Chapter 3

Data Analysis

The analysis of the measured track data is described in this chapter. First, information regarding source and content of the measured track data is discussed, followed by the evaluation and analysis methods used. Finally, the results of the data analysis and their application to TRAKVU are described.

3.1 Track Data

The measured track data that were used in this analysis were taken from the test tracks at the Transportation Technology Center, Inc. (TTCI) in Pueblo, Colorado. The data were collected by ENSCO, Inc., using the FRA Track Geometry Measurement Vehicle (TGMV) T-10, from October 13, 1997 to October 16, 1997. The TGMV sampled once every foot (or every 30 cm), and recorded the curvature, gauge, cross level, left alignment, right alignment, left profile, and right profile. These data were then provided to TTCI in four data files, one for each day's measurements. Each data file contained the track geometry data of one or more of the TTCI test tracks, shown in Fig. 3.1. The test tracks that were measured were the Railroad Test Track (RTT), Transit Test Track (TTT), Turn Around Loop (Balloon Loop), Precision Test Track (PTT), Wheel/Rail Mechanism Track (WRM), and the FAST High Tonnage Loop (HTL). These tracks constitute most of the 48 miles of railroad tracks at TTCI devoted to the testing of locomotives, vehicles, track components, and safety and signal devices.

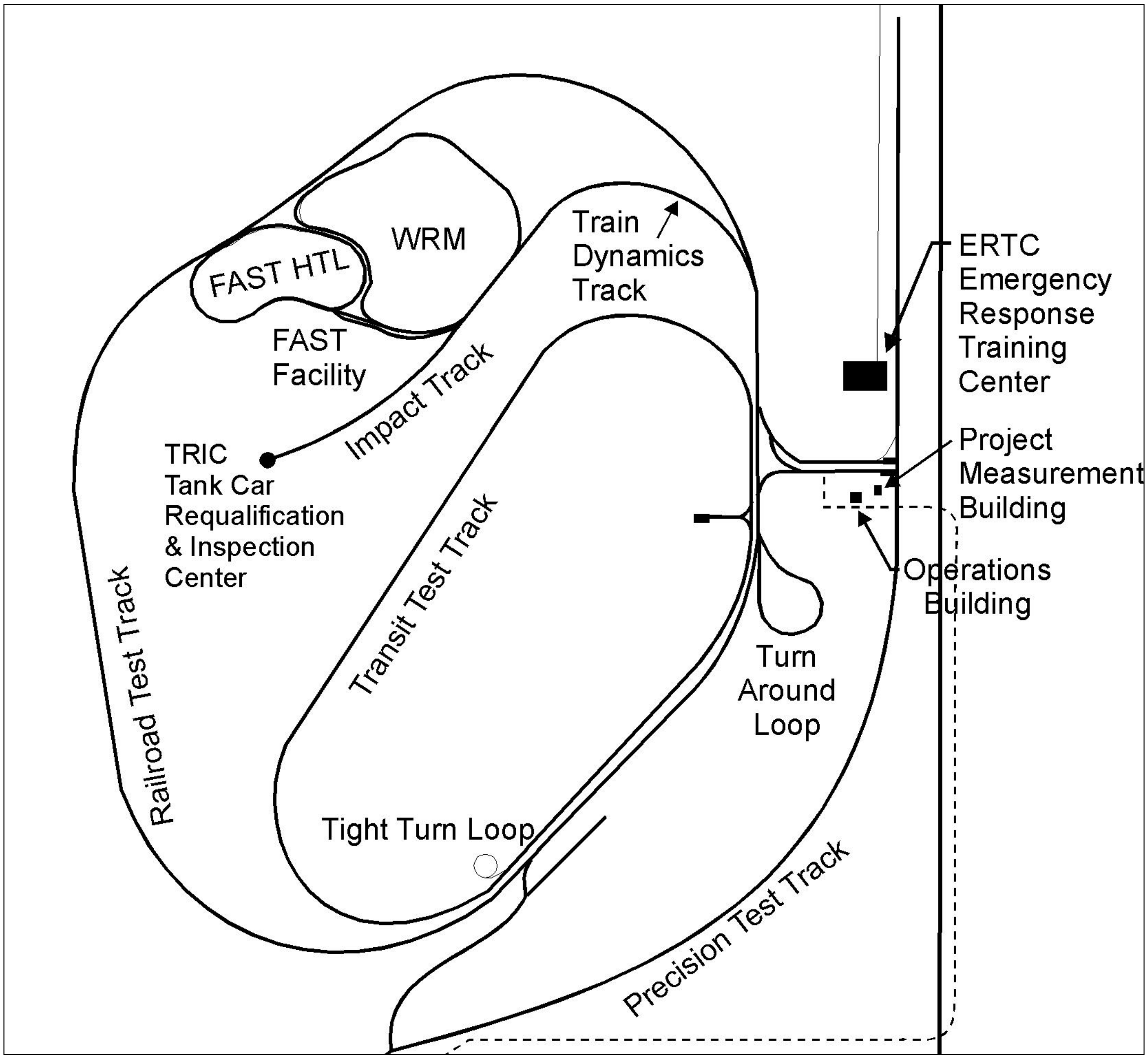


Figure 3.1 Test Tracks at the Transportation Technology Center, Inc. in Pueblo, CO

3.2 Evaluate Track Data

The first step in evaluating the measured track data was to identify the test loops at TTCI from the files provided. This was accomplished by plotting each file using MATLAB [10], and comparing the curvature and superelevation plots to the plots on the track charts from TTCI [11], such as that shown in Fig. 3.2.

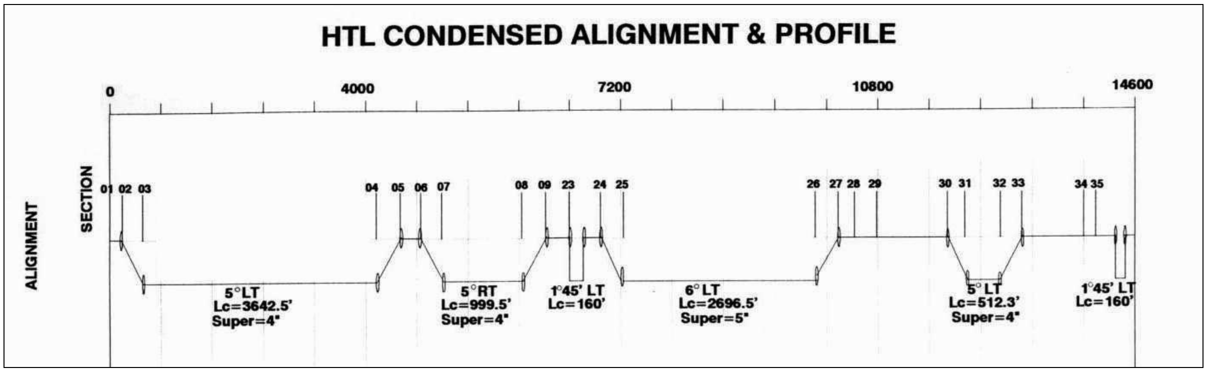


Figure 3.2 Track Chart of High Tonnage Loop from the Transportation Technology Center, Inc. [11]

By comparing the length of track, degree of curvature, length of curve, and amount of superelevation, the data were matched to the appropriate track. The start and end points of each track were identified, and used to separate each track into its own file for ease of analysis.

MATLAB M-files[1] were used to load the data from each track and plot the various elements of the track geometry. While viewing the alignment and profile data, “dead spots,” or periods where the data was absent, were discovered. These areas, shown in Fig. 3.3, were caused when the measurement car was traveling too slowly to accurately measure the track. Since the dead spots do not contain any data, they were ignored in our statistical and frequency analysis. The analysis was performed for five test tracks: three class-4 tracks (the Balloon Loop, WRM, and HTL); one class-5 track (the TTT); and one class-6 track (the RTT). No analysis was performed for the PTT since it contained several large dead spots, and the remaining data were not sufficient for a valid statistical or frequency analysis.

[1] All M-Files used for this study are included in Appendix A.

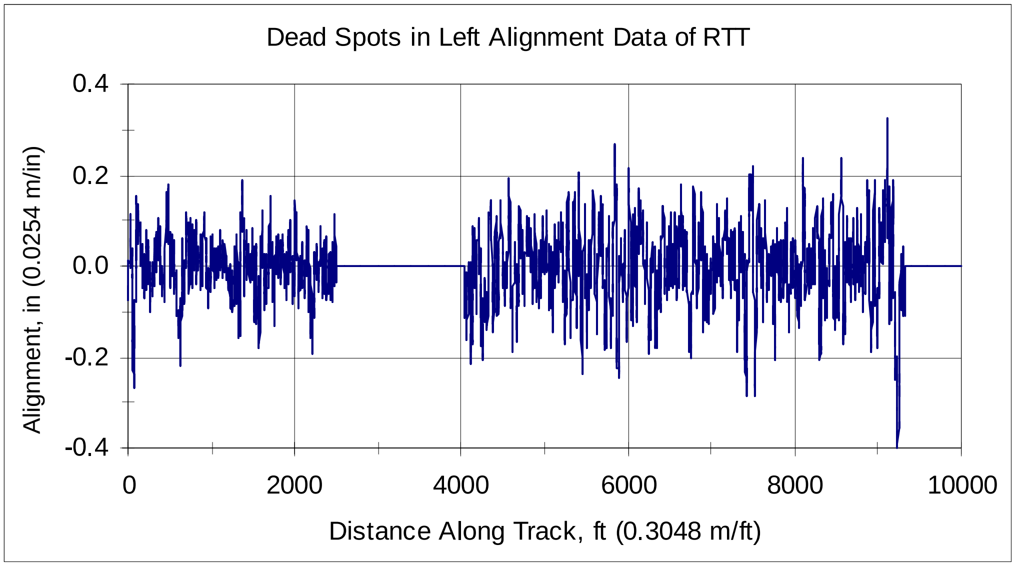


Figure 3.3 Dead Spots in the Left Alignment Data of the Railroad Test Track

3.3 Analysis

The purpose of the data analysis was to identify any general characteristics that exist in the various track classes. To discern any trends in the data, both a statistical and a frequency analysis were performed on the alignment and profile data. We chose to examine the irregularities in the alignment and profile data because they are representative of all the track geometry data. The track perturbations associated with the curvature and gauge data are contained within the alignment data, as described earlier in Section 2.3. The same is true for the cross level data, which is contained within the profile data. The mean and standard deviation were examined for each track to establish the statistical characteristics of the tracks. Further, a frequency analysis of the track data was conducted to examine the frequency characteristics for each class of track.

