Climate Change Forecast

Holistic study of change in Global Land/Ocean temperature and Associated factors

Team 4

Disha Sugandhi

Chakradhar Rajineni

Biheng Yang

Divya Verma

Chandrashila Chattopadhyay

TABLE OF CONTENTS

ABSTRACT:

Global Warming is the major contributor that would reshape the world in the next century. The global temperature is the observed increase in the average, maximum and minimum temperatures of the Earth. The report would explore the temperature changes in major cities with highest population, coastal cities, cities, and global average temperatures on the land. The report also explores the major reasons for global temperatures are fossil fuel burning, CO2 concentration to lead to greenhouse effect. The approach taken here is time series Analysis and forecasting to comprehend the phenomenon and make better predictions for the future. The data sets are collected from Kaggle and algorithms like exponential smoothing, univariate and multivariate SARIMAX is used with exogenous factors like population, co2 concentration, fossil fuel consumption are considered for the forecasting. Temperature timeseries from different cities, countries and global level are used. Initially, the approach is creating a model with the past data and results are extrapolated for the future. The report considers only few factors that could affect the temperature change in future and therefore, cannot be considered as accurate projections for the future. The report can be used as exhibition of timeseries analysis and forecasting.

1. BACKGROUND

Weather is affecting our life subtly from all aspects, like transportation, human activity and even our mood. With impact on key sectors like agriculture, water resources and economics, climate plays an influential role in the human life cycle. Extreme events such as floods, droughts and cyclones affect lives and livelihoods, and often result in millions worth of damages. If we could fully prepare for this in advance, the cost will decrease dramatically. Thus, implementing policies based on temperature change forecasts will help us mitigate these unforeseen damages. We are developing a model to predict the average Land temperature and performing climate data visualizations which would empower policymakers, researchers, media and other stakeholders to gather insights on global progress of climate change.

2. PROBLEM STATEMENT

Recent years, the concept of “global warming” has become universal. Someone believes it’s a great threat to all creators. It’s a sign to tell people to act environmentally friendly. The first one is to explore whether global warming is really happening or only a myth. Based on the extra factors: population, CO_2 (and GHG) emission, etc, we will pick several cities most or least relevant to a certain factor. Then we conducted a further exploration on whether these factors have an impact on climate change and how much it affects.

3. LITERATURE REVIEW

The current rate of warming has been unprecedented over the last millennium (IPCC 2007). Average global temperature is projected to increase from 1 to 6 °C from 1990 to 2100. Increases across much of the United States may be even greater (Houghton and others 2001). Causes of climate change include changes in atmospheric concentrations of greenhouse gases and aerosols, and changes in land cover. The concentration of carbon dioxide has risen dramatically since preindustrial times. Models of the carbon cycle project that the concentration of carbon dioxide will increase from 379 parts per million (ppm) in 1995 to 540 to 970 ppm by 2100 (IPCC 2001, 2007). Even if human sources of greenhouse gas emissions were stabilized, warming would continue for centuries owing to time lags associated with climate processes and feedback (IPCC 2007).

4. DATASET VIEW

The dataset is obtained from Kaggle but is originally from Berkeley Earth. We analyze each variable by looking at the distribution, verifying the variable data types relevance and planning to perform data preprocessing (check the correlations between the variables treatment of nulls, handling outliers if any). The objective of the dataset is to predict the average temperature in coming years, based on certain diagnostic measurements.

Dataset Variable Description

The dataset has seven variables having 6 predictor variables and 1 target variable (forecasted average temperature).

Following are the detailed description of each variable:

Sl.No.	Variable Name	Data Type	Variable Description
1	Dt	Date	Date range is from 1850 – 2013
2	Average Temperature	Decimal	Average land temperature for each city w.r.t mentioned date. No missing values
3	Average Temperature Uncertainty	Decimal	Error margin related to average temperature calculated
4	City	String	This column has 3448 cities all over the world
5	Country	String	This column has 159 countries all over the world
6	Latitude	String	Latitude measure of cities
7	Longitude	String	Longitude measure of cities
8	Land Average Temperature	Decimal	Average Temperature recorded in Land for Global level
9	Land and Ocean Average Temperature	Decimal	Average Temperature Global level
10	Land Max Temperature	Decimal	Maximum temperature recorded in a month
11	Land Min Temperature	Decimal	Minimal temperature recorded in a month

5.

6. OUR GOAL AND APPROACH

7. To validate the hypothesis whether climate change is the biggest threat of our age or it is a myth based on unreliable scientific analysis.

Technical Approach

Model is developed in Python 3 (using IDE: Jupyter (Anaconda)), SAS, Tableau using Timeseries Analysis techniques and algorithms like SARIMAX, exponential smoothing to forecast the temperatures for the future.

Criteria: To explore global warming, we wanted to explore the changes in temperature overtime but also see if the whole world is affected by similar amounts. So, for exploring and forecasting, we have chosen multiple criteria like

1. Coastal cities- cities that are close to sea, to explore the changes at sea level.
2. Most populated - cities that have the most population, which will lead to more carbon emissions and pollution.
3. Cities on same longitude- cities that are on the same longitude to check by location metrics.
4. Most affected cities
5. Global average

Exploration goal:

We are going to analyze the forecast in each of the criteria, see the trend, see the change by annual with factors like Co2 emissions, co2 concentration, population, and fossil fuel usage. We are going to use our learnings from the class by using Time series Analysis to explore the reality of climate change and prepare forecasts until 2025.

Modeling Approach:

We used the SEMMA process to be able to better understand the data and create a model. A detailed explanation of the SEMMA process has been given below.

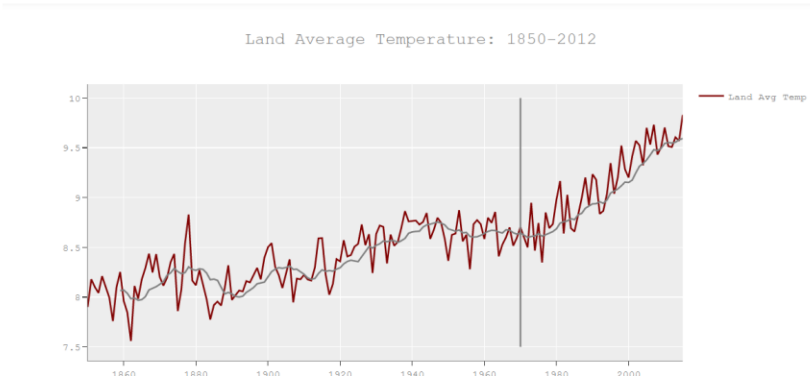
SAMPLE

As the original dataset has more than 3000 cities worldwide, we sampled the data on the basis of different factors associated with climate change such as Cities with huge population, Cities near sea, Cities with least population and low industrialization etc. For the modelling we have selected the last 1 year as test data, 5 years as validation data and remaining to train the model.

Data Exploration:

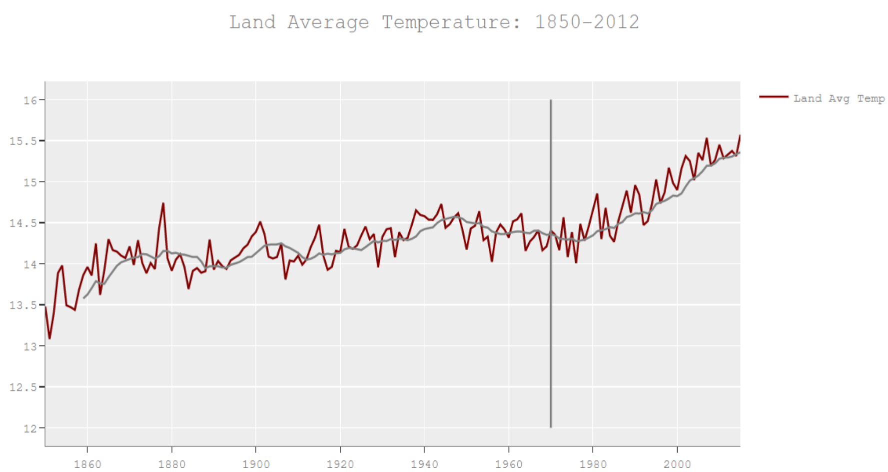
In the Data exploration part, we want to explore the Land Average Temperature, Land Max Temperature and Land Min Temperature with variables like Country, Continent, Co2 emissions, population, and Fossil Fuel extraction.

Variation in Average Temperatures:



As we can see, the average temperatures in 1850 was 8 degrees, with overtime has increased in 2015 to 9.5 degrees. As the Moving Average line (10 years) suggests there is an upward trend with the Average Temperature.

Variation in Maximum Temperature:

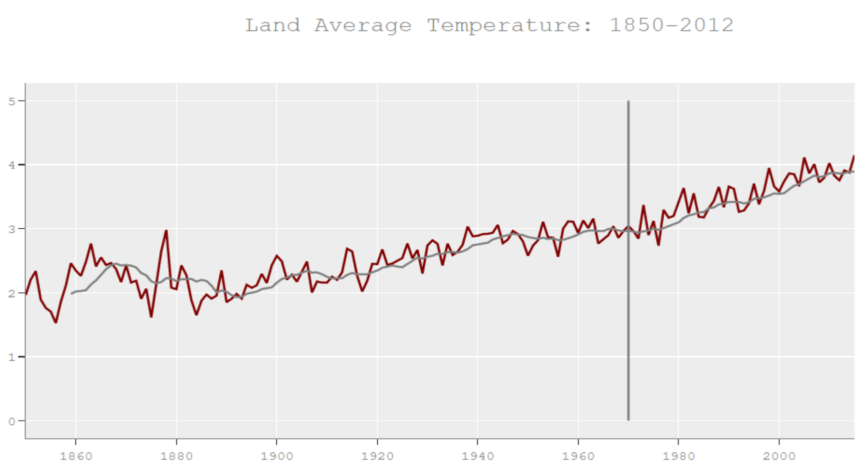


The graph max temperature increased from 13 to 15.6 which is 2.6 degree Celsius. The observation to see here is with Average temperature rise is only 1.5 degree Celsius.

This confirms that the extreme temperatures are raising, which would mean, people are affected more in the hot places.

8.

9. Variation in Minimum Temperatures:

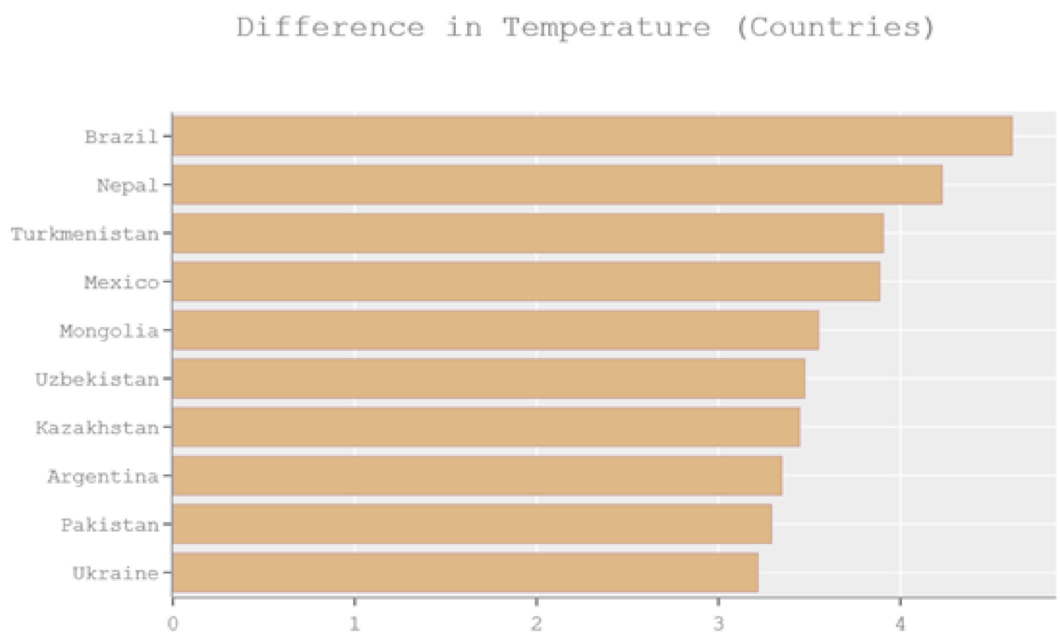


The minimum temperature also rose by 2.5 degree Celsius, which is causing the extension of beautiful glaciers form the planet.

10.

Countries rank - from the biggest to smallest increase in temperature

Another approach to understanding climate change and its factors could be to analyze the areas that have maximum temperature change over the years and then identifying the major events or trends that took place in those areas which can potentially affect climate in the longer run. We averaged the temperature of all the cities by year and then grouped them by the countries, regions and continents and grouped them by year for different analysis. To avoid any outlier, we ranked the countries based on their rise in temperature from mean over the years.

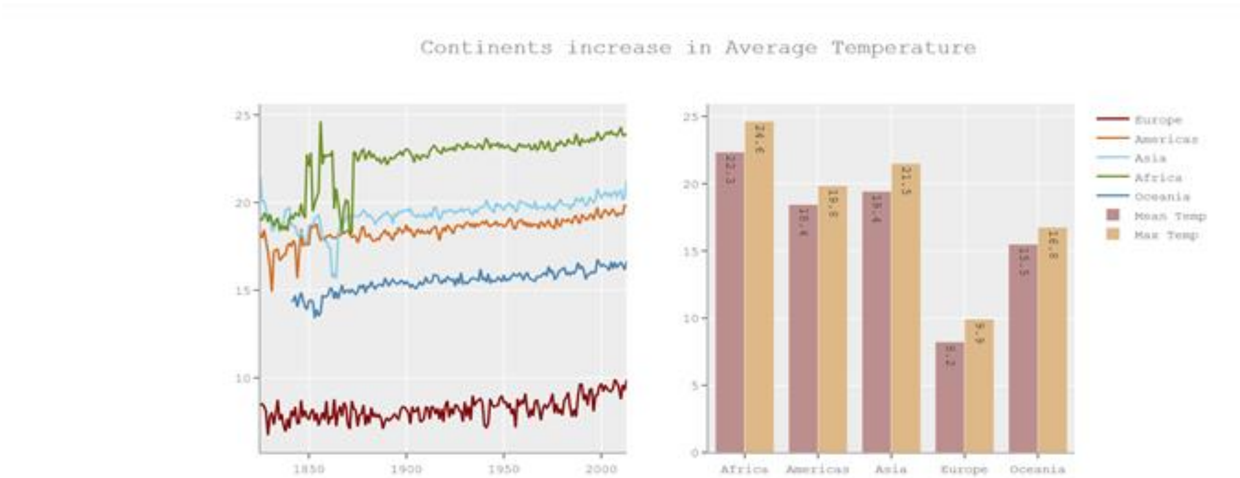


We have listed below the top 10 countries that have seen maximum change in their climate. Further analyzing the economy and environmental factors of those areas. We found some major events in countries like below that they need to focus on:

- **Brazil and Argentina** – We found that agriculture has increased by multiple folds in these areas which further have caused huge amounts of deforestation. The same can be seen from the decrease in forest cover % over the years.
- **Kazakhstan** – Increase in industrialization at an improper rate, without appropriate environmental guidelines. Also, Kazakhstan is a playground for different types of biological and nuclear weapons testing
- **Turkmenistan** - Desertification and drying of the Arab Sea, due to heavy agricultural practices.
- **Nepal** - Air Pollution is a major concern in Nepal. We find that although Nepal stands low in the population factor (i.e. rise in population), there other factors that we need to consider like increase in greenhouse and CO2 emission per capita to accurately measure climate change.

