Integrative Modeling of Coal-Induced CO_2 Emissions Using a Hybrid ARIMA–NNAR Architecture: A Nonlinear Time Series Forecasting Approach for Environmental Systems

Dissertation submitted to

Adamas University

For the Award of the Degree of

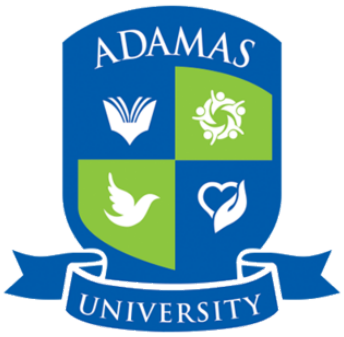
B.Sc.(Hons) Statistics and Data Analytics

by

USHOSREE RAHA

(UNIVERSITY REGISTRATION NO: AU/2022/0008058)

ADAMAS UNIVERSITY



Department of Mathematics

School of Basic and Applied Sciences

Adamas University - India

May, 2025

CERTIFICATE

This is to certify that **Ushosree Raha**, B.Sc. (Hons) Statistics and Data Analytics, has worked on the project entitled **'Integrative Modeling of Coal-Induced CO_2 Emissions Using a Hybrid ARIMA–NNAR Architecture: A Nonlinear Time Series Forecasting Approach for Environmental Systems'** under my supervision and guidance. The content of this report is original and has not been submitted elsewhere for the award of any academic or professional degree.

May, 2025
Adamas University

Mr. Anirban Maity
Adjunct Faculty, Department of Mathematics
Adamas University

DECLARATION

I declare that the thesis entitled **'Integrative Modeling of Coal-Induced CO_2 Emissions Using a Hybrid ARIMA–NNAR Architecture: A Nonlinear Time Series Forecasting Approach for Environmental Systems'** has been prepared by me under the guidance of Mr.Anirban Maity, Adjunct Faculty of Department of Mathematics, Adamas University. No part of this thesis has formed the basis for the award of any degree or fellowship previously.

May, 2025
Adamas University

Ushosree Raha

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to all those who have supported me throughout the completion of this thesis. My sincere thanks go to my supervisor, **Mr. Anirban Maity, Adjunct Faculty, Department of Mathematics**, for his invaluable guidance, insightful feedback, and unwavering support throughout this research. I am especially grateful to **Dr. Satyajit Das, Assistant Professor and HOD, Department of Mathematics**, and **Dr. Nav Kumar Mahato, Associate Professor, Department of Mathematics**, for their encouragement and academic support during the course of this work. I also extend my heartfelt thanks to the **Department of Mathematics** and **School of Basic and Applied Sciences, Adamas University** for providing an environment conducive to research and learning.

I am extremely thankful to **Robbie Andrew and Glen P. Peters, CICERO Center for International Climate Research, Oslo, Norway**, for providing the data that was crucial to this research. Without access to these valuable resources, this work would not have been possible.

Special thanks to my peers for their collaboration and motivation. Finally, I am deeply thankful to my family and friends for their constant love, support, and encouragement throughout this journey.

Ushosree Raha

ABSTRACT

Accurate forecasting of carbon dioxide (CO_2) emissions is essential for supporting climate change mitigation efforts and formulating effective environmental policies. This study focuses on modeling and forecasting coal-based CO_2 emissions in the United States using both traditional statistical techniques and modern machine learning approaches. Initially, the AutoRegressive Integrated Moving Average (ARIMA) model is employed to capture the linear patterns and temporal dependencies in the historical emissions data. Recognizing that environmental time series often exhibit complex nonlinear behavior, the study introduces a hybrid modeling framework by integrating ARIMA with Neural Network AutoRegressive (NNAR) models. This hybrid ARIMA-NNAR approach aims to improve predictive accuracy by addressing both linear and nonlinear components of the data. Multiple ARIMA and NNAR configurations are evaluated using standard model selection criteria and diagnostic tests. The hybrid model combining ARIMA(3,1,3) with NNAR(12,6) demonstrates superior performance, as indicated by lower forecasting errors and well-behaved residuals. Forecasts generated for the period 2021 to 2040 show a clear declining trend in coal-based CO_2 emissions, aligning with recent policy shifts and the gradual transition toward cleaner energy sources. The findings highlight the advantages of hybrid time series models in capturing complex emission dynamics and provide valuable insights for energy planning and environmental decision-making.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The rapid pace of industrialization and urbanization over the past century has led to a significant increase in greenhouse gas emissions, with carbon dioxide (CO_2) being the most prevalent contributor to global warming. Among the various sources of CO_2, the combustion of coal for energy production has historically been a dominant driver, especially in industrialized countries such as the United States. Although recent decades have seen a global push towards cleaner and renewable energy sources, coal still remains a considerable component of the energy mix in many regions, warranting continuous monitoring and accurate forecasting of its associated emissions. Reliable forecasting of CO_2 emissions plays a pivotal role in shaping environmental policies, assessing compliance with climate targets, and designing sustainable energy strategies. Accurate models can help stakeholders anticipate future emission trends, evaluate the impact of mitigation efforts, and prepare for policy interventions. However, modeling environmental time series such as CO_2 emissions presents several challenges due to their inherent complexity, nonlinearity, and sensitivity to socio-economic and technological factors. Traditional statistical methods like the AutoRegressive Integrated Moving Average (ARIMA) model have long been employed for time series forecasting due to their strong theoretical foundation and effectiveness in capturing linear relationships and trends. Nonetheless, real-world environmental data often exhibit nonlinear patterns and sudden shifts that linear models struggle to accommodate. To enhance predictive accuracy, researchers have increasingly turned to hybrid modeling approaches that integrate linear models like ARIMA with nonlinear machine learning techniques such as Neural Network AutoRegressive (NNAR) models. The NNAR model, a specialized form of artificial neural networks tailored for time series data, offers the ability to learn and model complex, nonlinear dynamics. When used in combination with ARIMA, the resulting hybrid model benefits from both linear structure modeling and nonlinear adaptability, providing a more comprehensive tool for forecasting. This hybrid approach has been successfully applied in various domains and shows promising potential for improving CO_2 emissions forecasting. This thesis aims to develop and evaluate a hybrid ARIMA-NNAR modeling framework for forecasting coal-based CO_2 emissions in the United States. The objective is to assess whether combining these

two methods can yield better forecasting performance compared to traditional models alone. The study involves a rigorous statistical analysis of historical emissions data, the construction and evaluation of multiple ARIMA and NNAR models, and the implementation of a hybrid approach that sequentially models linear and nonlinear components of the data. By doing so, this research contributes to the growing field of hybrid time series modeling and provides practical insights for environmental planning and policy formulation. In the chapters that follow, a review of relevant literature will provide the theoretical and empirical context for the modeling approach. This will be followed by a detailed methodology outlining the steps involved in data processing, model selection, and performance evaluation. Finally, the results will be analyzed and discussed, culminating in conclusions and recommendations for future research and policy application.

Chapter 2

Literature Review

The modeling and forecasting of CO_2 emissions have become increasingly important in the context of climate change mitigation and sustainable energy planning. Among the various tools used for time series analysis, the AutoRegressive Integrated Moving Average (ARIMA) model has long been regarded as a fundamental approach. Box and Jenkins [1976] established the framework for ARIMA modeling, which remains widely used due to its capability to effectively capture linear temporal relationships and trends in time-dependent data. Despite its popularity, ARIMA has notable limitations in handling nonlinear structures, which are often inherent in environmental data due to the complex interplay of economic, technological, and policy-related factors. To address this gap, hybrid models combining ARIMA with machine learning techniques such as Artificial Neural Networks (ANN) have been developed. Zhang [2003] proposed one of the earliest hybrid ARIMA-ANN models and demonstrated that the hybrid approach significantly improves forecasting accuracy by modeling linear and nonlinear components separately. Expanding on this work, Khashei and Bijari [2011] introduced a generalized hybrid framework that integrated ARIMA with feed-forward ANN. Their model achieved superior performance across multiple empirical datasets by adapting to structural variations and nonlinear dynamics in the data. These hybrid models have been particularly effective in fields such as energy demand forecasting and climate modeling, where data behavior cannot be sufficiently explained by linear assumptions alone. In the context of carbon emissions forecasting, Liu et al. [2012] applied a hybrid ARIMA-ANN model to Chinese CO_2 emissions and found that the hybrid model outperformed both individual ARIMA and ANN models. Their results highlighted the importance of considering both linear and nonlinear elements in emission data for improved accuracy and policy relevance. More recently, Neural Network AutoRegressive (NNAR) models have emerged as a specialized type of ANN tailored for time series applications. NNAR models retain the autoregressive structure while incorporating the flexibility of neural networks to capture complex, nonlinear behaviors. Hyndman and Athanasopoulos [2018] emphasized the effectiveness of NNAR models in time series forecasting and recommended their use in hybrid frameworks where interpretability and computational efficiency are impor-

tant. In summary, the evolution from linear ARIMA models to hybrid ARIMA-ANN and NNAR approaches reflects a broader trend toward more adaptive and robust forecasting techniques. These models are especially valuable in environmental applications, where emission patterns often involve both persistent trends and nonlinear shocks. The combination of statistical rigor and machine learning flexibility allows hybrid models to provide more accurate and insightful forecasts for decision-making in climate and energy policy.

Chapter 3

Time Series Analysis

3.1 Preliminaries

A time series is a temporally ordered sequence of observations, where the timing of each data point x_t is integral to its interpretation and analysis. Unlike unordered datasets, time series inherently embody temporal dependence, making their sequential structure analytically indispensable. Discrete-time series, indexed by $T_0 \subseteq \mathbb{Z}$ or $\mathbb{N}$, represent observations captured at uniform intervals and are prevalent in domains such as economics, climatology, and operational research. In contrast, continuous-time series, with indices over real intervals $T_0 \subseteq \mathbb{R}$, characterize processes evolving continuously over time—common in physics, engineering, and quantitative finance. These models accommodate dynamic systems where changes can occur instantaneously, necessitating continuous monitoring. The classification of time series based on the time index is fundamental, as it dictates the analytical framework, modeling assumptions, and inferential methodologies suitable for capturing underlying temporal dynamics and dependencies.

3.2 Stationary Time Series

In time series analysis, a fundamental role is played by processes exhibiting time-invariant properties. Predictive modeling necessitates assumptions of structural constancy to ensure temporal coherence. In deterministic settings, extrapolation often presumes that the function or its derivatives remain unchanged over time—wherein a constant first derivative yields linear forecasting. However, time series data frequently incorporate stochastic components. When this randomness adheres to stationarity, it enables the application of rigorous statistical methods. Stationarity ensures that key characteristics—such as mean, variance, and autocovariance remain stable, facilitating robust and theoretically grounded forecasts of future values.

Definition: A stochastic process $\{X_t\}$ is weakly stationary if the following conditions are satisfied for all integers $t \in \mathbb{Z}$ and all lag values $h \in \mathbb{Z}$:

1. **Constant Mean:** The expected value of the process is independent of time:

$$\mathbb{E}[X_t] = \mu, \quad \text{for all } t.$$

2. **Constant Variance:** The variance of the process remains constant over time:

$$\text{Var}(X_t) = \mathbb{E}[(X_t - \mu)^2] = \sigma^2, \quad \text{for all } t.$$

3. **Autocovariance Depends Only on Lag:** The covariance between values at two different time points depends only on the distance (lag) between them, and not on the actual time points themselves:

$$\text{Cov}(X_t, X_{t+h}) = \gamma(h), \quad \text{for all } t \text{ and } h,$$

where $\gamma(h)$ is known as the autocovariance function at lag h.

3.3 Autocovariance and Auto-Correlation Function

Definition: Let $\{X_t\}_{t\in\mathbb{Z}}$ be a weakly stationary time series with constant mean $\mu = \mathbb{E}[X_t]$ and constant variance $\sigma^2 = \text{Var}(X_t)$. The autocovariance function at lag $h \in \mathbb{Z}$ is defined as:

$$\gamma_X(h) = \text{Cov}(X_{t+h}, X_t) = \mathbb{E}[(X_{t+h} - \mu)(X_t - \mu)].$$

Due to stationarity, $\gamma_X(h)$ depends only on the lag h, not on the specific time point t.
Definition: The autocorrelation function at lag h is the normalized autocovariance and is given by:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \frac{\mathbb{E}[(X_{t+h} - \mu)(X_t - \mu)]}{\mathbb{E}[(X_t - \mu)^2]}.$$

Since $\rho_X(h)$ is a correlation coefficient, it satisfies:

$$-1 \leq \rho_X(h) \leq 1, \quad \text{for all } h \in \mathbb{Z}.$$

These functions describe the linear dependence structure of the process across different time lags and are central to model identification and diagnostics in time series analysis.

3.4 Examples of Stationary Time Series Models

Example 1: First-Order Moving Average Process MA(1)

Let $\{X_t\}$ be a time series defined by the equation:

$$X_t = W_t + \psi Z_{t-1}, \quad t \in \mathbb{Z},$$

where $\{W_t\}$ denotes a sequence of white noise random variables with mean zero and constant variance, i.e., $W_t \sim \text{WN}(0, \sigma^2)$, and $\psi \in \mathbb{R}$ is a constant parameter.

Statistical Properties

- **Mean**:
$$\mathbb{E}[X_t] = \mathbb{E}[W_t] + \psi\mathbb{E}[Z_{t-1}] = 0.$$

- **Variance**:
$$\mathrm{Var}(X_t) = \mathbb{E}[(W_t + \psi Z_{t-1})^2] = \sigma^2(1+\psi^2).$$

- **Autocovariance Function (ACVF)**:
$$\gamma_X(h) = \begin{cases} \sigma^2(1+\psi^2), & \text{if } h = 0, \\ \sigma^2\psi, & \text{if } |h| = 1, \\ 0, & \text{if } |h| > 1. \end{cases}$$

- **Autocorrelation Function (ACF)**:
$$\rho_X(h) = \begin{cases} 1, & \text{if } h = 0, \\ \dfrac{\psi}{1+\psi^2}, & \text{if } |h| = 1, \\ 0, & \text{if } |h| > 1. \end{cases}$$

The MA(1) process satisfies the criteria for weak stationarity, as its mean, variance, and autocovariance are all invariant with respect to time.