3.3.1 Statistical Analysis

The statistical analysis was performed in two different ways. The first was calculating the various statistical elements of the alignment and profile data for each track. MATLAB was used to determine the sample mean and standard deviation for the alignment and profile of each track [12], according to

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} \tag{3.1}$$

$$S_x = \sqrt{\frac{\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2}{n-1}} \tag{3.2}$$

The sample mean and standard deviation were calculated to provide an estimate of the mean and standard deviation of the entire population of tracks. This method proved inadequate for the class 5 and 6 tracks because we had data for only one class 5 and one class 6 track. Therefore, a second method for statistical analysis was established. Although the TTT is the only class 5 track and the RTT is the only class 6 track, they are over 48,000 feet (14,630 m) and 70,000 feet (21,336 m) long, respectively. In addition to their long length, the irregularities associated with the alignment and profile are normally distributed, as seen in the histogram of the TTT data in Fig. 3.4. The TTT and RTT where then divided up into short sections. Each section was then statistically evaluated to determine its characteristics. In fact, each track was long enough and sufficiently normally distributed to allow us to use this method for all tracks. Each track was divided into 1000-foot (305 m) sections, and the means and standard deviations were calculated for each section. These sections could then be used to compare the statistical properties of the different track classes. Finally, to obtain a representative statistical value for each class of track, the means and standard deviations of each track within a class were averaged together. These representative values are shown in Section 3.4.1 as the “1000-ft sections.”

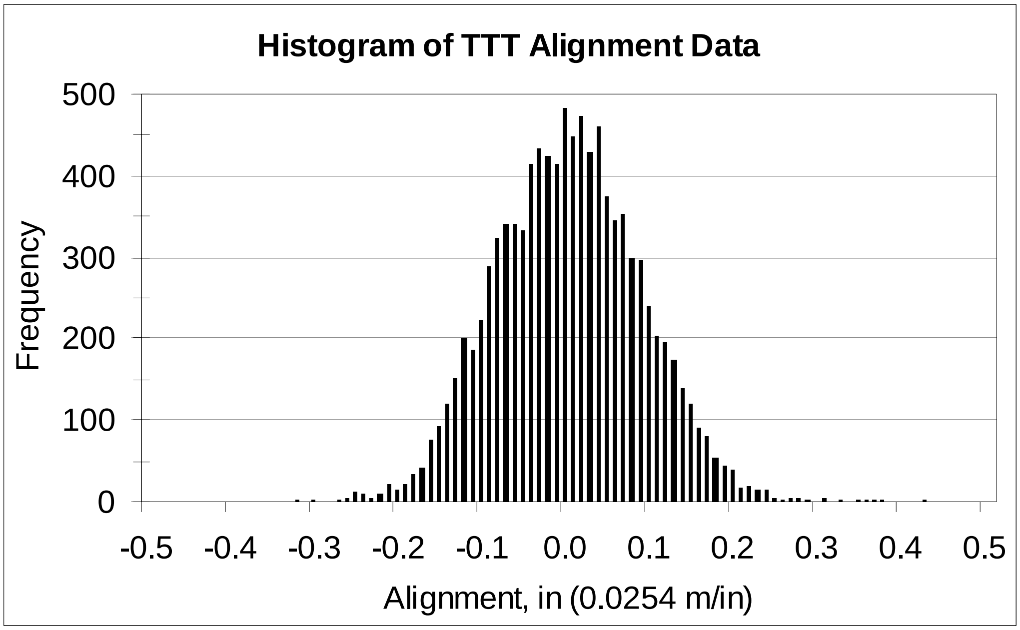


Figure 3.4 Histogram of Transit Test Track Alignment Data

3.3.2 Frequency Analysis

To determine if any similarities exist among the frequency compositions of the various tracks and track classes, a frequency analysis was performed. It was decided that an auto spectrum would best represent the data [13]. Several steps were necessary for performing the frequency analysis: choosing the number of windows and sample size, selecting a window shape and the amount of overlap, and performing a discrete Fourier transform (FFT) on the windows and calculating the single-sided auto spectrum. It was determined that at least twenty windows would be needed to calculate a reliable average, which was used, along with the fact that each track has a different size, to choose the sample size for each track. For the mid-size track, the HTL, a sample size of 1024 was selected, allowing for 27 averages. For the remaining tracks, the sample sizes were either a multiple or factor of 1024. For example, the Balloon Loop, the shortest track, had a sample size of 512 for 30 averages, and the RTT, the longest track, had a sample size of 4096 for 32 averages. This allowed each track to have a sufficient number of averages for an enhanced frequency trace. The sample rate of the data is 1 sample/ft, which yields a Nyquist frequency of .5 cycles/ft, or a wavelength of 2 feet (.6

m). Aliasing is avoided since the magnitudes of the frequencies above .5 cycles/ft are very small.

One problem that occurs while performing an FFT is leakage. Leakage is a distortion in the spectral representation of data caused by non-periodic functions in the sample time window. By applying time domain windows to the data, the data are forced to be periodic and leakage can be reduced. For this reason, a Hanning window was applied to the data prior to performing the FFT. A Hanning window is defined by

$$H = \sin^2(\pi * T) \tag{3.3}$$

Although applying a window, such as the Hanning window, reduces leakage, it can distort or eliminate some of the data points. One way of reducing this effect is by overlapping the data windows, as shown in Fig. 3.5. By using an overlap of 50%, we are able to minimize the data distortion.

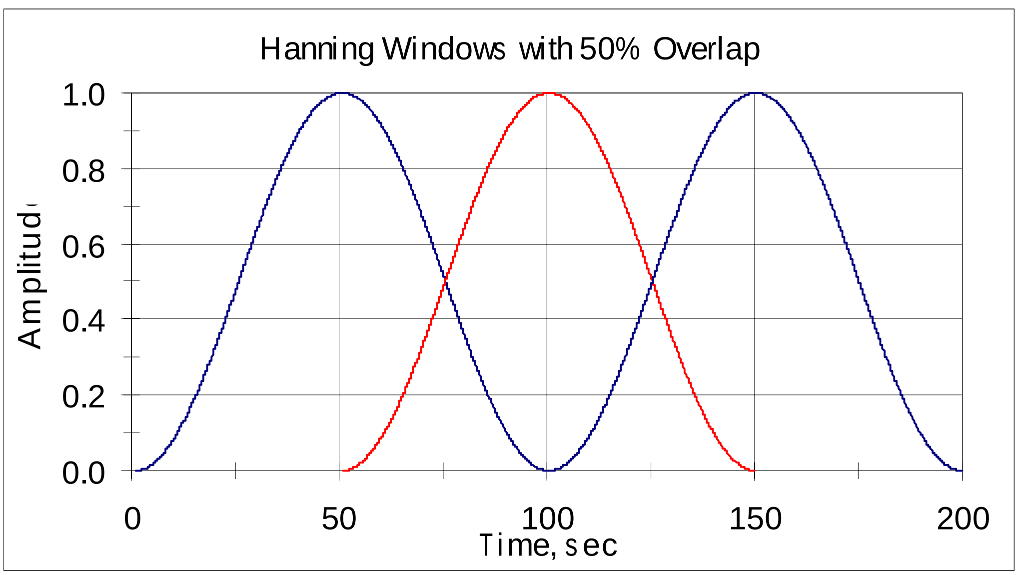


Figure 3.5 Three Hanning Windows with 50% Overlap

To examine the frequency content of each track, MATLAB was used to perform a discrete Fourier transform (FFT) on each window or block of data. Finally, the single-sided auto spectrum (or power spectrum) was calculated and averaged. The auto spectrum of a discrete frequency spectrum, such as $X_n(f)$, is defined as

$$G_{xx} = \frac{2}{N}\sum_{n=1}^{N} X_n^*(f)X_n(f) \tag{3.4}$$

where $X_n^*(f)$ indicates the complex conjugate of $X_n(f)$. The constant N denotes the number of windows or blocks, and the 2 is used to calculate the single-sided auto spectrum. This produces an auto spectrum with the units of $(\text{inches})^2$. When viewing the data, it is necessary to properly scale it. Equation 3.5 is used to scale the auto spectrum, where the constant nfft denotes the number of spectral lines [13].

$$A = \frac{2}{\text{nfft}}\sqrt{2 * Gxx} \tag{3.5}$$

The results were then plotted, as shown in Section 3.4.2, to check for common frequencies. The x-axis in the plots represents the number of samples per foot rather than samples per second (or Hz), which one is accustomed to seeing in a frequency spectrum. In this case, the sample rate of the recorded data is one sample per foot. The samples per foot can be thought of as spatial frequencies with units of cycles per foot (1/ft), which is the inverse of wavelength.

3.4 Results

After performing the analysis on the track data, the results were examined for any trends that may have existed. The statistical analysis determined common standard deviations for each class of track, while the frequency analysis showed that there were no common frequencies among the various tracks; rather, a common bandwidth existed for the frequency content of all classes of tracks.

3.4.1 Statistical Results

The statistical analysis provided some expected results. The means of the alignment and profile data for each class of track, shown in Table 3.1, are all zero. The representative values for the means of the 1000-foot sections, shown in Table 3.1, are also all zero.

