PSD Analysis for Mount Meager Seismic Data

Student: Hongyi Su
UCID: 30053908

Supervisor: Prof. Dr. Hersh Gilbert

Submitted to
Department of Geoscience
University of Calgary

Date: January 6, 2019

Abstract

In seismic data processing, observational seismologists often convert seismic data from the time domain to the frequency domain to investigate the amplitude spectrum. In this thesis, we used recorded field data from Mt. Meager, BC, Canada, to perform power spectrum density (PSD) analysis for filtering and denoising purposes. The primary goal is to identify and quantify the source of noise. To do this, we first convert the observed data to the frequency domain using the FFT algorithm. Then divide the frequency band into two components, the high-frequency component (15–100 Hz) and the low-frequency component (1–15 Hz). After that, we calculated the PSD value for each station every 30 minutes. We also developed a MATLAB program to plot power distribution for all seismometers. We have discovered many interesting findings through this analysis. First, the seismometers located near rivers and roads show greater amplitude as compared to the seismometers in quiet areas. Second, traffic is the main source of noise because the seismometers near roads have lower PSD values during the night while traffic is neglectable. Third, the hydrometric data recorded by the nearest water gauge can be related to observed seismic data. We also identified signals that exhibit the Doppler effect, which could be caused by our research helicopter dropping us off and flying away. We also did several other comparisons and analyses to track and distinguish other possible noise contributors, such as glacier melting, animal activities, wind, etc.

Background and data acquisition

The Garibaldi Volcanic Belt, in southwestern British Columbia, is a zone of recent volcanism in the northernmost segment of the Cascade Arc. The volcanic belt also hosts numerous hot springs. Fluids with temperatures greater than 250 °C at Mt. Meager make it the warmest geothermal resource in Canada (GSC, 2019). To better characterize this geothermal resource, we have installed a dense array of close to 60 seismometers across Mt. Meager that operated for several weeks during the summer of 2019. The study area lies between the latitudes of 50.5°N to 50.7°N and the longitudes of -123.65°W to -123.25°W (Figure 1). These instruments are comprised of Inova Hawk nodal systems connected to 10 Hz three-component geophones. Since we want to investigate the structure related to the geothermal fluids and the heat source, the depth range of our targets lies within the shallow crust. Sampling Mt. Meager in its entirety allows us to characterize the crustal structures associated with the area of high geothermal heat and how those vary across the mountain. The observations from this array will be used to identify the distribution of low seismic wavespeeds that can mark the pathways of geothermal fluids as well as magma chambers. Figure 2, 3, and 4 are examples of our recorded seismic data. Figure 2 shows the recorded vertical component (HHZ) data of five seismometers within an approximately 15-second window on July 20th, 2019, and Figure 3 shows the same dataset but within an approximately 10-minute time window. Figure 4 shows the primary results of a detected seismic event among the approximately 650 GB of recorded data. The glimpse of local seismicity offered by this brief seismic deployment provides constraints on the pattern of faults and fractures within Mt. Meager. Determining the distribution of these cracks will allow for constraints on the distribution of fluid circulation through the geothermal system.

Method

Seismic data always contain noise. Therefore, denoising is a necessary step for seismic data processing. Seismic data processing algorithms can often be implemented more easily in the frequency domain than in the time domain. Power Spectral Density (PSD) is defined as the square of the amplitude per frequency interval, in the unit of Watt/Hz. PSD is a measure of the 'power' contained in a signal at a given frequency (Karchewski, 2019). We use Equation [1] to calculate the PSD value (Peters, 2012)

$$PSD = \frac{|FFT(s)|^2}{df} \quad [1]$$

Where *s* is the amplitude in the time domain, and *df* is the frequency interval.

FFT refers to Fast Fourier Transform Algorithm. PSD is a measure of the 'power' contained in a signal at a given frequency. If we take the integral of a range of frequencies, then we get the 'energy' that is contained in this frequency range. For simplification, we only consider two frequency bands. We assign the low-frequency range contains frequencies from 1 Hz to 15 Hz, and the high-frequency range contains frequencies from 15 Hz to 100 Hz. We calculated the PSD value for every 30 minutes of data for plotting. The Inova Hawk system has a sampling frequency of 250 Hz, which means it records 250 discrete data points per second. A 30-minute-long dataset would have 450,000 data points in total. Equations [2] and [3] show how theoretically we calculated the PSD values that we used for plotting in discrete Fourier transform form,

$$PSD_{1-15} = \int_{1}^{15} \left| \sum_{k=0}^{450000-1} s_k e^{-i \cdot 2\pi f \cdot k \cdot 0.004} \right|^2 df \quad [2]$$

Where PSD_{1-15} represent the 'power' value for our low-frequency range (1-15Hz) contained in the signal and the unit is in Watt.

$$PSD_{15-100} = \int_{15}^{100} \left| \sum_{k=0}^{450000-1} s_k e^{-i \cdot 2\pi f \cdot k \cdot 0.004} \right|^2 df \quad [3]$$

Where PSD_{15-100} represent the 'power' value for our high-frequency range (15-100Hz) contained in the signal and the unit is in Watt.

We use the Noise Toolkit from Incorporated Research Institutions for Seismology (IRIS) to compute the PSD value. It is composed of three open-source Python script bundles. The script used to compute the PSD value is called ntk_computePSD.py (IRIS DMC, 2014). From the script ntk_computePSD.py, we used the following equation is used to convert FFT to PSD.

$$PSD = N \times \frac{FFT(s)^2}{f_N}$$

[4]

Where N is equally spaced frequency samples and f_N is the Nyquist frequency, which can be calculated by Equation [5],

$$f_N = \frac{1}{2\times\Delta t} \quad [5]$$

This processing procedure involves three basic steps. The first one is pre-processing; in this step, we read waveform data in MATLAB, then remove the instrument response, detrending, demeaning, and tapering; next, we compute the PSD value using the Fast Fourier Transformation (FFT) algorithm and generate a PSD log for every 60 minutes of data; For the last part, we extract PSDs and Power Density Functions (PDFs) from the PSD log. In the second stage, we take a logarithm of the PSD value for each station. Figure 5 and Figure 6 show the plot for the logarithm of the PSD value for the low-frequency range (1–15 Hz) and the high-frequency range (15–100 Hz) respectively. Figure 7 and Figure 8 show the zoomed plots for the low (1–15 Hz) and high (15–100 Hz) frequency ranges. Finally, we plotted each station based on its PSD value for every 30 minutes on a 256-colour map system, where a value of 1 represents the darkest colour (black) and a value of 256 represents the brightest colour (white). The core of the colour mapping is to relate the PSD value proportionally to the integer value from 1 to 256. We first make the 'amplitude' of the colour value all positive by simply subtracting the minimum value, then normalize this value by dividing it by the maximum value and multiplying it by 256.

$$color\_val = \frac{log(PSD) - min\,(log\,(PSD))}{max\,(log\,(PSD))} \times 256 \quad [6]$$

Where color_val represents the calculated colour value of each seismometer for mapping. Since there is a possibility that color_val could equal 0 through our calculation method, and it will break the program, so we assigned the color_val equal to 1 if it was 0. We can also control the colour contrast by changing the lower bound value and upper bound value. The larger the gap between the lower bound and the upper bound, the less contrast in colour will be. Figure 9 shows one of the maps that we have made, the time is at 01:30:00 on July 13th, 2019 (UTC), where the lower bound is -24.5, and the upper bound is -10. The brighter colour means a higher PSD value. Figure 10 shows the same date map with the same PSD data, the only difference is we chose the lower bound to be -28 instead of -24.5 and the upper bound to be -5 instead of -10. As we can see in Figure 10, using this wider range in colour values causes the plot to have less contrast due to the increased gap between the lower and the upper bound. We use the narrower range to emphasize the difference between high and low power stations.

Discussion and Result

Our goal is to find noise sources. And because almost half of our seismometers were installed nearby rivers and roads, we naturally suspect that traffic on the roads and fluid flow in the rivers are the main noise contributors. Diurnal variations in temperatures would be expected to cause changes to the amount of meltwater flowing from glaciers and patterns of wind around Mt. Meager. Another source of noise could come from wildlife activities in the research field as we have 'luckily' encountered a couple of Grizzly bears at the station sites where we installed them.

To test the traffic and river flow hypotheses. We separated the seismometers into two groups. One group is for the stations located near the rivers and roads, and the other group is for the stations far away from the rivers and roads. We selected 10 seismometers for each group. Table 1 listed the station names for each group. The stations shown in bold correspond to those presented in Figure 11 and Figure 12.

Near river and road	Far away from river and road
MM42	MM03
MM43	MM06
MM07	MM28
MM17	**MM18**
MM25	MM11
MM31	**MM10**
MM22	MM41
MM29	MM24
MM00	MM15
MM23	MM16

Table 1, divides stations into two categories, ones are near rivers and roads, and the other ones are far away from rivers and roads

We found that the stations near the rivers and roads have higher PSD values and the ones further away have in general lower PSD values. Figure 11 and Figure 12 show the PSD plots for stations MM10 and MM18 (both stations are far from the road), the former showing the frequency range of 1–15 Hz and the latter showing the frequency range of 15–100 Hz. Differently, Figure 13 and Figure 14 show the PSD plots for stations MM42 and MM43 (both stations are near the road). We noticed that the two stations that are far away from the roads have lower 'Power' than the two near the roads. Moreover, the magnitude of energy level for 15–100 Hz is always higher than 1–15 Hz, and this is probably because the 15–100 Hz range includes more individual frequency components than the 1–15 Hz range. Therefore, the first conclusion is the river and road indeed cause a lot of noise.

