CLIMATE CHANGE TIME SERIES

Introduction

Many believe that climate change is one of the biggest threats to the humanity, and that massive changes must be made in how we live to revert it and/or reduce its impact. At the same time, there are also many skeptics, who either do not believe in the threat per se, or that it is caused by human activity, or do not see a need (or possibility) to do much about it. This issue and the debate are, of course, multi-faceted, but in this project, we will look at the statistical evidence for global warming – a largely scientific, fact based view on the situation. The data comes from two well-known sources (NASA and UK Met Office).

Executive Summary

- Our base model predicts a 1.48°C (NASA) and 1.19°C (UKMET) global mean temperature increase by the year 2100 which is lower than the IPCC 2 - 4°C increase projections. However, our alternate approach using Co2 as a regressor is predicting a 2.7 °C increase in global mean by 2100.

- Our model accuracy (measured by MAE or absolute temperature differences) is anywhere between 0.1 °C ~ 0.2 °C. However, when these models are used to predict future values the accuracy ranges between 0.3 °C ~ 0.5 °C.

- Results from our cross validation shows that predicting 10 years into the future before 1967 resulted in a significantly accurate forecast as compared to the present time. We believe this might be to due increasing temperature and significant variance post 1980.

- Actual temperature used in Climate Bet is obtained from University of Alabama. The readings are from Satellite as oppose to Surface and Sea Reading in NASA and UKMET dataset. The readings are significantly different.

- We have performed t-test hypothesis on NASA and UKMET dataset. Our results show that both the datasets are significantly different. This is also apparent in the visual shown under EDA section of the report. The readings have briefly reconciled during 1920-1950 but NASA reading after that is on higher side.

- There is a very significant statistical correlation between Co2 emissions (ppm) and climate change (~0.91) which has started to weaken in recent years.

- Our Explanatory Data Analysis (EDA) shows that temperature in Northern Hemisphere is on a constant increase in last 4 decades and is significantly above global mean temperature or Southern Hemisphere. Our qualitative research shows this is due to having more land masses in north (south has more oceans) and ocean current transporting heat from South to North but there is more to this. However, this doesn't affect the model in any way because all analyses are done on global means.

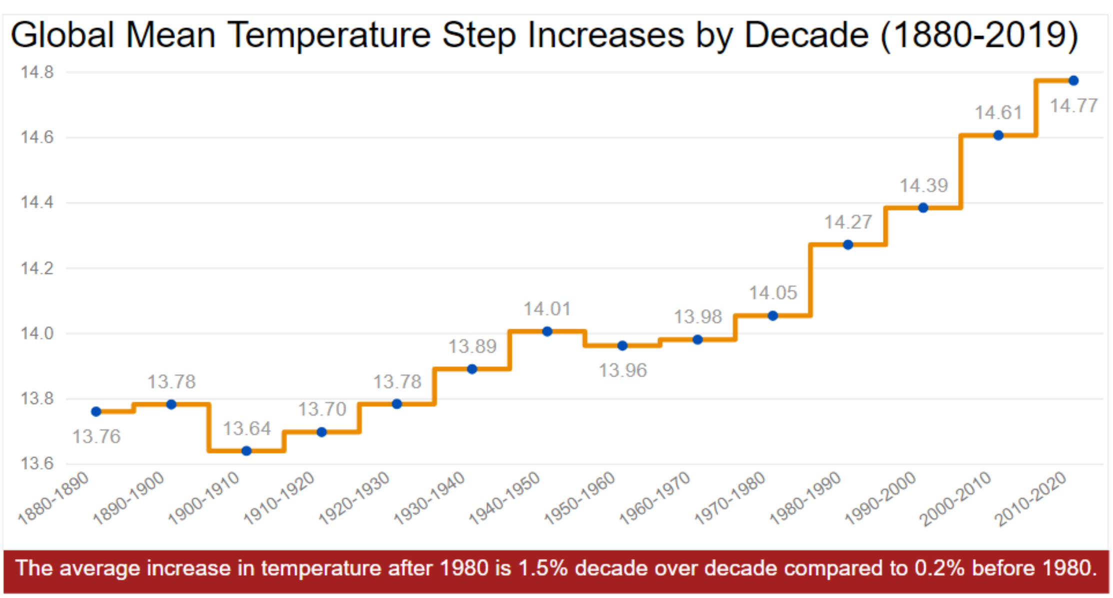
PROJECT DATASETS

1. **GISTEMP Team, 2019**: *GISS Surface Temperature Analysis (GISTEMP)*. NASA Goddard Institute for Space Studies. Dataset accessed 2019-04-29 at https://data.giss.nasa.gov/gistemp/. *This has been referred as 'NASA' throughout the document.*

2. **Climatic Research Unit** (University of East Anglia) in conjunction with the Hadley Centre (UK Met Office). https://crudata.uea.ac.uk/cru/data/temperature/. *This has been referred as 'UK Met' throughout the document.*

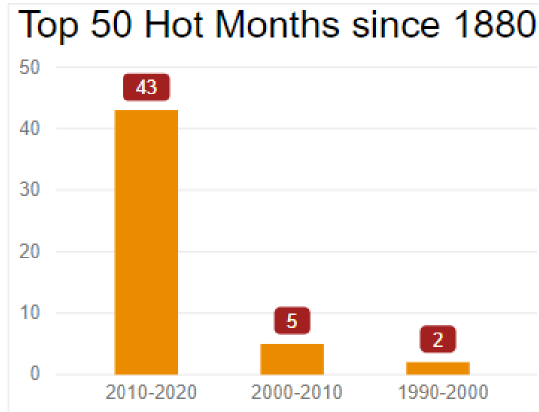
EXPOLARATORY DATA ANALYSIS (EDA)

To gain an initial understanding of the global temperature datasets, the historical trends of global mean temperature were plotted. Instead of a traditional line chart, the 139 years of data has been aggregated into decades with the intention of analyzing the step increases. All visuals presented in this document are built using the NASA dataset, unless mentioned otherwise.



The chart shows a series of sharp decade over decade increases in global mean temperatures post 1980 with a record high increase of 0.2 °C in 2000-2010.

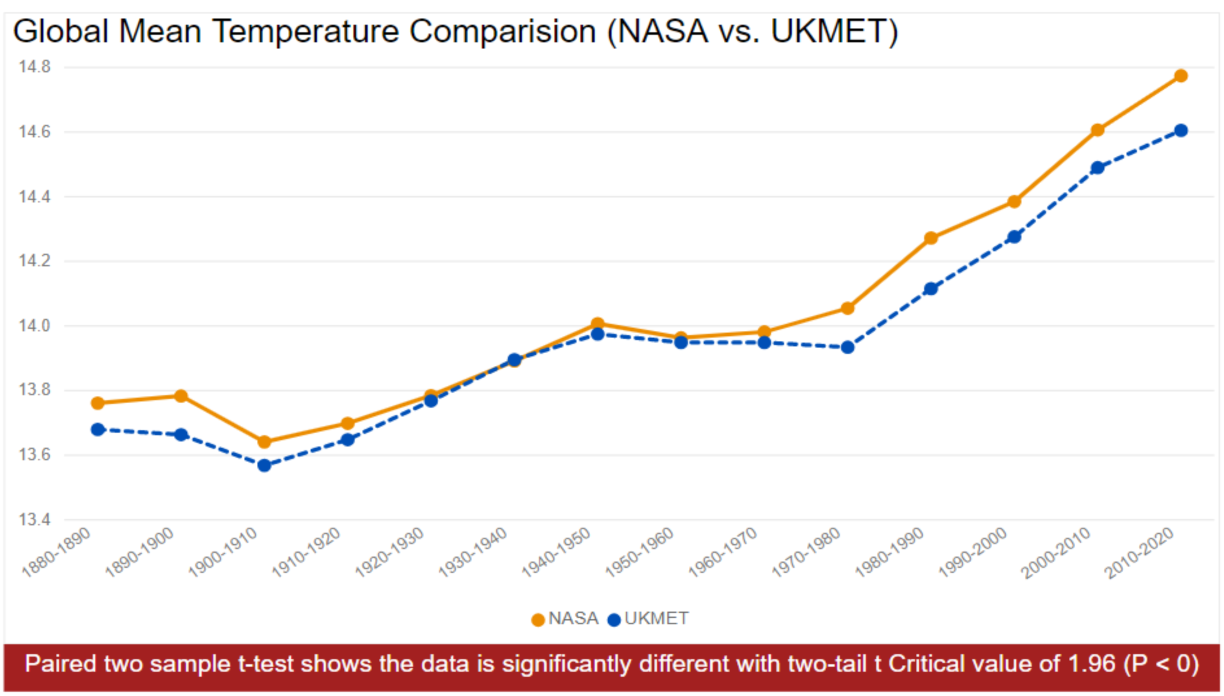
A review of the top 50 warmest months since 1880 supports the increasing trend as 43 out of top 50 hottest months since 1880 occurred within the present decade (2010-2020).



Comparison of NASA and UKMET Datasets

In order to identify if any initial significant differences exist between the two global temperature datasets, a paired two sample t-test was performed with a null hypothesis that assumes the true mean difference (µd) equals zero. The result suggests that both the datasets are significantly different (P value less than 0.00).

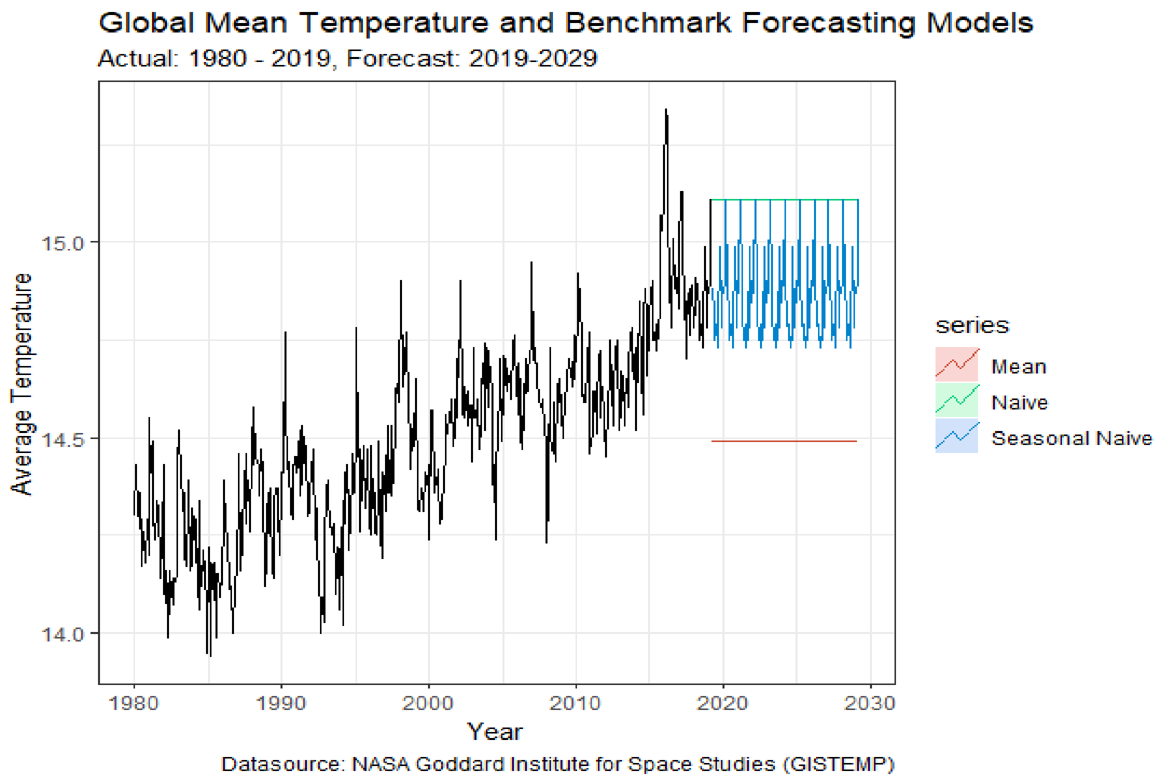
The chart below shows that both datasets were closely reconciled during 1930 – 1970. From 1980 onwards however, the NASA measurements are slightly higher which could impact the modeling process.



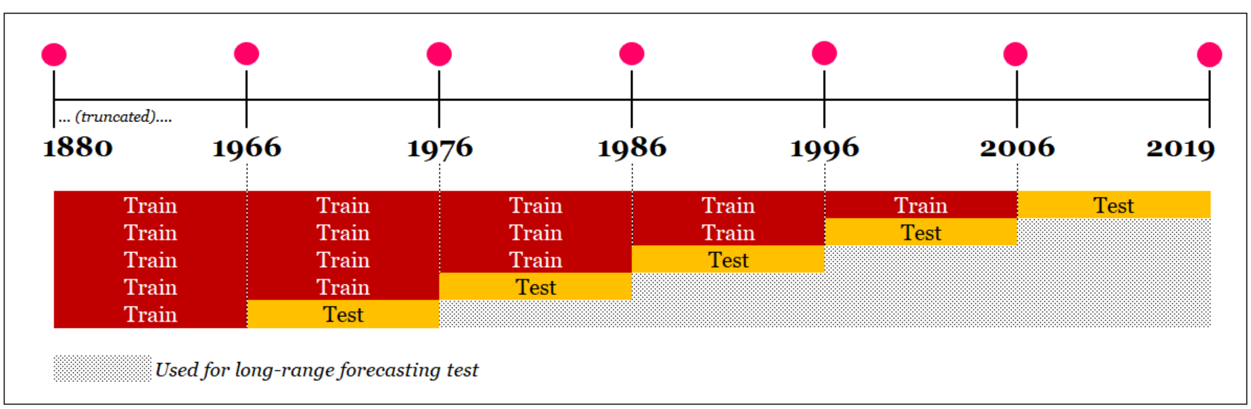
MODELLING PROCESS

Relying on a single model to predict temperatures that could likely occur over 80 years into the future would be very little evidence to base any significant insights on. A more robust approach was employed which involved running five different time series models over each dataset.

The first benchmark model utilized mean, naïve and seasonal naïve methods and was trained using the complete dataset. For clarity purposes the below graph has been truncated.



The results were then tested for accuracy and cross validated over five partitions which cover different time period intervals.



The below tables show that in almost all partitions ARIMA had the best performance which made it was the primary model of choice for building the future year temperature point predictions.

NASA: Test Period: 10 Year Interval (End not constant)

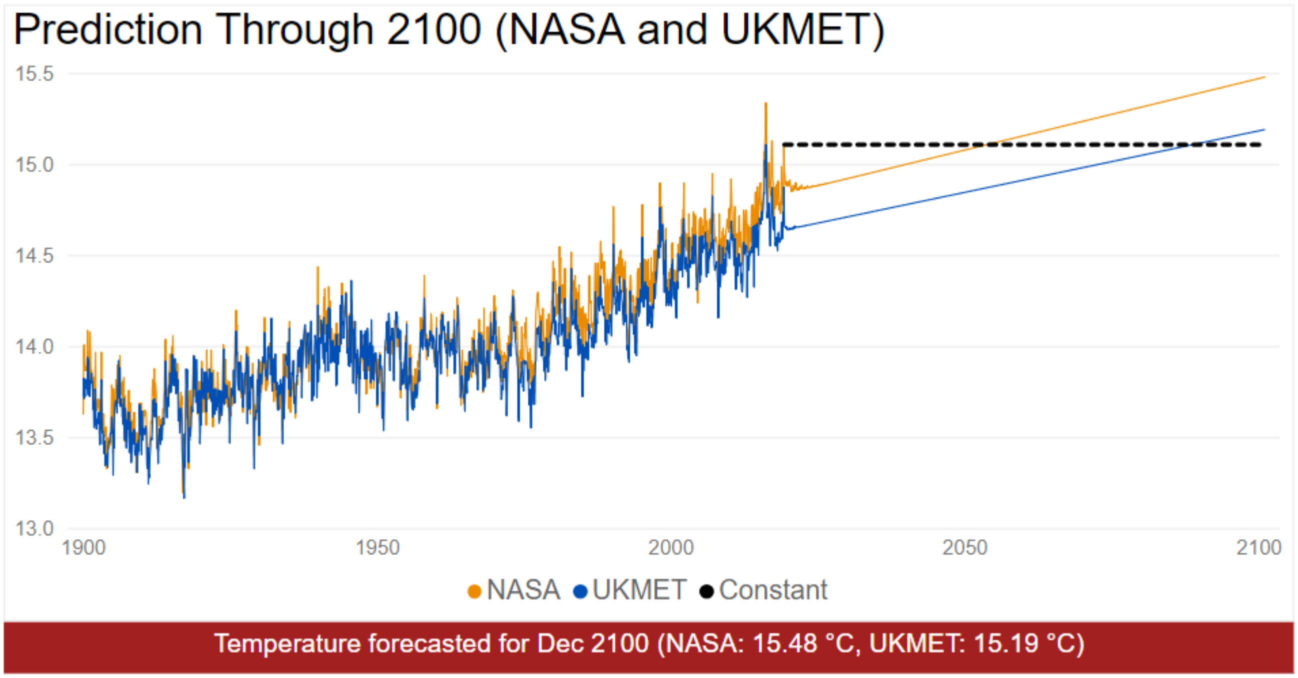
CV	Train Period	Test Period	**AAA**	**ANN**	**MMM**	**TBATS**	**ARIMA**
1	1/1880-12/2006	1/2007-3/2019	0.141	0.141	0.140	0.164	0.138
2	1/1880-12/1996	1/1997-12/2006	0.195	0.194	0.204	0.260	0.185
3	1/1880-12/1986	1/1987-12/1996	0.219	0.216	0.232	0.227	0.173
4	1/1880-12/1976	1/1977-12/1986	0.225	0.210	0.219	0.218	0.216
5	1/1880-12/1966	1/1967-12/1976	0.108	0.105	0.110	0.105	0.103

UKMET: Test Period: 10 Year Interval (End not constant)

CV	Train Period	Test Period	**AAA**	**ANN**	**MMM**	**TBATS**	**ARIMA**
1	1/1880-12/2006	1/2007-3/2019	0.136	0.128	0.131	0.130	0.112
2	1/1880-12/1996	1/1997-12/2006	0.250	0.260	0.260	0.252	0.232
3	1/1880-12/1986	1/1987-12/1996	0.188	0.197	0.194	0.188	0.168
4	1/1880-12/1976	1/1977-12/1986	0.192	0.203	0.197	0.176	0.172
5	1/1880-12/1966	1/1967-12/1976	0.108	0.107	0.108	0.110	0.122

1) Projected global temperatures through year 2100

The primary model predicts a 1.48°C (NASA) and 1.19°C (UKMET) global mean temperature increase by the year 2100 which is lower than the IPCC 2 - 4°C increase projections.