Table 3.1 Track Class Means

Track Class	Alignment		Profile	
	Mean	Mean 1000-ft Sections	Mean	Mean 1000-ft Sections
4	0.00 in (0.00 cm)	0.00 in (0.00 cm)	0.00 in (0.00 cm)	0.00 in (0.00 cm)
5	0.00 in (0.00 cm)	0.00 in (0.00 cm)	0.00 in (0.00 cm)	0.00 in (0.00 cm)
6	0.00 in (0.00 cm)	0.00 in (0.00 cm)	0.00 in (0.00 cm)	0.00 in (0.00 cm)

Since the calculated means of the alignment and profile data are all zero, then the irregularities that exist in the alignment and profile of the tracks are all equally distributed about zero. This indicates that the important value obtained from the statistical analysis will be the standard deviation. The standard deviation is a measure of the deviation of the data from the mean value. Therefore, a small standard deviation means that the irregularities are smaller, where a larger standard deviation means that the irregularities are larger and more dispersed about the mean. The calculated standard deviations of the alignment and profile data for each track are shown in Table 3.2. Although the standard deviations of the profile data for the class 4 tracks are different, they are closer to each other when compared with the values for the class 5 and 6 tracks. This is also true for the alignment data. This difference across the track classes yields a distinctive value, or range of values, for the standard deviations of each class of track.

Table 3.2 Standard Deviations of Alignment and Profile of All Tracks

	Class 4			Class 5	Class 6
	Balloon Loop	HTL	WRM	TTT	RTT
Alignment	0.30 in (0.76 cm)	0.26 in (0.66 m)	0.23 in (0.58 cm)	0.11 in (0.28 cm)	0.10 in (0.25 cm)
Profile	0.16 in (0.41 cm)	0.17 in (0.43 cm)	0.14 in (0.36 cm)	0.11 in (0.28 cm)	0.07 in (0.18 cm)

To obtain a representative value for the standard deviation for each class, the standard deviations for each track in each class were averaged together. The standard deviations of the alignment and profile data for each track class are shown in Table 3.3. Also shown in Table 3.3 are the values for the standard deviation for the 1000-foot (305 m) sections. These values are slightly lower than the values obtained over the whole track. This can be explained by the fact that when dividing the tracks into the 1000-foot (305 m) sections, they did not divide evenly. Consequently, the data that were not included in the sections was not used.

The data verifies common expectations from different classes of track regarding the track geometric irregularities. For instance, it is observed that the standard deviations for different parameters decrease with track class, indicating smaller irregularities for higher classes of tracks, as is fully anticipated.

Table 3.3 Track Class Standard Deviations

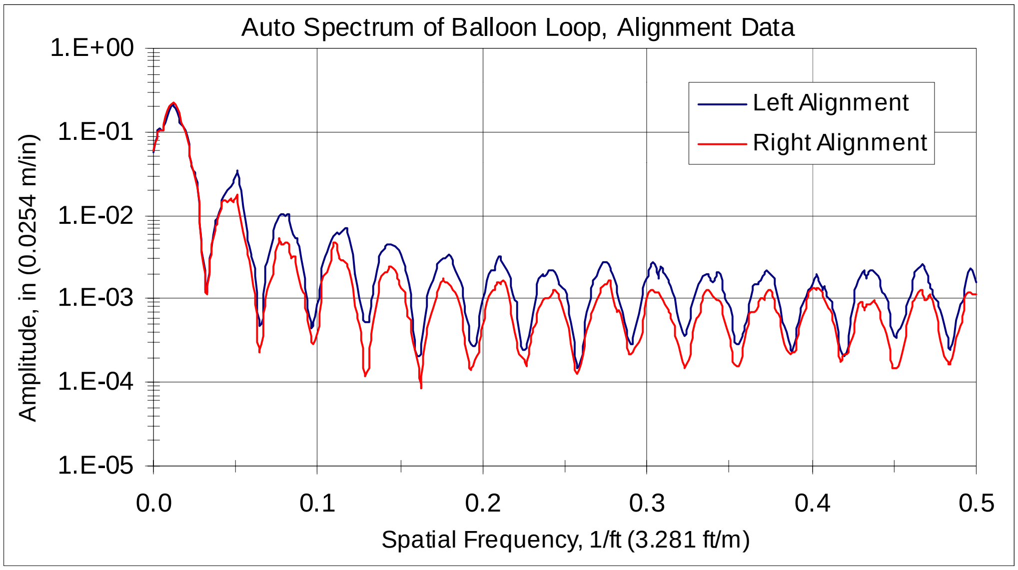
Track Class	Alignment		Profile	
	Standard Deviation	Standard Deviation 1000-ft Sections	Standard Deviation	Standard Deviation 1000-ft Sections
4	0.26 in (0.66 cm)	0.25 in (0.64 cm)	0.15 in (0.38 cm)	0.15 in (0.38 cm)
5	0.11 in (0.28 cm)	0.10 in (0.25 cm)	0.11 in (0.28 cm)	0.10 in (0.25 cm)
6	0.10 in (0.25 cm)	0.09 in (0.23 cm)	0.07 in (0.18 cm)	0.06 in (0.15 cm)

3.4.2 Frequency Results

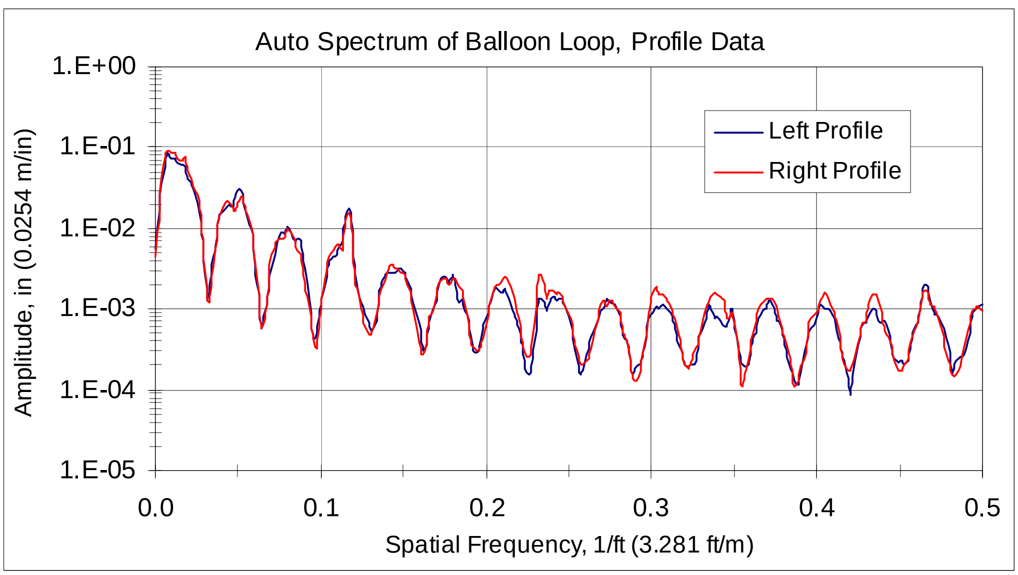
The results generated from the frequency analysis are shown in Figs. 3.6-3.10. Each figure consists of two parts, one presenting the auto spectrum of the alignment data and the other showing the auto spectrum of the profile data. The x-axis corresponds to spatial frequency with units 1/ft, while the y-axis corresponds to the magnitude of the auto spectrum with units of inches. The y-axis is displayed in logarithmic scale so that all the plots may be viewed with the same scaling.

Several observations can be made from the frequency results. The scalloped effect, or humps, evident in the figures will be explained in Section 3.4.3. Most notable is that the peaks of the auto spectrums that are above the noise floor are grouped within a common area. In other words, the frequency content of each plot appears to be contained within a common bandwidth. The auto spectra for the right and left tracks are almost the same indicating nearly identical irregularities for each.

Finally, it can also be noted that the amplitudes of the peaks are smaller for the TTT and smaller still for the RTT. This is expected considering that the TTT and RTT are higher-class tracks, and thus have smaller irregularities. This reduction in the amplitudes of the irregularities was also observed in the statistical analysis with the reduction in standard deviations.



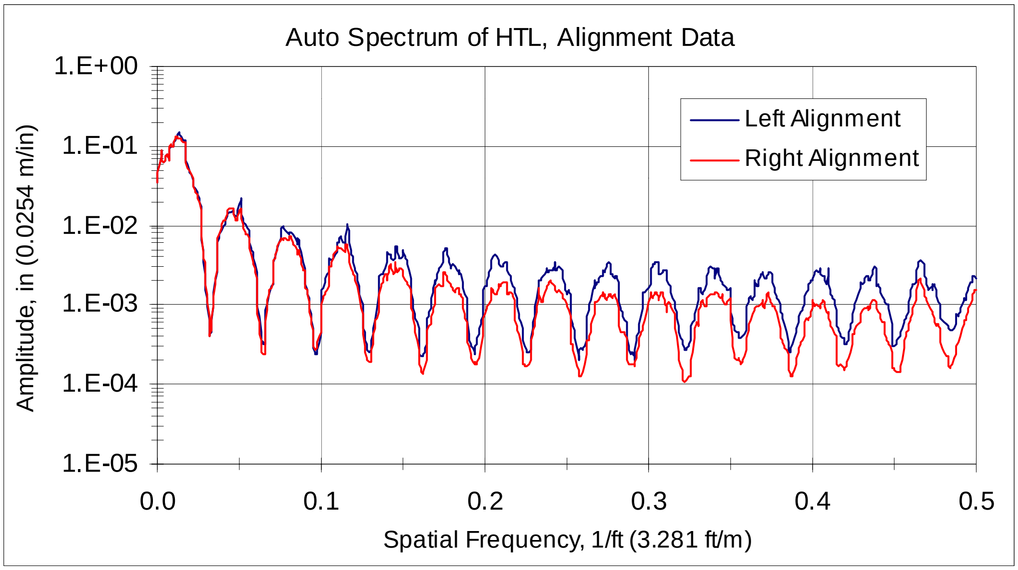
(a)



