Climate Change: Earth Surface Temperature Forecasting of Canada Using Time Series Modelling

SYDE631- Time Series Modelling

Course Instructor - Professor Keith W. Hipel

Shubham Koundinya
Student Id -20718539.
Masters in Engineering ,
Electrical and Computer Engineering .

Abstract:

Climate Change is one of the greatest ever threat humanity has faced . The average temperature of the Earth's surface increased by an estimated 0.7°C since the beginning of the 20th century and, according to the most recent projections of the Intergovernmental Panel on Climate Change, could rise by 1.6–4.3°C compared to a,1850–1900 baseline by 2100[6]. With this aim, the project uses Time Series and Statistics Modelling as a tool to make future forecasts on Earth Surface temperature data specific to Canada. The main contributions of the project are two fold – forecasting Yearly Average Temperature of Canada for next ten years and forecasting monthly temperature of Canada for the next ten years. These forecasts/predictions can help in sound decision making and raise general awareness about the Climate Change issue, especially increase in Earth Surface Temperature. The sources of both average yearly temperature data for Canada from 1820-1912, and the monthly data for Canada , from Jan 1820 to Dec2013 is Berkley Earth [4].

1. Introduction:

1.1 Motivation and Importance:

Climate change and Increase in Temperature are one of the major challenges of our time and adds considerable stress to our societies and to the environment. From shifting weather patterns that threaten food production, to rising sea levels that increase the risk of catastrophic flooding, the impacts of climate change are global in scope and unprecedented in scale. Without drastic action today, adapting to these impacts in the future will be more difficult and costlier [1].

The Intergovernmental Panel on Climate Change's major report on the science of climate change, published on 2 February 2007, predicts that temperatures will rise by 2°C and 4.5°C by 2100, relative to pre-industrial temperatures [2]. Below are estimates of the impacts for each degree of temperature rise. They are based on a major report published by Nicholas Stern, chief British government economist in October 2006. [2].

The changes of even 1 °C, 2 °C, have a huge impact. Below are some of the issues highlighted by the published report on the slightest of the changes and their impacts

According to the report[2], 1° C increase can cause the following:

- Small Andean glaciers disappear, threatening water supplies for 50 million people
- Cereal yields in temperate regions increase slightly
- At least 30,000 people die every year from climate-related diseases, but winter mortality in Northern Europe and US drops
- 80% of coral reefs are bleached, including the Great Barrier Reef.
- The Atlantic thermohaline circulation starts to weaken

According to the report[2], 2°C increase can cause the following:

- Water availability in some vulnerable regions (Southern Africa and Mediterranean) could drop by 20% – 30%.
- Crop yields in Africa drop by 5% – 10%
- 40 – 60 million more people are exposed to malaria in Africa
- Up to 10 million more people are affected by coastal flooding
- Arctic species, including the polar bear and caribou, run a high risk of extinction
- The Greenland ice sheet could begin an irreversible melt

According to the report [2], a 3°C increase can cause the following:

- In southern Europe, serious drought happens occurs every 10 years.
- 1 to 4 billion more people suffer water shortages; up to 5 billion gain water but they could suffer increased floods40 – 60 million more people are exposed to malaria in Africa

- Another 150 to 500 million people are at risk of hunger (if carbon fertilization is weak).
- 1 to 3 million more people die from malnutrition.
- The risk of abrupt changes in monsoons climbs
- There is a higher risk that the West Antarctic ice sheet, and the Atlantic thermohaline circulation, will collapse

Keep these issue of increase in Earth Surface Temperature data , the project aims to use Time Series Modelling as a tool to study the increase in surface temperature data of Canada . After the modelling future temperature forecasts will be done. Such forecasts can help authorities to take necessary actions to fight against the issue of Climate Change and help in sound decision making. Two main aims of the project include :

1. Analyze the average yearly temperature of Canada from 1820-1912 and forecast the average yearly Temperature values for the next 10 years.
2. Analyze the monthly Canadian Temperature from Jan 1820 to Dec2013 and forecast the monthly weather for Canadian Temperature for the next 10 years. More details about the goals are covered in section 2 background.

<u>1.2 Data Sources:</u>

The source for the data is: Kaggle.com – a Data Science Competition Website , which further attributes the data source as http://berkeleyearth.org/ - which is the true source of data for this project. Berkeley Earth was conceived by Richard and Elizabeth Muller in early 2010 [4]. The organization have a group of scientists to analyze the Earth's surface temperature record. Below Fig 1.2 on left side, shows the screenshot of Average Yearly Data to forecast future climate and on right shows the screenshot of monthly data to forecast the monthly data of Canada for the next 10 years. Data specific to Canada was extracted from the whole dataset. Further details about data are present in the appendix section.

Years	Average Yearly Temperature
1820	-6.171666667
1821	-5.560333333
1822	-5.820166667
1823	-6.323416667
1824	-5.171583333
1825	-5.1515
1826	-5.1655
1827	-4.64575
1828	-5.096833333
1829	-5.794416667
1830	-4.868833333
1831	-5.775
1832	-5.927833333
1833	-5.248333333
1834	-5.507416667
1835	-6.423333333
1836	-6.068083333
1837	-6.02775
1838	-6.23225
1839	-5.261916667
1840	-5.652416667
1841	-5.715
1842	-5.408916667

Months	Averagetemperature
1820-01-01	5.013
1820-02-01	3.825
1820-03-01	3.347
1820-04-01	3.079
1820-05-01	3.149
1820-06-01	2.791
1820-07-01	3.062
1820-08-01	2.951
1820-09-01	3.009
1820-10-01	3.061
1820-11-01	3.365
1820-12-01	3.563
1821-01-01	4.166
1821-02-01	4.379
1821-03-01	3.34
1821-04-01	3.091
1821-05-01	3.225
1821-06-01	3.342
1821-07-01	2.973
1821-08-01	2.98
1821-09-01	2.788
1821-10-01	2.986
1821-11-01	2.735
1821-12-01	2.954
1822-01-01	3.765

Fig 1.2 on the left shows the sample Average yearly Data that will be analyzed and used for making future predictions. Figure on the right shows, the Average Monthly Data , that will be analyzed and used for making future predictions.

1.3 Types of Models:

Since the project deals with two types of data :

1. Average Yearly Temperature data
2. Monthly seasonal Temperature data,

Hence the project considers ARIMA class of models, for forecasting the future values. After the Exploratory Data Analysis (mentioned in details in the upcoming sections), it was found that the two data sources have trends/ seasonality, hence two class of Models were considered – ARIMA and Seasonal ARIMA Models.