The road and river are always together, in the next step, we want to separate these two factors. We assume the traffic on the road will reduce dramatically during the night but the fluid flow would not stop no matter what time. Therefore, we compared the PSD map of local time at noon (12:00:00) and local time at midnight (00:00:00). If the PSD map shows significantly lower 'power' during the night, then we should conclude the traffic on the road is the dominating source of the noise and vice versa. Figure 15 and Figure 16 are the PSD maps at noon and at midnight, note that we use the UTC time, so the local time needs to subtract by 8 hours. By comparing Figure 15 and Figure 16, we cannot see a big difference. Therefore, we conclude that the river is a major source of noise in our seismic recordings.

To reinforce the conclusion that fluid flows in the river are the main source of the noise. We compare our seismic data with recorded hydrometric data from the nearest water gauge. Figures 17 and 18 show a close relationship between the seismic energy and the rivers' discharge and water level. The three plotted stations (MM03, MM06, and MM28) in Figure 17 are located near the glacier, where we expected almost no human activities (such as walking, driving, etc.). And the two plotted stations in Figure 18 (MM23 and MM27) are near the road, and the PSD plots for those stations appear much spikier compared to the glacier stations. Nevertheless, the seismic energy for these two groups of stations is both related to the discharge or the water level.

During our field campaign, we saw there were some construction workers for a hot spring resort in the area. Therefore, we want to investigate if the construction team contributes to the noise in our data too. We assume the construction workers only work on weekdays. Hence, we compared the PSD map between Monday and Sunday. If the energy level is much higher on Monday, then we could conclude the construction workers are also one source of the noise. Figure 19 and Figure 21 show the maps for July 14th and July 21st (weekends), and Figure 20 and Figure 22 show the maps for July 15th and July 22nd. All these figures show the local time at 10 a.m. when the construction crew is supposed to be at work. We found there was no big difference between weekdays and weekends. Therefore, we refute the hypothesis that the construction team caused some noise in our data.

Last but not the least, aeroplanes and helicopters could also contribute to the noise. From the data in Figure 23, we found a bizarre dispersion pattern with arrival times increasing and then decreasing with decreasing frequencies. This dispersion is not seismic, to get around this, we suspect these patterns are associated with the aeroplanes and helicopters flying around in the area. A study by Meng et al. (2017) also found a similar phenomenon. We also discovered the Doppler Effect in our recorded dataset, which possibly causes by the helicopters transporting us from one station to another. Doppler Effect means the apparent change in the frequency of a wave caused by relative motion between the source of the wave and the observer, in this case, the source of the wave is the helicopter, and the observer is the seismometer. Jet planes would travel at around 800 – 1000 km per hour, we estimated the travelling speed of the object using stations at different locations (MM12, MM14 and MM18), and conclude that the object caused the Doppler Effect in Figure 24 is the helicopter transports us.

Conclusion

In this study, we have calculated PSD values from our field-recorded data for two separate frequency ranges, [1–15] Hz and [15 –100] Hz. We then did a series of analyses in space and time to find the main source of noise in our seismic data. We found that the seismometers far away from rivers and roads have lower PSD magnitudes than the ones that are near and speculated that the river is a major noise contributor to our seismic recordings. We further investigate this hypothesis with observed hydrometric data at the nearest water gauge. The correlation between the seismic energy and the water discharge supported our hypothesis. Moreover, we refute the hypothesis that the construction team in the study area contribute to noise in our data. Furthermore, we also detected certain periods in the waveform data that exhibit the Doppler effect, and based on the travelling speed of the object, we suspect that this is from our helicopter.

By comparing the PSD maps, we got a general sense of how the seismic energy is distributed at each station. A future investigation is needed to make better decisions on filtering parameters. For example, we need to know how fast the river flows to minimize the noise caused by the river.

Figures

Figure 1, Location of seismometers, some stations are far away from the river and the road, such as MM18 and MM10, which are circled in blue. And some stations are near the river and road, such as MM42 and MM43, which are circled in white. Also, there is a river gauge located down the river to the east of the map; we did not include it here because it is outside of our map boundary.

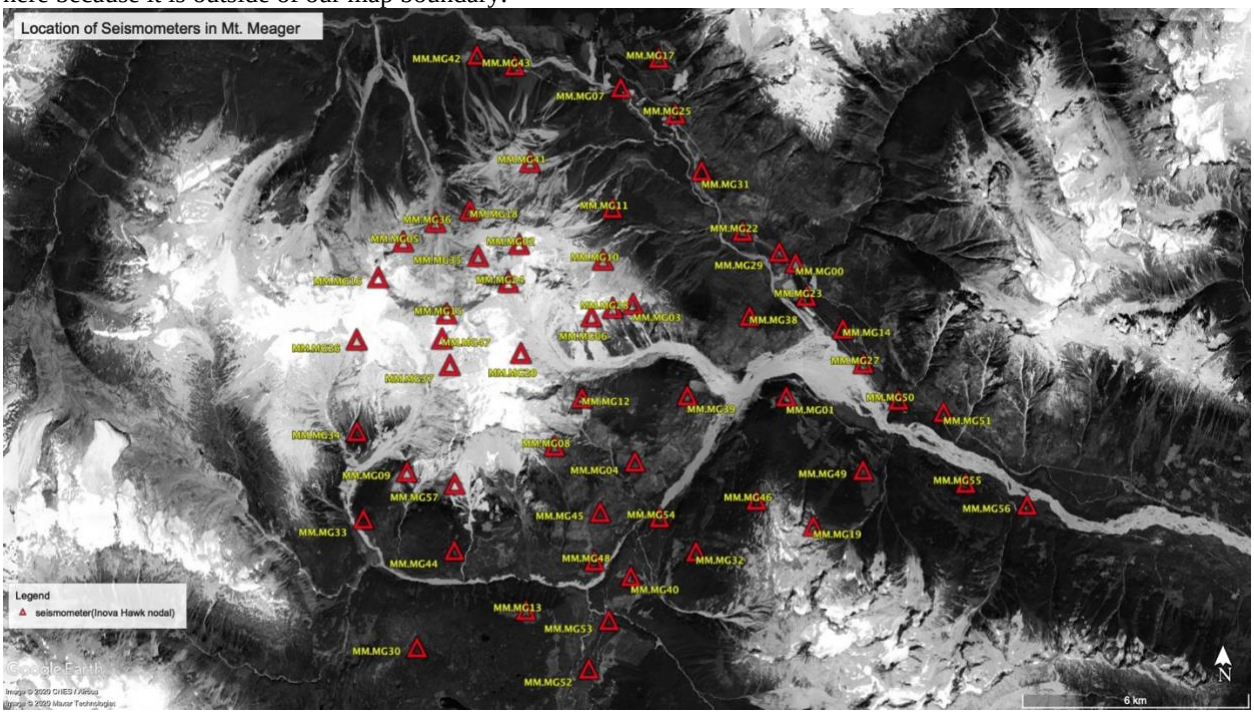


Figure 2, Recorded Mt. Meager data displayed in a window for 15 seconds in Snuffler (Pyrocko), no filter applied.

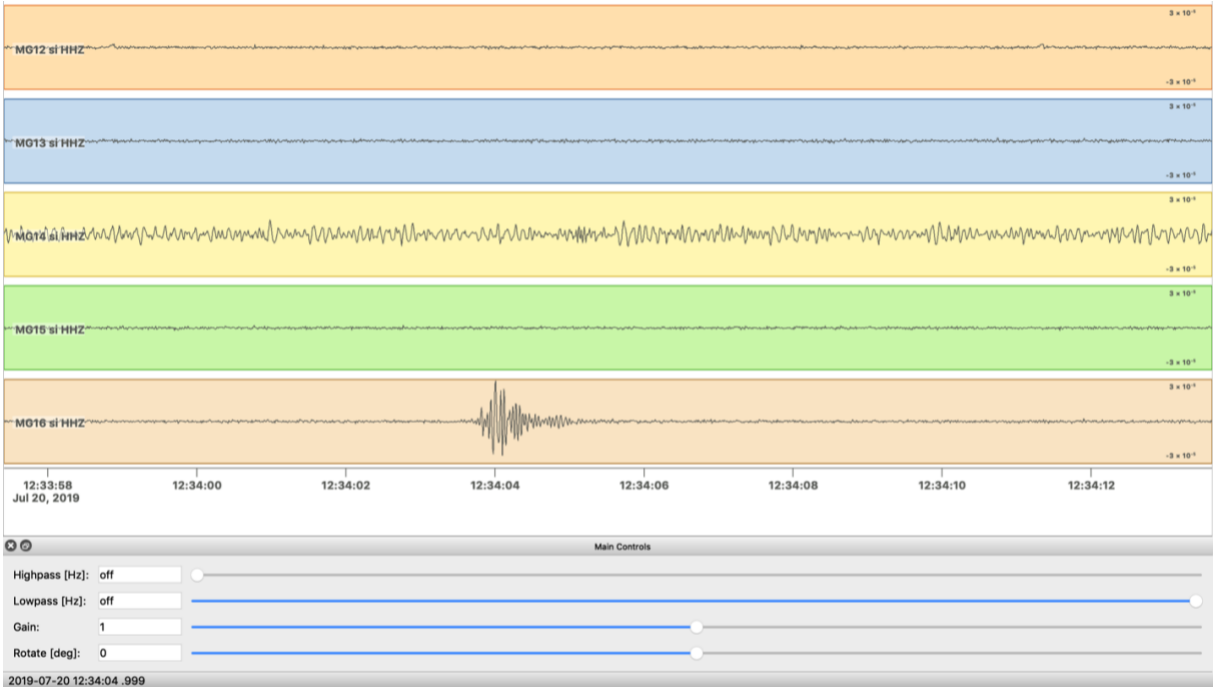


Figure 3, Recorded Mt. Meager data displayed in a window for 10 minutes in Snuffler (Pyrocko), no filter applied.

