TIME SERIES ANALYSIS LECTURE 1

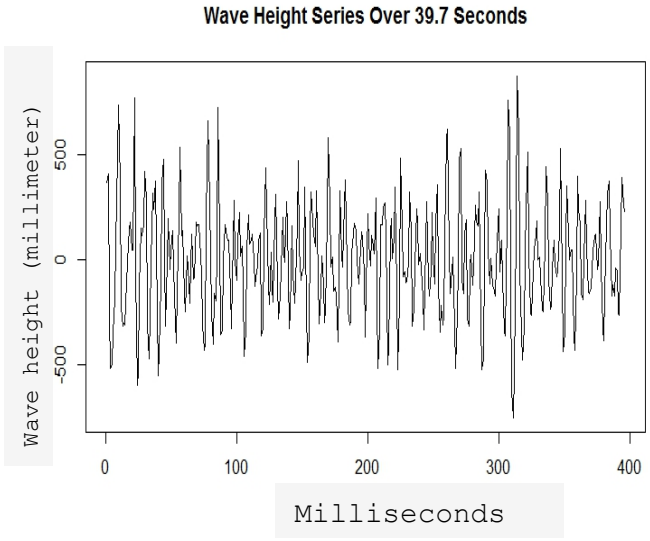
datascience@berkeley

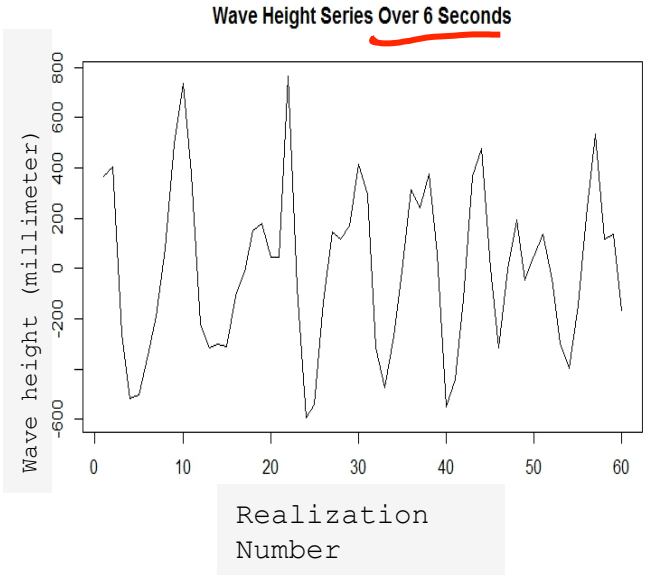
Examining Time Series Correlation—Autocorrelation Function: Example 1

Example: Wave Height

Let's go through a few examples to illustrate the concepts just covered.

- In this example, we use the wave height data provided in the textbook, *Introductory Time Series With R*.
- It is a time series of wave height, measured in millimeters (mm) relative to still water.
- When considering a time series, we need to pay attention to the sampling interval. In this example, the sampling interval is 0.1 second and the total record length is 39.7 seconds, giving almost 400 observations.
- As mentioned, we should always make a time series plot when analyzing a time series.
- The graph at the top displays the entire (sampled) series over the 39.7 seconds of recording.





Example: Wave Height (2)

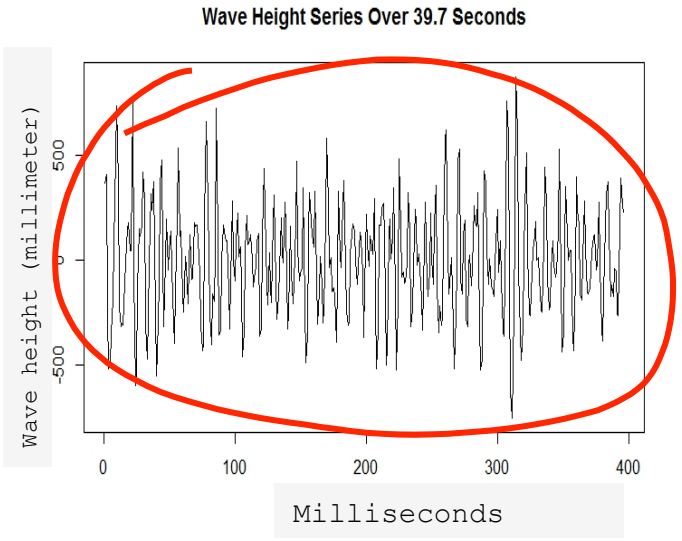
- When graphing a time series (or any other plots for that matter), ensure it includes a title and the axes are well-labeled. In a time series plot (or t-plot) the x-axis is time period (or frequency). Make sure the correct time period or interval is recorded to assist readers in understanding the graphs.

