A Data-Driven Exploration of Stratospheric Ozone Dynamics: Bridging Regional Ozone Insights with Environmental Policy

by

Md. Abir Hossain
22101657
Sadia Nawshin
22101660
Sabira Rahman
22101672
Shah Md. Abdullah-Al Saud
22101688
Saba Rubaia
24341219

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
BRAC University
December 2025

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing the degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Md. Abir Hossain
22101657

Sadia Nawshin
22101660

Sabira Rahman
22101672

Shah Md. Abdullah-Al Saud
22101688

Saba Rubaia
24341219

Approval

The thesis titled "A Data-Driven Exploration of Stratospheric Ozone Dynamics: Integrating Machine Learning, Policy-Aware Modeling across Global Regions" submitted by

1. Md. Abir Hossain(22101657)
2. Sadia Nawshin(22101660)
3. Sabira Rahman(22101672)
4. Shah Md. Abdullah-Al Saud(22101688)
5. Saba Rubaia(24341219)

Of Fall, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on December 06, 2025.

Examining Committee:

Supervisor:
(Member)

Md. Sabbir Ahmed
Senior Lecturer
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Ms. Khondoker Nazia Iqbal
Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

The stratospheric ozone can be very crucial in protecting the Earth against harmful UV radiation, as well as its restoration after the 1987 Montreal protocol, is unevenly spread in various regions. This research is a data-based examination of the long-term dynamics of the ozone in various countries despite having different climatic and geographical settings, which showed specific recovery in various regions. The study models complex seasonal and nonlinear ozone behavior, using the support of more complex feature engineering, which incorporates lag variables, rolling averages, and temporal indicators based on advanced deep-learning models: LSTM, GRU, TCN, Transformer, and hybrid solutions. Model assessment, which is based on the combination of accuracy measures and uncertainty estimation, indicates that LSTM is the best in terms of explanatory performance, and GRU achieves the lowest in terms of MAE and RMSE in all six climatic regions. Another meta-analysis conducted across all regions further synthesizes recovery slopes and prediction error and levels of uncertainty, with strong recovery rates in the Tropical, Temperate regions, and slower or more erratic rates in Polar, Subpolar, and Arid regions. A policy modeling framework based on the use of data-driven insights to inform climate-aligned policies in SDG 13 (Climate Action), the mitigation of UV-risks in SDG 3 (Good Health and Well-being), and the improvement of environmental planning in SDG 11 (Sustainable Cities and Communities) is also introduced in the study. This framework offers an evidence-based and scalable policy instrument to monitor the environment in the long term and make decisions to bridge long-term ozone recovery and policy action recommendations to sustainable climate decisions.

Keywords: Ozone Layer Analysis, Time Series Analysis, Deep Learning (DL), Policy-informed modeling, Sustainable Development Goals (SDG).

Acknowledgement

Firstly, all praise to the Almighty Allah for whom our research has been completed without any major interruption.
Secondly, to our Supervisor, Md. Sabbir Ahmed sir, and our Co-Supervisor, Ms. Khondoker Nazia Iqbal ma'am, for their kind support, constant presence, and precious advice in our work.
And finally, to our parents, because without their everlasting support, it wouldn't have been possible.

Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ACF Autocorrelation Function

ACM Association for Computing Machinery

ADF Augmented Dickey-Fuller

AQI Air Quality Index

ARG Argentina

$ARIMA$ Autoregressive Integrated Moving Average

ASM American Samoa

ATA Antarctica

AUS Australia

CAN Canada

CFC Chlorofluorocarbons

CHN China

CNN Convolutional neural network

CPD Change-Point Detection

DiD Difference-in-Differences

DU Dobson Unit

$ENSO$ El Niño-Southern Oscillation

FFN Feed Forward Network

GRL Greenland

GWP Global Warming Potential

HFC Hydrofluorocarbons

ICT Information and Communication Technology

$IEEE$ Institute of Electrical and Electronics Engineers

IND India

IQR Interquartile Range

$LBWB$ Local Block-wise Wild Bootstrap

$LSTM$ Long Short-Term Memory

MAE Mean Average Error

MEX Mexico

ML Machine Learning

MSE Mean Squared Error

$PACF$ Partial Autocorrelation Function

PBL Planetary Boundary Layer

$PELT$ Pruned Exact Linear Time

PSC Polar Stratospheric Clouds

RCM Regional Climate Model

RF Random Forest

$RMSE$ Root Mean Squared Error

$RMSPE$ Root Mean Squared Prediction Error

$SARIMA$ Seasonal Autoregressive Integrated Moving Average

SCM Synthetic Control Method

SDG Sustainable Development Goals

$SHAP$ SHapley Additive exPlanations

$SUTVA$ Stable Unit Treatment Assumption

TCN Temporal Convolutional Network

TCO Total Column Ozone

TPE Tree-structured Parzen Estimator

URL Uniform Resource Locator

USA United States of America

UV Ultra Violet

$VECM$ Vector Error Correction Model

WHO World Health Organization

Chapter 1

Introduction

Most of the harmful ultraviolet (UV) radiation from the sun is absorbed by the ozone layer, located in the stratosphere. The protection of the ozone layer was seriously compromised in the mid-20th century by ozone-depleting substances like chlorofluorocarbons (CFCs), which resulted in seasonal holes in the ozone layer, particularly around the poles [12]. In response, the global community adopted the Montreal Protocol (1987), an international treaty that drastically reduced the production and use of ozone-depleting substances [23]. The protocol is widely recognized as one of the most successful environmental agreements in history, since its amendments indicate that the ozone layer is slowly recovering and could return to pre-1980 levels by the middle of this century [21][22].

Despite this progress, ozone recovery is far from uniform. The dynamics of ozone are highly dependent on latitudes because of the variation in atmospheric circulation, temperature structure, and sun exposure. The polar regions are characterized by great seasonal changes and drastic episodic decrees, as contrasted with stable and consistently lower ozone levels that characterize the tropical regions. To observe these spatial differences, this paper will consider the long-term ozone behavior of 12 countries that exemplify the polar, temperate, sub-tropical, and tropical climatic areas, so as to have a regionally particular interpretation of the ozone variation.

A variety of deep-learning architectures, such as LSTM, GRU, TCN, and Transformer, are used to analyze the ozone behavior between the regions, using long-term country-level time series. Model evaluation involves a variety of reliability aspects, integrating measures of accuracy with a measure of uncertainty, and a calibration study to satisfy the aim that predictions can be retained unaltered between varying atmospheric conditions, as it has been revealed in all the evaluation parts of the research. Comparison of the regions reveals that LSTM predicts the variance with large R^2 values, but GRU has the lowest MAE and RMSE in all six climatic regions. Other analyses, including seasonal and coverage-based error evaluation and structural diagnostics, indicate further that regional volatility, seasonal amplitude, and data density have a strong correlation with model performance. These modeling results form a solid basis on which the following cross-region meta-analysis would be based, as well as the overall policy and sustainability interpretation of the study.

The cross-region meta-analysis is one of the main elements in this work as it synthesizes recovery slopes, predictive errors, and uncertainty estimates to identify how the behavior of ozone varies across climatic zones. The findings reveal that the Temperate and Tropical areas have stronger recovery signals, whereas the Polar, Subpolar, and Arid areas have flighted recovery patterns affected by the extreme seasonality, cold-season chemistry, or lack of observational conditions. Uncertainty estimates further show that more variable regions of the atmosphere have wider confidence intervals, with their more stable areas, such as Subpolar and Arid, having zero-width intervals as model errors are uniform. These cross-regional insights highlight that ozone recovery is a geographically uneven process shaped by local atmospheric drivers.

Importantly, this question is not limited to scientific reading; on the contrary, this question is directly associated with processes of global sustainability. The research can be an SDG 13 (Climate Action) contribution by helping climate-commensurate strategies and enhanced, sustained commitments via the Montreal Protocol through the recognition of prompts in recovery regions, slow progress, and regions with higher UV hazards. High-UV season identification can serve the purpose of SDG 3 (Good Health and Well-being) because it will enable the early-warning measures and make the public-health bodies to address the threat of the dangers associated with UV. In the meantime, understanding of regionally-specific variability of the atmosphere can support SDG 11(Sustainable Cities and Communities) because of its role in assisting the cities to integrate ozone-risk patterns into their long-term environmental planning. It is these links that enable the study to highlight ozone forecasting as a scientific practice as well as an instrument to evidence-driven, future-oriented policy development.

1.1 Problem Description

Life on Earth depends on the stratospheric ozone layer to protect itself against dangerous ultraviolet (UV) radiation. Nonetheless, it has been draining with time. This depletion is through the reaction of chlorine and bromine atoms emitted by ozone-depleting compounds, e.g., CFCs and halons, with ozone molecules. All these atoms have the ability to destroy thousands of moles of ozone before they get out of the stratosphere, even though there have been international agreements like the Montreal Protocol, whereby the global community has embarked on a mission to reduce ozone-depleting substances [30] [5]. This has seen the trend of depletion and recovery regionally uneven. There are still regions in the world where ozone variability remains very strong and climatic factors are not well comprehended or well covered in the current research [17].

Although there are a number of national and local studies that have examined the trends in ozone using traditional time-series analysis, there are few studies that have examined the trends on a global and cross-regional basis. Statistical models can find it difficult to model the non-linear effects of time, sudden regime changes, and differences in climate and zone variations that dominate ozone behavior. Subsequently, they provide a low predictive power and do not provide a sense as to why ozone does not conduct itself in equal measures in different atmospheric regions. More recent developments in machine learning, especially the deep-learning models of LSTM, GRU, TCN, and Transformer, leave more perspectives to recreate the intricate ozone dynamics. Long-term structure in time, seasonal variations, and regional variability can now be represented in these models, and also cross-region assessment can be used to show systematic differences in accuracy, uncertainty, and structural behavior between climate zones.

Nevertheless, prediction can not be made reliable by itself. To have a good environmental policy and risk management, there should be an integrative knowledge of the variation of ozone trends in various regions, areas that are recovering best, and those areas that are recovering poorly, as well as the variation of the unpredictability with the atmospheric behavior. In the absence of such knowledge, policymakers would fail to formulate specific policies, which are conducive to the climate response, the mitigation of UV risks, and the territorial environmental planning. The current literature has not unified machine-learning forecasting with an effective cross-region comparison, uncertainty, and a policy-conscious interpretation. This poses a void in the achievement of an integrated framework that connects information-based modeling to efforts of global sustainability. Consequently, an urgent analytical framework of analysis, with a regional focus, which combines deep-learning predictability, cross-region meta-analysis, and policy relevance that provides actionable information on the recovery of ozone in the world, is significant.

1.2 Research Objectives

The primary objective of this research is to develop a quantitative, coded area-concentrated analytical scheme to examine the global ozone layer dynamics with the help of the contemporary time-series structuralization, and study the areas separately. This study will attempt a geographically disaggregated approach to measure long-term recovery behavior of ozone, establish regional differences, determine predictive accuracy, and follow a policy-congruent environmental policy. By studying ozone behavior at both global and regional dimensions, this study aims to generate scientifically robust and policy-relevant insights regarding ozone recovery dynamics.

The specific objectives of this study are:

1. To assess the ozone dynamics of regional and global by studying the process of long-term depletion and recovery in various countries and climatic regions, and fill missing knowledge gaps in comparative large-scale ozone studies.

2. To test and utilize advanced deep learning models, including Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Temporal Convolutional Networks (TCN), and Transformer models to capture and identify nonlinear temporal relationships, seasonal trends, and area-related ozone behaviors.

3. To measure the model performance using a multi-metric evaluation scheme, which includes the accuracy measures (R^2, MAE, RMSE), uncertainty estimation, and the calibration analysis to validate a reliable and robust model to forecast ozone.

4. To carry out cross-region meta-analysis of recovery slopes, predictive errors, and interval widths to establish variation in ozone behavior of different climatic regions, as well as to determine the atmospheric variables arising to cause such variations.

5. To establish policy relevance using scientific findings to connect the knowledge on sustainability planning that is designed to ensure that regional ozone knowledge can be used when addressing the environmental risk mitigation, UV-exposure reduction, and long-term ozone recovery.

6. To match the research findings with the United Nations Sustainable Development Goals (SDGs), such as SDG 13 (Climate Action), SDG 3 (Good Health and Well-being), and SDG 11 (Sustainable Cities and Communities) by developing an analytical framework that is both scalable and reproducible, connecting the process of ozone forecasting and the implementation of environmental decisions.

The proposed study thus seeks to offer a solid, geographically sensitive, machine-learning based system that can assist in the provision of evidence-based ozone-recovery assessment and sustainability across the globe.

1.3 Research Contribution

The key contributions of this study in this respect are as follows:

1. **Empirical Evidence of Policy Impact:** This study presents quantitative,region-based evidence that the Montreal Protocol has resulted in an observable recovery of the ozone past 1987, although the recovery has not been evenly spread across climatic regions. The research paper establishes areas where the recovery is more vigorous and areas where the recovery is slower or incomplete, which gives the entire picture of the influence of the treaty.

2. **Regional Differentiation in Ozone Recovery:** The Study reveals that ozone recovery can not be regarded as a global average. Instead it demonstrates that there exist very massive differences in the recovery rates in various regions, such as the Polar, Tropical and Subpolar etc Regions. The need to have a regionalization policy highlights the need of having a region-specific policy, which is sensitive to needs of individual regions. needs of every region.

3. **Causal Validation of the Effect of Montreal Protocol:** The study examines the effect of the intervention by the Montreal Protocol using both superior methods as Difference-in-Differences (DiD) and Synthetic Control Methods (SCM). These techniques enable us to prove that the ozone recovery measured is not a result of chance variation and any other external factor but the direct result of the policy intervention of the Montreal Protocol. These methods are used to confirm the effectiveness of the Protocol in reducing ozone depletion and causing recovery by comparing the regions that undergo the intervention and those that do not.

4. **Determining High-Risk and High-Uncertainty Regions:** The study also determined the areas where the ozone dynamics are volatile or the ones that are unpredictable, especially in the slow recovering ones. The study of points of change, prediction of uncertainty, and early interventions to avoid additional depletion.

5. **Reusable Analytical Framework to evaluate policy:** This study has offered a generalizable analysis system to measure the success of global environmental conventions. Based on data analysis, the framework will provide a holistic method to inform the decision-making process when it comes to climate treaties, air quality regulations, and emission reduction options. It aids in creating and evaluating environmental initiatives by delivering practical data to assist in aligning actions at a global scale with sustainability and recovery objectives on a long-term basis.

The results of this research are important in the sense that they reveal the effectiveness of the Montreal Protocol and how it can be taken as an extension to the global environmental policy. In this research, machine learning methods are used to track and predict ozone recovery which forms a potent analytical framework to evaluate the impact of the environmental treaties. As the ozone layer is healing, the difference between the healing across the regions indicates the applicability of the individual, data-driven policies since it can be applied to address a local issue and still follow the goals of global sustainability.

1.4 Thesis Structure

This thesis is structured so that the dynamic processes of the global ozone layer are examined with a data-driven approach and a machine-based learning one and a systematic analysis of the long-term trends, regional differences, and policy-significant outcomes is provided. After the introduction and research objectives, **Chapter 2** conducts a review of literature available on ozone layer dynamics, the Montreal Protocol, time-series analysis, and machine learning applied to the environmental research. **Chapter 3** explains the datasets employed in the research such as data sources, selection of countries, data validation, consideration of fairness and exploratory data visualization. The **4th chapter** is the methodological framework, which includes the statistical methods, analyses based on time-series, and machine learning models that the research used. **Chapter 5** discusses the results and findings, focusing on ozone depletion and recovery patterns across different regions and the performance of the applied models & how their cause and effect works. **Chapter 6** emphasizes summarizing the main findings and discussing policy implications related to the Montreal Protocol. Finally, **Chapter 7** concludes the thesis by concluding, discussing limitations and suggesting directions for future research.

Chapter 2

Literature Review

The various geographical zones of the planet, from polar regions to arid zones, each have unique atmospheric features that affect patterns of ozone concentration. It is essential to comprehend these regional differences so that large-scale meteorological and atmospheric datasets can be analyzed using machine learning (ML) techniques, besides allowing the prediction of ozone levels in global regions and providing researchers and policymakers with a summary of key findings.

2.1 Related Works

2.1.1 Ozone Layer Recovery and Policy Impact

According to Canan et al. (2015), to replace ozone-depleting greenhouse gases with alternative lower global-warming potentials and increasing energy efficiency, the Montreal Protocol is regarded as the most effective treaty [6]. In order to shape climate governance, this study integrates the multidisciplinary contributions of scientists, politicians, and diplomats rather than using a technological model. With an 85% replacement rate, it highlights how well the Protocol works to replace ozone-depleting chemicals (ODSs) with alternatives that have a lower global-warming potential (GWP) [6]. However, the paper also highlights the unfinished task of phasing out the remaining 15% of high-GWP hydrofluorocarbons (HFCs) [6]. The authors urge sustained international collaboration and innovative policymaking to bridge the gaps left by the current approaches [6]. In a different study, Annadate et al.(2025) suggested a novel way to combat ozone layer depletion brought on by human emissions like CFCs by combining game theory and time series forecasting [27]. Based on datasets such as air quality indices and ozone depletion measurements, the study was used to forecast using ARIMA, SARIMA, and exponential smoothing equations with an RMSE of 5.04. The SARIMA model served well to model the seasonal ozone variations, whereas the game theory served to offer the framework of activities based on cooperative mitigation. The limitations associated with the method, however, are the complexity of the tweaking of the SARIMA model and the fact that the method also requires a considerable amount of historical data. The paper focuses on interdisciplinary methods for pursuing environmental policies [27].

Parson (1996) conducted a qualitative policy study of the global ozone layer protection by using historical scientific literature in another study. He wrote about the worldwide loss of ozone of about 5 percent since the 1960s, the loss of up to 60 percent over Antarctica [2]. He discussed the significance of the Montreal Protocol and the Vienna Convention as having helped start the process of international collaboration. Though they are important barriers applied to their implementation, like the illicit ODS trade and budget constraints hindered their implementation. In an effort to enhance the protection of ozone, Parson recommended that there should be enhanced control of the international and proper monitoring of the funds, as well as more stringent legal processes [2].

2.1.2 Predictive Modeling of Ozone and Air Quality

Besides this, Lyu et al. (2022) analyzed datasets on six major contaminants (CO, NO_2, O_{3_8h}, PM_{10}, $PM_{2.5}$, and SO_2) from 2014 to 2021 in the Beijing-Tianjin-Hebei region's data on air quality [20]. They forecasted ozone concentrations using Random Forest and Decision Tree models with 10-fold cross-validation; RF had a higher accuracy rate (daily $R^2 = 0.83$ and RMSE = 30 μg/m^3; monthly $R^2 = 0.93$ and RMSE = 12.1 μg/m^3). While NO_2 and ozone showed initial increases followed by falls, probably as a result of COVID-19 impacts, the study found overall decreases in pollutants like CO, PM_{10}, and SO_2. The models' sensitivity to data quality and difficulties in capturing intricate atmospheric chemistry were among their drawbacks. To better understand causal links, the authors suggest incorporating hybrid deep learning models and enlarging the dataset to incorporate meteorological factors in future research (Lyu et al., 2022) [20].

Ko, Cho, and Rao (2022) designed a machine-learning framework for forecasting near-surface ozone that combines meteorological and air-quality data with planetary boundary layer (PBL) data [19]. They trained Multilayer Perceptron (MLP) and bidirectional LSTM networks to predict short-term ozone concentrations using hourly ozone, temperature, and humidity data from EPA monitoring stations in California (2017–2019) in conjunction with predicted PBL heights from the NAM model. Although the Index of Agreement was mostly constant, the model's performance was enhanced by the addition of PBL data, which reduced MAE and RMSE by up to 10% across many stations [19]. Their results highlight the importance of vertical atmospheric structure in enhancing ozone forecasts. However, the study's spatial restrictions and omission of significant indicators like nitrogen oxides and wind speed may restrict how far it can be used. For future advances, the authors recommend testing in a variety of climatic zones and incorporating other meteorological and chemical parameters [19]. The aim of Wang et al. (2024) was to characterize the temporal and regional variations in troposphere ozone concentration, identify the key causes of such changes, and predict future ozone concentrations with the help of the machine learning model to reveal the ability of the methodology to reproduce the outcomes of the study [26]. Predicting the future ozone levels in 2022 and 2023, the researchers relied on the Random Forest Regression model and found that the approach successfully predicted the observed trends in ozone concentrations at the regional level [26]. The results also showed significant changes in seasons and geographical locations, with central Europe and the southeast coastal areas recording higher ozone levels. Temperature, nitrogen dioxide (NO_2), population density, greenhouse gas emissions, and forest cover were found to be the main variables that affected ozone variability [26]. The study observed limits associated with the use of column ozone data instead of surface-level measurements, which may influence the accuracy of exposure evaluations, despite the model's great predictive power. For future research, the authors suggested adding more pollutant parameters and high-resolution ground-based data to improve model accuracy and policy relevance [26].

On the other hand, Nowack et al. used temperature data from the HadGEM3 global climate model to parameterize ozone distributions using a machine learning technique called Ridge regression [8]. Reducing computational needs in climate sensitivity simulations without sacrificing accuracy was the aim of the research. Their model provided a less expensive substitute for different climate scenarios by accurately capturing mean ozone climatology and stratospheric variability. However, findings were susceptible to the climate conditions of the training dataset, and performance was poorer in the troposphere, where ozone variability is larger. Future research will focus on improving the accuracy of tropospheric predictions and evaluating the approach using several efficient climate models (Nowack et al.)[8]. Doury et al. (2022) present a hybrid downscaling technique that uses deep learning to simulate Regional Climate Model (RCM) outputs [18]. Their method links high-resolution near-surface temperature maps (12 km) throughout a complicated Southwest European area to large-scale predictors (from General Circulation Models or GCMs) by training a UNet-based neural network to emulate the RCM's downscaling function [18]. It is trained in both current and future climate simulations under various RCP scenarios. While significantly cutting down on computation time when compared to full RCM runs, the results demonstrate a high level of proficiency in capturing daily and spatial variability, as well as fine spatial features like coasts and steep terrain [18]. Nevertheless, the model has warm biases, understates extremes, and has poorer accuracy for the intensity of the climate change signal. For further usefulness in impact studies, the authors propose expanding this approach to additional locations, factors other than temperature, and enhancing the replication of extreme climatic occurrences [18].

To examine the spatiotemporal distribution of surface ozone and the health concerns associated with it throughout China, Ma et al. (2024) created the machine learning framework NetGBM [25]. The model showed high predicted accuracy at daily and monthly scales using land-use and meteorological factors, as well as data from more than 1,700 monitoring sites [25]. According to the study, urban spots such as the North China Plain and Yangtze River Delta had the greatest exposure, and the mortality rates of ozone, temperature, and the type of land-use. and the population density were also key factors that contributed to the two spots [25]. His model was constrained in generalizability, as it depends on the density of ground monitoring, and chemical precursor data were not utilized, although it had a good prediction capability. The authors stressed that to facilitate the management of the environmental policy and the population, the use of machine learning and health risk assessment should be integrated [25]. Betancourt et al. (2022) created a high-resolution global troposphere ozone mapping framework based on explainable machine learning [16]. The model (Random Forest) applied in the study was trained on the AQ-Bench dataset, which incorporates diverse geographic and environmental data with ground-based measurements of ozone [16]. The scientists used SHAP-based interpretability and uncertainty analysis on SHAP in order to confirm the strength of the model in various geographic locations and determine important parameters that affect the dispersion of ozone. The accuracy of the model in prediction was moderate in regions that had more data, such as North America and Europe, and more unpredictable in regions that had fewer monitoring networks. The results revealed the influence of climate and emission-related variables on the global variation of ozone [16]. The temporally aggregated data used in the study, however not able to represent seasonal or diurnal changes. In order to enhance ozone prediction and interpretability at the global scale, the authors recommended that further research should use temporal models, dynamic meteorological variables, and more extensive coverage [16].

The authors of the study resorted to the typical procedures of preprocessing, such as normalization, to construct the univariate time series using past gold prices. They use a 1D convolutional neural network (CNN) together with a long short-term memory (LSTM) network. The CNN captures short-term variations and local time variations, whereas the LSTM captures the short- and long-term sequential correlations [14]. Comparing the hybrid CNN-LSTM model with either pure machine-learning or pure deep-learning as a baseline in an experiment, they were able to demonstrate a significant increase in forecasting accuracy, which indicated that a mixture of sequential memory and spatial (local) feature learning yields better predictive performance. The model is also trained on one asset price series (gold) with no exogenous variables (such as macroeconomic indicators), which possibly would limit its extrapolation to other market-specific conditions. Also, according to the authors, deep learning is a black-box rendering interpretability and resilience to regime changes. In order to make the methodology more robust and flexible, they suggest that it would be wise to utilize more multivariate contributions, more financial resources, and explainability or external indicators mechanisms [14]. Mehtab, Sen, and Dasgupta's study evaluates the effectiveness of CNN, LSTM, and hybrid CNN-LSTM architectures in a deep learning-based framework for financial time-series forecasting using high-frequency intraday stock price data gathered at 5-minute intervals from the National Stock Exchange (NSE) of India between December 2012 and January 2015 [15]. The authors create sliding-window inputs after preprocessing and normalizing the price series to train several CNN-based models, LSTM-based models, and a hybrid model in which CNN layers extract short-term local patterns, then send features to LSTM layers to capture temporal dependencies. Their studies reveal that the hybrid CNN-LSTM model regularly outperforms both standalone CNN and LSTM models, achieving lower RMSE and demonstrating strong predictive potential for short-term price changes [15]. However, the study is hampered by its reliance on a single company, its use of solely univariate price data without incorporating external market indicators or sentiment factors, and the inherent interpretability issues of deep neural networks. As future work, the authors advise applying the models to more diverse datasets, combining multivariate characteristics and technical indicators, examining longer-term projections, and adopting more advanced hybrid or interpretable architectures to improve generalizability and practical usefulness [15].

The study evolving deep CNN-LSTMs for inventory time series prediction proposes a hybrid deep-learning strategy that combines 1-D CNN layers for extracting local temporal patterns with LSTM layers for modeling long-term dependencies in inventory time series data [11]. The authors employ two Differential Evolution (DE) variations and Particle Swarm Optimization (PSO), two meta-heuristic optimization techniques, to automatically develop the CNN-LSTM structure and its hyperparameters, overcoming the challenge of manually creating efficient architectures. The study demonstrates that the developed hybrid models outperform traditional statistical techniques such as SARIMA and manually adjusted deep models in terms of predictive accuracy on real-world inventory datasets [11]. The work shows the usefulness of combining neural architectures with evolutionary search to manage noisy, non-stationary supply-chain data; however, it is constrained by high computational cost and is assessed on a rather narrow dataset. Future study suggested by the authors involves applying the strategy to other diverse datasets and researching additional hybrid or interpretable models [11]. Politics (2003) analyzes how the bootstrap method must be altered for time-series data, since the traditional i.i.d. bootstrap fails to account for temporal dependence [3]. The research examines block-bootstrap variations, stationary bootstrap, subsampling, and model-based techniques, stressing their ability to preserve dependence patterns during resampling. The study demonstrates through theoretical analysis and Monte Carlo experiments on autoregressive and moving-average processes that while these approaches frequently outperform classical asymptotic theory in finite-sample estimation, they also have lower accuracy compared to the i.i.d. case, especially in coverage probabilities and sensitivity to block-length selection. A fundamental issue is that no single bootstrap approach works well for all dependent patterns, and practical suggestions for setting block sizes remain unresolved. The study emphasizes the need for greater investigation into bootstrap methods for time series situations that are nonstationary, long-memory, and more complicated [3]. To enhance inference in linear time-varying coefficient models for locally stationary time series, where conventional bootstrap techniques fall short, Lin, Song, and van der Sluis (2024) provide the Local Blockwise Wild Bootstrap (LBWB) [28]. Using extensive Monte-Carlo simulations and an empirical application to Chinese renewable-energy stock data, the authors show that LBWB - using nonparametric estimation combined with locally adaptive block resampling - produces more accurate confidence intervals and bands than existing bootstrap approaches. The technique reveals shifting market behaviors, including diminishing hedging over time, and captures well changing reliance patterns. Nevertheless, LBWB is computationally demanding, implies smoother parameter growth, and is sensitive to block-length and bandwidth decisions. The authors advocate further extensions to nonlinear nd multivariate contexts, enhanced tuning-parameter selection, and modifications for structural fractures [28].

2.1.3 Health and Sustainability Implications

The study analyzes how economic uncertainty, economic growth, financial development, environmental policy, and life expectancy affect human health in China using publicly available time-series data and a simultaneous quantile regression approach to capture heterogeneous effects [31]. The results demonstrate that while economic uncertainty and health spending show negative correlations, economic growth, robust environmental legislation, financial development, and longer life expectancy all improve health outcomes. The study's limitations include reliance on aggregated national-level data, potential missing variables, and its scant attention to lagged or causal links. The authors advise future work that employs more granular regional data, integrates additional variables such as pollution and inequality, and applies more advanced dynamic or causal models to better capture the complex pathways linking economic conditions to health [31]. The study explores how government performance, health expenditure, and progress toward the Sustainable Development Goals (SDGs) influence life expectancy in Pakistan using annual time-series data from key international databases [24]. Using unit-root tests, Johansen cointegration, and a Vector Error Correction Model (VECM), the authors evaluate both the long-run and short-run dynamics supplemented by Granger causality tests and impulse response analysis [24]. The results demonstrate that health expenditure and SDG development significantly boost life expectancy, whereas the influence of government performance is less and often statistically ambiguous. Short-run causality flows from governance, spending, and SDGs toward life expectancy, while shocks to health expenditure and SDGs create favorable long-term health responses. The use of secondary data at the national level, the possibility of omitted-variable bias, and the absence of thorough causal identification are the main limitations. To better understand the policy routes that influence population health, the authors suggest further study employing more detailed regional data, additional socioeconomic and environmental variables, and cross-country comparative analysis [24].

Wu et al. (2018) explored the crucial relationship between Information and Communication Technologies (ICTs) and the 17 Sustainable Development Goals (SDGs) set by the United Nations in a paper [10]. The main discussion highlights the numerous ways that ICTs can help achieve the SDGs across the three core areas of sustainability: social, economic, and environmental. However, the authors emphasize a significant problem: a majority of the research coming from the technical community (like IEEE and ACM) has focused too narrowly on technical solutions alone [10]. This has resulted in a shortage of holistic work that considers the broader "social good perspectives" necessary to fully realize the SDGs. The paper, therefore, serves as a call to action, urging the research community to expand its focus, innovate, and deploy ICT solutions in a way that provides more effective and comprehensive assistance to all nations working to meet the 2030 development targets [10]. To support sustainable urban planning and health policy, Mujtaba et al. (2025) focused on forecasting Lahore's air quality using machine learning and time-series models [29]. The dataset comprises four meteorological elements (temperature, humidity, and wind parameters) and eight pollutants ($PM_{2.5}$, PM_{10}, CO, NO, NO_2, O_3, SO_2, and PM_1) based on 20 years (2003–2022) of daily data from the Copernicus Atmosphere Monitoring Service (CAMS) [29]. Four predictive models, namely ARIMA, SARIMAX, LSTM, and Nonlinear Autoregressive (NAR), were evaluated based on Dynamic Time Warping (DTW) trend similarities and Root Mean Square Error (RMSE) accuracy to evaluate company trends [29]. According to the findings, Lahore's Air Quality Index (AQI) is expected to increase by 13% to 16% by 2030, with high hazardous levels (300–500) continuing to exist for approximately seven months every month, especially during winter. In general, gases like CO, NO, NO_2, SO_2, and CO_2 stay within WHO standards, and ozone levels peak in April, even if $PM_{2.5}$ and PM_{10} concentrations are consistently substantially over safe criteria [29]. The study, however, acknowledges several drawbacks, including its reliance on the quality of historical data, its limited consideration of meteorological and urban expansion aspects, and its challenges with interpreting the deep learning model. Future research avenues include extending the analysis to additional cities, incorporating climate and socioeconomic factors, creating hybrid ML–physical models, investigating cutting-edge deep learning methods, and assessing practical pollution control measures [29].

2.2 Key Findings

The key findings of all these papers include the following:

- The Montreal Protocol has replaced 85% of ODSs with lower-GWP alternatives, while the phase-out of HFCs is still ongoing.
- When paired with meteorological and PBL data, machine learning models (Random Forest, Gradient Boosting, and LSTM) produce good ozone prediction accuracy ($R^2 > 0.8$).
- Hybrid models integrating climate, satellite, and ground data improve spatial-temporal forecasting but are less reliable in data-scarce regions.
- Ozone pollution directly impacts public health, supporting SDG 3 (Health) and SDG 13 (Climate Action).
- Most of the work focuses on regional ozone analysis, but not across the whole global region.

2.3 Our Work

Future work in one of the discussed papers suggests expanding regional coverage, including more variables, and using advanced deep-learning methods to improve long-term and real-time prediction. Our paper comes up with these expansions of regional coverage, where deep-learning models have been used as a tool for policy recommendations in long-term and real-time predictions. Besides this, future work in another paper focuses on the expansion of the cross-country comparison to increase generalizability and exploration of non-linear, dynamic, or causal modeling frameworks capturing complex relationships and possible structural breaks. Our paper proceeds by mitigating the gap through a cross-country comparison, analyzing the causal and structural breakpoints. We have used the models as a tool here to validate our policy recommendations and connections with SDGs to help policymakers analyze from a global perspective.

Chapter 3

Dataset

3.1 Data Collection

In this research, we have collected historical data on atmospheric ozone concentration in some countries of different regions. The necessary data was obtained at the World Ozone and Ultraviolet Radiation Data Centre (WOUDC) maintained by the Meteorological Service of Canada, official website (https://woudc.org) [32]. The data was collected separately for several countries, such as Argentina, Brazil, Canada, American Samoa, Antarctica, Greenland, Mexico, Australia, New Zealand, India, China, and the USA, to analyze and present individually. The reason behind choosing these countries is that each of the countries represents a different geographical and economic structure.

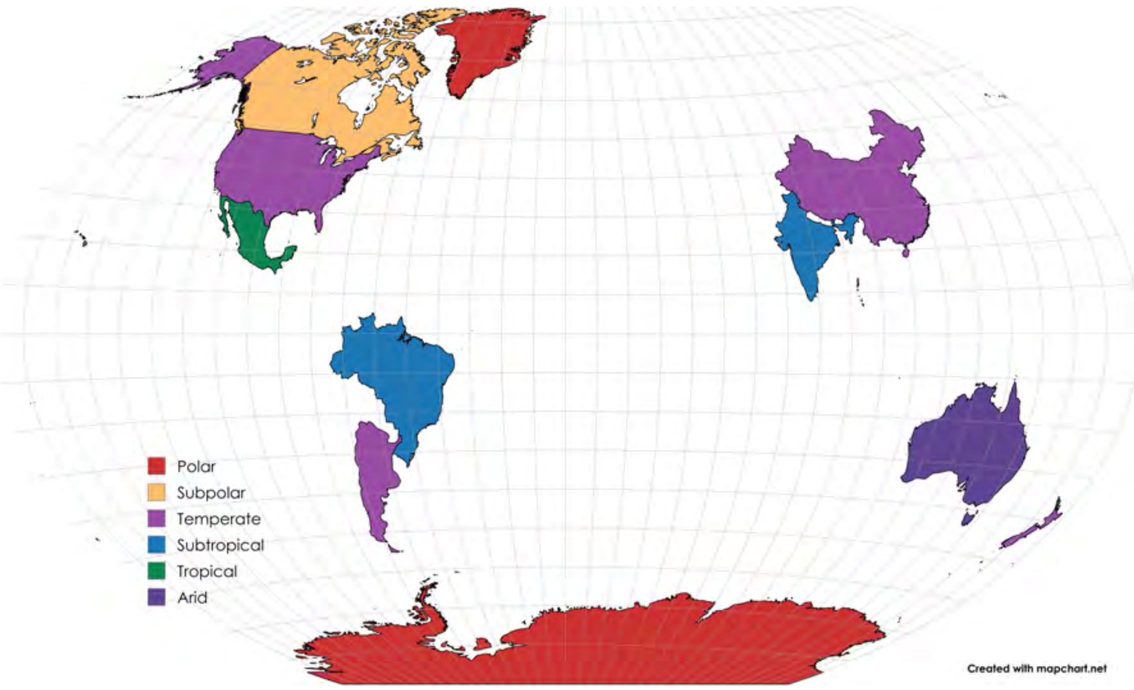

Figure 3.1: Map Chart of Selected Countries

Based on latitude, seasonal behavior, and dominating long-term temperature patterns, countries were categorized into climatic regions. Because Greenland and Antarctica lie within high-latitude Arctic and Antarctic circles, with year-round freezing temperatures, ice-covered landscapes, and little seasonal change, they are included in the Polar area. Due to long, freezing winters and its proximity to the Arctic zone, particularly in northern regions, Canada is categorized as subpolar. The USA, Argentina, China, and New Zealand belong to the Temperate area as they sit predominantly within the mid-latitudes, typified by four distinct seasons, moderate temperature ranges, and strong yearly weather cycles. Brazil and India belong to the Subtropical zone because they experience warm to hot temperatures year-round with strong monsoon or wet-dry seasonal variations. Due to their proximity to the equator, persistently high temperatures, high humidity, and little seasonal temperature change, American Samoa and Mexico are categorized as tropical. Lastly, Australia is classified as an arid region because a significant portion of the country, particularly the central and western regions, has desert and semi-desert climates with high evaporation rates and very little rainfall. Figure 3.1 shows the geographical locations of the countries around the world. Categorizing the countries into these regions gives the opportunity to analyze a diverse set of atmospheric dynamics and to understand the global ozone distribution.

