# Time Series Report on
# Temperature Prediction of Austin

**Akshay Kumar Varanasi**

## Introduction

Accurate weather prediction is important for planning our day to day activities. People have tried to predict weather since many decades ago. Weather prediction can be used in many places like for scheduling flights or events, for farmers to plan their planting and harvest. It can be used to warn people or take actions in case of weather calamities even before they occur.

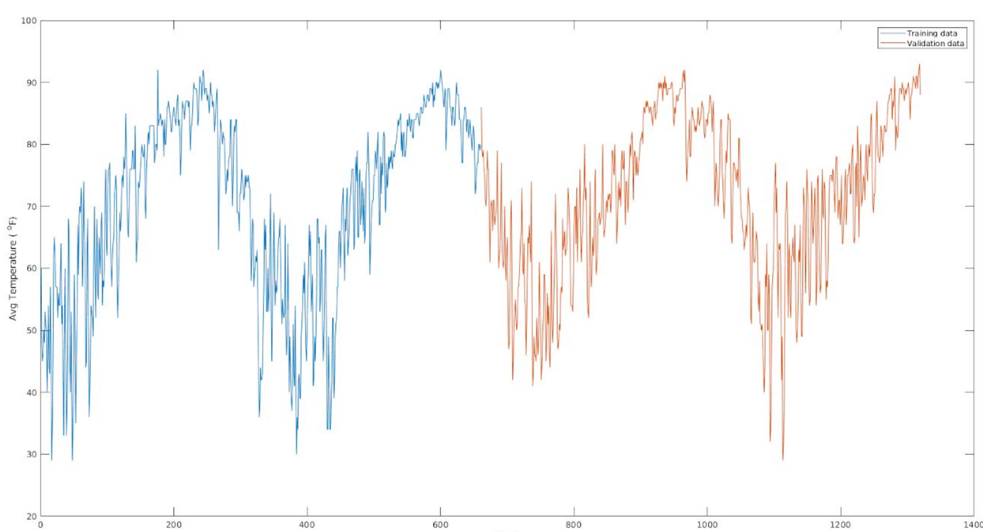
My attempt through this project is to use Time series analysis to predict temperature of city Austin. In our case, we are predicting Temperature which can be used for daily activities like whether it will be cool or hot tomorrow? Do we need to wear sweater or T-shirt? etc.

## Dataset

For this project, Austin Weather Dataset from Kaggle was used which was obtained from WeatherUnderground.com, at the Austin KATT station.
https://www.kaggle.com/grubenm/austin-weather

This dataset contains data for every date from 2013-12-21 to 2017-07-31. (1319 data points). For training, we used 50% data, which is around 660 data points and validated on the remaining data i.e used to check our forecast was good or not.

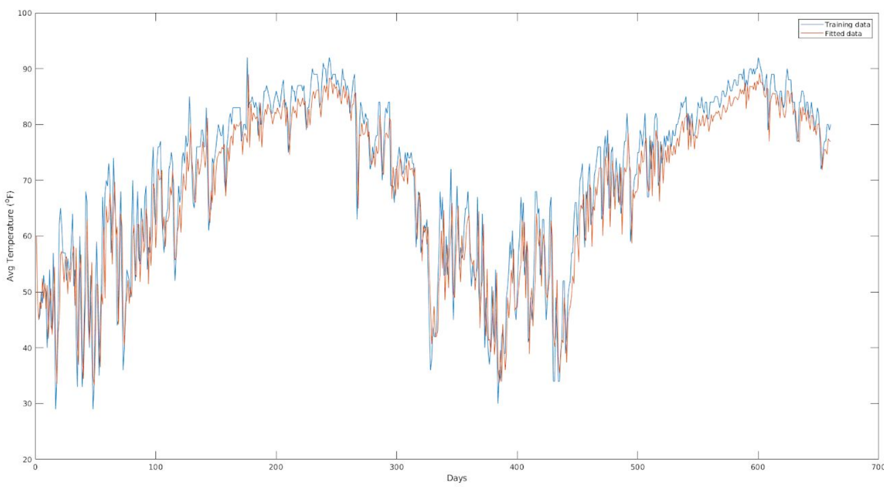
**Plot of Training and Validation Data of Temperature**



