STAT 221 Final Project - Mammoth Snow Depth

John Rapp Farnes | 405461225

3/17 2020

Contents

% TODO - Prediction på known värden - Tillbaka trend på pred

- Kommentera stora residualer? datum?
- Log: större skillnad större värden, plotta
- Periodogram residualer
- Tapering

1 Introduction

1.1 Background

Mammoth is a mountain in Northern California 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 and recieves over 1 million anuual visitors. For Southern Californian residents, 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. **??**

As people familiar with alpine sports know, one of the most important conditions for the sport is the snow depth at the mountain, as this affects which runs are open and how "good" the skiing is. As such, the purpose of this report is to **model the weekly 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 give the data easily in a downloadable form, hence the data was obtained through injecting JavaScript into a browser client that took the data from the browser JavaScript enviroment and printed it in a PDF 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 incehs. Opon 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 cleaned.

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 data. 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, creating 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 not snow on Mammoth during the summer. The resulting data after cleaning is shown in figure **??**, featuring 431 weeks. The availability of data per month is displayed in **??**, showing that data exists for the most part Dec-May, with less than 50% Jun-Oct, reaching ~80% in November.

2 Analyis

Looking at the grapth, 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. To study other properties than this obvious observation, time-series methods were applied.

2.1 Series properties

First, the properties of the time-series were studies. Figure ref{fig:acf_series} shows the ACF and PACF of the series, where the ACF shows periodic behaviour with length 1 year that tails off, and the PACF cuts off after lag 2, 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 ref{fig:period_series}), together with a 4 year period, both 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 random chance.

The data does not appear to be stationary as it has an obvious yearly trend, and ARIMA models can not be applied. Hence the data must first be detrended.

2.2 Detrending

2.3 Detrended series properties

2.4 Model fitting

2.5 Model interpretation

2.6 Model evaluation

2.6.1 Residual analysis

2.7 Frequency domain

2.8 Alternative model

3 Results

3.1 Conclusion

3.2 Dicsussion

3.3 Next steps

4 Appendix - Figures and graphs

5 References

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

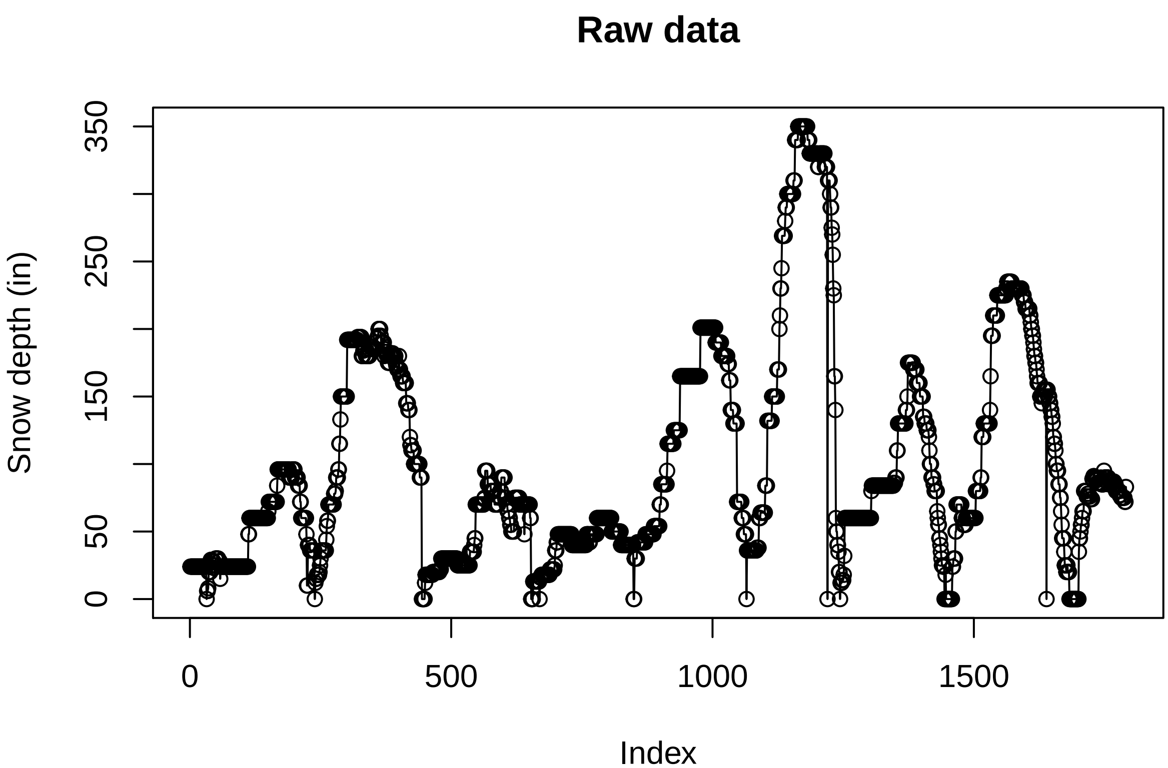
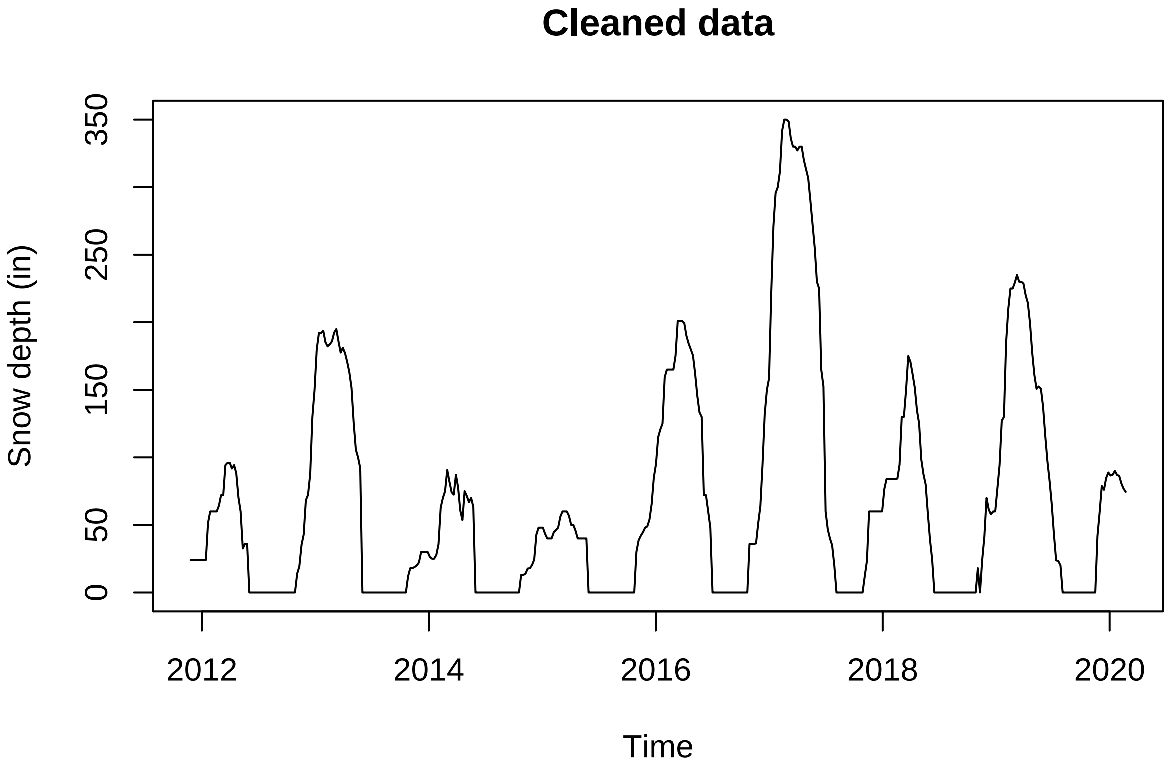
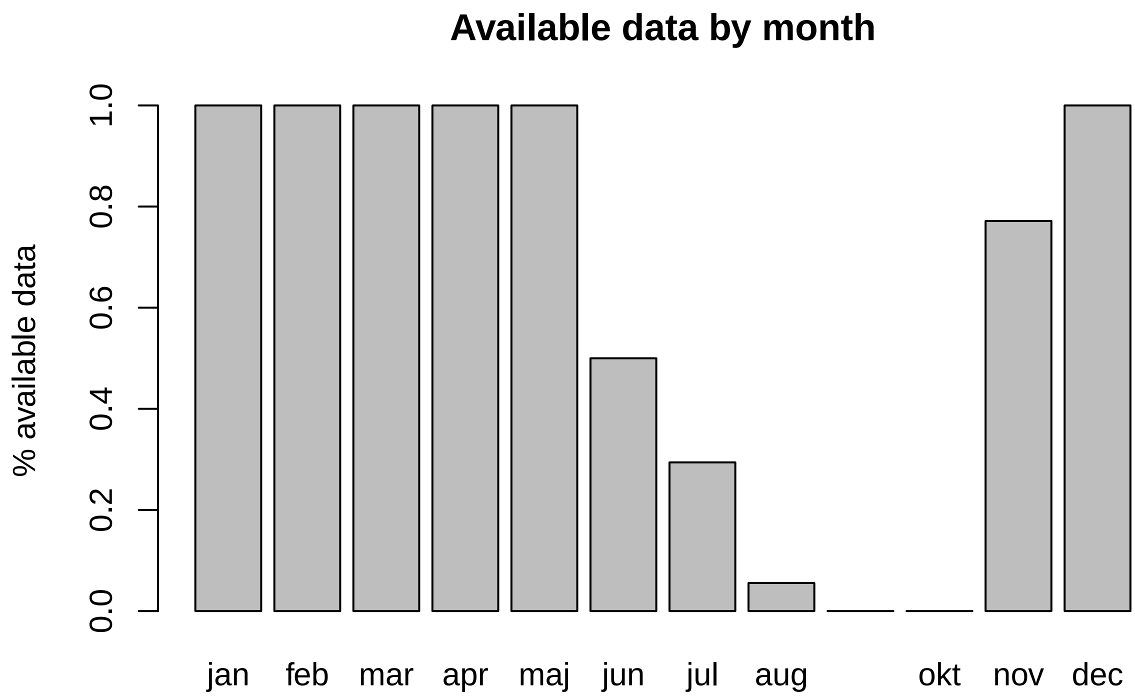