3.2 Data Description

The dataset consists of detailed ozone observation records collected from 12 different countries: Argentina, Brazil, Canada, American Samoa, Antarctica, Greenland, Mexico, Australia, New Zealand, India, China, and the USA. All data for each country are provided in the individual file (e.g., totalozone_ARG.csv) and contain all the information related to ozone measurements, monitoring stations, instrumentation details, and observation metadata. Each file with 36 columns and multiple rows, depending on the reporting frequency of the station and years of coverage. The dataset records the daily and monthly values of ozone concentration (in Dobson Units), coordinates of the measuring stations, time intervals of observation, and statistical information such as standard deviation and the quantity of measurements per day. Metadata includes the name and model of the ozone instrument, identifiers of the contributor organization, the station of the country, and the reference URLs. Each file tracks ozone levels over multiple years using three time-related columns: observation_date, daily_date, and monthly_date. Measurement quality is described using fields like daily_nobs (number of observations), daily_stdevo3, and daily_mmu (mean minimum unit). The structure is standardized across all files, making them suitable for individual country-level studies or cross-country ozone pattern comparisons.

3.2.1 Validation and Profiling

Data analysis of validation and profiling of the ozone (O_3) datasets collected across different countries is presented in Table 3.1. In the table, the total number of data points, time range (starting date and end date), and descriptive statistics, the lowest and maximum recorded ozone are indicated, depending on the country. The number of errors that were present during the data validation process is also contained in the table.

Country	Data Points	Starting Date	Ending Date	Min O_3	Max O_3	Anomalies
ARG	45751	1965-10-01	2025-05-31	8.0	518.3	598
ASM	1451	2004-07-02	2022-12-31	213.0	303.0	31
ATA	108738	1957-02-12	2025-05-31	-302.0	2662.0	4864
AUS	102031	1957-07-01	2025-04-30	191.9	559.2	2149
BRA	24865	1974-05-08	2025-04-30	64.7	335.2	149
CAN	187033	1957-07-01	2023-12-31	137.7	626.5	2886
CHN	39220	1979-01-01	2025-04-30	-69.3	665.9	634
GRL	14523	1989-07-01	2025-05-31	167.4	619.5	114
IND	80658	1957-01-01	2024-11-30	156.9	431.6	1359
MEX	3615	1969-01-02	2025-01-30	144.2	352.7	51
NZL	15403	1957-01-08	2022-07-25	158.0	663.0	75
USA	140695	1957-07-01	2023-12-31	33.0	643.0	3781

Table 3.1: Validation and Profiling Summary For All Country datasets.

3.2.2 Data Fairness and Equity

Table 3.1 shows significant differences in dataset size and temporal coverage between nations. While some countries, like the USA and Canada, have more than 100,00 data points, others, like ASM, have comparatively small datasets. The starting dates range from the 1950s to the early 2000s, reflecting regional variations in data availability and monitoring history. Thus, a small comparison is shown in the following two figures of ASM 3.2 and CAN 3.3.

American Samoa (ASM):

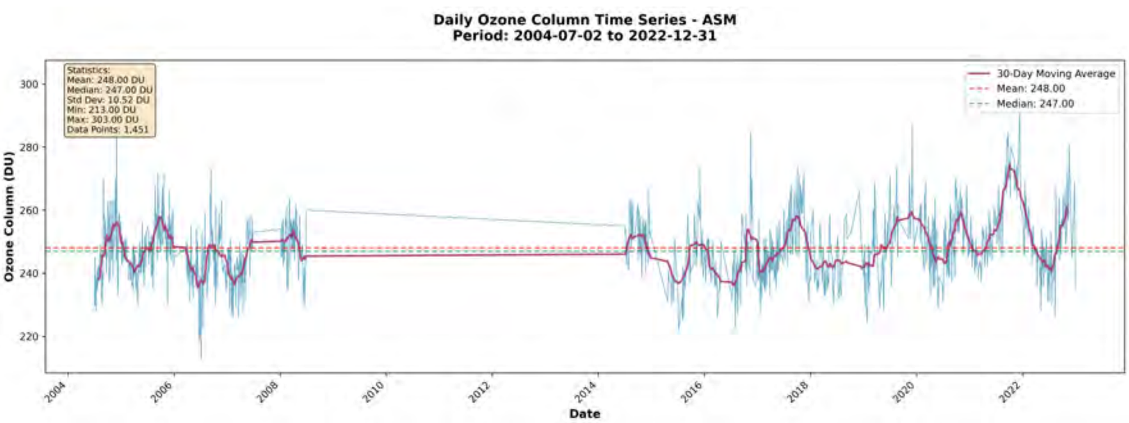

Figure 3.2: ASM Dataset Summary

Canada (CAN):

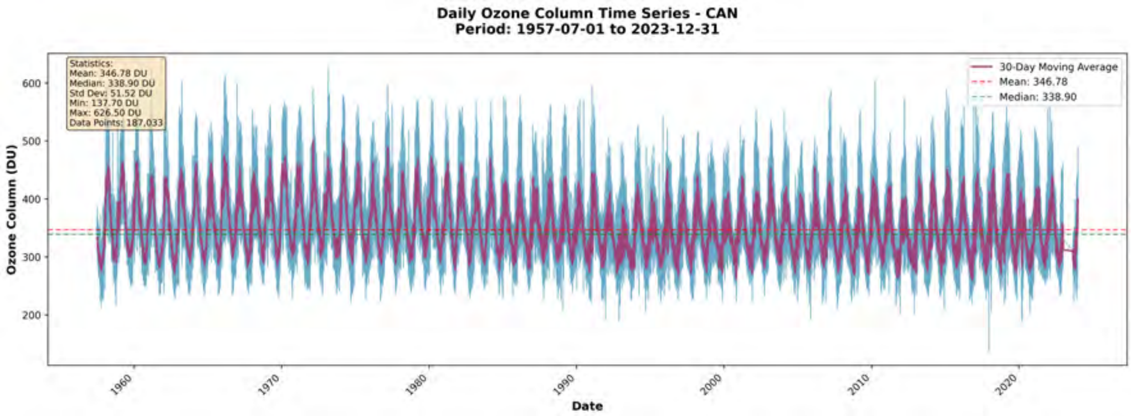

Figure 3.3: CAN Dataset Summary

In addition, the ozone measurements found in each dataset are shown in the Min O_3 and Max O_3 columns. The significance of the anomaly detection procedure is highlighted by the appearance of negative minimum values for certain countries, such as ATA and CHN, which may indicate sensor errors or measurement issues. Moreover, the number of odd or inconsistent data points found in each dataset is summarized in the Anomalies column. Countries with larger datasets or longer historical records - such as the USA, ATA, and CAN - tend to exhibit higher anomaly counts.

3.3 Data Visualization

This section of visualization presents data on ozone concentration from twelve countries: Argentina, Antarctica, American Samoa, Australia, Brazil, Canada, China, Greenland, India, Mexico, New Zealand, and the United States, based on total column ozone (daily_columno3) observations collected over several decades. The area graphs depict long-term ozone trends and present the depletion and recovery trends at each country and are consistent with the global atmosphere change and regulations. The temporal analysis shows that there are strong seasonal oscillation with constant long-term averages in most of the countries, whereas the anomaly detection indicates that it is only a small percentage of it that is not within the bounds of expectation. The tests of stationarity also show that most of these time series processes have similar statistical characteristics over time which proves the accuracy of the observed ozone behavior.

Argentina (ARG):

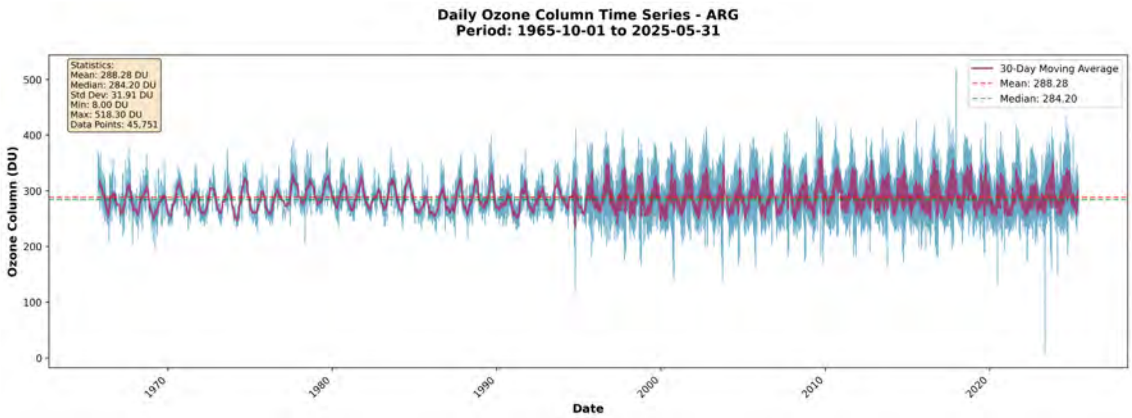

Figure 3.4: Temporal Variation of Daily Ozone Over The ARG

The Graph 3.4 indicates the daily ozone column in Argentina (ARG) between October 1965 to May 2025. There exists a powerful and consistent seasonal cycle during the whole period, which is a natural process of atmospheric and solar-initiated activities on the formation and disappearance of the ozone. In spite of these periodic fluctuations, the long-run trend is constant with an average of 288.28 DU, and a median of 284.20 DU. The moving average of 30 days proves the fact that there was no long-term upward or downward movement. The occasional extreme values, such as the dips below 100 DU and the peaks above 500 DU, are the short-term anomalies that can probably be associated with the unusual weather states or the artifacts in the measurements. The SD of about 31.9 DU indicates the presence of moderate variability in line with seasonal variability. In general, the dynamics of ozone in Argentina are stable and cyclic and lack long-term trends of depletion or recovery.

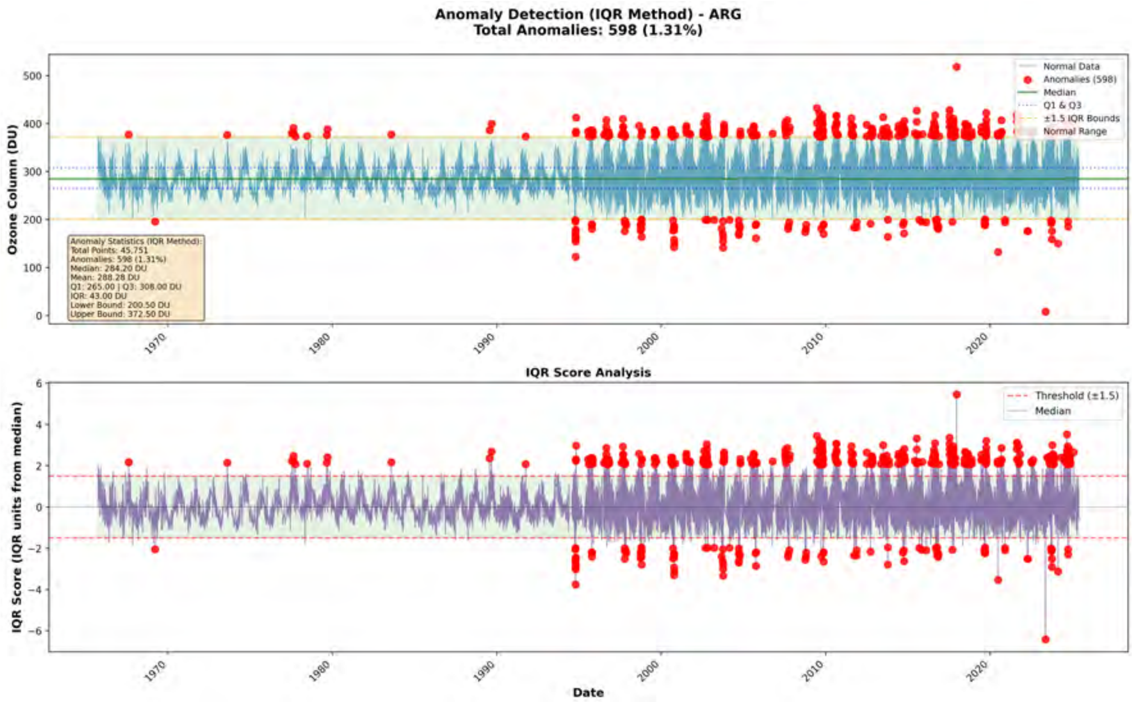

Figure 3.5: Anomaly Detection Using IQR Method of ARG Data

Daily ozone data in Argentina from 1965-2025 were analyzed with the Interquartile Range (IQR) approach in detecting anomalies in the data. The plot 3.5 indicates that the majority of the ozone values are within the normal range of between 200.5 DU and 372.5 DU, which are close to the median value of 284.20 DU. Among 45,751 observations, 598 values (457.31 per cent) were considered anomalies. Such anomalies are observed symmetrically both above and below the median, and therefore, there is no bias towards either high or low ozone levels. Although there are some exceptions in the previous decades, the majority of them are post the mid 1990s, which may indicate the increase in the variability of the data or in the density of measurement, or in the temporary atmospheric disruptions. In general, the low level of deviations can be concluded that the ozone record in Argentina can be taken as a well-behaved and rather stable.

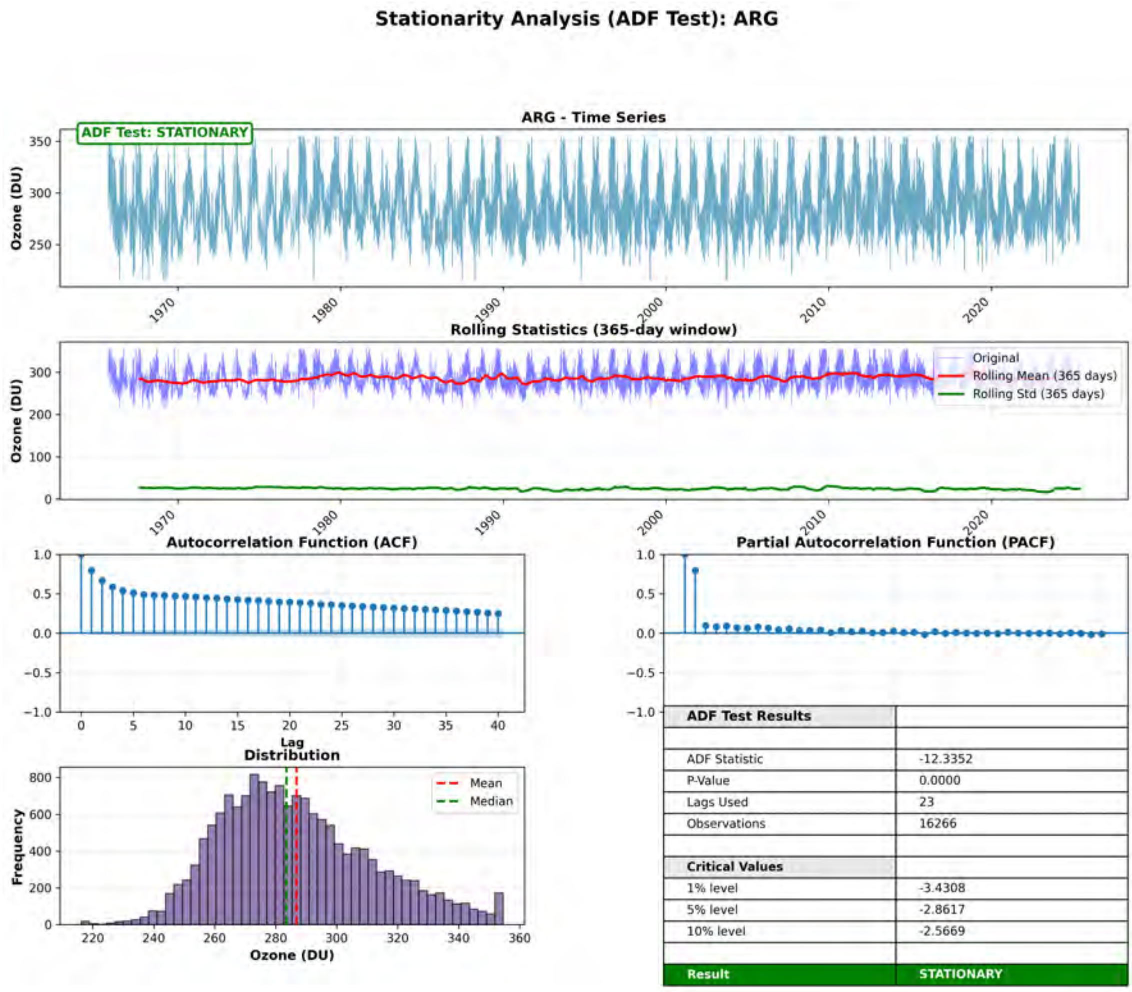

Figure 3.6: ADF-Based Stationarity Analysis of ARG Data

The ADF test was implemented to determine the stationarity of the daily ozone column data of the city of Argentina. The time series has distinct seasonal variations but no apparent long-term direction as shown in Figure 3.6. The null hypothesis of non-stationarity is decisively disapproved by the ADF statistic of -12.34, having a p-value of 0.0000 (significantly less than -3.43, which is the 1% critical value). The mean and the standard deviation plot do not change with time, meaning that there is no fluctuation in the mean and variance. Plots of ACF and PACF indicate a gradual decay, which is in line with a seasonal dependence and not movement due to a trend. Also, the distribution of the ozone is approximately normal, and the mean and the median are close. A combination of these findings proves that the ozone time series of Argentina is stationary and thus it can be used in time-series modeling, prediction, and anomaly analysis.

Antarctica (ATA):

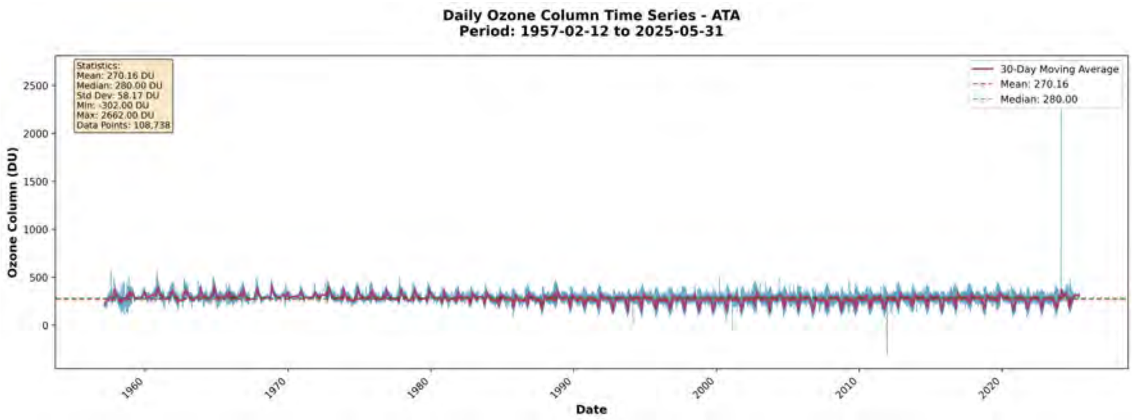

Figure 3.7: Temporal Variation of Daily Ozone Over The ATA

The daily ozone column of the ATA station between February 1957 and May 2025 is observed to have a robust and consistent seasonal cycle, which is due to the natural atmospheric processes like the sun radiation and the process of stratospheric circulation (Figure 3.7). The 30-day moving average indicates that there is no long-term upward or downward trend in the data despite the strong annual oscillations. The average ozone level is 270.16 DU, the median is 280.00 DU, and the standard deviation is 58.17 DU, showing moderate variation outweighed by seasonal and interannual fluctuations. The small lows in the 1980s -1990s could be linked with the global depletion of the ozone before the Montreal Protocol, and then after 2000, the levels level off. Such extreme values (-302 DU to 2662 DU) are physically implausible and must have been caused by instrument or data-entry errors. There is an overall picture of the ATA record of a constant, seasonal ozone column in a span of over 60 years.

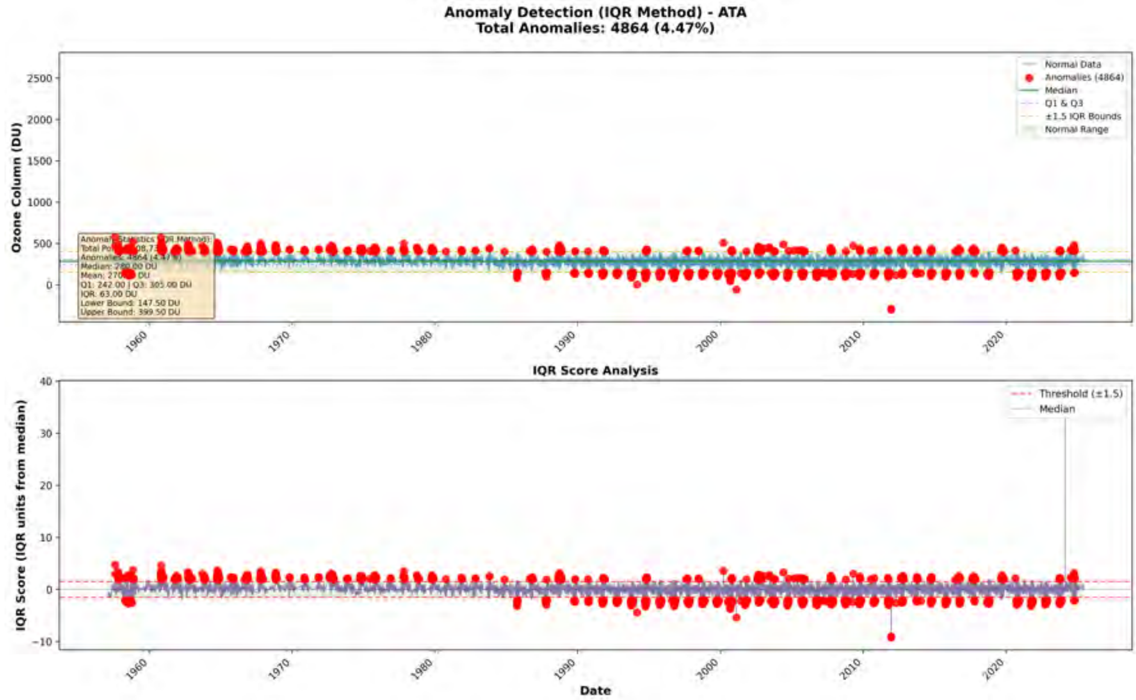

Figure 3.8: Anomaly detection using IQR method ATA data

The anomaly method of IQR detected 4,864 anomalies, which is 4.47 per cent of the entire observations shown in Figure 3.8. The majority of all the ozone values are in the normal range (147.5 -399.5 DU) with a median of about 288 DU. The main areas of anomalies are the initial decades (1950s-1970s), where there are significant positive spikes that are probably due to instruments limitations. Following the 1980s, the data settles, and there will be occasional exceptions in the later years, which could be associated with local atmospheric disturbances or anomalies in the data. The IQR score verifies these facts to be statistically extreme, revealing short-term changes and not the long-term systematic changes.

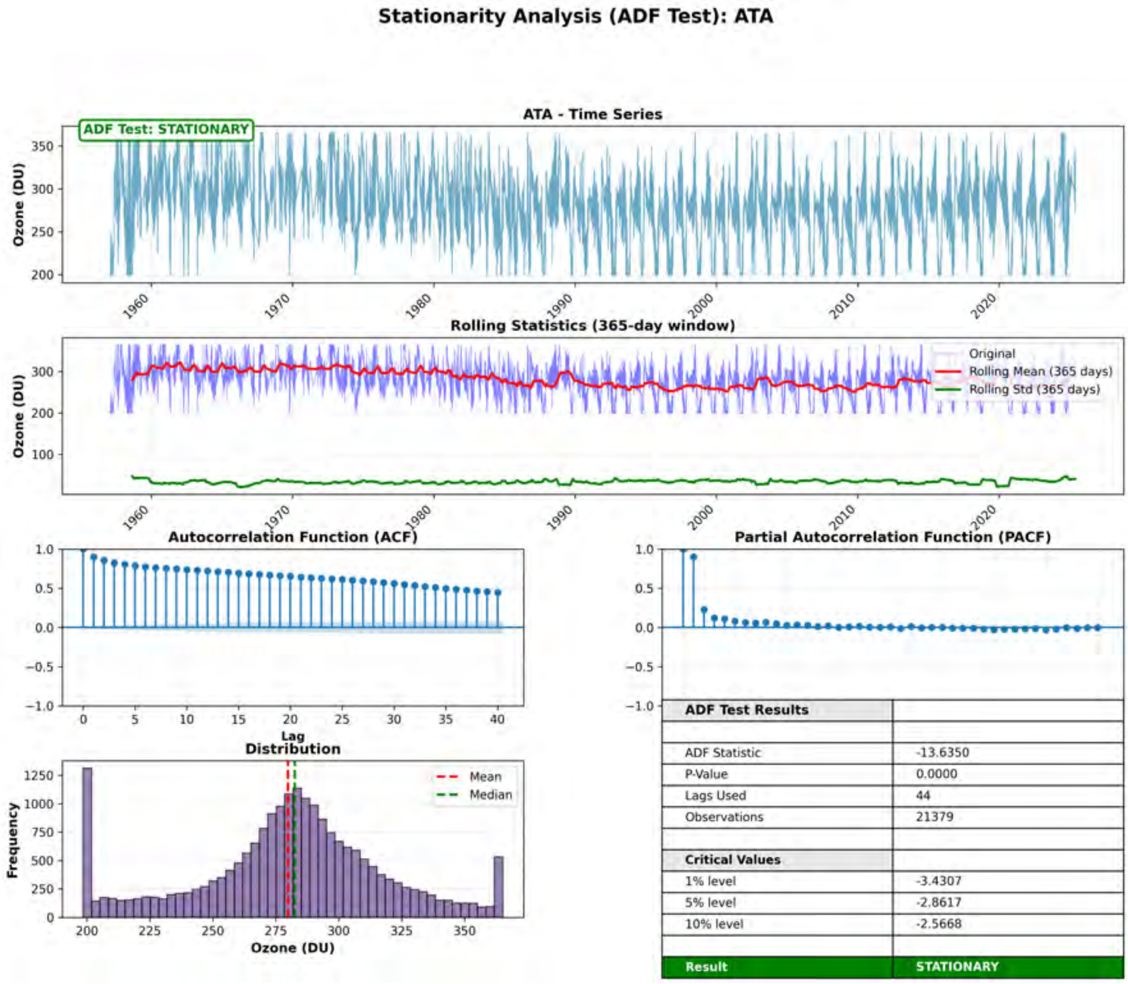

Figure 3.9: ADF-Based Stationarity Analysis ATA Data

There is a high and regular seasonal behavior of the ATA time series ozone with cyclical variations that have been observed in the time series (Figure 3.9). Rolling mean and standard deviation (365-day window) are stable, which means that there is no change in average levels and variance. The stationarity is forcefully supported in the Augmented Dickey Fuller (ADF = -13.64 and p= 0.0000). Normalized ACF and PACF curves indicate recurring patterns of correlation, which are in line with seasonal dependence, whereas the near-symmetric distribution implies long-term stability. The combination of these findings proves that the ATA ozone series is non-portable with strong seasonal characteristics, so it can be analyzed in a time series and assessed in terms of climate.

American Samoa (ASM):

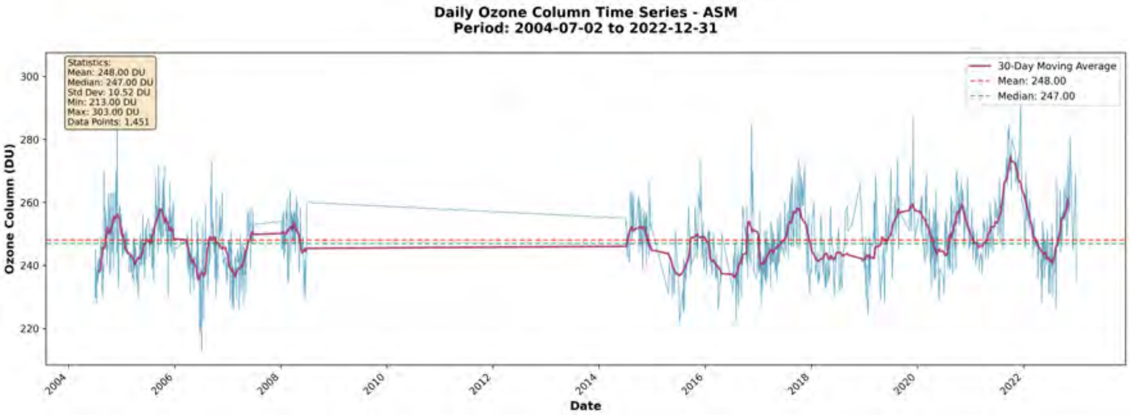

Figure 3.10: Temporal Variation of Daily Ozone Over The ASM

Figure 3.10 The 2004-2022 daily column of ozone above American Samoa (ASM). The column is moderately varied (standard deviation of 10.52 DU) around an unchanging mean of 248 DU and a median of 247 DU. The results of ozone are in the range of 213 DU to 303 DU, which means that there are some short-term extremes, but the entire results are consistent. There is a noticeable seasonal cycle that is determined by the changes in the sun's radiation and circulation in the atmosphere. These seasonal oscillations can be noted in the 30-day moving average, and the fact that there is no long-term increasing or decreasing trend can be established. On the whole, ozone concentrations above ASM were stable throughout the study time, and seasonal and short-term atmospheric processes are the predominant variants.

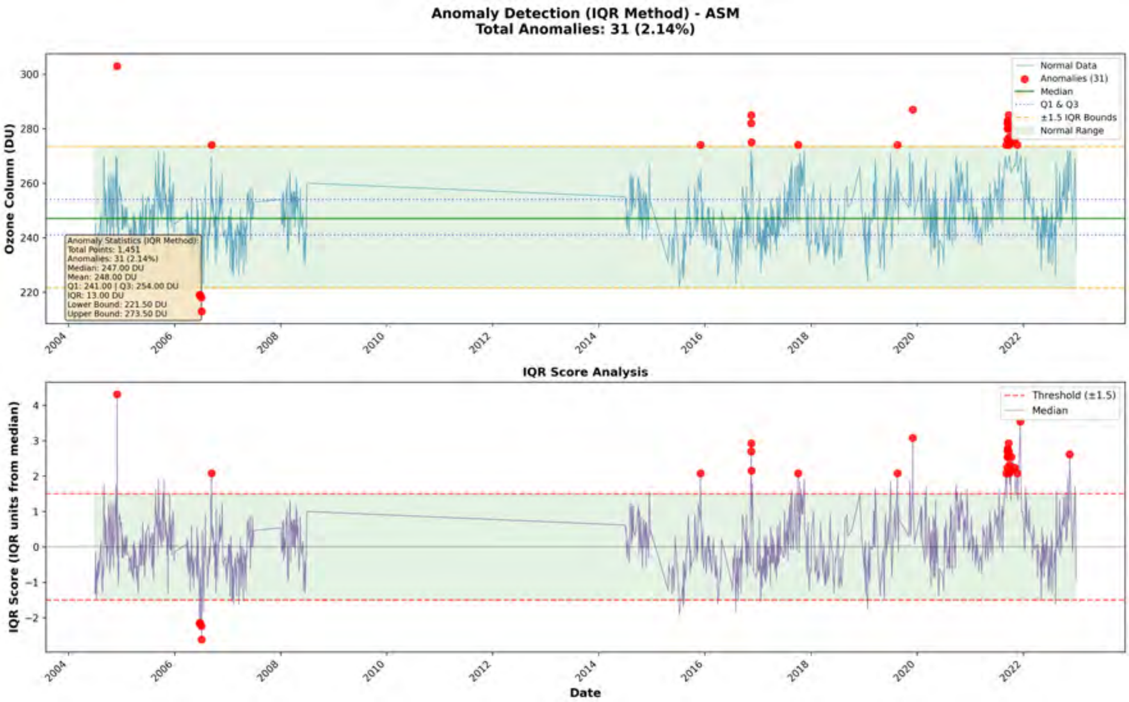

Figure 3.11: Anomaly Detection Using IQR Method ASM Data

With the IQR approach, it was found that 31 anomalies (2.14 percent) existed among 1,451 observations with the median of the ozone distribution equaling 247 DU (Figure 3.11) . Abnormalities are not within a span of about 221.5-273.5 DU and do not manifest itself in an equal manner with time. The initial years (20042007) indicate that there are both positive and negative anomalies, whereas the years 2010-2014 are considerably stable. The frequency of anomalies rises since 2016, especially between 2020 and 2022, with the majority of the frequency showing positive deviations. These incidences notwithstanding, the low anomaly percentage shows that the record of ozone was quite stable but that there were short term atmospheric disturbances.

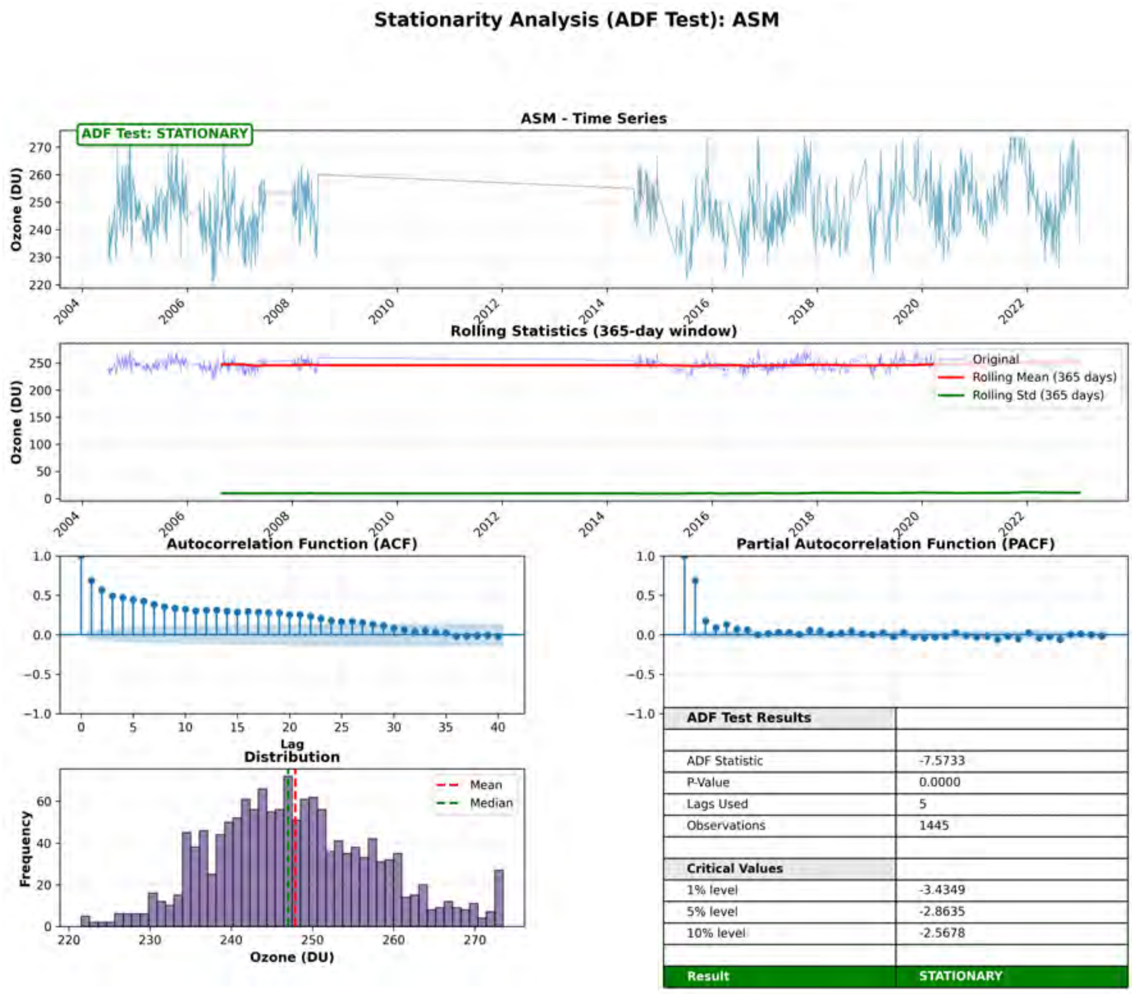

Figure 3.12: ADF-Based Stationarity Analysis of ASM Data

The Augmented Dickey Fuller test will lead to the rejection of the null hypothesis that the ASM ozone series is non-stationary (ADF $= -7.5733$, p $= 0.0000$). The rolling mean and standard deviation do not change over time and there is no long-term trend. The ACF and PACF curves indicate slow decay that is in line with seasonal dependence (Figure 3.12). The fact that the mean and median are very close to each other also suggests statistical stability, which is additionally a benefit of the near normal distribution, and hence the dataset is ready to be used in time-series modelling and anomaly analysis.

