Astronomical Time Series Analysis

ABSTRACT

In this project, we aim to statistically analyze the various aspects of the time-series data generated from astronomical sources namely LINEAR and LIGO Dataset. The components we aim to analyse are:

- Forecasting
- Seasonality and Trend Detection
- Power Spectral Density Estimation
- Frequency Estimation

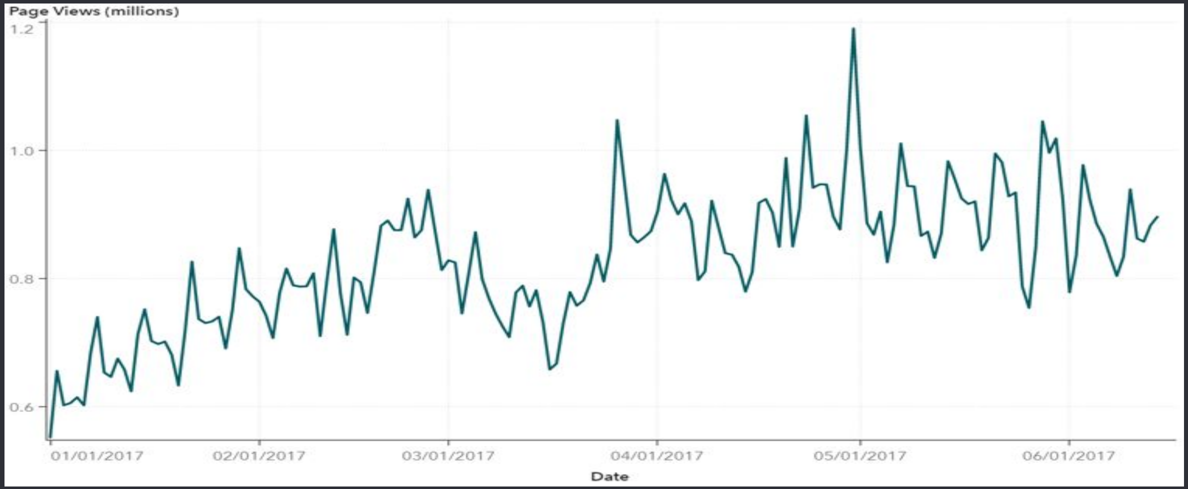
DATASETS USED

- **LIGO DATASET**
  - We put into use the time series released from LIGO (Laser Interferometer Gravitational Observatory) Hanford, sampled at 4096 Hz.
  - The data is calibrated such that the the wave signals have units of dimensionless strain ($\Delta L/L$).
- **LINEAR DATASET**
  - LINEAR (Lincoln Near-Earth Asteroid Research) is a project collaborated to detect and catalog the near-Earth asteroids or Near-Earth objects(NEO).
  - The dataset consists of time series of 7010 different objects, which are assigned some ID, so as to get spectra for a corresponding object.

KEY TERMS

TIME SERIES

- A Time Series consists of various data points plotted over a course of time. In general view, it consists of sequence of data, evenly spaced over in time.
- The time series are mostly used in statistics, signal processing, econometrics, weather forecasting, astronomy and many other fields and comprises of the application of statistical techniques to analyse the time series data and extract useful information from it.



POWER SPECTRAL DENSITY

- The power spectrum $S_{xx}(f)$ of a time series $x(t)$ describes the distribution of power into frequency components composing the signal.
- When the energy of the signal is concentrated around a finite time interval i.e. if its total energy is finite, we compute the Energy Spectral Density.
- **Power Spectral Density (PSD)** applies to signals existing over all time, or over a time period large enough that it could as well have been over an infinite time interval.

FAST FOURIER TRANSFORM

- A fast Fourier transform (FFT) is an algorithm that computes the Discrete Fourier transform(DFT) of a sequence.
- The DFT is obtained by decomposing a sequence of values into components of different frequencies which is often a slow process in the practical world.
- An FFT rapidly computes such transformations and as a result, it manages to reduce the complexity of computing the DFT.