Continents Increase in Average Temperature

When we look at the trend of average temperature in different continents, ignoring the initial fluctuations as the temp data for south America and south Africa were not accurately recorded in the 1850

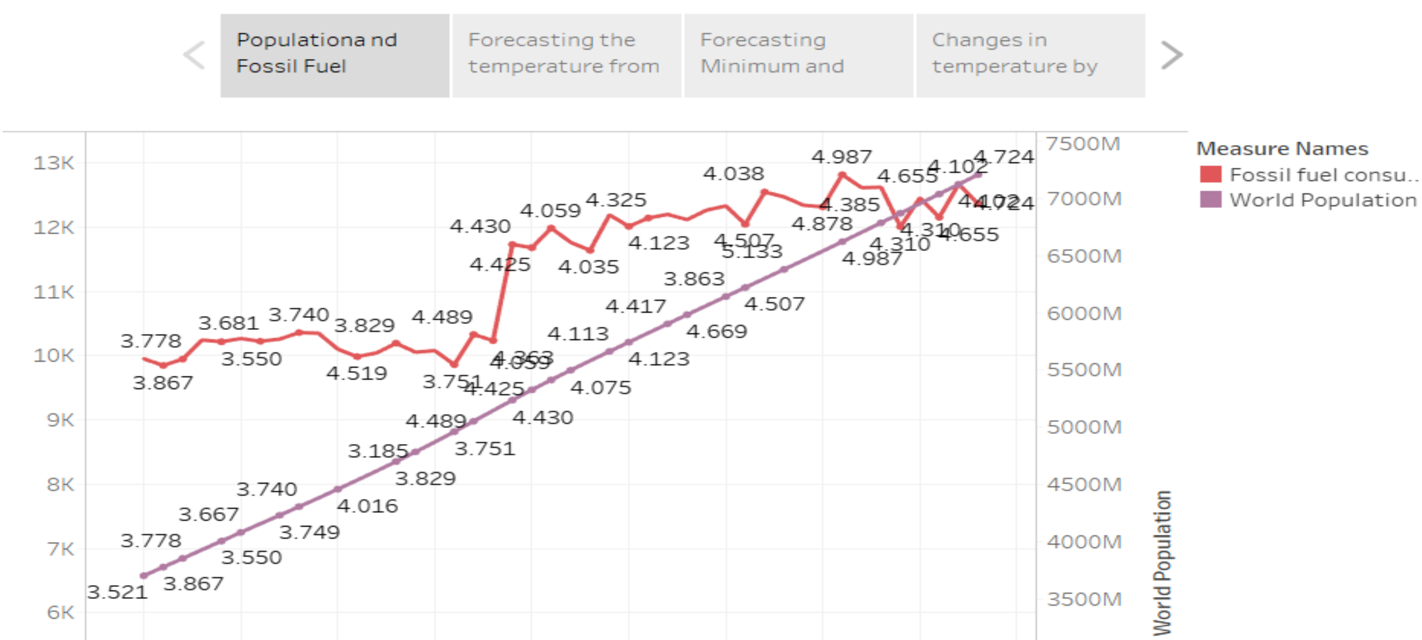


’s.

We see that the Temperature for all the continents have increased and All the continents look to have increased in Temperature in about the same pace. Africa has the highest average temperature and maximum rise (2.3 degrees) followed by Asia (with a rise of 2.1 degrees). One major factor for Higher temperature in Africa is its geological position. Population is the primary factor for high temperature in Asia. In contrast to Africa and Asia, we also see that the EU has the least temperature.

Temperature with Trend of Fossil fuels and Population:

https://public.tableau.com/profile/chakradhar6487#!/vizhome/PopulationandFossilFuelExtractionwithAverageTemperaturechange/GlobalTemperaturechangewithPopulationandFossilFuels?publish=yes



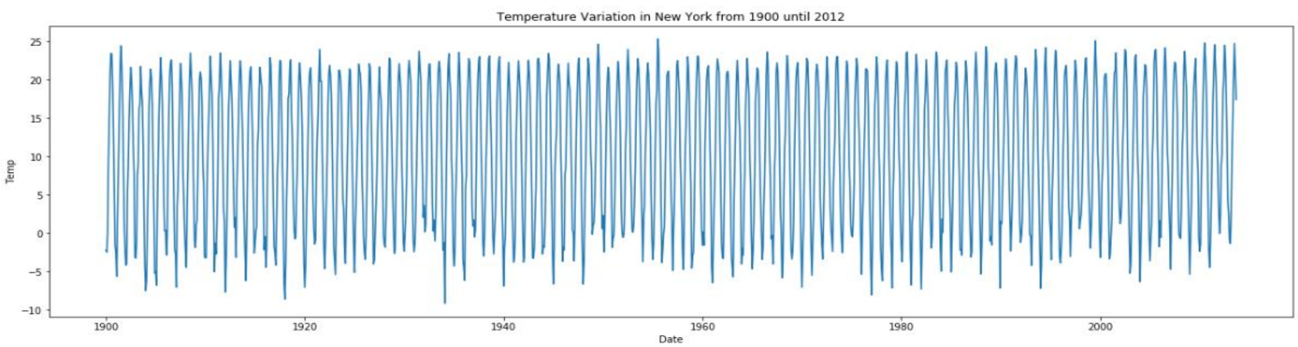
We can here see that as the population grew, after rapid consumption of energy started in 1970's, we can see how temperature has increase by at least 1 degree Celsius within last 45 years, where earlier the rise was 0.5 degree Celsius for 120 years.

11. NEWYORK DATASET-Most Populated cities

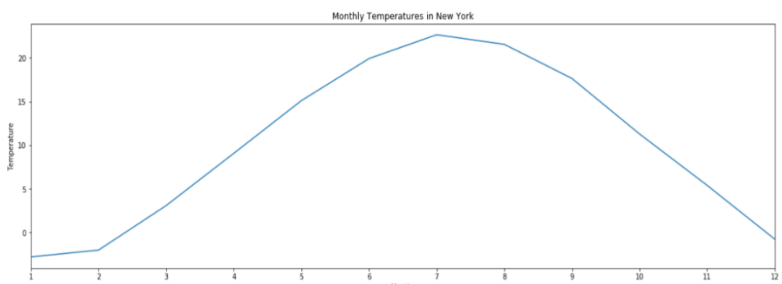
Missing Values- Data cleaning:
Our original dataset had temperature variation from 1765 to 2013 but most of the rows from 1765 – 1899 did not have temperature values available so we sampled our dataset using temperatures from 1900 till 2013. Additionally, we ignored September 2013 temperature as this was null in our original dataset. Temperature variation was monitored based on cities.

Data exploration: To understand the trend and seasonality in climate change using temperature variation over 100 years as the outcome variable

Temperature variation in New York from 1900 to 2012:

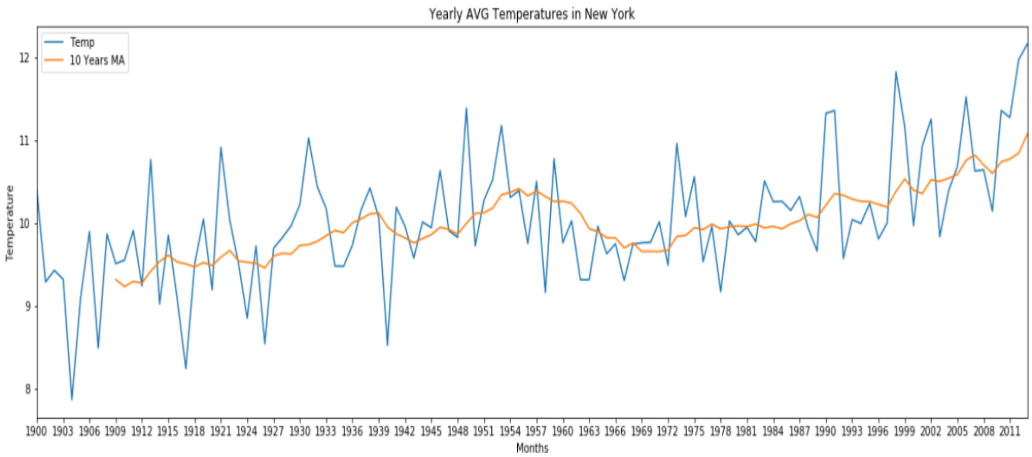


Yearly Seasonality in New York temperature series:



Series follow a parabolic curve - lower temperatures during January, February, March, November and December and higher temperatures between April and October.

Trend in New York temperature series using Moving average:



We can confirm that there is a constant increasing trend with slight decrease during 1969 to 1986 and that the average temperature increased from 8º to 12º, that is 50% increase in over 100 years!

Data preparation (splitting)

We split the data into training (model building), validation (model checking) and testing (final evaluation) dataset to ensure the model built will not overfit the new data.

Here, we split the data as

Training set- from year 1900 – September 2008- 1305

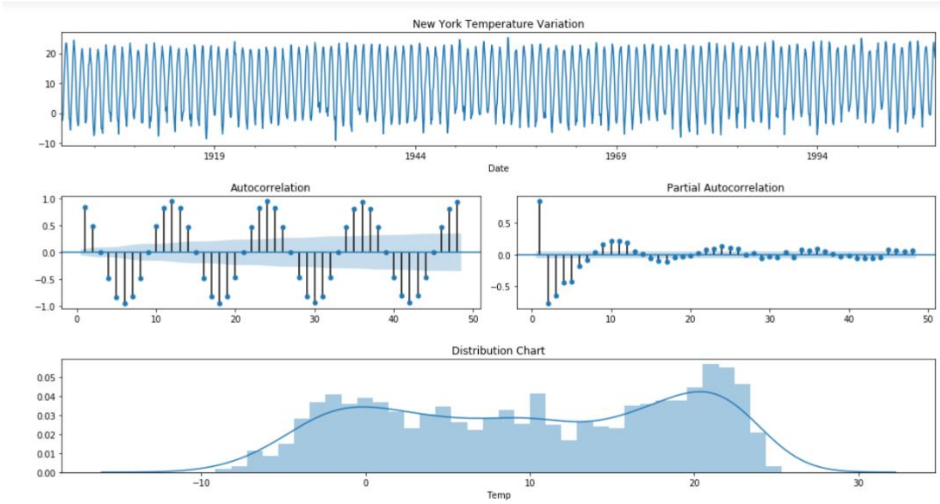
Validation set– October 2008 - September 2012-48

Testing set– October 2012 – September 2013-12

1. **SERIES ANALYSIS:** As we are dealing with time series, we want to make sure our original series is **stationary** and no influence of **white noise** to ensure that the model forecast using this time series is valid. Stationary test was done using the Augmented Dickey-Fuller test to check trend and seasonality in the original series. We used the Ljung-Box test to check white noise in the original series and residual analysis.

Temperature series analysis for most populated cities (New York in this case)

Stationary test

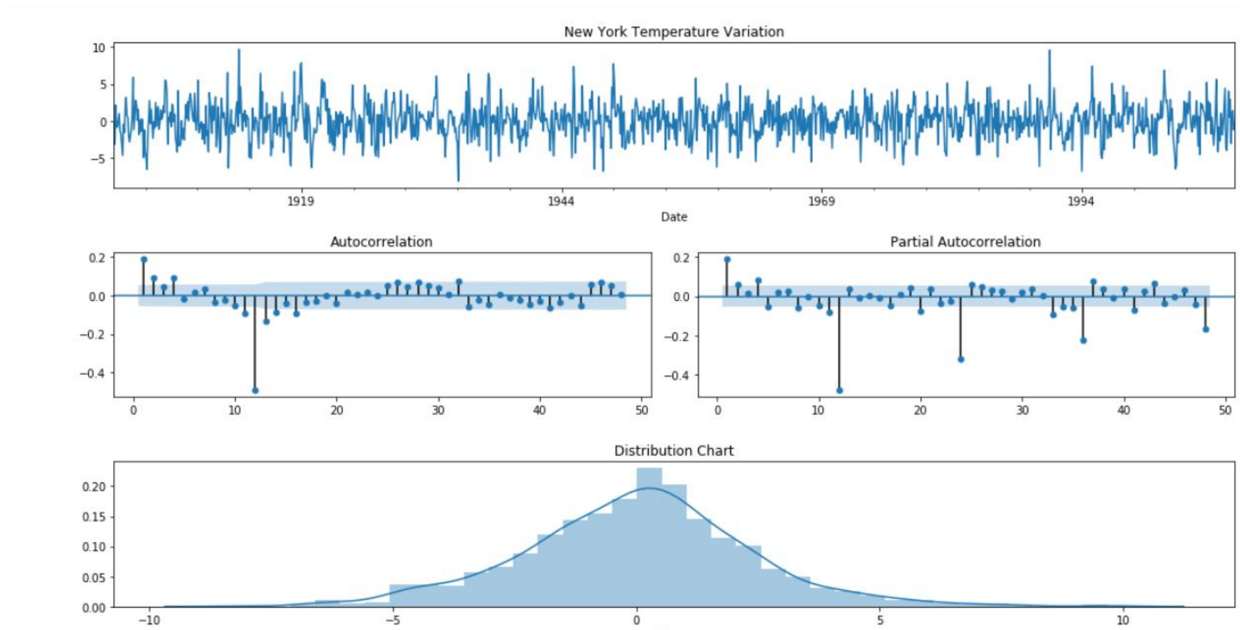


The series has an interesting behavior, there is a sequential significative negative autocorrelation starting at lag 6 and repeating each 12 months, it's because of the difference in the seasons, if today is winter with cold temperatures in 6 months we will have higher temperatures in the summer, that's why the negative autocorrelation occurs. These temperatures usually walk in opposite directions. Also, from lag 12 and sequentially from every 12 lags there is a significant positive autocorrelation.

The PACF shows a positive spike in the first lag and a drop to negative PACF for following lags.

This behavior between the ACF and PACF plots suggests an AR(1) model and also a first seasonal difference (Y(t)-Y(t-12)).

As we see some seasonality in ACF checking stationarity with lag = 12



As the plots above showed, the first ACF lags have a gradual decay with one spurious significant spike at lag 12, while the PACF drops under the confidence interval after the second lag, this is an AR signature with a parameter of 2, so this is an AR(2) model. Initially we are going to work with the following (p,d,q) orders: (2, 0, 0), and with the following seasonal (P, D, Q, S) orders (0,1,0,12) and as the series has a clear uptrend we are going to use it in the model.

White Noise test:

sm.stats.acorr_ljungbox(NY['Temp'], lags=[12],boxpierce=True)

```
Out[125]: (array([7597.22169562]), array([0.]), array([7547.29511899]), array([0.]))
```

The 2nd list in the returned array is p-values from the Ljung test and the 4th list is the p-values from the Box pierce test. Since p-value < .05, we reject the null hypothesis: Series is white noise. Both tests agree that the original series (each of its first 12 lags with > 95% confidence level) are not white noise.

MODELING:

Univariate model

As we have seen above, the data is seasonal with cycle of 12. Therefore, we will be using SARIMAX algorithm and predict the temperature. The important factors in developing a SARIMAX model are order and seasonal order. To select the order, we will be using AIC and BIC score. The range for the order is taken from ACF and PACF graphs. From fig M.1, we can see that p=1 and q=1 with d=1 is the order for the SARIMAX model.