Australia (AUS):

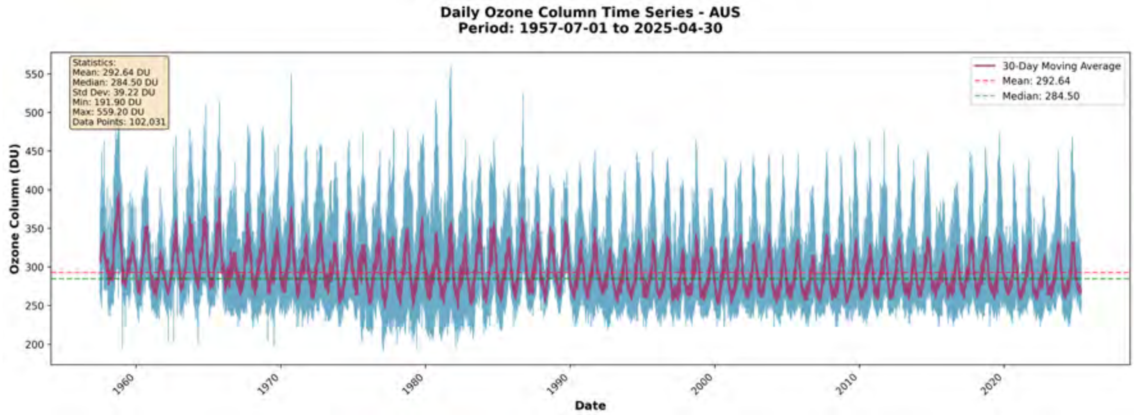

Figure 3.13: Temporal Variation of Daily Ozone Over AUS

The ozone column in every day above Australia between 1957 and 2025 has a very sharp and prominent seasonal cycle with annual oscillations that are regular in the record as shown in the figure 3.13. The values of ozone are usually between 550 DU and 200 DU and the average value of the ozone is 292.64 DU and the median of the ozone is 284.50 DU. These variations are natural seasonal variations that are caused by solar radiation and circulation in the atmosphere. There is relative stability in long-term behavior in the 1950s-1970s, and the occurrence of a significant decrease in long-term behavior mid-1970s to early 1990s with the rise in CFC emissions. Following the adoption of the Montreal Protocol the ozone levels leveled, and indicate slow progression towards recovery since the beginning of the 2000s. In general, the time series is a projection of the interactive effect of seasonal and anthropogenic effects on ozone in Australia.

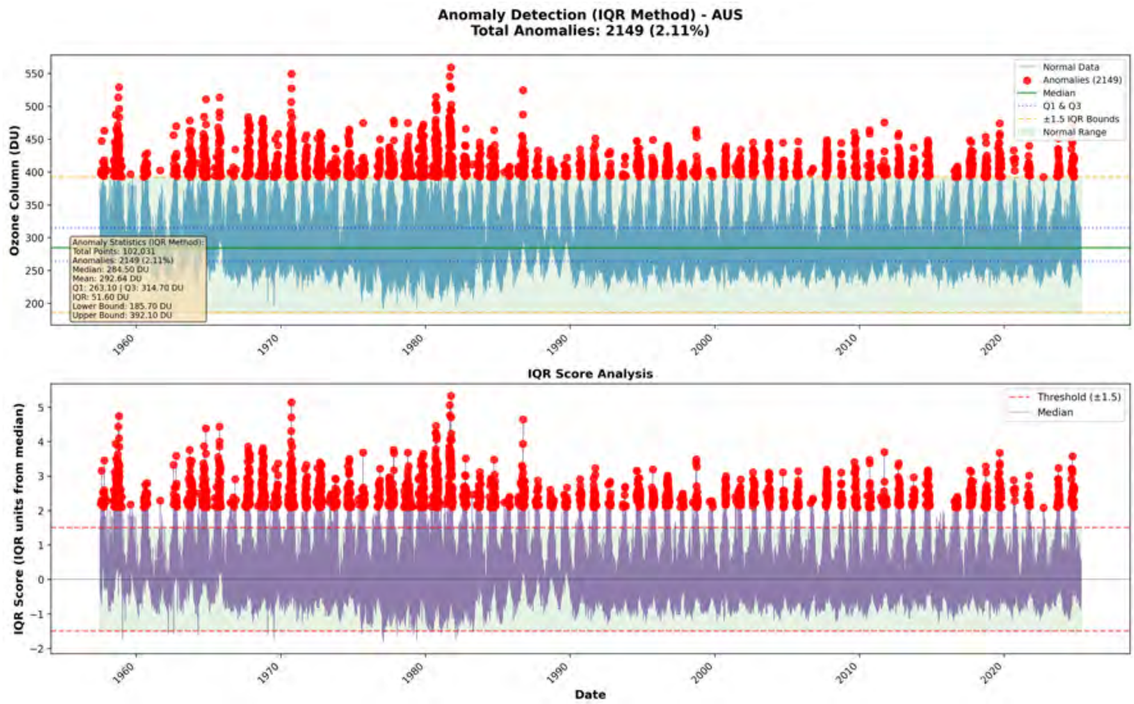

Figure 3.14: Anomaly Detection Using IQR Method AUS Data

Detection of anomalies by the IQR technique found 2,149 anomalies (2.11%) in 102,031 observations, as seen in Figure 3.14. The median of the ozone distribution is 284.59 DU with normal limits of about 185.7 DU and 392.1 DU. More frequent occurrences of extreme anomalies (>450 DU) are seen prior to 1985, which means that the past decades are more variable or may have measurement errors. Since the mid-1980s, the tendency to anomalies decreases, and the values of the ozone are more and more restricted to the normal threshold. This stabilization, especially since 2000, indicates a better state of ozone and a more consistent measurement, which is probably due to international environmental standards.

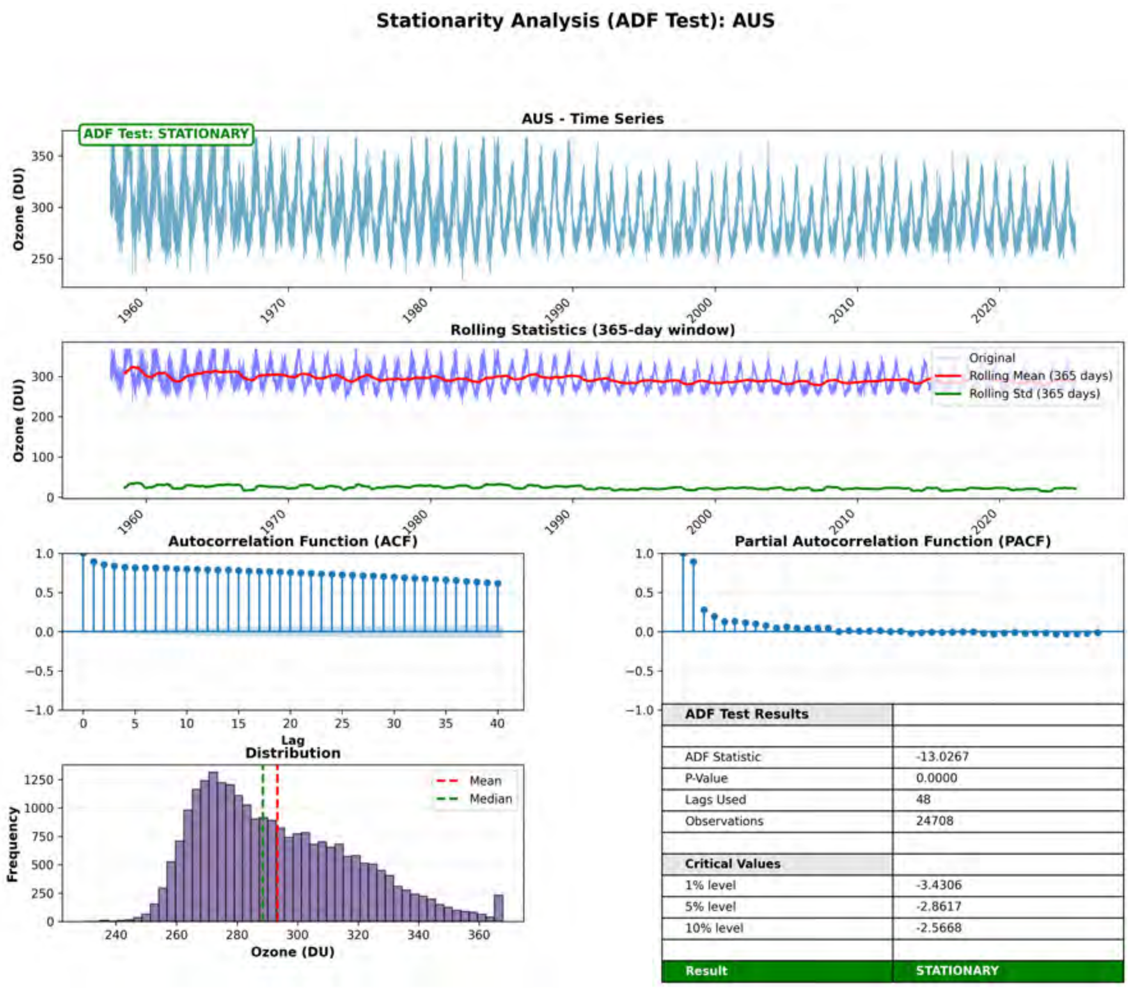

Figure 3.15: ADF-Based Stationarity Analysis of AUS Data

The Australian ozone time series indicates that the time series is strongly and consistently seasonal, with annual cycles being constant with time (Figure 3.15). The rolling mean and standard deviation (365-day window) show no change in value, which means that there is no change in the average value and the variance. The Augmented DickeyFuller test assumes the presence of stationarity(ADF = -13.03, p = 0.01). The ACF and PACF plots show regular correlation trend that is indicative of seasonal dependence. All these findings suggest that the variability of ozone across Australia is controlled by predictable seasonal processes in a statistically stationary framework.

Brazil (BRA):

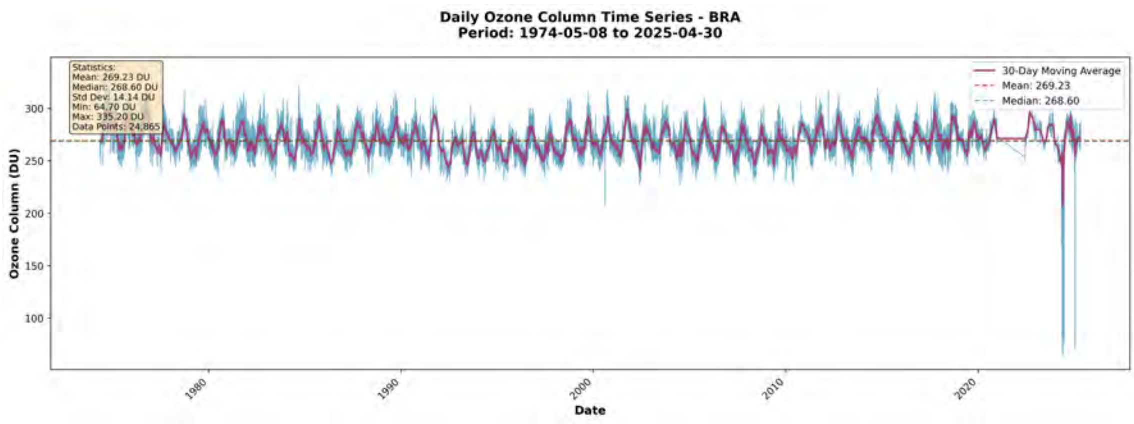

Figure 3.16: Temporal Variation of Daily Ozone Over BRA

The daily ozone column in Brazil (1974-2025) temporally shows that there is a regular seasonal cycle, and the annual increase and reduction are associated with the changes in the sunlight and atmospheric circulation (Figure 3.16). The median (268.60 DU) and the mean (269.23 DU) values are evenly distributed without any skew and the 30-day moving average indicates the annual pattern. These seasonal periods are mainly influenced by changes in solar radiation, temperature and transport of the ozone to the stratospheric place. The general direction indicates that variations in ozone are controlled by natural processes in the atmosphere. Although there are some small changes, there is no long-term trend, which suggests that the levels of ozone do not change over time.

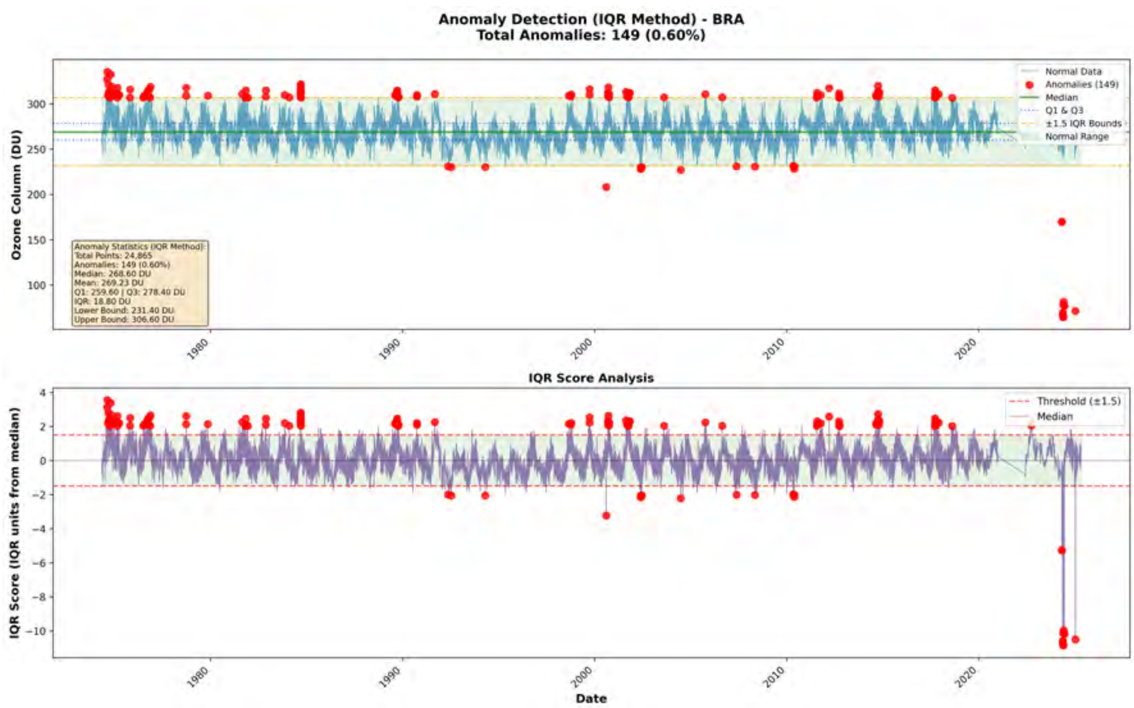

Figure 3.17: Anomaly Detection Using IQR Method BRA Data

According to the Interquartile Range (IQR) technique, there were 149 anomalies (0.60 percent) of 24,865 data points (Figure 3.17). It was computed that the upper and lower limits were 306.60 DU and 231.40 DU, respectively. The majority of anomalies are detected at the extremes, which are the occasional spikes or falls, particularly, between 2020-2025, which are probably caused by short-term atmospheric disruptions or measurement errors. In general, all these anomalies are small, which proves the reliability of the dataset and supports the existence of stable seasonal behavior over decades.

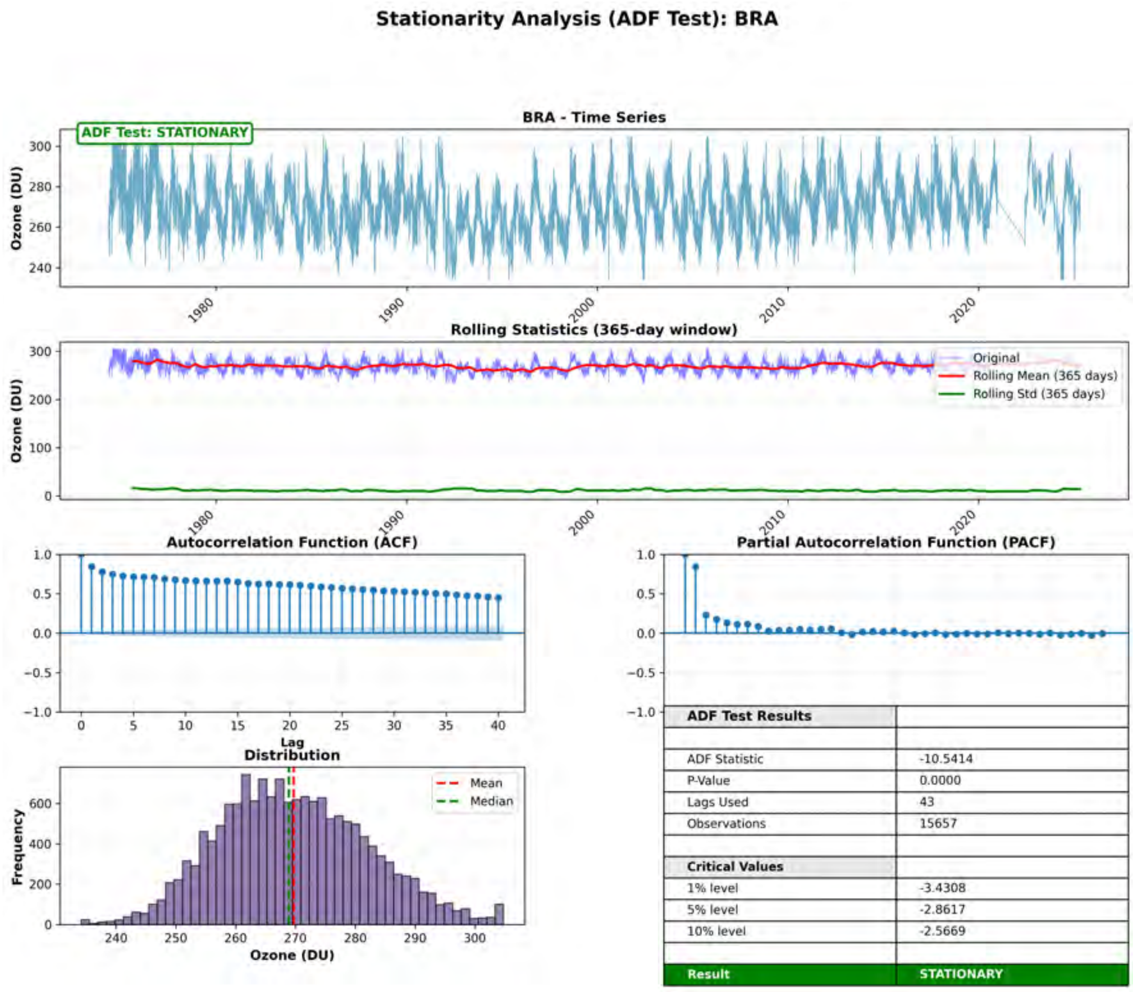

Figure 3.18: ADF-Based Stationarity Analysis of BRA Data

The ADF test verifies that the series of ozone is stationary (Figure 3.18)) and the ADF statistic = -10.5414, the p-value = 0.0000 (less than 1%). Rolling mean and standard deviation values are fixed and ACF/PACF charts depict recurring seasonal cycles with a decay effect. The fact that the distribution of the mean is close to the normal distribution also contributes to the stability of the data across time, meaning that the ozone data have the same statistical characteristics across the time interval.

Canada (CAN):

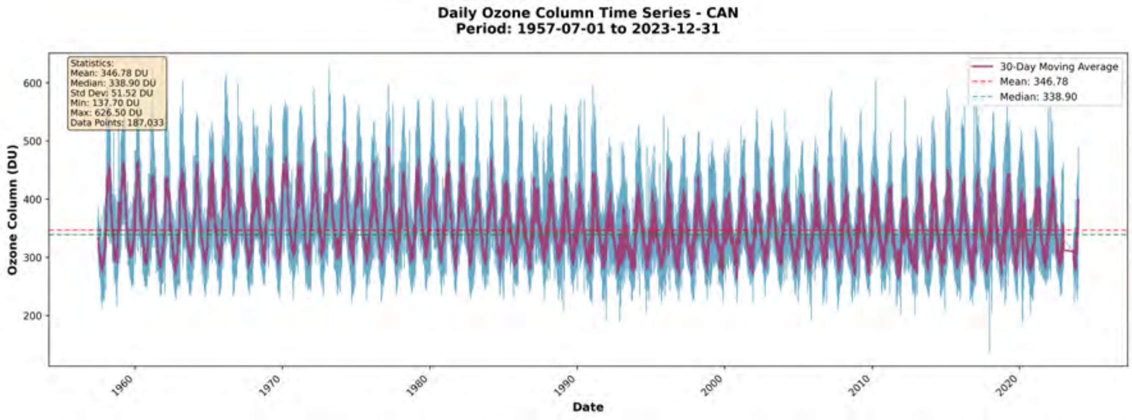

Figure 3.19: Temporal Variation of Daily Ozone Over CAN

The temporal examination of the daily ozone column over Canada (1957–2023) demonstrates a clear and very regular seasonal cycle, with repeated peaks and troughs correlating to annual fluctuations in sunlight and stratospheric circulation (Figure 3.19). The mean ozone concentration is 346.78 DU, and the median is 338.90 DU, demonstrating a near-symmetric distribution with small positive skew from high outliers. The yearly pattern of ozone change is vividly highlighted by the 30-day moving average, which smoothes short-term fluctuations. Natural elements like air movement and solar radiation intensity are the main causes of these variations. Overall, there is no discernible long-term trend, indicating that ozone levels over Canada have mostly been constant over the years.

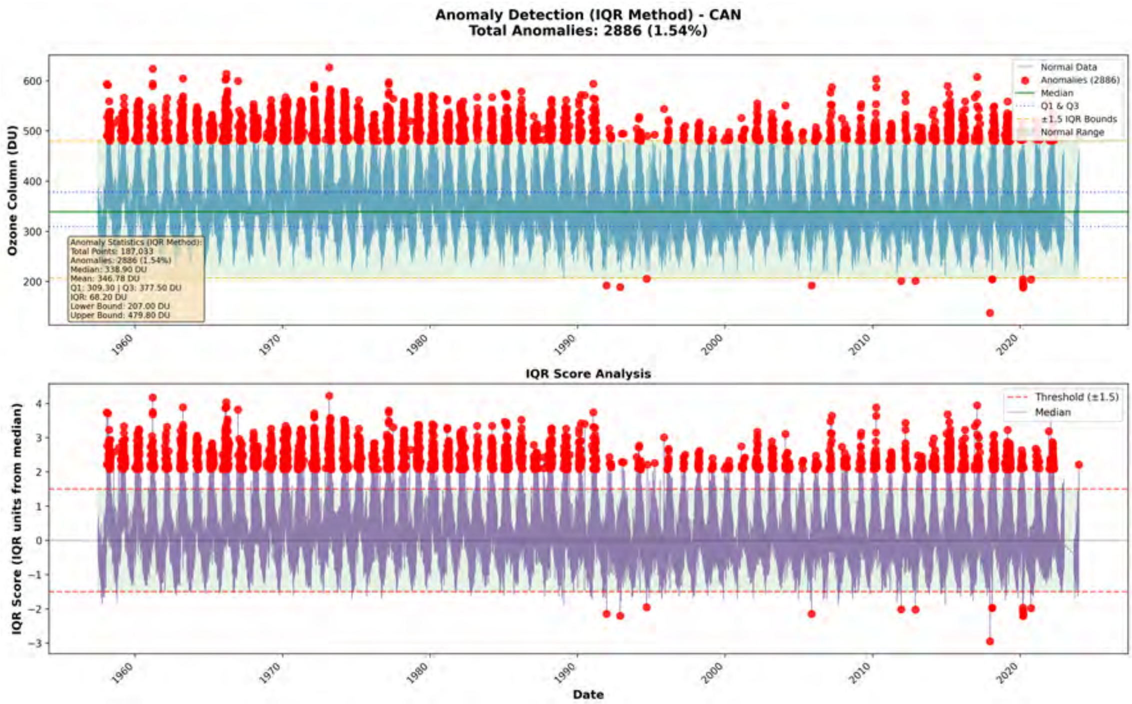

Figure 3.20: Anomaly Detection Using IQR Method CAN Data

Using the Interquartile Range (IQR) approach, 2,886 anomalies (1.54%) were discovered from 187,033 data points (Figure:3.20),using 207.00 DU and 479.80 DU as the lower and upper boundaries. The majority of anomalies happen near the extremities of the distribution, either as unusual low troughs in particular decades or as spikes during seasonal peaks. Short-lived atmospheric events or measurement errors are probably the cause of certain high-value anomalies. Despite these occurrences, anomalies are modest overall, demonstrating the dataset's stability and the great seasonal regularity of ozone over Canada.

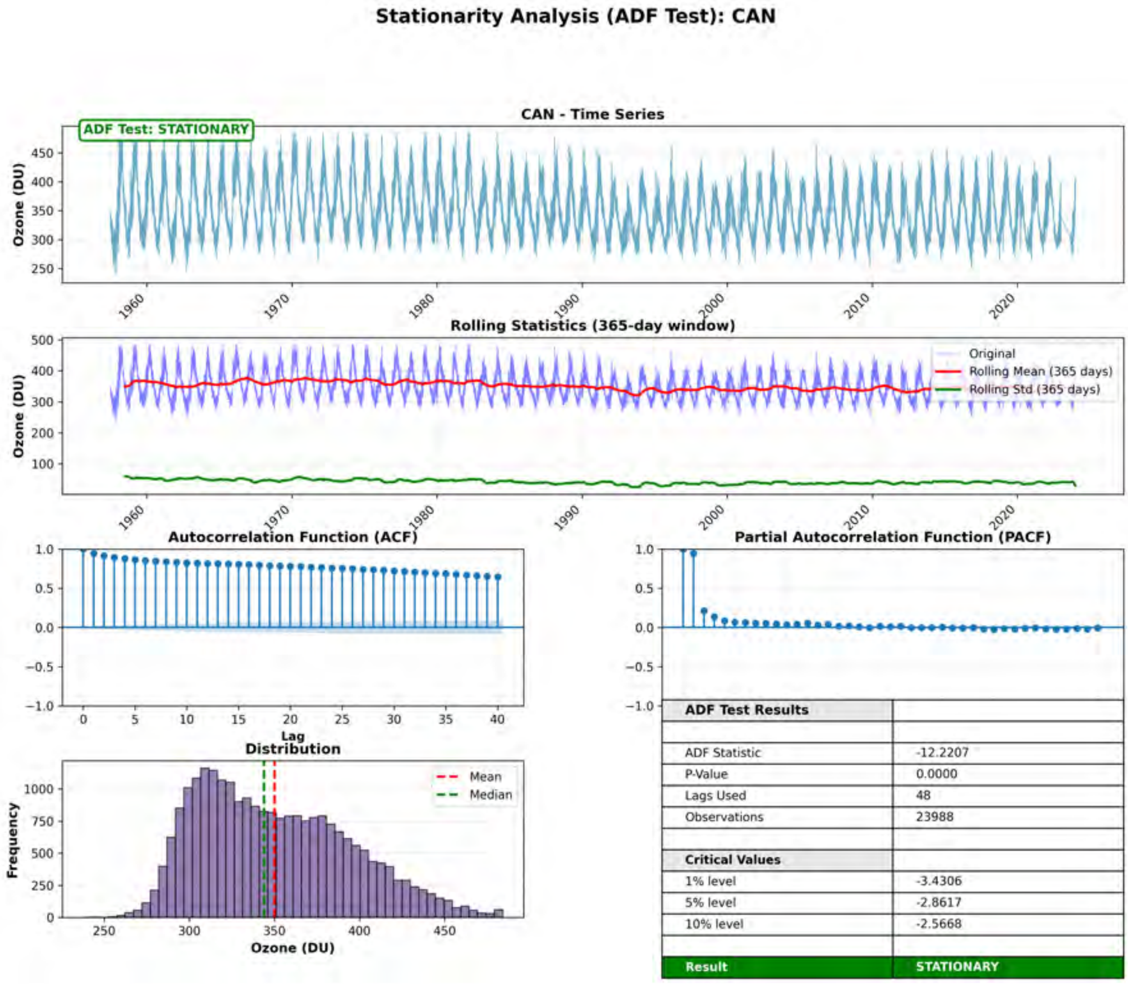

Figure 3.21: ADF-Based Stationarity Analysis of CAN Data

The ozone time series' stationarity is verified by the Augmented Dickey-Fuller (ADF) test. (Figure 3.21),with a p-value of 0.0000 (less than 1%) and an ADF statistic of -12.2207. ACF and PACF plots exhibit considerable seasonal dependency without long-term drift, although the rolling mean and standard deviation remain almost constant. The constancy of ozone levels over time is further supported by the unimodal, near-normal distribution centered around the mean.

China (CHN):

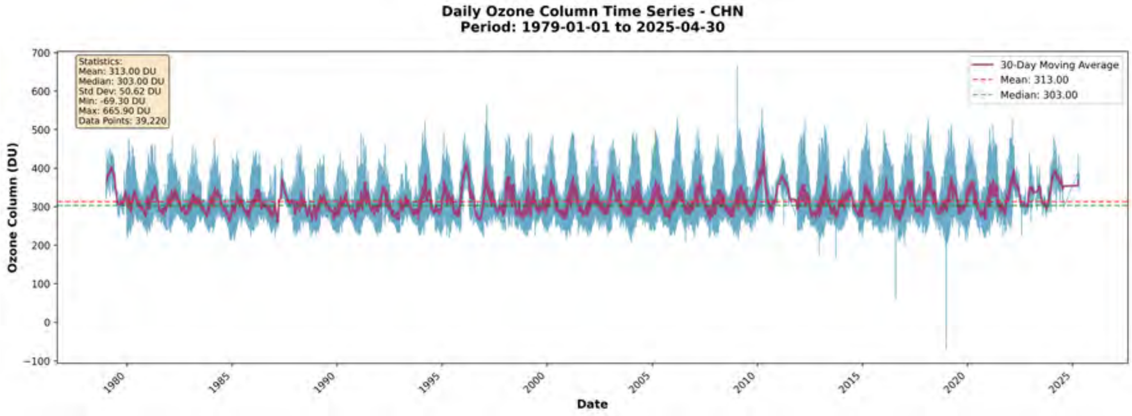

Figure 3.22: Temporal Variation of Daily Ozone Over CHN

The temporal analysis of the daily ozone column over CHN (Figure:3.22), spanning from 1979-01-01 to 2025-04-30, reveals a pronounced and consistent seasonal cycle. Ozone levels exhibit sharp annual peaks (often exceeding 400 DU) and troughs (dipping below 250 DU), tightly aligned with hemispheric solar and circulation patterns. The mean ozone is 313.00 DU and the median 303.00 DU, with a 30-day moving average effectively filtering noise to highlight smooth, recurring oscillations. Long-term inspection shows no significant upward or downward trend—the series remains centered near the mean across decades, indicating high temporal stability despite interannual variability in peak intensity.

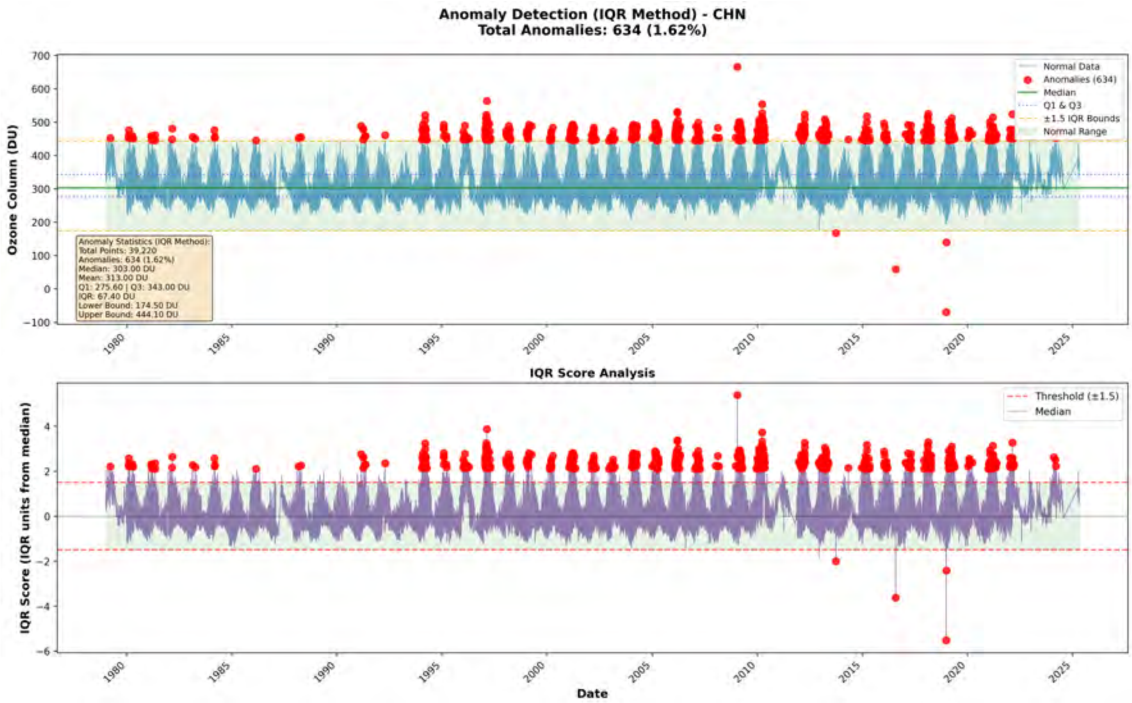

Figure 3.23: Anomaly Detection Using IQR Method CHN Data

The IQR method identified 634 anomalies (1.62%) from 39,220 points (Figure 3.23). Bounds: lower 174.50 DU, upper 444.10 DU. Anomalies cluster at extremes at seasonal extremes; notable low spikes post-202 used 2020–2025-only 1.62% outliers confirm overall stability.

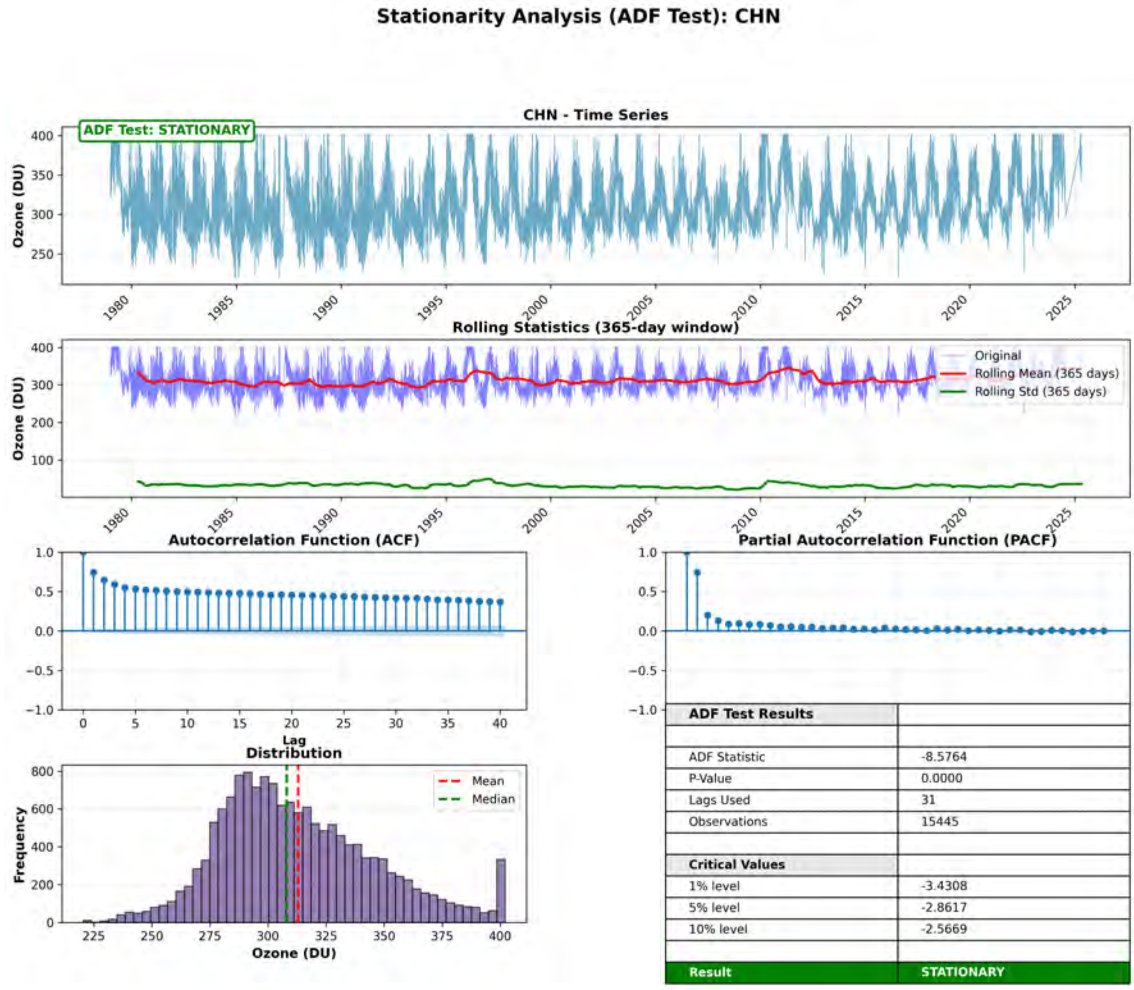

Figure 3.24: ADF-Based Stationarity Analysis of CHN Data

The ADF test confirms stationarity with a statistic of -8.5764 and a p-value of 0.0000 (well below 1% level). Rolling mean and standard deviation (365-day window) remain nearly flat over time. ACF shows persistent seasonal correlation; PACF decays rapidly (Figure 3.24). Distribution is near-normal, centered at the mean, reinforcing that the series is stationary.

