Filtering Noisy Data Using Band-Passed Filter, Prediction Filter and Matched Filter

Fazlie Latib, *Department of Geoscience, University of Calgary*

Abstract

To analyze and understand the different type of filters used in seismic processing, a code was developed in MATLAB to filter noisy measured data provided with band-pass, prediction and matched filters. A function is built to perform convolution of two vectors through frequency domain multiplication. Two types of band-passed filters were used which were the Ormsby and Butterworth filters. The chosen high pass and low pass frequencies for both filters were found from the plot of the measured or original data in the frequency domain. Both band-pass filters managed to significantly remove the noise from the original data. The Butterworth filter was used as the preferred band-passed filtered result of the original data. A short prediction filter of length 30 and a long prediction filter of length 2000 were built using half of the original data. The short prediction filter did a better job in predicting the first half and second of the data while the long prediction filter did poorly in predicting the second half of the data. A matched filter was applied to both original data and band-passed data and the results matched the band-passed data well.

Background / Theory

In seismic survey, collected seismic data are usually contaminated with noise, which refers to any unwanted features in the data. These unwanted "noise" features generally provide little or no information about the subsurface. Natural sources such as wind or anthropogenic sources such as construction can generate noise at frequencies outside the seismic band.

To remove undesirable portions of data during seismic processing, a process called filtering can be applied to the data. Filtering is done by specifying a set of limits to eliminate unwanted portions of seismic data, commonly based on frequency or amplitude, to enhance the signal-to-noise ratio of the data. Generally, seismic data processing is conducted in the frequency domain where it significantly reduces the amplitude of data at noise-dominated frequency and does not change the amplitude of data at signal-dominated frequencies.

Band-pass filter is used to filter both low and high frequency noise. This is the most common type of filter for band-limited seismic data. A steep cutoff from general band-pass filter results in undesirable "ringing" or Gibbs effect in time domain which is caused by the missing nearby frequencies at about the cutoff frequency (Figure 1).

A gentle roll-off prevents prominent time-domain artifacts from forming. A simple approach to reducing these Gibbs effects is to have a linear ramp-off instead of an abrupt cutoff. This results in an Ormsby band-pass filter (Figure 2). The equation of each line in the Ormsby filter is shown in Figure 2.

Some artifacts are associated with derivative discontinuities and cannot be avoided with Ormsby. An alternative with no derivative discontinuities is called the Butterworth filter (Figure 3). The form is somewhat more sophisticated where it involves asymptotic convergence to zero and unity at low or high frequencies. The Butterworth filter has two parameters which is the cutoff frequency, f, and the order, n, for each low and high pass. The cutoff frequency defines the frequency at which amplitude is reduced to ~70.7% of maximum while the order controls the steepness of the filter. The low pass, B_l, and high pass, B_h, can be calculated using Equation 1 and 2 respectively.

$$B_l = \sqrt{\frac{1}{1 + w_l^{2n_l}}} \text{ where } w_l = \frac{f}{f_l} \tag{1}$$

$$B_h = \sqrt{\frac{w_h^{2n_h}}{1 + w_h^{2n_h}}} \text{ where } w_h = \frac{f}{f_h} \tag{2}$$

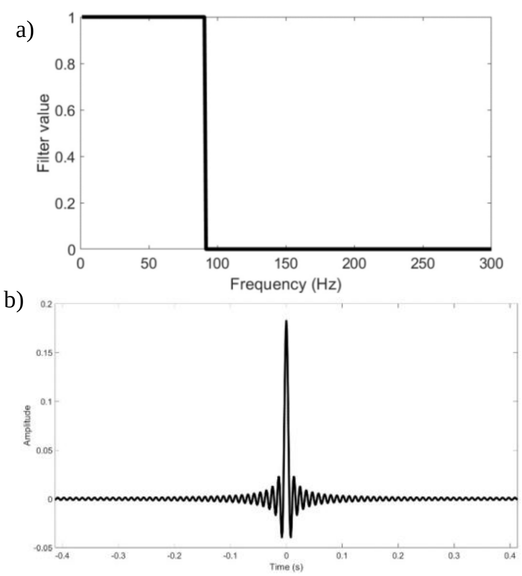

Figure 1 a) The steep cutoff in frequency domain, and b) the ringing effect in time domain. (Keating, S., 2021)

