Project Report

On

TIME SERIES ANALYSIS ON SURFACE TEMPERATURE IN ERNAKULAM

Submitted

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

APPLIED STATISTICS AND DATA ANALYTICS

by

AKHILA K PRAMOD

Reg No. SM22AS002

(2022-2024)

Under the supervision of

ANAKHA KURIAKOSE

DEPARTMENT OF MATHEMATICS AND STATISTICS (SF)

ST. TERESA'S COLLEGE (AUTONOMOUS)

ERNAKULAM, KOCHI – 682011

APRIL 2024

ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

CERTIFICATE

This is to certify that the dissertation entitled, **' TIME SERIES ANALYSIS ON SURFACE TEMPERATURE IN ERNAKULAM'** is a bonafide record of the work done by **Ms. AKHILA K PRAMOD** under my guidance as partial fulfillment of the award of the degree of **Master of Science in Applied Statistics and Data Analytics** at St. Teresa's College (Autonomous), Ernakulam affiliated to Mahatma Gandhi University, Kottayam. No part of this work has been submitted for any other degree elsewhere.

Date: 29/04/2024

Place: Ernakulam

Anakha Kuriakose
Assistant Professor,
Department of Mathematics and Statistics (SF)
St. Teresa''s College (Autonomous)
Ernakulam

Nisha Oommen
Assistant Professor & HOD
Department of Mathematics and Statistics (SF)
St. Teresa's College (Autonomous)
Ernakulam

External Examiners

1. ....
Chinu Joseph

2. ....
LAKSHMI SURESH.

DECLARATION

I hereby declare that the work presented in this project is based on the original work done by me under the guidance of **ANAKHA KURIAKOSE**, Assistant Professor, Department of Mathematics and Statistics, St. Teresa's College (Autonomous), Ernakulam and has not been included in any other project submitted previously for the award of any degree.

Ernakulam

Date: 29/04/2024

AKHILA K PRAMOD

SM22AS002

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to several individuals who encouraged me to carry out this work. Their continuous valuable input, feedback, and guidance throughout this study helped me to complete the work up to this stage.

Furthermore, I am very grateful to my project guide Anakha Kuriakose for the immense help during the period of work.

In addition, the very energetic and competitive atmosphere of the Department had much to do with this work. I acknowledge with thanks to the faculty, teaching, and non-teaching staff of the department and colleagues.

I am also very thankful to HOD Mrs. Nisha Oomen for their valuable suggestions and critical examination of work during the process.

Ernakulam

Date: 29/04/2024

AKHILA K PRAMOD

SM22AS002

ABSTRACT

Understanding climate change is of prime importance nowadays. Since heatwaves, droughts, global warming, etc... are becoming major threats to both humans and ecosystems. This study understands the trends of maximum and minimum temperatures in Ernakulam, Kerala over the past 13 years and forecasts 2-year maximum and minimum temperature values using the Time Series model SARIMA (seasonal Autoregressive Integrated Moving Average). For analyzing the trend patterns of past and future data Seasonal Mann-Kendall test, Sen's Slope, and Rolling average analysis are used. The SARIMA model captures the seasonality and non-seasonality of the data. Since the data exhibited seasonality Seasonal Mann-Kendall test was implemented, to capture seasonality the data was divided into three seasons commonly found in Kerala. Sen's slope was used because it is a powerful tool against outliers to capture the long-term trend, and moving average analysis to analyze the trend throughout the past and future values over time. The study concluded that a slightly increasing trend of the maximum and minimum data foreseeing climatic changes in the future, showing further factors should be considered for a broader analysis regarding this issue.

ST.TERESA'S COLLEGE (AUTONOMOUS) ERNAKULAM

Certificate of Plagiarism Check for Dissertation

Author Name	AKHILA K PRAMOD
Course of Study	M.Sc. Applied Statistics & Data Analytics
Name of Guide	Ms. ANAKHA KURIAKOSE
Department	Post Graduate Mathematics & Statistics
Acceptable Maximum Limit	20%
Submitted By	library@teresas.ac.in
Paper Title	TIME SERIES ANALYSIS ON SURFACE TEMPERATURE IN ERNAKULAM
Similarity	0% AI 2%
Paper ID	1685669
Submission Date	2024-04-23 13:06:02

Signature of Student

Signature of Guide

Checked By
College Librarian

ST. TERESA'S COLLEGE LIBRARY ERNAKULAM

* This report has been generated by DrillBit Anti-Plagiarism Software

CONTENTS

CERTIFICATE ………………………………………………………ii

DECLARATION ………………………………………………………iii

ACKNOWLEDGEMENT ………………………………………………iv

ABSTRACT ……………………………………………………………v

CONTENTS …………………………………………………………vi

LIST OF FIGURES …………………………………………………viii

LIST OF TABLES ……………………………………………………ix

1. INTRODUCTION **1**

1.1 ABOUT THE DATA 1

1.2 OBJECTIVES 2

2. LITERATURE REVIEW **3**

3. MATERIALS AND METHODOLOGY **6**

3.1 TIME SERIES ANALYSIS 6

3.1.1 ARIMA 7

3.1.2 SARIMA 7

3.1.3 ACF & PACF 8

3.1.4 AIC 8

3.1.5 ADF TEST 9

3.1.6 SEASONAL MANN-KENDALL TEST 9

3.1.7 SEN'S SLOPE 9

3.1.8 ROLLING AVERAGE ANALYSIS 10

3.2 SOFTWARE USED 10

3.3 MODEL EVALUATION 11

4. ANALYSIS **12**

4.1 MAXIMUM TEMPERATURE

4.1.1 TIME SERIES GRAPH 12

4.1.2 SEASONAL DECOMPOSITION 13
4.1.3 ADF TEST 13
4.1.4 ACF AND PACF 14
4.1.5 SARIMA MODEL & AIC VALUES 14
4.1.6 DIAGNOSTIC CHECK 16
4.1.7 IN SAMPLE FORECAST 18
4.1.8 FORECASTING 19
4.1.9 SEASONAL M-K TEST & SEN'S SLOPE 20
4.1.10 ROLLING AVERAGE 21
4.1.11 MSE & RMSE 21
4.2 MINIMUM TEMPERATURE 22
4.2.1 TIME SERIES GRAPH 22
4.2.2 SEASONAL DECOMPOSITION 22
4.2.3 ADF TEST 23
4.2.4 ACF AND PACF 23
4.2.5 SARIMA MODEL & AIC VALUES 24
4.2.6 DIAGNOSTIC CHECK 25
4.2.7 IN SAMPLE FORECAST 27
4.2.8 FORECASTING 28
4.2.9 SEASONAL M-K TEST & SEN'S SLOPE 29
4.2.10 ROLLING AVERAGE 30
4.2.11 MSE & RMSE 30
5. CONCLUSION **31**
6. REFERENCE **32**

LIST OF FIGURES

FIG NO.	DESCRIPTION	PG NO.
4.1	MAXIMUM TEMPERATURE	
Fig 4.1.1	TIME SERIES GRAPH	12
Fig 4.1.2	SEASONAL DECOMPOSITION	13
Fig 4.1.3	ACF & PACF	14
Fig 4.1.4	DIAGNOSTIC CHECK	16
Fig 4.1.5	IN SAMPLE FORECAST	18
Fig 4.1.6	FORECASTING	20
Fig 4.1.7	ROLLING AVERAGE	21
4.2	MINIMUM TEMPERATURE	
Fig 4.2.1	TIME SERIES GRAPH	22
Fig 4.2.2	SEASONAL DECOMPOSITION	22
Fig 4.2.3	ACF & PACF	23
Fig 4.2.4	DIAGNOSTIC CHECK	25
Fig 4.2.5	IN SAMPLE FORECAST	27
Fig 4.2.6	FORECASTING	29
Fig 4.2.7	ROLLING AVERAGE	30

LIST OF TABLES

TABLE NO.	DESCRIPTION	PG NO.
4.1	MAXIMUM TEMPERATURE	
Table 4.1.1	SARIMA MODEL & AIC VALUES	14
Table 4.1.2	COEFFICIENTS	15
Table 4.1.3	IN SAMPLE FORECAST	18
Table 4.1.4	FORECASTING	19
Table 4.1.5	MSE & RMSE	21
4.2	MINIMUM TEMPERATURE	
Table 4.2.1	SARIMA MODEL & AIC VALUES	24
Table 4.2.2	COEFFICIENTS	24
Table 4.2.3	IN SAMPLE FORECAST	27
Table 4.2.4	FORECASTING	28
Table 4.2.5	MSE & RMSE	30

CHAPTER-1

INTRODUCTION

"Time series analysis on Surface Temperature in Ernakulam" is a study of prime importance in today's era where global warming, severe heatwaves, droughts, and floods have become major challenges over the years. This project is an effort made to predict future values and to analyze the trend of the minimum and maximum temperatures in Ernakulam.

Surface temperature refers to the temperature on Earth's surface. Understanding surface temperature is essential for the prediction and analysis of climate change, and environmental impacts, and to come up with effective mitigation steps to overcome the challenges. A rise in the surface temperature may lead to issues like global warming, severe heatwaves, droughts, etc...

Kerala is known for its Tropical monsoon climate, where the state experiences climate variability throughout the year. Namely during the summer season, the state experiences high temperatures ranging from 35°C or more corresponding to the region, while humidity remains constant throughout the monsoon season. But during the past few years, there has been a significant increase in the temperature and a decrease in the rainfall during the summer as well as the monsoon seasons in many regions leading to droughts and heatwaves being reported.

Climate variability cannot be concluded only with the maximum and minimum temperatures but there are many other factors like spatial and temporal patterns through which, a specific conclusion can be obtained. Still understanding the maximum and minimum temperature is important to help the government or any other individuals to take mitigation steps to avoid the severe climate variabilities.

1.1 ABOUT THE DATA

The data for the study is collected from the website of the Indian Meteorological Department (IMD)[https://mausam.imd.gov.in/] which contains the monthly data of the maximum and minimum temperatures of the station Kochi (43356) during the period from January 2010 to December 2023.

1.2 OBJECTIVES

The main objective is as follows:

1. To predict the future maximum and minimum temperature values using SARIMA in Ernakulam.
2. To analyze the trend using the Seasonal Mann-Kendall test and Sen's Slope.
3. To analyze the long-term trend using the Rolling (Moving) Average.

