Data Driven Analysis of Major Coastal City Sea Levels in Japan

University of Kent

Author: Aridj Chenak (AC978), Joseph Florentino (JF491), Mitchel Berry (MB2011), Emanuela Vischetti (EV205) , Joshua Bhawanlall (JNB23)

MAST5957: Year in Data Analytics Project

Abstract

Through the use of 5 stations' data from the Permanent Service for Mean Sea Level dataset, we attempt to answer three research questions: 1) "How have the sea levels changed throughout the years in Tokyo, Japan?" 2) "How does this compare to the other three most densely populated cities in Japan?", and 3) "Is it possible to predict future tidal trends using the data available?". A literature review is considered to include this research's place in the wider literature. After data cleaning and EDA, analysis of seasonality and stationarity of the time series was carried out to understand the current trends. Autoregressive integrated moving average, and autocorrelation function were used to create a model for forecasting the Tokyo mean sea level trends in the next decade and attempt to answer the last research questions. Results show that the sea level values are slowly ascending, predictive models were created to answer the second and third questions. Implications for local populations and limitations of the dataset are discussed.

1. Introduction

The topic of sea levels and their rise is very topical today. Particularly in the contemporary political climate, where climate news and debates have been taking place and consumer choices are being flooded with green alternatives and advice columns on changes to implement to reduce individual carbon footprint, while corporations and big oil and gas companies continue polluting with little to no repercussions. It is important to understand and analyze the effect that pollution is having and consider the implications for populations. Our research attempts to utilize different academic fields to answer our research questions, which are outlined below.

1.1 The PSMSL Dataset

The PSMSL is the global data bank for long-term sea-level change information from tide gauges and bottom pressure recorders (PSMSL, 2022). Each station record comprises of two sets of data, each with four similar variables differing by frequency of collection:

- *Year/Year-Month*: indicates the year (annual dataset) or year plus month (monthly dataset) the sea level was recorded
- *Mean Sea Level*: average sea level obtained for the year/month in millimeters (mm)
- *Missing Days*: in the monthly data set, this variable is numeric and counts the number of days between the first and last day of the month with no data sea level data obtained; if set to 99, the value presented has been interpolated. On the other hand, the variable is a character “Y” or “N”: “Y” indicates the absence of 1-30 days worth of data in the year; if more than 2 months of data is deemed missing, the mean sea level value is set to -99999

- *Flag for Attention*: a mean tidal level value is presented instead of a mean sea level

Two thirds of the stations in the PSMSL database have been standardized by the definition of a common datum level of each station; the datum level is relative to 7 m below the mean sea level, and all the readings obtained from any period and tide gauge benchmark are related to this datum (AMS, 2012). This allowed for the data to be appropriate for the creation of time series values of sea level measurements. PSMSL made the formulation of this revised local reference (RLR) database possible through datum history supplied by the source authority. Another section of the PSMSL database consists of metric records, which have been advised against including in time series analysis by PSMSL. The stations from Japan investigated in this report were confirmed to be a part of the RLR database, therefore suitable for time series analysis (PSMSL, 2022).

1.2 Research Question and Hypothesis

In the brief for this analysis project we were given the instructions to perform analysis on the PSMSL Dataset. During February 2022 we submitted a project proposal outlining our initial plan in regards to carrying out said analysis. In this proposal the comprehensive aim for our project at the time was summarized: To investigate the tidal characteristics of coasts and their potential influence in Japan, Italy, and Scotland.

With this in mind, the research questions outlined in our initial project brief were:

1. “How have the sea levels changed throughout the years in major coastal areas?”

2. “Is it possible to predict future tidal trends using the data available?”
3. “Does a relationship between greenhouse gas emissions and sea levels exist?”

Investigating ‘sea level change in major coastal areas’ and ‘predicting future tidal trends’ initially required choosing specific locations with tide gauge stations that we could examine from the PSMSL dataset. Defining the common factor between these locations as countries with coastal cities that have “large amounts of coasts' and “coastal areas with high risks of effect from sea level” allowed us to choose locations based on these conditions. Within our initial project plan we concluded that the countries that aligned with our research criteria and that would be the focal point of our examination would be Italy, Scotland and Japan.

The correlation was based on our initial research, however after conducting further research (post proposal) we first found the evidence we used to choose these areas to be of low quality and not substantiated enough. Additionally there were not substantial enough common factors between these coastlines for us to only investigate these three regions, however investigating more regions than would not fit within the timeframe or the scope of our research project. Changing the regions our initial analysis focused on by altering our project's criteria to focus solely on a single location, allowed us to alter the project to fit the timeframe and scope of the analysis project, without substantially changing the initial focus of investigating the tide characteristics and trends in the sea level.

Japan was chosen because it is an island with a large amount of coastline, and high concentration of stations. 139 tide gauge stations are listed within the PSMSLs scope for Japan's 35,000 km of coastline (REF). In conjunction with this it was also the most interesting to us as a group choosing from our initial locations. We focused our analysis on the station in Tokyo and made comparisons with the top 3 most populated cities in Japan. Using this source (REF) we found that the top 3 cities by population in Japan were Tokyo, Osaka, and Nagoya. We looked at the time series data for stations within these areas.

In the initial plan, the third research question was used to compare a second database containing regional greenhouse gas data to the same areas we were looking at sea levels at in order to compare and model a possible relationship between these two variables. However, after further research we realized it was impossible for us (within our limited understanding of the processes of emissions and relative effect, and the scope of the project) to analyze the effect that local greenhouse gas emissions had on a specific countries/cities mean sea level as it has a holistic global effect. We decided to instead prioritize the comparison between the cities within Japan, as within the literature search we learned the potential scale of impact on the population due to the incredibly high density population of the island.

Considering these changes, our updated research questions were:

1. “How have the sea levels changed throughout the years in Tokyo, Japan?”
2. How does this compare to the other three most densely populated cities in Japan?”
3. “Is it possible to predict future tidal trends in Tokyo using the data available?”

2. Literature Review

The literature on the topic of sea levels is plentiful and significant. There are lots of themes and studies spanning decades of activity. Similarities and differences portraying the phenomenon and all its surrounding aspects in different ways according to the region of the world, climate, season etc. This creates a plethora of information available to analyze and explore. However, with that significant amount of information, comparison and critical thinking are needed to recognise which results are faulty and which are most closely related to reality. In this review, the general themes of sea levels and their rise will be explored first, then more specifically, studies into rising sea levels in Japan and the literature surrounding it.

Rising sea levels are one of the most well-known and discussed effects of climate change, they are often brought up in climate contexts and conversations. This is reflected in the literature, where we notice a significant amount of studies, research papers, and reviews attempting to explore questions of human impact and of flora and fauna endangerment among others. A major theme in the literature includes the investigation of sea levels in different areas of the world. A significant amount are focused around North America, such as the Gulf of Mexico and US atlantic coasts (Knowles, 2010; Orson, Panageotou and Leatherman, 1985; Comeaux, Allison and Bianchi, 2012). Other areas touched upon in studies are the Vietnam coast and the English channel, for example (Smajgl et al., 2015; Haigh, Nicholls and Wells, 2011). Besides the geographical similarities and differences, one thing that unites all these studies is that the main focus and objective of the research is the consideration of impacts of rising sea levels on local populations (ibid; Nicholls, 2002). However, differences arise in the more specific methods of exploration: one study “identifies policies” and containment measures for rising sea levels, another just peruses “historical changes and predicted changes”, another again uses data to “create a model for extreme water levels” (ibid; Hall et al, 2019). The last two look at marshlands and the biological changes

created by rising sea levels and the effect on flora and fauna of seeping saltwater into freshwater environments.

Research more specific to the rise in sea levels rather than on the levels generally includes, similarly to previous studies mentioned, the significant focus on the coastal impacts of such a phenomenon. By coastal impacts, we mean discussion of "sediment shifts", microflora upset, marsh submersion, human culture and population effects, "saltwater penetration into coastal aquifers", infrastructure risk and erosion, among others (FitzGerald et al, 2008; Cazenave and Cozannet, 2014; Nicholls and Cazenave, 2010; Morris et al, 2002; Kirwan et al, 2010; Gornitz, 1991; Heberger et al, 2009; Gornitz, 1990). Differences in this theme in the literature at this point include, also similarly to the previous theme discussed, the approach taken to the exploration of the theme of sea level rise. Some of these explorations utilize projections, create theoretical models, coastal risk databases, and predictions gathered from past data (ibid). This shows the intent of the researchers to look at the future for possible trends and to prepare and inform policy if necessary. The addition of the detailed study of populations and contexts, along with the collection of primary research studies in the form of reviews present in the body of literature adds an additional layer of reliability to the results regarding the effects and risks associated with rising sea levels and, more specifically, risks associated with coastal areas (ibid). Another, more general study into rising sea levels looks at the impact of this natural change on the "spread of infectious vector-borne diseases" and the effect that a global temperature rise could enhance this negative process (Ramasamy and Surendran, 2011).

The importance of these themes builds a foundation of knowledge off of which we can begin basing our research: this means that we are aware of the themes covered by previous academics, that we are aware of roughly what information can be gathered, but also simply what interests and what are the current issues researchers are attempting to fix or debate. From this brief summary we know that coastal areas around

the world are of significant interest to climate and sea level researchers, seeing as to the amount of research present on it; we also know that the human aspect and risks associated with rising sea levels are of interest, we will touch upon that in later sections.

In regards to sea levels and trends in the Japanese context, the one we are considering in this report, some interesting themes and results come up in the literature. There seems to be a commonality between studies where the Japanese coast is divided into areas with specific sea level characteristics and trends. Nakano and Yamada discuss the presence of five regions “of similar mean sea level deviations”, they did so using the Permanent Service for Mean Sea Level, the same dataset we are using (1975). The three stations we are considering from the PSMSL are multiple in Tokyo and one each in Nagoya and Osaka, the results discussed in Nakano and Yamada indicate that the mean sea level is lower than in the west coast of Japan and part of the Southern coast. Another study, carried out by Shoji came to the same conclusion, claiming as well that the Japanese coast is divided into “distincts” which follow similar internal patterns (1961). Another conclusion the authors come to is that the main cause of this deviance is the “effect of the deflecting force of the earth's rotation on currents”. Similarly, Oh et al. discuss the seasonal sea level oscillations in the Japan Sea and, using the PSMSL, find significant oscillations between summer and winter, making up the majority of the oscillation. However, Church et al. in a more recent study disagreed with this claim, they discussed more in-depth about the “major contributions to 20th and 21st century sea-level rise [being] a result of ocean thermal expansion and the melting of glaciers and ice caps” (2008). Kang et al., agrees with this concept, finding results that indicate that “sea level rise in the EJS is mainly due to thermal expansion” (2005). As a result of these rising sea levels, an increasing amount of studies set in Japan have discussed coastal loss due to SLR, such as Udo and Takeda, where they projected future beach loss, and Banno and Kuriyama, who developed a shoreline model and found that shoreline would retreat about 20 meters by the end of the century (2017; 2014). Finally, Cao, Esteban and Mino adopt

these results of shore retreat into a discussion of their possible effect on infrastructure, more specifically wastewater treatment plants (2020).

Overall, our report will add to the literature in a significant way by utilizing the PSMSL, the most thorough, coherent and long-running dataset providing information on sea levels, to answer our research question. The information provided in the literature informed our questions partially as well by providing us with background information on the area of the world we considered. The consideration of implications to local populations described in the literature will also inform some of our own considerations.

3. Methodology

3.1 Choosing Stations for Analysis

We determined that there were 6 stations in the top 3 cities by population in Japan, and every city chosen had one or more tide gauge stations. We disregarded the non datum controlled data (low quality datasets with a quality control flag on the PSMSL website) which included the 'Nagoya' station. With these exceptions each region had at least one station recording reliable tide gauge data, and by this method we found 5 stations which we defined as the top 5 tide gauge stations by surrounding population. The monthly and yearly mean sea level data from these stations would form the basis of our analysis. Regions that used more than one station to record data (due to discontinuation of stations recording in the same area) were counted as one station i.e Tokyo I,II,III counted as a single station for Tokyo.

The top 5 cities in Japan by population, and the respective stations used:

1. Tokyo City - Tokyo I, II, III - Station ID 881, 1222, 1545
2. Osaka City - Osaka - Station ID 1099
3. Nagoya City - Nagoya II - Station ID 1498

3.2 Pre-processing the PSMSL Data

In order to do any analysis the data was first imported into R, formatted as a dataframe and cleaned.

To import the data, R was used to directly access each station's csv data via the PSMSL website. Manipulating a specific section of a predefined url by altering to different station codes (a unique identifier for each station the PSMSL website) allowed us to download a comma delimited text file for each station. This CSV file was then stored within a variable as a data frame within the R Markdown code, for each region of interest a separate data frame was created within a markdown file i.e. Tokyo I,II,III stored in one data frame file Tokyo.

The four variables contained in each of the station data frames are “YearMonth", "Mean_Sea_Level", "Missing_Days_Flag","Flag_For_Attention”. As Year and Month were combined into one variable it needed to be split into two in order to do further analysis. Documentation on the steps necessary to perform the operation to separate the combined YearMonth variable was listed within the PSMSL explanation of the dataset. This conversion was done in R through splitting the dataframe by index, converting the character values to numeric variables, and performing the necessary operations.

Once the variables had been split the next step was to clean the data frame. To achieve this mean sea level values of -99999 were removed (and replaced with NAs),this value denotes two month or more of missing values. We included data that had up to one month of missing values. Two variables related to the quality control (and by proxy cleaning) of the dataset. There were no instances of a ‘Flag_For_Attention’ as we removed any stations with quality control flags. In regards to the variable ‘Missing_Days_Flag’ any interpolated data (labeled as the value

99) was kept as the level of interpolation used would be higher than the scope of our project and we aimed to not to remove any information that could be valuable to our analysis.

3.3 Methods of EDA

Dataset Summary

In order to tackle the numerical summaries all the stations data frames were converted to an eXtensible Time Series(xts) so it could be merged into one data frame and call the "profiling_num" function which gives insight into the stations: average mean, standard deviation, variation coefficient, percentiles (01, 05, 25, 50, 75, 95 and 99), skewness, kurtosis, interquartile range, range(80, 98).

Time Series Plots

Each station's dataframe was converted into a time series object. Using the library 'xts' the time series objects were converted to 'extensible time series objects', replacing the time series index with time values. Regions with one active but multiple discontinued stations, like Tokyo I, II and III were combined using two different methods. The first method merged the extensible time series data into multiple columns allowing the one region's separate stations to be stored in a single 'xts' object. The second method merged this same data into one column allowing the time series to be compared with other regions' stations in EDA. For analysis and prediction however, a 'ts' object was strictly needed; therefore, the three Tokyo stations were also bound into a single time series. Extensible time series plots were chosen to visualize the

time series data. These two methods achieved two individual types of plots, one to inspect the mean sea level of a region by itself, and one to compare multiple regions in the same plot.

3.4 Methods of Analysis

Imputation of Missing Data

The methods of time series analysis proposed require the dataset be complete with no missing values. The combination of Tokyo I, II and III revealed a void of data records in the years 1964 - 1968 inclusive, which was during the transfer of station I to another area in the same region. Simple removal of such a large fragment of data would result in major inaccuracies in predictions and analysis. Imputation of the void is hence done through seasonal splitting due to the potential seasonality of the time series: the Tokyo time series is split into multiple seasons; each season is then imputed separately (Moritz, 2021).

Seasonality and Stationarity

When looking at the time series, if it can be determined that there is both a trend and seasonality, the trend must be removed in order to analyze the seasonality. The seasonality in regards to our analysis is recurring mean sea level patterns across seasons (a.k.a seasonality) (Anish, 2020). This was firstly examined in the time series plots. To remove the trend we can use differencing on the original time series which creates a difference-stationary time series which can be stored in a separate object.

An ADF or Augmented Dickey-Fuller test can be used to test the p-value of the initial series, which numerically verifies stationarity. If the p-value > 0.01 then the series is not stationary and differencing can be used. Performing an ADF test on each iteration of differencing until the p-value reaches 0.01 allow's the difference-stationary time series to be easily identified.

In our case a seasonal plot and a sub series seasonal plot was used to visualize the stationary data. In addition to this the seasonality is also investigated through box and seasonal plots, whereby the time series values are plotted against a season for each year. This allows for the identification of any seasonal trends.

Decomposition

Decomposition is used to separate the time series data into three components: its trend, its seasonality and its irregularities. Plotting this decomposition allows us to visualize these three components that form the time series individually, the product of which is the original time series. The method of decomposition chosen is X11 decomposition. It is capable of handling both additive and multiplicative types of time series sparing the difficulty and time to determine the type of each time series using R. Seasonal variations are also made further obvious in the seasonal plots produced in comparison with classical decomposition (Dagum & Bianconcini, 2016).

Modeling Identification

To model the data, three methods of time series analyses were employed. Seasonal Naive (Benchmark Method), Exponential Smoothing (ETS) and ARIMA methods were used to create models from stations time series data. Comparing three models allowed us to iterate through fits until we found the model that suited the data the best, and allowed for more accurate statistical forecasting.

The Seasonal Naive model benchmarks by forecasting the future values by setting it equal to the last observed value for the same 'season' of the year Sultana (2018). Exponential Smoothing uses individual weightings of each data point when calculating set period moving average of the time series data in order to extract the trend and thus forecast. ARIMA models or AutoRegressive Integrated Moving Average are another type of univariate time series forecasting that uses a time series past values (lags/ lagged forecast errors) to forecast future values, it also requires seasonal data. It is a 'Auto Regression' linear model that uses its lags in order to make predictions. The parameters are 'p' - order of the AR term, 'q' - order of the moving average term MA (Prabhakaran, 2022).

For the Seasonal Naive and ETS method the first difference stationary series was used, for the ARIMA model the original time series was used as the first difference series was automatically calculated within the method. In order to find the Seasonal Naive fit snaive() was used, for ETS the ets() function was used and to iterate through models and find the best ARIMA

fit auto.arima was used. Comparisons between the residual plots of each of the models allowed us to determine the closest 'fitting' model.

Autocorrelation Function (ACF)

The AutoCorrelation Function (ACF) is plotted in order to determine the correlation coefficient between two values within a time series. In our case measuring the degree of similarity between the mean sea levels current monthly time series values and the lagged past values for each month over time. The function requires no NA values within the Data.

This method requires using stationary time series data and aids our understanding of the seasonality within the data and identifying patterns.

In the first iteration of our modeling process we intended to only use a manual ARIMA method and the ACF/Partial ACF (PACF) was to be used to select the type of ARIMA method we would use, however this was mitigated in further iterations by using the 'auto.arima' function which fitted the best ARIMA model using 5 parameters. This function could then be used to plot residual plots, the ACF graph contained in these residual plots would be a factor used to visually compare the fits of the models. The auto function automatically determines the best 'p' and 'q' values for the ARIMA model. If the coefficient falls outside of the 95% confidence interval at any lag interval, there is a statistically significant result that can be examined/discussed.

Forecasting

We used the library ‘forecast’ in order to use the models to forecast the future data. Forecasting by the seasonal naive method would yield fewer years of prediction than ARIMA and ETS.

4. Analysis and Results

4.1 Exploratory Data Analysis

Numerical Summaries

To gain an overview of the data EDA was firstly conducted on the data for the three stations that were chosen (Tokyo_Mean_Sea_Level is a combination of all three Tokyo stations). The average mean for Tokyo lies around 7.011m with not much deviation between the years as the standard deviation is about 0.09m which demonstrates that the data is not very volatile and quite consistent with increases/decreases of the sea levels. In comparison, Osaka seems to have a slightly higher sea level average out of the three and also a higher standard deviation (that is statistically insignificant enough that it would not influence the data by much either way).

The skewness for the stations showcases that the data is fairly distributed as the values are in the range of -0.5 and 0.5 and with a negative excess kurtosis (platykurtic distribution) demonstrates that stations won't have any extreme positive or negative values, further proving the standard deviation values that all the sea levels means are quite close in range.

The difference between the 25th (p_25) and 75th(p_75) percentile is only 0.137 m (iqr) which is slightly above standard deviation.

```
##                  variable     mean   std_dev variation_coef    p_01   p_05
## 1  var.Tokyo_Mean_Sea_Level 7011.035  99.92457     0.01425247 6779.48 6842.6
## 2  var.Osaka_Mean_Sea_Level 7054.468 151.68481     0.02150195 6711.83 6801.0
## 3 var.Nagoya_Mean_Sea_Level 7017.298 124.64104     0.01776197 6776.72 6822.0
##      p_25 p_50   p_75   p_95    p_99    skewness kurtosis    iqr
## 1 6947.00 7011 7084.0 7167.4 7233.92 -0.14896626 2.699927 137.00
## 2 6948.75 7050 7161.5 7302.0 7372.34 -0.06002842 2.597972 212.75
## 3 6920.00 7019 7104.0 7226.2 7303.80  0.11225833 2.314068 184.00
##             range_98     range_80
## 1 [6779.48, 7233.92] [6876, 7139]
## 2 [6711.83, 7372.34] [6862, 7258]
## 3  [6776.72, 7303.8] [6849, 7180]
```

Figure 1 - Numerical Summaries of Tokyo, Osaka and Nagoya II Stations

Time Series Plots

Tokyo

Tokyo, the capital, is the most densely populated region in Japan. The sea level is recorded by three stations, two of which have been decommissioned. Tokyo I station recorded data from 1957 - 1963, Tokyo II station recorded data between 1968-1982, and Tokyo III station recorded data from 1982 - 2021.