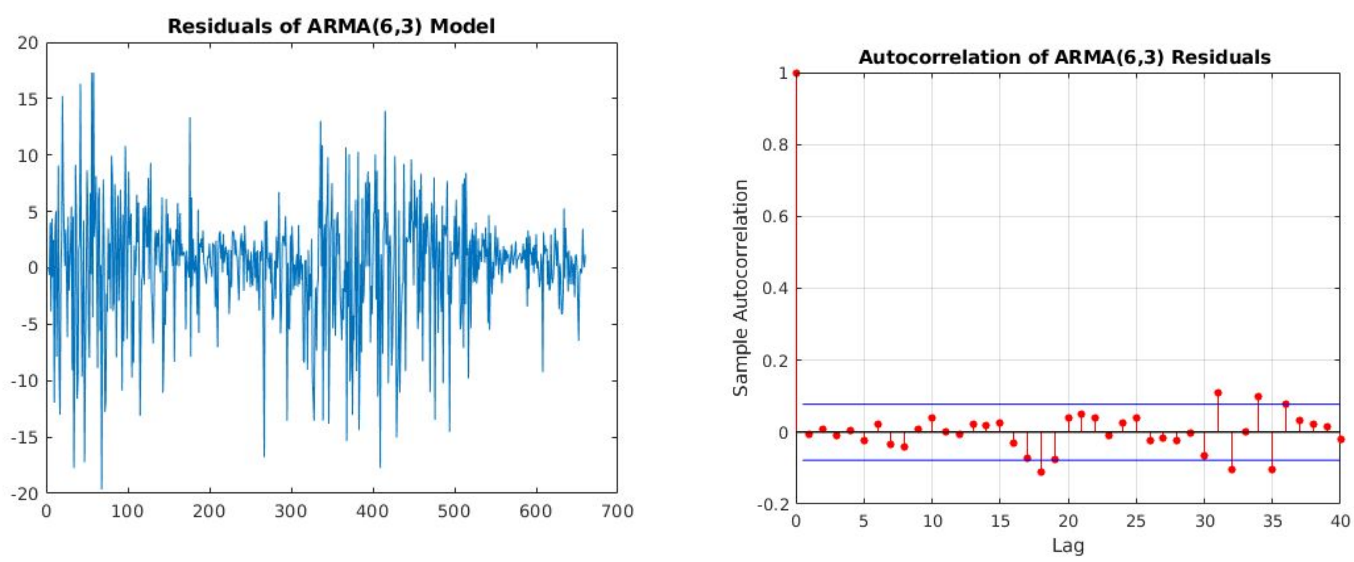
## Stationary Model

First, we tried to fit Stationary Time Series model using ARMA(2n,2n-1) technique and F-test was used as the model criterion. In this model our mean is constant with respect to time since we assume time series is stationary and our data varies around it. When, we use ARMA(2n,2n-1) technique we get ARMA(6,3) model as adequate with RSS value 1.7312e+04.

