ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

TELECOMMUNICATION ENGINEERING GRADUATE PROGRAM

Modeling GSM Spectrum Occupancy Using Time Series Analysis: The case of Ethio telecom

By: Anteneh Atnafu

Adviser: Dr. Murad Ridwan

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfilment of the Requirements for the Degree of Master Science in Telecommunication Network Engineering.

September 2021

Addis Ababa, Ethiopian

ADDIS ABABA INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING
TELECOMMUNICATION ENGINEERING GRADUATE PROGRAM

Modeling GSM Spectrum Occupancy using Time Series Analysis: The case of Ethio telecom

By Anteneh Atnafu

Signed by the Examining Committee:

Internal Examiner	______	Signature	______	Date	______
External Examiner	______	Signature	______	Date	______
Adviser	Dr. Murad Ridwan	Signature	______	Date	______
Co-Adviser	______	Signature	______	Date	______

____
Dean, School of Electrical and Computer Engineering

Declaration

I, the undersigned, declare that this thesis is my original work, has not been presented for a degree in this or any other university, and all sources of materials used for the thesis have been fully acknowledged.

Anteneh Atnafu

Name

Signature

Place: Addis Ababa University, Ethiopia

Submission Date: September 2021

This master thesis has been submitted for examination with my approval as a university adviser.

Dr. Murad Ridwan

Name

Signature

Abstract

Spectrum occupancy models can help you make better use of the radio spectrum. It has also been extensively researched in recent decades as it is critical for developing new regulations for spectrum allocation for future technologies as well as monitoring the activities that take place on that spectrum. Understanding the amount of available spectrum is critical for future wireless technologies that want to address the so-called spectrum scarcity issue. Spectrum occupancy measurements provide critical data for frequency planning and optimization, as well as assist in smart decision-making. The goal of time series modeling is to collect and thoroughly examine previous data from a time series in order to construct an appropriate model that accurately captures the series' intrinsic structure. This thesis examines three types of time series analysis methods: Holt-winters, Seasonal Auto Regressive Integrated Moving Average (SARIMA), and SARIMA eXogenous regresses (SARIMAX) based models, as well as their inherent prediction strengths and weaknesses. Time series modeling principles such as trend, stationarity, seasonality, residual, and so on have also been covered. To assess the accuracy rate, we fitted multiple models to a time series using five primary metrics. Among the methods used are mean square error, mean absolute error (MAE), root-mean-square error, mean absolute percentage error, and R-squared. At 1800MHz, the maximum spectrum occupancy is 60.35%, and at 900MHz, it is 44.71%. For 900MHz MAE, the SARIMAX model produced better predictions (35.34% and 50.1% lower than the SARIMA and Holt-Winter models, respectively), while for 1800MHz, the SARIMAX model produced 42.6% and 52.6% lower than the SARIMA and Holt-Winter models, respectively.

***Keywords*:** Spectrum Occupancy, Holt-Winters, SARIMA, and SARIMAX.

Acknowledgements

First and foremost, praise and thanks to God, the Almighty, for His blessings throughout my thesis.

I would like to thank Dr. Murad Ridawn for his invaluable assistance and insights leading to the writing of this paper.

I would like also to thank Dr. Beneyam Berehanu and Dr. -Eng. Yihenew Wondie for their great feedback and excellent encouragement and guidance throughout the thesis process.

I would like to thank ethiotelecom for giving me the opportunity to pursue this Master's degree.

I would also like to thank Huawei RF staff Taddel for their excellent support.

Most importantly, I am grateful for my family's and friend's unconditional, unequivocal, and loving support.

Table of Contents

Abstract . . . iv

Acknowledgements . . . v

Table of Contents . . . vii

List of Figures . . . viii

List of Tables . . . x

List of Acronyms . . . xi

1 Introduction . . . 1
- 1.1 Background . . . 1
- 1.2 Statement of The Problem . . . 2
- 1.3 Objectives . . . 3
  - 1.3.1 General Objective . . . 3
  - 1.3.2 Specific Objectives . . . 3
- 1.4 Methodology . . . 4
- 1.5 Scope and Limitation . . . 5
  - 1.5.1 Scope of The Thesis . . . 5
  - 1.5.2 Limitation of The Thesis . . . 5
- 1.6 Related Work . . . 5
- 1.7 Contribution . . . 6
- 1.8 Thesis Layout . . . 7

2 GSM Network and Spectrum Overview . . . 8
- 2.1 Introduction to GSM . . . 8
- 2.2 GSM Technology . . . 9
  - 2.2.1 GSM Network . . . 9
  - 2.2.2 GSM Protocol Structure . . . 12
  - 2.2.3 Frame Structure and Radio Channels . . . 13
- 2.3 EGPRS System . . . 16
  - 2.3.1 Channels in EGPRS Network . . . 17
  - 2.3.2 Coding Schemes . . . 17
- 2.4 Frequency Spectrum . . . 19
  - 2.4.1 Introduction to Spectrum . . . 19
- 2.5 Spectrum Refarming . . . 20
  - 2.5.1 Static Spectrum Refarming . . . 21
  - 2.5.2 Dynamic Spectrum Refarming . . . 22

3 Time Series 23
3.1 Introduction to Time Series 23
3.2 Characteristics Time-Series 23
3.2.1 Stationarity in Time-Series Data 24
3.2.2 Techniques of Making Data Stationery 25
3.3 Statistics Unit Root Tests 27
3.3.1 P-Value 27
3.3.2 Augmented Dickey-Fuller Test 28
3.3.3 Kwiatkowski-Phillips-Schmidt-Shin Test 28
3.4 Autocorrelation and Partial Autocorrelation Functions 29
3.5 Auto-Regressive Models 31
3.6 Moving Average 31
3.7 Autoregressive Integrated Moving Average 32
3.8 Seasonal ARIMA 33
3.9 SARIMAX 34
3.10 Exponential Smoothing Methods 34
3.10.1 Simple Exponential Smoothing 35
3.10.2 Double Exponential Smoothing 35
3.10.3 Triple Exponential Smoothing 36
3.11 Estimation and Order Selection 36
3.11.1 Maximum Likelihood Estimation 36
3.11.2 Information Criteria 37
3.12 Diagnostic Checking 37
3.13 Evaluation Metrics 38

4 Result and Discussion 40
4.1 The Data 40
4.2 Check for Stationarity 43
4.3 Autoregressive and Moving Average Orders 47
4.4 Triple Exponential Smoothing Holt-Winters 51
4.5 Estimation and Diagnostic Checking of Models 52
4.6 Prediction Comparison 56

5 Conclusion and Recommendation 60
5.1 Conclusion 60
5.2 Recommendation(Relevance) for Future Work 61

References 66

Appendices 67

List of Figures

1.1 System Model/Methodology 4
2.1 GSM /EDGE Architecture[14] 9
2.2 GSM INTERFACE PROTOCOL MODEL[17] 12
2.3 GSM Frame Structure[18] 14
2.4 FDMA channel[20] 14
2.5 Time-Frequency Structure of Physical Channel[20] 15
2.6 Sandwich Method[24] 22
2.7 Edge Method[24] 22
4.1 Google view of the study area 40
4.2 Spectrum Occupancy 900MHz plot four months data 41
4.3 Spectrum Occupancy 1800MHz plot four months data 42
4.4 Spectrum Occupancy1800MHz and 900MHz plot one month 42
4.5 Distribution plot of 1800MHz and 900MHz spectrum data 42
4.6 Lag 1 and 24 Correlation of 1800MHz and 900MHz 43
4.7 1800MHz occupancy decomposition 44
4.8 900MHz occupancy decomposition 44
4.9 Rolling mean & std stationary test 1800MHz occupancy data 44
4.10 Rolling mean & std stationary test 900MHz occupancy data 45
4.11 ACF and PACF first difference of 1800MHz spectrum data 49
4.12 ACF and PACF of 900MHz spectrum data 49
4.13 ACF & PACF of seasonal difference of 1800MHz 49
4.14 ACF & PACF of seasonal difference 900MHz 50
4.15 ACF & PACF of First-Order Differencing seasonal 50
4.16 ACF & PACF of First-Order Differencing seasonal 900MHz 50
4.17 Standardized residual SARIMA 900MHz 54
4.18 ACF plot of residual SARIMA 900MHz 54

4.19 Standardized residual SARIMA 1800MHz 54
4.20 ACF plot of residual SARIMA 1800MHz 55
4.21 Standardized residual SARIMAX 900MHz 55
4.22 ACF plot of residual SARIMAX 900MHz 55
4.23 Standardized residual SARIMAX 1800MHz 56
4.24 ACF plot of residual SARIMAX 1800MHz 56
4.25 prediction vs actual SARIMA 900MHz 57
4.26 Prediction vs Actual SARIMAX 900MHz 57
4.27 Prediction vs Actual holt-winters 900MHz 58
4.28 Prediction vs Actual SARIMA 1800MHz 58
4.29 prediction vs actual SARIMAX 1800MHz 59
4.30 Prediction vs Actual holt-winters 1800MHz 59

List of Tables

2.1 Logical control channels in GSM 16
2.2 EGPRS channels 17
2.3 Coding schemes for the GPRS network 17
2.4 Coding schemes for the EGPRS network 18
2.5 Spectrum used by ethiotelecom[5] 20
3.1 ACF and PACF to identify the orders of SARMA(p, q) × (P, Q)s 31
4.1 Show the detailed spectrum channel assigned to GSM 40
4.2 Descriptive Statics of Data 41
4.3 ADF test with test-statistic and p-value 45
4.4 KPSS test with test-statistic and rejection value 46
4.5 SARIMA(p, d, q) (P, D, Q)24 48
4.6 SARIMAX (p, d, q) (P, D, Q)24 49
4.7 Prediction Evaluation metrics for Model Holt-Winters 51
4.8 Hot-Winters fitted parameter 51
4.9 SARIMA (1, 0, 2) x (1, 1, 2,)24 fitted for 900MHz 52
4.10 SARIMA (2, 1, 2) x (2, 1, 3)24 fitted for 1800MHz 53
4.11 SARIMAX (1, 0, 2) x (1, 1, 1)24 fitted for 900MHz 53
4.12 SARIMAX (1, 0, 3) x (1, 1, 2)24 fitted for 1800MHz 53
4.13 Summary prediction Evaluation three models 58
4.14 Summary of Evaluation metrics the models accuracy 59

List of Acronyms

2G	Second Generation
3GPP	3rd Generation Partnership Project
ACF	Auto Correlation Function
ADF	Augmented Dickey-Fuller
AIC	Akaike information criterion
AMPS	Advanced Mobile Phone System
AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
ARQ	Automatic Repeat Request
AUC	Authentication Centre
BIC	Bayesian information criterion
BSC	Base Station Controller
BSS	Base Station Subsystem
BTS	Base Transceiver Station
CDMA	Code Division Multiple Access
CS	Coding Schemes
DSR	Dynamic Spectrum Refarming
ECA	Ethiopian Communication Authority
EDGE	Enhanced Data Rates for GSM Evolution
EGPRS	Enhanced General Packet Radio Service
EIR	Equipment Identity Registers
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
GGSN	Gateway GPRS Support Node

GMSK	Gaussian Minimum Shift Keying
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HARQ	Hybrid Automatic Repeat Request
HLR	Home Location Register
IMEI	International Mobile Equipment Identity
ITU	International Telecommunication Union
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
LAPD	Link Access Procedure on D channel
LLR	Loge-Likelihood Ratio
LMR	Land Mobile Radio
LTE	Long Term Evolution
MA	Moving Average
MAD	Median Absolute Deviation
MAE	Mean Absolute Error
MANE	Mean Absolute Normalized Error
MAPE	Mean Absolute Percentage Error
MCS	Modulation Coding Schemes
MMS	Multi Media Service
MSC	Mobile Switching Centre
MSE	Mean Square Error
MTP	Message Transfer Part
NMS	Network Management System
NMT	Nordic Mobile Telephone
NSS	Network Switching Subsystem
OMC	Operation and Maintenance Subsystem
PACF	Partial Auto Correlation function
PLMN	Public land mobile network
RMSE	Root Mean Square Error
SARIMA	Seasonal ARIMA
SARIMAX	SARIMA eXogenous regresses

SGSN	Serving GPRS Support Node
SIM	Subscriber Identification Module
SMS	Short Message Service
SSE	Sum Square Error
SSR	Static Spectrum Refarming
TACS	Total Allowable Catch System
TDMA	Time Division Multiple Access
TDNN	Time Delay Neural Networks
UMTS	Universal Terrestrial Universal Mobile Telecommunications System
VLR	Visitor Location Register

1 | Introduction

1.1 Background

Radio spectrum is a unique natural resource because it is non-exhaustible, but it is regarded as a scarce resource owing to how we use it. As a result, competent agencies such as the ITU have been established to ensure the sensible, fair, and economical use of the spectrum. Their objective is to ensure that wireless systems operate without interruption around the world by enforcing radio regulations and regional agreements. Current wireless communication systems rely heavily on fixed radio resource allocation. This means that a section of the radio spectrum is allocated or licensed to third parties on a long-term basis and that it covers enormous areas, such as entire countries. This approach provides ultimate protection from harmful interference in an allocated radio band, but as many measurement studies show, the fixed radio frequency allocation may result in significant under-utilization of the radio spectrum due to very sporadic usage across different geographical regions and periods[1][2].

The rising usage of radio-based technologies in society, as well as the enormous prospects for social development that these technologies afford, underscore the significance of radio-frequency spectrum and national spectrum management systems. Technological advancement has consistently opened the door to a lot of new spectrum applications, resulting in increased interest in, and demand for, finite spectrum resources. Increased demand necessitates efficient spectrum utilization and the implementation of appropriate spectrum management techniques. Modern data handling capabilities and technical analyses are critical in this framework to suit the wide range of possible users requesting spectrum access.

Wireless networks are experiencing a huge transformation right now. According to current trends, the number and variety of wireless devices, as well as spectrum demand, are on the rise. The radio frequency spectrum is, unfortunately, a finite resource. As a result, some parts of the spectrum are heavily utilised while others are entirely ignored. It is critical to monitor and understand spectrum resource consumption in order to improve and regulate radio spectrum utilization[3][4].

Spectrum occupancy measurement studies seek to quantify the amount of time that a specific frequency channel or frequency band is occupied in a given area, hence defining the channel's utilization rate. Measurements can be used to analyse the current state of spectrum utilization as well as the availability of spectrum for other users. Furthermore, the measurement results provide significant information to regulators as well as operators such as ethiotelecom about the efficiency of the existing usage of spectrum allocations. Spectrum measurements are also carried out to increase the accuracy of spectrum databases, which can be used to keep track of spectrum utilization statistics to promote spectrum sharing in an operational setting[4].

To optimize limited resources, current spectrum allocation strategies must be studied and actual spectrum occupancy information must be studied to provide an understanding and prediction of spectrum user activity, which is an essential step toward improving accuracy and decision-making processes to assess the feasibility of efficient technology.

In this thesis, we measure GSM 900MHz and 1800MHz spectrum occupancy models, which target the usage patterns of the spectral bands used by users of the ethiotelecom network in Addis Ababa, using time series analysis methods, specifically three algorithms: exponential smoothing (Holt-Winter), SARIMA, and SARIMAX. To assist the company's optimization and spectral impact efforts by providing the current spectrum occupancy model and prediction information for the GSM network.

1.2 Statement of The Problem

Spectrum is a limited resource, but its efficient use is the primary goal of operators. However, in the case of state monopoly operators such as ethiotelecom, scarcity of spectrum is not the main concern at this time. The ECA is, however, in the process of granting licenses to two additional telecom operators. As a result, ethiotelecom must consider possibilities to improve spectral efficiency approaches. Ethiotelecom uses 70% of the 900MHz spectrum for GSM and UMTS and 50% of the 1800MHz spectrum for GSM and LTE[5]. With the entrance of two new carriers, the spectrum distribution is projected to shift. Before attempting to evaluate the potential improvements attainable through the usage of additional efficient technology, we need to grasp and recognize the actual occupancy of the current GSM bands currently allotted to each service based on network performance data from that analysis result. Occupancy of the GSM Spectrum Understanding which frequency is used at which time and space is critical to finding a suitable solution to the approaching spectrum sharing effect.

1.3 Objectives

1.3.1 General Objective

The goal of this thesis is to model and predict ethiotelecom GSM spectrum occupancy and provide information that contributes to the company's solution effort to overcome the coming extra operator's spectrum sharing influence.

1.3.2 Specific Objectives

The specific objectives of this thesis are:

- To compare different time series modeling techniques and select the simplest and accurate one.
- To show spectrum occupancy in different time and frequency.
- To model the real-time spectrum occupancy mathematical.
- To demonstrate a statistical description of spectrum occupancy data analysis.

1.4 Methodology

The following methodology was used to carry out this thesis: The first is a literature review, in which different academic research regarding the spectrum occupancy measurement of different spectrum bands based on time series analysis was reviewed in order to learn more about time series models related books were studied. The second part involves collecting measured data in the GSM spectrum occupancy, which is assigned to GSM at 900 MHz and 1800 MHz. The Ethio Telecom Performance Reporting System (PRS) is used to collect data. Third, as shown in figure 1.1, method identification can be accomplished using the series pattern, time plot, and correlogram ACF and PACF to identify the model. After identifying the tentative model, the parameters are estimated and tested for statistical significance using the information criteria AIC and BIC. The correlogram of the residuals from the estimated model should be a white noise process in the diagnosis process, and the residual ACF should be plotted in the rang. If the residuals continue to be significantly correlated, a new model should be identified and diagnosed. When a model is chosen, it is used to predict. Time series analysis provides numerous opportunities for detecting, describing, and modeling time series. Finally, to comprehend the planning and decision-making process, select the best model from seasonal ARIMA, SRAIMAX, and exponential smoothing modeling methods. The forecast performance of the models is then compared to some commonly used error measures such as MAE, MSE, RMSE, MAPE, and R-Squared. Finally, we will hold a discussion based on the outcome, and then we will conclude. Based on the results, some optimization was carried out.

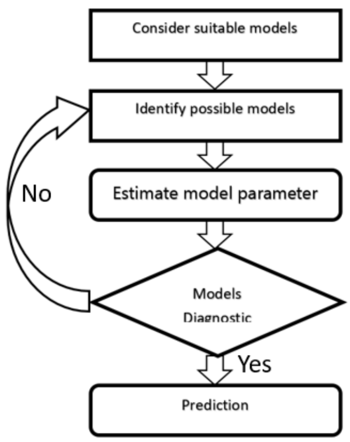

Figure 1.1: System Model/Methodology

1.5 Scope and Limitation

1.5.1 Scope of The Thesis

The scope of this thesis is to investigate how to measure the occupancy of the GSM frequency spectrum band currently assigned to ethiotelecom as well as potential solutions.

1.5.2 Limitation of The Thesis

The limitations of this thesis in the analysis of spectrum occupancy case study data should be considered in Addis Abeba, and the study should only consider spectrum allocated in GSM..

1.6 Related Work

Based on the constraints of the privies work, the research[6] develops a model and prediction that will aid in the creation of a cognitive radio platform. To analyse the data, use 850MHz spectrum data and divide it into three groups (high, medium, and low) using SARIMA time series analysis. Data evaluation is based on available time and channel occupancy in various MAE and AIC classes. The results of the research show that the MAE of the channel model with low occupancy outperforms the channel models with medium and high occupancy.

Durham University is conducting research[7] to estimate current spectral occupancy in the 100-2500 MHz band in order to examine the feasibility of cognitive radio technology in the UK and to build statistical models for more general estimation and prediction of spectrum occupancy. Time series analysis (SARIMA) can be used to analyse data based on its attributes. Based on the AIC used to quantify the goodness of fit of the model, the results show that SARIMA is the best fit model for GSM 935 - 960MHz band data, while ARMA is the best fit model for TV 470 – 590 MHz band data. The paper's findings reveal that spectrum occupancy is low and that up to 2 MHz of bandwidth is accessible up to

80% of the time, which is important for cognitive radio.

Based on the limitations of previous works, such as the measurement location instead of just one measurement, why not take many measurements, such as four measurements at the same time[8]? Use a high-performance disc one antenna (AOR DA5000) with a frequency range of 700-3000 MHz and Omni-directional horizontal receiving capabilities. Time series analysis utilizing the Rohde & Schwarz FSH6 Portable Spectrum Analyser (Doha, Qatar). The data is analysed using AR. In addition, the paper uses the AIC for model fit parameter and MRE for model evaluation. The results suggest that AR (3) is the best fit model for representing the data.

To examine a comparison of the predictive capability of time-series models and machine learning algorithms based on the behaviour of the data set. Also[9], consider the forecast of stationary traffic in a TV band versus non-stationary traffic in a cellular band. Use a Sirio SD 1 300N Omni-directional disc one antenna, an RF explorer-based mini-spectrum analyser, and a week's worth of data. The TV band has a spectrum range of 150-750 MHz, while the cellular 900MHz band has a spectrum range of 850-1 300 MHz for modeling and prediction, the research employs both time series analysis and ML-based techniques. According to the results, both time-series modeling and ML-based prediction are reasonably excellent at forecasting spectrum occupancy based on MAPE and MANE.

In this research[10], we want to know if adding a non-linear component time series prediction method to LMR spectrum measurement provides any significant benefits over the linear prediction method. The study compares TDNN and SARIMA using the LMR spectrum 138MHz to 941MHz and the algorithm applied. For evaluation metrics, use mean square error. The results of the paper reveal that both TDNN and SRAIMA fit the model well, but MAE TDNN predicts the data better.

1.7 Contribution

The main contributions of this Thesis are:

- Describe in mathematical model that represents the real-time GSM spectrum occupancy data.

- The thesis's output aids in the decision-making process regarding which spectral efficient technology to use and when and where the company will implement the viable solution.
- We can perform machine learning-based deep optimization based on the findings of this thesis.