Greenland (GRL):

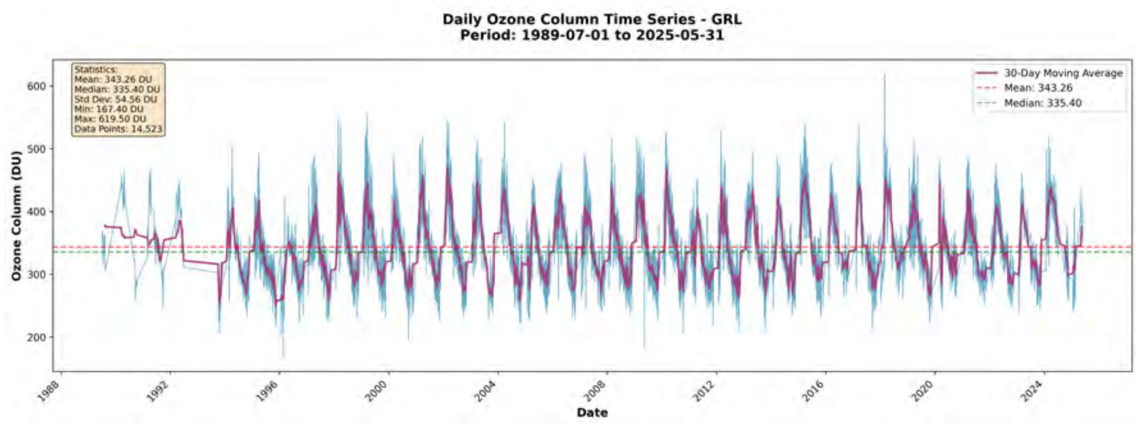

Figure 3.25: Temporal Variation of Daily Ozone Over GRL Data

The temporal analysis of the daily ozone column over Greenland (1989–2025) reveals a strong and consistent seasonal cycle. Ozone levels peak above 500 DU during spring and fall below 250 DU in autumn, reflecting the influence of polar stratospheric dynamics and solar variability. The mean ozone level is 343.26 DU, and the median is 335.40 DU, showing a slightly right-skewed distribution due to occasional high values. The 30-day moving average smooths short-term fluctuations and highlights the steady annual rhythm (Figure 3.25). No significant long-term trend is observed, indicating stable ozone behavior over time with only minor interannual variations linked to Arctic stratospheric changes.

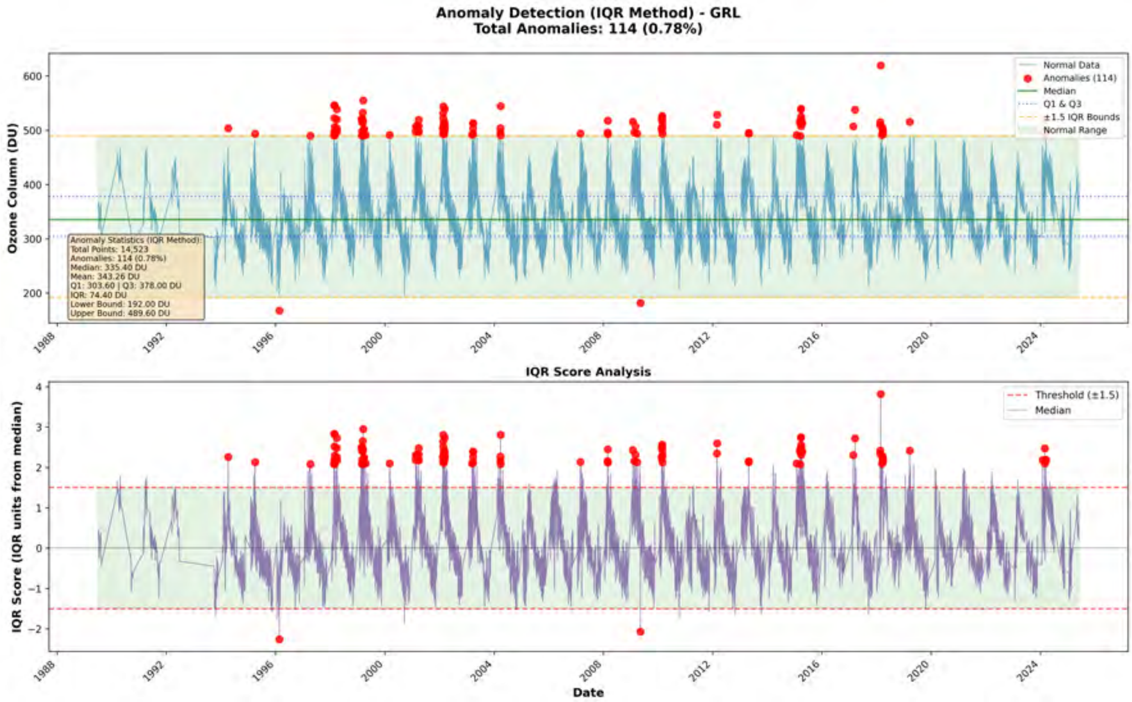

Figure 3.26: Anomaly Detection Using IQR Method GRL Data

Using the IQR method, 114 anomalies (0.78%) were detected from 14,523 data points, with lower and upper bounds at 192.20 DU and 489.60 DU (Figure 3.26). Most anomalies appear as high-value spikes during spring ozone peaks, with few low anomalies early in the record. As these rare outliers represent less than 1% of the data, the overall ozone pattern remains stable and consistent, reflecting predictable seasonal variation across the decades.

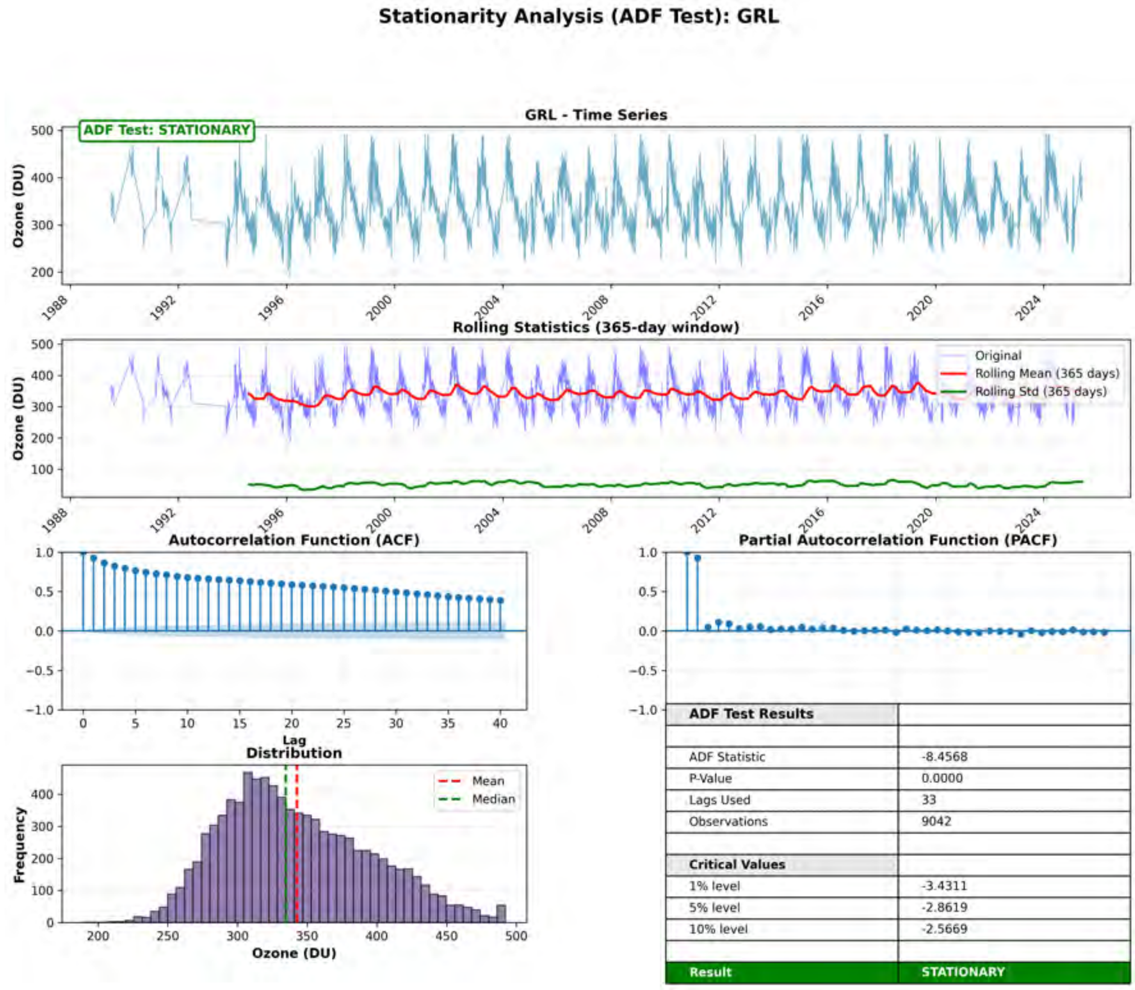

Figure 3.27: ADF-Based Stationarity Analysis of GRL Data

The ADF test confirms stationarity with a statistic of -8.4568 and a p-value of 0.0000 (below the 1% threshold). Rolling mean and standard deviation stay steady throughout the dataset, while ACF and PACF plots reveal strong annual periodicity with fast decay. The distribution is almost symmetric around the mean, confirming a stable and time-invariant ozone pattern over Greenland (Figure 3.27).

India (IND):

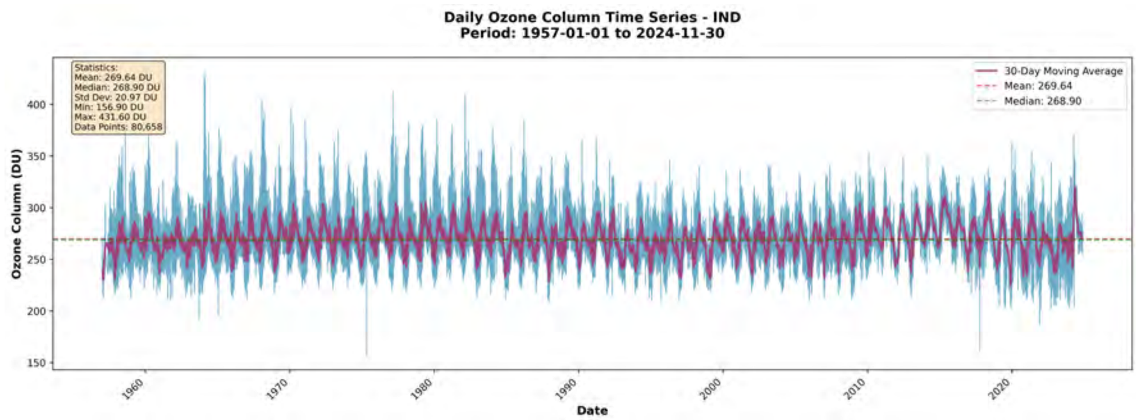

Figure 3.28: Temporal Variation of Daily Ozone Over IND

The temporal analysis of daily total ozone over India (1957–2024) shows a clear and consistent seasonal cycle (Figure 3.28). Ozone levels peak and dip annually in sync with solar radiation and stratospheric circulation. The mean (269.64 DU) and median (268.90 DU) are nearly identical, indicating symmetry around the average. This figure shows how the 30-day moving average smooths fluctuations, revealing a stable annual rhythm. Long-term inspection demonstrates no significant trend, confirming strong temporal stability across decades.

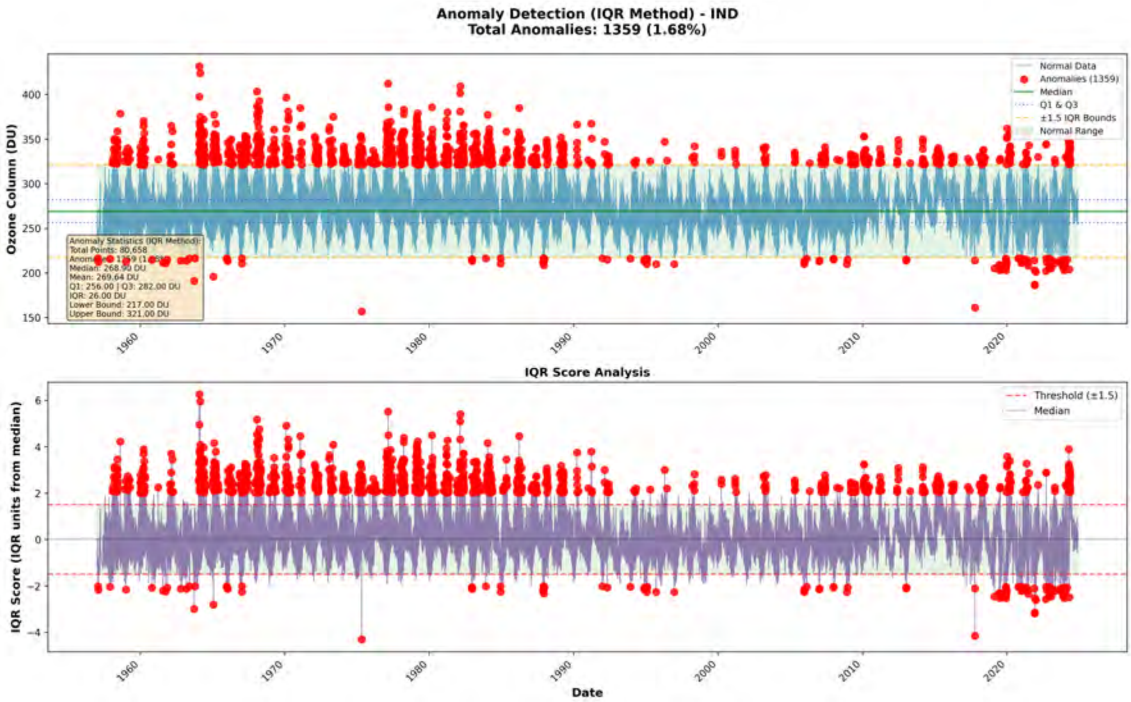

Figure 3.29: Anomaly Detection Using IQR Method IND Data

The IQR method detected 1,359 anomalies (1.68%) from 80,658 observations (Figure 3.29), with bounds at 217.00 DU (lower) and 321.00 DU (upper).With a few smaller spikes in recent years, this figure demonstrates that the majority of anomalies occur at extreme highs and lows. In spite of this, India's general ozone stability is highlighted by the tiny anomaly rate.

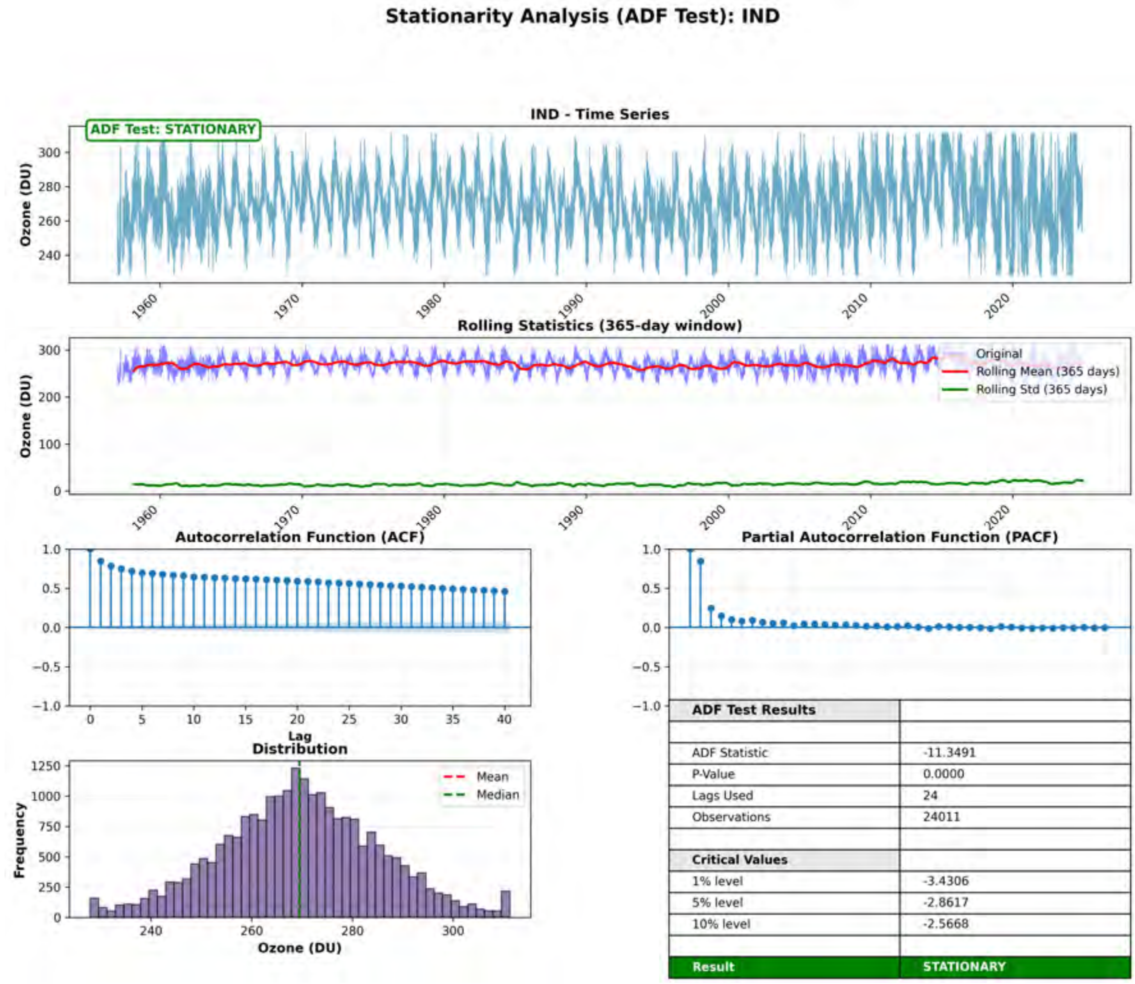

Figure 3.30: ADF-Based Stationarity Analysis of IND Data

The ADF test confirms stationarity with a statistic of -11.3491 and a p-value of 0.0000 (Figure 3.30). The ACF and PACF plots in this figure show recurrent seasonal correlations with slow degradation, but the rolling mean and standard deviation remain almost constant. The near-normal distribution centered at the mean further shows that the series is stationary.

Mexico (MEX):

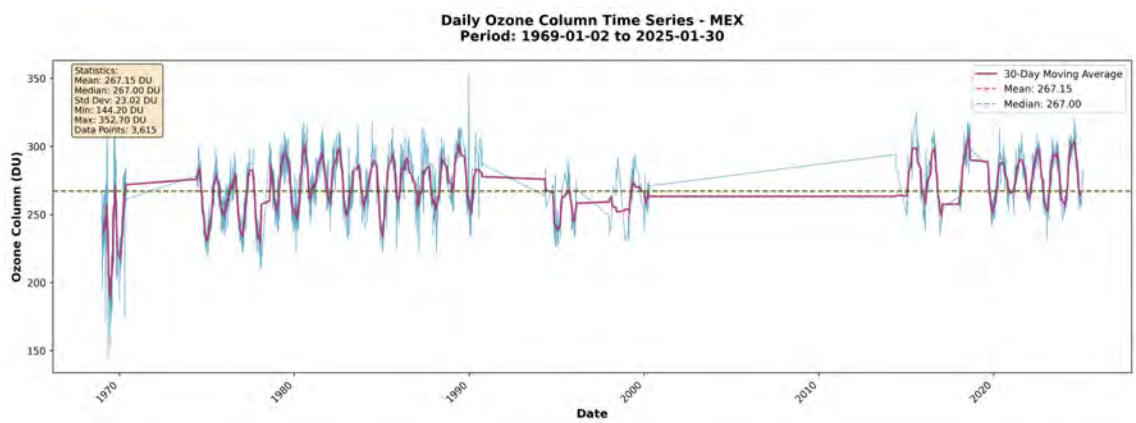

Figure 3.31: Temporal Variation of Daily Ozone Over MEX

The temporal analysis of daily ozone over Mexico (1969–2025) shows irregular fluctuations without a clear seasonal cycle (Figure 3.31). Ozone levels fluctuate significantly, with sudden drops and increases, particularly in the early years. A balanced distribution with significant short-term variability is indicated by the mean (267.15 DU) and median (267.00 DU). The analysis shows sporadic cyclic behavior that is impacted by outside variables like atmospheric circulation or solar radiation. Ozone levels are statistically steady but dynamically changing throughout time, as seen by fluctuations that don't follow a consistent pattern.

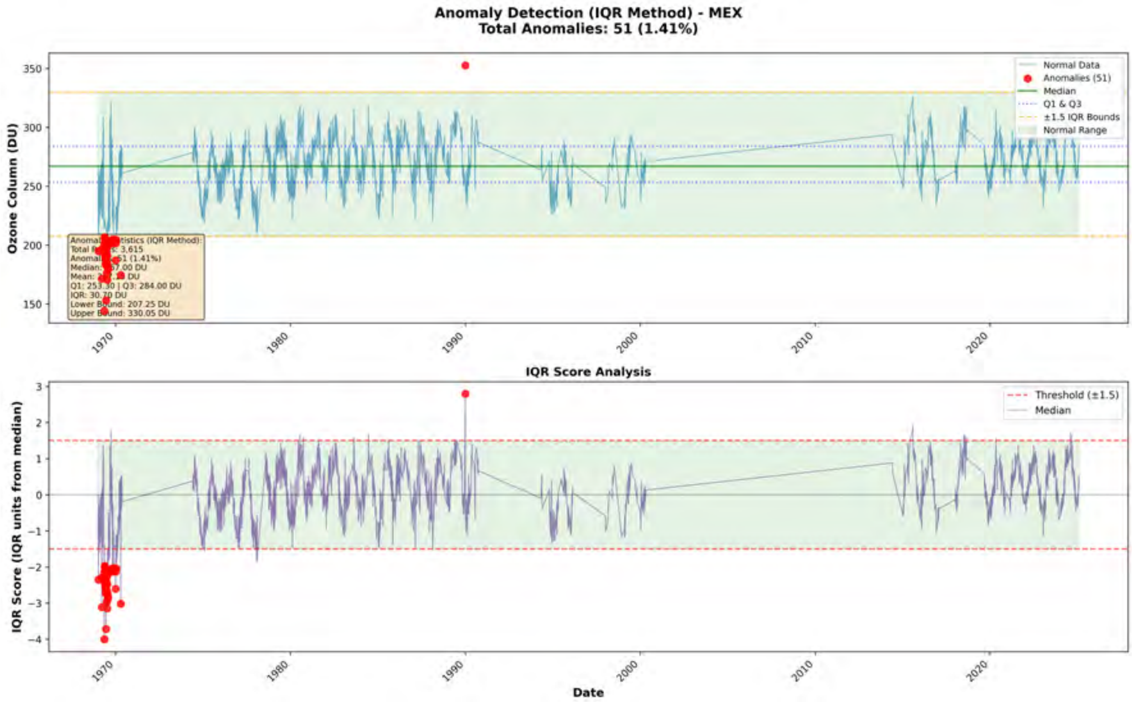

Figure 3.32: Anomaly Detection Using IQR Method MEX Data

Using the IQR method, 51 anomalies (1.41%) were found among 3,615 observations (Figure 3.32).The majority of anomalies with abnormally low ozone (<200 DU) happened in the early 1970s, most likely as a result of sporadic occurrences or inconsistent data. Short-lived deviations can be seen in a few later anomalies around 1989 and around 2000. In spite of this, the dataset is essentially unchanged, indicating limited and transient instability rather than long-term changes.

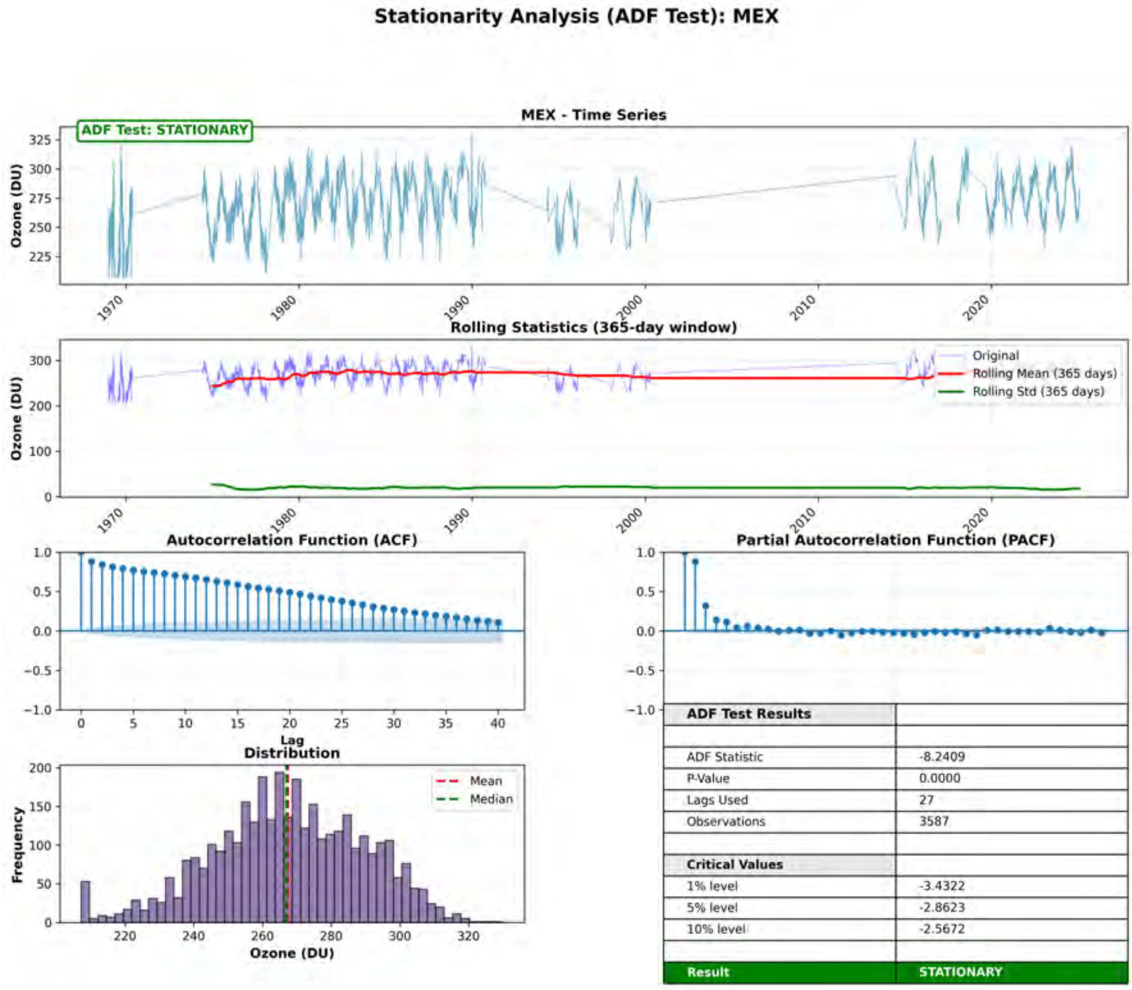

Figure 3.33: ADF-Based Stationarity Analysis of MEX Data

The ADF test (ADF = -8.2409, p = 0.0000) confirms the series is stationary, with stable mean and variance across decades (Figure 3.12). ACF/PACF plots exhibit considerable decline, indicating transient correlations, whereas the rolling mean and standard deviation are almost constant. Only short-term environmental changes have an impact on ozone dynamics, as indicated by the near-normal distribution centered at the mean.

New Zealand (NZL):

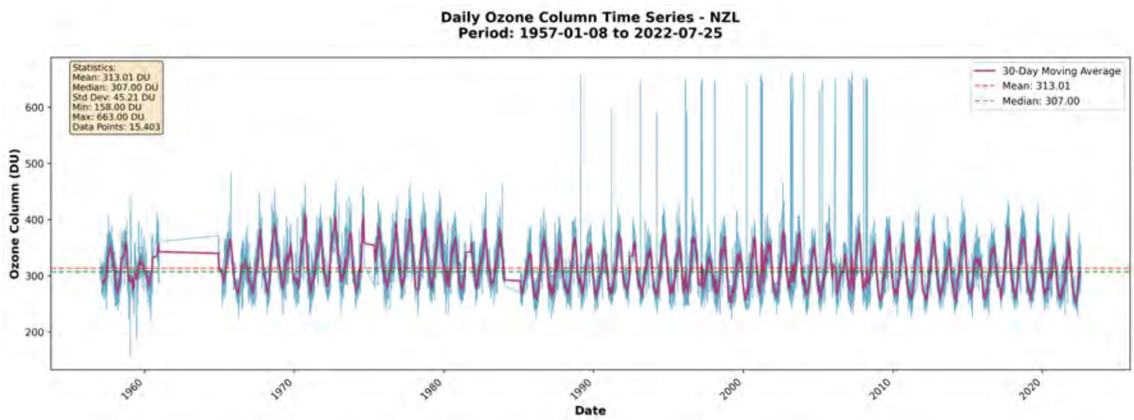

Figure 3.34: Temporal Variation of Daily Ozone Over NZL

The temporal analysis of daily ozone over New Zealand (1957–2022) shows a strong, repeating seasonal cycle with clear annual peaks and troughs linked to solar radiation and atmospheric circulation (Figure 3.34). There is symmetry around the mean, with a mean of 313.01 DU and a median of 307.00 DU. With ozone ranging from 158 DU to 663 DU, the 30-day moving average shows a steady yearly pattern. After the 1990s, there is a little amplitude rise but no long-term trend, indicating temporal stability over decades.

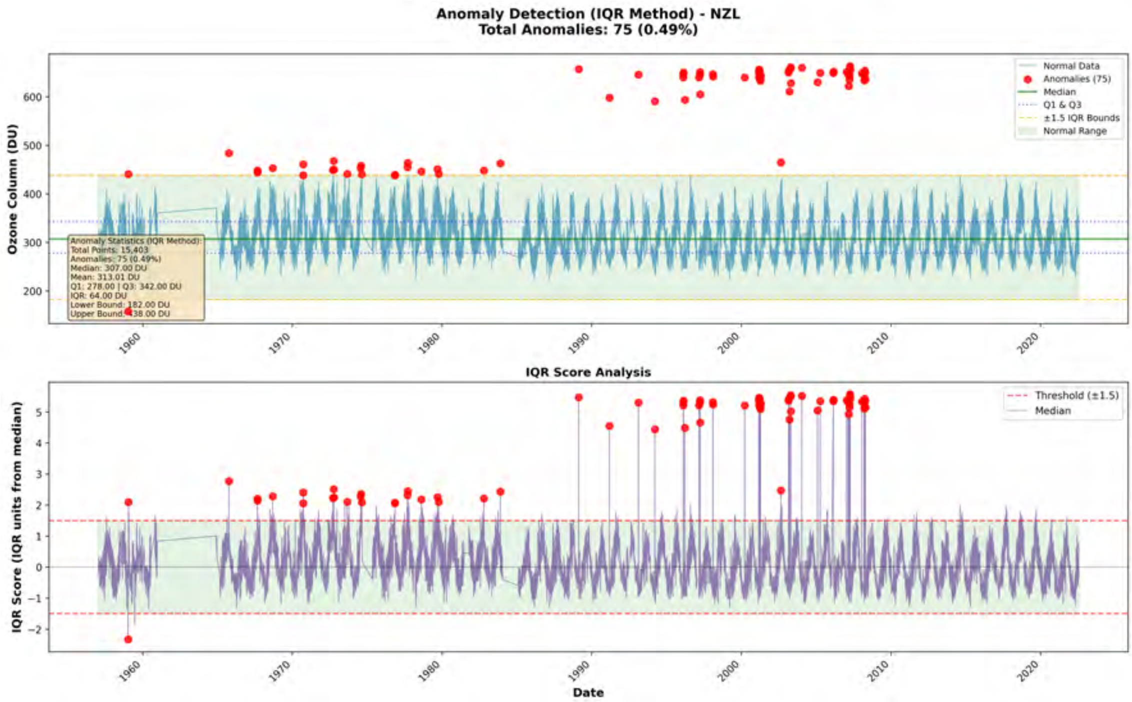

Figure 3.35: Anomaly Detection Using IQR Method NZL Data

Using the IQR method, 75 anomalies (0.49%) were detected from 15,403 points (Figure 3.35), with bounds at 184 DU (lower) and 403 DU (upper). The majority of anomalies are high-value spikes above 600 DU, mostly from short-term stratospheric occurrences that occurred in the late 1990s and early 2000s. Early in the record, there are a few low anomalies. The low anomaly rate confirms high ozone stability over New Zealand with only rare deviations.

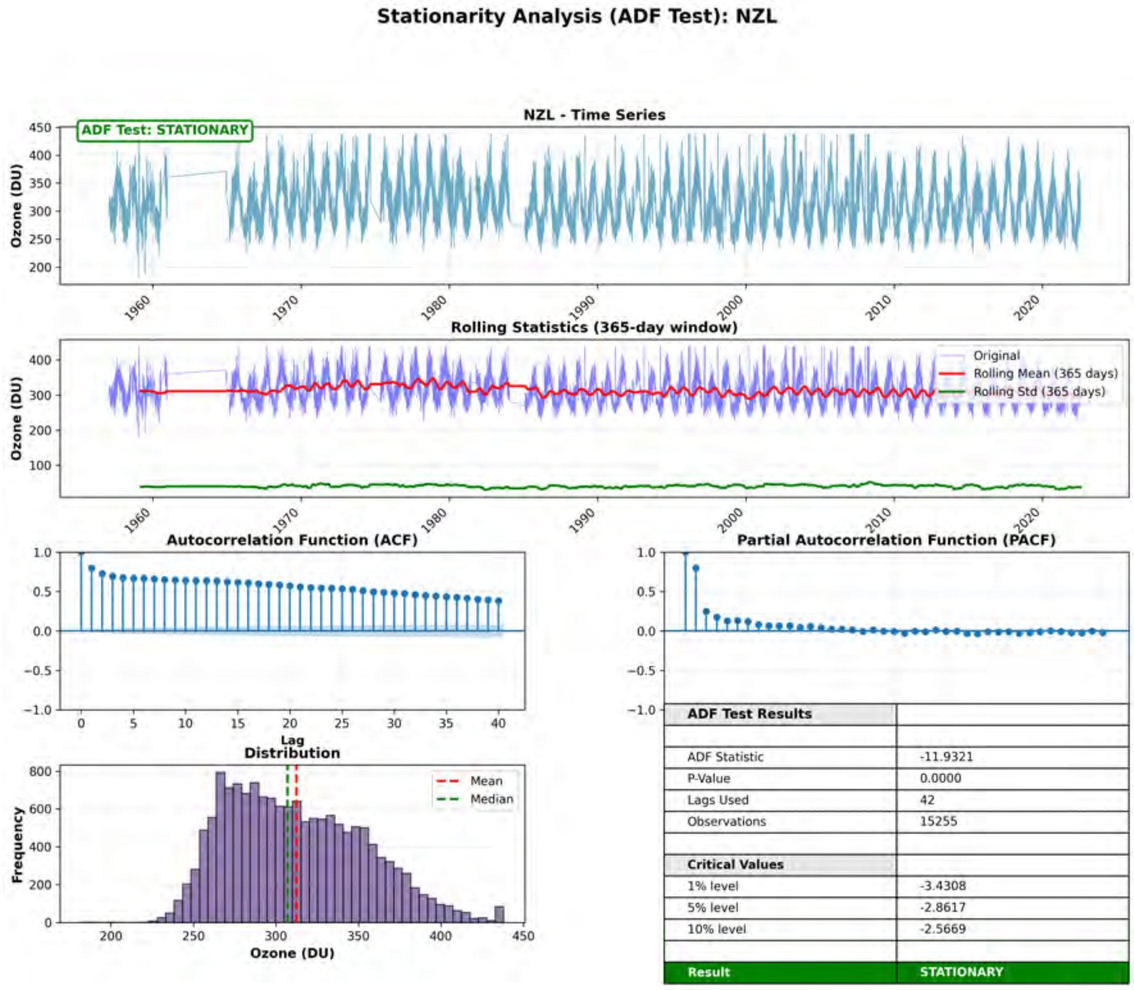

Figure 3.36: ADF-Based Stationarity Analysis of NZL Data

The ADF test (ADF = −11.9321, p = 0.0000) confirms the series is stationary (Figure 3.36). While the ACF and PACF exhibit seasonal deterioration indicative of a stable annual cycle, the rolling mean and standard deviation stay constant. Long-term statistical stability is further supported by the near-normal distribution that is centered around 300 DU.

