University of Patras
Mechanical and Aeronautical Engineering

Laboratory of stochastics mechanical signals and automation

Stochastic signals and systems

Name: Stratos Gkoltsios
Number: 246525
Year: 5

Course Topic

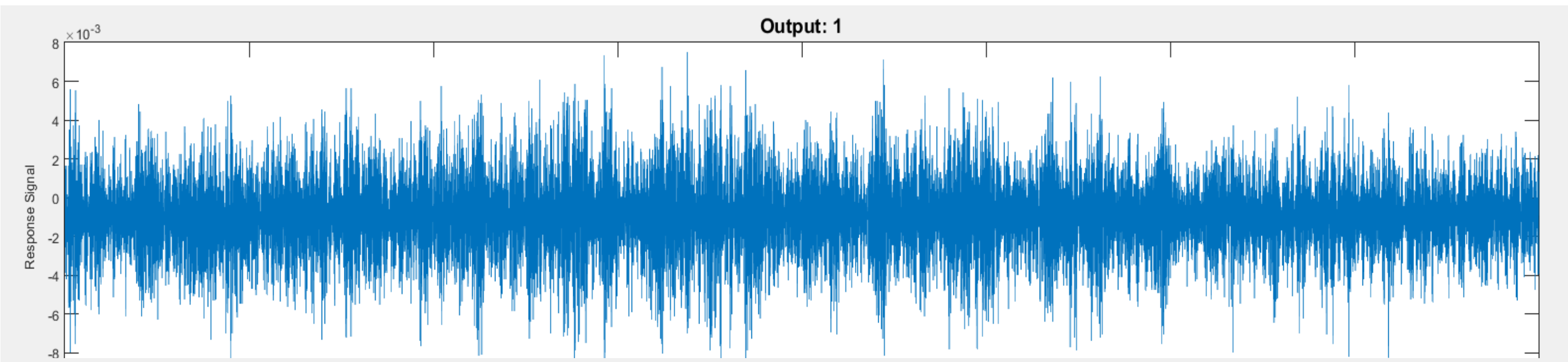
Aircraft Stabilizer 128Hz Experiment

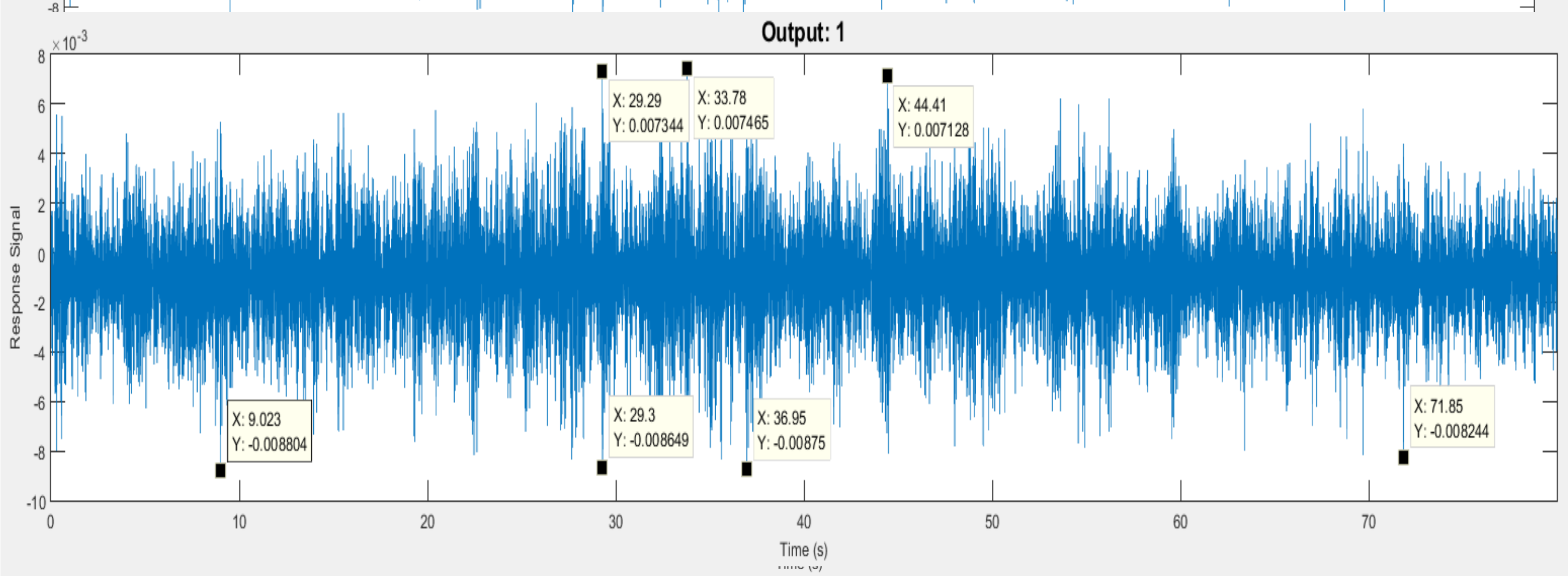
Image taken from: Random–vibration–based damage detection and precise localization on a lab–scale aircraft stabilizer structure via the Generalized Functional Model Based Method

Part A – Preliminary

A1. Preliminary

In the experiment, signals from the healthy state of an aircraft horizontal stabilizer were recorded. The stimulation was performed with a pseudorandom force and two acceleration responses were recorded at different points of the structure. The sampling frequency was 128 Hz and the length of each signal is N = 10240 samples. The signals have not undergone any pre-processing.



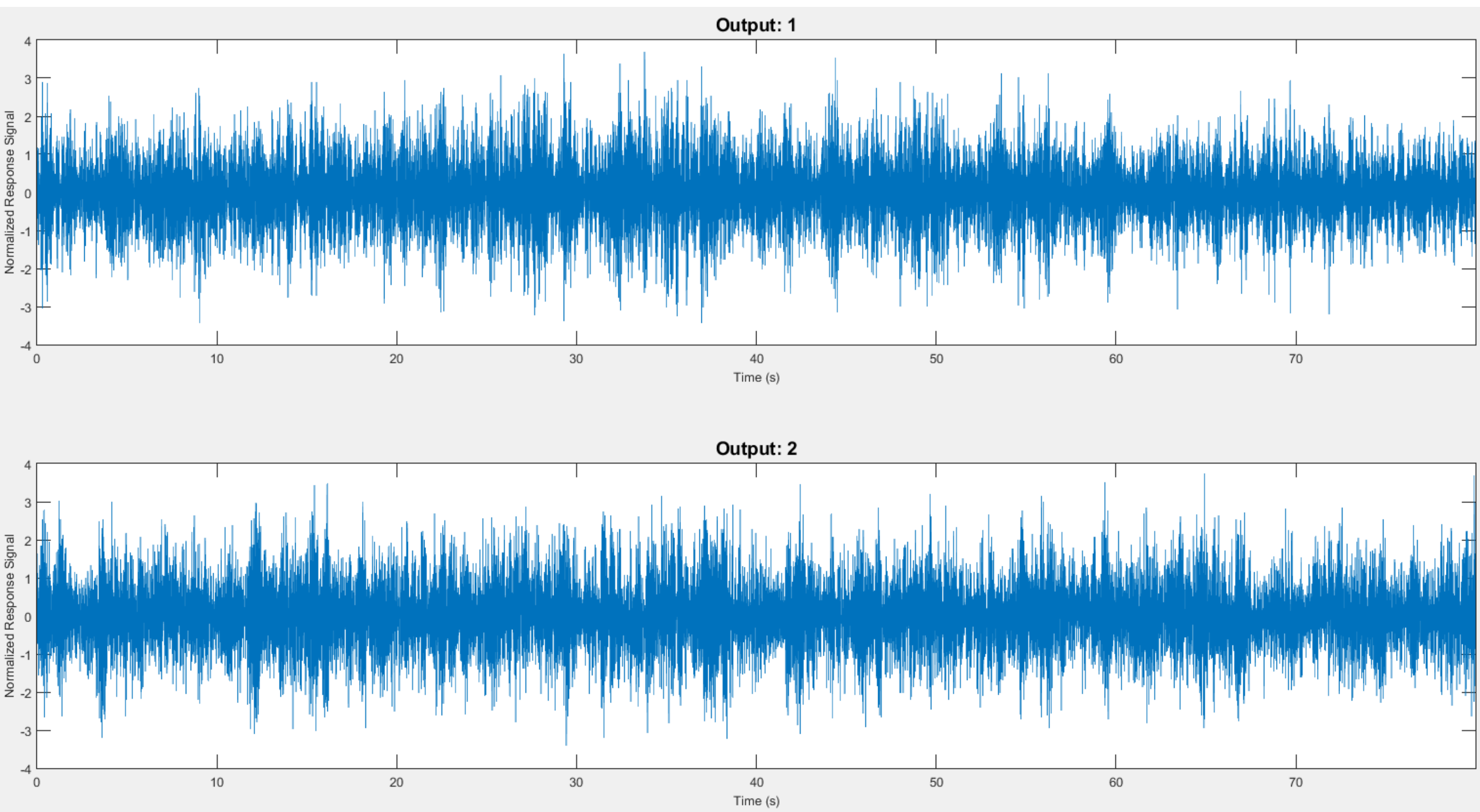


From the two systems we can observe that the median values are not zero. In the first signal the prices are moving around -1 and in the second around 1. It is clear that the median value remains steady in all the length of the signal and there is no steady fluctuation in the values relatively to time. That is a first sign of stationarity. Further below examining the first signal we can show some max values so it can be clear that there is no fluctuation in the max values. The biggest negative value is for t = 9.023s and the acceleration a = -8.8 x 10 $^{-3}$m/s^2 and the biggest positive for t= 33.78 the a = 7.47 x 10^{-3}m/s^2

A2. Normalization

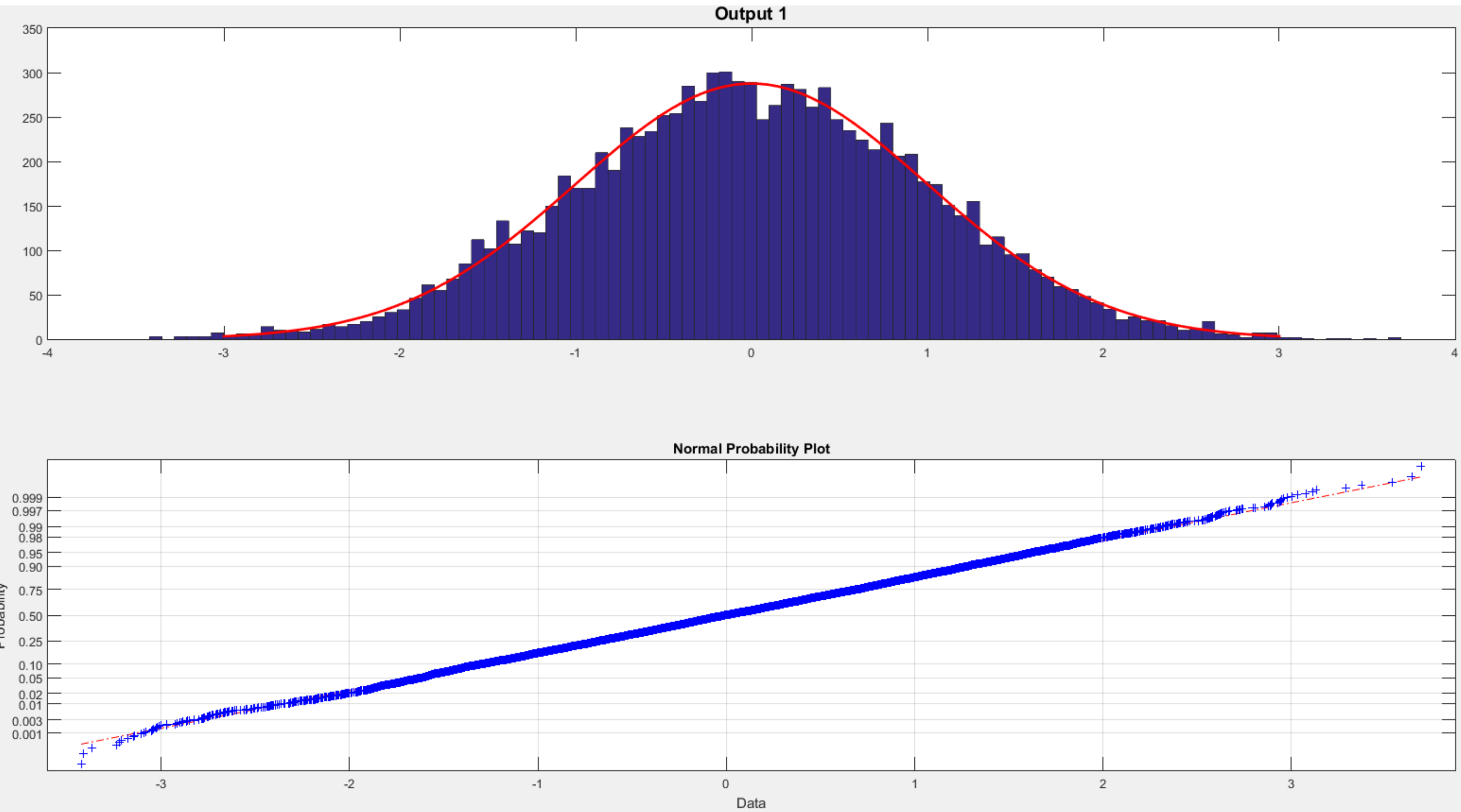
Since the signal appears stationary, it will be centered to remove its average value and move to zero. This will be done with the "detrend" command. Then for numerical reasons the signal will be normalized where the signal will be divided by its standard deviation. The standard deviation will be found with the "std" command.

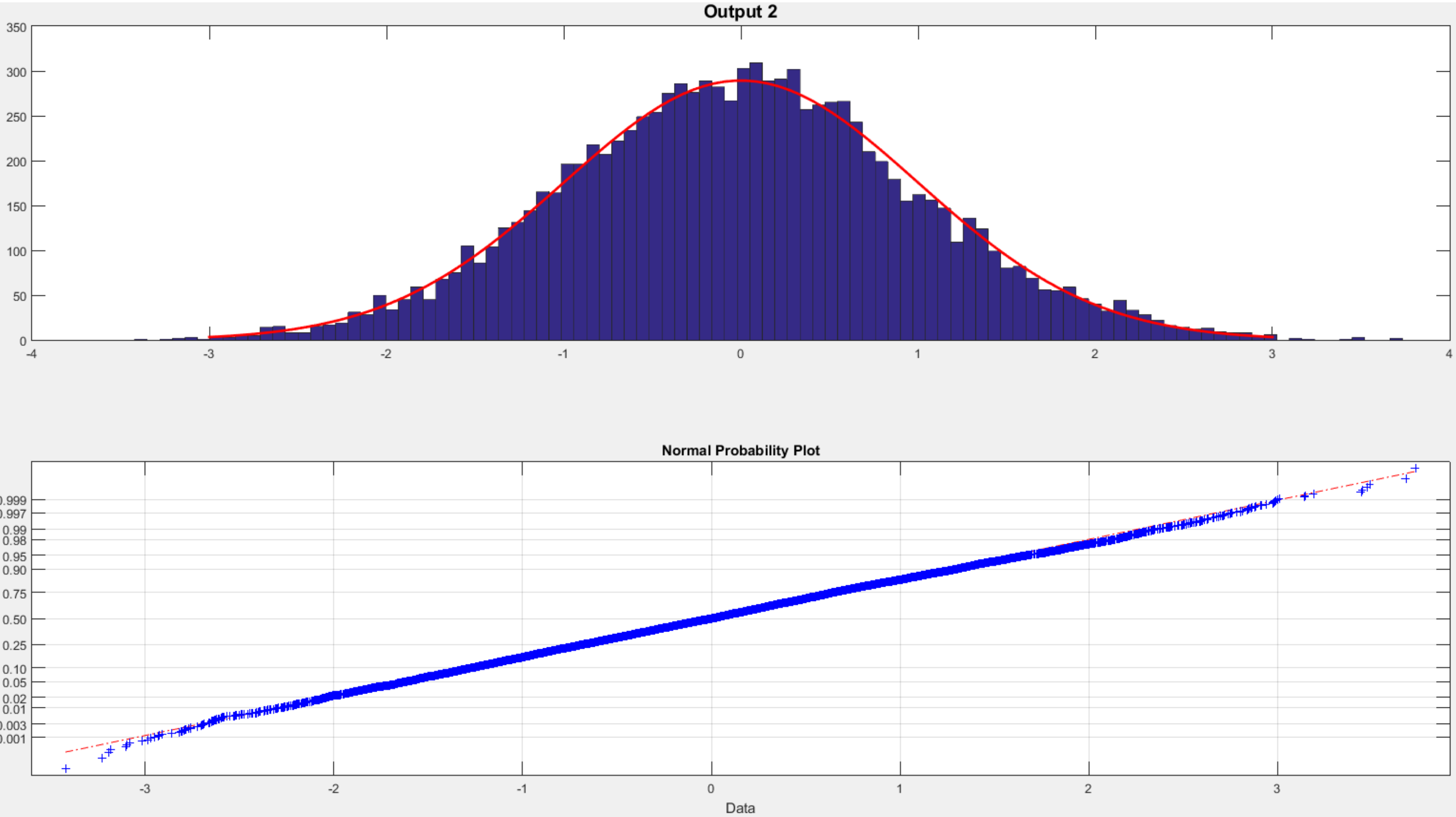
Here are the normalized signals:



A3. First analysis in the time field

To calculate the histogram, we divide the range of acceleration values into √N segments. The histogram serves as an empirical probability density function to examine the normality of the signal.
Another check for signal normality is the normal distribution plot. Below are these diagrams for each of the two acceleration signals:

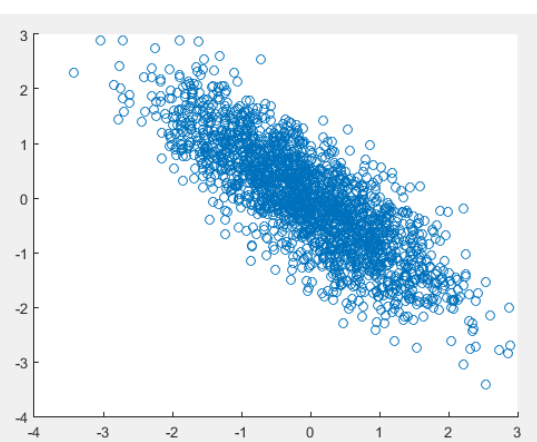


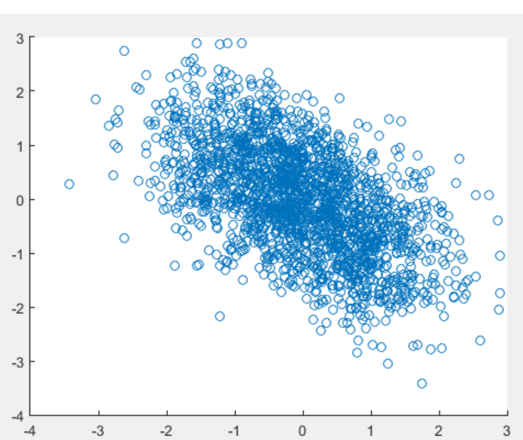


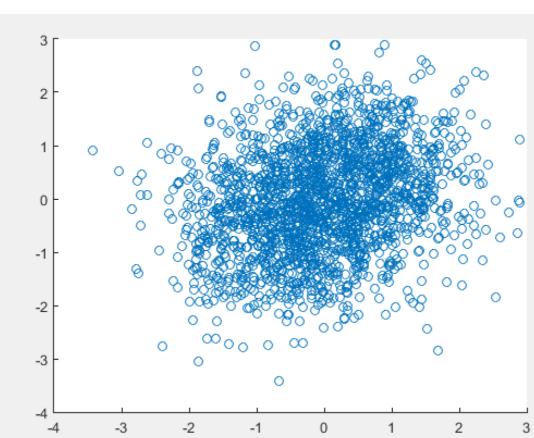
It is quite clear that the histogram is fairly close to the normal distribution in both signals, although there seems to be a slight skewness which seems to be opposite in one signal than the other. The "normal probability plot" graph is almost perfectly linear, which shows us that the signal follows a normal distribution.
Finally, we examine the skewness of the time series with the “skewness” command

Skewness (output 1) = -0.0181
Skewness (output 2) = 0.0667
As we see the skews are indeed very small, which is another element for the normality of the signal.

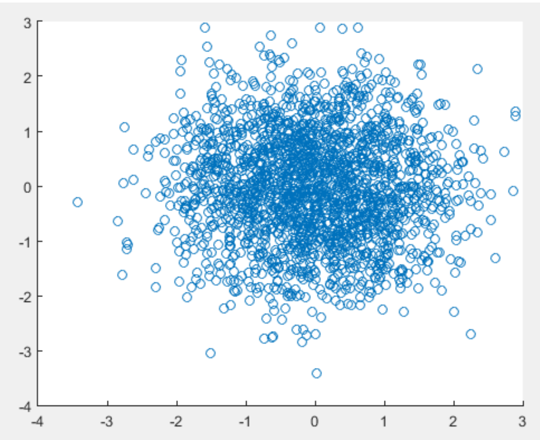
Lag = 1 Lag = 10 Lag = 20



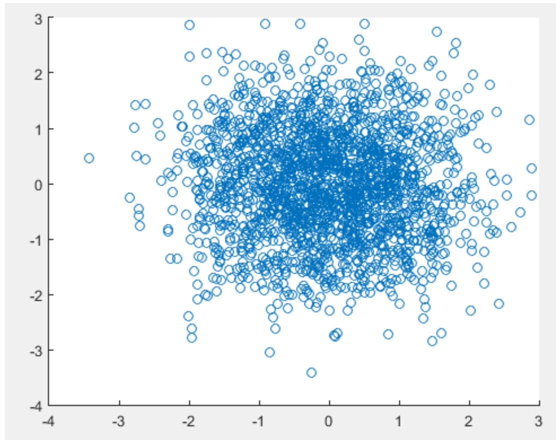




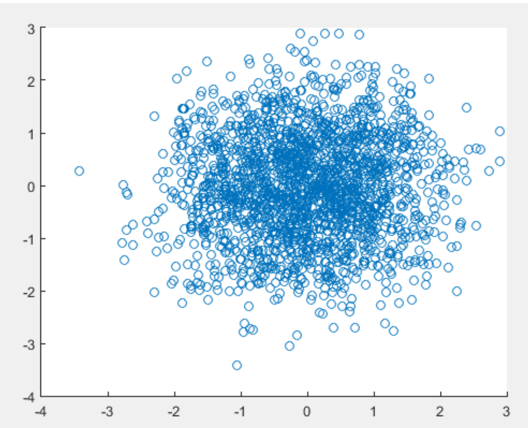
Lag = 50



Lag = 70



Lag = 100

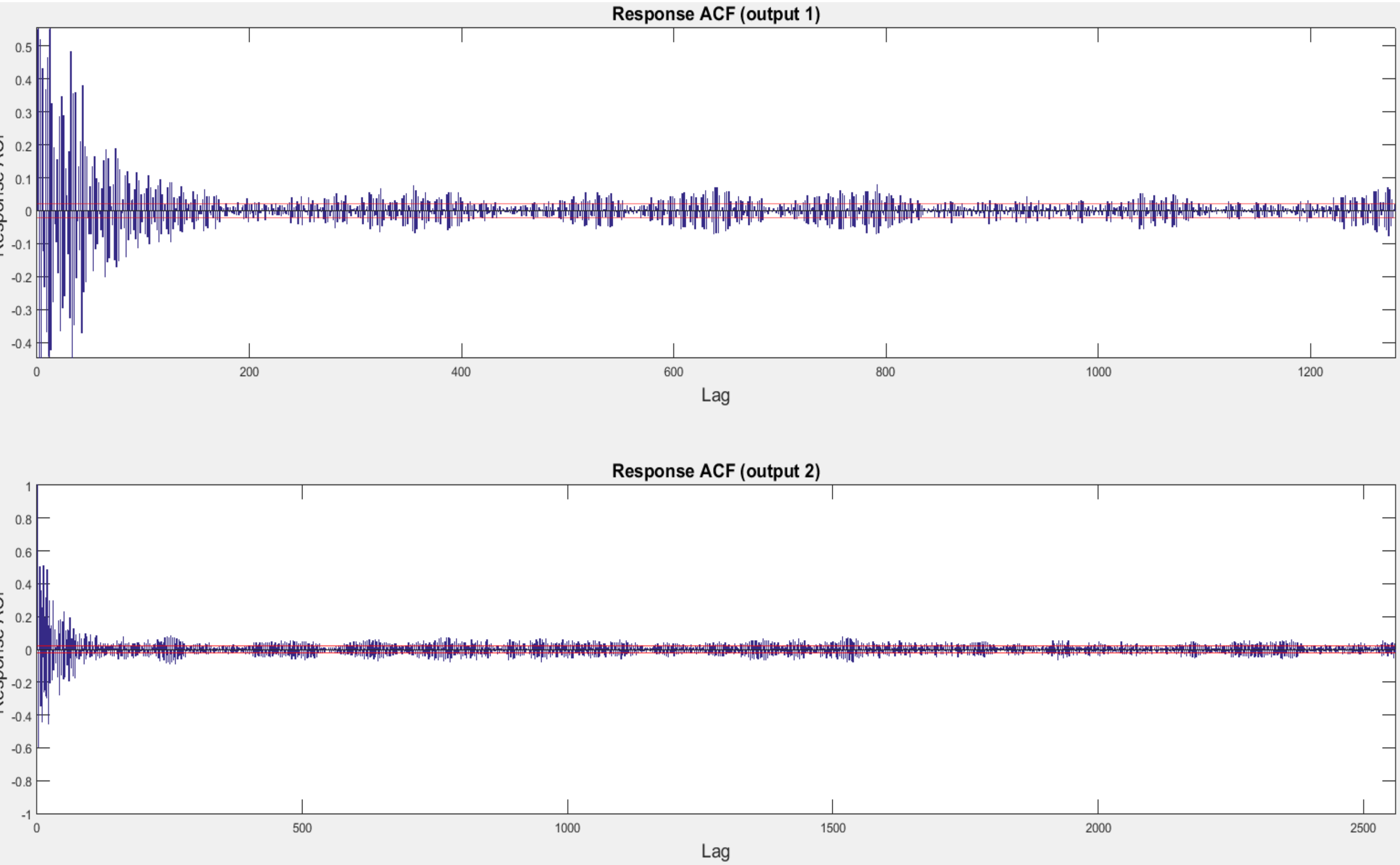


As can be seen from the scatter diagrams for different lags, the capacity has already started to be lost from lag 50 and from then we cannot distinguish with the eye a geometric deviation from the circle.

Section B - non-parametric analysis

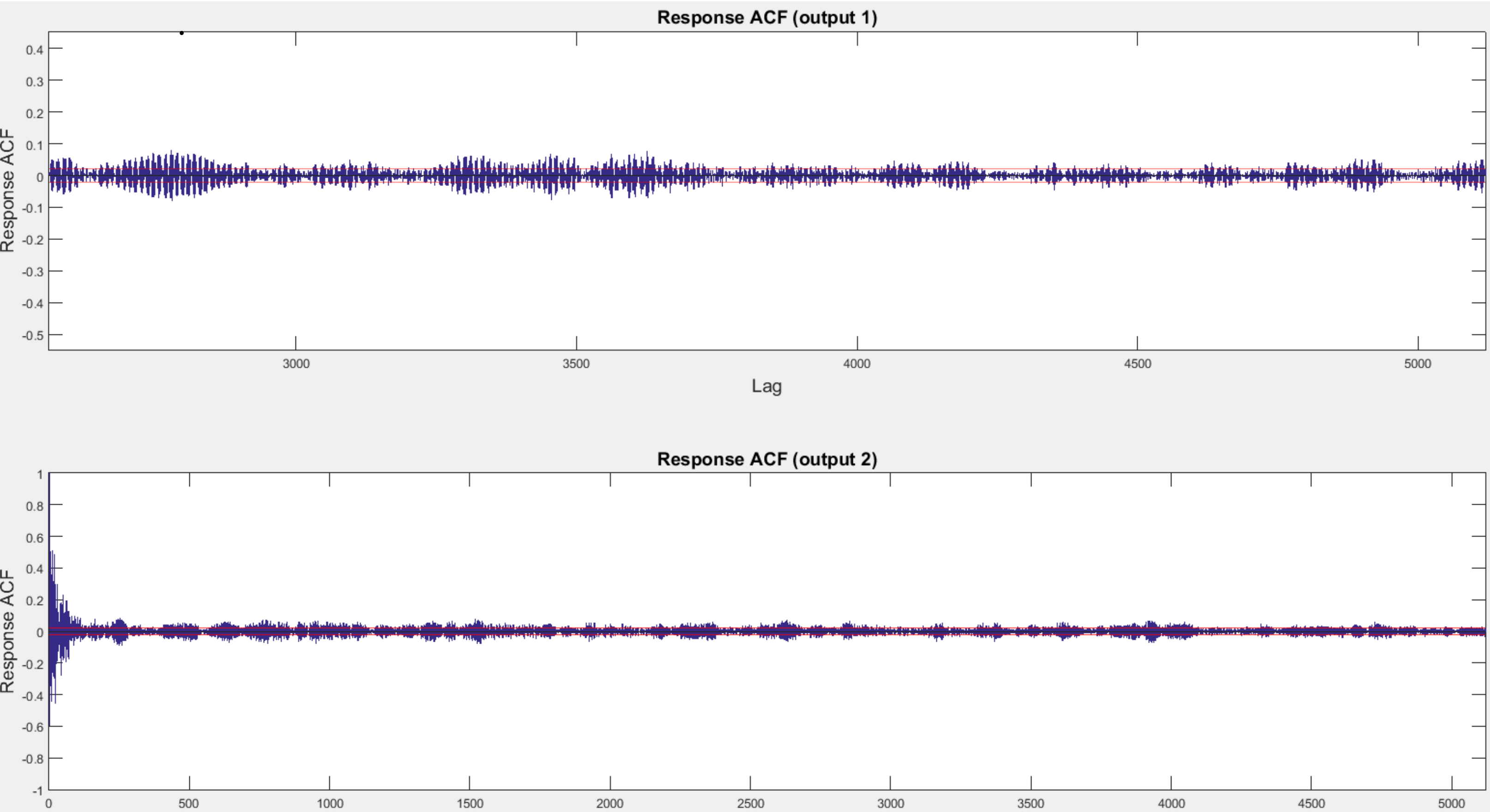
B1. Nonparametric analysis in the time domain

Then the estimation of the reduced autocovariance for Lag κmax < N/4 was examined, so that the estimator is as unbiased as possible... so for κmax = 2560 we have:



**A closer “zoom” was chosen in the first diagram so the values are displayed more clearly.

From the above graphs, it is evident that the autocorrelation of the signal never falls below the statistical significance limits ($1.96/\sqrt{N}$). In fact, while around Lag = 200 the autocorrelation seems to tend towards zero, it then rises again. This is unusual, as the autocorrelation should gradually decrease over time. In order to mitigate the risk of our estimator being biased, we also conduct a check for k_max = N/2 = 51

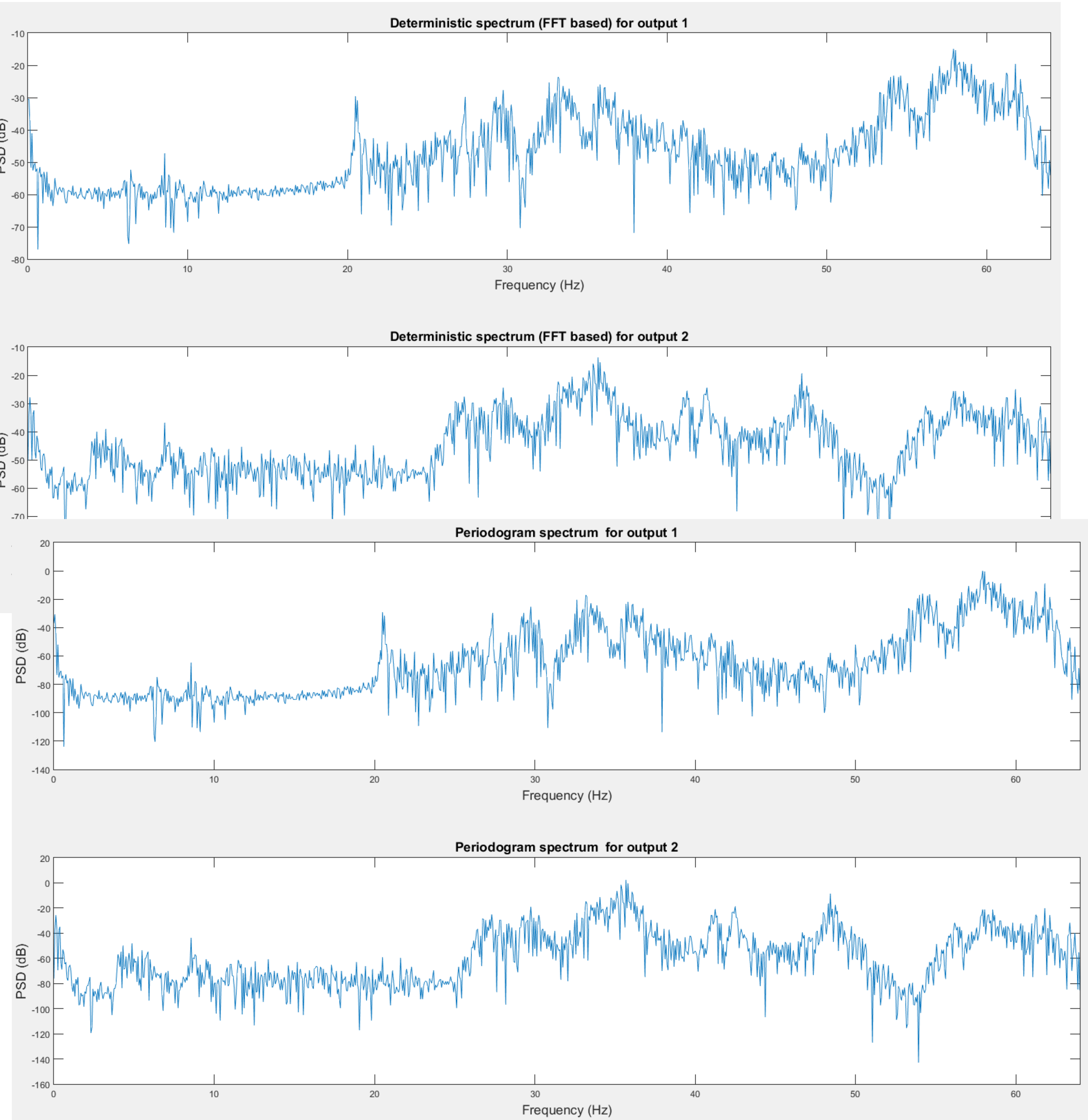


As can be seen, even for a very large Lag, several ACF values exceed the statistical limits, which shows that there is capacity in the system even after a long time.