Testing the fitted model on validation set and calculating RMSE:

```
#ARIMA(1, 1, 1)x(0, 1, 1, 12)12 - AIC:4933.10428 is the best
# Testing the walk-forward function on the validation set
val['Pred'] = walk_forward(train['Temp'], val['Temp'], ((1,1,1),(0,1,1,12),'c'))
```

```
# Measuring the error of the prediction
rmse_pred = measure_rmse(val['Temp'], val['Pred'])

print(f"The RMSE of the SARIMA(1,1,1),(0,1,1,12),'c' model is {round(rmse_pred,4)} celsius degrees")
print(f"It's a decrease of {round((rmse_pred/rmse_base-1)*100,2)}% in the RMSE")
```

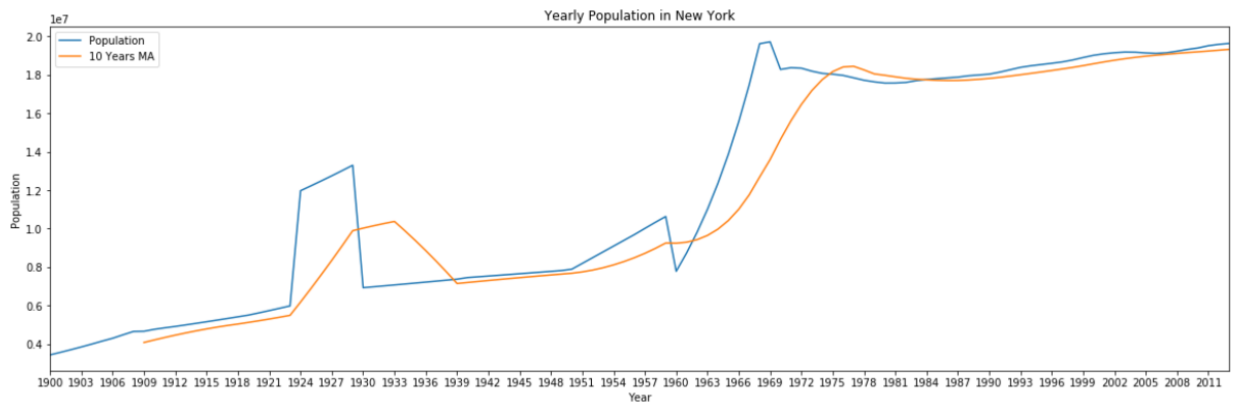
```
The RMSE of the SARIMA(1,1,1),(0,1,1,12),'c' model is 1.5138 celsius degrees
It's a decrease of -69.15% in the RMSE
```

Model with exogenous variable as Population

Before we create the model using any exogenous variable we checked if the series itself has white noise and check for stationarity was done

Exploring Population series

From the below graph we can confirm that there is a constant increasing trend with slight decrease during 1969 to 1986 and that the average population increased from 0.4*10^7 to 2*10^7 in over 100 years!



White Noise test for Population series

```
#Checking whether Population series has white noise
sm.stats.acorr_ljungbox(NY_pop['Population'], lags=[12],boxpierce=True)
```

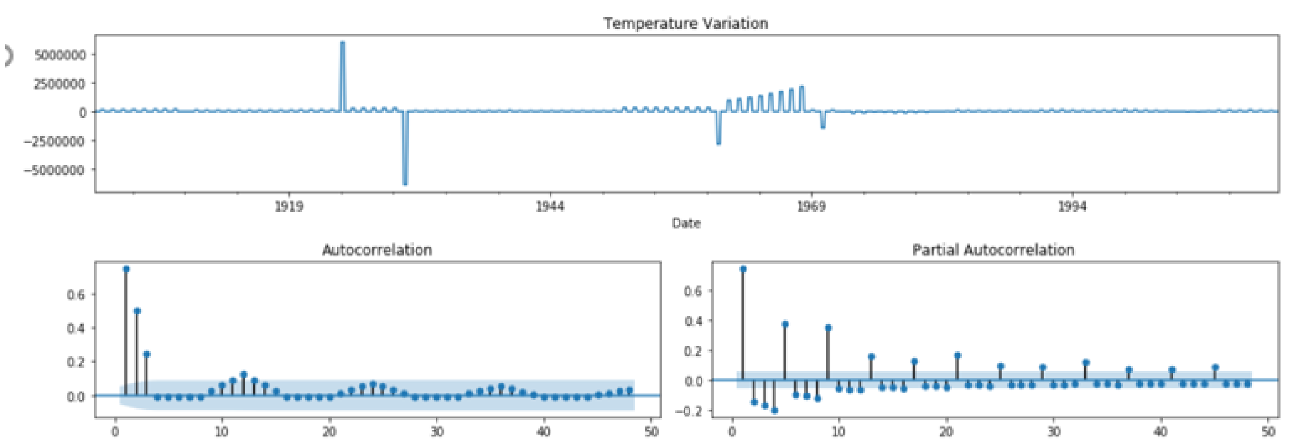
```
(array([15928.01682375]), array([0.]), array([15829.61227337]), array([0.]))
```

As the 2nd list in the returned array is p-values from the ljungbox test and the 4th list is the p-values from the boxpierce test. Since p-value < .05,we reject the null hypothesis: Series is white noise. Both tests agree that original series(each of its first 12 lags with > 95% confidence level) are not white noise .

Stationary test for Population series

As the original series isn't stationary, we applied first differencing to make this stationary.

As the plots above showed, ACF has significant lags till lag 3 followed by a kind of pattern from lag 11. But those are not that significant so we can ignore that pattern as of now. PACF becomes negatively correlated and exponentially decays and becomes insignificant after lag 11 so this shows a pattern of MA(1) model or we can try till MA(11) just to check if there any significance of lag =11 or that it's a spurious one.



Model with Population as a regressor

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	1305
Model:	SARIMAX(1, 0, 1)x(0, 1, [1, 2], 12)	**Log Likelihood**	-2618.274
Date:	Sun, 26 Apr 2020	**AIC**	5250.547
Time:	18:42:20	**BIC**	5286.700
Sample:	01-01-1900	**HQIC**	5264.116
	- 09-01-2008		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.0069	0.010	0.694	0.488	-0.013	0.026
Population	-4.804e-08	6.99e-08	-0.687	0.492	-1.85e-07	8.89e-08
ar.L1	0.4921	0.200	2.455	0.014	0.099	0.885
ma.L1	-0.3113	0.219	-1.422	0.155	-0.740	0.118
ma.S.L12	-0.8229	0.047	-17.424	0.000	-0.915	-0.730
ma.S.L24	-0.0527	0.045	-1.160	0.246	-0.142	0.036
sigma2	5.2701	0.298	17.694	0.000	4.686	5.854

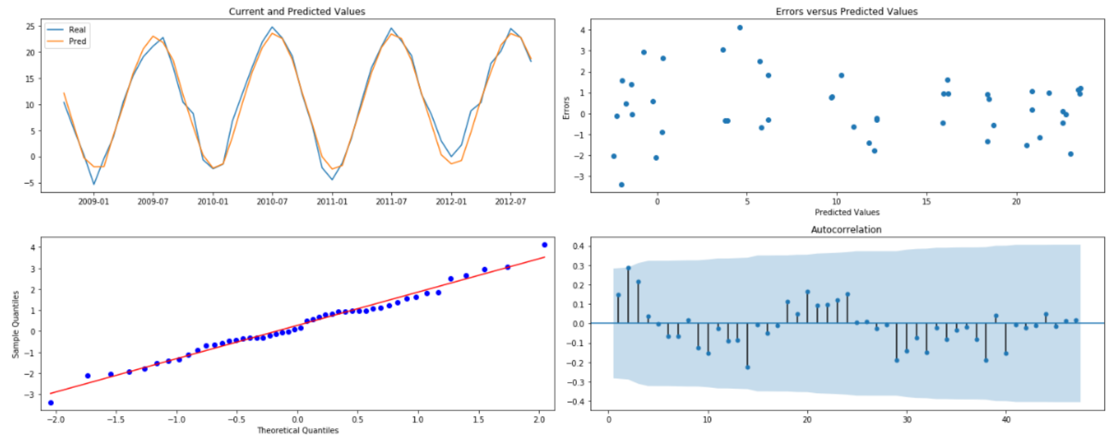
Ljung-Box (Q):	64.14	**Jarque-Bera (JB):**	52.82
Prob(Q):	0.01	**Prob(JB):**	0.00
Heteroskedasticity (H):	0.92	**Skew:**	-0.07
Prob(H) (two-sided):	0.38	**Kurtosis:**	3.98

SARIMAX (1, 0, 1) x (0, 1, [1, 2], 12) with AIC: 5250.547 is the best model we get after hyperparameter tuning.

But the coefficient of population is very less than 0.05 so this regressor has very little contribution in the model's prediction. This might be because the Population data is yearly however rest of the data i.e. temperature variable is monthly.

Residual Analysis on univariate:

```
In [105]:  plot_error(val)
```



Analyzing the plots above we can see that the predictions fit very well on the current values.

- The Error vs Predicted values have a linear distribution.
- The QQ Plot shows a normal pattern with some little outliers and,
- The autocorrelation plot shows no significant lags (outside 95% Confidence Interval).

White Noise test on residuals:

sm.stats.acorr_ljungbox(val['Error'], lags=[12],boxpierce=True)

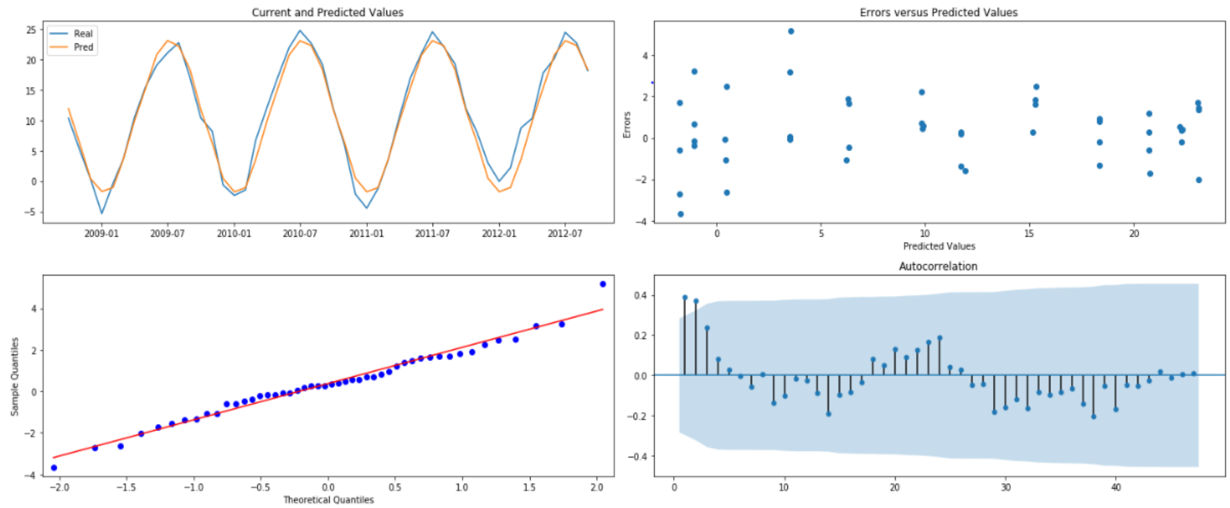
```
Out[122]: (array([24.5339867]),
           array([0.43142599]),
           array([17.55427811]),
           array([0.82407999]))
```

The 2nd list in the returned array is p-values from the Ljungbox test and the 4th list is the p-values from the Boxpierce test. Since p-value > .05, both tests agree that residuals (each of its first 12 lags with > 95% confidence level) are white noise. This is a good indication of a valid model.

Residual analysis for model using Population as regressor

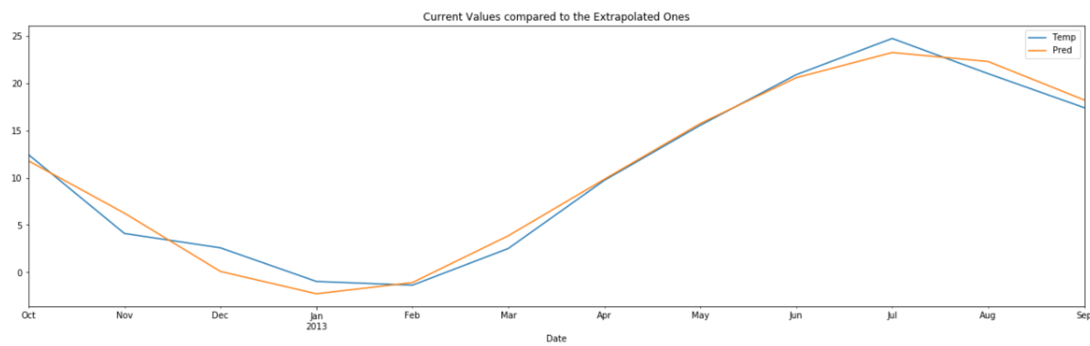
Analyzing the residual plots below we can see that the predictions fit very well on the current values. The Error vs Predicted values have a linear distribution. The QQ Plot shows a normal pattern with some little outliers and, But the autocorrelation plot shows lag=1 and lag=2 as significant (outside 95% Confidence Interval). Hence, the model after adding population parameter is less superior than the model without exogenous variable.

```
plot_error(valpop)
```



#Predictions on test set using Univariate model:

In [111]:
```
test[['Temp', 'Pred']].plot(figsize=(22,6))
plt.title('Current Values compared to the Extrapolated Ones')
plt.show()
```



Predictions follow the same pattern as the original time series. Evaluating RMSE on test set:

In [126]:
```
#Evaluating the model with the RMSE in the test set (baseline against the extrapolation):
test_baseline = test['Temp'].shift()

test_baseline[0] = test['Temp'][0]

rmse_test_base = measure_rmse(test['Temp'],test_baseline)
rmse_test_extrap = measure_rmse(test['Temp'], test['Pred'])

print(f'The baseline RMSE for the test baseline was {round(rmse_test_base,2)} celsius degrees')
print(f'The baseline RMSE for the test extrapolation was {round(rmse_test_extrap,2)} celsius degrees')
print(f'That is an improvement of {-round((rmse_test_extrap/rmse_test_base-1)*100,2)}%')
```

```
The baseline RMSE for the test baseline was 4.61 celsius degrees
The baseline RMSE for the test extrapolation was 1.28 celsius degrees
That is an improvement of 72.36%
```

Predictions on test set for model using Population as regressor:

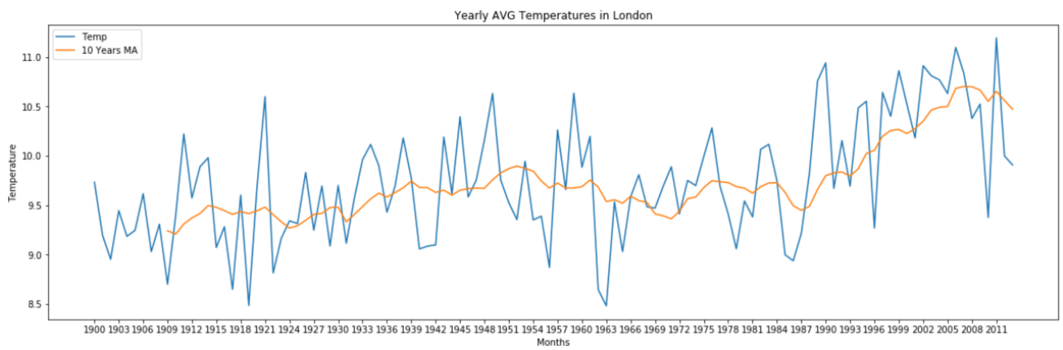
As the residual analysis of the model with Population as regressor wasn't satisfactory because autocorrelation plots have significant lags, we didn't use this model for forecasting temperature change.

12. Analyzing data Country wise:

Here we clustered our data set based on longitude and latitude and selected UK as a country and most populated city London.

Trend in Countries temperature series using Moving average:

```
year_avg = pd.pivot_table(gtemp, values='Temp', index='year', aggfunc='mean')
year_avg['10 Years MA'] = year_avg['Temp'].rolling(10).mean()
year_avg[['Temp','10 Years MA']].plot(figsize=(20,6))
plt.title('Yearly AVG Temperatures in London')
plt.xlabel('Months')
plt.ylabel('Temperature')
plt.xticks([x for x in range(1900,2013,3)])
plt.show()
```



We can confirm that there is a constant increasing trend with slight decrease during 1969 to 1986 and that the average temperature increased from 9º to 10.5º

Data splitting

We split the data into training (model building), validation (model checking) and testing (final evaluation) dataset to ensure the model built will not overfit the new data.

