INSTITUTO SUPERIOR TECNICO

MENO

MODELLING OF SEAWAVES

Project 1: Statistical and spectral analysis of waves

2021/2022 academic year, 1° semester

DAVIDE MELOZZI

IST1102230

INTRODUCTION

The goal of this research is to conduct statistical and spectral analysis of wave elevation time-series by examining a sample of heave movements from a buoy installed along the coast of Rio Grande do Sul, Brazil with coordinates Lat: 31.57, Lon: -49.87. This report presents the processes and equations used to solve the challenges outlined in the project specifications, always in the context of the applied physical-mathematical modeling covered in class.

To do so, keep in mind that the sea surface is a complicated structure governed by fluid mechanics equations. However, analytically describing these rules is challenging. It will be regarded a stochastic process, which allows the statistical aspects of waves to be evaluated through the frequency and probability domains. The spectral and statistical analyses compensate for the lack of comprehensive knowledge of wave genesis and propagation mechanics. This enables us to anticipate numerous statistical features with fair accuracy, which is required to complete this report.

The sample data is saved in a file with the .tsr extension entitled "W200908311900," with each line including the corresponding data of "heave", "etaEW", and "etaNS" at an instant in time (first column) (second, third and fourth columns, respectively). According to the data, the sample interval is 0.78 seconds, which translates to a frequency of 1.28 hertz

The MATLAB software was used to do all of the computations and is possible to consult at the end of the report.

PART 1

1. Mean, dispersion (variance), skewness, and kurtosis (first four moments).

Having obtained the data from the file, we wish to investigate the worth of these four first moments in relation to the data of the buoy's wave elevations. In this situation, we'll simply evaluate the heave motion because we're interested in the wave surface height.
The average of the wave heights is defined as the mean wave height, while the frequency of the wave is defined as the dispersion. Skewness is a measure of asymmetry, while kurtosis indicates how steep the distribution is.

The **probabilistic moments** are measuring methods that allow us to understand how the distribution along an aleatory variable accommodates. In order to calculate a continuous aleatory variable, such as this case, is given the follow equation where $f(x)$ is a probability density function:

$$M_X(t) = \int_{-\infty}^{+\infty} e^{(tx)} f(x) dx$$

The first four probabilistic moments are: mean, variance, skewness, and kurtosis respectively, given by MATLAB and demonstrated through the formulas above

Once the wave distribution must be sinusoidal, the values of mean, approach to zero. The following equation will produce the **mean surface** elevation over time:

$$\underline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

N is the number of elements in the data (number of samples) and xi is the surface elevation measured at a specific time.

The **dispersion** is a measurement to the second power of the predicted value of the gap between the data and its mean value.

Variance can be used as a quick way to analyze the data collected before proceeding with the full analysis and is given by:

$$var = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \underline{x})^2$$

Standard deviation:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} (x_i - \underline{x})^2}$$

Skewness and **Kurtosis** are dimensionless shape factors of the symmetry/asymmetry of the normal distribution and are given respectively by:

$$Skew = \frac{1}{N}\sum_{i=1}^{N} (\frac{x_i - \underline{x}}{\sigma})^3$$

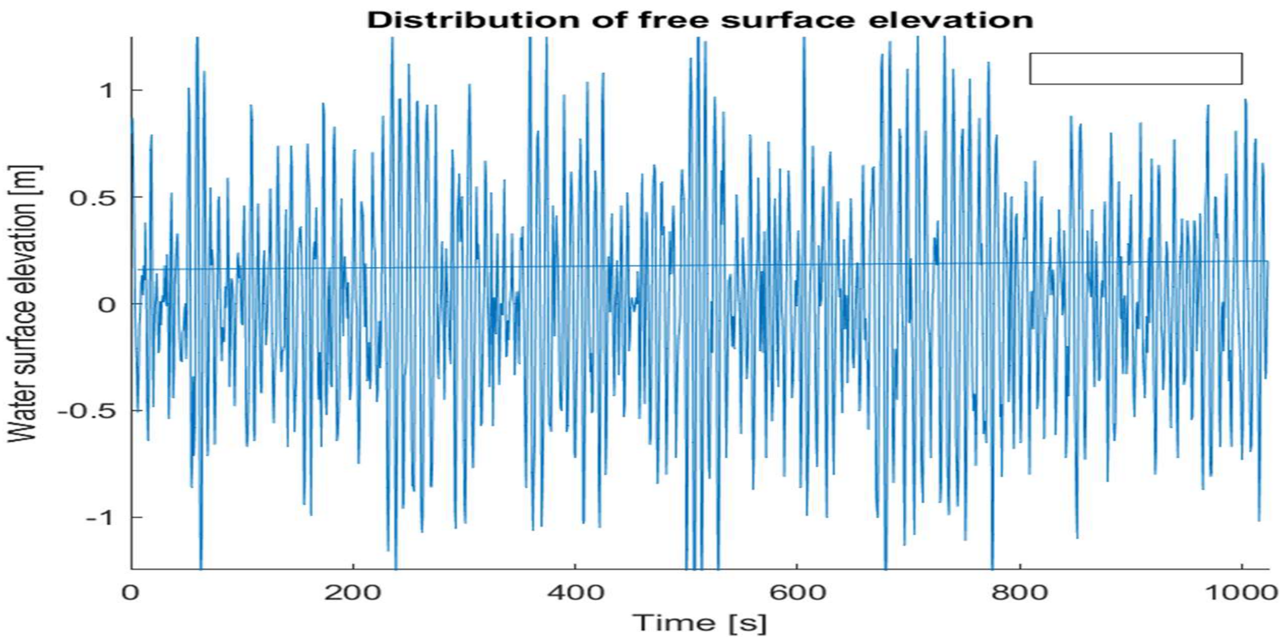
$$Kurt = \frac{1}{N}\sum_{i=1}^{N} \left((\frac{x_i - \underline{x}}{\sigma})^4\right)$$

Skewness is a helpful function that may be used to determine whether or not a data distribution is symmetric. If the data distribution is symmetric, the skewness is near to zero (or equal to zero if absolutely symmetric), and if the data distribution is more concentrated to the right of the mean value, the sample is right skewed (and the parameter positive). In the opposite instance, it will be left-skewed with a negative skewness.

Kurtosis is a measure of how flat the data distribution is in comparison to a normal distribution. If the kurtosis is equal to zero (mesokurtic), the data distribution flattens to the same extent as the normal distribution. However, if it is less than zero (platykurtic), the data distribution is flatter than the normal distribution. Finally, if the kurtosis is greater than zero (leptokurtic), the graph will be more bloated.