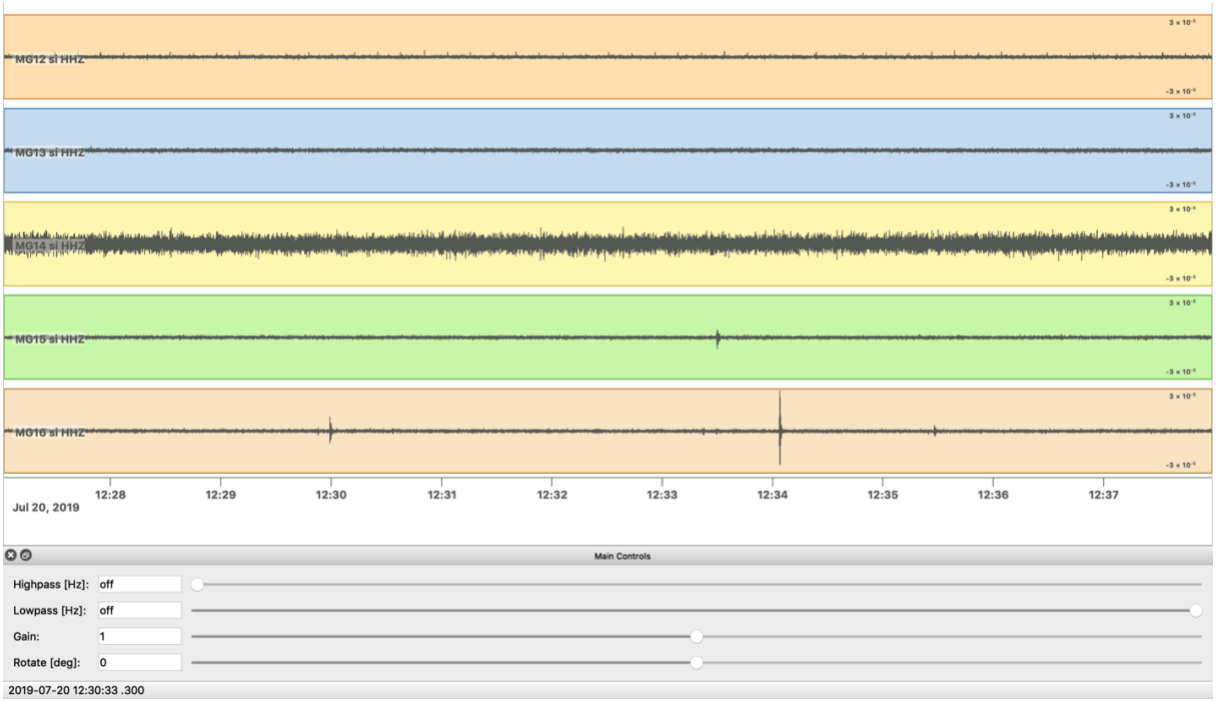


Figure 4, the seismogram shows two recorded events on July 26, 2019, in the Mt. Meager area, BC in SAC. The waves from one event are travelling from the north (top) to the south (bottom). Then, a second event, which follows this first event, displays a signal in which the waves propagate from the south (bottom) to the north (top), and creates this 'V' shape in our recorded seismogram.

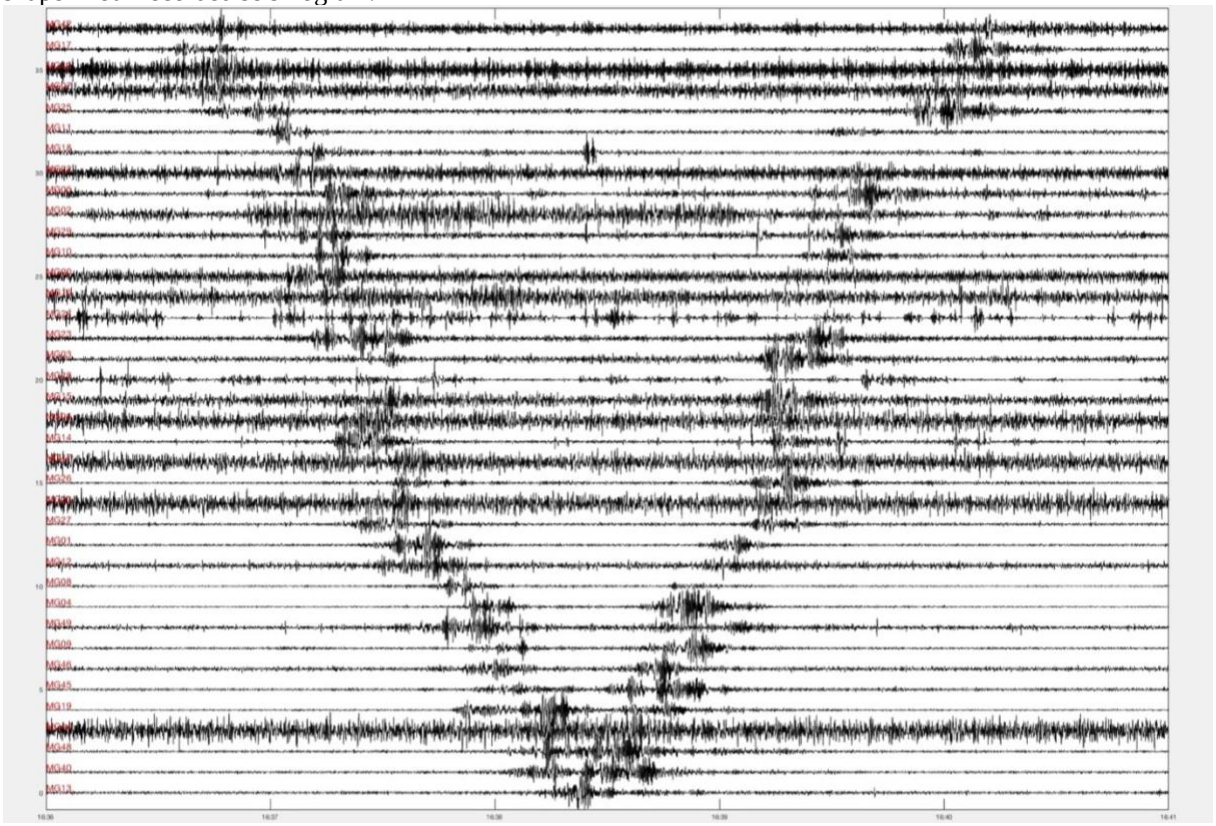


Figure 5, log (PSD) versus time for the low-frequency range (1 − 15 Hz)

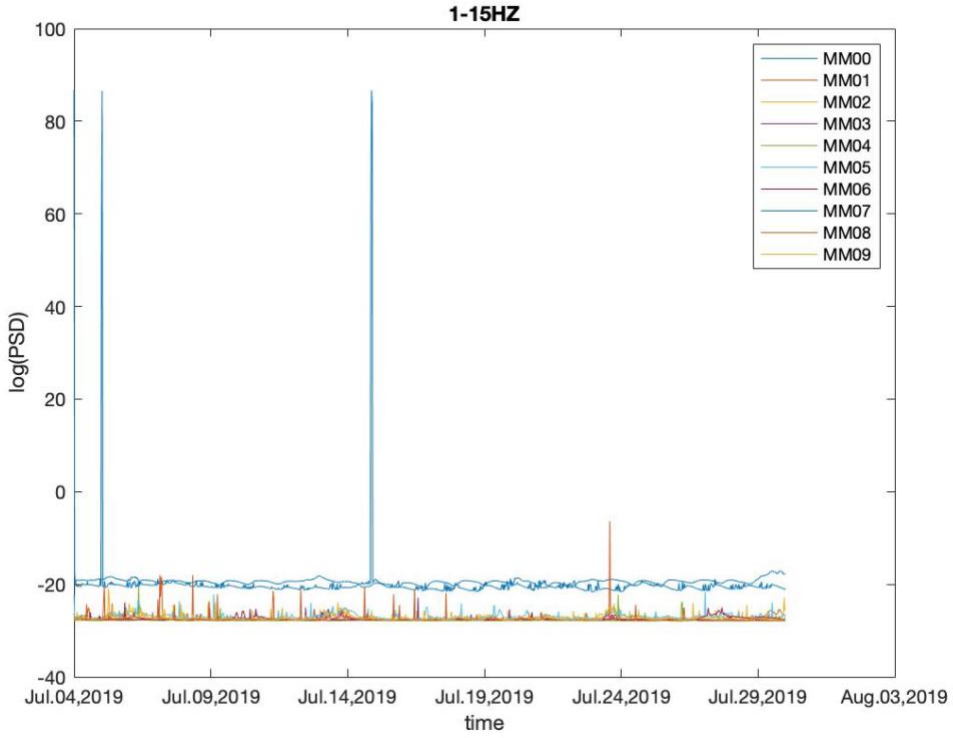


Figure 6, log (PSD) versus time for the high-frequency range (15 − 100 Hz)

