Catalog of Northern California earthquakes recorded by DAS

Siyuan Yuan, Eileen R. Martin, Jason P. Chang, Steve Cole, and Biondo Biondi

ABSTRACT

We catalogued more than 800 seismic events recorded at Stanford Distributed Acoustic Sensing Array (SDASA-1) from September 2016 to August 2017. The catalog is being continuously updated as new events occur. We have developed open-source interfaces so that users can query the database and extract earthquake recordings efficiently. Pulling the data via the interfaces, we performed signal repeatability analyses for blasts at nearby Permanente Quarry and nearby weak earthquakes from Ladera and Felt Lake. We found that geographically close events could have repeatable signals in terms of S-wave arrivals and surface-wave phase changes. With rich event recordings, the catalog enables us to extract and characterize distant and weak events, which we will use to quantify our array's sensitivity and study event detection and noise attenuation algorithms in future work.

INTRODUCTION

Distributed Acoustic Sensing (DAS) holds great promise for application in cost-effective monitoring of microseismic signals and detecting earthquakes. Unlike the mainstream work previously done using DAS (e.g., using fibers in wells or burying the array in trenches), Stanford DAS Array-1 (SDASA-1) uses a fiber-optic cable laying in an already existing polyvinyl chloride (PVC) conduit buried in the ground that makes the installation more convenient and economic (Martin et al., 2017; Biondi et al., 2017). Understanding how well this type of cost-effective array records seismic events is one of our objectives. Some preliminary event analyses have been done. Through analysis of six seismic events recorded by SDASA-1, Biondi et al. (2017) show the recorded data can provide us with valuable information, thereby demonstrating the suitability of our DAS array for event detection. We would like to get a better understanding of our array by further analyzing many more events. The location of SDASA-1 in a tectonically active location allows us to record hundreds of events per year.

Due to the large number of the events, manual data management and extraction become infeasible. Therefore we built an event catalog system to manage the data automatically. The catalog was built based on the United States Geographical Survey (USGS) online database, and we also provide a program that can pull the broadband

data recorded at the Jasper Ridge Seismic Station (JRSC) station for comparative studies. Additionally, we offer open-source interfaces for the users to select only events of interest and to extract the event data recorded by DAS efficiently. Using the interfaces, we performed the signal repeatability analysis in a more efficient manner. Here we analyze data from blasts at Permanente Quarry and several natural seismic events. This analysis shows some repeatable patterns.

In the next section, we give an overview of the passive data from which we extract earthquake recordings. Then we show how we manage the passive data, how to build the catalog, and how the catalog interfaces work. We then provide examples using the database to perform repeatability analyses for 67 quarry blasts, two nearby natural earthquakes from Ladera and two natural earthquakes from Felt Lake; and we present the repeatable wave arrival timings and waveform patterns we found.

OVERVIEW OF PASSIVE DATA

Martin et al. (2017) and Biondi et al. (2017) show the SDASA-1 array design, geometry, and experiment setup. We used passive data recorded by SDASA-1 to extract earthquake recordings. To interpret our results in the following sections, we first give an overview of the data.

The data have been recorded by 626 channels along the SDASA-1 since September 2016. SDASA-1 is a double-loop array with the layout outlined by the red line in Figure 1. At the end of the first loop, two fibers in the same jacket are spliced end-to-end. We numbered the channels in the interlacing order with a 4.08-meter effective spacing. The data are recorded by each channel on both fibers simultaneously.

There are two recording modes: active and passive. For the active mode, the sampling rate is 2500 samples per second; while 50 samples per second is the rate for the passive mode. The active data have 4x denser channel spacing: 1.02 meters. Active data and passive data require different processing procedures. Because the active mode accounts for a negligible proportion of recordings compared to the passive mode, we focus on the passive mode data with the 25 Hz Nyquist rate for now.

MANAGEMENT OF PASSIVE DATA

Our first step in data management was to organize files into smaller, easily navigated subdirectories. The passive data files initially show up in a directory called /data/biondo/DAS in SEG-Y format and are named based on the UTC time stamp of its first sample. For a limited number of files this could be the end of the story, but even simple file system operations like an *ls* become painfully slow when thousands of files are in one directory. Therefore, a python script called *checkOrganizeFiles.py* is run regularly to move files into subdirectories named as /data/biondo/DAS/*year*/*month*/*day*. For example, a file starting on December 10, 2016 at 13:01:54 and 351 millisec-

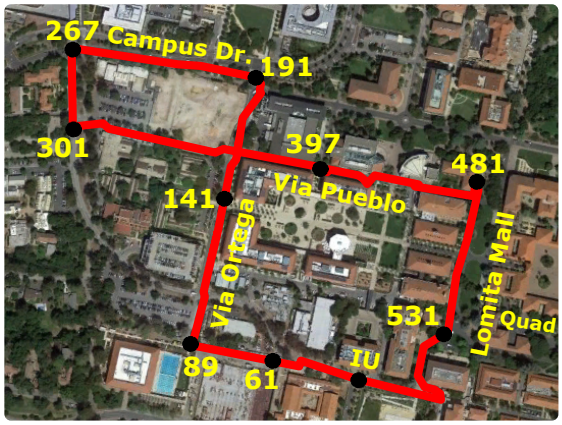

Figure 1: The layout of the fiber following telecommunications conduits overlaid on the map. The longest linear section is roughly 600 meters wide. Some deviations from straight lines had to occur due to existing conduit geometry constraints. [**NR**]

onds would show up as /data/biondo/DAS/cbt_processed_20161210_130154.351+0000.sgy and the next time *checkOrganizeFiles.py* is run, it will move to /data/biondo/DAS/2016/12/10/ cbt_processed_20161210_130154.351+0000.sgy.

