# Climate Change & Weather Prediction using Time Series

Haja Amir Rahman (p2100803)

School of Computing

Singapore Polytechnic, Singapore

amirrahman517804@gmail.com

*Abstract*—The study is focused on the prediction of different geographical components of weather such as Temperature, WindGustSpeed, WindSpeed, Humidity, Pressure. Recent shifts in the Australian climate including both higher temperatures and lower winter rainfall, have had significant effects on the agriculture sector. Despite these recent trends, there remains uncertainty over the future climate and its potential impacts on Australian farm businesses. Australia's climate has been experiencing continued warming, an increase in extreme fire weather and length of the fire season, declining rainfall in the southeast and southwest of the continent, and rising sea levels.

## I. INTRODUCTION

The Bureau of Meteorology is Australia's national weather, climate and water agency. Its expertise and services assist Australians in dealing with the harsh realities of their natural environment, including drought, floods, fires, storms, tsunami and tropical cyclones. Through regular forecasts, warnings, monitoring and advice spanning the Australian region and Antarctic territory, the Bureau provides one of the most fundamental and widely used services of government.

The Bureau contributes to national social, economic, cultural and environmental goals by providing observational, meteorological, hydrological and oceanographic services and by undertaking research into science and environment related issues in support of its operations and services.

BOM weather stations are assigned a weather station ID. Observation data is collected per weather (observation) station.

## II. METHODOLOGY

A. DATA COLLECTION

The 2 datasets used for this study were obtained from Kaggle. One is observation data and one is a normal dataset. The first observation dataset contains a date range of 1999-2018 while the second one contains a date range of 2007 to 2008. The dataset contains 2979 records of Weather data. Some EDA was done with the observation data as the date range is longer but the actual Time Series Algorithms were applied to the second dataset as it has more geographical features to analyze and make conclusions with.

The following figure shows the columns and their definitions for the second dataset:

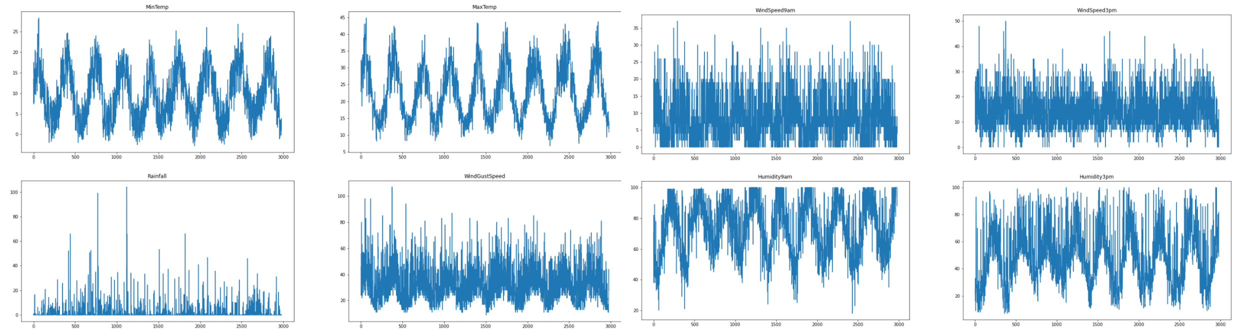
## Data Dictionary

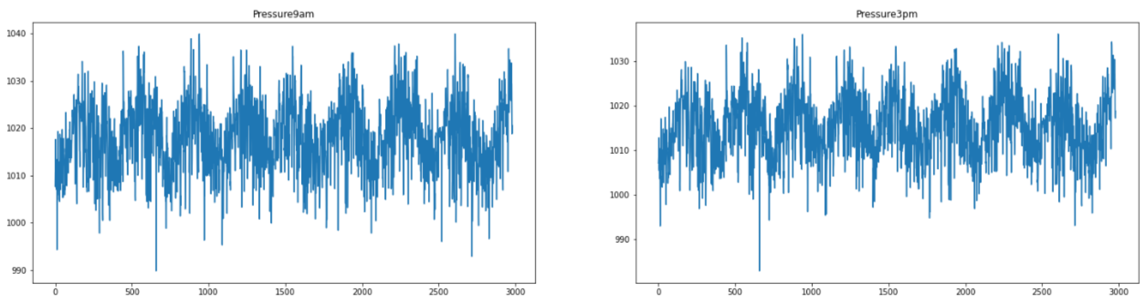
The training dataset consist of 2979 observations from March 2008 to January 2017 with daily records for the various attributes.

