Final Project

Aasha Reddy

3/16/2022

Introduction and Objectives

Most climate research cites warmer temperatures, more severe storms, increased drought, etc. Studies have shown that urban centers are generally experiencing a gradual rise in temperature (Tuholske et. al., 2021). However, climate change does not impact all cities in the same way. In New Delhi, India, studies have actually shown that temperatures are getting cooler. From 2016 to 2020, the mean annual temperature of Delhi dropped by 1.3 degrees Celcius (European Centre for Medium-Range Weather Forecasts), and the data shows that summer temperatures in Delhi have been slowly declining.

This is an important phenomenon to understand, as India is particularly prone to hydrometeorological disasters, which are increasing due to climate change (Prashar et. al, 2012). 80% of the country's population resides in extremely climate-vulnerable areas, and the frequency of extreme weather events has increased by about 200% since 2005 (Council on Energy, Environment and Water). Abnormal fluctuations in temperatures due to climate change have also accounted for over 700,000 deaths annually in India, about 94% of which are caused by extreme cold temperatures (Zhao et. al, 2019). In New Delhi, cooler weather has also been attributed to the buildup of smog and worsening air quality since 2016 (Yandap and Rawal, 2016).

In this paper, we aim to evaluate the claim that temperature has been getting cooler in New Delhi by building a time series model of monthly average temperature from 1996 to 2017. We will also evaluate trends in visibility (as a proxy for smog) and dew point temperature, and attempt to forecast weather patterns in the future.

Data

The dataset we use captures daily weather in New Delhi, India from 11/1/1996 to 4/27/2017, for average daily temperature, dew point temperature, humidity and visibility (https://www.kaggle.com/jonathanbouchet/new-delhi-20-years-of-weather-data/data?select=testset.csv). The data itself comes from Weather Underground, which provides daily weather data in New Delhi collected from the Safdarjung Airport Station.

The below tables summarize the data structure and show the first 10 rows of the dataset.

Table 1: Data Structure Summary

	Unit	Min	Max	Mean
date	Date	1996-11-01	2017-04-01	–
temp	Degrees Celcius	11.80560	36.13810	25.12
humidity	%	21.42045	88.96679	58.31
dew_pt_temp	Degrees Celcius	4.823890	27.593315	16.06
visibility	Km	0.7531557	3.9446691	2.13

Table 2: First 10 rows of Monthly New Delhi Weather Data

date	temp	humidity	dew_pt_temp	visibility
1996-11-01	19.43894	48.25248	7.252475	2.516174
1996-12-01	13.81679	58.71057	4.823890	1.918325
1997-01-01	12.87112	75.49068	7.910077	1.444385
1997-02-01	15.77878	62.43165	7.460432	2.354215
1997-03-01	21.13704	61.88704	12.512963	3.030060
1997-04-01	26.30693	50.52805	13.578254	3.814613
1997-05-01	29.94982	45.10036	14.607527	3.741715
1997-06-01	31.80620	64.49612	23.379845	3.708051
1997-07-01	31.14792	81.98305	27.300462	3.689882
1997-08-01	29.53608	83.04124	26.127148	3.419749

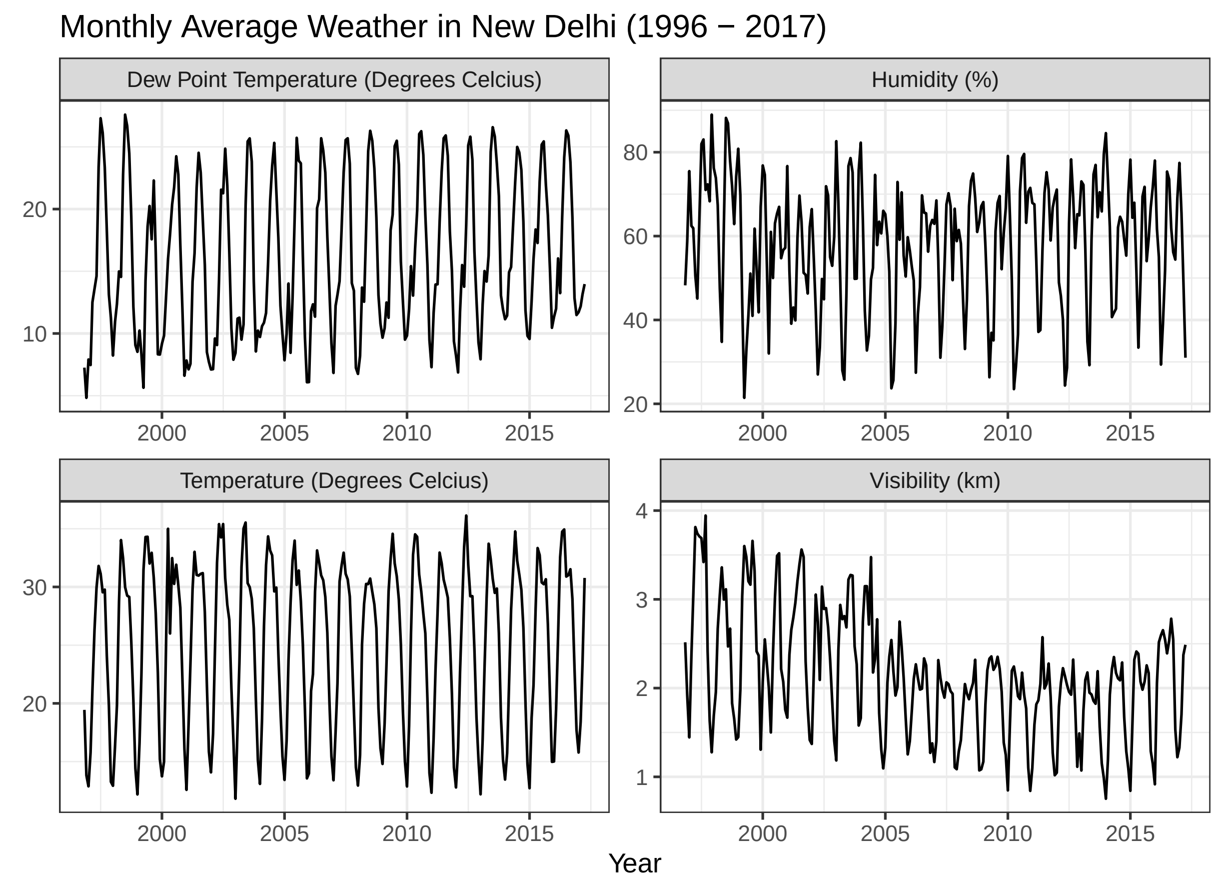
Analysis

Data Cleaning and Exploratory Analysis

First we clean the data. The raw data comes in an hourly format, so first we transform to a daily average between 1996 and 2017. We note that there are no issues with NA values in this dataset. However, there was one error in the Visilibity series due to an outlier. The maximum value of the Visibility series was 146.7511 km, which is about 76 standard deviations above the median value. We assume that this is an error in measurement and input the temperature fromt he day before. After this imputation, there are no other outliers we need to consider. We also find that these series may be too sensitive, and we also note that daily weather patterns may be of less interest than monthly average weather patterns. Thus, we further transform the data from daily average weather patters to monthly average weather patterns. Below we can see the time series for monthly average dew point temperature, humidity, temperature, and visibility in New Delhi between November 1996 to April 2017.

