MACHINE LEARNING-BASED PATH LOSS MODELS FOR HETEROGENEOUS RADIO NETWORK PLANNING IN A SMART CAMPUS

POPOOLA, SEGUN ISAIAH

Matriculation Number: 16PCK01420

B.Tech. Electronic and Electrical Engineering (LAUTECH, Ogbomoso, Nigeria)

July, 2018

MACHINE LEARNING-BASED PATH LOSS MODELS FOR HETEROGENEOUS RADIO NETWORK PLANNING IN A SMART CAMPUS

POPOOLA, SEGUN ISAIAH

Matriculation Number: 16PCK01420

B.Tech. Electronic and Electrical Engineering (LAUTECH, Ogbomoso, Nigeria)

A DISSERTATION SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND INFORMATION ENGINEERING, COLLEGE OF ENGINEERING IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF ENGINEERING (M.ENG) DEGREE IN INFORMATION AND COMMUNICATION ENGINEERING

July, 2018

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfilment of the requirements for the award of **Master of Engineering (M.Eng)** degree in the Department of **Electrical and Information Engineering,** College of Engineering, Covenant University, Ota, Ogun State, Nigeria.

Mr. J. A. Philip ……..

(Secretary, School of Postgraduate Studies) Signature & Date

Prof. A. H. Adebayo ……..

(Dean, School of Postgraduate Studies) Signature & Date

DECLARATION

I, **POPOOLA, SEGUN ISAIAH** (16PCK01420), declare that this M.Eng dissertation titled "Machine Learning-Based Path Loss Models for Heterogeneous Radio Network Planning in a Smart Campus" was carried out by me under the supervision of Prof. AAA. Atayero of the Department of **Electrical and Information Engineering**, Covenant University Ota, Ogun State, Nigeria. I attest that this dissertation has not been presented either wholly or in part for the award of any degree elsewhere. All sources of scholarly information used in this dissertation are duly acknowledged.

POPOOLA, SEGUN ISAIAH ……

Signature & Date

CERTIFICATION

We certify that the dissertation titled "Machine Learning-Based Path Loss Models for Heterogeneous Radio Network Planning in a Smart Campus" is an original work carried out by **POPOOLA, SEGUN ISAIAH** with Matriculation Number **16PCK01420,** in the Department of **Electrical and Information Engineering**, College of Engineering, Covenant University, Ota, Ogun State, Nigeria, under the supervision of Prof. AAA. Atayero. We have examined the work and found it acceptable for the award of **Master of Engineering (M.Eng)** degree in **Information and Communication Engineering**.

Prof. AAA. Atayero	**……**
(Supervisor)	Signature & Date
Dr. A. U. Adoghe	**……**
(Head of Department)	Signature & Date
Prof. A. A. Ayeni	**……**
(External Examiner)	Signature & Date
Prof. A. H. Adebayo	**……**
(Dean, School of Postgraduate Studies)	Signature & Date

DEDICATION

This work is dedicated to divine Trinity: the Father, the Son, and the Holy Spirit.

ACKNOWLEDGEMENTS

Above all, I appreciate the grace I received from God, who gives wisdom liberally without measure. There is a spirit in man, and the inspiration of the Almighty gives me understanding. All the glory be to God.

I acknowledge the immense support I receive from my parents, Mr. & Mrs. Popoola, and siblings, Tosin and Funso. I also wish to appreciate the support and encouragement I received from the family of Bro. & Sis. Abodunrin Adegbola, and Engr. Abiodun Abimbola Ogunjimi.

Special thanks to my erudite and loving supervisor, Prof. AAA. Atayero, for his support and commitment to the successful completion of this research project. I sincerely the coordinators of the Post-Graduate (PG) programmes in the Department of Electrical and Information Engineering, Dr. Ayokunle A. Awelewa, and Dr. Joke A. Badejo for their relentless efforts. I also wish to express my profound gratitude to the immediate past and present Heads of Department of Electrical and Information Engineering, Dr. Victor O. Matthews, and Prof. A. U. Adoghe. The guidance and support received from Prof. Emmanuel Adetiba is equally worthy of appreciation.

Finally, I acknowledge and appreciate the Chancellor, Dr. David Oyedepo, the Board of Regents, the School of Postgraduate Studies (SPS) under the leadership of Prof. A. H. Adebayo, and the management of Covenant University led by the Vice-Chancellor, Prof. AAA. Atayero, for providing a conducive and safe environment that encouraged sound academic research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ALGORITHMS

LIST OF FIGURES

LIST OF ABBREVIATIONS

3D	Three-dimensional
5G	Fifth Generation
AM	Amplitude Modulation
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
BFG	BFGS Quasi Newton
BTS	Base Transceiver Station
CDMA	Code Division Multiple Access
CGB	Conjugate Gradient with Powell/Beale Restarts
CGF	Fletcher-Powell Conjugate Gradient
CGP	Polak-Ribiere Conjugate Gradient
COST	COopération européenne dans le domaine de la recherche Scientifique et Technique
CW	Continuous Wave
DE	Differential Evolution
DCS	Digital Cellular System
DSR	Design Science Research
DTM	Digital Terrain Map
ELM	Extreme Learning Machine
GA	Genetic Algorithm

GDX	Variable Learning Rate Backpropagation
GHz	Giga Hertz
GIS	Geographic Information System
GPS	Global Positioning System
GSM	Global System for Mobile communications
HEI	Higher Education Institution
HF	High Frequency
ICT	Information and Communication Technology
IEEE	Institute of Electrical and Electronic Engineering
IoT	Internet of Things
IMT	International Mobile Telecommunications
ITU	International Telecommunication Union
ITU-R	International Telecommunication Union-Radio
LF	Low Frequency
LM	Levenberg-Marquardt
logsig	Logarithmic sigmoid function
LOS	Line of Sight
LRNN	Layer Recurrent Neural Network
LTE	Long Term Evolution
M2M	Machine-to-Machine
MAE	Mean Absolute Error

MATLAB	MATrix LABoratory
MHz	Mega Hertz
MLP-NN	Multi-Layer Perceptron Neural Network
MMDS	Multipoint Microwave Distribution System
MSE	Mean Square Error
N/A	Not Applicable
NLOS	Non-Line of Sight
OSS	One Step Secant
PC	Personal Computer
PCS	Personal Communication System
purelin	Linear activation function
QoS	Quality of Service
R	Correlation coefficient
RAM	Random Access Memory
RAN	Radio Access Network
RBF	Radial Basis Function
RMSE	Root Mean Square Error
RP	Resilient Backpropagation
RSS	Received Signal Strength
SCG	Scaled Conjugate Gradient
SED	Standard Error Deviation

SHF	Supper High Frequency
SPM	Standard Propagation Model
SUI	Stanford University Interim
SVM	Support Vector Machine
tansig	Hyperbolic tangent function
TETRA	TErrestrial Trunked Radio
TV	Television
UHF	Ultra-High Frequency
UMTS	Universal Mobile Telecommunications System
USB	Universal Serial Board
VHF	Very High Frequency
VLF	Very Low Frequency
WEKA	Waikato Environment for Knowledge Analysis
WiMax	Worldwide Interoperability for Microwave Access
Wi-Fi	Wireless Fidelity

ABSTRACT

An easy-to-use and accurate multi-frequency path loss model is a necessary tool for heterogeneous radio network planning and optimization towards achieving a smart campus. The learning ability in artificial intelligence may be exploited to reduce computational complexity and to improve prediction accuracy. In this research project, an optimal heterogeneous model was developed for path loss predictions in a typical university campus propagation environment using machine learning approach. Radio signal measurements were conducted within the campus of Covenant University, Ota, Nigeria to obtain the logs of signal path loss at 900, 1800, and 2100 MHz. Different path loss prediction models were developed based on Artificial Neural Network (ANN) and Support Vector Machine (SVM) learning algorithms. The prediction accuracy and generalization ability of the ANN-based model, which has seven input nodes (distance, frequency, clutter height, elevation, altitude, latitude, and longitude), single hidden layer with 43 neurons and logarithmic sigmoid (*logsig*) activation function, and a single output neuron (for path loss variable) with tangent hyperbolic sigmoid (*tansig*) activation function, was found to be the best when compared to the prediction outputs of SVM-based model, and popular empirical models (i.e. Okumura-Hata, COST 231, ECC-33, and Egli). The ANN-based path loss model was trained based on Levenberg-Marquardt learning (LM) learning algorithm. The prediction outputs of the ANN-based path loss model has the lowest Root Mean Square Error (RMSE) of 4.480 dB, Standard Error Deviation (SED) of 4.479 dB, and the highest value of correlation coefficient (R) of 0.917, relative to the measured path loss values. This finding was further validated by the results of Analysis of Variance (ANOVA) and multiple comparison post-hoc tests. In essence, ANN-based path loss model was found to be the optimal model for heterogeneous radio network planning, deployment, and optimization in a smart campus propagation environment.

Keywords: Path Loss Model; Heterogeneous Radio Network; Artificial Neural Network (ANN); Support Vector Machine (SVM); Radio Network Planning and Optimization (RNP/O); Smart Campus

CHAPTER ONE

INTRODUCTION

1.1. Background of the Study

Executives, managers, and administrators in Higher Education Institutions (HEIs) may increase system efficiency and reduce running cost of activities on campus by exploiting the potentials of emerging enabling technologies such as Internet of Things (IoT), cloud computing, Big Data analytics, artificial intelligence, and fifth generation (5G) wireless communications (Adamkó, 2017; Bakken et al., 2017; Ogawa & Shimizu, 2017; Uskov et al., 2017). For example, location intelligence data gathering can help to optimize services, safety, space management, and asset utilization. Intelligent parking services can be provided to optimize driving route, reduce traffic congestion and minimize incidence of vehicle theft on university campuses through combined deployment of Wireless Fidelity (Wi-Fi) communication networks, high resolution video cameras, real-time video analytics, and sensor-enabled parking management applications. Also, the use of smart lighting systems in classrooms, administrative and residential buildings, and street lights will help in optimizing energy efficiency, and reduce carbon footprint. In addition, challenges in the areas of waste management, retail service delivery, transportation, water quality, lecture delivery, and attendance capturing can be handled through efficient system automation. Ultimately, the concept of 'smart campus' is tailored towards achieving a sustainable educational system by improving learning, working, and living experiences of students, staff, and faculty through the use of Information and Communication Technologies (ICTs).

Mobile communication systems provide the network infrastructure required for efficient and seamless delivery of IoT-enabled smart applications and services. Communication between sensor nodes, humans, objects, buildings, and machines can be established through different radio access technologies such as Wi-Fi, Global System for Mobile communications (GSM), Digital Cellular System (DCS), Universal Mobile Telecommunications System (UMTS), and Long Term Evolution (LTE). According to 2017 Cisco Visual Networking Index Report (Cisco, 2017), the demand for mobile connectivity has continued to increase at an exponential rate across the globe. For example, the average monthly mobile data traffic in 2016 was estimated to be 7.2 billion Gigabytes (an equivalent of 7.2 Exabytes). A high proportion of this exponential growth in global mobile data traffic was attributed to smartphones, IoT device, sensor node, and Machine-to-

Machine (M2M) wireless connections. Within the next three years, the number of mobile-connected devices is expected to leap to an estimated total of 12 billion. Hence, the number of devices, machines, vehicles, objects, applications, processes, and services that will require mobile connectivity for their proper functioning will soon outnumber the projected global human population of 7.8 billion by 2021.

In the light of the above, deployment of more Radio Access Networks (RANs) have become expedient to meet the exponential growth in number and demands of mobile users without compromising the expected Quality of Service (QoS). On the other hand, commissioning additional base stations, especially on campuses that are driven towards system 'smartness', will address the problem of radio coverage holes in order to improve customer satisfaction. This will, in return, boost the revenue base of mobile network operators. However, the need for greater cellular network capacity will drastically increase the rate of deployment of base stations, making the determination of suitable locations more difficult.

The design of mobile communication networks requires a good knowledge of the wireless channel (N. Faruk, Y. A. Adediran, & A. A. Ayeni, 2013; Oseni, Popoola, Enumah, & Gordian, 2014). It largely determines the transmission rate and the quality of signal propagation due to its complexity and randomness (Sotirios P Sotiroudis & Katherine Siakavara, 2015). Interactions between radiated electromagnetic waves and physical objects in wireless propagation environment often result in reflection on large plane surfaces, scattering from surfaces of small size relative to the wavelength of transmission, transmission through dense materials like walls or floors, or shadowing by obstacles such as buildings and foliage. Therefore, radio waves that are transmitted by the base station antennas reach mobile devices through different propagation paths, depending on the environment. This often results in signal fading, which may be in small or large scale (Rappaport, 1996). Small-scale signal fading occurs due to rapid fluctuations of Received Signal Strength (RSS) over a short period of time and small distance (Phillips, Sicker, & Grunwald, 2013); conversely, large-scale signal fading takes place as average signal strength changes over a large distance between the base station and the mobile station (Nasir Faruk, Y. A. Adediran, & A. A. Ayeni, 2013). The effect of large-scale fading is also known as *path loss*.

Path loss prediction models are vital tools for radio coverage estimation, determination of base station location, frequency allocation, antenna selection, and interference feasibility studies during

radio network planning (Popoola, Badejo, Ojewande, & Atayero, 2017). Prior to actual mobile network deployment, radio engineers use these models to understand wireless channel characteristics and to predict signal attenuation. An accurate path loss prediction model is required to estimate the strength of radio signal received as the mobile user moves farther away from the serving base station (Fernández Anitzine, Romo Argota, & Fontán, 2012). Radio propagation models help in deciding the optimum locations for base stations. Correct choice of base station locations will ensure the delivery of highest possible data rate (Rappaport, 1996). The use of path loss models for radio network planning and optimization usually saves costs and time (Isabona, Konyeha, Chinule, & Isaiah, 2013; Ostlin, Zepernick, & Suzuki, 2010).

Propagation models can be broadly organized into two categories, namely: deterministic and empirical models. Deterministic models are based on theoretical principles of diffraction (Luebbers, 1984), ray tracing (Mohtashami & Shishegar, 2012), integral equation (Hufford, 1952), and parabolic equation (Zelley & Constantinou, 1999); while empirical models are based on practical measurements conducted in a particular environment. Although deterministic models are more accurate, they lack computational efficiency. Empirical models such as Okumura-Hata model (Hata, 1980), *COopération européenne dans le domaine de la recherche Scientifique et Technique* (COST) 231 model (V Erceg, 1999), and standard propagation model (S. I. Popoola & O. F. Oseni, 2014b) are easy to implement with satisfactory computational efficiency in terms of time and cost. However, they are not as accurate as deterministic models because they do not effectively account for the unique geographical configurations of the propagation environment. Meanwhile, the reliability of the radio access network depends on the accuracy of the propagation model employed. Hence, the need for significant improvement in the prediction accuracy of empirical models while maintaining model simplicity and ease of use.

Radio propagation environments have been widely categorized into rural, suburban, and urban (Rappaport, 1996). These environments are composed of varying unique geographical features with different altitude, terrain height information, land usage data, building shape and height information, and building surface characteristics. Of all available empirical models, previous research works (Al Salameh & Al-Zu'bi, 2015; Faruk, Ayeni, & Adediran, 2013; Ibhaze, Ajose, Atayero, & Idachaba, 2016; Nimavat & Kulkarni, 2012; Oseni, Popoola, Abolade, & Adegbola, 2014a; S. I. Popoola & O. F. Oseni, 2014a; Rath, Verma, Simha, & Karandikar, 2016) have

identified the Hata, COST 231, and SPM path loss prediction models as appropriate for radio network planning in the 1800 MHz band. These models accounts for propagation path loss based on radio parameters including heights of transmitter and receiver antennas, frequency of transmission, and distance between the base station and the mobile station. However, the presence of various sources of clutter in the propagation environment contributes largely to propagation path loss (Oseni, Popoola, Enumah, et al., 2014). Modelling large scale channel fading without consideration for altitude, land use, and clutter height results in path loss predictions with large deviation from real measurement value poles (Mitra & Reddy, 1987).

Okumura-Hata model is an empirical formulation of the graphical path loss data that was collected at 150-1500 MHz band (Hata, 1980). The separation distance between the transmitter and the receiver ranges from one to 20 km. The appropriateness of the empirical model for path loss prediction in practical environments has been widely investigated in the literature. Medeisis and Kajackas (2000) investigated the suitability of Okumura-Hata model for path loss prediction in different Very High Frequency (VHF) and Ultra-High Frequency (UHF) bands. Although the empirical model performed fairly well in a built-up environment, the prediction error was significant in a rural propagation environment. A least square technique was applied to the model to reduce the high prediction error. The prediction accuracy of Okumura-Hata model was enhanced by Schneider, Lambrecht, and Baier (1996) with the details of the morphology and buildings in the wireless channel. The findings of the authors showed that better prediction accuracy will be obtained if morphological data and building data are incorporated into the model for rural/suburban and urban environments respectively. Farhoud, El-Keyi, and Sultan (2013) examined the applicability of Okumura-Hata model in the GSM 900 MHz band and introduced correction factors to improve the accuracy of the model for different regions in Egypt.

Akhoondzadeh-Asl and Noori (2007) suggested another way of defining the antenna height of the base station in Okumura-Hata model. The empirical model was adapted for path loss predictions by performing a cubic regression on field measurement data that was collected by Nadir and Ahmad (2010). Mardeni and Pey (2010) optimized Okumura-Hata model for urban outdoor coverage in the Code Division Multiple Access (CDMA) system in Malaysia. Major differences were found in the parameters of Okumura-Hata model when it was applied to railway environment at 900 MHz and this findings were reported by Cota, Serrador, Vieira, Beire, and Rodrigues (2013).

Begovic, Behlilovic, and Avdic (2012) evaluated the applicability of a set of empirical models for WiMAX coverage planning at 3.5 GHz. Adeyemo, Ogunremi, and Ojedokun (2016) optimized Okumura-Hata model for LTE signal attenuation in Lagos, Nigeria using the Least Square method. The model was also tuned for TErrestrial Trunked RAdio (TETRA) mobile radio applications in Saudi Arabia as reported by Alamoud and Schütz (2012).

COST 231 extends Okumura-Hata model to cover the frequency range of 1500 to 2000 MHz (V Erceg, 1999). The transmitter antenna height and the receiver antenna height can be in the range of 30-200 m and 1-10 m respectively. In the study conducted in dense urban areas at 1800 MHz by Verma and Saini (2016), COST 231 model had the lowest RMSE with the most acceptable SED when compared to free space, Stanford University Interim (SUI), and ECC-33 models. In addition, SPM was developed based on the Hata path loss formulas (Popoola, Atayero, Faruk, Calafate, Adetiba, et al., 2017; Popoola, Atayero, Faruk, Calafate, Olawoyin, et al., 2017). It determines the large-scale fading of received signal strength over a distance range of 1–20 km. Therefore, it is appropriate for mobile channel characterization of popular cellular technologies such as GSM. Although distance is usually expressed in *km* in Hata formulas, Standard Propagation Model (SPM) accepts distance values in meters. SPM ignores the effects of diffraction, clutter, and terrain. It assumed that appropriate settings of the parameters which account for only one clutter class will cater for the influence of these external factors on signal propagation. The correction function for the mobile receiver antenna height was also ignored for $h_r \leq 1.5$ m since it has negligible values for an average mobile antenna height.

The learning ability in artificial intelligence may be exploited to reduce computational complexity and to improve prediction accuracy. Machine learning techniques may be exploited for path loss predictions in rural and urban propagation environments (Salman et al., 2017). Machine learning is an adaptive statistical tool that can be used to solve both regression and classification problems. The capability of machine learning techniques to model complex nonlinear functional relationships provides an opportunity to improve the accuracy of empirical path loss models with better computational efficiency.

1.2. Statement of the Problem

Signal path loss is very difficult to predict, especially at radio frequencies where signal propagation is achieved largely through Non-Line of Sight (NLOS) condition (Ileana Popescu, Nafornita, &

Constantinou, 2005). Despite the rich benefits that radio propagation path loss models offer, most of the reliable path loss models are computationally expensive to deploy (Axtell, 2000). Also, these models require detailed information about the intended propagation environment to produce accurate predictions (Huschka, 1994; McLeod, Bai, & Meyer, 2013). The behaviour of a wireless channel depends largely on the constituent properties of the propagation environment. Whereas, propagation environments differ from places due to the variations in land morphology, terrain, and clutters (Ileana Popescu et al., 2005). Therefore, radio propagation models should be designed to adequately represent the wireless channel where radio signals are intended to be propagated.

The following research gaps were identified to the best of my knowledge:

a) Although several path loss models have been developed in the literature, most of the models were not designed for heterogeneous (multi-frequency bands) radio network planning and optimization.
b) Different radio propagation techniques have been considered but the unique terrain factors of a university campus were not adequately accounted for in existing path loss models. Particularly, the behaviour of radio signals within the context of Nigerian propagation environment is yet to be investigated in the context of heterogeneous networks.
c) There is a need to experimentally determine: (a) the most appropriate kinds and number of input parameters that are required (b) the effect of input data normalization (c) the most suitable transfer functions at the hidden and output layers (d) the correct learning algorithm and (e) the optimum number of hidden neurons towards achieving an optimal machine learning-based path loss model for heterogeneous radio network planning in a smart campus.

1.3. Aim and Objectives

1.3.1. Aim

This research project is aimed at developing an optimal path loss prediction model for heterogeneous radio network planning, deployment, and optimization in a smart campus propagation environment using machine learning approach.

1.3.2. Objectives

The objectives of this research project are to:

a) Acquire and store large heterogeneous datasets of field measured received signal strength values and path loss values at varying longitude, latitude, altitude, elevation, clutter height, distance, and radio frequencies in a typical smart campus propagation environment.
b) Develop path loss prediction models for heterogeneous network planning and optimization in a smart campus using Artificial Neural Network (ANN) and Support Vector Machine (SVM) learning algorithms.
c) Evaluate the prediction accuracy and generalization ability of widely used empirical path loss models and the developed models with respect to the field measured path loss values.
d) Identify the most suitable path loss prediction model for heterogeneous radio network planning, deployment, and optimization in a smart campus propagation environment.

1.4. Research Methodology

a) An extensive field measurement campaign was conducted to obtain Received Signal Strength (RSS) values and path loss values at varying longitude, latitude, altitude, elevation, clutter height, distance, and available radio frequencies (900, 1800, and 2100 MHz) within the campus of Covenant University, Ota, Nigeria.
b) ANN-based path loss models were developed and implemented in MATrix LABoratory (MATLAB 2016a) software by: (i) varying the number of input variable requirement; (ii) evaluating the effect of input data normalization; (iii) changing the transfer functions at the hidden and output layers; (iv) varying the learning algorithm employed; and (v) increasing the number of hidden neuron to achieve the best prediction accuracy and generalization ability. SVM-based path loss model was developed and implemented in Waikato Environment for Knowledge Analysis (WEKA) software using SMOreg algorithm.
c) The prediction accuracy and generalization ability of Hata, COST 231, ECC-33, Egli, ANN-based, and SVM-based path loss models were evaluated based on the following statistical metrics and tests: Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), Standard Error Deviation (SED), correlation coefficient (R), Analysis of Variance (ANOVA), and multiple comparison post-hoc test.

d) The path loss prediction model with the least prediction error and the highest correlation relative to the measured path loss was considered as optimal for heterogeneous radio network planning, deployment, and optimization in a smart campus propagation environment.

1.5. Justification for the Research

There is a continuous exponential increase in the number of wireless devices, mobile applications and services that require connectivity to cellular networks in smart campus applications and service delivery. With the emerging trends in IoT and M2M communication technologies, more than 100 billion smart devices and sensors are expected to be connected by 2020. The capacity of the state-of-the-art cellular networks will not be sufficient to meet the high user requirements of future wireless communications. One of the easiest ways of increasing cellular network capacity is to deploy more base stations in a given coverage area. This will be done in the future to guarantee good quality of signal reception at every point within the coverage area.

Accurate and efficient path loss models are highly essential for signal power predictions at different points within the coverage area during radio network planning. Radio network engineers rely on path loss models to: (a) determine optimal locations of base stations; (b) achieve best possible data rates; (c) estimate radio coverage; (d) determine the required transmission power; (e) aid appropriate selection of antenna height and pattern; (f) perform efficient frequency allocation; (g) conduct radio network optimization to ensure acceptable quality of service; and (h) perform interference feasibility studies.

The practical deployment of the contribution of this research project will enhance the efficient and seamless delivery of IoT-enabled smart applications and services in a university campus. The provision of reliable heterogeneous radio networks in a university campus will promote sustainable development in the long run.

1.6. Scope of Study

The development of machine-learning based path loss prediction model for heterogeneous radio network planning, deployment, and optimization is limited to GSM, DCS, and UMTS wireless

systems. Also, the use case considered in this research project is peculiar to the geographic terrain and land morphology of Covenant University campus.

1.7. Limitation of the Research

A drive test approach was used for radio signal measurement instead of a Continuous Wave (CW) measurement procedure because of unavailability of a wide-spectrum signal generator and a corresponding spectrum analyzer. Therefore, the control of radio network parameters is limited in drive test approach because it involves the use of commercial (already deployed) BTS. Likewise, the scope of this work does not cover current and emerging wireless technologies such as LTE, LTE-Advanced, and 5G radio networks.

1.8. Organization of Dissertation Chapters

This research project is organized into five chapters as follows: Chapter one presents the contextual background of the study, the problem statement of research, significance of the study, aim and objectives, research methodology, scope of the research project, and the application areas of contributions. Chapter two presents the theoretical background of the study and reviewed existing related work that are available in literature. Chapter three explains the methodology adopted in this research project. This chapter covers the materials and methods of field measurement campaign, model development process, and model performance evaluation. Results obtained are presented in Chapter four. The main findings of the research study are discussed in Chapter five. Chapter six presents the summary of the research study, recommendations, and contributions to scientific knowledge.

