University of Essex
Department of Mathematical Sciences

MA981: DISSERTATION

Global Temperature Forecasting Using Comparative Analysis of ARIMA Model On Two Different Datasets

Deepanshu Lamba
2211688

Supervisor: Dmitry V Savostyanov

November 24, 2023
Colchester

Abstract

This study scrutinize a comprehensive analysis of global temperature variations by using **ARIMA(AutoRegressive Integrated Moving Average)** model on two different datasets, named as ***dataset A*** and ***dataset B***. The study depicts variations of temperature from 1961 to 2022 (this data is taken from ***dataset A***) and from 1850 to 2015 (this data is taken from ***dataset B***), focusing on different aspects like global temperature change, trends in temperature in decade wise, patterns and anomalies in tempreature changes and a comparisions between developed and developing nations. Furthermore, the research delves into trend analysis of temperature over time, shows uncertainties of temperature records over the years, seasonal temperature variations and fluctuations in temperature on different continents. The study also discusses about the previous temperature records and how it is moving upwards in each and every year and it also depicts about the rise in global warming every year and according to **IPCC (Intergovernmental Panel on Climate Change)** and 2015 **Paris COP21 Agreement**, why there is a need to put the limit on temperature at 1.5 degree Celsius.

The ARIMA model is evaluated by using various statisical methods, which includes seasonlity check, stationarity test by using **ADF** test and model diagnostics. The **MSE (Mean Squared Error)**, **RMSE (Root Mean Squared Error)** and **MAE (Mean Absolute Error)** provides the model's performance, how well the model is fitted, how good or bad the results are, can be analysed by these errors scores. Moreover, the study also depicts the use of **hyperparameter tuning** on the ARIMA model in forecasting task. The results are compared with and without the hyperparameter tuning on both datasets, by using the results of **MSE, RMSE** and **MAE**.

Contents

1 Introduction **8**

1.1 Background . . . 8

1.2 Objectives . . . 14

1.3 Research Questions . . . 14

2 Literature Review **16**

3 Methodology **21**

3.1 Data Collection . . . 21

3.2 Preprocessing and cleaning . . . 24

3.3 Descriptive Analysis . . . 28

3.4 ARIMA Model . . . 30

3.5 Data Visualizations . . . 35

3.5.1 Box-plot of Temperature Change by Decade in the United Kingdom . 35

3.5.2 Seasonal Analysis: Average Temperature Changes Across Months . . 36

3.5.3 Comparative Analysis: Land vs Land and Ocean Temperature Trends 37

3.5.4 Correlation Analysis: Heatmap . . . 38

3.5.5 Yearly Temperature Difference (Max - Min): . . . 39

4 Results **41**

4.1 *Answer 1*: Global average temperature change from 1961-2022. . . . 41

4.2 *Answer 2*: Surface temperature throughout the decades. . . . 42

4.3 *Answer 3*: Patterns or anomalies in the temperature changes over the years. . 43

4.4 *Answer 4*: Trends difference between developed and developing nations. . . . 44

4.5 *Answer 5*: Trend Analysis: Temperature Over Time. . . . 45

4.6 *Answer 6*: Uncertainties of temperature over time. . . . 46

4.7 *Answer 7*: Temperature changes in decades. 48
4.8 *Answer 8*: Global average land temperature by season. 49
4.9 **Answer 9**: Temperature changes in different continents. 50
4.10 *Answer 10*: Findings of ARIMA model on both datasets. 51

5 Conclusions **70**

List of Figures

1.1 The image shows that global warming does indicate temperature rise all over the place [1]. . . . 8
1.2 (Left) Carbon emissions throughout $20th$ Century, (Right) Project Temperature rise [2, 21]. . . . 10
1.3 Types of greenhouse gases [3]. . . . 11
1.4 Change in heat in the ocean from $1993-2022$ in the top 2300 feet. or 700 metres. Upto 6 watts per square metres heat rised, shown by dark orange [33]. . . . 12
1.5 Warming records over the years [1]. . . . 13

3.1 The image shows first 10 rows of the dataset 1. . . . 22
3.2 The image shows first 5 rows of the dataset 2. . . . 24
3.3 The image shows first 5 rows of the dataset 2. . . . 26
3.4 A general ARIMA image [71]. . . . 30
3.5 ACF and PACF plot samples [76]. . . . 34
3.6 UK temperature change in decades. . . . 36
3.7 Monthly temperature change. . . . 36
3.8 Land vs Land and Ocean Temperature Trends. . . . 37
3.9 Correlation Analysis 38
3.10 Yearly Temperature Difference (Max - Min) 39

4.1 Temperature change from 1961-2022. . . . 41
4.2 Surface temperature throughout the decades. . . . 43
4.3 Patterns or anomalies in the temperature changes. . . . 44
4.4 Surface temperature change in developed countries. . . . 45
4.5 Surface temperature change in developing countries. . . . 46
4.6 Land temperature over the years. . . . 47

4.7 Uncertainties of temperature over the years. 47
4.8 Average temperatur change in decades. 48
4.9 Average land temperature change in season. 49
4.10 Average temperature change in different continents. 50
4.11 Seasonality showing. 51
4.12 ACF & PACF plots. 54
4.13 Model summary result. 55
4.14 Model summary result. 57
4.15 Diagnostic plot. 58
4.16 Model summary after hyperparameter tuning. 59
4.17 Forecast after hyperparameter tuning. 60
4.18 Diagnostics after hyperparameter tuning. 60
4.19 Diagnostics after hyperparameter tuning. 61
4.20 ACF and PACF plot of ***dataset B***. 63
4.21 Model summary of ***dataset B***. 63
4.22 Forecast plot of ***dataset B***. 65
4.23 Residual plot of ***dataset B***. 65
4.24 Model summary after **hyperparameters**. 66
4.25 Forecast after **hyperparameters**. 67
4.26 Diagnostic plot after **hyperparameters**. 67
4.27 Next 5 years forecasting on the UK data. 69

List of Tables

3.1 Number of Missing Values in each Column 25
3.2 Percentage (%) of Missing Values in each Column 25
3.3 Annual Temperature Change 26
3.4 Temperature Data (Part 1) 27
3.5 Temperature Data (Part 2) 27
3.6 Descriptive Statistics of UK Temperature Change 28
3.7 Descriptive Statistics of Dataset 2 (Part 1) 29
3.8 Descriptive Statistics of Dataset 2 (Part 2) 29

4.1 ADF Test Results 52
4.2 ADF Test Results after differencing. 53
4.3 ADF test result on ***dataset B***. 62
4.4 ADF result after differencing on dataset B. 62

CHAPTER 1

Introduction

1.1 Background

Earth's temperature has been rising from the past few decades [2, 7]. Rising temperature is one of the main problems of the Earth [2]. Storms, heat waves, floods, and droughts are just a few of the calamities that are getting worse due to rising temperatures. Warmer temperatures provide an atmosphere with increased capacity to absorb, hold, and release water, altering weather patterns to make wet regions wetter and dry ones drier [13].

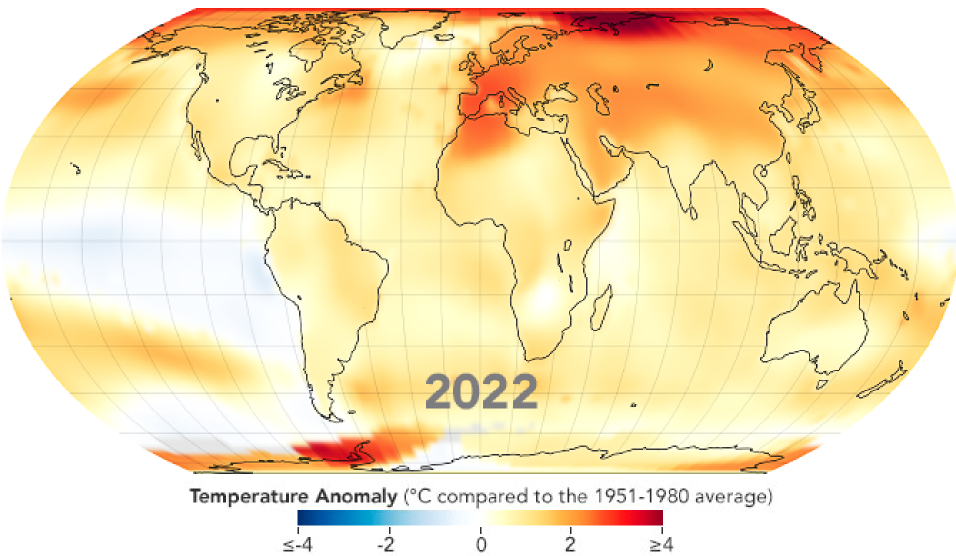

Figure 1.1: The image shows that global warming does indicate temperature rise all over the place [1].

In one part, temperature might face a 5 degree Celsius rise, while at the same time it might face a downfall of 2 degrees Celsius. For instance, if one place on the globe having extraordinarily mild winters and in other part might cancel out abnormally hot winters. As water does not soaks heat quickly as compared to land [1].

The more heat causing temperature rise in regional and seasonal aspects, minimizing the sea ice and snow layer, causing strong rainfall, natural environment for few plants and animals is changing [2]. As compared to ocean locations, the land locations are heating more quickly [2]. The air temperature of Earth has been increasing ever since the Industrial Revolution. The data depicts that the human tasks or activities are the main reason behing the increase of greenhouse gases, one of the reasons for the warming of planet, albeit natural variability also depict a role.

Since the pre-industrial or industrial revolution, temperature of earth has been rising [1, 5]. Since 1880, the Earth's temperature has been rising. Based on the scientists' report at **NASA's Goddard Institute for Space Studies (GISS)** [1], since 1880 by at least 1.1 degree Celsius (1.9 degree Fahrenheit) temperature on Earth has increased. Furthermore, similar analysis showed by the climate report by [2, 8, 9], that around 1 degree Celsius or roughly 2 degrees Fahrenheit increased in the land temperature of Earth, which has appeared since the pre-industrial era, which is $1880 - 1990$ [1, 6].

According to the **Annual Climate Report 2021** of **NOAA's** [2], since 1981 the rate of temperature increase has been more than double what it was before with a rise of 0.32 degrees Fahrenheit or 0.18 degrees Celsius, per decade [2, 7, 17]. It seems casual, but it is showing a serious rise in heat buildup [2]. Because of increase or rise of 2 degrees Celsius in the temperature in contrast with the temperature in pre-industrial times is related with perilous detrimental influence on the habitat and human well-being or physical condition and well being, consists of much high elevated hazard that perilous and perhaps devastating changes in the global environment will happen [3]. Due to this, the global community has realized the necessity to maintain warming under 2 degrees Celsius and seek an attempt to restrict it to 1.5 degree Celsius [3, 5]. By keeping global warming to 1.5 degrees Celsius, almost 420 million lesser people would be often defenceless to intense hot waves and about 65 million lesser people would be defenceless to unexpected hot waves [5].

Global warming is one of the reasons in the global temperature hike. The Earth's temperature and climate are being affected by the ignition of fossil fuels, fragmenting down forests and animal rearing or farming livestock [2, 3, 17]. An extensive volume of greenhouse gases generate because of this, which add to atmosphere and causing the hike in global warming and the greenhouse effect [3].

The heat on the planet, from the Arctic Pole and to the Antarctic Pole, the heat is evolving rapidly [10]. The hike of more than 1.6 degrees Celsius Fahrenheit or 0.9 degree Celsius average temperature, even higher in fragile polar region, has been experiencing globally [10]. The consequences of global warming are becoming visible right now, the influence of increasing temperatures is not delayed for some extensive future. Glaciers and sea ice are melting by the heat, changing precipitation figures and animal locomotion, moving animals from place to place [11, 10].

The path of future human activities will be responsible for the seriousness of results induced by climate change [2, 4, 10, 12, 13, 17]. Increased greenhouse gas ejaculation will result in more climate peaks and extensive detrimental effects that will affect the entire globe [12]. Nevertheless, those future consequences are contingent upon the total volume of carbon dioxide we release [12]. Consequently, if we can minimize ejaculations, we might be able to prevent some of the worst impacts [3, 12]. Particularly the carbon contamination we induce by flaming fossil fuels and by tearing down the forests we avert pollution capture [2, 13]. The cause of warming the planet is that we deliver carbon dioxide, methane, soot and other pollutants into the environment to play as a blanket and trap the sun's heat [13].

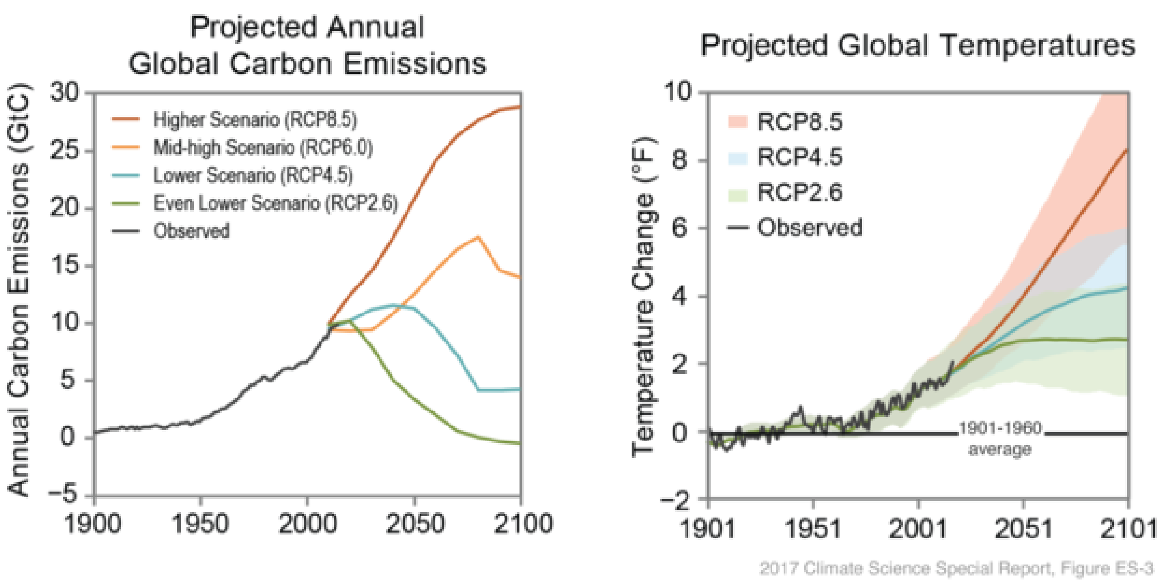

Figure 1.2: (Left) Carbon emissions throughout $20th$ Century, (Right) Project Temperature rise [2, 21].

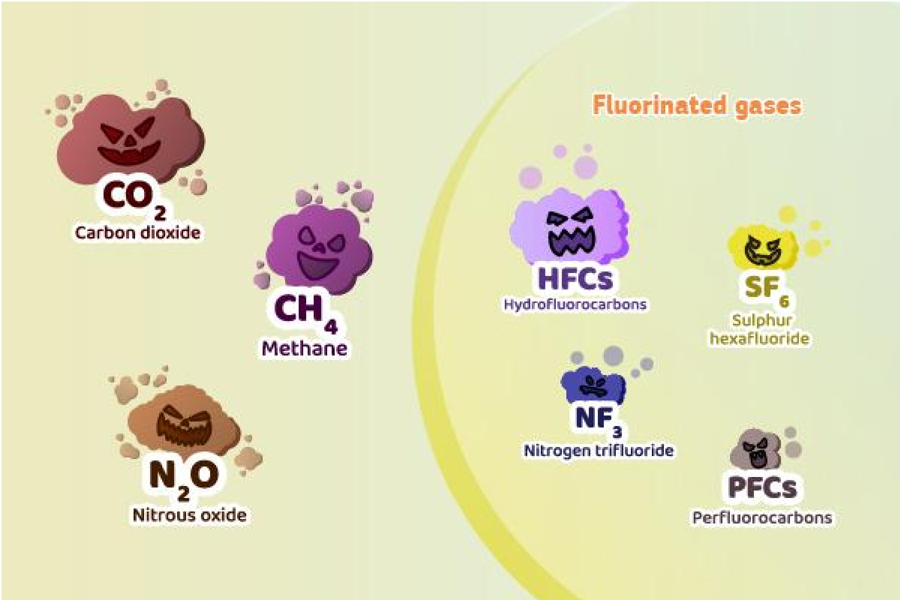

Figure 1.3: Types of greenhouse gases [3].

Greenhouse effect is the main reason for climate change [3]. Few gases in the atmosphere of Earth play similar to the glass in a greenhouse, absorbing the heat of the sun and creating global warming by stopping it from leaking back into space [3]. Some gases produced by human activities such as CO_2 and methane [3]. Human activity is the main creator of emissions that contribute to global warming [2, 3, 4, 10].

Some other greenhouse gases which are in lesser volume are ejaculated by human activities [3]. Methane has a lesser atmospheric timespan, but is extra fierce greenhouse gas than CO_2 [3]. Just like CO_2, nitrous oxide is a long lasting greenhouse gas that assembles in the atmosphere over millenia to decades. Aerosols like soot are instances of non-greenhouse gas impurities that have varying warming and cooling effects. They are also related to other concerns like awful air quality. CO_2 is the biggest participant in global warming, which is caused by human activities [2, 3, 10]. In the atmosphere, the involvement of CO_2 had increased to 48% above its preindustrial level (before 1750) [3].

The World's coastlines will have serious alteration due to this boosting in the sea level [4]. Sea levels rise as a result of warm water growth and ice sheets or glaciers defrosting [10]. The ocean's high heat is distributed unevenly, the southern hemisphere facing highest ocean warming, which contributes to the Antarctic ice shelf's subterranean melting [30, 31]. Thinning of ice shelves and sea ice is associated with heating of sea water. Eventually, marine ecosystems and human livelihoods are threatened by sea boiling [32].

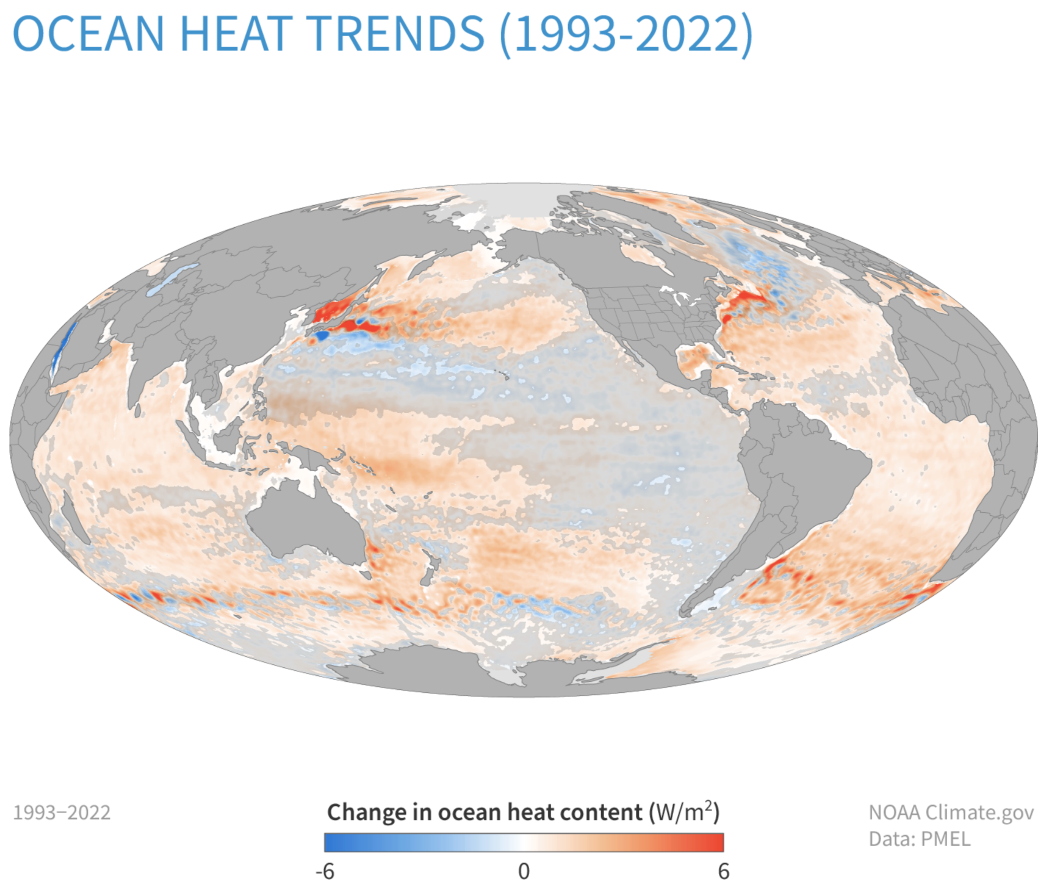

Figure 1.4: Change in heat in the ocean from $1993-2022$ in the top 2300 feet. or 700 metres. Upto 6 watts per square metres heat rised, shown by dark orange [33].

The rise in the sea level will keep on going, even after the limit of 1.5 degree Celsius, as the heat is already absorbed by the oceans [5]. As compared to 2 degrees Celsius, the rise is lower at 1.5 degree Celsius projected to be 0.33 feet or 0.1 metre [5, 12]. If the warming hits 2 degrees Celsius, over two-third percent of Earth's coastlines will face the increase in sea level higher than 0.66 feet or 0.2 metre, causing coastal flooding, beach erosion and other things [5, 10, 11, 12].

The mass of ice sheets of Antarctica and Greenland have declined dramatically [11, 23, 28, 29, 35, 36]. As per the report of **NASA's Gravity Recovery and Climate Experiment**, it appears that per year, around 286 billion tons of ice of Greenland has vanished [4, 23, 28]. More water goes inside the ocean from ice layers and glaciers when the temperature increases and ice melts and the sea water heats up and extends in quantity. As per the **IPCC(Intergovernmental Panel on Climate Change)** report [15], in the past 100 years, in elevating the average sea level of the planet by 4 to 8 inches or 10 to 20 centimetre, this merged effect has depicted an important part [4, 12, 29]. The southern ocean can be disturbed by defrosting ice from Antarctica by constructing a coating of freshwater that captures heat underneath, causing ocean warming [4]. The ocean's overall temperature will rise by this, the absorption of CO_2 will be less from the atmosphere because of this. Ultimately, more CO_2 will stay around in the atmosphere, causing global warming [4]. The **IPCC report** proclaims [15], with a huge warming level between 1.5 degrees and 2 degrees Celsius, instability in the Antarctic ice layer and the irreparable damage of the Greenland ice layer might go to higher than 2 meter hike in sea level more than a time span of centuries [5, 6, 15].