Further organization is based on the idea of *fileSet* objects, defined in *fileSet.py*. A *fileSet* is a collection of files named using a particular grammar that allows times to be associated with the names of files. Given a file name with that grammar, the *fileSet* can interpret the start time of the file (as a python *datetime* object), and given a start time, the *fileSet* can generate the name of a file that starts at that time. To define a *fileSet*, one first creates a list of strings, some of which are keywords tied to parts of the timestamp. For instance, when files initially show up in /data/biondo/DAS, the *fileSet* grammar that describes them is ['/data/biondo/DAS/cbt_processed_', 'year4','month','day','_','hour24start0','minute','second', '.','millisecond', '+0000.sgy']. The *fileSet* that describes files that have already been organized into subdirectories is ['/data/biondo/DAS/','year4','/','month','/','day','/cbt_processed_', 'year4', 'month', 'day','_','hour24start0', 'minute','second','.','millisecond','+0000.sgy'].

However, we do not always know the start time of every file. We just know that we want files that contain time A to time B. One clunky way to do this is every time you want data between A and B, you list all the files in the /data/biondo/DAS/*year*/*month*/*day* subdirectory for each *year*/*month*/*day* between A and B, use the *fileSet* described above to check their start times, also check the SEGY header with the file length, then generate a list of files. Instead, we streamline the process by organizing consecutive files into *regularFileSet* objects, which are a type of *fileSet* that describes files that

are the same length, same sample rate, the end of one file is right before the start of the next, and all are named using the same grammar. The *regularFileSet* object is defined by the start time of the first file, the number of consecutive files, the grammar for naming, and the length in seconds of the files. A *regularFileSet* also tracks the end of its last file, can return the name of the file in that *regularFileSet* containing a particular time, and can build a list of file names that include the data between time A and time B.

Generally, the interrogator chugs along without interruption, so we often have *regularFileSets* made up of hundreds or thousands of files. But we needed to automate the tracking of all these *regularFileSets* because there are occasional hiccups: a software license gets renewed, a power outage, a switch between active and passive recording, among other reasons. To do this, we have written a script called *fillSQLiteDB.py* to build and add on to a SQL database *regFileSetLog.sqlite* where each row is a distinct *regularFileSet*. Since there are typically hundreds or thousands of files in a *regularFileSet*, this means that the SQL database contains relatively little metadata, making it easy to navigate and figure out which file names should be used to study a particular event or time window.

The *fillSQLiteDB.py* script requires the *sqlite3* package and accepts three command-line arguments. The first is a text file containing the paths to directories containing the SEG-Y files to be added to the SQL database. While users can specify the directories directly containing the SEG-Y files, they can also specify directories containing subdirectories of SEG-Y files. Additionally, users can include multiple directories to search through by specifying them on different lines. As an example, if users would like to add SEG-Y files from all of August and from only the first day of September, the text file would look like:

```
/data/biondo/DAS/2017/08/
/data/biondo/DAS/2017/09/01/
```

Note that if there are already some files from August in the database, the script will be able to append *regularFileSets* to the existing database entries. The second argument is the path of the directory containing the database. By default, the database will be located in an associated subdirectory *log* and named *regFileSetLog.sqlite*. The third argument is a flag for whether to print out status updates to screen (1 for yes, 0 for no). The first few entries of the SQL database are shown in Figure 2.

EVENT CATALOG AND INTERFACES

The essential idea of building the event catalog was that for each event around the Bay area since September 2016, we locate the passive data using the *regFileSetLog.sqlite* SQL database and python tools described in the previous section, preprocess and save the data; and then record the data file paths and file names along with the metadata

	startTime	endTime	secondsBetweenFiles	nFiles
0	2016-09-02 17:38:53.923000	2016-09-02 17:43:53.923000	60.0000	5
1	2016-09-02 17:43:53.923000	2016-09-02 17:44:54.923000	61.0000	1
2	2016-09-02 17:44:54.932000	2016-09-07 00:08:54.932000	60.0000	6144
3	2016-09-07 01:06:54.932000	2016-09-12 13:05:54.932000	60.0000	7919
4	2016-09-12 13:05:54.932000	2016-09-12 13:06:17.752000	22.8200	1
5	2016-09-12 13:27:43.376000	2016-09-17 16:38:43.376000	60.0000	7391
6	2016-09-17 16:38:43.374000	2016-09-17 16:39:43.374000	60.0000	1
7	2016-09-17 16:39:43.373000	2016-09-17 16:41:43.373000	60.0000	2
8	2016-09-17 16:41:43.372000	2016-09-17 16:42:43.372000	60.0000	1
9	2016-09-17 16:42:43.374000	2016-09-17 16:43:43.374000	60.0000	1

Figure 2: First nine entries of the SQL database for passive data management. [**NR**]

in our event SQL database *event_log.sqlite*. We also created interfaces for the users to efficiently select the events of interest and extract the event data.

To start with a reliable set of events, metadata are pulled from the USGS online database. These metadata include the event start time, magnitude, and location of the epicenter. We calculate the three-dimensional (3D) distance from the epicenter to our array for each event. With the start time and estimated time range (proportional to the distance to our array) for each event, and taking advantage of data managing tools in the previous section, we locate the raw data on our storage system. The original data are proportional to strain. We take their temporal derivative to convert them to strain rate, which has a broader spectrum, useful for visualizing events (Martin et al., 2017). We then high-pass the data by setting the low-cut frequency to be 8/(event time range), considering that the closer events are to us the greater proportion high-frequency components arise in the spectrum. Last, but not least, at each time slice we calculate the median data rate and subtract the median value from the data rate recorded by each channel to compensate for possible laser drift occurring on all channels at once. We save the preprocessed data into both a SEG-Y file and a SEPLib file format organized into a /data/biondo/DAS/EventCatalog/Data/*year*/*month*/*day* folder. The file names and data paths were recorded in our event catalog SQL database. We have cataloged more than 800 events ranging from September 2016 to August 2017.

The interface of the event catalog takes time range, magnitude range, epicenter depth range, distance to SDASA-1 range, and azimuth angle to SDASA-1 range as inputs to filter the events and return the corresponding metadata, data paths, and

file names for both of the SEG-Y and SEPLib data files to users. Users can choose to either set every input option or choose any subset of the options of particular interest. They may save a limited amount of data to their local directories for further analysis. In the next section, we show how we used our catalog for signal repeatability analyses.