Based on the plot, it does not seem like temperature has a drastically decreasing trend, but it will be interesting to build a time series model that may be able to detect smaller change in temperature over time. One goal of the project is to also understand the difference in dew point temperature versus temperature. Meteorologists and climate change researchers often focus on dew point temperature, which is the temperature at which air is saturated with water vapor. In the below plot, we see that in cooler months, dew point temperature seems to be rising slightly over time. Visibility does seem to have a decreasing trend from 1996 to 2015, and potentially a slight upward trend after 2015. It is difficult to detect any trend in the humidity series based on the below plot. We decide not to focus on humidity, as meteorologists and climate researchers tend not to focus on this measure as a proxy for climate change.

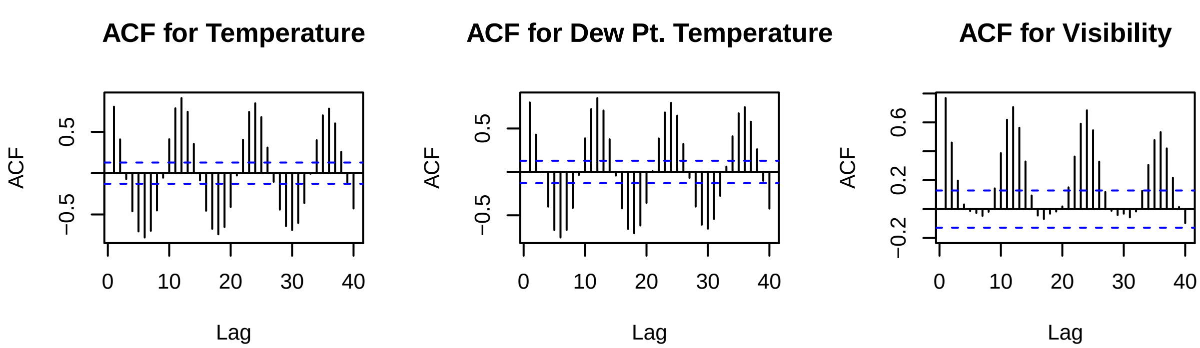
From the below plots, we note a yearly seasonality. It does not look like there are multiple seasonalities to consider, but we will examine this further.

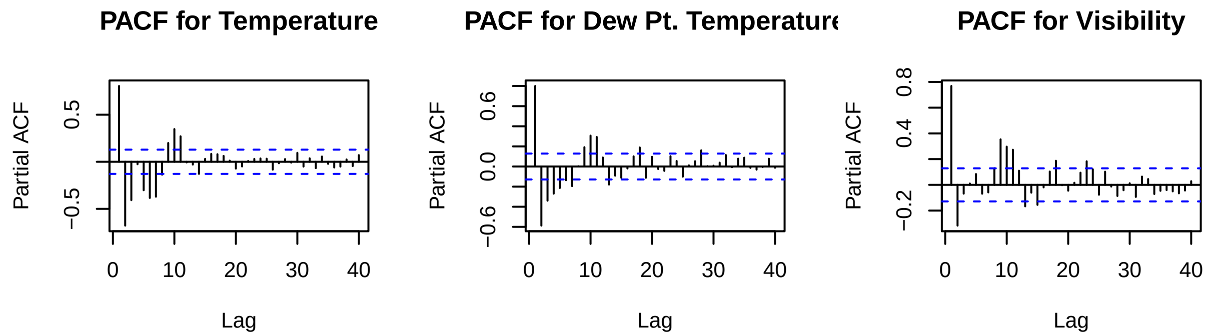


Methods and Models

To evaluate what type of time series best represents each weather series, we take November 1996 to March 2016 as a training dataset, with April 2016 to April 2017 as our test dataset. We will then choose a final model based on performance on the test set. We will refit the best model on all data, and forecast weather patterns from May 2017 to May 2020.

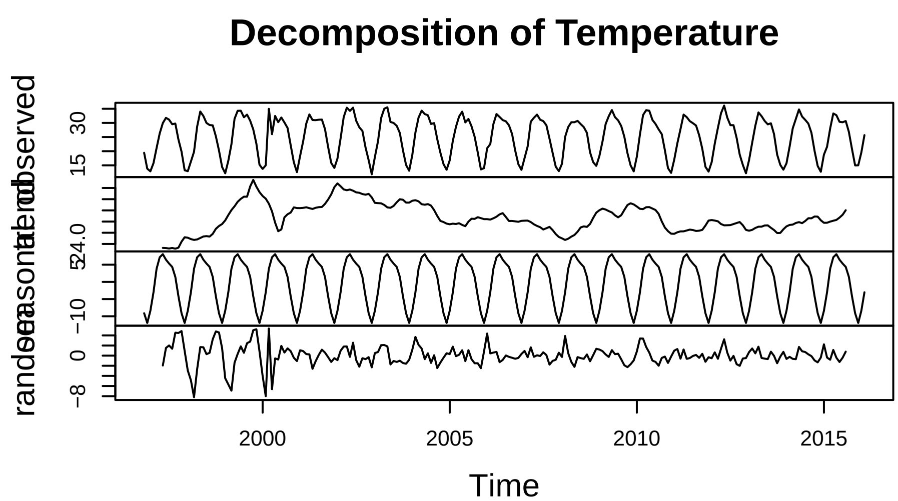
First we examine the ACF and PACF of each series for the training data. From the ACF plots we can see clear seasonality in each of the three series. We note that the ACF for visibility shows much smaller negative correlations than the ACF plots for temperature or dew point temperature. For each of the series', the PACFs plots show a slow exponential decay over time.