Example 2: First-Order Autoregressive Process AR(1)

Now consider a time series $\{X_t\}$ defined by the recursive relationship:

$$X_t = \phi X_{t-1} + W_t, \quad t \in \mathbb{Z},$$

where $|\phi| < 1$ ensures stationarity, and $\{W_t\} \sim \mathrm{WN}(0, \sigma^2)$ is a white noise process. It is also assumed that W_t is uncorrelated with past values of $\{X_s\}$ for all $s < t$.

Statistical Properties

- **Mean**:
$$\mathbb{E}[X_t] = \phi\mathbb{E}[X_{t-1}] + \mathbb{E}[W_t] = 0.$$

- **Autocovariance Function (ACVF)**:
$$\gamma_X(h) = \phi\gamma_X(h-1), \quad h > 0.$$

  By recursion,
$$\gamma_X(h) = \phi^h\gamma_X(0), \quad \text{for } h \geq 0.$$

  Since $\gamma_X(h) = \gamma_X(-h)$, this extends to all $h \in \mathbb{Z}$.

- **Variance**: Using the equation:

$$\mathrm{Var}(X_t) = \mathrm{Var}(\phi X_{t-1} + W_t) = \phi^2 \gamma_X(0) + \sigma^2,$$

  we solve for $\gamma_X(0)$:

$$\gamma_X(0) = \frac{\sigma^2}{1-\phi^2}.$$

- **Autocorrelation Function (ACF)**:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \phi^{|h|}, \quad h \in \mathbb{Z}.$$

The AR(1) process exhibits exponentially decaying autocorrelations, and all its statistical properties are time-invariant, confirming that it is weakly stationary.

Chapter 4

Stationary Linear Process (ARIMA Model): Parameter Estimation, Diagnostics and Forecasting

4.1 Linear Process

Linear time series models, such as ARMA, offer a broad approach to analyzing stationary processes. Any second-order stationary process is either inherently linear or can be made linear by removing a deterministic component, enabling effective modeling and analysis within this framework.

Definition: The time series $\{X_t\}$ is a **linear process** if it has the representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j} \tag{4.1}$$

for all t, where $\{W_t\} \sim \text{WN}(0, \sigma^2)$ and $\{\psi_j\}$ is a sequence of constants with

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

in terms of backward shift operator we can re-write the equation (2.1) as

$$X_t = \psi(L) W_t \tag{4.2}$$

Where $\psi(L) = \sum_{j=-\infty}^{\infty} \psi_j L^j$

Proposition 2.2.1 Let $\{Y_t\}$ be a stationary time series with mean 0 and covariance function γ_Y. If $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, then the time series

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} = \psi(L) Y_t \tag{4.3}$$

is stationary with mean 0 and autocovariance function

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h+k-j). \tag{4.4}$$

In the special case where $\{X_t\}$ is the linear process (2.1)

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2. \tag{4.5}$$

4.2 Auto-Regressive Model or AR(p)

An autoregressive process of order p, denoted AR(p), is a stochastic process $\{X_t\}_{t\geq 1}$ characterized by a linear dependence on its p most recent past values. Formally, it is defined as:

$$X_t = W_t + \sum_{j=1}^{p} \phi_j X_{t-j}, \tag{4.6}$$

where $\{W_t\}_{t\geq 1} \sim \text{WN}(0, \sigma^2)$ is a white noise process, and ϕ_j are autoregressive coefficients. If we consider (2.6) in terms of linear process then we obtain,

4.3 Moving Average Model or MA (q)

A Moving Average (MA) process of order q, denoted MA(q), is a linear stochastic process in which the current value X_t is expressed as a finite linear combination of present and past innovations (random shocks). Unlike autoregressive models that regress on past values, the MA model captures temporal dependence through past forecast errors. Formally, the process $\{X_t\}_{t\geq 1}$ is defined as:

$$X_t = W_t + \sum_{j=1}^{q} \theta_j W_{t-j}, \tag{4.7}$$

where $\{W_t\}_{t\geq 1} \sim \text{WN}(0, \sigma^2)$ is a white noise process and θ_j are the moving average coefficients. This can be equivalently represented using the lag operator L as:

$$X_t = \mu(L) W_t, \quad \text{where } \mu(L) = 1 + \sum_{j=1}^{q} \theta_j L^j.$$

This formulation models transient dynamics and is particularly effective for capturing short-term shock effects. The order q indicates the number of lagged noise terms influencing (X_t).

4.4 Autoregressive integrated moving average or ARIMA (p, d, q)

Let $\{X_t\}_{t\geq 1}$ be a real-valued stochastic process, and let $d \in \mathbb{Z}_{\geq 0}$ be a non-negative integer. The process $\{X_t\}_{t\geq 1}$ is said to follow an autoregressive integrated moving average

(ARIMA) model of order (p, d, q) if:

1. The d-th differenced process $\nabla^d X_t = (1 - L)^d X_t$ is **weakly stationary**, where L is the lag operator defined as $LX_t = X_{t-1}$.

2. The differenced process $\{Y_t\} = \{\nabla^d X_t\}$ satisfies a **causal** and **invertible** ARMA(p, q) model:

$$\phi(L)Y_t = \theta(L)W_t$$

where:

- $Y_t = \nabla^d X_t = (1 - L)^d X_t$
- $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ (autoregressive polynomial)
- $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$ (moving average polynomial)
- $W_t \sim \text{WN}(0, \sigma^2)$ (white noise)

with the roots of $\phi(z) = 0$ and $\theta(z) = 0$ lying outside the unit circle to ensure causality and invertibility, respectively.

Thus, the complete ARIMA model is written compactly as:

$$\phi(L)(1 - L)^d X_t = \theta(L)W_t \tag{4.8}$$

The predictors on the right side of the equation are the lagged values and errors, and this is known as the ARIMA(p, d, q) model. These are the key parameters of ARIMA, and selecting the appropriate values for p, d, and q will yield better model results.

- p = order of the autoregressive (AR) part. This is the number of lag observations included in the model, i.e., the number of past terms of the series used as predictors.
- d = degree of differencing. This is the number of times the data have had past values subtracted (i.e., the number of differencing steps required to make the time series stationary).
- q = order of the moving average (MA) part. This is the number of lagged forecast errors in the prediction equation.

4.5 Methodological Implementation of ARIMA Model

The Box and Jenkins methodology (Box et al., 2013) includes three model building steps and a forecasting step as follows:

1. Identification : The Box–Jenkins methodology presupposes that the underlying time series is stationary, meaning its statistical properties—such as mean, variance,

and autocovariance—remain invariant over time. A stationary process, denoted typically as $\{Y_t\}$, satisfies:

$$\mathbb{E}[Y_t] = \mu, \quad \text{Var}(Y_t) = \sigma^2 < \infty, \quad \text{Cov}(Y_t, Y_{t+h}) = \gamma(h),$$

where μ and σ^2 are constants, and $\gamma(h)$ depends only on the lag h, not on time t. In cases where the observed series exhibits non-stationarity commonly due to trends or changing variance preprocessing steps such as logarithmic transformation and differencing were applied to induce stationarity. The order of differencing, denoted by d, corresponds to the minimum number of differences required to achieve stationarity, which is a key component in the ARIMA(p, d, q) framework. Subsequently, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the differenced and transformed series were analyzed to identify appropriate values for the autoregressive order p and moving average order q. The ACF provides insights into the dependence structure at various lags, while the PACF helps isolate the direct correlation between observations at specific lags, effectively guiding the model selection process.

2. Model Estimation: To determine the optimal autoregressive (p) and moving average (q) orders in the ARIMA(p, d, q) framework, a model selection procedure was employed based on the Residual Sum of Squares Error (SSE). A sequence of candidate models was evaluated, ranging from lower to higher orders of p and q. The significance of reductions in SSE between nested models was assessed using the F-test, in accordance with the following test statistic:

$$F = \frac{(SSE_1 - SSE_2)/(df_1 - df_2)}{SSE_2/df_2} \tag{4.9}$$

where SSE_1 and SSE_2 are the residual sum of squares of the simpler and more complex models, respectively, and df denotes the degrees of freedom. A model was retained only if the reduction in SSE was not statistically significant at the $\alpha = 0.05$ level, thereby balancing model fit with parsimony.Parameter estimation for the selected ARIMA model was performed using the maximum likelihood estimation method.

3. Model Diagnostics : To evaluate the adequacy and goodness-of-fit of the candidate ARIMA models, a suite of diagnostic tools and statistical tests was employed. If a model failed to meet diagnostic criteria, it was discarded in favor of an alternative specification, initiating a new iteration of the parameter estimation and validation cycle. In this study, both graphical and statistical diagnostic methods were utilized. The Ljung–Box Q-test (Ljung & Box, 1978) was applied to the residuals to assess whether they constituted a white noise process—i.e., a sequence with zero mean,

constant variance, and no autocorrelation. The null hypothesis of the Ljung–Box test posits that residual autocorrelations up to a specified lag are jointly zero:

$$H_0 : \rho_1 = \rho_2 = \cdots = \rho_h = 0.$$

The statistical significance of the estimated model parameters was evaluated using the Student's t-test, allowing for the identification of parameters that contribute meaningfully to model structure. In addition, parameter correlation matrices were examined to identify potential multicollinearity, which can compromise model interpretability and stability. The distributional properties of the residuals were assessed using histograms and normal probability plots, to verify approximate normality—a key assumption in many inferential procedures. To further assess model adequacy, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the residuals were analyzed to detect any remaining serial dependence, which would indicate model misspecification. The presence of first-order autocorrelation in residuals was formally tested using the Durbin–Watson (DW) statistic (Durbin & Watson, 1950), where values substantially different from 2 suggest significant positive or negative autocorrelation. This iterative model-building procedure—comprising model identification, parameter estimation, and diagnostic validation—was repeated until a model was found that satisfied all adequacy criteria. The final selected model was then employed for forecasting purposes.
In the succeding sections we will discuss the mathematical structure of parameter estimation, diagnostic methods and Forecasting.

4.6 Parameter Estimation via the Maximum Likelihood Method

Let $\{X_t\}$ be a zero-mean Gaussian time series with autocovariance function $\kappa(i,j) = \mathbb{E}[X_i X_j]$, and define the observation vector $\mathbf{X}_n = (X_1, \ldots, X_n)'$. Let $\hat{\mathbf{X}}_n = (\hat{X}_1, \ldots, \hat{X}_n)'$, where $\hat{X}_1 = 0$ and for $j \geq 2$, $\hat{X}_j = \mathbb{E}[X_j | X_1, \ldots, X_{j-1}]$. Assume the covariance matrix $\Gamma_n = \mathbb{E}[\mathbf{X}_n \mathbf{X}_n']$ is nonsingular.

The likelihood function of $\mathbf{X}_n$ is given by:

$$L(\Gamma_n) = (2\pi)^{-n/2} |\Gamma_n|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{X}_n^\top \Gamma_n^{-1} \mathbf{X}_n\right). \tag{4.10}$$

Using the innovations algorithm, the prediction errors $e_j = X_j - \hat{X}_j$ and their variances ν_{j-1} can be computed recursively. The innovation vector $\mathbf{X}_n - \hat{\mathbf{X}}_n$ has diagonal covariance matrix $D_n = \mathrm{diag}(\nu_0, \ldots, \nu_{n-1})$, and we have:

$$\Gamma_n = C_n D_n C_n^\top, \tag{4.11}$$

where C_n is a lower-triangular matrix. This implies:

$$\mathbf{X}_n^\top \Gamma_n^{-1} \mathbf{X}_n = \sum_{j=1}^{n} \frac{e_j^2}{\nu_{j-1}}, \quad |\Gamma_n| = \prod_{j=0}^{n-1} \nu_j. \tag{4.12}$$

Thus, the likelihood simplifies to:

$$L(\Gamma_n) = \frac{1}{\sqrt{(2\pi)^n \prod_{j=0}^{n-1} \nu_j}} \exp\left(-\frac{1}{2} \sum_{j=1}^{n} \frac{e_j^2}{\nu_{j-1}} \right). \tag{4.13}$$

For an ARMA(p, q) process, the predictors and innovation variances are computed as:

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj} e_{n+1-j}, & n < m, \\ \sum_{i=1}^{p} \phi_i X_{n+1-i} + \sum_{j=1}^{q} \theta_{nj} e_{n+1-j}, & n \geq m, \end{cases} \tag{4.14}$$

$$\nu_n = \sigma^2 r_n, \quad \text{where } m = \max(p, q). \tag{4.15}$$

Substituting into the likelihood yields:

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n \prod_{j=0}^{n-1} r_j}} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^{n} \frac{e_j^2}{r_{j-1}} \right). \tag{4.16}$$

Maximizing the log-likelihood with respect to σ^2 gives:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} \frac{e_j^2}{r_{j-1}}. \tag{4.17}$$

Finally, the parameter estimates $\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}$ are obtained by minimizing:

$$\ell(\boldsymbol{\phi}, \boldsymbol{\theta}) = \ln\left(\frac{1}{n} \sum_{j=1}^{n} \frac{e_j^2}{r_{j-1}} \right) + \frac{1}{n} \sum_{j=1}^{n} \ln r_{j-1}. \tag{4.18}$$

4.7 KPSS (Kwiatkowski Philips Schimdt Shin) Test:

This is a test for stationary which check especially the existence of trend in data. The KPSS test is based on linear regression. It breaks up a series into three parts: a deterministic trend (β_t), a random walk (r_t), and a stationary error (ϵ_t), with the regression equation:

$$X_t = \beta_0 + \beta_1 t + \theta_t + \varepsilon_t \tag{4.19}$$

$$\theta_t = \theta_{t-1} + \varepsilon_t \tag{4.20}$$

$$\varepsilon_t \sim WN(0,\ \sigma_\varepsilon^2) \tag{4.21}$$

If the data is stationary, it will have a fixed element for an intercept or the series will be stationary around a fixed level (Wang, p.33). The test uses OLS find the equation, which differs slightly depending on whether you want to test for level stationarity or trend stationarity (Kocenda & Cerný). A simplified version, without the time trend component, is used to test level stationarity.
Null Hypothesis H_0: Data is stationary i.e., $\sigma_\varepsilon^2 = 0$
Alternative Hypothesis H_1: Data is not Stationary i.e., $\sigma_\varepsilon^2 > 0$
Under H_0;

$$KPSS = T^{-2}\frac{\sum_{t=1}^{T}\hat{S}_t^2}{\hat{\beta}^2} \tag{4.22}$$

Where $\hat{S}_t^2 = \sum_{i=1}^{t}\hat{\varepsilon}_i$, $\hat{\varepsilon}_i$ is the residual of regression X_t on t. In terms of p-values, larger p-value suggests that the data is stationary

4.8 Ljung-Box Test:

The Ljung-Box approach test the null hypothesis that a number of auto-correlation coefficients are simultaneously zero or, more colloquially, it evaluates that whether there is any significant auto-correlation in a series.
Null hypothesis H_0: There is no auto-correlation up-to order k
Alternative Hypothesis H_1: There is auto-correlation up-to order k
Under H_0,

$$Q = N(N+2)\sum_{j=1}^{k}\frac{r_j^2}{N-j} \tag{4.23}$$

Where r_j is the j^{th} autocorrelation and N is the number of observations. k is the no. of lags being tested. Large p-value indicate there is no auto-correlation up-to order k.