CHAPTER TWO

LITERATURE REVIEW

2.1. Introduction

In this chapter, the theoretical background of this research project is thoroughly reviewed to properly situate the findings of this research study within scientific body of knowledge. First, the fundamentals of radio propagation modelling is discussed in the context of wireless communication systems which operate at UHF bands. Then, various existing empirical path loss prediction models are discussed. Furthermore, model calibration process and the different techniques are explained. Finally, the review of related works are presented.

2.2. Fundamentals of Wireless Communication Systems

Communication primarily deals with the electrical transmission and reception of information signal from one point to another via a physical medium. This is done through a succession of processes to eliminate the distance barrier between the point of transmission and the receiver's location. Communication systems overcome the distance barrier, making the diverse information needed by different users available and accessible. These systems have found numerous applications in telephony, telegraphy, radio and television broadcasting, weather forecasting, internet services etc.

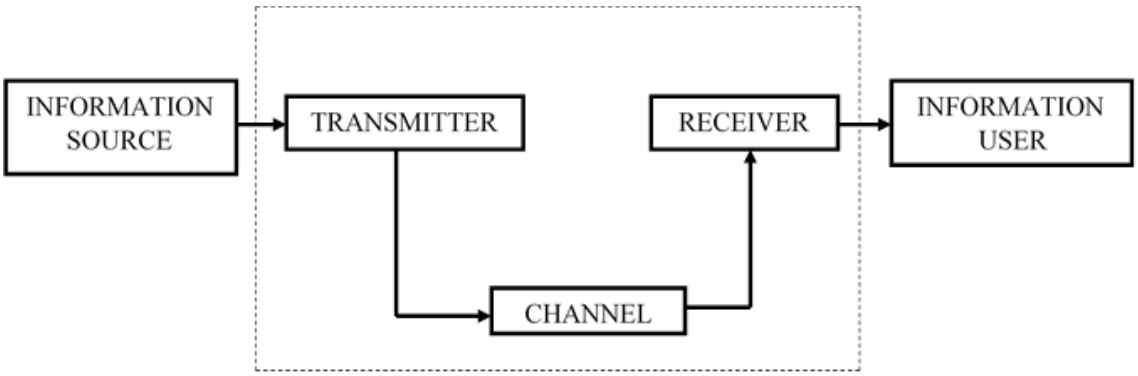

Figure 2.1. Elements of a Communication System

Irrespective of the form of communication adopted, there are three basic elements of a typical communication system namely (Proakis & Salehi, 2007): the *transmitter*, the *channel* and the *receiver* as shown in Figure 1. The *channel* is the physical medium that links the transmitter located

at a point relative to the receiver(s) situated at a distance away. The *transmitter* converts the original information signal into a form suitable for transmission over the channel through a *modulation* process while the *receiver* recovers the information signal from the modulated signal received through a reverse process known as *demodulation* (Sklar, 2001). Thus, the channel, being the interface between the two ends, plays an active role in the overall performance and efficiency of the communication system. It can be viewed objectively that the communication channel is pivotal to the efficient performance and operation of communication system since its properties determines the capacity and quality of service offered by the system. Depending on the mode of transmission, the communication channel may be a *guided propagation channel* or a *free propagation channel* (Haykin, 1988). Twisted pair cables channels, co-axial cables, and optical fibres are guided propagation channels while broadcast channel, mobile radio channel and satellite channel are free propagation channels.

Wireless channel introduces mobility into the public telecommunication network because of its ability to propagate through space without physical connections. This channel has a unique feature of propagating electromagnetic waves via NLOS paths. The propagation mechanisms of this channel are achieved through reflection, scattering, refraction, polarization and diffraction of electromagnetic waves over and around the edges and surfaces of the obstacles and obstructions present in the propagation environment (Aragon-Zavala, 2008). The transmitted signal reaches the mobile receiver through several paths. Hence, the mobile radio channel is a *multipath* propagation channel (Parsons, 2000). Here, multiple copies of the transmitted signal reach the mobile receiver from different directions and with different time delays. The propagation paths are of different amplitudes, phases, path lengths, and angles of arrival. Consequently, the received signal strength varies significantly with the location of the mobile receiver in a very complicated manner. The wireless channel is a random and linear time-varying channel because of its statistical nature (Simon & Alouini, 2005).

Wireless communication involves the electrical transmission and reception of voice and data using electromagnetic waves in an open space. The information signal is conveyed over a well-defined frequency band to the receiver(s). In wireless communication, two primary resources are employed namely: the transmitted power and the channel bandwidth (Proakis, Salehi, Zhou, & Li, 1994). The transmitted power is the average power of the transmitted signal while the channel bandwidth

is the band of frequencies allocated for the transmission of the message signal. The basic system design objective is to optimize the two limited resources as efficiently as possible. The frequency spectrum is a limited resource which is efficiently shared among numerous applications so as to avoid signal interference and it is usually auctioned by the government. Likewise, the mobile network operators do not have the luxury of increasing the transmitted power level beyond the maximum tolerable radiated power limit set to protect the people living close to the base station equipment from exposure to radiation hazards (Ahlbom et al., 2004; Repacholi, 1998, 2001).

Radio waves have an advantage of the ability to travel over long distances usually without a clear Line Of Sight (LOS). This is because they can penetrate into buildings. They can be easily generated and may be used for both indoor and outdoor communications. Radio waves are omnidirectional such that they can travel in all directions. They can be narrowly focused at high frequencies (greater than 100 MHz) using parabolic antennas. They are frequency-dependent in that they behave more like light at higher frequencies wherein they have direct paths and difficulty in passing through obstacles. The wave gets absorbed easily by rain, dust, snow, fogs etc. whereas they behave more like radio at lower frequencies, passing through obstacles and the power falls off sharply with distance from the source. This depends on the path loss exponent of the propagation environment. One of the basic challenges of radio waves is that they are more prone to interference from other sources (Chernov & Silverman, 1960; Rappaport & Sandhu, 1994).

Air is a frequency-selective channel. Information signal transmitted at different frequencies exhibits different behaviour when propagated through the free space. The rate of attenuation of the received signal strength increases as the frequency of transmission increases over the same distance. The propagation of electromagnetic waves is adversely affected by the atmospheric condition of the propagation environment. Electromagnetic waves of higher frequencies with few millimetre wavelengths get attenuated easily as the size of the wavelength of transmission tends towards the size of atmospheric agents such as rain droplets, dust, snow, fogs etc. Also, the design of high frequency communication systems is comparatively expensive in terms of cost and technicality; but they are indispensably used because the lower frequency bands are already assigned for other applications according to International Telecommunication Union (ITU) regulations (ITU, 2008).

At Very Low Frequency (VLF), Low Frequency (LF) and Medium Frequency (MF) bands, radio waves propagate along the ground. Here, the curvature of the earth is taken into consideration. This mechanism is adopted in Amplitude Modulation (AM) broadcasting which uses the MF band. At High Frequency (HF) bands, the ground waves tend to be absorbed by the earth. Thus, the propagation mechanism utilized is the ionospheric reflections (Bracewell, Budden, Ratcliffe, Straker, & Weekes, 1951; Kelso, 1964). The waves that reach the ionosphere (100 – 500 km above the earth surface) are reflected and sent back to the earth. The ionospheric reflection produces a multipath propagation. The vector sum of the multiple copies of the transmitted signals at the receiver results in a faded signal. The VHF transmission is restricted to a LOS transmission due to the higher frequency of transmission. Directional antennas are used to focus the transmitted signal from point to point. The waves follow direct path such that the LOS path must be strictly maintained, else the transmission path will be lost. The reflected waves interfere with the original signal which results in fading (Budden, 2009).

The field of wireless communications has been explored through a surge of research activities since the 1960s. The system has introduced mobility such that the users can easily roam anywhere, anytime. This eliminates the huge cost of cabling and its infrastructures. The freedom from wires also makes the installation easy and neat. Likewise, wireless communication network accommodates more users and provides global coverage.

Nevertheless, the wireless channel is faced with some technical challenges. There are some fundamental aspects that make wireless communication challenging than its wire line counterpart. The wireless channel exhibits multipath propagation characteristics which subject the transmitted signal to *fading* (Turin, Jana, Martin, & Winters, 2001). Also, the channel is highly susceptible to noise, interference and data corruption. The ease of access to the network without a need for wire connection makes the channel less secure. Thus, it requires a stronger security mechanism to protect data and/or bandwidth. Consequently, the susceptibility of the mobile radio channel to both signal fading and interference has made efficient radio network planning a vital part of the pre-deployment process. Cellular mobile operators and vendors have acknowledged the need for an efficient network of better QOS through a systematic radio network design process. This has brought radio network planning & optimization into a sharp focus (Laiho, Wacker, & Novosad, 2006; Lempiäinen & Manninen, 2003).

The reliability of the radio network planning process to produce a cost-effective deployment of Base Transceiver Stations (BTS) at the best possible sites for optimal network coverage largely depend on the degree of accuracy of the propagation prediction models in the network planning tools adopted (Oseni, Popoola, Abolade, & Adegbola, 2014b; S. Popoola & O. Oseni, 2014; S. I. Popoola & O. F. Oseni, 2014b). Thus, the accuracy of the propagation prediction models plays a pivotal role in obtaining an optimal network performance. The prediction models estimates the received signal strength at different locations of the area given the necessary radio network parameters. The reliability of the accuracy of this depends to a great extent on the degree to which the unique localized features in the area under study are taken into consideration in the radio propagation models.

The main cause of poor radio network deployment strategies is rooted in the inability of the engineer to accurately represent the propagation environment terrain, clutter type and the presence of physical features in the propagation models; whereas, these factors influence the propagation mechanism of the transmitted signals. Consequently, all these are responsible for the transmission path loss in the radio network. The limitation above has proven to be responsible for the wide variation between the results of the predicted received signal strengths and the actual signal strengths measured on live network deployed.

Through the years, much has changed in the field of wireless mobile communication. Despite these changes, the fundamental importance of propagation has continued undiminished. All the wireless systems are subjected to the variations imposed by the wireless channel, and a good understanding of these variations is needed to determine how far and fast can voice and data signals be transmitted and to answer the basic question of the number of users that the system can efficiently support.

Moreover, all the effects of the antennas (both transmit and receive antennas) and the propagation environment through which electromagnetic waves propagates must be accounted for in order to understand and analyze the performance of wireless communication systems. Although the general features of the propagation environment are taken into consideration in the planning tool, it is proven that there are yet some minor but consequential characteristics which uniquely distinguish the area from other similar areas in other places.

2.3 Mathematical Modelling of Radio Propagation Channel

Radio propagation model is a mathematical formulation for the characterization of radio wave propagation as a function of frequency of transmission, distance and other conditions that influence the behaviour of the radio channel in a given propagation environment (Abhayawardhana, Wassell, Crosby, Sellars, & Brown, 2005). Models are usually developed to predict the behaviour of propagation for all similar links under similar constraints. It provides a platform for simulating the behavioural characteristics of the radio channel before the proper deployment of the cellular mobile network. This is necessary because the mobile communication systems are expensive to deploy and any deficiency in the network planning can lead to an unnecessary cost expenses as a corrective measure. Path loss models are useful planning tools which allow the radio network designer to reach network optimal levels for the base station deployment and configuration while meeting the expected service level requirement. In order to explore the capacity of transmission in a wireless environment and also to develop suitable algorithms, there is a need to understand the concept of mathematical model of the environment. Propagation models are used in the design and development of wireless communication networks (Nawrocki, Aghvami, & Dohler, 2006).

Modelling the propagation channel has been one of the most difficult parts of mobile radio system design. The wireless channel places fundamental limitations on the performance of wireless communication system. This makes the channel very hostile to radio propagation. The tremendous growth of wireless communication systems and especially mobile radio system requires radio coverage prediction models that provide accurate results and fast processing time for several types of environments. This includes a large number of parameters describing the propagation environment. Due to the complexity and instability of radio transmission in wireless communication channels, it is very important to get the knowledge about the characteristics of wireless communication and use the suitable propagation models.

Radio propagation models have been developed as suitable, low cost and convenient system design alternatives since site measurements are costly. Channel modelling is required to predict path loss associated with the design of cellular network base stations, as this informs the design engineers how much power a transmitter need to radiate so as to service a given cell site. A typical network consists of a transmitter, a receiver and the surrounding environment. A model can be used for a certain frequency band to predict, to a high degree of accuracy, the behaviour of radio signal in a

particular environment/terrain. The performance of a communication system depends on *design parameters* whose values can be selected by the system designer and *environmental parameters* over which the designer has no control (Jeruchim, Balaban, & Shanmugan, 2006)

Path loss is the reduction in power density of an electromagnetic wave as it propagates through space. This is influenced by terrain contours, environment (urban or rural, vegetation and foliage), propagation medium (dry or moist air), the distance between the transmitter and the receiver, and the height and location of antennas. Path loss is a major component in the analysis and design of the link budget of a telecommunication system (S. I. Popoola & O. F. Oseni, 2014a).

Radio propagation prediction is one of the fundamentals of radio network planning. The coverage planning, frequency planning, capacity planning and interference analysis all rely on the propagation predictions. It is therefore vital that the propagation prediction models are as accurate as possible, taking into account the practical limitations of the propagation environment. The efficiency of the radio network planning process depends largely on the degree of accuracy of the propagation prediction models used. The attenuation of radio waves varies with different terrain and clutter types. The topography of the planned area, the location of the city, roads, water bodies, buildings and other hotspots have direct implication on the propagation predictions. Over the years, various propagation path loss models have been developed for the assessment of the performance of wireless communication systems for high quality of service delivery.

2.3.1 Free Space Path Loss Model

Development of radio propagation models can be retraced back to Friis' free space simple transmission equation of 1940s. Harald T. Friis emphasized the utility of the following simple transmission for a radio circuit made up of a transmitting antenna and a receiving antenna in free space (Friis, 1946). The ratio of transmitted power to received power is given by Equation (2.1):

$$\frac{P_r}{P_t} = \frac{A_r A_t}{d^2 \lambda^2} \tag{2.1}$$

Where P_r = Power fed into the transmitting antenna at its input terminals

P_t = Power available at the output terminals of the receiving antenna

A_r = Effective area of the receiving antenna

A_t = Effective area of the transmitting antenna

D = Distance between antennas

ƛ = Wavelength

The effective area of any antenna (A_{eff}), whether transmitting or receiving, is defined for the condition in which the antenna is used to receive a linearly polarized, plane electromagnetic wave. This is represented by Equation (2.2):

$$A_{eff} = \frac{P_r}{P_0} \tag{2.2}$$

In this case, P_o is the power flow per unit area of the incident field at the antenna. Thus, the P_r in Equation (2.2) shows that the received power is equal to the power flow through an area that is equal to the effective area of the antenna.

The power flow per unit area at the distance (d) from the transmitter is given by Equation (2.3):

$$P_0 = \frac{P_t}{4\pi d^2} \tag{2.3}$$

Substituting Equation (2.3) into Equation (2.2) gives Equation (2.4):

$$P_r = \frac{P_t A_{eff}}{4\pi d^2} \tag{2.4}$$

Replacing the isotropic transmitting antenna with a transmitting antenna of effective area A_t will increase the received power by the ratio $\frac{A_t}{A_{iso}}$ such that:

$$\frac{P_r}{P_t} = \frac{A_r A_t}{4\pi d^2 A_{iso}} \tag{2.5}$$

The need for Equation (2.5) is as a result of the fact that the transmitting antenna is not practically an ideal isotropic radiator.

Antennas were characterized by their effective areas; but in practice, antennas are characterized by their gains (G) relative to an ideal isotropic radiator as given by Equation (2.6) (Schelkunoff & Friis, 1952):

$$A = \frac{ƛ^2 G}{4\pi} \tag{2.6}$$

Substituting Equation (2.6) into Equation (2.5) gives Equation (2.7):

$$\frac{P_r}{P_t} = \frac{\lambda^2 G_t}{4\pi} \cdot \frac{\lambda^2 G_r}{4\pi} \cdot \frac{1}{\lambda^2 d^2}$$

$$\frac{P_r}{P_t} = G_t G_r \left(\frac{\lambda}{4\pi d}\right)^2 \quad (2.7)$$

Therefore, the inverse square relationship between path loss and distance is given by Equation (2.8):

$$\text{Path loss} \quad \alpha \quad \frac{1}{d^2} \quad (2.8)$$

The Friis' free space path loss is, therefore, expressed in decibel by Equation (2.9):

$$P_r(\text{dB}) = P_t(\text{dB}) + G_t(\text{dB}) + G_r(\text{dB}) + 20 \log\left(\frac{\lambda}{4\pi \text{d}}\right)$$

$$Path\ Loss\ (dB) = 32.44 + 20log(f) + 20log(d) \quad (2.9)$$

The free space propagation equation shows that the path loss (which is the ratio of the power received to the power transmitted) is inversely proportional to the square of the distance between the transmitter and the receiver. This model predicts that the received power decays as a function of the Transmitter-Receiver separation distance (Schelkunoff & Friis, 1952).

However, Friis' free space path loss model is only a valid predictor for the received power (P_r) for the values of *d* which are in the far-field region of the transmitting antenna. The far-field of a transmitting antenna is defined as *the region beyond the far-field distance 'd_f' which is related to the largest linear dimension of the transmitting antenna aperture and the carrier wavelength.*

W.D Lewis made a theoretical study of transmission between large antennas of equal areas with plane phase fonts at their apertures and he found that Equation (2.1) is only valid within a few per cent when Equation (2.10):

$$d_f \geq \frac{2d^2}{\lambda} \quad (2.10)$$

Where *d* is the largest linear dimension of either of the antennas.

Additionally, to be in the far-field region, Equation (2.11) and Equation (2.12) must be satisfied.

$$d_f \gg d \tag{2.11}$$

$$d_f \gg \lambda \tag{2.12}$$

More so, the free space propagation model is used to predict received signal strength when the transmitter and the receiver have a clear, unobstructed LOS between them. The equation only applies to free space conditions with no consideration for environmental influence and physical obstructions. Satellite communication systems and microwave LOS links typically undergo free space propagation.

Practically, this cannot be adopted for terrestrial wireless communication system deployment in urban areas which are known to be characterized with various material obstructions to the propagation of radio waves. These radio waves propagates through environments where they are reflected, scattered, and diffracted by walls, terrains, buildings and other objects.

2.3.2 Maxwell Equation Approach to Radio Propagation Modelling

Wireless channel poses severe challenge as a medium for reliable high-speed communication. It is not only susceptible to noise and interference; but also to other channel impairments which changes with time in unpredictable ways due to users' movement in mobile communication systems (Goldsmith, 2005).

The ultimate details of this propagation can be obtained by solving Maxwell's equations with boundary conditions that express the physical characteristics of these obstructing objects. This analysis involves the calculation of the radar cross section of the large and complex structures involved. The following assumptions are made in the development of the propagation model using Maxwell's equations:

1. The propagated signals are in the UHF and Super High Frequency (SHF) bands of 0.3-3 GHz and 3-30 GHz respectively.
2. The transmission distances on the earth are small enough so as to be affected by the earth's curvature.

3. All transmitted and received signals are considered to be real since modulators and demodulators are built using oscillators that generate real sinusoids (and are not complex exponentials).

4. Channels are typically modelled as having a complex frequency response due to the nature of the Fourier Transform.

The transmitted signal is thus represented by Equation (2.13):

$$s(t) = f[u(t)e^{j(2\pi f_c t + \emptyset_0)}]$$

$$= \{f[u(t)]\cos(2\pi f_c t + \emptyset_0)\} - \{g[u(t)]\sin(2\pi f_c t + \emptyset_0)\}$$

$$s(t) = x(t)\cos(2\pi f_c t + \emptyset_0) - y(t)sin(2\pi f_c t + \emptyset_0) \quad (2.13)$$

Where $u(t) = x(t) + jy(t)$ is a complex baseband signal with in-phase component $x(t) = f[u(t)]$, quadrature component $y(t) = g[u(t)]$, bandwidth B, and power P_u. The signal $u(t)$ is called the complex envelope of $s(t)$.

The power in the transmitted signal $s(t)$ is given by Equation (2.14):

$$P_t = \frac{P_u}{2} \quad (2.14)$$

The received signal will be of the form given by Equation (2.15):

$$r(t) = f[v(t)e^{j(2\pi f_c t + \emptyset_0)}] \quad (2.15)$$

Where the complex baseband signal $y(t)$ will depend on the channel through which $s(t)$ propagates.

Suppose $s(t)$ of power P_t is transmitted through a given channel, with corresponding received signal $r(t)$ of power P_r, where P_r is averaged over any random variations due to shadowing. Then, linear path loss of the channel is defined as the ratio of transmit power to receive power as given by Equation (2.16).

$$Path\ Loss = \frac{P_t}{P_r} \quad (2.16)$$

Path loss of the channel is defined as the decibel value of the linear path loss or, equivalently, the difference between the transmitted and received power as given by Equation (2.17):

$$Path\ Loss = 10\ log\left(\frac{P_t}{P_r}\right) \tag{2.17}$$

However, the analysis involves the calculation of the radar cross section of the large and complex structures. This is responsible for many difficulties frequently encountered in the process of calculation because the necessary parameters are not available on most occasions. To this effect, various approximation techniques have been developed to characterize signal propagation without resorting to Maxwell equations.

2.3.3 Empirical Radio Propagation Path Loss Models

Empirical models are those based on observations and measurements alone. These models are mainly used to predict the path loss, but models that predict rain-fade and multipath have also been proposed (Seybold, 2005). The deterministic models make use of the laws governing electromagnetic wave propagation to determine the received signal power at a particular location. Deterministic models often require a complete three-dimensional (3D) map of the propagation environment. An example of a deterministic model is a ray tracing model (Glassner, 1989; Schaubach, Davis, & Rappaport, 1992). Stochastic models, on the other hand, model the environment as a series of random variables (Abhayawardhana et al., 2005). These models are the least accurate but require the least information about the environment and use much less processing power to generate predictions.

The empirical models account in principle, for all the major mechanisms which are encountered in macro-cell prediction. However, to use such models would require detailed knowledge of the location, dimension and constitutive parameters of every tree, building and terrain feature in the area to be covered. This is too complex and would anyway yield an unnecessary amount of detail. One of the appropriate ways of accounting for these complex effects is through an empirical model. To design and develop such model, an extensive set of actual path loss measurements is made, and an appropriate function is fitted to the measurements, with parameters derived for the particular environment, frequency and antenna heights so as to minimize the error between the model and the measurements. Each measurement represents an average of a set of samples taken over a small area (around 10–50m), in order to filter out the effects of fast fading (Clarke, 1968). The model can then be used to design systems operated in similar environments to the original measurements.

2.3.3.1 Okumura-Hata Path Loss Model

Okumura's model is one of the most widely used empirical models for signal prediction in urban areas (Hata, 1980). This model is applicable for frequencies in the range 150 MHz to 1920 MHz and distances of 1 km to 100 km. The base station antenna height ranges from 30 m to 1000 m. The frequency range can be extrapolated up to 3 GHz for distances ranging from 1 km to 100 km.

Hata made extensive measurements in Tokyo city in Japan and present the results in an empirical path loss model. This model is discovered to be reasonably accurate that it is still in use in wireless communication industry, especially in Japan where the measurement campaigns were done. The propagation model is characterized with the following variables:

(a) frequency of operation

(b) transmitter and receiver antenna heights

(c) terrain parameters

The resulting simple path loss equation is given by Equation (2.18) (Hata, 1980):

$$Path\ Loss\ (dB) = L_f + A_{mu}(f,d) - G(h_t) - G(h_r) - G_{AREA} \qquad (2.18)$$

Where L_f = free Space Path Loss

$A_{mu}(f,d)$ = median attenuation relative to free space

$G(h_t)$ = base station antenna height gain factor

$G(h_r)$ = mobile station antenna height gain factor

G_{AREA} = gain due to the type of environment

f = frequency in MHz

h_t = transmitter antenna height in meters

h_r = receiver antenna height in meters

d = distance between transmitter and receiver antennas

$$G(h_t) = 20\ log\left(\frac{h_t}{200}\right) \qquad \text{for } 1000\text{m} < h_t < 3\text{m} \qquad (2.19)$$

$$G(h_r) = 10\ log\left(\frac{h_r}{3}\right) \quad \text{for} \quad h_r \leq 3m \tag{2.20}$$

$$G(h_r) = 20\ log\left(\frac{h_r}{3}\right) \quad \text{for} \quad 10m > h_r > 3m \tag{2.21}$$

The set of curves below provide the values of $A_{mu}(f,d)$ and G_{AREA} at different frequency of operation between 150 MHz and 3000 MHz. These curves as shown in the figure below were developed from extensive measurements using vertical Omni-directional antennas at both the base and mobile stations, and are plotted as a function of frequency in the range 100 MHz to 1920 MHz and as a function of distance from the base station in the range 1 km to 100 km. To determine path loss using Okumura's model, the free space path loss between the points of interest is first determined, and then $A_{mu}(f,d)$, as read from the curves, is added to it along with correction factors to account for the type of terrain.