| Columns | Descriptions |
|---|---|
| Date | (DD/MM/YYYY) |
| Location | The common name of the location of the weather station |
| MinTemp | The minimum temperature in degrees celsius |
| MaxTemp | The maximum temperature in degrees celsius |
| Rainfall | The amount of rainfall recorded for the day in mm |
| Evaporation | The so-called Class A pan evaporation (mm) in the 24 hours to 9am |
| Sunshine | The number of hours of bright sunshine in the day. |
| WindGustDir | The direction of the strongest wind gust in the 24 hours to midnight |
| WindGustSpeed | The speed (km/h) of the strongest wind gust in the 24 hours to midnight |
| WindDir9am | Direction of the wind at 9am |
| WindDir3pm | Direction of the wind at 3pm |
| WindSpeed9am | Wind speed (km/hr) averaged over 10 minutes prior to 9am |
| WindSpeed3pm | Wind speed (km/hr) averaged over 10 minutes prior to 3pm |
| Humidity9am | Humidity (percent) at 9am |
| Humidity3pm | Humidity (percent) at 3pm |
| Pressure9am | Atmospheric pressure (hpa) reduced to mean sea level at 9am |
| Pressure3pm | Atmospheric pressure (hpa) reduced to mean sea level at 3pm |
| Cloud9am | Fraction of sky obscured by cloud at 9am. This is measured in "oktas", which are a unit of eigths. It records how many |
| Cloud3pm | Fraction of sky obscured by cloud (in "oktas": eighths) at 3pm. See Cload9am for a description of the values |
| Temp9am | Temperature (degrees C) at 9am |
| Temp3pm | Temperature (degrees C) at 3pm |
| RainToday | Boolean: 1 if precipitation (mm) in the 24 hours to 9am exceeds 1mm, otherwise 0 |
| Risk_MM | The amount of next day rain in mm. Used to create response variable RainTomorrow. A kind of measure of the "risk". |
| RainTomorrow | The target variable. Did it rain tomorrow? |

Figure1

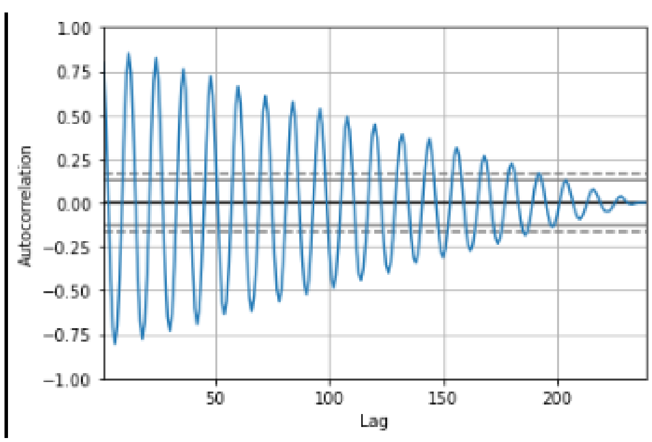
# III. EXPLORATORY DATA ANALYSIS (EDA)



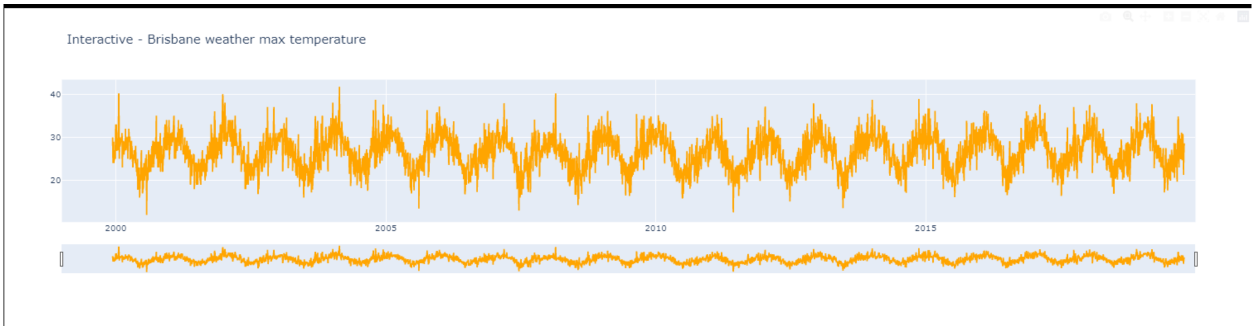




The above visualizations show us the overall trend of the different geographical factors: MinTemp, MaxTemp, Rainfall, WindGustSpeed, WindSpeed9am, WindSpeed3pm, Humidity9am, Humidity3pm, Pressure9am, Pressure3pm respectively. Since these are factors of weather and climate change, it is seasonal thus the smooth trend for most of the factors. However we can see that Rainfall does not have a seasonal or smooth trend.

For the observation dataset, I have created an autocorrelation plot using pandas.plotting.



We can see that there is a correlation around every 12 months (which makes sense due to seasons). Let's have a look at the data in an interactive way - you can try it out in the jupyter notebook.




# IV. DATA PREPROCESSING

For the second dataset which is the main dataset, I did some data preprocessing to drop some trivial geographical factors such as Cloud measurement in the morning and at noon.