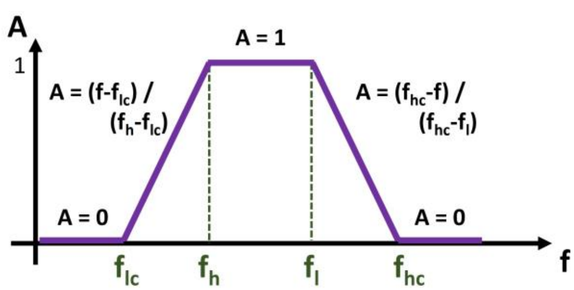

Figure 2 Ormsby band-pass filter and the equations representing each straight lines in constructing the filter. (Keating, S., 2021)

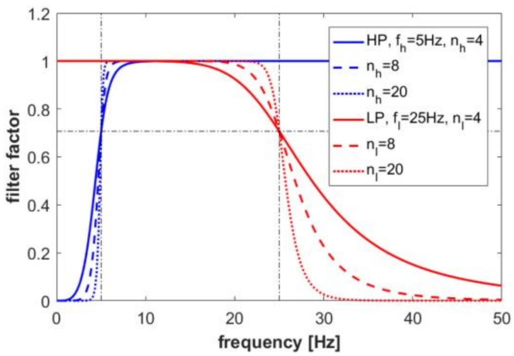

Figure 3 Butterworth band-pass filter and the examples of using different magnitudes of order, *n*. (Keating, S., 2021)

Prediction filter is one example of least-squares filter application. It is used to predict the next entry in a time series. The filter aimed to minimize the discrepancy between predicted and measured signal at the next time sample. Ideally, the prediction filter designed using s_1, s_2, … , s_{N-1}, s_N will continue to be an effective predictor for s_{N+1}, s_{N+2}, … . The prediction filter can be found by using Equation 3 where *S* is the signal vector and *y* is the target vector.

$$f = (S^T S)^{-1} S^T y \tag{3}$$

A matched filter is designed to maximize the ratio between the signal response and the noise response after filtering. The matched filter can be built by inverting or flipping the source wavelet.

From previous lab, a convolution is performed by convolving a reflectivity series with a source wavelet (Russell, B. H., 1988). Another way of performing convolution is by transforming both the reflectivity series and wavelet from time domain into frequency domain through Fourier transform. After multiplying both results, a convolution is completed after applying an inverse Fourier transform to the results of the multiplication. This process is shown in Figure 4.

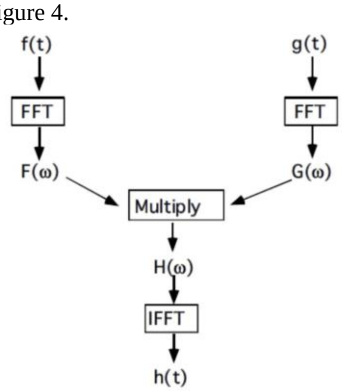

Figure 4 Convolution process through frequency domain.

Methods / Algorithm

The algorithm or code for this study is provided separately from the paper. For this study, the data file used to read by MATLAB contains measured data in “data”, the associated times in “t”, the source wavelet in “wavelet”, and the associated times in “tw”.

A function is written to perform the convolution of two vectors through multiplication of the vectors in frequency domain. The provided wavelet is convolved with a delta function time series of length 500 using the function.

The Ormsby filter is built based on Figure 2 while the Butterworth filter is formed using Equation 1 and 2. Both band-passed filters is applied to the original data by multiplying the filter with the original data in the frequency domain. The original data in time domain is transformed into frequency domain through Fourier transform.

A short prediction filter of length 30 and a long prediction filter of length 2000 is built using the first half of the original data. The short and long filters is applied to the original data and the chosen band-passed data.

A matched filter is built by flipping the source wavelet provided. This filter is then applied to the original data and the band-passed data.

Results

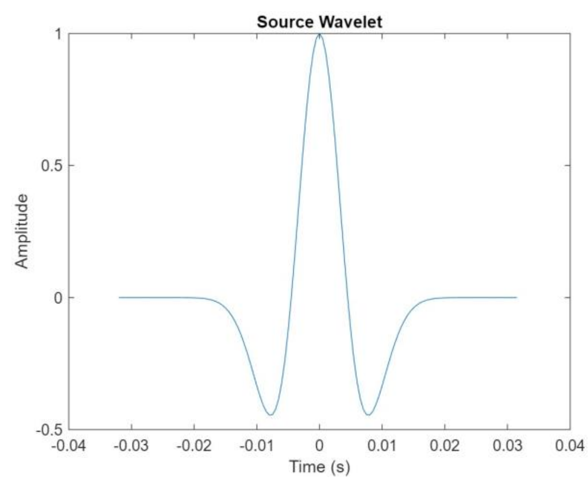

Figure 5 The provided source wavelet.