CHAPTER-2

LITERATURE REVIEW

This chapter includes the project-related findings on Surface Temperature, where different techniques and analyzing tools, are used for the prediction of climate variabilities in different regions is founded.

Hingane et al. (1985) analyzed surface air temperature variation in India over the period 1901-1982. Collecting data from 73 widespread stations all over India, seasonal and annual temperature anomalies were obtained indicating variation over different parts of the country. The long-term trends were analyzed by a linear trend. The result reveals an evident warming trend in the mean annual temperatures of the West coast, interior peninsula, north-central, and northeast regions. The warming trend is caused by the post-monsoon and winter seasons. The post-1940 cooling in the Northern Hemisphere is not evident, suggesting a peculiar climatic response.

Kothawale and Rupa (2005) from previous studies analyzed that there are significant deviations in warming trends over India, solely contributed by the maximum temperatures. Utilizing data up to 2003 and focused mainly on the last three decades revealed noticeable changes in the temperature trends. All India's mean annual temperature shows a significant warming trend around the 1901-2003 period. Furthermore, rising trends are shown during the monsoon seasons too, resulting in reduced seasonal temperature, and both daytime and nighttime show evident warming trends.

Gopakumar et al. (2011) examined a significant shift in the rainfall and temperature trends in Kerala, where southwest monsoon rainfall in Kerala is declining and an increase in post-monsoon season is seen. Despite Kerala being classified as a heavy rainfall zone, it's been changing into dryness within the humid climates. It is known by examining that warming trends are significantly increasing. Coastal areas like Pampadumpara show an increase in rainfall and temperature, while areas like Ambalavayal showcase a declining trend. Hence the study concludes the vulnerability of plantation crops due to climate variability (floods and droughts), rather than long-term changes in temperature and rainfall.

Subash and Sikka (2014) investigated the trends in rainfall and temperature and, also the possibility of having any rational relationship between the trends across the homogeneous regions of India. It showcases a substantial increase in the annual temperature and annual rainfall over the homogeneous regions except the North East. Except in October, monthly analysis reveals no notable relationship between increasing rainfall and maximum temperature trends where seasonal patterns are concerned over meteorological subdivisions of India. Despite this, it can be concluded that the significant spatial and temporal dependence implicit relationship between rainfall and temperature trends.

Basha et al. (2017) examined the factors influencing Surface Temperature (ST) variations across India obtained from the Coupled Model Inter-comparison phase 5 (CMIP5) during the 20th and 21st centuries. In the 20th century, Greenhouse Gases (GHH) and Land Use (LU) became the major factors affecting global warming whereas Anthropogenic Aerosols (AA) hindered the warming rate. By the end of the 21st century, Representative Concentration Pathways (RCP) 8.5 shows a dramatic increase. The study shows an upsurge in ST in the 21st century during the summers up to the year 2050. The analysis finds extreme heat and cold events under RCP2.5 for the period 2080-2099 relative to the 1986-2006 base period. Nevertheless, the study suggests that controlling regional AA and LU can further reduce the warming across India.

Ross et al. (2018) study utilizes a new thorough surface temperature dataset of India, covering seven decades to examine the impact of global warming. To analyze the temperature patterns data is segmented into pre-monsoon, monsoon, and post-monsoon. The mean of minimum, maximum, and daily mean temperatures over the 2000s are compared to those of the 1950s revealing a constant warming trend across northwestern and southern India, while northeastern and central India showcases a cooling trend. The potential impact of these temperature patterns on agriculture is considered.

Arulbalaji et al. (2020) considered rapid urbanization and unsustainable developments to lead to significant environmental threats, especially in areas that are densely populated, tropical and subtropical regions. This paper examined the spatiotemporal dynamics and associated changes in the Land Surface Temperature (LST) in Thiruvananthapuram. The dataset contains Landsat imageries from 1988,2000, and 2019, where the study reveals a swift decrease in barren land and vegetation cover, including an increase in the built-up area. An increasing trend in LST indicates a rapid mitigation process to extend the green cover in these areas and to consider environment-friendly developments.

Kabbilawsh et al. (2020) proposed a temperature forecasting model using Seasonal Autoregressive Integrated Moving Average (SARIMA), for the state of Kerala. Where the mean maximum and mean minimum temperature data spanning 47 years collected from seven stations were analyzed. The data displays seasonality and trend due to the data's five-decade span. Hence the key concept is to determine the non-stationarity of data and to transform it into stationary data, this is done using Seasonal and Trend decomposition using Loess technique and Kwiatkowski-Phillips-Schmidt-Shin test. The conversion of non-stationarity to stationary is dealt with seasonal and first-order differencing. Eight out of the fourteen time series datasets revealed SARIMA as the ideal forecasting model.

Vijay and Varija (2022) address a 119-year spatiotemporal assessment of climate variability in Kerala, with gridded datasets of daily rainfall and temperature from various stations across Kerala. Using the Mann-Kendall test and its modified versions, indicating negative trends. More than 83% of the grid points indicate a declining trend over the mean annual and southwest monsoon seasons. Both annual and seasonal show an increasing trend in the mean minimum and maximum temperatures. K-means clustering reveals vulnerability to water scarcity in non-rainy seasons.

Valappil et al. (2023) study utilizes Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) data over a period of 40 years (1980-2019) used to analyze the trend patterns in daily average temperature across Kerala, India. Statistical tests like Mann-Kendall (MK), Spearman's Rho (RS), Pettitt tests, and Theil-Sen's method reveal significant trends in daily average temperatures. The study indicates an increase of 1.35℃, Among the stations Pathanamthitta, Idukki, and Kollam districts show a higher rate of increase in the daily average temperature. The study concludes that high attention should be given where the region's vulnerability to drought and agriculture should be considered.

CHAPTER-3

MATERIALS AND METHODOLOGY

3.1 TIME SERIES ANALYSIS

Time series analysis is an important statistical method that collects data points at regular intervals to discover underlying patterns and trends in the data. This method is widely used in various industries like financial, economic, medical, weather, etc… as it encourages making informed decisions and accurate predictions based on historical data. In time series analysis two types of time series data can be looked up to stationary and non-stationary. For the smooth processing of data, they should be stationary because stationary data is stable and has a constant mean, variance, and trends. If the data encountered is non-stationary it won't be stable, the data comes in contact is of seasonal, trend, and irregular fluctuations. This can be eliminated using differencing which in turn makes the data stationary for effective modeling.

The major four components of the Time series are:

- **TREND:** Trend is the long-term movements or changes that are found in the data. The trend can be linear, increasing, decreasing, stable, and non-linear over time. Thus understanding the trend of data is necessary because it can give a basic understanding of underlying patterns in the data.
- **SEASONALITY:** It refers to the periodic fluctuations that happen periodically i.e., daily, weekly, monthly, or yearly. Seasonality can be influenced by many external factors according to the data taken. Whenever time series analysis is considered techniques such as seasonal decomposition becomes an essential step to remove the seasonality from the data for a better processing of the data. Seasonality shows the overall behavioral pattern and underlying trends in the data.
- **CYCLICALITY:** Unlike seasonality which has a fixed period, cyclicality refers to the periodic fluctuations in the data which does not have a fixed period but shows a regular pattern over time. Business and economic fields undergo cyclic fluctuations, and identifying cyclicality in the business cycles can help identify the underlying social and economic dynamics.
- **RANDOM (IRREGULAR) FLUCTUATIONS:** These are the irregular fluctuations that happen rarely or randomly over time, these don't have any

seasonality, trend, or cyclicality but still can provide valuable pieces of information through time series modeling and forecasting. Examples of irregular fluctuations are earthquakes, cyclones, etc…

Different types of Time series models include ARIMA (Autoregressive Integrated Moving Average) and seasonal ARIMA model (SARIMA) which are useful in analyzing and forecasting data with seasonal patterns.

3.1.1 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)

ARIMA model is a powerful and widely used statistical tool along the time series data.

Autoregressive (AR) component: which considers the difference between the past and current observations. i.e., the effect of previous observation in the current data, where ‘p’ is the parameter

Integrated (I) component: ‘d’ being the parameter that represents the number of differencing needed for the data to achieve stationarity.

Moving Average (MA) component: It is the linear combination of the current data with the errors (residuals) of past data. Where ‘q’ is the parameter.

3.1.2 SEASONAL ARIMA (SARIMA)

SARIMA is the extension of ARIMA (Autoregressive Integrated Moving Average), which captures the seasonal pattern or variations in the data. The parameter and model for SARIMA are represented as (p, d, q) × (P, D, Q, s), s being the seasonal interval (mainly the value for s is 12, 24, etc…), (P, D, Q) the seasonal AR, MA, and degree of seasonal differencing. Grid search method, and information criteria (AIC, BIC) are typical methods used to select the best model. Components of SARIMA,

Seasonal AutoRegressive (SAR) Component: It captures the seasonality of the data, by comparing the relationship between current and past observations at the same lag. SAR component denoted as ‘P’

Seasonal Integrated (SI) Component: Like the integrated component this is also used to differentiate the data to stationary. It is denoted as ‘D’.

Seasonal Moving Average (SMA) Component: This is the relationship between the current observations and the past errors, usually denoted as 'Q'.

Hence the SARIMA model is the combination of seasonal and non-seasonal components, Therefore the equation for the SARIMA model is **(p, d, q) × (P, D, Q, s)**. By comprising seasonal components it extends the shortcomings of ARIMA models to capture the seasonal patterns, making it well-known among researchers and analysts in multiple fields.

3.1.3 ACF & PACF:

The Autocorrelation Function (ACF) is the correlation between each observation with its lag values within the same series.ACF has lag **'k'** which is the number of time periods between the observations being compared. The function of the lag k is denoted as $\mathbf{p_k}$. By identifying the number of lags outside the range ACF can be used to determine the order of the moving average components (q). Seasonality in the data can be derived from the ACF plot from multiples of the seasonal period.

Whereas the Partial Autocorrelation function (PACF) is the correlation between observations after excluding the effects of in-between observations or intermediate lags. The PACF plot determines the order of the autoregressive component (p). The PACF at lag k indicates a direct relationship between the lags after removing the effects of the short lags.

Combining both ACF and PACF can provide insights into the underlying patterns and choose an appropriate model for analysis and forecasting.

