STAT 221 Final Project - Mammoth Snow Depth

John Rapp Farnes | 405461225

3/17 2020

Contents

% TODO - Tillbaka trend på pred

- Tapering? Ge periodogram lite kärlek
- Figurtexter/labels/axes

1 Introduction

1.1 Background

Mammoth Mountain is situated in Northern California and is know for its great alpine ski and snowboarding conditions. For this purpose, the mountain features a ski resort with the same name. This resort has has more than 3,500 acres of ski-able terrain and is serviced by 28 lifts, recieving over 1 million annual visitors. For Southern Californian residents especially, the mountain is of interest as it is one of the closest "good" ski resorts, about a 4-6 hour drive away from Los Angeles. **??**

Most people familiar with alpine sports considers snow depth one of the most important conditions for the sport, as this affects which runs are open and how "good" the skiing is considered. As decisions about travelling to a ski resort generally are done in advance and require some planning (e.g. booking a cabin), being able to predicts future conditions are be useful to the alpine skiier. In addition, the entire ski economy of a mountain such as Mammoth, including the resort, workers as well as restaurants and stores in the sorrounding city, face uncertainty over how many visitors the mountain will get a given week or year as this drives revenue. As visitorships likely is correlated with snow depth, forecasting it would be an important tool also for those stake holders. This paper will attempt to do just that: **model the snow depth at Mammoth mountain in order to make predictions on future skiing conditions**.

1.2 Data set

Data on historic snow depth at Mammoth were obtained from the reporting of Mammoth Mountain Ski Area, through a third party website **??**. The website does not provide the data easily in a downloadable format, hence the data was obtained through injecting JavaScript into a browser client that extracted, extracting the data from the browser JavaScript enviroment and printing it in a JSON format. This raw data is shown in figure 1, featuring 1791 recordings from 2011-12-01 through 2020-03-02 of daily snow depth measured in inches. Upon looking at the graph, two issues with the raw data are found: 1. The dates in the off-season (outside of the winter months) are not included, rather the years are concatinated together in a single time-frame 2. some values within the recorded period are missing and reported as 0. Hence, the data needed to be further processed and cleaned before fitting any models on it.

1.3 Cleaning data

The first step in cleaning the data was to include the missing dates in order to capture the full time-frame of the series The next step was handling the missing values, both in the off-season as well as the missing recorded values. In order to handle the missing recorded values, as well as to make lower variance predictions on snow depth further in the future than a couple of days, the data was aggregated and averaged (disregarding the missing values) per week, resulting in a weekly time series. This week was defined as starting a Saturday, as this a day of interest for many weekend skieers. The off-season missing values were replaced by 0s, as this is an accurate description of the snow depth during those months – there rarely is any snow on Mammoth during the summer. The resulting data after cleaning is shown in figure 2, featuring 431 weeks. The availability of data per month is displayed in figure 3, showing that data exists for the most part Dec-May, with less than 50% Jun-Oct, again reaching ~80% in November.

2 Analyis

Looking at the grapt in figure 2, it is clear that the ski season of 2020 has a far greater snow depth than prior years, an unfortunate fact for skiiers this season. In order to study other properties than this obvious observation, time-series methods were applied.

2.1 Series properties

Figure 12 plots the ACF and PACF of the series, it where the ACF shows periodic behaviour with a 1 year period that appears to be slowly tailing off, while the PACF has 2 (barely) significant values and then cuts off, followed by a significant value at lag 1 year. This implying that a seasonal AR model may be a good description of the series. The 1 year period is easily seen also in the periodogram (figure 5), together with a 4 year period, both being significant peaks. The 4 year period may be an artifact of the data being recorded for 4 years and these years having a pattern of yearly depth by chance.

The data does not appear to be stationary as it has an obvious yearly trend and a clea pattern. As such, the data must first be detrended before ARIMA models can be applied.

2.2 Detrending

The most obvious trend in the data that may be removed is the 1 year seasonal trend. This trend can be seen in figure 6 which shows the average snow depth during the different weeks of the year (week 1 through week 53). This graph has a clear sinusoidal shape. As such, a first attempt at detrending the data was made by subracting the average snow depth of the week of year to every data point. The result of this may be seen in figure 7. Looking at the graph, it does however not appear to be stationary. One artifact of this detrending method is that the different years experience different levels of snow, causing years with less snow to have a clear negative valley and years with more snow to have a positive peak. Looking at the snow depth by week of the year for all years simoultaneaosly, seein in figure 8, makes it clear that the snow level is very different each year and that the average hence is not a good predictor.

One way to mitigate this fact is to study values in relation to the peak snow level of the year. The assumption one makes is that the dynamics that determine what the peak snow depth, or how high the snow peak will be a given year, is different from the dynamics that govern how its snowing and thawing during the year, dividing into a makro and micro model. With this approach, one model could be used to predict peak snow depth a given year, e.g. assuming that they are independent and follow a certain distribution. Under this assumption, predicting the snow depth for the rest of a season may be done by only looking at how the current and past values have been in relation to the peak snow depth during the year, which is the approach implemented in this paper. Figure 9 shows the snow depth relative to the peak value of the season, from this the average of the week of the year of every data point is subtracted in order to remove seasonality. This model assumes that the dynamics of when and how much it snows is similar from year to year. The resulting detrended graph is shown in figure 10 both with the off-seasons as 0s, and with missing values ommited, all ski seasons shows next to each other. This series finally looks fairly stationary, while still having some features that can modelled with time-series methods. As the series however appears to still have an upwards linear trend, the final detrended data is created by removing this linear trend, shown in figure 11.

As for the missing values in the data set, there are multiple ways to handle them the most obvious being either setting them to 0 or removing those dates from the data. As the lenght of the seasons differ from year to year, removing the off-seasons from the data makes seasonality measures such as the periodogram harder to interpret. As models were fit to both data sets showed similar performance, the one with missing values as 0s is considered going forward.

2.3 Detrended series properties

Figure **??** shows the ACF and PACF of the detrended series. The ACF is tailing of faster for this series, however still with a peak after one year, impying that the yearly feature has not been fully taken out. The PACF only has a single (barely) significant value then cuts off. This implies that a low order AR model may be appropriate, again potentially with a yearly seasonal component. The periodogram, seen in figure 13, has 5 significant peaks (in blue), with period ~2.1, ~0.69, ~0.44, ~0.28 and ~0.20 years. Figure 14 shows that the peaks has a clear pattern, the first having frequency ~0.48, and then being fairly evenly spaced. This means that the detrended series still has periodic components to it, however the frequencies are not easily interpretable as to what they describe. Figure 15 plots the parametric periodogram, showing the periodogram of the fitting the best fitting AR(p) model. For the detrended series, an AR(2) fit the best, having no clear peaks but a low frequency spectrum. These two periodograms both tell us is that the noise in the data is mostly low frequency, and therefore fairly smooth.

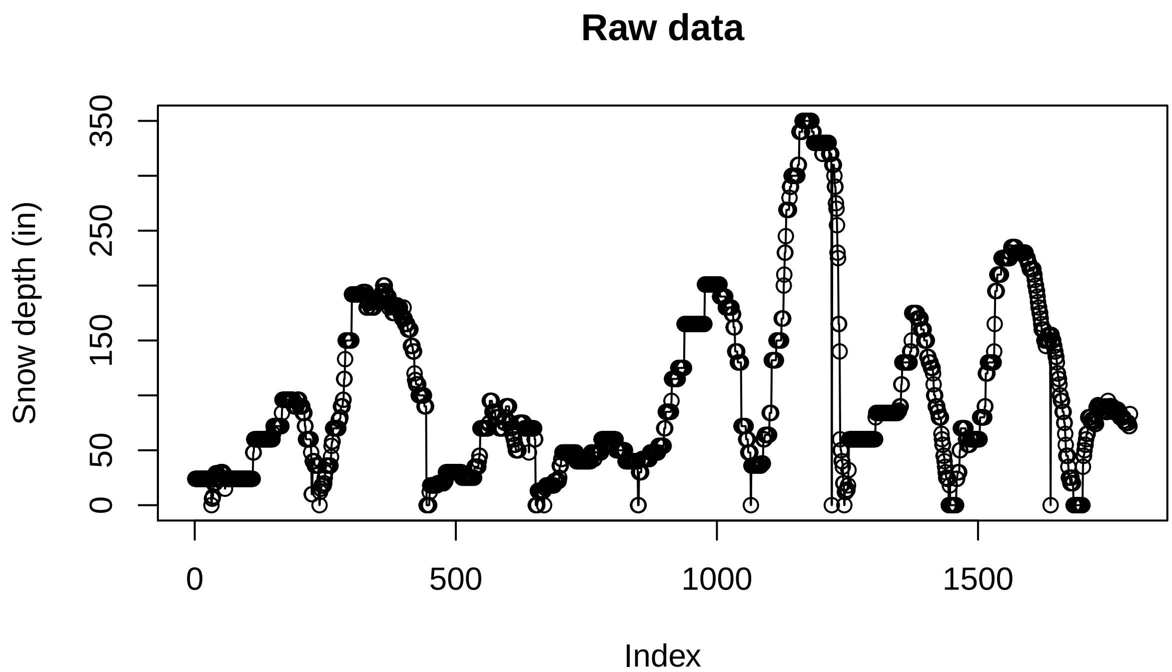


Figure 1: Raw Mammoth snow depth data

2.4 Model fitting and interpretation

In order to fit a (seasonal) ARIMA model to the detrended data, different models with different values of p, q and d as well as seasonal components were compared and evaluated based on their AIC-value. This process yielded two candidates ARIMA(2,0,1)(0,0,1)[52] ARIMA(2,0,1)