Figure 1: Raw Mammoth snow depth data


Cleaned data
Snow depth (in)
0
50
150
250
350
2012
2014
2016
2018
2020
Time


Available data by month
% available data
0.0
0.2
0.4
0.6
0.8
1.0
jan
feb
mar
apr
maj
jun
jul
aug
okt
nov
dec

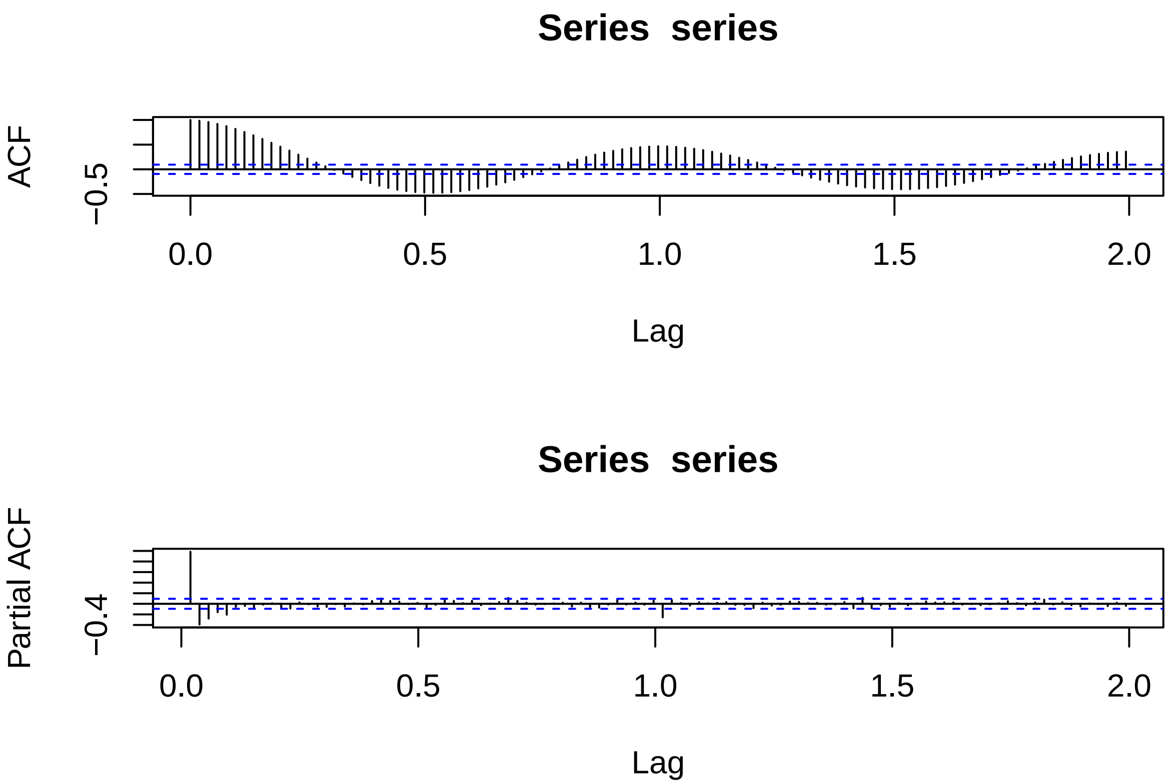
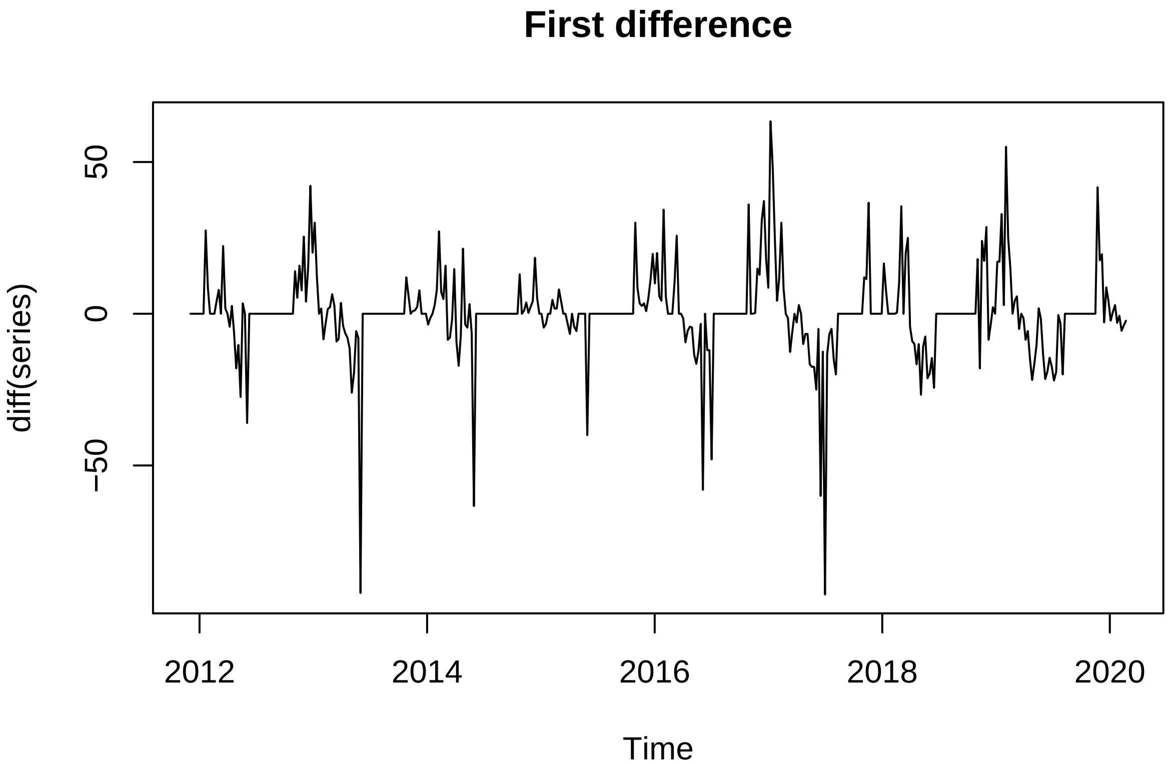
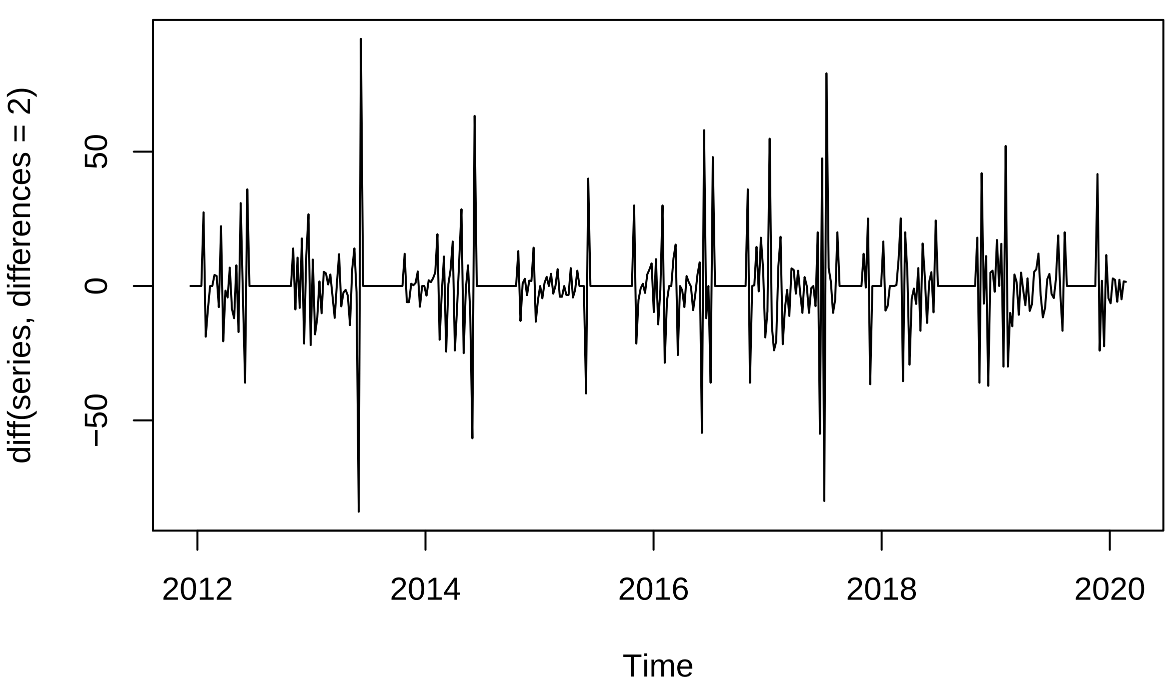
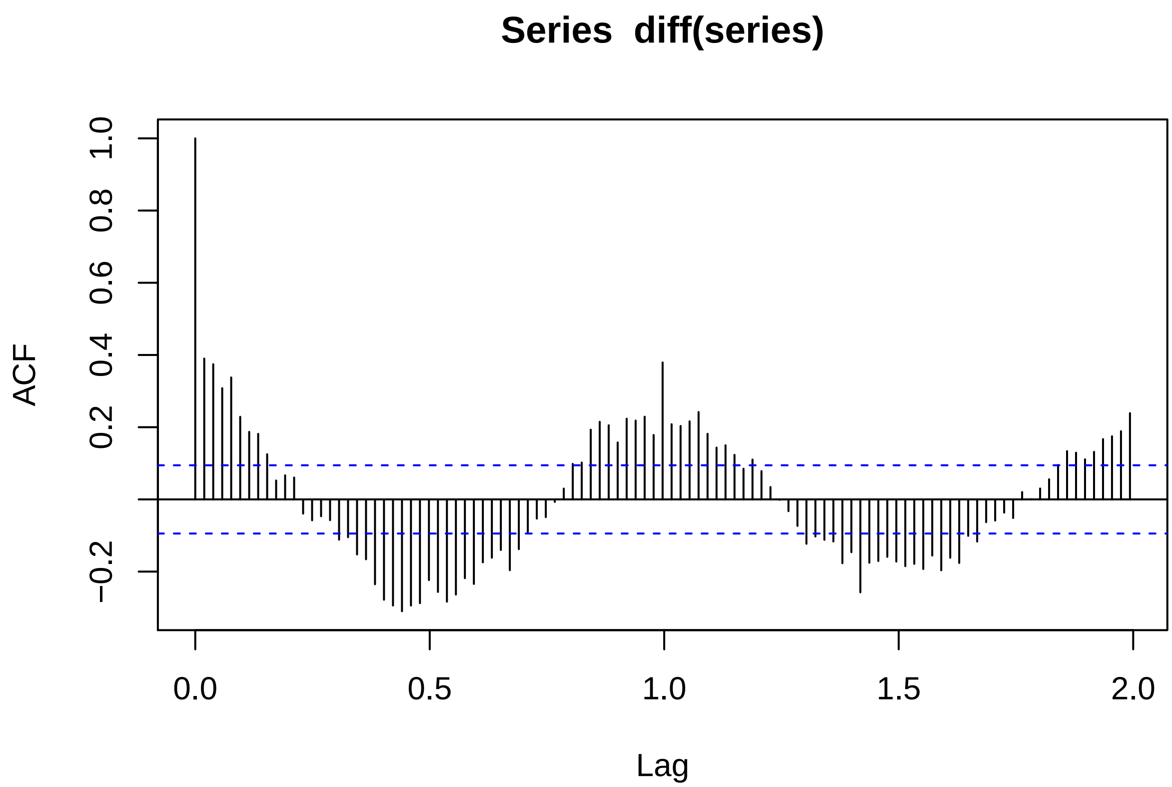

