Stat 261 Research Paper

Harrison Jia

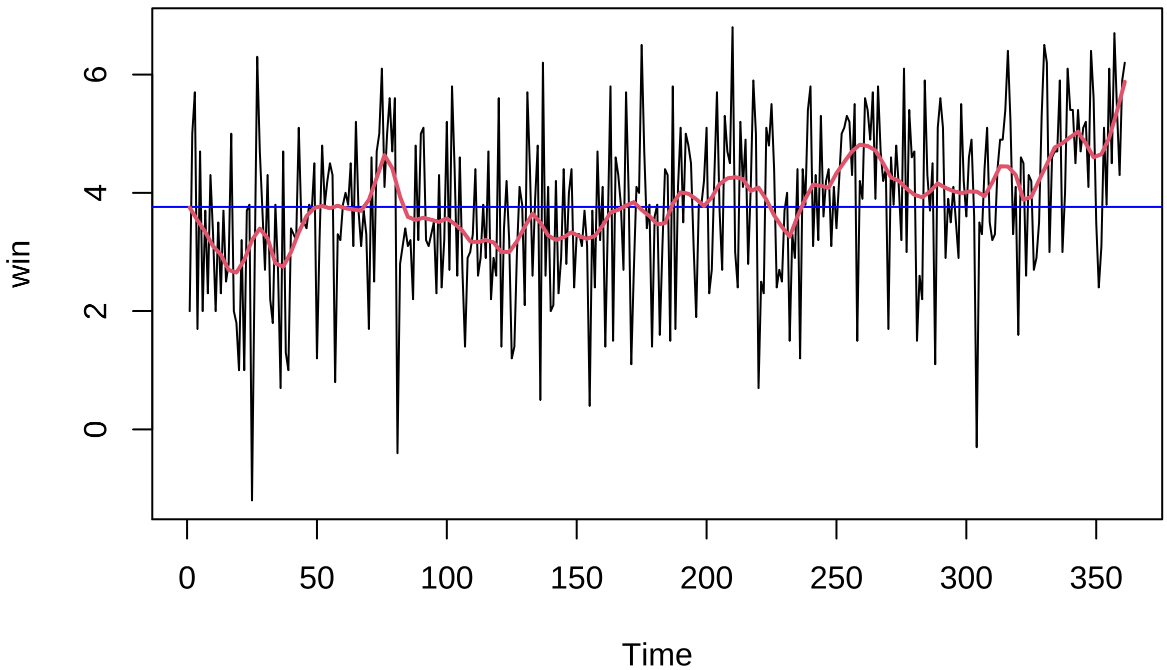
12/5/2020

Introduction

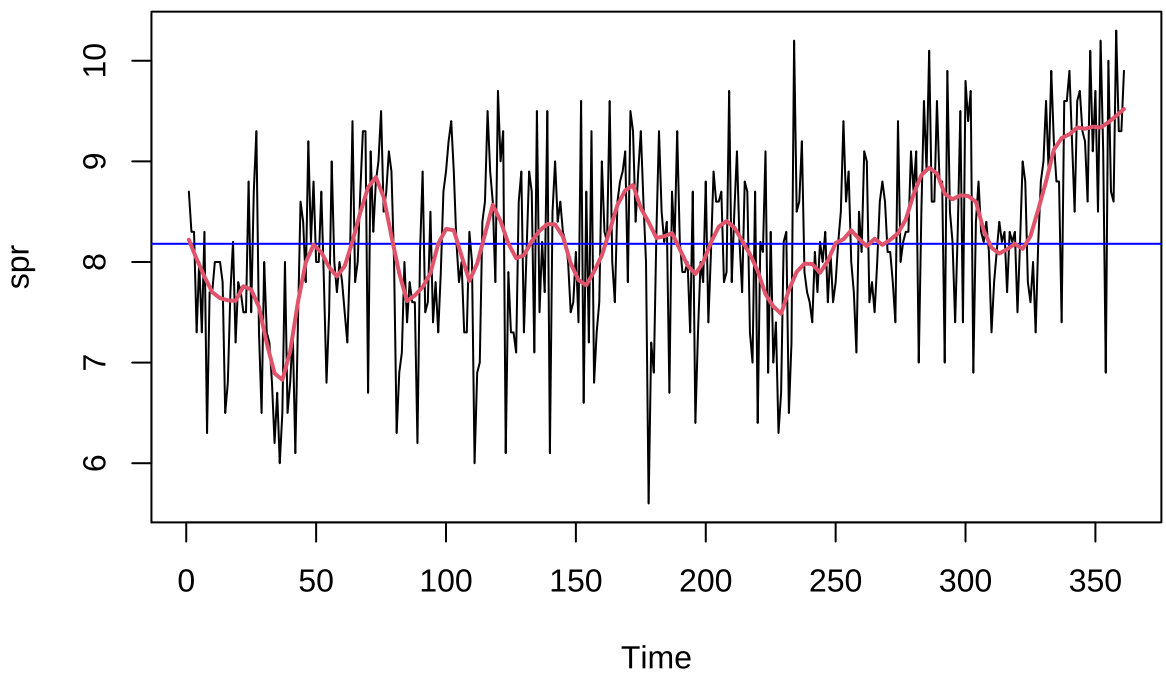
Over the course of time since the beginning of the industrial revolution, temperatures of the earth has largely been seen to increase at a steady pace (1). Scientists tend to attribute this to effects of this revolution and the emissions of various types of gas in large quantities. These large amounts of emissions are generally sought as a primary reason for the meteoric rise in global temperature. However, even such a well established theory has its own critics and many have been much more vocal since the relsease of new data from the NOAA and scientists there concluding that there global warming had no slowed down (2) - at which critics who have seen the data and compared them to predicted models claim there is a clear "slow down" relative to what was being predicted. While global warming as a concept in tandem the larger ideas of climate change is a tough one to tackle without accounting for a multitude of types of data, we will look to use data that that we do have in our disposal to further understand at the very least a localized sample of this trend or lack there of. To be more precise, we will use governmental data from the United Kingdom that tracks seasonal average temperature in Central England from 1659 to 2020. Perhaps, we may be able to see a trend that may help us answer the question at hand.

Analysis and Graphs

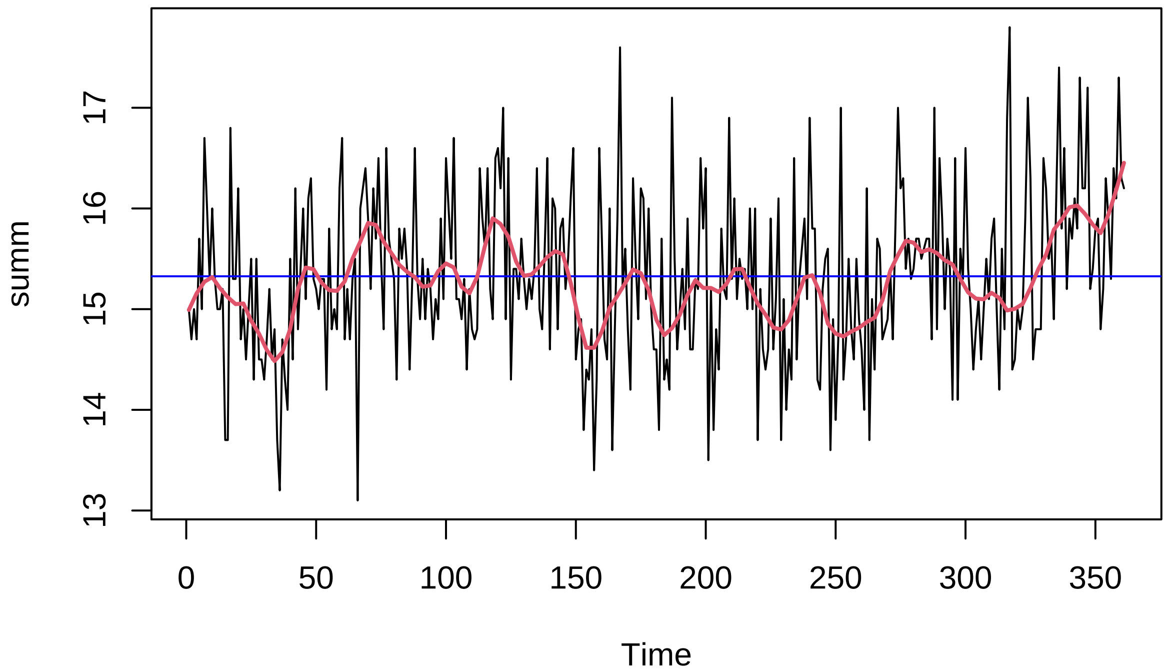
To begin our search, we will start by examining each of the four given time series given to us (each resembling a season). We will look for any patterns we may see in general with out eyes first as well as issues we may need to adjust for, such as heteroskedastic variance. Before we get started, a quick look at the table itself, we can notice that there is an obscenely large magnitude for the first winter and as such will choose to begin from the year 1660 for all seasons in order to keep the sample sizes consistent.



