Bombing Operation Analysis During World War Ⅱ

Name: Yukang Xu

Ⅰ. Introduction

Strategic bombing during World War II was the sustained aerial attack on railways, harbours, cities, workers' and civilian housing, and industrial districts in enemy territory during World War II. Strategic bombing is a military strategy which is distinct from both close air support of ground forces and tactical air power.

During World War II, it was believed by many military strategists of air power that major victories could be won by attacking industrial and political infrastructure, rather than purely military targets. Strategic bombing often involved bombing areas inhabited by civilians and some campaigns were deliberately designed to target civilian populations in order to terrorize and disrupt their usual activities.

The effect of strategic bombing was highly debated during and after the war. Both the Luftwaffe and RAF failed to deliver a knockout blow by destroying enemy morale. However some argued that strategic bombing of non-military targets could significantly reduce enemy industrial capacity and production and in the opinion of its interwar period proponents, the surrender of Japan vindicated strategic bombing.

Ⅱ. Data Analysis

1. Identify the dataset

As I mentioned at introduction, we use multiple data sources.

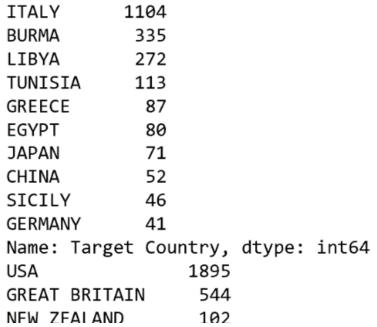
a. Aerial Bombing Operations in WW2Shortly, this data includes bombing operations. For example, USA who use ponte olivo airfield bomb Germany (Berlin) with A36 air craft in 1945.
b. Weather Conditions in WW2Shortly, weather conditions during ww2. For example, according to George Town weather station, average temperature is 23.88 in 1/7/1942.This data set has 2 subset in it. First one includes weather station locations like country, latitude and longitude. Second one includes measured min, max and mean temperatures from weather stations.

Since we have so many information, we only keep useful information in three data resources. Based on that, we analyzed every variable then we make a table as below.

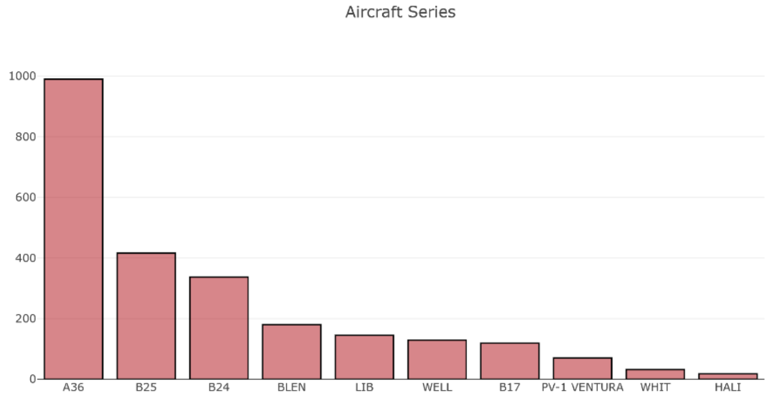
Name of spreadsheet	Description	Variables
Operation	Airfield bombing information	Mission date; takeoff and target country, longitude, altitude
Weather station location		Station longitude, altitude
Summary of weather	Weather in one specific date	Temperature, Precipitation, wind speed

2. Explore the dataset

To explore our dataset, we need to pick the important information at first. In this case, most useful information for us is about target country and set-off country. From the graphs below, we know that US made most of airfield bombing missions. Their destinations are most in Italy and Burma. A36 is their favourite choice.

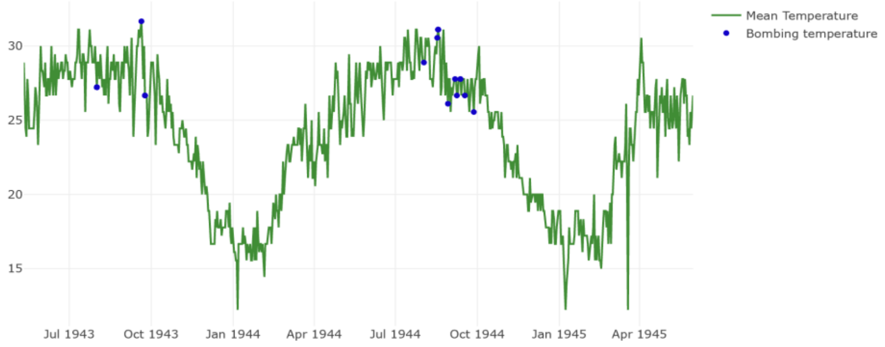


```
ITALY      1104
BURMA       335
LIBYA       272
TUNISIA     113
GREECE       87
EGYPT        80
JAPAN        71
CHINA        52
SICILY       46
GERMANY      41
Name: Target Country, dtype: int64
USA                1895
GREAT BRITAIN       544
NEW ZEALAND         102
```



