Time Series Modeling for Temperature Forecasting

Prem Amal[1],

Abstract
The study involves the analysis of time series data of eight parameters related to temperature forecasting. To carry out the analysis, we utilized the statistical modeling techniques provided by the stats models and forecasting libraries. The analysis involved exploring the temporal properties of each variable, assessing their statistical stationarity, and performing appropriate transformations to ensure the models we fit are reliable.To fit the models, we used various time series modeling approaches, including ARIMAX, SARIMAX, and VAR models. We then compared the performance of these models using different goodness-of-fit metrics and selected the best-performing model for each variable.Furthermore, we employed the selected models to forecast the values of the variables for a future period and evaluated the accuracy of the predictions. The results of our analysis demonstrate the effectiveness of time series modeling for temperature forecasting, providing insights into the future trends of the analyzed variables.In summary, the study highlights the importance of utilizing time series models for temperature forecasting and provides a comprehensive guide for researchers and practitioners in the field. Additionally, the study emphasizes the significance of pre-processing, model selection, and evaluation stages in ensuring reliable and accurate temperature forecasting.

Contents

1. Introduction

1.1 Problem Statement

- The ability to accurately predict temperature based on weather conditions is critical for various industries and applications, from agriculture and energy to transportation and public health.
- In this study, the aim is to develop a predictive model for temperature forecasting using Time Series Analysis techniques.
- Specifically, the study objective is to determine how time series analysis can be used to forecast temperature and test a few hypotheses that can provide valuable insights for stakeholders. The analysis will examine the relationship between temperature and other weather parameters, such as humidity, wind, and dewpoint, and determine the extent to which these variables can be used to reliably forecast temperature.
- The study also involves plotting the time series of temperature data and summarizing the au-

tocorrelations, cross-correlations, and spectral analysis for the frequency domain. Additionally, various linear dynamic models will be fitted to the data and compared, including ARIMAX, SARIMAX, and VAR, to determine the most effective approach for forecasting temperature.
- The insights generated from this study can be used by policymakers, city planners, and other stakeholders to make informed decisions about resource allocation, infrastructure planning, and public safety measures in response to changing weather patterns.

1.2 Hypotheses

- H1: Temperature exhibits a significant seasonal pattern over the years.
- H2: Temperature at a given day is positively correlated with the temperature on the previous day.
- H3: Temperature is influenced by various weather parameters, such as humidity, dew point, and wind speed, as well as their past values, including the lagged temperature.

2. Data Collection

2.1 Dataset

The dataset used in this analysis contains daily weather data for Boston from July 2012 to August 2015. The purpose of the weather dataset could be to provide information on the daily weather conditions in Boston over a three-year period. This information can be useful for a variety of purposes, such as:

Climate research: The dataset can be used to study long-term trends in temperature, precipitation, and other weather variables in Boston, which can help in understanding climate patterns in the region.

Agricultural planning: The dataset can be used by farmers to plan irrigation schedules, determine the best time to plant crops, and predict weather-related risks such as drought or flooding.

The purpose of the dataset colud be to provide a valuable resource for researchers, policymakers, and other stakeholders who are interested in understanding weather patterns and their impact on various aspects of society.

2.2 Data Description

In this study, data was collected from Kaggle, specifically from the ”Weather Data - Boston Jul 2012-Aug 2015” dataset provided by Naveen Pandian. The dataset contains daily weather measurements, including temperature, humidity, dew point, wind, precipitation, and date, from July 2012 to August 2015. The data was collected from a weather station located in Boston, Massachusetts, USA. The dataset is publicly available and was downloaded in CSV format for further analysis.The dataset consists of 1157 rows of daily data and 8 features:

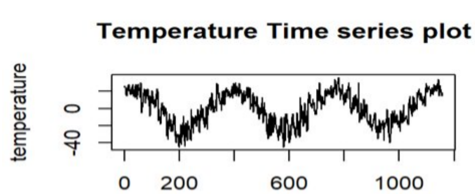
- Precipitation: The amount of rainfall in millimeters
- Day: Day of the month
- Month: Month of the year
- Year: Year
- Temperature: The average temperature over the day in Fahrenheit
- Dewpoint: The average dew point over the day in Fahrenheit
- Humidity: Relative percent amount of moisture in the air.
- Wind: The average wind speed in miles per hour.