**Plot shows training data and data fitted using ARMA(6,3)**



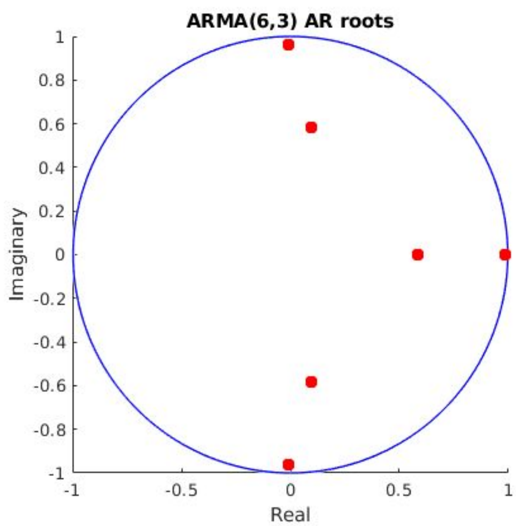

**Plot shows ARMA(6,3) residuals and residual autocorrelation**

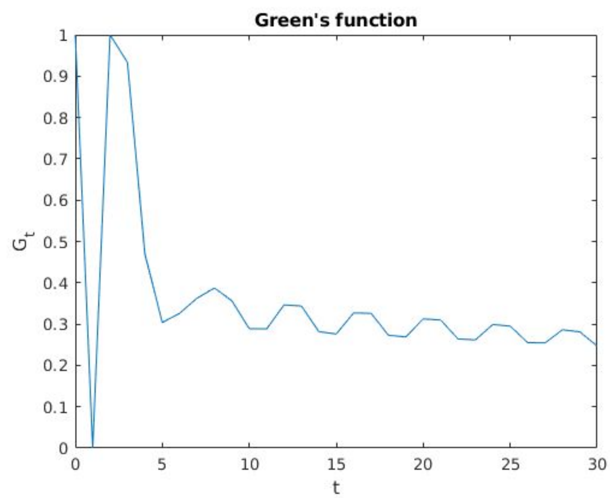


After we get the the model, we look at the AR roots. Out of 6 roots, 2 roots are real and 4 are complex. Among real roots, 1 root is 0.9903 ≅ 1 so stochastic trend was checked it was found that it exits. RSS=1.7374e+04. When we look at complex roots, they give period of 3.9748 ≅ 4 and 4.4614 ≅ 4.5. Seasonality of these periods was checked and it was found that it does not exists. It makes sense as temperature does not repeat every 4/5 days.

From the green's function, we see that the system is stable.

**Plot shows ARMA(6,3) AR roots on unit circle and Green's function**

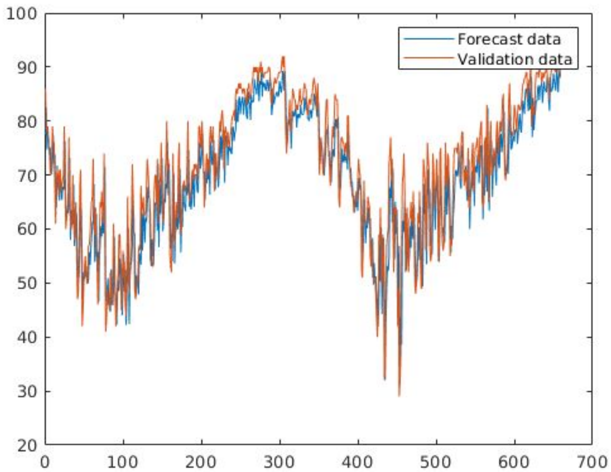


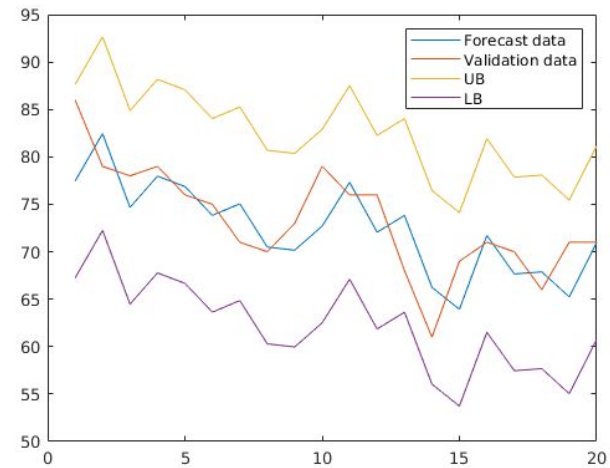




## Stationary Model Forecast

We would like to know how good is the model by forecasting and checking with the validation data. We see that stationary model forecasts reasonable well, when we do rolling 1-step ahead forecast as the RSS = 1.9632e+04,STD=5.2048, MSE=27.09 values are reasonable.

**Plot shows 1-step rolling forecast on validation data and zoomed version for 20 points**




## Non Stationary Model

In Non Stationary Model, we assume mean is not constant with time. So first we try to fit the data with different types of curves like linear, polynomial, sinusoidal etc. We find that following sinusoidal curve is a good fit with RMSE = 7.2470 .