As a conclusion, we could say that perhaps some poles of the system are very close to the limits of the unit cycle, as a result of which the time constants are very long and the capacity is not lost quickly.

B2. Nonparametric analysis in the frequency domain

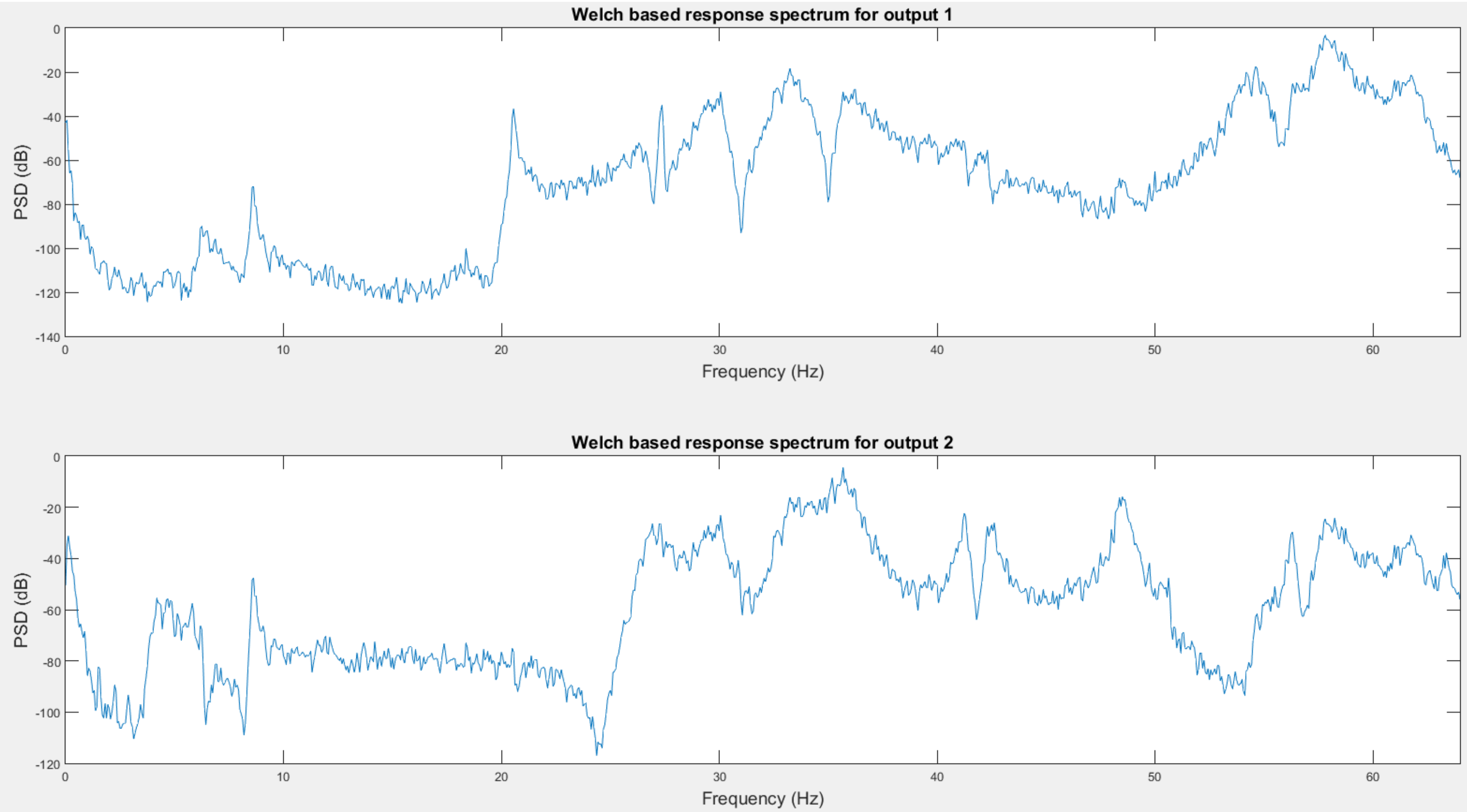
For the non-parametric analysis in the frequency domain the discrete Fourier transform (FFT) is used where it is an estimate of the spectrum of the signal. As a better estimate, the periodogram is also given through the "periodogram" command. For both estimations, 2048 samples of the signal were used, as for more, the estimations become very distorted since there is no consistency in the estimator.


Deterministic spectrum (FFT based) for output 1
PSD (dB)
Frequency (Hz)
Deterministic spectrum (FFT based) for output 2
PSD (dB)
Periodogram spectrum for output 1
PSD (dB)
Frequency (Hz)
Periodogram spectrum for output 2
PSD (dB)
Frequency (Hz)

Observing the diagrams we can see in principle that the improvement of the periodogram by the FFT is present but quite small. There is still a lot of curl, so it's not considered a good enough estimate of the spectrum. Also, although we can distinguish some main frequencies, it would be better to examine them with a better spectrum estimator.

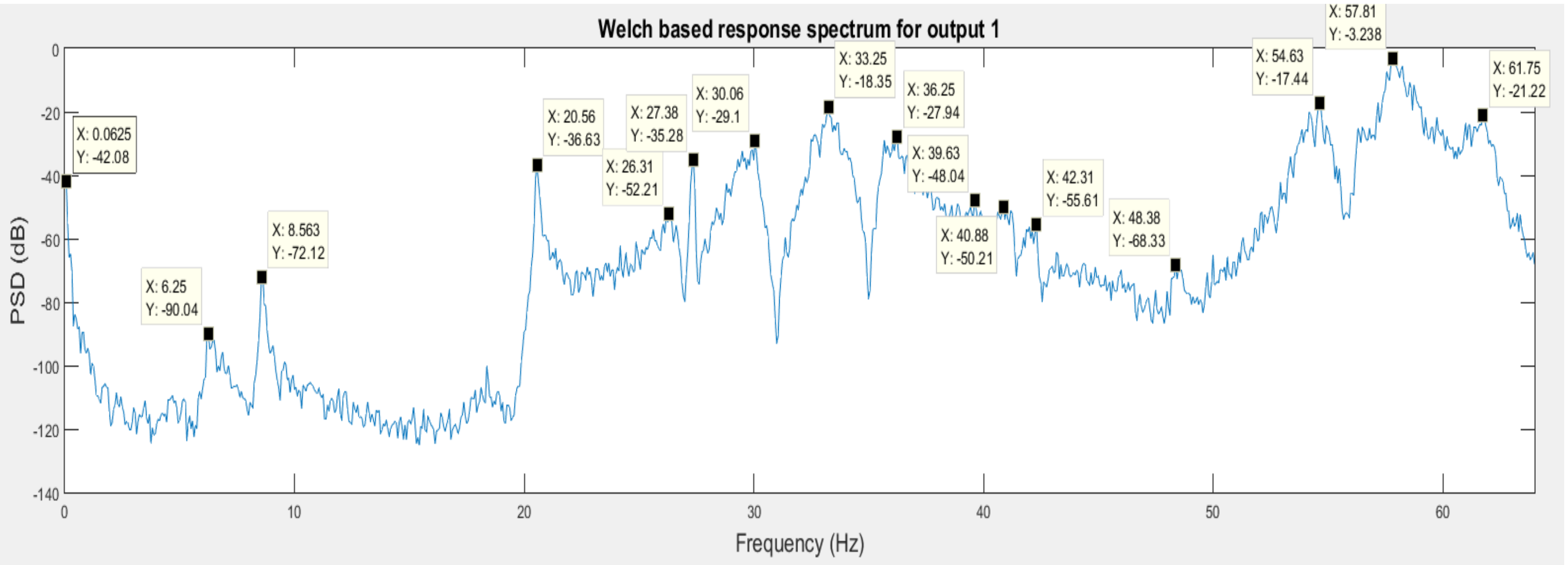
The resolution (δf) of both diagrams is the same $\delta f = Fs/L = 128/2048 = 0.0625$ Hz or $\delta\omega = 2\pi Fs/L = 0.393$ rad/s

In order to have a comparison between the methods, we will estimate the spectrum using the Welch method (pwelch) with a window size of L = 2048 samples and an overlap rate of 75%.



****The resolution is the same as above. Here, it is clear that the estimator is much better, and the spectrum is much more clearly distinguishable. Additionally, we can now much more easily identify the main frequencies as well as the system's zeros.

For a better visualization of the results, we will consider only the 1st signal.



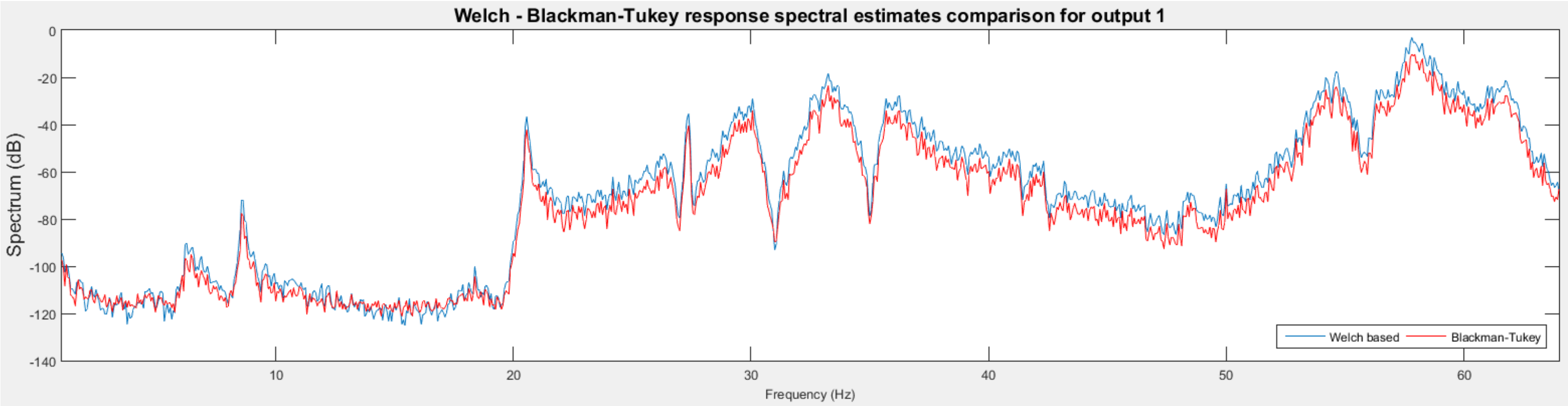
By examining the main frequencies, we can make the following observations:

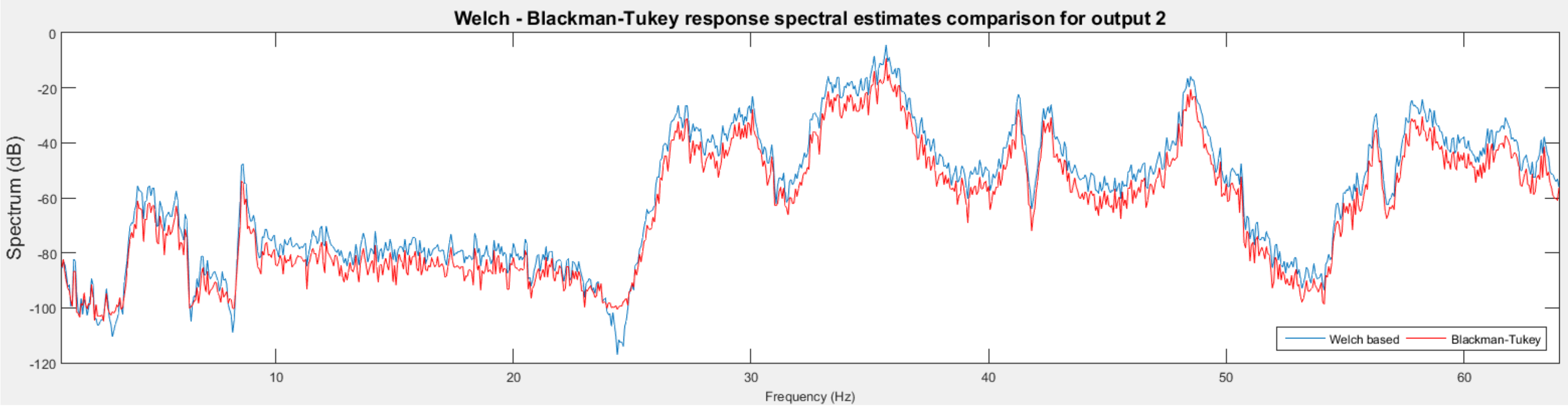
Firstly, there is an increase in the spectrum (peak) at a very low frequency. This may indicate a lack of stationarity or a non-zero mean.

We can also observe that the system exhibits several high frequencies. This could be due to the fact that it’s an aircraft stabilizer, which implies a relatively lightweight construction.

Finally, although the frequencies displayed are the most obvious ones, we could consider other peaks as potential main frequencies. Perhaps with a slight improvement in spectral estimation, we could confirm certain frequencies more accurately.

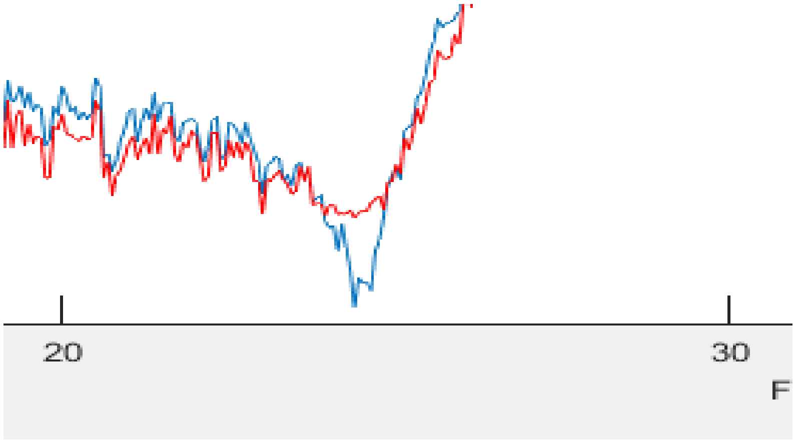
Next, we compare the Welch method calculated above with the Blackman-Tukey method.