2.5 Model evaluation

2.5.1 Residual analysis

2.6 Frequency domain

3 Results

3.1 Conclusion

3.2 Dicsussion

3.3 Next steps

4 Appendix - Figures and graphs

```
## initial  value -2.019724
```

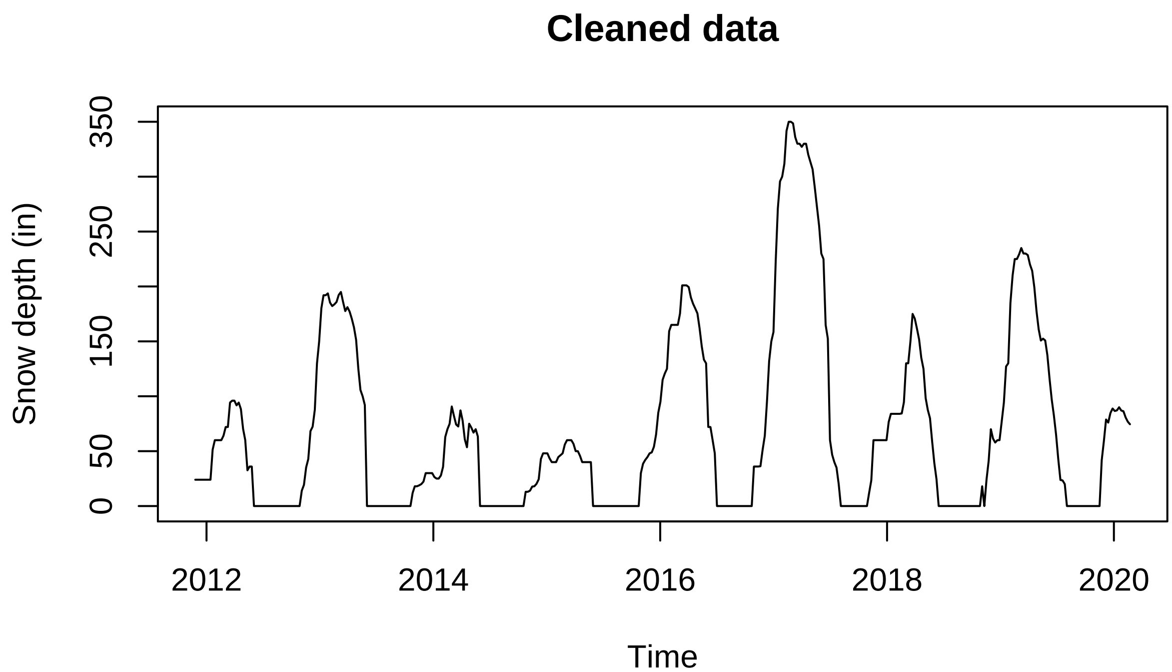


Figure 2: Mammoth after initial cleaning

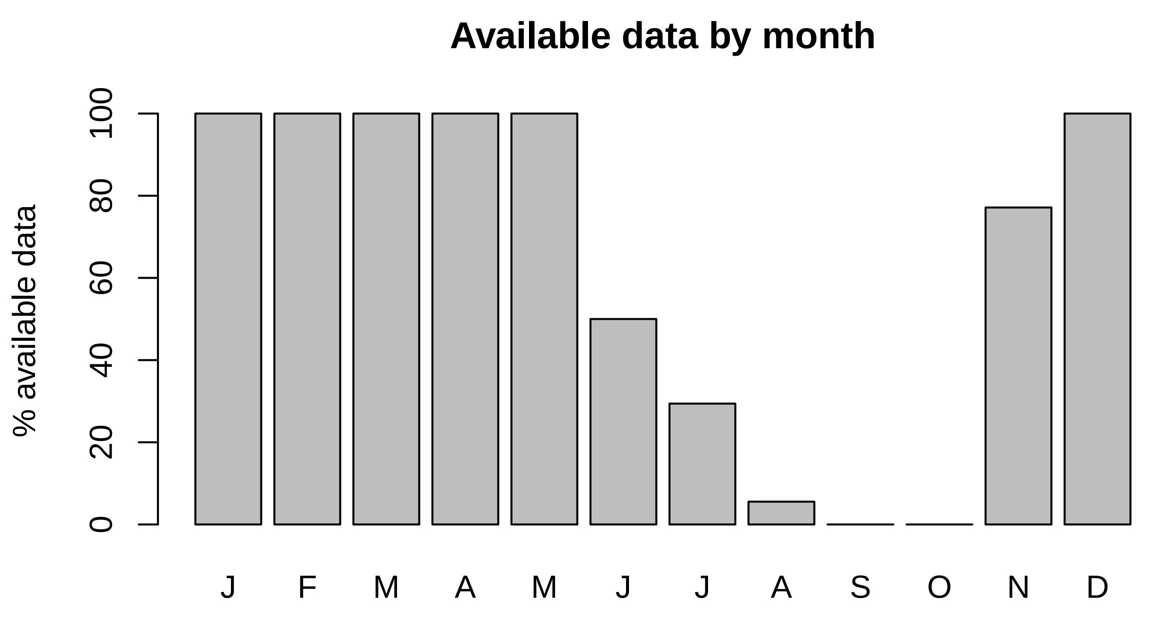


Figure 3: Percent of weeks where data was available, by month

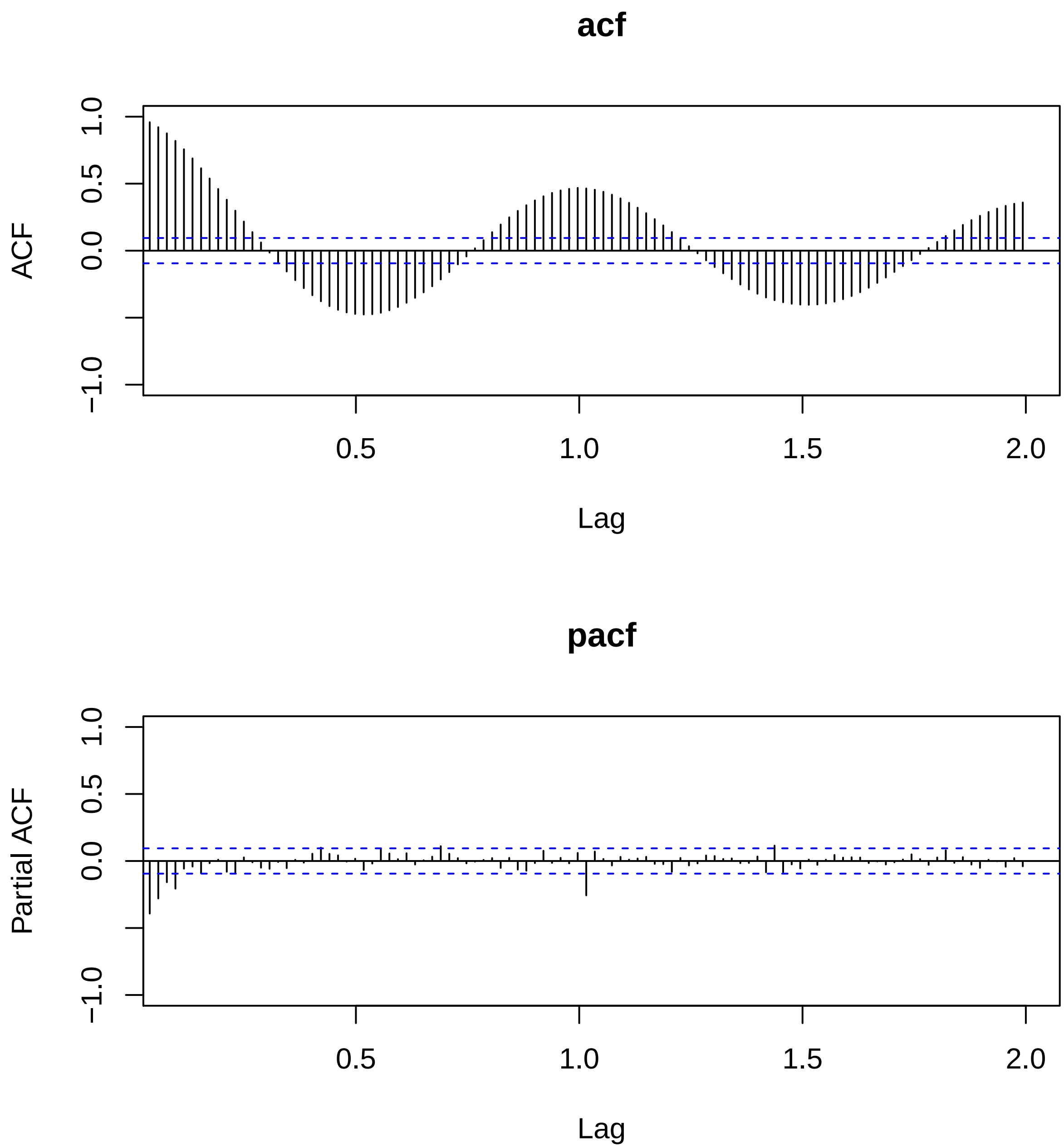
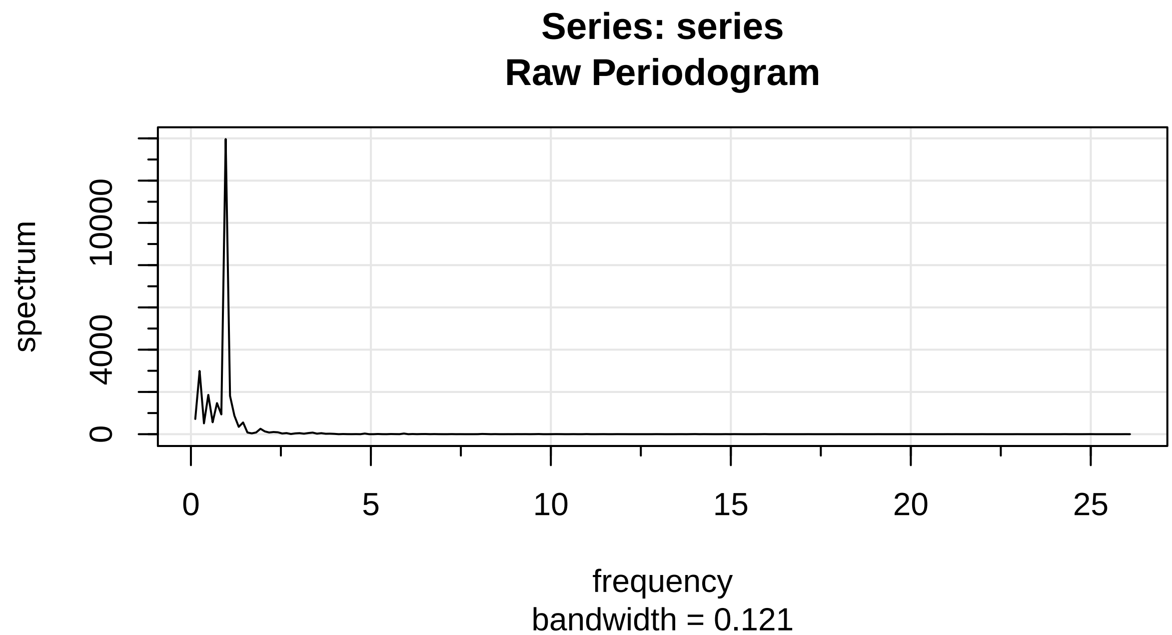


Figure 4: ACF and PACF for the cleaned time-series



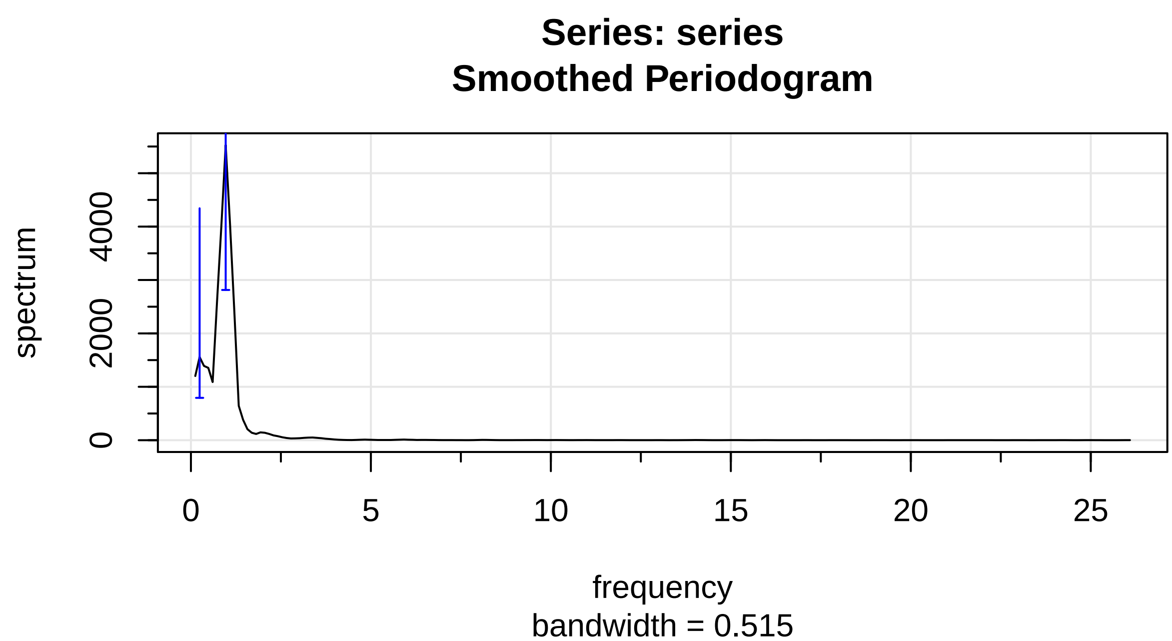


Figure 5: Periodogram for the cleaned data, peeks at period 1 and 4 years shown with a 95% significance interval

