

LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY  
- LIGO -  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

<b>Document Type</b>	<b>LIGO-T060011-00-Z</b>	2006/03/30
<b>PSU Data Conditioning, S2-S4</b>		
Tiffany Summerscales, Mike Ashley, Bethany Bongiorno, Lee Samuel Finn, Eric Rothhoff, Keith Thorne, Matthew Tibbits, Daniel Tomcavage, Kristina Zaleski		

*Distribution of this draft:*  
LIGO Scientific Collaboration

**California Institute of Technology**  
**LIGO Project - MS 51-33**  
**Pasadena, CA 91125**  
Phone (626) 395-2129  
Fax (626) 304-9834  
E-mail: [info@ligo.caltech.edu](mailto:info@ligo.caltech.edu)

**Massachusetts Institute of Technology**  
**LIGO Project - MS 20B-145**  
**Cambridge, MA 01239**  
Phone (617) 253-4824  
Fax (617) 253-7014  
E-mail: [info@ligo.mit.edu](mailto:info@ligo.mit.edu)

WWW: <http://www.ligo.caltech.edu/>

## Abstract

We present the data conditioning procedure used by the Penn State group in the burst search pipeline for the BlockNormal event trigger generator. This data conditioning procedure was used in the analysis of S2, S3 and S4 data. The procedure consists of several steps which whiten the data and break it up into several non-overlapping frequency bands. For the case of S3 we calculate a figure of merit for all frequency bands and show that the data conditioning procedure is successful in improving the whiteness of the data most of the time.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Downsampling</b>	<b>3</b>
<b>3</b>	<b>Line removal with a Kalman filter</b>	<b>5</b>
3.1	Application to Narrow-Band LIGO Features . . . . .	8
3.2	Calculating $A$ , $C$ , $W$ , and $V$ . . . . .	10
<b>4</b>	<b>Shaping Filter</b>	<b>13</b>
<b>5</b>	<b>Basebanding</b>	<b>17</b>
<b>6</b>	<b>Line Regression</b>	<b>20</b>
<b>7</b>	<b>Final Whitening</b>	<b>22</b>
<b>8</b>	<b>Startup and Ending Transients</b>	<b>24</b>
<b>9</b>	<b>Effectiveness of Data Conditioning</b>	<b>25</b>
<b>A</b>	<b>Kalman filtered line parameters</b>	<b>32</b>
<b>B</b>	<b>Frequency Band Definitions and Filters</b>	<b>43</b>
<b>C</b>	<b>Line Regression Parameters</b>	<b>45</b>
<b>D</b>	<b>Data conditioning PSDs</b>	<b>49</b>
<b>E</b>	<b>Rayleigh FOM Distributions</b>	<b>65</b>

## 1 Introduction

In the search for gravitational waves, we do not want to be restricted to those solely from well modeled sources like neutron star inspirals. We also want to keep our eyes open to the unexpected and to sources that are not as well modeled, such as core-collapse supernovae and black hole mergers. The Burst Upper Limit Group has been charged with this difficult task of looking for

gravitational waves of unknown character. The only strategy capable of finding any short-duration signal that may be present is to search for places where the detector data differ in a statistical fashion from the noise.

The LIGO data analysis group at Penn State, which participates in the search for unmodeled bursts, has developed an algorithm for finding sections of data that have different statistics from the noise. The algorithm is called BlockNormal [1]. BlockNormal assumes that the data are normally distributed and breaks the data into blocks of time where the data can be described by a single mean and variance. If the mean or variance of a block of data exceeds a certain threshold, those data are flagged as possibly containing a gravitational wave event.

It is assumed that the noise that BlockNormal is characterizing has no correlations, or in other words is white. White data are data that have equal power at all frequencies. However, the data produced by the LIGO interferometers are far from white (see Figure 1). There are many sharp features in frequency (lines) due to instrumental resonances. For example, the wires suspending each of the mirrors vibrate at their resonant frequencies (also known as violin modes since the wires are vibrating like the strings of a violin). Consequently, the mirrors are moved at these violin modes, changing the arm lengths and injecting power at these frequencies into the data. There are features due to the 60Hz oscillations of the U.S. power grid and there are mirror oscillations that are intentionally injected for calibration purposes. In addition, the general shape of the noise in frequency space is dominated by seismic noise which increases with low frequency and an increase in noise with high frequency that is related to shot noise (caused by the discrete nature of photons).

In order to remove these instrumental and environmental artifacts and produce whitened data suitable for BlockNormal, we have designed a data conditioning procedure. This procedure not only removes lines and whitens the data but breaks the data into frequency bands. Breaking the data into frequency bands simplifies the whitening process by reducing the number of features that must be dealt with at any one time. It also means that when BlockNormal identifies a possible event, we know the range of frequencies to assign to the event based on the band it was found in.

The data conditioning procedure is described in detail in the following sections along with the ordering necessary to simplify successive steps. The procedure was applied in its entirety to LIGO data from both the second and third science runs (S2 and S3). A shortened version that excluded line removal was applied to the data from the fourth science run (S4). The conditioned LIGO data from these science runs was analyzed by BlockNormal.

## 2 Downsampling

LIGO data are sampled at 16384 Hz, resulting in very large amounts of data that are computationally expensive to whiten. Therefore, the first step in our data conditioning procedure is to downsample the data. Downsampling reduces the sampling rate (number of data samples per second) by a constant factor. The downsampling applied to the S2-S4 data reduced the sample rate from 16384Hz to 4096Hz. This significantly reduces the size of the data which can greatly reduce the computational cost of the next, Kalman filtering step as well as simplify the highpass shaping filter described in Section 4. Since discretely sampled data contains information about frequencies up to half the sample rate (half the sample rate is called the Nyquist frequency), downsampling will throw away some high frequency information. We therefore select the highest frequency that is of interest (2048 Hz) and downsample the data to set the new Nyquist to that highest frequency. Dis-

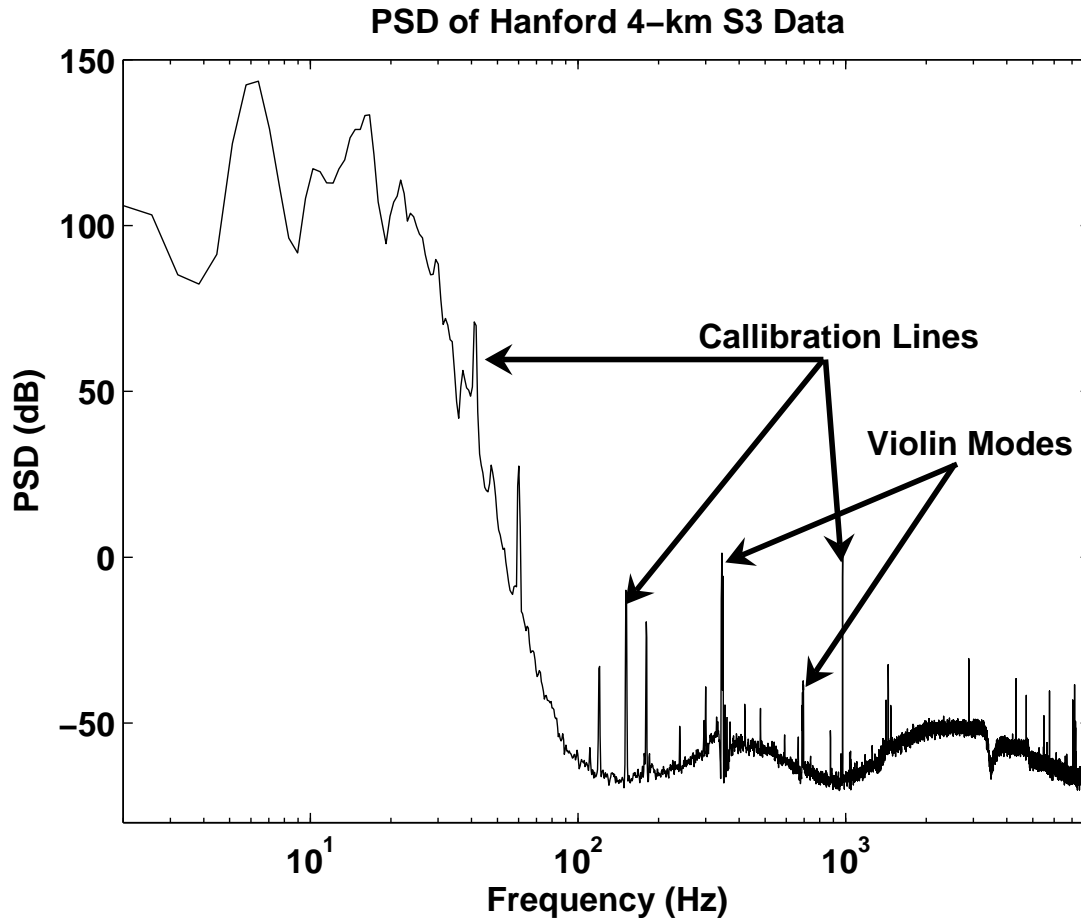


Figure 1: Power Spectral Density (PSD) of data from the Hanford, 4-km detector taken during the third science run (S3). There is a large increase in power at low frequencies caused by seismic noise as well as many narrow frequency features. Some of the narrow “lines” are caused by violin modes, which are mirror suspension wire resonances and their harmonics. Some lines are the result of the mirrors being moved at certain frequencies for calibration purposes. Most of the remaining lines are power lines which are harmonics of the 60 Hz oscillations of the U.S. power grid.

carding information at very high frequencies does not reduce our chances of seeing gravitational waves much because the detectors are less sensitive at these frequencies [2, 3, 4, 5].

The process of downsampling involves more than throwing away data points to achieve a smaller number of samples. First, a lowpass filter that suppresses the frequencies above the new Nyquist frequency must be applied to prevent aliasing. Aliasing can occur when a high frequency component of a signal is under-sampled. When this happens, there are not enough samples to distinguish between a signal at a frequency less than or equal to Nyquist and one that is above Nyquist so that the power at above-Nyquist frequencies is erroneously attributed to lower frequencies (see Figure 2). Once the above-Nyquist frequencies have been suppressed, it is safe to discard samples or interpolate between them to achieve the new sample rate [6].

### 3 Line removal with a Kalman filter

For the next step in our process, we wish to remove some of the narrow frequency features (lines) present in the data. Some lines are caused by processes whose effects are measured by environmental or other monitors. For example, the 60 Hz power oscillations that cause the power lines at multiples of 60 Hz in the data are measured by voltmeters and magnetometers positioned around the detectors. Removal of these lines, based on these measurements, is discussed in Section 6. The majority of lines, however, are caused by processes that cannot be directly measured and may not even be understood. To remove these lines we use a Kalman filter which gives us a way of modeling the process creating each line.

The Kalman filter was originally developed in 1960 by R. E. Kalman [7]. The filter is a set of equations that allow us to use a simple model of the process to predict what part of the data is due to the line [8, 9, 10]. It is assumed that the part of the data to be captured by the filter is caused by a system which is not directly observed. (We assume that the system causing each line can be approximated as a weakly damped harmonic oscillator that is excited by random noise.) At each time step  $k$  the system can be described by a state vector  $\psi[k]$ . (In the model of the harmonic oscillator, the state vector could describe the position and momentum of the oscillator.) The state evolves from one time step to the next via

$$\psi[k] = \mathbf{A} \cdot \psi[k-1] + \mathbf{w}[k], \quad (1)$$

where  $\mathbf{A}$  is the evolution operator matrix determined by the nature of the system and  $\mathbf{w}[k]$  is the white noise that is exciting the system. This noise is also called the process noise.

The state of the system is unknown, but affects the output of the detector which is observed. The observable (data) at time step (sample)  $k$  is expressed as the vector  $\mathbf{z}[k]$ . The dependence of the observable on the state can be expressed as:

$$\mathbf{z}[k] = \mathbf{C} \cdot \psi[k] + \mathbf{v}[k], \quad (2)$$

where  $\mathbf{C}$  is a linear transformation called the measurement operator and  $\mathbf{v}[k]$  is an additional, white noise background. This noise background is known as the measurement noise, which in the case of LIGO data is primarily caused by seismic noise at lower frequencies and laser shot noise at higher frequencies. Approximating the measurement noise background  $\mathbf{v}[k]$  as white would not be valid if the observable  $\mathbf{z}[k]$  was the data as produced by the detectors. (LIGO data minus a line are far from white.) However, if we take as our observable the data restricted to a small region in

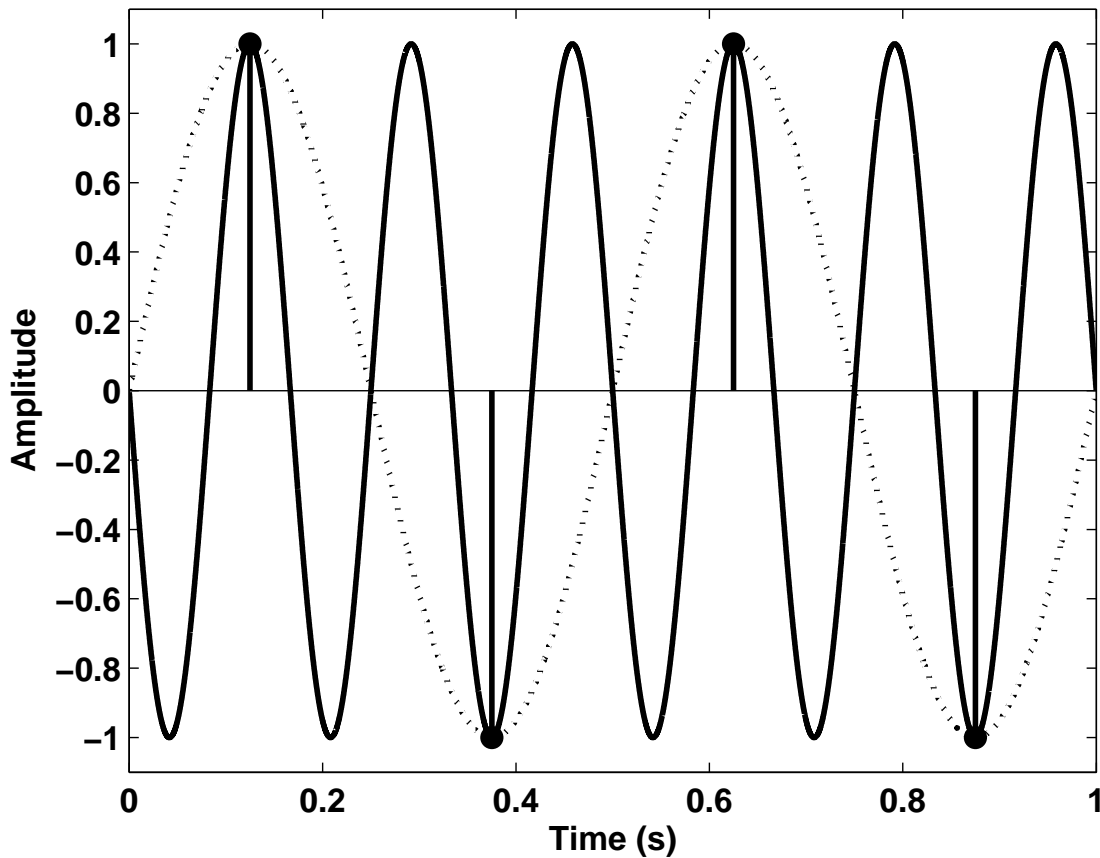


Figure 2: Aliasing example. Suppose a signal is sampled at a rate of 4 samples per second (black dots in the figure above). The sampled data only contain information on frequencies up to the Nyquist frequency which in this example is 2 Hz. If the signal being sampled is a 6 Hz sine wave, as shown above, the samples will have exactly the same values as if the signal were a 2 Hz sine wave. The power in the 6 Hz sine will be aliased or mistakenly attributed to the 2 Hz frequency since there is no way to distinguish the two frequencies once the sampling has taken place.

frequency about the line, as described in Section 3.1, the background is approximately white. The  $\mathbf{C} \cdot \boldsymbol{\psi}[k]$  term of Equation 2 is the part of the observable due to the line and is the part that we wish to isolate and remove.

The process and measurement noise,  $\mathbf{w}[k]$  and  $\mathbf{v}[k]$ , are both white and zero-mean and are therefore completely characterized by their covariance matrices

$$\mathbf{W} := E[\mathbf{w} \cdot \mathbf{w}^T] \quad (3)$$

$$\mathbf{V} := E[\mathbf{v} \cdot \mathbf{v}^T], \quad (4)$$

where  $E[\cdot]$  denotes the expectation operator.

Suppose we have an estimate  $\hat{\boldsymbol{\psi}}[k-1]$  of the system's state at the sample  $k-1$ . The error in this estimate is

$$\hat{\mathbf{e}}[k-1] := \boldsymbol{\psi}[k-1] - \hat{\boldsymbol{\psi}}[k-1]. \quad (5)$$

The variance of the error is

$$\hat{\mathbf{P}}[k-1] := E[\hat{\mathbf{e}}[k-1] \cdot \hat{\mathbf{e}}^T[k-1]]. \quad (6)$$

The state estimate can then be evolved to the next time sample  $k$  by using the evolution operator

$$\tilde{\boldsymbol{\psi}}[k] := \mathbf{A} \cdot \hat{\boldsymbol{\psi}}[k-1]. \quad (7)$$

Because we started with an estimate of the state that had some error, there will be error in  $\tilde{\boldsymbol{\psi}}[k]$  due to the error in the previous state estimate and due to process noise. The total error in  $\tilde{\boldsymbol{\psi}}[k]$  is defined to be

$$\tilde{\mathbf{e}}[k] := \boldsymbol{\psi}[k] - \tilde{\boldsymbol{\psi}}[k], \quad (8)$$

with a second moment of the error

$$\begin{aligned} \tilde{\mathbf{P}}[k] &:= E[\tilde{\mathbf{e}}[k] \cdot \tilde{\mathbf{e}}^T[k]] \\ &= \mathbf{A} \cdot \hat{\mathbf{P}}[k-1] \cdot \mathbf{A}^T + \mathbf{W}. \end{aligned} \quad (9)$$

Now we can calculate the observable  $\tilde{\mathbf{z}}[k]$  based on the estimated, evolved state  $\tilde{\boldsymbol{\psi}}[k]$

$$\tilde{\mathbf{z}}[k] = \mathbf{C} \cdot \tilde{\boldsymbol{\psi}}[k]. \quad (10)$$

Note that, while  $\tilde{\mathbf{z}}[k]$  is derived from an estimate of the state and therefore contains some error, it is free of the background measurement noise present in the measured observable  $\mathbf{z}[k]$ . Thus  $\tilde{\mathbf{z}}[k]$  is an estimate of the part of the observable due to the line, the part we wish to isolate and remove.

We can now calculate a corrected estimate of the state  $\hat{\boldsymbol{\psi}}[k]$  based on the difference between  $\tilde{\mathbf{z}}[k]$  and  $\mathbf{z}[k]$

$$\hat{\boldsymbol{\psi}}[k] = \tilde{\boldsymbol{\psi}}[k] + \mathbf{K}[k] \cdot (\mathbf{z}[k] - \tilde{\mathbf{z}}[k]), \quad (11)$$

where  $\mathbf{K}$  is the Kalman gain, described below. This corrected state estimate has an associated error and second moment of the error  $\hat{\mathbf{P}}[k]$  where, according to [11],

$$\begin{aligned} \hat{\mathbf{P}}[k] &= (\mathbf{I} - \mathbf{K}[k] \cdot \mathbf{C}) \cdot \tilde{\mathbf{P}}[k] \cdot (\mathbf{I} - \mathbf{K}[k] \cdot \mathbf{C})^T + \mathbf{K}[k] \cdot \mathbf{V} \cdot \mathbf{K}[k]^T \\ &= (\mathbf{I} - \mathbf{K}[k] \cdot \mathbf{C}) \cdot \tilde{\mathbf{P}}[k]. \end{aligned} \quad (12)$$

The Kalman gain is a linear operator that minimizes the error in the state. It is found by minimizing  $\text{tr}\hat{\mathbf{P}}[k]$  which is a measure of the total error in the corrected estimate of the state. According to [9], this results in

$$\mathbf{K}[k] = \tilde{\mathbf{P}}[k] \cdot \mathbf{C}^T / (\mathbf{V} + \mathbf{C} \cdot \tilde{\mathbf{P}}[k] \cdot \mathbf{C}^T). \quad (13)$$

The procedure for finding the portion of the observable due to the line,  $\tilde{\mathbf{z}}[k]$ , is as follows:

Begin with an estimate of the initial state  $\hat{\boldsymbol{\psi}}[0]$  and initial second moment of the error  $\hat{\mathbf{P}}[0]$ . A reasonable choice is to set  $\hat{\boldsymbol{\psi}}[0]$  equal to an array of zeros and  $\hat{\mathbf{P}}[0] = \mathbf{W}$ . Now for  $k = 1$  to the length of  $\mathbf{z}$ :

1. Evolve the estimate of the state via Equation 7 and find the error second moment using Equation 9.
2. Calculate  $\tilde{\mathbf{z}}[k]$  using Equation 10.
3. Find the Kalman gain via Equation 13.
4. Update the state and error second moment with Equations 11 and 12.

We apply a Kalman filter to the lines that are within the frequency band that will be selected when the data are broken into frequency bands. In some cases, lines just outside the frequency band are also removed so that they will not be aliased into the band of interest in the process of restricting the data to a band (see Appendix A). Kalman filtering is applied separately to each line or small group of lines if they are closely clustered in frequency.

### 3.1 Application to Narrow-Band LIGO Features

As mentioned in Section 3 the detector data are not appropriate to use as the observable  $\mathbf{z}$  in the Kalman filtering process. The data minus the line being removed are not white as is assumed for the background process noise. In order to get an observable that consists only of the line and an approximately white background, it is necessary to restrict attention to a narrow frequency range about the line.

For each line or set of lines, a small frequency band is chosen and designated by two numbers:  $f_c$ , the central frequency, and  $\Delta f$ , the width of the band. The band then spans the frequencies  $f_c - \Delta f/2$  to  $f_c + \Delta f/2$ . The band must include the line frequency or frequencies  $f_0$  ( $f_0$  is a vector of frequencies in the case of multiple lines) and should be chosen so that the power spectrum within the band is relatively flat and does not include any lines other than those being removed (see Figure 3). Typical values we use for  $\Delta f$  are 8-16 Hz. A list of Kalman filtered lines along with the chosen values of  $f_c$  and  $\Delta f$  is given in Appendix A

To restrict the data to the chosen band, the data are shifted down in frequency (heterodyned) by  $f_c$  Hz. This shift is done by multiplying the full data,  $g[k]$ , by a complex exponential with frequency  $f_c$ . That is for  $k = 1$  to the length of  $g[k]$

$$g'[k] = 2 \cdot \exp(-2\pi i k f_c / f_s) g[k], \quad (14)$$

where  $g'[k]$  are the shifted data and  $f_s$  is the sample rate of the original data. After the shift,  $f_c$  is at zero frequency and both  $f_c + \Delta f/2$  and  $f_c - \Delta f/2$  are shifted to  $\Delta f/2$ .



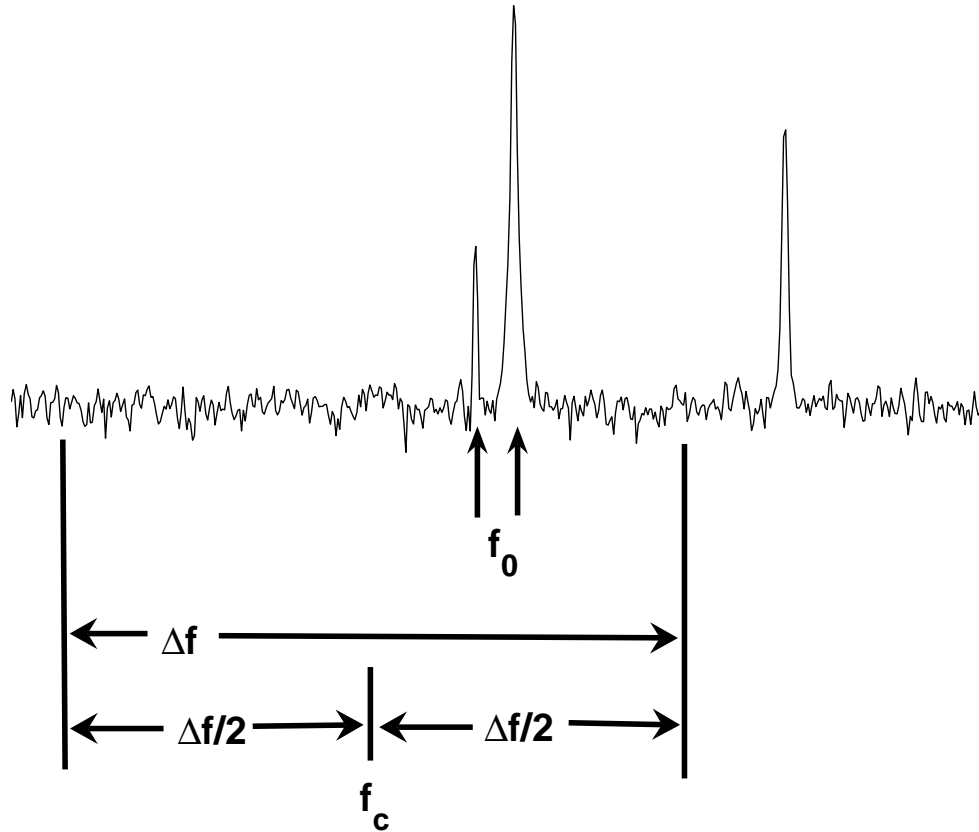


Figure 3: Diagram showing the specification of frequency-related variables needed for Kalman filtering. In this case, we want to filter the two lines near the center of the diagram.  $f_0$  is a two element vector that contains the central frequencies of these two lines.  $f_c$  and  $\Delta f$  are used to specify a frequency band that includes the lines of interest and some of the white background but excludes any other frequency features.

Next, the shifted data are downsampled to a new sample rate of  $\Delta f$ . This downsampling throws out all frequencies outside the band of interest. The downsampled data, labeled  $z$ , are complex. We construct the real vector series  $\mathbf{z}[k]$ :

$$\mathbf{z}[k] = \begin{pmatrix} \Re(z)[k] \\ \Im(z)[k] \end{pmatrix}, \quad (15)$$

which is the observable that undergoes the Kalman filtering procedure.

After Kalman filtering is done and the line component of the observable,  $\tilde{\mathbf{z}}[k]$ , is obtained it must be put through a process inverse to that described above before it can be subtracted from the data. First,  $\tilde{\mathbf{z}}[k]$  is turned into a complex series:

$$\tilde{z}[k] = \tilde{\mathbf{z}}[1, k] + i \cdot \tilde{\mathbf{z}}[2, k], \quad (16)$$

where  $\tilde{\mathbf{z}}[1, k]$  is the element from the first row,  $k$ th column of  $\tilde{\mathbf{z}}$  and  $\tilde{\mathbf{z}}[2, k]$  is from the second row,  $k$ th column (see Equation 15).

Then  $\tilde{z}[k]$  is resampled to a new sample rate  $f_s$ , where  $f_s$  is the sample rate of the data prior to Kalman filtering. These resampled data, denoted  $\tilde{g}'[k]$ , are shifted up in frequency by the amount  $f_c$ , the central frequency of the band used to find the observable:

$$\tilde{g}[k] = \exp(2\pi i k f_c / f_s) \tilde{g}'[k]. \quad (17)$$

Finally, the real part of the shifted data is taken:

$$\tilde{g}[k] = \Re \tilde{g}[k]. \quad (18)$$

$\tilde{g}[k]$  is now the part of the data due to the line. Subtracting  $\tilde{g}[k]$  from the data,  $g[k]$  removes the line from the data, as shown in Figure 4.

### 3.2 Calculating $\mathbf{A}$ , $\mathbf{C}$ , $\mathbf{W}$ , and $\mathbf{V}$

Before we can apply the Kalman filter, we must find the values of the matrices  $\mathbf{A}$ ,  $\mathbf{C}$ ,  $\mathbf{W}$ , and  $\mathbf{V}$ . Detailed derivations of the expressions given below may be found in [9].

We can write the state vector at each time sample as a real vector such that

$$\boldsymbol{\psi}[k] = \begin{pmatrix} \Re(\psi[k]) \\ \Im(\psi[k]) \end{pmatrix}. \quad (19)$$

The evolution operator  $\mathbf{A}$ , which controls the evolution of the state vector according to

$$\boldsymbol{\psi}[k+1] = \mathbf{A} \cdot \boldsymbol{\psi}[k] + \mathbf{w}[k], \quad (20)$$

can be calculated via  $\mathbf{A} = \exp(\mathbf{A}' / \Delta f)$  where  $\exp$  denotes matrix exponentiation and

$$\mathbf{A}' = \begin{pmatrix} -\omega_0 / (2Q) & \omega_c - \omega_0 \sqrt{1 - 1 / (4Q^2)} \\ -\omega_c + \omega_0 \sqrt{1 - 1 / (4Q^2)} & -\omega_0 / (2Q) \end{pmatrix}. \quad (21)$$

The angular frequencies  $\omega_0 = 2\pi f_0$  and  $\omega_c = 2\pi f_c$  are the frequency of the line and central frequency of the band about the line that we use in creating the observable, while  $Q$  is a damping

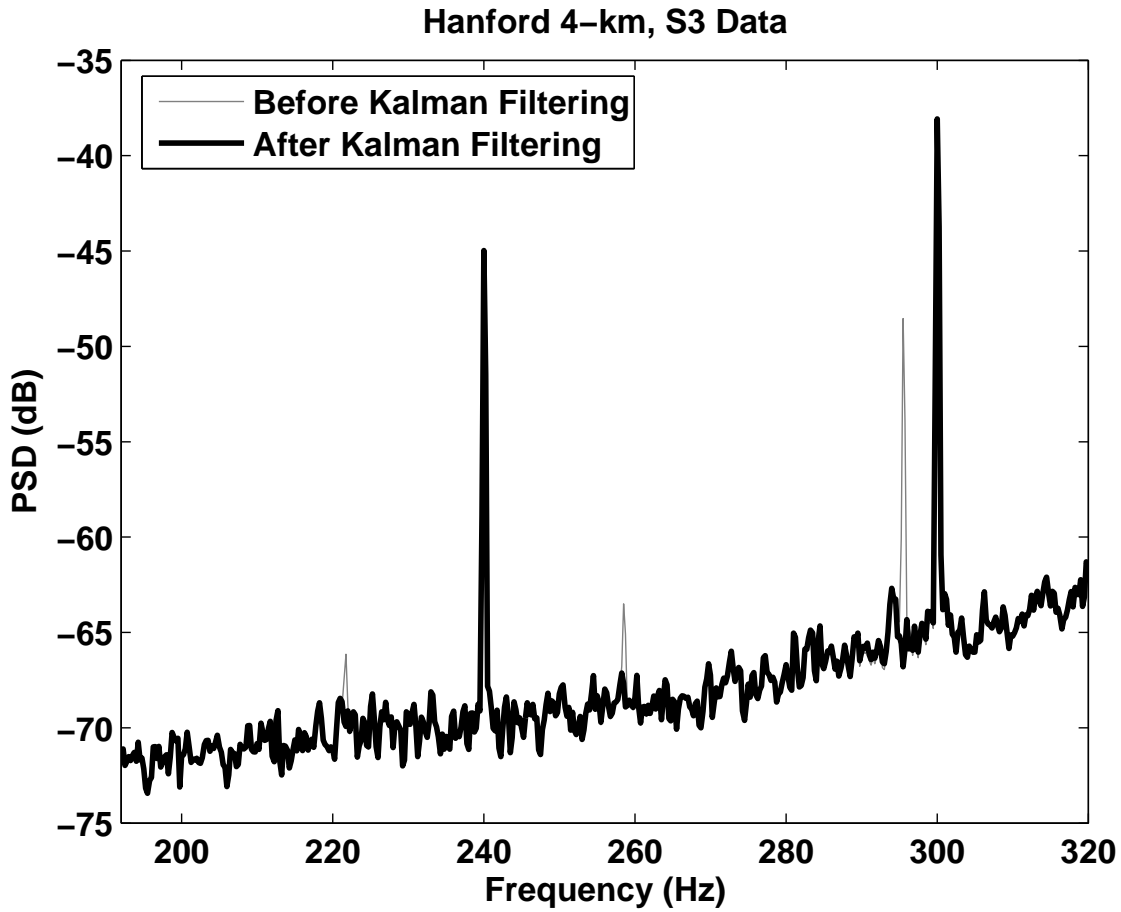


Figure 4: Line removal via Kalman filtering in the 192-320 Hz band of data taken during the third science run (S3) with the Hanford 4-km detector. The two lines at 240 and 300 Hz that are not removed are power lines that will be dealt with in the later, line regression step.

constant.  $Q$  and  $f_0$  are found for each line by calculating high resolution power spectra for each detector.  $f_0$  is the frequency corresponding to the maximum of the line and  $Q$  is measured as  $f_0$  divided by the width of the line at half the maximum.  $f_0$  and  $Q$  for each Kalman filtered line are given in Appendix A.

If multiple lines are being removed together, the expressions for the state and evolution operator become

$$\Psi = \begin{pmatrix} \Re(\psi_1) \\ \Im(\psi_1) \\ \vdots \\ \Re(\psi_N) \\ \Im(\psi_N) \end{pmatrix} \quad (22)$$

and

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_N \end{pmatrix}, \quad (23)$$

where  $\psi_1 \dots \psi_N$  and  $\mathbf{A}_1 \dots \mathbf{A}_N$  are the state and evolution operators, respectively, for each of the individual lines.

Next, is the measurement operator  $\mathbf{C}$  which relates the state to the observable

$$\mathbf{z}[k] = \mathbf{C} \cdot \boldsymbol{\psi}[k] + \mathbf{v}[k]. \quad (24)$$

We let the state be the shifted amplitude of the mode causing the line so the measurement operator  $\mathbf{C}$  is just the identity matrix. For multiple lines,  $\mathbf{C}$  is a matrix of repeats of the identity matrix

$$\mathbf{C} = (\mathbf{C}_1 \dots \mathbf{C}_N) = \begin{pmatrix} 1 & 0 & \dots & 1 & 0 \\ 0 & 1 & \dots & 0 & 1 \end{pmatrix}. \quad (25)$$

The measurement noise covariance matrix  $\mathbf{V}$  is found by measuring the average value of the power spectrum of the observable  $\mathbf{z}$  away from the line. This average value,  $\bar{P}_m$ , is equivalent to the total measurement noise  $\text{tr} \mathbf{V}$ . We find  $\bar{P}_m$  by summing the power spectrum over frequencies more than 2 Hz away from the line or lines being removed. Also, since the observable  $\mathbf{z}$  is a vector with real and imaginary parts separated, we take the power spectrum of the two parts separately and then add the means of the two spectra to get  $\bar{P}_m$ . The measurement noise covariance matrix is then

$$\mathbf{V} = \frac{1}{2} \bar{P}_m \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (26)$$

The process noise covariance matrix  $\mathbf{W}$ , on the other hand, is related to the power spectrum at frequencies dominated by the line  $P_z$  via the expression [11]

$$P_z(\omega) = \frac{4Q^2 \text{tr} \mathbf{W}}{\omega_0^2 + (2Q(\omega + \omega_c) - \omega_0(4Q^2 - 1)^{1/2})^2}. \quad (27)$$

The process noise covariance matrix is found by integrating both sides of Equation 27 over the range of frequencies where the line is dominant. The expression on the right hand side is so strongly peaked about  $\omega_0$  that integrating over the line frequencies is approximately equivalent to

integrating over the entire range from  $\omega = -\infty$  to  $\omega = \infty$ . Also, since  $P_z$  is the power spectrum of the observable  $z$ , which are data that have been shifted down in frequency by the amount  $\omega_c$ , the value of  $\omega_c$  in Equation 27 is zero. The integration of the right side is thus simplified and results in

$$\int P_z(\omega)d\omega = (2\pi Q/\omega_0) \text{tr}\mathbf{W}. \quad (28)$$

When evaluating this expression, it is important to remember that the frequency shift used in producing the observable changes the values of  $\omega_0$  and  $Q$ . ( $\omega_0(\text{shifted}) = \omega_0 - \omega_c$ ,  $Q(\text{shifted}) = (f_0 - f_c)/\Delta f$ ). It is also important to make sure that one calculates a power spectrum that preserves the absolute magnitude of the signal creating the line.

Once the power spectra for the real and imaginary parts of the observable  $z$  have been found, each is integrated over the line-dominated frequencies. This is done by multiplying the magnitude of the power spectra at each frequency by the frequency bin width and then summing.  $P_I$  is computed as the mean of the real and imaginary integrated power spectra.  $\text{tr}\mathbf{W}$  is found via Equation 28 and the process noise covariance matrix,  $\mathbf{W}$  is

$$\mathbf{W} = \frac{\omega_0 P_I}{4\pi Q} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (29)$$

If multiple lines are being filtered at once,  $\mathbf{W}$  becomes

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 & & \\ & \ddots & \\ & & \mathbf{W}_N \end{pmatrix}, \quad (30)$$

where  $\mathbf{W}_1 \dots \mathbf{W}_N$  are the process noise covariance matrices for each of the individual lines.

It is important to note that  $\mathbf{W}$ , as a measure of the power contained in a line, is a quantity that can vary greatly over time. The resonances that produce the lines can be excited over short timescales and lines that have large amplitudes at some times can be completely absent at others. To have an accurate value for the process noise it is important to measure it periodically. For the S2 and S3 data, following collaboration specified protocols [12, 13], the process noise was calculated on segments of data whose start times were spaced 6370 seconds apart and were 300 seconds long. When Kalman filtering the lines from a segment of data, we select the  $\mathbf{W}$  measurement from the segment with a midpoint closest to the midpoint time of the data we are filtering.

## 4 Shaping Filter

As can be seen in Figure 1, there is a huge amount of power at low frequencies. These frequencies are lower than the lowest edges of any of the frequency bands we are interested in (128 Hz for S2, S3 and 96 Hz for S4) but this high power can still cause problems. During the basebanding step when the data are restricted to a frequency band, as described in Section 5, the low frequency part of the data gets shifted to high frequencies and then the data are downsampled. If the power at low frequencies is not reduced before basebanding, aliasing can occur when this downsampling takes place. The next step in our data conditioning procedure is application of a filter that reduces power at low frequencies.

Generally speaking, a filter is a system that modifies the frequency structure of a signal, amplifying or suppressing some frequencies relative to others. A filter can be expressed as either a series of numbers  $x[n]$ , where  $-\infty < n < \infty$ , or in the frequency domain as a frequency response  $X(\omega)$ , where  $-\pi \leq \omega \leq \pi$ . The series and frequency response representations are related to each other via the discrete Fourier and inverse Fourier transforms:

$$X(\omega) = \sum_{n=-\infty}^{\infty} x[n]e^{-i\omega n}, \quad (31)$$

and

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega)e^{i\omega n}d\omega. \quad (32)$$

Applying the filter in the time domain involves convolving the data  $d[n]$  with the filter sequence  $x[n]$  to obtain a filtered signal  $y[n]$  or

$$y[n] = d[n] * x[n] = \sum_{k=-\infty}^{\infty} d[k]x[n - k]. \quad (33)$$

Convolution in the time domain is equivalent to multiplication in the frequency domain so filtering can also be performed by multiplying the filter frequency response by the Fourier transform of the data or

$$Y(\omega) = D(\omega)X(\omega), \quad (34)$$

where  $Y(\omega)$  and  $D(\omega)$  are the Fourier transforms of the filtered data and original data, respectively.

If the data being filtered are zero everywhere but a single sample  $p$  where its value is one it is known as a unit impulse signal, written as

$$d[n] = \begin{cases} 1 & n = p \\ 0 & n \neq p \end{cases}, \quad (35)$$

or

$$d[n] = \delta[n - p]. \quad (36)$$

Filtering the unit impulse signal with the filter  $x[n]$  results in

$$y[n] = \sum_{k=-\infty}^{\infty} \delta[k - p]x[n - k] = x[n + p], \quad (37)$$

which is simply the filter sequence delayed by  $p$  samples. For this reason,  $x[n]$  is also called the impulse response of a filter. During filtering, any sample of data will be smeared over a number of samples equal to the length of the impulse response.

The shaping filter that reduces the power at low frequencies should have a frequency profile that changes at the lowest frequency of interest  $f_l$  ( $f_l = 128$  Hz for S2, for example). Above  $f_l$ , the filter should leave the data as unchanged as possible and below  $f_l$ , the power of the data should be reduced so that it is less than any frequency above  $f_l$ . At the same time, we want to use a filter that is as short as possible to avoid smearing any burst signals over a large number of samples, making them harder to detect.

The best filter for our purposes would be an ideal highpass filter, in other words, a filter whose frequency response is zero below the cutoff  $f_l$  and one above it

$$X(f) = \begin{cases} 0 & f < f_l \\ 1 & f \geq f_l \end{cases} . \quad (38)$$

As stated above, applying the filter is equivalent to multiplying the filter's frequency response  $X(f)$  by the Fourier transform of the data  $D(f)$  so that the filtered data  $Y(f) = X(f)D(f)$  and

$$Y(f) = \begin{cases} 0 & f < f_l \\ D(f) & f > f_l \end{cases} . \quad (39)$$

The ideal highpass filter completely suppresses frequencies below  $f_l$  and leaves those above  $f_l$  unchanged.