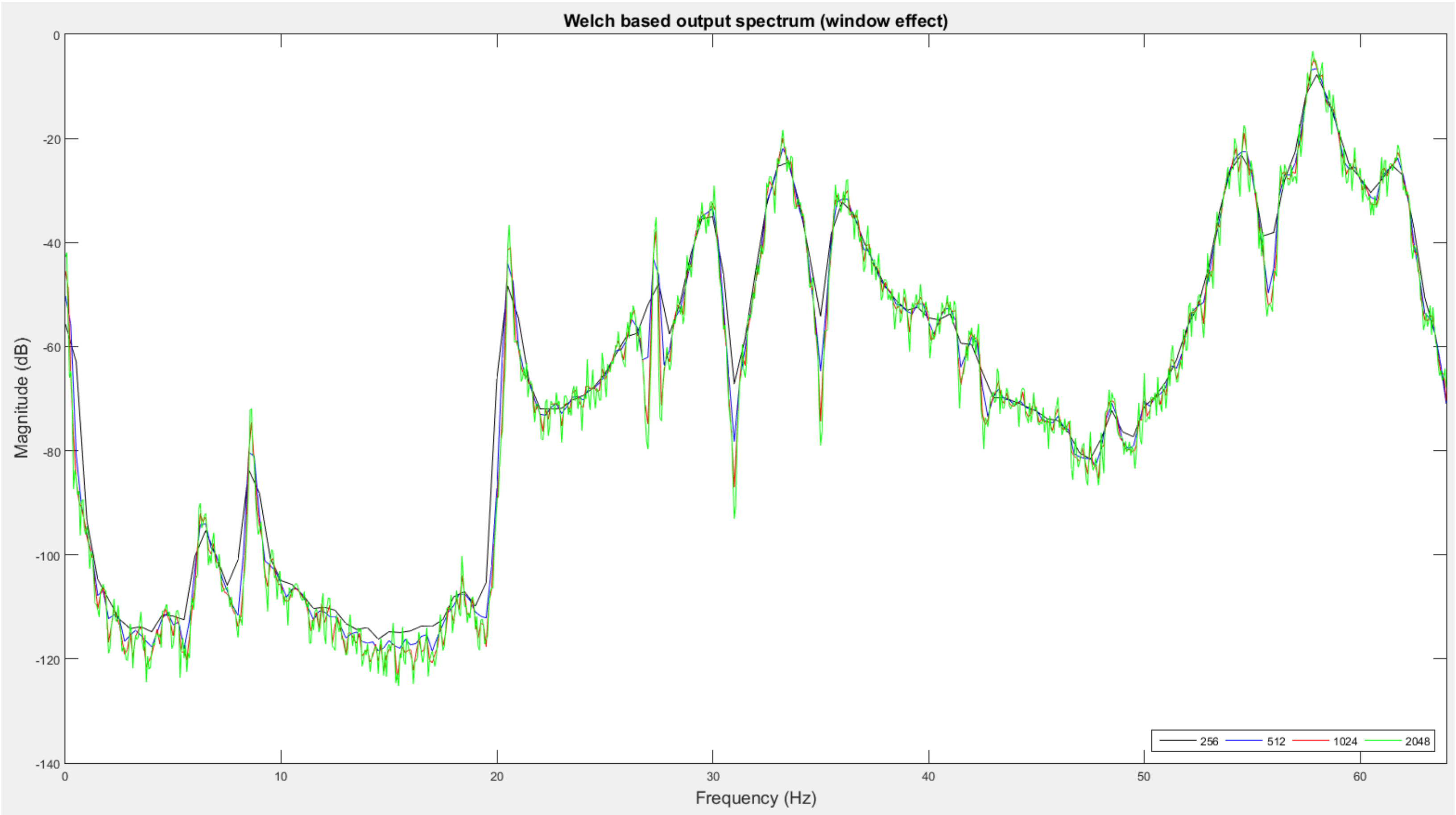


As is evident, the two methods show great similarity in their accuracy. The Welch method appears to display the values at a larger scale, but this is not significant since the same eigenfrequencies appear to be depicted by both methods. A noticeable difference is in the second signal, where the Welch method seems to detect a zero with greater accuracy.

.



Finally, we compare data windows of different sizes as well as different overlap percentages.



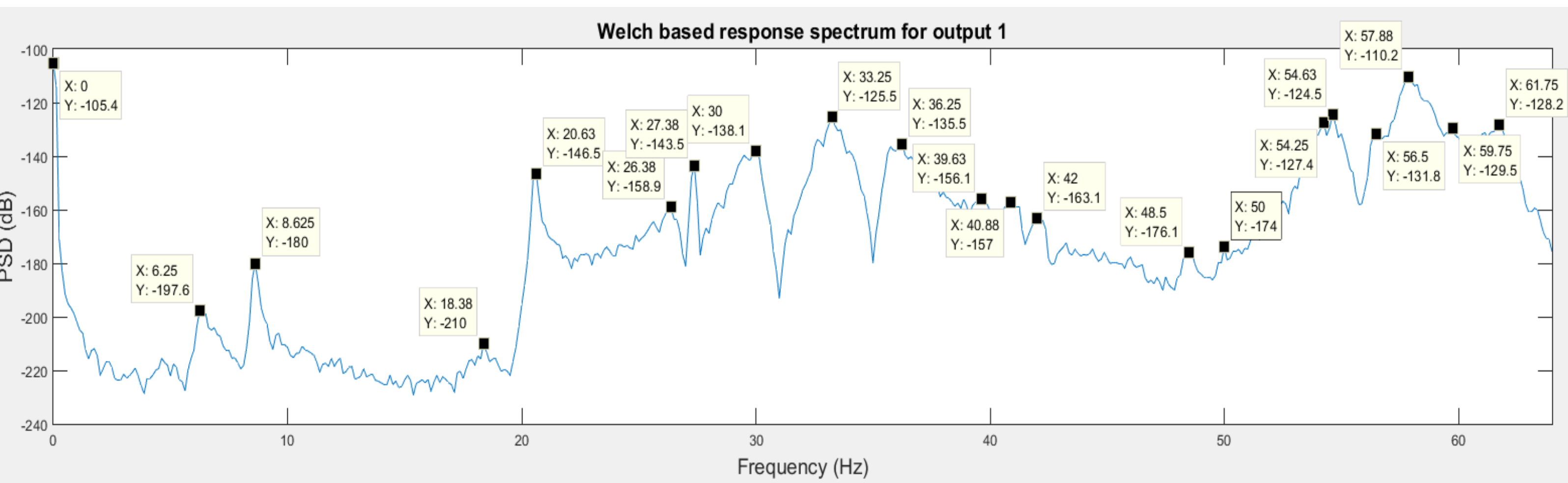
The resolutions for each window are as follows:

L = 256 : δf = 0.5 Hz
L = 512 : δf = 0.25 Hz
L = 1024 : δf = 0.125 Hz
L = 2048 : δf = 0.0625 Hz

Resolution essentially indicates the step by which the frequencies of the spectral estimation change. If it is too large, there is a risk of missing certain eigenfrequencies if they fall between two successive resolution steps. This would cause the main frequency information to be split between its two neighboring frequencies. Therefore, it is clear that we want the highest possible resolution.

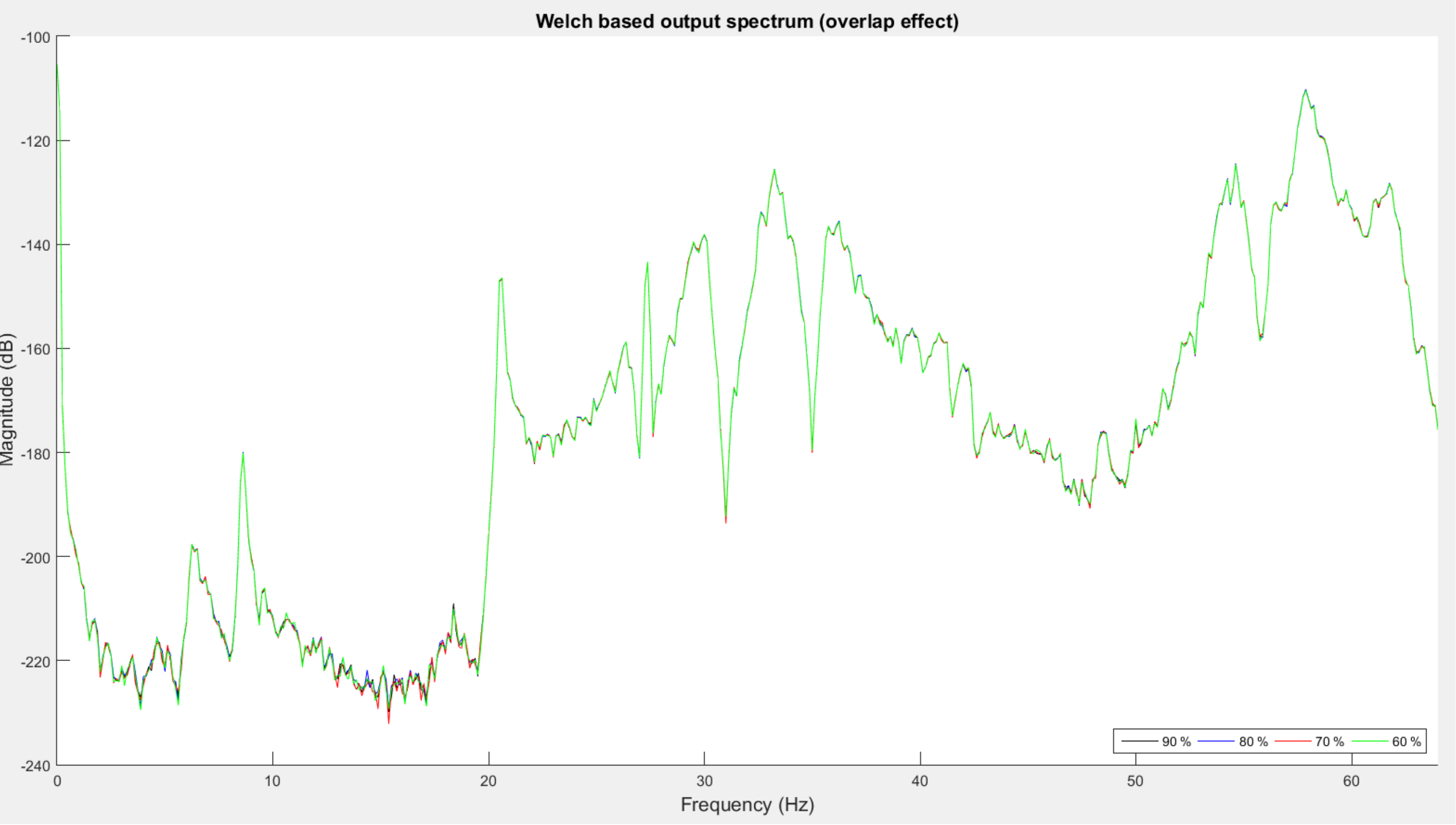
However, this results in an increase in the spectrum's variance. Thus, using the “windowing” technique, we select the appropriate resolution.

In this case, we see that even with a window of L = 1024 and δf = 0.125Hz no eigenfrequencies seem to be lost, while the spectrum becomes much clearer. Therefore, this will be selected as the optimal data window for re-evaluating the main frequencies.



By examining the eigenfrequencies presented here, we can infer more frequencies than with the estimation for a window of L = 2048, as the stochasticity has decreased significantly. To be certain, however, we will examine the spectrum using parameterization with an ARMA model. For accuracy, since we can clearly distinguish 10 eigenfrequencies, we will begin modeling with ARMA(20,20) or higher. It is also quite clear here that the diagram takes a high value at zero frequency, which may suggest a lack of stationarity.

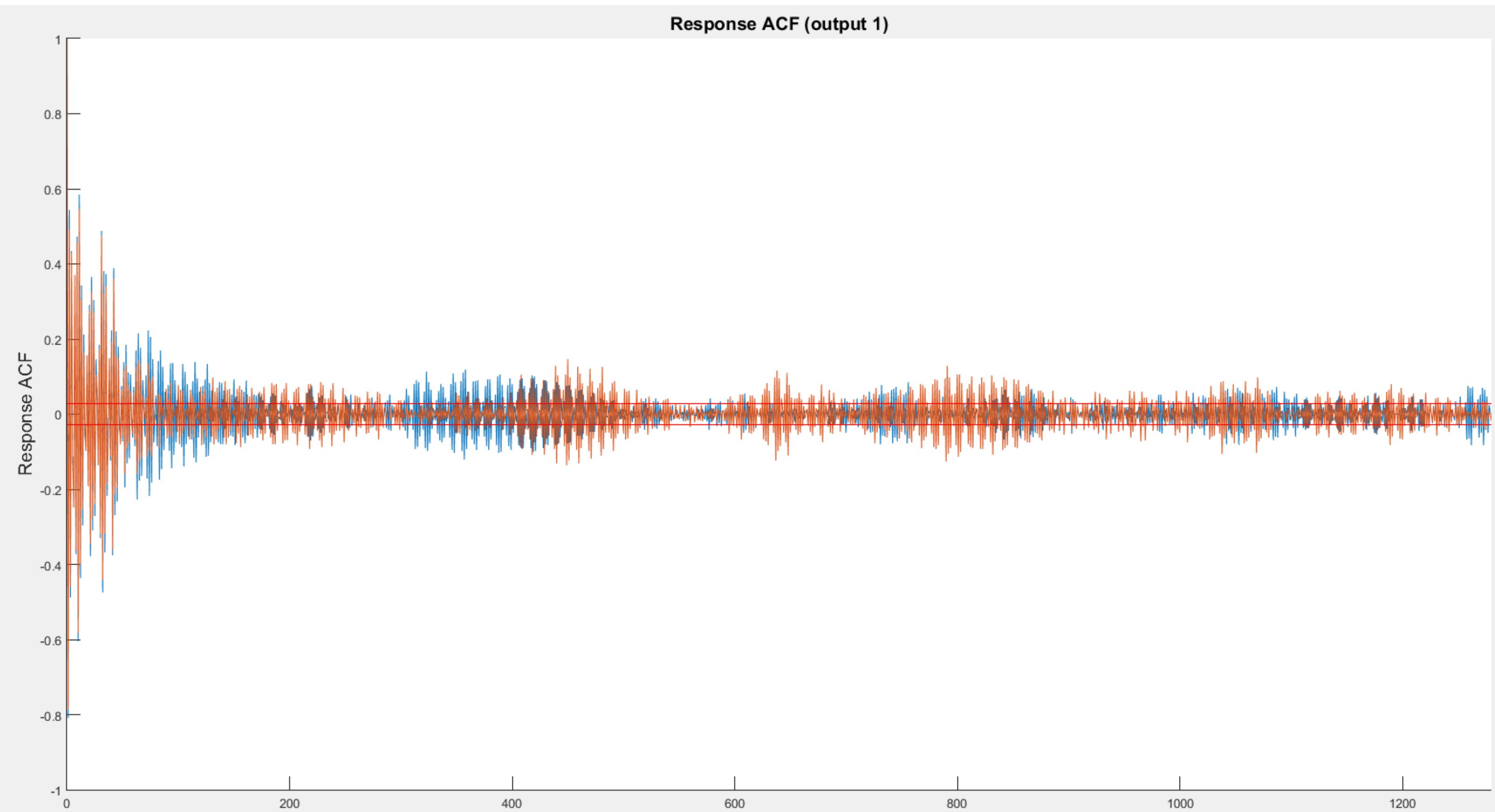
**Finally, we assess the quality of the spectral estimation for different overlap percentages:



The diagram was created with a data window of L = 1024 samples. The difference appears to be minimal.

B3. Stationarity test

We split the first signal into two equal parts with N = 5120 samples each. Then, we examine the normalized autocovariance in the same diagram. The test is conducted up to k_{max}= N/4 = 1280 samples.



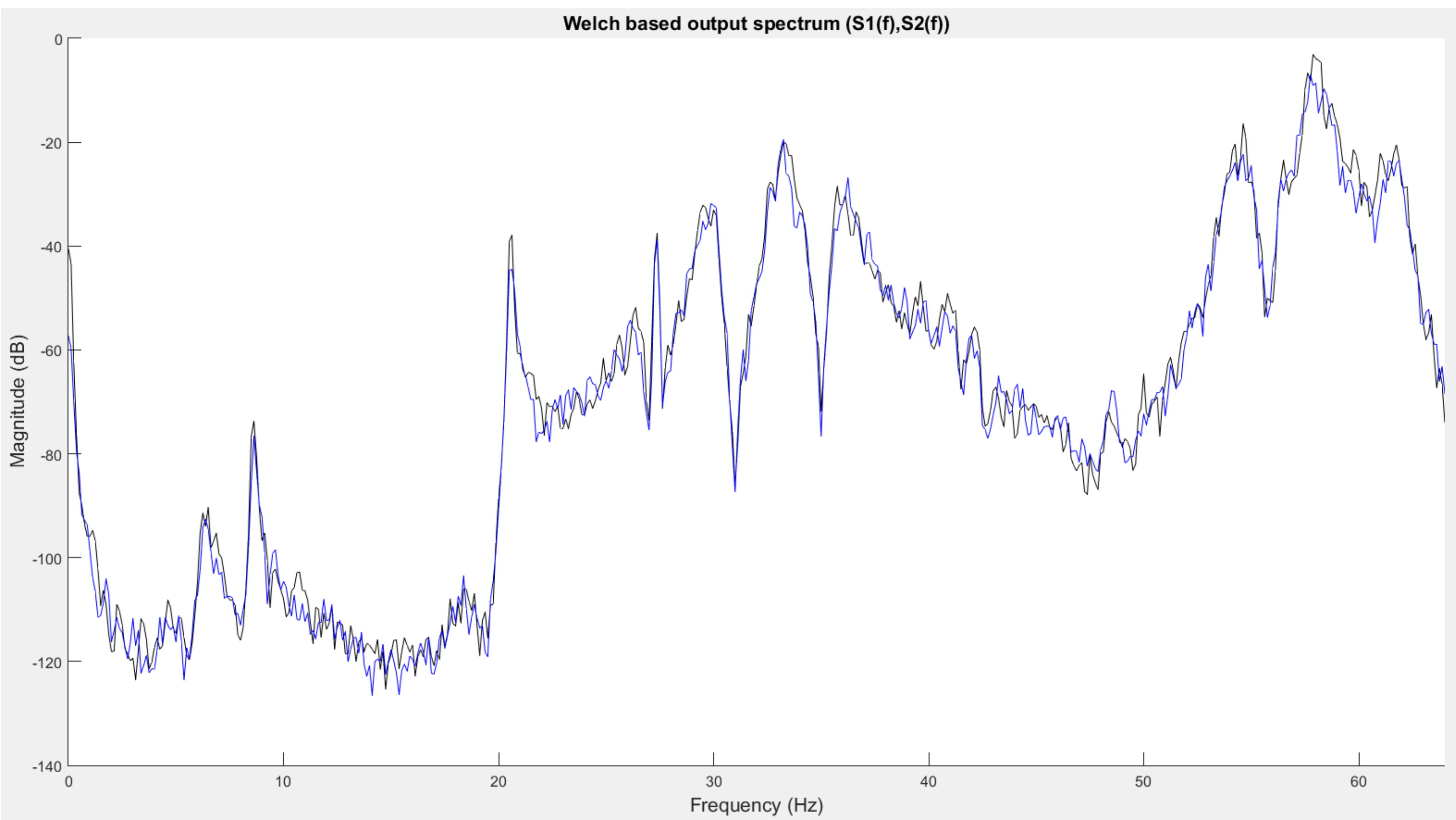
It is now quite clear that the signal is non-stationary, as the autocovariance depends not only on the lag but also on time. Additionally, it appears that the autocovariance differs even for small lags (from lag 70 and above, the difference becomes evident).