At this time, we focus on the operation between US and Burma. The Burma campaign was a series of battles fought in the British colony of Burma, primarily involving the forces of the British Empire and China, with support from the United States, against the invading forces of Imperial Japan, Thailand. The climate of the region is dominated by the seasonal monsoon rains, which allowed effective campaigning for only just over half of each year. This, together with other factors such as famine and disorder in British India and the priority given by the Allies to the defeat of Nazi Germany, prolonged the campaign and divided it into different phases from 1942-1945.

Mean Temperature --- Bombing Dates and Mean Temperature at this Date



After exploring the connection between weather and mission date, mission usually happened with high temperature. dry heat would raise the risks of heat-related illness. heavy work must limit working hours. If they were taking rest, they may get together. That will be easier for bombing operation.

3. Clean the dataset

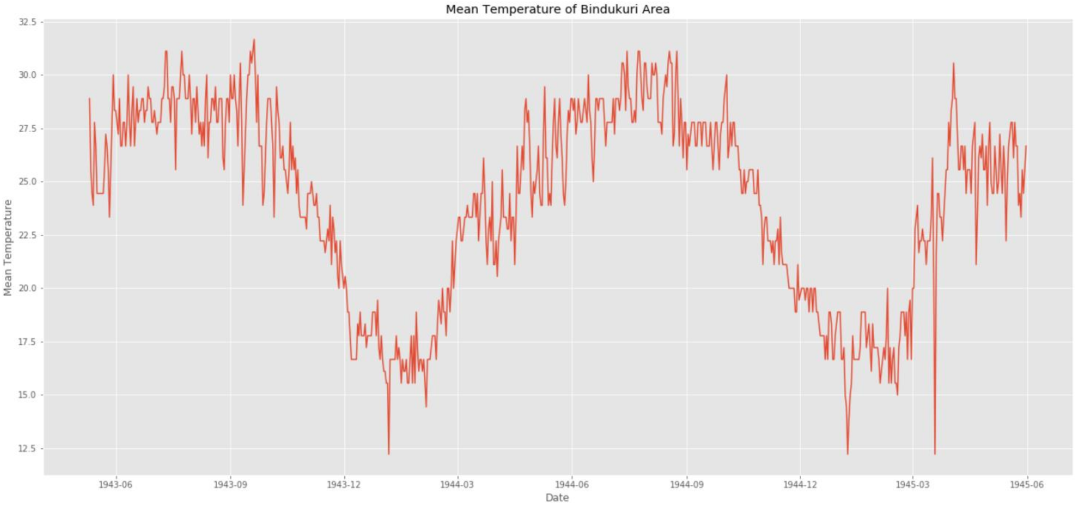
a. Aerial Bombing data includes a lot of NaN value. Instead of using them, I drop some NaN values. It does not only remove the uncertainty but it is also data visualization process.
Drop countries that are NaN;
Drop if target longitude is NaN;
Drop if takeoff longitude is NaN;
Drop unused features

b. Weather Condition data does not need any cleaning. According to exploratory data analysis and visualization, we will choose certain location to examine deeper. However, let us put our data variables what we use only.

4.Stationary Test

I list all important things as below in stationary test including stationary character, how to detect and solve.

Stationarity rule	Description	How to detect	How to solve
Constant mean	Follow one specific pattern	Pattern detect: duller test and kpss test	Differencing
Autocovariance does not depend on time			
Constant variance	Not stable	Variance and mean: rolling statistic	Transformation



One more thing, Rolling statistic is to smooth out short-term fluctuations and highlight longer-term trends. As you can see from plot above, our time series has seasonal variation. In summer, mean temperature is higher and in winter mean temperature is lower for each year. Now let us check stationary of time series. We can check stationarity using the following methods: duller test, kpss test and rolling statistic graph.