1.8 Thesis Layout

This thesis is divided into six chapters. An introduction, a problem statement, general and particular aims, related work, and the methodology employed in this thesis comprise Chapter One. The second chapter covers the fundamentals of GSM/EDGE technology, channel organization, and the spectrum overview and spectrum refarming concepts, as well as refarming methodologies (SSR and DSR). Chapter three discusses time-series features, time-series data, data decomposition, and time series analysis tools, which include exponential smoothing procedures. Chapter four consists of data visualization of the algorithms' results and subsequent analysis depending on the outcomes. Finally, Chapter five presents the results of the conclusion, some recommendations, significance, and future work.

2 | GSM Network and Spectrum Overview

2.1 Introduction to GSM

Because of the incompatibility of the various systems in place, the European Commission initiated a series of discussions aimed at changing the then-existing telecommunication regulatory framework, resulting in a more harmonized environment and the development of a common market for telecommunication services and equipment. In the early 1990s, digital transmission technology became available, ushering in the next generation system known as the Second Generation Mobile System [11][12].

Three primary benefits of 2G networks over their predecessors were:

- Phone conversations that are digitally encrypted, at least between the mobile phone and the cellular base station, but not necessarily throughout the network.
- Significantly better use of the radio frequency spectrum, allowing for more users per frequency band.
- Data services for mobile devices, beginning with SMS text messages.

2G technologies enabled services such as text messaging, image messaging, and MMS. Following the introduction of 2G, the previous mobile wireless network systems were renamed 1G retroactively. 2G network radio signals are digital, whereas 1G network radio signals are analog. Both systems use digital signaling to connect radio towers (which listen for devices) to the rest of the mobile system. With General Packet Radio Service, the theoretical maximum transfer speed is 40 Kbit/s, and with EDGE, the theoretical maximum transfer speed is 384 Kbit/s. Time division multiple access-based GSM, which originated in Europe but was widely used outside of Japan and North America, was the most widely used 2G technology. [13].

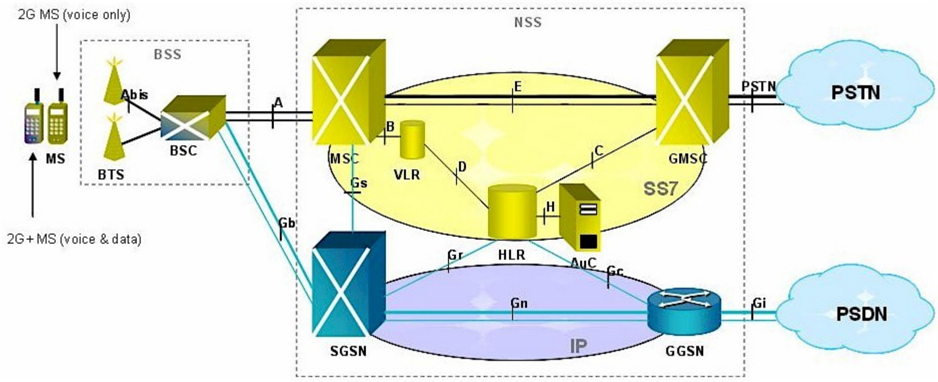

Figure 2.1: GSM /EDGE Architecture[14]

2.2 GSM Technology

2.2.1 GSM Network

Let us understand the key components of the GSM network. The latter consists of three main domains: The Mobile Station, the Base Station Sub-system, the Network Sub-system and the Network Management System, as shown in Figure 2.1.

Network Switching Subsystem

NSS is the core element of network switching which interfaces with subscriber services for voice and data.NSS Main components are:

Mobile Switching Centre: - This is the core of a GSM network. MSC performs a switching function, that is, it connects PLMN subscribers to subscribers in other networks. MSC manages PLMN subscribers' paging access, channel assignment, call connection, traffic control, billing, and base station management. It provides interfaces to other functional entities, interfaces to other networks, and interfaces to other MSC.

Home Location Register: -Each mobile subscriber's information is stored in the HLR. Each subscriber mobile contains data such as the type of subscription, services that the user can use, the subscriber's current location, and the status of the mobile equipment. The database in the HLR remains intact and unchanged until the subscription expires. It provides MSC with routing information for call setup. HLR may cover multiple MSC service areas or the entire PLMN.

Visitor Location Register: - VLR stores all subscriber information in its coverage region and provides call setup conditions to registered mobile customers. VLR, as a dynamic database, must communicate a huge volume of data with HLR to assure data veracity. When an MS exits the controlling region of a VLR, it registers in a different VLR. The subscriber's temporary records are deleted by the original VLR. VLR is integrated into MSC.

Equipment Identity Registers: - Each item of mobile equipment is identified by a number known as the International Mobile Equipment Identity. This number is installed during the manufacturing of the equipment and indicates compliance with GSM standards. As a result, whenever a call is made, the network checks the identification number, and if it is not found on the approved list of authorized equipment, access is denied. This list of authorized numbers is contained in the EIR and allows the International Mobile Equipment Identity to be verified.

Authentication Centre: -The Authentication Centre (AUC or AC) bears "responsibility" for network policing. This contains all of the data required to protect the network from false subscribers and to protect regular subscribers' calls. In the GSM standards, for example, there are two major keys: one for mobile user encryption and the other for mobile user authentication. The encryption keys are stored in both the mobile equipment and the AUC, and the data is protected from unauthorized access[15].

Base Station Subsystem

BSS serves as a bridge between NSS and MS. It performs radio channel management and wireless reception and transmission. Base Station Controller and Base Transceiver Station are the main components of BSS.

Base Station Controller: - Located between MSC and BTS, it controls and manages more than one BTS. It performs radio channel assignments. BTS and MS transmit power control and inter-cell handover. BSC is also small a switch that converges and connects the local network with the MSC through A-interface. A-bis interface connects BTS to BSC.

Base Transceiver Station: - BTS is wireless transceiver equipment controlled by the BSC in BSS. BTS carries radio transmission. It performs a wired-related wireless conversion, radio diversity, radio channel encryption, and hopping. Um, the interface connects BTS to MS[16].

Operation and Maintenance Subsystem

OMS is the operation and maintenance part of GSM. Functional units in GSM are connected to OMS internal networks. OMS monitors various functional units in the GSM network, submits status reports, and performs fault diagnoses.

Mobile Station

MS is subscriber equipment in GSM, it can be vehicle installed or hand portable. MS consists of mobile equipment and SIM. Mobile equipment processes voice signals to receive and transmits radio signals.

SIM stores all information required for identifying a subscriber and security information, preventing unauthorized subscribers. Mobile equipment cannot access the GSM network without a SIM card.

2.2.2 GSM Protocol Structure

2G cellular mobile network GSM adopts Open System Interconnection model to define its protocol structure. which defines the interfaces and protocols between MS and MSC?

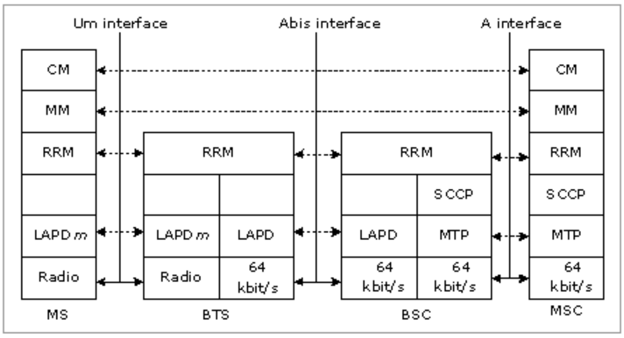

Figure 2.2: GSM INTERFACE PROTOCOL MODEL[17]

OSI reference model is a hierarchical structure. According to the hierarchy concept, the communication process can be divided into several logical layers from the lowest to highest layer. In different systems, the entities in the same layer that exchange information for the same purpose are called peer entities. Entities in adjacent layers interact with each other through the common layer. The lower layers provide services to higher layers. The services provided by layer N are a combination of the services and functions provided by the layers below it show Figure 2,2.

- The first layer of Um interface protocol is the physical layer, which is marked as L1 and is the lowest. L1 provides basic radio channels for the information transmission of higher layers.
- The second layer L2 is the data link layer, which is marked as LAPD. It covers various data transmission structures and controls data transmission.
- The application layer is the third highest layer L3. It covers various messages and programs and controls services. L3 includes Radio Resource Management, Mobility Management, and Call Connection Management.
- A-bis interface protocol is slightly different from Um interface protocol. Its physical layer is 64 kbps landline, and the link layer is LAPD.

- The first layer of the A-interface protocol is a 64 kbps landline, and the second layer is the MTP, which is part of the Common Channel Signalling7 network. MTP consists of many network protocols and centralizes all link-layer protocols. The signaling connection control part and MTP together represent a network layer protocol on A-interface[17].

2.2.3 Frame Structure and Radio Channels

GSM air interface uses TDMA based frame structure. Communication services are obtained by the transmission of information using logical channels on physical channels. Mapping between the logical channel and physical channel is the process that arranges the information to be sent to the suitable TDMA frames and timeslots[18][19].

Radio Frame Structure: -Five levels of GSM radio frame structure are timeslot, TDMA frame, multi-frame, super-frame, and hyper frame show in Figure 2.3.

- The timeslot is the basic unit of a physical channel.
- TDMA frame consists of eight timeslots. It is a basic unit occupying carrier bandwidth. Each carrier has eight timeslots.
- There are two types of multi frames.
  - One type of multi-frame consists of 26 TDMA frames. This type of multi-frame is used in TCH, SACCH, and FACCH.
  - The other type of multi-frame consists of 51 TDMA frames. This type of multi-frame is used in BCCH, CCCH, and SDCCH.
- The super frame is a consecutive 51 x 26 TDMA frame. It consists of 51 26-multi-frames or 26 51-multi-frames.

- The hyper-frame consists of 2,048 super frames.

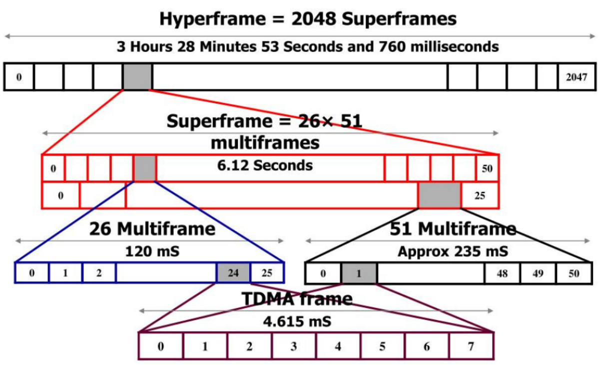

Figure 2.3: GSM Frame Structure[18]

Physical Channel

GSM adopts mixed technology of Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA). GSM features high-frequency utilization.

Frequency-Multiplexing Structure: -Enables 124 carrier frequencies (carriers for short) to be assigned to the uplink (from the MS to the BTS) 890 MHz – 915 MHz or downlink (from the BTS to the MS) 935 MHz – 960 MHz in the GSM900 band. The interval between carriers is 200 kHz. Carriers in the uplink and downlink are in pairs called duplex communication modes. The interval between duplex receiving and transmitting carrier pair is 45MHz. The respective 200 kHz bandwidth is kept as a guard band for the neighbouring systems in the frequency band. If the carrier frequencies on the uplink are denoted by Fu and those on the downlink as Fd then the GSM band can be defined as.

$$Fu\,(n) = 890.2MHz + 0.2\,(n-1)\,MHz \;\; (1 < n < 124) \tag{2.1}$$

$$Fd\,(n) = 935.2MHz + 0.2\,(n-1)\,MHz \;\; (1 < n < 124) \tag{2.2}$$

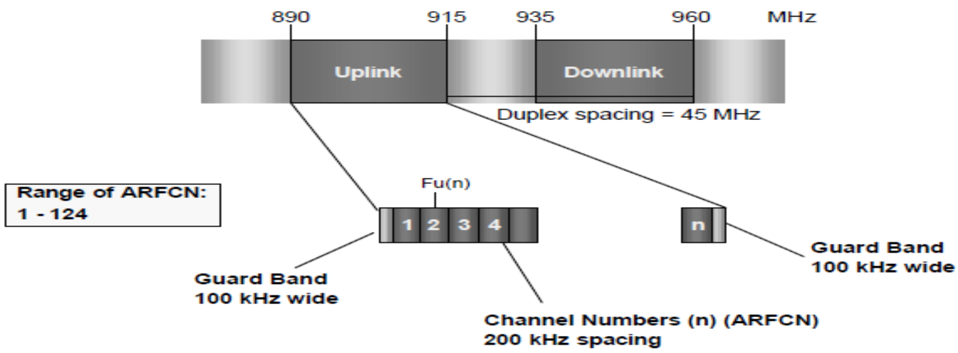

Figure 2.4: FDMA channel[20]

Time-Multiplexing Structure: -enables each carrier of the GSM900 band to be divided into eight-time segments. Each time segment is called a timeslot Figure 2.5. This type of timeslot is called a channel or a physical channel. Eight consecutive timeslots on a carrier constitute a TDMA frame, that is, a carrier of GSM provides eight physical channels. Eight timeslots in the TDMA frame are called physical channels[20].

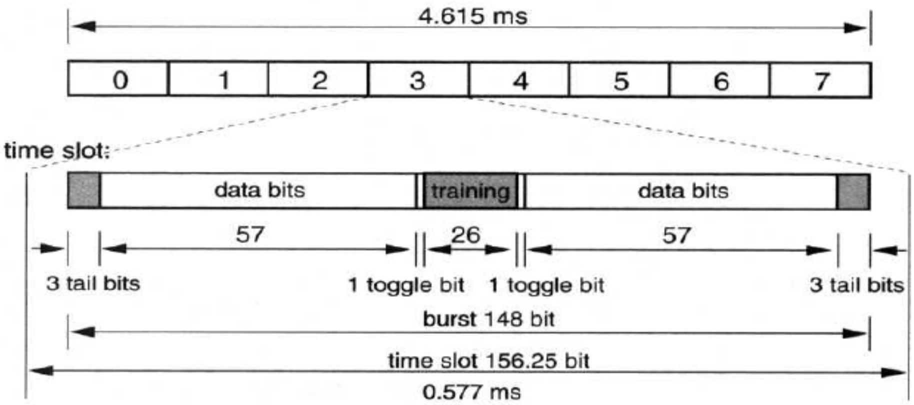

Figure 2.5: Time-Frequency Structure of Physical Channel[20]

Logical Channels

The GSM recommendations define several logical channels for data transmission based on this principle, dividing them into two main groups: traffic channels and control channels.

A. Traffic Channels (TCH)

Traffic Channels are logical channels over which user information is exchanged between mobile users during a connection. Speech and data are digitally transmitted on these channels using different coding methods. Different transmission capacities are required depending on the type of service used (e.g., speech transmission, short-message service, data transfer, facsimile).

Full-Rate Traffic Channels

- Full-rate voice channels: the output data rate of the voice encoder is 13 Kbit/s. Channel coding increases the effective transmission rate to 22.8 Kbit/s.

- Full-rate data channels: the payload data with data rates of 9.6, 4.8, or 2.4 Kbit/s are encoded with Forwarding Error Correction codes and transmitted with an effective data rate of 22.8 Kbit/s.

Half-Rate Traffic Channels

- Half-rate voice channels: voice encoding with a data rate as low as 6.5 Kbit/s is feasible. Channel coding increases, the transmitted data rate to 11.4 Kbit/s.
- Half-rate data channels: payload data with rates of 4.8 or 2.4 kbit/s can be encoded with an FEC code, which leads to an effective transmission rate of 11.4 kbit/s.

B. Control Channels(CCH)

Control channel are basically of two types: Common Control Channels (CCCHs) and Dedicated Control Channels (DCHs)(see in Table 2.1).

Table 2.1: Logical control channels in GSM

Channel	**Abbreviation**	**Function/Application**
Access Grant Channel(DL)	AGCH	Resource allocation is subscriber access authorization
Broadcast Common Control Channel(DL)	BCCH	Dissemination of general information
Cell Broadcast Channel(DL)	CBCH	Transmits the cell broadcast message
Fast Associated Control Channel	FACCH	For user network signaling
Paging Channel(DL)	PCH	Paging for a mobile terminal
Random Access Channel(DL)	RACH	Resource request made by a mobile terminal
Slow Associated Control Channel	SACCH	Used for transport of radio layer parameters
Standalone Dedicated Control Channel	SDCCH	For user network signaling
Synchronization Channel(DL)	SCH	Synchronization of a mobile terminal

2.3 EGPRS System

GSM introduces two new equipment to support GPRS: Serving GPRS Support Node and Gateway GPRS Support Node. BSC is added with Packet Control Units, and concerned BSS software is upgraded. SGSN provides similar functions as MSC. It performs GPRS channel assignment, mobility management, encryption, and charging. GGSN provides various interfaces. It supports interconnection with external Public Data Networks like Internet and X.25, and other PLMNs. Using SGSN and GGSN, operators can construct a GPRS backbone network based on the current transmission network. By reconstructing the current GSM network, operators can easily provide both circuit and packet services, and fully utilize radio resources and network terrestrial resources.

2.3.1 Channels in EGPRS Network

Moving onto the existing channels in the GSM network which are voice channels, there are some new channels in a GPRS network that are related to the packet functionality. These logical channels are allocated to the physical channel called the PDCH (Table 2.2). Channel allocation in the EGPRS (EDGE) network is similar to that of the GPRS network. The PACCH is the only associated channel when physical resources are assigned while the BCC H, PCH, RACH, and AGCH are signaling channels.

Table 2.2: EGPRS channels

Channel	Abbreviation	Function/Application
Packet Broadcast Control Channel(DL)	PBCCH	Broadcast system information specific to packet data
Packet Access Grant Channel(DL)	PAGCH	Notifies that mobile about resource assignment before actual packet transfer
Packet Notification Channel(DL)	PNCH	Used for sending information to multiple mobile stations
Packet Paging Channel(DL)	PPCH	Pages a mobile station before a packet transfer process begins
Packet Random Access Channel(UL)	PRACH	Used by the mobile station for initialization of the uplink packet transfer.
Packet Common Control Channel	PCCCH	Contain logical channels for common control signaling
Packet Data Traffic Channel	PDTCH	Channel temporarily used for data transfer
Packet Associated Control Channel	PACCH	Used for signaling information transfer for a given mobile

Table 2.3: Coding schemes for the GPRS network

Coding Schemes	Code Rate	Date Rate (Kbps)	Date Rate (Kbps) RLC/MAC
CS-1	0.1	9.05	8
CS-2	∼2/3	13.4	12
CS-3	∼3/4	15.6	14.4
Cs-4	1	21.4	20

2.3.2 Coding Schemes

The GPRS network has four coding algorithms (see Table 2.3), which are CS-1, CS-2, CS-3, and CS-4. For forward-error correction, CS-1 employs half-rate convolution coding with a data rate of 9.05 kbps. The CS-2 and CS-3 are similar to the CS-1 but for the introduction of 'puncturing.' This method boosts data throughput at the expense of redundancy. With no FEC in CS-4, the data rate is increased even more. Because the

EGPRS system is an upgrade to the GPRS system, data speeds increase to 59.6 kbps. The first four are still GSMK, whereas the latter five are 8-PSK (see Table 2.4). The modulation used in GPRS systems is GMSK, whereas the modulation used in EGPRS systems is 8-PSK. Aside from the data rates scheme, a few of other elements (link adaptability and incremental redundancy) make EGPRS systems quite appealing[21].

Link adaptation is used in changing environmental conditions to maximize channel throughput while minimizing latency. As a result, the link adaptation features attempt to maintain signal quality under adverse conditions. By automatically adjusting the overall amount of transmitted redundancy to radio channel conditions, incremental redundancy improves throughput. This is accomplished by employing two techniques: ARQ and FEC. The 'power control feature' in EGPRS systems is more difficult to implement than in GSM systems due to the addition of data. Uplink power control is used to reduce interference (and thus mobile battery life), whereas downlink power control is used to reduce BTS power and thus network interference. TBF(Temporary Block Flow) is another concept used in EGPRS networks. [11][12].

Table 2.4: Coding schemes for the EGPRS network

MCS	**Modulation**	**User Rate(Kbps)**
1	GMSK	8.8
2	GMSK	11.2
3	GMSK	14.8
4	GMSK	17.6
5	8-PSK	22.4
6	8-PSK	29.6
7	8-PSK	44.8
8	8-PSK	54.4
9	8-PSK	59.2

2.4 Frequency Spectrum

2.4.1 Introduction to Spectrum

Cellular networks in different ITU areas use different radio frequencies (Americas, Europe, Africa, and Asia). AMPS, which operated in the 800 MHz frequency band, was America's first commercial standard for mobile connections. The first widespread automatic mobile network in Europe's Nordic countries was based on the NMT-450MHz standard, which operated in the 450 MHz band. Mobile service providers struggled to keep up with the growing number of customers as mobile phones became more popular and affordable. They needed to upgrade their existing networks and eventually implement new standards, which were frequently based on different frequencies. TACS at 900 MHz has been adopted as part of the GSM standard by some European countries and Japan. Carriers acquired licenses in the 1,800 MHz band as demand grew[22].

Not all radio frequencies are treated equally, and mobile network operators need access to a variety of frequency bands in order to provide inexpensive, high-quality mobile broadband services with outstanding coverage.The basic harmonized bands for mobile generally fall between the frequency range of 400MHz to 5GHz, with the lower range offering greater coverage and the higher range providing greater capacity[23].

The frequency bands used in mobile networks today have been designated for mobile services globally by the ITU Radio Communication Sector and have been harmonized on a regional or global scale. Before commercial deployment, they are standardized by 3GPP. The most commonly used current bands are listed below. Despite the fact that different countries in different regions have adopted different combinations of those bands, regional and global harmonization has created economies of scale, making mobile services and handsets more affordable.