2) Point Predictions

GLOBAL MEAN TEMPERATURE PREDICTION THROUGH 2100

	NASA			UKMET		
Period	**Point Prediction**	**Low 90**	**High 90**	**Point Prediction**	**Low 90**	**High 90**
JAN 2030	14.92	14.59	15.25	14.71	14.40	15.03
JUL 2030	14.93	14.59	15.26	14.72	14.40	15.03
JAN 2050	15.08	14.66	15.50	14.85	14.46	15.24
JUL 2050	15.08	14.67	15.50	14.85	14.46	15.24
JAN 2100	15.47	14.90	16.05	15.19	14.65	15.72
JUL 2100	15.48	14.90	16.05	15.19	14.65	15.72

ALTERNATIVE METHOD

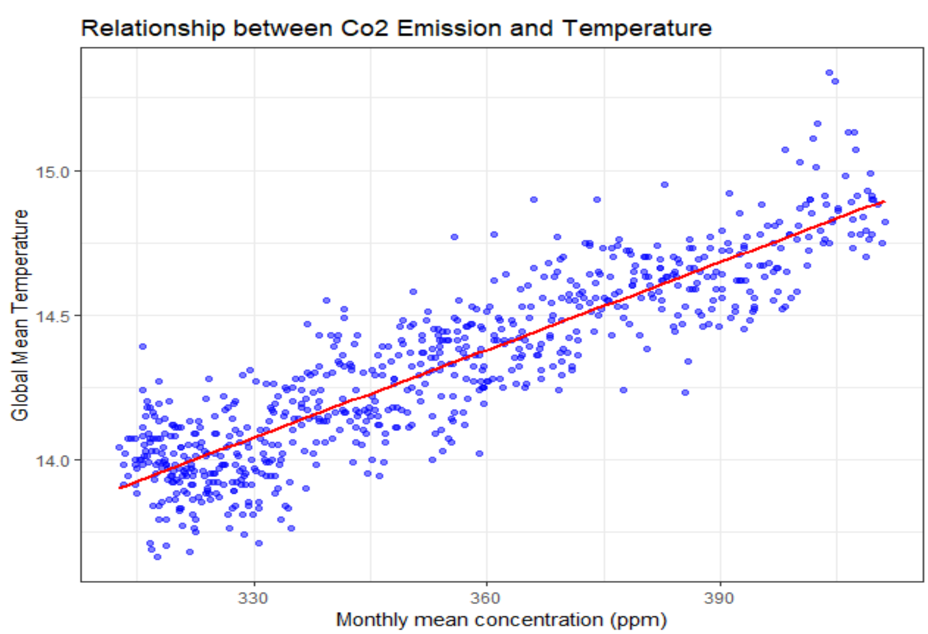
The July 2100 point predictions indicate a 0.65 to 2.05 °C increase in global mean temperate (90% confidence intervals) which is lower than the 2-4°C projections made by climate science organizations such as the IPCC.

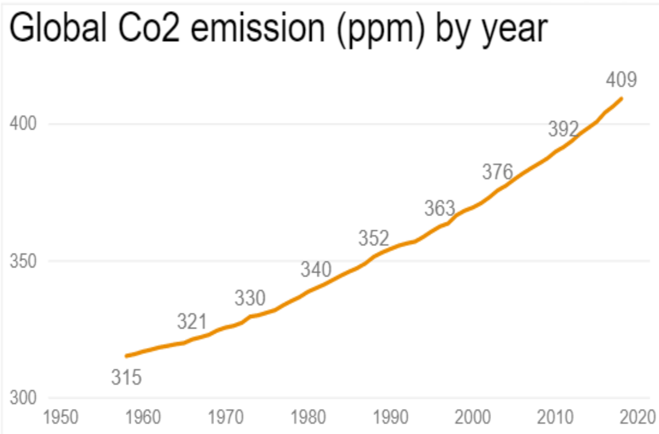
Upon further investigation it becomes apparent that relying solely on past temperature data is insufficient when trying to predict climate change since there are other important variables that should be taken into account. The most notable of these would be CO2 emissions since it is widely believed that human activity which adds more carbon to the atmosphere will increase the rate of warming.

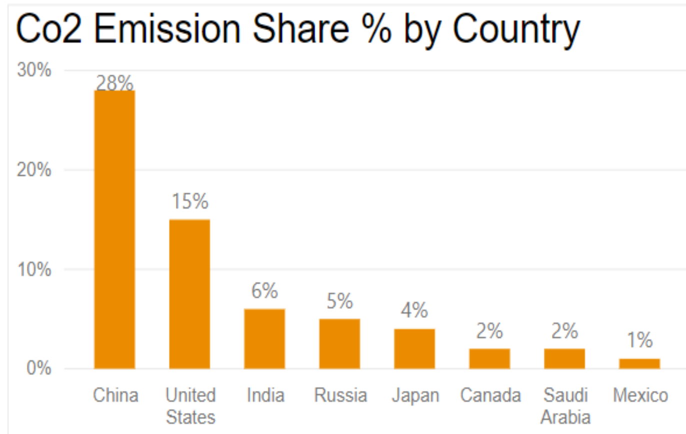
To account for this, monthly atmospheric Co2 emission data has been obtained from **NOAA-ESRL Global Monitoring at C02.earth**. The unit of measure is parts per million (ppm) which measures the concentration of carbon in the atmosphere: https://www.co2.earth/monthly-co2#noaa.

Exploration

Initial exploration shows that there is a very strong relationship between temperature increase and Co2 emissions. The statistical correlation is ~ 0.9 which is highly significant.



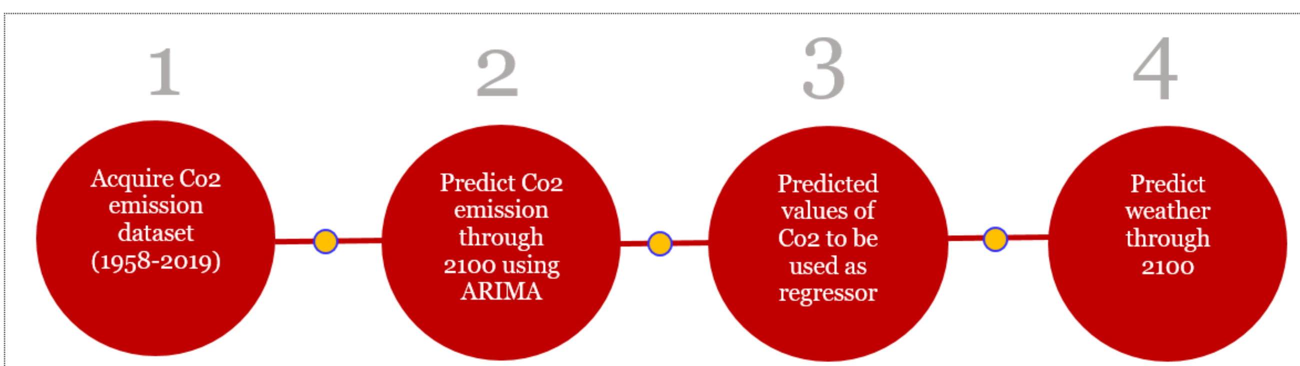




This trend is enough information to justify building an alternate approach model to predict temperature through 2100 using Co2 emissions. Other variables such as changes in the Earth's orbit, strength of the sun, greenhouse gases, plate tectonics and volcanic eruptions could also potentially be factored in but are beyond the scope of this analysis.

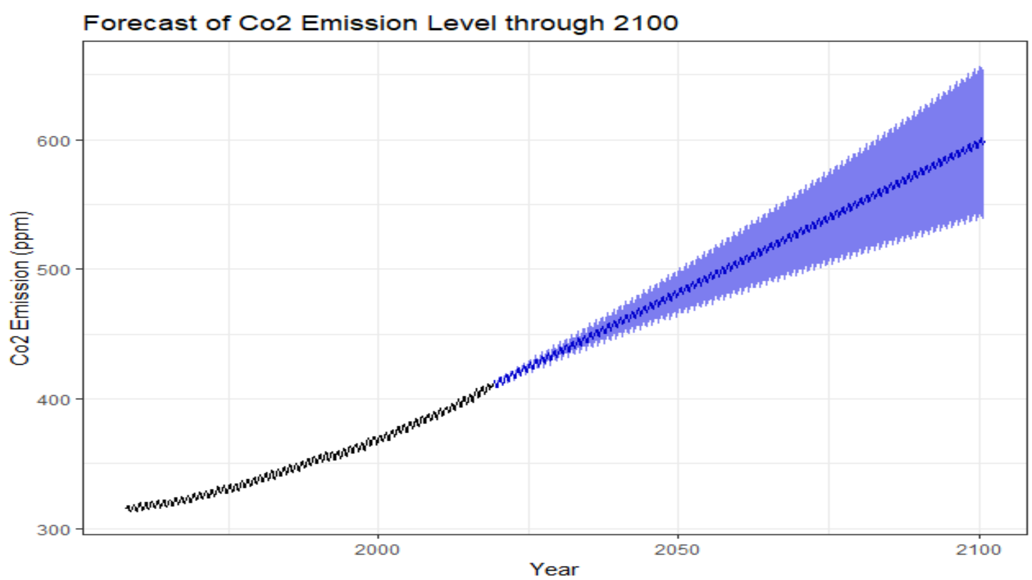
CO2 Regressor

To use CO2 as a regressor, we have built a model to predict future emissions through year 2100 and passed them as a parameter matrix into the original temperature model. The process is summarized below:



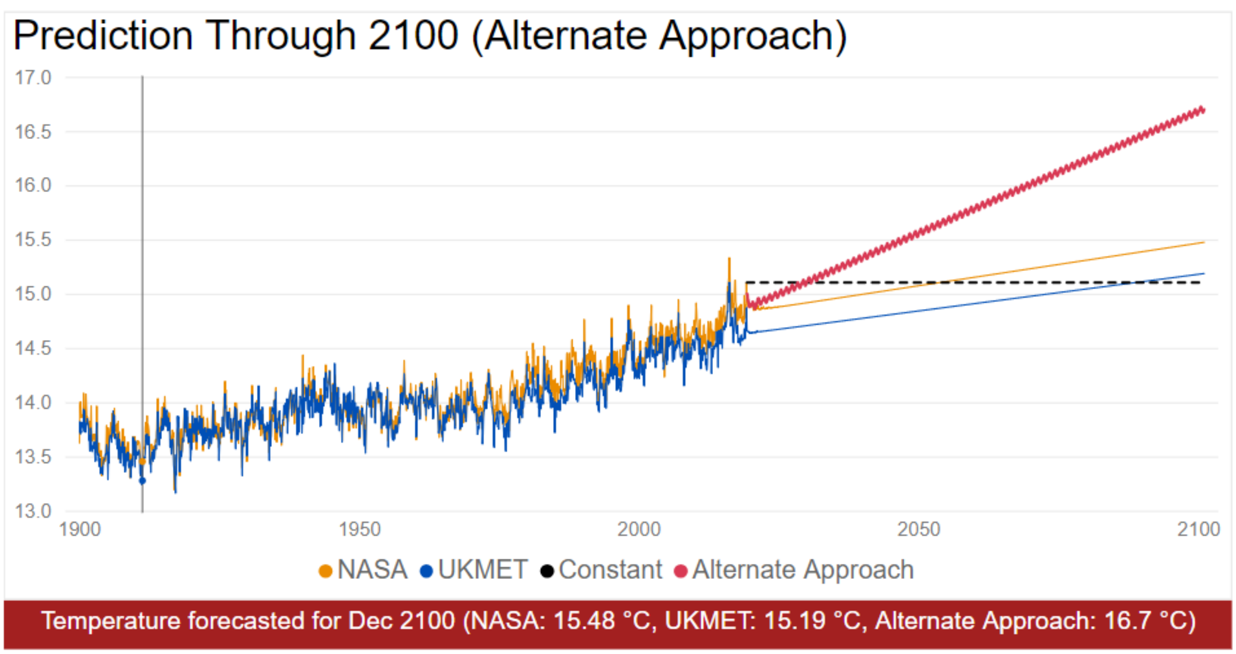
The model predicts the sharp rising level of Co2 emission based on historical data from 1958 onwards. The increasing trend is consistent with the forecasts produced by co2.earth (https://www.co2.earth/2100-projections). The Co2 concentration level in 2100 matches very closely with 2°C pathway published on October 27, 2015 which shows that the temperature will rise by 2°C in 2100 with a given concentration level of ~ 670 ppm (ARIMA prediction is 610 ppm).

Note that predicted CO2 emissions do not factor in efforts to reduce carbon such as electric vehicles and government policy.



Alternative Approach Point Predictions

Adding the CO2 emission predicted growth rates as a regressor results in a 2.7 °C average global temperature increase prediction which is quite a bit higher than original ARIMA model and much closer to the 3°C IPCC projection.



GLOBAL MEAN TEMPERATURE PREDICTION THROUGH 2100

	Alternative (C02 Regressor)
Period	**Point Prediction**
JAN 2030	15.12
JUL 2030	15.13
JAN 2050	15.57
JUL 2050	15.58
JAN 2100	16.69
JUL 2100	16.70

3) Kingston, Ontario Temperature Predictions

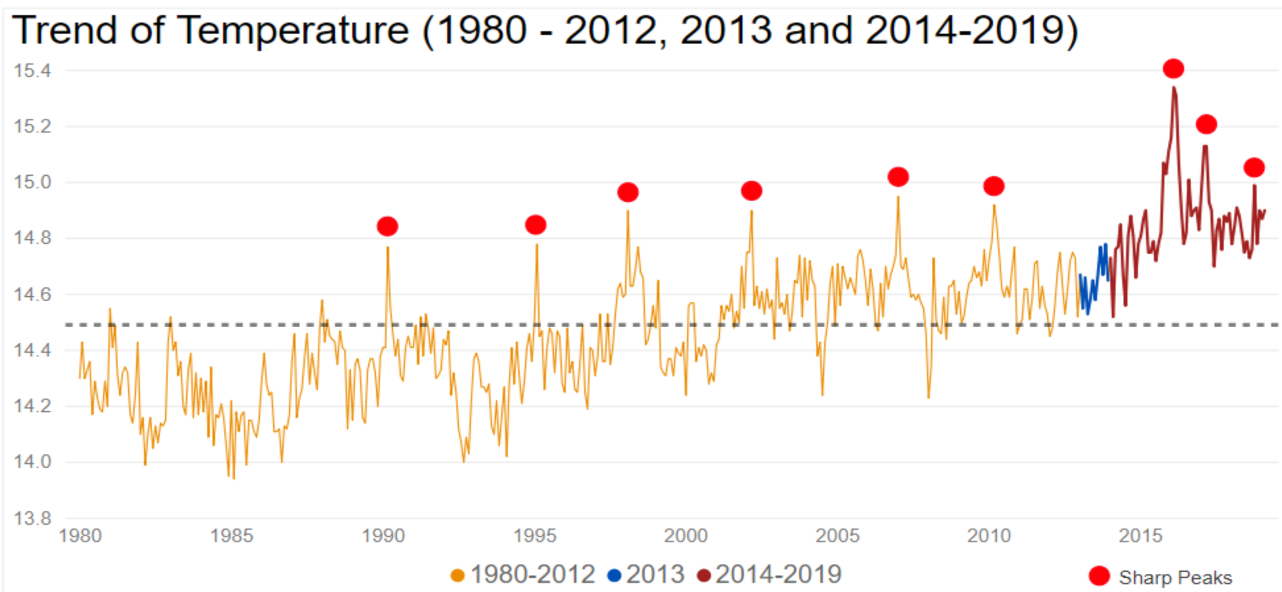
Weather station data for Kingston, Ontario (postal code K7L 3N6) has been acquired through the original two data sources. Missing values were imputed using Environment and Climate Change Canada datasets available from:
http://climate.weather.gc.ca/historical_data/search_historic_data_e.html