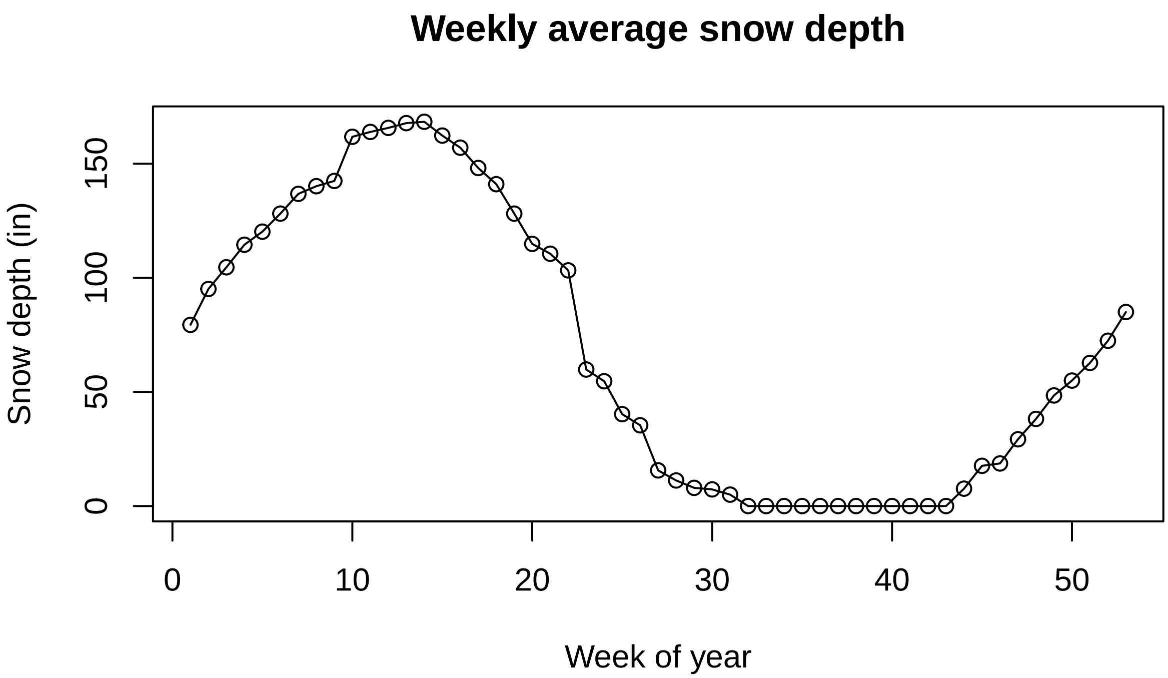


Figure 6: Average snow depth by week of the year

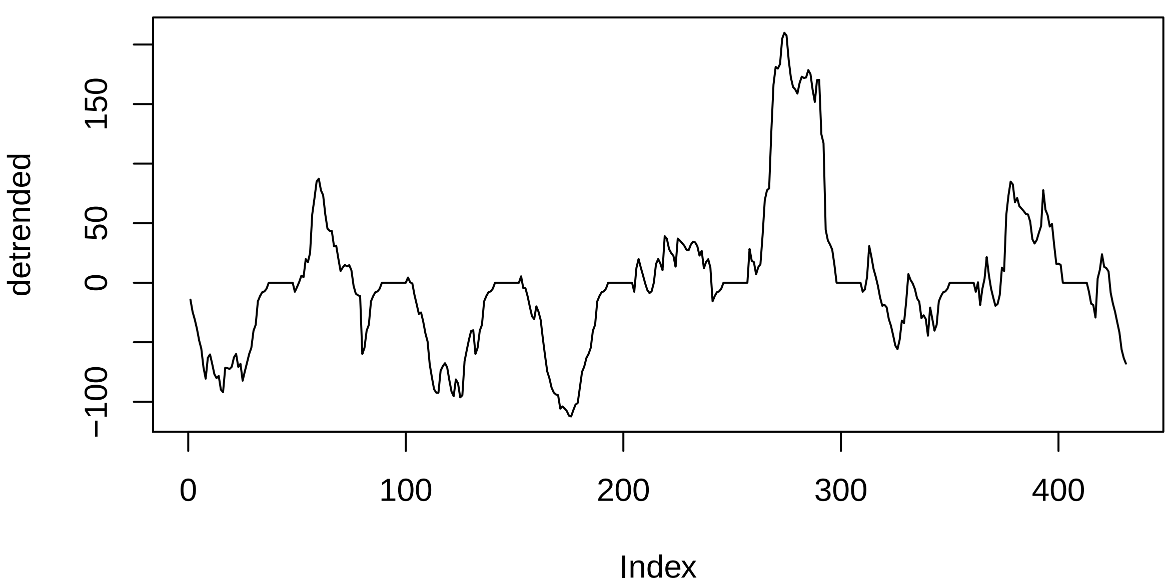


Figure 7: Snow depth detrended by subtracting average of week of the year

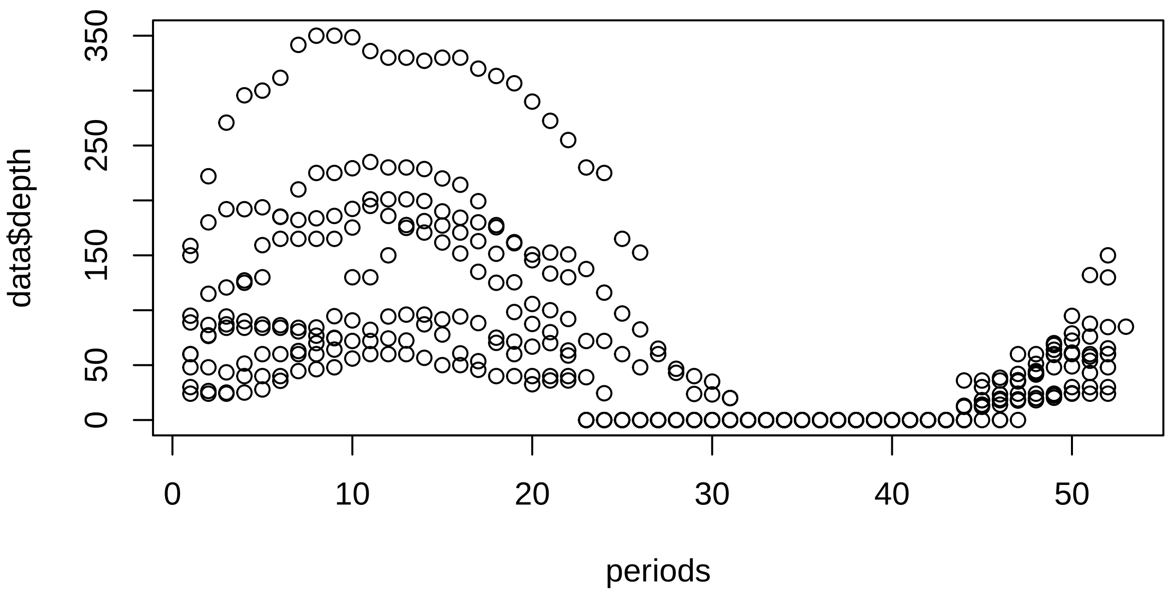


Figure 8: Weekly snow depth of all years

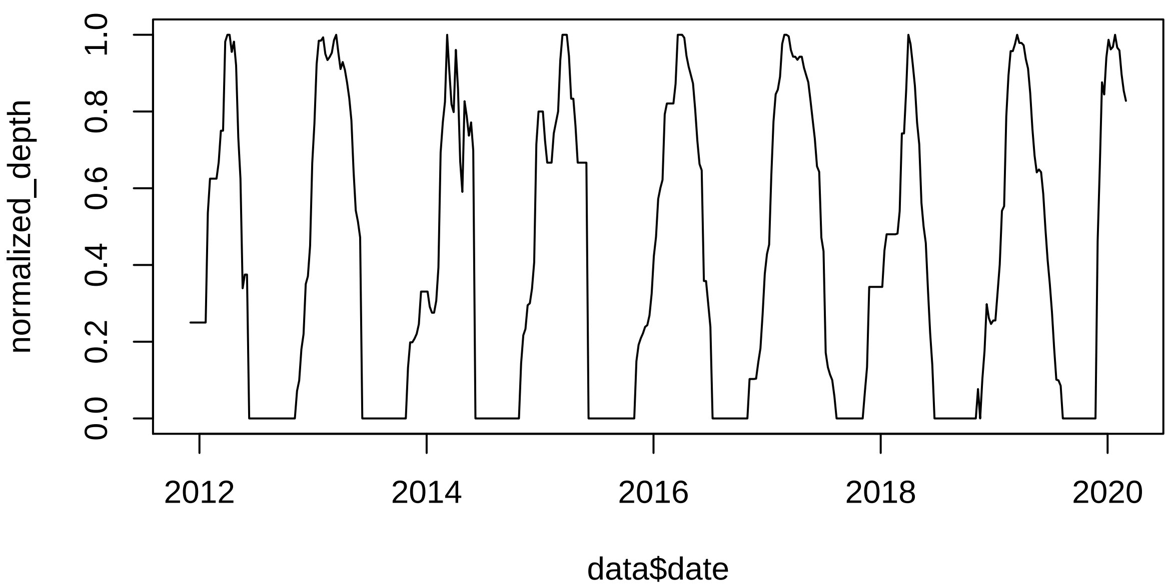


Figure 9: First step of detrending the data, taking the values relative to the peak of the season