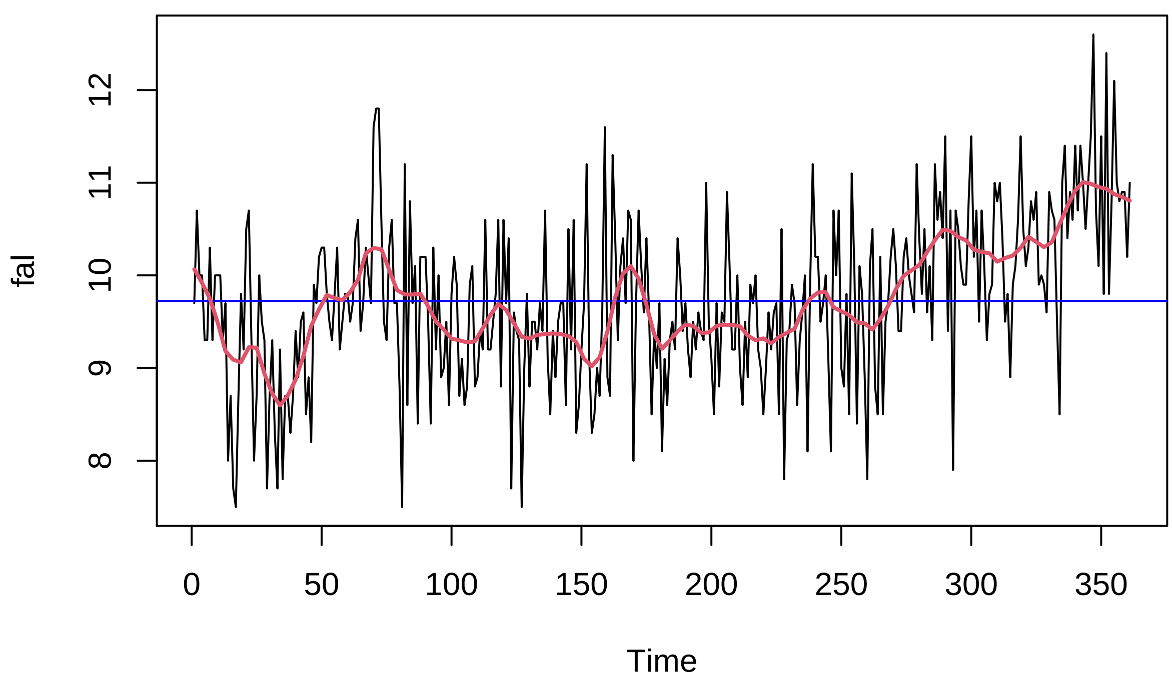
Here we have the mean of the time series for the winter months in blue. In addition, smoothed trend using lowess has been added to the graph with approximately decade pattern. Interresting to note that the smooothed line stays above the average for over the past 50 years and before that had been hovering around the mean.



This is the same thing but with the spring months. There is a much clear “jump” or rise towards the end tail of the data in the past 20-30 years.

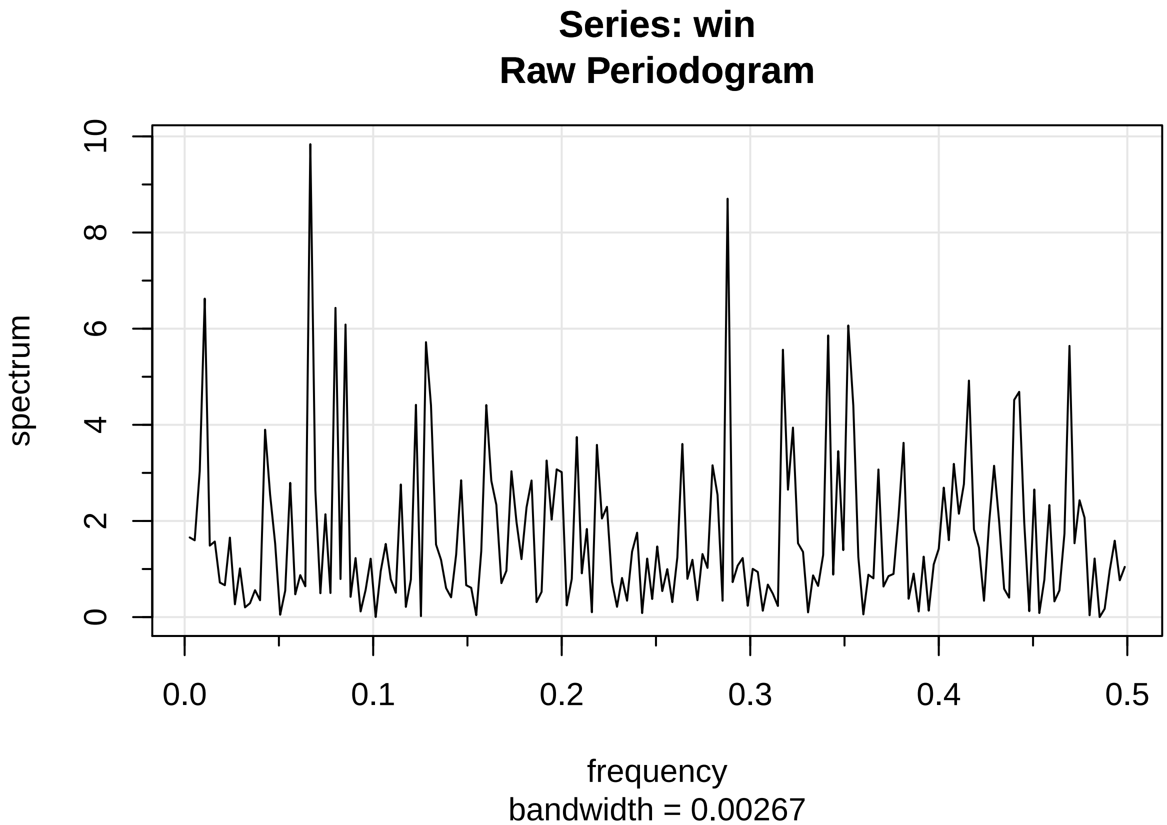


Summer Months, again we still see a jump and stagnant of the values near the end of the series; perhaps there is a raising of temperature pattern here?



Fall Months - same conclusions can be made but it is interesting to note that the values seem to be aove the curve for over 70 years here.

So far at a general glance of the time series we can see a general upwards trend within the last 20-25 years for all four months. In order to attempt to answer the question at hand, we will try to fit models on each of the four series with the last 20 years removed and then plot the perdiciton on top of the time series to see if there is any discrepancy from the perdiction model and its confidence interval. Meanwhile the rest of the time the data showed more of an oscillation around the mean. There seems to be alot of shifting up and down and so it may be hard to see anything clear in a periodogram with this data and may focus more on looking towards fitting time domain trends rather than frequency domain ones. We can see this in the following graph.