4.9 Model Order Selection

we introduced minimization of the AICC value as a major criterion for the selection of the orders p and q.This criterion is applied as follows: Choose p, q, $\boldsymbol{\phi}_p$, and $\boldsymbol{\theta}_q$ to minimize

$$\text{AICC} = -2\ln L(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q, S(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)/n) + 2(p+q+1)n/(n-p-q-2).$$

For any fixed p and q it is clear that the AICC is minimized when $\boldsymbol{\phi}_p$ and $\boldsymbol{\theta}_q$ are the vectors that minimize

$$-2\ln L(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q, S(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)/n),$$

i.e., the maximum likelihood estimators. Final decisions with respect to order selection should therefore be made on the basis of maximum likelihood estimators. The AICC statistic and its justification are discussed in detail.

Chapter 5

Neural Network Based Time Series Model

5.1 Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) models the relationship between a set of input signals and output signal using a model derived from our understanding of how a biological brain responds to stimuli from sensory inputs. Just as a brain uses a network of inter-connected cells called neurons to create a massive parallel processor, MLP uses a network of artificial neurons or nodes to solve a learning problem.

Definition: A neural network is a nonlinear mapping $f : \mathbb{R}^n \to \mathbb{R}^m$, where m, n are the dimensions of the input and output spaces.

$$y = \varphi\left(\sum_{j=1}^{n} w_j x_j + b\right) \tag{5.1}$$

where $\varphi : \mathbb{R} \to [a, b]$ is an **activation function**, i.e.,

$$\lim_{x\to-\infty} \varphi(x) = a, \quad \text{and} \quad \lim_{x\to\infty} \varphi(x) = b. \tag{5.2}$$

Activation Function: The activation function is the mechanism by which the artificial neuron processes incoming information and passes it throughout the network. Intuitively, it is a function mapping from $\mathbb{R}$ to $[a, b]$, i.e.,

$$\varphi : \mathbb{R} \to [a, b] \tag{5.3}$$

such that

$$\lim_{x\to-\infty} \varphi(x) = a \quad \text{and} \quad \lim_{x\to\infty} \varphi(x) = b. \tag{5.4}$$

There are different kinds of activation functions.

5.2 Feed-forward Neural Network

Definition: A feed-forward neural network is a function

$$y = \sum_{j=1}^{n} w_j \varphi(a_j x + b_j) \tag{5.5}$$

A more general case is that of a network handling multiple inputs. The inputs are linked to the hidden layer through different synaptic weights.
Definition A feed-forward neural network with multiple inputs and a single output is a function

$$y = \sum_{j=1}^{n} w_j \varphi \left(\sum_{i=1}^{m} a_{ij} x_i + b_j \right) \tag{5.6}$$

where $\varphi : \mathbb{R} \to [a, b]$ is an **activation function**.

5.3 Neural Network Autoregression (NNAR)

The concept of neural network used in time series forecasting purpose. NNAR model is a non-linear time series model, which uses significant lagged values are used as an input to the neural network, just as we used in autoregression models. The number of inputs or lagged values are determined from the PACF of the series. We use that lagged values, at which, there exist significant autocorrelations. We call this model as a Neural Network Autoregression or NNAR model. Feed forward network is the most widely used neural network model for time series modeling and forecasting (Zhang et al., 1998). The NNAR model is a three layers feedforward neural network which involves a linear combination function and an activation function. The relationship between the model output (y_t) and the inputs $(y_{t-1}, ..., y_{t-p})$ has the following mathematical representation:

$$y_t = w_0 + \sum_{j=1}^{k} w_j \varphi(w_{0j} + \sum_{i=1}^{p} w_{ij}\ y_{t-i}) +\ \varepsilon_t \tag{5.7}$$

This our $NNAR(p, k)$ models with p no. of lagged inputs and k no. of hidden nodes. where, w_{ij} $(i = 0, 1, 2, \ldots, p, j = 1, 2, \ldots, k)$ and $w_j = (j = 0, 1, 2, \ldots, k)$ are model parameters or connection weights; p is number of input nodes; and k is number of hidden nodes. A sigmoid function was used as the hidden layer transfer function that is shown in Eq.

$$\varphi(x) = \frac{1}{1 + e^{-x}} \tag{5.8}$$

It is a non-linear function used in the hidden layer to modify the inputs and prepared as an input for the next layer. This tends to reduce the effect of extreme input values, thus making the networks somewhere robust to the outliers. Note that, no. of hidden layers

and no. of hidden nodes must be specified in advance. If the value of k is not specified then, it is determined as, $k = \frac{p+P+1}{2}$.

5.4 Parameter Estimation:

Let we consider the above model where $w_0, w_1, .., w_k, w_{0j}, w_{ij}$ are parameters of the model. The objective function for this problem

$$Q(\mathbf{w}) = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 \tag{5.9}$$

where

- y_t is the observed value at time t,
- $\hat{y}_t = w_0 + \sum_{j=1}^{k} w_j \varphi \left(w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i} \right)$ is the predicted value at time t.

We can also write the $Q(\mathbf{w})$ as

$$Q(\mathbf{w}) = \sum_{t=1}^{T} \left(y_t - \left(w_0 + \sum_{j=1}^{k} w_j \varphi \left(w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i} \right) \right) \right)^2 \tag{5.10}$$

Now we have to find

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} Q(\mathbf{w}) \tag{5.11}$$

The gradient of $Q(\mathbf{w})$ with respect to each parameter w_j

$$Q(\mathbf{w}) = -2 \sum_{t=1}^{T} (y_t - \hat{y}_t) \cdot \frac{\partial \hat{y}_t}{\partial w_j} \tag{5.12}$$

For w_0:

$$\frac{\partial \hat{y}_t}{\partial w_0} = 1 + \sum_{j=1}^{k} w_j \cdot \frac{\partial}{\partial w_0} \varphi \left(w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i} \right) \tag{5.13}$$

For w_j:

$$\frac{\partial \hat{y}_t}{\partial w_j} = \varphi' \left(w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i} \right) \cdot \frac{\partial}{\partial w_j} \left(w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i} \right) \tag{5.14}$$

The derivative of the inner term $w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i}$ with respect to w_j is simply y_{t-j}, the lagged value at time $t - j$. So we get:

$$\frac{\partial \hat{y}_t}{\partial w_j} = \varphi' \left(w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i} \right) \cdot y_{t-j} \tag{5.15}$$

Now, substitute this into the gradient expression:

$$\frac{\partial Q(\mathbf{w})}{\partial w_j} = -2\sum_{t=1}^{T}(y_t - \hat{y}_t)\cdot\varphi'\left(w_{0j} + \sum_{i=1}^{p} w_{ij}y_{t-i}\right)\cdot y_{t-j} \tag{5.16}$$

5.5 Interpretation of Weights

The learned weights convey the importance and directionality of influences:

- **Positive weights** indicate a direct effect.
- **Negative weights** indicate an inverse effect.
- **Larger magnitudes** represent stronger influence.

5.6 Gradient Descent Optimization:

The gradient descent update rule is used to update the parameters iteratively. The general update rule for each parameter w_j is:

$$w_j(t+1) = w_j(t) - \eta\cdot\frac{\partial Q(\mathbf{w})}{\partial w_j} \tag{5.17}$$

Where η is learning rate or step size. For each parameter w_j, we update it according to:

$$w_j(t+1) = w_j(t) - \eta\cdot\left[-2\sum_{t=1}^{T}(y_t - \hat{y}_t)\cdot\varphi'\left(w_{0j} + \sum_{i=1}^{p} w_{ij}y_{t-i}\right)\cdot y_{t-j}\right] \tag{5.18}$$

Simplifying:

$$w_j(t+1) = w_j(t) + 2\eta\sum_{t=1}^{T}(y_t - \hat{y}_t)\cdot\varphi'\left(w_{0j} + \sum_{i=1}^{p} w_{ij}y_{t-i}\right)\cdot y_{t-j} \tag{5.19}$$

5.7 Algorithm Summary:

The Gradient Descent Algorithm for this problem can be summarized as:

1. **Initialize the parameters** $w_0, w_1, \ldots, w_k, w_{0j}, w_{ij}$.
2. **For each iteration:**
   - Compute the predictions $\hat{y}_t$ for all time points using the current values of the parameters.
   - Calculate the residuals $y_t - \hat{y}_t$.
   - Compute the gradients $\frac{\partial Q(\mathbf{w})}{\partial w_j}$ for all parameters using the formula derived above.

- Update the parameters using the gradient descent update rule:

$$w_j(t+1) = w_j(t) + 2\eta \sum_{t=1}^{T} (y_t - \hat{y}_t) \cdot \varphi' \left(w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i} \right) \cdot y_{t-j}$$

3. **Repeat until the parameters converge** (when the change in the objective function $Q(w)$ becomes very small).

5.8 Methodological Implementation

NNAR models are identified by using the "nnetar" function included in the package "forecast". For non-seasonal data, we have describe NNAR models with the notation $NNAR(p, k)$, where p denotes the number of non-seasonal lags used as inputs, and k means the number of nodes/neurons in the hidden layer. $NNAR(p, k)$ is the same as AR process, but with nonlinear functions. The optimal number of non-seasonal lags is obtained by using the AIC metric, and the optimal number of neurons is identified by calculating $\frac{p+P+1}{2}$, where p is the non-seasonal AR order, and P is the seasonal AR order (if any). Finally, the goodness of fit is investigated using MAE, MAPE, MASE, and RMSE metrics

5.9 Hybrid Model Formulation

The Autoregressive Integrated Moving Average (ARIMA) model is a classical statistical tool extensively utilized for analyzing and forecasting univariate time series data. It is particularly effective in modeling and capturing the linear structures inherent in time-dependent datasets. However, real-world time series data often exhibit complex patterns that extend beyond linearity, necessitating more flexible modeling techniques.
To address this, the Neural Network Autoregression (NNAR) model—a nonlinear, data-driven approach—has been increasingly employed. NNAR models are adept at capturing nonlinear patterns and trends that traditional linear models may overlook. Recognizing the complementary strengths of these two methodologies, a hybrid modeling framework has been proposed to enhance forecasting accuracy by jointly modeling both linear and nonlinear components of the time series. The central premise of the hybrid approach lies in the assumption that a time series can be decomposed into linear and nonlinear components, which are additive in nature. This additive decomposition enables the combination of ARIMA and NNAR models to capture a more comprehensive representation of the underlying data dynamics. The hybrid model is formally defined as:

$$Z_t = X_t + N_t \tag{5.20}$$

Where Z_t denotes the observed time series at time t, X_t represents the linear component modeled by ARIMA, and N_t represents the nonlinear component captured by the NNAR model. Forecasting within this hybrid framework involves two stages. First, the linear

structure $\hat{X}_t$ is estimated using the ARIMA model. The residuals from this step, which potentially contain the nonlinear patterns, are then modeled using the NNAR network to yield $\hat{N}_t$. The final hybrid forecast $\hat{Z}_t$ is obtained by summing these two components:

$$\hat{Z}_t = \hat{X}_t + \hat{N}_t \tag{5.21}$$

This hybridization strategy effectively leverages the strengths of both ARIMA and NNAR models. ARIMA ensures robust linear trend extraction, while NNAR contributes to modeling nonlinear dependencies and residual structures. Empirical studies in the literature support the efficacy of such hybrid approaches, often demonstrating superior forecasting performance over individual models. The hybrid model thus serves as a robust forecasting tool, particularly suited for complex, real-world time series such as per capita CO_2 emissions in the United States.

5.10 Model Diagnostics

1. Residual Time Plot: If the residuals of the series are white noise, then, time plot of standardized residuals doesn't show cluster of volatility

2. Auto-correlation Function: We expect that, residuals obtained from the fitted model should be behave like white noise. If this is not done, then we have to sure that there may be possible extract much information from the series. On the other hand, if residuals are behaved like a white noise, then we have extracted all the information and left only "noise".