In general, signals below 1 GHz are lower-frequency and travel further and penetrate structures better. These frequencies are commonly referred to as coverage bands because an operator can serve a larger region with a single base station. These frequencies are ideal for providing low-cost mobile broadband services in rural areas. The amount of spectrum used by a wireless connection for data or voice conversations is known as the channel bandwidth, and At higher frequencies, broader channel bandwidths are more readily available., such as 1.8 GHz and above. These frequencies are frequently referred to as capacity bands. Deploying a network that uses these higher-frequency bands necessitates the installation of more base stations to cover the same area, necessitating additional investment. On the other hand, these bands can accommodate more mobile broadband traffic and better speeds, making them useful in densely populated locations.

Ethiotelecom use both coverage band (<1GHZ) and capacity band (>1GHZ) by statically assigned with services(see Table 2.5).

Table 2.5: Spectrum used by ethiotelecom[5]

Band	UL(MHz)	DL(MHz)	Total BW(MHz)	Current Usage (MHz)	Use
450 MHz	410-430	450-470	20MHz X 2	2.5MHz	CDMA 1x
850 MHz	824-849	869-894	25MHz X 2	8.7MHz	CDMA EV-DO
900MHz	890-915	935-960	25MHz X 2	17.5MHz	UMTS + GSM
1800MHz	1710-1785	1805-1880	75MHz X 2	37.5MHz	LTE + GSM
2100MHz	1920-1980	2110-2170	60MHz X 2	20MHz	UMTS
2600 MHz	2500-2570	2620-2690	70MHz X 2	20MHz	LTE

2.5 Spectrum Refarming

Spectrum redeployment (spectrum refarming) is a combination of administrative, financial, and technical measures aimed at removing users or equipment of the existing frequency assignments either completely or partially from a frequency band. The frequency band may then be allocated to the same or different service(s). These measures may be implemented in short, medium, or long timescales. Spectrum refarming is one of several solutions to improve spectral efficiency. The spectral efficiency (in bit/s/Hz) of a RAT is its ability to transmit a certain throughput in a given amount of spectrum. Spectrum re-farming improves spectral efficiency as new RAT's are more spectral-efficient than the legacy ones, usually due to breakthrough innovations.

Types of spectrum refarming

For frequency reallocation, there are two mainstream methods commonly used,

2.5.1 Static Spectrum Refarming

The SSR refers to a static partitioning of the spectrum being utilized by a legacy network to new technology (e.g. LTE). The bandwidth allocated for a new network is fixed and hence the changes in the traffic demand are not entertained[24].

There are two types of static spectrum refarming:

1. Full Refarming (One-off Refarming).
2. Partial Refarming (Interim Refarming).

Full Refarming (One-off Refarming): - It is appropriate for mobile operators who have well-developed GSM and UMTS networks and abundant spectral resources. The number of GSM/GPRS subscribers is declining as 3G services expand, and these subscribers are migrating to 3G networks, resulting in fewer loads on the GSM network at 1800MHz. As a result, the 1800MHz spectrum can be fully re-farmed for use in LTE networks, while GSM900 is used for all voice services.

Partial Refarming (Interim Refarming): - is suitable for the operators with limited spectral resources who have no UMTS or other networks and have difficulty in subscriber migration or who have a large number of GSM subscribers that will remain stable in the short term. These operators need to consider how to retain the existing subscribers and provide competitive high-speed mobile data access. They may therefore phase in the band refarming to example LTE by 5MHz, 10MHz, or 20MHz spectrum bandwidth.

Partial Refarming is done by two methods:

- The Sandwich Method
- The Edge Allocation Method.

Sandwich Method: - It releases the middle portion of the some spectrum to another band of an operator by some BW for example 5MHz, 10MHz, or 20MHz spectrum bandwidth. Portions on both ends are still used by GSM (see Figure 2.6).

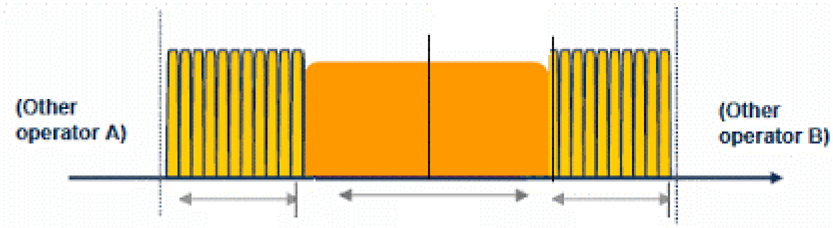

Figure 2.6: Sandwich Method[24]

Edge Allocation Method: - It transfers either end of the operator's band to the other band while keeping the other end for GSM. Given the GSM frequency reuse plan, interference between frequencies, and, in particular, interference with other operators, the sandwich method is recommended. Controlling interference between GSM and other bands, such as LTE, can be done within an operator's frequency band without requiring coordination with neighboring bands of other operators (see Figure 2.7).

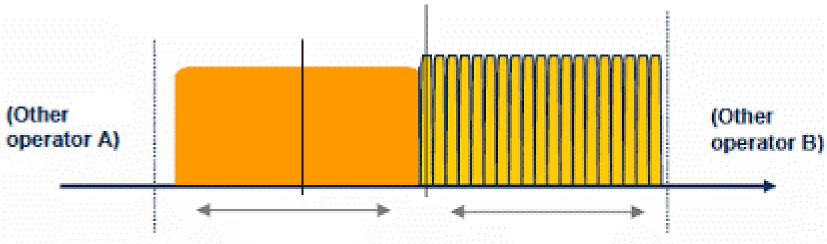

Figure 2.7: Edge Method[24]

2.5.2 Dynamic Spectrum Refarming

The DSR approach allows for the flexible coexistence of GSM, LTE, and UMTS. Another important focus is on improving legacy technologies' spectral efficiency, allowing DSR solutions to allocate maximum spectrum and, thus, maximum data traffic to the most efficient layer. This approach has aided in the delivery of a DSR solution that does not degrade the quality of service provided by legacy technology users—we avoided the temptation to concoct a "blind" spectrum sharing solution to pass off as DSR.

So ethiotelecom use static and from static use partial refarming technique in U900MHz and LTE 1800MHz spectrum.

3 | Time Series

3.1 Introduction to Time Series

We can define a time series as a sequential or chronologically order set of data point. Typically, these data point are measured over successive times. in mathematical notation, it can be expressed as random variable x_t with t denoting the time dimension$(t \geq 0)$. If multiple variables are expressed in a time series at a time ‘t’ then it is a multivariate time series otherwise it is uni variate. The value of x_t can be either continuous or discrete. Mathematical and statistical analysis performed on this kind of data to find hidden patterns and meaningful insight is called time-series analysis. Time-series modeling techniques are used to understand past patterns from the data and try to forecast future horizons[25].

3.2 Characteristics Time-Series

There are four characteristics in time series namely trend, cyclicality, seasonality and irregularity.

Trend:-long-term upward or downward movement.

Cyclicality:- It is not the same as seasonal variation. The cycle is the variation of time series data’s autoregressive component. Cycles occur over longer time periods, such as every 6–10 years, whereas seasonal variation occurs over shorter time periods.

Seasonality:-variation in data caused by seasonal effects.For example Ice cream sales are high in summer, heating oil sales are high in winter but low in summer.

Irregularity:-a component that is left after other components have been calculated and removed from time series data. It is randomly, identically and independently distributed. We can use seasonal decomposition from a statistics model to breakdown the data into its constituent pieces, taking into account whether the series is additive or multiplicative.The additive model works with linear trends of time-series data such as changes constantly over time. The additive model formula is as follows:

$$Y[t] = T[t] + S[t] + c[t] + e[t] \tag{3.1}$$

The multiplicative model works with a nonlinear type of data such as quadratic or exponential. The multiplicative model formula is as follows:

$$Y[t] = T[t] * S[t] * c[t] * e[t] \tag{3.2}$$

3.2.1 Stationarity in Time-Series Data

A stochastic process's stationarity can be interpreted as a type of statistical equilibrium. As a result, the statistical properties of the process are not time dependent. Except for inherent stochastic fluctuation, stationary stochastic models are typically designed so that the mean level, variance, and standard deviation are time independent. The stationarity assumption may reflect reality in addition to reducing the mathematical complexity of a stochastic model[26][27].

We have to make time-series data stationary before fitting a model. We can make time-series stationary by transforming the data. Usually, differencing is used to make the data stationary.

So, how can we test whether a time series data is stationary or not?

1. Visualization of the time series plots and identify trends or seasonality.
2. Divide the data into 3 different sets and calculate the mean and variance for each set and confirm whether the mean and variance for each set are substantially different

or not.

3. The statistical tests.

3.2.2 Techniques of Making Data Stationery

A. Differencing (lag difference)

Is the transformation of a time series in order to stabilize the mean. There are several methods for identifying time series, such as a line plot, which depicts the series over time. Seasonality and random walk can be seen within these trends, indicating a change over time, and this behavior is referred to as a non-stationary time series[28].

The clear rule states that it needs to remove trends and seasonality in the data from the training time series forecasting model. The following are the reasons for differencing:

- To convert non stationary data into a stationary time series.
- To remove seasonal trends.

B. Random Walk

Is the behavior that occurs when the time series exhibits irregular growth. When growth is not constant over time, it is easier to predict the change from one period Y_t to the next Y_{t-1}. Moving from left to right in a different timestamp, the movement that is independent of the variable takes a random step up or down.

So, a change between two consecutive observations in the time series can be written as follows:

$$\hat{Y}_t = Y_t - Y_{t-1} \tag{3.3}$$

When the differenced series contains white noise (ε_t), the formula can be written as follows:

$$Y_t - Y_{t-1} = \varepsilon_t \tag{3.4}$$

$$Y_t = Y_{t-1} + \varepsilon_t \tag{3.5}$$

where ε_t = white noise. This is known as a normal random walk. A random walk represents that a time series is non-stationary. A random walk has mostly long duration's of trend ups and down and uncertain and unpredictable changes. Notes:

- First-order differencing in a time series will remove a linear trend (i.e., differences=1).
- Second-order differencing will remove a quadratic trend (i.e. differences=2).
- In addition, first-order differencing in a time series at a lag equal to the period will remove a seasonal trend.

C. Second-Order Differencing

Is a technique used to make first-order differencing data stationary when the first-order differencing has failed. So, it's necessary to apply second-order differencing to obtain a stationary series.

$$Y_t'' = Y_t' - Y_{t-1}' \tag{3.6}$$

$$= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \tag{3.7}$$

$$= Y_t - 2Y_{t-1} + Y_{t-2} \tag{3.8}$$

This is second-order differencing, Y", which has t-2 values. These are known as double changes in the original series.

D. Seasonal Difference

Is a difference between an observation and the corresponding observation from the previous year.

E. First-Order Differencing For Seasonal Data

Seasonal differencing is similar to first-order differencing in that we are calculating current and previous observation differences for the same season. It can be written as follows:

$$Y_t' = Y_t - Y_{t-s} \tag{3.9}$$

where s is the number of seasons, which is known as lag-s differencing. We can make a difference in observation after s-lag periods. If we observed white noise in seasonal differencing, then the model can be written as follows:

$$Y_t = Y_{t-s} + \varepsilon_t \tag{3.10}$$

When we try to forecast the value with this approach, it is equal to the last observation from that season. This model gives seasonal naive forecasting results.

3.3 Statistics Unit Root Tests

A statistical test relies heavily on data assumptions. We can also make some assumptions about time-series data. To accomplish this, we must determine whether or not a time-series null hypothesis has been rejected. As a result, we can elucidate the result for a specific problem to be meaningful. The following methods are included in statistical unit root tests, each of which includes a unit root stationary test[29].

3.3.1 P-Value

We must deal with the null hypothesis (H0), which is not stationary, and the alternative hypothesis (H1), which is the stationary time series, in the unit root stationary test. The results of this unit root stationary test will be in p-value format. As a result, understanding how to interpret the p-value is critical. The p-value is used to determine the significance of a result. If the p-value is less than the cutoff, we reject the null hypothesis, indicating that the time series is stationary. If the p-value exceeds the threshold but we are unable to reject the null hypothesis, the time series is non stationary. An ADF test looks at the test statistic, the p-value, and the critical values found at 1%, 2.5%, 5%, and 10% confidence intervals. The following shows that if the significance level is set to 0.05, then the corresponding confidence level is 95%[30][31].

- P-value > 0.05 Fail to reject the null hypothesis(H0): - It indicates that data has unique roots and time series is non-stationary.
- P-value < = 0.05 Reject the null hypothesis(H0): - It indicates that data does not have unique roots and time series is stationary.

3.3.2 Augmented Dickey-Fuller Test

The augmented Dickey-Fuller (ADF) test is an important and credible statistical test for stationary checking. It can be used to determine the existence of the unit root in a domain of the series, as well as to determine whether the time series is stationary. The null hypothesis (H0) that the unit root does not exist in a time series is included in the ADF test. For the stationary version of the time series, the alternative hypothesis is used. It is an enhanced version of the Dickey-Fuller test and is best suited for large and complex time-series data sets[32].

This test uses a negative number. The more negative the number is, the higher the chance of rejecting the hypothesis that there is a unit root at some level of confidence.

Null Hypothesis: - Failed to reject, it recommends the time series has a unique root and it has a non-stationary structure, which means time-dependent structure.

Alternative Hypothesis: - Reject null hypothesis, it recommends the time series has not unique root and it has stationary structure, which means not-time dependent structure.

3.3.3 Kwiatkowski-Phillips-Schmidt-Shin Test

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests are used to compare the null hypothesis that the time series is stationary around a conclusive trend to the alternative of a unit root. The presence of a unit root is the alternative hypothesis, which is a reverse version of the ADF test. Furthermore, in a KPSS test, the absence of a unit root is proof of trend stationarity rather than stationarity. This is primarily used for trend stationary tests. Trend stationary data is a series that does not have a unit root but exhibits a trend[33].

Null Hypothesis: - it represents that no unit root and stationary time series.

Alternative Hypothesis: - it represents unit root and non-stationary time series.

3.4 Autocorrelation and Partial Autocorrelation Functions

The ACF and PACF analyses are required to determine a suitable model for a given time series data. These statistical measures reflect the relationship between the observations in a time series. Plotting the ACF and PACF against consecutive time lags is frequently useful for modeling and forecasting purposes. These plots aid in the determination of the order of AR and MA terms. Their mathematical definitions are given below: For a time, seriesx(t), t = 0,1, 2... the Auto covariance at lag k is defined as[33]:

$$\gamma = cov\left(x_t, x_{t+k}\right) = E\left[\left(x_t - \mu\right)\left(x_{t+k} - \mu\right)\right] \tag{3.11}$$

The Auto correlation Coefficient at lag k is defined as:

$$\rho_k = \frac{\rho}{\gamma_o} \tag{3.12}$$

Hereϕ is the mean of the time series, i.e. $\mu = E[x_t]$. The autocovariance at lag zero i.e γ_0 is the variance of the time series. From the definition, it is clear that the autocorrelation coefficient γ_k is dimensionless and so is independent of the scale of measurement. Also, $-1 \leq \rho_k \leq 1$. Statisticians Box and Jenkins termed ρ_k as the theoretical Auto covariance Function (ACVF) andρ_k as the theoretical Autocorrelation Function (ACF). Another measure, known as the Partial Autocorrelation Function (PACF) is used to measure the correlation between an observation k period ago and the current observation, after controlling for observations at intermediate lags (i.e. at lags < k). At lag 1, PACF (1) is the same as ACF (1).

Normally, the stochastic process governing a time series is unknown, so determining the actual or theoretical ACF and PACF values is impossible. Rather, these values are to be estimated using the training data, i.e. the available time series. The estimated ACF and PACF values from the training data are referred to as sample ACF and PACF, respectively[34][35].

As given in, the most appropriate sample estimate for the ACVF at lag k is:

$$c_k = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \mu)\left(x_{t+\mu} - \mu\right) \tag{3.13}$$

Then the estimate for the sample ACF at lag k is given by

$$r_k = \frac{c_k}{c_o} \tag{3.14}$$

Here x(t), t = 0,1,2, ... is the training series of size n with mean ϕ. The sample ACF plot, as explained by Box and Jenkins, is useful in determining the type of model to fit a time series of length N. Because ACF is symmetrical about lag zero, plotting the sample ACF for positive lags from lag one on wards to a maximum lag of about N/4 is all that is required. The sample PACF plot aids in determining an AR process's maximum order.

We use ACF and PACF to determine the best order for the AR(p) and MA(q) values models. Examine the PACF plot for AR order p and select a lag value with a significant correlation factor before correlations become insignificant. Look at the ACF plot and do the same for MA order q. Remember that these values should only be obtained from the ACF and PACF plots of stationary time series, not the above plots. The ACF and PACF plots shown above are the non-stationary original data plots(see in table 3.1).

The AR and MA components of an ARIMA model can be determined using ACF and PACF plots. The ACF and PACF plots can be used to determine both the seasonal and non-seasonal AR and MA components[36][37].

Table 3.1: ACF and PACF to identify the orders of SARMA(p, q) × (P, Q)s

Model	ACF	PACF
AR(p)	Exponentially Decay	Significant till the p lags
MA(q)	Significant till the q lags	Exponentially Decay
AR(P)	Exponentially Decay at each m lag	Significant at Pm lags
MA(Q)	Significant at Qi lags	Exponentially Decay at each m lag

3.5 Auto-Regressive Models

The AR model of this section describes how an observation directly depends upon one or more previous measurements plus an error noise term. This form of a time series model is intuitively appealing and has been widely applied to data set in many different fields. The autoregressive model uses only past behaviour data to forecast the value[38].

AR models use past values to forecast as shown here:

$$\hat{Y}_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t \tag{3.15}$$

Where ε_t is Prediction error at time step t. ϕ Weighting parameter vector for the lagged AR terms This model is known as the AR(p) model, where p is the order for the autoregressive model. The AR model is easy to use to handle a wide range of time-series models.

3.6 Moving Average

The MR model describes how an observation depends upon the current error noise term as well as one or more previous innovations. A moving average forecasts future points by using an average of several past data points. The moving average model practices past forecast errors:

$$\hat{Y}_t = \mu_o + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} \tag{3.16}$$

Where ε_{t-n} is Prediction error at time step t, θ is Weighting parameter vector for the lagged error terms. where$\varepsilon_t$is white noise. This model is known as the MA(q) model, where q is ordered for the moving average model. The MA model is easy to use to handle a wide range of time-series models.

3.7 Autoregressive Integrated Moving Average

Autoregressive integrated moving average also called ARIMA (p, d, q) is a forecasting equation that can make time-series stationary with the help of differencing and log techniques when required. A time series that should be differentiated to be stationary is an integrated (d) (I) series. Lags of the stationary series are classified as autoregressive (p), which is designated in (AR) terms. Lags of the forecast errors are classified as moving averages (q), which are identified in (MA) terms.

$$\hat{Y}_t = \mu + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{3.17}$$

OR

$$(1 - \phi_1 B - ... - \phi_p B^p) Y_t = \mu + \left(1 + \theta_1 B + ... + \theta_1 B + ... + \theta_q B^q\right) \varepsilon_t \tag{3.18}$$

Where μ is constant ϕ_k is the AR coefficient at lag k θ_k is the MR coefficient at lag k

A non-seasonal ARIMA model is called an ARIMA (p, d, q) model, where:

AR: p is the number of autoregressive terms.

I: d is the number of non-seasonal differences needed for stationarity.

MA: q is the number of lagged forecast errors in the prediction equation.

Back shift Notation For ARIMA

ARIMA is a method among several used for forecasting uni-variate variables, which uses information obtained from the variable itself to predict its trend. The variables are regressed on their past values. AR(p) is where p equals the order of autocorrelation (designates weighted moving average over past observations) I (d), where d is the order of integration (differencing), which indicates a linear trend or polynomial trend. MA(q) is where q equals the order of moving averages (designates weighted moving average over past errors). ARIMA is made up of two models: AR and MA.

3.8 Seasonal ARIMA

Seasonal ARIMA (SARIMA) is a technique of ARIMA, where the seasonal component can be handled in uni-variate time-series data. It adjoins three new hyper parameters to lay down AR(P), I(D), and MA(Q) for the seasonality component of a time series. SARIMA allows for the occurrence of seasonality in a series.

$$\Phi_P\left(B^s\right)\phi_p\left(B\right)\left(1-B\right)^d\left(1-B^s\right)^D Y_t = \theta_q\left(B\right)\Theta_Q\left(B^s\right)\varepsilon_t \tag{3.19}$$

where B is the back shift operator (i.e.$By_t = y_{t-1}$), and

$$\Phi_P\left(B^s\right) = 1-\Phi_1 B^s - \Phi_2 s^{2s} - \ldots - \Phi_P B^{Ps} \tag{3.20}$$

$$\phi_p\left(B\right) = 1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p \tag{3.21}$$

$$\theta_q\left(B\right) = 1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q \tag{3.22}$$

$$\Theta_Q\left(B^s\right) = 1-B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs} \tag{3.23}$$

The seasonal ARIMA model combines both nonseasonal and seasonal components in a multiplicative model. The notation can be defined as follows:

ARIMA (p, d, q) X (P, D, Q) m, m is the number of observations per year.

Three trend elements need to be configured. It is the same as the ARIMA model, as shown here:

(p, d, q) is a non-seasonal component, as shown here:

- p: Trend autoregressive order.
- d: Trend differencing order.

- q: Trend moving average order.

(P, Q, D) is a seasonal component. Four seasonal components are not part of the ARIMA model that are essential to be configured.

- P: Seasonal autoregressive order.
- D: Seasonal differencing order.
- Q: Seasonal moving average order.
- m: Timestamp for single-season order.

3.9 SARIMAX

The SARIMAX model is a SARIMA model with external influencing variables, called SARIMAX (p, d, q) (P, D, Q) m (X), where X is the vector of exogenous variables. The exogenous variables perhaps modeled by the multi-linear regression equation are articulated as follows:

$$(1-\phi_1 B)(1-\Phi_1 B^s)(1-B)(1-B^s)Y_t = (1+\theta_1 B)(1+\Theta_1 B^s)\varepsilon_t\left(X_{k,t}\right) \tag{3.24}$$

where $X_{1,t}, X_{1,t} \ldots X_{k,t}$are observations of k number of exogenous variables corresponding to the dependent variable.