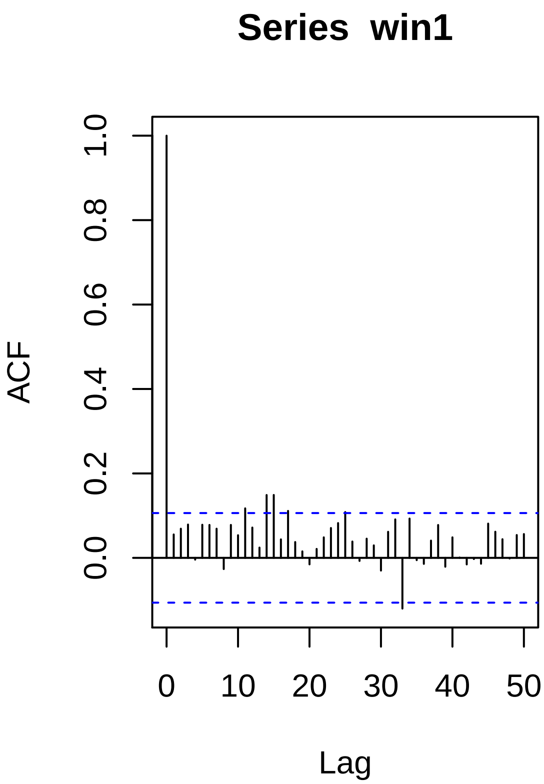


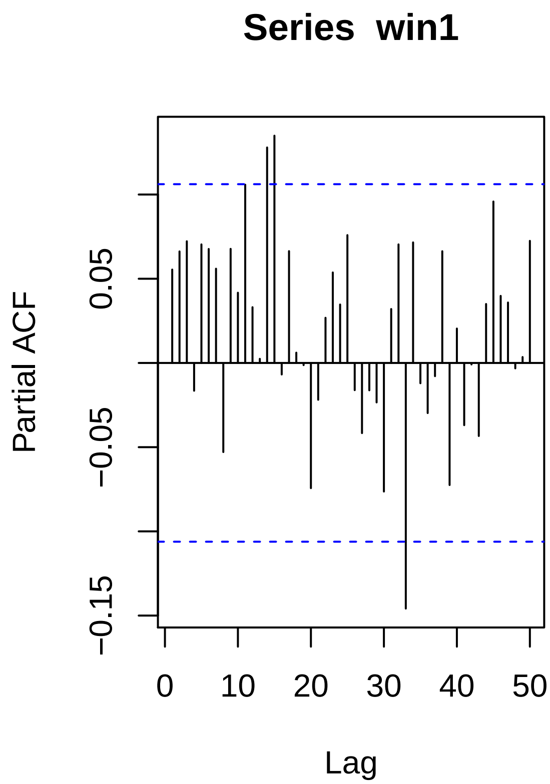

```
## [1] 0.06666667
```

```
## [1] 0.288
```

While there are a few that stand out, it is hard to say which ones we omit as well. This would tell us there are a lot of frequencies that we may need to account for. 1 big trend seems to be near 0.066..., or $\frac{1}{15}$ or 24 years or so and another large trend seems to be near 0.288 or about 12 years or so. Many articles talk about a decadal cycle of the ocean temerpature and currents as a big factor for changes in temperature. There may be some sort of thing similar going on here. This could also be from the activities of the sun. We won't dive any deeper on analysis of the frequencies for though the values of the biggest peak do seem like what was expected there are too much noise involved for us to fully get anywhere. Log transformations do not help this either.

Next we will examine the ACF and PACF of each to understand the models better.





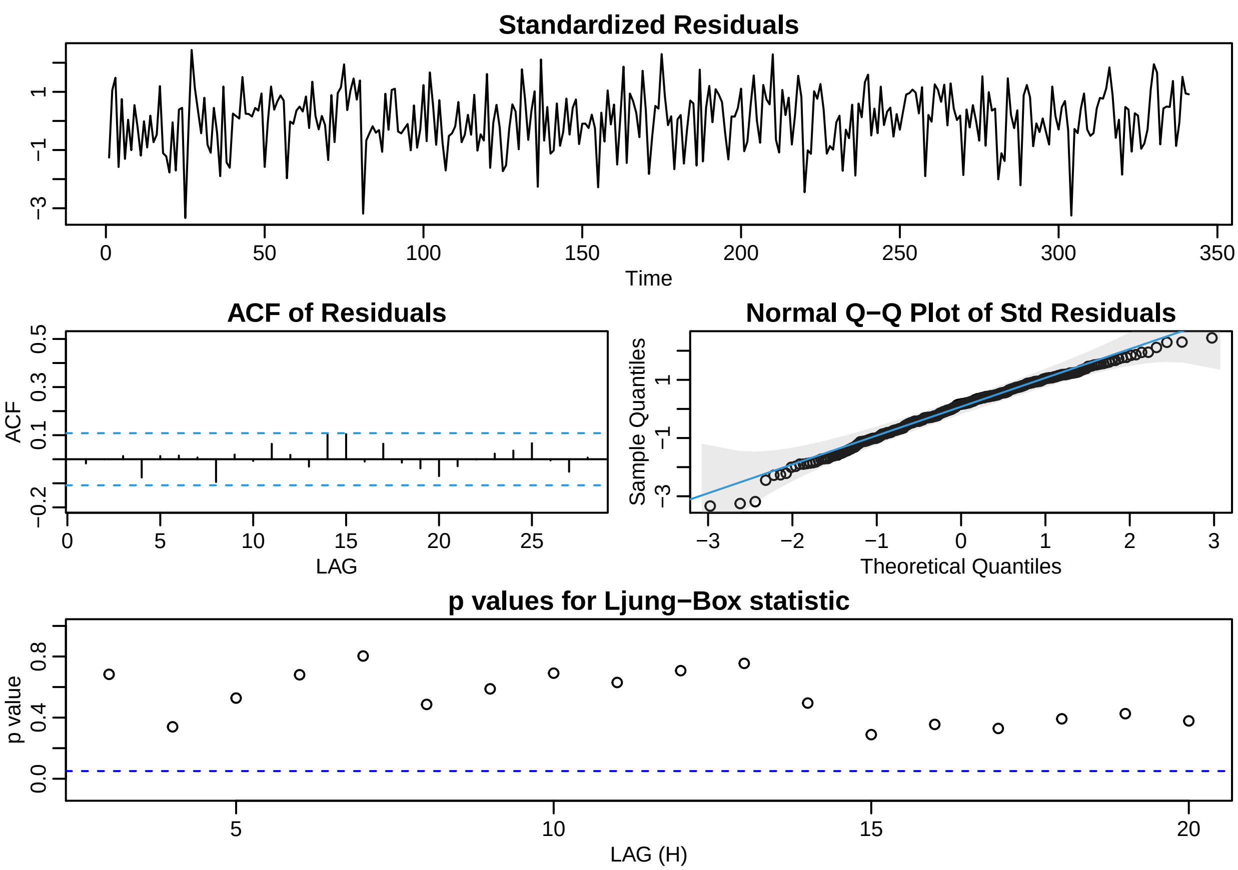
When looking at the ACF and the PACF, we cannot really see much of a semblence of either an AR or MA process. ACF drops after 0 while so does PACF in a sense. This is beginning to look like a white noise model of $x_t = w_t$

```
##
## Call:
## arima(x = win1, order = c(1, 0, 1))
##
## Coefficients:
##          ar1      ma1  intercept
##       0.9927  -0.9614     3.7226
## s.e.  0.0113   0.0242     0.2957
##
## sigma^2 estimated as 1.752:  log likelihood = -579.82,  aic = 1167.63
##
## Training set error measures:
##                      ME     RMSE      MAE       MPE     MAPE     MASE
## Training set 0.04707037 1.323751 1.063314 -10.48516 49.35178 0.723922
##                     ACF1
## Training set -0.01735659
```

```
## initial  value 0.301095
## iter   2 value 0.300053
## iter   3 value 0.299536
## iter   4 value 0.299535
## iter   5 value 0.299533
## iter   5 value 0.299533
## iter   5 value 0.299533
## final  value 0.299533
## converged
```

```
## initial  value 0.300489
## iter   2 value 0.300480
## iter   3 value 0.300463
## iter   4 value 0.300419
## iter   5 value 0.300391
## iter   6 value 0.300193
## iter   7 value 0.300184
## iter   8 value 0.299727
## iter   9 value 0.299545
## iter  10 value 0.299309
## iter  11 value 0.299186
## iter  12 value 0.299165
## iter  13 value 0.298945
## iter  14 value 0.298164
## iter  15 value 0.297954
## iter  16 value 0.297664
## iter  17 value 0.297209
## iter  18 value 0.296887
## iter  19 value 0.296842
## iter  20 value 0.294940
## iter  21 value 0.294818
## iter  22 value 0.294122
## iter  23 value 0.293873
## iter  24 value 0.292544
## iter  25 value 0.287245
## iter  26 value 0.285951
## iter  27 value 0.285367
## iter  28 value 0.285204
## iter  29 value 0.284896
## iter  30 value 0.284024
## iter  31 value 0.283290
## iter  32 value 0.281955
## iter  33 value 0.281817
## iter  34 value 0.281784
## iter  35 value 0.281782
## iter  36 value 0.281769
## iter  37 value 0.281548
## iter  38 value 0.281516
## iter  39 value 0.281505
## iter  40 value 0.281499
## iter  41 value 0.281487
## iter  42 value 0.281486
## iter  43 value 0.281482
## iter  44 value 0.281471
## iter  45 value 0.281450
## iter  46 value 0.281425
## iter  47 value 0.281411
## iter  48 value 0.281409
## iter  49 value 0.281409
## iter  50 value 0.281408
## iter  51 value 0.281408
## iter  52 value 0.281408
## iter  53 value 0.281407
## iter  54 value 0.281405
```

```
## iter  55 value 0.281404
## iter  56 value 0.281404
## iter  57 value 0.281404
## iter  58 value 0.281404
## iter  59 value 0.281404
## iter  60 value 0.281403
## iter  61 value 0.281403
## iter  62 value 0.281403
## iter  63 value 0.281403
## iter  64 value 0.281403
## iter  65 value 0.281403
## iter  66 value 0.281403
## iter  67 value 0.281403
## iter  68 value 0.281402
## iter  69 value 0.281402
## iter  70 value 0.281402
## iter  71 value 0.281402
## iter  71 value 0.281402
## iter  71 value 0.281402
## final  value 0.281402
## converged
```