WELCH’S METHOD

- This method proposed by Welch is often used as an estimator of the PSD.
- The method consists of dividing the time series data into (possibly overlapping) segments, computing a modified periodogram of each segment, and then averaging the PSD estimates resulting in Welch's PSD estimate.
- Although overlap between segments introduces redundant information, this effect is diminished by the use of a non-rectangular window, which reduces the importance or weight given to the end samples of segments

WINDOW FUNCTION

- A Window Function is a mathematical function that is zero-valued outside of some chosen interval, normally symmetric around the middle of the interval, usually near a maximum in the middle, and usually tapering away from the middle.
- When another function is multiplied by a window function, the product is zero-valued outside the interval: all that is left is the part where they overlap, the "view through the window".
- In actual practice, the segment of data within the window is first isolated, and then only that data is multiplied by the window function values.

TYPES OF WINDOWS

- **Hanning Window** - This window is used in order to smoothen the discontinuities at the beginning and end of the sampled signal.
- **Top Hat Window -** Top Hat Window is typically employed on data where frequency peaks are distinct and well separated from each other.

The Hanning Window is preferred when the frequency peaks are not guaranteed to be well separated as it is less likely to cause individual peaks to be lost in the spectrum whereas if the frequency peaks are distinct and well separated from each other Top Hat Window is employed.

TIME SERIES DECOMPOSITION

- ❖ A useful abstraction for selecting forecasting methods is to break a time series down into systematic and unsystematic components.
- ❖ A given time series is thought to consist of three systematic components including level, trend, seasonality, and one non-systematic component called noise.
- ❖ These components are defined as follows:-
  - ▹ Level - The average value in the series
  - ▹ Trend - The increasing or decreasing value in the series
  - ▹ Seasonality - The repeating short-term cycle in the series
  - ▹ Noise - The random variation in the series
- ❖ A series is thought to be an **Additive** model or a **Multiplicative** model of these components.

LOMB-SCARGLE PERIODOGRAM

- The Lomb-Scargle periodogram is a standard method to search for periodicity in unevenly sampled time series data. It is a well-known algorithm for detecting and characterizing periodicity in unevenly sampled time-series.
- The Lomb-Scargle periodogram is a method that allows efficient computation of a Fourier-like power spectrum estimator from such unevenly sampled data, resulting in an intuitive means of determining the period oscillation.
- The Lomb-Scargle periodogram corresponds to a single sinusoidal model, $y(t) = a \sin(wt) + b \cos(wt)$, where t is time and w is angular frequency (= 2f).
- The model is linear with respect to coefficients a and b, and non-linear only with respect to frequency w.

AUTOCORRELATION FUNCTION

- The sample autocorrelation function (ACF) for a series gives correlations between the series x_t and lagged values of the series for lags of 1, 2, 3, and so on.
- Let x_t denote the value of a time series at time *t*. The ACF of the series gives correlations between x_t and x_{t-h} for h = 1, 2, 3, etc. Theoretically, the autocorrelation between x_t and x_{t-h} equals:

$$\mathrm{ACF} = \frac{\mathrm{Covariance}(\mathrm{x_t}, \mathrm{x_{t-h}})}{\mathrm{Std.Dev.}(\mathrm{x_t}) * \mathrm{Std.Dev.}(\mathrm{x_{t-h}})}$$

PARTIAL AUTOCORRELATION FUNCTION

- The PACF is a conditional correlation which examines the correlation between two variables under the assumption that we know and taken into account the values of some other set of variables.
- For instance, consider a regression context in which y is the response variable and x_1, x_2 and x_3 are predictor variables. The partial correlation between y and x_3 is the correlation between the variables determined taking into account how both y and x_3 are related to x_1 and x_2.
- We can define the partial correlation just described as:

$$\mathrm{PACF} = \frac{\mathrm{Covariance}(y, x_3 | x_1, x_2)}{\sqrt{\mathrm{Variance}(y | x_1, x_2) * \mathrm{Variance}(x_3 | x_1, x_2)}}$$

AUTOREGRESSIVE MODEL

- An autoregressive (AR) model is a representation of a type of random process.
- As such, it is used to describe certain time-varying processes in nature, economics, etc.
- It specifies that the output variable depends linearly on its own previous values.
- The notation AR(p) indicates an autoregressive model of order p and is defined as:

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$$

MOVING AVERAGE MODEL

- In time series analysis, the Moving-Average model(MA model) is a common approach for modeling univariate time series.
- It species that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) term.
- Together with the AR model. It is a special case and key component of ARIMA model of time series.
- The notation MA(q) refers to the moving average model of order q:

$$\mathbf{X_t} = \mu + \epsilon_\mathbf{t} + \theta_\mathbf{1}\epsilon_\mathbf{t-1} + \cdots + \theta_\mathbf{q}\varepsilon_\mathbf{t-q}$$

ARIMA MODELLING

- An ARIMA model is a class of statistical models for analyzing and forecasting time series data that explicitly caters to a suite of standard structures in time series data, and provides a simple yet powerful method for making skillful time series forecasts.
- In this x(t), we express as a function of past value(s) of x and/or past errors (as well as a present time error). When we forecast a value past the end of the series, we might need values from the observed series on the right side of the equation or we might, in theory, need values that aren't yet observed.
- Any ARIMA model can be converted to an infinite order MA model:

$$x_t - \mu = w_t + \Psi_1 w_{t-1} + \Psi_2 w_{t-2} + \cdots + \Psi_k w_{t-k} + \dots \tag{1}$$

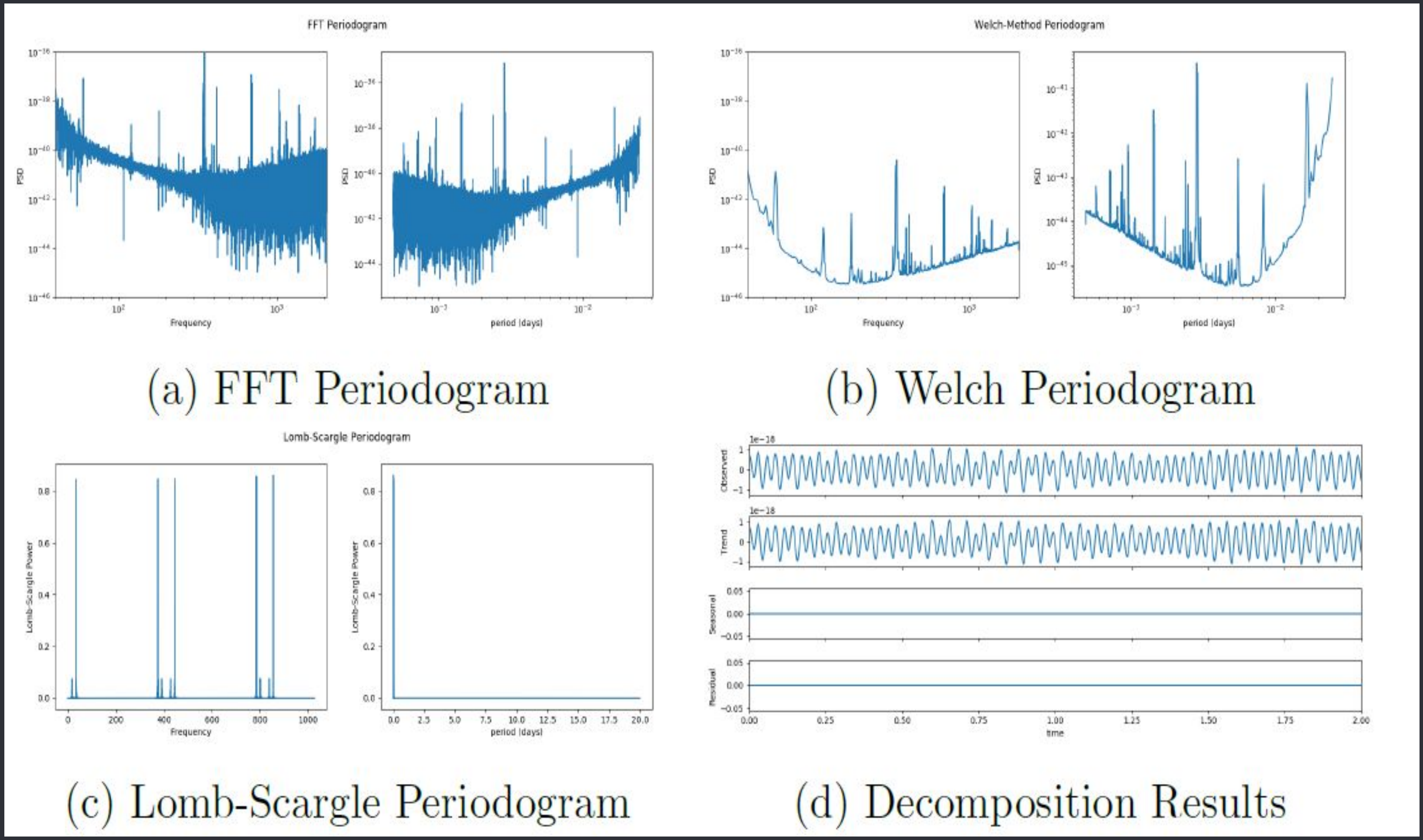
$$= \sum_{j=0}^{\infty} \Psi_j w_{t-j} \text{ where} \Psi_0 \equiv 1 \tag{2}$$