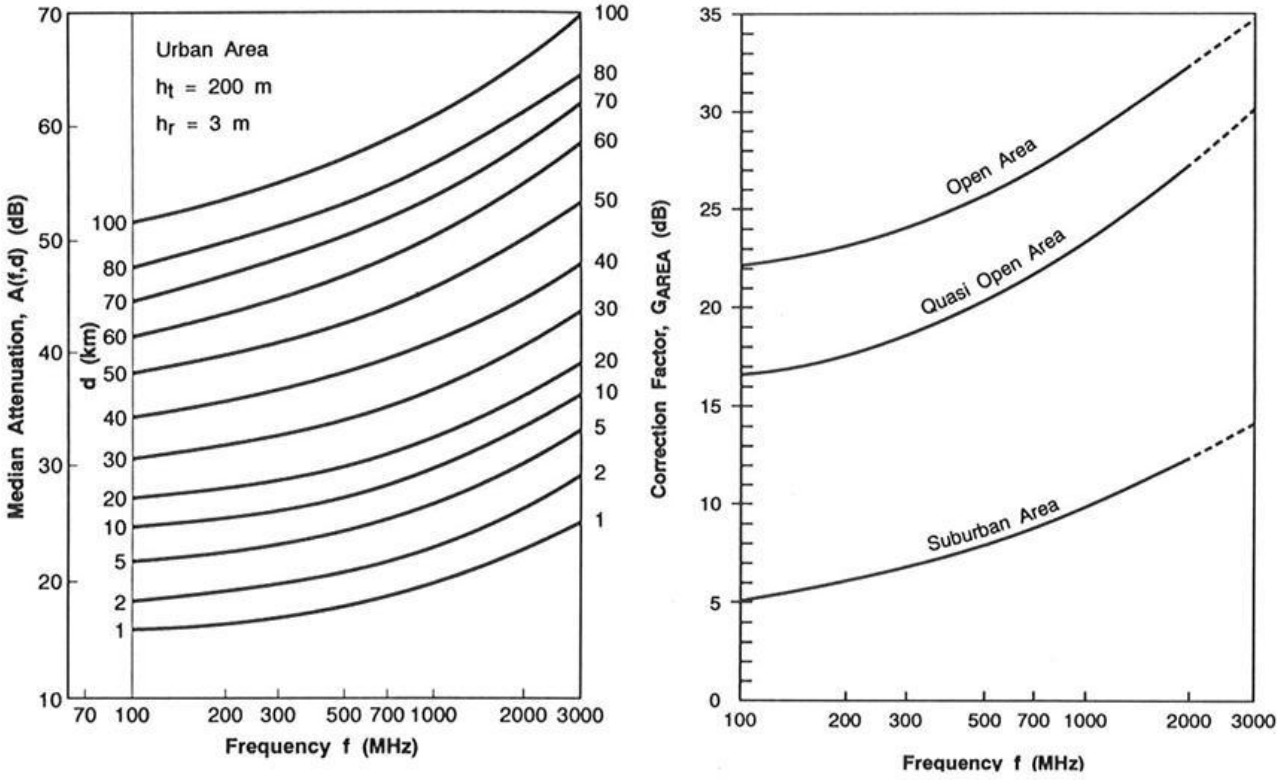

Figure 2.2. Median Attenuation and Area Gain Factor (Hata, 1980)

Okumura Hata path loss model is wholly based on measured data and does not provide any analytical explanation. This model is considered to be among the simplest and best in terms of accuracy in path loss prediction for cellular and land mobile radio system in cluttered

environments. Typically, the SED between the predicted and measured path loss values is between 10 dB and 14 dB. The major disadvantage with the model is its slow response to rapid changes in terrain. Therefore, the model is fairly good in urban and suburban areas, but not as good in rural areas.

2.3.3.2 Hata Path Loss Model

Hata (1980) developed a model which is an empirical formulation of the graphical path loss data provided by Okumura; and is valid from 150 MHz to 1500 MHz. Here, the urban area propagation path loss is presented as a standard formula and correction equations are provided for application to other situations. The standard formula for median path loss in urban areas is given by Equation (2.22):

$$PL_{urban}(dB) = 69.55 + 26.16\log(f_c) - 13.82\log(h_t) - a(h_r) + [44.9 - 6.55\log(h_t)]\log(d) \quad (2.22)$$

where f_c = Frequency (in MHz) from 150 MHz to 1500 MHz

h_t = Effective transmitter antenna height (in metres): 30 m to 200 m

h_r = Effective receiver antenna height (in metres): 1 m to 10 m

d = Separation distance (in km): 1 km to 20 km

$a(h_r)$ = Correction factor for effective mobile antenna height

For a small to medium-sized city,

$$a(h_r) = [1.1\log(f_c) - 0.7]h_r - [1.56\log(f_c) - 0.8] \quad (2.23)$$

For a large city,

$$a(h_r) = 8.29\,[\log(1.54h_r)]^2 - 1.1 \quad \text{for} \quad f_c \leq 300 \text{ MHz} \quad (2.24)$$

$$a(h_r) = 3.2\,[\log(11.75h_r)]^2 - 4.97 \quad \text{for} \quad f_c \geq 300 \text{ MHz} \quad (2.25)$$

For a suburban area,

$$PL_{suburban} = PL_{urban}(dB) - 2[\log(\tfrac{f_c}{28})]^2 - 5.4 \quad (2.26)$$

For an open rural area,

$$PL_{rural} = PL_{urban}(dB) - 4.78[\log\ (f_c)]^2 - 18.33\log(f_c) - 40.98 \quad (2.27)$$

The predictions of the Hata model compare very closely with the original Okumura model, as long as *d* exceeds 1 km. The formula shows that the Base Station antenna height has a strong effect on the degree of path loss because it appears as a multiplicative factor for the logarithmic distance. This model is well suited for large cell mobile systems, but not Personal Communication System (PCS) which have cells of the order of 1 km radius.

Unfortunately, Okumura model does not lend itself well to computer analysis. For this reason, Hata (1980) established parametric equations that fit the curves reported by Okumura. Using Hata's model, an engineer need only to plug the parameters into the empirically derived formulas.

2.3.3.3 COST-231 Extensions

COST extended the Hata model for scientific frequencies of interest (Action, 1999). The model, which was renamed COST – Hata model, is applicable for only cases in which the antenna heights are above the rooftops of the surrounding buildings. The COST committee also suggests a combination of work done by Walfisch and Ikegami to come up with a model called COST–Walfisch–Ikegami (or COST-WI), characterizing urban environments. Here, the heights of buildings, the widths of roads, the building spacing, and the road orientation relative to the transmitter are taken into account. All of the parameters are characteristic of the entire coverage area.

2.3.3.4 COST 231–Hata Path Loss Model

COST 231 has extended Hata's model to the frequency band of 1500 MHz ≤ f_c ≤ 2000 MHz by analyzing Okumura's propagation curves in the upper frequency band. The proposed model for path loss is given by Equation (2.28):

$$PL(dB) = 46.3 + 33.9\log(f_c) - 13.82\log(h_t) - a(h_r) + [44.9 - 6.55\log(h_t)]\log(d) + C_m \quad (2.28)$$

For small to medium-sized city,

$$a(h_r) = [1.1\log(f_c) - 0.7]h_r - [1.56\log(f_c) - 0.8] \quad (2.29)$$

For a large city,

$$a(h_r) = 8.29\,[\log(1.54h_r)]^2 - 1.1 \quad \text{for} \quad f_c \leq 300\ \text{MHz} \tag{2.30}$$

$$a(h_r) = 3.2\,[\log(11.75h_r)]^2 - 4.97 \quad \text{for} \quad f_c \geq 300\ \text{MHz} \tag{2.31}$$

and,

$$C_m = \begin{cases} 0\ \text{dB} & \text{for medium} - \text{sized city and suburban areas} \\ 3\ \text{dB} & \text{for metropolitan areas} \end{cases} \tag{2.32}$$

Range of parameters

f : 1500 – 2000 MHz

h_t : 30 – 200 m

h_r : 1 – 10 m

d : 1 km – 20 km

To show the COST-231 Hata model in a simpler form, the model (Ray, 2007) is given by Equation (2.33):

$$Path\ Loss\ (dB) = L_o + nlog(f) - 13.82\log(h_b) - C_H + [\sigma - 6.55\log(h_b)]\log(d) + C_m \tag{2.33}$$

Where L = Median path loss in decibel (dB)

f = Frequency of transmission, in megahertz (MHz)

h_b = Base station antenna height, in meters (m)

d = Link distance, in kilometer (km)

C_H = Mobile station antenna height correction factor

and

L_0 = 46.3

σ = 44.9

n = 33.9

2.3.3.5 Standard Propagation Model

Standard Propagation Model (SPM) is based on the Hata formulas and is suitable for predictions in the 150 – 3500 MHz frequency band over long distances ranging from 1 – 20 km. It is best suited to GSM 900 and GSM 180, UMTS, CDMA 2000, Worldwide Interoperability for Microwave Access (WiMAX) and LTE radio technologies (Oseni, Popoola, et al., 2014b; S. Popoola & O. Oseni, 2014; S. I. Popoola & O. F. Oseni, 2014b)

The model is based on the formula given by Equation (2.34):

$$P_r = P_t - \{K_1 + K_2 \log(d) + K_3 \log(h_t) + K_4 . Diffraction\ Loss + K_5\ log(d).log(h_t) + K_6 . h_r + K_7 \log(h_r) + K_{clutter} . f_{clutter} + K_{hill}\} \quad (2.34)$$

Where,

P_r = Received power in dBm

P_t = Transmitted power (EIRP) in dBm

K_1 = Constant offset in dB

K_2 = Multiplying factor for $\log(d)$

d = Distance between the receiver and the transmitter in metres

K_3 = multiplying factor for $\log(h_t)$

h_t = Effective transmitter antenna height in metres

K_4 = Multiplying factor for diffraction calculation

K_5 = Multiplying factor for $log(d).log(h_t)$

K_6 = Multiplying factor for h_r

K_7 = Multiplying factor for $\log(h_r)$

h_r = Effective mobile receiver antenna height in metres

$K_{clutter}$ = Multiplying factor for $f_{clutter}$

$f_{clutter}$ = Average of the weighted losses due to clutter

K_{hill} = Corrective factor for hilly region

The SPM formula is derived from the basic Hata formula given by Equation (2.35):

$$PL\ (dB) = A_1 + A_2 log(f) + A_3 log(h_t) + [B_1 + B_2 log(h_t) + B_3 . h_t][log(d)] - a(h_r) - C_{clutter} \quad (2.35)$$

Where,

$A_1 \ldots B_3$: Hata parameters

f : Frequency in MHz

h_t : Effective transmitter antenna height in metres

d : Distance in Km

$a(h_r)$: Mobile receiver antenna height in metres

$C_{clutter}$: Clutter correction function

It was observed that the distance in Hata formula is in km as opposed to the SPM, where the distance is given in metres. The typical values of the Hata parameters are:

$$A_1 = \begin{cases} 69.55 & for\ 900\ MHz \\ 46.30 & for\ 1800\ MHz \end{cases}$$

$$A_2 = \begin{cases} 26.16 & for\ 900\ MHz \\ 33.90 & for\ 1800\ MHz \end{cases}$$

$A_3 = -13.82$

$B_1 = 44.90$

$B_2 = -6.55$

$B_3 = 0$

Thus, for GSM 900,

$$PL\ (dB) = 69.55 + 26.16\ log(f) - 13.82\ log(h_t) + [44.9 - 6.55\ log(h_t)][log(d)] - a(h_r) - C_{clutter} \quad (2.36)$$

For GSM 1800,

$$PL\ (dB) = \ 46.3 \ + \ 33.9\ log(f) - 13.82\ log(h_t) + \ [44.9 - 6.55\ log(h_t)][log(d)] - a(h_r) - \ C_{clutter} \quad (2.37)$$

It was noted that the influence of diffraction and clutter correction can be ignored since the consideration is on standard formulas. With the appropriate settings of the A_1 and K_1, and taking only one clutter class into consideration, the clutter correction factor can be set to zero without altering the validity of the equations. The correction function for the mobile receiver antenna height can also be ignored for $h_r \leq 1.5$ m since it has negligible values for realistic mobile antenna heights. Then, the Hata formula becomes Equation (2.38):

$$PL\ (dB) = \ A_1 + \ A_2 log(f) + \ A_3 log(h_t) + \ [B_1 + \ B_2 log(h_t)][log(d)] \quad (2.38)$$

The SPM formula can also be reduced to:

$$PL\ (dB) = K_1 + \ K_2 \log(d) + \ K_3 \log(h_t) + \ K_5\ log(d).log(h_t) + \ K_6.h_r + \ K_7 \log(h_r) \quad (2.39)$$

Rewriting the simplified Hata equation with distance in metres as in SPM we have,

$$PL\ (dB) = \ A_1 + \ A_2 log(f) + \ A_3 log(h_t) + \ [B_1 + \ B_2 log(h_t)]\left[log\left(\frac{d}{1000}\right)\right]$$

$$PL\ (dB) = \ A_1 + \ A_2 log(f) + \ A_3 log(h_t) + \ [B_1 + \ B_2 log(h_t)][log(d\ x\ 10^{-3})]$$

$$PL\ (dB) = \ A_1 + \ A_2 log(f) + \ A_3 log(h_t) + \ B_1 log(d\ x\ 10^{-3}) + \ B_2 log(h_t).log(d\ x\ 10^{-3})$$

$$PL\ (dB) = \ A_1 + \ A_2 log(f) + \ A_3 log(h_t) + \ B_1 log(d) + \ B_1 log(10^{-3}) + \ B_2 log(h_t).[log(d) + \ log(10^{-3})]$$

$$PL\ (dB) = \ A_1 + \ A_2 log(f) + \ A_3 log(h_t) + \ B_1 log(d) - 3B_1 + \ [B_2 log(h_t)][log(d\ ) - 3]$$

$$PL\ (dB) = \ A_1 + \ A_2 log(f) - 3B_1 + \ [A_3 - 3B_2][log(h_t)] + \ B_1 log(d) + \ B_2 log(h_t).log(d) \quad (2.40)$$

Equating the coefficients of the Equation (2.39) and Equation (2.40):

$$K_1 = \ A_1 + \ A_2 log(f) - 3B_1$$

$$K_2 = \ B_1$$

$$K_1 = A_3 - 3B_2$$

$$K_5 = B_2$$

$$K_6 = K_7 = 0$$

Therefore, the SPM formula for GSM 900 is given by Equation (2.41):

$$PL\ (dB) = 12.5 + 44.9\ log(d) + 5.83\ log(h_t) - [6.55\ log(d).log(h_t)] \quad (2.41)$$

Also, the SPM formula for DCS 1800 is given by Equation (2.42):

$$PL\ (dB) = 22 + 44.9\ log(d) + 5.83\ log(h_t) - [6.55\ log(d).log(h_t)] \quad (2.42)$$

2.3.3.6 Bertoni - Walfisch Path Loss Model

Bertoni – Walfisch model is the first model which takes into account the effects of buildings on radio propagation channel in path loss modeling (Walfisch & Bertoni, 1988). The model assumes that propagation takes place over rows of buildings having equal heights and equal spacing arranged in a perfect grid. The model has been adopted by ITU for International Mobile Telecommunications (IMT-2000)/3G standard.

Walfisch & Bertoni (1998) proposed a semi-empirical model that is applicable to propagation through buildings in built-up environments. The model assumes building heights to be uniformly distributed and the separation between buildings are equal. The figure below illustrates the building geometry and parameters in the Bertoni - Walfisch model.

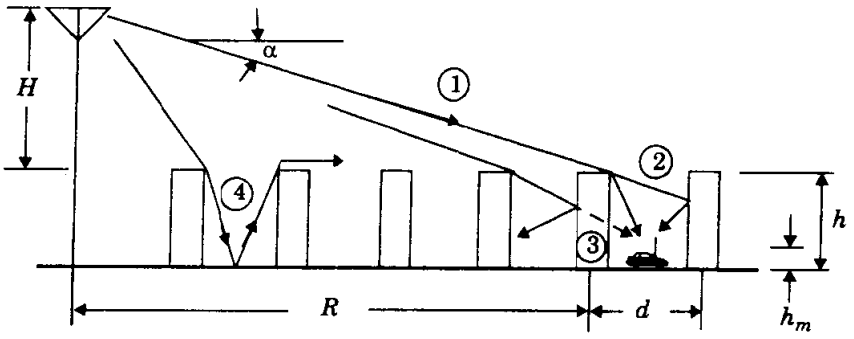

Figure 2.3. Propagation Geometry for Bertoni - Walfisch Model (Walfisch & Bertoni, 1988)

The model has three components namely:

(a) the free space loss (L_f);

(b) the loss due to diffraction by rooftops (L_{rt}); and

(c) the diffraction and scatter loss from rooftop down the street (L_{rs})

Therefore,

$$PL(dB) = L_f + L_{rf} + L_{rs} \tag{2.43}$$

Where $L_f = \left(\frac{\lambda}{4\pi r^2}\right)$

$$L_f(dB) = 32.44 + 20log(f) + 20log(d)$$

$$L_{rf} = \left\{0.1\left(\frac{\sin\alpha\sqrt{d/\lambda}}{0.03}\right)^{0.9}\right\}^2$$

Here, *sin α* is in terms of the transmitter height (h_t), the building height (h_b) and the distance R as given by:

$$\sin\alpha = \frac{h_t - h_b}{R}$$

$$L_{rf} = 0.01\left(\frac{h_t - h_b}{R}\right)^{1.8}\left(\frac{d}{\lambda}\right)^{0.9}$$

$$L_{rs} = \left(\frac{\lambda\rho}{h_b - h_r}\right).\left(\frac{1}{2\pi^2}\right)$$

The total path loss is thus given by Equation (2.44):

$$PL(dB) = 89.5 - 10\log\left\{\frac{\rho d^{0.9}}{(h_b - h_r)^2}\right\} + 21\log(f_m) - 18\log(h_t - h_b) + 38\log R \tag{2.44}$$

where $\rho = \sqrt{\left(\frac{d}{2}\right)^2 + (h_b - h_r)^2}$

f_m : frequency in MHz

h_t : transmitter antenna height in metres

h_r : receiver antenna height in metres

d : separating distance between buildings in metres

R : distance between transmitter and the receiver in metres

2.3.3.7 Stanford University Interim (SUI) Model

Institute of Electrical and Electronic Engineering (IEEE) 802.16 Broadband Wireless Access working group proposed the standards for the frequency band below 11GHz containing the channel model developed by Stanford University namely SUI model (Milanovic, Rimac-Drlje, & Bejuk, 2007). This prediction model comes from the extension of Hata model with frequency larger than 1900 MHz. The correction parameters are allowed to extend this model up to 3.5 GHz band. In the USA, this model is defined for the Multipoint Microwave Distribution System (MMDS) for the frequency band from 2.5GHz to 2.7GHz.

The following are the range of parameters involved:

Base station (transmitter) antenna height	:	10 – 80m
Mobile station (receiver) antenna height	:	2 – 10m
Cell radius	:	0.1 – 8km

The SUI model describes three types of terrain, they are terrain A, B and C. there is no declaration about any particular environment. The basic path loss expression is given as:

$$PL(dB) = A + 10\gamma\, log\left(\frac{d}{d_o}\right) + X_f + X_h + S \qquad for\ d > d_o \tag{2.45}$$

where d = distance between T_x and R_x antennas in metres

d_o = 100 m

λ = wavelength in metres

X_f = correction factor for frequency above 2 GHz (in MHz)

X_h = correction factor for receiver antenna height in metres

S = correction factor for shadowing (dB)

γ = path loss exponent

$$A = 20\, log\left(\frac{4\pi d_0}{\lambda}\right)$$

$$\gamma = a - bh_r\left(\frac{c}{h_r}\right)$$

The parameter h_r is the receiving antenna height in metres which is between 10m and 80m. The constants *a*, *b* and *c* depend on the type of terrain.

Note that,

$\gamma = 2$ for free space propagation

$3 < \gamma < 5$ for Urban NLOS

$\gamma > 5$ for indoor propagation

Table 2.1. SUI Model Parameters (Milanovic et al., 2007)

Model Parameter	**Terrain A**	**Terrain B**	**Terrain C**
A	4.6	4.0	3.6
b (m^{-1})	0.0075	0.0065	0.005
C (m)	12.6	17.1	20

$$X_f = 6\,log\left(\frac{f}{2000}\right)$$

$$X_h = \begin{cases} -10.8\log\left(\frac{h_r}{2000}\right) & for\ terrain\ A\ and\ B \\ -20\log\left(\frac{h_r}{2000}\right) & for\ terrain\ C \end{cases}$$

Where *f* is the frequency of operation in MHz

For the above correction factors, this model is extensively used for the path loss prediction of all three types of terrain in rural, urban and suburban environments.

2.3.3.8 ECC-33 Path Loss Model

Hata – Okumura model is one of the most extensively used empirical propagation models which are based on the Okumura model (Rappaport, 1996). This model is suited for the UHF band. The tentatively proposed propagation model of Hata-Okumura model with report (Abhayawardhana et al., 2005) is referred to as ECC-33 model. In this model, path loss is given by Equation (2.46):

$$PL(dB) = A_{fs} + A_{bm} - G_t - G_r \quad (2.46)$$

Where A_{fs} = free space attenuation

A_{bm} = basic median path loss

G_t = transmitter antenna height gain factor

G_r = receiver antenna height gain factor

$$A_{fs} = 92.4 + 20\log(d) + 20\log(d)$$

$$A_{bm} = 20.41 + 9.83\log(d) + 7.894\log(f) + 9.56\,[\log(f)]^2$$

$$G_t = \left[log\left(\frac{h_t}{200}\right)\right][13.958 + 5.8\,(\log d)^2]$$

For medium-sized cities,

$$G_r = [42.57 + 13.7\log(f)][\log(h_r) - 0.585]$$

For large cities,

$$G_r = 0.759h_r - 1.862$$

2.3.3.9 COST-231 Walfisch-Ikegami (WI) Model

This model is most suitable for flat suburban and urban areas that have uniform building height (Har, Watson, & Chadney, 1999).

The equation of the proposed model is given by Equation (2.47) and Equation (2.48):

$$PL(dB) = 42.6 + 26\log(f) + 20\log(d) \qquad for\ LOS\ condition \quad (2.47)$$

$$PL(dB) = L_f + L_{rf} + L_{rs} \qquad for\ urban\ and\ suburban\ if L_{rts} + L_{msd} > 0 \quad (2.48)$$

L_f = free space path loss

L_{rts} = roof top to street diffraction

L_{msd} = multi-screen diffraction loss

$$L_f(dB) = 32.44 + 20\ log\ (f) + 20\ log\ (d)$$

$$L_{rts} = -16.9 - 10\log\,(w) + 10\log\,(f) + 20\log\,(h_r) + L_{or} \qquad for\ h_{roof} > h_r$$

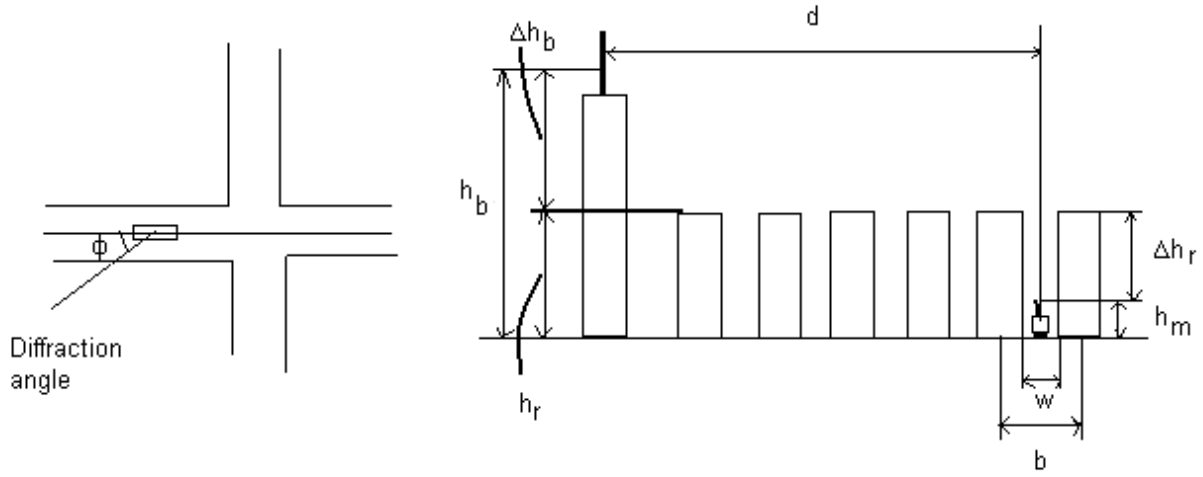

Figure 2.4. Diffraction Angle and Urban Scenario (Har et al., 1999)

where

$$L_{or} = \begin{cases} -10 + 0.354\phi & for \quad 0 \le \phi \le 35 \\ 2.5 + 0.075(\phi - 35) & for \quad 35 \le \phi \le 55 \\ 4 - 0.114(\phi - 55) & for\ 55 \le \phi \le 90 \end{cases}$$

Note that $\Delta h_r = h_{roof} - h_r$

$$\Delta h_{base} = h_{base} - h_{roof}$$

$$L_{msd} = L_{bsh} + K_a + K_d \log(d) + K_f \log(f) - 9\log(f) - 9\log(B) \qquad for\ L_{msd} > 0$$

$$L_{msd} = 0 \qquad for\ L_{msd} < 0$$

Where $L_{bsh} = \begin{cases} -18\log(1 + \Delta h_{base}) & for\ h_{base} > h_{roof} \\ 0 & for\ h_{base} \le h_{roof} \end{cases}$

$K_a = 54 \qquad for\ h_{base} > h_{roof}$

$K_a = 54 - 0.8\Delta h_{base} \qquad for\ d \ge 0.5km\ and\ h_{base} \le h_{roof}$

$K_a = 54 - 0.8\Delta h_{base}\left(\frac{d}{0.5}\right) \qquad for\ d < 0.5km\ and\ h_{base} \le h_{roof}$

$$K_d = \begin{cases} 18 & for\ \ h_{base} > h_{roof} \\ 18 \mp 15\left(\frac{\Delta h_{base}}{h_{roof}}\right) & for\ h_{base} \leq h_{roof} \end{cases}$$

$$K_f = \begin{cases} -4 + 0.7\left(\frac{f}{925} - 1\right) & for\ suburban\ or\ medium - sized\ cities \\ -4 + 1.5\left(\frac{f}{925} - 1\right) & for\ urban\ cities \end{cases}$$

where f : frequency in GHz

B : building to building distance

d : T_x- R_x separation distance

w : street width

ϕ : street orientation angle with respect to direct radio path

2.3.3.10 Ericsson Model

This model also stands on the modified Okumura - Hata model to allow room for changing in parameters according to the propagation environment. Path loss according to this model is given by (Milanovic et al., 2007):

$$PL(dB) = a_0 + a_1 \log(d) + a_2 \log(h_b) + [a_3 \log(h_b) \log(d)] - 3.2[\log(11.75h_r)^2] + g(f) \quad (2.49)$$

Where $g(f) = 44.49 \log(f) - 4.78[\log(f)]^2$

f = frequency in MHz

Table 2.2. Values of Parameters for Ericsson Model (Milanovic et al., 2007)

Environment	**a_0**	**a_1**	**a_2**	**a_3**
Urban	36.2	30.2	12.0	0.1
Suburban	43.20	68.93	12.0	0.1
Rural	45.95	100.6	12.0	0.1

2.3.3.11 Lee Path Loss Model

Lee path loss model has been widely used in the prediction of path loss in macro cell applications, particularly for systems operating near 900 MHz and for ranges greater than 1.6 km (Vinko Erceg et al., 1999). The Lee model specifies distinct parameters for varying region types. Lee model should not be expected to be accurate outside a relatively narrow range of frequencies near 900 MHz.