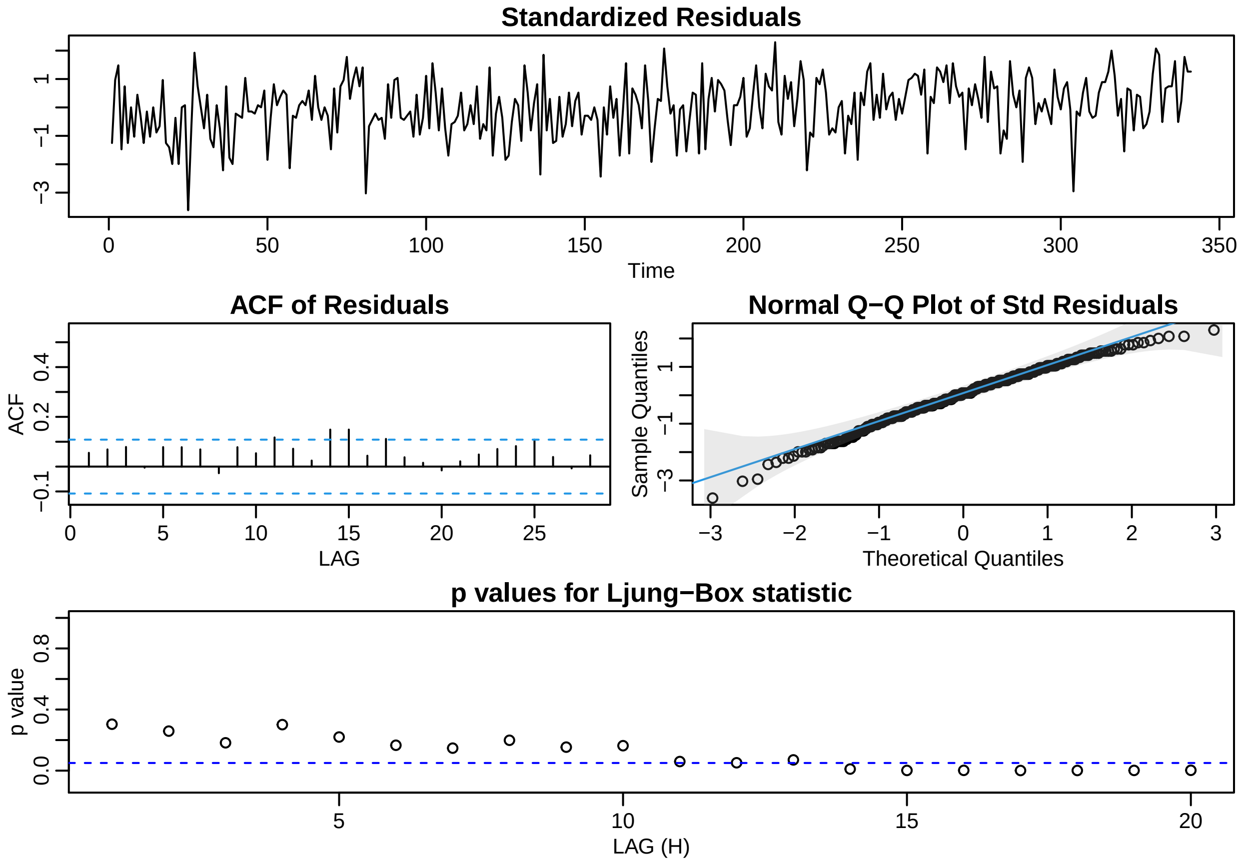


Looking at the AMRA(1,1) model it seems to confirm my suspision and worries. Perhaps the model is with the coefficients as estimated, but it is important to note that both values are close to each other in that perhaps the model was a redundant one: $(1-\phi B)x_t = (1-\phi B)w_t$ and we may cancel them out to get back to what was said before. It is hard to think of another method to do here as something like differencing may overdo the model and reintroduce an MA(1) model that would create dependence that does not exactly exist in the model currently. The plot of the residuals seem okay however, they fit normality quite well.

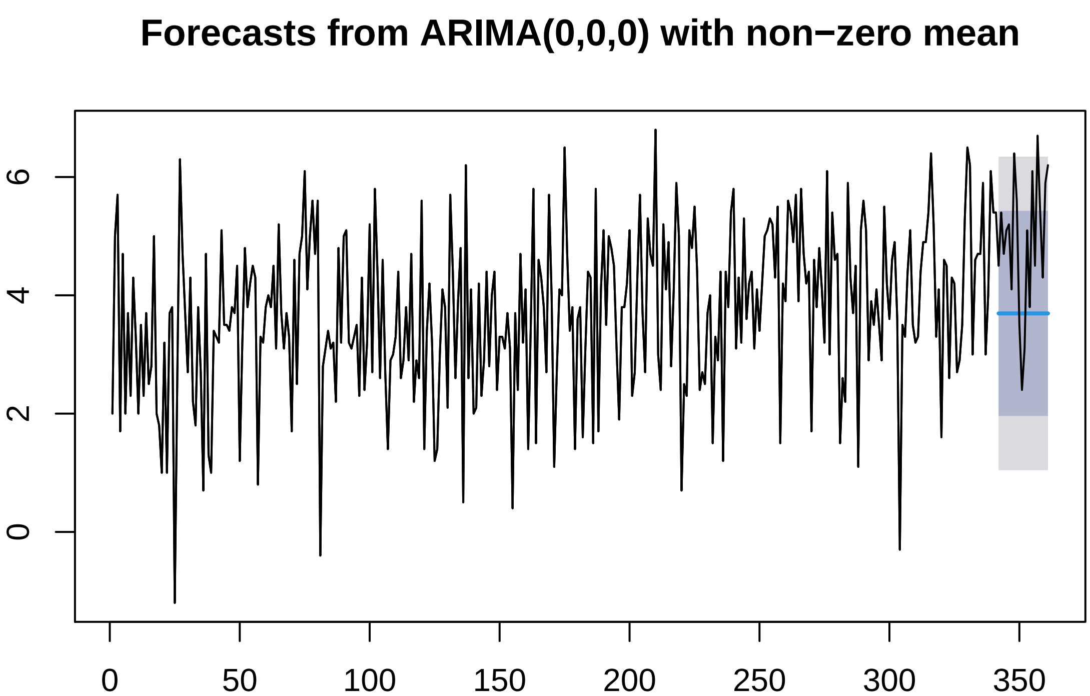
Instead we could look to examine an ARMA(0,0) model or the white noise model onto the winter time series.

```
##
## Call:
## arima(x = win1, order = c(0, 0, 0))
##
## Coefficients:
##       intercept
##          3.6941
## s.e.     0.0732
##
## sigma^2 estimated as 1.829:  log likelihood = -586.82,  aic = 1177.63
##
## Training set error measures:
##                         ME    RMSE      MAE       MPE     MAPE      MASE
## Training set -1.521718e-15 1.35247 1.081437 -13.34813 50.67545 0.7362608
##                    ACF1
## Training set 0.0554362
```

