

# Monthly mean total sunspot number analysis and forecast

Time Series and Forecasting - Final Project

Klismam F. Pereira

26/01/2021

---

## Introduction

**Sunspots** are earth sized dark areas that occur at the surface of the Sun, and are commonly found in pairs that have magnetic fields pointing in diverging directions. In them, the magnetic field is higher than anywhere else on the Sun's photosphere (i.e. surface), and 2500 times stronger than Earth's own field. They appear as darker spots because they're colder than the surrounding area - while sunspots temperature is around 3593.33 °C, the Sun's surface is around 5537.78 °C. These spots are usually composed of a darker region called *umbra*, and a lighter surrounding region called *penumbra* (Weather Forecast Office, National Weather Service n.d. [1]).

Sunspots are strongly related to **solar flares** and to **geomagnetic storms** that occur here on earth. Solar flares are gigantic explosions that happen near sunspots, and can release up to a billion megatons of TNT of energy. They also emit x-rays and magnetic fields that cause said geomagnetic storms. As a result, during sunspot maxima, more solar flares and more geomagnetic storms happen. It is quite important to understand when said events take place, since **their effects on our daily lives can be quite disruptive**, among which are the increase in Northern and Southern lights, possible disruption in radio transmissions and power grids, damaging electronics in satellites, etc (Weather Forecast Office, National Weather Service n.d. [1]).

Given their impact, it is needless to stress how important it is to understand their mechanisms. For that end, the **sunspot number** is used. The sunspot number, also called sunspot index, is the oldest solar activity index used to characterize the Sun's **eleven year cyclical behavior**. It can be found in the abstract of several papers, and has application in many fields of study such as climatology, meteorology, space physics, etc (Weather Forecast Office, National Weather Service n.d. [1], Berghmans et al. n.d. [2]).

It's history begins when sunspots started being systematically observed in the 17th century after the invention of the telescope. Years later, in 1843 Heinrich Schwabe published his discovery on the ten-year cycle of the number of sunspots, as a result of studying his recordings of the number of sunspots observed daily since 1826. The effects of the solar magnetic cycle had already been observed in 1803, though, when Ritter reported that auroras are more commonly sighted during specific time intervals, nowadays called solar maxima (Berghmans et al. n.d. [2]).

Only in 1852 Edward Sabine, chief British promoter of magnetic studies, pointed out the "coincidence" between the period and epochs of minima and maxima described by H. Schwabe and the ones found for magnetic variation, and in the same year Rudolf Wolf presented his results showing that the actual period was of 11.1 years. He also derived spot numbers from scattered data back to 1749. Later, with more observatories at disposal and interested in the matter, the continuity of counts is preserved through joint effort (Berghmans et al. n.d. [2]).

In 1980, the production of the sunspot index, i.e. the **International Sunspot Number** becomes responsibility of the Sunspot Index Data Center (SIDC), of the Royal Observatory of Belgium. Under their

administration, the number of contributing stations doubled in a few years, improving accuracy and stability of the international sunspot number, and in the 2000, the SIDC became a Regional Warning Center for the Western Europe, which is a space weather forecast and monitoring center of the International Space Environment Service (Berghmans et al. n.d. [2]).

Space weather is an emerging science with great and direct relevance to technological systems. Beyond the already noted impacts, solar influences might have an important effect on the evolution of the earth's climate, thus understanding the solar cyclic activity is important to humanity as whole (Weather Forecast Office, National Weather Service n.d. [1], Berghmans et al. n.d. [2]).

## Objective

Given the importance of such an interesting subject, this project's aim is to explore the monthly mean total sunspot number using time series analysis tools, explain some of the series behavior by fitting models from the ARIMA and Exponential Smoothing families, and finally obtain forecasts based on the best of these models, chosen through criteria such as the information criterion and forecast metrics.

All data process, analysis, and model fitting is done with *R*, a language and environment for statistical computing and graphics.

## The time series

The time series consists of the monthly mean total sunspot number from 01/1749 to 11/2020. As described in the introduction section, the earliest records were compiled by Rudolph Wolf. The monthly mean total sunspot number is obtained by calculating the average of the number of daily total sunspot number over each month (Berghmans et al. n.d. [2], Sunspot data from the World Data Center SILSO [3]).

The data set contains seven features - Year, Month, Date in fraction of year, Monthly mean total sunspot number, Monthly mean standard deviation of the input sunspot numbers, Number of observations used to compute the monthly mean total sunspot number, Definitive/Provisional marker (1 means the value is definitive, 0 means its still provisional) (Sunspot data from the World Data Center SILSO [3]).

From the seven features, three are used in this project:

- Year
- Month
- Monthly mean total sunspot number

---

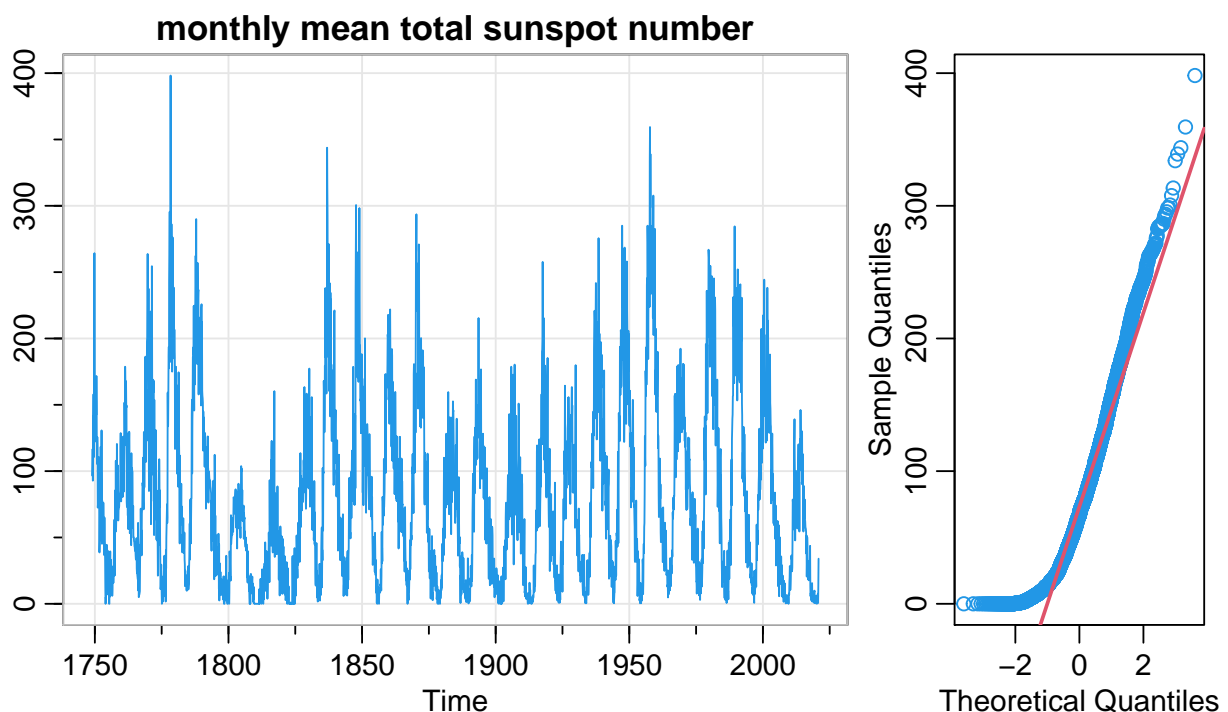
## Exploratory Data Analysis

In this section, the time series is explored graphically, and its main characteristics are described. The monthly mean total sunspot number for the two first and last years are shown.

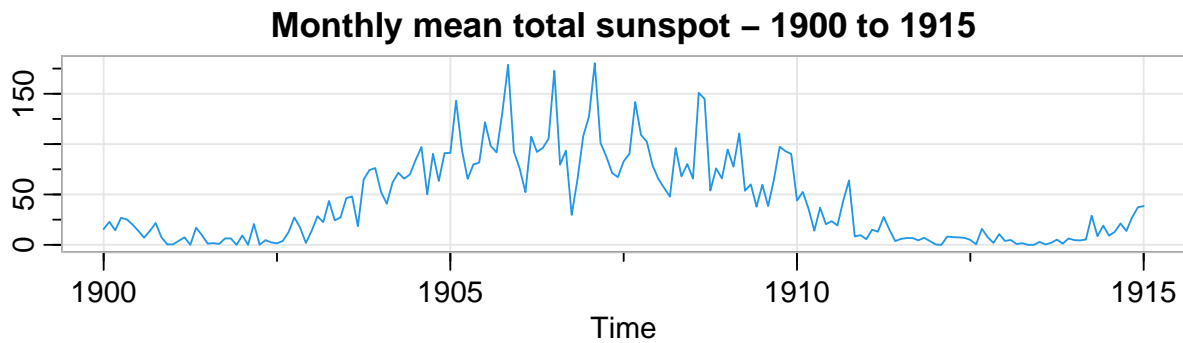
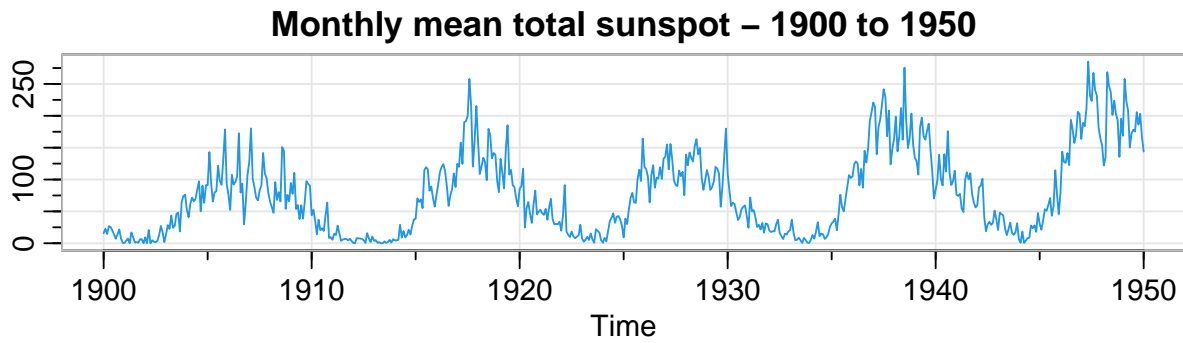
##	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
## 1749	96.7	104.3	116.7	92.8	141.7	139.2	158.0	110.5	126.5	125.8	264.3	142.0
## 1750	122.2	126.5	148.7	147.2	150.0	166.7	142.3	171.7	152.0	109.5	105.5	125.7

##	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
## 2019	7.7	0.8	9.4	9.1	9.9	1.2	0.9	0.5	1.1	0.4	0.5	1.5
## 2020	6.2	0.2	1.5	5.2	0.2	5.8	6.3	7.6	0.7	14.4	34.0	

