Temperature Anomalies in Global Climate Change

The movie *"The Revenant"* helped Alejandro G. Iñárritu win his second Academy award for Best directing. To film his western survival drama, Alejandro and his production team have to travel to the southern tip of Argentina just to find some snow to shoot their final battle scene. Climate change is real and global warming is serious, it is the most urgent threat facing the entire human race. Yet, some people still consider global warming as a joke or even some conspiracies from the big corporations. In this project, our team want to help address this problem by analyzing historical climate temperature anomalies and assessing different models in forecasting temperatures.

Data Description

1. What is temperature anomaly?

Scientists usually combine air temperatures above land and ocean surface temperatures collected by ships, satellites and buoys to graph a full picture of Earth's temperature. Each daily temperature from both land and sea will then be used to compare with the normal value of that site in the same period of time, which usually are long-term average over 30-year period. The difference between the actual temperature and reference value is called the temperature anomaly.

Temperature anomaly is an important criterion in climate study for it is often used to evaluate how temperature varies over time. Intuitively, a positive anomaly means warmer temperatures than the long-term expectation, and a negative anomaly indicate that the observed temperature is lower than the reference value. Scientists then calculate the average temperature anomaly of each month, which further used to get the seasonal and annual anomalies.

2. Selection of data set

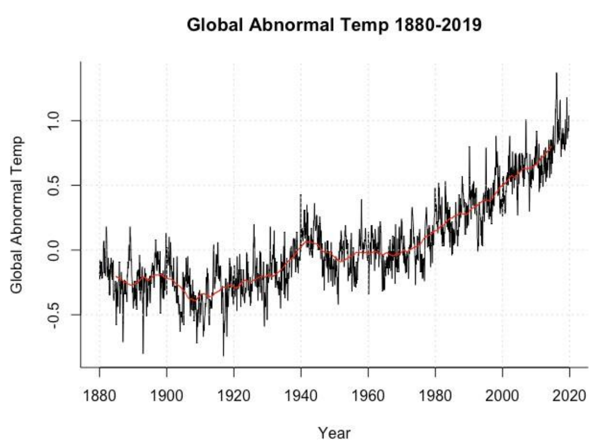
There are four major data sets used by meteorologists to study global temperatures, HadCRUT by Hadley Centre located in Exeter, UK, JMA by the Japanese Meteorological Agency, GISTEMP by NASA's Goddard Institute for Space Science, and MLOST by NOAA (National Oceanic and Atmospheric Administration). We choose GISTEMP's Combined Land-Surface Air and Sea-Surface Water Temperature

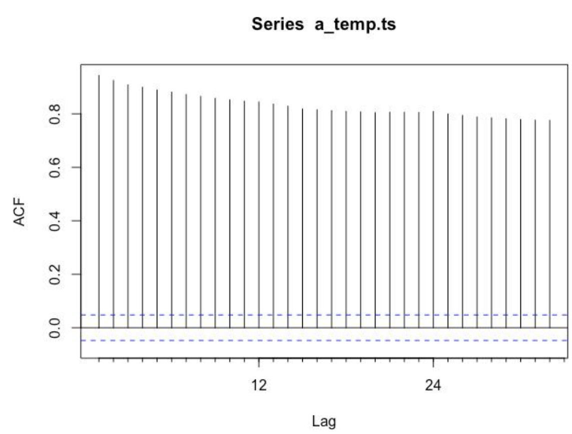
Anomalies (Land-Ocean Temperature Index, LOTI) with global-mean monthly, seasonal, and annual means from 1880 to 2019. For this data set "GISTEMP.csv", the index is 0.01 degrees Celcius; besides Year, there are 18 columns in the dataset, the first 12 columns are the 12 months, the next two columns (J-D and D-N) are annul means, while the last four columns are seasonal means. We select GISTEMP because it has the widest coverage that measuring 99% of the world, thus it does an excellent job on handling the remote part of the world.

3. Data pre-processing

Before we imported the original data file into R, we streamlined the data by preserving only the monthly temperature anomalies, eliminating all other non-necessary numbers. There are only two columns in "abnomral_temp" file. One is year index starting from 1 to 1678, representing Jan 1880 to Oct 2019. The other is monthly temperature anomalies in 0.001 degrees Celcius.

