

DIGITIZATION OF ANALOG SEISMOGRAMS

By

Kaitlynn Mary Stibitz

A THESIS

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Geological Sciences – Master of Science

2023

## **ABSTRACT**

The recovery and digitization of analog seismograms is critical for research into historical seismological events. Analog seismogram digitization is a difficult and complex problem and requires standards to successfully recover information from the analog media. This study investigates proposed standards for the digitization of analog seismograms. For this investigation, 'white noise' synthetic seismograms were used, with known frequency content that emulates analog records. The synthetic signal was modified to test variables such as scan resolution, interpolation algorithms, amplitude, line thickness, etc. After digitization, the digital seismograms were compared back to the original synthetic seismogram. Effectiveness of scan density can be quantified by the simplicity of digitization and waveform accuracy. Low scan resolutions adversely affect waveform accuracy and ultimately the frequency recovery. For example, a 200 DPI image can recover signals up to 2.5 Hz whereas a 600 DPI image can recover up to about 8 HZ, assuming an original recording speed of 60 mm/minute. Variability of the waveform thickness can change due to the focus of the recording beam/pen. Wider signal traces reduce the probability of accurately recovering high frequency signals due to hidden signals in the overlapping traces. We also observed the recoverable signal from low amplitude analog traces. Signals that exceed five times the width of the analog trace can be recovered within 3db of the true reference amplitude.

## ACKNOWLEDGEMENTS

I would first like to thank Dr. Kevin Mackey for taking a chance on me back in 2014, which allowed me to become an undergraduate researcher. I would not be where I am today if it were not for your constructive feedback, advice, and friendship. I would also like to thank my parents and my friends for always allowing me to bounce ideas off of them and for always supporting me in all of my endeavors. I would like to thank Dr. Kazuya Fujita for helping me become a better writer and scientist by your feedback, questions, and guidance. Thank you for helping me find and translate Russian literature and for being a valuable resource. I would also like to extend my gratitude to Dr. Jeffrey Freymueller and Dr. Min Chen for serving as vital members of my committee. Dan Burk, thank you for providing the framework for numerous programs that were used in this project as well as taking the extra time to help me comprehend study material. This project would not have been possible without the digitization efforts from my numerous colleagues in Russia, Kazakhstan, and Kyrgyzstan. Thank you to Chris Witte at Michigan State for also assisting in numerous digitizations for this project.

The financial assistance for my master's degree was jointly provided by the US Department of State and the Air Force Research Laboratory. I would like to thank both agencies who assisted my continued education and the field work in numerous countries. Lastly, I would like to thank my husband, John, who was my emotional and supportive rock during this whole process. Thank you for believing in me and reminding me to never give up.

## TABLE OF CONTENTS

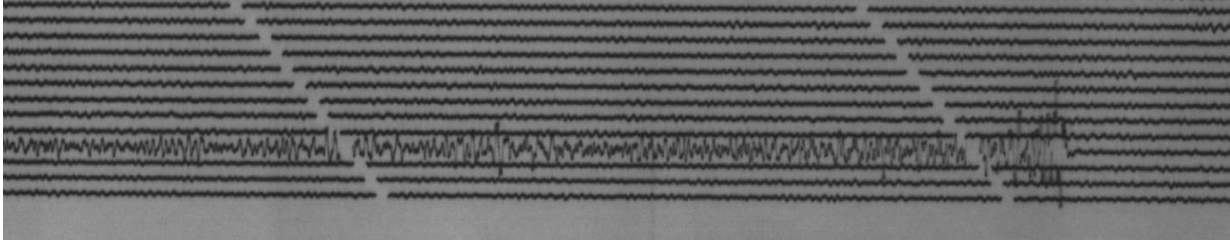
|   |     |
|---|-----|
| INTRODUCTION .....  | 1   |
| PREVIOUS WORK.....  | 4   |
| Early Digitizations.....                                    | 4   |
| Modern Digitizations .....                                  | 12  |
| Summary .....   | 21  |
| METHODOLOGY .....   | 22  |
| Interpolation Method .....                                  | 30  |
| Seismogram Sample Rate .....                                | 36  |
| Synthetic Seismogram Development and Variables Studied..... | 38  |
| DISCUSSION OF RESULTS .....                                 | 48  |
| Effects of DPI.....   | 48  |
| Image Resolution .....                                      | 48  |
| Theoretical Sine Test .....                                 | 58  |
| Technician Variability .....                                | 62  |
| Image Resolution Test .....                                 | 66  |
| Waveform Thickness Test .....                               | 74  |
| Amplitude Test.....   | 85  |
| Co-located Stations.....                                    | 92  |
| CONCLUSIONS.....  | 96  |
| REFERENCES .....  | 99  |
| APPENDIX .....  | 101 |

## INTRODUCTION

The global archives of analog seismological data are of critical importance as they contain the only recordings for many of Earth's historic large earthquakes; they contain many seismograms from the era of nuclear testing (mid 1940s to 1990s), which was largely completed by the time most stations were converted to digital. Analog seismograms are an underexploited resource, although “[they] constitute an irreplaceable dataset for the quantitative investigation and understanding of the planet's long-term seismicity” (Okal, 2015). Most of the world's seismological data up until the late 20<sup>th</sup> century was recorded on analog media, such as paper, photographic film, or magnetic tape (Figure 1). A pen or a stylus would inscribe the ground motion on various forms of media which in turn produced different levels of signal definition due to the nature of the recording method. For example, an ink pen seismogram has sharp lines compared to the photosensitive paper where a light beam has a potential for being out of focus and draws ‘fuzzy’ signal. Almost all seismological stations were converted to digital recording of seismometers in the 1990s and early 2000s. The analog nature of the older analog data prevents modern digital processing techniques from being used for their analysis. The analog data archives are also aging, and as such the analog data is at a high risk for loss due to degradation, disposal, and/or destruction.

An option to modernize these analog records and to save them from further deterioration and loss, is to scan them into high resolution image files. Scanned seismograms will not only preserve the raw data but allow the re-creation of the original analog waveform into a digital signal, which will then allow the data to be digitally processed. Digitization is conducted by manually selecting points at the peak, trough, or a change of slope in the signal or by a computer automatically identifying and tracing the signal and producing a digital waveform.

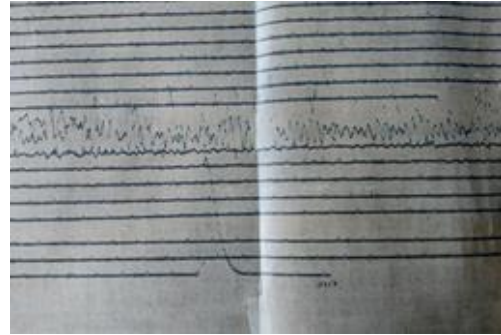
A.



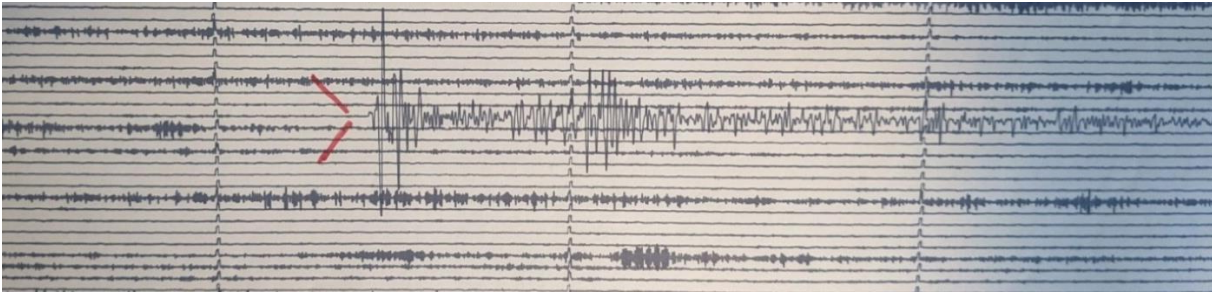
B.



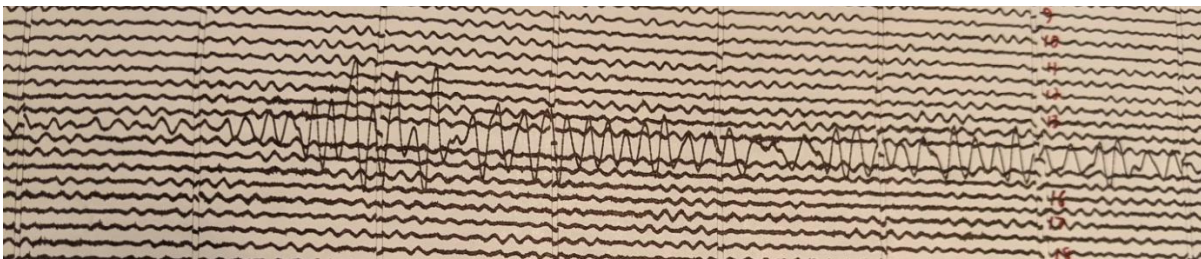
C.



D.



E.



**Figure 1. Examples of different types of analog seismograms. A) Photopaper, B) Microfilm, C) Printed seismogram from microfilm, D) Ink pen, E) Thermal paper (Photographs from Michigan State University archives).**

Numerous projects around the world have started to preserve, scan, and digitize analog seismograms. However, the successful digital recovery and usage of these signals is dependent on the accurate and complete recovery of information from the analog seismogram.

Unfortunately, there are no existing standards or recommendations on how to achieve the level of accuracy required nor, has ‘the level of accuracy required’ even been defined. For example, if a waveform is of high frequency, and the known signal contains data up to 10 Hertz (Hz), there has been no published guidance on the reliable recovery of this data. The goal for digitizing seismograms is to achieve the most precise waveform as possible retaining the frequencies, amplitudes, and signal integrity, all while limiting any artifacts from arising in the digitization. That cannot be done unless the factors within the digitization process are fully understood.

Kemerait et al. (1981) claims that “The [seismogram] database dependence is not only on the quantity and distribution of data but also on the quality of the data”, but as a community that openly shares data and data management practices, it surprisingly fails in setting criteria for obtaining quality analog data digitizations that maintain both good signal quality and frequency response. This thesis examines variables within the digitization process to achieve quality analog digitizations practices and provides recommendations for digitization to the seismological community.

## **PREVIOUS WORK**

### **Early Digitizations**

Early digitizations of analog seismograms utilized the physical seismogram where waveforms were digitized by hand either with a millimeter scale or on a digitization table. Seismograms were attached to a digitization table where a technician would manually select points with a stylus, or a puck, to recreate the seismic signal. An example of someone digitizing is displayed in Figure 2. Digitization is completed by aligning a small crosshair in a magnifying glass inside the puck and the point of interest on the seismogram. Once aligned, the person digitizing will define the point which is then translated onto a computer (Indian Institute of Science).



A.

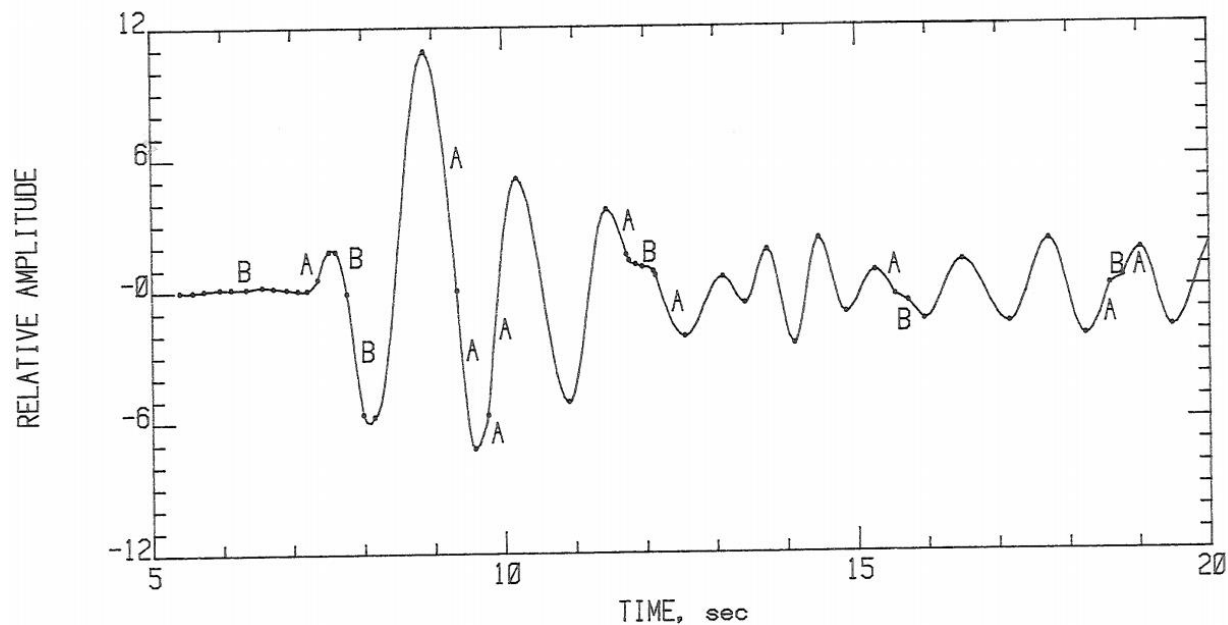


B.



**Figure 2. A) Photograph of a technician digitizing a map on an old-fashioned digitization table (photograph from Indian Institute of Science). B) Digitizer puck used for selecting points in the digitization (photograph from Eastmancuts, 2013).**

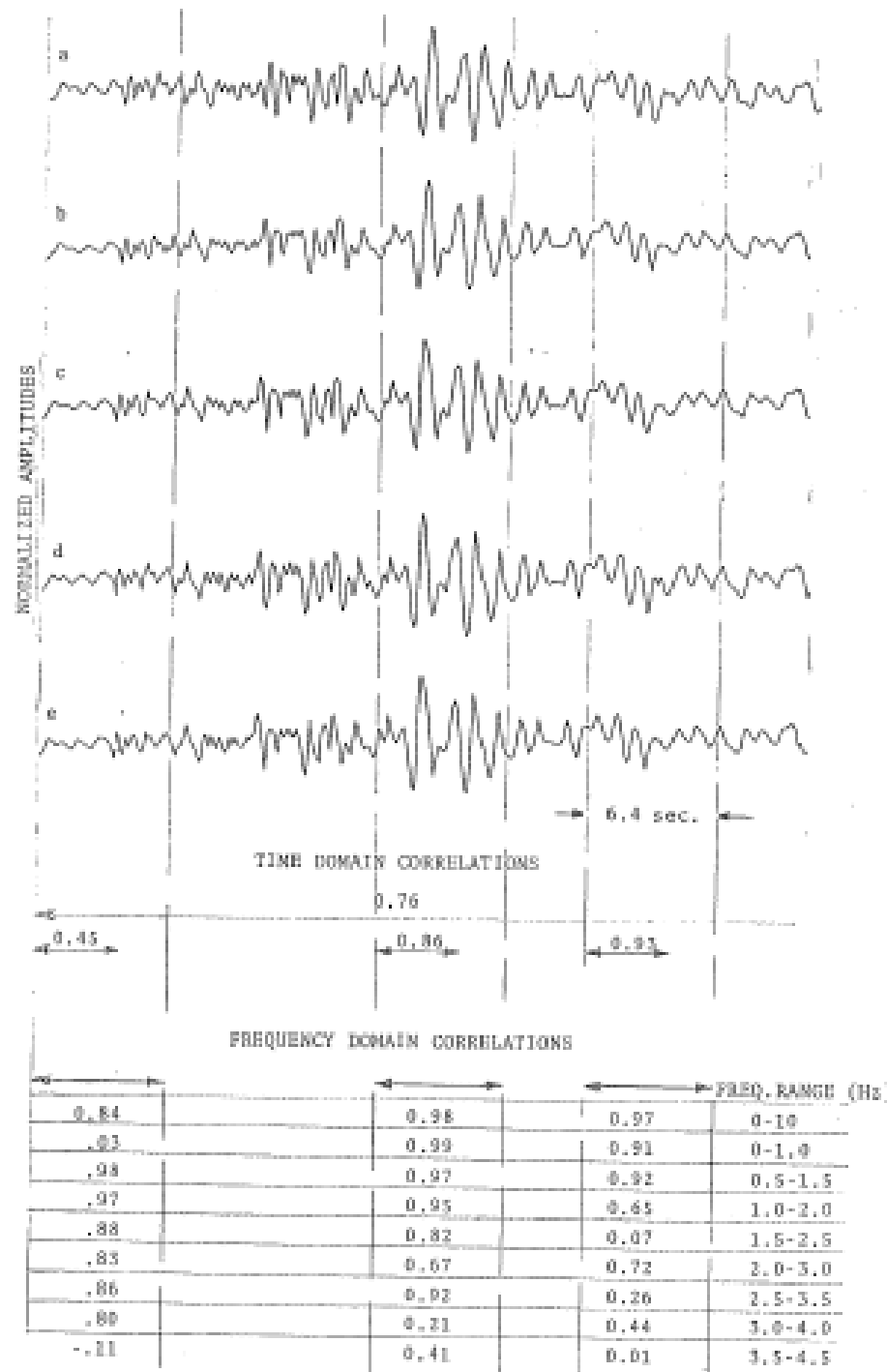
Chiburis et al. (1980) digitized 35-mm and 70-mm film chip seismograms from the World-Wide Standard Seismograph Network (WWSSN) by hand on a digitization table. For digitization, the film chips were enlarged to produce a copy of the seismogram from which (x, y) coordinates were selected throughout the waveform. This authors initially considered an automated digitizer for this project, however ultimately decided against it due to photos needing significant alterations and ‘retouching’ from overlapping signals or variances in the beam brightness. Coordinate points that were selected in the digitization process were interpolated using three different methods (four-point Lagrangian,  $\frac{1}{2}$  cosine, and  $\frac{1}{4}$  cosine) to help recreate the sinusoidal shape of the waveform. Figure 3 illustrates a digitization with the application of the interpolation techniques on the raw coordinate points from digitizing. The small dots illustrate the selected points during digitization. The letters denote the interpolation technique applied in the specific section of the waveform. Testing different methods ensured a suitable replica of the original waveform. By visual inspection, the original waveform was overlain by the interpolated method to compare the fit. If a particular section had noticeable discrepancies of amplitude or frequency recovery, it was sent back for revision.



**Figure 3. Application of interpolation techniques for digitization points. Points (shown as small dots) were selected during digitization of a 70-mm film chip seismogram. Each letter signifies the interpolation function in that section of the digitization. A) continuous mode, and B)  $\frac{1}{2}$  cosine interpolation. (Figure from Chiburis et al., 1980).**

Kemerait et al. (1981) recognized the usefulness of digitizing analog seismic data; however, they questioned how “good” the resulting digitization were. They created a series of synthetic seismograms to help quantify the “goodness” of a digitized seismogram, without specifically isolating variables within the digitization process, such as sampling rate, DPI (Number of pixels per inch of image) etc. The authors state how digitization is a complex process, with three identified sources of error. The first source of error is the inconsistency of a user’s digitization experience. Second, in order to achieve a quality digitization a program must have the ability to fit an appropriate curve to the data and apply correct interpolation methods. The final source of error in digitizing is the high potential for signal distortion. This could happen when a print from a film chip is made, or a scan is magnified before digitization which can both create and/or amplify a distortion.

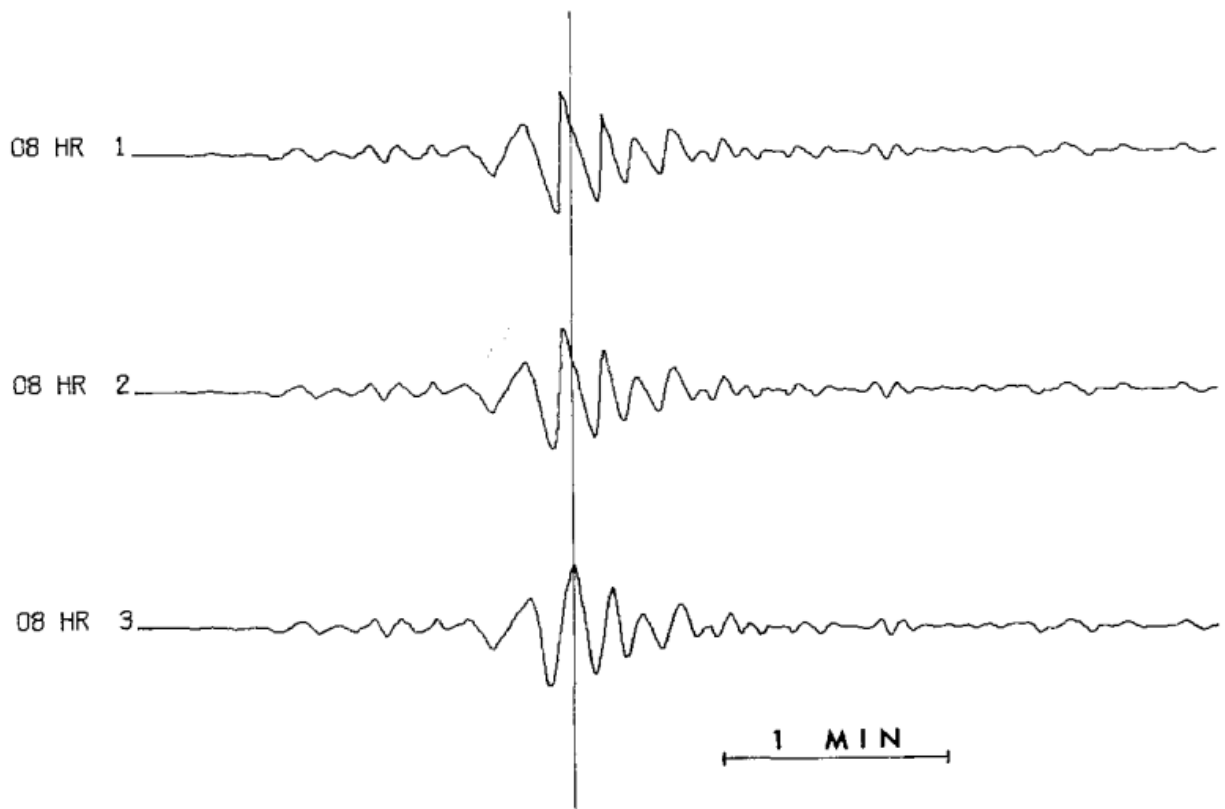
To examine a digitization’s potential, Kemerait et al. (1981) used synthetic seismic data and hand-digitized analog records. The process for digitizing records were the same described in Chiburis et al. (1980). Synthetic signals generated at 5 Hz were used to examine how well the generated synthetic signals were relative to hand-digitizations completed by four individuals. The five signals were compared against one another for their frequency response. Figure 4 illustrates the signals, and the table below describes the frequency response. The topmost waveform (a) is the synthetic waveform followed by the hand-digitized signals (b-e). The frequency recovery between the four waveforms showed a strong correlation between the frequencies of 0 – 3 Hz with a value of 0.89. From this examination, the authors of this study concluded that hand-digitizations do indeed produce adequate signals for data analysis.



**Figure 4. Comparison of synthetic seismogram (a) with four hand-digitized analog records (b-e). The table below compares their frequency response between different frequency ranges (Figure from Kemerait et al., 1981).**

James and Linde (1971) examined the range of WWSSN microfilm digitizations where they digitized one seismogram three times with the x-axis at different orientations. One digitization had the x-axis parallel to the trace, a second had the x-axis perpendicular to the seismogram drum axis, and a third where the x-axis was oriented to the direction of the galvanometer swing. If inappropriately aligned in the digitization process, the resulting digitization would have significant skew (Figure 5). The first and second traces, which correlate to the first two digitization methods, display distortion in the waveform as compared to the third trace where the x-axis was oriented to the direction of the galvanometer swing. In the third digitization method, the digitizing device was tilted at an angle similar to the swing of the galvanometer to limit the amount of distortion in the digitization.

Singh (1983) compared digitized WWSSN microfilm chips to long period seismograms from High Gain Long Period (HGLP) and Seismic Research Observatories (SRO) networks for studies on anisotropy. Paper records were created from a microfilm reader-printer where seismic traces were digitized using a semi-automatic D-Mac digitizer where coordinates were selected along the trace. Some microfilm records were deemed unsuitable for digitization as they had too thin of a line or had overlapping traces which made it hard to follow in order to recreate the trace. Singh (1983) confirmed the same digitization alignment errors as James and Linde (1971). The x-axis of the digitizing device must be parallel to the swing of the galvanometer. Both Singh (1983) and James and Linde (1971) agreed that if the seismogram alignment is incorrect during digitization, the waveform would have significant errors and negatively impact geophysical studies.



**Figure 5. Example of distortion in a digitized WWSSN microfilm chip. Each line shows a separate digitization method where the x-axis was oriented differently. Trace 1 was digitized with the x-axis parallel to the trace. Trace 2 was digitized with the x-axis perpendicular to the trace, and Trace 3 was digitized with the x-axis oriented in the direction of the galvanometer swing. (Figure taken from James and Linde (1971)).**

Early digitizations would sometimes use printed copies of magnified film chip seismograms. With this method of digitization, there are multiple layers of distortion. Initially, film chips have a potential for optical lens focal point distortion from the camera that took the original photo of the seismogram. Information deviating away from the focal point may have a distorted perspective in the data. Secondly, the seismogram would then be magnified and printed, which is yet another potential source of distortion, for easier digitization. In some instances, the magnified film chip copy would be scanned and sent to a user for digitization. Having a copy of a copy is several steps removed from the original source of the data and has a high potential for distortion and data misrepresentation in the digitization process. To limit these deformations, seismograms should be taken from storage, and scanned at a high resolution on a scanner directly at the seismic network and returned to storage for safe keeping. Having the original seismogram be full scale and have no focal point distortions can greatly improve the chances of retaining the fidelity of the seismic information in the digitization.

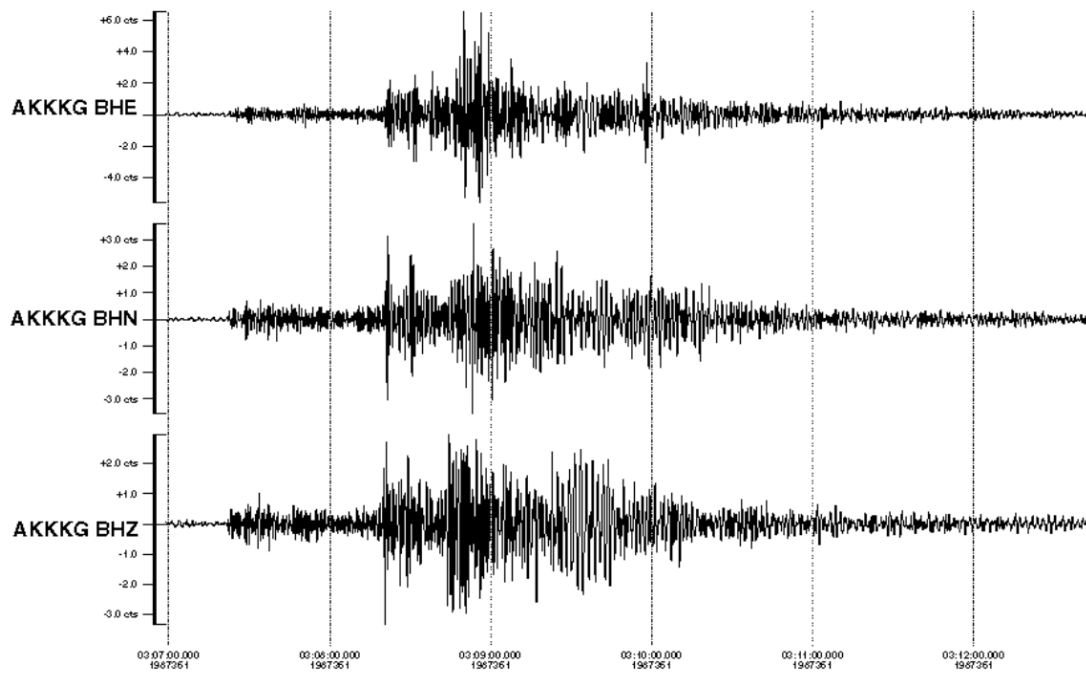
### **Modern Digitizations**

Modern seismogram digitization efforts utilize different processes, depending on the original recording media. For media such as paper or photographic film, the process uses a scanned image of the original waveform. The resulting digitization is done on a computer either manually, where an operator selects all points used to reconstruct the trace, or automatically where the images are digitally processed by a computer algorithm to recognize and follow the trace. The automatic techniques typically require human oversight to correct any errors such as adjusting the timing and correcting the trace. Magnetic tapes are another type of analog data that contain recorded seismic data, however, techniques of recovering this analog data and digitization from this form of media are not discussed here. Much of the basic theory such as



relationships between sample rate and frequency recovery remain the same. Many digitization projects had to develop their own digitization software due to the lack of a ‘one-size-fits-all’ program where numerous types of analog media could be digitized by the same program.