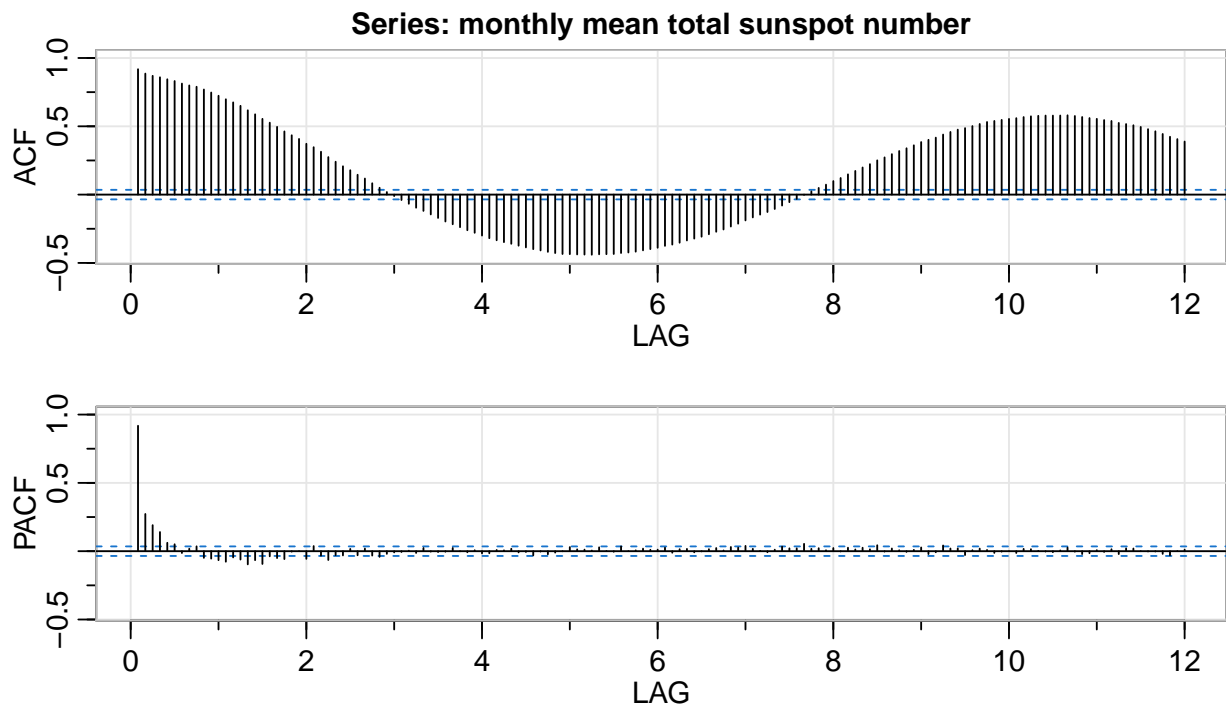
The following plot represents the complete time series and its respective QQ-plot. It seems 25 sunspot maxima occurred since 1750, and latest entries suggest the latest minimum is already past. The cycles observed are compatible with the long 11 year period description given in the introduction section. There is no clear trend, and the cycles aren't ruled by the more common yearly seasonality found in several time series. Differences in the maxima magnitudes imply heteroscedasticity, or non-constant variance. There are no discontinuities or abrupt changes in this specific time period. The QQ-plot denotes departure from normality in the lower and upper quantiles. The prior might be due to the sunspot number having a minimum of zero counts.



Two smaller periods of the time series are plotted below. The first graphic presents a window from 1900 to 1950, where the 11 years cycle is more evident. The second focus on the period between 1900 and 1915, in what it seems to be a full cycle. There appears to be some cyclic behavior within the 11 year cycle, but its frequency can't be pinpointed to yearly seasonality.



Now, both ACF and PACF graphics for 144 lags are shown (i.e. twelve years).



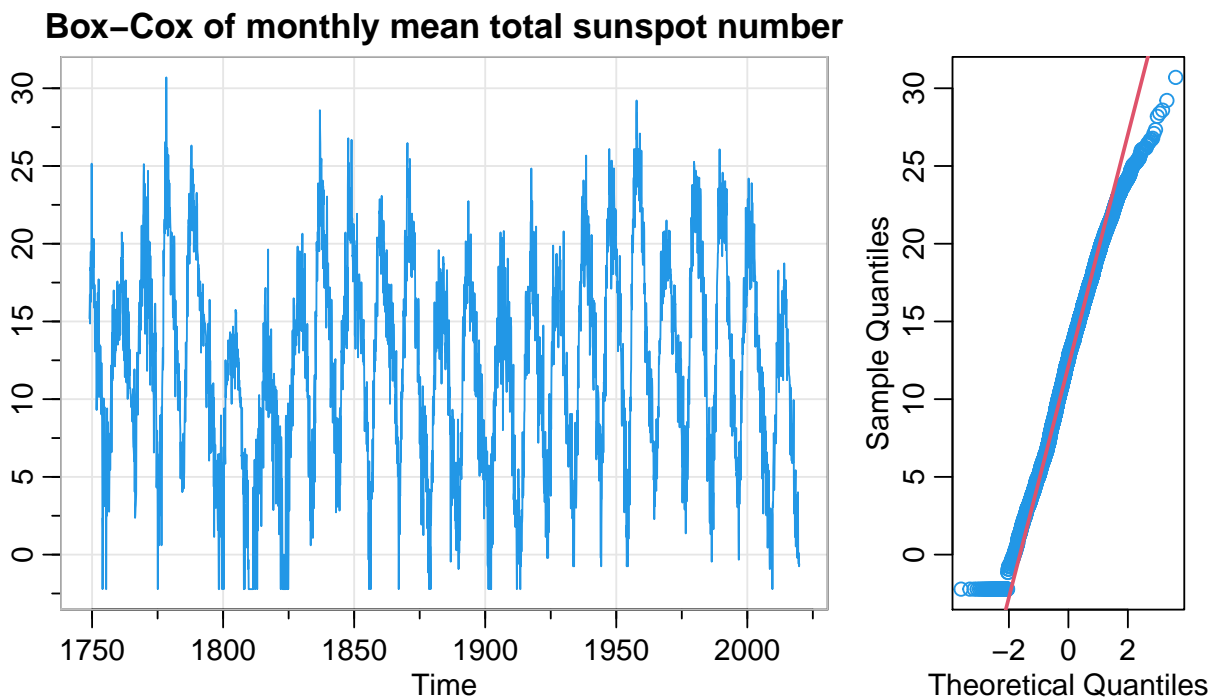
The ACF presents high correlations that do not seem to be decaying to zero, even though 144 lags are shown. This is characteristic of a long-memory process, in which correlations persist for a long time (Shumway & Stoffer, 2015 [4]).

Correlations change signals at around the three years mark (i.e. 36 lags) and reach a new negative maximum at around 5 years (i.e. 60 lags). The correlations change signal again and arrive at positive maximum close to 11 years (i.e. 132 lags). At a first glance, this ACF plot may imply some of the long cyclical behavior described in previous paragraphs, with high correlations occurring 5 and 11 years apart. The PACF presents high correlations at lags one to four, and there are also smaller partial correlations that are outside the significance threshold in the negative side of the plot.

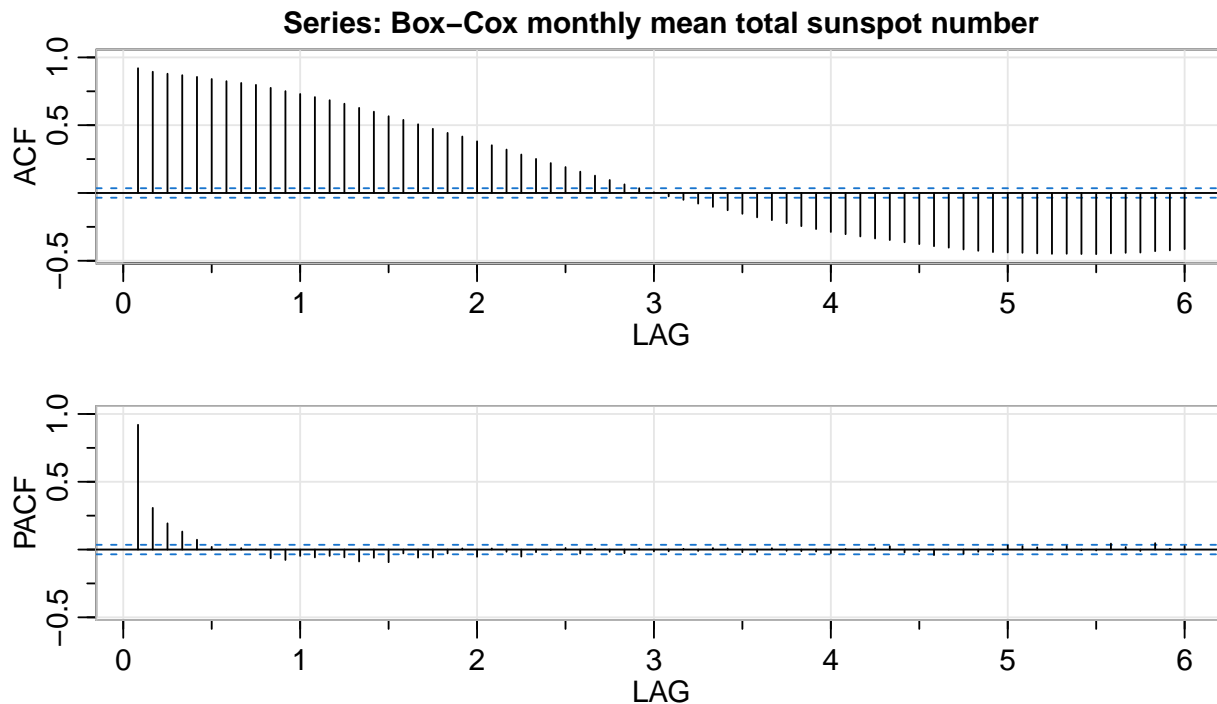
## Box-Cox transformation and train/test split

First, the time series is divided into train and test data sets in order to allow forecast evaluation. The test set comprises the last twelve months of observations - from 12/19 to 11/20 - and the train set has data from 01/1749 to 11/2019.

A Box-Cox transformation will be applied to stabilize variability, and thus improve approximation to normality. The plot below shows the Box-Cox transformed time series and its QQ-plot. Variability seems reduced, and the sample quantiles tend to comply with the theoretical, with the exception of the lower quantiles under two standard deviations. The parameter  $\lambda$  for the Box-Cox transformation is equal to 0.45.



The ACF and PACF plots of the Box-Cox transformed series are shown below. Their behavior is remarkably similar to what was observed and described in the original series correlation plots.



As before, the ACF plot indicates a long-memory process, for which an ARFIMA model might be appropriate, while the PACF plots exhibits significant correlations for lags one to four. It is not clear which degrees of a simpler auto-regressive (AR), moving average (MA) or ARIMA model would be an interesting first approach. Thus, experiments using the Box-Jenkins method for model selection, as well as the functions *auto.arima* and *arfima* (both from *R* package *forecast*) are described in the next section.

## ARIMA models experimentation and selection

### Box-Jenkins method

In this section the Box-Jenkins method is used to select and fit an ARIMA model to train data. Overall, the method consists on the following steps (Holmes et al. 2021[5]):

1. Evaluate stationarity
2. Selection of the differencing level ( $d$ ) – to fix stationarity problems
3. Selection of the AR level ( $p$ )
4. Selection of the MA level ( $q$ )
5. Parameter estimation
6. Model checking

The output below presents the results of the Augmented Dickey–Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) unit root tests, used to aid in the selection of a proper difference order. The low p-value for the ADF test suggests evidence in favor of the alternative hypothesis that the time-series is stationary.