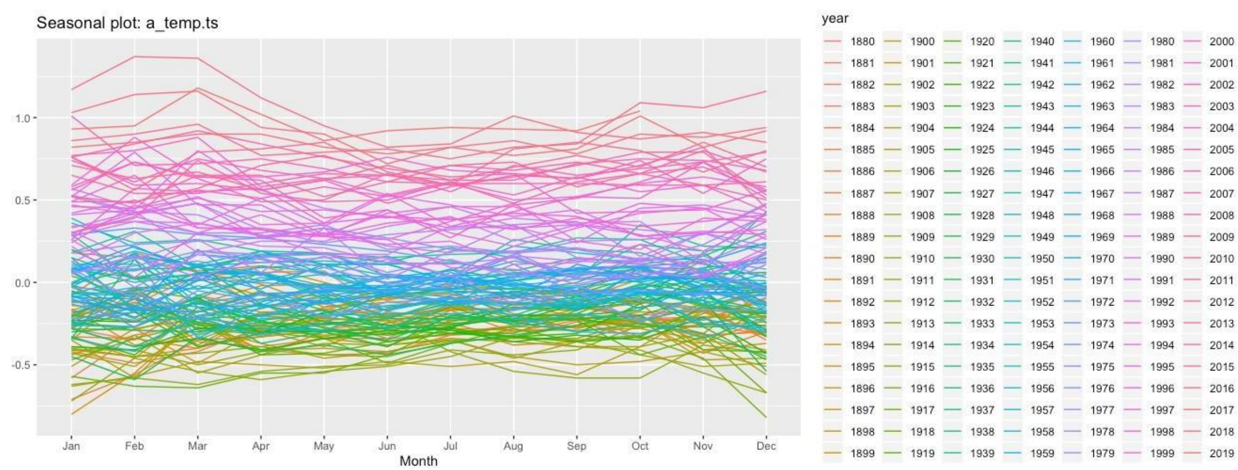
Data Visualization

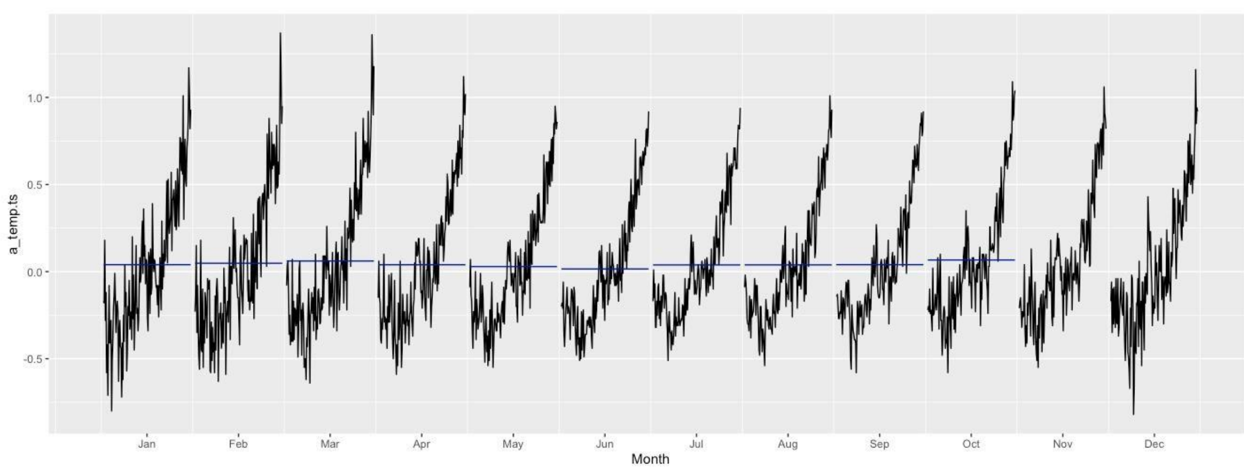




By plotting the line chart and autocorrelation, we can see that the level of temperature anomalies ranges from -1 to 1.5 degrees Celcius. Intuitively, we would rather have a negative temperature anomaly than positive, because more ice doesn't do much harm but when ice begins melting, the whole ecosystem will be affected. In the line chart, there is a clear turning point showing that after year 1980, temperature anomalies are all positive. It implies an upward trend for global temperature and the earth is losing its temperature-restoring power. To observe the trend in detail, we use a mid-point MA filter with 10-year rolling window to visualize the trend. As shown by the red line in plot, we can conclude that the abnormal temperature was stable before 1900 but decreased slightly till 1910. Then the data has an overall upward trend at a steadily increased rate till present, except the period from 1930 to 1950. During this special period, the abnormal temperature increased at a higher speed in prior ten years but decreased a bit in later ten years.

