Final Project – Statistics and Signal Processing

Author: Lúcia Moreira

Date: 07/02/2020

Subject: ***Gravitational waves discovery***

Summary

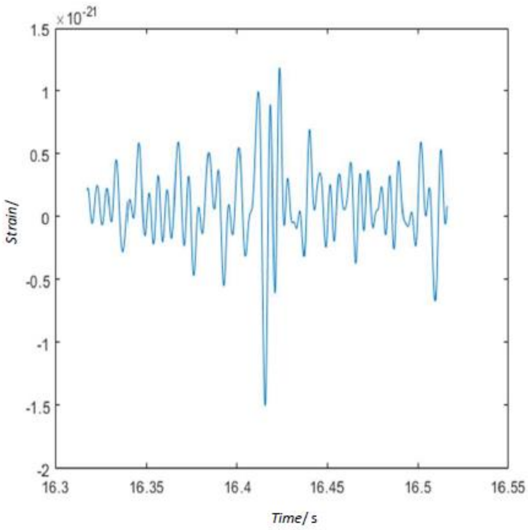
This project is a continuation of the topic initiated on the reports prepared during the continuous evaluation phase (please see Reports 1 and 2). Project nr. 1 considered the thematic involved in the signal herein approached, shortly: introduction to the theme, physical sensors involved, physical understanding of the strain length involved, signal denoising, periodogram and power density spectrum analysis. Report nr. 2 considered the parametric ARMA modelling of the strain signal. Due to the nature of the signal with multiple frequencies involved, ARMA modelling showed to fit the data quite poorly, and as expected. However, due the pedagogical interest of the understanding of such procedure the parametric approach showed to be relevant.

Regarding the final project, here there is a focus on the non-parametrical modelling of the strain signal using continuous wavelets (frequency-domain) and exponential smoothing (time-domain). Moreover, and also so due to the pedagogical relevance of the topic, GARCH modelling is also considered. Non-parametric model analysis for both the wavelet method and exponential smoothing provided a very good fitting to the data set. In its turn, GARCH parametric modelling despite improving the ARMA fitting provided the worse fitting to the data set, as expected.

Introduction

As described in the previous reports [1,2], the first detection of a gravitational wave signal occurred on September, 14th 2015 at the LIGO detectors in USA related to the merge of two black holes 1.34 billion light years away. The strain signal from the Livingston detector (L1) herein considered was filtered with the band pass [50 Hz, 250 Hz], followed by a notch to remove the 60 Hz interference from the AC electrical grid and its harmonics, as previously analyzed in the first report [1]. Without performing filtering and de-notching only random noise is observed (please see Report nr. 1).

The filtered signal around the binary black hole merge event is reprinted in Fig. 1 [1] (left), together with the respective power density spectrum (PDS) and periodogram, showing a strong magnitude signal at 180 Hz that explains the major variance in this data set.



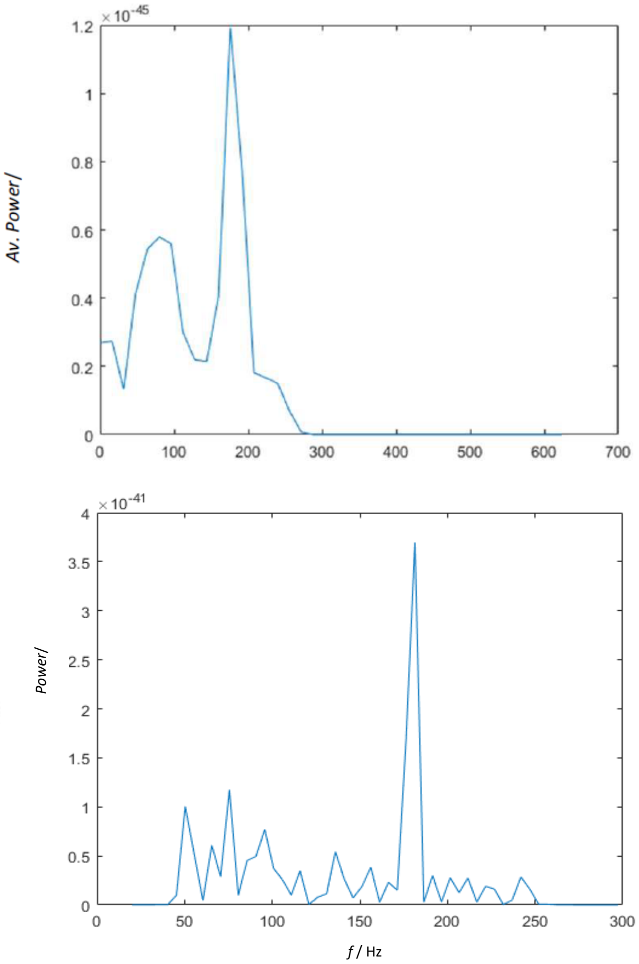


Fig. 1 - Signal amplification around the 16 s (left). PDS from the Welch´s method (top right). Periodogram (bottom right)

