Instantaneous phase-based statistical method for detecting seismic events with application to Groningen gas field data

Oleg V. Poliannikov and Michael C. Fehler*
Earth Resources Laboratory, Department of Earth, Atmospheric, and Planetary Sciences, MIT

SUMMARY

We propose a new method for detecting seismic events recorded by an array of receivers. The method uses an approximate moveout of a known master event to stack the normalized traces containing only the instantaneous phase. We show that the instantaneous phase stack is a stationary Gaussian process, which makes event detection a classical statistical problem of finding outliers in a sample of multidimensional Gaussian vectors. Application to Groningen field data shows a very good performance of our detector.

INTRODUCTION

The Groningen gas field has been experiencing induced seismicity most likely associated with compaction and stress redistribution due to gas production. A large shallow borehole-monitoring array was installed in late 2015, and is continuously recording seismic activity in the area. Although a large number of events have been recorded, detected, located, and cataloged, it is clear that many more recorded events remain undetected. Detecting smaller events, such as aftershocks or precursors, could shed additional light on geomechanical processes responsible for seismicity.

The goal of this paper is to present an event detection method that takes advantage of the acquisition available at Groningen and of work that has already been done. At the same time the method is quite general and readily portable to many other fields. Plans to eventually apply this method to months, and possibly years worth of data, collected over several hundreds of stations impose important constraints. The method must be fast, and in particular not rely on repeated forward modeling in a complicated velocity model. The method should ideally be fully automatic, require a set of seismic traces as input and output events in a format that allows easy quality control. A large number of faults and a very complicated velocity structure yield a variety of source mechanisms even in a small volume, so waveform template matching seems to be a poor choice from the outset. The method should be flexible to accommodate non-stationary background noise that in addition to conventional seismic events may include other signals as well as various cultural and technological noise.

Our method starts with a master event. This typically has a relatively high magnitude, very good SNR, it has a cataloged location and origin time. Master events in the Groningen catalog span a large area (although not uniformly). They provide us examples of test moveout curves that represent not just them but also foreshocks and aftershocks originating in their vicinity. If a phase coherence, to be precisely defined, is observed along a reasonable moveout then the arrival of an event may be declared. We measure phase coherence by dividing recorded data by their envelopes, and stacking them along the test moveout curve. We show the resulting phase stack can be modeled as a stationary (non-white) Gaussian process. Events whose kinematics are similar to those of the master event create outliers in the phase stack that can be analyzed using standard statistical tools.

We process 4 hours of data surrounding two catalog event times, and calculated phase stacks of bandpass filtered data. We demonstrate that those phase stacks are realizations of the stationary Gaussian process with sporadic outliers on top of that. By looking at the most prominent outliers, we find 12 more events (in addition to the 2 catalog events) within that time period.

INSTANTANEOUS PHASE STACK

Our event detection method is general but we will illustrate each step with an application to the 4 hour dataset recorded at the Groningen gas field. Our goal is to find events in a location close to the hypocenter of a known master event. Detecting aftershocks and possibly foreshocks will allow us to better understand geomechanical processes that led to and followed the large event. Two events, among many others, were recorded by multiple stations in Groningen. A magnitude 1.9 event on 2016-11-01 at 00:12:28 was followed by a magnitude 2.2 event that occurred on 2016-11-01 at 00:57:46.

The stations we used are closest to the estimated epicenter: G67, G23, G29, G19, G24. Figure 1 shows a map view of the event epicenters and the 5 closest stations. Each station here is actually a borehole that consists of four three-component receivers at depths 50, 100, 150, and 200 m. We will focus on detecting the P-wave and

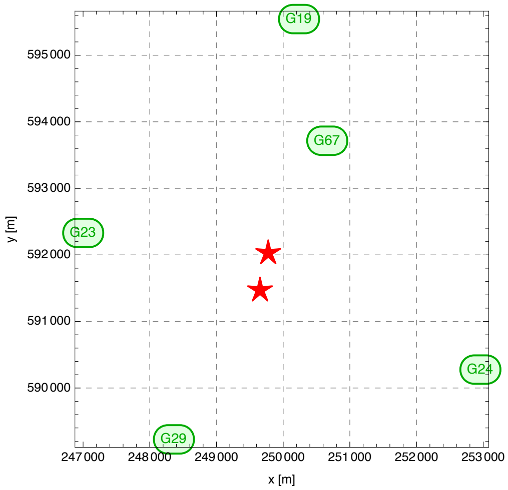

Figure 1: *Catalog event epicenters and 5 nearby stations' locations*

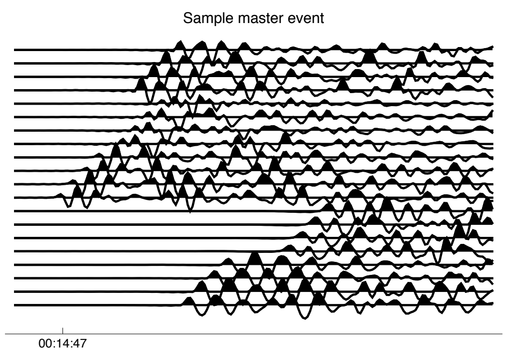

Figure 2: *Recorded data corresponding to the first of the two catalog events. The other event has similar kinematics but different waveforms.*

