SAND2023-12003C

HAMPEL FILTERING OF POWER SPECTRAL DENSITIES TO REMOVE SPURIOUS SINE TONES IN RANDOM VIBRATION DATA

James Woodall and Vit Babuška
Sandia National Laboratories*
Albuquerque, NM 87185

ABSTRACT

Collecting accurate data during vibration testing requires clean data. One of the first steps in a vibration test is the characterization of the noise floor of the measurement chain. This step is important so that artifacts of the measurement chain do not pollute the measured data, particularly at low signal to noise ratios. During a recent transportation vibration test conducted by Sandia National Laboratories, the data acquisition system introduced narrow band spurious signals that were large enough to be observable in the ASDs of the accelerations collected during the test. The spurious noise tones had to be removed from the ASDs collected during the driving phase of the test. A novel approach based on frequency domain median filtering of the ASDs was used to clean the measured data. Three methods were developed, and all were able to remove the sine tones. The results show that frequency domain median filtering is a viable, straightforward way to remove certain types of measurement noise artifacts without biasing the data.

I. INTRODUCTION

Vibration environments are usually represented as Autospectral Densities (ASDs), which quantify the spectral content and intensity of a given time history. The ASDs of a quiescent system characterize the noise floor, which when compared to the ASD of the system in a dynamic environment, provide a signal-to-noise ratio (SNR) spectrum. This ratio indicates which elements of the dynamic system ASDs are relevant with respect to the background noise so that artifacts of the measurement chain are not given undue consideration in the processing and analysis of the data. The artifacts that may manifest in the noise floor need to then be rectified since they could otherwise show up in the dynamic system ASDs, contaminating it and influencing conclusions that may be drawn from the data.

Usually, measurement noise is a random phenomenon and therefore can be reduced by averaging. The variance of the estimated (i.e., measured) ASD is inversely proportional to the number of averages used to compute the ASD [1]. This relationship means that for additive noise, the SNR increases with the number of segments because the noise level decreases with the number of segments.

Averaging does not reduce the levels of noise sources that are not random. For example, 60 Hz electrical noise artifacts sometimes pollute data if the measurement instrumentation is not properly grounded. The poor ground appears as a 60 Hz tone in the spectral data, often with higher

* Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

harmonics as well. Ideally this artifact would be removed by proper grounding prior to collecting test data.

Sometimes deterministic noise artifacts of unknown origin appear in the data, and—since they were not removed prior to collecting data—their contributions to the ASDs should be removed. Deterministic artifacts, such as sine tones from rotating machinery or electrical noise, are often removed with time-domain filtering. Usually this removal is done with digital filters such as the Vold-Kalman Order Tracking filter [2,3] for tracking sinusoidal signals or with notch filters that use sine tone estimation with FIR (Finite Impulse Response) or IIR (Infinite Impulse Response) filters [4] to attenuate a signal at a specific frequency.

Deployment of these time domain methods requires a-priori knowledge of the artifacts to be filtered out or a disturbance rejection adaptive control system. If one does not anticipate the presence of deterministic artifacts, then the data may be corrupted from pollution by these artifacts. If the complete time domain data from a test are available, the deterministic artifacts can be removed by filtering during post processing. However, there may be real data that are of interest at these frequencies as well, and time domain filtering runs the risk of attenuating real data.

This paper describes frequency domain methods to eliminate deterministic noise artifacts (i.e., sine tones). The methods are applicable when the time history of a given test record is not available after a test. For longer duration tests, it is beneficial to have a means of mitigating noise artifacts from ASDs directly. The methods described herein use the noise floor of a test to locate artifacts in the frequency domain to mitigate their pollution of the test ASDs. This approach depends on the assumption that noise artifacts will pollute the ASDs additively, and that they are easily identified in a system's noise floor where they are the only deterministic artifact.

This remainder of this paper is organized into five sections. First the motivation for the work is described in the next section, Section II. The approaches are based on median filtering in the frequency domain, so the basics of median filters are presented in Section III. The approaches investigated, and their performance are described in Section IV. Section V gives conclusions and recommendations.

II. MOTIVATION

The motivation for this work was a transportation vibration test, called an Over-the Road (OTR) test, whose objective was the establishment of transportation environments for the cargo. Toward this end, a range of sensors collected acceleration data at a variety of locations as the vehicle operated on several different road types over a range of speeds. The overall duration of the test was a few hours and data were recorded at a high sample rate. For a general analysis, the data were divided into three distinct phases. The first of these is the *engine-off quiescent phase*. Data from this phase were used to determine the noise floor of the test, as the system was assumed to be perfectly static. The second phase is the *engine-on static phase*, when the vehicle was not in motion but had been turned on. The final phase is the *driving phase*, which accounts for all data while the vehicle was in motion. The transportation vibration environments, both random vibration and shock environments, are derived from data measured during this phase.

In many of the recorded channels, sine-tone spikes cropped up in the ASDs during the quiescent phase as well as in the other phases of the test. Their manifestation in a nominally quiescent environment corresponds to no known physical cause, and consequently they are considered to be noise artifacts from the measurement chain rather than meaningful data. Figure 1 presents the noise floor for two channels, illustrating the contaminating sine-tone spikes.