Data analysis

Modelling of the gravitational wave signal was performed by both parametric (ARIMA + GARCH and only GARCH) and non-parametric modelling (continuous wavelet transform and exponential smoothing).

Parametric modelling

Previously, report nr.2 [2] considered ARMA(p,q) modelling of the filtered signal performed in Matlab. ARMA modelling included auto-regressive *p* orders from 0 to 5 simultaneously with moving average *q* orders also from 0 to 5, corresponding to the total fitting of 36 different ARMA models. Such parametric ARMA modelling showed not to provide a good fit to the gravitational wave signal due to the time-frequency dependency in the strain signal.

Herein, an extended ARMA model fitting was revisited now using using AIC and BIC criteria and performed in R using “astsa” and “forecast” libraries. For that, ARMA modelling now included auto-regressive *p* orders from 0 to 9 simultaneously with moving average *q* orders also from 0 to 9, corresponding now to the total fitting of 100 different ARMA models using sarima() function. Data set was scaled by $1x10^{22}$ and a first difference introduced otherwise R would assume zero values for the whole dataset and sarima() function would return an error, respectively.

Based on AIC and BIC, an ARMA(4,5) model is obtained, indicating that previously in report nr. 2 [2], the range of order in the ARMA modeling performed in Matlab from 0 to 5 was indeed a good choice. However, and as previously observed, an ARMA(4, 5) anyway provides a quite bad fitting to this dataset. However, this step is necessary for the next stage of considering a GARCH model to the residuals of the ARMA(4,5) model and evaluate if an improvement in the parametric fitting is obtained.

Fig. 2 shows the partial auto correlation (PACF) and autocorrelation (ACF) functions for the squared residuals of the ARMA(4,5) model. Indeed, PACF and ACF shows some relevant cluster correlations justifying a GARCH modelling approach, majorly at lower lags, indicating that an ARCH(1) process seems to be quite evident (strong signal at lag 1 in PACF).

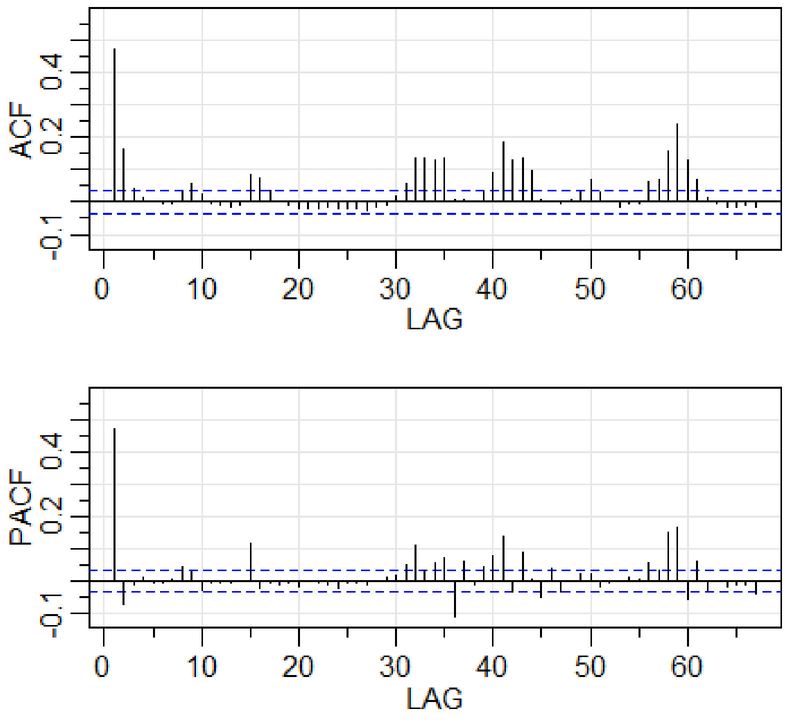


Fig. 2 - ACF and PACF analysis of the squared ARMA(4,5) model residuals.