Here, we split the data as

Training set- from year 1900 – September 2008

Validation set– October 2008 - September 2012

Testing set– October 2012 – September 2013

```
train = gtemp[:-60].copy()
val = gtemp[-60:-12].copy()
test = gtemp[-12:].copy()
print(f'Training set: {train.count()} ,validation set : {val.count()}, Testing set : {test.count()}')
```

```
Training set: Temp      1305
City      1305
month     1305
year      1305
dtype: int64 ,validation set : Temp     48
City      48
month     48
year      48
dtype: int64, Testing set : Temp     11
City      12
month     12
year      12
dtype: int64
```

```
def measure_rmse(y_true, y_pred):
    return sqrt(mean_squared_error(y_true,y_pred))

# Using the function with the baseline values
rmse_base = measure_rmse(val.iloc[1:,0],baseline)
print(f'The RMSE of the baseline that we will try to diminish is {round(rmse_base,4)} celsius degrees')
```

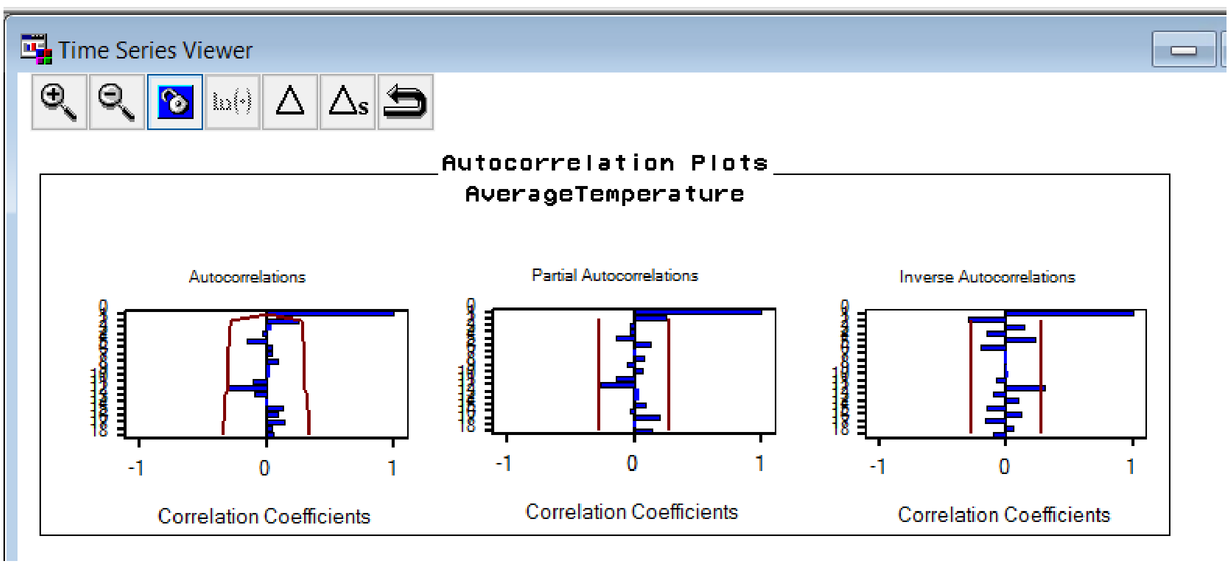
```
The RMSE of the baseline that we will try to diminish is 8.1455 celsius degrees
```

We checked the RMSE of our trained data set is 8.14 Celsius degree.

We want to make sure our original series is stationary and no influence of white noise to ensure that the model forecast using this time series is valid. We used the Ljung-Box test to check white noise in the original series and residual analysis. Checking the stationarity of the series:

Temperature series analysis for most Countries (UK and Russia in this case)

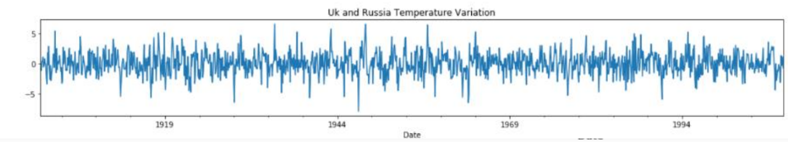
ACF AND PACF

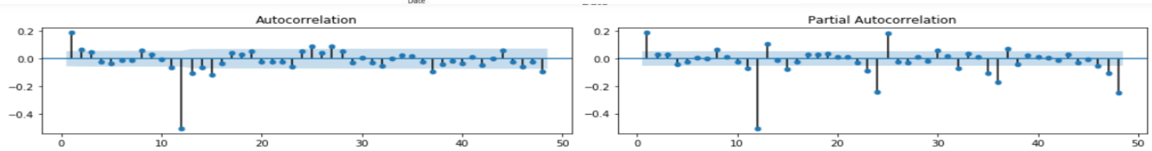


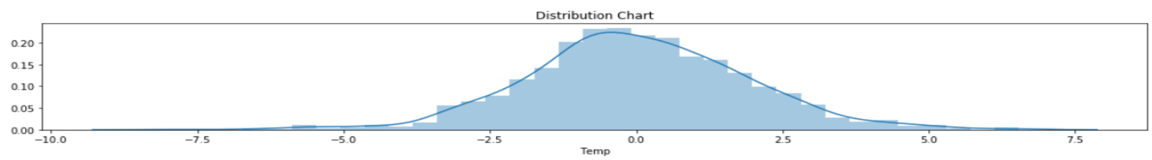
The series has an interesting behavior, there is a sequential significative negative autocorrelation starting at lag 6 and repeating each 12 months, it's because of the difference in the seasons, if today is winter with cold temperatures in 6 months we will have higher temperatures in the summer, that's why the negative autocorrelation occurs. These temperatures usually walk in opposite directions. Also, from lag 12 and sequentially from every 12 lags there is a significant positive autocorrelation. The PACF shows a positive spike in the first lag and a drop to negative PACF for following lags. This behavior between the ACF and PACF plots suggests an AR(1) model and also a first seasonal difference (Y(t)-Y(t-12)).As we see some seasonality in ACF checking stationarity with lag = 12. As the plots above showed, the first ACF lags have a gradual decay with one spurious significant spike at lag 12, while the PACF drops under the confidence interval after the second lag, this is an AR signature with a parameter of 2, so this is an AR(2) model. Initially we are going to work with the following (p,d,q) orders: (2, 0, 0), and with the following seasonal (P, D, Q, S) orders (0,1,0,12) and as the series has a clear uptrend we are going to use it in the model.

We tried to diminish this effect and hence used the differencing method.

```
In [164]:  check_stationarity(train['Temp'].diff(12).dropna())

Results of Dickey-Fuller Test:
Test Statistic                   -12.0324
p-value                            0.0000
Lags Used                         23.0000
Number of Observations Used     1269.0000
Critical Value (1%)               -3.4355
Critical Value (5%)               -2.8638
Critical Value (10%)              -2.5680
dtype: float64

The Test Statistics is lower than the Critical Value of 5%.
The serie seems to be stationary
```



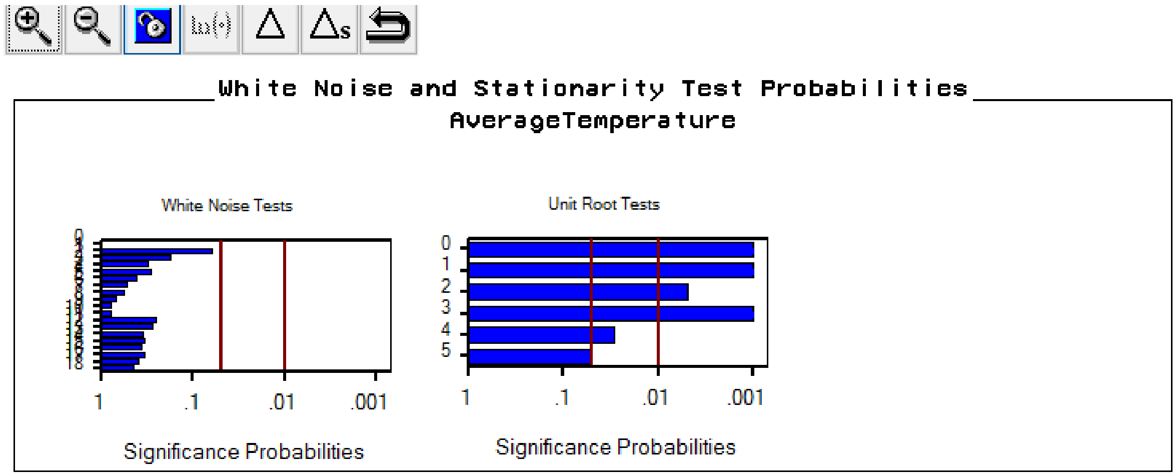





```
5]:  # Excluding the first line, as it has NaN values
     train['Temp'].dropna(inplace=True)
     baseline = val['Temp'].shift()
     baseline.dropna(inplace=True)
     baseline.head()

:[165]: Date
     2008-11-01    9.634
     2008-12-01    7.156
     2009-01-01    3.605
     2009-02-01    2.769
     2009-03-01    4.088
     Name: Temp, dtype: float64
```

Here, you can see the effect has been removed and now the time series has become stationary.

White Noise test



The 2nd list in the returned array is p-values from the Ljung test and the 4th list is the p-values from the Box pierce test. Since p-value < .05, we reject the null hypothesis: Series is white noise. Both tests agree that the original series (each of its first 12 lags with > 95% confidence level) are not white noise.

Models:

Model based on country with region same latitude and longitude (UK):

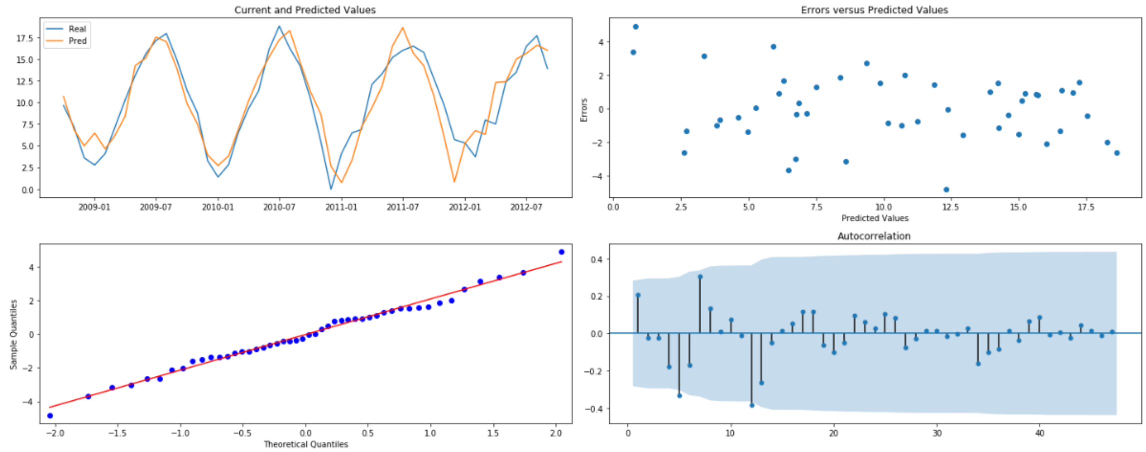
Univariate model (without adding any regressor Co2 and average temperature uncertainty):

As we see seasonality in series, we used a function to create Seasonal ARIMA models and we checked multiple Seasonal ARIMA models using various combinations of 'p' and 'q' and compared their AICs to get the best combination of 'p' and 'q'. For doing this we used the following function. From fig C.1, We can see the ARIMA(1,0,1)x(1,1,1,12) has the lowest AIC values: 11790.93. The highlighted one has the least AIC and hence the best model among the rest. Testing the fitted model on validation set and calculating RMSE:

```
In [16]:  # Let's test it in the validation set
          val['Pred'] = walk_forward(train['Temp'], val['Temp'], ((1,0,1),(1,1,1,12),'c'))

In [17]:  # Measuring the error of the prediction
          rmse_pred = measure_rmse(val['Temp'], val['Pred'])

          print(f"The RMSE of the SARIMA(1,0,1),(1,1,1,12),'c' model was {round(rmse_pred,4)} celsius degrees")
          print(f"It's a decrease of {round((rmse_pred/rmse_base-1)*100,2)}% in the RMSE")

          The RMSE of the SARIMA(1,0,1),(1,1,1,12),'c' model was 1.9802 celsius degrees
          It's a decrease of -33.65% in the RMSE

In [18]:  # Creating the error column
          val['Error'] = val['Temp'] - val['Pred']
```

Residual Analysis

21]:
```
plot_error(new)
```

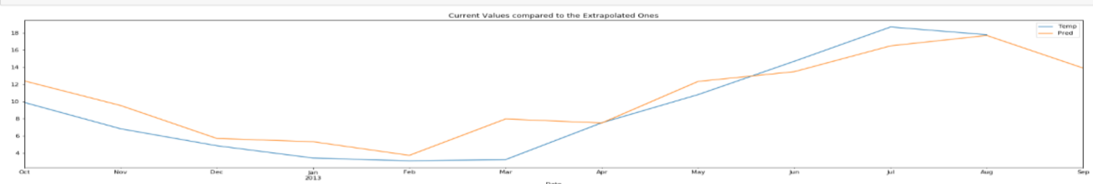


Analyzing the plots above we can see that the predictions fit very well on the current values.

- The Error vs Predicted values have a linear distribution.
- The QQ Plot shows a normal pattern with some little outliers and,
- The autocorrelation plot shows no significant lags (outside 95% Confidence Interval).

Prediction and extrapolated plots

```
test[['Temp', 'Pred']].plot(figsize=(28,6))
plt.title('Current Values compared to the Extrapolated Ones')
plt.show()
```



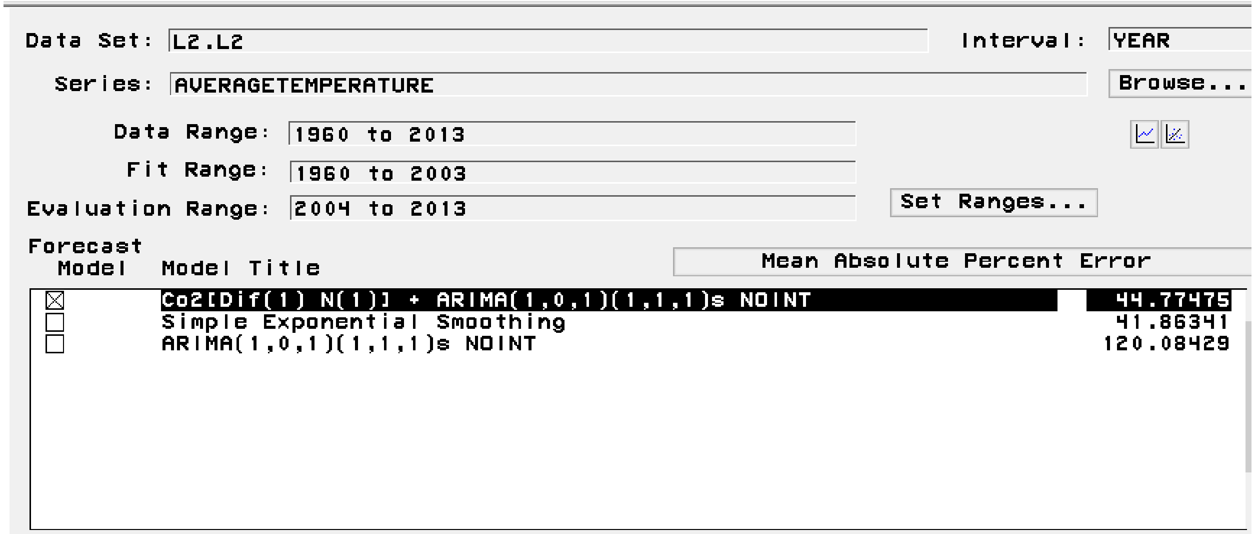
The RMSE for the extrapolated plots is 1.84 degree Celsius for the univariate model prediction.