| Date | MinTemp | MaxTemp | Rainfall | WindGustSpeed | WindSpeed9am | WindSpeed3pm | Humidity9am | Humidity3pm | Pressure9am | Pressure3pm |
|---|---|---|---|---|---|---|---|---|---|---|
| 2008-12-01 | 13.4 | 22.9 | 0.6 | 44.0 | 20.0 | 24.0 | 71.0 | 22.0 | 1007.7 | 1007.1 |
| 2008-12-02 | 7.4 | 25.1 | 0.0 | 44.0 | 4.0 | 22.0 | 44.0 | 25.0 | 1010.6 | 1007.8 |
| 2008-12-03 | 12.9 | 25.7 | 0.0 | 46.0 | 19.0 | 26.0 | 38.0 | 30.0 | 1007.6 | 1008.7 |
| 2008-12-04 | 9.2 | 28.0 | 0.0 | 24.0 | 11.0 | 9.0 | 45.0 | 16.0 | 1017.6 | 1012.8 |
| 2008-12-05 | 17.5 | 32.3 | 1.0 | 41.0 | 7.0 | 20.0 | 82.0 | 33.0 | 1010.8 | 1006.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2017-06-21 | 1.2 | 15.2 | 0.4 | 15.0 | 0.0 | 2.0 | 100.0 | 62.0 | 1029.4 | 1026.7 |
| 2017-06-22 | 0.8 | 13.4 | 0.0 | 17.0 | 6.0 | 0.0 | 100.0 | 66.0 | 1029.4 | 1025.9 |
| 2017-06-23 | 1.1 | 11.9 | 0.0 | 44.0 | 9.0 | 2.0 | 100.0 | 81.0 | 1022.3 | 1017.7 |
| 2017-06-24 | 1.1 | 14.1 | 0.2 | 28.0 | 4.0 | 15.0 | 100.0 | 49.0 | 1018.8 | 1017.2 |
| 2017-06-25 | 3.9 | 10.9 | 0.0 | 28.0 | 6.0 | 0.0 | 88.0 | 82.0 | 1020.5 | 1018.8 |

Remaining are these 10 columns which will be passed through the baseline ARIMA model to see how our prediction, RMSE and AIC scores are.

Before passing the finalized dataset through the baseline model, since this is time series and though this is seasonal data, I conducted the Dicky Fuller Test for Stationary Series. Turns out all the features are indeed stationary so there is no differencing required.

## Dickey Fuller Test for Stationarity Series

To model a time-series using ARMA model, we need to ensure the time-series is stationary, that is the the **mean, variance and covariance does not vary with time.**

To verify stationarity, we will be utilising Augmented Dickey Test at Significant Value, $p = 0.05$ with,

$$H_0 : \text{Time-Series is Non-Stationary}$$
$$H_1 : \text{Time-Series is Stationary}$$

```
p_values = []
for colname in data:
    coldata = data[colname]
    p_values.append((colname, adfuller(coldata.values)[1])) # Run Adfuller test to determine is the time-series Stationary

_ = pd.DataFrame(p_values, columns=["Colname", "P value"])
_["stationary"] = _["P value"]<0.05
_
```

✓ 1.1s

| | Colname | P value | stationary |
|---|---|---|---|
| 0 | MinTemp | 2.024920e-03 | True |
| 1 | MaxTemp | 1.226176e-02 | True |
| 2 | Rainfall | 0.000000e+00 | True |
| 3 | WindGustSpeed | 1.270442e-05 | True |
| 4 | WindSpeed9am | 1.317602e-07 | True |
| 5 | WindSpeed3pm | 1.800013e-08 | True |
| 6 | Humidity9am | 9.463469e-03 | True |
| 7 | Humidity3pm | 3.116801e-03 | True |
| 8 | Pressure9am | 1.276637e-08 | True |
| 9 | Pressure3pm | 1.210534e-08 | True |

Figure 3

Next we have the PACF plots for the selected features so that we can decide an appropriate order for the prediction model.