Unfortunately, this type of ideal filter with its infinitely sharp cutoff is not realizable. Since the impulse response can be calculated from the frequency response via

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{i\omega n} d\omega, \quad (40)$$

where  $\omega = 2\pi f/f_s$  and  $f_s$  is the sample rate, the impulse response of the ideal filter is

$$x[n] = \frac{1}{2\pi} \left[ \int_{-\pi}^{-\omega_l} e^{i\omega n} d\omega + \int_{\omega_l}^{\pi} e^{i\omega n} d\omega \right] \quad (41)$$

$$x[n] = \frac{-\sin(\omega_l n)}{\pi n} \quad (42)$$

for all  $-\infty < n < \infty$ . The impulse response of the ideal filter is infinitely long. What is worse,  $x[n]$  is not absolutely summable, that is  $|x[n]|$  summed over all  $n$  is not less than infinity. This means that when we try to build our ideal filter, it will not converge uniformly for all values of  $\omega$ . To see this, consider the Fourier transform of our ideal filter impulse response:

$$X(\omega) = \sum_{n=-\infty}^{\infty} x[n] e^{-i\omega n} \quad (43)$$

$$= \sum_{n=-\infty}^{\infty} \frac{-\sin(\omega_l n)}{\pi n} e^{-i\omega n}. \quad (44)$$

If we let  $X_M(\omega)$  be the sum of a finite number of terms, or

$$X_M(\omega) = \sum_{n=-M}^M \frac{-\sin(\omega_l n)}{\pi n} e^{-i\omega n}, \quad (45)$$

Figure 5 illustrates the behavior of the sum as  $M$  increases. With increasing  $M$  the transition at  $\omega = \omega_l$  becomes sharper and the ripples near  $\omega = \omega_l$  become more rapid. In fact, the magnitude of the first ripple on either side of  $\omega_l$  does not decrease with larger  $M$ , a behavior known as the Gibbs phenomenon.

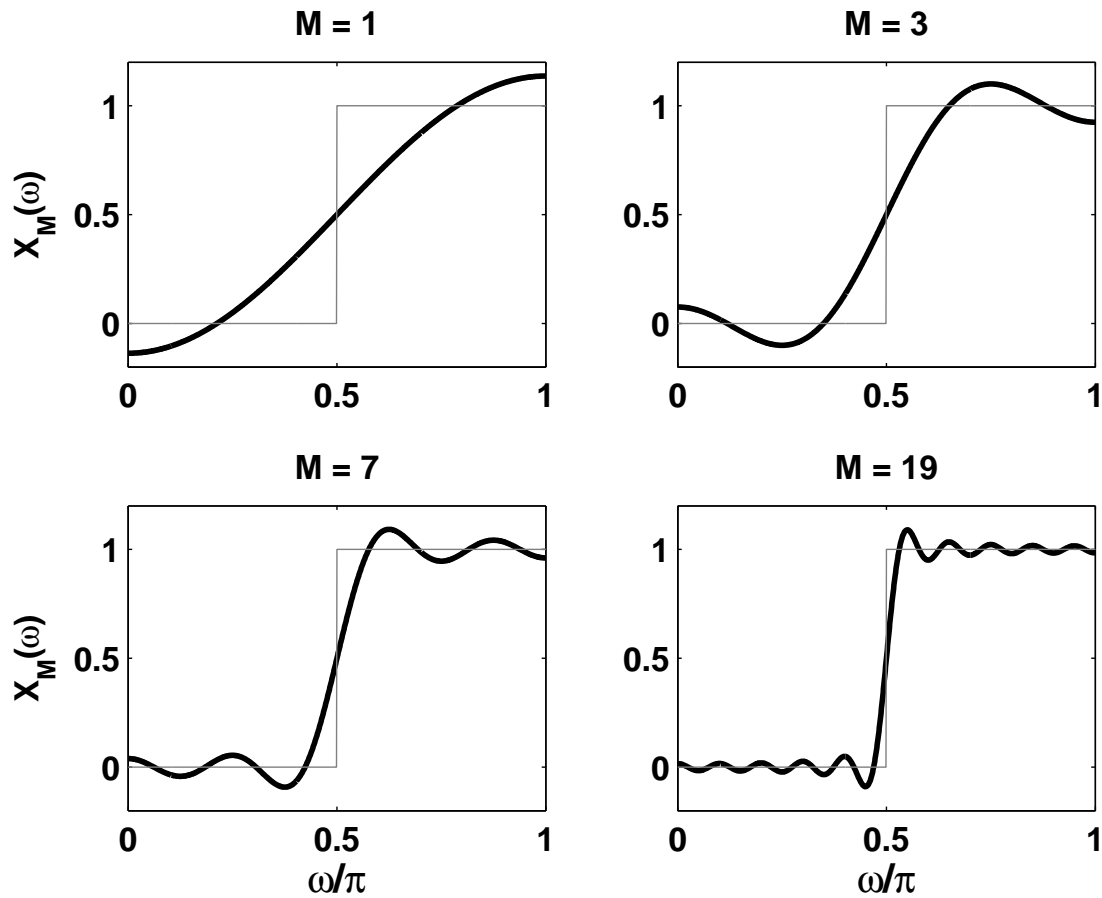


Figure 5: Result of summing different number of terms of  $X_M(\omega)$  from Equation 45 with  $\omega_l = 0.5\pi$ . As  $M$  increases, the frequency response  $X_M(\omega)$  (dark line) approaches that of the ideal highpass filter (light line). The transition at  $\omega_l = 0.5\pi$  becomes sharper with higher  $M$  but the magnitude of the first ripple on either side of the transition does not decrease.



We now see that if we are to apply a real-world highpass filter having a finite length, we are going to have to deal with imperfections in its frequency response. The frequency response will have three regions, a stopband where the response is nearly zero, a passband where the response is close to one and a transition band where the response transitions smoothly between the two extremes. A frequency response with these bands is shown in Figure 6. Plus, there will be some ripple in both the passband and stopband. The goal is to build a shaping filter whose passband begins at  $f_l$ . The transition band should be as wide as possible to lower the filter length but it cannot be too low or the high power at low frequencies will not be suppressed. Also, the filter must be designed so that the ripple is small, especially in the stopband. We design the necessary highpass filter using the Parks-McClellan algorithm. This algorithm returns an optimum filter for a given transition band and given errors in the stopband and passband [6].

The effect of the shaping filter on S3 data from the Hanford 4-km detector is shown in Figure 7.

## 5 Basebanding

Trying to remove all of the artifacts and correlations from the data at all of the frequencies where LIGO is sensitive to gravitational waves would be a difficult task and require very long and complex filters. The necessary filters could decrease our sensitivity to gravitational wave bursts by smearing the signals out over a large number of samples. To make the task of whitening simpler, requiring shorter filters, we break the data into frequency bands, each containing a smaller number of features requiring removal. The bands can also be chosen to avoid artifacts that are difficult to remove. For example, the data from S2-S4 were broken into frequency bands that avoided the violin modes (resonances of mirror suspension wires) and their harmonics (see Appendix B). A second advantage to breaking the data into separate bands is that any candidate gravitational wave events found in a band can be automatically associated with a narrow range of frequencies.

We call the process of breaking the data into frequency bands basebanding. The goal of basebanding is to completely restrict the data in frequency to a band in the original data. After basebanding, the interval of zero frequency to Nyquist frequency is completely taken up by a band of interest. For example, if we want to restrict the data to a band running from 128 Hz to 256 Hz, the basebanding procedure will return data with these frequencies shifted to 0 Hz - 128 Hz and with a sample rate of 256 Hz.

Basebanding has a number of steps and the order in which they are applied is very important. This procedure must be carried out for each frequency band of central frequency  $f_c$  and width  $\Delta f$ .

1. Shift by  $f_c$  - For each band with central frequency  $f_c$  and frequency width  $\Delta f$ , heterodyne (shift down in frequency) the data by  $f_c$ . For original data  $g[k]$  and shifted data  $g'[k]$

$$g'[k] = \exp(-2\pi i k f_c / \Delta f) g[k] \quad (46)$$

for  $k = 1$  to the length of  $g[k]$ . This will have the effect of shifting the data down in frequency by  $f_c$  so that the band of interest will be centered at zero frequency. To see how this shifting works, consider data consisting of a single frequency  $\omega_0$  or  $g[k] = \exp(i\omega_0 k)$ . Applying Equation 46 with  $2\pi f_c / \Delta f = \omega_c$  will result in  $g'[k] = \exp(-i\omega_c k) \exp(i\omega_0 k) = \exp(i(\omega_0 - \omega_c)k)$  or data shifted to the frequency  $\omega_0 - \omega_c$ .

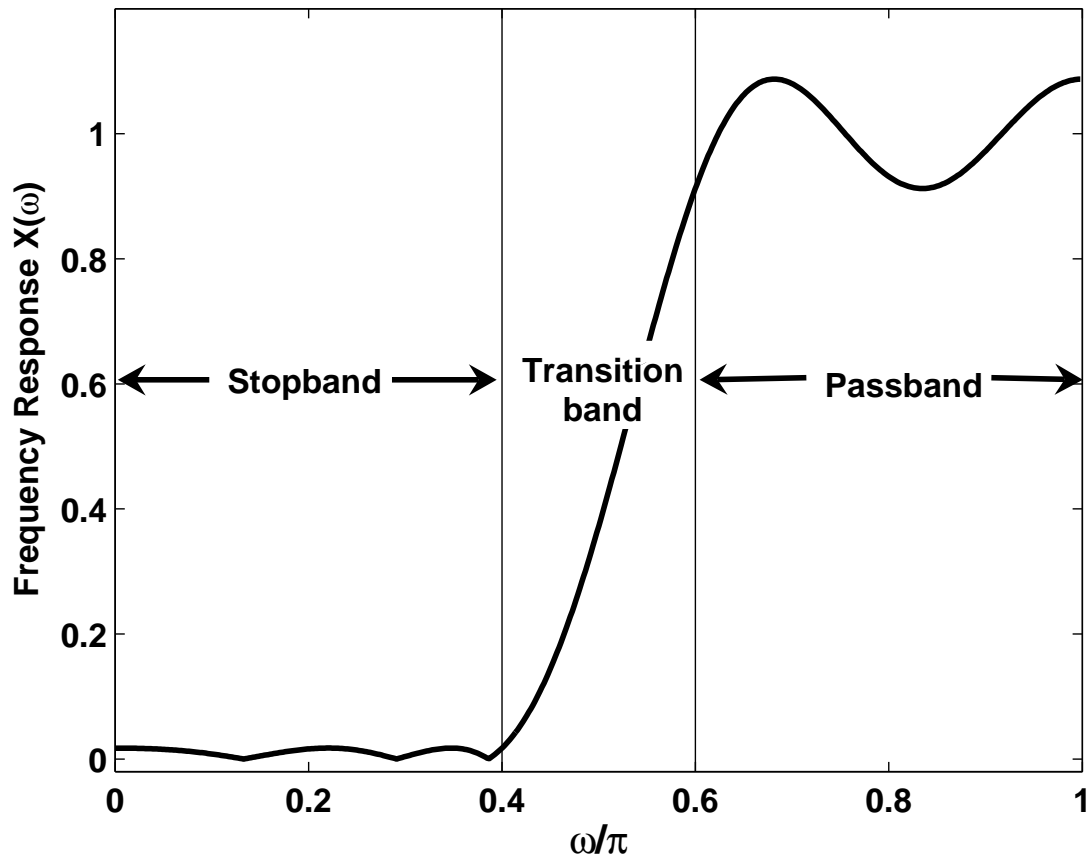


Figure 6: Frequency response of realistic highpass filter. The frequency response consists of three frequency regions: a stopband running from  $0$ - $0.4\pi$  where the response is close to zero, a passband running from  $0.6\pi$ - $\pi$  where the response is close to one and a transition band where the response transitions smoothly between the two extremes. The stopband has an allowed error (maximum ripple height) of  $0.02$  while the passband's allowed error is  $0.1$ . This filter was designed using the Parks-McClellan algorithm.

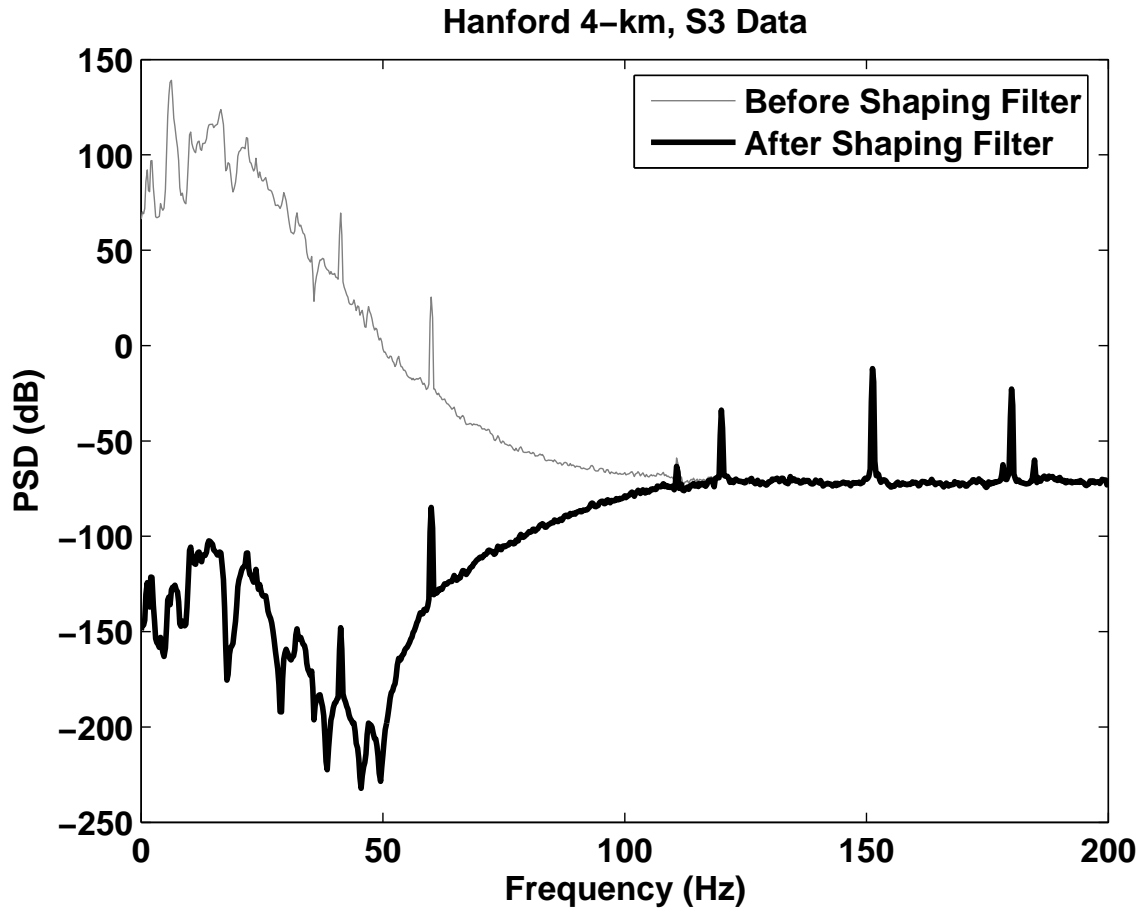


Figure 7: The effect of the shaping filter on third science run (S3) data from the Hanford 4-km detector. The power below our lowest frequency of interest (128 Hz) is well suppressed by the filter.

2. Downsample - Downsample to move the Nyquist frequency to  $\Delta f$ . This will make the design of the filter used in the next step much easier since now the filter's transition band as a fraction of the Nyquist frequency will be much larger.
3. Lowpass - A lowpass filter (having a frequency response near one below a cutoff and near zero above a cutoff frequency) is applied with a cutoff frequency equal to  $\Delta f/2$ , removing the frequencies outside the desired band (see Figure 8). Therefore, by first shifting the data in frequency, we are able to effectively bandpass with a lowpass filter, which is easier to design. The lowpass filters used for the S2-S4 data were created using the Parks-McClellan algorithm (see Section 4). The filters we used are different for each frequency band, but are the same across detectors. The differences are required by both the differences in frequency width of the bands and the features that lie outside the bands that necessitate different falloff rates so that they will not be aliased into the band. The filter parameters are given in Appendix B.
4. Shift up in frequency by half the bandwidth - Shift the data up by the frequency equal to  $\Delta f/2$ . This shift will move the low end of the desired band to zero frequency and the high end to the frequency equal to  $\Delta f$ . For  $g[k]$  equal to the data before shifting and  $g'[k]$  equal to the data after

$$g'[k] = \exp(2\pi i k/4)g[k]. \quad (47)$$

Now the entire range of available frequencies, from zero to Nyquist is taken up by the band of interest (see Figure 8).

5. Take the real part - Finally, take the real part of the complex data. The complex part will be approximately zero after the previous step. Taking the real part will result in a real time series just as the data before basebanding was real.

## 6 Line Regression

The next step in our data conditioning procedure is to remove the lines that were not removed via Kalman filtering. These lines can be removed by building filters that capture the correlations between the detector output channel and other channels such as environmental channels that monitor the interferometers' surroundings. Power lines and calibration lines, for example, can often be regressed from the data using filters applied to other channels and then subtracting the filtered channel data from the detector output data. The calibration lines are, of course, correlated strongly with the channel which measures the excitation applied to the mirrors to create the calibration lines. The power lines at multiples of 60 Hz that are caused by the AC oscillations of the power grid are correlated with channels from environmental monitors like voltmeters and radiometers and sometimes with optical channels.

To remove the power and calibration lines, the environmental or excitation channel is first resampled to a Nyquist frequency of 2048 Hz and then put through the same data conditioning steps as the detector output including application of the shaping filter and basebanding but excluding Kalman line removal.

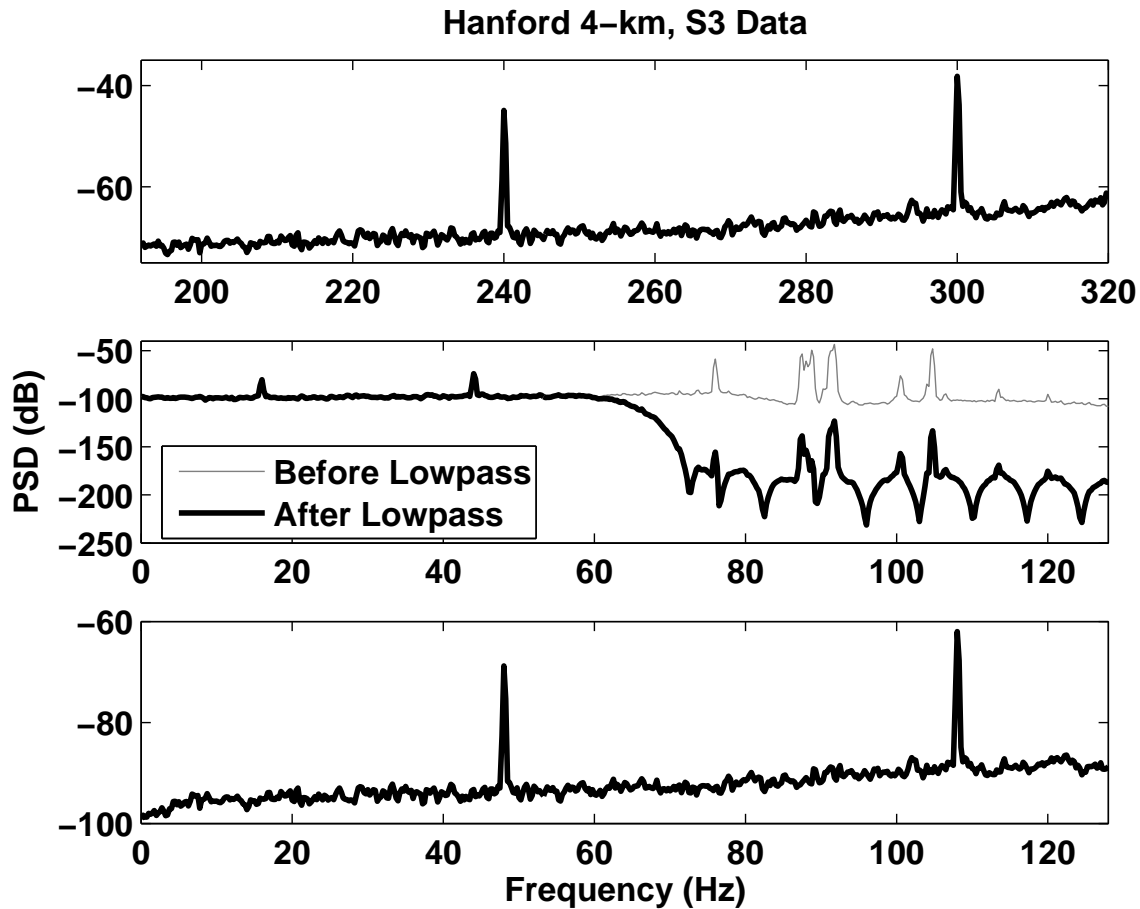


Figure 8: The above three plots show the basebanding process performed on the 192-320Hz band of Hanford-4km, S3 data. The top plot is a power spectra of the pre-basebanded data from 192-320 Hz. The middle plot shows the effect of the lowpass filter applied to the shifted, resampled data. The lowpass filter suppresses the features outside the desired band. Finally, the bottom plot shows the basebanded data which ranges in frequency from 0-128 Hz.

For each band, an ARX model is found that relates the basebanded detector data to the basebanded environmental or excitation data. An ARX model has the form  $A(q)y(t) = B(q)u(t-k) + e(t)$  or

$$y[t] + a[1]y[t-1] + \dots + a[n]y[t-n] = b[1]u[t-k] + \dots + b[m]u[t-k-m] + e[t], \quad (48)$$

where  $y[t]$  is the output data,  $u[t]$  is the input data,  $e[t]$  is white noise,  $k$  is a delay parameter, the  $a[1] \dots a[n]$  relate current output data to past output samples and the  $b[1] \dots b[m]$  relate input data to output data [14]. In the case of line removal,  $y[t]$  is the basebanded detector data and  $u[t]$  is the basebanded environmental or excitation data. Since we are only interested in how detector data are related to excitation or environmental data and not in how they are related to past detector data samples  $n = 0$ , there are no  $a$  coefficients. We also assume there is no delay between the channels so  $k = 0$ .

Once the data have been fit to the ARX model and the least squares estimate of the  $b$  coefficients have been found, the power or calibration lines can be removed by first filtering the basebanded environmental or excitation data with the  $b$  coefficients and then subtracting the result from the detector data or  $d[t] = y[t] - b[n] * u[t]$  where  $d[t]$  is the basebanded detector data with the lines removed.

Appendix C gives a list of the lines removed via regression for the three interferometers along with the regression channel used and the filter order  $m$ . As with the process noise measurements, the necessary regression filters change with time. Therefore, these filters were also calculated on 300 seconds of data spaced every 6370 seconds and the filters from the closest segment were used for a given data stretch.

The effect of line regression on the 192-320 Hz band of Hanford 4-km, S3 data is shown in Figure 9.

## 7 Final Whitening

Now that we have mostly removed the prominent lines and other artifacts from the data, all that remains is to smooth out any frequency features that have survived thus far and remove any general increases or decreases in the noise with frequency. This last step in our data conditioning procedure is accomplished with a final whitening filter.

The final whitening filter is created by finding an AR model for the basebanded, line-filtered data and using the model as a filter to remove remaining features. AR is a linear difference equation which has the form  $A(q)x(t) = e(t)$  or

$$x[t] + a[1]x[t-1] + a[2]x[t-2] + \dots + a[n]x[t-n] = e[t] \quad (49)$$

[14]. The use of this model assumes that elements of the time series are made up of a component which can be predicted from past behavior and a white noise component. If the time series is filtered with the AR coefficients, the result is just the white noise component,  $e(t)$ . For the filter applied to the S1-S4 LIGO data, we set the number of  $a$  coefficients,  $n$  to 60 for each band. This choice of order was made by noticing that an increase in order above 60 did not result in substantially better whitening.

Before calculating the whitening filter, it is important to remove a small piece of data from the beginning and end of the data to avoid startup and ending transients (described in Section 8). These are created by previous steps in the data conditioning process.

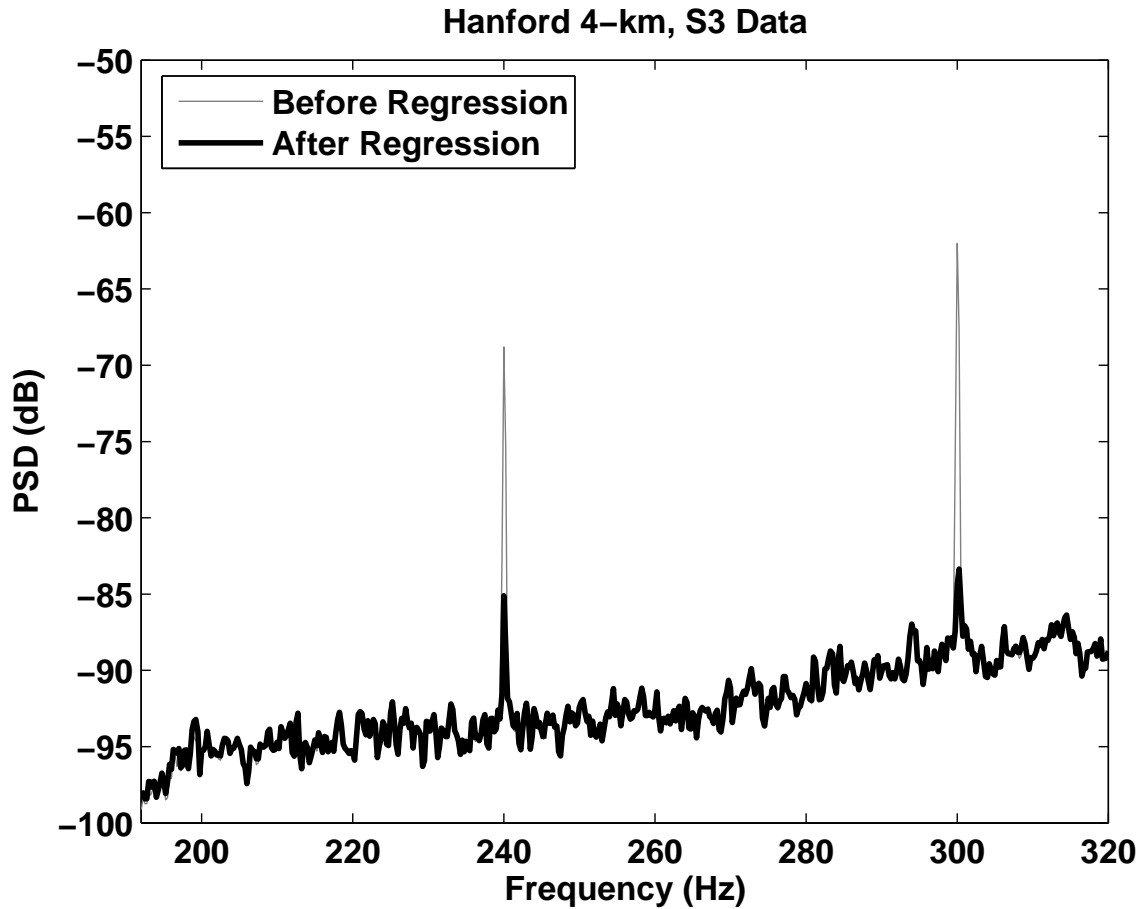


Figure 9: The effect of line regression on the 192-320 Hz band of Hanford 4-km, S3 data using a voltmeter channel. The two power lines at 240 and 300 Hz are greatly reduced. This plot shows basebanded data but the frequencies have been mapped to corresponding frequencies in non-basebanded data for ease of comparison.

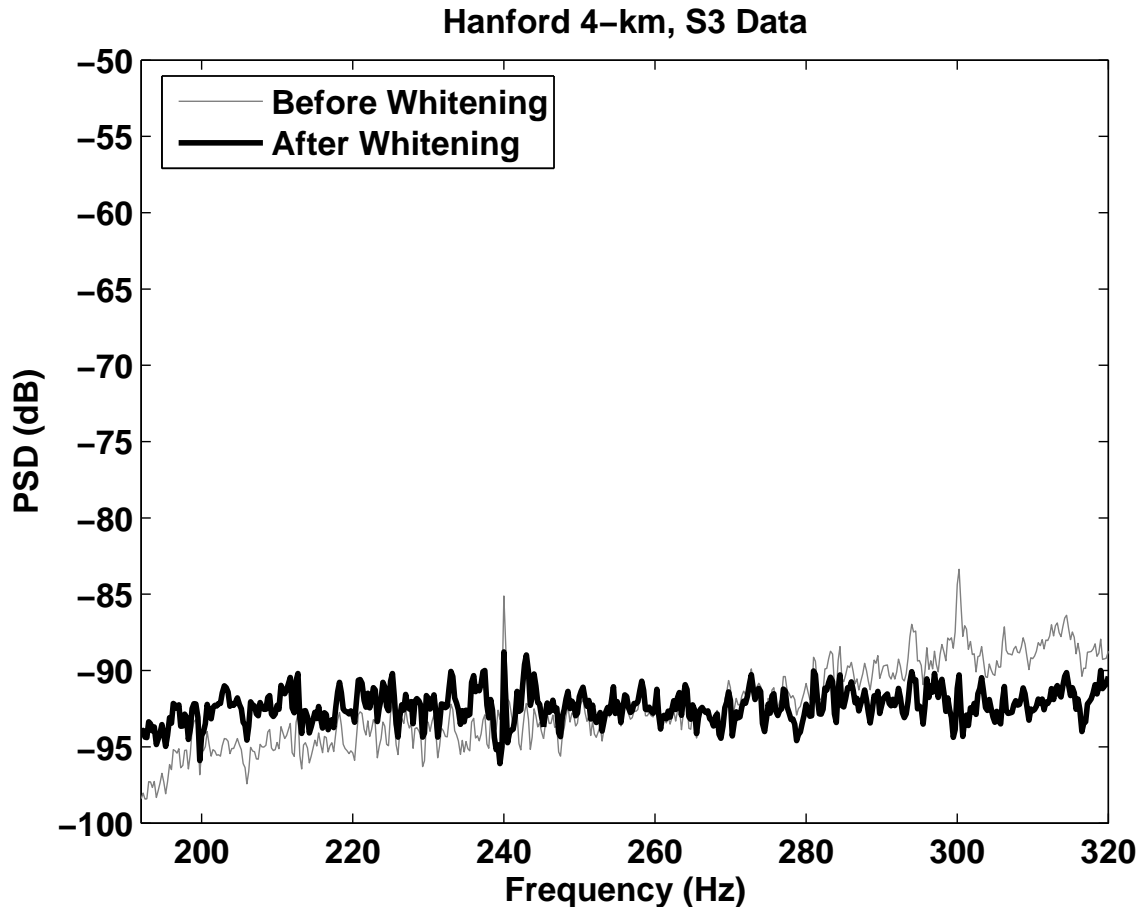


Figure 10: The effect of whitening on the 192-320 Hz band of Hanford 4-km, S3 data. The general increase in power with frequency has been flattened. This plot shows basebanded data but the frequencies have been mapped to corresponding frequencies in the non-basebanded data for ease of comparison.

Once again, the whitening filter was found using 300 seconds of data every 6370 seconds and the filter from the nearest segment is used to whiten data. An example of whitening is shown in Figure 10.

## 8 Startup and Ending Transients

Several of the data conditioning steps involve filtering the data. Filtering usually involves zero padding, or adding zeros to the beginning of the data so that the filtered data is of the same length as the initial data. While zero padding preserves length, it can cause the first samples of the filtered data to behave strangely. (For example, these samples can have abnormally large or small amplitudes). As a result, it is a good idea to discard these affected samples at the end of the data conditioning procedure.

To find the length of this startup transient, we calculate the length of the impulse response of



Freq. Band (Hz)	Startup Imp. Resp.		Ending Imp. Resp.	
	Samples	Seconds	Samples	Seconds
128 - 192	27	0.2109	26	0.2031
192 - 320	36	0.1406	35	0.1367
384 - 512	34	0.1328	33	0.1289
512 - 640	34	0.1328	33	0.1289
704 - 832	43	0.1680	42	0.1641
832 - 1024	59	0.1536	58	0.1510
1065 - 1365	64	0.1067	63	0.1050
1408 - 1708	64	0.1067	63	0.1050
1758 - 2048	62	0.1069	61	0.1052
890 - 940	23	0.2875	22	0.2200

Table 1: Startup and ending impulse response lengths for S3 data conditioning and all bands. The impulse responses were found by applying data conditioning to a vector of all zeros except for a unit impulse at the first (startup) or last (ending) sample. Kalman filtering and line regression were both turned off.

the data conditioning procedure. The impulse response is found by applying the data conditioning procedure to a time series with a one for its first sample and all other samples equal to zero. The data conditioning filters will smear out the one in the first sample. The number of nonzero samples in the conditioned time series will be equal to the number of samples influenced by the zero padding. There is also an end transient caused by filtering during basebanding. The length of this transient can be found by applying data conditioning to a vector with a unit impulse for its last sample. When calculating the impulse responses, Kalman filtering and line regression are turned off since these steps do not involve direct filtering of the data but instead apply filters to copies of the data or other channels that are then subtracted from the data.

The startup and ending impulse responses for S3 data conditioning on the 192-320 Hz band are shown in Figures 11 and 12. The lengths of the transients for all bands for S3 are given in Table 1. According to this table, removing a few tenths of a second from the beginning and end of conditioned data should be sufficient to avoid data conditioning transients.

## 9 Effectiveness of Data Conditioning

Data conditioning involves many choices, from what lines will be removed with a Kalman filter to what environmental or interferometer channels will be used in regression to filter orders. Poor choices could result in ineffective data conditioning. Plus, line strength measurements as well as filter calculations are done on data segments spaced every 6370 seconds and these filters are applied to data away from these segments where the noise characteristics may have changed. It is important to know if applying the data conditioning procedure is actually working and producing nearly white data or if there is a failure either as a result of a poor parameter choice or a bad section of data that should not be used in an analysis.

The simplest way to see if data are being whitened is to look at a power spectral density (PSD) plot and see if the conditioned data has a PSD that has nearly the same magnitude for all frequen-

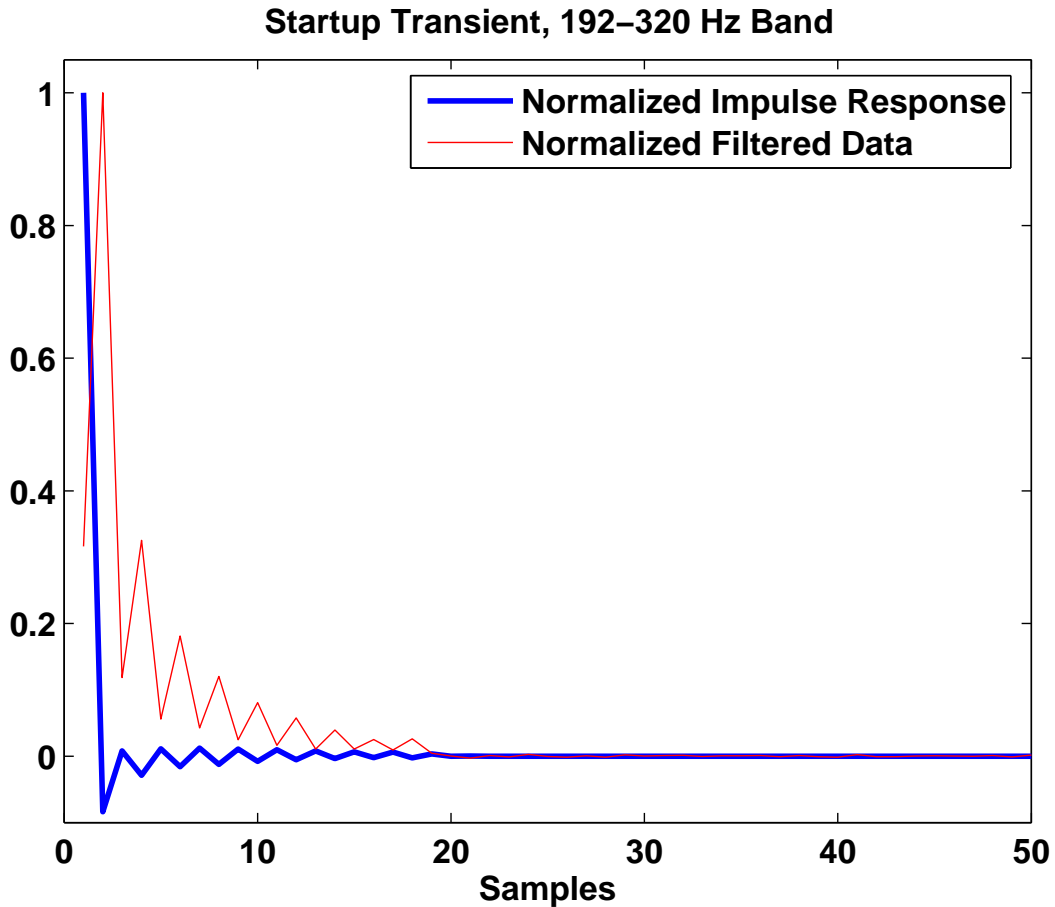


Figure 11: Startup impulse response for S3 data conditioning on the 192-320 Hz band as well as startup transient for conditioned data. The impulse response was found by applying data conditioning to a vector with one as its first element and zeros for all other elements. Kalman filtering and line regression were both turned off. The impulse response is the same length as the startup transient for the filtered data. Both the impulse response and filtered data have been normalized so that their largest samples are equal to one for ease of comparison.

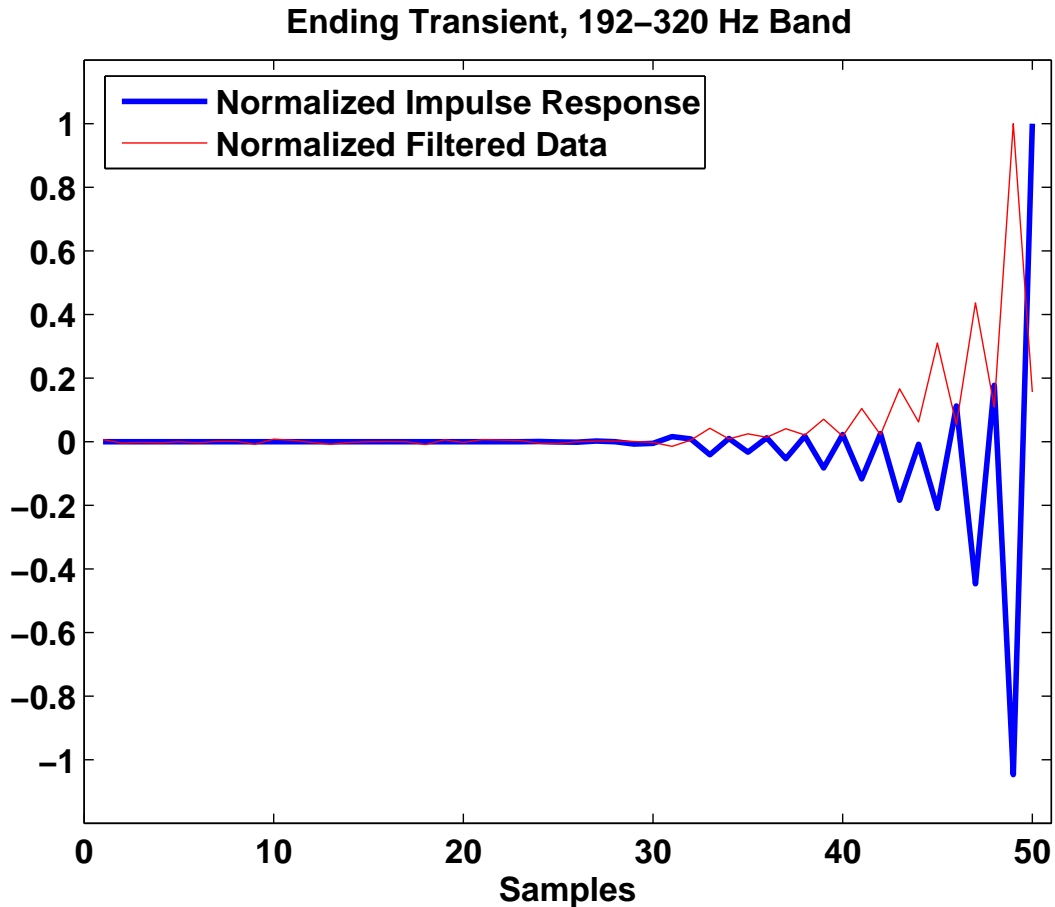


Figure 12: Ending impulse response for S3 data conditioning on the 192-320 Hz band as well as ending transient for conditioned data. The impulse response was found by applying data conditioning to a vector with one as its last element and zeros for all other elements. Kalman filtering and line regression were both turned off. The impulse response is the same length as the ending transient for the filtered data. Both the impulse response and filtered data have been normalized so that their largest samples are equal to one for ease of comparison.