The autocorrelation graph indicates very high positive correlation even when lag is greater than 33, which means that the global temperature follows a positive feedback loop --- a high increase in global temperature in the past years is likely to result in a continuous increase in the future years. This is really worrying as we just find out that the global climate is not a self-stabilizing system; any small effect on temperature will be exaggerated, causing continuous greater change in the future.





To explore the seasonality, we use 'ggseasonplot()' and 'ggsubseriesplot()'. From the plots above, we do not find distinct seasonality. The pattern in sub series plot may look like seasonality at first glance, but it is actually the result of a large amount of data points condensed together. They have the same long-term trend but their shapes are not matching. This can be partially explained by the autocorrelation graph because it shows that the future value is correlated with every month's data in the past, not just seasonal. It is an interesting difference between temperature and temperature anomaly; we know that temperature has seasonality for sure but when it comes to temperature anomaly, all previous months are important.

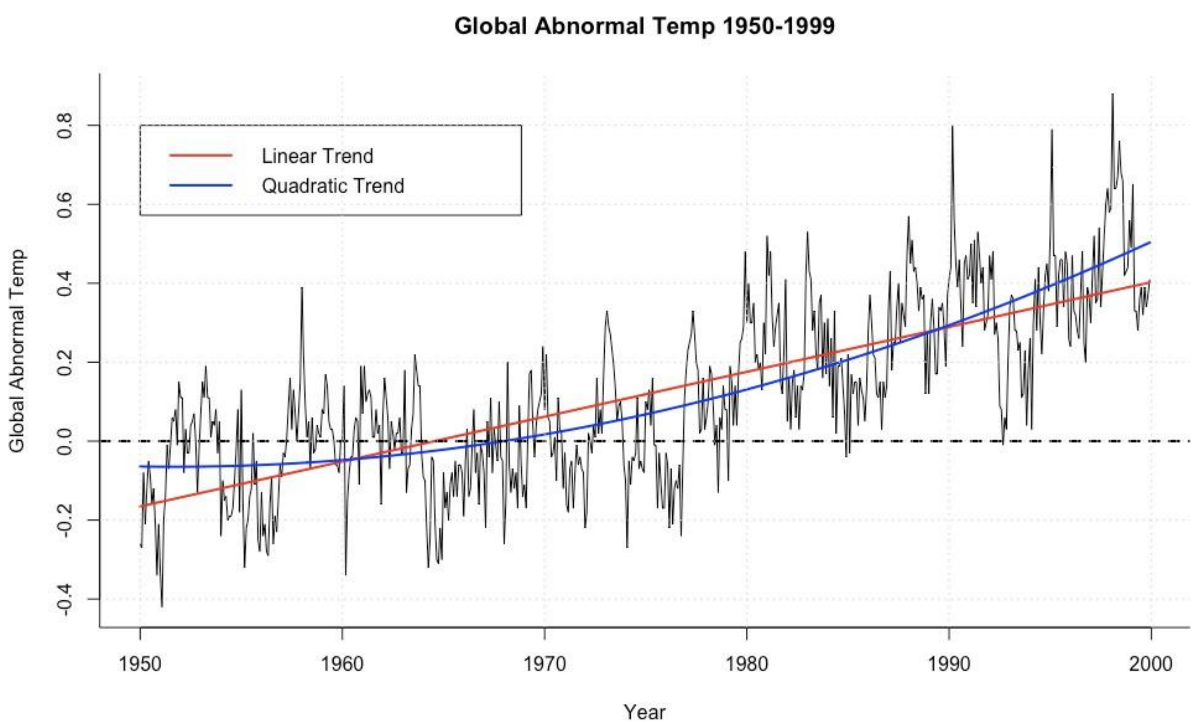
Model Comparison And Interpretation

We have put all of our thinking process and coding logic in the R file as comments so that we can explain our code step by step. Therefore, it is highly recommended to go through our R file before reading the following sections.

Our analysis focuses on the recent warming period starting from 1950. Specifically, the training set is from 1950 to 1999 and the validation set is from 2000 to October 2019.