Hypothesizing about the cause of the signal's non-stationarity, one could say it may be due to a change in the direction of the stabilizer fin during the experiment. This would lead to a different geometric mass distribution, resulting in changes in the body's response. Another explanation could be a malfunction in the structure.

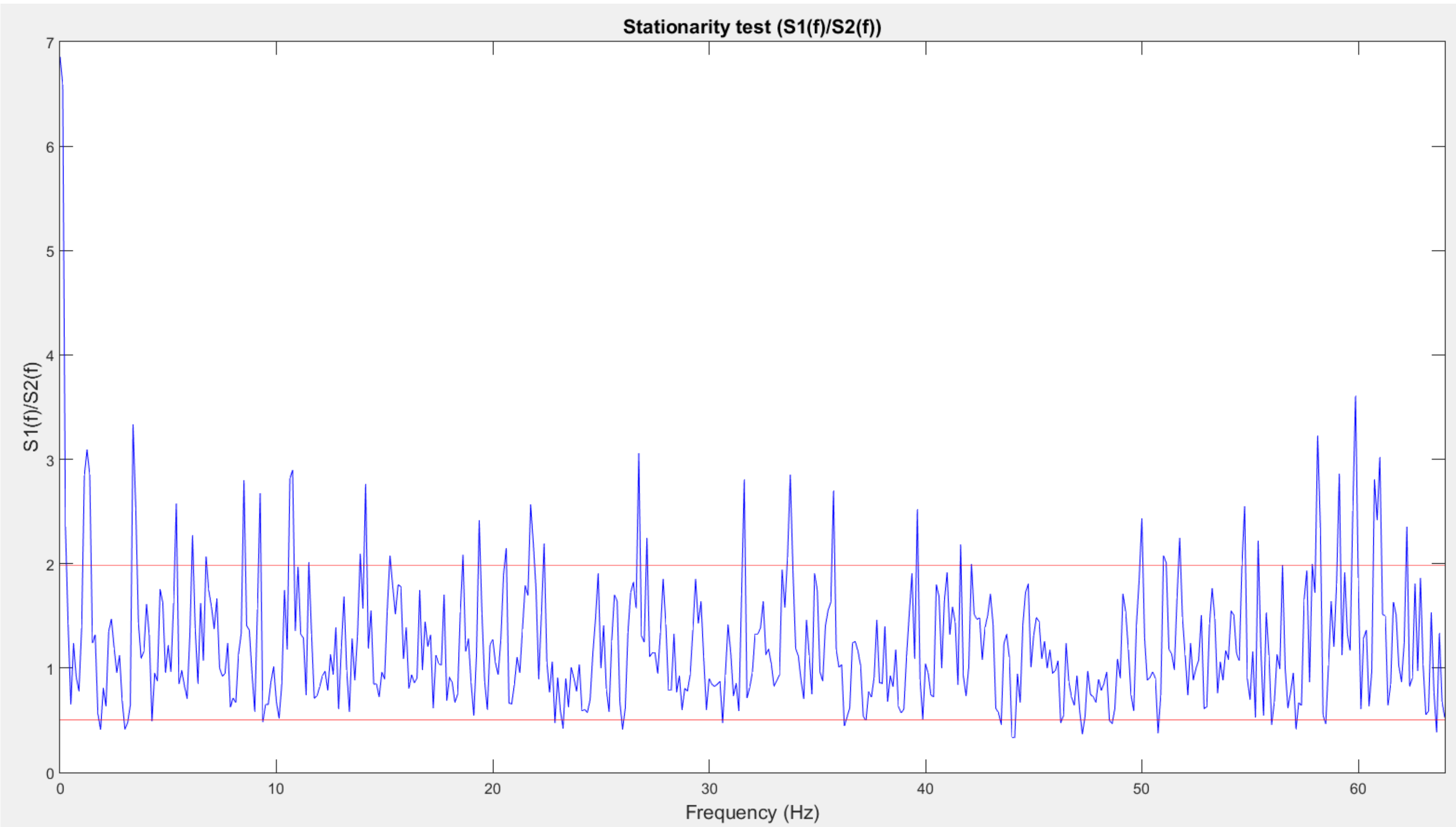
For the stationarity test via Welch's spectrum estimation, we divide the signal into two equal parts N = 5120 and use a window of L = 1024 samples. With an overlap percentage of 75%, we advance by D = 256 samples. Thus, our data is divided into $\kappa = (N - L) / D + 1 = 17$ segments.

To consider the signal as stationary, 95% of the values in the ratio $\frac{S1(f)}{S2(f)}$ diagram should fall within the significance limits of the F(2k,2k) distribution, where 2k is the degrees of freedom of the distribution. For a = 0.05, the statistical significance limits for the F distribution are as follows:

$P(F \leq f_1) = \alpha/2 = 0.025 \rightarrow \quad f_1 = 0.5$
$P(F \leq f_2) = 1 - \alpha/2 = 0.975 \rightarrow \quad f_2 = 1.98$



From the graph of the two estimated spectra (for data ranges [1, 5120] and [5121, 10240]), we observe minor differences where there should be none if the signal were stationary, as the spectrum should not depend on time. Subsequently, the spectra are divided for confirmation.



It is quite evident that the values exceeding the significance limits of the F distribution are well above 5%. Therefore, as confirmed above, we cannot consider our signal stationary. Nevertheless, within the scope of this course, we will attempt to model it parametrically with ARMA models, even though this is not optimal.

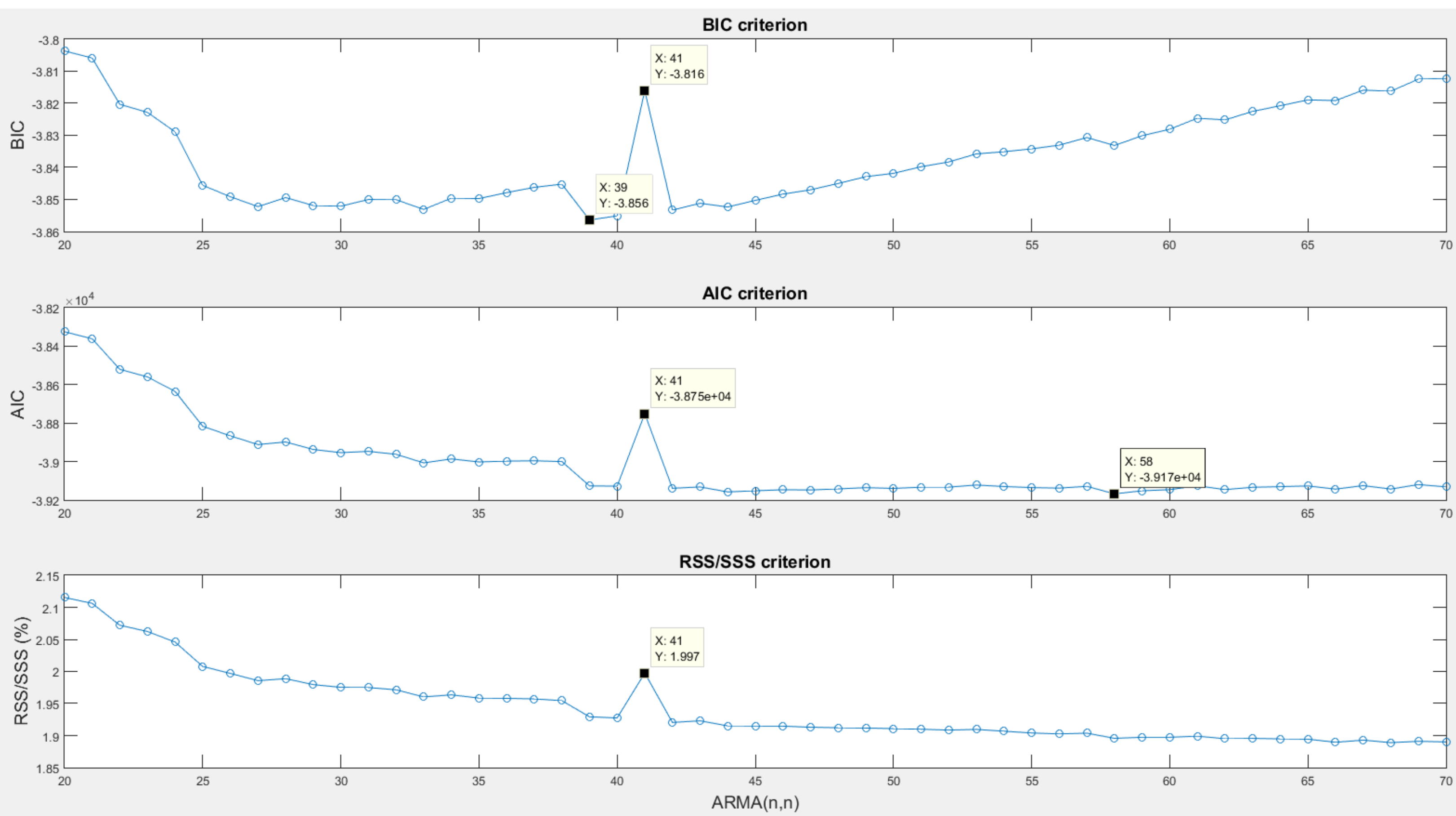
Section C – parametric analysis ARMA

C1. Preliminary

Initially, we will divide the signal into two segments. The first, which will be the larger segment, will be the estimation segment (N_e=10000 samples), and the second will be the validation segment (N_v=240 samples).

C2. ARMA model estimation

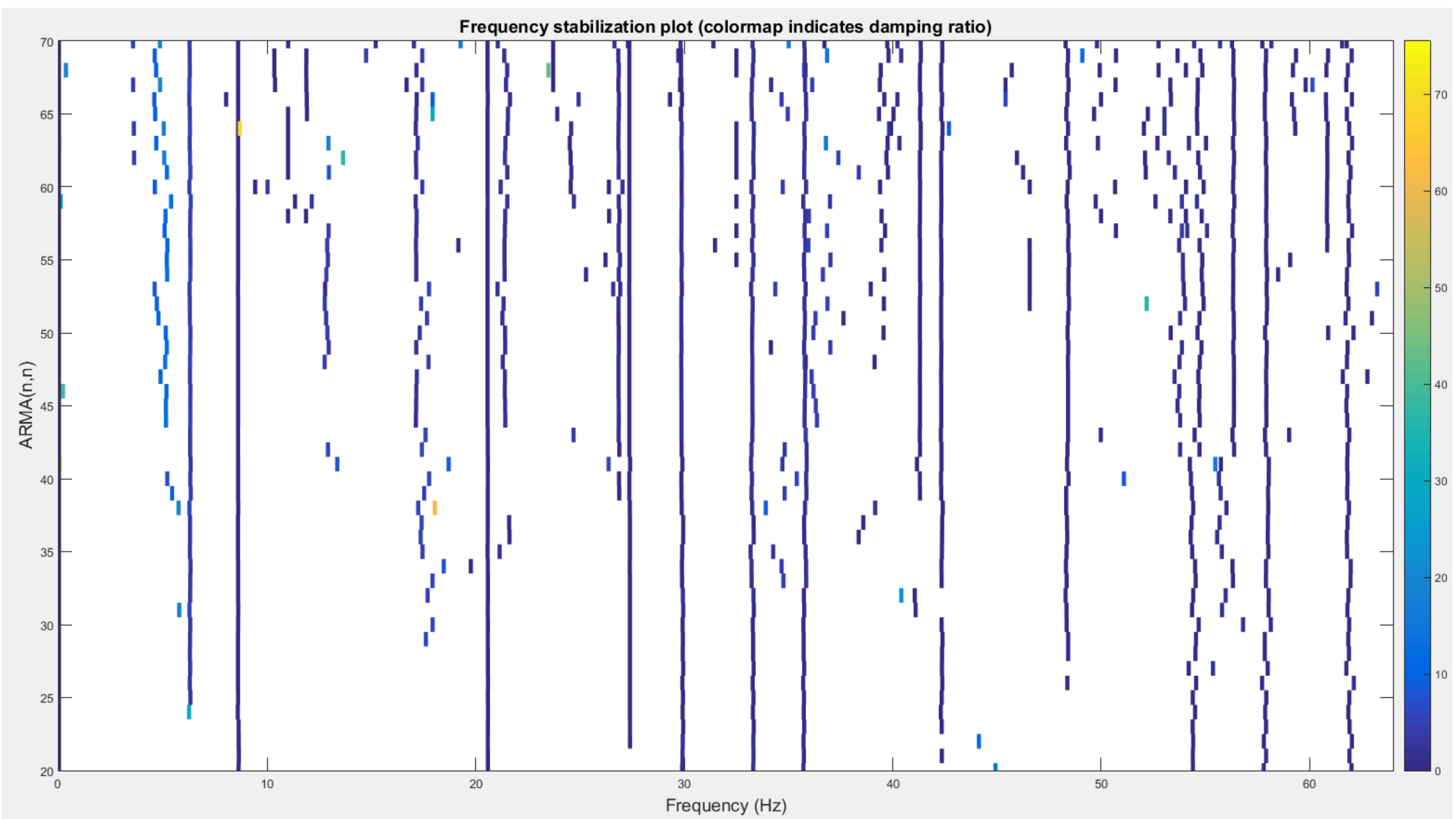
Using the estimation segment, we evaluate ARMA(K,K) models through Akaike and Bayesian criteria, as well as the RSS/SSS ratio (Residual Sum of Squares divided by Series Sum of Squares). Our evaluation starts at K=20 since we estimated ten apparent eigenfrequencies in the Welch spectrum, and continues up to K=80



The criteria indicate a sharp rise at K=41, which should not occur and may be a sign of numerical issues. We also observe fluctuations up to K=39 in the BIC criterion. The optimal order of the model, based on the BIC criterion, is chosen as K=39, where the lowest value is found. For K=39, we have SPP = 10000/39 = 256.4 >> 15

Finally, from the RSS/SSS criterion plot, we can conclude that our signal, though non-stationary, is quite deterministic, meaning it has strong dynamics. This is evident from the low values in the plot.

Next, we perform a check to identify the system's dominant frequencies using the frequency stabilization diagram. This allows us to examine which natural frequencies consistently appear for each change in the model's K parameters



As expected, with the increase in the model order, more frequencies appear. However, many of these only appear temporarily for a particular AR order and then do not reappear. From the graph, we can easily discern that certain natural frequencies consistently appear for each K, while others only for specific model order values. These pseudo-frequencies may be due to measurement noise or the non-absolute whiteness of the input signal. They could also be attributed to the lack of stationarity in the signal.

Another piece of information provided by the graph is that the damping factor is extremely low ($\zeta<0.1$) for almost all frequencies, which is expected given that we are dealing with a metallic structure.

Selecting a model order of K=39, we calculate the system's natural frequencies and damping ratios and then compare them with those in the frequency stabilization diagram.

Natural Frequencies (Hz)	Damping Ratios (%)
5.4288	10.5132
6.3230	4.6874
8.5959	0.6548
17.5236	4.6657
20.5753	0.2409
26.8797	1.1469
27.3662	0.1094
29.8677	1.1448
33.2264	1.4336
34.8163	3.5531
35.8611	1.2351
41.3080	0.2350
42.3289	0.3673
48.3249	0.3841
54.3630	0.9017
55.7327	1.2241
58.0081	0.8299
61.8055	0.6551

***In bold and underlined are the 'true' natural frequencies that clearly appear for each value of K.**

In bold but not underlined are the frequencies that do not appear for all model orders but seem to converge as the order increases, so we can also consider these as 'true' natural frequencies.