Freq. Band (Hz)	% FOM < 1		
	Hanford 4-km	Hanford 2-km	Livingston 4-km
128 - 192	100.0	100.0	99.9
192 - 320	99.6	100.0	99.8
384 - 512	99.6	99.7	98.2
512 - 640	99.6	100.0	98.6
704 - 832	96.1	97.0	98.7
832 - 1024	100.0	100.0	100.0
1065 - 1365	100.0	100.0	100.0
1408 - 1708	100.0	99.9	99.0
1758 - 2048	100.0	98.5	87.5
890 - 940	81.4	99.6	100.0

Table 2: Percentage successful data conditioning. Shown above are the percentages of S3 data segments for each detector and frequency band where the figure-of-merit (FOM) was less than 1. A FOM less than 1 indicates that the Kalman filtering, line regression and whitening portions of data conditioning were successful in producing whiter data.

cies. Plots showing the effectiveness of data conditioning on a segment of S3 data from the Hanford 4-km detector and all frequency bands are shown in Appendix D. Looking at PSDs by eye, besides being merely qualitative, would be impossibly time consuming if applied to a significant fraction of the huge amount of data produced by the three LIGO detectors. We need a quantitative measure of data conditioning performance that can be applied quickly.

A simple measure of data conditioning effectiveness has been proposed in [15]. For this method, the Rayleigh statistic  $\mathcal{R}$  is calculated before and after data conditioning is applied. The Rayleigh statistic is a measure of the flatness of the PSD and is calculated as  $\mathcal{R}_{PSD} = \sigma_{PSD}/\mu_{PSD}$  where  $\sigma_{PSD}$  and  $\mu_{PSD}$  are the variance and mean of the PSD, respectively. The flatter the PSD, the smaller  $\mathcal{R}_{PSD}$ . Therefore, a figure-of-merit (FOM) for the data conditioning is

$$\text{FOM} = \frac{\mathcal{R}_{PSD'}}{\mathcal{R}_{PSD}} = \frac{\sigma_{PSD'}/\mu_{PSD'}}{\sigma_{PSD}/\mu_{PSD}}, \quad (50)$$

where  $PSD$  is the power spectral density calculated on data that has only been treated with the shaping filter and then basebanded and  $PSD'$  is data that has undergone the entire data conditioning process. If full data conditioning including Kalman filtering, line regression and whitening produces data that are whiter than data that have only been restricted to the same frequency band, then  $\text{FOM} < 1$ .

The FOM was calculated for data from all three LIGO detectors taken during S3 and for all frequency bands used in S3. For each detector, the data that passed quality cuts recommended by the Detector Characterization Working Group [16, 17, 18, 19] (the cuts ensure the detector was in a good, operating state) were broken into 300 second long non-overlapping segments. Any segment less than 300 seconds long was discarded. Data conditioning was applied to each segment and the FOM calculated. The percentages of S3 segments for which data conditioning was successful are shown in Table 2. For most segments, the FOM was less than one.

Figures 13 and 14 show the FOM per segment and a histogram of FOM values for the 192-320 Hz band of the Hanford 4-km detector. Corresponding plots for all frequency bands and

detectors are given in Appendix E. In the 192-320 Hz band for the Hanford 4-km detector, the FOM was less than one for 9,814 (or 99.6%) of the 9,854 segments tested in S3. Looking closely at the remaining 40 segments with high FOM values, data conditioning was found to fail for two reasons. In the first 31 segments, the regression of power lines failed and the strength of a power line was increased. Usually this occurred when the 240 Hz line disappeared or mostly disappeared from the data but did not disappear from the regression channel. The disappearance of the line may have been caused by fans, motors, or heaters switching on or off and changing the magnetic environment around the mirrors, for example. This change in relationship between the data and regression channels occurred sometime between these segments and the playground segment on which the regression filter was calculated. To fix the power line regression, filters would have to be calculated closer to these segments. The remaining 9 segments with high FOM values had “glitches” which are very brief (usually just a few samples) very high amplitude pulses in the data of instrumental origin. Most data segments containing glitches are excluded by data quality cuts but the ones in our 9 segments had evaded these cuts. The data conditioning procedure cannot remove the effects of glitches and these segments should be avoided (or vetoed) by any analysis looking for gravitational wave bursts.

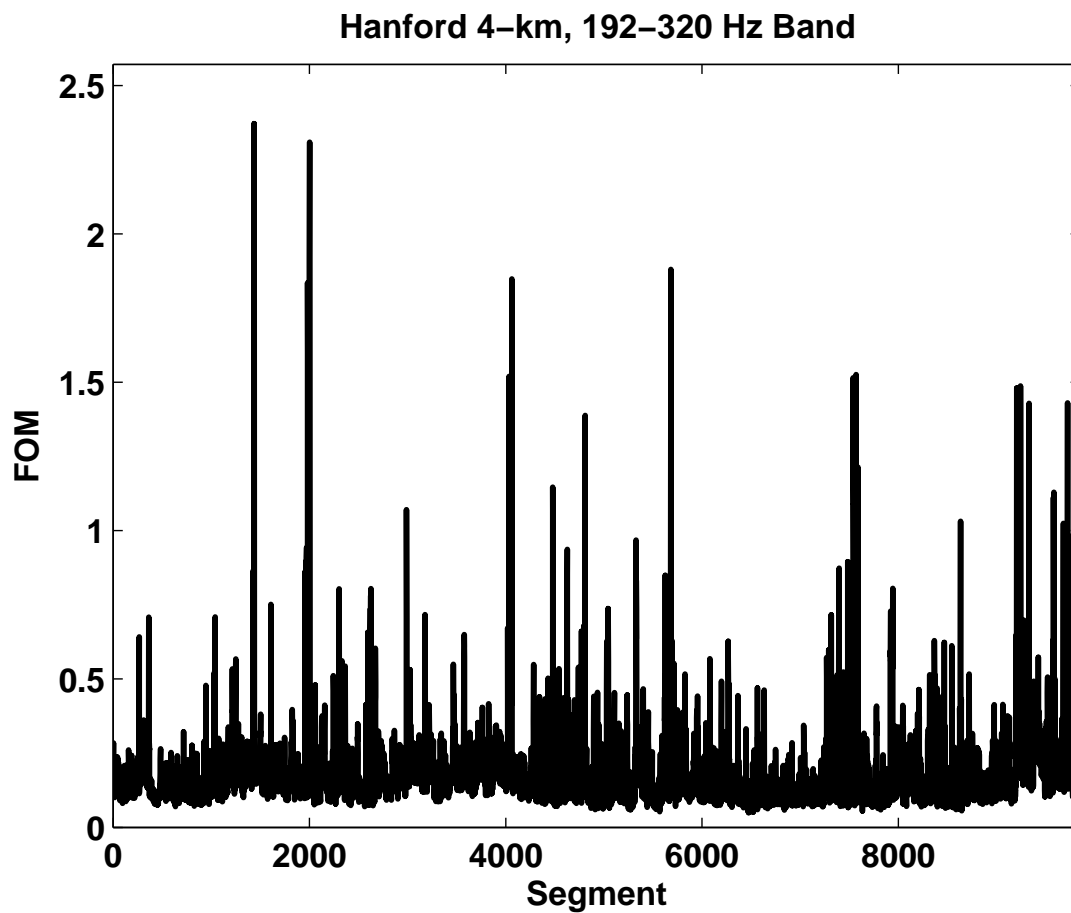


Figure 13: FOM per segment for the Hanford 4-km, S3 data, 192-320 Hz band. The segments for which  $FOM > 1$  are distributed throughout the run.

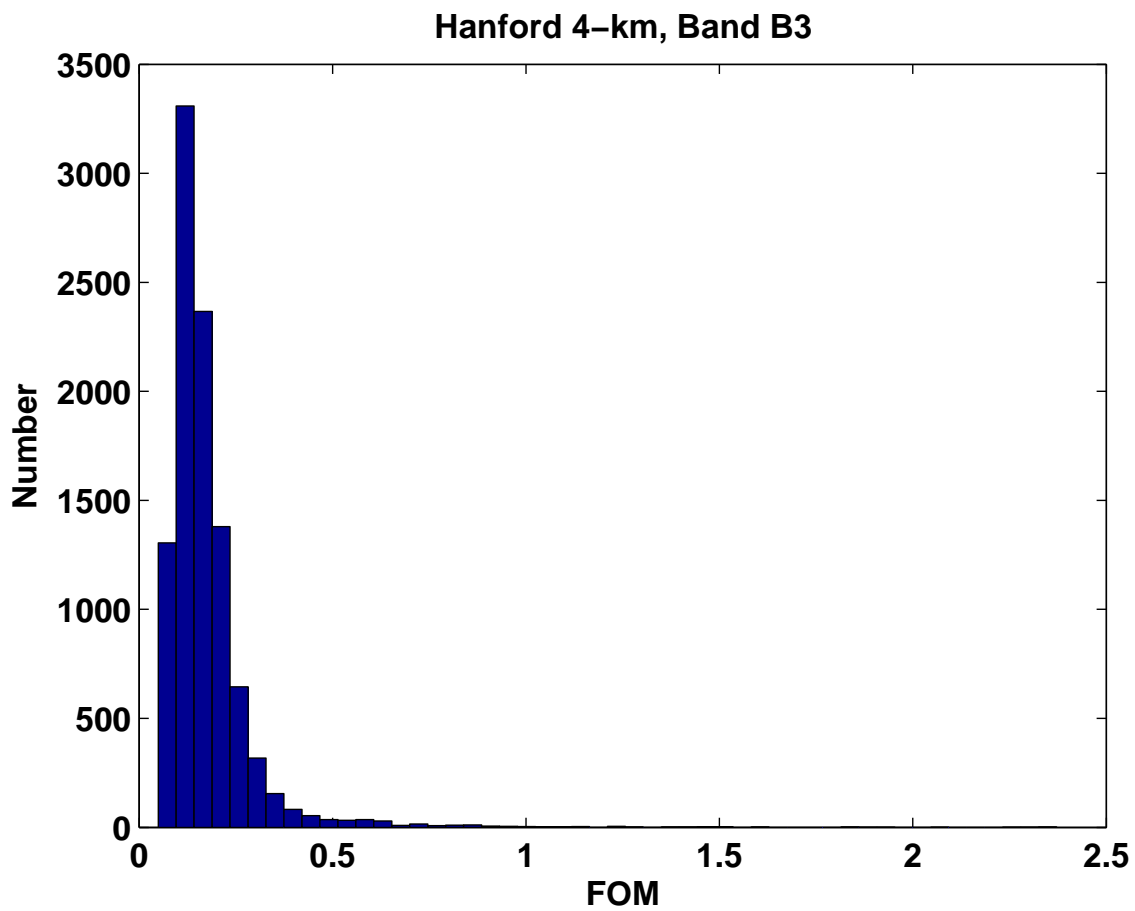


Figure 14: Histogram of FOM values for the Hanford 4-km, S3 data, 192-320 Hz band. Most segments have  $FOM < 1$  but there is a tail of FOM values that extend above 1.

## A Kalman filtered line parameters

The following tables specify the Kalman filtering parameters used during the S2 and S3 runs. The first column of each table gives the frequency band for which each line was removed. Frequency band definitions are given in Appendix B. The second column gives the line number. Lines with the same band and line number are removed together with the same Kalman filter. As for the other columns,  $f_0$  is the frequency and  $Q$  the quality factor of each line.  $f_c$  and  $\Delta f$  are the central frequency and size of the frequency band upon which the filter acts (see Figure 3).  $v$  is the measurement noise of the background close to the line such that the measurement noise covariance matrix

$$\mathbf{V} = v \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (51)$$



Band	Index	$f_0$	$f_c$	$Q$	$\Delta f$	$v$
A2	1	184.718	184	31000	8	3.312e-04
B2	1	258.606	258	32000	8	1.706e-06
B2	2	295.550	295	49000	8	1.154e-05
C2	1	406.380	406	51000	8	1.063e-06
C2	2	465.694	465	78000	8	5.161e-07
D2	1	546.068	546	91000	8	5.282e-07
D2	2	564.070	566	94000	16	9.738e-07
D2	2	566.102	566	142000	16	9.738e-07
D2	2	568.050	566	95000	16	9.738e-07
D2	2	568.102	566	142000	16	9.738e-07
D2	3	591.098	589	74000	8	3.995e-07
E2	1	775.816	775	78000	8	5.773e-07
E2	2	812.760	812	68000	8	3.443e-07
E2	3	886.646	886	148000	8	3.106e-07
E2	4	931.386	931	233000	8	3.247e-07
F2	1	1366.803	1369	10000	8	3.543e-07
F2	2	1108.308	1108	277000	8	2.657e-07
F2	3	1182.194	1182	296000	8	2.833e-07
F2	4	1256.082	1254	628000	8	3.850e-07
F2	5	1329.970	1330	332000	8	3.385e-07
G2	1	1403.856	1403	351000	8	3.273e-07
G2	2	1551.632	1551	388000	8	3.630e-07
G2	3	1625.482	1628	406000	8	3.952e-07
G2	4	1699.404	1699	425000	8	3.948e-07
H2	1	1773.292	1773	443000	8	3.725e-07
H2	2	1847.180	1847	462000	8	3.256e-07
H2	3	1862.750	1866	465000	8	3.435e-07
H2	3	1862.772	1866	465000	8	3.435e-07

Table 3: Hanford 4-km, S2 Kalman line parameters

Band	Index	$f_0$	$f_c$	$Q$	$\Delta f$	$\nu$
A2	1	162.000	162	405000	16	2.186e-05
B2	1	295.800	295	49000	16	6.849e-07
C2	1	406.724	406	102000	8	4.927e-07
C2	2	465.708	465	78000	8	5.676e-07
C2	3	486.000	486	122000	8	2.859e-07
D2	1	510.720	510	800	4	1.500e-06
D2	2	591.598	591	99000	8	1.902e-07
E2	1	698.454	698	134000	8	1.202e-07
E2	1	698.565	698	152000	8	1.202e-07
E2	1	698.643	698	152000	8	1.202e-07
E2	1	699.378	698	175000	8	1.202e-07
E2	2	931.418	931	155000	8	8.416e-08
E2	3	961.848	961	160000	16	6.777e-08
E2	4	972.000	972	162000	16	7.122e-08
E2	5	988.080	988	35000	8	6.957e-08
F2	1	1257.138	1254	314000	8	5.746e-08
G2	1	1719.500	1719	4000	8	8.180e-08
G2	2	1504.790	1506	5000	8	5.541e-08
G2	3	1528.240	1528	5000	8	7.736e-08
G2	4	1496.400	1496	5000	8	8.103e-08
G2	5	1453.790	1453	182000	8	7.085e-08
G2	5	1453.802	1453	103000	8	7.085e-08
G2	6	1463.540	1463	10000	8	5.882e-08
G2	7	1472.570	1472	5000	8	6.129e-08
G2	8	1481.470	1481	5000	8	6.348e-08
G2	9	1626.890	1626	406000	8	7.015e-08
G2	10	1689.160	1689	5000	8	7.923e-08
H2	1	1830.470	1830	5000	8	6.234e-08
H2	2	1862.834	1866	465000	8	6.956e-08
H2	3	1996.644	1996	499000	8	4.234e-08
H2	4	1976.196	1974	494000	8	7.671e-08
H2	5	1902.600	1903	5000	8	7.636e-08
H2	5	1903.500	1903	5000	8	7.636e-08

Table 4: Hanford 2-km, S2 Kalman line parameters

Band	Index	$f_0$	$f_c$	$Q$	$\Delta f$	$v$
A2	1	140.605	140	2200	8	1.062e-07
A2	2	187.580	187	19000	8	3.299e-08
B2	1	221.206	221	55000	8	7.070e-08
B2	2	234.474	234	20000	8	1.649e-08
B2	3	281.370	281	20000	8	2.263e-08
B2	4	294.940	294	37000	8	8.962e-08
D2	1	516.146	516	86000	8	7.346e-08
D2	2	589.882	589	147000	16	4.128e-08
E2	1	737.352	737	184000	16	7.254e-08
E2	2	811.088	811	135000	16	8.782e-08
E2	3	884.822	884	147000	16	8.561e-08
E2	4	958.558	958	240000	16	7.763e-08
F2	1	1079.750	1080	3000	8	3.251e-08
F2	2	1106.028	1106	553000	8	3.191e-08
F2	3	1139.650	1140	3000	8	3.150e-08
F2	4	1179.764	1180	295000	8	3.249e-08
F2	5	1199.650	1200	3000	8	3.144e-08
F2	6	1243.700	1243.5	300000	8	5.670e-08
F2	7	1247.160	1247	300000	8	2.503e-08
F2	8	1260.000	1260	3000	8	3.449e-08
F2	9	1319.630	1319.63	3000	8	3.656e-08
F2	10	1327.234	1327	313000	8	3.532e-08

Table 5: Livingston 4-km, S2 Kalman line parameters

Band	Index	$f_0$	$f_c$	$Q$	$\Delta f$	$\nu$
G2	1	1400.940	1400.97	350000	8	4.026e-08
G2	2	1407.100	1407.1	3500	8	4.177e-08
G2	3	1412.070	1412.07	300000	8	4.170e-08
G2	4	1415.020	1415	300000	8	3.095e-08
G2	5	1428.000	1428	500000	8	7.034e-08
G2	6	1431.500	1432	600000	8	2.771e-10
G2	7	1437.600	1437.6	359000	8	4.829e-08
G2	8	1439.530	1440	3000	8	4.316e-08
G2	9	1474.704	1478	737000	8	4.863e-08
G2	10	1500.000	1500	3000	8	4.743e-08
G2	11	1548.440	1548	387000	8	4.658e-08
G2	12	1559.500	1559.5	3000	8	5.006e-08
G2	13	1612.320	1612.32	50000	8	8.461e-10
G2	14	1615.880	1616	60000	8	3.588e-10
G2	15	1619.950	1619.95	80000	8	5.179e-10
G2	16	1620.200	1620.2	3000	8	5.463e-08
G2	17	1622.174	1622.17	405000	8	9.248e-08
G2	18	1680.000	1680	300	8	8.957e-08
G2	19	1695.910	1695.91	847000	8	9.165e-08
H2	1	1769.646	1769.65	442000	8	5.394e-08
H2	2	1799.500	1799.5	6000	8	5.335e-08
H2	3	1843.380	1843.38	921000	8	4.973e-08
H2	4	1860.100	1860.1	3000	8	4.797e-08
H2	5	1917.116	1917.12	479000	8	4.021e-08
H2	6	1920.000	1920	3000	8	4.027e-08
H2	7	1979.500	1979.5	3000	8	3.024e-08
H2	8	1990.850	1990.85	497000	8	3.012e-08
H2	9	2031.414	2028	1013000	8	2.911e-08
H2	10	2032.600	2034	1013000	8	2.142e-08
H2	11	2039.500	2039.5	3000	8	2.754e-08

Table 6: Livingston 4-km, S2 Kalman line parameters continued

Band	Index	$f_0$	$f_c$	$Q$	$\Delta f$	$v$
A3	1	184.721	184.50	57439	8	8.5812e-07
A3	2	178.344	178.00	5360	8	7.8817e-07
B3	1	295.553	294.50	90460	12	3.3275e-06
B3	1	296.286	294.50	59248	12	3.3275e-06
B3	1	294.820	294.50	59754	12	3.3275e-06
B3	2	258.609	258.25	78708	8	9.3666e-07
B3	3	221.665	221.50	67922	8	7.1649e-07
C3	1	406.385	405.50	121361	16	2.6095e-06
C3	1	407.119	405.50	69779	16	2.6095e-06
C3	1	410.098	405.50	27171	16	2.6095e-06
C3	1	405.653	405.50	71470	16	2.6095e-06
C3	2	443.330	443.00	132723	8	1.7300e-06
C3	3	465.693	465.50	121331	8	1.7060e-06
D3	1	591.106	590.25	177343	12	1.8413e-06
D3	1	590.373	590.25	124736	12	1.8413e-06
D3	1	591.840	590.25	122069	12	1.8413e-06
D3	2	517.218	517.00	151213	8	1.5553e-06
D3	3	628.050	627.75	183616	8	1.0631e-06
D3	4	546.052	545.75	104658	8	1.4594e-06
D3	5	566.102	563.75	132537	16	2.4088e-06
D3	5	564.067	563.75	111928	16	2.4088e-06
D3	5	568.101	563.75	139885	16	2.4088e-06
D3	6	554.162	554.00	145531	8	1.4378e-06
D3	7	535.056	534.75	32872	8	1.5853e-06
D3	7	534.900	534.75	81605	8	1.5853e-06
E3	1	738.883	738.00	220593	12	1.3913e-06
E3	1	739.616	738.00	134574	12	1.3913e-06
E3	1	738.150	738.00	142383	12	1.3913e-06
E3	2	775.827	775.50	230807	8	7.9951e-07
E3	3	812.771	812.50	241657	8	7.3486e-07
F3	1	886.660	885.75	245345	12	1.0426e-06
F3	1	887.393	885.75	165479	12	1.0426e-06
F3	1	885.926	885.75	169739	12	1.0426e-06
F3	2	931.385	931.00	249564	8	7.0324e-07
F3	3	997.491	997.25	283184	8	7.3332e-07
F3	4	923.601	923.25	257940	8	6.9436e-07
F3	5	947.750	947.50	295644	8	7.1309e-07
F3	6	849.715	849.50	176666	8	7.0975e-07
F3	7	902.250	902.00	226767	8	6.9790e-07
F3	8	880.339	880.00	147200	8	7.8983e-07

Table 7: Hanford 4-km, S3 Kalman line parameters

Band	Index	$f_0$	$f_c$	$Q$	$\Delta f$	$v$
G3	1	1256.100	1255.25	353120	12	1.7324e-06
G3	1	1256.834	1255.25	244113	12	1.7324e-06
G3	1	1255.368	1255.25	251989	12	1.7324e-06
G3	2	1329.989	1329.00	370711	12	1.9393e-06
G3	2	1330.723	1329.00	248832	12	1.9393e-06
G3	2	1329.256	1329.00	256971	12	1.9393e-06
G3	3	1182.212	1182.00	334731	8	1.0211e-06
G3	4	1108.324	1108.00	315056	8	8.6505e-07
G3	5	1293.045	1292.75	347216	8	1.2347e-06
G3	6	1071.380	1071.00	297310	8	7.9759e-07
G3	7	1145.268	1145.00	308879	8	9.2550e-07
G3	8	1219.156	1219.00	317695	8	1.0936e-06
G3	9	1360.250	1360.00	340112	8	1.5423e-06
G3	10	1239.750	1239.50	359985	8	1.1568e-06
H3	1	1442.000	1441.75	534074	12	3.2950e-06
H3	1	1442.733	1441.75	300611	12	3.2950e-06
H3	2	1625.542	1625.25	432878	8	2.5889e-06
H3	3	1699.430	1699.25	441708	8	2.8208e-06
H3	4	1551.653	1551.50	419329	8	2.2518e-06
H3	5	1663.750	1662.25	514580	12	4.0169e-06
H3	5	1662.486	1662.25	395461	12	4.0169e-06
H3	6	1477.765	1477.50	370571	8	5.0368e-06
H3	7	1672.750	1672.50	531144	8	2.6860e-06
H3	8	1577.250	1577.00	487712	8	2.3030e-06
H3	9	1426.684	1426.50	153660	8	2.2018e-06
H3	10	1486.204	1486.00	274980	8	2.1738e-06
I3	1	1773.318	1773.00	462004	8	3.7692e-06
I3	2	1847.207	1847.00	471483	8	2.7871e-06
I3	3	1994.983	1994.75	512935	8	1.8392e-06
I3	4	1862.771	1862.50	426324	8	2.7065e-06
R3	1	931.386	931.25	294878	12	9.6572e-07
R3	1	933.750	931.25	200081	12	9.6572e-07
R3	2	923.608	923.50	241796	16	1.3367e-06
R3	2	927.250	923.50	217876	16	1.3367e-06

Table 8: Hanford 4-km, S3 Kalman line parameters continued

Band	Index	$f_0$	$f_c$	$Q$	$\Delta f$	$v$
A3	1	184.870	184.50	59067	8	1.7559e-06
A3	2	137.625	137.25	8695	8	2.5259e-06
A3	2	137.439	137.25	8575	8	2.5259e-06
A3	2	137.784	137.25	8364	8	2.5259e-06
A3	3	155.642	155.25	7833	8	3.2137e-05
A3	4	150.355	150.00	11641	8	2.5948e-06
B3	1	295.792	295.50	88803	8	1.7635e-06
B3	2	258.819	257.50	69907	12	1.6850e-06
B3	2	257.620	257.50	15151	12	1.6850e-06
B3	2	259.403	257.50	19570	12	1.6850e-06
B3	3	221.844	221.50	64724	8	2.9281e-06
B3	4	267.000	266.75	98617	8	1.0981e-06
C3	1	432.999	431.50	81801	16	8.5591e-07
C3	1	433.173	431.50	24919	16	8.5591e-07
C3	1	432.829	431.50	25124	16	8.5591e-07
C3	1	432.231	431.50	30828	16	8.5591e-07
C3	1	433.761	431.50	30817	16	8.5591e-07
C3	1	434.398	431.50	23383	16	8.5591e-07
C3	1	431.597	431.50	24312	16	8.5591e-07
C3	2	406.715	406.50	122006	8	4.4874e-07
C3	3	443.689	443.50	130522	8	4.1669e-07
C3	4	400.002	399.75	81434	8	4.6002e-07
D3	1	591.585	591.25	176834	8	4.3434e-07
D3	2	517.637	517.50	123820	16	7.5280e-07
D3	2	520.865	517.50	121268	16	7.5280e-07
D3	3	554.611	554.25	140740	8	4.8240e-07
D3	4	512.726	512.50	132403	8	3.9184e-07
D3	5	529.003	528.75	111044	8	4.0161e-07
D3	6	537.142	536.75	125030	8	3.1701e-06
D3	7	545.279	545.00	141549	8	4.8657e-07
D3	8	569.695	569.50	127039	8	4.5525e-07
D3	9	534.000	533.75	189520	8	4.2397e-07
D3	10	551.299	551.00	44174	8	4.3647e-07
D3	11	570.772	570.50	118490	8	4.5411e-07

Table 9: Hanford 2-km, S3 Kalman line parameters

Band	Index	$f_0$	$f_c$	$Q$	$\Delta f$	$v$
E3	1	739.481	739.25	211902	12	1.4206e-07
E3	1	740.604	739.25	190191	12	1.4206e-07
E3	2	776.455	773.00	213934	16	3.3788e-07
E3	2	773.158	773.00	192801	16	3.3788e-07
E3	3	813.429	813.25	212025	8	7.6753e-08
E3	4	716.189	716.00	150390	8	2.5970e-07
E3	5	708.050	707.75	168302	8	1.0928e-07
E3	6	765.020	764.75	197130	8	8.9535e-08
E3	7	724.327	724.00	150323	8	2.4429e-07
E3	8	732.467	732.25	175248	8	9.8033e-08
E3	9	801.000	800.75	269610	8	7.8696e-08
F3	1	866.000	865.75	267239	8	8.1725e-08
F3	2	1000.000	998.00	361785	12	7.7418e-08
F3	2	998.299	998.00	263271	12	7.7418e-08
F3	3	887.377	887.00	251798	8	6.2372e-08
G3	1	1331.066	1330.75	379092	8	5.0779e-08
G3	2	1298.999	1298.75	359094	8	4.9757e-08
G3	3	1257.118	1256.75	357617	8	8.9992e-08
G3	4	1072.247	1072.00	297874	8	5.5602e-08
G3	5	1183.169	1183.00	325128	8	5.1539e-08
H3	1	1626.858	1626.50	467533	8	6.0395e-08
H3	2	1700.806	1700.50	461920	8	6.6547e-08
H3	3	1552.910	1552.75	426357	8	5.8778e-08
H3	4	1478.962	1478.75	372932	8	5.4200e-08
H3	5	1587.099	1586.75	139230	8	7.8578e-07
H3	6	1512.723	1512.50	67244	8	8.8186e-08
H3	7	1474.468	1474.25	250750	8	6.6713e-08
H3	8	1418.361	1418.00	273790	8	5.9116e-08
R3	1	903.373	903.00	194739	8	2.6961e-07
R3	2	931.418	927.50	250128	16	1.1510e-07
R3	2	927.790	927.50	243060	16	1.1510e-07
R3	3	911.514	911.25	166713	8	7.2944e-08
R3	4	895.236	895.00	230290	8	6.5898e-08
R3	5	935.928	935.75	227913	8	6.8618e-08
R3	6	919.652	919.50	196327	8	5.9984e-08

Table 10: Hanford 2-km, S3 Kalman line parameters continued



Band	Index	$f_0$	$f_c$	$Q$	$\Delta f$	$v$
A3	1	143.147	142.75	44455	8	1.6043e-06
A3	2	190.862	190.50	50761	8	2.0054e-06
A3	3	131.857	131.50	21600	8	1.7299e-06
B3	1	237.788	237.50	21486	16	4.0554e-06
B3	1	241.544	237.50	13344	16	4.0554e-06
B3	1	241.960	237.50	26395	16	4.0554e-06
B3	2	251.756	251.50	14331	8	2.3387e-06
B3	3	294.937	294.75	79712	8	2.2760e-06
B3	4	221.203	221.00	54617	8	2.2223e-06
B3	5	242.169	242.00	39484	8	2.4708e-06
B3	6	258.070	257.75	56511	8	2.3401e-06
C3	1	405.538	405.25	135179	8	1.7829e-06
C3	2	442.405	442.25	118199	8	1.6988e-06
D3	1	516.139	515.75	118847	8	1.3204e-06
D3	2	589.874	589.50	174962	8	1.1622e-06
D3	3	626.742	626.50	181664	8	1.2686e-06
D3	4	553.007	552.75	146169	8	1.2159e-06
E3	1	737.342	737.00	215957	8	1.0476e-06
E3	2	811.076	810.75	240573	8	9.4177e-07
E3	3	774.210	774.00	202976	8	9.8769e-07
F3	1	884.803	884.50	213849	8	9.1640e-07
F3	2	995.410	995.25	269029	8	9.7339e-07
F3	3	921.672	921.50	253206	32	2.8254e-06
F3	3	925.606	921.50	112878	32	2.8254e-06
F3	3	929.264	921.50	101837	32	2.8254e-06
F3	3	929.726	921.50	36147	32	2.8254e-06
F3	3	933.609	921.50	193828	32	2.8254e-06
F3	4	847.944	847.75	250623	8	9.4388e-07
F3	5	926.090	925.75	35257	8	1.5033e-06
F3	5	926.049	925.75	142469	8	1.5033e-06
F3	6	939.556	939.25	65398	8	9.5486e-07
F3	7	915.825	915.50	43819	8	9.1476e-07
F3	8	958.543	958.25	225380	8	9.6067e-07

Table 11: Livingston 4-km, S3 Kalman line parameters

Band	Index	$f_0$	$f_c$	$Q$	$\Delta f$	$v$
G3	1	1106.013	1105.75	302425	8	1.1182e-06
G3	2	1327.216	1327.00	388724	8	1.8624e-06
G3	3	1253.482	1253.25	378205	8	1.5780e-06
G3	4	1179.747	1179.50	342665	8	1.3382e-06
G3	5	1142.880	1142.50	312506	8	1.2131e-06
G3	6	1216.615	1216.25	299869	8	1.4450e-06
G3	7	1364.084	1363.75	231200	8	1.9808e-06
G3	8	1069.147	1068.75	311402	8	1.0640e-06
G3	9	1334.164	1334.00	115762	8	1.9289e-06
H3	1	1695.887	1691.75	418000	16	6.2473e-06
H3	1	1692.000	1691.75	143086	16	6.2473e-06
H3	1	1692.439	1691.75	131536	16	6.2473e-06
H3	2	1474.684	1474.50	355958	8	2.4964e-06
H3	3	1548.419	1548.25	399960	8	2.7818e-06
H3	4	1622.153	1622.00	445296	8	2.8605e-06
H3	5	1511.551	1511.25	345779	8	2.6414e-06
H3	6	1437.819	1437.50	359454	8	2.6093e-06
H3	7	1585.289	1585.00	332578	8	2.8504e-06
I3	1	1769.621	1769.25	463945	8	2.4232e-06
I3	2	1843.356	1843.00	492499	8	1.9138e-06
I3	3	1917.090	1916.75	530420	8	1.7582e-06
I3	4	1957.703	1957.50	431212	8	1.6844e-06
I3	5	2031.437	2031.25	447453	8	1.7804e-06
I3	6	1867.219	1867.00	369746	8	1.7910e-06
I3	7	1806.487	1806.25	470439	8	2.1561e-06
I3	8	1990.826	1990.50	530886	8	1.6808e-06
I3	9	1761.500	1758.25	338749	16	4.9036e-06
I3	9	1758.472	1758.25	254851	16	4.9036e-06
R3	1	921.679	921.50	237196	8	9.2760e-07
R3	2	929.316	926.00	64017	16	1.8036e-06
R3	2	926.096	926.00	49790	16	1.8036e-06
R3	3	933.610	933.25	229576	8	9.6617e-07
R3	4	939.562	939.25	27715	8	9.5488e-07
R3	5	915.814	915.50	42268	8	9.1521e-07

Table 12: Livingston 4-km, S3 Kalman line parameters continued

Band	$B_{lo}$	$B_{hi}$	$n$	$T_{lo}$	$T_{hi}$
A2	128	192	52	0.4688	0.5469
B2	192	320	56	0.4688	0.5391
C2	384	512	32	0.5000	0.6250
D2	512	640	56	0.4688	0.5391
E2	704	1024	70	0.5000	0.5563
F2	1065	1365	48	0.4833	0.5667
G2	1408	1708	48	0.4833	0.5667
H2	1758	2048	40	0.5000	0.5862
R2	900	930	40	0.5000	0.6000

Table 13: S2 Bands

Band	$B_{lo}$	$B_{hi}$	$n$	$T_{lo}$	$T_{hi}$
A3	128	192	26	0.4375	0.5938
B3	192	320	36	0.4531	0.5625
C3	384	512	32	0.4531	0.5781
D3	512	640	32	0.4531	0.5781
E3	704	832	50	0.4688	0.5469
F3	832	1024	76	0.4792	0.5312
G3	1065	1365	74	0.4800	0.5333
H3	1408	1708	74	0.4800	0.5333
I3	1758	2048	72	0.4793	0.5345
R3	890	940	20	0.4200	0.6200

Table 14: S3 Bands

## B Frequency Band Definitions and Filters

The following tables give the band definitions and specify parameters for the lowpass filters used in basebanding. Each band covers the frequency range from  $B_{lo}$  to  $B_{hi}$  Hz. There are gaps in frequency coverage around the violin modes and violin mode harmonics. The violin modes have a lot of power and can vary considerably in power over short timescales, making them difficult to remove. Therefore, we have chosen to avoid them. The R band is a narrow band chosen to correspond to the frequency range of some bar detectors, allowing comparison with their results.

The lowpass filter used in the basebanding step of data conditioning is created using Matlab's `firpm` function. The filter `b` is calculated via the command `b = firpm(n, f, a, w)` where `n` is the length of the filter, `f = [0,  $T_{lo}$ ,  $T_{hi}$ , 1]` is a vector of frequencies (as fraction of Nyquist) and `a = [1, 1, 0, 0]` are the amplitudes of the filter's frequency response at these frequencies. `w = [1, 1]` specifies the allowed amount of passband and stopband ripple (see [6]). `n`,  $T_{lo}$ , and  $T_{hi}$  are given in the tables.

Band	$B_{lo}$	$B_{hi}$	$n$	$T_{lo}$	$T_{hi}$
A4	96	192	38	0.4583	0.5625
B4	192	320	32	0.4531	0.5781
C4	384	512	28	0.4531	0.5938
D4	512	640	28	0.4531	0.5938
E4	704	832	50	0.4688	0.5469
F4	832	1024	76	0.4792	0.5312
G4	1065	1365	74	0.4800	0.5333
H4	1408	1708	74	0.4800	0.5333
I4	1758	2048	72	0.4793	0.5345
R4	890	940	20	0.4375	0.5833

Table 15: S4 Bands

## C Line Regression Parameters

The following tables give the parameters used in line regression for S2 and S3 for each detector. The parameters are given by frequency band and frequency band definitions may be found in Appendix B. There are two types of lines that are regressed: power lines at 60 Hz and harmonics caused by the AC oscillations of the US power grid and calibration lines which are injected into the interferometers by wiggling mirrors and used for calibration purposes. Parameters used to remove power lines are marked as “power” in the table while those used to remove calibration lines are designated “calib”.

When line regression is performed, data from an environmental or interferometer channel is put through the same conditioning and bandlimiting procedure (minus Kalman filtering) as the data from the gravitational wave channel of the detector. An ARX model is then found that relates the regression channel data to the detector data. The regression channel is filtered using the ARX model and the result subtracted from the detector data. If the bandlimited detector data is stored in the variable `band_data` and the bandlimited regression data in `band_reg`, the lines can be regressed using the following MATLAB commands:

```
data= iddata(band_data, band_reg);
model = arx(data, orders);
b = model.b;
filt_data = band_data-filter(b,1,band_reg);
```

`orders` is a vector whos three elements are given in the tables below.

Two channels are listed for some of the power line regression parameters because, in those cases, regression based on no single channel could remove both the even and odd harmonics. When two channels are listed, the power lines are subtracted as much as possible with the first channel and then the filtered data is used to find the  $b$  coefficients for the second channel:

```
data1 = iddata(band_data', band_reg1');
modell = arx(data1, orders1);
b1 = modell.b;
filt_data1 = band_data-filter(b1,1,band_reg1);
data2 = iddata(filt_data1', band_reg2');
model2 = arx(data2, orders2);
b2 = model2.b;
```

When the data are filtered, both sets of channels and filters are used:

```
filt_data = band_data-filter(b1,1,band_reg1) ...
- filter(b2,1,band_reg2);
```

This regression approach, where the filter for the second channel is dependent on regression applied with the first channel is specified by the column labeled `Dep.` in the tables below. A nonzero `Dep.` value indicates that dependent regression is used and the number given is the Index value of the regression parameters that should be applied first.

As can be seen from Tables C and C, regression is not used to remove power lines from Livingston 4-km data above 1024 Hz. Unfortunately, no suitable channel could be found. L1:LSC-POB\_Q was not correlated enough with the power lines to be used to remove them and the environmental channels contain no information at those frequencies because they are sampled at 2048Hz.

Band	Index	Type	Channel	orders	Dep.
A2	1	power	H0:PEM-LVEA2_V1	0 6 0	0
A2	2	calib	H1:LSC-ETMX_EXC_DAQ	0 6 0	0
B2	1	power	H0:PEM-LVEA2_V1	0 20 0	0
C2	1	power	H0:PEM-LVEA2_V1	0 16 0	0
D2	1	power	H0:PEM-LVEA2_V1	0 8 0	0
E2	1	power	H0:PEM-LVEA2_V1	0 90 0	0
E2	2	calib	H1:LSC-ETMX_EXC_DAQ	0 8 0	0
F2	1	power	H1:LSC-POB_Q	0 90 0	0
G2	1	power	H1:LSC-POB_Q	0 90 0	0
H2	1	power	H1:LSC-POB_Q	0 90 0	0
R2	1	power	H0:PEM-LVEA2_V1	0 10 0	0

Table 16: Hanford 4-km, S2 line regression parameters

Band	Index	Type	Channel	orders	Dep.
A2	1	power	H0:PEM-LVEA2_V3	0 6 0	0
A2	2	calib	H2:LSC-ETMX_EXC_DAQ	0 6 0	0
B2	1	power	H0:PEM-LVEA2_V3	0 40 0	0
C2	1	power	H0:PEM-LVEA2_V3	0 50 0	0
D2	1	power	H0:PEM-LVEA2_V3	0 40 0	0
E2	1	power	H0:PEM-LVEA2_V3	0 90 0	0
E2	2	calib	H2:LSC-ETMX_EXC_DAQ	0 8 0	0
F2	1	power	H2:LSC-POB_Q	0 90 0	0
G2	1	power	H2:LSC-POB_Q	0 90 0	0
H2	1	power	H2:LSC-POB_Q	0 90 0	0
R2	1	power	H0:PEM-LVEA2_V3	0 10 0	0

Table 17: Hanford 2-km, S2 line regression parameters

Instead, these lines are treated with a Kalman filter which does not perform as well as regression on power lines but is able to accomplish some reduction of power.