3.10 Exponential Smoothing Methods

This is a common method for generating a smoothed Time Series. Whereas past observations in Single Moving Averages are weighted equally, Exponential Smoothing assigns decreasing weights exponentially as the observation gets older. In other words, recent observations are given more weight in forecasting than older ones. In the case of moving averages, the weights assigned to the observations are the same and equal to 1/N. However, one or more smoothing parameters must be determined (or estimated) in exponential smoothing, and these decisions determine the weights assigned to the observations. This technique is more efficient when time-series data moves slowly over time. It incorporates errors, trends, and seasonal factors into the smoothing parameter computation. These components are combined in either an additive or multiplicative manner[39].

3.10.1 Simple Exponential Smoothing

The simplest of the exponentially smoothing methods is naturally called "simple exponential smoothing" (SES). This method is suitable for forecasting data with no clear trend or seasonal pattern.

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + (1-\alpha)^2 Y_{t-2} \tag{3.25}$$

$$l_t = \alpha Y_t + (1-\alpha)l_{t-1} \tag{3.26}$$

where l_t smoothed observation or exponential weighted moving average, Y_t stands for the original observation and α is called the smoothing parameter of level.

3.10.2 Double Exponential Smoothing

Double Exponential Smoothing extended simple exponential smoothing to allow the forecasting of data with a trend. Forecast = estimate level + trend at the most recent time point.

$$\hat{Y}_{t+1} = l_t + kb_t \tag{3.27}$$

$$l_t = \alpha Y_t + (1-\alpha)Y_{t-1}\ 0 \leq \alpha \leq 1 \tag{3.28}$$

$$b_t = \beta(l_t - 1_{t-1}) + (1-\beta)b_{t-1}\ 0 \leq \beta \leq 1 \tag{3.29}$$

Where l_t is estimated of level and b_t is the trend estimated,α is the smoothing parameter for the level and β is the smoothing parameter for trend.

3.10.3 Triple Exponential Smoothing

Triple exponential smoothing is a forecasting method that enforces exponential smoothing three times. This is the extension of double exponential smoothing (Hoilt's) method when seasonality is found in the data This triple exponential smoothing is also known as the Holt-Winters method. Forecasting equation

$$\hat{Y}_{t+1} = l_t + b_t + s_{t-m(k+1)} \tag{3.30}$$

$$l_t = \alpha\,(Y_t - s_{t-m}) + \alpha\,(1-\alpha)\,Y_{t-1}\; 0 \leq \beta \leq 1 \tag{3.31}$$

$$b_t = \beta\,(l_t - l_{t-1}) + (1-\beta)\,b_{t-1}\; 0 \leq \beta \leq 1 \tag{3.32}$$

$$s_t = \gamma(Y_t - l_t - b_{t-1}) + (1-\gamma)\,s_{t-m}\; 0 \leq \gamma \leq 1 \tag{3.33}$$

Where l_t is estimated of level and b_t is the trend estimated and s_t is the seasonal estimated, α is the smoothing parameter for the level and β is the smoothing parameter for trend and γ is the smoothing parameter for seasonality.

3.11 Estimation and Order Selection

3.11.1 Maximum Likelihood Estimation

Once the model order has been identified (i.e., the values of p, d and q), we need to estimate the parameters c ,$\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$. When estimates the ARIMA model, it uses maximum likelihood estimation (MLE). This technique finds the values of the parameters which maximize the probability of obtaining the data that we have observed. For ARIMA models, MLE is very similar to the least squares estimates that would be obtained by minimizing[26].

$$MLE = \sum_{t-1}^{T} e_t^2 \tag{3.34}$$

Where; $e^2$is the error between the actual and the estimated data.

3.11.2 Information Criteria

Akaike's Information Criterion (AIC), which was useful in selecting predictors for regression is also useful for determining the order of an ARIMA model. It can be written as:

$$AIC = -2log(1) + 2(k) \tag{3.35}$$

$$AIC_c = AIC + \frac{2k^2 + 2k}{n - k - 1} \tag{3.36}$$

$$BIC = -2log(1) + \ln(n-)k \tag{3.37}$$

Where K = p+q+1 l = likelihood of the data n = sample size

It is important to note that these information criteria tend not to be good guides to selecting the appropriate order of differencing (d) of a model, but only for selecting the values of p and q. This is because the differencing changes the data on which the likelihood is computed, making the AIC values between models with different orders of differencing not comparable. So we need to use some other approach to choose d, and then we can use the AICc to select p and q.

3.12 Diagnostic Checking

The "residuals" in a time series model are what is left over after fitting a model. For many (but not all) time series models, the residuals are equal to the difference between the observations and the corresponding fitted values:

$$\varepsilon_t = Y_t - \hat{Y}_{t-1} \tag{3.38}$$

Residuals are useful in checking whether a model has adequately captured the information in the data. A good forecasting method will yield residuals with the following properties:

Assumptions

1. ε_t uncorrelated. If they are n't, then information left in residuals that should be used in computing forecasts.

2. ε_t have mean zero. If they don't, then forecasts are biased.

Useful properties (for Forecast intervals)

1. ε_t have constant variance.

2. ε_t are normally distributed.

White noise data is uncorrelated across time with zero mean and constant variance. (Technically, we require independence as well.) Think of white noise as completely uninteresting with no predictable patterns. In addition to looking at the ACF plot, we can also do a more formal test for auto-correlation by considering a whole set of r_k auto-correlation on lag k values as a group, rather than treating each one separately.

For uncorrelated data, we would expect each auto-correlation to be close to zero. Sampling distribution of$r_k$for white noise data is asymptotically N(0,1/T). T is sample size.

- 95% of all r_k for white noise must lie within$\pm\frac{1.96}{\sqrt{T}}$.
- If this is not the case, the series is probably not WN.
- Common to plot lines at $\pm\frac{1.96}{\sqrt{T}}$ when plotting ACF. These are the critical values.

3.13 Evaluation Metrics

Create a function that has all the required evaluation metrics, which will give us the results in one go. This function helps us understand how far off our forecasts are against the actual.

Mean squared error (MSE) tells you how close a regression line is to a set of points. It does this by taking the distances from the points to the regression line (these distances are the errors) and squaring them. The closer to zero the error is, the better the model.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{3.39}$$

Mean absolute error (MAE) measures the average magnitude of the errors in a set of predictions, without considering their direction. The closer to zero the error is, the better the model.

$$MAE = \frac{\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|}{n} \tag{3.40}$$

Root mean square error (RMSE) is a quadratic scoring rule that also measures the average magnitude of the error. It's the square root of the average of squared differences between prediction and actual observation. The closer to zero the error is, the better the model.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n}} \tag{3.41}$$

Mean absolute percentage error (MAPE) is a statistical measure of how accurate a forecast system is. It is a measure in terms of percentage. It is mostly used for time-series forecasting. The closer to zero the error is, the better the model.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \tag{3.42}$$

R-squared determines the proportion of variance in the dependent variable that can be explained by the independent variable. The closer to zero the error is, the better the model.

$$R - Squared = \frac{SS_{regression}}{SS_{total}} \tag{3.43}$$

Where y_i: - Actual value, $\hat{Y}_i$: - prediction value, n : - Total sample data, $SS_{regression}$: - is the sum of squares due to regression (explained sum of squares), and SS_{total} : - is the total sum of squares.

4 | Result and Discussion

4.1 The Data

The information is derived from the ethiotelecom Performance Reporting System (PRS). The research region is Addis Ababa Bole BSC 05, and the spectrum chosen to explore is the 1800MHz and 900MHz spectrum assigned to ethiotelecom for GSM service. The data was collected hourly from October 18, 2020, to February 14, 2021, a period of four months, and a total of 789 cells were included in the study. Table 4.1 and Figure 4.1 show the detailed spectrum and the study area.

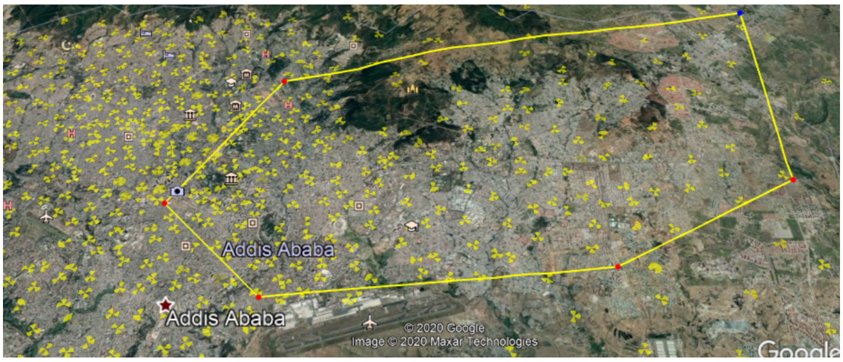

Figure 4.1: Google view of the study area

Table 4.1: Show the detailed spectrum channel assigned to GSM

Spectrum	**DL in MHz**	**Spectrum assigned to GSM**	**Bandwidth**	**Channel assigned**	**Study area**	
					No site	No cell
900MHz	935 - 960	947.6MHz - 959.8MHz	12MHz	61	134	402
1800MHz	1805 - 1880	1842.6MHz – 1851.4MHz	8.8MHz	44	129	387
		1871.6MHz - 1879.8MHz	8.2MHz	41		

After data pre-processing with Excel and Python, the two spectrum detail statically described as below Table 4.2. To visualization spectrum occupancy of both 1800MHz and

900MHz let plot with python in show Figure 4.2 to 4.5

The spectrum occupancy data is time-series data the is a correlation between lag orders figure show correlation lags order 1 and 24 have a strong correlation in 24 lags order seasonal period shown in Figure 4.6.

Table 4.2: Descriptive Statics of Data

Descriptive statics	**1800MHz**	**900MHz**
count	2880	2880
Mean	26.59368465	17.85808223
median	35.6135044	22.78355167
Maximum	60.35384534	44.70626258
Minimum	0.422016012	0.424584404
Standard deviation	17.88005539	11.7100353
Variance	319.6963808	137.1249268
5-th percentile	0.693050217	0.631371651
95-th percentile	48.37223578	34.07332064
Kurtosis	-1.493688107	-1.284137967
Skewness	-0.377943746	-0.286091652
Median Absolute Deviation (MAD)	9.97311089	7.36987093

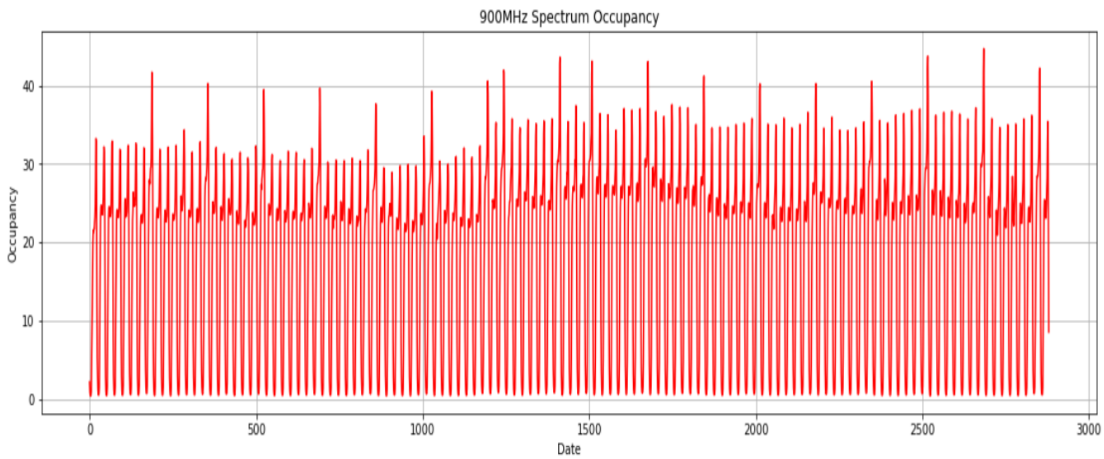

Figure 4.2: Spectrum Occupancy 900MHz plot four months data

And also the distribution of data is a plot in Figure 5.5

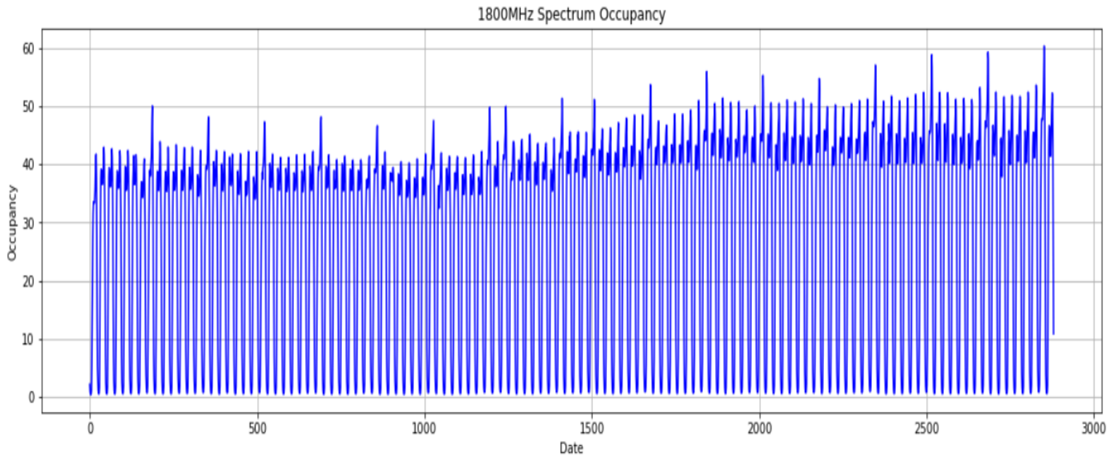

Figure 4.3: Spectrum Occupancy 1800MHz plot four months data

Figure 4.4: Spectrum Occupancy1800MHz and 900MHz plot one month

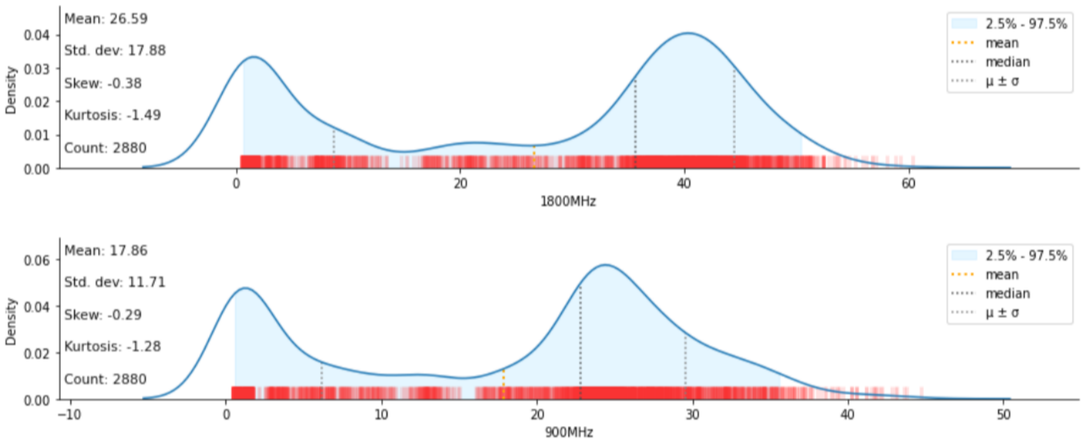

Figure 4.5: Distribution plot of 1800MHz and 900MHz spectrum data

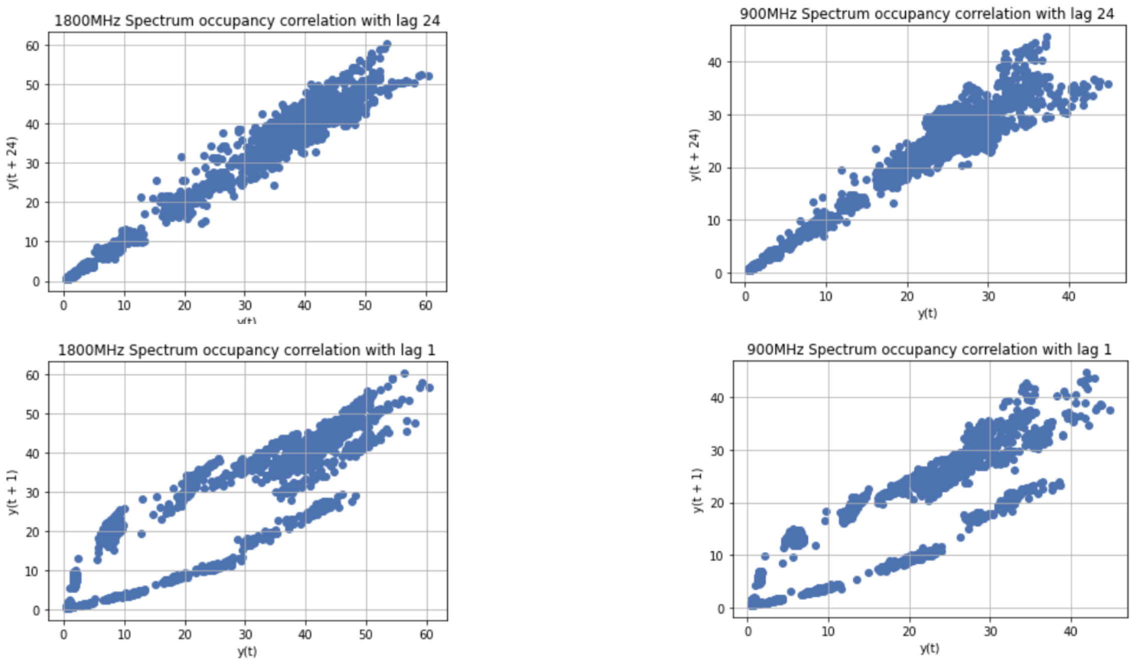

Figure 4.6: Lag 1 and 24 Correlation of 1800MHz and 900MHz

4.2 Check for Stationarity

Most time series models require stationary data. If a time series' statistical features, such as mean, variance, and covariance, remain constant across time, it is said to be stationary. Plotting the data, performing a visual analysis, and using a statistical test are the formal approaches to verify this. Time series decomposition is a statistical activity that divides a time series into numerous components, each representing one of the underlying groups of patterns.

Visual: - We may apply the decomposition method, which allows us to individually view seasonality in both data sets daily, and the trend in 1800MHz spectrum data is increasing, while the trend in 900MHz spectrum data is relatively horizontal, as shown in Figures 4.7 and 4.8.

The plot above shows the seasonal component of the data, as well as the data's independent upward and downward trend. The graphic demonstrates that the data exhibits both a weak trend and substantial seasonality. That suggests it is very likely not stationary.

Plotting Rolling Statistics - We can examine the moving average or moving variance to determine if it varies over time. Moving average and variance mean that we will take

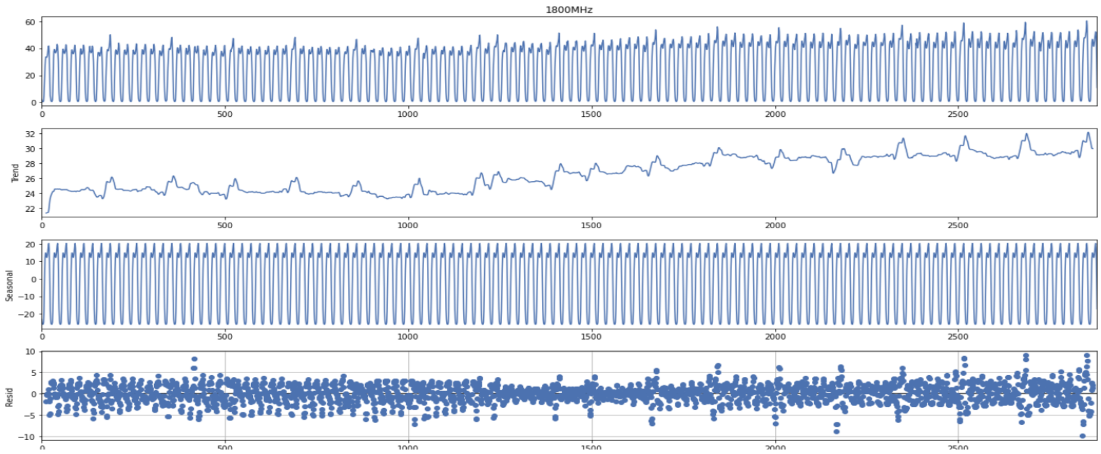

Figure 4.7: 1800MHz occupancy decomposition

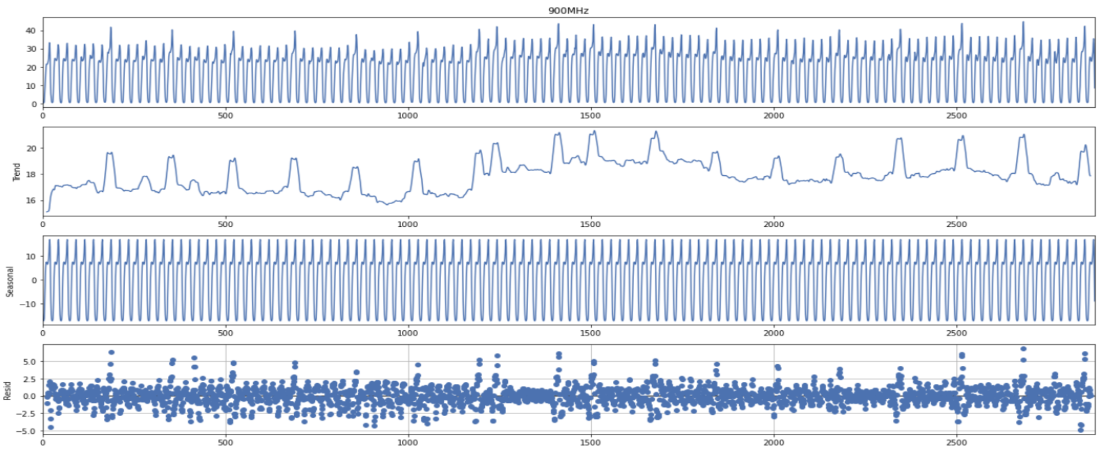

Figure 4.8: 900MHz occupancy decomposition

the average and variance of the hourly-based data at any moment "to," but this is more of a visual strategy. Figures 4.9 and 4.10 demonstrate the rolling and standard tests of stationary.