Series series
ACF
−0.5
0.0
0.5
1.0
1.5
2.0
Lag
Series series
Partial ACF
−0.4
0.0
0.5
1.0
1.5
2.0
Lag

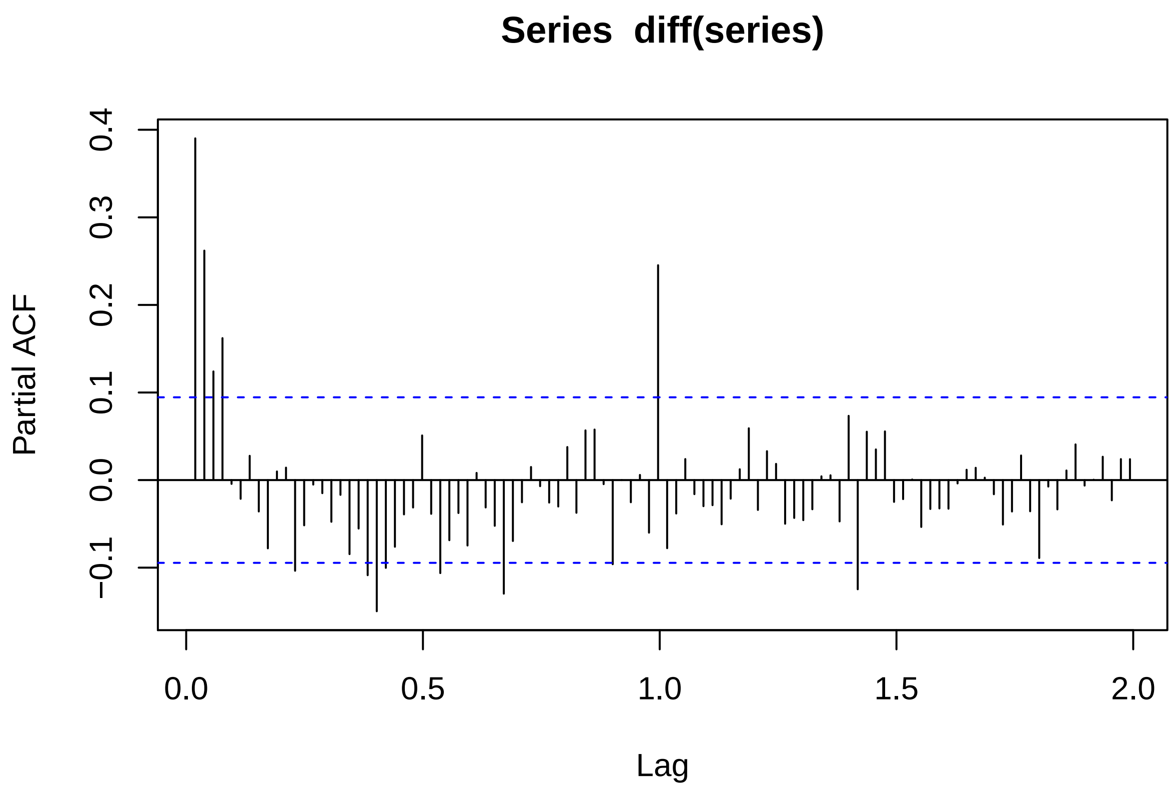
Figure 2: HEJ

6 ACF


First difference
50
0
−50
diff(series)
2012
2014
2016
2018
2020
Time

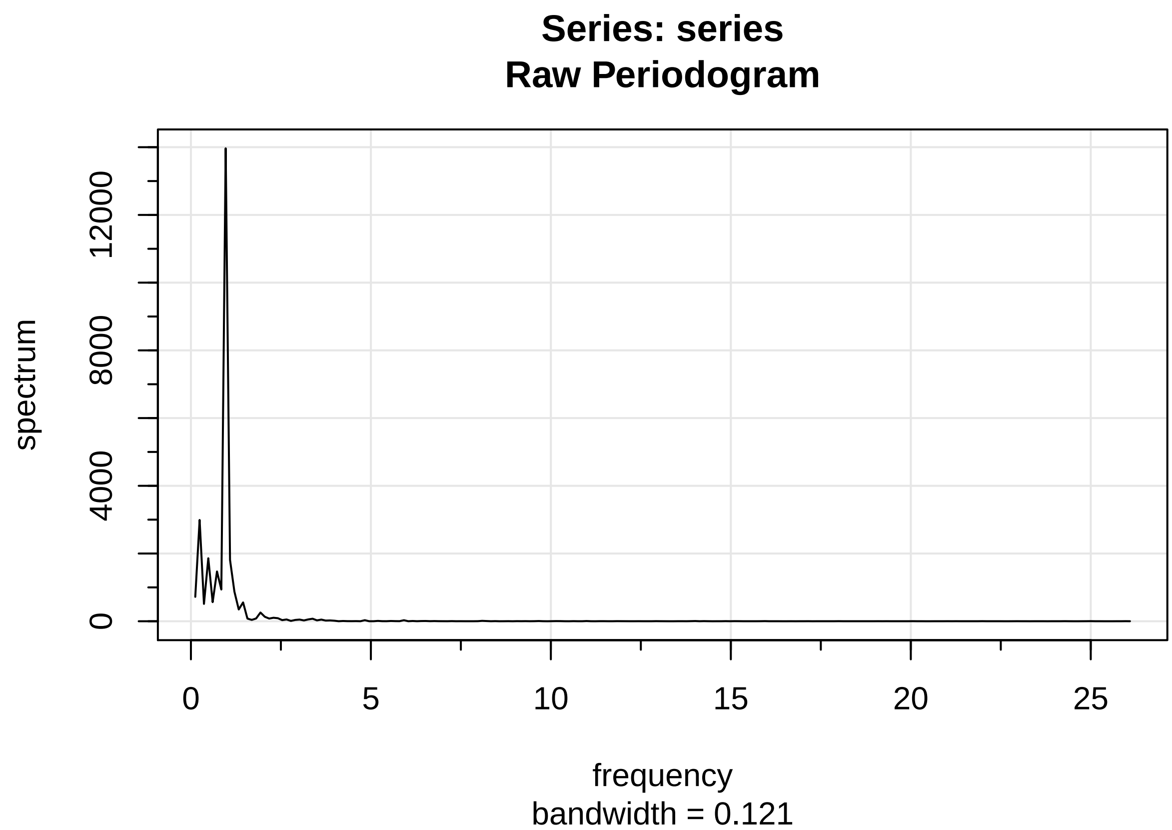

50
0
−50
diff(series, differences = 2)
2012
2014
2016
2018
2020
Time


Series diff(series)
ACF
1.0
0.8
0.6
0.4
0.2
−0.2
0.0
0.5
1.0
1.5
2.0
Lag



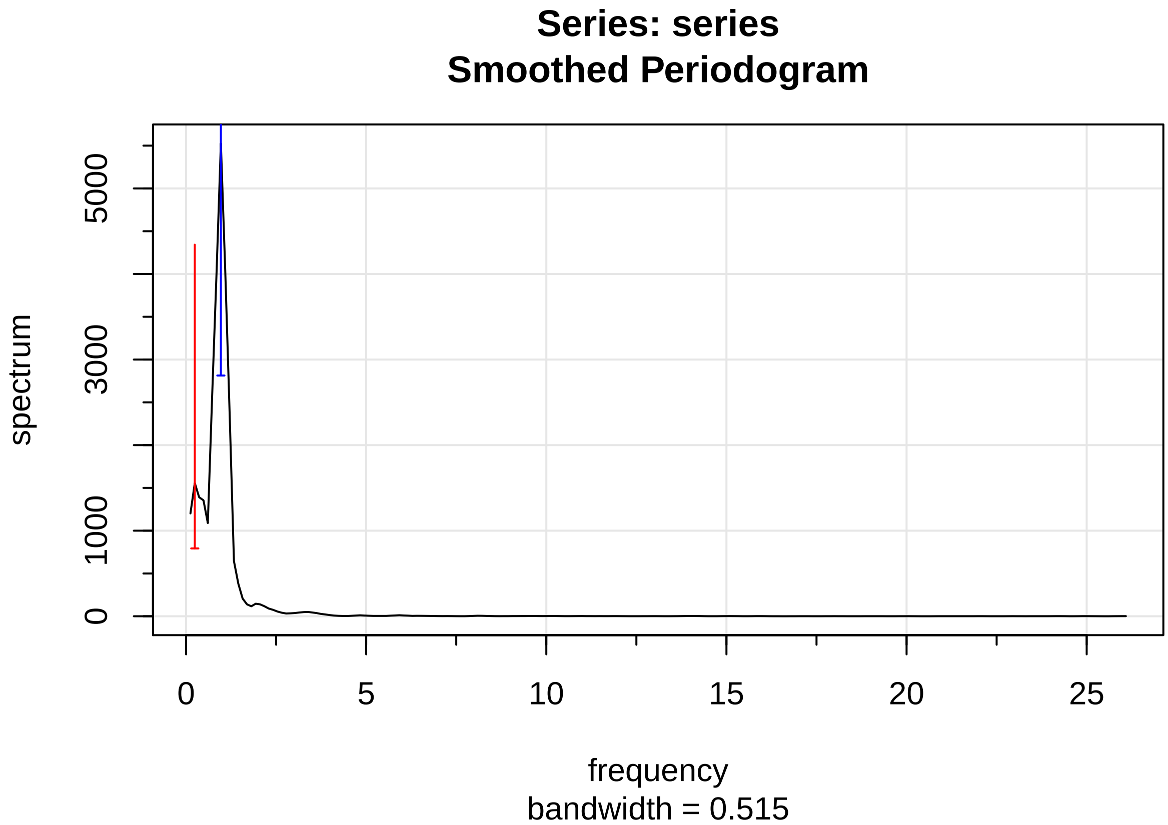

```
## Warning in adf.test(series): p-value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  series
## Dickey-Fuller = -5.5486, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

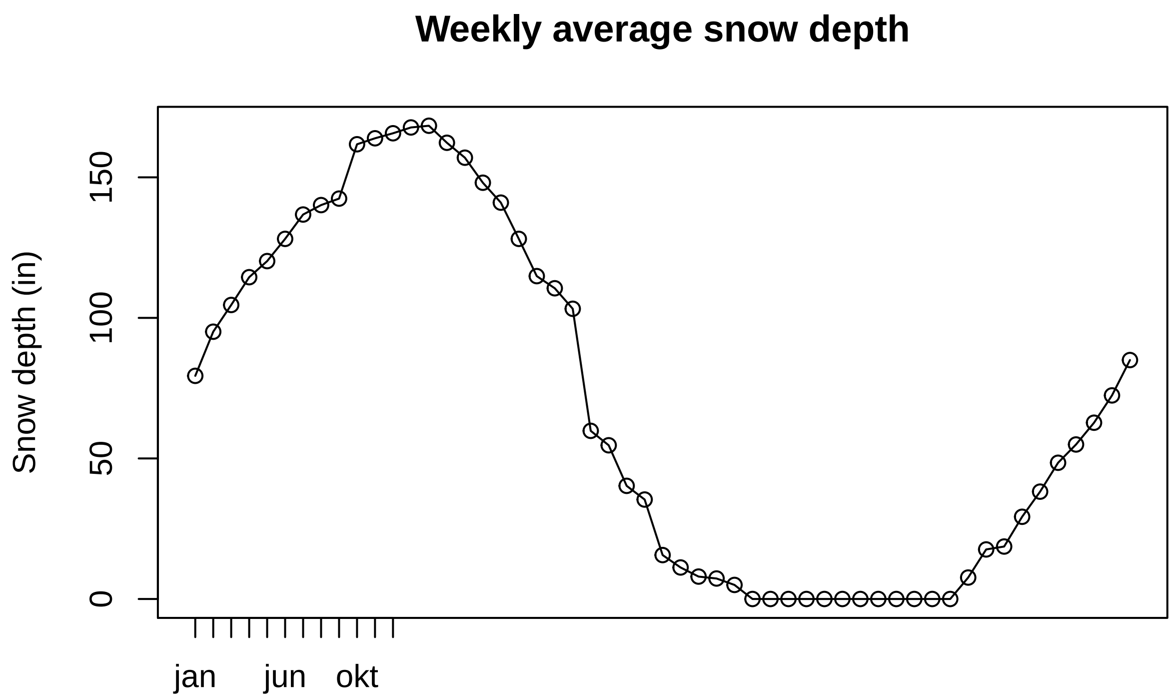