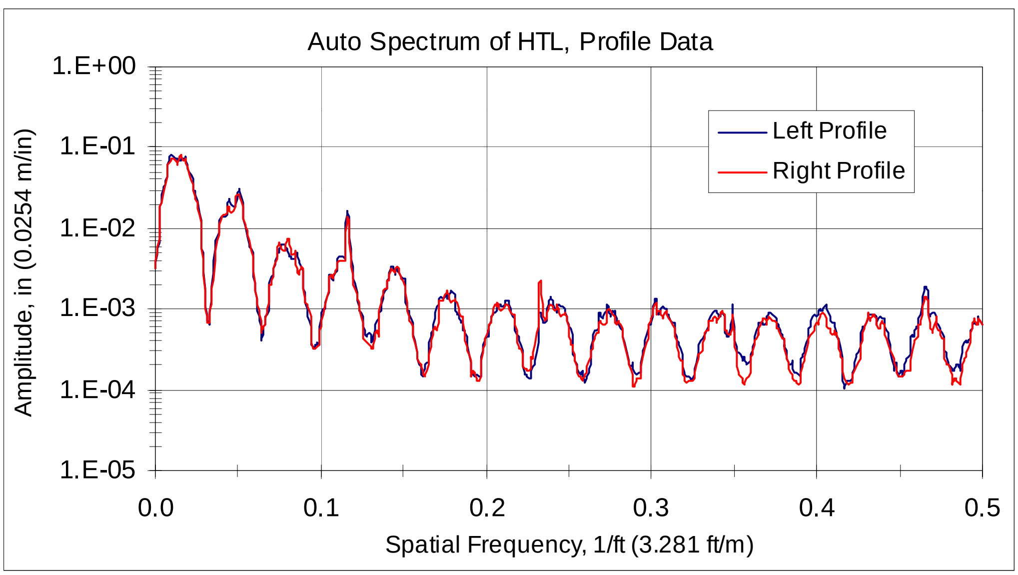
(b)

Figure 3.6 Auto Spectrum of Balloon Loop Track Geometry Data;

(a) Alignment Data, (b) Profile Data



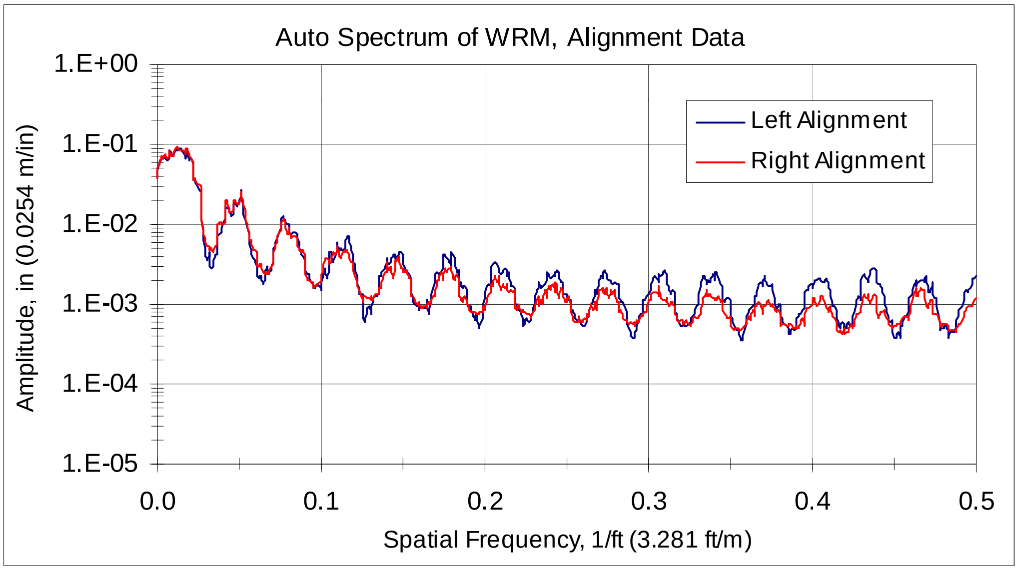
(a)



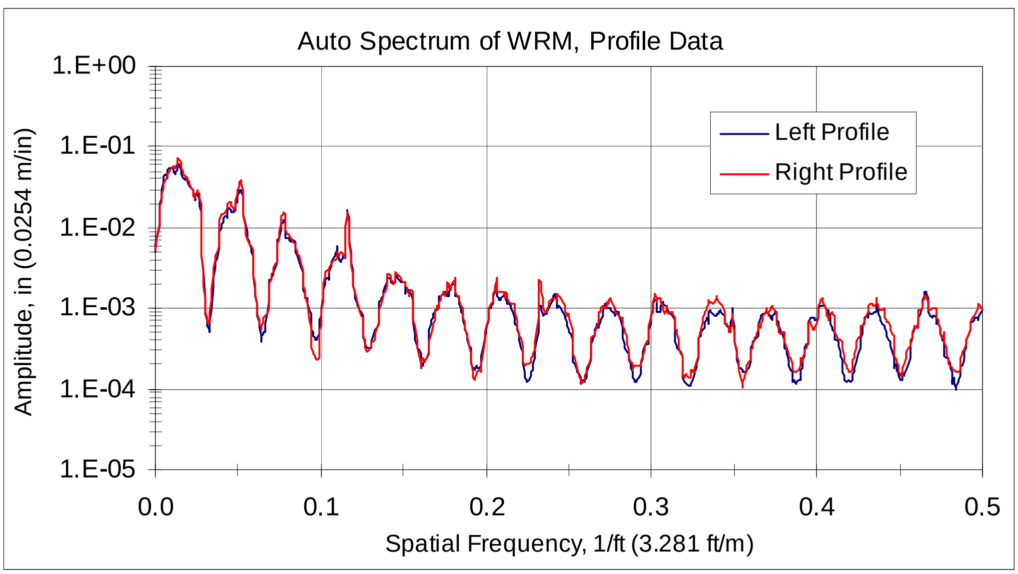
(b)

Figure 3.7 Auto Spectrum of HTL Track Geometry Data;

(a) Alignment Data, (b) Profile Data



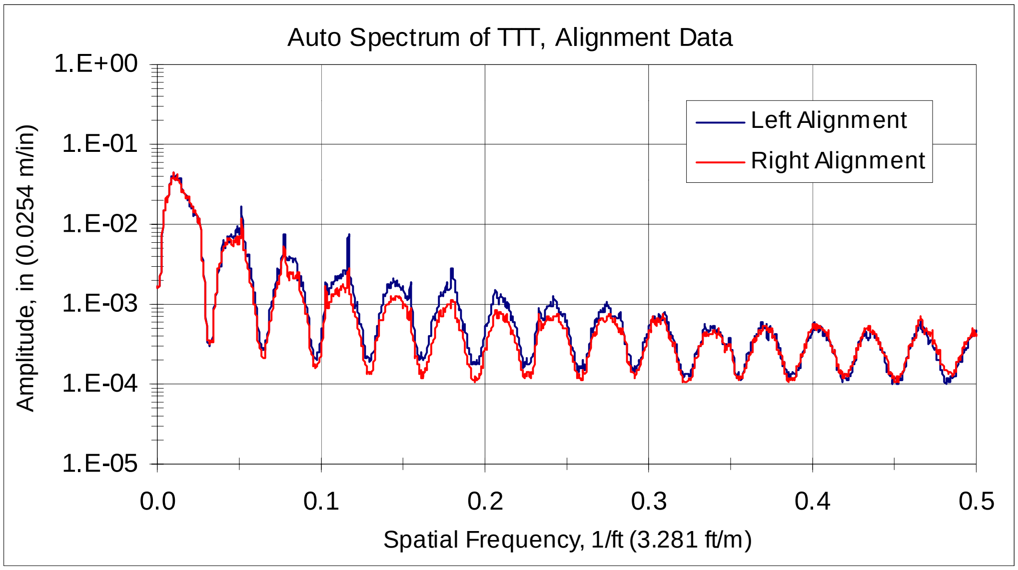
(a)



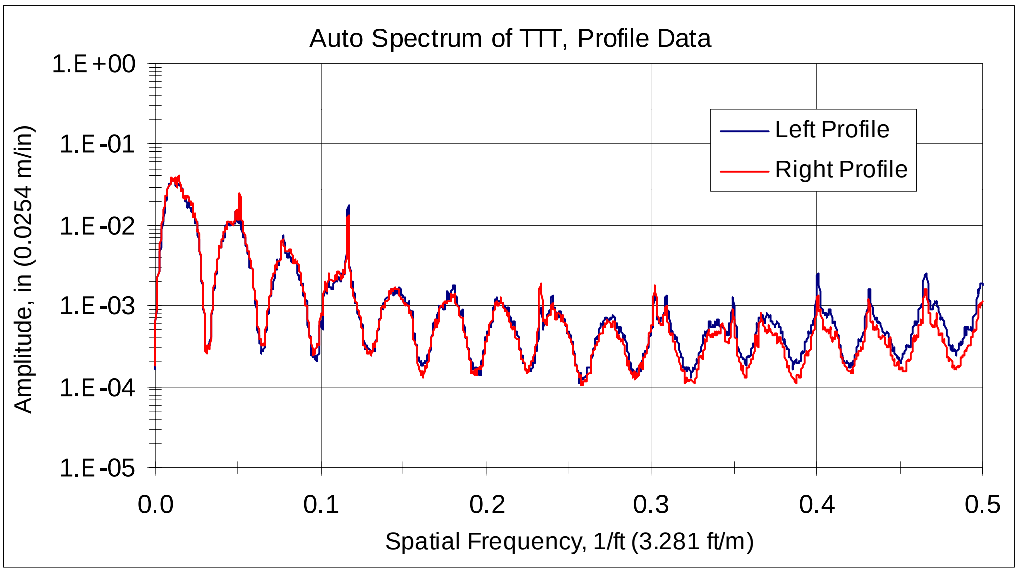
(b)

Figure 3.8 Auto Spectrum of WRM Track Geometry Data;