SARIMA Model with Co2 and Average Temperature Uncertainty

Model with Co2 as a regressor

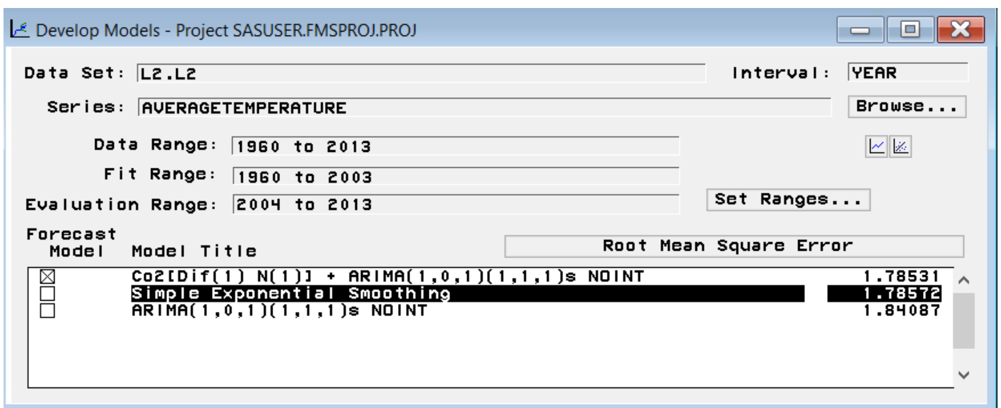
SARIMA (1, 0, 1)x(1, 1, [1, 2], 12) with AIC: 11790.93 is the best model we get after hyperparameter tuning.

MAPE error for the models

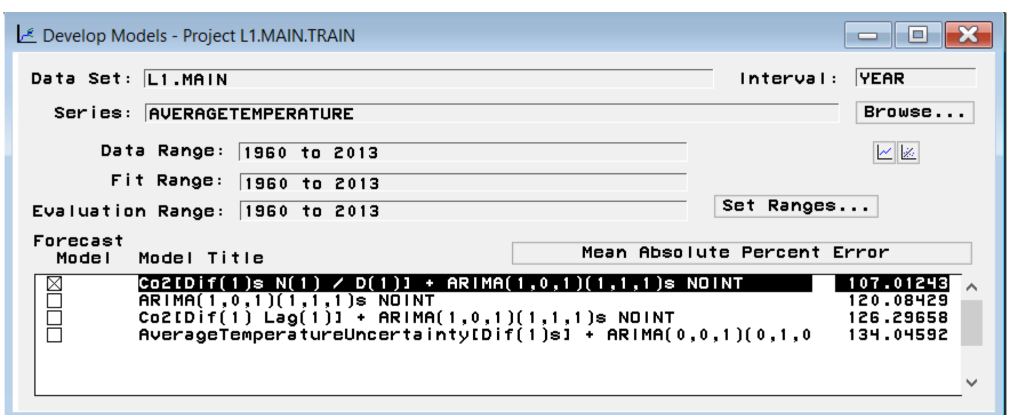


RMSE for the below models:Clearly visible that the model is performing better when Co2 as a regressor is added.

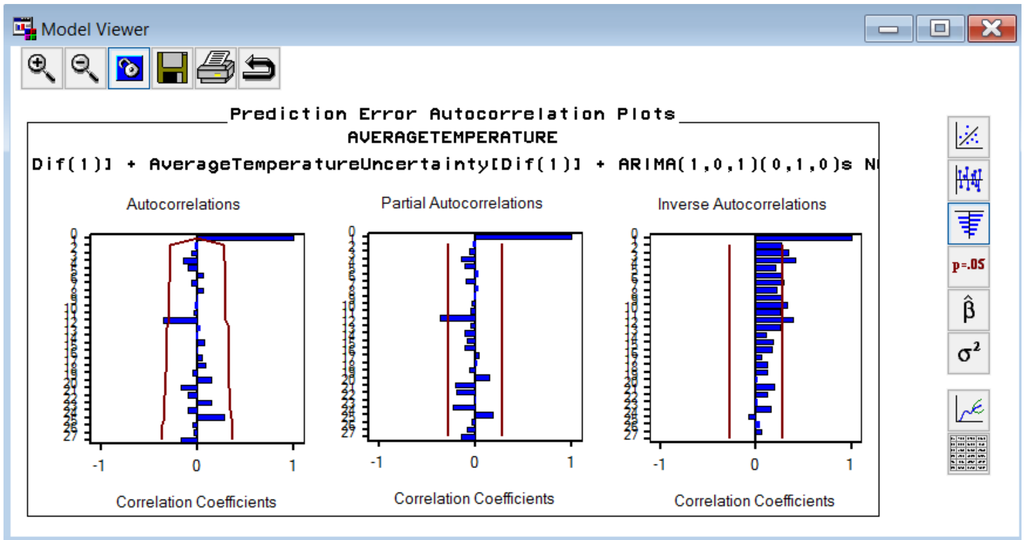
Tried to add the denominator = 1in the best model and then refitted the model to the holdout samples



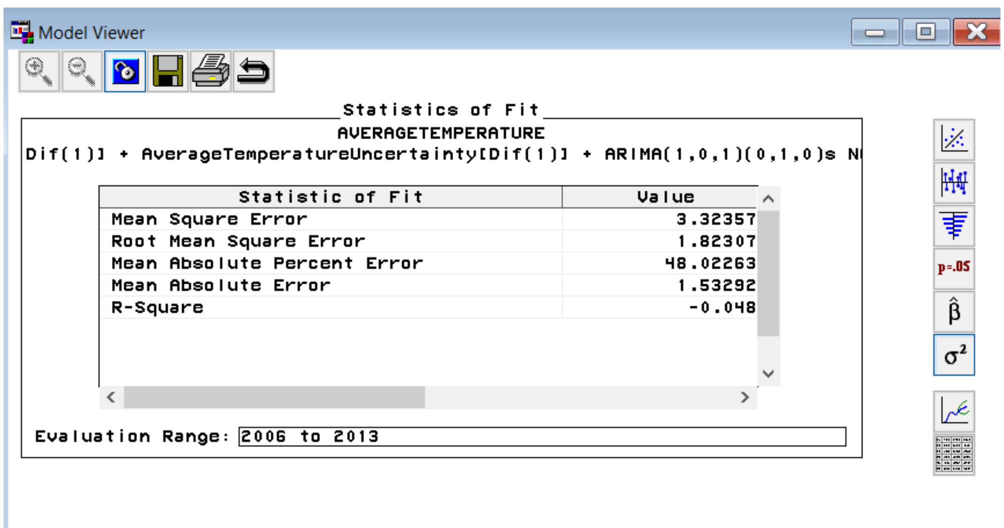
MAPE increased when model is refitted to the hold out samples.



ACF and PACF plots for CO2 and average temperature uncertainty as regressor in our best model.

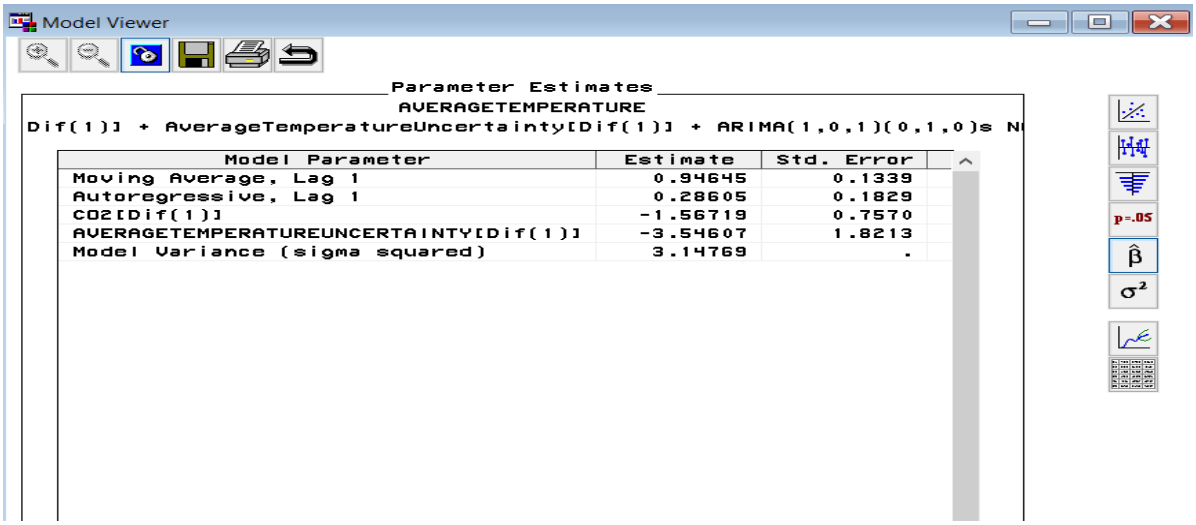


We can see that the model performance has been decreased.



Statistic of Fit	Value
Mean Square Error	3.32357
Root Mean Square Error	1.82307
Mean Absolute Percent Error	48.02263
Mean Absolute Error	1.53292
R-Square	-0.048

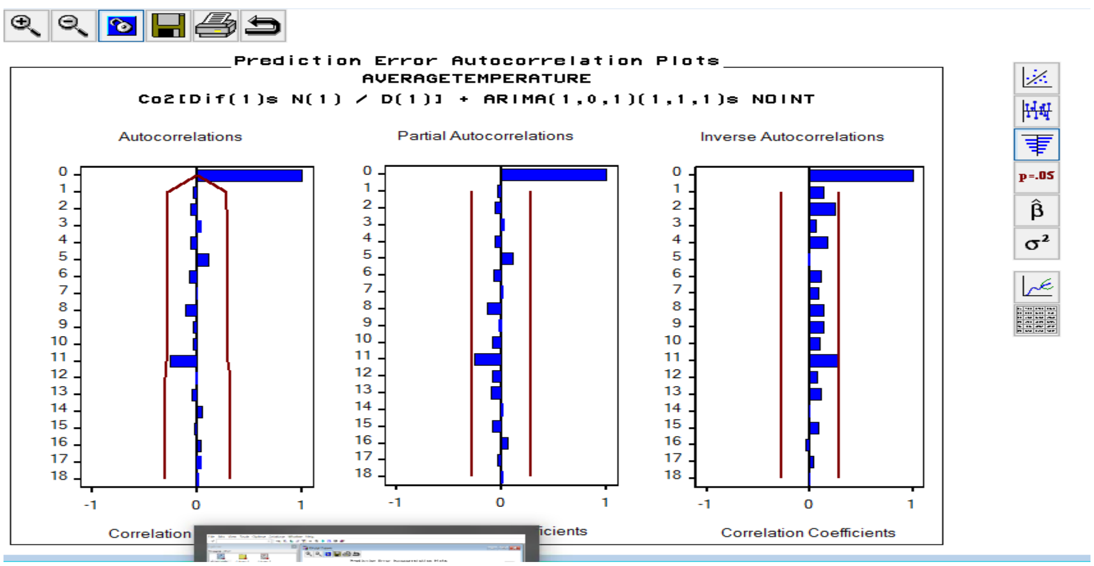
Parameter estimates of the model.


Model Viewer

Parameter Estimates
AVERAGETEMPERATURE
Dif(1)] + AverageTemperatureUncertainty[Dif(1)] + ARIMA(1,0,1)(0,1,0)s N

Model Parameter	Estimate	Std. Error
Moving Average, Lag 1	0.94645	0.1339
Autoregressive, Lag 1	0.28605	0.1829
CO2[Dif(1)]	-1.56719	0.7570
AVERAGETEMPERATUREUNCERTAINTY[Dif(1)]	-3.54607	1.8213
Model Variance (sigma squared)	3.14769	.

ACF and PACF of the best model.



Parameter estimates of the best model.

Model Viewer

Parameter Estimates
AVERAGETEMPERATURE
Co2[Dif(1)s N(1) / D(1)] + ARIMA(1,0,1)(1,1,1)s NOINT

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Moving Average, Lag 1	0.92572	2292	0.000404	0.9997
Seasonal Moving Average, Lag 1	0.92566	2294	0.000404	0.9997
Autoregressive, Lag 1	0.84092	2.4435	0.3441	0.7324
Seasonal Autoregressive, Lag 1	0.28380	0.3322	0.8543	0.3976
CO2[Dif(1)s N(1) / D(1)]	-1.45940	0.7982	-1.8284	0.0743
CO2[Dif(1)s N(1) / D(1)] Num1	0.80157	1.1648	0.6882	0.4949
CO2[Dif(1)s N(1) / D(1)] Den1	0.21749	0.7401	0.2939	0.7702
Model Variance (sigma squared)	3.53651	.	.	.

Fit Range: 1960 to 2013

Statistics of fit for the best model.

Statistics of Fit
AVERAGETEMPERATURE
Co2[Dif(1)s N(1) / D(1)] + ARIMA(1,0,1)(1,1,1)s NOINT

Statistic of Fit	Value
Mean Square Error	3.43507
Root Mean Square Error	1.85339
Mean Absolute Percent Error	107.01243
Mean Absolute Error	1.47934
R-Square	0.011

From the parameter estimates we can see the RMSE is almost constant but slightly improved. But the p-value of the regressor Co2 is greater than 0.05. Hence, this is not a significant parameter in our model.

More exploration on the effect of CO_2 and GHG emissions

--Cities selection

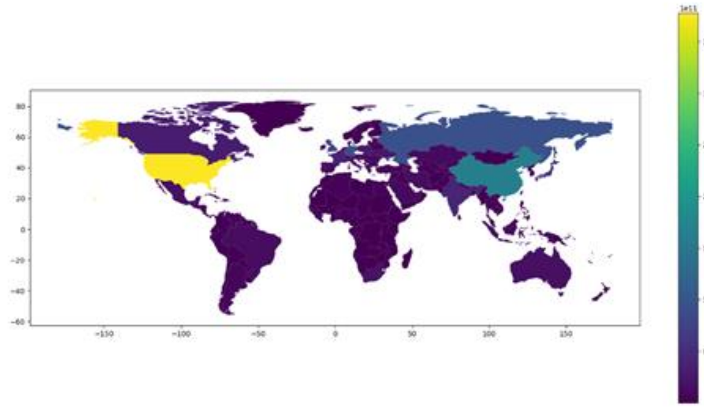


Figure 1: CO2 & GHG Emission of global countries and regions

Figure 1 shows CO_2 and GHG(Greenhouse Gas) emissions of global countries and regions. Apparently CO_2 and GHG emissions in the United States and China are much higher than that of other countries and regions. So we chose Peking as one of our objects since it's the capital of a country with very large CO_2 and GHG emissions.