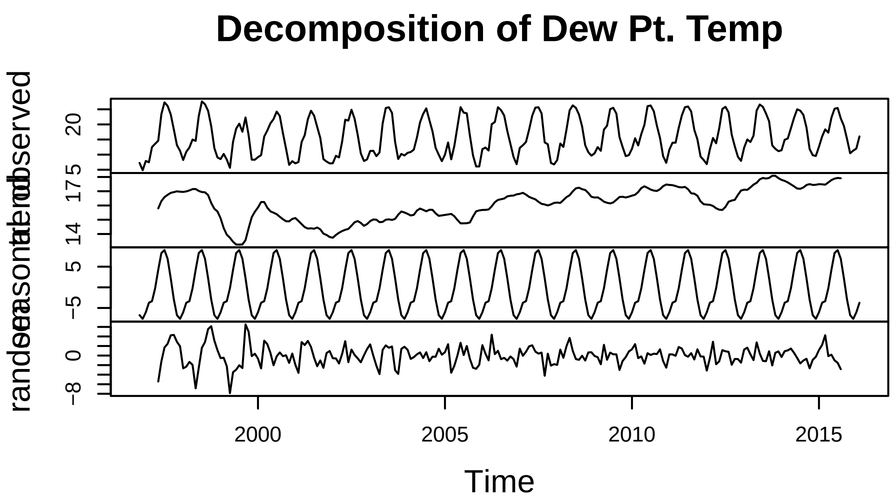


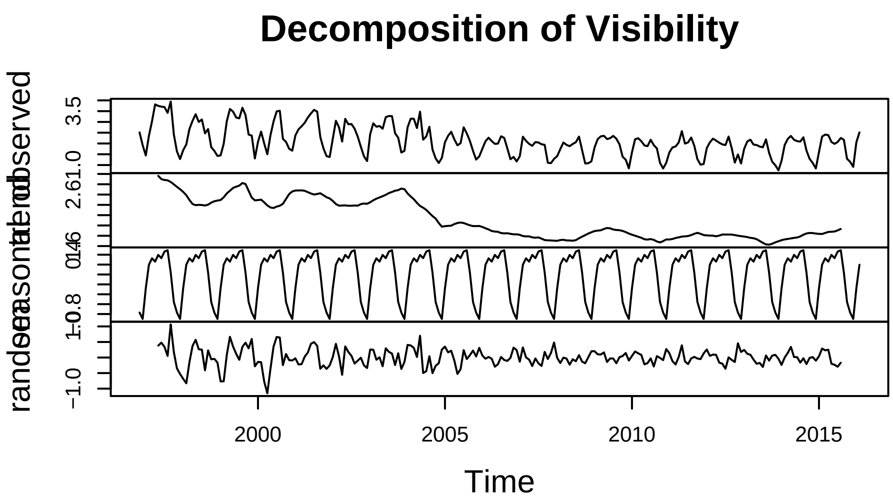


Manually Fitting Seasonal ARIMA Models

We first fit a seasonal ARIMA to all three series: temperature, dew point temperature, and visibility. We fit these models manually by examining the ACF and PACF plot for differenced series and seasonal lags. First we need to decompose each series. From the below decomposition plots, we can more easily see the increasing trend of dew point temperature over time. We also see a clear decrease in visibility over time. It is more difficult to assess any trend for change in temperature over time. We also see a clear yearly seasonality for all three series.







We then create non-seasonal series for temperature, dew point temperature, and visibility. We find that the non-seasonal temperature series does not need to be differenced, the dew point temperature series needs to be differenced once, and the visibility series needs to be differenced once.

```
## Number of differencing needed for Temperature:  0
```

```
## Number of differencing needed for Dew Pt. Temp.:  1
```

```
## Number of differencing needed for Visibility:  1
```

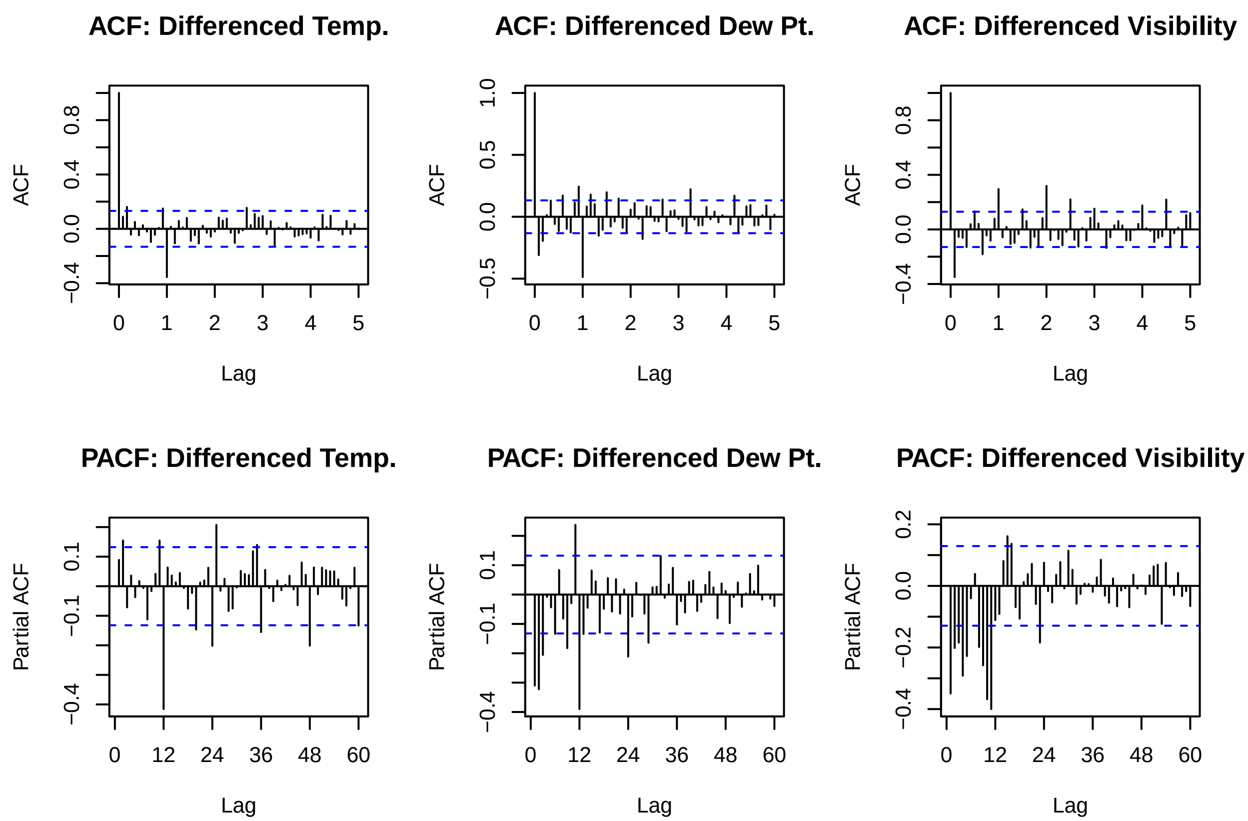
Next we will figure out how many times we need to difference the seasonal component of each series, and find that each series needs to be differenced once. We then look at the both seasonal and non-seasonal combined differenced series to determine the orders for our seasonal ARIMA models.

```
## Number of seasonal differencing needed for Temperature:  1
```

```
## Number of seasonal differencing needed for Dew Pt. Temp.:  1
```