Between 2005 - 2011, a large-scale digitization project between Lamont-Doherty Observatory of Columbia University, USA (LDEO) and the Institute of Geophysical Research (IGR) in Kazakhstan, digitized more than 6000 records of nuclear and chemical explosions in and around Central Asia (Sokolova, 2015). Technicians digitized photopaper seismograms using a software program developed by the California Institute of Technology, known as *NXSCAN*. *NXSCAN* is a semi-automatic digitization program that requires the use of a 1980’s Sun Microsystems workstation (NXSCAN, 1992). Scanned seismograms were uploaded into the program where a line-following algorithm digitized the waveform with a sampling rate of 40 samples per second. Digitized seismograms were utilized for multiple geophysical studies such as regional travel-time curves and seismic attenuation of shear waves (e.g., Richards et al., 2015; Sokolova, 2015). Figure 7 illustrates a sample seismogram with three digitized components.



**Figure 7. Digitized three-component seismogram showing a nuclear explosion. Analog seismogram digitized using semi-automatic program, *NxScan*. Three components are shown for the station Ak-Kiya (AKK), in Kyrgyzstan (KG) with East-West as the top trace, North-South middle, and vertical as the bottom trace. (Figure from Richards et al., 2015).**

The Berkeley Seismological Laboratory (BSL) started scanning their analog seismogram archive, consisting of over 1 million seismograms, in 2003. Scientists and researchers within the lab understood the significance of digitizing analog records not only for its data importance, but the fact that scanning analog seismograms would further ensure the preservation and protection of the seismic data on the photographic and smoke paper records should they deteriorate. Researchers like Bromirski and Chuang (2003) attempted to use the older digitization software from LDEO, *NXSCAN*, but due to computer constraints the need for a new software system was realized. They developed a digitization software program called *SeisDig*. This new program utilized a MATLAB interface which provided the flexibility of use by various types of computers. Seismograms were digitized from a scanned 400 Dots per Inch (DPI) seismograms, a 4 sample per second sampling interval, and a spline interpolation method using *SeisDig* (Bromirski and Chuang, 2003).

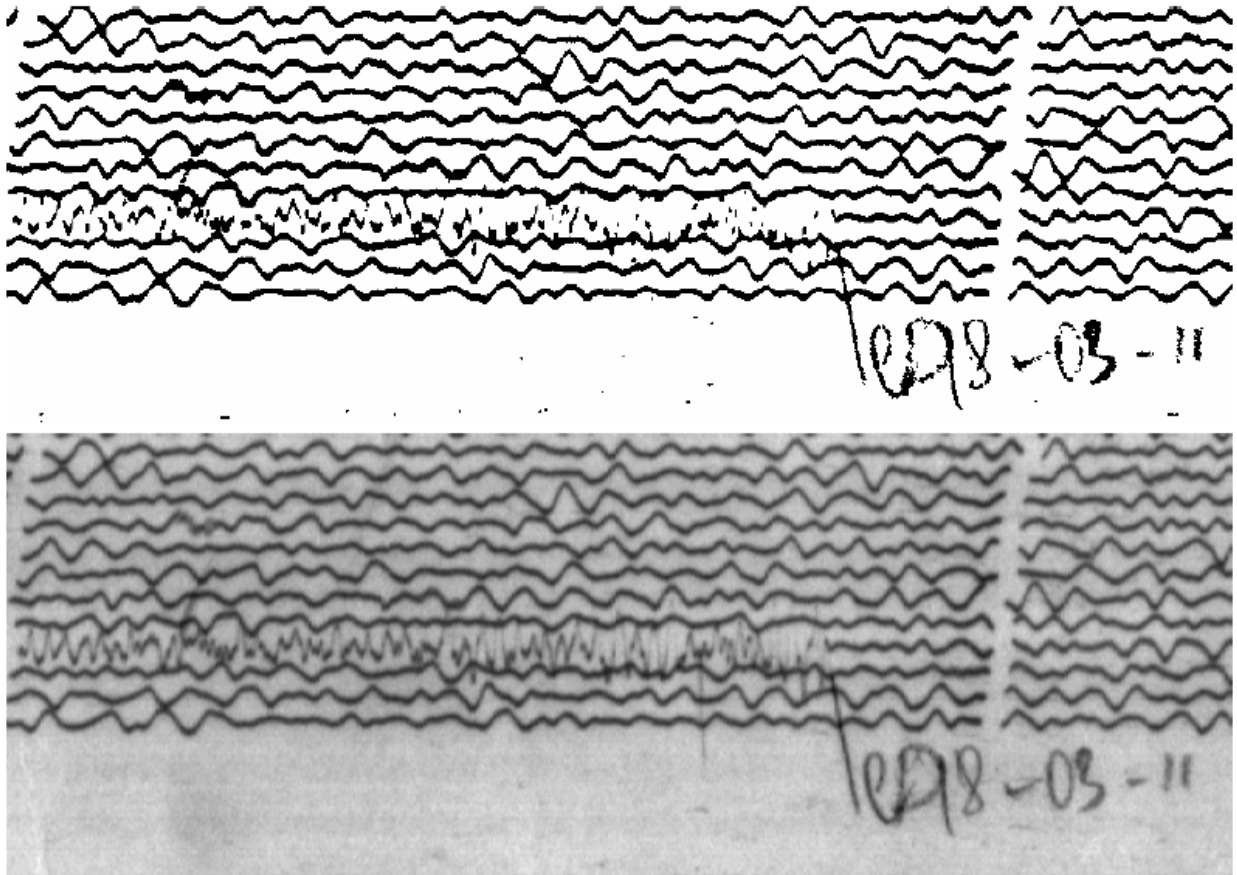
In 2001, the Istituto Nazionale di Geofisica E Vulcanologia (INGV) in Italy initiated a Europe-wide project, Progetto SISMOS (SISMOgrammi Storici), to locate, scan, digitize, and archive historical analog seismograms. A collection of countries around the Mediterranean gathered records at their observatories and sent them to SISMOS for scanning and preservation. Historic seismograms were brought to a ‘scanning laboratory’ where high resolution scanners would scan the entire seismogram at 1016 DPI (Michellini et al., 2005). Some seismograms were scanned at lower image resolutions, such as 200 or 600 DPI, that was dependent on the scanner used (Okal, 2015). A high resolution was recommended however, because it prevented the loss of important seismic information. During SISMOS, a new program that vectorized seismograms, *Teseo*, later named *Teseo2*, was developed in 2005. Seismic signals were traced by connected vectors that were representative of a piecewise cubic Bézier curve that recreated

the signal. The program allowed for manual or automatic digitization of seismograms (Pintore et al., 2005).

Ishii et al. (2015) at Harvard University began a large digitization project in which seismograms from the local Harvard station were scanned and digitized using a Harvard-developed semi-automatic digitization software program, *DigitSeis*. Upon digitization, the seismograms were uploaded to an online archive with seismic records dating back to 1933. Seismograms were scanned at image resolutions between 800 to 1200 DPI, depending on the scanner used. The authors explored higher DPI resolutions but were limited by the scanning time per image. For example, scanning a 2400 DPI image would take more than 30 minutes to scan compared to 1200 DPI which only took 5 minutes. In determining the DPI recommendations for their program, the authors made no other justification for their DPI selection for *DigitSeis*. A comprehensive evaluation of *DigitSeis* completed by the author is found in the Appendix where synthetic seismograms were put through a manual and automatic digitization process and compared for their frequency recovery.

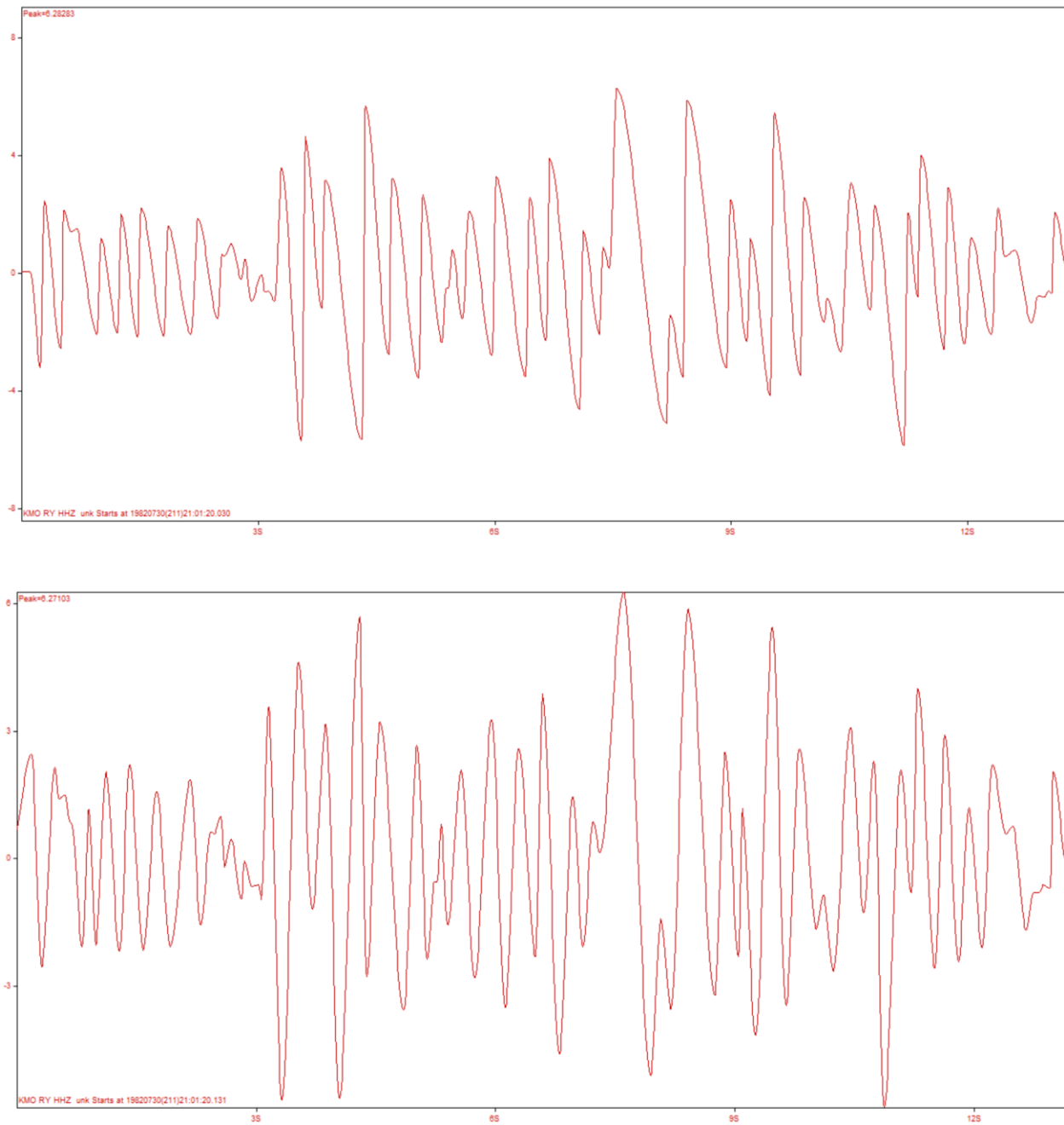
Yu et al. (2017) described several Chinese analog seismogram digitization projects that digitized and cataloged hundreds of thousands of seismic records and maps. Varying image resolutions, from 300 to 600 DPI, were used in their digitization projects. The authors also investigated the effects of image resolution and file size. They concluded that if the resolution was too low, information would be lost and if the resolution was too high the file size would be too large. The recommendation was made that in order to maintain a balance between frequency recovery and file size, an image resolution of 600 DPI was best. However, this recommendation is dependent on a seismogram's recording speed, which was not described by the authors.

Currently, Michigan State University (MSU) and the Geophysical Survey of Russia are collaborating on an ongoing effort of the collection and digitization of Peaceful Nuclear Explosion (PNE) analog seismograms from across the Former Soviet Union. Due to a majority of the detonation sites being in seismically stable regions of Russia, these analog records are of great interest to researchers. After locating and scanning the seismograms of interest, the seismograms were digitized in a manual digitization program developed by the Institute of Petroleum Geology and Geophysics in Novosibirsk, called *Wavetrack*. Mackey et al. (2009) observed differences in signal quality from scanned images. Several of the Russian-scanned seismograms were in black and white and showed low signal definition as compared to seismograms scanned by Mackey that were in grayscale that showed better clarity of the seismic signal (Figure 8). If too much detail is lost, the seismogram becomes un-digitizable because the signal is not detectable in the digitization process.



**Figure 8. Comparison between scanned image color. Top image illustrates a black and white image, and the bottom is grayscale. Note the differences in the signal definition between the two images (Top photo: GSRAS, 2001, bottom photo: Mackey et al., 2009).**

Other efforts at MSU showed the effect of skew on digitization. A situation where the seismic event is close to the recording station will result in the recording media having both high signal frequencies and amplitudes. This can create a problem in digitizing waveforms as the signals can have noticeable skew if there is a slight angle from the horizontal to the waveform. If left untouched, the waveform will become tilted during digitization. Factors that can introduce skewed waveforms into the digitization include misalignment of the recording pen to the recording media and the orientation of the paper relative to the scanner or photocopier. Figure 9 illustrates the effect of digitizing skewed waveforms. The top figure displays the signal leaning to the left which is a direct result from a skewed signal. The misalignment of the seismogram as originally digitized was 0.17 degrees. To correct this situation, the scanned seismogram must be realigned in a photo editing software, so the waveforms digitized are orthogonal to the time axis. The bottom figure in Figure 9 shows the waveform corrected after image rotation (K. Mackey, personnel communications).



**Figure 9. Digitized section of a seismogram with visible skew. Top figure: digitized section of a waveform with left tilting slopes. Bottom figure: Corrected waveform post modification with skew removed (figure from K. Mackey, personnel communications).**



## Summary

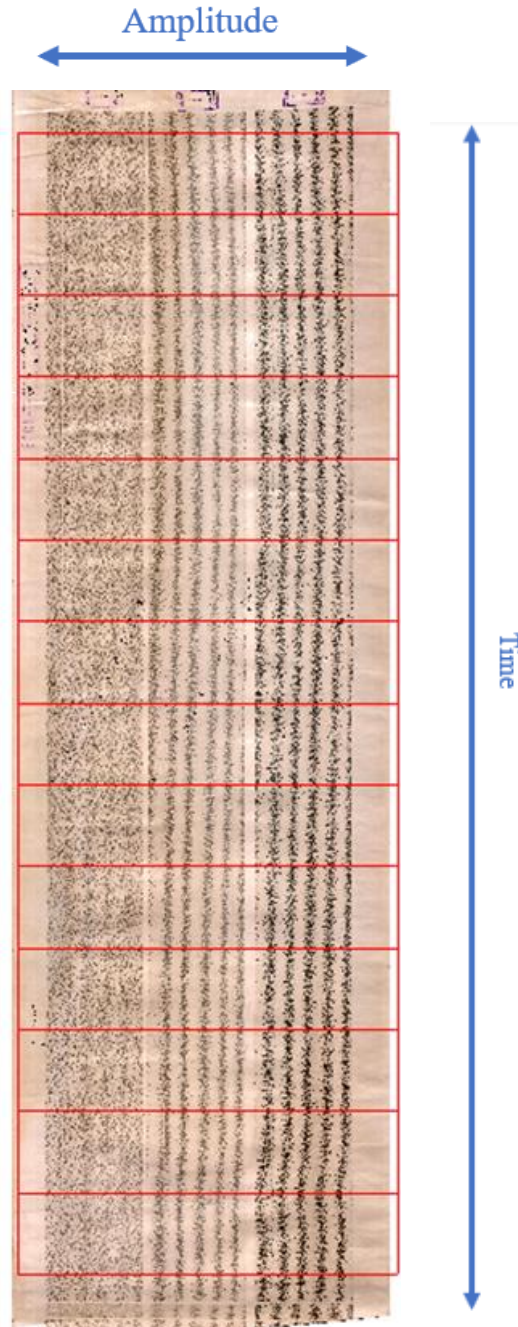
The process of digitizing analog seismograms has been an ongoing, evolving process for many decades. Due to the advancement of technology, digitization has progressed from using the physical seismogram or a copy on a digitization table to computer-based methods, like *Wavetrack* and *DigitSeis*, where the seismogram is digitized on a digital workstation from an optical scanned image. Upon review of previous digitization projects, one observation is clear: there are no standards for digitizing analog seismograms. No prior discussions have occurred about the variables and parameters within the digitization process that affect the quality and accuracy of digitized analog data. Some unknown variables include sampling rate, interpolation method, the DPI needed to retain a certain frequency, recording rate, the waveform thickness line as well as many more. The accurate use of digitized data is only as good as the digitization process. Standards need to be set, or at least discussed in length, for the seismological community to better utilize the unique datasets available from analog seismograms and to identify the limitations of previously produced digital datasets and studies based on them.

## METHODOLOGY

This section describes the digitization process and post-processing used in this research and includes background as to why certain procedures were done. For this thesis, most of the research was conducted using the *Wavetrack* software. In this digitization process, a grid is overlaid on the seismic image where inside, the trace of interest is digitized. The signal amplitude is on the x-axis and time is on the y-axis. Within the grid, there are horizontal lines that correlate to minute marks. These lines can be moved to accurately mark the beginning of each minute. In some seismograms that were recorded on a drum, the rotational speed was not always constant, causing the length of minutes and corresponding time scale to be variable. The ability to individually use a variable timing grid largely corrects the variability of the time axis and thus retains accurate timing in the digitization. The rotational speed of seismic recorders varies due to both environmental factors and the mechanics of the recording system. Other digitization programs lack this feature of adjustable minute marks, which makes *Wavetrack* an optimal software.

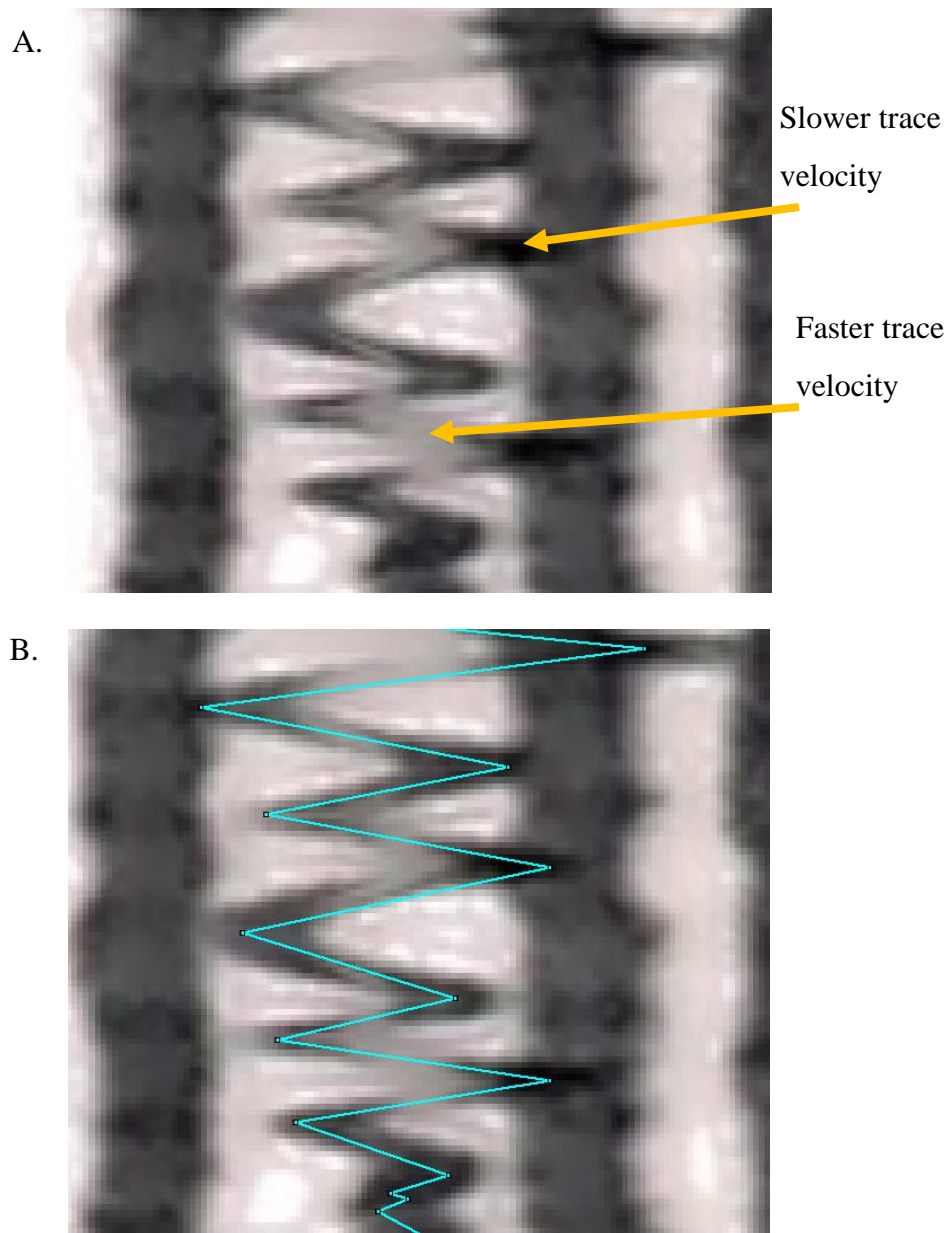
An example of the digitization grid in *Wavetrack* is shown in Figure 10 where the grid spans the entire length and width of the seismogram. This example shows a grid length of fourteen minutes and a width of 30 centimeters (cm) to accommodate seismograms with a 60 mm/minute recording speed. These dimensions are customizable to fit any sized seismogram and recording speed. The seismogram in this example has a width of 30cm and a template having this predetermined parameter ensures that the digitized amplitudes are truthfully preserved. Due to the nature of some of the types of analog seismograms, the recording media is wrapped around a rotating drum so once unraveled, the fifteenth minute is broken apart. Digitizing the broken minute is possible, but additional image processing and digitizing is

necessary to accurately merge the data together. For example, a small section, or the entire seismogram will need to be appended to the end of the original image to extend the trace of interest and ultimately the length of the digitization. Thorough image processing is required for this step to accurately align the original and appended images together. Once appended, a new *Wavetrack* grid can be put on the newly extended image and digitized as normal.



**Figure 10. View within Wavetrack of a scanned seismogram and digitization grid which is shown in red. Amplitude and time are retained in the grid to produce accurate digitizations. Amplitudes are measured in the x-axis and time is measured on the y-axis. The example seismogram is Soviet seismogram where time runs right to left.**

Within the digitization grid, the user recreates the signal by selecting points (or click points) along the trace. To accurately select click points, the user must be cognizant of the waveform thickness and exposure levels on the paper. Photographic seismograms have a light beam tracing the ground motion or signal velocity (or acceleration on a strong motion sensor) on photosensitive paper. For example, if the ground motion, or signal velocity, were fast, the light beam has less contact or exposure on the paper resulting in a lighter and thinner trace on the seismogram. Likewise, slower velocity signals have a darker and thicker appearance on the seismogram due to the light beam having more exposure on the paper. The velocity changes can be observed in click points that show the peaks, troughs, and points where the slope changes within the signal in Figure 11A. When choosing click points in the digitization process, the user must select the center points of the light beam trace and not the edges. Distortion will be created if edges are selected due to beam focus or trace thickness. With slower trace velocities having a darker and potentially wider trace, it is best to select the center of these regions as this will reflect the movement of the light beam more accurately. An example of accurately selected click points are shown in Figure 11B.

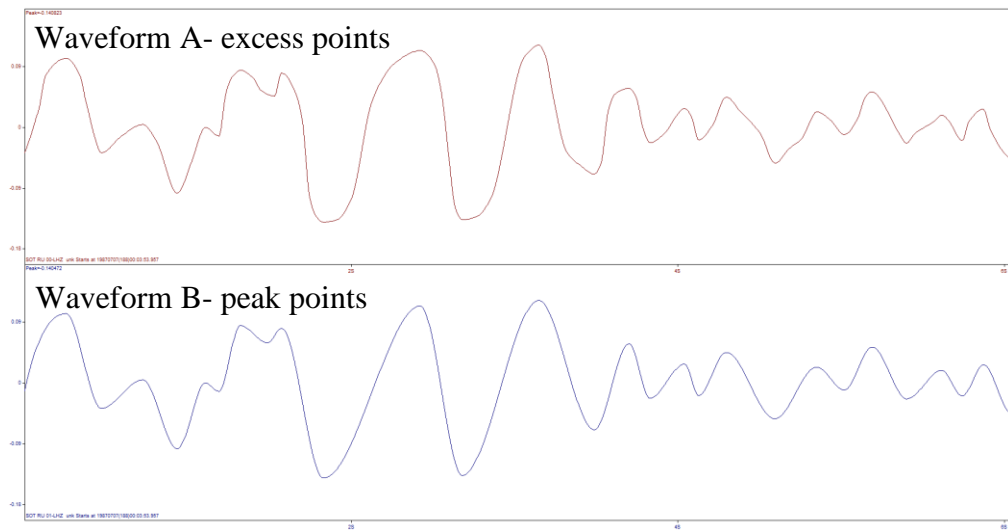
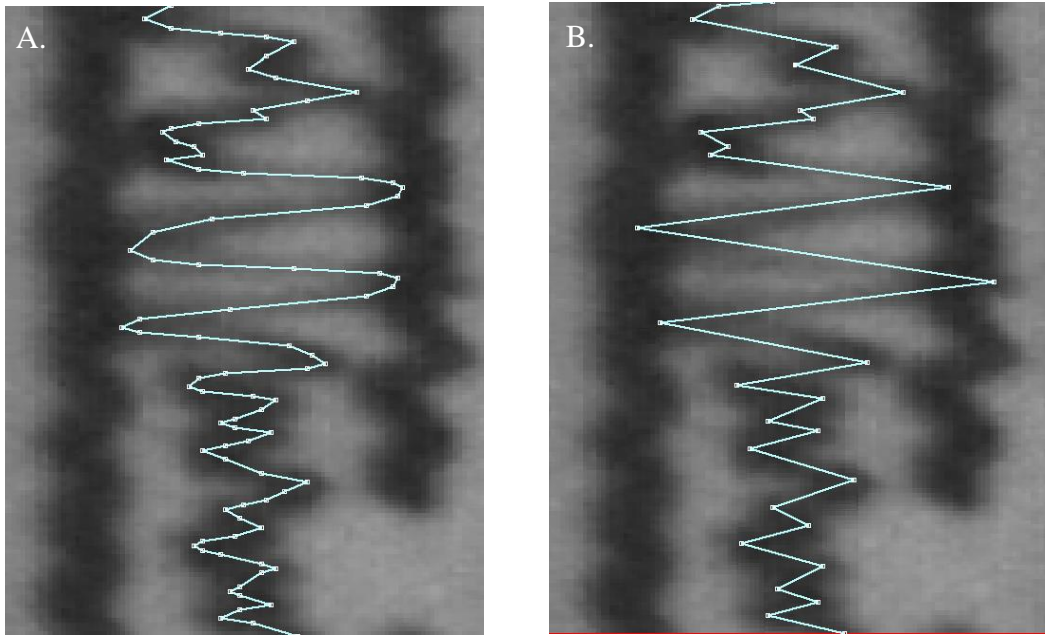


**Figure 11. Signal velocity relates to the waveform's brightness. Areas where the trace velocity was slow which causing the light beam to have more exposure with the paper appear as darker regions in the waveform. Lighter sections highlight areas where the trace velocity was fast (Shown as orange arrows in 11A). The blue line in 11B show click points selected at the center of the trace to mimic the true movement of the light beam.**

The user must also be mindful to select points only at peaks and inflections along the trace. *Wavetrack* interprets these click points as a series of line segments and exports a linear fit to the digitization of the seismic waveform. As this creates an unnatural shape, post-digitization processing is needed. A curve fitting algorithm called Piecewise Cubic Hermite Interpolating Polynomial, or PCHIP, is used to fit a realistic curve to the click points selected in the digitization process (Fritsch and Carlson, 1980). The waveform generated by the PCHIP algorithm passes through each click point and retains amplitudes and as such this algorithm is the preferred way to reinterpret waveforms digitized by *Wavetrack*. An analysis of testing different curve fitting algorithms is explained later. The waveform is then exported with a 100 samples per second sampling rate resulting in a realistic digital waveform. Choosing additional click points is tempting by novice digitizers, but these extra click points do not recover the waveform as well and create artifacts with the curve fitting algorithm. A comparison between a section of a digitized seismogram with excess click points (A) and one with only points chosen at the peaks, troughs, and any changes of slope between the peaks (B) are displayed in Figure 12. Below the waveforms are digitized signals with the application of the PCHIP interpolation. Note how Waveform A, the signal with additional points, has a smoothed-rectangular shape compared to Waveform B, which has only the peaks chosen. The more sinusoidal shape of Waveform B is a more realistic seismic signal. Yellow highlighted regions display sections of the waveform with noticeable differences in shape. Further discussions regarding the examination of different curve algorithms and deciding digitization sample rates are described later in this section. Other digitization procedures that utilize automatized routines or a different interpolation algorithm may need to approach point selection differently. The above description relates to the *Wavetrack* software and post-processing of waveforms in use here at MSU. However, having a thorough

understanding of the steps of the digitization and post-processing of waveforms is the only way to accurately recover analog waveforms.