Figure 2 - Monthly Mean Sea Level in Tokyo, 1957 - 2021.

The time series plot above uses different colors to denote the time period of each station activity, it is visually evident that with each new station that records data for Tokyo, the station's lifespan increases.

Comparisons with other Station

Tokyo, when compared to the station in the city with the next highest population, Osaka.

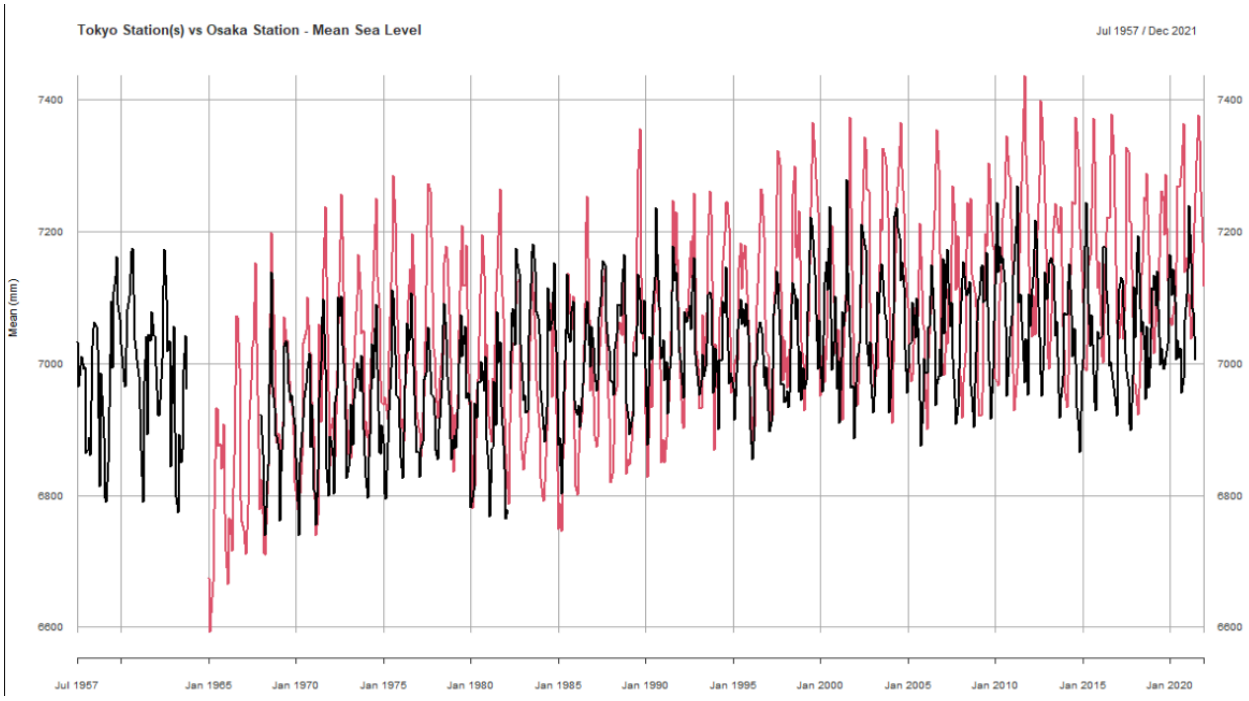


Figure 3 - Tokyo vs Osaka, Monthly mean sea level.

This plot demonstrates that Tokyo's time series data spans a longer range than that of Osaka although it contains NA values whereas Osaka does not, as a result of the data points for Osaka looks more spread out and the Tokyo data points more grouped up. Additionally, it shows that on average Osaka has a higher mean sea level. Lastly, the sea level for Tokyo tends to be a more steady increase/decrease whilst Osaka's seems to be more volatile as it can reach a new peak record and then drop 200mm the following year.

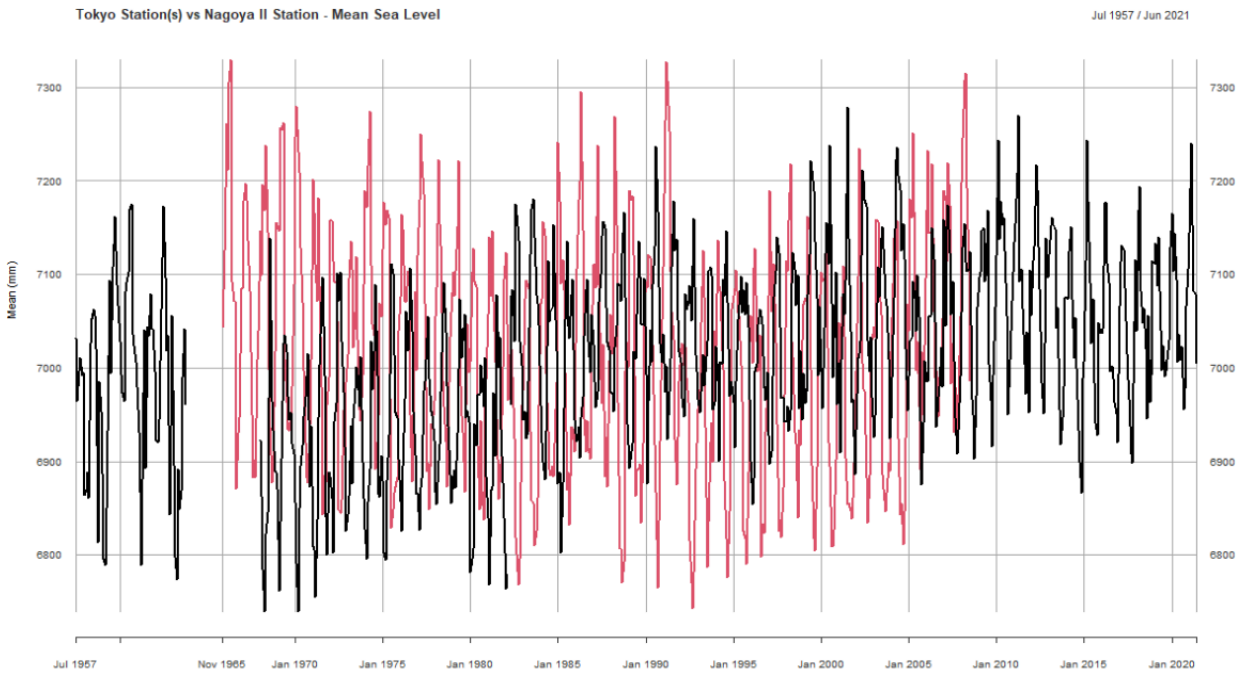


Figure 4 - Tokyo vs Nagoya II, Monthly mean sea level.

This plot shows us that Nagoya is a station that does not obtain data regularly and can sometimes be unreliable as it has not recorded new data since 2008. Reason being, that although Nagoya has two stations it could not be merged the same way as Tokyo's station because Nagoya I had been marked with a “Quality Check Flag” in the PSMSL website.

Onto the comparison and similarities between the two cities they both have a clear and steady trend, that the Osaka station did not have as the trend was an ascending one. However, although most of the data from Nagoya II seems to follow a trend it has quite a few years in which the data points deviate quite a lot from the normal trend.

Seasonality

An initial investigation of the seasonal effects in the combined Tokyo stations is done through the observation of its time series monthly box-plots across the years shown in figure 5 plotted alongside the boxplots of the Osaka and Nagoya II stations for comparison.

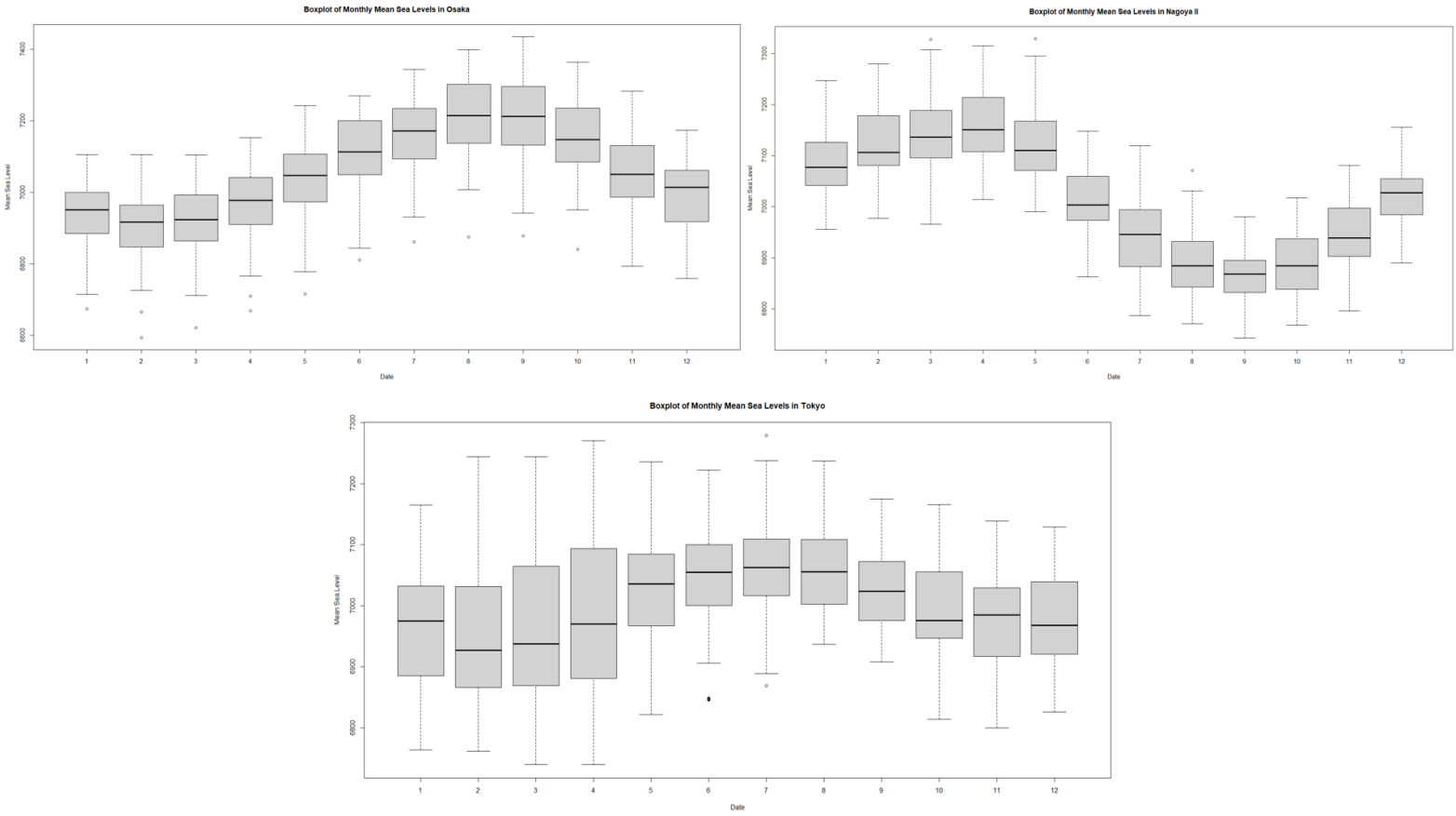


Figure 5 - Monthly Mean Sea Level Boxplots - Top: Osaka and Nagoya II; Tokyo (Bottom)

Considering box height, Tokyo demonstrates major variance in mean sea level throughout the months January to April, and becoming less varied the rest of the year on average. A low mean sea level is also often recorded in the first 4 months of the year, followed by an incline during summer, and decreasing again in the autumn; a slight increase occurs during the winter entering the new year as well. A similar seasonal effect, but less varied, is shown by Osaka,

whereby the mean sea level lowers at the beginning of the year. In this station however, the increase is observed from spring and into late summer, peaking in September instead of July. A decrease in sea level during autumn through winter is observed in this station as well.

Nagoya II on the other hand showcases a trend that might be considered opposing to both the Tokyo and Osaka stations. Unlike the previous stations, Nagoya II on average records an increase in mean sea level from the start of the year till spring; a declining trend follows in summer, and into autumn and winter.

4.2 Analysis of Stations Time Series Data

Imputation

As stated in the methodology section, seasonal split imputation was done to attempt to fill in the lost mean sea level data for the years 1964 - 1968 in the combined dataset for the three Tokyo stations. Figure 6 shows the filled gap in the Tokyo time series.

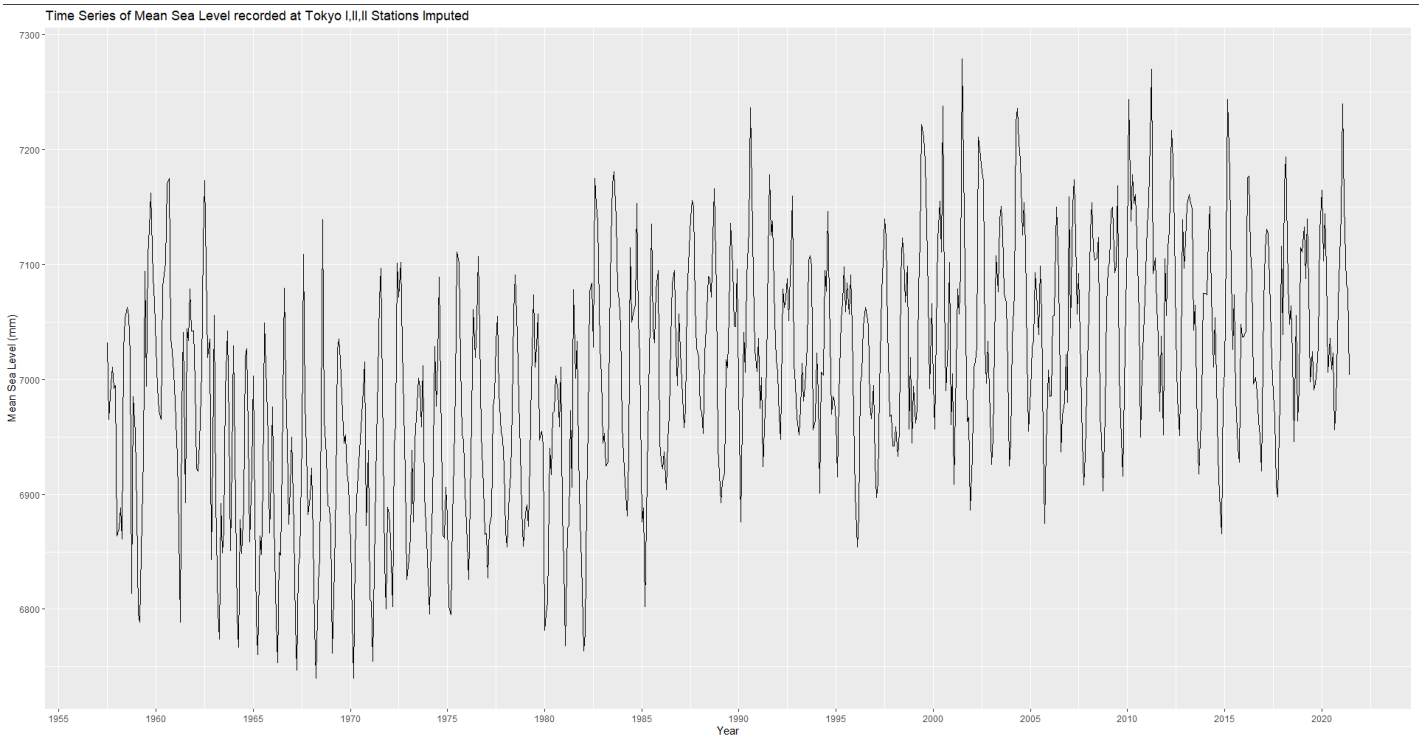


Figure 6 - Imputed Time Series Plot of the Mean Sea Levels from Tokyo Stations I, II and III

A brief visual interpretation of the imputed line suggests an overall decreasing mean sea level trend beginning in 1964 and ending in around 1966 with constant seasonal amplitude, and spiking up once again into 1968 with an increasing seasonal amplitude. The time series waves that fill the gap appear relatively consistently patterned in comparison to the recorded values

perhaps due to the nature of the imputation algorithm producing artificial and potentially ideal values.

Stationarity - Autocorrelations

Prior to further seasonality analysis, and later decomposition, it must be identified whether the Tokyo time series is non-stationary, which allows successful differencing into a stationary time series, a crucial step in the creation of seasonal plots. A visual assessment of stationarity is shown in figure 7 as a correlogram (a.k.a autocorrelation graph).

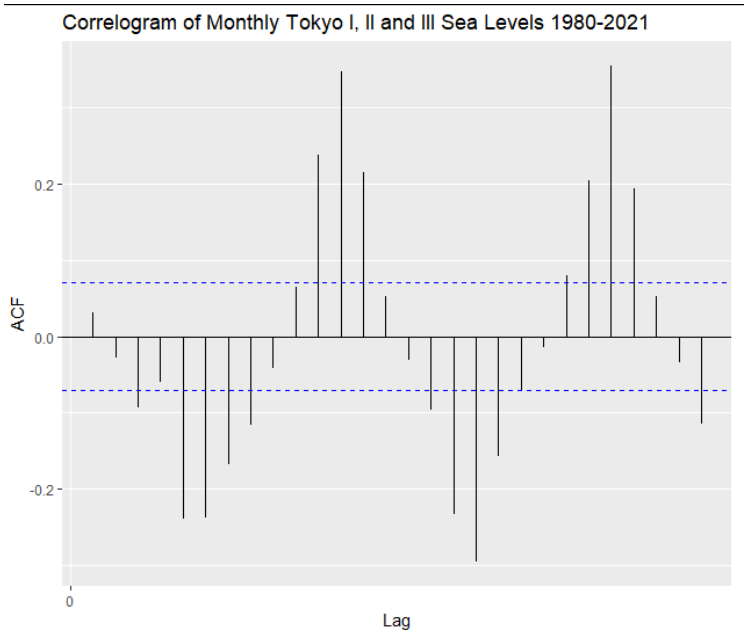


Figure 7 - Tokyo I II II Autocorrelation Graph

A notable observation in the graph is that the ACF value does not instantly drop to 0, but rather gradually, which is an indicator of a non-stationary time series (Kwiatkowsi et al. 1992). This, however, is not a satisfactory indicator, which leads to the next stationarity test result in the upcoming section. ACF plots for Nagoya Osaka can be found in appendices O and P respectively.

Augmented Dickey-Fuller Test

```
        Augmented Dickey-Fuller Test

data:  tokalltc
Dickey-Fuller = -3.3803, Lag order = 12, p-value = 0.05684
alternative hypothesis: stationary

Warning in adf.test(difftokyo, k = 12) :
  p-value smaller than printed p-value

        Augmented Dickey-Fuller Test

data:  difftokyo
Dickey-Fuller = -13.36, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary
```

Figure 8 - Example of ADF test with Tokyo time series vs single difference time series.

In order to guarantee that differencing the time series data for a station results in stationarity, and to work out how many times the series needs to be different ADF tests were used. Above, the tokyo time series variable ‘tokalltc’ returns a p-value of 0.05684 showing that it is not stationary. After differencing it once, variable ‘difftokyo’ it can be seen that the p-value returned is less than 0.01, resulting in confirmation of the stationarity of the differenced time series data.

Seasonality and Sub Series Seasonality Plots

The inconsistency in seasons showcased in the seasonal plot in figure 9 demonstrates the variation of seasonal trend recorded throughout the years by Tokyo I, II and II; this is in contrast to the seasonal plots produced by Nagoya II (Appendix L) and Osaka (Appendix K). Various

overlapping rising peaks in mean sea level can be seen during May, and many declining peaks during October and November. This conforms with the boxplot observations made in EDA.