PROCESS IMPLEMENTED

LIGO DATASET

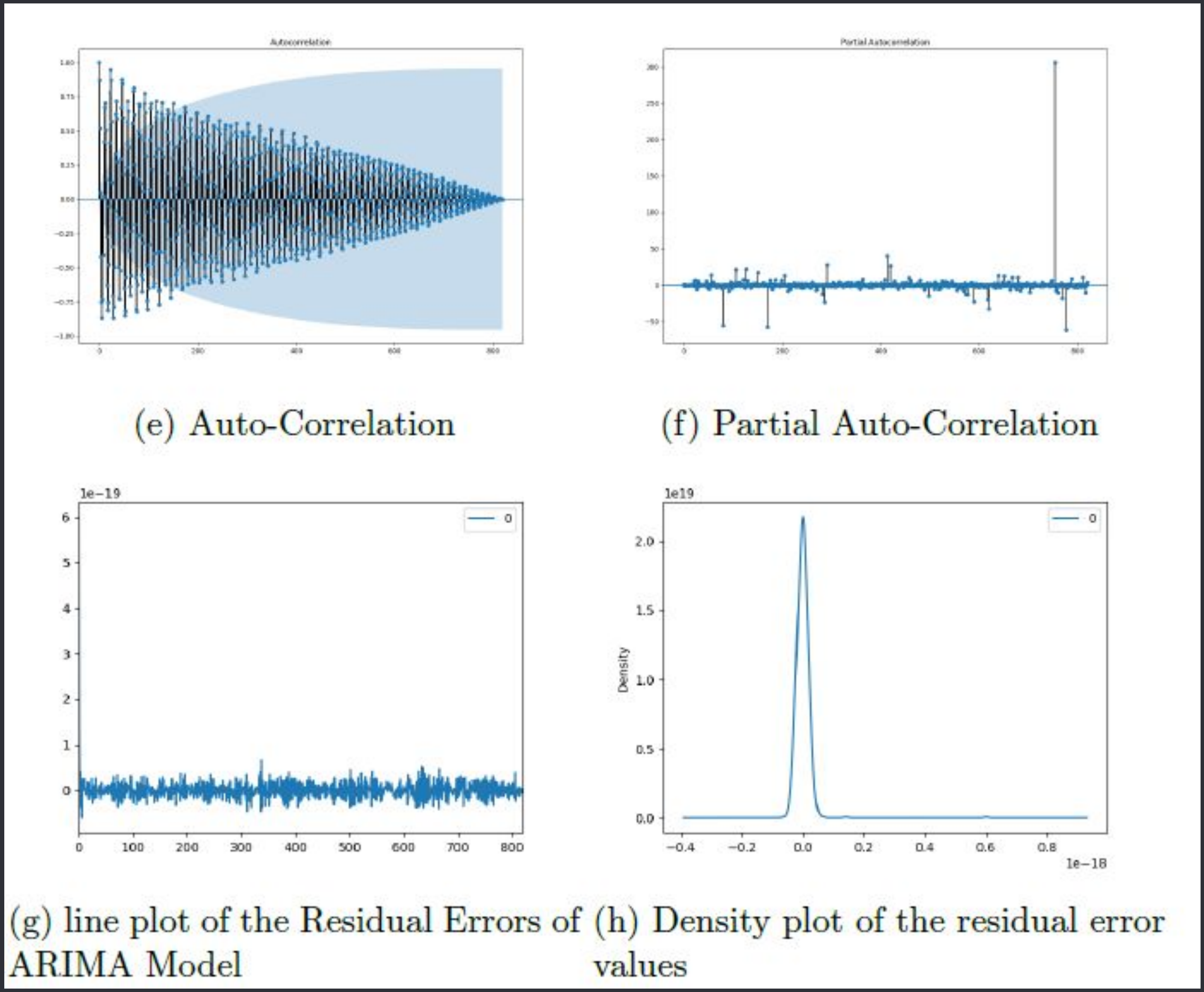
- We have extracted a section of the data for our purpose.
- The data is then decomposed into their respective seasonality and trends so as to gain better insight into the time series. For the Power Spectral Density, all the possible methods (viz. FFT, Welch's method and Lomb-Scargle method) have been applied.
- For the ARIMA modelling, we first obtained the plot of the autocorrelation and partial autocorrelation.
- The ACF gives us an estimate of the order of the MA model, while the PACF gives an estimate of the AR modelling. The results of both of them are combined so as to create the required ARIMA model.