In accordance with **NOAA's Global Climate Report**, since 1880, the $6th$ hottest year was the year 2022 [2, 7]. The surface temperature of 2022 was 0.86 degree Celsius or 1.55 degree Fahrenheit which is hotter or higher than the $20th$ century of 13.9 degrees Celsius or 57 degree Fahrenheit and higher than 1.06 degrees Celsius or 1.90 degrees Fahrenheit [2, 7]. It is less than the record set in 2016, which was 0.13 degree Celsius or 0.23 degree Fahrenheit and it is just 0.02 degree Celsius or 0.04 degree Fahrenheit more than the previous year's (which was year 2021) value, which now classifies as the $7th$ highest [2, 7]. Since 2010, 10 warmest years occurred in the record of 143 years, nine hottest years were the previous nine years from 2014 to 2022 [2].

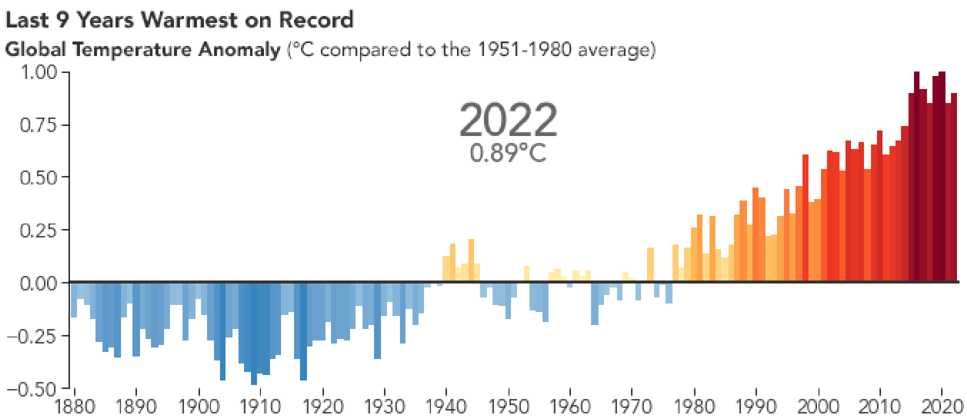

Figure 1.5: Warming records over the years [1].

As per the **NCEI**, in 2022 every month was rated within the ten hottest of that particular month, notwithstanding the frosting effect from the **La Niña** (also known as **"El Niño Southern Oscillation"**) weather trend in the tropical Pacific [7, 36]. November was the coldest month, which was 0.75 degree Celsius or 1.35 degree Fahrenheit hotter than normal [2]. At $+1.10$ degrees Celsius or $+1.98$ degrees Fahrenheit, the Northern Hemisphere's land temperature in 2022 was likewise the $6th$ hottest year on account. Simultaneously, the $7th$ hottest year for the Southern Hemisphere with 0.61 degree Celsius or 1.10 degree Fahrenheit [7].

1.2 Objectives

The main purpose of this research is to analyse the previous records and information of the global temperature and to perform **ARIMA** model on the available dataset. In order to do so, we have gone some deep research and have gone through some previous researches related to this topic. We performed **ARIMA** model on two different dataset, which are available online can easily accessible for anyone who wants to do some work related to this. The exact and proper information about the dataset is given in the **"Data Collection"** section. The main motive of using the **ARIMA** model is to check that if the **ARIMA** model gives the better or nearly result to actual data or not. The outcomes of this task is discussed further in the paper.

1.3 Research Questions

Some problems can be identified from this research. Those are mentioned below:

- **Q1:** How is the global average temperature change from $1961 - 2022$?
- **Q2:** What variations have there been in surface temperature throughout the decades?
- **Q3:** Are there any noticeable patterns or anomalies in the temperature changes over the years?
- **Q4:** Do temperature trends differ between developed and developing nations?
- **Q5:** What trends of temperature can be seen over time?
- **Q6:** What are the uncertainties of temperature over time?
- **Q7:** How is the temperature changes in decades?
- **Q8:** How do temperature changes vary across different seasons globally and in specific regions?

- **Q9:** How annual temperature changes in different continents over the time?
- **Q10:** What are the results or outcomes of ARIMA model performed on both datasets?

These are discussed in the **Result** (*Chapter 4*) later.

CHAPTER 2

Literature Review

Global temperature has been increasing every day rapidly [2]. It is one of the mutual problems for each and every nation of this world. Due to the practical importance of weather forecasting in meteorology and the wide public interest in scientific study, it has been one of the most difficult problems to solve globally [38]. Few studies have been done on this topic. Various reasons are described for the increase of the temperature such as greenhouse gas, global warming and others ("Causes of climate change" [3]). Discussions about the causes of increasing temperature have been done and the methods and techniques and research to the problems. For making the prediction, several methods or models were used in some previous research.

Few decades back, forecasting temperature was a challenge [38]. The scrutiny inspects the factors of climate change conversations and the consequences for setting temperature goals. In determining a pre-industrial reference period and predicting the change in global temperature since then are discussed in [38]. The research implies that the most proper information for the preindustrial period is $1720 - 1800$, notwithstanding details from this period is finite [38, 39, 40]. The period of $1850 - 1900$ is recognised as a feasible replacement [42, 43]. The research forecast an escalate in the temperature from the limit between 0.55 degree Celsius to 0.80 degree Celsius from preindustrial to $1986 - 2005$ time span [38]. The primary year in which temperature of Earth was higher than 1 degree Celsius higher than pre-industrial period was 2015 [38]. In accordance with **Paris COP21 Agreement** 2015, the goal is to keep the temperature of Earth under 2 degrees Celsius higher than pre-industrial levels and seeking attempts to restrict the temperature rise to 1.5 degrees Celsius higher than pre-industrial marks [39, 40, 41]. The actual global temperature may be higher than previously stated. The writer also discussed radiative forcings for determining pre-industrial period [38, 44]. In calculating historical temperature fluctuations and the outcomes for policy, the challenges and difficulties are included.

A sequence of trustworthy global temperature evaluation started in the $1850 - 1880$ lapse of time [4]. After 1940, the average annual temperature rose, but was fairly constant until 1975. Since 1975, it has risen by more or less 0.15 degree Celsius to 0.20 degree Celsius per decennary, to at least 1.1 degree Celsius or 1.9 degree Fahrenheit, which is above 1880 levels [1, 4]. The present yearly **GMST** (global mean surface temperature) is near 15 degree Celsius or 59 degrees Fahrenheit [14], nevertheless monthly temperatures can fluctuate approximately 2 degrees Celsius or 4 degrees Fahrenheit more or less this data [4]. According to the **Annual Climate Report 2021** of **NOAA's**, as per the analysis of the **United Nations (**2015**)**, to keep the average global temperature "under 2 degree Celsius over pre-industrial levels and seek attempt to control the hike in temperature to 1.5 degree Celsius, which is over pre-industrial levels" aim of the **2015 Paris COP21 Agreement** [1, 11]).

Proofs display that the $2010s$ were warmer than any other decennary on data and since the $1960s$ every decennary has aggregated hotter than the last one [13]. The climate mechanism of Earth is changing by this warming, containing its land atmosphere, ice and ocean, in a wide-reaching approach [12, 13]. By 2020, its attentiveness into the aerosphere had increased to 48% above its pre-industrial level, which means here 1750 [3]. $2011 - 2020$ was the hottest decennary observed, with average temperature of Earth getting to 1.1 degree Celsius more than pre-industrial levels in 2019 [3]. According to **IPCC report** [15], since the pre-industrial period, people actions are considered to have raised the average temperature of Earth by almost 1 degree Celsius or 1.8 degree Fahrenheit [2, 6, 12], a figure that is now rising by 0.2 degree Celsius or 0.36 Fahrenheit, each decade [6]. At that pace, within the upcoming few decades between the year 2030 and the year 2052, global warming is expected to touch 1.5 degree Celsius or around 3 degrees Fahrenheit above pre-industrial levels, with a perfect approximation of around year 2040 [6, 12, 17]. These transitions will strike all regions of Earth [12].

A study has been done on land and ocean temperature merged together and did the forecast and solved or trying to solve the related issues [45]. The analysis investigated five datasets, three for surface temperature (from **NASA/GISS** [37], **NOAA/NCDC** [46, 47] and **HadCRU** [48]) and two for lower troposphere (from **RSS** [49] and **UAH** [50]) which are depend on satellite microwave sensors. These five datasets manifested global warming patterns from 1979 to 2010, varying from 0.014 to 0.018 degree Celsius per annum [45]. The scrutiny exhibits how familiar elements like **ENSO** or **El Niño**, solar variations and volcanic aerosols, the global warming signal became more obvious [45, 51, 52]. The research proposes that the lower troposphere is highly responsive to **El Niño** and to fiery effects than temperature of land [53]. Sea level hike, decreased Arctic sea ice extent, elevated ice loss from Greenland and ice sheets of Antarctica were the causes responsible for volcanoes and a broad range of warming associated influences [51]. Man-made heating are the dominant sources for these warming indications [45]. This generated the necessity to entice human induced climate variation crucially, as in upcoming decades further temperature rises are envisioned [45].

Global temperature is rising upwards. A study shows that a few years back, 2014, 2015 and 2016 were the hottest years, also dealing with the pitfalls in recent trends in these years [54]. This utilised 5 different datasets for displaying the patterns or trends in global warming. The author states that there is no arithmetical proof for an elevation of global warming from the immediate temperatures data and neither there was a slowdown period with noteworthy pace of warming [54, 55, 56]. Instead, they propose that since the $1970s$ the temperature records are in tune with a stable warming pattern, along the side of casual temporary variance [54]. Entrenched for a limited record-hot years it debates declaration of a quick elevation in global warming and disproves the concept that there was a **"hiatus"** or **"pause"** in global warming between 1998 and 2014 [55, 56, 57]. For affirming the notable warming slowdown, it utilized few statistical evaluation techniques, displaying that these arguments do not grasp up under careful statistical study [54]. The study suggests that since 1970 there has been no measurable alter in global warming and that recent variations in temperature patterns are not statistically abnormal in contrast to the everlasting patterns and its changeability [54].

Various natural factors of Earth are affecting, as global surface temperature increases [58]. With Earth's temperature increasing significantly since 1880, particularly from 1981 onwards with double the rate at 0.32 degree Fahrenheit or 0.18 degree Celsius, the research outlines the accelerated rate of global warming [2, 7, 17, 58]. According to **NOAA** data, the year 2021 was the sixth highest on record [58, 62]. The detrimental chain reaction could happen to the ecosystem due to this warming [58]. The negative impacts of change in climate on ecosystems and human communities, including the severe effects of storms, droughts and rising sea levels [10, 11, 58]. In order to slow down global warming it is crucial to cut greenhouse gas emissions, discussed in the research. The paper explores the prediction of global temperature changes from 1961 to 2020 through various machine learning algorithms, by using the **NASA-GISS** temperature data [58]. With an **RMSE** score of 0.3998 degree Celsius and a maximum error of 3.5 degrees Celsius, the **extra trees algorithm** was found to be the most effective for prediction, on the other hand **Bayesian Ridge** performed the least effectively [58, 59, 60, 61]. Reducing human emissions to mitigate the impact of global warming, discussed in the study and also the potential of machine learning algorithms for comprehending and forecasting changes in the Earth's temperature [58].

During the Earth's 4.6 billion year record, the height sea levels have increased and declined rapidly [29]. With the rising Earth's temperature, the extended heat has been soaked by the oceans, since 1969, the heating of 0.22 degree Celsius or 0.4 degree Fahrenheit can be seen on the top 700 metres of ocean [28]. Nevertheless, rising global temperatures have caused the rise in the global sea level recently [4]. From 1901 to 2015, during the $20th$ century, sea surface temperatures increased steadily with an average rise of 0.13 degree Fahrenheit per decade [34]. The reason for this growth is absorption of solar energy by greenhouse gases, which raises temperatures and sea levels. Stronger tropical storms may arise from these alterations which might have serious impacts on coastal communities, comprising property loss and losing life [4, 11].

For grasping the processes of the world's climate and avert disastrous weather tragedies [63]. The research indicates the establishment and implementation of a latest machine learning method, the **LSTM-AdaBoost amalgamation** model, for forecasting small and mid-term daily **SSTA (Sea Surface Temperature Anomalies)** [63, 68]. This model combines the **LSTM** deep recurrent neural network, which is good at capturing **long-term dependencies** but is susceptible to overfitting [67], with the **AdaBoost ensemble learning** model, widely known for its strong forecast ability and resistance to overfitting [63, 64, 65, 68]. The method eradicates seasonality, trains the **LSTM** and **AdaBoost** models separately and makes utilize of historical **SSTA** (SST anomaly) satellite data [63]. The technique desires to minimize forecasting blunders and enhance precision, by averaging the forecasts from these two models [63]. A case study of **East china sea**, in the matter of several error statistics, the

LSTM-AdaBoost model has been found that it conquer **AdaBoost** and **LSTM** individually and in addition to other models such as SVR, BPNN and even a stacking **LSTM-AdaBoost** model, validating its effectiveness for SST forecasts that are critical for apprehension and reducing the outcome of climate change and serious weather incidents [63, 66]. To forecast more atmospheric, environmental and oceanic aspects, this technique could be expanded in the future [63].

In contrast to these studies, a straightforward correlation between fluctuations in global sea-level and fluctuations in global mean temperature [69]. By utilizing the synthetic data for the previous millennium and the upcoming century [69]. From 1800 to 2000, when implemented to perceived data, it depicts severe correlations by explaining 98% of the variance, by taking into consideration anthropogenic hydrologic contributions [69, 51]. According to the **IPCC (Intergovernmental Panel on Climate Change)**, for future temperature scenarios, between 75 cm to 190 cm from 1990 to 2100, the correlation predicts an acceleration in the sea-level [69, 51].

CHAPTER 3

Methodology

3.1 Data Collection

Primary: The first dataset named **"Annual Surface Temperature Change"** occurs to be an extensive gathering of data associated with variation in surface temperature throughout several countries.

The dataset consists of various columns, mentioned below:

- **'ObjectId':** A special determiner for every row.
- **'Country':** Countries or regions names.
- **'ISO2 & ISO3':** Country code in two letters and three letters, respectively.
- **'Indicator':** Represent the variety of data, which in this scenario, change in temperature with reference to a baseline period.
- **'Unit':** The measurement unit here is in Degree Celsius.
- **'Source':** Food and Agriculture Organization of the United Nations (**FAO**), which is the place of origin of the data.
- **'CTS_Code & CTS_Name':** Climate Tracking System (CTS), codes and names of it mentioned.

- **'CTS_Full_Descriptor':** A complete illustration of the CTS type to which the data corresponds.

- **'Yearly columns from '$F1961$' to '$F2022$':** These columns display the surface temperature change annually for each and every country from the years 1961 to 2022.

The dataset gives a factual point of view on how in individual countries the surface temperatures have altered over time, which is very essential for comprehending the effect of climate change.

	ObjectId	Country	ISO2	ISO3	Indicator	Unit	Source	CTS_Code	CTS_Name	CTS_Full_Descriptor	...	F2013	F2014	F2015	F2016	F2017	F2018	F2019	F2020	F2021	F2022
0	1	Afghanistan, Islamic Rep. of	AF	AFG	Temperature change with respect to a baseline ...	Degree Celsius	Food and Agriculture Organization of the Unite...	ECCS	Surface Temperature Change	Environment, Climate Change, Climate Indicator...	...	1.281	0.456	1.093	1.555	1.540	1.544	0.910	0.498	1.327	2.012
1	2	Albania	AL	ALB	Temperature change with respect to a baseline ...	Degree Celsius	Food and Agriculture Organization of the Unite...	ECCS	Surface Temperature Change	Environment, Climate Change, Climate Indicator...	...	1.333	1.198	1.569	1.464	1.121	2.028	1.675	1.498	1.536	1.518
2	3	Algeria	DZ	DZA	Temperature change with respect to a baseline ...	Degree Celsius	Food and Agriculture Organization of the Unite...	ECCS	Surface Temperature Change	Environment, Climate Change, Climate Indicator...	...	1.192	1.690	1.121	1.757	1.512	1.210	1.115	1.926	2.330	1.688
3	4	American Samoa	AS	ASM	Temperature change with respect to a baseline ...	Degree Celsius	Food and Agriculture Organization of the Unite...	ECCS	Surface Temperature Change	Environment, Climate Change, Climate Indicator...	...	1.257	1.170	1.009	1.539	1.435	1.189	1.539	1.430	1.268	1.256
4	5	Andorra, Principality of	AD	AND	Temperature change with respect to a baseline ...	Degree Celsius	Food and Agriculture Organization of the Unite...	ECCS	Surface Temperature Change	Environment, Climate Change, Climate Indicator...	...	0.831	1.946	1.690	1.990	1.925	1.919	1.964	2.562	1.533	3.243
5	6	Angola	AO	AGO	Temperature change with respect to a baseline ...	Degree Celsius	Food and Agriculture Organization of the Unite...	ECCS	Surface Temperature Change	Environment, Climate Change, Climate Indicator...	...	1.044	0.828	1.331	1.609	0.870	1.395	1.752	1.162	1.553	1.212
6	7	Anguilla	AI	AIA	Temperature change with respect to a baseline ...	Degree Celsius	Food and Agriculture Organization of the Unite...	ECCS	Surface Temperature Change	Environment, Climate Change, Climate Indicator...	...	0.770	0.814	1.051	1.125	0.960	0.664	0.843	1.224	0.893	0.839
7	8	Antigua and Barbuda	AG	ATG	Temperature change with respect to a baseline ...	Degree Celsius	Food and Agriculture Organization of the Unite...	ECCS	Surface Temperature Change	Environment, Climate Change, Climate Indicator...	...	0.783	0.744	1.035	1.097	0.958	0.627	0.797	1.131	0.862	0.770
8	9	Argentina	AR	ARG	Temperature change with respect to a baseline ...	Degree Celsius	Food and Agriculture Organization of the Unite...	ECCS	Surface Temperature Change	Environment, Climate Change, Climate Indicator...	...	0.442	0.951	0.957	0.488	1.095	0.878	0.760	1.123	1.031	0.643
9	10	Armenia, Rep. of	AM	ARM	Temperature change with respect to a baseline ...	Degree Celsius	Food and Agriculture Organization of the Unite...	ECCS	Surface Temperature Change	Environment, Climate Change, Climate Indicator...	...	1.407	1.283	1.931	1.356	0.889	2.772	1.859	1.954	2.087	1.707

10 rows × 72 columns

Figure 3.1: The image shows first 10 rows of the dataset 1.

The dataset is taken from this website: IMF Climate Data. It is an internet source that provides a span of climate associated datasets or information. In order to use for research, policy making and educational intentions, this website is appropriately created to give allowance to a variety of datasets that can be used. These datasets usually cover a broad spectrum of climate signals, such as variations in temperature, trends of precipitation and other climate associated variables that are crucial for comprehending the effects and patterns of climate change globally. For anyone fascinated in climate research and environmental inspection, this website possibly features contents for envisaging and examining the data, making it a

precious support.
Secondary: In the second dataset, there are historical temperature records in the **Global Climate Change Data project dataset**.

Here is a quick synopsis of its composition and contents:

- **"Date(dt)":** Begins in 1750 for the worldwide land and ocean temperatures and 1850 for the maximum and minimum temperatures of surface and global ocean and land temperatures.
- **"LandAverageTemperature":** The global surface temperature in average in Celsius for the given date.
- **"LandAverageTemperatureUncertainty":** The uncertainty related with the land average temperature, the approximately average land temperature of the 95% confidence interval.
- **"LandMaxTemperature":** The maximum global land temperature for the specified date (not accessible for beginning records), available in Celsius.
- **"LandMaxTemperatureUncertainty":** The uncertainty related with the maximum land temperature, the approximately maximum land temperature of the 95% confidence interval.
- **"LandMinTemperature":** It is available in Celsius, the minimum global land temperature for the mentioned date(also not available for beginning records).
- **"LandMinTemperatureUncertainty":** Related uncertainty with the minimum land temperature, the roughly minimum land temperature of the 95% confidence interval.
- **"LandAndOceanAverageTemperature":** The amalgamated global average temperature over land and ocean (not accessible for beginning records), available in Celsius.
- **"LandAndOceanAverageTemperatureUncertainty":** The uncertainty correlated with the merged land and ocean average temperature, the roughly global average land and ocean temperature of the 95% confidence interval.

	dt	LandAverageTemperature	LandAverageTemperatureUncertainty	LandMaxTemperature	LandMaxTemperatureUncertainty	LandMinTemperature	LandMinTemperatureUncertainty	LandAndOceanAverageTemperature
0	1750-01-01	3.034	3.574	NaN	NaN	NaN	NaN	NaN
1	1750-02-01	3.083	3.702	NaN	NaN	NaN	NaN	NaN
2	1750-03-01	5.626	3.076	NaN	NaN	NaN	NaN	NaN
3	1750-04-01	8.490	2.451	NaN	NaN	NaN	NaN	NaN
4	1750-05-01	11.573	2.072	NaN	NaN	NaN	NaN	NaN
5	1750-06-01	12.937	1.724	NaN	NaN	NaN	NaN	NaN
6	1750-07-01	15.868	1.911	NaN	NaN	NaN	NaN	NaN
7	1750-08-01	14.750	2.231	NaN	NaN	NaN	NaN	NaN
8	1750-09-01	11.413	2.637	NaN	NaN	NaN	NaN	NaN
9	1750-10-01	6.367	2.668	NaN	NaN	NaN	NaN	NaN

Figure 3.2: The image shows first 5 rows of the dataset 2.

This dataset is taken from the website: Global Climate Change Data. It starts from the year 1750, it seems to be a comprehensive data of global temperature variation over a long period of time. Both land and ocean temperature observation are embraced in it, albeit in the beginning data the ocean associated data appears to be unavailable. A perception into the authenticity and changeability of the observations are given by the involvement of uncertainty measures for every temperature reading. This dataset can be helpful in examining long lasting climate patterns and variations.

3.2 Preprocessing and cleaning

After loading the datasets, first we check if there is any null values or **NAN** values or missing values in the dataset.

For first dataset:

In the dataset first, we did some preprocessing steps to make the data for compatible for performing various tasks. In the first step, we checked that if there is any missing values, by using the command **"checkmissing = mydata1.isnull().sum()"** which gives the count of missing values of each column that tells the amount of **NAN** values in the columns. Below Table 3.1 shows the result of the step.

Then we also checked the percentage of **NAN** values for each column in the dataset, by using the command **"checkmissper = mydata1.isnull().mean() * 100"**. By just multiplying with 100 it gives the percentage of the values. The below Table 3.2 mentioned table shows the result of it.