Band	Index	Type	Channel	orders	Dep.
A2	1	power	L0:PEM-LVEA_V1	0 6 0	0
A2	2	calib	L1:LSC-ETMX_EXC_DAQ	0 6 0	0
B2	1	power	L0:PEM-RADIO_LVEA	0 50 0	0
B2	2	power	L0:PEM-LVEA_V1	0 50 0	1
C2	1	power	L0:PEM-RADIO_LVEA	0 50 0	0
C2	2	power	L0:PEM-LVEA_V1	0 50 0	1
D2	1	power	L0:PEM-RADIO_LVEA	0 70 0	0
D2	2	power	L0:PEM-LVEA_V1	0 70 0	1
E2	1	power	L0:PEM-RADIO_LVEA	0 90 0	0
E2	2	power	L0:PEM-LVEA_V1	0 90 0	1
E2	3	power	L1:LSC-ETMX_EXC_DAQ	0 8 0	1
R2	1	power	L0:PEM-LVEA_V1	0 10 0	0
R2	2	calib	L1:LSC-ETMX_EXC_DAQ	0 10 0	0

Table 18: Livingston 4-km, S2 line regression parameters

Band	Index	Type	Channel	orders	Dep.
A3	1	power	H0:PEM-LVEA2_V1	0 2 0	0
A3	2	calib	H1:LSC-DARM_CTRL_EXC_DAQ	0 10 0	0
B3	1	power	H0:PEM-LVEA2_V1	0 150 0	0
C3	1	power	H0:PEM-LVEA2_V1	0 16 0	0
D3	1	power	H0:PEM-LVEA2_V1	0 8 0	0
E3	1	power	H0:PEM-LVEA2_V1	0 150 0	0
F3	1	calib	H1:LSC-DARM_CTRL_EXC_DAQ	0 4 0	0
F3	2	power	H0:PEM-LVEA2_V1	0 50 0	1
G3	1	power	H1:LSC-POB_Q	0 90 0	0
H3	1	power	H1:LSC-POB_Q	0 90 0	0
I3	1	power	H1:LSC-POB_Q	0 10 0	0
R3	1	power	H0:PEM-LVEA2_V1	0 10 0	0

Table 19: Hanford 4-km, S3 line regression parameters

Band	Index	Type	Channel	orders	Dep.
A3	1	power	H0:PEM-LVEA2_V3	0 2 0	0
A3	2	calib	H2:LSC-DARM_CTRL_EXC_DAQ	0 6 0	0
B3	1	power	H0:PEM-LVEA2_V3	0 50 0	0
C3	1	power	H0:PEM-LVEA2_V3	0 50 0	0
D3	1	power	H0:PEM-LVEA2_V3	0 40 0	0
E3	1	power	H0:PEM-LVEA2_V3	0 50 0	0
F3	1	calib	H2:LSC-DARM_CTRL_EXC_DAQ	0 10 0	0
F3	2	power	H0:PEM-LVEA2_V3	0 150 0	0
G3	1	power	H2:LSC-POB_Q	0 150 0	0
H3	1	power	H2:LSC-POB_Q	0 150 0	0
H3	2	calib	H2:LSC-DARM_CTRL_EXC_DAQ	0 10 0	0
R3	1	power	H0:PEM-LVEA2_V3	0 50 0	0

Table 20: Hanford 2-km, S3 line regression parameters

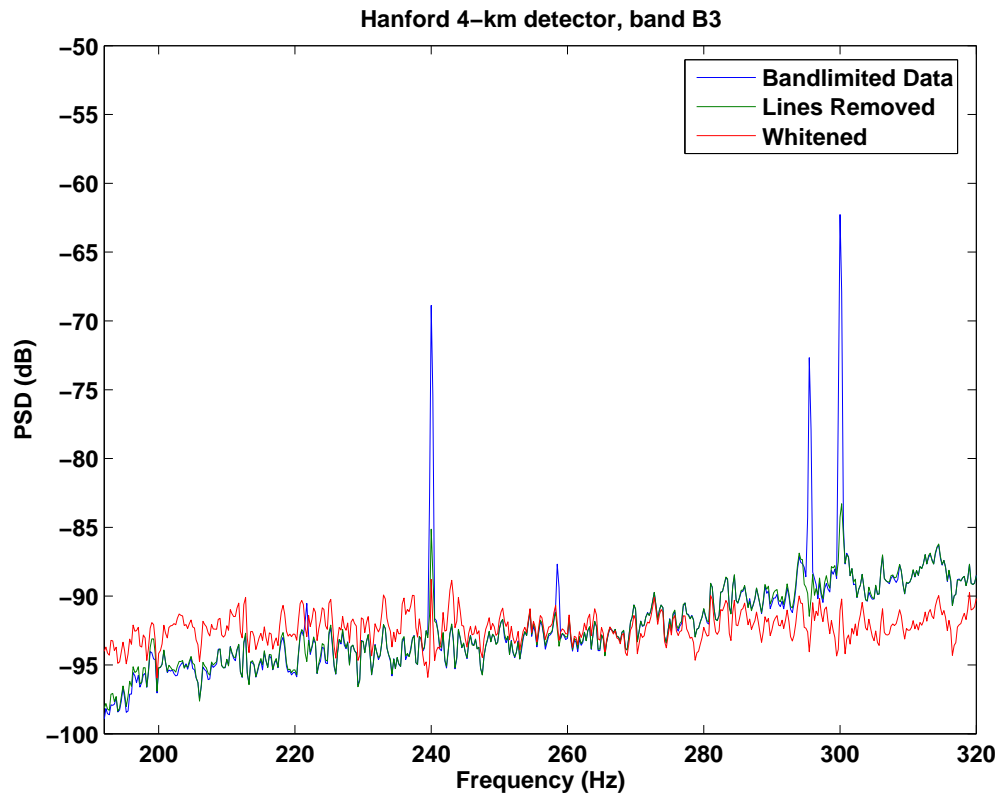
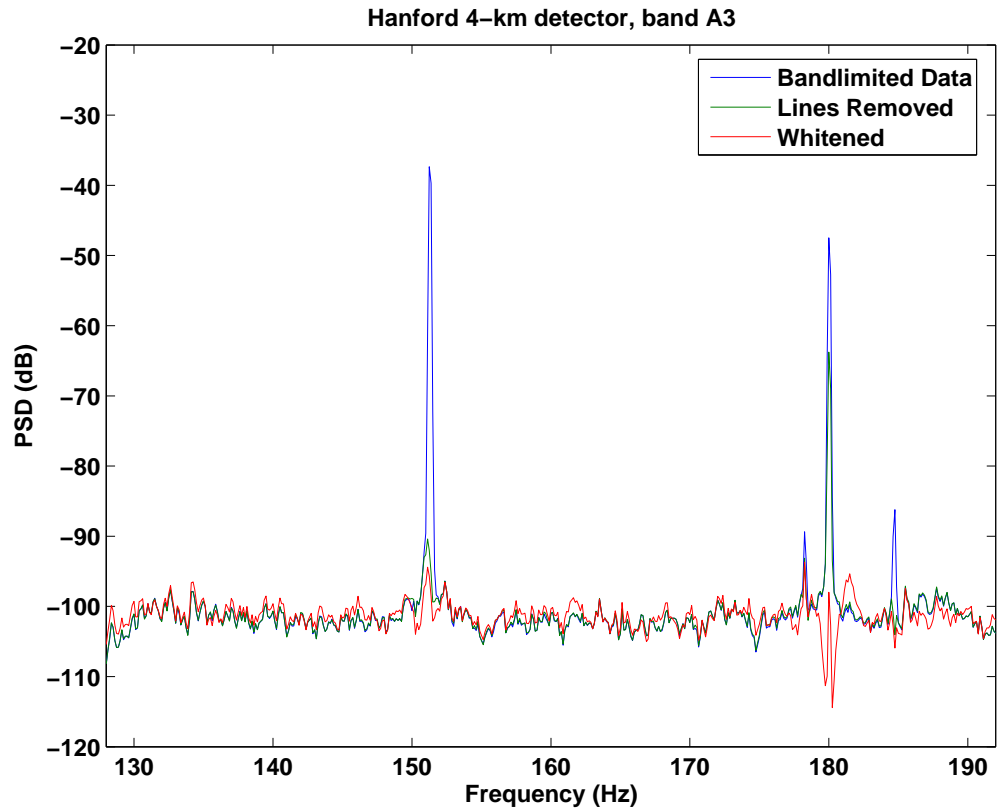
Band	Index	Type	Channel	orders	Dep.
A3	1	power	L0:PEM-LVEA_V1	0 2 0	0
A3	2	calib	L1:LSC-DARM_CTRL_EXC_DAQ	0 22 0	0
B3	1	power	L0:PEM-RADIO_LVEA	0 2 0	0
B3	2	power	L0:PEM-LVEA_V1	0 2 0	1
C3	1	power	L0:PEM-RADIO_LVEA	0 2 0	0
C3	2	power	L0:PEM-LVEA_V1	0 50 0	1
D3	1	power	L0:PEM-RADIO_LVEA	0 8 0	0
D3	2	power	L0:PEM-LVEA_V1	0 50 0	1
E3	1	power	L0:PEM-RADIO_LVEA	0 90 0	0
E3	2	power	L0:PEM-LVEA_V1	0 90 0	1
F3	1	power	L0:PEM-RADIO_LVEA	0 90 0	0
F3	2	power	L0:PEM-LVEA_V1	0 90 0	1
F3	3	calib	L1:LSC-DARM_CTRL_EXC_DAQ	0 4 0	0
R3	1	power	L0:PEM-LVEA_V1	0 10 0	0
R3	2	calib	L1:LSC-DARM_CTRL_EXC_DAQ	0 10 0	0

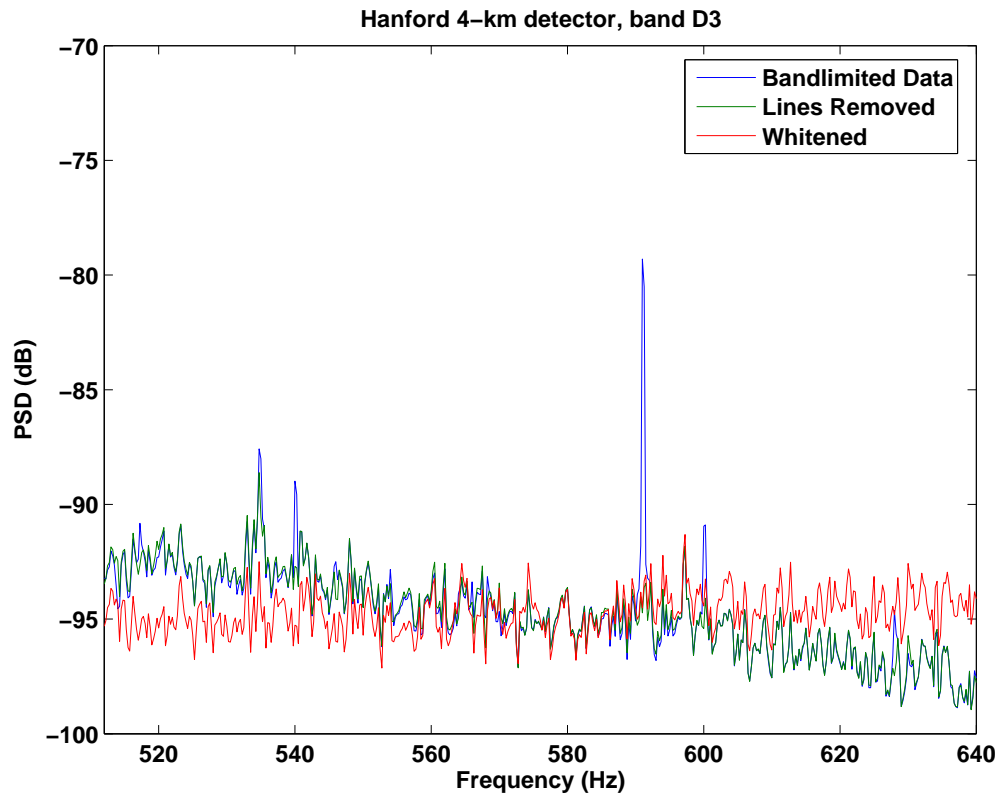
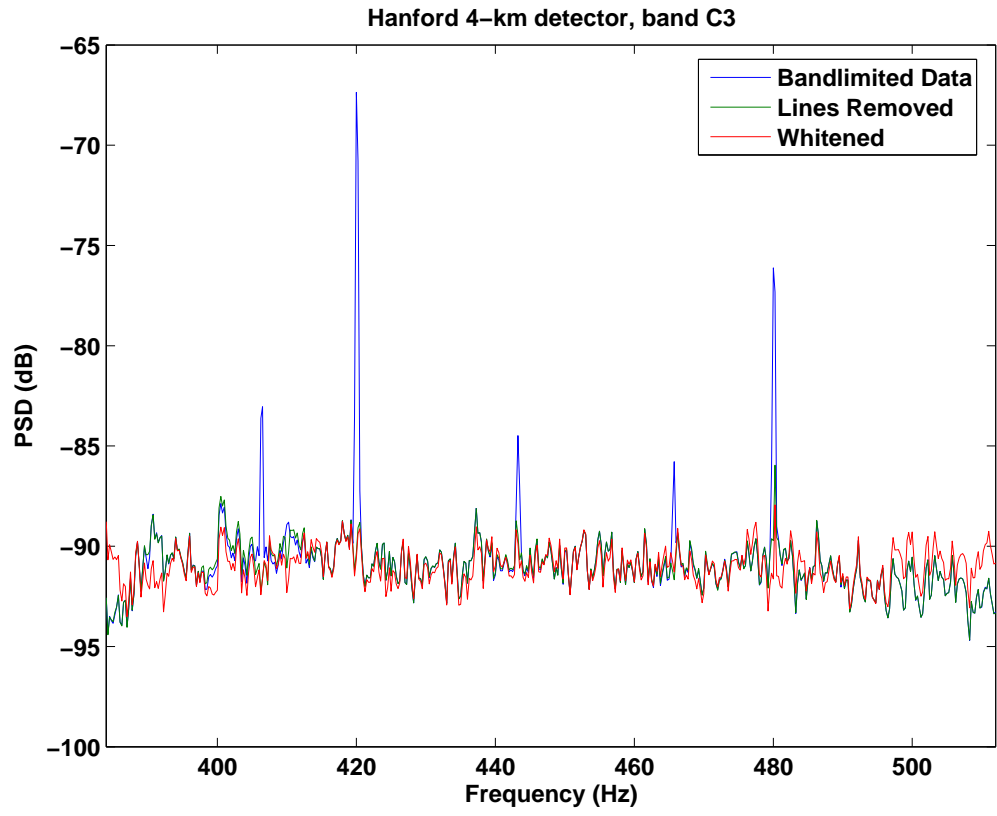
Table 21: Livingston 4-km, S3 line regression parameters

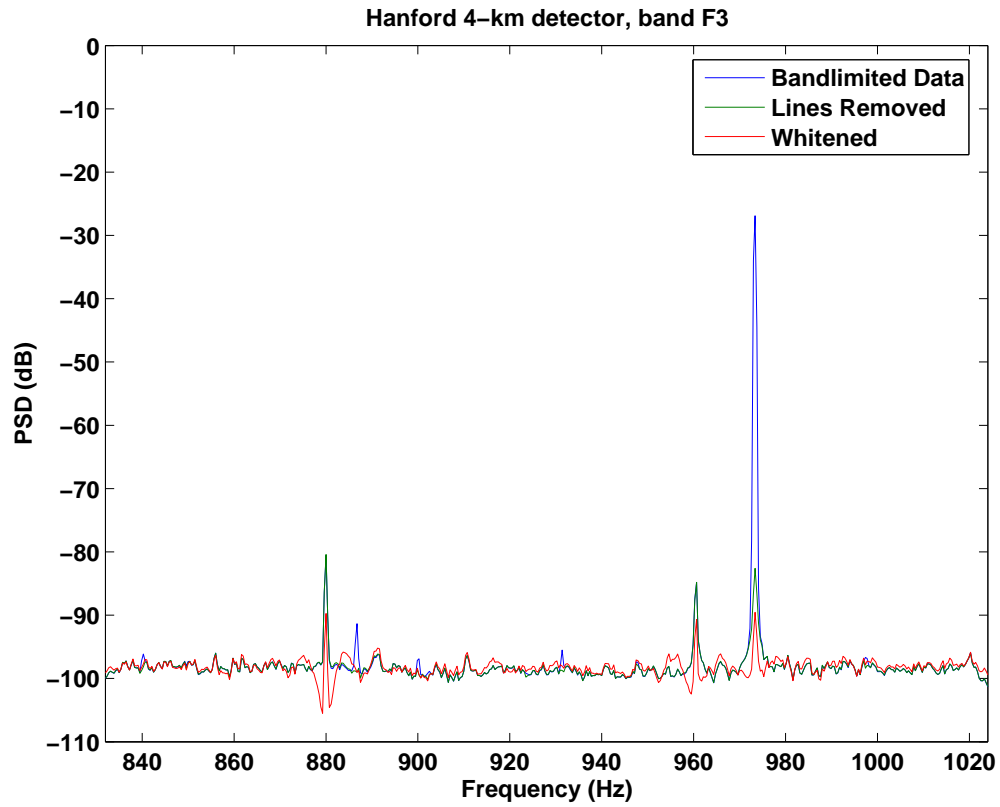
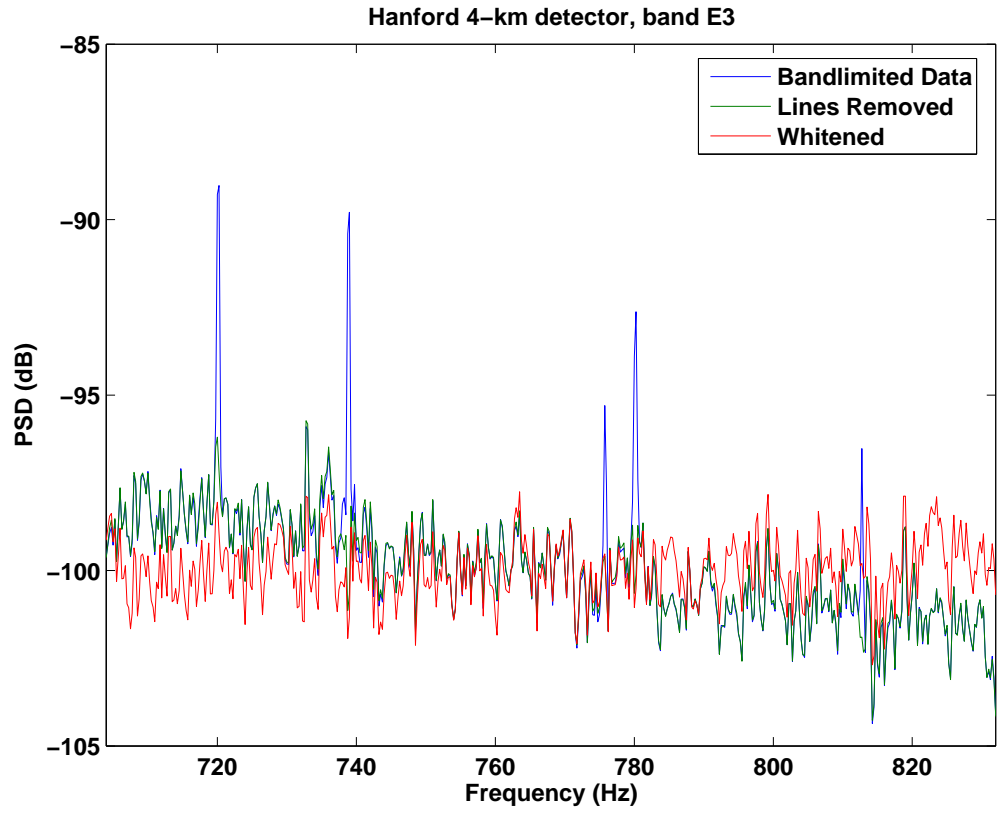


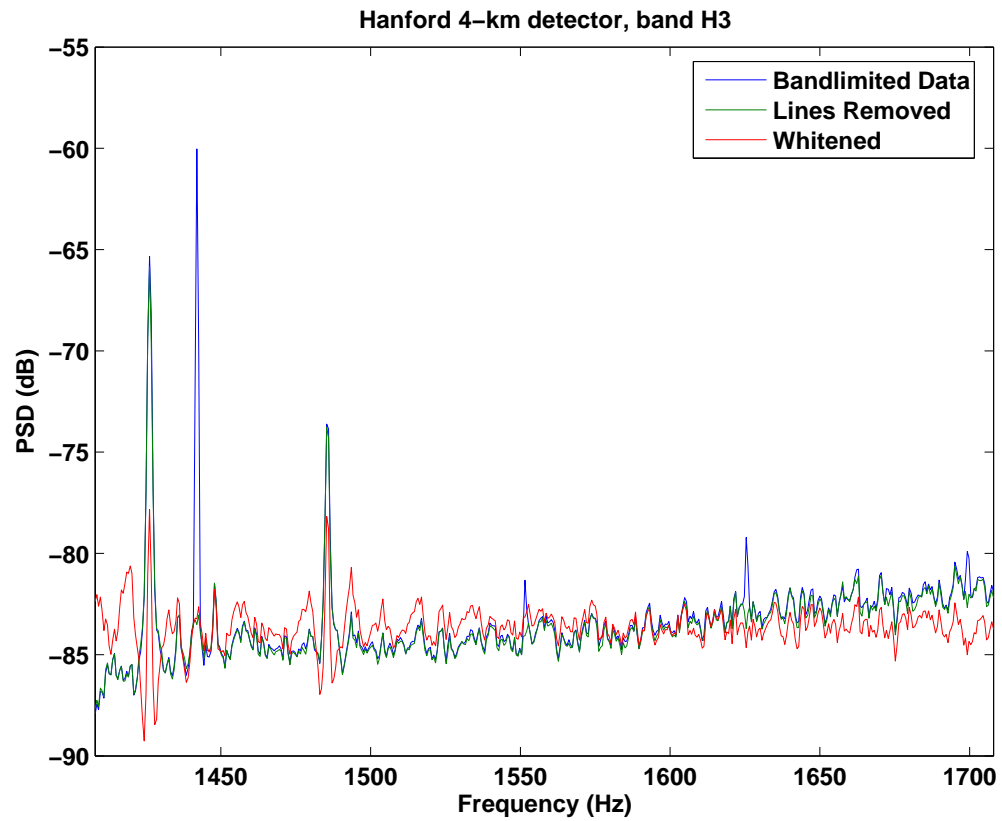
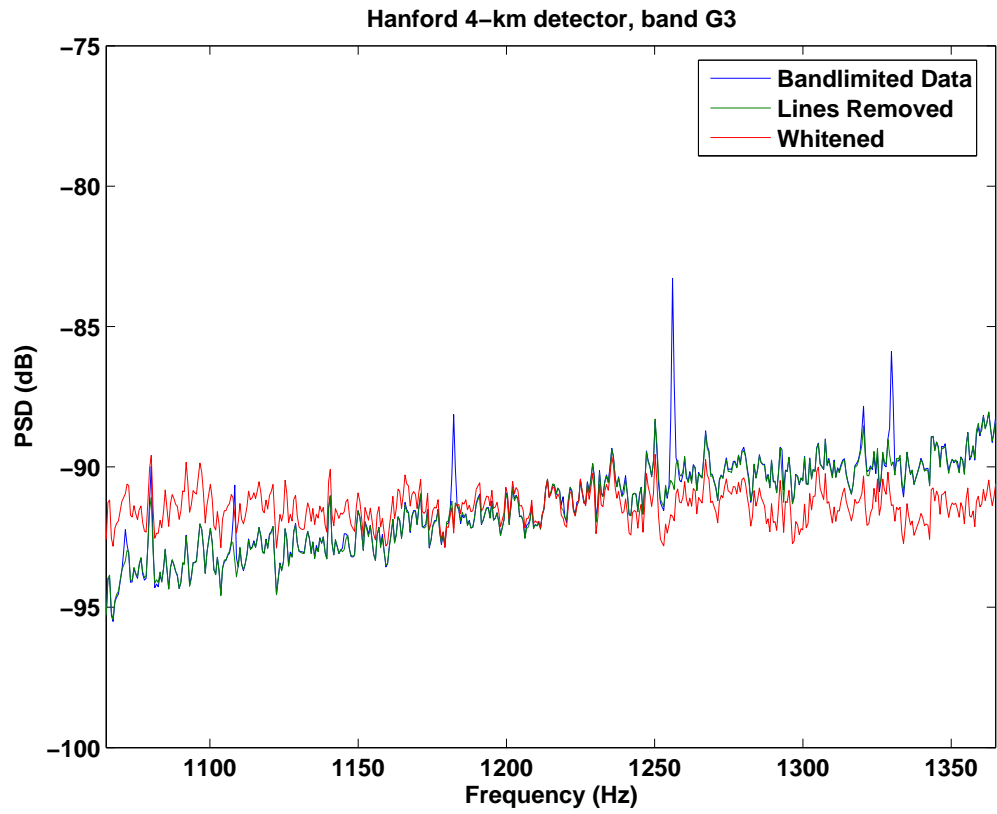
## **D Data conditioning PSDs**

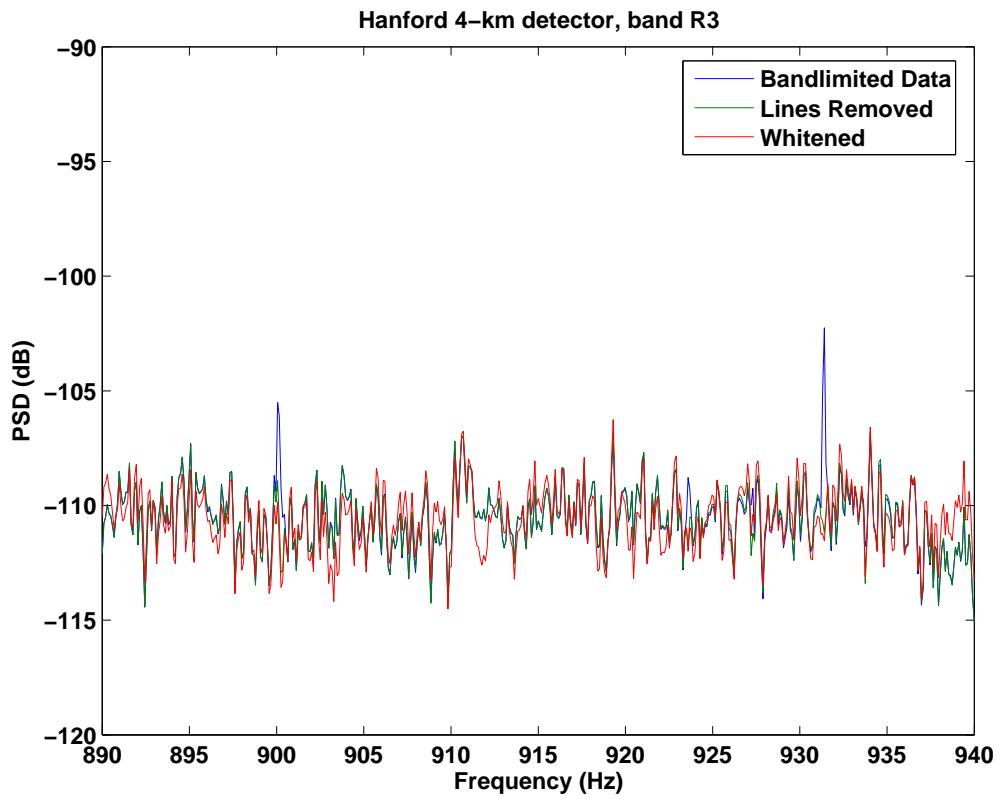
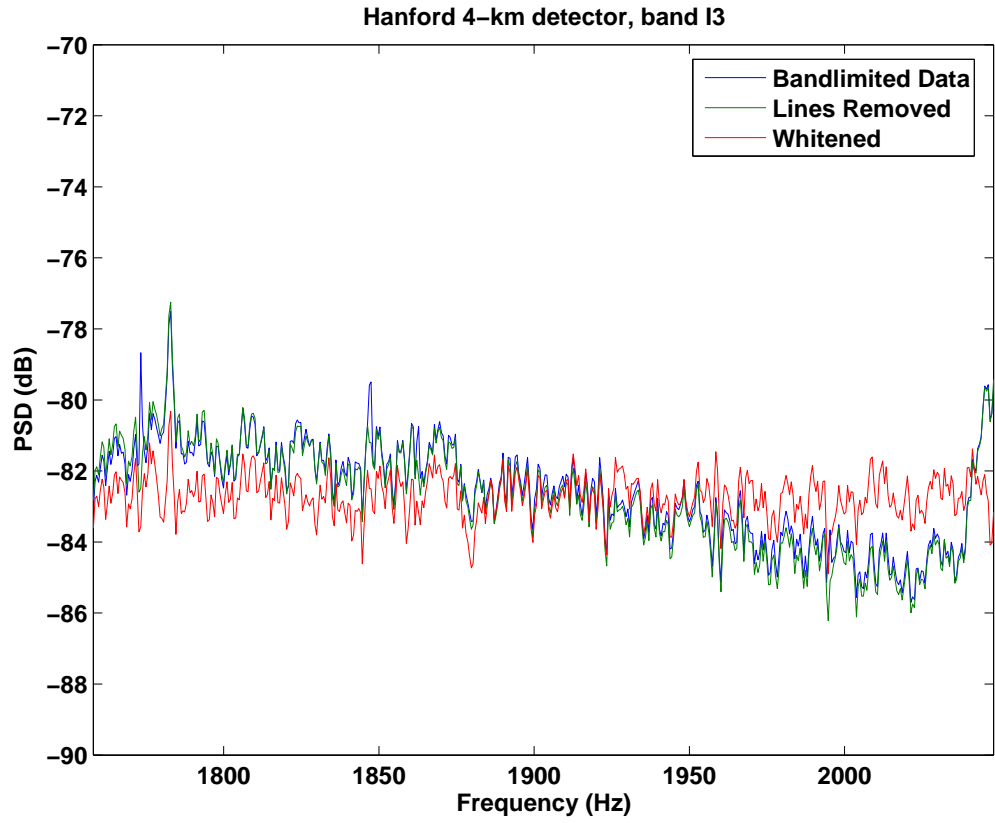
Below are power spectral density (PSD) plots that show the effects of different data conditioning steps. Each plot shows the effects of data conditioning for a different frequency band and detector. The frequency band definitions are given in Appendix B. The plots are made with data from GPS time 756715873 which is 300 seconds in duration. The data conditioning uses filters calculated on the previous 300 seconds. In each plot, the blue curve is the PSD of data that has been base-banded only, the green curve shows the effect of line removal including Kalman filtering and line regression, and the red curve is the PSD of the fully conditioned data including whitening.

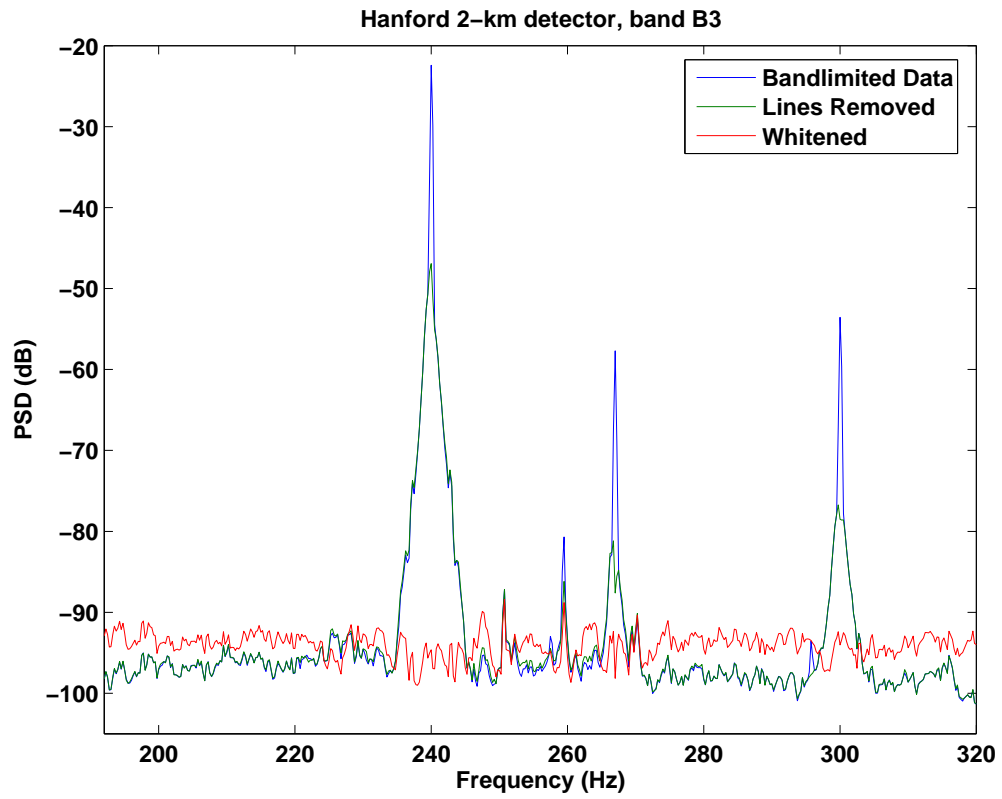
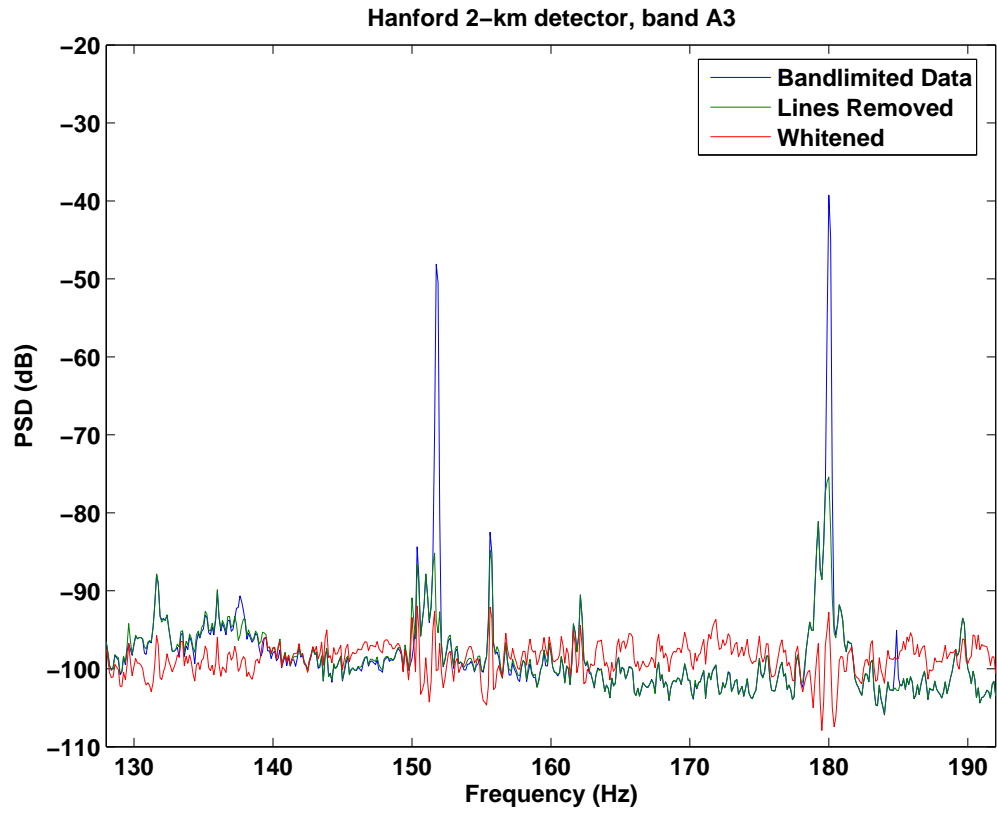


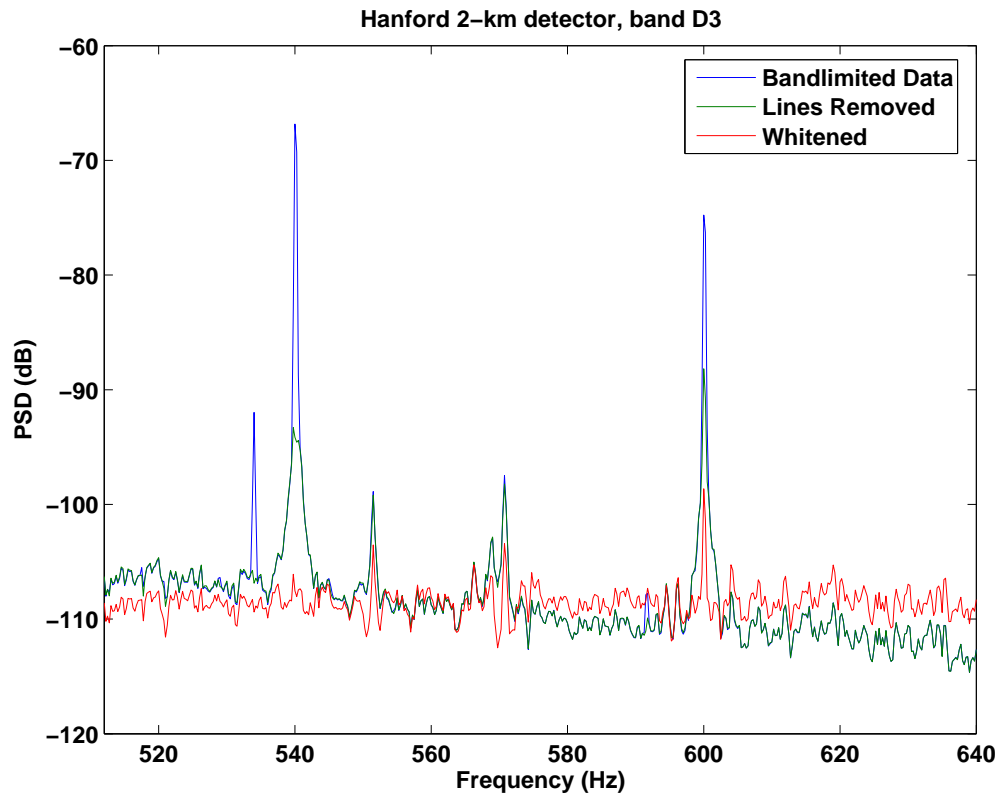
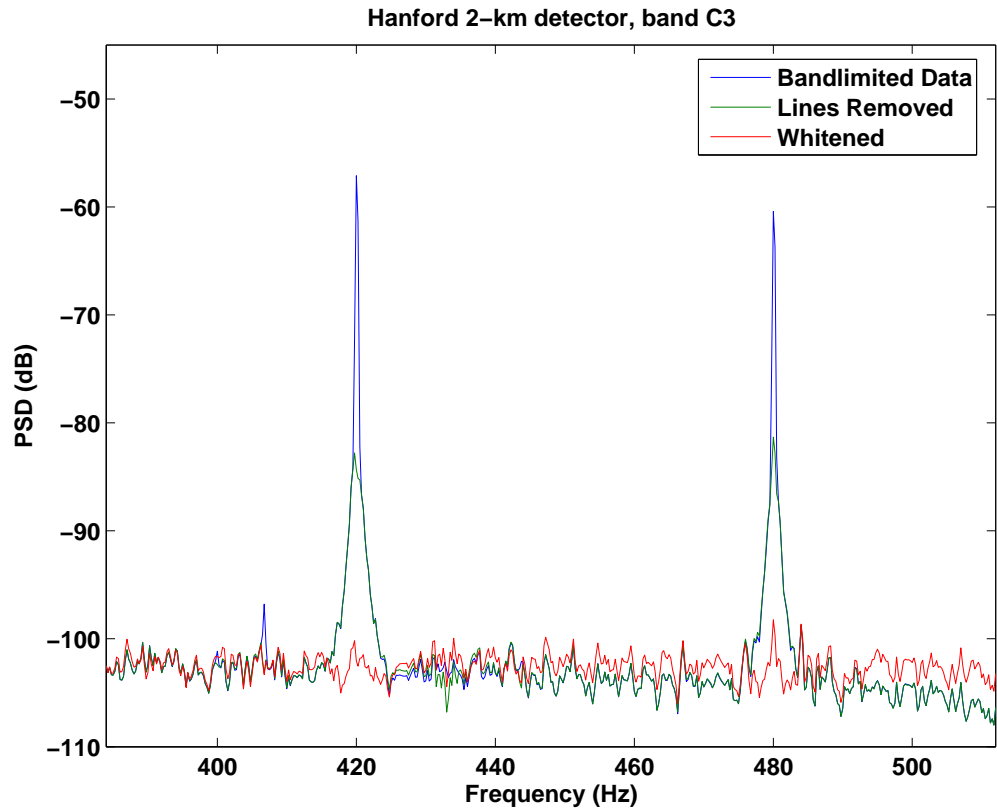