RESULTS OBTAINED - 1



(a) FFT Periodogram
(b) Welch Periodogram
(c) Lomb-Scargle Periodogram
(d) Decomposition Results

RESULTS OBTAINED - 2



(e) Auto-Correlation

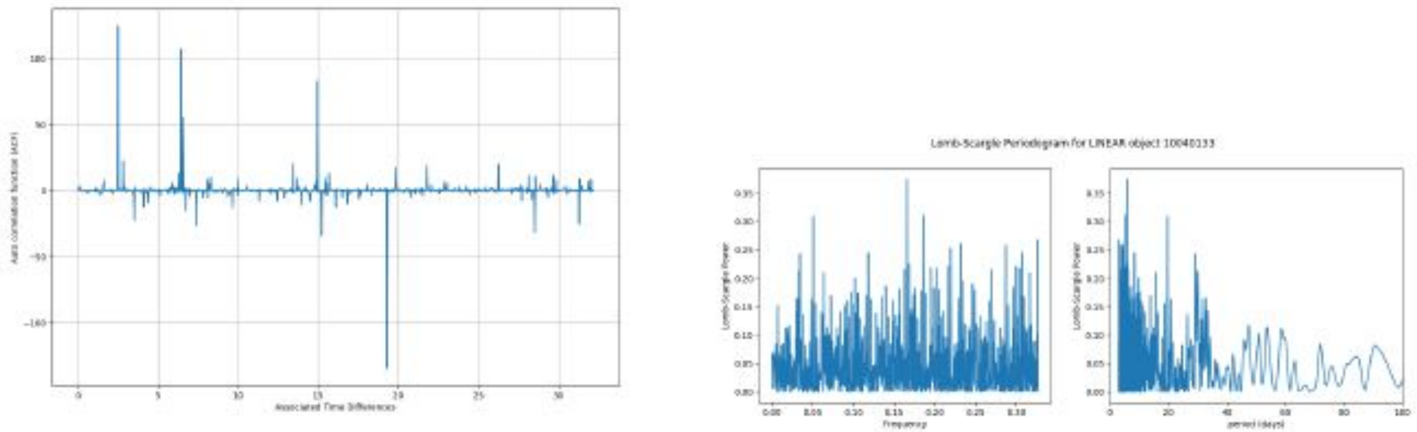
(f) Partial Auto-Correlation

(g) line plot of the Residual Errors of ARIMA Model

(h) Density plot of the residual error values

LINEAR DATASET

- For the LINEAR data, we have only the option of applying the Lomb-scargle periodogram method so as to create the required Power Spectral Density.



(a) Auto Correlation function (b) Lomb-Scargle Periodogram

CONCLUSION

- Through this project, we aim to analyse and experiment over the astronomical data, using the time series data generated over several experiments and observations.
- We aim to explore and analyse different method of statistical techniques of time series analysis viz. Decomposition, Spectral Analysis and Forecasting.
- Our aim is to deal with big, uneven data, and apply different statistical methods to generate insight into the data.

THANK YOU!