Column	Number of Missing Values
ObjectId	0
Country	0
ISO2	2
ISO3	0
Indicator	0
::::	::
F2018	12
F2019	12
F2020	13
F2021	12
F2022	12

Table 3.1: Number of Missing Values in each Column

Column	Percentage of Missing Values (%)
ObjectId	0.000000
Country	0.000000
ISO2	0.888889
ISO3	0.000000
Indicator	0.000000
::::	::
F2018	5.333333
F2019	5.333333
F2020	5.777778
F2021	5.333333
F2022	5.333333

Table 3.2: Percentage (%) of Missing Values in each Column

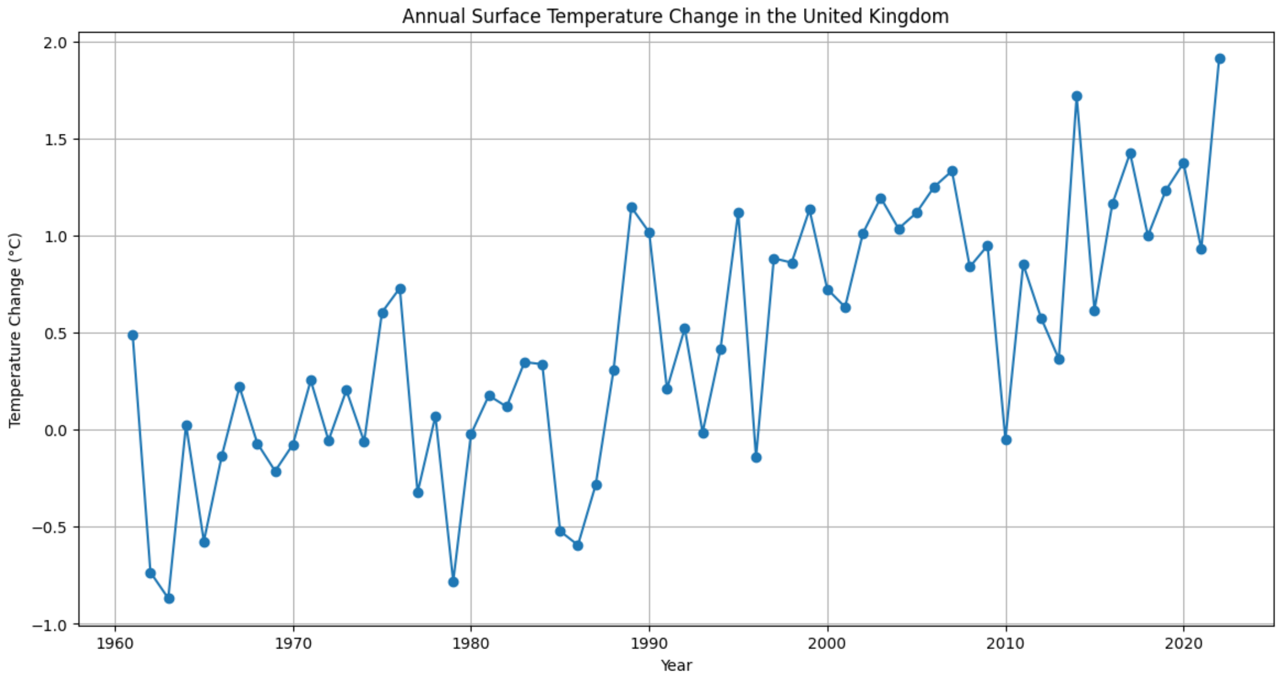

Figure 3.3: The image shows first 5 rows of the dataset 2.

As this dataset, shows the temperature values of different countries. It is showing country wise for the time period of 1961 to 2022. So, in order to perform ARIMA model or any other model, we needed to get the data of a particular country to perform tasks. So, we take out the UK data from the dataset and perform the model operation on this new dataframe. The table 3.3 and the figure 3.3 shows the result of it. The figure 3.3 shows the UK temperature change over the years from 1960 to 2022. As it is clearly visible that temperature goes upwards.

Year	UK Temperature_Change
1961	0.491
1962	-0.734
1963	-0.868
1964	0.025
1965	-0.575

Table 3.3: Annual Temperature Change

For Second dataset: In the second dataset, first we change **"Date (dt)"** column into timeseries format, by using the command:

```
data['dt'] = pd.to_datetime(data['dt'])
```

```
data.set_index('dt', inplace=True)
```

Then after that, we remove the data from 1750 to 1850 from the dataset, for better analysis. As from 1750 to 1850 out of 13 columns or features only 3 is filled with values. So in order to do better tasks, we needed to remove this part from the dataset. For doing this, we used this command:

```
data_post_1850 = data[data['year'] >= 1850]
data_post_1850.head()
```

Afterwards, we check the null values in the newly formed dataset. After performing the null values operation, it depicts that there is no **NAN** values in the new dataset. So, which means we can use it efficiently and easily. There will be no problem with missing gaps. Below tables 3.4 and 3.5 shows the newly formed dataset.

dt	LandAvgTemp	LandMaxTemp	LandMinTemp
1850-01-01	0.749	8.242	-3.206
1850-02-01	3.071	9.970	-2.291
1850-03-01	4.954	10.347	-1.905
1850-04-01	7.217	12.934	1.018
1850-05-01	10.004	15.655	3.811

Table 3.4: Temperature Data (Part 1)

dt	LandAvgTempUncertainty	LandMaxTempUncertainty	LandAndOceanAvgTemp
1850-01-01	1.105	1.738	12.833
1850-02-01	1.275	3.007	13.588
1850-03-01	0.955	2.401	14.043
1850-04-01	0.665	1.004	14.667
1850-05-01	0.617	2.406	15.507

Table 3.5: Temperature Data (Part 2)

3.3 Descriptive Analysis

Analysis of first dataset:

Statistic	Year	UK Temperature_Change
count	62	62
mean	1991.5	0.46
std	18.04	0.65
min	1961	-0.86
25%	1976.25	-0.04
50%	1991.5	0.45
75%	2006.75	1
max	2022	1.91

Table 3.6: Descriptive Statistics of UK Temperature Change

For doing the descriptive analysis or inspection, we calculated the various statistical things, by using the command **"mydata1.describe()"**. The summary statistics for a extracted dataset for a particular country United Kindom contains various statistical values. It is showing the information about all the statistical values of both columns **"Year"** and **"UK Temperature Change"**. A short explanation is given below:

There are total 62 counts in the table 3.6 result. The table shows the result of temperature in the UK from 1961 to 2022. The mean temperature in Celsius of the UK is 0.46. Temperature change of 0.65 is the standard deviation. The temperature at $25th$ percentile is -0.04 in degree Celisus. At $50th$ percentile is 0.45 in degre Celsius, in year 1991, which is middle of the data. And at $75th$ percentile is 1 in degree Celsius. The maximum temperature change is 1.91 which was in the year 2022.

For understanding the basic description of the dataset, these statistics are very useful. A sense of the distribution of the temperature change over the time is showing by the quartile values ($25\%, 50\%$ and 70%).

Analysis of second dataset:

The descriptive analysis in table 3.7 & table 3.8 shows the various statistical values of each and every columns in the dataset. It shows the count data of each column.

Statistic	LandAvgTemp	LandAvgTempUnc	LandMaxTemp	LandMaxTempUnc
count	1992	3180	1992	1992
mean	8.57	0.94	14.35	0.48
std	4.26	1.1	4.31	0.59
min	0.40	0.03	5.9	0.04
25%	4.43	0.18	10.21	0.14
50%	8.85	0.39	14.76	0.25
75%	12.85	1.42	18.45	0.54
max	15.49	7.88	21.32	4.37

Table 3.7: Descriptive Statistics of Dataset 2 (Part 1)

Statistic	LandMinTemp	LandMinTempUnc	LandOceanAvgTemp	LandOceanATU
count	1992	1992	1992	1992
mean	2.74	0.43	15.21	0.12
std	4.15	0.44	1.27	0.07
min	-5.40	0.045	12.47	0.04
25%	-1.33	0.15	14.04	0.06
50%	2.94	0.28	15.25	0.12
75%	6.78	0.45	16.39	0.15
max	9.71	3.49	17.61	0.45

Table 3.8: Descriptive Statistics of Dataset 2 (Part 2)

Mean value can also be seen from the above table, it means it shows the average value of that column.

The standard deviation shows that how far is the data is spread out from the actual values. Here, in this case, three columns have higher standard deviation (**LandAvgTemp, LandMaxTemp** and **LandMinTemp**) as compared to other columns. That suggests, that these columns have more spread data.

The quartiles are shows different meanings of the dataset. The 25% shows the first quartile or upper part of the dataset. The 50% percentile shows the middle or we can say median of the dataset. The 75% percentile shows the second quartile or lower part of the dataset. With the help of these, we can get idea that our data is skewed or not. Our data is right skewed if the mean is higher than the median. Here in this case, four columns have higher mean value than the median and four have smaller mean value than median. That means, four have right skewed and four have wrong skewed.

3.4 ARIMA Model

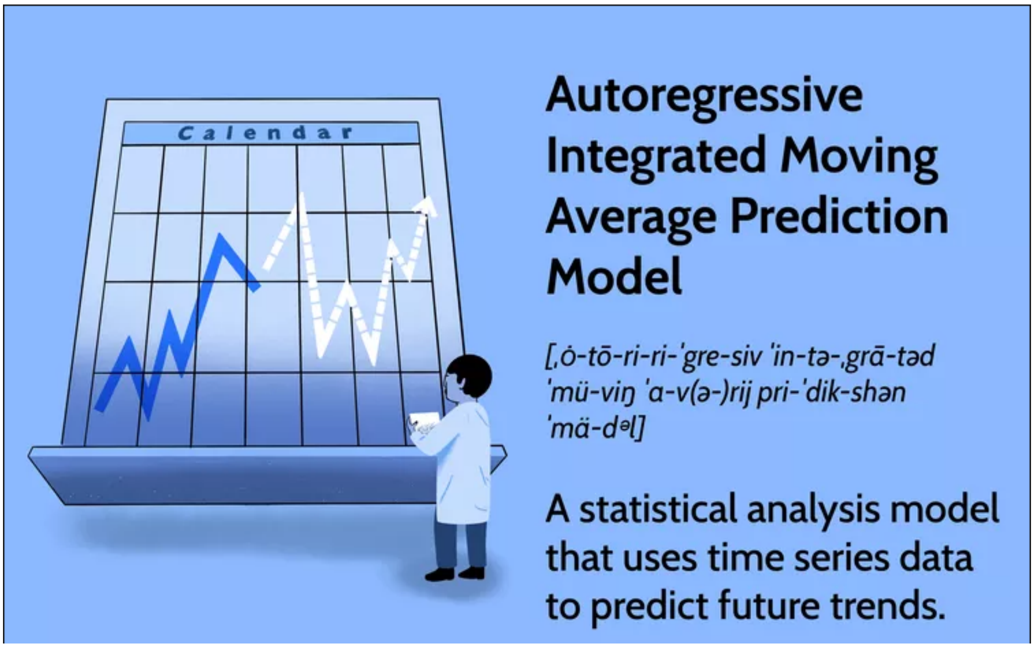

Figure 3.4: A general ARIMA image [71].

The Arima model is one of the well known models for time series forecasting [70]. It stands for **AutoRegressive Integrated Moving Average model** [70, 71, 72, 74, 77]. It is a mathematical model which is used for analyzing and forecasting time series data [71, 72, 74, 76, 77]. It forecast the future values or trends based on the past values or historical values or previous values or trends [71, 72, 77]. In other words, it uses its own lags to make the prediction [77]. It is a more enhanced or generalized version of **ARMA (AutoRegressive Moving Average)**, with the addition of "Integration" [72, 74].

The ARIMA model can be evaluated using the **Box-Jenkins** approach or method [72, 73, 74]. This model was invented by two statisticians, named **George Box** and **Gwilym Jenkins** in 1970 [73, 74]. This has the name after these two statisticians. It follows the iterative approach consists of 3 steps:

- Identification

- Estimation
- Diagnostic Checking

The **reason** we used ARIMA model, because it provides good result on a time series data. Forecasting with ARIMA is more appropriate. However, the results are depend on the type of datasets, sometimes the data is more complex because of that model's performance is not so good. Exceptions are always there.

Basic understanding of the model:

An AutoRegressive Integrated Moving Average model is actually a type of regression analysis, as it uses its **own lags** to work, from one variable relative to other changing variables it measures the strength [71, 72, 77].

We can get a better point of view after going through with its components:

- **AR (AutoRegression):** This part of the ARIMA model suggests that it uses the altering variable that reverts its own lagged or prior values [71, 72]. A vulnerable connection between a measurement and several lagged findings [74]. Only rely on its lag [77]. Equation is below [70, 77]:

  $$X_t = c + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \epsilon_1 \tag{3.4.1}$$

  where X_t = lag,

  X_{t-1} = first lag,

  α_1 = coefficient of X_{t-1},

  ϵ_t = white noise [70]

  Autoregressive model of p order.

- **Integration:** It refers to the differencing of the data. Means differencing of original observations so that data can be stationary [71]. In simple words, the data is changed with a difference of actual values to the previous values [71, 72, 74]. The main purpose of doing the differencing is to **make the data stationary**. It can be done more than once. Equation is below [72]:

  $$y'_t = y_t - y_{t-1} \tag{3.4.2}$$

Sometimes it must be necessary to do it twice based on situation, that can be written in mathematically as [72]:

$$y_t^* = y_t' - y_{t-1}' \tag{3.4.3}$$

$$= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \tag{3.4.4}$$

$$= y_t - 2y_{t-1} + y_{t-2} \tag{3.4.5}$$

- **MA (Moving Average):** A moving average model takes past or previous forecast errors in a regression like model, rather than using previous data of the forecast variable [70]. It uses the reliance between finding and residual mistakes from a moving average model used on lagged observations [71, 74]. Basically, this part represents that the regression error is simply a mixing of error variables whose outcomes happened simultaneously and at various points in the past [72]. Similarly, Y_t in a simple MA (moving average) model uses forecast errors that have lagged [77]. Equation is shown below [77]:

$$X_t = c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q} \tag{3.4.6}$$

where ε_t = error term,

ε_{t-1} = second error term,

θ_1 = moving average parameter.

The complete model equation would be like [72]:

$$X_t - \alpha_1 X_{t-1} - \ldots - \alpha_p X_{t-p} = \varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q} \tag{3.4.7}$$

or equivalent to

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \tag{3.4.8}$$

ARIMA Parameters:

The parameters of the **ARIMA** model are **p, d** and **q**. Every individual element in the **ARIMA** model has a standard notation and functions as a parameter [71], to indicates the kind of **ARIMA** model used, where the integer values substitute with the parameters [71, 74]

Parameters definition:

p: lag order, number of lag observations, **AR** term [71, 72, 74, 76, 77].

d: degree of differencing or number of times data being differenced [71, 72, 74, 76, 77].

q: order of moving average, moving average window size, **MA** term [71, 72, 74, 76, 77].

Stationarity Check:

For applying the **ARIMA** model, we need to check the stationarity or seasonality in the dataset. In order to do so, we need to do the differencing of the data if the data is not stationary [71, 77]. A stationarity model shows that there is a persistence or stability in the dataset [71].

There is one method or function by which we can check the stationarity of the dataset. The function is called **ADF (Augmented Dickey Fuller)** test. The Hypothesis for this method are:

- For **non-stationary** time series: H_0 (Null Hypothesis).
- For **stationary** dataset: H_1 (Alternate Hypothesis).

We must define a significance mark, which is usually set at **0.05** [76]. This can be checked with this test by comparing the value **p-value** from the result of the test. If the **p-value is higher** than the significance level, then we can say that we **fail to reject** the **null hypothesis** and if the **p-value is lower** than the significance value, then we can **reject** the **null hypothesis**, which confirms that our data is now stationary [76]. We just need to keep one thing in mind, do not try to do over differencing, over-differenced data can be stationary which affects the result [77].

Determining the order:

In order to do so, we need to find the order of regression (p) and the order of moving average (q) with the help of **ACF (Autocorrelation Function)** and **PACF (Partial Autocorrelation Function)** [71, 72, 73, 76].

- **ACF:** Have direct and indirect effects [76]. The plot shows the relation of lag values of the observations. On the x-axis it shows the lags and on the y-axis it shows the associated coefficient negative and positive correlation between -1 and 1 respectively [74, 75].

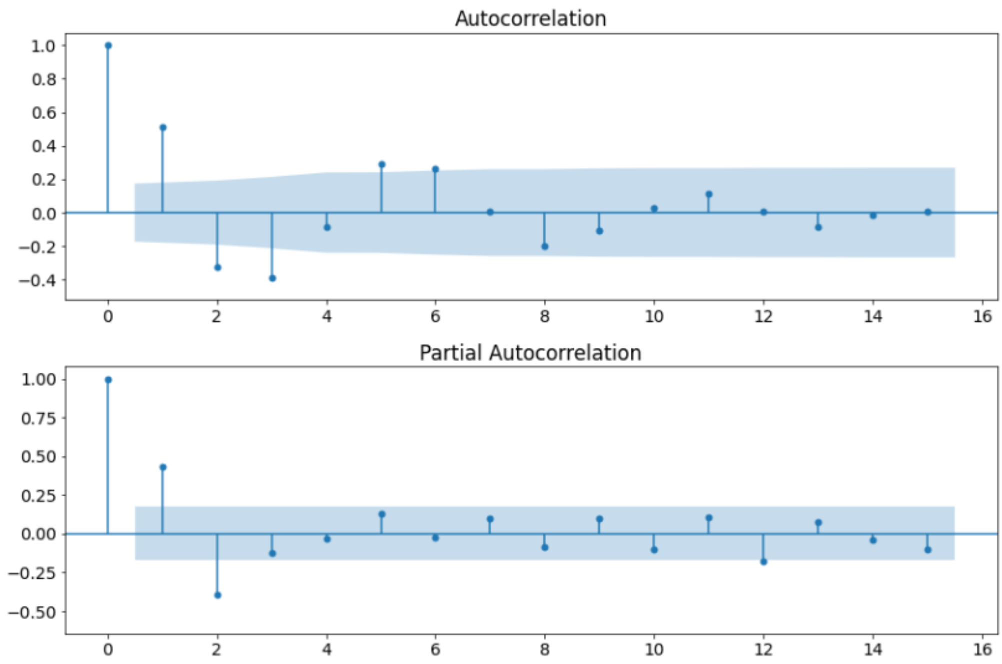

Figure 3.5: ACF and PACF plot samples [76].

- **PACF:** This plot also shows the relation of lag values of observations, but which are not explained by previous lagged values [74, 75]. It only depicts the direct relation [76].

To visualize the result we use **lollipop graphs**. These lollipops or spikes, on the x-axis, show the relation of each lag with the original values [76]. We just need to focus on the spikes which are outside the **blue-shaded** area.

- For **p value**, we use **PACF** plot.
- For **q value**, we use **ACF** plot.

An alternative method to find the order is **AIC (Akaike Information Criterion)** and **BIC (Bayesian Information Criterion)** [70, 72], which can be shown mathematically as [70, 72]:

$$AIC = -2\log(L) + 2(p + q + k) \tag{3.4.9}$$

where, L = lag of data,

p = AR order part,

q = MA order part,

If $k = 1$, then the **ARIMA** model has an intercept $(c = 0)$ and if $k = 0$, then no intercept in the model $(c = 0)$, [70, 72].

The **ARIMA** model with rectified **AIC** written as [70, 72]:

$$AIC_c = AIC + \frac{2(p+q+k)(p+q+k+1)}{T-p-q-k-1} \tag{3.4.10}$$

The **Bayesian Information Criterion** shown below [70, 72]:

$$BIC = AIC + (\log(T) - 2)(p+q+k) \tag{3.4.11}$$

The motive of this is to reduce the values of these terms for a good model [70, 72].

Pros and Cos of ARIMA model [72]:

Advantages:

- Suitable for short-term forecasting.
- Only required historical or previous data.
- Can model non-stationary data.

Disadvantage:

- Not very good at spotting turning points.
- Computational cost is high.
- Not suitable for everlasting forecasting.
- Arbitrary parameters.

3.5 Data Visualizations

3.5.1 Box-plot of Temperature Change by Decade in the United Kingdom

The figure 3.6 is indicating the United Kingdom's temperature in decades. In general, the box plot shows the spread of a set of data. It indicates the median (the middle value), the $25th$ and $75th$ percentiles (means division in 4 equal parts of data) and the outliers, which means values outside the group.

This box plot shows the United Kingdom's median temperature change has been rising over time. As compared to previous decades, the $2010s$ saw the most median temperature

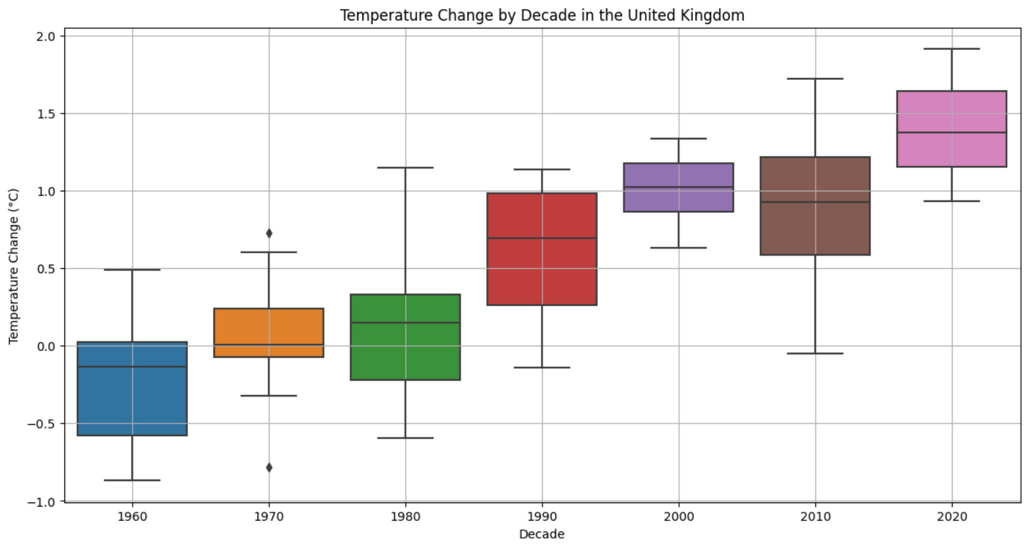

Figure 3.6: UK temperature change in decades.

change which was 1.5 degrees Celsius, higher than any other decade. The 1.0 and 2.0 are the $25th$ and $75th$ percentiles respectively for $2010s$. It means the variation range for the $2010s$ is from 1.0 to 2.0 degrees Celsius. It also appears that a lot of variation in temperature change has been happening from decade to decade.

It also shows some outliers in the data, like in the $1970s$ it can be visible. As it can be clearly understood that temperature was lower in initial decades as compared to last decades.