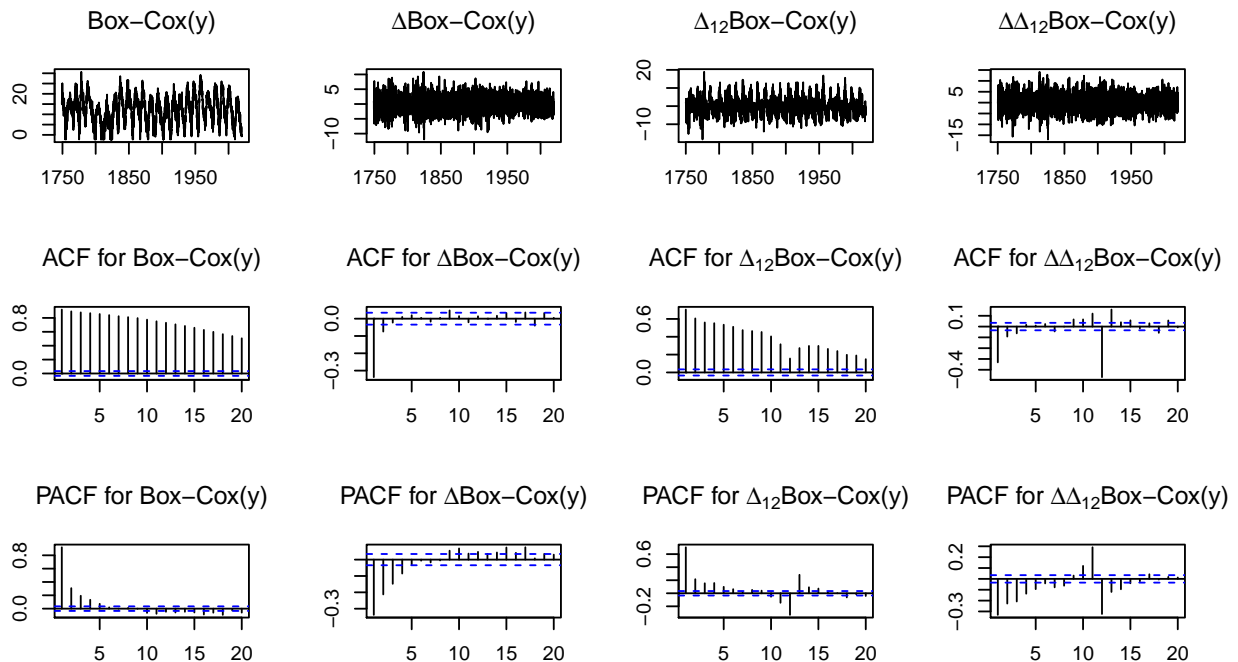
That said, the p-value equal to 0.01 from the KPSS test provides evidence against the null hypothesis that the time-series is trend stationary. Even though the ADF test suggested stationarity, the result of the KPSS test, ACF (correlations are not decaying to zero), and PACF plots indicate a non stationary series, in which some differencing is required (Hyndman & Athanasopoulos, 2018 [6]).

```
## Warning in adf.test(sunspot.BC, alternative = "s"): p-value smaller than printed
## p-value

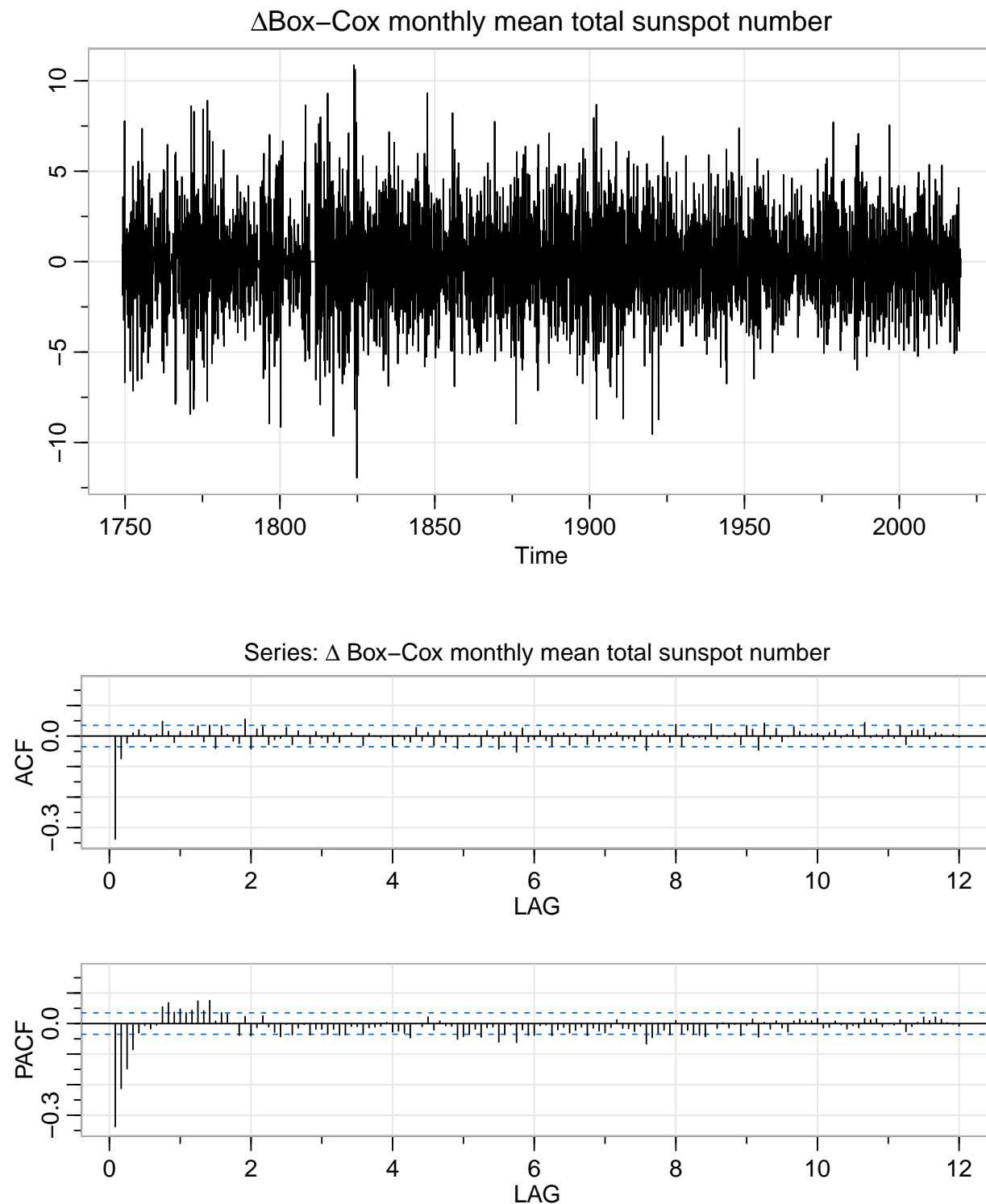
##
## Augmented Dickey-Fuller Test
##
## data: sunspot.BC
## Dickey-Fuller = -6.9026, Lag order = 14, p-value = 0.01
## alternative hypothesis: stationary

##
## KPSS Test for Trend Stationarity
##
## data: sunspot.BC
## KPSS Trend = 0.21294, Truncation lag parameter = 9, p-value = 0.01115
```

Below, several plots are shown to aid in the selection of a proper order for simple ( $d$ ) and seasonal ( $D$ ) differentiations. Note that until now seasonality is not a major feature of this series, and indeed the ACF and PACF plots related to seasonal differences (*i.e.* the ones that contain  $\Delta_{12}$ ) are not similar to the observed in a white noise process. Given the characteristics of the phenomena, it makes sense it is not ruled by earth's seasonality. Through the plots, it is decided that a simple differentiation ( $d = 1$ ) might suffice to explain correlation in the series.



Next, a first difference is applied to address the high correlations previously shown in the ACF plot. The time series, ACF, and PACF plots of the differenced series are shown again in more detail below. The time series plot resembles a white noise process - stationary, zero mean, approximately constant variance - while the ACF plot exhibits low correlations. That said, partial auto correlation values outside the significance threshold remain in the PACF plot.





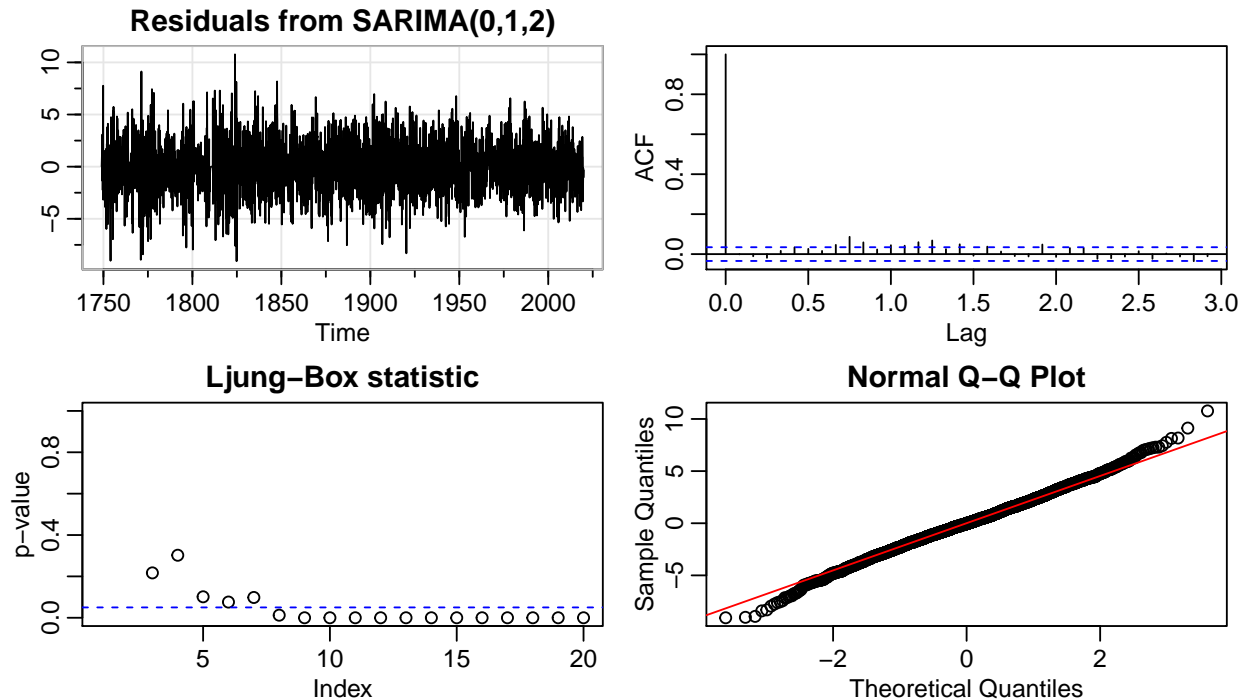
The plots hint at a close to stationary process. One may infer that an  $ARIMA(0,1,2)$  model might be appropriate to represent the time series since the PACF does not exhibit a cut-off while the ACF decays quickly, with a cut-off at lag two. From that, an experiment considering such a model is carried.

Nevertheless, the following output from the model fit reveals that even though both MA components have significant effects given that their confidence intervals do not contain zero, and the standardized residuals comply with a Normal distribution, their ACF exhibit correlations marginally outside of the significance threshold, and the Ljung-Box statistic presents low p-values, providing evidence that the residuals are in fact correlated.

```
## Series: traindat
## ARIMA(0,1,2)
## Box Cox transformation: lambda= 0.4503796
##
## Coefficients:
##          ma1      ma2
##      -0.4532  -0.0843
## s.e.   0.0175   0.0165
##
## sigma^2 estimated as 5.641:  log likelihood=-7422.08
## AIC=14850.16  AICc=14850.17  BIC=14868.42
```

