OCE 2901: Surf Engineering Analysis -
Spring 2022
MWF: 10:00-10:50 AM
Melbourne Beach, Ocean Ave.

Florida Institute of Technology
High Tech with a Human Touch™

Instructor Information

Robert J. Weaver, Ph.D.
Office: Link 205
Hours: M 11:00 am -noon; T R 11:00 am - 11:50 am; or by appointment
Phone: 7273
Email: rjweaver@fit.edu (preferred method of contact)

This Syllabus is Tentative and subject to modification throughout the semester.

Course Description

OCE 2901: Surf Engineering Analysis

The course focuses on the analysis of the physics of waves in the surfzone. During the field intensive experience students will learn about the physical processes of waves and wave riding through instrument deployment, data collection and analysis. Post processing of the collected field data will expose the relationships between waves, board design and wave rider.

Course Objectives

The objective of this course is to provide the students with experience in field data collection and data analysis, as well as a basic understanding of water waves and the physics behind surfing, including force balances, buoyancy and hydrodynamic drag. Students will not learn to surf. At the end of this course, students should be able to design a field experiment to collect pertinent data, explain what a wave is and how waves are generated, and describe the forces at work on a floating body on the face of a wave.

During the first 2 weeks of classes, all students will be administered a swim test overseen by Dr. Weaver. Students who pass the swim test will be allowed to enter the ocean to collect data either as a swimmer or as a surfer.

Upon successful completion of this course, the student should be able to:

1. Identify common terminology used in the coastal environment,
2. Explain and apply basic coastal hydrodynamic principles,
3. Discuss properties and characteristics of water waves,
4. Qualitatively and quantitatively describe coastal wave environment,
5. Identify engineering applications in the coastal environment,
6. Deploy and retrieve instrumentation in the surfzone
7. Analyze field data using time series analysis techniques

Course Information

Pre-requisites: Calculus

Req'd. Textbook:

- ** Karimpour, A., *FUNDAMENTALS of DATA SCIENCE with MATLAB*. 2020, ISBN-10: 1-7352410-1-6, ISBN-13:978-1-7352410-1-2
- Karimpour, A., *OCEAN WAVE DATA ANALYSIS, Introduction to Time Series Analysis, Signal Processing, and Wave Prediction*, 2018, ISBN-13: 978-0-692-10997-7

** for students without previous MATLAB Experience

Rec. Textbook:

- Butt, T. and Russell, P., *Surf Science*, University of Hawaii Press, 2002
- Emery, W.J. and Thompson, R.E., *Data Analysis Methods in Physical Oceanography*, 2nd Edition, Elsevier, 2004.
- CEM (Coastal Engineering Manual) US Army Corps of Engineers.
- Dean, R.G. and Dalrymple, R.A., *Water wave Mechanics for Engineers and Scientists*. World Scientific, 1991.
- Munson, Young, Okiishi, and Huebsch, *Fundamentals of Fluid Mechanics, 7th edition*, John Wiley & Sons

Field Trips: We will be meeting at the beach nearly every other week when the field data collection begins. Dates TBD.

Course Policies

1. Attendance is Mandatory as the course is based on extensive field data collection requiring the presence of all students. Successful completion of this course will require students to travel to the beach regularly for meetings, and to come prepared having read the appropriate textbook sections ahead of time.
2. The student is expected to come to class prepared:
    a. For field data collection this includes each student providing themselves: sunscreen, hat, sun protection, water, snack, and towel.
    b. For classroom sessions this included reading the assigned material and being prepared to discuss the data.
3. Quizzes will be administered periodically in order to evaluate course preparation.
4. Exams: There will be a total of two (2) exams given during the course, with the third occurring during the scheduled final exam period. The two term exams may consist of an "in class" portion and a "take home" portion, the exact format is to be determined.
5. Make-up exams will be administered only under extraordinary conditions. An alternative examination procedure may be considered by the instructor, who reserves the right to determine what an "extraordinary condition" is or is not.
6. Field reports will be due one week after each field deployment. The report will document the in detail the field experience, including the deployment of instruments, data collection process, and retrieval of instruments.
7. Lab reports will be due one week after field data is downloaded in the lab and provided to the students either via canvas or through direct communication.
8. Homework assignments will be due at the beginning of class

Student Evaluation

A student's final letter-grade will be determined based on the following ten-point scale:
A: 90 – 100%, **B**: 80 – 89%, **C**: 70 – 79%, **D**: 60 – 69%, **F**: < 60%

Student performance will be evaluated through field and lab reports, homework assignments, quizzes, mid-term exam and comprehensive final exam according to the following distribution:

Field/Lab Participation	25%
Homework/Quiz	25%
Mid Term	25%
Final Experiment Report	25%

Academic Honor Code (Plagiarism and Cheating)

Academic dishonesty includes: plagiarism; cheating—giving, receiving, or sharing information during an in-class, take-home, or on-line exam, test, or quiz, using unauthorized material (like notes) during an exam, submitting the same paper (or different versions of what is substantially the same paper) for more than one course or different sections of the same course—fabricating written work, sources, research and/or results; helping another student commit an act of academic dishonesty; and lying to protect another student who has committed an act of academic dishonesty.

According to Florida Tech's *Student Handbook*, "all forms of academic dishonesty, including cheating, fabrication, facilitating academic dishonesty and plagiarism . . . are subject to disciplinary action up to and including suspension or expulsion from the university."

Title IX

The federal law prohibiting sex discrimination in educational institutions is Title IX of the Educational Amendments Act of 1972. Title IX prohibits discrimination on the basis of sex under any education program or activity operated by an institution receiving or benefiting from federal financial assistance. Sexual harassment, which includes sexual violence, is a form of sex discrimination. To report a violation, please contact the Director of Security at extension 8111.

OCE 2901 Surf Engineering Analysis-
MW: 10:00-11:30 AM
Link (217/218)
Melbourne Beach, Ocean Ave.

Florida Institute of Technology
High Tech with a Human Touch™

Tentative Course Schedule

Week	Dates		Topic
	Part One: Introduction		
1		Classroom	Course Introduction – group assignments – Physics of Surfing – schedule SWIM TEST
2		Classroom	
3		Classroom	Focus on instrument deployment and data collection methodology in the field
4		**Field** / Classroom	Field Dry Run (Video)
5		Classroom	Focus on instrument deployment and data collection methodology in the field
6		**Field**	
	Part Two: Field Data Collection		
7		Classroom	Focus on correct deployment / collection and actual data collection
8		**Field**	Data collection
9		Classroom	**SPRING BREAK**
10		**Field**	Data collection
11		Classroom	Data download debriefing Data collection
12		Classroom	Wave
13		**Field**	Data collection
14		Classroom	Dimensional analysis
15		**Field**	Data collection
16		Classroom	Analysis of data and field program
17	Friday May 6th	**FIELD!**	**Final Exam Week Exam**

Field Plan

The field component of this course will involve students being exposed to the surfzone and nearshore region of the Atlantic Ocean at Paradise Park in Indialantic. The worst case scenario is that one of our students drowns; the second worst case is that one of our students gets attacked by a shark. However unlikely, we must prepare for these events and take the steps to make these events less likely.

In order to ensure the safety of the class during the field experiments the following steps are being taken prior to entering the water.

1. Every student enrolled in the course will be required to take a swim test at Clemente Pool. The swim test will include a 400 yd swim, a 25 yd underwater swim, and a 10 minute tread water.
    a. This swim test will identify the students who are weak swimmers and should not be allowed into the water during the data collection. Those students who are not swimming during the field experiment will take the responsibility for the land based tasks.
2. The number of students that will be in the water at any time will be limited to maximum of 4 or 5 depending on enrollment. At all times there will be a land based observer tracking the students in the water collecting the data.
3. At any time the students are in the water Dr. Weaver, a certified AAUS Science Diver /Rescue Diver will be in the water along with a graduate student who is AAUS / rescue diver trained or lifeguard trained.
4. There will always be a land based spotter who is trained in ocean rescue and has the means to lend assistance to a swimmer in need.

A typical field experiment will consist of 6 steps, pre-experiment briefing, instrument deployment, data collection, instrument retrieval, post-deployment debriefing, clean-up.

- During the pre-experiment briefing all safety procedures will be covered, and the land based spotters will be assigned.
- During the instrument deployment the following will take place:
    - 2 submersible pressure gauges will be swam out and deployed. One will be deployed outside the surfzone, and the second will be deployed just inside the surfzone. These instruments will be attached to small cinderblocks which will be tethered to a retrieval buoy. In order to get the gauges out, they will be deployed off of a small seakayak or surfboard.
    - Once the pressure gauges are in place, two instrumented surfboards will be deployed in the surfzone, by experienced surfers.
- During Data Collection experienced surfers will be identified from the enrolled students, to join Dr. Weaver and the experienced graduate students, and local surf expert who will also be participating
    - The surfers will position themselves at the point of incipient breaking, or just outside, and proceed to ride waves, using various styles which will produce different accelerations, measured by the on board accelerometers mounted in the surfboards.
    - While the surfers are collecting their data, they will be filmed by an in-water swimmer equipped with a Go-Pro4 and a personal floatation device.   This students role is to obtain videography of the surfers to determine how much of the board is in contact with the water as the surfer ‘catches’ the wave.
    - There will also be on-land videography using a tripod mounted digital camera, operated by the land based crew.
- Instrument Retrieval will commence after approximately 1.5 hours of data collection.
    - The surfers will return to the shore

    - The strong swimmers will retrieve the submerged pressure gauges.
    - Instruments will be taken to the on-site showers and rinsed, and the boards packed away.
- Once all materials are packed up, a post-experiment debriefing will take place, with special attention paid toward improving the experience for the next deployment
- Clean-up will begin at the beach and finish with the boards put away in the SEA lab Link 217

First Aid kits will be taken with us on each deployment, and cell phones will be ready and charged in order to call 911 in the case of an emergency.

OCE 2901: Surf Engineering Analysis -
MWF: 10:00-10:50 AM
Link (217/218)
Melbourne Beach, Ocean Ave.

Florida Institute of Technology
High Tech with a Human Touch™

SURF ENGINEERING ANALYSIS (SEA); Robert J. Weaver, Ph.D.

https://youtu.be/vUt8fWLqXGg?t=1178 :Video Time 19:38- 23:10 : commercial : 26:52-

This course implements: Physics, Mathematics, Basic Fluid and Wave Mechanics, Data Analysis, Field Experimental Design, Instrumentation

The course, SEA, will provide students with a fundamental working knowledge of data utility and data analysis through a minds-on and hands-on immersive experience, involving surfing on instrumented custom built surf boards to collect the data and subsequent analysis of the data in the lab using modern analysis tools (MatLab, R, etc.).

TIME

The course time will be divided into 2 parts of the semester:

1. Experimental Design
2. Data Collection and Analysis

In the first part, Experimental Design, the students will be tasked with identifying the important variables for understanding the physics of surfing. During the process, they will be introduced to concepts in fluid and wave mechanics and reinforcing the material learned in Physics. A series of questions will be posed:

Curiosity

- Why would one seek to develop a relationship between the surfer and the waves?
- How may society benefit from the results of this experiment?

Creating Value

- What value is there in such a relationship?
- What are the variables involved?
- What data might be needed to develop a relationship between surfer and wave?

Curiosity

- How might that data be collected?

Connections

- What instruments are available to collect that data?
- What methods can be used to deploy instruments and collect the data?
- How might we get the data from the instruments to the computer for analysis?
- What methods are available to analyze the time series data?
- What relationships exist between the surfer and the natural environment?

The students will design a field experiment to collect the data needed to solve the problem posed.

During the second phase the students will split time between classroom lectures and data analysis and field data collection experiments based on the experimental design developed in the first phase of the course.

Throughout the class the experimental design will be tweaked based on the implementation of the plan in the field. At the end of the semester the students will provide a final design for the experiment incorporating lessons learned.

The lectures will cover basic near-shore processes, physics of surfing, and data analysis techniques. Student will learn MatLab through a series of homework and assignments coupled with out of classroom self-guided tutorials.

LOCATION:

The course will use available space on campus for the classroom lectures. Students will have to seek out resources on campus, most likely the library, where they can learn MatLab and complete the data analysis assignments.

During field deployment days the students will meet on campus to prep the field trailer and then class will be held at Melbourne Beach, Ocean Avenue Beach Park. For three hours, the students will employ their field experiment designs to collect the data to be later analyzed in the class.

EDUCATORS:

The role of the professor will be to guide and supervise the students through the learning process. The professor will provide the required background knowledge to get the students moving down the correct path.

ASSESMENT:

The students will be assessed on their ability to design an experiment-apply the experimental design, collect data, and analyze the data. The Course will end with the students submitting a field experiment report including a set of 3 non-dimensional parameters used to elucidate the relationship between the waves and the surfers. Along the way, assignments will be required to assess their learning of the basic time series analysis techniques. Assignments will also provide the students with the opportunity to gain insight into the possible applications of the data that will be collected.

- Participation will be assessed based on each student's ability to self-motivate.

ACTIVITES:

Course activities in the semester start with a Problem Framing Exercise, with the goal of a having the students arrive at an experimental design plan. In the first week the entire class will come together to take a swim test, establishing the role of the individual students during the course.   Brainstorming exercises will used regularly throughout the course.

During the field experiment, students will work in groups to attain a common goal, the collection of field data for later analysis in the class. After each field experiment, the class will meet for a debriefing, covering lessons learned and modifications to the field plan.