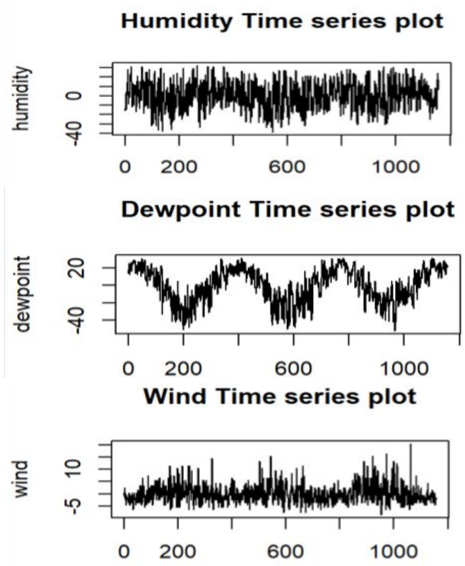
precipitation	day	month	year	temperature	dewpoint	humidity	wind
1	1	8	2015	79	58	50	10
0	2	8	2015	78	54	52	11
0	3	8	2015	79	64	67	13
0.49	4	8	2015	78	66	68	9
0	5	8	2015	75	57	58	11
0	6	8	2015	74	53	51	9
0	7	8	2015	69	57	70	8
0	8	8	2015	69	57	68	6
0	9	8	2015	68	59	80	9
0	10	8	2015	73	59	67	10
0.83	11	8	2015	68	67	94	10
0	12	8	2015	77	63	70	9

3. Methodology and Results

3.1 Time Series Plot

- The images below show the original Time Series plots for temperature, humidity, dewpoint and wind





- The plotted graphs of temperature and dewpoint time series data reveal significant trends and seasonal patterns.
- This proves Hypothesis 1 (H1) Temperature exhibits a significant seasonal pattern over the years to be correct.
- The mean and variance of these features do not appear to be constant, suggesting that the data may be non-stationary. On the other hand, the plots of humidity and wind show some visual signs of stationarity.
- To confirm whether the data is stationary or non-stationary, different statistical tests, such as the ADF test and KPSS test, are conducted.

3.2 Stationarity Checks

ADF (Augmented Dickey-Fuller) Test

- The ADF test is employed to assess the stationarity of a time series. It determines if the series possesses a unit root, which suggests non-stationarity. The test produces a p-value that is used to assess the statistical significance of the results.
- The null hypothesis postulates the existence of a unit root in the time series, implying non-stationarity. Conversely, the alternative hypothesis assumes that the time series is stationary and does not have a unit root.
- If the p-value is less than a pre-determined significance level (usually 0.05), the null hypothesis is rejected indicating the time series is not non-stationary.
- To confirm, if the time series is stationary, we perform KPSS test.

KPSS (Kwiatkowski-Phillips-Schmidt-Shin) Test

- The KPSS test is used to assess the stationarity of a time series. It determines if the series has a trend or not, which suggests non-stationarity. The test produces a p-value that is used to assess the statistical significance of the results.
- The null hypothesis assumes that the time series is stationary, while the alternative hypothesis assumes that the time series has a trend, indicating non-stationarity.
- If the p-value is less than or equal to the significance level, the null hypothesis is rejected, suggesting non-stationarity in the time series.

Test Results

- Temperature

```
> adf.test(temperature)

	Augmented Dickey-Fuller Test

data:  temperature
Dickey-Fuller = -2.7497, Lag order = 10, p-value = 0.2609
alternative hypothesis: stationary
```

  ADF test has a p-value of 0.2 which is greater than 0.05, hence we cannot reject the null hypothesis of non-stationarity. Performing KPSS test

```
> kpss.test(temperature)

	KPSS Test for Level Stationarity

data:  temperature
KPSS Level = 0.44655, Truncation lag parameter = 7,
p-value = 0.05709
```

  The KPSS test results in a p-value of 0.05 hence we can reject the null hypothesis stating time series is stationary. From the above two tests, it can be seen that Temperature is non-stationary.