SIGNAL REPEATABILITY ANALYSIS

Because weak events tend to be contaminated by local noise, especially during daytime, we are always interested in whether or not the data recorded are from the actual events. A nearby quarry blast (around 13km from SDASA-1) and some frequently occurring nearby earthquakes provided us with an opportunity to assess signal repeatability of the arrivals. To validate the events observed in DAS data, we used the data recorded at the Jasper Ridge Seismic Station (JRSC station) by a broadband seismometer managed by the Berkeley Digital Seismic Network. Data from JRSC are available online. The Jasper Ridge station is located approximately 6.4 km from our DAS array. Because near-surface conditions are different below our array and JRSC and ray paths are different, the waveforms were not directly comparable. However, JRSC data provided a rough indication of the arrival time and relative strength of the signal corresponding to different arrivals (i.e., P-waves, S-waves, and surface waves).

Permanente Quarry Blasts

Biondi et al. (2017) show the repeatable patterns from two quarry blasts. With the earthquake catalog tools presented in the last section, we quickly pulled quarry blast data by setting the geographical region (latitude range and longitude range around the quarry blast) through the catalog interface. We identified a total of 67 quarry blasts recorded by SDASA-1. The maximum distance, by which these events were separated is less than 800 meters, which means they are good candidates for waveform repeatability assessment. Indeed we did see satisfactory repeatability between a number of events. However, some events, especially weak ones, gave us no recognizable patterns. Here we first present our analyses of six representative events showing clear repeatability. The following list provides the main information for the six events in order of increasing magnitude. The magnitude (M) and depth (z) of each event are based on the online USGS database. Distance (Δ) to DAS array of each event is calculated based on the 3D distance between SDASA-1 and the epicenter.

- *Blast#1*: September 23, 2016 – M 1.47 – z=-0.31 km – Δ=13.43 km
- *Blast#2*: September 29, 2016 – M1.53 – z=-0.31 km – Δ=13.17 km
- *Blast#3*: November 28, 2016 – M1.64 – z=-0.22 km – Δ=13.38 km
- *Blast#4*: October 12, 2016 – M1.68 – z=-0.28 km – Δ=13.21 km

- *Blast#5*: May 23, 2017 – M1.73 – z=-0.31 km – Δ=13.46 km
- *Blast#6*: February 09, 2017 – M1.87 – z=-0.26 km – Δ=13.51km

To give the reader a sense of these SDASA-1 event records, the top panels of Figure 3 and Figure 4 show the data for *Blast#1* and *Blast#2* respectively, after bandpassing from 0.23 to 2.0 Hz. The origin of the time axis (0 seconds) is the event start time provided by the USGS online database. We use the same convention for the following data displays. The trace at the bottom shows the data recorded by the vertical component of broadband data recorded by JRSC (JRSC-BH data). This trace was bandpassed with the same filter parameters as the DAS data. The two events have similar surface-wave kinematics. We observe that these two events have a repeatable surface-wave arrival at approximately 9 seconds. Furthermore, compared with JRSC-BH data, for these quarry blast events, SDASA-1 records more reliable data.

To make the interpretation more straightforward, for each of the events, we stacked the envelopes of the traces after normalization by maximum absolute value. Figure 5 shows the stacking results of these blasts on the same plot, with each color line corresponding to each blast. Clearly, their waveforms have much in common. All the waveforms have amplitude jumps at around 8 seconds, which we interpret as S-wave arrivals. They all have amplitude peaks at around 12 seconds, which can be explained as arrivals of surface waves with strong energy.

The following events (*Blast#7* to *Blast#10*), however, represent events with no evident regular patterns, as stacking results (traces stacked in the same way as *Blast#1* to*Blast#6*) shown in Figure 6. The reason may be that these events were over-whelmed by locally-generated noise.

- *Blast#7*: April 12, 2017 – M 1.52 – z=-0.31 km – Δ=13.49 km
- *Blast#8*: September 13, 2016 – M1.58 – z=-0.23 km – Δ=13.47 km
- *Blast#9*: May 04, 2017 – M1.65 – z=-0.31 km – Δ=13.49 km
- *Blast#10*: December 13, 2016 – M1.66 – z=-0.28 km – Δ=13.56 km

Besides artificial events, we pulled natural earthquake recordings using our catalog interfaces. We observed satisfactorily repeatable signals for nearby events in terms of the arrival times of P-waves, S-waves, and surface-waves. We show the analysis for two earthquakes near Ladera, CA and two earthquakes at Felt Lake near the Stanford campus in the next two subsections.

Ladera earthquakes

The main information for the two Ladera events is listed in the following:

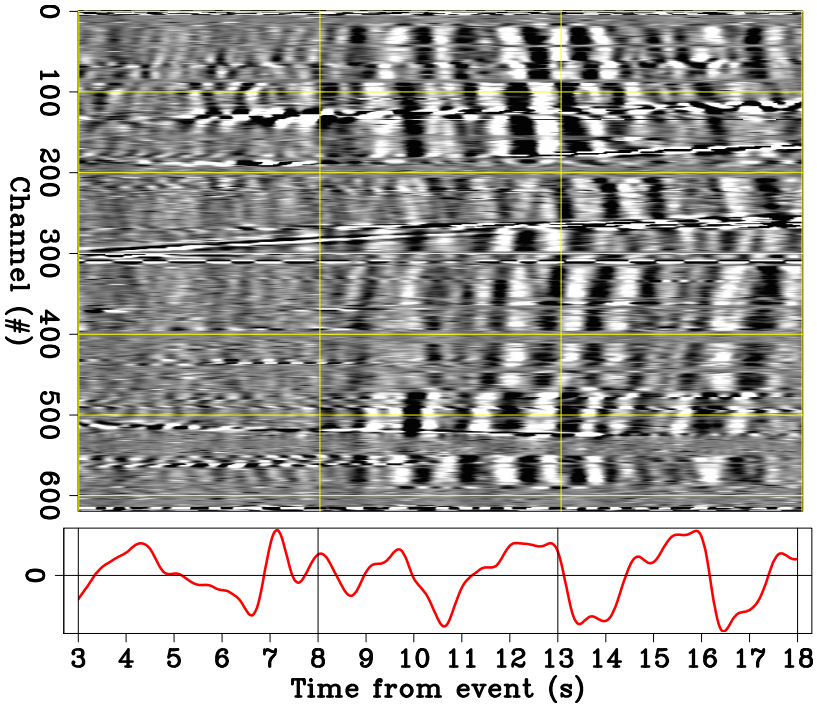

Figure 3: *Blast#1* Data bandpassing (0.23-2.0 Hz). Top: DAS array data; Bottom: JRSC-BH vertical component. The time origin (0 seconds) is the event time according to the USGS and channel numbers correspond to the markings on Figure 1. [**CR**]

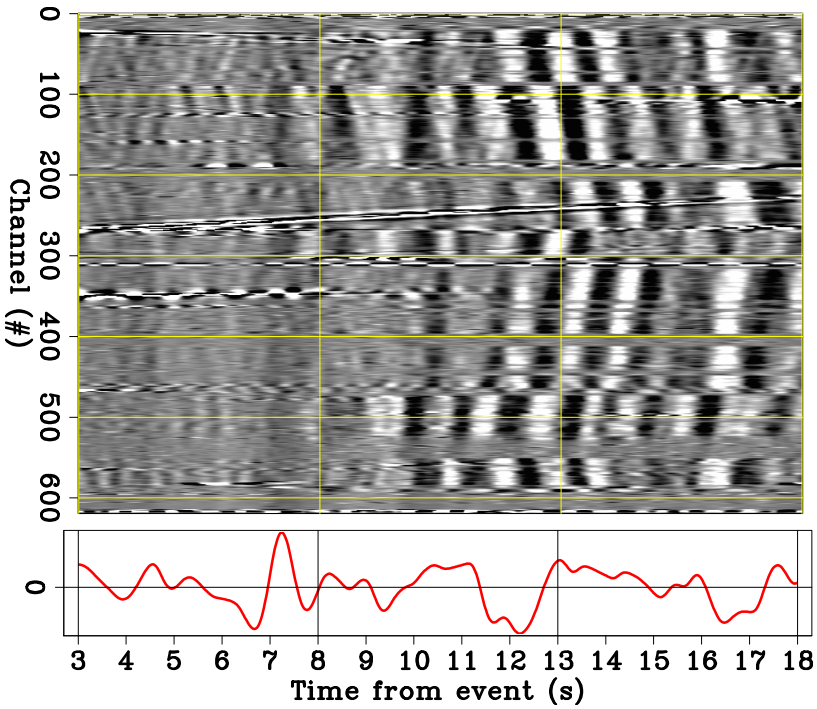

Figure 4: *Blast#2* Data after bandpassing (0.23-2.0 Hz). Top: SDASA-1 array data; Bottom: JRSC-BH vertical component. The time origin and channel numbers follow the same convention as Figure 3 [**CR**]

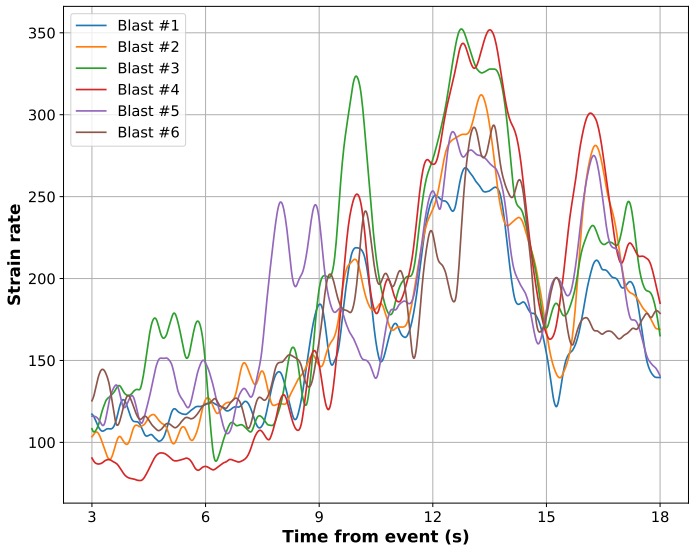

Figure 5: Stacks of the envelopes of the normalized traces for *Blast#1* to *Blast#6* after bandpassing (0.23-2 Hz). The time-axis origin (0 seconds) is the time of the event according to the USGS online database. [**CR**]

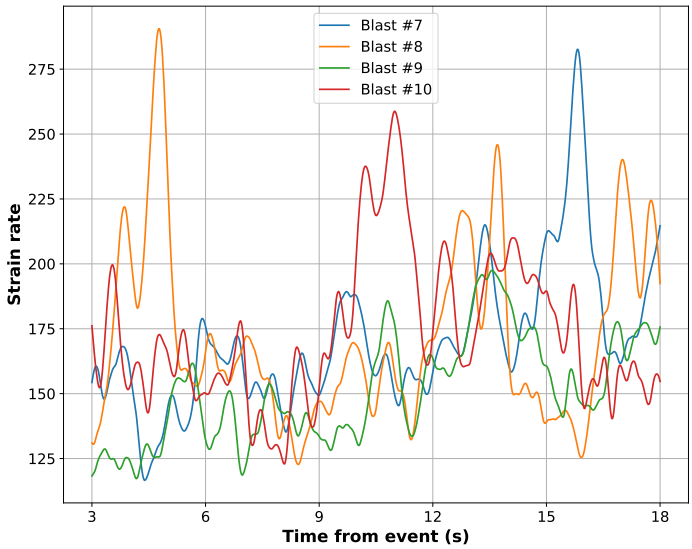

Figure 6: Stacks of the envelopes of the normalized traces for less clear events, *Blast#7* to *Blast#10* blasts after bandpassing (0.23-2.0 Hz). The time-axis origin (0 seconds) is the time of the event according to the USGS online database. [**CR**]