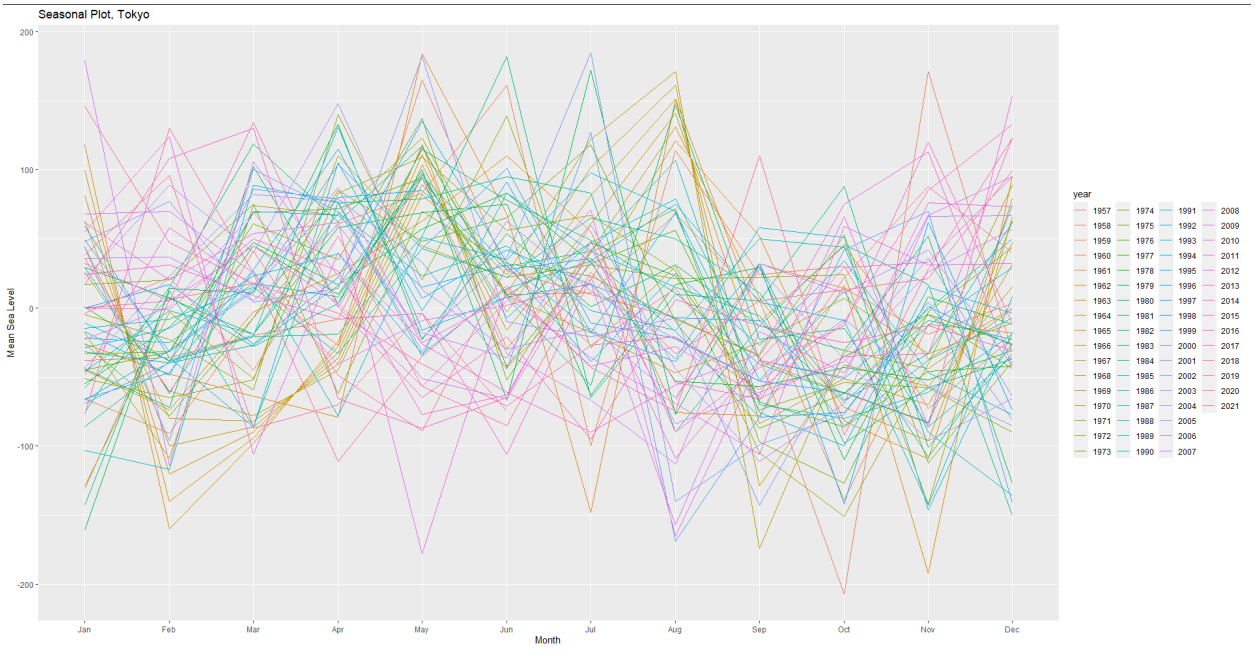


Figure 9 - Seasonality Plot, Tokyo

In the subseries seasonal plot shown in figure 10, the large seasonal amplitudes make the seasonal nature of the time series less obvious. Nonetheless, it can be seen that the pattern observed in both the seasonal plot and boxplot is resulted here as well. The same conclusion can be applied to the Nagoya II (appendix M) and Osaka (appendix N) subseries seasonal plots.

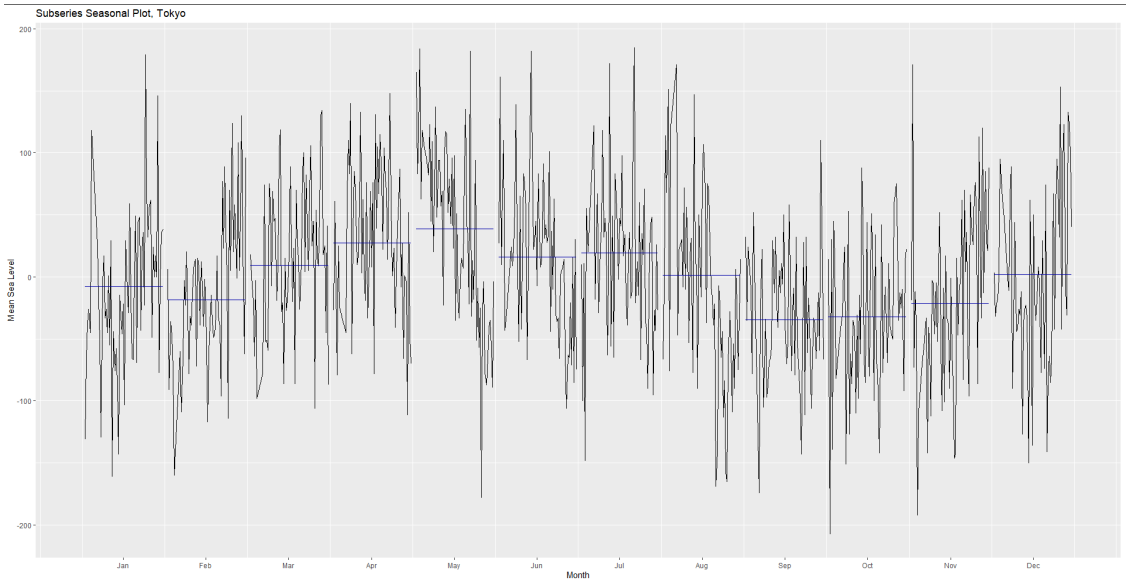


Figure 10 - Subseries Seasonality Plot, Tokyo

Time Series Decomposition

The Tokyo station time series decomposition (figure 11) splits the time series into its trend, seasonal component and remainder component.

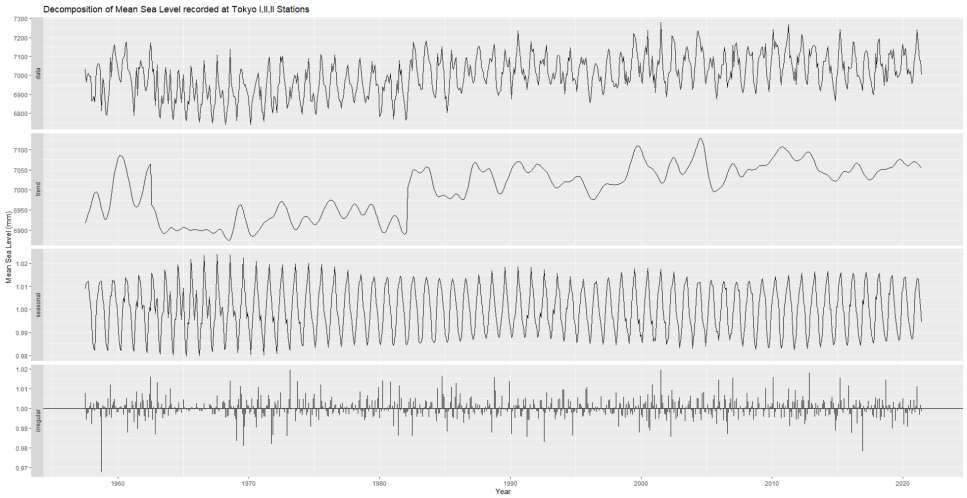


Figure 11 - Decomposition of the Tokyo Time Series Data.

The seasonal component shows that the mean sea level seasons changed drastically in the initial 20 years recorded, but remained more consistent in years to come. The next significant component is the trend, isolated and layered over the combined time series in figure 12.

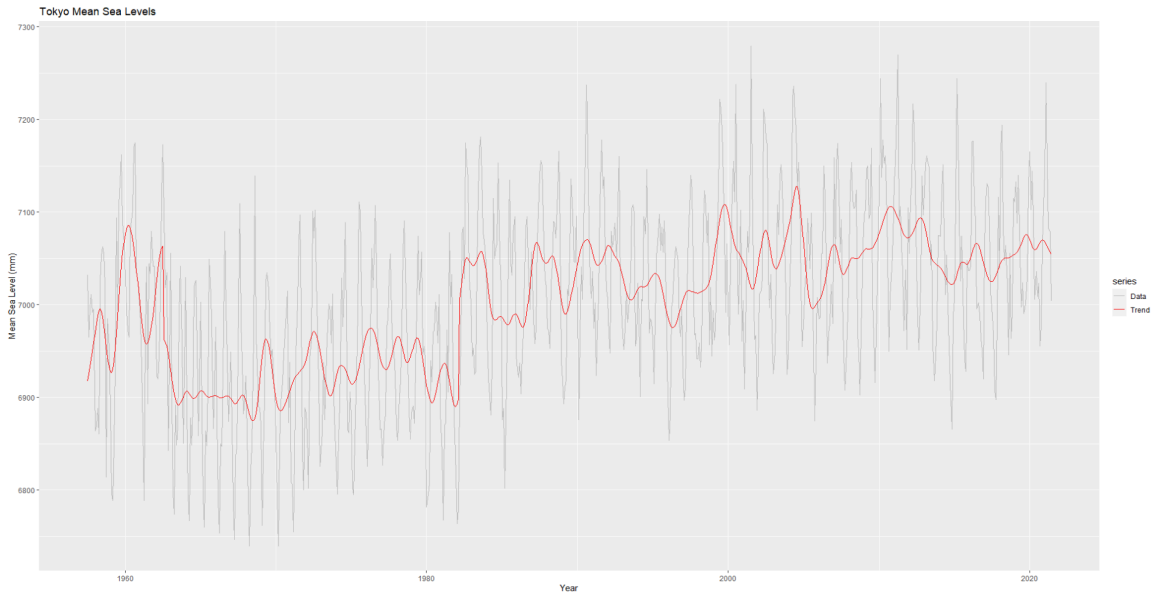


Figure 12- Tokyo Time Series Trend Layered atop its Time Series Plot

A negative mean sea level trend is initially observed in the first ~5 years which later remains constant between ~6880 mm and 6975 mm for ~17 years. A sharp peak occurs after, followed by what seems to be an increase reaching an overall mean sea level above previous years.

The decomposition graph for Osaka (appendix I) showed a trend with higher positivity throughout the years, while Nagoya II (appendix J) did not show a distinguishable pattern. Both, however, share a consistent seasonal trend.

Modeling - Seasonal Naive, ETS, ARIMA

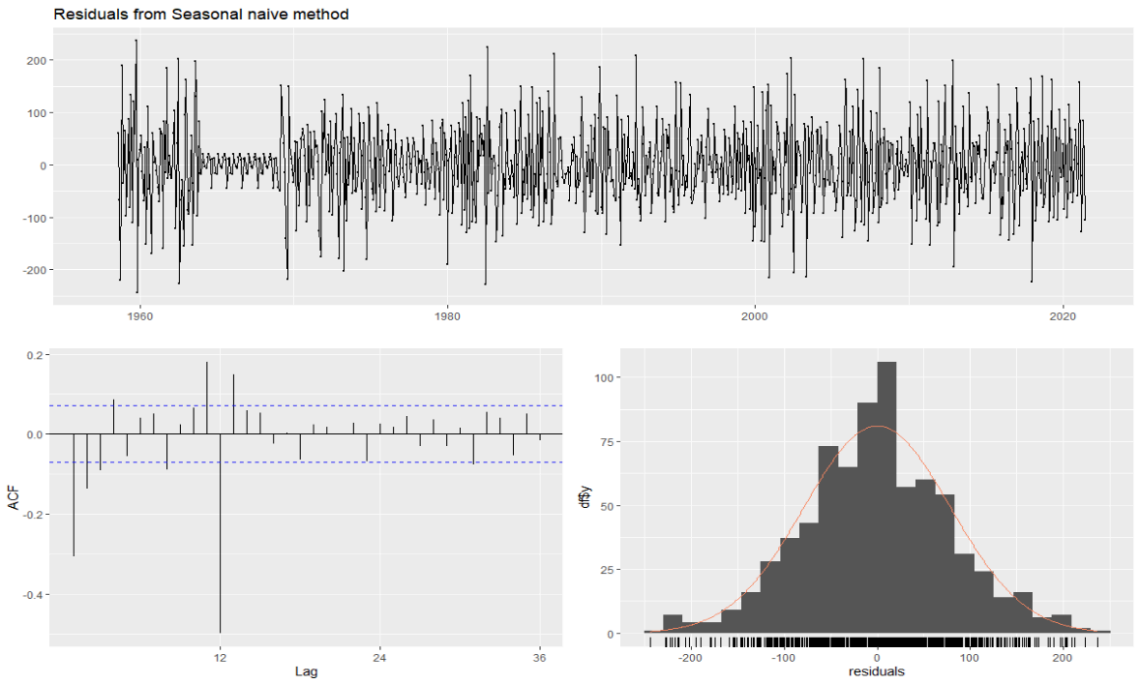


Figure 13 - Residual plots by modeling with the Seasonal Naive Method, Tokyo.

```
Forecast method: Seasonal naive method

Model Information:
Call: snaive(y = difftokyo)

Residual sd: 78.9422

Error measures:
                     ME     RMSE      MAE MPE MAPE MASE       ACF1
Training set -0.08344371 78.94221 61.15629 NaN  Inf    1 -0.3073381
```

Figure 14 - Forecast model output, Seasonal Naive Method, Tokyo.

```
Forecasts:
> checkresiduals(fittokyo)

        Ljung-Box test

data:  Residuals from Seasonal naive method
Q* = 360.14, df = 24, p-value < 2.2e-16

Model df: 0.   Total lags used: 24
```

Figure 15 - Ljung-Box test, Seasonal Naive Method, Tokyo.

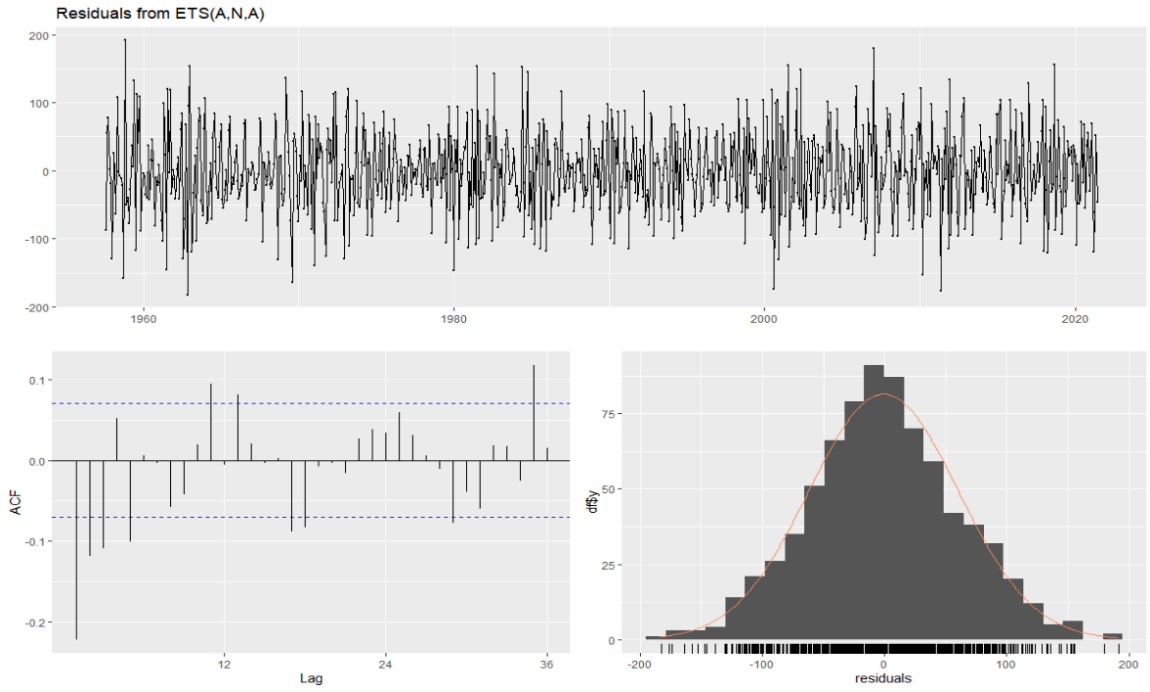


Figure 16 - Residual plots by modeling with the ETS, Tokyo

```
ETS(A,N,A)

Call:
 ets(y = difftokyo)

  Smoothing parameters:
    alpha = 1e-04
    gamma = 0.1862

  Initial states:
    l = 6.1247
    s = 12.0041 22.153 51.2372 30.339 15.1593 -25.2958
           -8.2382 15.3967 -27.191 -70.1032 -29.5598 14.0986

  sigma:  61.6419

     AIC     AICc      BIC
11432.80 11433.44 11502.43

Training set error measures:
                    ME     RMSE      MAE MPE MAPE      MASE       ACF1
Training set -0.571199 61.07673 48.18747 NaN  Inf 0.7879397 -0.2217992
```

Figure 17 - Forecast model output, ETS, Tokyo.

```
        Ljung-Box test

data:  Residuals from ETS(A,N,A)
Q* = 98.988, df = 10, p-value < 2.2e-16

Model df: 14.   Total lags used: 24
```

Figure 18 - Ljung-Box test, ETS, Tokyo.

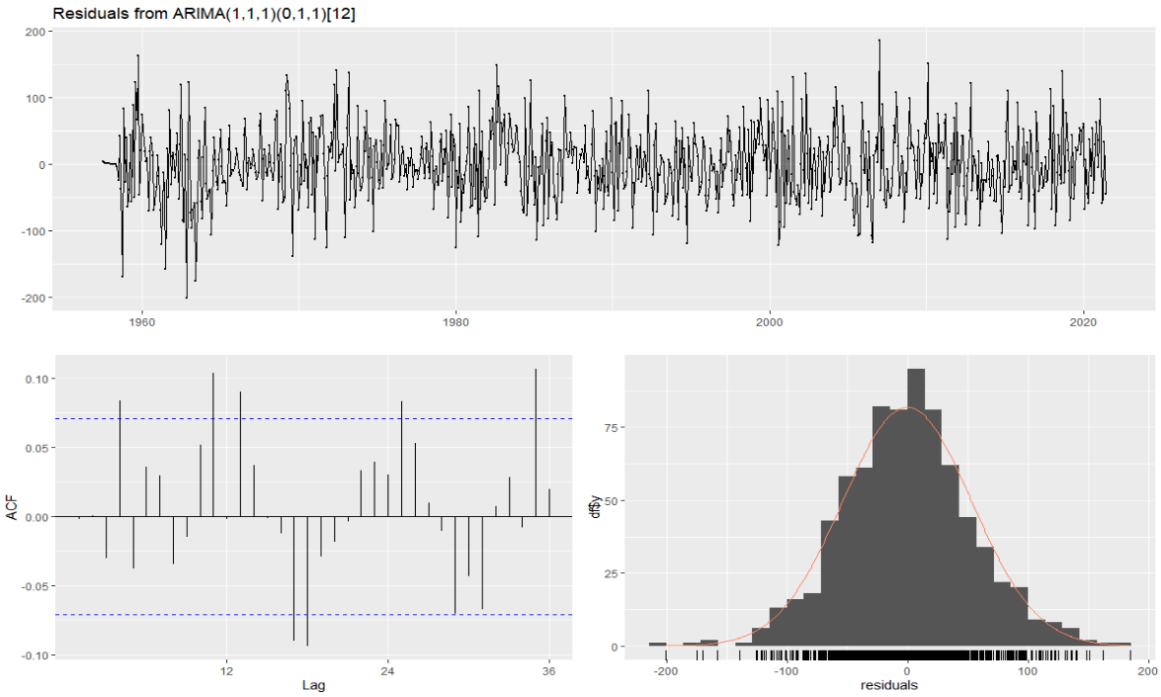


Figure 19 - Residual plots by modeling with the ARIMA method, Tokyo

```
 Best model: ARIMA(1,1,1)(0,1,1)[12]

> print(summary(fit_arimatokyo))
Series: tokalltc
ARIMA(1,1,1)(0,1,1)[12]

Coefficients:
          ar1      ma1     sma1
      0.5032  -0.9744  -0.7454
s.e.  0.0392   0.0171   0.0284

sigma^2 = 2926:  log likelihood = -4089.25
AIC=8186.5   AICc=8186.55   BIC=8205

Training set error measures:
                    ME     RMSE      MAE           MPE      MAPE     MASE         ACF1
Training set -0.3541815 53.52366 41.63983 -0.009142639 0.5940047 0.745029 -0.001714622
```

Figure 20 - Best auto.arima() fit and Forecast model output, ARIMA, Tokyo.

```
        Ljung-Box test

data:  Residuals from ARIMA(1,1,1)(0,1,1)[12]
Q* = 45.158, df = 21, p-value = 0.001651

Model df: 3.   Total lags used: 24
```

Figure 21 - Ljung-Box test, ARIMA, Tokyo.

Within the summary output of the model's residual standard deviation determines how well the data fits the model, the closer the value is to zero the better the fit of the model. The ACF shows the residuals, or the error terms that cannot be explained by the model. If there are many residuals it denotes that there is autocorrelation meaning that information in the data that the model is not using in the most efficient way. The Ljung-Box Test tests the hypothesis that there is autocorrelation within the time series, thus the null hypothesis is that the residuals are independently distributed (Koalatea.io, 2022). The output produces a p-value which can also be used to determine whether the time series data includes an autocorrelation.

The seasonal naive method is the benchmark test as it fits the data only based on only the previous seasons data rather than all past data holistically. This makes it the most inaccurate of all the models, however it was a good initial point to create a model that we can use to compare against the ETS model and the ARIMA model. For the Seasonal Naive method the standard deviation output is 78.9422. The ACF plot has 8 residuals that fall outside of the 95% confidence interval. The p-value result of the Ljung-Box test for this model is 2.2e-16, which is

smaller than 0.1 determining that the null hypothesis can be rejected and that the time series contains autocorrelation. This would make sense as it is the benchmark model which we expected to fit the least well.

The ETS model output standard deviation output is 61.6419, which is closer to zero than the Seasonal Naive model indicating a better fit. The ACF plot here shows 10 residuals falling outside of the confidence interval, more than that of the previous model. The p-value result is 2.2e-16 which explains the ACF plot as there is definite autocorrelation here.