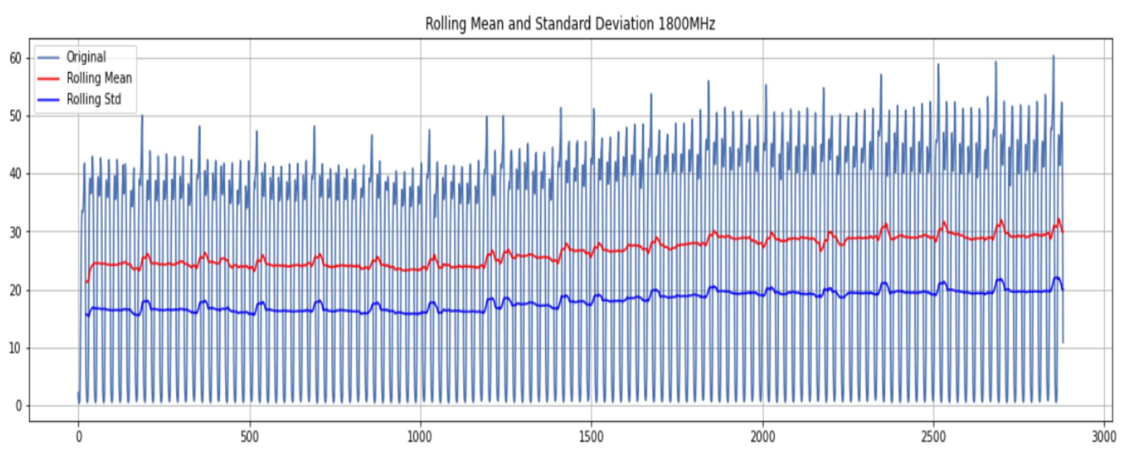

Figure 4.9: Rolling mean & std stationary test 1800MHz occupancy data

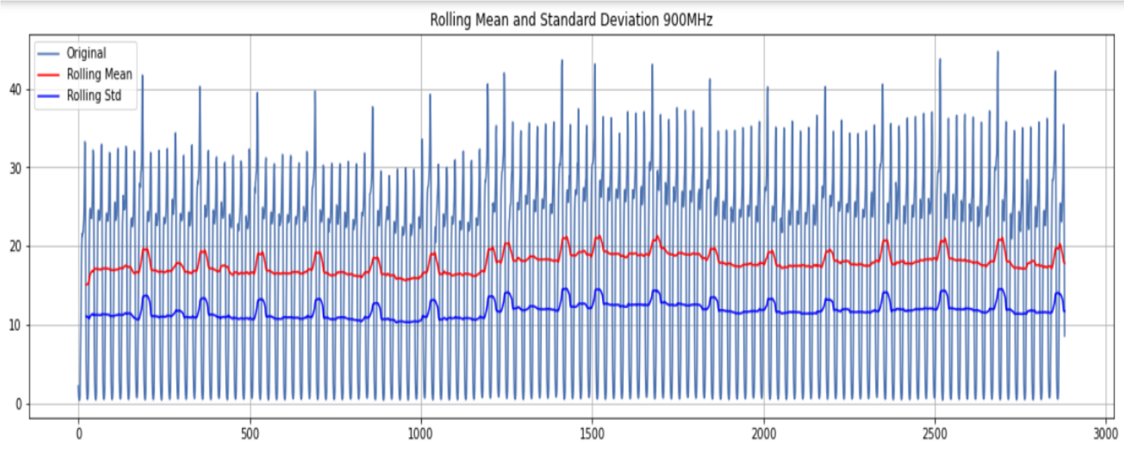

Figure 4.10: Rolling mean & std stationary test 900MHz occupancy data

We can see that the mean and standard deviations have slight variations with time in the 1800MHz spectrum data but are more or less consistent with the time in the 900MHz spectrum data.

Statistical test: We will use the Augmented Dickey-Fuller and KPSS Hypothesis Testing (see Tables 4.3 and 4.4) to confirm our visual observation on the above plot.

Null Hypothesis: The series is not stationary.

Alternate Hypothesis: The series is stationary.

Table 4.3: ADF test with test-statistic and p-value

Descriptive statics	**1800MHz**	**900MHz**
Test Statistic	-2.774202	-6.85E+00
p-value	0.062054	1.74E-09
No Lags Used	25	26
Number of Observations Used	2854	2853
Critical Value (1%)	-3.432643	-3.43E+00
Critical Value (5%)	-2.862553	-2.86E+00
Critical Value (10%)	-2.567309	-2.57E+00

The null hypothesis cannot be rejected based on the significance level of 0.05 and the p-value of the ADF test. As a result, at 1800MHz, the series is non-stationary, and the null hypothesis can be rejected. As a result, the series is stationary at 900MHz. Furthermore, the test statistic value of 10% of the critical value of 1800MHz spectrum data indicates that the data is stationary, and all critical values of 900MHz spectrum data below the test statistic value indicate that the data is stationary.

Table 4.4: KPSS test with test-statistic and rejection value

Descriptive statics	**1800MHz**	**900MHz**
Test Statistic	9.231523	2.915155
p-value	0.01	0.01
No Lags Used	24	23
Critical Value (1%)	0.739	0.739
Critical Value (2.5%)	0.574	0.574
Critical Value (5%)	0.463	0.463
Critical Value (10%)	0.347	0.347

There is evidence for rejecting the null hypothesis in favour of the alternative based on the significance level of 0.05 and the p-value of the KPSS test. As a result, the series is non-stationary according to the KPSS test in both the 1800MHz and 900MHz spectrum occupancy data.

It is always better to apply both the tests so that it can be ensured that the series is truly stationary. Possible outcomes of applying these stationary tests are as follows:

Case 1: Both tests conclude that the series is not stationary - The series is not stationary.
Case 2: Both tests conclude that the series is stationary - The series is stationary.

Case 3: KPSS indicates stationarity and ADF indicates non-stationarity - The series is trend stationary. The trend needs to be removed to make series strict stationary. The detrended series is checked for stationarity.

Case 4: KPSS indicates non-stationarity and ADF indicates stationarity - The series is different stationery. Differencing is to be used to make series stationery. The differenced series is checked for stationarity.

Here, due to the difference in the results from the ADF test and KPSS test, it can be inferred that the series is different stationery and not strictly stationary. The series can be detrended by differencing or by model fitting.

4.3 Autoregressive and Moving Average Orders

This section will determine the appropriate autoregressive orders p and P, as well as moving average orders q and Q. The first step is to visually inspect the correlogram for each well-differentiated process. The results in Tables 4.5 and 4.6 will be utilized as a guideline for what orders the autocorrelation function (ACF) and partial autocorrelation function (PACF) indicate. The findings will be shown for each spectrum occupancy, followed by the ordering chosen using the log-like likelihoods ratio and the information criterion specified as AIC and BIC.

The Dickey-Fuller test rejected the null hypothesis that a unit root is present after the first difference (d = 1) in the 1800MHz spectrum data. Figure 4.11. We can see from the plot that there is no discernible trend, hence the mean is constant and the variance is quite stable. The only issue remaining is seasonality, which must be dealt with before modelling. To do so, consider the "seasonal difference" shown in Figure 4.13, which is a straightforward subtraction of the series from itself with a lag equal to the seasonal period, which is 24. The ACF and the PACF It's a lot better now that the obvious seasonality has gone. The autocorrelation function, on the other hand, still has too many substantial lags. To get rid of them, we'll use first-order seasonal differences and subtract the seasonal difference from itself with lag 1. Figure 4.15.

Figures 4.11 and 4.15 show the ACF and PACF of the 1800 MHz spectrum occupancy.The clear importance of both functions in the first three lags and the subsequent non-The next lag suggests that p and q could be equal, but it is also difficult to discern with mixed processes beyond that. Seasonality is also shown because every 24 lag is significant for the autocorrelation function, as are the 24 and 48 lags in the case of partial autocorrelation. For the seasonal orders, the clear indicators of every 24 lag for the ACF and the PACF indicate that at least one of P and Q should be greater than one.

Initially, the Dickey-Fuller test rejected the null hypothesis that a unit root test was performed on 900 MHz spectrum data. As a result, there is no discernible relationship between the ACF and the PACF in Figure 4.12. The ACF has a distinct sine wave pattern with period 24, which corresponds to seasonal data with period 24. As a result, seasonal difference Figure 5.14 is required after the obvious seasonality has vanished. The autocorrelation function, on the other hand, still has too many substantial lags. To get rid of them, we'll use first-order seasonal differences, as shown in Figure 4.16, by subtracting the series seasonal difference from itself with lag 1. The sine wave pattern is weaker in the PACF, but it is significant during the 3, 6, and 24 lags before it fades. This implies that the seasonal lag orders for Q should be one greater than those for P. It is difficult to tell anything about potential orders p and q, although at least one number greater than zero should exist. It is difficult to predict which orders should be filled, but the procedure should include both autoregressive and moving average orders. Now that we know how to set the initial parameters, let's have a look at the final plot once again and set the parameters:

- p is most probably 2 for both 900MHz and 3 for 1800MHz since it is the last significant lag on the PACF, after which, most others are not significant.
- d equals 0 for 900MHz 1 for 1800MHz because we had first differences.
- q should be somewhere around max 3 for both as well as seen on the ACF.
- P might be 2, for both since 24^{th} and 48^{th} lags are somewhat significant on the PACF.
- D again equals 1for both because we performed seasonal differentiation.
- Q is probably maxed 3 for both The 24^{th} and 48^{th} lag on ACF is significant.

Based on log-likelihood value and Information criteria (AIC and BIC) select the best fit model parameter p, d, P, Q.

Table 4.5: SARIMA(p, d, q) (P, D, Q)24

Spectrum	Model	LLR	AIC	BIC
900MHz	SARIMA(1, 0, 2)x(1, 1, 2, 24)	-2111.54	4237.072	4277.195
1800MHz	SARIMA(2, 1, 2)x(2, 1, [1, 2, 3], 24)	-2544.06	5108.115	5165.43

Table 4.6: SARIMAX (p, d, q) (P, D, Q)24

Spectrum	Model	LLR	AIC	BIC
900MHz	SARIMAX(1, 0, 2)x(1, 1, [1], 24)	-812.931	1639.861	1679.985
1800MHz	SARIMAX(1, 0, 3)x(1, 1, [1, 2], 24)	-1315.94	2649.881	2701.468

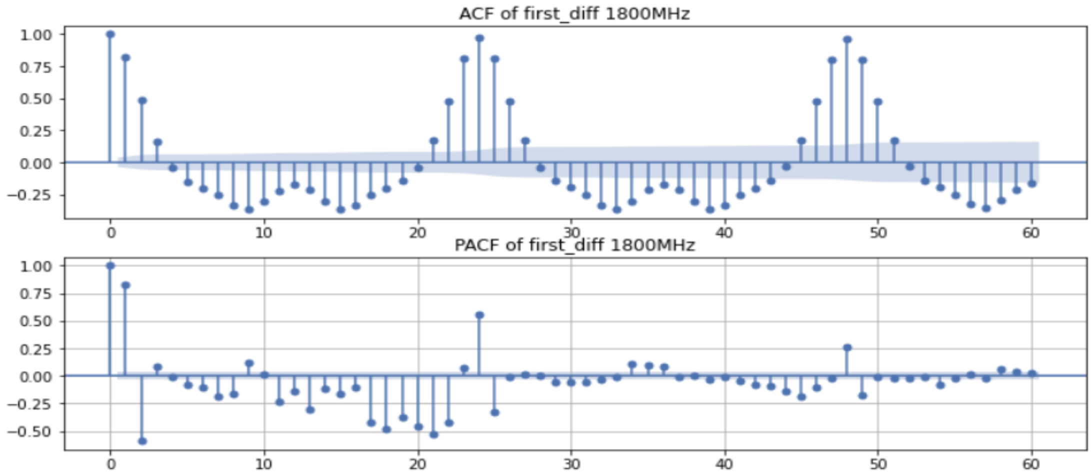

Figure 4.11: ACF and PACF first difference of 1800MHz spectrum data

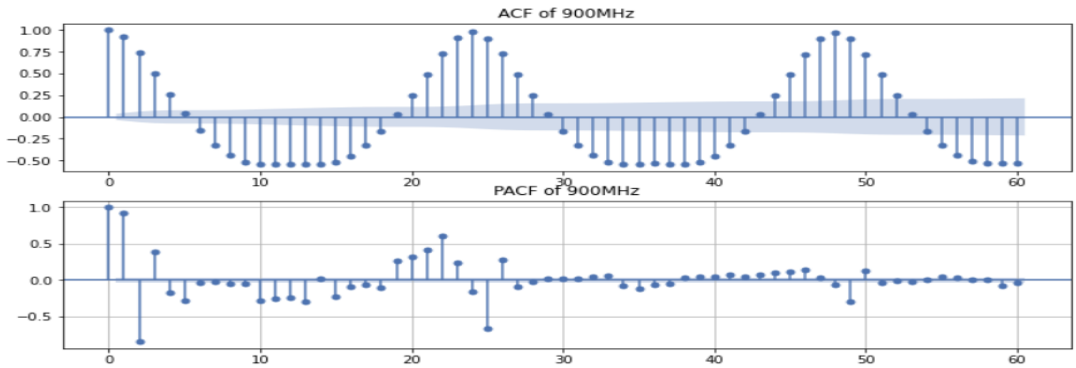

Figure 4.12: ACF and PACF of 900MHz spectrum data

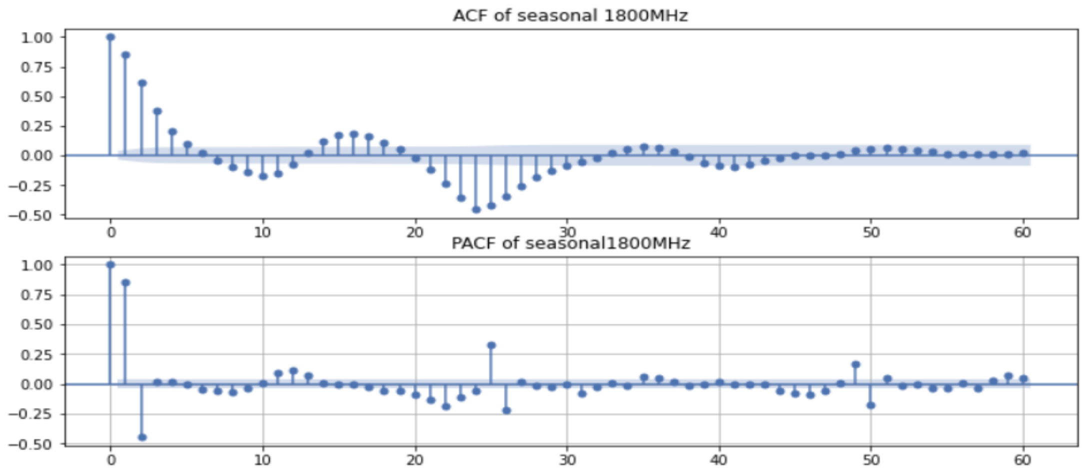

Figure 4.13: ACF & PACF of seasonal difference of 1800MHz

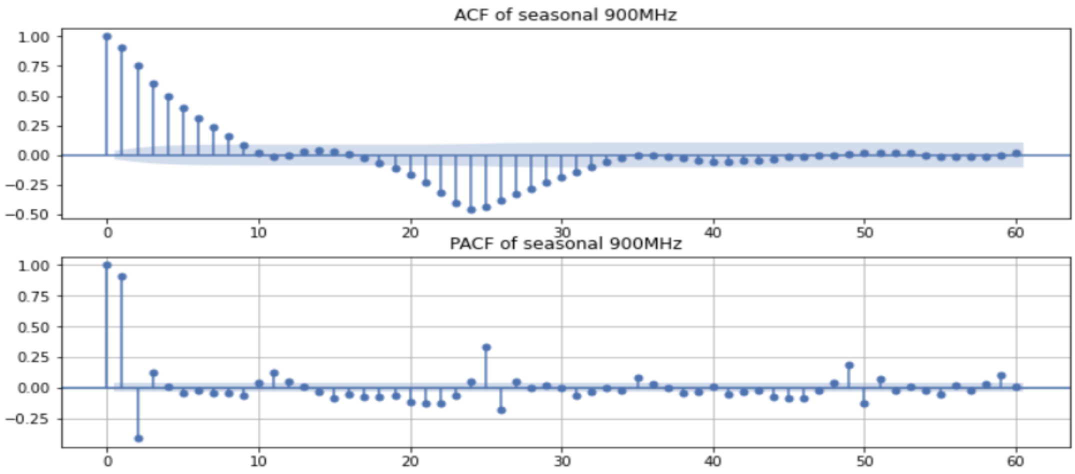

Figure 4.14: ACF & PACF of seasonal difference 900MHz

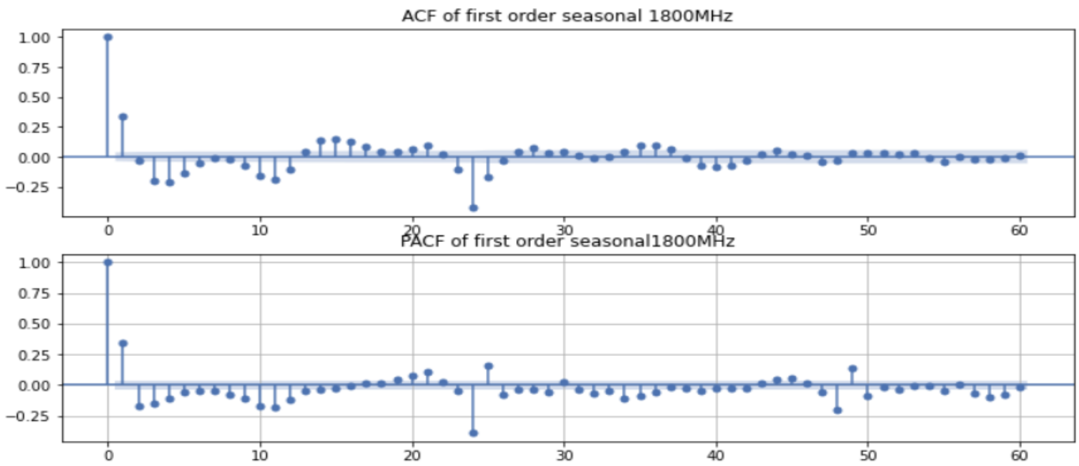

Figure 4.15: ACF & PACF of First-Order Differencing seasonal

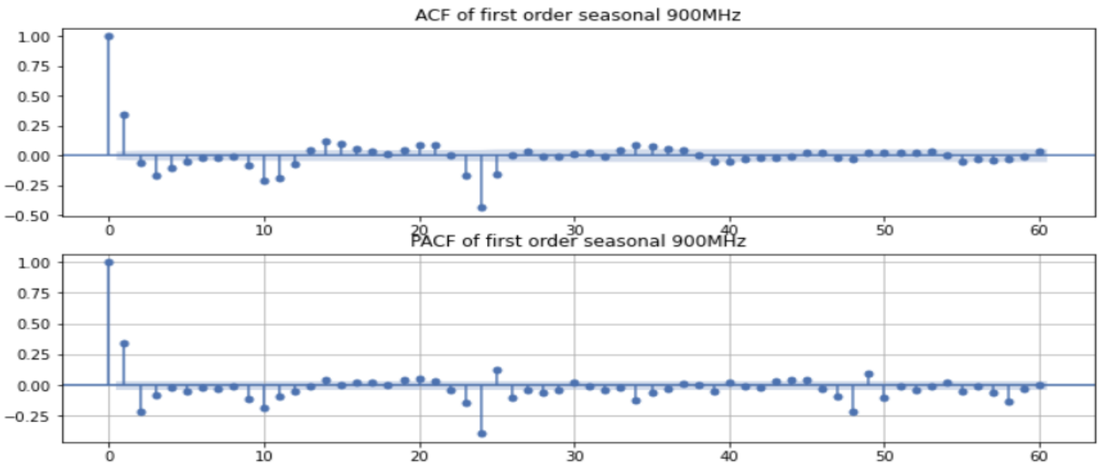

Figure 4.16: ACF & PACF of First-Order Differencing seasonal 900MHz

4.4 Triple Exponential Smoothing Holt-Winters

Intercept (level), trend, and seasonality are the three components of the triple exponential smoothing content. Seasonal components in the model will explain recurring variations around intercept and trend, and the length of the season, in other words, the period after which the variations repeat, will be provided. A separate component is for each observation in the season; for example, if the length of the season is 24 hours (daily seasonality), we will have 24 seasonal components, one for each hour of the day and an initial value for the intercept (level). A trend is also shown in Table 4.8.

Based on Sum Square Error (SSE) and Information Criteria fitted model selected after checking so many parameter possibilities Table 4.7.

Table 4.7: Prediction Evaluation metrics for Model Holt-Winters

Spectrum	**Model**	**SSE**	**AIC**	**BIC**	**AICc**
900MHz	Holt-Winters	1352.66	-1169.05	-1002.52	-1168.2
1800MHz	Holt-Winters	2371.773	124.795	291.325	125.668

Table 4.8: Hot-Winters fitted parameter

parameter	**Coefficient**	
	1800MHz	**900MHz**
α	0.8590816	0.9428782
β	1.1876e-08	2.3933e-09
γ	0.1409184	0.0571218
l_o	6.9228650	5.6123424
b_o	-0.8774664	-0.3773654

4.5 Estimation and Diagnostic Checking of Models

Tables 4.9, 10, 11, and 12 show the fitted SARIMA and SARIMAX models. The next stage is to extract the residuals and then undertake a visual inspection of each model before running various statistical tests that take into account their normality and randomness. Figures 4.17, 19, 21, and 23 show normalized residuals plots in time series analysis.