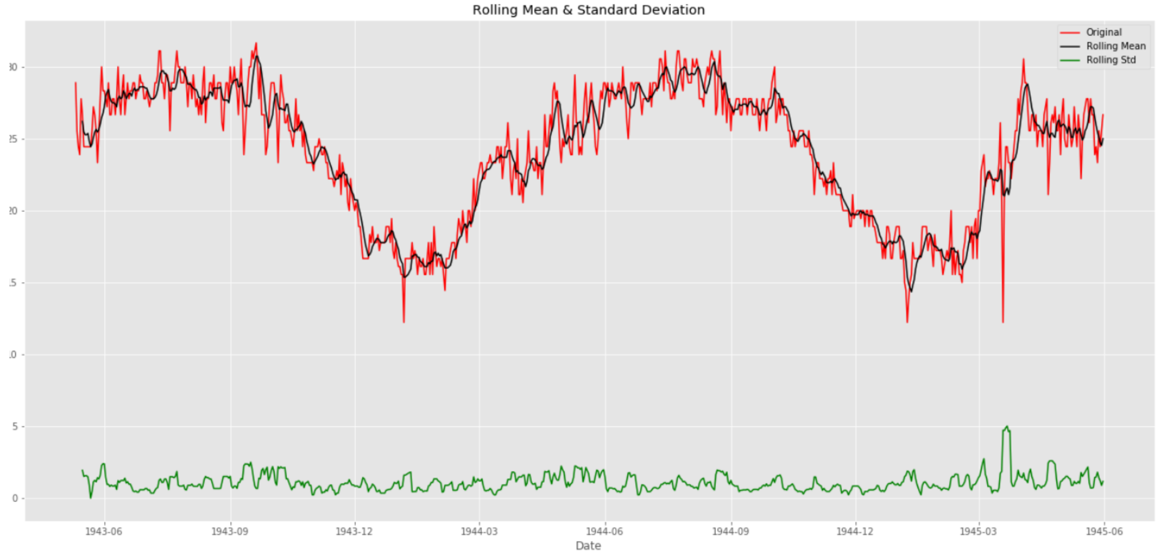
Plotting Rolling Statistics: We have a window lets say window size is 6 and then we find rolling mean and variance to check stationary.

Dickey-Fuller Test: The test results comprise of a Test Statistic and some Critical Values for difference confidence levels. If the test statistic is less than the critical value, we can say that time series is stationary.

KPSS Test: this is another test for checking the stationarity of a time series (slightly less popular than the Dickey Fuller test). The null and alternate hypothesis for the KPSS test is opposite that of the Dickey-Fuller Test, which often creates confusion.

Null Hypothesis: The process is trend stationary.

Alternate Hypothesis: The series has a unit root (series is not stationary).



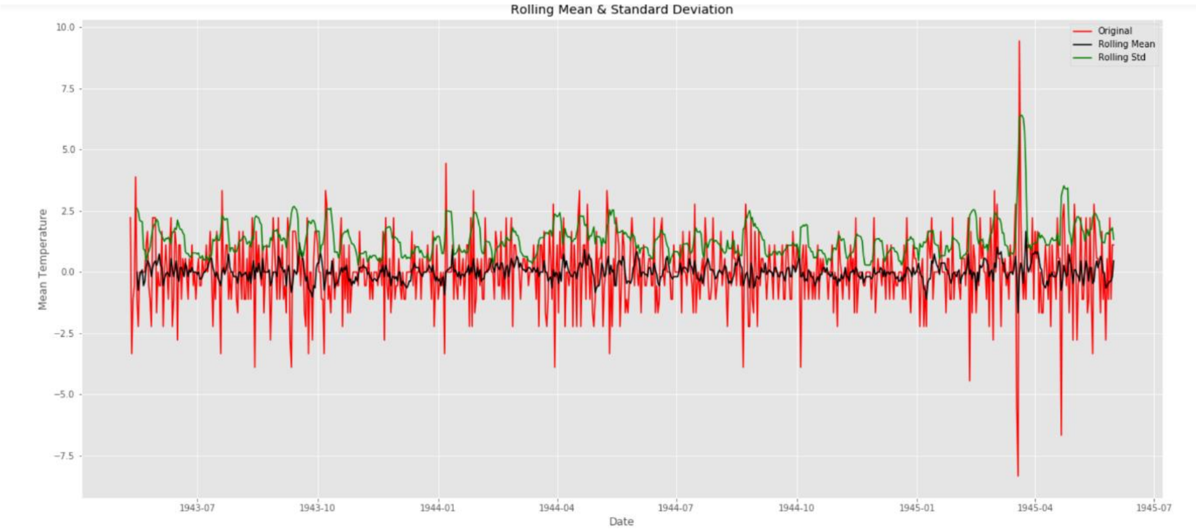
```
####################Dickey-Fuller test######################
Test statistic:  -1.4095966745887747
p-value:  0.577666802852636
Critical Values: {'1%': -3.439229783394421, '5%': -2.86545894814762, '10%': -2.5688568756191392}
####################KPSS test######################
Test Statistic            0.407370
p-value                   0.073978
Lags Used                20.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
dtype: float64
```