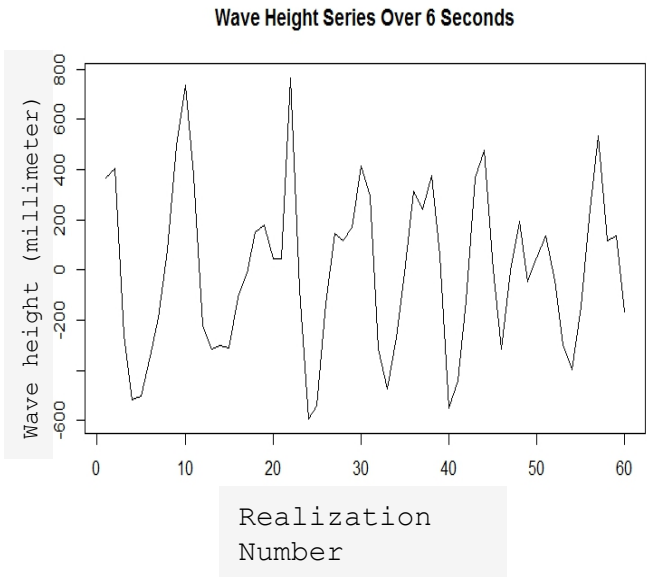
- In this example, the time period is measured in milliseconds and the unit of measurement (of wave height) is millimeters (mm).

- The wave height series does not appear to have any trends or "seasonal" components, so it is reasonable to assume that the series is a realization of a stationary process.

- In other words, a stationary process can be used to model this set of realizations.

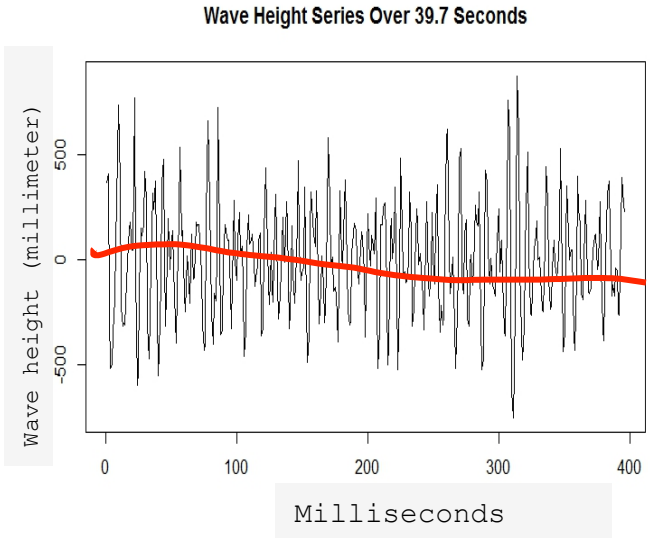
- The series also does not show any outliers.

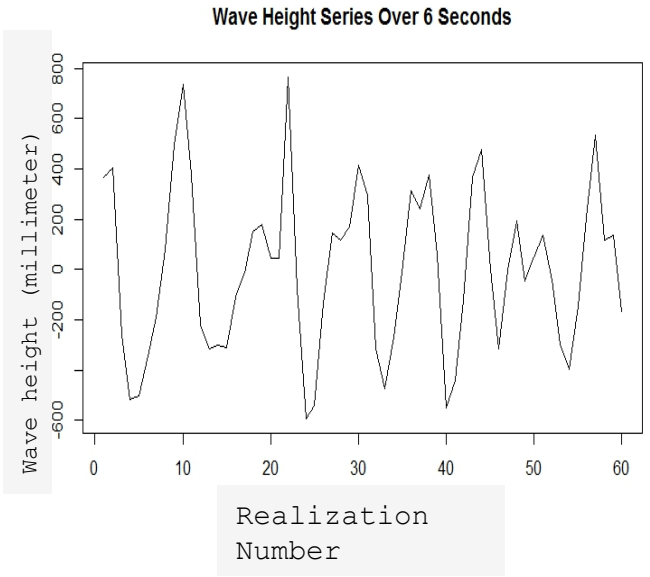




Example 1: Wave Height (3)