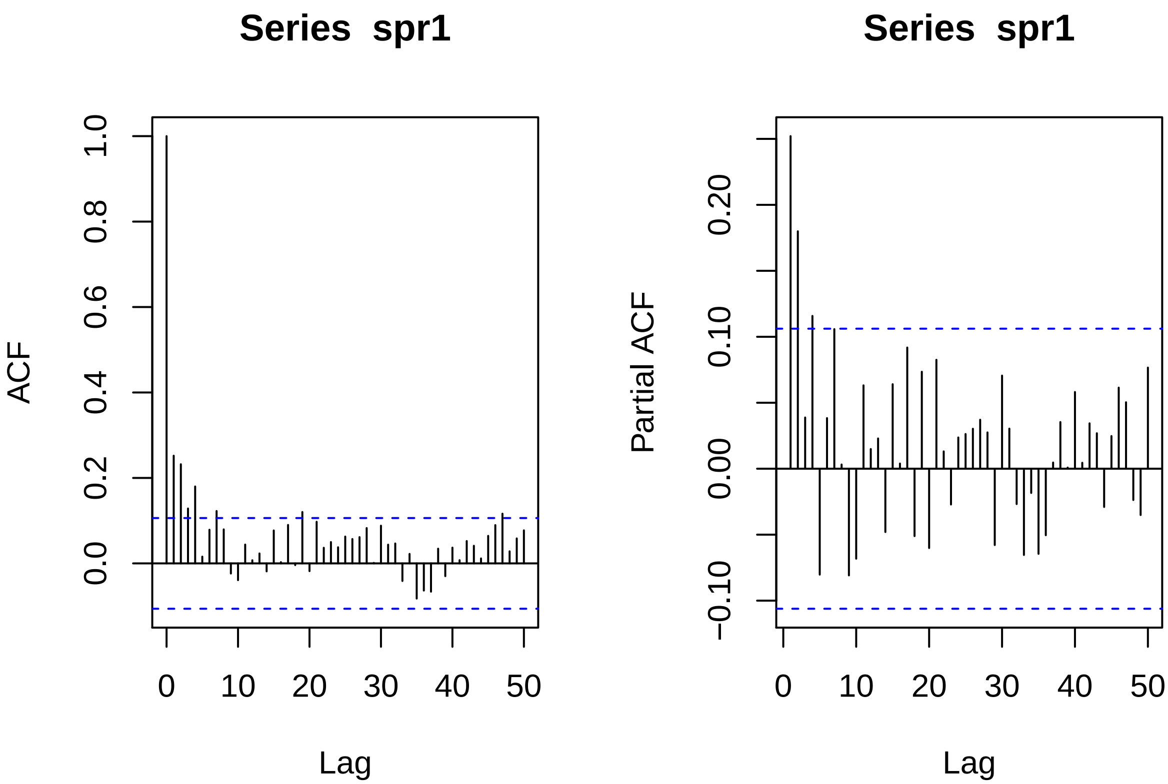
```
## initial  value 0.301932
## iter   1 value 0.301932
## final  value 0.301932
## converged
## initial  value 0.301932
## iter   1 value 0.301932
## final  value 0.301932
## converged
```



This model seems to fit well. We have low ME as well. The ACF of residuals look good but there may be some normality voilation as the Q-Q plot seems to be leaning towards the right - but this is still very close to normal still. Although the forecast of this type of model is simple to derive, we still still plot it and compare to what the actual values give us.



It seems as though again the predicted values are underestimating the true values a bit, but not much of it is too far outside of the confidence interval for statistical consern.

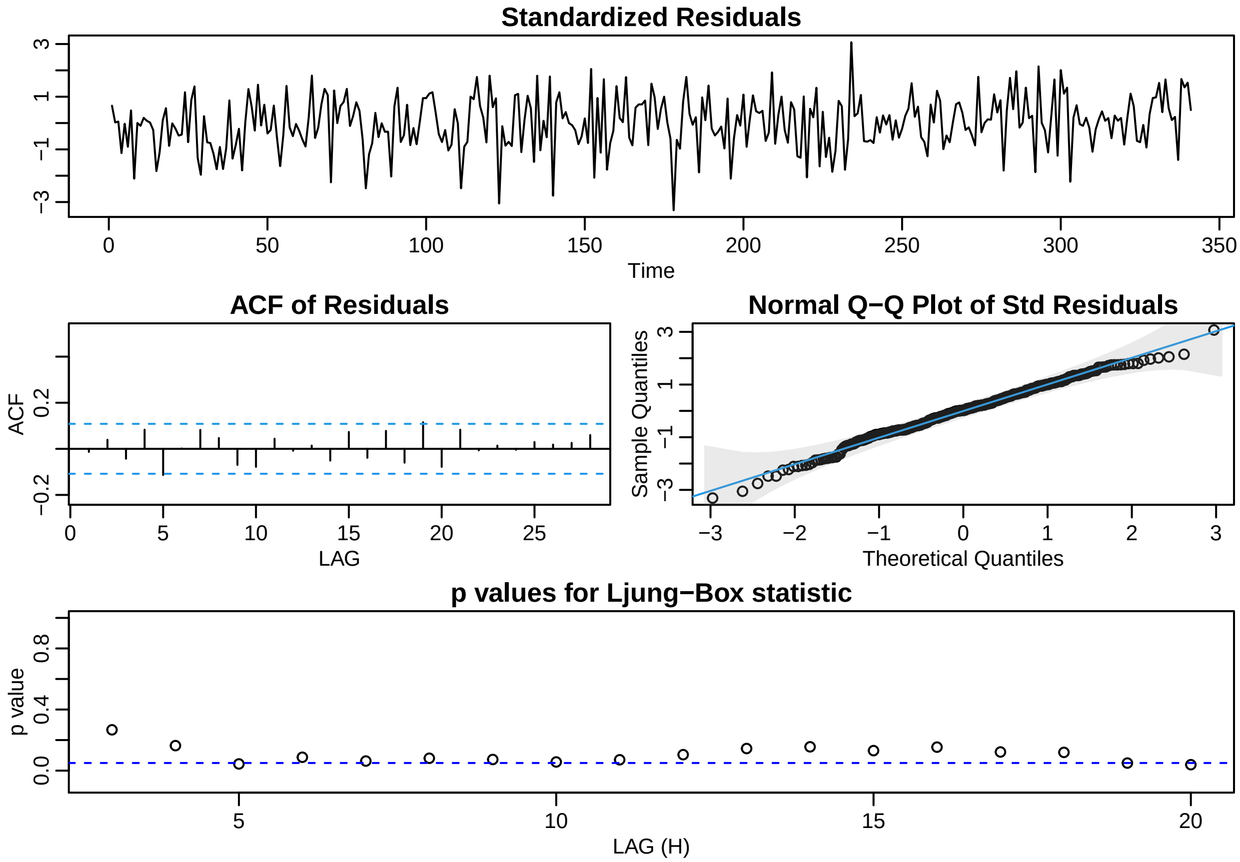


Here for spring, we can see that both ACF and PACF tend to trail off and thus it would seem that and ARMA model would be a good choice to examine for this time series.

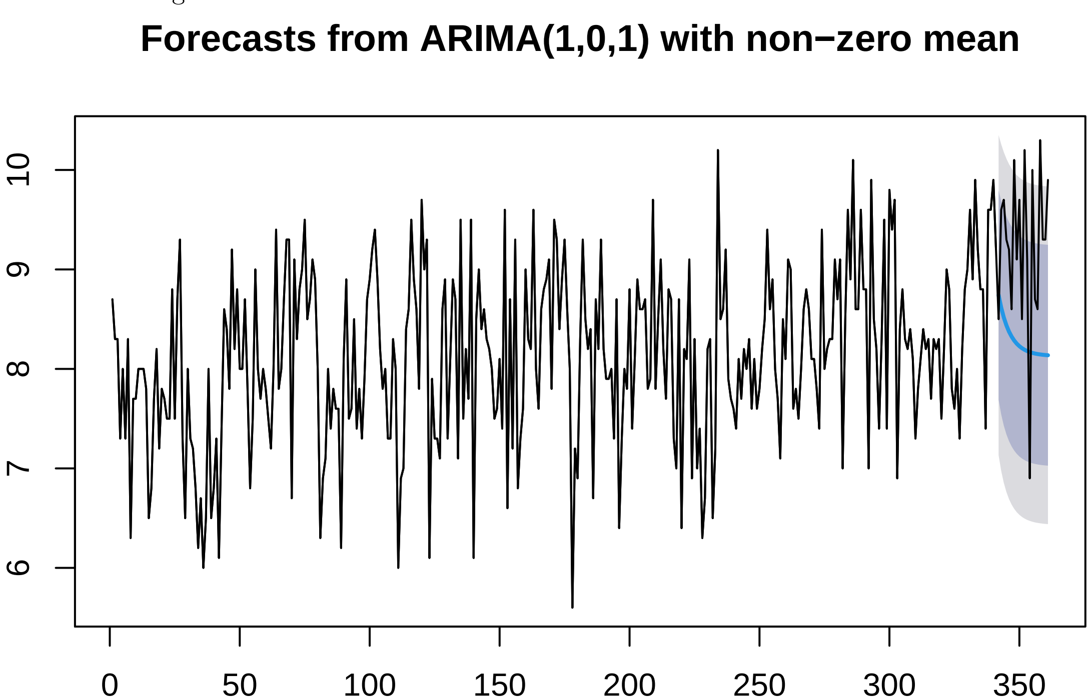
```
##
## Call:
## arima(x = spr1, order = c(1, 0, 1))
##
## Coefficients:
##          ar1      ma1  intercept
##       0.7963  -0.5921     8.1294
## s.e.  0.0855   0.1125     0.0885
##
## sigma^2 estimated as 0.6733:  log likelihood = -416.49,  aic = 840.99
##
## Training set error measures:
##                         ME      RMSE       MAE       MPE    MAPE      MASE
## Training set -0.0009566764 0.8205434 0.6459954 -1.093845 8.20676 0.7972357
##                     ACF1
## Training set -0.01326129
```

```
## initial  value -0.144361
## iter   2 value -0.164076
## iter   3 value -0.173090
## iter   4 value -0.174093
## iter   5 value -0.189357
## iter   6 value -0.192677
## iter   7 value -0.195270
## iter   8 value -0.196080
```