The residuals for all the Spectrum occupancy data appear to be highly homoscedastic, with a mean near to zero and a p-value of zero. Furthermore, in both spectrum data sets, the number of standardized residuals that lie outside the interval of 1.96 for both the SARIMAX and SARIMA models is less than 8%.

Figures 4.18, 20, 22, and 24 show the autocorrelation functions for the residuals. The values for the first 60 lags are reported, and the number of significant lags defined as those that fall outside the sample size-dependent interval $\pm\frac{2}{\sqrt{T}}$ is of interest. The significant 5 lags out of 60 lags in the SARIMA model are 8.3 percent, but the remaining 91.7 % are inside the sample site. On the other hand, in the SARIMAX model, 3 percent of the significant level 2 lags out of 60 lags are outside the sample size, but 97 % are within the sample size in both the 900MHz and 1800MHz data sets. This result is acceptable:

Table 4.9: SARIMA (1, 0, 2) x (1, 1, 2,)24 fitted for 900MHz

Parameter	**Coefficient**	**Standard error**
ϕ_1	0.8074	0.01
θ_1	0.5097	0.019
θ_2	0.1144	0.019
Φ_1	-0.8613	0.048
Θ_1	0.0612	0.044
Θ_2	-0.7586	0.035

Table 4.10: SARIMA (2, 1, 2) x (2, 1, 3)24 fitted for 1800MHz

Parameter	**Coefficient**	**Standard error**
ϕ_1	1.1471	0.036
ϕ_2	-0.3771	0.032
θ_1	-0.8831	0.038
θ_2	-0.1093	0.038
Φ_1	-1.0074	0.041
Φ_2	-0.7012	0.039
Θ_1	0.2682	0.035
Θ_2	-0.1213	0.031
Θ_3	-0.7214	0.025

Table 4.11: SARIMAX (1, 0, 2) x (1, 1, 1)24 fitted for 900MHz

Parameter	**Coefficient**	**Standard error**
ϕ_1	0.8834	0.008
θ_1	0.3304	0.015
θ_2	0.0592	0.017
Φ_1	0.0813	0.021
Θ_1	-0.8311	0.011

Table 4.12: SARIMAX (1, 0, 3) x (1, 1, 2)24 fitted for 1800MHz

Parameter	**Coefficient**	**Standard error**
ϕ_1	0.7803	0.019
θ_1	0.4363	0.021
θ_2	0.2038	0.028
θ_3	0.0697	0.026
Φ_1	-0.8493	0.046
Θ_1	0.0944	0.04
Θ_2	-0.7279	0.03

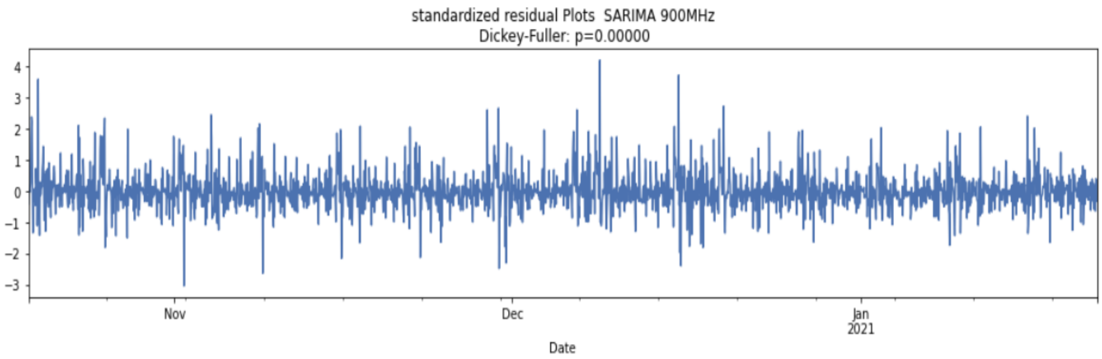

Figure 4.17: Standardized residual SARIMA 900MHz

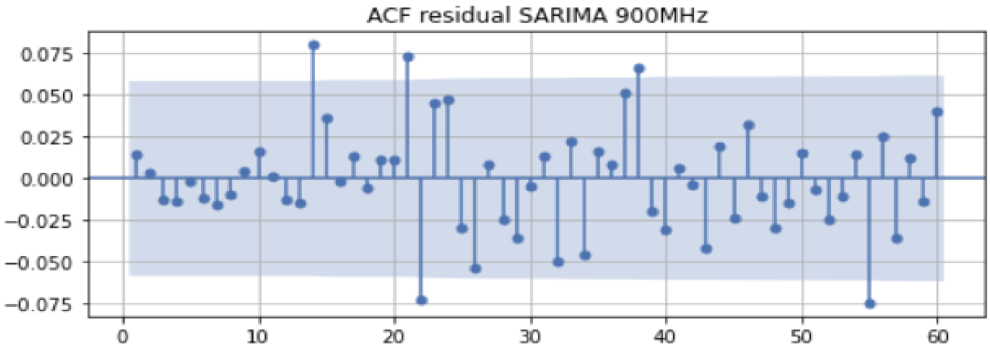

Figure 4.18: ACF plot of residual SARIMA 900MHz

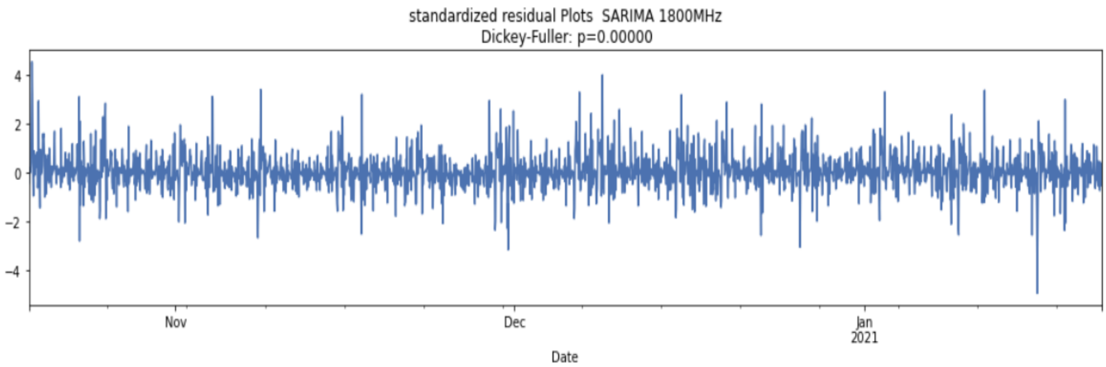

Figure 4.19: Standardized residual SARIMA 1800MHz

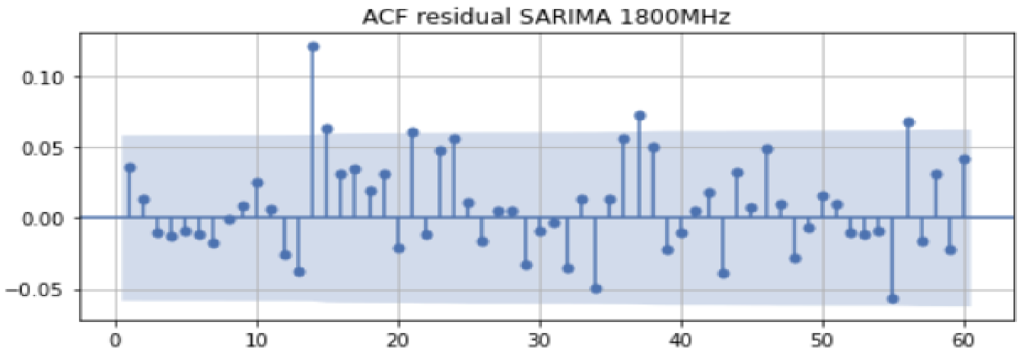

Figure 4.20: ACF plot of residual SARIMA 1800MHz

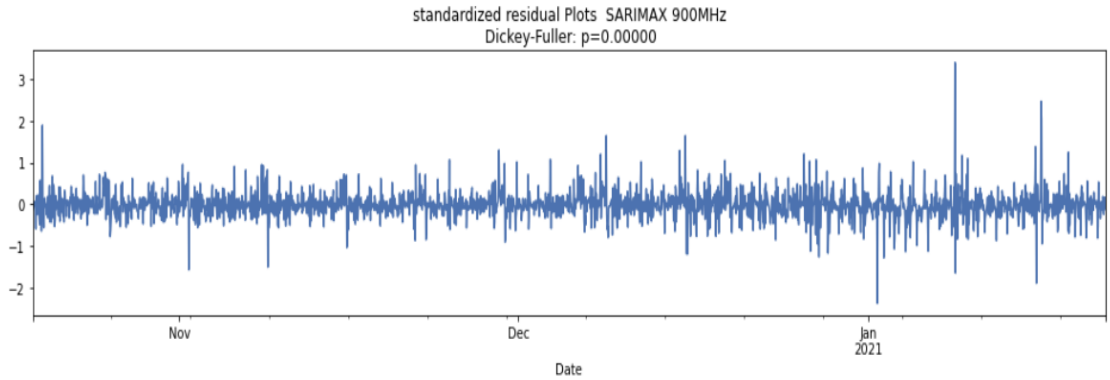

Figure 4.21: Standardized residual SARIMAX 900MHz

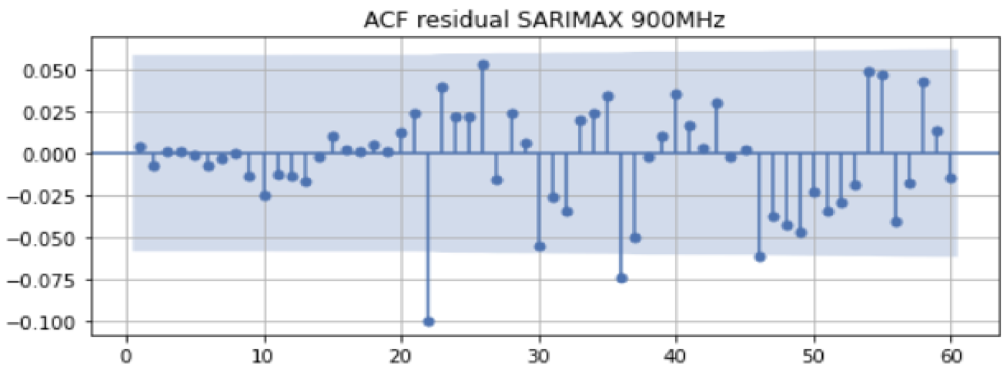

Figure 4.22: ACF plot of residual SARIMAX 900MHz

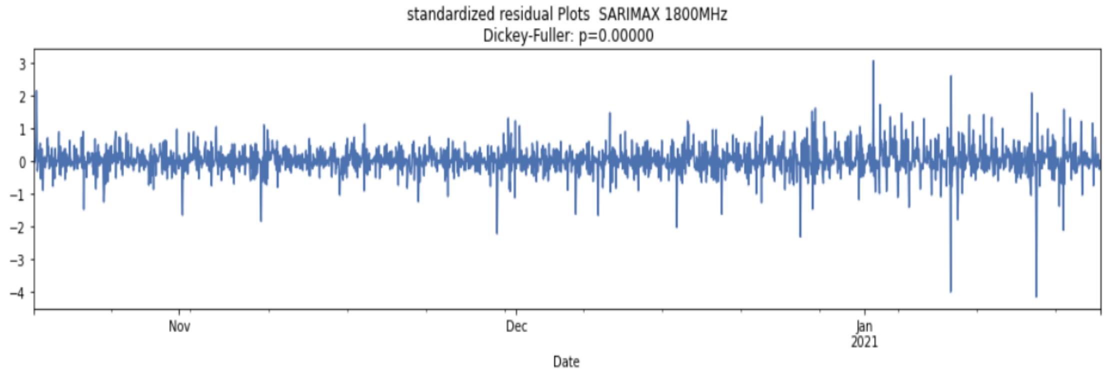

Figure 4.23: Standardized residual SARIMAX 1800MHz

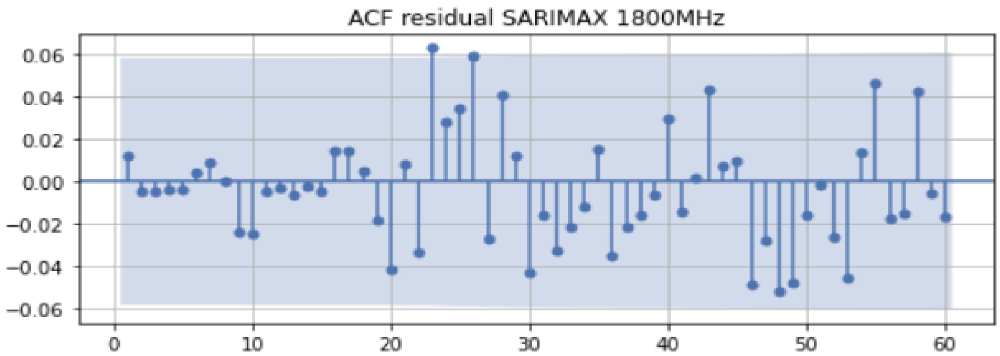

Figure 4.24: ACF plot of residual SARIMAX 1800MHz

4.6 Prediction Comparison

Figures 4.25, 26, 27, 28, 29, and 30 show the model prediction of spectrum occupancy at 1800 MHz and 900 MHz versus the actual results. All of the models appear to be well-fitted, with the forecast closely following the actual values. We employ test assessment metrics such as MSE, MAE, MAPE, RMSE, and R-Squared.

The evaluation metrics for the Holt-Winters models, SARIMA models, and SARIMAX models are listed in Table 4.13. The R squared statistics, which are close to one for each model, demonstrate that the models are well-fitted.

When the models are compared further, it can be shown that the error measures for all spectrum occupancy measurements are marginally smaller for the SARIMAX models than for the SARIMA and Holt-Winters models. The SARIMAX model appears to be more accurate at prediction in Table 4.14.

Demonstrate that the SARIMAX models' assessment metrics are smaller than those of the SARIMA and Holt-Winters models, implying that the tests are one-sided with the alternative hypothesis that the SARIMAX model is more accurate for both the 900MHz and 180MHz spectrum occupancy.

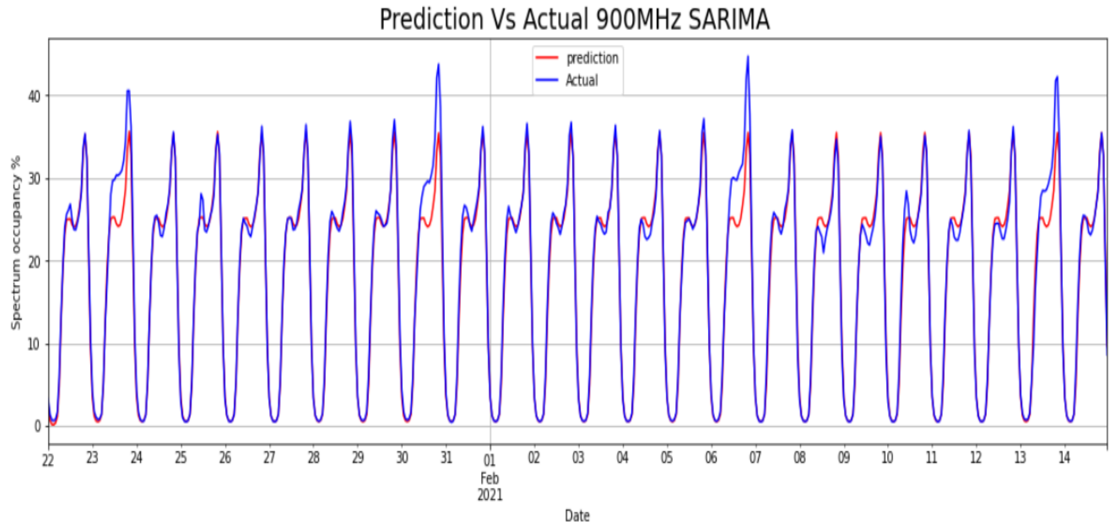

Figure 4.25: prediction vs actual SARIMA 900MHz

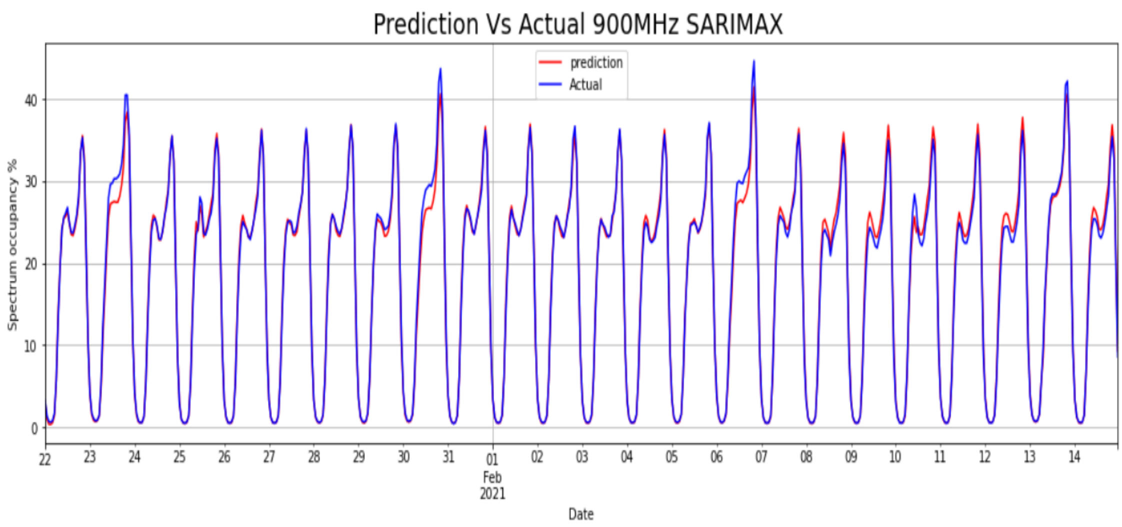

Figure 4.26: Prediction vs Actual SARIMAX 900MHz

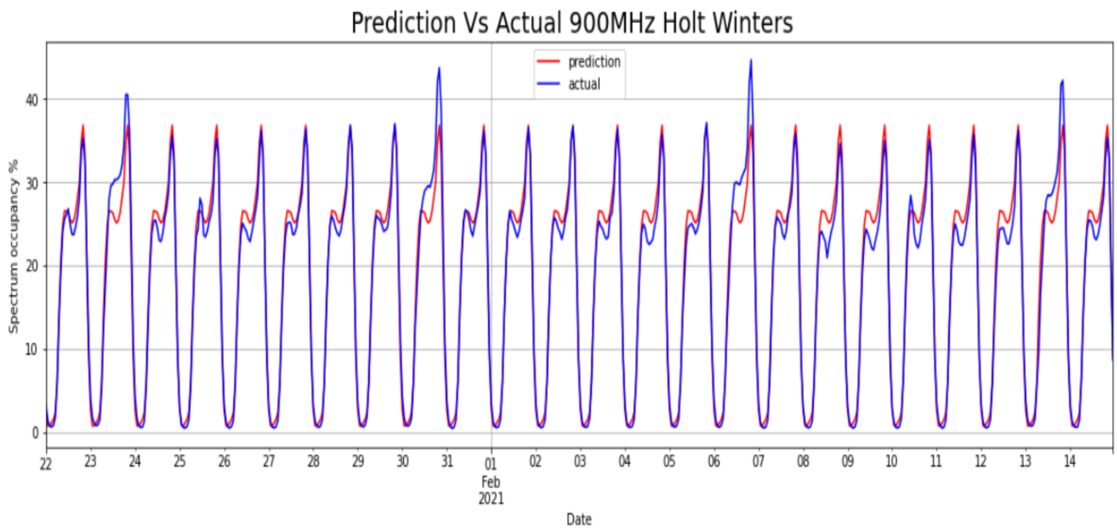

Figure 4.27: Prediction vs Actual holt-winters 900MHz

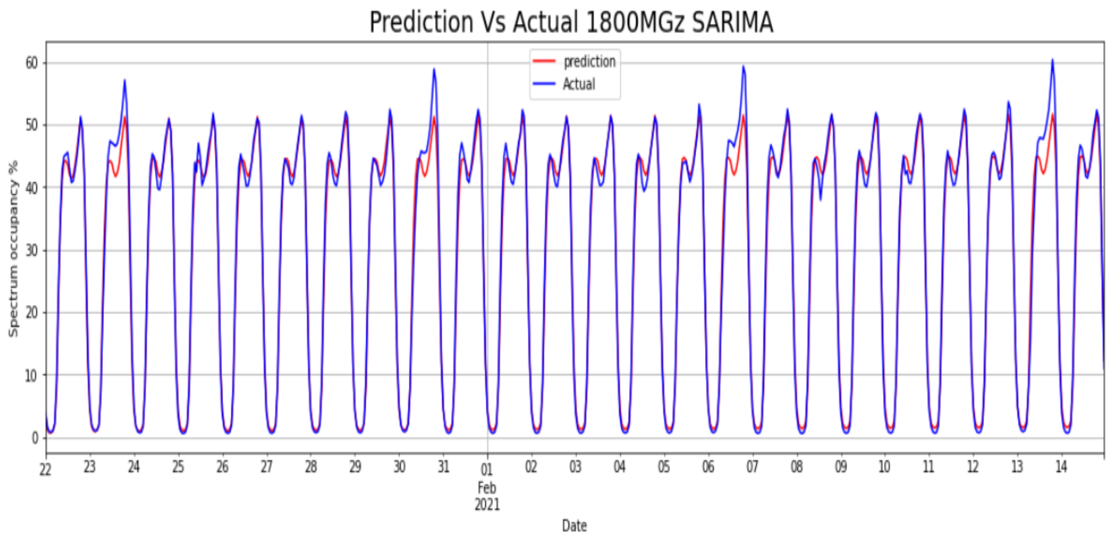

Figure 4.28: Prediction vs Actual SARIMA 1800MHz

Table 4.13: Summary prediction Evaluation three models

Metrics	Holt-Winters		SARIMA		SARIMAX	
	1800MHz	900Mhz	1800MHz	900Mhz	1800MHz	900Mhz
MSE	5.468695	3.234203	3.932156	3.200788	1.2443	0.9371
MAE	1.657788	1.33212	1.368839	0.997501	0.7857	0.6449
RMSE	2.338524	1.798389	1.982967	1.789075	1.1155	0.968
MAPE	19.87371	19.95908	16.03654	7.006883	3.5028	4.443
R-Squared	0.9856321	0.977682	0.989669	0.9779125	0.9967	0.9935