- Humidity

```
> adf.test(humidity)

	Augmented Dickey-Fuller Test

data:  humidity
Dickey-Fuller = -8.776, Lag order = 10, p-value =
0.01
alternative hypothesis: stationary
```

ADF test has a p-value of 0.01 which is less than 0.05, hence we can reject the null hypothesis of non-stationarity. This shows that Humidity is stationary.

- Dewpoint

```
> adf.test(dewpoint)

        Augmented Dickey-Fuller Test

data:  dewpoint
Dickey-Fuller = -3.1332, Lag order = 10, p-value =
0.09943
alternative hypothesis: stationary
```

ADF test has a p-value of 0.09 which is greater than 0.05, hence we cannot reject the null hypothesis of non-stationarity giving results similar to Temperature . Performing KPSS test

```
> kpss.test(dewpoint)

        KPSS Test for Level Stationarity

data:  dewpoint
KPSS Level = 0.46798, Truncation lag parameter = 7,
p-value = 0.04888
```

The KPSS test results in a p-value of 0.04 hence we can reject the null hypothesis stating time series is stationary. From the above two tests, it can be seen that Dewpoint is non-stationary.

- Wind

```
> adf.test(wind)

        Augmented Dickey-Fuller Test

data:  wind
Dickey-Fuller = -8.6165, Lag order = 10, p-value =
0.01
alternative hypothesis: stationary
```

ADF test has a p-value of 0.01 which is less than 0.05, hence we can reject the null hypothesis of non-stationarity. This shows that wind is stationary.

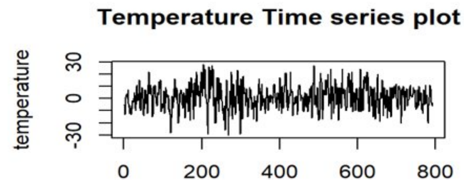
3.3 Differencing

- Differencing is a used to remove the trend and/or seasonality from time series. Here, differencing of order 1 is performed with lag 365 as the temperature and dewpoints are following an yearly trend.
- Differencing involves computing the difference between consecutive observations in a time series. The first difference of a time series is the difference between each observation and the previous observation one year ago.
- The process of differencing removes the trend and seasonality from the time series data, leaving behind only the random fluctuations or noise. This makes it easier to model and forecast the time series.

Results

- Temperature

  Time Series Plot



From the time series plot, it can be seen that after differencing there is no seasonality.

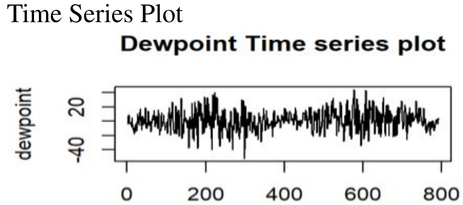
ADF Test

```
> adf.test(diff_temperature)

        Augmented Dickey-Fuller Test

data:  diff_temperature
Dickey-Fuller = -8.633, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary
```

ADF test has a p-value of 0.01 which is less than 0.05, hence we can reject the null hypothesis of non-stationarity. This shows that after differencing temperature has become stationary.

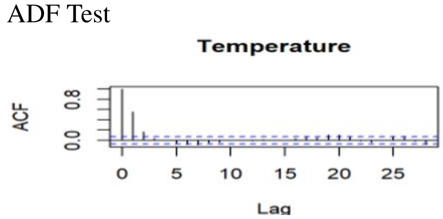
- Dewpoint

  Time Series Plot



From the above plot , we can see that there is no pattern/cycle or trend in the dewpoint after differencing.

ADF Test

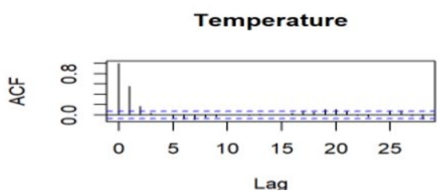


ADF test has a p-value of 0.01 which is less than 0.05, hence we can reject the null hypothesis of non-stationarity. As evidenced by the ADF test results, differences dewpoint is stationary. This suggests that the trend and seasonality have been successfully removed through differencing.

