TIME SERIES ANALYSIS AND QUALITY CONTROL OF POWER TRANSMISSION DATA

Case Study: Transmission Company of Nigeria (TCN), Omotosho Chapter, Ondo State

Presented By

OLUGBADE, DANIEL IYANU
MTH/2021/083

A PROJECT SUBMITTED TO THE DEPARTMENT OF STATISTICS, FACULTY OF SCIENCE, OBAFEMI AWOLOWO UNIVERSITY, ILE-IFE, OSUN STATE, NIGERIA,
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF
BACHELOR OF SCIENCE (B.Sc. HONS) DEGREE IN STATISTICS.

SEPTEMBER, 2025.

CERTIFICATION

This project has been read and certified as meeting the partial requirements for the award of the degree of Bachelor of Science (B.Sc.) in Statistics, Obafemi Awolowo University, Ife, Nigeria.

Prof. A. A. Olosunde (Supervisor)	Date
Dr. I. O. Ayodeji (Head of Department)	Date
External Moderator	Date.

DEDICATION

I dedicate this project first to Almighty God—my source of wisdom, strength, and unfailing love. And to my family Mr and Mrs olugbade, whose constant prayers, sacrifices, and encouragement made this journey possible. This achievement is as much yours as it is mine.

ACKNOWLEDGEMENT

I give all glory to Almighty God for his grace, wisdom, and strength throughout the course of this project. Without His guidance, this work would not have been possible. My heartfelt appreciation goes to my family (The Olugbade's) and blood (JFash, Mindset) for their unwavering support, love, and encouragement during this journey. Your sacrifices and constant belief in me have been my greatest motivation. I also sincerely thank my supervisor, Prof. A.A. Olosunde, for his invaluable guidance, patience, and constructive advice, which have greatly enriched this work. My gratitude extends to the HOD, Dr. I.O. Ayodeji, and the entire lecturers in Statistics and Mathematics who have impacted knowledge on us individually towards the success of our programme. I am also sincerely thankful to my biggest brothers (G.O, TJ, and Murphy) and my friends, who stood by me with encouragement, understanding, and assistance in different ways during this journey. Finally, my appreciation also goes to all Transmission Company Of Nigeria (TCN) staff, Omotosho chapter, for allowing me to use their data in this project. This achievement would not have been possible without all of you. God bless you all.

ABSTRACT

This project focuses on the application of time series analysis and quality control techniques to power transmission data in order to monitor performance, detect irregularities, and improve operational efficiency. Daily load data were collected and methods to identify patterns, trends, and potential anomalies in power transmission. The study employed time series modeling techniques, including ARIMA forecasting, to predict future load behavior, while quality control tools such as control charts were used to evaluate system stability and highlight periods of deviation from expected performance. Results from the analysis revealed significant seasonal patterns and load variations that could impact transmission efficiency if not properly managed. The forecasts provided reliable short-term load predictions, while the control charts effectively identified outlier periods requiring prompt intervention. The findings demonstrate that integrating time series forecasting with quality control methods can enhance decision-making, reduce losses, and improve the reliability of power transmission networks. It is recommended that utility operators adopt these analytical techniques for continuous monitoring and proactive system management.

Contents

List of Figures

List of Tables

Chapter 1: Introduction

The purpose of statistics in a transmission company cannot be neglected since they setup the organisation with the aim of transmitting power in good quality that will satisfy the Distribution sector and the consumer of power.

Virtually all distribution sector and the consumer needs adequate power supply, uninterrupted power and fair regulation so as to check the conformance of the power supply. If not so, there will be need to embark on statistical tools called "statistical quality control" to check the satisfaction.

Cautiously, quality control should be applied as a means of fair regulation and prevent power being interrupted. This technique is being employed by all transmission company that knows the importance of their power supply in satisfying human use. The quality of power is a composite of many factors like regular supply, duration, measurement and less payment of power. Quality control is a staff function with the objectives of co-ordinating the transmission facilities in transmitting power in accordance to the specification. Statistical quality control is a very special statistical techniques employed by company to find the conformity of their supply to specification. It is usually applicable where there is a generating and transmitting company together.(olabiyiidris et al 2012) Transmission Company of Nigeria, omotosho 330kV substation was energized in 2007 and was commissioned by Dr. Olusegun Obasanjo GCFR under the administration of Dr. Olusegun Kokumon Agagu, Ondo State Governor in1 2007. It was one of the NIPP project under the then Administration, and was designed and constructed by a chinese construction company named China Machinery Engineering Company (CMEC). The omotosho Works Centre was commissioned with an installed capacity of 2/150 MVA Transformer with the maximum load of 240MW.The Transmission Company of Nigeria omotosho is a substation operating at 330/132/33kV,its received High Voltage Alternating Current (HVAC) of 330kV from benin generating station and Oshogbo National Grid station. The omotosho Works Centre is the Area Control Centre (ACC) and has two sub-transmission stations under its supervision, 132kV are sent to the substations under TCN omotosho which are

Transmission Company of Nigeria Akure ,operating at 132/33kV and Transmission Company of Nigeria, ondo operating at 132/33kV. In each of the stations, the four 33kV is now feeds to the districts or Distribution Companies as well as the special industries. The commissioned was based on the insufficient power supply in the vicinity of the ondo State and its laboring states. (olugbade 2018) The organisation consist of two phases: PHASE 1:The first generating station have 8 generating units (turbines) GT1-GT8.The phasel generating turbines has the capacity to generate power at its maximum peak of 335MegaWatt (335MW).Each turbine can generate 41.8MW at its maximum peak and the rate at which each turbine generate their power will be sum together to give the total load on phase 1 generating. PHASE 2: The phase two generating station have 4 generating unit (turbines) GT1-GT4. The unit has the capacity to generate 500MegaWatt (500MW) as its maximum peak, and each turbine has the capacity to generate 125MW.The phase two generating unit generate more power than the phase one generating unit. The combination of 3-4 turbine capacity in phase 1 is equivalent to a turbine capacity in phase 2.

Table 1.1: Definition Of Some Terms In Transmission Company Of Nigeria

S/N	**Term**	**Definition**
1	CIRCUIT BREAKER	A Circuit Breaker is an automatically operated electrical switching device designed to protect electrical circuits from damage caused by excess current from an overload or short circuit. Its basic function is to interrupt current flow after a fault is detected.
2	WAVE TRAP	Wave trap is a device designed in a parallel tuned inductor–capacitor tank circuit made to be resonant at the desired communication frequency. It is installed in the substation for trapping the high-frequency communication signal sent on the line from a remote substation and diverting it to the telecom panel in the substation control room.
3	SECONDARY INJECTOR	A method of connecting a secondary injection test set to a trip unit (trip device, overcurrent module, protection device, OCR, ETU, etc.) on a circuit breaker, VT, and CT, and injecting a simulated current to prove it works at different levels.
4	LEAKAGE CURRENT TESTER (CLAMP ON)	This measures leakage current, which most commonly flows in the insulation surrounding conductors and in the filters protecting electronic equipment around the home or office.

S/N	Term	Definition
5	AUTO-TRANSFORMER	A transformer is a static device that transfers electric energy from one circuit to another without changing frequency. An autotransformer is a single-winding transformer with taps. With primary voltage applied to the primary terminals, the required secondary voltage from zero volts to the rated primary volts can be obtained from the secondary by varying the taps.
6	RELAYS	An automatic protective device designed to trip a circuit breaker when a fault is detected. Traditional protective relays were electromagnetic devices, relying on coils and moving parts to detect abnormal operating conditions such as overcurrent, overvoltage, reverse power flow, over-frequency, and under-frequency.
7	LIGHTNING ARRESTER	A device used on electric power and telecommunication systems to protect insulation and conductors from the damaging effects of lightning. Arresters bypass voltage surges that are harmful to connected loads and the system. When a lightning (or similar switching) surge travels along the power line to the arrester, the surge current is diverted through the arrester—typically to earth.
8	ISOLATORS	Mechanical switching devices used to disconnect a portion of a circuit. They are operated when the load is already cut off. Like circuit breakers, isolators are used in making and breaking the circuit.
9	INSTRUMENT TRANSFORMERS	High-accuracy electrical devices used to isolate or transform voltage or current levels. Commonly used to operate instruments or metering from high-voltage or high-current circuits, safely isolating secondary control circuitry from those high levels.

S/N	Term	Definition
10	ISOLATING ROD	A tool used to carry out isolation manually; some isolators can be operated remotely (mostly for high voltage levels—132 kV and 330 kV).
11	LOG BOOK	A book in which activities in the system are recorded for proper system control, stability, and reference. Entries include feeder outages and reasons, application/issuance of station guarantee, reports from substations under area control, reports to the regional control centre, trouble reports, etc.
12	SYSTEM CONTROL AND DATA ACQUISITION (SCADA)	A solution for data acquisition, monitoring, and control over large geographical areas; it combines data acquisition and telemetry.
13	GAS KIT	Contains gas pipes and nozzles for gas filling, e.g., for SF_6 circuit breakers.
14	OIL DIELECTRIC STRENGTH TESTER (ODST)	Used to test the dielectric strength of transformer oil.
15	TYFOR	Used to draw and control any tree touching the conductor, with the aid of wire swing, to direct the tree away from the line.
16	CLAMP	Straight or T-clamps; fastening devices used to hold or secure objects tightly together to prevent movement or separation by applying inward pressure.
17	TURBINES	Generators that produce power using gas, wind, or water as the driving energy for the turbine.
18	COOLANT	A water reservoir/system that cools the turbines.

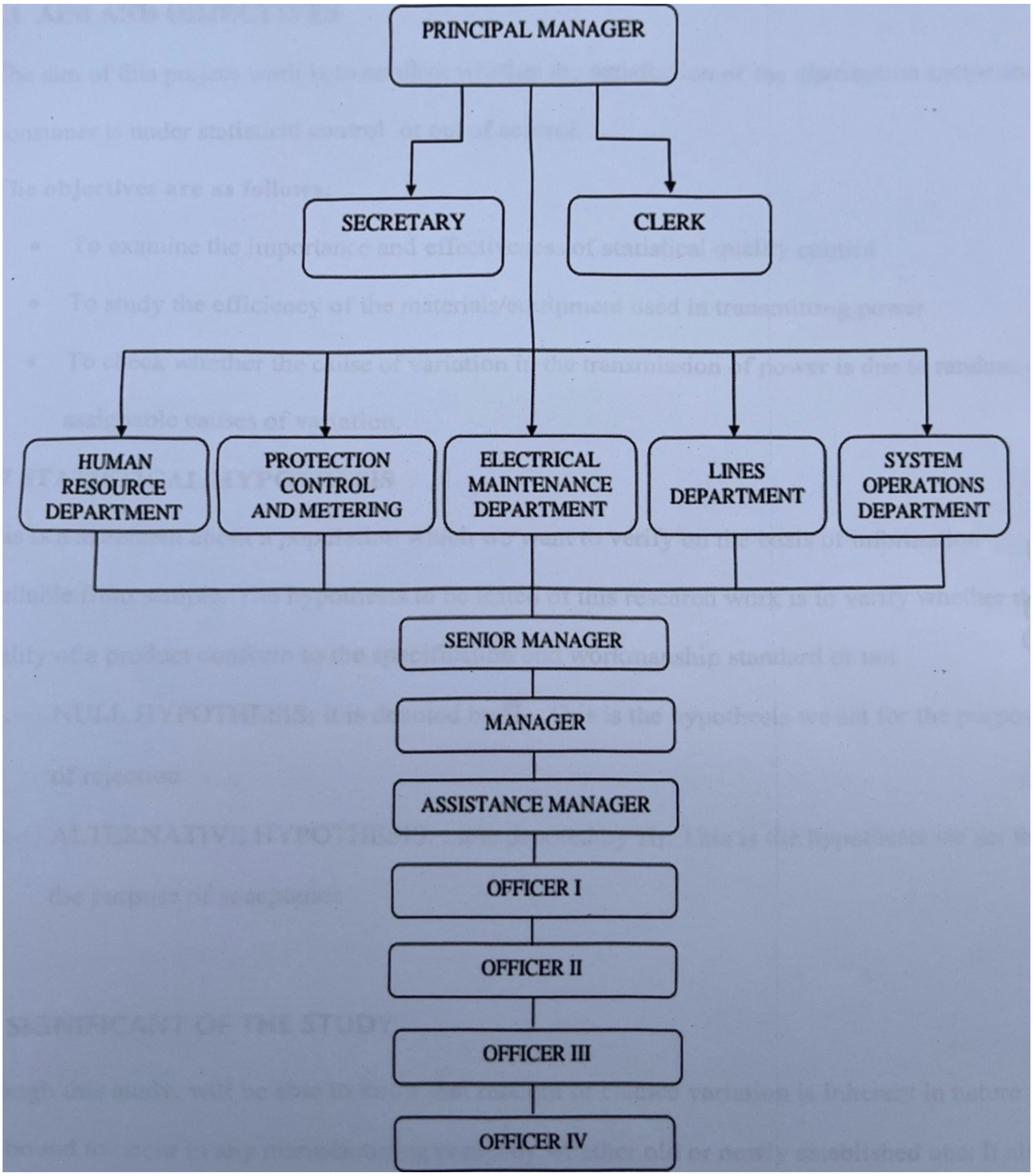

Figure 1.1: Organisation Structure of TCN

1.1 Aim and Objectives

The aim of this project is to analyse the time series behaviour of power transmission data to identify patterns, trends, and anomalies, and to apply quality control techniques to ensure accuracy, reliability, and operational efficiency in the power transmission system.

Objectives of the Study

The Objectives of the Study are to:

i. Derive a time series model for transmission of power data;

ii. Determine the upper and lower limit quality control for transmission of power;

iii. Apply the time series model to transmission of power; and

iv. Test the adequacy and validity of the time series.