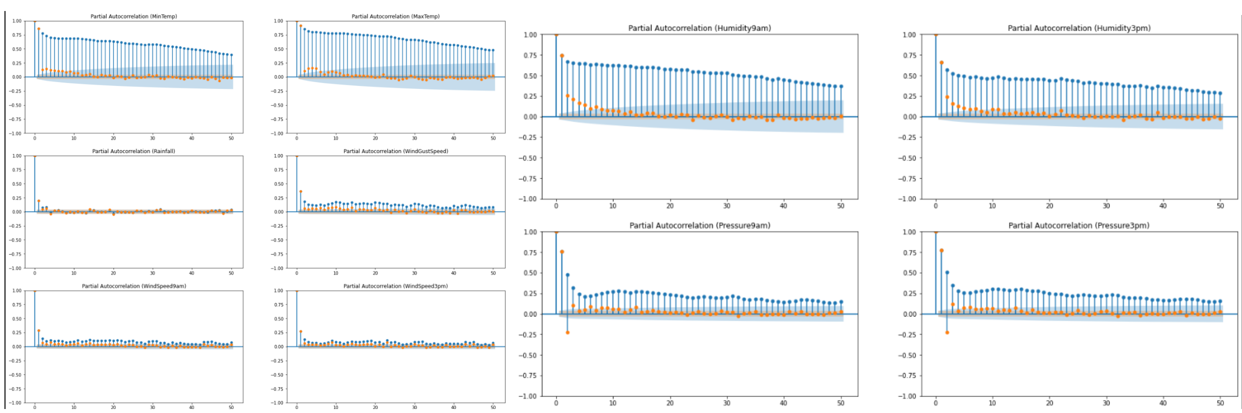
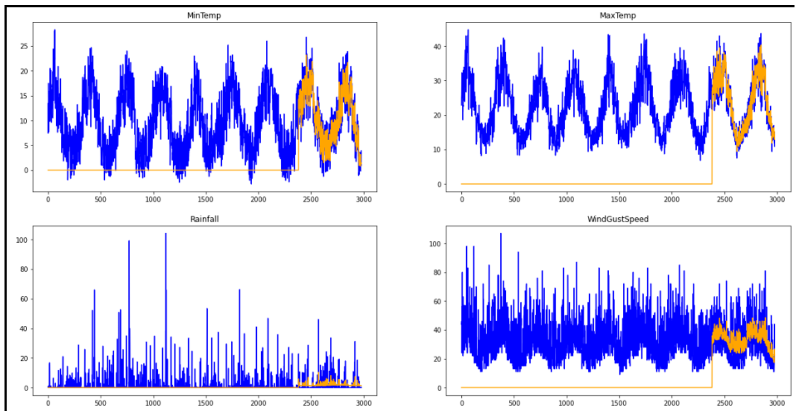
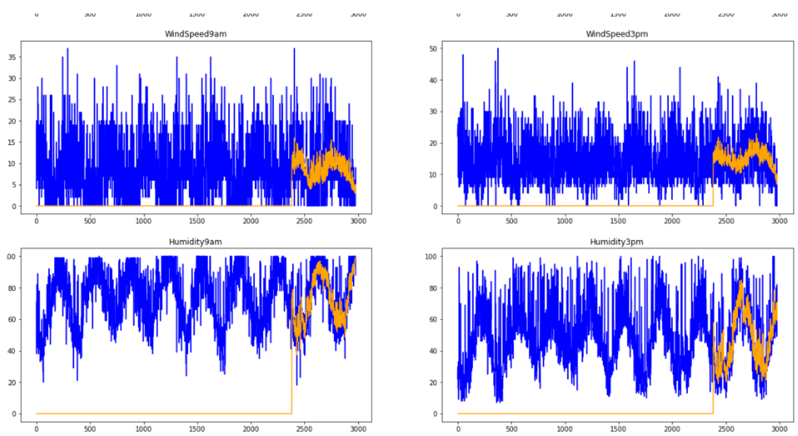


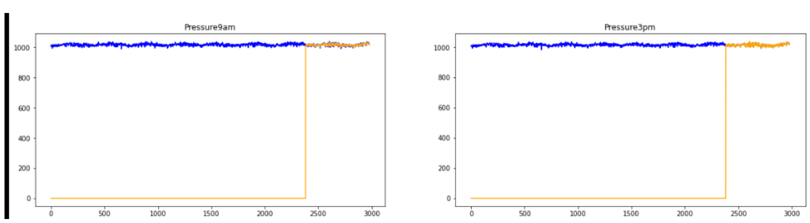

Figure 4

Based on this we have arrived at the orders of "orders = [(1,1,1),(2,0,3),(2,0,5),(5,1,3),(6,1,6),(6,2,6),(6,1,7)]". These orders do not correspond to the different features but are just there to test out first.

# V. MODELING (ARIMA & SARIMAX)




Above we have the baseline prediction results using an ARIMA model for the respective features. As we can see, the predictions are quite consistent with the actual data for all the features and there are no alarming anomalies.

# VI. HYPERPARAMETER TUNING

Now we move onto hyperparameter tuning. In addition to the interpreted the parameters based on ACF and PACF plot, I have also performed with hyperparameters tuning by using a customized GridSearch function in a for-loop as the time-series cross validation to find out the best set of hyperparameters for SARIMAX model by minimizing the Root Mean Squared Error (RMSE) and Akaike Information Criterion (AIC).

ARIMA Model Possible Hyperparameters:

| | colname | order | rmse | aic | mape |
|---|---|---|---|---|---|
| 60 | MinTemp | (6, 1, 7) | 2.914995 | 14832.359311 | 2.390319e+13 |
| 40 | MinTemp | (6, 1, 6) | 2.925516 | 14838.656970 | 2.423602e+13 |
| 10 | MinTemp | (2, 0, 3) | 2.933233 | 14849.236309 | 2.772110e+13 |
| 50 | MinTemp | (6, 2, 6) | 2.933983 | 14861.438737 | 2.263666e+13 |
| 30 | MinTemp | (5, 1, 3) | 2.938434 | 14854.670397 | 2.822358e+13 |
| ... | ... | ... | ... | ... | ... |
| 27 | Humidity3pm | (2, 0, 5) | 13.262674 | 24070.554665 | 2.369367e-01 |
| 57 | Humidity3pm | (6, 2, 6) | 13.268175 | 24092.567514 | 2.257825e-01 |
| 7 | Humidity3pm | (1, 1, 1) | 13.276105 | 24066.163240 | 2.318897e-01 |
| 37 | Humidity3pm | (5, 1, 3) | 13.278925 | 24071.416438 | 2.331931e-01 |
| 47 | Humidity3pm | (6, 1, 6) | 13.294638 | 24074.457846 | 2.329331e-01 |