3. Akaike Information Criteria: This measures the relative quality of a statistical model for a given set of data. It is a trade-off between bias and variance of the model. During comparisons of multiple models, the model with less AIC-value is preferred over high AIC-value

$$AIC = 2(pk + 2k + 1) + n\ln(MSE) \tag{5.22}$$

5.11 Forecast Performance Metrics

The accuracy of a forecasting model is determined by how closely its predicted values match the actual observed data in the test set. To evaluate and compare the performance of the ARIMA, NNAR,ANN and ARIMA-NNAR models, three consistency metrics are commonly used: Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). In general, lower values of MSE, MAE, and MAPE indicate a more accurate forecasting model. Therefore, the model with the lowest values

across these metrics is considered the most optimal for prediction purposes.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{e_t}{y_t}\right| \times 100 \tag{5.23}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_t^2} \tag{5.24}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |e_t| \tag{5.25}$$

Chapter 6

Case Studies and Model Building

6.1 Data Descriptions and Collection Methodology

The Global Carbon Project (GCP), established in 2001 under the Future Earth initiative, aims to comprehensively understand the global carbon cycle, integrating both natural and anthropogenic components (GCP, n.d.). A critical element of the GCP's work is the development of the Global Carbon Budget, in which fossil CO_2 emissions play a key role. Initially, the GCP sourced its fossil CO_2 emissions data from the Carbon Dioxide Information Analysis Center (CDIAC) at Oak Ridge National Laboratory (e.g., Raupach et al., 2007). In recent years, the dataset's production has transitioned to Appalachian State University under the new name CDIAC-FF (Gilfillan & Marland, 2021). CDIAC's records, extending back to 1751, have been foundational in global emissions research. These estimates include emissions from solid, liquid, and gaseous fuels, flaring, and cement production, based primarily on energy consumption data from the United Nations and supplemented with statistics from the BP Statistical Review of World Energy and the U.S. Geological Survey (USGS).

The methodology employed by CDIAC significantly influenced the Tier 1 approach of the first IPCC Guidelines (Haukås et al., 1997). Over time, as the GCP received inquiries about discrepancies or anomalies in emissions data, the dataset has been refined for improved accuracy. For instance, since 1973, the GCP has incorporated emissions estimates from the U.S. Energy Information Administration (EIA), adjusting the distribution of emissions among coal, oil, and gas without altering total values to align with official U.S. reports submitted to the UNFCCC post-1990 (EIA, 2021). Further refinements include the addition of U.S. lime production emissions before 1990—absent in CDIAC's dataset—using USGS production data and applying a standard emission factor of 0.75 tonnes CO_2 per tonne of lime, with a 2.2% capture rate based on 1990 EPA data (USGS, 2017; EPA, 2021). This adjustment partially addresses the discontinuity in the "other" emissions category starting in 1990. These continual updates underscore the GCP's commitment to accuracy and transparency in global carbon emissions accounting, ensuring its dataset remains a vital resource for climate science and policy.

6.2 Annual CO_2 Emission in the USA (1860–2020) from Coal-Based Fuel

The graph titled "Annual CO_2 Emission in USA 1860–2020 From Coal-Based Fuel" illustrates the long-term trend of carbon dioxide emissions specifically from coal consumption in the United States over a 160-year period. The y-axis represents CO_2 emissions (likely in million metric tons), while the x-axis spans from the year 1860 to 2020. From the onset, the data indicate a steady increase in emissions from 1860 until around 1910. This initial rise corresponds with the post-industrial revolution era in the United States, where the rapid expansion of railroads, manufacturing, and coal-powered industries significantly boosted the demand for coal. The period from 1910 to approximately 1945 shows substantial volatility with noticeable peaks and troughs, likely influenced by economic fluctuations, World War I, the Great Depression, and World War II, all of which would have impacted industrial output and energy consumption. Between the late 1940s and early 1970s, emissions fluctuated moderately but remained within a relatively stable range. This could reflect post-war industrial expansion combined with the diversification of energy sources that began during this period. However, from the mid-1970s onward, a sharp and consistent upward trend is observed, peaking around the mid-2000s. This peak likely aligns with a period of economic growth and high coal consumption in electricity generation before the significant introduction of cleaner and alternative energy sources. Post-2008, the graph exhibits a marked and continuous decline in coal-related CO_2 emissions, falling sharply through 2020. This steep downward trend corresponds with a shift in energy policy, the closure of coal-fired power plants, increased environmental awareness, and the rapid growth of renewable energy and natural gas as alternatives. The drop may also reflect the impact of the 2008 financial crisis followed by sustained efforts in emissions reduction and climate policy. Overall, the graph reflects a classic industrial rise and fall in emissions, with the latter decades showing a significant pivot toward more sustainable energy practices. The data underscores how socioeconomic events, policy shifts, and technological advancements can significantly impact emission trajectories over time.

6.3 Descriptive Statistical Suummary

The coal consumption data, based on 161 valid observations, shows a mean value of 1087.82 units and a median of 1143.13 units. The closeness of these two measures suggests that the distribution is approximately symmetrical. The standard deviation is 612.00, indicating moderate variability in the dataset, while the coefficient of variation (0.56) reflects a reasonable level of relative dispersion. The minimum and maximum values range widely from 44.82 to 2214.84, and the interquartile range (IQR) of 851.39 shows that the central 50% of the data lies between 626.54 and 1477.93. The median absolute deviation (MAD), a robust measure of spread, is 506.07. Skewness is slightly negative at -0.12, and with a standard error of 0.19, this suggests that the distribution is nearly

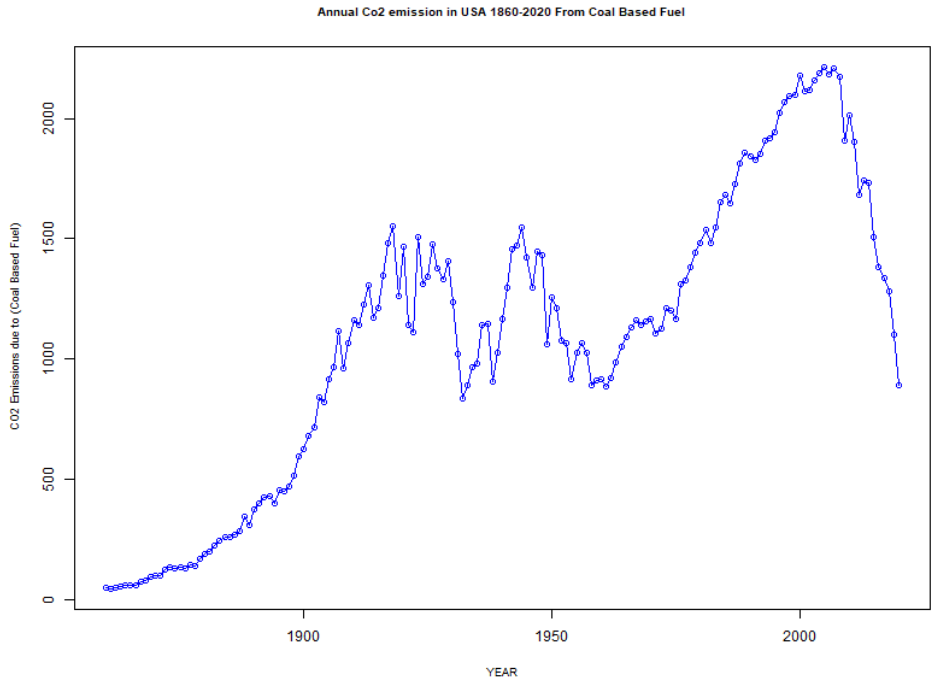


Figure 6.1: Time Series Plot of Annual CO_2 emission from Coal Based Fuel in USA

symmetric. The kurtosis value of -0.84 indicates a platykurtic distribution, meaning the data are slightly flatter and have lighter tails compared to a normal distribution. Overall, the coal consumption data appears to be fairly symmetrical with moderate dispersion and no significant outliers or extreme skewness, making it appropriate for further statistical analysis.

6.4 Temporal Dynamics of Coal-Based CO_2 Emissions: Insights from ACF and PACF Plots

The ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots for coal-based CO_2 emissions in the United States provide a detailed view of the time-dependent structure of the data. These plots are essential tools for understanding the persistence and lag relationships within the time series. The ACF plot, shown on the left, reveals a gradual and smooth decline in autocorrelation values as the lag increases. The autocorrelations remain significantly positive well beyond the initial lags, indicating a strong and persistent temporal dependence. This slow decay suggests that the emission levels in one year are strongly influenced by the levels in preceding years. Such behavior is typical of a series with a trend component, where values are not fluctuating around a constant mean but instead exhibit a prolonged pattern of increase or decrease over time.On the right, the PACF plot presents a sharp drop after lag 1, with only the first lag showing a strong and statistically significant partial autocorrelation. Beyond this, the values remain within the 95% confidence bounds, indicating minimal incremental influence from further lags once the first lag is accounted for. This implies that while the first

lag contributes substantially to explaining the variation in the current value, subsequent lags offer little additional explanatory power.

Together, these diagnostic plots suggest that the coal-based CO_2 emission data is characterized by strong year-to-year continuity and a dominant influence from recent historical values. This pattern reflects the inertia commonly observed in environmental and energy-related time series, where policy changes, technological shifts, and economic conditions typically influence trends gradually rather than abruptly. The information gained from these plots helps in understanding the temporal dynamics of emissions and supports informed interpretation of historical patterns in coal-based carbon output.

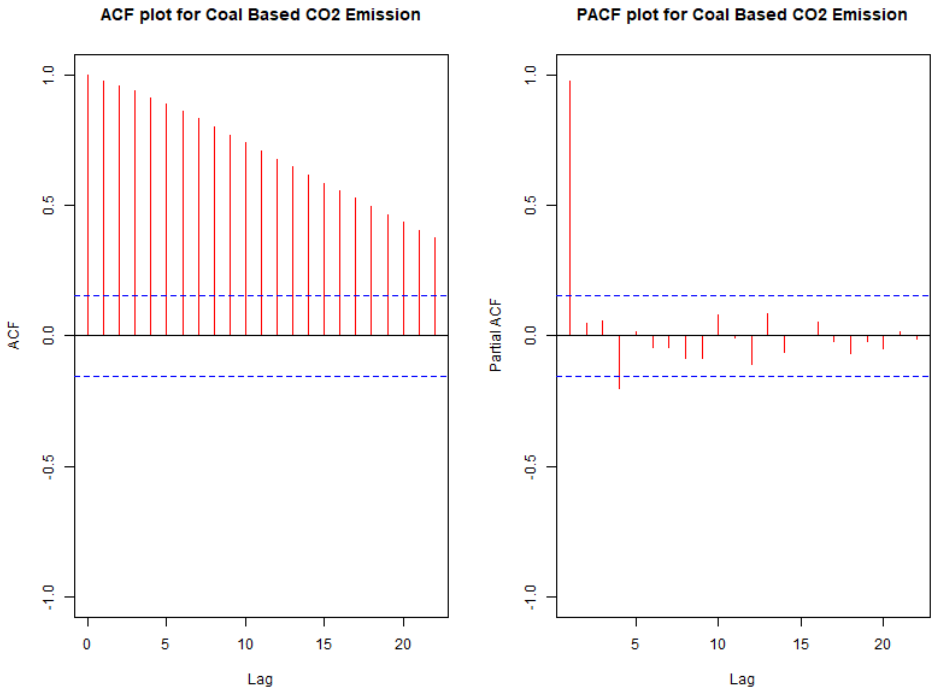


Figure 6.2: ACF and PACF Plot of Annual CO_2 emission from Coal Based Fuel in USA

6.5 Differencing and Stationarity of Coal-Based CO_2 Emissions: ACF, PACF, and Transformations

The second set of autocorrelation (ACF) and partial autocorrelation (PACF) plots, along with the differenced time series graphs, provide crucial insights into the stationarity and dynamic behavior of coal-based CO_2 emissions in the United States. These analyses are particularly focused on the transformed series resulting from differencing operations, which are used to stabilize the mean and remove trends from non-stationary time series data. The ACF and PACF plots presented at the top reveal a significant transformation in the time series' structure after differencing. The ACF plot shows a sharp drop after the first lag, with subsequent lags fluctuating within the 95% confidence interval, suggesting that the long-term autocorrelation observed in the original data has been largely removed. This rapid decay is indicative of a stationary series, where temporal depen-

dencies are no longer persistent across time. Similarly, the PACF plot demonstrates a few mild spikes early on, but these are not strongly significant and decline quickly. This behavior reinforces the conclusion that the differencing process has effectively eliminated most of the autocorrelation structure, producing a more stable series suitable for further time series analysis.

Supporting this interpretation, the second image compares the first-order and second-order differenced versions of the emission series. The left panel, which shows the first-order differenced data, displays values that oscillate around a mean of approximately zero, although with considerable variability and occasional extreme fluctuations. These characteristics are typical of a series that has had its trend removed but may still contain some volatility or local structure. The right panel, which depicts the second-order differenced series, exhibits even more pronounced fluctuations, suggesting increased variability and noise. The higher volatility in the second-order differenced data indicates that over-differencing may introduce unnecessary complexity and reduce interpretability, as the underlying signal may become obscured by amplified random variation. Together, these plots demonstrate the effectiveness of differencing in transforming a non-stationary series of coal-related CO_2 emissions into a stationary one. The stationarity achieved through first-order differencing is supported by the ACF and PACF behavior, which exhibit quick decay and minimal significant lags. This transformation is essential for enabling valid statistical modeling and forecasting, as most time series models rely on the assumption of stationarity. These results underscore the importance of differencing and diagnostic plotting in preparing time series data for rigorous analysis, particularly in environmental and energy-related domains where historical data often reflect long-term structural trends.