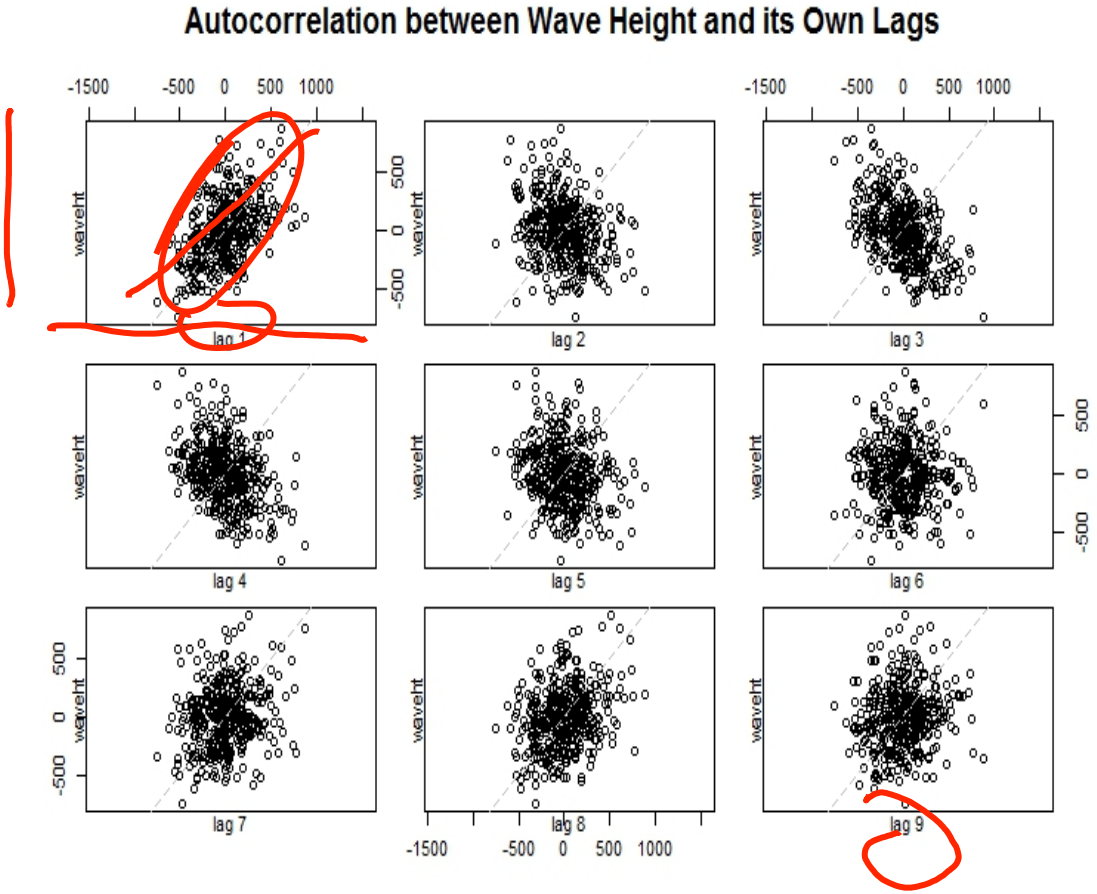
- The graph at the bottom displays a subset of the series, showing only the first 60 recorded wave heights.
- Note that the series appears to fluctuate around a constant mean.
- The series also appears to be generated by an underlying process that has a constant variance.
- Confirming from the graph at the top, the series appears to be a realization from a process with both mean and variance stationary.
- The skip-consecutive values appear to be relatively similar, mimicking that of a rough sea.
- It has quasi-periodicity but no fixed frequency.