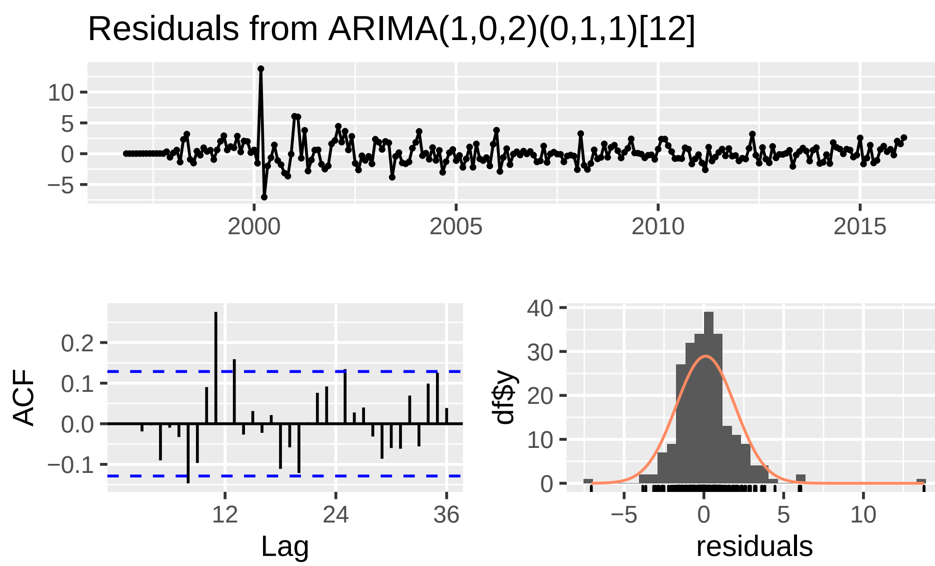
```
## Number of seasonal differencing needed for Visibility:  1
```

Next we examine the ACFs and PACFs plots for the differenced series'. For the Temperature series based on the ACF and PACF plots, we will fit an ARIMA(p = 1, d = 0, q = 2)(P = 0, D = 1, Q = 1)_[s=12]. For the Dew Point Temperature series, we will fit an ARIMA(p = 1, d = 1, q = 1)(P = 0, D = 1, Q = 1). For the Visibility series, we will fit an ARIMA(p = 1, d = 1, q = 1)(P = 1, D = 1, Q = 1). We go ahead and fit these models then examine the fit.



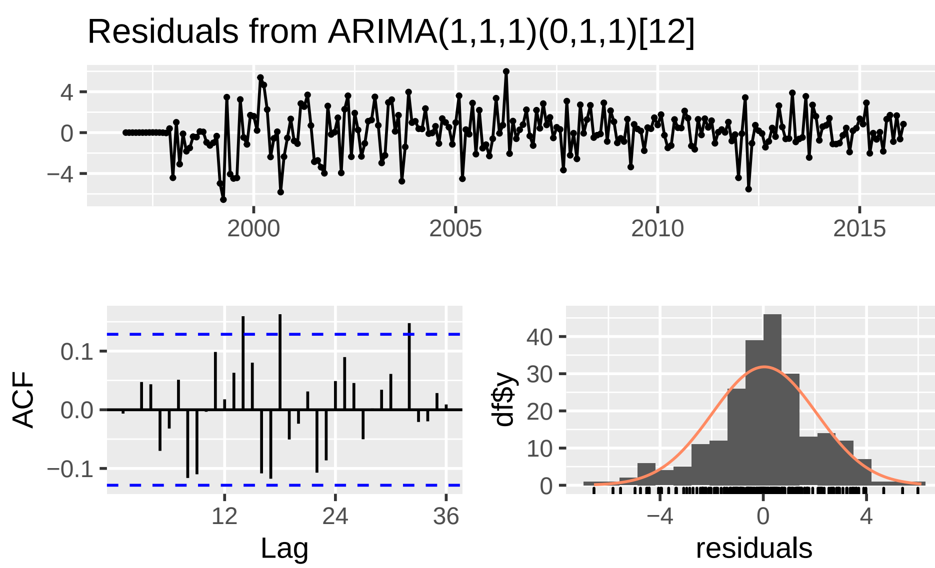
We see below that residuals for our model on temperature and visibility actually do not look so random, and there looks to still be some seasonality patterns. The residuals for the dew point temperature series look more random with no pattern. We note this as a limitation of manually fitting seasonal ARIMA models. The methodology used above to choose the order of the models based on PACF and ACF plots is often difficult and can be prone to errors. In the next phase of modeling, we will compare the manual models built here with autofitted seasonal ARIMA models.

Lastly, we obtain test accuracy for the model of temperature, dew point temperature, and visibility. We will compare these test accuracies to other models in the results section of our paper. To obtain test accuracy, we use the model to forecast for our test period of April 2016 to April 2017, and then compute accuracy against our true values from April 2016 to April 2017.