1. Linear and quadratic models

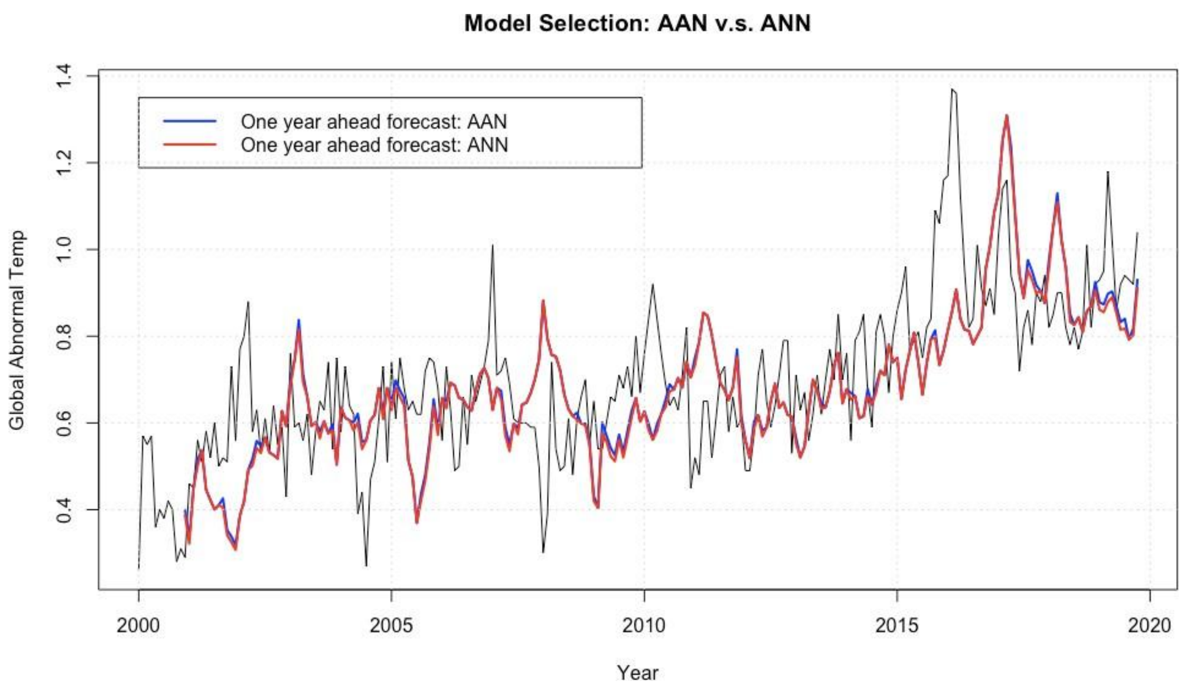
We start with fitting the sample linear and quadratic trend onto the training data.



The fitted linear and quadratic lines clearly show an uptrend in abnormal temperatures. We use these two linear models to forecast in the validation period, and compare the results with the actual values to calculate residuals - Diebold/Mariano test shows the quadratic trend model statistically fits the validation data better, though the difference is not big.[1] The R-squares in validation are respectively -0.58 and 0.32 with linear model and quadratic model, which further indicates the quadratic model is better than linear.

2. Exponential filters

As discussed in the data visualization section, an uptrend is observed in the data but the seasonality is not distinct, thus a Holt filter ("AAN") is the best filter based on our observation. However, the optimal model generated by R is actually an "ANN". To compare the two models, we conduct a recursive, one year ahead forecast to evaluate their performance in the validation period. It turns out that the Holt filter does perform better in the validation period in accuracy statistics and D/M test. (Appendix 1)



3. ARMA and ARIMA

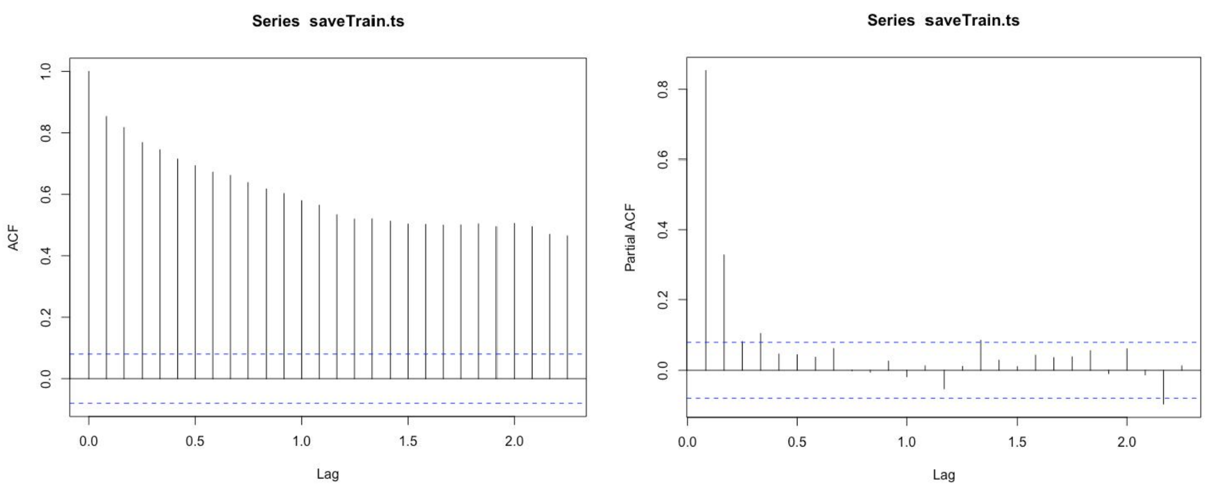
We first use Dicky/Fuller test to evaluate the existence of unit root. D/F test 1 is rejected at 5% and test 3 is strongly rejected at 1%[2]. This partially contradictory test result suggests 1) our data is stationary (no unit