- *Lad#1*: August 10, 2017 – M 1.63 – z=+3.63 km – Δ=4.62 km
- *Lad#2*: August 10, 2017 – M 1.79 – z=+3.25 km – Δ=4.62 km

These two events are very close to each other both in time (26 minutes apart) and in location (3D distance 456-meters apart). In addition, they were very close to Stanford, and both happened at night (local time), which means there should be less anthropogenic noise near the DAS array. The top panel of Figure 7 and Figure 8 show the data bandpassed between 0.25 Hz and 20 Hz for the two events, respectively. The bottom panels show the vertical component of JRSC-BH for comparison. We did see great repeatabilities for arrival timings. Although there is a little time shift, both events have the P-wave arriving at approximately 1.5 seconds, and mix of S-wave and surface-wave arrivals at approximately 3.0 seconds.

We scaled both events in terms of the maximum value for each trace. Then we trace-wise cross-correlated the scaled data of these two events from 1.4 seconds to 5 seconds after their USGS-picked start times, and stacked the results, shown in Figure 9. The peak value is at 0.078 seconds. We advanced the scaled data of *Lad#1* by 0.078 seconds, as shown in Figure 10. Figure 11 shows the scaled data of *Lad#2*. The difference between the two events is shown in Figure 12. We can see that significant cancelations between these events' waveforms happen after P wave arrives (around 1.5 seconds) showing great repeatability. Figure 13 shows the Fast Fourier Transform (FFT) spectra of these two events' scaled data and of their difference. The Nyquist rate of SDASA-1 recording is 25 Hz. Thus, we cannot see higher frequency content without risk of aliasing. The spectra of the two events are consistent with each other, especially for frequencies below 15 Hz. Their difference's spectrum shifts towards high frequency content, which shows that lower frequencies are more repeatable.

Additionally, we estimated the energy for each event by summing up the squared amplitude of all traces ranging from 1.48 seconds to 4.5 seconds. We obtained an estimated energy ratio (*Lad#2* to *Lad#1*) of 1.15. For comparison, we calculated the energy ratio by $10^{1.79}/10^{1.63}$, where 1.79 is the magnitude of *Lad#2* and 1.63 is that of *Lad#1*. The result is 1.45. The difference between the two numbers (1.15 and 1.45) could be explained as a combination of higher levels of noise in our recordings and the fact that we do not record complete three component data.

Felt Lake earthquakes

Repeatability is not just observable at night. Here we show the recording of an earthquake doublet occurring on Stanford's campus near Felt Lake, 4.2 km from the DAS. The distance between their epicenters was estimated by USGS to be close to 100 m, and they occurred one minute apart just prior to 13:00, local time. Figure 14 shows the data recorded by our DAS array and the Jasper Ridge Seismic Station broadband seismometer corresponding to the first, and stronger (magnitude 1.34) of

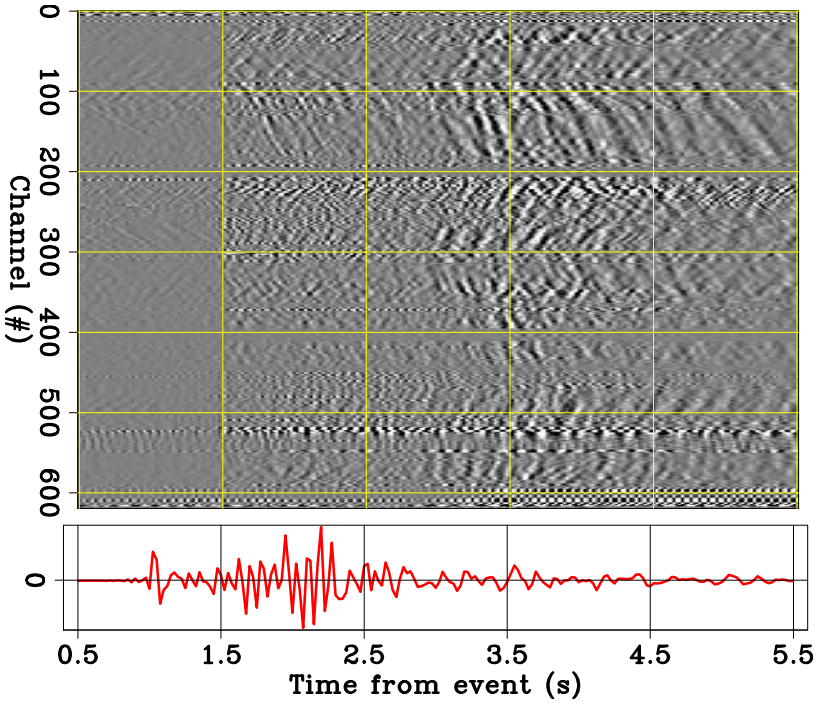

Figure 7: *Lad#1* Data bandpassing (0.25-20 Hz). Top: DAS array data; Bottom: JRSC-BH vertical component. The time origin (0 seconds) is the event time according to USGS. [**CR**]