```
## [1] 1.034908
```

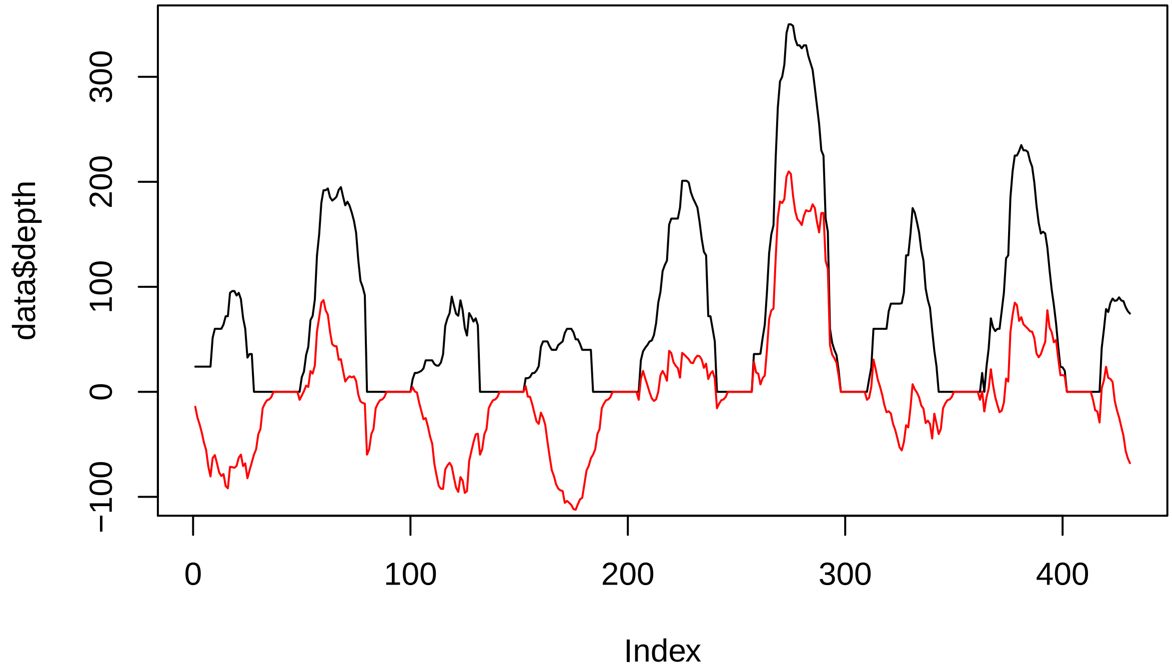
```
## [1] 4.13963
```


Series: series
Smoothed Periodogram
spectrum
0
1000
3000
5000
0
5
10
15
20
25
frequency
bandwidth = 0.515

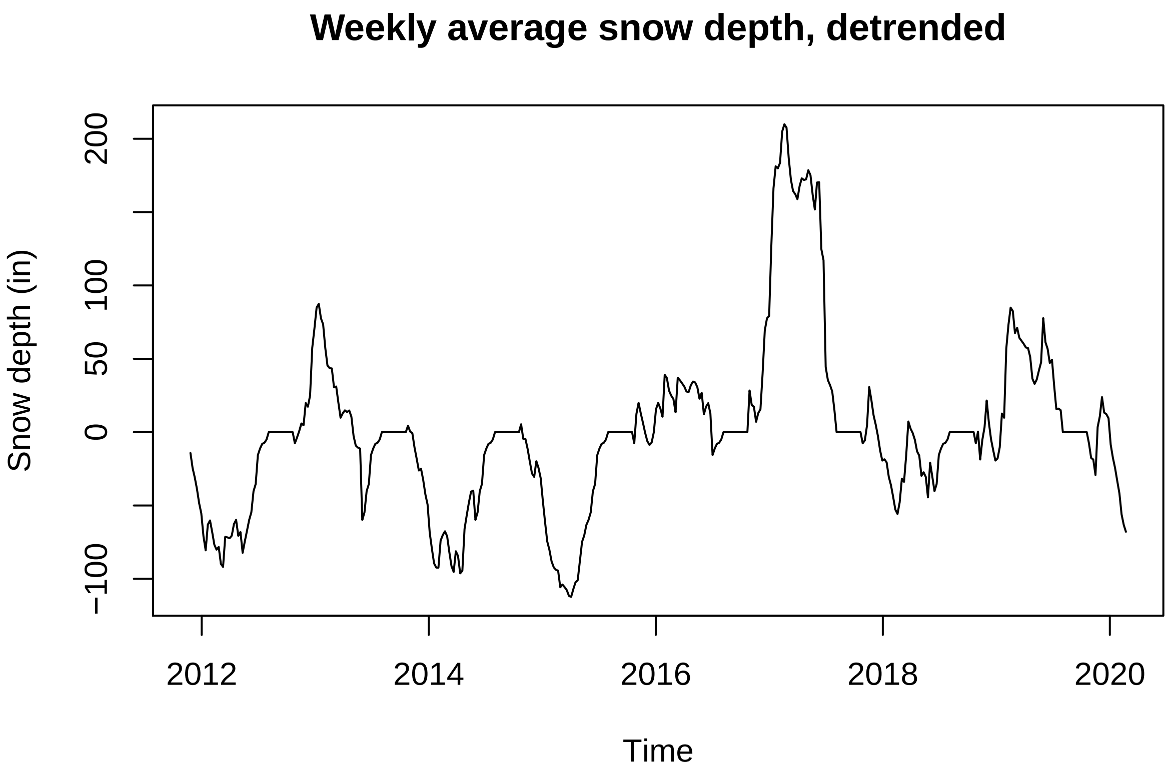
7 Detrend



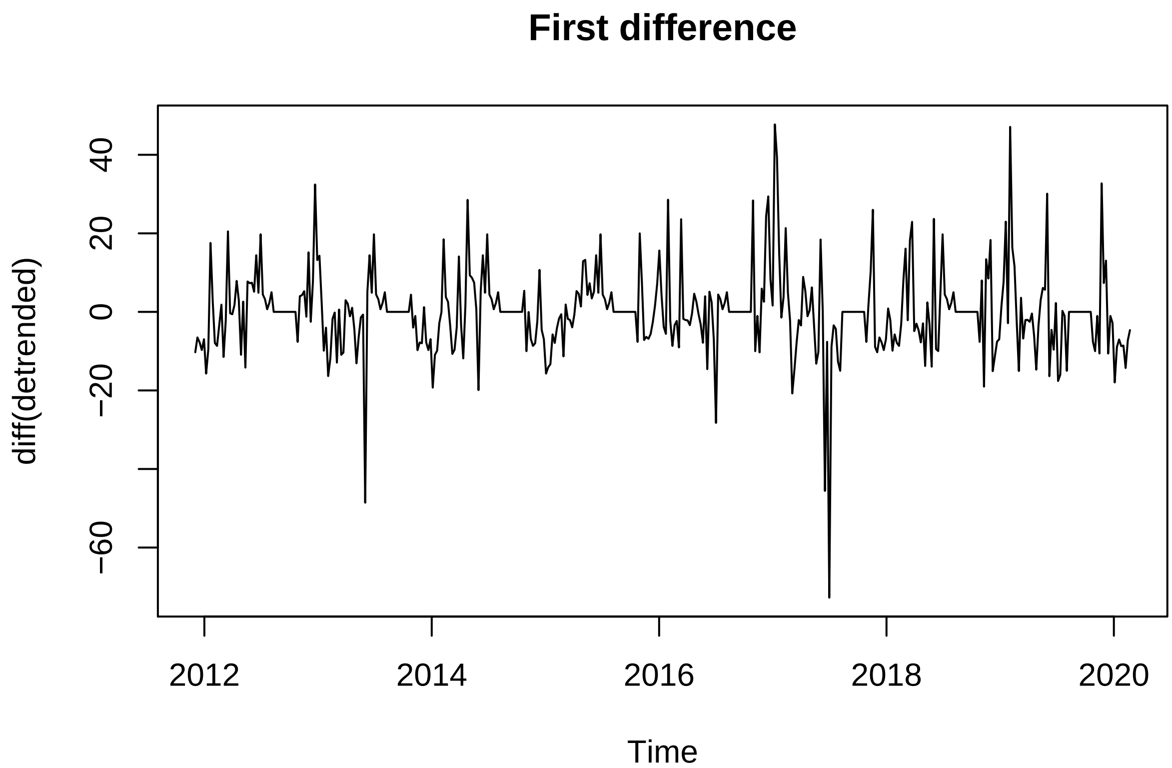
Weekly average snow depth
Snow depth (in)
0
50
100
150
jan
jun
okt



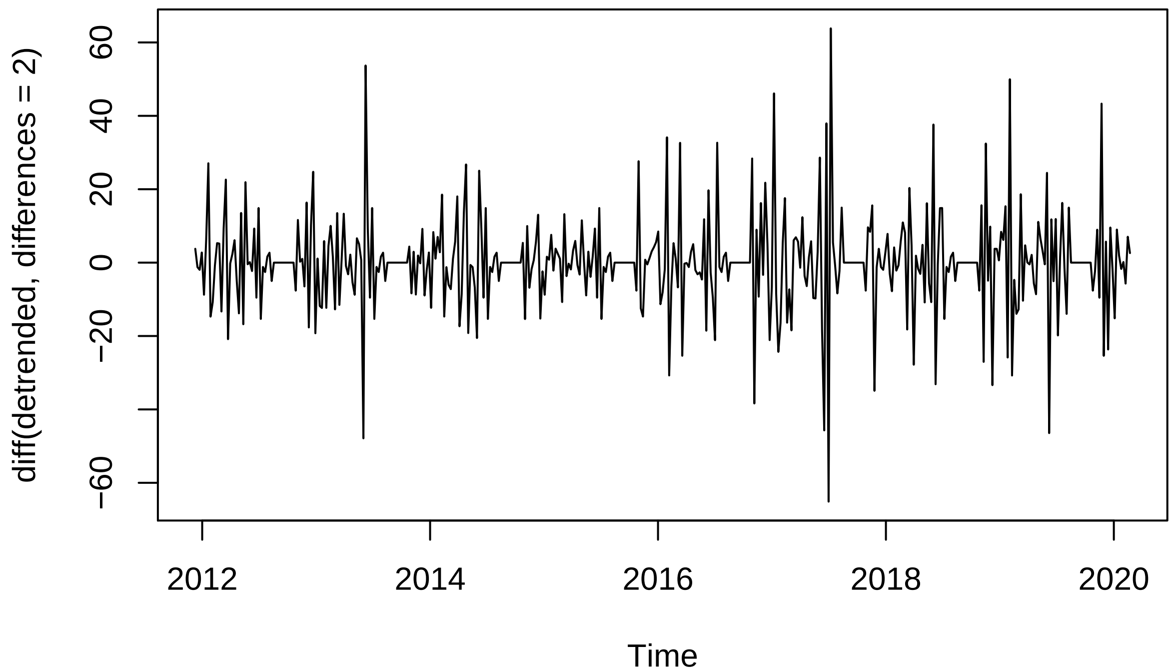
data$depth
−100
0
100
200
300
0
100
200
300
400
Index