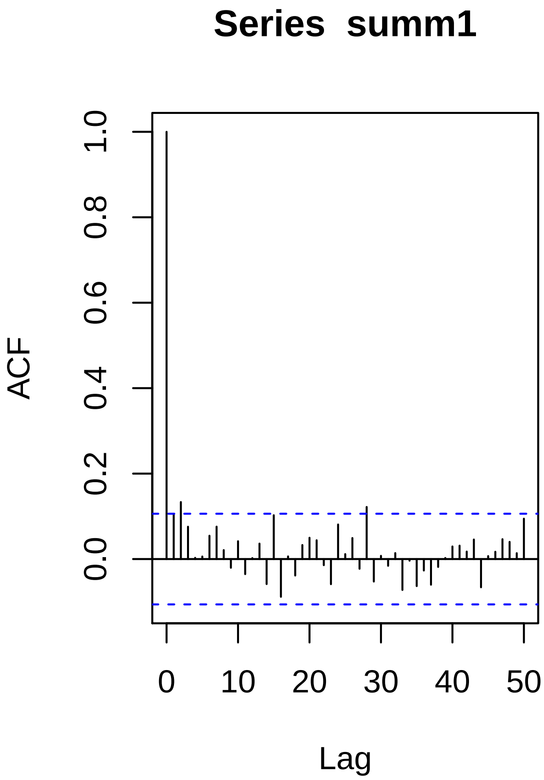
```
## iter   9 value -0.196098
## iter  10 value -0.196109
## iter  11 value -0.196131
## iter  12 value -0.196135
## iter  13 value -0.196136
## iter  14 value -0.196136
## iter  15 value -0.196137
## iter  16 value -0.196137
## iter  17 value -0.196137
## iter  18 value -0.196137
## iter  19 value -0.196137
## iter  20 value -0.196137
## iter  21 value -0.196137
## iter  22 value -0.196137
## iter  22 value -0.196137
## final  value -0.196137
## converged
## initial  value -0.197491
## iter   2 value -0.197499
## iter   3 value -0.197508
## iter   4 value -0.197517
## iter   5 value -0.197536
## iter   6 value -0.197547
## iter   7 value -0.197550
## iter   8 value -0.197550
## iter   9 value -0.197550
## iter  10 value -0.197551
## iter  11 value -0.197552
## iter  12 value -0.197552
## iter  13 value -0.197552
## iter  14 value -0.197552
## iter  15 value -0.197552
## iter  15 value -0.197552
## final  value -0.197552
## converged
```

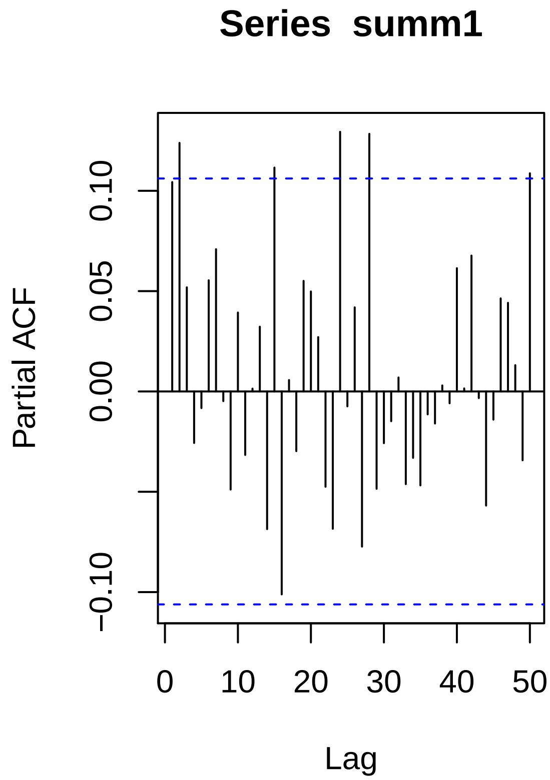


The coefficients here is much better compared to the winter series. The Q-Q plot of residuals seem to how normality of the residuals and the ACF shows no apparent devation from the model assumptions and thus seems to be a good fit for this series.



The predicted model and the actual values seem to deviate more and more as time goes on. There seems to be a underestimation of the true values in general indicating perhaps there may be some warming but until towards the final few years, majority of the values are within the confidence interval.



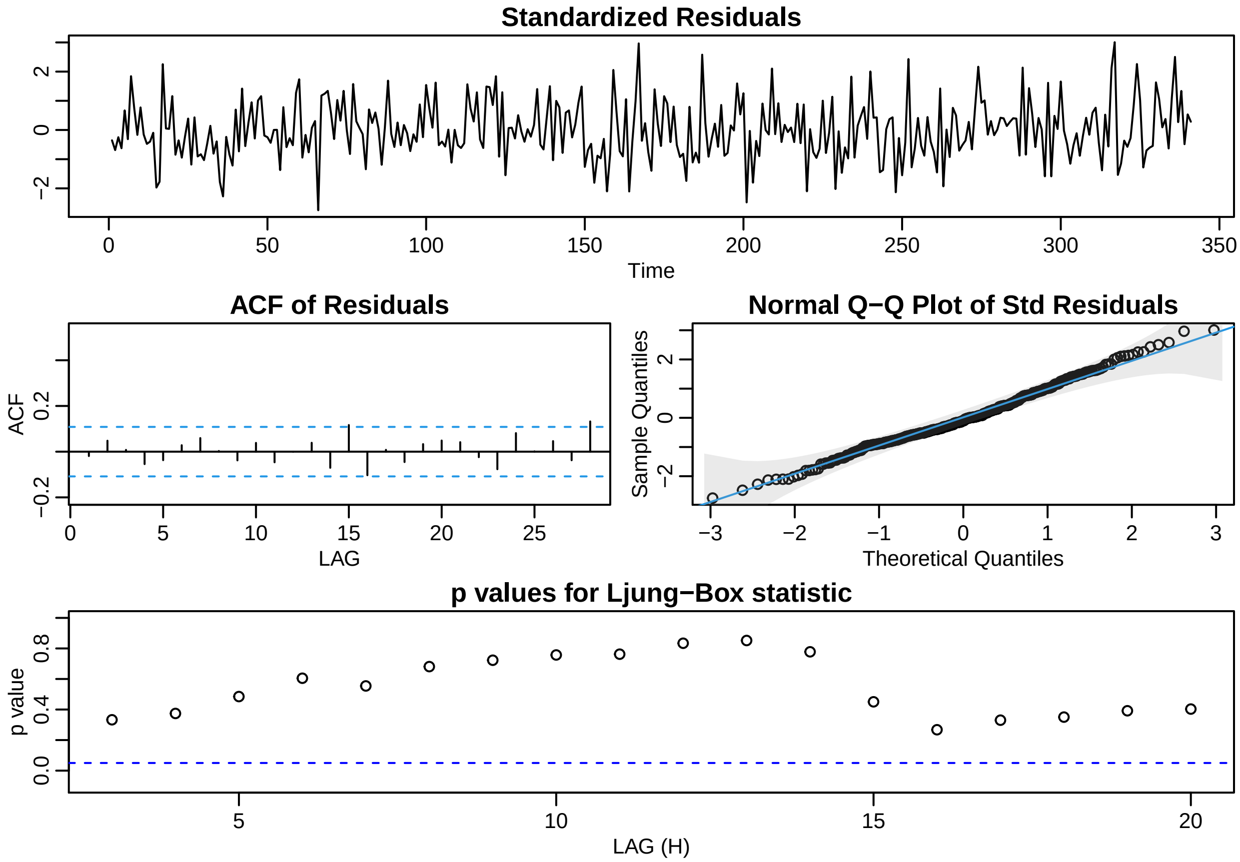


For the summer time series, we may have a similar issue as we saw with winter but there seems to be not as severe. We should look out for redundency if there is any and take action immediately.