**Figure 12. Comparison between a digitization with excess click points and one with only peaks chosen. In 12A, the blue digitized line illustrates click points following the line to recreate the signal whereas 12B only shows the peaks selected. Choosing only peaks, a curve fitting algorithm can recreate a more accurate signal.**

One major flaw of *Wavetrack* is that it does not retain the original click points of the digitization and as such a multi-step post-digitization process is necessary to have an archive of original click points in the instance of data recovery. The post-digitization processor back calculates the original user click points by finding the changes of slope in the digitization. Having a high digitization rate within *Wavetrack* allows for better point identification which are then used to fit a curve fitting algorithm to recreate the waveform.

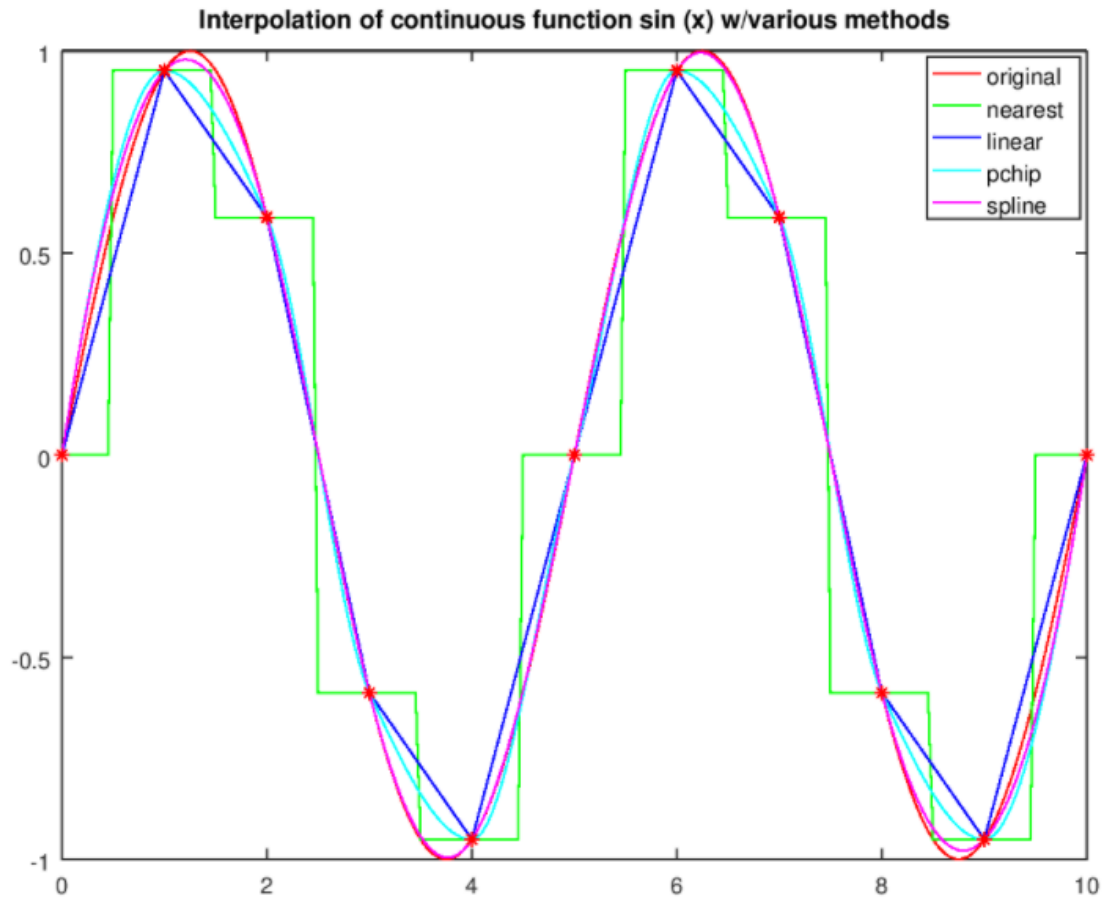
## **Interpolation Method**

As previously discussed, the *Wavetrack* program exports linear interpolations of the signal between the chosen click points. The goal of digitizing analog seismograms is to recover the original waveform and the only way to achieve these results is to apply different curve fitting algorithms to these discrete click points. Examining the effects of multiple interpolation methods in the frequency and time domain can help determine which curve fitting or smoothing algorithm best estimates the original signal (D. Burk, personnel communication, 2020). We interpolated the original click points using three methods: 1) a cubic Hermite spline, 2) a cubic spline, and 3) a Piecewise Cubic Hermite Interpolation Polynomial (Fritsch and Carlson, 1980). There are other curve fitting interpolation methods, however we chose to compare only three as these methods were readily available within Python Obspy.

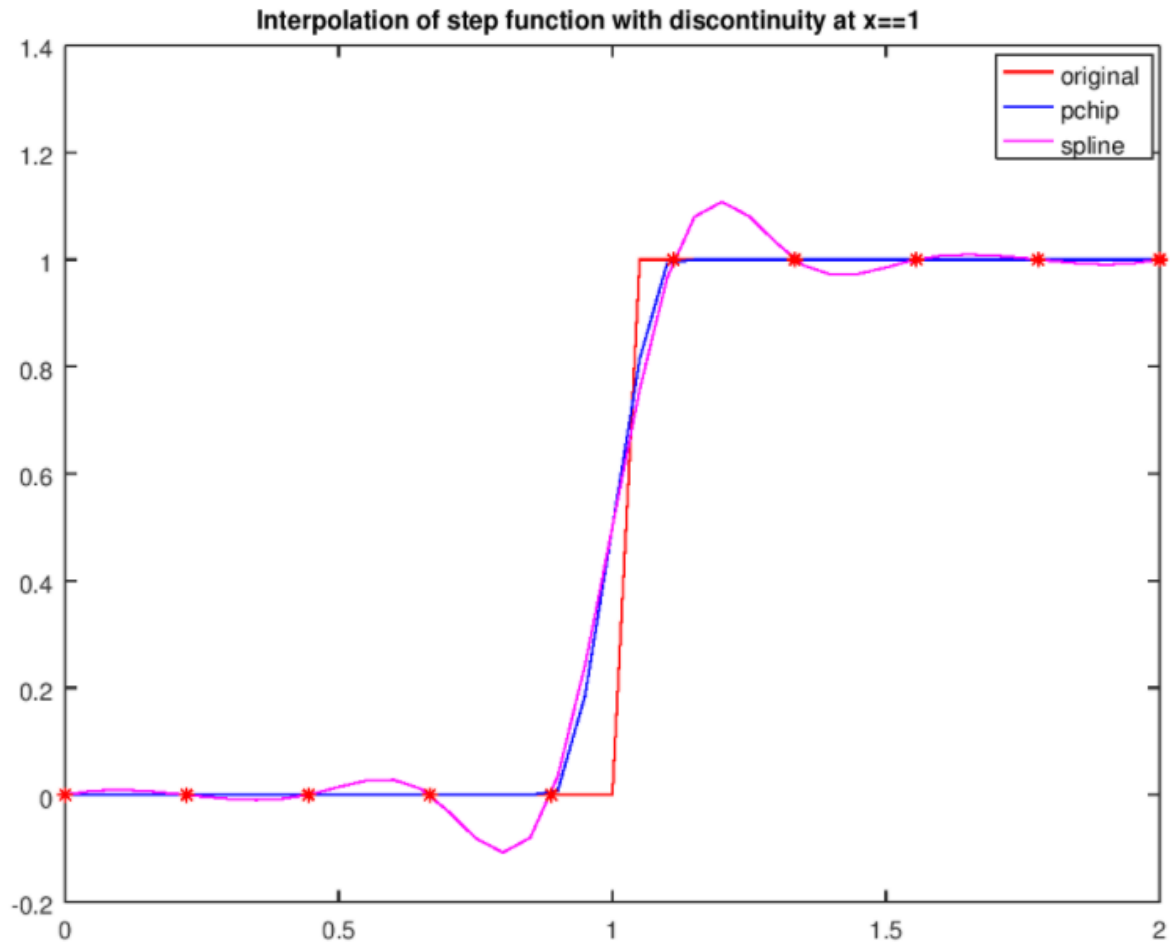
Interpolation methods display how the signal is modelled and how well the waveform is retained post processing. Each interpolation method has its own unique way to ‘draw’ the signal. Figure 13 illustrates how each interpolation method was drawn over an even time interval on a continuous sine function (red line). The red points denote the discrete points within the waveform selected by a user. Each colored line is a different interpolation method. The spline function (pink) displays a symmetric curve closely matching the original sine function. The

PCHIP interpolation (aqua) on the other hand aligns closer to the linear interpolation (dark blue) is asymmetric in shape. For this example, the spline function interpolates the signal better than the PCHIP.

Examining the PCHIP and spline function further, a basic step function was created and both interpolation methods were applied. Figure 14 shows a step function (red line), points within the signal which guide the interpolation algorithm in recovering the original signal (red points), and the two interpolation methods (PCHIP is blue, and spline is pink). Near the discontinuity at  $x = 1$ , the spline interpolation overshoots the amplitude of the step function whereas the PCHIP function is constrained and follows closer to the original function. For this example, PCHIP performs better interpolating the data because it follows closely to the original waveform.



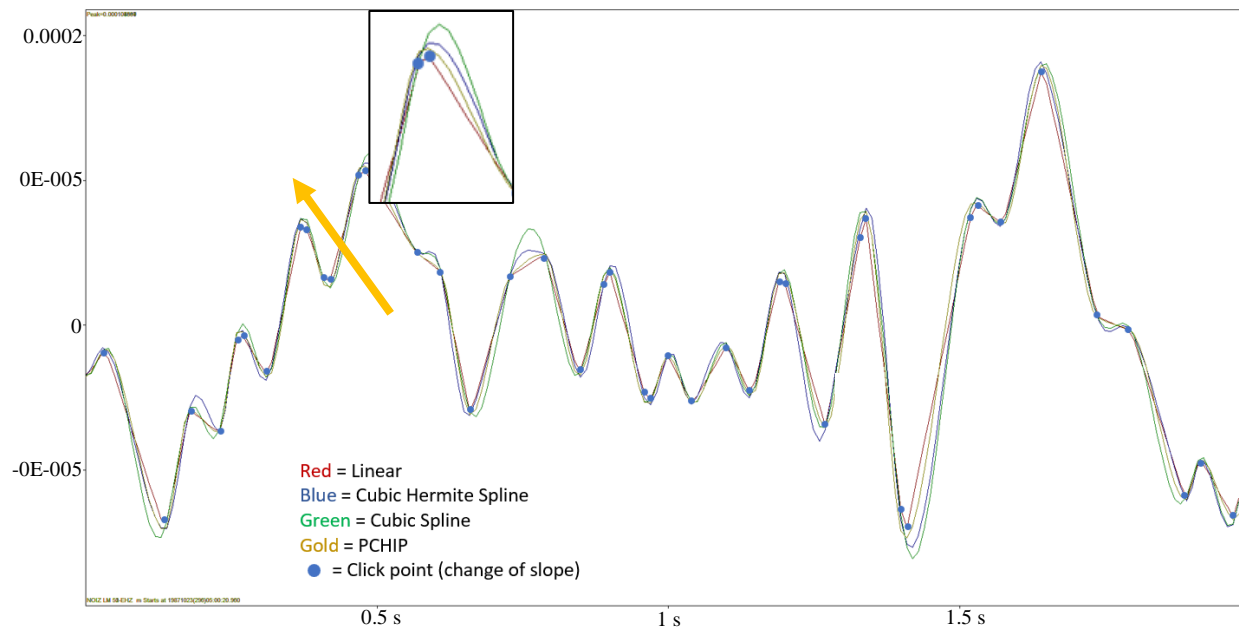
**Figure 13. Comparison of multiple curve fitting algorithms in the time domain. Red stars denote click points where there is a change of slope in the signal. The blue line represents a linear interpolation. The aqua line (PCHIP) and pink (Spline) interpolations highlight how each method estimates the curve (Octave Forge Community, 2017).**



**Figure 14. Step function with two curvilinear interpolation methods. A basic step function, outlined in red, with selected click points within the signal shown as red stars. A PCHIP interpolation method, shown in blue, preserves the shape of the original signal better than the spline interpolation, shown in pink (Octave Forge Community, 2017).**

The click points from *Wavetrack* are not distributed evenly in time, which is not compatible with desired digital processing. Figure 15 illustrates this effect. The blue dots are the click points chosen in *Wavetrack* with different colored lines showing the interpolation methods. The click points are the true peaks in the waveform as well as areas of a change in slope. The original linear interpolation (red) from *Wavetrack* still shows that it is not a good representation of the waveform. Both spline interpolations (blue and green) overshoot the peaks in the waveform while the PCHIP waveform (gold) follows closer to the original linear output and does not exceed the true amplitudes of the waveform. An inset within Figure 15 shows a peak in the waveform illustrating the overshooting peaks from the spline interpolations. For *Wavetrack's* points chosen at uneven time intervals, the spline functions try to apply a symmetric curve fit to the waveform. In this example, the PCHIP interpolation method is preferred due to accommodating the uneven time points which are representative of the peaks and troughs in the waveform all while not overshooting the amplitudes.

In the examples shown, both spline and PCHIP have situations where one may stand out over the other. The spline function works well in waveforms that have discrete points chosen at even time intervals and the points are mere guidelines to recover the original signal. This is usually the case for automatic and semi-automatic digitization programs. The PCHIP method on the other hand excels at situations where digitization points are representative of the true peaks, which happen at uneven time intervals, like our *Wavetrack* program. For seismograms digitized in this study, the PCHIP interpolation method was applied in the post processing step.



**Figure 15. Waveform with the application of different interpolation methods. The blue circles are the click points chosen in the digitization process. Red is the linear output from the manual digitization program *Wavetrack*. The blue and green lines are spline interpolations, and the gold line is the PCHIP interpolation. The inset illustrates a zoomed in peak of the waveform.**

## Seismogram Sample Rate

Sampling rate is defined as the number of samples per second in a continuous digital waveform signal. For seismogram digitization, the continuous analog signal is converted to be represented by a series of discrete points, or samples, each representing a specific time and amplitude. If a high enough sampling rate is used, the complexity of the signal will be better recovered because more points are used to define the shape of the signal. Too low of a sampling rate will yield a broader curve and potentially loss of high-frequency components of the signal due to limited points along the waveform. Figure 16 shows a wave with various sampling rates and better signal recovery with an increase in the sampling rate (shown as rectangles). To ensure an accurate waveform recovery with best signal retention, a high sampling rate is encouraged.

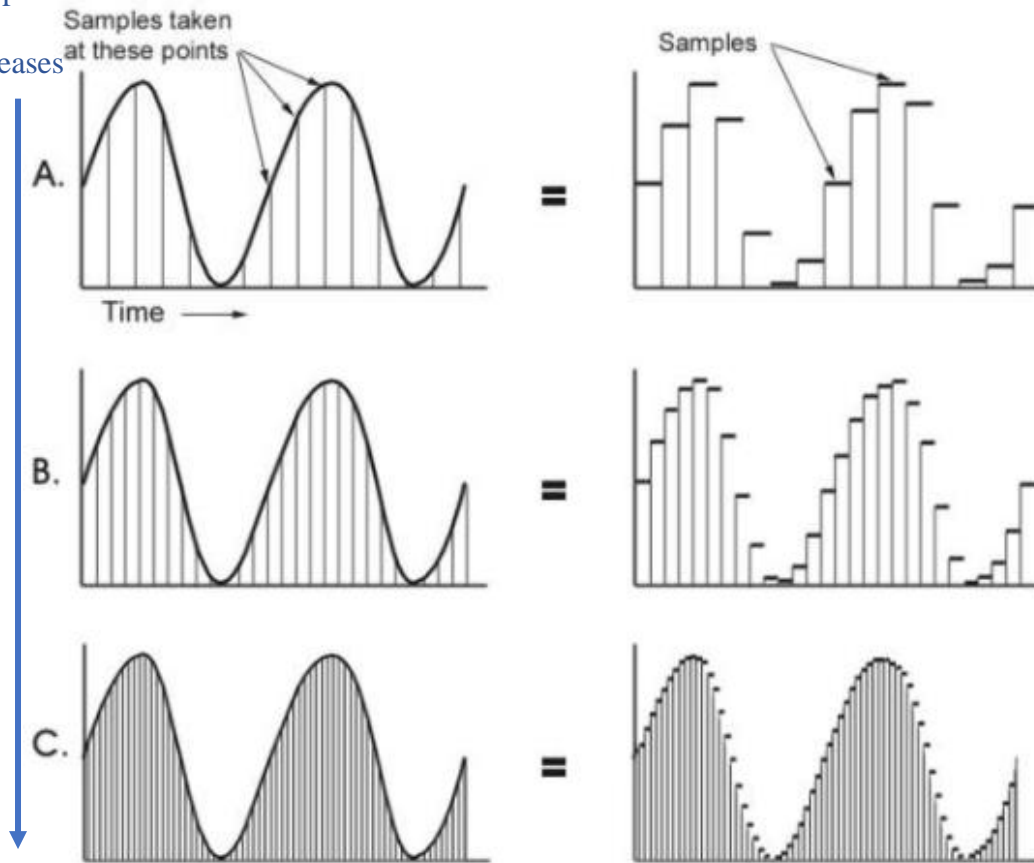
Within the same post-digitization processor that applies the PCHIP algorithm used in our research, the processor resamples the data at a 100 samples per second. This interval yields a Nyquist frequency of 50 Hz. This is five times above the upper limit of a 10 Hz response for a short period seismogram.



## Increasing Sample Rates

Sample Rate

Increases



**Figure 16. Sample rate, shown as rectangles, correlate to the recovery of the signal. A higher sampling rate will capture more complex frequencies and yield a better signal because there are more points along the line compared to lower sampling rates (Image modified from Brown, 2021).**

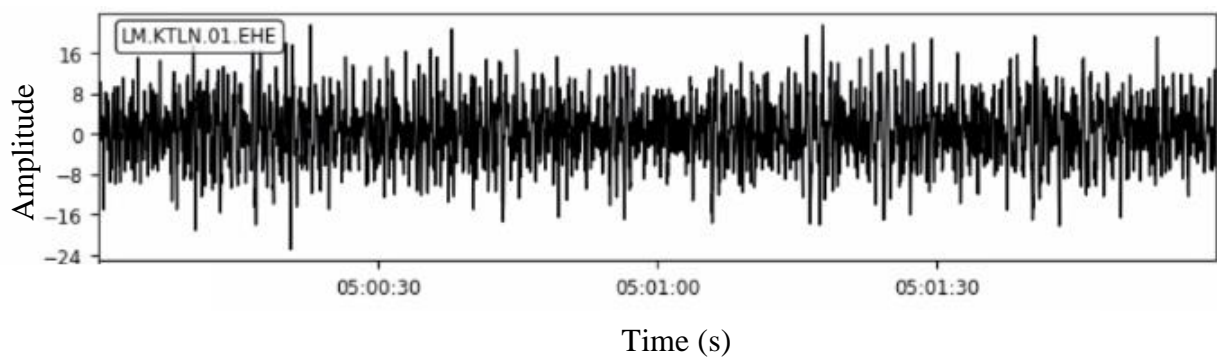
## Synthetic Seismogram Development and Variables Studied

The steps and information within the digitization process need to be accurate to allow reliable data processing. A series of tests were conducted that modified key variables in the digitization process where each change provided insight on how the frequency recovery in a digitized seismogram differed from the reference waveform. Analyzing these factors that influence the frequency recovery of a digitized seismogram allows researchers and scientists to better understand the digitization process and ultimately achieve accurate digitizations.

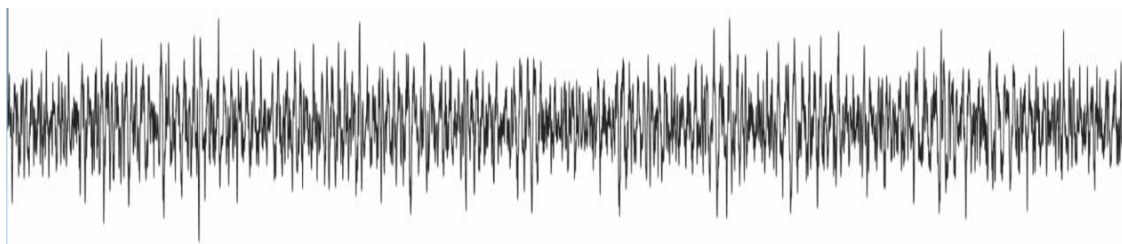
Synthetic seismograms were generated to examine each digitization variable independently. The synthetic seismograms were generated using a Python script developed by D. Burk where a white noise signal is created with a known frequency range in both the displacement and velocity spectrums. The program allows for the amplitude, trace thickness, and trace velocity to be changed. Currently, the code does not account for variations of trace thickness as a function of trace velocity. This is a future modification needed to mimic analog seismograms that occurs with some recording media. Generated waveforms were saved as a Miniseed file with an embedded network code, station name, location identifier, and channel imported in its header. The embedded seismic information allowed the comparison of multiple waveforms. An example of a white noise displacement signal is illustrated in Figure 17 with a known frequency response with the range of 1-12 Hz.

The Miniseed displacement signal was used as the reference signal. To simulate a scanned analog seismogram, the displacement signal was drawn on a blank image (i.e. an empty seismogram scan with no waveforms) then exported at 3000 DPI (see Figure 18). The reference signal did not go through the digitization process. *IrfanView*, a photographic processing software (Skiljan, 1996) was used to down-sample the original image into different image

resolutions for each digitization test. Each image was then digitized in *Wavetrack* using the process described earlier.



**Figure 17. Example of a synthetic seismogram generated with a Python script. This white noise signal contains known frequencies of 1-12 Hz.**

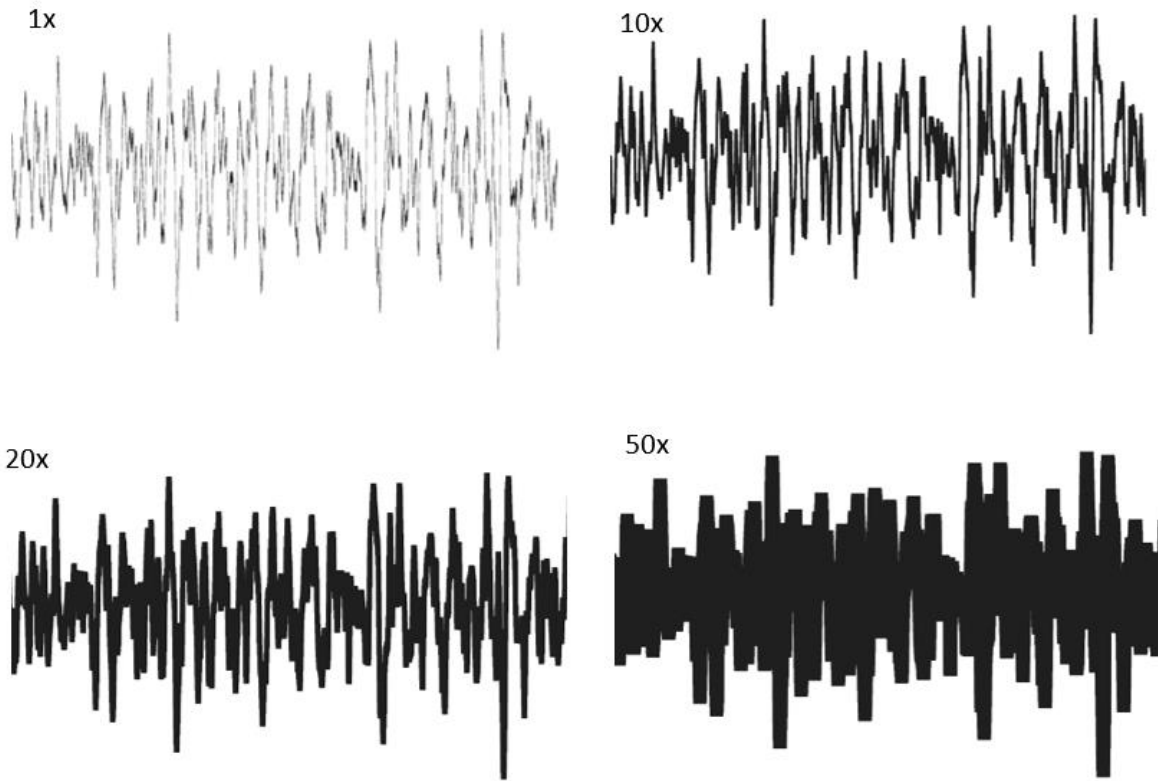


**Figure 18. Synthetic white noise seismic signal embedded on a blank image. This image can now be digitized in *Wavetrack*. This is the same waveform as in Figure 17.**

Four variables were examined in this thesis: 1) image resolution, 2) signal trace thickness, 3) waveform amplitude, and 4) Technician Variability. Three waveforms were created for first three variables to create an average result for each test. Analyzing technician variability (or experience) was conducted with a group of technicians digitizing a single waveform. For image resolution, the reference signal was modified to different image DPIs using *IrfanView*. Resolution extremes, both low and high, were chosen for this study to test the limits of seismogram digitization and its resulting frequency recovery. The impact of the image resolution on the frequency recovery of a digitized analog seismogram was studied by digitizing each seismogram and then comparing them back to the original Miniseed waveform that did not go through the digitization process. Waveforms were visually inspected for assessing frequency recovery in Power Spectral Density (PSD) graphs. These graphs show the distribution of energy as a function of frequency and are helpful in understanding which frequencies are strong or weak in a waveform (Cygnus Research International, n.d).

The second variable tested was the waveform thickness, where the thickness was modified to various widths. The original waveform thickness for the synthetic seismograms was 10 pixels wide, which is represented as 10x. Changing this number either higher or lower resulted in a shift in the waveform thickness. For example, a 50x waveform uses a 50-pixel wide waveform thickness. To keep some of the digitization variables constant, the image resolution was set to 600 DPI and the trace velocity was kept at a constant exposure rate throughout the seismogram. Exposure rate is closely related to the trace velocity and is the amount of time the recording media, for example, a light beam, has with the paper, tape, etc. which generates the waveform thickness on seismograms. More exposure yields a wider waveform thickness and relates to a slower trace velocity. A constant exposure rate was chosen to limit seismogram

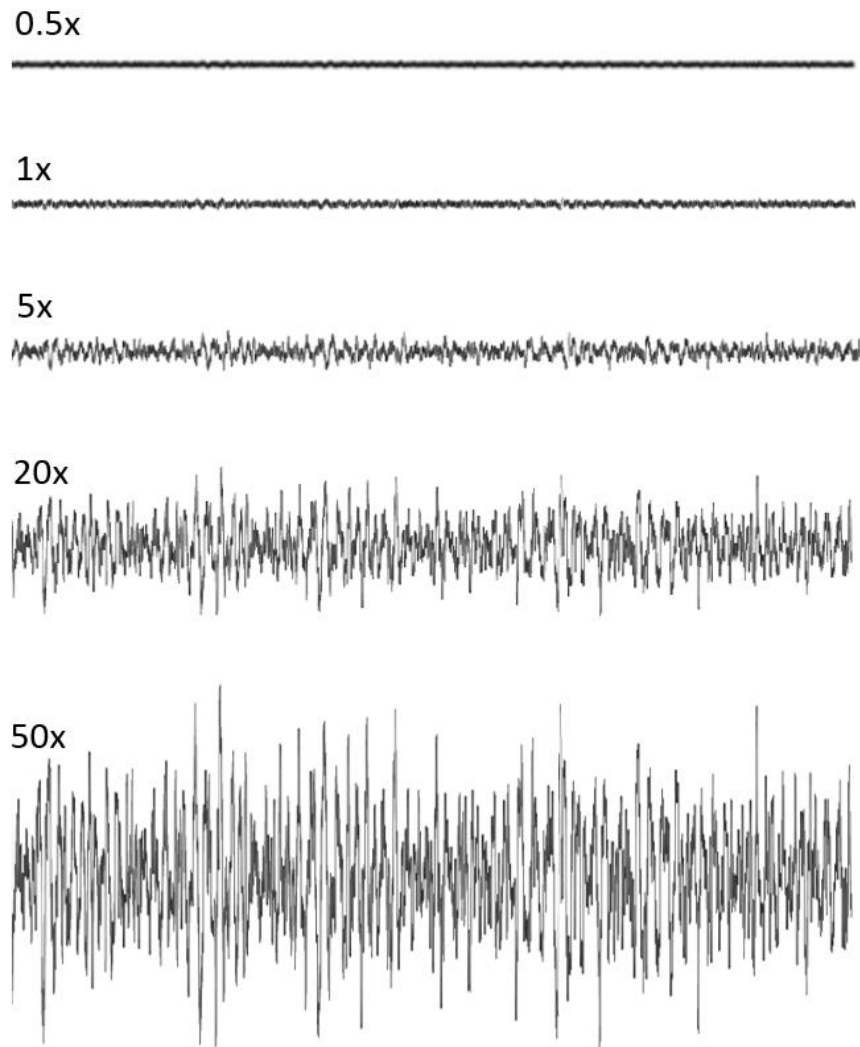
variables that influence seismogram digitization. Four waveforms were examined in the waveform thickness study. Figure 19 displays an example of a waveform for this test. Each waveform was generated four different times with various waveform thicknesses. Wider traces and high frequency signal have a high potential for overlap causing a reduction in signal recovery.



