

# Climate Change in Chicago

*Dustin Leatherman*

*2/29/2020*

## **Abstract**

The effects of climate change are being felt around the globe, sometimes more obviously in some areas than others. Chicago is famous for its harsh winters but as climate change sets in, has it affected Chicago winters? This is a look at 30 years of weekly averages for temperature data collected at O'Hare International Airport in Chicago, Illinois. Hourly and daily data was sourced by NOAA as part of its Local Climate Dataset program. Missing daily values were calculated from hourly values, which both were subsequently rolled up to a weekly granularity. The Data indicate that weekly Dry Bulb Temperatures have remained relatively stable over the past 30 years hinting at no significant change. The Box-Jenkins approach to ARIMA modeling is used to fit and forecast weekly average temperatures for the year 2020. Significant challenges were faced during model fitting as the number of transformations available were limited due to the possibility of negative forecasts, outliers, and the weekly granularity. Further analysis using deviations from mean temperature for the same time period may prove more insightful.

# Introduction

Climate is defined as the long-term average of weather, typically over a period of 30 years (Serge Planton 2016). Anthropogenic climate change is a well-known phenomenon that has captivated regional and national politics. This begs the question, how has Chicago's local climate changed over the past 30 years? Observations over a 30 year period from O'Hare International Airport have been collected and freely distributed by the NOAA. O'Hare's position makes it ideal for local climate data since it is far enough away from Lake Michigan to be free from the Lake Effect.

The questions at hand are: - Have Chicago Winters gotten warmer over the last 30 years? - Have Chicago Summers gotten warmer over the last 30 years? - What will the next year look like for Chicago based on the past 30?

Local Climatological Data (LCD) are summaries of climatological conditions from airport and other prominent weather stations managed by NWS, FAA, and DOD. The product includes hourly observations and associated remarks, and a record of hourly precipitation for the entire month. Also included are Weekly summaries summarizing temperature extremes, degree days, precipitation amounts and winds (*U.S. Local Climatological Data (Lcd)*, n.d.).

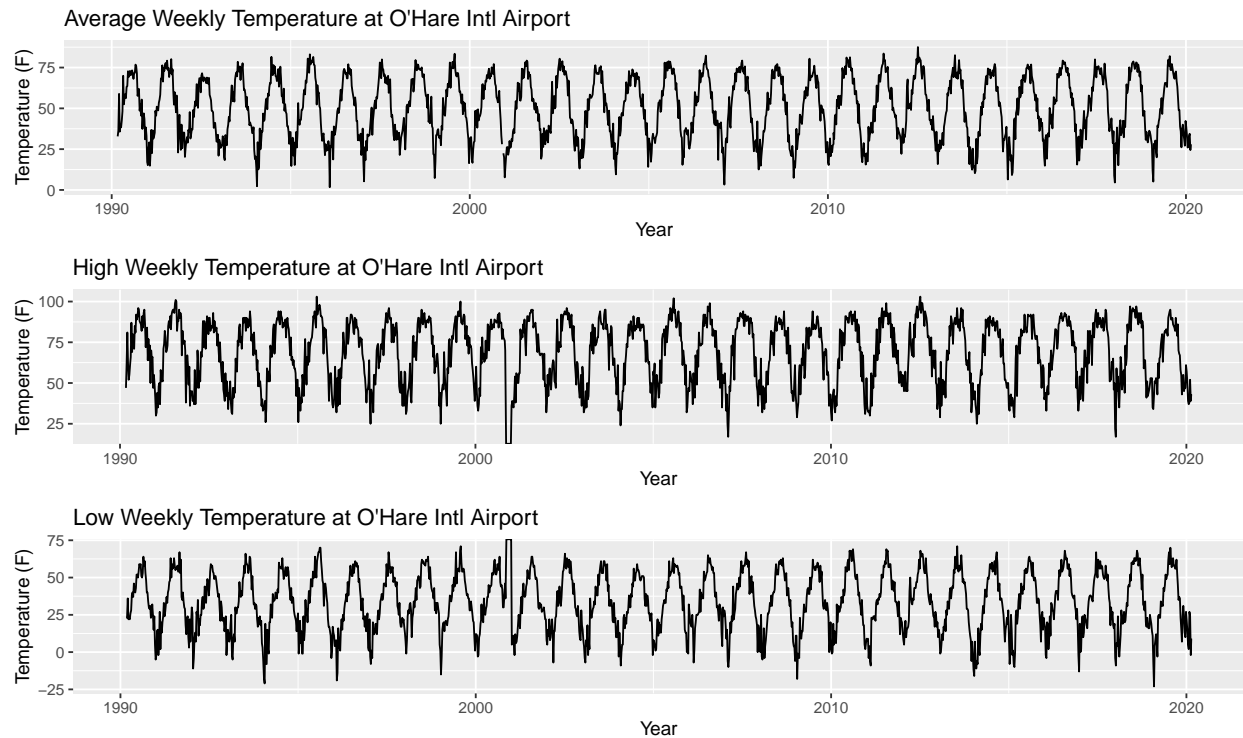
While there are many variables of interest in the data, the one most relevant to the questions at hand is Average Dry Bulb Temperature. Dry Bulb Temperature is the temperature of air measured by a thermometer freely exposed to the air, but shielded from radiation and moisture. This is considered one of the most important climate variables for human comfort (D.H Nall, n.d.). This is in direct comparison to the Wet Bulb Temperature which is measured on a thermometer wrapped in a damp cloth to mimic the temperature at 100% humidity. While Chicago can experience extreme bouts of humidity, that measure will not be taken into consideration.