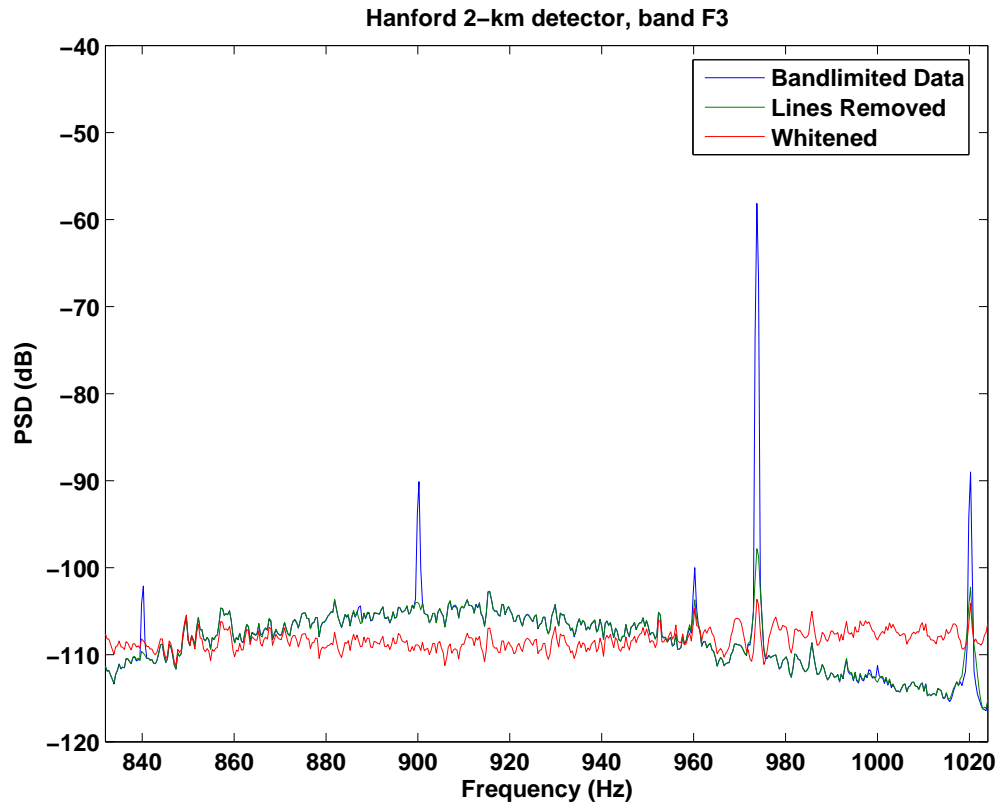
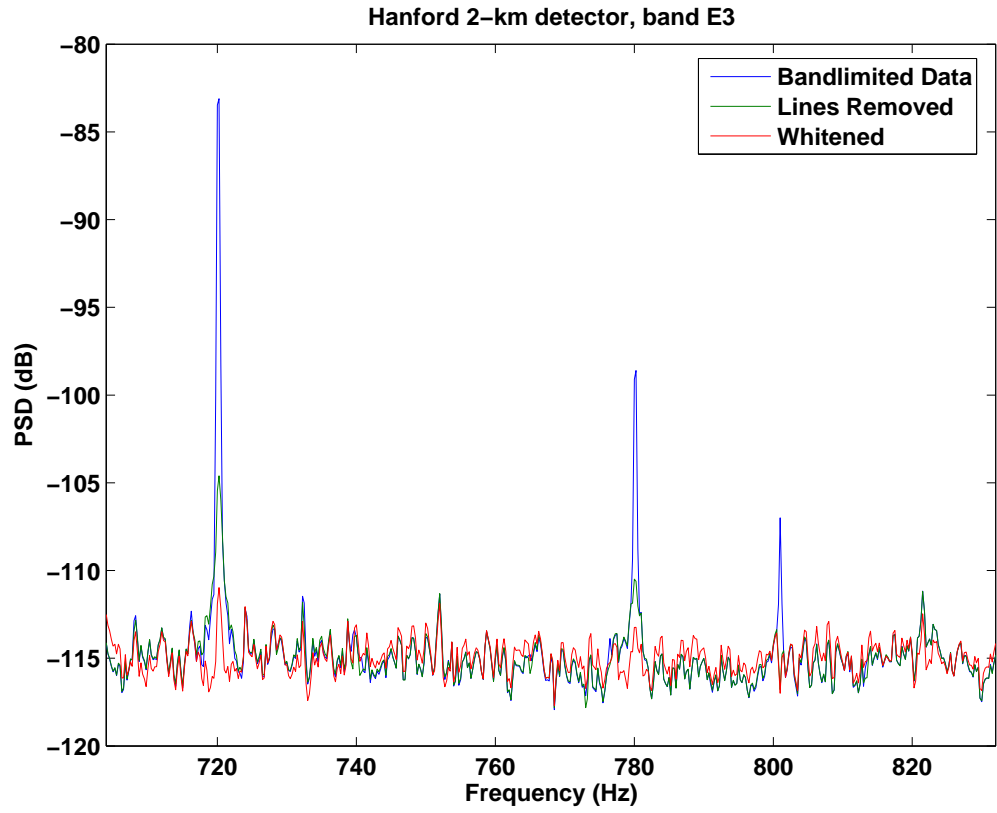


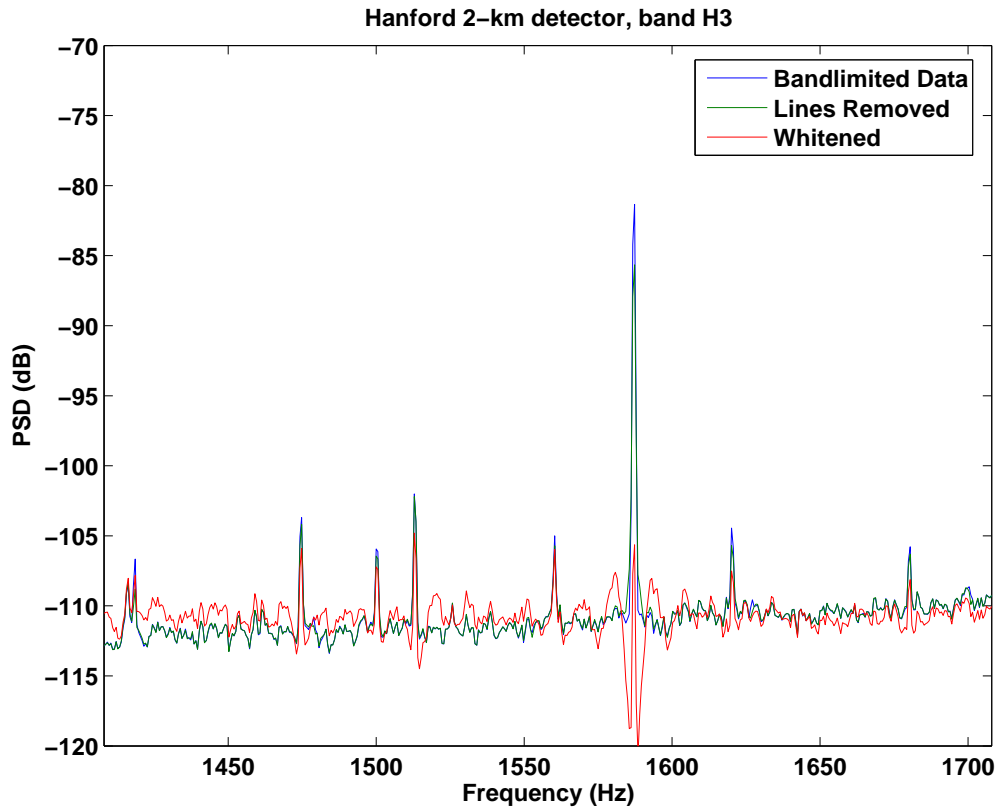
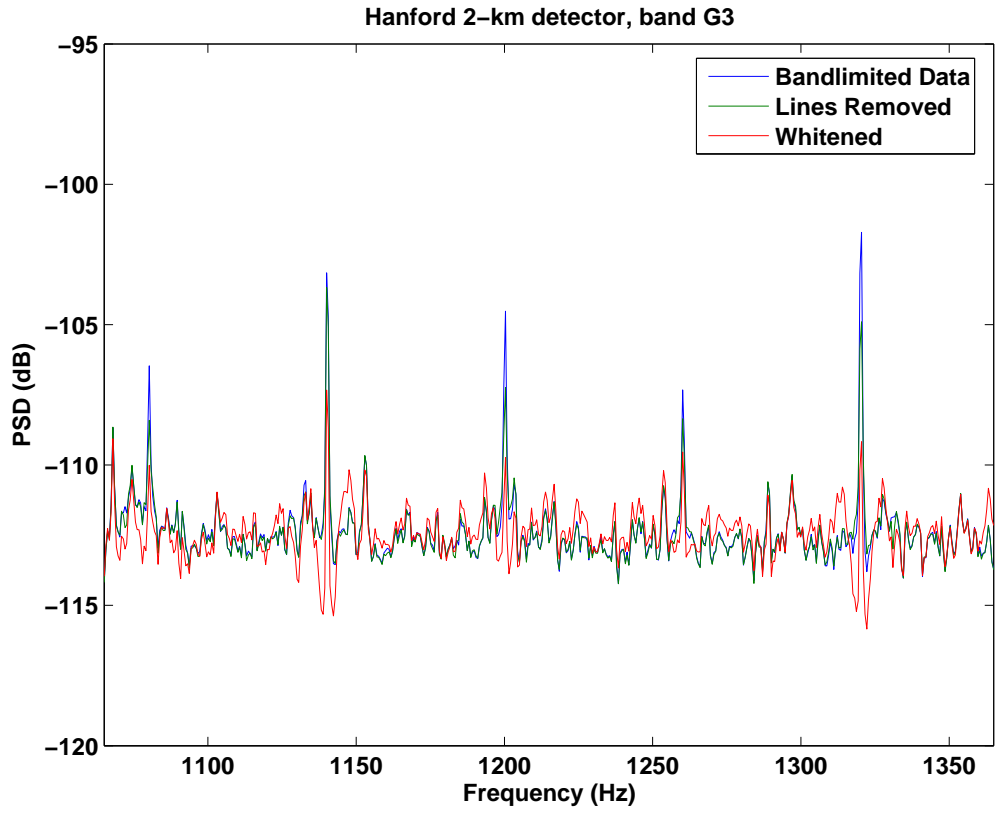


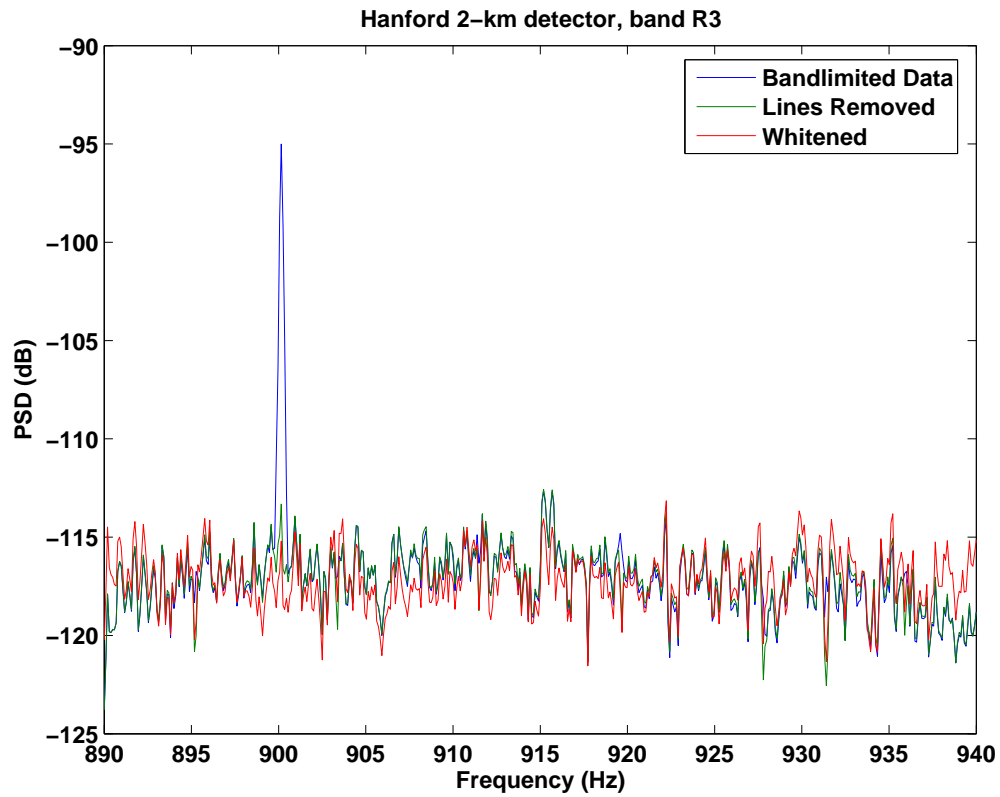
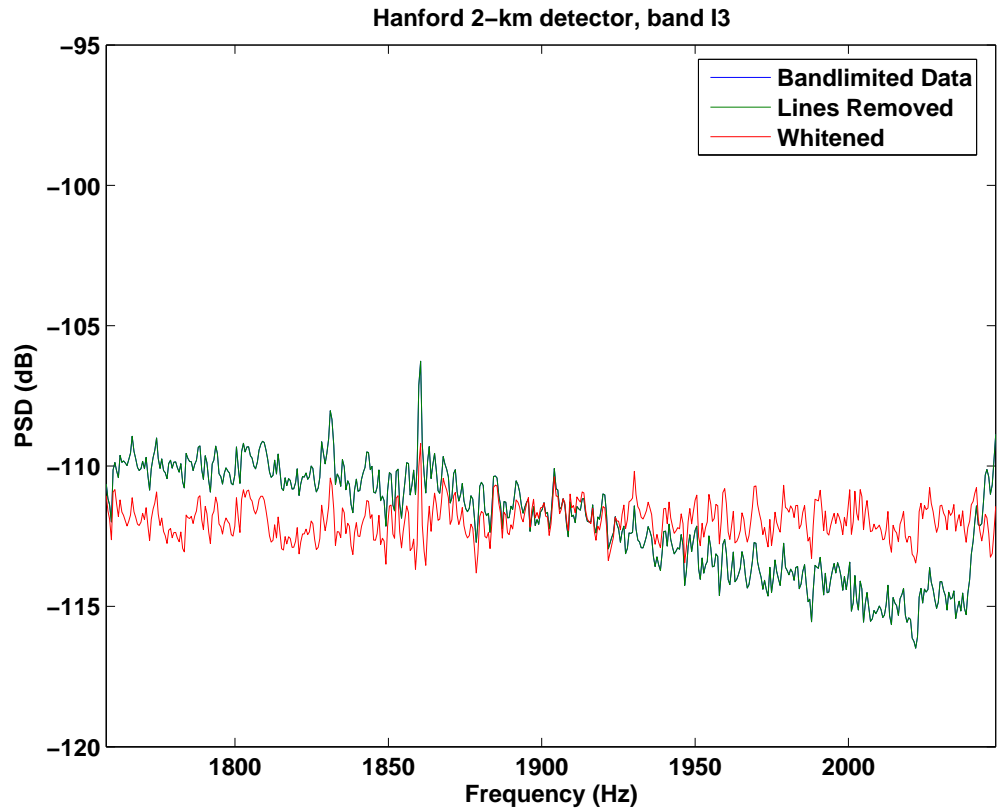


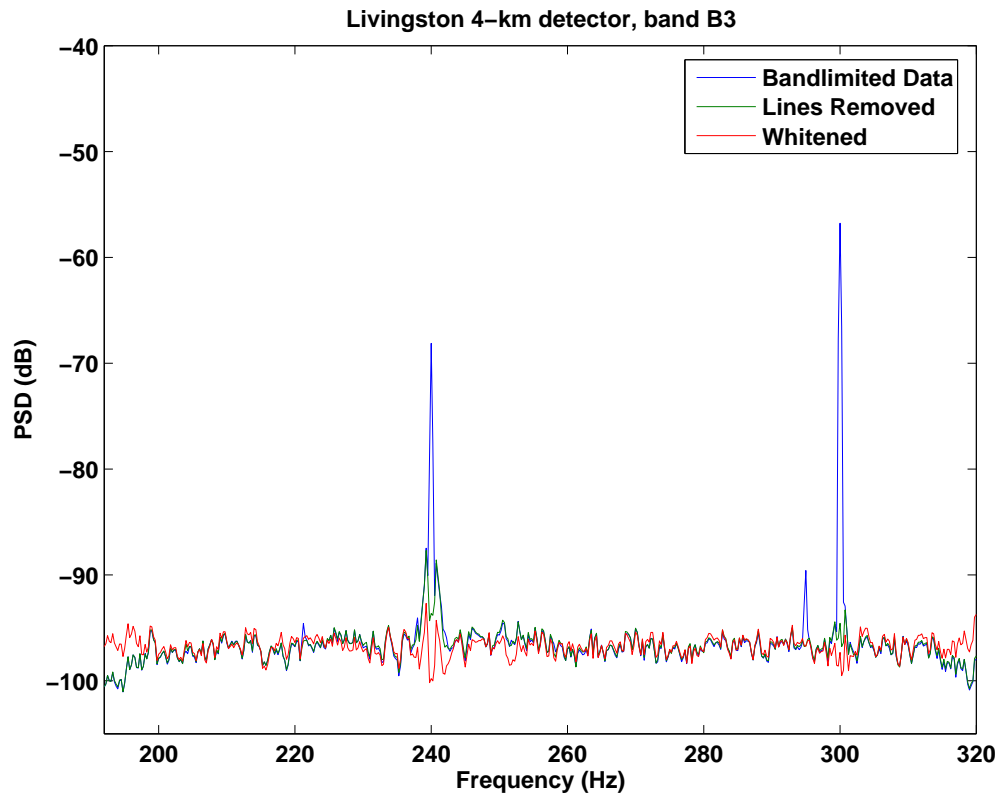
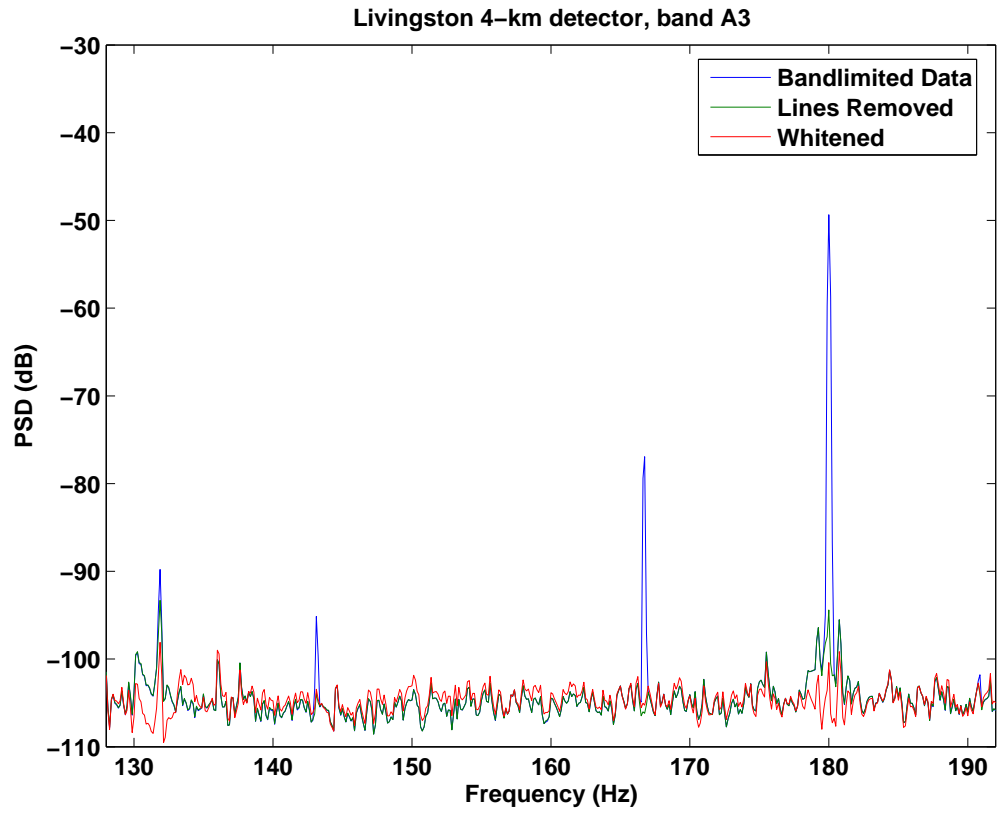


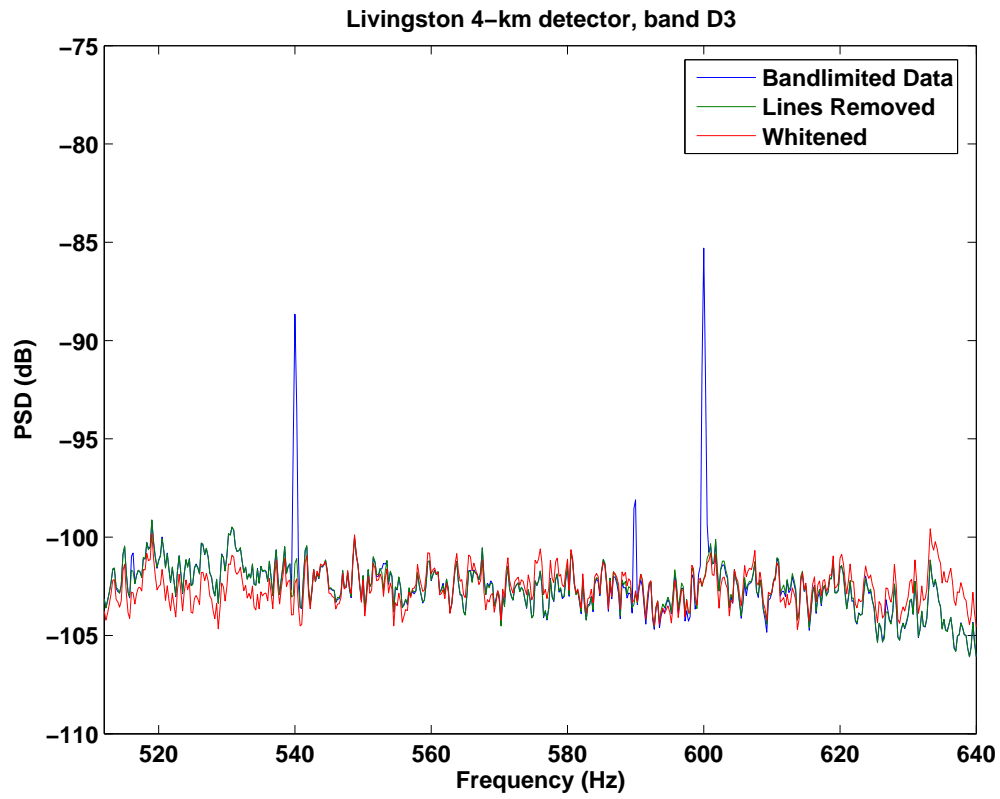
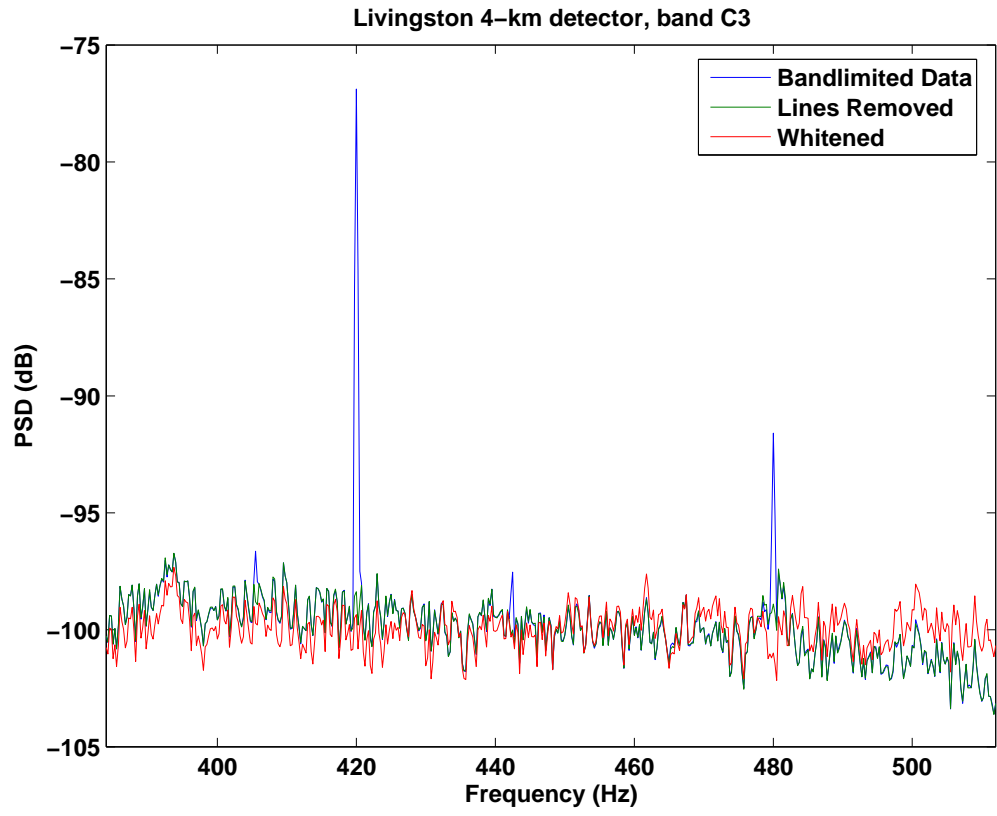


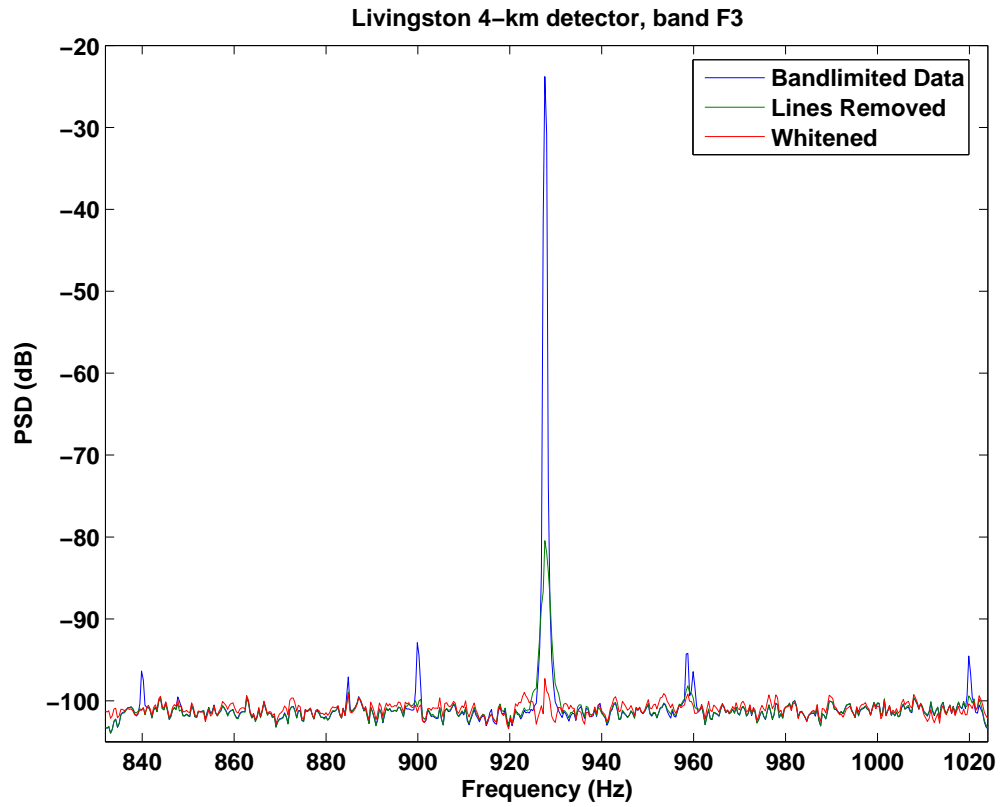
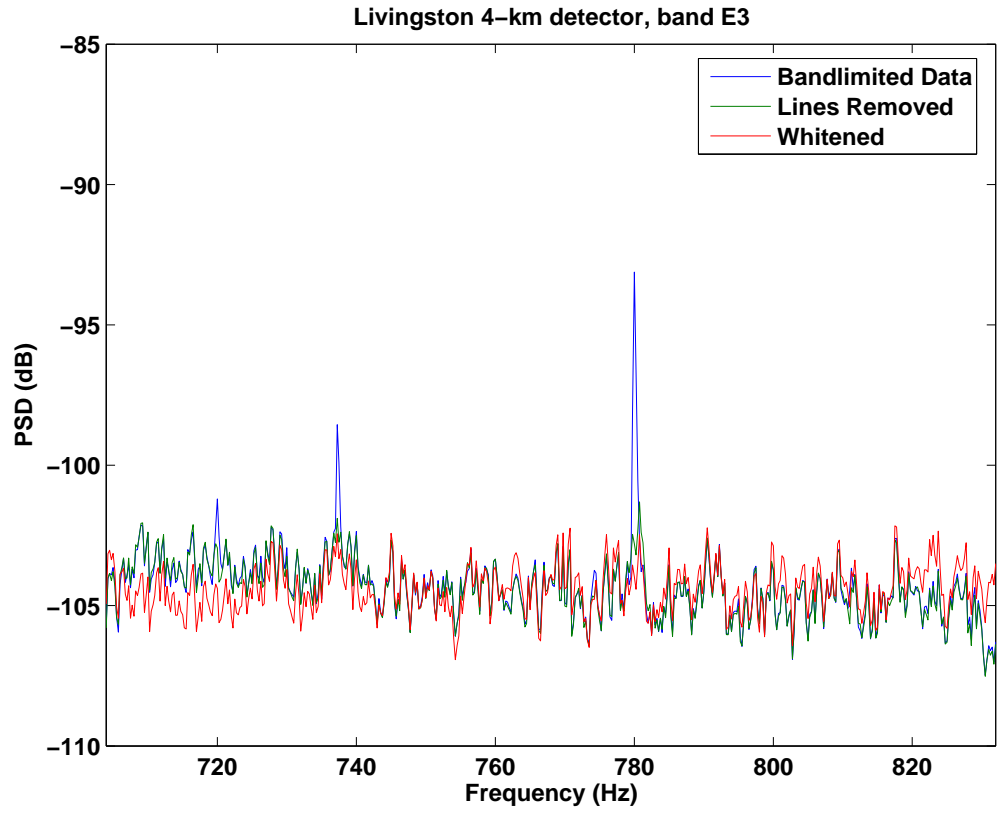


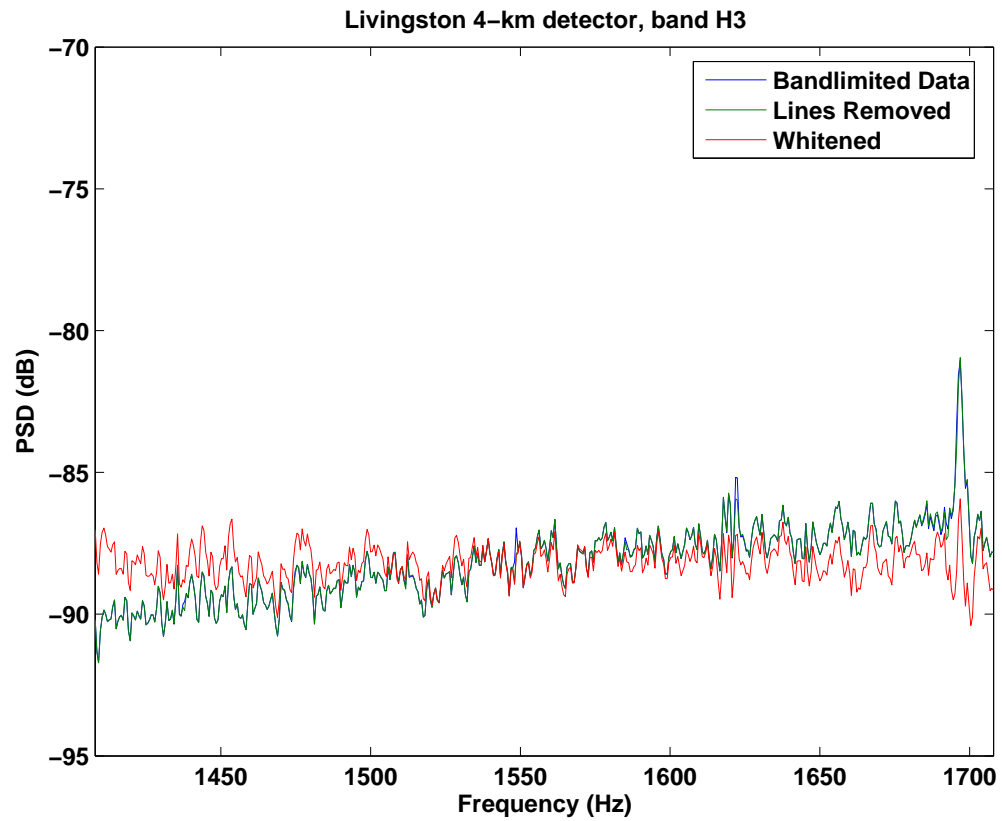
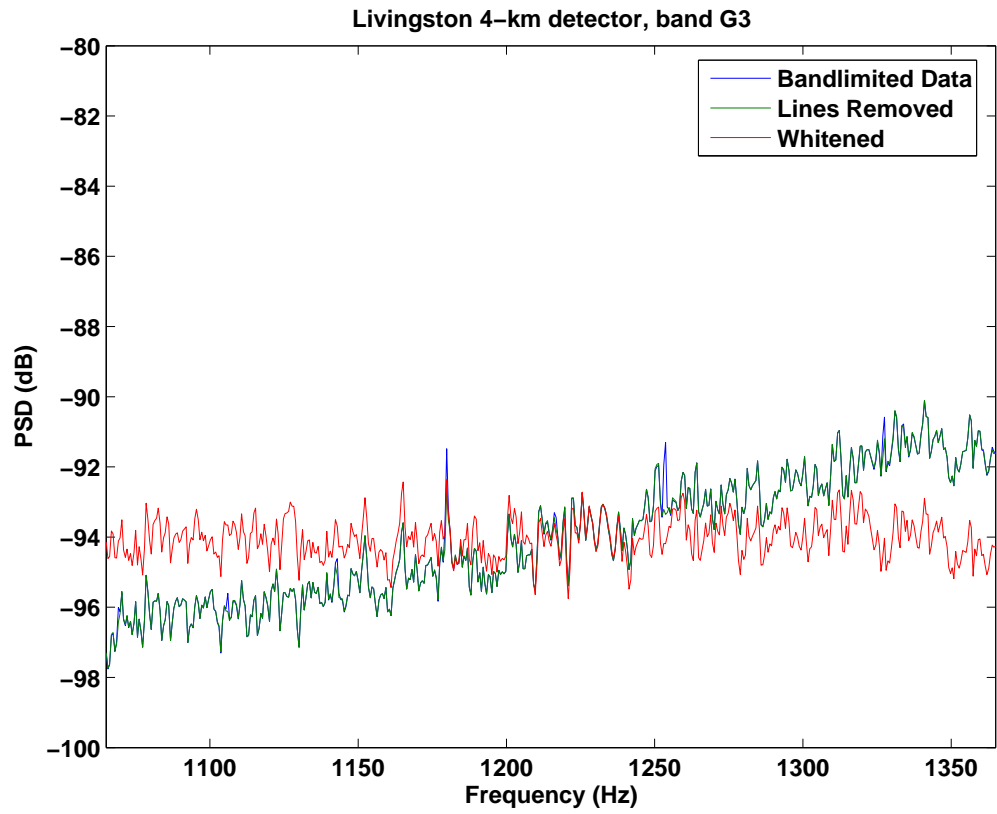


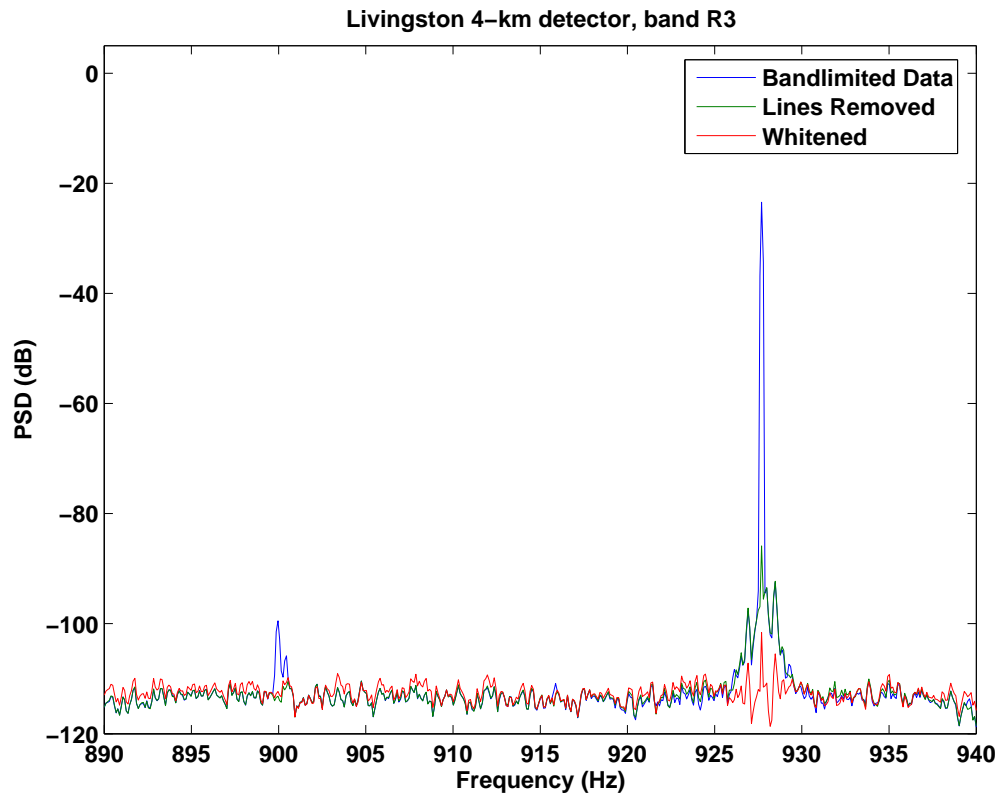
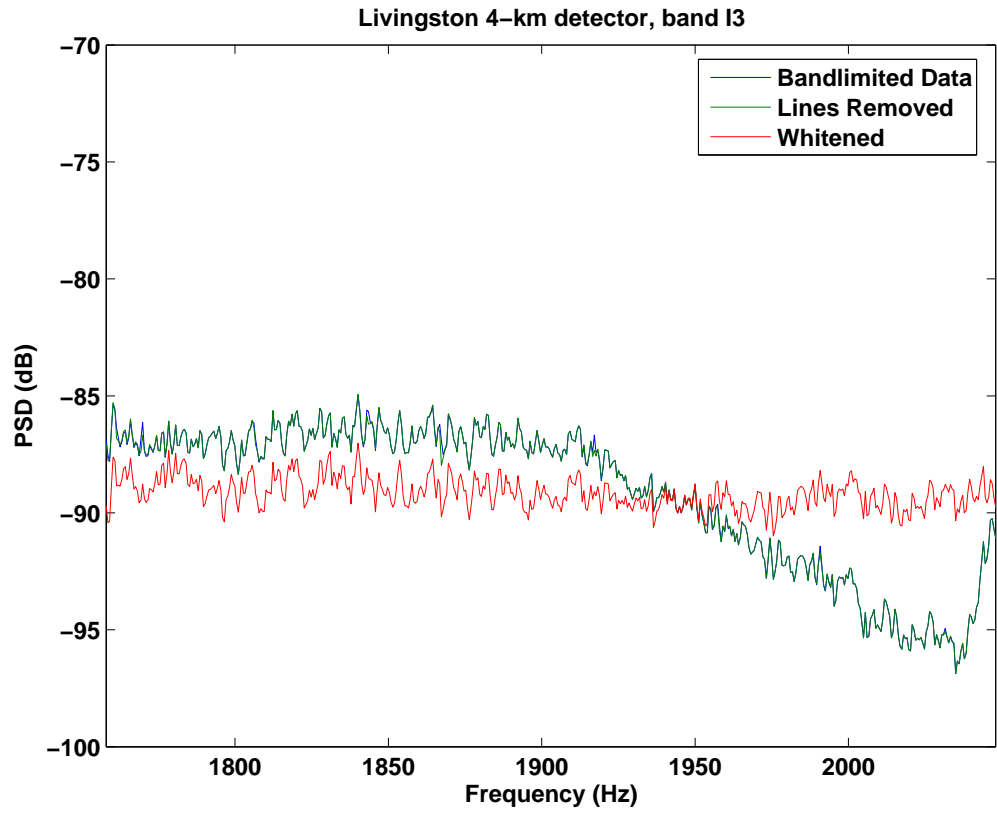














## E Rayleigh FOM Distributions

The Rayleigh figure-of-merit (FOM) is a measure of data conditioning performance. It is calculated via

$$\text{FOM} = \frac{\mathcal{R}_{PSD'}}{\mathcal{R}_{PSD}} = \frac{\sigma_{PSD'}/\mu_{PSD'}}{\sigma_{PSD}/\mu_{PSD}}, \quad (52)$$

where  $PSD$  is the power spectral density calculated on data that has only been treated with the shaping filter and then basebanded and  $PSD'$  is data that has undergone the entire data conditioning process. A  $\text{FOM} < 1$  indicates that the line removal and whitening portions of data conditioning have been successful.

Below are plots showing the FOM per segment and distributions of FOM values for each frequency band (frequency band definitions are given in Appendix B) and detector over all non-overlapping 300 second long S3 segments. For most of these segments, the FOM was less than one, indicating that data conditioning was successful.