**Figure 19. Synthetic seismogram with varied waveform thicknesses. These seismograms are generated with a frequency range of 1 – 12 Hz at an image resolution of 600 DPI. The greater the trace thickness, the higher the probability for concealed high frequency signal.**



The third variable that was examined assessed how well a waveform's amplitude was recovered in the digitization process depending on the amplitude of the signal. An original waveform with a pixel variation of  $\pm 825$  pixels from a zero line was generated (waveform 1x in Figure 20). A multiplier was applied to this number which either magnified or compressed the amplitudes of the waveform. If a number less than one was used for the multiplier, the amplitudes were compressed; however, a number less than 0.5 saw severe amplitude compression and signal discretization where the pixel spread was only a few pixels wide. The 0.5x images were almost a straight line and were not used in this study. Examples of the amplitude modified waveforms are illustrated in Figure 20.



**Figure 20. Portion of a 600 DPI synthetic seismogram with varied amplitude heights with white noise as the signal in a frequency range of 1 - 12 Hz.**

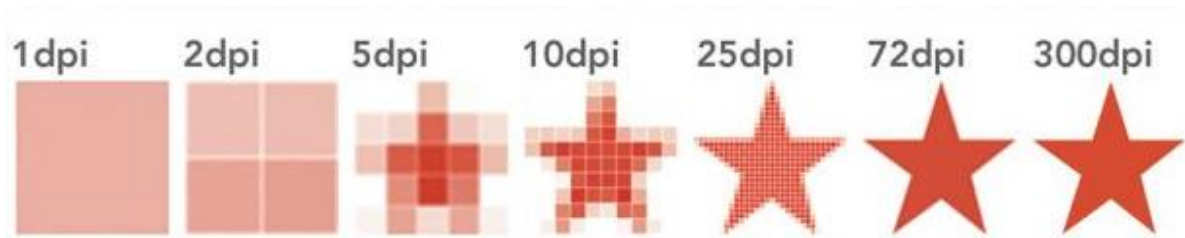
The digitization method was the same in all tests; however, different people performed the digitization. Some seismograms were digitized by the author; this created some known bias while digitizing as prior information was known about the individual tests and the reference waveforms. A ‘blind’ study was thus conducted to utilize over twenty technicians with various levels of digitization experience from independent organizations in Russia, Kazakhstan, and Kyrgyzstan. Each individual received one or two images for each test and were instructed to digitize the waveforms to the best of their ability. The individuals had no connection to each other, nor did they seek additional help with the digitization process. This ‘blind’ study group created a realistic situation where a research lab or institute digitizes analog seismograms. This group also established a way to quantify digitization experience (i.e., months digitizing and number of digitizations completed) and the quality (or frequency recovery) of a digitization. This is an important variable in digitizing seismograms as it introduces human influence and is another variable that is tested in this study. Referencing the digitization process below, the digitizations will be identified as either author digitized or blind digitized.

## DISCUSSION OF RESULTS

### Effects of DPI

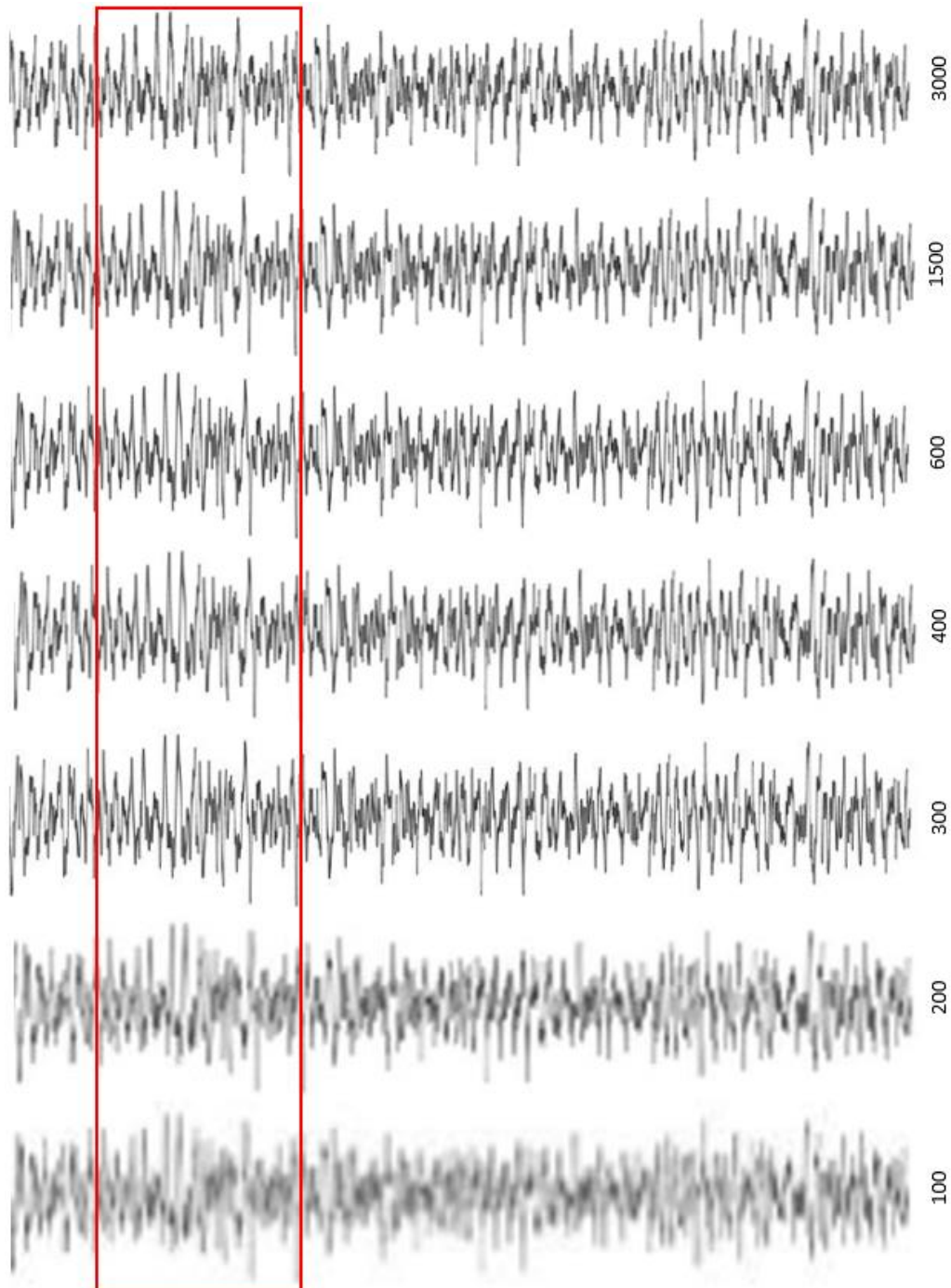
#### Image Resolution

To achieve good results, the copy/scan of the analog data to be digitized must be of good quality. One element to generate high quality digitized data is with a high-resolution image of the seismogram. An image with a higher DPI will have a higher pixel density and more detail of the original image retained, whereas a lower DPI image will feature a lower pixel density and retain less detail in the image. As the DPI decreases, the pixels become larger and coarser as they cover more area within the image, which ultimately decreases the confidence level of deciphering the contents of the image. To better visualize DPI uncertainty, consider a five-pointed star. As the individual pixels become coarser due to a decreasing DPI, the confidence in pinpointing each point in the star also decreases. Figure 21 illustrates how accurately identifying the five points of the star decrease as the DPI decreases.

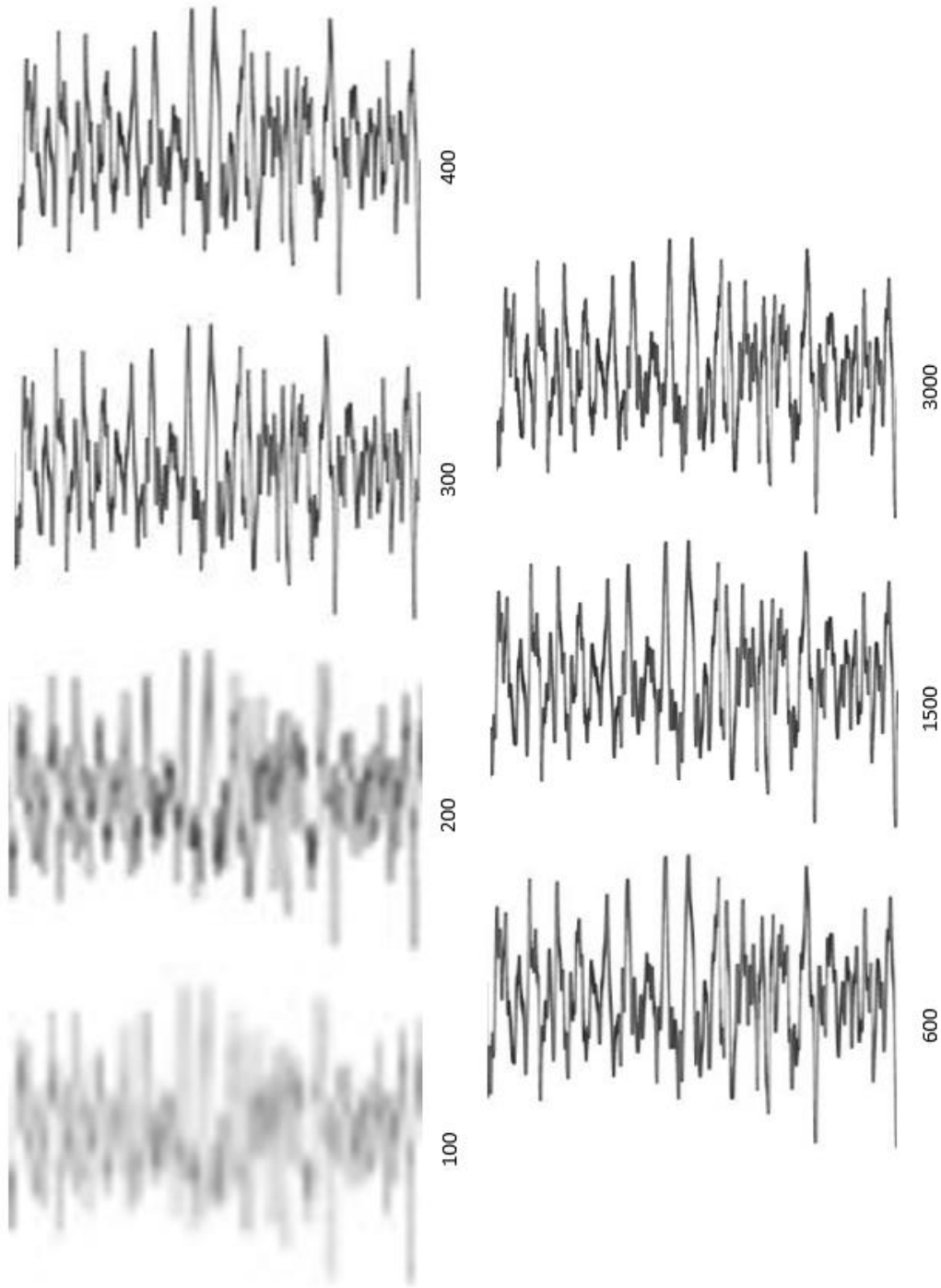


**Figure 21. Relationship between image resolution and image detail. Uncertainty in identifying the points of the star increase, like the click points of seismograms, as the image resolution decreases. The pixels become larger and coarser as they cover more area of the image. (Image modified from Toskey, 2018).**

Just as the star's five points are harder to identify as the DPI decreases, a similar result is noticeable in scanned seismograms. The image resolution significantly impacts the digitization quality, signal timing, and frequency response of a digitized waveform. In the digitization process, a user selects a peak or any point where there is a change of slope to recreate the signal. These points become harder to identify as the DPI decreases. Figure 22 illustrates how a scanned seismogram appears at several image resolutions that are common in modern scanners. Areas that have a change in slope within the black signal become harder to distinguish as the pixels become larger due to a lower image resolution and higher frequency signals become lost. The red box denotes a small area that is zoomed in in Figure 23. From afar, mid-range DPIs, like 300 and 400, may look reasonable to digitize. However, after zooming in closer in the image results in a 'fuzzier' picture which can raise difficulty identifying a slope change in the seismogram. This illustrates from an image perspective that lower and mid-range DPIs cannot accurately retain data for seismogram digitization.



**Figure 22. Relationship of image resolution and details within a seismogram. Scanned seismogram at various image resolutions, DPIs. As the image resolution decreases, the pixels within the image become larger and fuzzier which reduces the confidence in correctly identifying areas with a change of slope within the signal.**

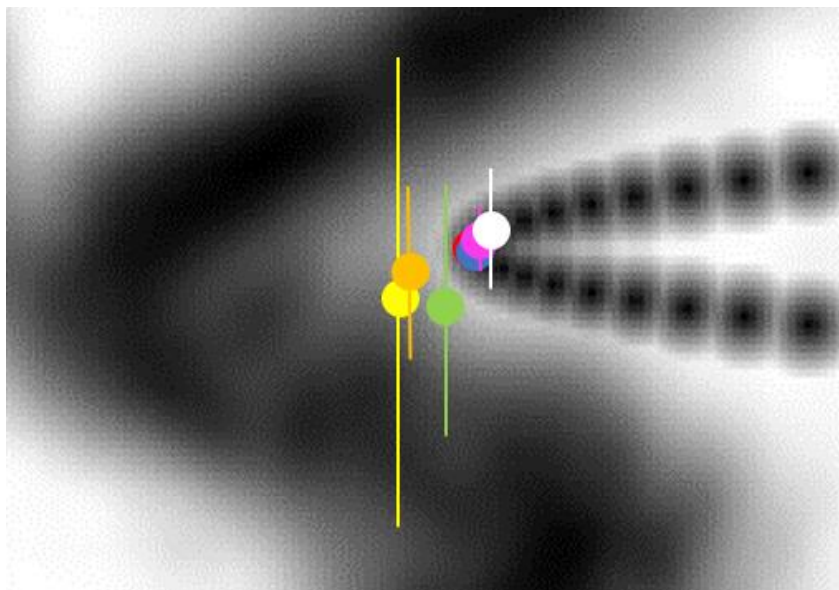


**Figure 23. Zoomed in section of a scanned seismogram. As the image resolution decreases, the pixels become larger which reduces the confidence in correctly identifying areas with a change in slope within the signal which is observed in the mid-range image resolutions like 300 and 400.**



Each point within the digitization is a coordinate representing time on the x-axis and amplitude on the y-axis. The width of the pixel represents time and as each pixel become larger due to lower image resolutions, the time per pixel also increases (shown in the table in Figure 24). A lower resolution image has the most time (seconds) per pixel,  $\Delta t$ . For example, a 100 DPI image of a seismogram that was recorded at 60 mm/minute has a  $\Delta t$  of 0.254 seconds/pixel which is a large uncertainty in recovering the time in the digitization compared to a 3000 DPI image with a  $\Delta t$  of 0.0085 seconds/pixel. Uncertainty will be higher with slower recording speeds (i.e., 120 mm/minute), or lower with faster recording speeds (i.e., 30 mm/minute). Low pixel densities force the click point selection of a signal peak to either be ahead of or behind the true peak which creates a time shift. A waveform superimposed with a click point at a peak, shown as colored dots, was picked at numerous DPI values, also shown in Figure 24. The dotted black line is the waveform of interest and is digitized while the adjacent black lines are different waveforms in the seismogram. The colored dots are shown for the peak on the dotted black line. Each color represents a DPI value, and the accompanying table lists the  $\Delta t$  values for each corresponding point. Color coded error bars show the range of each DPI's click point and for low DPIs, the magnitude of the variability is the highest. The error bars were determined by the pixel width for a given DPI. Click points are an intersection of two pixels and the error bar extent ranges from +/- one pixel intersection. By digitizing a seismogram at 100 DPI, a digitization's timing may be shifted by as much as +/- 0.25 seconds. In the end, this could affect seismological studies such as seismic phase picks used in geophysical analysis. Errors in the time position of peaks will also result in asymmetric waveforms that do not accurately represent the original seismogram. Digital analysis of the asymmetric waveforms will have incorrect frequency content, not filter properly, and generally result in larger errors. Since seismology, as

well as other scientific fields, requires accurately timed data, digitizations should be conducted only using higher DPIs to retain correct timing within the seismogram. Ultimately, having better timing in data will produce better geophysical studies.

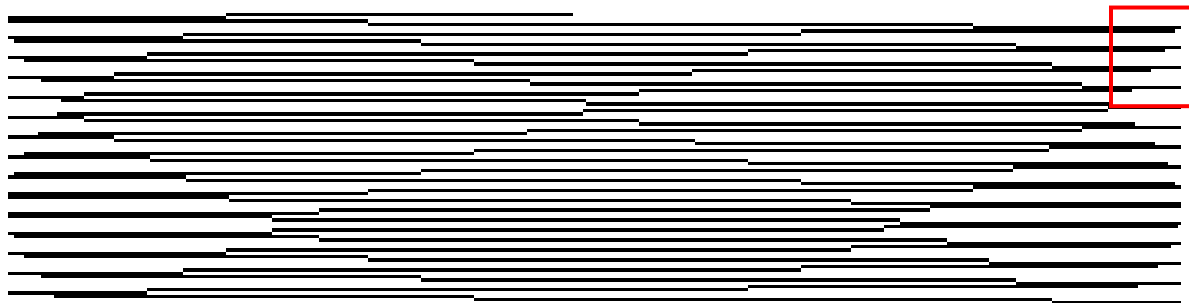


| DPI Value | Point Color | $\Delta t$ (seconds/pixel) |
|-----------|-------------|----------------------------|
| 100       | Yellow      | 0.254                      |
| 200       | Green       | 0.127                      |
| 300       | Orange      | 0.0846                     |
| 600       | White       | 0.0423                     |
| 1000      | Pink        | 0.0254                     |
| 1500      | Blue        | 0.0169                     |
| 3000      | Red         | 0.0085                     |

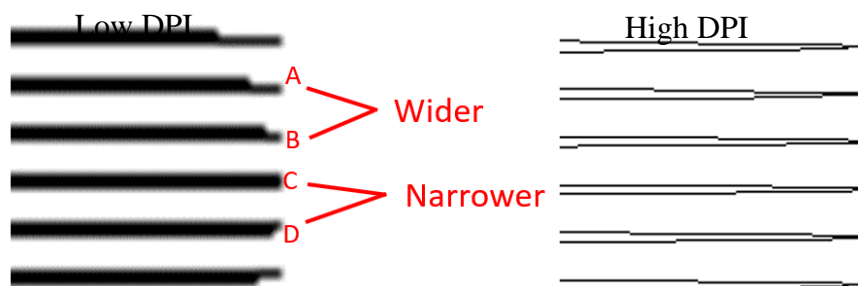
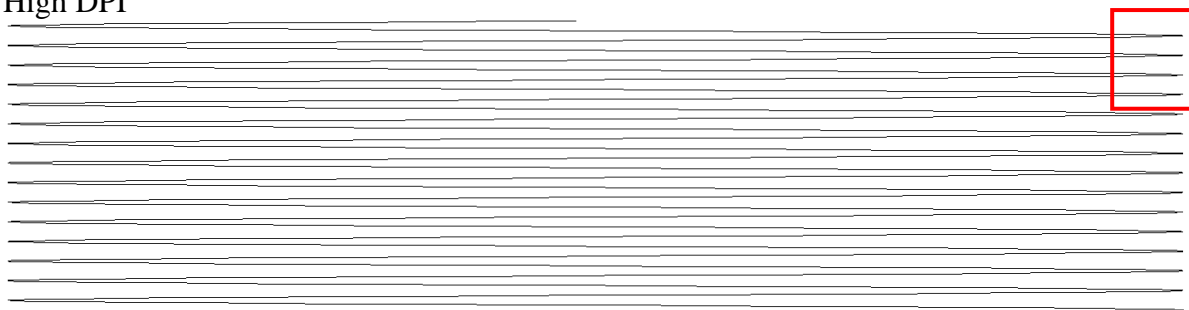
**Figure 24. Changes in time per pixel ( $\Delta t$ ) and Image Resolution (DPI) using an assumed recording velocity of 60 mm/minutes. Each DPI value is a different color for the peak location of the dotted black line. Its corresponding pixel  $\Delta t$  value is shown in the table. Error bars show the range of the click point location for the peak at a given DPI.**

Scan resolution also affects the frequency recovery of a signal. Low DPI values could force 'off peak' click point location selection within the digitization process. For example, if there is a combination of a thin trace thickness and a low scan resolution, the pixelization at the end of the line may join resulting in a distorted waveform. This causes a shift in the waveform peaks which transfers to distorted timing and overall shape of the waveform which then can result in erroneous frequencies. In Figure 25, a 2 Hz sine wave was generated at two different DPIs to illustrate how the waveform can become vertically distorted at high and low scan resolutions. This example demonstrates how the peaks become shifted if there is both a low DPI and a thin trace thickness (highlighted as red boxes in the figure). In the low DPI, the peak to peak spacing is variable compared to the higher DPI. There are sections that are wider and narrower due to where the pixels lie in the matrix in the low DPI image. If these peaks are selected during the digitization process, the resulting digitized waveform may have asymmetric signals and may recover incorrect frequencies from the original symmetric signal.

Low DPI



High DPI



**Figure 25. Spacing between the waveform peaks is shifted if there is a combination of low image resolution and a thin line thickness. Time is vertical. The red box highlights an area (shown below) of shifted pixels at the waveform peaks which create uneven spacing between the peaks.**

## Theoretical Sine Test

This test established the theoretical frequency recovery of a signal at a given DPI. Synthetic sine waves were created at different DPI values and digitized in *Wavetrack*. A ‘perfect scenario’ was formed for digitizing as the synthetic waves limited some of the digitization variables such as a uniform amplitude and trace velocity. Each synthetic sine wave was generated from the product of the time interval (0.01 seconds), pi ( $\pi$ ), and a multiplier. The base frequency of the sine wave was 0.5 Hz, the multiplier allowed the wave to easily transform into different frequencies. For example, using a multiplier of 2, the function would produce a 1 Hz sine wave. Each wave was exported at a specific DPI then digitized by the author in *Wavetrack*. Signal DPIs ranged from 72 DPI to 3000 DPI and a signal’s frequency ranged from 0.5 Hz to 12 Hz.

The digitized waveforms were examined for shape retention and frequency recovery and grouped into three categories: recoverable, recoverable but distorted, and not recoverable. The recoverable waveforms had no asymmetric signals and fully recovered the frequency. Recoverable but distorted waveforms were determined by noticing some asymmetry in the waveforms, but they were able to be digitized. Lastly the not recoverable waveforms were signals that could not recover the original frequency. This may be decided during the digitization process where the technician physically could not identify the signal from the background noise or if the digitized signal did not recover the original frequency. Categories are displayed in a matrix with respect to DPI in Table 1. This table also displays common seismometer instrument recording speeds.

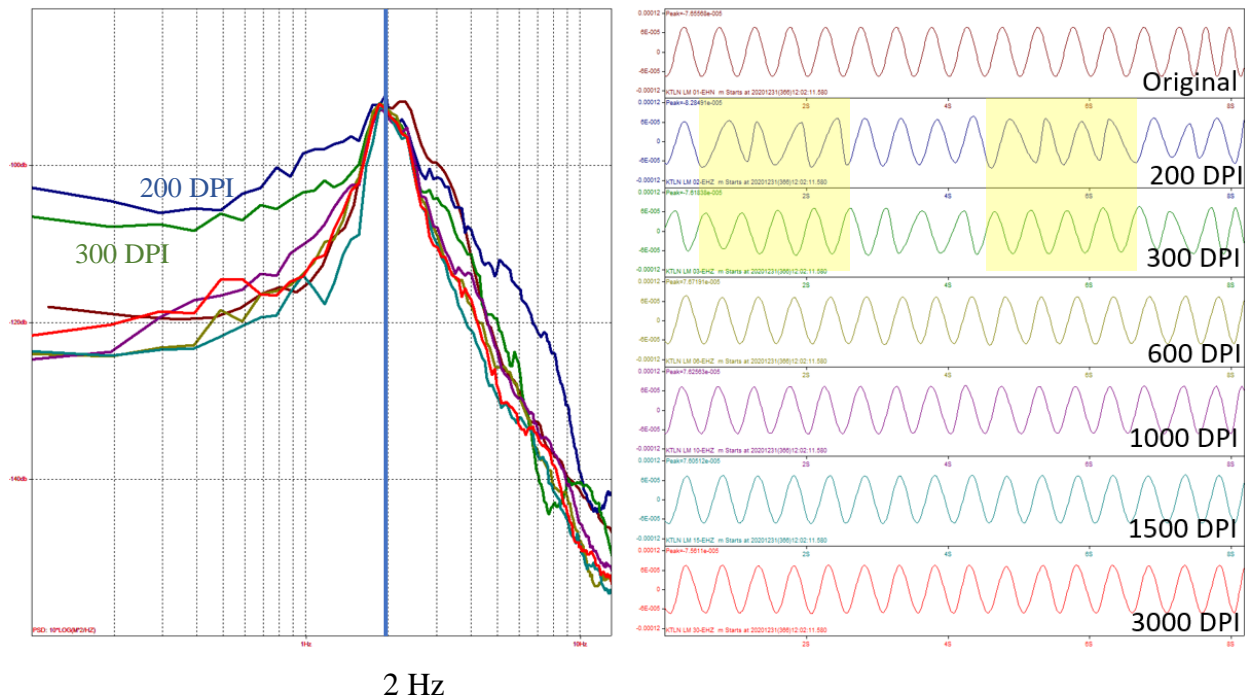
An example of a 2 Hz sine with numerous image resolutions are shown in Figure 26. An ideal 2 Hz signal would have a sharp peak at the 2 Hz line on the PSD graph (bolded in blue

in Figure 26) where the x-axis describes the frequency, and the y-axis describes decibels. Scans at various DPI values were digitized and compared against the ideal waveform to see their frequency recovery. If a signal shifted away from the ideal peak and has a broader crown in the PSD graph, that demonstrates that the digitization has additional noise in the system thus returning a non-pure 2 Hz frequency. A possible reason for added noise is due to asymmetric waves in the digitization. Distorted waveforms in the 200 and 300 DPI digitizations are highlighted as yellow boxes in Figure 26. The asymmetry in the digitization can also be seen in the PSD graph where the blue and green lines, correlating to the 200 and 300 DPI digitizations, have broader peaks around the 2 Hz line in the graph. Relating back to the theoretical frequency recovery matrix, the 200 and 300 DPI would be yellow and anything above 300 would be deemed green.

|                  | 30mm/<br>minute | 120mm/<br>minute | 60mm/<br>minute |     |      |     |      |     |     |     |     |     |     |     |     |     |     |      |      |      |
|------------------|-----------------|------------------|-----------------|-----|------|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|
| Resolution (DPI) | 6000            | 1500             | 3000            |     |      |     |      |     |     |     |     |     |     |     |     |     |     |      |      |      |
|                  | 3000            | 750              | 1500            |     |      |     |      |     |     |     |     |     |     |     |     |     |     |      |      |      |
|                  | 1200            | 300              | 600             |     |      |     |      |     |     |     |     |     |     |     |     |     |     |      |      |      |
|                  | 600             | 150              | 300             |     |      |     |      |     |     |     |     |     |     |     |     |     |     |      |      |      |
|                  | 400             | 100              | 200             |     |      |     |      |     |     |     |     |     |     |     |     |     |     |      |      |      |
|                  | 300             | 75               | 150             |     |      |     |      |     |     |     |     |     |     |     |     |     |     |      |      |      |
|                  | 200             | 50               | 100             |     |      |     |      |     |     |     |     |     |     |     |     |     |     |      |      |      |
|                  | 144             | 36               | 72              |     |      |     |      |     |     |     |     |     |     |     |     |     |     |      |      |      |
|                  |                 |                  |                 | 0.5 | 0.75 | 1.0 | 1.25 | 1.5 | 2.0 | 2.5 | 3.0 | 4.0 | 5.0 | 6.0 | 7.0 | 8.0 | 9.0 | 10.0 | 11.0 | 12.0 |
|                  |                 |                  | Frequency (Hz)  |     |      |     |      |     |     |     |     |     |     |     |     |     |     |      |      |      |