The Lee model is given by Equation (2.50) (Lee, 1985):

$$PL\ (dB) = L_0 + \ \delta \log(d) - 10 \log(F_A) \tag{2.50}$$

Where PL = the median path loss with unit in decibel (dB)

L_0 = the reference path loss along 1km; unit in dB

δ = the slope of the path loss curve; unit in dB

d = the distance on which the path loss is to be calculated; unit in metres (m)

F_A = the adjustment factor.

In a given location, the L_0 and δ parameters should be determined empirically through a set of measurements.

2.4 Model Calibration Techniques

Model calibration is a process in which an existing propagation prediction model is tuned (or adjusted) to accurately fit the actual measurement data collected on the network sites. This process is targeted at improving the degree of accuracy of the model so as to truly represent the real physical environment of the propagation scenarios (Lempiäinen & Manninen, 2003). A number of approaches to achieving this have over the years been developed.

2.4.1 Statistical Model Calibration Method

This method is based on the modification of the coefficients of the propagation prediction models. These coefficients are the parameters that have been found through statistical analysis to be responsible for the relationship of the models with the quantity to be predicted (Ghassemzadeh, Jana, Rice, Turin, & Tarokh, 2002; Mao, Anderson, & Fidan, 2007). In this approach, an

appropriate empirical propagation model is selected from the existing radio propagation models. Thereafter, the selected model is used to predict the received signal strengths of the given cellular mobile network. The actual field measurement data are collected on a live network through drive tests. Subsequently, the collected data is dutifully analyzed to filter out error-related measurement points that may affect the accuracy of the calibration process. This iteration is repeated several times to minimize the error between the predicted results obtained and the field measured data (Lempiäinen & Manninen, 2003).

The statistical method implicitly takes all the environmental factors into account regardless of whether they can be separately recognized. By direct implication, the accuracy of the approach depends not only on the accuracy of the collection procedures of the field measurement data, but also on the similarities between the environment under study and the environment where the measurements were taken. Nevertheless, the statistical calibration method is widely used because of its simplicity, flexibility and computational efficiency. It achieves reasonably good prediction results when the propagation environment under study is fairly homogenous and similar to the environment where the measurements were taken.

With the recent development of the automated received signal measurement devices and systems with Global Positioning System (GPS) logging, it has become relatively easy to record large amounts of measurement data.

2.4.2 Deterministic Model Calibration Method

This approach is based on the physical laws governing the interaction of EM waves with the physical elements of the propagation environment. These laws are purely derivatives of the Maxwell equations. Here, a suitable propagation model is selected. This requires detailed data describing the area morphology, topology, street orientation etc. The data are used in the tuning process such that the resulting model represents the real environment and in turn adjust the model coefficients. The iteration is repeated so as to minimize the errors in the model approximating the real environment (Rappaport & Sandhu, 1994). The modelling is simplified by the use of path loss approximation models for diffraction, reflection, refraction, and absorption.

For the deterministic model to be effective there is a need for a detailed description of the objects present in the propagation environment. Thus, the limitation of this method of calibration is that it

requires extensive and detailed information of the terrain and physical objects present in the environment under study. Consequently, it requires significant computational resources in order to account for all of such information for the propagation prediction calculations to be performed. The uniqueness of the approach is that it predicts the received signal strength at a precise point in space by taking into consideration the specific propagation environment's peculiarities and features involved. Thus, the method is considered to be site-specific.

2.4.3 Semi-Deterministic Model Calibration Method

This method requires more information about the propagation environment than the statistical method but less than that of the deterministic. It provides a tradeoff between the two earlier discussed approaches. Its inclusion of deterministic correction factors improves the accuracy of the statistical models. This method facilitates the statistical propagation model to simulate the real propagation environment as close as possible. Consequently, the resulting models are easy to implement but offer less efficiency when compared to statistical models. Nevertheless, they provide the optimal choice of tradeoffs between accuracy and ease of use of the resulting propagation models (Subrt & Pechac, 2011).

2.4.4 Site-Specific Model Calibration Method

This method is targeted at improving the reliability of radio network planning and also to increase the degree of accuracy of the existing propagation prediction models for better planning and maintenance of radio communication network. However, the approach is built around the earlier discussed methodologies (EC Mukubwa, Karien, & Chatelain, 2006). The concept of this method is to develop a way of tuning propagation models for a single cell as opposed to the current cluster tuning approach. The approach is appreciated by radio network operators in cases where there are problematic cell sites within the cluster. The prediction method can only be used for cells within a similar environment (EM Mukubwa, 2006).

Once the suitable propagation prediction model is selected, it must be calibrated to model a specific cell site. The calibration process involves modifying the model parameters to accurately approximate relevant measurement data. The propagation measurement data is used to calibrate these propagation models. These measurement data is obtained through actual field measurements taken at various locations throughout the cell. Precise measurement can be determined using a

GPS. Typically, a large amount of the measurement may be required to accurately calibrate the model. Once the data is collected, it is converted to a suitable format and used to characterize the cell site to its location (Lempiäinen & Manninen, 2003). The calibration process uses the field data collected to define parameters, variable coefficients and constants of equations used in modelling the cell coverage.

2.5. Review of Related Works

Machine learning offers useful statistical tools for complex nonlinear regression problems and thus can be exploited to obtain more accurate path loss predictions at low computational cost. Feed-forward neural networks have been widely applied to estimate path loss under diverse circumstances.

A. Bhuvaneshwari, Hemalatha, and Satyasavithri (2017) identified COST 231 as the most appropriate for path loss predictions at GSM 900 MHz in dense urban area of Hyderabad city. The authors reduced the prediction errors by implementing the empirical model using dynamic neural networks. The Layer Recurrent Neural Network (LRNN) that was trained based on Levenberg-Marquardt (LM) learning algorithm was found to produce the best results but the model training takes longer time. T. A. Benmus, Abboud, and Shatter (2016) developed a neural network model based on measurement collected at 900 MHz, 1800 MHz, and 2100 MHz in the dense urban, urban, dense suburban, suburban, and rural environments of the capital of Libya Tripoli. The developed model had a better prediction accuracy when compared to Hata model. Zineb and Ayadi (2016) developed a novel model for indoor radio propagation for commonly available frequency bands (GSM, UMTS and Wi-Fi) using Multi-Layer Perceptron Neural Network (MLP-NN) and back-propagation training algorithm.

Delos Angeles and Dadios (2016) trained and validated feed-forward neural network with Longley-Rice model simulation results instead of using field measurement data. The developed neural network model outperformed both free space and Egli models. Zaarour, Affes, Kandil, and Hakem (2015) addressed the challenge of developing a path loss model at 60 GHz for a confined and harsh environment such as an underground mine using MLP and Radial Basis Function (RBF) neural networks. S. P. Sotiroudis and K. Siakavara (2015) developed ensemble ANNs that produced good prediction accuracy with optimal input data. S. P. Sotiroudis, S. K. Goudos, K. A. Gotsis, K. Siakavara, and J. N. Sahalos (2013) applied Differential Evolution (DE) algorithms to

ANN-based path loss modeling. Fernández Anitzine et al. (2012) investigated different methods of selecting the training dataset that is just sufficient for accurate path loss predictions. Different neural networks were developed for LOS and NLOS scenarios. The review of other related works is summarized in Table 2.3.

Table 2.3. Review of Related Works

S|N	Author(s)	Year	Location of Study	Scenario	Frequency Band(s)	Clutter Class(es)	Contribution(s)	Comment(s)
1	Angeles and Dadios (2015)	2015	People's Park, Tagaytay City, Philippines	Outdoor	600 MHz Digital Television (TV) band	Arbitrary	The authors claimed that the developed ANN model performed better than empirical models with the ability to adapt to arbitrary environment	Datasets were not real field measurements but simulation results of Longley-Rice model
2	Ayadi, Zineb, and Tabbane (2017)	2017	Tunis City, Tunisia	Outdoor	UHF bands (450 MHz, 850 MHz, 1800 MHz, 2100 MHz, 2600 MHz)	Urban, suburban, and rural	A new model was developed, instead of calibrating an existing one, to achieve a better precision.	The performance of the developed model was only compared with standard propagation model and ITU-R P.1812-4 model
3	Tammam A Benmus,	2015	Great Tripoli, Libya	Outdoor	900 MHz, 1800 MHz,	Dense urban, urban, dense suburban,	A model that is suitable for Tripolis, Lybia was developed.	The accuracy of the model was

	Abboud, and Shatter (2015)				and 2100 MHz	suburban, and rural		only compared to that of Hata model
4	A Bhuvaneshwari, Hemalatha, and Satyasavithri (2016)	2016	Hyderabad City, India	Outdoor	GSM 900 frequency band	Dense urban	COST 231 Hata model was implemented using dynamic neural networks (Focused Time Delay Neural Network, Distributed Time Delay Neural Network, and Layer Recurrent Neural Network)	The accuracy of the model was at the cost of increased computation time.
5	Cerri, Cinalli, Michetti, and Russo (2004)	2014	Ancona, Italy	Indoor	900 MHz	Urban	The authors developed ANN model to account for the effect of buildings on path loss.	The network training samples were obtained by Ray Tracing simulation.
6	Cruz and Caluyo (2014)	2014	Philippines	Indoor	677 MHz	Urban	The authors developed a heuristic path loss model to estimate indoor losses caused by signal	The model was not validated by other heuristic techniques

							penetration inside residences	
7	Eichie, Oyedum, Ajewole, and Aibinu (2017b)	2017	Minna, Niger State, Nigeria	Outdoor	1800 MHz	Rural and suburban	ANN models were developed for rural and suburban routes in Minna and its environs.	Comparisons were only made with basic empirical models but not with other machine learning based approach.
8	Fernández Anitzine et al. (2012)	2012	Spain	Indoor, Outdoor	900 and 1800 MHz	Urban	The authors developed ANN model with optimum selection of training set.	Although the multipath effect was eliminated, the measurements contain slow channel variations.
9	Ferreira, Matos, and Silva (2016)	2016	Rio de Janeiro, Brazil	Outdoor	1140 MHz	Urban	Hybrid application of ITU-R model and ANN model for efficient prediction.	More measurements are needed from different types of regions.

10	Gschwendtner and Landstorfer (1996)	1996	Mannheim, Germany	Outdoor	N/A	Urban	A hybrid modeling approach was introduced using environmental parameters.	The results of the study was only compared with COST Walfisch-Ikegami model
11	Kalakh, Kandil, and Hakem (2012)	2012	N/A	Undergro und tunnel	3-10 GHz	Mine environment	Developed path loss model for Ultra-Wideband propagation channel	Neural network is a feasible solution for channel modeling in environments that are considered to be harsh.
12	McLeod, Bai, and Meyer	2010	United States of America	Outdoor	2.4 GHz	Rocky/mount ainous area	Additional performance was gained by implementing a trained model in a parallel manner with a Graphical Processing Unit.	The network is tasked to generalize solutions across all possible point tuples.

13	Ostlin et al. (2010)	2010	Australia	Outdoor	IS-95 CDMA band	Rural	Evaluated the prediction accuracy, time, and generalization ability of simple neuron model and feed-forward networks with different number of hidden layers and neurons.	The performances of the model was only compared with ITU-R P.1546 and Okumura-Hata model.
14	Ileana Popescu, Kanatas, Constantinou, and Naforniţă (2002)	2002	Kavala, Greece	Outdoor	1890 MHz	Urban	A general regression neural network model was designed and implemented.	Comparison was limited to empirical models.
15	I Popescu, Nafomita, Constantinou, Kanatas, and Moraitis (2001)	2001	Kavala, Greece	Outdoor	1890 MHz	Urban	Designed neural network models for LOS and NLOS cases	Comparison was limited to empirical models.

16	Ileana Popescu et al. (2005)	2005	Oai Village, Santorini Island, Greece	Outdoor	1890 MHz	Rural	Investigated error correction model.	Hybrid model proved to be highly adaptive
17	S. Sotiroudis, S. Goudos, K. Gotsis, K. Siakavara, and J. Sahalos (2013)	2013	Thessaloniki, Greece	Outdoor	N/A	Urban	Applied the combination of DE and LM algorithms.	The combination achieved produced convergence. The model was trained with Ray Tracing simulation results.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1. Introduction

The research methodology adopted in this study is the Design Science Research (DSR) approach. An overview of the research methodology is depicted using the flowchart shown in Figure 3.1. In this chapter, the physical environment of the study area is described as a typical wireless propagation channel of heterogeneous radio networks in a smart university campus. Radio signal measurements are conducted to obtain the strengths of radio signals received at varying separation distances between fourteen base station transmitters and two mobile receivers. Geographic and network information (i.e. longitude, latitude, elevation, altitude, frequency, clutter height, and RSS) recorded are stored in a local database. These data are further processed to remove duplicates and extraneous data points. Machine learning-based path loss models are developed using ANN and SVM techniques. Finally, prediction accuracy and generalization ability of popular empirical models (Hata, COST 231, ECC-33, and Egli), ANN-based models, and SVM models are evaluated using MAE, MSE, RMSE, SED, R, ANOVA, and multiple comparison post-hoc test.

3.2. Description of the Smart Campus Propagation Environment

Extensive field measurement campaign was conducted within the campus of Covenant University using drive test approach. The university campus is located along kilometer 10, Idi-Iroko road, Ota, Canaanland, Ogun State, Nigeria (Latitude 6°40'30.3"N, Longitude 3°09'46.3"E). The physical environment of Covenant University campus is made up of ultra-modern lecture halls, high-rise administrative buildings, four-storey residential buildings, high-rise research and innovation hub, a standard stadium complex, a modern health center facility, trees planted along roadsides, flowers, parks and gardens, street lights, signposts, and billboards. A satellite imagery of the physical environment is shown in Figure 3.2. Most of these physical structures have considerable heights such that they obstruct Line of Sight LOS but produce NLOS signal paths for wireless communications at radio frequencies.

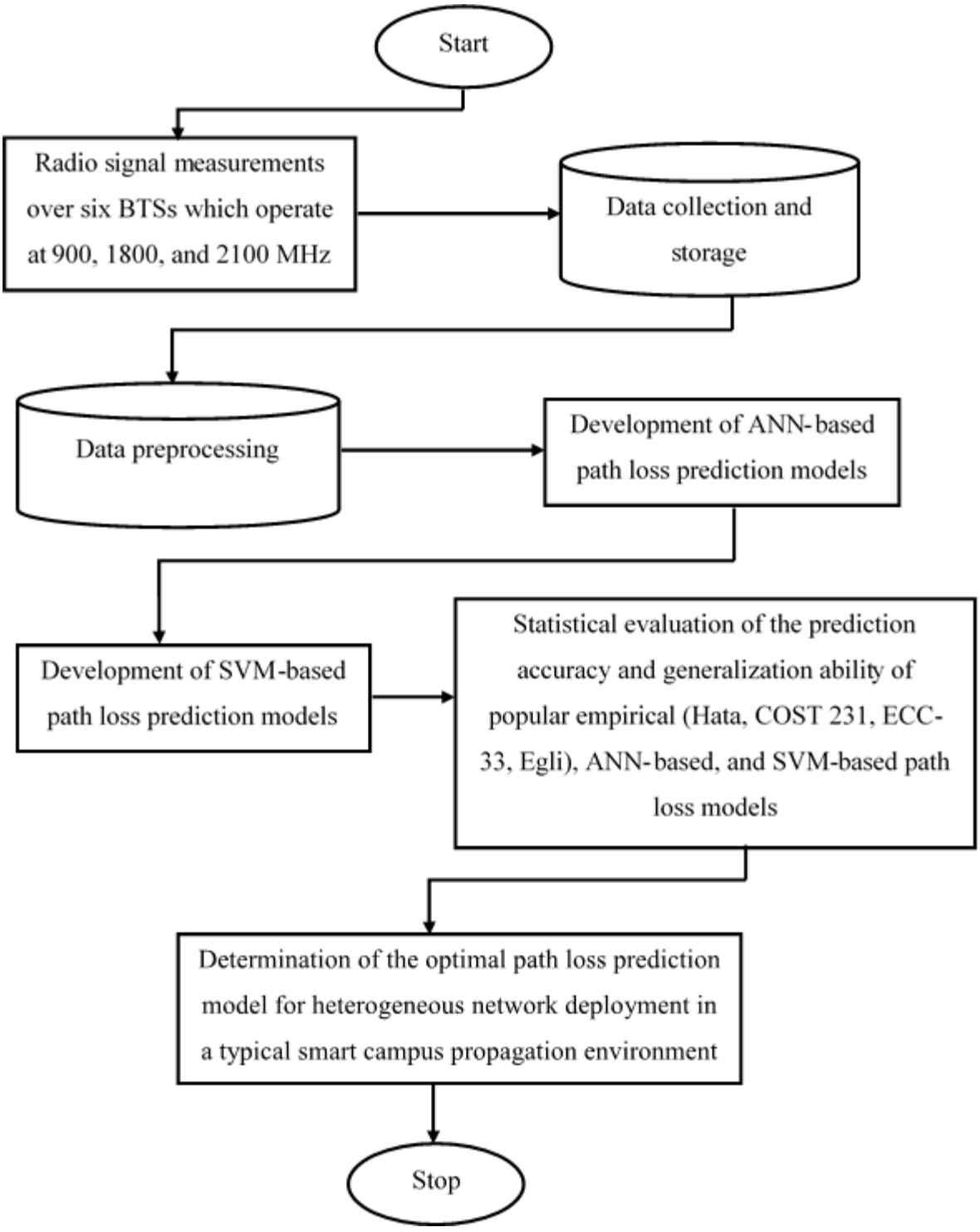

Figure 3.1. Flowchart of Research Methodology

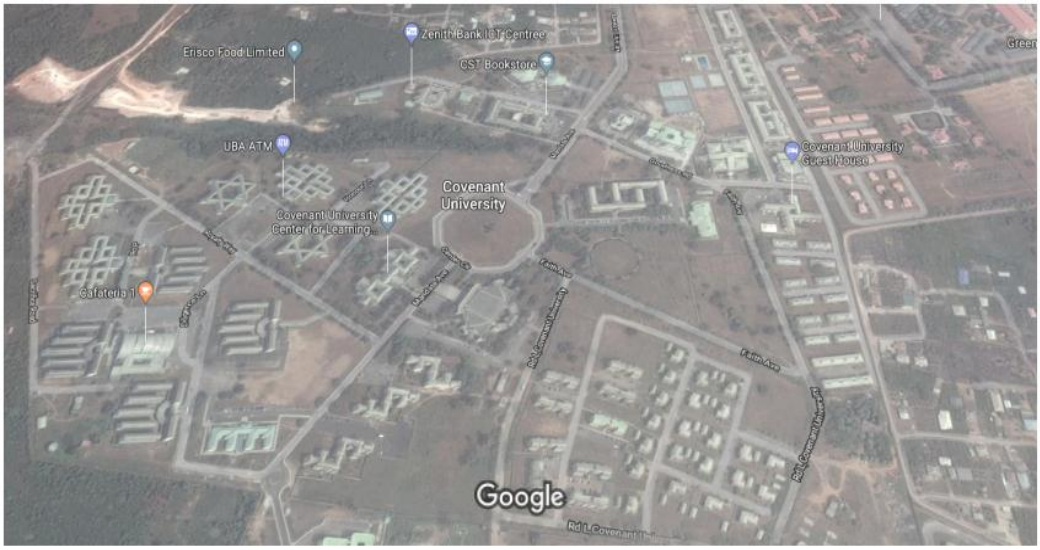

Figure 3.2. Physical environment of Covenant University campus

3.3. Radio Signal Measurement Campaign

A drive test experimental setup was designed for the field measurement campaign. The equipment, devices, and tools that constitute the experimental setup include: six commercial transceivers with fourteen (14) directional antennas, two mobile receivers, a GPS receiver, a radio signal measurement software that runs on a Personal Computer (PC), and a motor vehicle. Ericsson RBS 2216, Ericsson RBS 2116, and Ericsson RBS 6201 base station transceivers were used for radio signal transmission at 900, 1800, and 2100 MHz respectively. Sectorial antennas of 13 dBi gain, 120 degree horizontally polarized sector panel were used to radiate electromagnetic signals which emanate from Ericsson RBS 2216 transmitters. 18 dBi gain, 65 degree vertically polarized antennas were used for radio wave transmission at 1710-1880 MHz frequency range. 17 dBi gain, 90 degree vertically polarized antennas were utilized for radio propagation at 2090-2290 frequency range. Two Sony Ericsson w995 mobile phones, with processing speed of 369 MHz and a removable Li-Po 930 mAh battery each, were used for radio signal reception at 900, 1800, and 2100 MHz. A Universal Serial Board (USB) magnet mount GPS receiver, BU-353-S4, was used to track mobile receiver's location at a given time. A 64-bit Windows Operating System (OS), 4 GB Random Access Memory (RAM) laptop with Intel® Core™ i5, M520 @2.40 GHz central processing capacity was used for data logging and storage.

Radio signal measurements were conducted along 14 drive test survey routes within Covenant University campus in order to adequately represent the wireless channel characteristics of a typical smart campus propagation environment. Information about the geographic location and the altitude of the radio transmitters are provided in Table 3.1. The drive test survey routes were carefully planned using the terrain map of the study area shown in Figure 3.3. Strength of radio signals received (RSS) from respective transmitters were measured, recorded, and stored as the mobile receivers are driven along the each survey route using TEMS™ Investigation software developed by InfoVista®. The amount of radio signal power transmitted by each of the transmitters was 43 dB and the selected mobile receiver has a minimum sensitivity of -100 dBm.

The empirical measurements covered six (6) commercial transceivers with fourteen (14) directional antennas namely: A2GS1, A2GS2, A2GS3, A3GS1, AW3GS2, A3GS3, E2GS1, E2GS3, E3GS1, E3GS3, M2GS1, M2GS3, M3GS1, and M3GS3. Radio signal transmission and reception were performed at 900, 1800, and 2100 MHz operating frequencies, as expected of GSM, DCS, and UMTS wireless systems respectively, in the directions of the base station antennas. Continuous measurement of longitude, latitude, elevation, altitude, frequency, clutter height, and RSS were recorded.