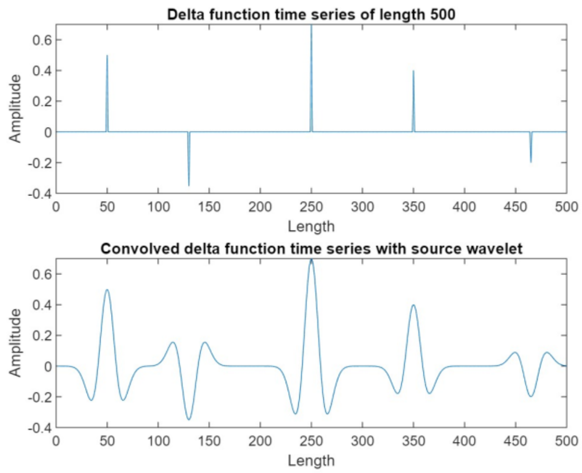

Figure 6 The delta function time series of length 500 convolved with the source wavelet via frequency domain multiplication.

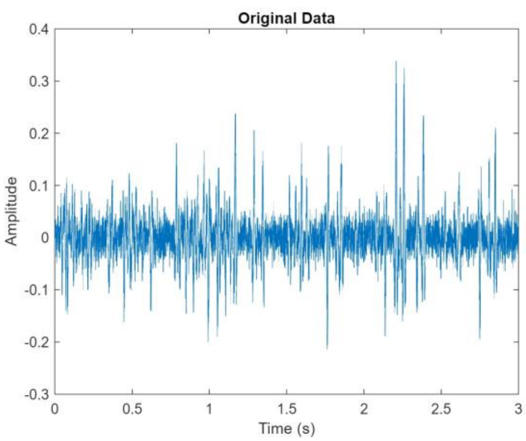

Figure 7 The original data in the time domain.

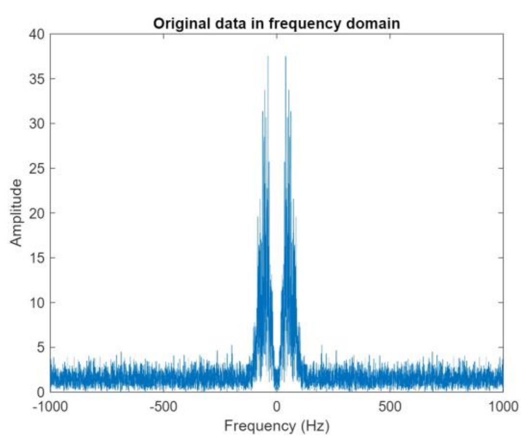

Figure 8 The original data in the frequency domain.

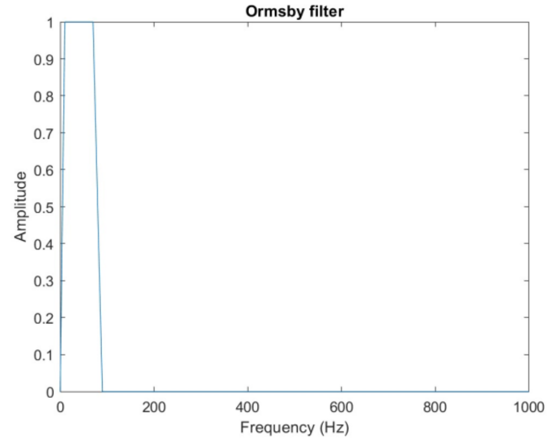

Figure 9 The Ormsby band-pass filter.

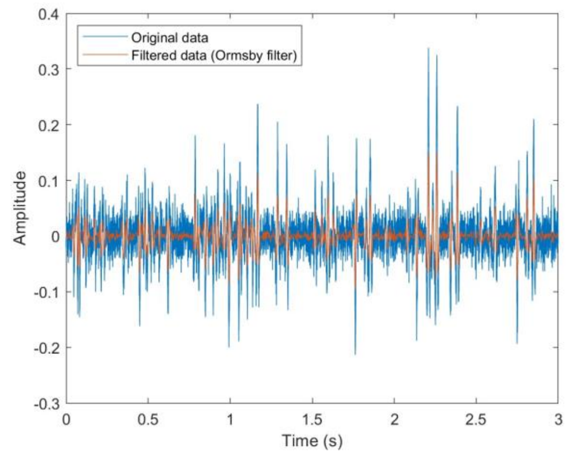

Figure 10 The filtered data using Ormsby filter plotted with the original data.