3.4 Correlations

Auto Correlations

- Autocorrelation is the correlation between a variable and its own lagged values. In other words, it measures the degree to which a variable is correlated with its past values.
- A positive autocorrelation indicates that observations at a particular lag are positively correlated, while a negative autocorrelation indicates that observations at a particular lag are negatively correlated. A value of zero indicates no correlation.
- A common approach to interpreting an ACF plot is to look for significant autocorrelations

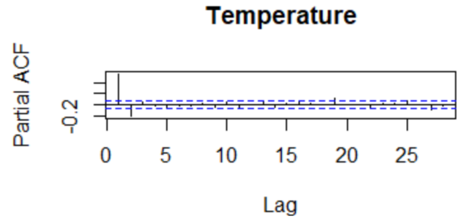


Results

- There are 3 positive spikes at lag=0,1,2, depicting a positive correlation.
- From the above ACF plots, we can see that after lag 2, there are no significant spikes showing temperature, a MA model of order 2.
- This provides evidence in support of hypothesis 2 (H2), which suggests that there is a positive correlation between the temperature on a particular day and the temperature on the previous day.

Partial Auto Correlations

- The PACF measures the correlation between the time series and its past lags after removing the effects of intervening lags. In other words, it captures the unique correlation between the time series and its past lags without the influence of other lags.
- A significant spike in the PACF plot at lag k suggests that the time series may be modeled as an AR model of order k.



Results

- There are 2 significant spikes at lag=1,2 showing temperature, an AR model of order 2.

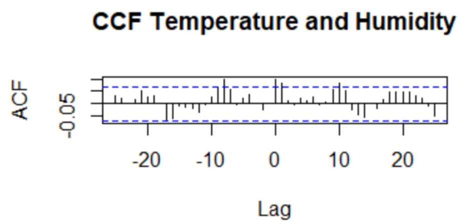
Cross Correlations

- Cross-correlation is a statistical tool used to measure the similarity between two-time series as a function of the time lag applied to one of them. It helps to identify if there is a relationship between two-time series and the lag between them.
- Cross-correlation is calculated by sliding one time series (the ”target” series) over another time series (the ”reference” series) and calculating the correlation coefficient between them at each lag.
- A positive lag indicates that the second variable is driving the first variable, while a negative lag indicates that the first variable is driving the second variable. On the positive lag side, if the spike is statistically significant and upwards that means the second variable is positively correlated to the first variable at that lag. An increase in the second variable will increase the value of the first variable i.e. positive correlation at that lag. In case of a downward spike, the decrease in the second variable will lead to an increase in the first variable ie a negative correlation at that lag.
- On the negative lag side, if the spike is statistically significant and upwards that means the first variable is positively correlated to the second variable at that lag, and the first variable is driving the second variable. An increase in the first variable will increase the value of the second variable i.e. positive correlation at that lag. In case of a downward spike, the decrease in the first variable will lead to an increase in the second variable ie a negative correlation at that lag.

Results

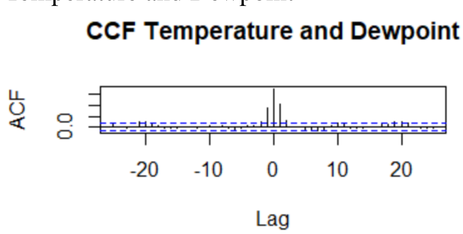
Temperature is the first variable and other features are the second variable, hence we are only interested in the positive lags showing how the second variable is driving the temperature.

- Temperature and Humidity



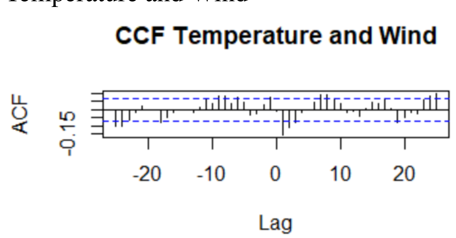
  At lag=0 we have a statistically significant positive spike showing at the same timepoint temperature and humidity are positively correlated. On positive lag sides, we have statistically significant positive spikes At lag= 1,10 showing humidity is driving the temperature and it is positively correlated with temperature.