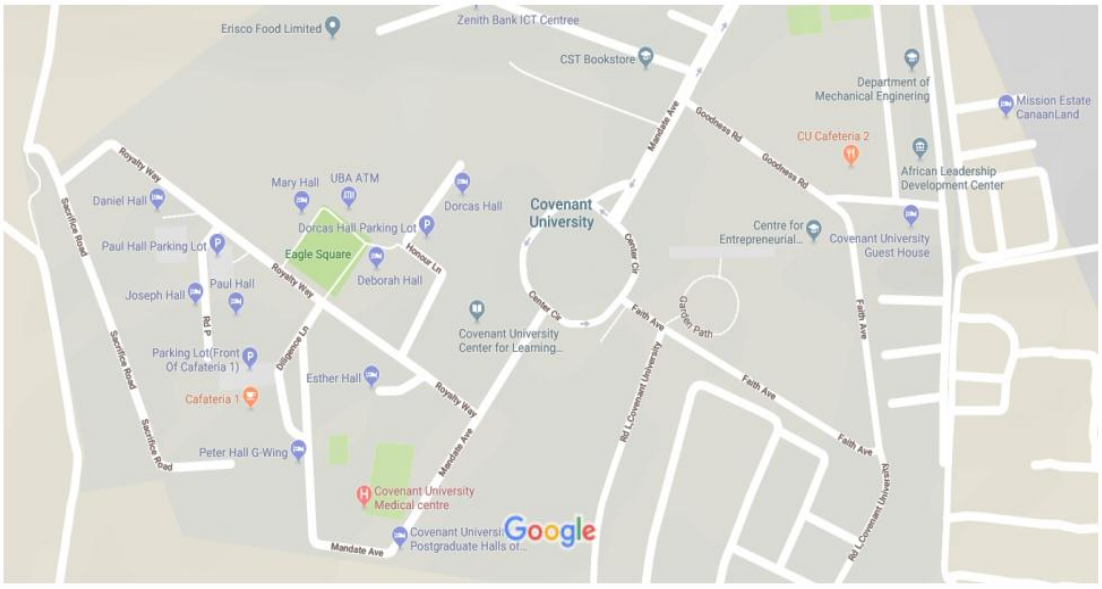

Figure 3.3. Drive test survey routes

Table 3.1. Geographic locations of base station transmitters

S|N	BTS ID	Longitude	Latitude	Altitude (m)
1	A2GS1	3.162867	6.675068	50
2	A2GS2	3.162867	6.675068	50
3	A2GS3	3.162867	6.675068	50
4	A3GS1	3.162867	6.675068	50
5	A3GS2	3.162867	6.675068	50
6	A3GS3	3.162867	6.675068	50
7	E2GS1	3.164015	6.675253	52
8	E2GS3	3.164015	6.675253	52
9	E3GS1	3.164015	6.675253	52
10	E3GS3	3.164015	6.675253	52
11	M2GS1	3.16393	6.675245	52
12	M2GS3	3.16393	6.675245	52
13	M3GS1	3.16393	6.675245	52
14	M3GS3	3.16393	6.675245	52

3.4. Data Preprocessing

Data collected through drive test (i.e. longitude, latitude, elevation, frequency, and RSS) were exported from TEMS™ Investigation software developed by InfoVista® into a spreadsheet file format. Mapping and location analysis of RSS data collected at 900, 1800, 2100 MHz radio frequencies within Covenant University campus was performed using a desktop Geographic Information System (GIS) software, MapInfo Pro™, produced by Pitney Bowes. Appropriate data filtering and sorting were performed using Microsoft Excel 2013 to remove data instance

duplicates. In this research project, data instance is considered to be duplicated if another data instance with the same values of longitude, latitude, elevation, altitude, frequency, clutter height, distance have been previously captured in the dataset. The whole experimental field measurement process was accurately represented in ATOLL™ v3.1 radio network planning software produced by Forsk®. Altitude and clutter height data for respective longitude and latitude points recorded were obtained from the Digital Terrain Map (DTM) of the propagation environment shown in Figure 3.4. Separation distances between the base station transmitters and the mobile receivers were computed for all the data instances using ATOLL software. Corresponding path loss values were calculated for each data instance of RSS using Equation (3.1):

$$PL_{measured,i}(dB) = 43\ dB - RSS_i \tag{3.1}$$

The complete filtered and sorted data with nine variables (i.e. longitude, latitude, elevation, altitude, frequency, clutter height, distance, RSS, and path loss) were randomly divided into two parts: 75% training dataset and 25% testing dataset for path loss model development and evaluation.

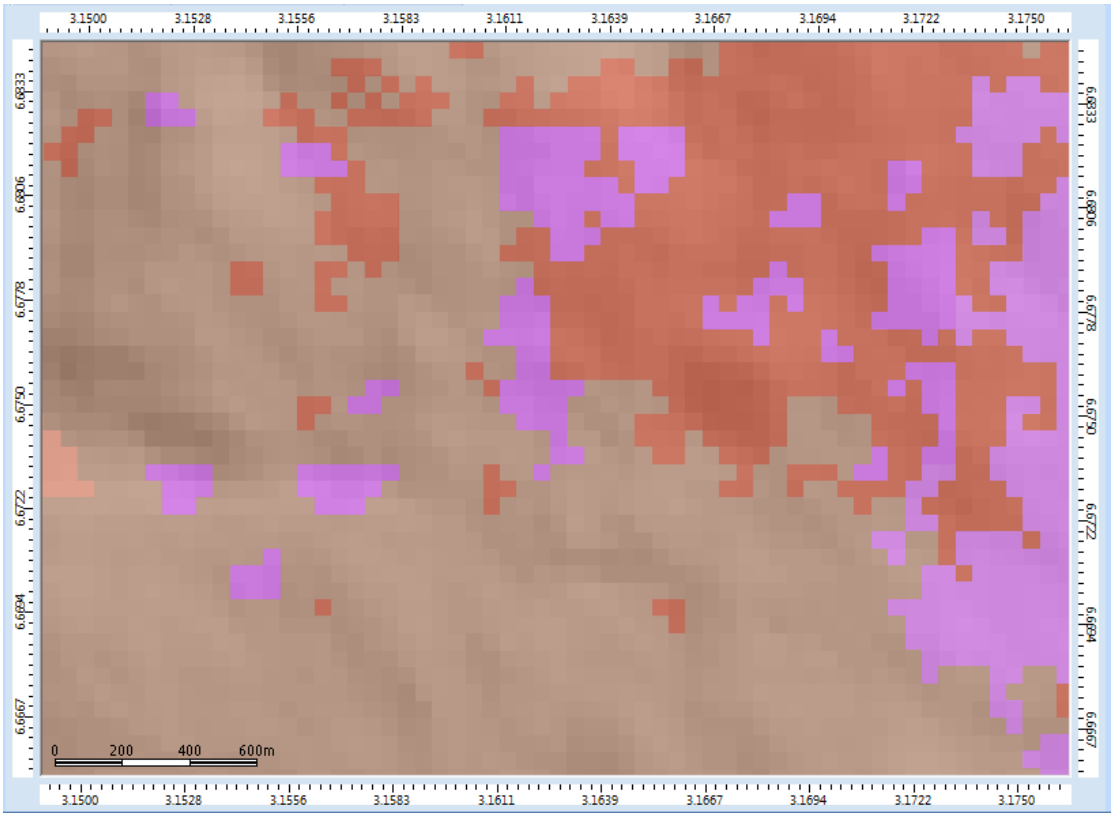

Figure 3.4. Digital Terrain Map of the propagation environment

3.5. Development of ANN-Based Path Loss Prediction Models

An optimal ANN-based path loss model is expected to produce high prediction accuracy and demonstrate good generalization ability. In order to ensure that these performance criteria are satisfied, series of experimentations were performed to determine the most appropriate neural network architecture for radio propagation modeling in a heterogeneous smart campus environment. Model development, validation, testing, and evaluation were performed using the Neural Network Toolbox in MATLAB 2016a produced by MathWorks Inc. First, the minimum input variable(s) required for ANN-based path loss prediction model development was established. Then, effect of input data normalization on prediction accuracy and generalization ability was investigated. Furthermore, the most suitable activation function for ANN-based path loss model development was determined. Also, the most appropriate neural network training algorithm was identified. Finally, the number of neuron required at the hidden layer was experimentally determined.

In order to determine the minimum input variable(s) that guarantee(s) sufficiently high prediction accuracy and good generalization ability, six scenarios were investigated through extensive experimentations. At first, ANN models were developed based on single input requirement (i.e. distance). Consequently, the number of input variable was systematically increased to include frequency, clutter height, elevation, altitude, latitude and longitude as presented in Algorithm 3.1. The output variable of the ANN-based path loss models is the path loss vector. The procedure highlighted in Algorithm 3.1 was repeated 10 times for each of the six use cases under investigation. Therefore, a total of 60 ANN-based path loss models were trained and tested to assess the set objective using training and testing datasets respectively.

Response of ANN-based path loss models to input data normalization was investigated by implementing the algorithm presented in Algorithm 3.2. In this case, seven input variables (i.e. distance, frequency, clutter height, elevation, altitude, latitude, longitude) were used for model development. All data instances of the input training and testing matrices were normalized (-1, 1) using min-max method. The mathematical formulation is given by Equation 3.2:

$$A_i = -0.9 + \left[\frac{1.8(K_i - K_{\min})}{K_{max} - K_{min}}\right] \tag{3.2}$$

Algorithm 3.1. Development of ANN Models with Varying Input Parameters

```
0  start
1      switch training input variable data, testing input variable data
2        case a: load distance; caseID =1;
3        case b: load distance, frequency; caseID =2;
4        case c: load distance, frequency, clutter height; caseID =3;
5        case d: load distance, frequency, clutter height, elevation; caseID =4;
6        case e: load distance, frequency, clutter height, elevation, altitude; caseID = 5;
7        case f: load distance, frequency, clutter height, elevation, altitude, latitude,
                    longitude; caseID =6;
8      end switch
9      REPEAT
10     load training and testing output variable (path loss) data; caseID =1;
11      for number of iteration = 1 to 10
12         configure a single hidden layer neural network model architecture
13           set number of hidden neuron to 10
14           set activation function at the hidden layer to hyperbolic tangent sigmoid
15           set activation function at the output layer to hyperbolic tangent sigmoid
16           set neural network training rule to Levenberg Marquardt
17           divide training dataset into 70% training, 15% validation, and 15% testing
18         train neural network model for path loss predictions
19           set learning rate to 0.1 and number of epochs to 1000
20           set maximum validation failures to 1000
21           set performance goal to zero
22         evaluate model prediction accuracy
23           compute training time, MAE, MSE, RMSE, SED, R
24         evaluate model generalization ability
25           compute MAE1, MSE1, RMSE1, SED1, R1; caseID =caseID + 1;
26     end for caseID; UNTIL caseID > 6;
27 stop
```

Algorithm 3.2. Effect of Input Data Normalization on ANN Model Performance

```
0  start
1       load training input variable data, testing input variable data
2            distance, frequency, clutter height, elevation, altitude, latitude, longitude
3       load training and testing output variable (path loss) data
4       switch raw input data, normalized input data
5            case a: load raw training input data, raw testing input data; caseID =1;
6            case b: normalize training input data, testing input data; caseID = 2;
7       end switch; caseID = 1;
8       REPEAT
9         for number of iteration = 1 to 10
10          configure a single hidden layer neural network model architecture
11            set number of hidden neuron to 10
12            set activation function at the hidden layer to hyperbolic tangent sigmoid
13            set activation function at the output layer to hyperbolic tangent sigmoid
14            set neural network training rule to Levenberg Marquardt
15            divide training dataset into 70% training, 15% validation, and 15% testing
16          train neural network model for path loss predictions
17            set learning rate to 0.1
18            set maximum number of epochs to 1000
19            set maximum validation failures to 1000
20            set performance goal to zero
21          evaluate model prediction accuracy
22            compute training time, MAE, MSE, RMSE, SED, R
23          evaluate model generalization ability
24            compute MAE1, MSE1, RMSE1, SED1, R1
25      end for; caseID = caseID + 1;
26      UNTIL caseID > 2;
27 stop
```

Where

A_i = normalized input matrix

K_{min} = column vector of minimum values in raw input matrix

K_{max} = column vector of maximum values in raw input matrix

At the end of the experimentations, a total of 20 ANN-based path loss models were developed: 10 of the models were developed based on raw input matrix while the remaining 10 models were developed based on normalized input matrix.

In the process of training neural network models, either of these three activation/transfer functions are most commonly employed at the hidden and output layers: linear (*purelin*), logarithmic sigmoid (*logsig*), and hyperbolic tangent sigmoid (*tansig*). The mathematical expressions for *logsig* and *tansig* activation functions are given by Equation 3.3 and Equation 3.4 respectively. Combination of these three transfer functions in pair resulted in nine use cases. The nine scenarios were experimented based on the algorithm presented in Algorithm 3.3. Ten iterations were performed in each case, and a total of 90 ANN-based path loss models were realized.

Table 3.2. Training algorithm for ANN model development

Acronymn	**Algorithm**	**Description**
LM	*trainlm*	Levenberg-Marquardt
BFG	*trainbfg*	BFGS Quasi Newton
RP	*trainrp*	Resilient Backpropagation
SCG	*transcg*	Scaled Conjugate Gradient
CGB	*traincgb*	Conjugate Gradient with Powell/Beale Restarts
CGF	*traincgf*	Fletcher-Powell Conjugate Gradient
CGP	*traincgp*	Polak-Ribiere Conjugate Gradient
OSS	*trainoss*	One Step Secant
GDX	*traingdx*	Variable Learning Rate Backpropagation

Algorithm 3.3. Development of ANN Models with Varying Activation Functions

```
0  start
1      load training input variable data, testing input variable data
2          distance, frequency, clutter height, elevation, altitude, latitude, longitude
3      load training and testing output variable (path loss) data
4      for number of iteration = 1 to 10
5          configure a single hidden layer neural network model architecture
6              set number of hidden neuron to 10
7              switch hidden layer and output layer activation functions
8                  case a: purelin, purelin; caseID = 1;
9                  case b: logsig, purelin; caseID = 2;
10                 case c: tansig, purelinI; caseID = 3;
11                 case d: purelin, logsig; caseID = 4;
12                 case e: logsig, logsig; caseID = 5;
13                 case f: tansig, logsig; caseID = 6;
14                 case g: purelin, tansig; caseID = 7;
15                 case h: logsig, tansig; caseID = 8;
16                 case i: tansig, tansig; caseID = 9;
17             end switch; caseID = caseID + 1; REPEAT
18             set neural network training rule to Levenberg Marquardt
19             divide training dataset into 70% training, 15% validation, and 15% testing
20         train neural network model for path loss predictions
21             set learning rate to 0.1, epochs to 1000, and validation failures to 1000
22         evaluate model prediction accuracy
23             compute training time, MAE, MSE, RMSE, SED, R
24         evaluate model generalization ability
25             compute MAE1, MSE1, RMSE1, SED1, R1
26     end for
27     caseID =caseID + 1; UNTIL caseID > 9;
28 stop
```

Algorithm 3.4. Development of ANN Models with Varying Learning Rules

```
0  start
1      load training input variable data, testing input variable data
2          distance, frequency, clutter height, elevation, altitude, latitude, longitude
3      load training and testing output variable (path loss) data
4      for number of iteration = 1 to 10
5          configure a single hidden layer neural network model architecture
6              set number of hidden neuron to 10
7              set activation function at the hidden layer to tansig
8              set activation function at the output layer to tansig
9              switch learning algorithm
10                 case a: trainlm
11                 case b: trainbfg
12                 case c: trainrp
13                 case d: trainscg
14                 case e: traincgb
15                 case f: traincgf
16                 case g: traincgp
17                 case h: trainoss
18                 case i: traingdx
19         end switch
20         divide training dataset into 70% training, 15% validation, and 15% testing
21         train neural network model for path loss predictions
22         set learning rate to 0.1, epochs to 1000, and validation failures to 1000
23         evaluate model prediction accuracy
24             compute training time, MAE, MSE, RMSE, SED, R
25         evaluate model generalization ability
26             compute MAE1, MSE1, RMSE1, SED1, R1
27     end for
28 stop
```

$$logsig(K) = \frac{1}{1+ e^{-K}} \tag{3.3}$$

$$tansig(K) = \frac{1}{(1+ e^{-2K})-1} \tag{3.4}$$

ANN models may be trained based on any of the learning rules presented in Table 3.2. 90 ANN-based path loss models were trained based on Levenberg-Marquardt, BFGS Quasi Newton, Resilient Backpropagation, Scaled Conjugate Gradient, Conjugate Gradient with Powell/Beale Restarts, Fletcher-Powell Conjugate Gradient, Polak-Ribiere Conjugate Gradient, One Step Secant, and Variable Learning Rate Backpropagation learning algorithms. The procedure taken for the experimentation is explained in Algorithm 3.4.

The choice of the number of neurons used at the hidden layer is a very important factor in avoiding poor model generalization (if number of hidden neuron is insufficient) and model complexity (if hidden neurons used are too many). The most suitable number of hidden neuron, which guarantees good model generalization as well as model simplicity, was determined through repeated experimentations. The number of hidden neuron was varied between 1 and 50 with an incremental step size of 1. The experimental process was carried out by implementing the algorithm described in Algorithm 3.5.

3.6. Development of SVM-Based Path Loss Prediction Models

Most influencing input variable attributes were selected using 10-fold validation approach. CFS Subset Evaluator and Greedy Stepwise methods were used to search and evaluate the influence of seven independent attributes (i.e. longitude, longitude, latitude, elevation, altitude, frequency, clutter height, and distance) on a dependent variable (path loss). These algorithms were implemented in a Java-based machine learning software, WEKA, produced at the University of Waikato, New Zealand. Furthermore, a SVM-based path loss model was developed using SMOreg regression algorithm based on the algorithm presented in Algorithm 3.6. Model parameters and kernel evaluations were obtained for path loss predictions in heterogeneous smart campus environment.

Algorithm 3.5. Development of ANN Models with Varying Number of Hidden Neuron

```
0  start
1      load training input variable data, testing input variable data
2          distance, frequency, clutter height, elevation, altitude, latitude, longitude
3      load training and testing output variable (path loss) data
4      for number of hidden neuron = 1 to 50
5          for number of iteration = 1 to 10
6              configure a single hidden layer neural network model architecture
7                  set number of hidden neuron to 10
8                  set activation function at the hidden layer to tansig
9                  set activation function at the output layer to tansig
10                 set neural network training rule to Levenberg Marquardt
11             divide training dataset into 70% training, 15% validation, and 15% testing
12             train neural network model for path loss predictions
13                 set learning rate to 0.1
14                 set maximum number of epochs to 1000
15                 set maximum validation failures to 1000
16                 set performance goal to zero
17             evaluate model prediction accuracy
18                 compute training time, MAE, MSE, RMSE, SED, R
19             evaluate model generalization ability
20                 compute MAE1, MSE1, RMSE1, SED1, R1
21         end for
22     end for
23 stop
```

Algorithm 3.6. Development of SVM Model with Polynomial Kernel Function

0 **start**

1 **load** training input vector (*p*), testing input vector ($\boldsymbol{p_t}$)

2 *distance, frequency, clutter height, elevation, altitude, latitude, longitude*

3 **load** training target vector (t) and testing target vector ($\boldsymbol{t_a}$)

4 **estimate** model coefficients *v*, *h*, *C*, σ

5 **compute** $\boldsymbol{regularization\ term} = \frac{1}{2}\|v^2\|$

6 **compute** $\boldsymbol{G_\sigma(t,z)} = \begin{cases} |\boldsymbol{d-y}| & -\boldsymbol{\sigma}|\boldsymbol{d-y}| \geq \boldsymbol{\sigma} \\ \boldsymbol{0} & \boldsymbol{otherwise} \end{cases}$

7 **compute** $\boldsymbol{empirical\ error} = \boldsymbol{Q} \times \frac{1}{m}\sum_{k=1}^{m} \boldsymbol{G}(\boldsymbol{t_k}, \boldsymbol{z_k})$

8 **minimize the regularized risk function**

9 $\boldsymbol{L_{SVM}(Q)} = \boldsymbol{empirical\ error} + \boldsymbol{regularization\ term}$

10 **add** positive slack variables $\boldsymbol{u_k}, \boldsymbol{u_k^*}$

11 **minimize** $\boldsymbol{L_{SVM}(v,u^*)} = \boldsymbol{empirical\ error} + (\boldsymbol{Q}\ \sum_{k=1}^{m}(\boldsymbol{u_k} + \boldsymbol{u_k^*})$

12 subjected to $\boldsymbol{t_k} - \boldsymbol{vf(p_k)} - \boldsymbol{h_k} \leq \boldsymbol{\sigma} + \boldsymbol{u_k}$

13 $\boldsymbol{vf(p_k)} + \boldsymbol{h_k} - \boldsymbol{t_k} \leq \boldsymbol{\sigma} + \boldsymbol{u_k^*}, \quad \boldsymbol{u_k^*} \geq \boldsymbol{0}$

14 **introduce Lagrange multipliers**

15 **compute** kernel function $\boldsymbol{R(p,p_k)} = \boldsymbol{f(p_k)^* f(p_i)}$

16 **compute** $\boldsymbol{\Phi}(\mathbf{p}, \boldsymbol{e_k}, \boldsymbol{e_k^*}) = \sum_{k=1}^{m}(\boldsymbol{e_k} - \boldsymbol{e_k^*})\,\boldsymbol{R(p,p_k)} + \boldsymbol{h}$

17 **define** kernel parameter *b*

18 **apply polynomial kernel function**

19 $\boldsymbol{K(p_k,p_i)} = (\boldsymbol{p_k}^* \boldsymbol{p_i} + \boldsymbol{1})^{\boldsymbol{b}}$

20 **evaluate model prediction accuracy**

21 **compute** *training time, MAE, MSE, RMSE, SED, R*

22 **evaluate model generalization ability**

23 **compute** *MAE, MSE, RMSE, SED, R*

24 **end for**

25 **end for**

26 **stop**

3.7. Statistical Evaluation of Empirical, ANN, and SVM Path Loss Models

Prediction accuracy and generalization ability of all the developed ANN-based path loss models and the SVM-based path loss models were evaluated based on their MAE, MSE, RMSE, SED, and R, with respect to the path loss values in both training data and testing data respectively. The values of MAE, MSE, RMSE, SED, and R were obtained by using Equations (3.5)-(3.9) respectively.

$$MAE\ (dB) = \frac{1}{k}\sum_{i=1}^{k}(Path\ Loss_{measured,i} - Path\ Loss_{predicted,i}) \tag{3.5}$$

$$MSE\ (dB) = \frac{1}{k}\sum_{i=1}^{k}(Path\ Loss_{measured,i} - Path\ Loss_{predicted,i})^2 \tag{3.6}$$

$$RMSE\ (dB) = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(Path\ Loss_{measured,i} - Path\ Loss_{predicted,i})^2} \tag{3.7}$$

$$Standard\ Deviation\ (dB) = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(\left|Path\ Loss_{measured,i} - Path\ Loss_{predicted,i}\right| - mean\ error)^2} \tag{3.8}$$

$$R = \frac{\sum_{i=1}^{k}(M_i - P_{mean})^2 - \sum_{i=1}^{k}(P_i - M_i)^2}{\sum_{i=1}^{k}(M_i - P_{mean})^2} \tag{3.9}$$

Empirical models (Hata, COST 231, ECC-33, and Egli) which are commonly used were employed for path loss predictions based on the distance input vector provided in training and testing datasets. In addition, ANOVA and multiple comparison post-hoc tests were performed to understand whether the differences between the mean measured path loss, and the mean path loss predictions produced by Hata model, COST 231 model, ECC-33 model, and Egli model, optimal ANN-based model, and SVM-based model are statistically significant. ANOVA and multiple comparison post-hoc analyses were performed using statistical toolbox in MATLAB 2016a.

CHAPTER FOUR

RESULTS

4.1. Introduction

In this chapter, the results of this research project are presented and discussed. First, data obtained through drive test field measurement campaign are preprocessed and divided into training and testing datasets. Exploration of these data are presented in great details using various visualization techniques such as boxplots, tables, scatter plots, and frequency distribution histograms. Simple linear regression and correlation analyses are performed to understand the relationship between the nine attributes considered in this research project namely: longitude, latitude, elevation, altitude, frequency, clutter height, distance, RSS, and path loss. More importantly, results of the ANN-based model and SVM-based model are presented. Finally, the results of the statistical evaluation of the empirical, ANN, and SVM path loss models are discussed.

4.2. Results of Radio Signal Measurement Campaign and Data Preprocessing

Data collected through field measurement campaign are considered extensive and sufficient for model development, validation, and testing. A large dataset containing hundreds of thousand data instances of longitude, latitude, elevation, altitude, frequency, clutter height, distance, and RSS was obtained and analyzed. RSS data collected at 900, 1800, and 2100 MHz were depicted using MapInfo Professional software and shown in Figures 4.1-4.3 respectively. Information about the quantitative results obtained during data collection are presented in Table 4.1. A total of 123,985 raw data instances were logged with an average of 8,856 data instances per antenna. 105,120 data instances were found to be duplicates and they were removed. The remaining 18,865 unique data instances were curated for model development and evaluation. The mean number of unique data instances available along the survey routes of each of the fourteen sectors is 1,348.