1.2 Statistical Hypothesis

i. **Null Hypothesis ($\mathbf{H_0}$):** The proposed model gives the best fit for the data.

ii. **Alternative Hypothesis ($\mathbf{H_1}$):** The proposed model does not give the best fit for the data.

1.3 Significance of the Study

Through this study, we will be able to test the adequacy and validity of the time series model on transmission of power and recognize that random or chance variation is inherent in nature and bound to occur in any manufacturing company, whether old or newly established. It also exposes the researcher and management of the company to the various quality control measures needed to improve capacity performance with the application of MR-Chart and $\overline{X}$-Chart.

1.4 Scope of the Study

The project work is majorly designed to cover the application of time series and statistical quality control in transmission of power in order to improve utilization. The study identifies the importance of statistical quality control and time series during transmission processes, and the problems encountered in achieving their aim.

1.5 Abbreviation of Some Terms

Y_t Original value of Time series

S_t Seasonal variation

C_t Cyclical variation

I_t Irregular variation

$T = a + bx$ Trend equation

$T \times SI$ Using least square method to forecast

SI Seasonal indices

MA Moving Average

Q.T Quarterly Total

Q.M Quarterly Mean

A.M Adjusted Mean

$\overline{X}$ Average of all the sample

R Range

$\overline{R}$ Average range

P Proportion of defectives

$\overline{P}$ Average proportion of defectives

C Number of defectives

$\overline{C}$ Average number of defectives

N Number of observations

UCL Upper Control Limit

LCL Lower Control Limit

AQL Acceptable Quality Level

AOQ Average Outgoing Quality

AOQL Average Outgoing Quality Limit

LTDP Lot Tolerance Percent Defective

1.6 Definition of Some Statistical Terms

Accuracy The degree of conformance of a product to the specification.

Sampling A technique for selecting a part of a population for study.

Variable Figures arising from the random examination of sample products.

Conform The extent or degree to which a product meets pre-established standards of specification.

Performance The primary operating characteristics of the product (e.g., acceleration).

Sampling Plans Necessary for ensuring sampling is fully representative.

Acceptance Quality Level The level of quality considered to be good and desirable to be accepted.

Producer's Risk (α) The probability of rejecting a lot of relatively good quality; also the probability of committing Type I error.

Consumer's Risk (β) The probability of accepting a lot of relatively poor quality; also the probability of committing Type II error.

Process A series of actions or operations that transform inputs to outputs.

Chapter 2: Literature review

2.1 BRIEF HISTORY ON POWER & GRID EXTENSION

Over 1.6 billion people of the world have no access to electricity due to high capital cost required for extending the grid and the utility grid is hard pressed to meet the growing demand of urban areas with lowest priority to rural areas.

Deshmukh and Bilolikar (2006) studied the feasibility of grid extension and distributed generation considering biomass and diesel-based generation options and BEP (break even point) based optimization suggested it as a cheaper option than grid extension.

Monteiro et al. (2005) studied the impact of integration of distributed resources on electricity distribution using a spatial support system.

Geographical information systems (GIS) was used by Zhou Quan et al. (2002) to develop mathematical models of the substation location and capacity optimization and proposed a new multi-period optimal selection algorithm of substation that can determine the reasonable location, capacity and time of substation operation.

Khator and Leung (1997) reviewed the models related to the planning of substations and/or distribution feeders under two major groups: planning under normal conditions and planning for emergency, and found the power distribution planning as difficult to ensure substation capacity (transformer capacity) and feeder capacity (distribution capacity) to meet the load demands.

A new methodology was developed by Monteiro et al. (2005) for automated route selection for the construction of new power lines based on GIS considering environmental constraints, operation and maintenance, and equipment installation costs associated with the slope of the terrain crossed by the power lines.

Jewell, Grossardt and Bailey (2006) developed a new method to reduce public opposition to new lines by way of public participation in transmission line routing decisions and hence the time needed for the approval of new line construction.

A fast algorithm was presented by Tram and Wall (1988) to help select the proper conductors for feeder expansion plans including selection of optimal conductor type for each feeder segment to maintain an acceptable voltage profile along the entire feeder and minimize the capital investment and the cost of feeder losses.

Urban, Bandars and Moll (2009) developed six different scenarios of rural electrification for the period 2005–2030 using a regional energy model (REM) to assess the effect of greenhouse gases (GHGs), primary energy use, and the costs, and compared the business-as-usual (BAU) scenario with different rural electrification scenarios based on electricity from renewable, diesel, and the grid. The results indicated that rural electrification with renewable energy tends to be the most cost-effective option.

Kaundinya, Balachandra and Ravindranath (2009) studied the modeling and analysis of economic, environmental, and technical feasibilities of both grid-connected and standalone systems as decentralized power options, which was restricted to annualized life cycle cost (ALCC) methods.

The literature reveals that no work has been done on the design of mini-grid in India for connecting all the cluster SHPs and other renewable energy power projects together as well as with the main grid, wherever available, to electrify the remote rural areas not connected with the utility grid. Therefore, the design of the mini-grid for nine SHPs of Bageshwar District has been undertaken and the results are presented in this paper.

2.2 Reviews on Statistical Quality Control and Time Series

2.2 Statistical Quality Control

2.2.1 Meaning of Statistical Quality Control

Statistical quality can be said to be a very special technique; it's a literature of its own. To know what is meant by statistical quality control, the definition of terms like quality and control cannot be over-emphasized.

Quality in business, engineering, and manufacturing has an interpretation as non-inferiority or superiority of something. The quality of a product or service refers to perception of degree to consumer's expectation which also comprises many quality characteristics like strength, colour, weight, conformance, etc.

Also, Juran (1974) defined quality in his book *Quality Control Handbook* as fitness for use in all sectors of human institutions such as industries, schools, hospitals, companies, and government agencies engaged in providing products or services for human beings. This includes useful delivery date and fitness for use in general. If the goods and services respond to the overall needs of users, they are said to be "stability."

Control means exercising, restraining, or directing influence over some factors of operation, producing, and packaging in an industrial setup.

Karmel also defines statistical quality control in applied statistics as an important application of the theory of sampling and significance in the industrial field in order to reduce variation in production processes.

Finally, statistical quality control can be defined as a very special statistical technique employed by industries to find out conformity of the product to specification.

2.2.2 Need for Quality Control

The needs for statistical quality are as follows:

i. Difference in the abilities of workers;

ii. Difference of raw material;

iii. Difference among machines;

iv. Wear and tear of machines; and

v. Level of commitment of labour according to supervision and welfare scheme.

2.2.3 Importance of Quality Control

Statistical quality control is highly important in checking the production process and directing production for the following reasons:

a) Cost of productionis lowered due to smooth running b) Profit maximization is feasible through smooth production at low cost c) Improvement of workers relationship is enhanced accordingly d) Marketing prices are according lowered due to lower production cost e) Limitation of products becomes almost impossible because good standard has been maintained all along f) It eliminates unnecessary waste e.g. discarding some finished in products which are bad after all g) Optimum sales are achieved because of the low cost and good production

2.2.4 Causes of Variation

It is possible that two identical products produced one after the other, on the same machine, under the same condition, may still look different. This may be due to chance variation or as a result of assignable variation.

2.2.5 Chance/Random Causes of Variation

This is otherwise known as random variation. It is an inherent and inevitable variation in any production process. Chance variation cannot be controlled or eliminated. A production process is in a state of statistical control if it is confined to chance variation only.

2.2.6 Assignable Causes of Variation

Processes that are out of control are operating in the presence of *Assignable Causes of Variation*. When the variability present in a production process is confined to chance variation, the process is said to be in a state of manageable statistical control. However, when the variability is not due to chance (random) but due to some specific and identifiable causes—which are not normally distributed but sporadically distributed—it is known as an assignable cause of variation.

One or more of the following may cause the variation:

i. Poorly trained operator;

ii. Raw material of inferior quality;

iii. Raw material from different batches and of varying quality;

iv. Faulty machine settings; and

v. Faulty machine due to worn-out parts.

It is rare that a production system will be free from all the assignable variations listed above. For this reason, a systematic method of detecting serious deviation from the state of statistical control is needed. Control charts are designed for this purpose.

2.2.7 Acceptance Sampling

Sampling plans are required in two distinct situations. The first situation occurs when the operation is in progress and the producer is checking the control situation to ensure conformance of the products to specifications. For this situation, control charts are used as effective tools. The inspection conducted here is called *process inspection.*

The second situation arises when the manufacturer receives goods to be used in his own production process, or when a consumer or distributor requires assurance that the goods conform to specifications before usage or resale. The inspection conducted at this level is called *acceptance inspection.* Since goods typically arrive in containers, bottles, boxes, or barrels, acceptance sampling becomes desirable in such cases.

It is important to note that acceptance sampling is not an attempt to control quality, but rather a decision-making tool used to accept or reject lots.

2.3 Review on Time Series

Time series analysis is an important area of study with significant applications in industry and government. This is not surprising, as organizations in Nigeria and around the world rely on forecasting to ensure the smooth running of their operations.

Many statisticians and authors have contributed valuable works in this area through books, journals, and articles, which have greatly informed research such as this study.

Below are some notable definitions and perspectives on time series:

i. Murray R. Spiegel (1992, *Metric Edition*, *Theory and Problems of Statistics*) described time series as "a set of observations taken at specific times, usually at equal intervals."

ii. Biyi Afonja (1985, *Introduction to Statistics*) defined time series as "the study of methods of collecting and analyzing data in such a way as to minimize uncertainty in conclusions drawn from the data, while being able to assess the degree of such uncertainty."

iii. Leonard J. Kazmier (*Theory and Problems of Business Statistics*) defined time series as "a set of observed values, such as production or sales data, recorded for a sequentially ordered series of time periods."

iv. C. P. Gupta (*Introduction to Statistical Methods*) defined time series as "a set of data pertaining to the values of a comparable variable at different times."

v. Robert D. Mason and Douglas A. Lind (1990, *Statistical Techniques in Business and Economics*) defined time series as "the collection of data recorded over a period of time, usually on a weekly, monthly, quarterly, or yearly basis."

2.3.1 Types of Time Series Data

There are four distinct types of time series data, namely:

i. **Economic Time Series** – These include data related to economic activities such as income, employment, price indices, and gross domestic product (GDP).

ii. **Physical Time Series** – These represent natural or physical measurements such as rainfall, temperature, river flow, or agricultural yields.

iii. **Marketing Time Series** – These consist of data related to sales, advertising expenditure, market demand, and consumer preferences.

iv. **Demographic Time Series** – These include population-related data such as birth rates, death rates, migration, and age distribution over time.

2.3.2 Time Plot

A time plot is the observation of a time series of a variable, where the values are plotted on the vertical (Y-axis) against time on the horizontal (X-axis).

The time plot clearly reveals important factors of the series such as trend, seasonality, and discontinuity. It is also useful for investigating past behavior and forecasting future performance of a business organization. In other words, past information is recorded, analyzed, and projected into the future to aid production planning.

Furthermore, a time plot indicates the presence of turning points, which signify changes in trend. These turning points reflect upward and downward movements that transform the series over time.

2.3.3 Components of a Time Series

A time series is generally composed of four main components:

i. **Trend:** The long-term movement or direction of the data over an extended period, which may be upward, downward, or stable.

ii. **Seasonal (Regular) Variation:** Short-term, regular, and predictable patterns that repeat at fixed intervals, such as monthly, quarterly, or yearly.

iii. **Cyclical Variation:** Long-term oscillations that occur over periods longer than one year, often linked to business or economic cycles.

iv. **Irregular Variation:** Random, unpredictable fluctuations caused by unusual or unforeseen events, such as strikes, natural disasters, or sudden market shocks.

2.3.4 Time Series Model

In dealing with time series, there are two main types of models, namely:

1. **Additive Model:** Also called the independent model. It assumes that the values of observed data are the sum of its four components, expressed as:

$$Y_t = T_t + S_t + C_t + I_t$$

   where Y_t = observed value, T_t = trend, S_t = seasonal component, C_t = cyclical component, and I_t = irregular variation.

2. **Multiplicative Model:** Also called the dependent model. It assumes that the observed data is the product of its four components, expressed as:

$$Y_t = T_t \times S_t \times C_t \times I_t$$

   where Y_t = observed value, T_t = trend, S_t = seasonal component, C_t = cyclical component, and I_t = irregular variation.

Both models illustrate the relationship between the trend, seasonal, cyclical, and irregular variations in a time series.

Chapter 3: Research Methodology

3.0 INTRODUCTION

This study adopts a quantitative, analytical research design focusing on historical power transmission data. The objective is to identify trends, detect anomalies and assess system reliability using time series analysis and quality control techniques (Box et al., 2016; Montgomery, 2020). Historical data on voltage levels, power loads, line losses, and outage records were obtained from the daily output load on each turbines in phase 1 and phase 2. The data were first cleaned by handling missing values through interpolation, removing corrupted entries, and detecting outliers using statistical thresholds (Hyndman & Athanasopoulos, 2018). Time indexing was applied to ensure chronological order, and normalization was used to maintain consistency across all variables.

Time series analysis was employed to identify long-term trends, detect seasonal patterns, and forecast future performance using models such ARIMA, and exponential smoothing. In parallel, statistical quality control methods including X-bar and R control charts as well as process capability analysis were used to monitor system stability and ensure that transmission parameters remained within specified limits (Montgomery, 2020).

3.1 Data Collection

The research utilizes secondary data, specifically daily power transmission records obtained from the relevant transmission system operator or energy authority. Data covering at least 30 days will be gathered to capture trends, seasonal variations, and stochastic behaviour (Box et al., 2016).

Prior to analysis, the dataset will undergo screening and preprocessing, which includes:

i.Identifying and treating missing values using interpolation or imoutation methods.

ii.Removing outliers or anomalies using statistical detection techniques.

iii.Verifying data consistency and accuracy by cross checking with operational logs.