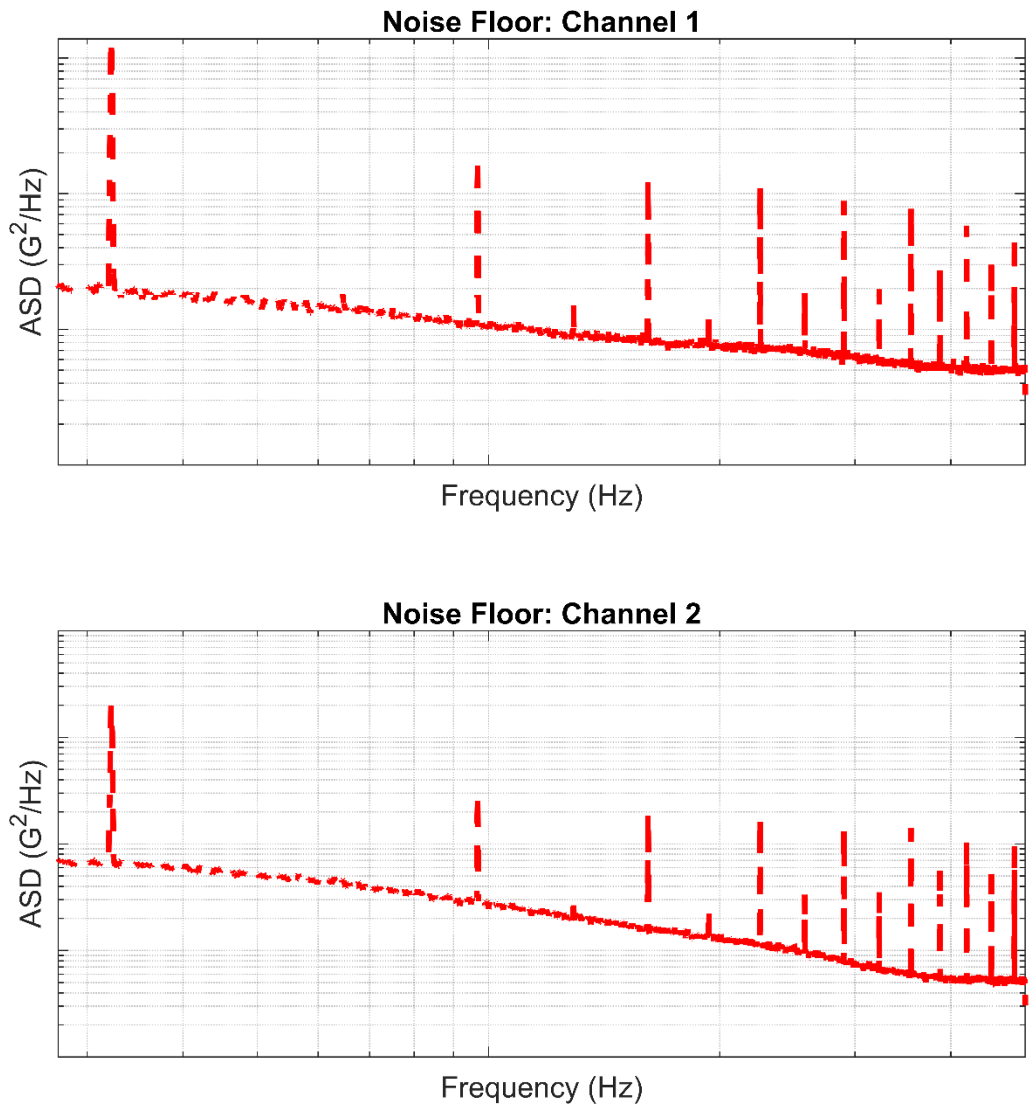

Figure 1: Channel 1 & 2 Noise Floors

Since these spikes are not representative of the true vibration environment, they should be removed from the driving phase data prior to analysis so as to not contaminate the data measured in the driving phase. This paper describes methods of sine tone spike removal in the frequency domain using median filters.

III. MEDIAN FILTERS

Median filters remove noise from signals, smoothing the data while preserving edges better than many other data-smoothing techniques. These filters work by sliding a window, most often defined by an odd number of points, over a set of data and replacing the central value for each window with the window's median value. Median filters are often used effectively in image processing to remove "popcorn noise" or hot pixels, at the cost of some reduction in sharpness and clarity. This paper employs a Hampel Filter, which builds off the concept of a median filter by adding an element of outlier detection. Developed by the statistician Frank Hampel, this filter checks the central value of each window against that window's standard deviation. If the central value is greater than the predetermined number of standard deviations, $k\sigma_j$, then the value is identified as an outlier and replaced like a traditional median filtered value.

$$y_j = \begin{cases} x_j \text{ if } |x_j - m_j| \leq k\sigma_j \\ m_j \text{ if } |x_j - m_j| > k\sigma_j \end{cases} \tag{1}$$

where σ_j is the standard deviation of the data in the window.

Most often k is assigned a value of 3 or 4. The window length should be long enough so that any outliers do not bias the standard deviation estimate. This conditional application of a median filter allows the filter's strength to shine by mitigating outliers without unnecessarily smoothing legitimate data. Only two parameters need to be determined to employ the Hampel Filter: the window length, and the number of standard deviations, k, that a value must exceed to be considered an outlier.

IV. APPROACHES

Noise Floor Removal

This method is the reference method and consists simply of reducing the Driving Phase ASDs by the measured noise floor across all frequencies. At frequencies with a high signal-to-noise ratio, the reduction will have little overall effect beyond reducing the sine-tone spikes in the Driving Phase by those in the noise floor, while frequencies with low signal-to-noise ratio will be under-represented. In most cases caution should already be exercised in the frequencies with low signal-to-noise ratio, as contamination from any low-level measurement chain artifacts is more likely to corrupt the data in subtle ways, reducing the confidence about the veracity of the environments at those frequencies. Figure 2 depicts the SNR for the two representative channels chosen as examples in this paper, while Figure 3 shows the Driving Phase ASDs, the noise floors, and the results of applying the Noise Floor Removal method. Note that the SNR drops at the sine tone frequencies.

