Group 7

DSC 425 Time Series: Climate Change

By:
Ashley Naratdam
David Van
Sebastian Zdarowski

Abstract:

The overall goal in analyzing this dataset is to model and predict the global land/ocean temperature within a time series. In this paper, various modeling techniques are used to forecast global average land temperature, global average land and ocean temperature, and average U.S. land temperature. Data for the global average land temperature spans from 1753-2015. Several SARIMA models were fit and compared using the BIC and RMSE scores. The SARIMA model with the best performance was an ARIMA(3,0,1) (1,1,1)[12] model which had an RMSE of 0.86. A neural network which combined two LSTM and two convolutional layers was also built to model the data. The neural network had an even lower RMSE of 0.687. Overall, both models performed well at this prediction task, with the neural network outperforming the SARIMA Model. The impact of ocean temperatures was also taken into consideration and separate models were built for the global average land and ocean temperature whose data spanned from 1850-2016. The best performing model, based on AIC and RMSE scores, was an ARIMA(6,1,1)(1,1,1)[12] model which had an RMSE of 0.27. Analysis of the model's residuals showed no autocorrelation and confirmed that the model was appropriate. A closer look was taken at the average temperature for the United States. Multiple SARIMA models were compared using the AIC. The model ARIMA(2,1,1)(0,1,1)[12] performed best with an AIC of 2603.96. A residual analysis ensured the absence of serial correlation in the model's residuals. A linear model was also fit to the U.S. temperature data. The SARIMA model had a slightly lower RMSE score; however, both models did a good job of forecasting temperatures.

Introduction:

The ability to predict future temperatures should be useful when assessing the potential impact of climate change. Climate change is an issue that our world currently faces. It has already had an impact with the average temperature rising across the earth. By being able to build functional time series models, we will allow ourselves to have a better understanding of the impact of global warming and find out the trend of rising temperatures is true. The dataset that we have obtained for this analysis was found on Kaggle, the datasource itself was put together by the Berkeley Earth Surface Temperature Study which combines 1.6 billion temperature reports from 16 pre-existing archives. We will be focusing on three different aspects of the data to see if this problem is able to be modeled properly. Global land temperature, global land and ocean temperature and average United States temperature. These will be used to see if its possible to create models that are able to forecast effectively. By having multiple datasets within this large dataset it will allow us to conduct multiple experiments to see if this is a problem that can be done with forecasting. The methods that models will be built will be elaborated on within the methods/results section of this paper.

Dataset description:

Global Temperatures:Dates(1750-2015) (3192 Rows)

LandAverageTemperature
LandAverageTemperatureUncertainty
LandMaxTemperature
LandMaxTemperatureUncertainty
LandMinTemperature
LandMinTemperatureUncertainty
LandAndOceanAverageTemperature
LandAndOceanAverageTemperatureUncertainty

US Dataset(3 Columns excluding Date) Dates(1750:2013) (2941 rows of data)

AverageTemperature
AverageTemperatureUncertainty
Country

*Link:*https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data

Literature Review:

Time-Series Modeling and Short Term Prediction of Annual Temperature Trend on Coast Libya Using the Box-Jenkins ARIMA Model [AN 1]

ARIMA modeling has also been used to forecast annual temperatures in Libya. The article opens by discussing the already observable effects of global warming. There are notable differences between the surface air temperature change on smaller, regional scales when compared to global changes. Factors such as land cover make regional patterns more variable. It is necessary to look into regional changes as this has varying impacts on the environment and the economies of different countries. Certain regions will experience more significant increases in temperature when compared to the global average increase. In this study, temperatures were collected from 19 coastal locations in Libya which have a Mediterranean climate. ARIMA models were built using SPSS.

After stationarizing the data, the model parameters were then determined using the ACF and PACF of the differenced data. After fitting a model, the residuals were analyzed. The Ljung-Box test was performed to determine if the model's residuals were independent. The ACF and PACF plots were also examined for the final models (ARIMA(3,1,2) and ARIMA (3,2,3)) to ensure no serial correlation was present among the residuals. The residuals for both models appeared to be white noise. The R-squared values were 0.8 and 0.79, which indicated that both did a good job of explaining variation in the data.

Predictions were then made for the 21st century's second decade. The model did well at predicting temperatures when compared to the observed values. In this particular instance, an ARIMA model without a seasonal component was adequate.

SARIMA modeling of the monthly average maximum and minimum temperatures in the eastern plateau region of India [AN 2]

In this paper, a SARIMA model was fitted to monthly average temperatures in Giridih, India. This was done using datasets for both the average maximum temperatures and average minimum temperatures from 1990-2011. The paper presents the need to model and forecast surface air temperatures for agricultural purposes, particularly in India's eastern plateau area. Furthermore, temperature is an input for several other environmental models related to crop growth, making forecasting crucial.

The SARIMA model, or Seasonal Autoregressive Integrated Moving Average model, is appropriate for data with seasonal patterns. In this study, only 22 years were taken into account. For this reason, the data was assumed to be without yearly trend. It was assumed that seasonality was responsible for variation.

To start, the necessary differences were taken to stationarize the data before determining the model order from the ACF and PACF. The order of p (non-seasonal autoregressive terms) and order of q (non-seasonal moving average terms) are initially selected based on the number of spikes in the PACF and ACF plots, respectively. The seasonal terms are also determined from the plots based on spikes at lags that are multiples of the frequency (in this case, 12).

Once model building began, the different models were compared using AIC, BIC and AICc scores and the model with the lowest scores was picked. The model's standardized residuals were then analyzed. The normality and the autocorrelations of the residuals were determined. The presence of autocorrelation among the residuals would indicate an inadequate model. To make sure this was not the case, the Ljung-Box test was performed using several lags. Overall, for both models (maximum average temperature and minimum average temperature) proved to be appropriate based on the diagnostic tests. There was no significant autocorrelation visible in the ACF plots and nearly all of the Ljung-Box test p-values were all greater than 0.5.

Forecasts were made for the next 12 months and compared to the observed temperatures. Except for one month, which had some extremely high temperatures, the observations were within the 95% prediction interval.

ARIMA (1; 0; 2) × (0; 1; 1)[12] was selected as the ideal model for forecasting the average maximum temperatures whereas ARIMA (0; 1; 1) × (1; 1; 1)[12] performed best for the average minimum. This study confirmed that Seasonal ARIMA is a great approach to monthly temperature forecasting. The extreme and unexpected heat in June, however, highlights the need for a different approach when it comes to extreme weather events.

Time-series is a sequence of observations measured sequentially through time. These observations in this problem is considered continuous through time, from 1750-2015.Numerous forecasting models have been proposed to find an effective method that can be applied to practical situations. These techniques mostly rely on complex statistics, artificial intelligence techniques, and large amounts of meteorological and topographic. The most popular method is ARIMA/SARIMA. Aside from traditional models in the time-series forecasting domain is the rise of Neural Network, especially LSTM (Long-Short Term Memory) [DV 1].
ARIMA/SARIMA is regarded as a smooth technique, and it is applicable when the data is reasonably long and the correlation between past observation is stable.
LSTM is the recurrent networks use their feedback connections to store representations of recent input events in the form of activation . The recurrent networks with Long Short-Term Memory have emerged as an effective and scalable model for several learning problems related to sequential data [DV 2]. Convolutional Neural Network is a type of network that is used when data related to the task is spatial in nature or exists in a grid-like topology. 1D CNNs is a new form of Convolutional Neural Networks, which had been proposed and immediately achieved the state-of the-art performance levels in several applications such as personalized biomedical data classification and early diagnosis, structural health monitoring, anomaly detection and identification in power electronics and motor-fault detection [DV 2]. Based on the major advantage of real-time and low-cost implementation, 1D CNNs is able to merge with other Neural Network method to increase the performance.
In this problem, ARIMA/SARIMA will be built to achieve the best performance in forecasting. Also, neural network is built by a combination of Convolutional Neural Network and LSTM Network. These two models will be compared to see whether or not the difference is significant or not.