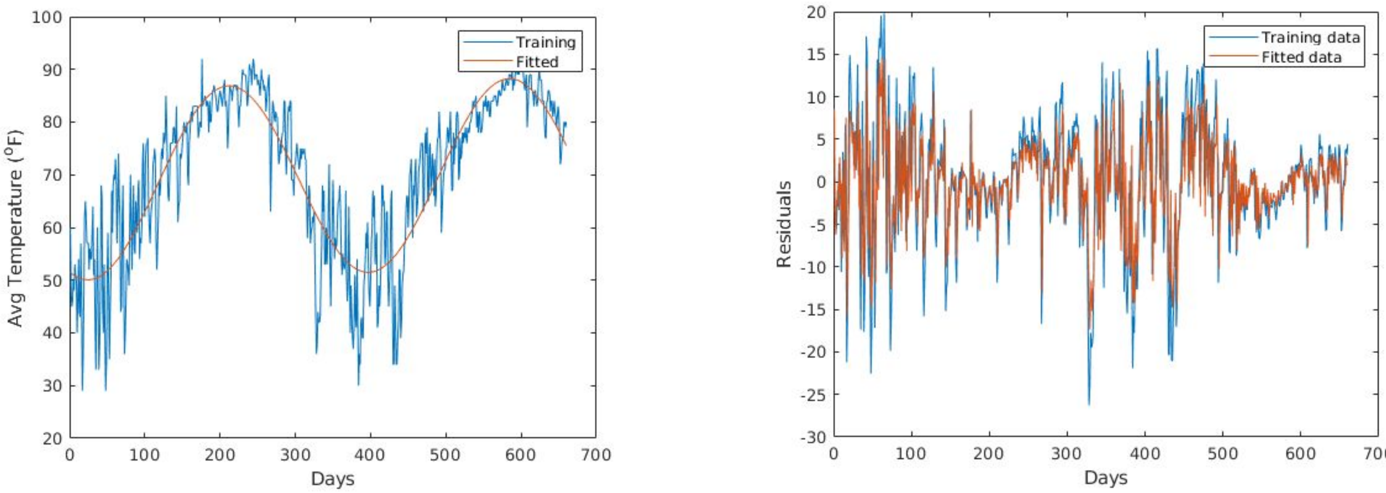
a1*sin(b1*x+c1)+a2*sin(b2*x+c2)

```
Coefficients (with 95% confidence bounds):
  a1 =       77.59  (-324, 479.2)
  b1 =    0.000105  (-0.002595, 0.002805)
  c1 =       1.069  (-8.322, 10.46)
  a2 =       18.04  (17.22, 18.85)
  b2 =     0.01686  (0.01656, 0.01717)
  c2 =      -1.992  (-2.111, -1.874)
```

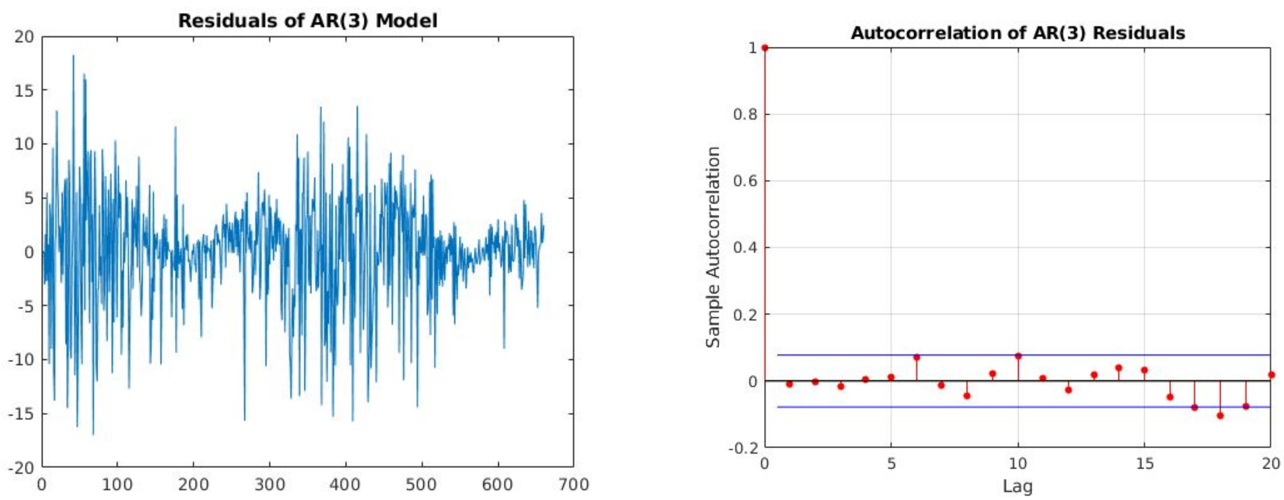
If we carefully look at b2 value, we see that periodicity is around 370. This makes sense as it is yearly periodic (365). Hence, there is strong deterministic trend.