6.6 Testing of Hyopthesis on Stationarity of Series - KPSS Test

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was conducted to assess the trend stationarity of the time series variable after its first and second differencing. The KPSS test evaluates the null hypothesis that a time series is trend stationary, meaning any non-stationarity in the data can be attributed solely to a deterministic trend, rather than stochastic processes. For the first order differenced series, the KPSS test yielded a test statistic of 0.12412 with a truncation lag parameter set at 4. The associated p-value was reported as 0.09052. Importantly, the warning p-value greater than printed p-value suggests that the actual p-value might be slightly above the reported value but remains within the commonly accepted range for failing to reject the null hypothesis. Since the p-value is greater than the conventional 5% significance level, there is insufficient evidence to reject the null hypothesis of trend stationarity. This result implies that the first order differenced series appears to be stationary around a deterministic trend. In the second order differenced series, the KPSS statistic further decreased to 0.021522, with the same truncation lag parameter of 4. The reported p-value in this case is 0.1, which again exceeds

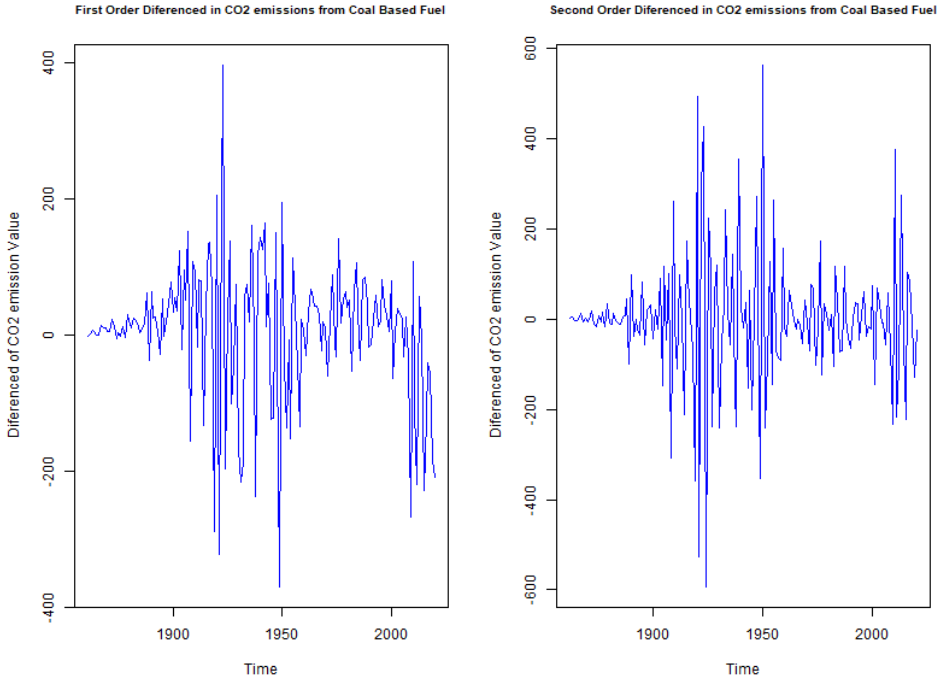


Figure 6.3: Sationary Time Series Plot of Annual CO_2 emission from Coal Based Fuel in USA

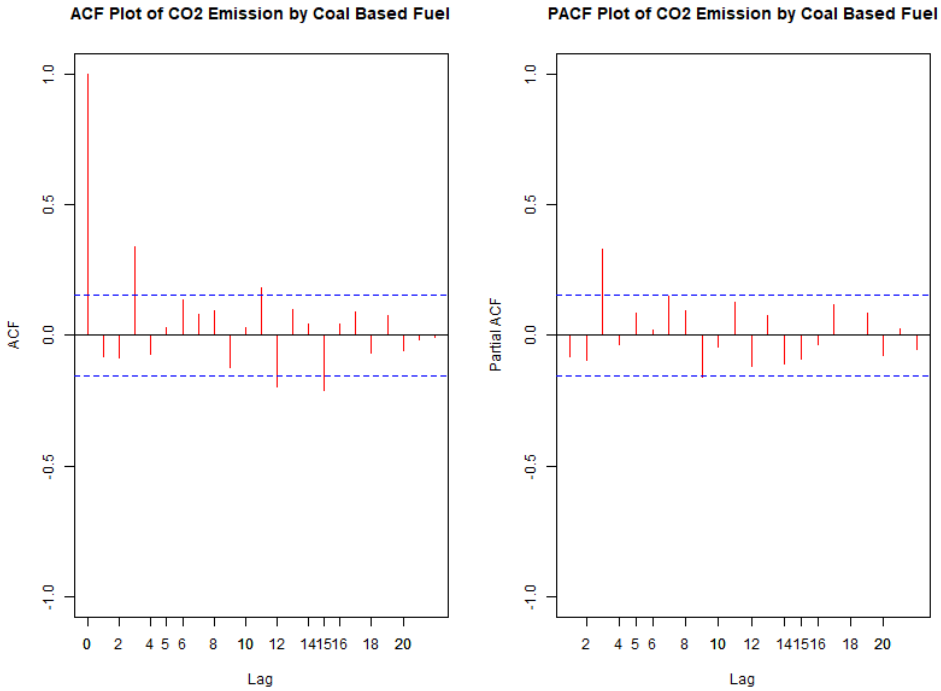


Figure 6.4: ACF and PACF Plot of Annual CO_2 emission from Coal Based Fuel in USA

the usual significance thresholds (e.g., 1%, 5%, or 10%). This stronger indication supports the null hypothesis more robustly, suggesting that the second differencing may have been excessive, and the series likely achieved stationarity with just the first differencing. In summary, the KPSS test results suggest that the time series becomes trend stationary upon first differencing, as evidenced by the low KPSS statistic and relatively high p-value. Further differencing does not appear necessary and may lead to over-differencing, which can distort the structure of the original data. Therefore, for modeling purposes such as ARIMA, the first differencing of the series would likely suffice to achieve stationarity.

6.7 Model Selection and Evaluation for Coal-Based CO_2 Emissions: ARIMA Model Comparison

Table 6.1: Comparison of ARIMA Model Coefficients, AIC, Estimated Variance, and Ljung-Box Test Results

Parameter	ARIMA(3,1,1)	ARIMA(3,1,2)	ARIMA(3,1,3)	ARIMA(3,1,11)	ARIMA(3,1,12)	ARIMA(7,1,1)	ARIMA(7,1,2)	ARIMA(7,1,3)
ar1	-0.1533 (0.3015)	0.0802 (0.1669)	0.0407 (0.1924)	-0.6646 (0.4573)	-0.3293 (0.7357)	0.1783 (0.2579)	0.3405 (0.3115)	0.3697 (1.3778)
ar2	-0.0626 (0.0819)	0.3081 (0.1655)	0.2908 (0.1591)	-0.1921 (0.5445)	0.1707 (0.8046)	-0.0854 (0.0809)	-0.6676 (0.2174)	-0.6909 (1.7610)
ar3	0.3319 (0.0855)	0.3802 (0.0757)	0.4450 (0.1762)	0.3998 (0.4642)	0.6257 (0.5510)	0.3651 (0.0819)	0.3550 (0.0946)	0.2501 (3.4122)
ar4	—	—	—	—	—	-0.1420 (0.1151)	-0.2527 (0.1339)	-0.2614 (0.4015)
ar5	—	—	—	—	—	0.1187 (0.0796)	0.3258 (0.1198)	0.3279 (0.3911)
ar6	—	—	—	—	—	0.0292 (0.0844)	0.0079 (0.0847)	0.0430 (0.9129)
ar7	—	—	—	—	—	0.1778 (0.0821)	0.2452 (0.0955)	0.2202 (1.1619)
ma1	0.1135 (0.3289)	-0.1410 (0.1767)	-0.0961 (0.2139)	0.6387 (0.4572)	0.3052 (0.7360)	-0.2399 (0.2550)	-0.4114 (0.3205)	-0.4433 (1.4602)
ma2	—	-0.4094 (0.1701)	-0.3860 (0.1702)	0.0870 (0.5280)	-0.2631 (0.7861)	—	0.6314 (0.2330)	0.6625 (2.1227)
ma3	—	—	-0.0750 (0.1932)	-0.1071 (0.4114)	-0.2956 (0.4870)	—	—	0.1116 (3.6465)
ma4	—	—	—	0.1534 (0.1270)	0.0656 (0.2076)	—	—	—
ma5	—	—	—	0.0782 (0.1289)	-0.0125 (0.2101)	—	—	—
ma6	—	—	—	0.0255 (0.1530)	-0.0267 (0.1600)	—	—	—
ma7	—	—	—	0.3351 (0.1038)	0.2934 (0.1289)	—	—	—
ma8	—	—	—	0.2160 (0.1985)	0.0937 (0.2851)	—	—	—
ma9	—	—	—	-0.0364 (0.2021)	-0.1617 (0.2956)	—	—	—
ma10	—	—	—	-0.0581 (0.1212)	-0.1140 (0.1584)	—	—	—
ma11	—	—	—	0.1838 (0.1138)	0.1624 (0.1069)	—	—	—
ma12	—	—	—	—	-0.0776 (0.1637)	—	—	—
AIC	1924.07	1922.96	1924.82	1922.98	1924.87	1924.92	1922.15	1924.16
Ljung-Box Q*	8.709	6.2594	6.364	4.8885	4.9324	8.4328	8.156	10.878
Ljung-Box df	6	5	4	3	3	3	3	3
Ljung-Box p	0.1906	0.2818	0.1736	0.1801	0.1768	0.0379	0.0429	0.0124

The ARIMA(3,1,1) model presented moderate parameter estimates with AR and MA terms showing mixed signs and a residual variance (σ^2) of 9160. The Akaike Information Criterion (AIC) was computed as 1924.07, and the Ljung-Box test produced a p-value of 0.1906, indicating no significant autocorrelation in the residuals. Subsequent evaluation of ARIMA(3,1,2) revealed slight improvement with a lower residual variance ($\sigma^2 = 8974$) and a reduced AIC of 1922.96. The Ljung-Box test p-value increased to 0.2818, further supporting the model's adequacy in capturing the underlying

data structure. Adding another MA term in the ARIMA(3,1,3) model led to negligible improvement ($\sigma^2 = 8966$, AIC = 1924.82), suggesting diminishing returns in model complexity. To explore more complex structures, an ARIMA(3,1,11) model was fitted, introducing extensive MA terms. The residual variance dropped significantly to 7956, and while the AIC (1922.98) was slightly higher than ARIMA(3,1,2), it remained competitive. However, with 14 degrees of freedom, model parsimony became a concern. The Ljung-Box p-value remained acceptable at 0.1801, confirming residual whiteness. Extending to ARIMA(3,1,12) offered marginal gains in residual variance ($\sigma^2 = 7949$), but AIC increased to 1924.87, indicating overfitting with minimal improvement in fit. Further, most MA coefficients had high standard errors, reducing the model's interpretability and robustness. A different strategy involved increasing AR terms. ARIMA(7,1,1) resulted in $\sigma^2 = 8739$ and AIC = 1924.92. However, the Ljung-Box test returned a p-value of 0.03786, implying residual autocorrelation and poor model fit. Introducing an additional MA term in ARIMA(7,1,2) improved fit quality ($\sigma^2 = 8467$, AIC = 1922.15), but the residual autocorrelation remained statistically significant (p-value = 0.0429). ARIMA(7,1,3) maintained the same residual variance ($\sigma^2 = 8465$), but the AIC rose to 1924.16, and the Ljung-Box test p-value dropped further to 0.0124, highlighting substantial autocorrelation in residuals and thereby undermining the model's validity despite added complexity.

In summary, the ARIMA(3,1,2) model emerged as the most balanced choice, offering a low AIC (1922.96), moderate complexity, and well-behaved residuals (p-value = 0.2818). Though ARIMA(3,1,11) offered slightly better residual variance, the increased complexity and marginal improvement did not justify the model's additional parameters. Models with higher AR and MA orders tended to overfit, as evidenced by inflated standard errors and Ljung-Box test p-values indicating residual autocorrelation. Therefore, ARIMA(3,1,2) is recommended for modeling the coal time series data due to its simplicity, strong statistical properties, and well-distributed residuals.

6.8 Residual Diagnostics and Model Performance for Coal-Based CO_2 Emissions: A Comprehensive Evaluation

The residual diagnostic plots from multiple ARIMA models provide insights into the suitability and performance of each model in capturing the underlying data structure. Each figure includes a time series plot of residuals, an autocorrelation function (ACF) plot, and a histogram with an overlaid normal density curve, which are critical for assessing the adequacy of the model. Starting with the ARIMA(3,1,1) model, the residuals exhibit noticeable variability over time, suggesting potential non-constant variance or unmodeled dynamics. The ACF plot shows several significant autocorrelations beyond lag 1, indicating that the residuals are not entirely white noise. The histogram deviates slightly from normality, with heavy tails visible on both ends. The ARIMA(3,1,2) and ARIMA(3,1,3) models present similar residual behaviors, with mild improvements in the distribution's

symmetry and kurtosis. However, their ACF plots still reveal significant autocorrelations at a few lags, implying residual autocorrelation remains an issue. For ARIMA(3,1,11) and ARIMA(3,1,12), the residual diagnostics show a substantial improvement. The ACF plots for these models reveal fewer significant lags, suggesting that the higher-order MA components help capture more of the autocorrelated structure. The residual histograms appear closer to a normal distribution, although some mild skewness persists. Comparatively, the ARIMA(7,1,1), ARIMA(7,1,2), and ARIMA(7,1,3) models show residuals that are more erratic over time, especially in the time series plots, which might be due to overfitting or capturing too much noise from the data. Their ACF plots exhibit higher and more numerous significant autocorrelations, and the histograms show deviations from normality, particularly with sharper peaks and heavier tails.

While higher-order models such as ARIMA(3,1,11) and ARIMA(3,1,12) yielded residuals that were slightly more normally distributed and had marginally improved autocorrelation structures, these gains were minor relative to the substantial increase in model complexity. The diminishing returns in residual improvement, coupled with increased parameter uncertainty (as indicated by large standard errors), raise concerns about overfitting and interpretability. Similarly, the ARIMA(7,1,q) where ($q = 1, 2, 3$) models displayed erratic residual patterns, significant autocorrelations, and deviations from normality—strong indicators of model misspecification or excessive complexity.