- Temperature and Dewpoint



  At lag=0 we have a statistically significant positive spike showing at the same timepoint temperature and humidity are positively correlated. On positive lag sides, we have statistically significant positive spikes At lag= 1,2,20 showing dewpoint is driving the temperature and it is positively correlated with temperature.

- Temperature and Wind

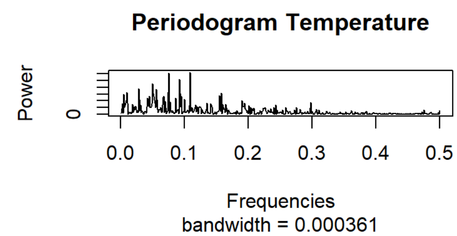


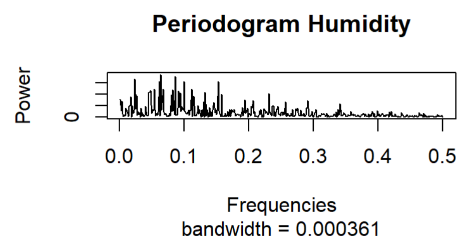
  On positive lag sides, we have statistically significant negative spikes At lag= 1,2,3,7,8,19 showing wind is driving the temperature and it is negatively correlated with temperature.

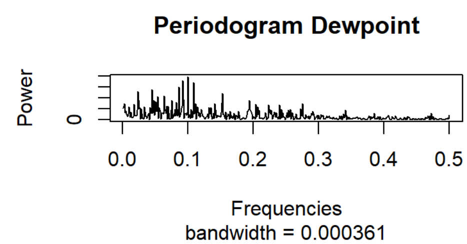
3.5 Spectral Analysis

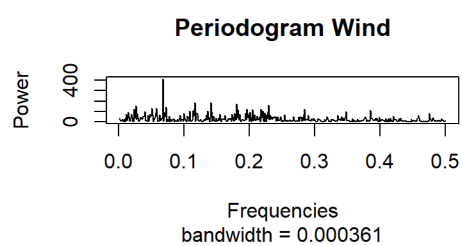
- Spectral analysis is a method of analyzing signals in the frequency domain, which involves decomposing a time series into its underlying frequency components.
- A periodogram is a tool used in spectral analysis to estimate the power spectral density (PSD) of a signal. The PSD represents the distribution of power of the signal across different frequencies.

Results









- The presence of multiple peaks in the periodogram suggests that the data contains several periodic components or cycles. The peak shows a high amplitude/ rise in the initial frequencies, then the peak then decays gradually.
- The initial peak at a certain frequency may indicate the presence of a dominant frequency

component in the data, while the smaller peaks at other frequencies could represent additional, less dominant frequency components.

3.6 Model Selection and Fitting

The order of the AR and MA components of a time series can be estimated using ACF, PACF, and CCF plots. Additionally, the CCF plot can be used to investigate the relationships and correlations between a target variable, such as temperature, and other features such as humidity, dew point, and wind. However, it is important to note that the output of correlations should not be solely relied upon, hence fitting various models with different lagged inputs and different orders of lags.

ARIMAX

Four different models have been created, namely:

- Model 1 (Complex Model with three lags)
- Model 2 (Complex Model with two lags)
- Model 3 (Average Model with a single lag)
- Model 4 (Noise Model)

The model incorporates features such as humidity, dew point, and wind, with their lagged values up to lag 3. Furthermore, temperature with lagged values of 1, 2, and 3 are also included as input features in order to forecast the temperature.

Observations and Results

For each model, all the possible combinations of AR, MA, and ARMA models are tried from orders 0 to 6. AIC, BIC metric is used to evaluate the best models. The best combination for each model is shown below.