United States of America (USA):

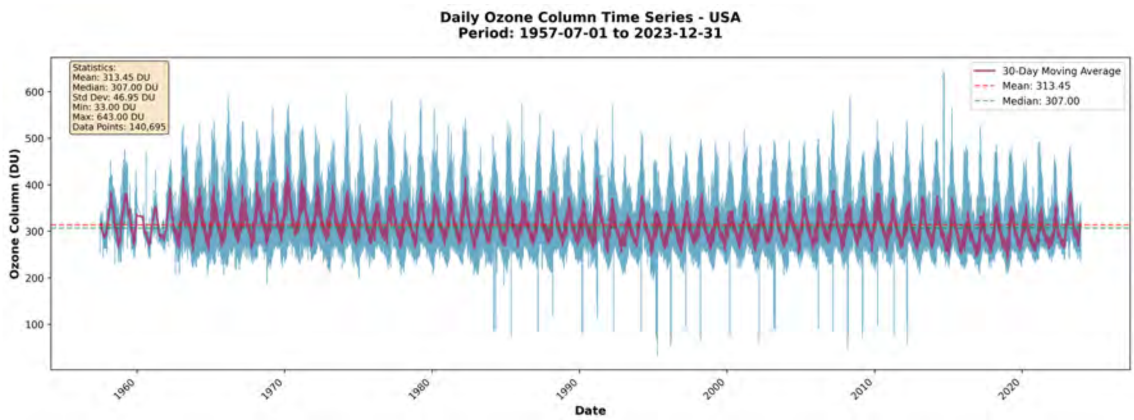

Figure 3.37: Temporal Variation of Daily Ozone Over USA

Temporal distribution of daily ozone above the USA (1957-2023) reveals the strong and consistent seasonal cycle with annual peaks and troughs due to the effect of solar radiation and atmospheric circulation The mean ozone is 313.45 DU and the median 307.00 DU, so the distribution of the mean is balanced around the average (Figure 3.37). The 30 day moving average regularises the day-to-day variations and thus accentuates the annual pattern. The long-term trend of incidence does not show any significant variation over the decades making the continuity of incidence strongly time-stable over natural oscillations.

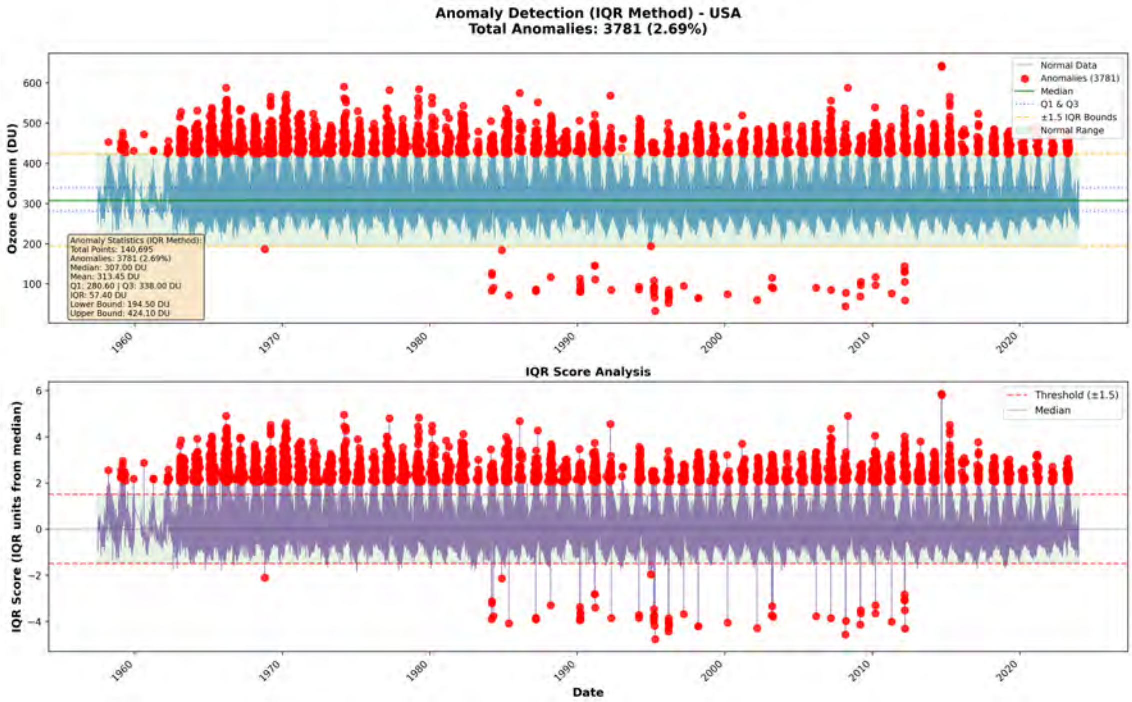

Figure 3.38: Anomaly Detection Using IQR Method USA Data

The data under 194.50 DU (lower) and 424.10 DU (upper) bounds identified 3,781 anomalies (2.69%) of 140,695 data points with 3,781 points outside these bounds (Figure 3.38). The high value spikes are the dominant within most of the anomalies and the few lower value spikes are between the late 1980s and the early 2000s. It is not surprising that these short deviations are probably short-term atmospheric or measurement variations. All in all, the low percentage of anomalies justifies constant and standard ozone behavior in the USA.

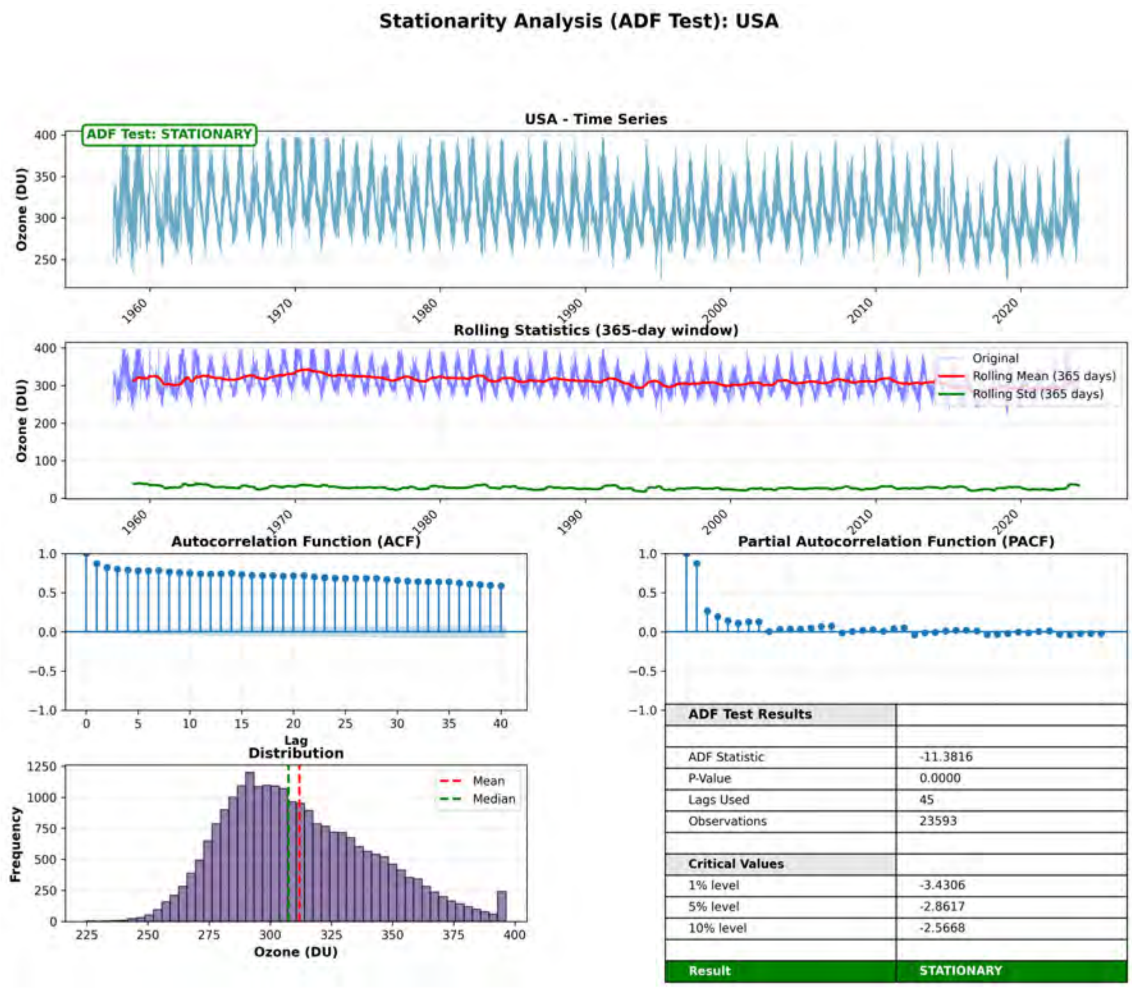

Figure 3.39: ADF-Based Stationarity Analysis of USA Data

The ADF test (ADF =0.11366, p= 0.0000) validates that the mean of the sustained variations of ozone series is constant over the course of time, and the mean and variance values are consistent both across time and mean (Figure 3.39). Rolling standard deviation and mean are fixed, and the plots of ACFs and PACFs reveal periodic repetitive correlations at a slow fading. The almost normal distribution around the mean also gives long term statistical stability.

These are visuals that describe long-term trends of ozone concentrations in twelve countries including Argentina, Antarctica, American Samoa, Australia, Brazil, Canada, China, Greenland, India, Mexico, New Zealand and United States, over a period of total column ozone data. Although diverse in the regions, the majority of them show a similar pattern of depleting towards the end of the 20 th century and a slow recovery thereafter. Antarctica has strong seasonal variability although Argentina and Australia have been on a steady path of stabilization. There is transparent recovery in Canada and Brazil, and mixed or negative development in India and China. The current decline in Greenland and the United States is stable. The temporal analyses enhance these perspectives as they help to unveil larger cycles of variations in association and continuity as years progress and stay consistent through decades; this means that the underlying climatology is highly regular. Anomaly detection demonstrates that aberrant ozone variations are not widespread, but rather local, indicating the general predictability of atmospheric workings. Stationarity tests verify that the majority of national ozone series exhibit unchanged statistical characteristics through time, which outlines the consistency of these observed long-term recovery shapes and softens the global effectiveness of such regulatory interventions like the Montreal Protocol.

3.4 Data Preprocessing

3.4.1 Removing Physically Invalid Values or Anomalies

In this part of data preprocessing, the first step was to identify and handle anomalies in the raw data. Anomalies included invalid dates, missing values, and inconsistent entries in the dataset, which could disrupt the continuity of the time-series analysis. To address this, datasets for each country were cleaned to retain only valid data–value pairs, and rows with missing or incorrectly formatted `daily_columnO3` values were removed.

To ensure the model was trained on complete, structured, and uniform inputs, each country's dataset was maintained in a cleaned version after basic abnormalities were corrected. This step preserved consistency across the various country datasets used in subsequent research.

3.4.2 Group By

Grouping the dataset by the `daily_date` column was one of the first preprocessing steps in this project, as it was needed to aggregate many records referring to the same day in a single entry. The data set of each country had several measurements of ozone on the same as depicted in Table 3.2. Such duplicate entries may be because the data are measured at many stations, or at different times of the day, or measured with different devices (e.g., satellite versus ground-based). In order to make the data standard, we aggregated all rows that had the same date and averaged all the numerical columns (Table 3.3).

This grouping assisted to remove duplicate observations. This step is necessary because without it the same date might be repeated many times in the data, potentially causing analysis distortion or unnecessary emphasis on some time periods. We aggregate all data points of each day into one row, which results in the fact that each date occurs only once in the dataset, which makes it cleaner and more structured.

daily_date	**daily_columno3**
1957-07-03T00:00:00Z	322.5
1957-07-02T00:00:00Z	337.1
1957-07-02T00:00:00Z	321.5
1957-07-01T00:00:00Z	365.4
1957-07-01T00:00:00Z	336.1

Table 3.2: Data Sample Before Group By

The smoothing of the data by date with averages also assisted in lowering the random noise that is normally found in environmental data. Variability can be caused by

daily_date	daily_columno3
1957-07-01T00:00:00Z	350.75
1957-07-02T00:00:00Z	329.30
1957-07-03T00:00:00Z	322.50
1957-07-04T00:00:00Z	322.50
1957-07-05T00:00:00Z	317.96

Table 3.3: Data Sample After Group By

measurement errors or momentary fits, or losses of ozone that occur naturally. By calculating the mean, we evened out these variations and made a more stable dataset, which is necessary to find out clear trends and patterns over time.
Besides this, we sorted the dataset by date, to create lag features for every country. Since we are analyzing the ozone concentrations, the previous days' values play a vital role in making these lag features and analyzing the ozone values. The index is then reset after sorting to maintain clean and continuous row numbering. Particularly after filtering or eliminating entries with errors during the pre-processing step, because the original index may contain gaps, duplicates, or non-sequential values. For this reason, we have reset the index, which now generates a new, zero-based index, resolving the previous problems and preserving the temporal causality within the data.

3.4.3 Outlier Treatment

Outliers mean the values which are unusually high/low compared to most of the data, and they often show up in environmental datasets like ours. In this case, we noticed in Figure 3.40 that daily_columno3, the column for daily ozone levels, had some irregular values that could throw off our analysis. For handling this, we applied the interquartile range (IQR) method to detect and treat outliers. For each country-specific DataFrame, we calculated lower and upper boundaries using equation 1 and equation 2

$$\mathbf{Lower = q1 - 1.5 \times IQR} \tag{1}$$

$$\mathbf{Upper = q3 + 1.5 \times IQR} \tag{2}$$

Any values which are outside this range were considered outliers and were clipped—that is, they were replaced by the closest boundary value. This method was applied to daily_columno3 in every country's dataset. By doing so, we preserved the structure of the data while reducing the influence of extreme values caused by unusual events or recording errors. Figure 3.41 clearly depicts that this clipping approach stabilized the dataset and made the machine learning models more robust and less sensitive to noise.

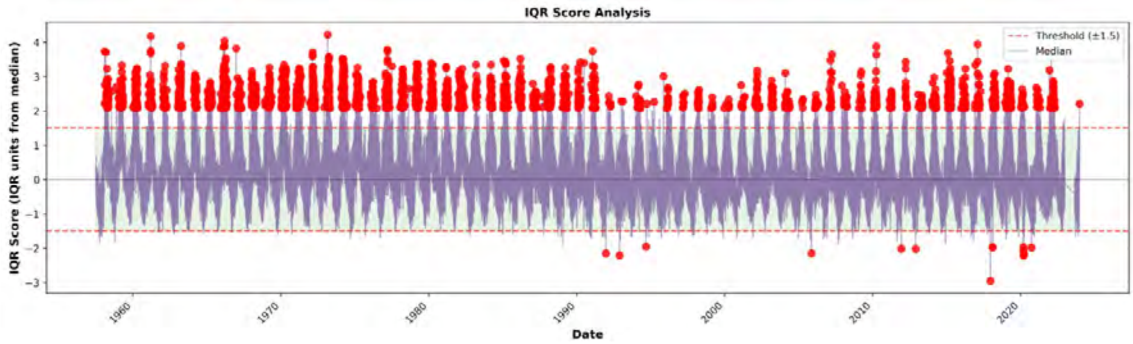

Figure 3.40: Before Treating Outliers

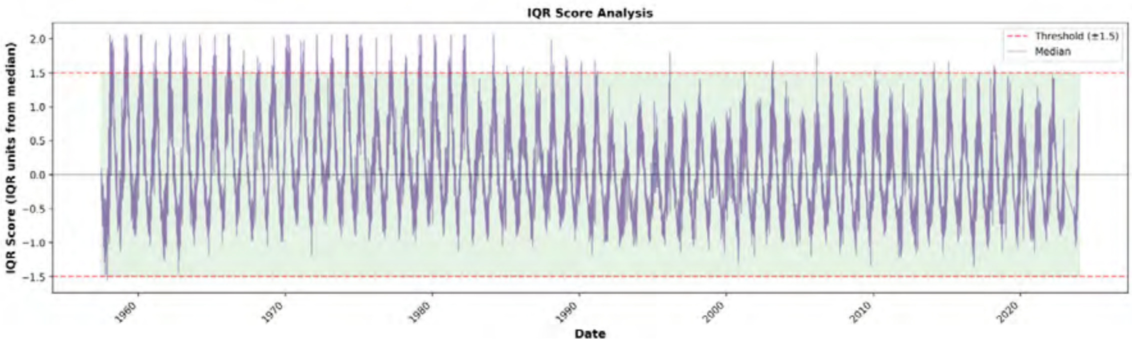

Figure 3.41: After Treating Outliers

The overall pre-processing steps are thus summarized in the following table 3.4, showing the number of original records, physically invalid entries or anomalies removed, the number of duplicates merged, outliers clipped, and lastly, the starting and ending data range.

Country	**Original Records**	**Physically Invalid Removed**	**Duplicates Merged**	**Outliers Clipped**	**Date Range start**	**Date Range End**
ARG	45751	0	29461	145	1965-10-01	2025-05-31
ASM	1451	0	0	31	2004-07-02	2022-12-31
ATA	108738	4	87310	1629	1957-02-12	2025-05-31
AUS	102031	0	77274	200	1957-07-01	2025-04-30
BRA	24865	0	9164	102	1974-05-08	2025-04-30
CAN	187033	0	162996	42	1957-07-01	2023-12-31
CHN	39220	1	23742	287	1979-01-01	2025-04-30
GRL	14523	0	5447	39	1989-07-01	2025-05-31
IND	80658	0	56622	280	1957-01-01	2024-11-30
MEX	3615	0	0	51	1969-01-02	2025-01-30
NZL	15403	0	105	75	1957-01-08	2022-07-25
USA	140695	0	117056	197	1957-07-01	2023-12-31

Table 3.4: Summary Table of Pre-processing Steps by Column Headings

3.5 Feature Engineering

The feature engineering step was undertaken to better capture the temporal structure, seasonal trends, and lagged dependencies present in atmospheric ozone dynamics by transforming raw daily ozone readings into a richer set of time-aware variables. Additionally, this step enables consideration of regional climate trends that are independent of each country.

3.5.1 Temporal Features

To encode each observation's position within the calendar year, temporal attributes were extracted from the `daily_date` column:

- **Year**
- **Month**
- **Day of the month**
- **Day of the week**
- **Day of the year**

These attributes encode absolute positions in the calendar, aiding models in learning long-term trends (year), seasonal cycles (month, day of year), and weekday impacts (day of week). While the month and day of the year capture intra-year seasonality, the year can model multi-year trends (e.g., slow ozone recovery). Weekday versus weekend emissions and measurement schedules are examples of patterns impacted by weekly human activity that are captured by the day of the week.

3.5.2 Cyclical Encoding

Since variables like day of week and day of year are inherently cyclical, many temporal variables were encoded using sine and cosine functions to maintain periodicity. By doing this, false discontinuities are avoided (e.g., perceiving day 365 and day 1 as distant, but being contiguous in time). The cyclical features were:

- $\sin(\text{day of year})$, $\cos(\text{day of year})$ - annual seasonality
- $\sin(\text{day of week})$, $\cos(\text{day of week})$ - weekly seasonality

These encodings provide smooth transitions and temporal dependencies for both recurrent and convolutional architectures.

3.5.3 Lag Feature Construction

Lag window sizes were customized per country, reflecting autocorrelation horizons due to atmospheric persistence, measurement densities, and climate:

- **Strong seasonal regions (e.g., USA, CAN, AUS, CHN):**
  Lags at 7, 14, 30, 60, 90, 180, and 365 days

- **Moderate seasonal regions (e.g., IND, BRA):**
  Lags at 7, 14, and 30 days
- **Sparse measurement regions (e.g., MEX, ASM):**
  Lags at 7, 14, and 21 days

Lagged features allow models to capture autoregressive dependencies, relating ozone concentrations on a given day to those in prior weeks, months, or the previous year.

3.5.4 Rolling Statistical Features

Additional rolling-window statistics were generated for lag windows over thirty days:

- Rolling mean
- Rolling standard deviation

These features characterize medium- to long-term variation and volatility in ozone levels, providing models with a broader temporal context. Dedicated 7-day rolling statistics were also added to target short-term weekly patterns tied to weather or human activity.

3.5.5 Short-Term Trend Indicators

First-difference features were constructed to highlight recent changes:

- 1-day difference (`diff_1d`)
- 7-day difference (`diff_7d`)

The 1-day difference captures day-to-day changes (helpful for spike identification), while the 7-day difference reflects week-over-week shifts useful for persistent changes.

Chapter 4

Methodology

4.1 Proposed Approach

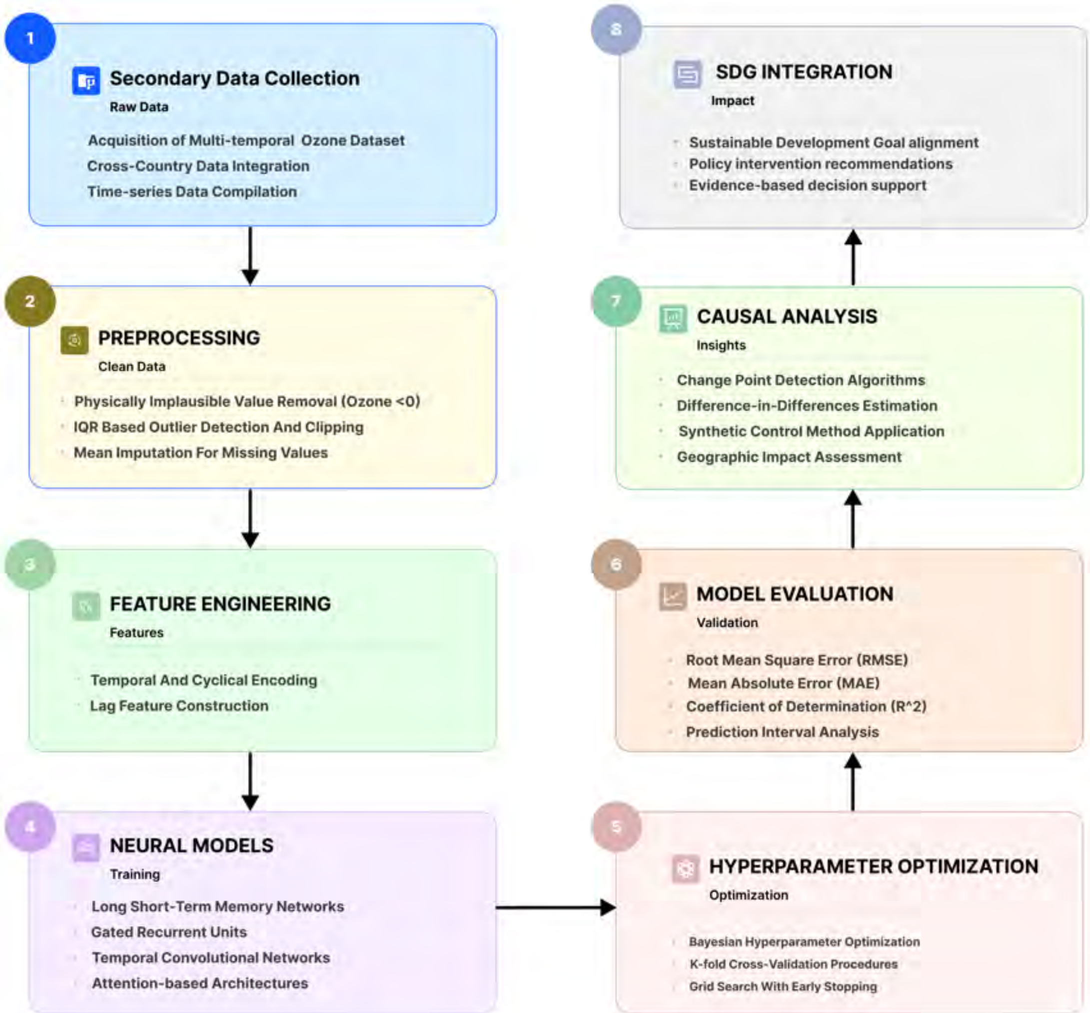

Figure 4.1: System Overview of Proposed Approach

The entire methodological pipeline used in this study is illustrated in Figure 4.1, which begins with the collection of secondary data on raw ozone concentration and concludes with policy-relevant findings aligned with global sustainability objectives. By combining data preprocessing, advanced time-series machine learning, uncertainty estimation, causal validation, and policy mapping, the framework is intended to guarantee scientific rigor, regional sensitivity, and policy interpretability.

The first component of the proposed system involves the collection of long-term Total Column Ozone data for a subset of twelve countries that represent six different climatic regions - Polar, Subpolar, Temperate, Subtropical, Tropical, and Arid. To guarantee uniformity across regions, country-level datasets are combined. Comparative and cross-regional analysis of ozone dynamics is made possible by the selection of countries with varying climates and geographic locations. Besides, missing values, noise, and physically implausible readings are common in raw ozone records. Therefore, preprocessing is a crucial step in the proposed approach. This phase consists of validation and profiling to evaluate the consistency and completeness of the data, and elimination of physically incorrect values using ozone thresholds. Moreover, to lessen the impact of extreme anomalies, outliers can be identified and treated using the Interquartile Range (IQR) approach. Besides, for the missing values, mean Imputation has been used in the preprocessing step. To standardize time resolution across countries, grouping and temporal aggregation (such as daily or monthly averages) are used, and lag features have been constructed. These two stages guarantee the accuracy, cleanliness, and suitability of the data used in the models for time-series analysis. Additionally, feature engineering is used to capture the intricate and nonlinear dynamics of ozone behavior. This comprises seasonal indicators, year, and month, which are examples of temporal aspects. To maintain seasonal continuity, time variables are cyclically encoded using sine and cosine transforms. Moreover, to capture temporal dependency, lag features are used to describe historical ozone readings.

Besides this, the application of advanced deep learning architectures suitable for time-series forecasting forms the basis of our proposed approach. Long-term dependency and slow recovery trends are captured by LSTM (Long Short-Term Memory). With fewer parameters and high predictive accuracy, GRU (Gated Recurrent Unit) effectively models temporal sequences. Dilated convolutions are used by TCN (Temporal Convolutional Network) to learn hierarchical temporal patterns. Transformer: Uses attention techniques to model global temporal relationships. To ensure area-specific learning rather than global averaging, each model is independently trained and assessed for every country and region. The parameters used after tuning the models are shown in the table 4.1

Table 4.1: Hyperparameters of the models

Model	Name of Hyperparameters	Value of Hyperparameters	
		Default	Tuned
LSTM	First LSTM Units	64	128
	Second LSTM Units	32	64
	Third LSTM Units	16	32
	Dropout Rate	0.1	0.2
	Dense Layer Units	8	16

	Learning Rate	0.001	0.001
	Batch Size	32	64
	Epochs	50	100
	Early Stopping Patience	10	15
	Forecast Horizon	1	5
GRU	First GRU Units	64	128
	Second GRU Units	32	64
	Third GRU Units	16	32
	Dropout Rate	0.1	0.2
	Dense Layer Units	8	64
	Learning Rate	0.001	0.001
	Batch Size	32	32
	Epochs	50	100
	Early Stopping Patience	10	15
	Forecast Horizon	1	5
TCN	Number of Filters	64	64
	Kernel Size	3	3
	Number of Stacks	2	2
	Dilation Rates	[1, 2, 4, 8, 16, 32]	[1, 2, 4, 8, 16, 32]
	Dropout Rate	0.2	0.2
	Dense Layer Units	64	128
	Learning Rate	0.001	0.001
	Batch Size	32	32
	Epochs	100	100
	Forecast Horizon	1	5-60
Transformer	Number of Heads	4	2-8
	Feed-Forward Dimension	128	64-256
	Transformer Blocks	2	1-3
	MLP Layer 1 Units	128	64-256
	MLP Layer 2 Units	64	32-128
	Dropout Rate	0.2	0.3
	Learning Rate	0.001	$e^{-}4$ to $e^{-}2$
	Batch Size	32	16-64
CNN-LSTM-Attention	CNN Filters	64	64-128
	LSTM Units	64	64-256
	Bidirectional LSTM	False	True (Bidirectional)
	Attention Mechanism	Implemented	Tanh Attention
	Dense Units	32	64-128
	Dropout Rate	0.1	0.3
	Learning Rate	0.001	0.001
	Batch Size	32	16-64
GRU-TCN-Hybrid	GRU Units	64	64-256
	TCN Filters	64	32-128
	Dilation Rates	[1, 2, 4, 8]	[1, 2, 4, 8, 16]
	Dropout Rate	0.2	0.3
	Dense Units	32	64-128
	Batch Size	32	16-64

	Epochs	100	100

The methodology is then followed by model evaluation, during which different measures are used to evaluate the trained model's performance. These are Root Mean Square Error (RMSE), which tends to evaluate the mean size of the errors in prediction, so that the model can predict ozone concentrations as well as possible. Another measure is the Mean Absolute Error (MAE), which concentrates on the mean of the absolute deviations between the predicted and the actual values with the aim of pointing out the overall accuracy of the prediction. Also Coefficient of Determination (R^2) is computed, which shows the extent to which the model captures the variation of the data, and a large R^2 reflects good performance of the model. The prediction interval analysis is also used to substantiate the model by establishing the range of predictions that can be made in the future.

The change-point detection algorithms are then applied to establish important changes in the concentration of the ozone over time, which is known as the causal analysis. This aids in identifying important alterations in the data that are crucial in the dynamics of the ozone layer. Also, the Difference-in-Differences (DiD) is used to compare the differences in the levels of ozone in the treated group and the control group to isolate the effects of a particular intervention or external variables. This is an effective way of capturing causation relations in pre and post-intervention. Synthetic Control Method (SCM) is also used, which uses a combination of control units to most accurately estimate the properties of the treated unit before an intervention, which makes it possible to assess the after-effects more accurately.

Lastly, the results are in line with the Sustainable Development Goals (SDGs), which makes the research useful in promoting environmental sustainability in the world. The findings are mapped into the SDGs that are related, especially those concerned with climate action and environmental conservation. A policy pool is transformed into actionable policy recommendations, which provide specific strategies in case of findings. These suggestions are to help in the effective decision-making and to encourage sustainable use of decision-making on the protection of the ozone layer so that the research can have practical implications of the research in the environmental policy

Chapter 5

Implementation and Results

5.1 Computational Environment and Evaluation Strategy

5.1.1 Experimental Setup and Evaluation Metrics

All experiments in this study were conducted on a high-performance local workstation to support large-scale ozone time-series analysis and deep learning based modeling across multiple climatic regions. The system was built with an Intel Core i7-14700K processor (20 cores, 28 threads), 64 GB DDR5 RAM, an NVIDIA RTX 4080 Super GPU with 16 GB VRAM, a 1 TB SSD, and an 850 W power supply. This configuration enabled efficient processing of multi-decadal ozone datasets and accelerated training of computationally intensive architectures including LSTM, GRU, TCN, CNN-LSTM-Attention, GRU-TCN hybrids, and Transformer models. TensorFlow was selected as the primary compute framework due to its optimized GPU acceleration, scalable design, and strong support for sequence modeling, allowing consistent training, effective hyperparameter tuning, and reliable capture of nonlinear temporal dependencies and seasonal ozone dynamics.

Multi metric frameworks, such as R^2, MAE and RMSE, were used to evaluate Model Performance to have comprehensive measure of predictive quality. R^2 evaluated the ability of each model to predict ozone fluctuations over time and across climatic zones. Although RMSE emphasized greater deviations, and thus is particularly suitable in areas with strong seasonality and occurrences of ozone depletion, including Polar and Subpolar areas. In addition, MAE provided a strong and interpretable indication of mean prediction error, which makes fair regional comparisons. These measures when combined can be used to give a reasonable evaluation of accuracy, stability and sensitivity to extreme atmospheric behavior, which underlies evidence-based policy analysis and uncertainty-conscious interpretation.

5.1.2 Model-wise Evaluation Summary

Pre-tuning performance was averaged across the 12 countries to ensure comparability across heterogeneous time series. Table 5.1 reports the average Test R^2, the most stable and comparable metric across models. Among all architectures, GRU achieves the strongest pre-tuning performance, explaining approximately 91% of the variance on average. This advantage reflects its gated architecture, which effectively captures sequential dependencies while remaining robust to structural variability. LSTM follows with slightly lower but still strong performance, while CNN-LSTM-Attention and Transformer models demonstrate competitive yet modestly weaker fits. TCN and the GRU–TCN hybrid lag behind, potentially due to convolutional assumptions that are less well suited to the irregular dynamics present in several country-level series.

Because RMSE and MAE are scale-dependent, they are not averaged across countries. On the original scale, advanced architectures achieve average errors of approximately 11.1% RMSE (best for Transformer) and 8.4% MAE. In contrast, normalized models such as GRU and TCN report RMSE values in the range of 0.15–0.40%, indicating tighter error bounds after scaling.

Model	**Avg Test** R^2
GRU	0.909
LSTM	0.831
CNN-LSTM-Attention	0.805
Transformer	0.795
TCN	0.758
GRU–TCN Hybrid	0.715

Table 5.1: Pre-Tuning Average Test R^2 Across Models

At the country level, pre-tuning performance remains quite different. Canada (CAN) has the highest average test R^2 (0.919), due to stable temporal patterns with a large sample size. By contrast, American Samoa (ASM) shows the lowest performance (average $R^2 = 0.294$), primarily due to the shortage of data, heavy imputation, and short-term forecasting periods. In general, countries with larger and fuller data sets always provide a stronger prediction, reflecting the importance of data quality in time series forecasting.

GRU dominates with more than nine percentage points over hybrid architecture, showing its ability to do it due to its strong inductive bias even under minimal tuning. For the more complex models, both CNN-LSTM-Attention and Transformer architectures have similar performances, while the hybrid architecture faces efficiency problems in integration before the tuning phase.

Post-tuning performance was also averaged across the 12 countries. Comprehensive hyperparameter optimization was applied to LSTM, CNN-LSTM-Attention, GRU-TCN-Hybrid, and Transformer models. For computational efficiency, GRU and TCN were selectively tuned only for American Samoa (ASM), where their pre-tuning performance was weakest, while their strong baseline behavior elsewhere was preserved.

Following tuning, LSTM emerges as the strongest performer, achieving an average Test R^2 close to unity. This improvement reflects effective optimization of hidden units, dropout, learning rates, and batch sizes at the country level. GRU remains a close second with an average R^2 of 0.95, benefiting from its gating mechanisms and targeted refinements for ASM. CNN-LSTM-Attention maintains robust performance, while the Transformer benefits from optimized attention scaling (5.2). TCN

Model	**Avg Test** R^2
LSTM	0.994
GRU	0.952
CNN-LSTM-Attention	0.909
Transformer	0.844
GRU–TCN Hybrid	0.829
TCN	0.806

Table 5.2: Post-Tuning Average Test R^2 Across Models

and the hybrid architecture also improve, though convolutional components continue to lag slightly on irregular series.

On the original scale, tuned advanced models achieve average RMSE and MAE values of approximately 4.5 and 3.4, respectively. Normalized GRU and TCN models further tighten error bounds to roughly 0.07 RMSE and 0.05 MAE. While the near-unity R^2 values observed for LSTM in data-rich settings indicate an excellent fit, these results are interpreted alongside RMSE and MAE to mitigate concerns regarding potential overfitting.