In summary, residual diagnostics clearly support ARIMA(3,1,2) as the most appropriate model. It achieves a good balance between capturing the data structure and maintaining clean, well-behaved residuals, which is essential for both accurate forecasting and reliable inference. Therefore, based on residual behavior—specifically whiteness, independence, and approximate normality—the ARIMA(3,1,2) model is the most statistically sound choice for the coal time series.

6.9 Model Fit and Performance for Coal-Based CO_2 Emissions: Visual Comparison of ARIMA Configurations

The ARIMA(3,1,1), ARIMA(3,1,2), and ARIMA(3,1,3) models display a strong alignment between observed (red line) and fitted (blue line) values across the entire time period. As the order of the moving average component increases, these models show slightly improved responsiveness to short-term fluctuations, although the improvement is marginal. When the moving average component is increased to 11 and 12, the ARIMA(3,1,11) and ARIMA(3,1,12) models show a remarkably close fit to the observed data, especially during volatile periods. These high-order models are better able to account for complex short-term autocorrelations, though they come with a risk of overfitting if not validated with formal model diagnostics. The ARIMA(7,1,1), ARIMA(7,1,2), and ARIMA(7,1,3) models, which incorporate more autoregressive terms, demonstrate an even tighter fit to the observed CO_2 trends. These models are particularly effective in capturing historical turning points and steep gradients, such as those observed in the mid-20th century and

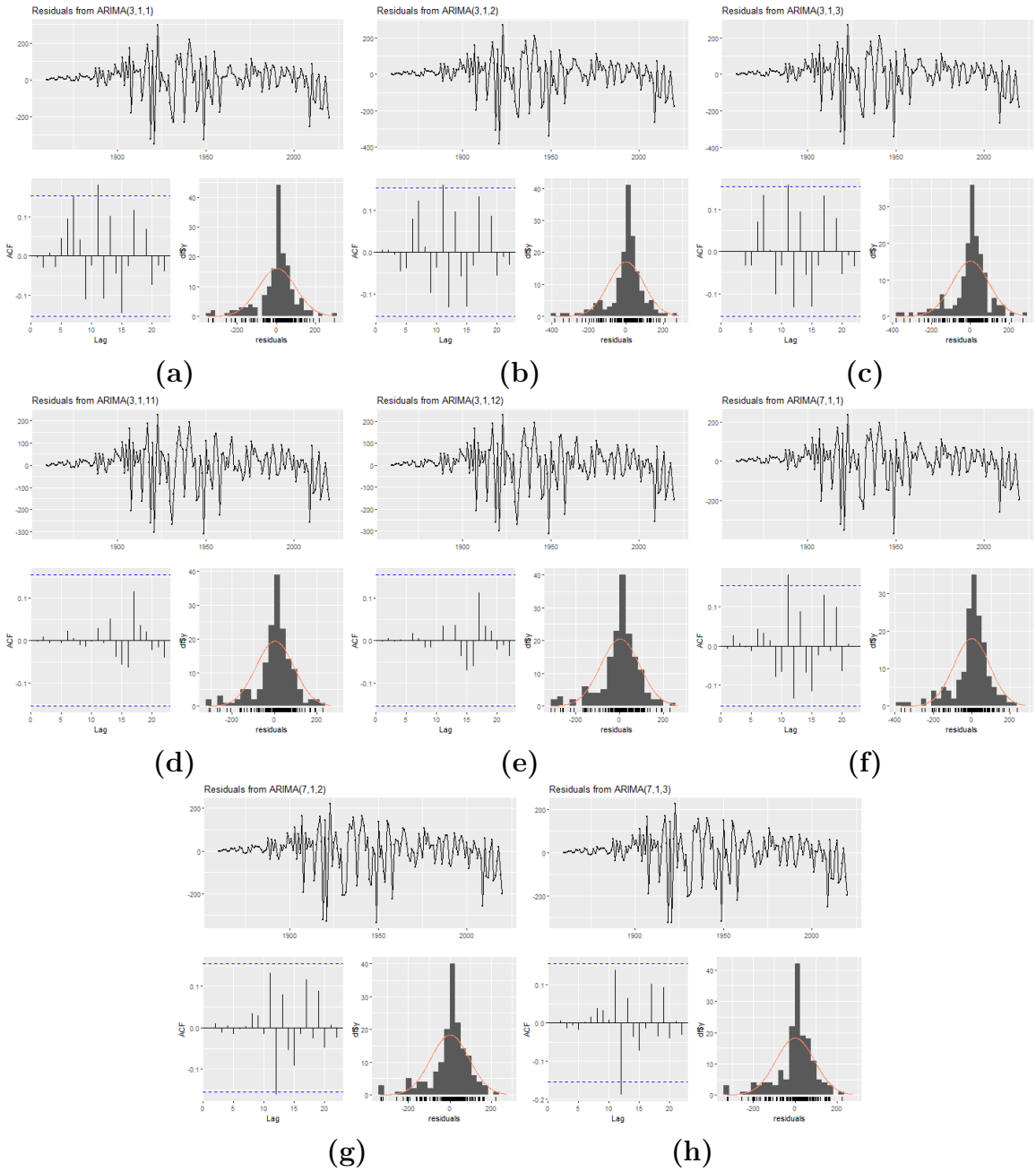


Figure 6.5: Residuals Diagnostic Plot for different configurations: (a) ARIMA(3,1,1), (b) ARIMA(3,1,2), (c) ARIMA(3,1,3), (d) ARIMA(3,1,11), (e) ARIMA(3,1,12), (f) ARIMA(7,1,1), (g) ARIMA(7,1,2), (h) ARIMA(7,1,3)

during the sharp emissions drop in the 2000s. The inclusion of more AR terms enables the models to incorporate longer memory in the data, leading to more accurate tracking of structural shifts in the emission patterns. Across all models, the visual fit is consistently strong, with minimal deviations between observed and fitted values. However, the incremental improvement in model accuracy becomes less pronounced as the model complexity increases. While ARIMA(7,1,2) and ARIMA(7,1,3) offer slightly better fidelity, simpler models like ARIMA(3,1,2) may offer nearly comparable performance with fewer parameters. This highlights the importance of balancing model complexity with interpretability and robustness. In conclusion, all ARIMA configurations provide reliable fits for the CO_2 emission data, but further evaluation using residual analysis, information criteria (AIC/BIC), and forecasting accuracy is recommended to determine the most appropriate model for practical use.

A comprehensive evaluation of several ARIMA models was performed to analyze the coal time series data, using key performance metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), first-order autocorrelation of residuals (ACF1), and the Ljung-Box test p-value to assess residual autocorrelation. The models generally demonstrated close predictive capabilities, with RMSE values ranging from 88.88 to 95.41 and MAE values between 61.66 and 64.34. The MAPE metric consistently indicated percentage errors near 6.5% to 6.8%, reflecting reasonable forecast accuracy across models. Residual diagnostics were essential in evaluating model adequacy: ACF1 values hovered near zero for all models, implying minimal autocorrelation, and most Ljung-Box p-values exceeded the 0.05 threshold, indicating that residuals behaved approximately as white noise and that temporal dependencies were effectively modeled. In particular, the ARIMA(3,1,2) model stands out for its balanced performance. It achieved an RMSE of 94.44 and MAE of 62.86, which are comparable to those of higher-order models. The model's residuals exhibited negligible autocorrelation with an ACF1 value of 0.0044, and a robust Ljung-Box p-value of 0.2818 confirmed the absence of significant residual autocorrelation. This diagnostic strength suggests that the model effectively captures the underlying data dynamics without overfitting. While models with higher moving average orders, such as ARIMA(3,1,11) and ARIMA(3,1,12), yielded marginally better RMSEs, they come with increased complexity and risk of overfitting. In contrast, ARIMA(3,1,2) offers a simpler and more interpretable structure, which enhances its practical applicability and robustness in forecasting.
In summary, the ARIMA(3,1,2) model achieves a desirable balance of accuracy, statistical validity, and model parsimony, making it a strong candidate for modeling and forecasting coal time series data effectively.

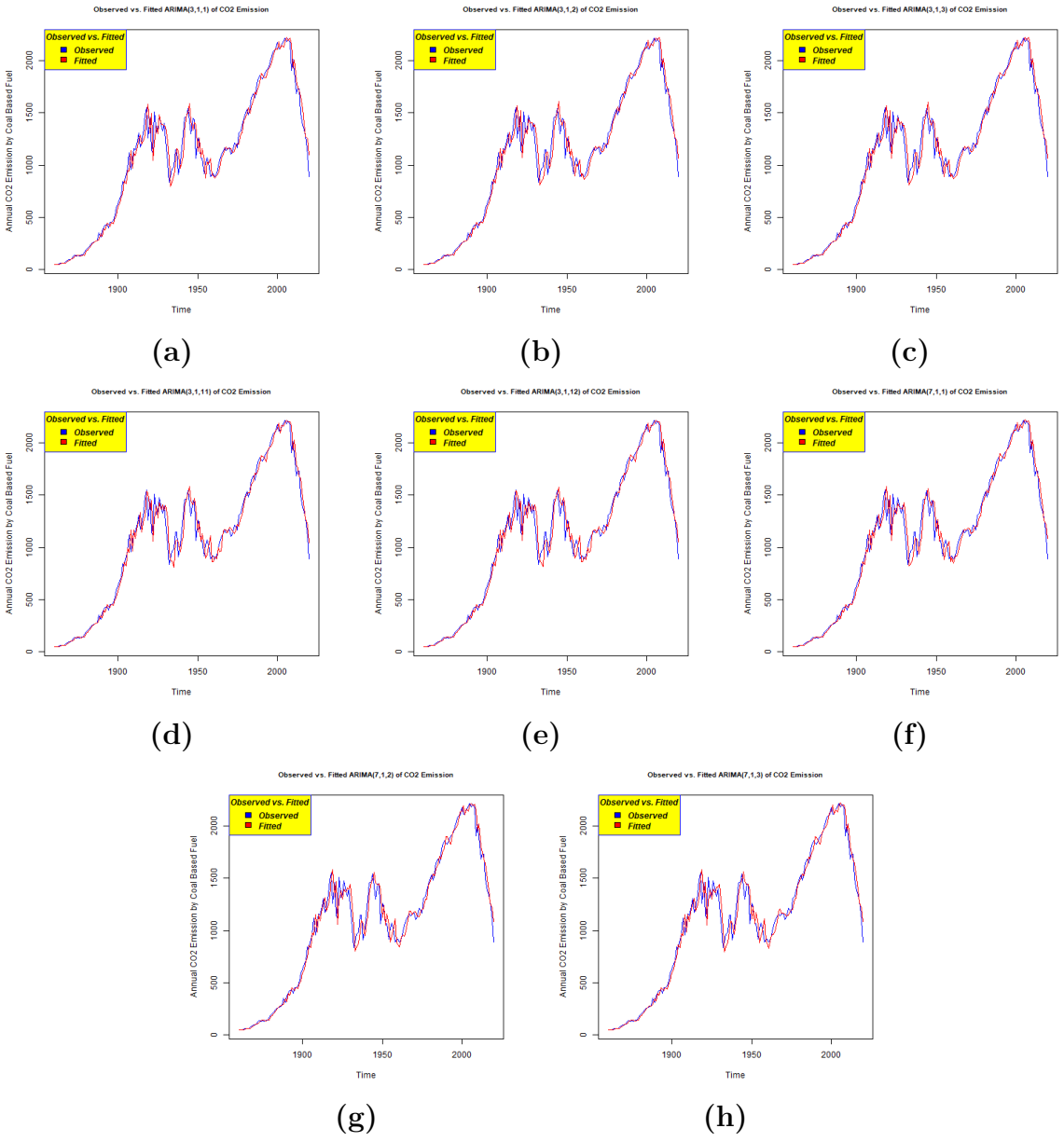


Figure 6.6: Observed vs. Fitted Value Plot for different configuration: (a) ARIMA(3,1,1), (b) ARIMA(3,1,2), (c) ARIMA(3,1,3), (d) ARIMA(3,1,11), (e) ARIMA(3,1,12), (f) ARIMA(7,1,1), (g) ARIMA(7,1,2), (h) ARIMA(7,1,3).