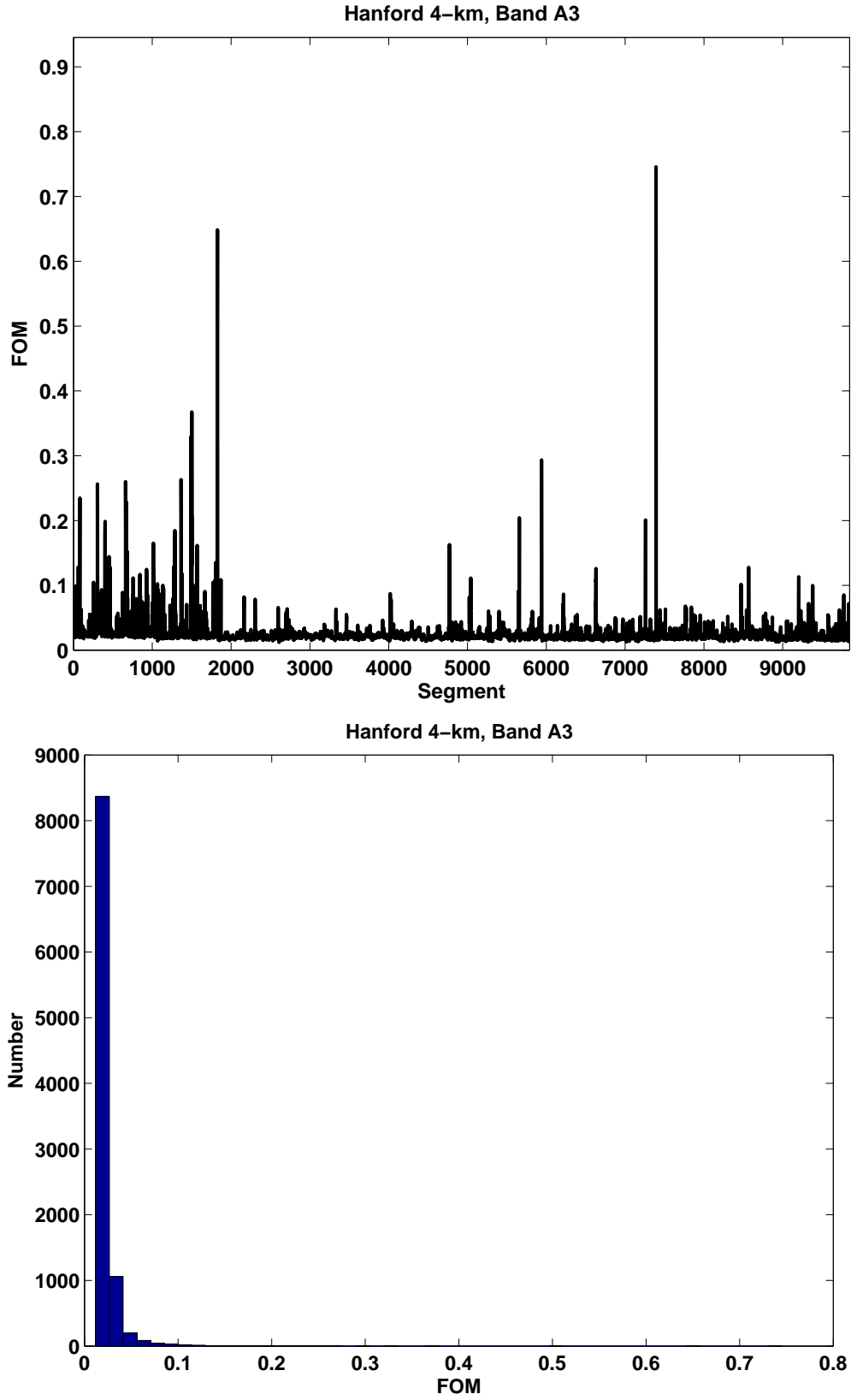


Figure 15: Data conditioning FOM for Hanford 4-km detector, band A3

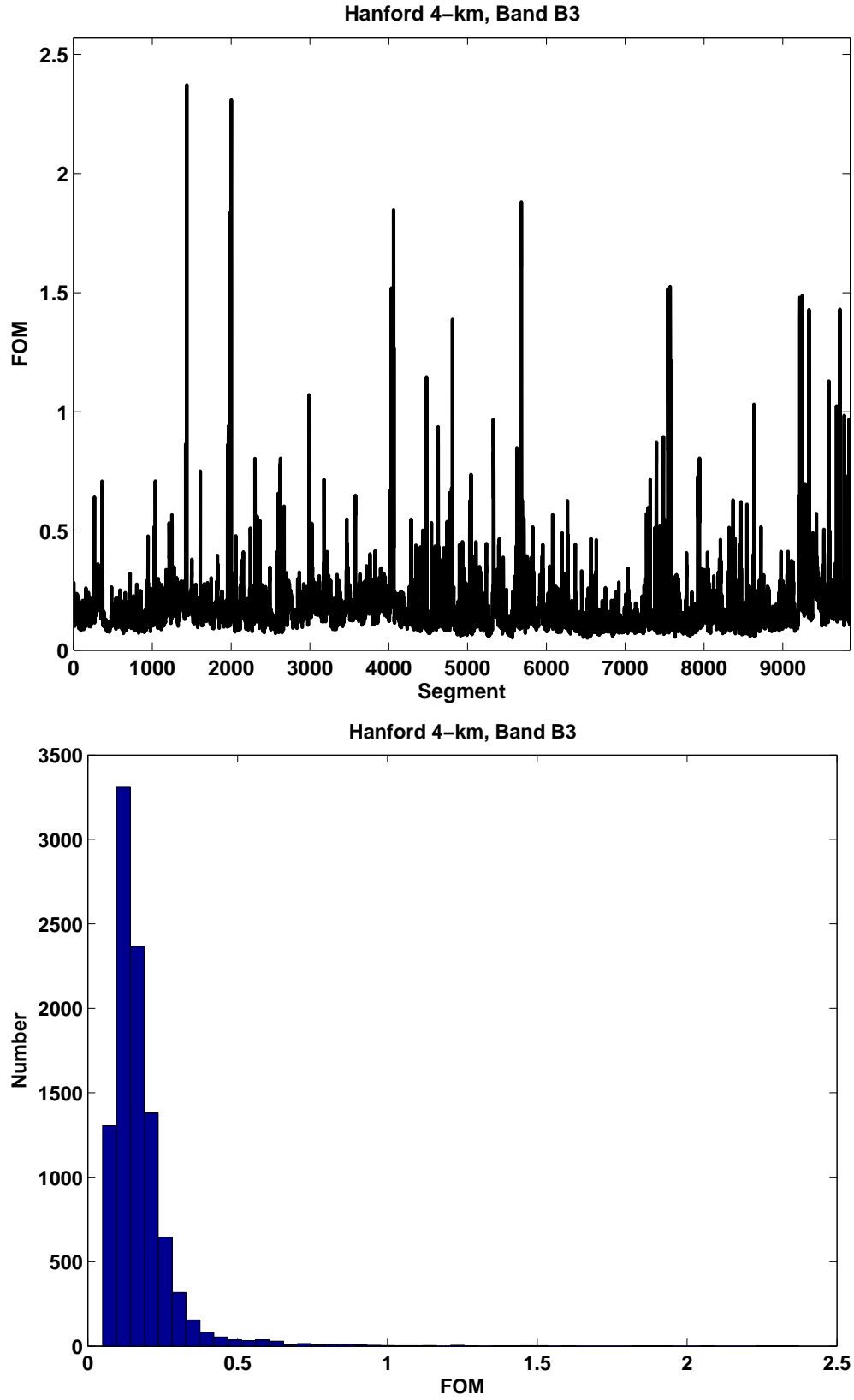


Figure 16: Data conditioning FOM for Hanford 4-km detector, band B3

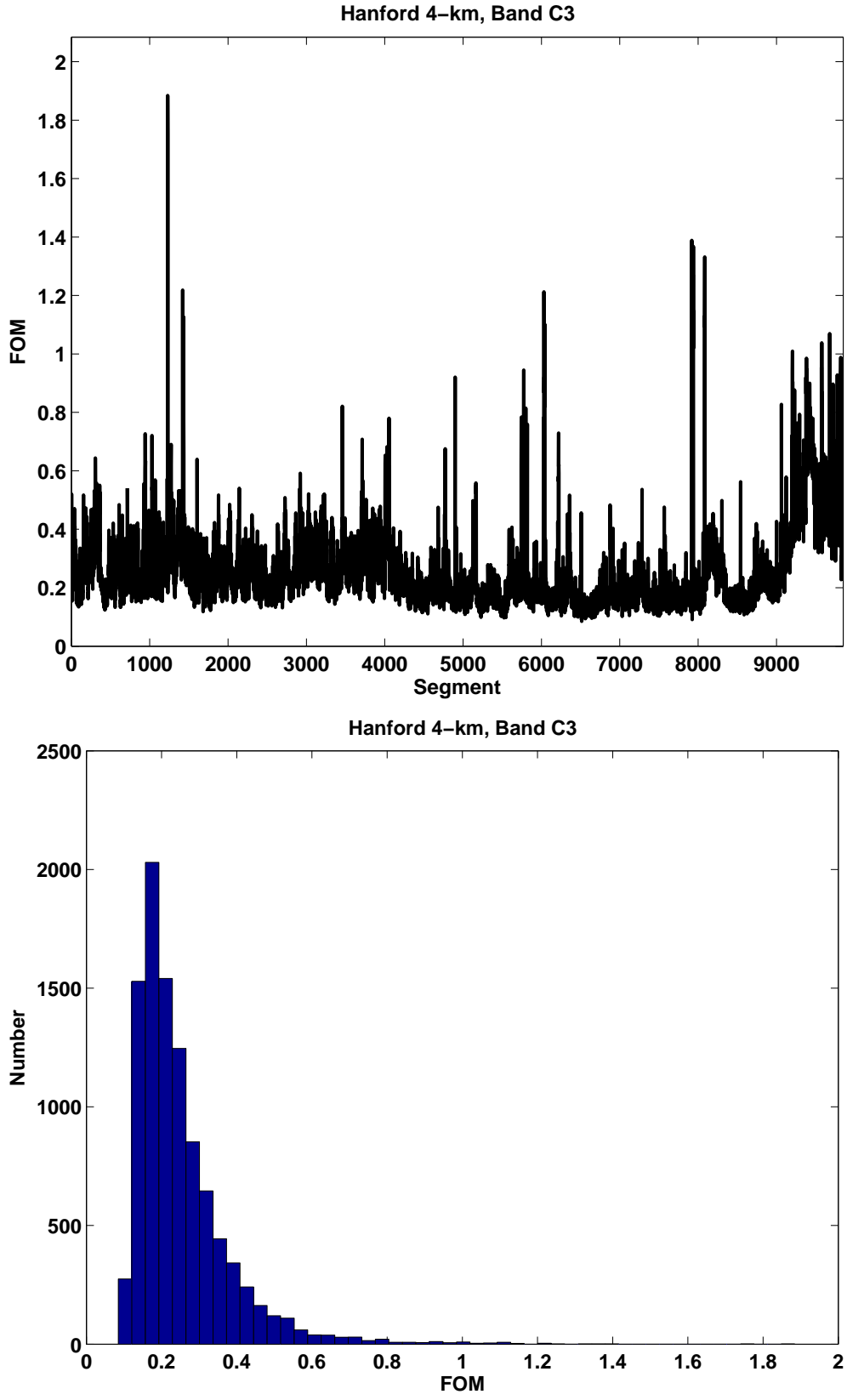


Figure 17: Data conditioning FOM for Hanford 4-km detector, band C3

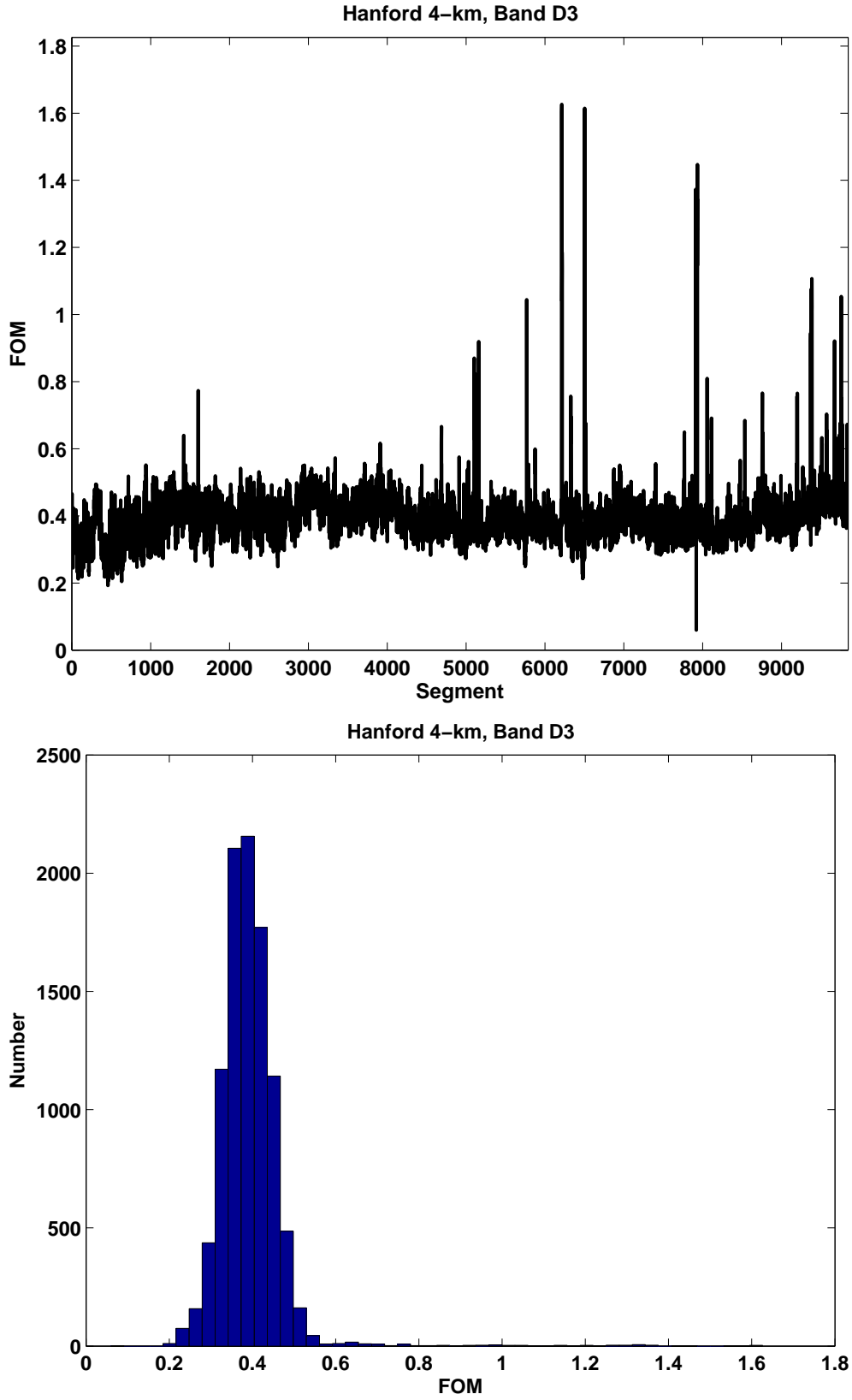


Figure 18: Data conditioning FOM for Hanford 4-km detector, band D3

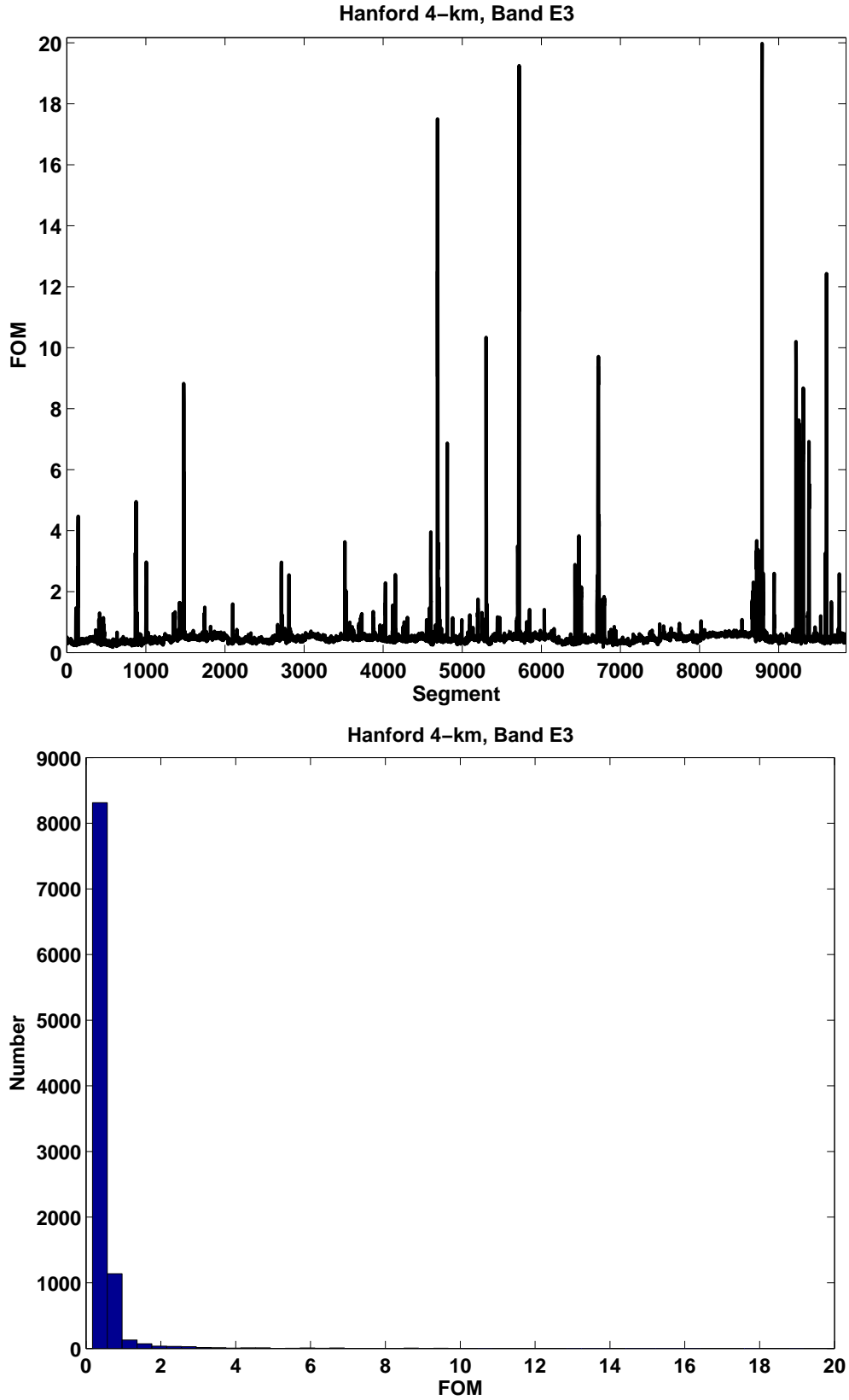


Figure 19: Data conditioning FOM for Hanford 4-km detector, band E3

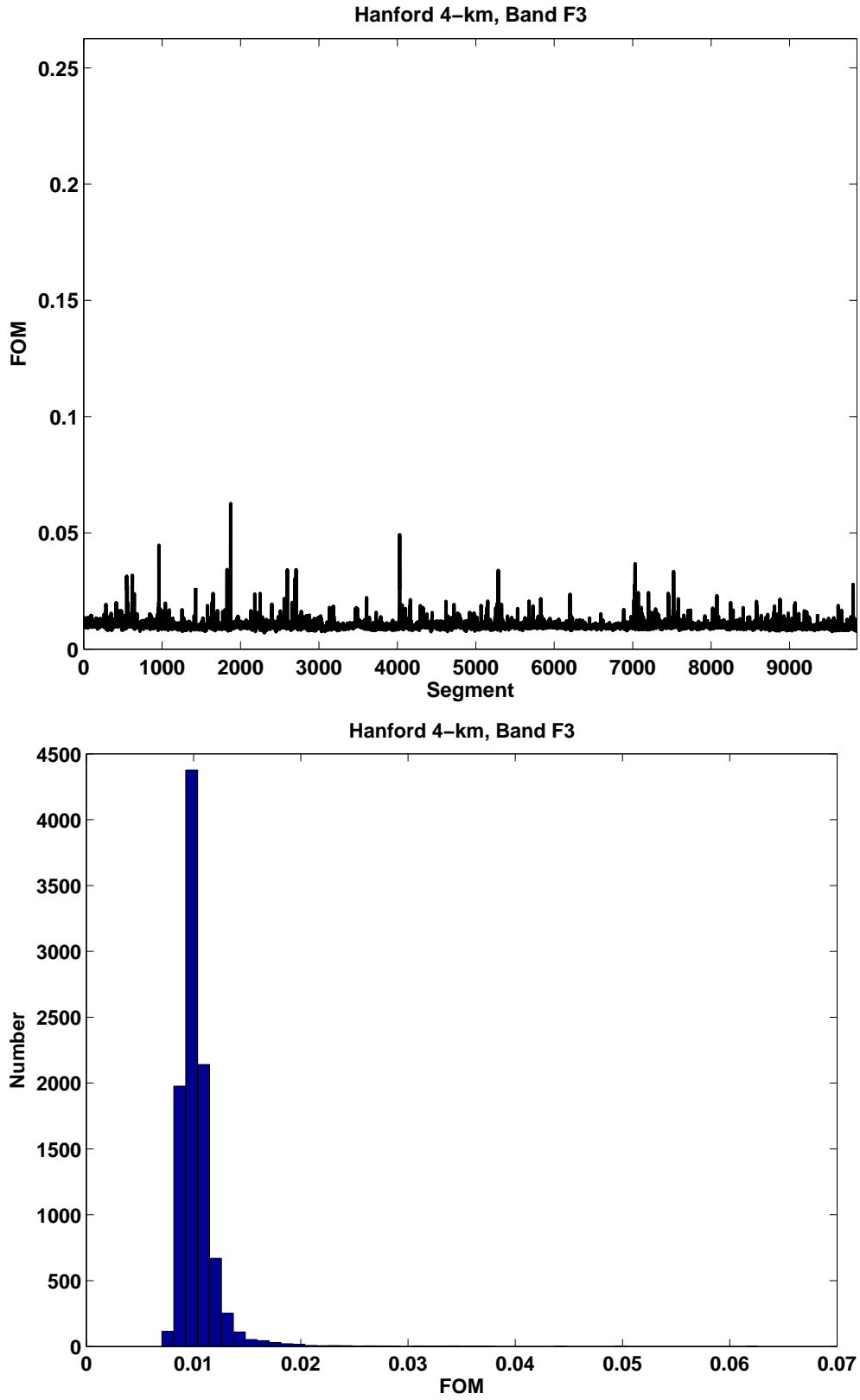


Figure 20: Data conditioning FOM for Hanford 4-km detector, band F3

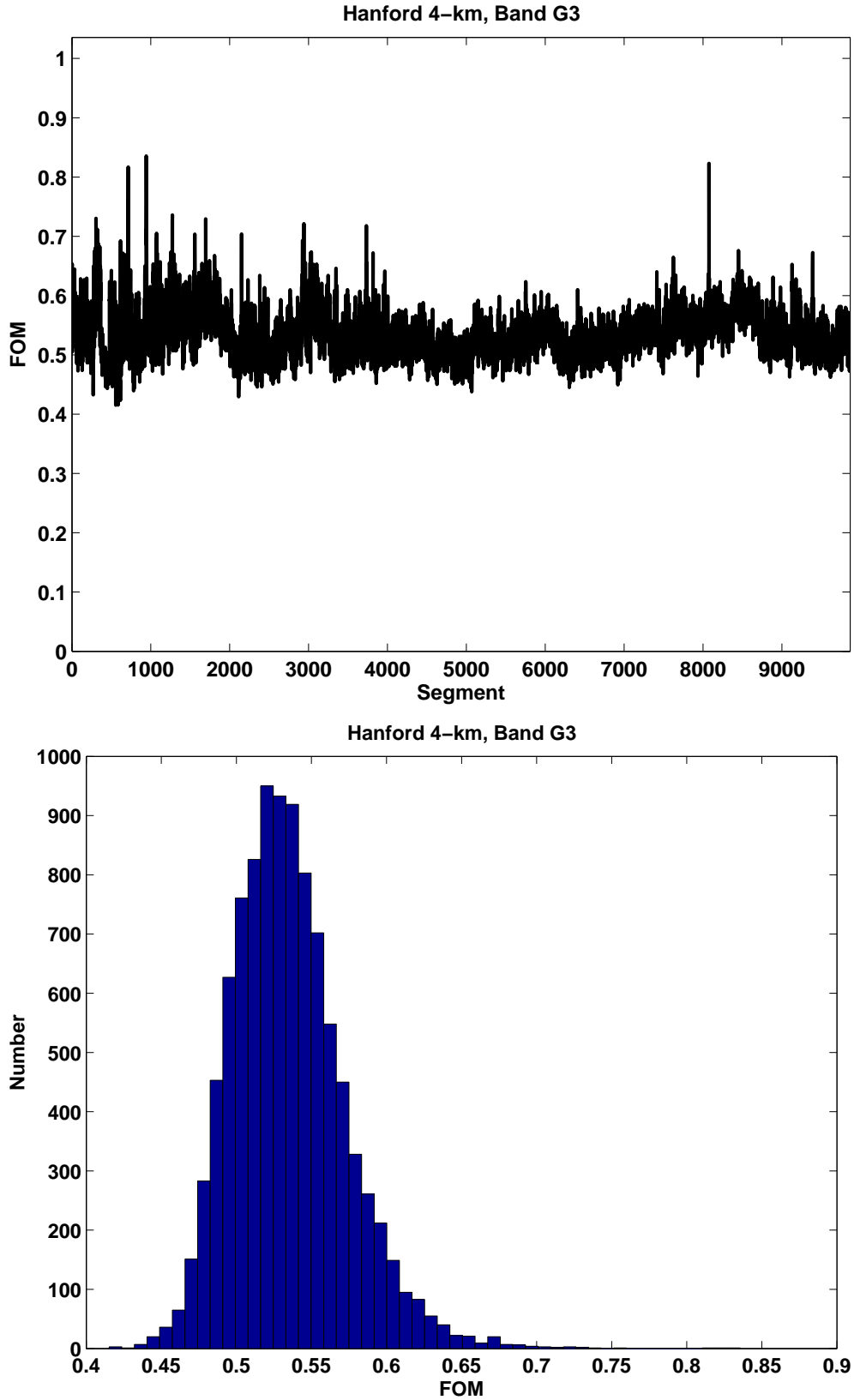


Figure 21: Data conditioning FOM for Hanford 4-km detector, band G3



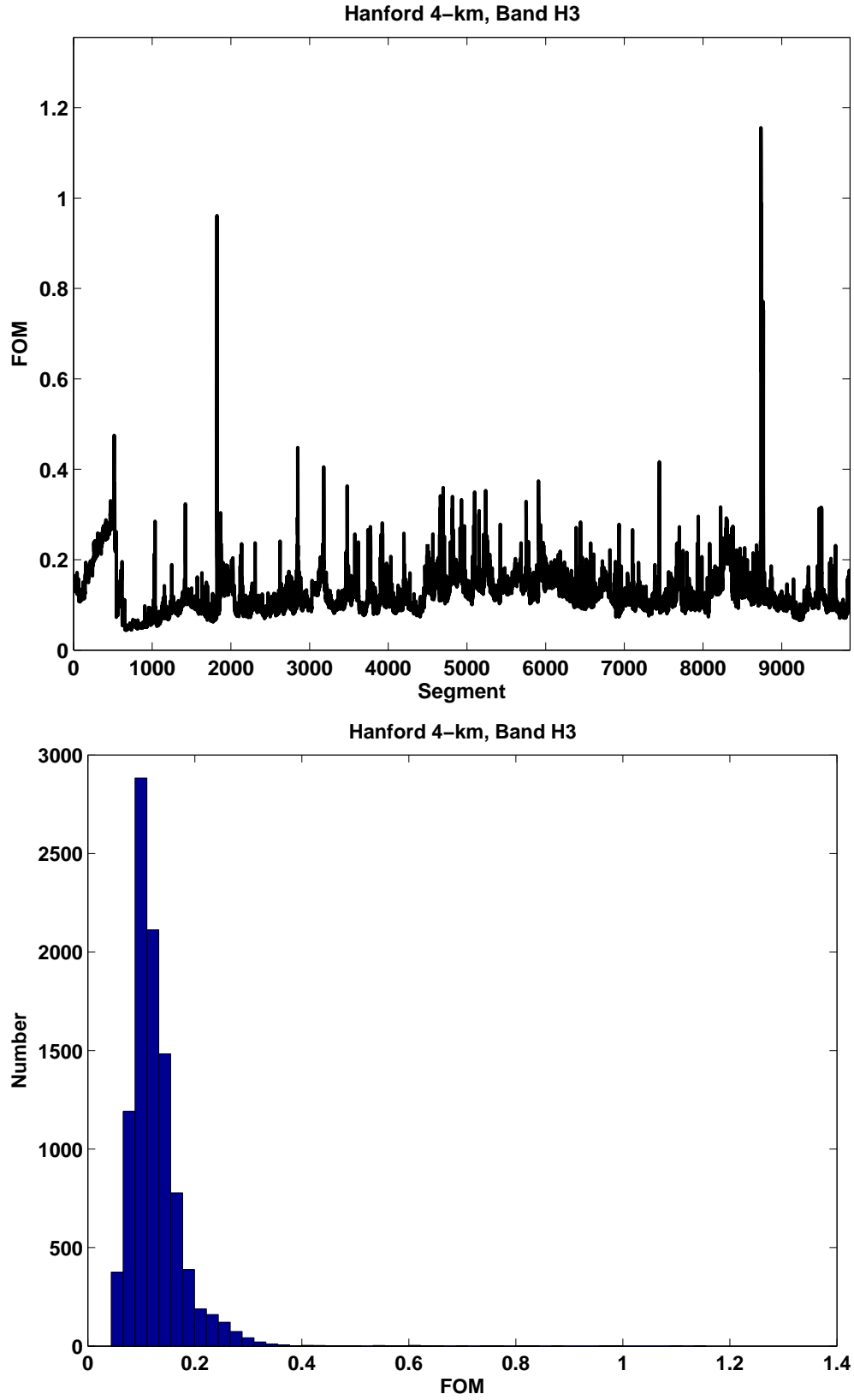


Figure 22: Data conditioning FOM for Hanford 4-km detector, band H3

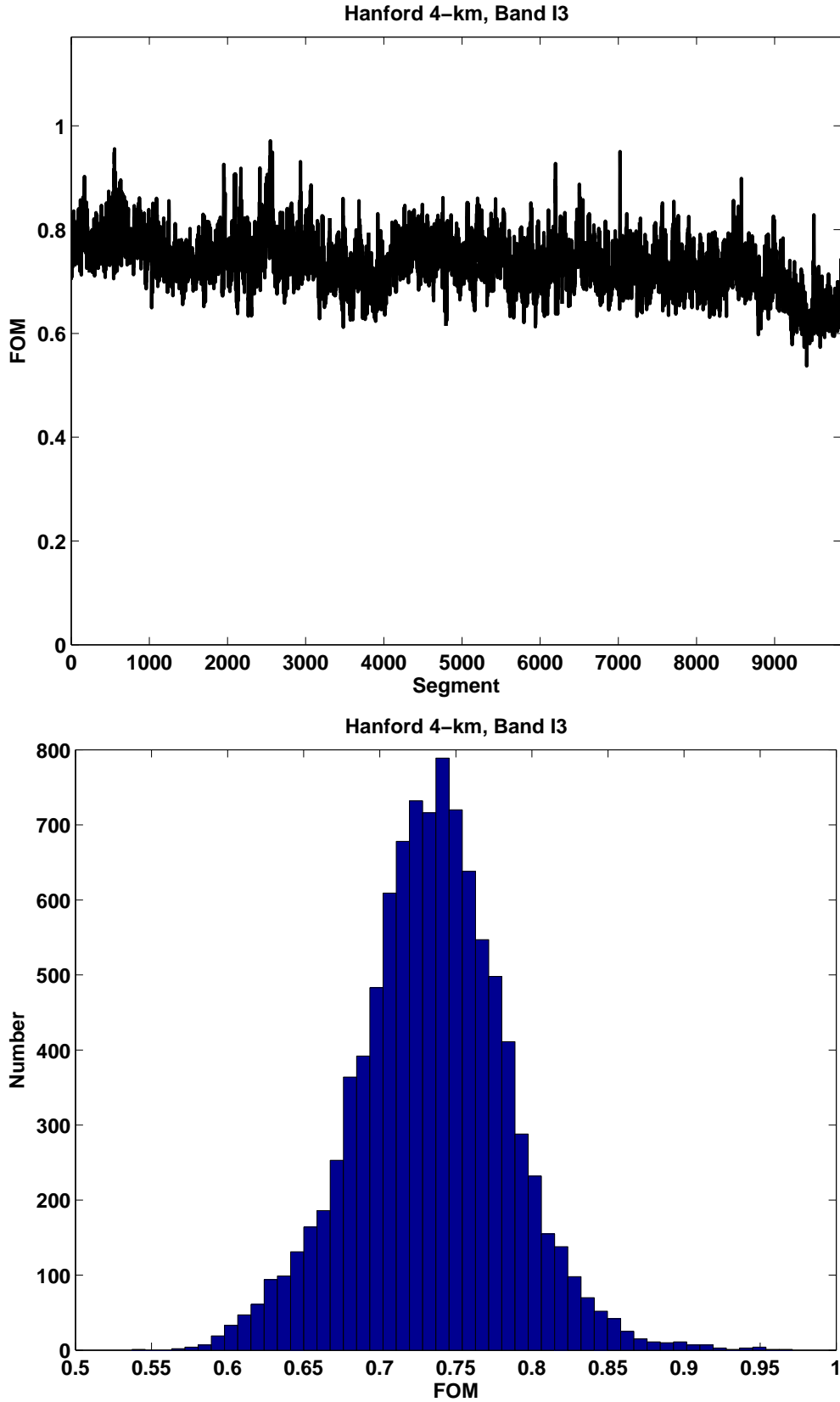


Figure 23: Data conditioning FOM for Hanford 4-km detector, band I3

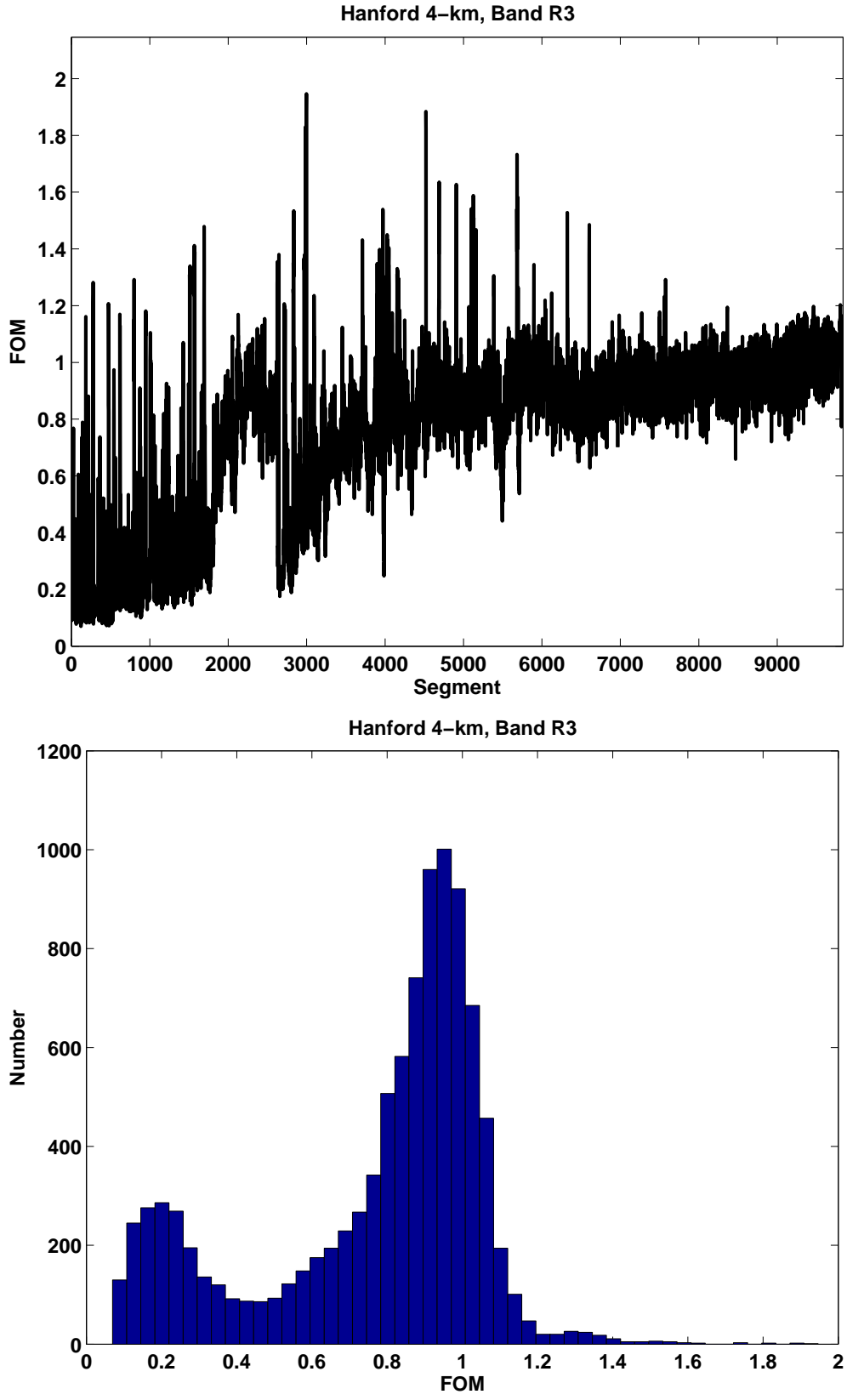


Figure 24: Data conditioning FOM for Hanford 4-km detector, band R3

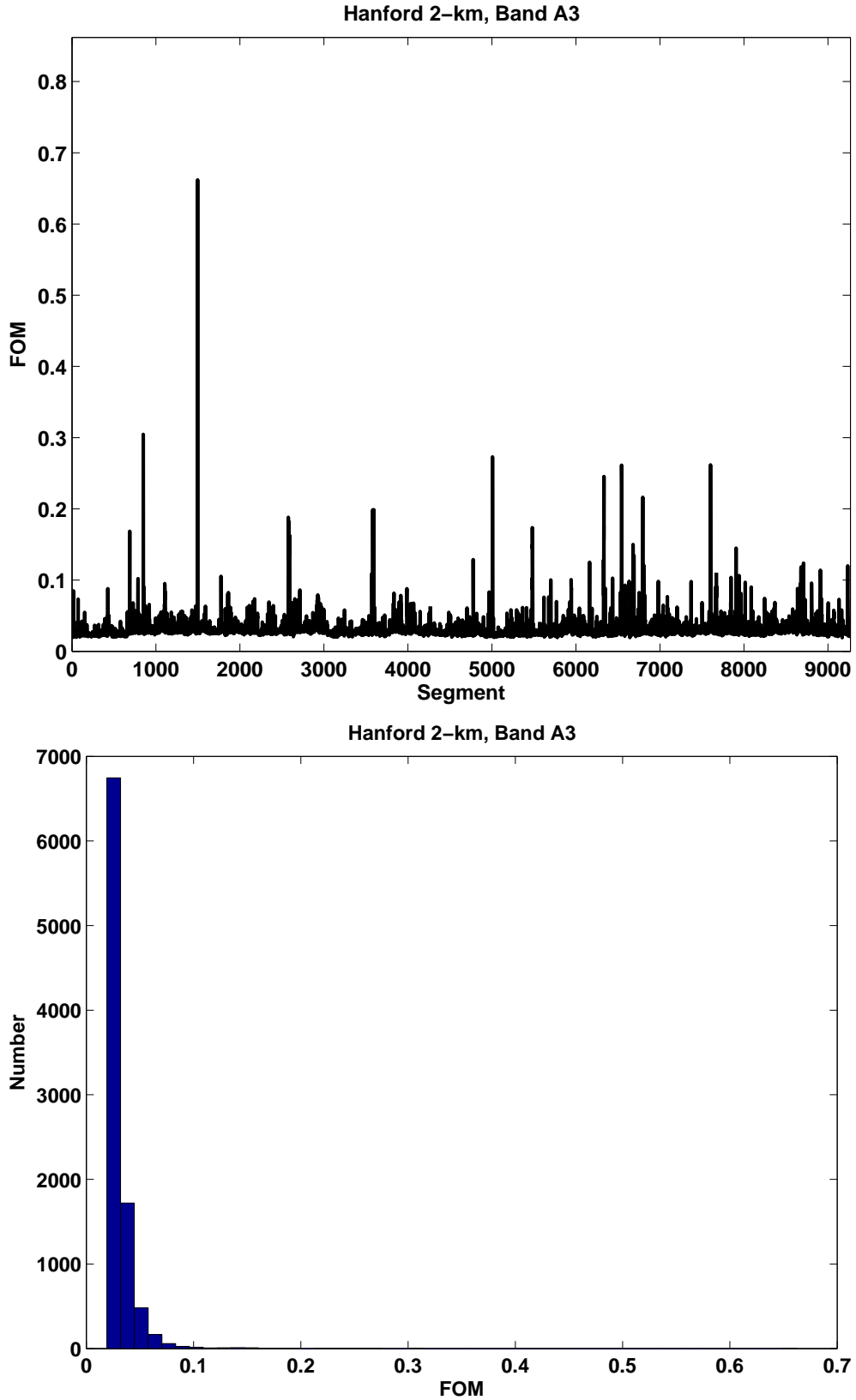


Figure 25: Data conditioning FOM for Hanford 2-km detector, band A3

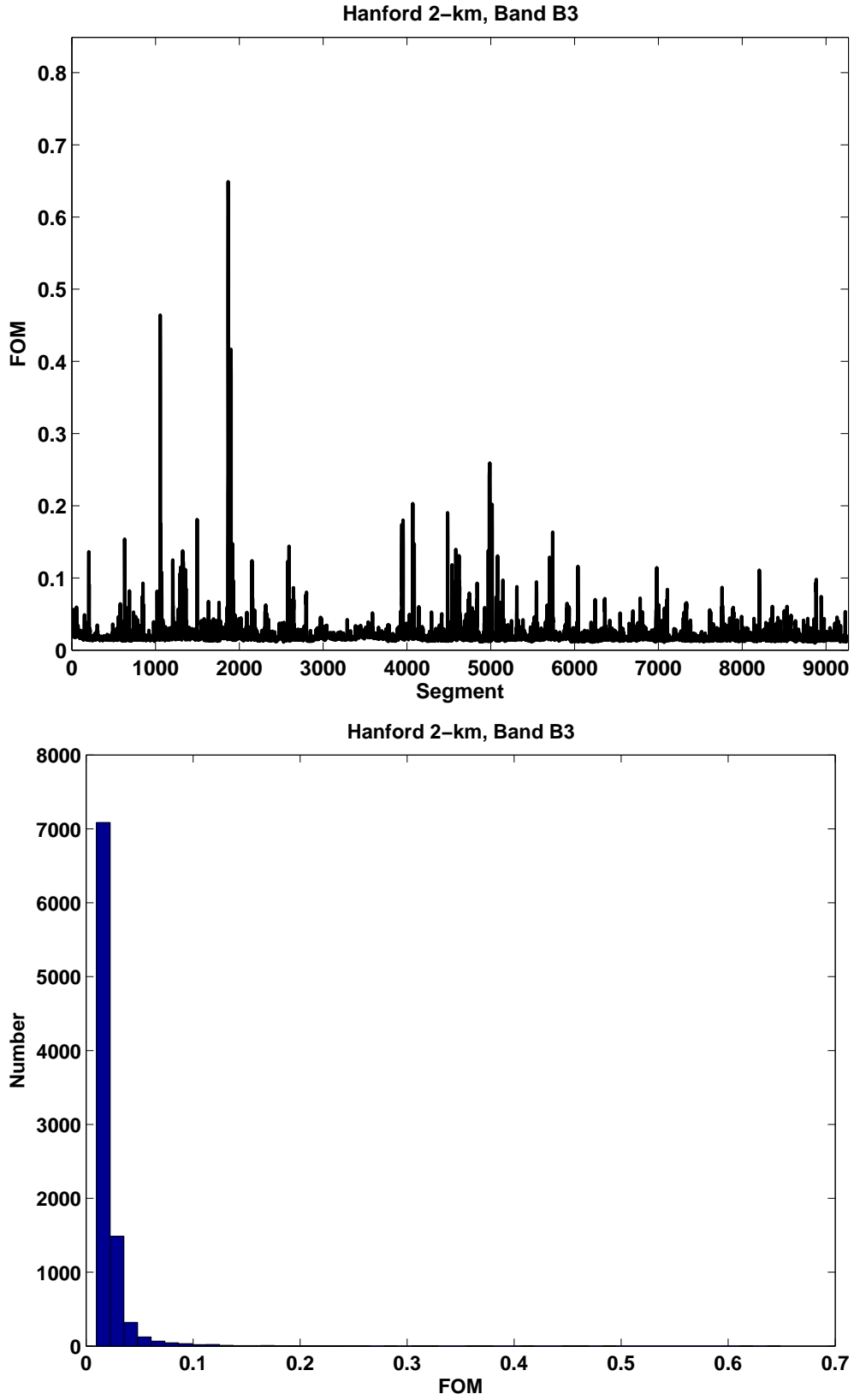


Figure 26: Data conditioning FOM for Hanford 2-km detector, band B3

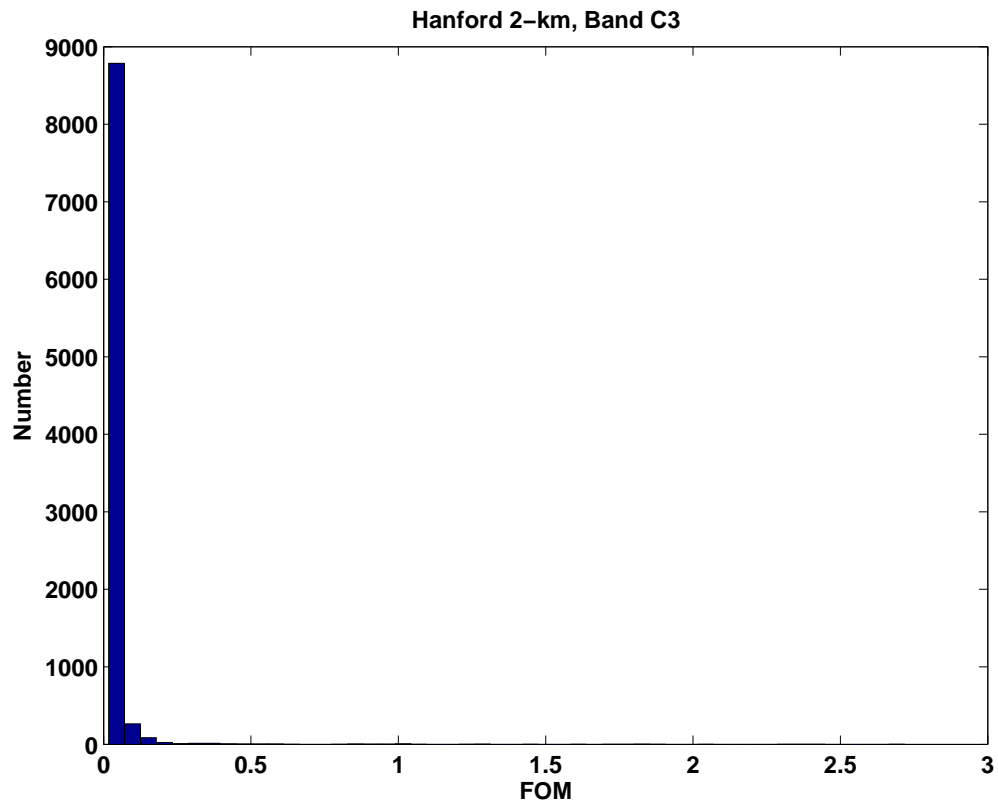
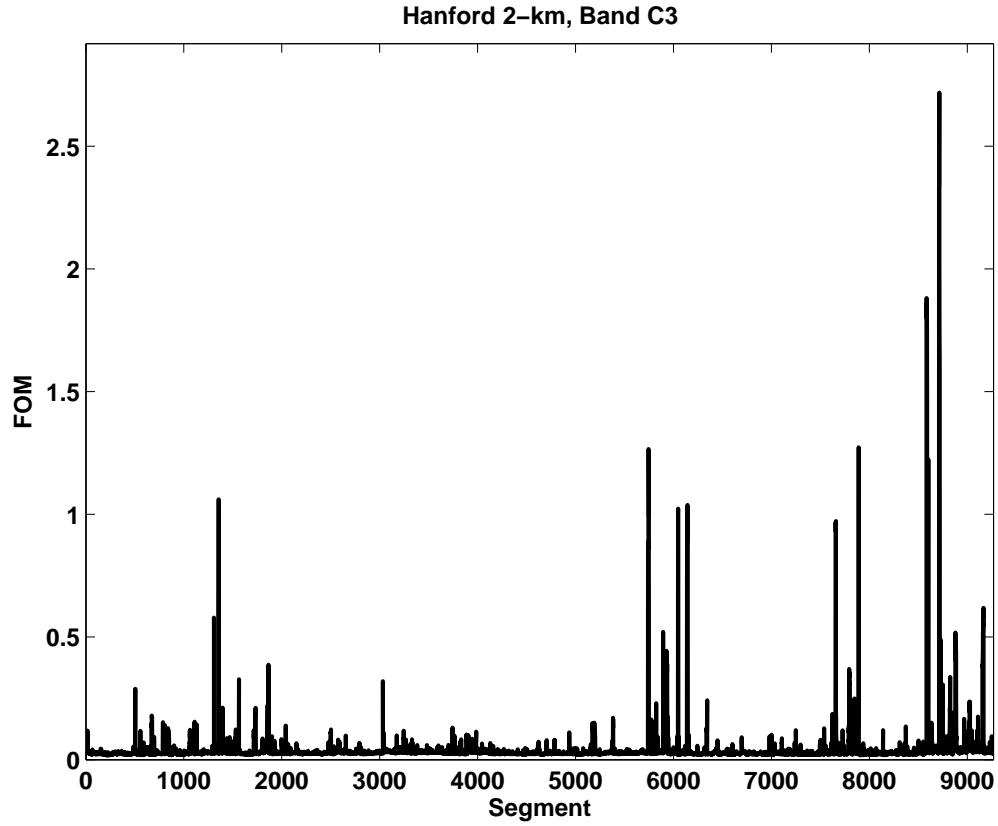


Figure 27: Data conditioning FOM for Hanford 2-km detector, band C3

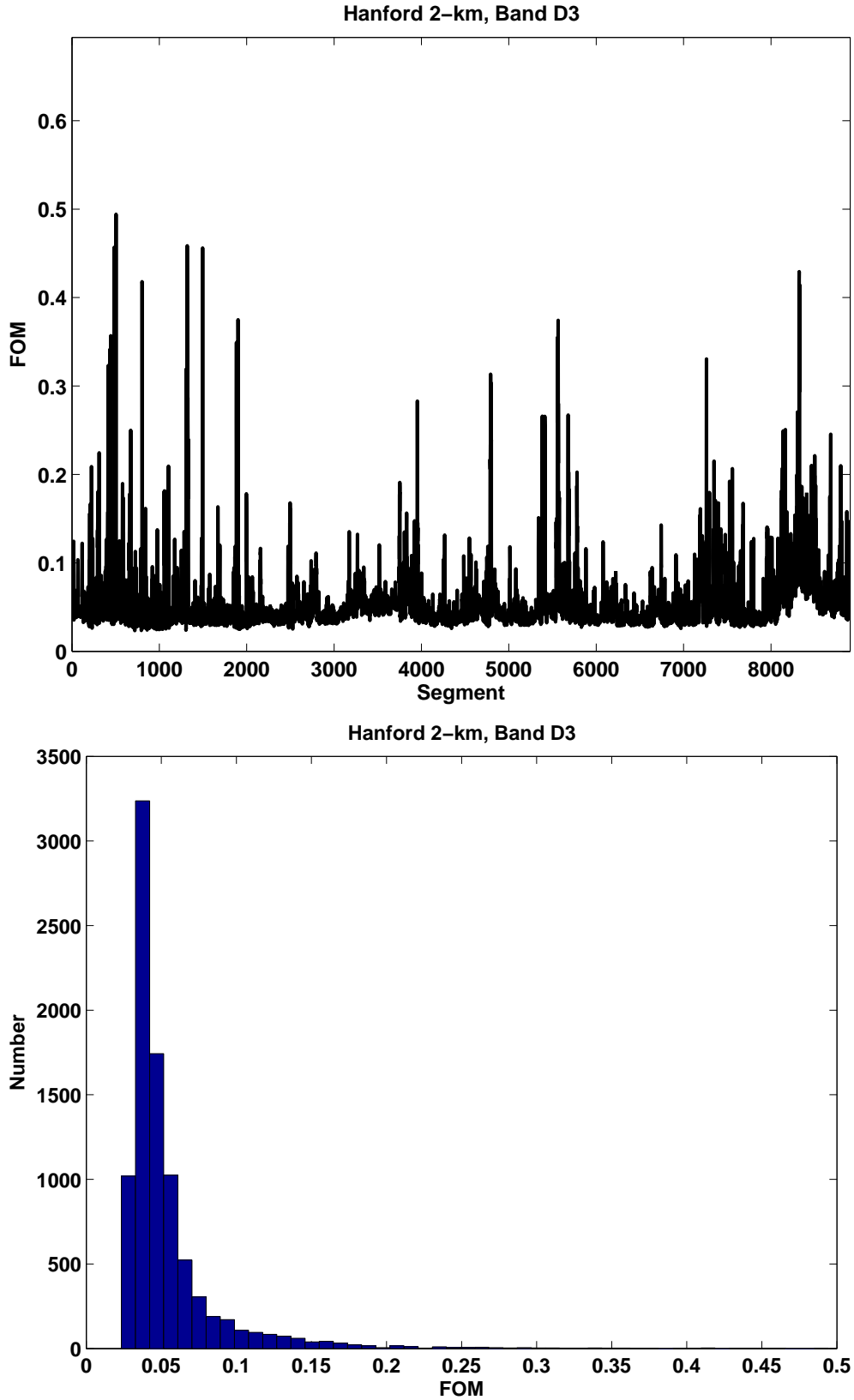


Figure 28: Data conditioning FOM for Hanford 2-km detector, band D3

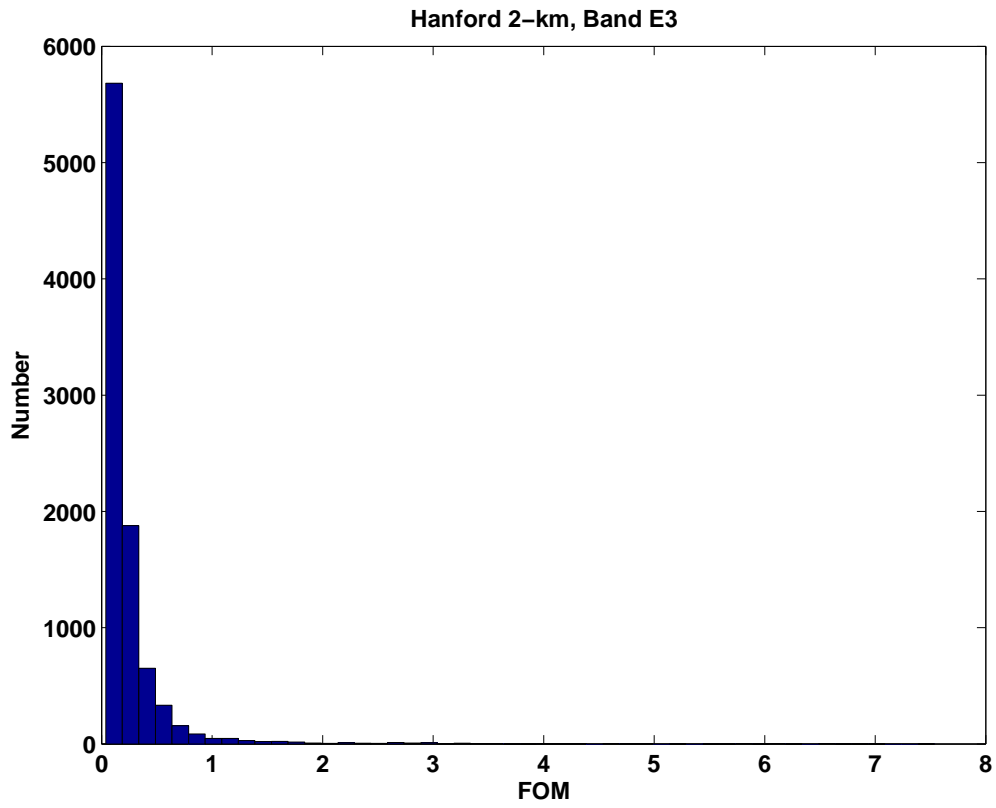
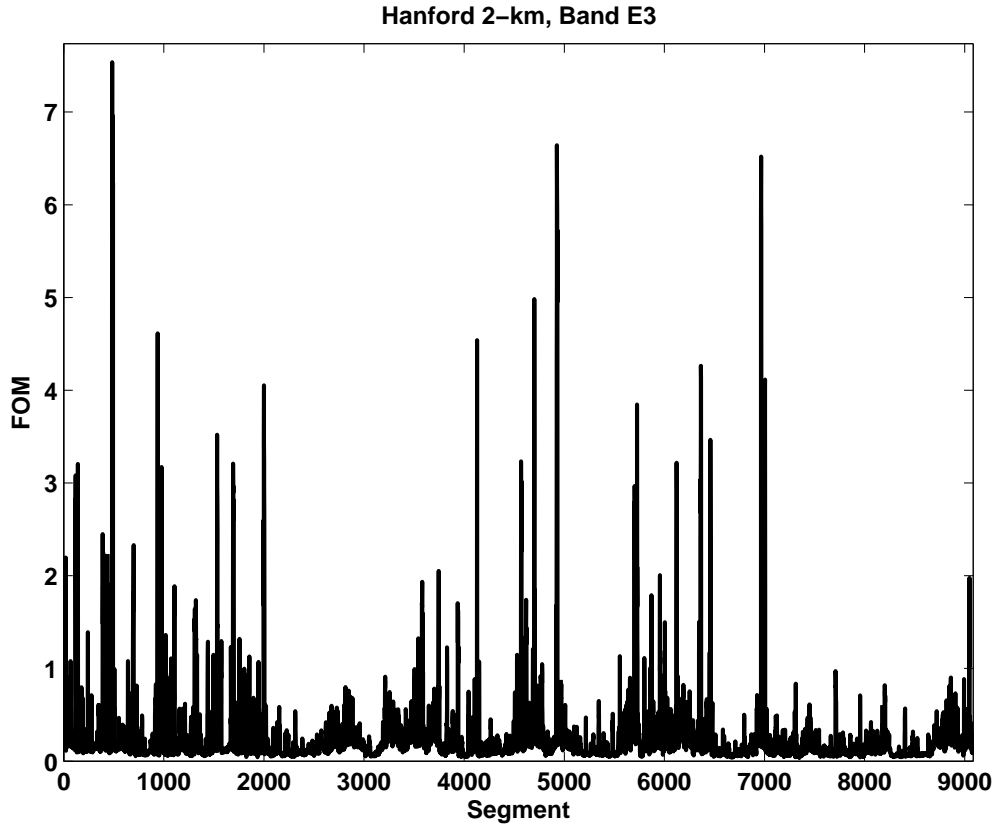