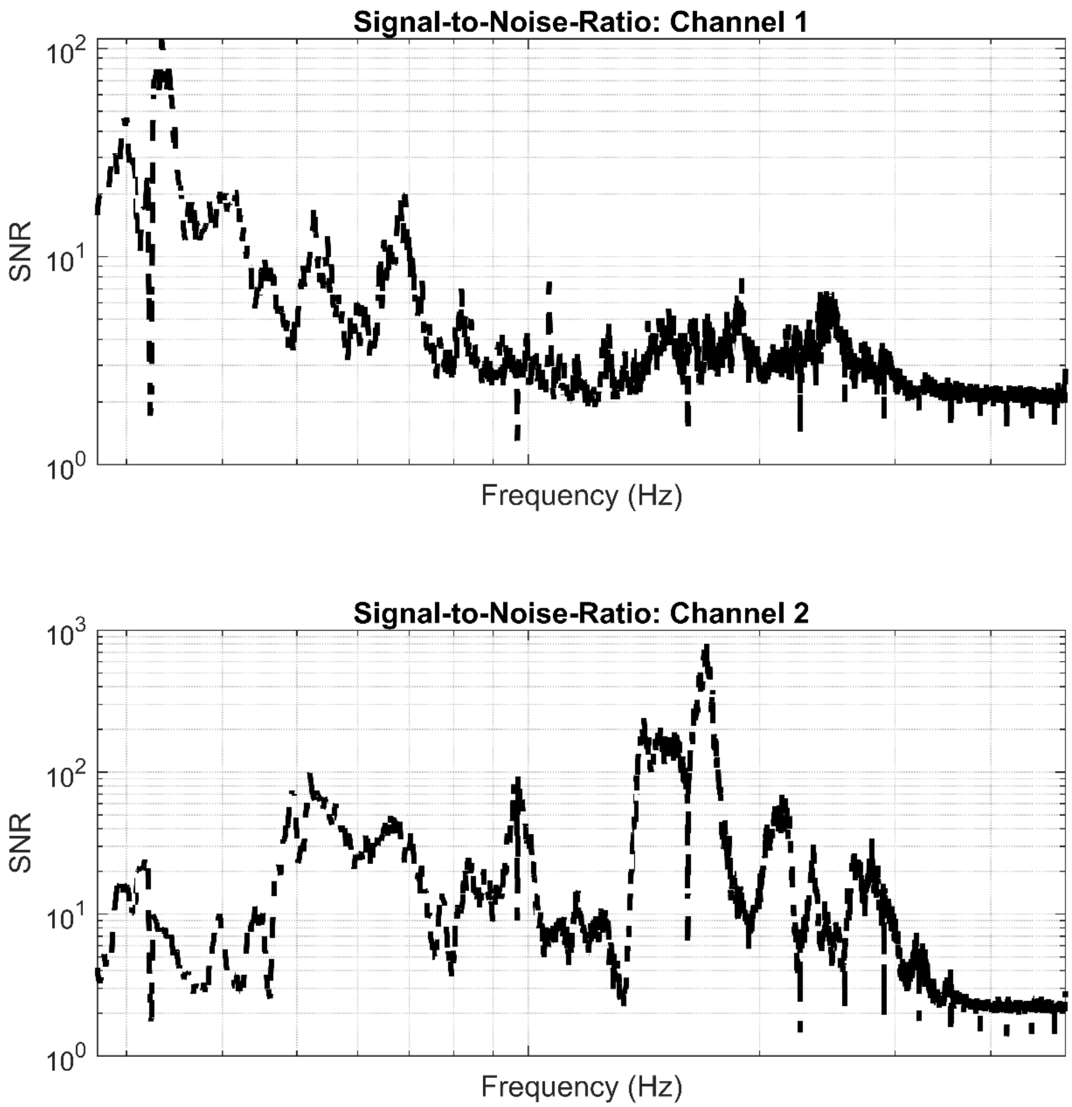

Figure 2: Signal-to-Noise Ratios

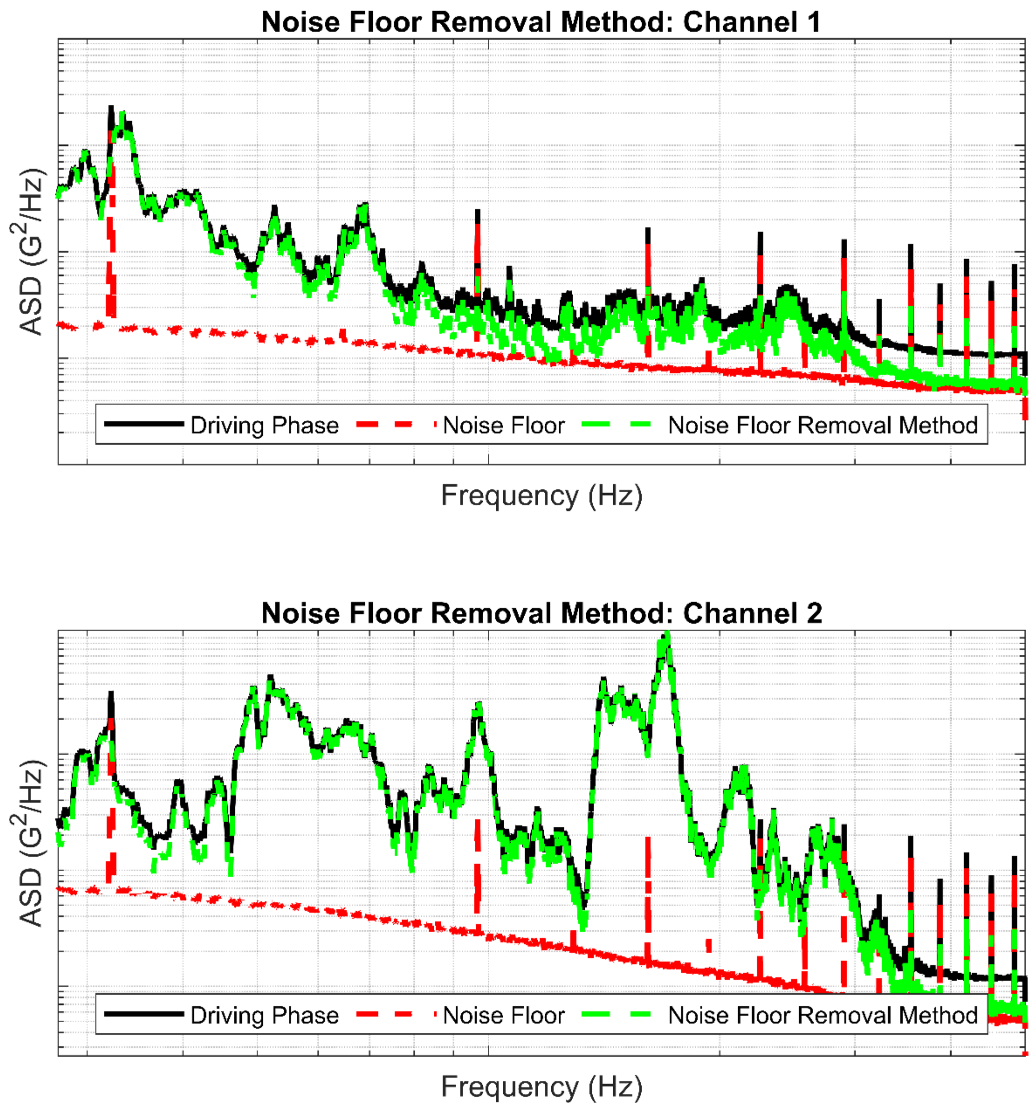

Figure 3: Noise Floor Removal Results

For the frequencies with the highest Signal-to-Noise Ratio, the Noise Floor Removal method does a reasonable job of mitigating the sine-tone spikes. However, these are also the frequencies where it is hard to determine the extent of the contamination by eye. Low SNR frequencies still maintain a distinctive spike, while also notably reducing the magnitude of the signal at surrounding frequencies. Even at mid-level SNR frequencies, a residual spike is left, protruding from an obviously reduced curve. This method diminishes the SNR of the Driving Phase ASD by 1.0 across all frequencies-as a general rule it seems to mitigate the sine-tone spike artifacts adequately when the SNR is above 10 but struggles to be effective when the Driving Phase ASD has a lower magnitude.

Filtered Noise Floor

This method approaches the problem similarly to the Noise Floor Removal method, but rather than applying a blanket reduction of the Driving Phase ASD with the noise floor, only the frequencies with the sine-tone spikes are affected. The reduction of only the sine tone spikes is achieved through the use of a Hampel Filter, which is applied to the noise floor (NF) in order to obtain a Filtered Noise Floor (NF_{DS}). All subsequent methods likewise depend on the derivation of this Filtered Noise Floor. The Hampel Filter applied used a window of 33 points, replacing values beyond four standard deviations of that window with its median value. Figure 4 illustrates the Filtered Noise Floors against their progenitors.