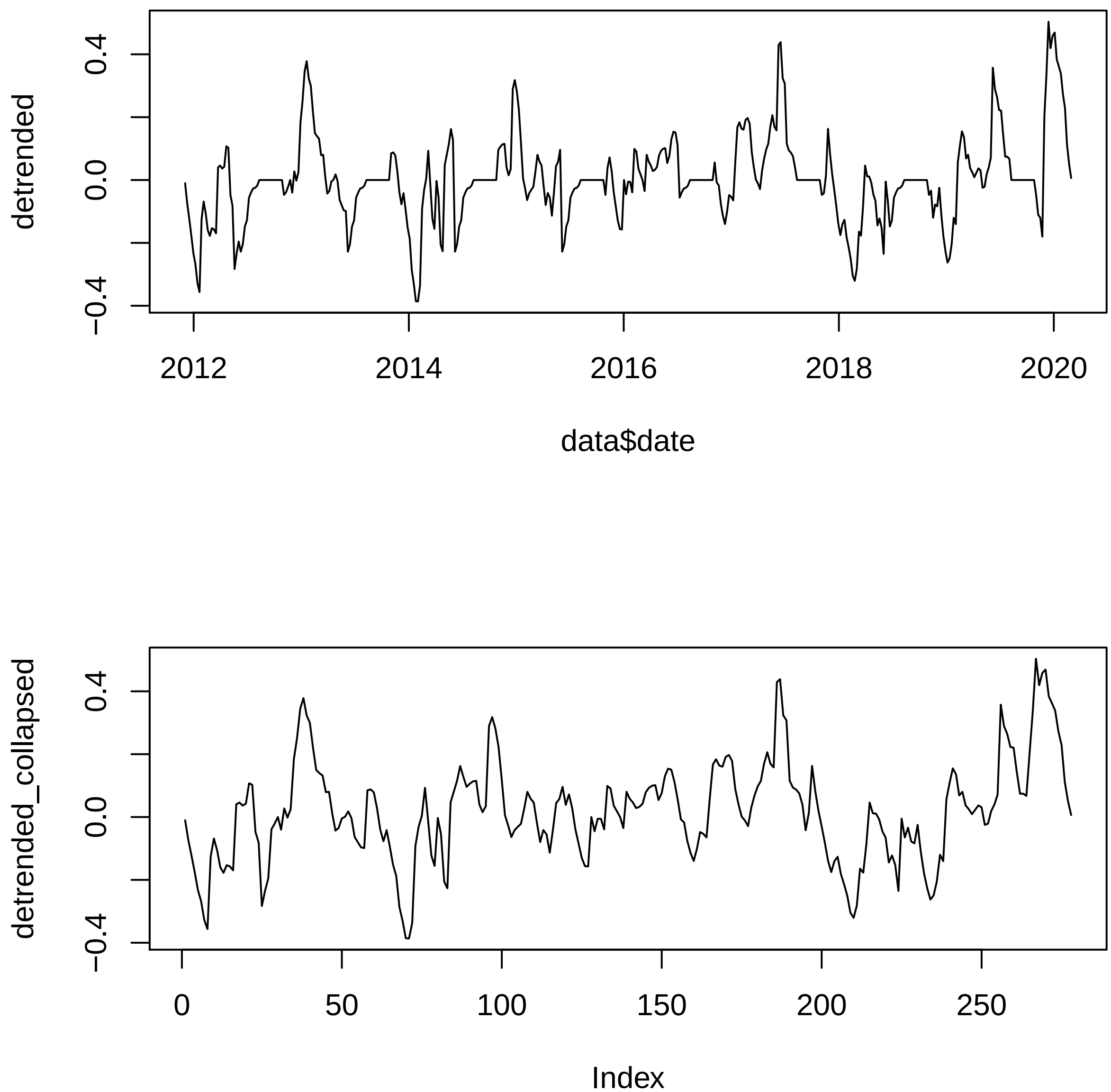


Figure 10: Detrended data after subracting the yearly trend

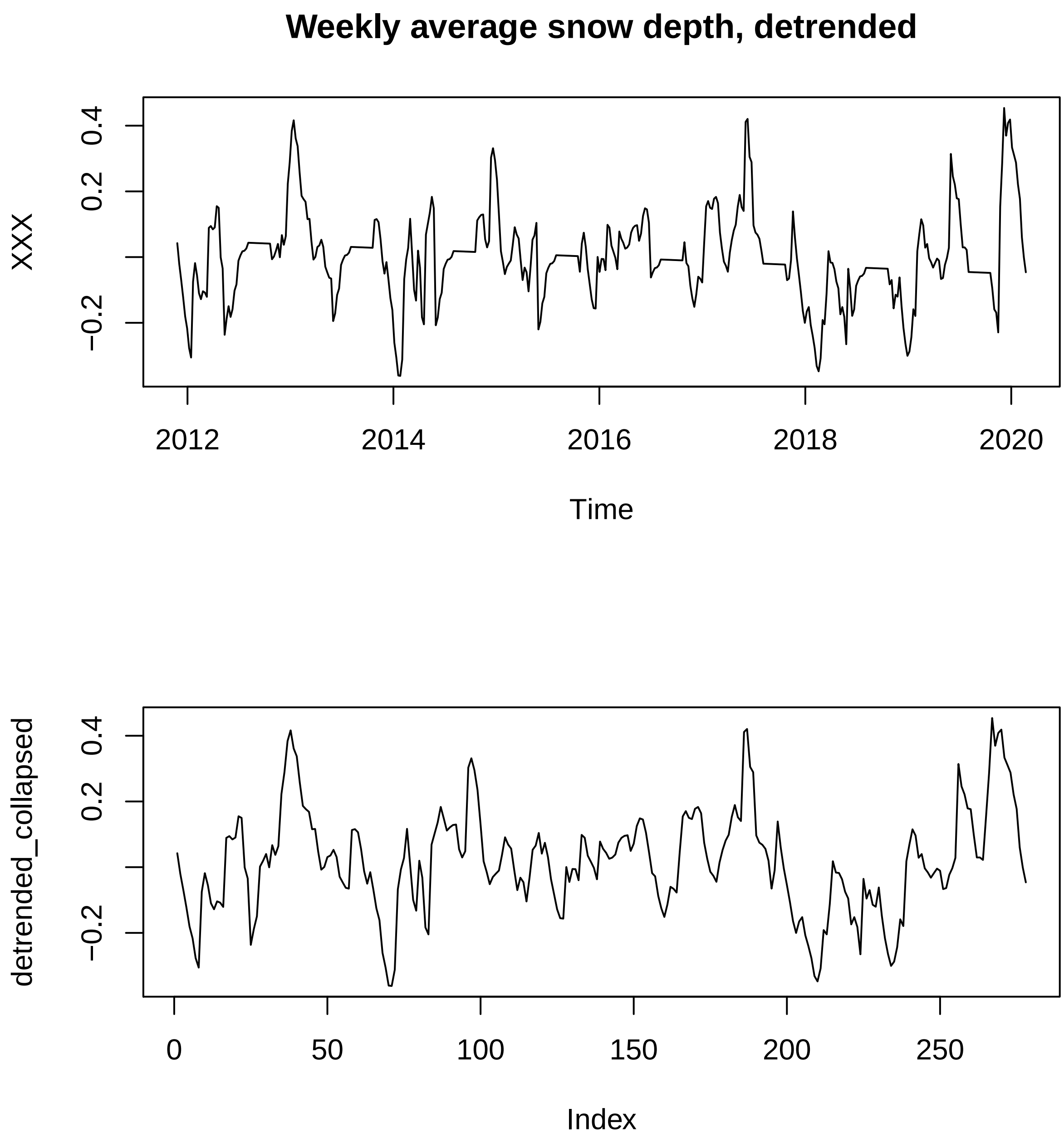


Figure 11: Final detrended data

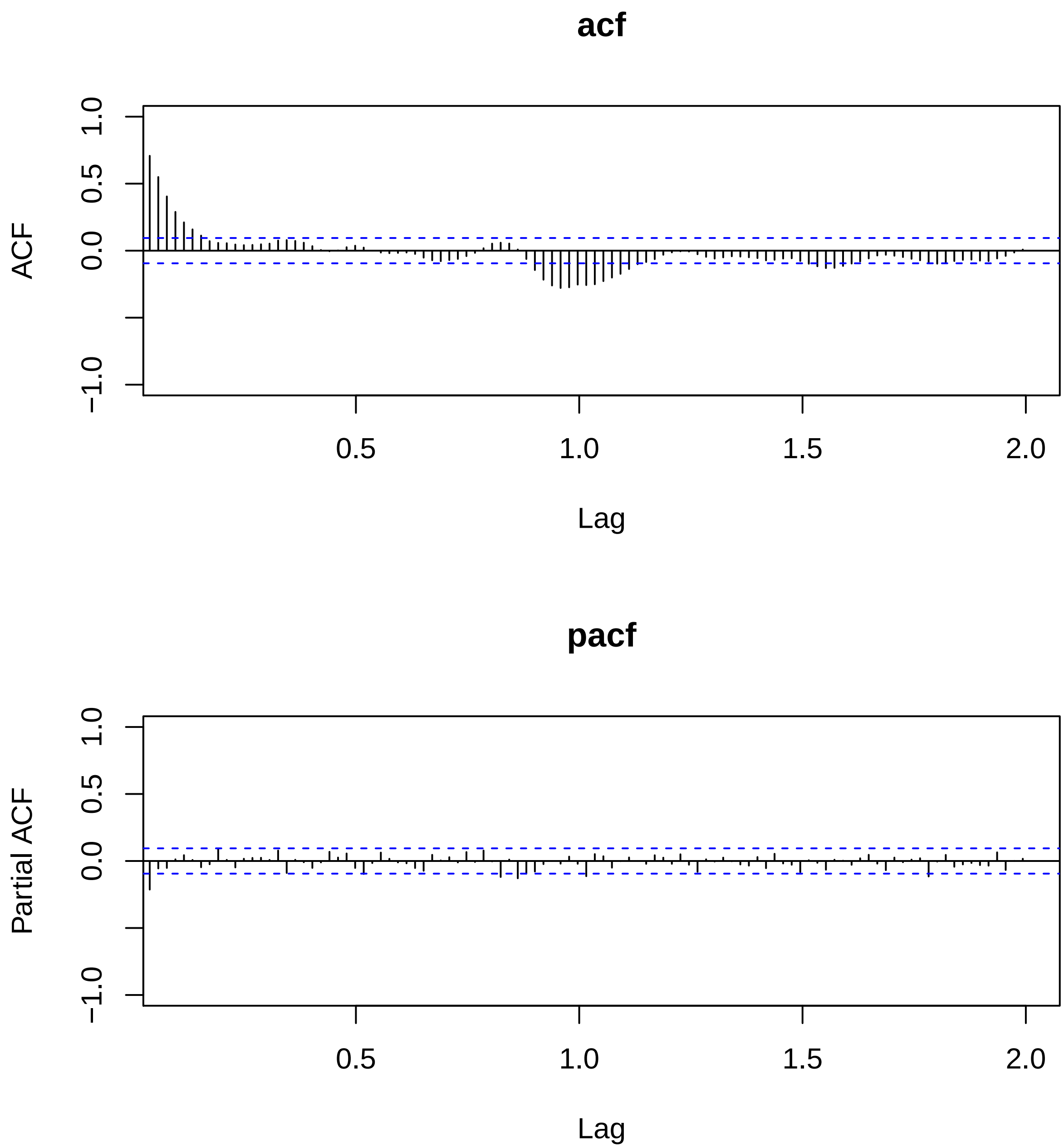


Figure 12: ACF and PACF for the detrended time-series

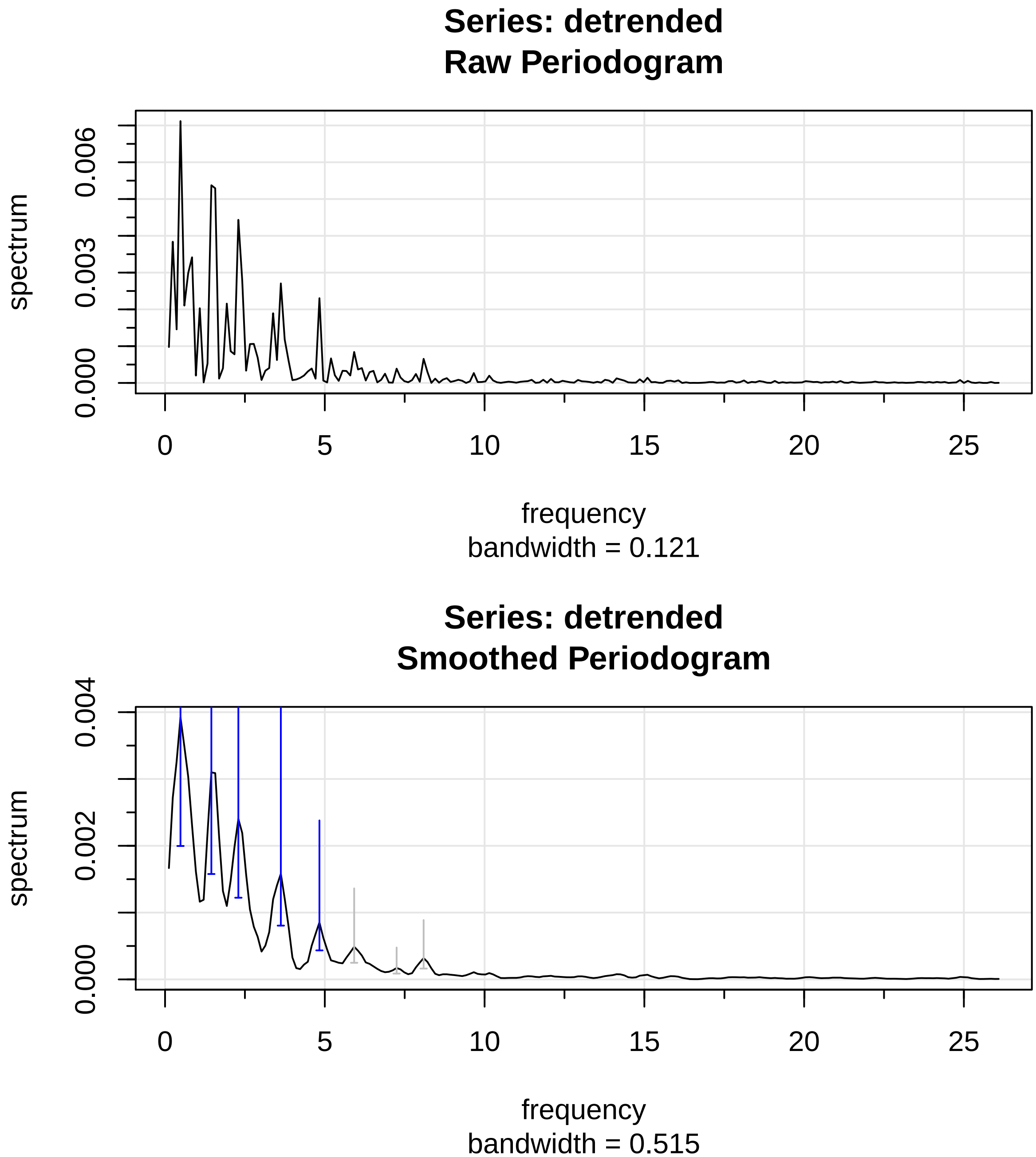