Probabilistics Moments	Values
1st moment – Mean [m]	0.0012
2nd – Standard deviation [m]	0.5062
3rd moment – Variance [m^2]	0.2567
4th moment – Skewness	-0.0287
5th moment - Kurtosis	2.8820

The MATLAB representation of the waves surface elevation in a time series is:



2. The distribution of the free surface elevation (taking ordinates of the whole data) and compare with the Gaussian distribution and Longuet-Higgins.

We'd want to spread our data probability distribution so that we may examine and compare it to the Gaussian (or normal) distribution and the Longuet-Higgins functions. In general, the Gaussian distribution is a sort of continuous probability distribution for a real-valued random variable.
Provide Longuet, it is known that the probability function is derived from the cumulate function. The density is calculated applying the function density which is a process similar to the orthogonal polynomials. These polynomials to the Gaussian distribution are the Hermite's polynomials where the $formul$a is given by:

$$f(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} 3 \cdot \left[1 + \frac{m_3}{3!} H_3(z) + \frac{m_4 - 3}{4!} H_4(z) + \frac{m_5 - 10m_3}{5!} H_5(z) + \frac{m_6 - 15m_4 + 30}{6!} H_6(z) + \cdots\right]$$

Once the mean of surface elevation is zero, the Longuet-Higgins distributions gets the follow equation:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}} \cdot \left[1 + \frac{\lambda_3}{3!} H_3\left(\frac{x}{\sigma}\right) + \frac{\lambda_4}{4!} H_4\left(\frac{x}{\sigma}\right) + \frac{\lambda_3^2}{72} H_6\left(\frac{x}{\sigma}\right)\right]$$

Where

$$H_3 = x^3 - 3 \cdot x$$

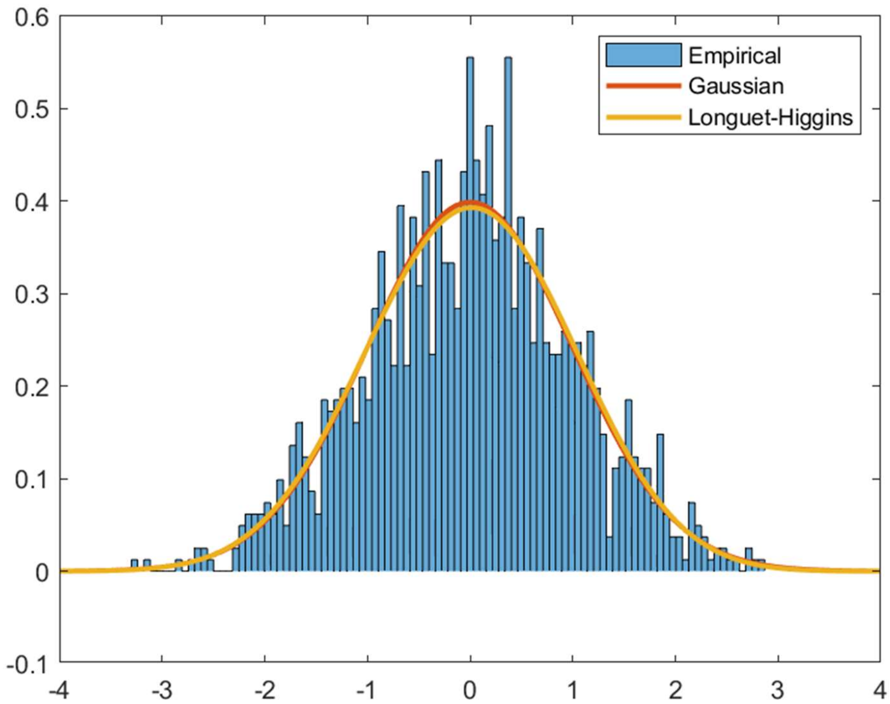
$$H_4 = x^4 - 6 \cdot x^2 + 3$$

$$H_6 = x^6 - 15 \cdot x^4 + 45 \cdot x^2 - 15$$

And

$$x = \frac{Surface\ Elevation}{Standart\ Deviation}$$

In comparison to the Gaussian distribution, the Longuet-Higgins distribution considers non-linear effects, and if they are equal to zero, the Longuet-Higgins distribution is the same as the Gaussian distribution.

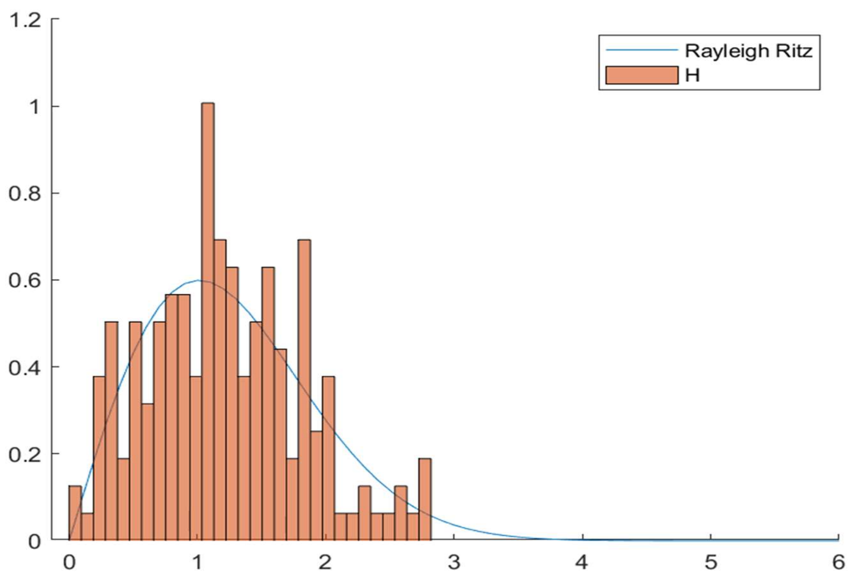


A normal distribution with a zero mean can be used to depict the wave elevation distribution. The actual empirical distribution and the other two techniques show some disparity. We can observe that for larger and smaller wave heights, the distribution closely matches the histogram, but for medium values, the empirical distribution and the Gaussian or Longuet-Higgins distributions do not provide a decent approximation. Although the Longuet-Higgins model matches the actual distribution better since it takes into account the non-linear impact, the fact that a variable is represented by an anormal distribution indicates that it is linear.
Also the approximation of the Longuet-Higgings distributions seems to be slightly better than the Gaussian distribution. This does not show much symmetry and that is why Gaussian distribution should not be the best option to use for this case.
Once the Gaussian distribution occurs in the nonexistence of factors affecting the idealized random process, this distribution should be used only for deep waters. Contrary, the Longuet-Higgins distribution considers the irregularities of the wave which is more appropriate in intermediate and shallow water analysis.