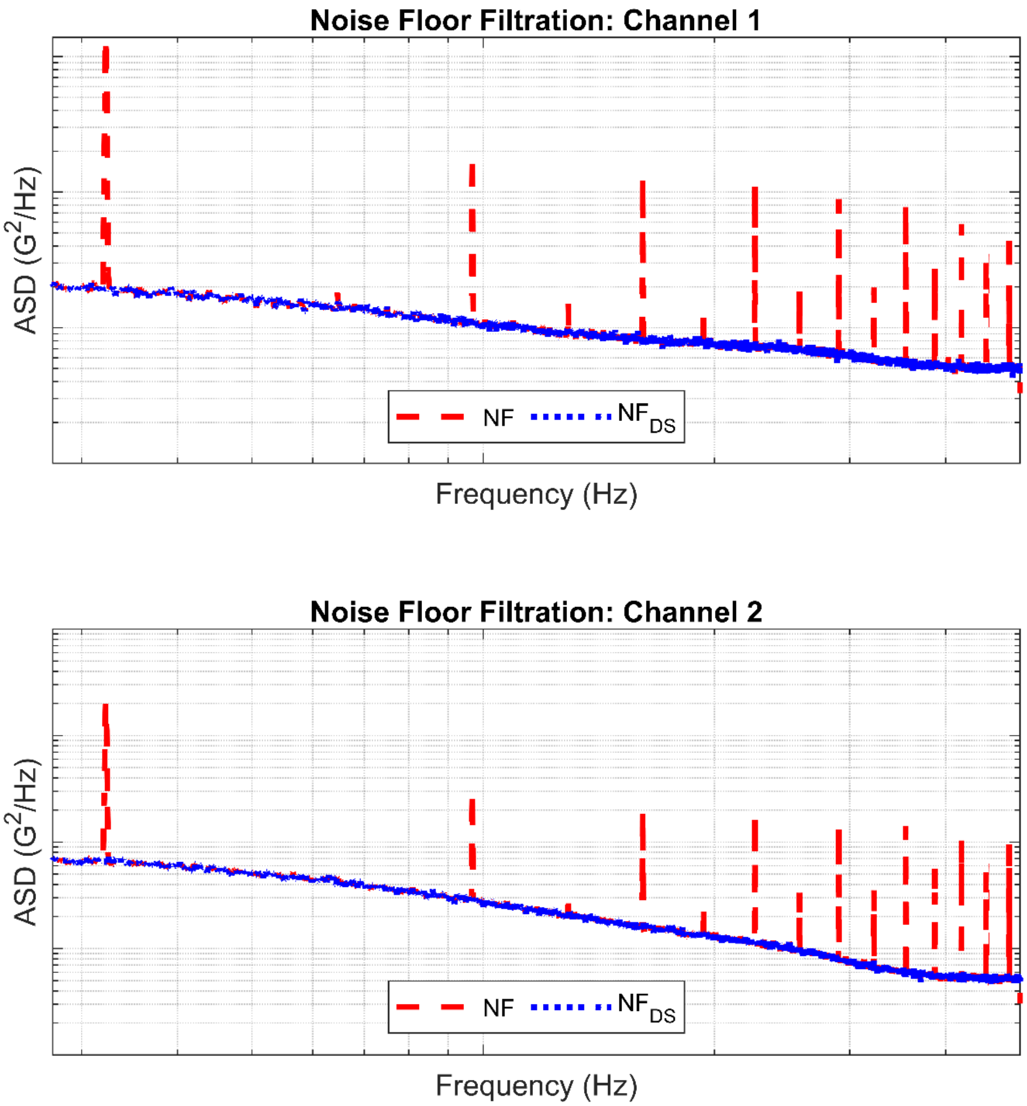

Figure 4: Filtered Noise Floors

The Driving Phase ASD is then reduced by the difference between the original Noise Floor and the Filtered Noise Floor, as per Eq. (2):

$$ASD_{DS} = ASD - (NF - NF_{DS}) \tag{2}$$

where ASD_{DS} is the Driving Phase ASD after the sine-tone spikes have been mitigated, ASD is the original Driving Phase ASD, and NF and NF_{DS} are the original and Filtered Noise Floors, respectively. This approach removes just the sine tones, avoids the pitfall of the Noise Floor

Removal method in altering frequencies that are untouched by the noise artifacts, and reduces their manifestation by the extent to which they diverge from the original Noise Floor. Figure 5 presents the results of applying this method.

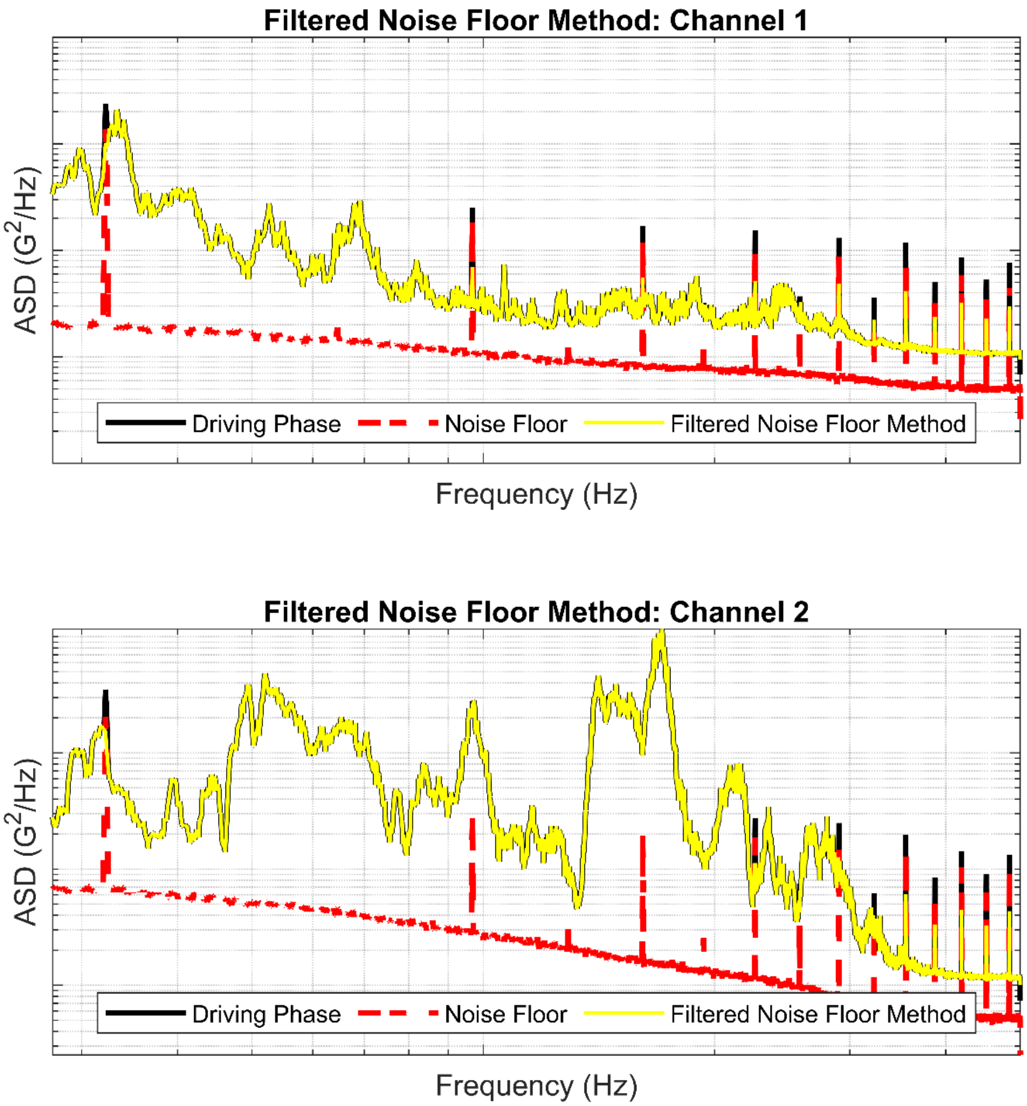

Figure 5: Filtered Noise Floor Method Results

Due to the Hampel Filter's identification of the noise artifacts, the Driving Phase ASD experienced significantly less alteration than with the Filtered Noise Floor method. While the spikes were greatly reduced, there are still residual peaks at frequencies with noise artifacts. These appear most severe at frequencies with low SNR, in which log-scale ASDs are traditionally viewed to draw out the differences at low magnitudes.