KINGSTON WEATHER PREDICTION THROUGH 2100

	NASA			UKMET		
Period	**Point Prediction**	**Low 90**	**High 90**	**Point Prediction**	**Low 90**	**High 90**
JAN 2030	-6.99	-10.31	-3.67	-7.26	-10.66	-3.84
JUL 2030	20.91	17.58	24.24	21.09	17.68	24.50
JAN 2050	-6.99	-10.58	-3.41	-7.26	-10.98	-3.52
JUL 2050	20.91	17.32	24.40	21.09	17.35	24.82
JAN 2100	-6.99	-11.15	-2.83	-7.25	-11.68	-2.83
JUL 2100	20.91	16.75	25.08	21.09	16.65	25.52

4) 2013 Temperature stabilization

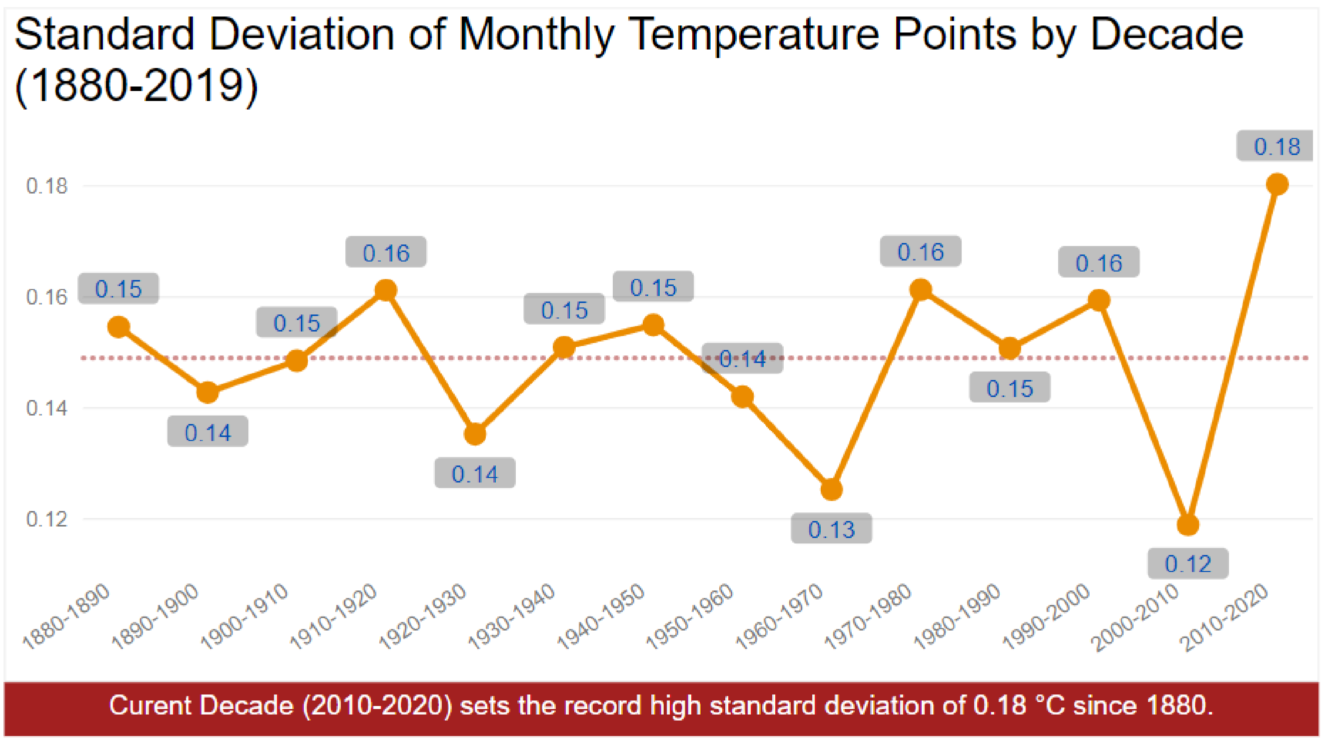
Temperatures did stabilize relative to prior periods during 2013 but this is a very small timeframe from which to draw broad conclusions. The year over year increase in 2013 was 0.03 °C compared to the overall mean change of 0.01 °C with a standard deviation of 0.11 °C. Based on this, the probability of such a change is well over 57% which is not statistically significant enough to conclude that global warming had plateaued.



2013 and 2014 was period of relative stability. However, flucttutations are observed more frequently from Sep 2015 with record high of 15.34 °C in Feb 2016. Fluctuations are not only seveare but more frequent as well.

5) Temperature fluctuations

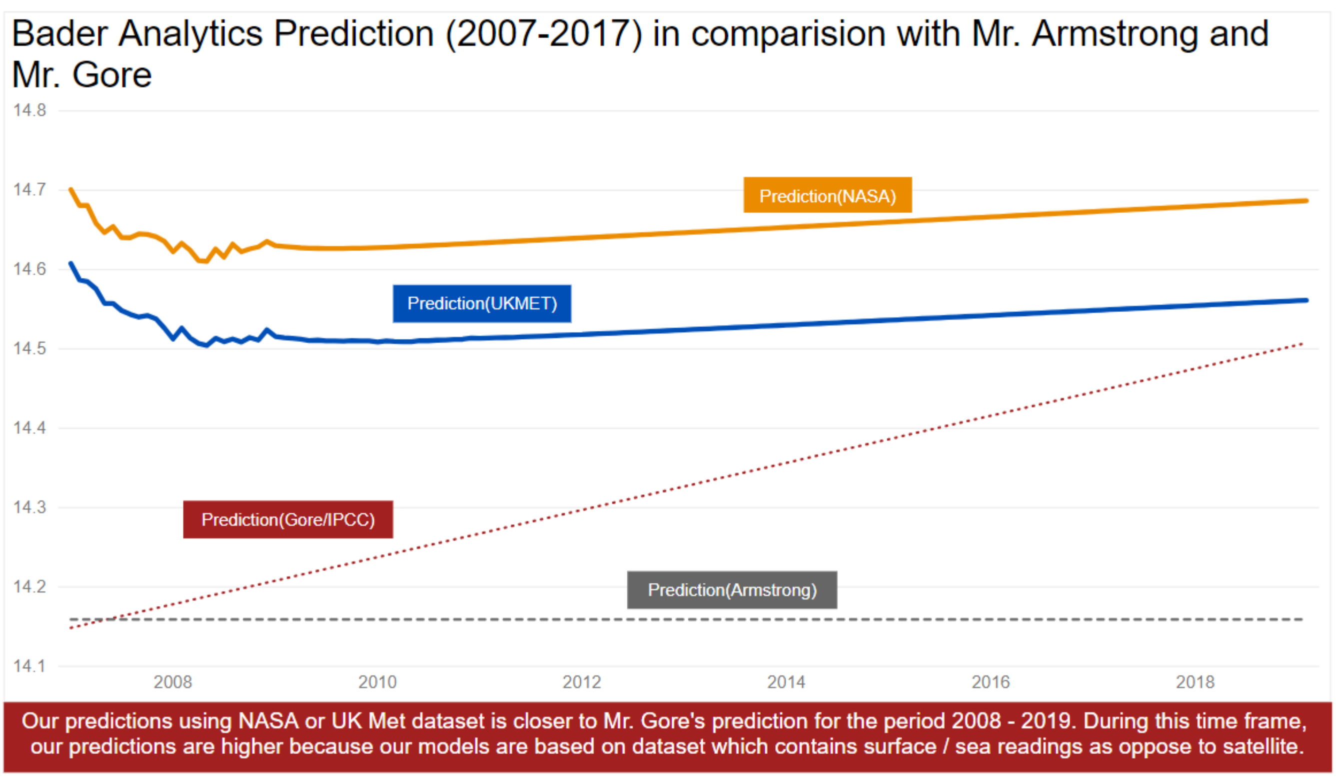
Although 2013 was a year of relative stability it is not a long enough period of time to arrive at any significant conclusions. In fact, not only did the fluctuations became larger post 2013 but they also become more frequent during 2015-2019. An analysis of the standard deviation of monthly temperatures by decade confirms that fluctuations are getting larger and more frequent in the most recent set of years.



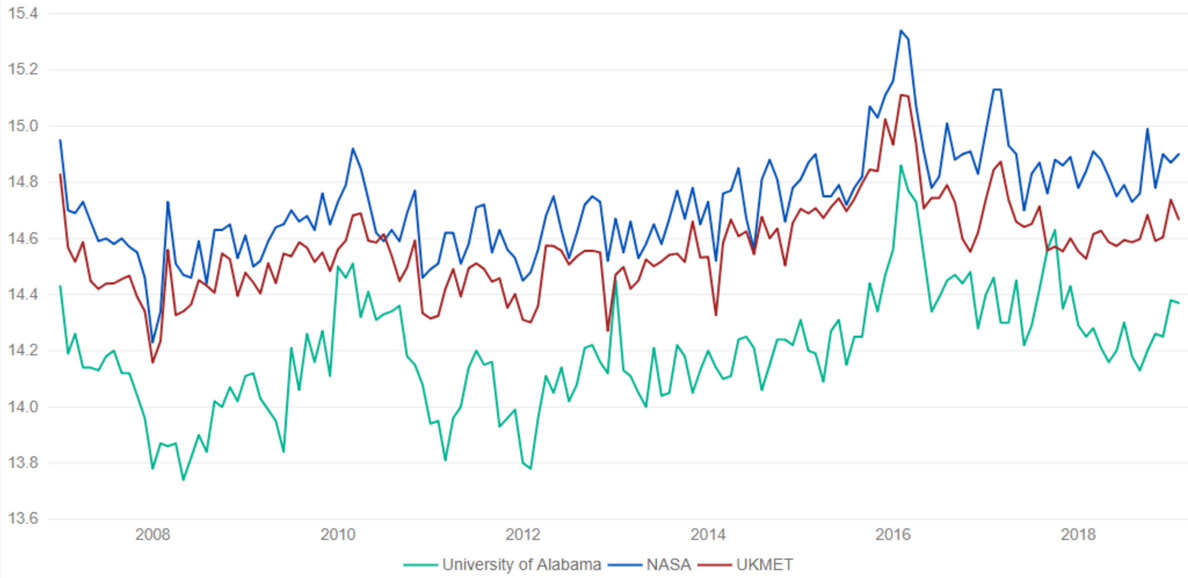
6) Gore vs. Armstrong 'Climate Bet' 2007-2017 Predictions

Training the original model on the pre-2007 NASA and UKMET data results in predictions that are closer to the Gore/IPCC projections. An important observation is that the models did not produce the same rate of constant growth but were closer to the actual values. The NASA predictions had a cumulate absolute error of 47.25°C vs UKMET 29.84°C.

One important note is that the climate bet data referenced on theclimatebet.org was obtained from the University of Alabama at Huntsville (UAH). It uses satellite-based temperature measurements whereas NASA and UKMET use surface and sea weather stations. This is highly relevant since the UAH actual temperature readings are lower than both NASA and UKMET which would favour Armstrong's no trend claim. This is likely an example of pre-selecting data that supports a low temperature position in order to discredit the existence and/or severity of climate change.



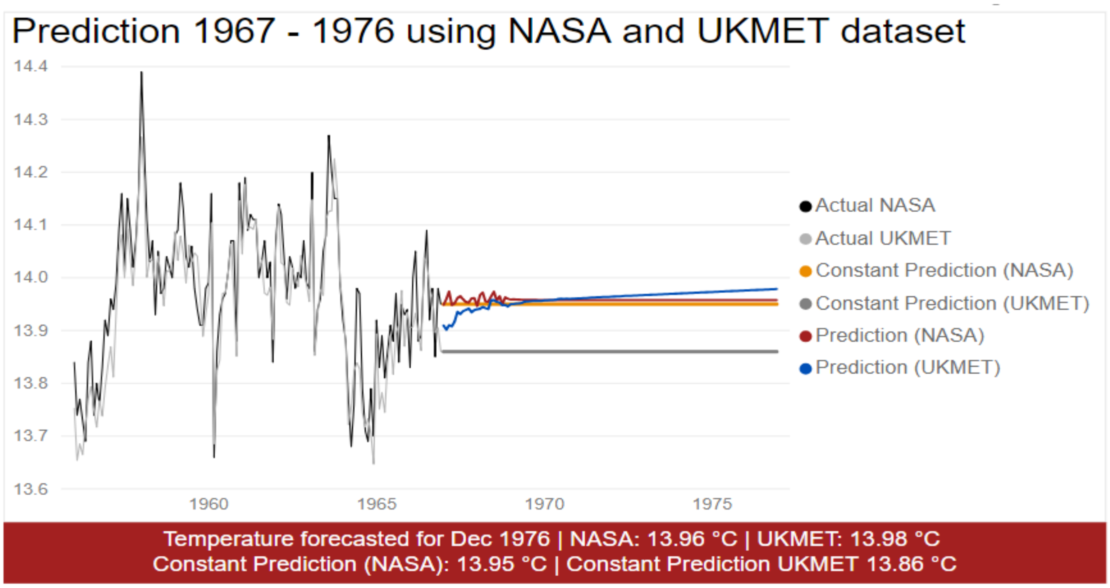
Actual Temperature Readings by NASA, UKMet, UoA (2007 - To date)



Temperature readings between three different sources are significantly different. University of Alabama's reading are based on satellite temperature as oppose to land surface and sea temperature which is used by NASA and UK Met.

7) 1966-1976 Prediction

The primary model was re-run using all years prior to 1966 to predict the next 10 years. The NASA based prediction had a MAE of .1033 while UKMET was .1215. The results in the chart below show that the model is close to a constant temperature prediction which is in line with expectations since the global average temperature did not begin to sharply increase until after 1980. Also of note is that the models did not capture the same level of fluctuations that are seen in the actual values but this is partially due to the chart scaling.



APPENDIX

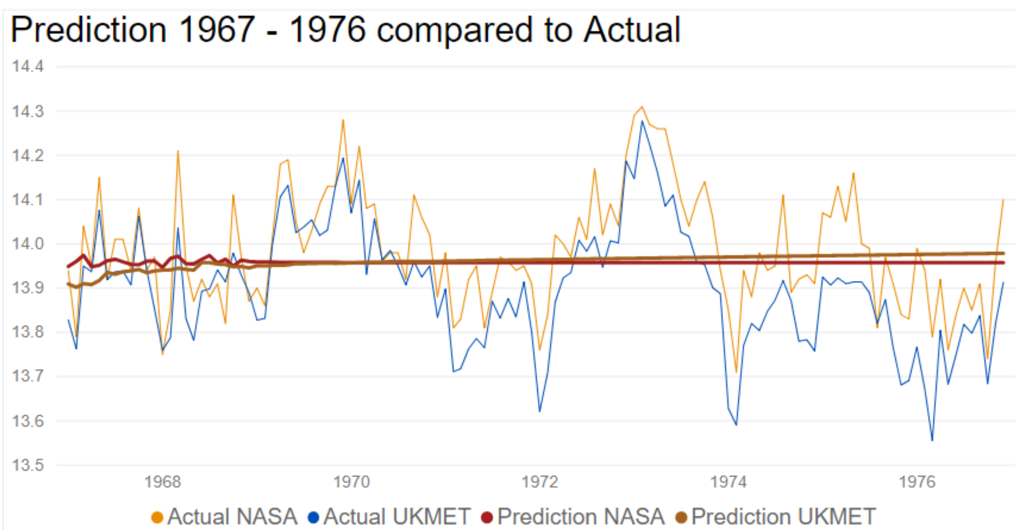
Model Process: Long Range Results.