3.5.2 Seasonal Analysis: Average Temperature Changes Across Months

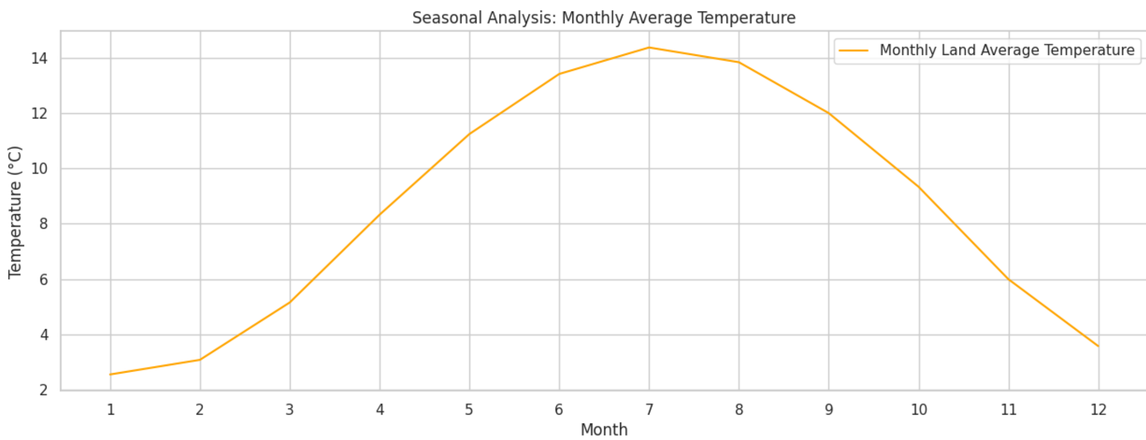

Figure 3.7: Monthly temperature change.

The figure 3.7 showing the monthly average temperature change. The x-axis or horizontal axis is showing months and showing months from $1st$ to $12th$ means from January to December. The y-axis shows the temperature in degrees Celsius.

The 3.7 has only a single orange line which highlights the "Monthly Land Average Temperature", which for every month plots the average land temperature. It shows months from $1st$ to $12th$, meaning from January to December. The line states that in the beginning months, the temperature was lower than slowly increasing to its peak in summer months (warmer months) and then going down in December for the winter season. Temperate regions with four distinct seasons are reflected by a typical temperature curve.

In the beginning of the year and the ending of the year, the temperature was lowest and highest in the middle.

3.5.3 Comparative Analysis: Land vs Land and Ocean Temperature Trends

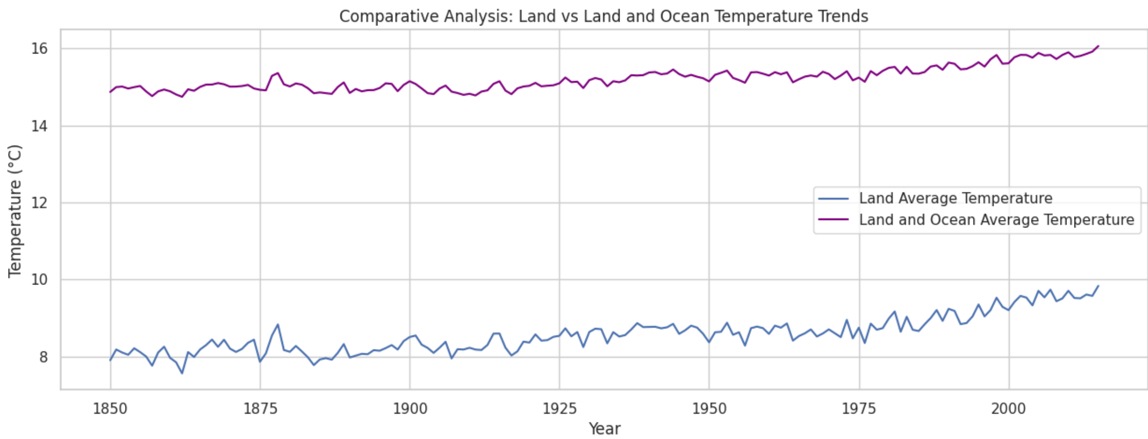

Figure 3.8: Land vs Land and Ocean Temperature Trends.

A figure 3.8 shows the graph of comparative analysis between land and land and ocean temperature trends. It is plotting two sets of temperature data.

- **Land Average Temperature:** It depicts the average land temperature only. It is depicted in a **blue** color. It depicts changeability with some fluctuations over the years, but it appears to rise in later years.
- **Land and Ocean Average Temperature:** It shows the combination of both land and ocean temperature together. It is depicted in **magenta** color. It represent the upward trend over time and also it is more stable than the land only temperature.

The x-axis or horizontal axis depicts years from $1850s$ to nearly the early $2000s$. The y-axis depicts the temperature in degrees Celsius. Both the land and land ocean temperature have risen over a period of time. The variable in land average temperature has more than land ocean. The land ocean temperature is more steady and less vulnerable to precise fluctuations while showing a rise.

3.5.4 Correlation Analysis: Heatmap

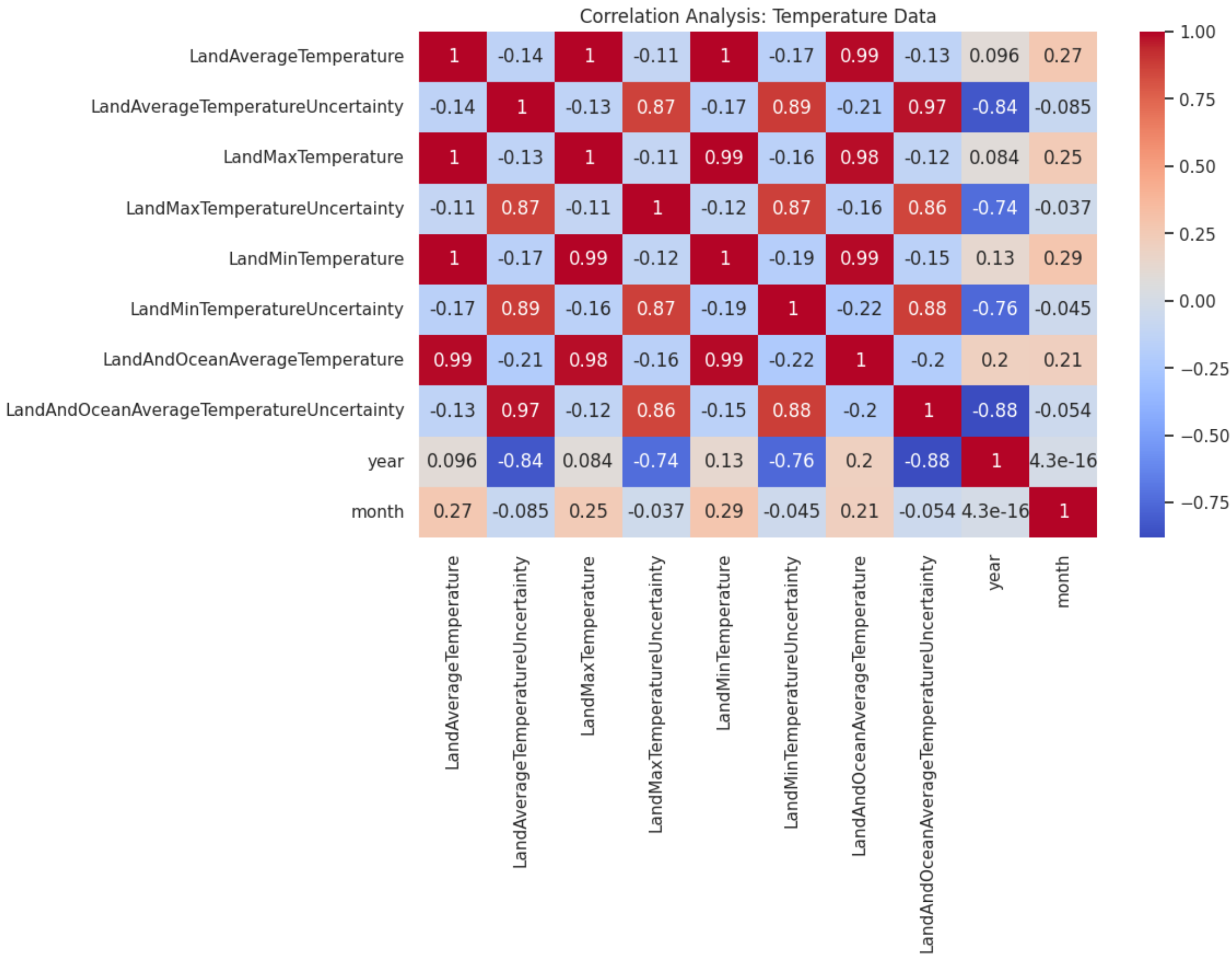

Figure 3.9: Correlation Analysis

The figure 3.9 heat map shows correlation between several variables associated with temperature data. A correlation coefficient between variables is depicted by the correlation matrix or heatmap. The correlation between two variables is depicted by each and every cell of the map. From -1 to 1 is the range of it. When two variables exhibit a strong positive correlation, near to 1, it means that they often increase or decrease together. One variable rises as the other falls if the connection is strong and near to -1. No relation between the variables if it's close to 0.

The correlation depicted by color scale on the right, a perfect positive correlation is at 1 which is red, no correlation at 0 (white) and a proper negative correlation at -1 (blue). A very high positive correlation is shown by the variables like **LandAverageTemperture** and **LandMaxTemperature**, besides **LandMinTemperature** and **LandAndOceanAverageTemperature**. A heatmap is very useful for fastly recognizing how strongly different variables are associated with each other.

3.5.5 Yearly Temperature Difference (Max - Min):

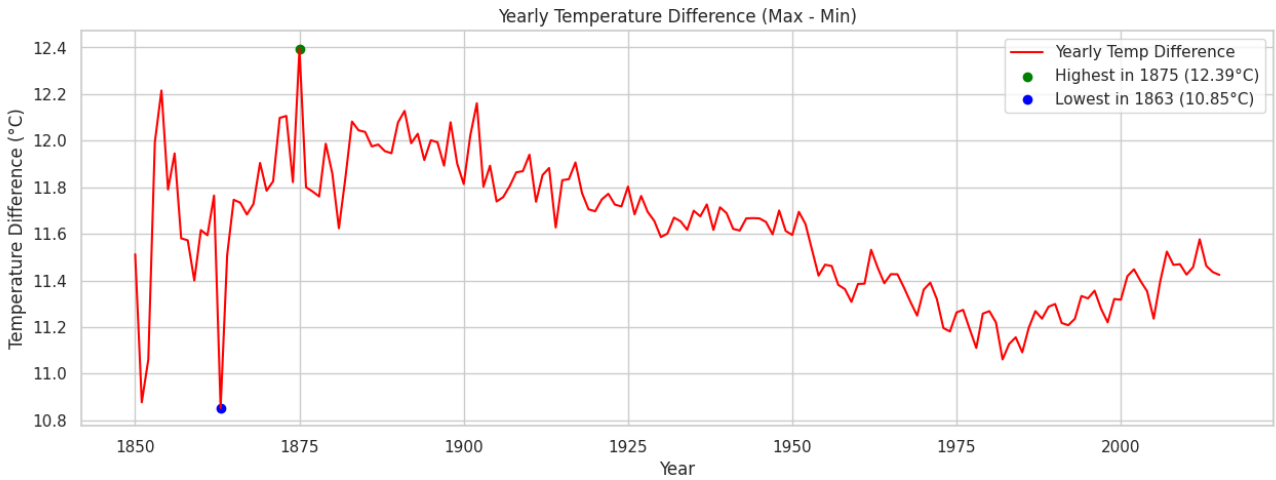

Figure 3.10: Yearly Temperature Difference (Max - Min)

The figure 3.10 the difference between the maximum and minimum temperatures observed every year over a period of time. The x-axis depicts the year from 1850 to the $2000s$, and y-axis depicts the temperature difference in degrees Celsius.

The red line depicts the range of yearly temperatures for every year, which is calculated by subtracting the maximum and minimum observed temperatures for that year.

There are two precise data points featured on the graph.

- **Green Dot:** It depicts that the year 1875 had the highest range at 12.39 degrees Celsius.
- **Blue Dot:** It depicts the smallest temperature range at 10.85 degrees Celsius in the year 1863.

With few years exhibiting a larger spread between their highest and lowest temperature than others, the line graph depicts variation in the yearly temperature range. A little declining pattern overall, indicating that the difference between the highest and lowest temperatures in a year might be reducing over time.

CHAPTER 4

Results

In this section, we discussed the research questions which are mentioned in the **"Introduction"** chapter. Here, we are explaining the answer of those questions with some visualizations.

4.1 *Answer 1*: Global average temperature change from 1961-2022.

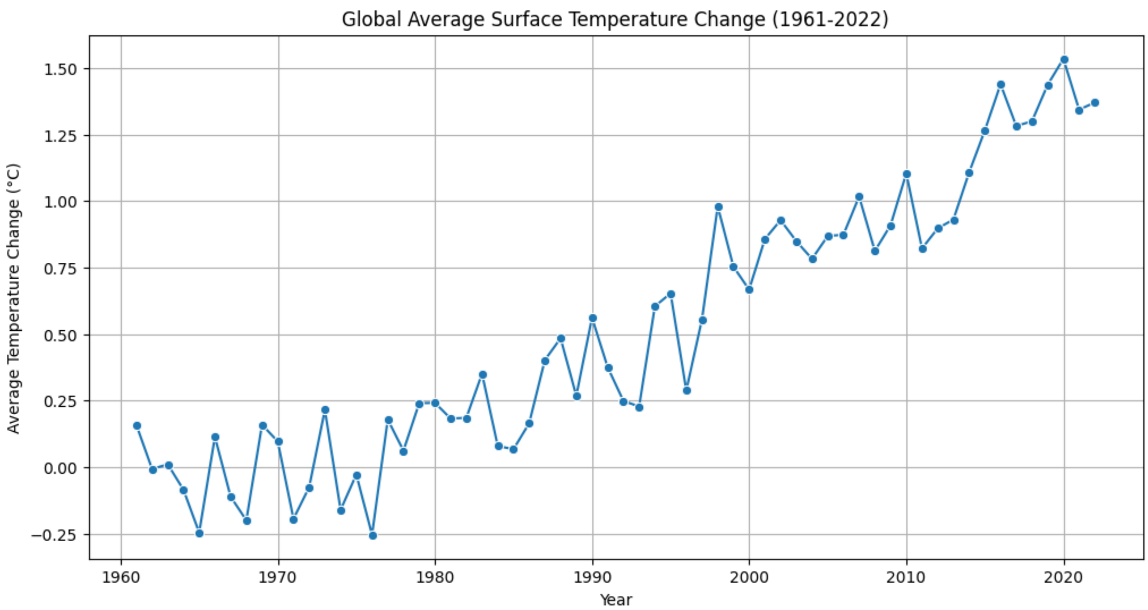

Figure 4.1: Temperature change from 1961-2022.

The figure 4.1 depicts a line graph of the global average surface temperature change from 1961-2022. The x-axis or horizontal axis indicates the years from 1961 to 2022. Here are the key points:

- The x-axis or horizontal axis indicates the years from 1961 to 2022.
- The y-axis or vertical axis shows the average temperature variations in degrees Celsius.
- For each year, the average temperature change is indicated by the line markers on the graph.
- The variations in the line show that there is some variety from year to year.
- As per the line trend, over this period we can see the global surface temperature has been rising up.
- It suggests that the recent years have been more hotter than the start of dataset in 1961, hotter years are on the graph appear to be towards to the end.

A particular concern of some scientists and environmentalists is the rise in the temperature strongly signalling global warming due to its severe impacts on ecosystems, sea levels and others.

4.2 *Answer 2*: Surface temperature throughout the decades.

The figure 4.2 depicts a bar graph of average surface temperature with respect to decades, ranging from the $1960s$ to $2020s$. The y-axis depicts the average temperature change and the x-axis depicts the decades from 1960-2022.

Each and every bar depicts the decade and the average temperature change is depicted by the height of the bar of that decade. The colour code bars are in the plot, for visual differentiation without any extra meaning to the colour itself. The height of the bars depicts an upward pattern that says each decade has faced a higher average temperature change compared to the previous. The $2020s$ has the biggest bar, that says the highest temperature change has been recorded on the year. By decades the visualization eases the data, making it easier to see the abiding trend without the year to year variations that might look in yearly data.

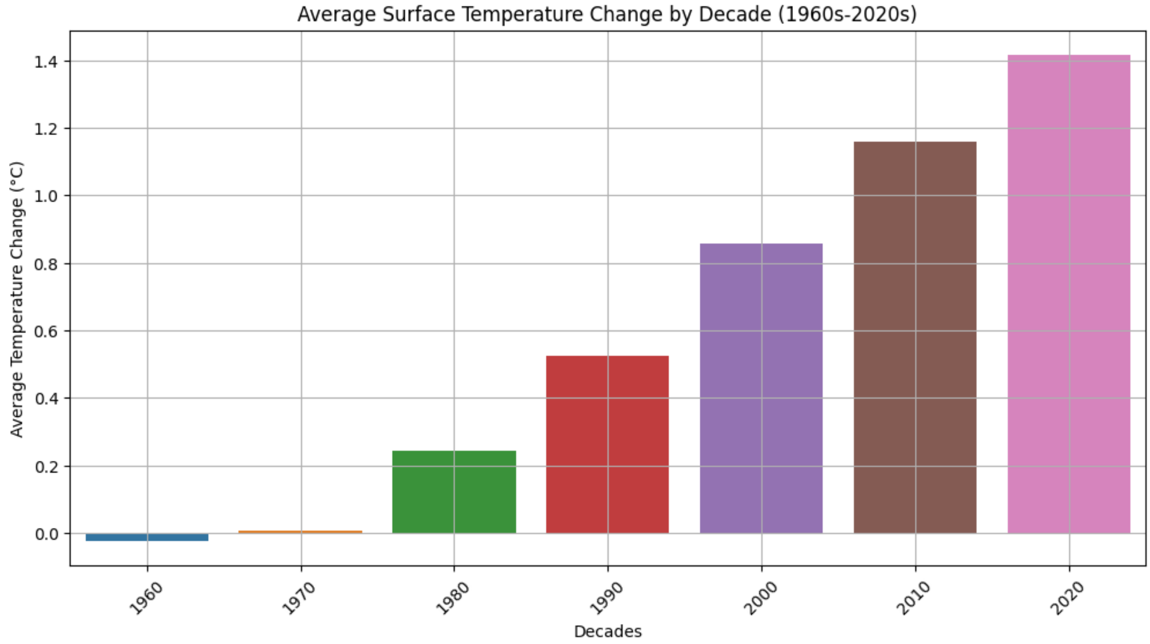

Figure 4.2: Surface temperature throughout the decades.

4.3 *Answer 3*: Patterns or anomalies in the temperature changes over the years.

The figure 4.3 depicts the trace of global average surface temperature change from 1961 to 2022. The blue line depicts the average temperature change for every year. The lower outlier threshold showed by the red dashed line and the upper outlier threshold showed by the green dashed line. The y-axis depicts the average temperature change in degree Celsius. On the x-axis depicts the years.

The graph depicts the general rising pattern over the time period even after the temperature changes from year to year. After the late $1970s$, the average temperature change was constantly above 0 degree Celsius. The fact that the data points stay inside the upper and lower outlier criterion, that as per the defined threshold, the yearly values do not belong in any outlier category. Even though no data points above the upper outlier threshold, it might show an important variation from the average that would be very high. Same with lower outliers threshold shows a negative variation from the average, but no data points below this threshold.

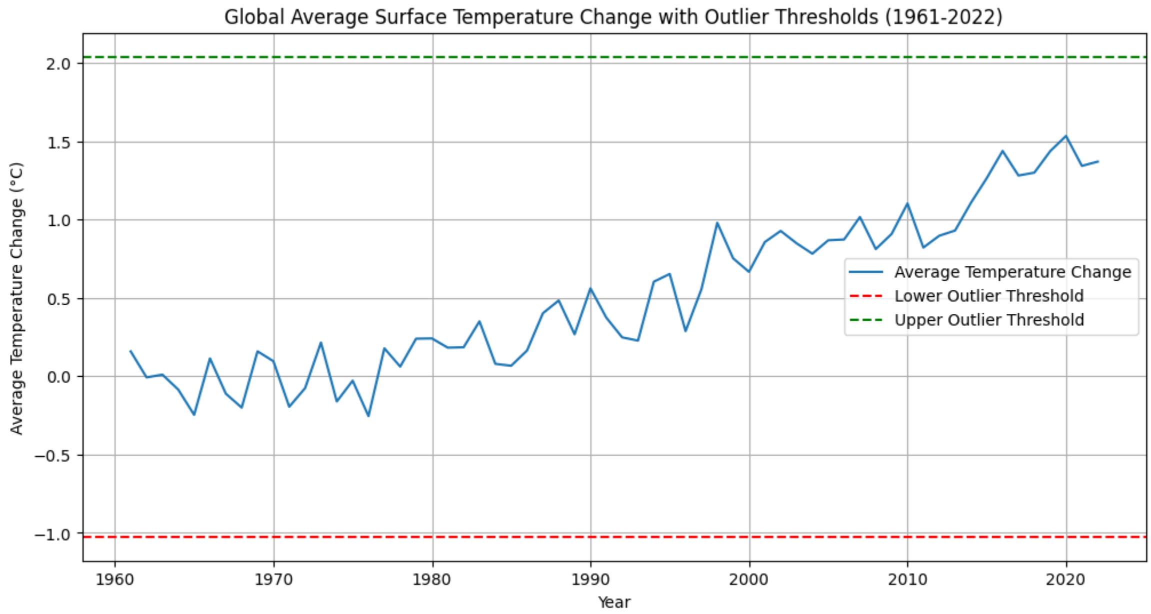

Figure 4.3: Patterns or anomalies in the temperature changes.

4.4 *Answer 4*: Trends difference between developed and developing nations.

The two graphs show the surface temperature changes across different countries, classified as **developed** and **developing countries**.

Developed countries graph:

This figure 4.4 surface temperature variation from 1960 to 2022 for the developed countries like the United States, Germany, Japan, Australia and Canada. Representation is done by different coloured lines of each country. On x-axis years are plotted and on y-axis change in temperature in degrees Celsius is plotted. Each and every country faced variations in temperature changes over time, with a single pattern of change matches with each other.There are obvious times when the temperature changes more and less, showing variations year to year. It seems like there are few sharp increases and decreases at several points, especially in the data for Canada and Australia. In general, an upward trend might appear in temperature changes for these countries, proposing a rise in average temperatures over the decades.