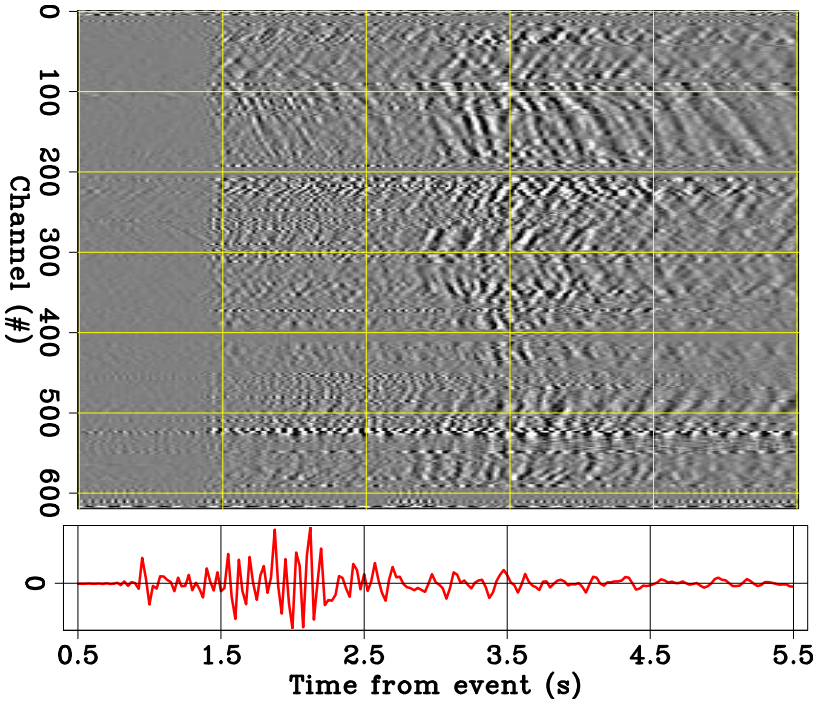

Figure 8: *Lad#2* Data bandpassing (0.25-20 Hz). Top: DAS array data; Bottom: JRSC-BH vertical component. The time origin (0 seconds) is the event time according to USGS. [**CR**]

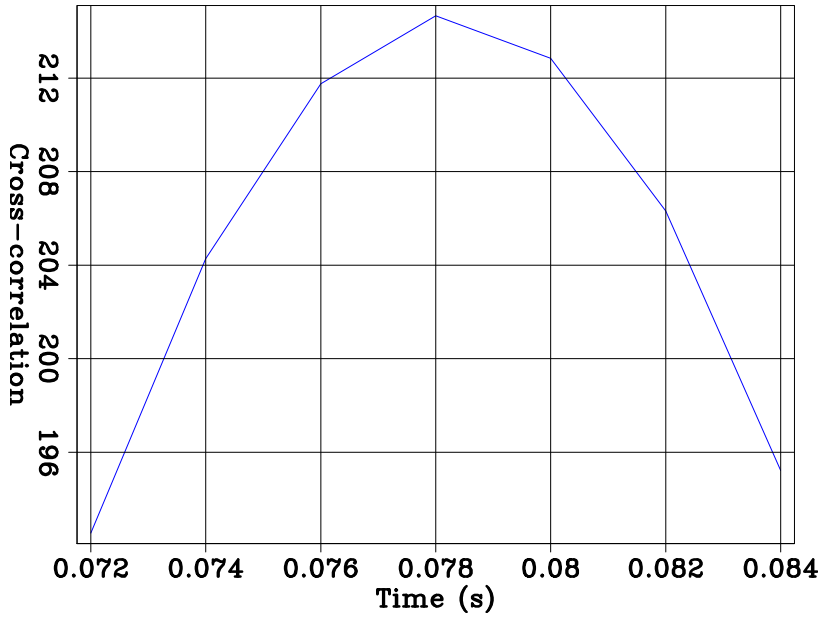

Figure 9: Stack of the tracewise cross-correlations between the bandpassing (0.25-20 Hz) data of *Lad#1* and *Lad#2*. [**CR**]

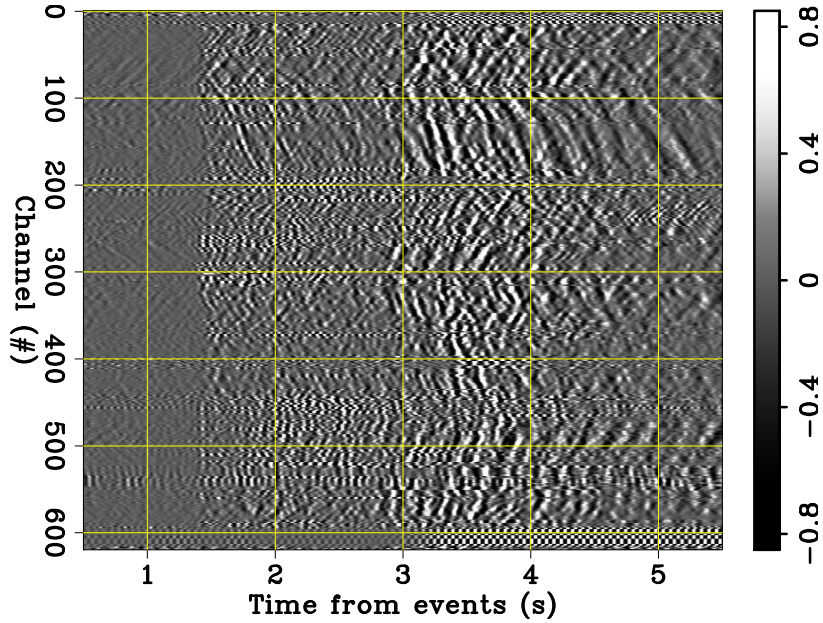

Figure 10: *Lad#1* data after bandpassing (0.25-20 Hz), scaling each trace by its maximum value, and shifting forward by 0.078 seconds. The time origin (0 seconds) is 0.078 seconds from *Lad#1* start time based on USGS. [**CR**]