Table 1: 95% Confidence Interval for estimates

	2.5 %	97.5 %
ma1	-0.4875870	-0.418902
ma2	-0.1165392	-0.052003

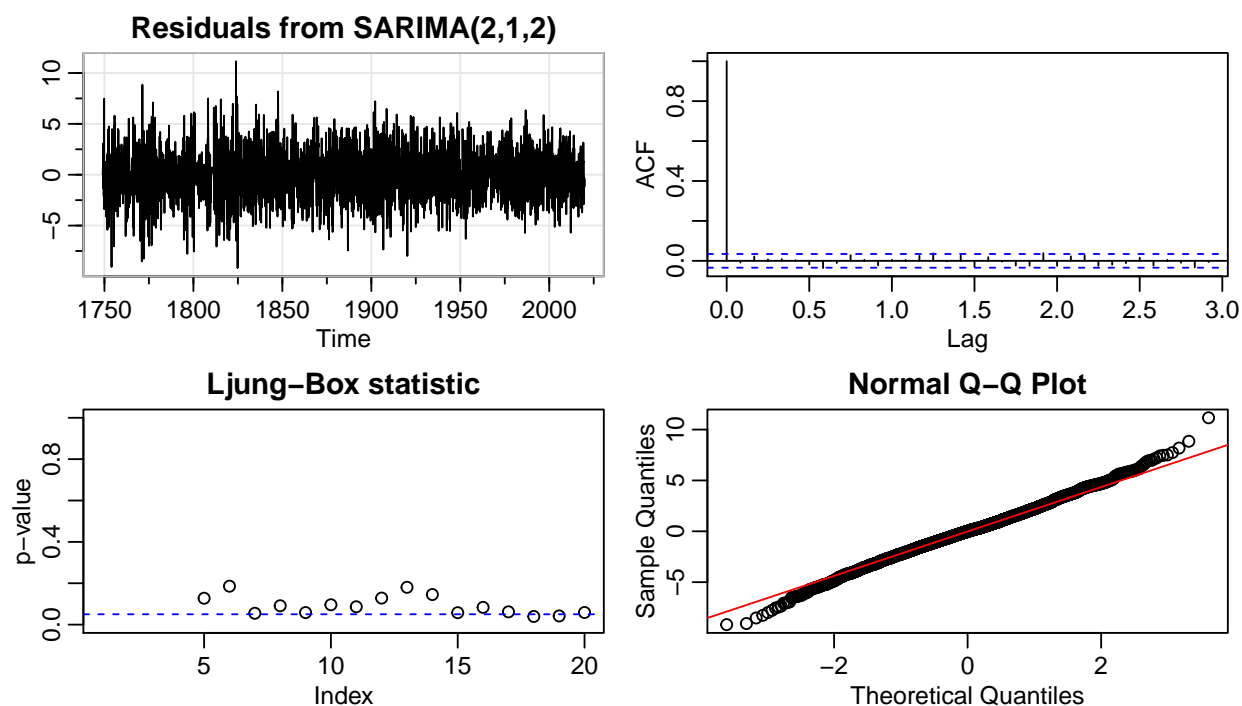


Experiments were carried considering different orders for the AR component of the model. The output below is relative to an  $ARIMA(2,1,2)$  model, in which the effect of all coefficients is significant, but the Ljung-Box statistic signals for some correlation among residuals at lags 18 and 19 (*i.e.* p-values under the significance threshold).

```
## Series: traindat
## ARIMA(2,1,2)
## Box Cox transformation: lambda= 0.4503796
##
## Coefficients:
##          ar1          ar2          ma1          ma2
##          1.2599   -0.3118   -1.7315    0.7730
## s.e.    0.0305    0.0287    0.0218    0.0207
##
## sigma^2 estimated as 5.5:  log likelihood=-7380.04
## AIC=14770.09   AICc=14770.11   BIC=14800.52
```

Table 2: 95% Confidence Interval for estimates

	2.5 %	97.5 %
ar1	1.2002382	1.3196316
ar2	-0.3681168	-0.2555225
ma1	-1.7741400	-1.6888416
ma2	0.7325040	0.8135200

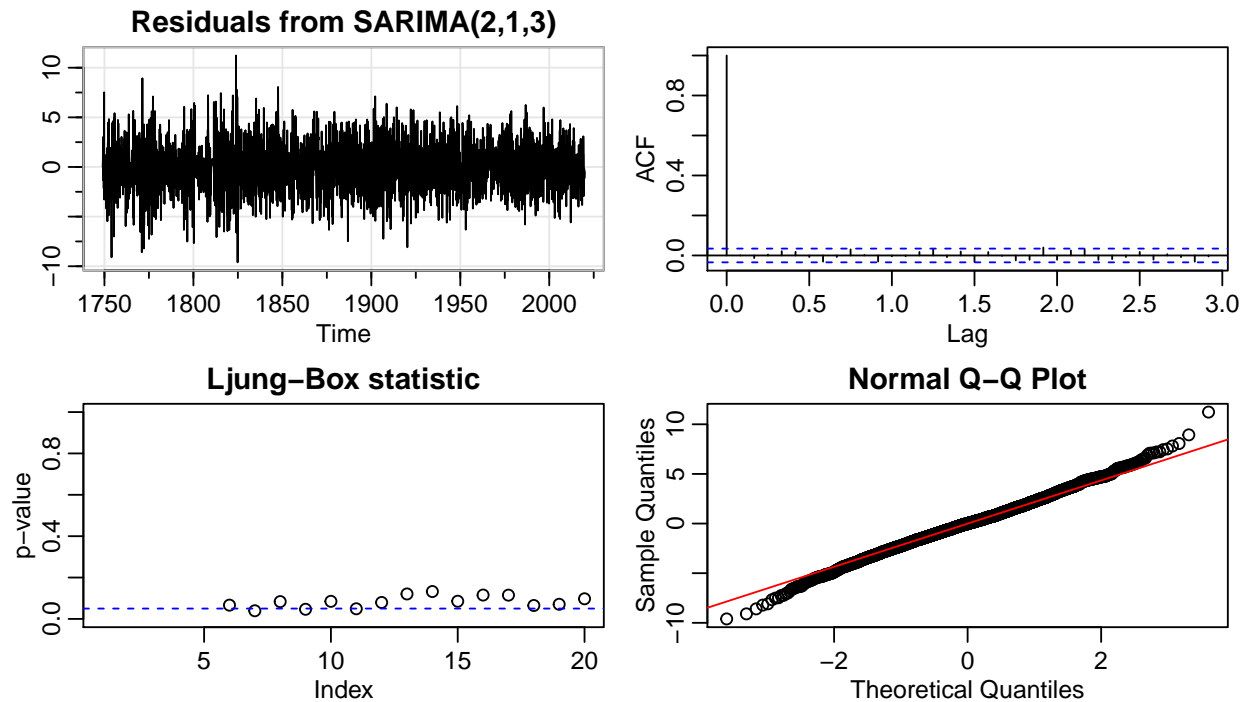


Below, an  $ARIMA(2,1,3)$  model is also fit. As with  $ARIMA(2,1,2)$ , the effect of all estimates is significant, but the Ljung-Box statistic show marginally significant correlation among residuals. That said, both models might be useful and shall be tested in forecasting.

```
## Series: traindat
## ARIMA(2,1,3)
## Box Cox transformation: lambda= 0.4503796
##
## Coefficients:
##          ar1      ar2      ma1      ma2      ma3
##          1.4774 -0.5198 -1.9602  1.1388 -0.1454
## s.e.    0.0899  0.0857  0.0953  0.1536  0.0625
##
## sigma^2 estimated as 5.491:  log likelihood=-7377.08
## AIC=14766.17  AICc=14766.19  BIC=14802.69
```

Table 3: 95% Confidence Interval for estimates

	2.5 %	97.5 %
ar1	1.3011972	1.6536441
ar2	-0.6877365	-0.3518878
ma1	-2.1469175	-1.7734419
ma2	0.8377922	1.4398402
ma3	-0.2678583	-0.0230097



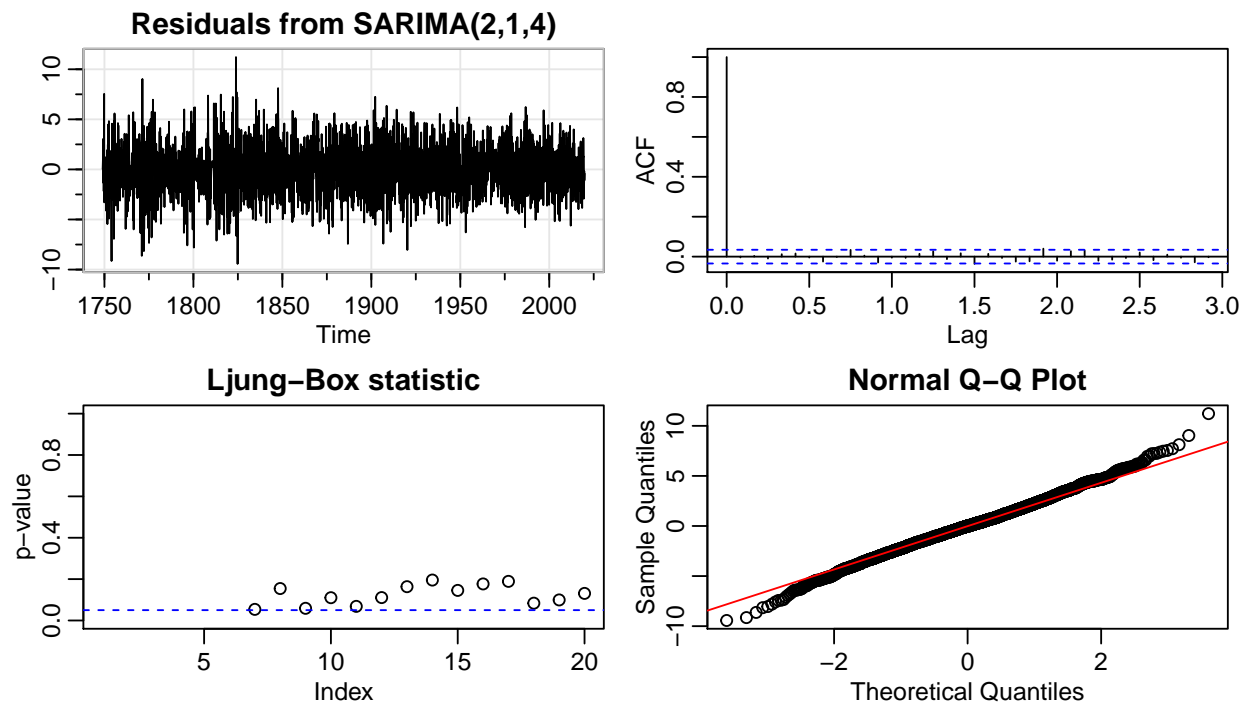
An  $ARIMA(2,1,4)$  model is also fit. Its  $ma_4$  coefficient is not significant because the 95% confidence interval contains zero, thus this model is rejected as it would potentially cause overfitting.

```
## Series: traindat
## ARIMA(2,1,4)
## Box Cox transformation: lambda= 0.4503796
```

```
##
## Coefficients:
##          ar1      ar2      ma1      ma2      ma3      ma4
##          1.6655 -0.6953 -2.1445  1.3800 -0.1565 -0.0563
## s.e.    0.1129  0.1063  0.1144  0.1636  0.0487  0.0310
##
## sigma^2 estimated as 5.486:  log likelihood=-7374.93
## AIC=14763.86  AICc=14763.89  BIC=14806.46
```

Table 4: 95% Confidence Interval for estimates

	2.5 %	97.5 %
ar1	1.4441432	1.8868108
ar2	-0.9035873	-0.4869425
ma1	-2.3687804	-1.9202827
ma2	1.0593544	1.7006236
ma3	-0.2519988	-0.0609748
ma4	-0.1170282	0.0043929