1.4 Overview of the Upcoming Sections:

Section 2 covers more background to the Project and highlights the importance of the specific data analyzed. Section 3 contains the Exploratory Data Analysis of both the monthly and the average yearly data. Section 4 follows the principle of three stages of model construction to fit the models to both the monthly and the average yearly data. Section 5 covers the Applications and Forecasts from both the monthly and the yearly data. Finally, Section 6 covers Conclusions followed by the section on References. Section 7 contains the Appendices and finally Section 8 contains the references.

All the simulations in the project have been done using the statistical Software – R Programming Language [8]

2. Background:

As seen from the section 1.1, increase in temperature is a grave concern globally. The project aims to study the Temperature Surface data of Canada in two ways - Average Yearly Temperature Data Across Canada. 2. Monthly data across Canada.

Few of the reports specific to Canada include - A report from Canada.ca [5], - Warming trends are seen consistently across Canada. Future warming will be accompanied by other changes, including the amount and distribution of rain, snow, and ice and the risk of extreme weather events such as heat waves, heavy rainfalls and related flooding, dry spells and/or droughts, and forest fires. In addition, Canada is a maritime nation with 8 of its 10 provinces and all three territories bordering on ocean waters (including Hudson Bay). Thus, many regions of Canada will also be affected by changing ocean environments, including changes in average and extreme sea level, wave regimes, and ice conditions. According to another report from Canada.ca, Five of the ten warmest years occurred within the last decade. From 1948 to 2014, a trend was detected in annual average temperature departures [7].

The project uses the earth surface temperature data specific to Canada from the Berkley Earth [4], with the aim to study trends and analyze the temperature. This study can help in sound decision making to fight against the issue of Climate Change/ specifically increase in Earth Surface Temperature data in Canada.

Thus the project has two main goals :

1. Analyze the average yearly temperature of Canada from 1820-1912 and forecast the average yearly Temperature values for the next 10 years.
2. Analyze the monthly Canadian Temperature from Jan 1820 to Dec2013 and forecast the monthly weather for Canadian Temperature for the next 10 years.

Both the yearly average and monthly forecasts will be presented. The average yearly data of Canada from 1820-1912 , and monthly Canadian Temperature from Jan 1820 to Dec2013, were obtained from Berkley Earth [4]. These are two separate datasets (Average Yearly) and (Monthly), hence two separate analysis is done.

3. Exploratory Data Analysis (EDA):

Since the two datasets have been analyzed (yearly - data from 1820-1912, and monthly data from Jan 1820 to Dec2013) hence two separate Exploratory Data Analysis have been conducted. Below section 3.1, contains the Exploratory Data Analysis for Yearly Average Temperature Data , and Section 3.2 contains EDA for monthly data.

3.1 <u>Exploratory Data Analysis for Average Yearly Data from 1820 -1912:</u>

This section includes the EDA for Average Yearly Canadian Average Temperature from 1820 – 1912. X axis shows the years and y axis shows the temperature. The figure 3.1 shows slight upward trend, The fig 3.2 Below captures the trend line, which is given by a red line. We can observe from figures 3.1 and 3.2 that there is a slight upward trend.

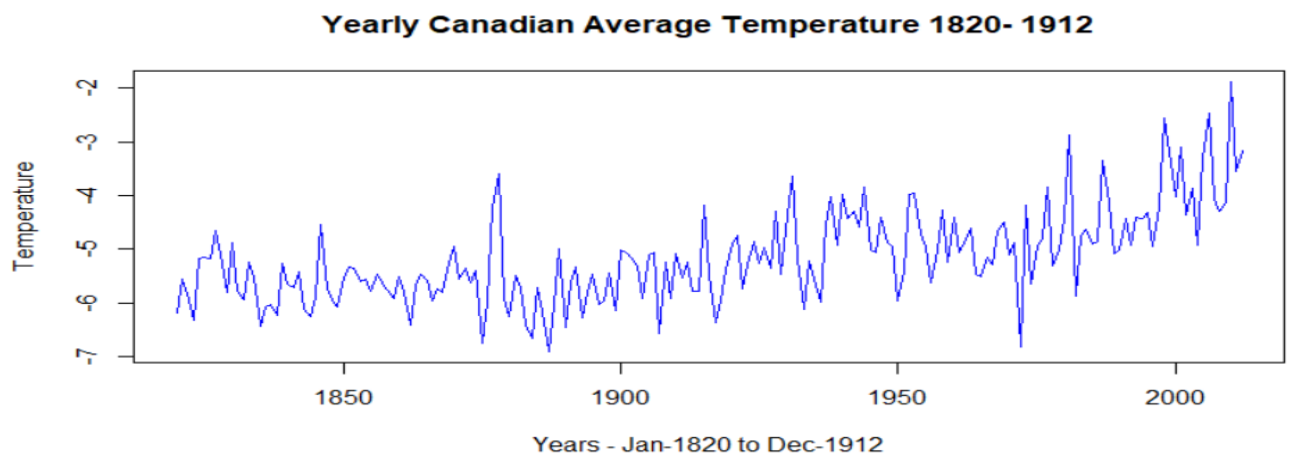

Fig 3.1.1 Yearly Canadian Average Temperature from 1820 – 1912

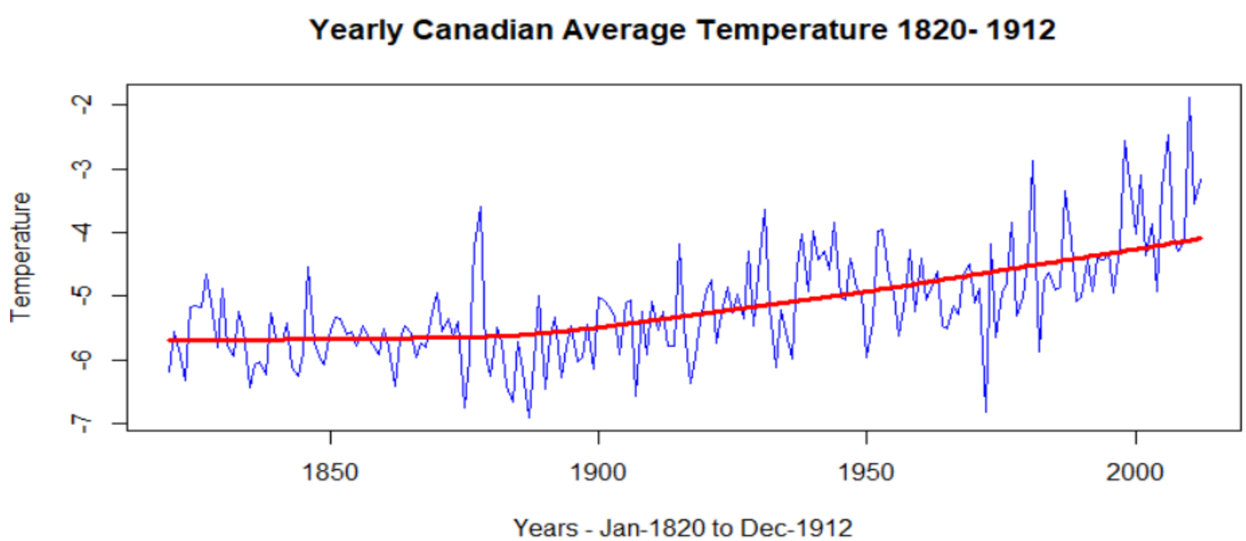

Fig 3.1.2 Yearly Canadian Average Temperature from 1820 – 1912, with trend line

ACF Plot for Yearly Canadian Average Temperature from 1820 – 1912:

ACF was also plotted for Yearly Canadian Average Temperature from 1820 – 1912 and it was seen to be dying off, hence this also confirms that the trend is present.

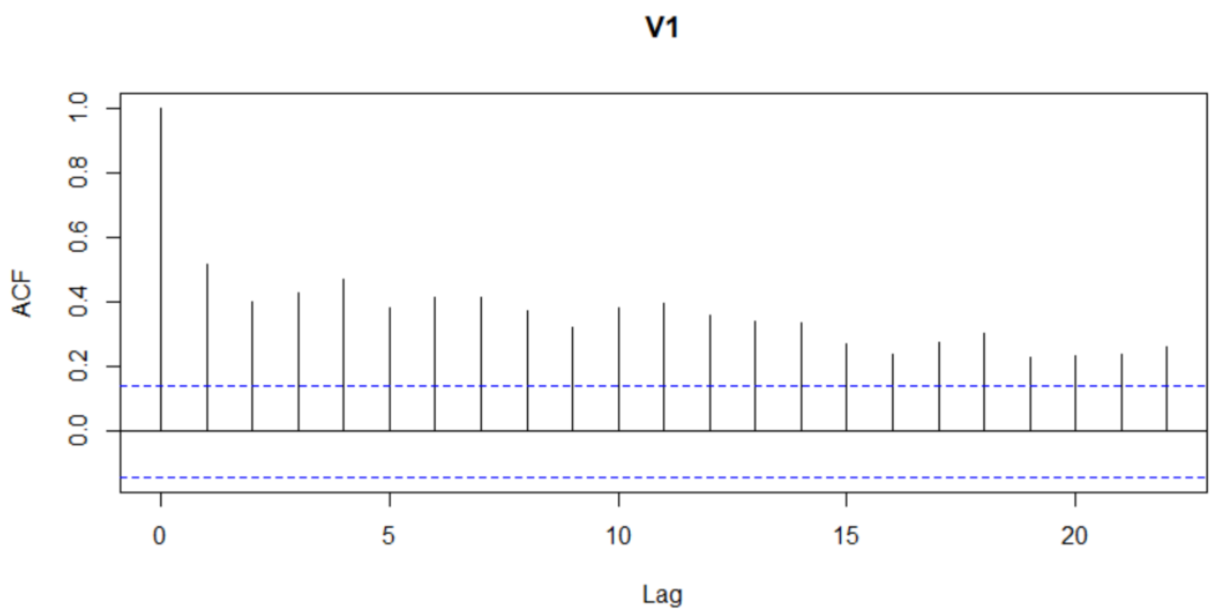

Fig 3.1.3 ACF for Yearly Canadian Average Temperature from 1820 – 1912

Augmented Dickey- Fuller Test On Yearly Canadian Average Temperature from 1820 – 1912:

To carry out further stationarity test, Augmented Dickey-Fuller Test was conducted in R statistical software, which further confirmed that the above (Average Temperature Yearly) series is stationary. Below are the results from the Augmented Dickey Fuller Test, conducted in R software. The details about Augmented Dickey Fuller test is mentioned in the Appendix.

Augmented Dickey-Fuller Test
data: myts_yearly
Dickey-Fuller = -3.3357, p-value = 0.067

As seen from the above plots fig 3.1 ,3.1.2 and the ACF plot 3.1.3 and the Augmented Dickey Fuller Test result, the series is non stationary with a upward trend.

Since upward trend is present in the above series a differencing of one was done. Augmented Dickey Fuller Test was performed after differencing the data set by one. The results are below and this confirms that the series after differencing is stationary.

Augmented Dickey- Fuller Test On Yearly Canadian Average Temperature from 1820 – 1912 after differencing of one:

Below are the results of the Augmented Dickey Fuller test after differencing of one was done on the yearly time series data.

Augmented Dickey-Fuller Test
data: diff(myts_yearly)
Dickey-Fuller = -9.6918, p-value = 0.01

From the EDA carried out in section 3.1 on the monthly on yearly Canadian average temperature from 1820 – 1912 , it can be confirmed that the original series has upwards trend, and differencing by one, removes that trend. Further Confirmatory Data Analysis on this Series is covered in section 4.

3.2 Exploratory Data Analysis for monthly Seasonal Data for Canada Jan 1820 to Sept 2013 :

This section includes the Exploratory Data Analysis done on Monthly Average Temperature for Canada. The below figure shows the data is seasonal and the red line shows a slight upward trend, in the figure. From the figures 3.2.1 and 3.2.2 we, can see a slight upward trend.

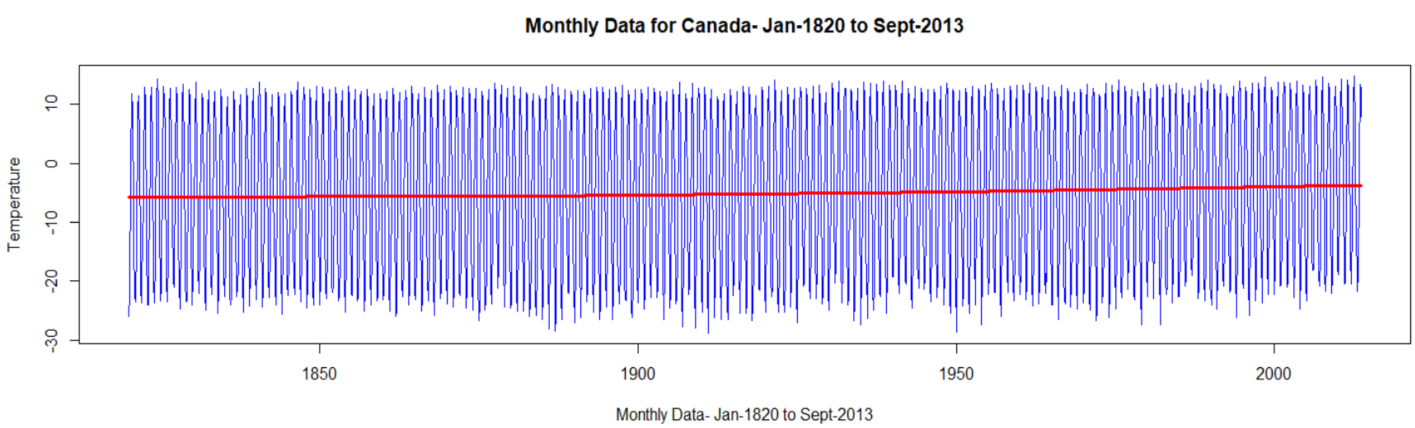

Fig 3.2.1 Monthly Canadian Average Temperature from Jan 1820 to Sept 2013

Since the above data, is little congested for viewing , figure 3.2.2 on the next page, shows another view of the monthly data.