After we get a good fit, we detrend the data using this curve and then we use ARMA(2n,2n-1) technique and F-test as the model criterion for finding the adequate ARMA model. When we do that we get AR(3) as the adequate model with RSS = 1.6763e+04.

**Plot shows trend and fitted data**



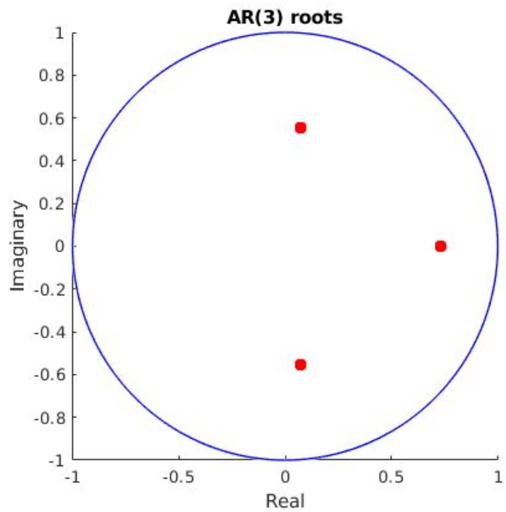

**Plot shows ARMA(6,3) residuals and residual autocorrelation**

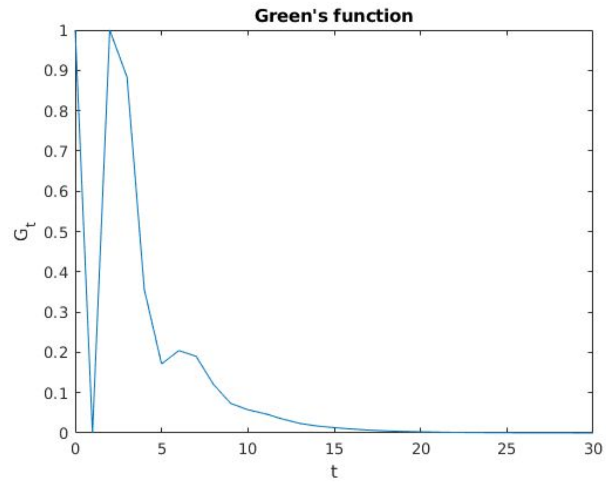



After we get AR(3) model, we look at its roots. We find that it has 1 real root and 2 complex roots. The one real root is 0.7345 so stochastic trend was not checked. The complex roots give period 4.3701 ≅ 4 or 4.5, so seasonality for this period was checked and we found that it does not exists.

From the green's function, we see that the system is really stable.