On the classroom days, time series data from previous experiments will be analyzed using MatLab as students learn basic techniques:

- Histogram and frequency distributions
- Probability
- Probability density function (PDF)
- Cumulative distribution function (CDF)
- Normal Distributions
- Central Limit theorem
- Estimation of mean and variance linear regression
- Time Series
- Fourier transform
- Ocean wave spectrum
- Signal filtering and windowing

CURIOSITY: Participation points will be awarded for students exhibiting curiosity

CONNECTIONS: The course will require connecting concepts from outside the classroom and seeing how they may apply to the course

CREATING VALUE: There is potential to create value by creating / designing a new instrument array or technique.

Not everyone can be a leader. Leadership must be clearly defined and followed on order to have a successful field experiment.   This must be done not on the day of the experiment, but in advance. Each team must have developed a detailed field experiment that can be followed to the letter while in the field. There should be no deviations from the field plan unless safety issues have been identified, or the entire team can agree at a field meeting.

PREFACE

This course is not about surfing as much as it is about learning to apply physics and science to a particular problem, and analyze the data that are collected. The goal of the course is to introduce underclassmen to upper level concepts in science technology and engineering in a fun and exciting way.

This introduction is accomplished through a field intensive data analysis course focused on the analysis of surf zone waves and surfing. Students develop a field plan in which they will deploy gauges to measure wave conditions, collect dynamic data by riding instrumented surf boards, and use videography to help capture the hydrodynamic parameters needed for analysis. Custom made surfboards are implemented, with instrumentation attached. Every 2 weeks the students manage a 3 hour field deployment at a local beach. The more a surfer grasps the physics behind surfing, the more fun he or she will have on the water, ie, helping a surfer determine where to position their center of gravity on the board. Additionally the data collected in this class could be used to determine optimal board design for any particular surfer to insure they have a fun day surfing for any given wave climate.

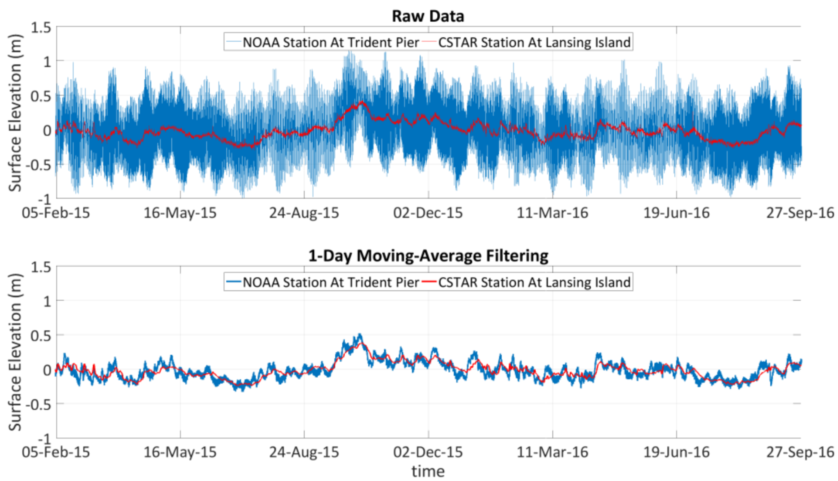
We will be surfing for science!

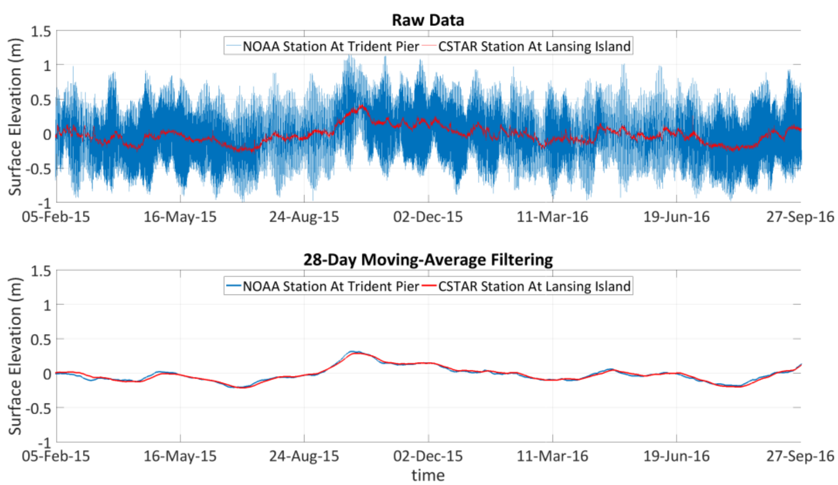
Students will learn basic data analysis techniques for time series analysis, addressing water wave properties and dynamic motion. During the course, students will reinforce the physics and mathematics that they have been taking as an underclassman to help them process the new introductory level information they are exposed to focusing on fluids and wave mechanics.

INTRODUCTION

What is Data Analysis?

Seek to identify relationships between physical (chemical, social, etc.) processes. For this course we look at waves and the interactions between waves and surfers.





Must have data to analyze, this data must be reliable and pertinent to the question that the researcher is seeking to answer.

Random Variable: can have any value from a given set of possible values

We call this a sample, event, or outcome

Types and classes of variables:

1. Nominal variable: has categories but no ranking or order
    - Animal
    - Vegetable
    - Mineral
    - Colors
    - Students randomly seated in a class
2. Ordinal Variable: ordered but no quantitative difference
    - Red is 1
    - Blue is 2
    - Green is 3
    - If I lined the class up and numbered each student instead of having you all randomly sit in a seat
3. Interval Variable: numerical units, #'s but not true zero
    - IQ no real ratio value
    - 150 vs 100, 150 is not 50% greater than 100
4. Ratio variable: units, true zero
    - Can represent with real numbers

Examples:

1. Engineering Programs
2. Groups
3. Sediment type classification

Clay	,	silt	,	sand	,	gravel	,	cobble	,	boulder	,	other
1	,	2	,	3	,	4	,	5	,	6	,	7

4. Sediment Diameter-Equivalent sphere diameter
    - Continuous variable ratio- real zero

$$0 \leq D < \infty$$

i.e. D=0.2mm

5. Wave Height
    a. Continuous variable within an accepted range of validity

Exercise 1:

Give an example not mentioned in class of each of the 4 types of variables listed:

Nominal, Ordinal, Interval, and Ratio

The set of all possible values is the ***Population***:

It is important for the Engineer to be able to identify or ***Define the Population***.

Take Temperature for example, the set of all temperatures is a poorly defined population.

Rather define temperature at a certain place and time. This gives us an Eulerian value of Temperature, a time series of values at a particular location.

Histogram and frequency distributions

Define the population:

The common objective is to determine the population or determine some of the parameters of the population. A population has:

- a true mean μ_x
- a true variance of σ_x^2.

  So we go out and compute the mean of a sample of the population containing N total members or measurements:

$$sample\ mean \equiv \overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$sample\ variance\ \equiv S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2$$

As the number of measurements, N, gets large then:

$$\overline{X}\ \rightarrow \mu_x$$
$$S^2 \rightarrow \sigma_x^2$$

See sample_wave_analysis_01.m figs 1, 11, 2, 3

We can use a histogram to describe a population:

Significant Wave Height, H_s, defined as the average of the largest 1/3 of the waves in a time series, can be estimated as:

$$H_s = 4\sigma_x$$

We describe populations using a histogram. A histogram is a bar plot of the number of occurrences of a particular population that fall into a specified range of values.

Read Section 3.2

Population distribution of the US:
0-10 years old, 11-20 years old, 21-30 years old …etc.

The intervals selected will play a role in determining the shape of the histogram. If there are N samples in the population, and we selected the histogram to N total bins, then is clear to see that each bin will have one sample represented. The frequency of occurrence for that value would be 1/N.

However, at the other extreme, if we have one bin for N samples, then all samples would be contained in that one bin and the frequency of occurrence for any sample being contained in that one bin is N/N or 1.

How to calculate a histogram,

- Order the samples by rank form smallest to largest.
- Find the extreme values (sample min and max)
- Calculate the Range of the samples, $R = max - min$
- Divide R, into classes or ***bins*** (usually 10-20)
    - # of bins ≡
        - $k = 1 + 3.32 \log_{10} N = 1 + \log_2 N\,; for\ N\ \geq 30 \text{ Sturges's Rule}$
        - $k = 2N^{\frac{1}{3}}\,;\ \text{Rice's Rule}$
- Draw bars or lines with a height equal to the number of samples in each bin.

Information from histograms:

- Unimodal or bimodal?
- Skewed?
- Symmetric?

Distributions

Discreet distributions represent a population whose possible number of samples is countable:

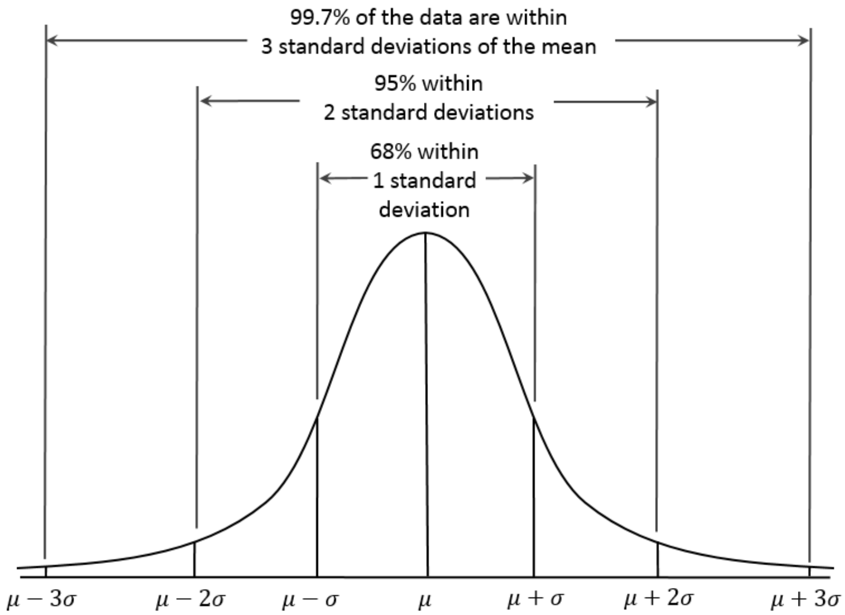
- Results from a coin toss (either head or tail)
- Roll of the Dice (one through 6)

Continuous Distribution

A population with an infinite number of possible sample values

- Height of the people in the US
- Atmospheric pressure
- Wave height

For a Normal Distribution:



Recall: compute the mean and variance of a sample of the population containing N total members or measurements:

$$sample\ mean \equiv \overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$sample\ variance\ \equiv S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2$$

As N gets large:

$$\overline{X}\ \rightarrow \mu_x$$
$$S^2 \rightarrow \sigma_x^2$$

Probability

$(x, f(x))$, represents a probability or frequency distribution, $f(x)$, is the probability. It is the chance that a value will occur within a given population.

$f(x)$ is always greater than or equal to zero and the sum of all probabilities is equal to 1.

$$P(x) \geq 0 \; or \; f(x) \geq 0$$

Discreet Distribution | Continuous Distribution

$$\sum P(x) = 1 \; or \; \int f(x) = 1$$

Discreet Distribution | Continuous Distribution

Probability that a sample taken from the population of discreet random variable, X, is equal to some value, x, is given by the following notation:

$$P(X = x) = f(x)$$

Probability

$(x, f(x))$ is the Probability of x given a population.

$$f(x) \geq 0$$
$$and$$
$$\sum_{all\,x} f(x) = 1$$

$$P(X = x) = f(x);$$

$$the\ probability, P, that\ an\ observation, X, is\ equal\ to\ some\ value, x, is\ f(x)$$

$(x, F(x))$ is the Cumulative Distribution for x given a population.

$$0 \leq F(x) \leq 1$$

$$P(X \leq x) = \sum_{all\,X \leq x} f(x) = F(x);$$

$$the\ probability, P, that\ an\ observation, X, will\ be\ less\ than\ or\ equal\ to\ some\ value, x, is\ F(x)$$

For a given sample data set the probability of any given value or range of values is the number of occurrences, n, divided by the total number of samples, N.

$$P = \frac{n}{N}$$

Ex. : Probability that a wave height is greater or equal to some value is given by:

$$P\left(H > \hat{H}\right) = \frac{n_{H>\hat{H}}}{N} \qquad or\ P\left(H \leq \hat{H}\right) = 1 - \frac{n_{H>\hat{H}}}{N}$$

$n_{H>\hat{H}}$ = number of waves for which H>$\hat{H}$

For a continuous distribution, i.e. the population of values is infinite and equal to the points on a line segment, $P(X = x) = f(x) = 0$. The probability of any one exact value is zero since there are infinite alternate values near the selected value.

Ex. Height of a person question how is $f(x) = 0$? Because we are looking for an exact height to n significant digits. If we round a person's height to the nearest inch then we can talk about finite probabilities.

Take example from water level data. Plot data in MATLAB.

See sample_wave_analysis_01.m figs 31,

Because of this we think of probabilities in term of ranges or intervals.

$$P = \int_{x-\Delta x}^{x+\Delta x} f(x)$$

Using PDF and CDF we can organize data and begin to analyze the data. The raw data can be used to fit a distribution.