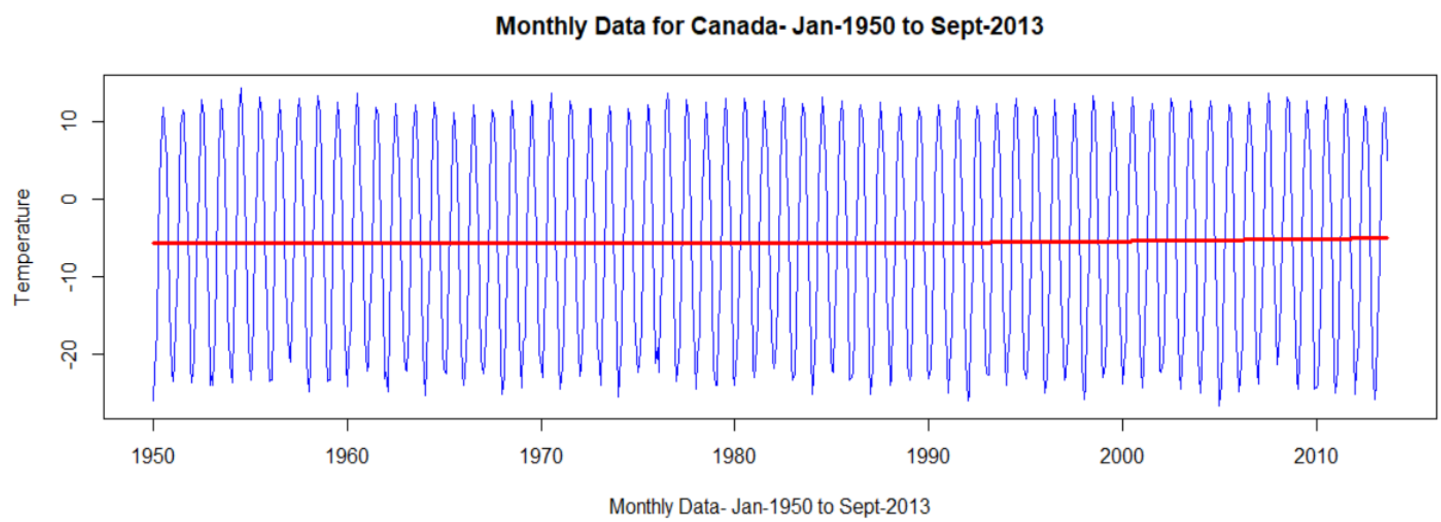

Fig 3.2.2 Monthly Canadian Average Temperature from Jan 1950 to Sept 2013

The monthly average data was further decomposed as shown by the additive decomposition formula below using the R statistical software.

Monthly Time series Data = Seasonal + Trend + Random

The figure 3.2.3, shows the decomposition of the trend and seasonality for monthly data for temperature of Canada.

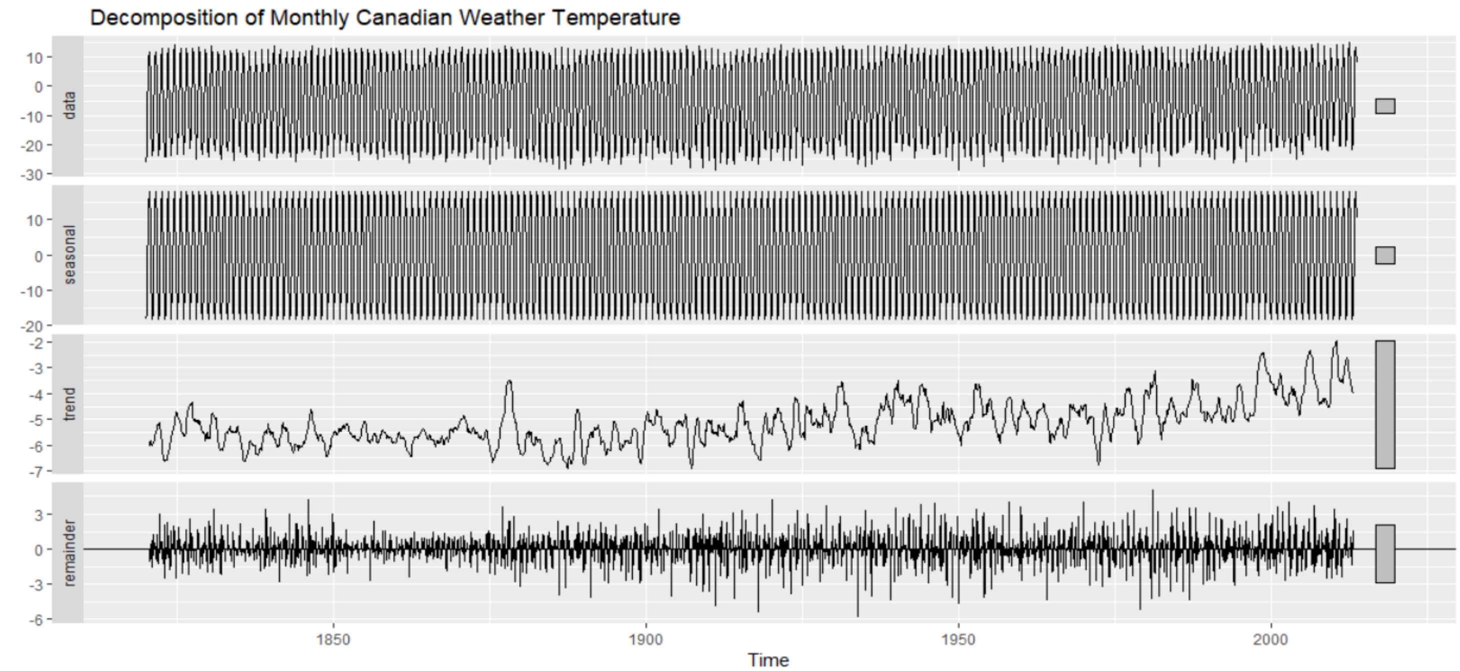

Fig 3.2.3 Decomposition of the Monthly Temperature Data for Canada from Jan 1820 to Sept 2013.