Difference of Roots

The Difference of Roots method shares a similar approach to the Filtered Noise Floor method but does not assume a linear relationship between the divergence of the spikes in the noise floor and that in the ASD. Instead, this method considers the fact that the parameters involved are powers of a quantity (G^2/Hz), and Eq. (2) is updated to:

$$ASD_{DS} = \left(ASD^{\frac{1}{2}} - \left(NF^{\frac{1}{2}} - NF_{DS}^{\frac{1}{2}}\right)\right)^2 \quad (3)$$

The results of applying this method are depicted in Figure 6.

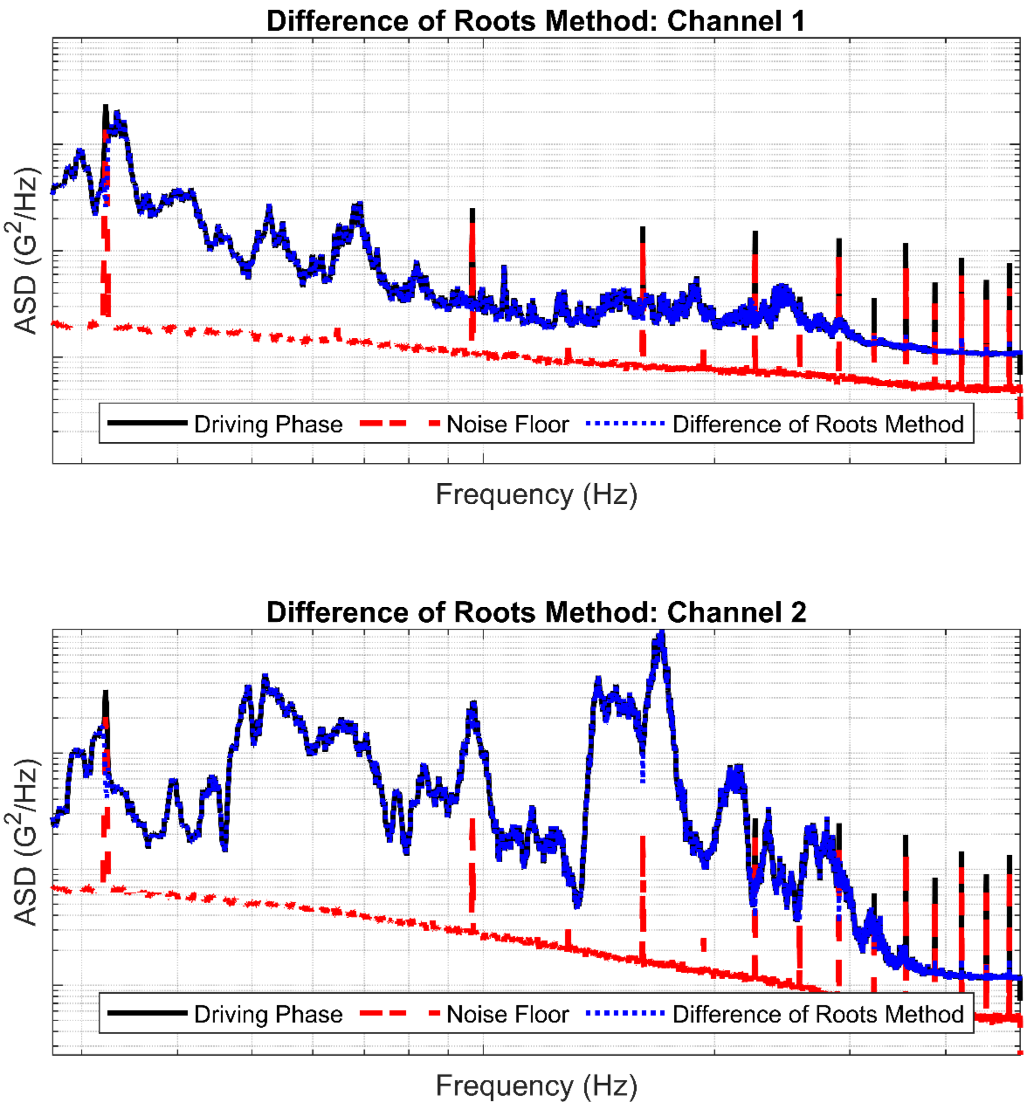

Figure 6: Difference of Roots Method Results

Contrasting the Noise Floor Removal and the Filtered Noise Floor methods, the Difference of Roots approach performs best at low-magnitude frequencies, leaving only very small residual peaks. However, this method appears to over-compensate at frequencies with high SNR, dragging the ASD lower than the surrounding frequencies would seem to indicate is appropriate. While this deviation can appear small, especially given the fact that it occurs in log-scale where the apparent differences are diminished, it still results in an under-conservative representation of the environment at that frequency, and some tests may be better served by leaving a large conservative discrepancy as opposed to creating a small, under-conservative error.

LS Model

The Least Squares (LS) Model method forgoes any assumptions about how the sine-tone spikes in the noise floor relate to those in the Driving Phase ASD, and instead takes a collection of sine-tone spike properties to create a LS model. Four parameters were taken from each of the sample frequencies that contain a sine-tone peak. These parameters are the magnitude of the spike in the noise floor (NF Pk), the expected magnitude at that frequency given the magnitude of the curve at surrounding frequencies (NF Ex), the magnitude of the spike in the Driving Phase (DP Pk), and the expected magnitude of the ASD at that frequency given the surrounding magnitudes. From these parameters, the difference of each curve's peak to its expected value was computed (NF Dif and DP Dif, respectively). These parameters are illustrated in Figure 7.