According to that, the choice of the best GARCH model to the residues was performed for a short range of orders using BIC criteria. For that, GARCH modelling included “auto-regressive”-type p orders from 1 to 5 simultaneously with “moving average”-type q orders from 0 to 5, corresponding to the total fitting of 30 different GARCH models using garch() function.

Orders higher than 5 were not included due to the increased problem of overfitting when higher orders are used. The negative log-likelihood output from this function was suitably converted to BIC.

BIC criterion indicates that a GARCH(1,1) model is the best fit for the ARMA(4,5) squared residuals. Summary of the parametric fitting with this model indicated that both “alfa” and “beta” coefficients in the GARCH model are statistically significant for a 95 % confidence level despite the “ar4” coefficient is not statistically significant. Analysis of the final fitting to the strain signal did not show any improvement to the single ARMA model when using GARCH in the residuals (not shown).

A final attempt on parametric modelling was considered by fitting exclusively a GARCH model to the original strain signal. BIC criterion was also used to estimate the GARCH orders ranging from “auto-regressive”-type p orders from 1 to 5 simultaneously with “moving average”-type q orders from 0 to 5. BIC criterion indicated that an ARCH(1) process provides a better fitting to the strain signal. Fig. 3 shows the ARCH(1) fitting to the data set. GARCH modelling indeed captures better the more random oscillating nature of the signal than an only ARMA process, but fitting is anyway quite poor. This analysis indicates that a parametric modelling in this dataset is not relevant due to the multiple frequencies involved.

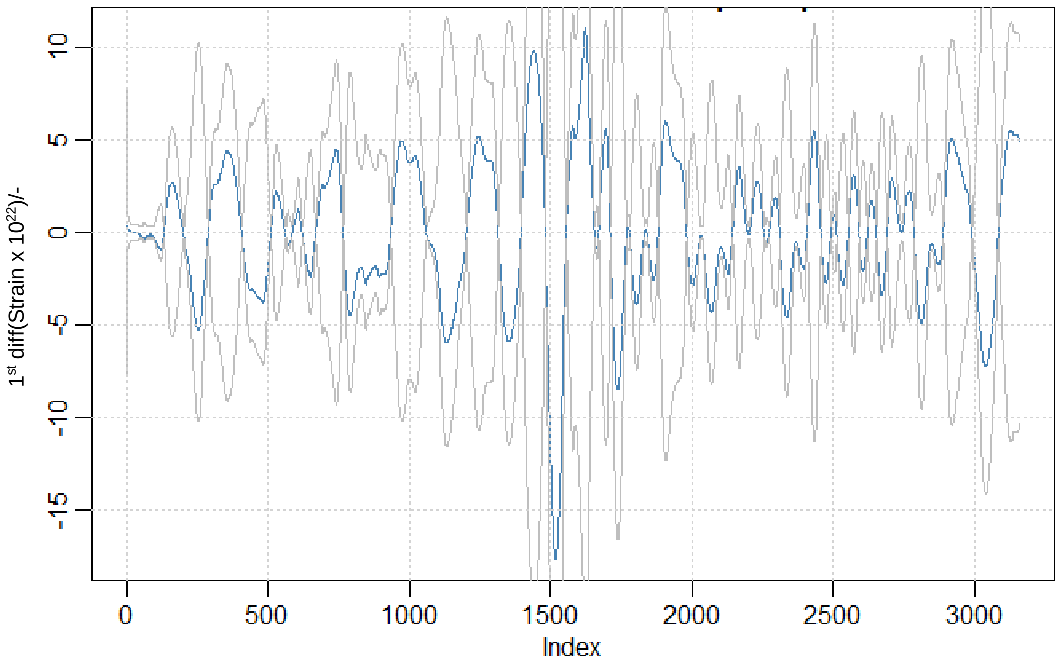


Fig. 3 – ARCH(1) fitting to the strain signal.

<u>Non-parametric modelling</u>

Exponential Smoothing [3]

In time-domain analysis, the concept behind simple exponential smoothing is to give larger weights to more recent observations than to observations from the distant past and the weights decay exponentially as the observations get older.