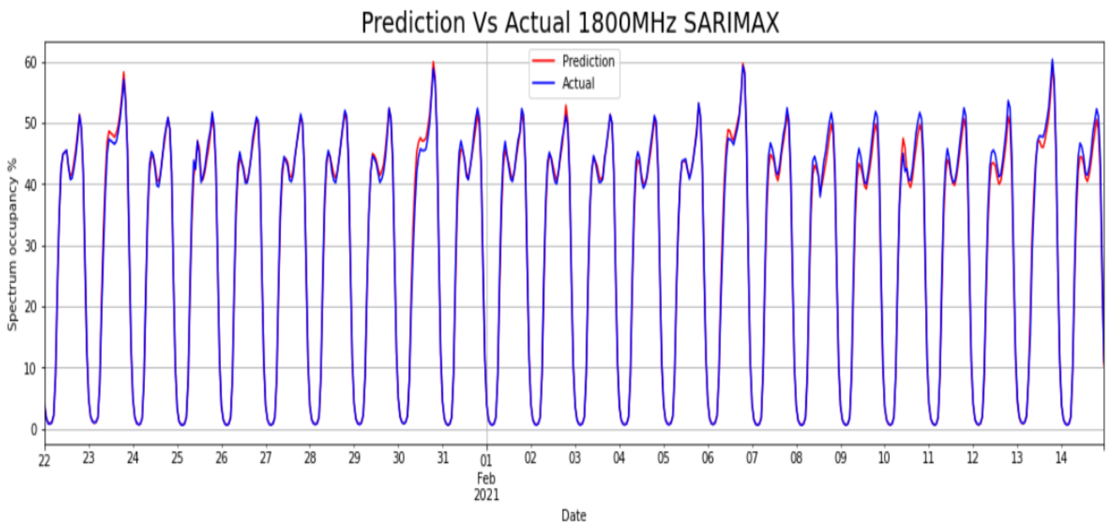

Figure 4.29: prediction vs actual SARIMAX 1800MHz

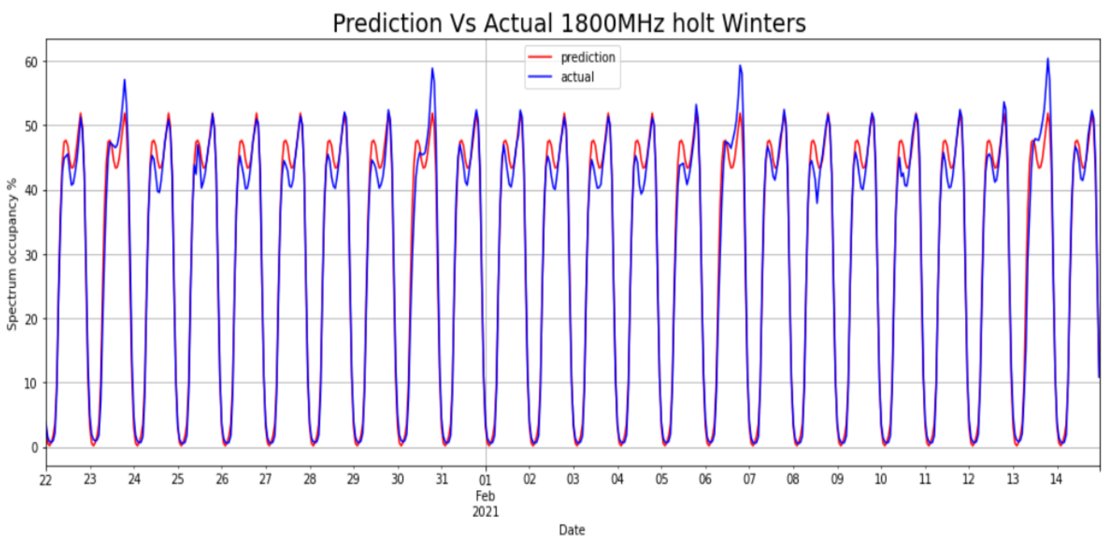

Figure 4.30: Prediction vs Actual holt-winters 1800MHz

Table 4.14: Summary of Evaluation metrics the models accuracy

Metrics	900MHz	1800MHz
Lowest MSE	SARIMAX	SARIMAX
Lowest MAE	SARIMAX	SARIMAX
Lowest RMSE	SARIMAX	SARIMAX
Lowest MAPE	SARIMAX	SARIMAX
Higher R- Squared	SARIMAX	SARIMAX

5 | Conclusion and Recommendation

5.1 Conclusion

This thesis describes a series of long-term observations of spectrum occupancy monitoring and modelling carried out in the 900MHz and 1800MHz bands assigned for GSM services in ethiotelecom in the Addis Ababa Bole BSC 05.

The measurement's goal was to uncover possible spectrum opportunities for additional spectral efficiency technologies. The results of these measurements show that the actual spectral occupancy assigned for GSM services in 900MHz is between 0.4245% and 44.71%, while it is between 0.422% and 60.354% in 1800MHz, indicating that the 900MHz spectrum is significantly lower than the 1800MHz spectrum.The typical occupancy ratios in these bands are 17.858% and 26.595%, indicating that there is a significant amount of unused spectrum in these bands.Furthermore, the percentage of spectrum occupancy is time, frequency, and space-dependent.

In this thesis, SARIMA and SARIMAX models, as well as the Holt-Winter (Triple exponential smoothing) model, are utilized to model spectrum occupancy on both the 900MHz and 1800MHz bands. According to the findings of the diagnostic testing, all models were found to be pretty well-fitted.

According to the graphical analysis of the forecast, all models appear to provide roughly comparable predictions. However, the error measures showed that the SARIMAX model produced better predictions for the 900MHz MAE of SARIMAX model is 35.34% and

50.1% lower than the SARIMA and Holt-Winter models, respectively, and for 1800MHz, SARIMAX was 42.6% and 52.6% lower than the SARIMA and Holt-Winter models, respectively. Through time series modelling techniques, it is studied which observable endogenous and exogenous factors can be employed as influential inputs for model identification.

Finally, based on the results, we can conclude that the spectrum occupancy at 900 MHz is less occupied than 1800 MHz, but the amount of occupancy depends on time, and the SARIMAX model is a better-fitted model to capture real-time spectrum occupancy data than SARIMA and Holt-Winter. As a result, the findings of this thesis will be extremely useful in the studies indicated in the following studies.

5.2 Recommendation(Relevance) for Future Work

Initially, the thesis goal is to provide information for future spectral efficiency activities to offset the incoming two new operators' spectrum allocation influences. The most allocated spectrum, according to the privacy spectrum overview chapter, is 900 MHz, which is utilized in GSM and UMTS. 70% of the spectrum has been assigned. Based on the results of the spectrum occupancy measurement at 900MHz, the maximum occupancy was 44.71% on this day, and the higher utilization of 88 cells out of 402 cells consumed more than 75%. The Bole BSC05 has a total of 4762 TRX assigned, of which 2291 TRX are in the 900MHz spectrum, hence the average TRX per cell is 2291/402 = 5.7 TRX per cell. 28 (31.8%) of the 88 cells are less than the BSC average TRX, 29 (32.9%) are Max TRX pre-cell 8TRX, and 31 (35.2%) are only one cell utilization higher than the site-level underutilized.

From a spectrum refarming standpoint, overview chapter ethiotelecom uses static from that partial spectrum refarming employed in U900MHz and LTE 1800MHz. So, in BSC 05, 134 sites used 900MHz spectrum, with 53 sites using U900MHz Plus 900GSM service and 81 sites using only 900GSM. So, out of 88 cells with occupancy greater than 75%, 41 cells (46.6 %) are from the U900MHz + 900GSM site; 10 cells are below the BSC TRX average, and 27 cells are underutilized at the site level. and 47 cells (53.4%)are from the

900GSM only site; 18 cell is below BSC TRX average, even though the allotted TRX is 8TRX, 25 cells are at or above 75% utilization.

The recommendation for overcoming the possible spectrum sharing most of the spectrum under utilization the first one deep optimization which is upgrading and downgrading TRX, the second one do refarming most of the site on the BSC 900MHz is used GSM service only add U900MHz (U111) reduced spectrum cognition, in addition, the ethiotelecom data shows smartphone penetration now more than 69% in Addis Ababa so addition UMTS cell in 900MHz (U222) also a solution. The last but not least possible solution feature enabling like push traffic to co-located DSC 1800MHz spectrum cell, increase half-rate traffic channel penetration, and enable VAMOS (Voice services over Adaptive Multi-user channels on One Slot) should be a solution.

This thesis focuses on spectrum occupancy for GSM only services. Additional analysis is necessary for other spectrum allotted to other services. Further investigation is needed for UMTS and LTE in the co-located site.

Furthermore, time series analysis will be used to investigate the thesis, and future machine learning combined with time-series algorithms must be tested to see if the prediction skill extends to another study field.

References

[1] M. Mehdawi, N. Riley, K. Paulson, A. Fanan, and M. Ammar, “Spectrum occupancy survey in hull-uk for cognitive radio applications: Measurement & analysis,” *International Journal of Scientific & Technology Research*, vol. 2, no. 4, pp. 231–236, 2013.

[2] M. Wellens, J. Riihijärvi, and P. Mähönen, “Empirical time and frequency domain models of spectrum use,” *Physical Communication*, vol. 2, no. 1-2, pp. 10–32, 2009.

[3] Z. Wang and S. Salous, “Time series arima model of spectrum occupancy for cognitive radio,” 2008.

[4] R. I. Chiang, G. B. Rowe, and K. W. Sowerby, “A quantitative analysis of spectral occupancy measurements for cognitive radio,” in *2007 IEEE 65th Vehicular Technology Conference-VTC2007-Spring*, IEEE, 2007, pp. 3016–3020.

[5] Ethiotelecom. (2019). “Gsm spectrum optimization document,” [Online]. Available: `www.intranet.ethiotelecom.et`.

[6] L. Pedraza, C. Hernandez, and E. Rodriguez, “Modeling of gsm spectrum based on seasonal arima model,” in *Proceedings of the 6th IEEE Latin-American Conference on Communications, Cartagena, Colombia*, 2014, pp. 5–7.

[7] Z. Wang and S. Salous, “Spectrum occupancy statistics and time series models for cognitive radio,” *Journal of signal processing systems*, vol. 62, no. 2, pp. 145–155, 2011.

[8] A. Gorcin, H. Celebi, K. A. Qaraqe, and H. Arslan, “An autoregressive approach for spectrum occupancy modeling and prediction based on synchronous measure-

ments," in *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, IEEE, 2011, pp. 705–709.

[9] A. Agarwal, A. S. Sengar, and R. Gangopadhyay, "Spectrum occupancy prediction for realistic traffic scenarios: Time series versus learning-based models," *Journal of Communications and Information Networks*, vol. 3, no. 2, pp. 35–42, 2018.

[10] S. Mohammadjafari, E. Kavurmacioglu, J. Maidens, and A. Bener, "Neural network based spectrum prediction in land mobile radio bands for iot deployments," in *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, IEEE, 2019, pp. 31–36.

[11] A. Mishra, "Fundamentals of cellular network planning and optimisation-[book review]," *Communications Engineer*, vol. 2, no. 6, pp. 46–46, 2005.

[12] A. R. Mishra, *Cellular technologies for emerging markets: 2G, 3G and beyond*. John Wiley & Sons, 2010.

[13] P. McGuiggan, *GPRS in Practice: A Companion to the Specifications*. John Wiley & Sons, 2005.

[14] GSMA. (2016). "Gsm architecture," [Online]. Available: `www.worldpress.com`.

[15] J. Eberspächer, H.-J. Vögel, C. Bettstetter, and C. Hartmann, *GSM-architecture, protocols and services*. John Wiley & Sons, 2008.

[16] T. Halonen, J. Romero, and J. Melero, *GSM, GPRS and EDGE performance: evolution towards 3G/UMTS*. John Wiley & Sons, 2004.

[17] ZTE. (2010). "Zxg10 ibsc base station controller technical manual," [Online]. Available: `www.zte.com`.

[18] A. R. Mishra, *Advanced cellular network planning and optimisation: 2G/2.5 G/3G... evolution to 4G*. John Wiley & Sons, 2007.

[19] M. Saily, G. Sebire, and E. Riddington, *GSM/EDGE: Evolution and performance*. John Wiley & Sons, 2011.

[20] B. H. Walke, *Mobile radio networks: networking and protocols*. John Wiley & Sons, Inc., 2001.

[21] E. Seurre, P. Savelli, and P.-J. Pietri, *EDGE for mobile internet*. Artech House, 2003.

[22] Huawei. (2013). "Whitepaper on spectrum," [Online]. Available: `www.huawei.com`.

[23] Huawei@. (2020). "Lte 1800mhz ecosystem drivers issue 1.0," [Online]. Available: `www.huawei.com`.

[24] L. HUAWEI TECHNOLOGIES CO., "Lte 1800mhz ecosystem drivers," *Communications Engineer*, no. 1, pp. 1–17, 2011.

[25] B. Vishwas and A. Patel, *Hands-on Time Series Analysis with Python*. Springer, 2020.

[26] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

[27] N. LLC. (1976). "The box-jenkins method," [Online]. Available: `https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/The_Box-Jenkins_Method.pdf`.

[28] P. J. Brockwell, P. J. Brockwell, R. A. Davis, and R. A. Davis, *Introduction to time series and forecasting*. Springer, 2016.

[29] J. D. Hamilton, *Time series analysis*. Princeton university press, 2020.

[30] R. Lawton *et al.*, "Time series analysis and its applications: Robert h. shumway and david s. stoffer; springer texts in statistics; 2000, springer-verlag;[uk pound] 55, us $79.95, isbn 0-387-98950-1," *International Journal of Forecasting*, vol. 17, no. 2, pp. 299–301, 2001.

[31] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[32] Ď. Peter and P. Silvia, "Arima vs. arimax–which approach is better to analyze and forecast macroeconomic time series," in *Proceedings of 30th International Conference Mathematical Methods in Economics*, vol. 2, 2012, pp. 136–140.

[33] R. Adhikari and R. K. Agrawal, "An introductory study on time series modeling and forecasting," *arXiv preprint arXiv:1302.6613*, 2013.

[34] A. B. Batista and P. S. d. M. Pires, “An approach to outlier detection and smoothing applied to a trajectography radar data,” *Journal of Aerospace Technology and Management*, vol. 6, pp. 237–248, 2014.

[35] J. Brownlee, “Introduction to time series forecasting with python,” *Jason Brownlee*, 2019.

[36] F. Nikolaisen Sävås, *Forecast comparison of models based on sarima and the kalman filter for inflation*, 2013.

[37] P. S. E. C. of Science. (). “Applied time se rices analysis,” [Online]. Available: `https://online.stat.psu.edu/stat510/lesson/4/4.1`.

[38] U. D. of Commerce. (). “Engineering stastics,” [Online]. Available: `https://www.itl.nist.gov/div898//handbook/index.htm`.

[39] H.-C. Changtzeng, D.-F. Chang, and Y.-H. LO, “Detecting concurrent relationships of selected time series data for arimax model,” *ICIC express letters. Part B, Applications: an international journal of research and surveys*, vol. 10, no. 10, pp. 937–944, 2019.

Appendices

Modeling GSM Spectrum Occupancy Using SARMA, SARMAX and Holt-Winters

Anteneh Atnafu
SCHOOL OF ELECTRICAL AND
COMPUTER ENGINEERING
ADDIS ABABA INSTITUTE OF TECHNOLOGY
Addis Ababa, Ethiopian
anteneh.atinafu@ethiotelecom.et

Dr. Murad Ridwan
SCHOOL OF ELECTRICAL AND
COMPUTER ENGINEERING
ADDIS ABABA INSTITUTE OF TECHNOLOGY
Addis Ababa, Ethiopian
muradridwan@gmail.com

***Abstract*—Spectrum occupancy models can help you make better use of the radio spectrum. It has also been extensively researched in recent decades as it is critical for developing new regulations for spectrum allocation for future technologies as well as monitoring the activities that take place on that spectrum. Understanding the amount of available spectrum is critical for future wireless technologies that want to address the so-called spectrum scarcity issue. Spectrum occupancy measurements provide critical data for frequency planning and optimization, as well as assist in smart decision-making. The goal of time series modeling is to collect and thoroughly examine previous data from a time series in order to construct an appropriate model that accurately captures the series' intrinsic structure. This paper examines three types of time series analysis methods: Holt-winters, Seasonal Auto Regressive Integrated Moving Average (SARIMA), and SARIMA eXogenous regresses (SARIMAX) based models, as well as their inherent prediction strengths and weaknesses. Time series modeling principles such as trend, stationarity, seasonality, residual, and so on have also been covered. To assess the accuracy rate, we fitted multiple models to a time series using five primary metrics. Among the methods used are mean square error, mean absolute error (MAE), root mean square error, mean absolute percentage error, and R-squared. At 1800MHz, the maximum spectrum occupancy is 60.35%, and at 900MHz, it is 44.71%. For 900MHz MAE, the SARIMAX model produced better predictions (35.34% and 50.1% lower than the SARIMA and Holt-Winter models, respectively), while for 1800MHz, the SARIMAX model produced 42.6% and 52.6% lower than the SARIMA and Holt-Winter models, respectively.**

***Index Terms*—Spectrum Occupancy, Holt-Winters, SARIMA, and SARIMAX**

I. Introduction

Radio spectrum is a unique natural resource because it is non-exhaustible, but it is regarded as a scarce resource owing to how we use it. As a result, competent agencies such as the ITU have been established to ensure the sensible, fair, and economical use of the spectrum. Their objective is to ensure that wireless systems operate without interruption around the world by enforcing radio regulations and regional agreements. Current wireless communication systems rely heavily on fixed radio resource allocation. This means that a section of the radio spectrum is allocated or licensed to third parties on a long-term basis and that it covers enormous areas, such as entire countries. This approach provides ultimate protection from harmful interference in an allocated radio band, but as many measurement studies show, the fixed radio frequency allocation may result in significant under-utilization of the radio spectrum due to very sporadic usage across different geographical regions and periods [1] [2].

The rising usage of radio-based technologies in society, as well as the enormous prospects for social development that these technologies afford, underscore the significance of radio-frequency spectrum and national spectrum management systems. Technological advancement has consistently opened the door to a lot of new spectrum applications, resulting in increased interest in, and demand for, finite spectrum resources. Increased demand necessitates efficient spectrum utilization and the implementation of appropriate spectrum management techniques. Modern data handling capabilities and technical analyses are critical in this framework to suit the wide range of possible users requesting spectrum access.

Wireless networks are experiencing a huge transformation right now. According to current trends, the number and variety of wireless devices, as well as spectrum demand, are on the rise. The radio frequency spectrum is, unfortunately, a finite resource. As a result, some parts of the spectrum are heavily utilised while others are entirely ignored. It is critical to monitor and understand spectrum resource consumption in order to improve and regulate radio spectrum utilization [3] [4].

Spectrum occupancy measurement studies seek to quantify the amount of time that a specific frequency channel or frequency band is occupied in a given area, hence defining the channel's utilization rate. Measurements can be used to analyse the current state of spectrum utilization as well as the availability of spectrum for other users. Furthermore, the measurement results provide significant information to regulators as well as operators such as ethiotelecom about the efficiency of the existing usage of spectrum allocations. Spectrum measurements are also carried out to increase the accuracy of spectrum databases, which can be used to keep track of spectrum utilization statistics to promote spectrum sharing in an operational setting.

In this paper, we measure GSM 900MHz and 1800MHz spectrum occupancy models, which target the usage patterns of the spectral bands used by users of the ethiotelcom network in Addis Ababa, using time series analysis methods, specifically three algorithms: exponential smoothing (Holt-Winter), SARIMA, and SARIMAX. to assist the company's optimization and spectral impact efforts by providing the current spectrum occupancy model and prediction information for the GSM network.

II. MEASUREMENT SETUP

The information is derived from the Ethiotelecom PRS system. The research region is Addis Ababa Bole BSC 05, and the spectrum chosen to explore is the 1800MHz and 900MHz spectrum assigned to ethiotelecom for GSM service. The data was collected hourly from October 18, 2020, to February 14, 2021, a period of four months, and a total of 789 cells were included in the study. Table I shows the detailed spectrum and the study area.

TABLE I
SHOW THE DETAILED SPECTRUM CHANNEL ASSIGNED TO GSM

Spectrum	DL in MHz	Spectrum assigned to GSM	Bandwidth
900MHz	935 - 960	947.6MHz - 959.8MHz	12MHz
1800MHz	1805 - 1880	1842.6MHz – 1851.4MHz	8.8MHz
		1871.6MHz - 1879.8MHz	8.2MHz

III. BASIC CONCEPTS OF TIME SERIES MODELING

We can define a time series as a sequential or chronologically order set of data point. Typically, these data point are measured over successive times. in mathematical notation, it can be expressed as random variable x_t with t denoting the time dimension$(t \geq 0)$. If multiple variables are expressed in a time series at a time 't' then it is a multivariate time series otherwise it is uni variate. The value of x_t can be either continuous or discrete. Mathematical and statistical analysis performed on this kind of data to find hidden patterns and meaningful insight is called time-series analysis. Time-series modeling techniques are used to understand past patterns from the data and try to forecast future horizons [5].