Developing countries graph:

The figure 4.5 shows the surface temperature changes over the exact time span but this time it is taking developing countries- India, Brazil, Nigeria, Vietnam and Indonesia. Just

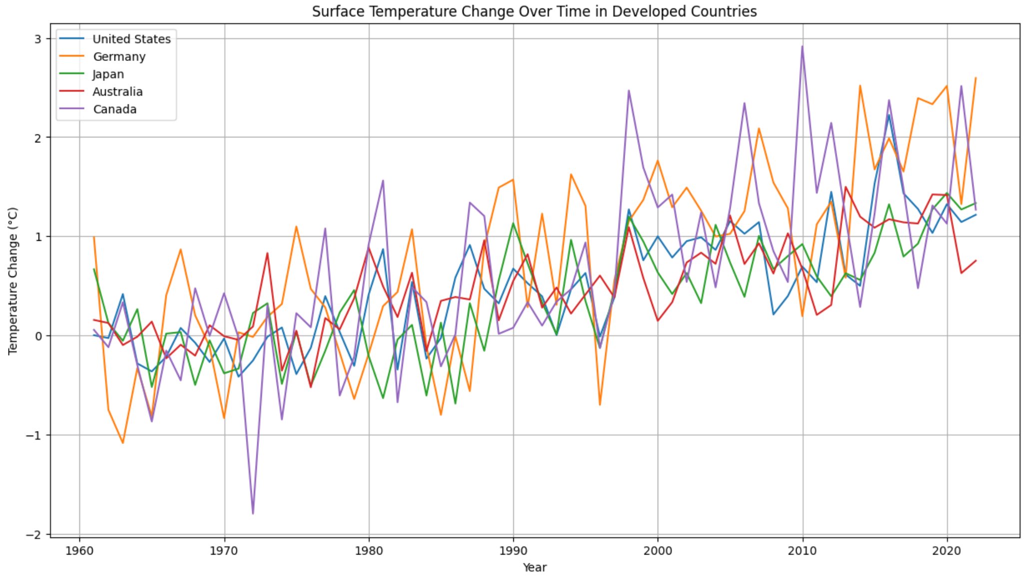

Figure 4.4: Surface temperature change in developed countries.

like the first graph, a different colour line is used for showing each country. It also uses the same temperature measurement over the years. It too display yearly variations with a general upward trend seen in lines of countries. As compared to developed countries' graph, fluctuations between the countries are more pronounced, which might in return different climate patterns in data collection.

Both show that surface temperature changes are not uniform, both show a general trend of increasing temperature over a time period. The developed countries show less extreme yearly variation as compared to developing countries. In these graphs, changes in temperature are influenced by the merging of climate trends and regional factors, for instance, deforestation, urbanization.

4.5 *Answer 5*: Trend Analysis: Temperature Over Time.

The figure 4.6 depicts a trend analysis to temperature over a period of time, it is mainly focusing on land temperatures. The x-axis depicts years from 1850 to 2015 and the y-axis depicts the temperature in degrees Celsius. The three lines presenting different temperatures measurement:

- **Land Average Temperature:** The blue line depicts the land average temperature over the years. Throughout the period it appears like it is comparatively steady with small

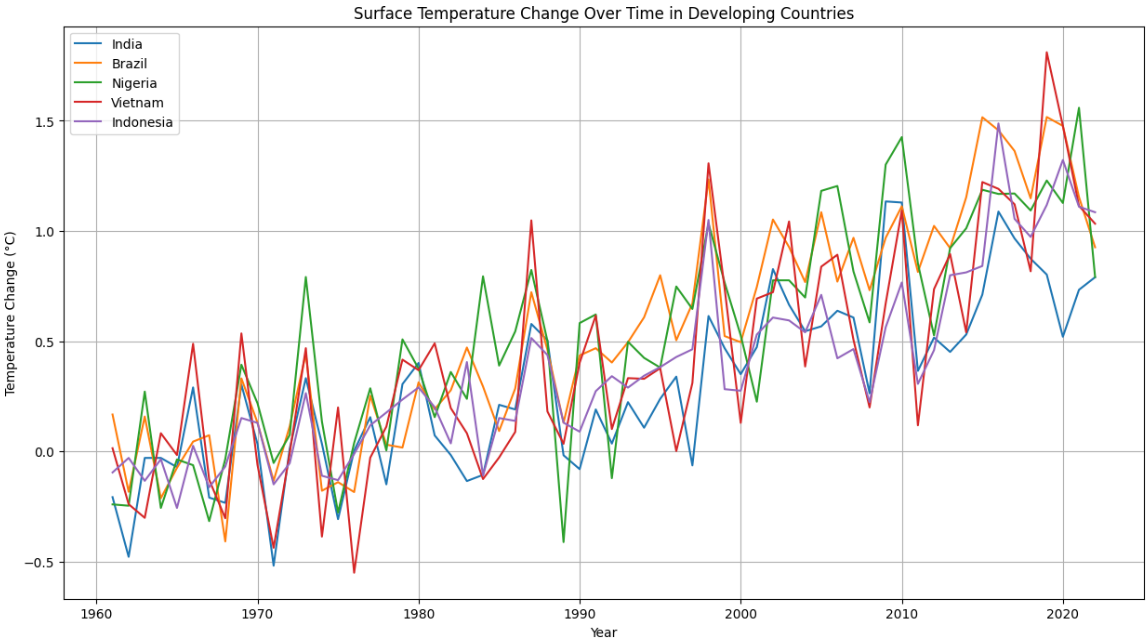

Figure 4.5: Surface temperature change in developing countries.

variations.

- **Land Max Temperature:** It is shown by the orange line on top. It too depicts the maximum land temperature recorded over the years. As compared to average and minimum land temperatures, it is persistently higher as expected. Without noteworthy upward or downward trends, it appears to be quite stable.

- **Land Min Temperature:** The bottom green line on the graph depicts the minimum temperature. In contrast with others, it is lower among all of them and depicts the stability over the time.

The graph depicts the record over 150 years. All of them depict stability in some years, a bit similar to each other but not entirely. This type of record is useful for climate studies, allowing experts to observe long term trends in temperature.

4.6 *Answer 6*: Uncertainties of temperature over time.

The figure 4.7 depicts the uncertainties or margins of error of temperature from 1850 to past 2000. The green line depicts the uncertainties of recorded temperatures, which is degrees Celsius. The x-axis shows the years and the y-axis shows the temperature uncertainties.

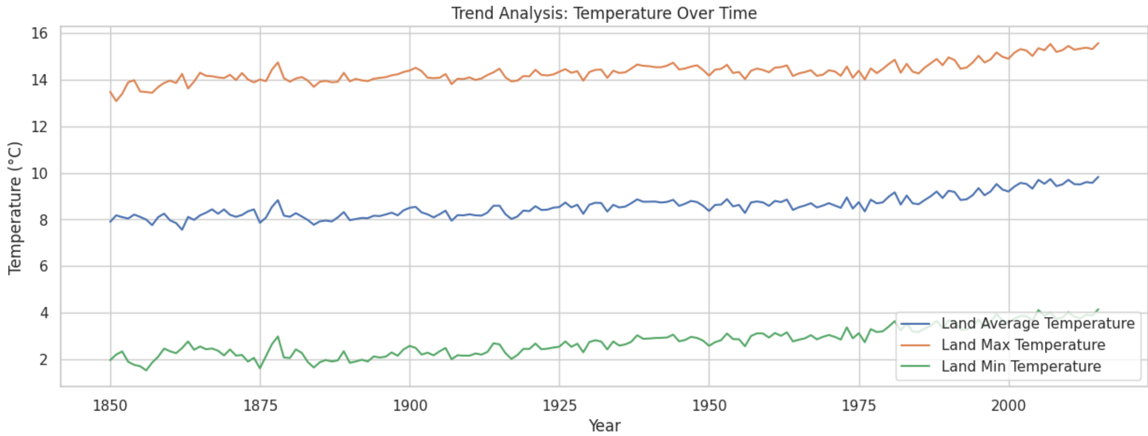

Figure 4.6: Land temperature over the years.

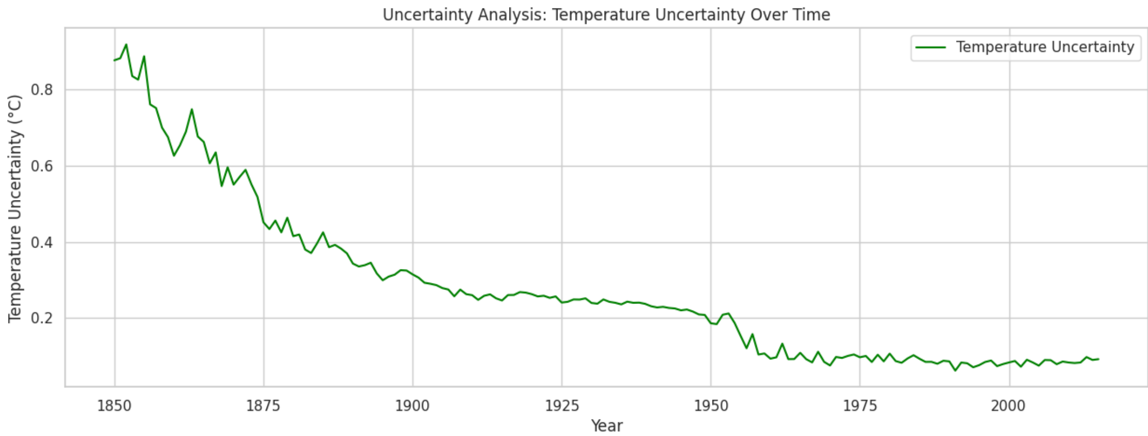

Figure 4.7: Uncertainties of temperature over the years.

- **Decreasing Uncertainty:** The line depicts the decline in temperature uncertainty. Starting with higher in the 1850 and then slowly reducing towards the year 2000.
- **Initial changes in record:** In the graph, we can see there is detectable fluctuation at the beginning of with few spikes. This might be possible because in the initial years we did not have many reliable measurement methods or techniques, with very few data sources.
- **Uncertainty stability:** In the later years, mainly after 1925, there seems to be a decrease in uncertainty; it appears to stabilize with fewer fluctuations. This stabilization and deduction in uncertainty might be possible because of improvements or new techniques of measurement and accessible to resources.

This study could be helpful for climate studies and environmental sciences. Decreasing uncertainties give confidence in temperature patterns used for climate studies and comprehension of long-term climate change impacts.

4.7 *Answer 7*: Temperature changes in decades.

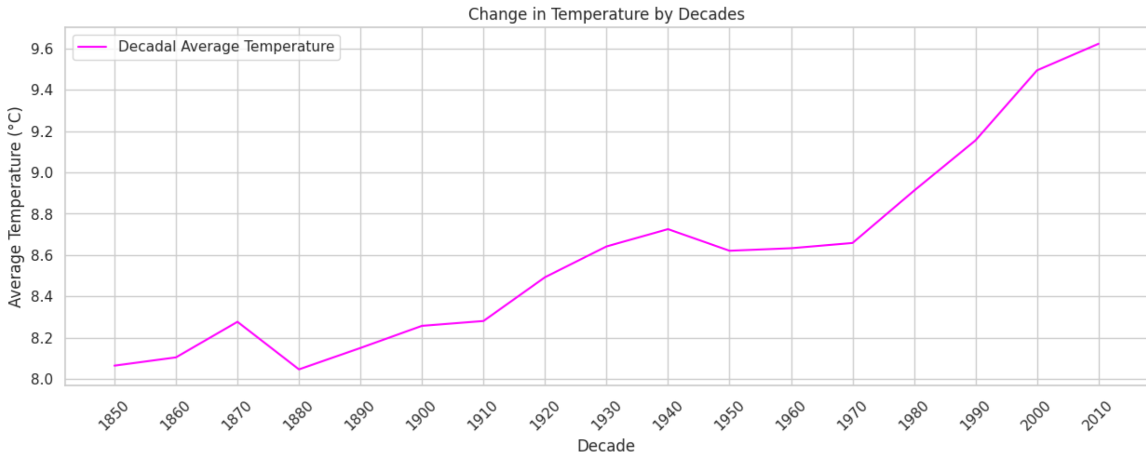

Figure 4.8: Average temperatur change in decades.

The figure 4.8 depict a change of temperature in decades, starting from the $1850s$ to the $2010s$. The graph is showing the mean of each decade, x-axis depicts the decades and the y-axis depicts the temperature in degrees Celsius. The magenta line depicts the values of temperature for each decade.

The line depicts that from the $1850s$ to $1910s$, with few fluctuations, the record was relatively flat. After the $1910s$, there was a hike in average temperature with continuous decades. The growth picked up very rapid momentum in the $1980s$ and peaked in the $2010s$, the last decade on the graph.

The graph gives an explanation of long-term climate change, especially the temperature rise in recent decades, which could be proof of growing global warming. It's crucial to remember that these are decadal averages, which eliminate any fluctuation from year to year and emphasize the longer-term trend of rising temperatures.

4.8 *Answer 8*: Global average land temperature by season.

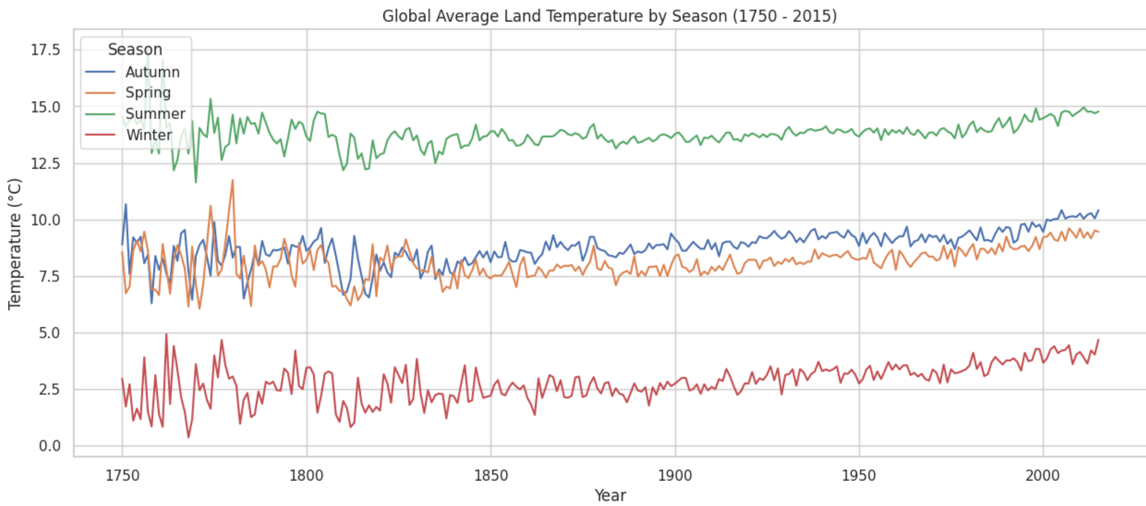

Figure 4.9: Average land temperature change in season.

The figure 4.9 depicts the temperature in four different seasons from 1750 to 2015, in four different lines, each line depicts the different seasons. The Green line means summer season, blue and orange means autumn and spring respectively and the red line depicts the winter season. Each depicts the temperature trend over the time, on the x-axis shows the years and on the y-axis shows the temperature in degrees Celsius.

Below is the brief summary:

- **Red line or winter season:** It begins at the minimum temperature range and depicts the extreme fluctuation over the time, with the later portion of the graph appearing to be increasing.
- **Blue and orange line or autumn and spring season:** These begin in the middle temperature range and follow similar trends, with spring slightly higher just after the beginning, then both follow similar trends.
- **Green line or summer season:** Highest temperature range among all of them, slowly increasing with the time.

Overall, the graph depicts that there is an upward trend followed by the temperature across all the seasons. The seasonal differences are also precise with summer being the hottest and winter the coolest. Basically, the graph depicts long-term climate change in different seasons.

4.9 Answer 9: Temperature changes in different continents.

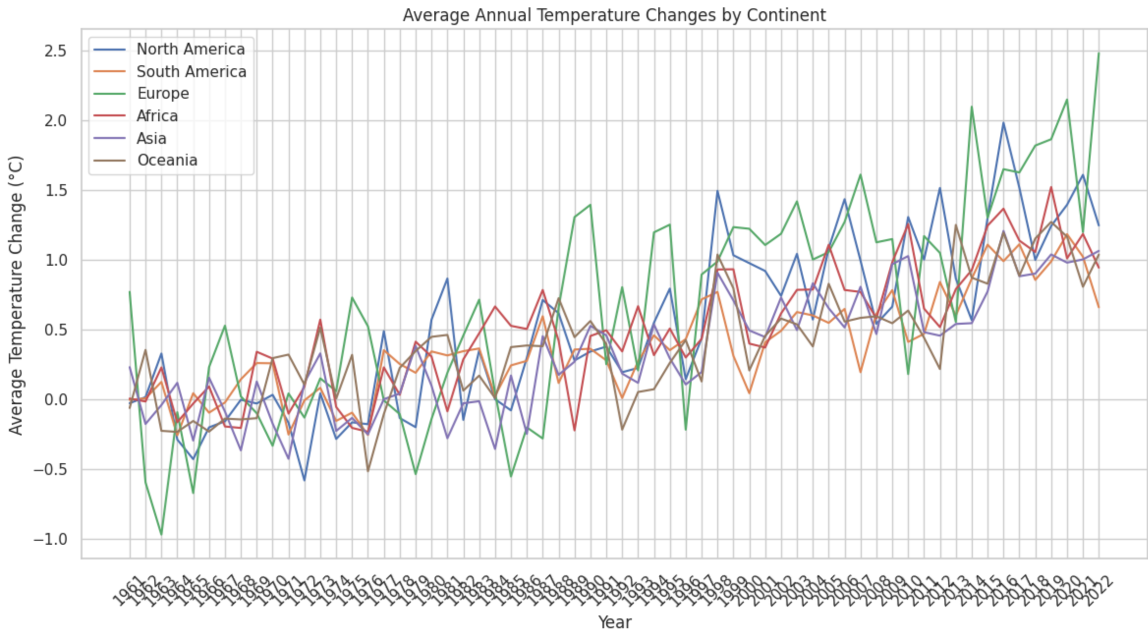

Figure 4.10: Average temperature change in different continents.

The figure 4.10 depicts the variations in temperature in degrees Celsius in different continents over the years. On the x-axis, it depicts the years from 1961 to 2022 and on the y-axis , it depicts the temperature values ranging from -1.0 degree Celsius to 2.5 degree Celsius. Each line indicates a different continent, with different colours, as shown on the top right corner in the figure.

The graph depicts that for all the continents the temperature variations are going upward, all following the upward trend. The pace of temperature change is not uniform in all continents, as the fluctuations records suggest. A rapid increase is shown by some continents than others. For instance, the green and blue line means Europe and Asia respectively, and depicts a sharp increase at the end. During the last few years, it seems that South America's line shows a downward trend. The most significant increases seem to be in Europe and North America, facing the highest peak in the end.

This data can be useful for climate change analysis, depiction how the world is experiencing temperature change in different regions.

4.10 *Answer 10*: Findings of ARIMA model on both datasets.

On dataset A:

First, after loading of dataset and preprocessing steps. We checked the stationarity of the dataset to implement further step of ARIMA model. The figure 4.11 is a time series decompo-

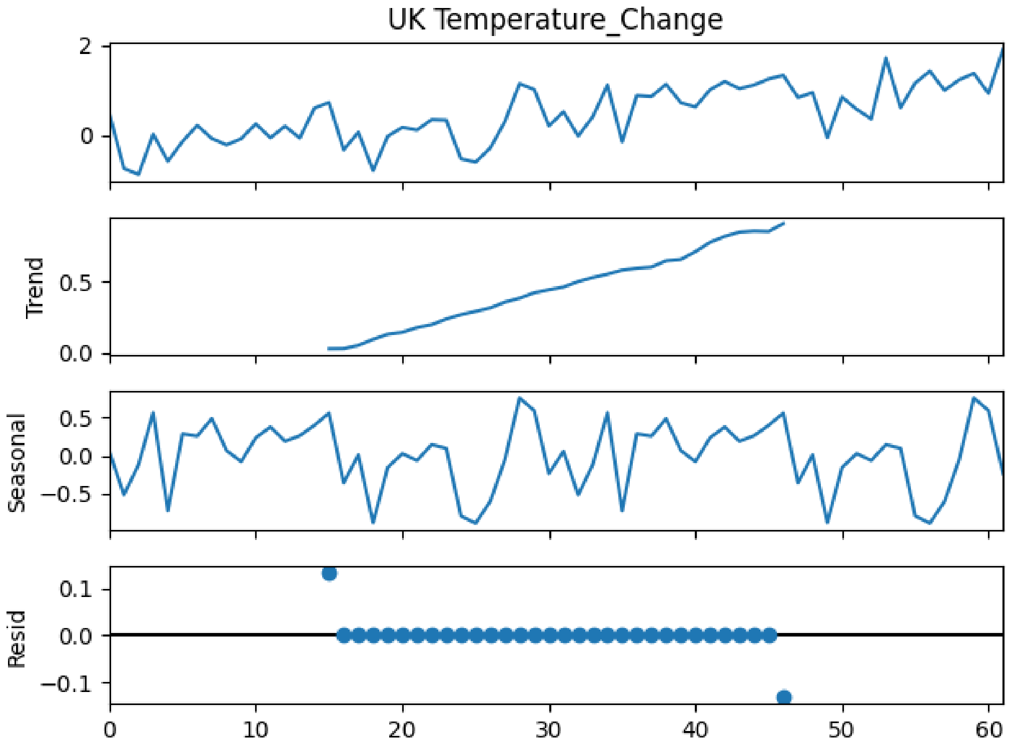

Figure 4.11: Seasonality showing.

sition plot. It analysis time series into several components:

- **Top plot:** It depicts the temperature change of the UK over the years.
- **Trend:** It depicts the long-term progression of the series, tells about the increasing or decreasing pattern of the series. Here in this case, it is going upward, which means temperature is increasing.
- **Seasonal:** It captures the seasonal pattern or trend of the data at regular intervals in a monthly, quarterly or yearly. Here in this, it depicts there is a fluctuation which seems to repeat itself, which means indicative of seasonal behaviour.

- **Residual:** It displays the irregularities or noise which are left over after trend and seasonal components have been removed. In this case, residuals are very less and lying around zero, which means trend and seasonal have covered almost all of it.

After the seasonality check **Figure 4.10**, we need to check the stationarity for that we have to do the differencing.

Stationarity check:

In order to do stationarity check, we have to do **ADF (Augmented Dickey-Fuller)test** table 4.1.

Statistic	Value
Test Statistic	-0.58
p-value	0.87
# Lags Used	9.00
Number of Observations Used	52.00
Critical Value (1%)	-3.56
Critical Value (5%)	-2.91
Critical Value (10%)	-2.59

Table 4.1: ADF Test Results

Test Statistic: The value of the test statistic is -0.586290, it is compared with all critical values. As in this case, the value is higher than all the critical values of -3.562879 at 1%, -2.918973 at 5% and -2.597393 at 10%. As the test statistic value is higher than all critical values, we can say that we fail to reject the null hypothesis.

P-value: The p-value is 0.874050, which is **higher than the significance level** of 0.05. So, we have **failed to reject the null hypothesis**. If the p-value is higher than the significance level, that means there is **strong evidence** that the null hypothesis is true. It shows that the series has a unit root and is **non-stationary**.

Lags used: While doing the **ADF** test, 9 lags were used. It affects the power of the test.