[1] DM = 6.1234, p-value = 3.773e-09

[2] Test 1: t-statistics -2.2756; critical value -2.58 at 1%, -1.95 at 5%
Test 3: t-statistics for phi3 34.1671; critical value 8.27 at 1%

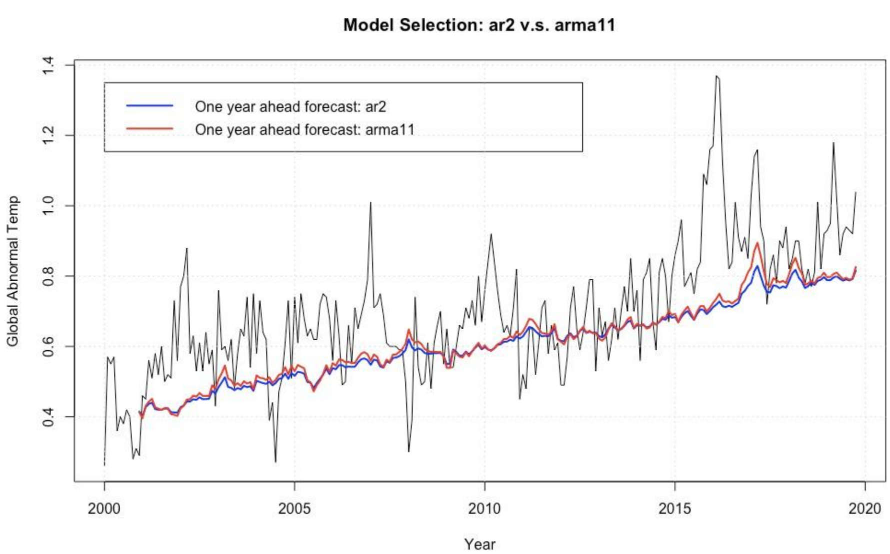
root), thus the “I” part in ARIMA should be set to 0 and 2) the existence of trend is ambiguous. Based on our observation from the plots, we believe our data has a trend, therefore all models in the following discussion include trends.

In the training period, the ACF graph decays slowly while the PACF drops to 0 after two monthly lags. This indicates AR(2) is likely to be our best choice.



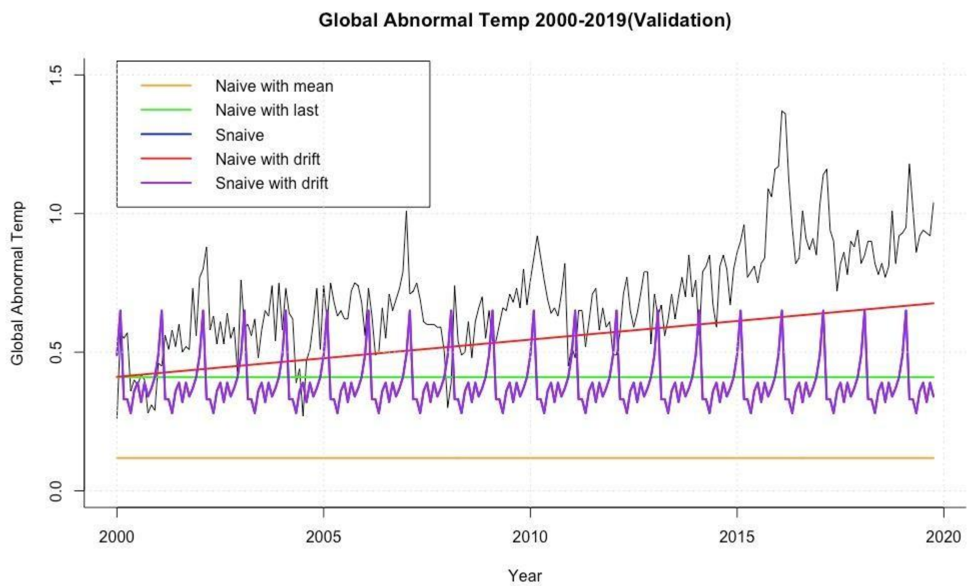
R’s optimal parameter selection is different. By executing the auto.arima function (with d=0 and trend), R generates an ARMA(1,1). This optimal output is the same regardless we set seasonal to true or false, which suggests the seasonal pattern is not a main concern in the training data.