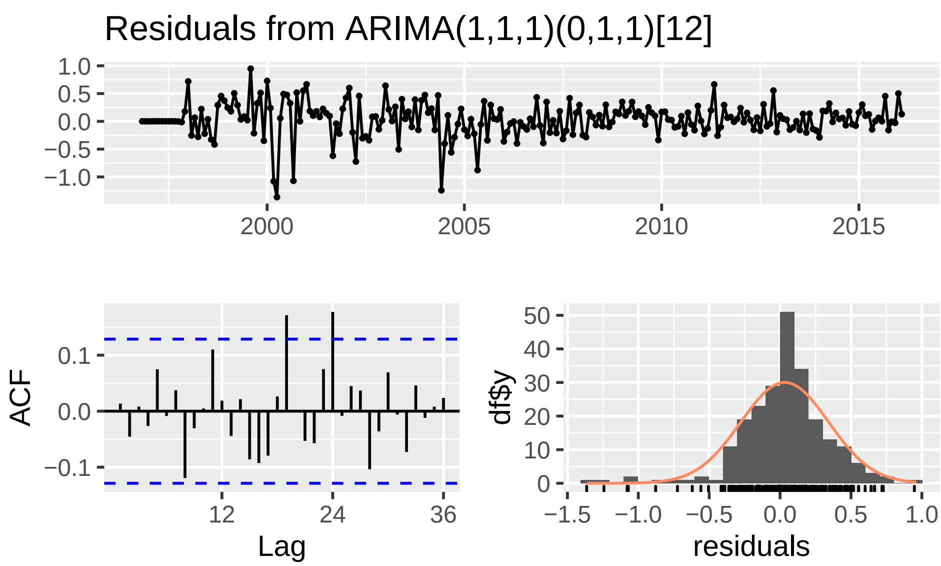
Temperature:

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,2)(0,1,1)[12]
## Q* = 48.802, df = 20, p-value = 0.000328
##
## Model df: 4.   Total lags used: 24
```



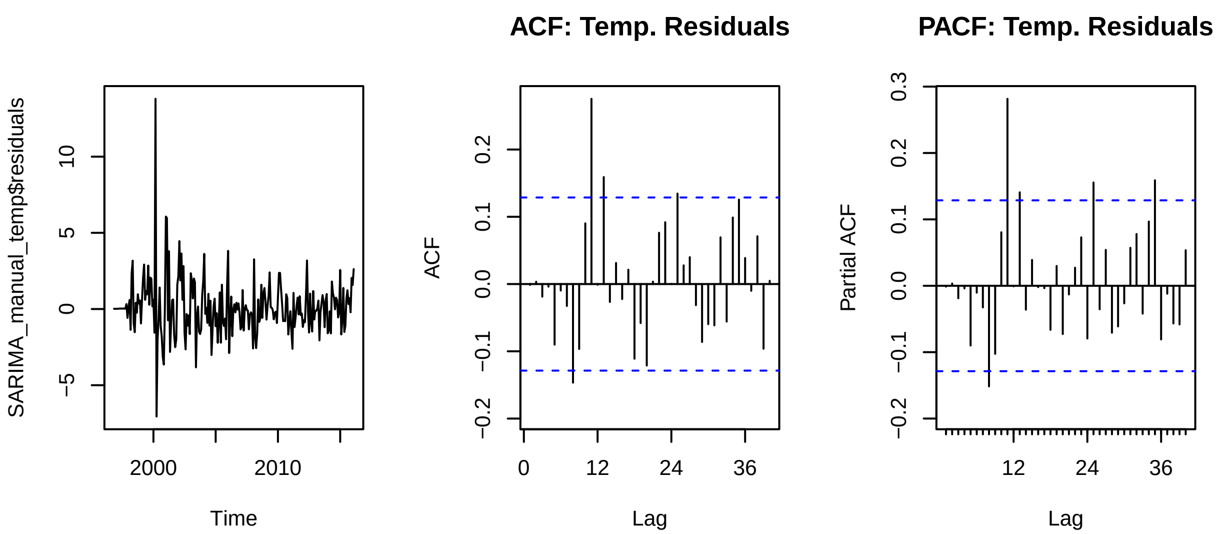
Dew Point Temperature:

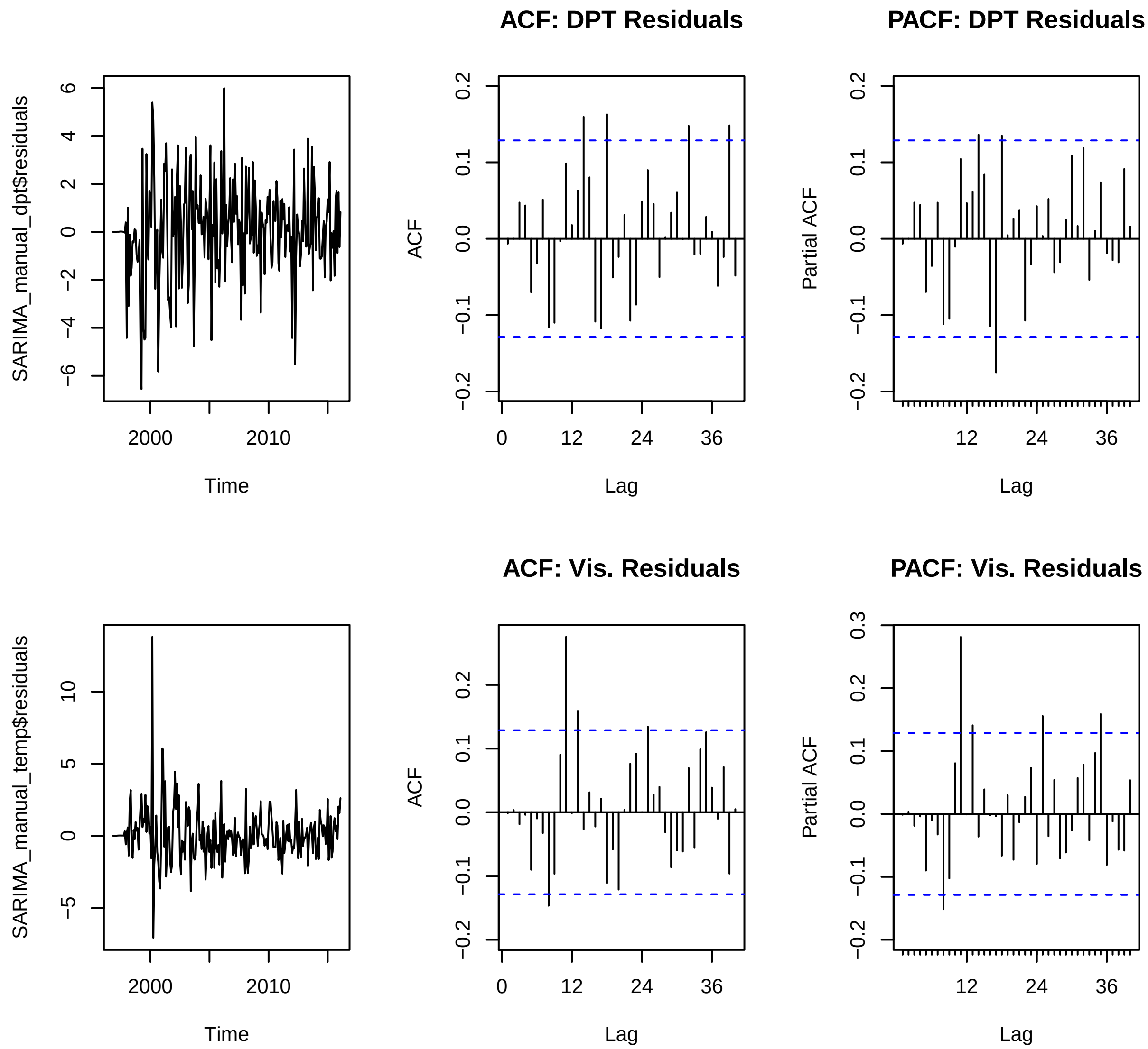
```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,1)(0,1,1)[12]
## Q* = 40.412, df = 21, p-value = 0.006625
##
## Model df: 3.   Total lags used: 24
```



Visibility:

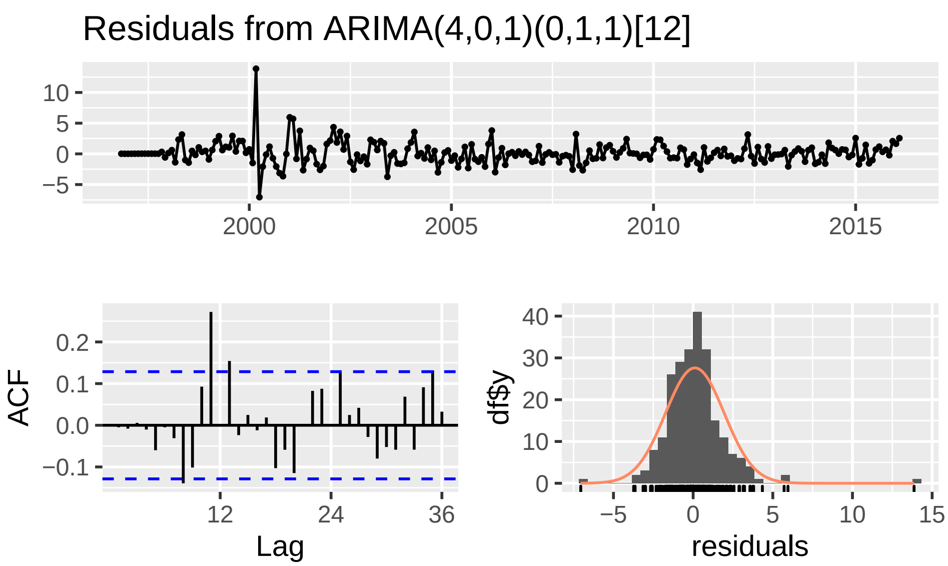
```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,1)(0,1,1)[12]
## Q* = 34.229, df = 21, p-value = 0.03425
##
## Model df: 3.   Total lags used: 24
```





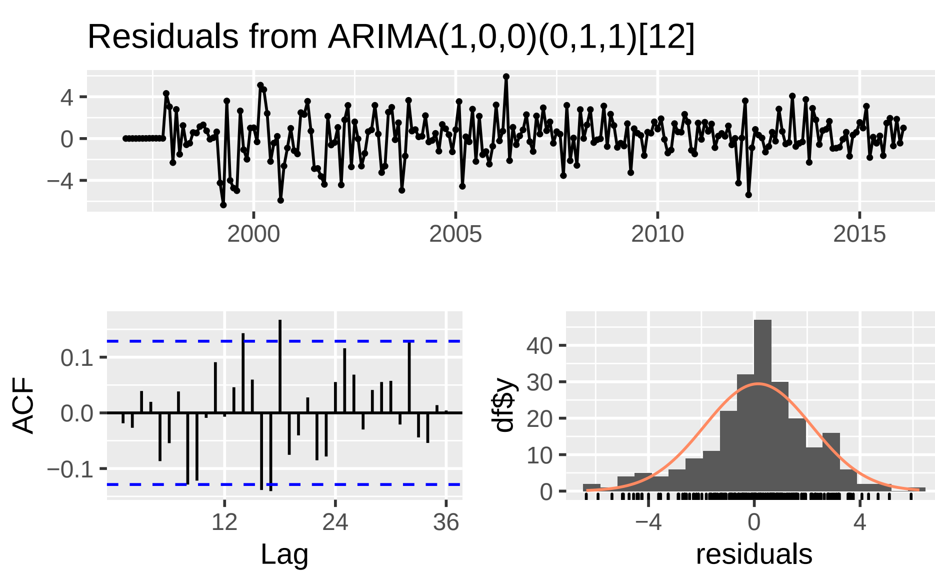
Automatically Fitting Seasonal ARIMA Models

Next we will fit seasonal ARIMA models using `auto.arima` and compare to the SARIMA models we fit manually above. It appears as though we still have some residual patterns, and the residual ACF plots look very similar to those of the manually fit SARIMA model.



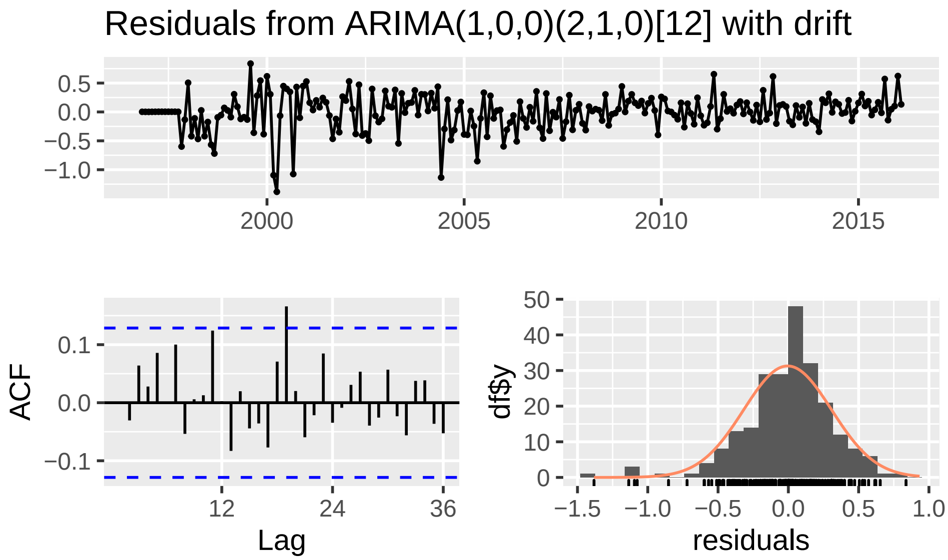
Temperature:

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(4,0,1)(0,1,1)[12]
## Q* = 45.652, df = 18, p-value = 0.0003334
##
## Model df: 6.   Total lags used: 24
```