(a) Alignment Data, (b) Profile Data



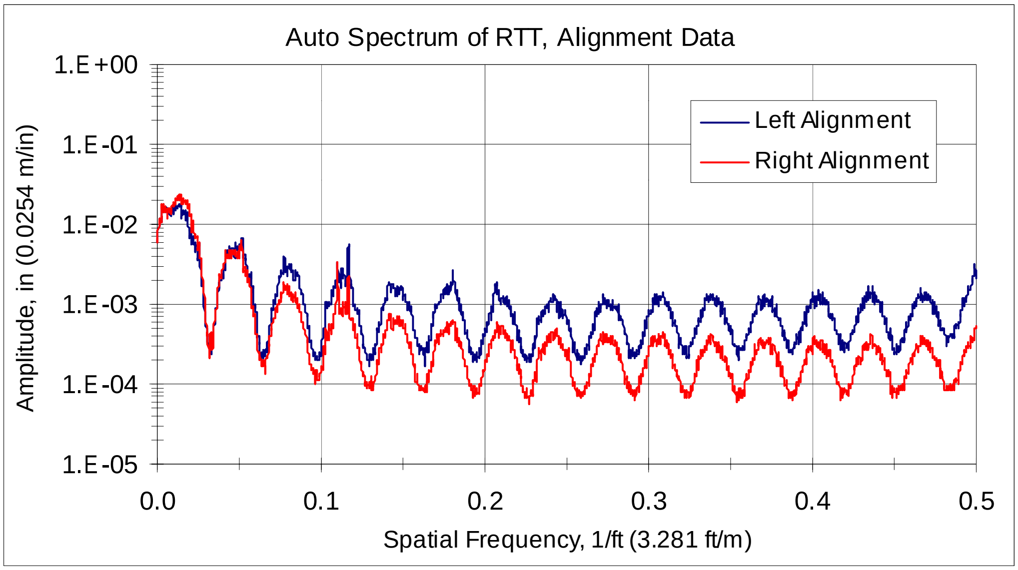
(a)



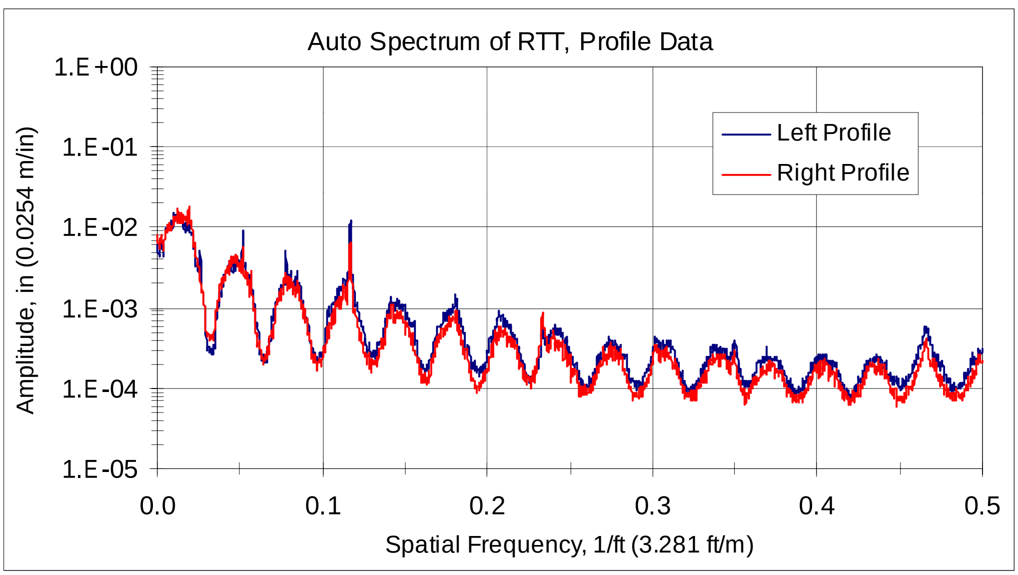
(b)

Figure 3.9 Auto Spectrum of TTT Track Geometry Data;

(a) Alignment Data, (b) Profile Data



(a)



(b)

Figure 3.10 Auto Spectrum of RTT Track Geometry Data;

(a) Alignment Data, (b) Profile Data

The next step is to compare the frequency content of tracks of the same class. Figures 3.11-3.14 display the auto spectra for the left alignment, right alignment, left profile, and right profile data. Each figure shows a comparison of all of the class 4 tracks: the Balloon Loop, the HTL, and the WRM. The plots are used to highlight any similarities in the frequency content of different tracks with the same class.