3.1.4 AIC (AKAIKE INFORMATION CRITERION)

Akaike Information Criterion is a popular statistical tool used to select the best-fit model. AIC avoids overfitting by giving a good fit. A good model is selected based on the lower AIC values. Mainly it is a comparison model.

The formula for AIC is:

$AIC = 2k - 2\ln(\hat{L})$

Where

k is the number of parameters.

$L^\wedge$ is the maximum value for the likelihood function.

3.1.5 AUGMENTED DICKEY-FULLER (ADF) TEST

ADF test is used to check whether the data is stationary or not. To check stationarity the hypotheses taken are:

H_0: Time series data is non-stationary

H_1: Time series data is stationary.

If p-value > 0.05, we fail to reject H_0.

This is used in complex trends. Stationarity is an important assumption in time series. ADF test is one of the crucial steps to find stationarity.

3.1.6 SEASONAL MANN-KENDALL TEST

The seasonal Mann-Kendall (Seasonal M-K) test is a non-parametric test mainly used to detect the trend in time series data. The Seasonal Mann-Kendall test is an augmentation to the traditional Mann-Kendall test, used to check whether the data statistically estimates the upward or downward trend.

The seasonal M-K test finds the monotonic trend in the data by comparing the ranks of the data points. Time series are grouped into different seasons such as weeks, months, or years. Then the Mann-Kendall test is applied to the sub-grouped seasons, and the significance of the trend is then calculated for each season.

3.1.7 SEN'S SLOPE

Theil-Sen estimator also known as Sen's Slope is a non-parametric test. This is commonly used in calculating the magnitude of trends in time series data. Sen's slope is sturdy to outliers and does not require the data to follow any specific distributions. Hence, making it appropriate for long-term data with irregular fluctuations. Sen's slope is more robust than traditional linear regression techniques. It is commonly used in trend analysis for environmental and hydrological studies where long-term trends, and underlying patterns are present.

3.1.8 ROLLING AVERAGE ANALYSIS

Rolling (Moving) average analysis is a statistical technique utilized to defuse variations or fluctuations in the time series data. A specific time interval or window size is given to identify the patterns and eliminate noise from the data. The window size depends upon the data and its time, for a monthly data 12 period can be the starting size. Moving averages smooth out short-term fluctuations in the data to obtain informed decisions hinged on underlying patterns or trends.

The rolling average analysis is to be similar to other types of moving averages such as simple, weighted, and exponential moving averages. However, the Rolling average is considered to be more flexible and adaptable since it allows adjustments in the window size as per the requirements.

3.2 SOFTWARE USED

Python is a widely recognized object-oriented, high-level programming language, known for its easy-to-use interface, modules, and packages. It is commonly used for data analysis, web development, machine learning, etc…

Here python is put into action for performing SARIMA which helps recognize the seasonal patterns and forecast future values. Seasonal Mann-Kendall test, Sen's Slope, and Moving Average Analysis are used to identify patterns, make informed decisions, and mitigate risks.

Excel is used to perform Data preprocessing steps, Excel allows users to import data to various sources. In this study, Excel was used to find and replace the null values, and to format Dates.

3.3 MODEL EVALUATION

Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are the most popular techniques used in statistics and machine learning for evaluating the performance of the models. The accuracy and reliability of the model prediction can be obtained using these methods. Lower MSE and RMSE values indicate a better model performance which in turn can be used for choosing a final model.

CHAPTER-4

ANALYSIS

'TIME SERIES ANALYSIS ON SURFACE TEMPERATURE IN ERNAKULAM' is a detailed analysis of the data collected from the IMD website of the maximum and minimum surface temperature in Ernakulam which spans 13 years i.e., from January 2010 to December 2023. The results, and discussions procured after the study are enclosed in this chapter.

Here, the chapter is divided into two parts maximum temperature and minimum temperature. Under each heading its detailed analysis and results are mentioned.

4.1 MAXIMUM TEMPERATURE

4.1.1 Time series graph

Fig 4.1.1 shows the time series graph of Maximum temperatures in Ernakulam from June 2010 to December 2023.

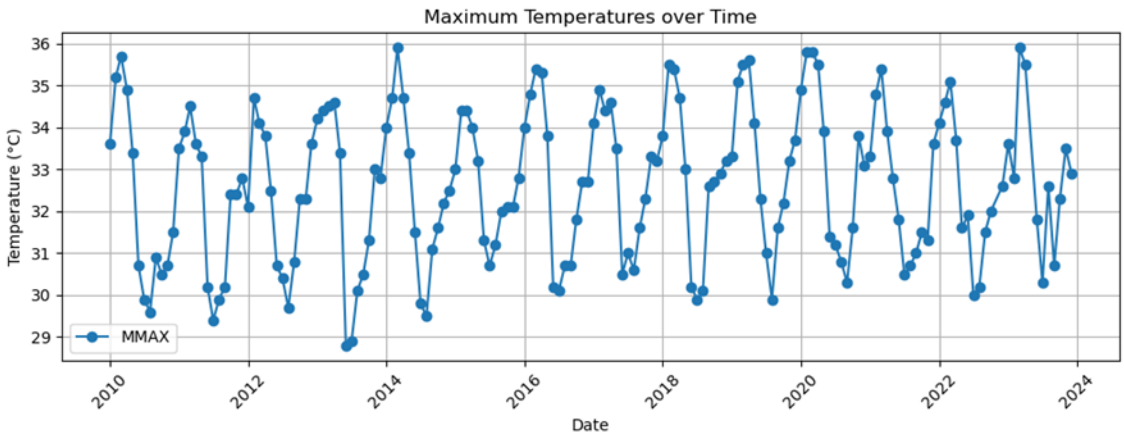


Fig 4.1.1

4.1.2 Seasonal Decomposition

The graph below shows seasonality and trend

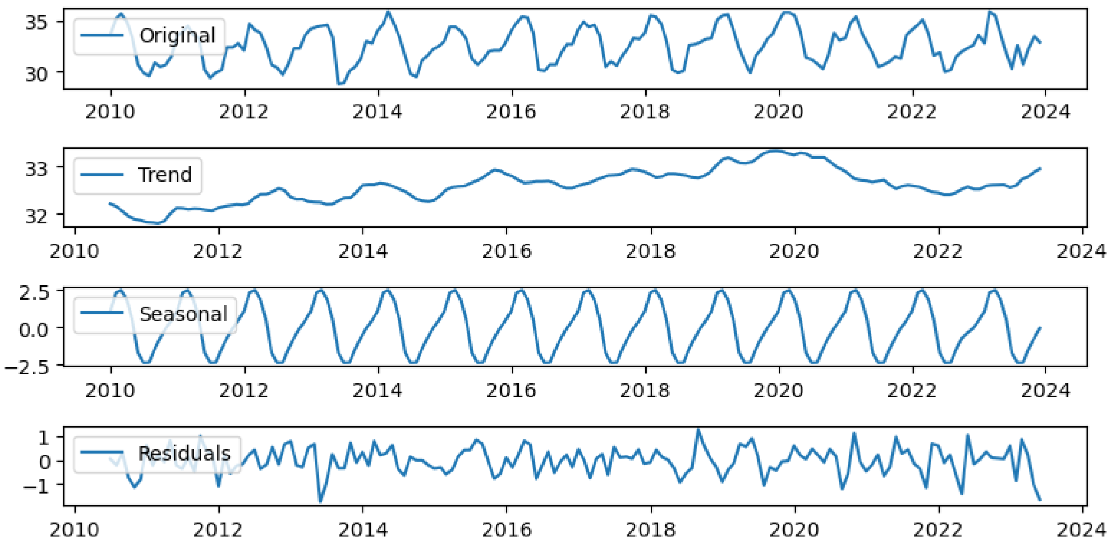


Fig 4.1.2

4.1.3 Augmented Dickey-Fuller test

The ADF test is done to check whether the data is stationary or not. For that generate a hypothesis,

H_0**:** The data is non-stationary

H_1**:** The data is stationary

The result obtained is,

ADF Statistic: -1.716599735688399

p-value: 0.4225807253061309

Here **p value > 0.05**

i.e., we fail to reject the null hypothesis.

Hence the data is non-stationary.

4.1.4 ACF AND PACF

The ACF and PACF graph of the maximum temperature is given below in Fig 4.1.3

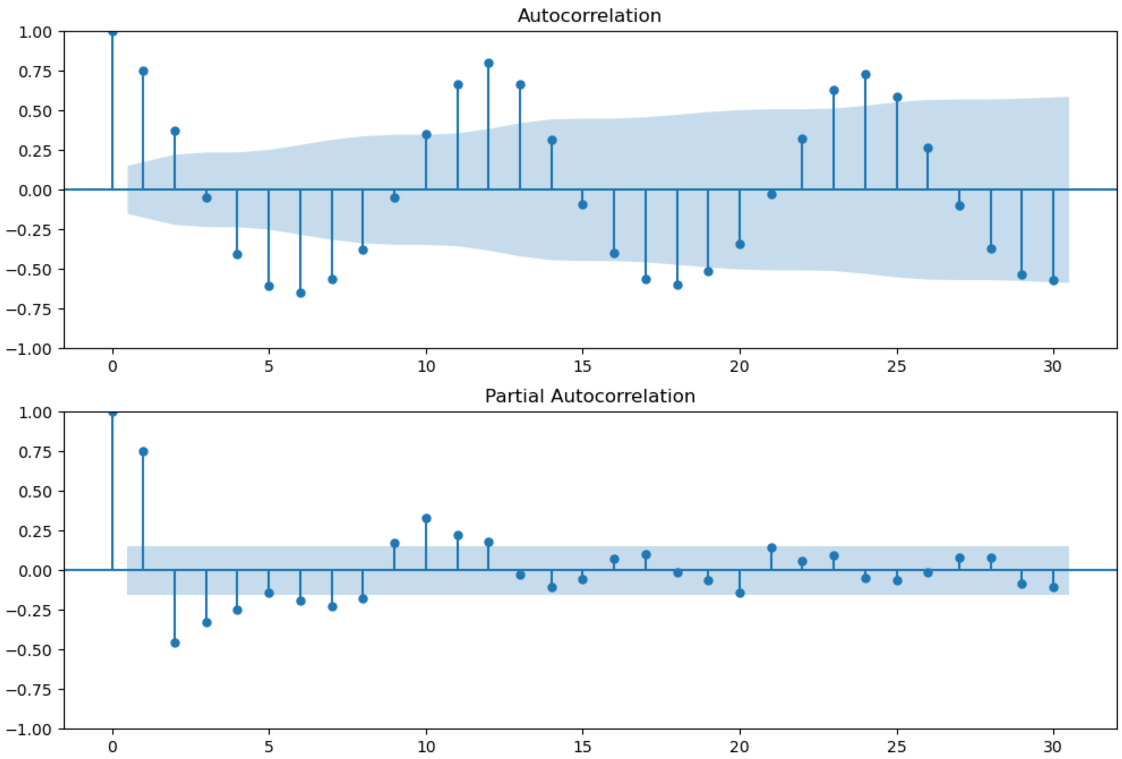


Fig 4.1.3

From the above figure, it is clear that an ACF of lag 16 and PACF of lag 10 is obtained.