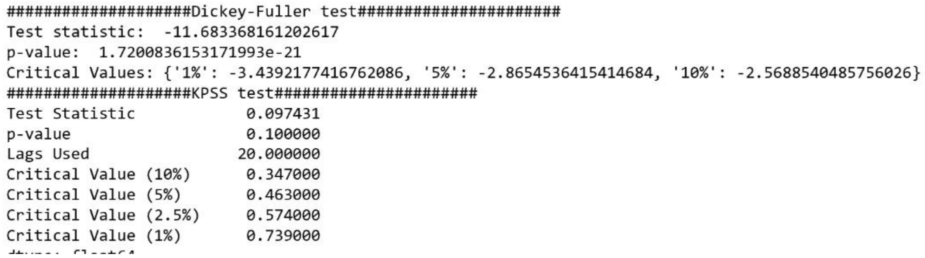
From the rolling statistic graph and statistical test, we find that the graph without constant mean did not pass the Dickey-Fuller test. So we have to deal with seasonal issue by differencing. Since variance appears to be stable, so we do not have to consider transformation except for differencing.

5.Dealing with non-stationary

Time series datasets may contain trends and seasonality, which may need to be removed prior to modeling. Trends can result in a varying mean over time, whereas seasonality can result in a changing variance over time, both which define a time series as being non-stationary. Stationary datasets are those that have a stable mean and variance, and are in turn much easier to model. Differencing is a popular and widely used data transform for making time series data stationary.

Differencing can help stabilize the mean of the time series by removing changes in the level of a time series, and so eliminating (or reducing) trend and seasonality. Differencing is performed by subtracting the previous observation from the current observation. The result after differencing is as below. Since rolling statistic graph appear to be stable and we pass both of statistical test, we can move on to the next part.





```
####################Dickey-Fuller test######################
Test statistic:  -11.683368161202617
p-value:  1.7200836153171993e-21
Critical Values: {'1%': -3.4392177416762086, '5%': -2.8654536415414684, '10%': -2.5688540485756026}
####################KPSS test######################
Test Statistic            0.097431
p-value                   0.100000
Lags Used                20.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
dtype: float64
```