A. Characteristics Time-Series

There are four characteristics in time series namely trend, cyclicality, seasonality and irregularity.

Trend:-long-term upward or downward movement.

Cyclicality:-It is not the same as seasonal variation. The cycle is the variation of time series data's autoregressive component. Cycles occur over longer time periods, such as every 6–10 years, whereas seasonal variation occurs over shorter time periods.

Seasonality:-variation in data caused by seasonal effects. Ice cream sales are high in summer, heating oil sales are high in winter but low in summer.

Irregularity:-a component that is left after other components have been calculated and removed from time series data. It is randomly, identically and independently distributed. We can use seasonal decomposition from a statistics model to breakdown the data into its constituent pieces, taking into account whether the series is additive or multiplicative.The additive model works with linear trends of time-series data such as changes constantly over time. The additive model formula is as follows:

$$Y\left[t\right]=T\left[t\right]+S\left[t\right]+c\left[t\right]+e\left[t\right] \tag{1}$$

The multiplicative model works with a nonlinear type of data such as quadratic or exponential. The multiplicative model formula is as follows:

$$Y\left[t\right]=T\left[t\right]*S\left[t\right]*c\left[t\right]*e\left[t\right] \tag{2}$$

B. Stationarity in Time-Series Data

A stochastic process's stationarity can be interpreted as a type of statistical equilibrium. As a result, the statistical properties of the process are not time dependent. Except for inherent stochastic fluctuation, stationary stochastic models are typically designed so that the mean level, variance, and standard deviation are time independent. The stationarity assumption may reflect reality in addition to reducing the mathematical complexity of a stochastic model [6] [7].

We have to make time-series data stationary before fitting a model. We can make time-series stationary by transforming the data. Usually, differencing is used to make the data stationary.

So, how can we test whether a time series data is stationary or not?

1) Visualization of the time series plots and identify trends or seasonality.
2) Divide the data into 3 different sets and calculate the mean and variance for each set and confirm whether the mean and variance for each set are substantially different or not.
3) The statistical tests.

C. Autocorrelation and Partial Autocorrelation Functions

The ACF and PACF analyses are required to determine a suitable model for a given time series of data. These statistical measures reflect the relationship between the observations in a time series. Plotting the ACF and PACF against consecutive time lags is frequently useful for modeling and forecasting purposes. These plots aid in the determination of the order of AR and MA terms [8].

We use ACF and PACF to choose the correct order for the AR (p) and MA (q) value models. For AR order, look at the PACF plot and choose a lag value that has a significant correlation factor before correlations get insignificant. For MA order q, look at the ACF plot and do the same. Don't forget, you should only get these values from the ACF and PACF plots of stationary time series, not the above plots. The ACF and PACF plots given above are the plots of the original data, which is non-stationary.

ACF and PACF plots allow you to determine the AR and MA components of an ARIMA model. Both the Seasonal and the non-Seasonal AR and MA components can be determined from the ACF and PACF plots(see Table II) [9] [10].

TABLE II
ACF AND PACF TO IDENTIFY THE ORDERS OF SARMA(P, Q) × (P, Q)S

Model	ACF	PACF
AR(p)	Exponentially Decay	Significant till the p lags
MA(q)	Significant till the q lags	Exponentially Decay
AR(P)	Exponentially Decay at each m lag	Significant at Pm lags
MA(Q)	Significant at Qi lags	Exponentially Decay at each m lag

IV. SYSTEM MODEL

A. SEASONAL ARIMA

Seasonal ARIMA (SARIMA) is a technique of ARIMA, where the seasonal component can be handled in univariate time-series data. It adjoins three new hyper parameters to lay down AR(P), I(D), and MA(Q) for the seasonality component of a time series. SARIMA allows for the occurrence of seasonality in a series [1] [11].

$$\Phi_P\left(B^s\right)\phi_p\left(B\right)\left(1-B\right)^d\left(1-B^s\right)^D Y_t=\theta_q\left(B\right)\Theta_Q\left(B^s\right)\varepsilon_t \tag{3}$$

where B is the backshift operator (i.e.$By_t = y_{t-1}$), and

$$\Phi_P\left(B^s\right)=1-\Phi_1 B^s-\Phi_2 s^{2s}-...-\Phi_P B^{Ps} \tag{4}$$

$$\phi_p\left(B\right)=1-\phi_1 B-\phi_2 B^2-...-\phi_p B^p \tag{5}$$

$$\theta_q\left(B\right)=1-\theta_1 B-\theta_2 B^2-...-\theta_q B^q \tag{6}$$

$$\Theta_Q\left(B^s\right)=1-B^s-\Theta_2 B^{2s}-...-\Theta_Q B^{Qs} \tag{7}$$

The seasonal ARIMA model combines both nonseasonal and seasonal components in a multiplicative model. The notation can be defined as follows:

ARIMA (p, d, q) X (P, D, Q) m, m is the number of observations per year.

Three trend elements need to be configured. It is the same as the ARIMA model, as shown here:

(p, d, q) is a non-seasonal component, as shown here:

- p: Trend autoregressive order
- d: Trend differencing order
- q: Trend moving average order

(P, Q, D) is a seasonal component Four seasonal components are not part of the ARIMA model that are essential to be configured.

- P: Seasonal autoregressive order
- D: Seasonal differencing order
- Q: Seasonal moving average order
- m: Timestamp for single-season order

B. SARIMAX

The SARIMAX model is a SARIMA model with external influencing variables, called SARIMAX (p, d, q) (P, D, Q) m (X), where X is the vector of exogenous variables. The exogenous variables perhaps modeled by the multilinear regression equation are articulated as follows:

$$\left(1-\phi_1 B\right)\left(1-\Phi_1 B^s\right)\left(1-B\right)\left(1-B^s\right)Y_t= \tag{8}$$

$$\left(1+\theta_1 B\right)\left(1+\Theta_1 B^s\right)\varepsilon_t\left(X_{k,t}\right) \tag{9}$$

where $X_{1,t}, X_{1,t}...X_{k,t}$are observations of k number of exogenous variables corresponding to the dependent variable.

C. Triple Exponential Smoothing(Holt-Winters)

Triple exponential smoothing is a forecasting method that enforces exponential smoothing three times. This is the extension of double exponential smoothing. Holt-Winter's method when seasonality is found in the data This triple exponential smoothing is also known as the Holt-Winters method [12]. Forecasting equation

$$\hat{Y}_{t+1}=l_t+b_t+s_{t-m(k+1)} \tag{10}$$

$$l_t=\alpha\left(Y_t-s_{t-m}\right)+\alpha\left(1-\alpha\right)Y_{t-1}\ 0\leq\beta\leq 1 \tag{11}$$

$$b_t=\beta\left(l_t-l_{t-1}\right)+\left(1-\beta\right)b_{t-1}\ 0\leq\beta\leq 1 \tag{12}$$

$$s_t=\gamma\left(Y_t-l_t-b_{t-1}\right)+\left(1-\gamma\right)s_{t-m}\ 0\leq\gamma\leq 1 \tag{13}$$

Where l_t is estimated of level and b_t is the trend estimated and s_t is the seasonal estimated, α is the smoothing parameter for the level and β is the smoothing parameter for trend and γ is the smoothing parameter for seasonality

V. ESTIMATION AND ORDER SELECTION

A. Maximum Likelihood Estimation

Once the model order has been identified (i.e., the values of p, d and q), we need to estimate the parameters c ,$\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$. When estimates the ARIMA model, it uses maximum likelihood estimation (MLE). This technique finds the values of the parameters which maximize the probability of obtaining the data that we have observed. For ARIMA models, MLE is very similar to the least squares estimates that would be obtained by minimizing [6].

$$MLE = \sum_{t-1}^{T} e_t^2 \tag{14}$$

Where; $e^2$is the error between the actual and the estimated data

B. Information Criteria

Akaike's Information Criterion (AIC), which was useful in selecting predictors for regression is also useful for determining the order of an ARIMA model. It can be written as

$$AIC = -2log\left(1\right) + 2\left(k\right) \tag{15}$$

$$AIC_c = AIC + \frac{2k^2 + 2k}{n - k - 1} \tag{16}$$

$$BIC = -2log\left(1\right) + \ln\left(n-\right)k \tag{17}$$

Where K = p+q+1 l = likelihood of the data n = sample size

It is important to note that these information criteria tend not to be good guides for selecting the appropriate order of difference (d) of a model, but only for selecting the values of p and q. This is because the differencing changes the data on which the likelihood is computed, making the AIC values between models with different orders of differencing not comparable. So we need to use some other approach to choose d, and then we can use the AICc to select p and q.

VI. Result and Discussion

Figure 1 and Table III show total visualization and description of the spectrum occupancy data.

TABLE III
DESCRIPTIVE STATICS OF DATA

Descriptive statics	1800MHz	900MHz
count	2880	2880
Mean	26.59368465	17.85808223
median	35.6135044	22.78355167
Maximum	60.35384534	44.70626258
Minimum	0.422016012	0.424584404
5-th percentile	0.693050217	0.631371651
95-th percentile	48.37223578	34.07332064

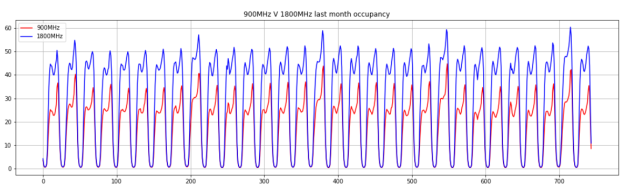

Fig. 1. Spectrum Occupancy1800MHz and 900MHz plot one month

A. Triple Exponential Smoothing Holt-Winters

Intercept (level), trend, and seasonality are the three components of the triple exponential smoothing content. Seasonal components in the model will explain recurring variations around intercept and trend, and the length of the season, in other words, the period after which the variations repeat, will be provided. There is a separate component for each observation in the season; for example, if the length of the season is 24 hours (daily seasonality) hourly data, we will have 24 seasonal components, one for each hour of the day and an initial value for the intercept (level), a trend is also shown in Tables V.

Based on Sum Square Error (SSE) and Information Criteria fitted model selected after checking so many parameter possibilities Table IV.

TABLE IV
HOLT-WINTERS FITTED PARAMETERS

Spectrum	Model	SSE	AIC	BIC
900MHz	Holt-Winters	1352.66	-1169.05	-1002.52
1800MHz	Holt-Winters	2371.773	124.795	291.325

TABLE V
HOT-WINTERS FITTED PARAMETER

parameter	Coefficient	
	1800MHz	900MHz
α	0.8590816	0.9428782
β	1.1876e-08	2.3933e-09
γ	0.1409184	0.0571218
l_o	6.9228650	5.6123424
b_o	-0.8774664	-0.3773654

B. SARIMA and SRAIMAX Order Selection

1) Autoregressive and Moving Average Orders: This section will determine the appropriate autoregressive orders p and P, as well as moving average orders q and Q. The first step is to visually inspect the correlogram for each well-differentiated process. The results in Table VI will be utilized as a guideline for what orders the autocorrelation function (ACF) and partial autocorrelation function (PACF) indicate. The findings will be shown for each spectrum occupancy, followed by the ordering chosen using the log-like likelihood ratio and the information criterion specified as AIC and BIC.

Figures 2 and 3 show sample plots the ACF and PACF of 900MHz and 1800MHz spectrum occupancy.

TABLE VI
SARIMA AND SARIMAX(P, D, Q) (P, D, Q)24

Spectrum	Model	LLR	AIC	BIC
900MHz	SARIMA(102)x(112,24)	-2111.54	4237.072	4277.195
	SARIMAX(102)x(111, 24)	-812.931	1639.861	1679.985
1800MHz	SARIMA(212)x(213,24)	-2544.06	5108.115	5165.43
	SARIMAX(103)x(112, 24)	-1315.94	2649.881	2701.468

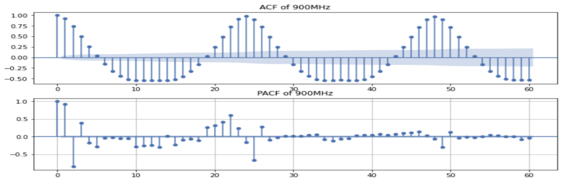

Fig. 2. ACF and PACF of 900MHz spectrum data

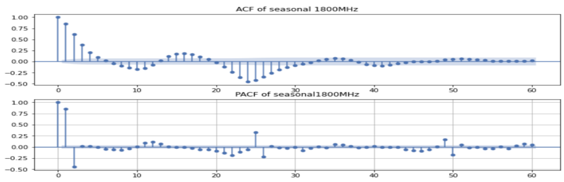

Fig. 3. ACF & PACF of seasonal difference of 1800MHz

C. Estimation and Diagnostic Checking of Models

Tables VII and VIII show the fitted SARIMA and SARIMAX models. The next stage is to extract the residuals and then undertake a visual inspection of each model before running various statistical tests that take into account their normality and randomness.

The residuals for all the Spectrum occupancy data appear to be highly homoscedastic, with a mean near to zero and a p-value of zero. Furthermore, in both spectrum data sets, the number of standardized residuals that lie outside the interval of 1.96 for both the SARIMAX and SARIMA models is less than 8%.

Figures 4 and 5 show the autocorrelation functions for the residuals. The values for the first 60 lags are reported, and the number of significant lags defined as those that fall outside the sample size-dependent interval $\pm\frac{2}{\sqrt{T}}$ is of interest. The significant 5 lags out of 60 lags in the SARIMA model are 8.3 percent, but the remaining 91.7 % are inside the sample site. On the other hand, in the SARIMAX model, 3 percent of the significant level 2 lags out of 60 lags are outside the sample size, but 97 % are within the sample size in both the 900MHz and 1800MHz data sets. This result is acceptable

TABLE VII
SARIMA FITTED PARAMETER

parameter	Coefficient	
	1800MHz	900MHz
ϕ_1	1.1471	0.8074
ϕ_2	-0.3771	-
θ_1	-0.8831	0.5097
θ_2	-0.1093	0.1144
Φ_1	-1.0074	-0.8613
Φ_2	-0.7012	-
Θ_1	0.2682	0.0612
Θ_2	-0.1213	0.7586
Θ_3	-0.7214	-

TABLE VIII
SARIMAX FITTED PARAMETER

parameter	Coefficient	
	1800MHz	900MHz
ϕ_1	0.7803	0.8834
θ_1	0.4363	0.3304
θ_2	0.2038	0.0592
θ_3	0.0697	-
Φ_1	-0.8493	0.0813
Θ_1	0.0944	-0.8311
Θ_2	-0.7279	-

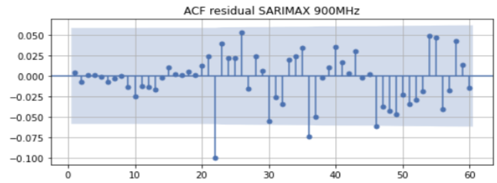

Fig. 4. ACF plot of residual SARIMAX 900MHz

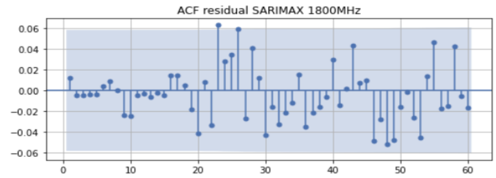

Fig. 5. ACF plot of residual SARIMAX 1800MHz

D. Prediction Comparison

Figures 6,7,8, and 9 show the model prediction of spectrum occupancy in 1800MHz and 900MHz versus the actual results. All of the models appear to be well-fitted, with the forecast closely following the actual values. We employ test assessment metrics such as MSE, MAE, MAPE, RMSE, and R-Squared.

The evaluation metrics for the Holt-Winters models, SARIMA models, and SARIMAX models are listed in Table IX. The R squared statistics, which are close to one for each model, demonstrate that the models are well-fitted.

When the models are compared further, it can be shown that the error measures for all spectrum occupancy measurements are marginally smaller for the SARIMAX models than for the SARIMA and Holt-Winters models. The SARIMAX model appears to be more accurate at prediction.

Demonstrate that all assessment metrics of the SARIMAX models are smaller than those of the SARIMA and Holt winters models, implying that the tests are one-sided with the alternative hypothesis that the SARIMAX model is more accurate for both 900MHz and 180MHz spectrum occupancy.

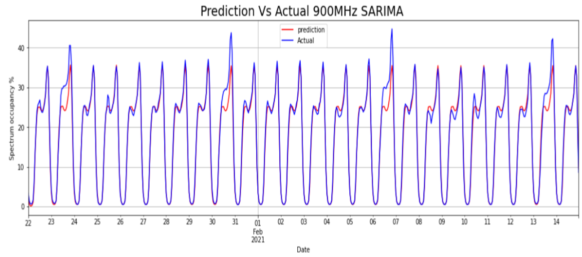

Fig. 6. prediction vs actual SARIMA 900MHz

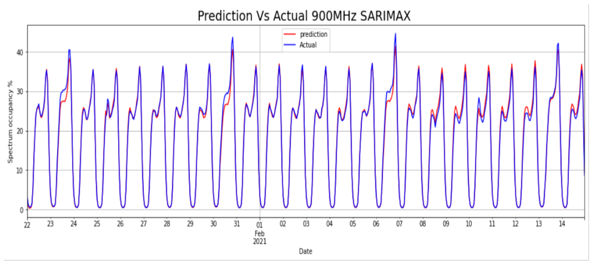

Fig. 7. Prediction vs Actual SARIMAX 900MHz

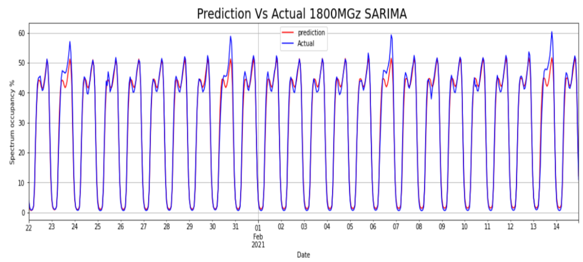

Fig. 8. Prediction vs Actual SARIMA 1800MHz

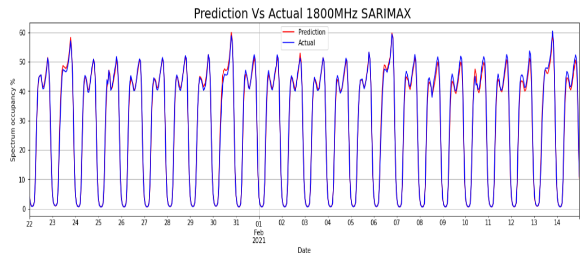

Fig. 9. prediction vs actual SARIMAX 1800MHz

TABLE IX
SUMMARY PREDICTION EVALUATION THREE MODELS

Metrics	Holt-Winters		SARIMA		SARIMAX	
	1800MHz	900Mhz	1800MHz	900Mhz	1800MHz	900Mhz
MSE	5.469	3.234	3.932	3.201	1.244	0.937
MAE	1.658	1.332	1.369	0.998	0.786	0.645
RMSE	2.339	1.798	1.983	1.79	1.116	0.968
MAPE	19.874	19.959	16.037	7.007	3.503	4.443
R-Squared	0.986	0.978	0.99	0.978	0.997	0.994

VII. CONCLUSION

The measurement's goal was to uncover possible spectrum opportunities for additional spectral efficiency technologies. According to the results of these measurements, the actual spectral occupancy assigned for GSM services in 900MHz ranges between 0.4245% and 44.71%, while it ranges between 0.422% and 60.354% in 1800MHz, indicating that the 900MHz spectrum is significantly lower than the 1800MHz spectrum. The typical occupancy ratios in these bands are 17.858% and 26.595%, indicating that there is a significant amount of unused spectrum in these bands. Furthermore, the percentage of spectrum occupancy is time, frequency, and space-dependent.

In this paper, the SARIMA and SARIMAX models, as well as the Holt-Winter (Triple exponential smoothing) model, are utilized to model spectrum occupancy in both the 900MHz and 1800MHz bands. According to the findings of the diagnostic testing, all models were found to be pretty well-fitted.

REFERENCES

[1] M. Mehdawi, N. Riley, K. Paulson, A. Fanan, and M. Ammar, "Spectrum occupancy survey in hull-uk for cognitive radio applications: measurement & analysis," *International Journal of Scientific & Technology Research*, vol. 2, no. 4, pp. 231–236, 2013.

[2] M. Wellens, J. Riihijärvi, and P. Mähönen, "Empirical time and frequency domain models of spectrum use," *Physical Communication*, vol. 2, no. 1-2, pp. 10–32, 2009.

[3] Z. Wang and S. Salous, "Time series arima model of spectrum occupancy for cognitive radio," 2008.

[4] R. I. Chiang, G. B. Rowe, and K. W. Sowerby, "A quantitative analysis of spectral occupancy measurements for cognitive radio," in *2007 IEEE 65th Vehicular Technology Conference-VTC2007-Spring*. IEEE, 2007, pp. 3016–3020.

[5] B. Vishwas and A. Patel, *Hands-on Time Series Analysis with Python*. Springer, 2020.

[6] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

[7] N. LLC. (1976) The box-jenkins method. [Online]. Available: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/

[8] R. Adhikari and R. K. Agrawal, "An introductory study on time series modeling and forecasting," *arXiv preprint arXiv:1302.6613*, 2013.

[9] F. Nikolaisen Sävås, "Forecast comparison of models based on sarima and the kalman filter for inflation," 2013.

[10] P. S. E. C. of Science. (2016) Applied time serices analysis. [Online]. Available: https://online.stat.psu.edu/stat510

[11] L. Pedraza, C. Hernandez, and E. Rodriguez, "Modeling of gsm spectrum based on seasonal arima model," in *Proceedings of the 6th IEEE Latin-American Conference on Communications, Cartagena, Colombia*, 2014, pp. 5–7.

[12] H.-C. Changtzeng, D.-F. Chang, and Y.-H. LO, "Detecting concurrent relationships of selected time series data for arimax model," *ICIC express letters. Part B, Applications: an international journal of research and surveys*, vol. 10, no. 10, pp. 937–944, 2019.