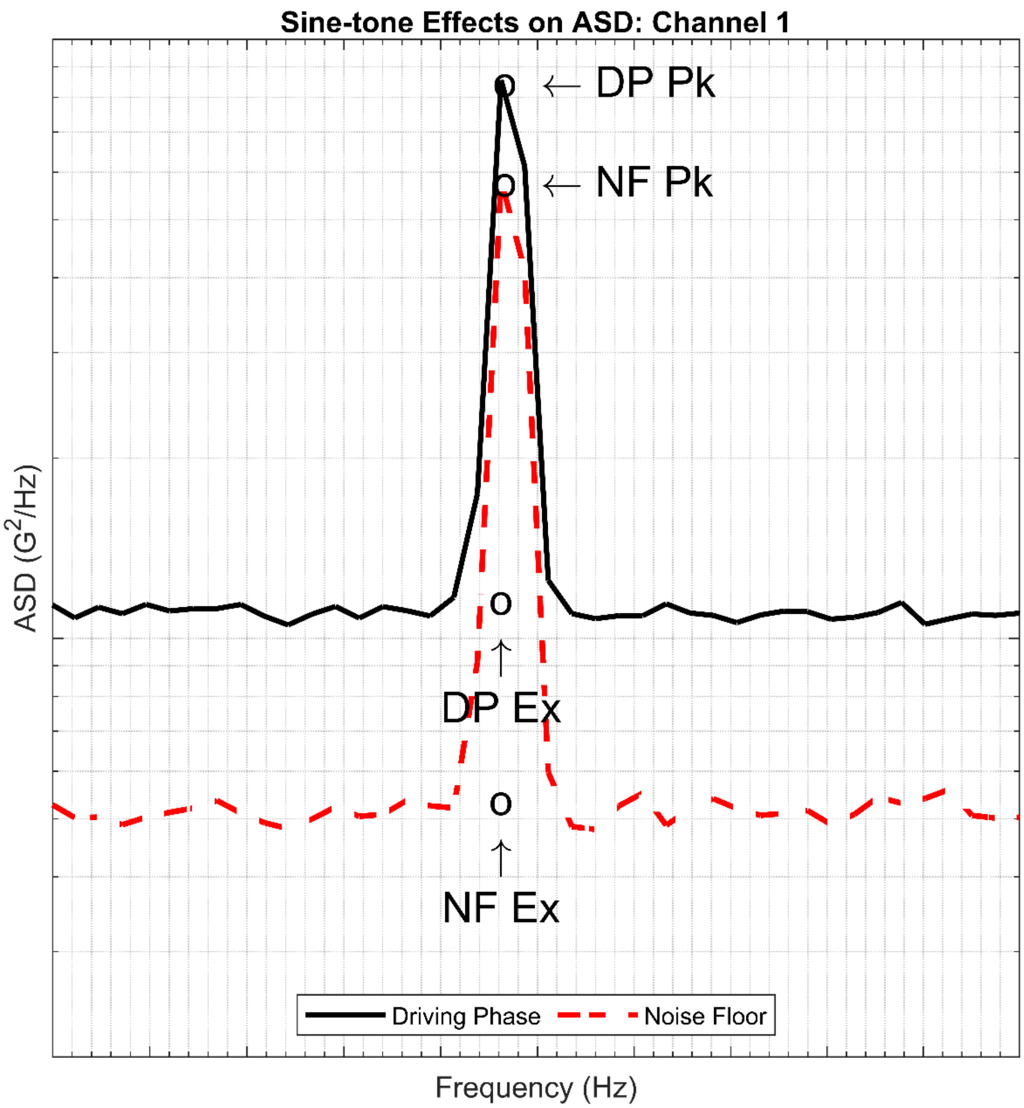

Figure 7: Collected Parameters

DP Dif was found to be linearly related to NF Dif on a loglog scale, so a linear regression model was created for the logs of these parameters to quantify that relationship. The scatter-plot of these logs for the samples taken is shown in Figure 8.

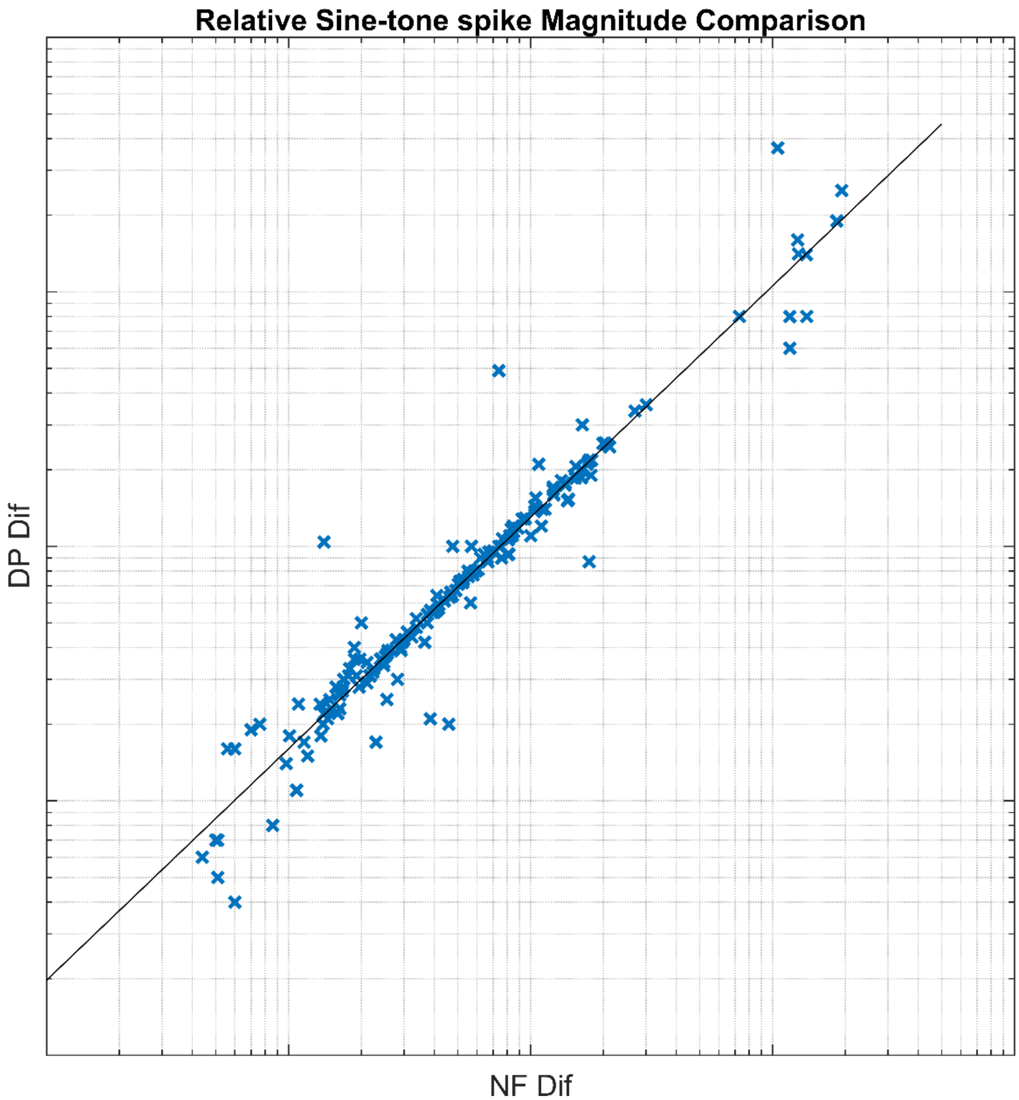

Figure 8: Log-scale Scatterplot of Approximate Spike Deviations

The Least Squares Linear Regression Model for this data takes the form

$$\log(DP\ Dif) = A * \log(NF\ Dif) + B \tag{4}$$

This expression is then transformed to be more readily applicable in linear space:

$$Y = X^A * e^B$$

Where X represents the NF Dif axis, Y is the DP Dif Axis, and A and B are the respective slope and intercept for the least squares regression model using the logs of NF Dif and DP Dif. Since Y represents the difference between the sine-tone peak and the expected value at that frequency, it is subtracted from the Driving Phase ASDs. The results of applying this method are shown in Figure 9.