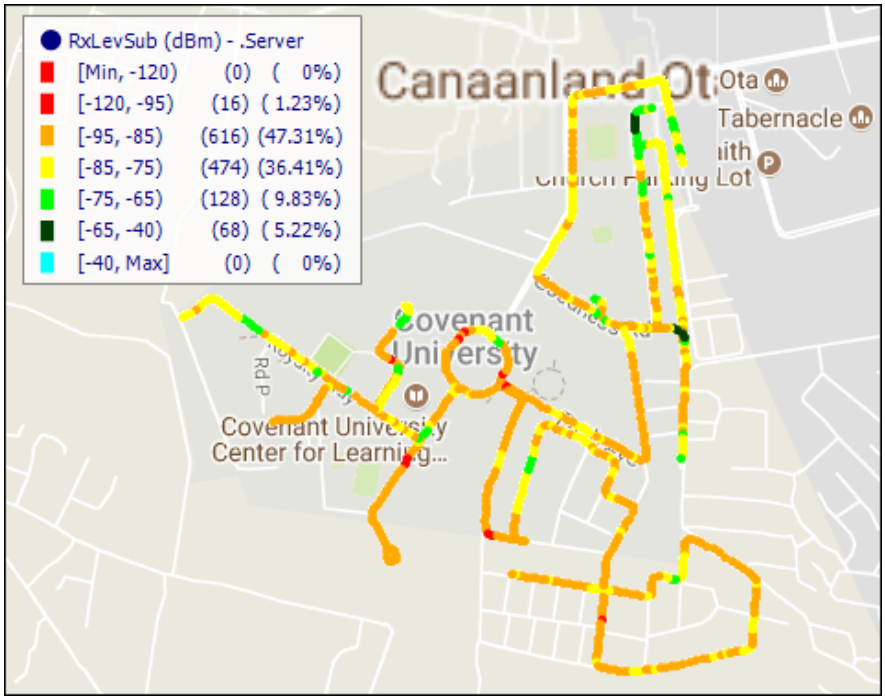

Figure 4.1. RSS data collected at 900 MHz

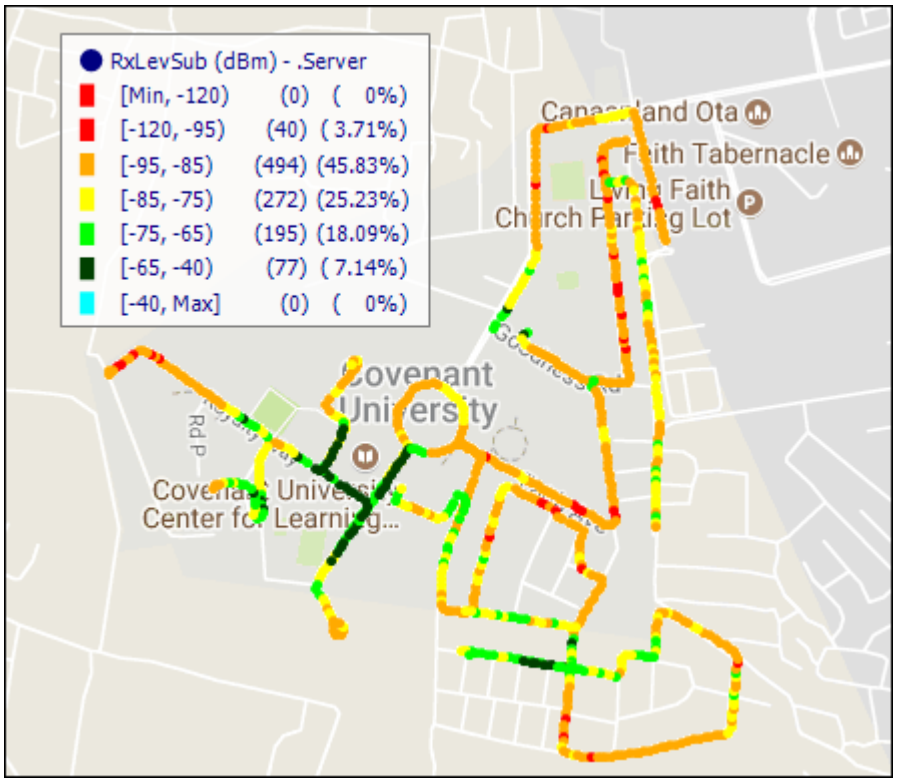

Figure 4.2. RSS data collected at 1800 MHz

Table 4.1. Quantitative summary of field measurement data

S|N	BTS ID	Raw data instances	Duplicate of data instances found	Unique data instances after filtering
1	A2GS1	2,284	1,626	658
2	A2GS2	3,918	3,168	750
3	A2GS3	4,838	3,388	1,450
4	A3GS1	5,551	4,632	919
5	A3GS2	8,139	7,414	725
6	A3GS3	11,555	9,687	1,868
7	E2GS1	11,028	9,067	1,961
8	E2GS3	6,591	4,837	1,754
9	E3GS1	24,371	22,274	2,097
10	E3GS3	18,319	15,828	2,491
11	M2GS1	4,228	3,439	789
12	M2GS3	6,597	5,052	1,545
13	M3GS1	4,123	3,734	389
14	M3GS3	12,443	10,974	1,469
Total		123,985	105,120	18,865

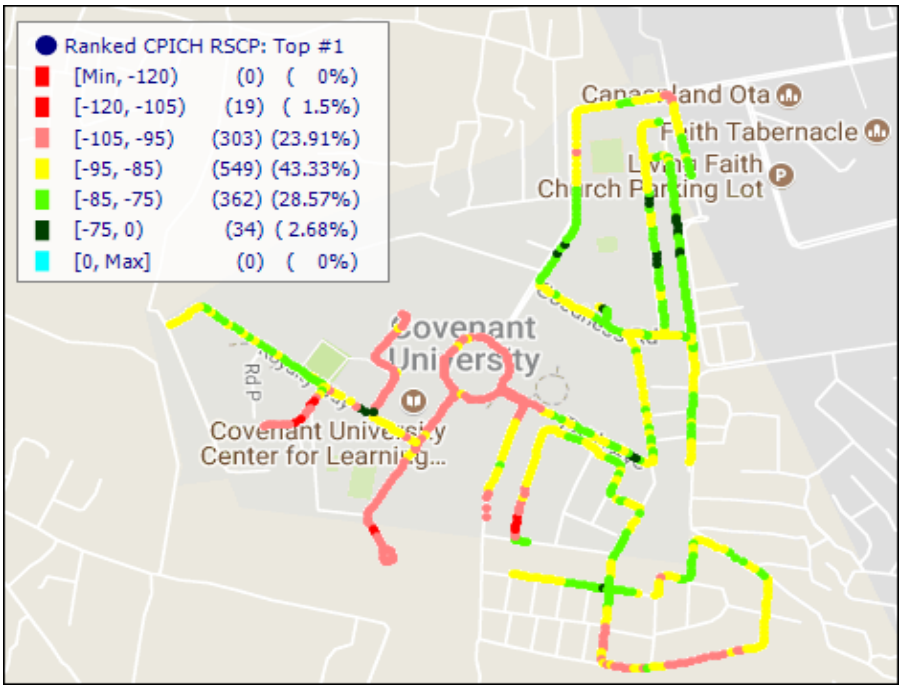

Figure 4.3. RSS data collected at 2100 MHz

4.3. Exploration of Field Measurement Data

The set of unique data instances was divided into two sub datasets (training dataset, and testing dataset). 75% of the complete dataset (i.e. 14,142 unique data instances) was used for model training while the remaining 25% (i.e. 4,714 unique data instances) was used for model evaluation and testing. The quantitative distribution of the training data and the testing data are described using their first-order descriptive statistics. Table 4.2 and Table 4.3 describe the distribution of longitude, latitude, elevation, altitude, frequency, clutter height, distance, RSS, and path loss measured in training data and testing data respectively, based on their mean, median, mode, SED, variance, kurtosis, skewness, range, minimum, and maximum. In order to understand the data distribution in each of the frequency bands considered in this research project, the univariate analysis was also performed on training data and testing data collected at 900, 1800, and 2100 MHz radio frequencies. The results of this analysis are presented in Tables 4.4-4.9. Total unique data instances of longitude, latitude, elevation, altitude, frequency, clutter height, distance, RSS, and path loss measured in training data obtained at 900, 1800, and 2100 MHz are 1,908, 5,058, and 7,176 respectively. On the other hand, 635, 1,687, and 2,392 unique data instances of longitude, latitude, elevation, altitude, frequency, clutter height, distance, RSS, and path loss were logged at 900, 1800, and 2100 MHz respectively

Table 4.2. Descriptive first-order statistics of training data

	Longitude	**Latitude**	**Elevation (m)**	**Altitude (m)**	**Frequency (MHz)**	**Clutter Height (m)**	**Distance (m)**	**RSS (dBm)**	**Path Loss (dB)**
Mean	3.16	6.67	51.32	56.56	1830.80	6.62	526.78	-80.32	122.32
Median	3.16	6.67	51.00	57.00	2100.00	6.00	450.00	-81.00	123
Mode	3.16	6.68	52.00	57.00	2100.00	6.00	380.00	-88.00	130
SED	0.00	0.00	4.85	5.43	392.45	3.72	327.79	11.32	11.32
Variance	0.00	0.00	23.52	29.54	154013.34	13.84	107444.11	128.04	128.04
Kurtosis	3.42	2.16	5.31	3.00	4.38	5.28	2.85	3.03	3.03
Skewness	0.06	0.05	-0.57	-0.11	-1.64	1.94	0.67	0.34	-0.34
Range	0.02	0.02	35.00	37.00	1200.00	12.00	1432.00	76.00	76
Minimum	3.15	6.67	30.00	41.00	900.00	4.00	1.00	-112.00	78
Maximum	3.17	6.68	65.00	78.00	2100.00	16.00	1433.00	-36.00	154
Data Points	14142	14142	14142	14142	14142	14142	14142	14142	14142

Table 4.3. Descriptive first-order statistics of testing data

	Longitude	**Latitude**	**Elevation (m)**	**Altitude (m)**	**Frequency (MHz)**	**Clutter Height (m)**	**Distance (m)**	**RSS (dBm)**	**Path Loss (dB)**
Mean	3.16	6.67	51.32	56.56	1830.99	6.60	526.59	-80.35	122.35
Median	3.16	6.67	51.00	57.00	2100.00	6.00	450.00	-81.00	123
Mode	3.16	6.68	52.00	55.00	2100.00	6.00	709.00	-77.00	119
SED	0.00	0.00	4.85	5.43	392.24	3.70	328.05	11.24	11.24
Variance	0.00	0.00	23.51	29.53	153851.46	13.72	107619.77	126.41	126.41
Kurtosis	3.42	2.16	5.32	3.00	4.39	5.35	2.85	3.03	3.03
Skewness	0.06	0.05	-0.57	-0.10	-1.64	1.96	0.67	0.35	-0.35
Range	0.02	0.02	35.00	37.00	1200.00	12.00	1432.00	74.00	74
Minimum	3.15	6.67	30.00	41.00	900.00	4.00	1.00	-112.00	80
Maximum	3.17	6.68	65.00	78.00	2100.00	16.00	1433.00	-38.00	154
Data Points	4714	4714	4714	4714	4714	4714	4714	4714	4714

Table 4.4. Descriptive first-order statistics of training data collected at 900 MHz

	Longitude	**Latitude**	**Elevation (m)**	**Altitude (m)**	**Clutter Height (m)**	**Distance (m)**	**RSS (dBm)**	**Path Loss (dB)**
Mean	3.1598	6.6735	50.08	55.07	7.66	665.87	-73.77	115.77
Median	3.1620	6.6721	51.00	53.00	6.00	698.00	-74.00	116.00
Mode	3.1515	6.6720	52.00	62.00	6.00	709.00	-71.00	106.00
SED	0.0036	0.0036	5.70	7.44	4.21	373.69	10.46	10.46
Variance	0.0000	0.0000	32.54	55.33	17.75	139641.39	109.32	109.32
Kurtosis	2.5090	2.3275	5.04	2.49	3.16	2.19	2.56	2.56
Skewness	-0.9201	0.7779	-1.38	0.35	1.39	0.22	0.03	-0.03
Range	0.0125	0.0124	28.00	37.00	12.00	1432.00	58.00	58.00
Minimum	3.1515	6.6690	30.00	41.00	4.00	1.00	-104.00	88.00
Maximum	3.1639	6.6814	58.00	78.00	16.00	1433.00	-46.00	146.00
Data Points	1908	1908	1908	1908	1908	1908	1908	1908

Table 4.5. Descriptive first-order statistics of testing data collected at 900 MHz

	Longitude	**Latitude**	**Elevation (m)**	**Altitude (m)**	**Clutter Height (m)**	**Distance (m)**	**RSS (dBm)**	**Path Loss (dB)**
Mean	3.1598	6.6735	50.08	55.08	7.64	665.86	-73.86	115.86
Median	3.1620	6.6721	51.00	53.00	6.00	699.00	-74.00	116.00
Mode	3.1515	6.6690	52.00	62.00	6.00	709.00	-76.00	118.00
SED	0.0036	0.0036	5.70	7.45	4.20	373.81	10.48	10.48
Variance	0.0000	0.0000	32.43	55.57	17.66	139730.34	109.81	109.81
Kurtosis	2.5063	2.3301	5.05	2.53	3.19	2.19	2.67	2.67
Skewness	-0.9185	0.7795	-1.38	0.37	1.41	0.22	0.15	-0.15
Range	0.0124	0.0124	28.00	37.00	12.00	1431.00	56.00	56.00
Minimum	3.1515	6.6690	30.00	41.00	4.00	2.00	-103.00	89.00
Maximum	3.1639	6.6814	58.00	78.00	16.00	1433.00	-47.00	145.00
Data Points	635	635	635	635	635	635	635	635

Table 4.6. Descriptive first-order statistics of training data collected at 1800 MHz

	Longitude	**Latitude**	**Elevation (m)**	**Altitude (m)**	**Clutter Height (m)**	**Distance (m)**	**RSS (dBm)**	**Path Loss (dB)**
Mean	3.1631	6.6755	52.40	58.57	6.47	509.99	-81.55	123.55
Median	3.1628	6.6755	52.00	59.00	6.00	418.50	-84.00	126.00
Mode	3.1629	6.6751	52.00	62.00	6.00	62.00	-88.00	130.00
SED	0.0041	0.0036	4.51	4.60	3.77	331.00	11.64	11.64
Variance	0.0000	0.0000	20.37	21.13	14.24	109563.32	135.43	135.43
Kurtosis	2.7126	2.6711	2.83	2.70	5.27	2.55	3.93	3.93
Skewness	0.4439	-0.3524	0.63	0.02	1.94	0.61	0.95	-0.95
Range	0.0175	0.0161	20.00	28.00	12.00	1325.00	69.00	69.00
Minimum	3.1559	6.6665	45.00	49.00	4.00	1.00	-105.00	78.00
Maximum	3.1734	6.6826	65.00	77.00	16.00	1326.00	-36.00	147.00
Data Points	5058	5058	5058	5058	5058	5058	5058	5058

Table 4.7. Descriptive first-order statistics of testing data collected at 1800 MHz

	Longitude	**Latitude**	**Elevation (m)**	**Altitude (m)**	**Clutter Height (m)**	**Distance (m)**	**RSS (dBm)**	**Path Loss (dB)**
Mean	3.1631	6.6755	52.40	58.56	6.44	510.37	-81.63	123.63
Median	3.1628	6.6755	52.00	59.00	6.00	419.00	-84.00	126.00
Mode	3.1630	6.6768	52.00	62.00	6.00	62.00	-90.00	132.00
SED	0.0041	0.0036	4.52	4.59	3.75	331.75	11.49	11.49
Variance	0.0000	0.0000	20.42	21.09	14.04	110058.62	132.10	132.10
Kurtosis	2.7142	2.6728	2.84	2.65	5.37	2.55	3.88	3.88
Skewness	0.4446	-0.3533	0.63	0.01	1.96	0.61	0.94	-0.94
Range	0.0175	0.0162	20.00	25.00	12.00	1322.00	67.00	67.00
Minimum	3.1559	6.6665	45.00	49.00	4.00	1.00	-105.00	80.00
Maximum	3.1734	6.6826	65.00	74.00	16.00	1323.00	-38.00	147.00
Data Points	1687	1687	1687	1687	1687	1687	1687	1687

Table 4.8. Descriptive first-order statistics of training data collected at 2100 MHz

	Longitude	**Latitude**	**Elevation (m)**	**Altitude (m)**	**Clutter Height (m)**	**Distance (m)**	**RSS (dBm)**	**Path Loss (dB)**
Mean	3.1618	6.6746	50.89	55.53	6.44	501.63	-81.20	123.20
Median	3.1623	6.6745	51.00	56.00	6.00	432.00	-81.00	123.00
Mode	3.1630	6.6752	52.00	57.00	6.00	379.00	-78.00	120.00
SED	0.0037	0.0031	4.68	4.91	3.49	302.40	10.70	10.70
Variance	0.0000	0.0000	21.94	24.12	12.20	91446.53	114.59	114.59
Kurtosis	3.3834	1.9825	5.57	2.84	6.29	3.43	2.61	2.61
Skewness	-0.1959	0.1563	-0.84	-0.24	2.15	0.81	-0.06	0.06
Range	0.0202	0.0124	34.00	26.00	12.00	1432.00	61.00	61.00
Minimum	3.1515	6.6690	30.00	41.00	4.00	1.00	-112.00	93.00
Maximum	3.1717	6.6814	64.00	67.00	16.00	1433.00	-51.00	154.00
Data Points	7176	7176	7176	7176	7176	7176	7176	7176

Table 4.9. Descriptive first-order statistics of testing data collected at 2100 MHz

	Longitude	**Latitude**	**Elevation (m)**	**Altitude (m)**	**Clutter Height (m)**	**Distance (m)**	**RSS (dBm)**	**Path Loss (dB)**
Mean	3.1618	6.6746	50.89	55.53	6.44	501.07	-81.17	123.17
Median	3.1623	6.6745	51.00	56.00	6.00	431.50	-81.00	123.00
Mode	3.1515	6.6706	52.00	57.00	6.00	355.00	-79.00	121.00
SED	0.0037	0.0031	4.68	4.91	3.48	302.47	10.67	10.67
Variance	0.0000	0.0000	21.92	24.14	12.13	91487.07	113.88	113.88
Kurtosis	3.3856	1.9845	5.60	2.85	6.33	3.43	2.60	2.60
Skewness	-0.2000	0.1562	-0.85	-0.24	2.16	0.81	-0.06	0.06
Range	0.0202	0.0124	34.00	26.00	12.00	1432.00	61.00	61.00
Minimum	3.1515	6.6690	30.00	41.00	4.00	1.00	-112.00	93.00
Maximum	3.1717	6.6814	64.00	67.00	16.00	1433.00	-51.00	154.00
Data Points	2392	2392	2392	2392	2392	2392	2392	2392

Exploratory summaries of the univariate data analysis are displayed using boxplot representations. Figures 4.4-4.11 display single batch data about longitude, latitude, elevation, altitude, clutter height, distance, RSS, and path loss measured respectively along 14 survey drive test routes. These graphical representations provide overview details about the location, spread, Skewness, and 'longtailedness' of each variable at a quick glance. In all the eight classes of data represented, there are significant variations in the middle cut lines (median), lengths of the boxes (spread), deviations of the median lines from the center of the boxes relative to the lengths of the boxes (skewness), and the distances between the ends of the whiskers relative to the length of the box ('longtailedness'). These significant variations provide the diversity of data instances needed for the training and testing of efficient machine learning models.