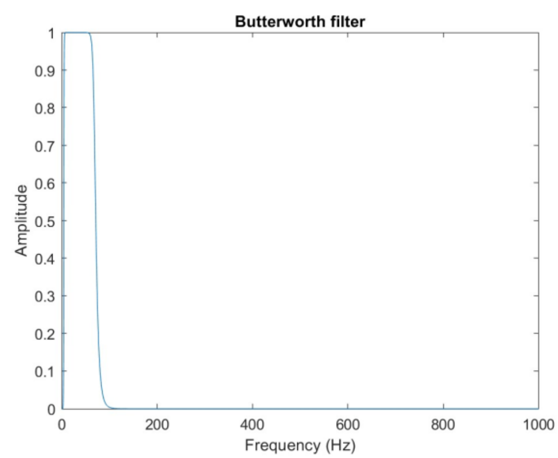

Figure 11 The Butterworth band-pass filter.

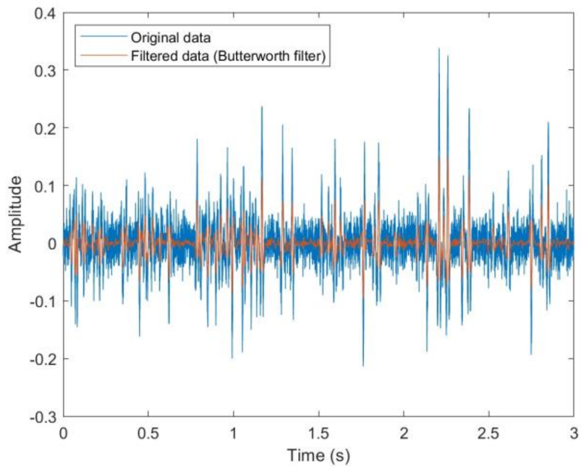

Figure 12 The filtered data using Butterworth filter plotted with the original data.

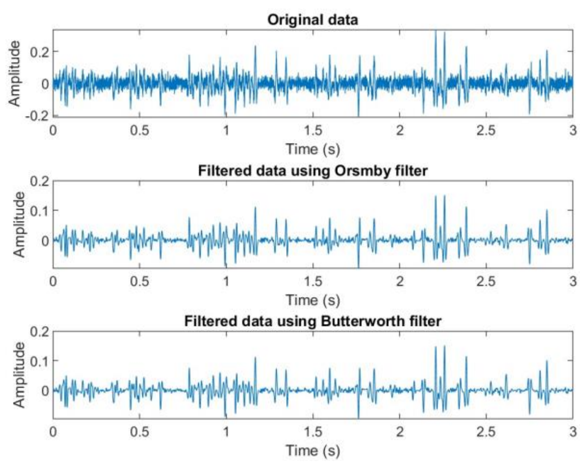

Figure 13 The original data and both filtered data using Ormsby and Butterworth filter.

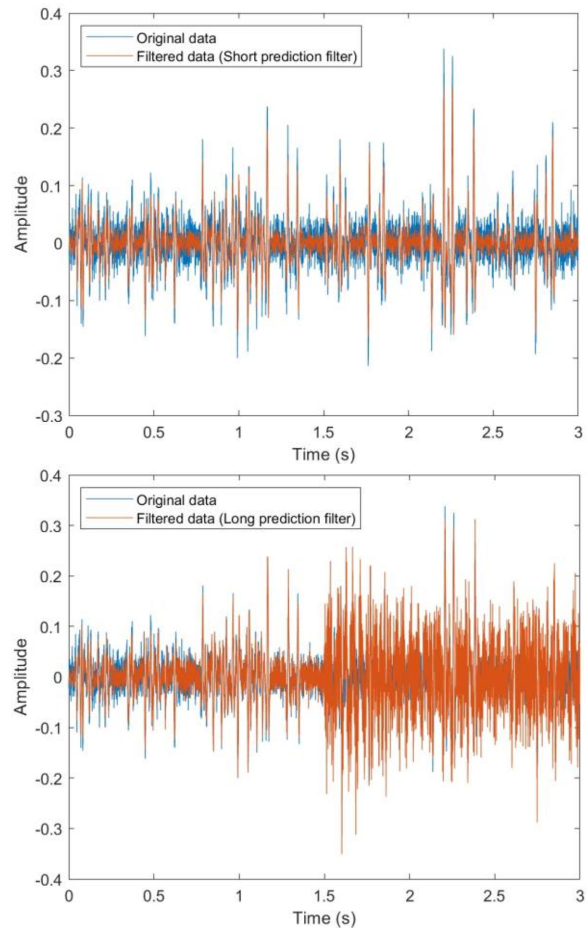

Figure 14 The filtered original data using the short and long prediction filter plotted with the original data.