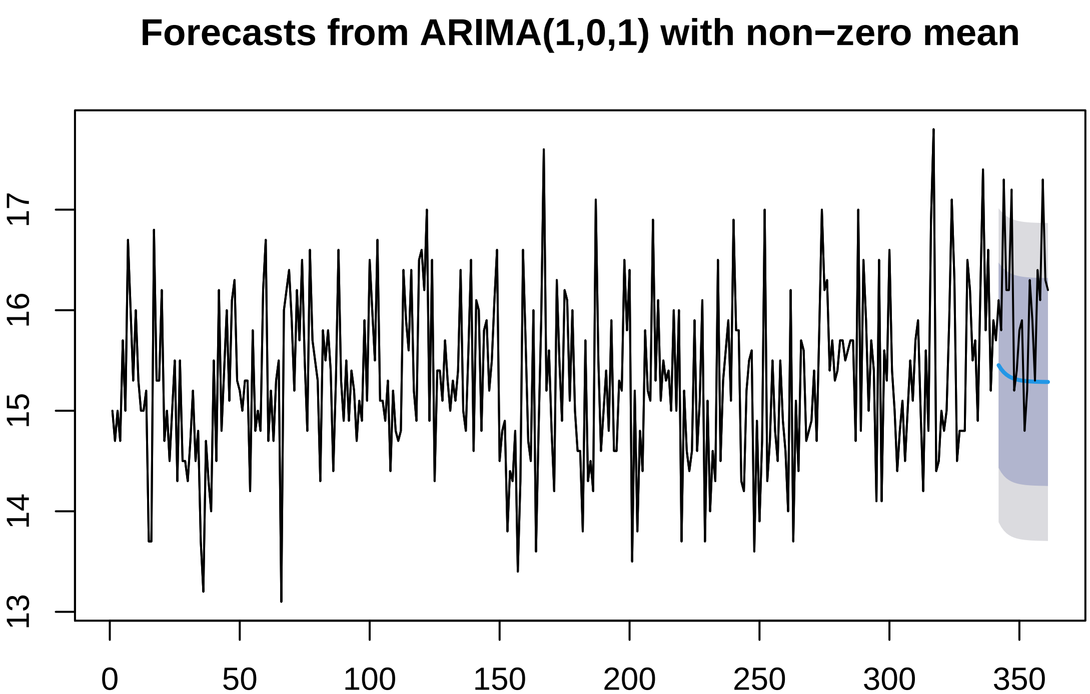
```
##
## Call:
## arima(x = summ1, order = c(1, 0, 1))
##
## Coefficients:
##          ar1      ma1  intercept
##       0.7626  -0.6585    15.2853
## s.e.  0.1677   0.1946     0.0618
##
## sigma^2 estimated as 0.6342:  log likelihood = -406.25,  aic = 820.5
##
## Training set error measures:
##                        ME      RMSE       MAE        MPE     MAPE      MASE
## Training set 0.0007814496 0.7963971 0.6285606 -0.2655233 4.113369 0.7382059
##                     ACF1
## Training set -0.01911802
```

```
## initial  value -0.213599
## iter   2 value -0.216963
## iter   3 value -0.218522
## iter   4 value -0.218601
## iter   5 value -0.221790
## iter   6 value -0.222708
## iter   7 value -0.224767
```

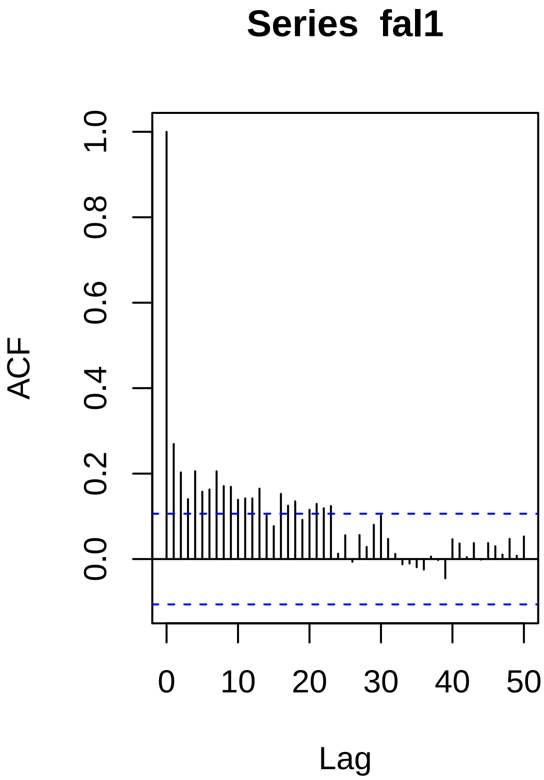
```
## iter   8 value -0.225654
## iter   9 value -0.225823
## iter  10 value -0.225924
## iter  11 value -0.226028
## iter  12 value -0.226274
## iter  13 value -0.226494
## iter  14 value -0.226586
## iter  15 value -0.226633
## iter  16 value -0.226637
## iter  17 value -0.226684
## iter  18 value -0.226717
## iter  19 value -0.226720
## iter  20 value -0.226721
## iter  21 value -0.226722
## iter  22 value -0.226723
## iter  23 value -0.226724
## iter  24 value -0.226724
## iter  25 value -0.226724
## iter  26 value -0.226725
## iter  27 value -0.226725
## iter  28 value -0.226725
## iter  29 value -0.226725
## iter  30 value -0.226725
## iter  31 value -0.226725
## iter  32 value -0.226725
## iter  33 value -0.226725
## iter  33 value -0.226725
## iter  33 value -0.226725
## final  value -0.226725
## converged
## initial  value -0.227578
## iter   2 value -0.227581
## iter   3 value -0.227589
## iter   4 value -0.227589
## iter   5 value -0.227589
## iter   6 value -0.227591
## iter   7 value -0.227591
## iter   8 value -0.227592
## iter   9 value -0.227592
## iter  10 value -0.227592
## iter  11 value -0.227592
## iter  11 value -0.227592
## iter  11 value -0.227592
## final  value -0.227592
## converged
```

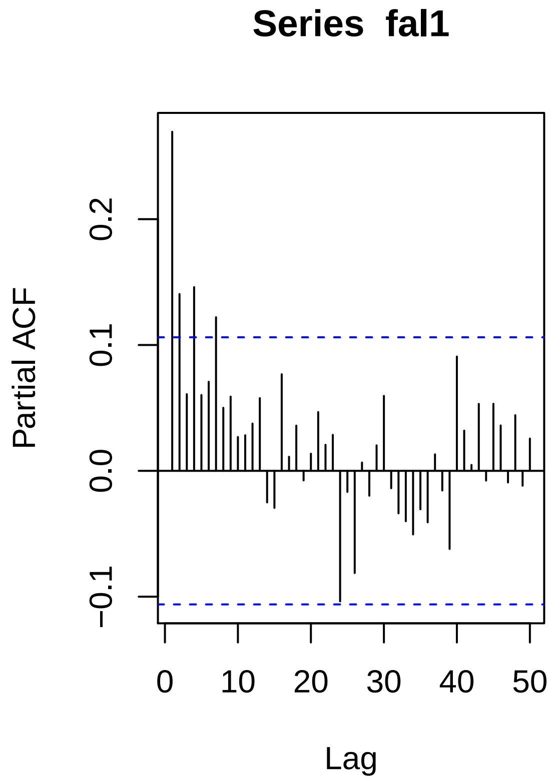


Values from the ARMA(1,1) again seem to fit well, more importantly the coefficients of AR and MA do seem different/further apart so redundency seems to be not an issue here.



Overall, the summer time series as a similar ACF and PACF as spring and thus we can fit an ARMA(1,1) as well. The fitted values seem to be good in that the values dont seem to indicate any redundency as it did in the winter series. Fitting the model prediciting 20 years forward, we can see that unlike the spring series, that the actual values fit very better with the predicted ones.