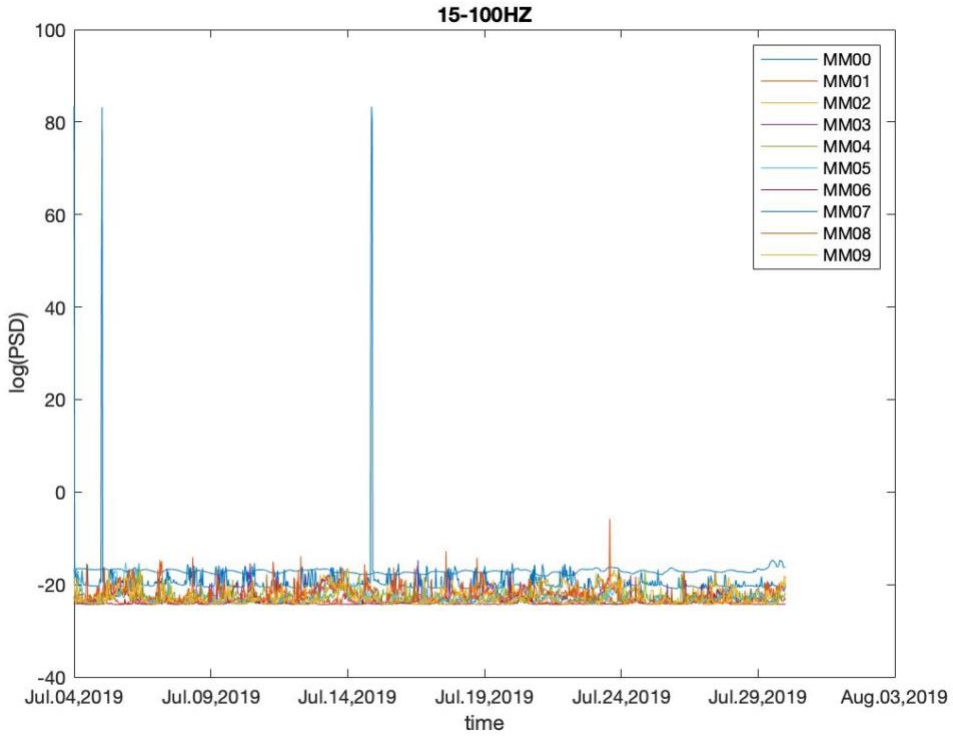


Figure 7, log (PSD) versus time for the low-frequency range (1 – 15 Hz), y-axis limited to [-28, -18]

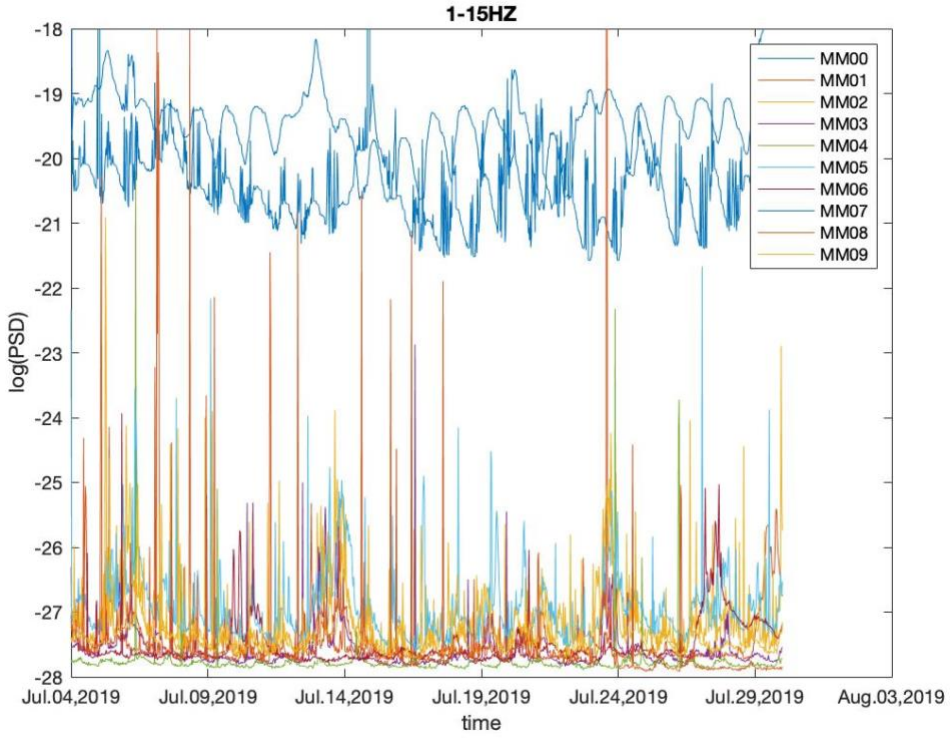


Figure 8, log (PSD) versus time for the high-frequency range (15 – 100 Hz), y-axis limited to [-25, -15]

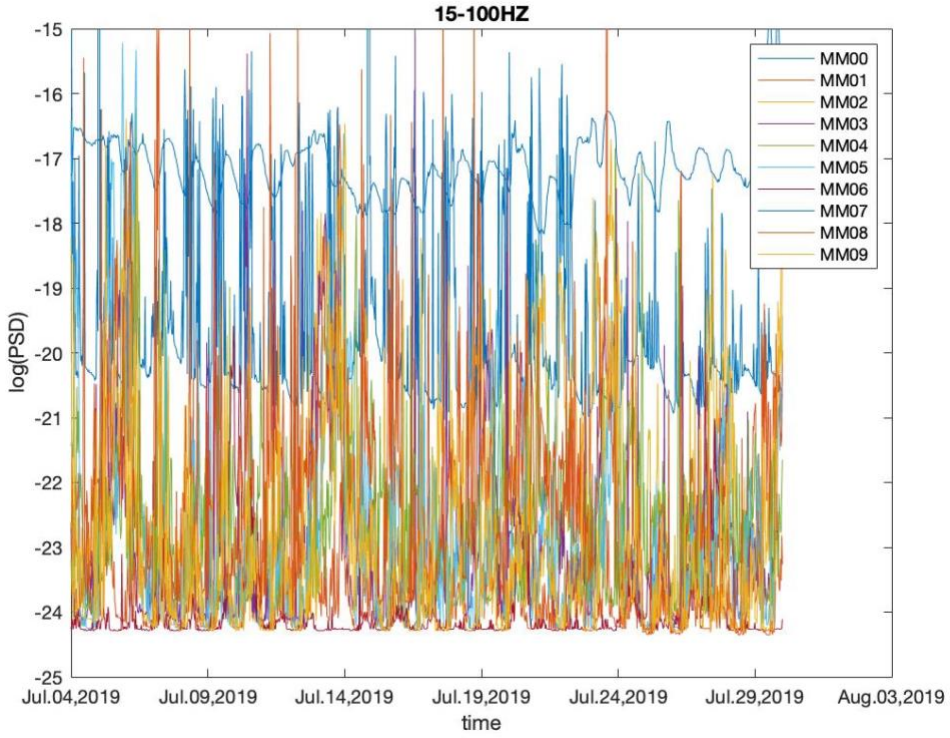


Figure 9, PSD map with an upper bound of -10 and a lower bound of -24.5.

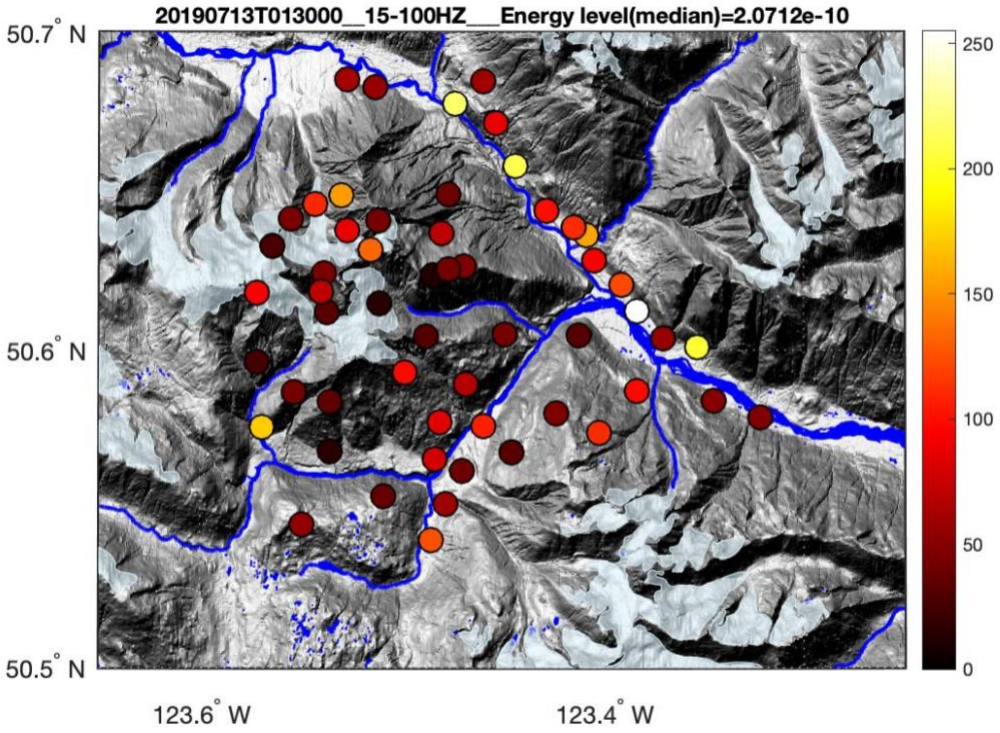


Figure 10, PSD map with an upper bound of -5 and a lower bound of -28.