3.2 Sources of Data

Historical research data may be obtained from either primary or secondary sources. For the purpose of this study, the data were collected from the Transmission Company of Nigeria (TCN), Omotosho chapter, Ondo State.

3.3 Data Type

The dataset obtained from the Transmission Company of Nigeria (TCN), Omotosho chapter, consists of daily load values of each generating turbine for Phase 1 and Phase 2. The data were provided in Microsoft Excel format for ease of preprocessing and analysis.

3.4 Analytical Methods

3.4.1 Time Series Analysis

Time series methods such as Autoregressive (AR), Moving Average (MA), and Autoregressive Integrated Moving Average (ARIMA) were applied to model the transmission data **hyndman2018forecasting**.

Model Selection: Competing models were compared using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to ensure parsimony and accuracy.

Trend Analysis: Moving averages and decomposition methods were used to detect long-term load behaviour.

Seasonality Detection: Autocorrelation functions (ACF) and spectral analysis were employed to identify recurring patterns.

Forecasting: ARIMA models were implemented to predict future daily load levels.

Parameter Estimation: The parameters of the ARIMA model were estimated following a structured procedure:

1. **Differencing to achieve stationarity:** If $d > 0$, the original series Y_t is differenced d times to obtain a stationary series:

$$W_t = \nabla^d Y_t$$

For example, - First difference: $W_t = Y_t - Y_{t-1}$ - Second difference: $W_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$

2. **ARMA(p, q) structure on differenced data:** Once differenced, the ARIMA model is just an ARMA model on W_t:

$$W_t = c + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \ldots + \phi_p W_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

where: c = intercept, ϕ_i = autoregressive (AR) coefficients, θ_j = moving average (MA) coefficients, ε_t = white noise error with variance σ^2.

3. **Compact Operator Notation:** Using the lag operator L (where $LY_t = Y_{t-1}$):

$$\phi(L)W_t = c + \theta(L)\varepsilon_t$$

with

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p, \quad \theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$$

Hence, the full ARIMA model can be expressed as:

$$\phi(L)(1-L)^d Y_t = c + \theta(L)\varepsilon_t$$

Model Selection and Evaluation: To determine the most parsimonious model, information-theoretic criteria such as AIC and BIC were employed:

$$AIC = 2k - 2\ln(\hat{L}), \qquad BIC = \ln(n)k - 2\ln(\hat{L})$$

where: k = number of parameters, $\hat{L}$ = maximum likelihood, n = sample size.

Interpretation: A model with lower AIC or BIC values is preferred. While AIC is more flexible, BIC applies a stricter penalty on complex models, particularly for large sample sizes. By applying both AIC and BIC, this study ensures that the selected model achieves accurate forecasting while avoiding overfitting.

3.4.2 Quality Control Techniques

To ensure reliable power transmission, Statistical Quality Control (SQC) techniques were applied. These methods help monitor system stability, detect abnormal variations, and maintain operational standards **montgomery2020introduction**.

Control Charts (X-bar, R): Shewhart control charts were constructed to establish Upper Control Limits (UCL) and Lower Control Limits (LCL), in line with standard statistical process control principles. These charts monitored transmission stability over time and signaled deviations beyond acceptable bounds.

Process Capability Analysis: The capability of the transmission process was assessed to verify whether the load parameters consistently remained within operational limits.

Statistical Process Control (SPC): Abnormal deviations in system behaviour were detected, providing early warnings of potential faults.

Process Monitoring: Variations were analyzed to distinguish between:

i. **Common causes** – systematic and inherent to the process.

ii. **Special causes** – unexpected disturbances indicating faults.

Operational Significance: Control limits provided benchmarks for early detection of transmission issues, enabling corrective action before severe faults occurred.

Control Chart Design Control charts combine graphical and numerical descriptions of data using sampling distributions. They act as a signal system to identify and correct disturbances when a process deviates from stability.

Moore **moore2009statistics** defines a control chart as a tool that monitors a process and alerts when the process has shifted, moving it out of control. Since variation is unavoidable, Dr. W. H. Shewhart established the theory that variation can be divided into:

i. Random variation (inherent, unavoidable).

ii. Assignable causes (specific, correctable).

If variation is due to assignable causes, the process is out of control; otherwise, it remains stable.

article enumitem

A Shewhart control chart is characterized by three horizontal lines:

i. Central control limit (CL) – indicating the desired process standard.

ii. Upper control limit (UCL).

iii. Lower control limit (LCL).

The UCL and LCL, often called *warning limits*, represent thresholds that a process must not exceed to remain acceptable.

3.5 Validation and Justification of Methodology

Validation of results was carried out through back-testing of forecast models against historical data. Performance was evaluated using accuracy metrics such as:

i. Mean Absolute Percentage Error (MAPE).

ii. Root Mean Square Error (RMSE) **shumway2017time**.

Additionally, expert reviews by power system engineers were sought to ensure that analytical findings aligned with practical operational realities.
Ethical considerations were observed by:

i. Maintaining data confidentiality.

ii. Using the data strictly for academic and analytical purposes.

Justification: The integration of time series forecasting and quality control techniques provides a comprehensive framework for monitoring, predicting, and improving power transmission reliability.
article enumitem

Tools and Software:

i. **Microsoft Excel** – used for preliminary data organization and visualization.

ii. **R Programming Language** – applied for statistical modeling, visualization, and computation.

3.6 Typical Control Chart

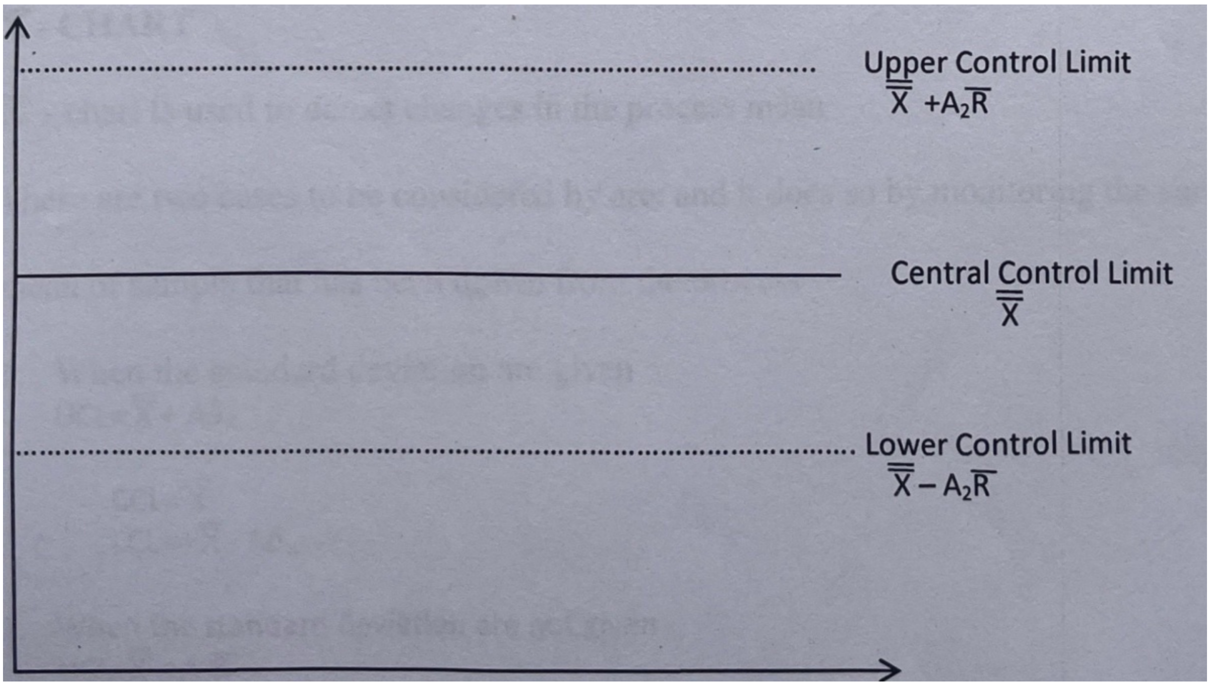

Figure 3.1: : X-BAR CHART

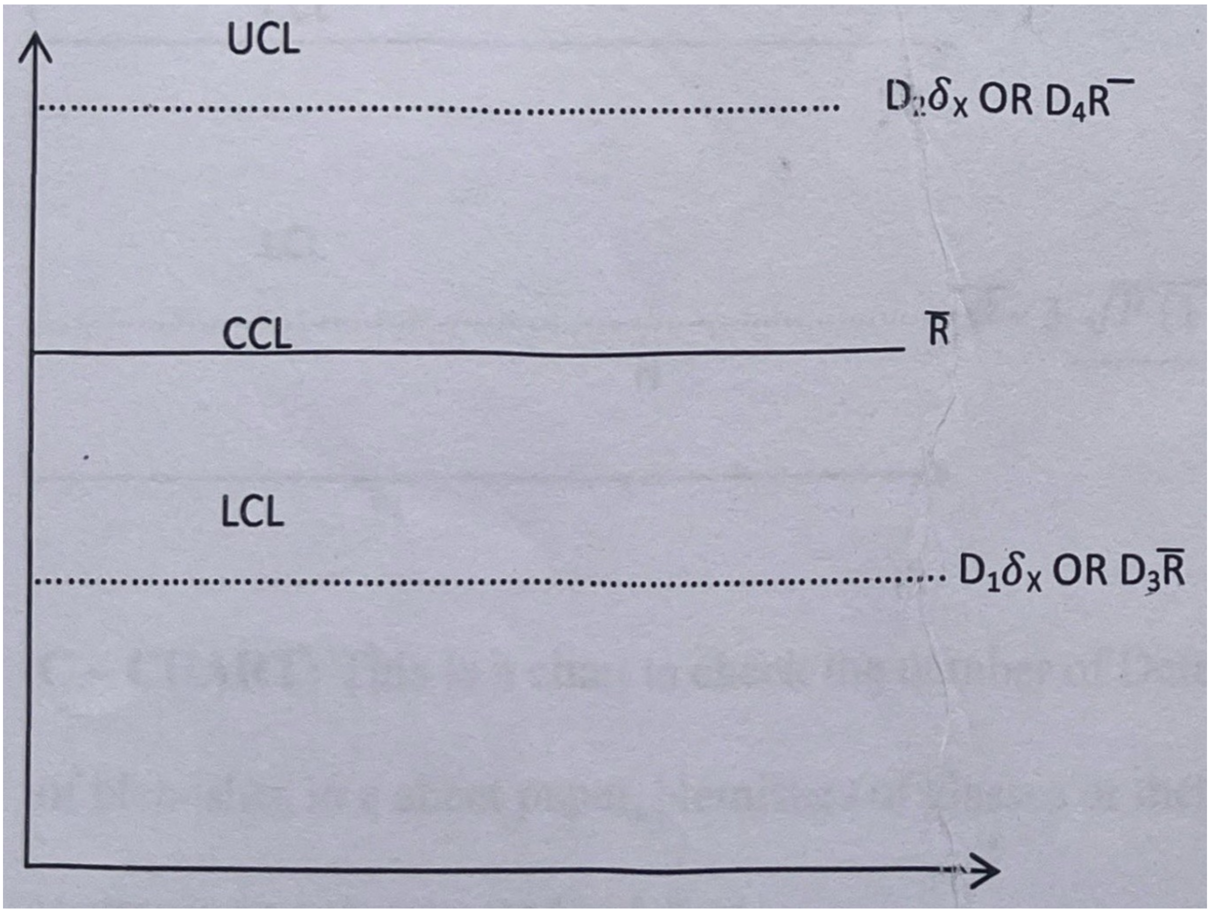

Figure 3.2: : R-BAR CHART

Chapter 4: Results and Discussion

4.1 Introduction

This chapter provides a comprehensive analysis of the power generation data collected across two operational phases, each involving multiple gas turbines. Phase One consists of eight turbines (GT1–GT8), while Phase Two focuses on four turbines (GT1–GT4). The analysis evaluates performance, efficiency, and operational stability across distinct observation periods, which may differ due to time, environmental conditions, maintenance schedules, or variations in operational load.

The analysis is structured to identify underlying trends, fluctuations, and quality concerns in the power generation process. Time series methods are employed to detect temporal patterns, seasonal effects, and long-term trends in the data for each turbine across both phases. Descriptive statistics—including mean, median, standard deviation, skewness, and kurtosis—are computed to summarize the central tendency, dispersion, and distributional characteristics of power outputs. These measures provide insights into the consistency and reliability of turbine performance.

In addition, quality control tools such as control charts are applied to assess process stability and detect assignable variations or anomalies. These charts help determine whether the generation process is statistically under control and guide corrective actions where necessary. Graphical approaches, including line plots and time series decomposition, further assist in visualizing operational behaviour over time.

The chapter is divided into two main sections corresponding to the two phases of data collection. Each section provides:

1. A descriptive statistical profile of turbine performance.

2. A time series examination of operational patterns.

3. A quality control assessment to evaluate process stability.

By segmenting the analysis into phases, differences in performance across turbines and operational periods are highlighted. This structured approach ensures a comprehensive understanding of power generation trends, facilitating informed decision-making in turbine maintenance and reliability management.

4.1.1 Descriptive Statistics

This section presents descriptive statistics for each phase of the data collection. The objective is to independently summarize the characteristics of the datasets and assess their variability, central tendencies, and distributional features. Analyzing each dataset separately enables the identification of unique behaviours, operational anomalies, and performance variations prior to conducting comparative or combined analyses.

Key statistical measures reported include:

i. **Mean and Median:** Indicators of central tendency in power output.

ii. **Standard Deviation:** Measure of variability in turbine performance.

iii. **Skewness:** Assessment of asymmetry in the power output distribution.

iv. **Kurtosis:** Evaluation of the peakedness and presence of outliers.