3. Identify Up-crossing and Down-crossing individual waves and associated periods. Represents the wave heights in a histogram and compare with the Rayleigh distribution and calculate the significant wave height (H1/3) and significant wave period. Take the maximum wave height and compare with the significant wave height and the mean wave height.

The zero-crossing analysis is the usual approach for calculating the short-term statistics of a wave record. For a suitably long record, either the so-called zero-upcrossing approach or the zero-downcrossing method may be employed; the statistics should be the same.
The Rayleigh distribution is one of the most often used probability distributions, is critical in modeling and analyzing life-time data in areas such as project effort loading modeling, survival and reliability analysis, communication theory, physical sciences, technology, and diagnostic imaging.
To do that, first must find the index of the zero up and down crossing, passing the surface elevation of time series to signal. If the difference between two consecutive signals were equal to 2, it means it is an up crossing, and if it is equal to - 2, it is a down crossing. Then, the wave height subtracting the min surface elevation to the max surface elevation within the index calculated must be calculate. Thereafter must be calculated the periods of individual waves subtracting the initial time to the final time such as the mean of the individual wave heights and finally the maximum wave height and the wave period.

There are some potential downsides to using this model, such as the negative probability density over some range of large negative surface displacements commonly observed in steep sea states, the requirement to know the higher-order statistical averages before using the model, and the distribution used in the statistical analysis of shallow water waves cannot be used to derive distributions for the wave elements, in contrast to the Gaussian distribution.



There is a clear distinction between the individual wave distribution and the Rayleigh distribution. Except in the case of exceptionally huge waves in deep water, which display horizontal asymmetry in the profile, the two definitions (up and down crossing) are predicted to be statistically equivalent. This is because the particular wave must be the same for both. However, it is crucial to note that the theoretical Rayleigh distribution has a peak that is somewhat to the left of the actual one.

In the time domain is possible to calculate two important characteristics: the significant height, which is one of the most used measures in real sea stats, (characterizes the wave heights, but not just the maximum and minimum amount), and the significant wave period.

The significant wave height is defined as the mean or average of the greatest one-third of the waves in the wave record, as denoted by the notation H1/3, which represents the average of the largest 33% of the waves in the record. The statistical distribution of ocean waves is influenced by significant wave height. The most common waves are shorter than Hs in height. This means that encountering a large wave is not all that often. However, statistically, a wave substantially larger than the significant wave is feasible.

In general, a Rayleigh distribution is a good approximation of the statistical distribution of individual wave heights.

The significant values was evaluated with the help of MATLAB, and with the help of the following table is possible to compute.

Wave height (Hs) [m]	1.88
Wave period (Tz) [s]	8.12
Maximum wave height (Hmax) [m]	2.80
Mean wave height (Hmean) [m]	1.21

The distribution of the individual waves were calculated just like the distribution of the surface elevation. The empirical distribution was achieved with the help of MATLAB along with the Rayleigh distribution of the individual wave heights. These distributions were computed for the zero up and down crossing.

PART 2

1. Compute and plot the autocorrelation function from the time series;

In this part 2 we have to introduce multiple factors, it is important to optimize and examine the spectrum of the data to understand the behavior in a time series. As is customary, the waves under consideration are assumed to be irregular, and so may be characterized as a collection of numerous distinct harmonic waves that can be analyzed using properties such as amplitude and frequency. The Fourier transform was used to convert from a time domain to a frequency domain. It is a well-known and often used mathematical expression for investigating waves of any description.

Thus, the function f(t) (the value of heave at each instant of time) is a discrete set of points rather than a continuous function. If we first assume that the discrete values of f(t) are deterministic (no randomness in their measurement) and have finite energy (as is physically predicted for deterministic buoy motions), we may describe its Discrete-Time Fourier Transform (DTFT).

However, the quantities of heaves experienced are not deterministic - the irregularity of genuine waves causes us to believe the wave process to be random. The mathematical instrument used to explain such behavior is a random sequence, which is made up of an ensemble of potential realizations, each with its own chance of occurrence. In this situation, the manifestations of a sequence of random heaves lack finite energy and so lack a valid DTFT. However, it is acceptable to suppose that the series has limited average power and, as a result, may be described by a mean "power spectral density."

We will see it better later.

The **autocorrelation function** can be utilized for two things:

- Non-randomness in data must be detected.
- If the data are not random, choose a suitable time series model.

When using autocorrelation to determine an acceptable time series model, autocorrelations are often shown for multiple delays.

Definition:

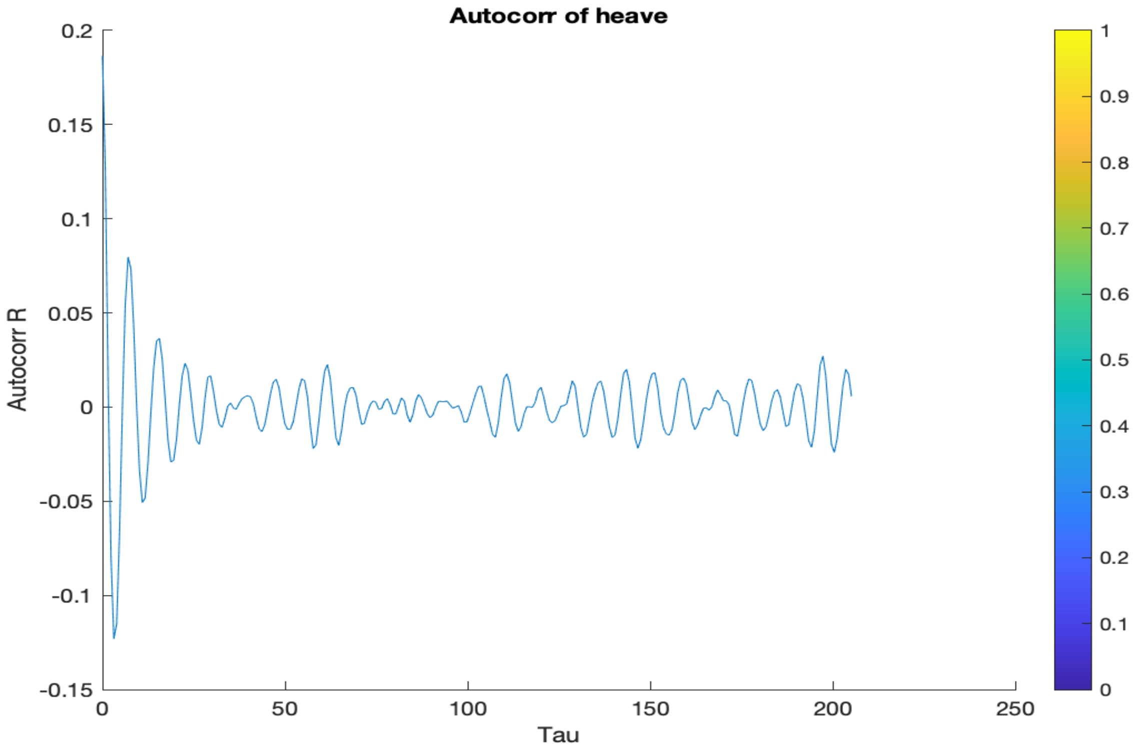
Given measurements, Y1, Y2, ..., YN at time X1, X2, ..., XN, the lag k autocorrelation function is defined as:

$$r_k=\sum_{i=1}^{N-k}(Y_i-\bar{Y})(Y_{i+k}-\bar{Y})/\sum_{i=1}^{N}(Y_i-\bar{Y})^2$$