For a given sample data set the probability of any given value or range of values is the number of occurrences, n, divided by the total number of samples, N.

$$P = \frac{n}{N}$$

Probability *mass* function (PMF)

The PMF is a representation of a set of ***discreet*** probability distribution of random variables

Probability *density* function (PDF)

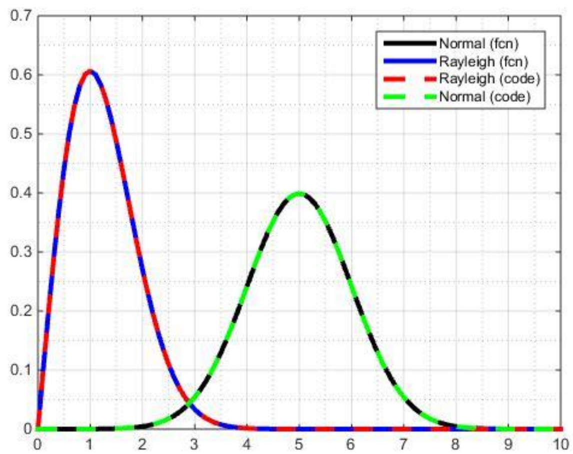
The PDF is a representation of a set of ***continuous*** probability distribution of random variables

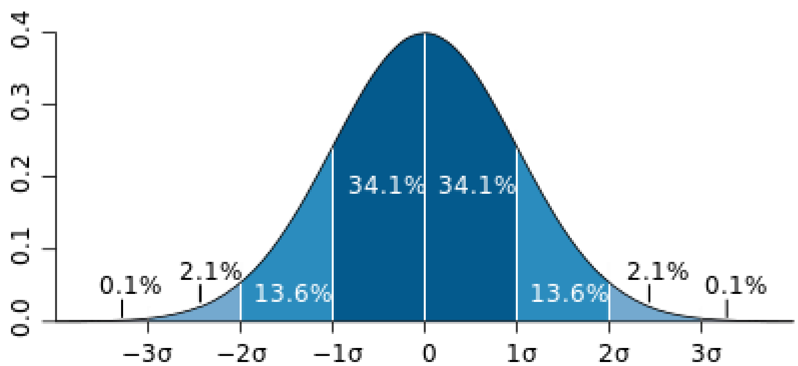
Probability that a sample taken from the population of continuous random variable, X, is equal to some value, x, is given by the following notation:

$$P(X = x) = f(x)$$

Shape of the PDF provides insight into the nature of the population.

Skewed left, more values in the smaller range, skewed right (J shaped) more values likely in the higher range.





Cumulative *distribution* function (CDF)

The cumulative probability function, or Cumulative distribution function (CDF), $F(x)$, is derived by examining the population and sorting the values based on the probability that a sample member will be less than a certain value.

The CDF is the fraction of the total sample that is less than or equal to a specified value, x.

$$F(x) = P(X \le x) = \sum_{X \le x} P(x) \text{ OR } \int_{-\infty}^{x} f(x)$$

It is the sum of all probabilities of values less than x occurring in the population.

$$F(-\infty) = 0 \text{ and } F(\infty) = 1$$

Assignment 3:

Roll a two of dice 100 times, and record the outcome of each die seperately in a table. Using MatLab, plot the histogram of the data and a CDF. Repeat the exercise; again roll the two dice 100 times. This time combine the outcomes with the first 100 outcomes and regenerate the histogram and CDF. Does the histogram change? Does the CDF change? Does your plot match Figure 3.3?

Uniform PDF

Every value between two limits is equally possible:

$$\begin{aligned} f(x) &= \frac{1}{x_2 - x_1} \quad ; for\ x_1 \le x \le x_2 \\ &= 0 \quad\quad\quad ; for\ x < x_1, x > x_2 \end{aligned}$$

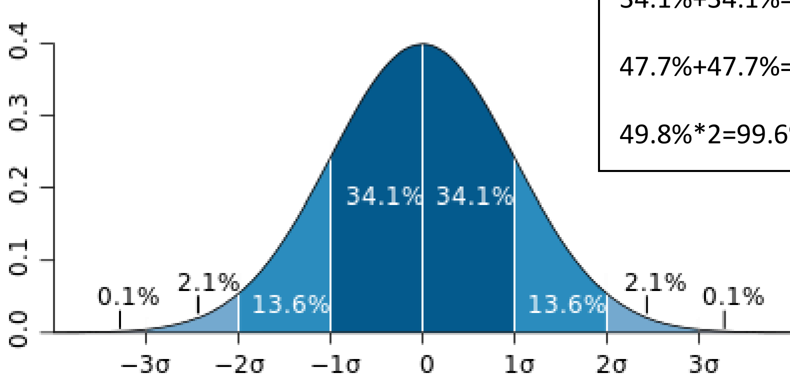
$$\begin{aligned} F(x) &= \frac{x - x_1}{x_2 - x_1}; for\ x_1 \le x \le x_2 \\ &= 1 \quad\quad\quad ; for\ x > x_2 \end{aligned}$$

Normal Distributions

The Normal or Gaussian distribution is a very common distribution that seems to capture natural occurrences very well.

The shape is symmetric about the mean and asymptotic toward $\pm\infty$.

This distribution is valid when many different independent influences contribute to stochastic variations.



34.1%+34.1%=68.2%

47.7%+47.7%=95.4%

49.8%*2=99.6%

Estimation of mean and variance linear regression

Data Acquisition and Recording

OCE 2901 Rea*d*:
Chapter 1 & 2 Karimpour

OCE 5586 Rea*d*:
Chapter 1 in Emory and Thompson: Sections 1.1, 1.2, 1.5, 1.6, 1.8 and anything in between or after which interests you.

Sampling interval and Nyquist Frequency

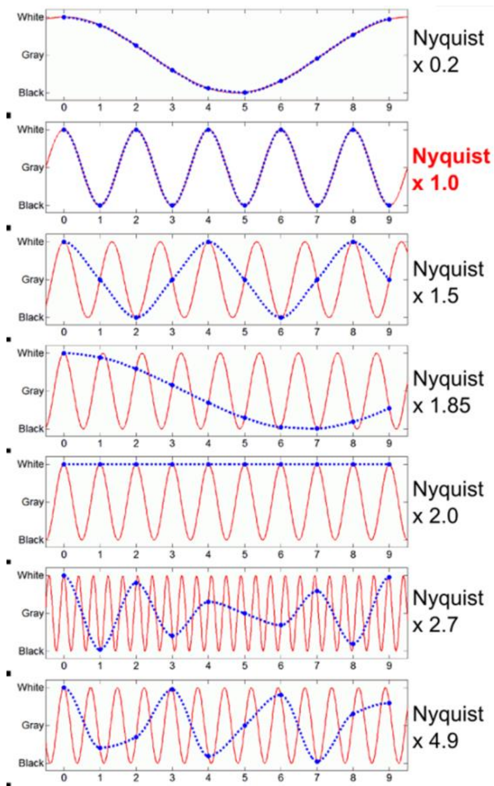
The limiting frequency (shortest wave or fastest response) at which a wave can be adequately resolved is called the Nyquist frequency, or folding frequency:

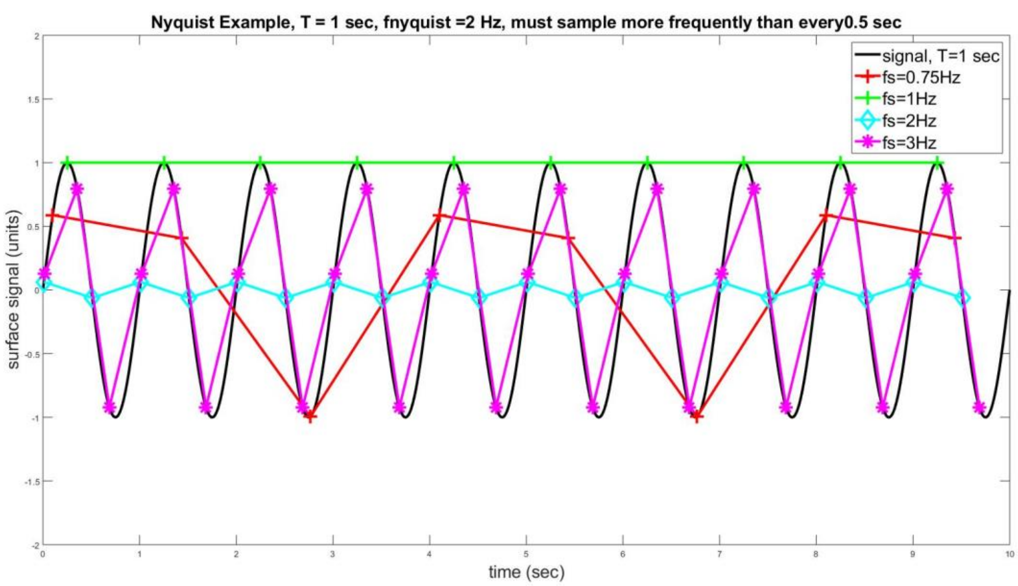
$$f_{Nyq} = \frac{1}{2\Delta t}$$

Nyquist_Example.m

$T=2\Delta t$ $\sigma=\pi/\Delta t$

For any higher frequency waves, (T smaller), the sampling rate is inadequate. **Researcher must design Δt to avoid aliasing.**





"As a general rule one should plan a measurement program based on the frequencies and wavenumbers of the parameters of interest over the study domain. " (E&T, sec. 1.2.1, pg. 4)

Instruments ***must*** be selected based on their abilities to adequately resolve the desired characteristic.

Sampling Duration

The duration of our sampling will determine the frequency resolution for the lowest frequency (longest wave or slowest response) that can be resolved (fundamental frequency), Δf or f_0. This is determined by the Δt and number of records.

$$\Delta f = f_0 = \frac{1}{T} = \frac{1}{N\Delta t}$$

Sample Space (Frequency Domain)

In theory we should be able to resolve all frequencies, f, between the fundamental frequency and the Nyquist frequency:

$$f_0 \leq f \leq f_N$$

Measurements length and interval will be determined by the capabilities of the instruments available. Should the instruments not be ideal for the characteristics that we are seeking to measure, one option is to measure at the highest frequency available (as often as we can) for as long as possible and then use data analysis techniques available such as averaging, smoothing sub-sampling, etc. to improve the statistics of the data collected.

Stastistics

A stochastic process is a set of random occurring variables defined on a population; a family of random variables. Many stochastic processes can be represented by time series. However, a stochastic process is by continuous while a time series is a set of discreet observations indexed by integers. A stochastic process may involve several related random variables.

A stochastic process is said to be *ergodic* if a sufficiently long finite random sample can be used to determine its statistical properties.

A process in which every sequence or sizable sample is equally representative of the statistical parameters of the whole.

Sampling Techniques

Continuous sampling vs. burst sampling,

Most data analysis methods have been developed to accept data at regular spaced sampling intervals.

At the root of data analysis are the concepts of data accuracy/precision, both spatial and temporal resolution, and statistical significance (the “so what” factor, statistical sampling theory).

In order to ensure the statistical significance of the data, data samplings must be taken such that each sample can be considered an independent realization of the system at that time.

If a group of measurements is highly correlated in space or time, then they cannot be independent; and a group of measurements that is completely uncorrelated must be independent of each other.

The degree of correlation can be an indicator for how two measurements are related. Often shifting one measurement in time would produce a more correlated relationship in that case there is a lag between the two data groups.

Chapter 2: Experimental Design

For this course we focus on collecting wave data, sea level data, bottom contour data, and dynamic data associated with the movements of the surfboard. Each of these data are required in order to develop a relationship between the surfer and the waves.

The techniques learned; however, will translate to any data collection experiment in which you may find yourself involved.

Design a Field Data Collection Experiment that will elucidate the connections between the nearshore wave climate and the dynamic motions of a surfer.

In this first part, Experimental Design, the students will be tasked with identifying the important variables for understanding the physics of surfing. During the process, they will be introduced to concepts in fluid and wave mechanics and reinforcing the material learned in Physics. A series of questions will be posed:

Curiosity

Creating Value

Curiosity

Connections

- Why would one seek to develop a relationship between the surfer and the waves?
- How may society benefit from the results of this experiment?
- What value is there in such a relationship?
- What are the variables involved?
- What data might be needed to develop a relationship between surfer and wave?
- How might that data be collected?
- What instruments are available to collect that data?
- What methods can be used to deploy instruments and collect the data?
- How might we get the data from the instruments to the computer for analysis?
- What methods are available to analyze the time series data?
- What relationships exist between the surfer and the natural environment?

The students will design a field experiment to collect the data needed to solve the problem posed.

The FIRST step in any experiment (or any other task in life) is to establish a well-defined ***population***

How do we do this?

Much as when we make a purchase with cash (anyone anymore?) we for a rough sum of the cost prior to checkout, and have some idea of the values before paying. Then we select the bills that would be needed and estimate the change, to be sure that we do not get ripped off.

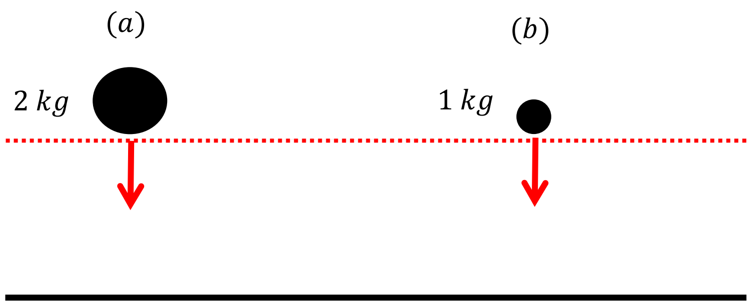
For any experiment we must have an idea (understanding) of what are about to measure and what the expected results will be. How do we accomplish this?