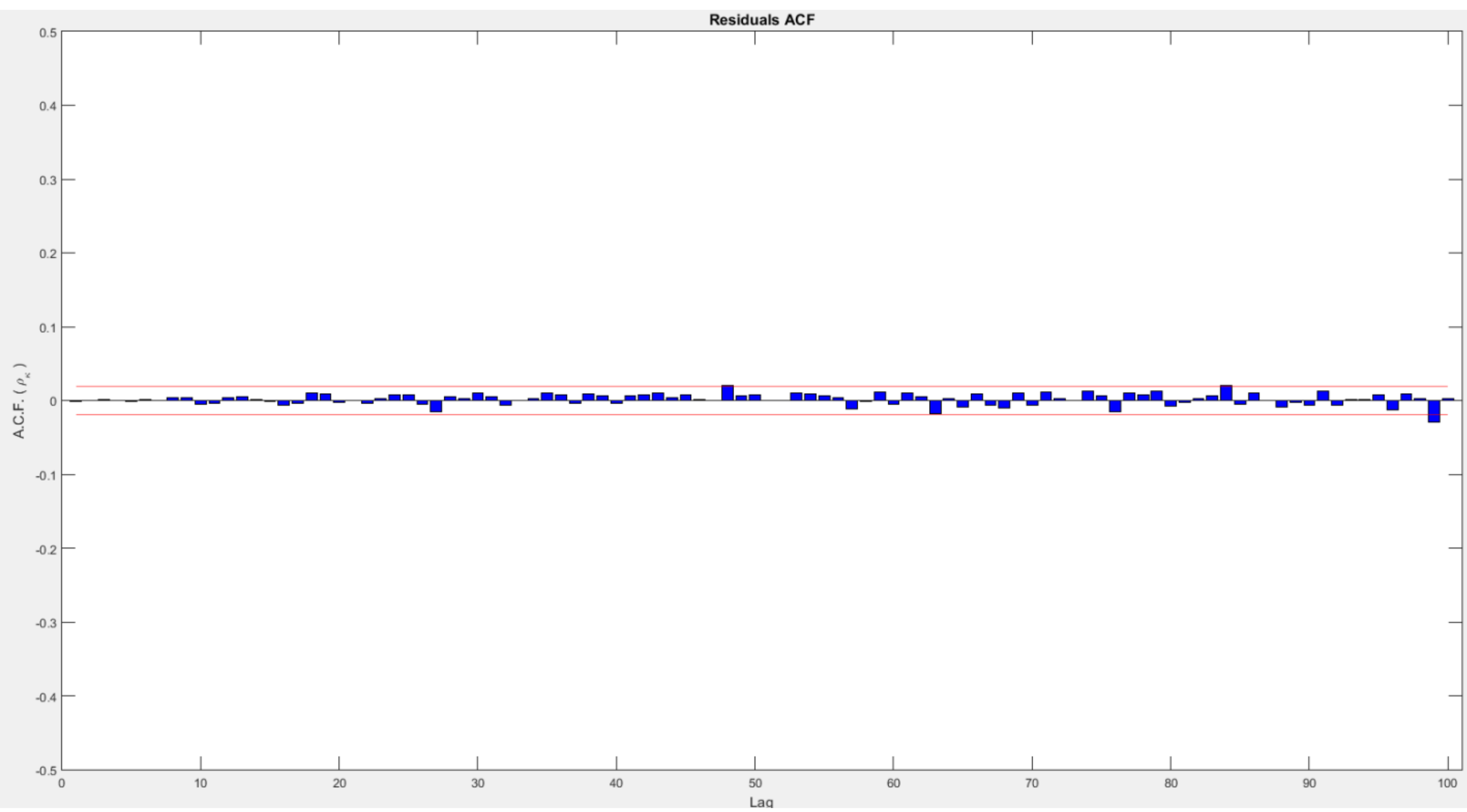
As we can see from the graph, for a model order of K = 39, we do not calculate many pseudo-frequencies as we would for a higher-order model, except for one (34.8163 Hz). However, there is a risk of not capturing some natural frequencies, as the graph shows certain frequencies converging at higher model orders.

Specifically, there appears to be a natural frequency in the range [21, 22] Hz that is not detected, as well as one in the range [56, 58] Hz. These frequencies could potentially be detected with a model order of K = 45, although this would introduce additional pseudo-frequencies.

C3. Validity check of the selected ARMA model.

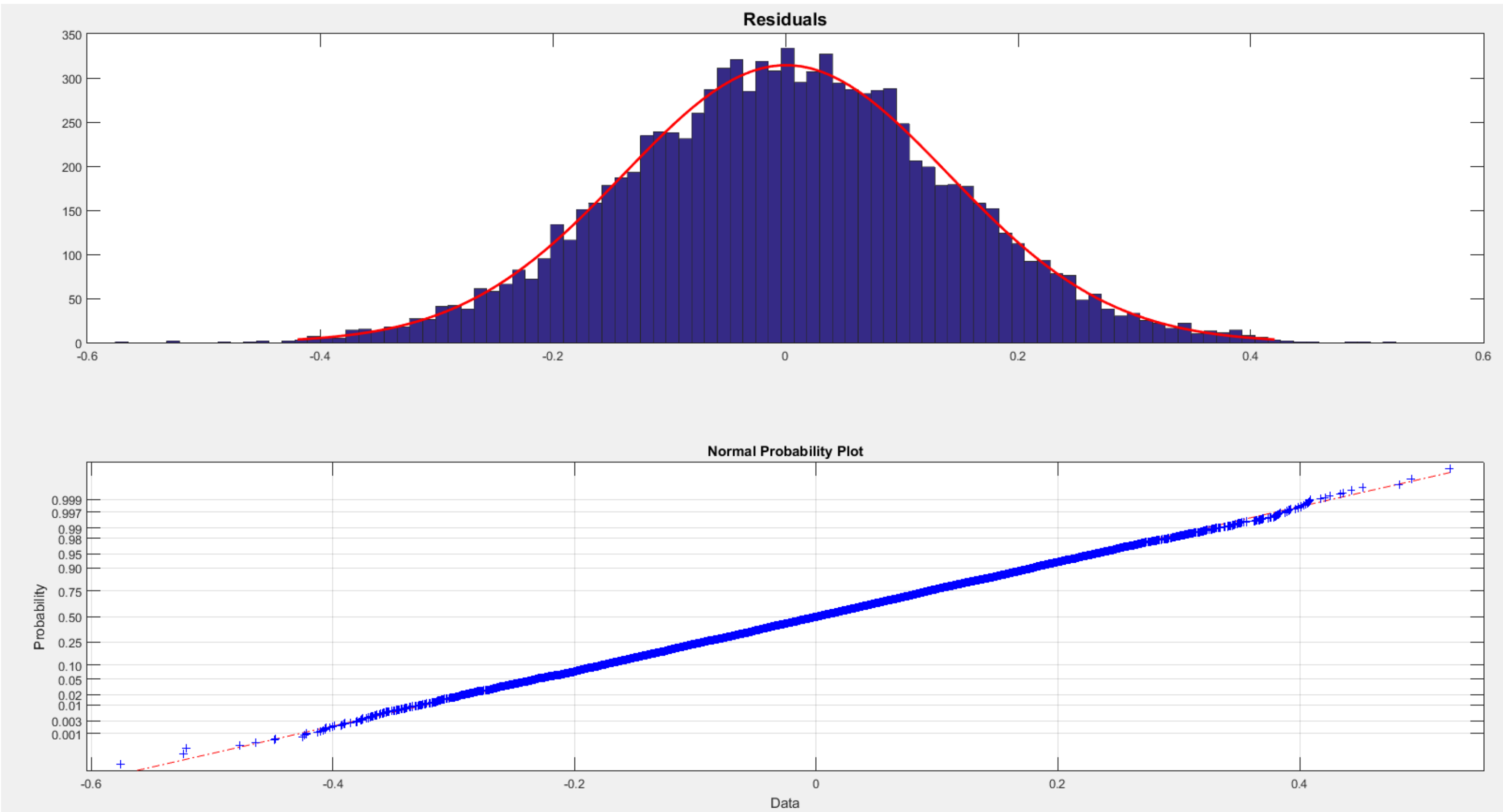
Έχοντας επιλέξει μοντέλο ARMA(39, 39) ,θα προβούμε σε έλεγχο εγκυρότητας του.

Initially, a plot of the normalized autocovariance of the residuals is created, with the statistical significance limits at a 5% error level.



As observed, there are a few values above the limits, though not by much. Therefore, we can accept that the residuals satisfy the whiteness assumption, with only a few values exceeding the statistical limits.

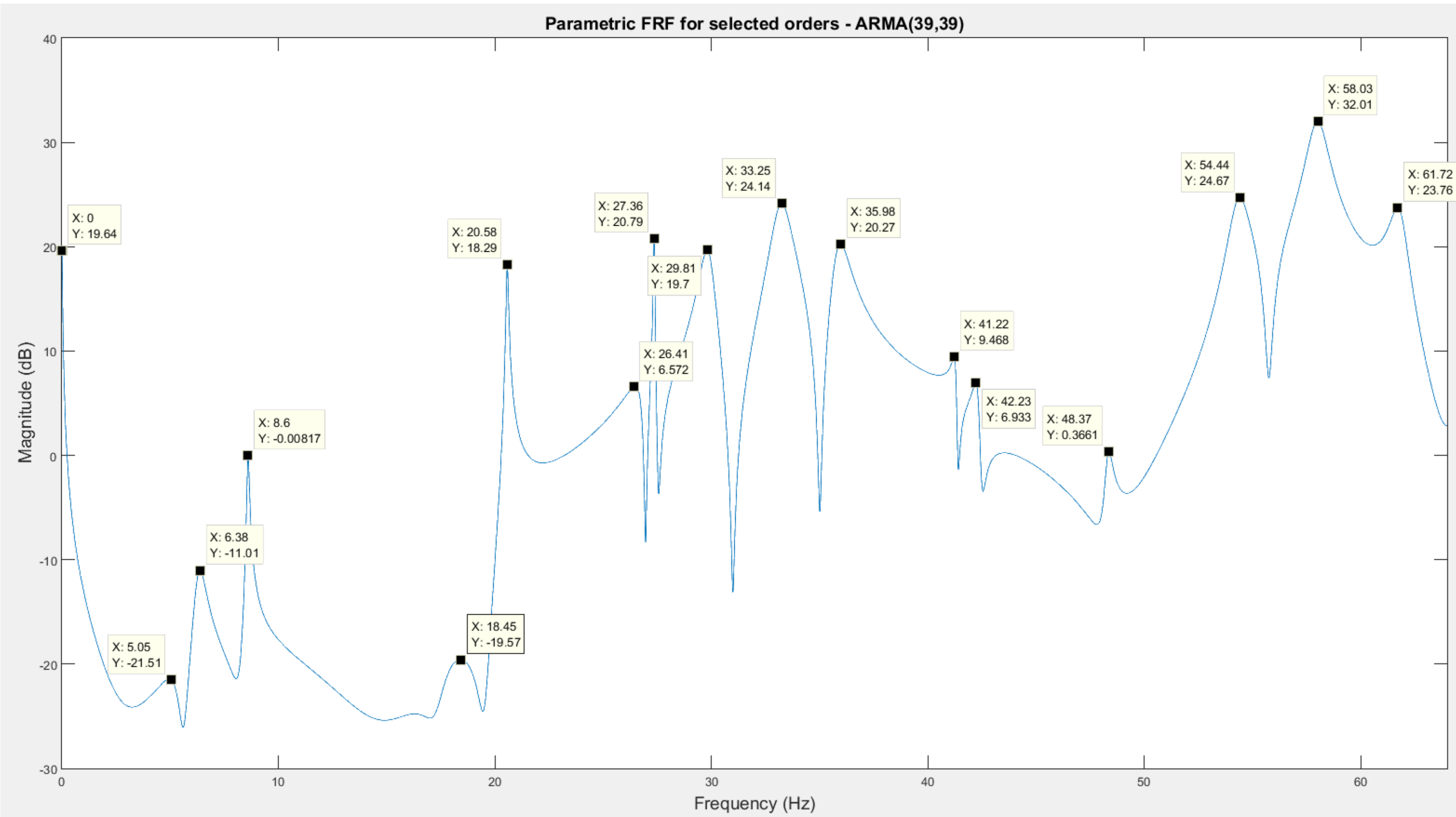
Next, we perform a normality check of the residuals using a histogram and a normal probability plot.



Both graphs indicate that the residuals closely approximate a normal distribution. From the above, it appears that the residuals largely resemble a Gaussian white noise signal, which is a good indication of the model's suitability.

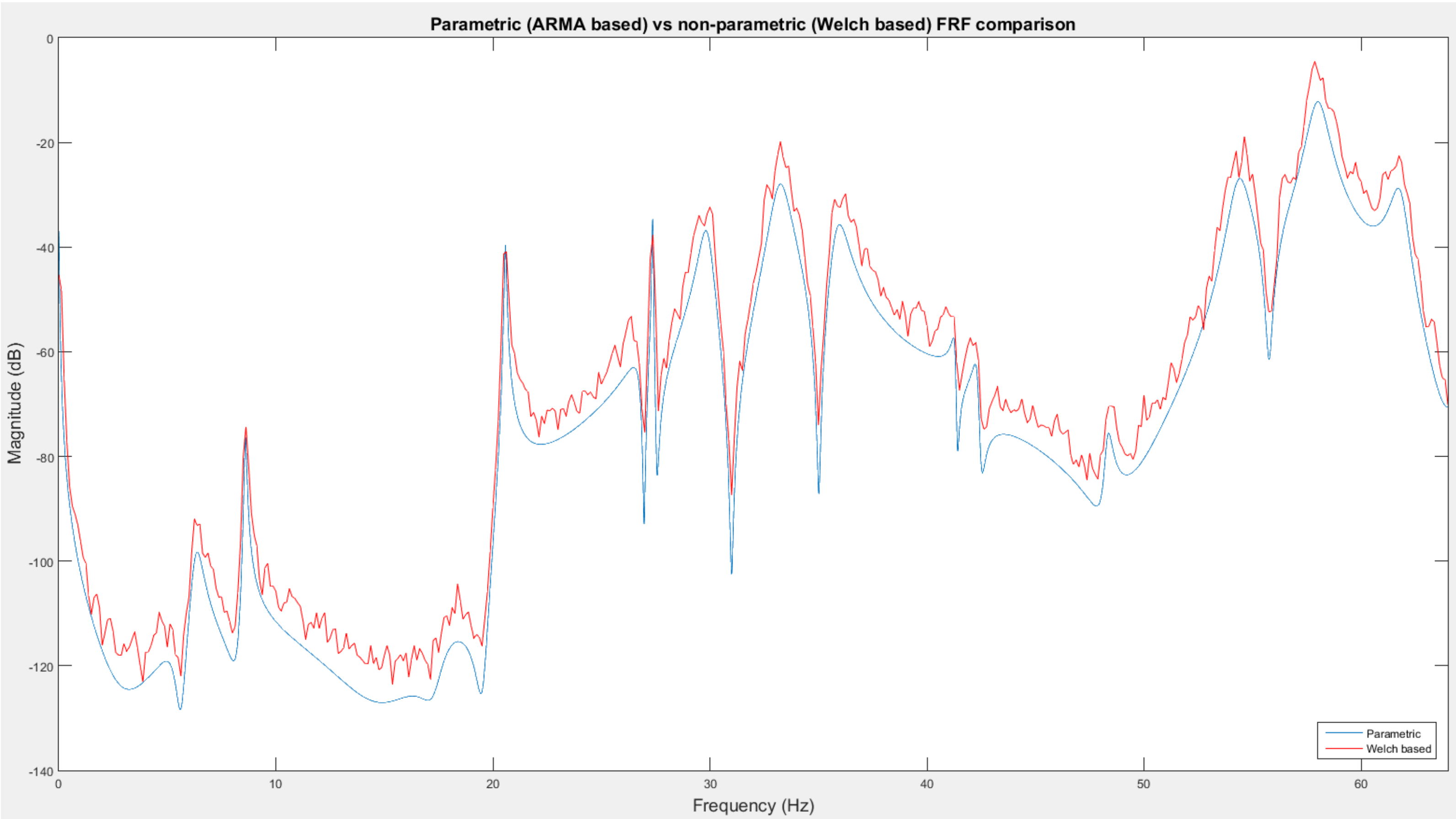
C4. Analysis of the Selected ARMA Model

The following are the graphs of the parametric spectrum estimation, along with a comparison with the non-parametric Welch method for a window of L=1024 and an overlap percentage of 90%