**Table 1. Theoretical recoverable frequencies at a specified DPI for different recording speeds. Green is defined as recoverable frequency, yellow describes recoverable but with distortion, and red denotes not recoverable.**



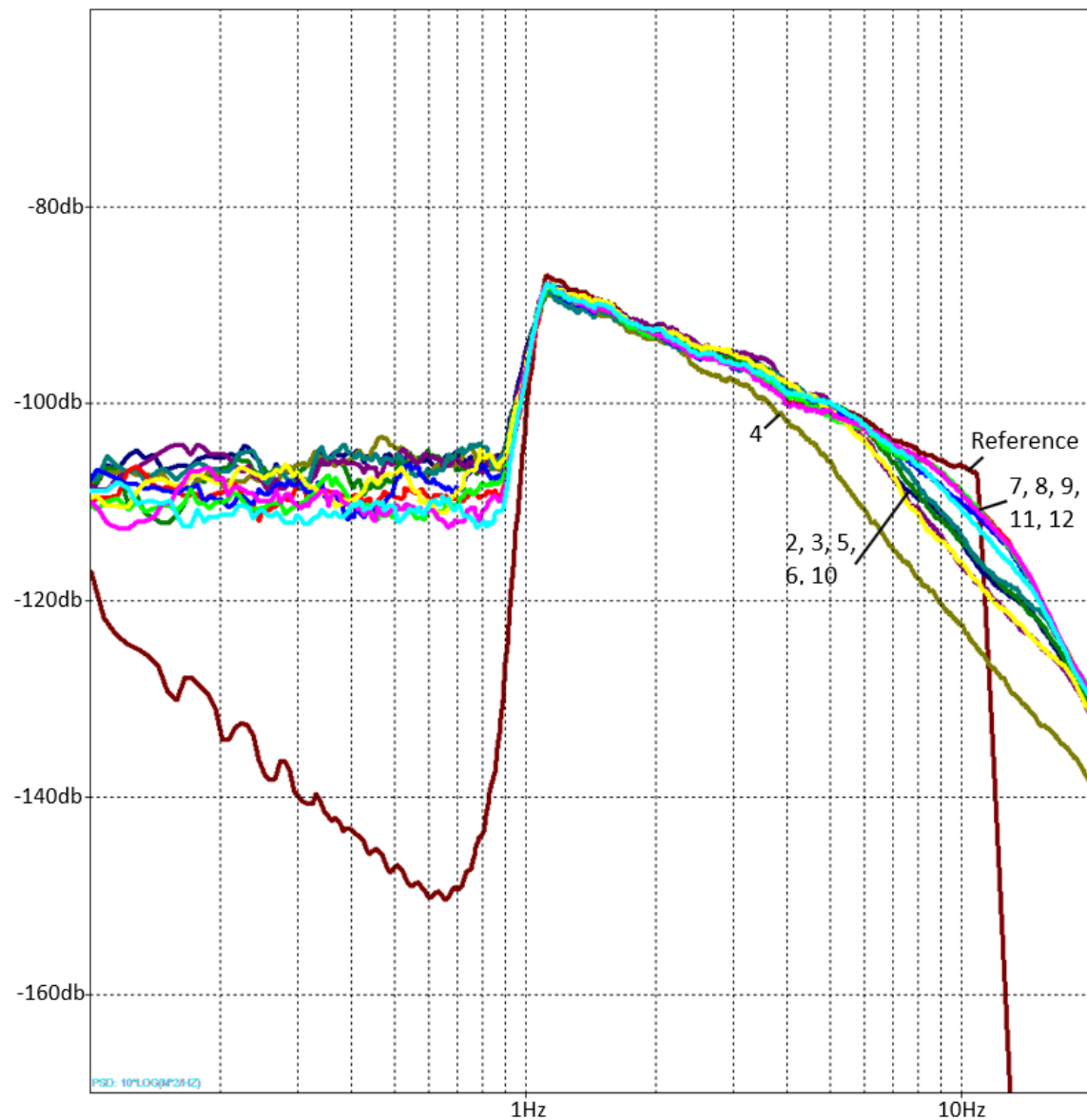


**Figure 26. Degradation of 2 Hz sine wave at various image resolutions for a recording speed of 60mm/minute. Digitizations are compared against a reference waveform for their appearance and their frequency response. Yellow regions highlight asymmetry in the digitizations due to low pixel densities. The Power Spectral Density (PSD) graph on the left shows the frequency recovery of the waveforms.**

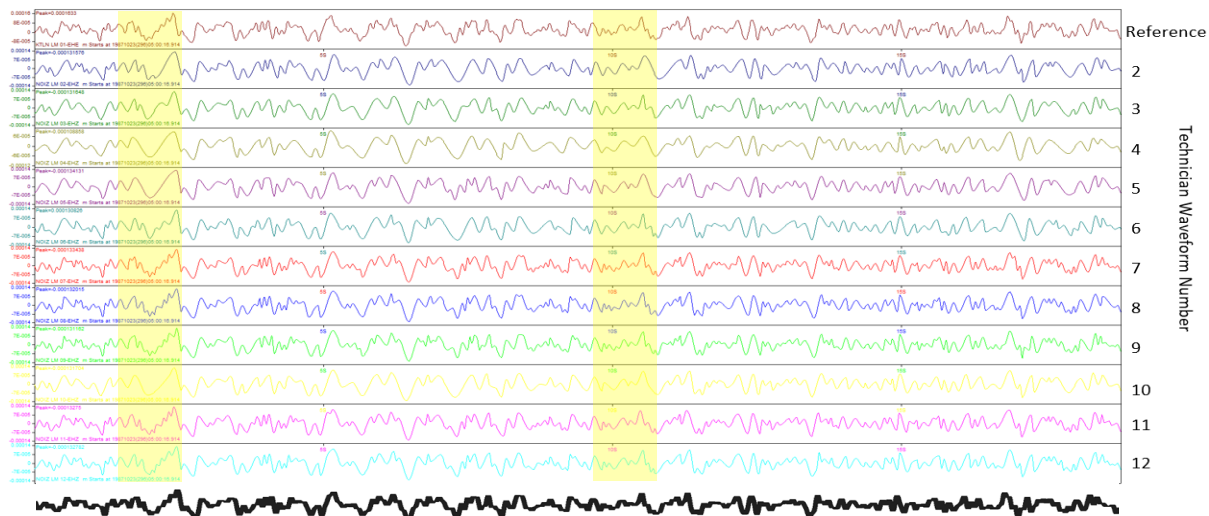
## Technician Variability

Kemerait et al. (1981) claimed that the inconsistencies of a user's digitization experience significantly impact the subsequent digitization. This is an almost unspoken variable that needs to be taken into consideration while digitizing analog seismograms. To quantify the variability in digitizations due to operator, a separate mini study examined eleven technicians and their ability to duplicate a 1 – 12 Hz 600 DPI waveform. This was a blind study where each technician independently digitized a waveform. Each waveform was assigned a number and compared against one another and a reference waveform.

From the PSD graph in Figure 27, there are three apparent groupings of technicians: needs revision, average, and excels. These groupings were based off the maximum recoverable frequencies determined in Figure 34 and explained in a later section. Technician 4 is in the 'needs revision' grouping, as they recovered between 3 – 4 Hz, which is less than 50% of the expected frequency recovery for a 1 – 12 Hz 600 DPI seismogram. Supplemental training and revision of the digitization could improve technician 4's future digitizations. The 'average' grouping contains technicians 2, 3, 5, 6, and 10, all of whom recovered up to 6 – 8 Hz. These technicians recovered what is expected for a 600 DPI seismogram. Lastly the 'excels' group has technicians 7, 8, 9, 11, and 12, as they recover frequencies up to 8 – 9 Hz. This grouping surpassed the expectations of what we expect a 600 DPI image should recover from a 1 – 12 Hz signal. Figure 28 highlights the digitizations in the time domain. Within the yellow regions, the amount of detail in the waveforms are apparent between the technicians which ultimately relate back to the recoverable frequency in the previous figure. For example, technicians 2 and 4 have less detail compared to technicians 7 and 11.



**Figure 27. PSD graph illustrating the eleven technician's digitizations of a 600 DPI 1 – 12 Hz waveform. Each technician's waveform was assigned a number and compared against one another and the reference. There are three groupings based on the frequency recovery of digitizations: needs revision (technician 4, recovers 3-4 Hz), average (technicians 2, 3, 5, 6, 10 and recovers 6-8 Hz), and excels (technicians 7, 8, 9, 11, and 12 recovers 8-9 Hz).**



**Figure 28. Comparison of a small section of a 1 – 12 Hz 600 DPI waveform that was digitized by eleven technicians. Each technician was assigned a number and compared against the reference. The yellow highlighted regions especially show variations of digitization detail. The waveform that the technicians digitized is shown below the waveforms.**

The eleven technicians had various level of experience (ranging from 0.5 years to 1.5 years) and have completed different quantities of digitizations (ranging from 10s to 100s). This grouping of technicians is very experienced in seismogram digitization so there is no correlation in this group of experience level and quality of digitization. One of the technicians who has one of the highest levels of experience and completed digitizations fell into the ‘average’ category while another technician who had a low level of experience exceeded the expected frequency recovery of 600 DPI seismograms. This may describe a technician’s ability to understand seismograms. It could be that a technician simply does not see some of the superimposed high frequency signals overlain on the lower frequency signals. Technicians need to grasp the nature of waveform mechanics and the influence of ground motion and how it translates to a seismogram. This understanding combined with adequate training will result in better seismogram digitization.

The time taken to digitize a waveform (or speed) was a factor that was not collected or examined in this study. A technician may have completed the task quickly and may have missed important information while others took a slower, more methodical, approach. This may have an influence on the recoverability of a waveform’s frequency content. Referring to Kemerait et al. (1981) statement again, they mention how good digitizations stem from a user’s digitization experience. While this statement is true and is observed this in this study, a better definition for a user/technician’s experience should be a mixture of involvement (i.e., number of digitizations and time overall digitizing), ability to understand waveforms, and the time it takes to digitize a single waveform all have an influence on the resulting digitization.

The variability in digitization quality between different technicians is important in this study. To account for technician variations in the following sections, each individual test is

evaluated using multiple independent and blind digitizations. The independent digitization waveforms are summed for an overall estimation of frequency recovery versus the variable under each test.

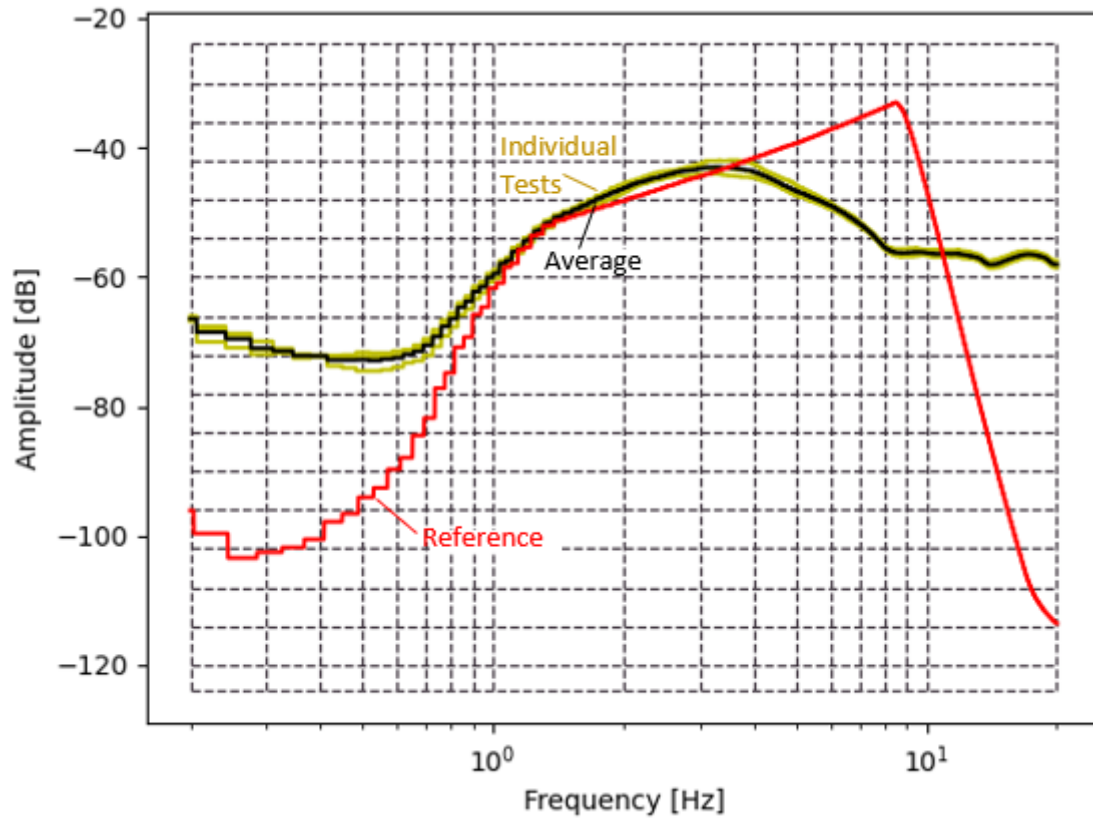
### **Image Resolution Test**

To better simulate analog seismograms, synthetic ‘white noise’ signals were generated with a known frequency range. With the combination of both a known frequency and a reference signal, it made it easier to determine if a specific waveform digitized at a certain DPI could recover the maximum frequency in a waveform. For this test, five image resolutions were used to evaluate the limitations of scan resolution and frequency recovery. The waveforms studied in this section assume a recording speed of 60 mm/minute. Faster recording speeds, such as 120 mm/minute, with the same DPI will have a better recovery because there will be more pixels per waveform cycle. Slower recording speeds, such as 30 mm/minute, will have poorer recovery because there will be fewer pixels per waveform cycle.

Three white noise signals containing frequencies of 1 – 12 Hz were generated at different image resolutions between 200 and 3000 DPI and were blind digitized from different institutions in *Wavetrack*. An example of one of the waveforms is shown in Figure 23. Each waveform was digitized independently by different institutions for statistical control. PSD graphs in Figures 29 through 33 illustrate the variability of each DPI and each technician’s ability to recover the waveform. The red line signifies the reference waveform, the three yellow lines denote individual digitizations, and the black line is the average of all three digitizations to create an overall estimate for the specific variable tested. For example, Figure 31 shows the frequency recovery for digitizations completed at 600 DPI. The individual tests for the 600 DPI trial indicate good recovery to 7 – 9 Hz, with the average around 8 Hz.

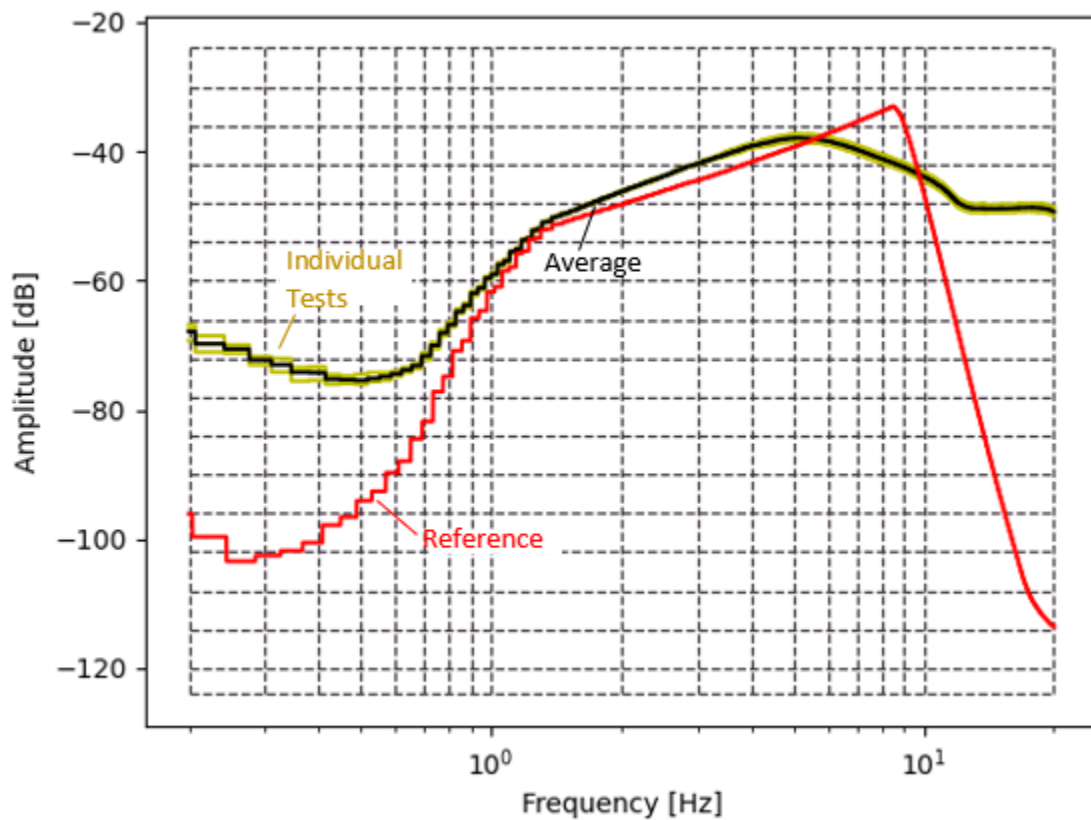
Each of the compiled waveforms at each DPI were combined in a final figure shown in Figure 34. For each DPI, a different color is used with the reference waveform shown as red. There are color-coded labels indicating the maximum recoverable frequency for each image resolution. The maximum recoverable frequency was determined whenever a line on the graph takes a sharp decline that suggests that that digitization is no longer recovering that frequency content. It can be expected that low image resolutions limit the ability to accurately choose points in the digitization process, thus low image resolutions will yield a poorer frequency recovery. For example, the maximum frequency for a 200 DPI image is between 3 – 4 Hz as shown in the PSD in Figure 34. This level of frequency recovery is insufficient for many quantitative results that could be done with digitized seismograms. Raising the image resolution to 300 DPI yields a frequency recovery between 5 -6 Hz and increasing the DPI even further to 600 DPI, returns up to approximately 7 – 8 Hz. The only waveforms from which the full range of 12 Hz were recovered were the two highest image resolutions: 1500 and 3000 DPI. There is no meaningful difference in the results between the 1500 and 3000 DPI trials. This indicates that digitizing a seismogram scanned at greater than 1500 DPI has no additional benefit.

It is suggested that using a minimum of 600 DPI for seismograms with a recording speed of 60 mm/minute will ensure sufficient frequency recover for most geophysical analysis. For circumstances that require a higher image resolution (i.e., the definition of the signal at 600 DPI still makes it difficult to accurately identify peaks), a recommendation can be made for scanning the seismogram at 1500 DPI if data storage space permits.

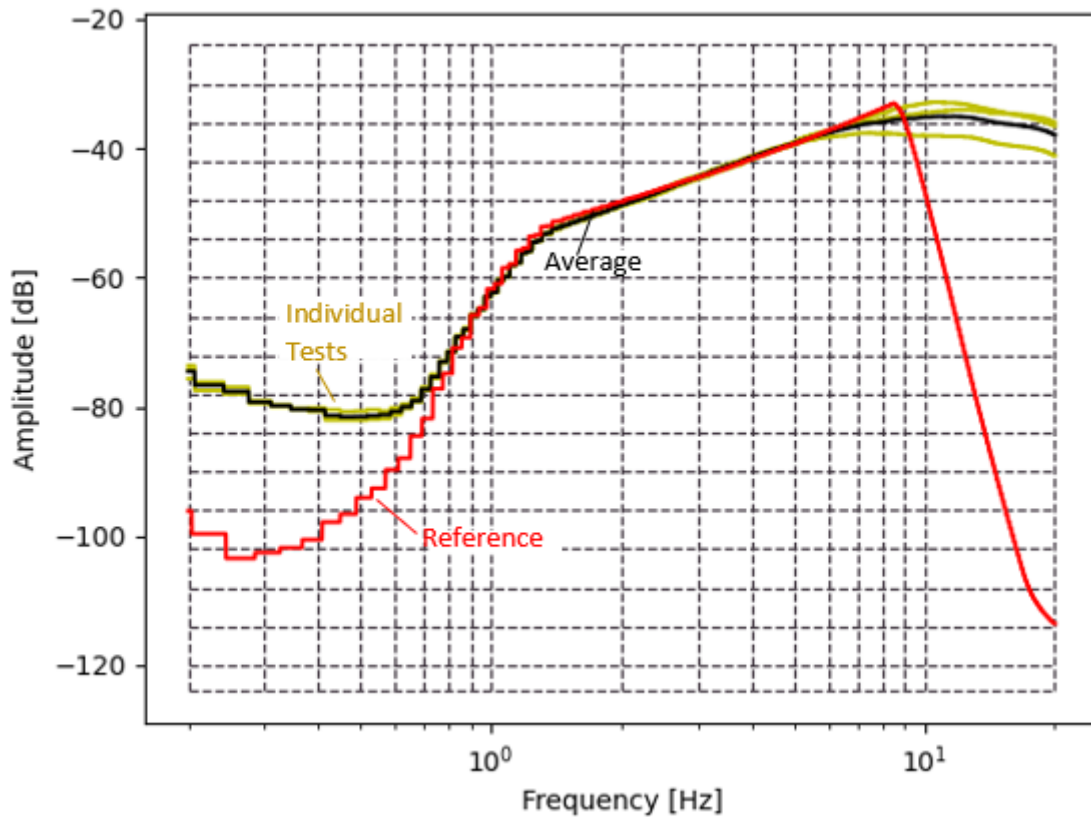


**Figure 29. PSD graph displaying the frequency recovery for a 200 DPI image for a signal with a frequency range of 1 – 12 Hz. The reference signal is in red, the yellow lines are the individual digitizations completed by independent technicians, and the black line is combination of the three yellow individual digitizations to make an average for 200 DPI.**

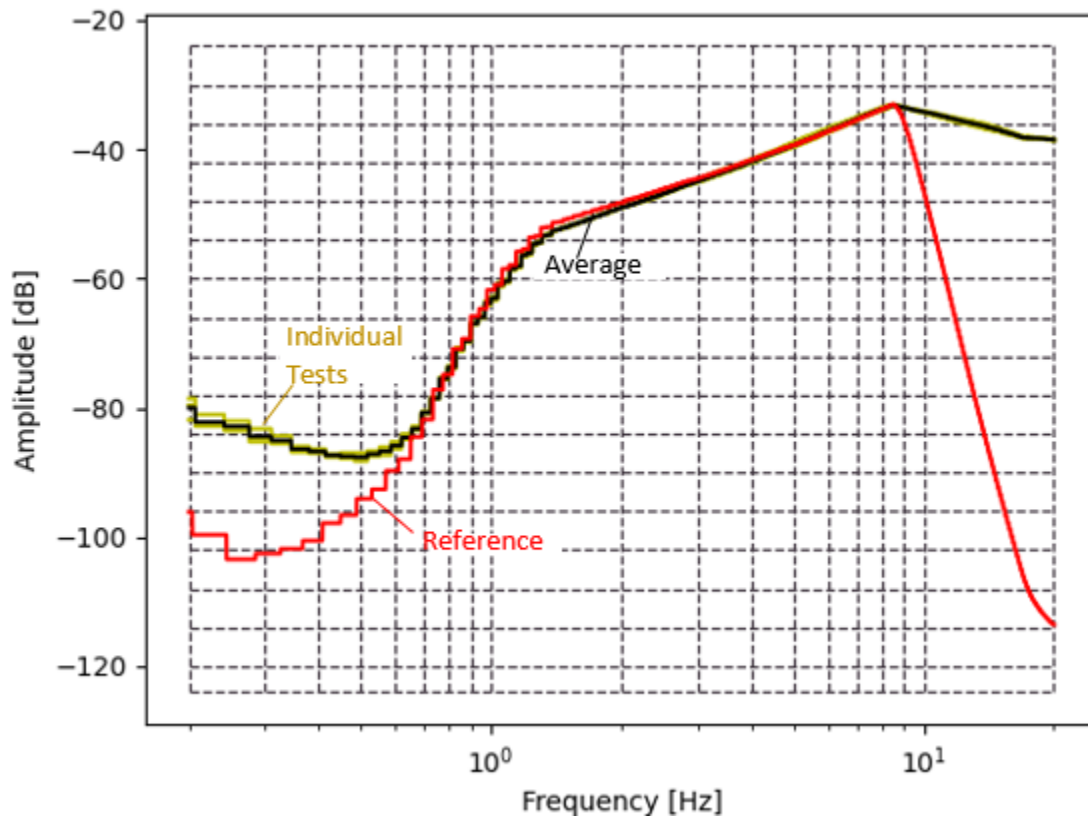




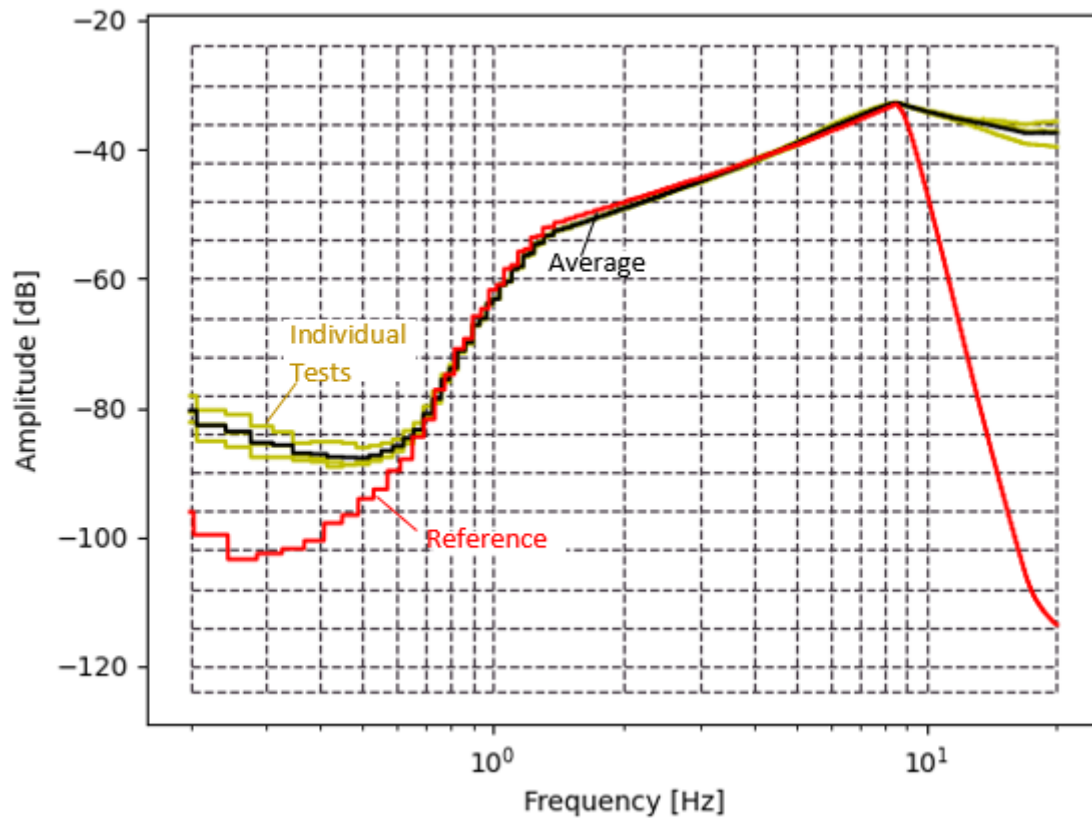
**Figure 30. PSD graph displaying the frequency recovery for a 300 DPI image for a signal with a frequency range of 1 – 12 Hz. The reference signal is in red, the yellow lines are the individual digitizations completed by independent technicians, and the black line is combination of the three yellow individual digitizations to make an average for 300 DPI.**