`auto.arima()`

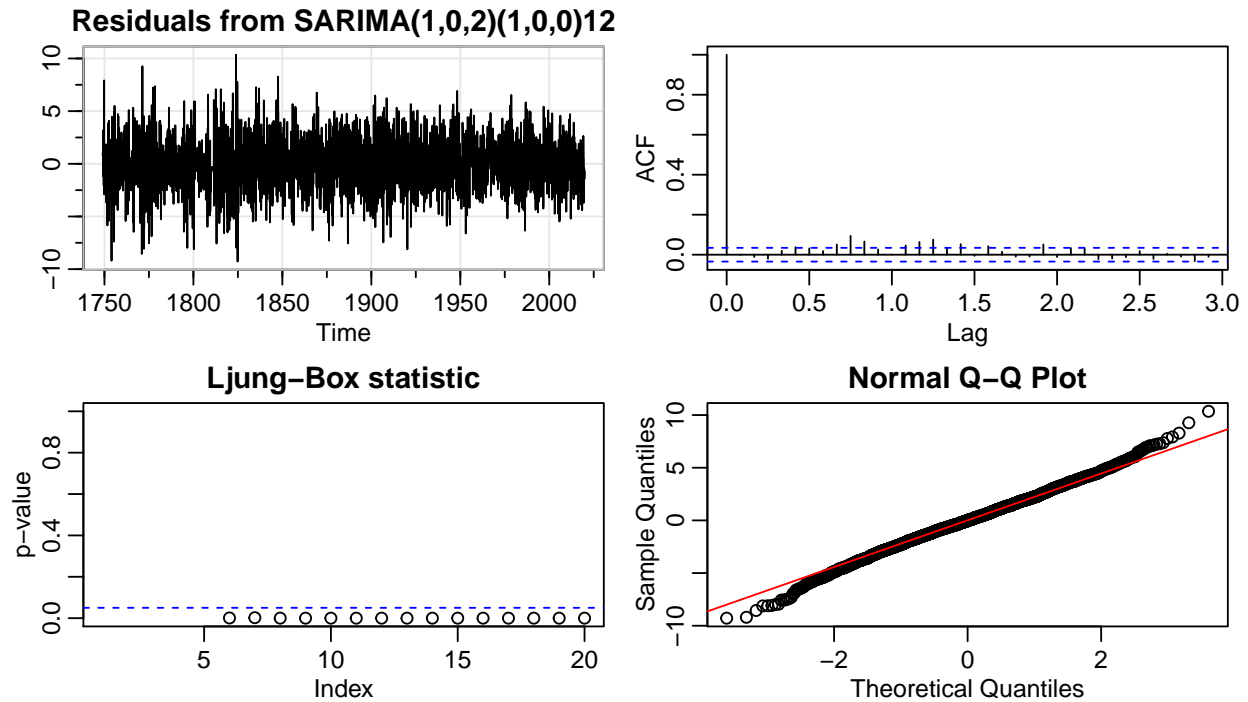
Next the function `auto.arima` from package *forecast* is tested in order to obtain a benchmark for our selections. It suggests a  $SARIMA(1,0,2)(1,0,0)_{12}$  model. As with other models evaluated, it fails to return uncorrelated residuals according to the Ljung-Box test.

```
## Series: traindat
## ARIMA(1,0,2)(1,0,0)[12] with non-zero mean
```

```
## Box Cox transformation: lambda= 0.4503796
##
## Coefficients:
##          ar1          ma1          ma2          sar1          mean
##          0.9836    -0.4480    -0.0835    0.0506    12.2536
## s.e.    0.0034     0.0179     0.0167    0.0180     1.2262
##
## sigma^2 estimated as 5.591:  log likelihood=-7409.34
## AIC=14830.68  AICc=14830.71  BIC=14867.21
```

Table 5: 95% Confidence Interval for estimates

	2.5 %	97.5 %
ar1	0.9769822	0.9902477
ma1	-0.4830709	-0.4128841
ma2	-0.1161754	-0.0508181
sar1	0.0152573	0.0859065
intercept	9.8503589	14.6568861



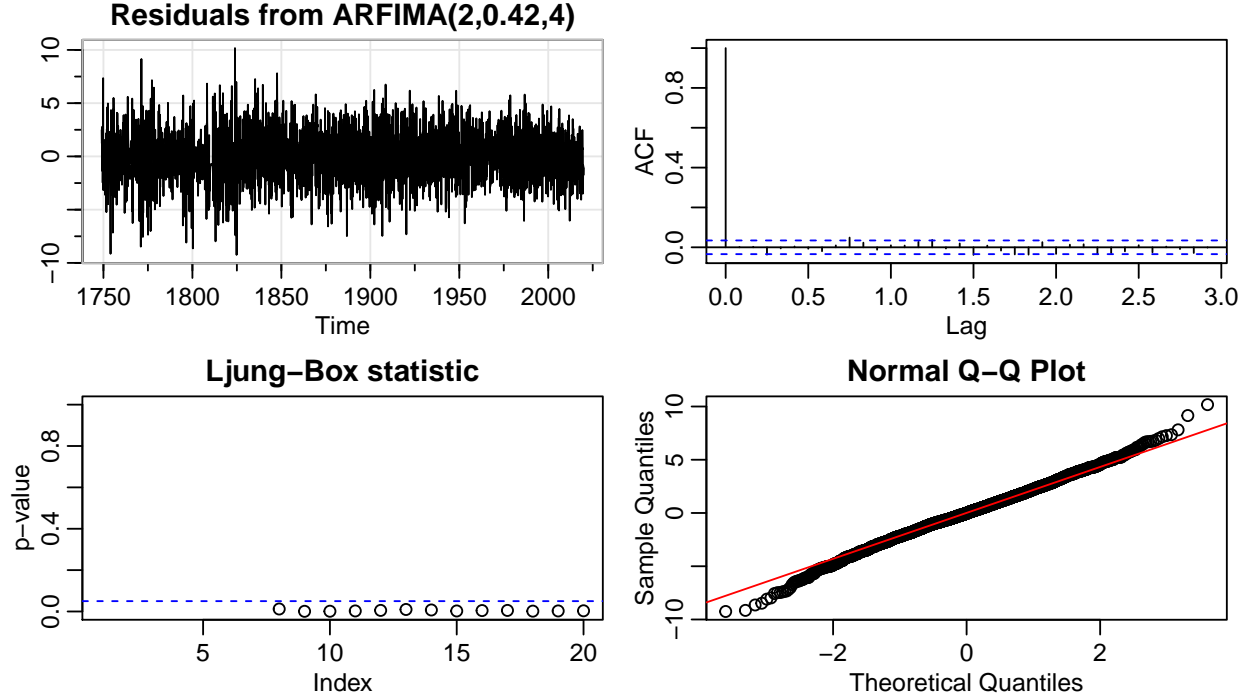
## ARFIMA

Finally, since the time series ACF plot presented characteristics of a long-memory process - correlations that do not decay to zero, persisting for a long time - a fractionally differenced ARFIMA model is fit. ARFIMA are useful in situations where differencing can be too harsh of an alteration. The function *arfima* from the package *forecast* returned an *ARFIMA*(2,0.42,4) model, in which all coefficients are significant. Nevertheless, the Ljung-Box statistic shows that residuals are correlated (Shumway & Stoffer, 2015 [4]).

```
##
## Call:
## arfima(y = traindat, drange = c(0, 0.5), estim = c("mle", "ls"), model = NULL, lambda = lambda)
##
## Coefficients:
##      Estimate Std. Error z value Pr(>|z|)
## d      0.4218310  0.0047692  88.449 < 2e-16 ***
## ar.ar1  1.9866600  0.0157133 126.432 < 2e-16 ***
## ar.ar2 -0.9891047  0.0169057 -58.507 < 2e-16 ***
## ma.ma1  1.9238807  0.0086239 223.088 < 2e-16 ***
## ma.ma2 -0.8973327  0.0100543 -89.249 < 2e-16 ***
## ma.ma3 -0.0843773  0.0190294  -4.434 9.25e-06 ***
## ma.ma4  0.0542123  0.0001704 318.114 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## sigma[eps] = 2.353479
## [d.tol = 0.0001221, M = 100, h = 7.794e-05]
## Log likelihood: -7396 ==> AIC = 14808.53 [8 deg.freedom]
```

Table 6: 95% Confidence Interval for estimates

	2.5 %	97.5 %
d	0.4124836	0.4311784
ar.ar1	1.9558625	2.0174575
ar.ar2	-1.0222394	-0.9559701
ma.ma1	1.9069782	1.9407832
ma.ma2	-0.9170388	-0.8776267
ma.ma3	-0.1216742	-0.0470804
ma.ma4	0.0538783	0.0545463



## ARIMA models selection

Models  $ARIMA(2,1,2)$  and  $ARIMA(2,1,3)$  display residuals marginally correlated, while the models fit by the functions *auto.arima* and *arfima* from the package *forecast*,  $SARIMA(1,0,2)(1,0,0)_{12}$  and  $ARFIMA(2,0.42,4)$  respectively, exhibit significant p-values for the Ljung-Box statistic. The residuals from all models present low correlations in the respective ACF plots, and adhere to the Gaussian distribution according to each QQ-plots. Even though the Ljung-Box statistic for some suggested rejection of the null hypothesis of uncorrelated residuals, all four models quoted in this paragraph will be used in forecasting and evaluation.

Models  $ARIMA(0,1,2)$  and  $ARIMA(2,1,4)$  will not be used. The first presents high correlation among residuals and the highest AIC (14868) among the simple differenced models fit. The second presents a non significant coefficient that may lead to overfitting.

The table below exhibits the chosen models' orders, number of parameters, as well as AIC, AICc, and BIC metrics. These models were fit using the package *forecast* to keep consistency among calculations.

Table 7: Chosen ARIMA models

Model	Order	N.par	AIC	AICc	BIC
mdl1	$ARIMA(2,1,2)$	4	14770.09	14770.11	14800.52
mdl2	$ARIMA(2,1,3)$	5	14766.17	14766.19	14802.69
mdl3	$SARIMA(1,0,2)(1,0,0)_{12}$	5	14830.68	14830.71	14867.21
mdl4	$ARFIMA(2,0.42,4)$	7	14808.53	14808.57	14857.22

Information criterion should only be compared among ARIMA models of the same orders of differencing (Hyndman & Athanasopoulos, 2018 [6]). The information criterion of the first differenced models are very similar -  $ARIMA(2,1,2)$  has a lower BIC while  $ARIMA(2,1,3)$  has lower AIC and AICc. All four models will be used in later sections to generate forecasts that can be evaluated regarding the test data set. Next, convenient Exponential Smoothing models are also chosen to enrich the discussion.

# Exponential Smoothing (ES) models experimentation and selection

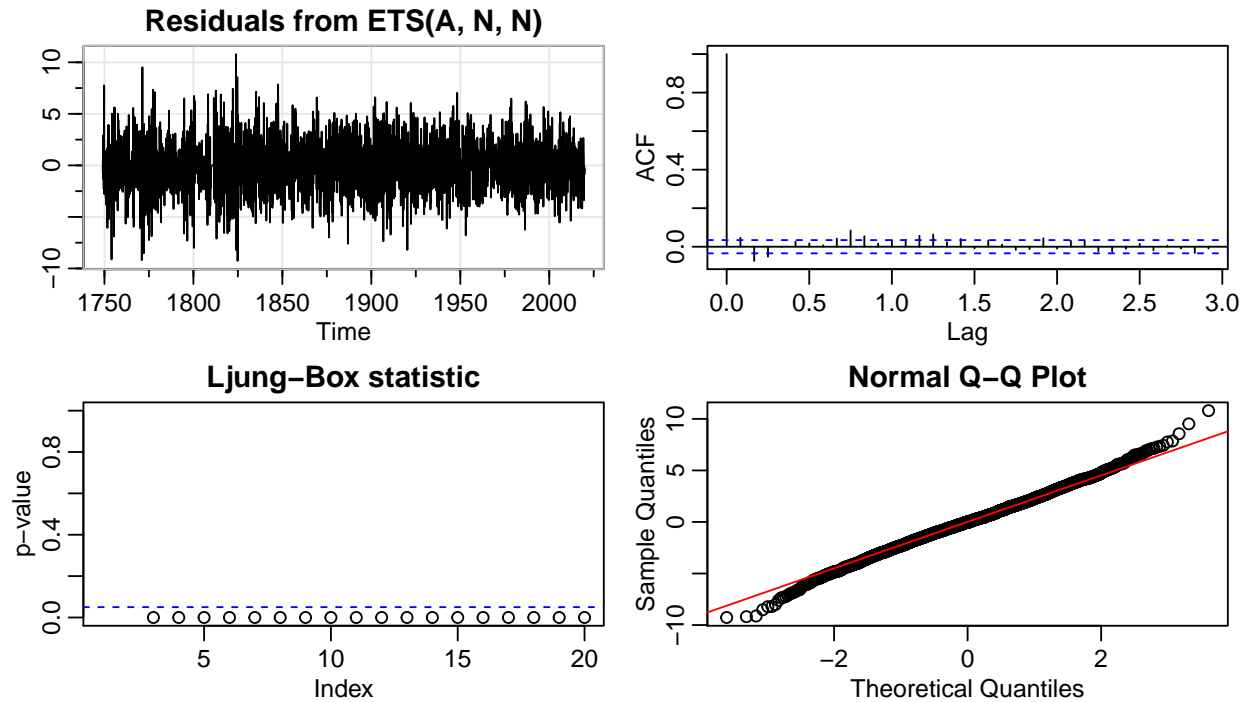
## Simple Exponential Smoothing (SES)

The simplest exponential smoothing method is convenient to forecast data without trend or seasonal patterns, which might be the case of the dataset in question. The forecasts of SES are equivalent to an ARIMA(0,1,1) model. The *ets* function of package *forecast* is used with model parameters set to “ZNN”, in which the first letter denotes the error type, the second denotes the trend type, and the third denotes the season type - “N” for none, “A” for additive, “M” for multiplicative, and “Z” for automatically selected. Model selection is based on the information criterion, while parameters and initial values are estimated by minimizing the sum of squared errors (Hyndman & Athanasopoulos, 2018 [6]).