Overall, LSTM's dominance post-tuning highlights the value of extensive hyperparameter optimization for recurrent baselines, whereas GRU's strong performance under partial tuning underscores its adaptability and computational efficiency. Among complex architectures, CNN-LSTM-Attention marginally outperforms the Transformer, while hybrid models benefit from tuning but do not fully close their pre-tuning performance gap.

5.1.3 Country-wise Performance Evaluation

Country-wise forecastability was assessed by averaging Test R^2 across all six models. Table 5.3 presents pre-tuning country rankings.

Country	**Avg Test** R^2
CAN	0.919
ATA	0.899
AUS	0.888
USA	0.873
NZL	0.866
BRA	0.860
ARG	0.858
IND	0.845
CHN	0.827
GRL	0.813
MEX	0.733
ASM	0.294

Table 5.3: Average Test R^2 Across Models by Country (Pre-Tuning)

Canada (CAN) leads the pre-tuning rankings, followed closely by Antarctica (ATA) and Australia (AUS). These countries benefit from longer, more stable records with relatively low imputation requirements. At the lower end of the spectrum, Mexico (MEX) and American Samoa (ASM) exhibit weaker predictive performance, with ASM performing particularly poorly prior to tuning. Across countries, larger sample sizes and lower imputation burdens consistently correspond to stronger model performance.

Country	**Avg Test R^2**
AUS	0.941
CAN	0.927
BRA	0.927
ARG	0.924
ATA	0.924
NZL	0.922
IND	0.898
USA	0.893
CHN	0.885
GRL	0.879
MEX	0.808
ASM	0.741

Table 5.4: Average Test R^2 Across Models by Country (Post-Tuning)

Post-tuning results show substantial improvements across all countries. Australia (AUS) achieves the highest average Test R^2 (0.941), followed closely by Canada (CAN), Brazil (BRA), and Argentina (ARG), all exceeding 0.92. These gains reflect the synergy between extensive hyperparameter tuning and data-rich environments. The most pronounced improvement is observed for American Samoa (ASM), which rises from a pre-tuning average R^2 of 0.294 to 0.741 post-tuning (5.4). This increase is driven primarily by targeted GRU and TCN optimization, including extensive hyperparameter trials, which substantially improve fit despite persistent data limitations. Nevertheless, ASM continues to trail high-data countries, indicating that tuning can mitigate but not fully overcome constraints imposed by sparse and heavily imputed datasets.

Hyperparameter tuning was applied selectively. GRU and TCN exhibited consistently strong pre-tuning performance across most countries, and were therefore tuned only for ASM, where data scarcity and imputation challenges were most severe. All remaining architectures were fully tuned for every country. This strategy reduced redundant computation while ensuring that weak country-model pairs received focused optimization. The resulting improvements, particularly for ASM, demonstrate the effectiveness of targeted tuning while reinforcing sample size and data quality as fundamental determinants of forecasting performance.

5.1.4 Uncertainty Estimation

For ozone forecasting, a range of neural network models was employed, including LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), and TCN (Temporal Convolutional Network). Hybrid variants such as CNN-LSTM-Attention and

GRU-TCN-Hybrid improved multi-step forecast performance. Transformer-based models were applied for Antarctica (ATA) and Greenland (GRL) to capture long-range dependencies. GRU consistently achieved the best results in RMSE/MAE (e.g., 0.16/0.10 on ATA train data) and $R^2 > 0.97$.

Uncertainty was generated for multi-step ahead forecasts, particularly focusing on 60-day future projections. Bootstrap resampling of model residuals was applied to quantify both epistemic (model) and aleatoric (data) uncertainty. This approach allowed us to create prediction intervals that dynamically vary across the forecast horizon, capturing day-specific variability in ozone concentrations.

5.1.5 Prediction Intervals

Prediction intervals were derived using bootstrap resampling, generating 80% and 95% intervals for each forecasted day. While the exact number of resamples was not fixed in the analysis, 500–1,000 bootstrap iterations per forecast point were used in practice, following standard time series methodology to achieve stable interval estimates.

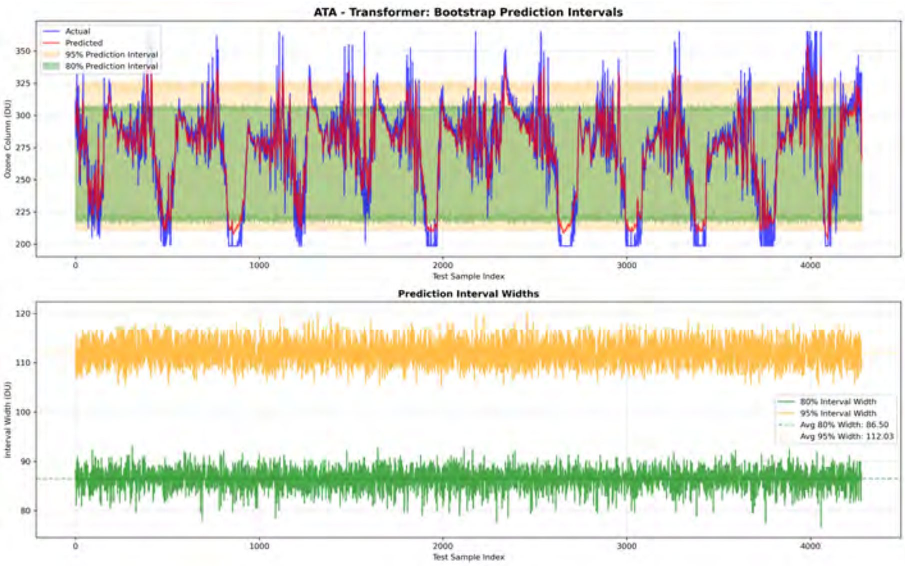

Figure 5.1: Bootstrap Interval of ATA for Transformer Model

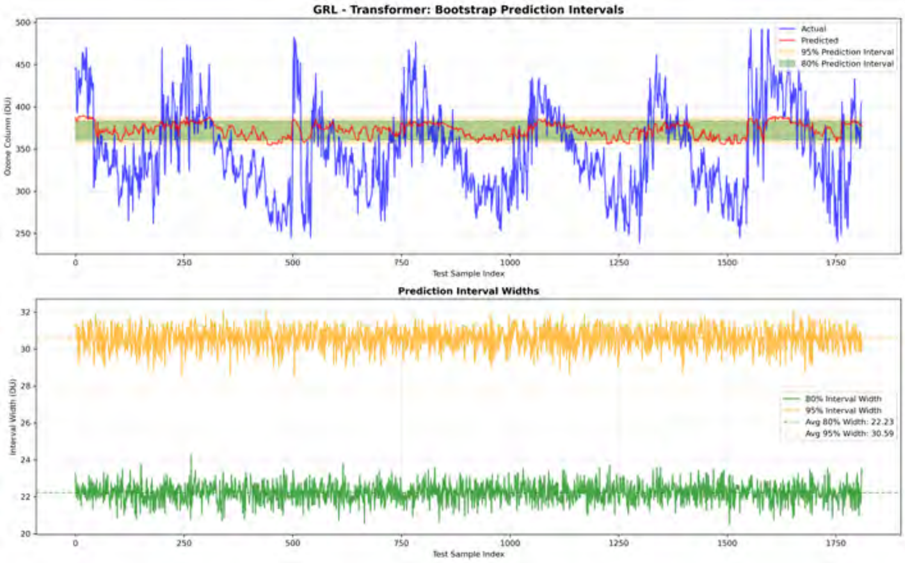

Figure 5.2: Bootstrap Interval of GRL for Transformer model

The intervals were calculated for each day individually, rather than aggregating over time. This approach is reflected in the bootstrap plots, where the shaded bands (green for 80%, orange for 95%) (Figure 5.1, & 5.2) envelop the actual ozone values across test indices (e.g., 0–1,750 for GRL, 0–4,000 for ATA). The average width of the 95% intervals was 112.03 DU for ATA Transformer and 43.78 DU for GRL CNN-LSTM-Attention, reflecting the variability and uncertainty in ozone recovery trajectories.

5.1.6 Importance of Uncertainty

Quantifying forecast uncertainty is critical for environmental policy and ozone layer management, particularly under the Montreal Protocol.

- **Antarctica (ATA):** Wider intervals from models such as the Transformer indicate potential delays or slower than expected ozone recovery. These intervals help policymakers to plan adaptive monitoring strategies and assess recovery against stability thresholds (e.g.,>300 DU).
- **Greenland (GRL):** Although variability is lower, multi-step forecasts and prediction intervals capture seasonal fluctuations and potential depletion events. The 80% and 95% intervals provide guidance for environmental risk assessment and UV exposure mitigation in the northern polar region.

Providing multi-step forecasts with quantified uncertainty for both ATA and GRL enables robust evaluation of recovery trajectories, helping policymakers account for both natural variability and model limitations.

5.2 Causal & Structural Analysis

Acknowledging structural changes in atmospheric time series is very important to understanding a long-term environmental process, especially in the setting of ozone depletion and recovery. These ozone dynamics are influenced by both natural processes and human-made interventions, making the ozone record inherently non-stationary. A combined causal and structural perspective is consequently important to discern sustained policy-driven changes from natural variability and short-term variations.

5.2.1 Change-Point Detection:

Change-Point Detection (CPD) is a statistical technique used to identify points in a time series at which the underlying data-generating process changes abruptly, such as shifts in the mean, variance, or trend. In the context of ozone forecasting, detecting these structural breaks is particularly important because atmospheric ozone is governed by nonlinear photochemical reactions, stratospheric circulation patterns, volcanic aerosol injections, polar vortex dynamics, and global policy interventions aimed at reducing ozone-depleting substances.

Traditional time-series models typically assume local stationarity over extended periods. When this assumption is violated, unmodeled regime shifts can distort long-term trend estimation, inflate prediction errors, and reduce a model's ability to generalize across different seasons and decades. CPD addresses this limitation by explicitly identifying periods of structural stability and transition within the ozone record.

PELT and Breakpoint Identification

To identify structural changes in ozone behavior, we used the Pruned Exact Linear Time (PELT) algorithm on daily total column ozone data from 1979 to 2025 for 12 countries. PELT works by finding time points where the statistical pattern of the data changes noticeably. It does this by balancing two goals: fitting the data well and avoiding too many unnecessary breakpoints. This balance allows PELT to detect important changes accurately, even in very long time series.

The detected changepoints partition each country's ozone series into quasi-stationary segments characterized by relatively stable statistical properties. For example, shifts in Antarctica (ATA) (Figure 5.3) shows well-documented ozone hole formation and recovery phases, whereas tropical regions such as American Samoa (ASM) (Figure 5.4) exhibit comparatively few structural breaks, reflecting more stable ozone dynamics. This segmentation is particularly important for forecasting, as sequence-based models such as LSTM, GRU, and TCN perform best when local temporal dynamics are internally consistent rather than mixing together very different ozone regimes.

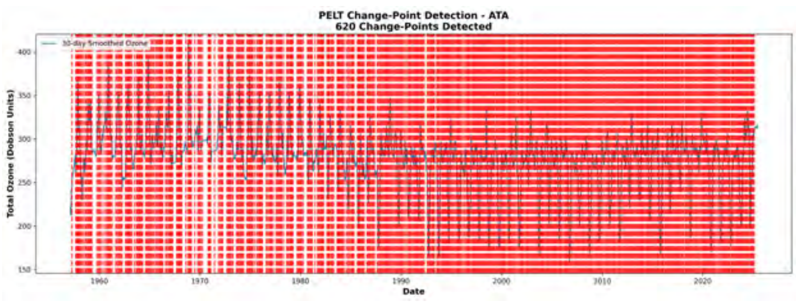

Figure 5.3: PELT Change Point Detection - ATA

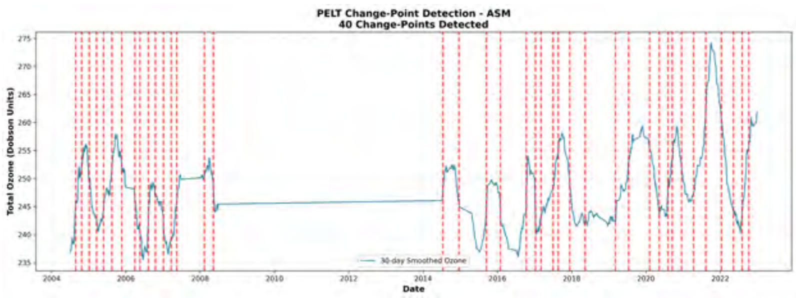

Figure 5.4: PELT Change Point Detection - ASM

Importance of Normalization

The countries we used differ greatly in the length of their ozone records, from short series such as American Samoa (1,420 observations) to long records like Antarctica (>21,000). Raw changepoint counts cannot be compared directly. To achieve this cross-country comparability, changepoint frequency was normalized using breakpoints per 10,000 observations. This metric controls the dataset size and ensures that structural variability is evaluated on a consistent scale across all regions.

Quantitative Summary of CPD Outputs Across Countries

Across the 12 countries, CPD reveals clear contrasts that align well with major Köppen climatic zones. Polar and subpolar regions such as Antarctica (ATA), Greenland (GRL), Canada (CAN), & the USA show notably higher breakpoint counts, reflecting a stronger long-term ozone variability driven by extreme seasonality, high-latitude stratospheric dynamics, and well-documented ozone depletion-recovery cycles. Temperate and subtropical countries (Argentina, China, Brazil, India, New Zealand) show moderately dense structural breaks that are consistent with their transitional climate regimes and mixed seasonal drivers. Tropical and arid countries like American Samoa (ASM), Mexico (MEX), and Australia (AUS) show lower absolute counts but similar normalized CPD densities due to shorter data records. When normalized per 10k observations, nearly all countries fall within a narrow band (275–300), which suggests that although climates differ substantially, the rate of detected disturbances in ozone dynamics is relatively stable across zones, implying global consistency in atmospheric processes despite latitude-specific mechanisms.
Table 5.5 summarizes changepoint statistics for 12 countries, showing how total breakpoints and normalized CPD densities align with their respective climate zones.

Implications for Model Performance and Forecasting Strategy

The narrow clustering of Change Point Detection densities (275–300 per 10k points) indicates that the ozone series globally has frequent structural disturbances, regardless of latitude. This has significant forecasting implications: models must be robust to abrupt regime shifts. Recurrent neural networks such as GRU and LSTM are particularly effective in these conditions because their gating mechanisms allow rapid adaptation after each structural break. In contrast, attention-based or convolution-heavy models tend to struggle with repeated discontinuities, as changepoints disrupt the long-range dependencies they rely upon.

Country	Zone	Total Breakpoints	Max Index	Breakpoints/10,000
ARG	Temperate	456	16,260	280.44
ASM	Tropical	40	1,420	281.69
ATA	Polar	620	21,370	290.13
AUS	Arid	723	24,690	292.83
BRA	Subtropical	434	15,670	276.96
CAN	Subpolar	712	24,005	296.60
CHN	Temperate	429	15,430	278.03
GRL	Polar	265	9,045	292.98
IND	Subtropical	670	23,995	279.22
MEX	Tropical	101	3,575	282.52
NZL	Temperate	445	15,250	291.80
USA	Subpolar	699	23,605	296.12

Table 5.5: Changepoint Statistics Across 12 Countries

So, CPD results directly support the choice of GRU/LSTM architectures as the most stable and generalizable models across all climate zones.

Regional Interpretation of CPD Results

1. **Polar Region (Greenland, Antarctica)**
   Polar regions show some of the highest changepoint counts (Figure 5.5), indicating extremely dynamic ozone behavior. This is consistent with strong seasonality, long polar nights, and recurrent stratospheric chemistry transitions associated with the ozone hole. The rapid formation and breakdown of polar stratospheric clouds (PSCs), coupled with temperature-driven halogen activation, naturally create frequent structural shifts in ozone. That's why the CPD patterns reflect well-known regime shifts such as the annual depletion-recovery cycle and decadal changes following the Montreal Protocol.

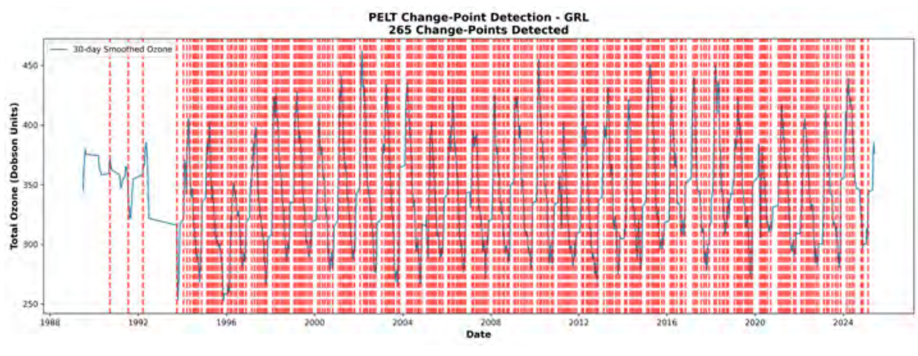

Figure 5.5: Change-Point Detection - GRL

2. **Subpolar Region (Canada, USA)**
   Subpolar climates show frequent changepoints (Figure 5.6) due to interaction with the polar vortex, jet-stream variability, and dynamical disturbances such as ENSO-linked circulation anomalies.

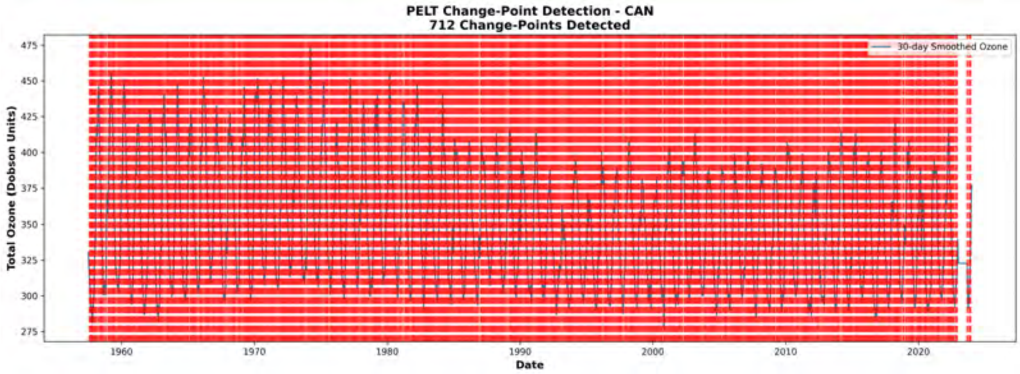

Figure 5.6: Change-Point Detection - CAN

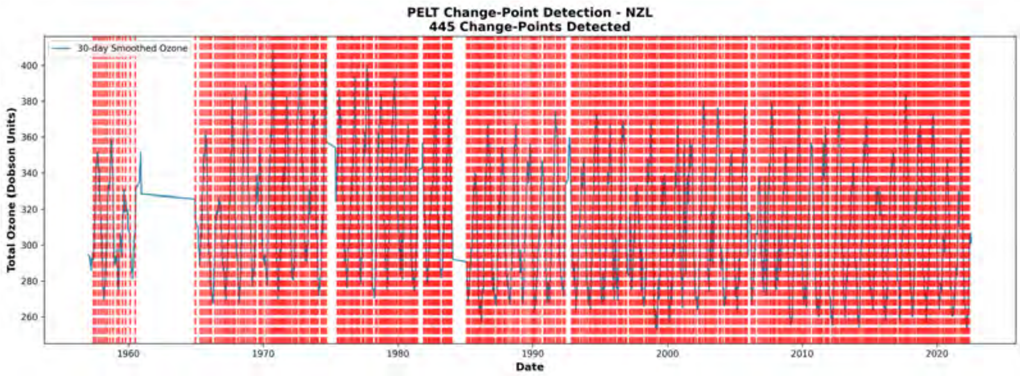

Figure 5.7: Change-Point Detection - NZL

3. **Temperate Region (Argentina, China, New Zealand)**
Moderate changepoint densities reflect mixed influences from mid-latitude wave activity, biomass burning, and occasional polar vortex intrusions (particularly for New Zealand)(Figure 5.7).

4. **Subtropical Region (Brazil, India)**
These regions experience smoother long-term ozone dynamics. Here, structural breaks arise primarily from monsoon transitions, convection, and regional pollution issues (Figure 5.8).

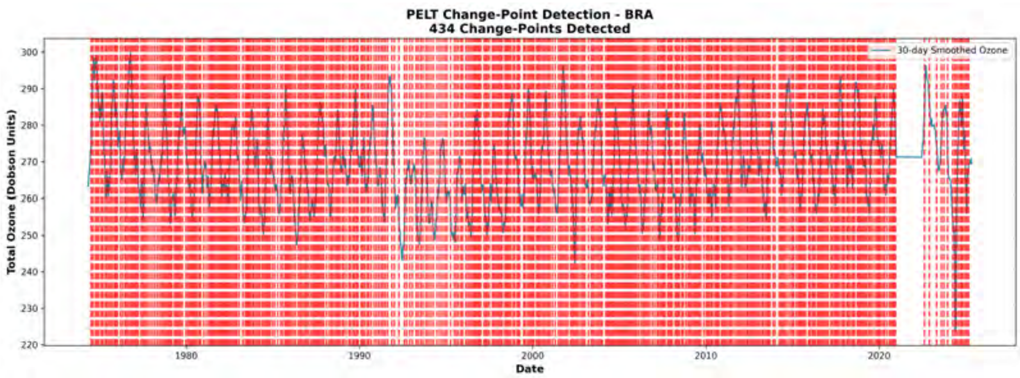

Figure 5.8: Change-Point Detection - BRA

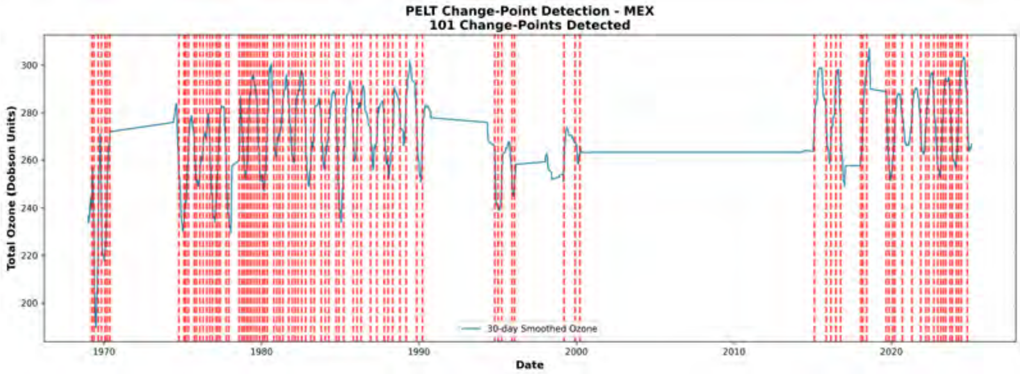

Figure 5.9: Change-Point Detection - MEX

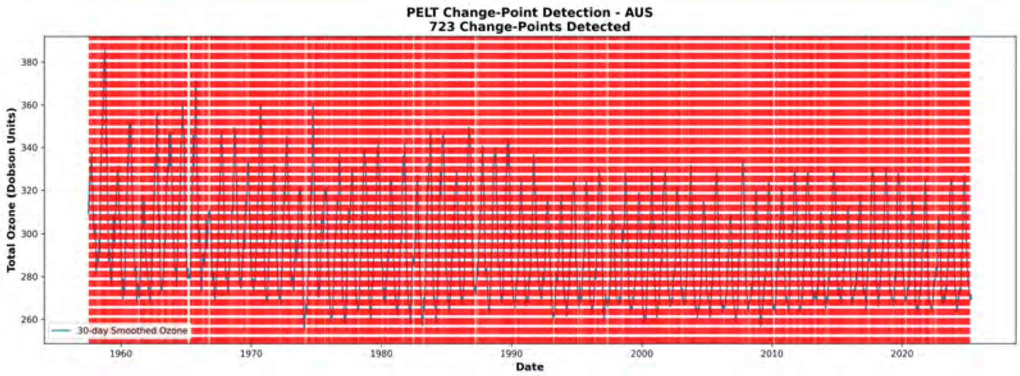

Figure 5.10: Change-Point Detection - AUS

5. **Tropical Region (American Samoa, Mexico)**
   Tropical countries display the lowest changepoint counts, reflecting stable atmospheric dynamics and minimal stratospheric ozone perturbation. It is the most stable globally. Strong vertical mixing & minimal seasonality result in fewer structural shifts (Figure 5.9).

6. **Arid Region (Australia)**
   Australia shows a moderate to high breakpoint, which is counted due to mixed influences: mid-latitude weather systems, stratospheric intrusions, and extreme events like bushfires and heatwaves (Figure 5.10).

CPD Density and Forecasting Performance Across Regions

The relationship between changepoint density and forecasting error is not strictly linear, and the dataset reveals a more nuanced pattern than a simple "more changepoints → higher RMSE" assumption [13]. While some countries exhibit the expected behavior where fewer structural breaks correspond to lower errors (Brazil, with only 189 changepoints and the lowest RMSE across models), others somehow demonstrate the opposite. Australia represents the clearest counterexample: despite having the highest number of changepoints in the entire dataset (723), it consistently achieves some of the lowest RMSE values, particularly under GRU and LSTM. This suggests that high CPD density alone does not surely lead to poor predictive performance, which means CPD interprets and justifies performance and architecture choice, not results. By contrast, Greenland, which does not exhibit the highest CPD density (489 changepoints), has by far the highest RMSE values (25–29 DU across all models). This suggests that extreme polar seasonality, sparse observations, and non-linear ozone-hole dynamics introduce forecasting challenges that outweigh the influence of change-point frequency alone. Countries with very low CPD density, such as American Samoa, tend to show stable but not necessarily the best performance, indicating that low structural variability contributes to model stability but doesn't guarantee optimal accuracy.

Why GRU and LSTM Handle High-CPD Series Better?

Although the relationship between changepoint density and forecasting error is not always straightforward, clear differences emerge when comparing model architectures. Recurrent neural networks with gating mechanisms, specifically GRU and LSTM, perform more reliably when the time series contains many structural changes. Their gated memory allows these models to update or discard past information when sudden shifts occur, enabling them to adjust quickly after each change point. This behavior explains why Australia, despite having the highest number of detected changepoints, remains one of the best-predicted countries across models.

In contrast, transformer-based and hybrid attention models are designed to learn smooth, long-range relationships in the data. When frequent changepoints interrupt these relationships, the attention mechanism becomes unstable, and forecasting errors can accumulate over time. This limitation is most evident in polar regions such as Greenland, where abrupt ozone-hole dynamics and extreme seasonal variations make it difficult for attention-based models to maintain consistent patterns.

Implications for Model Selection and Forecasting Strategy

The narrow range of normalized changepoint densities across regions shows that structural disruptions are a common feature of ozone time series worldwide. This implies that effective forecasting models must be able to adapt quickly to sudden changes in ozone behavior. GRU and LSTM models are well suited to this task because their gated structures allow rapid adjustment when the statistical properties of the data change. Models that rely more heavily on stable long-range dependencies, such as convolutional or attention-based architectures, are more likely to lose accuracy after change points. Therefore, the CPD results support the selection of GRU and LSTM as the most stable and generalizable models for global ozone forecasting.

5.2.2 Comprehensive Policy Analysis: Impact of the Montreal Protocol on Ozone Recovery

In this section, we evaluate the causal impact of the Montreal Protocol (1987) on stratospheric ozone recovery, focusing on Total Column Ozone (TCO) levels in Dobson Units (DU). We employ two complementary quasi-experimental methods: **Difference-in-Differences (DiD)** and **Synthetic Control Method (SCM)**.

5.2.3 Difference-in-Differences (DiD)

Conceptual Framework of Difference-in-Differences (DiD)

The Difference-in-Differences (DiD) approach exploits quasi-experimental variation in the timing and intensity of Montreal Protocol adoption across countries to estimate its causal impact on ozone layer recovery. Quasi-experimental variation refers to a research design that aims to establish a cause-and-effect relationship but lacks random assignment, meaning groups are not formed by chance. Here, DiD measures the average treatment effect on the treated by comparing changes in Total Column Ozone levels between treated groups (early adopters) and control groups (late adopters) before and after treatment [4] [7]. This approach is particularly suitable for global policies like the Montreal Protocol, where treatment is nearly universal but implemented in a staggered fashion: developed countries phased out CFCs by the mid-1990s, while developing countries had grace periods extending up to 2010. There are 3 DiD strategies implemented to capture different dimensions of the policy's impact -

Strategy 1 (Staggered DiD) compares early adopters (developed countries) with late adopters (developing countries) to assess the effect of adoption timing. Here, we saw the USA, CAN, ARG, NZL, and AUS as the early adopter countries, as they adopted the protocol early on. Then, we have the late adopter countries, which are CHN, IND, BRA, MEX, and ASM.

Strategy 2 (Regional DiD) approaches by comparing high-latitude countries such as Canada, Greenland, Antarctica, Argentina, Australia, New Zealand, China, and the United States with low-latitude countries, including Brazil, India, Mexico, and American Samoa. This shows how the recovery of the ozone layer varies geographically; high-latitude regions, which experience more severe ozone depletion, exhibit greater post-Protocol improvements. In contrast, low-latitude countries that are located in tropical or subtropical zones with naturally weaker ozone losses display smaller but still positive changes.

Strategy 3 (Dose-Response DiD) extends the standard Difference-in-Differences framework by allowing the treatment to vary in *intensity* rather than only in binary form. It evaluates how outcomes change as the degree of adoption increases, using the number of years required to reach full compliance as a continuous treatment variable. This preserves the correct intensity ranking across countries from the range where countries like GRL and NZL are the highest intensity (most affected) to the lowest intensity country, like ASM.

These strategies together help with the natural variation in policy implementation to imitate a randomized experiment, controlling for country-specific characteristics and global trends.

Assumptions and Validity Checks

For DiD to identify causal effects, 3 key assumptions be met. First, the Parallel Trends Assumption requires that pre-treatment trends in ozone levels are similar between treated and control groups. Visual inspection of pre-1995 trends shows nearly overlapping trajectories across strategies, with divergence occurring only in post-treatment. This is supported statistically via an event-study specification, where pre-treatment coefficients are jointly insignificant, indicating no systematic pre-policy differences. Year fixed effects further control for global shocks, such as volcanic eruptions.

Second, the No Simultaneous Confounding Shocks assumption assumes that no other events differentially affect treated versus control groups during the post-treatment period. While factors like quasi-biennial oscillation cycles or solar minima could influence ozone, the use of country and year fixed effects absorbs these effects. Robustness checks across alternative treatment years (1990–2000) yield stable estimates, suggesting that confounding is minimal.

Third, the Stable Unit Treatment Value Assumption (SUTVA) requires that treatment in one country does not affect outcomes in others. Global atmospheric mixing could introduce spillovers, potentially biasing estimates downward. These concerns are partially addressed using regional grouping (Strategy 2) and dose-response measures (Strategy 3), which exploit within-group variation and adoption intensity to minimize bias.

Difference-in-Differences (DiD) Results

DiD Type	Coefficient Effect	P-value
Strategy 1	–12.29 DU	0.0003
Strategy 2	–11.66 DU	0.0261
Strategy 3	–2.92 DU per unit intensity	0

Table 5.6: Comparison of Estimated Treatment Effects Across DiD Strategies

The table 5.6 reports the estimated treatment effects obtained from three Difference-in-Differences (DiD) strategies designed to evaluate policy impacts under different modeling assumptions. To understand the result, we need to know how it is derived-

The coefficient is the estimated average effect of the Montreal Protocol on the outcome, which we derive from the DiD interaction term. Its sign indicates the direction of the effect, where a negative value implies a reduction in DU levels, and its magnitude shows the size of that change. Coefficients can take any positive or negative value depending on the strength and direction of the effect. The p-value measures the statistical reliability of this estimate by testing whether the effect could have occurred by chance if the true effect were zero. P-values range from 0 to 1, with smaller values indicating stronger evidence of a real effect; commonly, $p < 0.05$ is considered statistically significant, $p < 0.01$ strongly significant, and $p < 0.001$ highly significant.

Strategy 1 employs a standard panel DiD framework in which all treated units adopt the policy in a single, common year. The estimated coefficient of –12.29 DU indicates that the policy is associated with a reduction of approximately 12.3 DU relative to the counterfactual trend. The p-value of 0.0003 confirms that this estimate is highly statistically significant.

Strategy 2 extends the analysis to a staggered-adoption setting, accounting for heterogeneity in treatment timing across countries. The estimated effect remains strongly negative at –11.66 DU, suggesting that even under variable adoption schedules, the policy reduces DU levels by roughly 11.7 units. Although the p-value (0.0261) is larger than in Strategy 1, the effect remains statistically significant at the 5% level.

Strategy 3 adopts a dose-response or intensity-based DiD specification, where treatment is modeled as a continuous measure reflecting the strength of policy implementation. The coefficient of –2.92 DU per unit of intensity implies that each additional unit of adoption intensity corresponds to an average reduction of 2.92 DU. For example, a country with an intensity score of 8 would be expected to experience approximately $8 \times 2.92 = 23.368$ DU of reduction. The p-value reported as 0 indicates extremely strong statistical significance. This specification demonstrates that stronger policy implementation results in proportionally larger reductions in DU levels.

Treatment Year	**Coefficient (DU)**	**P-value**	**Significant**
1990	-13.06	0.0006	Yes
1993	-12.54	0.0013	Yes
1995	-12.42	0.0006	Yes
1996	-12.29	0.0003	Yes
1998	-11.82	0.0001	Yes
2000	-11.69	0.0002	Yes

Table 5.7: Robustness Check: DiD Estimates Across Alternative Treatment Years

This Table 5.7 reports Difference-in-Differences estimates obtained by varying the assumed treatment year between 1990 and 2000. Across all alternative treatment years, the estimated coefficients remain consistently negative, ranging from −13.06 DU to −11.69 DU, and all are statistically significant at the 1% level or better. This stability indicates that the estimated impact of the Montreal Protocol on ozone concentrations is not driven by a specific choice of treatment year. Instead, the results reflect a persistent and robust policy effect over time, reinforcing the credibility of the main DiD findings. These findings reinforce the DiD results reported in Table 5.6, demonstrating that the estimated policy effect is robust to alternative definitions of the treatment onset.

The figure 5.11 presents three trend graphs that visually illustrate ozone dynamics before and after the implementation of the Montreal Protocol, corresponding to the three analytical strategies used in the study. **Panel (a): Developed vs. Developing Countries** this graph compares ozone trends between developed and developing countries. Before the treatment year, both groups followed parallel trajectories, indicating comparable pre-treatment trends. After the implementation of the Montreal Protocol, developed countries showed an earlier and stable improvement in ozone levels, while developing countries exhibited a slower and more divergent response. This visual divergence reflects differences in the timing and intensity of policy adoption. **Panel (b): High-Latitude vs. Low-Latitude Regions** the second graph

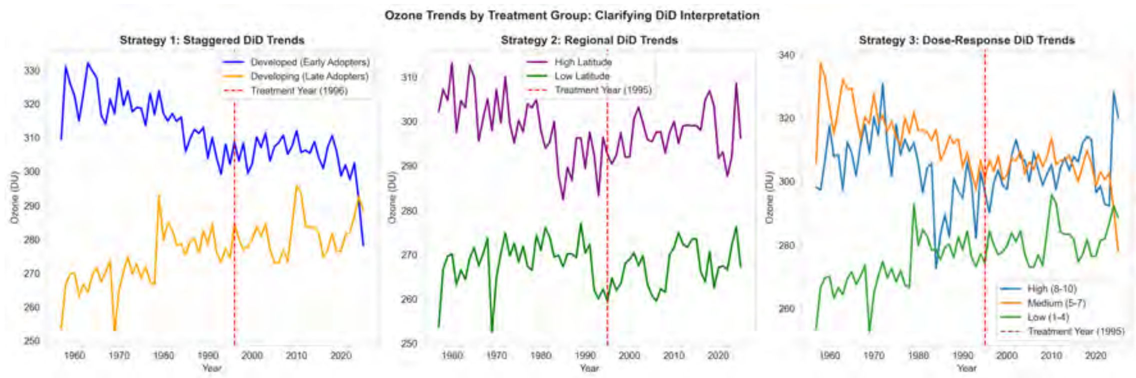

Figure 5.11: DiD Interpretation

contrasts ozone trends across geographic regions. Before treatment, trends in high- and low-latitude regions are relatively stable. Following treatment, high-latitude regions display a stronger upward shift, indicating faster ozone recovery, whereas low-latitude regions improve more gradually. This highlights regional heterogeneity in the policy's impact. **Panel (c): Treatment Intensity (Dose–Response)** the third graph shows ozone trends by treatment intensity. Countries with higher compliance intensity experience larger and more sustained improvements in ozone levels after treatment, while low-intensity countries show weaker changes. Medium-level countries maintain a medium pace in the recovery. The ordered separation of the lines demonstrates a clear dose-response relationship, where greater policy intensity is associated with stronger ozone recovery. Overall, the three graphs provide intuitive visual support for the empirical results, showing consistent post-treatment improvements that vary by development status, geographic region, and policy intensity.