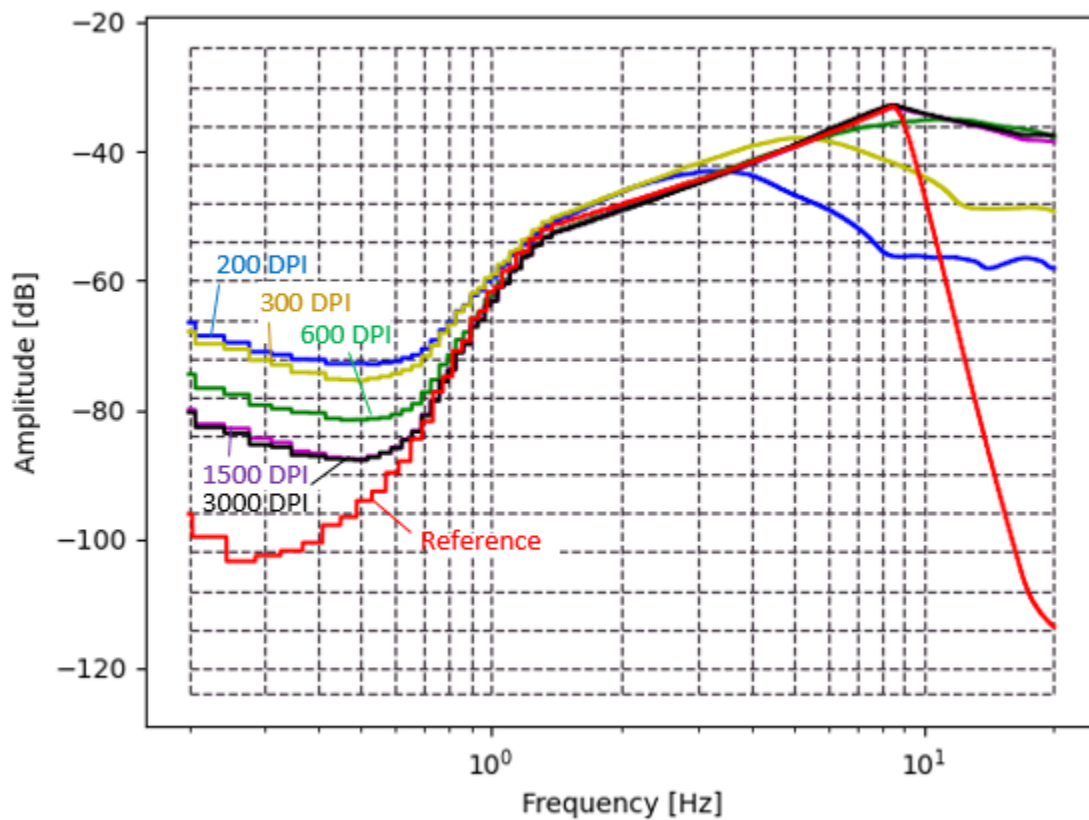
**Figure 31. Power Spectral Density (PSD) graph displaying the frequency recovery for a 600 DPI image for a signal with a frequency range of 1 – 12 Hz. The reference signal is in red, the yellow lines are the individual waveforms digitized by independent technicians, and the black line is combination of the three yellow individual digitizations to make an average for 600 DPI.**



**Figure 32. Power Spectral Density (PSD) graph displaying the frequency recovery for a 1500 DPI image for a signal with a frequency range of 1 – 12 Hz. The reference signal is in red, the yellow lines are the individual waveforms digitized by independent technicians, and the black line is combination of the three yellow individual waveforms to make an average for 1500 DPI.**



**Figure 33. Power Spectral Density (PSD) graph displaying the frequency recovery for a 3000 DPI image for a signal with a frequency range of 1 – 12 Hz. The reference signal is in red, the yellow lines are the individual waveforms digitized by independent technicians, and the black line is combination of the three yellow individual waveforms to make an average for 3000 DPI.**

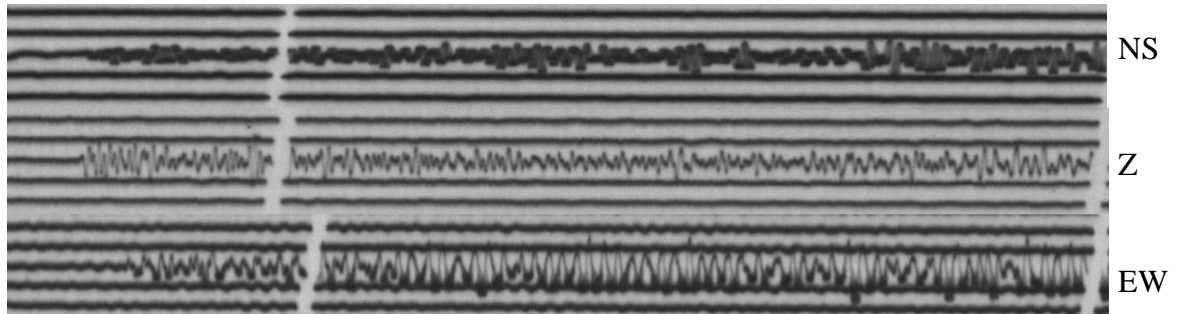


**Figure 34. Power Spectral Density (PSD) graph illustrating all the compiled waveforms for each image resolution against a 1 – 12 Hz reference signal. Each colored line is the average for a given DPI.**

## Waveform Thickness Test

Waveform thickness can be defined as the width of the seismic trace on a seismogram. This may vary for each seismic station and component. Seismic signals can appear thicker for many reasons that differ among recording media. On photopaper, the light beam can be out of focus causing the beam to appear fuzzy, or there can be a very high frequency (such as 50 or 60 Hz) line noise signal overprinted onto the seismic signal. The very fast oscillation of the galvanometer essentially fattens the seismic trace at all points. This can cause a ‘buffer’ surrounding the incoming ground motion data. The velocity of the recording pen (or light beam, stylus, etc. depending on the recording methodology) also influences the waveform thickness. The recording pen has less contact with the paper if it has a fast velocity. This results in a light and potentially thinner trace. The recording pen has more contact with the paper if the incoming ground motion data has a slower velocity but also in peaks and troughs where there is a change of slope. Figure 35 presents an example of a photopaper seismogram and how each component has a different waveform thickness. With ink recording, the nub of the pen may be worn down or a bit of fuzz could be stuck on the pen, both which will fatten the trace. A partially clogged pen may make a thinner than normal trace readable at low amplitudes but invisible at high amplitudes. A worn heat pen also produces a fatter than normal trace on thermal paper as it is in contact with the paper over a larger area. The waveform thickness on heat pens is also affected by the temperature control of the pen. If the pen is too hot, the trace will be much thicker, while if too low, it will be thin and faint. Additionally, if the thermal paper is not installed correctly and has areas that are not in contact with the underlying drum, waveforms will be thicker. The underlying drum acts as a heat sink that pulls heat away from the paper. If the paper is not in contact with the drum, the heat stays more in the paper making a locally wider waveform. While

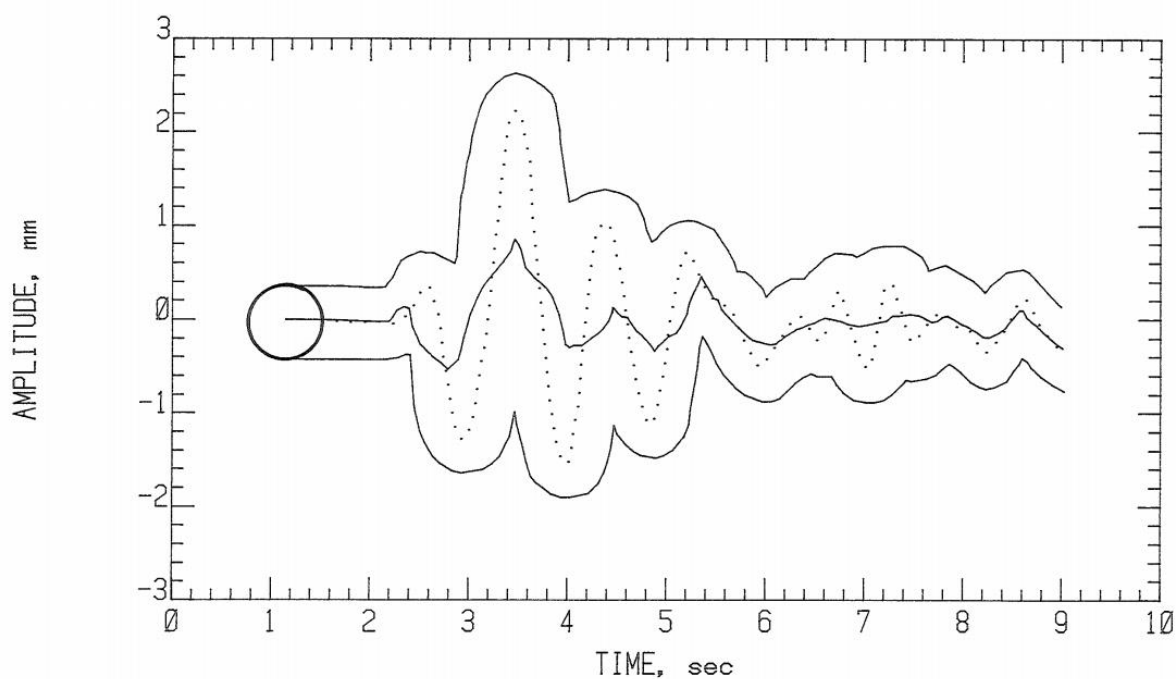
digitizing, it is increasingly difficult with wider waveforms to notice and detect seismic signal within this 'buffer' because seismic signals may be hidden.



**Figure 35. Analog photopaper seismogram displaying variations in signal waveform thicknesses. This seismogram exemplifies how variable the line thickness is due to the beam focus and changes in trace velocity. Each group of traces is a different component from the same seismogram.**



Chiburis et al. (1980) documented these situations while digitizing their WWSSN seismograms and noticed that waveform width is a major issue in the digitization process. They explain how an averaging effect takes place in a digitization where a program, or technician, follows the true center of the waveform. But by following the absolute center, the digitizer may miss points such as the true accurate peak and trough of the signal hidden in a wider waveform. Figure 36 illustrates this concept where the circle on the far left denotes a galvanometer light beam, the dotted line depicts the true signal, the two outer solid lines represent the waveform extent, and the single solid line in the middle illustrates the center averaged digitized signal. The extrema of the amplitude peaks are averaged due to a wide waveform width. This can guarantee that the digitized line will not represent the true seismic information, which is problematic because it does not capture the true amplitude and nature of the waveform. With an unrealistic waveform produced from the averaging effect, it can be assumed that frequency response of the waveform is also not faithful.



**Figure 36. Trace width and the digitizing 'averaging' effect. Circle on the left denotes the galvanometer light width, the dotted line illustrates the true signal, the outer black lines show the extent of the amplitude, and the black center line in the digitized waveform from computing the average of the amplitude extrema. (Figure from Chiburis et al., 1980).**

To examine how the waveform thickness affects the digitization process and the overall frequency response, three synthetic seismograms were created with a frequency range of 1 – 12 Hz and a 600 DPI image resolution. The process of altering the waveforms was described previously in the Methodology section with an example of one of the signals shown in Figure 19. These digitizations were completed as a blind digitization test and a compiled digitization was created for each waveform thickness.

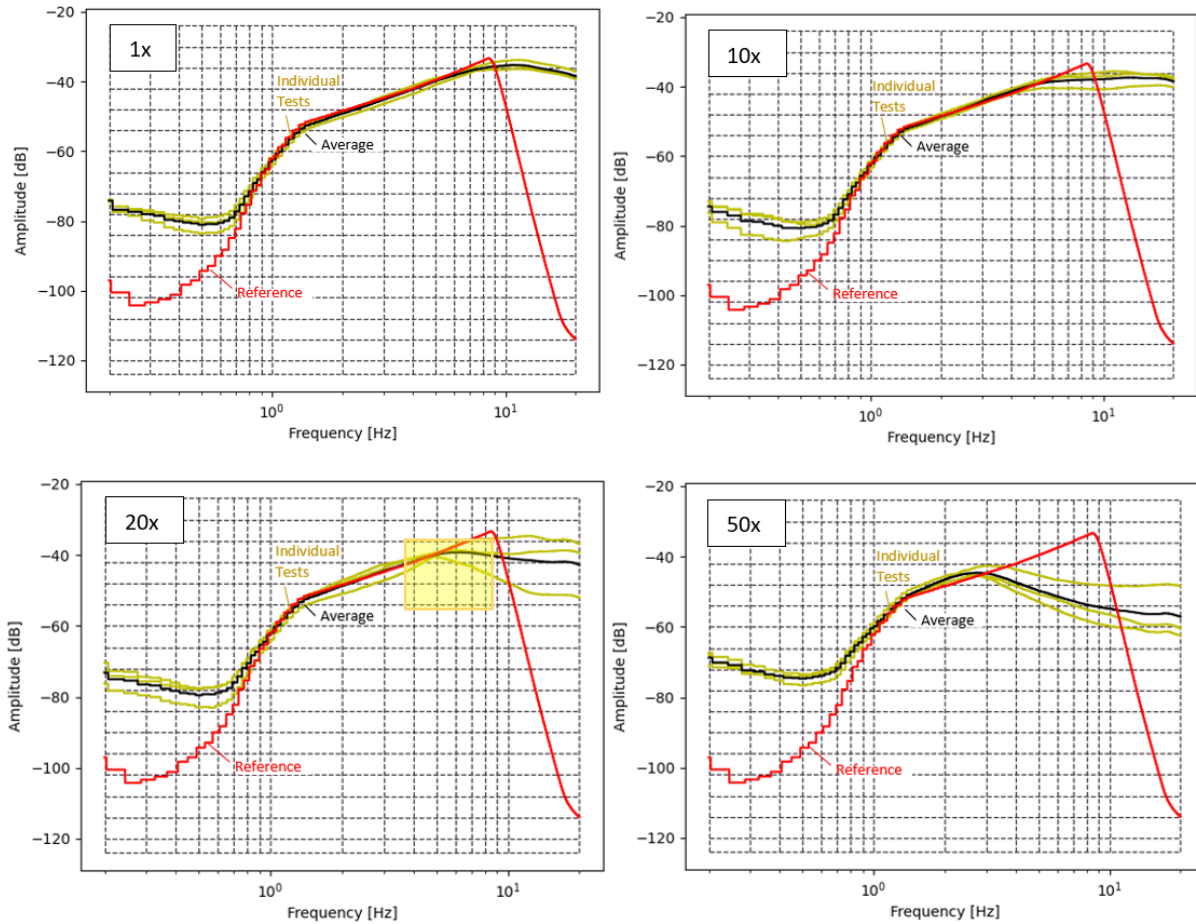
As the waveform becomes fatter, the point where the change of slope occurs becomes harder to identify in the seismic signal, especially for high frequency signals. Most waveforms are darker in color which increases the difficulty of identifying these points as they tend to blend, making them indistinguishable from adjacent waveforms. It is expected that as the waveform thickness becomes wider, the digitization quality will suffer and will not recover high frequency signals as well.

The PSD graphs in Figure 37 illustrate the separate waveform thicknesses. Within each graph, the reference signal is red, the three individual digitizations are yellow, and the compiled digitizations used as an estimation of the individual waveform thickness is black. The 1x and 10x waveforms show little variances between the technician's digitizations with an approximate frequency recovery of around 7 Hz. Both thicknesses are narrow enough to easily identify peaks and areas of a change of slope.

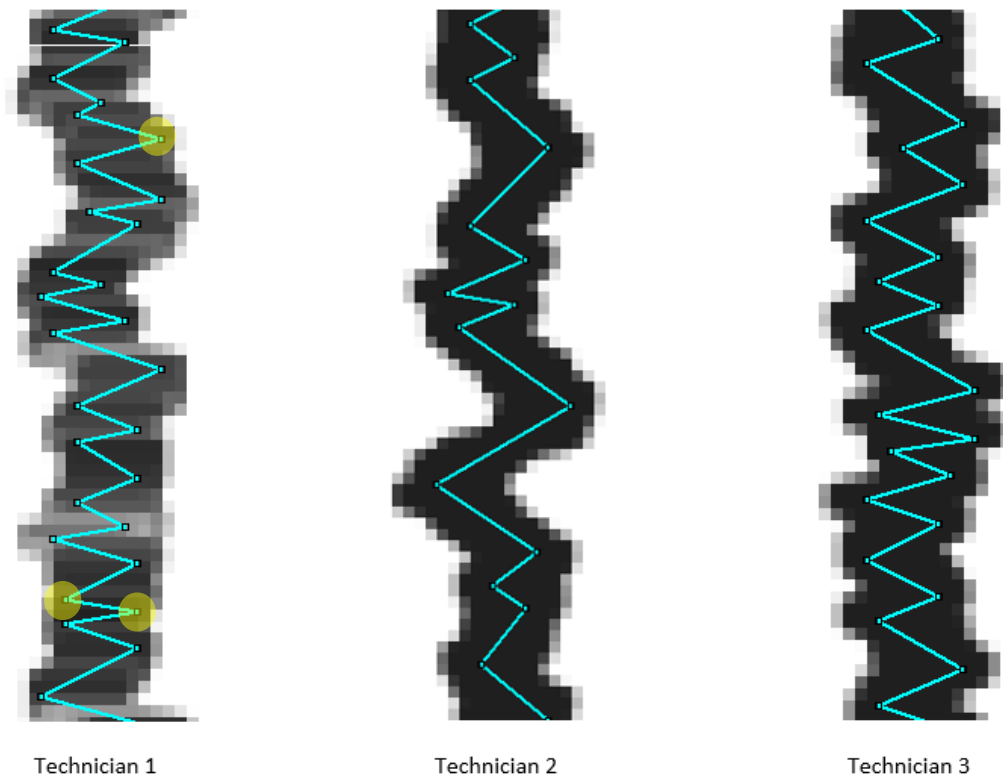
The 20x waveform reveals the most inconsistency between the technicians. Two technicians recovered around 5 Hz and the other slightly more around 8 Hz because the third technician may have selected their click points on the outer edge of the waveform. By doing this, the digitized signal amplitudes increase, and this may be the reason why there is an observed bump in the PSD graph around 5 Hz. The bump is highlighted yellow in the 20x PSD

and we see an averaged recovery for the 20x waveform is now around 7 Hz. Lastly, the 50x waveform has some variability, which is expected due to the nature of the wider waveform thickness and the technician's ability to see the signal. Recoverable frequencies for this waveform thickness ranged from 3 – 3.5 Hz.

Figure 38 displays the effect of a technician choosing the outside edge of the trace during digitization that correlates with the increase or bump in the PSD graph in Figure 37. For example, when a light beam slows down at a peak or trough, the beam has more exposure in this area causing a wider waveform thickness. To account for this, technicians should select in the middle of this area to collect the most accurate peak location. If technicians select the outer edge as their peak location (i.e., the click point) the true amplitude of the signal increases. Technician 1 in Figure 38 illustrates this problem. Areas of their digitization show peak locations that are too high and should be moved slightly more inward towards the center of the beam like Technicians 2 and 3. Some of the digitization points in Technician 1's waveform that demonstrate this problem are shown in yellow. Retaining accurate amplitude heights are vital for geophysical studies that utilize peak to peak signal measurements like earthquake-explosion discrimination studies.



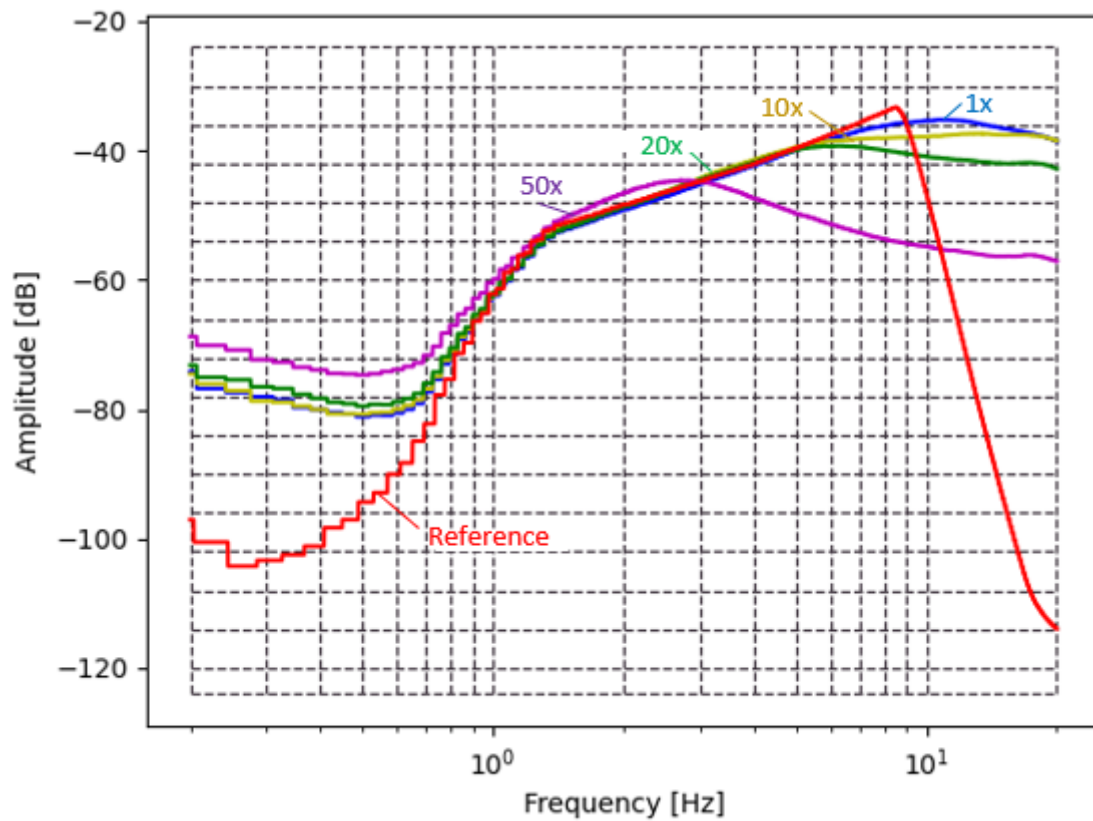
**Figure 37. Power Spectral Density (PSD) graphs illustrating the frequency recovery of each waveform thickness. The number in the upper left denotes the individual test. The red lines denote the reference signal, yellow lines are the three individual digitizations, and the black is the compiled final estimation digitization. Yellow box in the 20x PSD highlights a bump in a technician's recoverable frequency (see text for details).**



**Figure 38. Waveform comparison of three individuals who digitized a 1x amplitude waveform in *Wavetrack*. The blue line is the digitized trace. Technician 1 chose digitization point locations at the furthest edge of the waveform thickness (examples shown as yellow circles) more so than Technician's 2 and 3 where they selected points in the middle.**

Figure 39 shows the spectra of the final compiled waveforms on a PSD from each waveform thickness test and the reference waveform. Each color denotes a waveform thickness. From the PSDs, there are three waveform thicknesses that recover similar frequencies (1x, 10x, and 20x). The confidence that all three of these waveform thicknesses return the same signal is low. The 20x waveform is slightly skewed towards a higher recoverable frequency due to one technician selecting their click points on the outer edge of the waveform. This skewed the results for the maximum recoverable frequency for the 20x waveform making the average around 7 Hz. A more believable recovery for the 20x waveform based on the maximum recovery of the other two digitizations in this trial is around 5 Hz. The 50x waveform has low frequency recovery since areas containing high frequency signal are simply lost due to the wide waveform thickness.

With that being said, a majority of the world's analog seismograms' waveform thicknesses will typically fall between the 10x and 20x thickness. Digitizing a seismogram with this waveform size results in data recovery to between 5 – 7 Hz, which is acceptable for most geophysical analyses. There are extreme cases of seismograms having thin and broad thicknesses (1x and 50x) and in these situations, technician's should note the waveform thickness (especially 50x waveforms) due to the potential of having data recovery loss



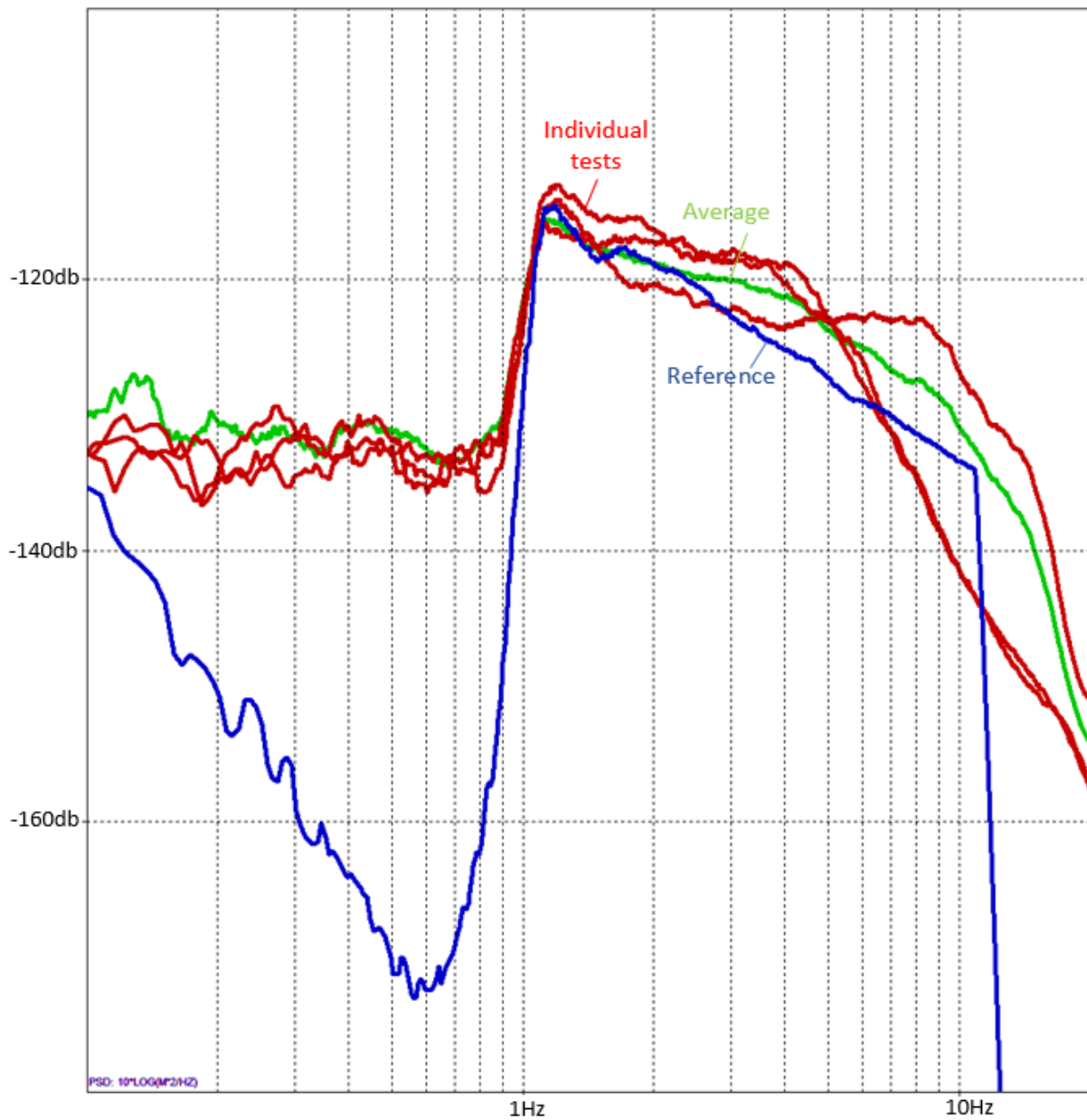
**Figure 39. Summary of the frequency recovery of digitizations with varied waveform thicknesses on a Power Spectral Density (PSD) graph. Each colored line is the average for a given waveform thickness. All digitized seismograms were compared against a reference signal to characterize the recoverable frequencies for each digitization.**



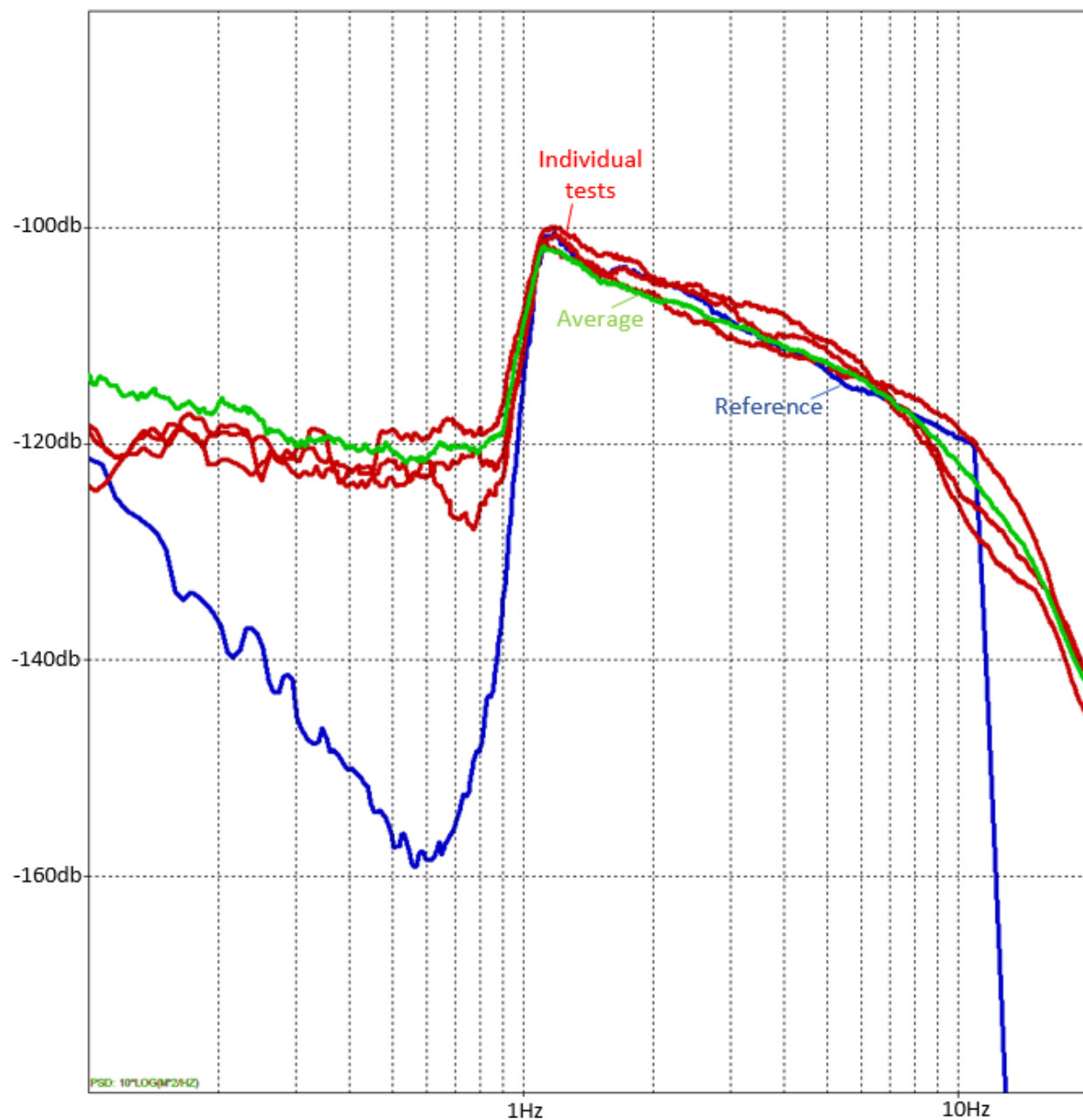
## **Amplitude Test**

A waveform was generated with a frequency range of 1-12 Hz with various amplitudes. These amplitudes ranged from 1x to 50x of the base amplitude. The process of generating these waveforms were described in the Methodology section and an example of each amplitude is shown in Figure 20. For this test, the seismograms were scanned at 600 DPI, and a final compiled waveform was used to show the ‘average’ of each individual variable’s result. Having a variety of amplitudes will examine how a waveform’s amplitude influences the frequency recovery.