Figure 13: Periodogram for the detrended data, peeks shown with a 95% significance interval, significant peaks marked in blue

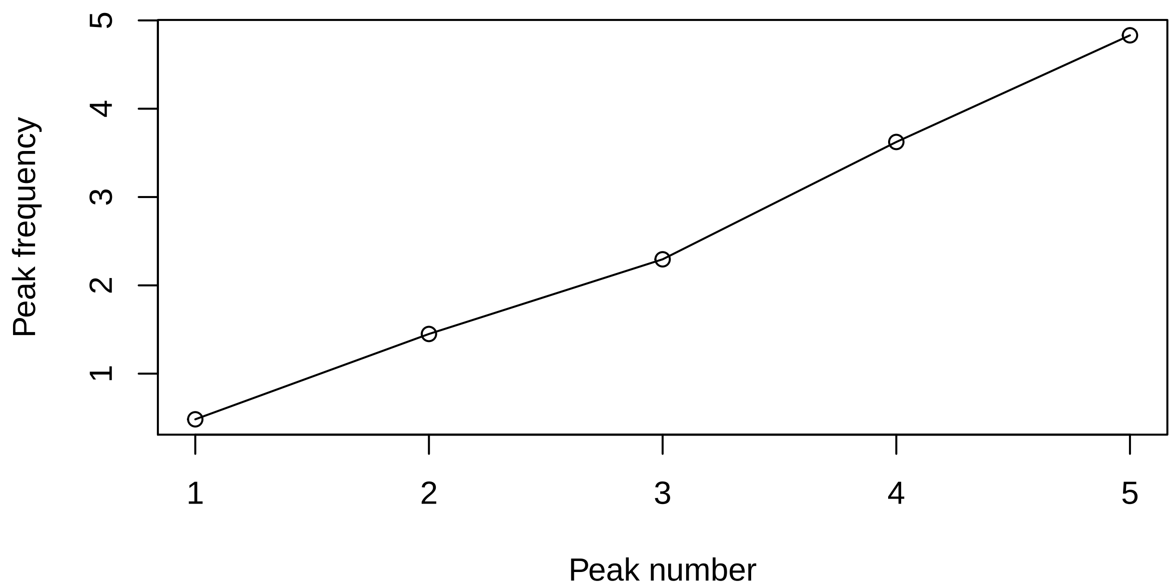


Figure 14: Frequencies of the significant peaks in the periodogram of the detrended data

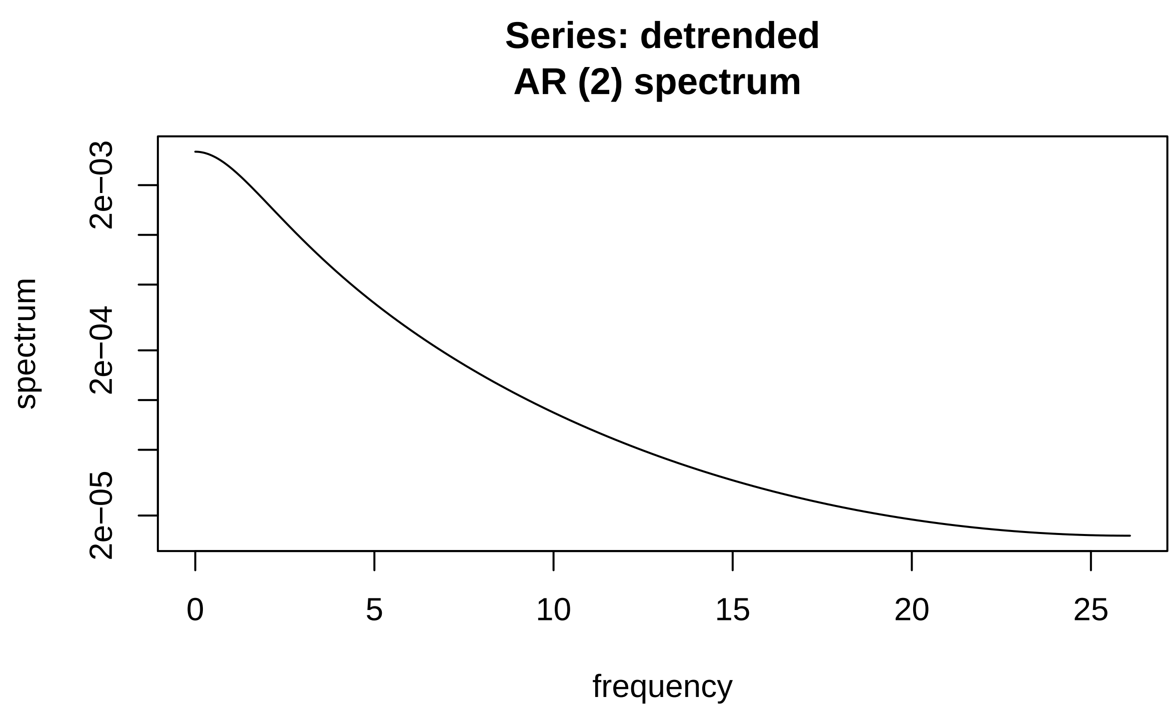
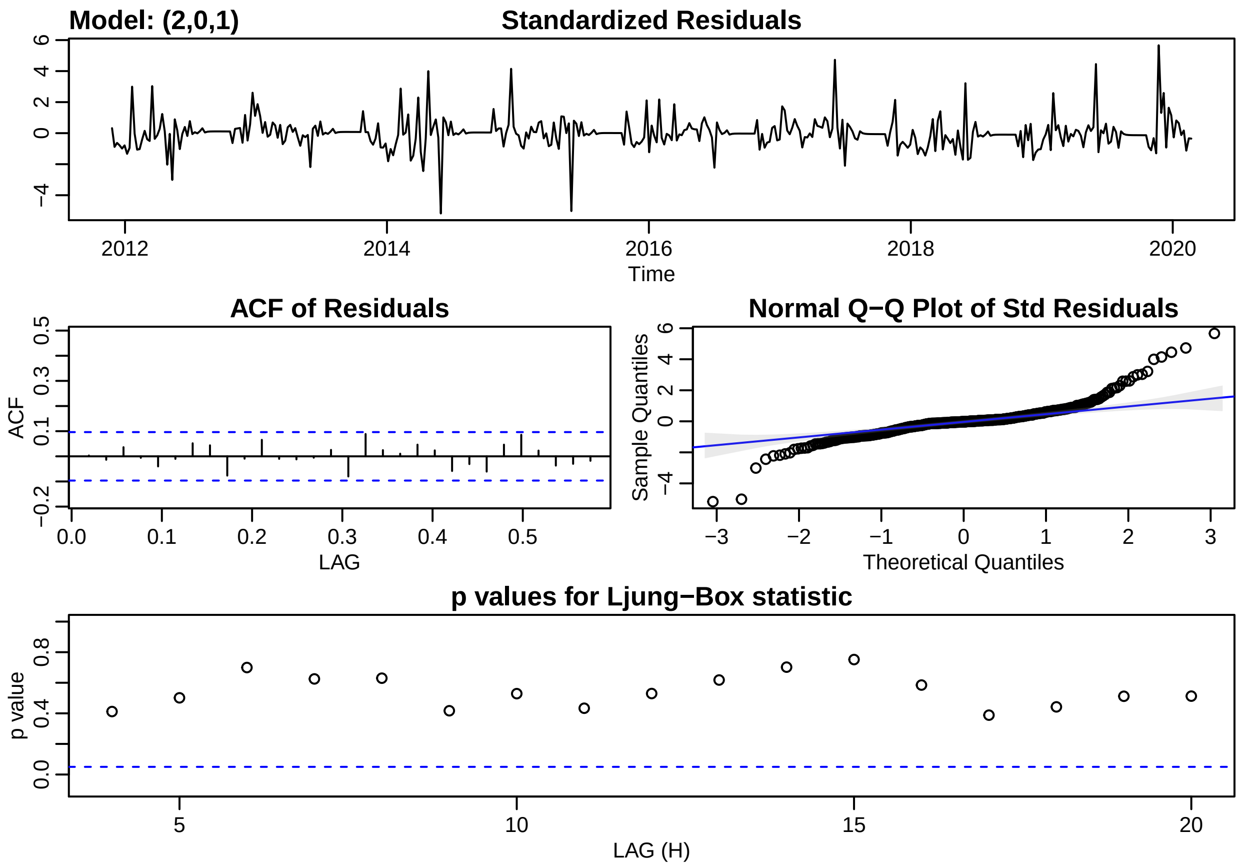