Although the time variable, X, is not utilized in the autocorrelation calculation, it is assumed that the observations are equally spaced. A correlation coefficient is autocorrelation. The correlation, however, is between two values of the same variable at periods Xi and Xi+k, rather than two independent variables. When using autocorrelation to identify non-randomness, just the initial autocorrelation is generally of importance. When using autocorrelation to determine an acceptable time series model, autocorrelations are often shown for multiple delays. One of the important assumptions in establishing whether a univariate statistical process is under control is randomness. If the assumptions of constant location and size, randomness, and fixed distribution are valid, the univariate process may be described as Yi=A0+Ei, where Ei is an error factor.

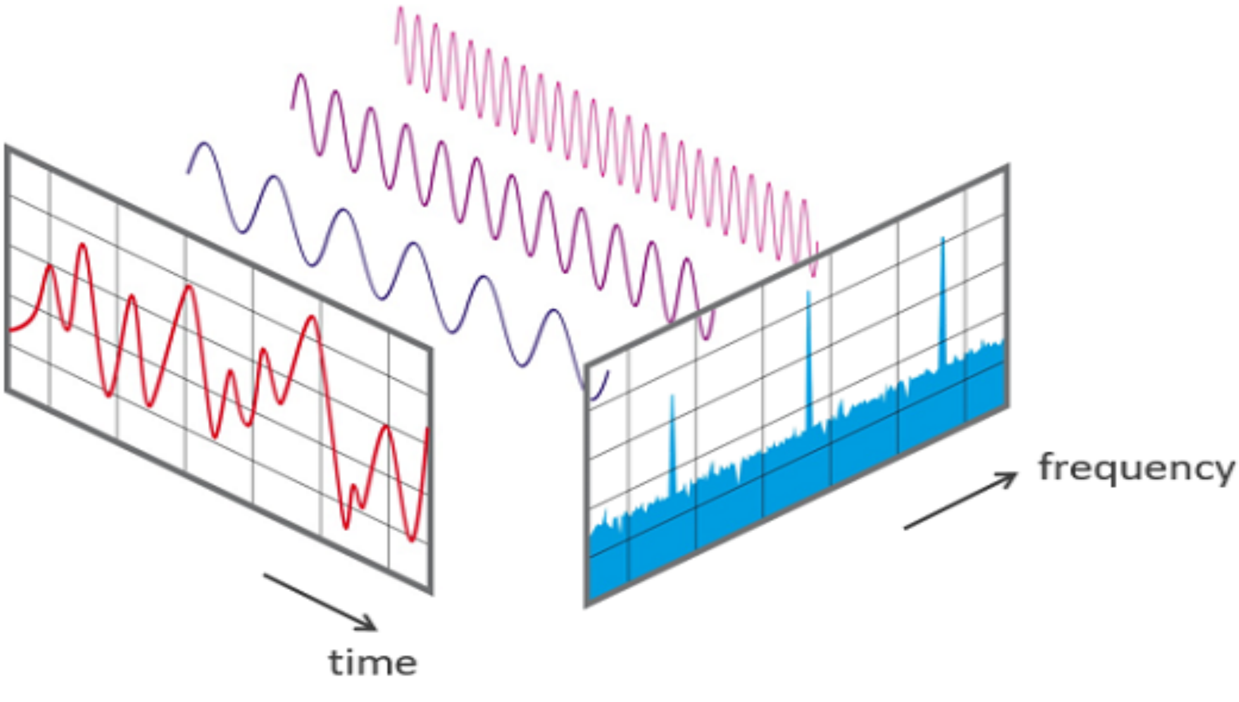
If the randomization assumption is violated, an alternative model must be utilized. This is usually a time series model or a non-linear model (with time as the independent variable).

The computation resolved permit to obtain the following plot for the autocorrelation function:

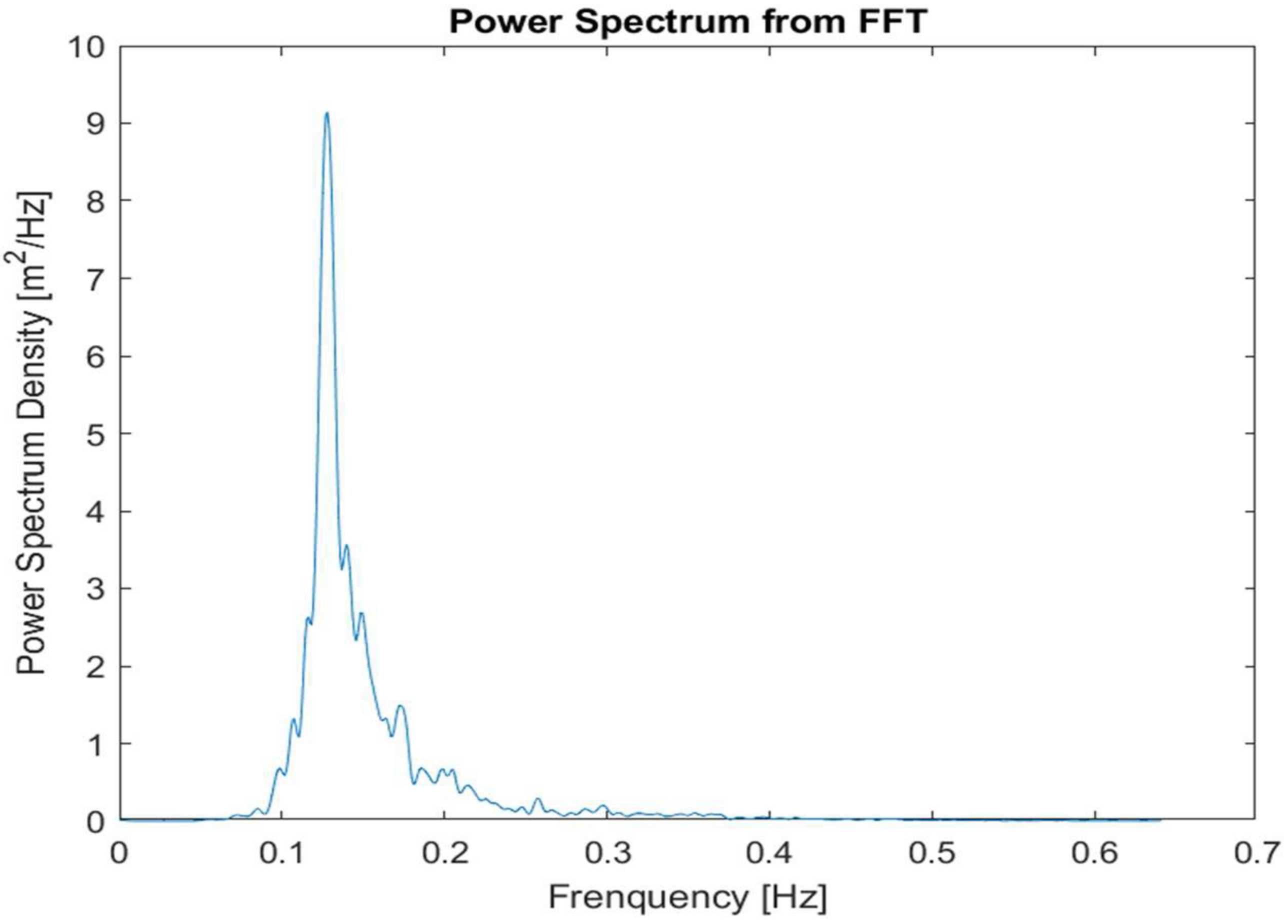


2. Using FFT, calculate and plot the power spectrum

A fast Fourier transform (FFT) is a method that computes a sequence's discrete Fourier transform (DFT) or its inverse (IDFT). Fourier analysis transforms a signal from its native domain (typically time or space) to a frequency domain representation and vice versa. Decomposing a series of values into components with various frequencies yields the DFT. This operation is helpful in many disciplines, but computing it straight from the definition is frequently too slow. Such changes are quickly computed by factorizing the DFT matrix into a product of sparse (mainly zero) elements. As a consequence, it manages to minimize the difficulty of computing the DFT from O(N2), which happens if the definition of DFT is simply applied, to O(NlogN), where N is the data size. The performance difference can be substantial, especially for large data sets where N may be in the millions.
Many FFT techniques are substantially more accurate in the presence of round-off error than directly or indirectly assessing the DFT specification. There are several FFT algorithms based on various published theories, ranging from simple complex-number arithmetic to group theory and number theory.
In our case is critical to define a sample interval "dt" that will be utilized throughout the project. In this manner, we will utilize the Power Spectrum Density to compare it with the following ways; we compute and plot the Power Spectrum Density in the next cell as the primary one.



In our case we obtained that power spectrum:



3. Reduce the noise of the resulting spectrum using the Daniell's method (play with the smoothing as you prefer) and then using the Bartlett and Welch's methods (partitioning the time-series using the degrees-of-freedom as you find more appropriate). Plot the relevant resulting spectra. Compare and discuss the differences involving the number of segments.

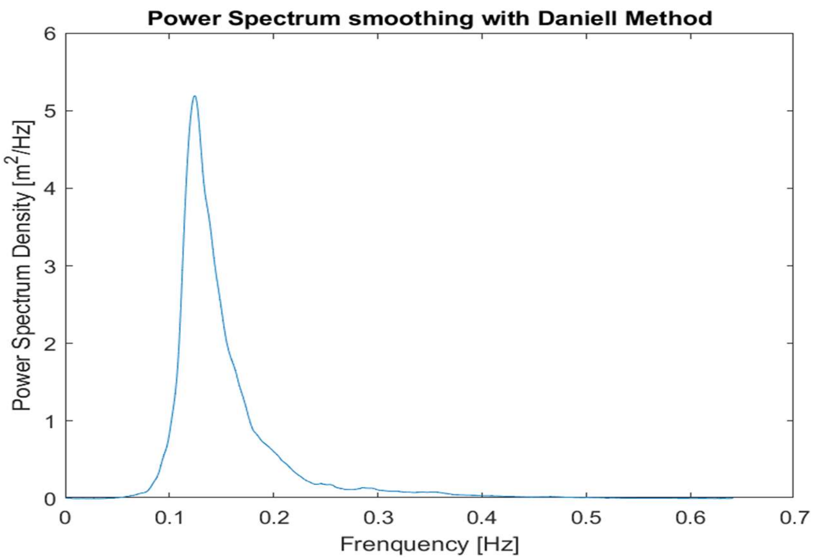
The resulting spectrum is very fractured and clearly susceptible to noise. In signal theory, noise is defined as undesirable and/or unexpected changes in the signal that may occur throughout the data acquisition process owing to capture, transmission, processing, or conversion errors. We want to lower the amount of noise in order to build a smoother spectrum that can then be compared to a theoretical one. To acquire the smoothing, we will "experiment" with the variables examined in MATLAB, such as the number of windows, overlapping, and degrees of freedom.

This "smoothing" procedure eventually leads to a typical problem: data loss. Yes, noise reduction will be employed for this reason, but it must not be overused or we will end up with a spectrum that does not allow for proper data presentation and collecting.

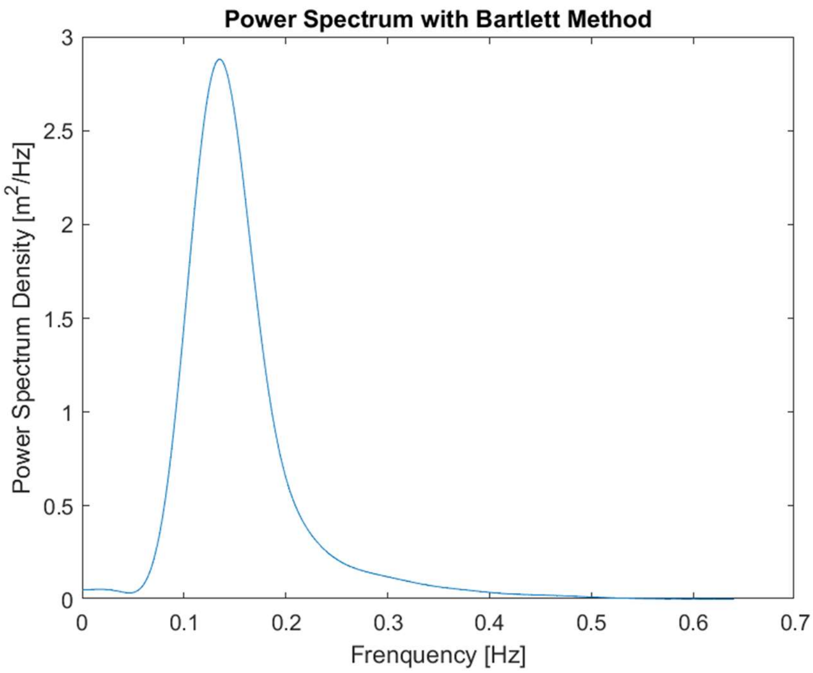
In the **Daniell method**, the variance of the periodogram is reduced by averaging it; but this smoothing of the periodogram causes elimination of the details of the periodogram. So, after calculating the periodogram, it is decomposed to approximate and detailed coefficients using a wavelet transformation:

$$a_k = \left\langle \hat{\rho}_p, \phi_{1,k} \right\rangle = \sum_k \hat{\rho}_p(x)\phi_{1,k}(x)$$