The final model used was an auto fitted ARIMA model. The parameters here are the original time series model, thee number of differences to be taken (in our case 1), the seasonality difference (in our case also 1, this removes the seasonality), boolean stepwise (set to FALSE as we are using a univariate time series and only a few iterations of the ARIMA model are necessary), approximation (set to FALSE as again only using a single time series and thus we do not need to save time by using approximation AIC) , and trace (set as true so all model iterations are printed to thee console). The standard deviation output here is squared but once the root is taken it outputs 54.09251, which is the best fit of all the models. The ACF plot shows 7 residuals outside of the confidence interval again, the least residuals out of all the models created. The p-value result of the Ljung-Box test is 0.0001651, which is significantly bigger than the previous results, however is still smaller than 0.1 denoting that whilst it has less autocorrelation than the other models, it still denotes that the residuals are not distributed independently and thus shows a serial correlation.

Forecasting

Tokyo Forecasting - ARIMA Model

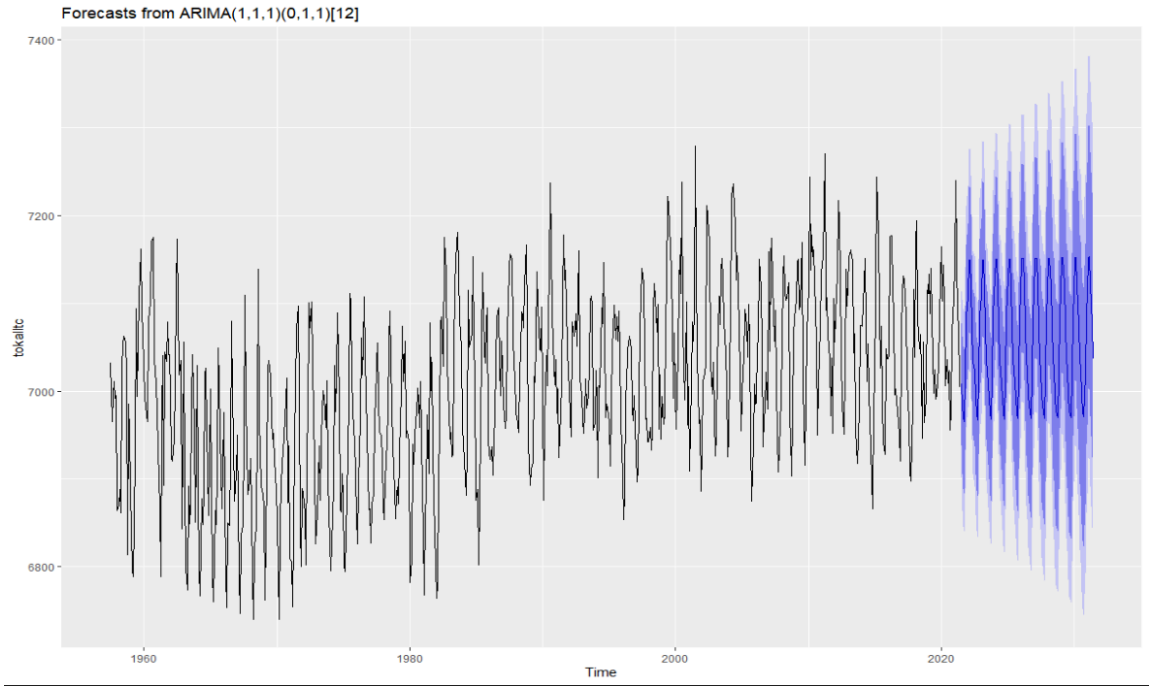


Figure 22 - Forecasting with the ARIMA, Tokyo

Forecasting was done using the ARIMA model that was created, as it provided the best fit with the least autocorrelation out of the three models attempted. Each model produced an output of type forecast. The forecast() function was used with the ARIMA Model that used all of the time series data for Tokyo.

The forecasting plots for the Seasonal Naive and ETS models using this method are included in the Appendix for context.

4.3 Further Analysis

Following the model creation and forecasting a different model was employed to check the accuracy of the method we used. Using the same auto.arima() method a model was created using a training set of 70% of the original Tokyo time series data, this included 538 months (44.83 years). This left a testing set of data of 230 months (19.16 years). These 230 months of the original Tokyo time series data could then be compared against a forecast of 230 months of the ARIMA model created.

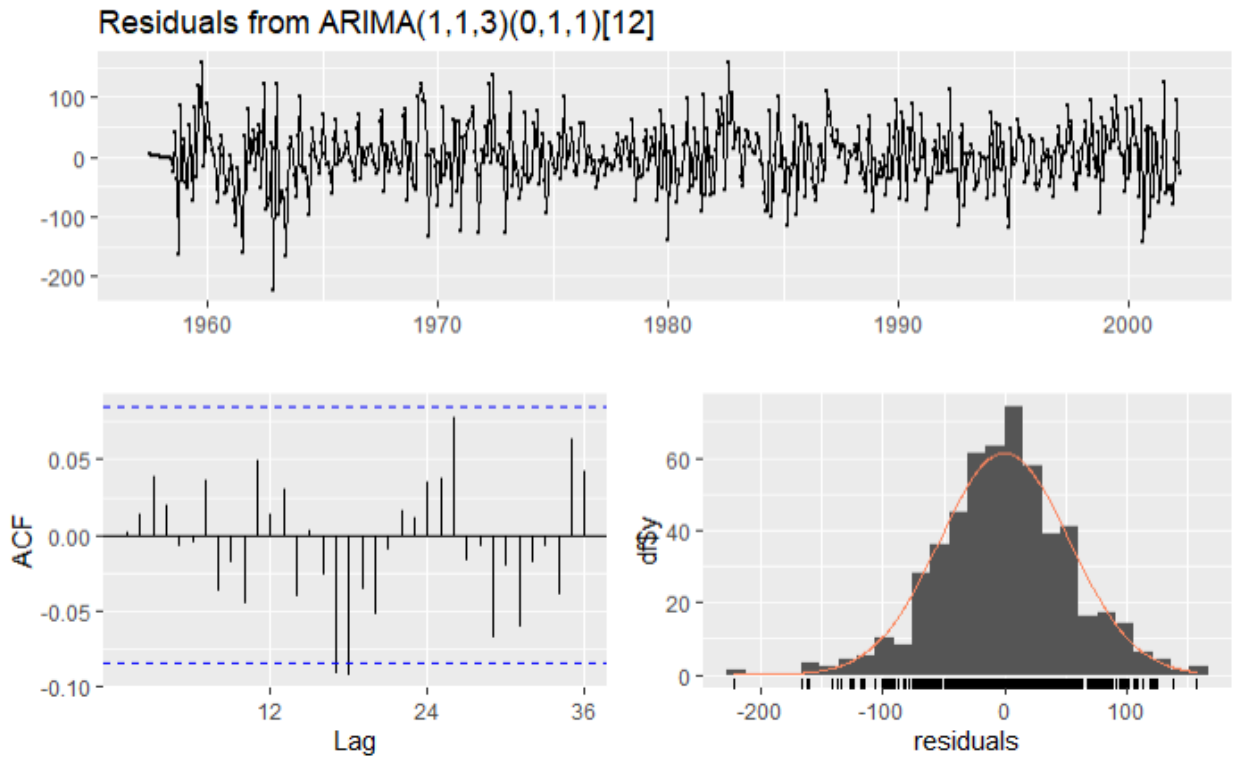


Figure 23 - Residual, ARIMA, Tokyo, Model Based on training data.

```
Forecast method: ARIMA(1,1,3)(0,1,1)[12]

Model Information:
Series: traintokyo
ARIMA(1,1,3)(0,1,1)[12]

Coefficients:
          ar1     ma1      ma2      ma3     sma1
      -0.5027  0.0621  -0.4924  -0.3401  -0.7855
s.e.   0.1540  0.1500   0.0826   0.0577   0.0413

sigma^2 = 2896:  log likelihood = -2841.31
AIC=5694.63   AICc=5694.79   BIC=5720.21

Error measures:
                    ME     RMSE      MAE          MPE      MAPE      MASE        ACF1
Training set -0.57416 52.90528 40.56868 -0.01240308 0.5808472 0.7169923 0.002378789
```

Figure 23 - ARIMA Method Output, Tokyo, Model Based on training data.

The standard deviation of the training data model is 53.8145, which is a better fit than all previous models. Additionally there are only 2 residuals outside of the confidence interval on the ACF plot which is a better value for autocorrelation than previous models also.

```
                    ME      RMSE      MAE          MPE
Training set -0.57416  52.90528 40.56868 -0.01240308
Test set     13.22058 109.63232 91.52918  0.17278866
                 MAPE      MASE        ACF1
Training set 0.5808472 0.7169923 0.002378789
Test set     1.2943213 1.6176449 0.753322052
             Theil's U
Training set        NA
Test set      1.666878
```

Figure 24 - Accuracy of the model against the real time series data.

Looking at the RMSE here we can see that the model provides a value of 52.90528 vs the real datas RMSE of 109.63232. The mean percentage error here is 0.01% denoting a highly accurate fitting.

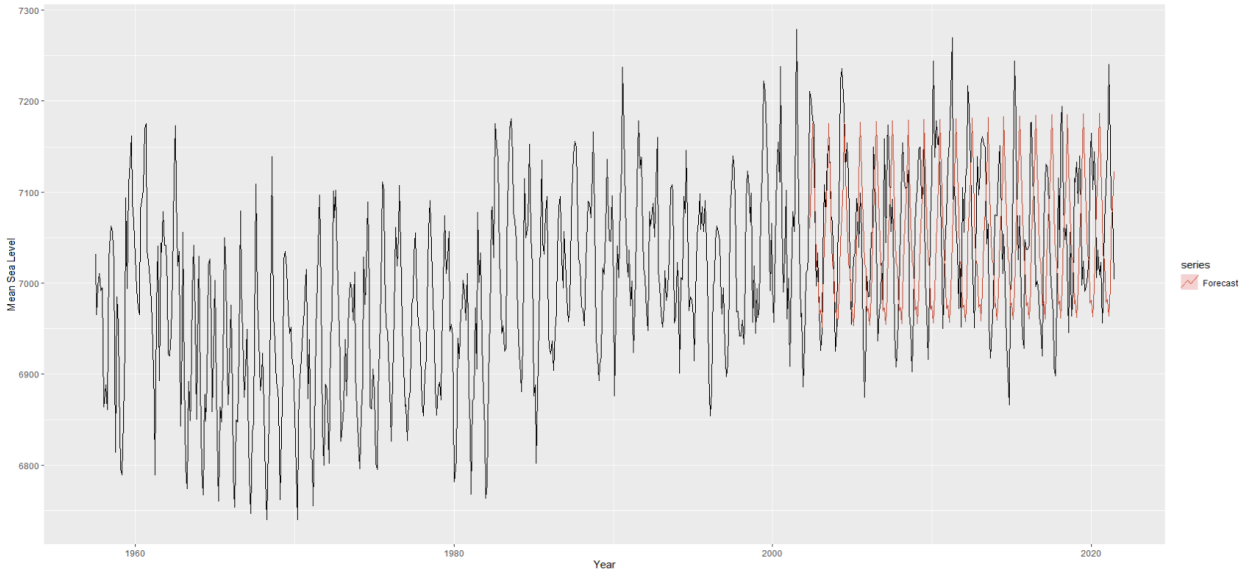


Figure 25 - Ljung-Box test, ETS, Tokyo.

Both the real time series data and the forecasted data can be visualized in the time series plot above. It can be seen that the real data often shows a higher amplitude than the forecast and that the forecast's seasonality is slightly out of phase with the real time results. However the fitting of the model is visually quite accurate as the forecasted data fits within the bounds of the real data.

5. Discussion and Conclusions

Discussion

This report explored a small fragment of the available tide gauge data in Japan recorded by stations scattered throughout the country. Stations of concern were limited to those in the top 3 highest populated areas in Japan, which consisted of Tokyo, Osaka and Nagoya in order, with Tokyo being the primary area of focus. Analysis of stationarity and seasonality of the sea level obtained for Tokyo, as well as Osaka and Nagoya was done, leading into modeling and forecasting the potential sea level changes 230 months into the future.

The seasonality demonstrated by Tokyo sea levels during different seasons of the year suggest that sea levels are affected by drastic weather and temperature changes; for instance, sea levels may be rising in Tokyo in summer due to increased temperatures that cause water bodies to expand. The opposite is true in winter, in which cooler temperatures result in water bodies contracting instead, and hence sea level decreases. Furthermore, perhaps the sea levels faced a higher trend after the 1980s potentially due to increased global warming.

Dataset Limitations

PSMSL data set suffers from three severe limitations: (1) the geographical distribution of reliable tide gauge stations is rather uneven with pronounced concentrations in some areas of the northern hemisphere (Europe, North America, Japan), and much fewer stations on the southern hemisphere where particularly few stations are located in Africa and in Antarctica; (2) the number of stations recording simultaneously at any time is far less than the total number of

stations with the maximum within the interval between 1958 and 1988; (3) the number of long records is extremely small and almost all of them originate from a few regions of the northern hemisphere.

Conclusions

1. The time series data for Tokyo shows that there is both a positive trend. and a seasonal trend. The sea level has been rising between 1982-2021.
2. A model has been created to forecast Tokyo's future sea levels, this model shows a very minor upward trend.
3. The forecast model created for 70% of Tokyo's existing data, when compared to the real data, denotes a mean percentage error of 0.1%, it is a fairly accurate model.

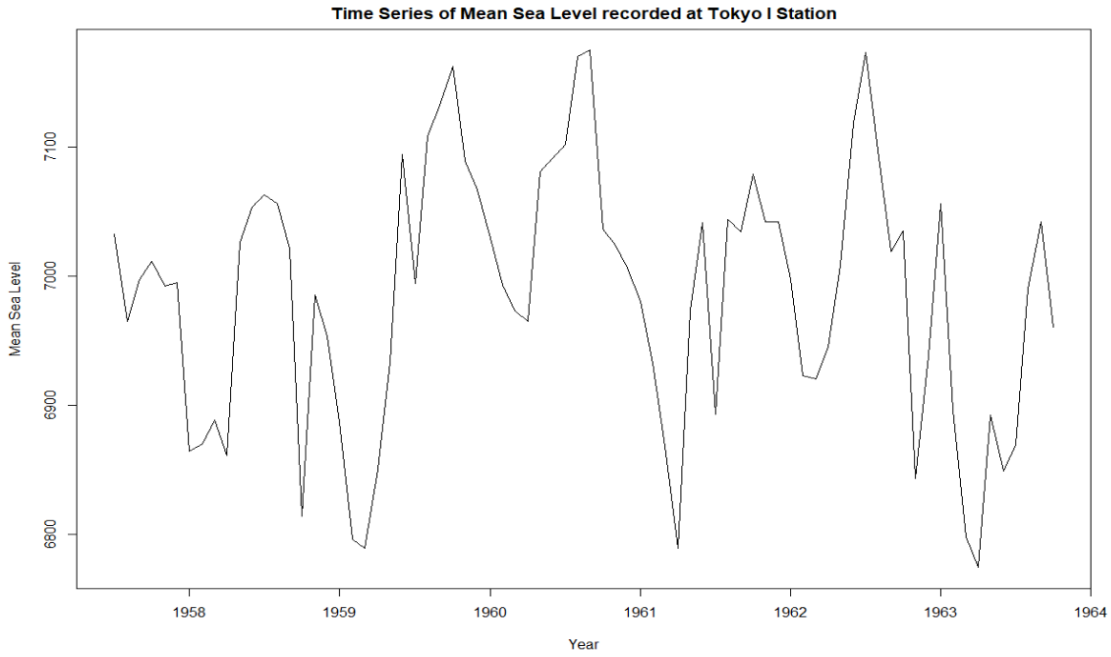
References

- Aubrey, D. G., & Emery, K. O. (1986). Relative sea levels of Japan from tide-gauge records. *Geological Society of America Bulletin*, *97*(2), 194-205.
- Anish, A., 2020. Time Series Analysis. [online] Medium. Available at: <https://medium.com/swlh/time-series-analysis-7006ea1c3326> [Accessed 24 May 2022]
- Banno, M., & Kuriyama, Y. (2014, October). Prediction of future shoreline change with sea-level rise and wave climate change at Hasaki, Japan. In *Proc. Int. Conf. In Coastal Eng, Seoul, South Korea*.
- Bower, M. (2019). Timescales of Global Tidal Flooding.
- Cao, A., Esteban, M., & Mino, T. (2020). Adapting wastewater treatment plants to sea level rise: learning from land subsidence in Tohoku, Japan. *Natural Hazards*, *103*(1), 885-902.
- Cazenave, A., & Cozannet, G. L. (2014). Sea level rise and its coastal impacts. *Earth's Future*, *2*(2), 15-34.
- Church, J. A., White, N. J., Aarup, T., Wilson, W. S., Woodworth, P. L., Domingues, C. M., ... & Lambeck, K. (2008). Understanding global sea levels: past, present and future. *Sustainability Science*, *3*(1), 9-22.
- Comeaux, R. S., Allison, M. A., & Bianchi, T. S. (2012). Mangrove expansion in the Gulf of Mexico with climate change: Implications for wetland health and resistance to rising sea levels. *Estuarine, Coastal and Shelf Science*, *96*, 81-95.
- Dagum, E. B., & Bianconcini, S. (2016). Seasonal adjustment methods and real time trend-cycle estimation. Springer. Available: https://otexts.com/fpp2/x11.html
- Douglas, B. C. (1991). Global sea level rise. *Journal of Geophysical Research: Oceans*, *96*(C4), 6981-6992.
- FitzGerald, D. M., Fenster, M. S., Argow, B. A., & Buynevich, I. V. (2008). Coastal impacts due to sea-level rise. *Annu. Rev. Earth Planet. Sci.*, *36*, 601-647.
- Gornitz, V. (1990). Vulnerability of the East Coast, USA to future sea level rise. *Journal of Coastal research*, 201-237.
- Gornitz, V. (1991). Global coastal hazards from future sea level rise. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *89*(4), 379-398.
- Gröger, M., & Plag, H. P. (1993). Estimations of a global sea level trend: limitations from the structure of the PSMSL global sea level data set. *Global and Planetary Change*, *8*(3), 161-179.
- Haigh, I., Nicholls, R., & Wells, N. (2011, June). Rising sea levels in the English Channel 1900 to 2100. In *Proceedings of the Institution of Civil Engineers-Maritime Engineering* (Vol. 164, No. 2, pp. 81-92). Thomas Telford Ltd.
- Haigh, I. D., Pickering, M. D., Green, J. M., Arbic, B. K., Arns, A., Dangendorf, S., ... & Woodworth, P. L. (2020). The tides they are a‐Changin': A comprehensive review of past and future nonastronomical changes in tides, their driving mechanisms, and future implications. *Reviews of Geophysics*, *58*(1), e2018RG000636.
- Hall, J. A., Weaver, C. P., Obeysekera, J., Crowell, M., Horton, R. M., Kopp, R. E., ... & White, K. D. (2019). Rising sea levels: Helping decision-makers confront the inevitable. *Coastal Management*, *47*(2), 127-150.