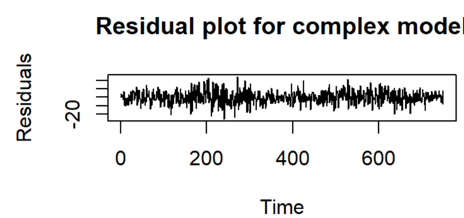
Models	Lagged Input	AR order	MA order
Model 1	3	5	2
Model 2	2	3	1
Model 3	1	5	5
Model 4	0	1	1

```
> AIC(complex_model_3lag)
[1] 5218.347
> BIC(complex_model_3lag)
[1] 5315.369
> AIC(complex_model_2lag)
[1] 5209.694
> BIC(complex_model_2lag)
[1] 5274.375
> AIC(average_model_1lag)
[1] 5200.05
> BIC(average_model_1lag)
[1] 5273.971
> AIC(simple_model)
[1] 4903.312
> BIC(simple_model)
[1] 4921.517
```

Based on the above results of AIC, and BIC, it can be seen that Average_model_lag_1 has the lowest AIC and BIC values. Also, the values are comparable to the complex_model_lag_2. Selecting the complex model for predicting/forecasting the temperature.

The dataset is divided into a training set and a testing set. Fitting the above model with training set data.

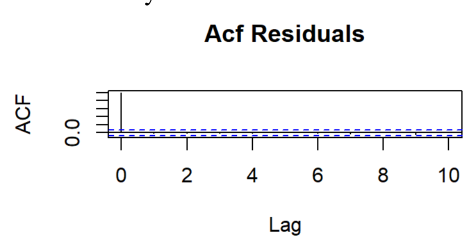
Checking residuals a white noise



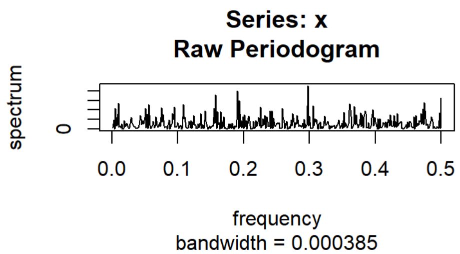
The plot does not show any seasonality and even seems stationary which can be confirmed using the ADF test.

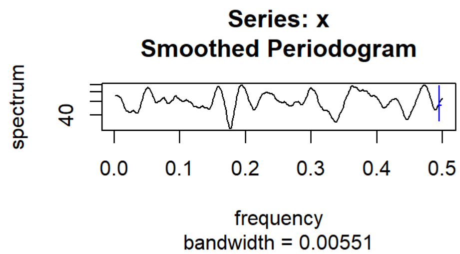
```
> adf.test(res_complex)

        Augmented Dickey-Fuller Test

data:  res_complex
Dickey-Fuller = -8.7951, Lag order = 9, p-value =
0.01
alternative hypothesis: stationary
```

Rejecting the null hypothesis, proving the residuals are stationary.



The autocovariance plot up to lag 10 has been plotted. The plot shows no significant lags. No lines are above or below the two blue lines i.e. at any lag the autocovariance among the residual show no significant values.



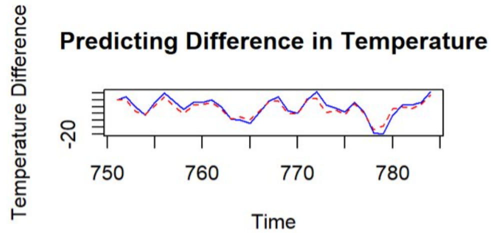


We see no pattern after smoothing the curve. From the periodogram and the smoothened periodogram, we can see some spikes at irregular intervals with no pattern or seasonality being captured in the residuals, showing characteristics of noise. Also, ACF confirms the same as there are no significant spikes. Hence, it seems that residuals are white noise.

Prediction/Forecasting

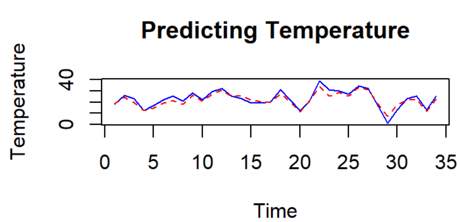
```
Call:
arima(x = diff_temperature[0:750], order = c(3, 0, 1), xreg = data_ari
ma_1[0:750,
    ])

Coefficients:
          ar1      ar2      ar3     ma1  intercept
      -0.9869  -0.4038  -0.1984  0.7036     0.3691
s.e.   0.1813   0.1636   0.0667  0.1879     0.1914
      temperature_lag1  temperature_lag2  humidity_lag1
                1.0329           -0.2798         0.0443
s.e.            0.1664            0.1389         0.0457
      humidity_lag2  dewpoint_lag1  dewpoint_lag2  wind_lag1
             0.0030        -0.0771        -0.0254     -0.289
s.e.         0.0454         0.0963         0.0962      0.057
      wind_lag2
         0.1363
s.e.     0.0634

sigma^2 estimated as 58.61:  log likelihood = -2590.85,  aic = 5209.69
```