**Plot shows ARMA(6,3) AR roots on unit circle and Green's function**

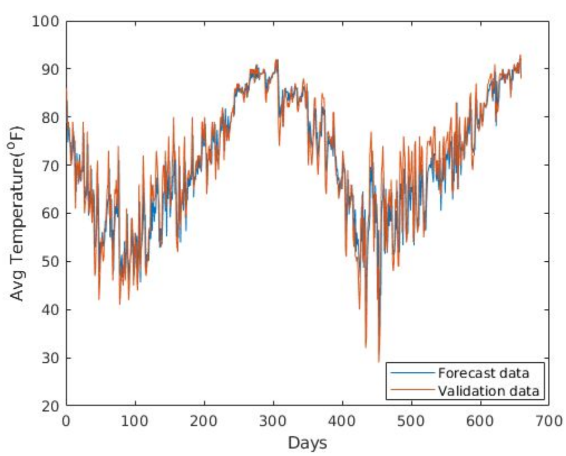


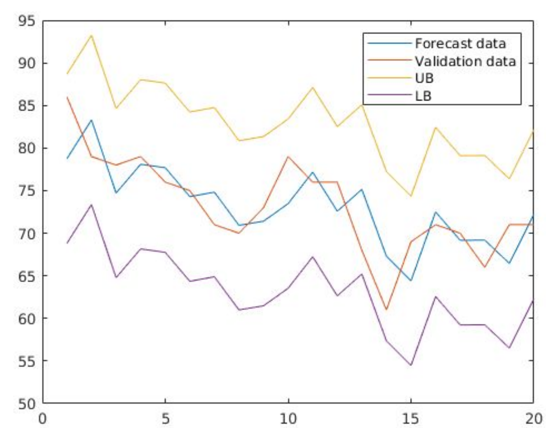




## Non-Stationary Model Forecast

We would like to know how good is the model by forecasting and checking with the validation data. Forecast values for rolling 1-step are really good when compared to stationary model as the RSS =1.6550e+04 , MSE=25.69, STD=5.0685 values are better.