Figure 15: Parametric periodogram of the detrended data, showing the periodogram of an AR(2)

```
## iter   2 value -2.319725
## iter   3 value -2.547063
## iter   4 value -2.660146
## iter   5 value -2.726968
## iter   6 value -2.741929
## iter   7 value -2.748877
## iter   8 value -2.749129
## iter   9 value -2.749377
## iter  10 value -2.750168
## iter  11 value -2.750486
## iter  12 value -2.750493
## iter  13 value -2.750900
## iter  14 value -2.751109
## iter  15 value -2.752079
## iter  16 value -2.754035
## iter  17 value -2.755802
## iter  18 value -2.757934
## iter  19 value -2.759598
## iter  20 value -2.759634
## iter  21 value -2.759709
## iter  22 value -2.759749
## iter  23 value -2.759752
## iter  24 value -2.759794
## iter  25 value -2.759806
## iter  26 value -2.759811
## iter  27 value -2.759811
## iter  27 value -2.759811
## iter  27 value -2.759811
## final  value -2.759811
## converged
## initial  value -2.758644
## iter   2 value -2.758648
## iter   3 value -2.758661
## iter   4 value -2.758669
## iter   5 value -2.758691
## iter   6 value -2.758705
## iter   7 value -2.758707
## iter   8 value -2.758708
## iter   8 value -2.758708
## iter   8 value -2.758708
## final  value -2.758708
## converged
```

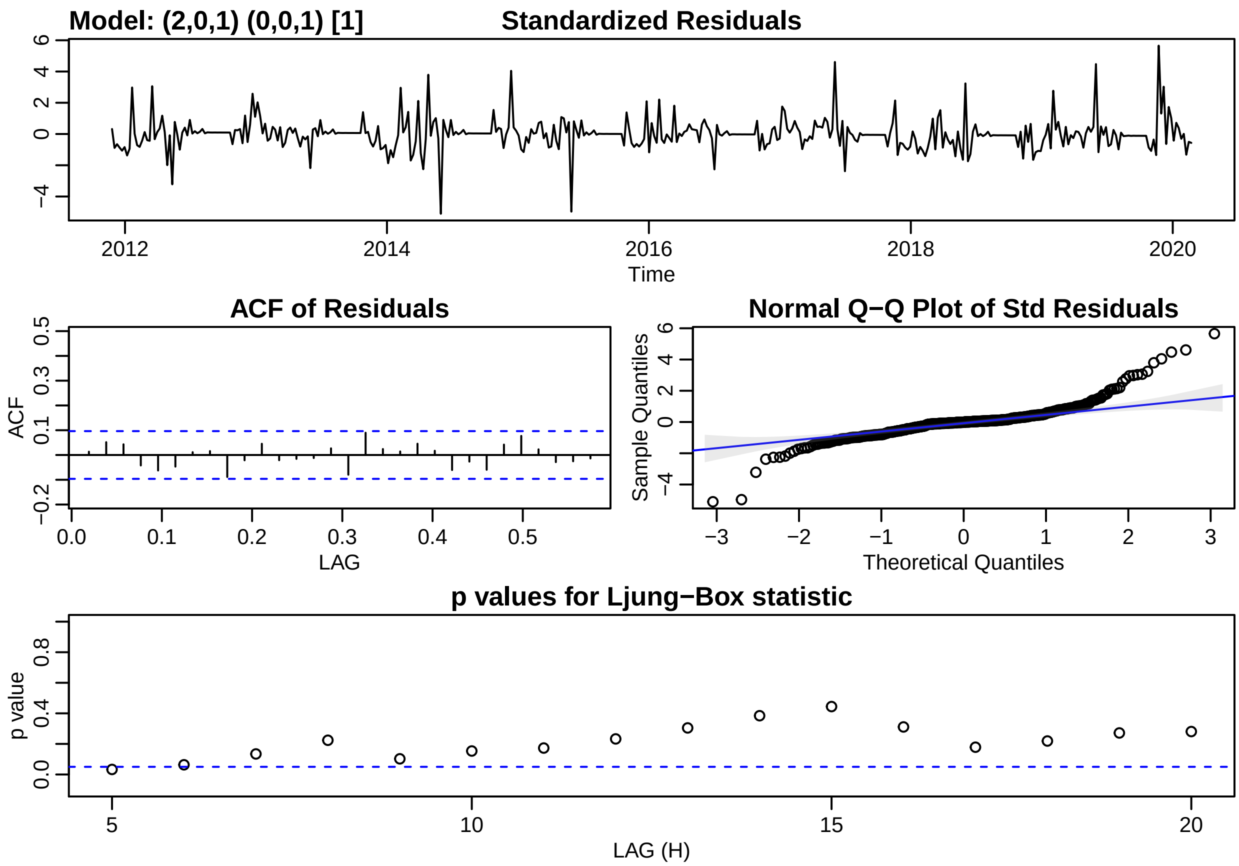



```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xmean, include.mean = FALSE, transform.pars = trans,
##     fixed = fixed, optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1      ar2      ma1    xmean
##       1.3655  -0.4827  -0.3252  -0.0001
## s.e.  0.1739   0.1495   0.1903   0.0174
##
## sigma^2 estimated as 0.004002:  log likelihood = 577.44,  aic = -1144.88
##
## $degrees_of_freedom
## [1] 427
##
## $ttable
##       Estimate     SE t.value p.value
## ar1     1.3655 0.1739  7.8513  0.0000
## ar2    -0.4827 0.1495 -3.2278  0.0013
## ma1    -0.3252 0.1903 -1.7087  0.0882
## xmean  -0.0001 0.0174 -0.0065  0.9948
##
## $AIC
## [1] -2.656336
```

```
##
## $AICc
## [1] -2.656118
##
## $BIC
## [1] -2.609166
```

```
## initial  value -2.019724
## iter   2 value -2.515325
## iter   3 value -2.734814
## iter   4 value -2.751073
## iter   5 value -2.752822
## iter   6 value -2.752865
## iter   7 value -2.752934
## iter   8 value -2.752934
## iter   9 value -2.752937
## iter  10 value -2.752941
## iter  11 value -2.752947
## iter  12 value -2.752951
## iter  13 value -2.752952
## iter  14 value -2.752953
## iter  15 value -2.752954
## iter  16 value -2.752958
## iter  17 value -2.752994
## iter  18 value -2.752997
## iter  19 value -2.752999
## iter  20 value -2.753002
## iter  21 value -2.753012
## iter  22 value -2.753024
## iter  23 value -2.753040
## iter  24 value -2.753047
## iter  25 value -2.753050
## iter  26 value -2.753050
## iter  27 value -2.753050
## iter  28 value -2.753050
## iter  29 value -2.753051
## iter  30 value -2.753051
## iter  31 value -2.753051
## iter  32 value -2.753051
## iter  32 value -2.753051
## iter  32 value -2.753051
## final  value -2.753051
## converged
## initial  value -2.753656
## iter   2 value -2.753659
## iter   3 value -2.753668
## iter   4 value -2.753682
## iter   5 value -2.753709
## iter   6 value -2.753747
## iter   7 value -2.753764
## iter   8 value -2.753770
## iter   9 value -2.753771
## iter  10 value -2.753771
## iter  11 value -2.753773
```

```
## iter  12 value -2.753774
## iter  13 value -2.753774
## iter  14 value -2.753775
## iter  14 value -2.753775
## iter  14 value -2.753775
## final  value -2.753775
## converged
```



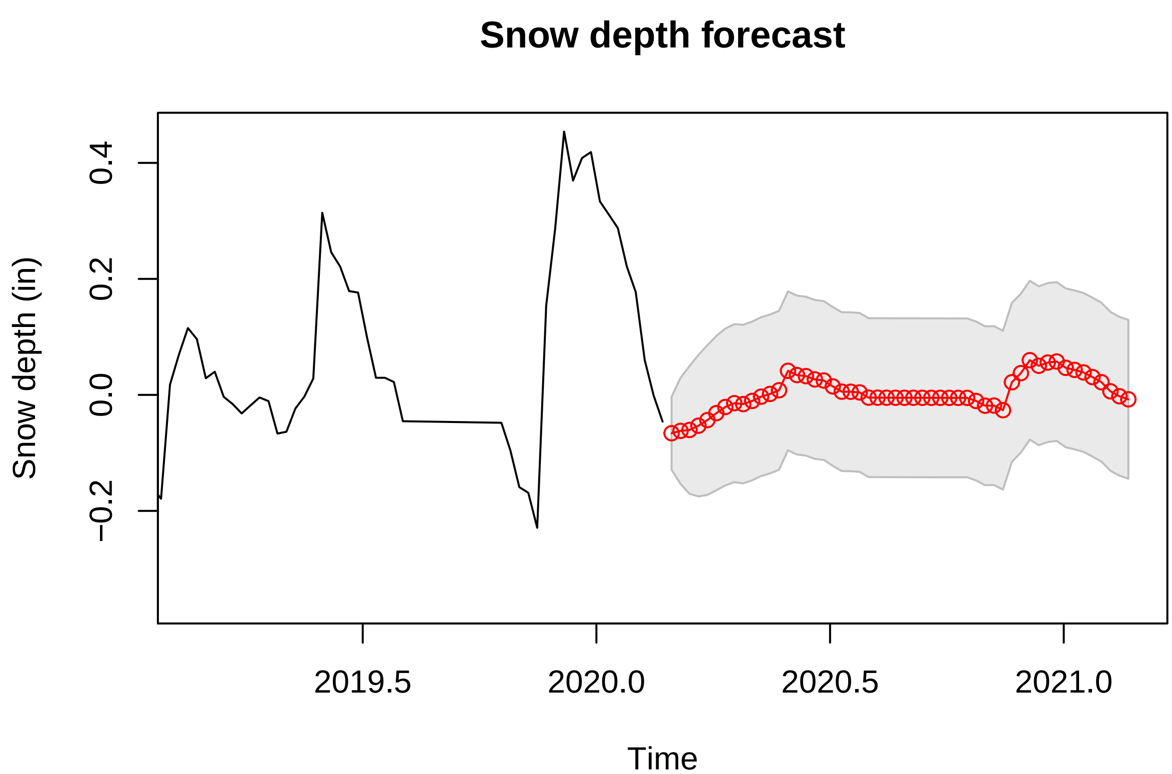

```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xmean, include.mean = FALSE, transform.pars = trans,
##     fixed = fixed, optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1     ar2     ma1    sma1   xmean
##       0.2922  0.4271  0.3726  0.3726  0.0003
## s.e.  0.2118  0.1600  0.3585  0.3585  0.0203
##
## sigma^2 estimated as 0.004042:  log likelihood = 575.31,  aic = -1138.63
##
## $degrees_of_freedom
## [1] 426
##
```