therefore will use only the vertical component.

Figure 2 shows recorded data around the arrival time of the first catalog event. We display 4 traces per station (borehole) one underneath the other, and repeat for all stations. The final gather thus shows 20 traces. The exact geometry of the acquisition array is not important.

The arrival times of both catalog events appear very similar. However, the polarities of the direct P-wave arrivals do not always match. We can conclude without a more detailed kinematic analysis that the two events originate in the same approximate area but they have different source mechanisms and source time functions. Estimated catalog locations confirm that prediction. Any aftershocks or precursors that occur in the same area will manifest themselves similarly in that they will tend to have a similar moveout. Their amplitude variation will not necessarily match that of the catalog events but it should display some coherence or else we would likely conclude that no event is present there.

Phase information of recorded data

Denote recorded amplitudes $d_j(t)$, $j \in \{1,\ldots,N_\mathrm{r}\}$, $t \in [t_0,t_f]$, where j is the receiver number, N_r is the total number of receivers, and $t \in [t_0,t_f]$ is the recording time. In our case we are working with a 4 hour window, so that $t_0 = 0$ and $t_f = 14400$ s. In practice, the data are available only at discrete times, possibly different for different receivers. We assume that they are resampled and/or interpolated in such a way that we can treat time as discrete or continuous as convenient. We have visually inspected spectrograms of our data and filtered them to the band of $5-45$ Hz away from the very low frequency and the electric noise bands while maintaining the band where the signal energy is mostly concentrated.

A master event has an estimated moveout $\{\tau_j\}_{j=1}^{N_\mathrm{r}}$, where τ_j is the arrival time of the master event to the j-th receiver. Because we will use the moveout only to shift traces, we can assume that $\tau_1 = 0$, and all other τ_j are delays relative to τ_1. In our processing, we have selected a small time window that contains the direct P-wave arrivals of the master event, and then estimated relative arrival times by cross-correlating envelopes defined below.

Our detection method is based on analyzing of the instantaneous phase of the data. We first define a complex-valued analytic signal

$$d_j^a(t) = d_j(t) + i\mathscr{H}[d_j](t), \tag{1}$$

where $i^2 = -1$, and $\mathscr{H}$ denotes the Hilbert transform. The envelope $e_j(t)$ and the instantaneous phase $\theta_j(t)$ of d_j are then the absolute value and the phase of the analytic signal, respectively:

$$d_j^a(t) = e_j(t)\exp\big[i\theta_j(t)\big]. \tag{2}$$

We see from the the above definitions that the ratio $\varphi_j(t) = d_j(t)/e_j(t) = \exp\big[i\theta_j(t)\big]$ satisfies $|\varphi_j(t)| \le 1$. It contains all phase information and virtually no amplitude information. Working with only the phase allows us to detect events of different magnitudes and normalize large period noise variations as well as equalize contributions of different receivers to the event detection functional as will become apparent soon.

In order to detect events whose moveouts resemble the moveout $\{\tau_j\}_{j=1}^{N_\mathrm{r}}$ of the given master event, we apply a moveout correction of the recorded data, $\tilde{d}_j(t) = d_j(t-\tau_j)$, so that the direct P-wave of detected events should be observed around the same (moveout-corrected) time. The moveout corrected phase $\tilde{\varphi}_j(t) = \varphi_j(t-\tau_j)$ should