These metrics provide a foundation for subsequent time series and quality control analysis, ensuring that both systematic patterns and unusual variations are properly identified and interpreted.

Table 4.1: **Summary Statistics for Phase 1**

GT	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
1	30.07	3.89	43.5	64.5	1.13	2.55
2	30.77	3.58	25.0	36.6	0.07	-1.00
3	30.22	3.28	25.0	35.6	0.20	-0.86
4	30.02	2.64	20.5	36.0	-1.34	4.75
5	30.25	4.02	25.2	35.6	0.10	-1.51
6	28.35	2.95	25.1	33.4	0.15	-1.63
7	30.82	3.93	20.5	35.5	-0.69	-0.40
8	30.84	3.61	25.3	36.8	0.10	-1.01

This table summarizes descriptive statistics for Phase 1 across eight GT variables. It highlights central tendencies, variability, and distribution shapes.

Means range from 28.35 (GT 6) to 30.84 (GT 8), indicating a relatively stable central trend. Standard deviations range from 2.64 (GT 4) to 4.02 (GT 5), showing moderate variability. GT 1 has an unusually high maximum (64.5), suggesting a potential anomaly. Skewness and kurtosis emphasize asymmetry and peakedness in GT 1 and GT 4.

Table 4.2: **Summary Statistics for Phase 2**

GT	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
1	92.90	3.48	90.0	102.4	1.76	1.86
2	93.13	9.84	68.8	104.7	-0.86	-0.24
3	91.09	7.63	77.9	110.5	0.46	-0.13
4	86.73	11.74	64.5	108.1	0.00	-0.91

Phase 2 includes four GT variables analyzed with key descriptive statistics.

GT 1: mean 92.90, low variability (std. dev. 3.48), positive skew (1.76), leptokurtic (1.86).

GT 2: mean 93.13, higher variability (9.84), negative skew (-0.86), slightly platykurtic (-0.24).

GT 3: mean 91.09, moderate variability (7.63), mild positive skew (0.46), near-normal kurtosis (-0.13).

GT 4: mean 86.73, highest variability (11.74), symmetric (skew 0.00), flat-topped (kurtosis -0.91).

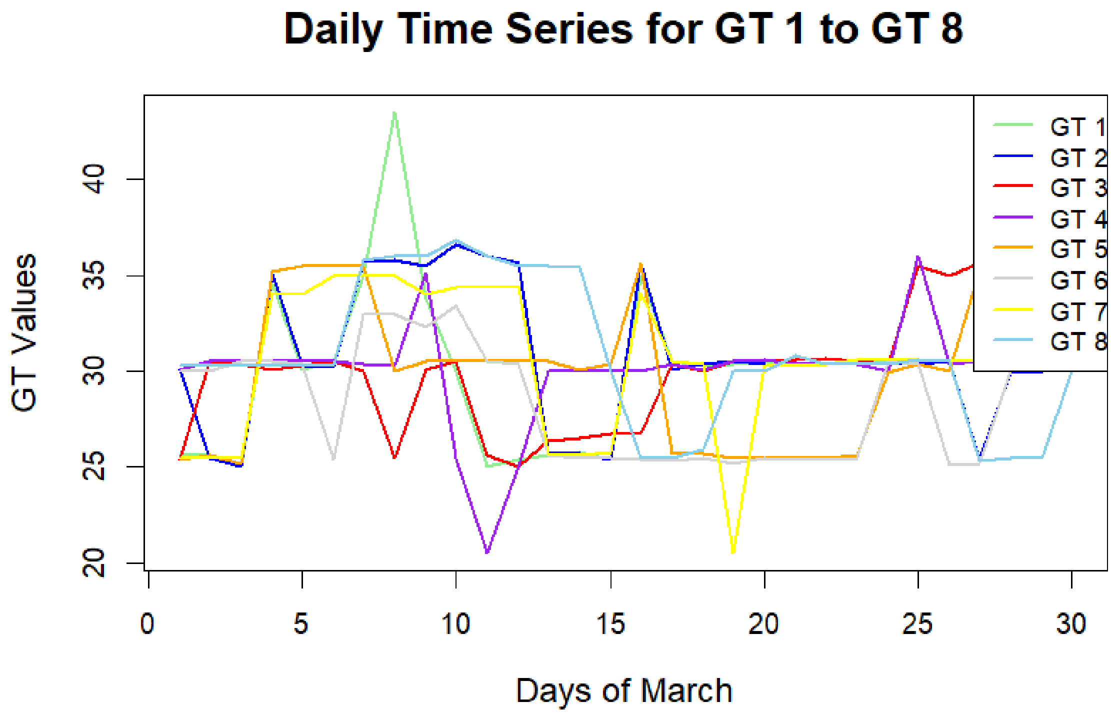

Figure 4.1: The line chart for Phase 1

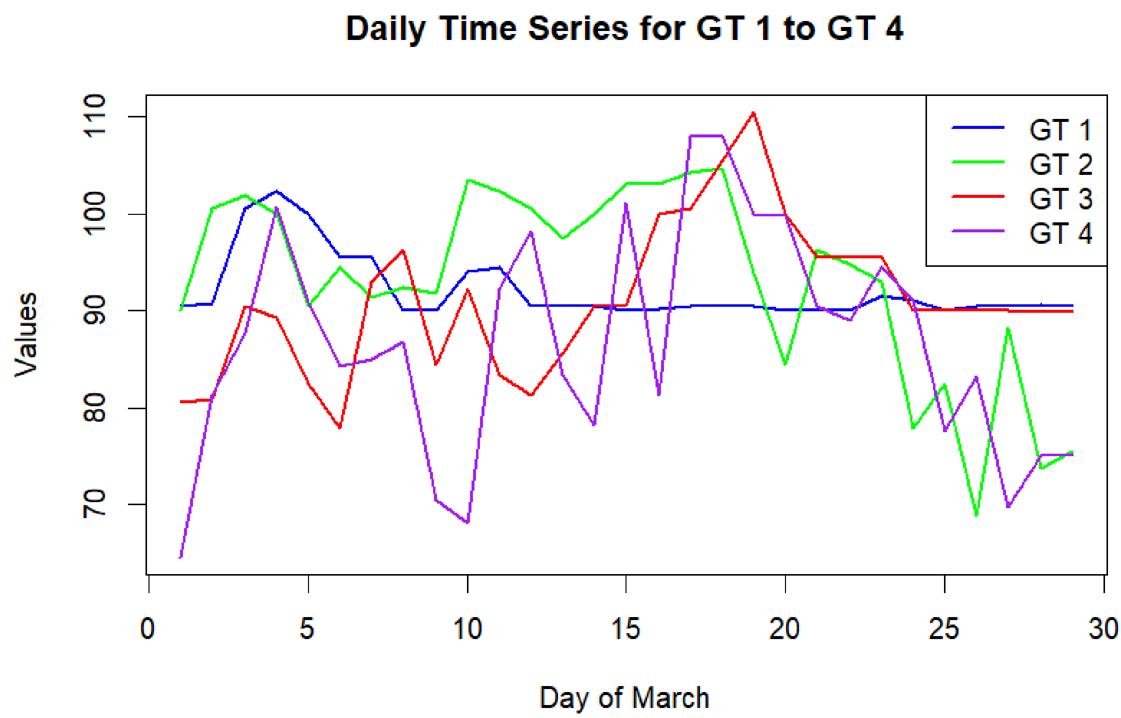

Figure 4.2: The line chart for Phase 2

The time series plot provides a comparative view of the daily values of GT 1 through GT 8 for the month of March. Each of the eight GT variables displays a unique pattern of fluctuation over the 30-day period, offering insight into the temporal dynamics and relationships between them. Among the variables, GT 1 stands out significantly due to a pronounced spike around the 8th day of March, where its value surged well above those of the other series, briefly exceeding 40. This sharp peak, which is not mirrored by the other GT variables, suggests either an isolated event or an anomaly affecting only GT 1. Outside of this outlier, GT 1's movement remains relatively moderate, aligning with the general flow of other series after mid-March. In contrast, GT 4 and GT 7 are characterized by high volatility throughout the month. GT 4 shows an abrupt dip around the 11th day, plunging below the 20-mark, which is the lowest point across all series. Similarly, GT 7 exhibits erratic swings, particularly a sharp drop between Days 15 and 18, followed by an immediate rebound. These irregular fluctuations may indicate external disruptions, sensitivity to certain conditions, or unstable underlying trends in these specific variables. GT 2, GT 5, and GT 8 demonstrate relatively stable and consistent behavior. Their values generally hover within a tight range of approximately 28 to 36, showing less pronounced peaks or troughs. GT 2, in particular, maintains a smooth trajectory, even during periods where other variables experience more turbulence. This steadiness suggests a level of resilience or insulation from the volatility affecting other GTs. Interestingly, a number of the GT series— notably GT 3, GT 4, GT 6, and GT 7—display a coordinated dip around the midpoint of the month, between Days 13 and 16. This simultaneous downward trend may suggest a common external factor influencing multiple variables at once. However, the degree of decline varies, with GT 4 and GT 7 experiencing the steepest drops. As the month progresses, particularly from Day 20 onward, a convergence of all eight GT series becomes apparent. During this final third of the month, the values stabilize and align more closely, indicating a period of general calm across all variables. The synchrony in their patterns during this phase may suggest the emergence of shared structural behaviour or the diminishing influence of earlier disturbances. Overall, the plot underscores the heterogeneity in the temporal behaviour of the GT variables. While some variables like GT 2 and GT 8 are relatively stable and may lend themselves more easily to forecasting models, others like GT 4 and GT 7 are far more erratic and may require transformation or closer examination before inclusion in predictive analysis. The outlier behaviour observed in GT 1 also calls for careful treatment to prevent distortion in any statistical modelling. This time series comparison provides a foundation for assessing the dynamics of each GT variable, setting the stage for stationarity testing, transformation, and eventual forecasting. Additionally, it highlights the importance of understanding not just individual variable behaviour, but also how these variables relate to each other—temporally synchronizing at times, diverging at others—and how this interplay may inform further analysis.

For Figure 4.2: Line chart for Phase 2 Overall, the period is characterized by significant volatility and divergence among the four entities. GT 1 (blue) largely maintained a relatively stable performance, primarily oscillating between 90 and 100 before settling around 90 towards the end of the month. GT 2 (green) displayed the most erratic behavior, with sharp peaks above 100 and notable dips below 70, particularly in the latter half of the month. GT 3 (red) exhibited a varied trend, starting around 80, peaking near 110 mid-month, and then stabilizing around 90 by month-end. GT 4 (purple) also showed considerable fluctuation, with initial strong performance above 100, followed by significant troughs below 70, before recovering somewhat by the 28th day. The first half of March saw GT 1 and GT 2 generally performing higher, often exceeding 90 and occasionally 100. In contrast, the latter half, particularly from day 20 onwards, witnessed a general decline or increased instability for GT 2 and GT 4, while GT 1 and GT 3 showed greater resilience and stabilization around the 90 mark. The period around day 25 highlights a shared dip across several GTs, most pronounced for GT 2 and GT 4.

4.2 Modelling the Time Series Forecast for Phase 1

To prepare the time series data for forecasting, stationarity was first assessed using the Augmented Dickey-Fuller (ADF) test. The ADF test was applied individually to each of the eight GT time series (GT 1 through GT 8) to evaluate whether their statistical properties, such as mean and variance, remain constant over time.

The results indicated that all series had p-values greater than the 0.05 significance level. For instance, GT 1 had a p-value of 0.4447, GT 2 recorded 0.419, and GT 3 showed 0.7672, with the remaining series displaying similarly high p-values. These outcomes provide insufficient evidence to reject the null hypothesis of non-stationarity. In other words, all eight GT series are non-stationary in their current form.

Given this finding, it was necessary to apply appropriate transformations to stabilize the series before forecasting. The `auto.arima()` function was employed, as it can automatically determine the required differencing and other transformations needed to make the series stationary and suitable for predictive modeling.

Since the p-values exceeded 0.05, the null hypothesis of non-stationarity was not rejected, confirming that the time series contain trends or other forms of non-stationary behavior. Addressing this non-stationarity is a critical step in ensuring reliable and accurate forecasts for the GT variables in

For the second phase, the `auto.arima` function was employed to identify the most appropriate models for each of the four series (GT 1 to GT 4). This automated approach was adopted because the original time series data were found to be non-stationary, meaning their statistical properties, such as mean and variance, change over time. `auto.arima` is particularly useful in such cases, as it automatically applies differencing to achieve stationarity before selecting the best-fitting ARIMA model based on model selection criteria like the Akaike Information Criterion (AIC).

For GT 1, the model selected by the algorithm was ARIMA(2,1,0). Although the specific values of the autoregressive (AR) coefficients were not provided in the output summary, the model achieved an AIC value of 115.0476, indicating a strong fit relative to more complex or less suitable alternatives.

In the case of GT 2, the best-fitting model was ARIMA(1,1,0), which also required first differencing to stabilize the series. The AIC for this model was 189.0873. While the AR(1) coefficient was not explicitly reported, the relatively high AIC compared to GT 1 suggests that GT 2 may have more noise or less predictable structure.

GT 3 was best modeled by a simple ARIMA(0,1,0) process, which corresponds to a random walk. This means the series is best described by differencing alone, without any autoregressive or moving average components. The AIC for this model was 178.8527, suggesting a moderately good fit given the simplicity of the model.

For GT 4, the model selected was ARIMA(1,0,0) with a non-zero mean, indicating that the series did not require differencing but did contain one autoregressive component. This model yielded the highest AIC value among the four series at 221.0612, implying that the model may not be as effective in capturing the structure of the data compared to the others.

Forecast for GT 1: The forecast for GT 1 (daily up to March 2025) shows a stationary trend, with no significant upward or downward movement expected. The forecasted values remain consistent around a central average (approximately 30), as reflected by the flat forecast line. The prediction intervals indicate a moderate level of uncertainty; however, the model suggests stability rather than growth or decline overall.