The output below presents the fit parameter  $\alpha = 0.4972$ , implying that some weight is given to older observations in the past, given that the rate at which the weights decrease is controlled by the parameter  $\alpha$ , the smoothing parameter for the level. The Ljung-Box statistic plot suggest correlation among residuals.

```
## ETS(A,N,N)
##
## Call:
## ets(y = traindat, model = "ZNN", lambda = lambda)
##
## Box-Cox transformation: lambda= 0.4504
##
## Smoothing parameters:
##   alpha = 0.4972
##
## Initial states:
##   l = 15.7331
##
## sigma: 2.3842
##
##           AIC      AICc      BIC
## 31943.22 31943.23 31961.48
```





## Holt's Linear Trend method

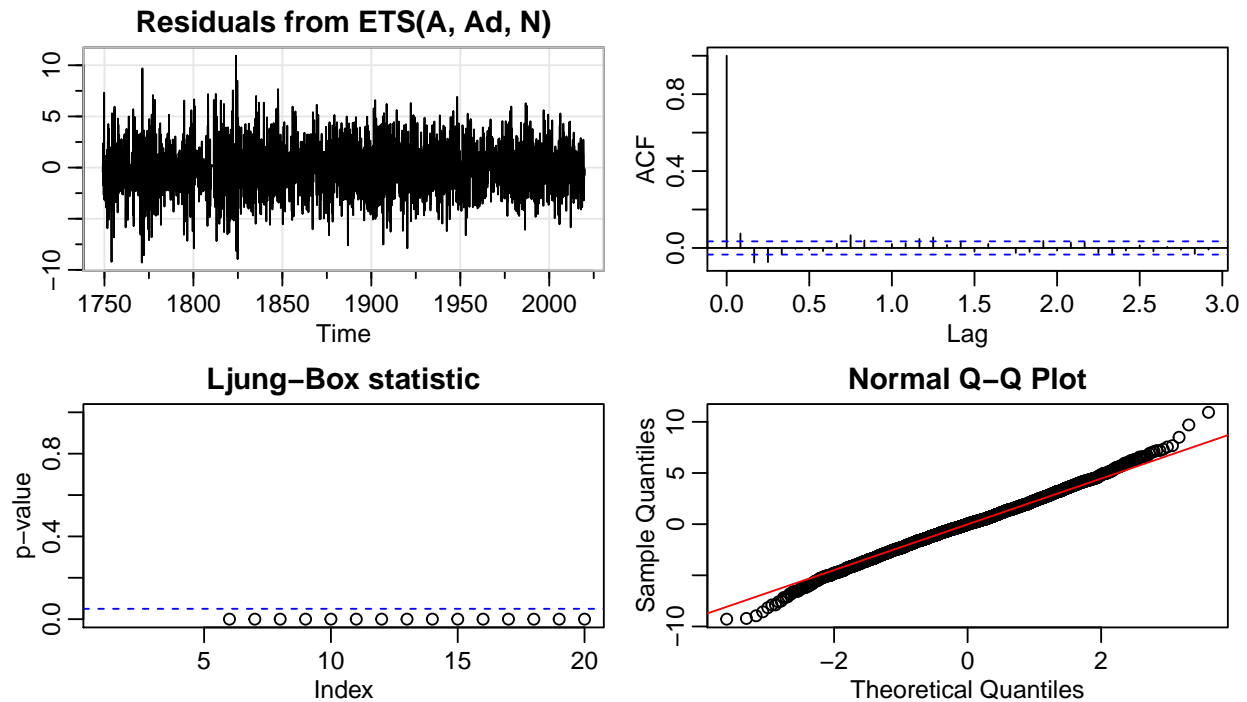
This extension of SES allows forecasting of data with trend. The forecasts of Holt's method are equivalent to an ARIMA(0,2,2) model. The *ets* function of package *forecast* is used with model parameters set to "ZAN" (Hyndman & Athanasopoulos, 2018 [6]).

The output below presents the fit smoothing parameter for the level  $\alpha = 0.4426$ , implying that some weight is given to older observations in the past, and the smoothing parameter for the trend  $\beta = 0.0159$ , which suggests the slope changes very little over time.

Note that the automatic selection determined a damped trend component, symbolized by the subscript  $d$  in ETS(A,Ad,N). The forecasts of Damped Holt's method are equivalent to an ARIMA(1,1,2) model. Sometimes damping is required because ETS methods tend to over-forecast. The estimated damping parameter  $\phi = 0.9412$  suggests the model is quite close to non-damped Holt's linear method. The Ljung-Box statistic plot suggest correlation among residuals.

```
## ETS(A,Ad,N)
##
## Call:
## ets(y = traindat, model = "ZAN", lambda = lambda)
##
## Box-Cox transformation: lambda= 0.4504
##
## Smoothing parameters:
##   alpha = 0.4426
##   beta  = 0.0159
##   phi   = 0.9412
##
## Initial states:
##   l = 15.2901
##   b = 0.34
```

```
##
##  sigma:  2.3789
##
##      AIC      AICc      BIC
## 31931.80 31931.83 31968.32
```



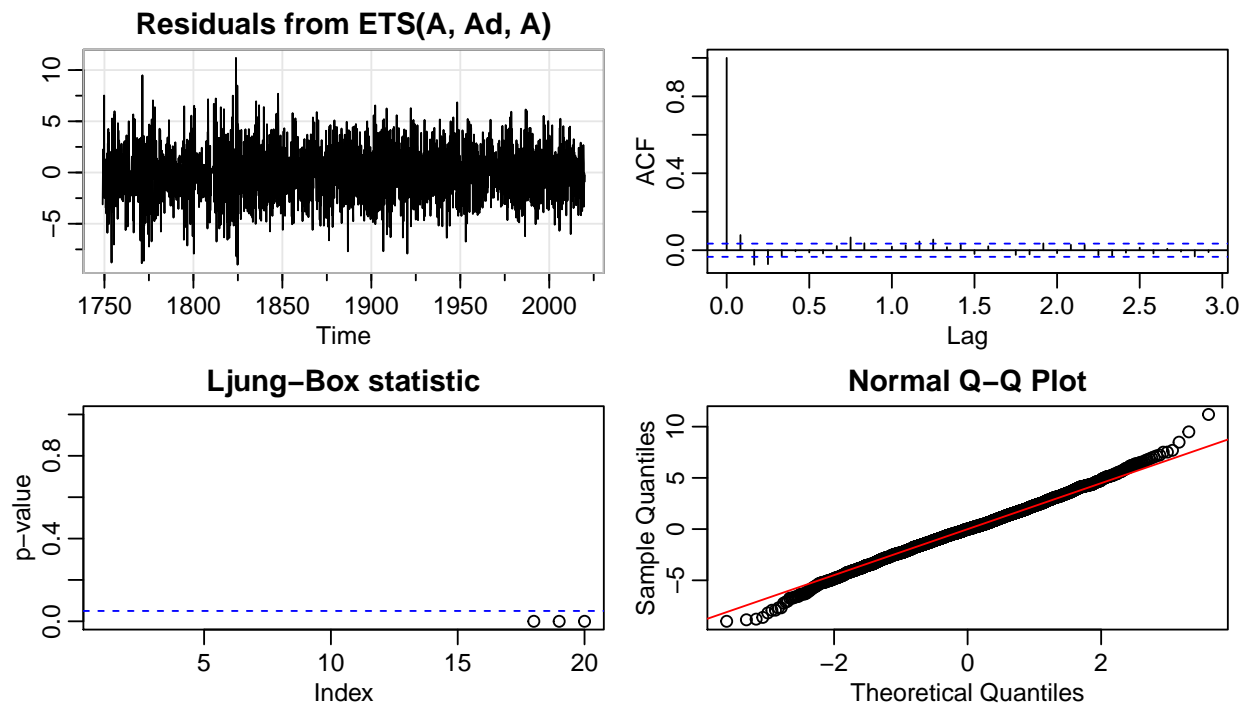
## Additive Holt-Winters' seasonal method

This extension of Holt's method allows for the forecast of time series with seasonal components. We've seen in previous sections that the dataset present no evident seasonality, thus this model might not be the most convenient. The forecasts of Holt-Winter's method are equivalent to an  $ARIMA(0, 1, m + 1)(0, 1, 0)_m$  model, where  $m$  is the frequency of the seasonality. The `ets` function of package `forecast` is used with model parameters set to "ZAA" (Hyndman & Athanasopoulos, 2018 [6]).

The output below presents the fit smoothing parameter for the level  $\alpha = 0.438$ , implying that some weight is given to older observations in the past, and the smoothing parameter for the trend  $\beta = 0.0163$ , which suggests the slope changes very little over time. Again, a damped trend component is selected, and the estimated damping parameter  $\phi = 0.9445$  suggests the model is quite close to non-damped one. The very small estimated smoothing parameter for the seasonality  $\gamma = 1e - 04$  means the seasonal component barely changes over time, thus the component might be neglected. The Ljung-Box statistic plot suggest correlation among residuals.

```
## ETS(A,Ad,A)
##
## Call:
## ets(y = traindat, model = "ZAA", lambda = lambda)
##
## Box-Cox transformation: lambda= 0.4504
##
```

```
## Smoothing parameters:
##   alpha = 0.438
##   beta  = 0.0163
##   gamma = 1e-04
##   phi   = 0.9445
##
## Initial states:
##   l = 18.5386
##   b = 0.1736
##   s = -0.1086 0.1776 0.1758 0.0281 0.0126 -0.0036
##         0.1592 0.1251 -0.0541 -0.0692 -0.0317 -0.4115
##
## sigma: 2.3815
##
##      AIC      AICc      BIC
## 31950.79 31951.00 32060.35
```



## ES models selection

All ES models showed correlated residuals that, nevertheless, adhere to Gaussian distribution quantiles. That said, the damped Additive Holt-Winters' seasonal method presented a very small  $\gamma$  estimate, thus it'll be discarded. The table below exhibits the remaining models' orders, number of parameters, as well as AIC, AICc, and BIC metrics. Both models present similar information criterion metrics, being the AIC and AICc of  $ETS(A, N, N)$  higher, and its BIC lower than the  $ETS(A, Ad, N)$  model. These models were fit using the package *forecast* to keep consistency among calculations. Both models are used in the next section, along with the chosen ARIMA models, to generate forecasts that can be evaluated regarding the test data set.

