows the event localization without assuming a velocity model". Again, this is misleading. In reality you should talk about epicentral location. You cannot fix the depth without a velocity model.
- line 341: hybrids show significantly larger apparent velocities than volcano-tectonic earthquakes? You said that in average, at the Fogo arrays the apparent velocity is 7.1 km/s for volcano-tectonic earthquakes from Brava (line 165), and 7.8-8.4 km/s for hybrids (line 178). I don't think this is "significantly larger", specially considering the uncertainties in these estimates.
- line 348: the mention of volcanic tremor in the conclusions of the paper is out of place.