Using SARIMA to Forecast Monthly Mean Surface Air Temperature in the Ashanti Region of Ghana[SZ 1]
Within this paper the researcher Boaheng focused on building adequate SARIMA models to forecast mean monthly surface air temperature within the Ashanti Region of Ghana for the 2014 year. Data that was used for the research paper was obtained from the department of Meteorology and Climatology in the Ashanti Region from the period of 1985 to 2013. Author went through 4 steps to obtain optimal model for forecasting. Model Identification: KPSS and ADF tests were performed to check for the data stationary condition to be satisfied. KPSS and ADF test obtained .1 and .01 values to show that time series was stationary and ready for model building. ACF and PACF plots were created to obtain model parameters. Seasonal difference was performed as well to obtain seasonal components for time series. 2nd step was model

building, researcher went through 6 SARIMA models to obtain model with lowest AIC and BIC values. Final model obtained was SARIMA(2,1,1)(1,1,2)[12] Goodness of Fit tests were performed, residuals were inspected to ensure that SARIMA model was modeling white noise within residuals. Forecasting was the final step, forecasting shown within the paper followed trends of previous mean air temperatures, researcher concluded that model was optimal for forecasting. Researcher didn't use test set or training set to test forecasting, paper is lacking in how final model was truly optimal. In the future, author should include some type of test set in order to validate model properly. Within this paper, we will be validating models with proper training and testing sets.

Global Warming and the Problem of Testing for Trend in Time Series Data[SZ 2]

Researches Woodward and Gray analyzed the trend within two global warming time series "Hansen and Lebedeff" & "IPCC Series" and look into determining if ARMA models are appropriate to capture global warming trends in the short run or if can be applied in the long term. Several models are applied, linear regression models, ARMA and ARIMA models to determine if trend can be captured. Once models are built, forecasting is done for all models that are completed. Then two tests are administered on the forecasting results; Linear trend and Monotonic trend. Authors conclude that after running these tests, that the ARIMA model would be a plausible model to be able to predict future trends within the dataset. If ARIMA model contains correlation present within the model, then there is a high chance that the trend that is captured would be incorrectly captured within model and would cause issues with forecasting accuracy.

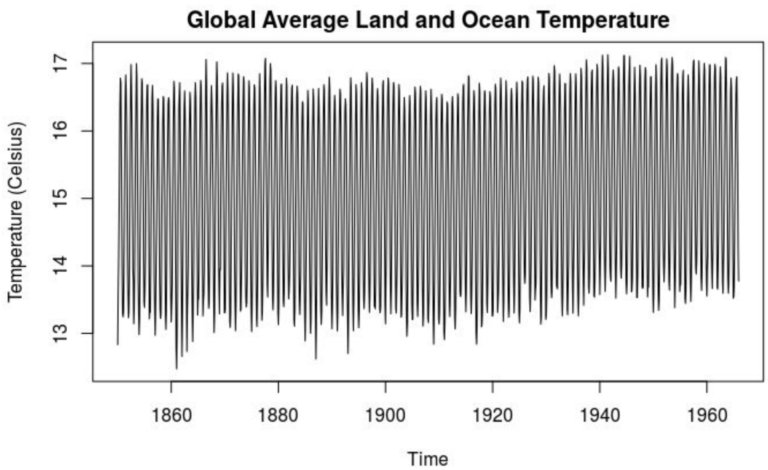
Methods & Results:

We will be examining different subsets of the global dataset and US dataset and applying different modeling techniques in order to build a proper forecasting model for each dataset. Some techniques that will be shared across the datasets are Autocorrelation function(ACF) and Partial Autocorrelation Function(PACF) in order to identify the proper model order of each model at various lags. The Ljung box test will also be applied to models to ensure that there is no correlation present within each model.We would use this techniques to ensure that our dataset fits the standard requirements for the different time series models. Each section will have a training and test set in order to properly evaluate each forecasting model. Sections will be broken up by as such: Ashley will be measuring the Global Ocean and Land Temperature, David will be focusing on the Global Land Temperature and Sebastian will focus on the United States average temperature.

Global Average Land and Ocean Temperature

Exploratory Analysis

Analysis was done on the variable “LandAndOceanAverageTemperature” for the global data set in order to explore trends and forecast the combined land and ocean temperatures. Data is present for this feature from 1850 to 2016. Temperatures were recorded once a month; In total, 1992 temperatures were recorded. A 70/30 split was performed to create training and testing sets before model building began. Below is a plot of the training data where a slight upward trend can be detected.



Prior to building a SARIMA model, the ACF and PACF plots of the stationarized training data must be examined. This will help determine the parameters for the model which include the number of seasonal and nonseasonal differences as well as the number of moving average and autoregressive terms to include.

The ACF and PACF plots of the training data are shown in Figure 2. In order to build SARIMA models, the level of differencing to be performed on the data must be determined. The sinusoidal pattern in the ACF plot indicates that there is seasonality in the temperatures. As such, a seasonal difference must be taken. A difference with a lag of 12 was taken since the frequency of monthly data is 12. In addition to this seasonal difference, a first-order difference was also taken to address the trend in the data.

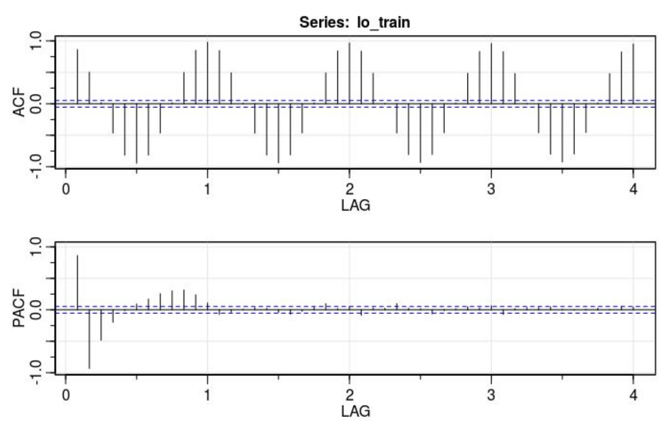


Figure 2

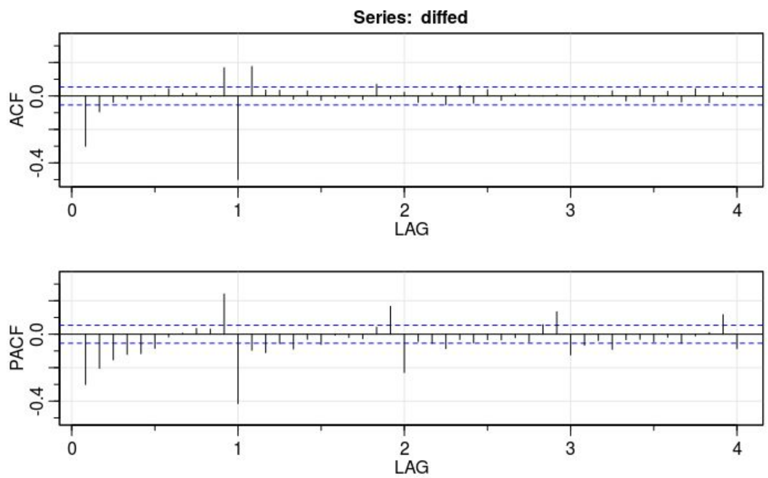


Figure 3

The ACF and PACF plots of the differenced data is shown in Figure 3. There is a spike at lag-1 and another spike at lag-12 in the ACF plot. This would suggest a SARIMA model with one non-seasonal MA term (q=1) and one seasonal MA term (Q=1). There are also spikes at lags 1 and 12 for the PACF plot, which indicates that 1 non-seasonal (p=1) and 1 season (P=1) AR term should be included in the model to start.