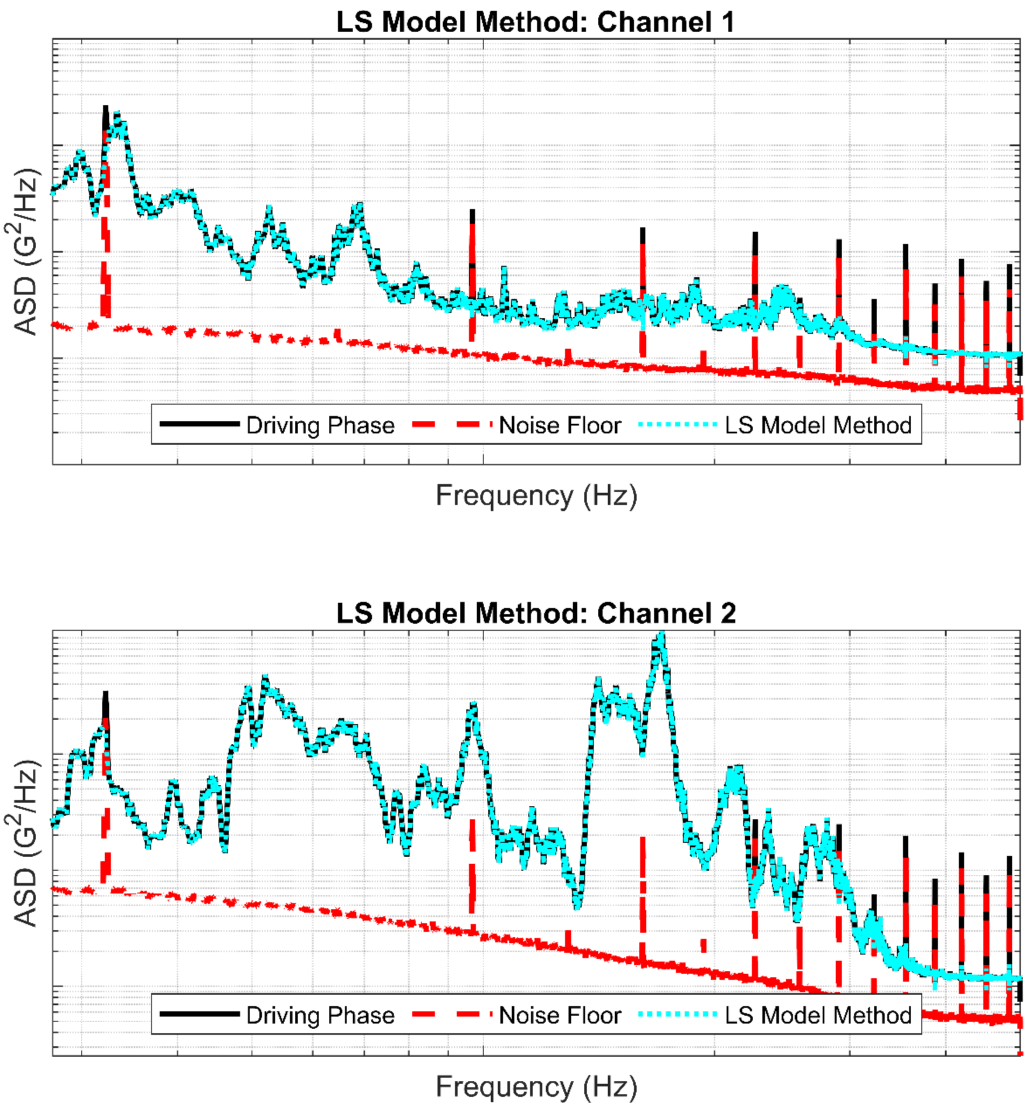

Figure 9: Least Squares Model Method Results

This method effectively brings down the artificial peaks, adjusting them to be believably within the range of expected values of each frequency given the curve behavior at adjacent frequencies. Both the sine-tone peaks at higher and lower SNR frequencies appear to be reasonably well mitigated, although a residual "jag" pattern can be left where each peak had been. This jag is most apparent at frequencies with low-magnitude spectral content, but still much more closely approximates the expected values than the other methods. However, this method is more reliant on preprocessing, requiring the collection of a representative sampling of the artifact parameters to derive A and B. Figure 10 presents all of the methods together, while Figure 11 zooms in on the labeled windows.