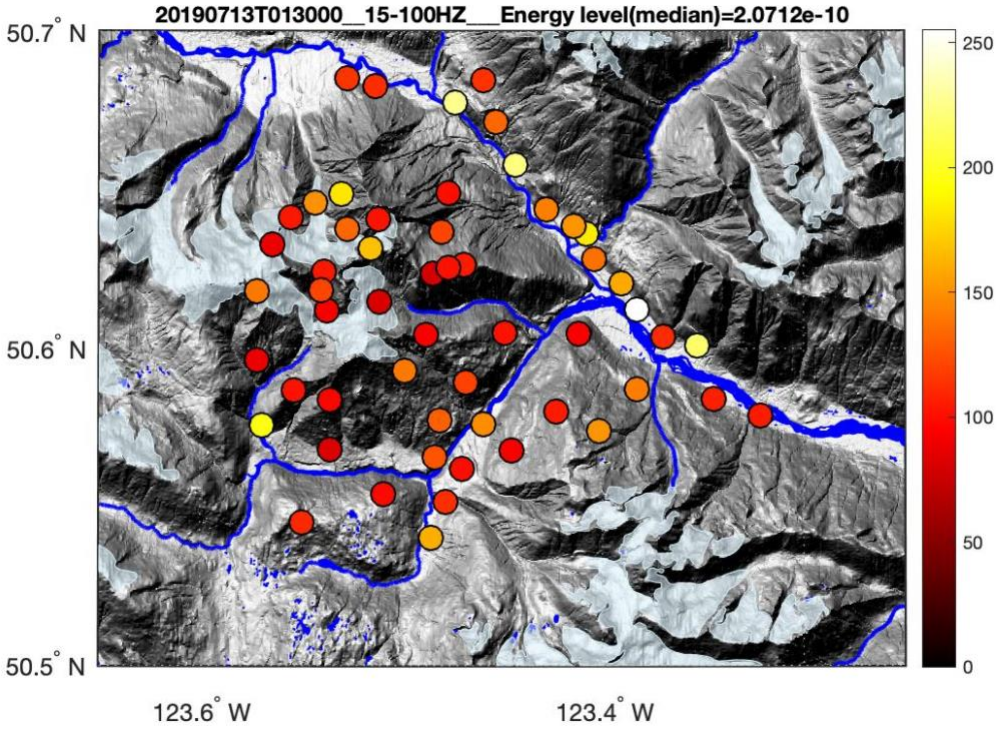


Figure 11, PSD plot (1–15 Hz) for stations MM10 and MM18, which are far from rivers or roads.

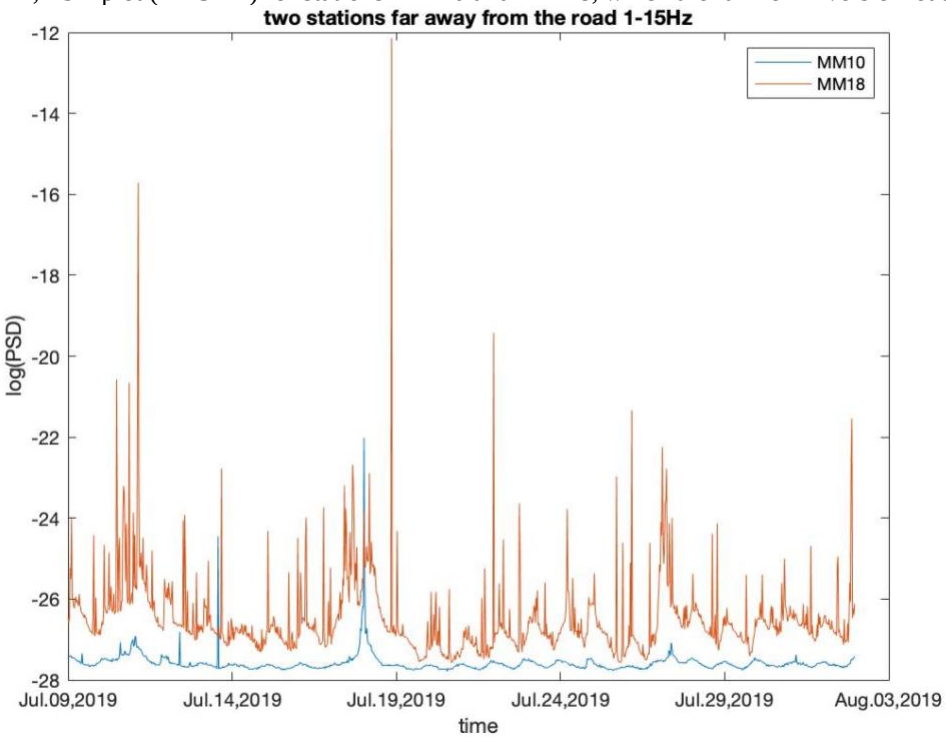


Figure 12, PSD plot (15–100 Hz) for stations MM10 and MM18, which are far from the rivers or roads.

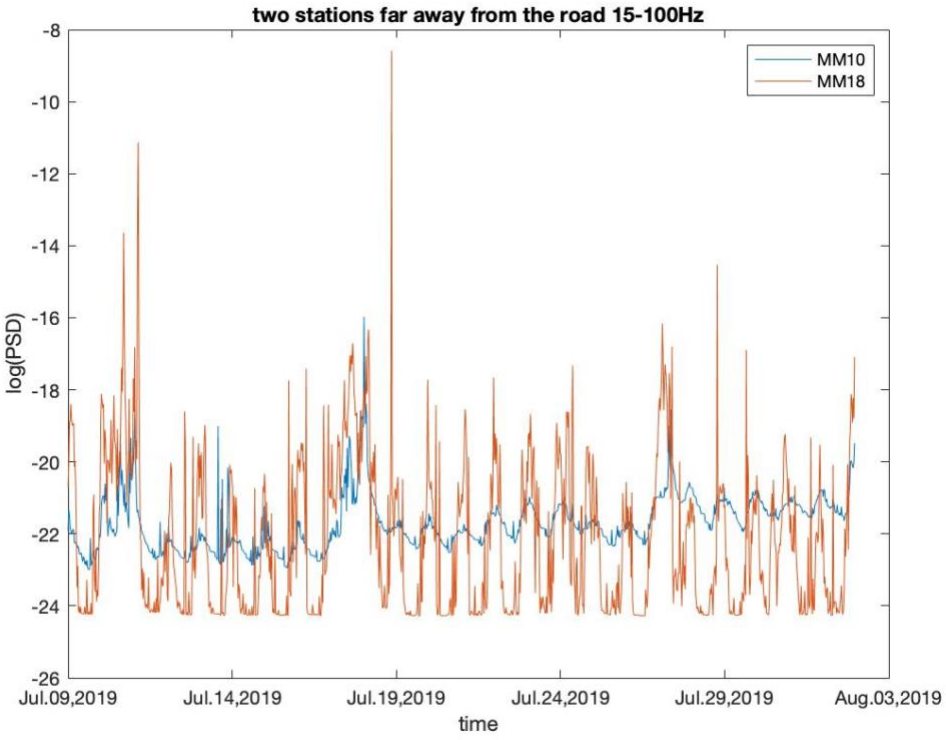


Figure 13, PSD plot (1–15 Hz) for stations MM42 and MM43, which are near rivers and roads.

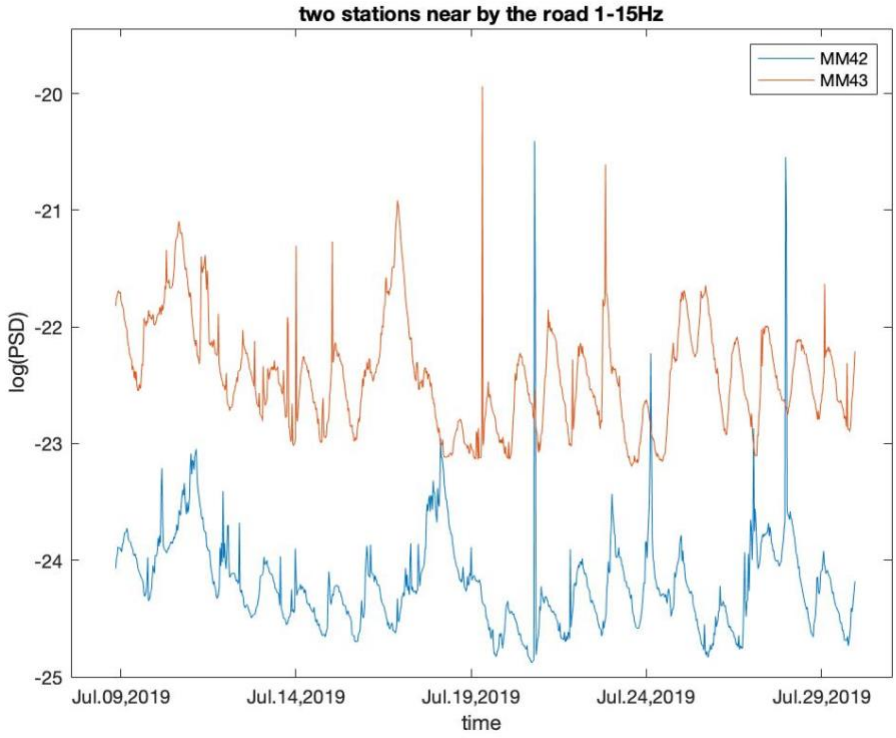


Figure 14, PSD plot (15–100 Hz) for stations MM42 and MM43, which are near rivers and roads.