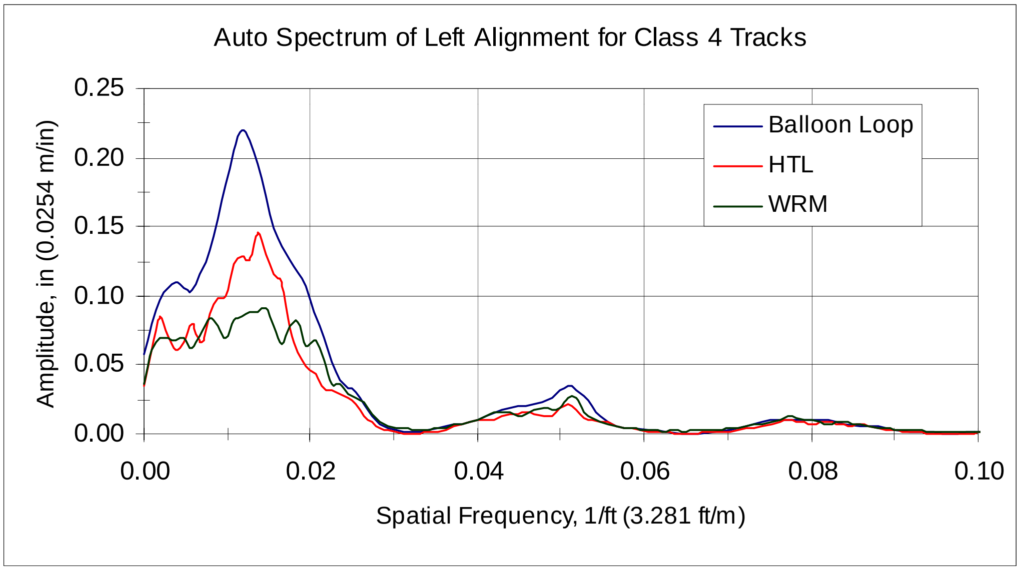


Figure 3.11 Comparison of Auto Spectrum of Left Alignment for Class 4 Tracks

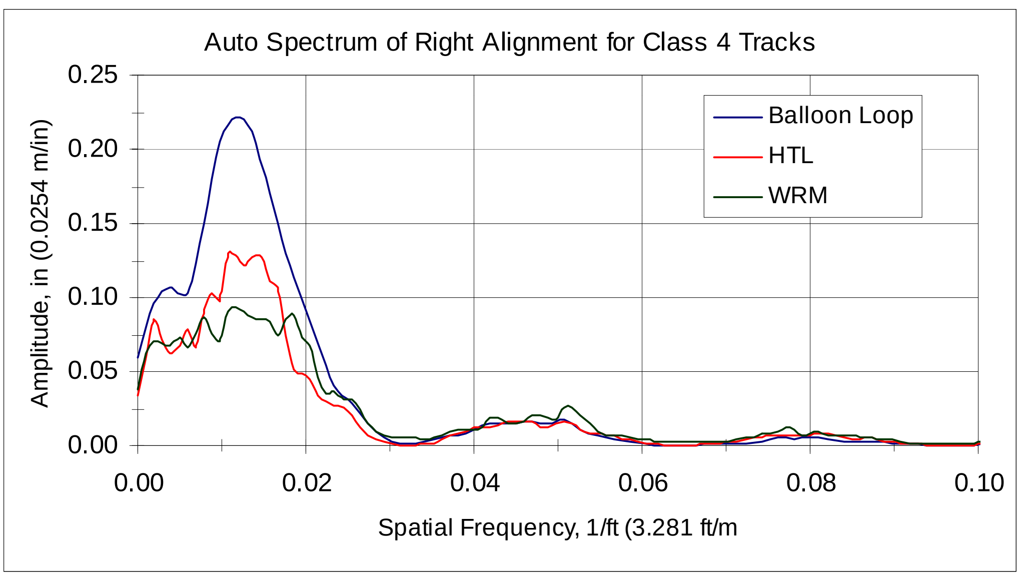


Figure 3.12 Comparison of Auto Spectrum of Right Alignment for Class 4 Tracks

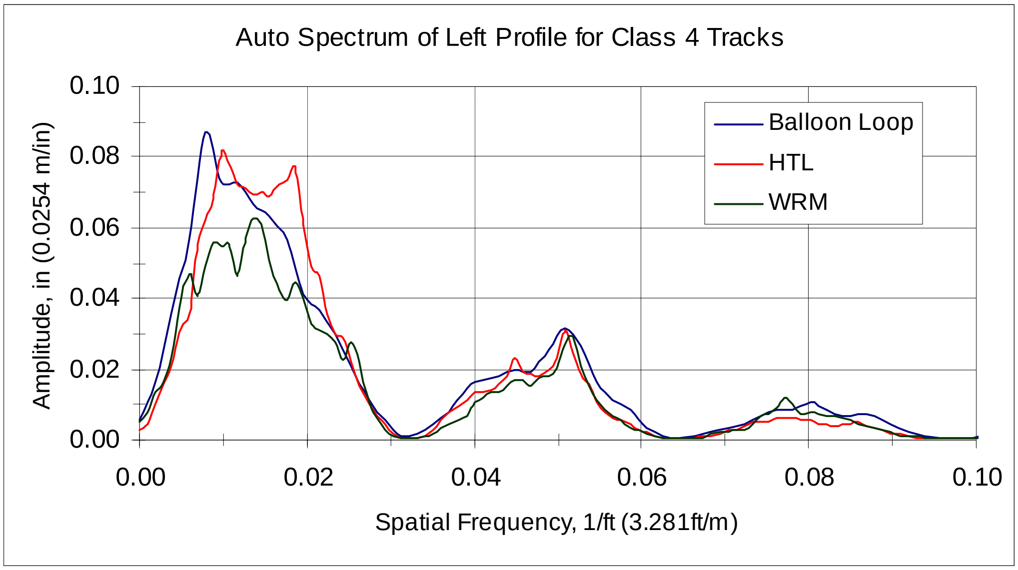


Figure 3.13 Comparison of Auto Spectrum of Left Profile for Class 4 Tracks

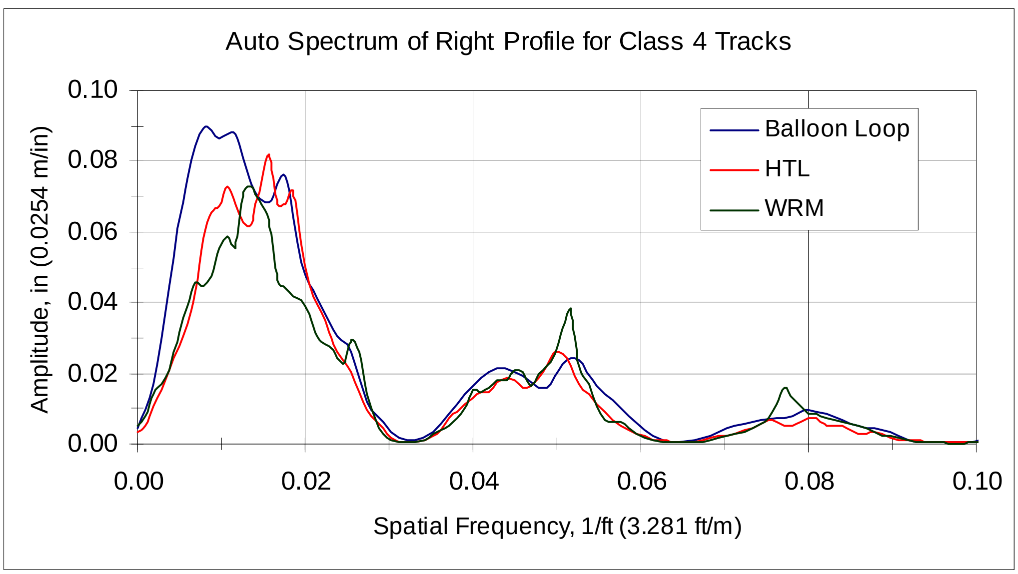


Figure 3.14 Comparison of Auto Spectrum of Right Profile for Class 4 Tracks

The plots clearly show that there are no common peaks among the class 4 track data. At first glance, it may appear as if there is a common peak for the left alignment at approximately 0.01 1/ft (0.03 1/m) for the Balloon Loop and HTL. A closer examination of the peak, however, reveals that the peak occurs at 0.01172 1/ft (0.03845 1/m) for the Balloon Loop and at 0.01367 1/ft (0.04485 1/m) for the HTL. These spatial frequencies correspond to wavelengths of 85.3 ft (26.0 m) and 73.2 ft (22.3 m), respectively, which are significantly different. Recognizing that the spectral resolution of the plots (i.e., the smallest discernible frequency) is 0.00195 1/ft (0.00640 1/m) for the Balloon Loop, and 0.00098 1/ft (0.00322 1/m) for the HTL, it is reasonable to conclude that the peaks are different. As observed before, there does appear to be a common bandwidth in which the peaks are contained. The bandwidth appears to be from 0.00 1/ft (0.00 1/m) to 0.03 1/ft (0.098 1/m), or comprising wavelengths larger than about 33 feet (10 m).

Finally, we compared the frequency contents of all of the tracks, as shown in Figs. 3.15-3.18. Each figure shows a comparison of the auto spectra for the Balloon Loop, the HTL, the WRM, the TTT, and the RTT.

Similar to the previous plots, there are no common peaks, or frequencies, in the various tracks, although a common frequency bandwidth is again apparent. Since there are no common frequencies among tracks of the same class or all classes, it can be concluded that the irregularities associated with the alignment and profile data are random. This is further validated by the normal distribution observed in the histogram plot of the irregularities, shown in Fig. 3.4.