UKMET: Test Period: Rolling (End constant at 3/2019 for long-range accuracy measure)

CV	Train Period	Test Period	AAA	ANN	MMM	TBATS	ARIMA
1	1/1880-12/2006	1/2007-3/2019	0.136	0.128	0.131	0.130	0.112
2	1/1880-12/1996	1/1997-3/2019	0.315	0.325	0.325	0.317	0.272
3	1/1880-12/1986	1/1987-3/2019	0.396	0.406	0.402	0.396	0.338
4	1/1880-12/1976	1/1977-3/2019	0.470	0.482	0.477	0.454	0.448
5	1/1880-12/1966	1/1967-3/2019	0.367	0.375	0.374	0.349	0.266

NASA: Test Period: Rolling (End constant at 3/2019 for long-range accuracy measure)

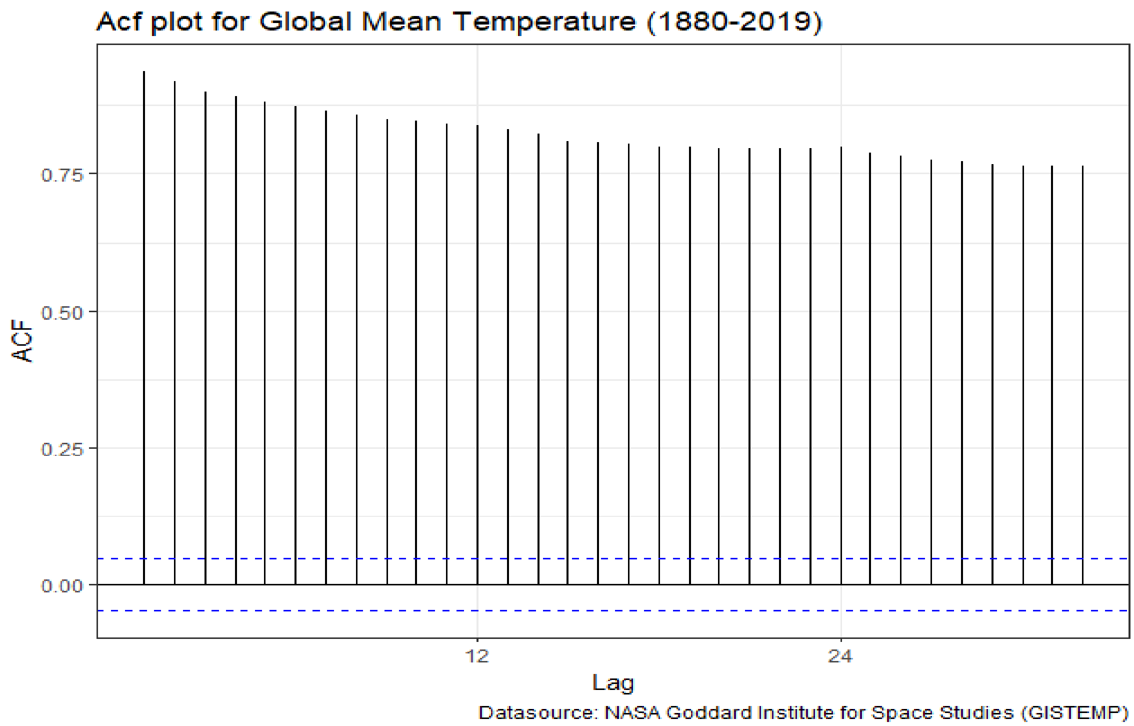
CV	Train Period	Test Period	AAA	ANN	MMM	TBATS	ARIMA
1	1/1880-12/2006	1/2007-3/2019	0.141	0.141	0.140	0.164	0.138
2	1/1880-12/1996	1/1997-3/2019	0.287	0.287	0.298	0.363	0.250
3	1/1880-12/1986	1/1987-3/2019	0.389	0.417	0.434	0.429	0.322
4	1/1880-12/1976	1/1977-3/2019	0.497	0.481	0.491	0.491	0.489
5	1/1880-12/1966	1/1967-3/2019	0.447	0.440	0.452	0.433	0.432

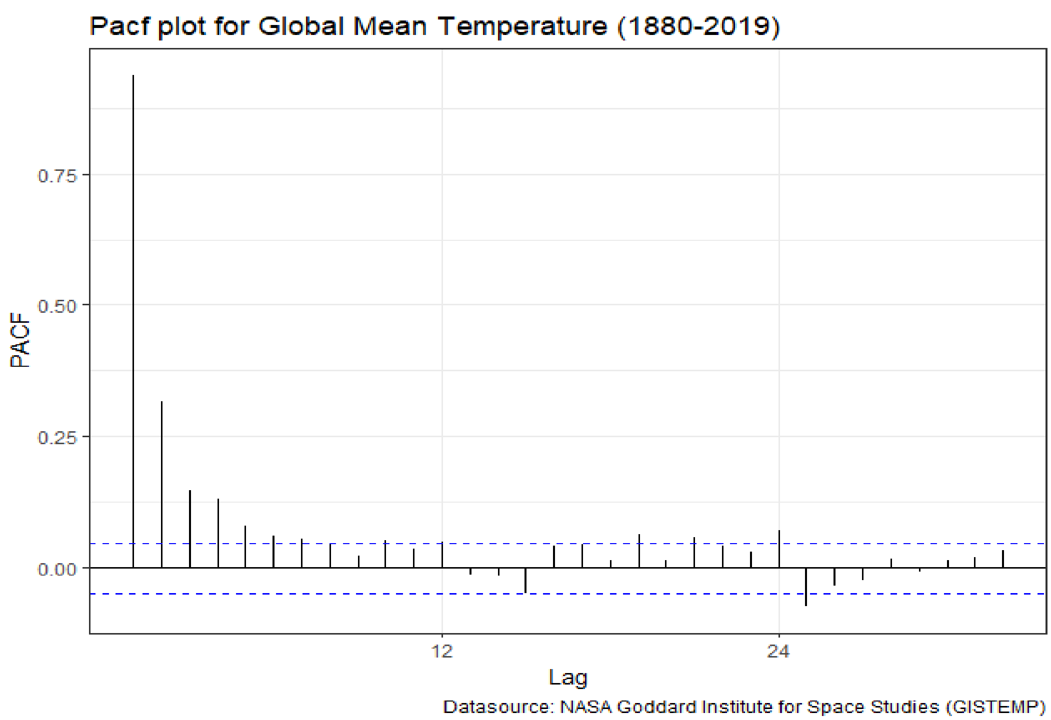


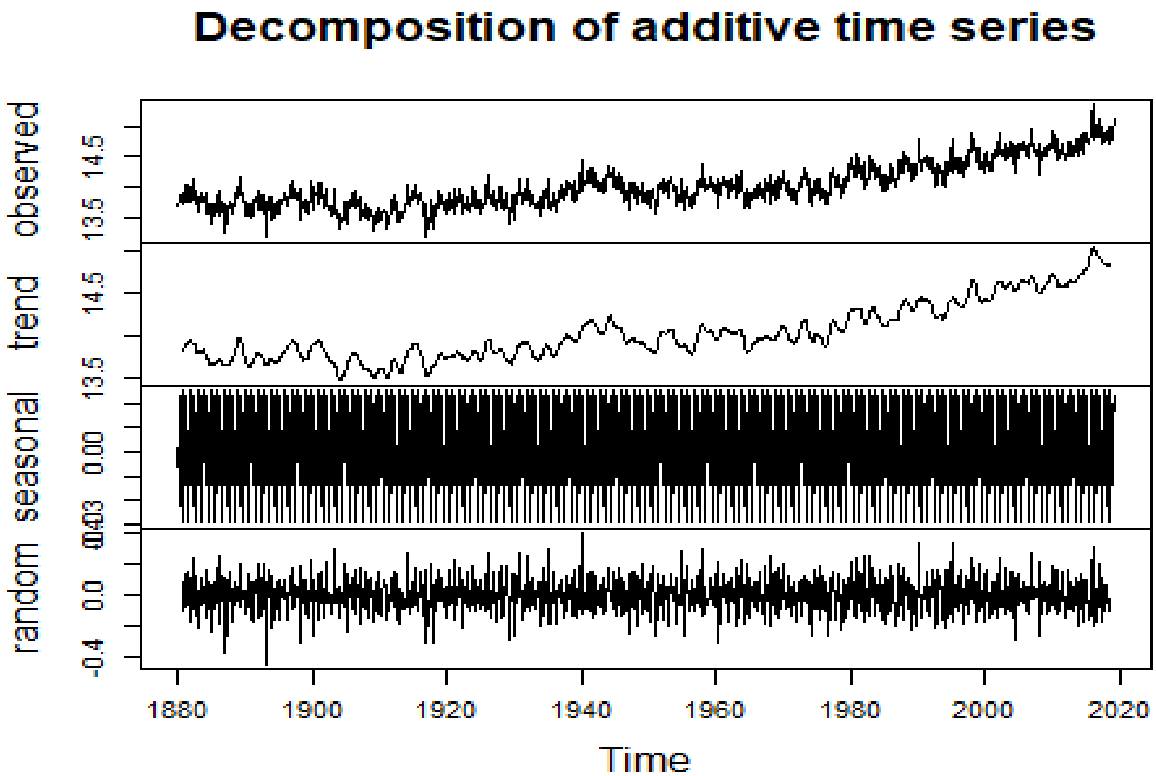
See Attached Excel File for Annexure 1-4

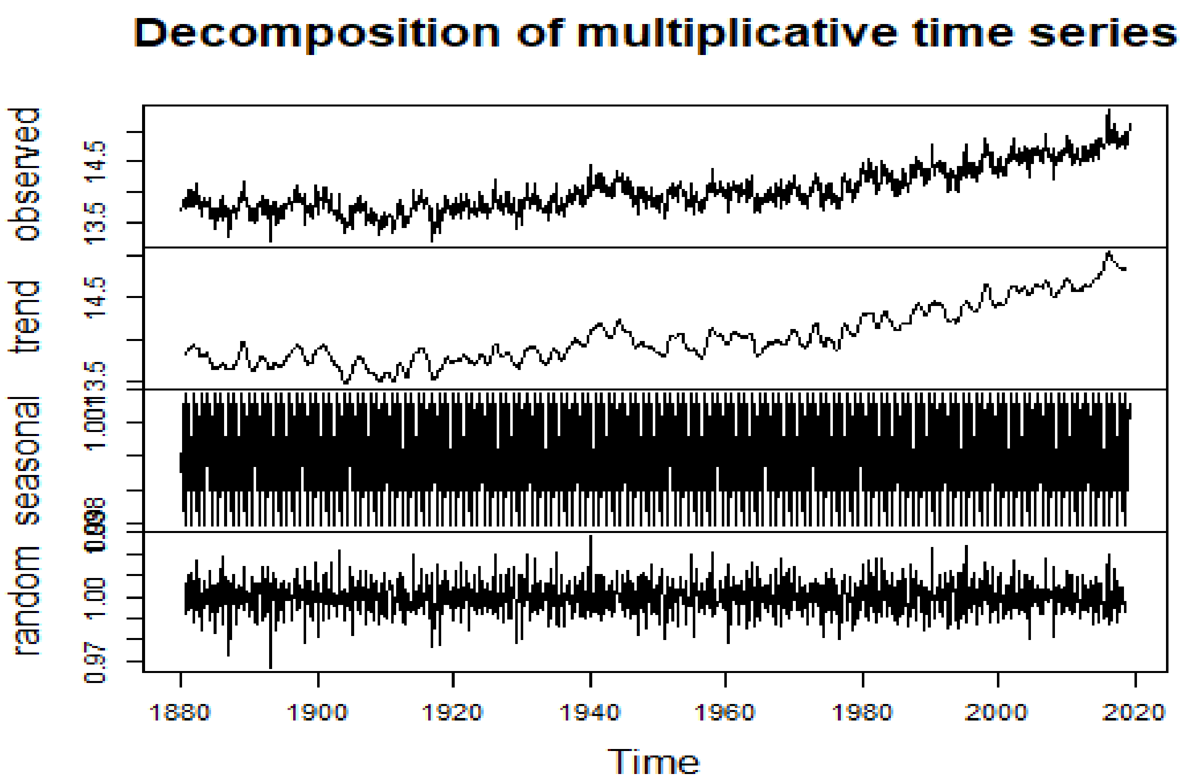
Annexure 5

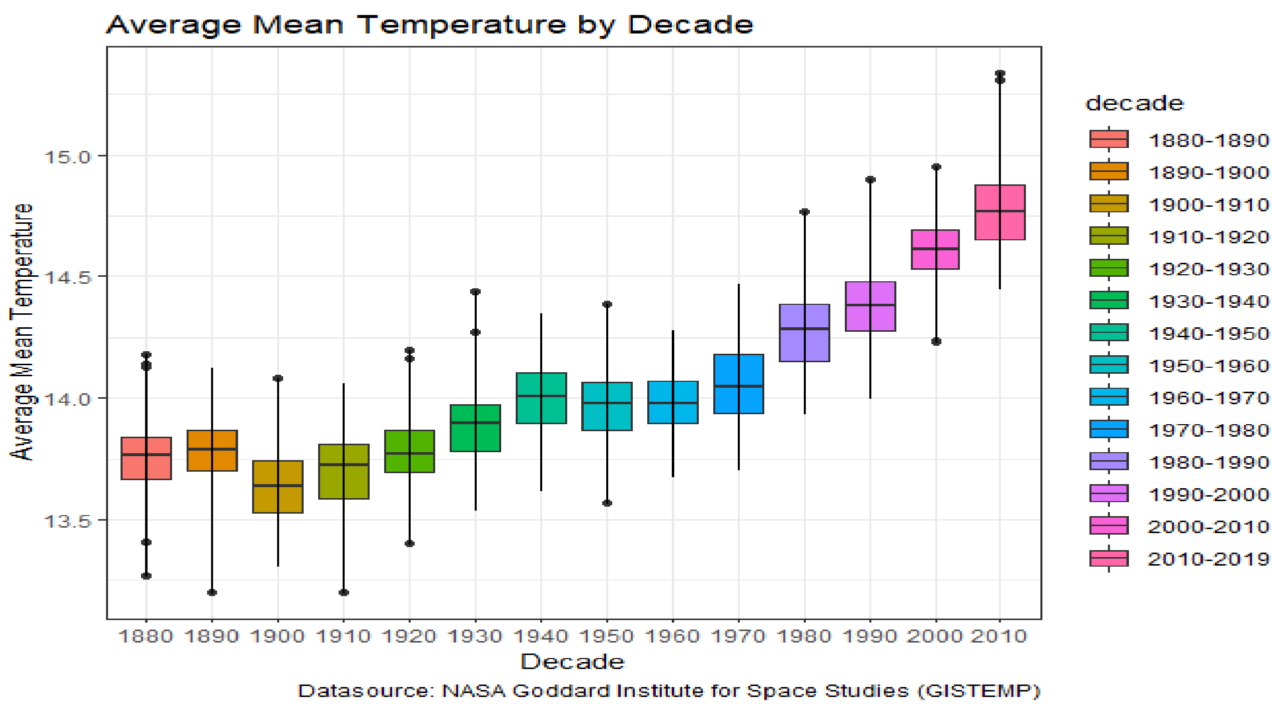
NASA Model



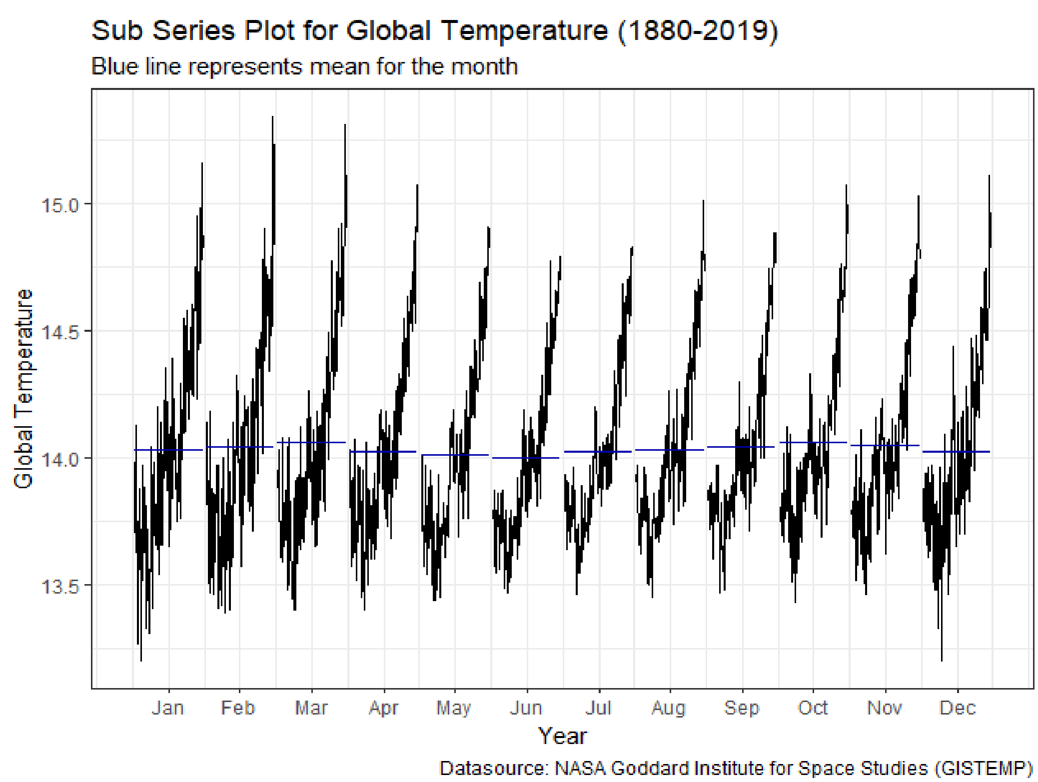



Decomposition of additive time series
observed
trend
seasonal
random
13.5
14.5
0.00
-0.4
0.0
1880
1900
1920
1940
1960
1980
2000
2020
Time