Number of Observations Used: It shows 52 observations are used in the test, which accounts for the lag terms.

Both outputs are in favour of failing to reject the null hypothesis. So, we need to do differencing to make the data stationary and have to check again by doing this test again.

After that, we can proceed further to implement the **ARIMA** model.

Differencing:

After differencing, we again did the ADF test (table 4.2) to check the stationarity. Here are the results:

Metric	**Value**
Test Statistic	-2.91
p-value	0.04
# Lags Used	8.00
Number of Observations Used	52.00
Critical Value (1%)	-3.56
Critical Value (5%)	-2.91
Critical Value (10%)	-2.59

Table 4.2: ADF Test Results after differencing.

Test Statistic: The value of the test statistic is -2.91 after the differencing, it is compared with all critical values. Now, the value is higher(or more negative) than the critical values of 1% and 5% (-3.56 and -2.91 respectively) but lower than at 10% which is -2.59. So, we can fail to reject the null hypothesis at 1% and 5% but can reject at 10%. However, the p-value is lower than significance(0.05), we have evidence to reject the hypothesis at 5%.

P-value: The p-value is 0.04 this time, which is **lower than the significance level** of 0.05. So, we can **reject the null hypothesis**.

Lags used: While doing the **ADF** test this time, 8 lags were used.

Number of Observations Used: It shows 52 observations are used in the test, just like before.

So, now we can reject the null hypothesis and proceed further in the model implementation.

Determining the order:

We have plotted the **ACF** and **PACF** plot figure 4.12 to find the order of the **ARIMA** model. The figure 4.12 shows two plots related to time series analysis: the **Autocorrelation Function (ACF)** plot on the top and the **Partial Autocorrelation Function (PACF)** plot on the bottom.

- **ACF plot:** It depicts the correlation between a time series and its lagged version. The x-axis shows the lags and the y-axis shows correlation coefficient, ranging from -1 to 1, to show negative and positive correlation. The blue shaded region shows the range

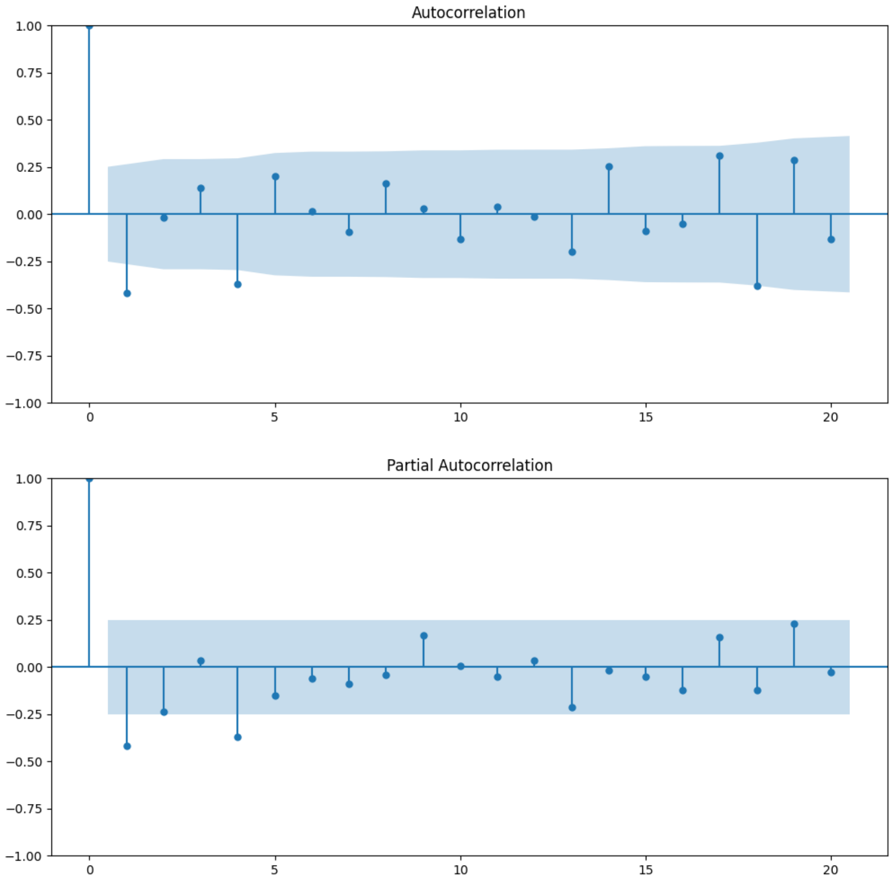

Figure 4.12: ACF & PACF plots.

inside which the correlations can be identified as statistically significant. If the values cross this region, it means that there is correlation at that particular lag. It helps to identify the MA (moving average) order, which is denoted by q in the order of ARIMA.

- **PACF plot:** It is almost similar to **ACF** except it depicts the correlation of the series with its lagged version, after deleting the effects of previous lags. It helps to identify the AR (autoregressive) order, which is denoted by p in the **ARIMA** order.

Model fitting:

Before applying the model, we have done all the necessary steps to perform the model. Now, we got the **orders** $(2, 1, 2)$ for fitting the model. After getting the order of the **ARIMA** model, we can apply the model to perform.

The figure 4.13 shows the result of ARIMA model. Below is the summary explanation:

Model Specification:

- **Dep. Variable:** Defining the dependent variable.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:     UK Temperature_Change   No. Observations:                   50
Model:                    ARIMA(2, 1, 2)   Log Likelihood                 -32.733
Date:                   Tue, 21 Nov 2023   AIC                             75.465
Time:                           05:23:24   BIC                             84.924
Sample:                                0   HQIC                            79.054
                                    - 50
Covariance Type:                     opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.9594      0.670      1.431      0.152      -0.354       2.273
ar.L2         -0.3431      0.222     -1.543      0.123      -0.779       0.093
ma.L1         -1.5547      0.654     -2.377      0.017      -2.836      -0.273
ma.L2          0.6669      0.546      1.221      0.222      -0.403       1.737
sigma2         0.2182      0.051      4.278      0.000       0.118       0.318
===================================================================================
Ljung-Box (L1) (Q):                   0.05   Jarque-Bera (JB):                 4.04
Prob(Q):                              0.83   Prob(JB):                         0.13
Heteroskedasticity (H):               1.11   Skew:                            -0.69
Prob(H) (two-sided):                  0.84   Kurtosis:                         3.26
===================================================================================
```

Figure 4.13: Model summary result.

- **Model:** It states that the model used $(2, 1, 2)$ order, **AR** term is 2, differencing is 1 and **MA** term is 2.
- **Date & time:** It displays the date & time of the implementation.
- **Sample:** It appears that model was trained on 50 values, as the size is not completely visible.
- **No. Observations:** Total 50 observations were used.

Model Fit Statistics:

- **Log Likelihood:** It tells about how well the model is fitted. The value of log likelihood is -32.73.
- **AIC:** In order to avoid overfitting, models with a penalty for the number of parameters are compared using the **Akaike Information Criterion**, which is 75.46.
- **BIC:** The **Bayesian Information Criterion**, same as AIC but larger penalty as comparable to the AIC, which is 84.92.
- **HQIC:** Another criteria model comparison, value is 79.05.

Parameter Estimates: The table has the lists of **AR** and **MA** terms along with their statistical results.

- **ar.L1** and **ar.L2**: coefficient of autoregressive component, two in total.
- **ma.L1** and **ma.L2**: coefficient of moving average component. It states that the first **MA** term is statistically significant ($p < 0.05$).

Diagnostics:

- **sigma2:** Estimated variance of residuals.
- **Ljung-Box Test:** An evaluation for the autocorrelation of residuals at different lag points. It indicates that no significant autocorrelation in the residuals, as the p-value ($Prob(Q)$) is 0.85.
- **Jarque-Bera Test:** To check the normality of residuals. As the $Prob(JB)$ is 0.13, residuals may be regarded as normally distributed.
- **Heteroskedasticity Test:** Checking the consistency of the variance of residuals over time. It indicates no significant heteroskedasticity because the $Prob(H)$ is 0.84.

Based on the above result, the model appears to be fair but not so good. Now, we have to plot the forecast result figure 4.14.

The forecast (figure 4.14) is not so good, we can check this with model diagnostic and evaluation, then we can do the **hyperparameter tuning** to see if the model improves the result or reducing the result.

Model Diagnostics:

The figure 4.15 shows four diagnostic plots which are used to check the model performance. Here is the 4 plots information below:

- **Top-Left(Standardized Residuals):** It depicts the differences between observed and predicted values, which are called standardized residuals, using standard deviation of residuals for scaling. The **residual should fluctuate around the zero** without any trend or pattern, this shows the model predictions are biased. In this values are revolve around zero but with more context, it suggests that there are patterns or outliers.

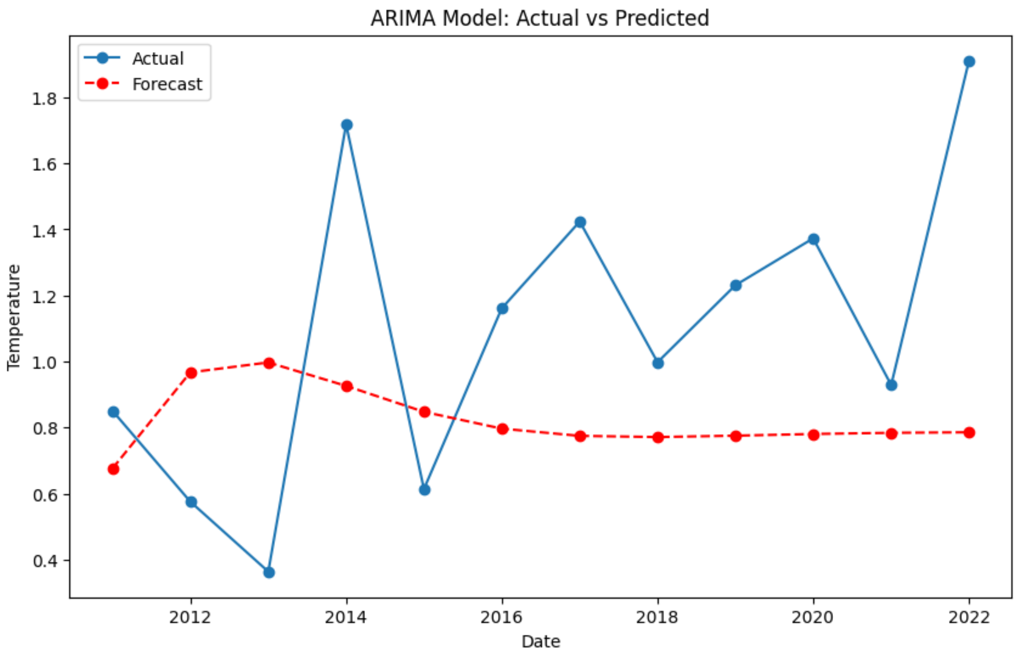

Figure 4.14: Model summary result.

- **Top-Right(Histogram with Kernel Density Estimate (KDE) and Normal Distribution Overlay)**: The blue bar shows the histogram of the residuals, displaying their distribution. The orange colour shows the **KDE**, providing a smooth estimate of the distribution. The standard distribution is shown by green colour, if residuals are distributed normally they will follow this. Basically, the residuals normality is given by this plot.

- **Bottom-Left(Normal Q-Q (Quantile-Quantile) Plot):** It equates the quantiles of the residuals to the quantiles of standard normal distribution. The points should revolve roughly along the red line. The **Q-Q** plot indicates that the residuals **almost follow the red line**, meaning they are almost normally distributed but not entirely, as some of them are away from the line.

- **Bottom-Right(Correlogram (or Autocorrelation Function - ACF):)** It depicts the autocorrelation of residuals at distinct lag values. The blue dots tell about the lags, and the shaded region tells about the confidence interval for statistical significance. If the dots are inside of it, it is not statistically significant.

Evaluation of model:

The result gives three metrics to evaluate the performance of the model:

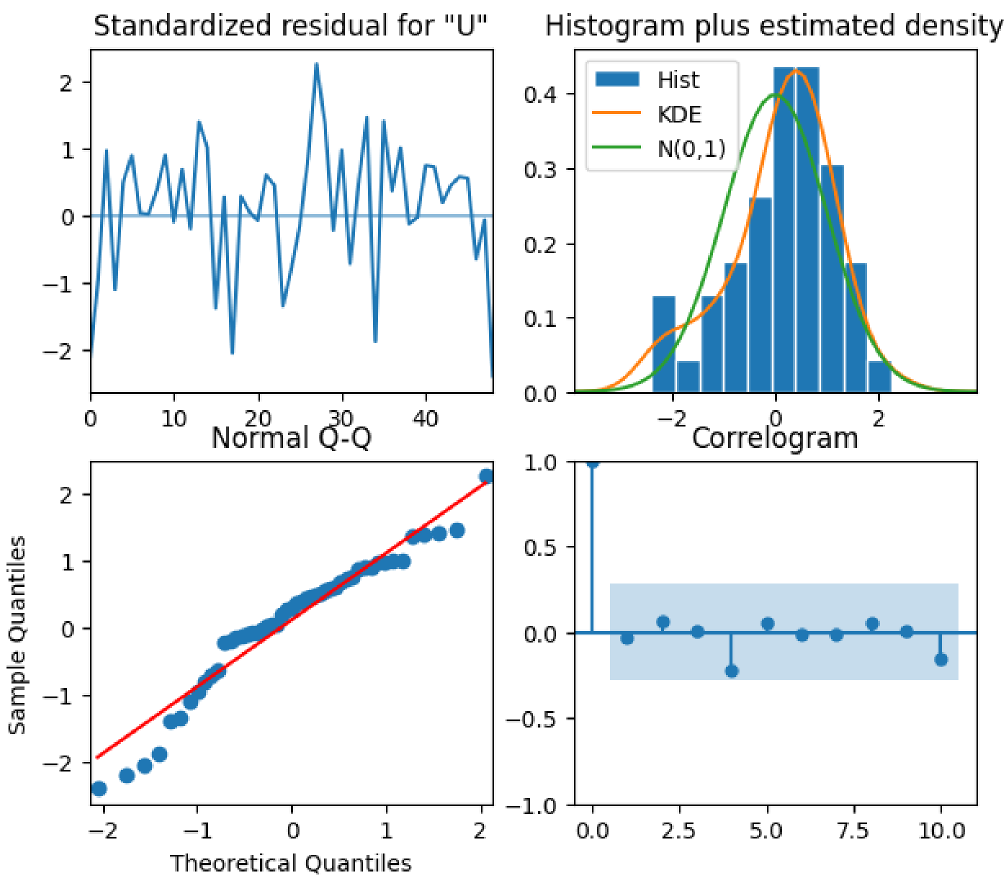

Figure 4.15: Diagnostic plot.

- **Mean Squared Error (MSE):** The average of the squares of errors. The difference between the predicted and actual values. Here, the **MSE** value is approximately 0.309.
- **Root Mean Squared Error (RMSE):** Square root of MSE, an error metric, is given in the same unit as the original, by taking the square root. The value of **RMSE** is 0.556.
- **Mean Absolute Error (MAE):** It gives the average of absolute errors(value of the difference between the predicted and actual). It doesn't square the errors, as **MSE** and **RMSE** do. The value of **MAE** is 0.481.

The smaller the values of these terms, the better the performance of the model will be. High values indicate poor results, and might do overfitting or underfitting the model.

In order to increase the performance of the model, we can do hyperparameter tuning to find the better optimized model.

Results after hyperparameter tuning:

After the hyperparameter tuning, results are slightly better than before. The figure 4.16

shows the model results. It shows that the results are pretty good. All the statistical values are point. Hyperparameter tuning gives a different order for ARIMA and it seems like it more better than the normal one.