4.1.5 SARIMA Models and AIC Values

The lag values obtained for the ACF and PACF were high due to this complexity to make the process easier grid search method is used. To find the best-fit model the corresponding lowest Akaike Information Criterion (AIC) from all the possible SARIMA models is considered. On that account, some of the SARIMA models with their corresponding AIC values are given below in Table 4.1.1

SL NO.	ARIMA (p,d,q) × $(P,D,Q)_{12}$	AIC
1.	(0, 1, 0) × (0, 0, 1)	468.482
2.	(0, 1, 0) × (0, 1, 1)	357.141
3.	(0, 1, 0) × (0, 1, 2)	336.293

4.	(0, 1, 1) × (1, 1, 1)	367.079
5.	(0, 1, 0) × (1, 1, 2)	336.293
6.	(0, 1, 0) × (2, 1, 2)	334.061
7.	(0, 1, 1) × (0, 1, 0)	408.098
8.	(0, 1, 1) × (0, 1, 1)	308.454
9.	(0, 1, 1) × (1, 1, 1)	314.945
10.	(0, 1, 1) × (2, 0, 1)	320.741
11.	(0, 1, 1) × (2, 1, 2)	291.711
12.	(0, 1, 2) × (0, 1, 0)	398.224
13.	**(0, 1, 2) × (1, 1, 2)**	**281.132**
14.	(0, 1, 2) × (1, 1, 0)	343.951
15.	(1, 1, 2) × (0, 1, 1)	301.488

Table 4.1.1

The best model is ARIMA (0, 1, 2) × (1, 1, 2)$_{12}$ with AIC Value 281.132

Coefficients:

	Coef	Std err	z	P > |z|	[0.025	0.975
ma.L1	-0.8073	0.128	-6.283	0.000	-1.059	-0.556
ma.L2	-0.2616	0.109	-2.396	0.017	-0.476	-0.048
Ar.S.L12	-0.5658	0.193	-2.928	0.003	-0.945	-0.187
ma.S.L12	-0.2417	0.227	-1.065	0.287	-0.687	0.203
ma.S.L24	-0.5418	0.207	-2.618	0.009	-0.947	-0.136
Sigma2	0.4080	0.049	8.335	0.000	0.312	0.504

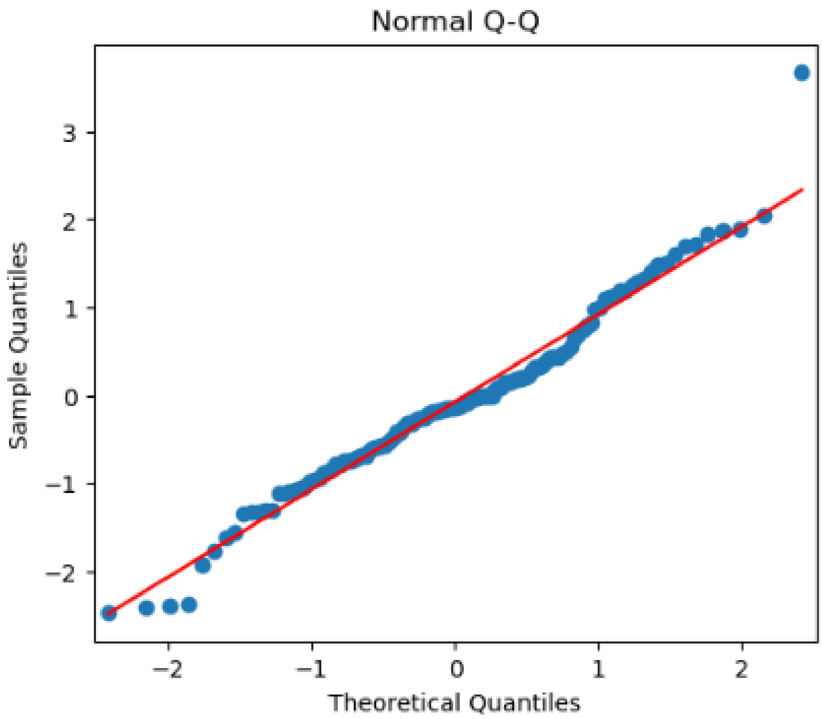
Table 4.1.2

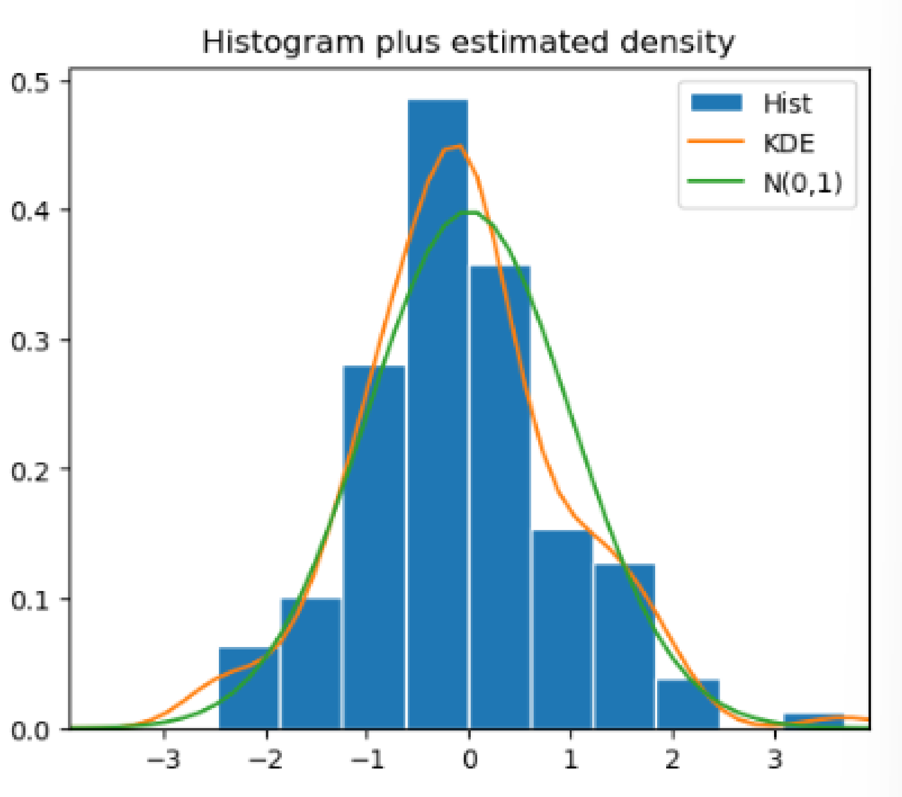
From the above table ma.S.L12 has a p-value greater than 0.05 indicating their statistical significance is less compared to the other lags which are highly statistically significant.

4.1.6 Diagnostic check

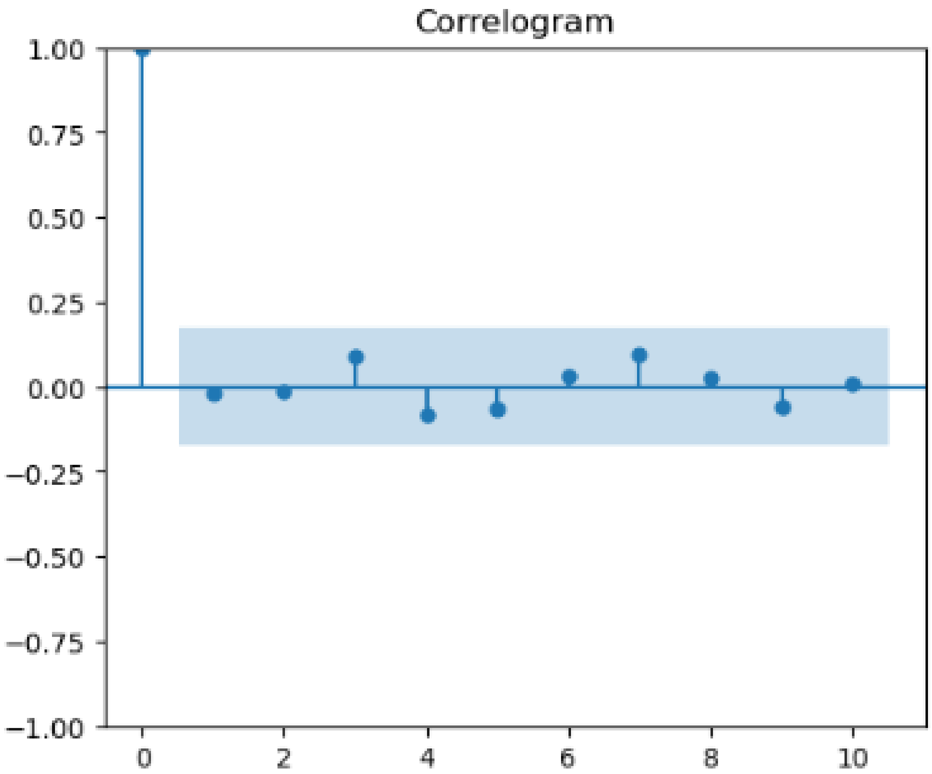
Diagnostic checking is essential to ensure the reliability, validity, and effectiveness of the statistical model selected. Given below are the different diagnostic processes. Different diagnostic plots are given below in Fig 4.1.4

In a Normal Q-Q plot, it can be seen that most of the data points are aligned close to the diagonal line suggesting the residuals are normally distributed. In the next figure, the normal curve and kernel density estimation (KDE) curve are closely matched to each other with the histogram being aligned along it smoothly indicating a normal distribution. In correlogram, it visualizes the correlation between time series and lagged values. Here the values within the range indicate that the values that lie within the range are statistically significant.