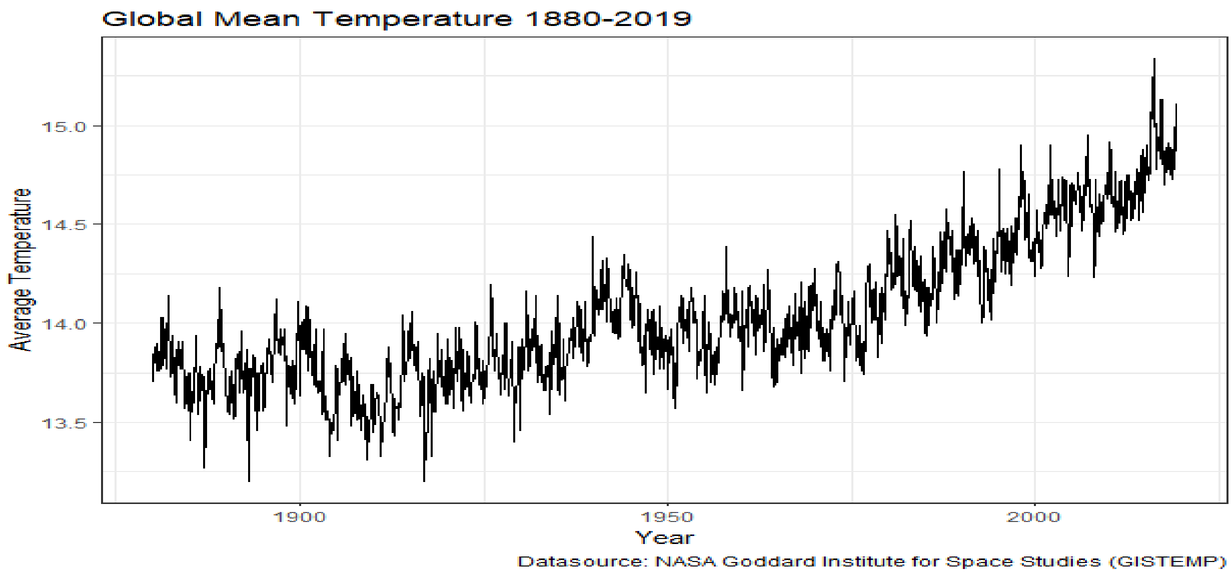

Decomposition of multiplicative time series
observed
trend
seasonal
random
13.5
14.5
1.00
0.97
1880
1900
1920
1940
1960
1980
2000
2020
Time

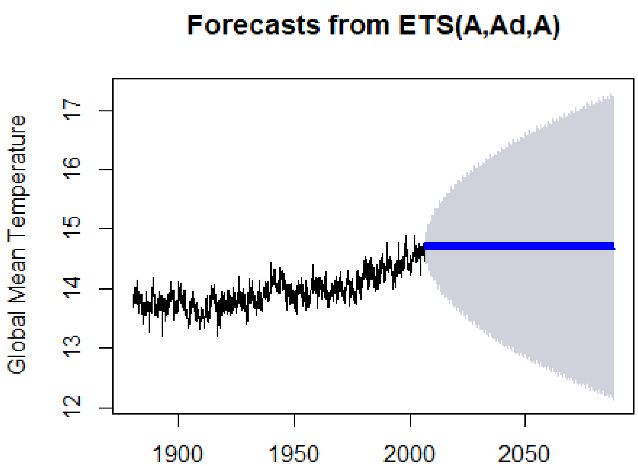


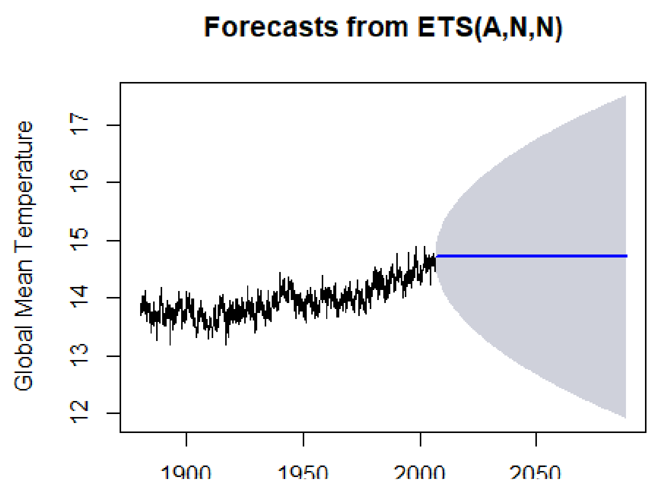
Average Mean Temperature by Decade
Average Mean Temperature
15.0
14.5
14.0
13.5
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
Decade
decade
1880-1890
1890-1900
1900-1910
1910-1920
1920-1930
1930-1940
1940-1950
1950-1960
1960-1970
1970-1980
1980-1990
1990-2000
2000-2010
2010-2019
Datasource: NASA Goddard Institute for Space Studies (GISTEMP)

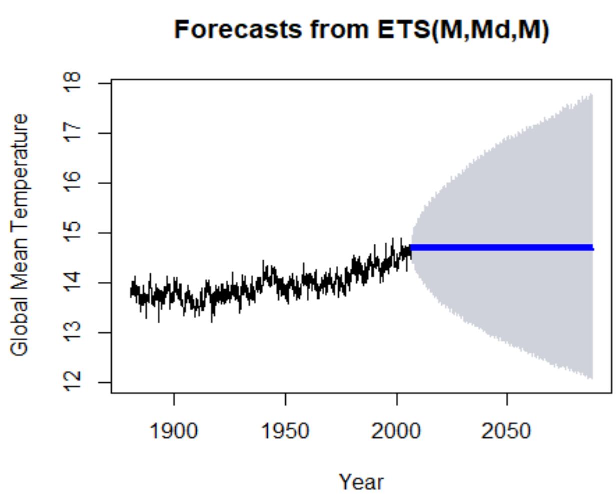


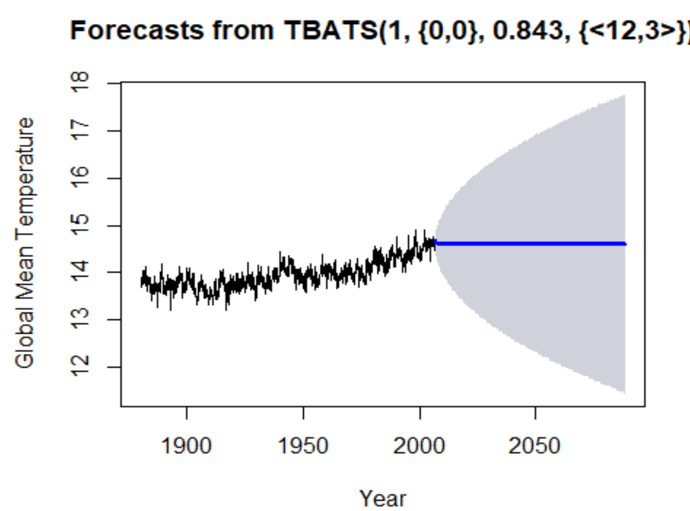
Sub Series Plot for Global Temperature (1880-2019)
Blue line represents mean for the month
Global Temperature
15.0
14.5
14.0
13.5
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Year
Datasource: NASA Goddard Institute for Space Studies (GISTEMP)

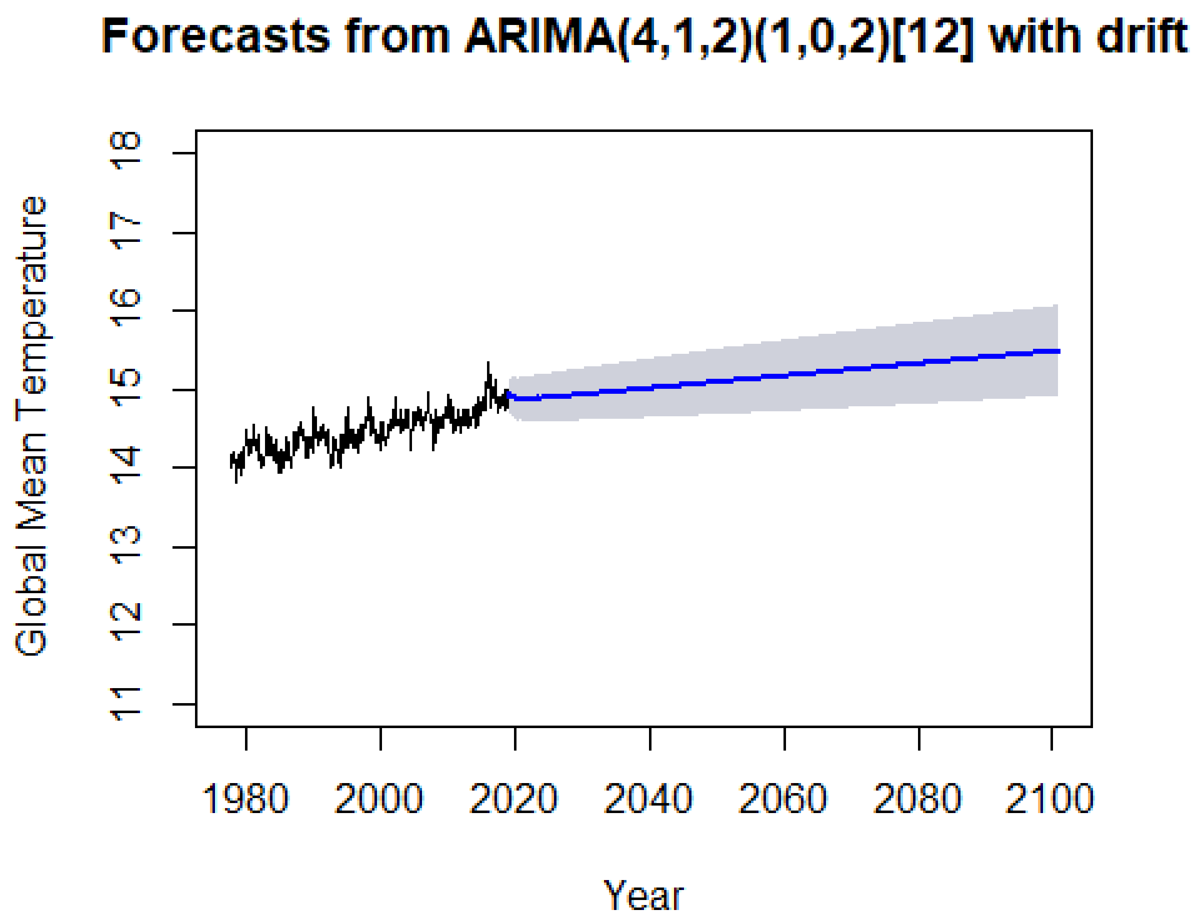

Global Mean Temperature 1880-2019
Average Temperature
15.0
14.5
14.0
13.5
1900
1950
2000
Year
Datasource: NASA Goddard Institute for Space Studies (GISTEMP)


Forecasts from ETS(A,Ad,A)
Global Mean Temperature
12
13
14
15
16
17
1900
1950
2000
2050

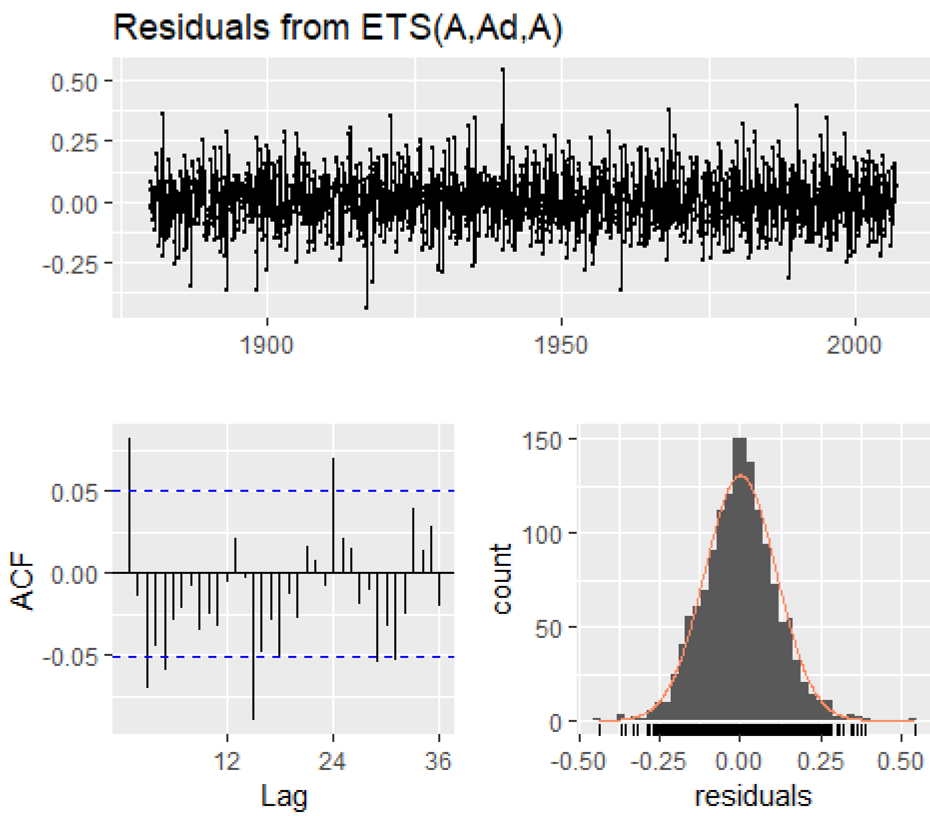

Forecasts from ETS(A,N,N)
Global Mean Temperature
12
13
14
15
16
17
1900
1950
2000
2050

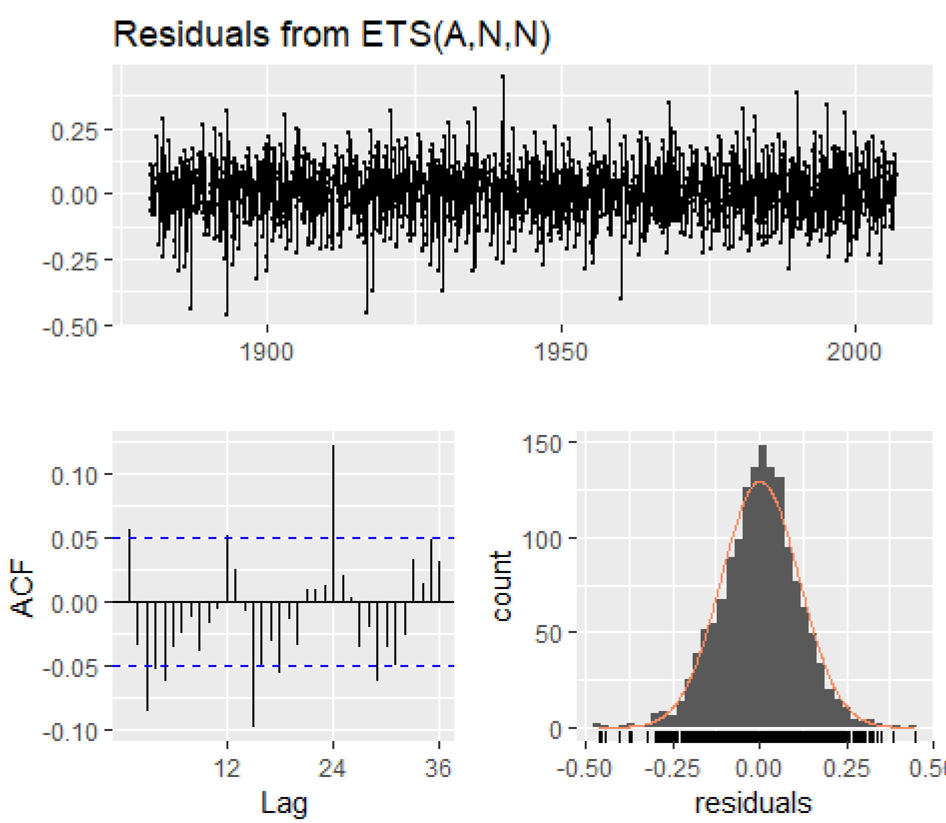

Forecasts from ETS(M,Md,M)
Global Mean Temperature
12
13
14
15
16
17
18
1900
1950
2000
2050
Year

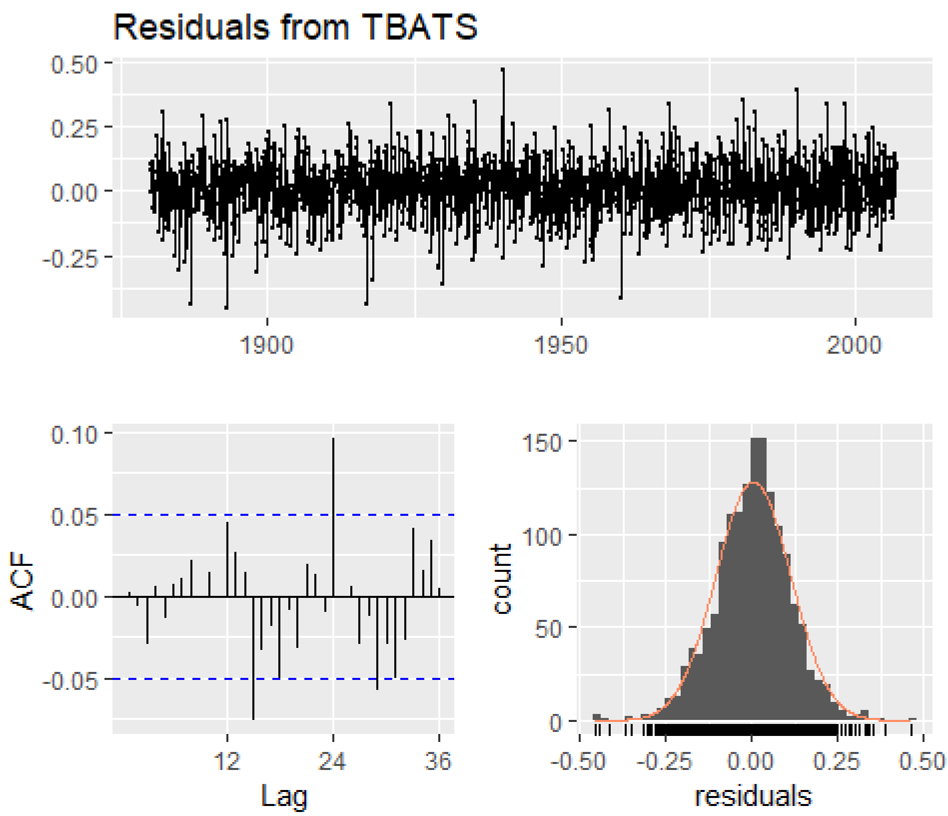

Forecasts from TBATS(1, {0,0}, 0.843, {<12,3>})
Global Mean Temperature
12
13
14
15
16
17
18
1900
1950
2000
2050
Year