Forecast for GT 2: Similarly, the forecast for GT 2 demonstrates a stationary trend. Historical data show more volatility than GT 1, with several sharp increases and decreases. However, as the series approaches the forecast point, it stabilises near the 30 mark. The projected forecast line remains flat, indicating that future values are expected to remain stable without significant change. The prediction intervals reinforce this stability by showing a consistent range of likely values. Despite earlier fluctuations, the model anticipates a steady performance of GT 2 in the coming period.

Forecast for GT 3: The forecast for GT 3, projecting daily values up to March 2025, indicates a period of uncertainty with no distinct upward or downward trend. Historical data shows a relatively steady increase in values up to about time point 40, after which the forecast begins. The forecast line, represented by the solid blue line, remains flat at around the 35 mark, suggesting that the central prediction is for GT 3 to maintain its current level over the forecast period. However, the fan-shaped confidence intervals—shaded in light and dark grey for 95% and 80% prediction intervals respectively—expand widely over time, indicating increasing uncertainty in the forecast. Although the average forecast is stationary, the growing width of the confidence bands suggests that actual future values could vary significantly. By the end of the forecast period, the possible range spans from below 0 to above 100. Overall, the forecast suggests a stationary trend in the average projection, but with increasing uncertainty over time rather than a clear upward or downward trajectory.

Forecast for GT 4: The forecast analysis for GT 4, covering daily data up to March 2025, reveals a stationary trend. Historical data show moderate fluctuations, with values generally ranging between 21 and 41. The forecasted portion, represented by a flat blue line, suggests that future values are expected to remain around a consistent average of approximately 28. The accompanying prediction intervals indicate moderate uncertainty, but there is no indication of any significant upward or downward movement. Overall, the forecast implies a stable outlook with no strong directional trend.

Forecast for GT 5: The forecast for GT 5 begins at a value of 26, then rises briefly to 31 before stabilising at this level for the remainder of the forecast period. This pattern suggests a modest short-term upward trend followed by a plateau, indicating stability in values through to March 2025. The prediction implies minimal volatility, with the values expected to remain steady overall without significant fluctuations.

Overall, the forecast indicates a modest upward trend in the short term, which then levels off, reflecting stabilization in the values through to March. The initial rise from

26 to 31 could imply a short-term recovery or growth, while the subsequent flat line at 31 signals an expectation of consistency and minimal volatility in the forecasted period. This pattern is useful for planning and decision-making, as it suggests that while a slight increase is anticipated initially, the values are likely to remain steady, without significant fluctuations, for the rest of the year.

Forecast for GT 6: The projecting daily values up to March 2025, indicates a period of uncertainty with no distinct upward or downward trend. There was no real trend to conclude in March as the spread of the data was random. The forecasted portion, representing the blue region, remains flat at around the 50mark, suggesting that the central prediction is for GT 6 to maintain its current level over the forecast period. However, the fan-shaped confidence intervals shaded in light and dark grey for 60 percentage and 90 percentage prediction intervals respectively expand widely over time. This widening indicates increasing uncertainty in the forecast as time progresses.

Forecast for GT 7: For GT 7, initial observations in March revealed significant volatility, with performance fluctuating between approximately 20 and 35 units. However, from around day 50, a strong stabilization trend emerges. The central forecast (blue line) indicates GT 7's performance will consistently settle at approximately 30-31 units. The accompanying 80 percent confidence interval (darker blue) and 95 percent confidence interval (lighter grey) narrow considerably post-volatility, reinforcing the predictability. This suggests that while minor variations are possible, GT 7 is highly likely to maintain a stable performance around the 30-31 unit mark for the duration of the forecast period, through March 2025.

Forecast for GT 8: For GT 8, the initial observations in March show relatively stable performance for GT 8, hovering around 30-35 units. The central forecast (blue line) indicates that GT 8's performance is expected to remain consistent at approximately 30 units for the entire forecast period, up to March 2025. However, unlike GT 7, the confidence intervals for GT 8 show a significant divergence over time.