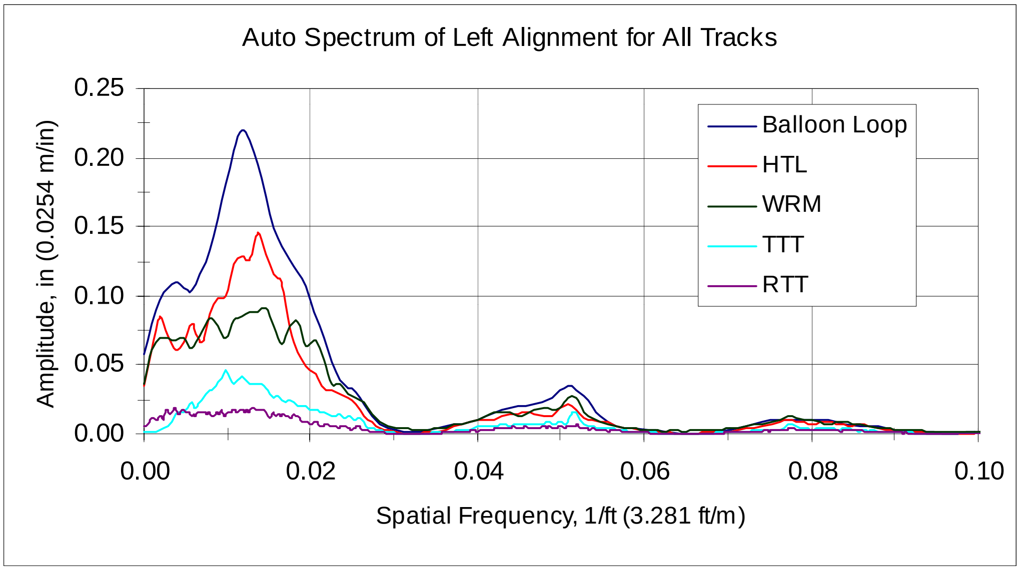


Figure 3.15 Comparison of Auto Spectrum of Left Alignment for All Tracks

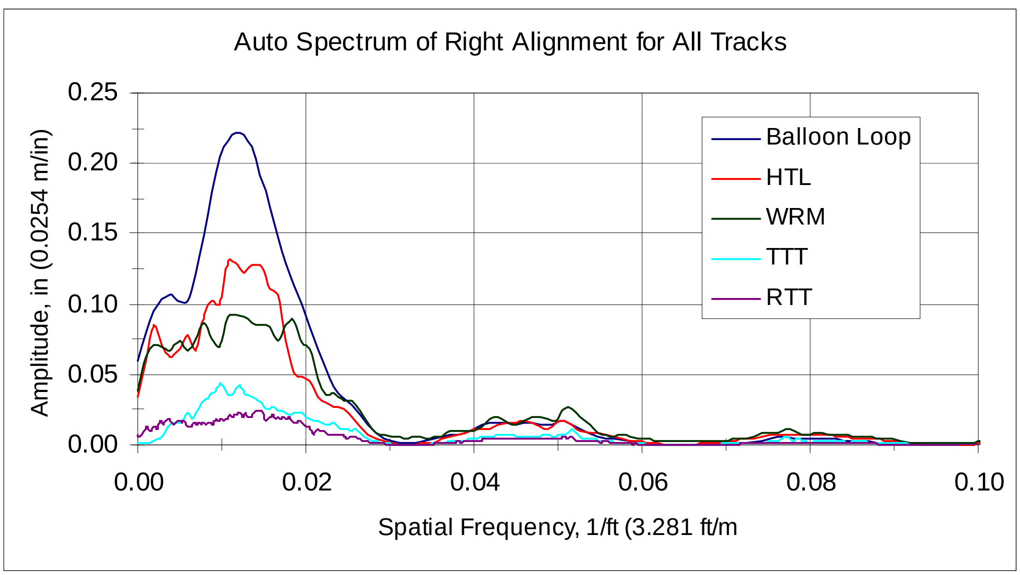


Figure 3.16 Comparison of Auto Spectrum of Right Alignment for All Tracks

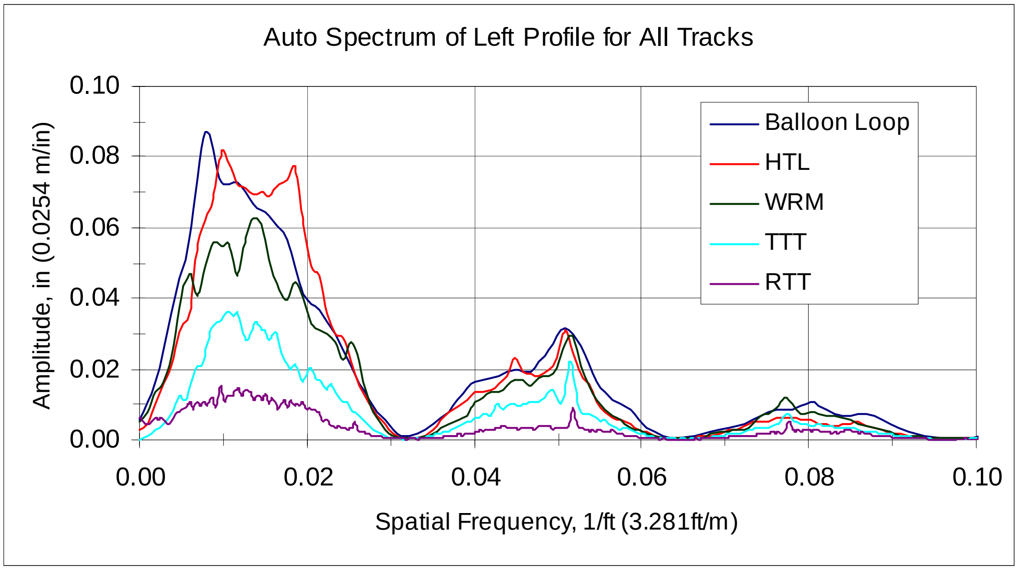


Figure 3.17 Comparison of Auto Spectrum of Left Profile for All Tracks

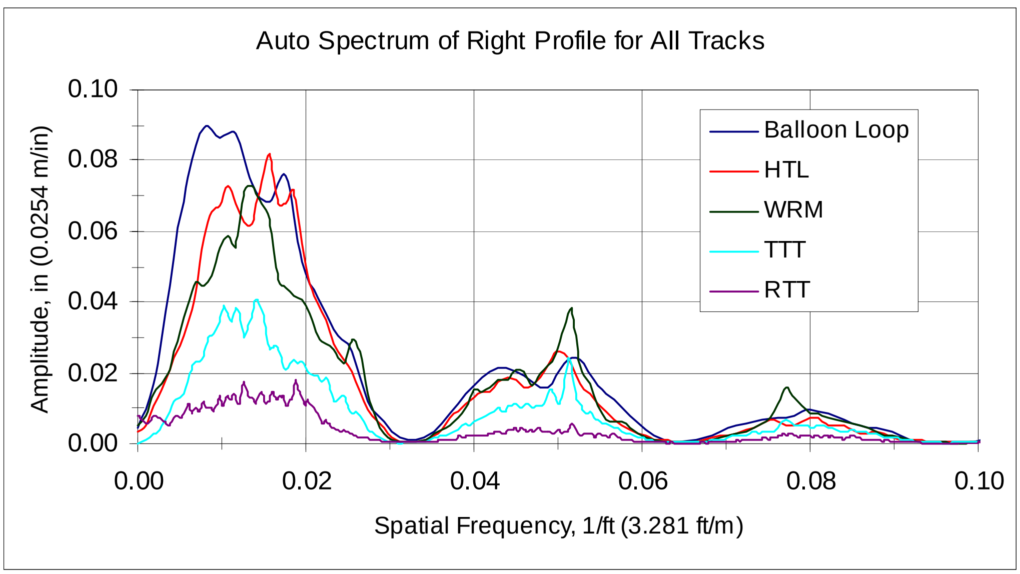


Figure 3.18 Comparison of Auto Spectrum of Right Profile for All Tracks

3.4.3 Effect of Data Averaging on the Frequency Plots

The track data that were used in this study was collected by ENSCO, Inc. The data were collected using their track geometry vehicle recording one data point every foot. The track geometry parameters were recorded using the 62-foot mid-cord offset (MCO) method. This method determines the deviation of the track by averaging the displacement of the track at two points, 62 feet (18.9 m) apart. This average value is then subtracted from the displacement of the track at the point midway between the two other points, as defined by Eq. (3.6). The track deviations were calculated for each track geometry parameter using this method.

$$\mathrm{deviation}_{i+31} = \mathrm{disp}_{i+31} - \left(\frac{\mathrm{disp}_{i} + \mathrm{disp}_{i+62}}{2} \right) \tag{3.6}$$

The MCO data collection method has a dramatic effect on the frequency content of the data. By averaging two points that are 62 feet apart, the magnitudes of all the wavelengths that are even sub-multiples of 62 ft (18.9 m), such as 31 ft (9.4 m) and 15.5 ft (4.7 m), are attenuated to zero. This occurs because the averaged value of the two points will always equal the displacement of the track at the midpoint, when the wavelengths are sub-multiples. Figure 3.19 shows an example signal with a 31 ft wavelength. A 62 ft cord is extended across the signal, where points A and B are the end points, and point C is the midpoint. The average of A and B is point C, which in this case, is located on the signal. If point C is subtracted from the value of the signal at that point, the resulting deviation will be zero.

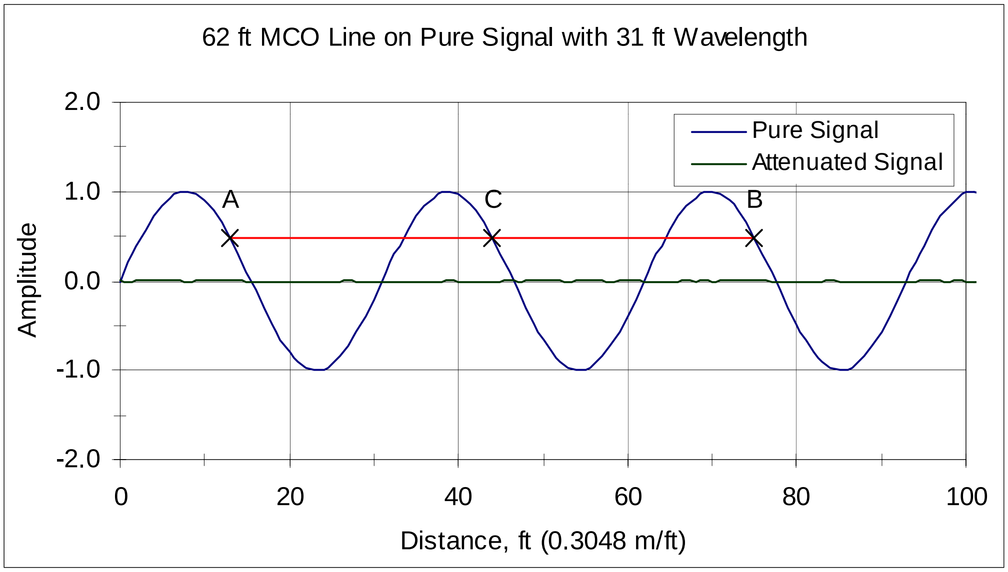


Figure 3.19 Result of 62 Foot Mid-Cord Offset Method on Signal with 31 Foot Wavelength

The MCO method also attenuates wavelengths larger than 62 ft. As the wavelength gets larger, the 62 ft cord becomes smaller with respect to the signal, and the difference between the actual signal and the midpoint C becomes smaller. This can be seen in Fig. 3.20, where a 62 ft cord is placed on a signal with a wavelength of 350 ft (106.7 m). The distance between point C and the signal is very small resulting in an attenuated signal.

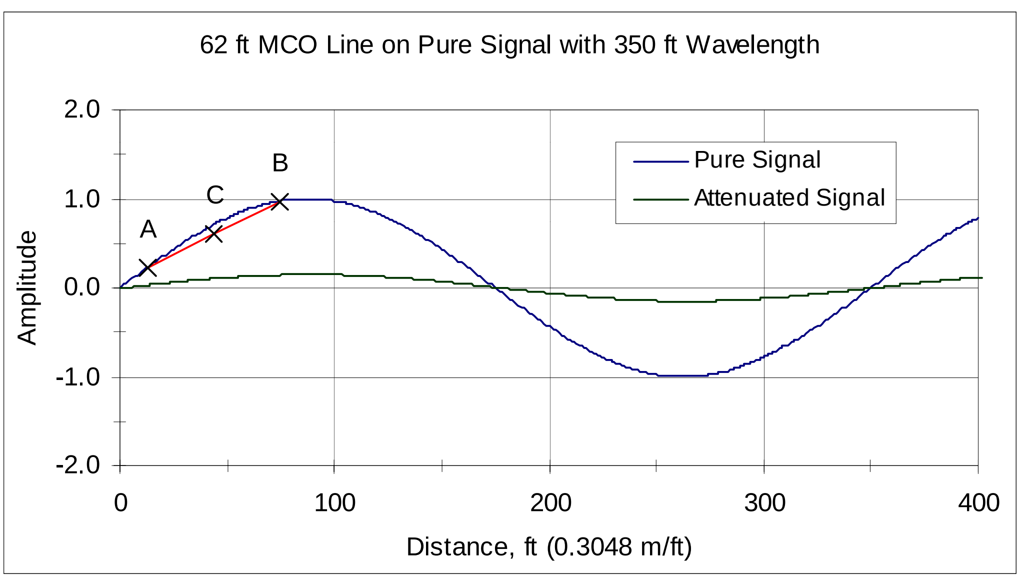


Figure 3.20 Result of 62 Foot Mid-Cord Offset Method on Signal with 350 Foot Wavelength

While attenuating some wavelengths, the MCO method will also add gain to some wavelengths. If the wavelength is an odd sub-multiple of 62 ft, such as 62 ft and 20.7ft, the difference between the actual signal and the point C, will be twice the original signal, as shown in Fig. 3.21.