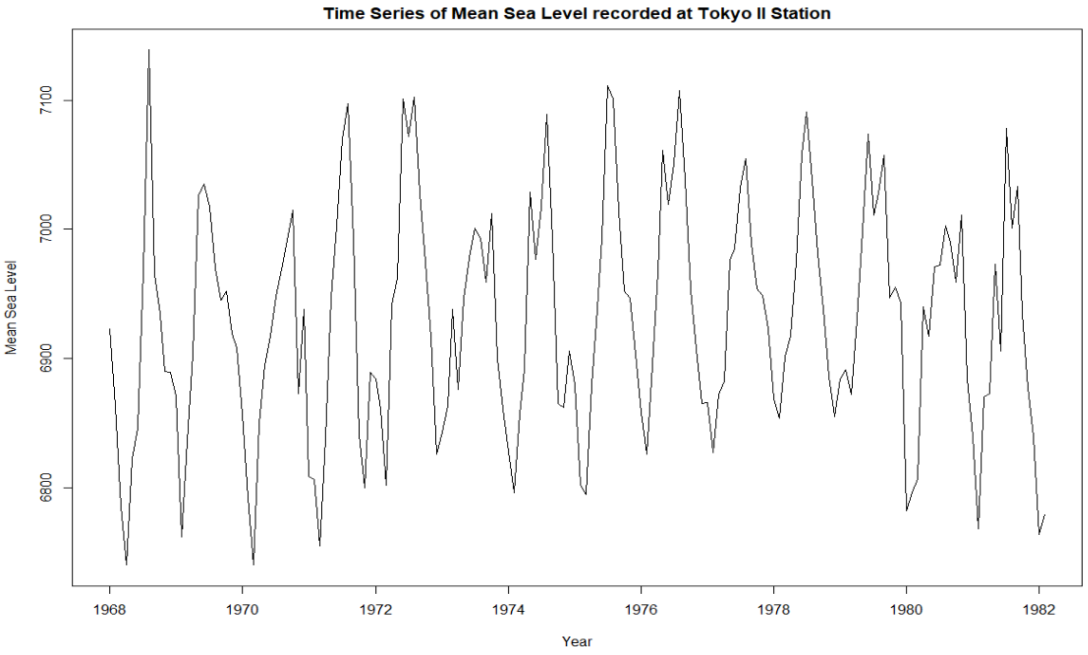
- Heberger, M., Cooley, H., Herrera, P., Gleick, P. H., & Moore, E. (2009). The impacts of sea-level rise on the California coast.
- Jevrejeva, S., Grinsted, A., Moore, J. C., & Holgate, S. (2006). Nonlinear trends and multiyear cycles in sea level records. *Journal of Geophysical Research: Oceans*, *111*(C9).
- Kang, S. K., Cherniawsky, J. Y., Foreman, M. G. G., Min, H. S., Kim, C. H., & Kang, H. W. (2005). Patterns of recent sea level rise in the East/Japan Sea from satellite altimetry and in situ data. *Journal of Geophysical Research: Oceans*, *110*(C7).
- Kirwan, M. L., Guntenspergen, G. R., d'Alpaos, A., Morris, J. T., Mudd, S. M., & Temmerman, S. (2010). Limits on the adaptability of coastal marshes to rising sea level. *Geophysical research letters*, *37*(23).
- Knowles, N. (2010). Potential inundation due to rising sea levels in the San Francisco Bay region. San Francisco Estuary and Watershed Science, 8(1).
- Koalatea.io. 2022. [online] Available at: <https://koalatea.io/r-ljung-box-test/> [Accessed 24 May 2022].
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics, 54(1-3), 159–178.
- Mimura, N. (2013). Sea-level rise caused by climate change and its implications for society. *Proceedings of the Japan Academy, Series B*, *89*(7), 281-301.
- Morris, J. T., Sundareshwar, P. V., Nietch, C. T., Kjerfve, B., & Cahoon, D. R. (2002). Responses of coastal wetlands to rising sea level. *Ecology*, *83*(10), 2869-2877
- Moritz, S., 2021. na_seasplit: Seasonally Splitted Missing Value Imputation in imputeTS: Time Series Missing Value Imputation. [online] Rdrr.io. Available at: <https://rdrr.io/cran/imputeTS/man/na_seasplit.html> [Accessed 24 May 2022].
- Nagai, R. et al. (2020). Tsunami risk hazard in Tokyo Bay: The challenge of future sea level rise. *International Journal of Disaster Risk Reduction* 45:101321.
- Nakano, M., & Yamada, S. (1975). On the mean sea levels at various locations along the coasts of Japan. *Journal of the Oceanographical Society of Japan*, *31*(2), 71-84.
- Nicholls, R.J. (2002) Rising sea levels: potential impacts and responses. In, Hester, R.E. and Harrison, R.M. (eds.) *Global Environment Change.* (Issues in Environmental Science and Technology, 17) Cambridge, UK. Royal Society of Chemistry, pp. 83-107.
- Nicholls, R. J., & Cazenave, A. (2010). Sea-level rise and its impact on coastal zones. *science*, *328*(5985), 1517-1520.
- N. Sultana and N. Sharma, "Statistical Models for Predicting Swine F1u Incidences in India," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018, pp. 134-138, doi: 10.1109/ICSCCC.2018.8703300.
- OH, I. S., RABINOVICH, A. B., PARK, M. S., & MANSUROV, R. N. (1993). Seasonal sea level oscillations in the East Sea (Sea of Japan). 한국해양학회지, *28*(1), 1-16.
- Orson, R., Panageotou, W., & Leatherman, S. P. (1985). Response of tidal salt marshes of the US Atlantic and Gulf coasts to rising sea levels. *Journal of Coastal Research*, 29-37.
- Permanent Service for Mean Sea Level (PSMSL), 2022, "Tide Gauge Data", Retrieved 09 May 2022 from http://www.psmsl.org/data/obtaining/.
- Pugh, D., & Woodworth, P. (2014). *Sea-level science: understanding tides, surges, tsunamis and mean sea-level changes*. Cambridge University Press.

- Prabhakaran, S., 2022. [online] Machinelearningplus.com. Available at: <https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/> [Accessed 20 May 2022].
- Ramasamy, R., & Surendran, S. N. (2011). Possible impact of rising sea levels on vector-borne infectious diseases. *BMC infectious diseases*, *11*(1), 1-6.
- Shoji, D. (1961). On the variations of the daily mean sea levels along the Japanese Islands. *Journal of the Oceanographical Society of Japan*, *17*(3), 141-152.
- Smajgl, A., Toan, T. Q., Nhan, D. K., Ward, J., Trung, N. H., Tri, L. Q., ... & Vu, P. T. (2015). Responding to rising sea levels in the Mekong Delta. *Nature Climate Change*, *5*(2), 167-174.
- Udo, K., & Takeda, Y. (2017). Projections of future beach loss in Japan due to sea-level rise and uncertainties in projected beach loss. *Coastal Engineering Journal*, *59*(02), 1740006.
- Woodworth, P. L. (1991). The permanent service for mean sea level and the global sea level observing system. *Journal of Coastal Research*, 699-710.
- Woodworth, P. L., & Player, R. (2003). The permanent service for mean sea level: An update to the 21stCentury. *Journal of Coastal Research*, 287-295.

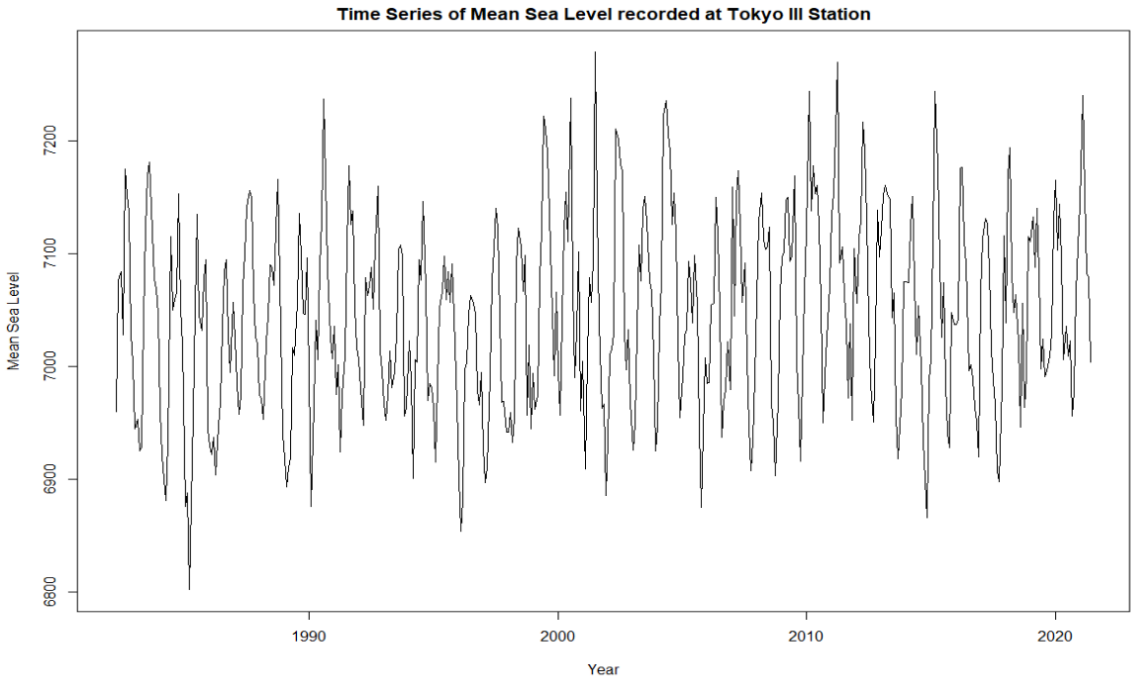
Appendix



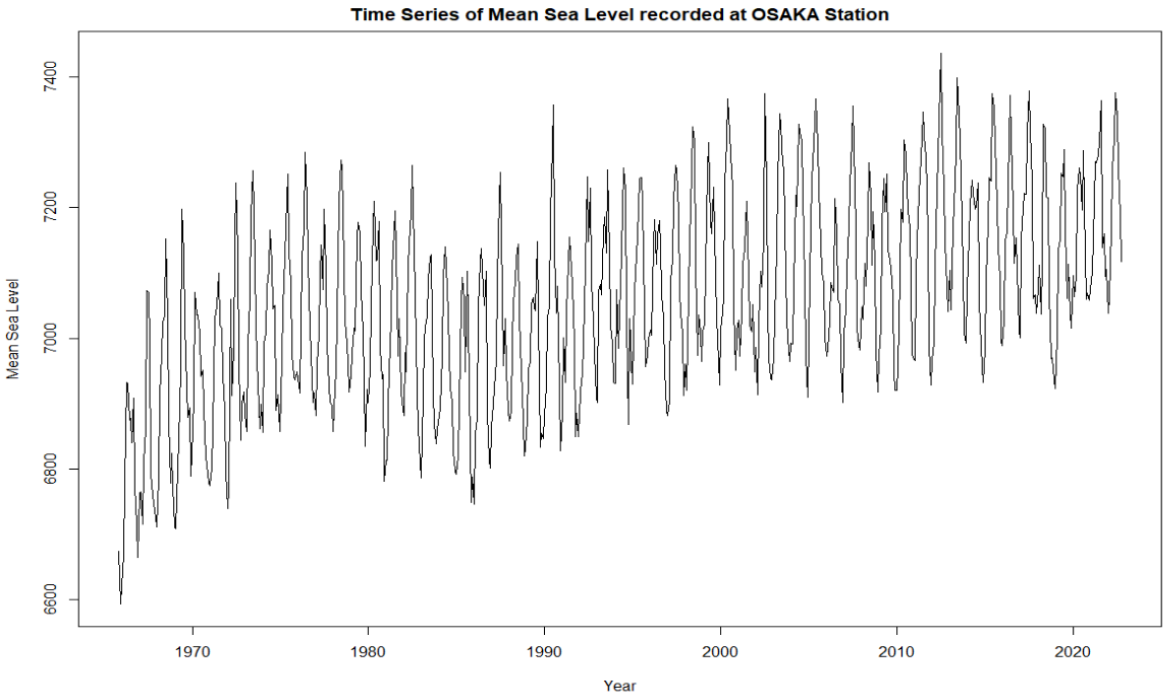
Appendix A - Mean sea level by month, Time series plot, Tokyo I



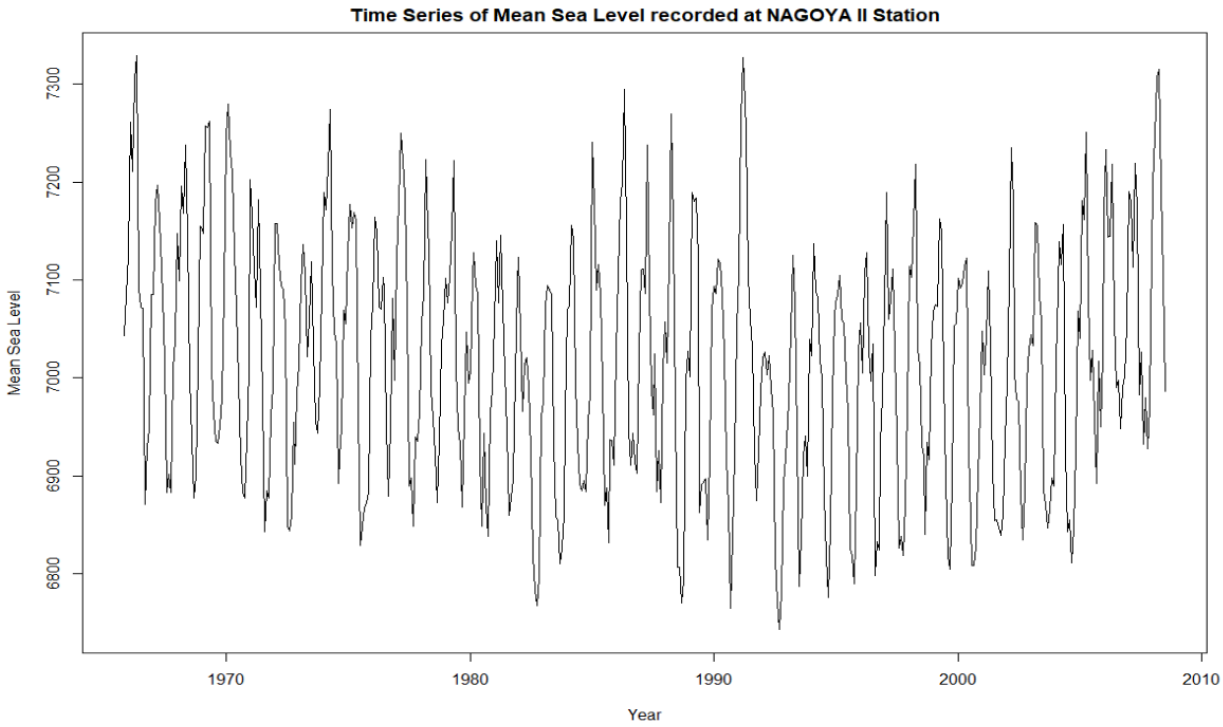
Appendix B - Mean sea level by month, Time series plot, Tokyo II



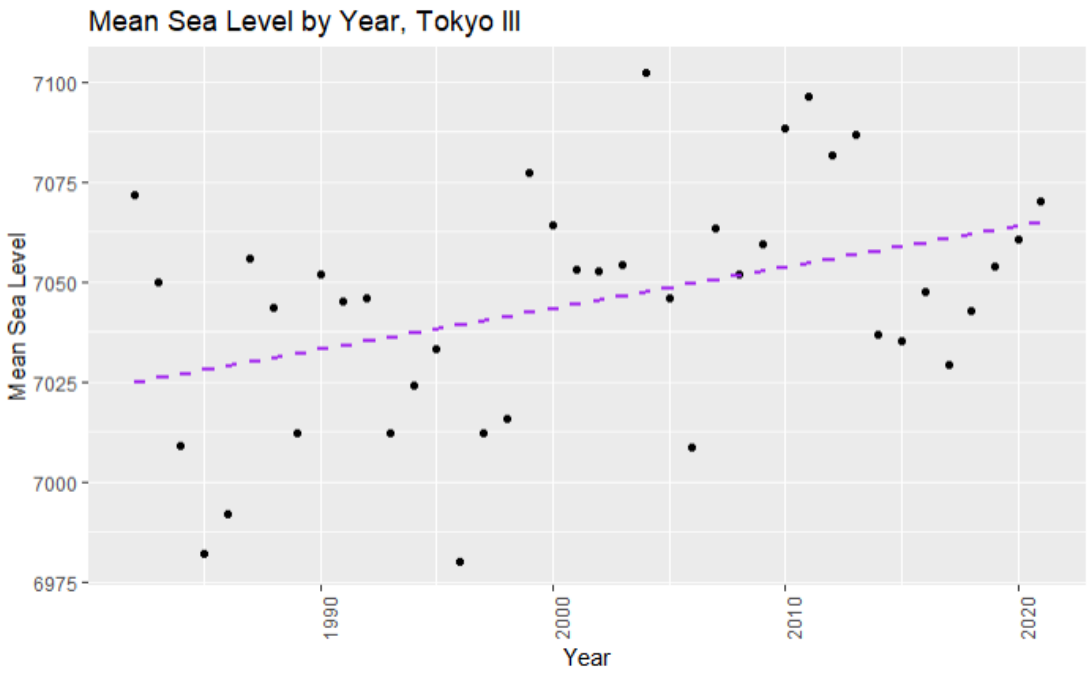
Appendix C - Mean sea level by month, Time series plot, Tokyo III



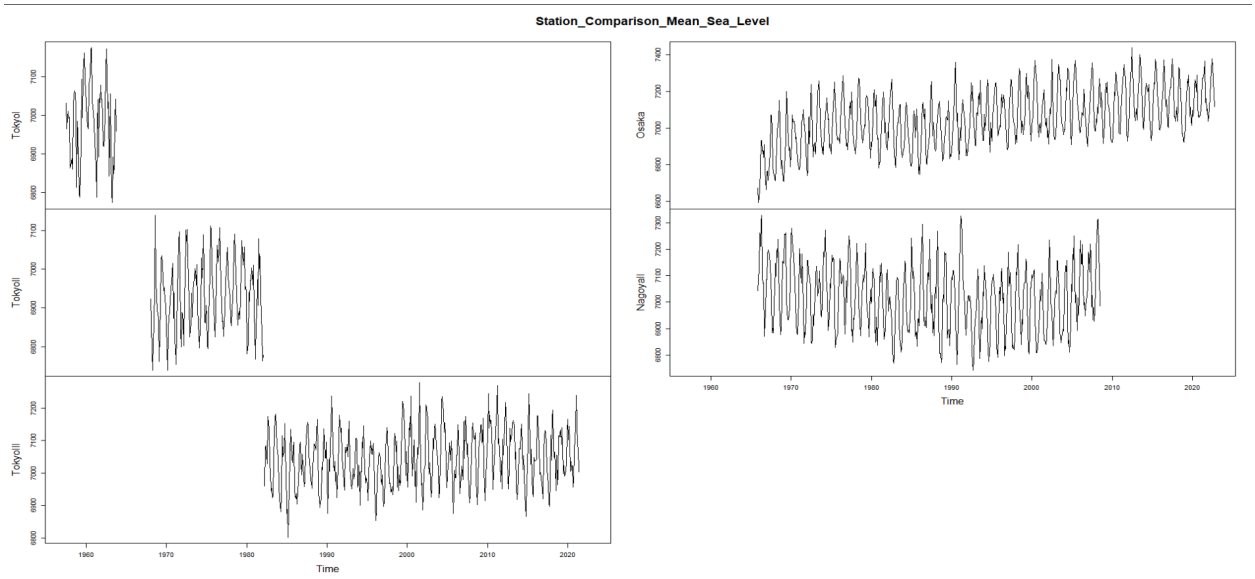
Appendix D - Mean sea level by month, Time series plot, Osaka



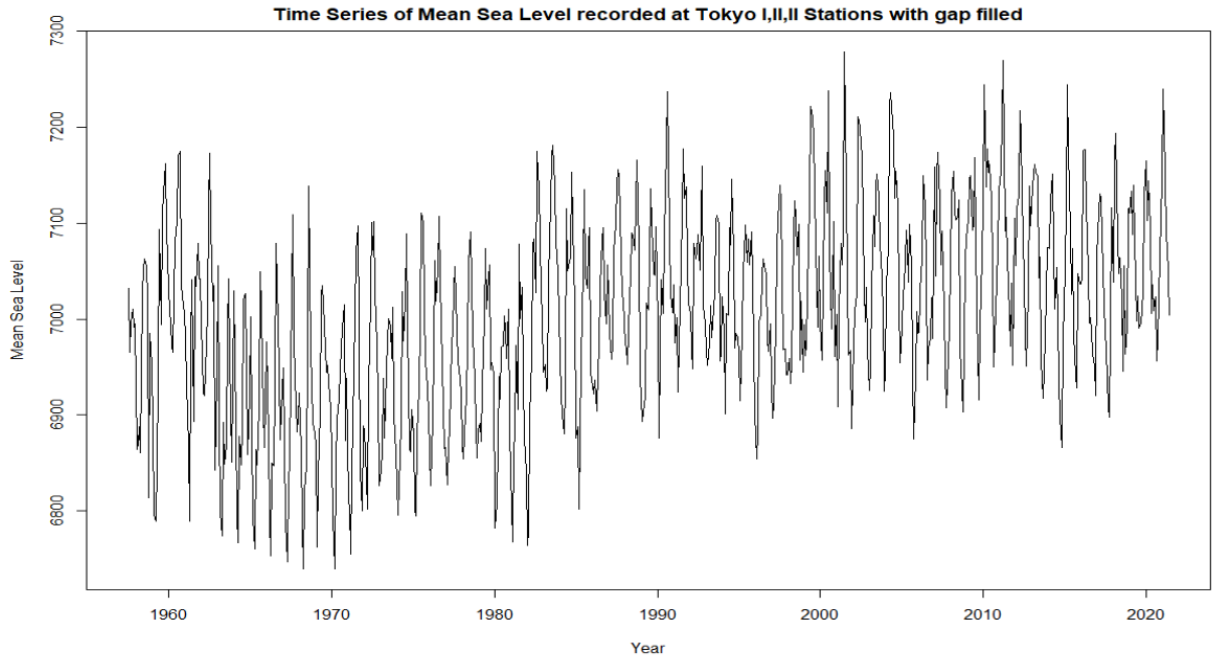
Appendix E - Mean sea level by month, Time series plot, Nagoya II