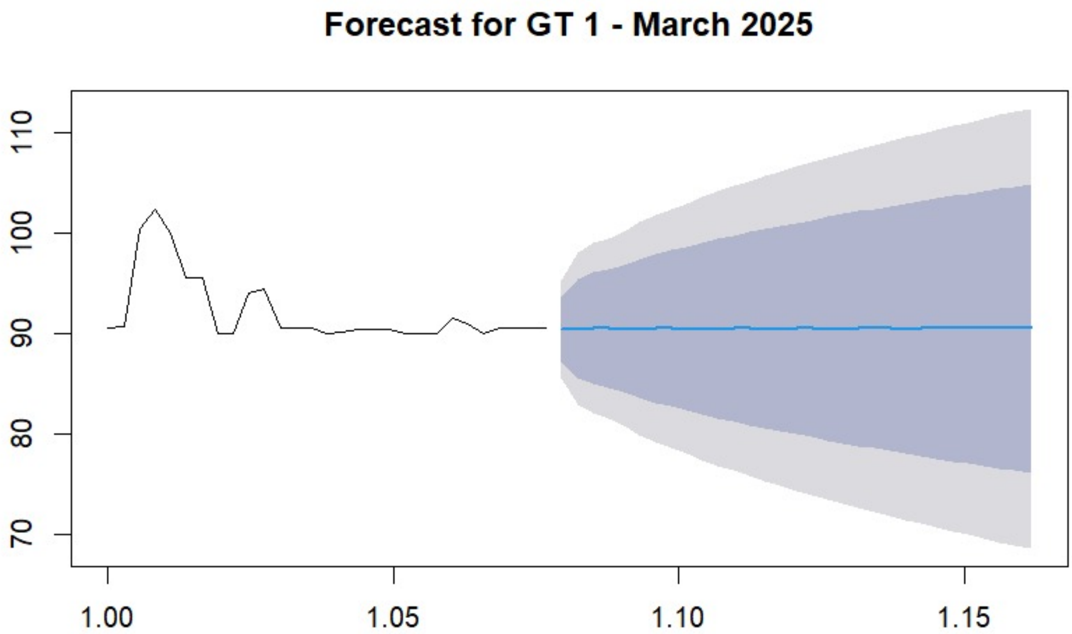

Figure 4.2.1: The line chart for Phase 1

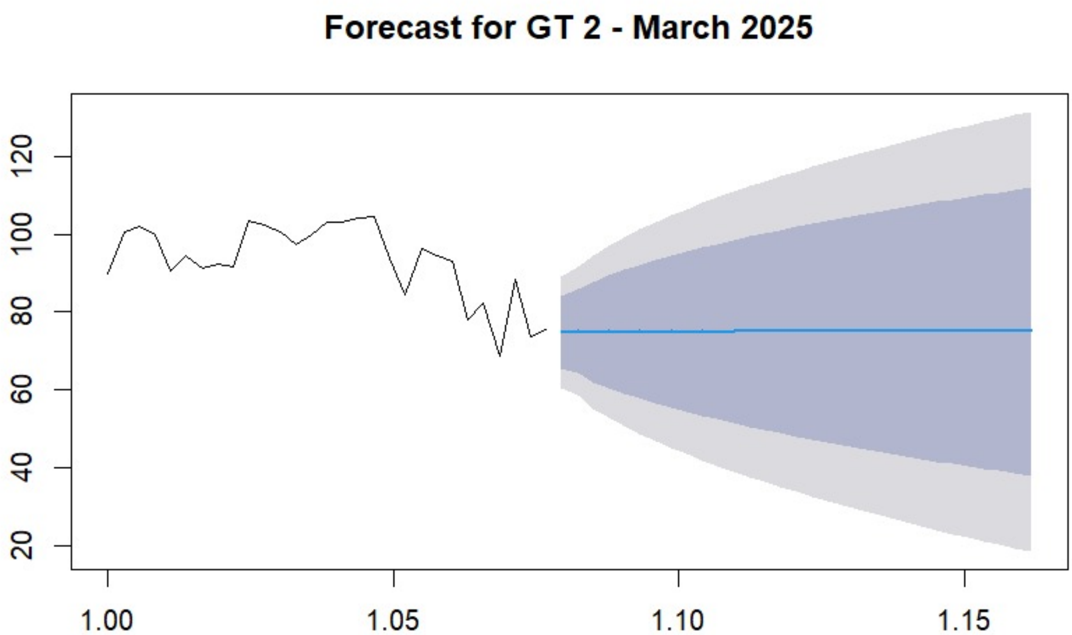

Figure 4.2.2: Forecast for GT 1 and GT 2 — Phase 1

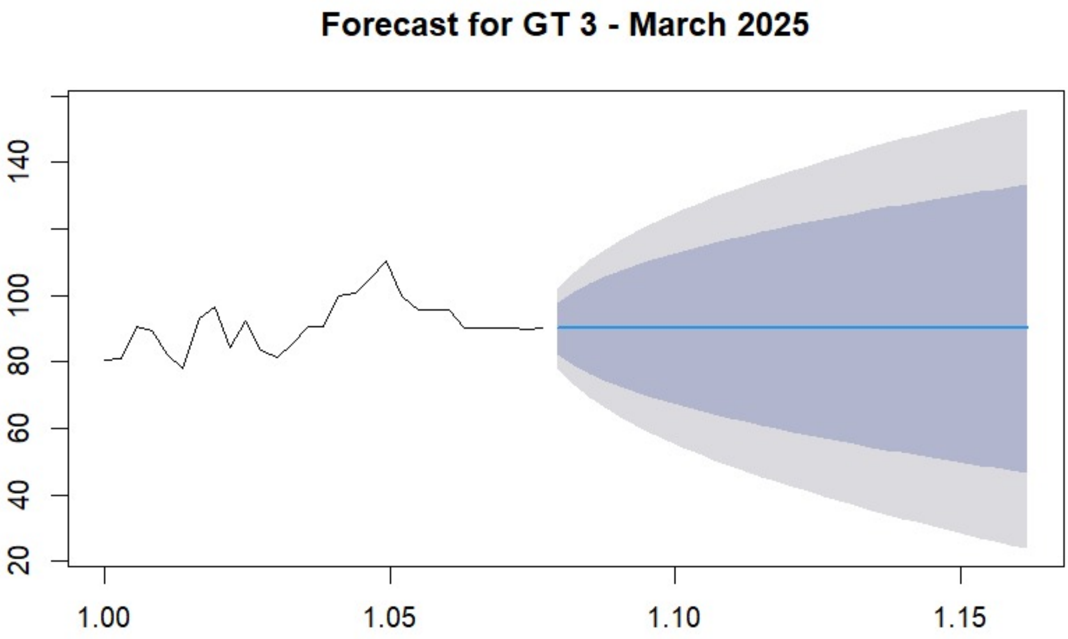

Figure 4.2.3: The line chart for Phase 1

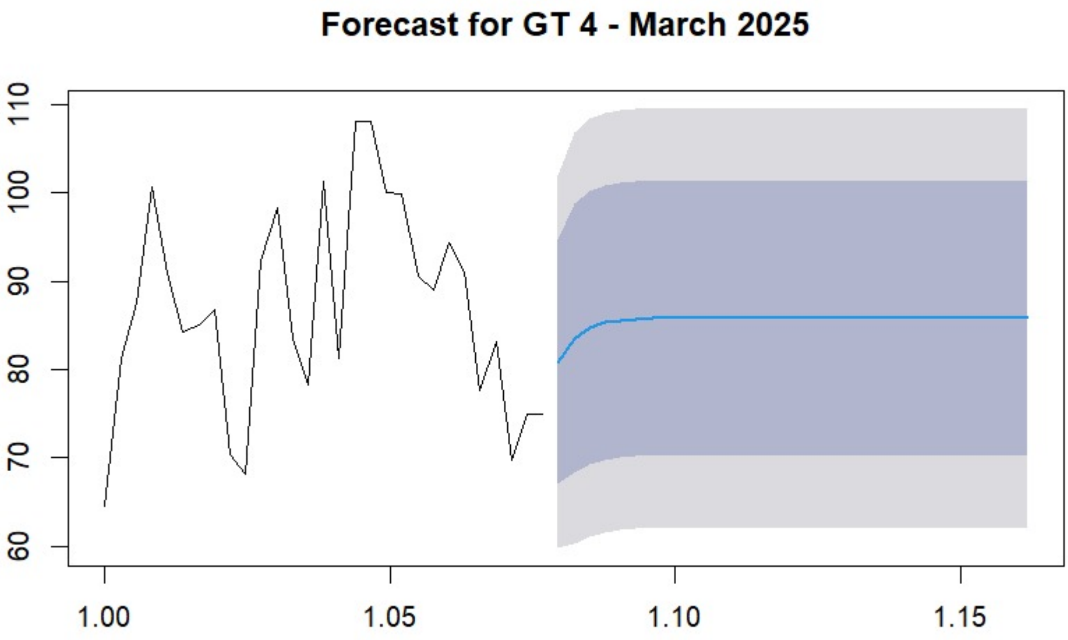

Figure 4.2.4: The line chart for Phase 1

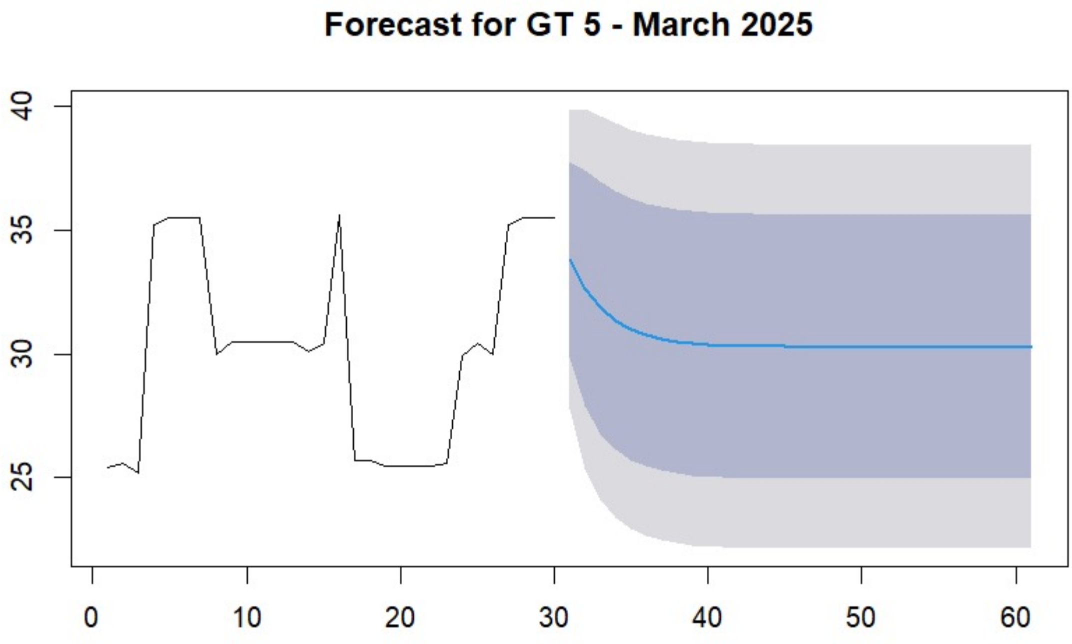

Figure 4.2.5: The line chart for Phase 1

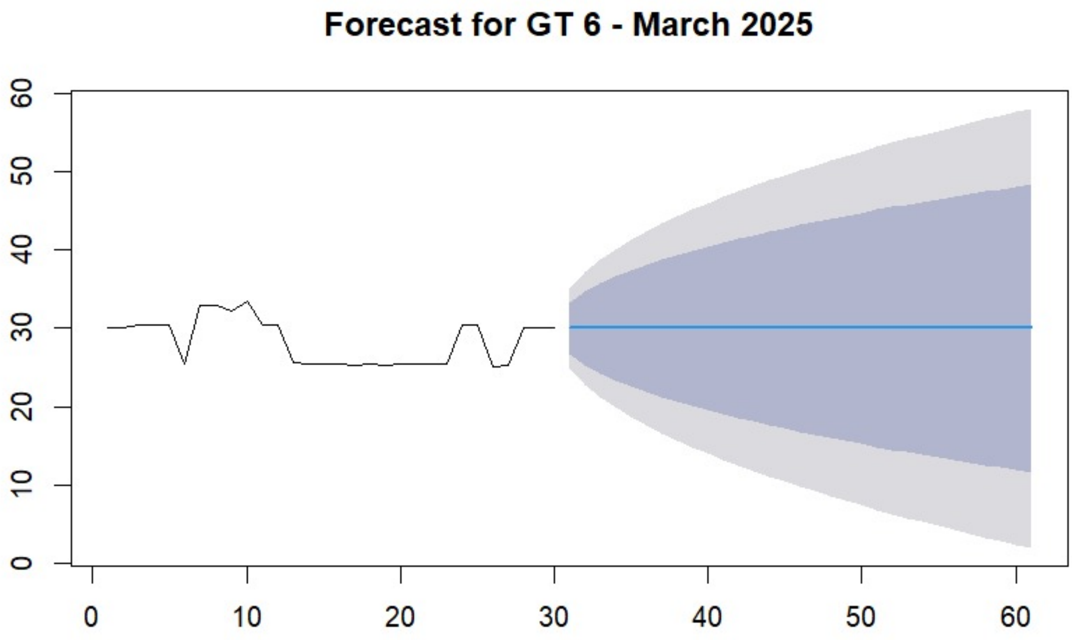

Figure 4.2.6: The line chart for Phase 1

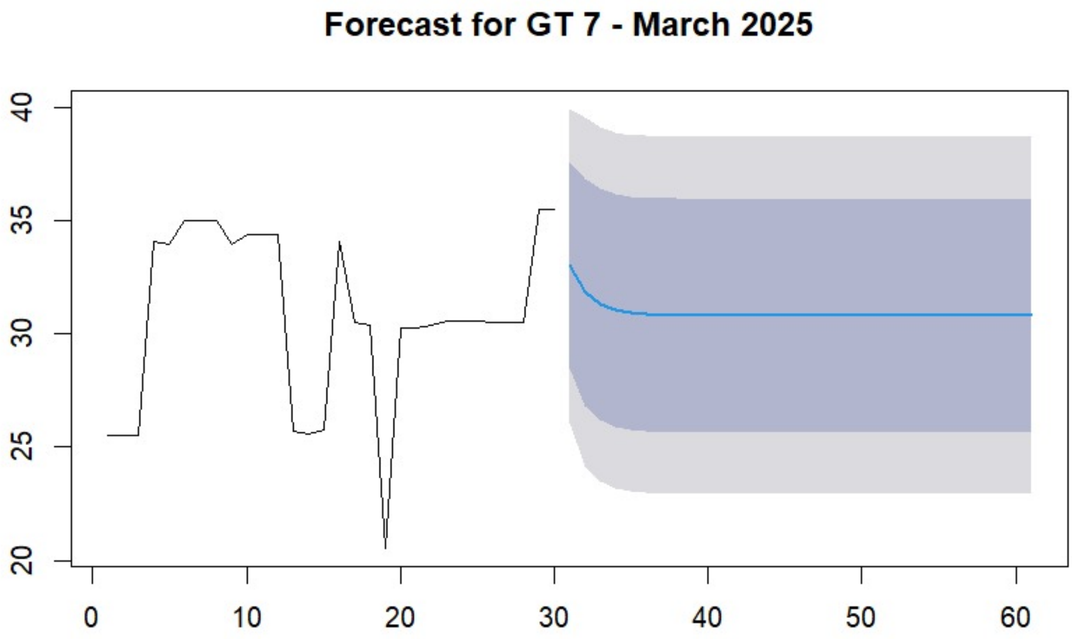

Figure 4.2.7: The line chart for Phase 1

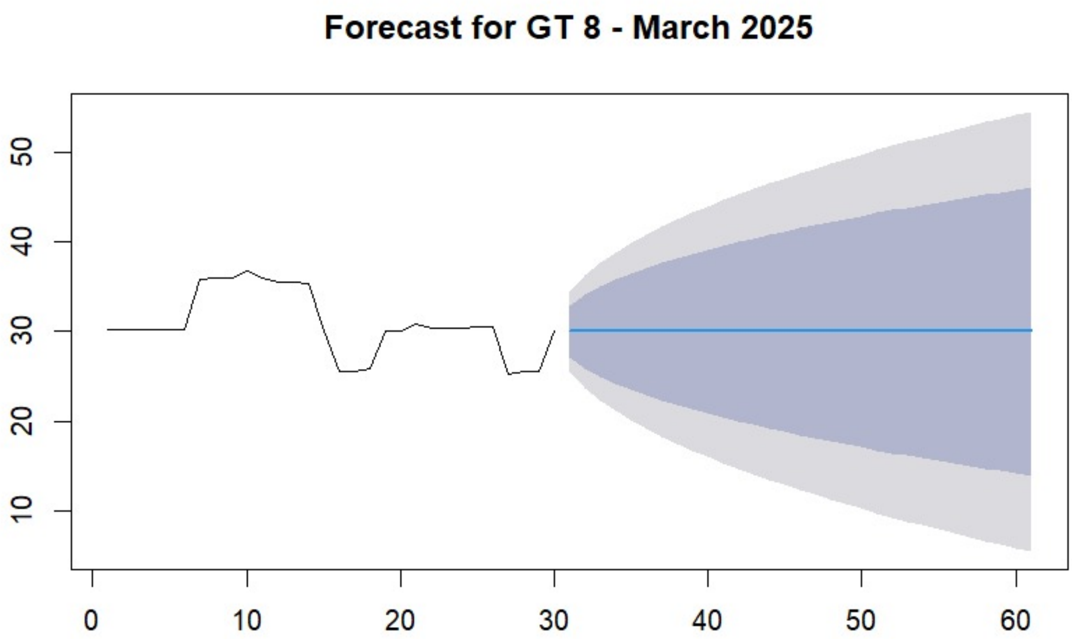

Figure 4.2.8: The line chart for Phase 1

Table 4.2.1: ARIMA Model Summary Table (auto.arima results) — Phase One

Series	ARIMA Model	Coefficient (AR/MA)	Mean	σ^2 (Residual Var.)	AIC	AICc	BIC
GT 1	ARIMA(1,0,0)	AR(1) = 0.4688	29.9609	12.16	164.27	165.19	168.47
GT 2	ARIMA(1,0,0)	AR(1) = 0.3758	30.7422	11.30	161.97	162.89	166.17
GT 3	ARIMA(0,1,0)	–	–	4.662	128.94	129.09	130.31
GT 4	ARIMA(0,0,1)	MA(1) = 0.4537	30.0223	5.874	142.42	143.34	146.62
GT 5	ARIMA(1,0,0)	AR(1) = 0.6726	30.2740	9.45	157.05	157.97	161.25
GT 6	ARIMA(0,1,0)	–	–	6.641	139.20	139.35	140.57
GT 7	ARIMA(1,0,0)	AR(1) = 0.4758	30.8049	12.54	165.19	166.11	169.39
GT 8	ARIMA(0,1,0)	–	–	5.048	131.25	131.40	132.62

Table 4.2.2: ARIMA Model Summary Table (auto.arima results) — Phase Two

Series	ARIMA Model	Coefficient (AR/MA)	Mean	σ^2 (Residual Var.)	AIC	AICc	BIC
GT 1	ARIMA(2,1,0)	MA(2) = 1	92.1034	4.9736	134.1540	136.7627	139.6232
GT 2	ARIMA(1,1,0)	AR(1) = 1, MA(1) = 1	93.1276	332.0178	194.0933	195.7599	198.0899
GT 3	ARIMA(0,1,0)	–	91.0897	258.3218	182.4598	182.9214	183.7920
GT 4	ARIMA(1,0,0)	MA(1) = 1	86.7345	268.3376	219.0870	220.0470	221.7514

4.3 Modelling the Time Series Forecast for Our Data — Phase Two

Forecast for GT 1: Looking back at the initial data, likely from March, GT 1 showed some moderate daily fluctuations, typically moving between 90 and 100 units before settling. As we project forward, the central forecast (represented by the blue line) indicates remarkable stability. GT 1's performance is consistently expected to hover around 90 units for the entire forecast period, suggesting a strong trend towards long-term average performance.

Even more broadly, the lighter-shaded area represents the 95% confidence interval, encompassing an even greater majority of expected daily results. This band will expand to a significant extent by the end of the year, potentially ranging from around 30 to 150 units. In essence, while GT 1 is predicted to maintain an average daily performance of 90, the increasing spread of these confidence bands highlights a growing uncertainty. This means that while 90 remains the most likely outcome, the further we look into the future towards March 2025, the greater the possibility that actual daily performance could deviate considerably, either above or below that 90-unit average.