Model Building

To start, a SARIMA model of the following order was built: ARIMA(1,1,1)(1,1,1)[12]. In the previous section, both a seasonal and a non-seasonal difference were taken to stationarize the data. As a result, d=1 and D=1 for this first model. Analysis of the ACF/PACF also indicated that a model with a seasonal and non-seasonal MA term as well as a seasonal and non-seasonal AR term should be tried. The AIC of the model was -1840.19 and the BIC was 1814.04. The significance of the model's coefficients were found, and it revealed the seasonal AR term to be insignificant, so it was removed. The resulting ARIMA(1,1,1)(0,1,1)[12] model had an AIC=-1842.13 and BIC=-1821.21.
Several models were built after this. The MA and SMA orders both remained the same (q=1 and Q=1) throughout. Varying number of AR terms were tried. In addition, the non-seasonal difference parameter was set to 0 at one point. The results of experimentation are summarized in Table 1. The first model in the table was the output of the auto.arima function.This gave the highest AIC. The RMSE was calculated for each model using the test data and forecasted temperatures.

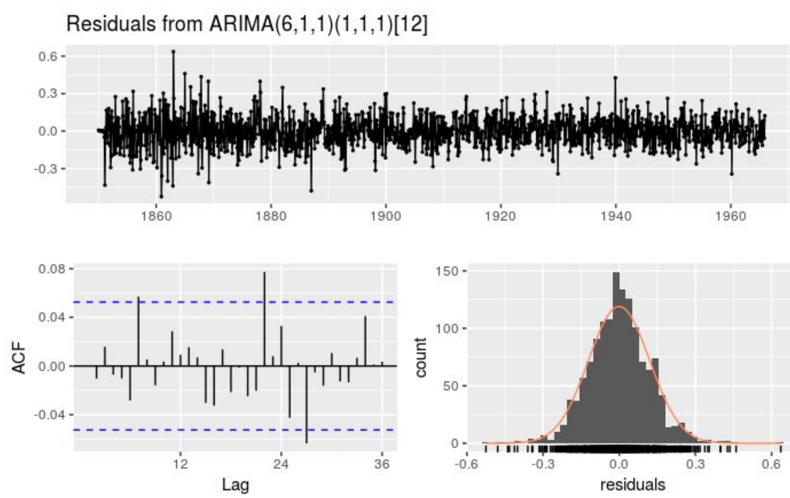
	AIC	**RMSE**
ARIMA(0,0,1)(0,1,1)[12] with drift	-1484.15	0.2782001
ARIMA(1,1,1)(1,1,1)[12]	-1840.19	0.3784482
ARIMA(1,1,1)(0,1,1)[12]	-1842.13	0.3810591

ARIMA(5,1,1)(1,1,1)[12]	-1854.4	0.2752052
ARIMA(6,1,1)(1,1,1)[12]	-1857.12	0.2700186
ARIMA(2,0,1)(0,1,1)[12]	-1862.29	0.4157216

Table 1

Overall, the ARIMA(6,1,1)(1,1,1)[12] was selected as the best performing model. Although it had a higher AIC than the ARIMA(2,0,1)(0,1,1)[12] model, it had the lowest RMSE score for the test data. Residual analysis was also performed for each model to determine if the model was appropriate.

The residuals of the ARIMA(6,1,1)(1,1,1)[12] model appear to be independent. The ACF plot of the residuals in Figure 4 does not show any serial correlation. The Ljung-Box test was also done for lag-5, lag-10, lag-15 and lag-20. The p-value for each test was greater than 0.5 which confirmed the absence of autocorrelation among the residuals.



Below are the coefficient for the model terms. AR3 was removed, but some insignificant terms were left in. This was because their removal led to models whose residuals were not independent.

	Estimate	**Pr(>|z|)**
ar1	0.5184512	< 2.2e-16 ***
ar2	0.0793911	0.005234 **
ar4	0.0546961	0.054263 .
ar5	0.0742000	0.015899 *
ar6	0.0500592	0.069256 .
ma1	-0.9866841	< 2.2e-16 ***
sar1	0.0255945	0.372636
sma1	-0.9683684	< 2.2e-16 ***

The forecast for the next 600 timesteps was then plotted in Figure 5.

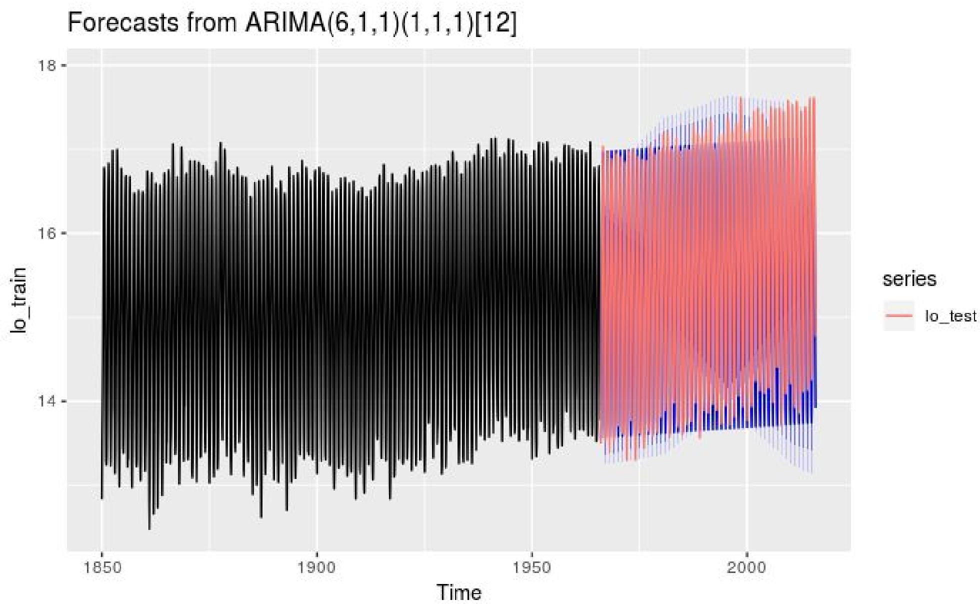


Figure 5

The black portion is the training data. The blue lines correspond to the predicted temperatures, and the pink lines are the actual values of the test data. The predictions seem decent; however, there seems to be more of an upward trend in the actual data.

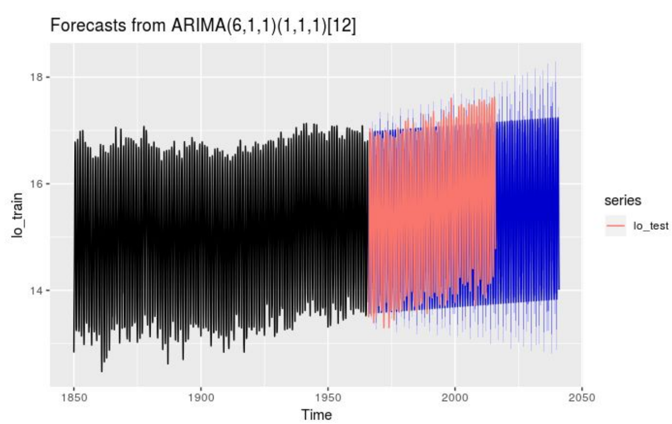


Figure 6

Figure 6 shows a forecast extending 300 months beyond the observed data. This model predicts a continuing upward trend. However, the test data seems to have a more dramatic upward trend which may continue into the predicted timesteps.