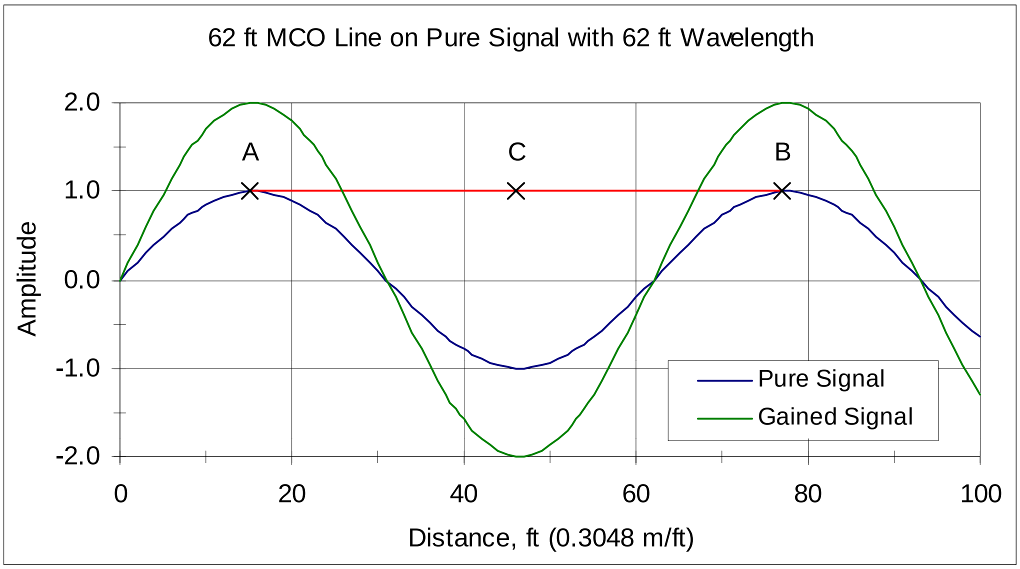
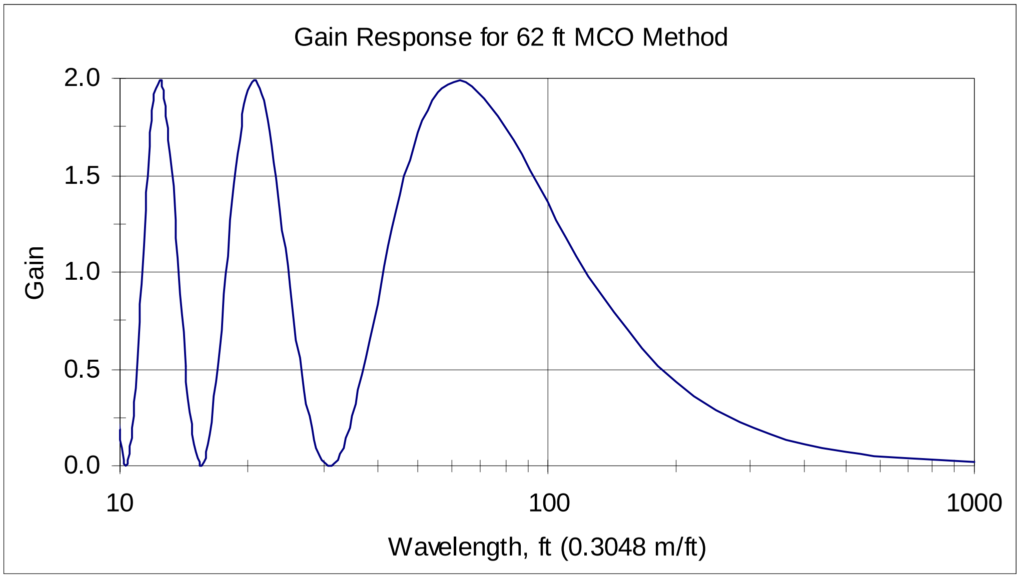
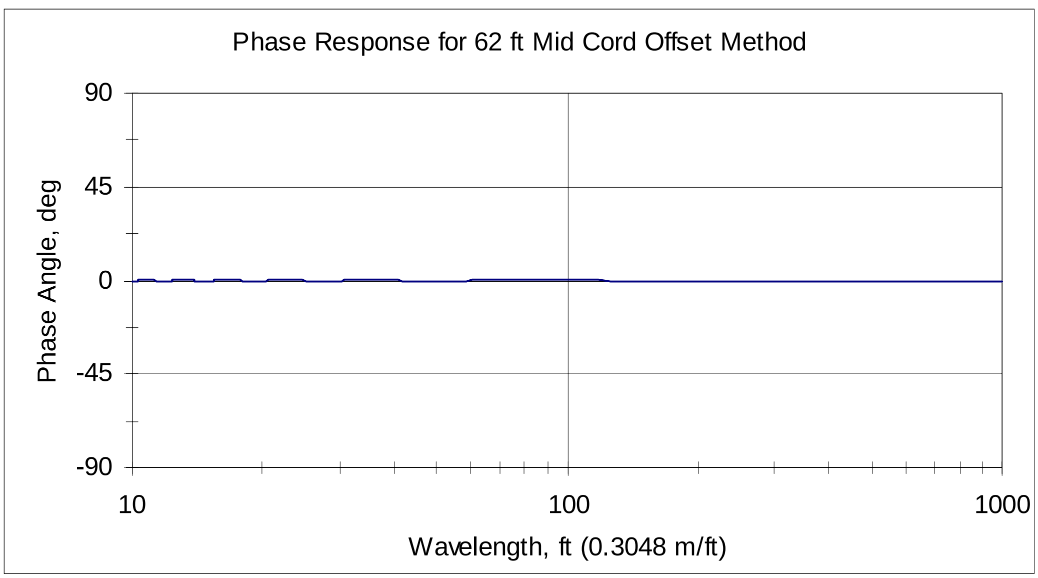


Figure 3.21 Result of 62 Foot Mid-Cord Offset Method on Signal with 62 Foot Wavelength

These characteristics of the MCO method attenuate some wavelengths from the resulting signal and add gain to other wavelengths causing the scalloping effect observed in the frequency plots of the measured track data. The gain and phase responses of the MCO method are shown in Fig. 3.22. Although the MCO method does not affect the phase, it does have notable effect on the gain response. The characteristics show that the phenomena observed in the frequency plots of the measured track data shown in Figs. 3.6-3.18, is a result of the MCO data collection technique.



(a)



(b)

Figure 3.22 Characteristics of 62 Foot Mid-Cord Offset Method; (a) Gain Response, (b) Phase Response

3.5 Representing Perturbations

By combining the results from the statistical analysis and frequency analysis, a method for representing the perturbations, or irregularities, in the software was established. The statistical analysis provided standard deviations for the alignment and profile inputs for each class of track, while the frequency analysis indicated the frequency content for the random data. Using this information, an algorithm for creating and applying the track irregularities was devised.

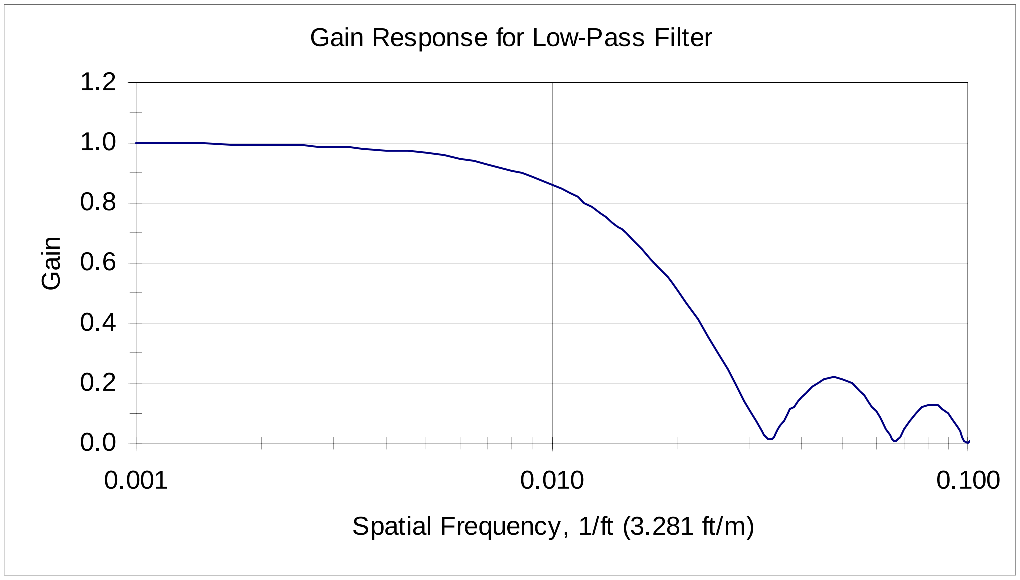
The first step was creating an array of normally distributed, random data to provide a base for which the remaining operations could be performed. The array size was selected to be dependent on the length of the track to be created. The data were then filtered to limit their frequency content to the bandwidth that we observed in our earlier data analysis.

We selected a simple filtering technique, using a moving average method. The moving average is a digital low-pass filter defined by Eq. (3.7). The cutoff frequency of the filter is determined by the number of values, N, included in the average. The cutoff frequency corresponds to where the magnitude of the signal has been attenuated by 0.7071. The filter works by summing N subsequent values and applying equal weighting of 1/N to each value. This is a very basic form of a finite impulse response (FIR) filter.

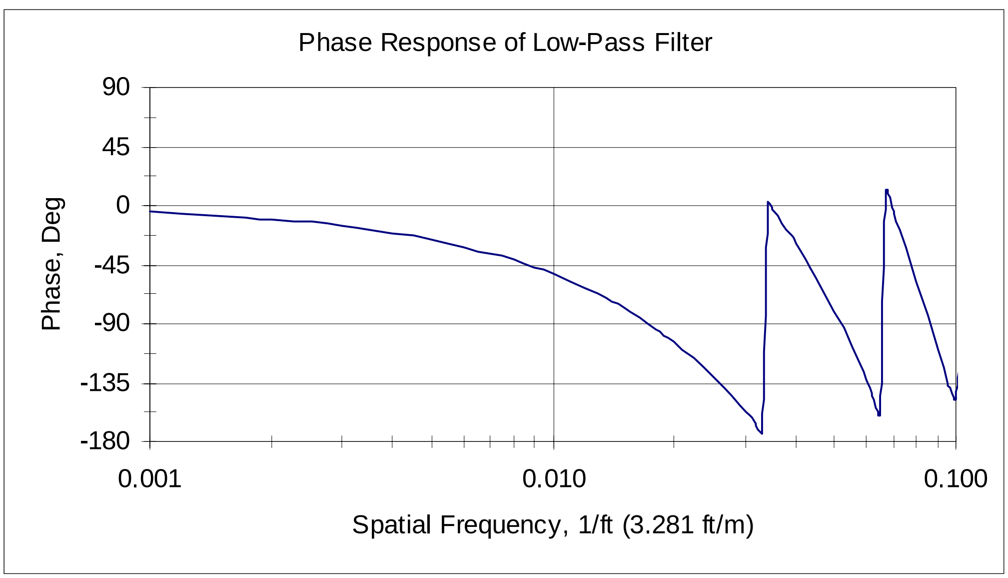
$$X_j = \frac{1}{N}\sum_{i=1}^{N} x_{(j+i-1)} \tag{3.7}$$

For the bandwidth of 0.0 to 0.03 cycles per foot (0.00 to 0.098 cycles per meter), thirty values (N=30) provided the best combination of filtering the data while still maintaining distinct frequencies.

The gain and phase responses of the filter are shown in Fig. 3.23. The responses were found by running signals with known amplitude and frequency through the filter. The gain response was determined by calculating the ratio of the amplitude of the filtered signal over the amplitude of the unfiltered signal. The ratio corresponded to the gain response of the filter at that frequency. The phase response was determined by calculating the phase angle of the filtered signal relative to the phase angle of the unfiltered signal, at the frequency of the signal. The gain response shows that the cutoff frequency of the filter is approximately 0.015 1/ft (0.049 1/m).



(a)



(b)

Figure 3.23 Characteristics of Finite Impulse Response Low-Pass Filter; (a) Gain Response, (b) Phase Response

Once the data were filtered, the appropriate standard deviation for the selected track class was applied to it by dividing the data by its standard deviation and multiplying the result by the standard deviation for the desired track. Next, the data were mean-zeroed to reflect what was observed in the track measurements.

The created track irregularities were now ready to analytically represent the actual measured track for NUCARS modeling.