```
                               SARIMAX Results
==================================================================================
Dep. Variable:     UK Temperature_Change   No. Observations:                   50
Model:                    ARIMA(5, 2, 5)   Log Likelihood                 -30.636
Date:                   Tue, 21 Nov 2023   AIC                             83.273
Time:                           05:39:11   BIC                            103.856
Sample:                                0   HQIC                            91.051
                                    - 50
Covariance Type:                     opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.6534      0.504     -3.278      0.001      -2.642      -0.665
ar.L2         -1.1871      0.638     -1.860      0.063      -2.438       0.064
ar.L3         -0.2645      0.760     -0.348      0.728      -1.754       1.225
ar.L4          0.0098      0.530      0.018      0.985      -1.030       1.049
ar.L5         -0.1912      0.322     -0.595      0.552      -0.822       0.439
ma.L1          0.1002    114.879      0.001      0.999    -225.059     225.259
ma.L2         -1.0651    144.570     -0.007      0.994    -284.417     282.287
ma.L3         -1.0804     10.263     -0.105      0.916     -21.196      19.035
ma.L4          0.0661    120.490      0.001      1.000    -236.090     236.223
ma.L5          0.9798    126.449      0.008      0.994    -246.855     248.814
sigma2         0.1479     19.099      0.008      0.994     -37.285      37.581
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                 2.41
Prob(Q):                              0.92   Prob(JB):                         0.30
Heteroskedasticity (H):               0.98   Skew:                            -0.45
Prob(H) (two-sided):                  0.97   Kurtosis:                         2.37
===================================================================================
```

Figure 4.16: Model summary after hyperparameter tuning.

The figure 4.17 shows that the model gives more better result after hyperparameter tuning. Almost similar to actual values at some point.

The diagnostics also shows the better results as compared to the normal one(before hyperparameters). The figure 4.18 depicts the diagnostics performance.

Evaluation of model after hyperparameters:

- **Mean Squared Error (MSE):** The **MSE** value is much better than the previous time. Now, this time it is 0.103.
- **Root Mean Squared Error (RMSE):** The value of **RMSE** is 0.321.
- **Mean Absolute Error (MAE):** The value of **MAE** is 0.264.

These values are actually good than before.

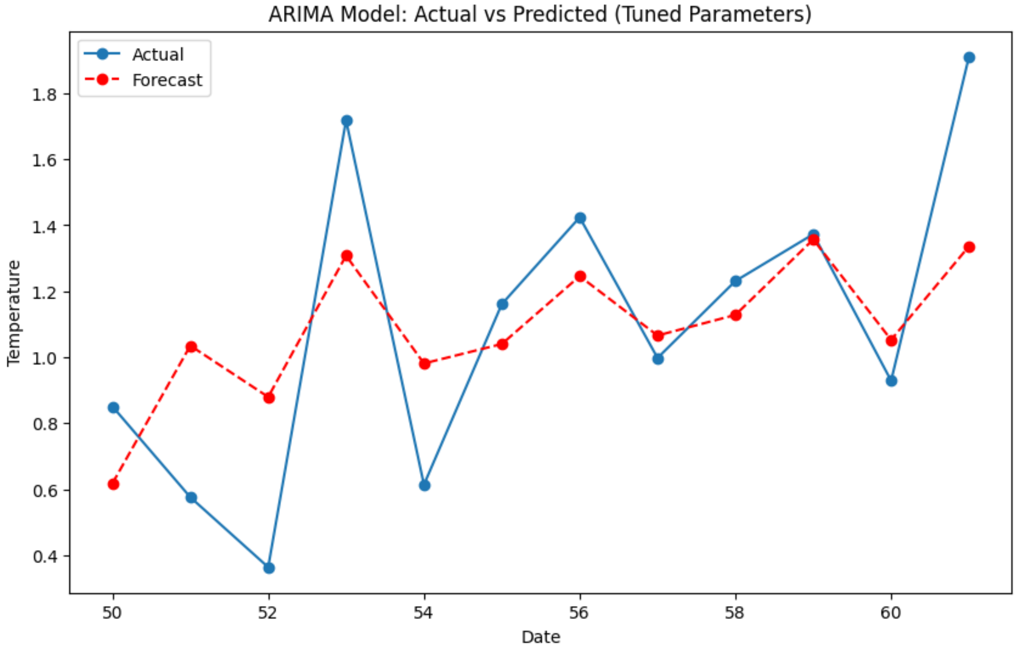

Figure 4.17: Forecast after hyperparameter tuning.

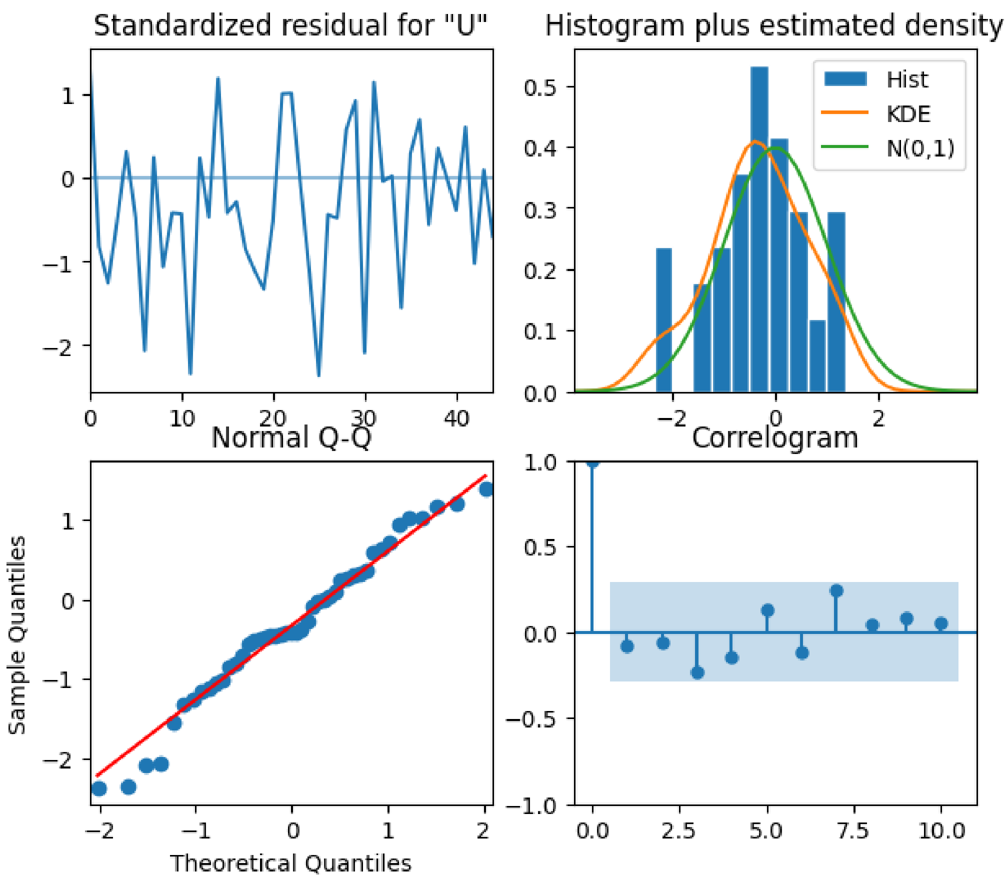

Figure 4.18: Diagnostics after hyperparameter tuning.

On dataset B: We performed the same process on ***dataset B*** as well. First we did preprocessing, then applied all the steps of ARIMA model.

Seasonality check: The figure 4.19 shows the seasonality trend of the ***dataset B***.

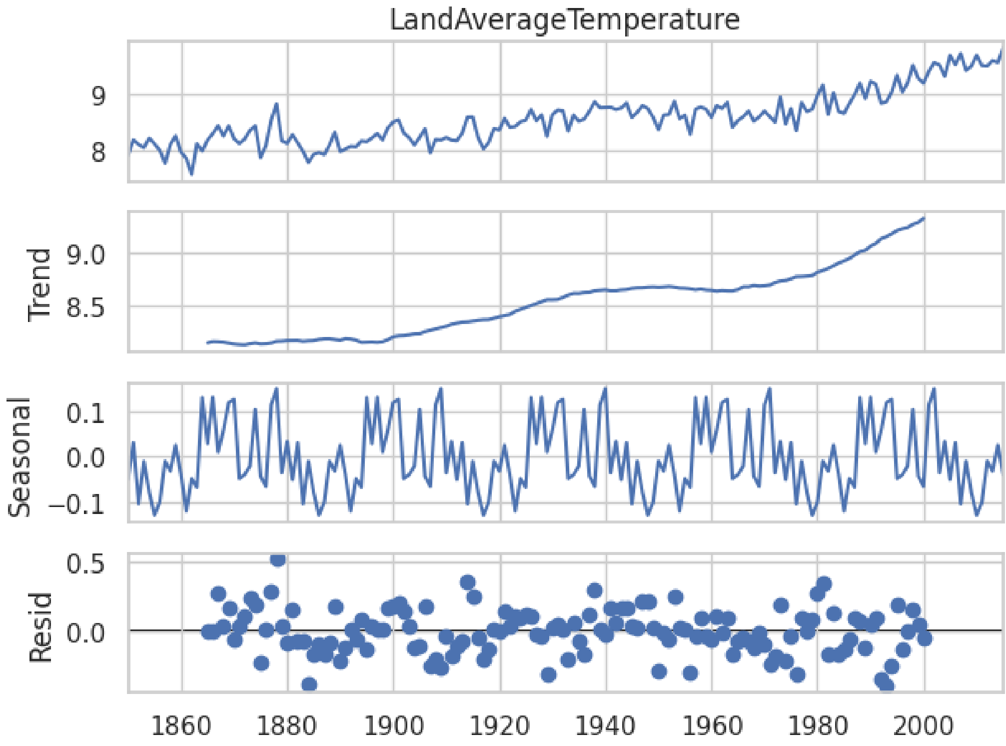

Figure 4.19: Diagnostics after hyperparameter tuning.

- **Top plot:** It depicts the temperature change of land average temperature over the years.
- **Trend:** Here in this case, it is going upward, which means temperature is increasing.
- **Seasonal:** Here in this, it depicts there is a fluctuation which seems to repeat itself, which means indicative of seasonal behaviour.
- **Residual:** In this case, residuals are very high as compared to ***dataset A***, and they are not revolve around zero.

Stationarity check: Now, we have to check the stationarity of the data. In order to do so, we have to apply **ADF()** test on this ***dataset B***. The table 4.3 shows that **p-value** is **higher than the significant level** (0.05), so we **failed to reject the null hypothesis** in this case. So, in order to **make the data stationary**, we need to do the differencing on it.

Statistic	Value
Test Statistic	-0.08
p-value	0.95
#Lags Used	3.00
Number of Observations Used	162.00
Critical Value (1%)	-3.47
Critical Value (5%)	-2.87
Critical Value (10%)	-2.57

Table 4.3: ADF test result on ***dataset B***.

Statistic	Value
Test Statistic	-1.353066×10^{1}
p-value	2.625627×10^{-25}
#Lags Used	2.000000×10^{0}
Number of Observations Used	1.620000×10^{2}
Critical Value (1%)	-3.471374×10^{0}
Critical Value (5%)	-2.879552×10^{0}
Critical Value (10%)	-2.576373×10^{0}

Table 4.4: ADF result after differencing on dataset B.

After the differencing, the data is stationary now. The figure 4.4 shows the result after differencing. Here, we can see now that **p-value** is now **smaller than the significance level** (0.05), which means now we can **reject the null hypothesis**. The test statistic value (-1.353066×10^{1}) now smaller than all the critical values(-3.471374×10^{0} at 1%, -2.879552×10^{0} at 5% and -2.576373×10^{0} at 10%).

Determining the order: Now, as before for finding the order, we need **ACF** and **PACF** plot, in order to get the **AR** order and **MA** order of the model. The figure 4.20 depicts the **ACF** and **PACF** information about the dataset. With this we can get the order for the model. After getting the orders$(3, 1, 2)$ we can apply **ARIMA** model on the ***dataset B***.

The figure 4.21 shows the statistical result of the model on the ***dataset B***.

Model Specification: Model order is $(3, 1, 2)$, it appears that the model trained on 132 values. Total number of observations are 132.

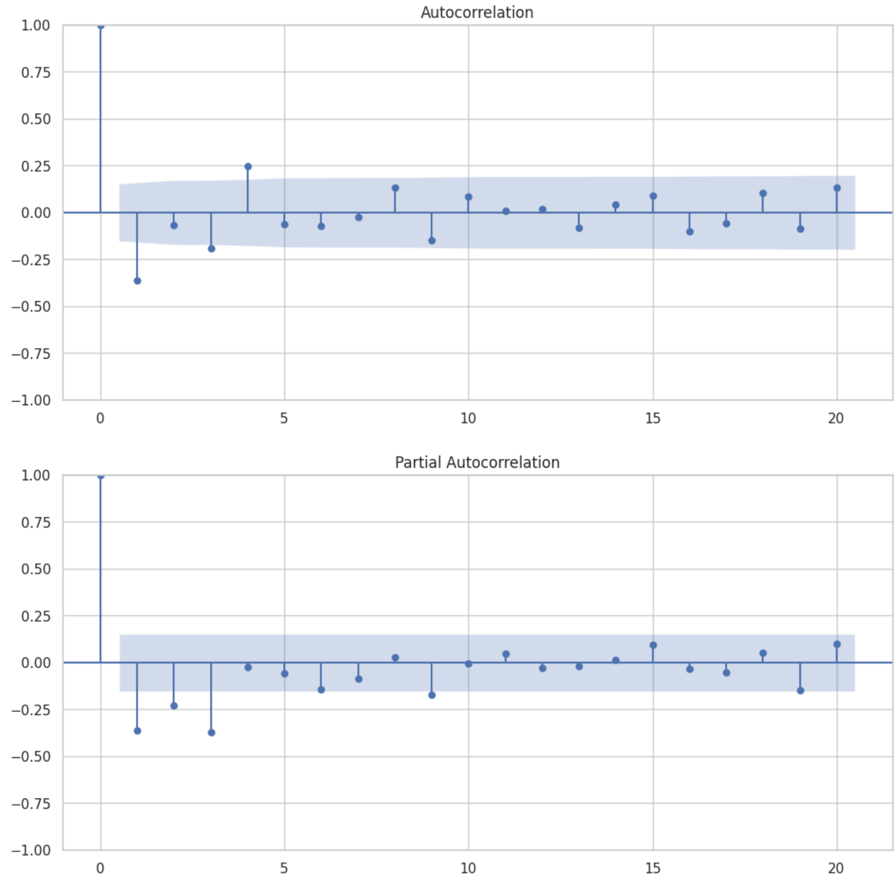

Figure 4.20: ACF and PACF plot of ***dataset B***.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:     LandAverageTemperature   No. Observations:             132
Model:                     ARIMA(3, 1, 2)   Log Likelihood             31.693
Date:                    Tue, 21 Nov 2023   AIC                       -51.385
Time:                            11:54:30   BIC                       -34.134
Sample:                                 0   HQIC                      -44.375
                                    - 132
Covariance Type:                      opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.4936      0.215     -2.298      0.022      -0.914      -0.073
ar.L2         -0.3829      0.102     -3.755      0.000      -0.583      -0.183
ar.L3         -0.4404      0.073     -6.041      0.000      -0.583      -0.297
ma.L1         -1.0034      2.097     -0.478      0.632      -5.114       3.107
ma.L2          0.0041      0.230      0.018      0.986      -0.447       0.455
sigma2         0.0341      0.073      0.470      0.638      -0.108       0.176
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                 0.93
Prob(Q):                              0.97   Prob(JB):                         0.63
Heteroskedasticity (H):               0.71   Skew:                            -0.12
Prob(H) (two-sided):                  0.26   Kurtosis:                         2.66
===================================================================================
```

Figure 4.21: Model summary of ***dataset B***.

Model Fit Statistics:

- **Log Likelihood:** The value of log likelihood is 31.69.
- **AIC:** The value of **Akaike Information Criterion** is -51.38.
- **BIC:** The **Bayesian Information Criterion**, same as **AIC** but larger penalty as comparable to the **AIC**, which is -34.13.
- **HQIC:** Another criteria model comparison, value is -44.37.

Parameter Estimates: The table has the lists of **AR** and **MA** terms along with their statistical results.

- **ar.L1, ar.L2** and **ar.L3**: coefficient of autoregressive component, two in total.
- **ma.L1 and ma.L2**: coefficient of moving average component. It states that the first **MA** term is statistically significant ($p < 0.05$).

Diagnostics:

- **sigma2:** Estimated variance of residuals.
- **Ljung-Box Test:** The $Prob(Q)$ value is 0.97.
- **Jarque-Bera Test:** The $Prob(JB)$ value is 0.63, residuals may be regarded as normally distributed.
- **Heteroskedasticity Test:** The heteroskedasticity value ($Prob(H)$) is 0.26.

Forecast the result: The figure 4.22 shows the forecast result on ***dataset B***. The red lines shows the predicted values, the blue line shows the actual values and the green line depicts the test data on which the prediction has done.

Below the figure 4.23 shows the diagnostic plot of ***dataset B***.

Model Diagnostics:

- **Top-Left(Standardized Residuals):** In this values are revolve around zero but with more context, it suggests that there are patterns or outliers.
- **Top-Right(Histogram with Kernel Density Estimate (KDE) and Normal Distribution Overlay)**: The orange line(KDE) and green line(ormal distribution) are almost similar, which means the model is performed well.

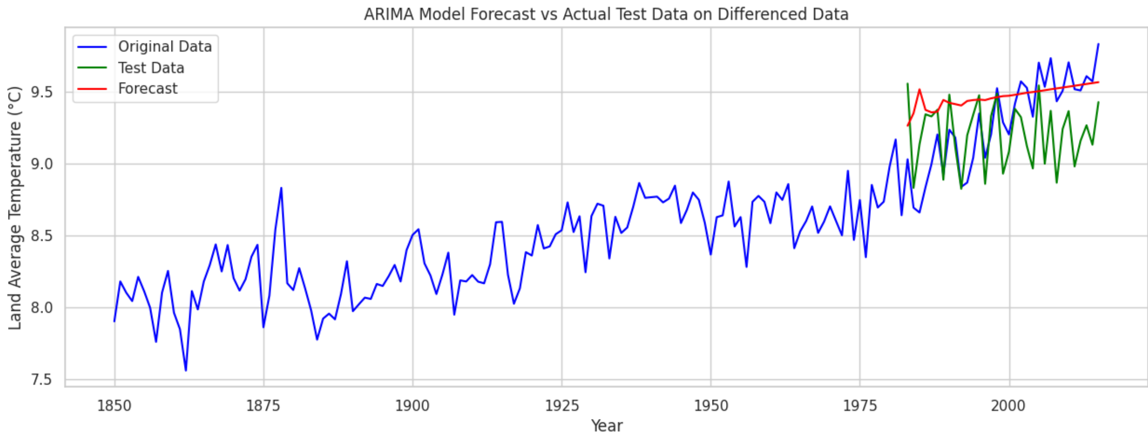

Figure 4.22: Forecast plot of ***dataset B***.

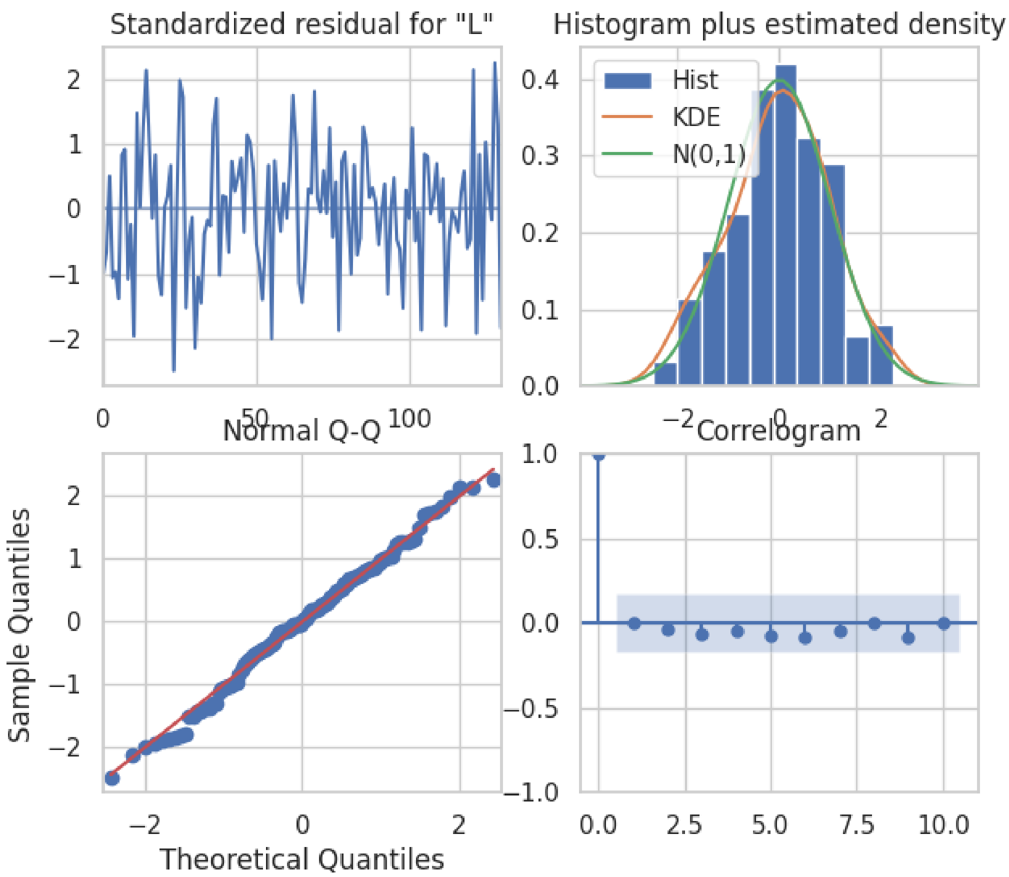

Figure 4.23: Residual plot of ***dataset B***.

- **Bottom-Left(Normal Q-Q (Quantile-Quantile) Plot):** The **Q-Q plot** indicates that the residuals follow the red line, which is a good thing.
- **Bottom-Right(Correlogram (or Autocorrelation Function - ACF):)** It depicts the autocorrelation of residuals at distinct lag values.

Evaluation of model:

- **Mean Squared Error (MSE):** The **MSE** value is 0.054, which is okay.
- **Root Mean Squared Error (RMSE):** The value of **RMSE** is 0.233.
- **Mean Absolute Error (MAE):** The value of **MAE** is 0.205.