We first use simplified (linearized) techniques to develop an analytic solution that can be solved. We understand that our measured solution will deviate from this simplified solution within some established error range. We can obtain an order of magnitude solution for what we should be expecting to measure.

Basic Physics:

$$F = ma$$

Given two spheres of differing sizes dropped from the same height, which is traveling faster when they reach the bottom?



We know from physics that the two balls will reach the bottom at the same time. That is the acceleration felt by both spheres is the same, $g = 9.81\ m/s^2$. Covering the same distance in the same time means that they are travelling at the same velocity.

$$V = V_0 + at$$
$$a = g = 9.81\ m/s^2$$

$V_0 = 0$ for both a and b, therefore the velocity is only a function of acceleration and time.

However the forces acting on each ball are not the same.

$$F_1 = m_1 a \quad and \quad F_2 = m_2 a$$

$$F_1 = 2kg * g \; and \; F_2 = 1\ kg * g$$
$$F_1 = 19.62\ kg\ m/s^2 \quad and \quad F_2 = 9.81\ kg\ m/s^2$$

Now consider the classical physics problem of a simultaneous fall. There are two ***identical*** balls, held at the same elevation above the ground. The balls are released at the same time. One ball is allowed to drop straight down to the ground and the other is released with an initial horizontal velocity, thus taking the ball into an arched trajectory. Recall from your physics lab that the two balls impact the ground at the same time!
This demonstration is used to show that horizontal and vertical motion is independent.
http://www.upscale.utoronto.ca/GeneralInterest/Harrison/Flash/ClassMechanics/TwoBallsGravity/TwoBallsGravity.html

The velocity and the elevation are related by the Bernoulli Equation. This is an energy conservation equation:

$$\frac{p}{\gamma} + z + \frac{V^2}{2g} = Const.$$

For the case of a free surface the absolute pressure experienced is the same along the path of the object, $p_{abs} = p_{atm}$, in our case atmospheric pressure. We will use gage pressure, $p_{gage} = p_{abs} - p_{atm}$, which is equal to zero at the free surface.

$$z + \frac{V^2}{2g} = 0$$

The total energy is a function of the momentum, multiplying by the mass and rearranging the equation we get:

$$mgz + \frac{1}{2}mV^2 = 0$$

Or conceptually, the potential energy + kinetic energy is constant, energy conservation.

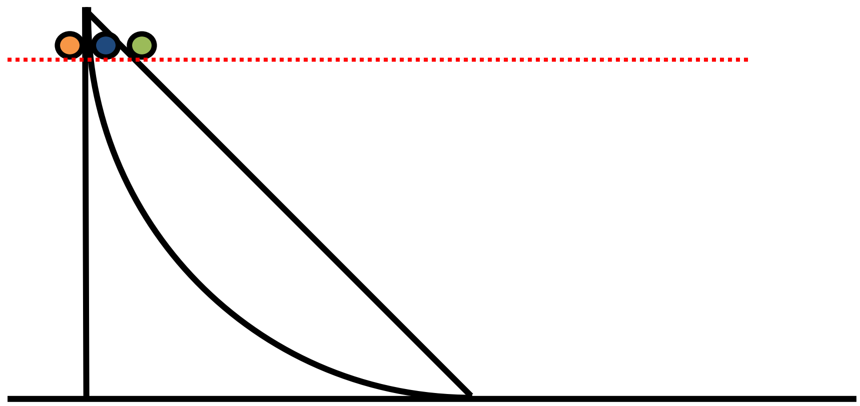
$$mgH = \frac{1}{2}mV^2$$

Note that the velocity at the end of the path is independent of the path taken, if we can ignore friction and drag. Velocity is only dependent on the initial height.

Time on the other hand can be a different story. The time it takes for an object to reach the bottom is dependent on the acceleration (gravity), as well as the angle of the path taken. For an object in contact with a surface, the gravitational component normal to that surface is countered by an equal and opposite force. For a floating body that force is the buoyant force.

Which ball is going faster when it reaches the bottom?

Which ball reaches the bottom first? (Brachistochrone problem)



If we first think about the conservation of energy, we see that the speed (kinetic energy) is governed only by the elevation (potential energy), and since they all start from the same height, the velocity at the bottom will be the same (neglecting friction).

In order to look at the time it takes to get to the bottom, we must revisit the equation for velocity.

$$V = V_0 + at$$

Both the initial velocity and the final velocity are the same for each case. To understand the time we must decompose the equation into the vector components.

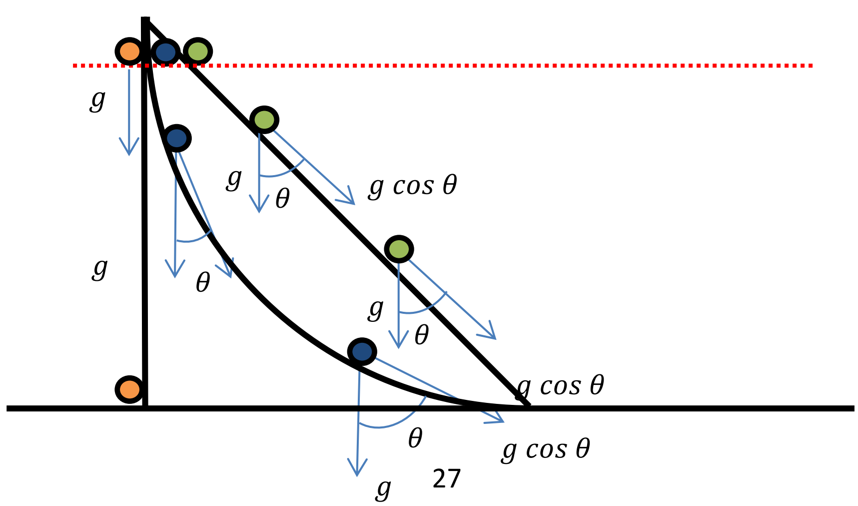
$$V = v_x\hat{\imath} + v_z\hat{k} \ and$$

$$a = (a_x\hat{\imath} + a_y\hat{k}) = a(\cos\theta\hat{\imath} + \sin\theta\hat{k}) =$$

$$g\cos\theta\hat{\imath} + g\sin\theta\hat{k}$$

$$v_x\hat{\imath} + v_z\hat{k} = v_{x0}\hat{\imath} + v_{z0}\hat{k} + (a_x\hat{\imath} + a_y\hat{k})t$$

$$v_x\hat{\imath} + v_z\hat{k} = (g\cos\theta\hat{\imath} + g\sin\theta\hat{k})t$$



- For the case of the first ball (orange), $\theta = 0$; therefore, $g \cos \theta = g(1) = g$, for all time, t.
- For the case of the third ball (green), $\theta \neq 0, \theta = C,$ some constant angle; therefore, constant acceleration less than that of gravity, $g \cos \theta < g$, for all time, t.
- For the middle case, the blue ball, $\theta \neq 0$ and $\theta \neq C$, the angle is initially zero, then non-zero and non-constant throughout the time t. Again we see that, $g \cos \theta < g$, however since θstarts out smaller than C, the ball drops faster than the green ball. Non-constant acceleration.

The waverider most closely follows the middle case.

In order to calculate the velocity of the surfer, we need the wave height, H. A dynamicist named McCowen in 1894, found a relationship between the breaker height, H_b and the water depth at breaking, h_b:

$$H_b = \kappa h_b; where\ \kappa = 0.4 - 1.2 \ \approx 0.78$$
$$H_b \approx 0.78\, h_b$$
$$h_b \approx 1.28 H_b$$

To simplify the problem, first let's assume that the we can take a snapshot of a wave just before breaking, this waveform can serve as a stationary waveform which the surfer will ride form the top of the wave to the trough.

From before we have our energy balance:

$$mgH = \frac{1}{2} mV^2$$

Solving for V we obtain:

$$V = \sqrt{2gH_b}$$

And thus the velocity of the surfer can be solved based on a given wave height. For typical waves in Florida we get bottom velocities given in Table 1.

Table 1: Estimated maximum velocities for typical waves observed on the Florida coast are provided in common units systems. These estimates assume the wave is stationary.

H(m)	V(m/s)	V(mph)	V(kph)
0.5	3.1	7.0	11.3
1.0	4.4	9.9	15.9
1.5	5.4	12.1	19.5
2.0	6.3	14.0	22.6
2.5	7.0	15.7	25.2
3.0	7.7	17.2	27.6
3.5	8.3	18.5	29.8
4.0	8.9	19.8	31.9
4.5	9.4	21.0	33.8
5.0	9.9	22.2	35.7

For those surfing with the latest tech, we can ground truth our estimates by comparing the recorded data with that in Table 1.

In nature waves are progressing toward the shore, on their way to certain demise. And before their lives extinguish, they let out one last burst of energy. The waves break, releasing their energy to the ocean and sea floor. We can calculate the speed of a wave moving toward shore. And once in the surf zone, once they are just about to break, we can estimate their speed by treating the solution as a shallow water wave solution.

For shallow water waves it can be shown that the wave speed, Celerity, C, is only a function of the water depth:

$$C_{sw} = \sqrt{gh}$$

$$C_b = \sqrt{gh_b} = \sqrt{1.28gH_b} = 1.13\sqrt{gH_b}$$

Using this relationship, we can calculate the forward speed of the wave as a function of the wave height.

This adds an extra term to our Energy balance. In order to 'catch' the wave a surfer must first 'catch-up' to the wave, that is, the surfer must be moving at a speed that is at or near the wave speed, celerity. Instead of starting from rest, with KE=0, the surfer has an initial momentum. This can be given by:

$$KE_I = \frac{1}{2}mV_I^2 = \frac{1}{2}m{C_b}^2$$

At the crest of the wave just as the surfer is about to begin the 'drop', we can now again balance the energy:

$$mgH_b + \frac{1}{2}m{C_b}^2 = \frac{1}{2}mV_s^2$$

Solving for the surfer velocity:

$$V = \sqrt{2gH_b} + C_b$$

$$V = \sqrt{2gH_b} + \sqrt{1.28gH_b}$$

$$V = \sqrt{3.28gH_b}$$

$$V = 1.81\sqrt{gH_b}$$

Using the formula for shallow water wave speed, we can modify Table 1 to include the forward motion of the wave. Results are presented in Table 2.

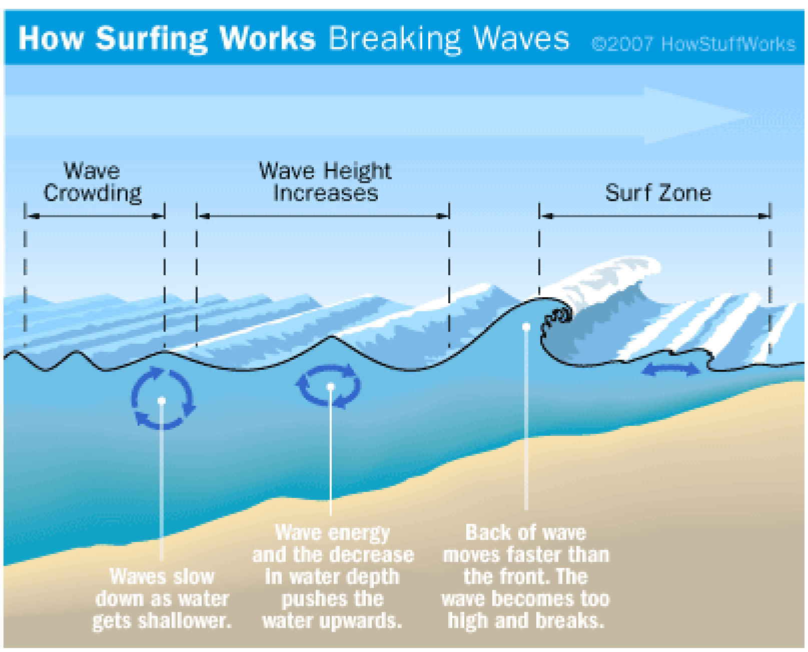
Table 2: Estimated maximum velocities for typical waves observed on the Florida coast are provided in common units systems. These estimates assume the wave is moving at the shallow water wave speed.

H(m)	Cb(m/s)	V(m/s)	V(mph)	V(kph)
0.5	2.5	4.0	9.0	14.4
1.0	3.5	5.7	12.7	20.4
1.5	4.3	6.9	15.5	25.0
2.0	5.0	8.0	17.9	28.9
2.5	5.6	9.0	20.1	32.3
3.0	6.1	9.8	22.0	35.4
3.5	6.6	10.6	23.7	38.2
4.0	7.1	11.3	25.4	40.8
4.5	7.5	12.0	26.9	43.3
5.0	7.9	12.7	28.4	45.7

Again, these values can be tested using the instrumentation deployed during the field deployments.

Using basic physics and wave theory, we can predict the velocities of the surfers as they drop to the bottom of a wave. In reality the bottom drop may not be the most fun ride. In general there is any number of paths down the face of a wave that a surfer may choose to take. If the waverider needs more speed, all they must do is shift their weight to direct the board to travel down the face instead of along it.