The simple exponential smoothing is suitable for data with no clear trend or seasonal pattern that was later extended by the Holt's linear method by two smoothing equations (one for the level and one for the trend, with α and β smoothing parameters). Later, the extended Holt's method was modified to also capture seasonality. The Holt-Winters seasonal method involves three smoothing equations — one for the level, one for the trend, and one for the seasonal component, with the corresponding smoothing parameters being α, β and γ.

In such models there are two approaches: one with additive errors and one with multiplicative errors. To distinguish between a model with additive errors and one with multiplicative errors, a third letter to the model classification system was added such as: (Error, Trend, Seasonal). The models can be estimated in R using the ets() function in the forecast package using the three-letter code. The possible inputs are "N" for none, "A" for additive, "M" for multiplicative, or "Z" for automatic selection. So, the possibilities for each component are: Error ={={A,M}}, Trend ={={N,A,Ad}d} and Seasonal ={={N,A,M}}.

Fig. 4 shows the output summary of the ets() fitting to the strain (without the first difference) dataset using automatic selection of the model. An additive model both for the level and trend and no seasonal component was the model automatically selected. Fig. 5 shows the fitting and original data where there is an outstanding fitting to the data.

```
ETS(A,A,N)

Call:
 ets(y = grav_ts, model = "ZZZ", damped = NULL, alpha = NULL,

 Call:
     beta = NULL, gamma = NULL, phi = NULL, additive.only = FALSE,

 Call:
     lambda = NULL, biasadj = FALSE, restrict = TRUE, allow.multiplicative.trend = FALSE)

  Smoothing parameters:
    alpha = 0.9997
    beta  = 0.9954

  Initial states:
    l = 2.0093
    b = 0.0927

  sigma:  0.0124

      AIC      AICc       BIC
-2235.175 -2235.157 -2204.742

Training set error measures:
                      ME       RMSE         MAE        MPE     MAPE       MASE      ACF1
Training set 0.0000192696 0.0124173 0.009384973 -0.1974347 1.559421 0.07044987 0.9954402
```

Fig. 4 – ETS(A,A,N) fitting summary to the strain signal.

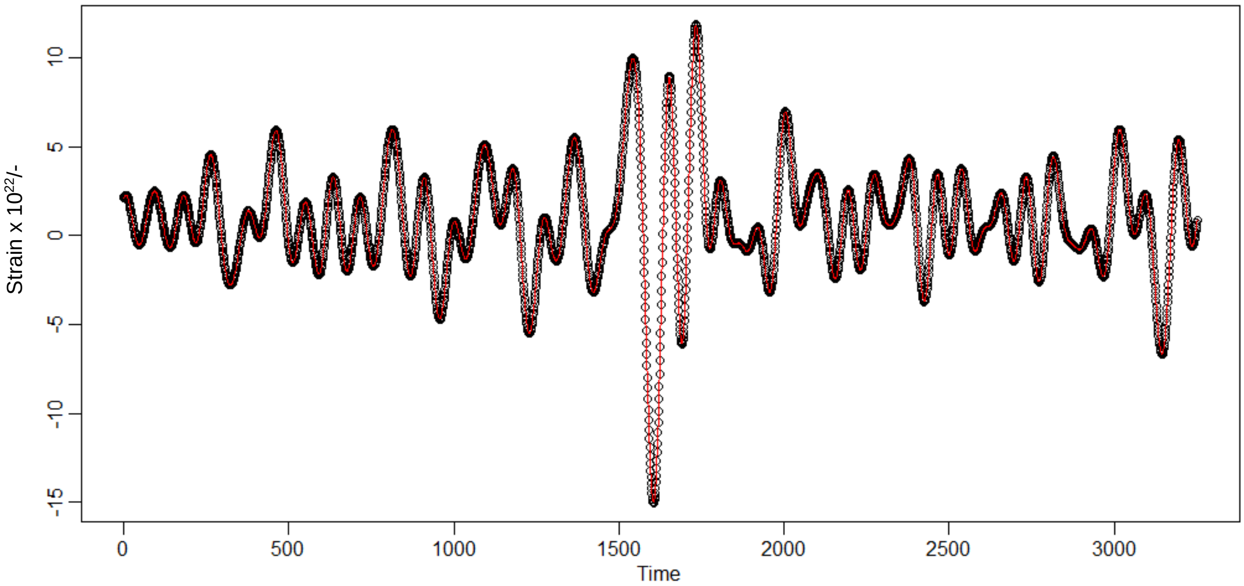


Fig. 5 – ETS(A,A,N) model fitting (red line) comparing with the strain signal (open black circle).