70 rows × 5 columns

SARIMAX Model Possible Hyperparameters:

| | colname | order | best rmse | aic |
|---|---|---|---|---|
| 0 | Humidity3pm | (6, 1, 7) | 13.258351 | 24080.301818 |
| 1 | Humidity9am | (6, 1, 7) | 10.099721 | 22409.379768 |
| 2 | MaxTemp | (2, 0, 5) | 2.951463 | 14998.142011 |
| 3 | MinTemp | (6, 1, 7) | 2.914995 | 14832.359311 |
| 4 | Pressure3pm | (6, 1, 7) | 4.456807 | 17159.269155 |
| 5 | Pressure9am | (6, 1, 7) | 4.917523 | 17536.740046 |
| 6 | Rainfall | (5, 1, 3) | 4.737037 | 19198.764235 |
| 7 | WindGustSpeed | (6, 1, 6) | 11.575267 | 23353.516119 |
| 8 | WindSpeed3pm | (6, 2, 6) | 6.595212 | 19966.531335 |
| 9 | WindSpeed9am | (2, 0, 3) | 6.339236 | 19486.606038 |

ARIMA Best Hyperparamters:

```python
best = results.loc[results.groupby("colname")["aic"].idxmin()]
best
```

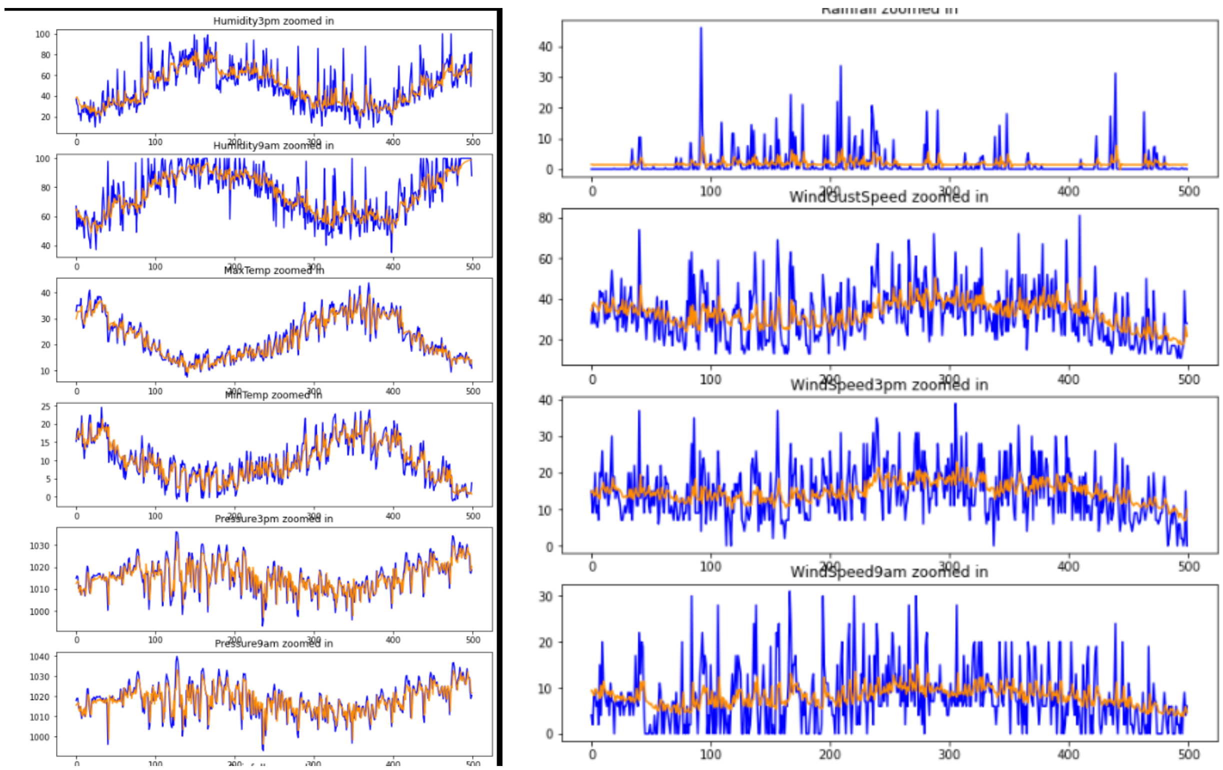
✓ 0.1s

| | colname | order | rmse | aic | mape |
|---|---|---|---|---|---|
| 7 | Humidity3pm | (1, 1, 1) | 13.276105 | 24066.163240 | 2.318897e-01 |
| 6 | Humidity9am | (1, 1, 1) | 10.115780 | 22402.700507 | 1.043831e-01 |
| 21 | MaxTemp | (2, 0, 5) | 2.949086 | 14993.863507 | 1.065957e-01 |
| 60 | MinTemp | (6, 1, 7) | 2.914995 | 14832.359311 | 2.390319e+13 |
| 39 | Pressure3pm | (5, 1, 3) | 4.474582 | 17154.278667 | 3.242843e-03 |
| 38 | Pressure9am | (5, 1, 3) | 4.939168 | 17536.135761 | 3.541815e-03 |
| 12 | Rainfall | (2, 0, 3) | 4.781712 | 19195.166232 | 4.739169e+15 |
| 33 | WindGustSpeed | (5, 1, 3) | 11.590366 | 23345.916269 | 3.164025e-01 |
| 5 | WindSpeed3pm | (1, 1, 1) | 6.617315 | 19935.777093 | 7.060598e+14 |
| 14 | WindSpeed9am | (2, 0, 3) | 6.329491 | 19470.765927 | 5.596466e+15 |