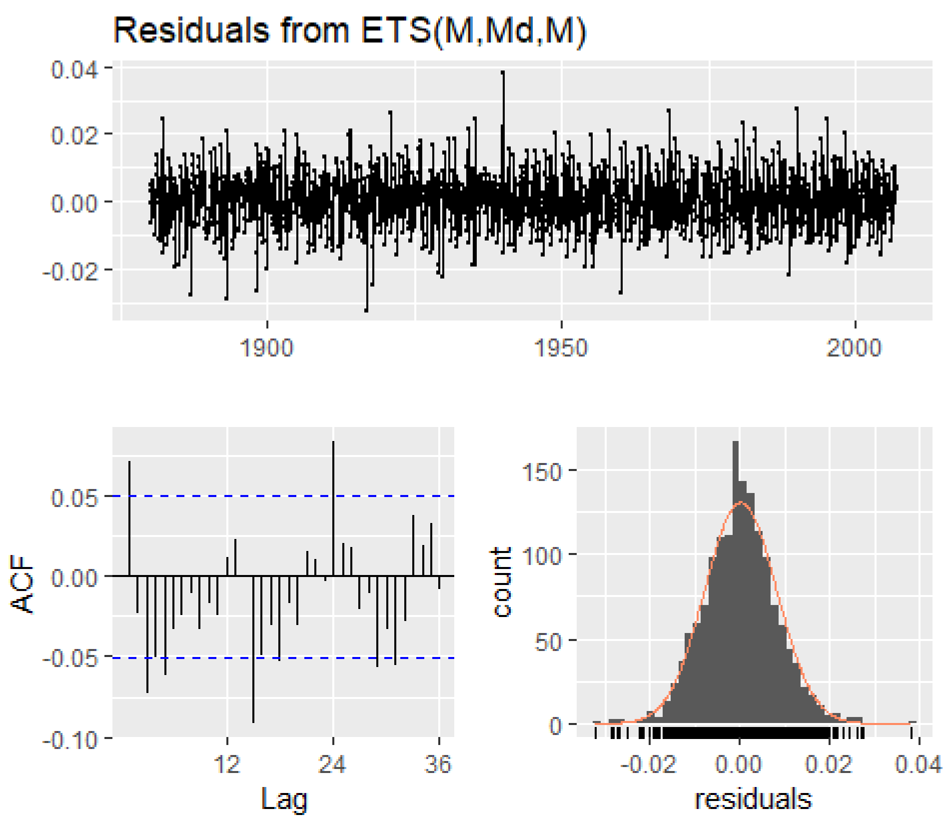


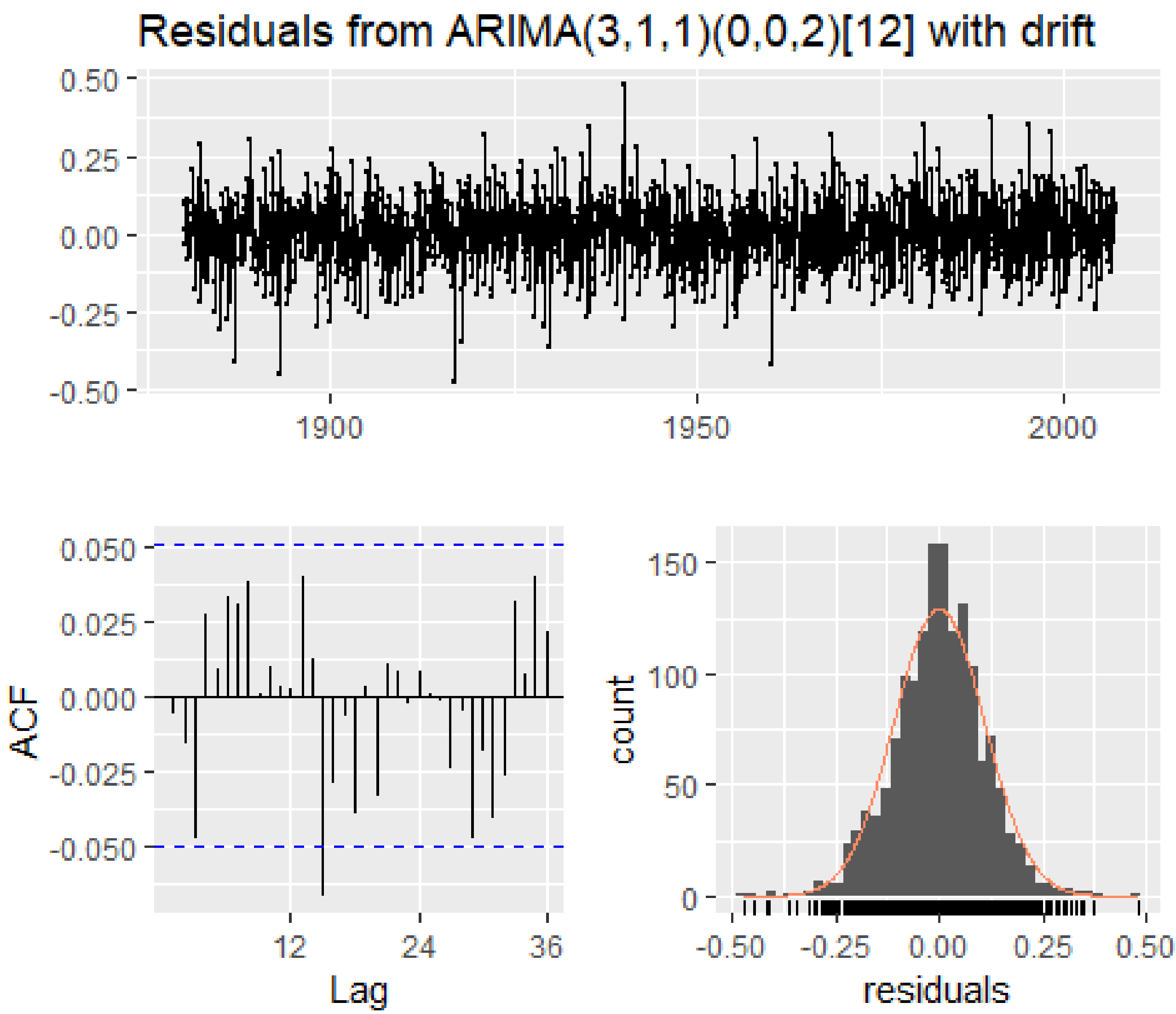
Forecasts from ARIMA(4,1,2)(1,0,2)[12] with drift
Global Mean Temperature
11
12
13
14
15
16
17
18
1980
2000
2020
2040
2060
2080
2100
Year


Residuals from ETS(A,Ad,A)
0.50
0.25
0.00
-0.25
1900
1950
2000
ACF
0.05
0.00
-0.05
12
24
36
Lag
count
150
100
50
0
-0.50
-0.25
0.00
0.25
0.50
residuals


Residuals from ETS(A,N,N)
0.25
0.00
-0.25
-0.50
1900
1950
2000
ACF
0.10
0.05
0.00
-0.05
-0.10
12
24
36
Lag
count
150
100
50
0
-0.50
-0.25
0.00
0.25
residuals

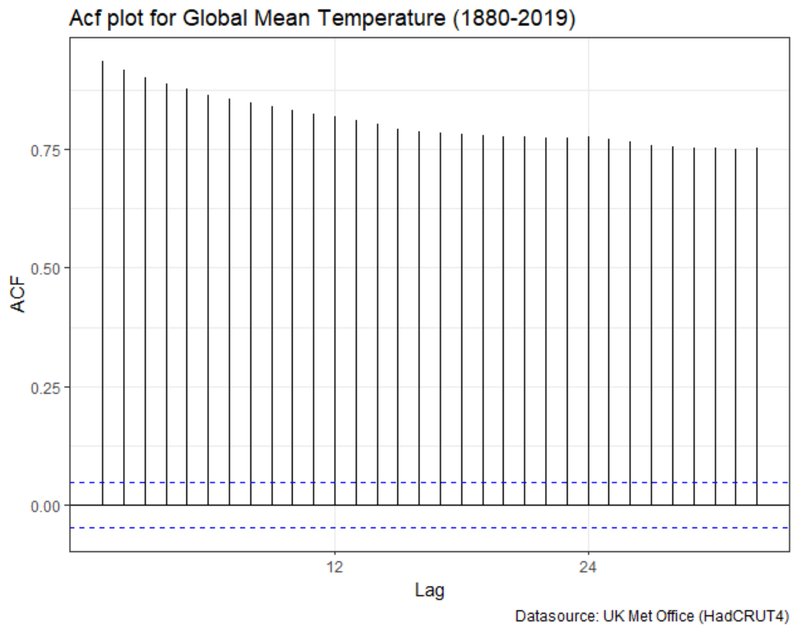

Residuals from TBATS
0.50
0.25
0.00
-0.25
1900
1950
2000
ACF
0.10
0.05
0.00
-0.05
12
24
36
Lag
count
150
100
50
0
-0.50
-0.25
0.00
0.25
0.50
residuals


Residuals from ETS(M,Md,M)
0.04
0.02
0.00
-0.02
1900
1950
2000
ACF
0.05
0.00
-0.05
-0.10
12
24
36
Lag
count
150
100
50
0
-0.02
0.00
0.02
0.04
residuals

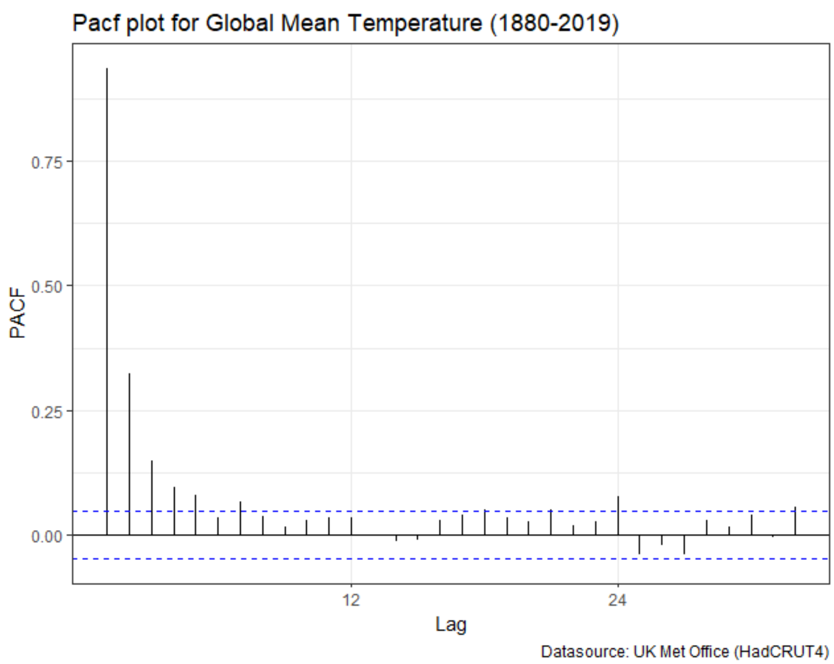


Residuals from ARIMA(3,1,1)(0,0,2)[12] with drift
0.50
0.25
0.00
-0.25
-0.50
1900
1950
2000
0.050
0.025
0.000
-0.025
-0.050
ACF
12
24
36
Lag
150
100
50
0
count
-0.50
-0.25
0.00
0.25
0.50
residuals

UKMET Model

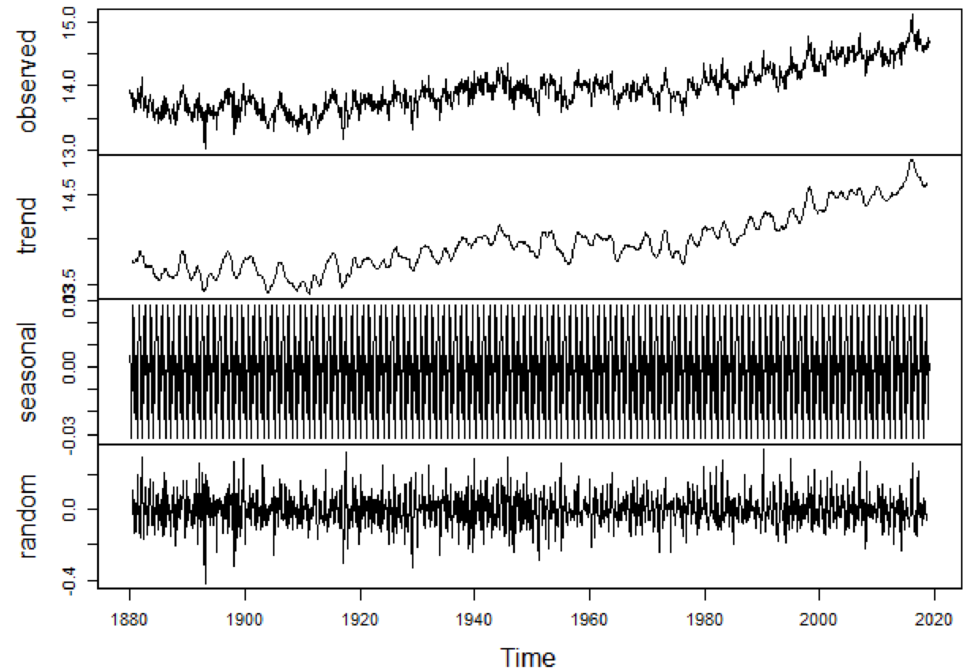


Acf plot for Global Mean Temperature (1880-2019)
ACF
0.75
0.50
0.25
0.00
12
24
Lag
Datasource: UK Met Office (HadCRUT4)

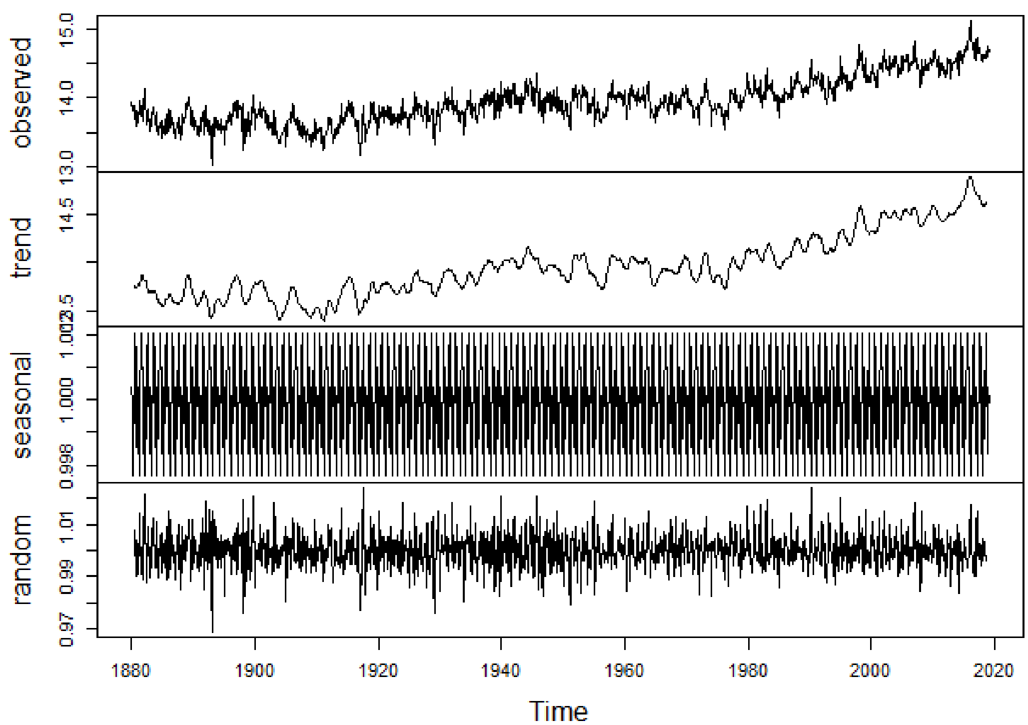


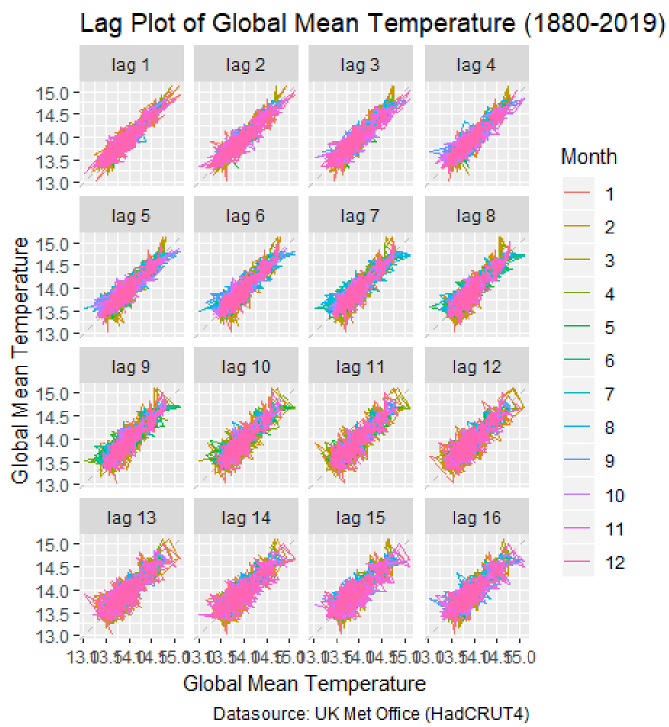
Pacf plot for Global Mean Temperature (1880-2019)
PACF
0.75
0.50
0.25
0.00
12
24
Lag
Datasource: UK Met Office (HadCRUT4)