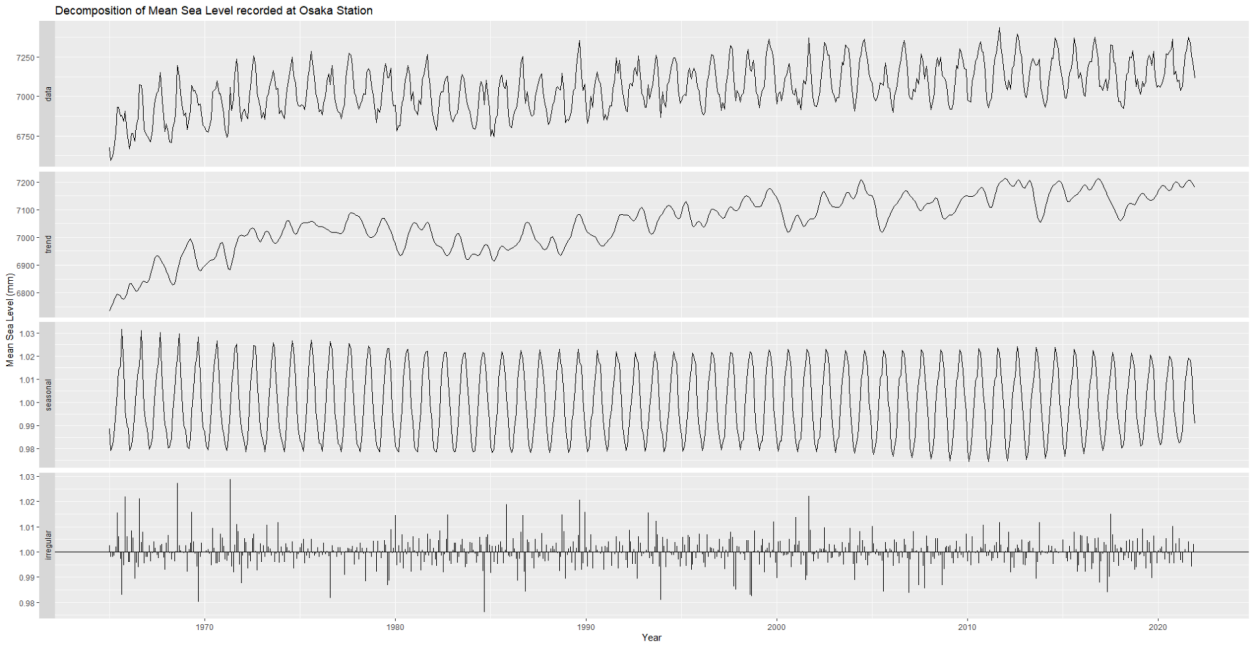
Appendix F - Mean sea level by month, Time series linear regression plot, Tokyo III



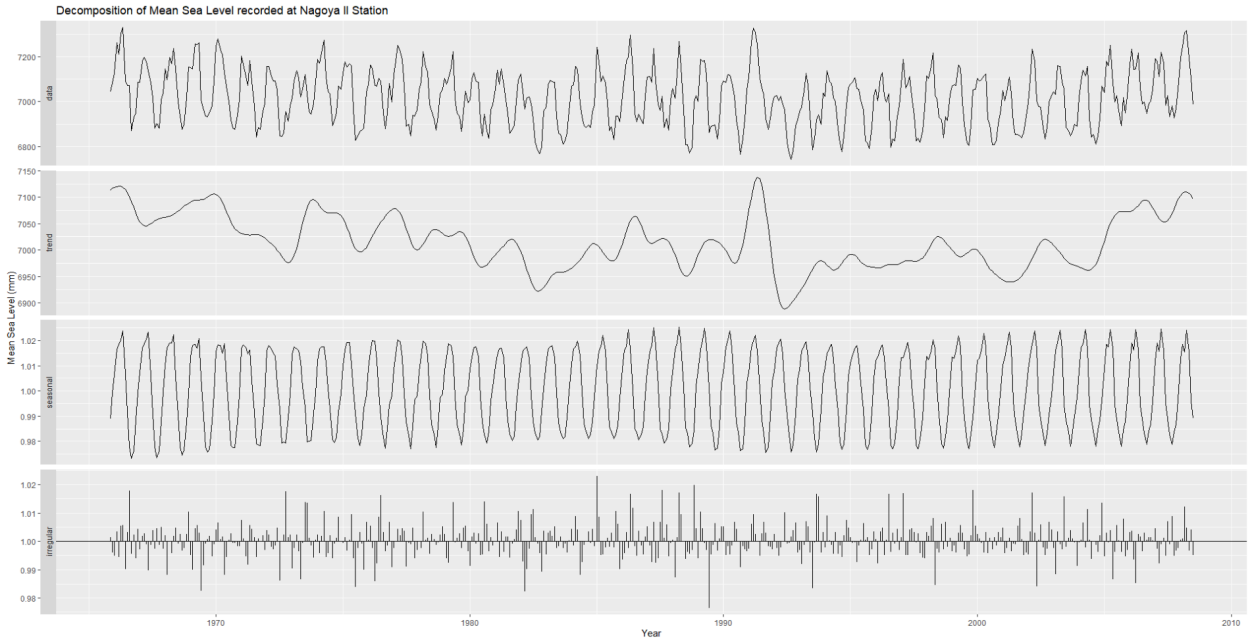
Appendix G - Time series plots of each stations data, comparison



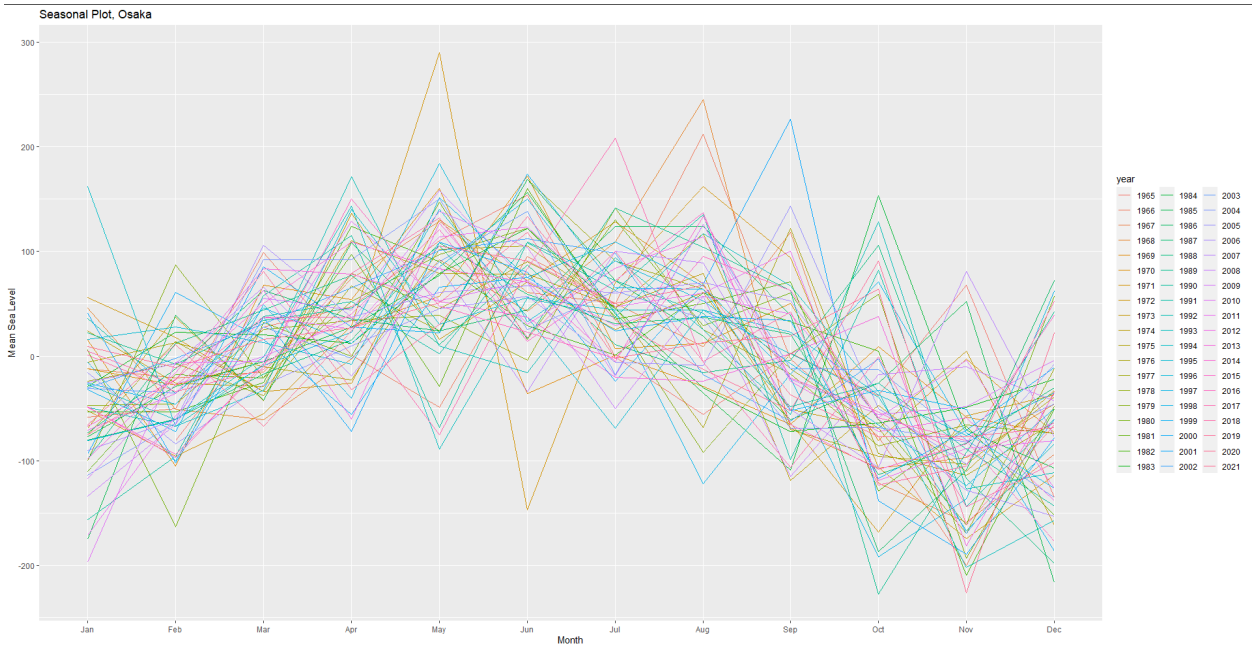
Appendix H - Imputation of the NA Tokyo Data, plotted.



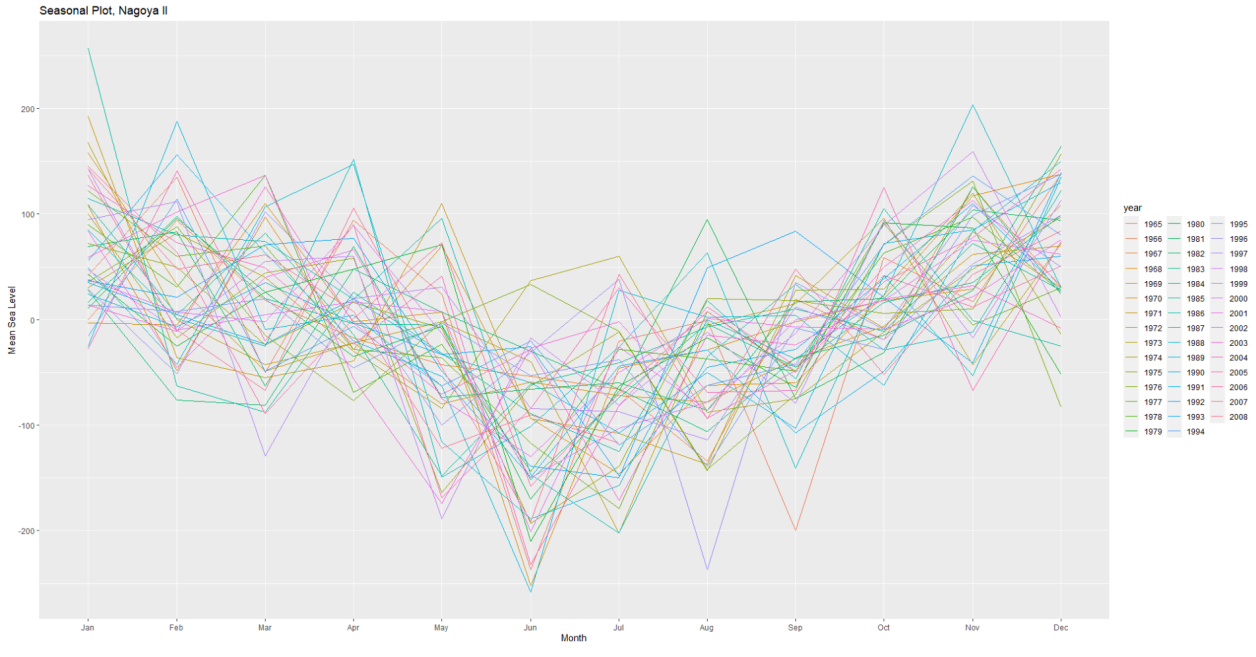
Appendix I - Decomposition of the Osaka Time Series Data.



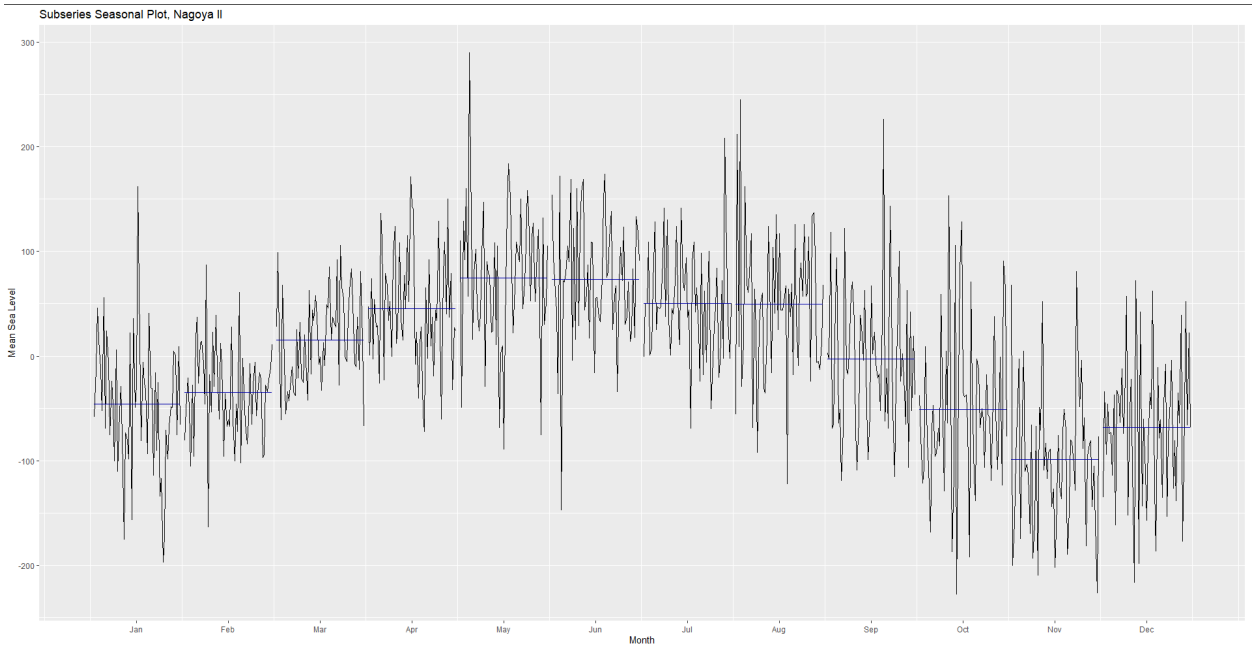
Appendix J - Decomposition of the Nagoya II Time Series Data.



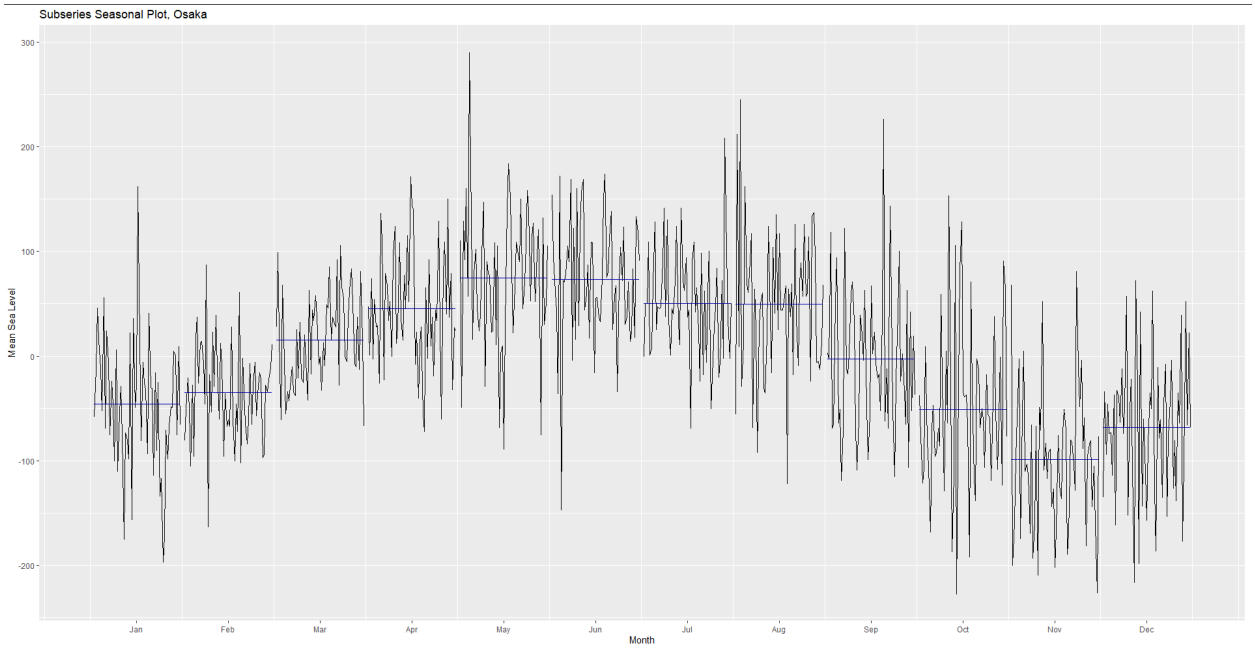
Appendix K - Seasonality Plot, Osaka



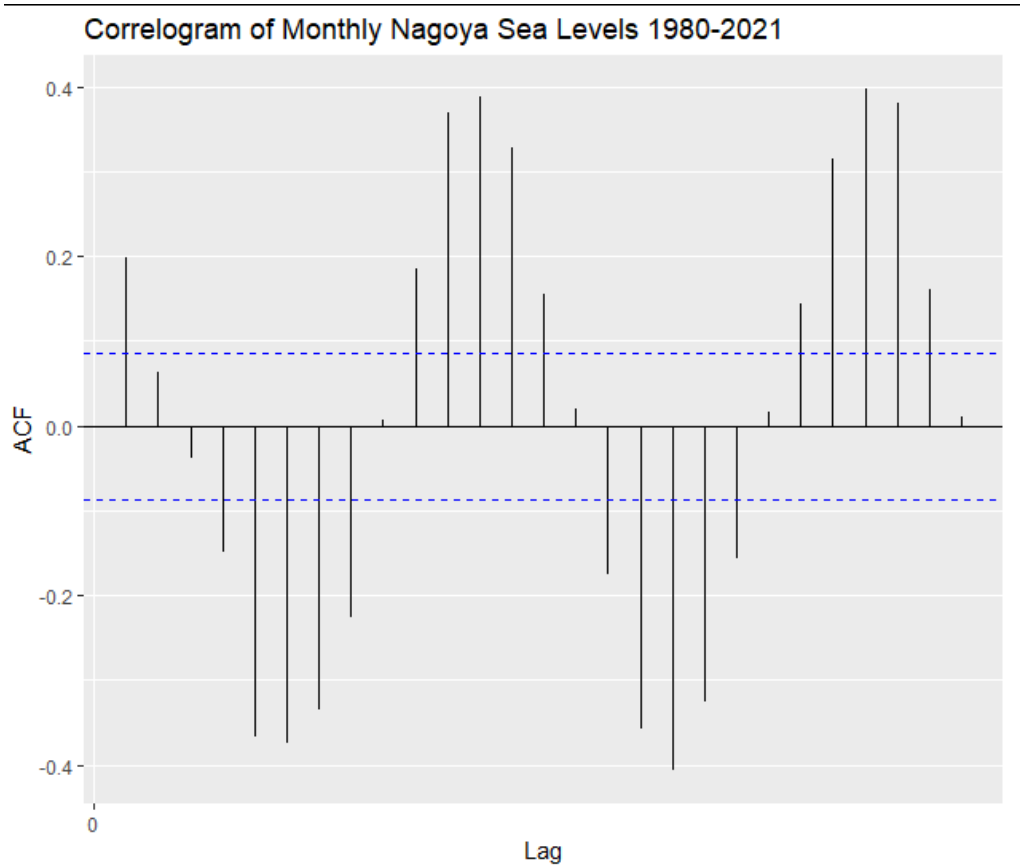
Appendix L - Seasonality Plot, Nagoya II



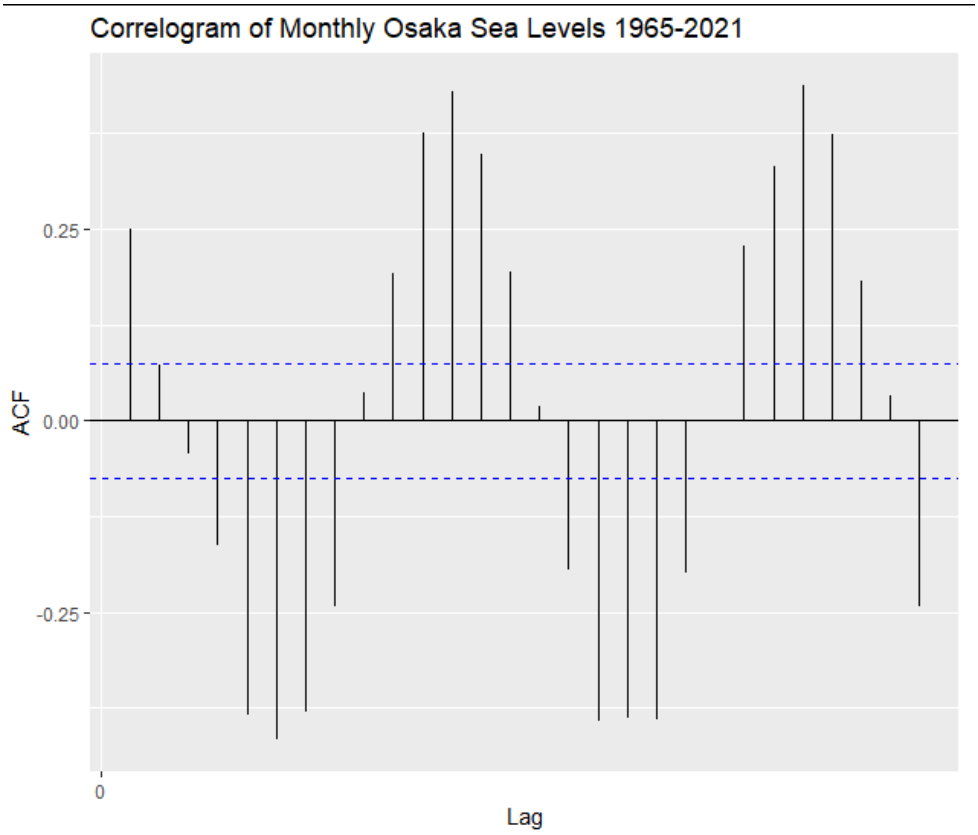
Appendix M - Subseries Seasonality Plot, Nagoya II