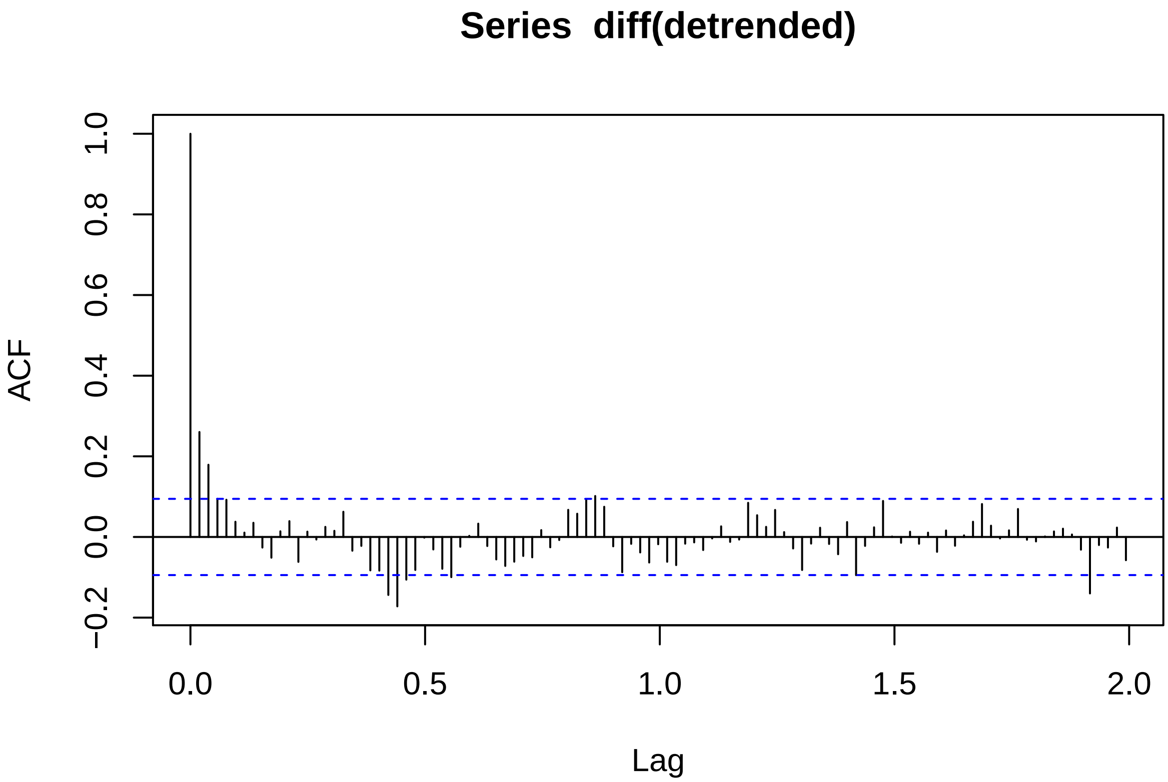


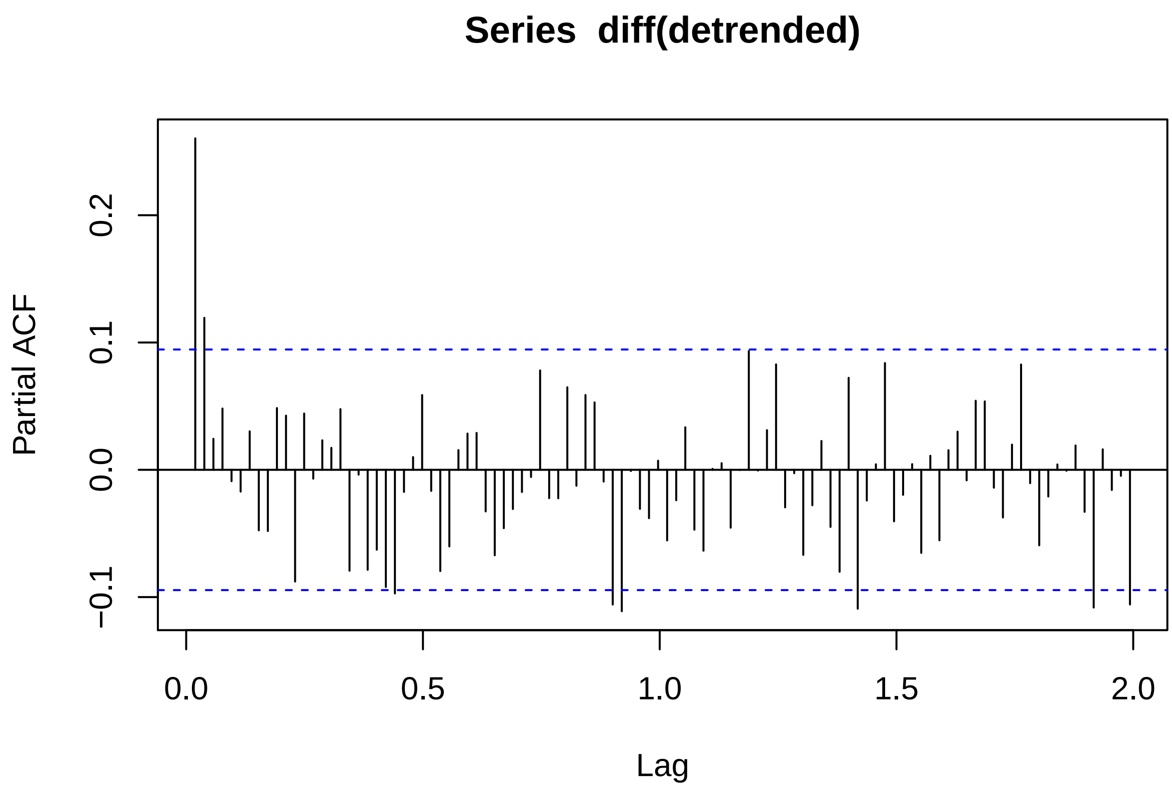
Weekly average snow depth, detrended
Snow depth (in)
200
100
50
0
−100
2012
2014
2016
2018
2020
Time



First difference
diff(detrended)
40
20
0
−20
−60
2012
2014
2016
2018
2020
Time


diff(detrended, differences = 2)
60
40
20
0
−20
−60
2012
2014
2016
2018
2020
Time


Series diff(detrended)
ACF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0.0
0.5
1.0
1.5
2.0
Lag

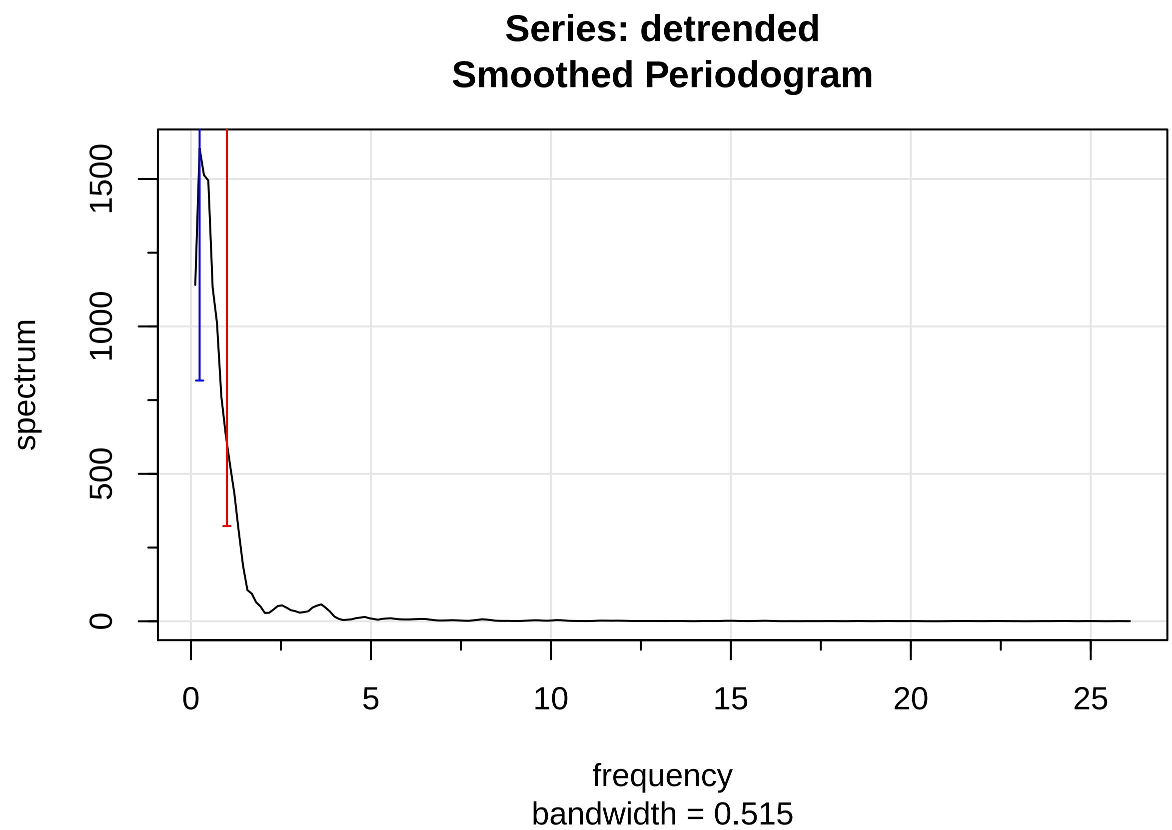