# Data Preperation

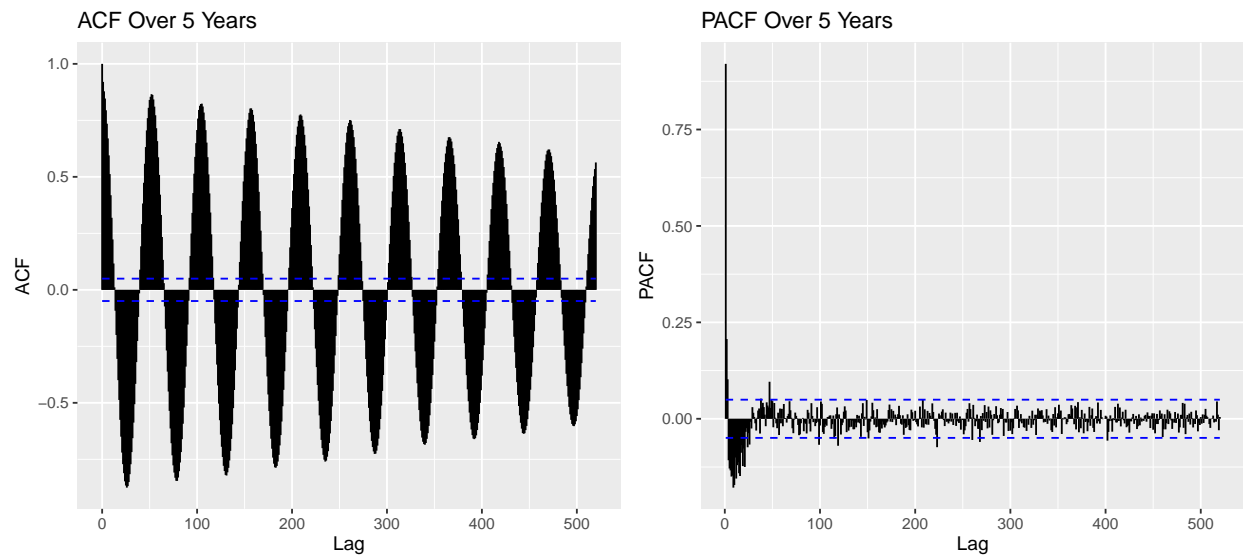
The data were collected from the U.S. Local Climatological Data (LCD) provided by the NOAA. The data contained three types of reports: hourly, daily, and monthly summaries. In addition, due to technical limitations of the NOAA's platform, only 10 years of measurements could be downloaded at a time. This required the merging the datasets, ensuring that no measurements overlapped for any given day. Furthermore, there were periods of missing hourly and daily values. Missing daily values were replaced with aggregations from hourly values. In the interest of not overfitting missing values with averages, any missing hourly values were ignored during the daily aggregations. It is worth noting that the average values provided by the daily summary were *approximate* to the calculated daily values. The daily summary temperatures were calculated using the arithmetic average of the max and min temperature for the day and taking the integer value. This discovery was made late in the analysis process and serves as a point for future improvement.

Initial analysis of this dataset was at the daily level but as one might expect, Seasonal ARIMA modeling becomes costly (and outright does not run) beyond a certain number of lags. Hence, the data was aggregated to the weekly grain for analytical purposes.