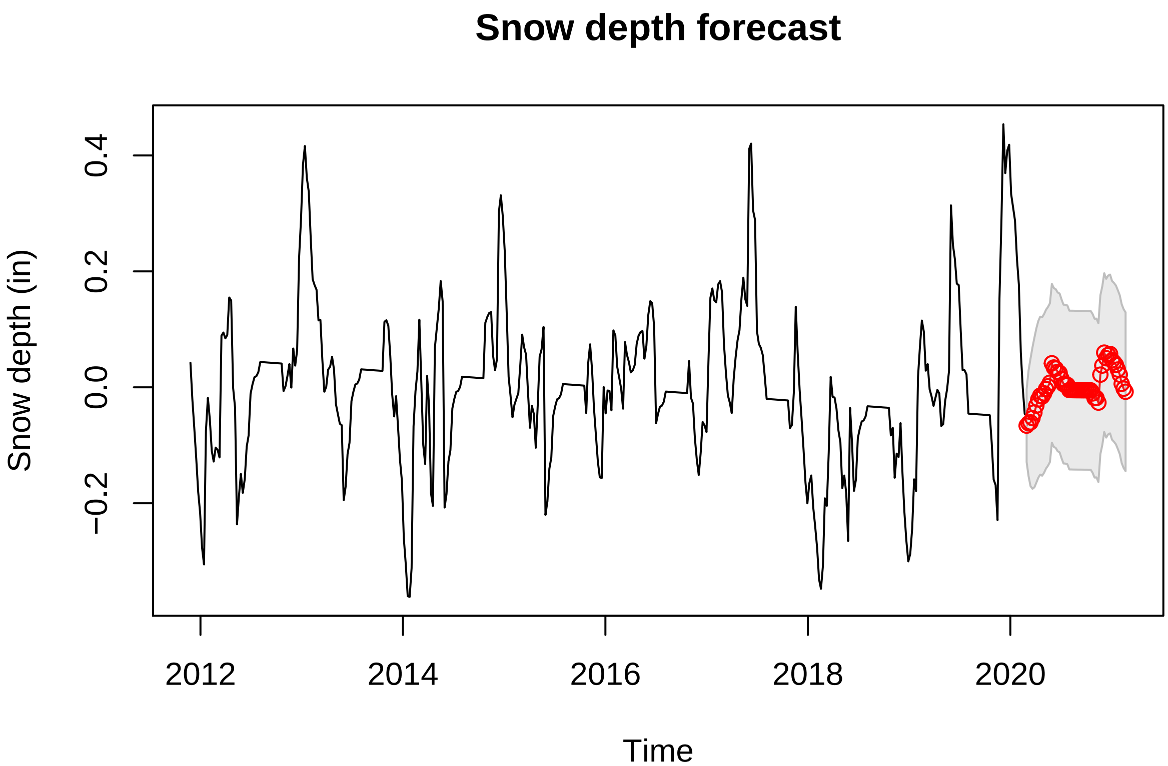
```
## $ttable
##       Estimate     SE t.value p.value
## ar1     0.2922 0.2118  1.3793  0.1685
## ar2     0.4271 0.1600  2.6689  0.0079
## ma1     0.3726 0.3585  1.0393  0.2993
## sma1    0.3726 0.3585  1.0393  0.2993
## xmean   0.0003 0.0203  0.0131  0.9895
##
## $AIC
## [1] -2.64183
##
## $AICc
## [1] -2.641502
##
## $BIC
## [1] -2.585225
```

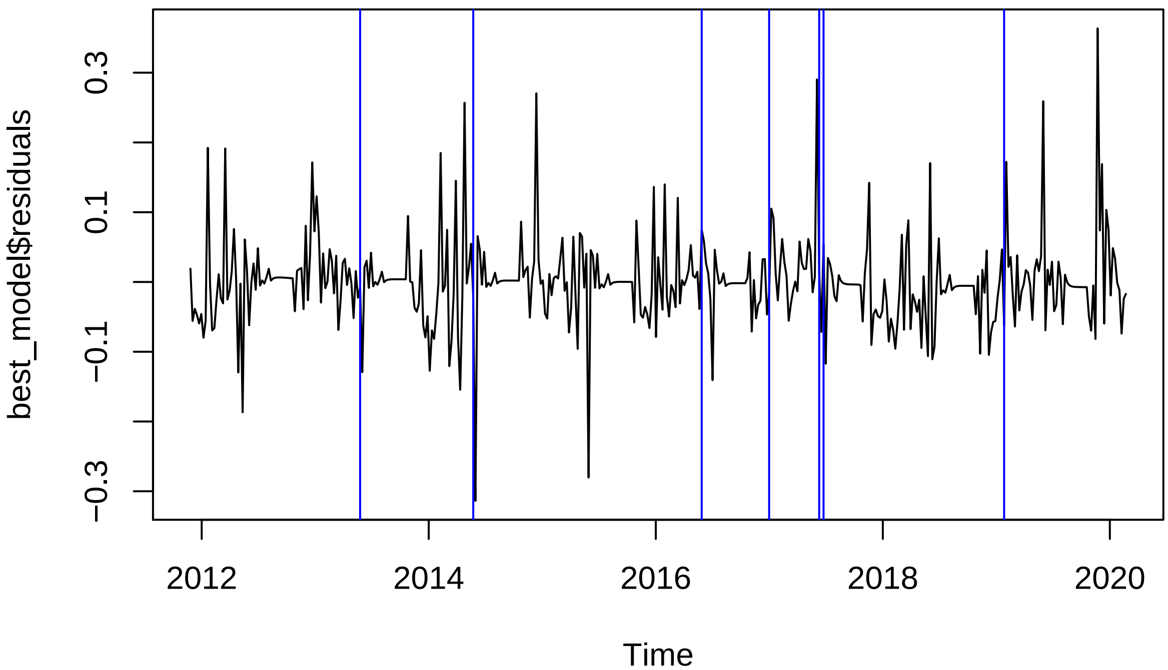
```
## [[1]]
##
## Call:
## arima(x = detrended, order = c(2, 0, 1), seasonal = list(order = c(0, 0, 1),
##     period = 52))
##
## Coefficients:
##          ar1      ar2      ma1    sma1  intercept
##       1.4321  -0.5351  -0.3756  0.1261     0.0004
## s.e.  0.1595   0.1381   0.1766  0.0549     0.0202
##
## sigma^2 estimated as 0.003945:  log likelihood = 580.04,  aic = -1148.08
##
## [[2]]
##
## Call:
## arima(x = detrended, order = c(2, 0, 2), seasonal = list(order = c(0, 0, 1),
##     period = 52))
##
## Coefficients:
##          ar1      ar2      ma1      ma2    sma1  intercept
##       1.4727  -0.5669  -0.4125  -0.0218  0.1263     0.0004
## s.e.  0.2116   0.1737   0.2144   0.0840  0.0549     0.0201
##
## sigma^2 estimated as 0.003945:  log likelihood = 580.07,  aic = -1146.14
##
## [[3]]
##
## Call:
## arima(x = detrended, order = c(2, 0, 1))
##
## Coefficients:
##          ar1      ar2      ma1  intercept
##       1.3655  -0.4827  -0.3252    -0.0001
## s.e.  0.1739   0.1495   0.1903     0.0174
##
## sigma^2 estimated as 0.004002:  log likelihood = 577.44,  aic = -1144.88
```

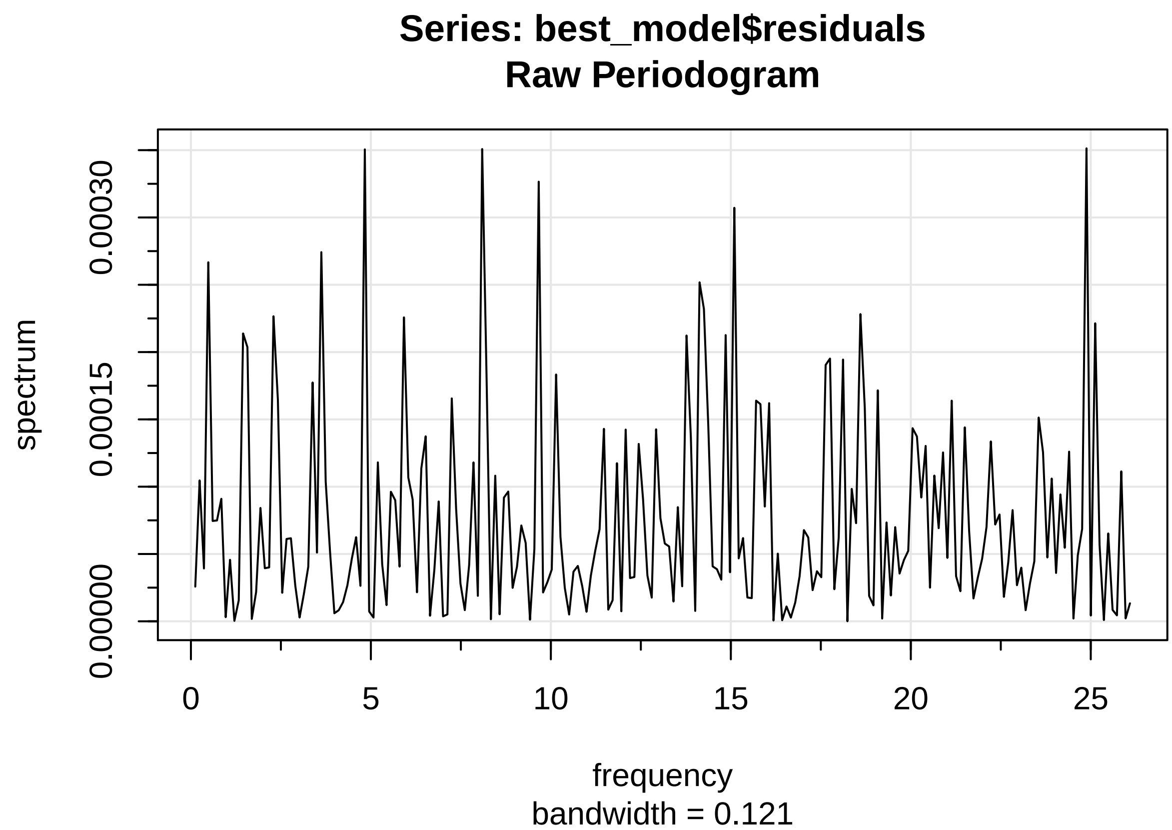
```
##
## [[4]]
##
## Call:
## arima(x = detrended, order = c(2, 0, 2))
##
## Coefficients:
##          ar1      ar2      ma1      ma2  intercept
##       1.4077  -0.5153  -0.3646  -0.0195     0.0000
## s.e.  0.2471   0.1999   0.2495   0.0899     0.0173
##
## sigma^2 estimated as 0.004002:  log likelihood = 577.46,  aic = -1142.93
```

```
## [[1]]
## [1] -1148.077
##
## [[2]]
## [1] -1146.144
##
## [[3]]
## [1] -1144.881
##
## [[4]]
## [1] -1142.925
```