Below figure 3.2.4, displays the PACF of Monthly Temperature Data :

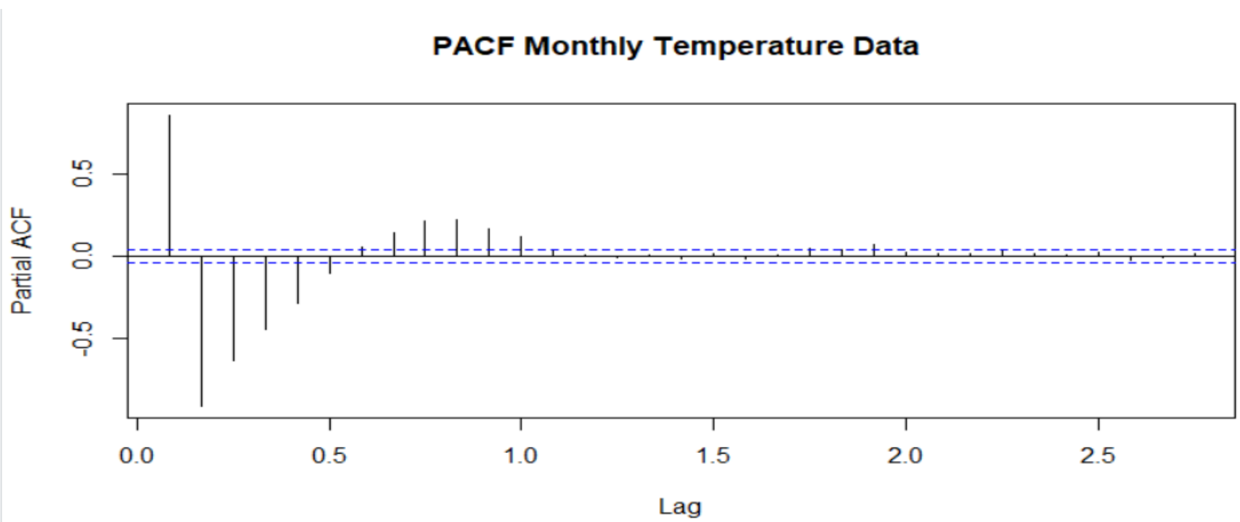

Fig 3.2.4 Fig. displaying PACF of Monthly Temperature Data for Canada from Jan 1820 to Sept 2013

4. <u>Confirmatory Data Analysis:</u>

As with the previous sections, Confirmatory Data Analysis was done on two different data sets. (yearly - data from 1820-1912, and monthly data from Jan 1820 to Dec2013) hence two separate Confirmatory Data Analysis have been conducted. Three stages of model construction as highlighted in the below diagram were considered for both the two data sets.

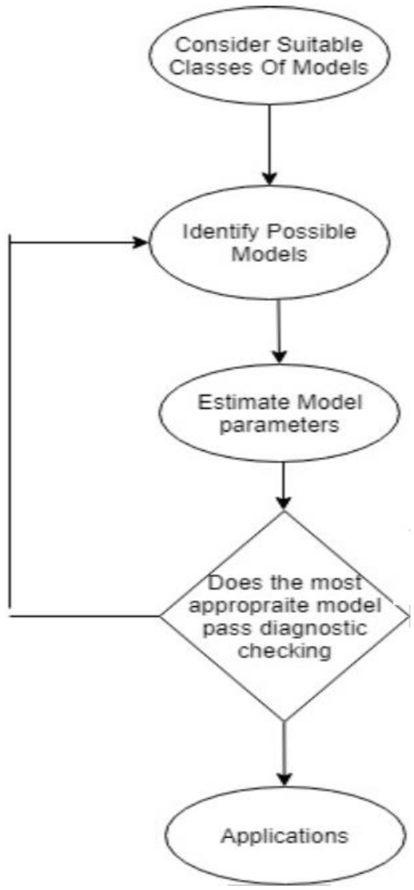

Fig. 4.1 Principles of three stages for Model Construction. Same principles were used for the Model construction for both the data sets.

4.1 Confirmatory Data Analysis for Average Yearly Data from 1820 -1912:

As show in the section 3 (EDA) - a differencing of one was needed for the yearly temperature series . Further the series becomes stationary after differencing of one, and since this is non-seasonal data, autoregressive integrated moving average (ARIMA) models were used. Non-seasonal ARIMA models are generally denoted ARIMA(p,d,q) where parameters p, d, and q are non-negative integers, p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (the number of times the data have had past values subtracted), and q is the order of the moving-average model. Below models were calibrated and the corresponding AIC values are highlighted below. The best model with the lowest AIC value was found to be ARIMA(0,1,1) with drift.

Calibrated Models	AIC
ARIMA(2,1,2)	: 386.2841
ARIMA(0,1,0)	: 477.6045
ARIMA(1,1,0)	: 449.3658
ARIMA(0,1,1)	: 385.0692
ARIMA(1,1,1)	: 385.9022
ARIMA(0,1,2)	: 385.3918
ARIMA(1,1,2)	: 386.5648

Best model: ARIMA(0,1,1)

To check the diagnostics of the best fitted model, residuals were plotted, which concluded that the best m odel is suitable for the given data.