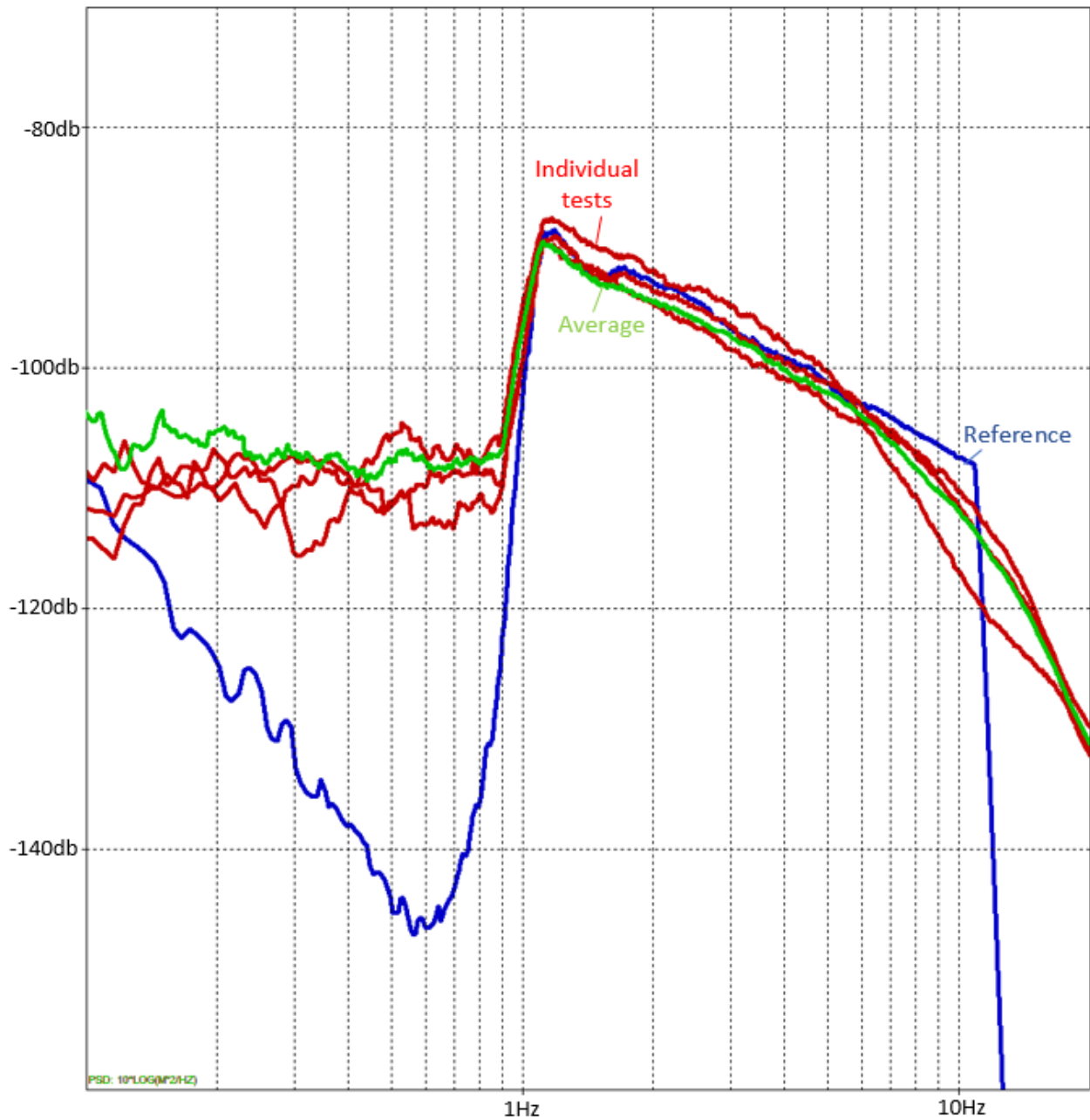
Figures 40 through 43 show PSD graphs for each amplitude test. The blue line is the reference waveform, the red lines are the three individual digitizations completed by independent technicians and the green line is the compiled or ‘averaged’ digitization. The 1x waveform in Figure 40 shows the most inconsistency in signal recovery due to the technician’s differing observations of the compressed signal. Alternatively, the 50x waveform has the least variability because the enlarged signals better define the waveform and retaining accurate points chosen at the peaks.



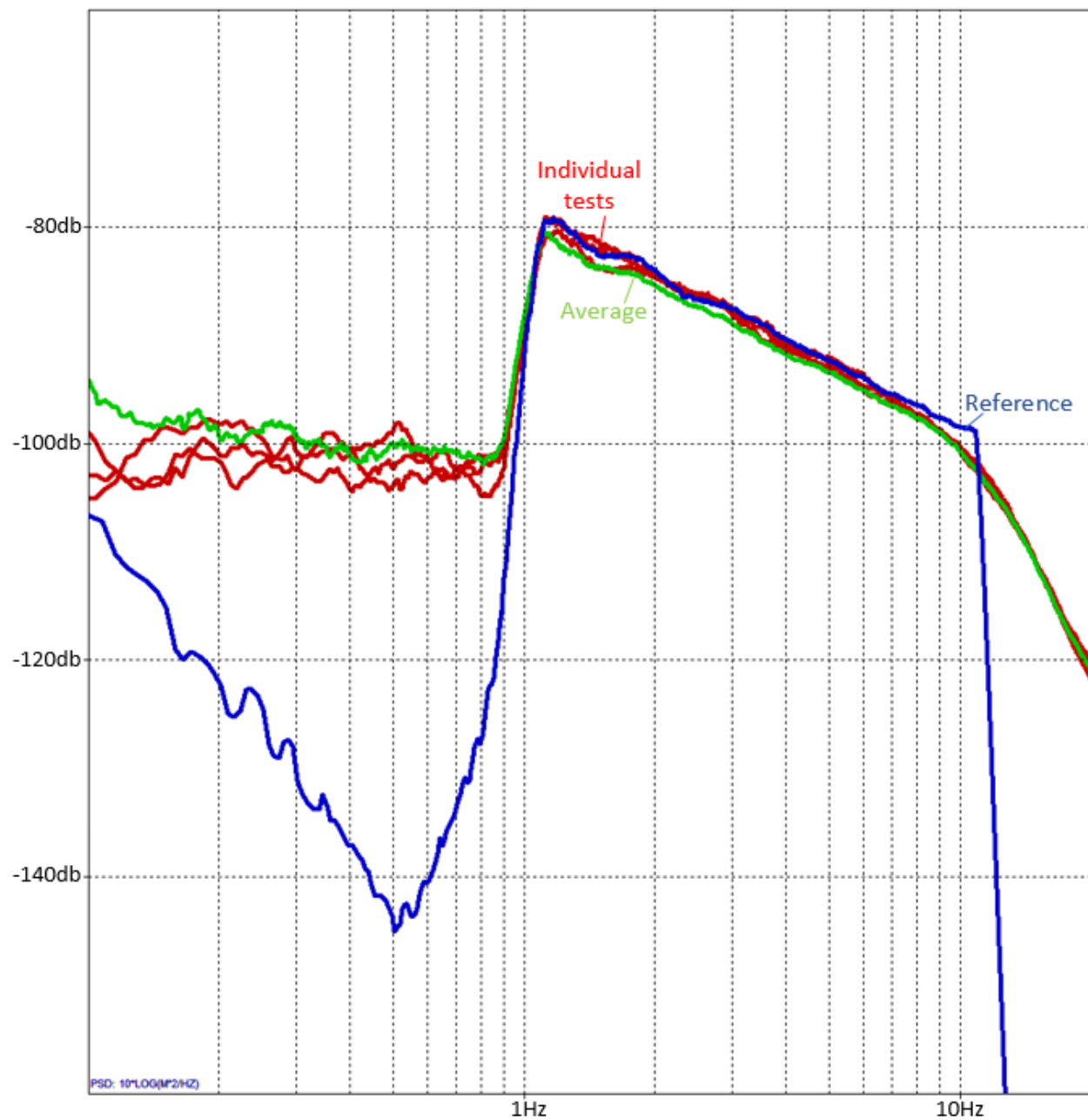
**Figure 40. PSD of a 1 -12 Hz 600 DPI waveform with a 1x amplification multiplier. The blue is the reference, red is the individual trials, and green is the compiled ‘averaged’ waveform.**



**Figure 41. PSD of a 1 -12 Hz 600 DPI waveform with a 5x amplification multiplier. The blue is the reference, red is the individual trials, and green is the compiled ‘averaged’ waveform.**



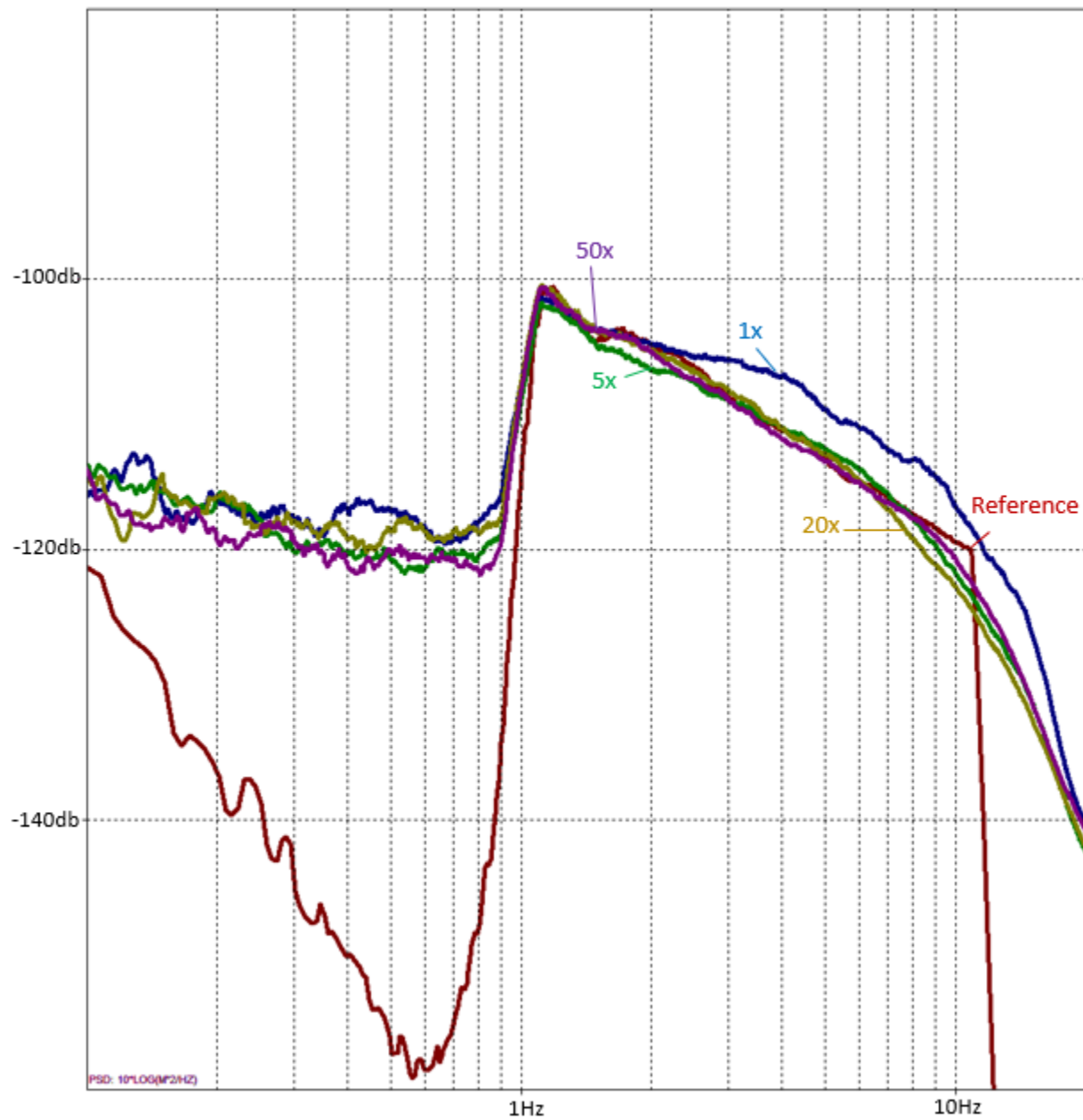
**Figure 42. PSD of a 1 -12 Hz 600 DPI waveform with a 20x amplification multiplier. The blue is the reference, red is the individual trials, and green is the compiled ‘averaged’ waveform.**



**Figure 43. PSD of a 1 -12 Hz 600 DPI waveform with a 50x amplification multiplier. The blue is the reference, red is the individual trials, and green is the compiled ‘averaged’ waveform.**

Figure 44 illustrates the frequency recovery with respect to signal amplitude. Each colored line is the compiled ‘averaged’ PSD for each amplitude trial. The PSD graph demonstrates that amplitude does not significantly impact a waveform’s frequency recovery. For this example, signals with amplitude that exceed five times the thickness of the recording line (5x, 20x, and 50x), are within 3db of the reference waveform. Having a digitization be within 3db of the reference signal is a good indicator that the digitizations are of good quality and are recovering accurate frequencies. Lower amplitude waveforms (like 1x) are still recoverable but need careful attention while digitizing. An increase in the 1x waveform is observed in the PSD. When amplitudes are compressed, the ease of correctly identifying and accurately selecting the true peaks and areas of a change in slope become increasingly difficult. A potential reason for the increase of the 1x waveform is that the technicians may have selected the outer edges of the waveform while digitizing (detailed in Figure 38). Selecting the outer edge of the waveform increases the true amplitude of the signal, which does not correlate with the original signal.

From the author’s personal experience with real-world seismograms, a signal waveform thickness between 1x and 5x correlates with general seismic background noise. Seismic events, depending on their magnitude, can show amplitude measurements well above the 20x waveform example in this study. Higher amplitudes run the risk of the signal overlapping the adjacent trace, component, or going off the page entirely. Many records were clipped for large earthquakes in the analog era. This potential problem increases the difficulty in accurately retaining amplitude locations. The chance of digitizing a higher amplitude event is great, other factors such as DPI and the focus of the signal beam play into the simplicity of digitization but also the overall frequency recovery.



**Figure 44. Power Spectral Density (PSD) graph demonstrating the frequency recovery of the various compiled ‘averaged’ signal amplitudes. Each colored line is a different amplitude level. The seismogram had an image resolution of 600 DPI with a frequency range of 1 – 12 Hz.**

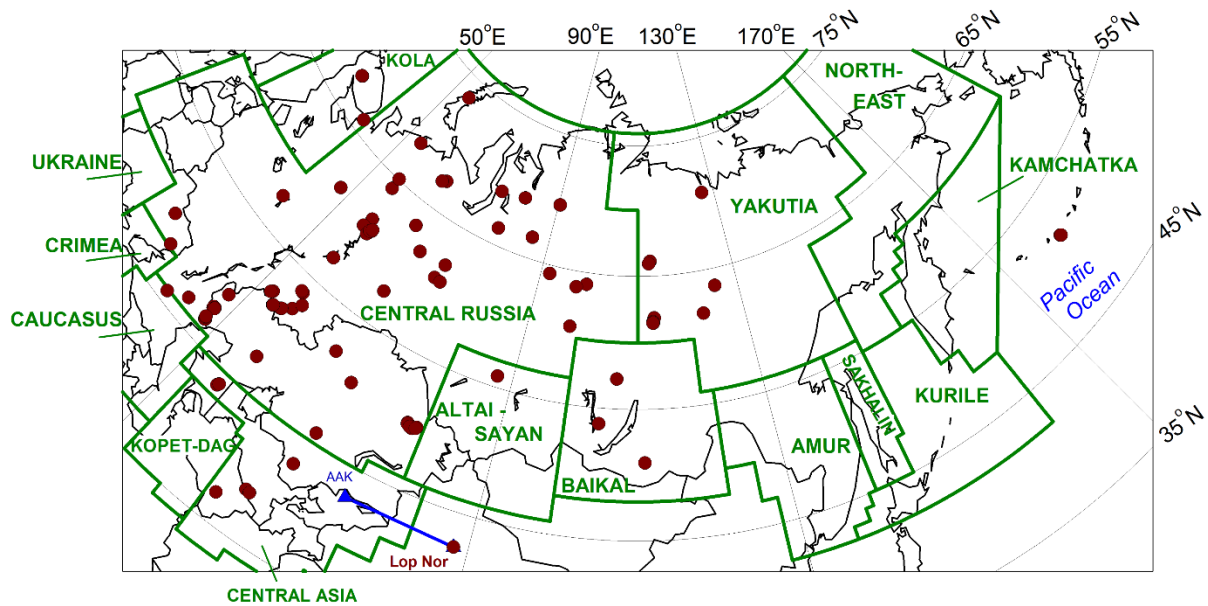
## Co-located Stations

Two co-located seismic instruments, one analog SKM-3 short period sensor and one broadband STS-1 sensor, from Ala-Archa (AAK), Kyrgyzstan were compared against one another to examine the trustworthiness of digitized analog seismograms and digital data. The seismic data for this comparison was from a Chinese Lop Nor nuclear test event. A map of station AAK and the detonation site are shown in Figure 45. The analog data was digitized by the author and the broadband data was downloaded from the Incorporated Research Institutions for Seismology (IRIS) digital seismogram database.

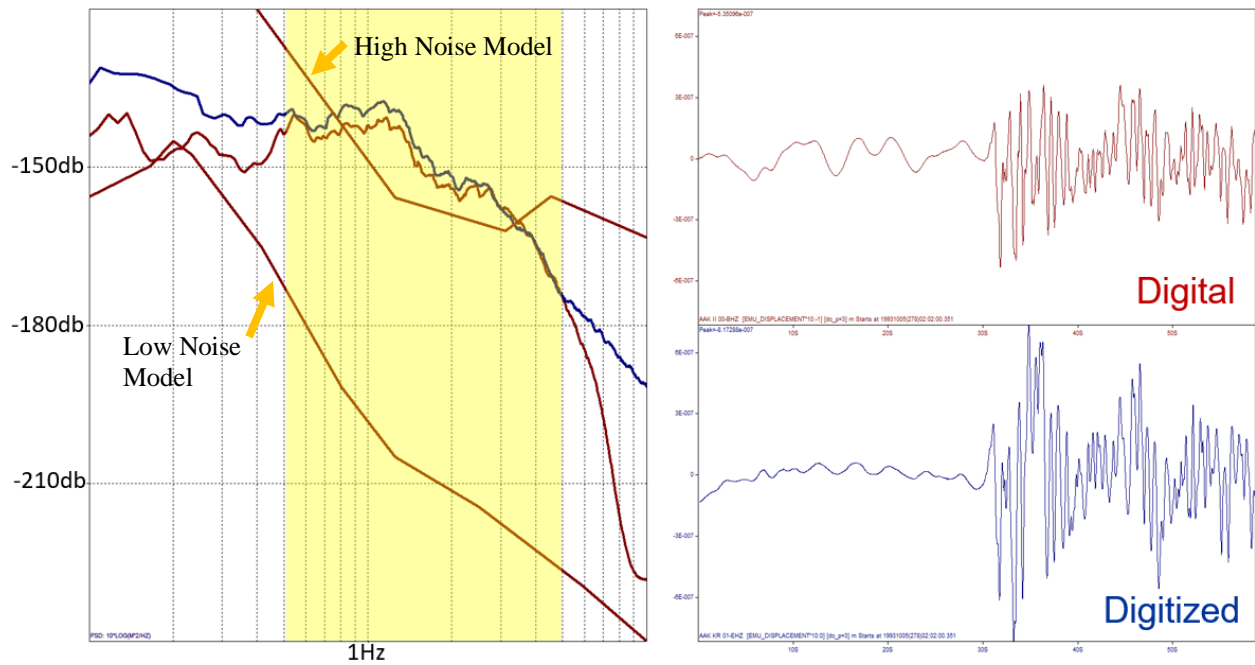
Figure 46 displays a comparison between the co-located instruments in the frequency domain. Within the PSD, there are labeled low and high noise models and these models are helpful in relating real-world data to the upper and lower bounds of seismic noise from the world's seismic stations. Both PSDs of the co-located instruments match well with one another between 0.5 - 5 Hz as shown in the yellow highlighted region on the PSD. Within this region, both waveforms fall within 3db of each other; however, above 5 Hz the broadband waveform takes a sharp decline. This roll off is due to the application of a lowpass filter on the broadband digital station. Due to the application of the low-pass filter, it is unknown whether the analog data continues to correlate with the digital data at higher frequencies. Another difference between the two waveforms is the sample rate. The digital waveform has a sample rate of 20 samples per second (sps), which we know is relatively low, but at the time of recording 20sps was thought to be adequate. The digitized waveform was a 600 DPI scan with a 100 sps sample rate. Even with these small differences, both waveform frequencies match well and show that the digitization of analog seismograms agree with digital acquired data. The data produced from digitizing is thus of high enough quality to be used in geophysical studies and processing.



Analog seismograms contain many variables that influence signal recovery. Variables such as signal amplitude and the beam focus account for the ability to see the signal. These are limitations put forth by the station at the time of recording. Scanning the image at an acceptable image resolution improves the clarity of the signal so the technician can ensure an accurate signal recovery. Situations where the amplitude extends off the page or the ink pen running out of ink, an increased image resolution will not improve the amount of data recovered simply due to variables that are out of the technician's control. Something like this shows that if one variable is impeded, data quality will suffer. On an analog seismogram, these variables can fluctuate between components. Alternatively, digital data does not have the same challenges as analog seismograms. Modern-day digital stations must monitor their storage capacity and battery power to ensure quality data collection. Overall, digitization is a complex problem, and a balance between waveform thickness, signal amplitude, and image resolution is needed for signal retention.



**Figure 45. Location for station with co-located instruments relative to the seismic event at the Lop Nor Chinese Nuclear Test Site used for comparison analysis. The blue triangle is the seismic station Ala-Archa (AAK), Kyrgyzstan and Lop Nor detonation site is red triangle. Additional detonation sites for Soviet and United States nuclear tests are shown as red circles.**



**Figure 46. PSD comparison of a broadband STS-1 sensor (red) and an analog SKM-3 sensor (blue). The High and Low Noise Models illustrate the bounds of seismic noise relative to the world's seismic stations. The yellow region highlights the frequencies where the two PSD curves waveform falls within 3db of one another between 0.5 – 5 Hz. An application of a low-pass filter on the broadband sensor is the reason for a strong signal roll off after 5 Hz.**

## CONCLUSIONS

The digitization of analog seismograms is a complex process with many variables affecting the ability to recover the original analog waveform and represent it in digital form with minimal loss of information. Some variables in the process are within our control, such as scan resolution, while other variables are natural limitations from the original record, like waveform thickness and amplitude. Nevertheless, it is necessary to understand the overall effect of the variables in the digitization process to both produce high quality digital waveforms as well as understand the limitations in the process and resulting data.

Scan DPI has a significant impact on the frequency recovery of digitizations. Higher DPI scanned seismograms yield higher recoverable frequencies in the digitization process. A DPI of at least 600 DPI is needed to achieve recoverable signal up to 8 Hz; however, if there is a need for retaining higher frequencies, a higher DPI will need to be used. Continual increases in scan resolution will not forever improve frequency recovery as evidenced by no change in recovery between scan resolutions of 1500 and 3000 DPI in one of the tests.

Regarding waveform thickness, the thinner the trace, the easier it is to recover signal. Narrower waveform thicknesses allow an easier selection of digitization points chosen at the peaks and areas where a change of slope occurs. If the signal is severely out of focus, frequency recovery is reduced, and maximum expected frequencies can range between 3 – 5 Hz. A majority of the world's seismograms will most likely fall between a 10x and 20x thickness described in this study. Digitizing a seismogram within this waveform size will result in a data recovery to between 5 – 7 Hz. Recovering frequencies in this range are acceptable for most geophysical analyses.

Signal amplitude does not have a significant influence on a waveform's frequency recovery for most seismograms. The true amplitudes of the signal are more easily chosen at higher amplitudes (like 20x and 50x in the examples described) compared to compressed amplitudes. Compressed amplitudes require additional attention because there is an increased chance that the technician may chose the outer edge of the signal beam and increase the true amplitude of the signal.

A technician's ability to recover the frequency and amplitude is an important factor in digitization and signal recovery. There are many facets that influence accurate recovery such as experience, understanding of waveform mechanics, speed or time taken to digitize, and care taken during digitization. From our group of technicians, there was no correlation between digitization experience and quality of their digitization. This is due to the reason that these technicians are veterans in analog seismogram digitization. I think there would be some separation if we took a technician just starting out and a technician who has been digitizing for a long time. Speed is important because technicians may speed through a digitization and miss click points in their digitization. Note that as a technician gains experience, they should digitize faster. The understanding of waveform mechanics and care while digitizing are additional items to consider because frankly some technicians cannot understand what seismograms should look like or understand ground motion so their digitizations will suffer from this. Lastly, does the technician care what their digitization looks like? Sometimes technicians will not put enough care or attention in their work which affects the resulting digitization.

The process of digitizing analog seismograms is meaningful, and worldwide efforts from institutions are needed to save and preserve these historic seismic. Analog seismograms contain vital information of large earthquakes and information from nuclear testing. For our manual

digitization program, a PCHIP interpolation is recommended with a high sampling rate because it maintains the original waveform's shape by utilizing the digitization points at each peak. The digitizations produced manually correlate well with digital data and are indeed usable for geophysical analysis

The complexity of digitization is not only a factor of seismogram variables in the digitization process but also the combination of human influences or technician ability. If one variable (DPI, signal amplitude, waveform thickness, and human influence) suffers, the data quality and the frequency recovery will be negatively affected. There needs to be a balance between all of the influences to achieve good digitizations. All of the factors can be on a sliding scale of importance depending on the research and data requirements for post processing.