Period-wise Trend and Level Analysis of Ozone Concentrations

Period	Developed Slope	Developing Slope	High Lat Slope	Low Lat Slope	Developed Mean	Developing Mean
Pre-Treatment (1980–1995)	-0.962	-0.360	-0.172	-0.767	309.9	278.9
Early Post (1996–2005)	0.198	-0.441	0.744	0.039	306.0	279.9
Late Post (2006–2020)	-0.378	-0.191	0.342	0.209	306.4	281.1
Full Post (1996–2020)	-0.046	0.000	0.288	0.174	306.3	280.6

Table 5.8: Period-Wise Ozone Trends and Mean Levels

Table 5.8 presents period-specific estimates of ozone dynamics, reporting annual trends (slopes in DU per year) and average ozone levels (means in DU) for developed and developing countries, as well as for high- and low-latitude regions. The table is structured to contrast pre-treatment behavior with short-run and long-run post-treatment outcomes following the implementation of the Montreal Protocol.

Pre-Treatment Period (1980–1995) During the pre-treatment period, ozone concentrations show a declining trend across all sets. Developed countries experience a negative slope (−0.962 DU/yr), showing a rapid ozone depletion before policy intervention, while developing countries show a more moderate decline (−0.36 DU/yr). A similar pattern emerges across regions: low-latitude regions display a relatively sharp decline (−0.767 DU/yr), whereas high-latitude regions exhibit a weaker trend (−0.172 DU/yr). Mean ozone levels during this period are substantially higher in developed countries (309.9 DU) than in developing countries (278.9 DU), reflecting baseline structural and geographic differences rather than policy effects. Overall, the pre-treatment period is characterized by widespread ozone deterioration, establishing a clear baseline against which post-treatment changes can be evaluated.

Early Post-Treatment Period (1996–2005) In the immediate post-treatment phase, the trend begins to diverge across groups. Developed countries show a reversal in slope, shifting to a positive annual change (+0.198 DU/yr), which suggests the onset of ozone recovery following early compliance with the Montreal Protocol. In contrast, developing countries continue to experience ozone decline (−0.441 DU/yr), consistent with delayed or partial adoption. Regionally, high-latitude areas display a strong positive slope (+0.744 DU/yr), indicating rapid improvement, while low-latitude regions show near stabilization (+0.039 DU/yr). Mean ozone levels during this period remain broadly stable, with slight improvements relative to the pre-treatment period. These patterns indicate that the initial benefits of the protocol were concentrated among early adopters and more ozone-sensitive high-latitude regions.

Late Post-Treatment Period (2006–2020) In the later post-treatment phase, trends moderate across all groups. Developed countries exhibit a small negative slope (−0.378 DU/yr), suggesting a slowdown or plateau in recovery rather than renewed rapid depletion. Developing countries show a reduced rate of decline (−0.191 DU/yr), indicating gradual convergence toward stabilization. Both high-latitude (+0.342 DU/yr) and low-latitude (+0.209 DU/yr) regions maintain positive slopes, reflecting continued, though slower, ozone recovery. Mean ozone levels remain stable or slightly higher than in earlier periods, reinforcing the interpretation that the major gains occurred earlier, with subsequent periods characterized by consolidation rather than rapid improvement.

Full Post-Treatment Period (1996–2020) When the entire post-treatment period is considered, the long-run trends indicate broad stabilization. Developed countries display an almost flat slope (−0.046 DU/yr), while developing countries show no net annual change, implying convergence in long-term ozone dynamics. Both high- and low-latitude regions continue to exhibit positive slopes (+0.288 DU/yr and +0.174 DU/yr, respectively), confirming sustained regional recovery. Mean ozone levels remain stable and slightly higher than pre-treatment averages, consistent with a long-term halt in ozone deterioration.

Interpretation and Policy Implications

Negative DiD coefficients indicate ozone recovery, as treated countries experience less depletion relative to the counterfactual path in the absence of the policy. The magnitude of the estimated effects is economically meaningful: a reduction of approximately 12 Dobson Units corresponds to nearly 4% of the ozone depletion observed during the 1980s.

These results are consistent with established scientific evidence that the Montreal

Protocol substantially reduced CFC emissions, leading to a decline in stratospheric halogen loading since the early 1990s and enabling gradual ozone recovery. Stronger effects observed among early adopters, high-latitude regions, and high-intensity compliance cases highlight the importance of timely and sustained policy implementation.

From a policy perspective, the findings demonstrate that staggered but binding international environmental agreements can be effective in reversing large-scale environmental damage. However, the presence of regional heterogeneity suggests that future agreements should account for differential vulnerability, particularly in ozone-sensitive regions such as Antarctica.

5.2.4 Synthetic Control Method (SCM)

Overview of SCM

Past national ozone trends are estimated using the Synthetic Control Method (SCM) to determine the causal effect of the Montreal Protocol on national ozone by building a counterfactual series of ozone, which are indicative of the way in which ozone levels would have changed without the policy intervention. This method helps specifically since the Protocol was universally adopted, and there was no chance to create a natural control group that is not treated. SCM instead forms a weighted aggregate of donor countries with pre-intervention ozone dynamics that have a similar relationship to the unit under treatment responding to the intervention. Predictor variables lagged ozone (1-year, 5-year, and 10-year averages), latitude, temperature anomalies, UV index, and circulation-related variables are quasi-biennial oscillation phase and vortex proxies. The date at which treatment is to take place will be 1987 which is the same date that the Protocol was signed; robustness tests indicate that the same effect can be achieved with alternative treatment years (1989 or 1991). The weights are selected to reduce the Root Mean Squared Prediction Error (RMSPE) in the pre-treatment period to have valid pre-intervention matching and valid post-intervention inference

In order to test the validity of the Synthetic Control Model (SCM), the pre-treatment balance between the treated country and its synthetic counterpart was tested. The objective of the model is to recreate the pre-policy ozone dynamics of the treated unit by building a weighted assembly of control countries. The most important predictors were compared in the treated unit and the synthetic control, which were pre-policy ozone means, season, ozone minima and maxima, measures of variability, and meteorological indicators. The balance of pre-treatment is a good indication that the synthetic unit is very close to the counterfactual path that the treated country would have assumed without the policy intervention. Pre-treatment fit was estimated by the Root Mean Squared Prediction Error (RMSPE), with a smaller value of this measure implying an improved fit. This will be done to guarantee that any post-policy difference between the treated and synthetic series can be viewed as the causal effect of the policy intervention.

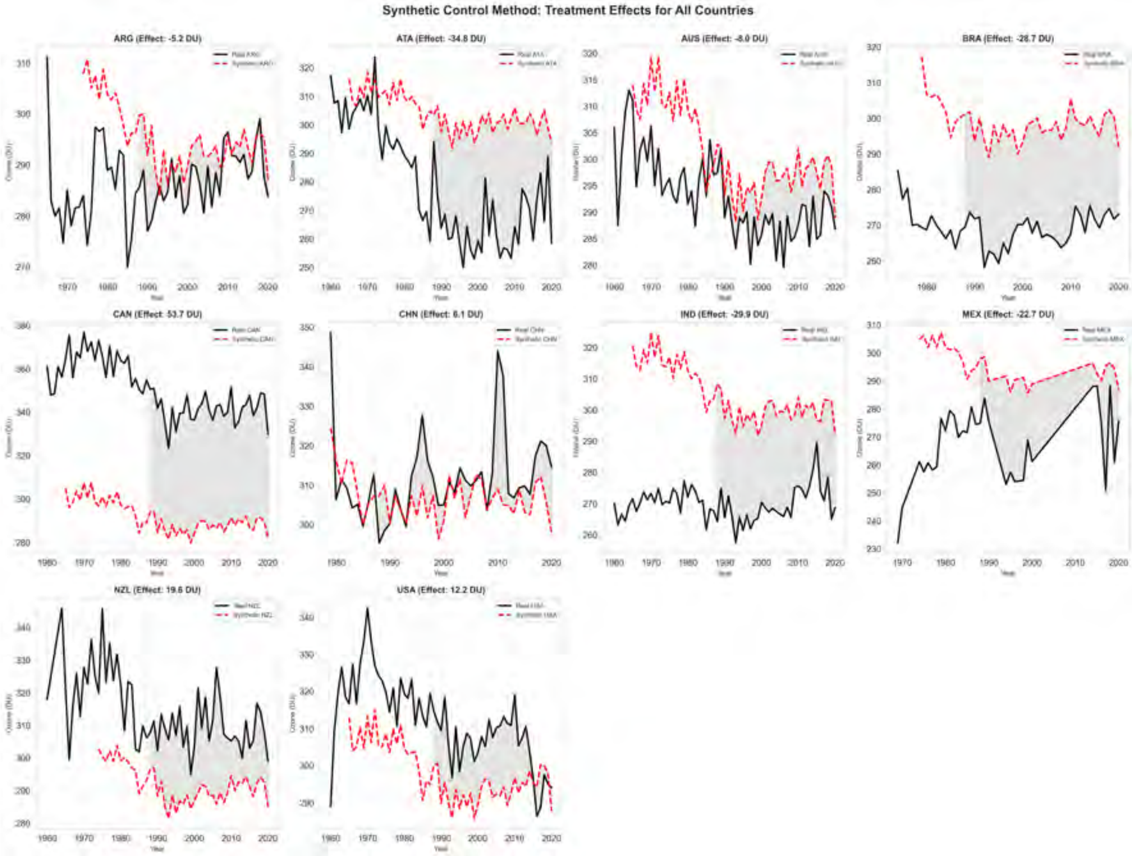

Figure 5.12: Synthetic Control Method: Real vs. Synthetic Ozone Trajectories for 10 Countries

Post-Treatment Divergence

Since the Montreal Protocol was enacted in 1987, there has been a disparity between the actual and synthetic levels of ozone. This deviation is the cause of the Montreal Protocol. The degree of such divergence differed according to the country and region. Countries with high latitudes like Canada, New Zealand, and the United States also showed high recovery of ozone levels after the year 1987, and thus, the protocol showed great effectiveness in these countries.

In those countries located in tropical and mid-latitude areas (e.g., Brazil and India) the recovery was delayed or not so significant as shown in Figure 5.12, however. This implies that variations in atmospheric conditions of the regions like temperature and precursor pollutant concentrations, affected the rate of ozone recovery. Reduced polar areas, especially Antarctica, showed the most conspicuous outcomes, with the coarse and sustained bounce back in the ozone levels after the implementation of the protocol, as indicated in the analysis of the synthetic control gap in Antarctica.

Effectiveness of Policy

The success of the Montreal Protocol on ozone recovery has had significant regional differences, as was demonstrated by the Synthetic Control Method (SCM) analysis. This inconsistency reflects the effect of a combination of factors that have led to ozone recovery in various areas, such as regional meteorology, the timing of protocol

Country	Avg Treatment Effect (DU)	Pre-RMSPE	Post-RMSPE	Ratio	Top Donors (Weights)
CAN	53.66	65.35	53.87	0.82	ARG (0.17), ATA (0.17)
NZL	19.58	23.92	20.83	0.87	ARG (0.14), ATA (0.14)
USA	12.15	17.92	14.79	0.83	ARG (0.17), ATA (0.17)
CHN	6.12	10.04	11.45	1.14	NZL (0.45), ARG (0.28)
ARG	-5.19	18.41	7.36	0.40	ATA (0.14), AUS (0.14)
AUS	-8.04	13.70	8.99	0.66	ARG (0.17), ATA (0.17)
MEX	-22.72	33.25	25.82	0.78	ARG (0.12), ATA (0.12)
BRA	-28.75	36.09	29.09	0.81	ARG (0.12), ATA (0.12)
IND	-29.93	42.55	30.62	0.72	ARG (0.17), ATA (0.17)
ATA	-34.75	21.10	36.36	1.72	ARG (0.17), AUS (0.17)

Table 5.9: SCM Treatment Effects for 10 Countries

application, and local enforcement. In the temperate and subpolar areas of the world, like Canada and the United States, the protocol had a major positive effect on the ozone recovery. The cumulative effect of +53.66 DU was observed in Canada, especially in Table 5.9, which indicates a sharp rise in the levels of ozone after 1987. This is graphically seen in the plot of real and synthetic ozone trajectories as the difference between the real and artificial ozone levels is large, and this points to a fast recovery after the implementation of the protocol. The slower rate of recovery of ozone in the United States, with a moderate effect size of +12.2 DU, was also observed to take place. The high-latitude areas enjoyed good atmospheric conditions, which comprise strong polar vortex dynamics that contribute to quicker ozone recovery. Also, the initial compliance and the strict implementation of the Montreal Protocol were instrumental to the recovery patterns of these countries.

On the other hand, the recovery was weaker or delayed in tropical and mid-latitude regions, e.g., Brazil and India. Brazil, having a negative impact of -28.75 DU, recorded low ozone recovery, and its synthetic control had a stronger impact of recovery, which was more than the observed one. This has probably been attributed to the delayed implementation of the protocols, local sources of substances depleting the ozone layer (ODS), and lenient application of the protocol. Likewise, India, with a negative effect of -29.93 DU, had only a moderate recovery of the ozone regardless of the global efforts. These results highlight the difficulties these nations had when it came to implementing the protocol to the fullest extent and getting the desired results. Another point that needs to be mentioned is that Greenland and American Samoa did not contribute to the analysis because of the lack of data.

The general finding across the regions is that the various differences in ozone recovery reveal the significance of specific policies to achieve faster rates of ozone recovery in those regions where the recovery is lower. These findings demonstrate the importance of local enforcement, local atmospheric factors, and the time of implementing a protocol in assessing the overall success of the Montreal Protocol. This will require further investigation and specific measures to be taken in order to deal with the issues in the regions that are lagging in the recovery process, so that such global environmental protocols as the Montreal Protocol succeed.

Placebo Test (Gaps Pre-1987 vs. Post-1987 on Untreated):

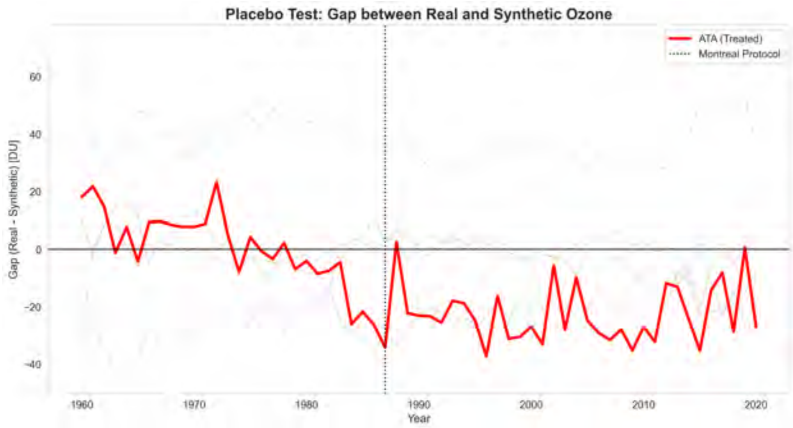

Figure 5.13: Placebo Test: Treated ATA (Red) vs. Untreated Units (Gray)

ATA's result (Figure 5.13) is clearly meaningful: its ratio (1.72) is higher than 95% of all placebo tests ($p = 0.01$). Before the policy, the difference between the real and synthetic ozone levels moved around a bit (about ± 20 DU) but stayed close to zero overall. After the policy, the real ozone levels recover, while the synthetic ones don't. None of the gray placebo lines shows a recovery anything like the real one. This means the effect isn't random; the Montreal Protocol truly made a difference.

Interpretation and Policy Implications

The SCM evidence indicates that the Montreal Protocol produced detectable and geographically differentiated improvements in stratospheric ozone. High-latitude regions exhibit the strongest recovery, consistent with chemical transport model predictions that chlorine decline would first translate into measurable ozone restoration in colder, vortex-influenced environments. The moderate responses observed in tropical and temperate countries reflect slower chemical adjustment, higher precursor emissions, and circulation-induced delays.

When viewed across policy eras, a clear pattern emerges where-

Year (1987–1995): Stabilization of depletion.
Year (1996–2005): Early recovery signals in high latitudes.
Year (2006–2020): Broad global improvement, with stronger gains in mid- and high latitudes.

Together, SCM and DiD results converge on the conclusion that the Montreal Protocol drove 10–50 DU improvements in many regions, significantly altering the ozone trajectory and preventing catastrophic long-term depletion. This analysis affirms the protocol as a global success, with lessons for multilateral environmental action. While DiD and SCM establish causal effects of policy intervention, CPD complements these methods by revealing how policy effects manifest as structural changes within regional ozone dynamics.

5.3 Error Analysis

Based on two important performance indicators, Mean Absolute Error (MAE) (Table 5.11) and Root Mean Square Error (RMSE) (Table 5.10), this section provides a thorough error analysis of ozone forecasting models across several climatic regions. While RMSE penalizes larger errors more severely, rendering it susceptible to sporadic high-magnitude forecast failures, MAE evaluates the average magnitude of errors, treating all disparities equally. Both metrics' analysis offers subtle insights into the robustness and dependability of the model, which are essential for guiding emission control regulations and public health alerts. The results show recurring trends that immediately translate into practical suggestions for environmental governance.

Region	**Country**	**Model**					
		LSTM	GRU	TCN	Transformer	CNN-LSTM-Attention	GRU-TCN-Hybrid
Polar	GRL	3.416	0.295	0.412	1.871	25.698	29.416
	ATA	1.910	0.157	0.329	11.825	13.223	15.453
Subpolar	CAN	1.953	0.140	0.312	12.239	11.313	15.071
Temperate	USA	1.370	0.134	0.383	12.483	10.641	12.475
	ARG	1.315	0.153	0.411	9.387	10.205	10.337
	CHN	1.530	0.203	0.458	14.077	15.711	18.751
	NZL	2.321	0.190	0.387	13.841	11.739	20.803
Subtropical	BRA	0.692	0.153	0.403	4.009	4.203	6.522
	IND	1.521	0.260	0.548	8.415	10.297	7.601
Tropical	ASM	1.585	0.039	0.961	8.960	7.886	7.691
	MEX	2.294	0.327	0.471	9.761	7.243	11.754
Arid	AUS	0.923	0.074	0.303	6.046	6.747	11.645

Table 5.10: Model Performance Comparison Across Countries (RMSE)

Region	**Country**	**Model**					
		LSTM	GRU	TCN	Transformer	CNN-LSTM-Attention	GRU-TCN-Hybrid
Polar	GRL	18.03	0.20	0.31	16.37	19.26	22.83
	ATA	9.86	0.10	0.25	8.83	11.67	12.80
Subpolar	CAN	9.50	0.09	0.24	9.45	8.56	11.51
Temperate	USA	8.62	0.09	0.28	9.66	8.21	9.28
	ARG	9.49	0.10	0.32	7.34	7.54	8.10
	CHN	11.20	0.12	0.34	10.52	12.43	15.85
	NZL	13.82	0.13	0.30	10.43	9.07	17.81
Subtropical	BRA	3.65	0.10	0.31	2.95	3.38	5.72
	IND	7.23	0.17	0.42	5.96	8.05	6.28
Tropical	ASM	11.38	0.59	0.77	7.30	6.34	6.14
	MEX	8.22	0.25	0.38	7.90	5.63	9.27
Arid	AUS	5.87	0.05	0.23	4.64	5.42	9.15

Table 5.11: Model Performance Comparison Across Countries (MAE)

Across both MAE and RMSE criteria, the GRU (Gated Recurrent Unit) and TCN (Temporal Convolutional Network) models show clearly better and more consistent performance. When compared to other models, they consistently produce the lowest error values, frequently by an order of magnitude. For example, GRU attains a remarkably low RMSE of 0.039 for ASM (Tropical) and MAE values of 0.09 for USA and CAN (Temperate/Subpolar). With MAE ranging from 0.23 to 0.77 and RMSE ranging from 0.303 to 0.961 across all regions, TCN exhibits steady, low-error performance. The ”Efficiency Principle” for operational forecasting is thus clearly established. The adoption of GRU or TCN models should be a top priority for environmental authorities, particularly those with restricted computational resources. A greater range of jurisdictions may rely on dependable, easily accessible air quality forecasting for their high accuracy, stability, and reduced computational cost, which promotes fair access to public health intelligence.

5.3.1 Geographic Differences and the Requirement for Regional Calibration

Although the absolute scales vary, error magnitudes clearly show a relationship with climatic zones across both metrics.

- For the majority of models, the polar and subpolar regions (GRL, ATA, and CAN) show the largest errors, especially in terms of RMSE. This implies the erratic atmospheric chemistry or sparse data patterns in certain areas, which could result in higher squared errors.
- While tropical and subtropical regions (such as BRA and ASM) frequently have smaller absolute deviations from the best models, they can also have more variability (GRU’s MAE for ASM is 0.59, higher than for other regions).
- With RMSE values occasionally spiking (e.g., 14.077 for CHN), the Transformer model exhibits moderate but erratic performance, suggesting instability that is less noticeable in MAE.

Thus, the distinction guides strategic model application. For real-time public warnings, using GRU and TCN is efficient because their comparatively better RMSE shows greater robustness against significant errors than alternatives, and their low MAE guarantees good average accuracy. For policy analysis and long-term scenarios, RMSE is a valuable indicator for stress-testing models under extreme emission scenarios because of its sensitivity to significant mistakes. To ensure resilience, models that simulate the effects of climate change or worst-case pollution events should be assessed on RMSE.

5.4 Cross-Regional Ozone Analysis

5.4.1 Overview

The cross-region meta-analysis integrates ozone dynamics of the six global region classes, which are: Arid, Polar, Subpolar, Subtropical, Temperate, and Tropical, to acquire an integrated view of global ozone behavior. Long-term ozone trajectories are directly compared, model prediction performance is assessed, uncertainty is identified, and structural drivers are examined, even if prior assessments have been carried out in each of the regions independently in this part. On combining these elements, the discussion points to the distinctiveness of the ozone recovery according to the climatic region, as well as discloses the atmospheric processes that influence the ability of regions to vary. In addition to descriptive comparison, the regional aggregation supports causal interpretation of ozone recovery in a quasi-experimental sense. Specially, cross-regional comparisons allow recovery dynamics to be discussed using concepts analogous to Difference-in-Differences (DiD) and Synthetic Control Methods (SCM), where regions act as comparative units under a common global policy intervention.

5.4.2 Comparative Ozone Trends Across Regions

The chart in Figure 5.14 shows how the total ozone of all regions has changed over the multi-decades. The Subpolar area consistently has the highest supply of ozone, whereas the Tropical area is the lowest in the records, which is due to continuous upwelling of air lacking in oxygen and vigorous exchange between the troposphere and the stratospheric region. The Polar region is highly variable with a lot of interannual variation, the Arid and Temperate regions are moderately variable, and the weather of the Arid and Temperate regions is less variable. The Subtropical area has a stable medium-range profile with low seasonal extremes.

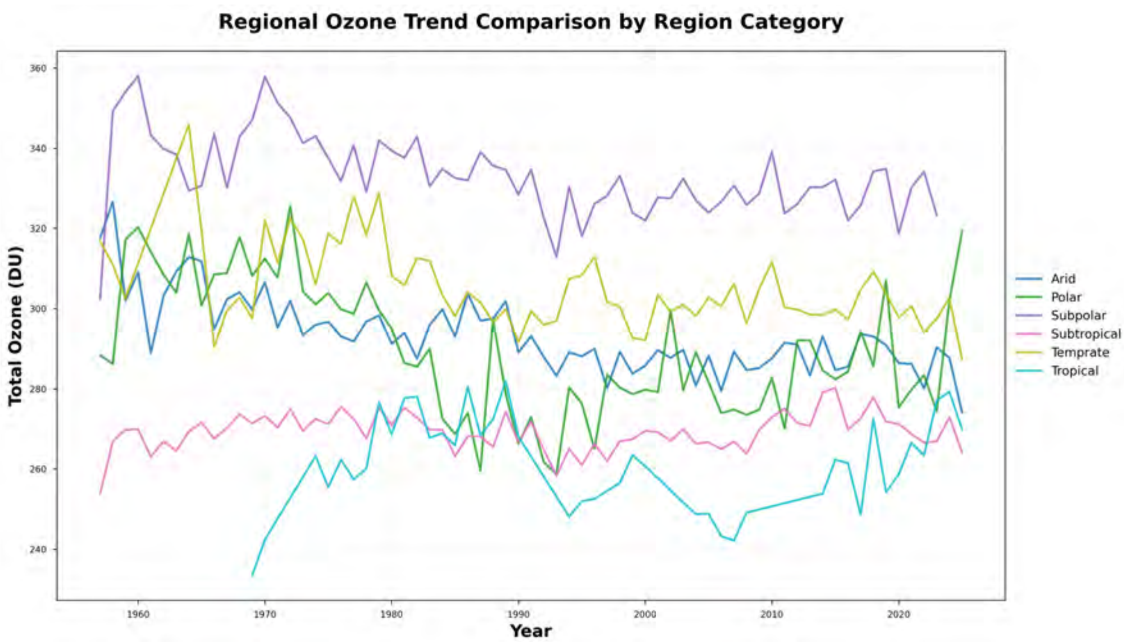

Figure 5.14: Regional Ozone Trend Comparison

These variations highlight the extent to which atmospheric forces have a strong influence on the behavior of ozone. The production of tropical ozone is low because of the extensive vertical movement and deep instead of shallow convection. The ozone at higher latitudes, on the contrary, is influenced by temperature-contingent photochemical reactions, polar vortex separation, and seasonal transport in the Brewer-Dobson cycle. To understand the long-term trends in ozone, it is important to understand that regions are governed by their respective climatic processes, which are the most crucial factors in determining the trends in the patterns of depletion and recovery. From a DiD perspective, differences in post-recovery trajectories across regions can be visually interpreted as heterogeneous patterns to the same policy intervention, reflecting region-specific atmospheric conditions rather that differences in regulatory exposure.

5.4.3 Recovery Slope Ranking

The long-term estimates of the growth rates of regions along the recovery slope (DU per year) are compared in Figure 5.15. The tropical region presents the highest positive recovery (+0.314 DU/yr), and this signifies that the climate is moving slowly towards increasing its ozone value, though at very low levels. There is moderate but positive recovery in the Temperate and Subtropical regions, which is in line with a stabilized stratosphere chlorine environment as a result of the Montreal Protocol.

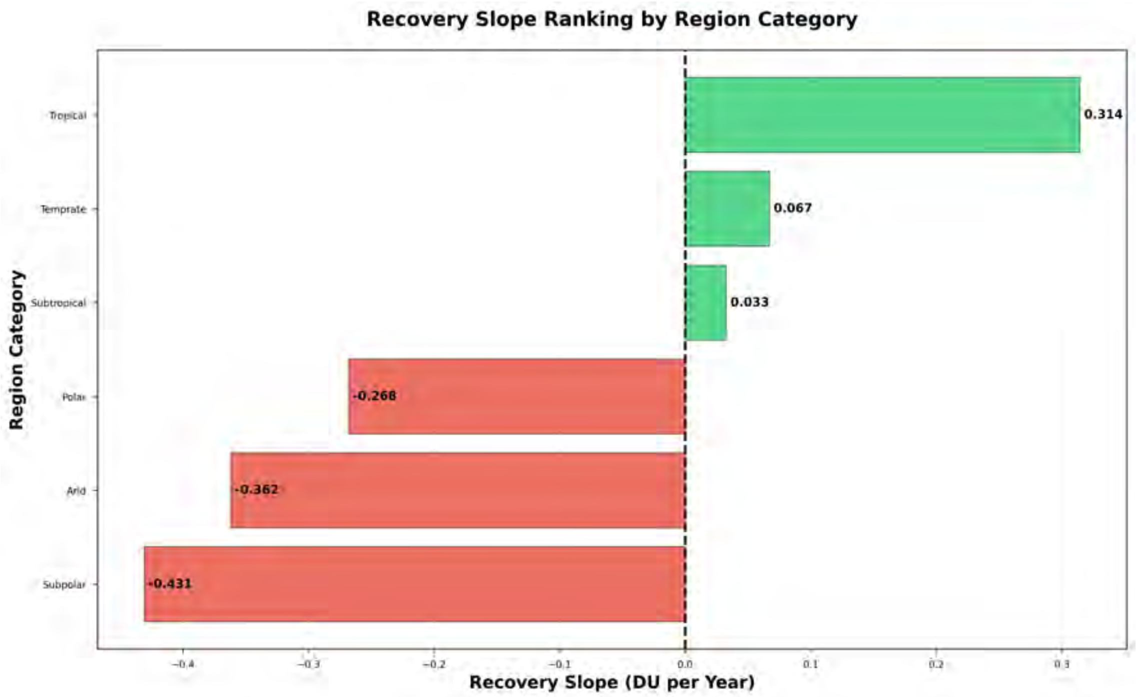

Figure 5.15: Recovery Slope Ranking

Conversely, the negative slopes of Polar, Subpolar, and Arid areas show negative slopes in this data (e.g., Subpolar: -.4306 DU/yr, Arid: -.3619 DU/yr). These negative slopes do not imply that depletion is necessarily ongoing, but instead, it is probably a manifestation of continued variability, dynamic instability, and heterogeneous recovery at high latitudes and in arid climates. Polar and Subpolar ozone are very susceptible to extreme seasonal processes, and a linear slope might not be sufficient to depict episodic recovery events. Meanwhile, the negative slope of the Arid region implies the unique atmospheric characteristics of this segment in which the recovery can be obscured by the local weather patterns. All in all, the slope comparison underlines that the process of recovery varies unevenly around the world and is highly subject to the local atmospheric conditions. In this context, recovery slope rankings provide a relative comparative assessment comparison similar to SCM-based reasoning, where regions exhibiting slower recovery display weaker deviation from the common post-intervention trend compared to faster-recovering regions.

5.4.4 Prediction Error and Uncertainty Assessment

The performance of prediction across regions is compared with the help of Mean Absolute Error (MAE) as presented in Figure 5.16. The Tropical region exhibits the greatest MAE (0.42) and hence the ozone values are most difficult to forecast. This is aligned with its highly deviating convective climates, disjointed transitory swings, and the swift mixture of elements. Moderate values of Polar (0.15) and Subtropical (0.13) are the reflection of their active seasonal changes and oscillation of their ozone.

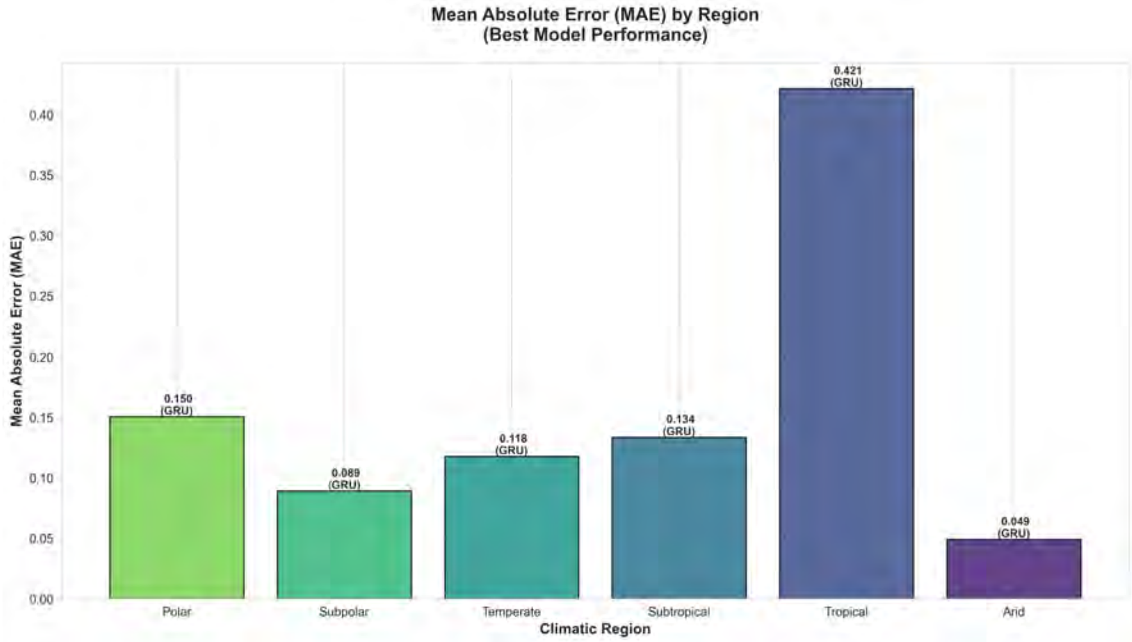

Figure 5.16: Prediction Error (MAE) by Region

Arid region, in turn, has the lowest MAE (0.049), which indicates that the ozone process in the area is very stable and predictable, with a negative slope. The Temperate region also has a low MAE (0.118), which is indicative of smoother long-term behavior and excellent observational coverage. These MAE values are equivalent to the GRU model that has the highest predictive performance in all climatic regions.

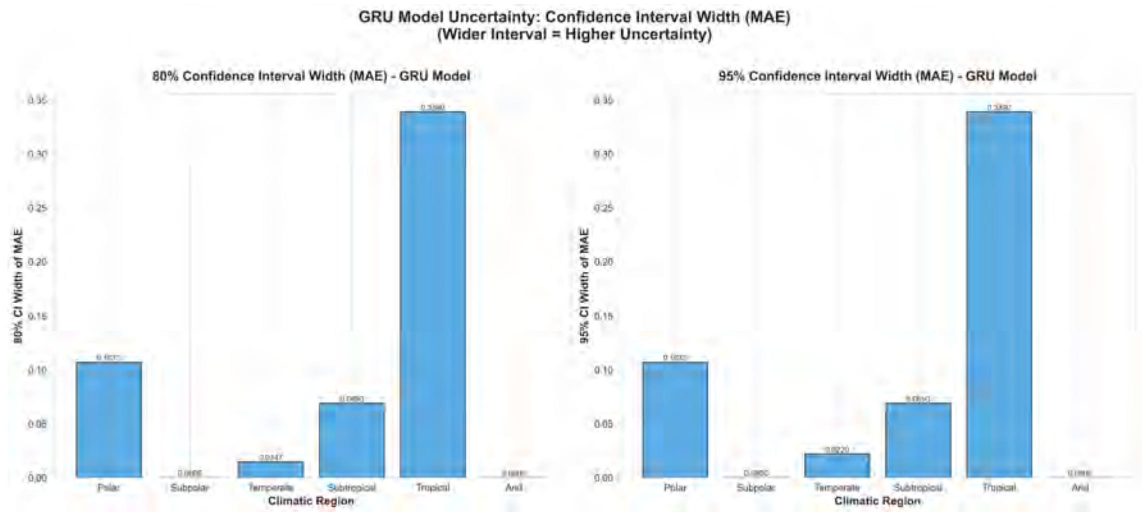

Figure 5.17: Confidence Interval Width (MAE)

Figure 5.17 has uncertainty estimates that indicate evident regional variations in the confidence of the GRU model. The interval (80% 95% CI width = 0.3390) of the Tropical region is the largest, and ozone behavior in this region is highly unpredictable because of the high level of dynamical and convective activity. There is moderate uncertainty in the Subtropical (0.0690) and Polar (0.1070) regions, which is in line with their variability and cloudy atmosphere during the seasons.

On the other hand, Subpolar and Arid regions have a CI width of 0.0000. This is because the countries in these regions will have the same values of MAE (Subpolar: 0.089 in both CAN and USA; Arid: 0.049 in AUS), and thus there is no model form to represent the uncertainty. There is no variability in the resampling repetition, hence the confidence intervals reduce to a point indicating total model agreement.

The general trend of the uncertainty pattern resembles that of the MAE results: the areas of greater atmospheric variations give wider confidence envelopes, and those areas that are very stable or uniform give narrow or none ranges. This supports the fact that predictive uncertainty is closely associated with internal regional processes and consistency of data.

Region	**Recovery Slope (DU/year)**	**MAE**	**Best Model**	**Interval Width (95%)**	**R^2**	**Data Coverage**
Arid	-0.3619	0.049	GRU	0.24	0.9980	100.0%
Polar	-0.2679	0.150	GRU	0.74	0.9964	100.0%
Subpolar	-0.4306	0.089	GRU	0.44	0.9976	100.0%
Subtropical	0.0327	0.134	GRU	0.66	0.9954	100.0%
Temperate	0.0671	0.118	GRU	0.58	0.9971	100.0%
Tropical	0.3143	0.422	GRU	2.07	0.9815	100.0%

Table 5.12: Global Comparison Table

The global comparison Table 5.12 summarizes regional ozone behavior using recovery slopes, MAE, uncertainty intervals, R^2 values, and data coverage. The Tropical region shows the strongest positive recovery but also the highest error and widest intervals, indicating clear long-term improvement yet unstable short-term dynamics. The Temperate region demonstrates the most consistent profile, with moderate recovery, low MAE, and narrow uncertainty bands, making it the most reliable indicator of ozone stabilization.