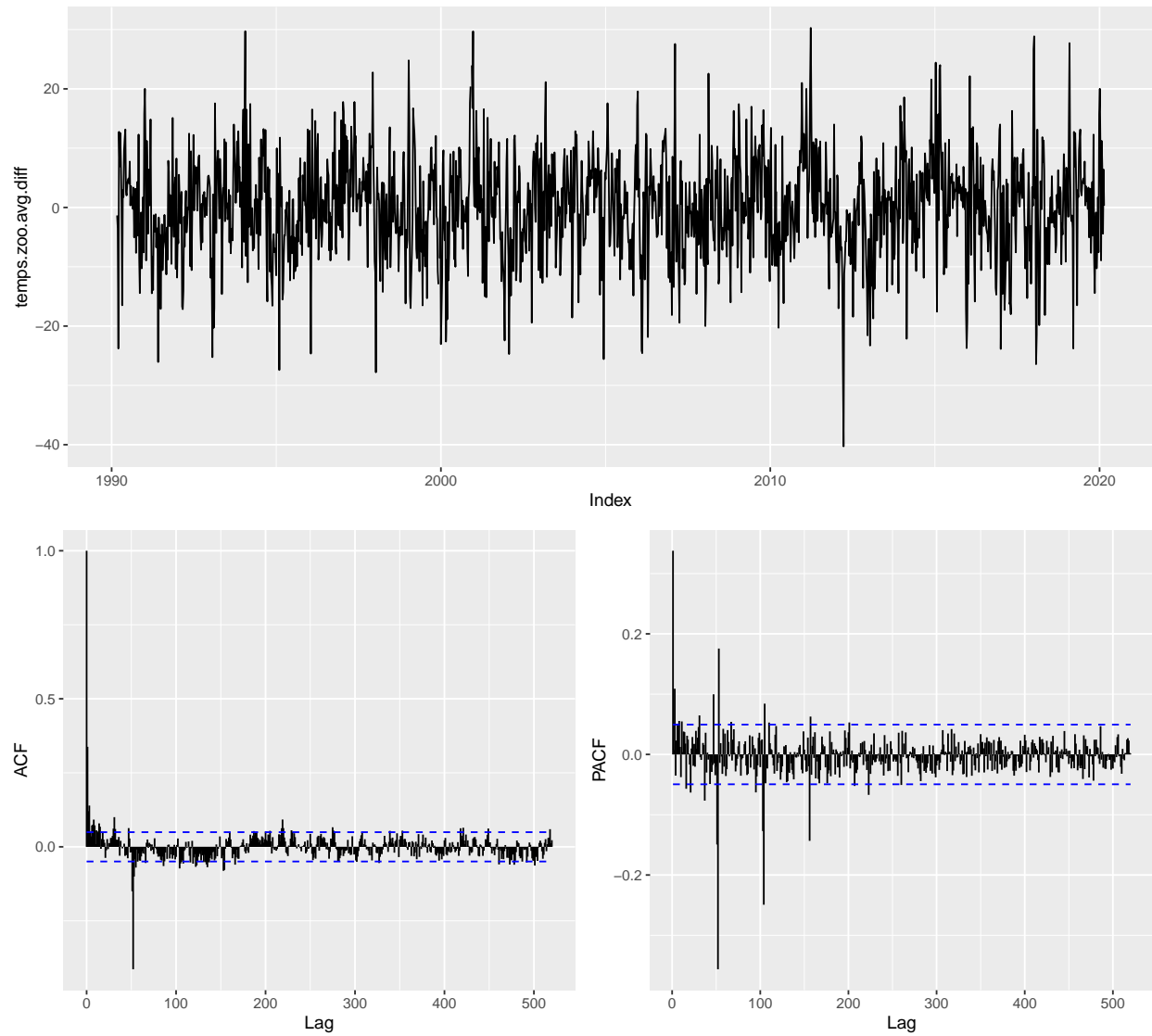
## Model Specification



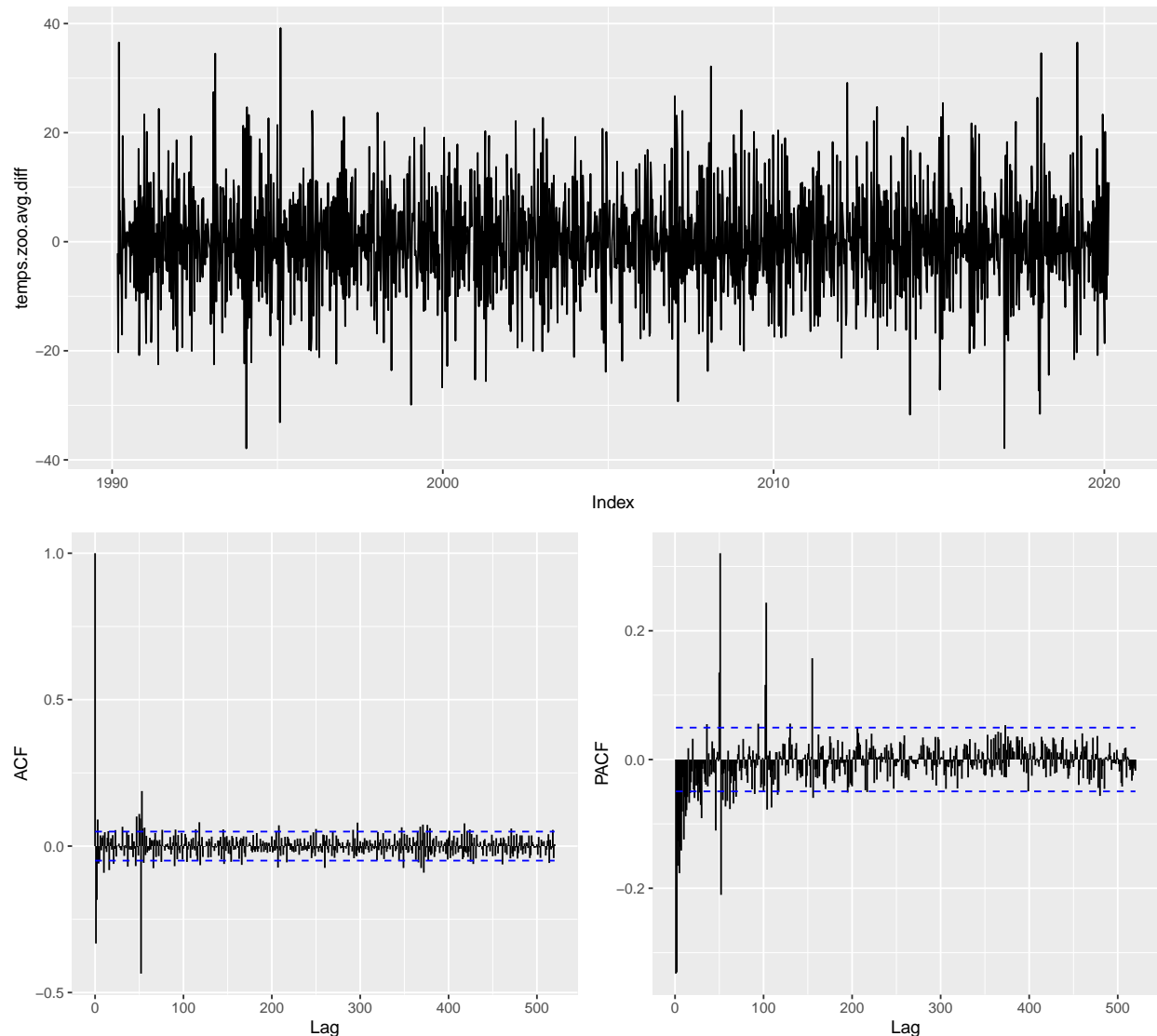
The peaks represent the summers and the valleys represent the winters. There doesn't appear to be a trend year over year but there appears to be more outliers with regards to extreme cold temperatures compared to extreme warm temperatures. Comparing the timeseries with Minimum and Maximum Dry Bulb Temperature show similar patterns. This graph highlights that Dry Bulb Temperature data is missing data for November and December of 2000. The series appears to be stationary with seasonality so differencing may not be needed.



As expected, there is seasonal correlation with the previous year present in the ACF plot and some correlation within the same year in the PACF plot. First, seasonality will be dealt with.



When differenced over one year, much of the seasonality is reduced. Since there doesn't appear to be a quadratic trend, a second difference is not needed. This indicates that Seasonal differencing of 1 is acceptable.



There are 4 significant bars on successive years on the PACF plot and 1 significant bar on the ACF plot indicating a SARIMA(p,d,q)x(4,1,1) may be an appropriate fit for a seasonal model. With regards to the non-seasonal portion, the ACF plot trails off to zero after the third lag indicating that an MA(3) model may be a possibility. The first lag is not significant but these models can still be run as options. Some possible orders include AR(1) - AR(5).

Since the seasonal Autoregressive (SAR) correlations were stronger than the Seasonal Moving Average (SMA) correlations, models without a SMA will be considered. Because AR and MA terms can work against each other, the approach for modeling will be to balance SMA with AR and SAR with MA terms. This will ensure that models are less likely to be overparameterized and provide conflicting output.

## Transformations (or lackthereof)

One particular principle worth sticking to is ensuring that any transformations that are applied do not remove the integrity of negative values. **While there are no negative weekly average Dry Bulb temperatures, any model should be able to predict them.** This rules out the log, square root, and power transformations. While log and square root are obvious, a power transformation cannot be applied because there is no way to *undo* a power value without knowing the original sign.

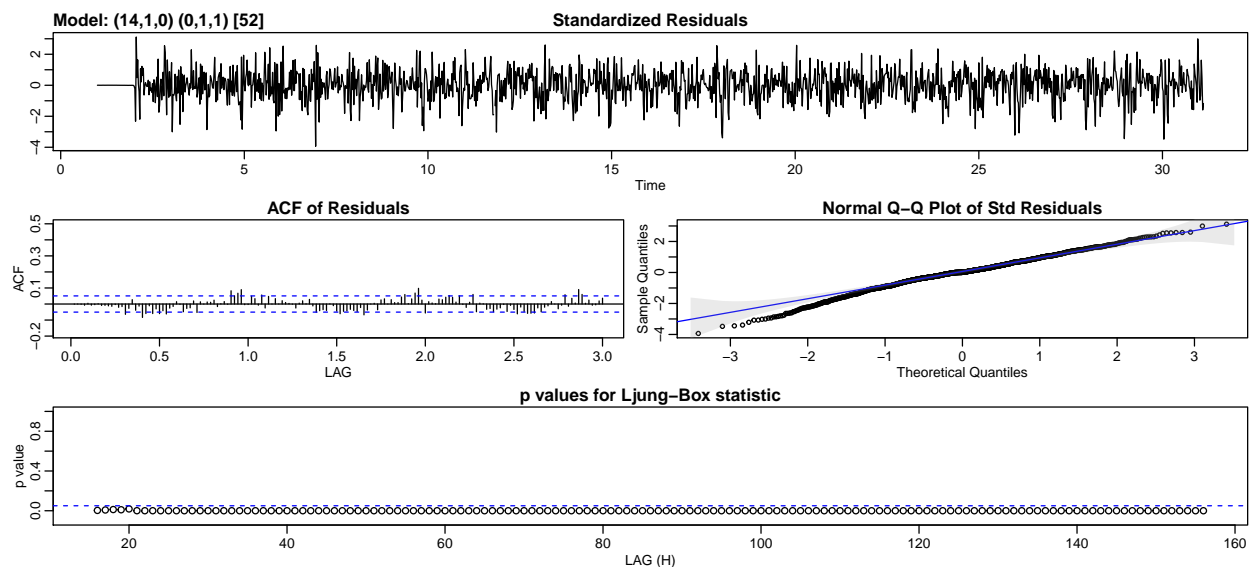
## Fitting and Diagnostics

Given previous observations, potential models include:

- SARIMA(14,1,0)x(0,1,1)
- SARIMA(8,1,0)x(0,1,1)
- SARIMA(0,1,3)x(3,1,0)
- SARIMA(0,1,5)x(3,1,0)
- SARIMA(14,1,0)x(0,1,3)
- SARIMA(8,1,0)x(0,1,3)

### SARIMA(14,1,0)x(0,1,1)

```
## initial value 2.287723
## iter 2 value 2.021136
## iter 3 value 2.002637
## iter 4 value 1.886124
## iter 5 value 1.870115
## iter 6 value 1.867264
## iter 7 value 1.863890
## iter 8 value 1.863291
## iter 9 value 1.863225
## iter 10 value 1.863220
## iter 11 value 1.863219
## iter 12 value 1.863219
## iter 12 value 1.863219
## iter 12 value 1.863219
## final value 1.863219
## converged
```



```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##      Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
```

```

##      optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1          ar2          ar3          ar4          ar5          ar6          ar7          ar8
##      -0.5794 -0.5415 -0.3853 -0.3634 -0.3319 -0.2804 -0.2449 -0.1729
## s.e.   0.0263  0.0314  0.0350  0.0372  0.0381  0.0393  0.0400  0.0396
##          ar9          ar10         ar11         ar12         ar13         ar14         sma1
##      -0.1462 -0.1758 -0.1063 -0.0985 -0.0672 -0.0182 -0.814
## s.e.   0.0387  0.0376  0.0363  0.0345  0.0309  0.0266  0.019
##
## sigma^2 estimated as 39.98:  log likelihood = -4965.9,  aic = 9963.81
##
## $degrees_of_freedom
## [1] 1498
##
## $ttable
##      Estimate      SE  t.value p.value
## ar1   -0.5794 0.0263 -22.0052 0.0000
## ar2   -0.5415 0.0314 -17.2569 0.0000
## ar3   -0.3853 0.0350 -11.0181 0.0000
## ar4   -0.3634 0.0372  -9.7624 0.0000
## ar5   -0.3319 0.0381  -8.7143 0.0000
## ar6   -0.2804 0.0393  -7.1354 0.0000
## ar7   -0.2449 0.0400  -6.1238 0.0000
## ar8   -0.1729 0.0396  -4.3663 0.0000
## ar9   -0.1462 0.0387  -3.7761 0.0002
## ar10  -0.1758 0.0376  -4.6785 0.0000
## ar11  -0.1063 0.0363  -2.9274 0.0035
## ar12  -0.0985 0.0345  -2.8524 0.0044
## ar13  -0.0672 0.0309  -2.1728 0.0300
## ar14  -0.0182 0.0266  -0.6833 0.4945
## sma1  -0.8140 0.0190 -42.9317 0.0000
##
## $AIC
## [1] 6.366651
##
## $AICc
## [1] 6.366849
##
## $BIC
## [1] 6.42106

```

There are a couple outliers shown in the Standardized residual plot. There is a clear pattern in the ACF of the residuals which has alternating significant correlations every 6 months. The residuals appear to deviate from normality but the sample size is large enough where that is not a concern. Most of the parameters except AR(14) are considered significant.

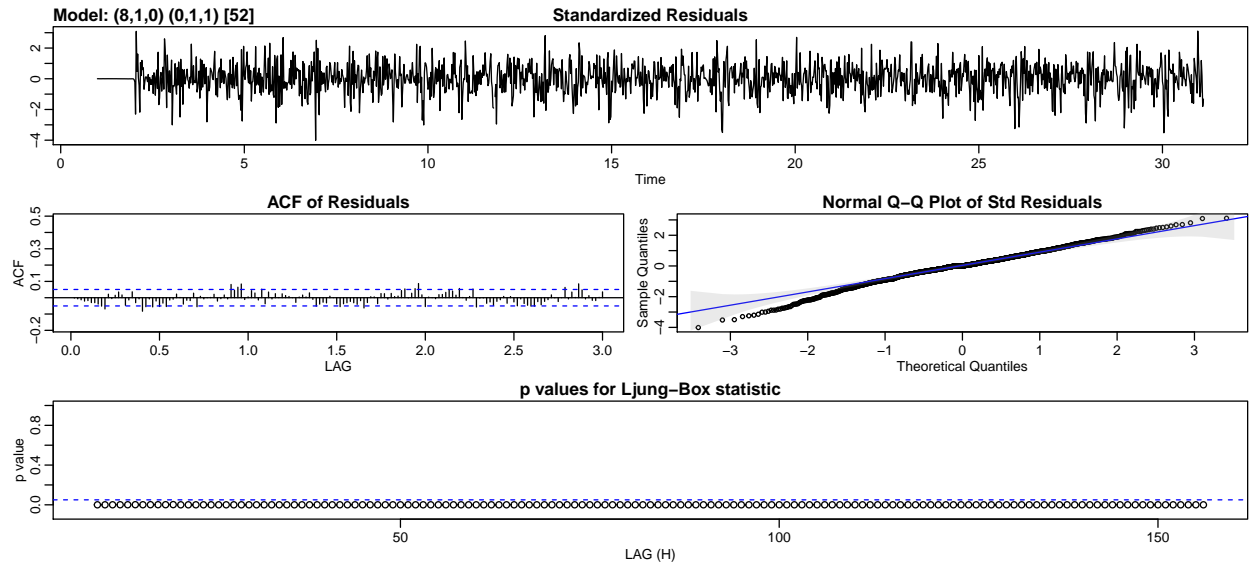
## SARIMA(8,1,0)x(0,1,1)

```

## initial  value 2.287723
## iter    2 value 2.021529
## iter    3 value 1.944409
## iter    4 value 1.890882
## iter    5 value 1.881374

```

```
## iter 6 value 1.873028
## iter 7 value 1.871365
## iter 8 value 1.871217
## iter 9 value 1.871198
## iter 10 value 1.871195
## iter 11 value 1.871195
## iter 11 value 1.871195
## iter 11 value 1.871195
## final value 1.871195
## converged
```



```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
##     optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##      ar1      ar2      ar3      ar4      ar5      ar6      ar7      ar8
## -0.5597 -0.5048 -0.3322 -0.2962 -0.2525 -0.1871 -0.1378 -0.0459
## s.e.    0.0261  0.0303  0.0329  0.0337  0.0334  0.0329  0.0304  0.0263
##      sma1
##     -0.8464
## s.e.    0.0176
##
## sigma^2 estimated as 40.38:  log likelihood = -4977.97,  aic = 9975.94
##
## $degrees_of_freedom
## [1] 1504
##
## $ttable
##      Estimate      SE  t.value p.value
## ar1   -0.5597 0.0261 -21.4602  0.0000
## ar2   -0.5048 0.0303 -16.6483  0.0000
## ar3   -0.3322 0.0329 -10.0915  0.0000
```