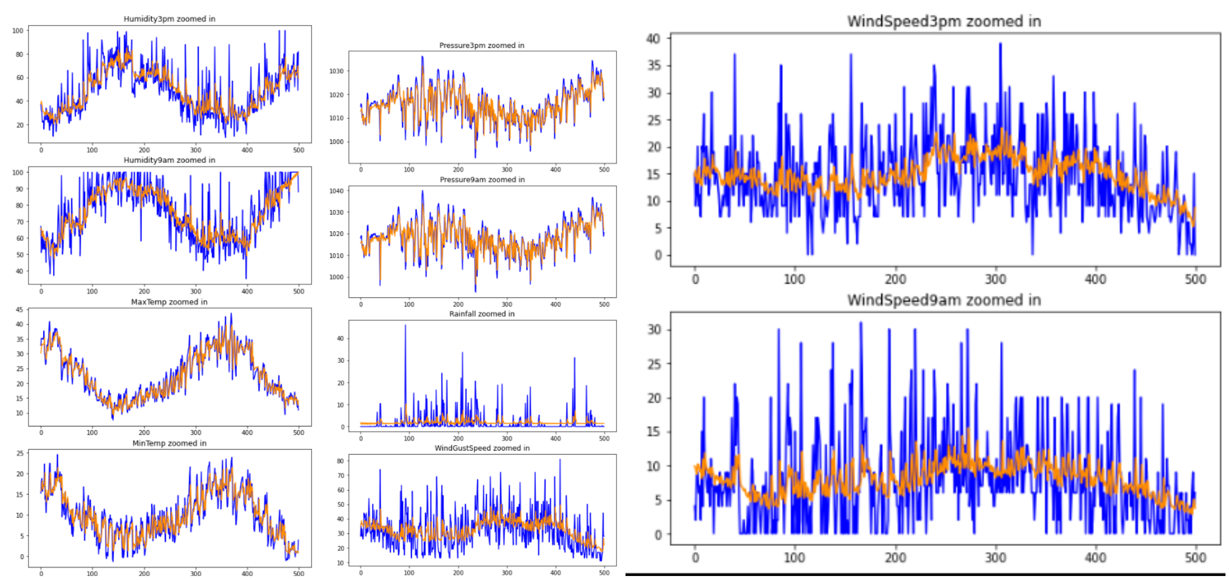
SARIMAX Best Hyperparameters:

| | colname | order | rmse | aic |
|---|---|---|---|---|
| 67 | Humidity3pm | (6, 1, 7) | 13.258351 | 24080.301818 |
| 66 | Humidity9am | (6, 1, 7) | 10.099721 | 22409.379768 |
| 21 | MaxTemp | (2, 0, 5) | 2.951463 | 14998.142011 |
| 60 | MinTemp | (6, 1, 7) | 2.914995 | 14832.359311 |
| 69 | Pressure3pm | (6, 1, 7) | 4.456807 | 17159.269155 |
| 68 | Pressure9am | (6, 1, 7) | 4.917523 | 17536.740046 |
| 32 | Rainfall | (5, 1, 3) | 4.737037 | 19198.764235 |
| 43 | WindGustSpeed | (6, 1, 6) | 11.575267 | 23353.516119 |
| 55 | WindSpeed3pm | (6, 2, 6) | 6.595212 | 19966.531335 |
| 14 | WindSpeed9am | (2, 0, 3) | 6.339236 | 19486.606038 |

ARIMA Prediction Results based on Hyperparameters:

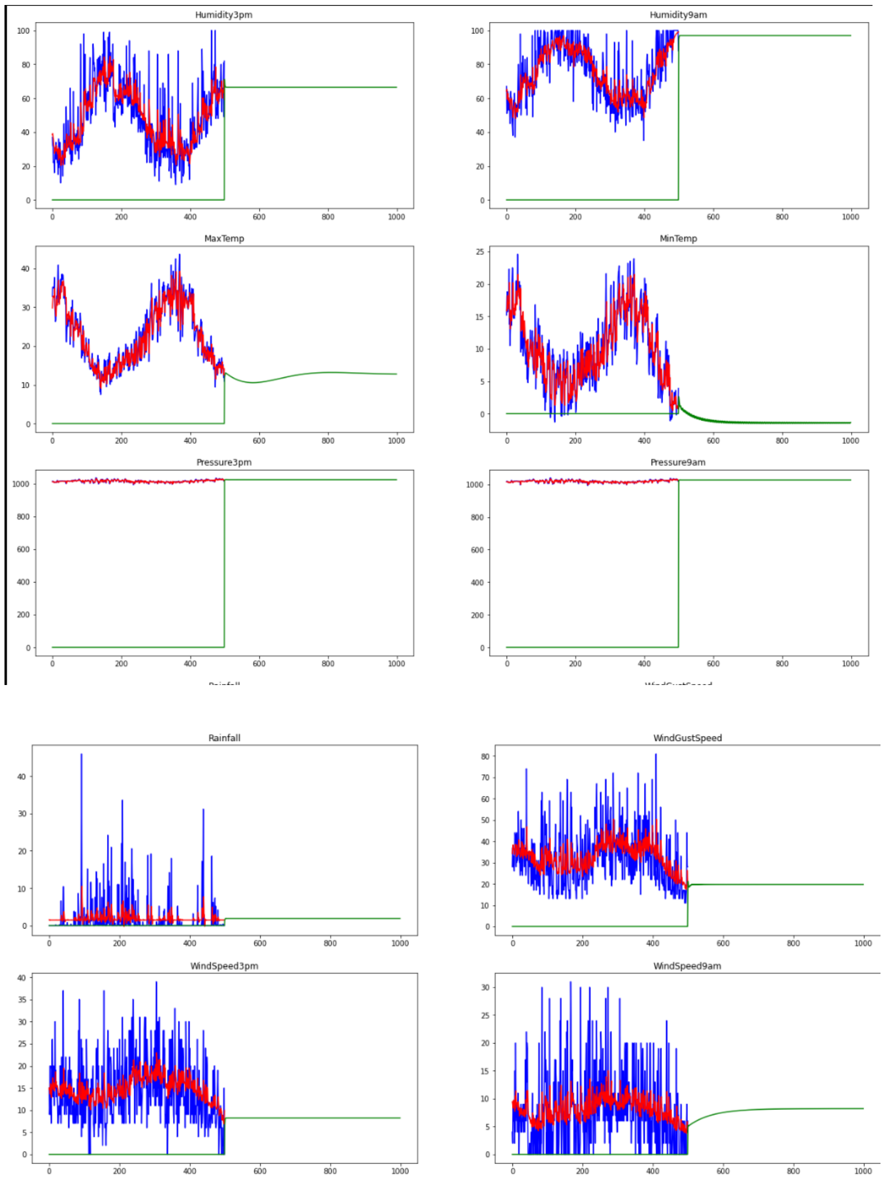



SARIMAX Prediction with Best Hyperparameters:



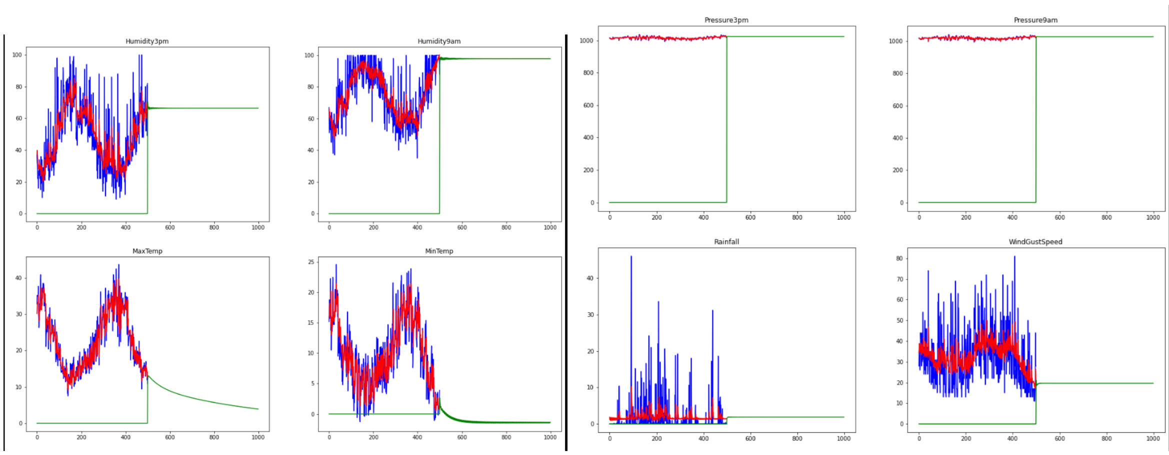

# VII. FORECASTING

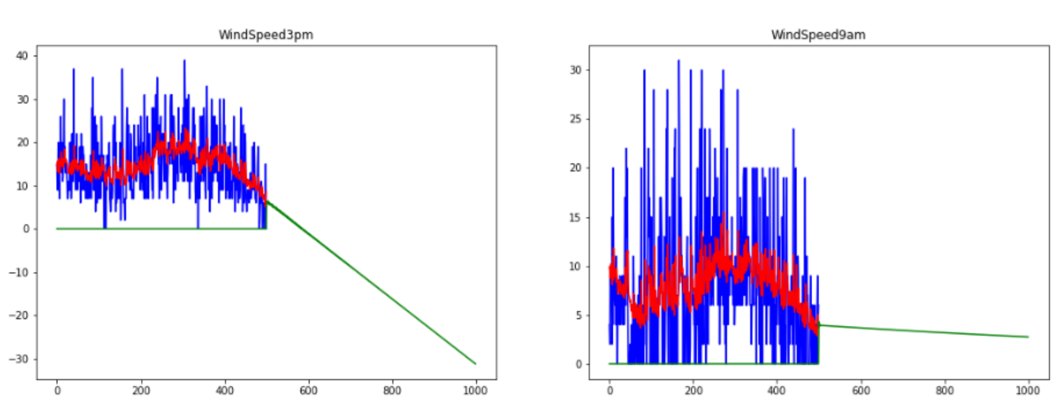
ARIMA Forecast:




I plotted the visualizations such that the forecast overlaps with the actual and predicted trends so that comparison is easier. We can see that Minimum, Maximum Temp, Pressure at 9am and Pressure at 3pm have a pretty consistent forecast. The rest do not have much consistent forecasts and after some research I found out that it is because of normal daily fluctuations. These factors are small but important factors that affect the weather daily and in the long term.