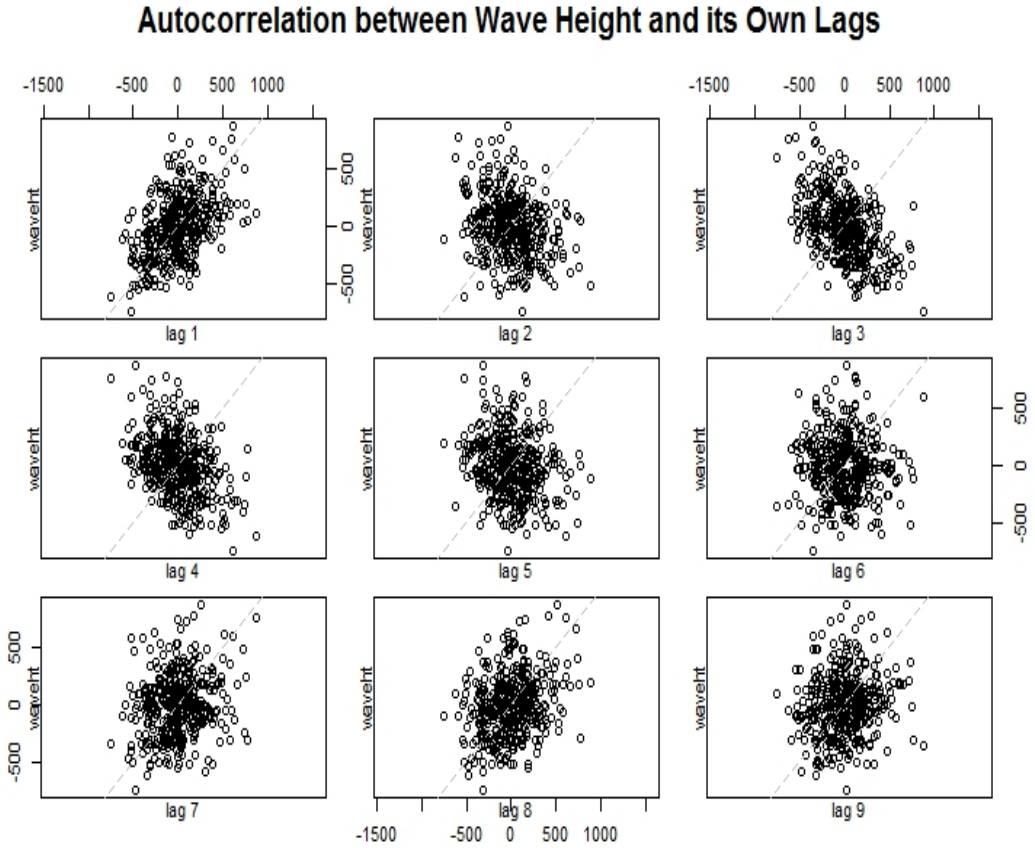
Example 1: Wave Height—Correlation With Lags

- A very useful method to visually inspect the dependency structure of a series is to plot the series against various values of its own lags using a scatter plot matrix.
- Each plot represents the wave height against a particular lag of itself.



Example 1: Wave Height—Correlation With Lags (2)

- Imagine this is a matrix. The scatter plot on the top left (or the (1,1) position of the matrix—first row and first column) can be used to examine the correlation between wave height and its first lag.
- The scatter plot in row 3, column 2 can be used to examine the correlation between wave height and its lag 8 values.