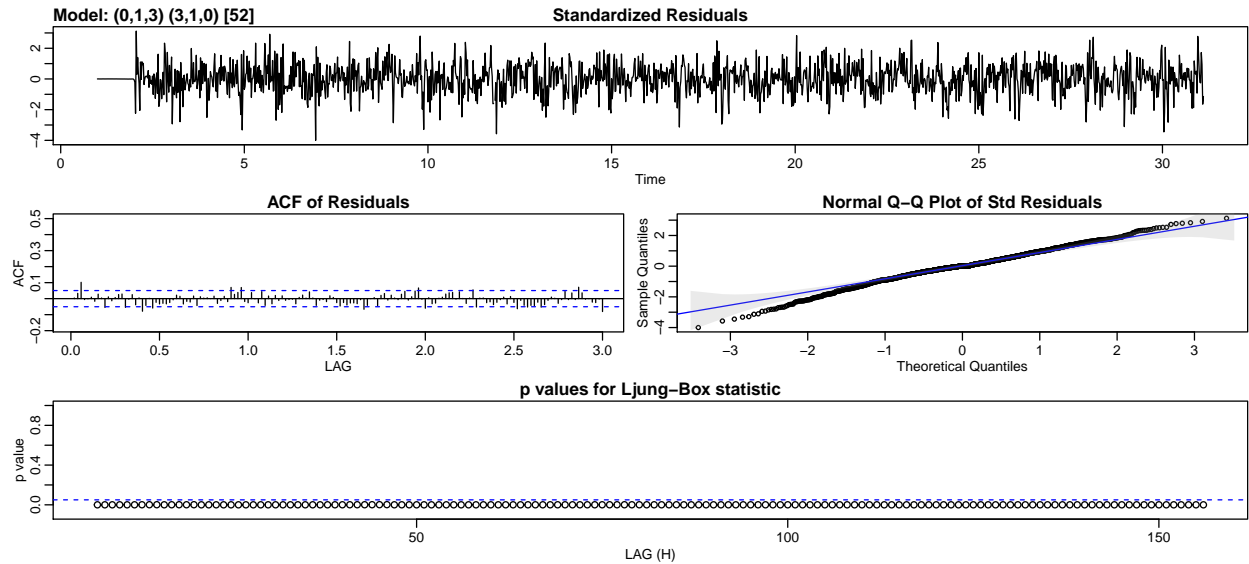
```
## ar4    -0.2962 0.0337 -8.7912 0.0000
## ar5    -0.2525 0.0334 -7.5667 0.0000
## ar6    -0.1871 0.0329 -5.6915 0.0000
## ar7    -0.1378 0.0304 -4.5329 0.0000
## ar8    -0.0459 0.0263 -1.7442 0.0813
## sma1   -0.8464 0.0176 -48.0935 0.0000
##
## $AIC
## [1] 6.374405
##
## $AICc
## [1] 6.374479
##
## $BIC
## [1] 6.40841
```

The plots here are similar to above indicating that while there are less parameters, this model does not appear to provide any improvement to the diagnostics.

MA can have the effect of *undoing* a difference while AR can have the effect of differencing. It is possible in these cases that including a SMA term has *undone* the differencing applied and the AR terms are not able to make up for it.

### SARIMA(0,1,3)x(3,1,0)

```
## initial value 2.287723
## iter 2 value 2.035861
## iter 3 value 2.034545
## iter 4 value 1.916806
## iter 5 value 1.909725
## iter 6 value 1.883256
## iter 7 value 1.881133
## iter 8 value 1.879283
## iter 9 value 1.878880
## iter 10 value 1.878822
## iter 11 value 1.878815
## iter 12 value 1.878814
## iter 13 value 1.878814
## iter 13 value 1.878814
## iter 13 value 1.878814
## final value 1.878814
## converged
```



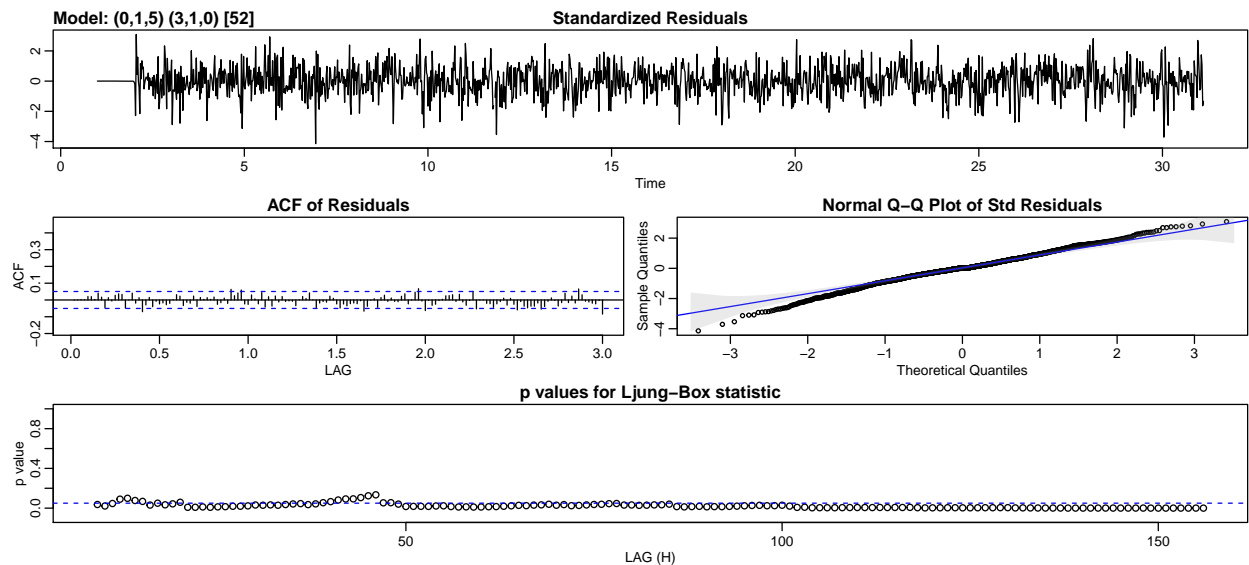
```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
##     optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ma1      ma2      ma3      sar1      sar2      sar3
##       -0.6233 -0.2673 -0.0466 -0.7322 -0.5340 -0.2611
## s.e.   0.0256   0.0325   0.0253   0.0256   0.0292   0.0264
##
## sigma^2 estimated as 41.69:  log likelihood = -4989.5,  aic = 9993
##
## $degrees_of_freedom
## [1] 1507
##
## $ttable
##      Estimate      SE  t.value p.value
## ma1   -0.6233 0.0256 -24.3369  0.000
## ma2   -0.2673 0.0325  -8.2349  0.000
## ma3   -0.0466 0.0253  -1.8467  0.065
## sar1  -0.7322 0.0256 -28.6107  0.000
## sar2  -0.5340 0.0292 -18.2795  0.000
## sar3  -0.2611 0.0264  -9.9044  0.000
##
## $AIC
## [1] 6.385303
##
## $AICc
## [1] 6.385337
##
## $BIC
## [1] 6.409107
```

A seasonal pattern is still present the residuals indicating that there is still a trend that has not been accounted

for in the data.