Histogram plus estimated density
Hist
KDE
N(0,1)
0.5
0.4
0.3
0.2
0.1
0.0
−3
−2
−1
0
1
2
3



Correlogram
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
2
4
6
8
10

Fig 4.1.4

4.1.7 In sample Forecast

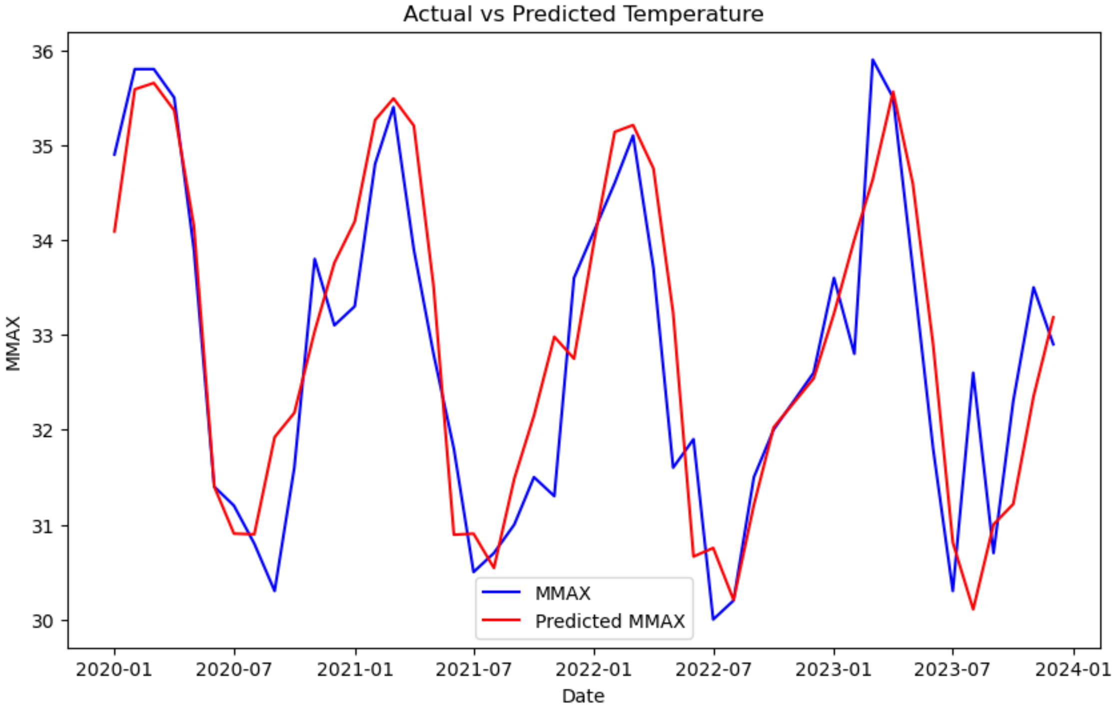


Fig 4.1.5

The in-sample forecast values for the year 2023 are given below in table 4.1.3

Date	MMAX	Pred_MMAX
JAN 2023	33.6	33.225
FEB 2023	32.8	34.004
MAR 2023	35.9	34.637
APR 2023	35.5	35.562
MAY 2023	33.7	34.595
JUN 2023	31.8	32.893
JUL 2023	30.3	30.809
AUG 2023	32.6	30.106
SEP 2023	30.7	31.001
OCT 2023	32.3	31.216
NOV 2023	33.5	32.352
DEC 2023	32.9	32.184

Table 4.1.3

4.1.8 Forecasting

The forecasted values of maximum temperature for the next 2 years i.e., from January 2024 to December 2025 are given below

Date	Forecast values	LCL	UCL
JAN 2024	33.320	31.994	34.646
FEB 2024	33.723	32.339	35.108
MAR 2024	35.312	33.925	36.699
APR 2024	35.446	34.057	36.836
MAY 2024	34.600	33.207	35.992
JUN 2024	33.159	31.764	34.554
JUL 2024	31.161	29.764	32.559
AUG 2024	31.015	29.615	32.416
SEP 2024	30.634	29.231	32.037
OCT 2024	31.644	30.238	33.049
NOV 2024	32.460	31.052	33.868
DEC 2024	32.851	31.440	34.262
JAN 2025	33.434	31.980	34.887
FEB 2025	33.936	32.473	35.399
MAR 2025	35.297	33.830	36.763
APR 2025	35.484	34.014	36.955
MAY 2025	34.716	33.242	36.190
JUN 2025	33.260	31.782	34.738
JUL 2025	31.294	29.812	32.775
AUG 2025	30.860	29.375	32.345
SEP 2025	30.655	29.167	32.144
OCT 2025	31.617	30.125	33.109
NOV 2025	32.403	30.907	33.898
DEC 2025	32.954	31.455	34.454

Table 4.1.4

The graphical representation of the maximum temperature is shown in Fig 4.1.6

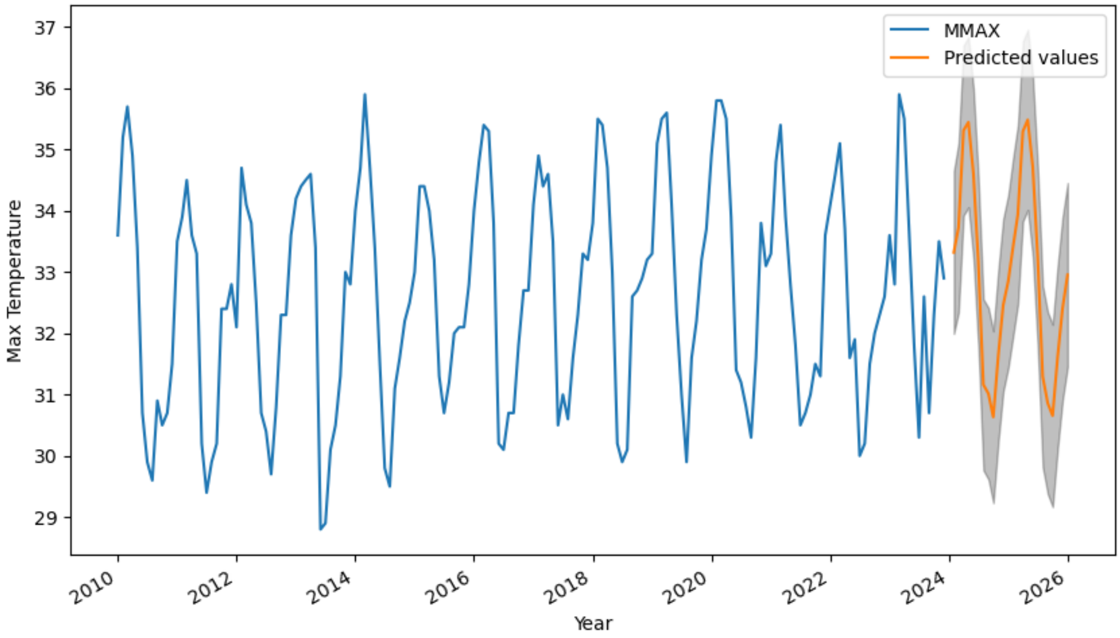


Fig 4.1.6

4.1.9 Seasonal Mann-Kendall Test & Sen's Slope Results:

Seasonal Mann-Kendall Test Results:

Season Winter: Tau = **0.2128**

Season Summer: Tau = **0.2954**

Season Monsoon: Tau = **-0.0168**

For performing the seasonal mann-kendall test consider the three different climates in Kerala, namely winter, summer, and monsoon. The result exhibits a moderate increase in the trend over the summer and winter seasons while the monsoon season indicates a weak decreasing trend.

Sen's Slope: **0.0033331551775202706**

An average increase of 0.00333℃ per month and since the value is positive indicates a positive trend. Sen's slope estimates the magnitude of the trend.

4.1.10 Rolling Average

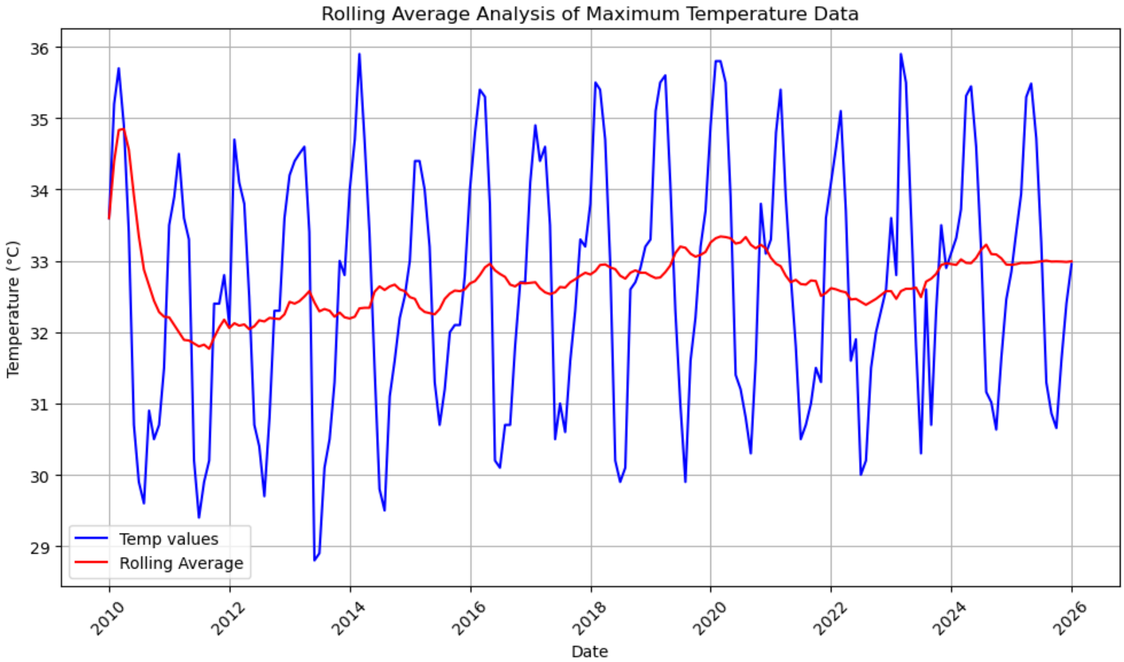


Fig 4.1.7

By smoothing out short-term variations rolling average makes it easier to recognize long-term trends and underlying patterns in the data. The above Fig 4.1.10 is the graphical representation of the rolling average over the past and forecasted values together indicating seasonal fluctuations and an increasing trend in the future.