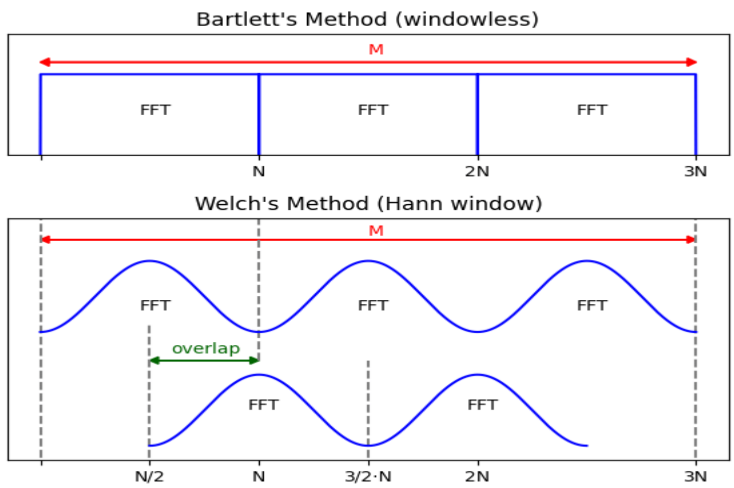
$$d_k = \left\langle \hat{\rho}_p, \psi_{1,k} \right\rangle = \sum_k \hat{\rho}_p(x)\psi_{1,k}(x)$$



Bartlett's approach (also known as the method of averaged periodograms) is used to introduce the noise reduction issue by taking measures that differ from the prior way while maintaining the same essential circumstances. The number of wave surface elevation data points is split into N equal length L segments, reminding that time series analysis is usefull to estimate power spectra. When compared to regular periodograms, it gives a means to lower the variance of the periodogram in return for a drop in resolution. A final estimate of the spectrum at a particular frequency is determined by averaging the estimations from non-overlapping regions of the original series' periodograms (at the same frequency).



A better illustration is indicated in the following picture



Welch's method (also called the periodogram method) for estimating power spectra is carried out by dividing the time signal into successive blocks, forming the periodogram for each block, and averaging.

Denote the m th windowed, zero-padded frame from the signal x by

$$x_m(n) \triangleq w(n)x(n+mR), \quad n = 0,1,\ldots,M-1,\ m = 0,1,\ldots,K-1,$$

where R is defined as the window *hop size*, and let K denote the number of available frames. Then the periodogram of the m th block is given by

$$P_{x_m,M}(\omega_k) = \frac{1}{M}\left|\text{FFT}_{N,k}(x_m)\right|^2 \triangleq \frac{1}{M}\left|\sum_{n=0}^{N-1} x_m(n)e^{-j2\pi nk/N}\right|^2$$

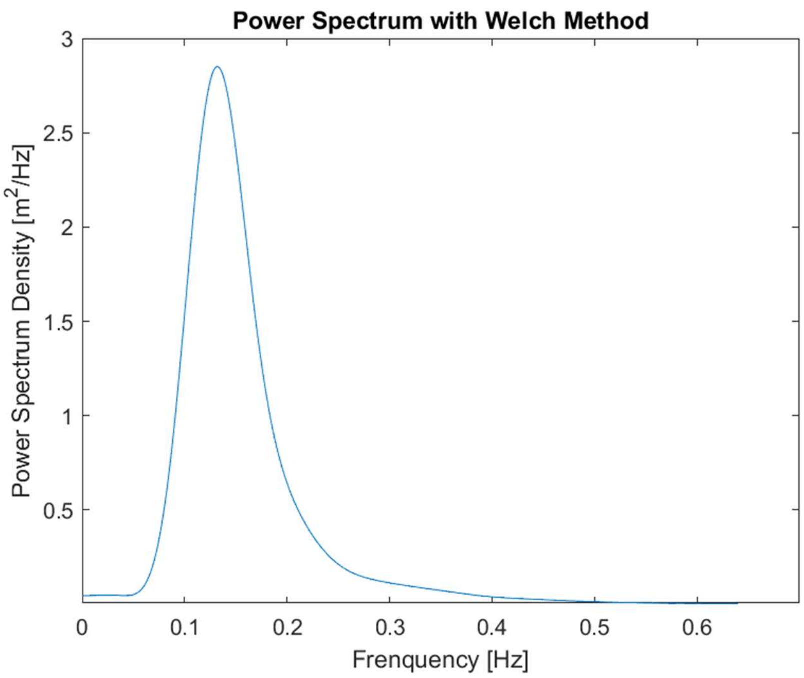
as before, and the Welch estimate of the power spectral density is given by

$$\hat{S}_x^W(\omega_k) \triangleq \frac{1}{K}\sum_{m=0}^{K-1} P_{x_m,M}(\omega_k).$$

In our case is important to analize some important aspects relative to a pratical way of application:

The signal is divided into N overlapping segments of length L, which are overlapped by D points. As a result, if D = L/2, the value we use in our case, the overlap is said to be 50%, which is a reasonable trade off between accurately estimating the signal power and not overcounting any of the data. If D = 0, there is no overlapping, so Welch's method becomes Bartlett's method. In other words, the overlapping segments are then windowed (functions that allow us to disperse the leakage spectrally) - a window function is applied to each individual segment (in our case, we use the Hanning window). As a consequence, the periodogram approach computes each windowed segment by performing a DTFT, squaring the result, and dividing it by L. The average of the calculated periodograms is used, which reduces the variation and therefore the noise. Welch's approach suppresses noise more than the Bartlett method, but it provides less frequency resolution. We used Welch's technique, taking into account the number of degrees of freedom and the effect this parameter has on the length of each segment, and therefore on noise reduction and resolution.