Continuous Wavelet Transform

The advantage of such non-parametric approach used in the frequency-domain is that the wavelet method is ideal for localized events and brings excellent temporal resolution across nearly the entire frequency range. Such localized event is exactly the main characteristic of the gravitational wave herein considered.

Fig. 6 shows the scalogram of the strain signal using the morse wavelet; a very defined high magnitude area can be observed related to the passage of the gravitational wave through the detectors.

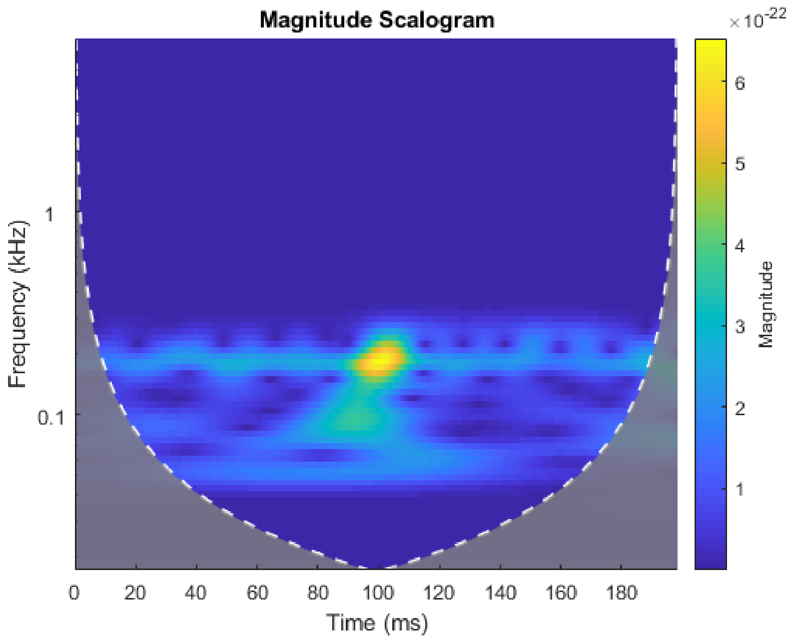


Fig. 6 – Scalogram of the strain signal.

The determination of the wavelet coefficients and comparison of the fitting to the data set was performed using the APP Wavelet Analyzer from Matlab. The Wavelet Coefficient Selection 1D Box was chosen and a haar wavelet with a level=1 was selected. Fig. 7 shows the fitting and coefficients. One sees that the synthesized and original signals match perfectly with just one single level. Fig. 7 shows the overlaid comparison of the fitting with the signal.