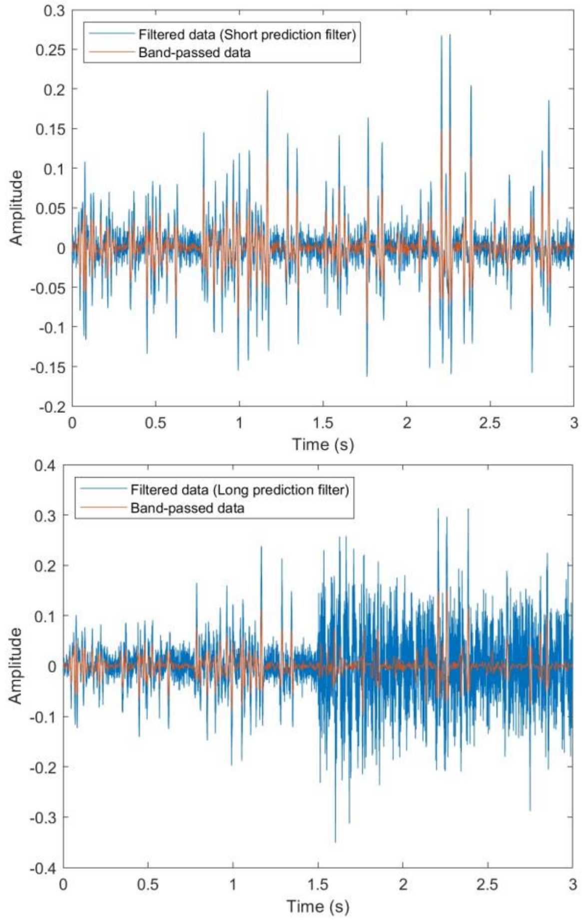

Figure 15 The filtered band-passed data using the short and long prediction filter plotted with the band-passed data.

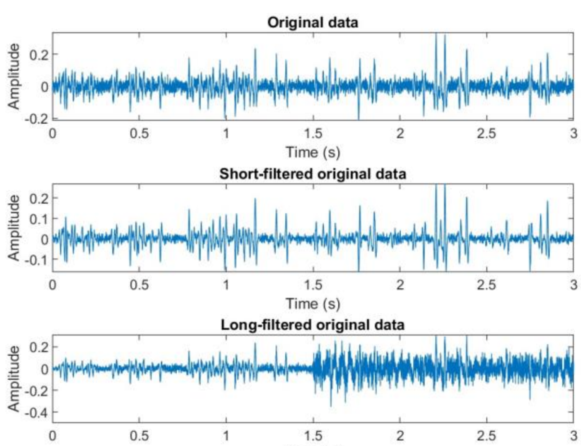

Figure 16 The original data and both filtered original data using short and long prediction filter.

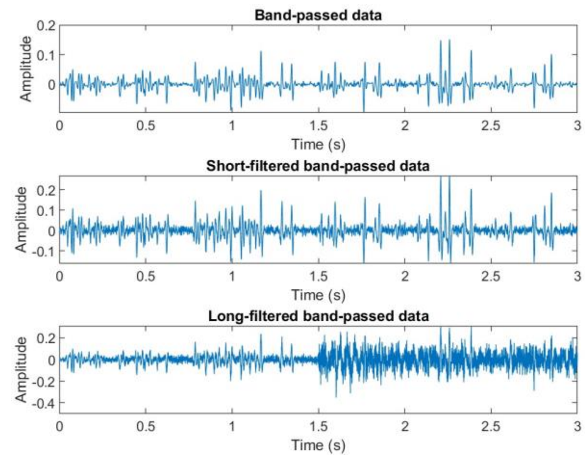

Figure 17 The band-passed data and both filtered band-passed data using short and long prediction filter.