We again conduct a recursive, one year ahead forecast to compare the two models. Both accuracy statistics and D/M test on residuals show ARMA(1,1) with trend is the better model.



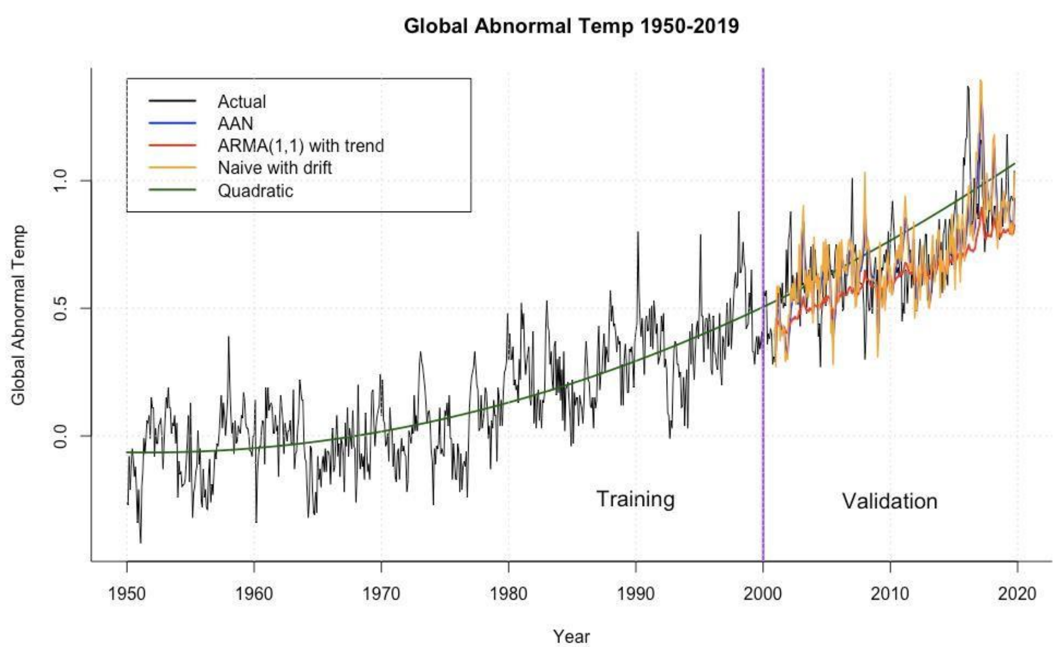
4. Naive models

Naive model with drift is apparently the best naive model.