The above equations assume that the surfer is dropping in and riding down the face of the wave, and neglects the drag from form and friction. In reality the surfer will be traveling along the face



of the wave and the path will depend on the peel angle of the breaking wave. The top speed for the surfer will reduced by the friction and form drag.

In a field experiment performed According to Dally (2001), the maximum sustainable speed for a surfer was found to be $6.03H_b$. This approximation is based on videogrammetry, analysis of the video of surfers.

Exercise:

For a given wave height the Dally equation produces a larger speed, why?

Acceleration is a function of time squared. The longer the surfer stays on the wave 'falling' under the pull of gravity, the faster the surfer will move.

Care must be taken in the experiment design in order to produce meaningful comparisons.

Experiment Design

Field Deployment:

A key component of this course is the development and execution of a field data collection plan. Designing a successful field deployment requires an in-depth understanding of the physics behind the data that is to be collected. Questions must be answered before the planning can begin:

- What is the question that is to be addressed?
- In general what is the physical phenomenon that is to be better understood?
- What is the specific data can be collected to help answer the question?
- What instrumentation is available to collect the data needed?
- What is the frequency of sampling that would be needed?
- What are the dangers that the operators may encounter during deployment, collection and retrieval?
- What is to be done with the data once it is downloaded?

Once the "What" has been answered then we must answer the "How":

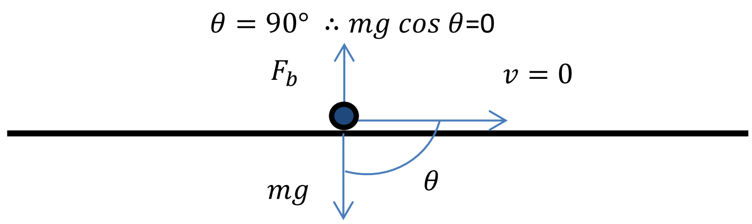
- How can we relate the important variables?
- How can this instrumentation be deployed and then later retrieved?
- How can the data be best processed to tease out the desired relationships?

The design of the field experiment is one of the main goals of this course.

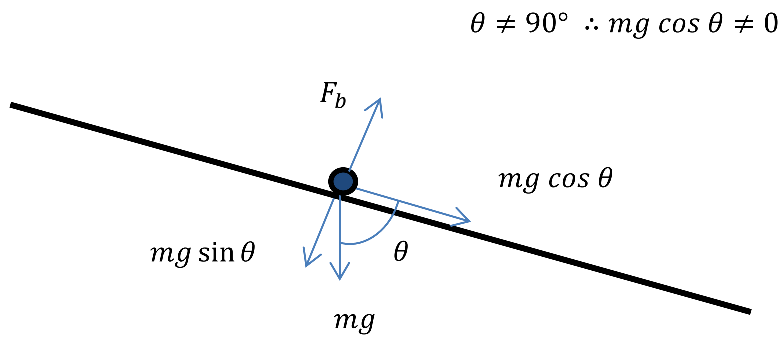
- **Initial field plan**
- **Updated field plan**

1. **Data Collection**
    - a. **Waves**
    - b. **Inertial data**
2. **Data Analysis**
    - a. **Waves**
    - b. **Inertial data**
3. **Final Field Plan**

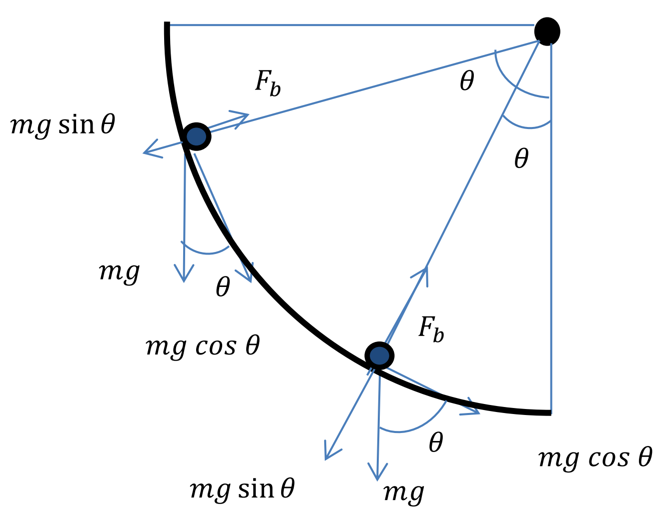
Let's now examine the force balance on a stationary horizontal plane:

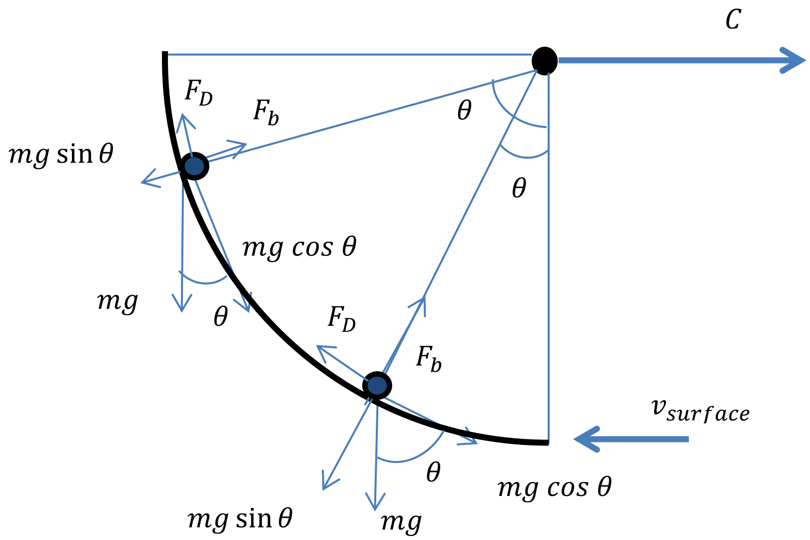


Then extend this to a stationary downward sloping plane:



Finally look at a curved surface:





Now take the movement down the curved surface and add in the moving reference frame and consider the surface particle velocities and the associated drag force

Quickly the problem can become quite complex. We handle this by decomposing the complex problem into the individual components and simplify the problem to solve for the variables of interest.

For example, the accelerations are needed in order to solve the force balance.

In the end we seek to develop a relationship between the variables in term of dimensionless parameters that account for the ***important*** scales that influence the problem. Who says what scales are important?

In order to arrive at such a parameter, we must examine each variable that may affect the system, and then determine the relative influence of those variables. A select few will rise to the level of most important, and those we will attempt to relate. This process is called Dimensional Analysis and you will learn more about this in your junior level Fluids course. Each dimensionless term generated is termed a Pi term (from the Buckingham Pi Theorem).

Project Variables:

One of the first steps during the design process is the determination of the relevant variables needed to address the problem of interest.

Exercise:

What can we measure?

- Dimensions of the surfboard
- Speed of surfer
- Accelerations in 3D
- Direction in 3D
- Hydrostatic Water Pressure
- Atmospheric pressure
- Weight of the surfer/surfboard

What can we calculate?

- Wave speed (Celerity)
- Some Forces
    - Analytic acceleration (frictionless)
- Total drag (difference between analytic velocity and actual velocity)
- Velocity of surfer
- Wave period
- Wave length
- Wave height
- Wave steepness
- Average bottom slope
- Surf similarity parameter (Iribarren Number)
- Water plane area

Can we guess as to which variables we may wish to relate, in Pi terms, at this time?

- Iribarren Number
- Surf board volume
- Water plane area
- Surfer/surfboard weight

Let us now think about how the real world situation of surfing is more complicated than the simple analytical scheme illustrated here:

What complexities can you think of?

- Surf board not parallel to the water surface
- Wave making drag as the rear end of the board is pushed into the wave face essentially "plowing" through the water.
- Wind drag may be significant
- Form and skin drag not measured
- Etc:

Instrument Calibration

Just as important as collecting the data, proper instrument calibration must be completed prior to deployment.

Exercise:

How will we calibrate the instruments required for this experiment?

Assignment 2:

Calibrate the instruments that will be used in this experiment. Split up into groups and have each group calibrate all instruments of a particular type.

Develop an instrument calibration experiment for each of the instruments to be used during deployment.

Develop the Field Experiment Design

Now it is up to you to begin the design process:

Things to remember:

- ***Safety FIRST!***
    - Absolutely the one most important factor is that no one gets injured or killed. We must be diligent in our design to incorporate safety at every step.
    - For every action that you propose in your design, you need to include a safety analysis, be sure to indicate all possible mechanisms for accident or injury and how you propose to mitigate the danger.
- Clear direction
    - Every step of your experiment design should have a clear direction for the researchers (students) to follow. There should be no ambiguity.
    - Indicate how you would eliminate ambiguity in the execution of each step..
- Clear purpose
    - Each step of the experiment needs to have a clearly articulated purpose. This will ensure that entire team is working toward the same goal.

Assignment 3:
Draft an experiment design to collect the necessary data in order to develop a relationship between the dynamic motions of a surfer and the incident wave climate.

Example dimensionless parameters :

- Forecasted wave coefficient:

$$H_{fwc} \equiv \frac{H_{measured}}{H_{forecast}}$$

For $H_{fwc} > 1$, the surfing is better than expected

- Steepness coefficient:

$$S_p \equiv \frac{\frac{H}{L}}{P}$$

$$relates\ the\ pitch\ of\ the\ board\ to\ the\ wave\ steepness$$

- Drop Sickness Coefficient

$$DSC \equiv \frac{a_z}{g}$$

- Surfer paddling coefficient

$$SPC \equiv \frac{V_{surfer}}{C}$$

DATA PROCESSING

Read Chapter 2 (Emory and Thompson)

The data collected by instrumentation is generally not available in the desired format. The electrical signal recorded must be converted to provide the desired information. During this process the instrument must be calibrated with known values to be sure that the conversion in not erroneous.

The calibration is generally performed in a controlled laboratory setting, and will be the focus of the session.

Once calibrated, the exact start and end times of deployment need to be recorded. This may be handled differently for the different instruments used. Ideally there are multiple methods used which will aid in the subsequent data processing. The most basic method of accomplishing this task is to assign the task of recording the start and stop times to one of team members. The only job this team member is to diligently take these records. We can determine the times by large shifts in the values of the data being recorded. Using our knowledge of the system and the parameters being tested, we can tease out the start and end times based on recorded values that are out of the range of the rest of the signal, provided that these are not anomalous data spikes. *For instance when deploying the submersible pressure gauge, we can accurately find the time of deployment by identifying the location in the signal* *where the pressure changes from atmospheric to hydrostatic, this time of the large pressure increase signals the time when the gauge is placed on the sea floor.*

Once the data is collected the researcher must perform a QA/QC (Quality Assurance/Quality Control), ***assure the quality of a product and ensure products and services meet expectations.***

This step is critical. During this step the data must be assured to have no anomalous errors by removing data spikes, and appropriately handling gaps in the data or time when the instrument has failed. Removing the data spikes requires a deep understanding of the data which is being collected, and how the system is expected to behave. Additionally, noise in the signal should be addressed during this step. The noise is handled by using a variety of statistical data analysis techniques (filtering is the most common).

For data gaps, one solution may be to interpolate between the recorded data points to "fill the gaps", however this must be done carefully as we know that the data created is fake. In many cases the gaps are simply filled with NAN's or Not-A-Number values. This allows the data user to use their best judgement as to how to handle the data gap, and ensures that there is no assumption of data validity at these times.

The QA can be handled either in the raw data state or after the data has been converted into the desired engineering format. For instance the pressure signal can be cleaned up or converted to water elevation and then cleaned up.

Once the data is QA'ed and QC'ed, we must be able to present it in a format that is widely understandable to the audience. The audience may simply be the researchers who have collected the data, or perhaps the data will be shared outside the research group.

Read Chapter 2.4.7, 2.4.8 (Emory and Thompson)

Statistics

In our experiment design, we have selected to use pressure sensors to record the change in water levels. These water level changes will be processed to provide the wave characteristics needed to advance the research into surfer wave relationship.

In order to make this conversion we must use time series methods of statistical analysis. We must first recognize that the water surface is a stochastic, random variable that can be described by a statistical distribution.

Time Series

For our work, we will be dealing primarily with time series of data for a group of variables that we track throughout the field experiment.

7 Wave Statistics and Spectra:

Single frequency waves are called monochromatic referencing light waves.

Up until now we have only been looking at monochromatic waves. We have derived solutions based on single frequency waves.

In reality, the sea surface is a superposition of many waves each having its own frequency, direction and amplitude.

In order to represent the sea surface we must superimpose many waves.

And to understand a sea surface we must decompose the raw elevation signal into the constituent components.

How do we extract the wave properties from a raw signal?

- Wave-by-wave analysis
    - Counting individual waves in a time series signal
    - Analog time series/Plot on paper
- Spectral analysis
    - This decomposition is performed by applying Fourier Transforms to the signal.
    - Digital time series/Numerical calculation

For our wave record the sea surface, η(t) is given by:

$$\eta(t) = \sum_{n=0}^{\infty} a_n \cos(\sigma_n t - \epsilon_n)$$

It is useful to arrange the data such that the researcher can gain insight into the characteristic frequencies of a signal.

The ability to plot the representative energy, a^2, vs frequency allows us to determine which frequencies are carrying the most energy in our signal.

Such a plot a_n^2 vs σ_n is called the energy spectrum. In reality we have waves of different σ and θ. When interested in direction we will calculate the directional wave spectrum.