Global Average Land Temperature

The dataset is split into training set and testing set equivalent with 70% and 30% of the original dataset. The training dataset is start from: 1753 - 2936. The testing set is start from 1937 - 2015. The training dataset will be used to stimulate for the two models, then the testing set will be used to test the performance of two models. For model evaluation, RMSE will be used.

ARIMA/SARIMA

The process of building ARIMA/SARIMA is based on understanding data nature. ACF and PACF is applied to identify the number of lags, the seasonality and how to modify the data to achieve the stationary stage. In the first stage, without modification, the data shows to have seasonality and trending (Fig.1)

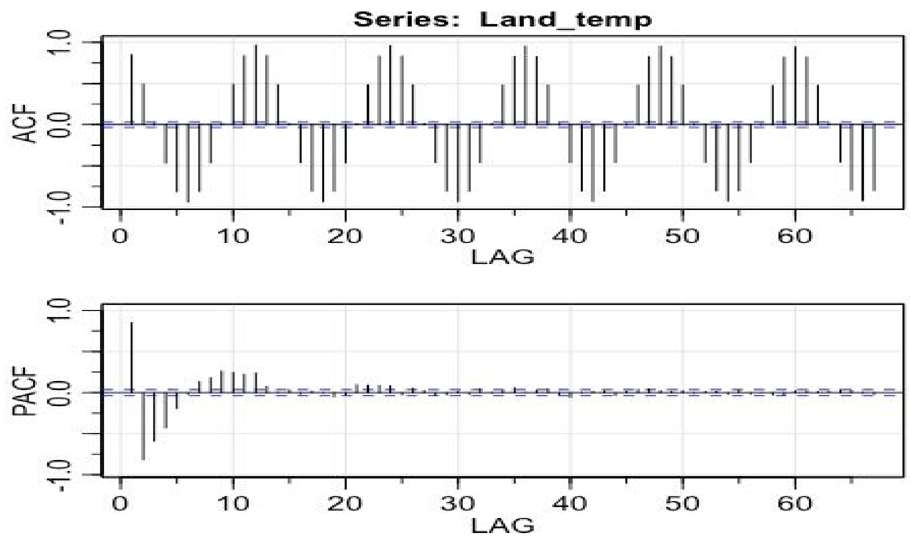


Fig.1

Fig.1 shows ACF contain trending and sequences repeated for every 10 lags frame.
The first difference is applied to the series.

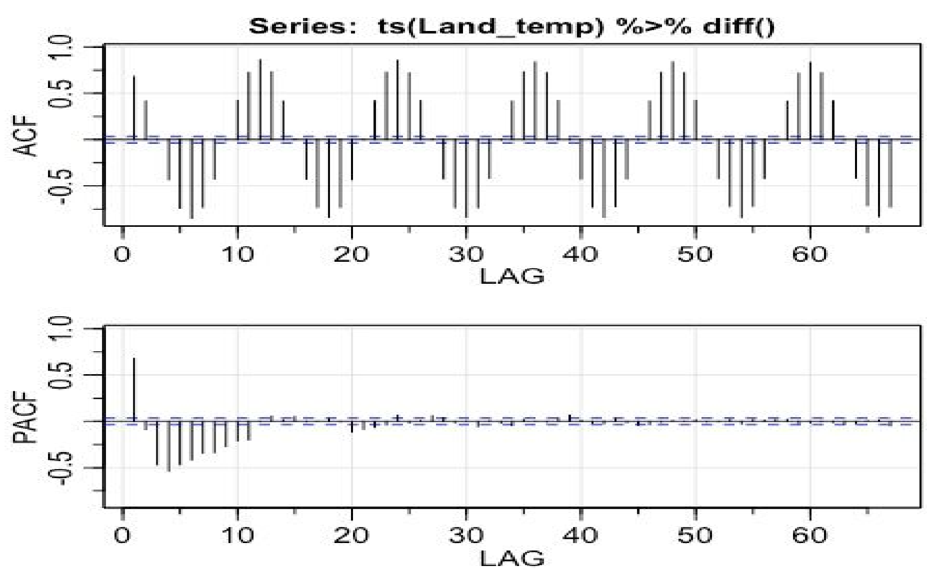


Fig.2

Fig.2 shows there is no significant change compared with the original. Even though there is a slightly different, ACF still shows the pattern or trend of all lags. Therefore, it is necessary to take a second difference with lag = 12 as seasonality.

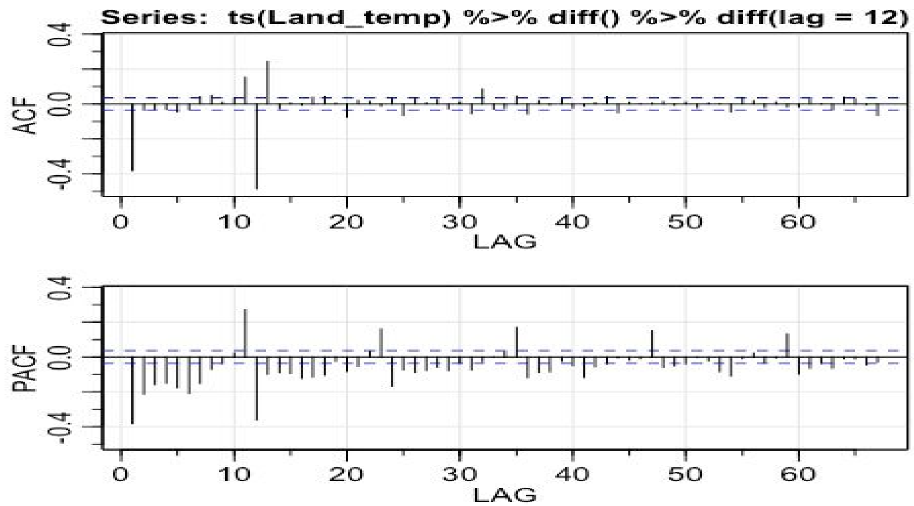


Fig.3

Fig.3 shows there is a significant change in ACF and PACF. ACF shows no trend or pattern among lags. In other words, the series is now achieving stationary status as shown in Fig 4