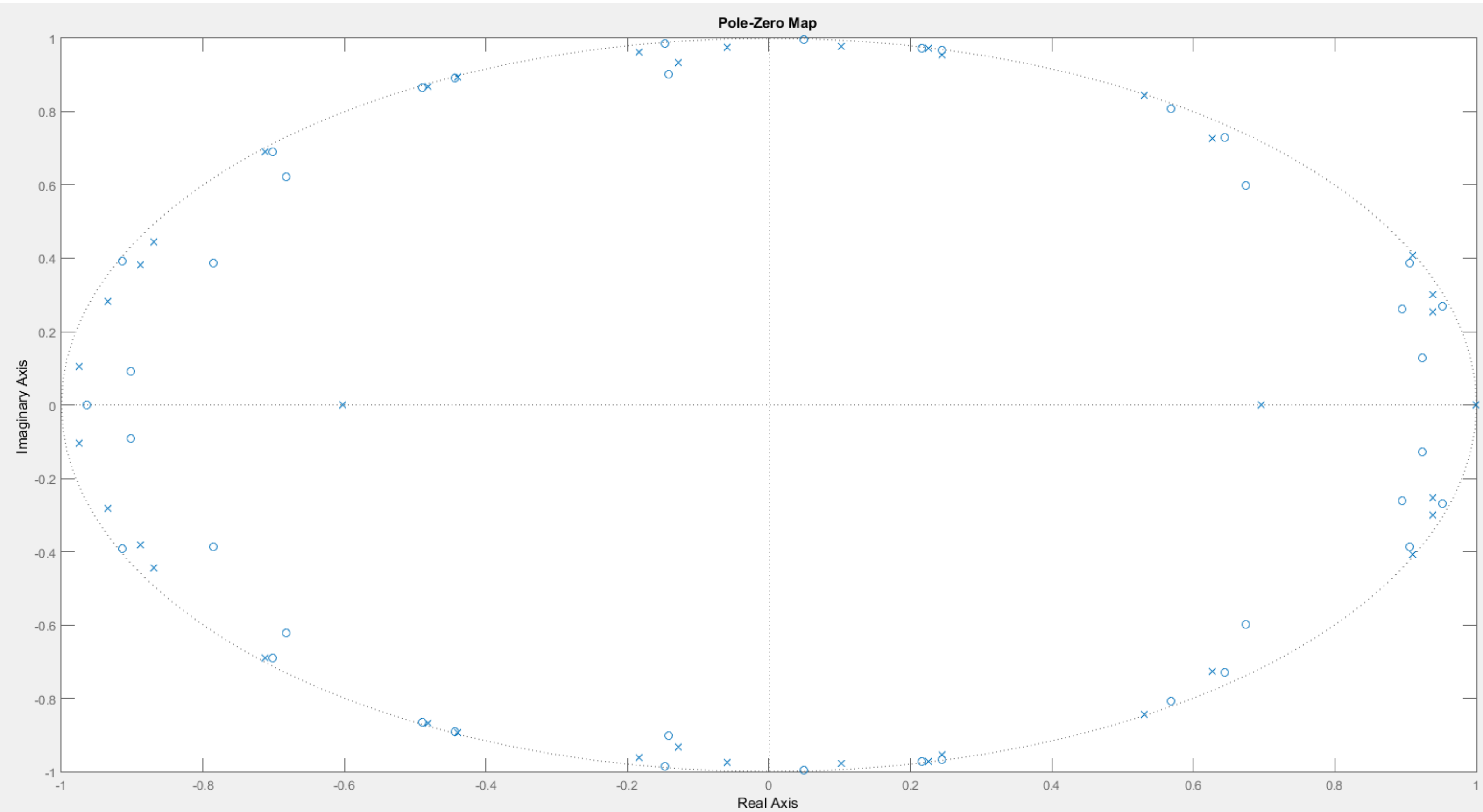
From the parametric spectrum estimation, compared with the table of calculated eigenfrequencies, it is now clear that the frequencies in bold and underlined stand out most prominently as peaks in the spectrum. The frequencies in bold but not underlined are visible as peaks in the spectrum but not as prominently. Finally, the frequencies (34.8163 Hz) and (55.7327 Hz) do not appear in the spectrum at all and can be considered pseudo-frequencies.

Finally, the two graphs are compared:

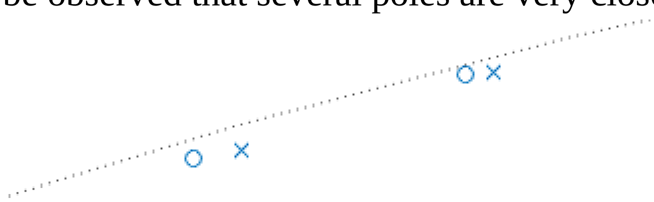


The comparison shows that the parametric spectrum estimation is much "cleaner." With the Welch method, there are indications of some additional peaks that may be natural frequencies or may result from the estimator's variance. To check for additional natural frequencies, a model with more parameters will be examined.

The representation of the poles and zeros of the system on the unit circle follows.

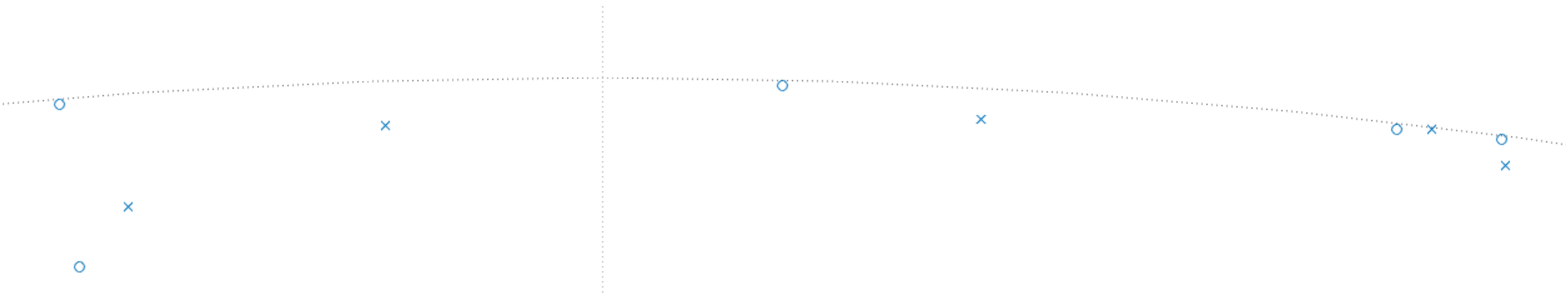


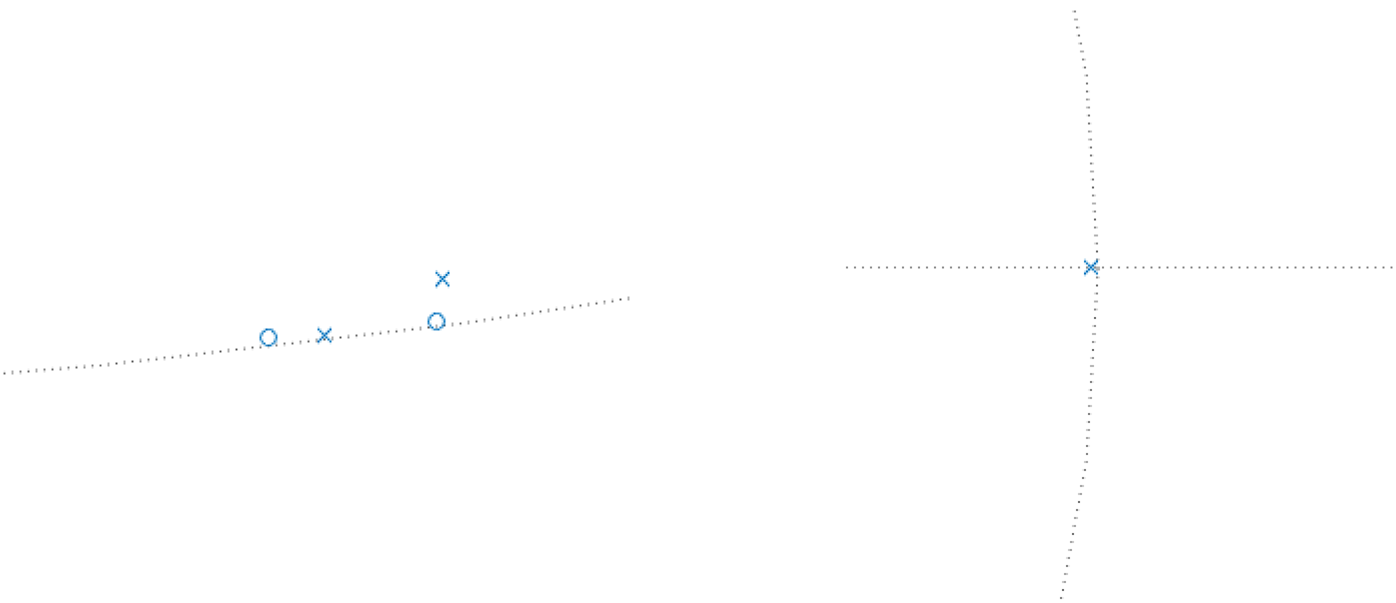
The graph shows that most poles and zeros are near or on the boundary of the unit circle. This indicates large time constants and strong dynamics. Additionally, it can be observed that several poles are very close to a zero.



However, since no pole coincides with a zero, there is no need to reduce the model's order due to over-parameterization.

Additionally, we observe poles and zeros that are extremely close to the unit circle, with some almost on it.

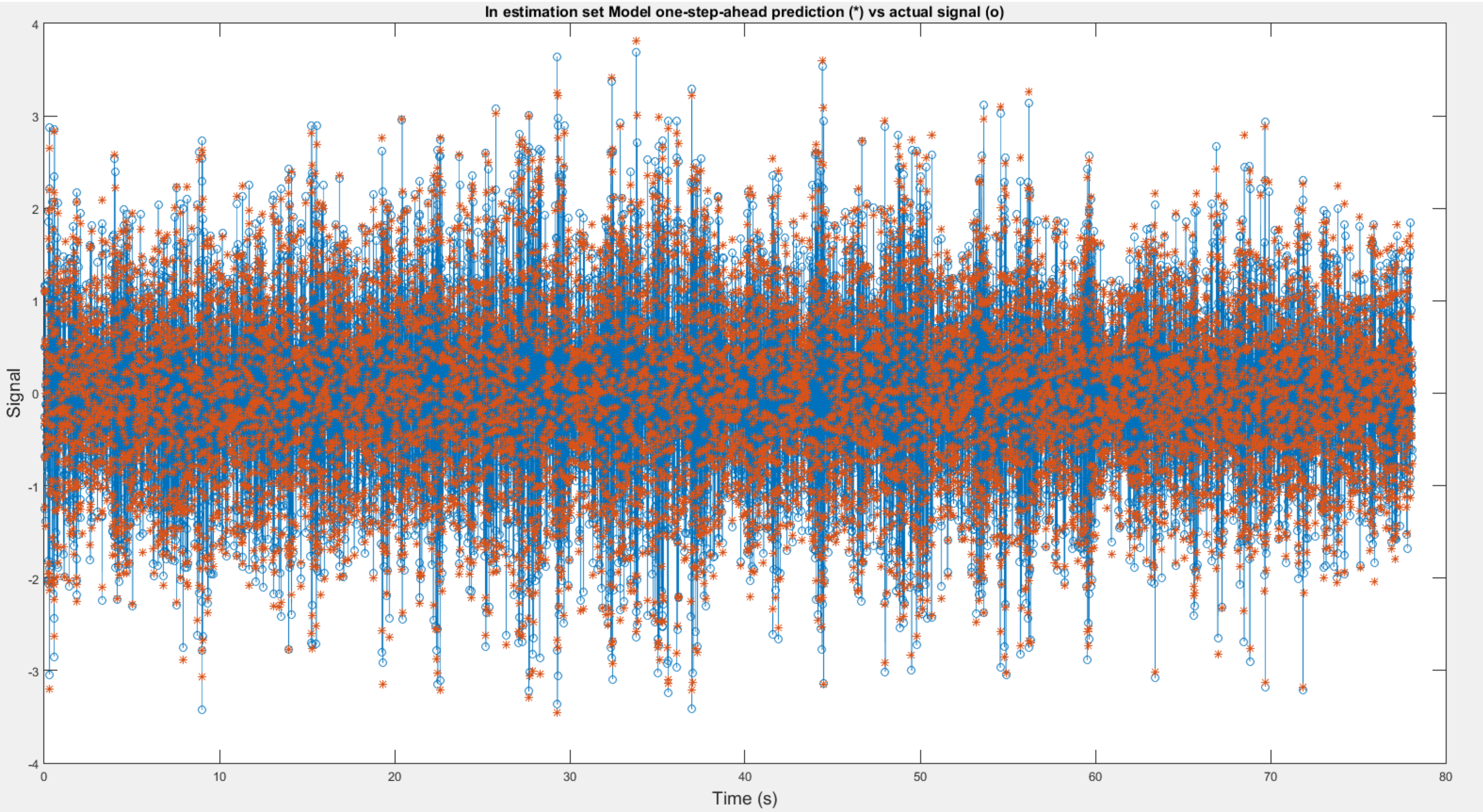




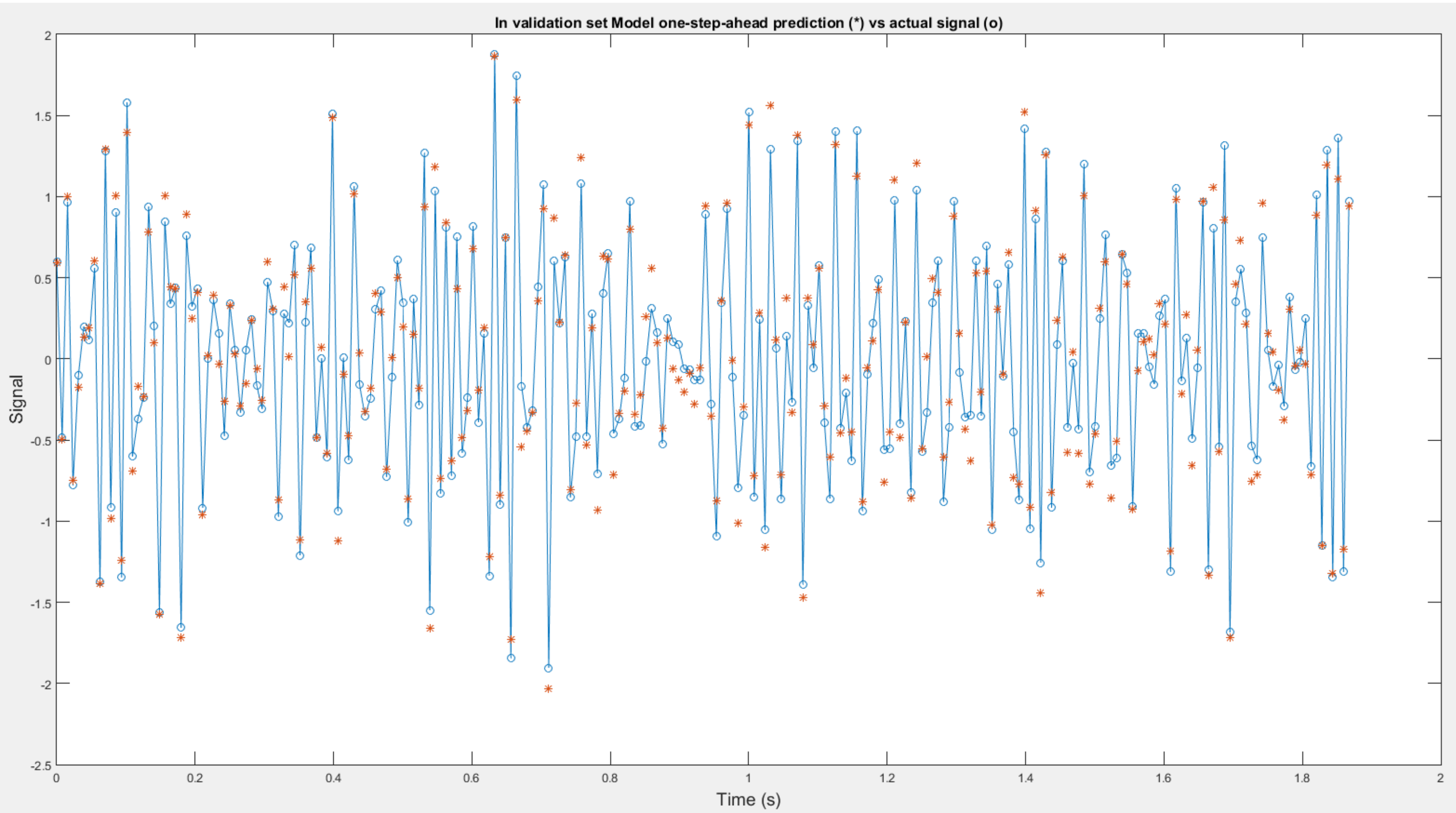
These poles are likely to create a relative instability in the system, which may explain the lack of stationarity.

C5. Prediction Based on the Selected Arma Model

For the one-step prediction based on the ARMA(39,39) model, both the estimation segment and the validation segment are used.