4.1.11 MSE & RMSE

The lower mean squared and root mean squared errors indicate a better, efficient model.

MSE	**0.72**
RMSE	**0.8493**

Table 4.1.5

4.2 MINIMUM TEMPERATURE

4.2.1 Time series plot

Fig 4.2.1 represents the minimum temperature over the years January 2010 to December 2023.

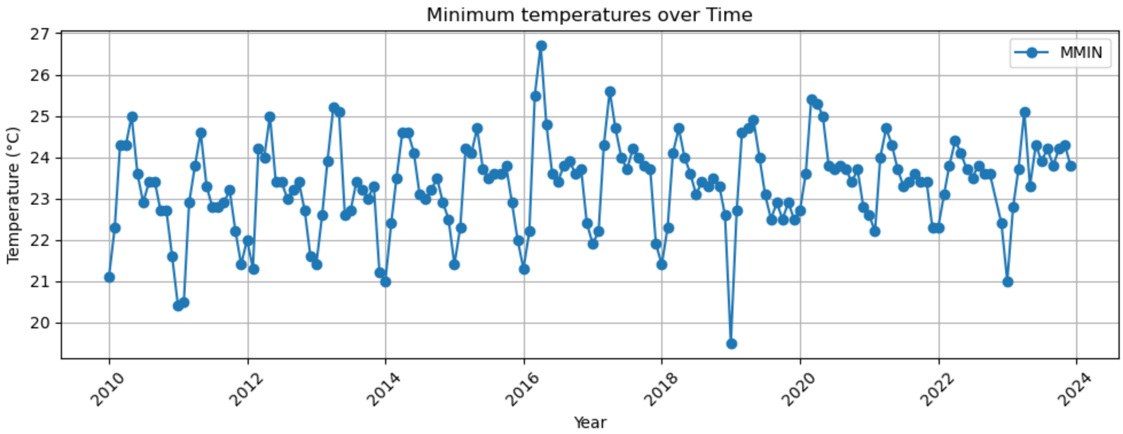


Fig 4.2.1

4.2.2 Seasonal Decomposition

From the figure below seasonality and trend are founded

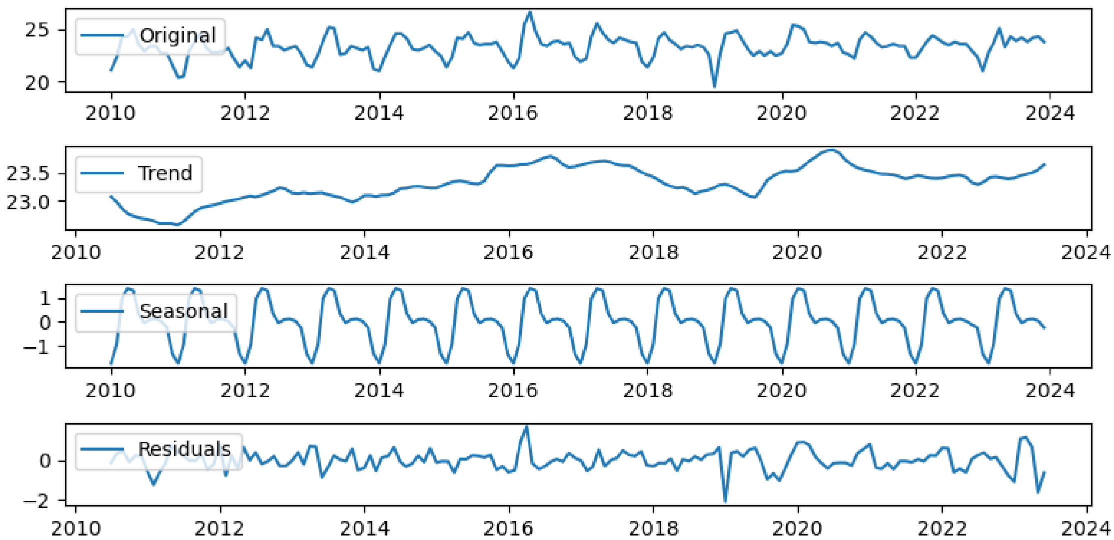


Fig 4.2.2

4.2.3 Augmented Dickey-Fuller test

Let, H_0**:** The data is non-stationary

H_1**:** The data is stationary

The result obtained is,

ADF Statistic: -2.074204486574692

p-value: 0.25501279109257136

Here **p value > 0.05**

i.e., we fail to reject the null hypothesis

Hence, the data is non-stationary

4.2.4 ACF AND PACF

From Fig 4.2.3 it seems that the ACF of lag10 and PACF of lag 6 is procured for the minimum temperature data.

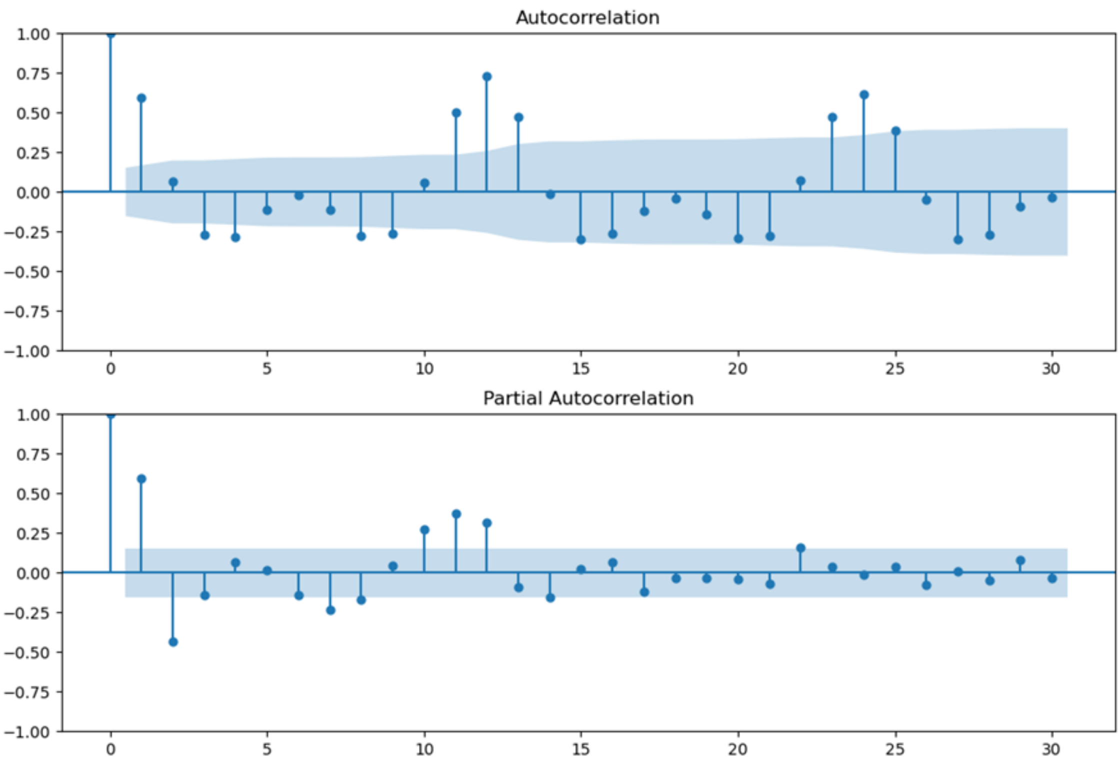


Fig 4.2.3

4.2.5 SARIMA Models and AIC Values

Here the model is decided based on the grid search method. The best-fit model is selected in correspondence with the lowest Akaike Information Criterion (AIC).

Therefore, the SARIMA models with AIC values are shown in Table 4.2.1

SL NO.	ARIMA (p,d,q) × $(P,D,Q)_{12}$	AIC
1.	(0, 0, 1) × (0, 0, 2)	969.55
2.	(0, 0, 1) × (0, 1, 2)	243.27
3.	(0, 0, 2) × (0, 1, 1)	259.21
4.	(0, 1, 0) × (1, 0, 1)	330.35
5.	(0, 1, 1) × (1, 1, 2)	245.49
6.	**(0, 1, 2) × (2, 1, 2)**	**234.38**
7.	(1, 0, 0) × (1, 0, 2)	314.10
8.	(1, 0, 1) × (0, 1, 2)	242.57
9.	(1, 0, 2) × (1, 1, 2)	243.41
10.	(1, 1, 0) × (0, 0, 1)	388.68
11.	(1, 1, 2) × (1, 1, 0)	290.98
12.	(0, 1, 1) × (2, 1, 0)	249.70
13.	(0, 1, 2) × (1, 1, 1)	266.85
14.	(1, 1, 1) × (2, 1, 2)	239.08
15.	(1, 1, 2) × (2, 0, 1)	264.14

Table 4.2.1

The best model is ARIMA (0, 1, 2) × $(2, 1, 2)_{12}$ with AIC Value 234.38

Coefficients:

	Coef	Std err	z	p>|z|	[0.025	0.975]
maL1	-0.7591	0.097	-7.862	0.000	-0.948	-0.570
ma.L2	-0.2844	0.079	-3.595	0.000	-0.439	-0.129
ar.S.L12	-0.2936	0.181	-1.626	0.104	-0.647	0.060