Decomposition of additive time series

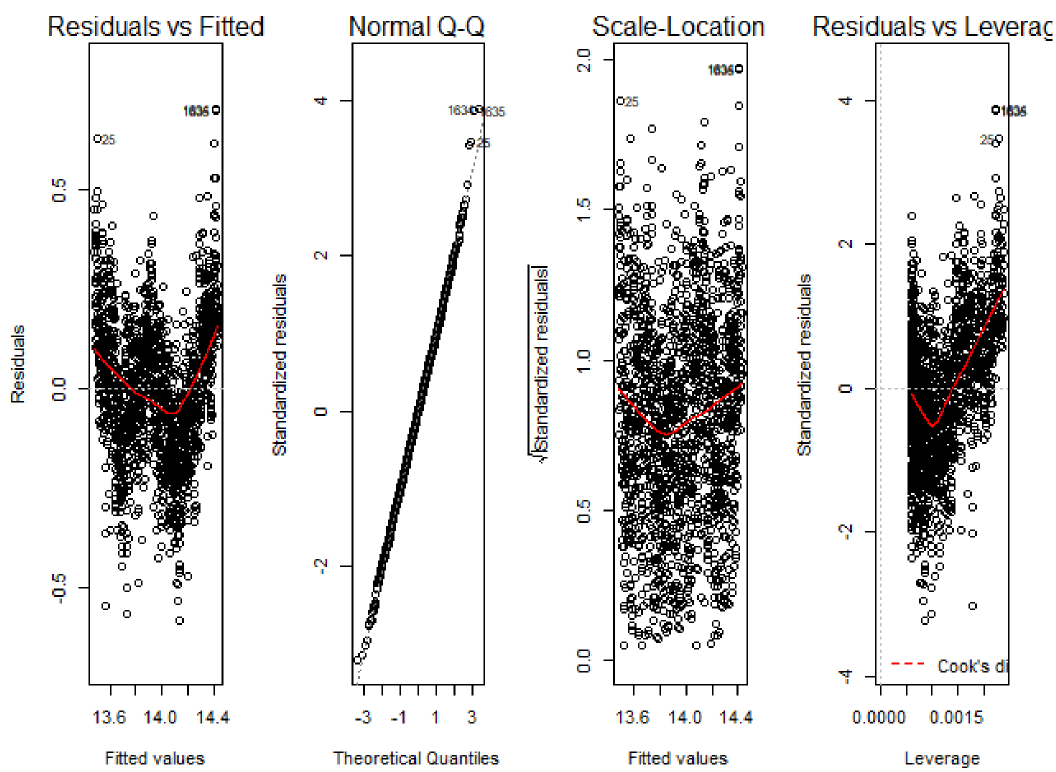


Decomposition of multiplicative time series

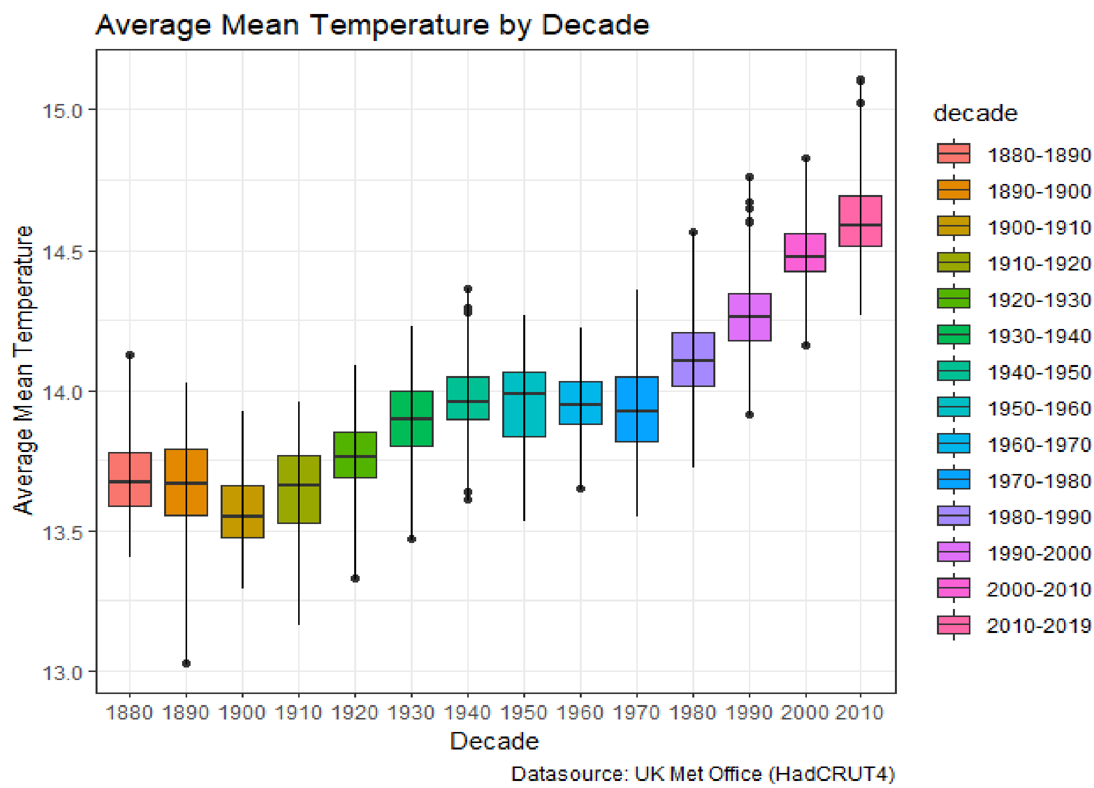


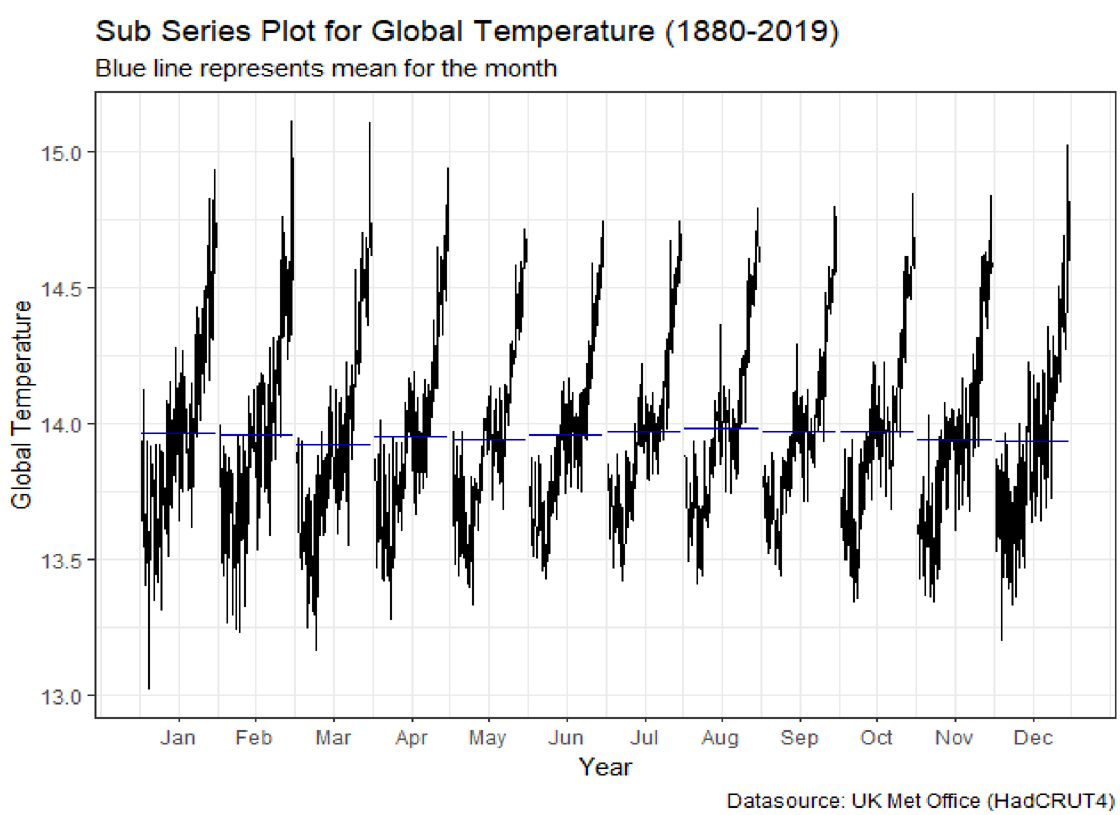


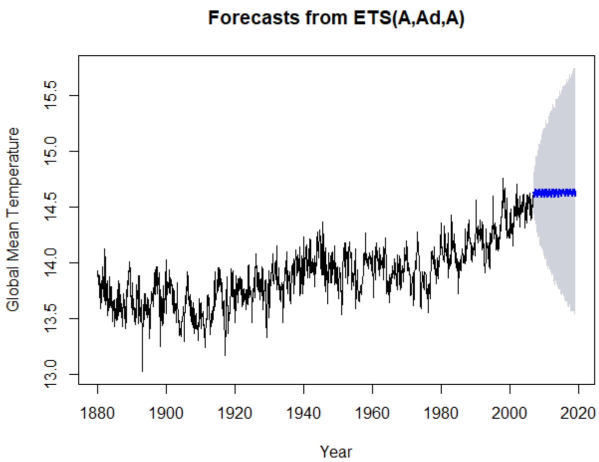
Lag Plot of Global Mean Temperature (1880-2019)
lag 1
lag 2
lag 3
lag 4
lag 5
lag 6
lag 7
lag 8
lag 9
lag 10
lag 11
lag 12
lag 13
lag 14
lag 15
lag 16
Global Mean Temperature
Month
1
2
3
4
5
6
7
8
9
10
11
12
Datasource: UK Met Office (HadCRUT4)

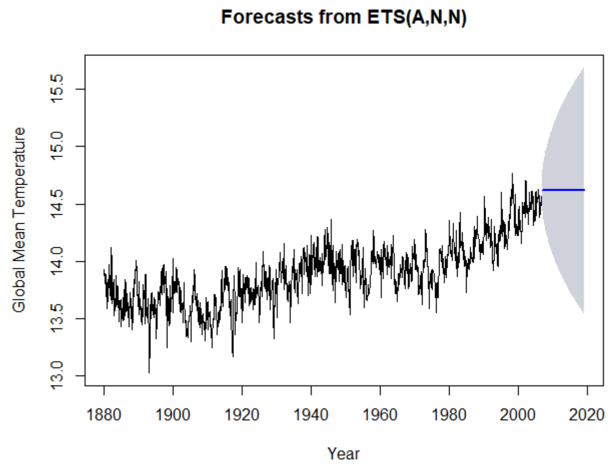


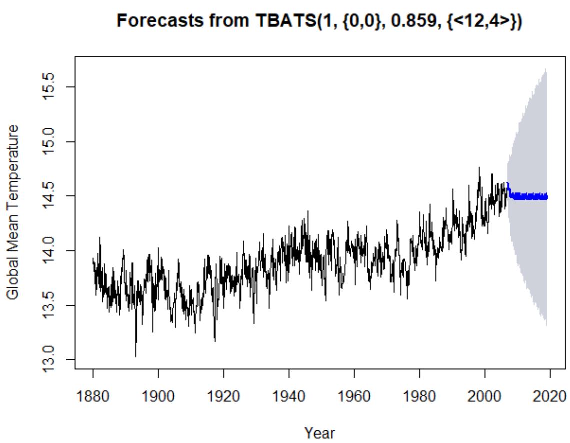
Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverag
Standardized residuals
Leverage
Cook's di

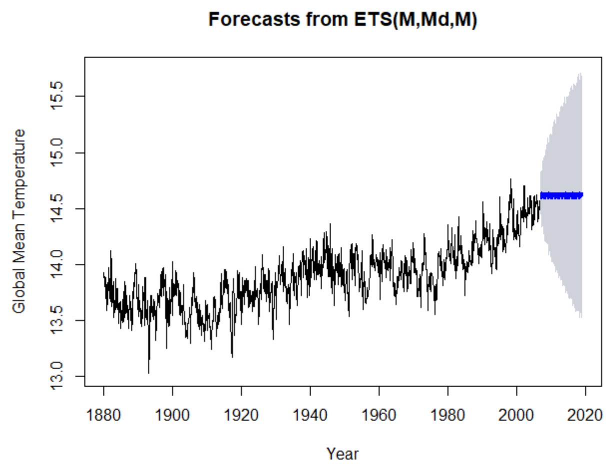

Average Mean Temperature by Decade
Average Mean Temperature
15.0
14.5
14.0
13.5
13.0
1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010
Decade
decade
1880-1890
1890-1900
1900-1910
1910-1920
1920-1930
1930-1940
1940-1950
1950-1960
1960-1970
1970-1980
1980-1990
1990-2000
2000-2010
2010-2019
Datasource: UK Met Office (HadCRUT4)

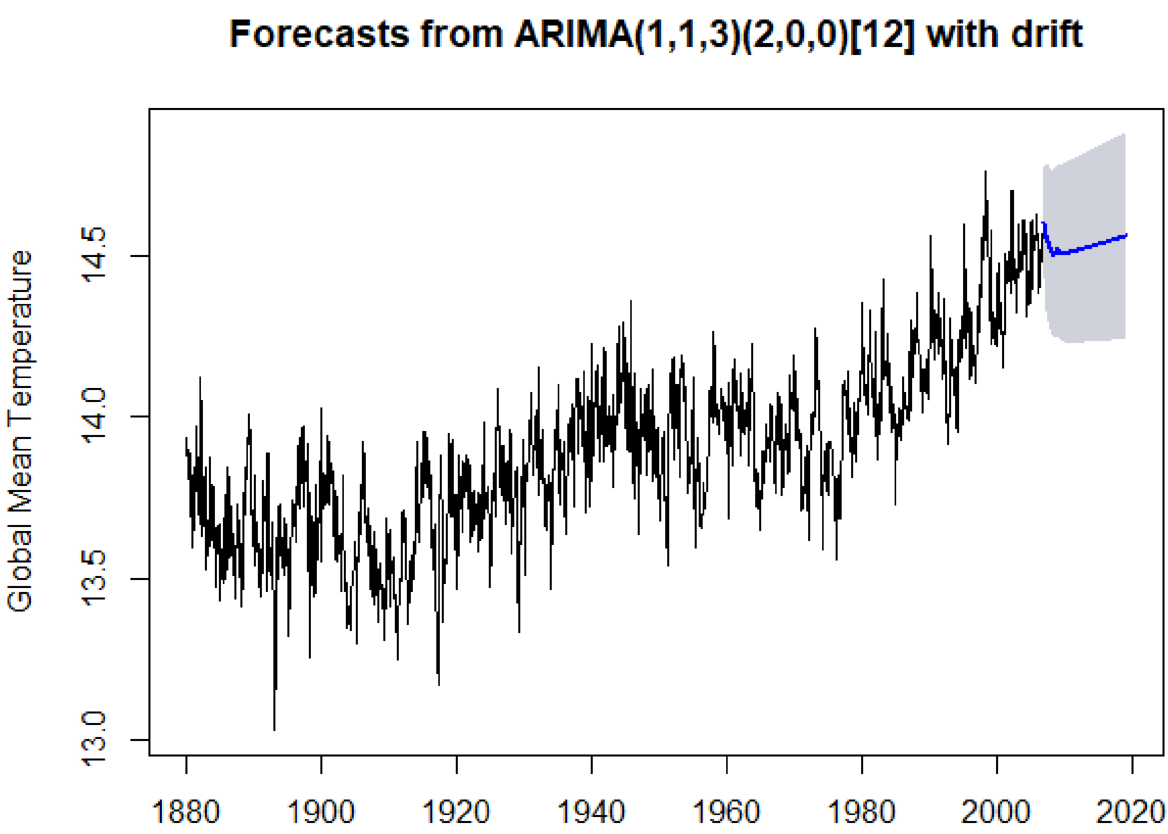

Sub Series Plot for Global Temperature (1880-2019)
Blue line represents mean for the month
Global Temperature
15.0
14.5
14.0
13.5
13.0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Year
Datasource: UK Met Office (HadCRUT4)