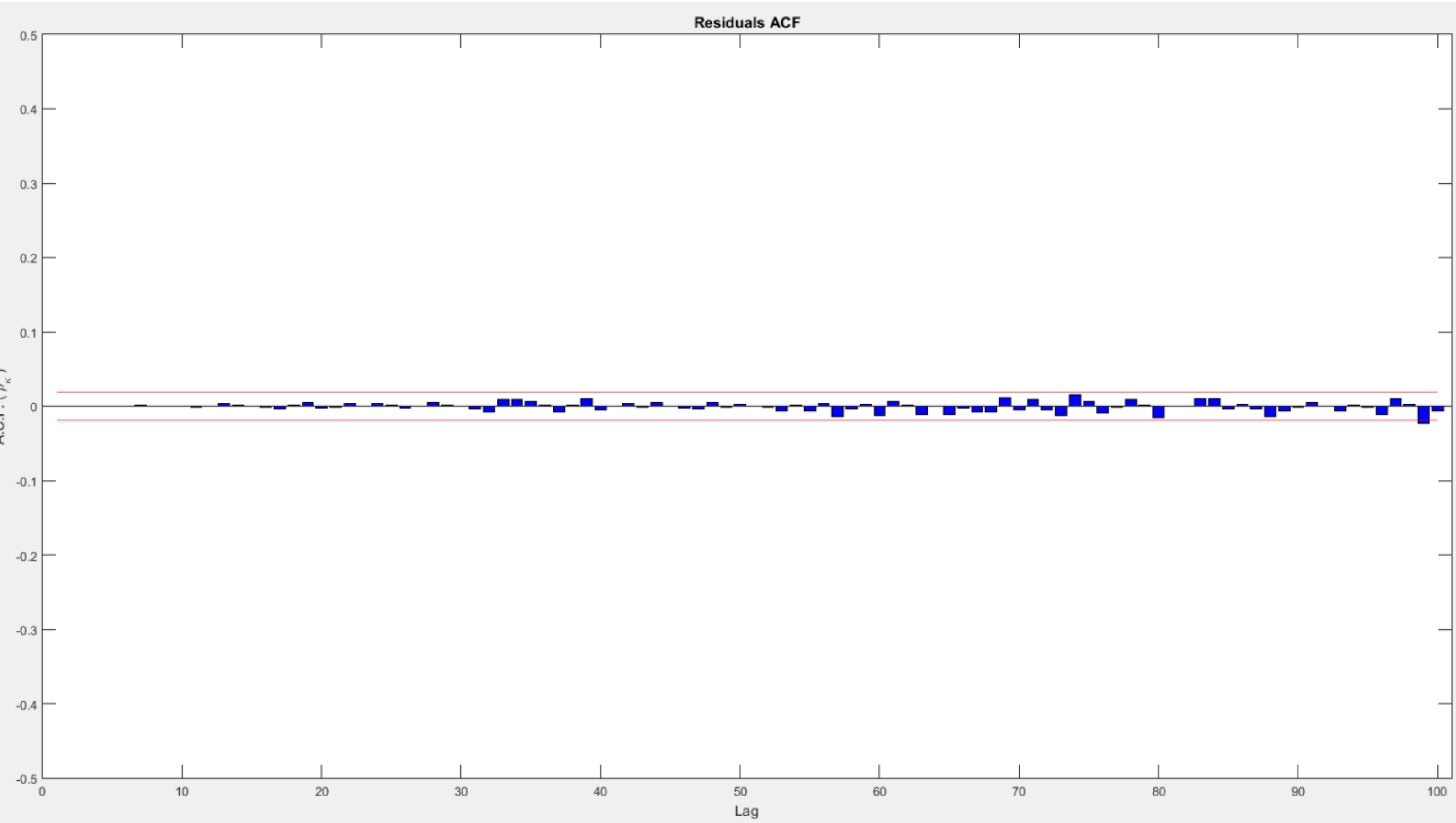


At a quick glance, the prediction appears to be quite good, but for greater clarity, we examine the prediction within the validation segment..

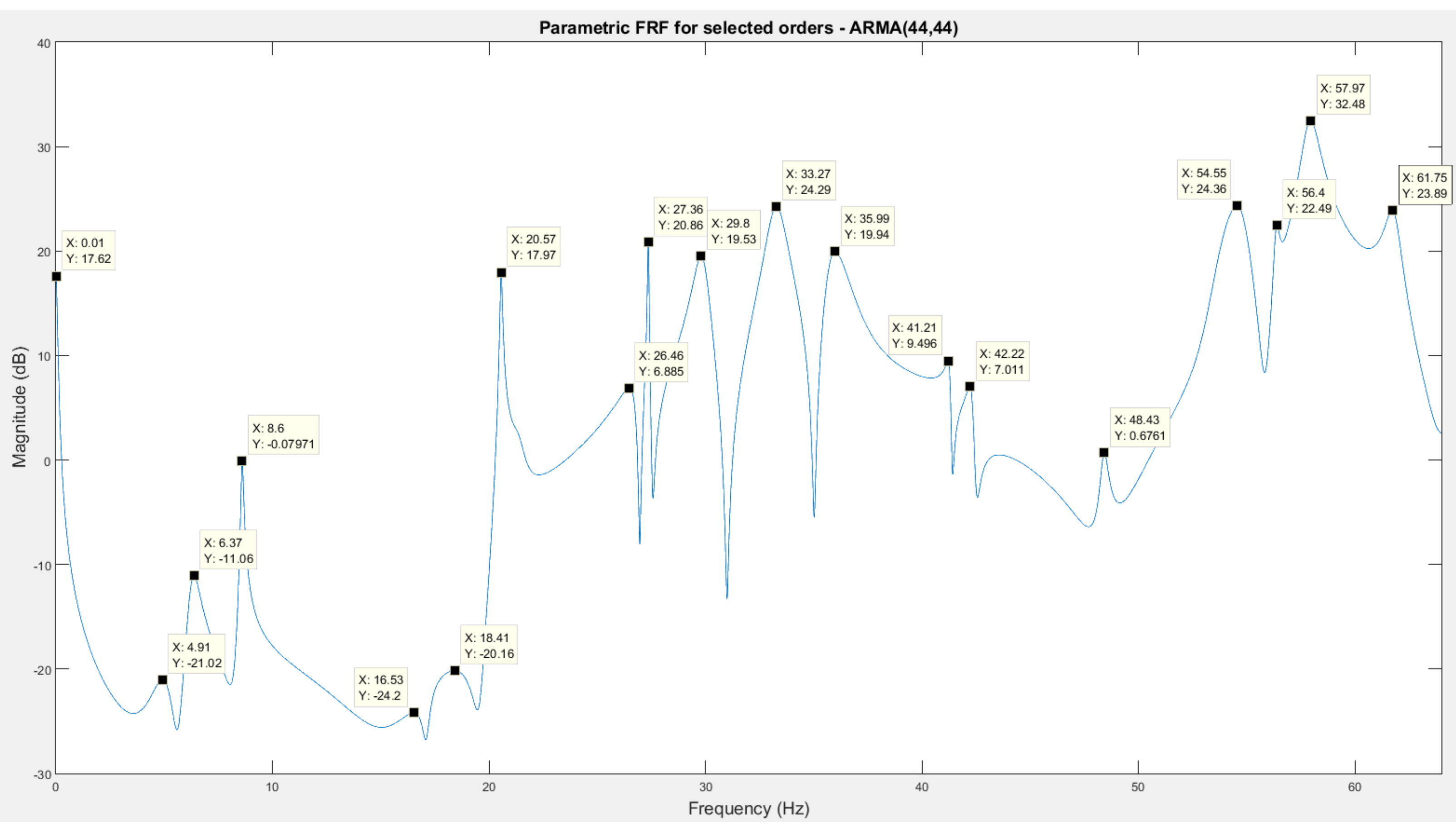


The prediction seems to approximate the signal quite well, and at certain points, it accurately matches the value. To confirm the model's predictive capability on the validation segment, a model with more parameters will also be examined.

Finally, it is noted that an alternative modeling approach will use an ARMA(44,44) model. For $K = 44$, the BIC criterion has the 3rd lowest value, and at this order, all natural frequencies appear to be detected from the frequency stabilization diagram. Additionally, there is a reduction in the residuals on the normalized autocovariance plot, indicating greater whiteness in the residuals.



Residuals ACF
Lag

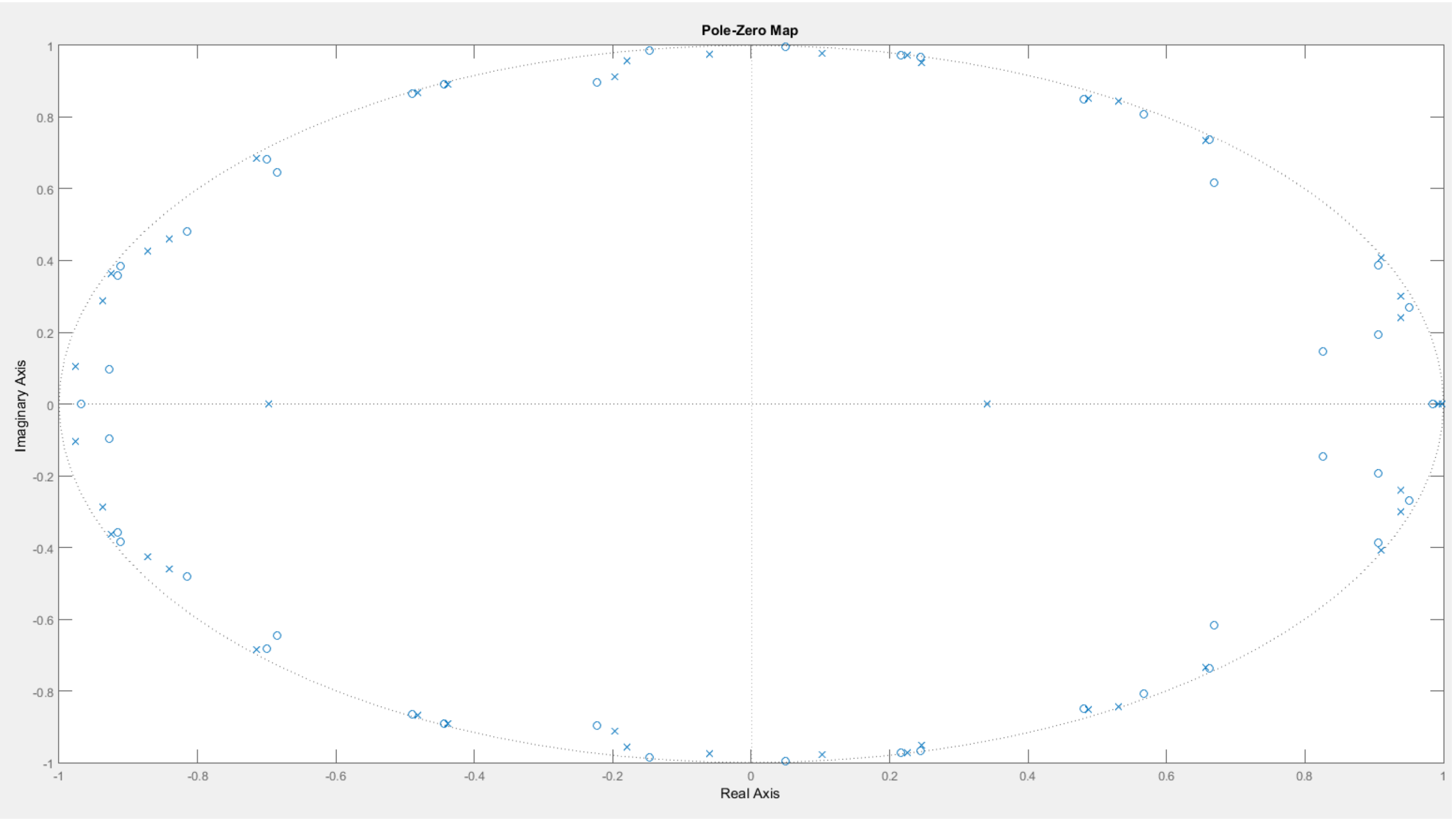


Parametric FRF for selected orders - ARMA(44,44)
Magnitude (dB)
Frequency (Hz)
X: 0.01
Y: 17.62
X: 4.91
Y: -21.02
X: 6.37
Y: -11.06
X: 8.6
Y: -0.07971
X: 16.53
Y: -24.2
X: 18.41
Y: -20.16
X: 20.57
Y: 17.97
X: 26.46
Y: 6.885
X: 27.36
Y: 20.86
X: 29.8
Y: 19.53
X: 33.27
Y: 24.29
X: 35.99
Y: 19.94
X: 41.21
Y: 9.496
X: 42.22
Y: 7.011
X: 48.43
Y: 0.6761
X: 54.55
Y: 24.36
X: 56.4
Y: 22.49
X: 57.97
Y: 32.48
X: 61.75
Y: 23.89

The natural frequencies that we were unable to estimate with the previous model are now computable.

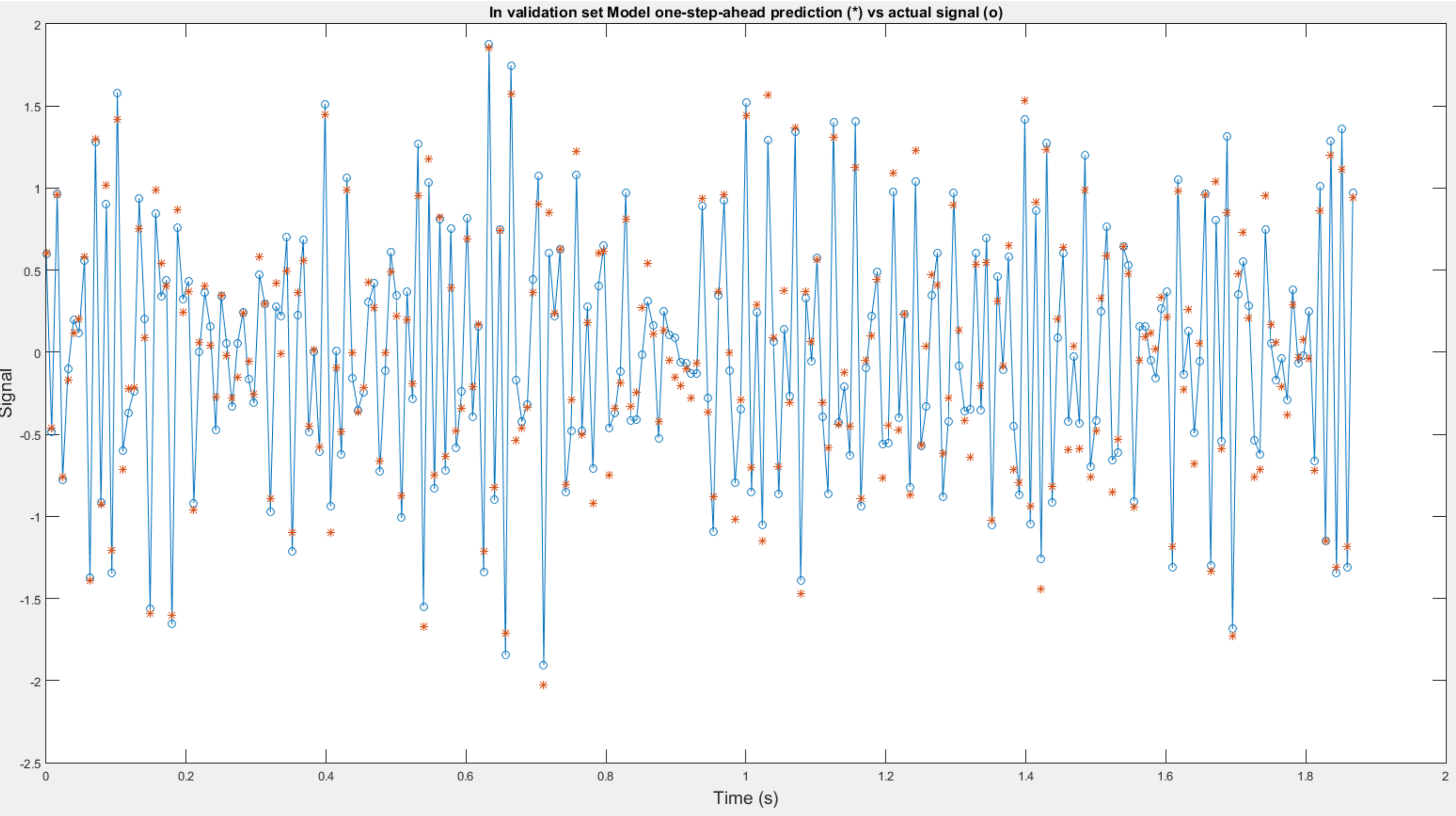
Natural Frequencies (Hz)	Damping Ratios (%)
5.1361	12.5332
6.2993	4.7247
8.5974	0.6737
17.1268	1.8056
20.5621	0.2495
21.4160	1.7832
26.8566	1.2140
27.3644	0.1094
29.8673	1.2139
33.2744	1.3908
35.7710	1.5348
36.3698	3.8884
41.3086	0.2608
42.3296	0.3952
48.4103	0.3970
53.7912	1.5775
54.7281	1.1128
56.3677	0.2489
57.9266	0.7263
61.8034	0.6003

Nevertheless, there does not appear to be any improvement in the spectrum, nor do additional eigenfrequencies appear. This indicates that the additional computed frequencies are pseudo-frequencies.



Examining the poles and zeros for this model, we observe that several poles have approached the zeros, though there is no overlap. Thus, the model is not over-parameterized. However, there may be a larger number of poles and zeros at the boundaries of the unit circle, which could create relative instabilities.

Finally, we also examine the one-step predictions for the ARMA(44,44) model to reach the final conclusion.



The prediction shows a slight improvement, with some points of the prediction being closer to the actual predicting values. This was expected, as the residuals and, therefore, their variance have been reduced.

In conclusion, the ARMA(44,44) model does not provide significant improvements in the system's modeling and instead calculates pseudo-frequencies. Therefore, the ARMA(39,39) model is chosen as optimal, unless maximum accuracy in prediction is the primary objective. It should also be noted that models for non-stationary stochastic processes or models with exogenous inputs might be better suited for modeling this signal.