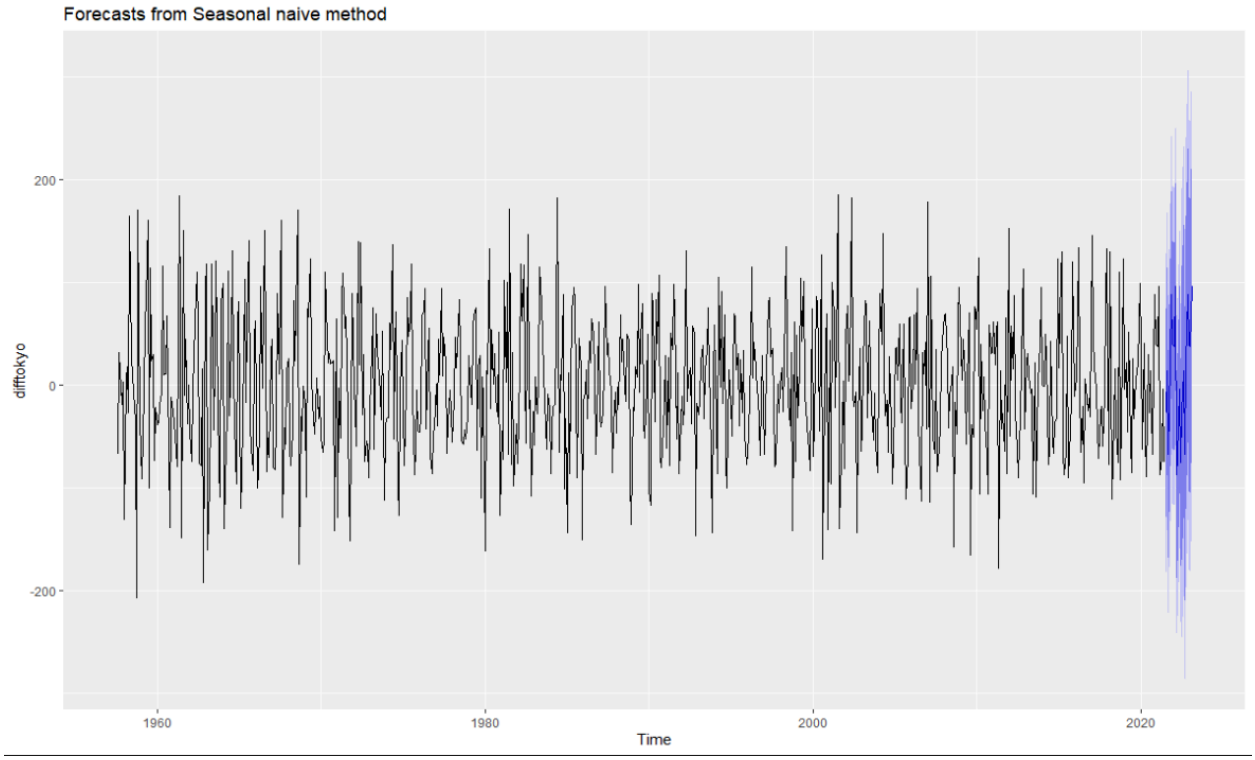
Appendix N - Subseries Seasonality Plot, Osaka



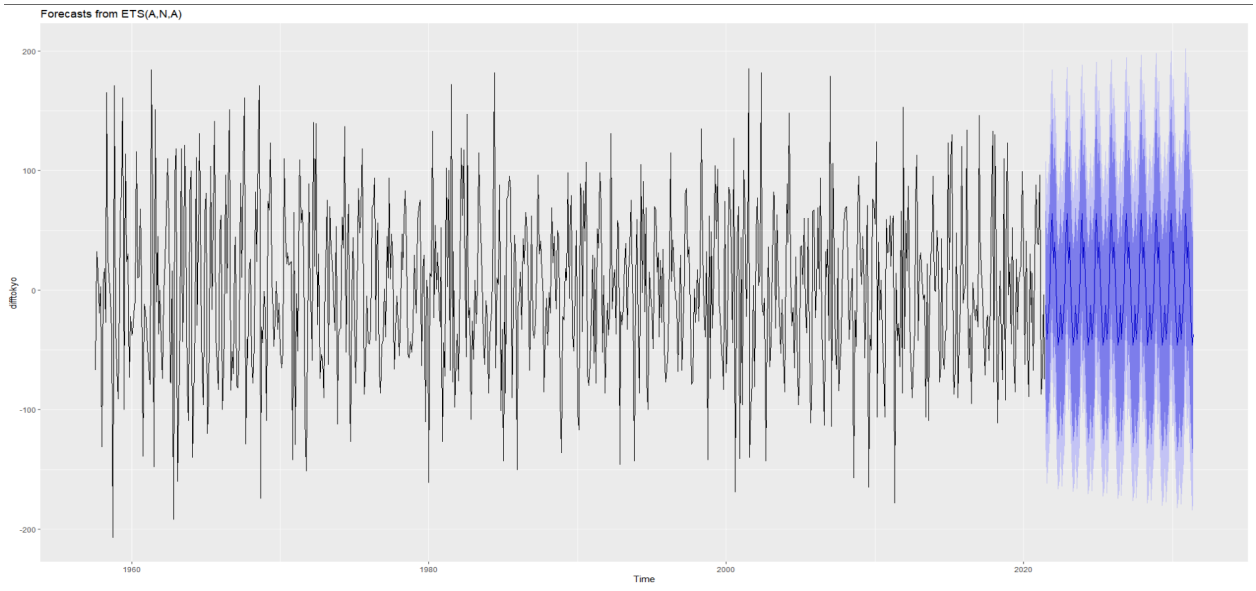
Appendix O - Autocorrelation Plot, Nagoya II