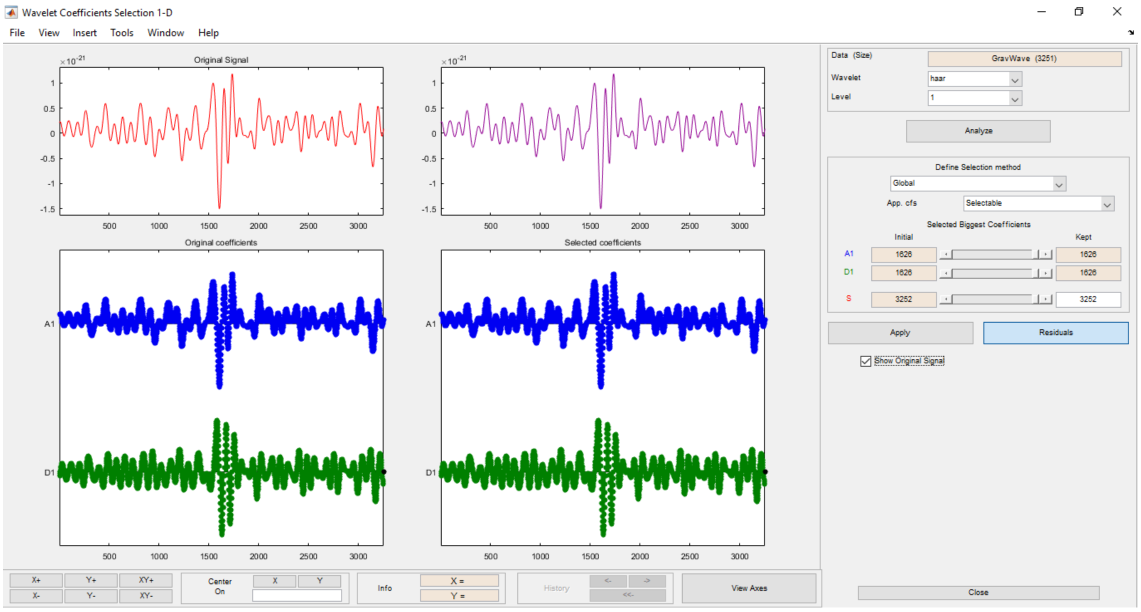


Fig. 7 – Wavelet Coefficient Selection 1-D App summary

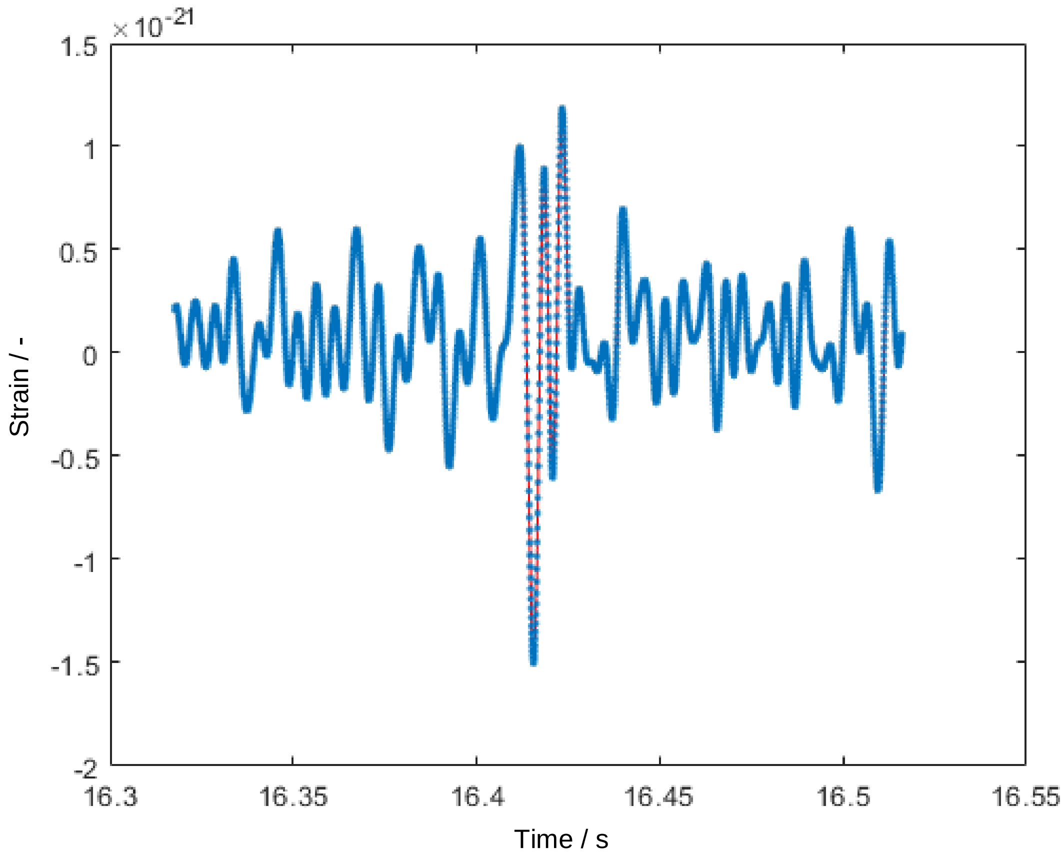


Fig. 8 - – Wavelet fitting (red line) comparing with the strain signal (blue points).

Conclusions

This project focused on the non-parametrical modelling of the strain signal using continuous wavelets (frequency-domain) and exponential smoothing (time-domain). Moreover, and also so due to the pedagogical relevance of the topic, GARCH modelling was also considered. Non-parametric model analysis for both the wavelet method and exponential smoothing provided a very good fitting to the data set. In its turn, GARCH parametric modelling despite improving the ARMA fitting provided the worse fitting to the data set, as expected.

References

[1] - Gravitational waves discovery, Work n. 1 – Statistics and Signal Processing, author: Lúcia Moreira, date: 10/11/2019.

[2] - Gravitational waves discovery, Work n. 2 – Statistics and Signal Processing, author: Lúcia Moreira, date: 30/11/2019.

[3] - Chapter 7 from the book “Forecasting Principles”, by Rob Hyndman.