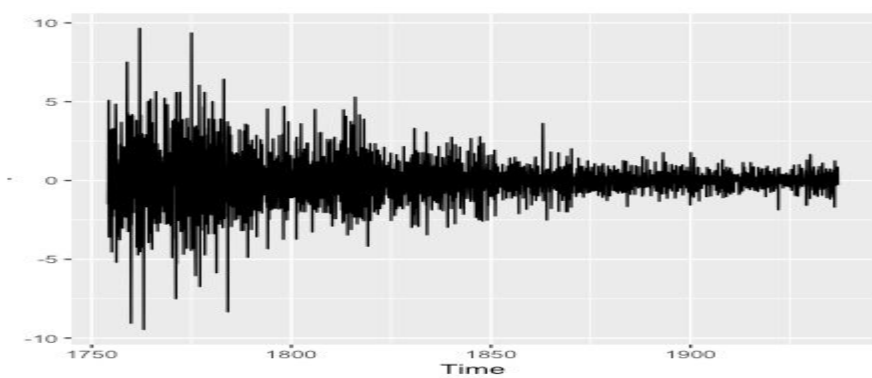


Fig.4

Fig.3, ACF plot shows a significant negative spike at lag 1 and lag 12. Also,at lag 11 and 13 show positive spikes, which is suggest either AR or MA. For MA, the model can be MA(1) and can be up to lag =12. For AR, the model can be AR(1). As PACF shows a slowly decaying until 0 of the coefficient as lags increase. From these characteristics of model, it is better to use SARIMA instead of ARIMA seen the series is involved with pattern or trend. The model then can be set as ARIMA(1,0,1)(1,0,1)[12] as the starting model. However, since auto.arima() is available, it makes more sense to try it first.

```
Series: temp_train
ARIMA(1,0,1)(1,1,0)[12]

Coefficients:
         ar1      ma1     sar1
      0.3798  -0.0721  -0.4740
s.e.  0.0555   0.0590   0.0189

sigma^2 estimated as 1.058:  log likelihood=-3177.49
AIC=6362.99   AICc=6363   BIC=6385.76
```

```
z test of coefficients:

      Estimate Std. Error  z value  Pr(>|z|)
ar1   0.379774   0.055534   6.8386 7.995e-12 ***
ma1  -0.072127   0.059024  -1.2220    0.2217
sar1 -0.473950   0.018908 -25.0663 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Auto.Arima() function recommend model Arima(1,0,1)(1,1,0)[12], in which achieve BIC score of 6385.76.

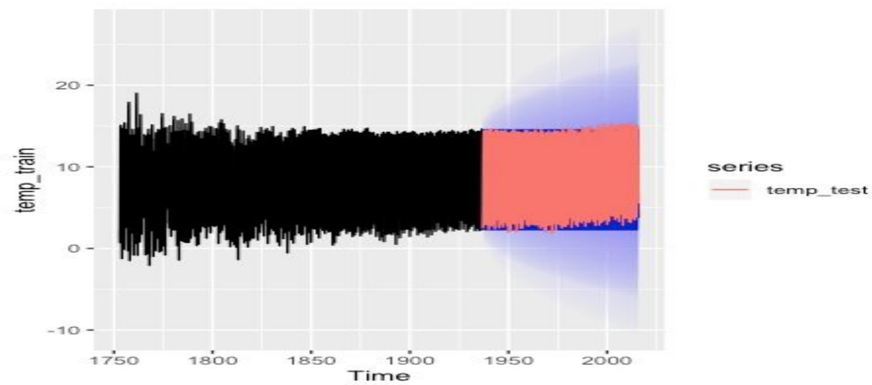


Fig.5

Fig 5 shows the prediction of model, the blueline, compares to the actual data, the redline. After 2000, there is a significant variance between predicted and actual.
As auto.arima() is the foundation. I create different models that can have better BIC score.

Model	BIC	RMSE
ARIMA(2,0,1) (1,1,1)[12]	5730.19	0.875
ARIMA(2,0,1) (1,1,2)[12]	5756.35	0.8779
ARIMA(3,0,1) (1,1,1)[12]	5714.66	0.86
ARIMA(3,1,1) (3,1,1)[12]	5735.62	0.87
ARIMA(3,1,1)(3,2,1)[12]	6327.75	0.986

The five models that are built have better BIC score compare to the auto.arima function. The best among them is the model ARIMA(3,0,1)(1,1,1)[12]. The model achieve BIC score 5714.66, which is the smallest BIC score overall.

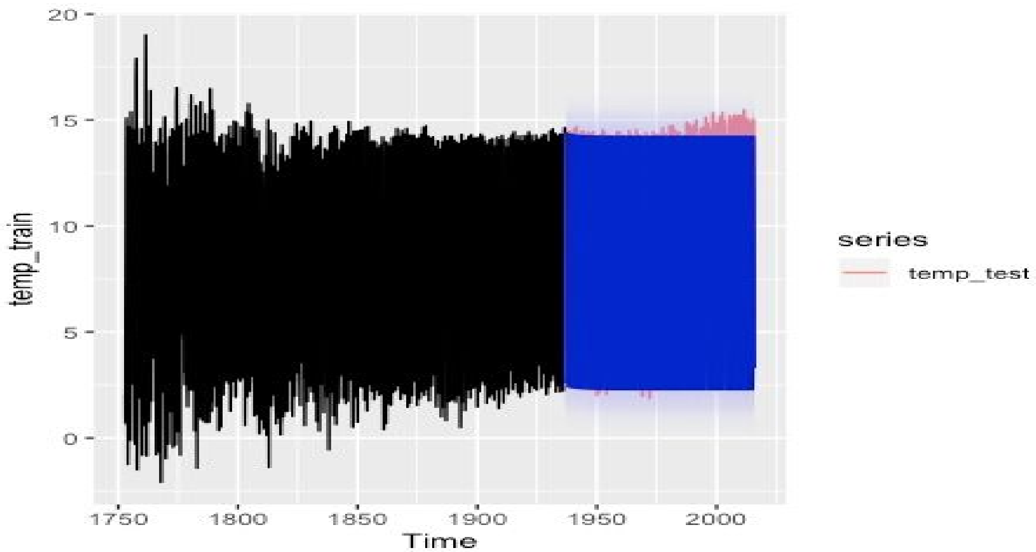


Fig.6

Fig.6 shows the prediction of test set from 1937 to 2015. The mean of prediction residuals is 0.0036, which is a very small residual in term of temperature different. There is a small difference between forecasting and actual after 2000. As the actual temperature rise up due to global warming, and the forecast could not catch that situation.

LSTM + Conv1D

This model was created based on the research of Siami [DV 2], which is the application of 1D Convolutional Neural Network. Understanding the use of LSTM, which was using a mixture of Real Time Recurrent Learning and Backpropagation through time. This model build is a network combined of two LSTM hidden layers with two Convolution1D. To be able to apply to this model. The training and testing dataset need to be transformed into time frames with variables are the 12 months and the target is the 13th month.

```
model6 = keras.Sequential()

model6.add(Conv1D(filters=32, kernel_size=3, activation='linear',
                  input_shape=(x_train.shape[1],x_train.shape[2])))
model6.add(Conv1D(filters=32, kernel_size=3, activation='linear'))
model6.add(Dropout(0.2))
model6.add(MaxPooling1D(pool_size=1))
model6.add(Flatten())
model6.add(RepeatVector(1))

model6.add(LSTM(20, recurrent_dropout=0.3, activation='linear', return_sequences=True,
                batch_input_shape=(x_train.shape[1],x_train.shape[2])))
model6.add(TimeDistributed(Dense(30,activation='linear')))
model6.add(TimeDistributed(Dense(1)))
model6.add(Dense(1))