Dew Point Temperature:

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0)(0,1,1)[12]
## Q* = 43.006, df = 22, p-value = 0.004711
##
## Model df: 2.   Total lags used: 24
```

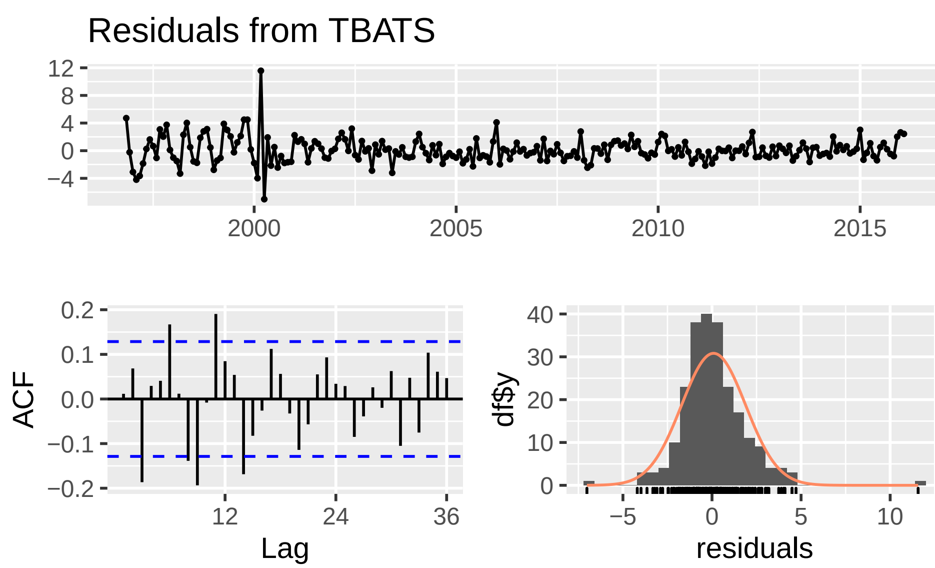


Visibility:

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0)(2,1,0)[12] with drift
## Q* = 25.816, df = 20, p-value = 0.172
##
## Model df: 4.   Total lags used: 24
```

TBATS Model

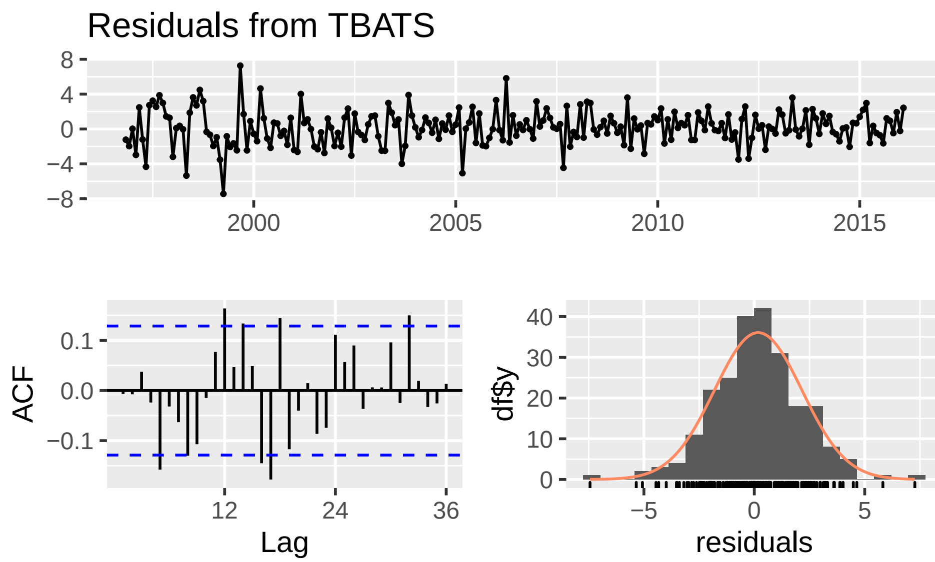
We have also tried TBATS models below, but as they do not fit well, we will not proceed with them for analysis.



Temperature:

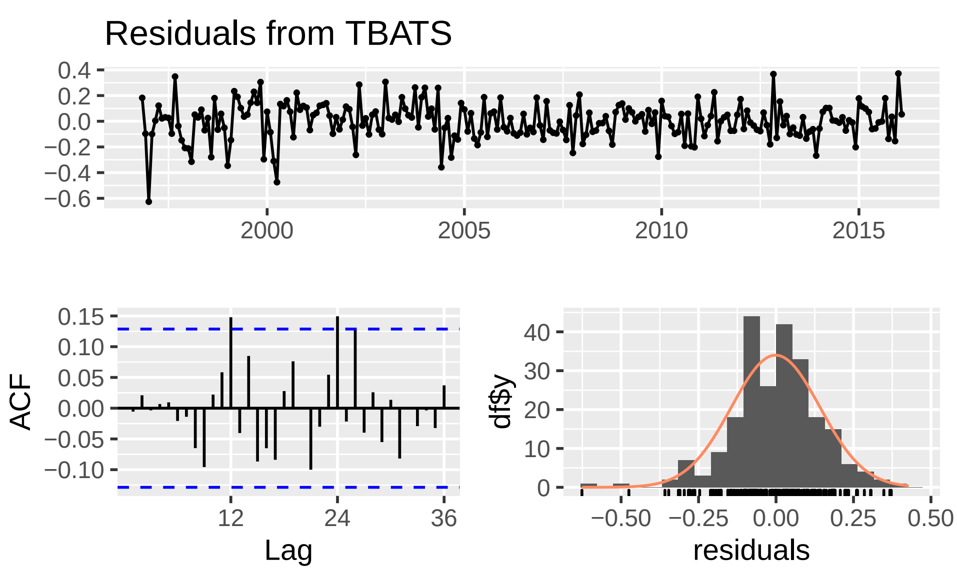
```
##
##  Ljung-Box test
##
## data:  Residuals from TBATS
```

```
## Q* = 65.146, df = 3, p-value = 4.663e-14
##
## Model df: 24.   Total lags used: 27
```



Dew Point Temperature:

```
##
##  Ljung-Box test
##
## data:  Residuals from TBATS
## Q* = 57.385, df = 6, p-value = 1.527e-10
##
## Model df: 18.   Total lags used: 24
```



Visibility:

```
##
##  Ljung-Box test
##
## data:  Residuals from TBATS
```

Table 3: Test MAPE for each model

Model	Temperature	Dew Point Temperature	Visibility
SARIMA Manual	15.56	17.36	20.27
SARIMA Autofit	6.44	8.97	17.26
TBATS	14.29	13.64	17.99

```
## Q* = 32.768, df = 3, p-value = 3.604e-07
##
## Model df: 24.   Total lags used: 27
```

Overall, we see that the residuals for the TBATS model on temperature appear to show some pattern and some seasonality, even though they do seem normally distributed. They actually look pretty similar to the residuals obtained from the SARIMA models above. The same appears to be true for the TBATS model on Dew Point Temperature as well as Visibility.