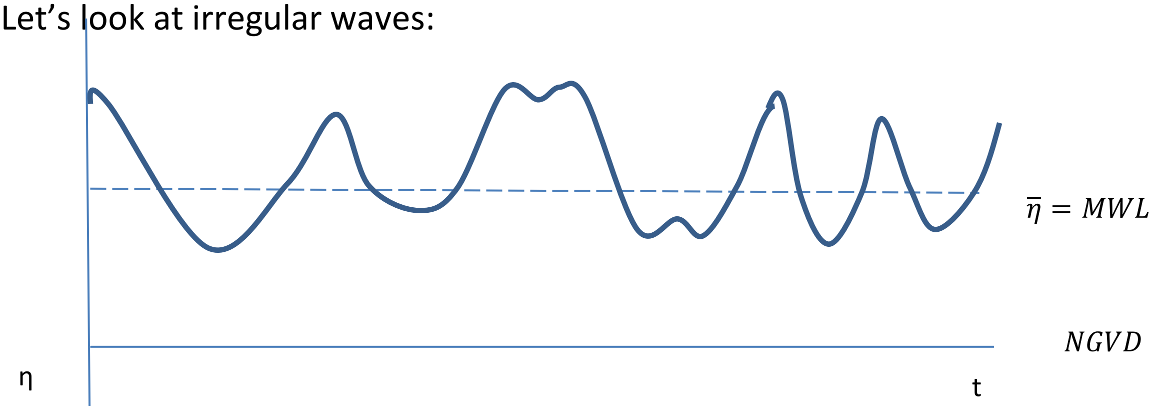
7.1 Wave-by-wave analysis

Divide up signal into individual 'waves'

Zero up crossing and zero down crossing

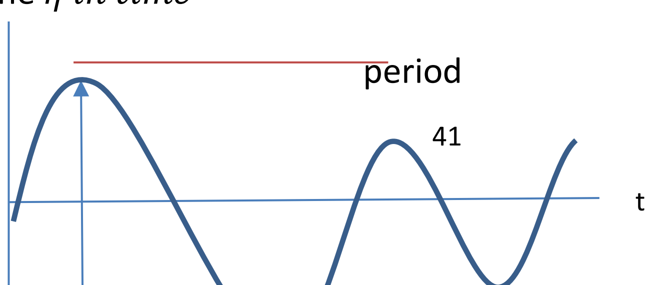
Need heights and periods for each wave

Let's look at irregular waves:



Superposition of many waves with different heights, frequency, phases, and directions

If we examine $\bar{\eta}$ *in time*



$\bar{\eta}$

height

We define a wave by height and period. From the signal we can pick out 'waves' by marking successive zero-upcrossings.

So let's go back to η above and we see that at this location, for this time duration we count 5 waves. Each has its own height, period. We can then collect this discrete data (height and period) and plot as bins (histograms).

In practice we record data from an analog plot then rank order (sort) the waves from largest to smallest, and finally bin the data into histograms.

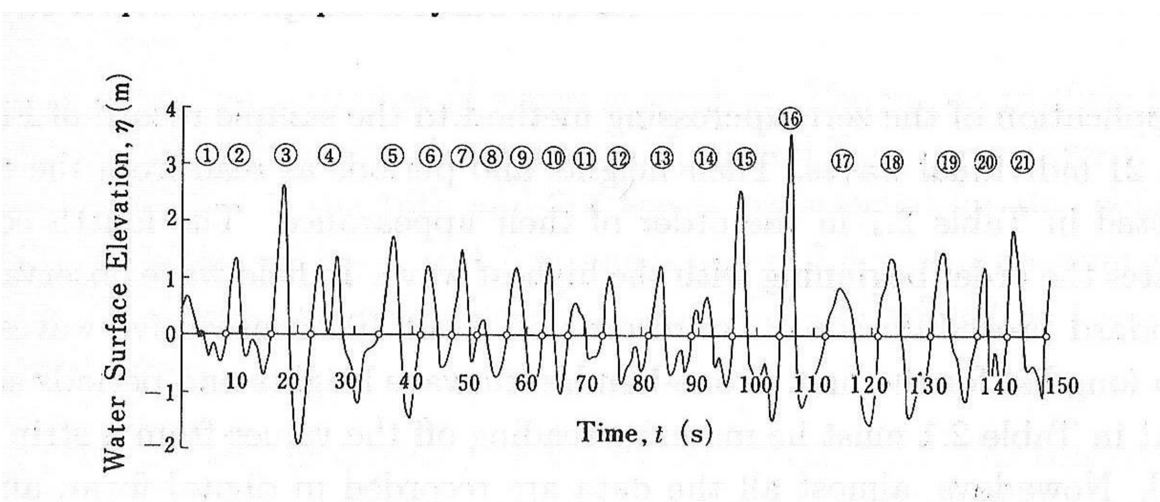
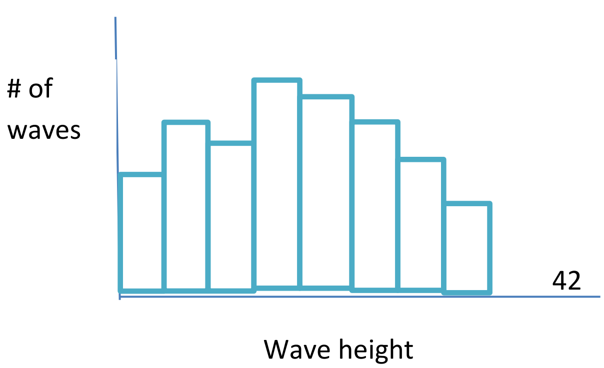


Fig. 2.2. Example of a wave record.



The histogram can be binned by height

Statistical measures can now be calculated.

Mean $\overline{H} = \frac{1}{n}\sum_i H_i$ (average height)

Root-mean squared wave height $H_{rms} = \sqrt{\frac{1}{n}\sum_i H_i^2}$ Measure of the magnitude of the varying wave height

Max

H_{rms} is always larger than $\overline{H}$

Min

Mean → average value

Mode → most common value

Median → half greater, half less

$Rms \equiv H_{rms} = \sqrt{\frac{1}{n}\sum_i H_i^2}$

Gives wave height with **<u>equivalent energy</u>** to the average energy in signal

Special statistics we like:

$H_{\frac{1}{n}} =$ the mean of the heighest 1/n waves

$H_{\frac{1}{3}} = H_s =$ significant wave height = the mean height of the highest 1/3 waves

$H_{\frac{1}{10}} =$ the mean height of the highest 1/10 waves

$H_1 = \overline{H} =$ the mean height of all waves

If N ≡ total number of waves,

The wave height distribution is defined by the probability.

Probability that a wave height is greater or equal to some value is given by:

$$P(H > \hat{H}) = \frac{n_{H>\hat{H}}}{N} \qquad \text{or } P(H \leq \hat{H}) = 1 - \frac{n_{H>\hat{H}}}{N}$$

$n_{H>\hat{H}}$ = number of waves for which H>$\hat{H}$

7.2 Discrete Wave Statistics:

$\bar{H}, H_{rms}, H_{\frac{1}{3}}$ $\rightarrow btwH_{\frac{1}{3}}$ ≡significant wave height

Kinsman (1965) determined a strong correlation between significant wave height and the visual wave height determined by the experienced observer.

Period can be defined as time between consecutive crests or zero crossing period (up or down). Now that the signal has been analyzed we can calculate probabilities. [Longuette-Higgins 1952 determined that the random distribution of waves followed known probability laws].

For a single, monochromatic wave signal

$\eta(t) = \left(\frac{H}{2}\right) cos\ \sigma t$; all waves have the same height

∴ $H_p=H_o$ for all P and $H_{rms}=H_o$

If we now add another monochromatic wave of the slightly different phase but the same wave height (as we did for constructing the wave group), we obtain a new signal of waves modulated by an envelope.

To determine H_p. since wave height decreases monotonically from max to min,

We average the wave height average from t=0 to P(π/Δσ)

$H_p = 4\ \frac{H_o}{p\pi} \sin\frac{p\pi}{2}$ and $H_{rms} = \sqrt{2}H_o$

So $H_p = \frac{2\sqrt{2}H_{rms}}{p\pi} \sin\frac{p\pi}{2}$ as p→ 0, $\sin(p) \rightarrow p$

$$\therefore \sin(P\pi/2) \rightarrow P\pi/2$$

As P→0, $H_p \rightarrow H_{max}$

$$H_{max} = \sqrt{2}H_{rms}$$

And recall $H_{rms} = \sqrt{2}H_0 \quad \therefore H_{max} = 2H_0$ as expected!

7.2.1 EXAMPLE:

What are $H_{max}, H_{\frac{1}{10}}, and\ H_1\ in\ terms\ of\ H_{rms}$

$$H_{max} = \sqrt{2}H_{rms} = 1.414H_{rms}$$

$$H_{\frac{1}{10}} = \frac{20\sqrt{2}H_{rms}}{\pi} sin\frac{\pi}{20} = 1.408H_{rms}$$

$$H_{\frac{1}{3}} = \frac{6\sqrt{2}}{\pi}H_{rms}\, sin\frac{\pi}{6} = 1.350H_{rms}$$

$$H_1 = \frac{2\sqrt{2}}{\pi}H_{rms}\, sin\frac{\pi}{2} = .9H_{rms}$$

Example 2: If you are an experienced observer and observe average wave height at the beach of 1.219m≈4 ft, what is the H_{max} and H_1?

Well it has been found that observed heights are correlated with $H_{\frac{1}{3}}$

$$\therefore 1.22m = H_{\frac{1}{3}} = 1.350H_{rms}$$

$$H_{rms} = 0.904m = 2.97ft$$

$$H_1 = .813m, 2.67ft$$

$$H_{max} = 1.28m, 4.2ft$$

7.3 Probability

Probability distribution (Probability mass function) P(x)= This function gives the frequency of occurrence of a possible outcome or value of a random variable.

It is important to note that:

$$0 \leq P(x)\ and \sum P(x) = 1,$$

- Probability is always greater than or equal to zero (non-negative)
- Sum of all probabilities is one

Another way to express probabilities is by the cumulative Probability C(x), this is the fraction of total population of events that a particular event is not exceeded.

$$C(x) = P\big(H \leq \widehat{H}\big) = 1 - \frac{n}{N}$$

Probability Density, p(x): this is the fraction of events in a population that is made up of a particular event. It represents the rate of change of a particular distribution. Or p(x)= $\frac{dC(x)}{dx}$. The two most common probabilities are Gaussian and Rayleigh Densities.

σ is stand dev of x

The Gaussian Probability Density is

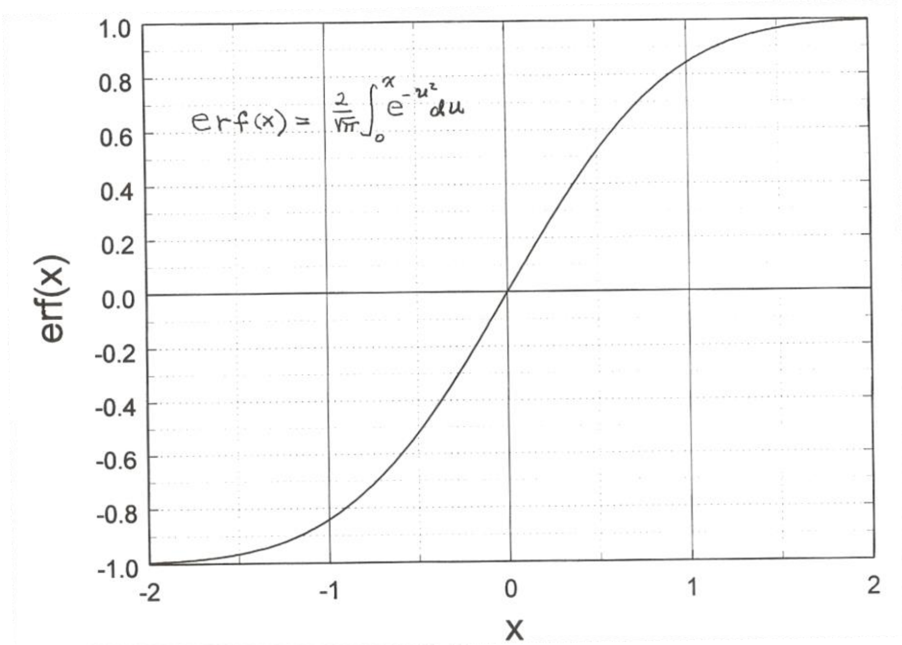
p(x)=$\frac{1}{\sigma_x\sqrt{2\pi}}\, e^{\frac{-(x-\mu_x)^2}{2\sigma_x^2}}$