**Plot shows 1-step rolling forecast on validation data and zoomed version for 20 points**

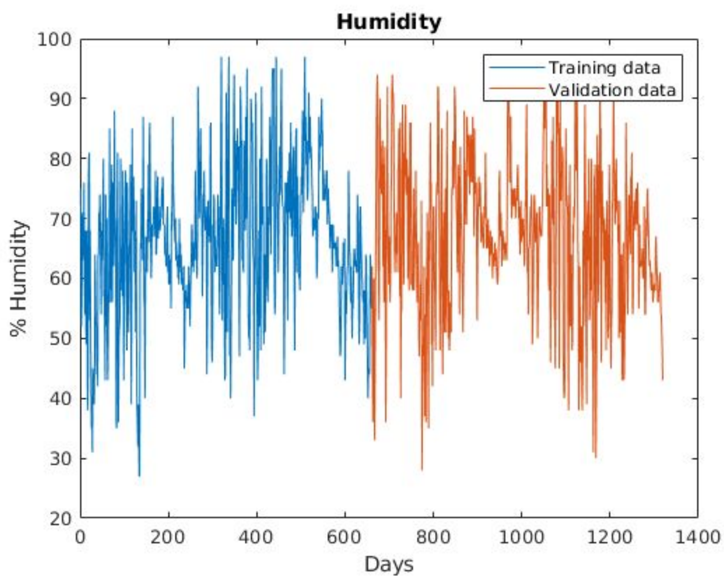


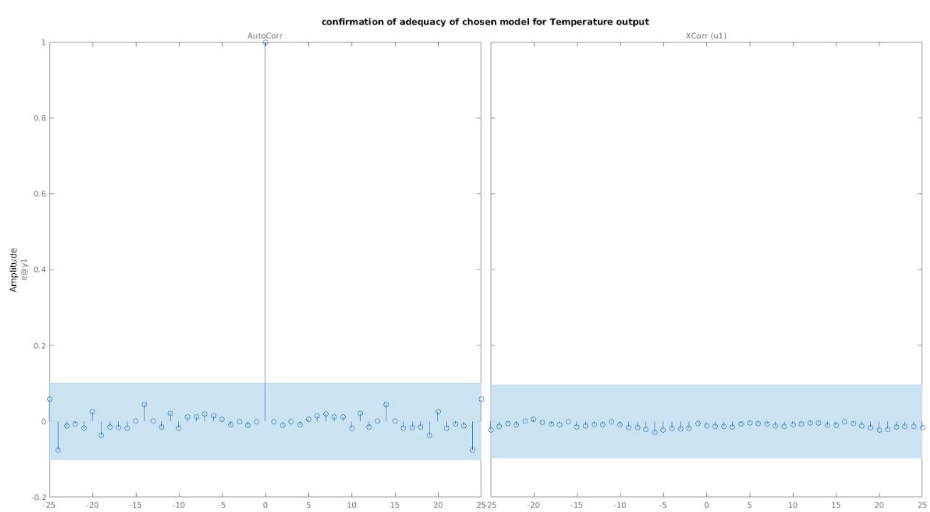
## ARMAV Model

In an attempt to improve the accuracy further, time series data of percentage humidity was also used to supplement the temperature data. Now using both the time series data, ARMAV model was used to fit the data. For temperature data, we used detrended data to improve the forecast results(It was tried on actual data also but it was quite bad). Using (n,n,n-1) strategy and F-test criterion for model selection, we get (12,12,11) model as adequate with RSS = 1.4083e+04 for humidity as input and temperature as output. Whereas for temperature as input and humidity as output, we get (8,8,7) model as adequate with RSS = 5.1232e+04.

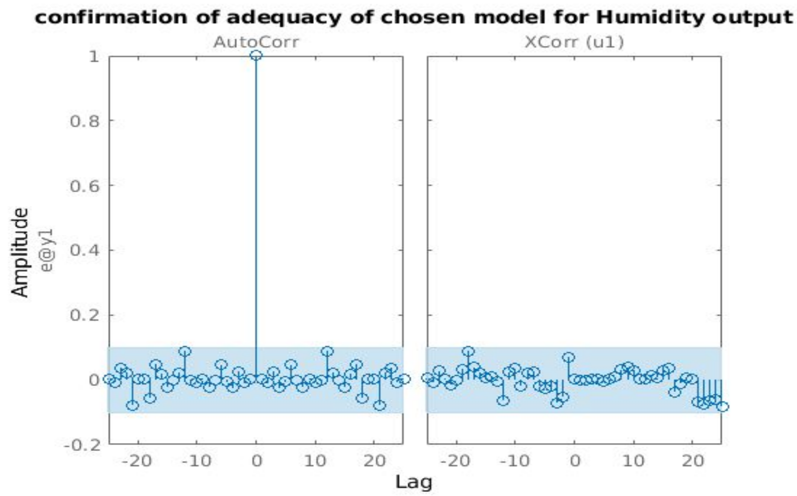
**. Plot of Training and Validation Data of Humidity**




**Plot shows ARMAV temperature residuals and residual autocorrelation and cross correlation**



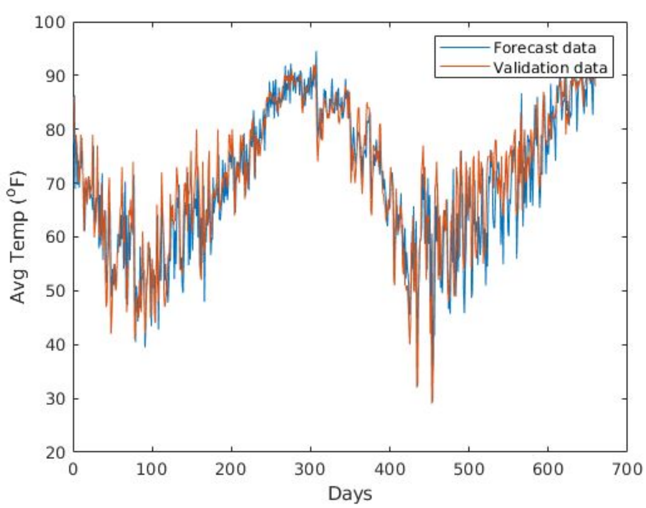
**Plot shows ARMAV humidity residuals and residual autocorrelation and cross correlation**