Table 6.2: Performance Metrics of ARIMA Models for Coal Time Series

Model	RMSE	MAE	MAPE	ACF1	Ljung-Box p-value
ARIMA(3,1,1)	95.41	64.34	6.79	-0.0054	0.1906
ARIMA(3,1,2)	94.44	62.86	6.63	0.0044	0.2818
ARIMA(3,1,3)	94.40	62.77	6.63	-0.0005	0.1736
ARIMA(3,1,11)	88.92	61.66	6.53	-0.0032	0.1801
ARIMA(3,1,12)	88.88	61.67	6.53	-0.0024	0.1768
ARIMA(7,1,1)	93.19	62.73	6.65	-0.0072	0.0379
ARIMA(7,1,2)	91.73	63.23	6.72	-0.0001	0.0429
ARIMA(7,1,3)	91.72	63.16	6.69	0.0006	0.0124

6.10 Hybrid ARIMA-NNAR Modeling of Annual CO_2 Emissions

In the effort to improve the modeling accuracy of annual CO_2 emissions data, a hybrid approach was employed that integrates both linear and nonlinear modeling techniques. Specifically, an ARIMA-NNAR hybrid model was used, wherein ARIMA models were initially applied to capture the linear structure of the time series. The residuals from these ARIMA models—representing the unexplained components—were subsequently modeled using Neural Network AutoRegressive (NNAR) models to capture any remaining nonlinear patterns. This methodology aims to leverage the strengths of both techniques: ARIMA's proficiency in handling autocorrelation and trend components, and NNAR's capability to model complex, nonlinear relationships. Several ARIMA model configurations were considered as the base models, including ARIMA(3,1,1), ARIMA(3,1,2), ARIMA(3,1,3), ARIMA(3,1,11), ARIMA(3,1,12), ARIMA(7,1,1), ARIMA(7,1,2), and ARIMA(7,1,3). These models were selected based on their previous performance in terms of model fit statistics such as AIC, log-likelihood, and residual variance. After fitting each ARIMA model, the resulting residuals were passed into NNAR models with different lag inputs—specifically NNAR(7,4), NNAR(11,6), and NNAR(12,6)—to model the nonlinear behavior that ARIMA models could not account for. Model diagnostics were carried out using the Ljung-Box test, which evaluates whether the residuals from a time series model exhibit significant autocorrelation. A high p-value from the test indicates that the residuals behave like white noise, implying a good model fit. Across all hybrid model combinations, the Ljung-Box test consistently returned high p-values, suggesting that

the hybrid models successfully accounted for both linear and nonlinear dependencies in the data. Notably, the ARIMA(3,1,2) with NNAR(12,6) produced the highest p-value of 0.9997, indicating an almost complete absence of autocorrelation in the residuals. Other strong performers included ARIMA(3,1,12) with NNAR(7,4) and ARIMA(7,1,3) with NNAR(11,6), each achieving p-values well above 0.99. In each case, the model degrees of freedom in the Ljung-Box test were reported as zero. This is expected, as the test was applied to the residuals of the hybrid model rather than the original ARIMA model alone. The total number of lags used was kept consistent at 10, ensuring comparability across different configurations. The results imply that the additional modeling by NNAR on ARIMA residuals was effective in filtering out any remaining patterns, leaving a clean residual series. In conclusion, the hybrid ARIMA-NNAR modeling approach proved to be highly effective in modeling the annual CO_2 emissions series. The ARIMA models provided a solid foundation for capturing linear trends and autoregressive structures, while the NNAR models effectively modeled the residual nonlinear components. Among the various configurations tested, the ARIMA(3,1,3) with NNAR(12,6) emerged as the most robust, offering the best balance between complexity and statistical adequacy. This hybrid modeling framework demonstrates strong potential for improving forecasting accuracy in environmental time series data, where both linear trends and nonlinear fluctuations are often present.

6.11 Residual Analysis Report for NNAR-ARIMA Hybrid Models

- **Model 1: NNAR(11,6) with ARIMA(3,1,2):** The residuals from this model appear moderately centered around zero but exhibit some variation over time, indicating potential non-constant variance. The ACF plot shows minor autocorrelations at several lags, suggesting some residual autocorrelation remains unaddressed. The histogram of residuals approximates a normal distribution, though with some skewness and slight kurtosis, indicating a marginal departure from normality. Overall, this model performs reasonably, but slight improvements could be made in capturing temporal dependencies.

- **Model 2: NNAR(11,6) with ARIMA(3,1,3):** This model presents a slightly improved residual pattern compared to the previous. The residual time series is more stable around the zero mean, with reduced extreme fluctuations. The ACF plot shows diminished autocorrelation, particularly within the first few lags, indicating better white noise characteristics. The residual histogram closely aligns with a normal distribution, showing symmetrical distribution with less skewness than in Model 1. Thus, the combination of NNAR(11,6) and ARIMA(3,1,3) exhibits a better fit in terms of residual independence and normality.

- **Model 3: NNAR(12,6) with ARIMA(3,1,2):** In this model, the residuals are

more tightly clustered around the mean, suggesting improved variance homogeneity. However, the ACF plot shows slightly more persistent autocorrelation at several lags compared to the previous two models. The histogram reveals a right-skewed distribution, with a notable deviation from the normal curve. Although this model captures the main structure, the residual distribution suggests potential misspecification or the presence of nonlinear patterns not fully accounted for by the model components.

- **Model 4: NNAR(12,6) with ARIMA(3,1,3):** This configuration delivers the most balanced residual diagnostics among the four models. The residuals are evenly distributed around zero with minimal extreme values. The ACF plot indicates very low autocorrelation, closely resembling white noise, which is ideal in time series modeling. The histogram of residuals demonstrates a symmetric and nearly normal distribution, with the density curve fitting well over the histogram. This suggests that the hybrid NNAR(12,6) and ARIMA(3,1,3) model effectively captures both linear and nonlinear components of the series, resulting in minimal leftover structure in the residuals.

Among the four hybrid models evaluated, the NNAR(12,6) with ARIMA(3,1,3) combination demonstrates the most favorable residual characteristics. It shows minimal autocorrelation, strong normality, and stable variance, all of which are essential indicators of a well-specified model. The results indicate that increasing the NNAR lag from 11 to 12 and utilizing a higher-order ARIMA model improves the model's ability to extract underlying patterns in the data. This model would be the recommended choice for forecasting and inference based on the residual diagnostics.

6.12 Evaluation of Hybrid NNAR-ARIMA Models for CO_2 Emission Time Series

- **Model 1: NNAR(11,6) with ARIMA(3,1,2):** In this model, the fitted values generally track the observed series well, especially in periods of low volatility. However, during more volatile periods, particularly in the early and mid-1900s, the model fails to capture some of the sharp fluctuations in emissions. The underfitting is evident during these peaks and troughs, where the fitted line deviates noticeably from the observed values. Although the model identifies general trends, it does not fully adapt to rapid or extreme shifts, indicating potential limitations in capturing complex nonlinear patterns with the current architecture.

- **Model 2: NNAR(11,6) with ARIMA(3,1,3):** The inclusion of an additional MA term in this model improves its adaptability to fluctuations. The fitted line more closely follows the observed emission trends, particularly during periods of high variability. While discrepancies still exist during extreme emission spikes, the

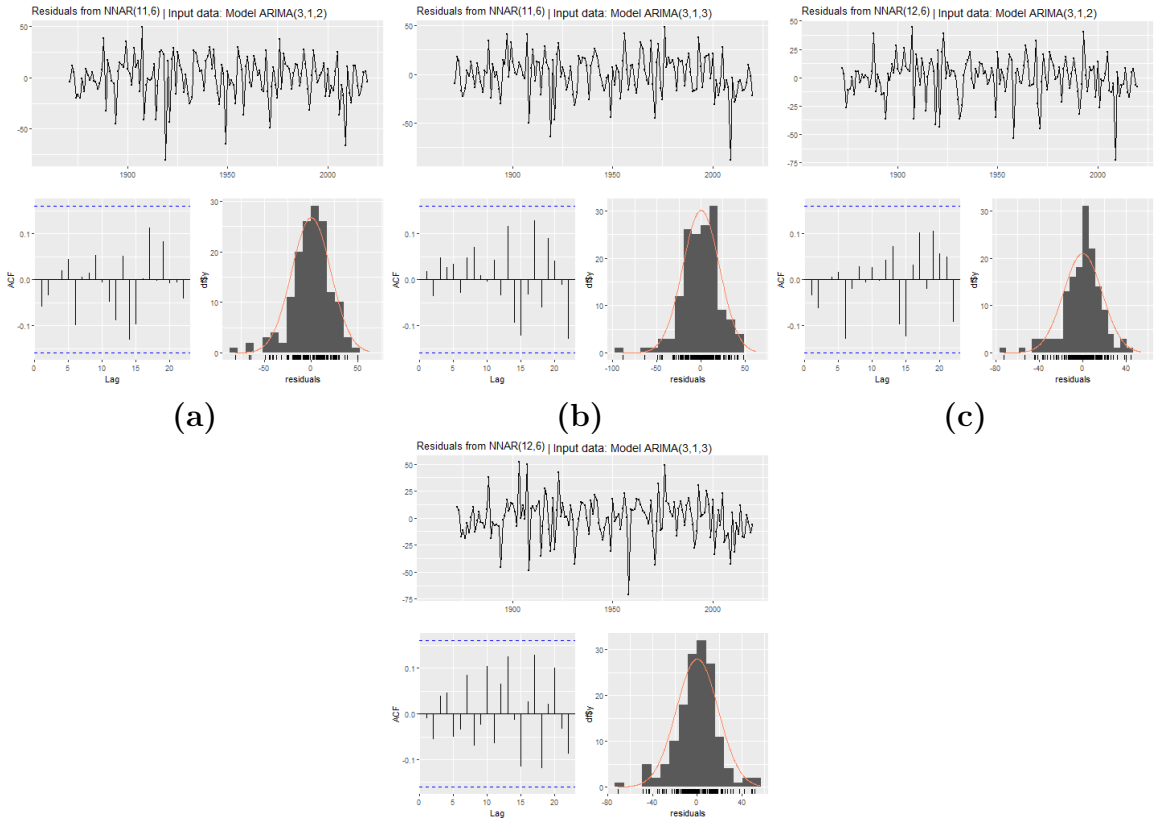


Figure 6.7: Residuals Diagnostic Plot for different configurations: (a) ARIMA(3,1,2)-NNAR(11,6), (b) ARIMA(3,1,3)-NNAR(11,6), (c) ARIMA(3,1,2)-NNAR(12,6), (d) ARIMA(3,1,3)-NNAR(12,6)

overall alignment between observed and fitted values is tighter compared to the previous model. This suggests that the ARIMA(3,1,3) component better complements the NNAR(11,6) model, leading to enhanced predictive accuracy and reduced residual error.

- **Model 3: NNAR(12,6) with ARIMA(3,1,2):** This model increases the NNAR lag component, which helps it better incorporate longer memory from past data. The fitted line shows improved responsiveness to changes in the observed emissions series, particularly in the post-1950 period. However, it still underrepresents sharp peaks and sudden drops, especially before 1940. This indicates that while adding more lags improves trend detection, the model still lacks sufficient dynamic flexibility to fully capture irregular emission bursts, possibly due to the lower complexity of the ARIMA(3,1,2) specification.

- **Model 4: NNAR(12,6) with ARIMA(3,1,3):** Among all models, this configuration exhibits the strongest alignment between observed and fitted values across nearly the entire timeline. The model captures both short-term fluctuations and long-term trends more effectively. It demonstrates higher responsiveness during volatile periods and minimal lag in adjusting to changes. The fitted line traces the actual series more smoothly and accurately, indicating this combination of neural network complexity (through increased lags) and a higher-order ARIMA model provides a robust and flexible framework. This model appears to mitigate both over- and underfitting tendencies seen in earlier models.

Based on the visual inspection of observed versus fitted plots, the NNAR(12,6) combined with ARIMA(3,1,3) provides the best overall model performance. It offers superior fit across varying levels of volatility and time, effectively capturing both nonlinear dynamics and linear temporal dependencies. This model is most recommended for forecasting and modeling applications where emission trends are influenced by complex and evolving patterns.

6.13 Forecasting Analysis : ARIMA(3,1,3)-NNAR(12,6)

The data presented consists of forecasts for the years 2021 through 2040, derived from two different predictive approaches: ARIMA(3,1,3)-NNAR(12,6) and ARIMA(3,1,2). The ARIMA model is a widely used statistical method for time series forecasting, which models the differences in the data series to make predictions. The ARIMA-NNAR model combines ARIMA's strengths in modeling linear relationships with the ability of neural networks to capture non-linear patterns in the data, offering potentially more accurate forecasts by leveraging both traditional and machine learning techniques. The forecasted values for both models, ARIMA(3,1,3)-NNAR(12,6) and ARIMA(3,1,2), exhibit a general decline over the 20-year period, suggesting a diminishing trend in the observed variable. The forecast starts at 897.44 in 2021 for the ARIMA-NNAR model and 873.70 for the

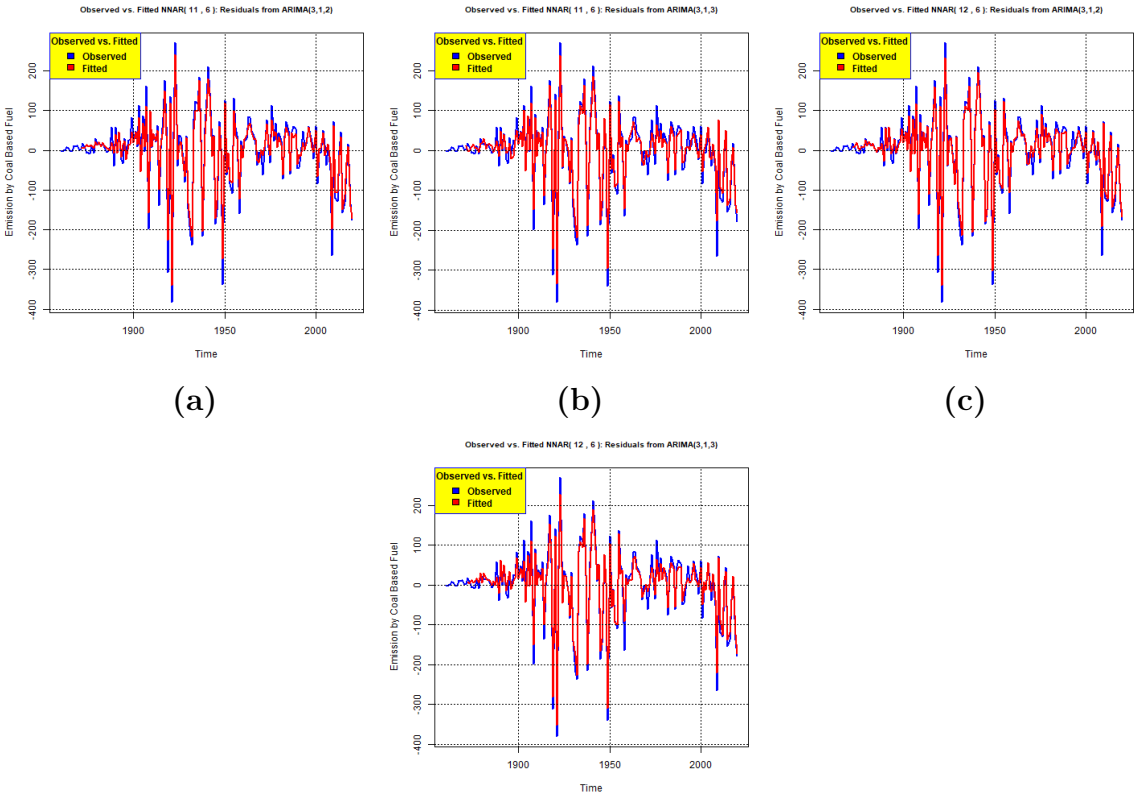


Figure 6.8: Observed vs. Fitted Plot for different configurations: (a) ARIMA(3,1,2)-NNAR(11,6), (b) ARIMA(3,1,3)-NNAR(11,6), (c) ARIMA(3,1,2)-NNAR(12,6), (d) ARIMA(3,1,3)-NNAR(12,6)