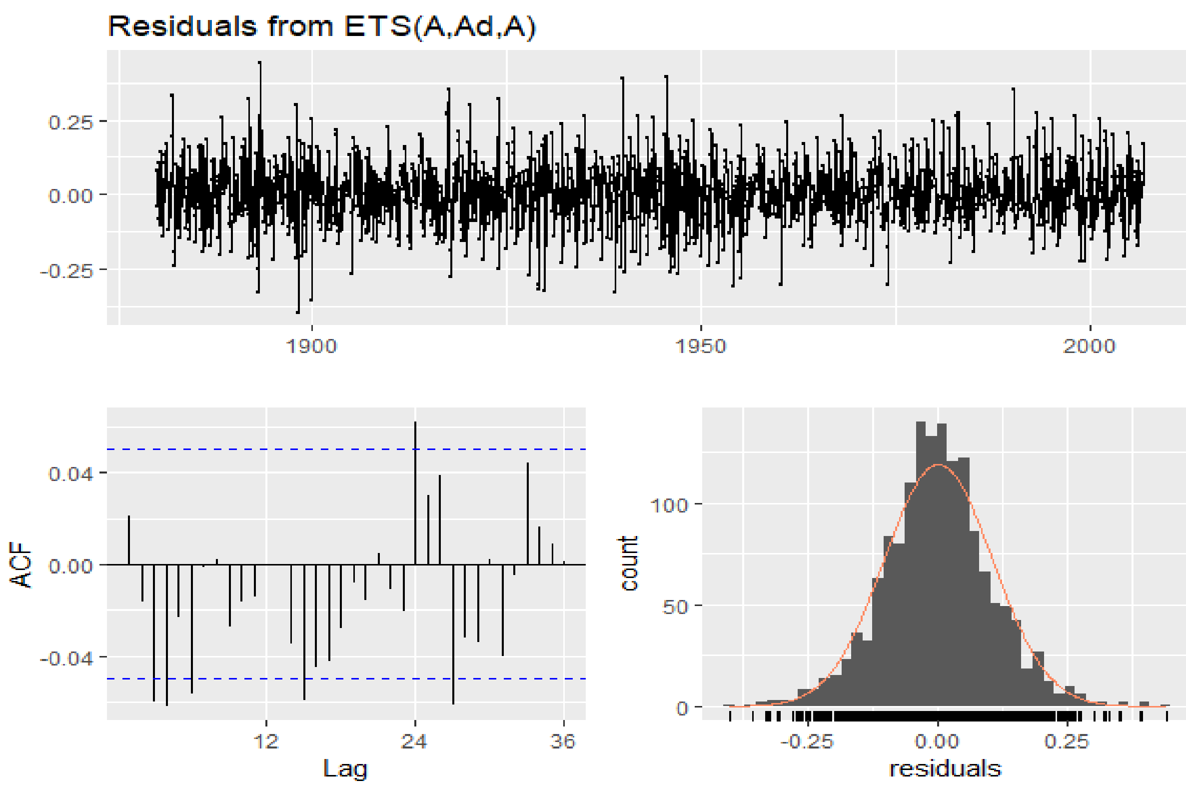

Forecasts from ETS(A,Ad,A)
Global Mean Temperature
Year
13.0
13.5
14.0
14.5
15.0
15.5
1880
1900
1920
1940
1960
1980
2000
2020


Forecasts from ETS(A,N,N)
Global Mean Temperature
Year
13.0
13.5
14.0
14.5
15.0
15.5
1880
1900
1920
1940
1960
1980
2000
2020

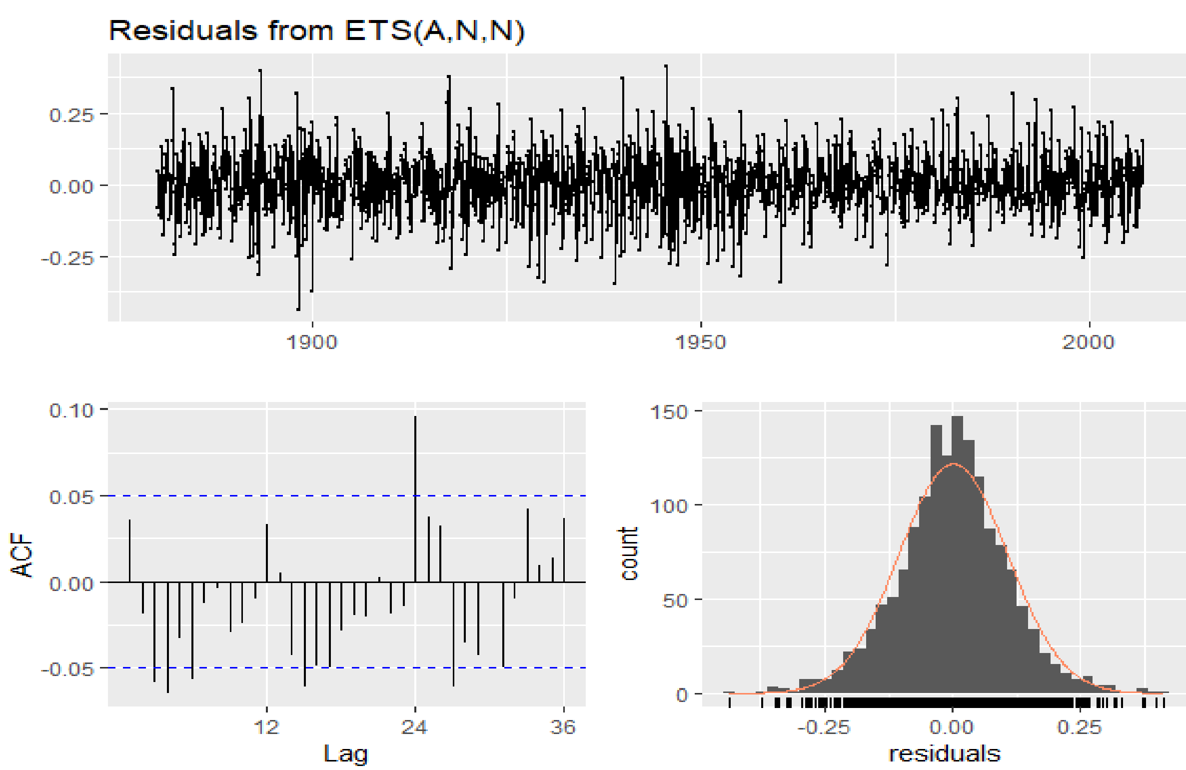

Forecasts from TBATS(1, {0,0}, 0.859, {<12,4>})
Global Mean Temperature
Year
13.0
13.5
14.0
14.5
15.0
15.5
1880
1900
1920
1940
1960
1980
2000
2020


Forecasts from ETS(M,Md,M)
Global Mean Temperature
Year
13.0
13.5
14.0
14.5
15.0
15.5
1880
1900
1920
1940
1960
1980
2000
2020


Forecasts from ARIMA(1,1,3)(2,0,0)[12] with drift
Global Mean Temperature
13.0
13.5
14.0
14.5
1880
1900
1920
1940
1960
1980
2000
2020

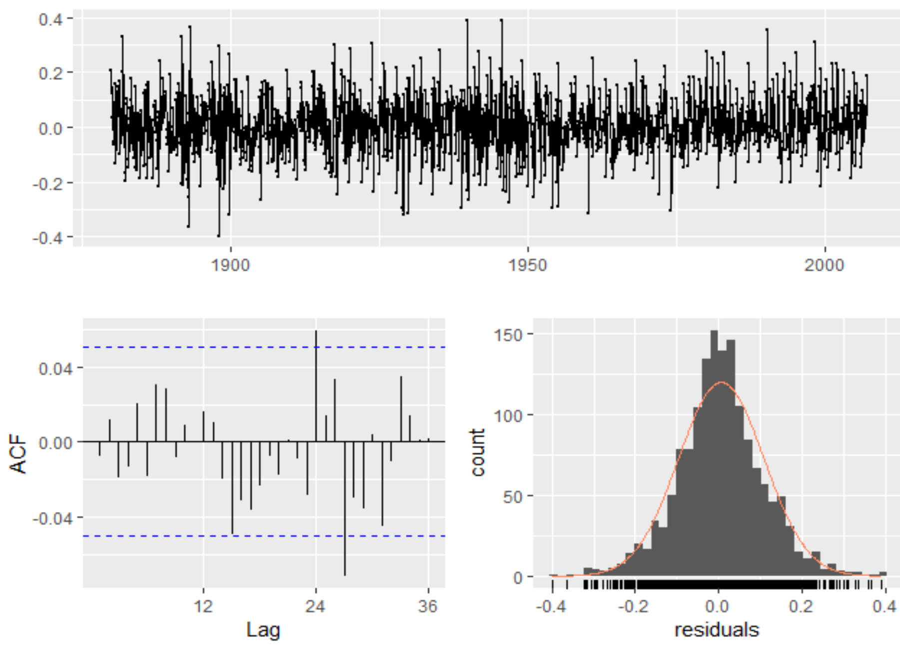


Residuals from ETS(A,Ad,A)
0.25
0.00
-0.25
1900
1950
2000
ACF
0.04
0.00
-0.04
12
24
36
Lag
count
100
50
0
-0.25
0.00
0.25
residuals

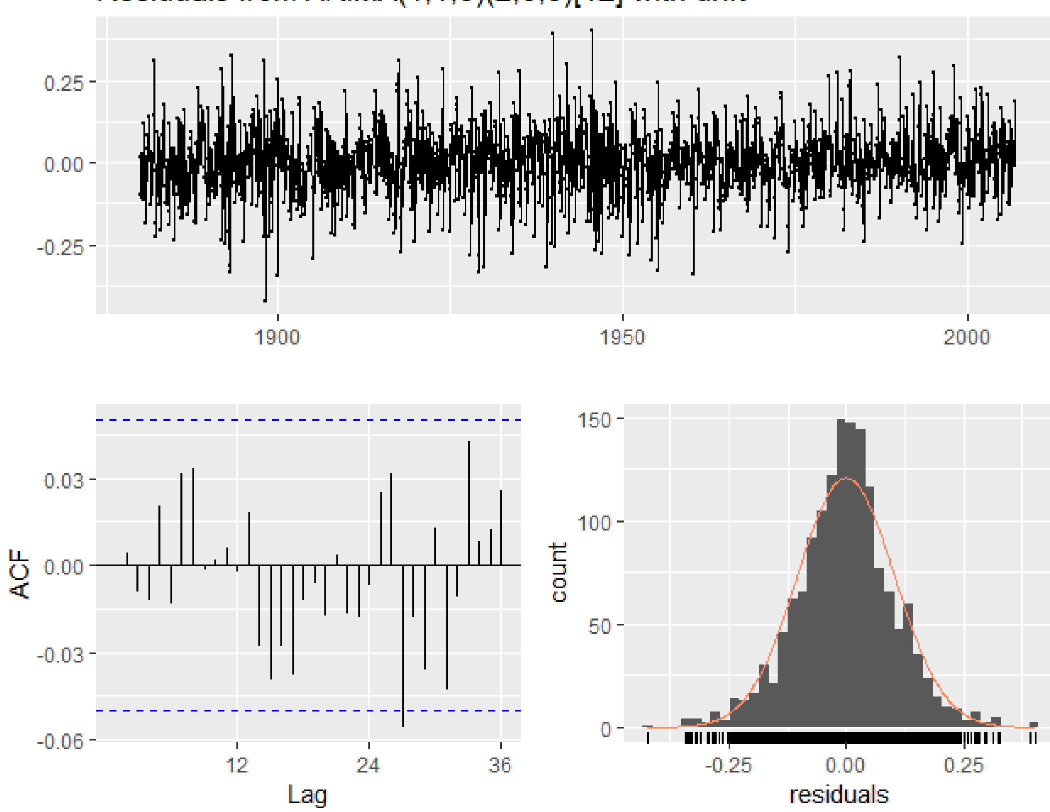


Residuals from ETS(A,N,N)
0.25
0.00
-0.25
1900
1950
2000
ACF
0.10
0.05
0.00
-0.05
12
24
36
Lag
count
150
100
50
0
-0.25
0.00
0.25
residuals

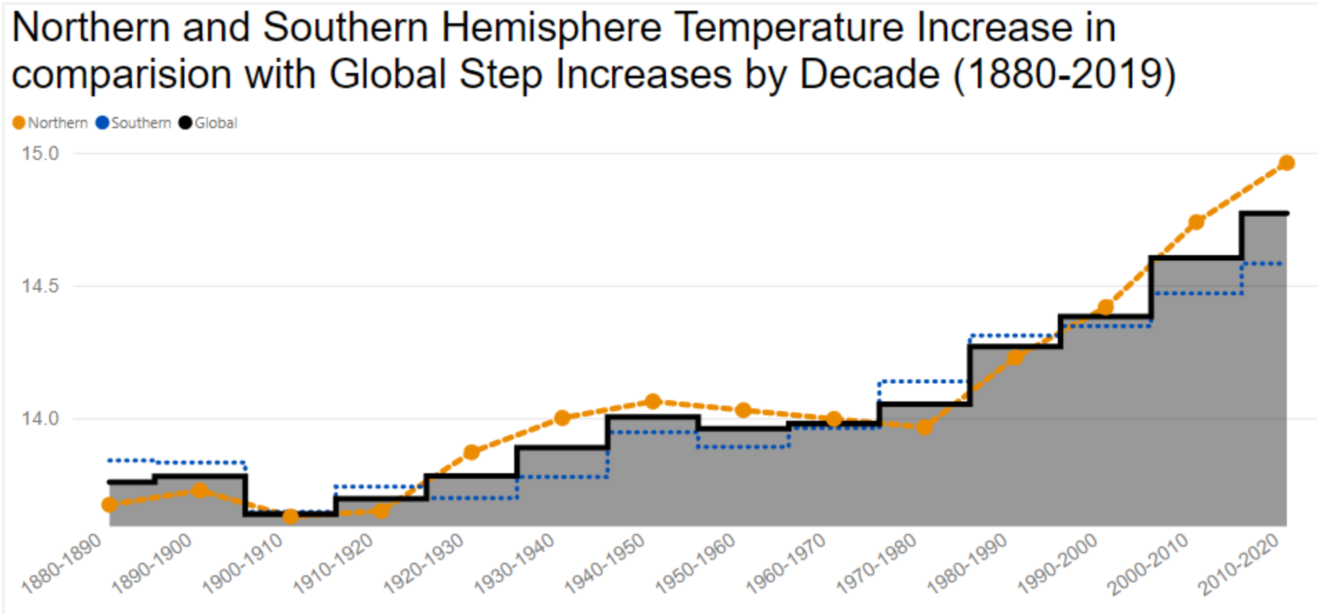
Residuals from TBATS



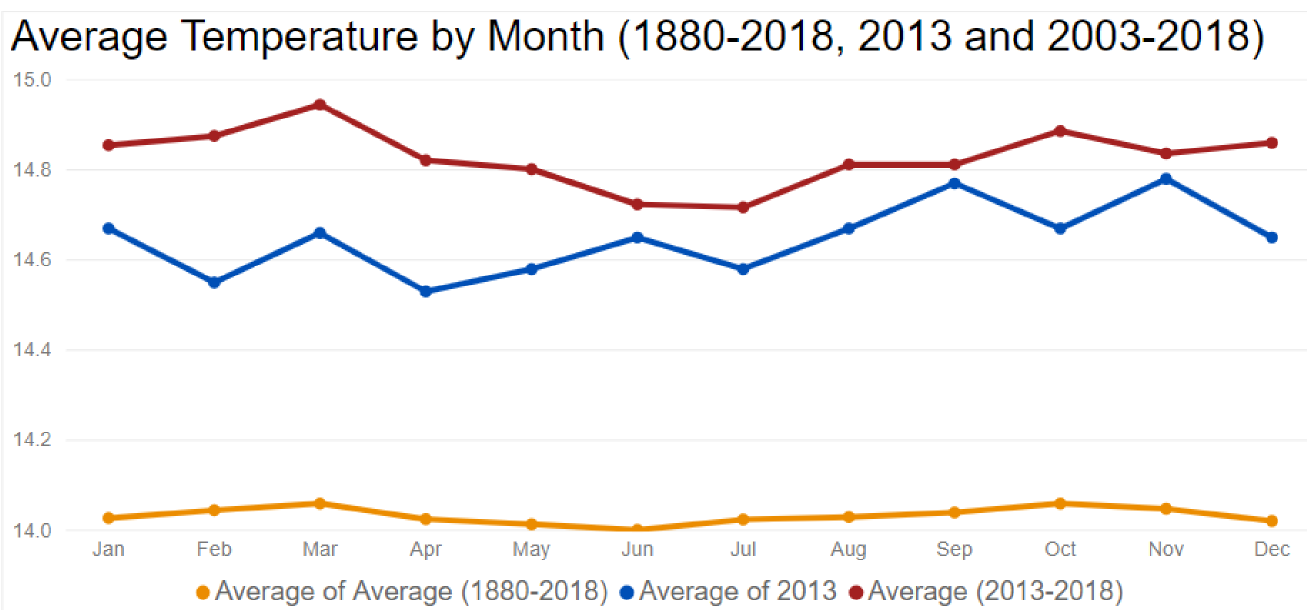
Residuals from ARIMA(1,1,3)(2,0,0)[12] with drift



Visuals from EDA (not shown in main document)



The average increase in temperature after 1980 in Northern Hemisphere is 0.25 °C each decade compared to 0.10 °C in Southern Hemisphere.



Average temperature by Month for 2013 and Average of 2013 to 2018, compared to Average of (1880-2018) is significantly higher.