```
##
##  Augmented Dickey-Fuller Test
##
## data:  detrended
## Dickey-Fuller = -3.2378, Lag order = 7, p-value = 0.08159
## alternative hypothesis: stationary
```

8 Periodogram

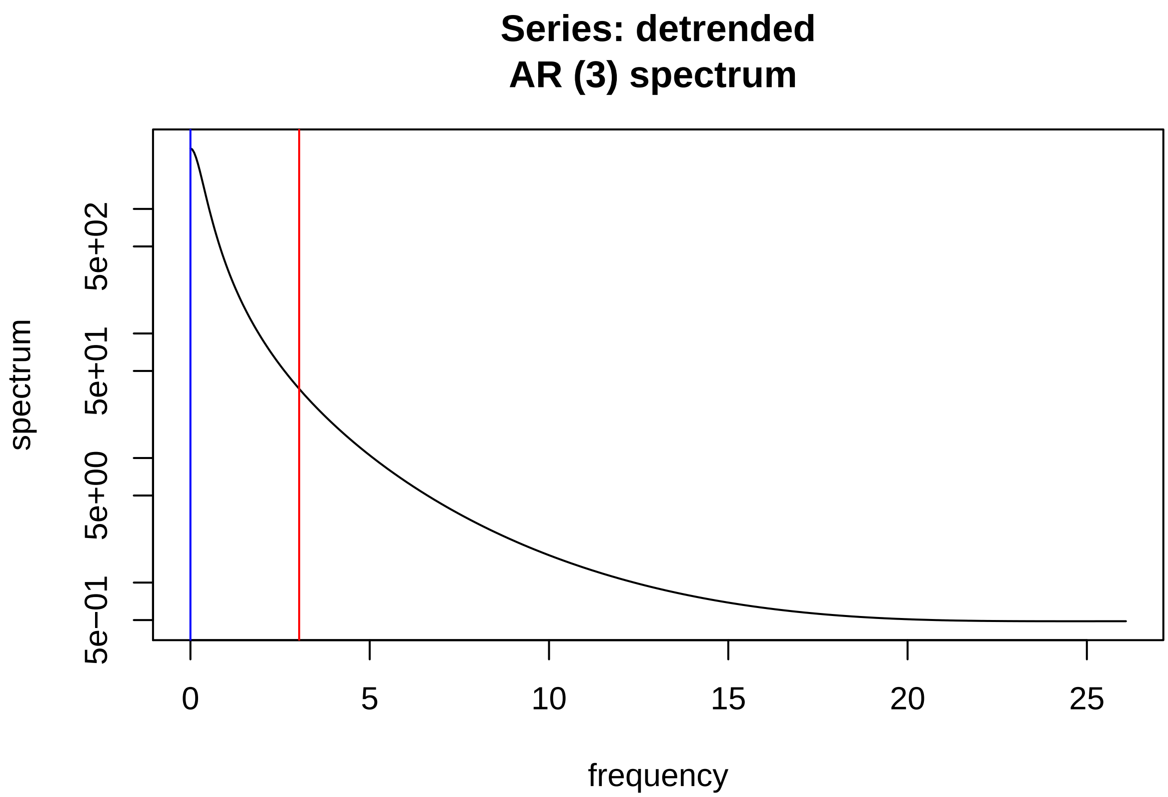
```
## [1] 4.13963
```