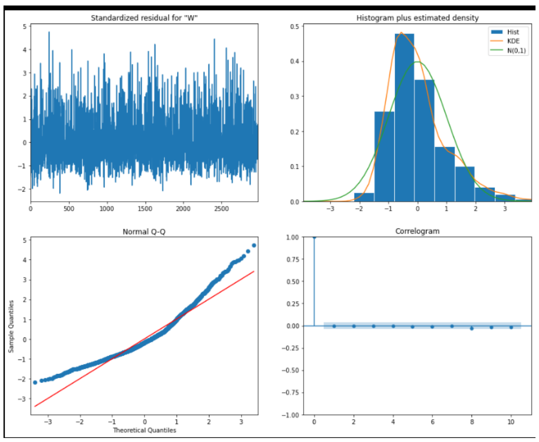
SARIMAX Forecast:




I did the same plotting for SARIMAX as I did for ARIMA so that comparison is easier. We can see that the forecasts for the different features affecting the weather are significantly more accurate and consistent than with ARIMA thus allowing me to choose the SARIMAX model as my final model.

SARIMAX Error Analysis and summarized results:




SARIMAX Results

| | | | |
|---|---|---|---|
| Dep. Variable: | WindSpeed9am | No. Observations: | 2979 |
| Model: | SARIMAX(2, 0, 3) | Log Likelihood | -9737.303 |
| Date: | Fri, 12 Aug 2022 | AIC | 19486.606 |
| Time: | 21:56:10 | BIC | 19522.602 |
| Sample: | 0 | HQIC | 19499.558 |
| | - 2979 | | |
| Covariance Type: | opg | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ar.L1 | 0.5350 | 0.309 | 1.731 | 0.083 | -0.071 | 1.141 |
| ar.L2 | 0.4639 | 0.309 | 1.502 | 0.133 | -0.141 | 1.069 |
| ma.L1 | -0.2930 | 0.308 | -0.950 | 0.342 | -0.897 | 0.311 |
| ma.L2 | -0.5039 | 0.242 | -2.081 | 0.037 | -0.979 | -0.029 |
| ma.L3 | -0.1249 | 0.052 | -2.414 | 0.016 | -0.226 | -0.024 |
| sigma2 | 40.3856 | 0.870 | 46.420 | 0.000 | 38.680 | 42.091 |

| | | | |
|---|---|---|---|
| Ljung-Box (L1) (Q): | 0.00 | Jarque-Bera (JB): | 768.36 |
| Prob(Q): | 0.94 | Prob(JB): | 0.00 |
| Heteroskedasticity (H): | 0.91 | Skew: | 1.06 |
| Prob(H) (two-sided): | 0.13 | Kurtosis: | 4.31 |

# VIII. CONCLUSION

In summary, I have presented a model that is able to forecast key Climate-Change and Weather statistics of new daily measurements of the different geographical factors of weather using multivariate SARIMAX model with RMSE of 2.951463 for Max Temperature, 10.099721 for Humidity at 9am,4. 917523 for Pressure at 9am, 4.737037 for Rainfall, 11.575267 for WindGustSpeed and 6.339236 for WindSpeed at 9am evaluated using a hold-out test set. However, there are still some gaps in my modeling as the data obtained only establish an upward trend. I am uncertain of the model performance when the climate situation in Australia subsides and the key factors fluctuate. Besides, a lower granularity modeling could be taken by modeling the Climate statistics for different cities in Australia such that the authorities could adopt a more targeted approach and channel the resources or implement a more advanced equipment to monitor the weather to maintain their results based on their policies. **On 27 July 2022, the Government introduced the Climate Change Bill 2022**. The Bill proposes to set Australia's greenhouse gas emissions reduction targets into law. It is similar to climate target laws enacted in other OECD nations and in four of Australia's states and territories.

Finally, I wish to publish the findings and deploy the model as an interactive web application as a reminder to urge all countries to abide the policies that they have decided to execute.

# IX. REFERENCES


Csiro.au. 2022. *State of the Climate - CSIRO*. [online] Available at: <https://www.csiro.au/en/research/environmental-impacts/climate-change/state-of-the-climate> [Accessed 12 August 2022].

Anon, Simulating the effects of climate change on the profitability of Australian Farms. *Simulating the effects of climate change on the profitability of Australian farms - DAFF*. Available at: https://www.agriculture.gov.au/abares/research-topics/working-papers/simulating-the-effects-of-climate-change [Accessed August 12, 2022].

corporateName=Commonwealth Parliament; address=Parliament House, C., 2022. Climate change bill 2022. *Home – Parliament of Australia*. Available at: https://www.aph.gov.au/About_Parliament/Parliamentary_departments/Parliamentary_Library/FlagPost/2022/July/Climate_Change_Bill_2022 [Accessed August 12, 2022].