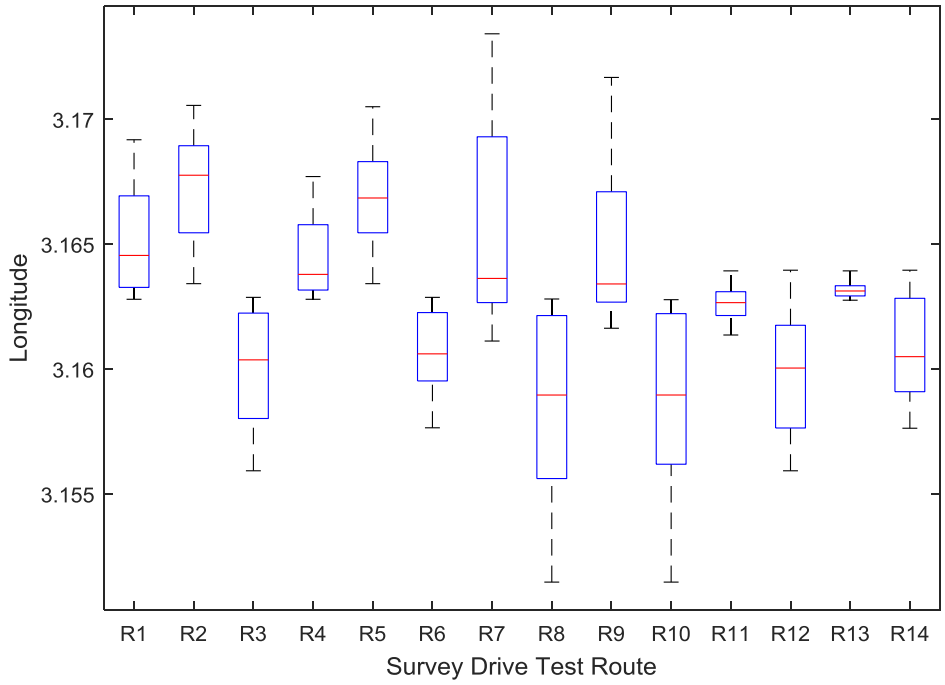

Figure 4.4. Boxplot of longitude data along 14 survey drive test routes

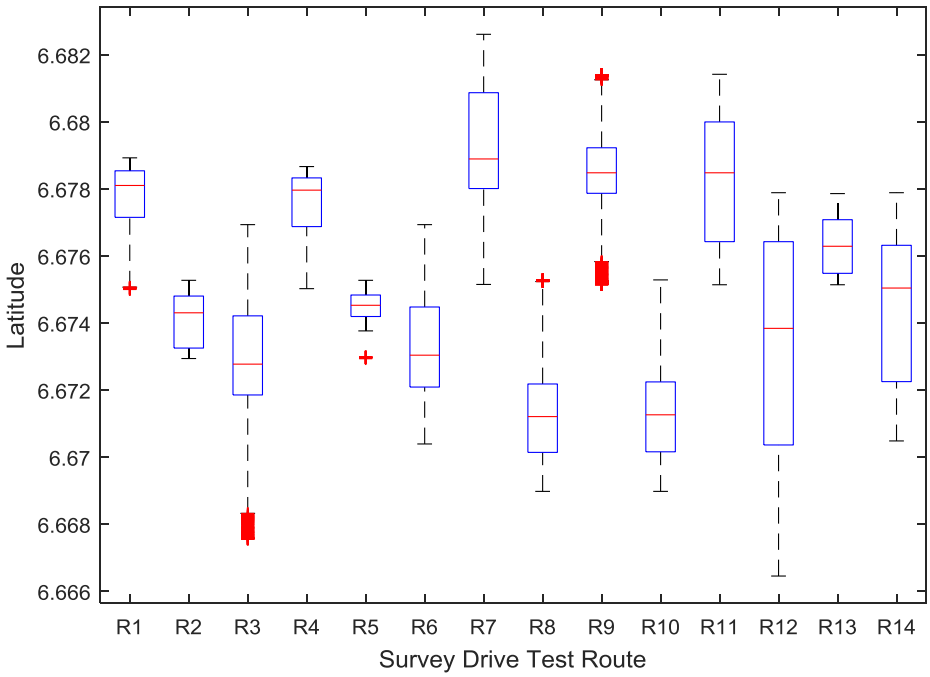

Figure 4.5. Boxplot of latitude data along 14 survey drive test routes

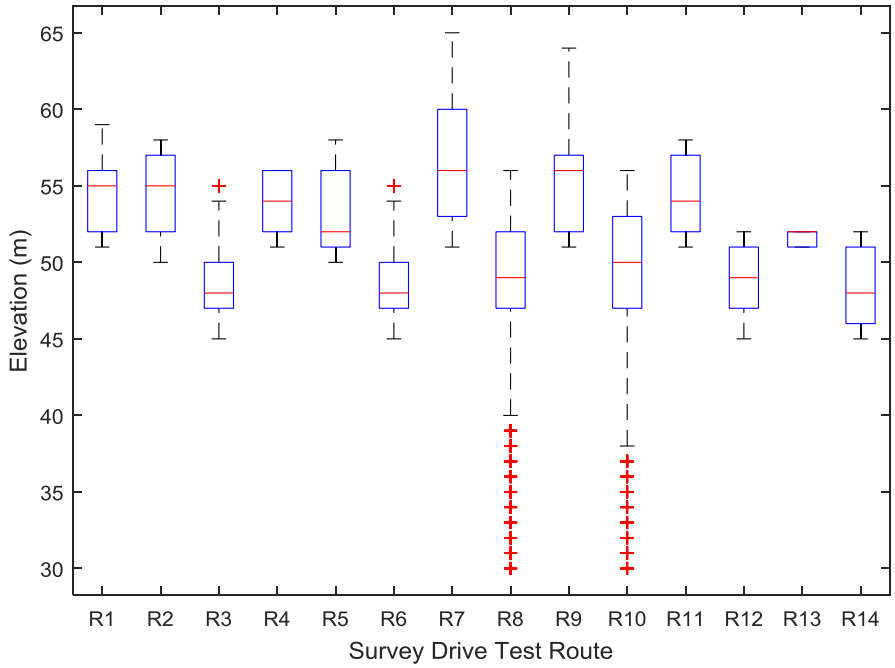

Figure 4.6. Boxplot of elevation data along 14 survey drive test routes

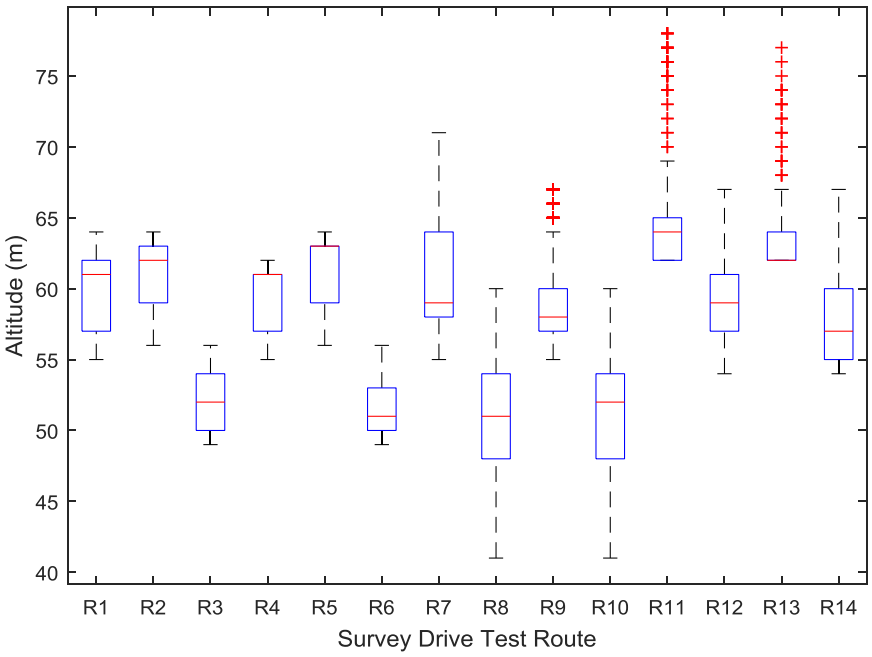

Figure 4.7. Boxplot of altitude data along 14 survey drive test routes

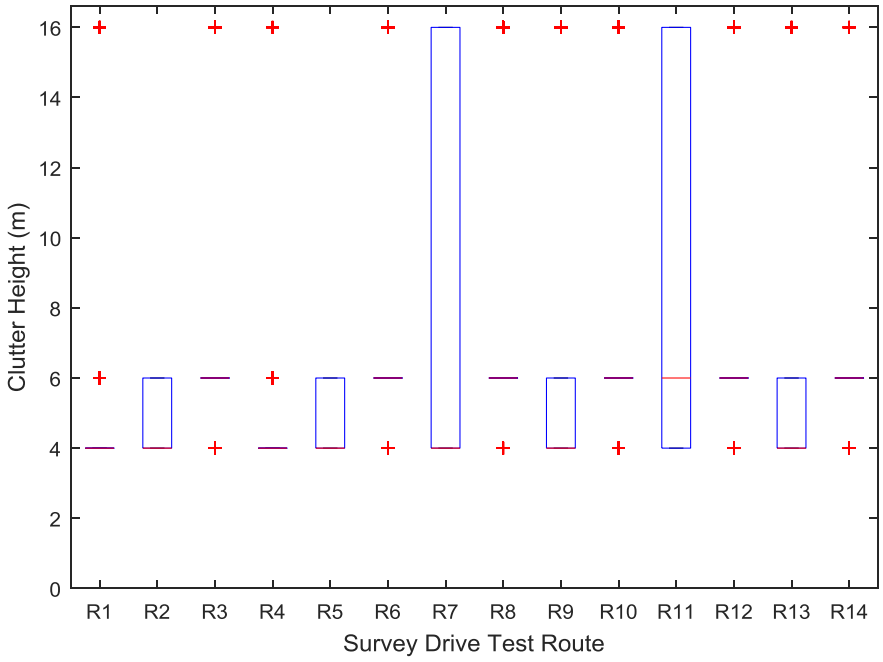

Figure 4.8. Boxplot of clutter height data along 14 survey drive test routes

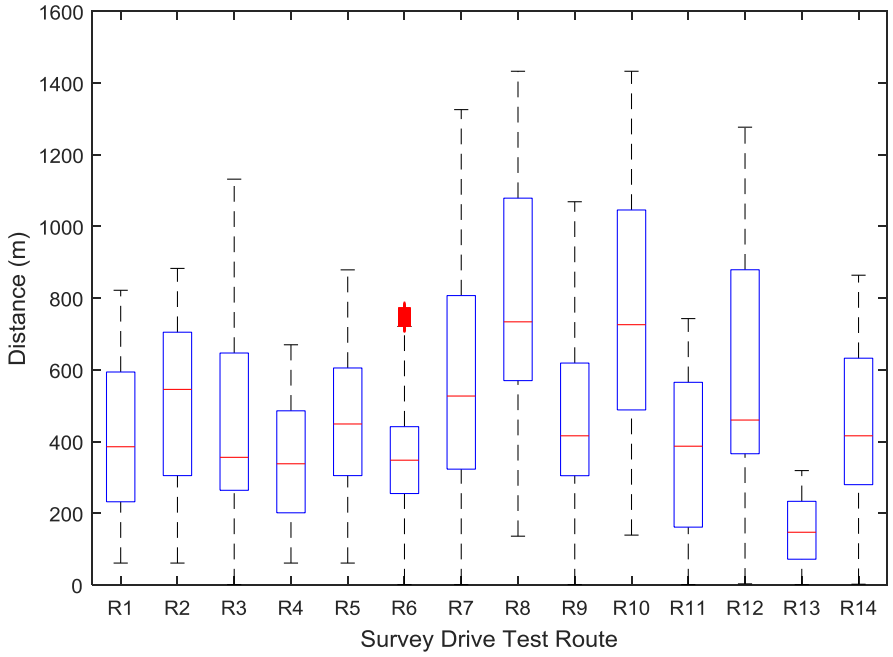

Figure 4.9. Boxplot of distance data along 14 survey drive test routes

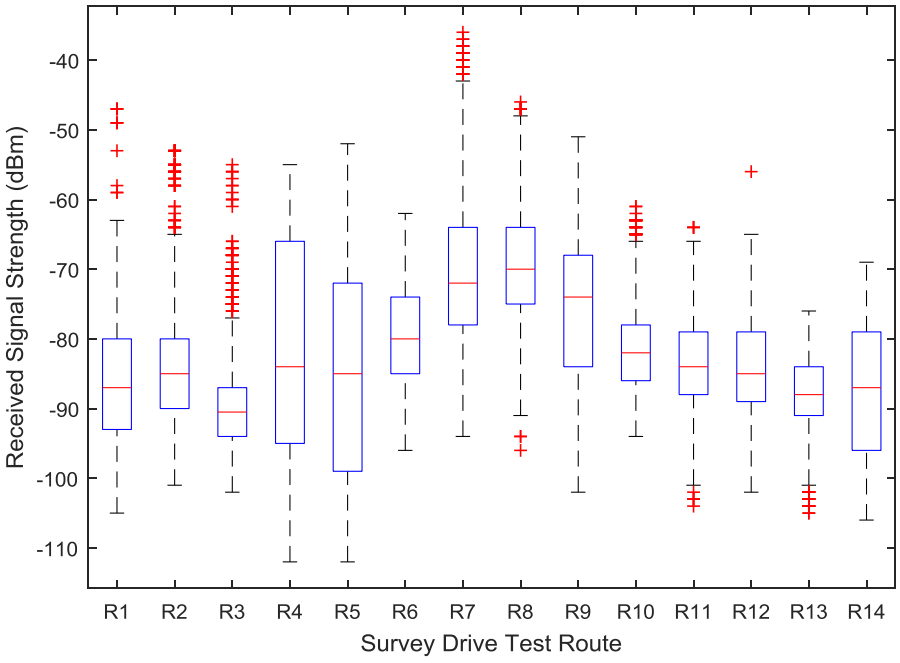

Figure 4.10. Boxplot of RSS data along 14 survey drive test routes

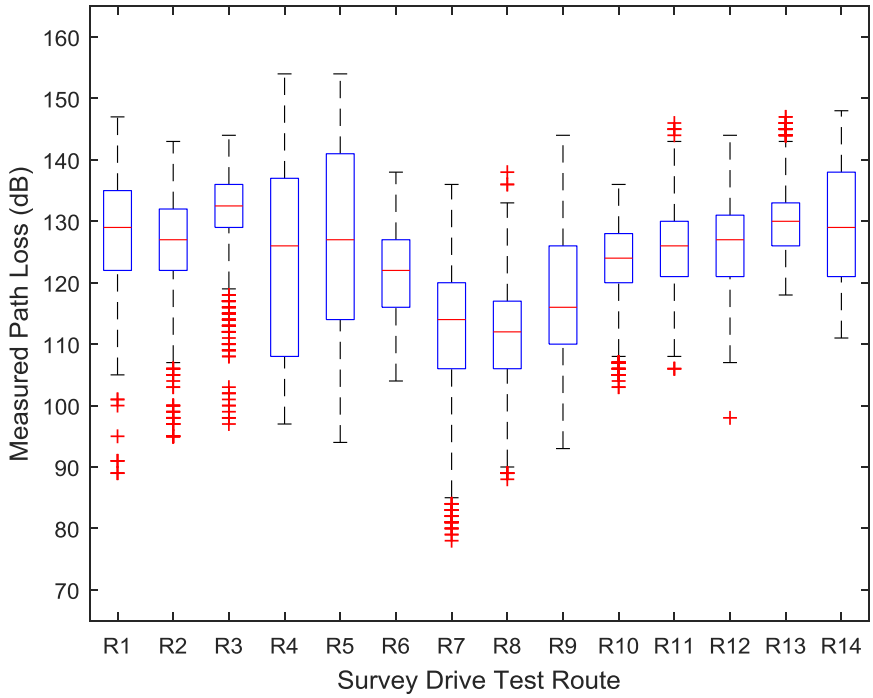

Figure 4.11. Boxplot of distance data along 14 survey drive test routes

Frequency distribution histograms of longitude, latitude, elevation, altitude, clutter height, distance, RSS, and path loss measured are shown in Figures 4.12-4.20 respectively to validate the assumption of data normality. The bell-shapes of the data distributions were determined by visual inspection.

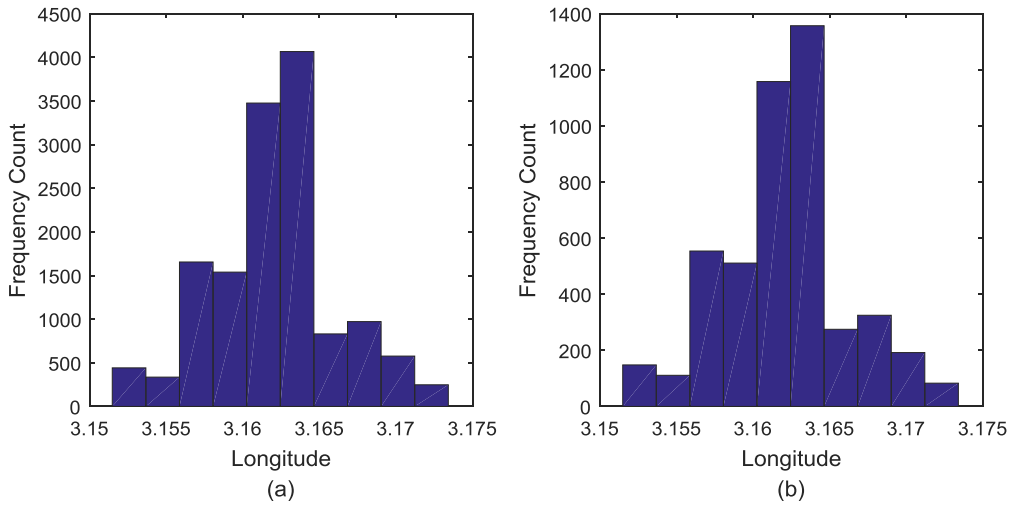

Figure 4.12. Frequency distribution histograms of longitude in (a) training data and (b) testing data

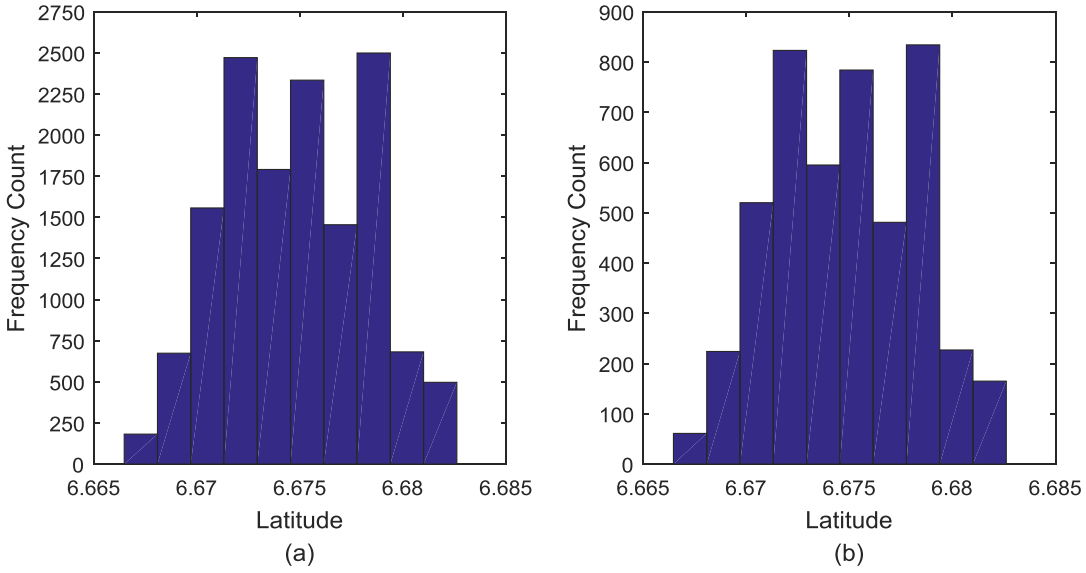

Figure 4.13. Frequency distribution histograms of latitude in (a) training data and (b) testing data

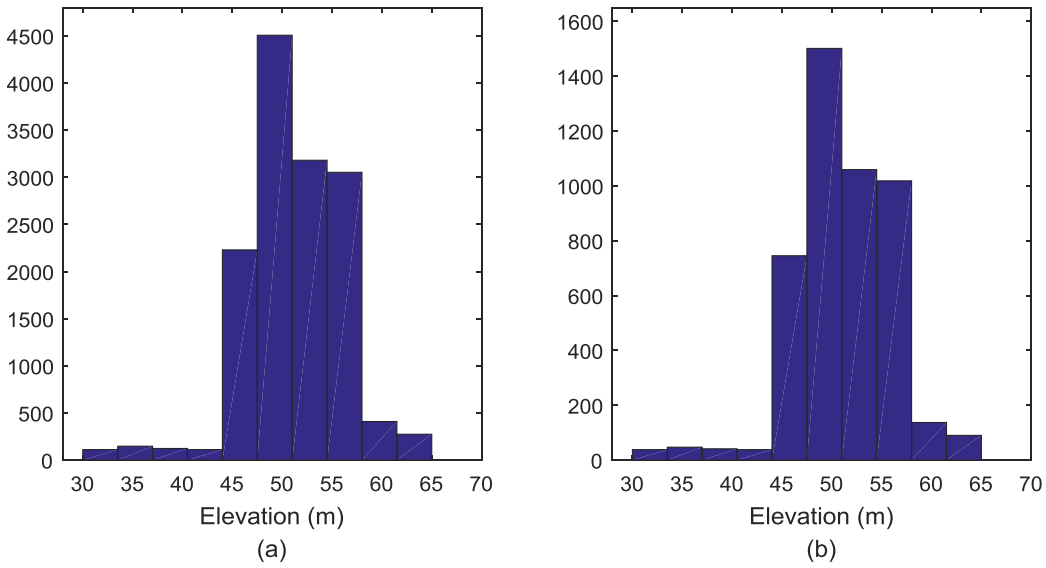

Figure 4.14. Frequency distribution histograms of elevation in (a) training data and (b) testing data

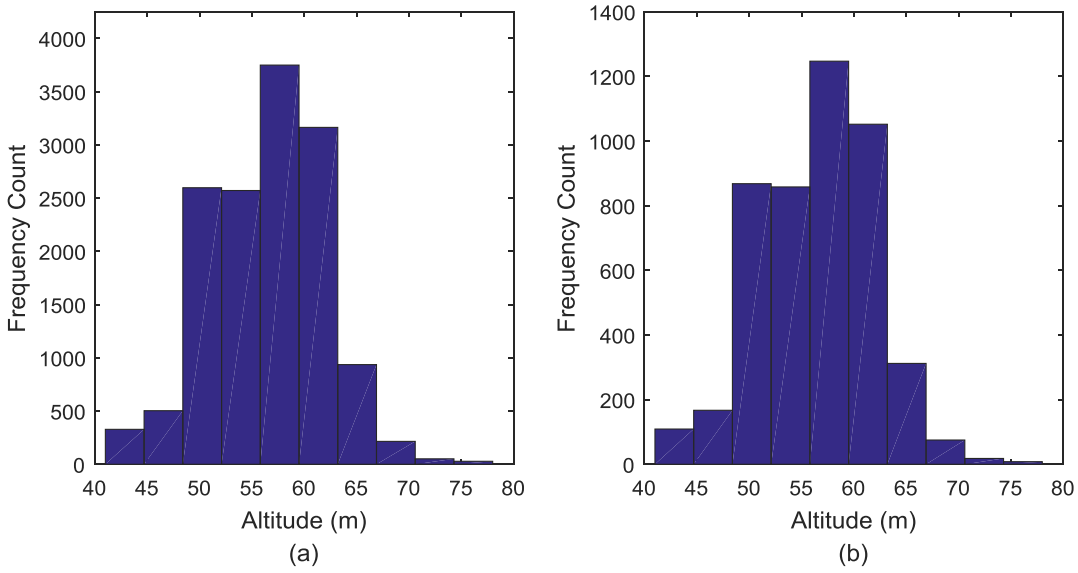

Figure 4.15. Frequency distribution histograms of altitude in (a) training data and (b) testing data

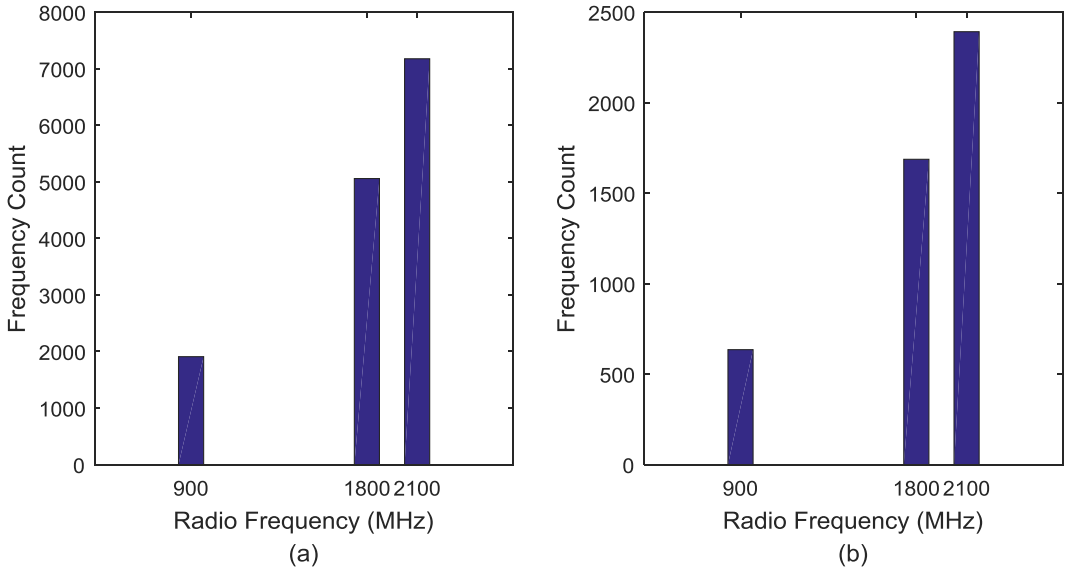

Figure 4.16. Frequency distribution histograms of radio frequency in (a) training data and (b) testing data

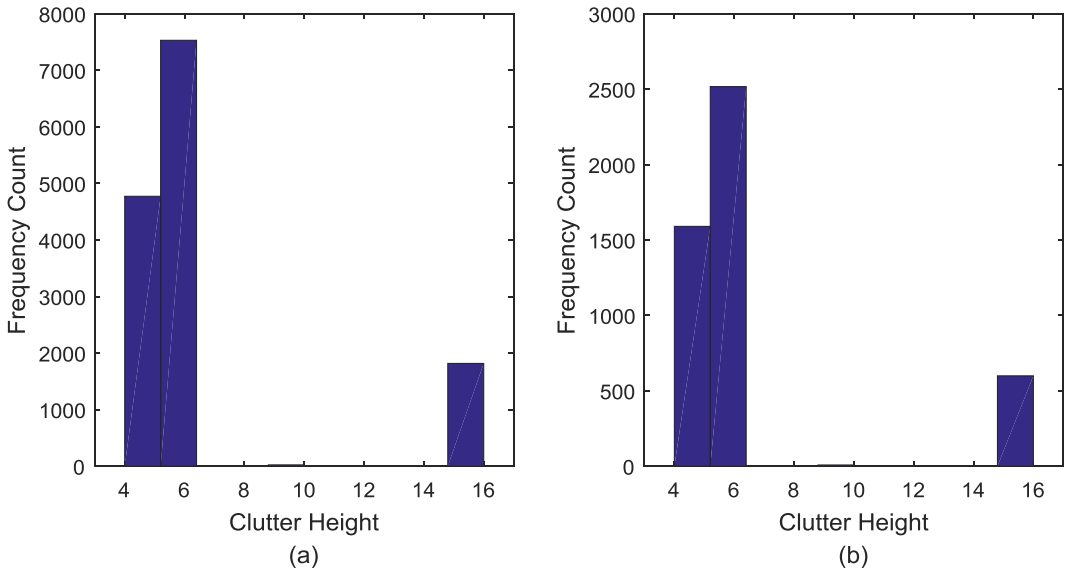

Figure 4.17. Frequency distribution histograms of clutter height in (a) training data and (b) testing data

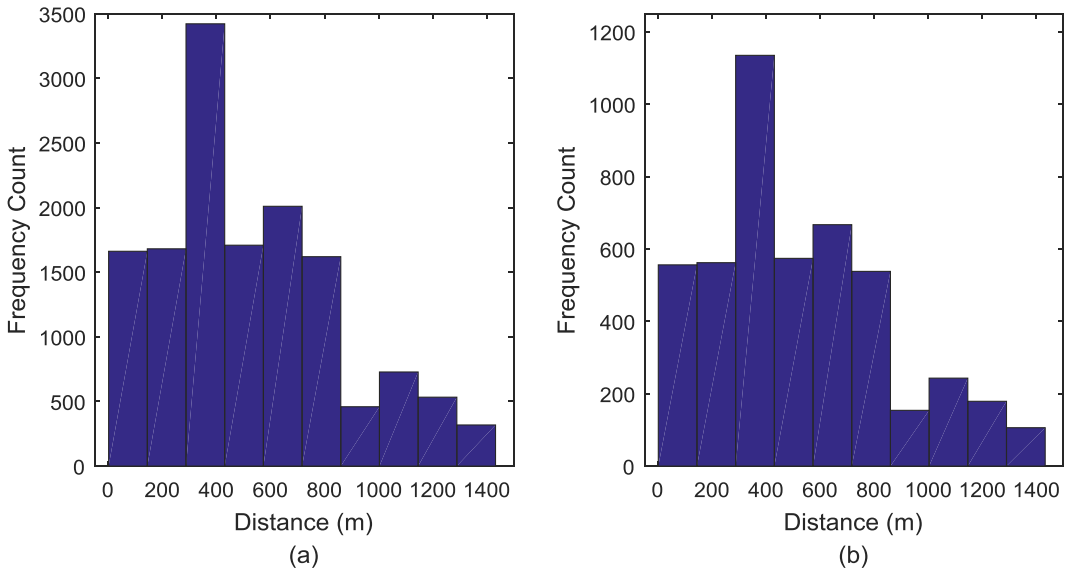

Figure 4.18. Frequency distribution histograms of distance in (a) training data and (b) testing data

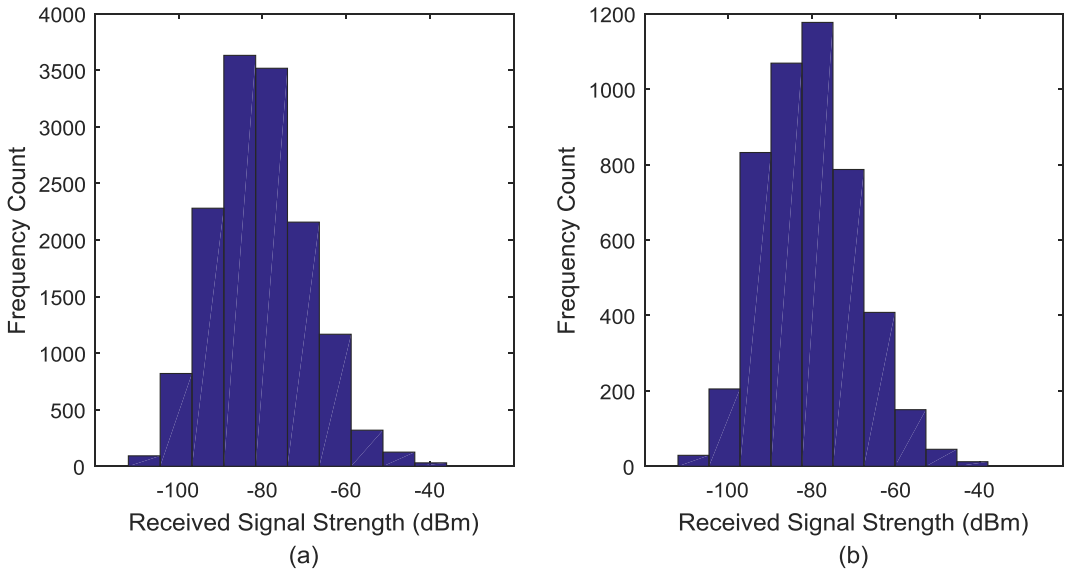

Figure 4.19. Frequency distribution histograms of RSS in (a) training data and (b) testing data

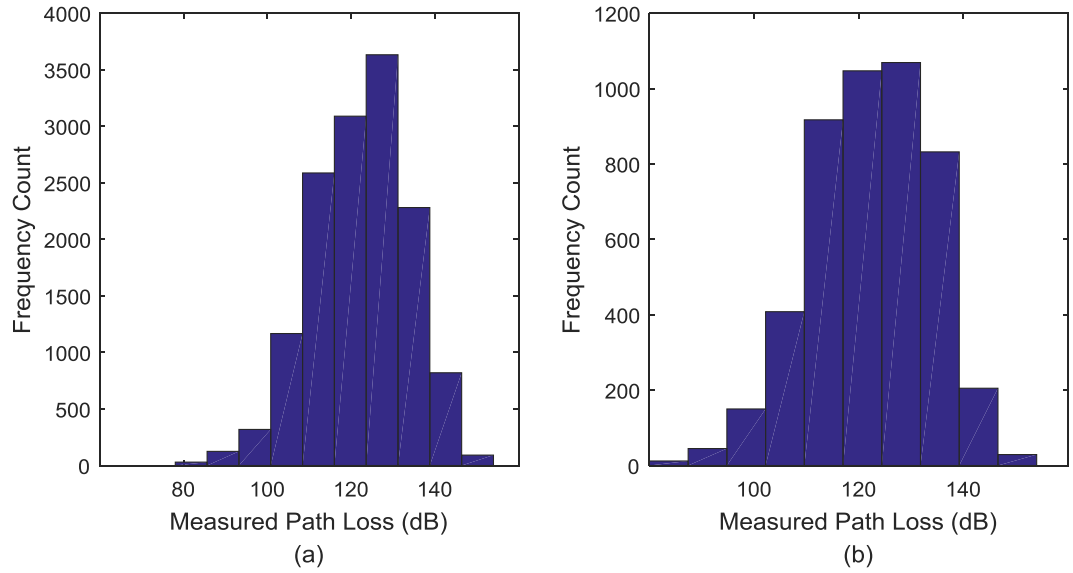

Figure 4.20. Frequency distribution histograms of measured path loss in (a) training data and (b) testing data

4.4. Statistical Analysis of Field Measurement Data

Simple linear regression was used to model the relationship between path loss (dependent variable) and each of the associated independent variables (longitude, latitude, elevation, altitude, clutter

height, distance, and RSS), one independent variable at a time. The unknown model parameters of the linear predictor functions were estimated from the complete data. The regression lines, linear model equations, and the correlation coefficients are shown in Figures 4.21-4.28. It was observed that latitude, elevation, altitude, radio frequency, and distance have direct relationships with path loss; while longitude, clutter height, and RSS are indirectly related to path loss. The correlation coefficients of the linear relationships between path loss and longitude, latitude, elevation, altitude, clutter height, distance, and RSS are 0.1042, -0.1123, 0.0692, 0.1775, 0.2086, -0.1197, 0.1704, and -1 respectively. In addition, cross-correlation coefficient matrix and the p-value matrix of the variables are presented in Table 4.10 and Table 4.11 respectively. Weak cross-correlations (i.e. correlation coefficients of less than 0.5) were observed among longitude, frequency, clutter height, distance, RSS, and path loss. On the other hand, longitude, latitude, elevation, and altitude were observed to be strongly correlated (i.e. correlation coefficients were greater than 0.5). Likewise, RSS has a strong inverse relationship with path loss. P-values of the relationships between the variables were computed to determine whether the correlation coefficients obtained are statistically significant (i.e. $p \leq 0.05$). Results presented in Table 4.11 show that all relationships are statistically significant except in the cases of clutter height versus elevation ($p = 0.0814$), frequency versus altitude ($p = 0.1226$), and clutter versus altitude ($p = 0.7519$).