Figure 29: Data conditioning FOM for Hanford 2-km detector, band E3



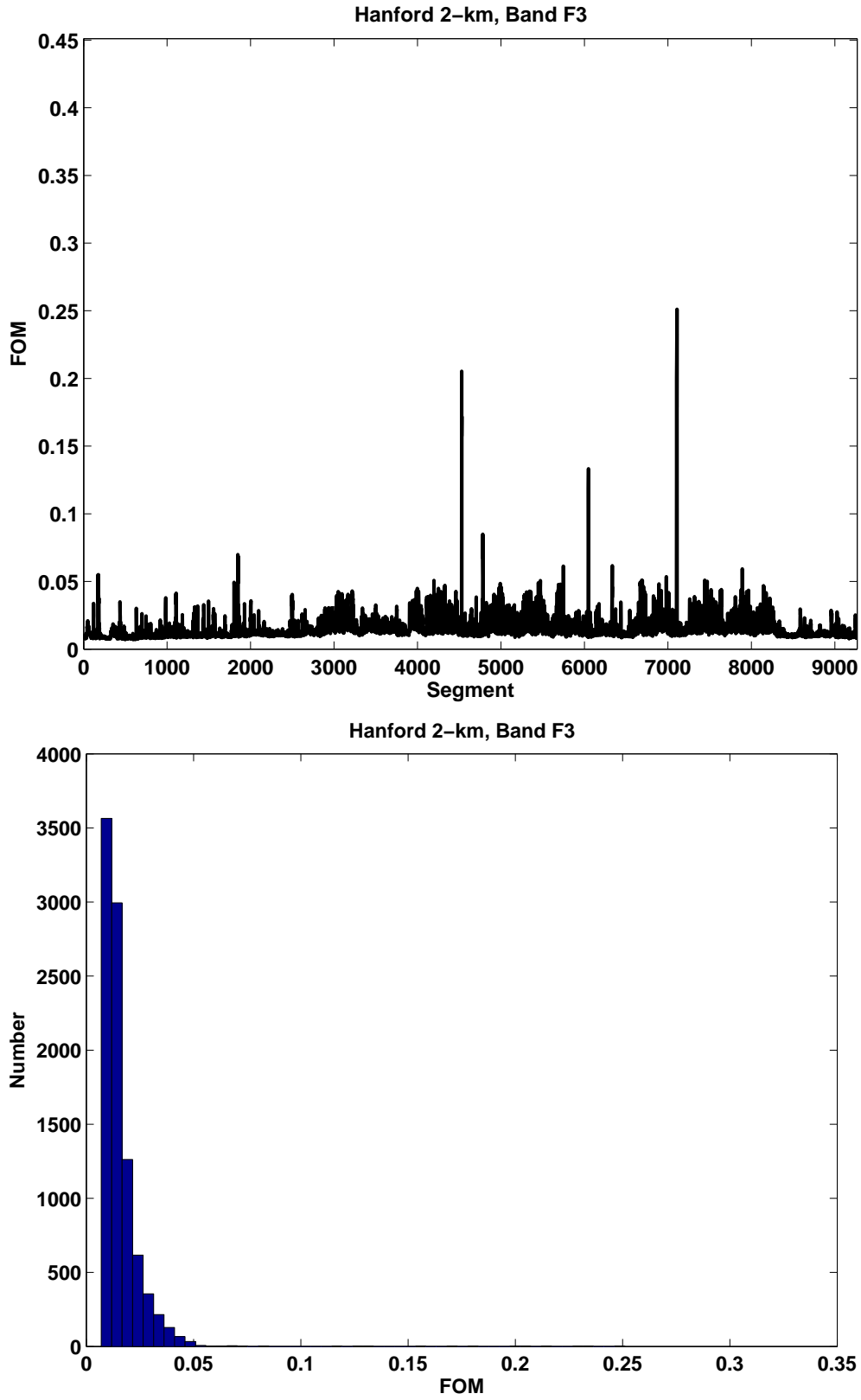


Figure 30: Data conditioning FOM for Hanford 2-km detector, band F3

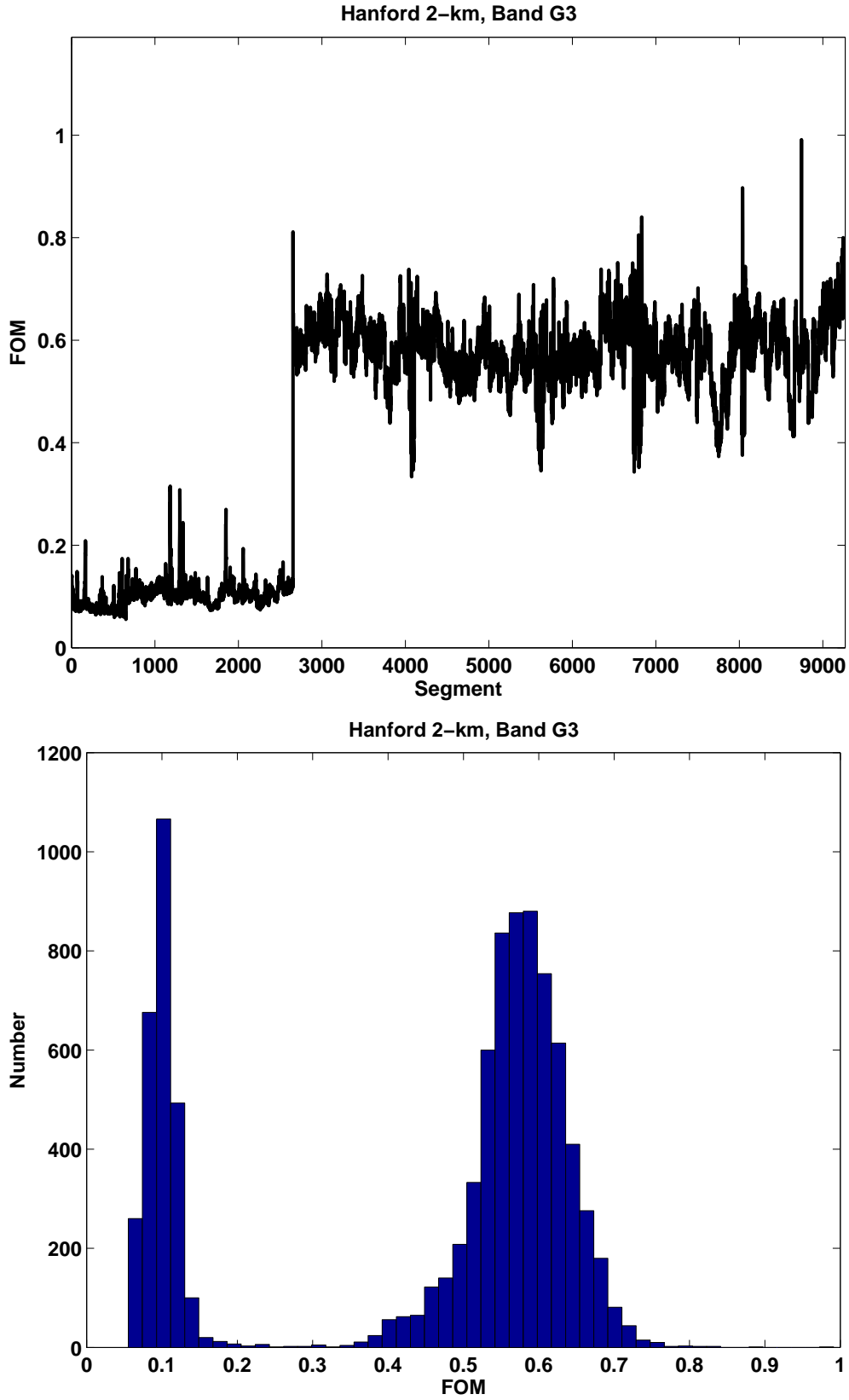


Figure 31: Data conditioning FOM for Hanford 2-km detector, band G3

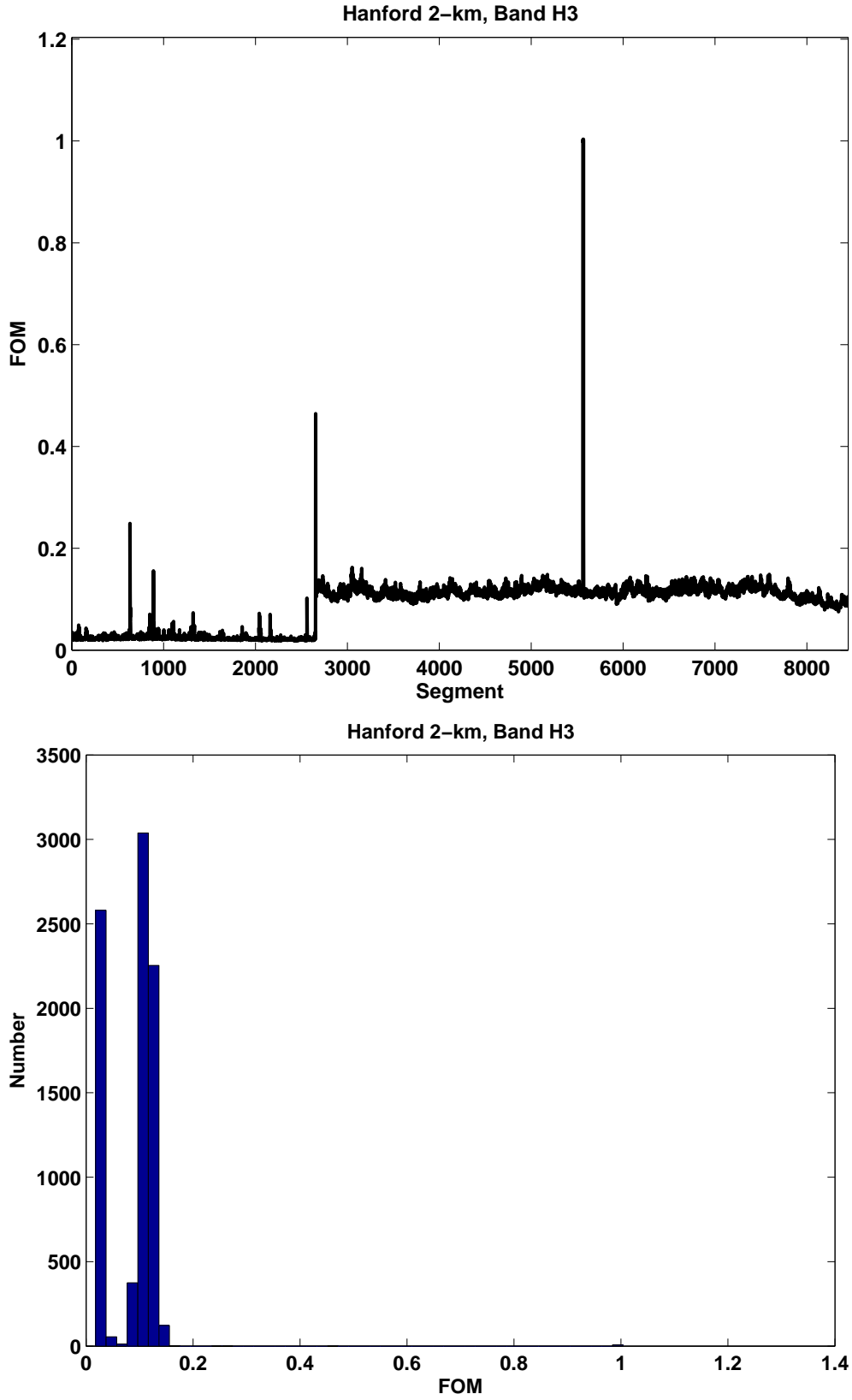


Figure 32: Data conditioning FOM for Hanford 2-km detector, band H3

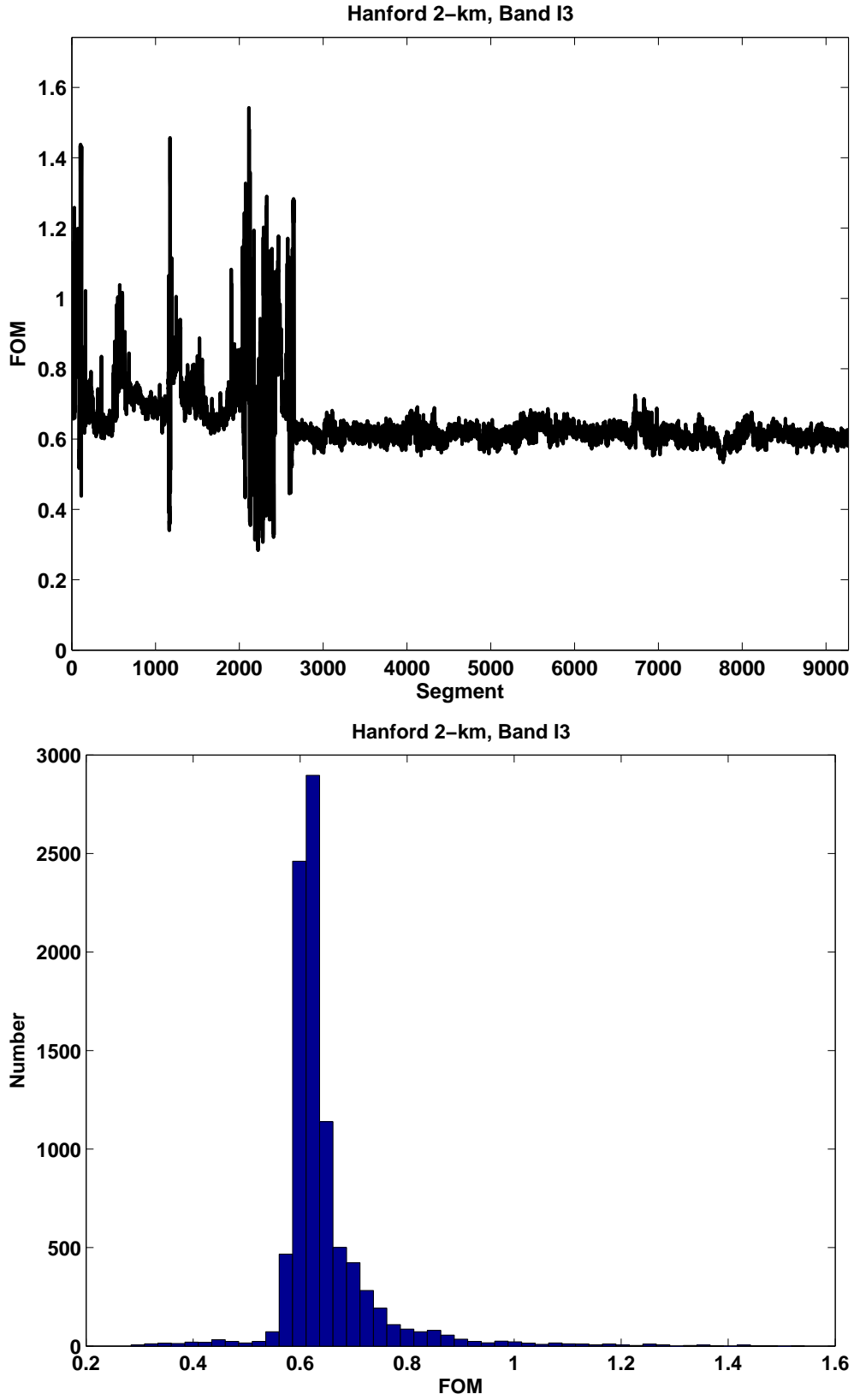


Figure 33: Data conditioning FOM for Hanford 2-km detector, band I3

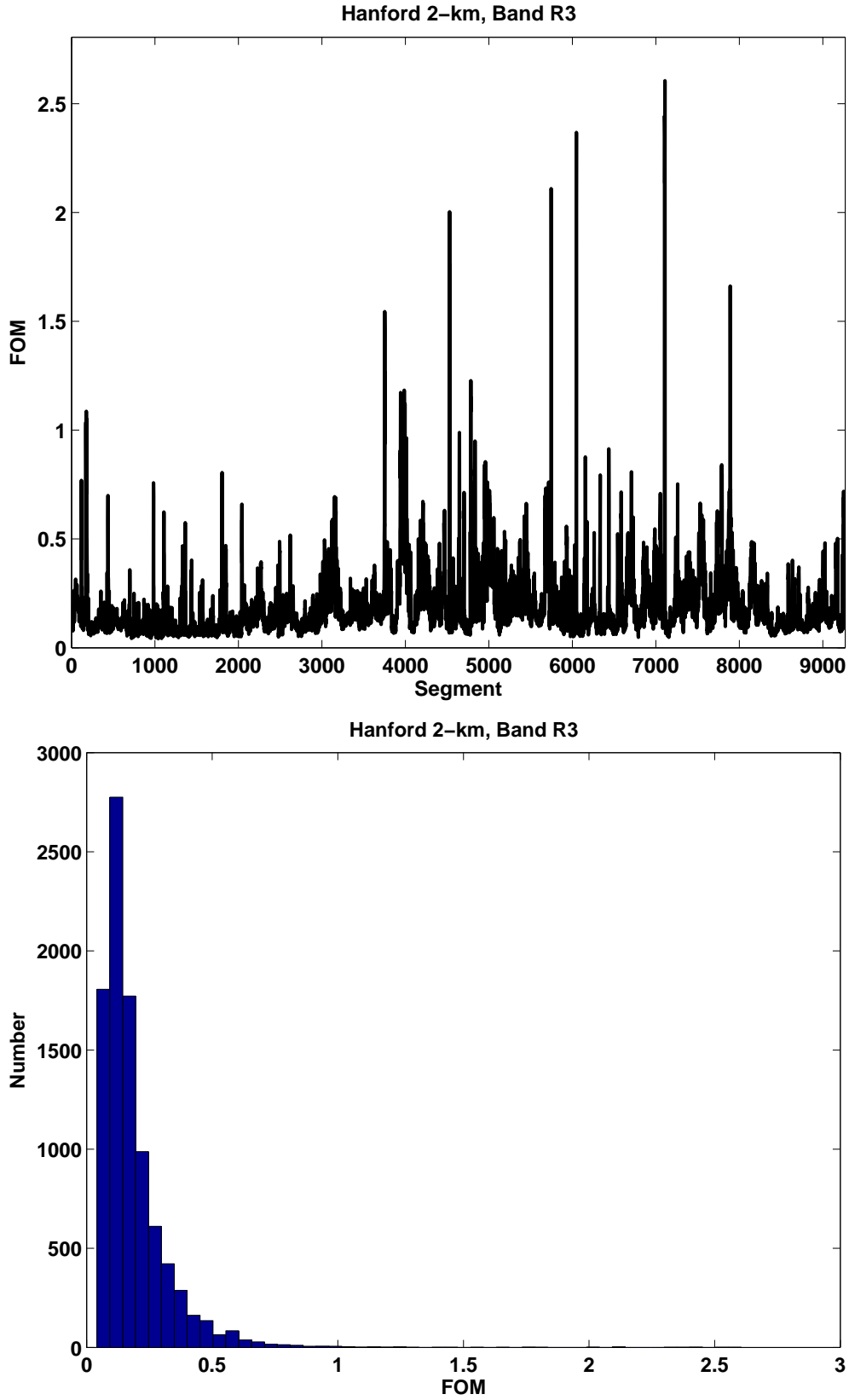


Figure 34: Data conditioning FOM for Hanford 2-km detector, band R3

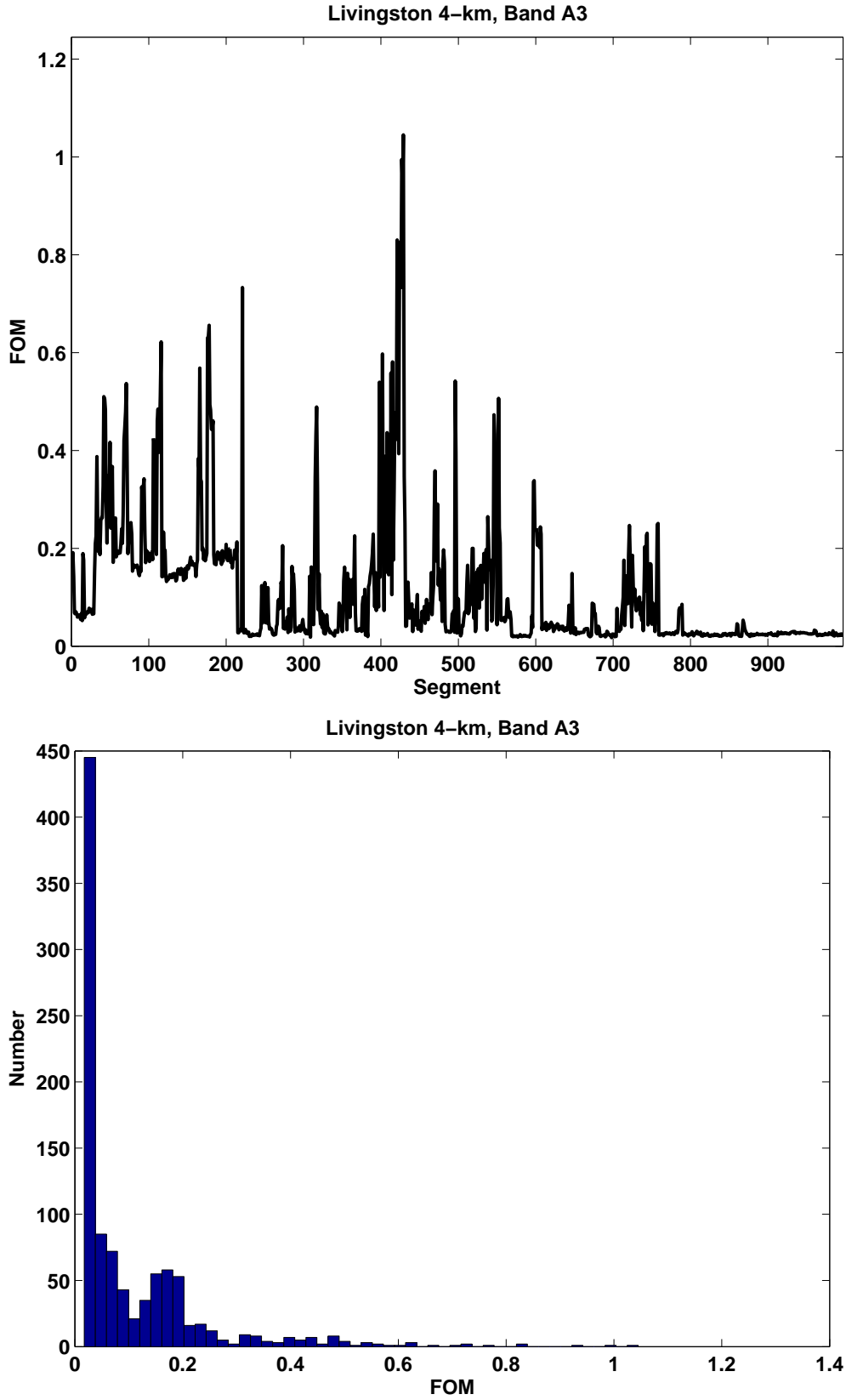


Figure 35: Data conditioning FOM for Livingston 4-km detector, band A3

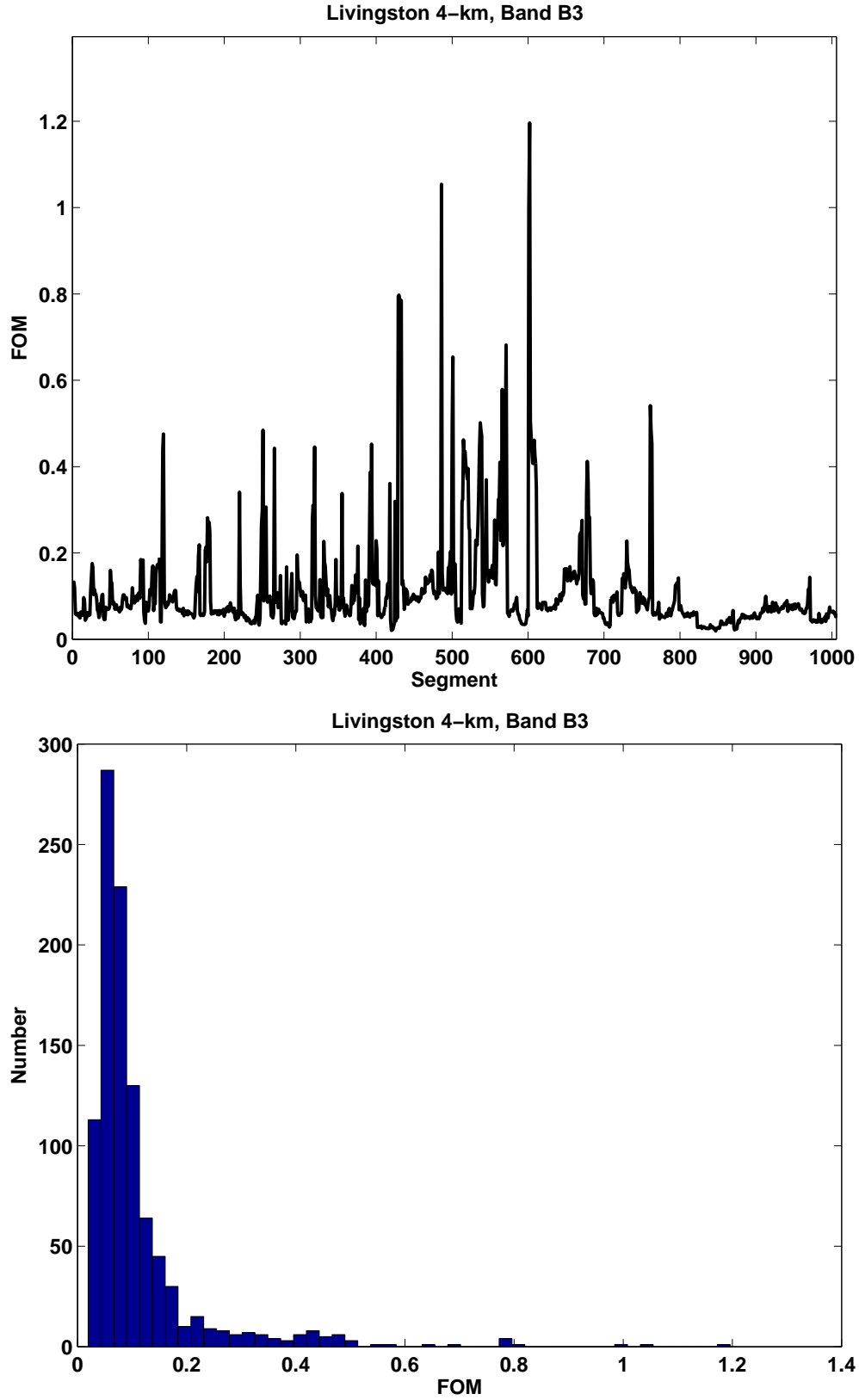


Figure 36: Data conditioning FOM for Livingston 4-km detector, band B3

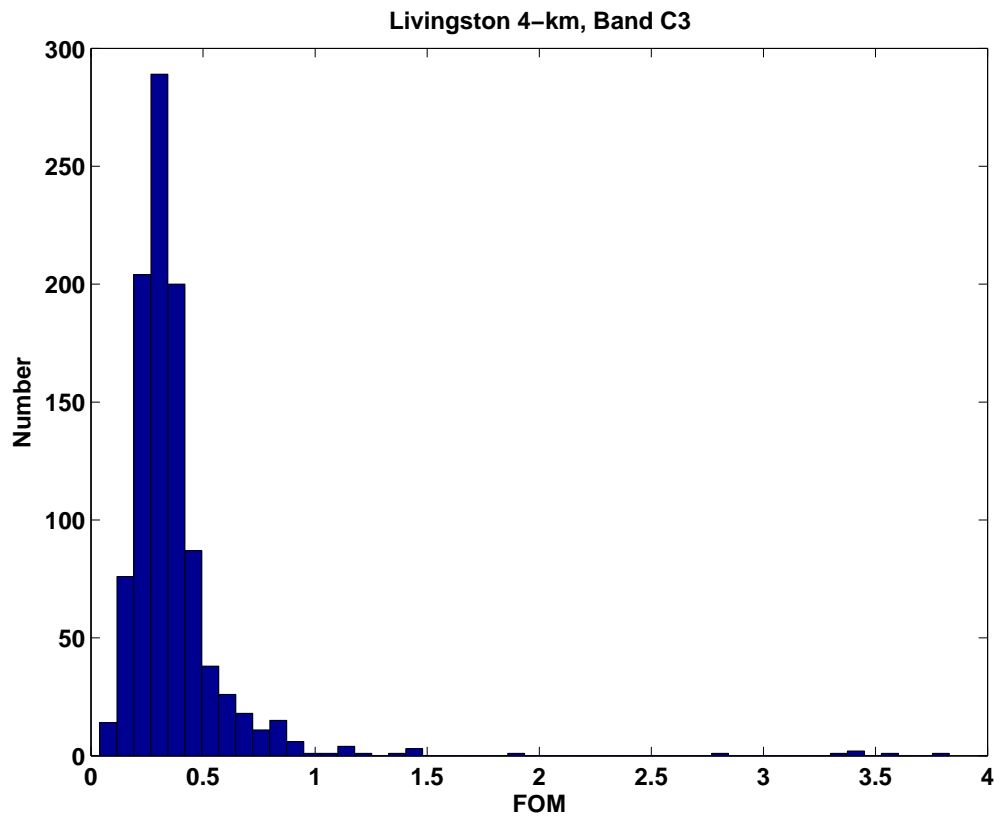
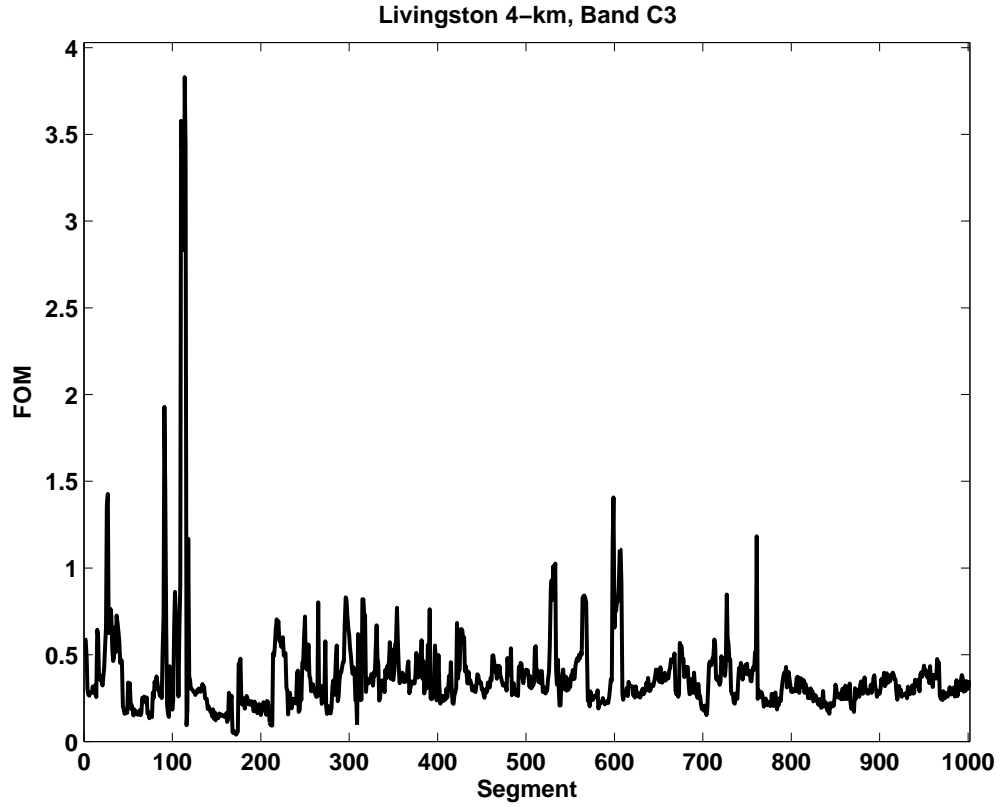


Figure 37: Data conditioning FOM for Livingston 4-km detector, band C3



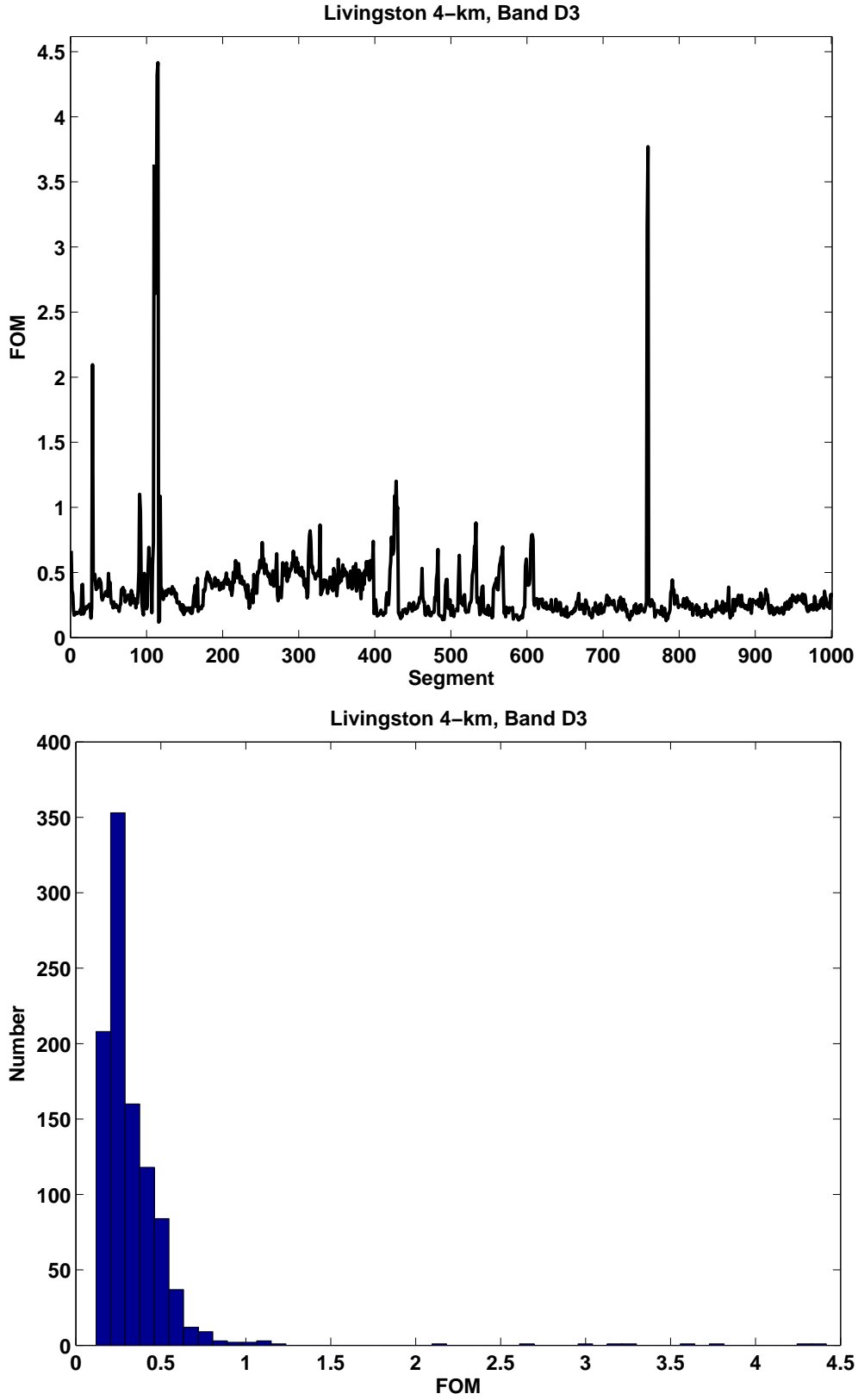


Figure 38: Data conditioning FOM for Livingston 4-km detector, band D3

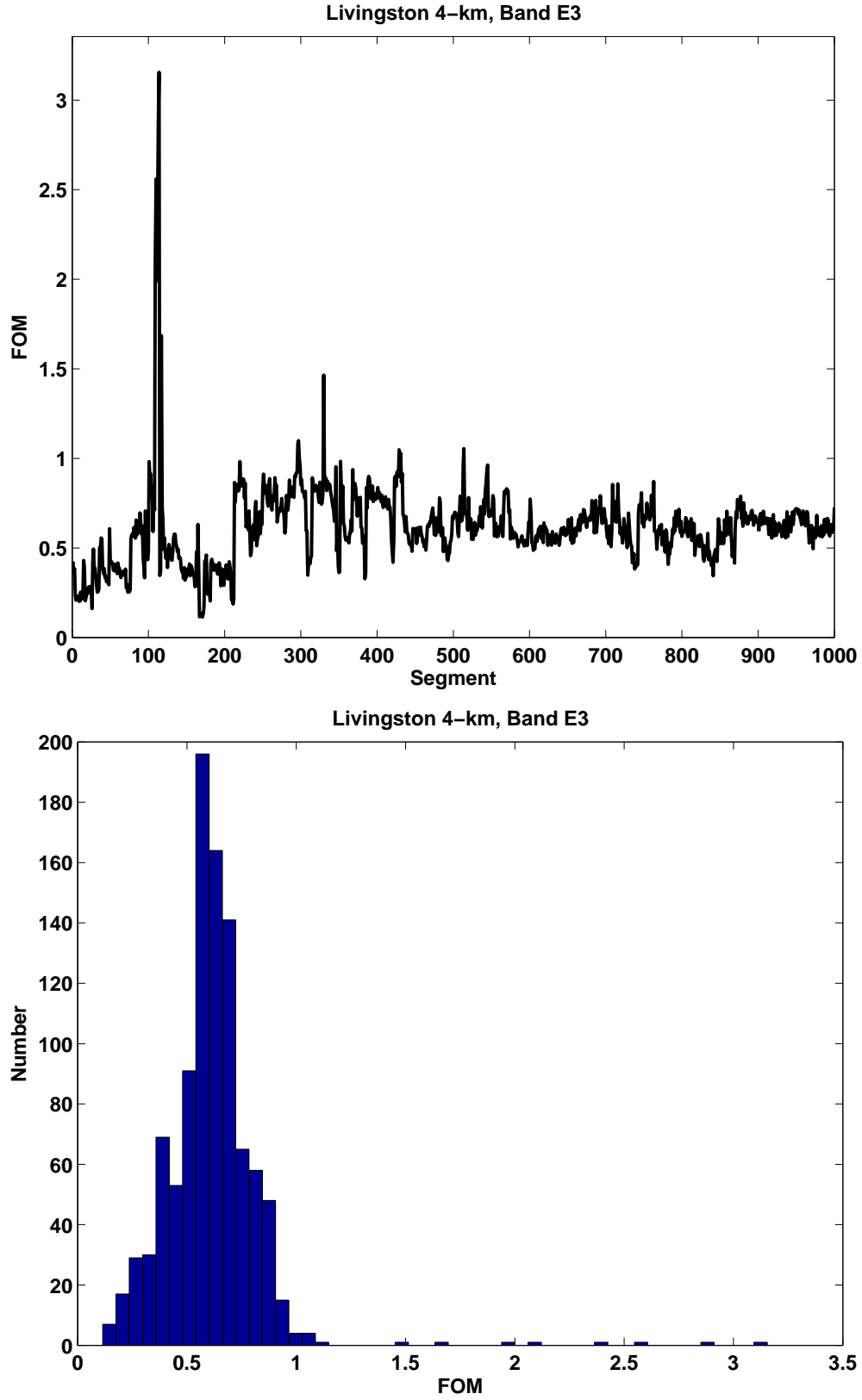


Figure 39: Data conditioning FOM for Livingston 4-km detector, band E3

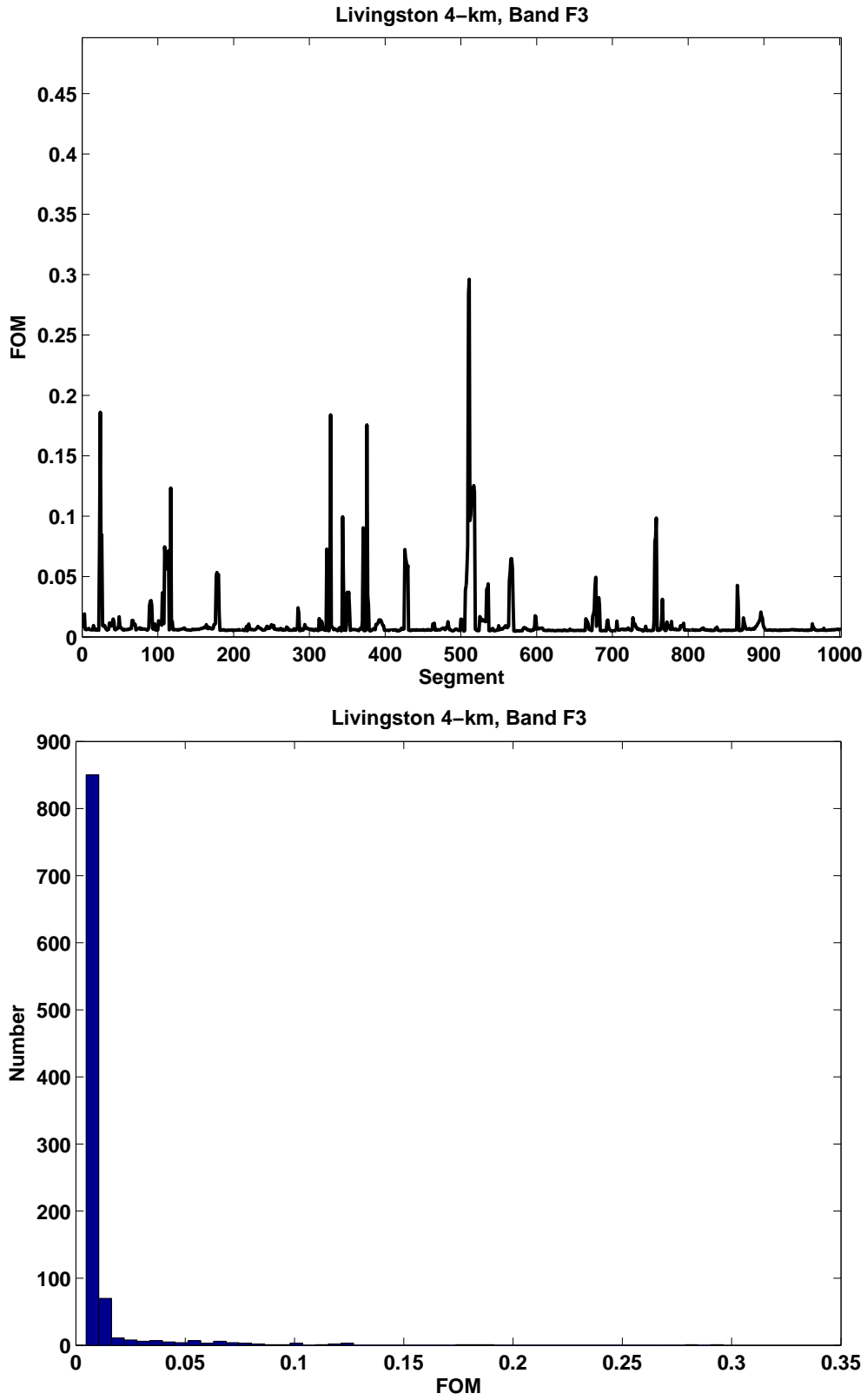


Figure 40: Data conditioning FOM for Livingston 4-km detector, band F3

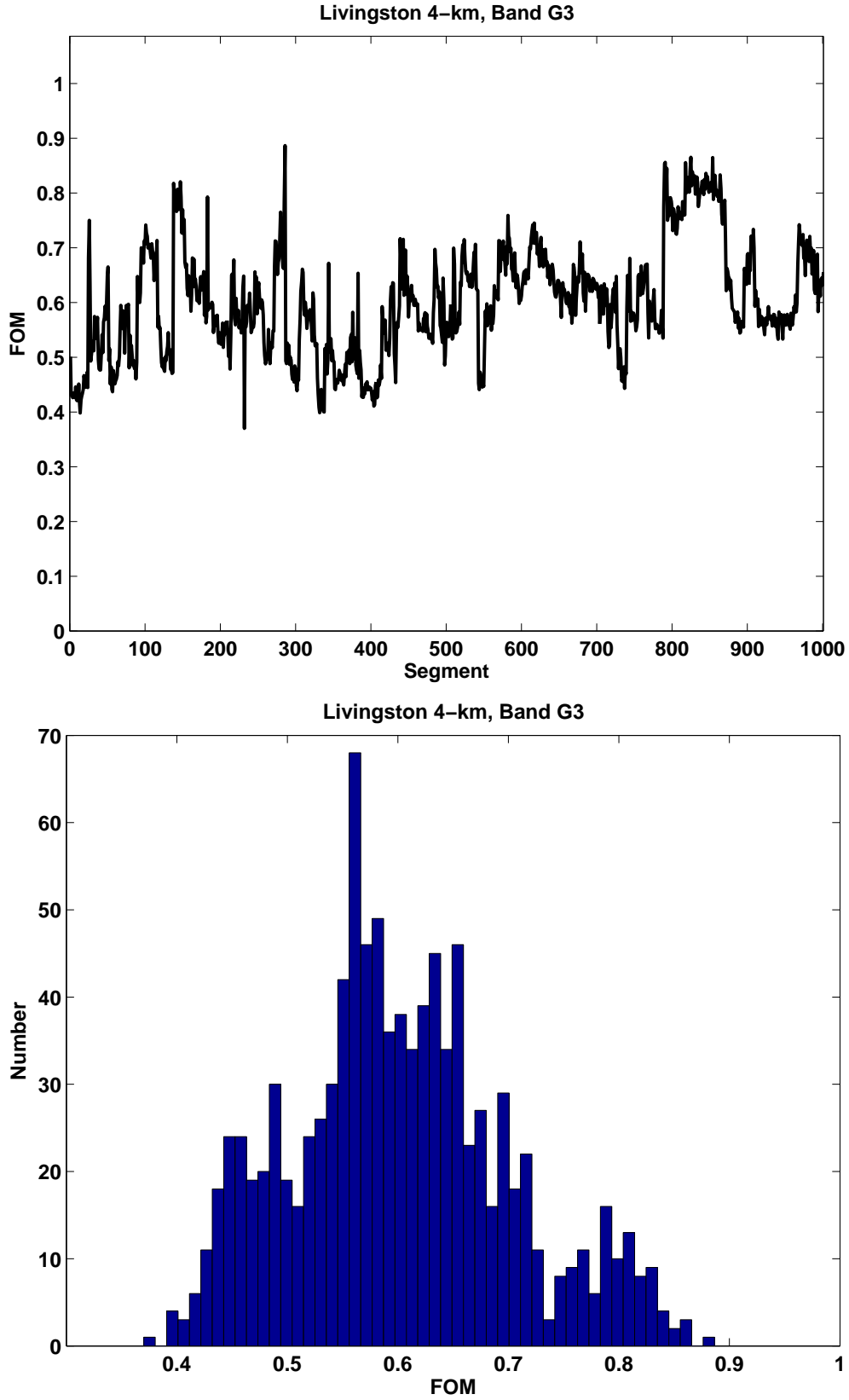


Figure 41: Data conditioning FOM for Livingston 4-km detector, band G3

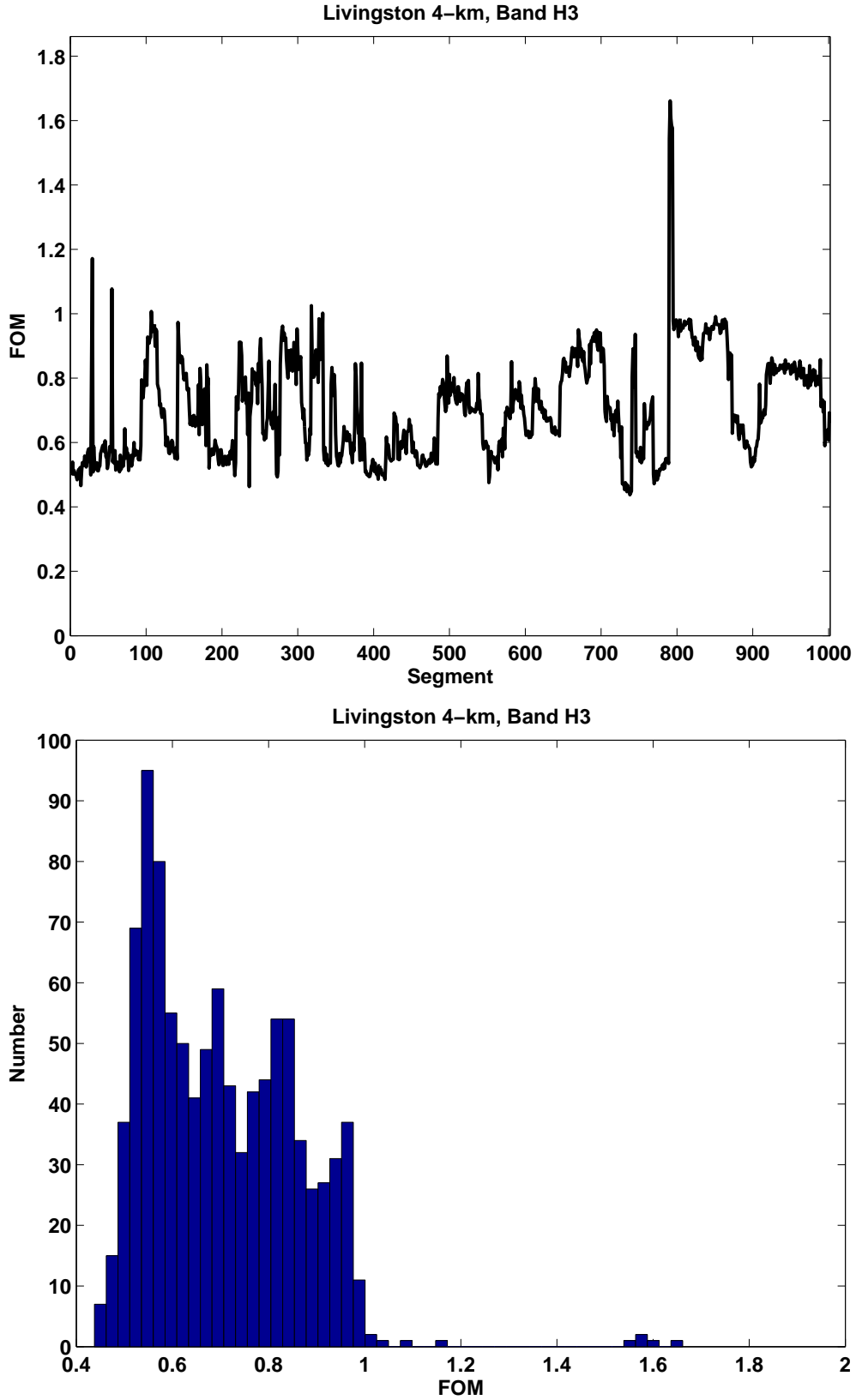


Figure 42: Data conditioning FOM for Livingston 4-km detector, band H3

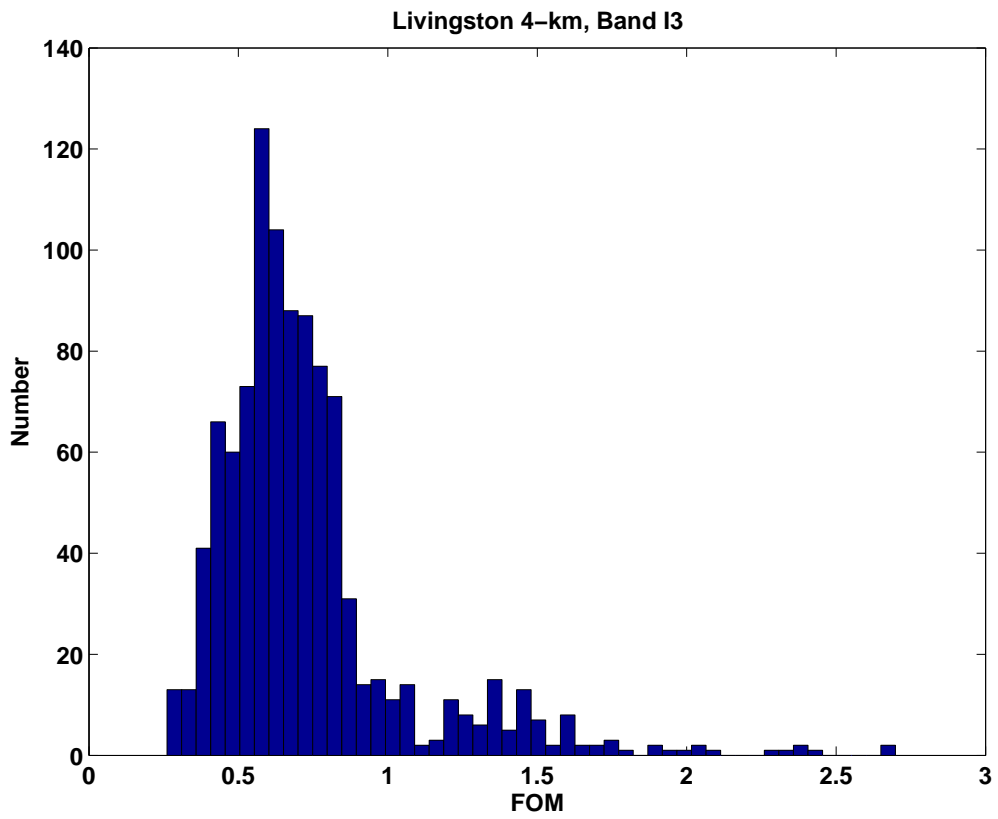
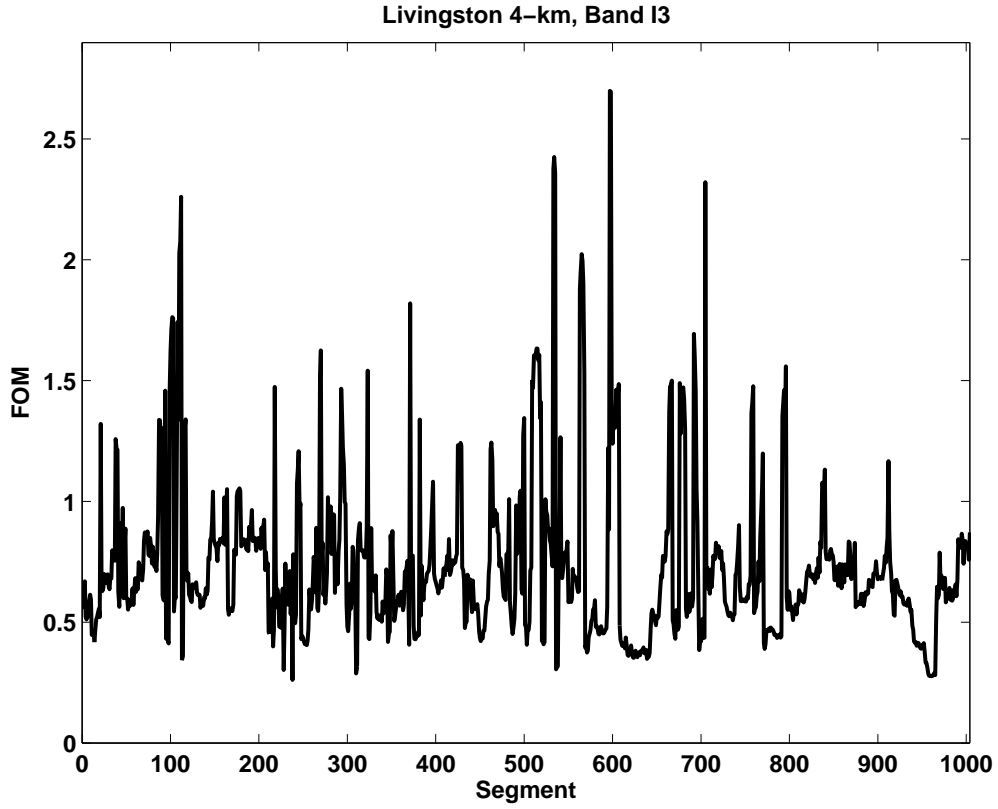


Figure 43: Data conditioning FOM for Livingston 4-km detector, band I3

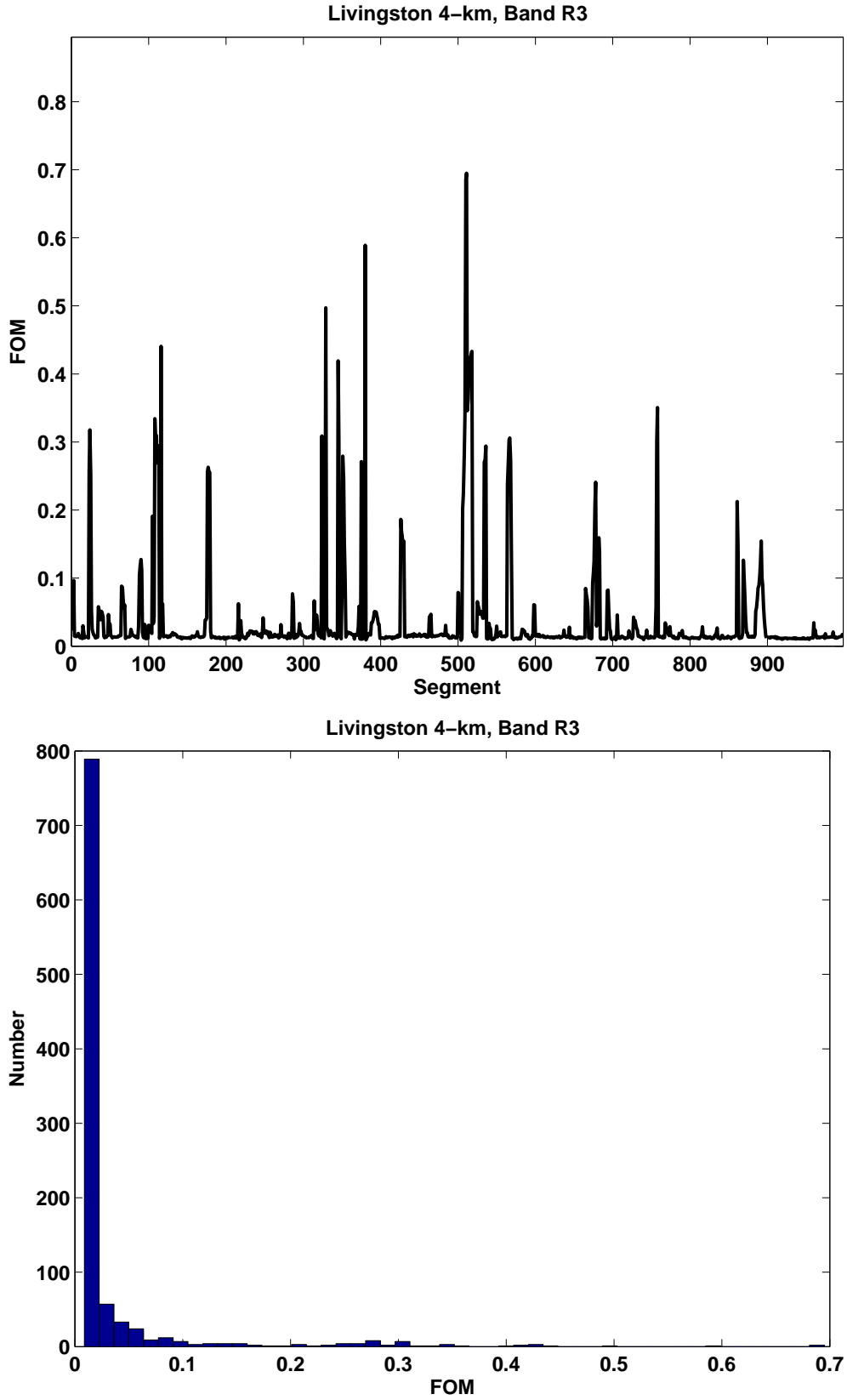


Figure 44: Data conditioning FOM for Livingston 4-km detector, band R3

## References

- [1] McNabb, J. W. C., Ashley, M., Finn, L. S., Rotthoff, E., Stuver, A., Summerscales, T., Sutton, P., Tibbits, M., Thorne, K., & Zaleski, K. *Overview of the BlockNormal event trigger generator* Class. Quant. Grav. **21**, S1705-S1710 (2004).
- [2] Lazzarini, A. *Strain Sensitivities for the LIGO Interferometers for S1* LIGO Presentation LIGO-G020461-01-E LIGO Document Control Center (2003).
- [3] Lazzarini, A. *Strain Sensitivities for the LIGO Interferometers during S2* LIGO Presentation LIGO-G030379-00-E LIGO Document Control Center (2003).