## REFERENCES

- Bromirski, P. D., & Chuang, S. (2003). SeisDig: Software to Digitize Analog Seismogram Images, User's Manual. Scripps Institution of Oceanography Technical Report. <http://escholarship.org/uc/item/76b2m74m>, 28 pp.
- Chiburis, E. F., Ahner, R. O., & Reinhardt, E. C. (1980). Procedures for Digitizing Seismograms. Indian Harbour Beach, FL. 44pp.
- Cygnus Research International (n.d). from <https://www.cygres.com/OcnPageE/Glosry/SpecE.html>
- Eastmancuts. (2013, February 6). Eastman-joining large pieces on digitizer. [Video]. YouTube. [https://www.youtube.com/watch?v=jvUQ\\_whKje0](https://www.youtube.com/watch?v=jvUQ_whKje0)
- Fritsch, F. N. & Carlson, R. E., (1980). Monotone Piecewise Cubic Interpolation. SIAM Journal on Numerical Analysis. 17. p.238–246.
- GSRAS. (2001). Calibration of the Seismic Stations of the Russian Academy of Sciences for CTBT Seismic Monitoring Purposes. Russian Academy of Sciences Geophysical Survey of Russia Technical Report. 30pp.
- Indian Institute of Science. (n.d.). Conducting a GIS Analysis: <http://wgbis.ces.iisc.ernet.in/envis/Remote/section156.htm>
- Ishii, M., Ishii, H., Bernier, B., & Bulat, E. (2015). Efforts to Recover and Digitize Analog Seismograms from Harvard-Adam- Dziewoński Observatory. Seismological Research Letters. 86(1), p.255-261.
- Kemerait, R. C., Kraft, G., Mott, J. S., & Dohner, E. (1981). A study of the hand-digitization process for digitizing short period seismic data. Indian Harbour Beach, FL. 38 pp.
- Mackey, K. G., Hartse, H., & Fujita, K. (2009). Final Report Analysis of Digitized Seismograms from Russian Geophysical Survey Stations of Soviet Peaceful Nuclear Explosions. (Report No. AFRI-RV-Ha-TR-2009-0000). Michigan State University, East Lansing, MI. 111 pp.
- Michelini, A., De Simoni, B., Amato, A., & Boschi, E. (2005). Collecting, Digitizing, and Distributing Historical Seismological Data. EOS. 86(28). p.261-266.
- Octave Forge Community. (2017, January 2). Function reference: Interp1. From <https://octave.sourceforge.io.octave/function/interp1.html>
- Okal, E. (2015). Historical seismograms: Preserving an endangered species. GeoResJ. 6. p.53-64.
- Pintore, S., Quintiliani, M., & Franceschi, D. (2005). Teseo: A vectoriser of historical seismograms. Computers & Geosciences. 31(10). p.1277-1285.

- Skiljan, I. (1996). Irfanview. [Computer software]. Developer. <https://www.irfanview.com/>
- Sokolova, I. (2015). Acoustic waves from atmospheric nuclear explosions recorded by infrasound and seismic stations of Kazakhstan. Poster T2.3-P3 presented at the CTBTO Snt 2015 Annual Meeting in Vienna, Austria.  
[https://www.ctbto.org/fileadmin/user\\_upload/SnT2015/SnT2015\\_Posters/T2.3-P3.pdf](https://www.ctbto.org/fileadmin/user_upload/SnT2015/SnT2015_Posters/T2.3-P3.pdf)
- Toskey, N. (2018, July 10). Image Resolution Explained [web log].  
<http://www.makingmediatoremember.com/learning/image-resolution-explained/>
- Yu, Z., Chaoyong, P., & Jiansi, Y. (2017). Historical Seismic Map Database and Sharing Platform. *Seismic and Geomagnetic Observations and Research*. 38(4). p.207-211.
- Zhang, J., Song, X., Li, Y., Richards, P. G., Sun, X., Waldhauser, F. (2005). Inner Core Differential Motion Confirmed by Earthquake Waveform Doublets. *Science*. 309(5739). p.1357-1360.



## APPENDIX

One of the holy grails for the digitization of seismograms is the development of a fully automated routine that will generate high quality waves with minimum operator input. In this thesis, I investigated parameters in the process using manual digitization techniques. However, I also investigated the possibility of conducting this research using the Harvard University developed semi-automatic digitization program called *DigitSeis* (Ishii et al., 2015). Although *DigitSeis* was not used in this research, I am providing an evaluation of the current state of the software (v1.5, 2020). Many of the variables discussed above are relevant for both manual and automated routines and affect the resulting digital waveforms.

*DigitSeis* is an image processing program where a line is traced along the seismogram. The image is classified or identified into three categories: noise, signal, and time marks. After careful identification by the user, the program digitizes the signal. If the original digitization was unsatisfactory, the user can manually go back and fix any data gaps in the digitization or incorrectly traced signals and re-digitize.

There are seven main steps to produce a digitized waveform in *DigitSeis*. First is image processing and uploading of the image into the program. *DigitSeis* requires JPEG images where the image has a white trace on a black background whereas *Wavetrack* required BMP images. If the images do not have the required image contrast, *DigitSeis* will adjust it automatically. Another process for preparing the image for digitization is cropping the image. As *DigitSeis* computes the pixel matrix within the image, it is strongly recommended to crop the image before the start of digitization as this will speed up processing time. The author found that cropping the image prior to importing the image into *DigitSeis* was the best method due to the large image size and long processing time.

The next step is determining the minute marks. Retaining accurate timing in analog seismogram digitization is vital for geophysical studies. *DigitSeis* can account for numerous types of time marks such as minute mark offsets, a few seconds of data that are above or below the normal trace line, and no time marks. A physical measurement of the pixel length of the time mark is used for later classification.

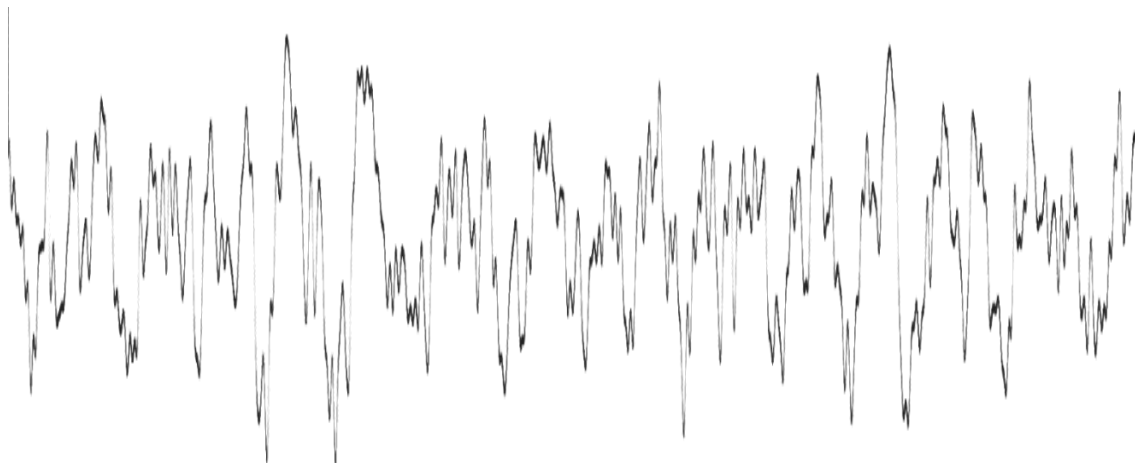
Classification and digitization are the next two steps for digitizing analog seismograms in *DigitSeis*. *DigitSeis* has a three-class system where it categorizes information into signal, noise, and time marks. Based on the prior pixel measurements of the time marks, *DigitSeis* manually calculates and displays the classification of the entire seismogram. The classification can be edited and re-classified to obtain an optimal classification for digitization. Using the pre-determined classifications, *DigitSeis* digitizes the signal portion of the waveform. Once digitized, the user can manually edit the digitization by merging several traces into one continuous waveform, correcting any digitized signals, and filling in any data gaps. If the original classification was unsatisfactory, the user can go back and edit the original classification and re-digitize.

The last two steps are determining the time and exporting the data file in the form of a SAC file. Each data line within the seismogram has two red bars that appear on either end of the digitization, and the user must click those points and enter in the date and time for that specific section of the seismogram. However, an obstacle for this step, which is certainly a suggested area of improvement for *DigitSeis*, is that the user must physically click on the red line in order to enter the time. For the digitizations in this small study, the red lines were on the edges of the viewing window making it challenging to define the time for the digitization. A suggestion for improvement is to set a tolerance around the red lines to allow easier selection. After the timing

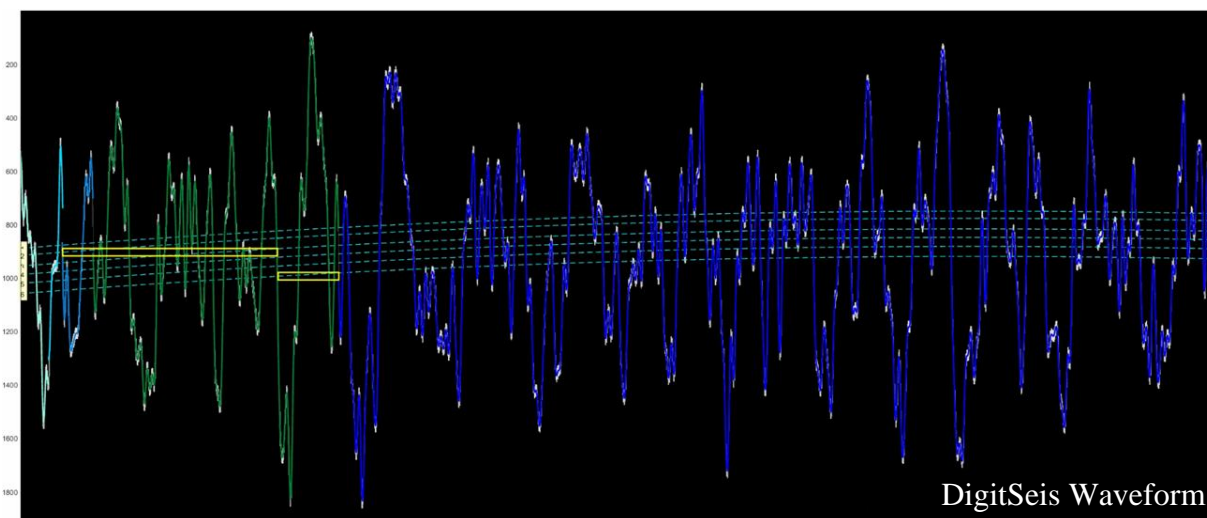
is determined the data is then exported as count of pixels in SAC format. The digitized data is now able to be viewed and processed on a computer.

This study assessed two synthetic white noise waveforms through the *DigitSeis* digitization process at 3000 DPI and compared the digitizations against reference signals. Each digitization was administered and edited by the author. One waveform was low frequency with a range of 0.1 – 2 Hz and the other was a higher frequency waveform with a range of 1 – 12 Hz. For the 0.1 – 2 Hz waveform, *DigitSeis* classified that waveform into six different data segments. These separate traces are the dashed blue lines and the change of signal color in Figure 47 while the yellow boxes denote data gap areas within the digitization. Manual modifications were needed to fix the data gaps and to merge the traces back again into one waveform.

As the lower frequency waveform is less complex, *DigitSeis* succeeded in recovering the full signal up to 2 Hz. The green is the reference signal that did not go through the digitization process and the red is the trace produced through *DigitSeis* (Figure 48). The PSD graph illustrates that both traces recovered the same frequency. There are subtle but detectable differences between the two, but overall, the traces compare well.

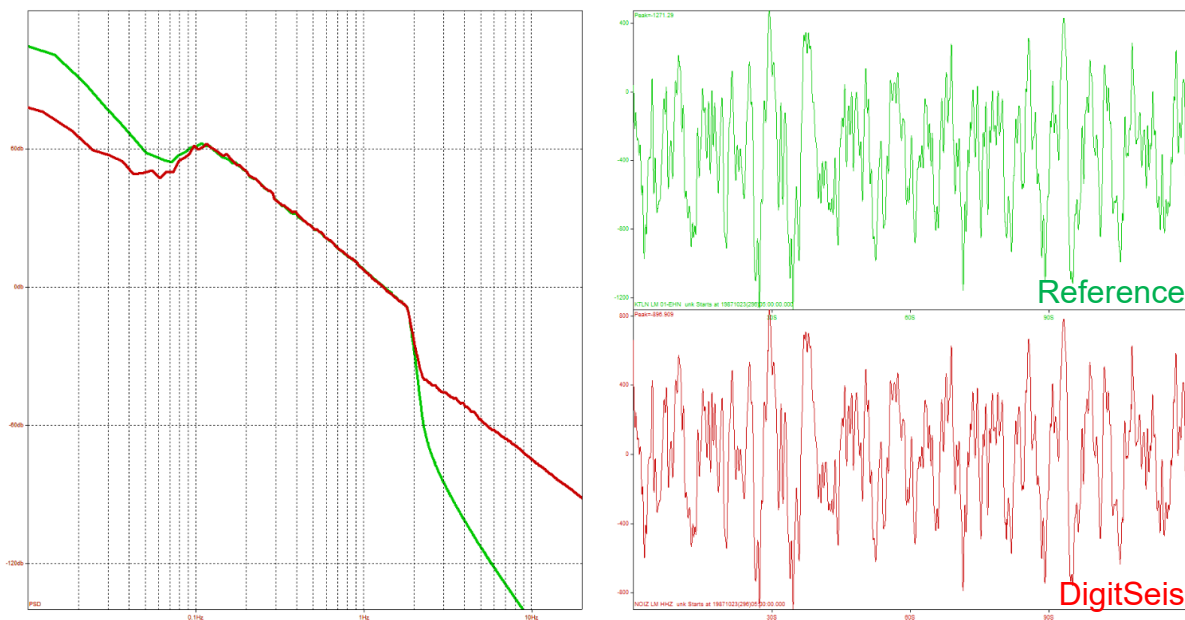


Reference Seismogram



DigitSeis Waveform

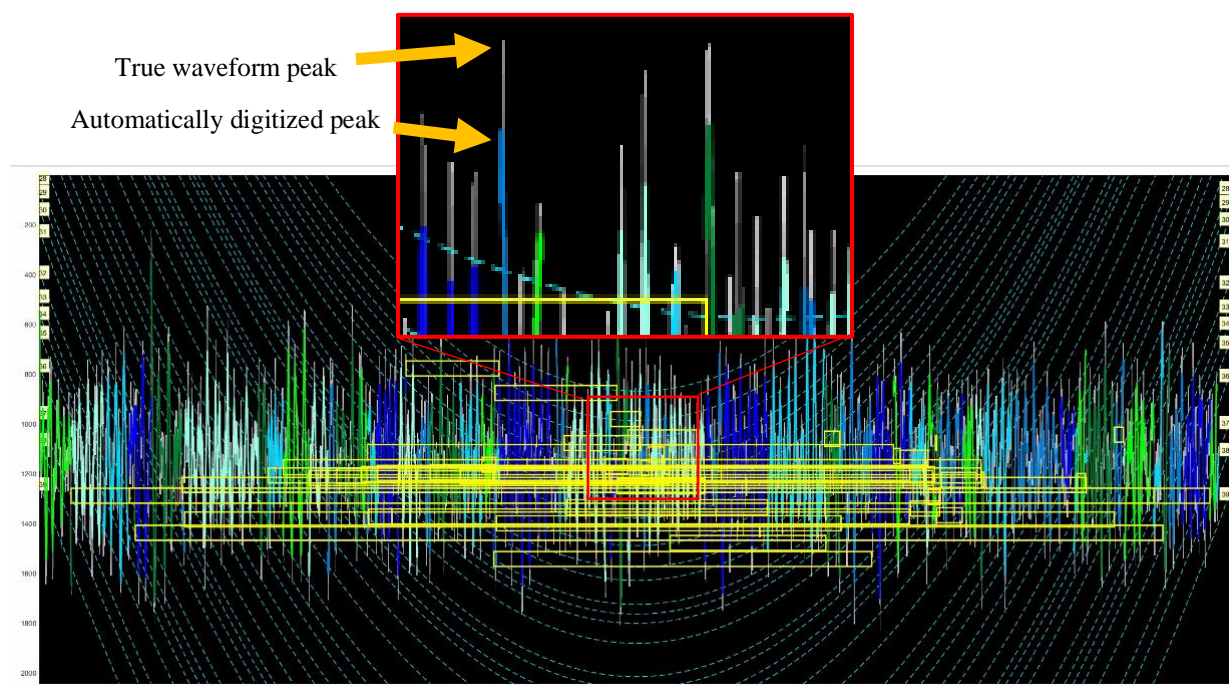
**Figure 47. Example of a digitized synthetic wave at 0.1 – 2 Hz in *DigitSeis*. The different colors of the waveform represent separate traces within the waveform. Yellow boxes denote data gaps in the digitization. Manual improvements are needed to fix the gaps and merge the traces.**



**Figure 48. Comparison between a low frequency signal (0.1 – 2 Hz) and its *DigitSeis* digitization. The green trace is the reference signal that did not go through the digitization process while the red trace was digitized using *DigitSeis*. The digitized signal recovers the full 2 Hz signal as seen in the PSD graph on the left.**

Both Figures 47 and 48 provide credible data that *DigitSeis* can fully recover waveforms up to 2 Hz. However, a higher frequency waveform with a frequency range of 1 – 12 Hz does not have the same result. This waveform was difficult to digitize because of its complexity. *DigitSeis* classified the trace into 38 individual traces, unlike the lower frequency waveform that only had six pieces (Figure 49). Significant modifications were needed to make this waveform usable for scientific studies.

If left in its current state, and not edited, the waveform has significant data loss. The red box in Figure 49 illustrates how the signal does not reach the full extent of a signal's amplitude. That is one source of error if left untouched. Another source of error is the 'averaging effect' algorithm within *DigitSeis*. For high frequency signals, the program cannot recognize and follow the trace due to a combination of complex signals and a wider trace thickness resulting in a muted waveform. It is suspected that *DigitSeis* considers the trace thickness when determining the trace during digitization. If a high frequency signal were present, the signal may be lost and not be accurately recovered. This observation is similar to the remark from Chibirus et al. (1980) where they noticed an 'averaging' effect of their WWSSN digitizations. The authors found that it was a "poor representation of the original signal." For this high frequency waveform example, the entire waveform was manually modified to enhance the amplitudes and adjust any subdued signals.



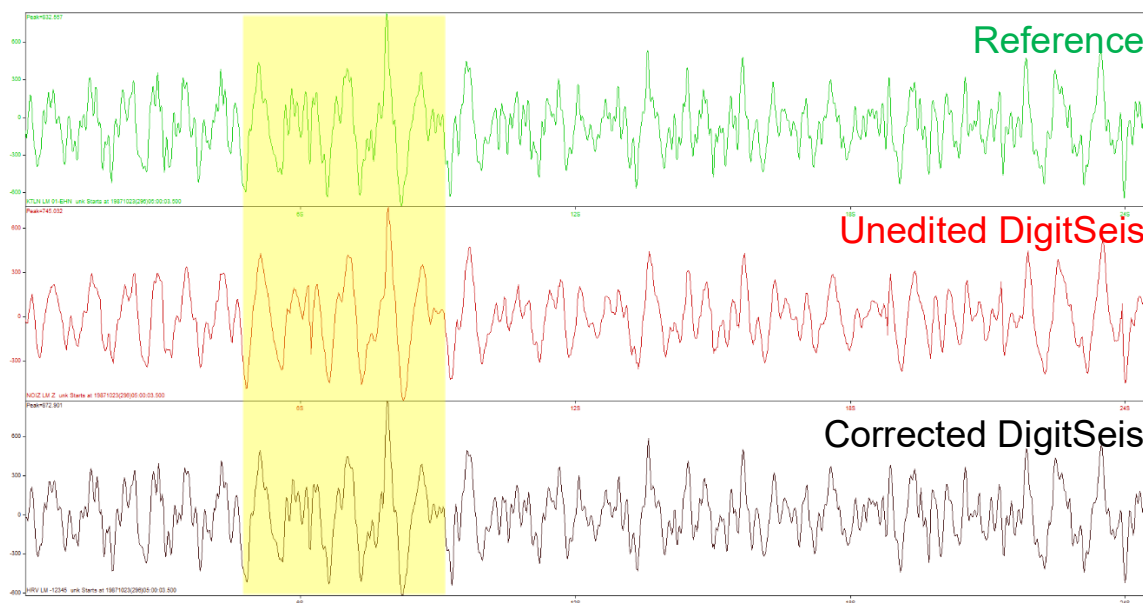
**Figure 49. Classification and initial digitization of high frequency waveform in *DigitSeis*. A 1 – 12 Hz signal was digitized in *DigitSeis*, and the program broke the signal into 38 different traces which are denoted by the blue dashed lines and the different colored signals. The yellow boxes denote data gaps within the waveform that need to be manually modified. The inset illustrates how the digitized trace does not continue to the full extent of the waveform’s amplitude.**

Figure 50 shows how a high frequency waveform appears pre- and post- editing. The un-edited signal is inconsistent with the reference signal and has severe loss of high frequency signal (shown in yellow highlighted region). There are two ways to edit a digitization in *DigitSeis*. The first method is to re-classify the objects in the waveform and re-digitize the waveform. The second method is more laborious where the user must manually comb through the digitization and update any errors. Within the manual method, reference points are selected to guide *DigitSeis* to fit a spline interpolation method between the points. Since the automatic algorithm selects discrete time points in the digitization (unlike *Wavetrack*, where a point is a peak or change of slope), a spline is ideal for this digitization program. If one point was adjusted, it affected the entire waveform in the editing window, thus smaller working windows for editing were suggested. Having smaller editing windows increased the time in the digitization process. The time spent improving the waveform is necessary because from Figure 51, the PSD graph illustrates the frequency recovery of the waveform pre- and post- editing. The edited version of the waveform (gold) follows strongly with the reference signal (blue) compared to the un-edited version (green). Also in the PSD, a manually digitized version of this waveform is shown in red. The *Wavetrack* digitization fully recovers the 12 Hz whereas the edited *DigitSeis* waveform recovers just below. I believe the difference in these waveforms is from my fluency with each digitization program. I have over seven years' experience with *Wavetrack* and only a few months with *DigitSeis*. To improve my results with *DigitSeis*, I could have spent more time updating and improving the waveform, but I am confident that *DigitSeis* returns a comparable waveform to *Wavetrack*.

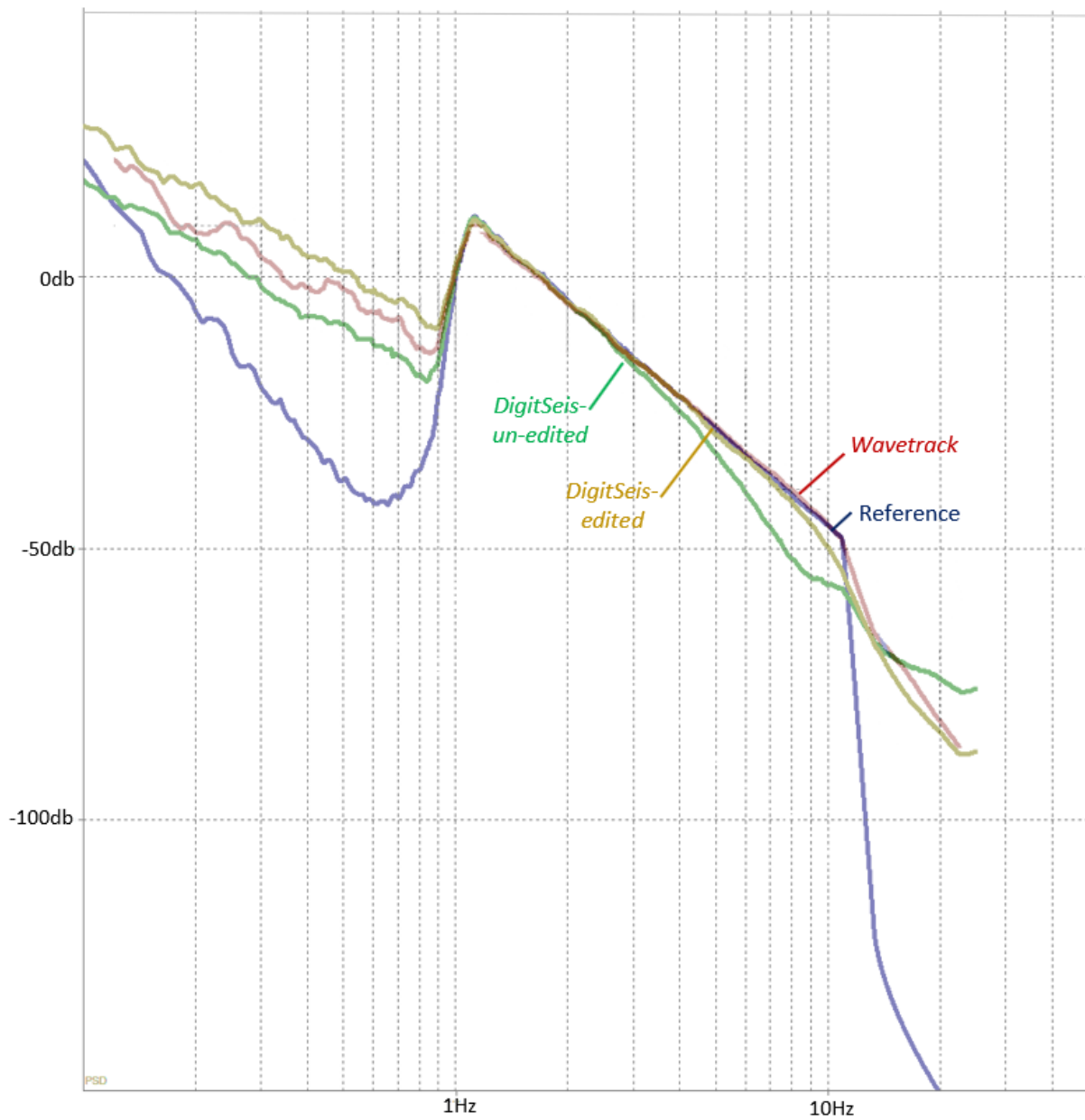
The approximate digitization time for both programs, *DigitSeis* and *Wavetrack*, are shown in Table 2. Lower frequency signals take significantly less time to digitize in both



programs compared to higher frequency seismograms. My final conclusions are that overall, *DigitSeis* produces good, quality digitizations; however, it struggles with complex waveforms. If a complex waveform is used, substantial time and effort is needed to produce a sufficient waveform for analysis. *Wavetrack* also produces good, quality seismograms and the digitization process is currently faster. For the complex waveform evaluated, it took approximately five times longer to achieve a quality waveform with *DigitSeis* as compared to *Wavetrack* due to the tedious process of implementing manual corrections. It is important to note that the synthetic waveforms used to test *DigitSeis* were single traces. They did not have adjacent waveform traces that are typical in a real-world seismogram. Adjacent traces crossing the waveform under digitization can interfere with the automatic digitization routine, requiring additional time for corrections. Both programs require significant training, but after an initial practice period you will achieve quality digitizations. This is the goal for any digitization project (either completed manually or automatically). *DigitSeis* is an open-source software program and readily available for download unlike *Wavetrack* which at this time is only used within Michigan State University and their research collaborators in Russia, Kazakhstan, and Kyrgyzstan. In the end, the program availability, complexity of the waveform, and time available for digitization should be considered when digitizing analog seismograms.



**Figure 50. Relationship between a reference waveform (green), an un-edited digitization in *DigitSeis* (red), and a corrected digitization in *DigitSeis* (black). If left un-edited, significant data loss can occur in the digitization and an example is shown in the yellow highlighted region. Significant manual modifications are needed to recover the lost data.**



**Figure 51. Power Spectral Density (PSD) graph illustrating the frequency recovery of a 1 – 12 Hz waveform that was digitized in two digitization programs: *DigitSeis* and *Wavetrack*.**

| <b>Waveform</b>            | <b>Digitization Program</b> | <b>Total Digitization time</b> |
|----------------------------|-----------------------------|--------------------------------|
| Low frequency (0.1 – 2 Hz) | <i>DigitSeis</i>            | ~ 1 hour                       |
|                            | <i>Wavetrack</i>            | ~ 1 hour                       |
| High frequency (1 – 12 Hz) | <i>DigitSeis</i>            | ~ 8 – 10 hours                 |
|                            | <i>Wavetrack</i>            | ~ 1.5 – 2 hours                |

**Table 2. Relative time it takes to digitize a low and high frequency signal in a manual digitization program, *Wavetrack*, and an automatic program, *DigitSeis*.**