model6.compile(loss='mse', optimizer='adam')
model_6 = model6.fit(x_train, y_train, epochs=50, batch_size = 30, verbose =1,
                     shuffle=False,callbacks=[early_stopping])
```

Fig.7

Fig 7 shows what the model look like while using python perform the calculation behind. This model has been used to predict the temperature of the 25th hours based on the 24 hours. The reason behind, I want to see how well ARIMA/SARIMA perform compares to the already built neural network that has been used from the different dataset. The result of the neural network:

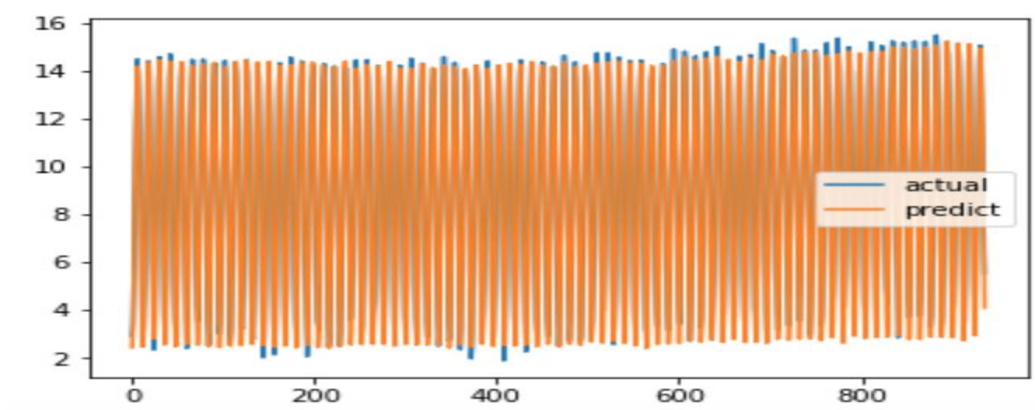


Fig.8I

Fig 8 is the prediction of the model compared to the actual testing data. The plot of two data is pretty close. To be precise, RMSE is performed to evaluate two models:

	SARIMA	LSTM
RMSE	0.86	0.687

RMSE of Neural Network model perform significantly smaller than SARIMA model. In other words, the neural network model perform pretty well in predicting the future event or showing that the average land temperature is getting higher, signal of global warming.

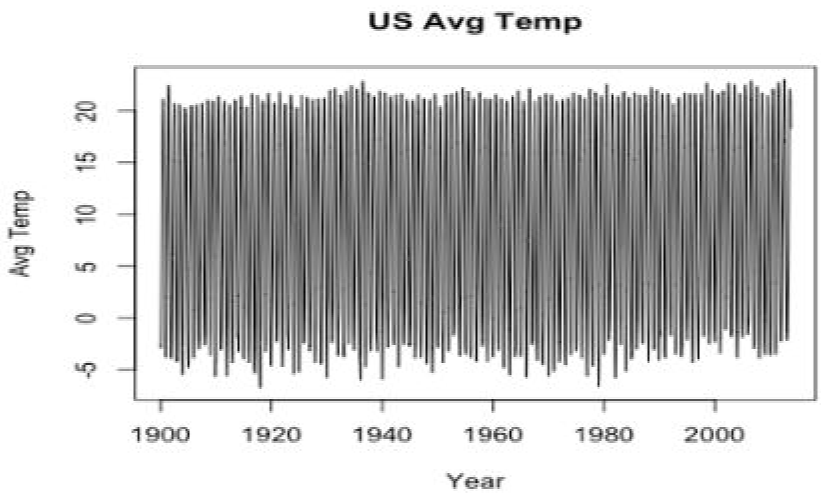
United States Average Land Temperature

For the United states dataset, we will be applying SARIMA models and a Linear Regression model in order to test if the dataset can be used to forecast for average united states temperature efficiently. The dataset will be split into 2 sections, training (70%) and testing(30%).

Data Exploratory:

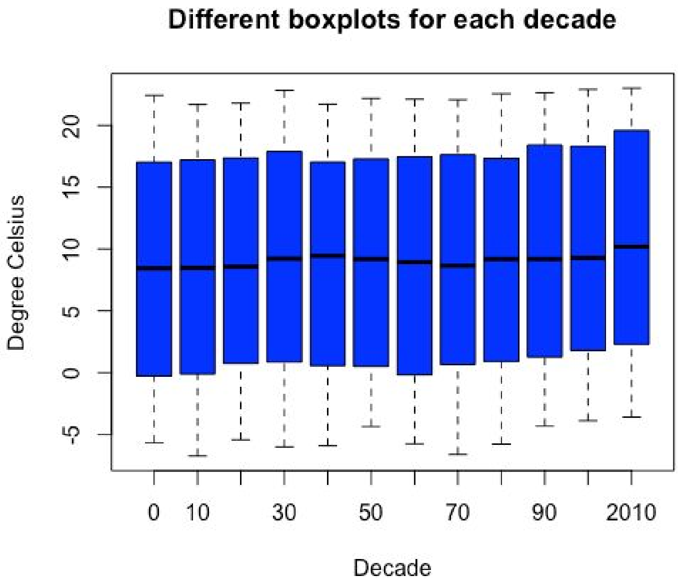
Data ranges from 1900, January 2013, September. Dates prior to 1900 were discarded as they were not as relevant to the problem at hand and wanted to see the impact of 1900 and forward of the average temperature within the United States. Frequency of the dataset is 12, due to this being a monthly dataset. The Maximum temperature is 23.08 C(73.54 F) and minimum temperature was -6.737 C (19.87 F)

Plot of the current dataset without any transformation:



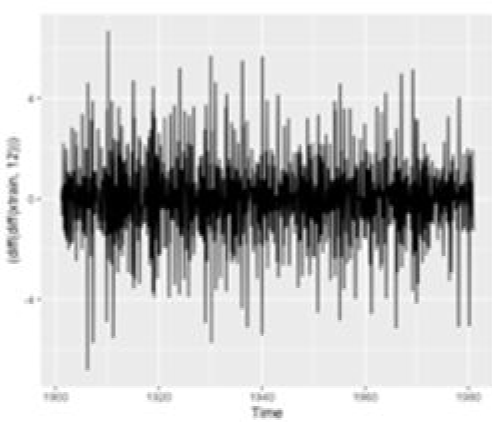
Fig(1 - SZ)

Within the current plot(Fig 1-SZ), we are unable to see anything of value truly besides temperature ranging from the max and min temperatures spanning across 113 years. We do see that the min temperature past 2000 is higher than the min temp at 1900. Slight trend is present within the dataset, as well as seasonality, we can tell this by seeing the up and down spikes occurring within each year. Within the box plot chart seen below(Fig B-SZ), the mean average temperature has been rising since 1900. Comparing boxplot of 2010 and 1900, we see that the avg temperature is higher in 2010 and the avg temperature is not as low as 1900 boxplot.

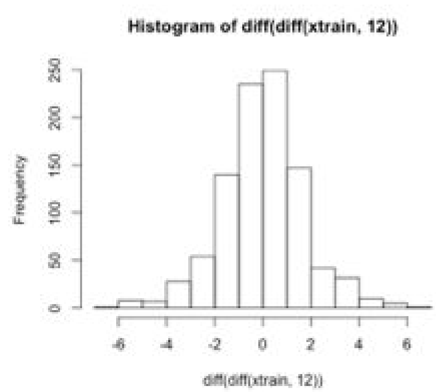


Fig(B-SZ)

Dataset needs to be pre processed due to it currently not being stationary. With modeling building, we need to have a stationary model in order to properly build a functioning model for forecasting. Model contains trend and seasonality, so difference needs to be taken twice to remove trend from the data and to remove seasonality. Below is a plot of the graph when difference is taken and another difference with lag 12 is applied. This plot(Fig 2 -SZ) shows that our data has become stationary, as the mean has centered around zero. Histogram analysis(Fig 3 -SZ) was also performed to see if the distribution followed normality, and with the histogram chart it is shown that it's following the bell shape curve implying normal distribution.

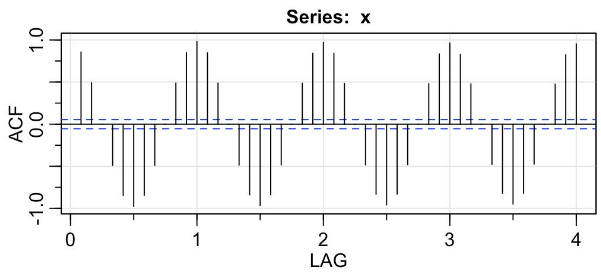


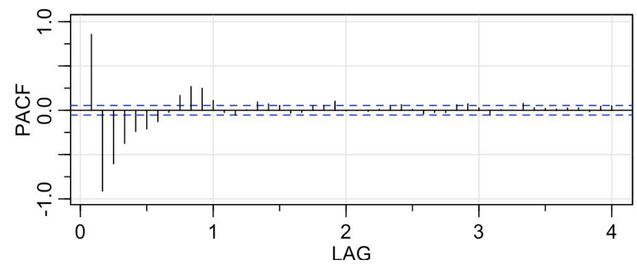
(Fig 2 -SZ)



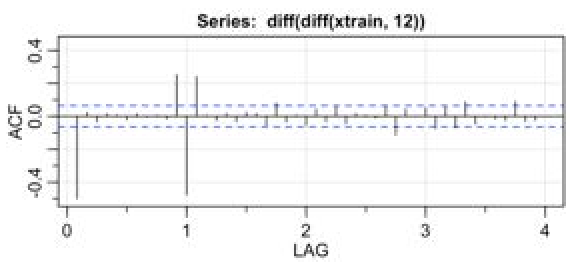
(Fig 3 -SZ)

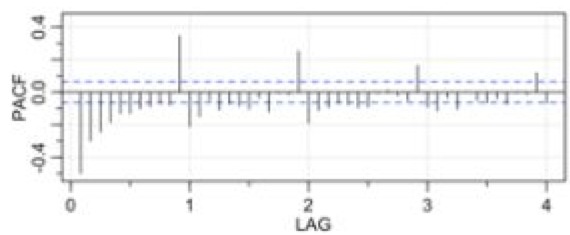
ACF & PACF analysis was needed to be done in order to find proper model order within the model.





Within the regular ACF plot without differencing we notice that we have a seasonal model by seeing the spikes occurring at ever half lag point. PACF plot shows we have 2 significant spikes at beginning possibly showing that p =2 is a possibility for model order. Within the twice differenced ACF plot, we noticed that we have 1 lag that before seasonal aspects, and another spike at lag 12, showing that seasonality will play a role within this chart. From this ACF plot, its concluded that q=1 and Q= 1.





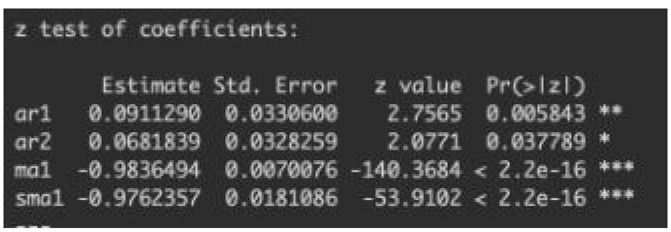
PACF chart is showing a significant spike at 1 which would imply that p=1, since this lag is larger than other spikes and will make up the difference for the other spikes. Its noticed that there is a significant lag spike at 12 as well, so there is seasonality present within the PACF plot, meaning that P = 1. Differences for the model order will be the following d=1, and D = 2.

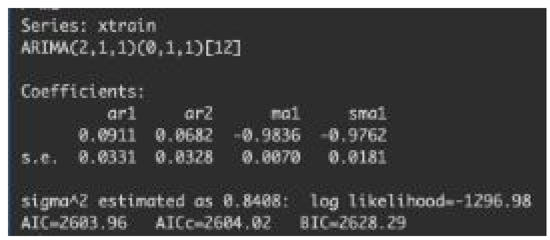
Modeling Building:

There will be 2 different models will be built in order to build proper forecasting models and to see which model outperforms the other in terms of forecasting and error. A Seasonal Autoregressive Integrated Moving Average model will be applied to the data. ARIMA is one of the most widely used forecasting methods for time series data forecasting, unfortunately there is a limitation when only applying an ARIMA model when it comes to seasonal data. The SARIMA model is an extension of this model that was built in order to handle seasonal elements within the data. While the ARIMA model contains 3 trend elements p, d and q, SARIMA model incorporates 4 additional elements P,D,Q and m(number of time steps used for a single seasonal period) to handle the seasonal aspect of the time series. Before building the model a good technique to apply is the auto.arima function with R, to see what kind of model is built using this function as a starting ground for the model build. The auto.arima function provided a model with parameters ARIMA(2,0,0)(2,1,0)[12] with an AIC value of 2819.9. While this is a good model for starting, lets see if we can make the model perform better with some fine tweaking of the parameters.

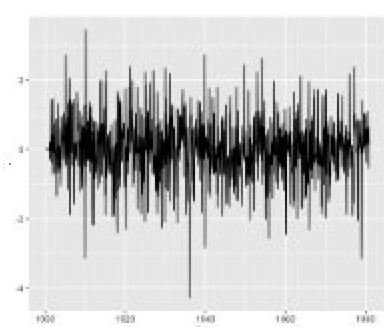
Sarima	Parameters	AIC Values
1st Build	(2,0,2)(2,1,0)[12]	3177.43
2nd Build	(2,0,2)(1,1,0)[12]	2957.81
3rd Build	(1,0,0)(2,1,0)[12]	2822.26
4th Build	(1,1,1)(1,1,1)[12]	2607.4
5th Build	(1,1,1)(1,1,1)[12]	2605.5
6th Build	(2,1,1)(0,1,1)[12]	2603.96

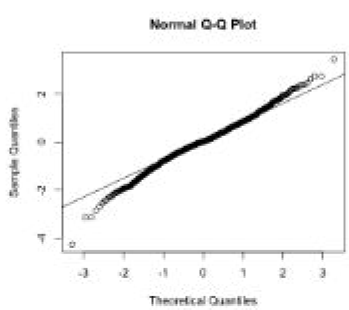
After running some tweaks on the model, the 6th build seemed to turn out to be the most optimal model. The parameters that we chose in the initial model building, seemed to be close to what the final model consisted off. P=1 was dropped due to no significance within the model. p=2 was obtained a better AIC score than p=1 so it was included within the final model and the variable proved to be significant in the coefficient test.

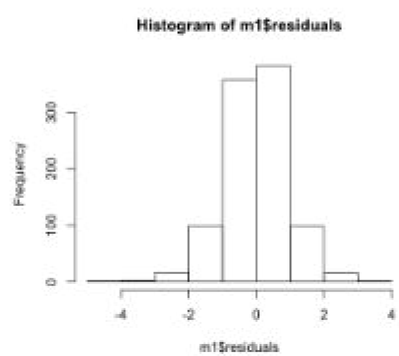