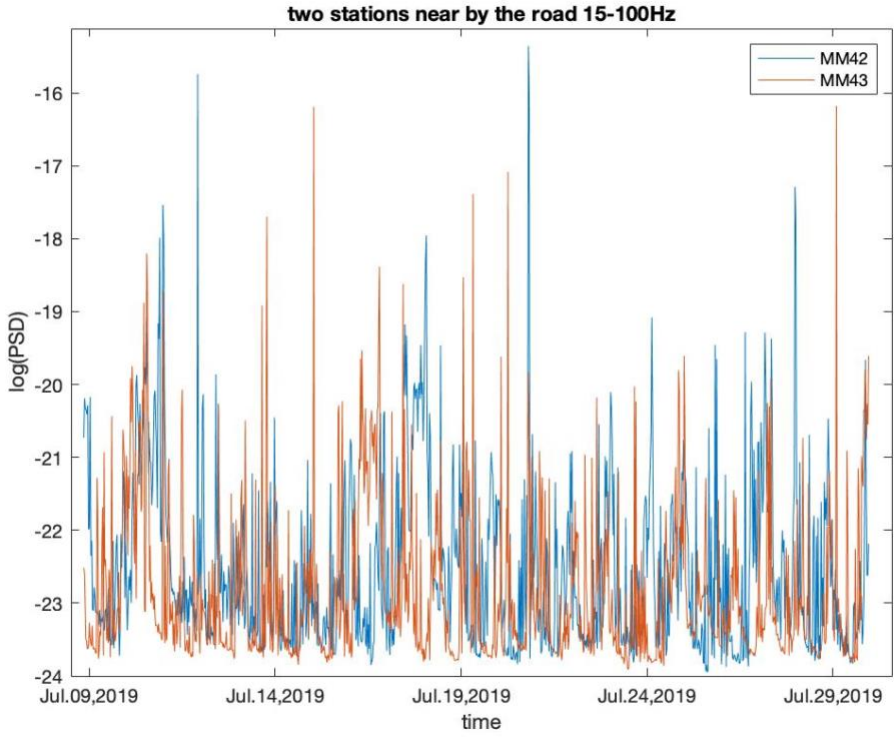


Figure 15. local time at 00:00:00 on 14 July 2019

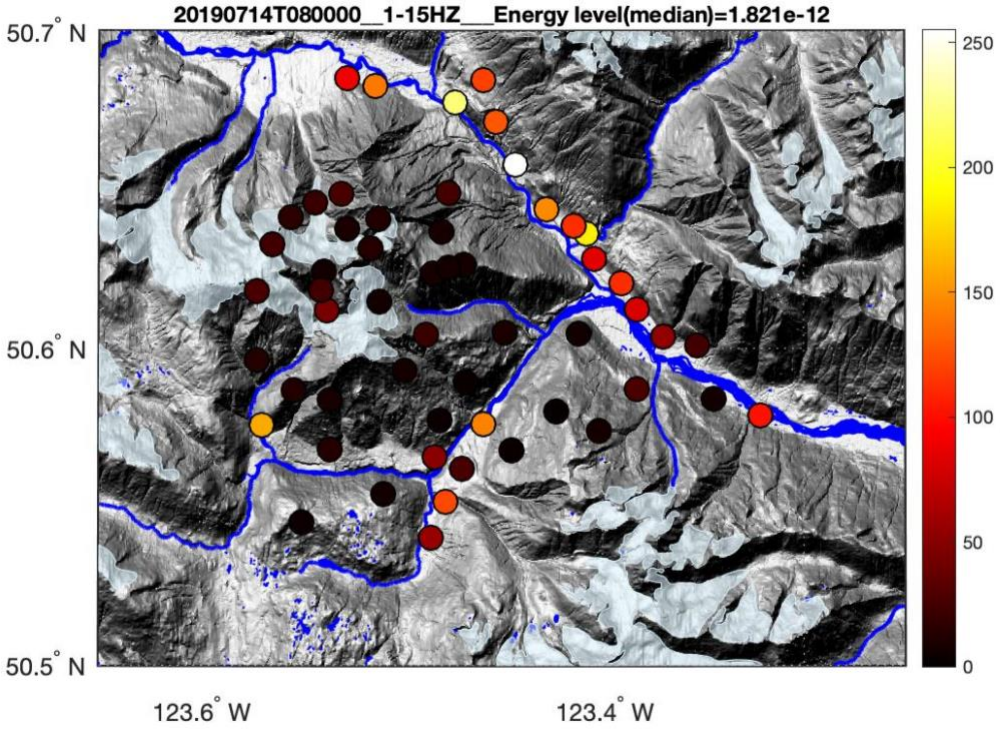


Figure 16. local time at 12:00:00 on 14 July 2019

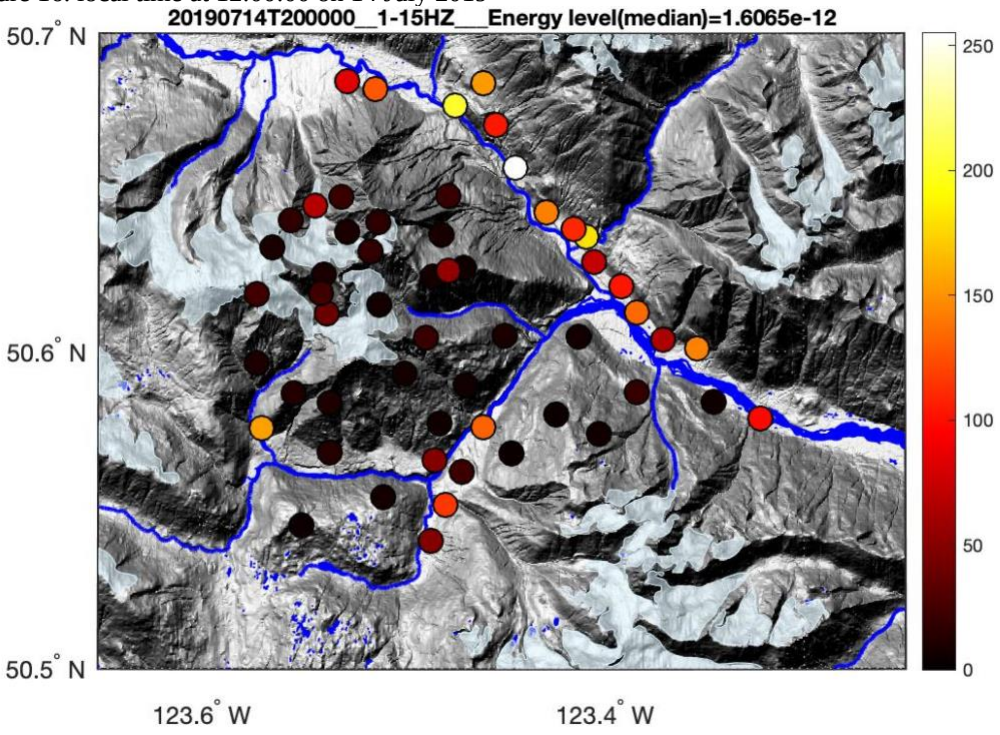


Figure 17, log (PSD) vs. time plots for stations MM06, MM03, MM28 and MM10 (in the glacier area), as well as discharge and water level plots from July 9 to August 3. The hydrometric data is provided by Environment Canada (2020), the water gauge is located by the Lillooet River near Pemberton, and the station number is 08MG005.

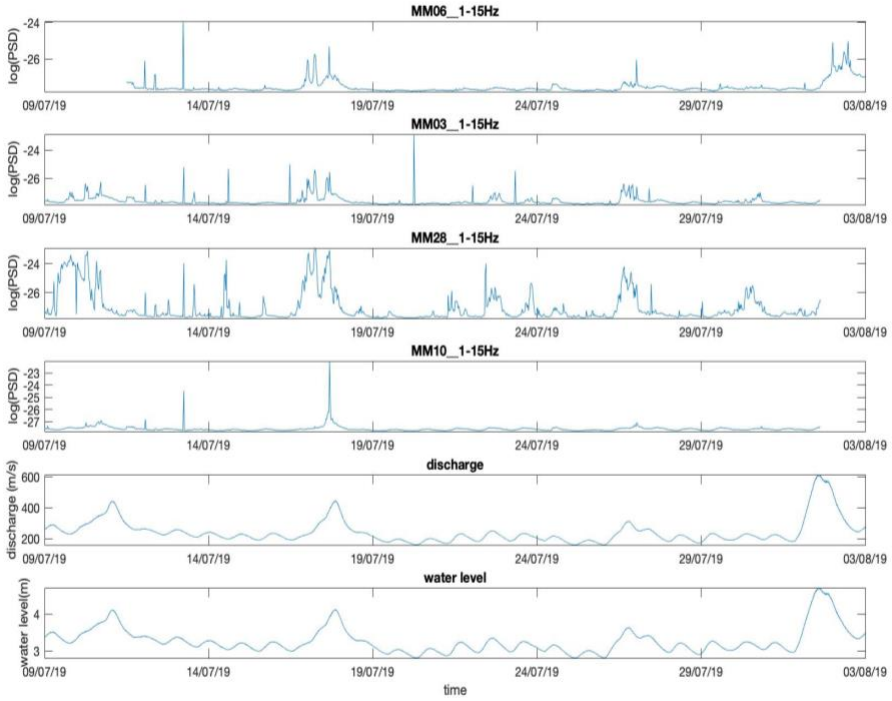


Figure 18, log (PSD) vs. time plots for stations MM23 and MM27 (nearby the road), as well as discharge and water level plots from July 9 to August 3. The hydrometric data is provided by Environment Canada (2020), the water gauge is located by the Lillooet River near Pemberton, and the station number is 08MG005.