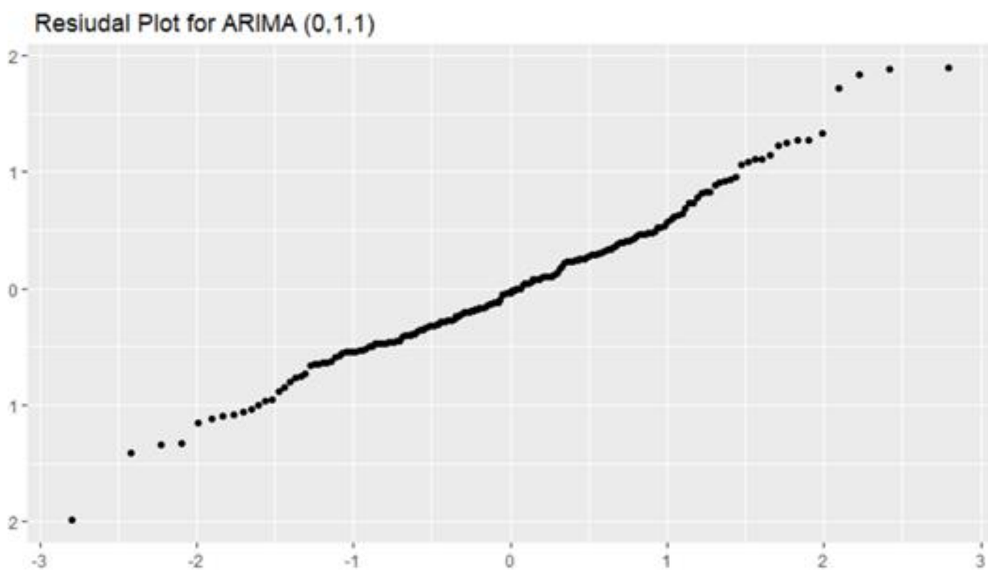

Fig 4.1 Shows the residuals of ARIMA(0,1,1) with drift after fitting the model to Yearly Data

4.2 <u>Confirmatory Data Analysis for monthly Seasonal Data of Canada Jan 1820 to Sept 2013</u>:

As shown in the section 3.2(EDA for monthly data) , monthly data has both trend and seasonality , hence Seasonal ARIMA models were used to model the monthly data .The seasonal SARIMA models are usually denoted ARIMA(p,d,q)(P,D,Q).. p, d, and q are non-negative integers, p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (the number of times the data have had past values subtracted), and q is the order of the moving-average model and the uppercase P,D,Q refer to the autoregressive, differencing, and moving average terms for the seasonal part of the ARIMA model.

Below table shows the range of fitted values and the corresponding the AIC criteria. The best model with the lowest AIC was found to be – ARIMA(1,0,1) (2,1,0), where Inf denotes Infinity.

Calibrated Models	**AIC**
ARIMA(2,0,2)(1,1,1)	: Inf
ARIMA(0,0,0)(0,1,0)	: 9737.035
ARIMA(1,0,0)(1,1,0)	: 9029.071
ARIMA(0,0,1)(0,1,1)	: Inf
ARIMA(0,0,0)(0,1,0)	: 9735.107
ARIMA(1,0,0)(0,1,0)	: 9614.177
ARIMA(1,0,0)(2,1,0)	: 8750.889
ARIMA(1,0,0)(2,1,1)	: Inf
ARIMA(0,0,0)(2,1,0)	: 8880.57
ARIMA(2,0,0)(2,1,0)	: 8730.806
ARIMA(2,0,1)(2,1,0)	: 8723.794
ARIMA(3,0,2)(2,1,0)	: 8727.596
ARIMA(2,0,1)(2,1,0)	: 8722.061
ARIMA(2,0,1)(1,1,0)	: 9003.065
ARIMA(2,0,1)(2,1,1)	: Inf
ARIMA(1,0,1)(2,1,0)	: 8720.859
ARIMA(1,0,0)(2,1,0)	: 8749.233
ARIMA(1,0,2)(2,1,0)	: 8722.161
ARIMA(0,0,0)(2,1,0)	: 8879.111
ARIMA(2,0,2)(2,1,0)	: 8724.107
ARIMA(1,0,1)(2,1,0)	: 8722.625
ARIMA(1,0,1)(1,1,0)	: 9002.597
ARIMA(1,0,1)(2,1,1)	: Inf
ARIMA(0,0,1)(2,1,0)	: 8773.477

Best model: ARIMA(1,0,1)(2,1,0)

The residual plot for the fitted SARIMA model (1,0,1) (2,1,0) is shown in the plot 4.2 confirms the proposed SARIMA (1,0,1) (2,1,0) has a reasonable performance over the other models.

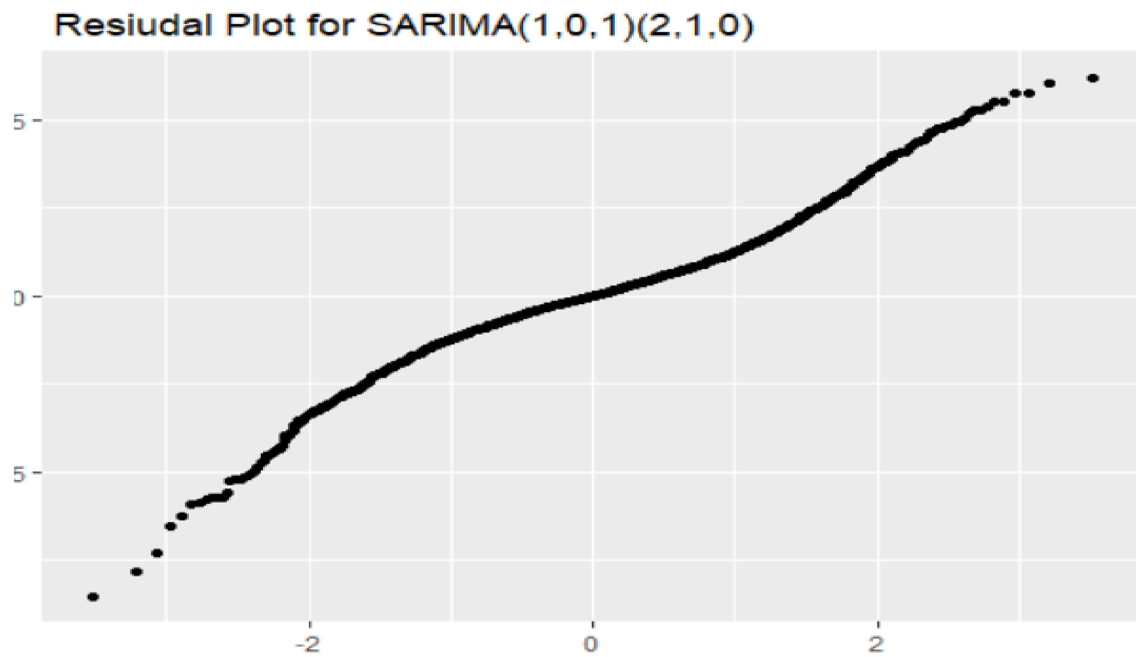

Fig 4.2 Shows the residuals SARIMA (1,0,1) (2,1,0) with drift after fitting the model to Monthly Data

5.**<u>Applications and Insights:</u>**

As with the other sections, the applications from the above two datasets, Forecasts can be done for both yearly and monthly data. Section 5.1 covers the yearly forecasts for average temperature for yearly data , and Section 5.2 consists of monthly forecasts for temperature.