## SARIMA(0,1,5)x(3,1,0)

```
## initial value 2.287723
## iter 2 value 2.028273
## iter 3 value 1.946430
## iter 4 value 1.922051
## iter 5 value 1.901569
## iter 6 value 1.894402
## iter 7 value 1.873683
## iter 8 value 1.872377
## iter 9 value 1.871891
## iter 10 value 1.871697
## iter 11 value 1.871665
## iter 12 value 1.871662
## iter 13 value 1.871662
## iter 14 value 1.871662
## iter 14 value 1.871662
## iter 14 value 1.871662
## final value 1.871662
## converged
```



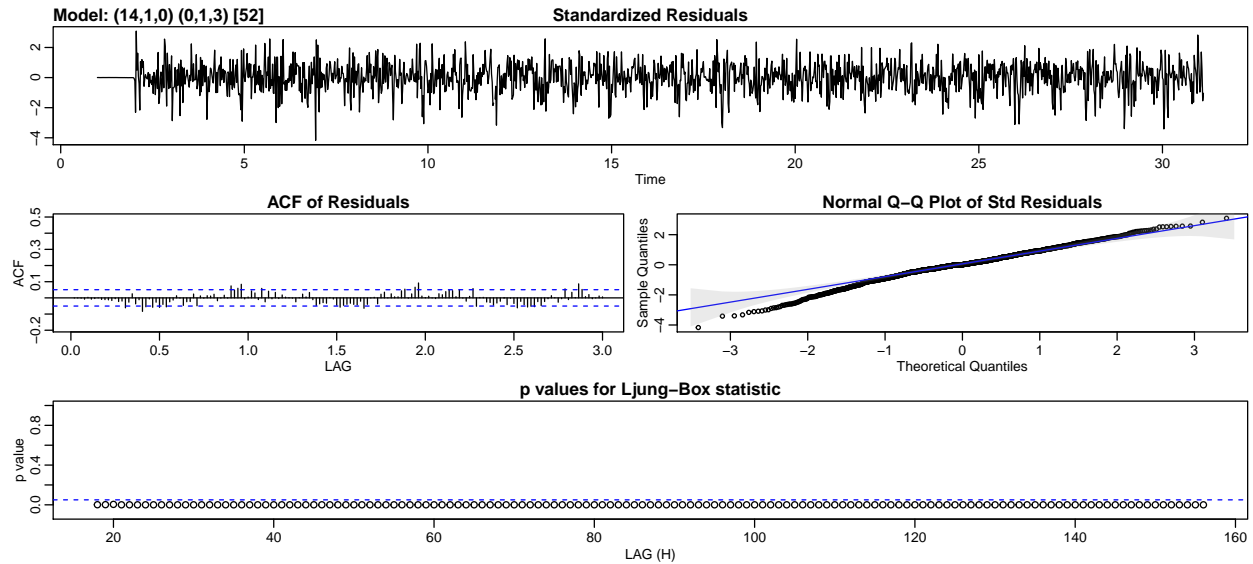
```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
##     optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ma1          ma2          ma3          ma4          ma5          sar1          sar2          sar3
##        -0.6233    -0.2362     0.0330    -0.0795    -0.0516    -0.7354    -0.5404    -0.2635
## s.e.    0.0258     0.0303     0.0309     0.0302     0.0263     0.0256     0.0292     0.0264
##
```

```
## sigma^2 estimated as 41.07:  log likelihood = -4978.68,  aic = 9975.36
##
## $degrees_of_freedom
## [1] 1505
##
## $ttable
##      Estimate      SE  t.value p.value
## ma1   -0.6233 0.0258 -24.1719 0.0000
## ma2   -0.2362 0.0303  -7.7930 0.0000
## ma3    0.0330 0.0309   1.0667 0.2863
## ma4   -0.0795 0.0302  -2.6318 0.0086
## ma5   -0.0516 0.0263  -1.9638 0.0497
## sar1  -0.7354 0.0256 -28.6785 0.0000
## sar2  -0.5404 0.0292 -18.5252 0.0000
## sar3  -0.2635 0.0264  -9.9855 0.0000
##
## $AIC
## [1] 6.37403
##
## $AICc
## [1] 6.374089
##
## $BIC
## [1] 6.404635
```

There are still a handful of outliers but with limited transformation options, not much can be done to mitigate their impact. There still appears to be a seasonal pattern in the residuals but it is less pronounced than the previous models. Normality is on par with the previous models. A small peak is visible in the Ljung-Box graph which is an improvement over a flat line. So far, this is the best model.

### SARIMA(14,1,0)x(0,1,3)

```
## initial  value 2.287723
## iter    2 value 2.012578
## iter    3 value 1.932547
## iter    4 value 1.891855
## iter    5 value 1.877578
## iter    6 value 1.864245
## iter    7 value 1.861225
## iter    8 value 1.860918
## iter    9 value 1.859813
## iter   10 value 1.859766
## iter   11 value 1.859761
## iter   12 value 1.859760
## iter   13 value 1.859760
## iter   13 value 1.859760
## iter   13 value 1.859760
## final   value 1.859760
## converged
```



```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##      Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
##      optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##      ar1      ar2      ar3      ar4      ar5      ar6      ar7      ar8
##    -0.5871 -0.5487 -0.3970 -0.3765 -0.3495 -0.2951 -0.2598 -0.1855
## s.e.   0.0263  0.0313  0.0349  0.0372  0.0383  0.0395  0.0403  0.0399
##      ar9      ar10     ar11     ar12     ar13     ar14     sma1     sma2
##    -0.1618 -0.1827 -0.1204 -0.1072 -0.072  -0.0185 -0.8233 -0.0505
## s.e.   0.0392  0.0377  0.0367  0.0347  0.031  0.0266  0.0276  0.0336
##      sma3
##      0.0862
## s.e.  0.0269
##
## sigma^2 estimated as 39.7:  log likelihood = -4960.67,  aic = 9957.34
##
## $degrees_of_freedom
## [1] 1496
##
## $ttable
##      Estimate      SE  t.value p.value
## ar1   -0.5871  0.0263 -22.3147  0.0000
## ar2   -0.5487  0.0313 -17.5244  0.0000
## ar3   -0.3970  0.0349 -11.3686  0.0000
## ar4   -0.3765  0.0372 -10.1207  0.0000
## ar5   -0.3495  0.0383  -9.1348  0.0000
## ar6   -0.2951  0.0395  -7.4786  0.0000
## ar7   -0.2598  0.0403  -6.4415  0.0000
## ar8   -0.1855  0.0399  -4.6546  0.0000
## ar9   -0.1618  0.0392  -4.1259  0.0000
## ar10  -0.1827  0.0377  -4.8442  0.0000
## ar11  -0.1204  0.0367  -3.2778  0.0011
```

```

## ar12  -0.1072  0.0347  -3.0893  0.0020
## ar13  -0.0720  0.0310  -2.3212  0.0204
## ar14  -0.0185  0.0266  -0.6960  0.4865
## sma1  -0.8233  0.0276 -29.8013  0.0000
## sma2  -0.0505  0.0336  -1.5019  0.1333
## sma3   0.0862  0.0269   3.2076  0.0014
##
## $AIC
## [1] 6.362519
##
## $AICc
## [1] 6.362772
##
## $BIC
## [1] 6.423729