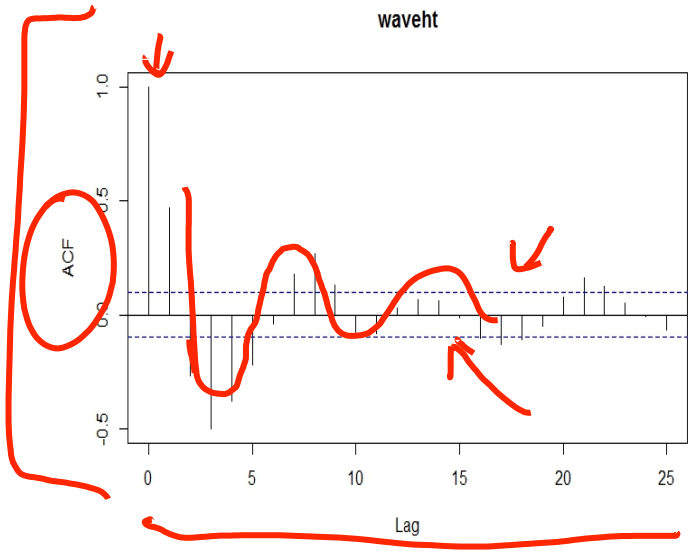


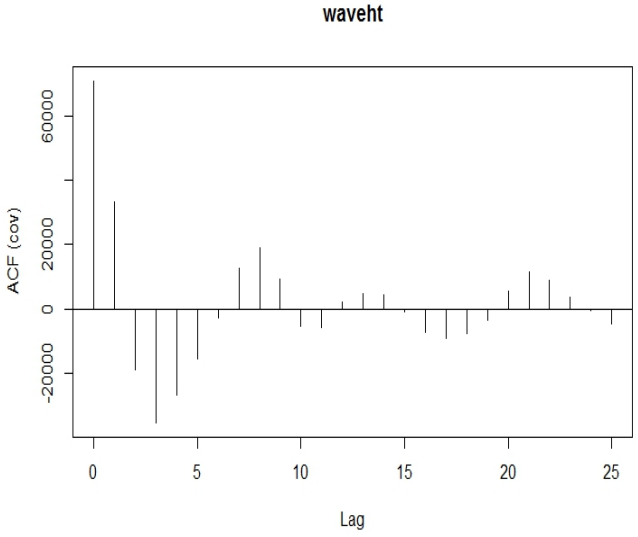
Example 1: Wave Height—Autocorrelation Function (3)

Recall that autocorrelation function takes the form:

$$\frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\frac{1}{T}\sum_{t=1}^{T-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

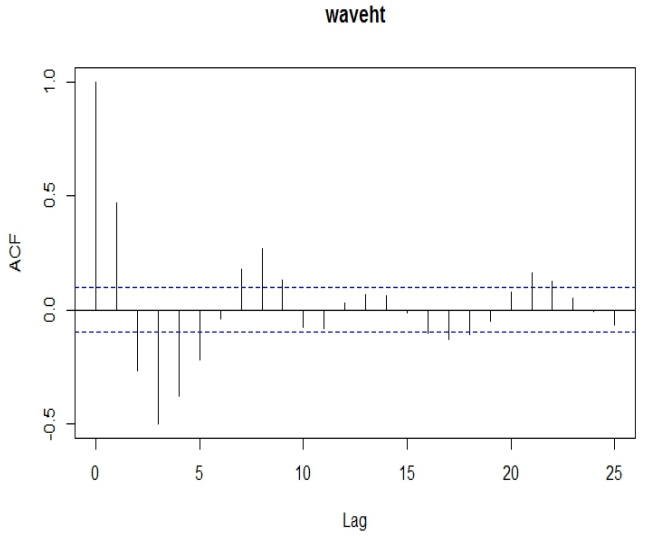
- The plots in the scatter plot matrix show both positive and negative correlations, which is confirmed by the autocorrelation (ACF) function (top graph) and autocovariance function graphs (bottom graph) on the right.
- The blue dotted lines represent the 95% confidence interval (CI) of the ACF.
- Notice that the CI of each of the autocorrelations are the same. It comes from the property of the AR model that both the conditional and unconditional variances are constant.
- The CI takes the form of $-\frac{1}{n} \pm \frac{2}{\sqrt{n}}$, which shrinks as the sample size increases.