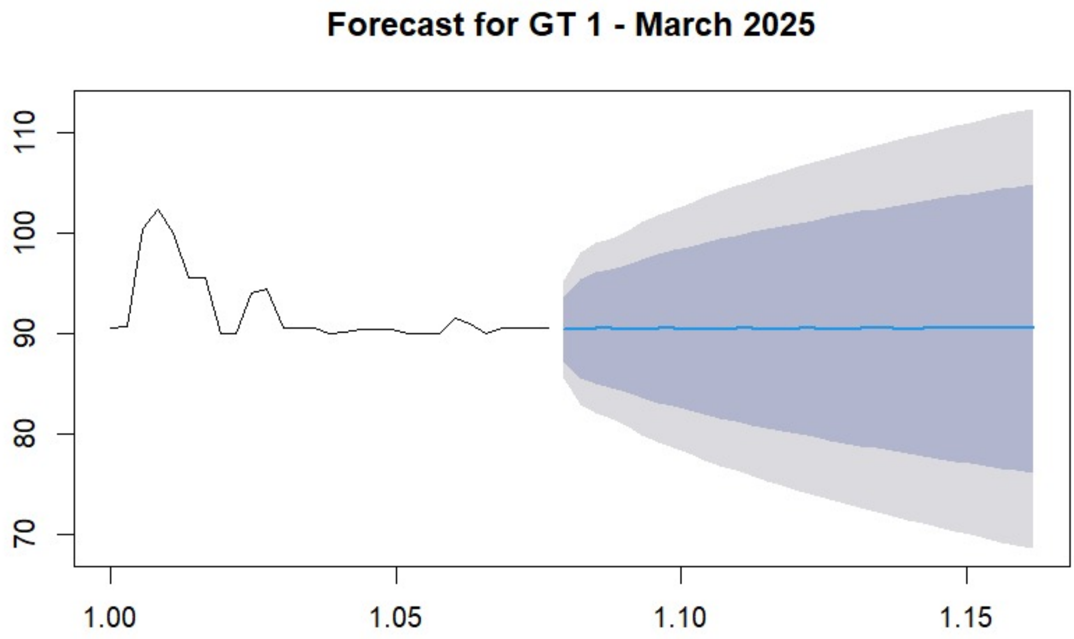

Figure 4.3.1: Forecast for GT 1 — Phase Two

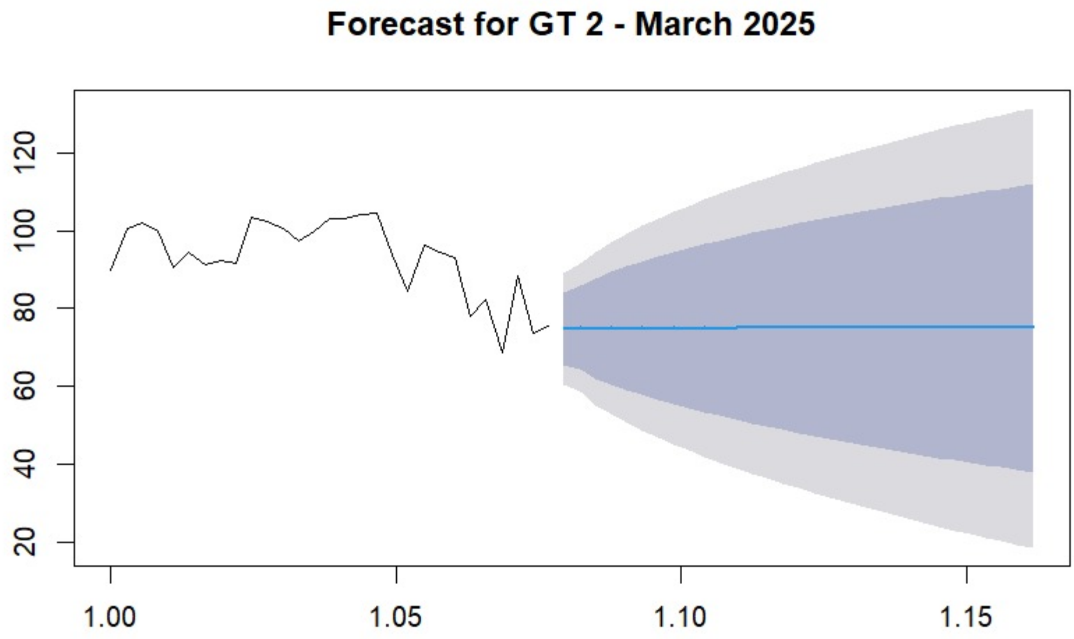

Figure 4.3.2: Forecast for GT 2 — Phase Two

Forecast for GT 2: The initial historical data, likely representing March, shows GT 2 exhibiting moderate fluctuations, primarily ranging from around 80 to 100 units. Following this initial period, the forecast model projects a stable central tendency. The most notable aspect of this forecast is the flat (stable) trend in the central prediction. The blue line, representing the forecast mean, indicates that GT 2's performance is expected to consistently hold around 70 units for the entire daily forecast period, extending to March 2025. This suggests no anticipated increase or decrease in its average performance over time.

However, the level of uncertainty surrounding this stable mean increases significantly over the forecast horizon. The darker shaded area indicates the 80% confidence interval, meaning 80% of future daily observations are expected to fall within this range. This interval widens progressively, expanding from a relatively tight band initially to span approximately 20 to 120 units by March 2025. The lighter-shaded area represents the 95% confidence interval, encompassing an even broader range of expected outcomes. This interval expands even more dramatically, reaching from below 0 to over 150 units by the end of the forecast period. It's crucial to note that the lower bound of the 95% interval extends into negative values towards the end of the forecast, implying a theoretical possibility of extremely low or even negative outcomes, though the central forecast remains positive.

Forecast for GT 3: The initial historical data, likely from March, shows GT 3 exhibiting some daily fluctuations, generally ranging from around 80 to 120 units. Following this initial period, the forecast model projects a strikingly stable central tendency. The most prominent feature of this forecast is the flat (stable) trend in the central prediction. The blue line, representing the forecast mean, indicates that GT 3's performance is expected to consistently hold around 90 units for the entire daily forecast period, extending to March 2025. This suggests no anticipated upward or downward movement in its average performance over the forecast horizon.

However, a significant observation is the increasing uncertainty surrounding this stable mean as the forecast extends. The darker shaded area, representing the 80% confidence interval, signifies that 80% of future daily observations are expected to fall within this range. This interval progressively widens over time, expanding from a relatively tight band initially to span approximately 20 to 160 units by March 2025. The lighter shaded

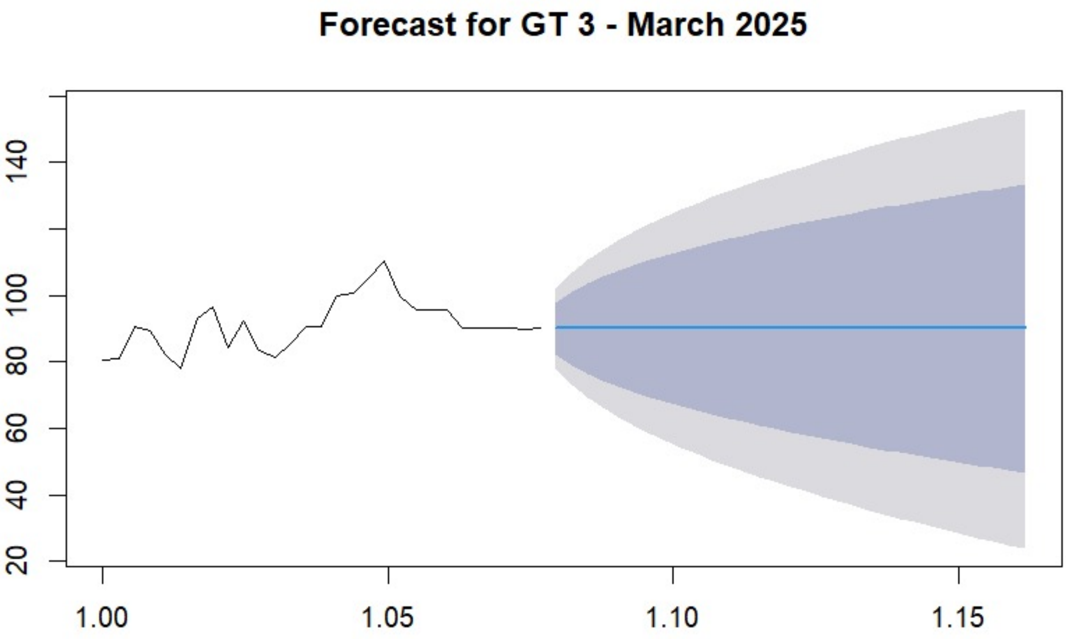

Figure 4.3.3: Forecast for GT 3 — Phase Two

area, representing the 95% confidence interval, encompasses an even broader range of expected outcomes. This interval expands even more dramatically, reaching from below -50 to over 200 units by the end of the forecast period. It is important to note that the lower bound of the 95% interval extends into negative values, implying a theoretical possibility of extremely low or negative outcomes, though the central forecast remains robustly positive at 90.

Forecast for GT 4: The initial historical data, likely from March, shows GT 4 experiencing notable fluctuations, with values ranging significantly, including dips below 70 and spikes above 100. Following this initial period of variability, the forecast model projects a stable central tendency. The blue line, representing the forecast mean, indicates that GT 4's performance is expected to consistently stabilize around 85 units for the entire daily forecast period, extending to March 2025. This suggests no anticipated upward or downward movement in its average performance over the forecast horizon.

However, the level of uncertainty around this stable mean is substantial and broadens over the forecast horizon. The darker shaded area, representing the 80% confidence

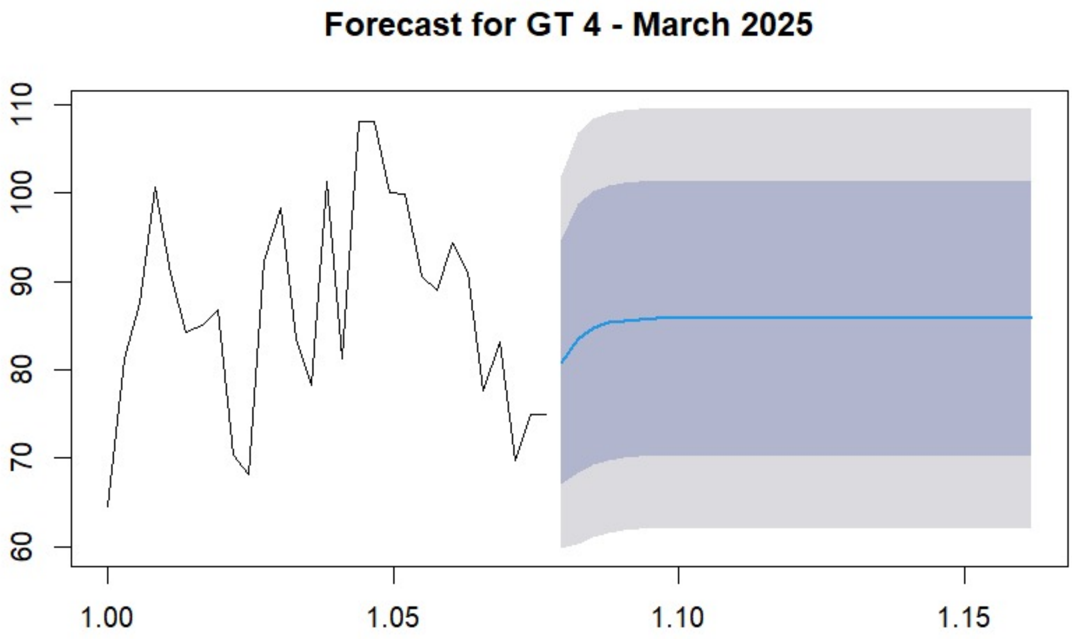

Figure 4.3.4: Forecast for GT 4 — Phase Two

interval, signifies that 80% of future daily observations are expected to fall within this range. This interval expands from a relatively tight band initially to span approximately 70 to 100 units by March 2025. The lighter shaded area, representing the 95% confidence interval, encompasses an even broader range of expected outcomes. This interval expands more dramatically, reaching from approximately 60 to 110 units by the end of the forecast period.

In conclusion, while GT 4 is forecast to maintain a stable average daily performance of approximately 85 units with no discernible upward or downward trend, the broadening width of the prediction intervals signifies a considerable and growing level of uncertainty. This implies that actual daily performance, despite the stable central prediction, could vary notably from this average as the forecast extends towards March 2025.

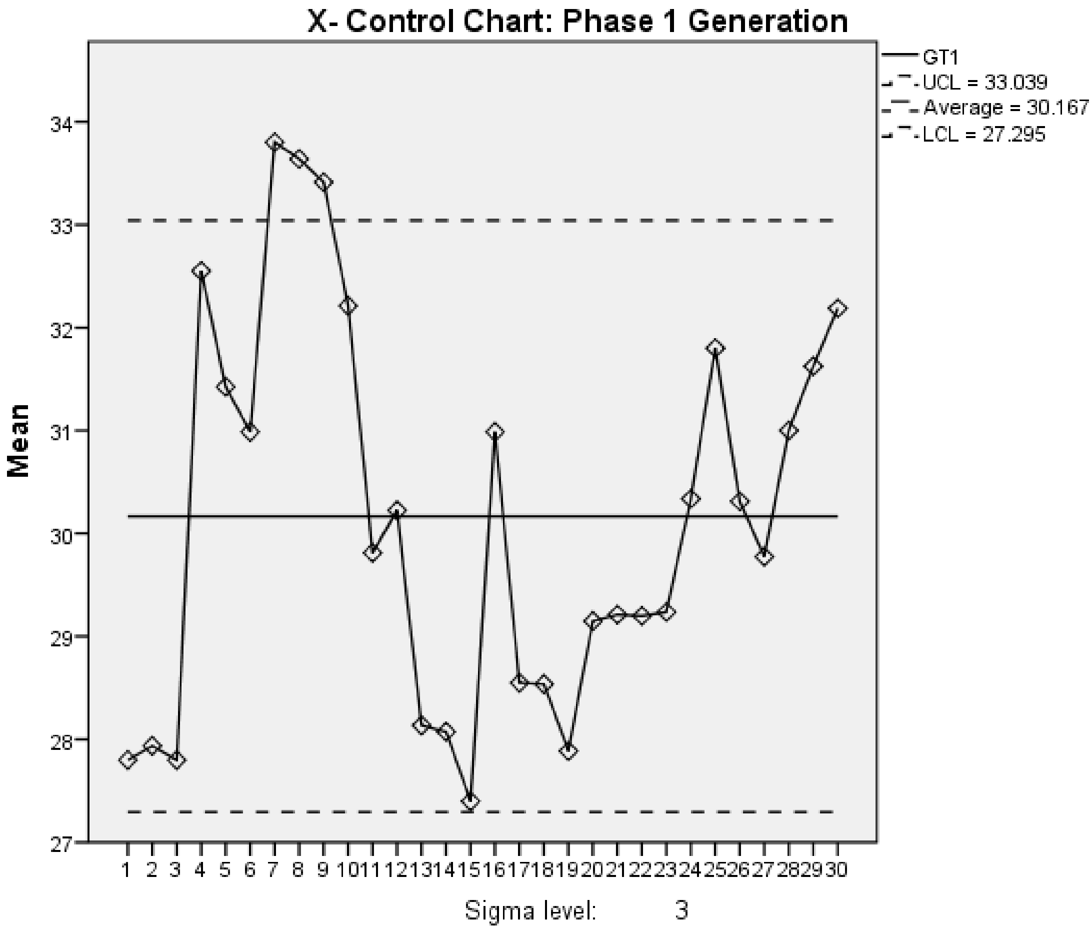

Figure 4.4.15: X-bar chart for GT 4 — Phase Two (Out of Control)

Interpretation: The X-bar chart shows some points outside the UCL and LCL, indicating the process is **out of control**.