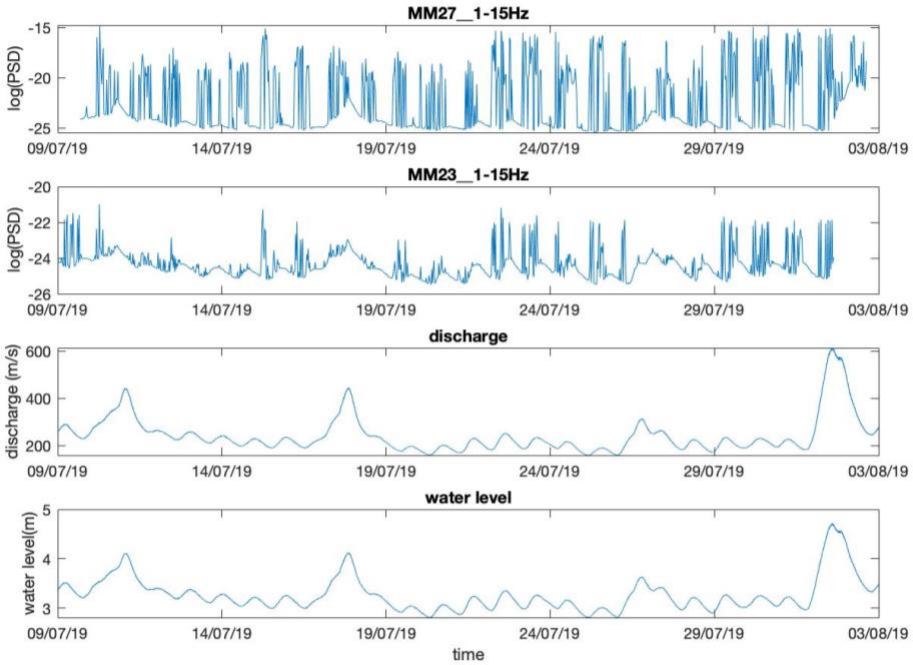


Figure 19, PSD map for 1-15 Hz of local time at 10:00:00 on July 14th (Sunday)

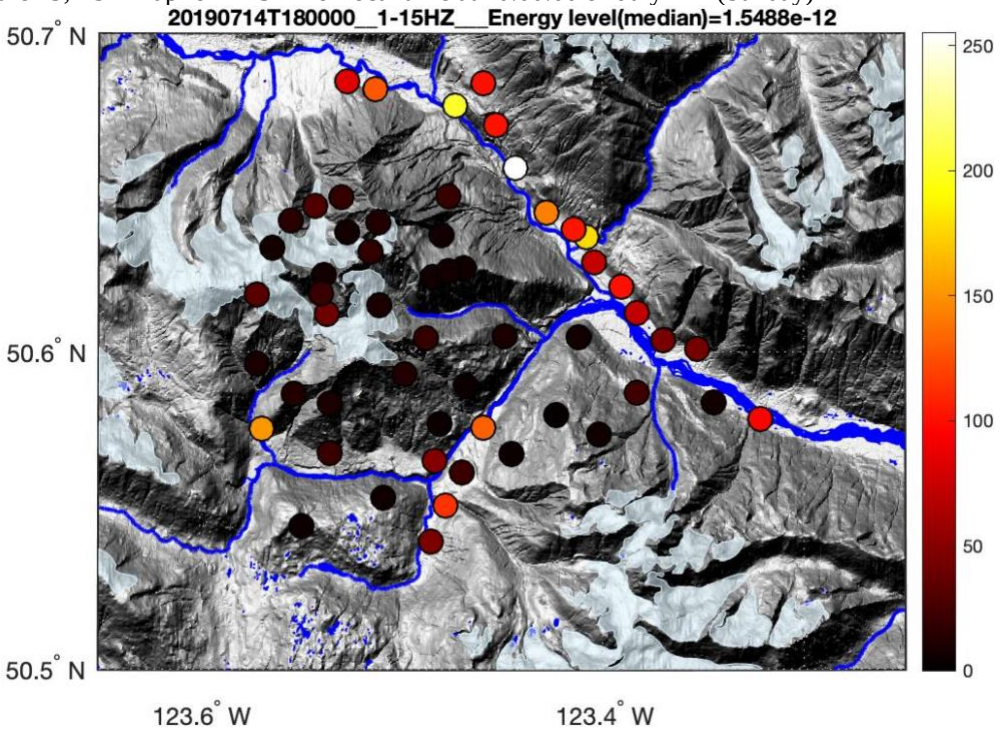


Figure 20, PSD map for 1-15 Hz of local time at 10:00:00 on July 15th (Monday)

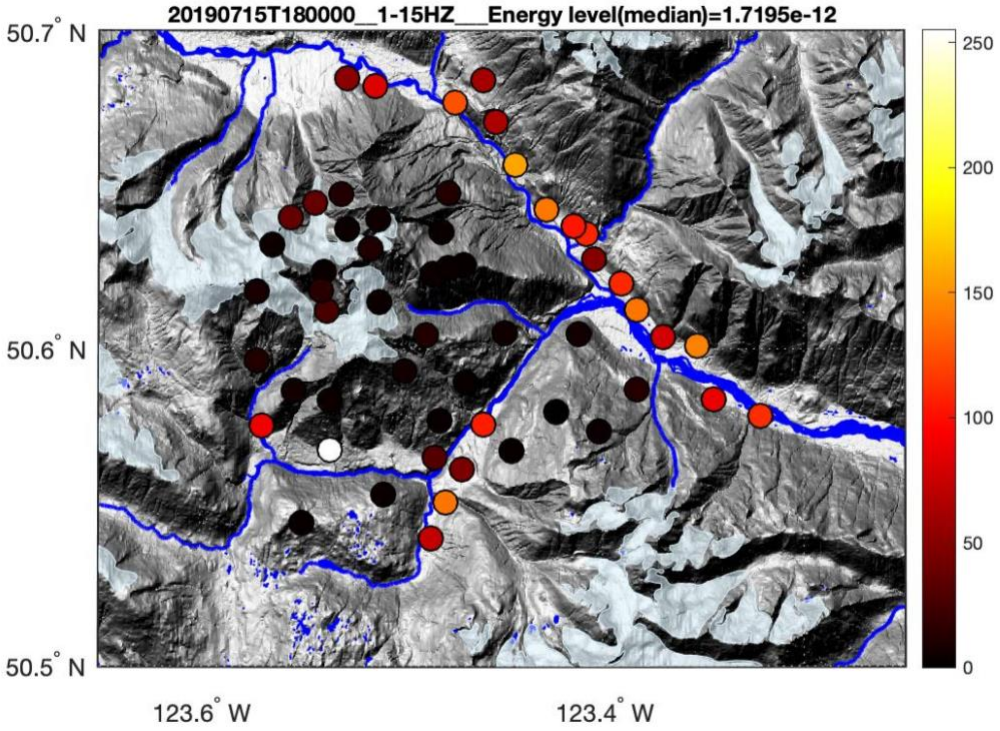


Figure 21, PSD for 1-15 Hz of local time at 10:00:00 on July 21 (Sunday)

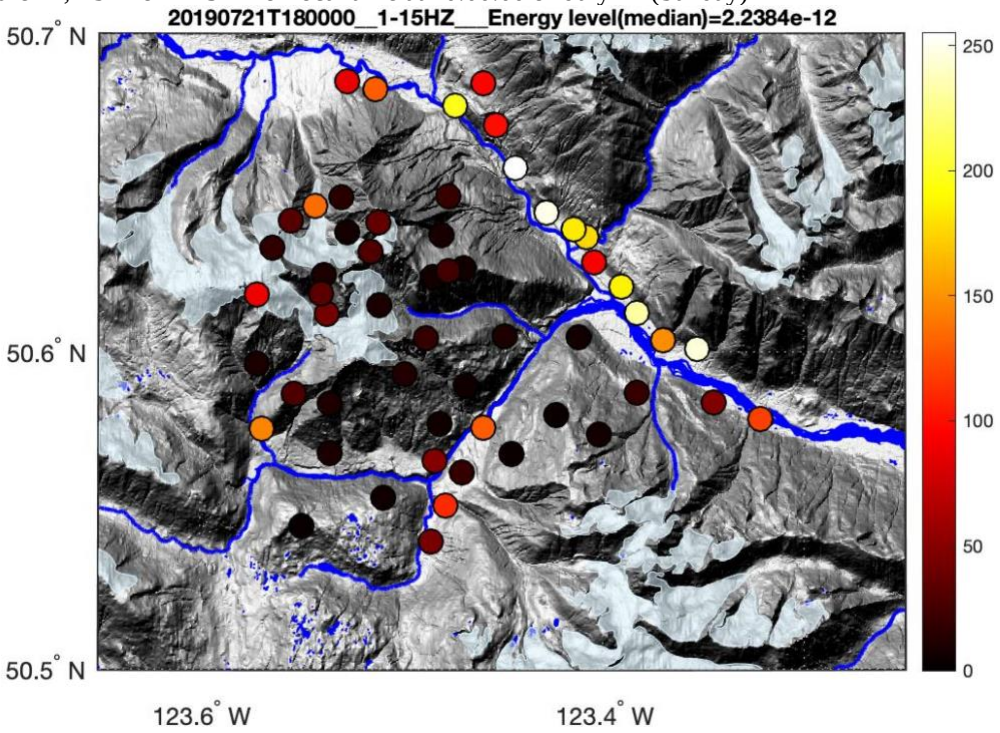


Figure 22, PSD map for 1-15 Hz of local time at 10:00:00 on July 22 (Monday)