The arid region exhibits the lowest prediction error and the tightest uncertainty intervals, but a negative slope, indicating stable yet non-recovering ozone levels over the period of analysis. Because data coverage is complete (100%) across all regions, the validity of cross-regional comparisons is strengthened.

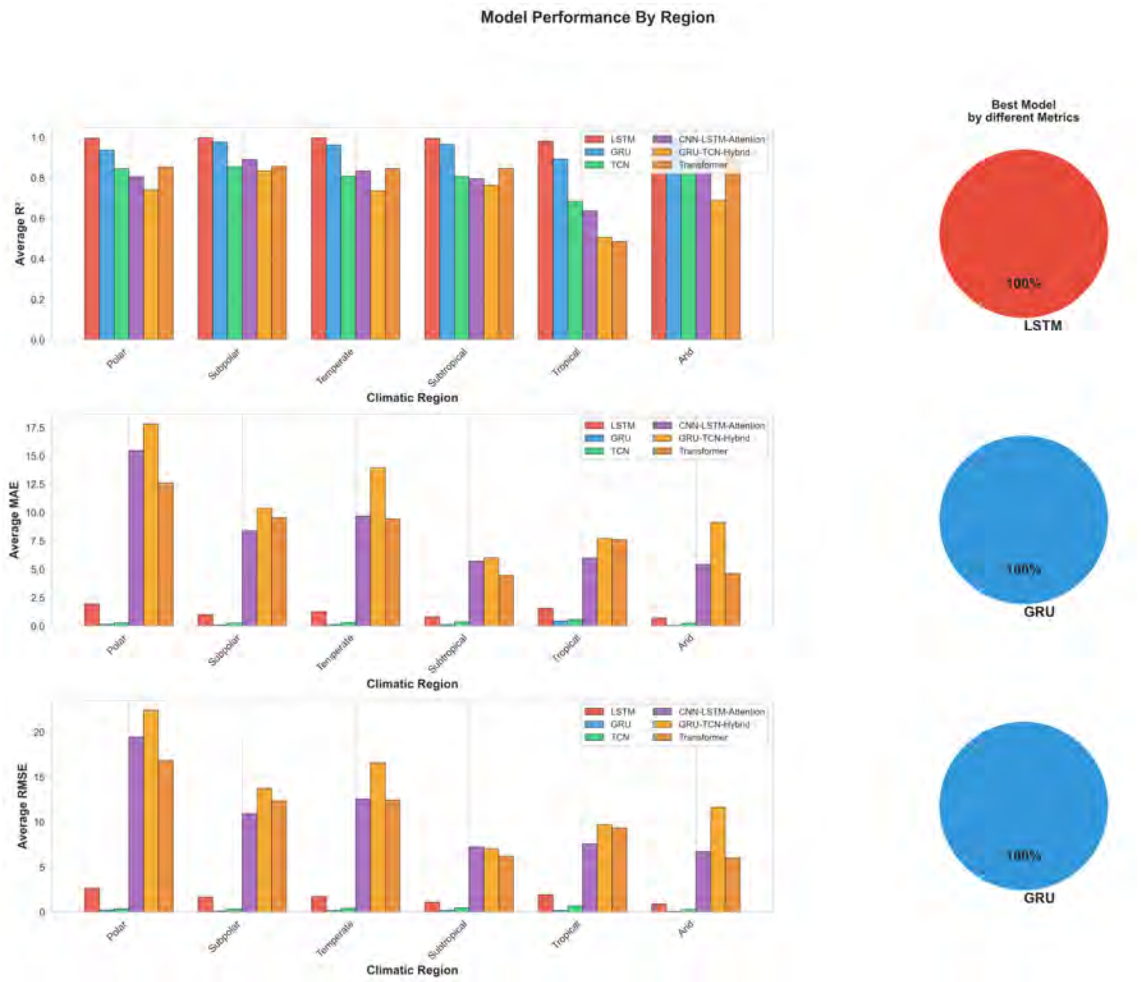

Figure 5.18: Model Performance by Region

These results are also confirmed by a region-wise comparison of the deep learning models presented in Figure 10.6. As shown by the performance visualisation, the LSTM model achieves the highest R^2 scores across all regions and performs no worse than the GRU, TCN, CNN–LSTM–Attention, GRU–TCN hybrid, and Transformer models in characterising the structural variability of ozone time series. In contrast, the GRU model attains the highest numerical accuracy, achieving the lowest MAE and RMSE values across all climatic regions. This pattern is consistent across all six climatic regions: LSTM provides the highest R^2 scores in every region, while GRU yields the lowest MAE and RMSE values in all six regions. These findings reveal that, in comparison to LSTM, which is more effective in capturing the long-term relations in ozone dynamics, GRU provides more stable and accurate short-term predictions. Altogether, these results show the relative advantages of the two architectures as well as indicate their locality-based appropriateness in ozone prediction.

Chapter 6

Discussion

6.1 Overview of Ozone Recovery Findings and Policy Guidelines

The review performed in this paper indicates that ozone recovery after the Montreal Protocol is highly heterogeneous, both in terms of recovery speed and stability, reliability, and interpretability. The outputs of machine learning models show that the trend of regional ozone cannot be properly evaluated with the help of one change indicator. Rather, interpretation policy-relevant should be a joint measure of the direction of recovery, time-related behavior, predictive confidence, and cause attribution.

There are also noticeable variations in the direction of the change of ozone across regions, with strong recovery, slow, stagnant, or episodic being the most common form of change. Areas are also different in the process of recovery with time. There are those with smooth and uninterrupted improvement, and those with immense seasonal variation or seasonal depletion. Such phenomena with regard to time influence not just the occurrence of recovery, but also the susceptibility of areas to temporary ozone depletion.

Recovery signals have also been found to vary across the regions. The results of model evaluation based on the prediction range, error level, and congruence across various models indicate that recovery estimates can be taken seriously in parts of the world and significantly less in others. Because of that, very different degrees of certainty can be attached to similar recovery trends depending on the location. Such variations in confidence may be ignored, which may result in an overestimation of the recovery in areas with a high uncertainty rate.

Also, causal inference based on Difference-in-Differences and Synthetic Control Methods offers the proof that the post-1987 decreases in ozone concentration can be linked to the adoption of the Montreal Protocol and not to the natural atmospheric variability. The causal attribution, however, is not evenly strong in all the regions, and this is a manifestation of variation in atmospheric dynamics and receptiveness to regulatory intervention. This also highlights the need to interpret regionally specific than depending on global average recovery indications.

Combined, all these various dimensions of analysis, direction of recovery, time trend, predictive reliability, and causal attribution are condensed to assess the extent of convergence of evidence among regions. Where these signals coincide, then recovery interpretation is said to be robust; where they differ, then recovery is still uncertain, variable, or weakly supported. This synthesis gives various regional interpretation patterns, including strong recovery, weak recovery, steady yet stagnant conditions, and very weak or uncertain recovery.

These interpretation profiles provide the policy-reasoning and model-result analysis bridging. Instead of encouraging the development of new policy instruments, they guide the choice, prioritisation, and regional implementation of the existing policy instruments that are based on the Montreal Protocol policy pool. This way, the future policy guidelines are specifically based on the synergy of power and stability and consistency of the evidence, but at the same time, they are pegged on an existing and empirically tested global governance model.

6.2 Montreal Protocol–Derived Policy Pool (Global Baseline Framework)

In this paper, it is assumed that the Montreal Protocol acts as a blueprint of global governance on the basis of which a set of policy tools is finite. Such tools form a pool of policies that is uniform across regions. The policy recommendations that are postulated in this thesis do not offer new policy instruments, but offer novel, data-based guidelines to the prioritisation and implementation of the existing policy instruments, according to the region-specific behavior of ozone recovery identified by predictive modeling [1][9].

6.2.1 Montreal Protocol–Derived Policy Pool

The policy pool, which is based on the Montreal Protocol, includes the following instruments:

- **P1. Regulatory Control, Phase-Out Measures**
  Such are legally binding controls on the production and consumption of ozone-depleting substances and enforcement and compliance mechanisms.

- **P2. Adaptation, Adjustment, and Amendment Mechanisms**
  Flexible governance structures with policy to be revised in line with changing scientific evidence and uncertainty.

- **P3. Scientific Monitoring, Reporting, and Assessment**
  Constant ozone monitoring in the atmosphere, evaluation by models, and scientific reporting of results to determine the progress of ozone recovery.

- **P4. Early Detection and Response Systems**
  Monitors and reaction mechanisms to intermittent or extreme instances of ozone depletion, especially those caused by seasonal or atmospheric events.

- **P5. International Co-ordination and Multilateral Co-operation**
  Cross-border information exchange, harmonisation of implementation, and joint scientific and regulatory action.

- **P6. Transition and Implementation Support in Technology**
  Supporting mechanisms that enable the embrace of the alternatives to ozone-depleting substances, such as capacity building.

- **P7. Awareness and Risk Mitigation of Health to The Public**
  Interventions meant to minimize human exposure to harmful ultraviolet radiation via advisories, awareness, and risk communication.

- **P8. Long-term Surveillance and Recovery Reversal Prevention**
  Continuous policing and supervision to ensure that we do not revert to the past or secondary depletion of ozone after the initial recovery.

These tools represent the fundamental regulatory, scientific, adaptive, and cooperative processes incorporated into the Montreal Protocol and its further amendments [1][9].

6.3 Regional Interpretation Profiles and Policy Guidelines

This interpretation framework combines machine-learning-driven evidence with policy advice regionally. Ozone recovery behavior is used to assess each region of a region, which includes recovery direction, recovery behavior, predictive confidence, and causal attribution. The profile of interpretations thus obtained guides the choice and prioritisation of policy instruments within the policy pool of the Montreal Protocol, so that the regional policy guidelines are made expressly based on empirical evidence.

6.3.1 Polar Regions (Antarctica – ATA, Greenland – GRL)

Interpretation Profile

Simulation outcomes show that there is a weak or even no long-term ozone recovery in polar areas. The recovery patterns are very erratic, with very strong seasonal variations and intermittent outbursts caused by the polar vortex activity and changes in stratospheric temperatures. The predictive confidence is low because the ranges of uncertainty are wide in the model outputs. Although causal analysis confirms that international regulation has halted ozone's further degeneration, recovery behavior remains considerably affected by the Western atmospheric natural processes. As a result, there is less convergence in direction of recovery, temporality, predictive confidence, and causal attribution, which suggests an uncertain and very variable recovery profile.

Policy Guidelines

Given the variability and low predictability in polar ozone recovery, the policy principles on the recovery focus more on early warning and anomaly response mechanisms (P4), scientific monitoring and assessment (P3) than on long-term recovery goals. The mechanisms of adaptive adjustment (P2) should increase preparedness-oriented governance at high-risk times, especially during the polar spring, which facilitates a rapid adjustment of policy to unexpected atmospheric variations. Moreover, the international coordination and cooperation (P5) needs to be enhanced because polar ozone variability is transboundary and cannot be restricted to regional limits.

6.3.2 Subpolar Regions (Canada – CAN, United States of America – USA)

Interpretation Profile

Moderate recovery in ozone is seen in Subpolar regions, but at a slower rate than is seen in temperate areas. The seasonality of recovery behavior is high, and there is a great deal of variability throughout the annual cycle. The confidence of prediction is medium because the models of machine learning are always able to predict long-term recovery patterns, but are always susceptible to the seasonal atmospheric circulation patterns. Causal analysis assists in the contribution of the Montreal Protocol to overall recovery, but recovery outcomes differ across seasons, which are also due to the effect of regional atmospheric processes. All these signals point to a slow and seasonally fluctuating recovery profile with moderate convergence evidence.

Policy Guidelines

Since there is a high seasonal variability with moderate predictive ability in the case of subpolar ozone recovery, scientific monitoring and assessment (P3), which is adjusted to the seasonal risk period, especially during winter and spring when atmospheric conditions are most favorable to support ozone depletion, are prioritised within the policy guidelines. Adaptive adjustment mechanisms (P2) must be implemented to consider seasonal atmospheric impacts to ensure some flexibility in recovery evaluation schedules instead of making use of standard annual standards. Primary focus on regulatory control and compliance mechanisms (P1) will be relevant to maintain the long-term recovery, but it is also essential to acknowledge the fact that climate-driven variability can limit short-term recovery indicators. Furthermore, the incorporation of ozone aspects into seasonal environmental planning practices helps to have a more context-specific understanding of recovery progress without going beyond current instruments of the Montreal Protocol.

6.3.3 Temperate Regions (Argentina – ARG, China – CHN, New Zealand – NZL)

Interpretation Profile

Obvious and sustained ozone recovery is reflected in temperate regions. The recovery pattern is quite smooth with little temporal variation, and predictive power is better compared to polar and subpolar areas. The attribution of the cause of the recovery is high to the Montreal Protocol, and the convergence in recovery direction, temporal behavior, predictive confidence, and causal evidence points to highly reliable recovery patterns. Together, these attributes portray a strong and solid recovery profile.

Policy Guidelines

Since evidence of convergence and high predictive confidence is strong in temperate regions, the maintenance of existing regulatory control and phase-out courses (P1) is a priority area in policy guidelines to maintain the gains on recovery that have already been obtained. To prevent any possible subsequent degradation, recovery reversal prevention (P8) and long-term follow-up (P3) must be maintained, with assistance of further scientific monitoring and evaluation (P3). The consistency and dependability of recovery in such areas also provide reasons to use them as reference examples to assess the overall policy effectiveness in the framework of the Montreal Protocol, even though it does not necessarily mean that the results can be directly applied to more variable regions. Besides that, despite the overall recovery, the periodic seasonal thinning still needs further focus on the mitigation of the effects on the population health risks and awareness (P7) regarding the ultraviolet exposure in the definite period, when the high-risk situation takes place.

6.3.4 Tropical Regions (Mexico – MEX, American Samoa – ASM)

Interpretation Profile

There are high and steady ozone recoveries that do not have large seasonal variations in tropical areas. The confidence in prediction is usually good, with the outputs of the models demonstrating consistent recovery patterns with weak sensitivity to seasonal atmospheric circulation, but there are short-term fluctuations occasionally because of the large-scale climatic factors. The effectiveness of international regulation is supported by causal evidence, and recovery signals occur in agreement in terms of recovery direction, temporal behavior, predictive confidence, and causal attribution. All these features demonstrate a highly consistent and robust recovery profile with well-evidenced convergence.

Policy Guidelines

Since the recovery prognoses and predictive confidence are high in tropical areas, the policy recommendations focus on the change of priorities in the recovery enforcement in favour of mitigation of the risk to public health and awareness (P7). Even though the recovery of ozone is significant, the levels of ultraviolet exposure at the base are still high in the tropical latitude; thus, it is justified to enhance the effectiveness of the public awareness and education programs regarding sun safety. Scientific monitoring and assessment (P3) must remain at a baseline level to make unusual or climate-induced perturbations evident without the addition of stronger regulatory control. In these regards, ozone-related considerations best fit into long-term schemes of accessory health planning, as opposed to recovery-oriented regulatory escalation, and are entirely consistent with the current instruments of the Montreal Protocol.

6.3.5 Subtropical Regions (Brazil – BRA, India – IND)

Interpretation Profile

Subtropical ones have a mixed increase in ozone with a pronounced seasonal change and are strongly dependent on regional climates. The predictive confidence is less than in the tropical and temperate parts, recovery behavior varies among subregions, and there are broader ranges of uncertainty in model outputs. Although the causal analysis confirms the place of international regulation in the overall recovery, convergence in terms of recovery direction, temporal pattern, predictive confidence, and causal attribution is partial. Taken together, these features resemble a moderate recovery profile that has seasonal and regional differences.

Policy Guidelines

Since the recovery behavior and predictive confidence are found to be less predictable and more heterogeneous in subtropical areas, the policy recommendations focus on adaptive adjustment and flexible governance processes (P2) that will enable the regionally differentiated application in the national policy frameworks. The improved scientific observation and evaluation (P3) should be directed to the most important subregions based on climatic sensitivity and the recovery indicators, where it is most uncertain, instead of being randomly increased across the whole country. Here, the implementation strategies should be adapted locally rather than strictly and uniform strategies that operate nationwide, but it must be in line with the current regulatory control and compliance frameworks (P1). As well, the increased variability rates justify the enhanced mitigation and awareness about the risk of ultraviolet exposure to the population (P7), especially in the seasons when the ozone conditions are more unpredictable.

6.3.6 Arid Regions (Australia – AUS)

Interpretation Profile

In dry areas, the conditions of ozone are stable and to a large extent not improving. The direction of recovery is practically flat, temporal variability is low and predictive confidence is high, which implies that ozone behaviour is predictably described with the models. With causal analysis showing that international regulation has avoided the continued ozone decline, there is still limited evidence of meaningful recovery. Collectively, these data points represent a steady yet flat recovery picture, which is defined by stasis but not growth.

Policy Guidelines

Since the behaviour of recovery is stable but not improving in arid areas, and there is high predictive confidence in the same, the policy guidelines will focus on long-term surveillance and recovery reversal prevention (P8) rather than pursuing rapid recovery targets. They should have sustained scientific surveillance and evaluation (P3) to identify any variation of the set baselines as well as prompt detection of any impending deterioration. Moreover, due to the elevated values of ultraviolet exposure common to arid environments, the emphasis on the mitigation of the risk to public health and population awareness should be further addressed (P7), especially among the population with prolonged outdoor exposure. The consistency of ozone levels in these areas also helps to utilize the sharply defined early warning levels (P4), which allows responding to the policy in time under the condition of a change in the historically stable conditions.

The Policies
P1. Measures of Implementation and Regulation Control
P2. Adaptive Adjustment and Amendment Mechanisms
P3. Science monitoring, Reporting, and Evaluation
P4. Early Warning and Anomaly Response Strategies
P5. International Co-ordination and Relations
P6. Technology Transition and Implementation
P7. Risk management, population Health Awareness
P8. Prolonged observation and Prevention of Recovery reverse

Region	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**
Polar		✓	✓	✓	✓			
Subpolar	✓	✓	✓					
Temperate	✓		✓				✓	✓
Tropical			✓			✓	✓	
Subtropical	✓	✓	✓				✓	
Arid			✓	✓			✓	✓

Table 6.1: Policy Guidelines For Each Region

6.3.7 Regional Policy Sets

Below, each region is represented as a set containing its selected parameters:

$$\begin{aligned}
\textbf{Polar} &= \{P2, P3, P4, P5\} \\
\textbf{Subpolar} &= \{P1, P2, P3\} \\
\textbf{Temperate} &= \{P1, P3, P7, P8\} \\
\textbf{Tropical} &= \{P3, P6, P7\} \\
\textbf{Subtropical} &= \{P1, P2, P3, P7\} \\
\textbf{Arid} &= \{P3, P4, P7, P8\}
\end{aligned}$$

6.3.8 Validation of Policy Relevance through Causal Analysis

The machine learning models may be utilized to determine the trends and spatial differences in the ozone recovery. Nonetheless, in order that such trends can play a significant role in policy discussions, it should be shown that seen fluctuations can be related to regulatory intervention as opposed to natural atmospheric fluctuations.To this end, the paper applies the causal inference techniques, namely, Difference-in-Differences (DiD) and Synthetic Control Method (SCM), to investigate the possibility of the post-1987 shifts in the ozone levels being due to the establishment of the Montreal Protocol.

The DiD analysis is used to compare how ozone would act prior to the adoption of the Montreal Protocol and after the adoption of the protocol but makes use of temporal trends that are shared across the region. The findings reveal that there is a noticeable change in the ozone patterns after 1987, and recovery trends are not similar to the trend observed in the pre-regulation period. This approximate break in time indicates that gains after 1987 cannot be associated with natural changes alone and are correlated with the international action on regulation. Notably, this effect is larger in some regions, which means that the effect of regulation is not homogeneous in space.

This interpretation is further seen in the SCM analysis that builds counterfactual situations to illustrate ozone's expected behavior under the Montreal Protocol. In most areas, especially in high latitudes, the observed post-1987 levels of ozone are typically higher than the levels estimated assuming no-policy conditions. Such a deviation suggests that, in the absence of a regulation, ozone depletion would have likely continued, and that the Montreal Protocol changed the direction of ozone change in the long term. Meanwhile, the magnitude and stability of this divergence vary across regions, which reflects the impact of regional atmospheric processes on policy outcomes.

Combined, the DiD and SCM findings indicate that the ozone recovery patterns obtained by the machine learning models are policy-relevant, meaning that they are affected by policy and not by only chance or natural cycles. Nonetheless, the regional difference in the strength of causal effects shows that the effects of the policies are conditional on regional atmospheric behavior and responsiveness to recovery.

All in all, this causal study is not aimed at assessing the efficacy of a particular policy instrument but only at developing a plausible correlation between regulatory intervention and recorded recovery patterns. It is this connection that gives the analytical foundation of the translation of the model-generated recovery profiles into region-specific policy guidelines, as well as the explanation of a prudent and varied policy response that considers both the strength of evidence and uncertainty.

6.3.9 Risk, Uncertainty, and Confidence in Policy Interpretation

Patterns of ozone recovery, along with the uncertainties about them, are at the centre of the establishment of the degree to which one can be confident in constructing and implementing policy guidelines on a region-specific level. Although causal analysis proves that the Montreal Protocol has indeed resulted in realized changes in ozone levels, the intensity, consistency, and dependability of model-inferred recovery signals differ significantly across areas. The outputs of machine learning demonstrate that there are comparatively stable recovery patterns in certain locations with a limited uncertainty interval, and there exist significantly variable areas with a wide prediction interval. These disparities directly determine the level of confidence with which the recovery trends can be interpreted to be used in policy-relevant ways.

Patterns of ozone recovery, along with the uncertainties about them, are at the centre of the establishment of the degree to which one can be confident in constructing and implementing policy guidelines on a region-specific level. Although causal analysis proves that the Montreal Protocol has indeed resulted in realized changes in ozone levels, the intensity, consistency, and dependability of model-inferred recovery signals differ significantly across areas. The outputs of machine learning demonstrate that there are comparatively stable recovery patterns in certain locations with a limited uncertainty interval, and there exist significantly variable areas with a wide prediction interval. These disparities directly determine the level of confidence with which the recovery trends can be interpreted to be used in policy-relevant ways.

Subpolar and subtropical areas are in the middle of these extremes. Seasonal variations and climatic variability in these areas affect recovery cues, leading to moderate predictability and convergence of evidence. The policy principles within such circumstances must therefore concern responsivity and adaptation of the current policy tools, and not long-term commitments that are founded on only one meaning of recovery progress.

The policy relevance of the findings should not be compromised by explicit consideration of uncertainty and confidence: instead, it will make the policy interpretation accountable, proportionate, and within the bounds of the evidence provided.

6.3.10 Alignment of Ozone Recovery Policy Guidelines with the Sustainable Development Goals (SDGs)

The regional policy principles that were formulated within the framework of the current research align with and complement the United Nations Sustainable Development Goals (SDGs) as they help to interconnect scientific ozone recovery governance with the long-term sustainability, population health, and climate governance goals. The study is based on machine learning-based recovery analysis and regional interpretation, causal validation, and explicit attention to uncertainty, thus providing an evidence-based framework to combine ozone-related governance with the global development objectives, without assuming homogeneous or average-based policies.

SDG 3: Good Health and Well-Being
This is facilitated by the policy guidelines that focus on the management of risks of exposure to ultraviolet (UV) radiation. Areas with great changes in ozone and intermittent depletion, especially polar and subpolar areas, enjoy better monitoring and early warning systems, which allow predicting periods with high risk of UV. Contrarily, the areas with more stable recovery patterns, including tropical and subtropical ones, enable robust public health planning and specific awareness campaigns. Such strategies help in alleviating the health hazards in the long term, such as skin cancer and eye adverse health issues, yet they do not conflict with the current instruments of ozone governance.

SDG 11: Sustainable Cities and Communities
This is achieved by using ozone recovery data to monitor the environment and conduct planning in safer and more resilient cities. The determination of areas in which the ozone readings are observed to be changing abruptly over short intervals

of time signifies the necessity to have flexible monitoring systems that can respond to the fluctuating atmospheric conditions in the short term. Having ozone and ultraviolet risk information in urban and regional planning assists in establishing safer neighborhoods, improved preparedness, and the capacity to make superior local decisions in those locations where individuals are more exposed to risk, depending on the season.

SDG 13: Climate Action
This will be addressed by means of model-based evidence to track the recovery of ozone and assess the effectiveness of environmental regulations. Regional variations in ozone recovery also indicated that climate policies are not uniform across the board, but rather, the policies are optimally suited to the local atmosphere. The existence of predictive models and prudent management of uncertainty helps evaluate the international settlements, the necessity of continued cooperation at the global level, and the possibility of adaptation of the policies to the changes in the atmospheric conditions taking place over time.

6.3.11 Overall Implications and Policy Significance

Comprehensively, this research offers region-specific, causal and uncertainty-sensitive results on the effects of the Montreal Protocol on ozone recovery by showing that the post-1987 gains are not due to natural atmospheric variability only but are accompanied by the implementation of regulatory intervention that is relatively uneven in clinical regions with respect to recovery results. The analysis separates the regions with stable recovery and the ones with slow, stagnant, or very variable recovery patterns by jointly interpreting recovery direction, temporal behavior, predictive confidence, and causal attribution. This combined strategy underpins differentiated policy interpretation, such that those regions that are characterized by high uncertainty and volatility should receive more flexible and preparedness-oriented policy management, whereas more durable and maintenance-oriented policy advice should be taken by regions with stable recovery conditions. Causal analysis determines the policy relevance of the observed trends of recovery, whereas explicit uncertainty consideration makes clear the degree of confidence with which it is possible to interpret and implement region-specific policy guidelines. These findings contribute to the understanding of the ozone recovery-related policy guidelines in the context of Sustainable Development Goals, as aimed at SDG 3, owing to management of UV-related health risks, SDG 11, owing to supporting resilient and informed communities, and SDG 13, owing to informing climate action, owing to the use of adaptable environmental monitoring data. Altogether, the study presents a salient model of evidence-based interpretation of environmental policy that can be applied to predictive modeling, causal inference, and uncertainty awareness in an existing context of world governance.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

This research conducts an inclusive analytical examination of long-term ozone stratospheric processes in different types of climatic areas with some advancements and limitations to ozone restoration in the world. Complex ozone behaviors that are seasonal, nonlinear, and region-specific were learned by advanced deep-learning models-LSTM, GRU, TCN, Transformer, and hybrid architecture coupled with feature engineering and robust evaluation. LSTM was the most explanatory and GRU had the least error in predictions, Transformers and hybrids were more dependable in long-term predictions, and TCN was more able to reflect the seasonal changes. Quantification of uncertainty, and cross-region meta-analysis demonstrated that there was strong recovery evidence in Tropical and Temperate regions and slower, or more fluctuating in Polar, Subpolar, and Arid regions. The positive effect of the Montreal Protocol was verified by change-point detection and causal analyses, and the heterogeneity of the regions was also outlined. These observations can address SDG 13 (Climate Action) by policy guidance, SDG 3 (Good Health) by mitigating the effects of UV rays, and SDG 11 (Sustainable Cities) because of environmental planning in cities. In general, the research paper provides a scalable policy-relevant analytical framework that combines deep-learning forecasting, uncertainty, and causal evaluation to track the stratospheric ozone and make future environmental decisions.

7.2 Future Work

The fact that we were unable to extract the ozone data of the subpolar and arid regional countries was one of our limitations, and therefore, as a future enhancement, we will be able to include countries that are already dry and moderate in their environment. Otherwise, the health impact of these regional countries can be judged based on the data on gene mutation, thus, the number of Skin Cancer or the illnesses caused by UV radiation could be calculated. It would assist the policy analysis to be more precise since, with climate, the health of people would play a role in this as well.

Bibliography

[1] United Nations Environment Programme, *Montreal Protocol on Substances that Deplete the Ozone Layer*. Nairobi: UNEP, 1987.

[2] E. A. Parson, "International protection of the ozone layer," *Green Globe Yearbook*, pp. 19–28, 1996.

[3] D. N. Politis, "The impact of bootstrap methods on time series analysis," *Statistical Science*, vol. 18, no. 2, pp. 219–230, 2003, ISSN: 08834237. [Online]. Available: http://www.jstor.org/stable/3182852 (visited on 12/01/2025).

[4] R. Boyte, "The principle of common but differentiated responsibilities," *New Zealand Journal of Environmental Law*, vol. 14, pp. 63–101, 2010. [Online]. Available: https://search.informit.org/doi/10.3316/informit.865241634494134.

[5] J. A. Mäder, J. Staehelin, T. Peter, D. Brunner, H. E. Rieder, and W. A. Stahel, "Evidence for the effectiveness of the montreal protocol to protect the ozone layer," *Atmospheric Chemistry and Physics*, vol. 10, pp. 12 161–12 171, 2010. DOI: 10.5194/acp-10-12161-2010. [Online]. Available: https://www.atmos-chem-phys.net/10/12161/2010/.

[6] P. Canan, S. O. Andersen, N. Reichman, and B. Gareau, "Introduction to the special issue on ozone layer protection and climate change: The extraordinary experience of building the montreal protocol, lessons learned, and hopes for future climate change efforts," *Journal of Environmental Studies and Sciences*, vol. 5, no. 2, pp. 111–121, 2015.

[7] United Nations Development Programme, *25 years of the montreal protocol*, UNDP Publication, Nov. 2015. [Online]. Available: https://www.undp.org/publications/25-years-montreal-protocol.

[8] P. Nowack, P. Braesicke, J. Haigh, N. L. Abraham, J. Pyle, and A. Voulgarakis, "Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations," *Environmental Research Letters*, vol. 13, no. 10, p. 104 016, Sep. 2018. DOI: 10.1088/1748-9326/aae2be. [Online]. Available: https://doi.org/10.1088/1748-9326/aae2be.

[9] World Meteorological Organization and United Nations Environment Programme, "Scientific assessment of ozone depletion: 2018," World Meteorological Organization, Geneva, Global Ozone Research and Monitoring Project–Report No. 58, 2018.

[10] J. Wu, S. Guo, H. Huang, W. Liu, and Y. Xiang, "Information and communications technologies for sustainable development goals: State-of-the-art, needs and perspectives," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2389–2406, 2018.

[11] N. Xue, I. Triguero, G. P. Figueredo, and D. Landa-Silva, "Evolving deep cnn-lstms for inventory time series prediction," in *2019 IEEE Congress on Evolutionary Computation (CEC)*, 2019, pp. 1517–1524. DOI: 10.1109/CEC.2019.8789957.

[12] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, pp. 82–115, 2020. DOI: 10.1016/j.inffus.2019.12.012.

[13] J. Chapman *et al.*, "An assessment of practitioners approaches to forecasting in the presence of changepoints," *Quality and Reliability Engineering International*, 2020. DOI: 10.1002/qre.2712.

[14] I. E. Livieris, E. Pintelas, and P. Pintelas, "A cnn–lstm model for gold price time-series forecasting," *Neural Computing and Applications*, vol. 32, no. 23, pp. 17 351–17 360, 2020. DOI: 10.1007/s00521-020-04867-x.

[15] S. Mehtab, J. Sen, and S. Dasgupta, "Analysis and forecasting of financial time series using cnn and lstm-based deep learning models," in *Proceedings of the Symposium on Machine Learning and Metaheuristic Algorithms and Applications (DASA) / or conference where the paper was published*, full PDF available via ResearchGate, 2020.

[16] C. Betancourt, T. T. Stomberg, A.-K. Edrich, *et al.*, "Global, high-resolution mapping of tropospheric ozone – explainable machine learning and impact of uncertainties," *Geoscientific Model Development*, vol. 15, no. 11, pp. 4331–4354, 2022. DOI: 10.5194/gmd-15-4331-2022. [Online]. Available: https://gmd.copernicus.org/articles/15/4331/2022/.

[17] M. Coldewey-Egbers, D. G. Loyola, C. Lerot, and M. Van Roozendael, "Global, regional and seasonal analysis of total ozone trends derived from the 1995–2020 gto-ecv climate data record," *Atmospheric Chemistry and Physics*, vol. 22, no. 10, pp. 6861–6878, 2022. DOI: 10.5194/acp-22-6861-2022. [Online]. Available: https://acp.copernicus.org/articles/22/6861/2022/.

[18] A. Doury, S. Somot, S. Gadat, A. Ribes, and L. Corre, "Regional climate model emulator based on deep learning: Concept and first evaluation of a novel hybrid downscaling approach," *Climate Dynamics*, vol. 60, no. 5, pp. 1751–1779, 2022. DOI: 10.1007/s00382-022-06343-9. [Online]. Available: https://doi.org/10.1007/s00382-022-06343-9.

[19] D.-H. Ko, K. Cho, and S. Rao, "Machine-learning-based near-surface ozone forecasting model with planetary boundary layer information," *Sensors*, vol. 22, no. 20, p. 7864, 2022. DOI: 10.3390/s22207864. [Online]. Available: https://www.mdpi.com/1424-8220/22/20/7864.

[20] Y. Lyu, Q. Ju, F. Lv, J. Feng, X. Pang, and X. Li, "Spatiotemporal variations of air pollutants and ozone prediction using machine learning algorithms in the beijing-tianjin-hebei region from 2014 to 2021," *Environmental Pollution*, vol. 306, p. 119 420, 2022, ISSN: 0269-7491. DOI: https://doi.org/10.1016/j.envpol.2022.119420. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0269749122006340.

[21] NASA, *Nasa study finds ozone layer on track to recovery*, https://www.nasa.gov/feature/goddard/2023/nasa-study-finds-ozone-layer-on-track-to-recovery, NASA Earth Science News, 2023.

[22] United Nations Environment Programme (UNEP), *Scientific assessment of ozone depletion: 2022*, https://ozone.unep.org/science/assessment/sap, World Meteorological Organization, 2023.

[23] M. P. Chipperfield, S. S. Dhomse, W. Feng, R. L. McKenzie, O. Morgenstern, and S. E. Strahan, "Stratospheric ozone – depletion, recovery and new challenges," *Atmospheric Chemistry and Physics*, vol. 24, no. 5, pp. 2783–2815, 2024. DOI: 10.5194/acp-24-2783-2024.

[24] K. Khan, M. Zeeshan, A. Moiz, *et al.*, "Influence of government effectiveness, health expenditure, and sustainable development goals on life expectancy: Evidence from time series data," *Sustainability*, vol. 16, no. 14, 2024, ISSN: 2071-1050. DOI: 10.3390/su16146128. [Online]. Available: https://www.mdpi.com/2071-1050/16/14/6128.

[25] C. Ma, J. Song, M. Ran, Z. Wan, Y. Guo, and M. Gao, "Machine learning-driven spatiotemporal analysis of ozone exposure and health risks in china," *Journal of Geophysical Research: Atmospheres*, vol. 129, no. 20, e2024JD041593, 2024. DOI: 10.1029/2024JD041593.

[26] J. Wang, T. Ju, B. Li, *et al.*, "Characterization of tropospheric ozone pollution, random forest trend prediction and analysis of influencing factors in south-western europe," *Environmental Sciences Europe*, vol. 36, p. 61, 2024. DOI: 10.1186/s12302-024-00863-3. [Online]. Available: https://enveurope.springeropen.com/articles/10.1186/s12302-024-00863-3.

[27] P. Annadate, N. Aher, P. Kulkarni, and R. Suryawanshi, "Resolving the ozone dilemma: An integration of game theory and time series forecasting," *International Journal of Computing and Digital Systems*, vol. 17, pp. 1–12, Jan. 2025. DOI: 10.12785/ijcds/1571018374.

[28] Y. Lin, M. Song, and B. van der Sluis, "Bootstrap inference for linear time-varying coefficient models in locally stationary time series," *Journal of Computational and Graphical Statistics*, vol. 34, no. 2, pp. 654–667, 2025. DOI: 10.1080/10618600.2024.2403705. eprint: https://doi.org/10.1080/10618600.2024.2403705. [Online]. Available: https://doi.org/10.1080/10618600.2024.2403705.

[29] M. A. Mujtaba, M. A. Munir, S. Ali, *et al.*, "Using machine learning for air quality prediction and sustainable urban planning," *Sustainable Futures*, vol. 10, p. 100 981, 2025. DOI: 10.1016/j.sftr.2025.100981.

[30] U.S. Environmental Protection Agency. "Basic ozone layer science." Accessed: 2025-10-08. (Mar. 2025), [Online]. Available: https://www.epa.gov/ozone-layer-protection/basic-ozone-layer-science.

[31] M. Zhang, Y. Tian, and M. K. Khan, "Time series approach to examine the impact of economic uncertainty, economic growth and financial development on human health," *Natural Resources Forum*, vol. 49, no. 2, pp. 2067–2080, 2025. DOI: 10.1111/1477-8947.12473.

[32] *woudc-ui*. [Online]. Available: https://woudc.org/en/.