- [4] Lazzarini, A. *S3 Best Strain Sensitivities* LIGO Presentation LIGO-G040023-00-E LIGO Document Control Center (2004).
- [5] Lazzarini, A. *S4 Strain Sensitivities for the LIGO Interferometers* LIGO Presentation LIGO-G050230-02-E LIGO Document Control Center (2005).
- [6] Oppenheim, A. V. & Schaffer, R. W. *Discrete-Time Signal Processing* Prentice Hall, Upper Saddle River, New Jersey, 486-501 (1999).
- [7] Kalman, R. E., *A New Approach to Linear Filtering and Prediction Problems* Transaction of the ASME – Journal of Basic Engineering, 35-45 (1960).
- [8] Welch, Greg & Bishop, Garry *An Introduction to the Kalman Filter* [http://www.cs.unc.edu/welch/media/pdf/kalman\\_intro.pdf](http://www.cs.unc.edu/welch/media/pdf/kalman_intro.pdf) (2004).
- [9] Finn, L. S., & Mukherjee, S. *Data conditioning for gravitational wave detectors: A Kalman filter for regressing suspension violin modes* Phys. Rev. D **63** (2001).
- [10] Brown, Robert Grover & Hwang, Patrick Y. C. *Introduction to Random Signals and Applied Kalman Filtering* John Wiley & Sons, New York (1997).
- [11] Finn, L. S., & Mukherjee, S. *Erratum: Data conditioning for gravitational wave detectors: A Kalman filter for regressing suspension violin modes* Phys. Rev. D **67** (2003).
- [12] Finn, L. S., & Saulson, P. *S2 Playground Selection* LIGO Technical Note LIGO-T030020-01-Z LIGO Document Control Center (2003).
- [13] Finn, L. S. *S3 Playground Selection* LIGO Technical Note LIGO-T030256-00-Z LIGO Document Control Center (2003).
- [14] Ljung, L. *System Identification Toolbox* The Mathworks, Inc. (2000).
- [15] Thorne, K. *Testing Data Conditioning used in BlockNormal Event Trigger Generator* LIGO Technical Note LIGO-T060028-00-Z LIGO Document Control Center (2006).
- [16] Zweizig, J. G. *S2/S3 Data Quality Flagging* LIGO Presentation LIGO-G030642-00-E LIGO Document Control Center (2003).



- [17] Zweizig, J. G. *S3 Data Quality Investigation* LIGO Presentation LIGO-G040136-00-E LIGO Document Control Center (2004).
- [18] Zweizig, J. G. *S4 Data Quality* LIGO Presentation LIGO-G050197-00-E, LIGO Document Control Center (2005).
- [19] Zweizig, J. G. *S3 Data Quality Review*  
[http://www.ligo.caltech.edu/%7Ejzweizig/S3\\_Data\\_Quality/index.html](http://www.ligo.caltech.edu/%7Ejzweizig/S3_Data_Quality/index.html)