5.1 Average Yearly Forecasts (using average yearly temperature of Canada from 1820-1912) :

The analysis in section 3.1 (EDA) and the best model in section 4.1 (Confirmatory Data Analysis) ,can be used to predict the average yearly temperature data of Canada . Below table , shows the average yearly forecasts from ARIMA(0,1,1) (which was found as the best model for the average temperature series) model. The table shows the point forecast for the future years , 2013- 2022. along with the confidence limits. Fig 5.1 shows the graph of the forecasts made on the yearly average temperature data.

Year	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2013	-3.461968	-4.288315	-2.635621	-4.725757	-2.198180
2014	-3.527341	-4.370822	-2.683861	-4.817334	-2.237349
2015	-3.474090	-4.317574	-2.630606	-4.764087	-2.184092
2016	-3.447959	-4.296258	-2.599661	-4.745320	-2.150598
2017	-3.442217	-4.299402	-2.585032	-4.753168	-2.131266
2018	-3.433490	-4.297779	-2.569201	-4.755306	-2.111674
2019	-3.421365	-4.291963	-2.550766	-4.752830	-2.089899
2020	-3.409495	-4.286550	-2.532441	-4.750834	-2.068156
2021	-3.398177	-4.281766	-2.514588	-4.749509	-2.046844
2022	-3.386853	-4.276906	-2.496800	-4.748071	-2.025634

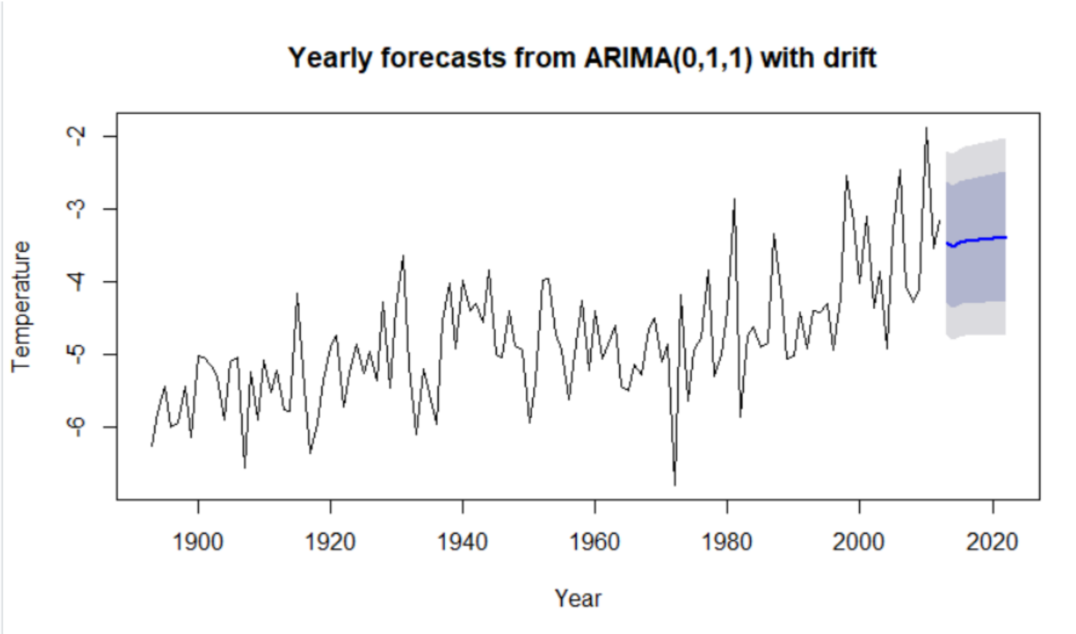

Fig 5.1 shows the average yearly forecast from the ARIMA(0,1,1) model for next 10 years

5.2 Monthly Forecasts (Canada Jan 1820 to Sept 2013):

Monthly Forecasts from the best model found as SARIMA model (1,0,1) (2,1,0), in section 4.2. The forecasts along with the confidence interval and the data are as shown in the below table

Month	**Year**	**Forecast**	**Lo-95**	**Hi-95**
Oct	2013	0.5803719	3.6963724	2.535629
Nov	2013	10.350539	13.528252	-7.172826
Dec	2013	17.131627	20.336742	13.926511
Jan	2014	20.847445	24.064825	17.630065
Feb	2014	18.640782	21.86367	15.417894
Mar	2014	14.471057	17.696422	11.245692
Apr	2014	7.0140406	10.240521	-3.78756
May	2014	3.3231854	0.0962031	6.550168
Jun	2014	10.588525	7.3613167	13.815733
Jul	2014	14.023961	10.796651	17.251271
Aug	2014	12.775756	9.5483999	16.003112
Sep	2014	7.8387641	4.6113876	11.066141
Oct	2014	0.7746997	-4.19702	2.647621
Month	**Year**	**Forecast**	**Lo-95**	**Hi-95**
Nov	2014	10.861659	14.291792	7.431527
Dec	2014	17.289331	20.722976	13.855687
Jan	2015	20.953054	24.388279	17.517829

Feb	2015	-17.98827	21.424206	14.552333
Mar	2015	14.034854	17.471111	10.598597
Apr	2015	6.6710671	10.107468	-3.234666
May	2015	3.4506514	0.0141854	6.887117
Jun	2015	10.793993	7.3574977	14.230488
Jul	2015	14.081224	10.644716	17.517732
Aug	2015	12.824607	9.3880929	16.261122
Sep	2015	7.8866966	4.4501796	11.323214
Oct	2015	1.0703612	-4.758495	2.617773
Nov	2015	11.128513	14.826454	-7.430571
Dec	2015	17.881963	21.584313	14.179614
Jan	2016	21.152882	24.857215	17.448549
Feb	2016	18.490763	22.195988	14.785537
Mar	2016	14.152923	17.858551	10.447296
Apr	2016	7.1734837	10.879293	-3.467675
May	2016	3.2343891	0.4715012	6.940279
Jun	2016	10.734998	7.0290711	14.440925
Jul	2016	13.809796	10.103852	17.515739
Aug	2016	12.744603	9.0386522	16.450554
Sep	2016	7.8844617	4.1785074	11.590416
Oct	2016	0.8159197	-4.973113	3.341274
Nov	2016	10.784706	14.959078	-6.610334
Dec	2016	17.450776	21.632861	13.268692
Jan	2017	20.989662	25.175216	16.804109
Feb	2017	-18.3925	22.579615	14.205385
Mar	2017	14.226024	18.413841	10.038206
Apr	2017	6.9700177	11.158152	-2.781884
May	2017	3.3288265	0.8594501	7.517103
Jun	2017	10.702718	6.5143776	14.891059
Jul	2017	13.96325	9.7748804	18.15162
Aug	2017	12.778962	8.5905797	16.967345
Sep	2017	7.8695871	3.6811985	12.057976
Oct	2017	0.8774896	5.3158476	3.560868
Nov	2017	10.913006	15.361218	-6.464794
Dec	2017	17.523971	21.976614	13.071329
Jan	2018	21.025701	25.480337	16.571065
Feb	2018	18.284241	22.739774	13.828707
Mar	2018	14.139305	18.595243	-9.683367
Apr	2018	6.9288086	11.384928	-2.472689
May	2018	3.3421964	1.1140054	7.798398