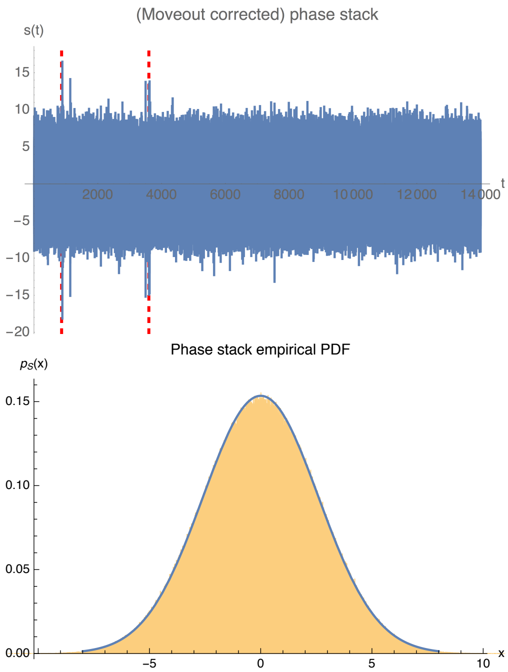

Figure 3: *(top) Phase stack as function of the moveout-corrected time, and (bottom) its histogram.*

correlate around the time of the P-wave arrival to the first receiver The phase at each trace is affected by the radiation pattern of the source and we may have to correct for the sign to ensure phase uniformity across the receivers. In our processing flow, we repeat the same algorithm for each possible combination of sign flips, $2^{N_\mathrm{r}-1}$ altogether. Observe that for our acquisition geometry with 5 boreholes and 4 levels in each borehole, the polarization sign will be the same for all levels of any given station because the polarity does not change from one level to another due to their proximity and a simple shallow part of the velocity model. Therefore we need to consider only a small number of $2^{5-1} = 16$ different sign combinations. If the number of station were larger, we would likely try to save computational time by assuming that the polarization only depends on the azimuth thereby also reducing the total number of sign combinations to a manageable level.

Single-point event detector

We will begin with a simplified (but fully functional) version of the event detector for presentation purposes. A slightly more complicated multidimensional version will follow immediately after. The time-shifted phase data $\tilde{\varphi}_j(t)$ is sampled at a sampling frequency F_s to yield a matrix

$$\phi_{jk} = \tilde{\varphi}_j(t_k),\ |\phi_{jk}| \leq 1. \tag{3}$$

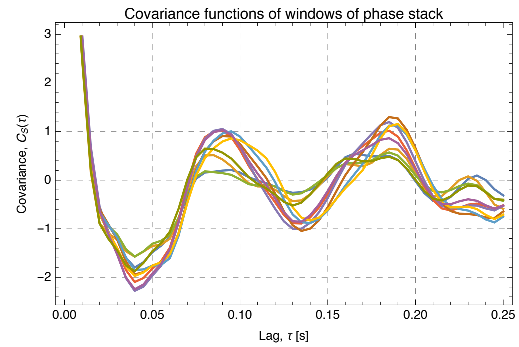

Figure 4: *Covariance functions of the phase stack calculated for ten non-overlapping windows are shown together.*

If t_k is an event time then for suitable polarization sign corrections, all $\{\phi_{jk}\}_{j=1}^{N_\mathrm{r}}$ will have the same signs. The sum of variables that have the same sign may reach $\pm N_\mathrm{r}$.

If t_k is away from any event arrival then $\{\phi_{jk}\}_{j=1}^{N_\mathrm{r}}$ will be a sequence of almost uncorrelated variables. The sum of such variables is approximately Gaussian according to the Central Limit Theorem. Define the instantaneous phase stack

$$s_k = \sum_{j=1}^{N_\mathrm{r}} \phi_{jk}. \tag{4}$$

Figure 3 shows the both the stack $\{s_k\}_{k=1}^{N_\mathrm{t}}$ and its histogram The two largest outliers marked with red dashed lines occur at the time of the catalog events. Other outliers from the Gaussian distribution are also clearly visible suggesting the presence of other events. When the phase stack exceeds an *a priori* chosen threshold, that serves as an indicator of the presence of an "event", whereas staying below the threshold would be interpreted as "noise". In essence, we first decide how many traces could have phases aligned just at the right time by a mere accident, and how many are enough to suspect that an event that is coherent across traces must have arrived.

Multidimensional event detector

From the statistical perspective, the single-point event detector implicitly assumes that the phase stack is an array of independent, identically distributed random variables. A threshold chosen to select outlier corresponds to some confidence interval centered at zero. Thresholding the phase stack is equivalent to testing a hypothesis that the observed sample is noise with a prescribed confidence. Values that are too far from zero are deemed highly unlikely and thus inconsistent with the noise-hypothesis.

The bandpass filtered recorded data are in practice correlated. They are continuous, and the signal value at the next sample is very close to the signal value at the current sample. The correlation remains relatively high for small time delays and it drops for larger time delays. We have taken the four-hour data, split it into 10 24-minute windows, and separately calculated the covariance function for each window. Figure 4 shows all 10 covariance functions plotted together. It is clear that the covariance function does not significantly change from one window to another. This suggests that we could model the phase stack as a stationary Gaussian process with an empirically calculated covariance function $C(t)$ shown in Figure 4. It is very important that the assumption of stationarity has not been simply adopted from the outset for theoretical tractability or other convenience. Our surprising finding is that the "noise" in the phase stack actually appears stationary with a fixed covariance (or correlation) function, which in turn indeed makes the outlier detection problem amenable to standard statistical analysis.