Based on the standard error associated with each feature, we have determined that some of the features in the model are statistically significant, while others are not. The statistically significant parameters include Ar1, Ar2, Ar3, Ma1, Temperature_lag1, Temperature_lag2, Wind_lag1, and Wind_lag2. On the other hand, the parameters that are not statistically significant include Humidity_lag1, Humidity_lag2, Dewpoint_lag1, Dewpoint_lag2, and the intercept. We also observed that temperature at lag 1 has a positive contribution to the model, indicating that with every increase in one unit of temperature lag 1, the temperature will increase positively. On the other hand, wind at lag 1 has a negative contribution to the model, which means that with every increase in one unit of wind lag 1, the temperature will decrease.



- The blue line represents the actual values, and the red line represents the predicted values.
- The plot above depicts the predicted values of the temperature differences. The temperature difference is defined as the temperature on a specific day of a given year minus the temperature on the same day of the previous year. For instance, the temperature difference for January 1st, 2014 would be calculated as the temperature on January 1st, 2014 minus the temperature on January 1st, 2013.
- The plot of predicted values shows a similar trend to the actual values, and the magnitude of the predicted values is also very close to that of the actual values.
- This observation is further supported by the low values of the MAE, MSE, and RMSE metrics, which indicate that the predictions are accurate and have small errors compared to the actual values.

```
> cat("MAE:", round(mae, 2), "\n")
MAE: 2.19
> cat("MSE:", round(mse, 2), "\n")
MSE: 7.09
> cat("RMSE:", round(rmse, 2), "\n")
RMSE: 2.66
```

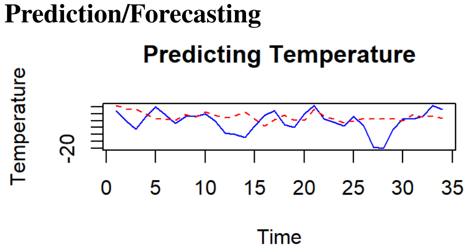


The above plot shows the temperature forecasted for a month. It can be seen that plot follows the trend and even the magnitude of the actual temperature.

VAR

- VAR (vector autoregression) model is fitted on the time series data that includes the differences of temperature, humidity, dew point, and wind. The data is split into training and testing sets, with the first 750 observations used for training and the remaining 34 observations used for testing.
- The VAR model is fitted on the training data using the VAR function from the vars package. The model has a lag order of 50 and a constant term. Then, the predict function is used to forecast the temperature values for the next 34 time points (i.e., the testing data).
- The model estimated the relationships between temperature, humidity, wind, and dewpoint and used them to predict the temperature values for future time periods.
- After testing various lag values ranging from 5 to 100, we identified that a lag value of 50 provided the lowest root mean square error (RMSE) among all the tested values. As a result, we selected the VAR model with a lag value of 50 for our analysis.
- We used a lag order of 50 to account for the potential long-term dependencies between the variables. This means that the model considers the values of all four variables from the previous 50 time periods to make a prediction for the current temperature value.
- Three metrics, MAE (mean absolute error), MSE (mean squared error), and RMSE (root mean squared error), are calculated to evaluate the accuracy of the predictions

Prediction/Forecasting



The actual and predicted temperature values for the testing period are plotted. The blue line represents the actual values, and the red line represents the predicted values. The initial part of the plot exhibits a resemblance to the trend in the data, and the magnitudes of the predicted values align with the actual values. However, there is a deviation at later lags. This is also supported by the RMSE, MSE, MAE metrics.

```
> cat("MAE:", round(mae_var, 2), "\n")
MAE: 6.31
> cat("MSE:", round(mse_var, 2), "\n")
MSE: 68.95
> cat("RMSE:", round(rmse_var, 2), "\n")
RMSE: 8.3
```