Table 8: Chosen ETS models

Model	Order	N.par	AIC	AICc	BIC
es1.ses	$ETS(A, N, N)$	2	31943.22	31943.23	31961.48
es2.holt	$ETS(A, Ad, N)$	5	31931.8	31931.83	31968.32

## Forecasting and evaluation

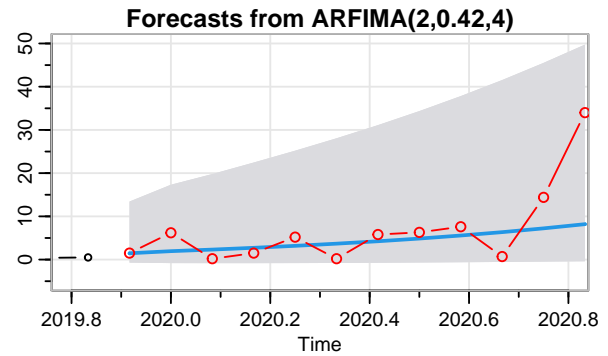
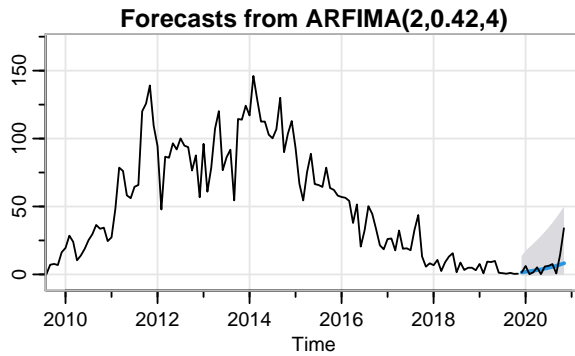
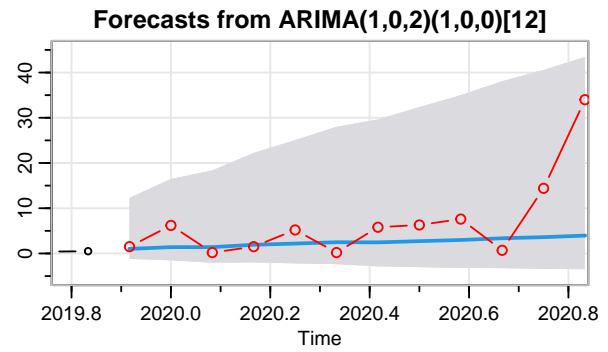
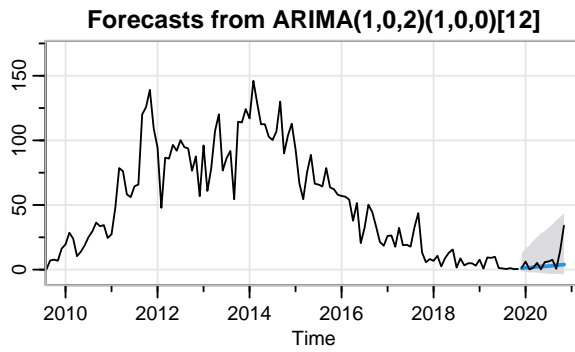
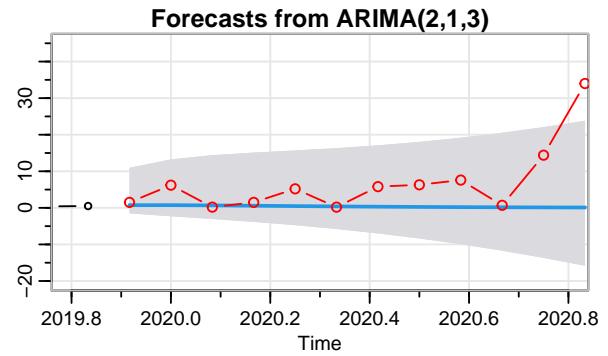
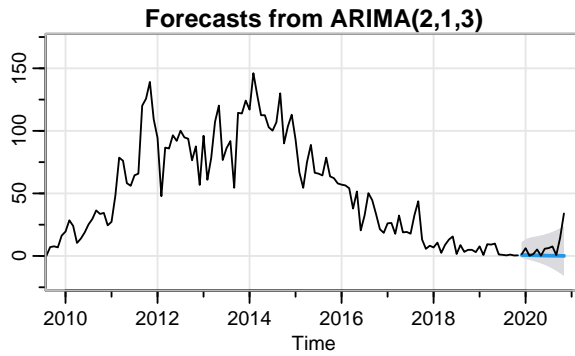
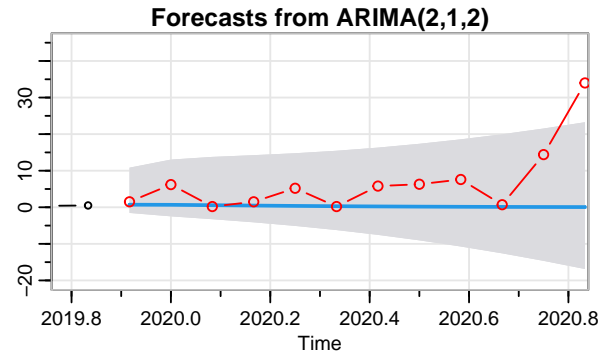
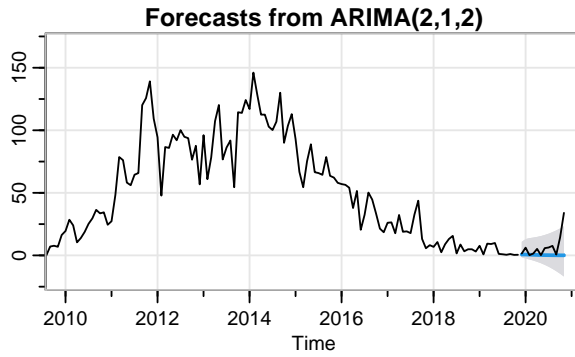
In this section twelve months forecasts are obtained from the chosen models are evaluated against test data through the scale dependent metrics Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), the scale-free error metric Mean Absolute Scaled Error (MASE), and Theil's U statistic. The twelve months forecasts as well as the test data are relative to the months of 12/2019 to 11/2020.

### Forecasting with ARIMA models

The plots below present the twelve months forecasts from chosen models of the ARIMA family as well as their 95% prediction intervals. For each model, the time series on the left presents observed data from 01/2010 to 11/2020, which represents the last observed 11 years sunspot cycle, while the time series on the right presents data from 11/2019 to 11/2020 (represented as 2019.8 and 2020.8 in the x axis). Both show forecasts as a blue line, and 95% prediction intervals as a shaded area in gray.

It seems models  $ARIMA(2, 1, 2)$  and  $ARIMA(2, 1, 3)$  forecasts underestimate the monthly mean total sunspot number, and fail to include the last observation in their respective prediction intervals. Their forecasts are very similar to the last case of the train set, relative to 11/2019.

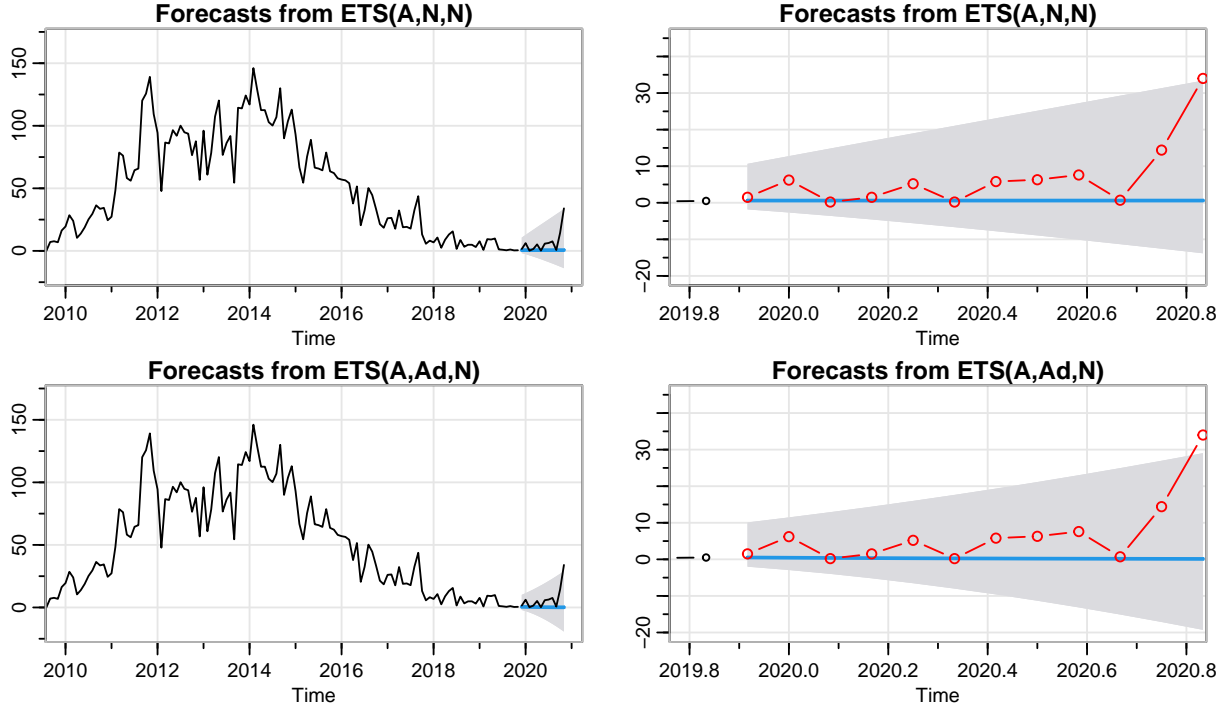
Models  $ARIMA(1, 0, 2)(1, 0, 0)_{12}$  and  $ARFIMA(2, 0.42, 4)$  appear to better represent the long cycles that are a characteristic of this time series, returning slowly increasing forecasts, and including all observations in their respective prediction intervals. Indeed, the fractionally differenced ARFIMA model may exhibit the best representation of the series dynamics thus far.



## Forecasting with ES models

Similar plots are shown below for the ES models. It occurs that both  $ETS(A, N, N)$  (SES) and  $ETS(A, Ad, N)$  (damped Holt's Linear Trend) models also underestimate the observed data and do not

enclose the last case in their prediction intervals. As with the  $ARIMA(p, 1, q)$  models, their forecasts are very similar to the last case of the train set.



## Evaluation

The following output displays each model's forecast scores in ascending order. As suspected, forecasts from  $ARFIMA(2, 0.42, 4)$  returned the best metrics, thus it is the best model to explain the series dynamics, *i.e.* the eleven year cycle. Note that the Ljung-Box test for the model did suggest correlation among residuals, but in practice the best model found is used even if it do not pass all tests (Hyndman & Athanasopoulos, 2018 [6]).

The  $SARIMA(1, 0, 2)(1, 0, 0)_{12}$  model performed better than the remaining ES and ARIMA models, which presented overall similar scores and could not take the cycles into account.

Table 9: Models' 12 month forecast scores against test data

Model	RMSE	MAE	MASE	Theil.s.U
$ARFIMA(2, 0.42, 4)$	8.16	4.73	0.13	0.42
$SARIMA(1, 0, 2)(1, 0, 0)_{12}$	9.62	5.61	0.15	0.66
$ETS(A, N, N)$	11.06	6.5	0.17	0.95
$ARIMA(2, 1, 3)$	11.29	6.68	0.18	1
$ARIMA(2, 1, 2)$	11.32	6.72	0.18	1.01
$ETS(A, Ad, N)$	11.31	6.74	0.18	1.01