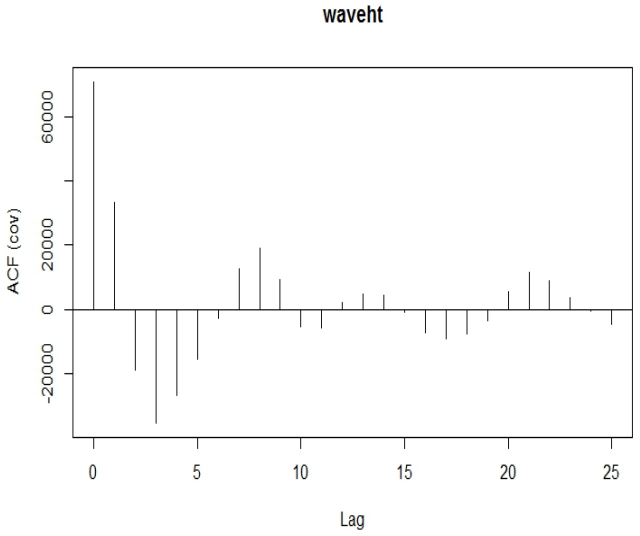




Example 1: Wave Height—Autocorrelation Function (4)

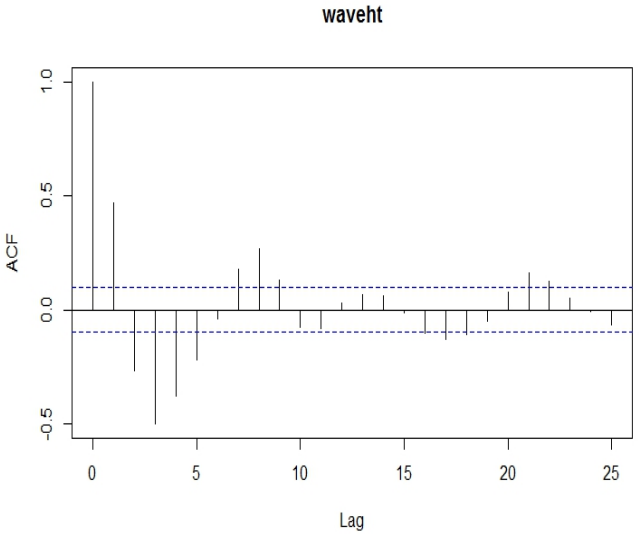
- Although in principle, when the estimated autocorrelation (at any lag k) falling outside the blue dotted line is evidence against the null hypothesis that the correlation at lag k is 0 at the 5% level, we should be very careful about interpreting multiple hypothesis tests. Even if all of the autocorrelations equal to 0 at all lags k, 5% of the estimated autocorrelations could still fall outside the blue lines, by chance. (Recall the hypothesis testing lectures in DATASCI W203.)

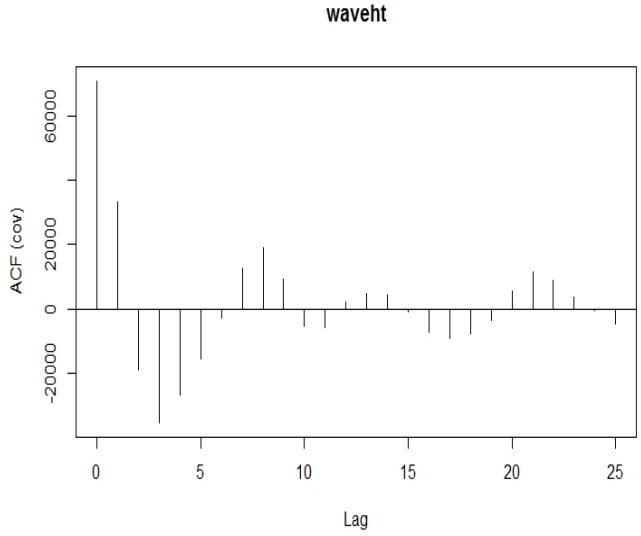




Example 1: Wave Height—Autocorrelation Function (5)

- The correlogram for wave heights has a wavelike shape that resembles that of a shrinking cosine function.
- This is typical of correlograms of time series generated by AR(2) process, as we will see in the next lecture.
- We will explore this aspect in details when we study
- Once we start studying stationary time series models in the next lecture, examining the correlogram (i.e., the graph of ACF function), along with a couple other measures, serves as an important for step to identify the order of an ARIMA model.





Berkeley SCHOOL OF INFORMATION