The coefficients of the model indicate the strength and direction of the relationships between the variables.

```
> coef(var_model)[[1]]
                        Estimate Std. Error      t value
diff_temperature.l1   0.7827075371 0.11781435  6.643567265
diff_humidity.l1      0.0263370688 0.05758090  0.457392469
diff_dewpoint.l1     -0.0570487867 0.12030992 -0.474181906
diff_wind.l1         -0.2394267809 0.07178324 -3.335413458
diff_temperature.l2  -0.2682248959 0.12313714 -2.178261560
diff_humidity.l2     -0.0161870869 0.05785031 -0.279809833
diff_dewpoint.l2     -0.0296620421 0.12053146 -0.246093780
diff_wind.l2          0.0664698010 0.07606673  0.873835418
```

A positive coefficient for the temperature variable at lag 1 would indicate that an increase in temperature lag 1 lead to an increase in temperature. Similarly, a negative coefficient for the wind variable at lag 1 would indicate that an increase in wind speed leads to a decrease in temperature.

```
                        Pr(>|t|)
diff_temperature.l1  8.028217e-11
diff_humidity.l1     6.475880e-01
diff_dewpoint.l1     6.355776e-01
diff_wind.l1         9.151670e-04
diff_temperature.l2  2.985398e-02
diff_humidity.l2     7.797394e-01
diff_dewpoint.l2     8.057108e-01
diff_wind.l2         3.826282e-01
```

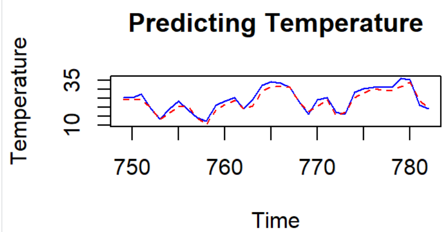
A p-value less than 0.05 is considered statistically significant, indicating that the estimated coefficient is unlikely to have arisen by chance. From the above figure, it can be seen that all the coefficients are statistically significant.

SARIMAX

- An ARIMA model is fitted to the training temperature data with external regressors. It is using the auto Arima function from the forecast package to automatically select the order of the ARIMA model based on minimizing the AIC criterion. The external regressors are the features of complex_model_2lag.
- The seasonal parameter is set to TRUE to indicate that the data has a seasonal pattern.
- Then, the model is used to predict the temperature for the next 34 time points and

the predictions are compared with the actual temperature values.

Prediction/Forecasting



- The blue line represents the actual values and the red line shows the predicted values.
- The predicted values exhibit a trend similar to that of the actual values, and their magnitudes are in close proximity as well.
- Moreover, this inference is reinforced by the low values of MAE, MSE, and RMSE metrics, indicating the accuracy of the predictions and small errors in comparison to the actual values.

```
> cat("MAE:", round(mae, 2), "\n")
MAE: 1.78
> cat("MSE:", round(mse, 2), "\n")
MSE: 4.64
> cat("RMSE:", round(rmse, 2), "\n")
RMSE: 2.15
```

3.7 Conclusion

- After evaluating the performance of ARIMAX, SARIMAX,and VAR models based on RMSE, MAE, MSE, and the comparison plot of predicted versus actual values, it can be concluded that the SARIMAX model outperforms the VAR model in temperature prediction.
- Based on the ARIMAX and SARIMAX model that was developed, it can be concluded that there is a significant correlation between the variables of lagged temperature, humidity, dew point, and wind speed with the temperature in Boston.
- The results suggest that the model provides a reasonable fit to the data and is a suitable model for predicting the temperature in Boston.
- Therefore, the initial hypothesis that the temperature in Boston is influenced by the lagged temperature, humidity, dew point, and wind speed is supported by the ARIMAX model.

4. References

- https://www.kaggle.com/datasets/naveenpandianv/weather-data-boston-jul-2012-aug-2015
- https://www.diva- portal.org/smash/get/diva2:1668033/FULLTEXT01.pdf
- https://jbusemey.pages.iu.edu/time/time_series.htm