Now, we applied **hyperparameter** tuning on **ARIMA** model to identify the optimized model. The figure 4.24 shows the model summary after the hyperparameter tuning.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:     LandAverageTemperature   No. Observations:              132
Model:                     ARIMA(0, 0, 1)   Log Likelihood              29.873
Date:                    Tue, 21 Nov 2023   AIC                        -53.745
Time:                            11:58:02   BIC                        -45.097
Sample:                                 0   HQIC                       -50.231
                                    - 132
Covariance Type:                      opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0060      0.004      1.432      0.152      -0.002       0.014
ma.L1         -0.7577      0.059    -12.817      0.000      -0.874      -0.642
sigma2         0.0370      0.005      7.865      0.000       0.028       0.046
===================================================================================
Ljung-Box (L1) (Q):                   3.06   Jarque-Bera (JB):                 0.59
Prob(Q):                              0.08   Prob(JB):                         0.75
Heteroskedasticity (H):               0.64   Skew:                            -0.16
Prob(H) (two-sided):                  0.14   Kurtosis:                         2.92
===================================================================================
```

Figure 4.24: Model summary after **hyperparameters**.

The figure 4.25 shows the forecast result after the hyperparameter tuning. It shows the predicted values by red line, actual values through blue line and test values by green line.

The figure 4.26 shows the diagnostic plot after hyperparameter tuning. It shows that standardized plot shows the lags around zero but also there are some outliers which are far from zero, which might effect the result. The Q-Q plot also states that the residuals are revolve around the red line, except for the few outliers.

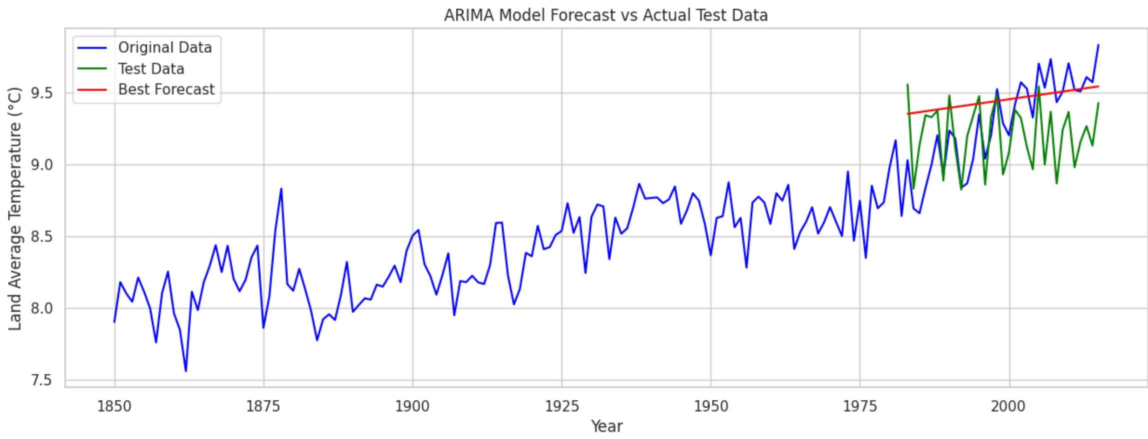

Figure 4.25: Forecast after **hyperparameters**.

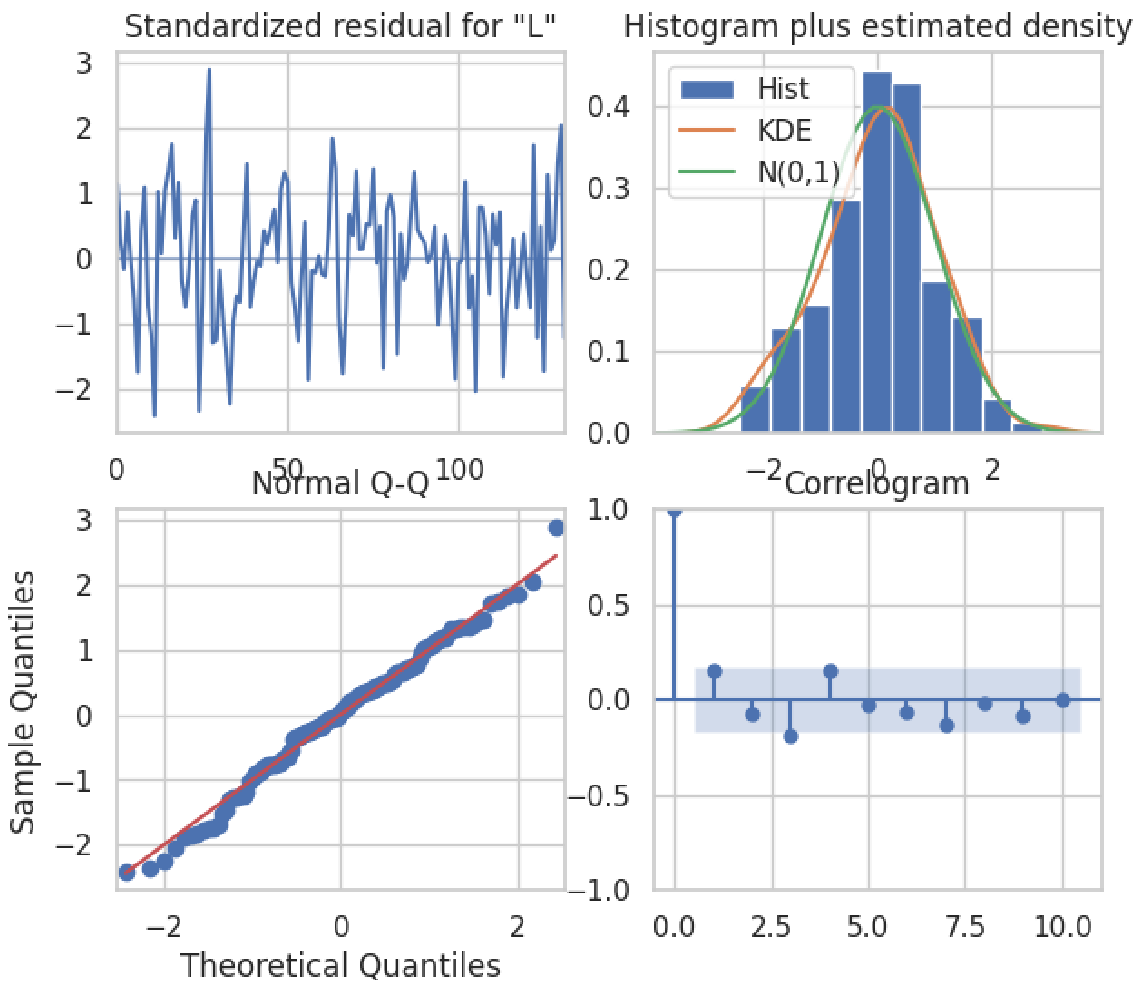

Figure 4.26: Diagnostic plot after **hyperparameters**.

Evaluation of model after hyperparameters:

- **Mean Squared Error (MSE):** The **MSE** value is much better than the previous time. Now, this time it is 0.069.
- **Root Mean Squared Error (RMSE):** The value of **RMSE** is 0.262.
- **Mean Absolute Error (MAE):** The value of **MAE** is 0.298.

Comparison between the results:

In similarity, we can say that the **ARIMA** model gave better result on both the kind of datasets. It gives the good forecast result and peformance was also good. If see the figure 4.14 and the figure 4.22 the results shows the similarities with each other. It states that the forecasted values first catches the similar trend as actual values, then later in the graph shows that the values went almost in a straight path. The evaluation score of both datasets was good initially before the hyperparameters tuning.

For ***dataset A***

- Mean Squared Error (MSE): 0.309
- Root Mean Squared Error (RMSE): 0.556
- Mean Absolute Error (MAE): 0.481

For ***dataset B***

- Mean Squared Error (MSE): 0.054
- Root Mean Squared Error (RMSE): 0.233
- Mean Absolute Error (MAE): 0.205

In contrast with each other, the ARIMA model performed more accurately on the ***dataset A*** as compare to ***dataset B***, after the implementation of hyperparameters. If we see the figure 4.17, the forecast after hyperparameters on ***dataset A*** , the forecast is more accurate as compared to the result of without hyperparameters and the figure 4.25, the forecast after hyperparamters on ***dataset B***, the forecast is similar to the result of before hyperparameters, just a minor changes. The evaluation score of both datasets has changed after the hyperparameters tuning: For ***dataset A***

- Mean Squared Error (MSE): 0.103
- Root Mean Squared Error (RMSE): 0.321
- Mean Absolute Error (MAE): 0.264

For ***dataset B***

- Mean Squared Error (MSE): 0.069
- Root Mean Squared Error (RMSE): 0.262
- Mean Absolute Error (MAE): 0.240

Forecast results for next 5 years

We take dataset A for doing forecast of next 5 years because as it is cleared that ARIMA worked well on ***dataset A***, evaluation score is more better on ***dataset A*** as compared to ***dataset B***. So based on evaluation score, we selected ***dataset A*** for doing the forecast. The figure 4.27 depicts the future forecast result.

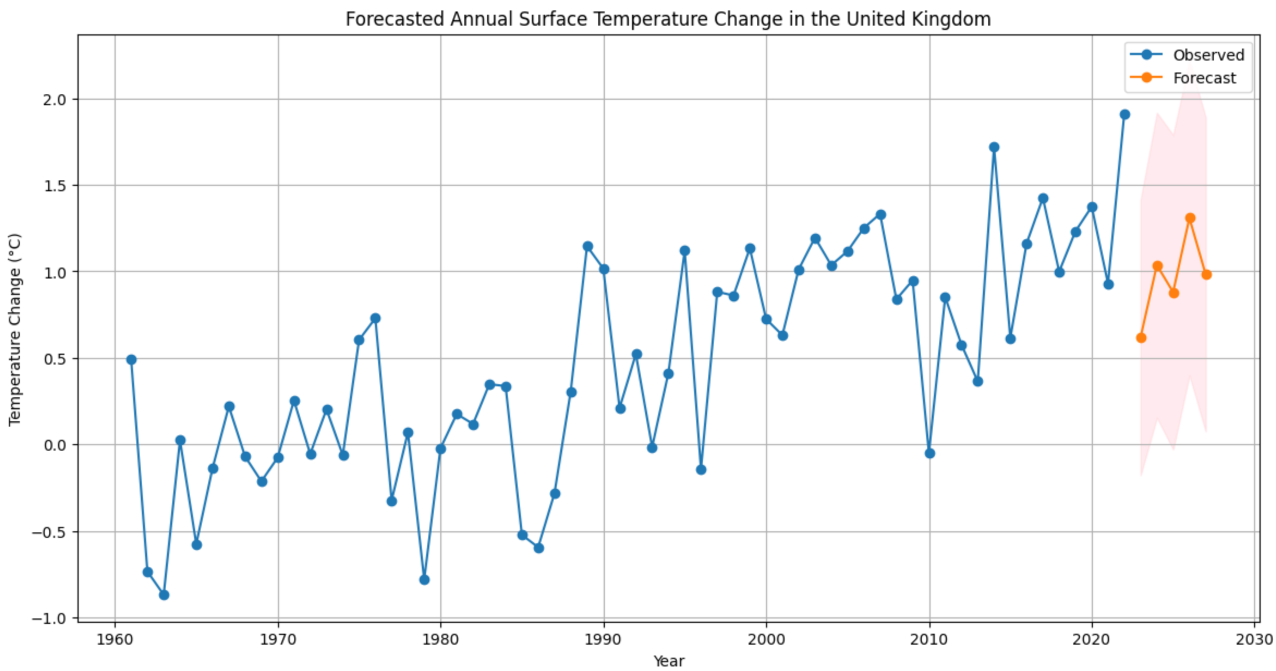

Figure 4.27: Next 5 years forecasting on the UK data.

CHAPTER 5

Conclusions

In conclusion, the research depicts the performance and results of the **ARIMA** model on two different datasets, named **dataset A** and **dataset B**. The performance is almost good, but not perfect. The performances are slightly different on both datasets. In similarity, on both the dataset, the **ARIMA** model worked just okay or not so good. But after the use of the **hyperparameter** tuning, the results changed. On **dataset A**, it performed really well, predicted values are almost similar to the actual values, while on the other hand, on **dataset B**, the performance reduced a little bit as compared to the results without **hyperparameter** tuning. The performances are analysed by the MSE, RMSE and MAE. These error scores depict the efficiency of the model, that how well the model performs or not. The lower these values are, the more is the good performance. The values of **dataset A** got reduced after the **hyperparameter** tuning, but the results are reversed for **dataset B**, the values increased a little bit. This might be possible because of the type of data or data values, sometimes that matters.

Moreover, we performed some analysis regarding global temperature change on both datasets. In all types of analyses, like trend analysis, comparison between developed and developing countries- it shows that developed countries have more hike in the temperature change, seasonal temperature change or continental temperature variations, yearly temperature difference with maximum and minimum values, comparative analysis between land and ocean temperature change and temperature change in different decades, all of them indicate a similar pattern. The values suggest that temperature is rising upwards. The reasons for increasing temperature are increasing global warming due to the effect of greenhouse gases, deforestation, burning fossil fuels or farming livestock and one of the main reasons is human activities. According to the **IPCC** and **2015 Paris COP21 Agreement** and other international level communities have confirmed that there is a need to put the **limit** 1.5 degrees Celsius on temperature. It should not increase 2 degrees Celsius, it might seem small but this is a huge increase as compared to temperature change in pre industrial time period, which is between 1720 to 1900. Rising temperature causes glaciers to melt and rise in sea-level, which can generate floods, plus other environmental problems. We should do things to reduce the temperature change globally.Planting more trees, doing environmental activities to reduce the temperature change.

Bibliography

[1] NASA. *World of Change: Global Temperatures*. https://earthobservatory.nasa.gov/world-of-change/global-temperatures.

[2] Lindsey, Rebecca, and Luann Dahlman. "Climate Change: Global Temperature | NOAA Climate.gov." *Climate.gov*, 18 January 2023. https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature.

[3] Climate Action. *Causes of climate change*. https://climate.ec.europa.eu/climate-change/causes-climate-change_en.

[4] Wikipedia contributors. *Global surface temperature — Wikipedia, The Free Encyclopedia*. 2023. https://en.wikipedia.org/w/index.php?title=Global_surface_temperature&oldid=1182362250. [Online; accessed 10-November-2023].

[5] Alan Buis. A Degree of Concern: Why Global Temperatures Matter. *NASA*. Oct 2020. https://climate.nasa.gov/news/2865/a-degree-of-concern-why-global-temperatures-matter/.

[6] Alan Buis. A Degree of Concern: Why Global Temperatures Matter. *NASA*, Jun 2019. https://climate.nasa.gov/news/2878/a-degree-of-concern-why-global-temperatures-matter/.

[7] NCEI Monitoring Info. *Annual 2022 Global Climate Report*. https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/202213 *Annual 2022 Global Climate Report | National Centers for Environmental Information (NCEI)*. Author: NCEI.Monitoring.Info@noaa.gov.

[8] Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Ruedy, R., and Zyss, D. Improvements in the GISTEMP uncertainty model. *J. Geophys. Res. Atmos.*, 2019, **124**(12), 6307–6326. doi: 10.1029/2018JD029522.

[9] GISTEMP Team 2023. GISS Surface Temperature Analysis (GISTEMP), version 4. NASA Goddard Institute for Space Studies. Dataset accessed 20YY-MM-DD at https://data.giss.nasa.gov/gistemp/.

[10] Environment 2022. *Global Warming Effects*. *National Geographic*, Feb 2022. https://www.nationalgeographic.com/environment/article/global-warming-effects.

[11] Climate Action 2023. *Consequences of Climate Change*. *Climate Action*, 2023. https://climate.ec.europa.eu/climate-change/consequences-climate-change_en.

[12] NASA 2023. *Effects of Climate Change*. *NASA*, Sep 2023. https://climate.nasa.gov/effects/.

[13] Melissa Denchak. *Are the Effects of Global Warming Really That Bad?*. *Consequences and Effects of Global Warming – What is the Impact?*, May 2022. https://www.nrdc.org/stories/are-effects-global-warming-really-bad.

[14] NASA. *Solar System Temperatures*. https://science.nasa.gov/resource/solar-system-temperatures/.

[15] IPCC. *Intergovernmental Panel on Climate Change*. *IPCC*. https://www.ipcc.ch/.

[16] Global Warming 1.5°C. *Global Warming of 1.5°C*. https://www.ipcc.ch/sr15/.

[17] Global Warming 1.5°C. *Chapter*. *SPM*. https://www.ipcc.ch/sr15/chapter/spm/.

[18] IPCC Chapter1. *Chapter 1 - Global Warming of 1.5°C - IPCC*. https://www.ipcc.ch/sr15/chapter/chapter-1/.

[19] IPCC Chapter2. *Chapter 2 - Global Warming of 1.5 Â°C - IPCC*. https://www.ipcc.ch/sr15/chapter/chapter-2/.

[20] IPCC 2022. Intergovernmental Panel on Climate Change (IPCC). *Annex I: Glossary*. In *Global Warming of 1.5°C: IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty* (pp. 541–562). Cambridge: Cambridge University Press, 2022. https://doi.org/10.1017/9781009157940.008.

[21] Usgcrp. *Executive Summary*. *Climate Science Special Report*. https://science2017.globalchange.gov/chapter/executive-summary/#fig-3.

[22] NCEI2021. NCEI.Monitoring.Info@noaa.gov. *Annual 2021 Global Climate Report*. *Annual 2021 Global Climate Report | National Centers for Environmental Information (NCEI)*. https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/202113.

[23] Wikipedia_2023. *Climate Change*. *Wikipedia*, Nov 2023. https://en.wikipedia.org/wiki/Climate_change.

[24] WisconsinDNR_ClimateChange *The Science of Climate Change | The World is Warming | Wisconsin DNR*. https://dnr.wisconsin.gov/climatechange/science.

[25] Addae Maxwell, Darkwah Williams Kweku1, Odum Bismark, Koomson Ato Desmond, Kwakye Benjamin Danso, Ewurabena Asante Oti-Mensah, Asenso Theophilus Quachie and Buanya Beryl Adormaa. *Greenhouse Effect, Greenhouse Gases and Their Impact on Global Warming*. https://www.researchgate.net/profile/Maxwell-Addae/publication/323223192_Greenhouse_Effect_Greenhouse_Gases_and_Their_Impact_on_Global_Warming/links/5ab29e670f7e9b4897c5933b/Greenhouse-Effect-Greenhouse-Gases-and-Their-Impact-on-Global-Warming.pdf.

[26] NASA_2014. *Secrets from the Past Point to Rapid Climate Change in the Future*. *NASA*, Sep 2014. https://climate.nasa.gov/news/649/secrets-from-the-past-point-to-rapid-climate-change-in-the-future/.

[27] Amanda MacMillan, Jeff Turrentine. *Global Warming 101*. *Definition, Facts, Causes and Effects of Global Warming*, Apr 2021. https://www.nrdc.org/stories/global-warming-101#warming.

[28] NASA_2023. *Evidence for Climate Change*. *NASA*, Sep 2023. https://climate.nasa.gov/evidence/.

[29] Wikipedia_2023. *Sea Level Rise*. *Wikipedia*, Nov 2023. https://en.wikipedia.org/wiki/Sea_level_rise.

[30] IUCN_2023. *Resources*. *IUCN*, Jan 2023. https://www.iucn.org/resources.

[31] IUCN_2018. *Ocean Warming*. *IUCN*, Dec 2018. https://web.archive.org/web/20200919112407/https://www.iucn.org/resources/issues-briefs/ocean-warming.

[32] Rebecca Lindsey and LuAnn Dahlman. *Climate Change: Ocean Heat Content*. *NOAA Climate.gov*. https://www.climate.gov/news-features/understanding-climate/climate-change-ocean-heat-content.

[33] NOAA Climate Gov. *NOAA Climate.gov Media*. *NOAA Climate.gov*. https://www.climate.gov/media/15576.

[34] EPASST. *Climate Change Indicators: Sea Surface Temperature | US EPA*. https://www.epa.gov/climate-indicators/climate-change-indicators-sea-surface-temperature.

[35] Wikipedia. *Antarctic Ice Sheet*. *Wikipedia*, Aug 2023. https://en.wikipedia.org/wiki/Antarctic_ice_sheet.

[36] Wikipedia. *Greenland Ice Sheet*. *Wikipedia*, Nov 2023. https://en.wikipedia.org/wiki/Greenland_ice_sheet.

[37] J. Hansen, R. Ruedy, M. Sato and K. Lo *Global Surface Temperature Change - Hansen - 2010 - Wiley Online Library*. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010RG000345.

[38] Ed Hawkins, Pablo Ortega, Emma Suckling, Andrew Schurer, Gabi Hegerl, Phil Jones, Manoj Joshi, Timothy J. Osborn, ValÃ©rie Masson-Delmotte, Juliette Mignot, et al. *Estimating Changes in Global Temperature since the Preindustrial Period*. *AMETSOC*, Sep 2017. https://journals.ametsoc.org/configurable/content/journals$002fbams$002f98$002f9$002fbams-d-16-0007.1.xml.

[39] UNFCCC_ParisAgreement *Adoption of the Paris Agreement - UNFCCC*. *Adoption of the Paris Agreement - UNFCCC*. https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf.

[40] UNFCCC_Convention *United Nations Framework Convention on Climate Change - UNFCCC*. *United Nations Framework Convention on Climate Change - UNFCCC*. https://unfccc.int/files/essential_background/background_publications_htmlpdf/application/pdf/conveng.pdf.

[41] Manoj Joshi, Ed Hawkins, Rowan Sutton, Jason Lowe, and David Frame. *Projections of When Temperature Change Will Exceed 2 Â°C above Pre-Industrial Levels*. *Nature News*, Oct 2011. https://www.nature.com/articles/nclimate1261.

[42] IPCC AR5. *Climate Change 2014: Synthesis Report*. *IPCC*. https://www.ipcc.ch/report/ar5/syr/.

[43] IPCC_SPM1. *SPM1 Summary for Policymakers - IPCC*. https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_SPM_FINAL.pdf.

[44] Malte Meinshausen, S. J. Smith, K. Calvin, J. S. Daniel, M. L. T. Kainuma, J-F. Lamarque, K. Matsumoto, S. A. Montzka, S. C. B. Raper, K. Riahi, et al. *The RCP Greenhouse Gas Concentrations and Their Extensions from 1765 to 2300 - Climatic Change*. *SpringerLink*, Aug 2011. https://link.springer.com/article/10.1007/s10584-011-0156-z.

[45] Grant Foster and Stefan Rahmstorf. *Global Temperature Evolution 1979-2010 - IOPscience*. *Global Temperature Evolution 1979-2010 - IOPscience*. https://iopscience.iop.org/article/10.1088/1748-9326/6/4/044022/meta.

[46] Thomas M. Smith and Richard W. Reynolds. *A Global Merged Land-Air-Sea Surface Temperature Reconstruction Based on Historical Observations (1880-1997)*. *AMETSOC*, Jun 2005. https://journals.ametsoc.org/view/journals/clim/18/12/jcli3362.1.xml?tab_body=abstract-display.

[47] Thomas M. Smith, Richard W. Reynolds, Thomas C. Peterson, and Jay Lawrimore. *Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis*

(1880-2006). *AMETSOC*, May 2008. https://journals.ametsoc.org/view/journals/clim/21/10/2007jcli2100.1.xml.

[48] P. Brohan, J. J. Kennedy, I. Harris, S. F. B. Tett and P. D. Jones. *Wiley_UncertaintyEstimates*. *Uncertainty Estimates in Regional and Global ... - Wiley Online Library*. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2005JD006548.

[49] Carl A. Mears and Frank J. Wentz. *Construction of the RSS v3.2 Lower-Tropospheric Temperature Dataset from the MSU and AMSU Microwave Sounders*. *AMETSOC*, Aug 2009. https://journals.ametsoc.org/view/journals/atot/26/8/2009jtecha1237_1.xml.

[50] John R. Christy, Roy W. Spencer, and William D. Braswell. *MSU Tropospheric Temperatures: Dataset Construction and Radiosonde Comparisons*. *AMETSOC*, Sep 2000. https://journals.ametsoc.org/view/journals/atot/17/9/1520-0426_2000_017_1153_mttdca_2_0_co_2.xml?tab_body=abstract-display.

[51] IPCC_AR4. *Climate Change 2007: The Physical Science Basis*. *IPCC*. https://www.ipcc.ch/report/ar4/wg1/.

[52] REGINALD E. NEWELL AND BRYAN C. WEARE *Science_TroposphericTemp*. *Factors Governing Tropospheric Mean Temperature | Science*. https://www.science.org/doi/10.1126/science.194.4272.1413.

[53] Caspar M. Ammann, Gerald A. Meehl, Warren M. Washington and Charles S. Zender. *A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate - Wiley Online Library*. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2003GL016875.

[54] Stefan Rahmstorf, Grant Foster and Niamh Cahill. *IOPScience_GlobalTempEvolution*. *Global Temperature Evolution: Recent Trends and Some Pitfalls - IOPscience*. https://iopscience.iop.org/article/10.1088/1748-9326/aa6825.

[55] Grant Foster and John Abraham *Variations US CLIVAR Variations US CLIVAR*. *Lack of evidence for a slowdown in global temperature*. https://usclivar.org/sites/default/files/documents/2015/Variations2015Summer-1_0.pdf.

[56] B. Rajaratnam, J. Romano, M. Tsiang, et al. *Debunking the Climate Hiatus*. *Climatic Change, 133, 129–140, 2015.* https://doi.org/10.1007/s10584-015-1495-y.

[57] Stephan Lewandowsky, James S. Risbey, and Naomi Oreskes. *The "Pause" in Global Warming: Turning a Routine Fluctuation into a Problem for Science*. *AMETSOC*, May 2016. https://journals.ametsoc.org/view/journals/bams/97/5/bams-d-14-00106.1.xml.

[58] Seyed Matin Malakouti. *Utilizing Time Series Data from 1961 to 2019 Recorded Around the World and Machine Learning to Create a Global Temperature Change Prediction Model*. *Case Studies in Chemical and Environmental Engineering*, Feb 2023. https://www.sciencedirect.com/science/article/pii/S2666016423000178.

[59] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*. *Advances in Neural Information Processing Systems*, Jan 1970. https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.

[60] Mingzhu Tang, Qi Zhao, Steven X. Ding, Huawei Wu, Linlin Li, Wen Long, and Bin Huang. *An Improved LightGBM Algorithm for Online Fault Detection of Wind Turbine Gearboxes*. *MDPI*, Feb 2020. https://www.mdpi.com/1996-1073/13/4/807.

[61] Hirofumi Michimae and Takeshi Emura. *Bayesian Ridge Estimators Based on Copula-Based Joint Prior Distributions for Regression Coefficients - Computational Statistics*. *SpringerLink*, Mar 2022. https://link.springer.com/article/10.1007/s00180-022-01213-8.

[62] NOAA_2021WarmestYear. *2021 Was World's 6th-Warmest Year on Record | National Oceanic and Atmospheric Administration*. https://www.noaa.gov/news/2021-was-worlds-6th-warmest-year-on-record.

[63] C. Xiao, N. Chen, C. Hu, K. Wang, J. Gong, and Z. Chen. *Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach*. https://www.sciencedirect.com/science/article/abs/pii/S0034425719303773.

[64] Jeremy Kun. *The boosting margin, or why boosting doesn't Overfit*. https://jeremykun.com/2015/09/21/the-boosting-margin-or-why-boosting-doesnt-overfit/.

[65] P. Bartlett, Y. Freund, W. S. Lee, and R. E. Schapire. *Boosting the margin: A new explanation for the effectiveness of voting methods*. https://projecteuclid.org/journals/annals-of-statistics/volume-26/issue-5/Boosting-the-margin--a-new-explanation-for-the-effectiveness/10.1214/aos/1024691352.full.

[66] D. H. Wolpert. *Stacked generalization*. *Neural Networks, Pergamon, Oct. 2005*. https://www.sciencedirect.com/science/article/abs/pii/S0893608005800231.

[67] J. Schmidhuber, S. Hochreiter, et al. *Long short-term memory*. *Long Johannes Kepler University Linz*. https://www.bioinf.jku.at/publications/older/2604.pdf.

[68] H. Liu, H.-q. Tian, Y.-f. Li, and L. Zhang. *Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions*. *Energy Conversion and Management, Pergamon, Jan. 2015*. https://www.sciencedirect.com/science/article/pii/S0196890414010929.

[69] M. Vermeer and S. Rahmstorf. *"Global sea level linked to global temperature"*. *PNAS, 2009*. Available: https://www.pnas.org/doi/abs/10.1073/pnas.0907765106

[70] Hyndman, Rob J and Athanasopoulos, George. *"Forecasting: Principles and Practice (2nd ed)"*. *OTexts*. https://otexts.com/fpp2/stationarity.html.

[71] Hayes, Adam. *"Autoregressive Integrated Moving Average (ARIMA) Prediction Model"*. *Investopedia*. https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp.

[72] Wikipedia. *"Autoregressive Integrated Moving Average"*. *Wikimedia Foundation, Oct 2023*. https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average.

[73] Wikipedia_2022. *"Box-Jenkins Method"*. *Wikimedia Foundation, Jul 2022*. https://en.wikipedia.org/wiki/Box%E2%80%93Jenkins_method.

[74] Brownlee, Jason. *"A Gentle Introduction to the Box-Jenkins Method for Time Series Forecasting". MachineLearningMastery.com, Aug 2020.* https://machinelearningmastery.com/gentle-introduction-box-jenkins-method-time-series-forecasting/.

[75] Sánchez, David Andrés. *"ARIMA for Time Series Forecasting - ML Pills". ML Pills - Machine Learning Pills, May 2023.* https://mlpills.dev/cheatsheets/cheatsheet-arima-for-time-series-forecasting/.

[76] Sánchez, David Andrés. *"Parameters Selection in ARIMA Models". ML Pills - Machine Learning Pills, Jan 2023.* https://mlpills.dev/time-series/parameters-selection-in-arima-models/.

[77] Prabhakaran, Selva. *"ARIMA Model - Complete Guide to Time Series Forecasting in Python: ML+". Machine Learning Plus, Sep 2023.* https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/#google_vignette.