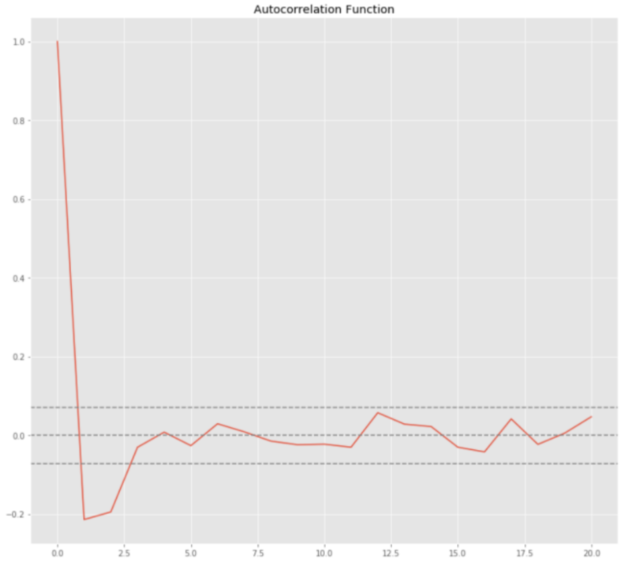
6. Model Development

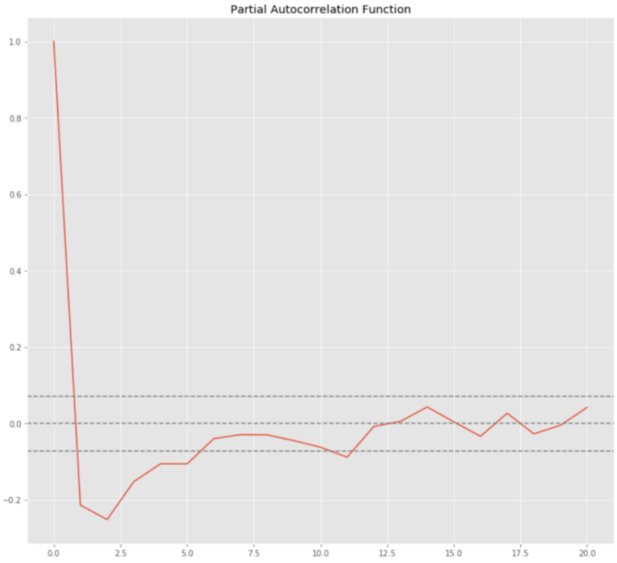
Our prediction method is ARIMA that is Auto-Regressive Integrated Moving Averages.

AR: Auto-Regressive (p): AR terms are just lags of dependent variable. For example, let us say p is 3, we will use x(t-1), x(t-2) and x(t-3) to predict x(t)
I: Integrated (d): These are the number of nonseasonal differences. For example, in our case we take the first order difference. So we pass that variable and put d=0
MA: Moving Averages (q): MA terms are lagged forecast errors in prediction equation.

(p,d,q) is parameters of ARIMA model. In order to choose p,d,q parameters we will use two different plots.

Autocorrelation Function (ACF): Measurement of the correlation between time series and lagged version of time series.
Partial Autocorrelation Function (PACF): This measures the correlation between the time series and lagged version of time series but after eliminating the variations already explained by the intervening comparisons. ACF and PACF graphs are as follows.





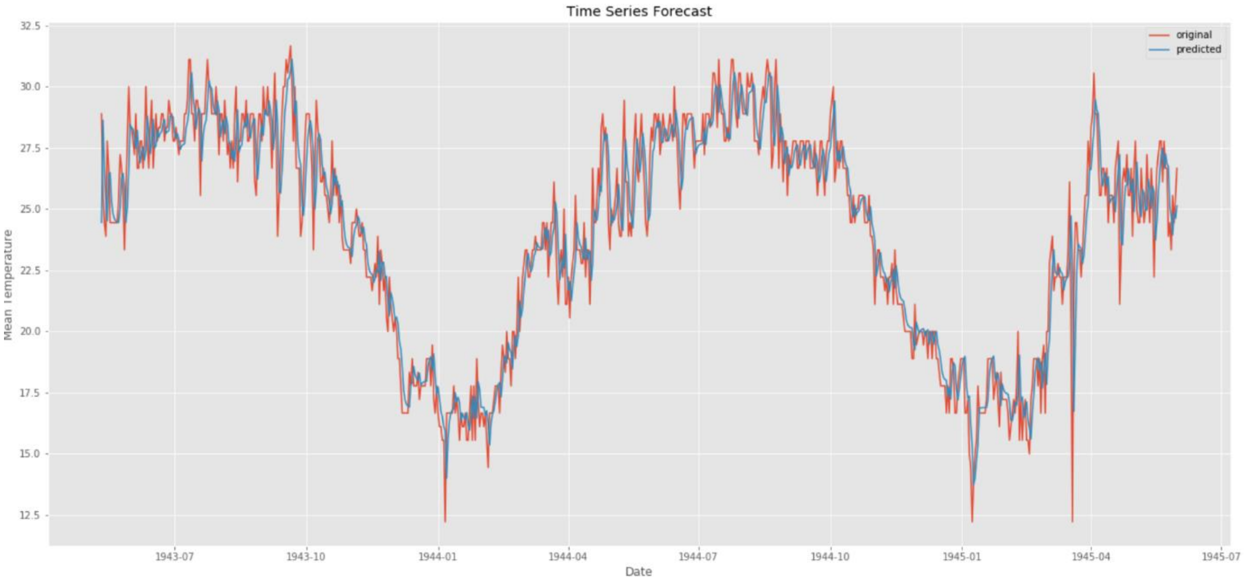
Two dotted lines are the confidence interevals. We use these lines to determine the ‘p’ and ‘q’ values

Choosing p: The lag value where the PACF chart crosses the upper confidence interval for the first time. p=1.
Choosing q: The lag value where the ACF chart crosses the upper confidence interval for the first time. q=1.

Now lets use (1,0,1) as parameters of ARIMA models and predict

Ⅲ. Prediction

Let us predict and visualize all path. It seems that our prediction makes sense. Do not forget our target: we can apply this model to predict bombing operation, sometimes happened when the temperature is high.


Time Series Forecast
original
predicted
Mean Temperature
Date