Appendix P - Autocorrelation Plot, Osaka



Appendix Q - Forecasting with the Seasonal Naive Model, Tokyo



Appendix R - Forecasting with the ETS Model, Tokyo

Appendix S - Code

```
---
title: "Dissertation"
author: "Joshua Bhawanlall, Aridj Chenak, Emmanuela Vischetti,
Joseph Florentino, Mitchel Berry"
date: "4/3/2022"

---

  # Year in Data Analytics - Group Project Preliminary Analysis
#




## Initializing Libraries ##



library(dplyr)
library(tidyr)
library(stringr)
library(tidyverse)
library(ggplot2)
```

```
library(plyr)

library(zoo)

library(lubridate)

library("reshape2")

library("tseries")

library(xts)

library("funModeling")

library("imputeTS") #ts imputation

library("tsbox") #ts binding

library(forecast)

library("seasonal") #seasonal adjust (may or may not need)




## Importing Data from PSMSL Website, Creating Dataframe,

Removing NAs ##



     ##STATION 881 - TOKYO I##


#Set URL for Data (Change number for station)
```

```
url881 <-
"https://psmsl.org/data/obtaining/rlr.monthly.data/881.rlrdata"
url881A <-
"https://psmsl.org/data/obtaining/rlr.annual.data/881.rlrdata"

#Set destination path for file to download, add "/output.csv" to
end of filepa/th (Downloads Folder)
destfile881 <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output881.csv"
destfile881A <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output881A.csv"

# Apply download.file function in R
download.file(url881, destfile881)
download.file(url881A, destfile881A)

#Read in txt station data .CSV into variable.
stationcsv881 <- read.table(url881, sep = ";", header = FALSE)
stationcsv881A <- read.table(url881A, sep = ";", header = FALSE)

# Manually assign the header names
```

```
names(stationcsv881) <- c("YearMonth", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")
names(stationcsv881A) <- c("Year", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")

#create temp dataframe to split Year/Month
df881 <-data.frame(stationcsv881)
df881A <-data.frame(stationcsv881A)

#Split Year/Month
df881 <- df881 %>% separate(YearMonth, into = c("Year",
"Month"),sep=4)

#Reformat Month to decimal
df881$Month = paste0("0",df881$Month)
df881$Month <- as.numeric(as.character(df881$Month))

#Convert decimal back into Month Number
df881$Month <- ((df881$Month)*12)+0.5
df881$Month <- round(df881$Month)
```

```
     ##STATION 1222 - TOKYO II##

#Set URL for Data (Change number for station)
url1222 <-
"https://psmsl.org/data/obtaining/rlr.monthly.data/1222.rlrdata"
url1222A <-
"https://psmsl.org/data/obtaining/rlr.annual.data/1222.rlrdata"

#Set destination path for file to download, add "/output.csv" to
end of filepa/th (Downloads Folder)
destfile1222 <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1222.csv"
destfile1222A <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1222A.csv"

# Apply download.file function in R
download.file(url1222, destfile1222)
download.file(url1222A, destfile1222A)

#Read in txt station data .CSV into variable.
stationcsv1222 <- read.table(destfile1222, sep = ";", header =
FALSE)
```

```
stationcsv1222A <- read.table(destfile1222A, sep = ";", header =
FALSE)

# Manually assign the header names
names(stationcsv1222) <- c("YearMonth", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")
names(stationcsv1222A) <- c("Year", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")

#create temp dataframe to split Year/Month
df1222 <-data.frame(stationcsv1222)
df1222A <-data.frame(stationcsv1222A)

#Split Year/Month
df1222 <- df1222 %>% separate(YearMonth, into = c("Year",
"Month"),sep=4)

#Reformat Month to decimal
df1222$Month = paste0("0",df1222$Month)
df1222$Month <- as.numeric(as.character(df1222$Month))

#Convert decimal back into Month Number
```

```
df1222$Month <- ((df1222$Month)*12)+0.5
df1222$Month <- round(df1222$Month)




     ##STATION 1545 - TOKYO III##


#Set URL for Data (Change number for station)
url1545  <-
"https://psmsl.org/data/obtaining/rlr.monthly.data/1545.rlrdata"
url1545A <-
"https://psmsl.org/data/obtaining/rlr.annual.data/1545.rlrdata"


#Set destination path for file to download, add "/output.csv" to
end of filepa/th (Downloads Folder)
destfile1545  <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1545.csv"
destfile1545A <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1545A.csv"


# Apply download.file function in R
```

```
download.file(url1545, destfile1545)
download.file(url1545A, destfile1545A)

#Read in txt station data .CSV into variable.
stationcsv1545  <- read.table(destfile1545, sep = ";", header =
FALSE)
stationcsv1545A <- read.table(destfile1545A, sep = ";", header =
FALSE)

# Manually assign the header names
names(stationcsv1545)  <- c("YearMonth", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")
names(stationcsv1545A) <- c("Year", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")

#create temp dataframe to split Year/Month
df1545  <-data.frame(stationcsv1545)
df1545A <-data.frame(stationcsv1545A)

#Split Year/Month
df1545 <- df1545 %>% separate(YearMonth, into = c("Year",
"Month"),sep=4)
```

```
#Reformat Month to decimal
df1545$Month = paste0("0",df1545$Month)
df1545$Month <- as.numeric(as.character(df1545$Month))

#Convert decimal back into Month Number
df1545$Month <- ((df1545$Month)*12)+0.5
df1545$Month <- round(df1545$Month)



     ##STATION 1094 - HAKATA##

#Set URL for Data (Change number for station)
url1094  <-
"https://psmsl.org/data/obtaining/rlr.monthly.data/1094.rlrdata"
url1094A <-
"https://psmsl.org/data/obtaining/rlr.annual.data/1094.rlrdata"

#Set destination path for file to download, add "/output.csv" to
end of filepa/th (Downloads Folder)
```

```
destfile1094  <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1094.csv"
destfile1094A <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1094A.csv"

# Apply download.file function in R
download.file(url1094, destfile1094)
download.file(url1094A, destfile1094A)

#Read in txt station data .CSV into variable.
stationcsv1094  <- read.table(url1094, sep = ";", header =
FALSE)
stationcsv1094A <- read.table(url1094A, sep = ";", header =
FALSE)

# Manually assign the header names
names(stationcsv1094)  <- c("YearMonth", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")
names(stationcsv1094A) <- c("Year", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")

#create temp dataframe to split Year/Month
```

```
df1094 <-data.frame(stationcsv1094)

df1094A <-data.frame(stationcsv1094A)


#Split Year/Month

df1094 <- df1094 %>% separate(YearMonth, into = c("Year",

"Month"),sep=4)


#Reformat Month to decimal

df1094$Month = paste0("0",df1094$Month)

df1094$Month <- as.numeric(as.character(df1094$Month))




#Convert decimal back into Month Number

df1094$Month <- ((df1094$Month)*12)+0.5

df1094$Month <- round(df1094$Month)




     ## 2nd Populated station (Osaka)


#Set URL for Data (Change number for station)
```

```
url1099  <-
"https://psmsl.org/data/obtaining/rlr.monthly.data/1099.rlrdata"
url1099A <-
"https://psmsl.org/data/obtaining/rlr.annual.data/1099.rlrdata"

#Set destination path for file to download, add "/output.csv" to
end of filepa/th (Downloads Folder)
destfile1099  <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1099.csv"
destfile1099A <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1099y.csv"

# Apply download.file function in R
download.file(url1099, destfile1099)
download.file(url1099A, destfile1099A)

#Read in txt station data .CSV into variable.
stationcsv1099  <- read.table(url1099, sep = ";", header =
FALSE)
stationcsv1099A <- read.table(url1099A, sep = ";", header =
FALSE)
```

```
# Manually assign the header names
names(stationcsv1099)  <- c("YearMonth", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")
names(stationcsv1099A) <- c("Year", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")

#create temp dataframe to split Year/Month
df1099 <-data.frame(stationcsv1099)
df1099A <-data.frame(stationcsv1099A)

#Split Year/Month
df1099 <- df1099 %>% separate(YearMonth, into = c("Year",
"Month"),sep=4)

#Reformat Month to decimal
df1099$Month = paste0("0",df1099$Month)
df1099$Month <- as.numeric(as.character(df1099$Month))

#Convert decimal back into Month Number
df1099$Month <- ((df1099$Month)*12)+0.5
df1099$Month <- round(df1099$Month)
```

```
     ##3rd Nagoya II



#Set URL for Data (Change number for station)
url1488  <-
"https://psmsl.org/data/obtaining/rlr.monthly.data/1488.rlrdata"
url1488A <-
"https://psmsl.org/data/obtaining/rlr.annual.data/1488.rlrdata"

#Set destination path for file to download, add "/output.csv" to
end of filepa/th (Downloads Folder)
destfile1488  <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1488.csv"
destfile1488A <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1488A.csv"

# Apply download.file function in R
download.file(url1488, destfile1488)
download.file(url1488A, destfile1488A)
```

```
#Read in txt station data .CSV into variable.
stationcsv1488  <- read.table(url1488, sep = ";", header =
FALSE)
stationcsv1488A <- read.table(url1488A, sep = ";", header =
FALSE)

# Manually assign the header names
names(stationcsv1488) <- c("YearMonth", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")
names(stationcsv1488A) <- c("Year", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")

#create temp dataframe to split Year/Month
df1488 <-data.frame(stationcsv1488)
df1488A <-data.frame(stationcsv1488A)

#Split Year/Month
df1488 <- df1488 %>% separate(YearMonth, into = c("Year",
"Month"),sep=4)

#Reformat Month to decimal
df1488$Month = paste0("0",df1488$Month)
```

```
df1488$Month <- as.numeric(as.character(df1488$Month))

#Convert decimal back into Month Number
df1488$Month <- ((df1488$Month)*12)+0.5
df1488$Month <- round(df1488$Month)



     ## 4th most populated station (kitakyushu part mozi)



#Set URL for Data (Change number for station)
url912  <-
"https://psmsl.org/data/obtaining/rlr.monthly.data/912.rlrdata"
url912A <-
"https://psmsl.org/data/obtaining/rlr.annual.data/912.rlrdata"
#Set destination path for file to download, add "/output.csv" to
end of filepa/th (Downloads Folder)
#
destfile912  <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output912.csv"
```

```
destfile912A <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output912A.csv"

# Apply download.file function in R
download.file(url912, destfile912)
download.file(url912A, destfile912A)

#Read in txt station data .CSV into variable.
stationcsv912  <- read.table(url912, sep = ";", header = FALSE)
stationcsv912A <- read.table(url912A, sep = ";", header = FALSE)

# Manually assign the header names
names(stationcsv912)  <- c("YearMonth", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")
names(stationcsv912A) <- c("Year", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")

#create temp dataframe to split Year/Month
df912 <-data.frame(stationcsv912)
df912A <-data.frame(stationcsv912A)

#Split Year/Month
```

```
df912 <- df912 %>% separate(YearMonth, into = c("Year",
"Month"),sep=4)

#Reformat Month to decimal
df912$Month = paste0("0",df912$Month)
df912$Month <- as.numeric(as.character(df912$Month))

#Convert decimal back into Month Number
df912$Month <- ((df912$Month)*12)+0.5
df912$Month <- round(df912$Month)
```

```
     ## 5th Shizuoko (yoizu)
```

```
#Set URL for Data (Change number for station)
url1438 <-
"https://psmsl.org/data/obtaining/rlr.monthly.data/1438.rlrdata"
url1438A <-
"https://psmsl.org/data/obtaining/rlr.annual.data/1438.rlrdata"
```

```
#Set destination path for file to download, add "/output.csv" to
end of filepa/th (Downloads Folder)
destfile1438 <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1438.csv"
destfile1438A <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1438A.csv"

# Apply download.file function in R
download.file(url1438, destfile1438)
download.file(url1438A, destfile1438A)

#Read in txt station data .CSV into variable.
stationcsv1438<- read.table(url1438, sep = ";", header = FALSE)
stationcsv1438A<- read.table(url1438A, sep = ";", header =
FALSE)

# Manually assign the header names
names(stationcsv1438) <- c("YearMonth", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")
names(stationcsv1438A) <- c("Year", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")
```

```
#create temp dataframe to split Year/Month
df1438 <-data.frame(stationcsv1438)
df1438A <-data.frame(stationcsv1438A)

#Split Year/Month
df1438 <- df1438 %>% separate(YearMonth, into = c("Year",
"Month"),sep=4)

#Reformat Month to decimal
df1438$Month = paste0("0",df1438$Month)
df1438$Month <- as.numeric(as.character(df1438$Month))

#Convert decimal back into Month Number
df1438$Month <- ((df1438$Month)*12)+0.5
df1438$Month <- round(df1438$Month)


     ## 5th Hamamatsu (Omaezaki ii)


#Set URL for Data (Change number for station)
```

```
url1263 <-
"https://psmsl.org/data/obtaining/rlr.monthly.data/1263.rlrdata"
url1263A <-
"https://psmsl.org/data/obtaining/rlr.annual.data/1263.rlrdata"

#Set destination path for file to download, add "/output.csv" to
end of filepa/th (Downloads Folder)
destfile1263 <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1263.csv"
destfile1263A <- "C:/Users/jflor/Desktop/Data
Analytics/MAST5957/Dissertation/output1263A.csv"

# Apply download.file function in R
download.file(url1263, destfile1263)
download.file(url1263A, destfile1263A)

#Read in txt station data .CSV into variable.
stationcsv1263<- read.table(url1263, sep = ";", header = FALSE)
stationcsv1263A<- read.table(url1263A, sep = ";", header =
FALSE)

# Manually assign the header names
```

```
names(stationcsv1263) <- c("YearMonth", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")
names(stationcsv1263A) <- c("Year", "Mean_Sea_Level",
"Missing_Days_Flag","Flag_For_Attention")

#create temp dataframe to split Year/Month
df1263 <-data.frame(stationcsv1263)
df1263A <-data.frame(stationcsv1263A)

#Split Year/Month
df1263 <- df1263 %>% separate(YearMonth, into = c("Year",
"Month"),sep=4)

#Reformat Month to decimal
df1263$Month = paste0("0",df1263$Month)
df1263$Month <- as.numeric(as.character(df1263$Month))

#Convert decimal back into Month Number
df1263$Month <- ((df1263$Month)*12)+0.5
df1263$Month <- round(df1263$Month)
```

```
  ## Removing NA's


df881no.na <- subset(df881, Mean_Sea_Level > -99999)

df1545no.na <- subset(df1545, Mean_Sea_Level > -99999)

df1222no.na <- subset(df1222, Mean_Sea_Level > -99999)

df1094no.na <- subset(df1094, Mean_Sea_Level > -99999)

df912no.na <- subset(df912, Mean_Sea_Level > -99999)

df1438no.na <- subset(df1438, Mean_Sea_Level > -99999)

df1263no.na <- subset(df1263, Mean_Sea_Level > -99999)


df881no.naA  <- subset(df881A, Mean_Sea_Level > -99999)

df1545no.naA <- subset(df1545A, Mean_Sea_Level > -99999)

df1222no.naA <- subset(df1222A, Mean_Sea_Level > -99999)

df1094no.naA <- subset(df1094A, Mean_Sea_Level > -99999)

df912no.naA  <- subset(df912A, Mean_Sea_Level > -99999)

df1438no.naA <- subset(df1438A, Mean_Sea_Level > -99999)

df1263no.naA <- subset(df1263A, Mean_Sea_Level > -99999)

df1099no.naA <- subset(df1099A, Mean_Sea_Level > -99999)

df1488no.naA <- subset(df1488A, Mean_Sea_Level > -99999)


  ## Scatter Plot - Mean Sea Level vs Year - Station 1545##
```

```
     # STATION 1545 #


#Find Mean Sea Level Value for each year
meanYear1545 <- ddply(df1545no.na,"Year",numcolwise(mean))
meanYear1545 <- meanYear1545[,-c(2,4,5)]
meanYear1545$Year <- as.numeric(meanYear1545$Year)


#Plot Mean Sea Level
meanYear1545 %>% ggplot(aes(x=Year, y=Mean_Sea_Level)) +
geom_point() + geom_smooth(method='lm', se=FALSE,col= 'purple',
linetype='dashed') + theme(axis.text.x = element_text(angle =
90)) + ggtitle("Mean Sea Level by Year, Tokyo III") +
  xlab("Year") + ylab("Mean Sea Level")




## Time Series Modelling ##
```

```
# Time series Modelling Station 881 #


df881.ts <- df881no.na[,-c(4,5)]

df881A.ts <- df881no.naA[,-c(3,4)]


df881.ts['Date'] <- paste(df881.ts$Year, df881.ts$Month,

sep="/")

df881.ts$Date <- ym(df881.ts$Date)


class(df881.ts$Date)

frequency(df881.ts)


df881.ts$Date <-as.Date(df881.ts$Date, format = "%m-%Y")


df881.ts<- df881.ts[,-c(1,2)]

df881.ts <- df881.ts[,c(2,1)]


timeseries881 <- ts(df881.ts[,2], start=c(1957,07),

frequency=12)


#frequency(df881temp1)
```

```
plot(timeseries881, xlab = "Year", ylab = "Mean Sea Level", main
= "Time Series of Mean Sea Level recorded at Tokyo I Station")

#start(timeseries881)
#end(timeseries881)



# Time series Modelling Station 1222 #

df1222.ts <- df1222no.na[,-c(4,5)]
df1222A.ts <- df1222no.naA[,-c(3,4)]

df1222.ts['Date'] <- paste(df1222.ts$Year, df1222.ts$Month,
sep="/")
df1222.ts$Date <- ym(df1222.ts$Date)

class(df1222.ts$Date)
frequency(df1222.ts)

df1222.ts$Date <-as.Date(df1222.ts$Date, format = "%m-%Y")
```

```
df1222.ts<- df1222.ts[,-c(1,2)]
df1222.ts <- df1222.ts[,c(2,1)]

timeseries1222 <- ts(df1222.ts[,2], start=c(1968,01),
frequency=12)

#frequency(df1222temp1)

plot(timeseries1222, xlab = "Year", ylab = "Mean Sea Level",
main = "Time Series of Mean Sea Level recorded at Tokyo II
Station")

#start(timeseries1222)
#end(timeseries1222)



# Time series Modelling Station 1545 #

df1545.ts <- df1545no.na[,-c(4,5)]
df1545A.ts <- df1545no.naA[,-c(3,4)]
```

```
df1545.ts['Date'] <- paste(df1545.ts$Year, df1545.ts$Month,
sep="/")
df1545.ts$Date <- ym(df1545.ts$Date)

class(df1545.ts$Date)
frequency(df1545.ts)

df1545.ts$Date <-as.Date(df1545.ts$Date, format = "%m-%Y")

df1545.ts<- df1545.ts[,-c(1,2)]
df1545.ts <- df1545.ts[,c(2,1)]

timeseries1545 <- ts(df1545.ts[,2], start=c(1982,04),
frequency=12)

#frequency(df1545temp1)

plot(timeseries1545, xlab = "Year", ylab = "Mean Sea Level",
main = "Time Series of Mean Sea Level recorded at Tokyo III
Station")

#start(timeseries1545)
```

```
#end(timeseries1545)


# Combining Tokyo Stations Data, Plotting combined Data #

combTokyo.ts <- cbind(timeseries881,timeseries1222,
timeseries1545) # please make sure the length of both your
timeseries
plot.ts(combTokyo.ts, plot.type = "single", xlab = "Year", ylab
= "Mean Sea Level", main = "Time Series of Mean Sea Level
recorded at Tokyo I,II,II Stations")



# Time series Modelling Station 1094 #

df1094.ts  <- df1094no.na[,-c(4,5)]
df1094A.ts <- df1094no.naA[,-c(3,4)]

df1094.ts['Date'] <- paste(df1094.ts$Year, df1094.ts$Month,
sep="/")
df1094.ts$Date <- ym(df1094.ts$Date)
```

```
class(df1094.ts$Date)
frequency(df1094.ts)

df1094.ts$Date <-as.Date(df1094.ts$Date, format = "%m-%Y")

df1094.ts<- df1094.ts[,-c(1,2)]
df1094.ts <- df1094.ts[,c(2,1)]

timeseries1094<- ts(df1094.ts[,2], start=c(1965,11),
frequency=12)

#frequency(df1094temp1)

plot(timeseries1094, xlab = "Year", ylab = "Mean Sea Level",
main = "Time Series of Mean Sea Level recorded at HAKATA
Station")

#start(timeseries1094)
#end(timeseries1094)
```

```
# osaka


# Time series Modelling Station 1099 #


df1099.ts <- df1099[,-c(4,5)]

df1099A.ts <- df1099no.naA[,-c(3,4)]


df1099.ts['Date'] <- paste(df1099.ts$Year, df1099.ts$Month,

sep="/")

df1099.ts$Date <- ym(df1099.ts$Date)


class(df1099.ts$Date)

frequency(df1099.ts)


df1099.ts$Date <-as.Date(df1099.ts$Date, format = "%m-%Y")


df1099.ts<- df1099.ts[,-c(1,2)]

df1099.ts <- df1099.ts[,c(2,1)]


timeseries1099<- ts(df1099.ts[,2], start=c(1965,1),

frequency=12)
```

```
#frequency(df1094temp1)

plot(timeseries1099, xlab = "Year", ylab = "Mean Sea Level",
main = "Time Series of Mean Sea Level recorded at OSAKA
Station")

#start(timeseries1094)
#end(timeseries1094)




# nagoya ii

# Time series Modelling Station 1488 #

df1488.ts  <- df1488[,-c(4,5)]
df1488A.ts <- df1488no.naA[,-c(3,4)]

df1488.ts['Date'] <- paste(df1488.ts$Year, df1488.ts$Month,
sep="/")
df1488.ts$Date <- ym(df1488.ts$Date)
```

```
class(df1488.ts$Date)

frequency(df1488.ts)


df1488.ts$Date <-as.Date(df1488.ts$Date, format = "%m-%Y")


df1488.ts<- df1488.ts[,-c(1,2)]

df1488.ts <- df1488.ts[,c(2,1)]


timeseries1488<- ts(df1488.ts[,2], start=c(1965,11),

frequency=12)


#frequency(df1094temp1)


plot(timeseries1488, xlab = "Year", ylab = "Mean Sea Level",

main = "Time Series of Mean Sea Level recorded at NAGOYA II

Station")


#start(timeseries1094)

#end(timeseries1094)
```

```
# mozi ts


# Time series Modelling Station 912 #


df912.ts <- df912no.na[,-c(4,5)]

df912A.ts <- df912no.naA[,-c(3,4)]


df912.ts['Date'] <- paste(df912.ts$Year, df912.ts$Month,

sep="/")

df912.ts$Date <- ym(df912.ts$Date)


class(df912.ts$Date)

frequency(df912.ts)


df912.ts$Date <-as.Date(df912.ts$Date, format = "%m-%Y")


df912.ts<- df912.ts[,-c(1,2)]

df912.ts <- df912.ts[,c(2,1)]


timeseries912<- ts(df912.ts[,2], start=c(1965,11), frequency=12)


#frequency(df1094temp1)
```

```
plot(timeseries912, xlab = "Year", ylab = "Mean Sea Level", main
= "Time Series of Mean Sea Level recorded at MOZI Station")

#start(timeseries1094)
#end(timeseries1094)

# yoizu



# Time series Modelling Station 1094 #

df1438.ts <- df1438no.na[,-c(4,5)]
df1438A.ts <- df1438no.naA[,-c(3,4)]

df1438.ts['Date'] <- paste(df1438.ts$Year, df1438.ts$Month,
sep="/")
df1438.ts$Date <- ym(df1438.ts$Date)

class(df1438.ts$Date)
frequency(df1438.ts)
```

```
df1438.ts$Date <-as.Date(df1438.ts$Date, format = "%m-%Y")

df1438.ts<- df1438.ts[,-c(1,2)]
df1438.ts <- df1438.ts[,c(2,1)]

timeseries1438<- ts(df1438.ts[,2], start=c(1965,11),
frequency=12)

#frequency(df1094temp1)

plot(timeseries1438, xlab = "Year", ylab = "Mean Sea Level",
main = "Time Series of Mean Sea Level recorded at YOIZU
Station")

#start(timeseries1094)
#end(timeseries1094)


# omaezaki ii

# Time series Modelling Station 1094 #
```

```
df1263.ts <- df1263no.na[,-c(4,5)]

df1263A.ts <- df1263no.naA[,-c(3,4)]


df1263.ts['Date'] <- paste(df1263.ts$Year, df1263.ts$Month,

sep="/")

df1263.ts$Date <- ym(df1263.ts$Date)


class(df1263.ts$Date)

frequency(df1263.ts)


df1263.ts$Date <-as.Date(df1263.ts$Date, format = "%m-%Y")


df1263.ts<- df1263.ts[,-c(1,2)]

df1263.ts <- df1263.ts[,c(2,1)]


timeseries1263<- ts(df1263.ts[,2], start=c(1965,11),

frequency=12)


#frequency(df1094temp1)
```

```
plot(timeseries1263, xlab = "Year", ylab = "Mean Sea Level",
main = "Time Series of Mean Sea Level recorded at OMAEZAKI II
Station")

#start(timeseries1094)
#end(timeseries1094)



# Time series Modelling Station 1094 #

df1094.ts  <- df1094no.na[,-c(4,5)]
df1094A.ts <- df1094no.naA[,-c(3,4)]

df1094.ts['Date'] <- paste(df1094.ts$Year, df1094.ts$Month,
sep="/")
df1094.ts$Date <- ym(df1094.ts$Date)

class(df1094.ts$Date)
frequency(df1094.ts)

df1094.ts$Date <-as.Date(df1094.ts$Date, format = "%m-%Y")
```

```
df1094.ts<- df1094.ts[,-c(1,2)]
df1094.ts <- df1094.ts[,c(2,1)]

timeseries1094<- ts(df1094.ts[,2], start=c(1965,11),
frequency=12)

#frequency(df1094temp1)

plot(timeseries1094, xlab = "Year", ylab = "Mean Sea Level",
main = "Time Series of Mean Sea Level recorded at HAKATA
Station")

#start(timeseries1094)
#end(timeseries1094)




# Plotting Tokyo 1,2 and 3 together #

Tokyo1.xts <- as.xts(timeseries881)
```

```
Tokyo2.xts <- as.xts(timeseries1222)

Tokyo3.xts <- as.xts(timeseries1545)


TokyoNew1.xts <-

merge(Tokyo1.xts,Tokyo2.xts,Tokyo3.xts,all=TRUE)


plot(TokyoNew1.xts, main = "Tokyo Stations I, II, III - Mean Sea

Level", xlab = "Years", ylab = "Mean")




#TokyoNew.xts <-

merge(Tokyo1.xts,Tokyo2.xts,Tokyo3.xts,all=TRUE)

TokyoNew2.xts <- list(Tokyo1.xts,Tokyo2.xts,Tokyo3.xts)


do.call.rbind <- function(lst) {

  while(length(lst) > 1) {

     idxlst <- seq(from=1, to=length(lst), by=2)

     lst <- lapply(idxlst, function(i) {

     if(i==length(lst)) { return(lst[[i]]) }

     return(rbind(lst[[i]], lst[[i+1]]))

     })
```

```
  }
  lst[[1]]
}

TokyoStationJoint <- do.call.rbind (TokyoNew2.xts)

station1099.xts <- as.xts(timeseries1099)

NEW <- merge(TokyoStationJoint, Osaka.xts,all=TRUE)

plot(NEW, main = "Tokyo Station(s) vs Osaka Station - Mean Sea
Level", xlab = "Years", ylab = "Mean (mm)")



## PLotting Tokyo Stations vs Hakar

tokyo.ts <- ts(c(timeseries881,timeseries1222,timeseries1545),
     # Combined time series object
               start = start(timeseries881),
               end = end(timeseries1545),
               frequency = 12)
```

```
plot.ts(tokyo.ts)


tsunion <-
ts.union(timeseries881,timeseries1222,timeseries1545,timeseries1
094)
tsnew <- ts.union(tokyo.ts,timeseries1094)


plot.ts(tsunion)
plot.ts(tsnew)




# Converting the data frames into eXtensibke Time Series


UpdatedColTS <- TokyoStationJoint
names(UpdatedColTS) <- c("Mean_Sea_Level")



Hakata.xts   <- as.xts(timeseries1094)
```

```
Osaka.xts <- as.xts(timeseries1099)

Nagoya.xts   <- as.xts(timeseries1488)

Mozi.xts  <- as.xts(timeseries912)

Yaizu.xts <- as.xts(timeseries1438)

Omaezaki.xts <- as.xts(timeseries1263)



# Merging the dataframes


Joint_Station <- merge(UpdatedColTS,Osaka.xts, Nagoya.xts,

                    all = TRUE,

                    fill = NA,

                    suffixes = NULL,

                    join = "outer",

                    retside = TRUE,

                    retclass = "xts",

                    tzone = NULL,

                    drop=NULL,

                    check.names=NULL)


names(Joint_Station) <- c("Tokyo_Mean_Sea_Level",

"Osaka_Mean_Sea_Level", "Nagoya_Mean_Sea_Level" )
```

```
# Numerical Summary

profiling_num(Joint_Station)









# decomposition trend



#osaka


#decomposing into overall and seasonal trend
decomosa<- decompose(timeseries1099, type = "mult")
#multiplicative
plot(decomosa)



#binding tokyo station time series and imputing them using
seasonal splitting
```

```
tokallt <- ts_bind(timeseries881,timeseries1222, timeseries1545)
tokalltc <- na_seasplit(tokallt)#Splits the times series into
seasons and afterwards performs imputation separately for each
of the resulting time series datasets (each containing the data
for one specific season).




#plot test
plot.ts(tokalltc, plot.type = "single", xlab = "Year", ylab =
"Mean Sea Level", main = "Time Series of Mean Sea Level recorded
at Tokyo I,II,II Stations with gap filled")


#decomposing
decomtok<-decompose(tokalltc) #type additive


#x11 decomposition>>>
x11tok<- seas(x=tokalltc, x11 = "")
```

```
autoplot(tokalltc)


#plotting decomposition
autoplot(decomtok)
autoplot(x11tok)


autoplot(tokalltc, series="Data") +
  autolayer(trendcycle(x11tok), series="Trend") +
  xlab("Year") + ylab("New orders index") +
  ggtitle("Tokyo Mean Sea Levels") +
  scale_colour_manual(values=c("gray","blue","red"),
                      breaks=c("Data","Seasonally
Adjusted","Trend"))



#x11 decomposition>>>
#tokyo
x11tok<- seas(x=tokalltc, x11 = "")
```

```
#osaka
x11osak <- seas(x = timeseries1099, x11 = "")

#nagoya ii
x11nag <- seas(x=timeseries1488, x11 = "")

#plotting decomposition
autoplot(x11tok) + labs(title = "Decomposition of Mean Sea Level
recorded at Tokyo I,II,II Stations") + ylab("Mean Sea Level
(mm)") +xlab("Year") #tokyo
autoplot(x11osak) + labs(title = "Decomposition of Mean Sea
Level recorded at Osaka Station") + ylab("Mean Sea Level (mm)")
+xlab("Year") #osaka
autoplot(x11nag) + labs(title = "Decomposition of Mean Sea Level
recorded at Nagoya II Station") + ylab("Mean Sea Level (mm)")
+xlab("Year") #nagoya ii

# auto correlation graphs


layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE),
```

```
     widths=c(3,1), heights=c(1,2))
#osaka
autoplot(acf(timeseries1099,plot=FALSE))+
labs(title="Correlogram of Monthly Osaka Sea Levels 1965-2021")
#nagoya ii
autoplot(acf(timeseries1488,plot=FALSE))+
labs(title="Correlogram of Monthly Nagoya Sea Levels 1980-2021")
#tokyo all
autoplot(acf(tokalltc,plot=FALSE))+ labs(title="Correlogram of
Monthly Tokyo I, II and III Sea Levels 1980-2021")



# Box plots



#osaka
boxplot(timeseries1099~cycle(timeseries1099),xlab = "Date")

#nagoya ii
boxplot(timeseries1488~cycle(timeseries1488),xlab = "Date")

#all tokyo
```

```
boxplot(tokalltc~cycle(tokalltc),xlab = "Date")


# indicate seasonal fluctuations

# Predictions


##OSAKA##

#### 1st Video Method ARIMA ####

#Plot Osaka time series
plot(timeseries1099)

#The data has a clear positive trend

adf.test(timeseries1099, k=12)

#P-value greater than 0.01 - 1%

diff1099 <- diff(timeseries1099, differences =1)
```

```
adf.test(diff1099, k=12)

#P-value is 0.01, First Difference removes trend and leaves
Stationary series, d term = 1

plot(diff1099)

#Note the time series looks stationary

Pacf(diff1099)

# p = 4

Acf(diff1099)

# q = 2

tsMod1099 <- Arima(y=timeseries1094, order = c(4,1,2))

#print model

print(tsMod1099)
```

```
#forecasting, h <- how far to forecast

forecast.1099 <- forecast(tsMod1099, h =24)
autoplot(forecast.1099)

##nagoya differencing

diff1488<- diff(timeseries1488, differences =1) #differencing
nagoya ii time series

#plotting seasonal plots



ggseasonplot(diff1488) +
  ggtitle("Seasonal Plot, Nagoya II") +ylab("Mean Sea Level")

ggsubseriesplot(diff1099) +
  ggtitle("Subseries Seasonal Plot, Nagoya II") + ylab("Mean Sea
Level")
```

```
#### 2nd Video Method Forecasting Seasonal Naive, ETS####

#Using first difference, removes trend - stationary series

#Seasonal plot

ggseasonplot(diff1099) +
  ggtitle("Seasonal Plot, Osaka") + ylab("Mean Sea Level")

#point out clear seasonal patterns

#subseries seasonal plot
ggsubseriesplot(diff1099) +
  ggtitle("Subseries Seasonal Plot, Osaka") + ylab("Mean Sea
Level")

#point out further seasonal patterns + averages.

ggseasonplot(difftokyo) +
  ggtitle("Seasonal Plot, Tokyo") + ylab("Mean Sea Level")
```

```
#point out clear seasonal patterns


#subseries seasonal plot
ggsubseriesplot(difftokyo) +
  ggtitle("Subseries Seasonal Plot, Tokyo") + ylab("Mean Sea
Level")

# Our time series has trend and seasonality.
#To remove the trend we take the first difference
#The first difference series still has seasonality



#Forecasting with benchmark method, Seasonal Naive method as
benchmark

fit1099 <- snaive(diff1099)
print(summary(fit1099))
checkresiduals(fit1099)

### Forecasting - Fit ETS (exponential time series model)
fit_ets1099 <- ets(diff1099)
print(summary(fit_ets1099))
```

```
checkresiduals(fit_ets1099)


## Auto Arima - second version
fit_arima1099 <-
auto.arima(timeseries1099,d=1,D=1,stepwise=FALSE,approximation=F
ALSE,trace=TRUE)
print(summary(fit_arima1099))
checkresiduals(fit_arima1099)


plotforecast1099 <- forecast(fit_arima1099, h=120)
autoplot(plotforecast1099, include=684)


plotforecasttokyoets <- forecast(fit_etstokyo, h=120)
autoplot(plotforecasttokyoets, include=768)


plotforecasttokyosnaive <- forecast(fittokyo, h=20)
autoplot(plotforecasttokyosnaive, include=768)



traintokyo <- subset(tokalltc, start=1, end=538)
testtokyo <- subset(tokalltc, start=538, end=768)
```

```
fit_arimatokyotrain <-
auto.arima(traintokyo,d=1,D=1,stepwise=FALSE,approximation=FALSE
,trace=TRUE)
print(summary(fit_arimatokyotrain))
checkresiduals(fit_arimatokyotrain)

#training set uses 70% or 538 months/44.83 years.
#leaving 230 months or 19.16 years to forecast and compare
against real results

forecasttraining <- forecast(fit_arimatokyotrain, h=230)



autoplot(tokalltc) +
  autolayer(forecasttraining, series = "Forecast", PI=FALSE) +
  xlab("Year") +
  ylab("Mean Sea Level")
guides(colour=guide_legend(title="Forecast"))

accuracy(forecasttraining, tokalltc)

summary(forecasttraining)
```

```
checkresiduals(forecasttraining)
```