```
z test of coefficients:

      Estimate Std. Error   z value  Pr(>|z|)
ar1   0.0911290  0.0330600    2.7565  0.005843 **
ar2   0.0681839  0.0328259    2.0771  0.037789 *
ma1  -0.9836494  0.0070076 -140.3684 < 2.2e-16 ***
sma1 -0.9762357  0.0181086  -53.9102 < 2.2e-16 ***
```

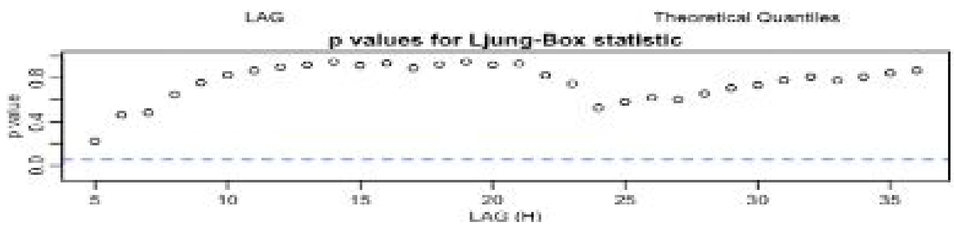

```
Series: xtrain
ARIMA(2,1,1)(0,1,1)[12]