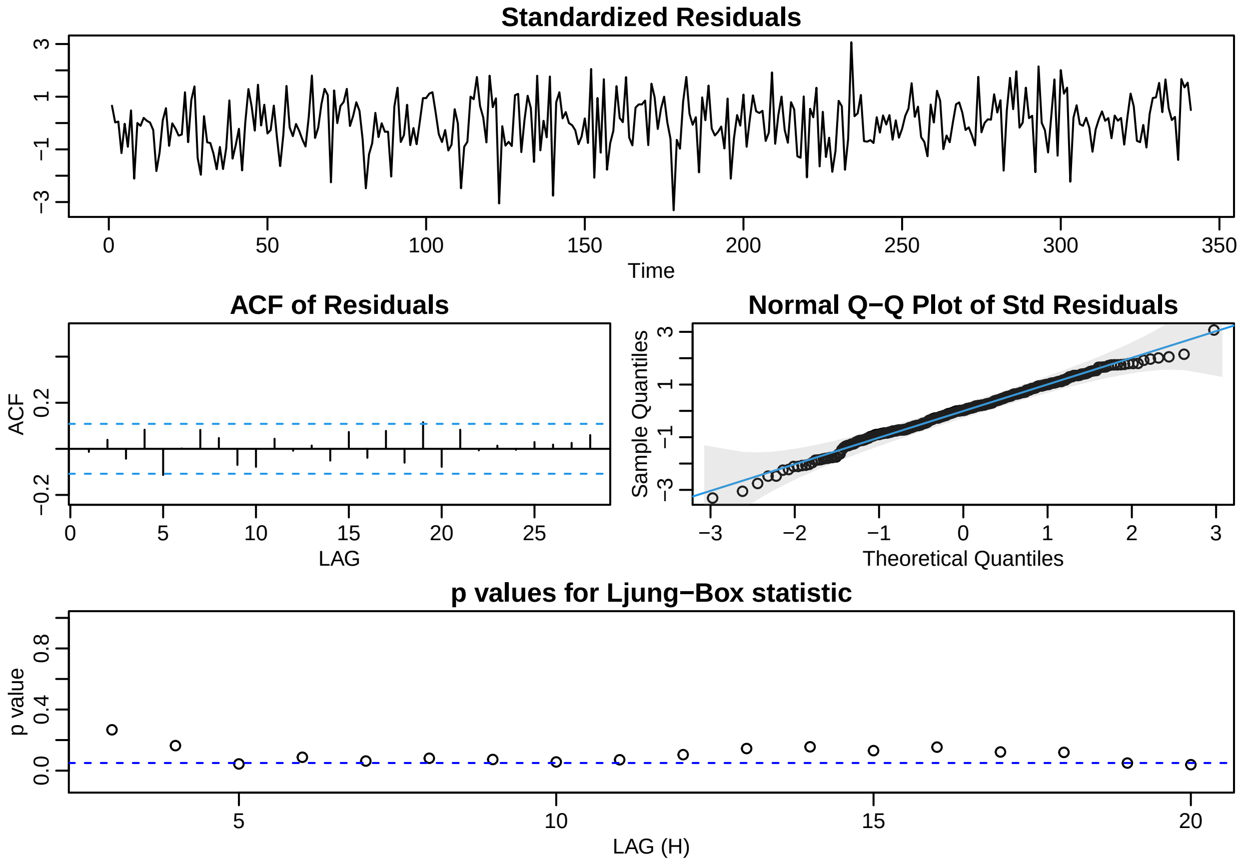


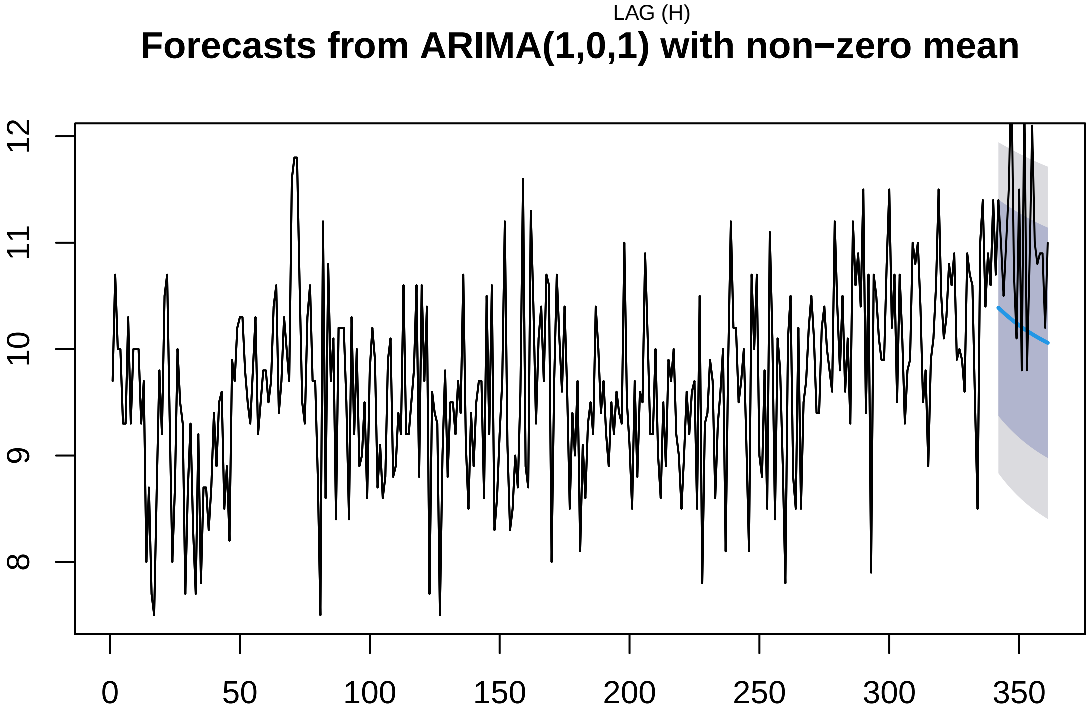



```
##
## Call:
## arima(x = fal1, order = c(1, 0, 1))
##
## Coefficients:
##          ar1      ma1  intercept
##       0.9659  -0.8553     9.7071
## s.e.  0.0242   0.0469     0.1752
##
## sigma^2 estimated as 0.6293:  log likelihood = -405.16,  aic = 818.31
##
## Training set error measures:
##                        ME      RMSE       MAE        MPE     MAPE      MASE
## Training set -0.001806253 0.7932953 0.6213771 -0.7210812 6.573969 0.7682481
##                    ACF1
## Training set 0.05749555
```

```
## initial  value -0.144361
## iter   2 value -0.164076
## iter   3 value -0.173090
## iter   4 value -0.174093
## iter   5 value -0.189357
## iter   6 value -0.192677
## iter   7 value -0.195270
## iter   8 value -0.196080
## iter   9 value -0.196098
## iter  10 value -0.196109
```

```
## iter  11 value -0.196131
## iter  12 value -0.196135
## iter  13 value -0.196136
## iter  14 value -0.196136
## iter  15 value -0.196137
## iter  16 value -0.196137
## iter  17 value -0.196137
## iter  18 value -0.196137
## iter  19 value -0.196137
## iter  20 value -0.196137
## iter  21 value -0.196137
## iter  22 value -0.196137
## iter  22 value -0.196137
## final  value -0.196137
## converged
## initial  value -0.197491
## iter   2 value -0.197499
## iter   3 value -0.197508
## iter   4 value -0.197517
## iter   5 value -0.197536
## iter   6 value -0.197547
## iter   7 value -0.197550
## iter   8 value -0.197550
## iter   9 value -0.197550
## iter  10 value -0.197551
## iter  11 value -0.197552
## iter  12 value -0.197552
## iter  13 value -0.197552
## iter  14 value -0.197552
## iter  15 value -0.197552
## iter  15 value -0.197552
## final  value -0.197552
## converged
```





Values again look good within the fitted ARMA(1,1) model, simiarly with normality within the residuals and the ACF of residuals look good as well. Model seemed to be good here. The forecasted model is slightly underestimating the actual values but majority of the values are within the confidence interval.

Conclusion

Overall, we can see that for most of the time series AMRA(1,1) model fit the series well. In addition, we can see from our forecast into the 20 years that the actual values tend to be within the confidence interval of the predicted. That being said the values tend towards a slight underestimation of the values (maybe this is a warming trend?). So far there is no sign of a slowing down here. In addition, the values all seem to show a trend upwards which seem to disagree with the idea of slowing down completely for all seasons. It seems that a great fit for all the models was an AMRA model as the ACF and PACF seemed to decay natually for most of the time series. Winter time series showed perhaps it doesn;t statistically differ from a white noise series. ARMA(1,1) for the other three fit the model well. Further analysis on other data such as ocean temperature or solar activity may be needed to fully understand if the raising effect we see here in the models are from those events or perhaps has global warming caught up to us - in terms of human activity.

References

1) https://www.climate.gov/news-features/climate-qa/why-did-earth%E2%80%99s-surface-temperature-stop-rising-past-decade
2) https://www.scientificamerican.com/article/did-global-warming-slow-down-in-the-2000s-or-not/
3) http://www.geo.umass.edu/faculty/bradley/jones1992a.pdf