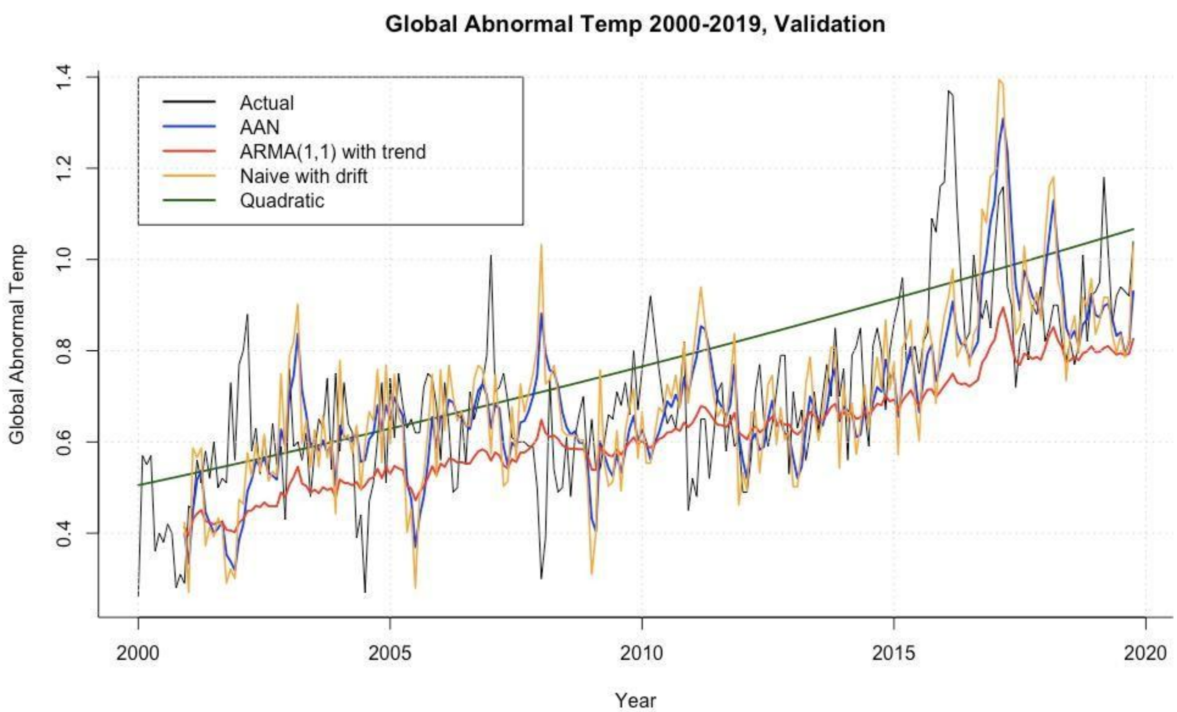


5. Final model selection

The following graph summarizes the best model from each category.



A closer look at the validation period:



From the accuracy statistics perspective, the AAN model is the best model to predict abnormal temperatures, while the ARMA(1,1) with trend is also a powerful candidate. The AAN model is slightly better than the ARMA model, but the difference is small. The RMSEs for AAN, ARMA, Naive and Quadratic are respectively 0.161, 0.163, 0.176 and 0.738. Other accuracy statistics such as RAE indicate the same result. (Appendix 3) From the previous analysis we know that our data contains strong autocorrelation patterns, meaning the value of a data point is heavily depending on previous data points. The AAN model focuses on more recent data, therefore it can be useful in short-term predictions. In fact, the alpha in the AAN model is 0.52, so the recent data points are significant to short-term predictions and this is one of the reasons why the AAN model has such a small RMSE.

From the D/M test perspective, the Quadratic model is the best. The D/M tests on residuals show the Quadratic model is better than the ARMA model and the ANN model, which is further better than the Naive model. (Appendix 3) This is a surprising result as the Quadratic model is a relatively “simple” model. One of the benefits of using a quadratic model is its inter-predictability. For example, our model can be written as:

$$Global\ Temperature\ Anomaly = -0.064 - 0.0000717t + 0.0000017t^2$$

$$,where\ t = 1\ for\ Jan\ 1950;\ t = 2\ for\ Feb\ 1950, \ldots \ldots$$

Because the model is so simple, even high school students can interpret it and understand how serious the situation is --- as time goes by, the abnormal temperatures are getting higher and higher. Besides, since quadratic models do not only focus on the most recent data, they are useful in catching long-term trends.

Summary

The two seemingly contradictory results show the strengths and limitations of our model evaluation tools. We believe there is no distinctive answer to which one is dominating the other, and the final model selection should depend on the goal of the research. **If the research aims to predict short-term levels**, such as forecasting whether next year's temperature will be favorable to agriculture, glaciers or coral reefs, the AAN model (alpha = 0.52, beta = 0.0005) can help generate an accurate prediction. Otherwise, **if the target is related to long-term trends or educational purposes**, such as supporting United Nations' climate plan or educating young kids, our grandparents or even presidents the importance of climate change, the fitted quadratic model is the best choice. Whichever model we choose contains a strong uptrend, which suggests global abnormal temperature is increasing at an accelerating speed. If the situation stays unchanged, then the climate issue will soon be out of control.

Appendix 1: exponential filter statistics

```
> print("Accuracy: AAN")
[1] "Accuracy: AAN"
> print(accuracy(saveValid.ts,fvec.aan.ts))
                 ME      RMSE      MAE       MPE    MAPE      ACF1 Theil's U
Test set -0.0255051 0.1607703 0.124295 -6.419477 19.3955 0.6628273  2.885659
> print("Accuracy: ANN")
[1] "Accuracy: ANN"
> print(accuracy(saveValid.ts,fvec.ann.ts))
                  ME      RMSE       MAE       MPE     MAPE      ACF1 Theil's U
Test set -0.03114319 0.1621313 0.1258912 -7.435259 19.92729 0.6633572  2.989217
> print("Diebold/Mariano")
[1] "Diebold/Mariano"
> print(dm.test(ferr.aan,ferr.ann))

        Diebold-Mariano Test

data:  ferr.aanferr.ann
DM = -2.3568, Forecast horizon = 1, Loss function power = 2, p-value = 0.01925
alternative hypothesis: two.sided
```