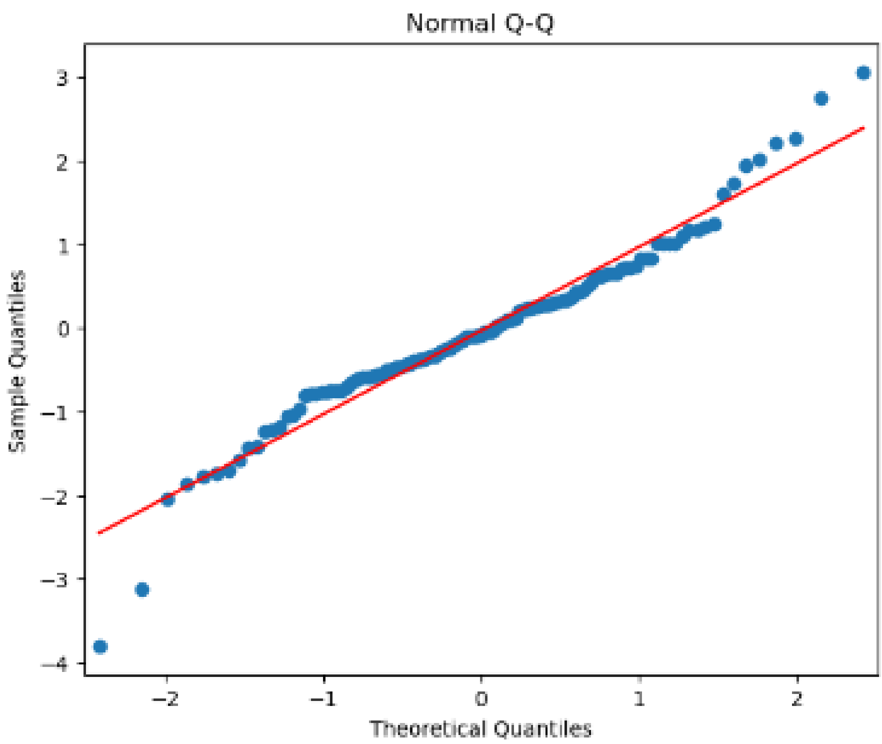
ar.S.L24	-0.3850	0.123	-3.142	0.002	-0.625	-0.145
ma.S.L12	-0.3543	0.190	-1.855	0.064	-0.727	0.020
ma.S.L24	-0.0613	0.225	-0.273	0.785	-0.501	0.379
Sigma2	0.3020	0.035	8.559	0.000	0.233	0.371

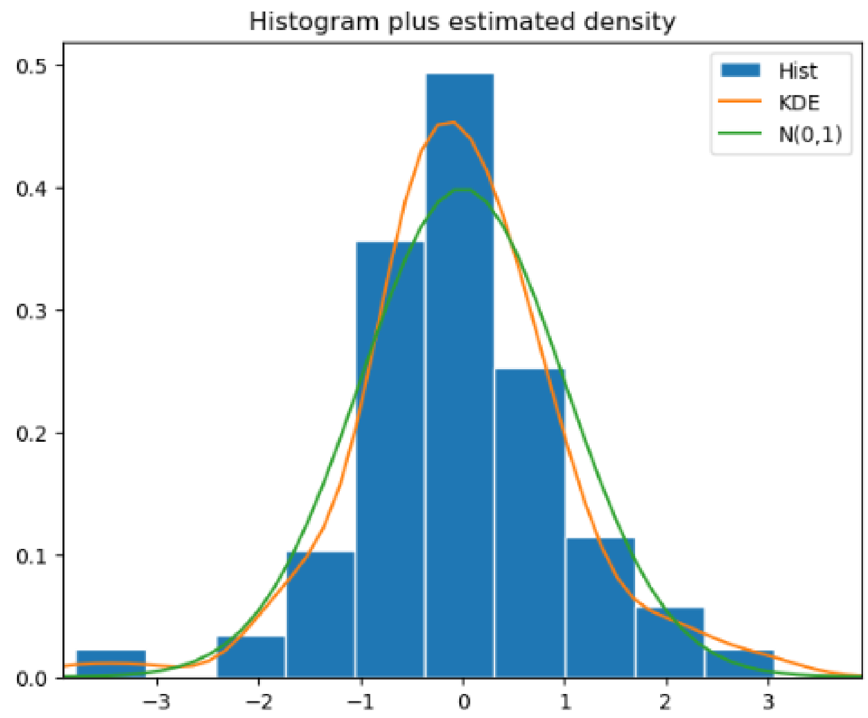
Table 4.2.2

The ar.S.L12, ma.S.L12, and ma.S.L24 are the ones that aren't statistically significant as their p-value is greater than 0.05.

4.2.6 Diagnostic Plots

Fig 4.2.4 shows the Normal Q-Q and Histogram plus estimated density graphs, Here the points are aligned close to the diagonal with some points deviating from the line indicating a deviation from the normal distribution. However in the histogram plus estimated density the normal distribution shows a bell curve aligned close to the histogram showing a smooth normal distribution. In the correlogram, the correlation between time series and lagged values is visualized. Here the values within the range indicate they are statistically significant.




Histogram plus estimated density
Hist
KDE
N(0,1)
0.5
0.4
0.3
0.2
0.1
0.0
−3
−2
−1
0
1
2
3

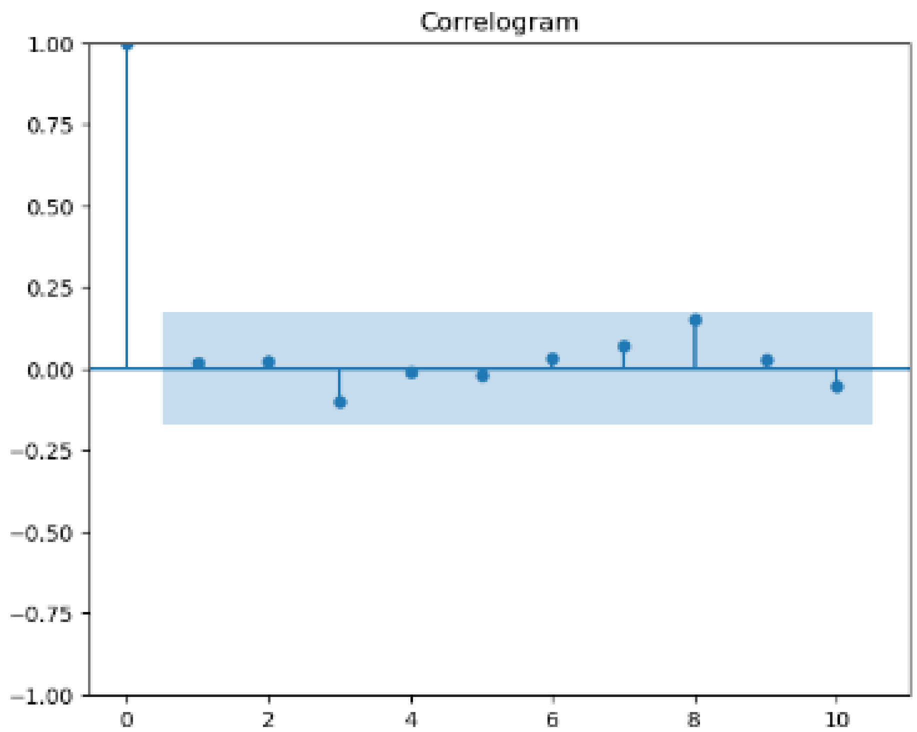

Correlogram
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
2
4
6
8
10

Fig 4.2.4

4.2.7 In Sample Forecast

The model's predictive capability against the original data is being plotted.

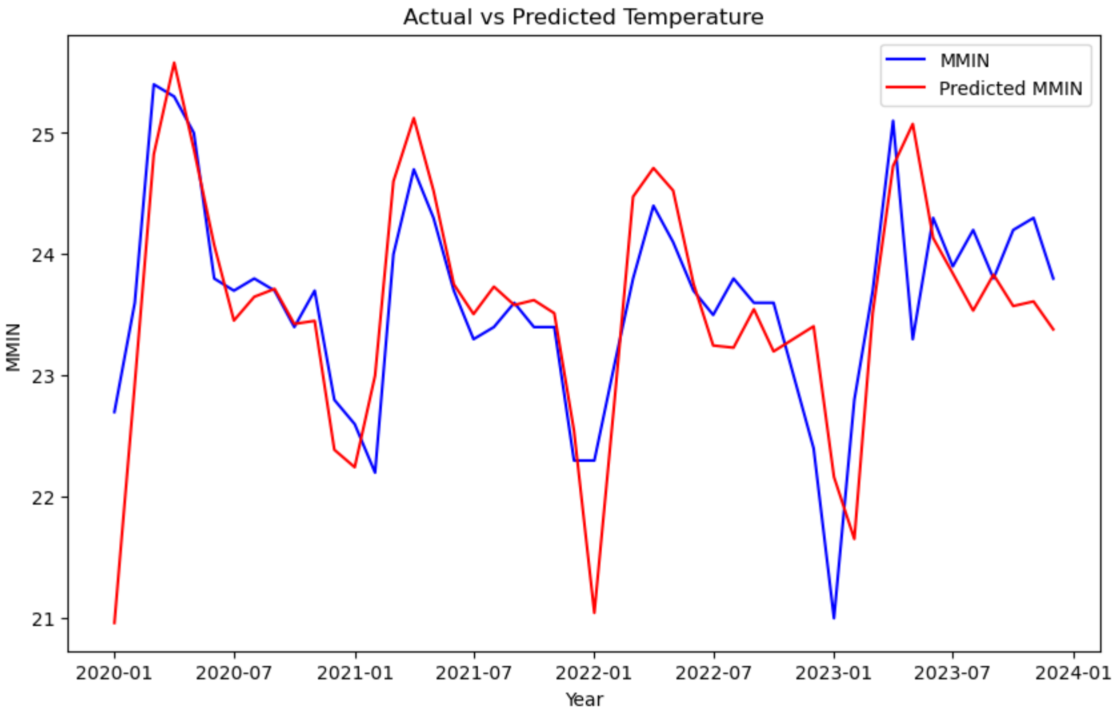


Fig 4.2.5

The in-sample forecast values for the year 2023 are shown in below Table 4.2.3

Dates	MMIN	Pred_MMIN
JAN 2023	21.0	22.164
FEB 2023	22.8	21.653
MAR 2023	23.7	23.515
APR 2023	25.1	24.727
MAY 2023	23.3	25.074
JUN 2023	24.3	24.134
JUL 2023	23.9	23.841
AUG 2023	24.2	23.536
SEP 2023	23.8	23.831
OCT 2023	24.2	23.573
NOV 2023	24.3	23.611
DEC 2023	23.8	23.381

Table 4.2.3

4.2.8 Forecasting

Table 4.2.4 contains the forecasted values from the year 2024 starting to the end of 2025.