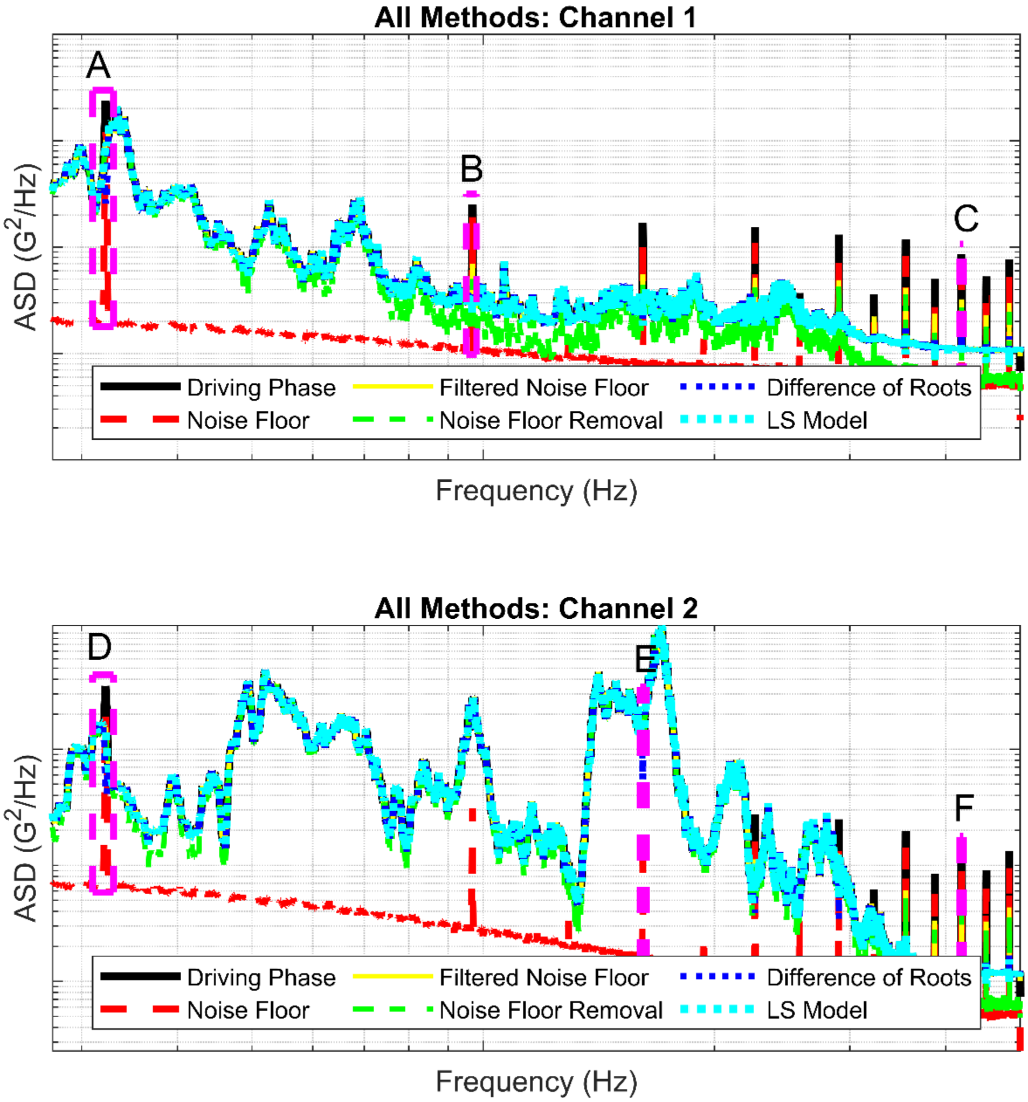

Figure 10: All Methods

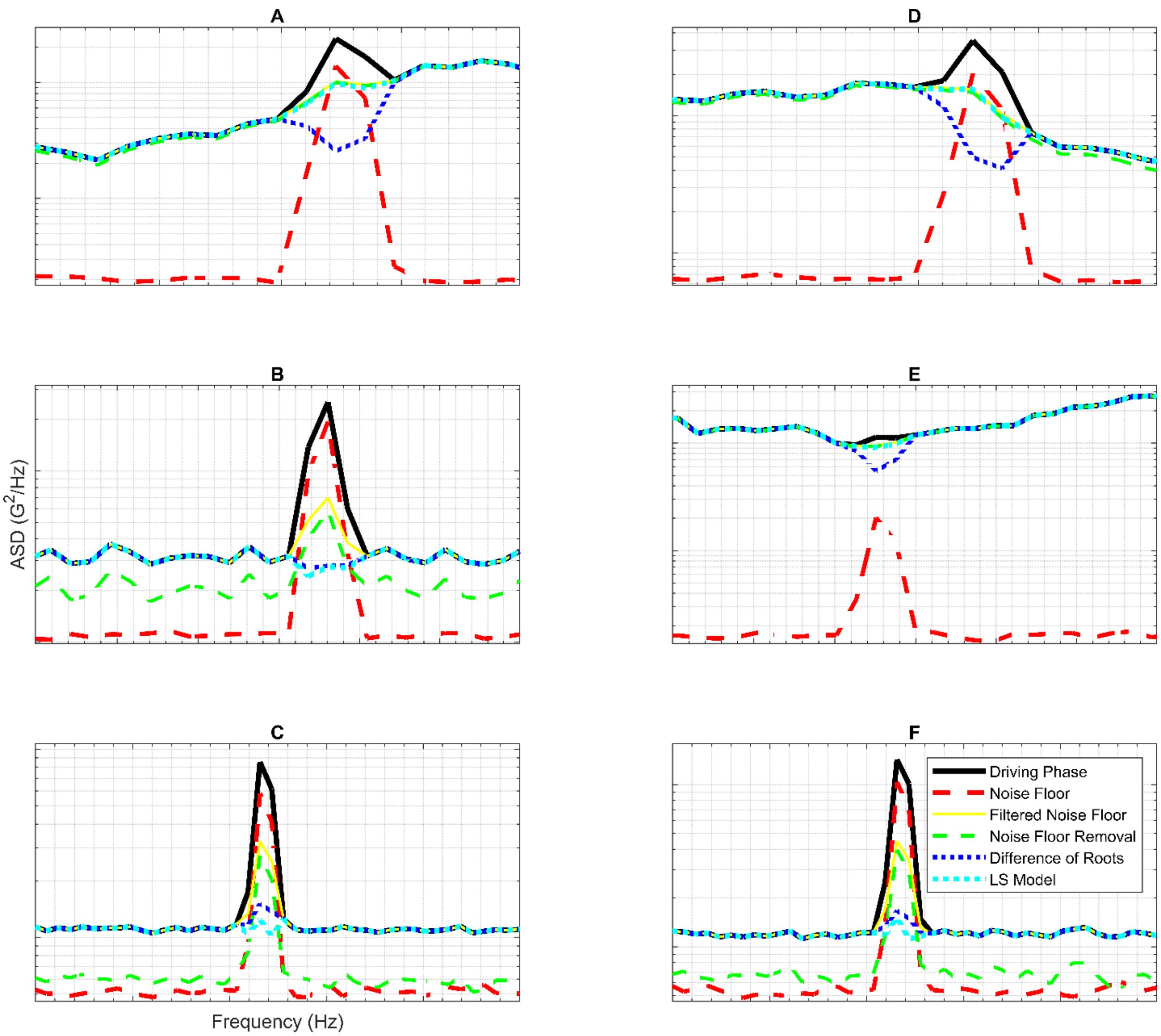

Figure 11: Spike-Remediation Comparison Windows from Spikes Identified in Figure 10

V. CONCLUSIONS

This paper presents frequency domain methods based on a Hampel Filter to determine the location and extent of artificial sine tones in the noise floor ASD and from there mitigate these sine tones in the ASDs of measurement phase data. Four approaches to removing the sine-tone artifacts in the Driving Phase of an OTR test are presented. The Noise Floor Removal method serves as the reference and is constituted simply by subtracting the measured noise floor from the Driving Phase ASDs across all frequencies. This method reduced the SNR by 1.0 and did little to diminish the relative sine-tone magnitudes. The Filtered Noise Floor method identifies and quantifies the noise artifacts in the noise floor with a Hampel Filter thereby only affecting the frequencies with the sine-tones. However, this method does leave some residual peaks. The Difference of Roots method works similarly to the Filtered Noise Floor but assumes the ASDs are combined as products of the square roots of the ASDs. This approach mitigates the sine-tones reasonably well at frequencies with low-magnitude spectral content but overcompensates at frequencies with higher magnitude spectral content, resulting in potential underestimation of ASDs where there is legitimate content. Finally, the Least Squares Model method creates a linear regression model of the sine-tone spikes in the Driving Phase ASDs based on those in the noise floor. This method performed well across a wide range of spectral magnitudes, although a small "jag" pattern tends to remain. Of the approaches illustrated in this paper, the Least Squares Model performed the best. While more reliable, this method is also dependent on more preprocessing to collect samples for generating the model.

These examples demonstrate the utility of applying a Hampel Filter in the frequency domain to identify sine-tone noise artifacts and mitigate their effects on measured ASD for that system. The results show that other Hampel Filter based methods could be developed for specific applications.

ACKNOWLEDGEMENTS

The authors are grateful for Mr. David Soine's insights about digital filters, Mr. Chad Heitman's suggestion to use the Hampel Filter on the OTR test data, and the helpful comments and technical review from their colleagues in the Sandia National Laboratories Environments Engineering Department.

This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan https://www.energy.gov/downloads/doe-public-access-plan.

REFERENCES

1. Kabe, A.M., and Sako, B.H., *Structural Dynamics – Fundamentals and Advanced Applications, Volume II,* Chapter 7, Academic Press, 2020.
2. Keller, T., "On the Use of Tracking Filters During Sine Vibration Testing, Sound and Vibration, January 2002.
3. Herlufsen, S., et al., "Characteristics of the Vold-Kalman Order Tracking Filter," Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, June 2000, Istanbul Turkey, DOI: 10.1109/ICASSP.2000.860254.
4. Verhaegen, M., and Verdult, V., *Filtering and System Identification – A Least Squares Approach,* Cambridge University Press, 2007.