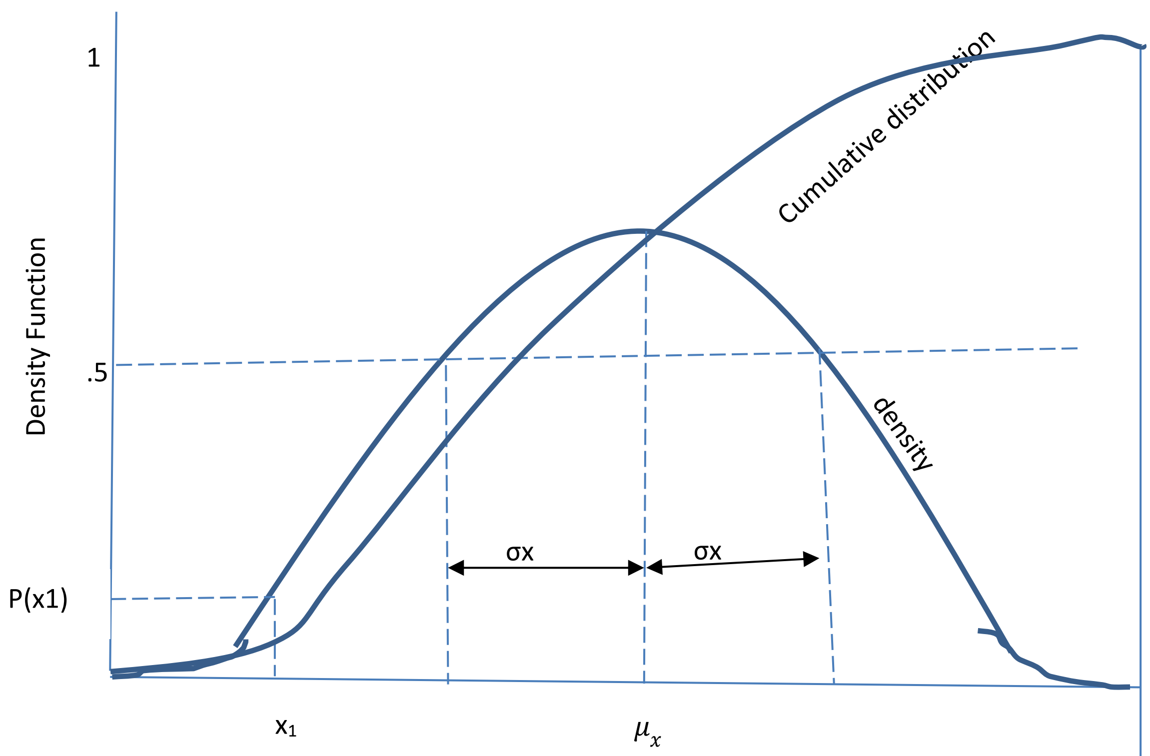
The integral of P(x) is the probability distribution *P*(x) for a zero mean, μ=0, and unit σ=1,

p(x)=$\frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$

P(x)=Φ$\left[\frac{x-\mu_x}{\sigma_x}\right] = \Phi = \int_0^x p(x)dx$

Where the integral $\Phi = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is known as the error function, erf(x)





The probability of exceedance Q(x) can be found as $Q(x(t)>x_1)=1-P\,(x(t)<x_1)=$

$1\text{-}\Phi\left[\frac{x-\mu_x}{\sigma_x}\right]$ which is the probability that x will exceed x_1 over time t.

Gaussian distribution is most useful for short term probabilities of free surface η(x,t).

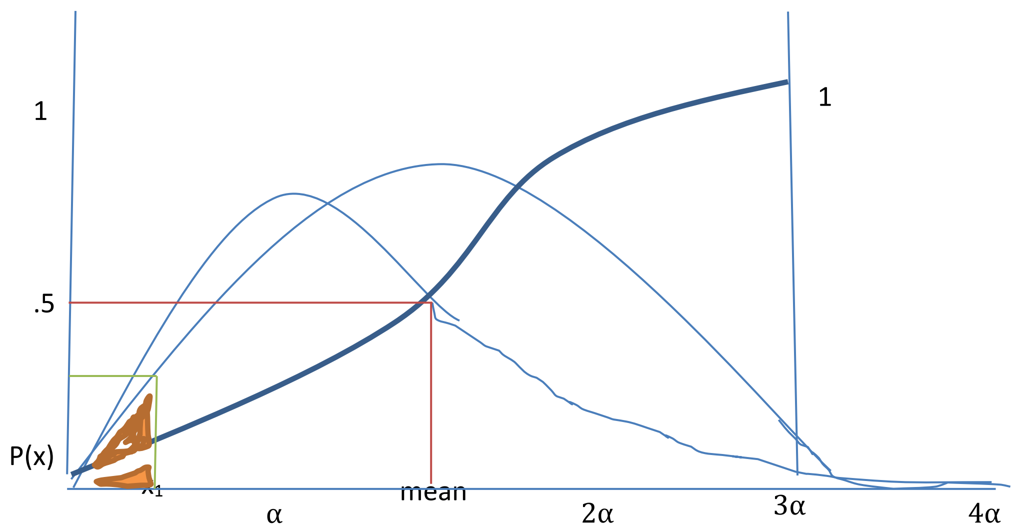
We are more concerned with wave height H rather than η. Longuet-Higgins (1952) examined statistically η(t) and found that both H and a followed the Rayleigh distribution.

Narrow Banded Spectra: **The Rayleigh Distribution**

$$p(\mathrm{x})=\frac{\pi x}{2\mu_x^2}\,e^{-\frac{\pi}{4}\left(\frac{x}{\mu_x}\right)^2}\quad \text{for } x\geq 0$$

$$P(\mathrm{x})=1-e^{-\frac{\pi}{4}\left(\frac{x}{\mu_x}\right)^2}\quad \text{for } x\geq 0$$

Rayleigh Dist. is asymmetric about mean



If we assume most wave energy falls under a small band of wave periods (narrow banded) the wave height probability density can be determined by

$$p(H) = \frac{2H}{H_{rms}^2} e^{\left(\frac{-H}{H_{rms}}\right)^2} \text{ or } P(H) = 1 - e^{-H^2/Hrms^2} \text{ or P(H>}\widehat{H}\text{)}$$

$$= e^{-\left(\frac{H}{H_{rms}}\right)^2}$$

Previously, P(H>$\widehat{H}$)= n/N for rank ordered grouping

$$\therefore \frac{n}{N} = e^{-\left(\frac{\widehat{H}}{Hrms}\right)^2}$$

N is the number of waves onto N total waves that have a height greater than or equal to $\widehat{H}$

We can rearrange this eq to find the height exceeded by the n largest waves in the group

$\widehat{H} = H_{rms}\sqrt{\ln\frac{n}{N}}$ $\quad$ $H_p = H_{rms}\sqrt{\ln\frac{1}{p}} \equiv$ height exceeding pN of the waves

Example: 400 waves in record

N=400

$H_{rms}=\sqrt{\frac{1}{N}\sum H_i^2}$

a) How many waves are expected to exceed H=2H_{rms}

$$n = Ne^{-\left(\frac{\widehat{H}}{H_{rms}}\right)^2} = 400e^{\left(-\frac{2H_{rms}}{H_{rms}}\right)^2} = 400e^{(-2)^2} = 400e^{-4} = 7.3, 7\ waves$$

$$\frac{n}{N} = \frac{7}{400} = 0.0175\ x\ 100\%$$

$$= 1.75\%\ of\ the\ total\ \#\ of\ waves\ exceeded\ 2H_{rms}$$

b) Height $\widehat{H}$ exceeded by ½ waves (n=N/2 or P=1/2)

$$H_{1/2} = H_{rms}\sqrt{\ln 2} = 0.833H_{rms}$$

c) Height exceeded by one wave? For $H_{1/N}$ we have P=1/N=1/400

$$H_{1/400} = H_{rms}\sqrt{\ln 400} = 2.45\ H_{rms}$$

As N ↑, Hmax ↑

Since PDF never actually reaches zero, rather decays asymptotically.

For a Rayleigh Distributed wave field (Table 7.1 D&D):

$$H_{\frac{1}{10}} = 1.80H_{rms}$$

$$H_{\frac{1}{3}} = 1.416\ H_{rms}$$

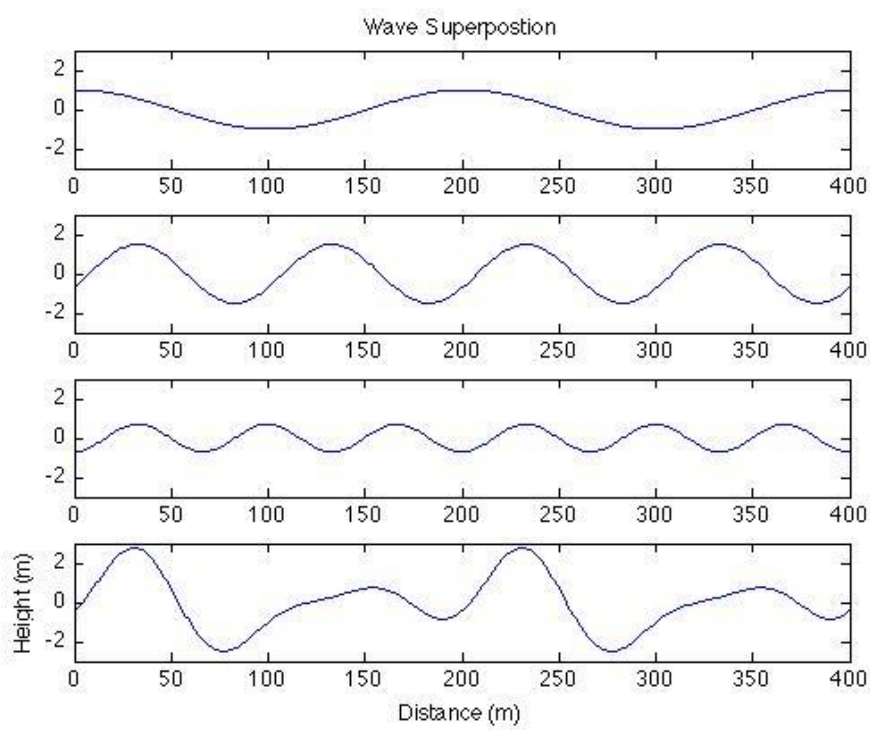
$$H_1 = 0.886H_{rms}$$

Or in general

$$\frac{H_p}{H_{rms}} = \sqrt{\ln\frac{1}{p}} + \frac{\sqrt{\pi}}{2p} erfc(\sqrt{\ln\frac{1}{p}}$$

$$WHere\ erfc\ is\ the\ complentary\ error\ function$$

Time series and Statistics (DD 7.1-7.2)



Histograms and Distributions (DD7.1-7.2)

7.4 Spectral Analysis

Any piecewise continuous function can be expressed by a sum of sines and cosines over some interval $t_o \rightarrow t_o + \tau$

Wave record sampled at regular discrete interval ($\Delta t \sim 0.5s$-$2s$) and finite in length: 20 min – 2 hr

Spectral method will bin record into frequencies 'automatically'. No need to sort the wave periods first.

Any periodic signal can be represented by an infinite sum of Fourier components:

$$\eta(t) = \sum_{n=0}^{\infty} a_n \cos n\sigma t + \ b_n \sin n\sigma t$$

In reality we have a limited series not an infinite, N not ∞, so for a finite record

$$\eta(t) = a_o + \sum_{n=1}^{N} a_n \cos n\sigma t + \ b_n \sin n\sigma t$$

where $a_n = \frac{2}{T}\int_t^{t+\tau} \eta(t) \cos n\sigma t \, dt$ even component

$b_n = \frac{2}{T}\int_t^{t+\tau} \eta(t) \sin n\sigma t \, dt$ odd component

and $a_o = \frac{1}{T}\int_t^{t+\tau} \eta(t) \, dt$ = average for mean value in record

a_o (Mean) usually subtracted out

Or in Complex notation:

$$\eta(t) = \sum_{n=-\infty}^{\infty} C_n e^{int}$$

$$C_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} \eta(t) e^{-int} dt$$

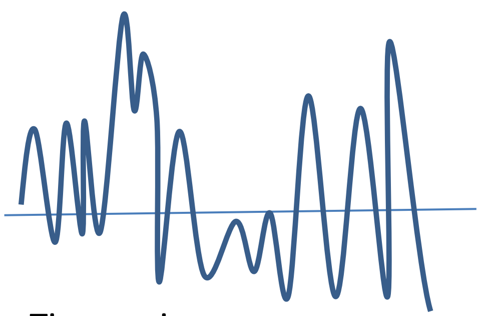
$$C_n = \frac{1}{2}(a_n - ib_n)$$

We typically use FFT (Cooley and Tukey,1965) which exploits symmetries to streamline calculations, to obtain a_n and b_n, 'fast former transform'

We will not get into theory behind Fourier sums and transforms for this class.

The FFT is a routine available in most canned mathematical programs like MatLab and Mathematica also in 'numerical recipes' (Press et al 1986)

FFT is most efficient when N is some power of 2
N= 512, 1024, 2084 or 4096 points
We treat FFT as a 'black box' so for us



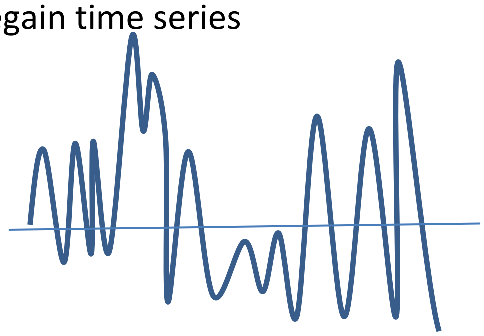
→ FFT→ N, a_n,b_n

Time series

Where N is the number of frequencies returned, $a_n and\ b_n$, are the coefficients for each of the N frequencies.