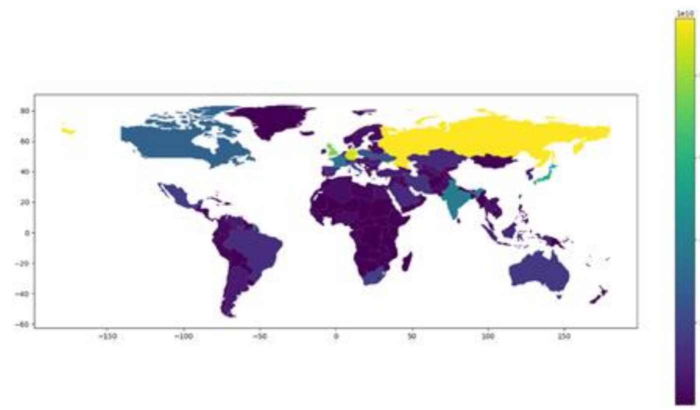
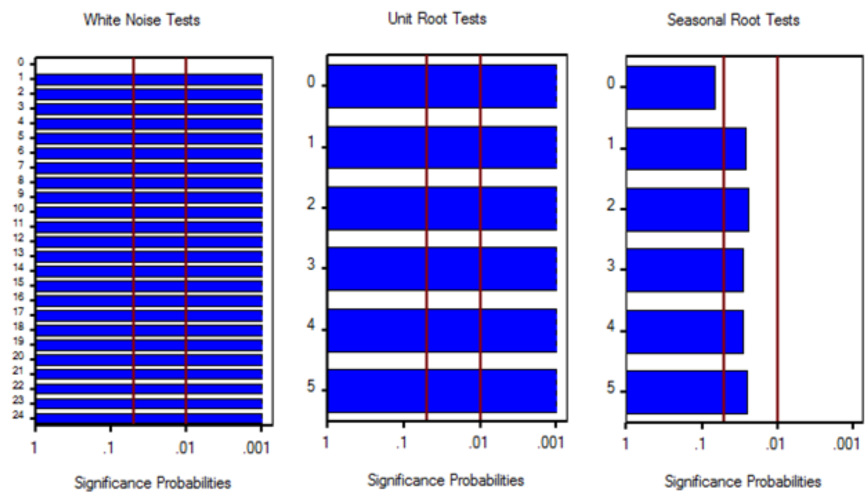


Figure 2: CO2 & GHG Emission of global countries and regions without USA and China

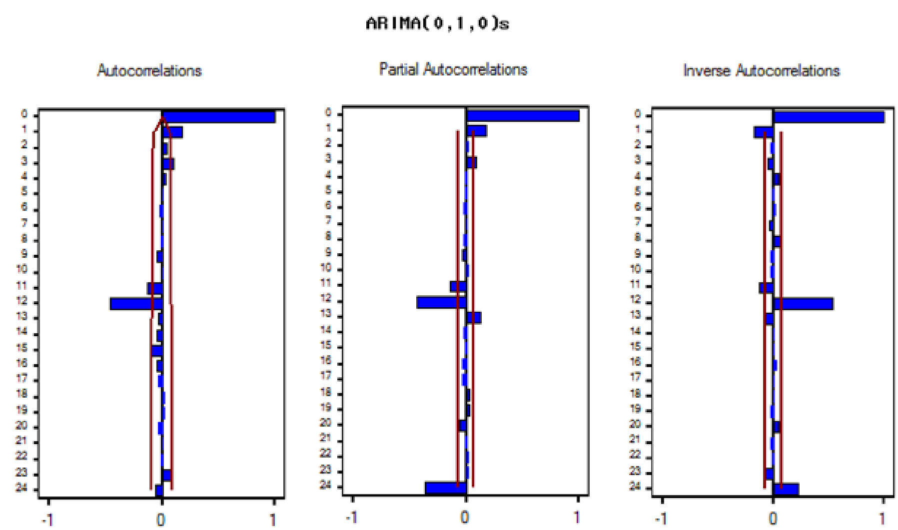
Figure 2 shows CO_2 and GHG(Greenhouse Gas) emissions of global countries and regions after removing the United States and Mainland of China for their numbers are so large. From this graph, it is easier to tell that the main area of Africa has much less CO_2 and GHG emissions. So we chose Mogadishu, capital of Somalia which is a country that does not produce much CO_2 and GHG.

TSFS on cities separately



From the original dataset of Peking, we could tell from figure 3 that obviously this is not white noise. And there is no clear sign to admit that the variation of temperature change in Peking has a trend. In addition, combined with our own experience, seasonal root test indicates that this data set follows a certain seasonality. The temperature tends to be high in summer and low in winter.

Figure 3: White noise and stationary test probabilities for Peking from 1950 to 2013



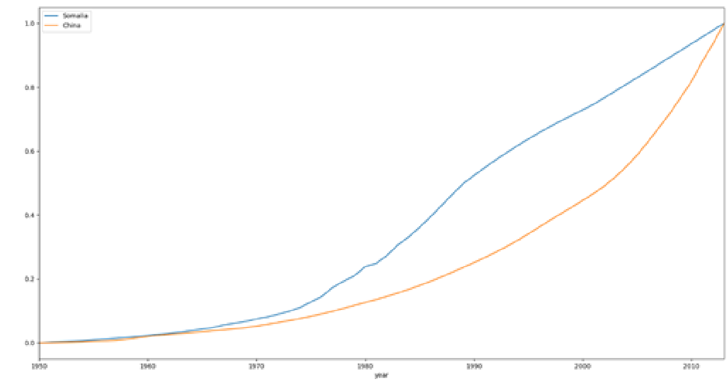
Therefore, from figure 4, the autocorrelations graph indicates that q is between 1 and 3. The rest two graphs indicate p is between 1 and 3. After several trials on the combination of different p and q, we found model ARIMA(1,0,0)(1,1,0)s is good and comparatively simple, whose RMSE is 1.501.

Figure 4: ARIMA(0,1,0)s on Peking from 1950 to 2013

Based on this model, we introduced CO_2 and GHG emissions as a dynamic regressor into it. This dynamic regressor was transformed into log format because the numbers in CO_2 and GHG emissions are much bigger compared to temperature which is normally between -20 and 50 degree. RMSE for this model is 1.498.

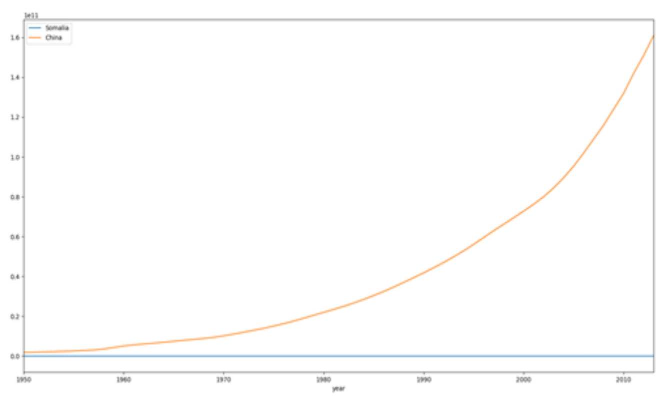
Same analysis and process was applied to Mogadishu, as well. Same results were found.

TSFS on the difference of two cities



From figure 5, we could tell that CO_2 and GHG Emissions in both cities increase over time. CO_2 and GHG emissions of Peking increase dramatically and that of Mogadishu increases steadily and slowly. The gap of CO_2 and GHG emission between these two cities also increases dramatically according to figure

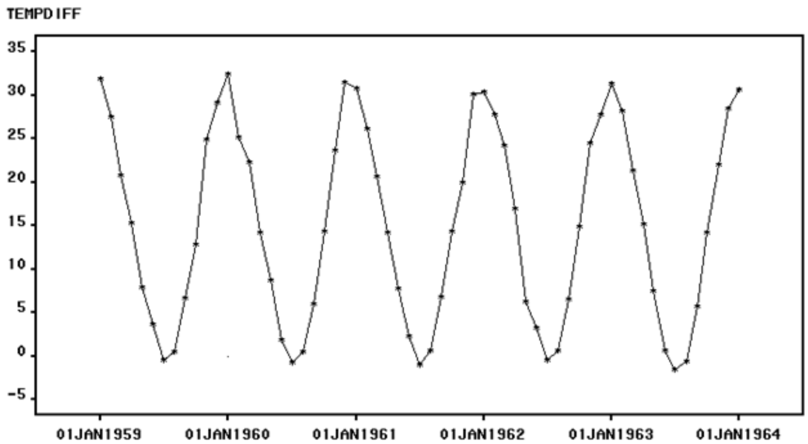
Figure 5: CO2 and GHG Emission of Peking and Mogadishu after standardized separately



In order to explore whether CO_2 and GHG emission has an effect on variation of average temperature of a region, we set the difference of the average temperature of two cities in every month as a target variable. And we set the difference of CO_2 and GHG emission of two cities as a regressor. From figure 7, the difference of the average temperature of two cities still follows seasonality.

Figure 6: CO2 and GHG Emission of Peking and Mogadishu

A good and simple TSFS model is ARIMA (1,0,1)(1,1,0)s, whose RMSE is 1.7395. Because the difference of CO_2 and GHG emission is so large. We introduced it as log regressor to model ARIMA (1,0,1)(1,1,0)s, which



only helped RMSE decrease by 0.002. And the P-value of this estimate is larger than 0.05, which means that the CO_2 and GHG emission does not have a clear effect on average temperature variation.

Figure 7: variation on the average temperature difference of Peking and Mogadishu

Therefore, we can come to a firm conclusion that CO_2 and GHG emission did not show much significance in having an impact on the average temperature of one country. It is probable that the flow of CO2 and GHG makes it hard to compare two cities only based on CO_2 and GHG emission difference.

13. Global Temperatures:

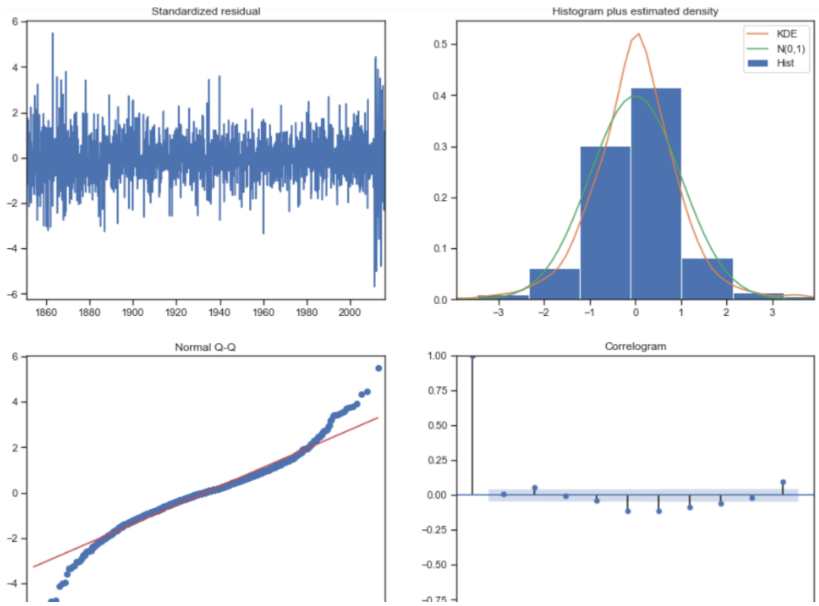
In this section, we will be predicting and extrapolating the prediction to predict the future for global average temperatures on the land. As again the series is seasonal.

From fig GT.1, we can see that the range for the order is 4 for p, as the ACF graph is tailing off. And for the seasonal order, we have used the seasonality factor as 12 and using the same ACF graphs, we can see from fig. GT 2 that P=0 and Q=1 will be our factors. With the range of (0,4), we have selected our model based on the lowest AIC and BIC which is an AR (4) model. Therefore p=2 and q=1 gives the best result with d=1.

The model is fitted with the data available till 2015 using the SARIMAX algorithm

```
                                 Statespace Model Results
==========================================================================================
Dep. Variable:                LandAverageTemperature   No. Observations:            1992
Model:            SARIMAX(2, 1, 1)x(0, 1, 1, 12)   Log Likelihood            -930.571
Date:                             Thu, 07 May 2020   AIC                       1871.141
Time:                                     05:55:49   BIC                       1899.093
Sample:                                 01-01-1850   HQIC                      1881.410
                                      - 12-01-2015
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.5433      0.015     36.323      0.000       0.514       0.573
ar.L2          0.0360      0.017      2.080      0.038       0.002       0.070
ma.L1         -0.9984      0.002   -496.403      0.000      -1.002      -0.994
ma.S.L12      -0.8312      0.008   -104.982      0.000      -0.847      -0.816
sigma2         0.1484      0.003     47.085      0.000       0.142       0.155
===================================================================================
Ljung-Box (Q):                      261.82   Jarque-Bera (JB):               926.31
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               1.04   Skew:                            -0.08
Prob(H) (two-sided):                  0.65   Kurtosis:                         6.35
===================================================================================
```

From the results, we can see that, for the Ljung-Box test the p value is less than 0.05, which means that the series is stationary. The results of Jarque-Bera test the p-value is less than 0.05, which means the residuals are not normally distributed.



Residual Analysis:

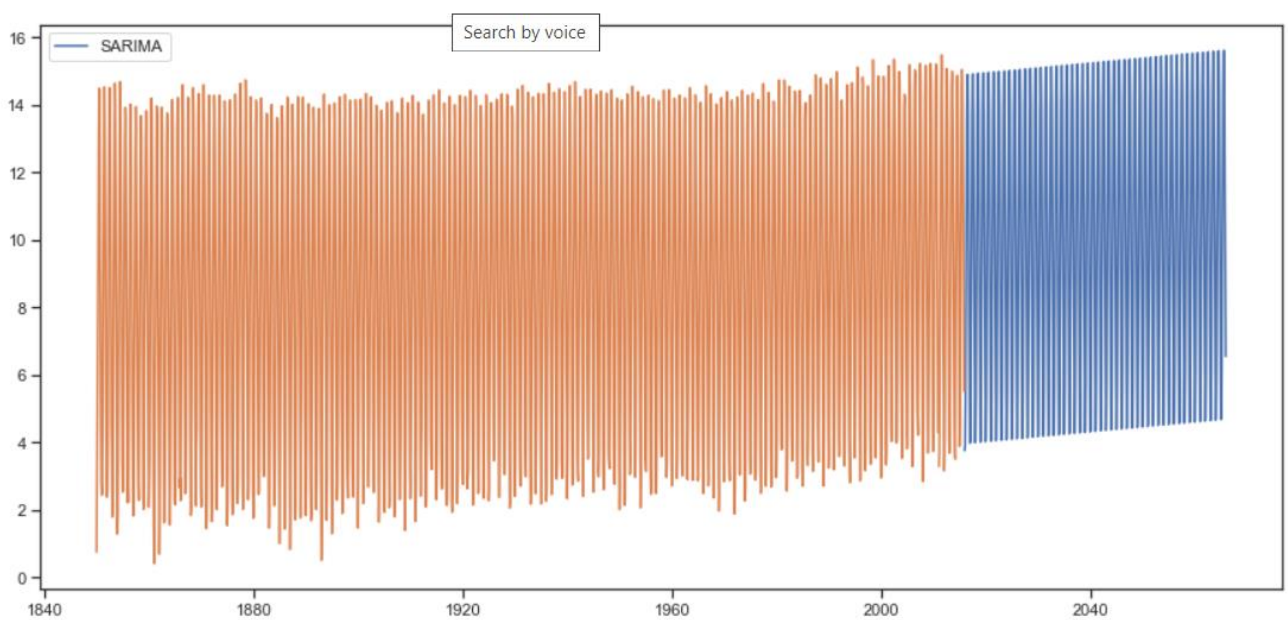
From the residual analysis, we can see a few outliers in the Normal Q-Q plot

```
mae = np.mean(np.abs(res.resid))

# Print mean absolute error
print(mae)
```

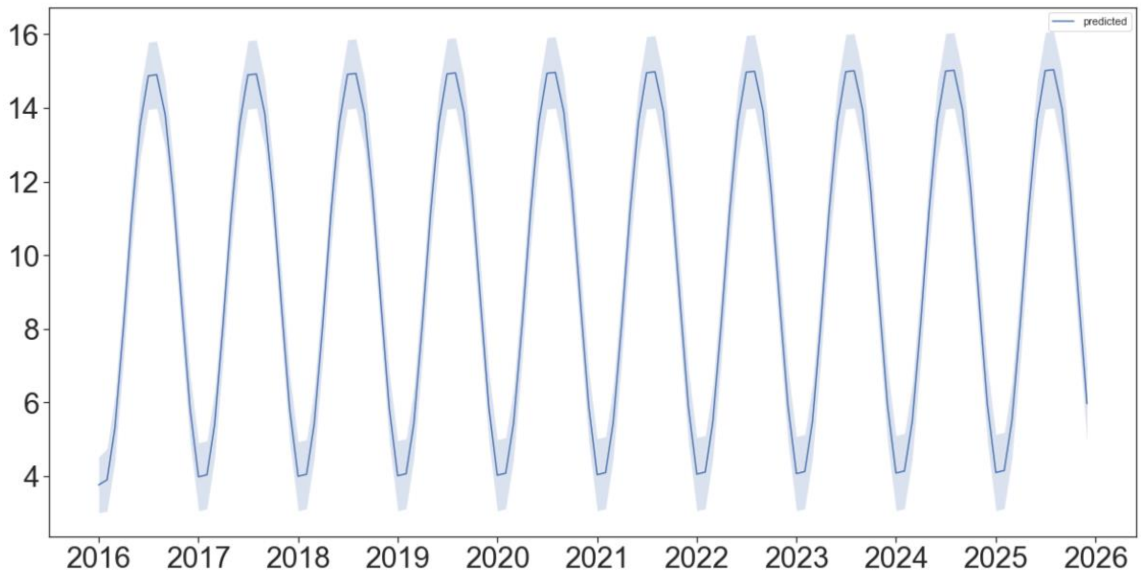
```
0.29044667135587393
```

The model is fitted with a mean absolute error of 0.29 degree with mean

Extrapolating for Next 50 years:



From the above figure, we can see that the model has captured both the trend and seasonality of the original series and with the confidence interval of 95%, the projections for next 10 years is presented below.