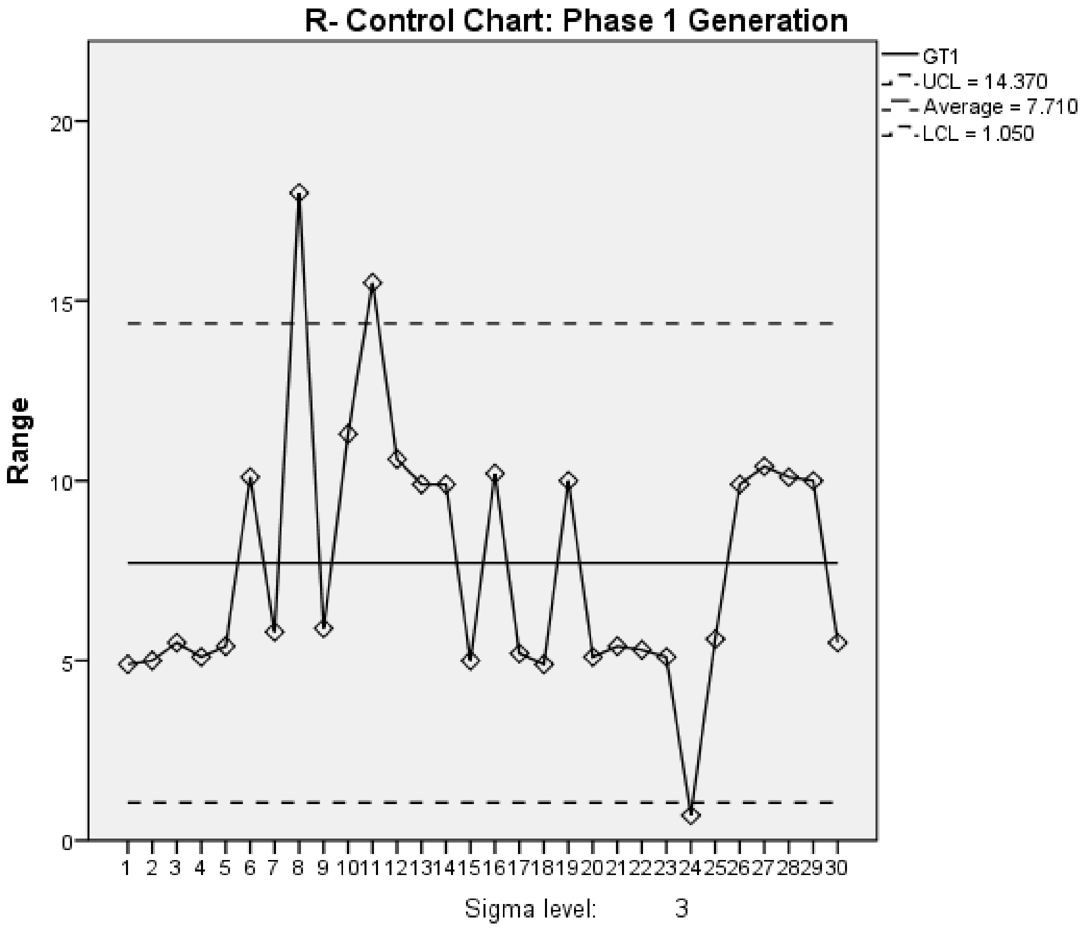

Figure 4.4.16: R-bar chart for GT 4 — Phase Two (Out of Control)

Interpretation: The R-bar chart shows some points outside UCL and LCL, confirming the process is **out of control**.

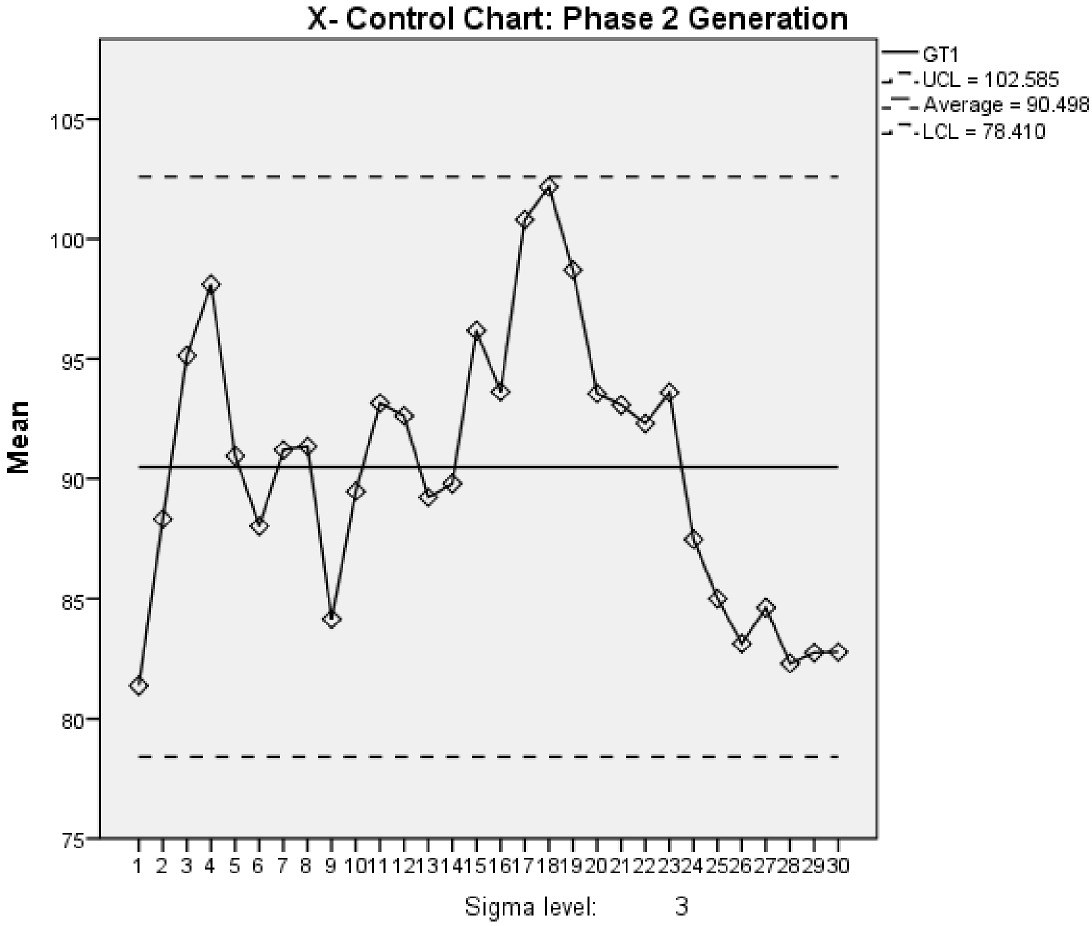

Figure 4.4.17: X-bar chart for GT 4 — Phase Two (In Control)

Interpretation: All points lie within UCL and LCL, so the process is **in control**.

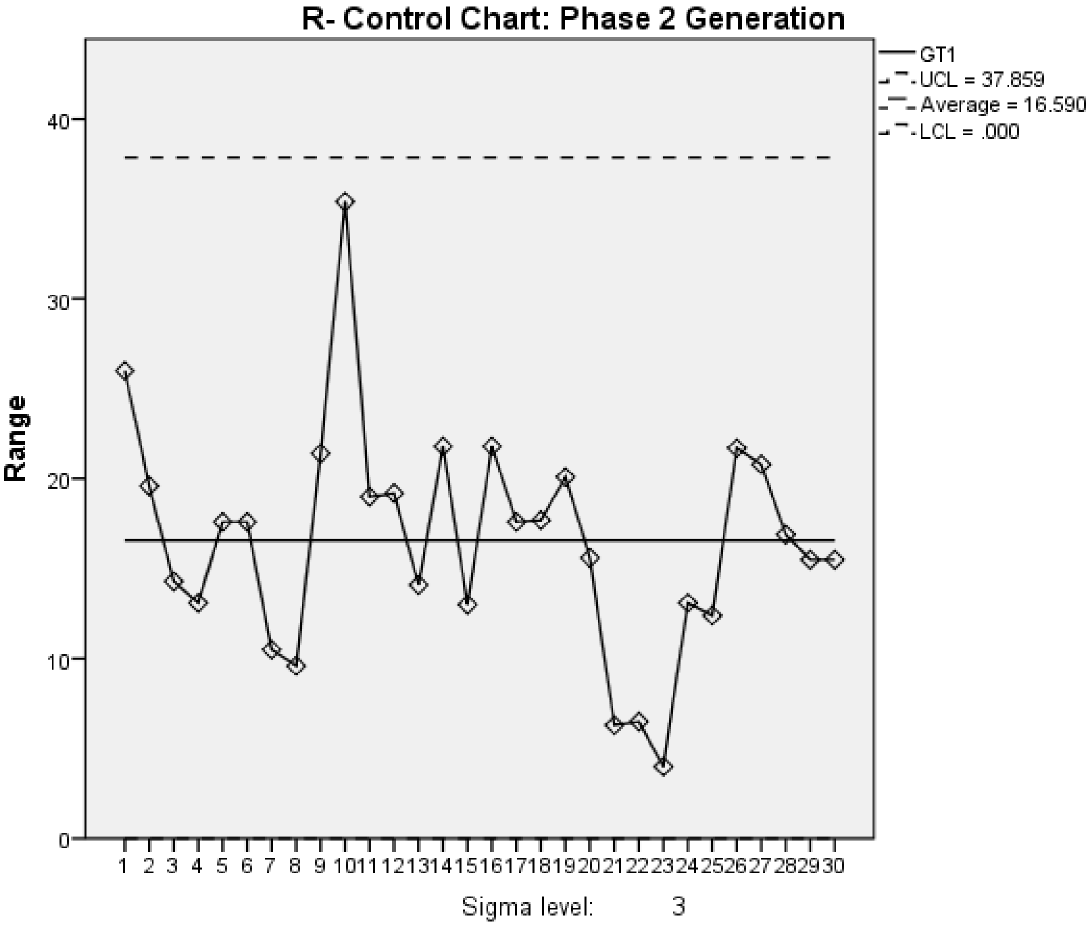

Figure 4.4.18: R-bar chart for GT 4 — Phase Two (In Control)

Interpretation: All points lie within control limits, confirming the process is **in control**.

Chapter 5: Summary, Conclusion and Recommendation

5.1 Summary

This study analyzed and predicted daily power transmission load data using time series techniques and statistical quality control (SQC) tools to assess system performance and reliability.

Data preprocessing involved cleaning, handling missing values, detecting outliers, and normalizing load records. Time series methods, such as ARIMA, were applied to identify trends, detect seasonality, and forecast future power demand. Statistical quality control techniques, including X-bar and R control charts and process capability analysis, were used to monitor load stability and ensure operational parameters remained within acceptable limits.

5.2 Conclusion

The results indicate that time series forecasting models can accurately predict future load patterns, helping operators anticipate demand fluctuations and improve transmission planning. Statistical quality control charts proved effective in detecting abnormal variations and potential faults in the power grid, thereby enhancing reliability. Integrating time series analysis with quality control techniques offers a robust framework for monitoring system performance, optimizing operational decisions, and reducing transmission losses.

5.3 Recommendations

Based on the study, the following recommendations are proposed:

1. **Adopt continuous monitoring:** Power utilities should implement real-time dashboards that combine time series forecasting with SQC tools to promptly detect anomalies.

2. **Model updating:** Forecasting models (e.g., ARIMA or SARIMA) should be periodically retrained with the latest load data to maintain accuracy.

3. **Preventive maintenance:** Use early fault detection from control charts to schedule timely equipment inspections and avoid unplanned outages.

4. **Capacity planning:** Apply load forecasts to guide infrastructure upgrades and improve grid stability during peak demand periods.

5. **Training and skill development:** Engineers and operators should receive training in time series modeling, quality control, and data analysis to maximize the benefits of these methods.

6. **Data quality improvement:** Ensure consistent and accurate data collection through modern SCADA systems and automated reporting tools.

References

Afonja, B. (1985). *Introduction to Statistics.* Ibadan: University Press.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., Ljung, G. M. (2016). *Time Series Analysis: Forecasting and Control.* Hoboken, NJ: Wiley.

Deshmukh, M., Bilolikar, R. (2006). Feasibility of grid extension and distributed generation considering biomass and diesel-based generation options. *Energy Policy Journal.*

Hyndman, R. J., Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). Melbourne: OTexts.

Juran, J. M. (1974). *Quality Control Handbook.* New York: McGraw-Hill.

Kazmier, L. J. (1993). *Theory and Problems of Business Statistics.* New York: McGraw-Hill.

Kaundinya, D. P., Balachandra, P., Ravindranath, N. H. (2009). Modeling and analysis of decentralized power options. *Renewable and Sustainable Energy Reviews, 13*(4), 852–858.

Khator, S. K., Leung, L. C. (1997). Power distribution planning under normal and emergency conditions. *IEEE Transactions on Power Systems.*

Monteiro, C., Ferreira, P., Bessa, R. (2005). Impact of integration of distributed resources on electricity distribution using GIS. *Electric Power Systems Research, 75*(2), 45–52.

Montgomery, D. C. (2020). *Introduction to Statistical Quality Control* (8th ed.). Hoboken, NJ: Wiley.

Moore, D. S., McCabe, G. P., Craig, B. A. (2017). *Introduction to the Practice of Statistics* (9th ed.). New York: W. H. Freeman.

Olabiyi, I., et al. (2012). *Statistical Quality Control Techniques in Power Transmission.* Lagos: Nigerian Journal of Power Systems.

Olugbade, A. (2018). *Transmission Company of Nigeria: Operational Review.* Abuja: TCN Publication.

P. Gupta. (2001). *Introduction to Statistical Methods.* New Delhi: Vikas Publishing House.

Shumway, R. H., Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples* (4th ed.). New York: Springer.

Spiegel, M. R. (1992). *Theory and Problems of Statistics (Metric Edition)*. New York: McGraw-Hill.

Urban, F., Benders, R. M., Moll, H. C. (2009). Renewable-based rural electrification: The most cost-effective option. *Energy Policy, 37*(1), 126–134.

Zhou, Q., et al. (2002). Substation location and capacity optimization using GIS. *International Journal of Electrical Power and Energy Systems, 24*(10), 813–819.