```
## [1] 1
```

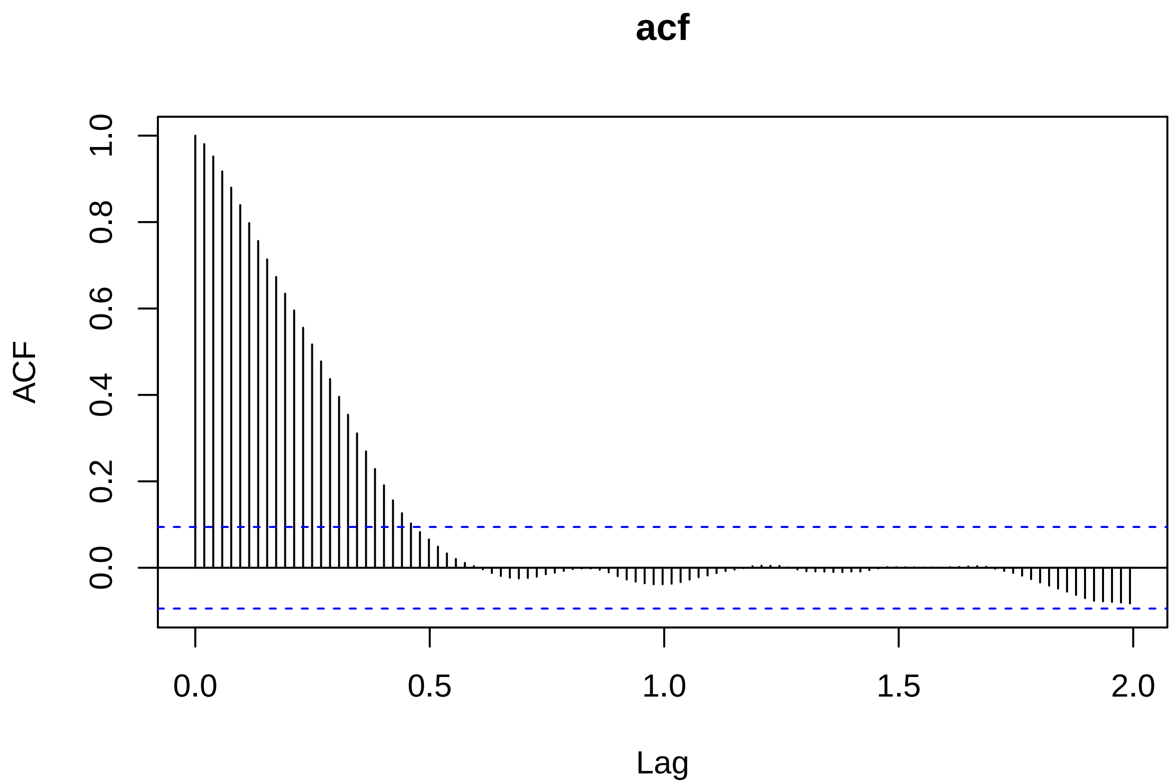



```
## [1] Inf
```

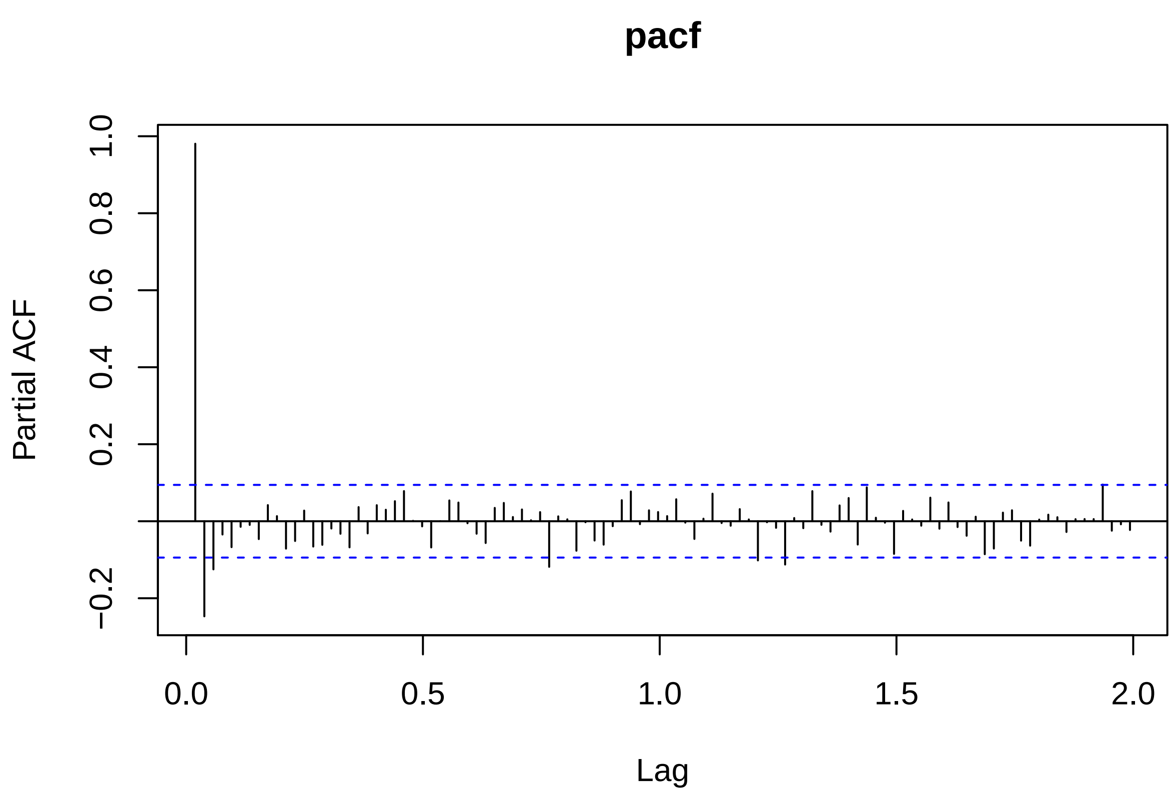
```
## [1] 0.3297694
```


Series: detrended
AR (3) spectrum
spectrum
5e−01
5e+00
5e+01
5e+02
0
5
10
15
20
25
frequency

9 ACF



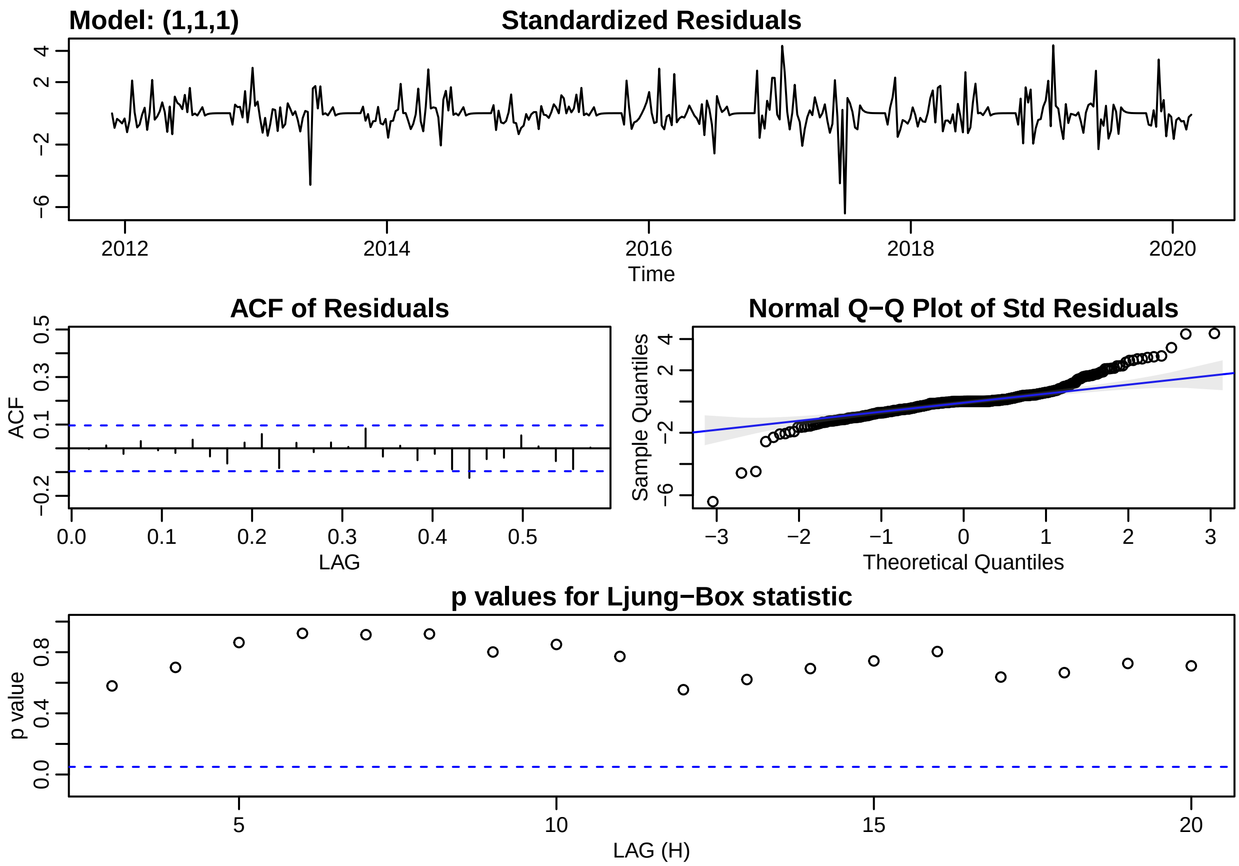
acf
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
1.5
2.0
Lag



pacf
Partial ACF
1.0
0.8
0.6
0.4
0.2
−0.2
0.0
0.5
1.0
1.5
2.0
Lag

10 Fit ARIMA