Comparing Accuracy of Models

Below we provide a comparison of the accuracy of each model using MAPE, which is the mean absolute percentage error. Recall that we trained our model on data from November 1996 to March 2016, and we use April 2016 to April 2017 as our test dataset. Thus, the MAPEs below are the MAPEs on the test dataset.

Based on the below table, we see that the SARIMA Autofit models have the lowest MAPE across the board, so we will go ahead and refit the SARIMA autofit models on all of the data from November 1996 to April 2017, and forecast from May 2017 to May 2020.

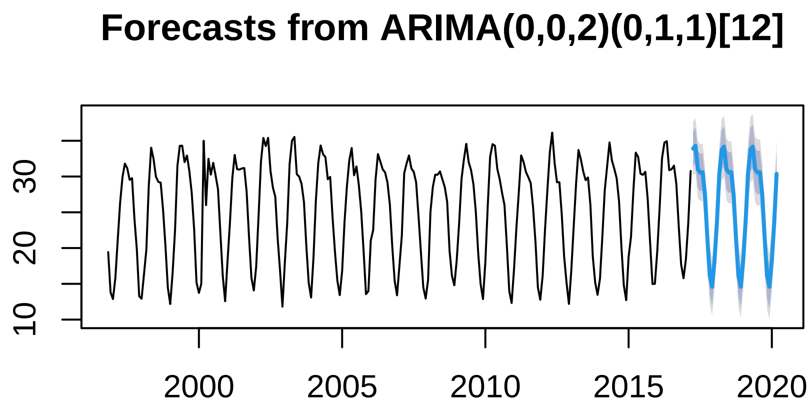
Summary and Conclusions

Forecasting Temperature, Dew Point Temperature, and Visibility

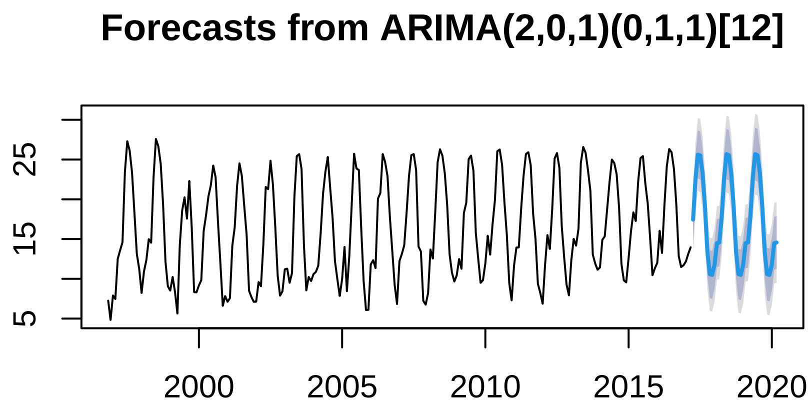
First we refit the models using all of the available data as mentioned above, and below we can see the three forecasts of temperature, dew point temperature, and visibility in New Delhi from May 2017 to May 2020.

For temperature in New Delhi, we do not see visual evidence from the plot forecast that temperatures in New Delhi are cooling. The forecasts plot from the dew point temperature plots suggest that in the colder months, dew point temperature is actually increasing slightly, though we cannot see any evidence of warmer months becoming cooler over time. For visibility, our forecast suggests that visiblity will continue to decrease over time.

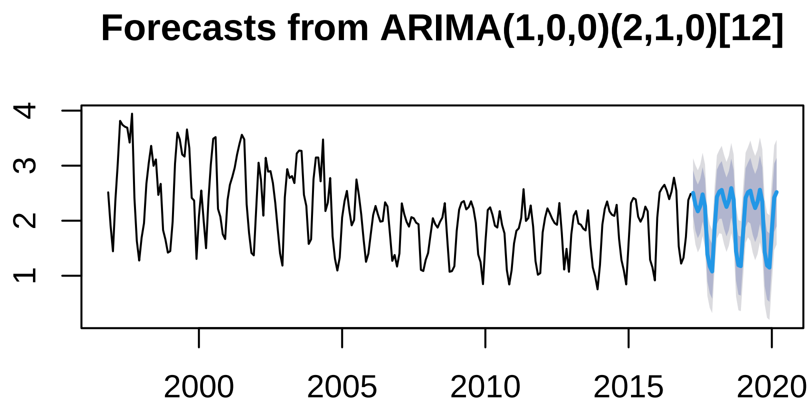
Temperature:



Dew Point Temperature:



Visibility:



There are of course a few limitations to note. First of all, the residuals of each of our models did suggest some violations of assumptions of our models, and thus I would not completely trust the forecasts from these models. In a more detailed analysis, I might try a few more models to see if I could find one that better represented the data. Additionally, the data used in this project is only until 2017. In a future analysis, I would like to obtain data up until the present and forecast a few years into the future to understand how temperature, dew point temperature, and visibility might change.

Sources

Tuholske, Cascade, et al. "Global urban population exposure to extreme heat." Proceedings of the National Academy of Sciences 118.41 (2021).

Zhao, Qi, et al. "Global, regional, and national burden of mortality associated with non-optimal ambient temperatures from 2000 to 2019: a three-stage modelling study." The Lancet Planetary Health 5.7 (2021): e415-e425.

Yadav, Sankalp, and Gautam Rawal. "The great Delhi smog." Indian Journal of Immunology and Respiratory Medicine 1.4 (2016): 78-79.

Prashar, Sunil, Rajib Shaw, and Yukiko Takeuchi. "Assessing the resilience of Delhi to climate-related disasters: A comprehensive approach." Natural Hazards 64.2 (2012): 1609-1624.

Nealon, Cory. "How will Climate Change Stress the Power Grid? Hint: Look at Dew Point Temperatures." News Center, University of Buffalo. https://www.buffalo.edu/news/releases/2018/09/030.html