```

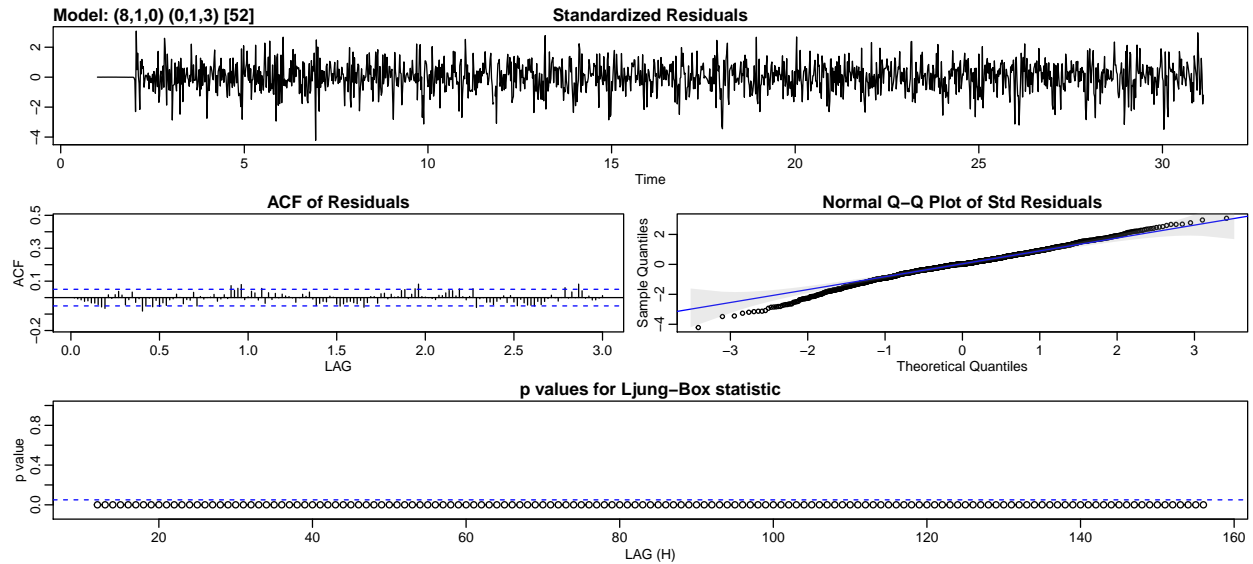
The Standardized Residuals, QQ Plot, and Ljung-Box plot match many of the previous models. A pronounced seasonal trend is present in the ACF Residual plot indicating that there is a trend in the data that has not been accounted for.

### **SARIMA(8,1,0)x(0,1,3)**

```

## initial  value 2.287723
## iter    2 value 2.012544
## iter    3 value 1.929380
## iter    4 value 1.893348
## iter    5 value 1.883087
## iter    6 value 1.871259
## iter    7 value 1.869111
## iter    8 value 1.868724
## iter    9 value 1.868405
## iter   10 value 1.868397
## iter   11 value 1.868396
## iter   12 value 1.868396
## iter   12 value 1.868396
## iter   12 value 1.868396
## final   value 1.868396
## converged

```



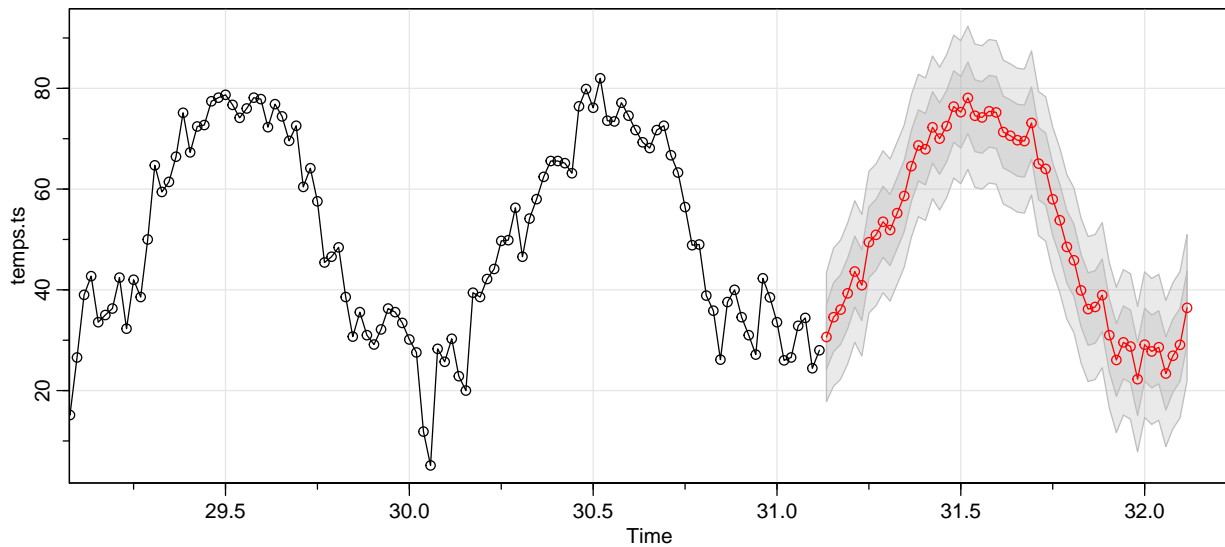
```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
##     optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1      ar2      ar3      ar4      ar5      ar6      ar7      ar8
##      -0.5646 -0.5082 -0.3390 -0.3033 -0.2622 -0.1927 -0.1412 -0.0473
## s.e.   0.0261  0.0303  0.0329  0.0336  0.0334  0.0329  0.0304  0.0263
##          sma1      sma2      sma3
##      -0.8462 -0.0588  0.0784
## s.e.   0.0274  0.0341  0.0269
##
## sigma^2 estimated as 40.18:  log likelihood = -4973.74,  aic = 9971.47
##
## $degrees_of_freedom
## [1] 1502
##
## $ttable
##      Estimate      SE  t.value p.value
## ar1   -0.5646 0.0261 -21.6599  0.0000
## ar2   -0.5082 0.0303 -16.7962  0.0000
## ar3   -0.3390 0.0329 -10.3134  0.0000
## ar4   -0.3033 0.0336  -9.0150  0.0000
## ar5   -0.2622 0.0334  -7.8496  0.0000
## ar6   -0.1927 0.0329  -5.8646  0.0000
## ar7   -0.1412 0.0304  -4.6430  0.0000
## ar8   -0.0473 0.0263  -1.8011  0.0719
## sma1  -0.8462 0.0274 -30.9301  0.0000
## sma2  -0.0588 0.0341  -1.7248  0.0848
## sma3   0.0784 0.0269   2.9197  0.0036
##
## $AIC
## [1] 6.371549
```

```
##
## $AICc
## [1] 6.371658
##
## $BIC
## [1] 6.412356
```