Coefficients:
         ar1     ar2      ma1     sma1
      0.0911  0.0682  -0.9836  -0.9762
s.e.  0.0331  0.0328   0.0070   0.0181

sigma^2 estimated as 0.8408:  log likelihood=-1296.98
AIC=2603.96   AICc=2604.02   BIC=2628.29
```

Now that the final model is built, residual checks are needed to be completed in order to ensure that a proper model has been built. Residual chart plot, shows that the model residuals are exhibiting white noise. Histogram residuals are showing bell shaped curve, still implying normality is achieved within the model. QQ-plot lines are mostly falling on the line, indicating our residuals are falling within normality.



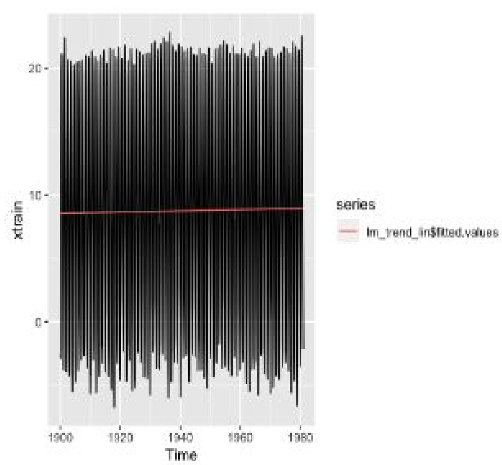




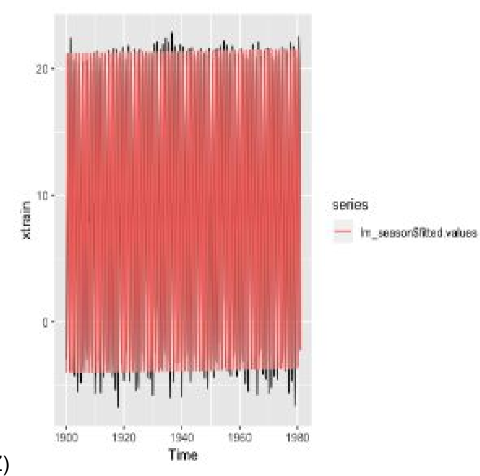


The p values for the Ljung-Box test statistics are looking good, since all values up to lag 35 are above .05 indicating that our model is not showing any significant correlation present.

A linear model was applied to the US dataset as well to see how well the model can handle this dataset, and in order to provide a comparison against the SARIMA model. Model was fitted with the US train dataset, same subset as the SARIMA model. Model was created using the function tslm() that is present in the time series package. This function is used to fit linear models to a time series that includes trend and seasonality. Building the model was straightforward as it required the training dataset, setting trend and seasonality to true within the function.



Fig(4-SZ)



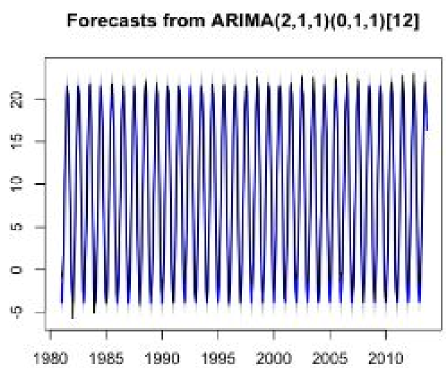
Fig(5-SZ)

Within the charts above, we are able to see the trend that is present within the linear model, its slightly going upwards as the average temperature within the US increases(Fig 4 - SZ). By applying the seasonal aspect into the model, we are able to see by figure(Fig 5 -SZ) that the seasonal component is an important part of this linear model. Since the seasonal fitted values, follow the trend of the training dataset. Model equation for final linear model: Model Equation: Yt = -4.021 + .0003355(Bt) + 1.978(S2) + 6.26(S3) + 11.87(S4) + 17.60(S5) + 22.60(S6) + 25.21(S7) + 24.24(S8) + 20.05(S9) + 13.60(S10) + 6.59(S11) + 1.43(S12) + et

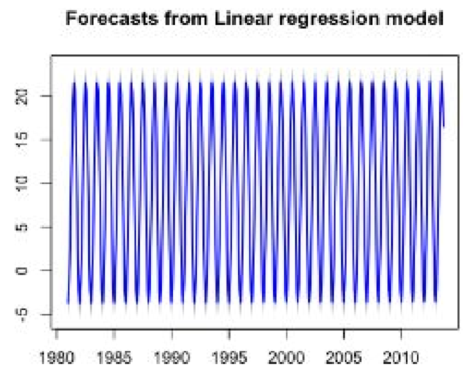
Forecasting comparison:

Both models were used for forecasting to see which model performed better. Both models performed relatively well in forecasting. The difference between both of these models was small.

It seems that the SARIMA model was able to capture the trend and seasonal aspect of the dataset quite well, as seen in figure(Fig -SZ). We are able to see that fitted lines form xtest(blue lines) fell almost exactly on the forecast prediction.



(Fig 6-SZ)



(Fig 7-SZ)

Linear model also performed quite well in terms of forecasting, seems that that model was quite stable and optimized for predictions(Fig 7-SZ). This dataset had no issue with forecasting or any outlying data points that would affect predictions.

RMSE	Scores
SARIMA	0.9088777
Linear	0.9162535

Both models scored close to each other in comparison to the RMSE scores across both models. SARIMA model performed slightly better in the RMSE score. But looking at both of the forecasting charts, they results are similar. The recommendation going forward would be to use the SARIMA model in forecasting due to the slightly lower RMSE score. In terms of model building and time, the linear model would be the option to go with, as this model didn't take quite as much time to build the model and was able to handle the dataset with similar results compared to the SARIMA model. In the future, if we were to do work on this US Average Data Set temperature, it would be beneficial to obtain another dataset, that had results from 2013 to 2019 and run a forecast for this model through that period to see how well the model would forecast in a longer period.

Conclusion:

In this project of predicting the future temperature, ARIMA/SARIMA is the focus model in predicting the test set. Each of us work on different dataset and have different opinion of how the model is going to be. Building ARIMA/SARIMA is the process of trial and error. Even though we have auto.arima() function helps, the model produced by auto.arima is not often the best one. There are a lot of work in understanding the series and be able to find the best parameters for the model.

We also work on different model to compare with the ARIMA/SARIMA model such as LSTM or Linear. As Sebastian work on the linear model, he found that there is not a significant difference in RMSE score between the two models, but ARIMA/SARIMA does prove that it achieves a better score. As Duy work on the LSTM, he found that LSTM perform significantly better than ARIMA/SARIMA. However, if account the time run and computational power, ARIMA/SARIMA perform better. Therefore, he wants to do more research about ARIMA/SARIMA to improve this model prediction. A SARIMA model also performed well when forecasting global average land and ocean temperature. For future work, it may be worthwhile to fit a model to the more recent data to see if this improves forecasting.

References

Ashley:
[AN 1] El-Mallah, E., & Elsharkawy, S. (2016). Time-Series Modeling and Short Term Prediction of Annual Temperature Trend on Coast Libya Using the Box-Jenkins ARIMA Model. *Advances in Research*, *6*(5), 1–11. doi: 10.9734/air/2016/24175
[AN 2] Saha, E., Hazra, A., & Banik, P. (2016). SARIMA modeling of the monthly average maximum and minimum temperatures in the eastern plateau region of India. *Mausam*, *67*(4), 841–848.

David:
[DV 1] A. E. Permanasari, I. Hidayah and I. A. Bustoni, "SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence," *2013 International Conference on Information Technology and Electrical Engineering (ICITEE)*, Yogyakarta, 2013, pp. 203-207. doi: 10.1109/ICITEED.2013.6676239
[DV 2] Siami Namini, Sima & Tavakoli, Neda & Siami Namin, Akbar. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. 1394-1401. 10.1109/ICMLA.2018.00227.

Sebastian:
[SZ 1]Boaheng, Micael Asamoah. “Using SARIMA to Forecast Monthly Mean Surface Air Temperature in the Ashanti Region of Ghana .” *International Journal of Statistics and Applications* , vol. 4, 2014, pp. 292–299., doi:10.5923/j.statistics.20140406.06.

[SZ 2]Woodward, Wayne A., and H. L. Gray. “Global Warming and the Problem of Testing for Trend in Time Series Data.” *Journal of Climate*, vol. 6, no. 5, 1993, pp. 953–962., doi:10.1175/1520-0442(1993)0062.0.co;2.