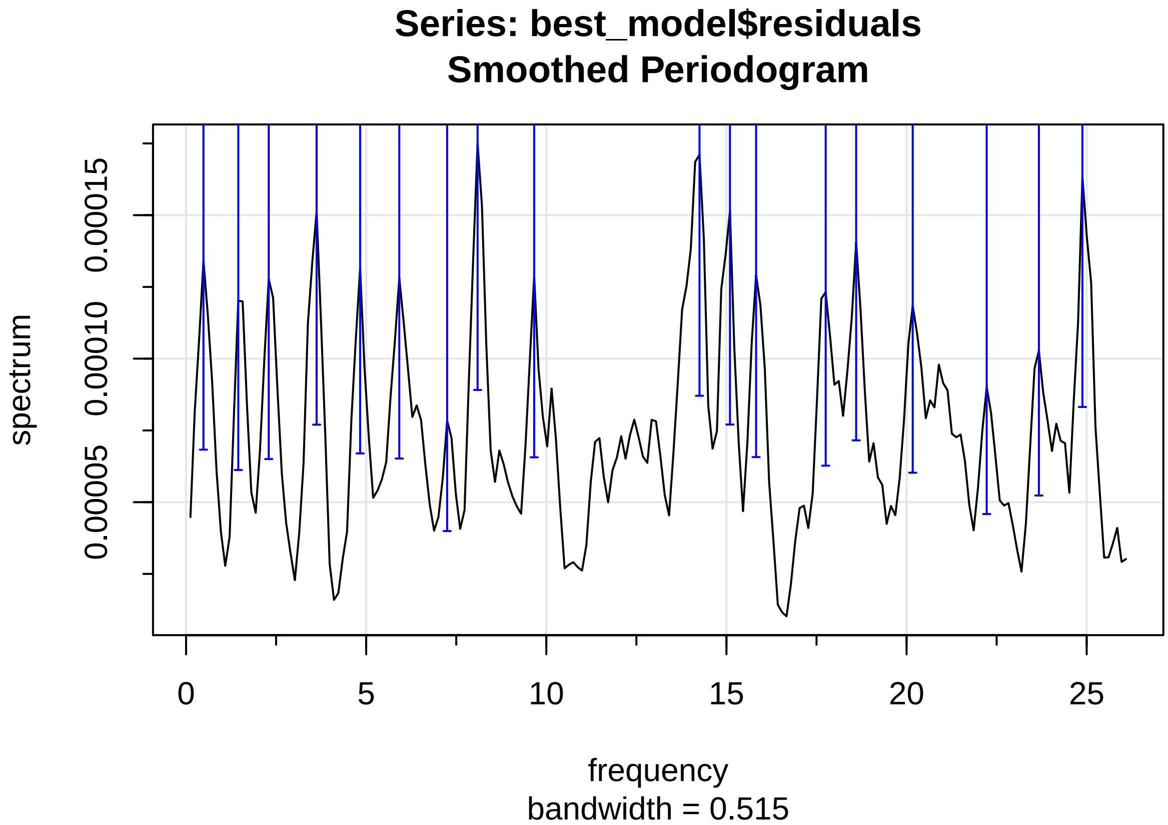


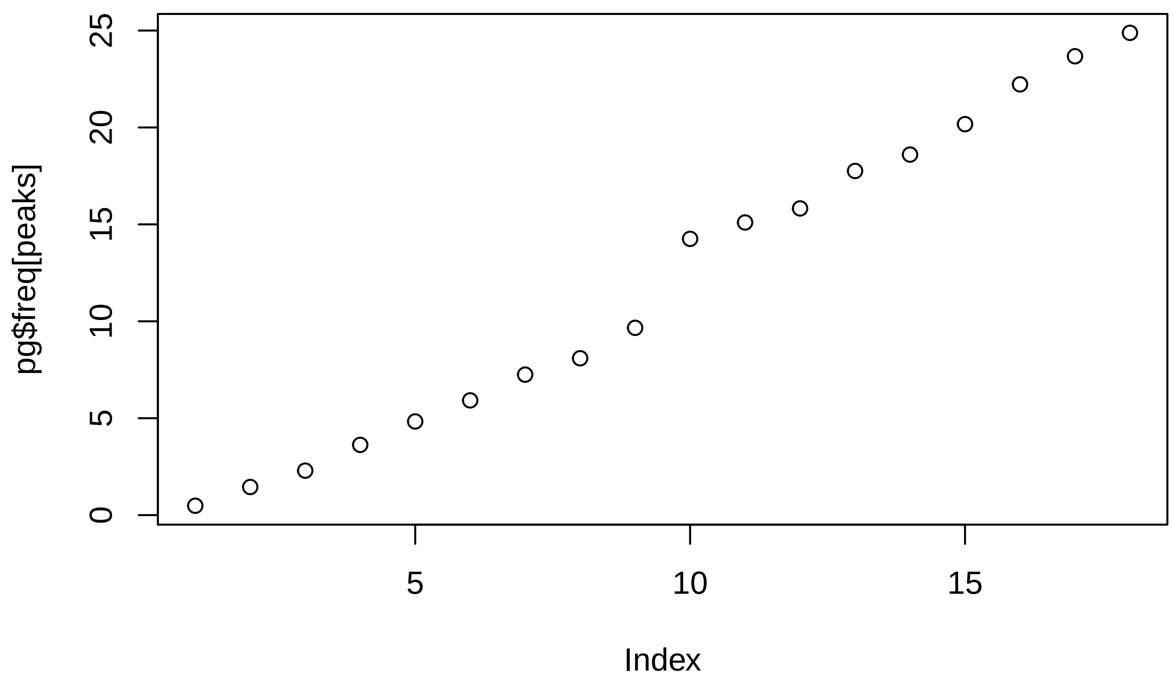

Snow depth forecast
Snow depth (in)
0.4
0.2
0.0
−0.2
2012
2014
2016
2018
2020
Time


best_model$residuals
0.3
0.1
−0.1
−0.3
2012
2014
2016
2018
2020
Time




```
##  [1] 2.06981520 0.68993840 0.43575057 0.27597536 0.20698152 0.16896451
##  [7] 0.13798768 0.12357106 0.10349076 0.07016323 0.06623409 0.06320046
## [13] 0.05632150 0.05376143 0.04957641 0.04499598 0.04224113 0.04019059
```


Series: best_model$residuals
Smoothed Periodogram
spectrum
0.00005
0.00010
0.00015
0
5
10
15
20
25
frequency
bandwidth = 0.515



pg$freq[peaks]
0
5
10
15
20
25
5
10
15
Index

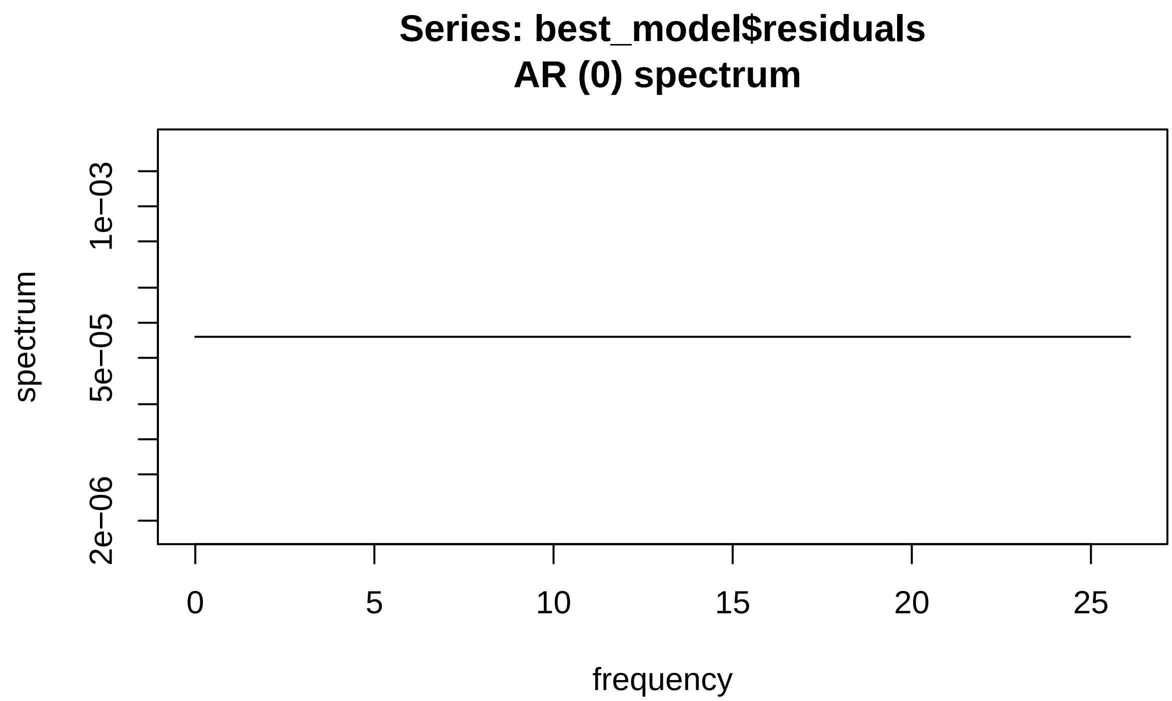
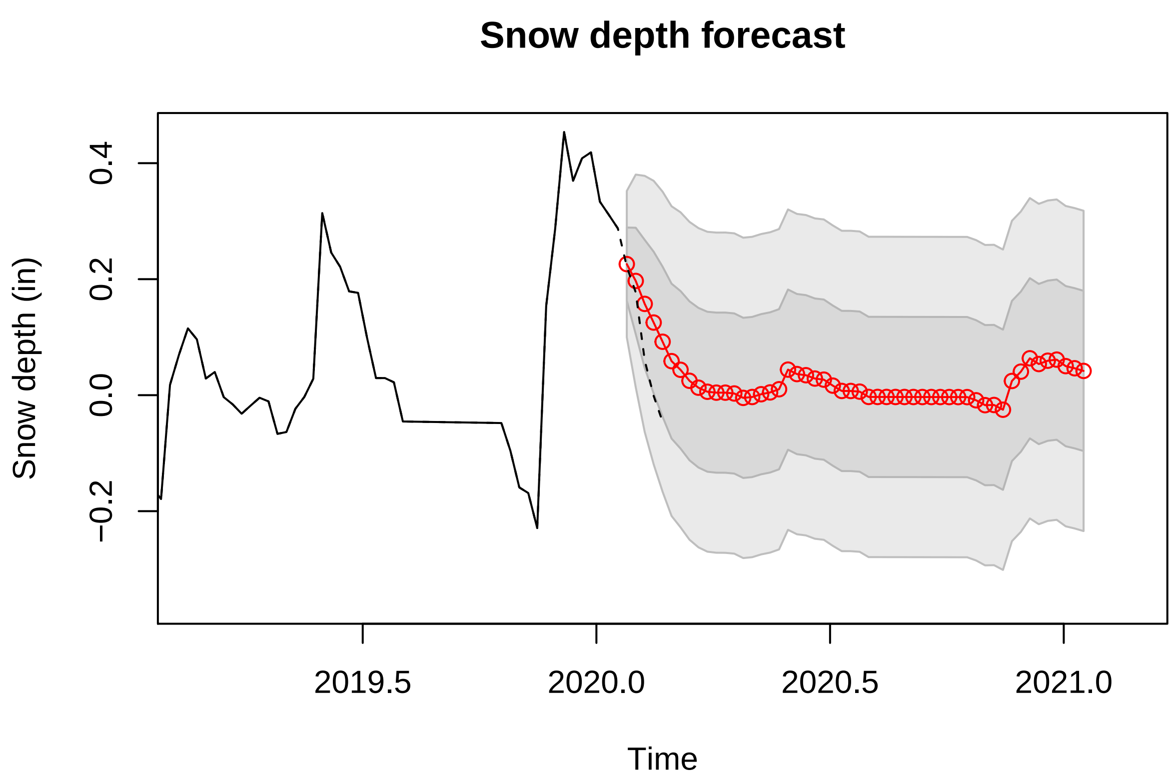


Figure 16: Periodogram of the best fitting AR(p) model of the detrended data



5 References

1. https://www.mammothmountain.com/