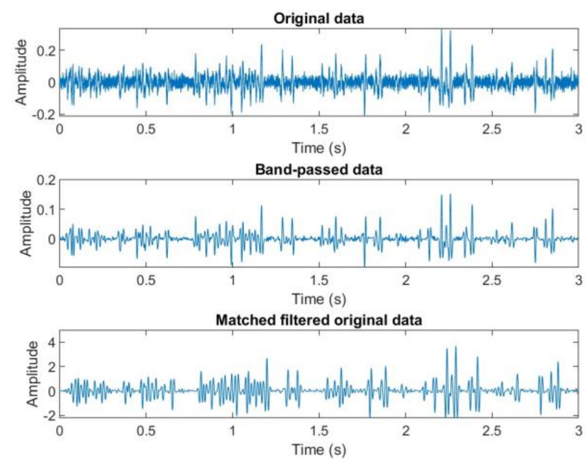

Figure 18 The original data, band-passed data and matched filtered original data.

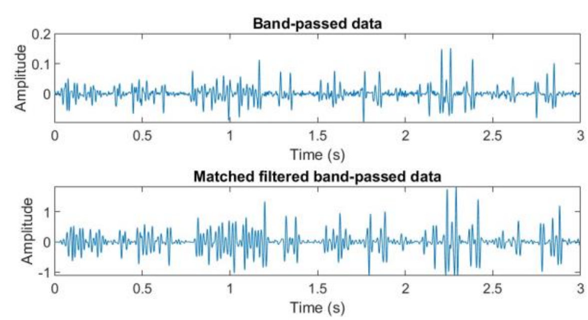

Figure 19 The band-passed data and matched filtered band-passed data.

Discussion

In Figure 6, the plot of the time series formed and the convolution of the time series with the source wavelet via frequency domain multiplication is shown. The maximum value of the convolution result can be seen occurred at the same time as the delta spike.

In Figure 13, the filtered data using Ormsby and Butterworth band-pass filters were plotted and compared with the original data. The Butterworth-filtered result were chosen as the preferred filtered result and will be use as the "band-passed data" for the subsequent analysis. The parameters for the Ormsby and Butterworth filters was set according to the plot of the original data in the frequency domain as shown in Figure 4. This plot was achieved after applying a Fourier transform to the original data in the time domain. From the figure, it can be observed that the dominant frequency of the signal lies between 0 to 100 Hz (big amplitude) while the frequency above 100 Hz (small amplitude) is considered as the noise. Therefore, the high cut frequency was set to be around 100 Hz.

Based on Figure 16 and 17, the short prediction filter does a better job in predicting the first half of the original data and the band-passed data than the long prediction filter. The short and long prediction filter was designed to optimally match the original data and hence, we want the prediction filter to match the original data. The quality of the prediction changes for the second half of the data (which was not use in designing the filter) is good using the short prediction filter but is very bad using the long prediction filter. This is because the long prediction filter is trying to predict the data at a considerably longer time ahead (in this case 2000 points or data ahead). This means that not all prediction filters can predict the data well and how well the prediction filter is based on how long the length of the filter. The better of the prediction filtering results which is the short-filtered band-passed data is very similar compared to the band-passed data except that the filtered results contained more noise than the band-passed data.

From Figure 18, the results of applying the matched filter to the original data matched almost completely with the band-passed data. The only prior information needed to design a match filter is to have the input wavelet information.

The results of applying the matched filter to the band-passed data is shown in Figure 19. The results agreed very well to the results from applying the matched filter to the original data where both results are very similar to the band-passed data. This shows that the matched filter is a great filter in improving the signal-to-noise ratio of the original data.

Conclusion

The process of convolution through frequency domain multiplication of reflectivity, *r*, with the source wavelet, *w*, gave us the seismic trace, *s*, or synthetic seismogram successfully. The "spikes" or amplitudes of the reflections in the time series were replaced by the shape of the source wavelet where the convolution result occurs at the same time as the delta spike. Both the Ormsby and Butterworth band-pass filters managed to remove the noise from the original data well. The short and long prediction filter built from the first half of the original data managed to predict the first half of the original and band-passed data well but only the short prediction filter did well in predicting the second half of the original and band-passed data. The results of applying the matched filter to the original and band-passed data matched the band-passed data very well.

References

Keating, Scott. (2021). GOPH 517 Topic 4 Lecture Note. University of Calgary, Calgary, AB.

Keating, Scott. (2021). GOPH 517 Topic 3 Lecture Note. University of Calgary, Calgary, AB.

Russell, B. H. (1988). "Part 2 - The Convolutional Model," *Course Notes Series* : 2-1-2-19, Introduction to Seismic Inversion Methods, Society of Exploration Geophysicists