14. Forecasting

Forecasting Using Exponential Smoothing by Tableau:

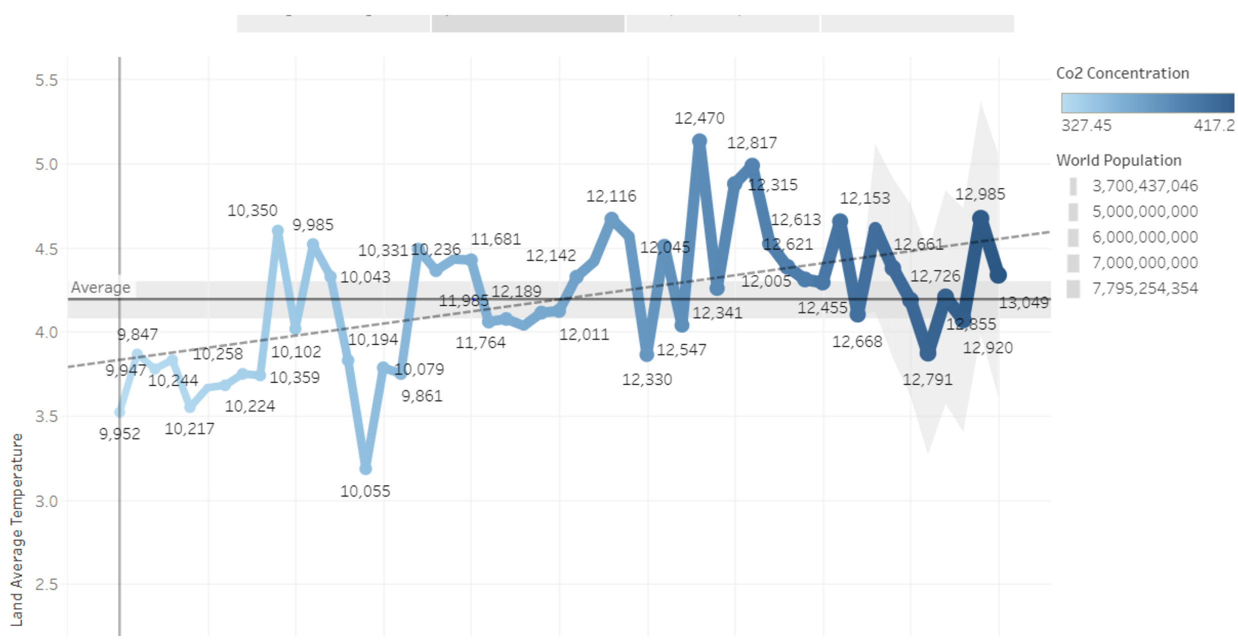


Tableau uses Exponential Smoothing to forecast. To interpret from the forecast, the forecast has not picked up the trend or it is learning too much from the past data. The forecast also has not picked up the seasonality factor. This also confirms that the forecast needs a lot of regressors that would affect the global warming and projections of all those factors.

Forecast of Country Temperature:

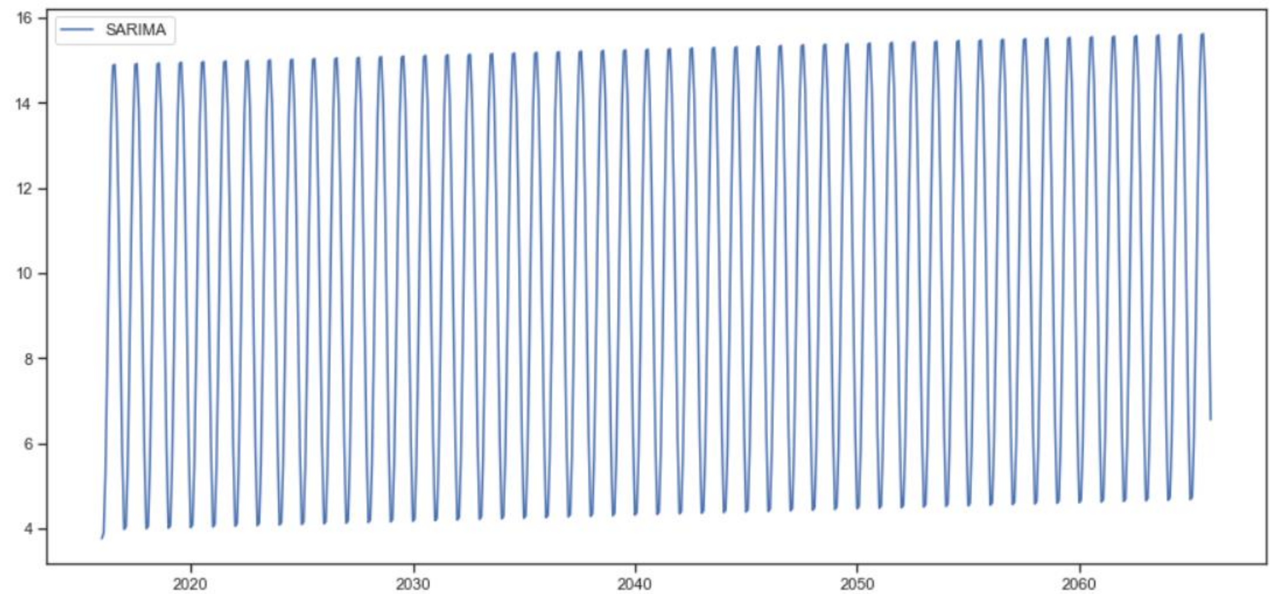
	DT	ACTUAL	Predicted value for AVERAGETEMPERATURE	Upper 95% Confidence Limit	Lower 95% Confidence Limit	Prediction error for AVERAGETEMPERATURE	Prediction standard error	Normalized prediction error for AVERAGETEMPERATURE	CO2[Dif(1)s N(1) / D(1)]
3	01/01/1963	-2.455	3.9022	8.5210	-0.7167	-6.3572	2.3566	-2.6976	11.2549
4	01/01/1964	2.225	-0.1276	4.1746	-4.4298	2.3526	2.1950	1.0718	11.2658
5	01/01/1965	3.214	1.6097	5.7515	-2.5321	1.6043	2.1132	0.7592	11.4562
6	01/01/1966	2.346	2.5480	6.5918	-1.4958	-0.2020	2.0632	-0.0979	11.3190
7	01/01/1967	3.954	3.0950	7.0722	-0.8822	0.8590	2.0292	0.4233	10.7841

```
abs(df['Normalized prediction error for AVERAGETEMPERATURE']).mean()
```

```
0.7597058823529411
```

With the forecasting SAS, we could predict the temperatures with an absolute error or 0.75 degree Celsius which is very good. We can therefore say that the model is working food.

Forecasting Global Temperatures for next 50 years:



From the SARIMAX model built above for the global temperature's dataset, we have forecasted the temperature for next 50 years. The forecast clearly shows an upward trend which confirms the fact and the average temperatures will increase about 1 degree in the next 50 years according to our model.

15. CONCLUSION

The model can accurately predict temperature variation for the next 2 years with decent RMSE. However, if we add Population as a regressor in the model the RMSE increases as we increase the number of years of prediction. This might have occurred as we had yearly population while the rest of the data was monthly. Additionally, we did not predict much into the future as that would reduce the credibility of forecasts as temperature changes depend more on various factors which occur daily rather than on monthly approximations.

The Analysis is a test of our Time Series skills to be executed using different technical skills like Python or SAP or Tableau and be able to use the process and steps to perform Time series analysis. The predictions are not to be used as climate changes would be affected by many other factors which are not used but this analysis has helped us in understanding better of Time Series Analysis to execute it on one of the toughest time series forecasts.

16.

17.

18. BUSINESS VALUE

Climate change is a long-studied topic and is still an active focus area. It is one important factor that influences every individual equally and every organization out there in one way or another.

Business value for Regulators and government:

This forecast in Global Warming would help cities to plan a cooling system architecture that would help it be viable to hold people who are already in the pollution clouds, also based on the rise in temperature and their preparedness it would help governments to plan in decreasing the CO_2 emission using renewable sources and using less fossil fuels. This will alarm the individuals for the immediate need for change in rules and regulations.

One Such example is of Delhi (India), where on most days the air quality index falls under the red alert zone, we see the government taking proper actions in response to an increase in greenhouse emission levels during Diwali Festival. This situation can be considered as a seasonal event having long term effect with gradual decay. Accordingly the government put an temporary ban on public sale of crackers during the season.

Again, a similar event was observed where sudden temperature rise was observed. Government studied the reasons for CO_2 emission in areas nearby. They observed the burning of large masses of land for agricultural use was the reason. And the government put out the 'Even-Odd' rule for vehicles for an immediate effect.

Similarly, this study will help us carefully plan and establish further industries and developments. As we see the areas with high temperature changes (like Kazakhstan, Brazil etc.), the government should immediately regularize agricultural activities. Also the development should be stopped or controlled or moved from areas like Shanghai(China, one of the most populated cities).

Business value for Industries:

Knowing about climate variations and its trend is not only important for the common public, governments but also has an impact on various businesses. On the one hand, it creates a series of new business risks. Besides the most obvious physical risks (for example, the operational impacts of extreme weather events, or supply shortages caused by water scarcity), companies are exposed to transition risks which arise from society's response to climate change, such as changes in technologies, markets and regulation that can increase business costs, undermine the viability of existing products or services, or affect asset values. Another climate-related risk for companies is the potential liability for emitting greenhouse gases (GHG). An increasing number of legal cases have been brought directly against fossil fuel companies and utilities in recent years, holding them accountable for the damaging effects of climate change.

In Insurance/banking Sector: Rising sea levels and the potential for increased incidences of catastrophic flooding will likely drive up both premiums and payouts, putting a strain on the insurance industry. For central banks and other supervisory authority's climate change will pose a risk to financial stability.

For Investors: As of 2018 we see there is already more than US$30 trillion in funds held in sustainable or green investments in the five major markets tracked by the Global Sustainable Investment Alliance, a rise of 34 per cent in just two years.

In Agricultural Industry: Though warmer temperatures can help crops grow more quickly, for many crops–like grains–the faster the growth, the less time seeds must mature, reducing worldwide yields. Also, the regulations on deforestation will affect the industry.

Energy: For some industries, the risks posed by global warming are mostly about governments' effort to slow it. Regulations on fossil fuels are likely to increase, threatening the lucrative oil, gas, and coal industries.

Tourism: In the long run, however, if winters grow shorter and warmer and the precipitation turns to rain, regions that depend on cold-weather tourism will lose a substantial amount of money.Because more precipitation is falling with the rising temperatures, there will be more snow during the colder months.

The companies will have an impact on their strategy, metrics-targets, governance, and risk management. This study will enable them in transforming how consumers, employees and shareholders evaluate companies and interact with them. In some cases, this can lead to a real shift, where business models need to be reevaluated. Companies do not only need to measure their exposure to climate-related risks and subsequently manage them, but also incorporate climate change in their strategic plans. Failure to do so will undermine the sustainability of their business. A thorough understanding of climate risks is rare and we see that the companies' climate responses focus primarily on measures that have a short-term cost-saving effect.This study will help companies to lay out a plan which will not only focus on short term savings but have long term and permanent changes.

Companies can aim to improve their resource productivity (for example by increasing energy efficiency), thereby reducing their costs. Secondly, it will help spur innovation, inspiring new products, and services. Thirdly, companies can enhance the resilience of their supply chains, for example by reducing reliance on price-volatile fossil fuels by shifting towards renewable energy. Together, these actions can foster competitiveness and unlock new market opportunities. Investors, majorly from energy-utility-mining, sector, tourism travel and transport and logistics too, will benefit from these studies. It will help them make the climate more central to their activities.

19.