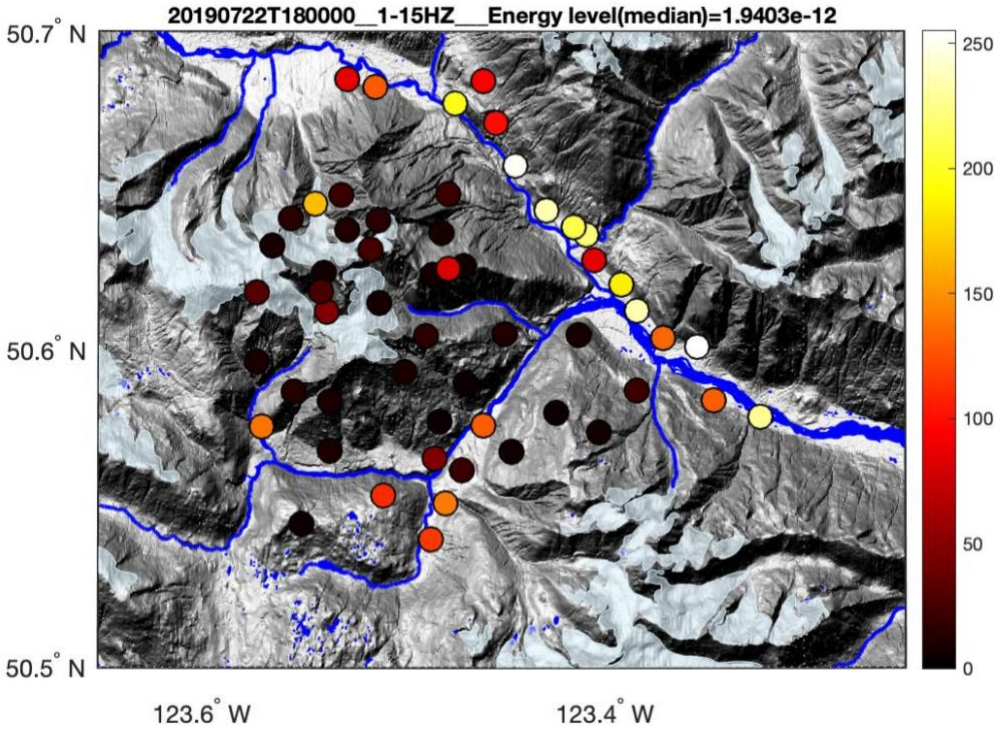


Figure 23, spectrum plots of stations MG12 and MG13 show dispersion.

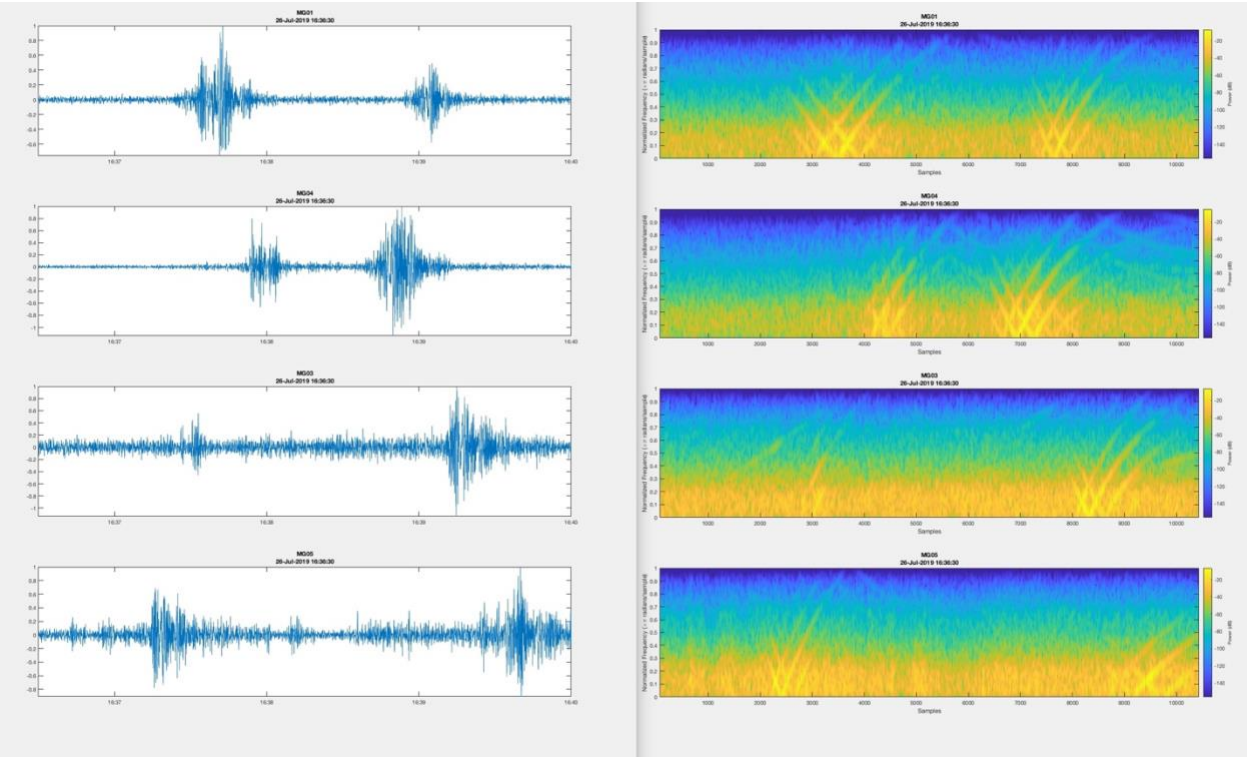
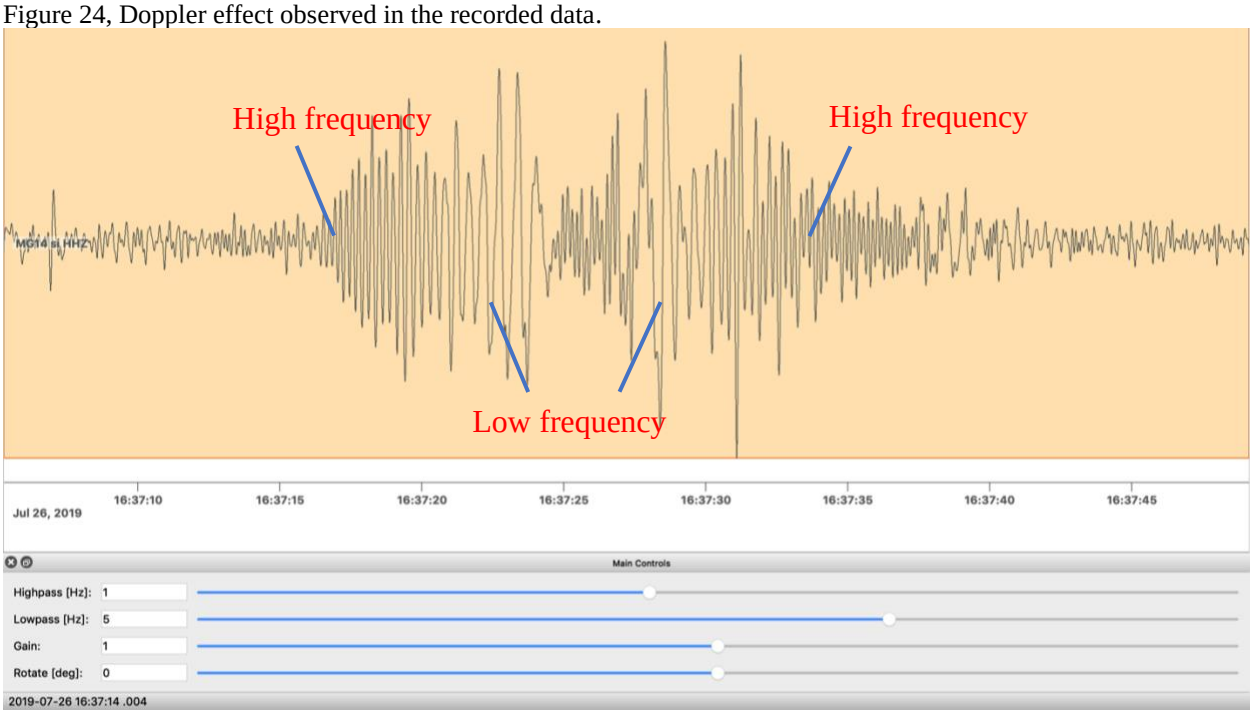


Figure 24, Doppler effect observed in the recorded data.



Reference

[1] Karchewski, B. (2019), GOPH 517-Time Series Analysis and 1D Signal Processing, Topic #3: Fourier Transformation: Geoscience Department, University of Calgary

[2] Peters, R. (2012), Tutorial on Power Spectral Density Calculations for Mechanical Oscillators, Physics Department, Mercer University

[3] Meng, H., & Ben-Zion, Y. (2018). Characteristics of aeroplanes and helicopters recorded by a dense seismic array near Anza California. Journal of Geophysical Research: Solid Earth, 123, 4783– 4797. https://doi.org/10.1029/2017JB015240

[4] Geological Survey Canada (2019). Geothermal resource potential of the Garibaldi volcanic belt, southwestern British Columbia.

Resources

[1] IRIS DMC, Data Services Products: Noise Toolkit PDF-PSD Noise Toolkit PDF/PSD bundle, https://doi.org/10.17611/DP/NTK.2.

[2] Environment Canada, Real-Time Hydrometric Data Search. https://wateroffice.ec.gc.ca/google_map/google_map_e.html?map_type=real_time&search_type=province&province=BC