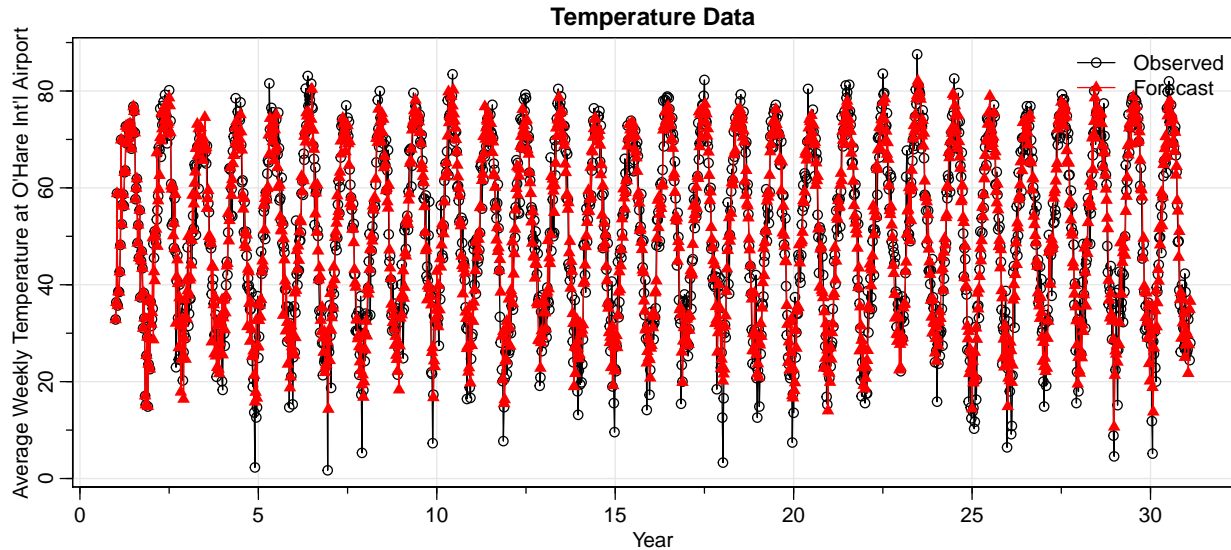
MODEL	AICc	AIC	BIC
SARIMA(0,1,5)x(3,1,0)	6.374089	6.374030	6.404635
SARIMA(8,1,0)x(0,1,1)	6.374479	6.374405	6.408410
SARIMA(0,1,3)x(3,1,0)	6.385337	6.385303	6.409107
SARIMA(8,1,0)x(0,1,3)	6.371658	6.371549	6.412356
SARIMA(14,1,0)x(0,1,1)	6.366849	6.366651	6.421060
SARIMA(14,1,0)x(0,1,3)	6.362772	6.362519	6.423729

## Forecasting

While not the most ideal in terms of diagnostics, the model of choice is SARIMA(0,1,5)x(3,1,0) with a BIC of 6.404635. It appears that the data shows a stronger affinity for Autoregressive models opposed to Moving Average for the Seasonal parts of the model while the non-seasonal portions can be modeled better with a moving average.







The predicted values combined with the confidence interval for the forecasted values very much mimic the peaks and valleys that have occurred before lending a sense of pragmatism to this model.

## Discussion

The questions initially posed were as follows:

- Have Chicago Winters gotten warmer over the last 30 years?
- Have Chicago Summers gotten warmer over the last 30 years?
- What will the next year look like for Chicago based on the past 30?

The first two questions can be answered with the initial time-series plot of the data. Chicago summers have appeared to generally stay fairly consistent with regards to average temperatures. One could argue that winter is also consistent with temperature but it is also consistent with the fact that there are cyclic spikes that cause the temperature to fluctuate in both directions. This was surprising at first but provides somewhat of a realistic explanation for comments regarding the perceived difference in winter now versus 30 years ago.

The last question was just answered. It will look much like it has the past 30 years, with a chance for it to be warmer or colder than average. This is not the most exciting conclusion to reach but it is comforting to know that while the winters may seem different, they are mostly different because due to our own experiences.

## Further Improvements

This analysis does not take into account other variables that are known to affect the perception of temperature. i.e. wind, precipitation, humidity, etc. It could be improved by analyzing those time series and incorporating them into this model for a “feels like” temperature.

This data could also be condensed to a smaller set to improve modeling time and throughput. Having temperature at the weekly grain makes the forecasts more pertinent but increases computation time. The same analysis would be interesting to do with against monthly average temperatures. While it would reduce forecasting power, it would make more complex models possible.

More research into transformations that can be applied to negative values is warranted. Research was done regarding possible options but nothing conclusive arose.

## Final Thoughts

This model proves somewhat useful as a tool of understanding regarding local impacts by climate change. We can ascertain that Dry Bulb Temperature has not significantly changed in the past 30 years, nor does our model think it will going forward.

D.H Nall. n.d. *Looking Across the Water: Climate-Adaptive Buildings in the United States & Europe*. The Construction Specifier.

Serge Planton. 2016. *Annex Iii. Glossary: IPCC – Intergovernmental Panel on Climate Change*. IPCC. [https://web.archive.org/web/20160524223615/http://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5\\_AnnexIII\\_FINAL.pdf](https://web.archive.org/web/20160524223615/http://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_AnnexIII_FINAL.pdf).

*U.S. Local Climatological Data (Lcd)*. n.d. <https://www.ncei.noaa.gov/metadata/geoportal/rest/metadata/item/gov.noaa.ncdc:C00684/html#>.