Generally, expanding the number of freedoms reduces the length of each section. After testing with the degrees of freedom by adjusting the number, is possible to estimated and established that 20 is a good quantity for obtaining a suitably smoothed power spectrum with adequate noise reduction and frequency resolution.



4. From final the spectrum estimated (with Welch method), calculate:

a) first 4 spectral moments.

b) spectral bandwidth coefficient, where ν=…

c) peak period TP and significant wave height HS from the spectrum.

d) Compare H1/3 obtained in time domain analysis with HS from the spectrum.

e) Compare the final spectrum with one theoretical spectral model of your choice.

The **Spectral Moments** are a useful mathematical technique extensively employed in signal processing analysis to compress temporal information into a few scalar numbers that can explain the signal distribution. By measuring the mean, standard deviation, skewness, and kurtosis of a signal, the first four statistical moments may be combined to properly define it. The n-th spectral moment of a one-sided (positive frequency) power spectral density is defined as:

$$m_n = \int_0^{\infty} f^n S(f) df$$

And calculated using MATLAB code:

m_0	0.2572
m_1	0.0397
m_2	0.0071
m_3	0.0017

The power spectrum is calculated in the frequency domain. This spectrum is called frequency spectrum because it represents the wave energy distribution with respect to the frequency alone. This distribution of energy comes from the infinite of wavelets with different frequencies and all of them have energy associated.
The **spectral bandwidth** coefficient measures the distribution of energy throughout wave frequencies and may be used for a variety of reasons, including evaluating the efficiency of wave energy converters and forecasting wave groupness. This coefficient may be calculated using the following formula:

$$\nu^2 = \frac{(m_0 m_2 - m_1^2)}{m_1^2}.$$

Spectral Bandwidth Coefficient
0.420054678267518

To calculate and find the **wave height and peak period** of the spectrum we just find the frequency with the highest comparable power spectrum density and apply the method to compute the peak period. According to the following formula, the important wave height is connected to the zeroth spectral moment:

$$H_s = 4 \times \sqrt{m_0}$$

$$T_P = \frac{2\pi}{\{\omega | \max S(\omega)\}}$$

Where T_p is the peak time of the spectrum, (1/frequency of the maximum of spectrum), and H_s is the significant wave height.
The frequency vector is calculated by MATLAB with the discretization of the time series and also the Welch method used to calculate the power spectrum suffers discretization.

Peak wave period and Significant wave height	
T_P $[s]$	7.57
H_S $[m]$	2.03

To investigate the disagreement between the significant wave weight in time domain and frequency domain, a simple calculus was used to determine the discrepancy in percentage of this significant values:

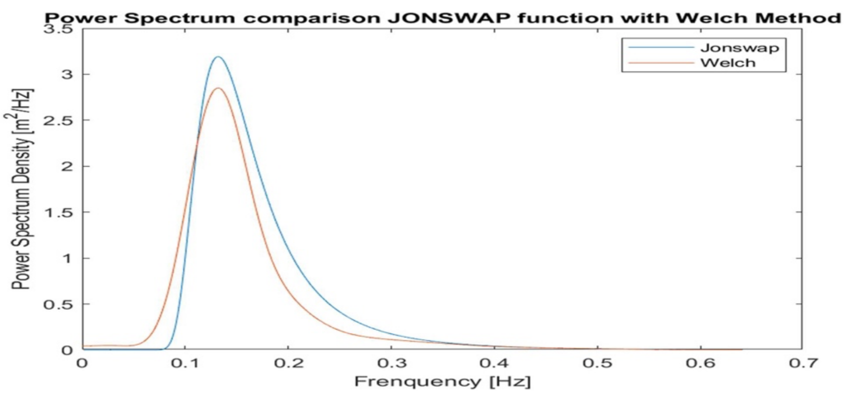
Wave height in time	2.0385
Wave time in frequency	1.8856
Discrepancy %	6.988

Comparing the different methods, in our case JONSWAP and WELCH is possible to realize that the energy spectrum is achieved with the Welch Method and seems to fit with the Jonswap spectrum which is a bit higher than the calculated, because of the peak enhancement factor that defines the sharpness of the spectrum.

One is is a modification on the Pierson and Moskowitz model, which assumed that if the wind blew steadily across a vast region for a long time, the waves would come into equilibrium with the wind. Nonetheless, it was discovered that in actuality, the wave spectrum is never entirely completed; it is continually evolving through nonlinear, wave-to-wave interactions, even over very long timeframes and distances. The empirical "additional peak enhancement" factor was added to the model to capture the reality of non-equilibrium, such that the JONSWAP model is just the Pierson-Moskowitz spectrum multiplied by this factor.

The JONSWAP spectrum function was introduced in MATLAB. The projected power spectrum densities predicted by the JONSWAP approach may then be calculated. Because we have discovered the spectrum using Welch's approach, it is important for the comparison to combine them.

$$S(\omega) = \frac{\alpha g^2}{\omega^5} e^{-\frac{5}{4}\left(\frac{\omega_m}{\omega}\right)^4} \qquad a = -\frac{(\omega-\omega_m)^2}{2(\sigma\omega_m)^2}$$



MATLAB CODE:

Part1

```
close all
clc
file_path ='W200908311900.tsr';
file_data = load(file_path); % Loading buoy data
Heaves = file_data(:, 2); % Heave motion
Time = file_data(:,1);
% Plots
figure
plot(Time,Heaves, 'LineWidth', 1);
title('Heave elevation on time');
xlabel('Time (s)');
ylabel('Water surface elevation (m)');

% Mean, Variance, Skewness and Kurtosis.
m = mean(Heaves);
v = var(Heaves);
skew = skewness(Heaves);
kurt = kurtosis(Heaves);
Table_1 = table(m,v,skew,kurt);
disp(Table_1);
sd = sqrt(v);
% Gauss and Longuet-Higgins relative distributions.
norm_Heave = (Heaves-m)/sqrt(v);
nbins = 100;
hist = histogram(norm_Heave, nbins, 'Normalization', 'pdf', 'DisplayName', ...
['Histogram, bins=', num2str(nbins)]);
xlist = (linspace(-4,4,length(norm_Heave))).';
gaussian = 1/sqrt(2*pi()) * exp(-(xlist).^2/2);
longuet_higgins = xlist;
H3 = longuet_higgins.^3 - 3*longuet_higgins;
H4 = longuet_higgins.^4 - 6*longuet_higgins.^2 + 3;
% Kurtosis
LongHigg = gaussian.*(1+skew/factorial(3)*H3+(kurt-3)/factorial(4)*H4);
figure(1), hold on
plot(xlist, gaussian, 'DisplayName', 'Gaussian', 'LineWidth', 2)
plot(xlist, LongHigg, 'DisplayName', 'Longuet-Higgins', 'LineWidth',2),
hold on
plot([-3, 3], hist)
legend('Empirical', 'Gaussian', 'Longuet-Higgins')
t=[0.78:0.78:1024,5]';
figure(2), hold on
plot(t, Heaves)
xlim([0 1025])
ylim([-1.25 1.25])
legend('')
title('Distribution of free surface elevation')
xlabel('Time [s]')
ylabel('Water surface elevation [m]')

% Comparison between Rayleigh distribution and Up and Down crossing
heave = Heaves-mean(Heaves);
heave = detrend(heave);
heave(heave==0) = 0.001;
heaveblocked = heave;
heaveblocked = sign(heaveblocked);
```

```
up_cross = find(diff(heaveblocked)==2); % Index for the upcrossing;
down_cross = find(diff(heaveblocked)==-2); %Index for the downcrossing
H=[];
T=[];
for indup=1:length(up_cross)-1
Wlength = heave(up_cross(indup):up_cross(indup+1));
H(indup) = max(Wlength) - min(Wlength);
T(indup) = 0.79*(length(Wlength)-1);
end
x = (0:0.1:6);
p = (x/(4*v)).*exp(-x.^2/(8*v));
figure(3), hold on
plot(x,p)
legend('Rayleigh Ritz','Wave heights')
Hhist = histogram(H, 30, 'Normalization', 'pdf');
% Calculation of the significant wave height and period with max&mean
sorted_H = sort(H);
sorted_P = sort(T);
highest_H3 = sorted_H(round((2/3)*length(sorted_H)):length(sorted_H));
highest_P3 = sorted_P(round(2/3*length(sorted_P)):length(sorted_P));

% Iteration for significant wave height and period
HSignificant = mean(highest_H3);
TSignificant = mean(highest_P3);

% H_max and H_mean
H_Max = max(H);
H_Mean = mean(H);
```

Part2:

```
% 2.1 Autocorrelation function
x = Heaves;
dt = 0.78;
tau_i = 0:round(1/5*length(x));
tau = tau_i *dt;
t = 1:length(x)-length(tau_i);
for i=1:length(tau)
R(i) = mean(x(t).*x(t+(tau_i(i))));
end
figure(4), hold on
plot(tau,R)
legend("Autocorrelation function")
xlabel("Tau")
ylabel("Autocorr R")
title("Autocorr of heave")
% Make the time-series a number multiple of 2
% Iteration in helping with the Bartlett and Welchs methods
if mod(length(Heaves),2) == 1
    Heaves(length(Heaves))=[];
end
n = length(Heaves); % Number of samples
dt=0.78; % Sampling interval

% Adjust the settings of Bartlett and Welch's approaches by adjusting the two
parameters below.
n_windows = 4; % Define the number of winows;
percentage_overlapping = 50; % Define percentage of overlapping;
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Do not change the next four rows
n_members_windows = length(Heaves)/n_windows;
noverlap = round(n_members_windows*percentage_overlapping/100);
[S,f]=spectrum(Heaves,length(Heaves),noverlap,hanning(n_members_windows),1/ dt);
S=2*dt*S(:,1);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Variable S contains the spectrum, variable f the frequency (in Hz)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% 2.2 FFT
figure('name','Power Spectrum from FFT');
plot(f,S);
title('Power Spectrum from FFT');
xlabel('Frenquency [Hz]');
ylabel('Power Spectrum Density [m^2/Hz]');
% 2.3 Noise reduction (Daniell, Bartlett, Welch)
% 2.3.1 Daniell's method
S_Daniell = S;
p = 15;
for i = (p + 1):length(S_Daniell) - p
S_Daniell(i) = sum(S_Daniell(i - p:i + p))/(2*p + 1)
end
figure('name','Power Spectrum smoothing with Daniell Method');
plot(f,S_Daniell);
title('Power Spectrum smoothing with Daniell Method'); xlabel('Frenquency
[Hz]');
ylabel('Power Spectrum Density [m^2/Hz]');
% Comparison and change of smoothing (p) with the real spectrum related to FFT
% 2.3.2. Bartlett's method
dgf = 20;
n_windows = fix(length(Heaves)/dgf/2);
% percentage_overlapping = 50;

% n_members_windows = length(Heaves)/n_windows;
noverlap = 0;
[S_Welch,f_Welch]=spectrum(Heaves,length(Heaves),noverlap,hanning(n_windows
),1/dt);
S_Welch = 2*dt*S_Welch(:,1);
figure('name','Power Spectrum smoothing with Bartlett Method');
plot(f_Welch,S_Welch);
title('Power Spectrum with Bartlett Method'); xlabel('Frenquency [Hz]');
ylabel('Power Spectrum Density [m^2/Hz]');
% Comparison
% 2.3.3. Welch Method
dgf = 20;
n_windows = fix(length(Heaves)/dgf/2);
percentage_overlapping = 50;
noverlap = round(n_windows*percentage_overlapping/100);
[S_Welch,f_Welch]=spectrum(Heaves,length(Heaves),noverlap,hanning(n_windows
),1/dt);
S_Welch = 2*dt*S_Welch(:,1);
figure('name','Power Spectrum smoothing with Welch Method');
plot(f_Welch,S_Welch);
title('Power Spectrum with Welch Method'); xlabel('Frenquency [Hz]');
ylabel('Power Spectrum Density [m^2/Hz]');
% Comparison
% 2.3 From the Welch spectrum calculate:
% 2.3.a. First spectral four moments
df = (f_Welch(length(f_Welch))- f_Welch(1))/length(f_Welch);
m0 = sum(S_Welch)*df;
m1 = sum(S_Welch.*f_Welch)*df;
```

```
m2 = sum(S_Welch.*f_Welch.^2)*df;
m3 = sum(S_Welch.*f_Welch.^3)*df;
% 2.3.b Spectra bandwidth
c = sqrt((m0*m2 - m1^2)/m1^2);
% 2.3.c
Hs = 4*sqrt(m0);
[S_max, index] = max(S_Welch);
fp = f_Welch(index);
Tp = 1/fp;
% 2.3.d Comparison different values
% 2.4 Comparison between the final spectrum and the theoretical result

% JONSWAP Spectrum ex.
S_Jonswap = jonswap_spectrum(f_Welch, Hs, Tp);
figure('name','Power Spectrum comparison JONSWAP function with Welch Method');
plot(f_Welch,S_Jonswap);
hold on;
plot(f_Welch, S_Welch);
title('Power Spectrum comparison JONSWAP function with Welch Method');
xlabel('Frenquency [Hz]');
ylabel('Power Spectrum Density [m^2/Hz]');
legend('Jonswap','Welch');
```