```
## initial  value 2.386026
## iter   2 value 2.364712
## iter   3 value 2.354355
## iter   4 value 2.353758
## iter   5 value 2.347833
## iter   6 value 2.344338
## iter   7 value 2.343570
## iter   8 value 2.342631
## iter   9 value 2.342502
## iter  10 value 2.342470
## iter  11 value 2.342466
## iter  12 value 2.342445
## iter  13 value 2.342444
## iter  14 value 2.342444
## iter  14 value 2.342444
## iter  14 value 2.342444
## final  value 2.342444
## converged
## initial  value 2.342904
## iter   2 value 2.342898
## iter   3 value 2.342893
## iter   4 value 2.342882
## iter   5 value 2.342877
## iter   6 value 2.342875
## iter   7 value 2.342875
## iter   8 value 2.342874
## iter   9 value 2.342874
## iter  10 value 2.342874
## iter  11 value 2.342873
## iter  12 value 2.342873
## iter  13 value 2.342873
## iter  13 value 2.342873
## iter  13 value 2.342873
## final  value 2.342873
## converged
```

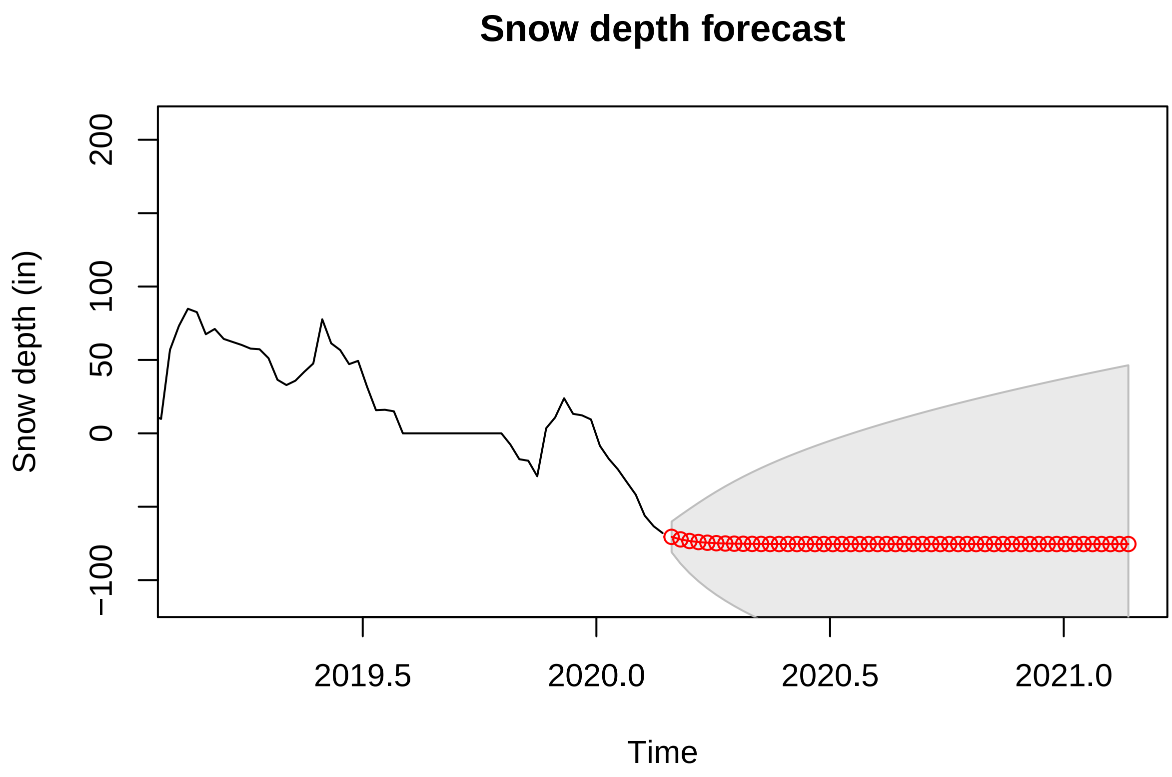


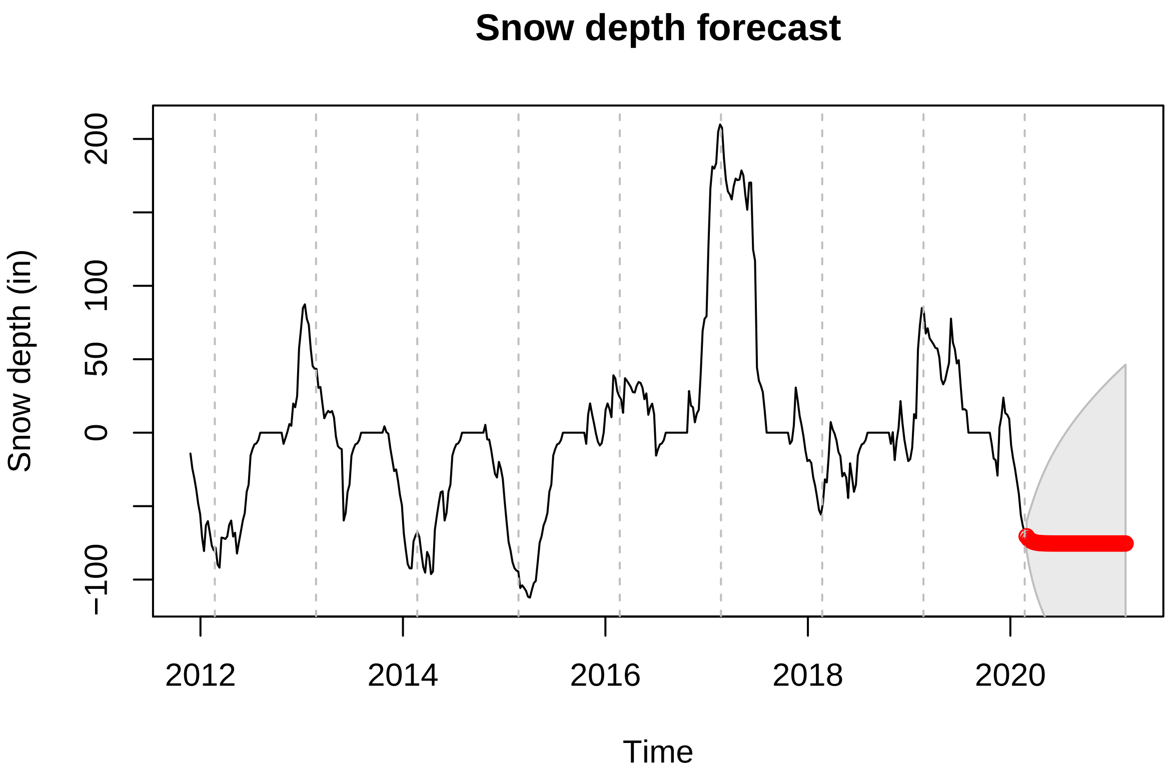

```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = constant, transform.pars = trans, fixed = fixed,
##     optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1      ma1  constant
##       0.6532  -0.4260   -0.1607
## s.e.  0.1126   0.1339    0.8289
##
## sigma^2 estimated as 108.4:  log likelihood = -1617.58,  aic = 3243.16
##
## $degrees_of_freedom
## [1] 427
##
## $ttable
##          Estimate     SE t.value p.value
## ar1        0.6532 0.1126  5.8016  0.0000
## ma1       -0.4260 0.1339 -3.1808  0.0016
## constant  -0.1607 0.8289 -0.1938  0.8464
##
## $AIC
## [1] 7.542228
##
```

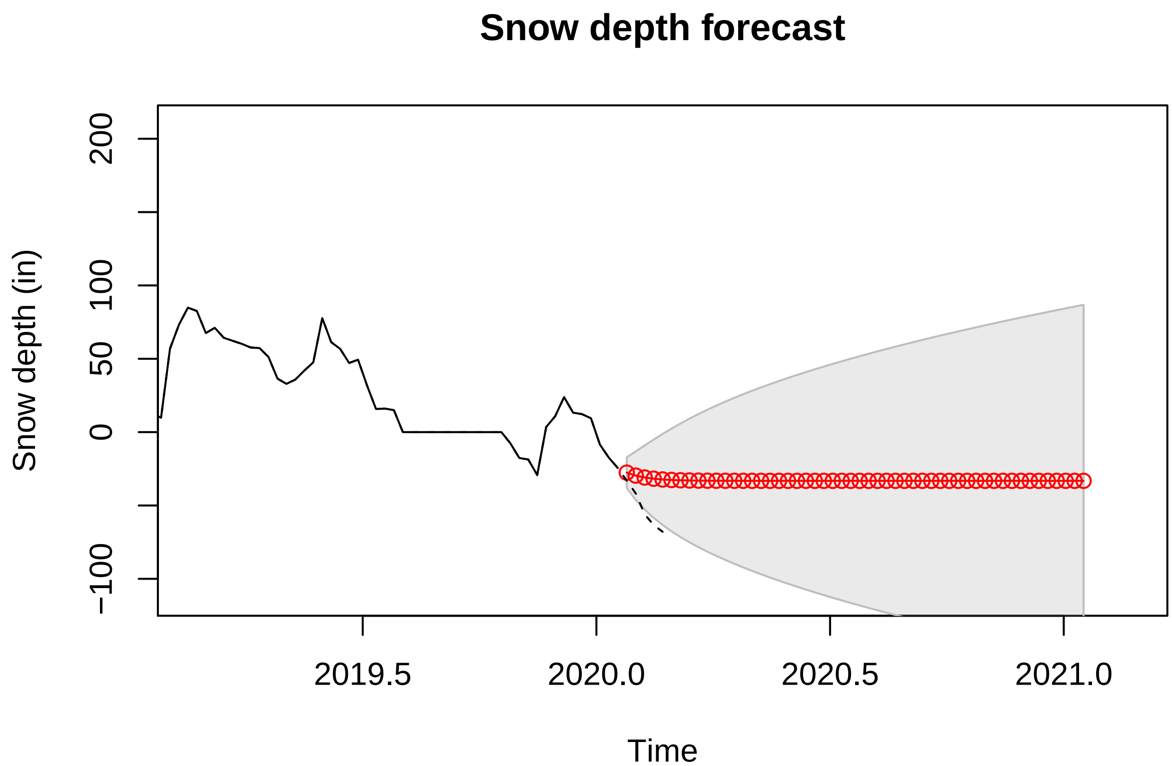
```
## $AICc
## [1] 7.542359
##
## $BIC
## [1] 7.580031
```

```
## [[1]]
##
## Call:
## arima(x = detrended, order = c(1, 1, 1))
##
## Coefficients:
##          ar1      ma1
##       0.6526  -0.4252
## s.e.  0.1127   0.1340
##
## sigma^2 estimated as 108.4:  log likelihood = -1617.6,  aic = 3241.2
```

```
## [[1]]
## [1] 3241.198
```




Snow depth forecast
Snow depth (in)
200
100
50
0
−100
2012
2014
2016
2018
2020
Time



Snow depth forecast
Snow depth (in)
200
100
50
0
−100
2019.5
2020.0
2020.5
2021.0
Time