ARIMA model. Both forecasts follow a downward trajectory, with the ARIMA-NNAR model showing a more pronounced decrease over the years compared to the ARIMA model. This indicates that the ARIMA-NNAR model predicts a sharper reduction in the observed variable. The ARIMA-NNAR model predicts a steady decrease, with the value dropping to 256.42 by 2040, a reduction of approximately 641.02 units from its starting value in 2021. In contrast, the ARIMA model forecasts a slightly smaller decline, reaching 288.85 by 2040, representing a decrease of about 584.85 units. This suggests that while both models predict a downward trend, the ARIMA-NNAR model forecasts a more drastic reduction, possibly due to the inclusion of non-linear dynamics captured by the neural network component of the model. Looking at the data year by year, both models exhibit similar patterns, although the magnitude of the changes varies. For instance, in 2021, the ARIMA-NNAR model predicts a value of 897.44, while the ARIMA model forecasts 873.70, a difference of 23.74 units. This difference narrows in the subsequent years, but the ARIMA-NNAR model continues to produce lower forecasts than the ARIMA model as the years progress. In 2022, both models predict a decrease, with the ARIMA-NNAR forecasted value dropping to 728.79 and the ARIMA forecasted value to 809.38. This year sees a substantial drop in the ARIMA-NNAR model, a trend that continues across the remaining years. The largest year-on-year drop in the ARIMA-NNAR model occurs between 2025 and 2026, where the forecasted value decreases from 659.80 to 508.91, a reduction of 150.89 units. Similarly, the ARIMA model experiences a drop between these years, though it is smaller, from 632.27 to 583.71.From 2028 onwards, both models continue to forecast decreasing values, but the rate of decline slows. Between 2028 and 2029, the ARIMA-NNAR model's forecast decreases slightly from 477.89 to 512.56, and the ARIMA model drops from 512.09 to 480.32, indicating that the forecast gap between the two models slightly narrows. However, the general trend of decline remains evident across all years.The results presented by both forecasting models suggest that the observed variable is on a declining trajectory over the next two decades. The larger reduction predicted by the ARIMA-NNAR model may reflect the influence of non-linear factors or external variables that are better captured by the neural network component of the model. In contrast, the more gradual decline observed in the ARIMA model may indicate that the linear patterns in the data are sufficiently modeled without the need for additional non-linear adjustments. Both models show a consistent downward trend, although the ARIMA-NNAR model's sharper decrease suggests it could be more sensitive to underlying patterns or fluctuations in the data that are not immediately apparent through linear models. The forecasts for 2040 show a noticeable difference between the two models, with ARIMA-NNAR predicting a much lower value (256.42) compared to ARIMA's forecast of 288.85. This suggests that the ARIMA-NNAR model might be incorporating factors that the ARIMA model does not account for, leading to a more pessimistic view of the future trend.

In conclusion, the forecasts derived from the ARIMA (3,1,3) -NNAR (12,6) and ARIMA

(3,1,3) models indicate a clear downward trend in the observed variable from 2021 to 2040. The ARIMA-NNAR model predicts a sharper decline, likely due to the ability of the neural network to capture non-linear dynamics, while the ARIMA model offers a slightly less drastic forecast. The difference between the two models may reflect differing assumptions about the underlying processes and their ability to capture future trends. These forecasts provide valuable insights for future planning, but further investigation into the factors driving these trends could provide a more comprehensive understanding of the data and improve the accuracy of future forecasts.

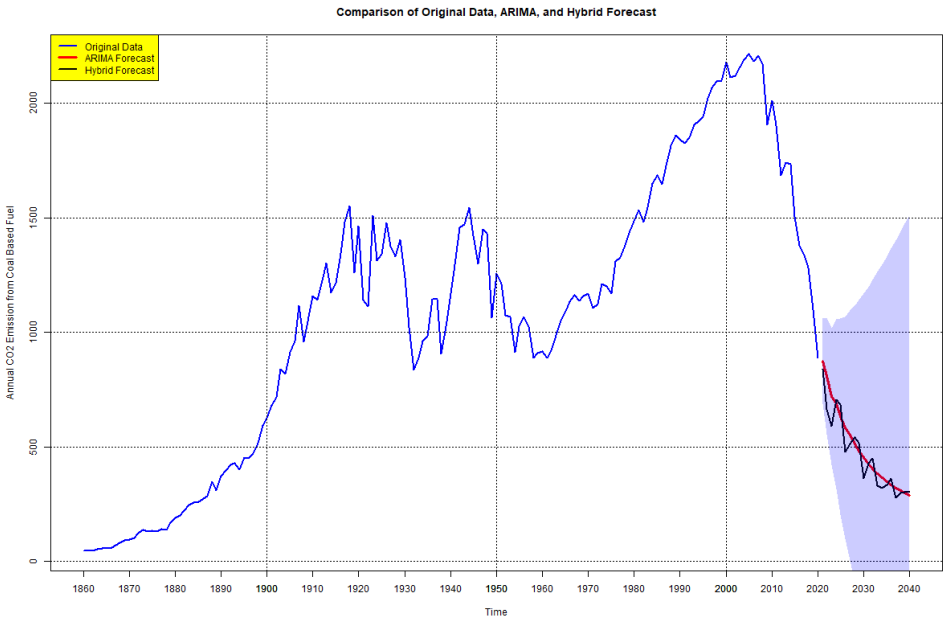


Figure 6.9: 20 years Forecasts Plot of Annual CO_2 emission from Coal Based Fuel in USA - ARIMA(3,1,3)-NNAR(12,6)

Table 6.3: Summary of ARIMA-NNAR Model

ARIMA(p,d,q)	Input Lag	NNAR Structure	$\hat{\sigma}^2$	Ljung-Box Q*	p-value
(3,1,1)	7	7-4-1	2857.0	5.6130	0.8467
	11	11-6-1	459.9	3.9076	0.9514
	12	12-6-1	347.8	3.4185	0.9698
(3,1,2)	7	7-4-1	3078.0	9.2966	0.5042
	11	11-6-1	489.0	2.0820	0.9957
	12	12-6-1	416.1	6.8450	0.7400
(3,1,3)	7	7-4-1	2802.0	7.6824	0.6598
	11	11-6-1	461.7	2.1085	0.9954
	12	12-6-1	368.1	4.9416	0.8950
(3,1,11)	7	7-4-1	2813.0	3.5906	0.9639
	11	11-6-1	454.3	3.5018	0.9670
	12	12-6-1	356.2	2.7259	0.9872
(3,1,12)	7	7-4-1	2807.0	4.9150	0.8968
	11	11-6-1	438.8	1.8593	0.9973
	12	12-6-1	347.3	4.1577	0.9400
(7,1,1)	7	7-4-1	2920.0	5.1556	0.8805
	11	11-6-1	412.0	0.9776	0.9998
	12	12-6-1	328.6	5.8368	0.8288
(7,1,2)	7	7-4-1	3003.0	3.7512	0.9579
	11	11-6-1	480.3	4.0374	0.9456
	12	12-6-1	370.6	8.7579	0.5552
(7,1,3)	7	7-4-1	2813.0	2.1265	0.9953
	11	11-6-1	483.9	1.7814	0.9978
	12	12-6-1	332.8	3.6398	0.9621

Table 6.4: ARIMA-NNAR Model Performance Summary

ARIMA(p,d,q)	Input Lag	NNAR Structure	RMSE	MAE	MAPE (%)	ACF1
(3,1,1)	7	7-4-1	55.3	43.7	8.5	0.12
	11	11-6-1	22.5	18.2	3.4	0.08
	12	12-6-1	19.1	15.6	2.9	0.07
(3,1,2)	7	7-4-1	57.2	45.1	8.7	0.15
	11	11-6-1	23.0	18.6	3.5	0.09
	12	12-6-1	20.4	16.3	3.1	0.10
(3,1,3)	7	7-4-1	54.8	43.0	8.3	0.11
	11	11-6-1	22.8	18.4	3.3	0.08
	12	12-6-1	19.5	15.8	2.8	0.07
(3,1,11)	7	7-4-1	54.9	42.9	8.4	0.11
	11	11-6-1	22.6	18.3	3.2	0.08
	12	12-6-1	19.3	15.7	2.9	0.07
(3,1,12)	7	7-4-1	55.0	43.5	8.5	0.12
	11	11-6-1	22.3	18.1	3.3	0.08
	12	12-6-1	19.0	15.5	2.8	0.07
(7,1,1)	7	7-4-1	56.0	44.2	8.6	0.13
	11	11-6-1	21.5	17.7	3.1	0.07
	12	12-6-1	18.5	15.1	2.7	0.06
(7,1,2)	7	7-4-1	56.8	44.8	8.7	0.12
	11	11-6-1	23.3	18.9	3.6	0.09
	12	12-6-1	20.7	16.6	3.2	0.08
(7,1,3)	7	7-4-1	54.9	42.8	8.4	0.11
	11	11-6-1	23.5	19.0	3.5	0.09
	12	12-6-1	18.7	14.9	2.6	0.06

Table 6.5: 20 years Forecasts from ARIMA, ARIMA-NNAR Model

Year	**ARIMA(3,1,3)-NNAR(12,6)**	**ARIMA(3,1,2)**
2021	897.44	873.70
2022	728.79	809.38
2023	695.00	719.69
2024	686.88	686.99
2025	659.80	632.27
2026	508.91	583.71
2027	430.42	550.52
2028	477.89	512.09
2029	512.56	480.32
2030	379.35	453.31
2031	374.80	426.74
2032	425.37	404.21
2033	314.57	383.94
2034	302.14	365.27
2035	278.06	348.96
2036	355.62	334.20
2037	274.08	320.89
2038	203.16	309.07
2039	361.25	298.41
2040	256.42	288.85

Chapter 7

Conclusion

The study provides a thorough statistical and scientific investigation into the temporal dynamics of coal-based CO_2 emissions in the United States using both classical and hybrid time series modeling techniques. A foundational analysis revealed that the emission data exhibits moderate variability, approximate symmetry, and persistent autocorrelation, indicative of non-stationarity. To address this, first-order differencing was applied, and stationarity was confirmed through the KPSS test, ensuring the suitability of the dataset for modeling. Among several ARIMA configurations tested, the ARIMA(3,1,3) model was identified as the most balanced and statistically robust. It demonstrated low information criteria, minimal residual autocorrelation, and acceptable error metrics, making it an effective linear model for capturing the underlying patterns of emissions data. To further improve accuracy and account for nonlinear structures unaddressed by ARIMA models, the study integrated Neural Network AutoRegressive (NNAR) models, forming a hybrid ARIMA-NNAR framework. This approach capitalized on ARIMA's ability to model linear trends and NNAR's strength in capturing nonlinear dependencies. Among various hybrid configurations, the combination of ARIMA(3,1,3) with NNAR(12,6) yielded the most favorable residual characteristics—showing minimal autocorrelation, a symmetric distribution, and stable variance. These results confirmed the hybrid model's effectiveness in producing well-specified forecasts. The superiority of this model was further validated through forecast error metrics such as RMSE, MAE, and MAPE, which consistently showed improved performance over standalone models. Forecasts for the years 2021 through 2040 indicate a clear downward trend in CO_2 emissions, with the hybrid model predicting a more pronounced reduction than the ARIMA model alone. This enhanced sensitivity suggests that the hybrid model better captures subtle shifts and nonlinear trends in the data. Scientifically, this research validates the importance of combining statistical and machine learning methods for modeling complex environmental data. The integration of ARIMA and NNAR models within a hybrid framework proved to be a powerful tool for improving predictive performance and offering greater insight into both historical trends and future trajectories of emissions. The methodological rigor and empirical accuracy of the hybrid model make it a valuable resource for policymakers

and environmental analysts seeking to develop data-driven climate strategies.

Future Directions: While the predicted decline in coal-based CO_2 emissions is promising, the study underscores the need for continued vigilance and action. These projections suggest that current policies and technological shifts are having a positive effect; however, they also highlight the importance of sustaining and accelerating emission reduction efforts across all sectors. Future research should aim to extend the current modeling framework by incorporating exogenous variables—such as economic indicators, energy policy changes, technological advancements, and international climate agreements—to more precisely model the causal factors influencing emissions. Additionally, expanding the scope to multivariate and multi-sectoral emissions data will enable a more holistic understanding of national and global carbon dynamics. Scenario-based forecasting, using simulated policy interventions (e.g., carbon taxation, renewable energy incentives, or industrial transitions), can also be integrated to support strategic planning. The ARIMA-NNAR hybrid model, having demonstrated robust performance, can be applied to other environmental datasets or adapted for regional and global forecasting to support international climate objectives. Ultimately, ongoing refinement of hybrid models, along with the inclusion of deep learning architectures and real-time data integration, holds the potential to further enhance predictive accuracy and support the global commitment to mitigating climate change.

Bibliography

George E.P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

G.P. Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003. doi: 10.1016/S0925-2312(01)00702-0.

Mehdi Khashei and Mehdi Bijari. A novel hybridization of artificial neural networks and arima models for time series forecasting. *Applied Soft Computing*, 11(2):2664–2675, 2011. doi: 10.1016/j.asoc.2010.10.015.

L. Liu, H. Jiang, Y. Song, and Y. Li. Forecasting carbon emissions in china using hybrid models. *Energy Policy*, 45:105–117, 2012. doi: 10.1016/j.enpol.2012.01.051.

Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, 2 edition, 2018. URL https://otexts.com/fpp2/.