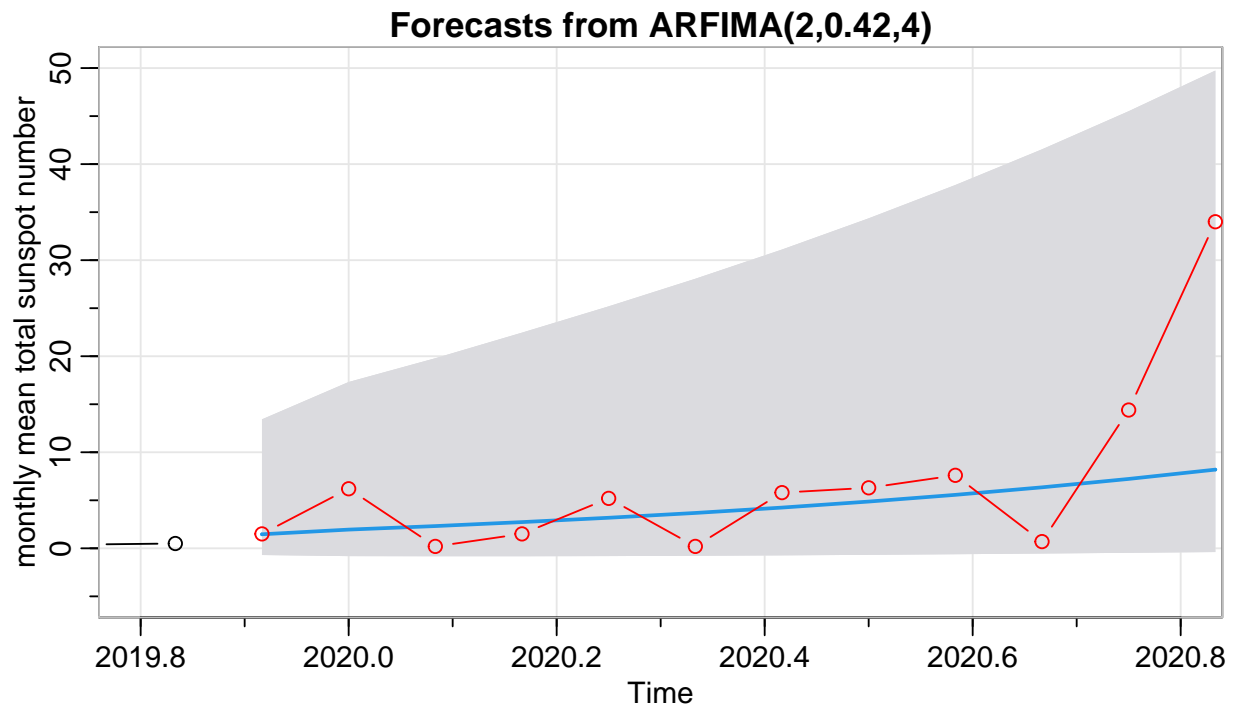
## ARFIMA model forecasts

In this final section, we take a closer look at the forecasts from the chosen  $ARFIMA(2, 0.42, 4)$  model. The table below presents the forecasts 95% prediction intervals ( $Lo.95$  and  $Hi.95$ ), a factor column (*contains.case*) that denotes if the observed value is within the interval, the observed value, the point forecast, and the absolute error. Note that percentage errors are not adequate considering that in some months the sunspot number is equal to or close to zero.

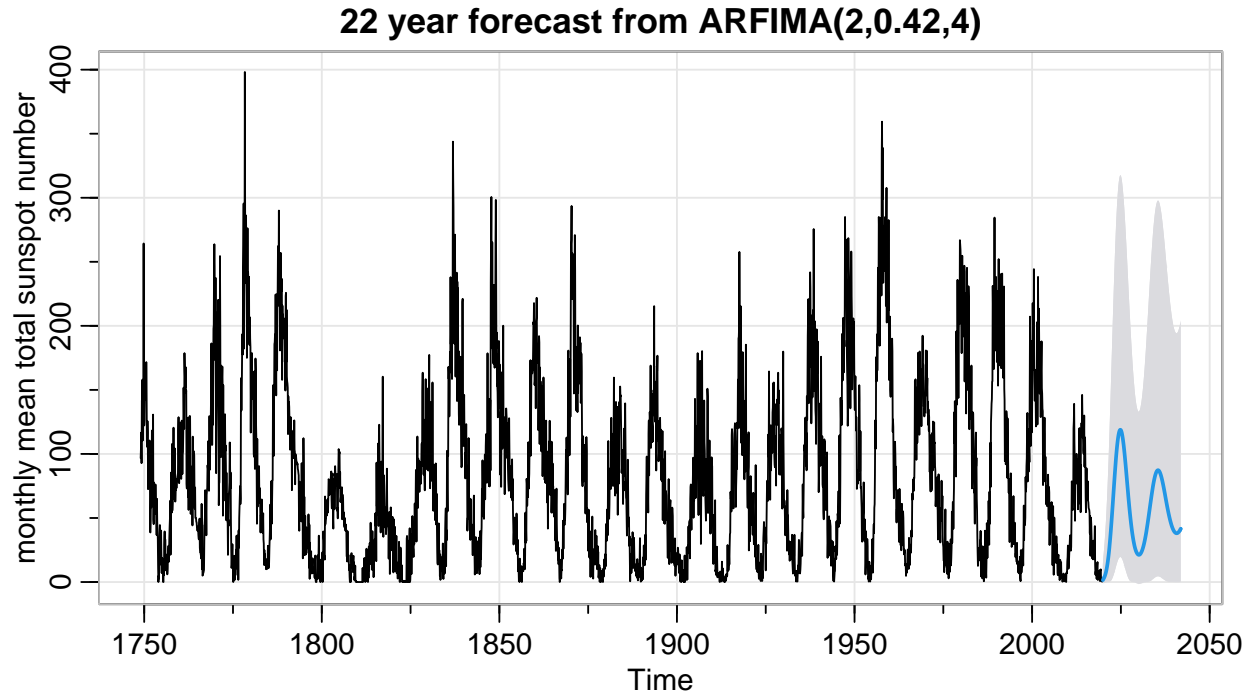
Table 10: ARFIMA(2,0.42,4) forecast analysis

Date	Lo.95	Hi.95	contains.case	Observed.value	Point.Forecast	error
12-2019	-0.6948	13.4278	1	1.5	1.46	0.04
01-2020	-0.8124	17.3148	1	6.2	1.95	4.25
02-2020	-0.8252	19.7911	1	0.2	2.32	2.12
03-2020	-0.8241	22.4528	1	1.5	2.73	1.23
04-2020	-0.807	25.1918	1	5.2	3.18	2.02
05-2020	-0.7756	28.0624	1	0.2	3.68	3.48
06-2020	-0.7317	31.1063	1	5.8	4.24	1.56
07-2020	-0.6767	34.3533	1	6.3	4.87	1.43
08-2020	-0.6126	37.8255	1	7.6	5.57	2.03
09-2020	-0.5413	41.5395	1	0.7	6.35	5.65
10-2020	-0.4653	45.5078	1	14.4	7.22	7.18
11-2020	-0.387	49.7397	1	34	8.19	25.81

All the cases are within the 95% prediction intervals. The absolute error for the predicted monthly mean sunspot number seems constant until September, when it increases suddenly. This might be explained by the new eleven year cycle maxima, which appears to start being formed by the end of 2020. The graphic below exhibits the forecast as a blue line along with the prediction interval as a shaded gray area.



Finally, the plot below represents a forecast for twenty two years after 11/2019. It demonstrates how the ARFIMA model is able to reproduce some of the long cyclic behavior observed in the original time series, effectively delineating the two eleven year cycles that are due to the period considered. Of course, forecasts tend to be less reliable the farther from the last observation, given their prediction intervals increase as the forecast horizon increases (Hyndman & Athanasopoulos, 2018 [6]).



## Conclusion

The sunspot number is an index used in several fields of study and applied sciences. The Sun's eleven year cyclical behavior is well documented, and has direct influence in the number of observed sunspots, in turn related to solar flares and to geomagnetic storms, which can affect our daily lives. In this project the mean monthly sunspot number time series was analyzed in order to select a convenient model to represent its dynamics and obtain reasonable forecasts.

The exploratory data analysis proved essential in guiding data processing, namely the Box-Cox transformation to stabilize variance, as well as identifying the time series major characteristics, particularly the long-memory behavior, in which correlation between lags extend far in time. Unit root tests Augmented Dickey-Fuller and Kwiatkowski-Phillips-Schmidt-Shin were used to determine the necessary order of differencing, and guided the manual selection of ARIMA family models. Automatic model selection functions were also used to determine a *SARIMA* and an *ARFIMA* model, the first acted as a benchmark to the manual selections, and the second was motivated by the quoted long-memory behavior observed.

According to the Ljung-Box test the  $ARIMA(p, 1, q)$  models chosen had some level of correlation among its residuals, while the *SARIMA* and *ARFIMA* models had significant p-values for all lags tested. The information criterion (AIC, AICc, and BIC) could only be used to determine a better model among those with equal difference order, specifically the  $ARIMA(2, 1, 2)$  and  $ARIMA(2, 1, 3)$  models, which showed very similar values.

Experiments were also carried with Exponential Smoothing models. More specifically, models based on the methods Simple Exponential Smoothing( $ETS(A, N, N)$ ) and damped Holt's Linear Trend ( $ETS(A, Ad, N)$ )



showed relevant smoothing parameter estimates, while the model based on Additive Holt-Winters' seasonal method displayed a very low estimate for the seasonal component, thus was discarded. The ES models also shown significant p-values regarding the Ljung-Box test. The information criterion of both models were very similar.

A twelve month forecast was obtained from all six models, *i.e.*  $ARIMA(2, 1, 2)$ ,  $ARIMA(2, 1, 3)$ ,  $SARIMA(1, 0, 2)(1, 0, 0)_{12}$ ,  $ARFIMA(2, 0.42, 4)$ ,  $ETS(A, N, N)$ , and  $ETS(A, Ad, N)$ . The forecasts were evaluated against the test data set, comprised of the mean monthly sunspot number from 12/2019 to 11/2020. The  $ARFIMA$  model performed better than the remaining, followed by the  $SARIMA$  model. ES models and the  $ARIMA$  models with order  $d = 1$  had similar metrics, and couldn't account for the eleven year cycle observed in the series, as their predictions were very similar to the last observations of the train data set.

Finally, a closer look was taken on forecasts obtained from the best model,  $ARFIMA(2, 0.42, 4)$ . The observed values were within the prediction intervals, and errors seemed to increase by the end of the forecast period, when a new solar maxima begins to form. That said, the fractionally differenced model is capable of reproducing the eleven year cycle to an extent, as seen in the last plot relative to a 22 years forecast.

It is interesting to note that although the  $ARFIMA$  model did not pass the Ljung-Box test, its forecasts displayed the best metrics among models. Therefore, it may be possible that yet another model could better explain the time series dynamics, and that the prediction intervals are incorrect. That said, in reality the best model might not pass all statistical tests, which do not make the model less useful or the tests less important (Hyndman & Athanasopoulos, 2018 [6]).

We conclude by remarking on the similar performance of  $ARIMA(p, 1, q)$  and  $ES$  models, even though they were not the most convenient choice considering the problem at hand. As described by Hyndman and Athanasopoulos (2018), there are  $ARIMA$  models that do not have an  $ES$  counterpart, and the reciprocal is true, thus surely each family of models can be successfully applied to different problems.

Future extensions of this project might consider improving on the best model found thus far, specifically regarding the correlation among residuals noted in the Ljung-Box test, and exploring different approaches for forecasting the sunspot number.

---

## REFERENCES

- [1] Weather Forecast Office, National Weather Service n.d., *The Sun and Sunspots, Sioux Falls, SD, United States of America*, viewed 23 December 2020, <https://www.weather.gov/fsd/sunspots>.
- [2] Berghmans, D., Van der Linden, R.A.M., Vanlommel, P., Clette, F., Robbrecht, E. n.d., *History of the Sunspot Index: 25 years SIDC*, SIDC, Royal Observatory of Belgium, Ringlaan -3- Av. Circulaire, B1180 Brussels, Belgium, viewed 23 December 2020, <http://www.sidc.be/silso/IMAGES/about/Berghmansetal2006.pdf>.
- [3] *Sunspot data from the World Data Center SILSO*, Royal Observatory of Belgium, Brussels, viewed 23 December 2020, <http://www.sidc.be/silso/datafiles>.
- [4] Shumway, R.H., & Stoffer, D.S., 2015, *Time Series Analysis and Its Applications With R Examples*, 4th edition, Springer, viewed 26 January 2021, <https://www.stat.pitt.edu/stoffer/tsa4/tsa4.pdf>.
- [5] Holmes, E.E., Scheuerell, M.D., Ward E.J., 2021, *Applied Time Series Analysis for Fisheries and Environmental Sciences*, NOAA Fisheries, Northwest Fisheries Science Center, 2725 Montlake Blvd E., Seattle, WA 98112, viewed 26 January 2021, <https://nwfsc-timeseries.github.io/atsa-labs/>.
- [6] Hyndman, R.J., & Athanasopoulos, G., 2018, *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2, viewed 26 January 2021, <https://otexts.com/fpp2/>.