6. Conclusions

The main findings of the project show how the Time Series Modelling can be used, as an effective tool to forecast future temperature . The project specifically considered two separate data sets – 1. Average Yearly Canadian Temperature 2. Monthly Canadian Temperature . Initially Exploratory Data Analysis was done to check for seasonality and trend. After which using three stages of model construction, various models were proposed for both the datasets. The best model for average yearly data set was ARIMA(0,1,1) with drift – and for the monthly seasonal data set was - ARIMA(1,0,1)(2,1,0). The best model were chosen according to AIC .

After the best models were constructed, future forecasts for both the yearly average temperature data and the monthly temperature data were made from a series of models.

These forecasts, can help humanity in sound decision making. The project covers temperature data specific to Canadian Climate , however the same forecasts can be made for other countries and cities.

Future opportunities may include using advanced Time Series Modelling approaches and Deep Learning Approaches. As an example, authors in [9], show how Deep Learning can be used for Time Series Forecasts. By using such new methods comparisons can be performed with proposed methods in the project.

7. Appendices:

7.1 Augmented Dickey Fuller Test:

Augmented Dickey–Fuller test (ADF) tests the null hypothesis that a unit root is present in a time series sample. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity. It is an augmented version of the Dickey–Fuller test for a larger and more complicated set of time series models. The testing procedure for the ADF test is :

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

where α is a constant, β the coefficient on a time trend and p the lag order of the autoregressive process. Imposing the constraints α =0 and β=0 corresponds to modelling a random walk and using the constraint β=0 corresponds to modeling a random walk with a drift.

By including lags of the order p the ADF formulation allows for higher-order autoregressive processes. This means that the lag length p must be determined when applying the test. The unit root test is then carried out under the null hypothesis γ =0, against the alternative hypothesis of , $\gamma < 0$. The intuition behind the test is that if the series is integrated then the lagged level of the series, will provide no relevant information in predicting the change in besides the one obtained in the lagged changes.

7.2 Data Sources - The source for the data is: www.kaggle.com – a Data Science Competition Website , which further attributes the data source as http://berkeleyearth.org/ -Hence Berkley Earth is the true source of data. Berkeley Earth is a Berkeley, California based independent non-profit focused on land temperature data analysis for climate science. Berkeley Earth was founded in early 2010 (originally called the Berkeley Earth Surface Temperature project) with the goal of addressing the major concerns from outside the scientific community regarding global warming and the instrumental temperature record[4].

The link to the Kaggle Website from where the data was taken is -
https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data

The link to Berkley Earth Website which is the true source of the data - http://berkeleyearth.org/source-files/.

Hence both websites contain the data. The data presented in these websites is for all the countries, however in this project we only analyzed data for Canada. Since the dataset is for 200 years , it has not been attached to the report. However a sample of both the yearly data and monthly data has been presented on next pages , in the section 7.2.1 and 7.2.2.

7.2.1. Sample Subset of Average Yearly Data .

Year	Average Temperature
1820	-6.171666667
1821	-5.560333333
1822	-5.820166667
1823	-6.323416667
1824	-5.171583333
1825	-5.1515
1826	-5.1655
1827	-4.64575
1828	-5.096833333
1829	-5.794416667
1830	-4.868833333
1831	-5.775
1832	-5.927833333
1833	-5.248333333
1834	-5.507416667
1835	-6.423333333
1836	-6.068083333
1837	-6.02775
1838	-6.23225
1839	-5.261916667
1840	-5.652416667
1841	-5.715
1842	-5.408916667
1843	-6.10775
1844	-6.255666667
1845	-5.8905
1846	-4.548083333
1847	-5.724666667
1848	-5.93175
1849	-6.069583333
1850	-5.53625
1851	-5.337916667
1852	-5.358
1853	-5.590916667
1854	-5.54575
1855	-5.778083333
1856	-5.46375
1857	-5.626583333
1858	-5.74625
1859	-5.90575
1860	-5.5
1861	-5.791416667
1862	-6.4155
1863	-5.663166667
1864	-5.450916667
1865	-5.56775
1866	-5.967916667
1867	-5.723
1868	-5.805
1869	-5.2675

7.2.2 Sample Subset of Monthly Data

Date	Monthly Temperature
1820-01-01	-25.981
1820-02-01	-22.379
1820-03-01	-16.937
1820-04-01	-6.173
1820-05-01	2.213
1820-06-01	9.547
1820-07-01	11.834
1820-08-01	9.259
1820-09-01	4.977
1820-10-01	-4.305
1820-11-01	-13.734
1820-12-01	-22.381
1821-01-01	-23.534
1821-02-01	-21.173
1821-03-01	-15.897
1821-04-01	-8.278
1821-05-01	2.125
1821-06-01	9.325
1821-07-01	11.489
1821-08-01	10.932
1821-09-01	3.933
1821-10-01	-3.476
1821-11-01	-10.85
1821-12-01	-21.32
1822-01-01	-23.63
1822-02-01	-20.431
1822-03-01	-12.923
1822-04-01	-7.402
1822-05-01	1.829
1822-06-01	9.333
1822-07-01	12.809
1822-08-01	10.779
1822-09-01	3.232
1822-10-01	-5.298
1822-11-01	-14.124
1822-12-01	-24.016
1823-01-01	-22.805
1823-02-01	-23.933
1823-03-01	-17.283

8. References:

1. United Nations , Climate Change Issue Report - https://www.newscientist.com/article/dn11089-the-impacts-of-rising-global-temperatures.
2. Catherine Brahic The impacts of rising global temperatures - https://www.newscientist.com/article/dn11089-the-impacts-of-rising-global-temperatures.
3. www.kaggle.com
4. http://berkeleyearth.org/
5. canada.ca
6. Ary Hoffman - Climate change and biodiversity -https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/temperature-change.html
7. Temperature change in Canada - https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/temperature-change.html.
8. https://www.r-project.org/
9. Xueheng Qiu ; Le Zhang ; Ye Ren ; P. N. Suganthan ; Gehan Amaratunga - Ensemble deep learning for regression and time series forecasting. https://ieeexplore.ieee.org/abstract/document/7015739