Appendix 2: ARMA statistics

```
> print("Accuracy: ar2")
[1] "Accuracy: ar2"
> print(accuracy(saveValid.ts,fvec.ar2.ts))
                 ME      RMSE       MAE       MPE     MAPE     ACF1 Theil's U
Test set -0.1022097 0.1690722 0.1315919 -17.28444 22.28166 0.600647  14.62399
> print("Accuracy: arma11")
[1] "Accuracy: arma11"
> print(accuracy(saveValid.ts,fvec.arma11.ts))
                  ME      RMSE       MAE       MPE     MAPE      ACF1 Theil's U
Test set -0.09301065 0.1632066 0.1259847 -15.62418 21.12239 0.5963838  10.23813
>
> print("Diebold/Mariano")
[1] "Diebold/Mariano"
> print(dm.test(ferr.ar2,ferr.arma11))

        Diebold-Mariano Test

data:  ferr.ar2ferr.arma11
DM = 5.5186, Forecast horizon = 1, Loss function power = 2, p-value = 8.939e-08
alternative hypothesis: two.sided
```

Appendix 3: final model comparison statistics

```
> print("Accuracy: AAN")
[1] "Accuracy: AAN"
> print(accuracy(saveValid.ts,fvec.aan.ts))
                 ME      RMSE      MAE       MPE    MAPE      ACF1 Theil's U
Test set -0.0255051 0.1607703 0.124295 -6.419477 19.3955 0.6628273  2.885659
> print("Accuracy: ARMA(1,1)")
[1] "Accuracy: ARMA(1,1)"
> print(accuracy(saveValid.ts,fvec.arma11.ts))
                  ME      RMSE       MAE       MPE     MAPE      ACF1 Theil's U
Test set -0.09301065 0.1632066 0.1259847 -15.62418 21.12239 0.5963838  10.23813
> print("Accuracy: Naive with Drift")
[1] "Accuracy: Naive with Drift"
> print(accuracy(saveValid.ts,fvec.naive.drift.ts))
                  ME      RMSE       MAE       MPE     MAPE      ACF1 Theil's U
Test set -0.01287961 0.1756587 0.1386359 -6.051118 21.82525 0.5301121  1.385068
> # Accuracy: AAN > ARMA(1,1) >> Naive with drift


> print(accuracy(saveValid.ts,ferr.q.ts))
                 ME      RMSE       MAE       MPE     MAPE      ACF1 Theil's U
Test set -0.7286455 0.7377157 0.7286455 -384.6683 3189.954 0.9850062  9.453806


> print(dm.test(ferr.aan,ferr.arma11))

        Diebold-Mariano Test

data:  ferr.aanferr.arma11
DM = -0.37456, Forecast horizon = 1, Loss function power = 2, p-value = 0.7083
alternative hypothesis: two.sided

> # No decisive conclusion
>
> print("Diebold/Mariano: AAN vs Naive with drift")
[1] "Diebold/Mariano: AAN vs Naive with drift"
> print(dm.test(ferr.aan,ferr.naive.drift))

        Diebold-Mariano Test

data:  ferr.aanferr.naive.drift
DM = -3.2291, Forecast horizon = 1, Loss function power = 2, p-value = 0.001418
alternative hypothesis: two.sided

> # AAN is better than naive with drift
>
> print("Diebold/Mariano: ARMA(1,1) vs Naive with drift")
[1] "Diebold/Mariano: ARMA(1,1) vs Naive with drift"
> print(dm.test(ferr.arma11,ferr.naive.drift))

        Diebold-Mariano Test

data:  ferr.arma11ferr.naive.drift
DM = -1.3653, Forecast horizon = 1, Loss function power = 2, p-value = 0.1735
alternative hypothesis: two.sided


> print(dm.test(ferr.q.ts,ferr.arma11.ts))

        Diebold-Mariano Test

data:  ferr.q.tsferr.arma11.ts
DM = -3.6766, Forecast horizon = 1, Loss function power = 2, p-value = 0.0002954
alternative hypothesis: two.sided

> print(dm.test(ferr.q.ts,ferr.aan.ts))

        Diebold-Mariano Test

data:  ferr.q.tsferr.aan.ts
DM = -3.2724, Forecast horizon = 1, Loss function power = 2, p-value = 0.001233
alternative hypothesis: two.sided
```