Date	Forecast values	LCL	UCL
JAN 2024	22.301	21.188	23.415
FEB 2024	22.511	21.345	23.676
MAR 2024	23.057	21.890	24.223
APR 2024	24.770	23.602	25.937
MAY 2024	24.443	23.275	25.612
JUN 2024	24.568	23.399	25.737
JUL 2024	23.924	22.754	25.095
AUG 2024	23.771	22.600	24.942
SEP 2024	23.653	22.482	24.825
OCT 2024	23.889	22.717	25.062
NOV 2024	23.777	22.603	24.951
DEC 2024	23.692	22.517	24.867
JAN 2025	22.541	21.286	23.795
FEB 2025	22.427	21.162	23.693
MAR 2025	23.057	21.790	24.324
APR 2025	24.412	23.144	25.681
MAY 2025	24.705	23.435	25.975
JUN 2025	24.507	23.235	25.779
JUL 2025	23.909	22.636	25.183
AUG 2025	23.661	22.386	24.936
SEP 2025	23.761	22.484	25.038
OCT 2025	23.785	22.507	25.064
NOV 2025	23.681	22.401	24.960
DEC 2025	23.218	21.937	24.500

Table 4.2.4

Given below is the graphical representation of the forecasted values of minimum temperature.

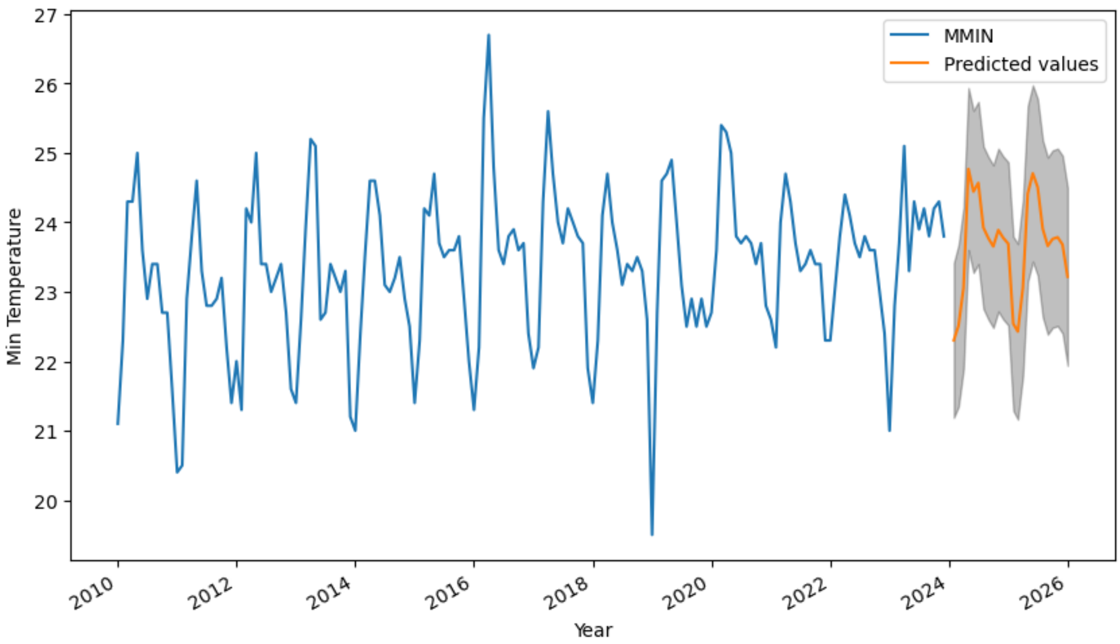


Fig 4.2.6

4.2.9 Seasonal Mann-Kendall Test & Sen’s Slope Results:

Seasonal Mann-Kendall Test Results:

Season Winter: Tau = **0.6161**

Season Summer: Tau = **-0.3150**

Season Monsoon: Tau = **0.4036**

For performing the seasonal mann-kendall test three different climates in Kerala, namely winter, summer, and monsoon are taken. The result exhibits a significant strong increase of trend over the winter season, and a moderate increase during monsoon, while the summer season indicates a moderate decreasing trend.

Sen's Slope: **0.0039999999999999915**

An average increase of 0.00399°C per month and since the value is positive indicates a positive trend.

4.2.10 Rolling Average

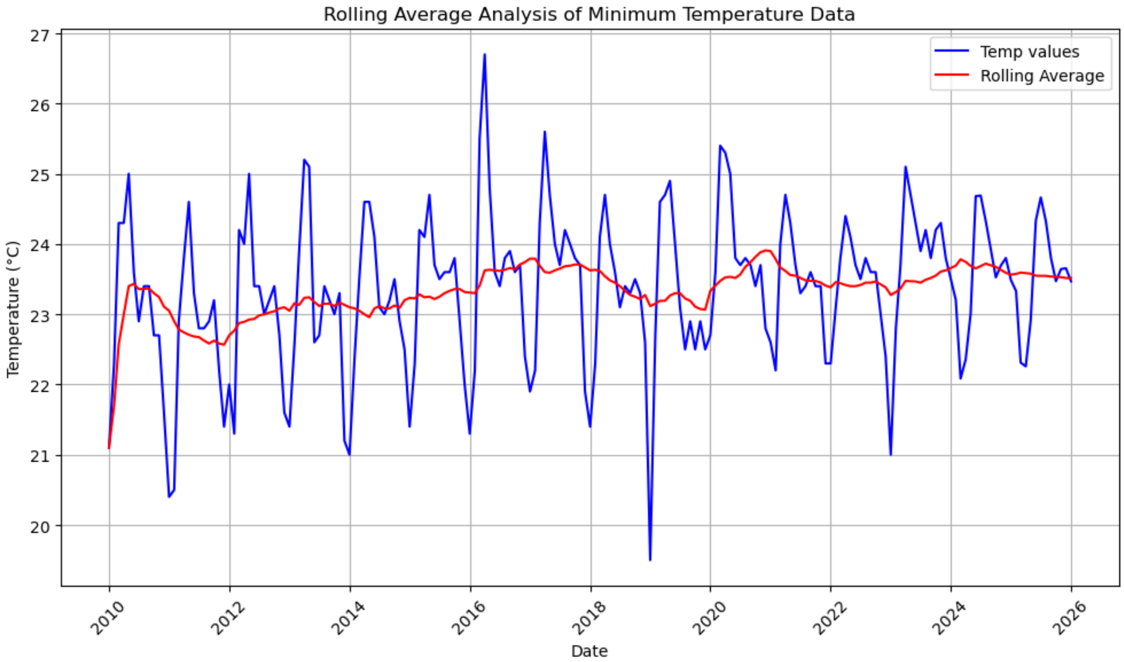


Fig 4.2.7

By defusing short-term fluctuations rolling average effortlessly exhibits a long-term trend, here a negligible increase in trend over time is obtained.

4.2.11 MSE & RMSE

The mean squared erroe and root mean squared errors are low values indicating an efficient model.

MSE	**0.37**
RMSE	**0.6083**

Table 4.2.5

CHAPTER -5

CONCLUSION

Based on the results gained after conducting the analysis it seems that there is an increase in the overall maximum and minimum temperatures over time. In detail, here in this study using the Seasonal Mann-Kendall test moderate positive trend during summer and winter, and a weak negative trend during monsoon is obtained for maximum temperature, whereas a strong positive trend during winter, a moderate positive trend during monsoon, and a moderate negative trend during summer is obtained for the minimum temperature. By Sen's Slope, both maximum and minimum temperature shows an increasing trend. i.e., a weak positive trend of maximum and a moderately positive trend of minimum temperatures. While considering the rolling average analysis the graph for both maximum and minimum temperatures shows a slight increase over time.

Hence, it can be concluded that the maximum temperature is increased to a slight extent. on the other hand, the minimum temperature is also increasing moderately i.e., the coldest temperatures are becoming less cold while the temperature is rising hotter than before. Therefore these trends suggest an impact on climate change, agriculture, human health, and ecosystems. Thus further studies and factors should be considered to make a concrete conclusion for the overarching complete understanding of the trends in Ernakulam. Additionally measures to mitigate the impact of climate change and the factors affecting it should be taken by the government and individuals for a better and healthier future ahead.

CHAPTER-6

REFERENCE

- Arulbalaji, P., Padmalal, D., & Maya, K. (2020). Impact of urbanization and land surface temperature changes in a coastal town in Kerala, India. *Environmental Earth Sciences*, *79*, 1-18.
- Basha, G., Kishore, P., Ratnam, M. V., Jayaraman, A., Agha, A., Ouarda, T. B., & Velicogna, I. (2017). Historical and projected surface temperature over India during the 20th and 21st century. *Scientific reports*, 7(1), 2987.
- Gopakumar, C. S., Prasada Rao, G. S. L. H. V., & Ram Mohan, H. S. (2011). Impacts of climate variability on agriculture in Kerala (Doctoral dissertation, Cochin University of Science & Technology).
- Hingane, L. S., Rupa, K., & Ramana, B. V. (1985). Long-term trends of surface air temperature in India. *Journal of Climatology*, 5(5), 521-528.
- Kabbilawsh, P., Sathish, D., & Chithra, N. R. (2020). Trend analysis and SARIMA forecasting of mean maximum and mean minimum monthly temperature for the state of Kerala, India. *Acta Geophysica*, 68(4), 1161-1174.
- Kothawale, D. R., & Rupa Kumar, K. (2005). On the recent changes in surface temperature trends over India. *Geophysical Research Letters*, *32*(18).
- Ross, R. S., Krishnamurti, T. N., Pattnaik, S., & Pai, D. S. (2018). Decadal surface temperature trends in India based on a new high-resolution data set. *Scientific reports*, 8(1), 1-10.
- Subash, N., & Sikka, A. K. (2014). Trend analysis of rainfall and temperature and its relationship over India. *Theoretical and applied climatology*, 117, 449-462.
- Valappil, N. K. M., Hamza, V., & de Oliveira Júnior, J. F. (2023). Evaluation of daily average temperature trends in Kerala, India, using MERRA-2 reanalysis data: a climate change perspective. *Environmental Science and Pollution Research*, *30*(10), 26663-26686.
- Vijay, A., & Varija, K. (2022). Machine learning–based assessment of long-term climate variability of Kerala. *Environmental Monitoring and Assessment*, *194*(7), 498.

ST. TERESA'S COLLEGE (AUTONOMOUS)
DEPARTMENT OF MATHEMATICS
AND STATISTICS
ERNAKULAM