And the inverse to regain time series

N, a_n b_n → FFT^{-1}→



Time series

Let's describe a time series at Δt intervals

So if we have a total record length of τ seconds and it is split or digitized into N evenly spaced, Δt, points:

$$N\Delta t = \tau$$

Or the total number of samples is,

$$N = \frac{\tau}{\Delta t}$$

Our continuous time record becomes a discreet number of points. In a finite discrete form (only N points):

$$\eta(t) \rightarrow \eta_j, j = 1, 2, 3, \ldots, N$$

$$\sigma = \frac{2\pi}{\tau}$$

where τ is the total time interval

$$2\pi$$

We can write

$$\eta(t) = \sum_{n=-N}^{N} F_n e^{in\sigma t}$$

$$F_n = \left(\frac{a_n - ib_n}{2}\right) \ and \ \sigma = \frac{2\pi}{\tau} = \frac{2\pi}{N\Delta t}$$

$$\eta(t) = \eta_j = \sum_{n=0}^{N-1} F_n e^{i\left(2\pi\left(\frac{nt_j}{N\Delta t}\right)\right)}$$

So if we know ALL F_n from n=0 to N-1 we can find the time record at time j. We can solve for the n^th^ coefficient F_n by:

$$F_n = \frac{1}{\tau}\int_t^{t+\tau} \eta(t)\, e^{-in\sigma t} dt = \frac{1}{\tau}\int_0^{\tau} \eta(t) e^{-2\pi i\left(\frac{nt}{N\Delta t}\right)} dt = \frac{a_n}{2} - i\frac{b_n}{2}$$

$$or\ for\ F_{-n} = F_n^* = \frac{a_n}{2} + i\frac{b_n}{2}$$

These are the finite discrete forms of the former transform pairs

We can use the equation above for F_n, to calculate a_n & b_n, however typically we will use FFT to do this.

Inputting a time series (N values) we get back a_n , b_n , one for each N frequency.

Our frequency resolution for the lowest frequency that can be resolved (fundamental frequency or Bandwith), Δf or f_0, is determined by the Δt and number of records.

$$\Delta f = f_0 = \frac{1}{\tau} = \frac{1}{N\Delta t}$$

For a fixed N and a finer sampling rate will lead to a larger Δf, or higher frequency resolution.

For a ‘good’ representation of infinite series using a finite record a_n & b_n, we need a large enough record length, N.

But how big is big enough?

How do we know if we have a long enough wavelength record?

If we take the average mean square value of our recorded time series function :

$$\frac{1}{\tau}\int_0^{t+\tau} \eta^2(t)dt$$

(average mean square value)

½ the sum of the squares of the Fourier coefficient should equal approx.:

$$a_o^2 + \frac{1}{2}\sum_n (a_n^2 + b_n^2) \approx \frac{1}{\tau}\int_t^{t+\tau}\left[a_o + \sum_n (a_n\cos(n\sigma t) + b_n\sin(n\sigma t)\right]^2 dt$$
$$= \frac{1}{\tau}\int_0^{t+\tau} \eta^2(t)dt$$

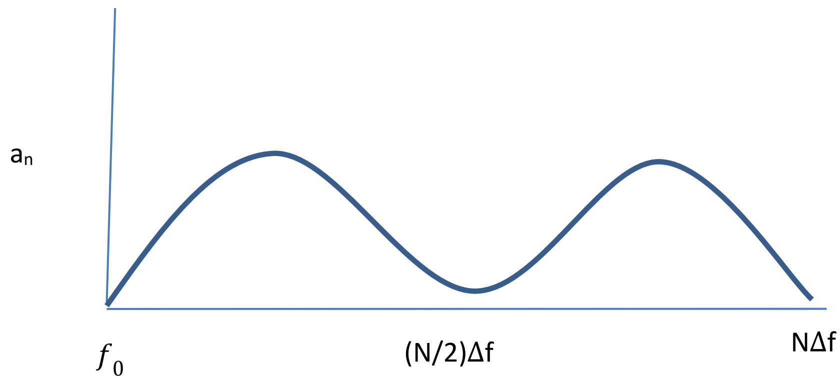
Parseval’s Theorm

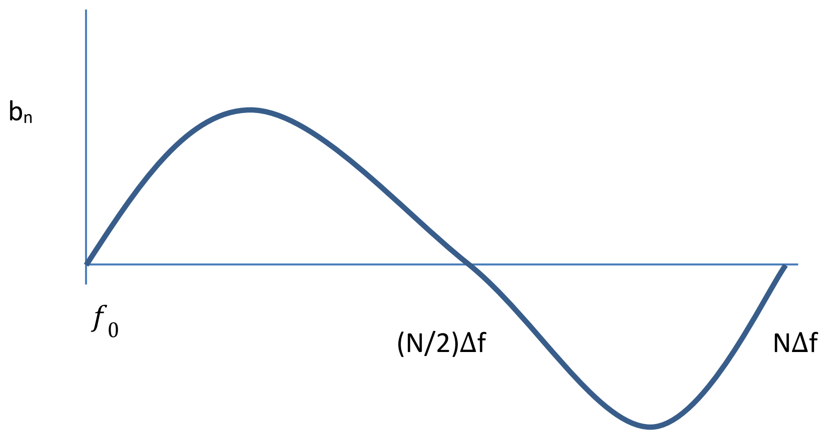
(Often de-mean the signal to get rid of a_o)

If not, then we need to add more numbers to the series, make N bigger.

What do a_n & b_n look like?

Most output is ordered from longest frequency to NΔf



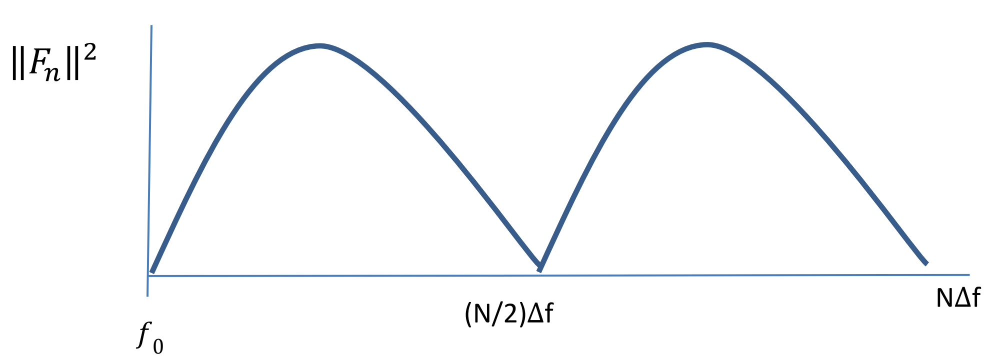


a's are exactly symmetric about (N/2)Δf, while the b_n are skew-symmetric about (N/2)Δf.

The results are typically presented by plots

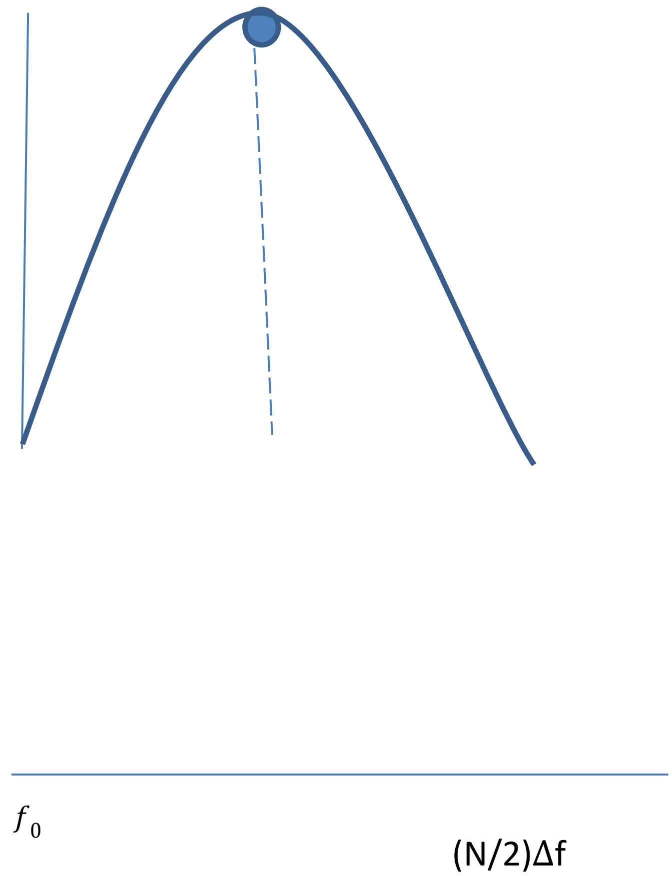
$$\|F_n\|^2 = \left(\frac{{a_n}^2}{4} + \frac{{b_n}^2}{4}\right)$$

Results in a two sided spectrum



Which is then typically ‘folded’ about (N/2)Δf.

$$2\|F_n\|^2 = \frac{a_n^2 + b_n^2}{2}$$



Results in a spectrum - a distribution of energy with frequency.

Most energy is contained in the spectral peak.

We can calculate the spectral density (or energy density).

$$S(f) = \frac{2\|F_n\|^2}{\Delta f}$$

For a free surface time series, in meters S(f) has units of m^2s or m^2/Hz

The variance $\sigma^2 = \overline{[\eta(t)]^2}$ (mean square of the surface elevation)

Let's connect the energy density spectrum to the energy of a wave.

Assume: $\eta(t) = a \sin\sigma t$

Calculating variance over 2π

$$\sigma^2 = \overline{[\eta(t)]^2} = \frac{1}{2\pi}\int_0^{2\pi} a^2 sin^2\sigma t \ d(\sigma t) = \frac{a^2}{2}$$

$\frac{a^2}{2} = 2\int_0^{\infty} S(f)df$ for a monochromatic wave.

The variance, wave energy and wave energy spectrum are related.

Given an energy spectrum S(f), the wave amplitude for a given frequency is

$$a(f) = \sqrt{2S(f)\Delta f}$$

For monochromatic wave show fig 6.11 pg 175

For a random wave show fig 6.12 pg 176

Directional spectra, 6.13, 6.14

The limiting frequency at which a wave can be adequately resolved is called the Nyquist frequency, or folding frequency

$$f_{Nyq} = \frac{1}{2\Delta t}$$

The shortest measurable limit is $T = 2\Delta t,\ \sigma = \pi/\Delta t$

For any higher frequency waves, (T smaller, f (or σ) bigger), the sampling rate is inadequate. Researcher must design Δt to avoid aliasing.

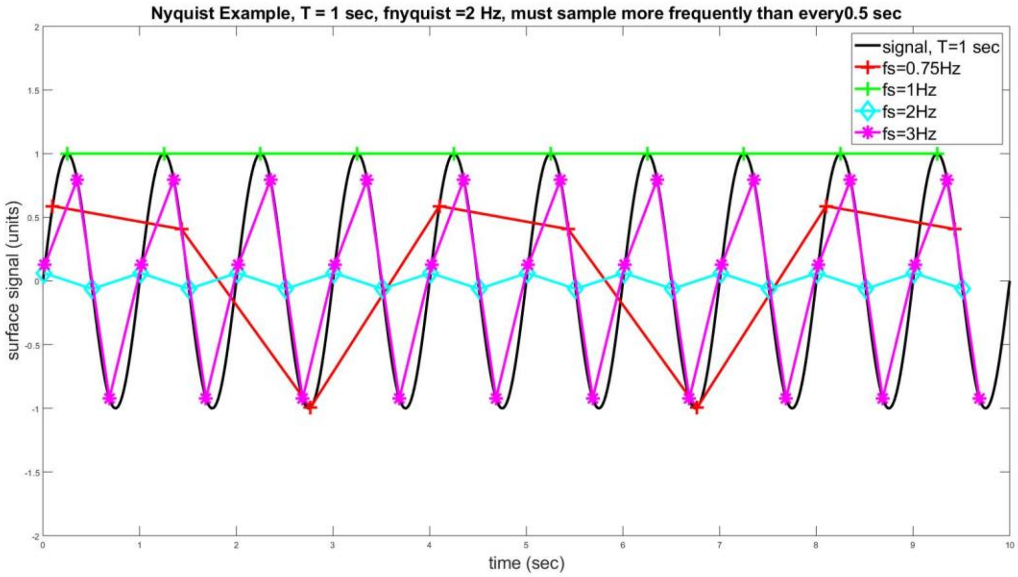
Sampling frequency must be greater than twice the highest frequency in the signal.

Nyquist frequency = smallest wave or maximum frequency can be represented given a sampling rate

Nyquist Rate = sampling rate needed to prevent aliasing

The sampling rate should be more than double the maximum frequency that we are trying to capture.

All frequencies less than (larger waves) the maximum frequency we are trying to measure will not be aliased. However the max. frequency could still be aliased:



.

Correlation

Read 3.13

Computing the correlation coefficient, r, allows researchers to determine how interrelated two signals may be by tracking how they co-vary in time or space.

r is always between +1 and -1. An r of +1 indicates that the signals are perfectly correlated and in phase with each other, while an r of -1 indicated that the signals are perfectly out of phase with each other. As r approaches zero, we find a plot to be a scattering of points with no relationship.

Computing the correlation coefficients enhances our ability to interpret the existence of influence that one variable may have on another, it does not indicate causation.

$$
\begin{aligned}
r &= \frac{1}{N-1}\sum_{i=1}^{N}\frac{(x_i-\overline{x})(y_i-\overline{y})}{s_x s_y} \\
&= \frac{C_{xy}}{s_x s_y}
\end{aligned}
$$

Where $C_{xy}$is the *covariance*, and *s* is the standard deviation.

Fourier transform

Ocean wave spectrum

Signal filtering and windowing