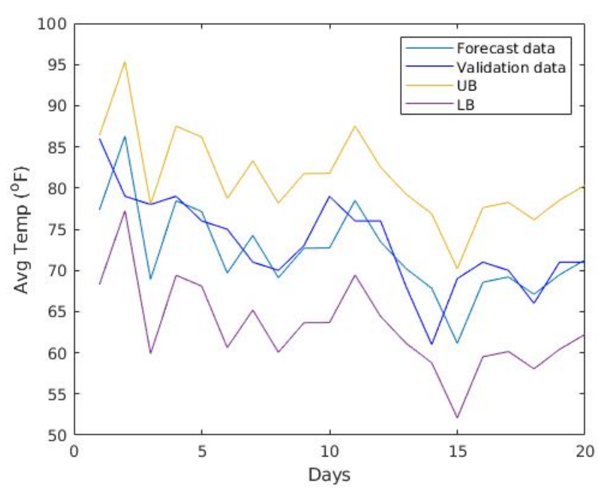



## ARMAV Model Forecast

Using the above model, we do 1-step rolling prediction for the validation data of temperature. Forecast RSS is 2.3642e+04 which is pretty bad compared to previous models, this may be because of no proper structure within Humidity data.

**Plot shows 1-step rolling forecast on validation data and zoomed version for 20 points**

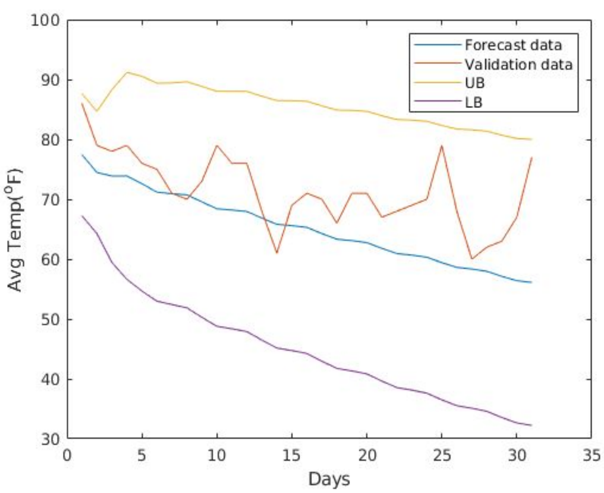


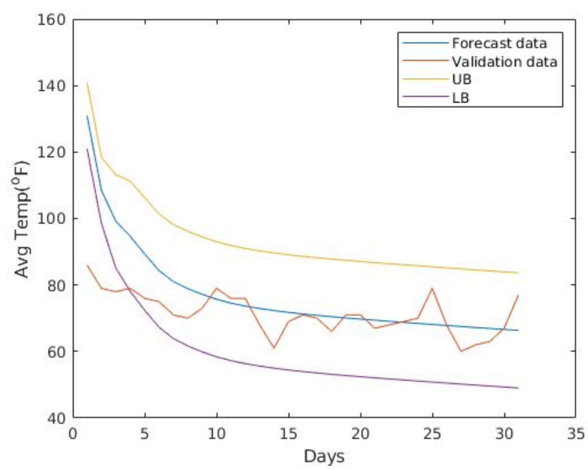



**Model Selected**

Among the models, non-stationary model performed really well followed by stationary model and ARMAV model. So using the first two good models, 1 to 31-steps ahead forecast was done. (One month temperature prediction)

**Plot shows 1 to 31-step forecast on validation data for 20 points (L: Stationary, R:Non-Stationary)**




We can clearly see that even here non-stationary model outperforms the stationary model.

**Conclusion**

Among the models, Non-stationary model seems to be the best model as it has the lowest forecast RSS of 1.655e+04 for 1-step rolling forecast and for 1 to 31 step forecast also the variance becomes constant after some time. This model outperforms other models, suggesting that there is strong deterministic trend.

Therefore the best model is Non-Stationary model with deterministic trend ( Yearly Periodic) and AR(3) model for the detrended data.

**References**


Pandit, S. M., & Wu, S. M. (1983). *Time series and system analysis with applications* (Vol. 3). New York: Wiley.