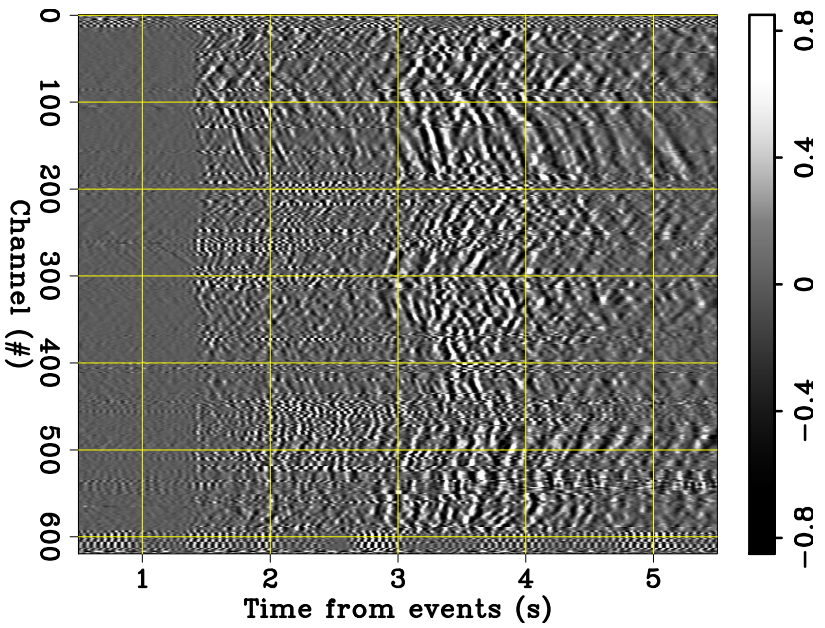

Figure 11: *Lad#2* data after bandpassing (0.25-20 Hz), and scaling each trace by its maximum value. The time origin and channel numbers follow the same convention as Figure 3. [**CR**]

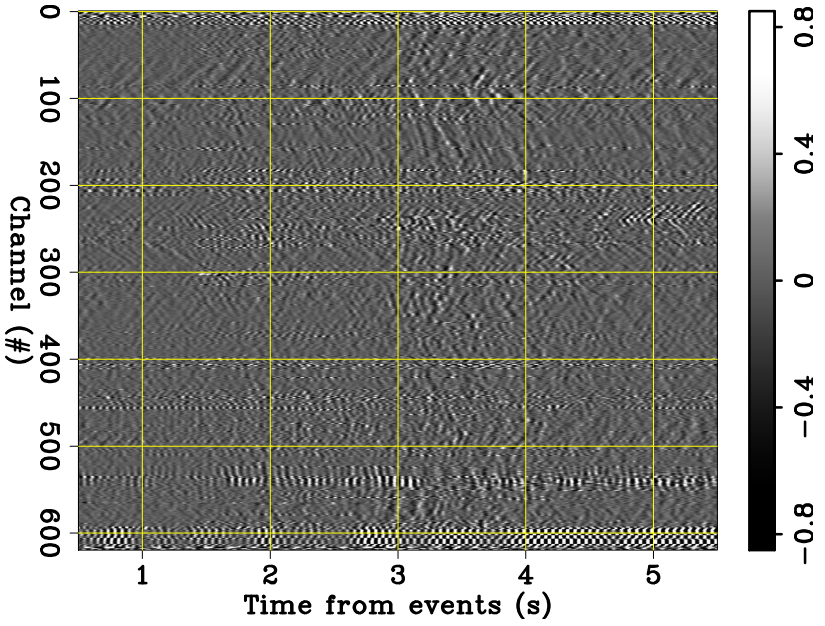

Figure 12: Difference between the trace-wise scaled data of *Lad#1* (shifted forward by 0.078 seconds) and *Lad#2*. The time-axis origin is the time of the event according to USGS's online database. [**CR**]

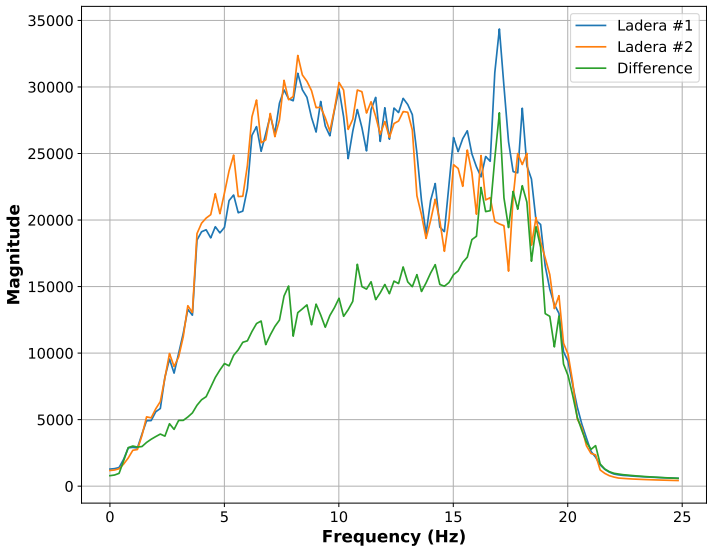

Figure 13: FFT spectra of *Lad#1*, *Lad#2*, and their difference. [**CR**]

the Felt Lake events. The DAS data were preprocessed via laser-noise attenuation, trace balancing, and bandpassing from 0.25 to 12.0 Hz. Figure 15 shows the recording corresponding to the second, and weakest (magnitude 0.95) of the Felt Lake events. The same preprocessing was applied as for the data showed in the previous figures.

- *Felt#1*: July 12, 2017 – M 1.34 – z=+3.24 km – Δ=5.45 km
- *Felt#2*: July 12, 2017 – M 0.95 – z=+3.05 km – Δ=5.34 km

The DAS data shown in Figure 14 and Figure 15 show strong and repeatable waveforms across the whole array starting at about 3.7 seconds. These are likely to be a mix of S-wave and surface-wave arrivals. The waveforms are complex because of the complexity of the near surface both close to the epicenter and in the vicinity of the DAS array. However, they stand out from the strong background noise. As expected, the signal-to-noise ratio is higher for the first stronger event than the second weaker event. There is no clear P-wave arrival visible in either of the two DAS recordings. The P-wave arrived at Jasper Ridge after approximately 1.4 seconds, as is clearly observable from the vertical-component trace shown in Figure 14. The corresponding vertical-component trace in Figure 15 shows a weaker, but still identifiable, P-wave arrival.

The Felt Lake examples demonstrate the repeatability of signals recorded by the DAS array, but they do not show clearly identifiable P and S phases, due in large part to strong anthropogenic noise generated on campus.