20. Appendix:

```
ARIMA(0, 0, 0)x(0, 0, 0, 12)12 - AIC:10499.676088614366
ARIMA(0, 0, 0)x(0, 0, 1, 12)12 - AIC:9040.219708247576
ARIMA(0, 0, 0)x(0, 1, 0, 12)12 - AIC:5870.243867099376
ARIMA(0, 0, 0)x(0, 1, 1, 12)12 - AIC:4997.726912291866
ARIMA(0, 0, 0)x(1, 0, 0, 12)12 - AIC:5865.952590193632
ARIMA(0, 0, 0)x(1, 0, 1, 12)12 - AIC:5047.851267438897
ARIMA(0, 0, 0)x(1, 1, 0, 12)12 - AIC:5457.3060845589625
ARIMA(0, 0, 0)x(1, 1, 1, 12)12 - AIC:4999.597996257966
ARIMA(0, 0, 1)x(0, 0, 0, 12)12 - AIC:9129.399358289784
ARIMA(0, 0, 1)x(0, 0, 1, 12)12 - AIC:8096.682863070602
ARIMA(0, 0, 1)x(0, 1, 0, 12)12 - AIC:5824.66511270374
ARIMA(0, 0, 1)x(0, 1, 1, 12)12 - AIC:4954.431900057392
ARIMA(0, 0, 1)x(1, 0, 0, 12)12 - AIC:5874.0901607519
ARIMA(0, 0, 1)x(1, 0, 1, 12)12 - AIC:5065.224673610315
ARIMA(0, 0, 1)x(1, 1, 0, 12)12 - AIC:5425.749703233335
ARIMA(0, 0, 1)x(1, 1, 1, 12)12 - AIC:4956.369230312844
ARIMA(0, 1, 0)x(0, 0, 0, 12)12 - AIC:7962.396618844026
ARIMA(0, 1, 0)x(0, 0, 1, 12)12 - AIC:7188.643052512962
ARIMA(0, 1, 0)x(0, 1, 0, 12)12 - AIC:6484.912591332938
ARIMA(0, 1, 0)x(0, 1, 1, 12)12 - AIC:5596.472033263604
ARIMA(0, 1, 0)x(1, 0, 0, 12)12 - AIC:6377.653183966428
ARIMA(0, 1, 0)x(1, 0, 1, 12)12 - AIC:5653.255116775461
ARIMA(0, 1, 0)x(1, 1, 0, 12)12 - AIC:6103.8526143574345
ARIMA(0, 1, 0)x(1, 1, 1, 12)12 - AIC:5598.471322889072
ARIMA(0, 1, 1)x(0, 0, 0, 12)12 - AIC:7554.609634274579
ARIMA(0, 1, 1)x(0, 0, 1, 12)12 - AIC:7081.800545416844
ARIMA(0, 1, 1)x(0, 1, 0, 12)12 - AIC:5868.685152525521
ARIMA(0, 1, 1)x(0, 1, 1, 12)12 - AIC:4970.037224586951
ARIMA(0, 1, 1)x(1, 0, 0, 12)12 - AIC:5858.937842355577
ARIMA(0, 1, 1)x(1, 0, 1, 12)12 - AIC:5024.736930417168
ARIMA(0, 1, 1)x(1, 1, 0, 12)12 - AIC:5463.101271876704
ARIMA(0, 1, 1)x(1, 1, 1, 12)12 - AIC:4977.165283093665
ARIMA(1, 0, 0)x(0, 0, 0, 12)12 - AIC:7922.786119038765
ARIMA(1, 0, 0)x(0, 0, 1, 12)12 - AIC:7129.949076291217
ARIMA(1, 0, 0)x(0, 1, 0, 12)12 - AIC:5823.53821008452
ARIMA(1, 0, 0)x(0, 1, 1, 12)12 - AIC:4948.93938013066
ARIMA(1, 0, 0)x(1, 0, 0, 12)12 - AIC:5811.423298526352
ARIMA(1, 0, 0)x(1, 0, 1, 12)12 - AIC:4999.5954521082185
ARIMA(1, 0, 0)x(1, 1, 0, 12)12 - AIC:5418.335030651312
ARIMA(1, 0, 0)x(1, 1, 1, 12)12 - AIC:4950.933774566156
ARIMA(1, 0, 1)x(0, 0, 0, 12)12 - AIC:7492.761165141585
ARIMA(1, 0, 1)x(0, 0, 1, 12)12 - AIC:7007.156426991347
ARIMA(1, 0, 1)x(0, 1, 0, 12)12 - AIC:5813.983450793861
ARIMA(1, 0, 1)x(0, 1, 1, 12)12 - AIC:4939.394315707705
ARIMA(1, 0, 1)x(1, 0, 0, 12)12 - AIC:5805.343462104465
ARIMA(1, 0, 1)x(1, 0, 1, 12)12 - AIC:4991.20930277140‌2
ARIMA(1, 0, 1)x(1, 1, 0, 12)12 - AIC:5417.67333803831
ARIMA(1, 0, 1)x(1, 1, 1, 12)12 - AIC:4941.347487353489
ARIMA(1, 1, 0)x(0, 0, 0, 12)12 - AIC:7301.27169550003
ARIMA(1, 1, 0)x(0, 0, 1, 12)12 - AIC:7006.428081442113
ARIMA(1, 1, 0)x(0, 1, 0, 12)12 - AIC:6208.737386311415
ARIMA(1, 1, 0)x(0, 1, 1, 12)12 - AIC:5321.355488776368
ARIMA(1, 1, 0)x(1, 0, 0, 12)12 - AIC:6157.956037294984
ARIMA(1, 1, 0)x(1, 0, 1, 12)12 - AIC:5375.301117677716
ARIMA(1, 1, 0)x(1, 1, 0, 12)12 - AIC:5829.099758162742
ARIMA(1, 1, 0)x(1, 1, 1, 12)12 - AIC:5323.353994625581
ARIMA(1, 1, 1)x(0, 0, 0, 12)12 - AIC:7298.121137218792
ARIMA(1, 1, 1)x(0, 0, 1, 12)12 - AIC:6983.40191210414
ARIMA(1, 1, 1)x(0, 1, 0, 12)12 - AIC:5821.245791733485
ARIMA(1, 1, 1)x(0, 1, 1, 12)12 - AIC:4933.10428583162
ARIMA(1, 1, 1)x(1, 0, 0, 12)12 - AIC:5800.5342038197405
ARIMA(1, 1, 1)x(1, 0, 1, 12)12 - AIC:4987.530089145281
ARIMA(1, 1, 1)x(1, 1, 0, 12)12 - AIC:5423.56695209944
ARIMA(1, 1, 1)x(1, 1, 1, 12)12 - AIC:4933.72776297273
```

Fig: M.1

```python
for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod = sm.tsa.statespace.SARIMAX(train['Temp'],order=param,seasonal_order=param_seasonal,enforce_stationarity=Fals
            results = mod.fit()
            print('ARIMA{}x{}12 - AIC:{}'.format(param,param_seasonal,results.aic))
        except:
            continue
```

```
ARIMA(0, 0, 0)x(0, 0, 0, 12)12 - AIC:20049.92717250891
ARIMA(0, 0, 0)x(0, 0, 1, 12)12 - AIC:17515.84974486111
ARIMA(0, 0, 0)x(0, 1, 0, 12)12 - AIC:13163.2500564536
ARIMA(0, 0, 0)x(0, 1, 1, 12)12 - AIC:12297.071361505401
ARIMA(0, 0, 0)x(1, 0, 0, 12)12 - AIC:13114.952519470324
ARIMA(0, 0, 0)x(1, 0, 1, 12)12 - AIC:12342.763843702793
ARIMA(0, 0, 0)x(1, 1, 0, 12)12 - AIC:12555.396911161226
ARIMA(0, 0, 0)x(1, 1, 1, 12)12 - AIC:12298.157044918102
ARIMA(0, 0, 1)x(0, 0, 0, 12)12 - AIC:17398.084127970356
ARIMA(0, 0, 1)x(0, 0, 1, 12)12 - AIC:15786.913021616907
ARIMA(0, 0, 1)x(0, 1, 0, 12)12 - AIC:12911.666293596261
ARIMA(0, 0, 1)x(0, 1, 1, 12)12 - AIC:11887.39492006458
ARIMA(0, 0, 1)x(1, 0, 0, 12)12 - AIC:12842.724561806961
ARIMA(0, 0, 1)x(1, 0, 1, 12)12 - AIC:11932.684637370417
ARIMA(0, 0, 1)x(1, 1, 0, 12)12 - AIC:12253.955351298298
ARIMA(0, 0, 1)x(1, 1, 1, 12)12 - AIC:11889.139719804776
ARIMA(0, 1, 0)x(0, 0, 0, 12)12 - AIC:15321.3670129935
ARIMA(0, 1, 0)x(0, 0, 1, 12)12 - AIC:14373.242175045863
ARIMA(0, 1, 0)x(0, 1, 0, 12)12 - AIC:14007.889825005255
ARIMA(0, 1, 0)x(0, 1, 1, 12)12 - AIC:12586.17767343687
ARIMA(0, 1, 0)x(1, 0, 0, 12)12 - AIC:13564.826643819633
ARIMA(0, 1, 0)x(1, 0, 1, 12)12 - AIC:12639.410860060612
ARIMA(0, 1, 0)x(1, 1, 0, 12)12 - AIC:13245.361617451112
```

```
ARIMA(1, 0, 0)x(0, 1, 0, 12)12 - AIC:12893.8055648447
ARIMA(1, 0, 0)x(0, 1, 1, 12)12 - AIC:11791.456132594025
ARIMA(1, 0, 0)x(1, 0, 0, 12)12 - AIC:12800.024468444677
ARIMA(1, 0, 0)x(1, 0, 1, 12)12 - AIC:11837.498572405037
ARIMA(1, 0, 0)x(1, 1, 0, 12)12 - AIC:12210.204236626007
ARIMA(1, 0, 0)x(1, 1, 1, 12)12 - AIC:11794.362360517458
ARIMA(1, 0, 1)x(0, 0, 0, 12)12 - AIC:14676.307964840897
ARIMA(1, 0, 1)x(0, 0, 1, 12)12 - AIC:14069.87464750081
ARIMA(1, 0, 1)x(0, 1, 0, 12)12 - AIC:12891.84535073262
ARIMA(1, 0, 1)x(0, 1, 1, 12)12 - AIC:11788.450949242102
ARIMA(1, 0, 1)x(1, 0, 0, 12)12 - AIC:12801.161519069332
ARIMA(1, 0, 1)x(1, 0, 1, 12)12 - AIC:11834.655033391966
ARIMA(1, 0, 1)x(1, 1, 0, 12)12 - AIC:12212.082056591773
ARIMA(1, 0, 1)x(1, 1, 1, 12)12 - AIC:11790.93886267654
ARIMA(1, 1, 0)x(0, 0, 0, 12)12 - AIC:14646.997043867417
ARIMA(1, 1, 0)x(0, 0, 1, 12)12 - AIC:14236.783719722502
ARIMA(1, 1, 0)x(0, 1, 0, 12)12 - AIC:13677.99718129121
ARIMA(1, 1, 0)x(0, 1, 1, 12)12 - AIC:12389.502566603118
ARIMA(1, 1, 0)x(1, 0, 0, 12)12 - AIC:13404.760948284149
ARIMA(1, 1, 0)x(1, 0, 1, 12)12 - AIC:12441.339178961398
ARIMA(1, 1, 0)x(1, 1, 0, 12)12 - AIC:12939.660863727511
ARIMA(1, 1, 0)x(1, 1, 1, 12)12 - AIC:12393.009884338606
ARIMA(1, 1, 1)x(0, 0, 0, 12)12 - AIC:14636.826417834072
ARIMA(1, 1, 1)x(0, 0, 1, 12)12 - AIC:14216.42385010931
ARIMA(1, 1, 1)x(0, 1, 0, 12)12 - AIC:12894.50327571739
```

Fig: C.1

```
test_baseline = test['Temp'].shift()
test_baseline[0] = test['Temp'][0]

final =test_baseline.drop(test_baseline.index[11])
final.head(13)

plate= test['Temp'].dropna()
plate.head(120)
rmse_test_base = measure_rmse(plate,final)


print(f'The baseline RMSE for the test baseline was {round(rmse_test_base,2)} celsius degrees')
Pred =test['Pred'].drop(test['Pred'].index[11])
Pred.head(13)
rmse_test_extrap = measure_rmse(plate, Pred)
print(f'The baseline RMSE for the test extrapolation was {round(rmse_test_extrap,2)} celsius degrees')
print(f'That is an improvement of {-round((rmse_test_extrap/rmse_test_base-1)*100,2)}%')
```

```
The baseline RMSE for the test baseline was 2.64 celsius degrees
The baseline RMSE for the test extrapolation was 2.12 celsius degrees
That is an improvement of 19.58%
```

Fig C.2

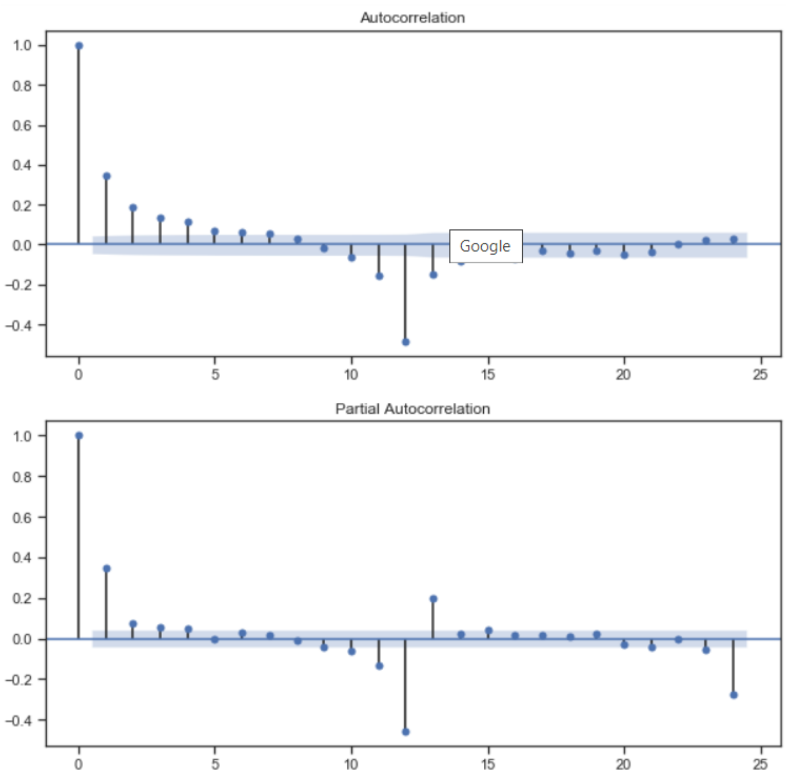
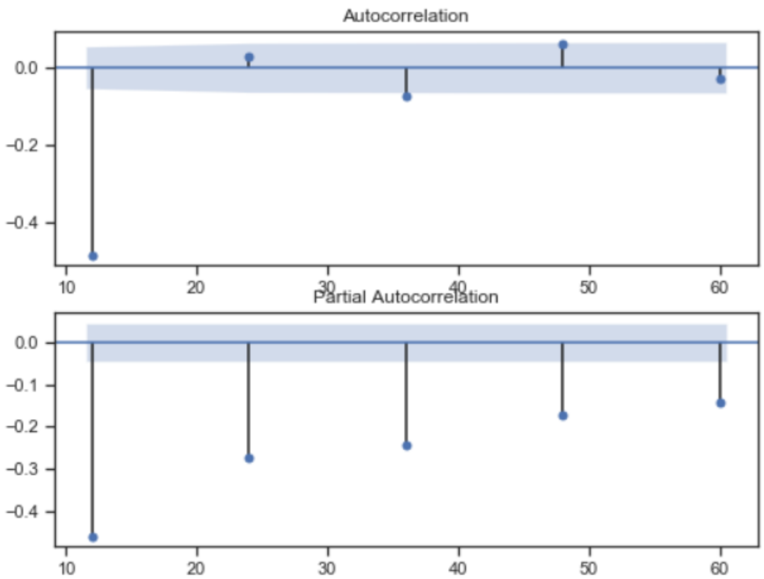


Fig: GT.1


Autocorrelation
0.0
-0.2
-0.4
10
20
30
40
50
60
Partial Autocorrelation
0.0
-0.1
-0.2
-0.3
-0.4
10
20
30
40
50
60

Fog GT.2

21. REFERENCES

1) http://cait.wri.org/
2) https://climate.nasa.gov/
3) http://berkeleyearth.org/land-and-ocean-data/
4) https://data.worldbank.org/indicator/AG.LND.FRST.ZS
5) https://www.worldometers.info/world-population/world-population-by-year/
6) https://www.worldometers.info/co2-emissions/co2-emissions-by-year/
7) https://data.worldbank.org/indicator/EG.USE.COMM.FO.Z
8) Houghton, J.T.; Ding, Y.; Griggs, D.J.; Noguer, M.; van der Linden, P.J.; Xiaosu, D., eds. 2001. Climate changes 2001: the scientific basis. Cambridge, United Kingdom: Cambridge University Press
9) Intergovernmental Panel on Climate Change [IPCC]. 2001. Chapter 5: Ecosystems and their goods and services/ 5.6.2.2.2. Pressures from diseases and insect herbivory.
10) Intergovernmental Panel on Climate Change [IPCC]. 2007. Summary for policymakers. In: Solomon, S.; Qin, D.; Manning, M.; Chen, Z.; Marquis, M.; Averyt, K.B.; Tignor, M.; Miller, H.L., eds. Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York: Cambridge University Press.
11) Mann, M.E.; Bradley, R.S.; Hughes, M.K. 1999. Northern hemisphere temperatures during the past millennium: inferences, uncertainties, and limitations. Geophysical Research Letters. 26(6): 759–762.
12) https://www.cnbc.com/2014/10/22/7-industries-at-greatest-risk-from-climate-change.html
13) https://www.nrdc.org/experts/rob-moore/climate-change-literature-made-waves-2019
14) https://www2.deloitte.com/us/en/insights/topics/strategy/impact-and-opportunities-of-climate-change-on-business.html