Fix an integer window size N_{w}. Then most vectors of consecutive phase stack values are samples from multidimensional zero-mean Gaussian distribution

$$(s_i,\ldots,s_{i+N_{\mathrm{w}}-1}) \sim \mathscr{N}(\mathbf{0}_{N_{\mathrm{w}}}, \Sigma_{N_{\mathrm{w}}\times N_{\mathrm{w}}}), \tag{5}$$

where

$$\Sigma_{N_{\mathrm{w}}\times N_{\mathrm{w}}} = (\Sigma_{kl})_{k,l=1}^{N_{\mathrm{w}}},\ \Sigma_{kl} = C(|k-l|/F_{\mathrm{s}}). \tag{6}$$

An outlier is a vector of values that can be a sample from the distribution given by Equation 5 only with a very low probability. More formally, a confidence region that contains $1-\alpha$ of the probability mass of that distribution is an ellipsoid given by

$$\boldsymbol{s}^{\top}\Sigma_{N_{\mathrm{w}}\times N_{\mathrm{w}}}\boldsymbol{s} = \chi^2_{N_{\mathrm{w}}}(\alpha), \tag{7}$$

where $\chi^2_{N_{\mathrm{w}}}(\alpha)$ is the α-quantile of the χ^2 distribution with N_{w} degrees of freedom. Phase stack windows of length N_{w} that lie outside of this N_{w}-dimensional confidence region are outliers much like one-dimensional ones we have seen in Figure 3.

APPLICATION TO GRONINGEN DATA

We finally apply the multidimensional event detector to the 4 hour of data. 14 events have been detected altogether, including the two known catalog events. Figure 5 shows some of these events in the order of decreasing confidence, which can be measure by the left hand side of Equation 7. Our results show that we are able to detect many more events than are currently in the catalog, and the SNR of some detected events is quite low.

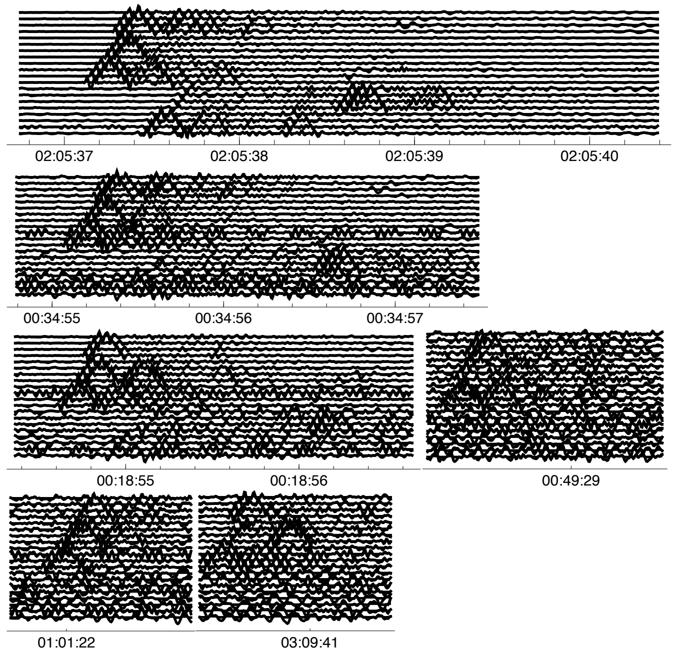

Figure 5: *Examples of detected events. Other events are omitted for space. The events are ordered by decreasing confidence.*

CONCLUSIONS

Most definitions of a seismic event either explicitly or implicitly include the notion of an expected moveout. Many events in the Groningen field that are already detected and located provide means to test different areas for smaller events.

Noise in field data is complicated and non-stationary. However, the instantaneous phase stack proposed in this work appears to behave as a stationary Gaussian process with zero mean and a covariance function easily computed from the recorded data. Events are then hidden among statistical outliers of the Gaussian process. They can be found using standard statistical analysis of Gaussian processes.

The performance on our method is quite good. We were able to find events that were neither in the catalog nor known to the authors from any other source.

ACKNOWLEDGEMENTS

We thank Yusuke Mukuhira and Brian Demartin for various fruitful discussions of this project. ExxonMobil provided support for this work under a project titled Fundamental Research in Induced Seismicity under master agreement EM09079.