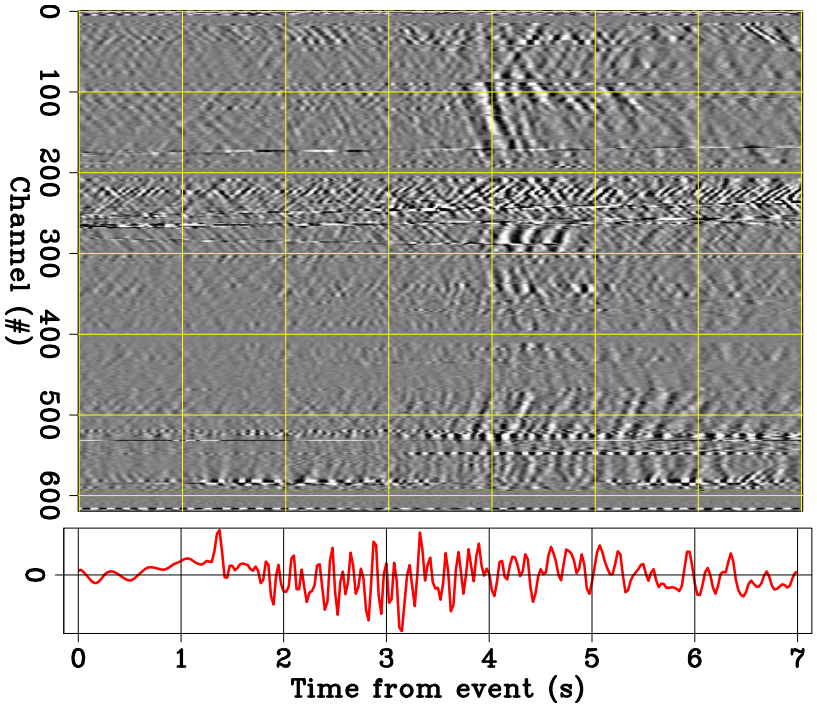

Figure 14: *Felt#1* Data bandpassing (0.23-12.0 Hz). Top: DAS array data; Bottom: JRSC-BH vertical component. The time origin (0 seconds) is the event time according to the USGS. [**CR**]

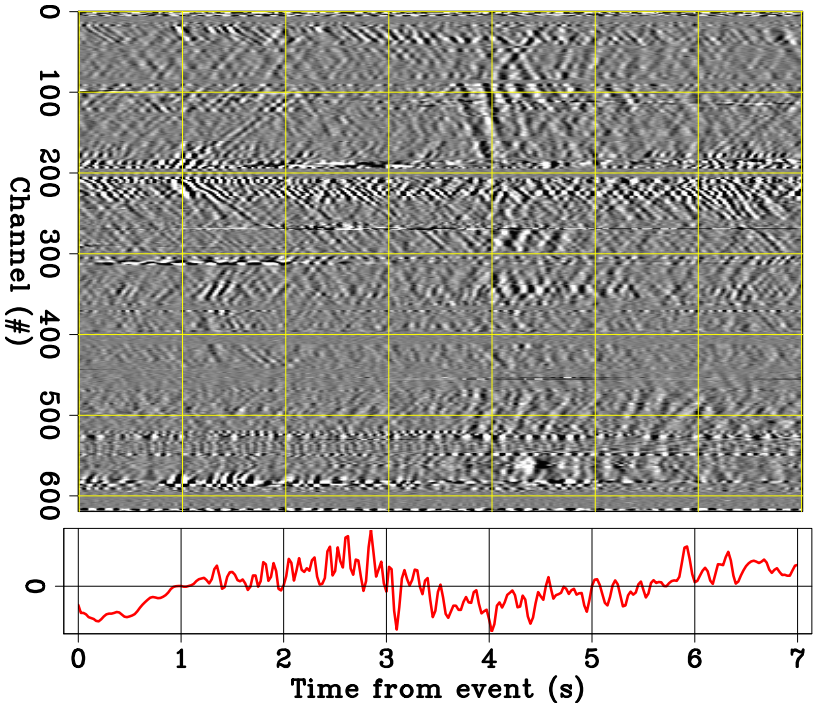

Figure 15: *Felt#2* Data bandpassing (0.23-12.0 Hz). Top: DAS array data; Bottom: JRSC-BH vertical component. The time origin (0 seconds) is the event time according to the USGS. [**CR**]

CONCLUSION AND FUTURE WORK

We have created a continuously updated event catalog for the passive data recorded by our DAS array. We have created interfaces that can be used to filter, access, and process the event data efficiently. Taking advantage of the catalog tools, we examined signal repeatability on both man-made events (quarry blasts) and natural events (Ladera and Feltlake). We found strong signal repeatability in terms of surface-wave arrival timings for Permanente Quarry blasts. And we noticed clearly repeatable P-wave, S-wave, and surface-wave arrivals for naturally occurring nearby events from Ladera and Felt Lake. In the future, we will work on estimating what distances and magnitudes are clearly detectable by the DAS array. We will use the catalog to extract data for distant and weak events. On these events, we will test common array methods for earthquake detection and location, including beamforming and Short-Term Averaging/Long-Term Averaging (STA/LTA) analysis in time and frequency. Then, further analysis should be done on detectability with methods tailored to small events (e.g., template matching).

ACKNOWLEDGEMENT

The interrogator unit was loaned to us by OptaSense, Inc. We thank S. Levin and S. Ronen for advice, F. Huot for sharing scripts pulling data from USGS. We thank Stanford IT and SEEES IT for assistance running the array, and Stanford Center for Computational Earth and Environmental Sciences for computing resources. E. Martin has also been supported in part by the DOE CSGF under grant number DE-FG02-97ER25308, and a Schlumberger Innovation Fellowship.

REFERENCES

Biondi, B., E. Martin, S. Cole, and M. Karrenbach, 2017, Earthquakes analysis using data recorded by the Stanford DAS Array: SEP-Report, **168**, 11–26.

Martin, E., B. Biondi, S. Cole, and M. Karrenbach, 2017, Overview of the Stanford DAS Array-1 (SDASA-1): SEP-Report, **168**, 1–10.