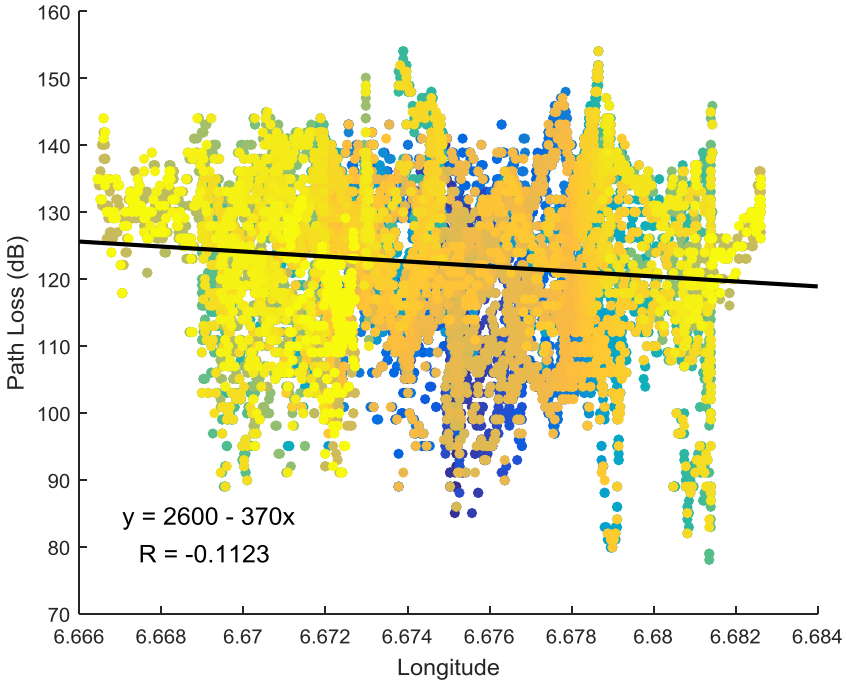

Figure 4.21. Scatter plot of path loss versus longitude

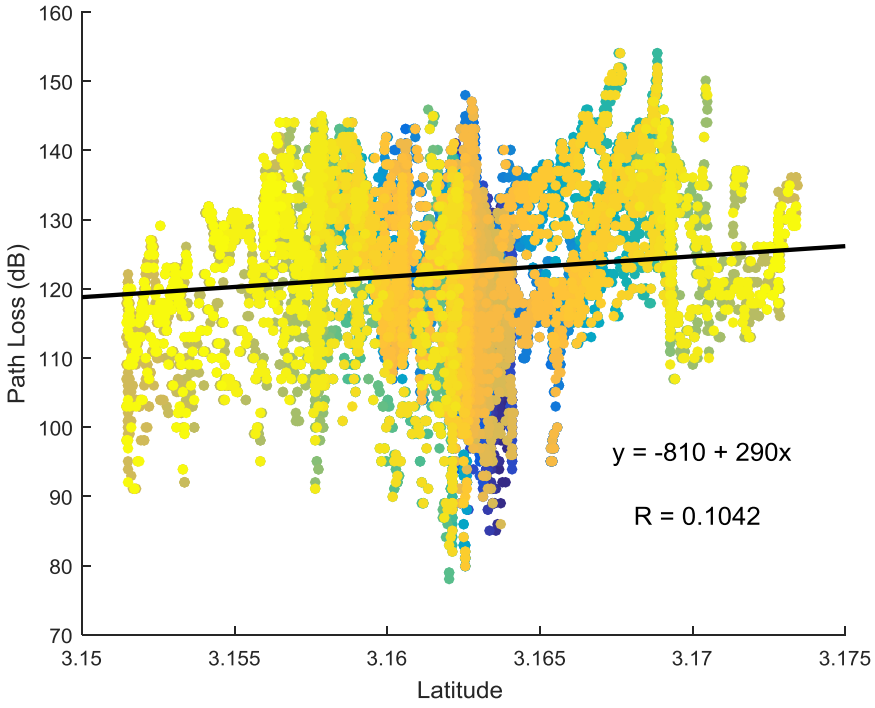

Figure 4.22. Scatter plot of path loss versus latitude

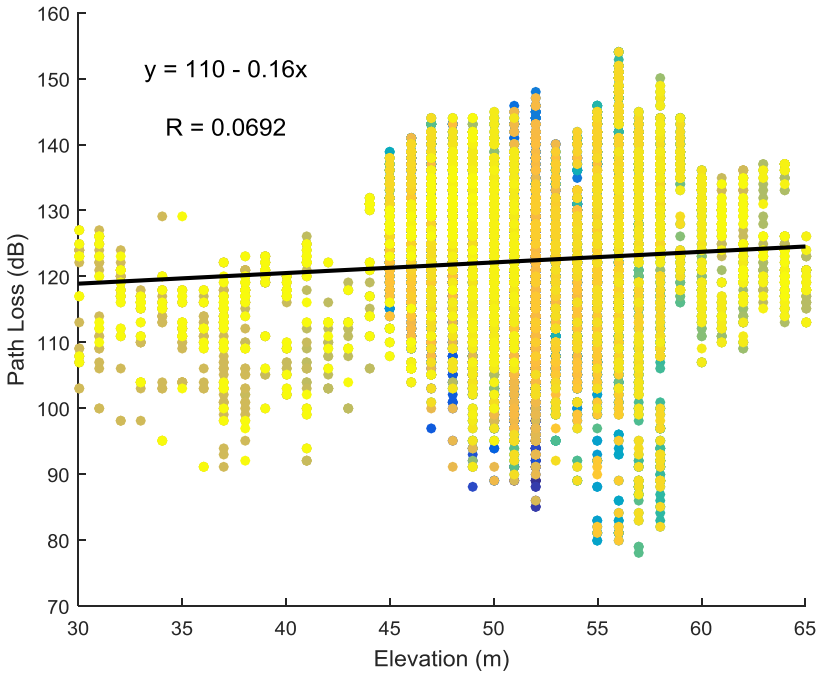

Figure 4.23. Scatter plot of path loss versus elevation

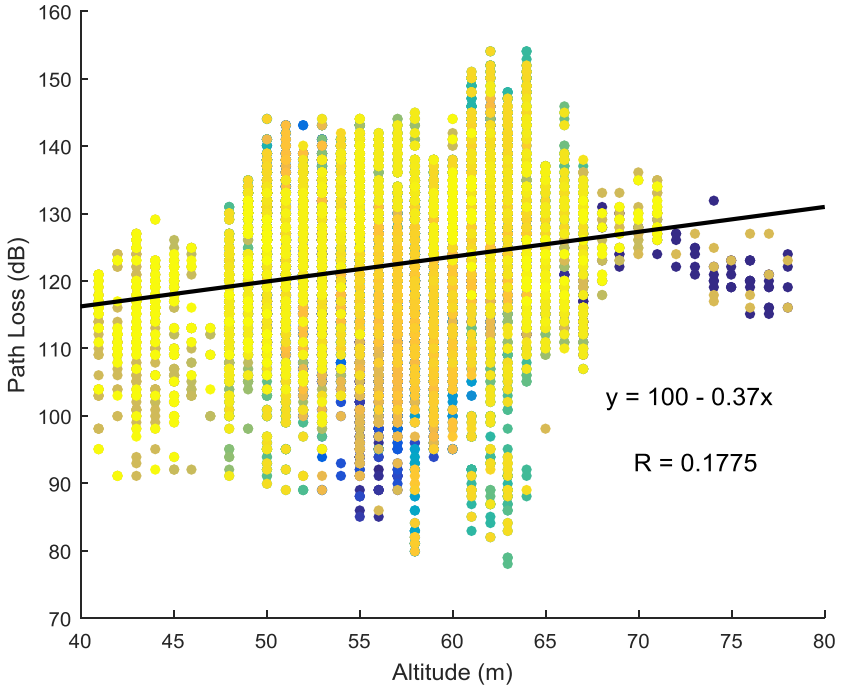

Figure 4.24. Scatter plot of path loss versus altitude

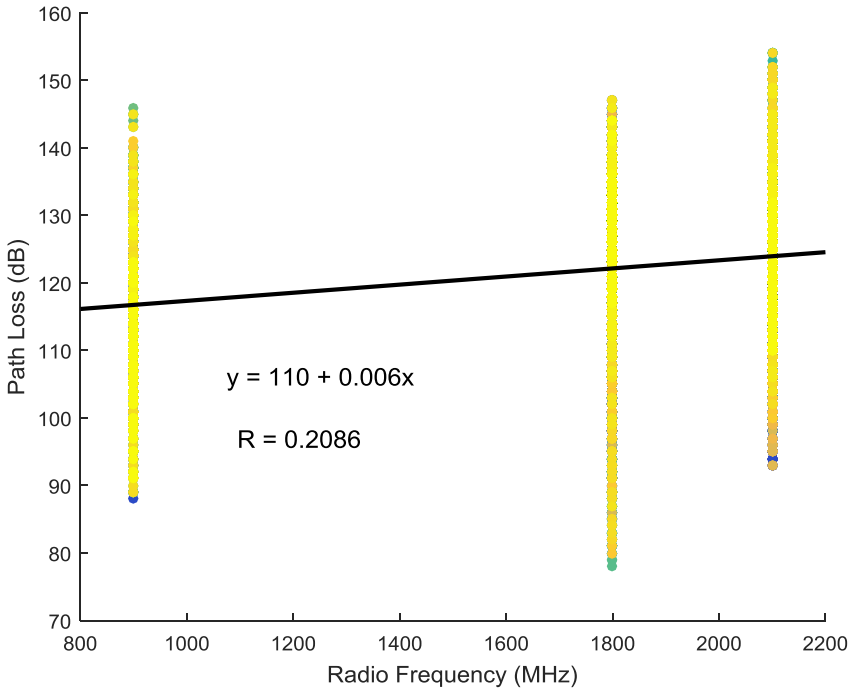

Figure 4.25. Scatter plot of path loss versus radio frequency

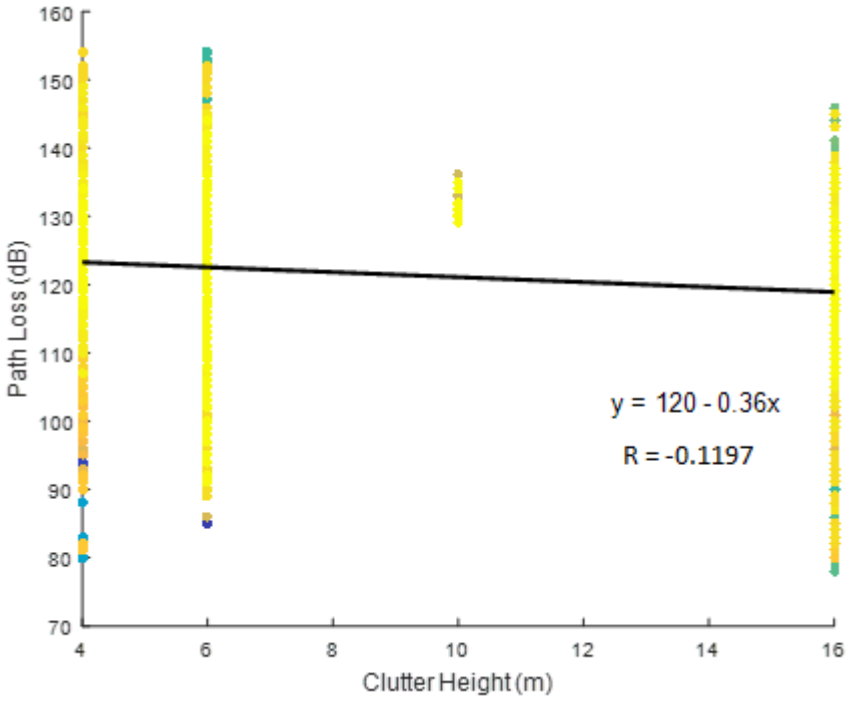

Figure 4.26. Scatter plot of path loss versus clutter height

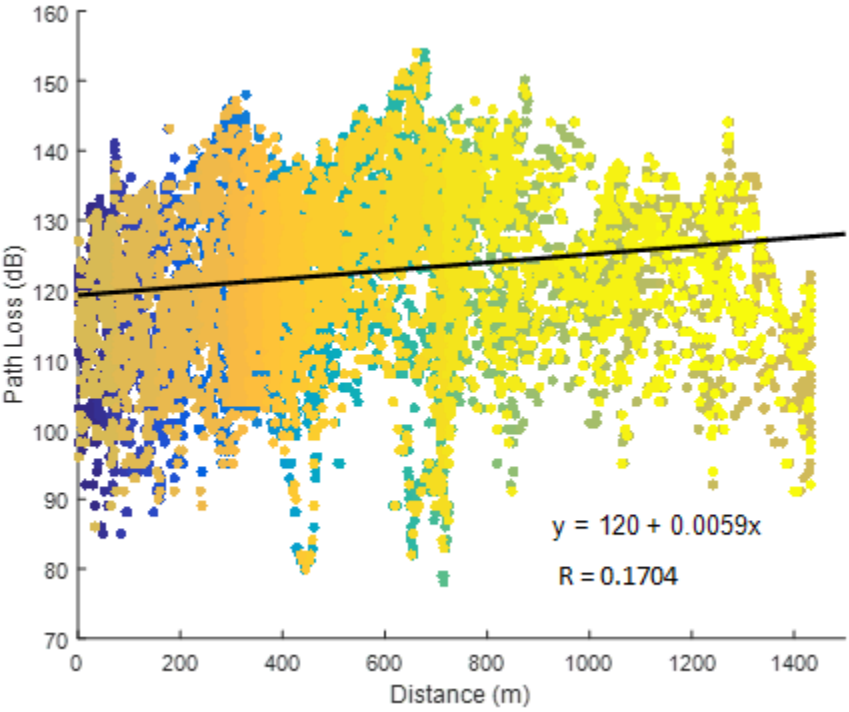

Figure 4.27. Scatter plot of path loss versus distance

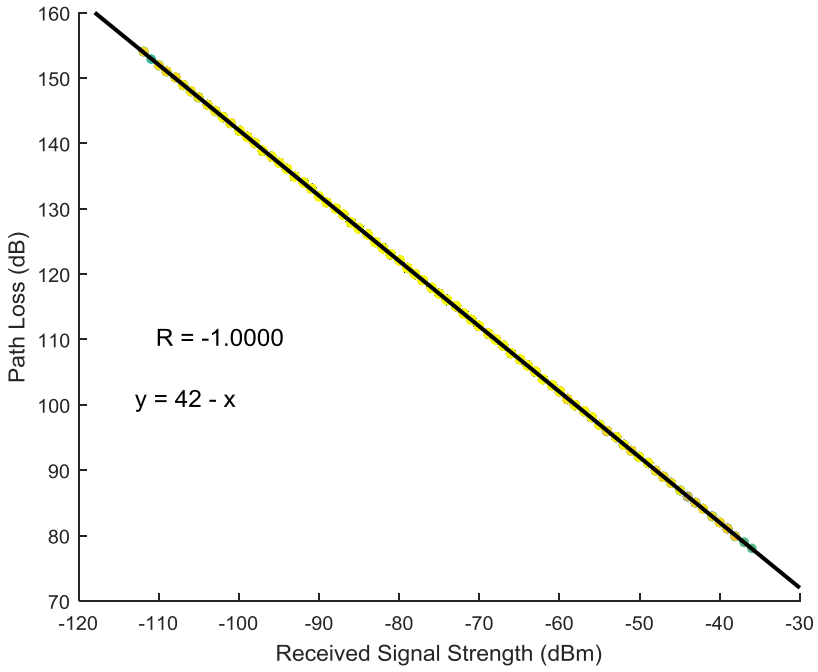

Figure 4.28. Scatter plot of path loss versus RSS

Table 4.10**.** Correlation coefficient matrix

	Longitude	**Latitude**	**Elevation (m)**	**Altitude (m)**	**Frequency (MHz)**	**Clutter Height (m)**	**Distance (m)**	**RSS (dBm)**	**Path Loss (dB)**
Longitude	1.0000	0.6555	0.8191	0.7402	0.1572	-0.0778	-0.3551	-0.1042	0.1042
Latitude	0.6555	1.0000	0.5678	0.6194	0.0961	0.1520	-0.3174	0.1123	-0.1123
Elevation (m)	0.8191	0.5678	1.0000	0.7485	0.0449	-0.0127	-0.1356	-0.0692	0.0692
Altitude (m)	0.7402	0.6194	0.7485	1.0000	0.0112	-0.0023	-0.2284	-0.1775	0.1775
Frequency (MHz)	0.1572	0.0961	0.0449	0.0112	1.0000	-0.1049	-0.1612	-0.2086	0.2086
Clutter Height (m)	-0.0778	0.1520	-0.0127	-0.0023	-0.1049	1.0000	0.1108	0.1197	-0.1197
Distance (m)	-0.3551	-0.3174	-0.1356	-0.2284	-0.1612	0.1108	1.0000	-0.1704	0.1704
RSS (dBm)	-0.1042	0.1123	-0.0692	-0.1775	-0.2086	0.1197	-0.1704	1.0000	-1.0000
PL (dB)	0.1042	-0.1123	0.0692	0.1775	0.2086	-0.1197	0.1704	-1.0000	1.0000

Table 4.11. P-value matrix

	Longitude	**Latitude**	**Elevation (m)**	**Altitude (m)**	**Frequency (MHz)**	**Clutter Height (m)**	**Distance (m)**	**RSS (dBm)**	**Path Loss (dB)**
Longitude	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Latitude	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Elevation (m)	0.0000	0.0000	1.0000	0.0000	0.0000	0.0814	0.0000	0.0000	0.0000
Altitude (m)	0.0000	0.0000	0.0000	1.0000	0.1226	0.7519	0.0000	0.0000	0.0000
Frequency (MHz)	0.0000	0.0000	0.0000	0.1226	1.0000	0.0000	0.0000	0.0000	0.0000
Clutter Height (m)	0.0000	0.0000	0.0814	0.7519	0.0000	1.0000	0.0000	0.0000	0.0000
Distance (m)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
RSS (dBm)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
PL (dB)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

4.5. ANN-Based Path Loss Models with Varying Input Data Requirements

Six different ANN-based path loss models were developed to determine the minimum input(s) that guarantee sufficiently high prediction accuracy and good generalization ability. The six ANN-based path loss models depend on a single input variable (i.e. distance), two input variables (i.e. distance and frequency), three input variables (i.e. distance, frequency, and clutter height), four input variables (i.e. distance, frequency, clutter height, and elevation), five input variables (distance, frequency, clutter height, elevation, and altitude), and seven input variables (i.e. distance, frequency, clutter height, elevation, altitude, latitude, and longitude) respectively. The prediction accuracy and the generalization ability of the six ANN-based path loss models were evaluated relative to the measured path loss values in the training and testing datasets respectively using MAE, MSE, RMSE, SED, and R. The results showing the prediction accuracy and the generalization ability of the developed ANN-based path loss models are presented in Table 4.12 and Table 4.13 respectively. When the prediction accuracy of the ANN-based path loss models was evaluated using training dataset, it was found that the prediction error significantly reduced (i.e. MAE, MSE, RMSE, and SED decreased from 8.202 dB to 4.553 dB, 107.813 dB to 35.657 dB, 10.383 dB to 5.966 dB, and 10.383 dB to 5.966 dB respectively) and the correlation between the measured path loss values and the predicted path loss values became stronger as the number of input variables increased from one to seven (i.e. R increased from 0.397 to 0.849). Regarding generalization ability, the testing dataset was used and it was observed that the generalization ability of ANN-based path loss model improved as more input variables are added from one to seven (i.e. MAE, MSE, RMSE, and SED decreased from 8.164 dB to 4.549 dB, 106.432 dB to 35.789 dB, 10.316 dB to 5.076 dB, and 10.317 dB to 5.977 dB respectively) and the correlations between the measured path loss values in both datasets and the predicted path loss values became stronger as the number of input variables increased from one to seven (i.e. R increased from 0.397 to 0.847, and 0.397 to 0.847 respectively). Conversely, the time required to train the ANN-based path loss models increased from 29.675 seconds to 105.886 seconds as the number of input variables increased from one to seven.

Table 4.12. Optimal path loss model with minimum input variable(s)

Case	Input Variable(s)	Training Time (seconds)	MAE (dB)	MSE (dB)	RMSE (dB)	SED (dB)	R
1	Distance	29.675	8.202	107.813	10.383	10.383	0.397
2	Distance, frequency	60.176	7.465	91.845	9.583	9.583	0.531
3	Distance, frequency, clutter height	67.779	6.416	69.801	8.354	8.355	0.674
4	Distance, frequency, clutter height, elevation	84.336	5.469	50.978	7.137	7.137	0.776
5	Distance, frequency, clutter height, elevation, altitude	80.799	5.339	53.311	7.173	7.163	0.777
6	Distance, frequency, clutter height, elevation, altitude, latitude, longitude	105.886	4.553	35.657	5.966	5.966	0.849

Table 4.13. Results of generalization ability testing

Case	Input Variable(s)	MAE (dB)	MSE (dB)	RMSE (dB)	SED (dB)	R
1	Distance	8.164	106.432	10.316	10.317	0.397
2	Distance, frequency	7.450	90.953	9.536	9.537	0.529
3	Distance, frequency, clutter height	6.414	69.168	8.316	8.317	0.673
4	Distance, frequency, clutter height, elevation	5.466	50.955	7.136	7.136	0.773
5	Distance, frequency, clutter height, elevation, altitude	5.335	53.233	7.170	7.161	0.774
6	Distance, frequency, clutter height, elevation, altitude, latitude, longitude	4.549	35.786	5.976	5.977	0.847

4.6. Effect of Input Data Normalization on ANN-Based Model Prediction Accuracy and Generalization Ability

The results of the experimentations performed to validate the effect of input data normalization on path loss prediction accuracy and generalization ability are presented in Table 4.14 and Table 4.15. It was observed that input data normalization significantly improved path loss prediction accuracy and it equally produced better generalization. The results in Table 4.14 show that the MAE, MSE, RMSE, and the SED between the measured path loss in the training dataset and the path loss values predicted by ANN-based models decreased from 4.553 dB to 4.448 dB, 35.6576 dB to 34.088 dB, 5.966 dB to 5.836 dB, and 5.966 dB to 5.836 dB respectively.

Similarly, the MAE, MSE, RMSE, and the SED between the measured path loss in the testing dataset and the path loss values predicted by ANN-based models reduced from 4.549 dB to 4.437 dB, 35.786 dB to 34.211 dB, 5.976 dB to 5.846 dB, and 5.976 dB to 5.846 dB respectively. In addition, the strong relationships between the measured path loss values in both datasets and the path loss values predicted by ANN-based models were improved, as the R value increased from 0.849 to 0.857, and 0.847 to 0.854 respectively. However, the input data normalization process further increased the training time in the development of the ANN-based models by 6.173 seconds. The complete results showing the effect of input data normalization on model generalization are presented in Table 4.15.

Table 4.14. Effect of input data normalization on path loss prediction accuracy

	Training Time (seconds)	**MAE (dB)**	**MSE (dB)**	**RMSE (dB)**	**SED (dB)**	**R**
Raw Data	105.886	4.553	35.657	5.966	5.966	0.849
Normalized Data	112.059	4.448	34.088	5.836	5.836	0.857

Table 4.15. Effect of input data normalization on generalization ability

	MAE (dB)	MSE (dB)	RMSE (dB)	SED (dB)	R
Raw Data	4.549	35.786	5.976	5.977	0.847
Normalized Data	4.437	34.211	5.846	5.847	0.854

4.7. ANN-Based Path Loss Models with Varying Transfer Functions

Nine ANN-based path loss models were developed to determine the most suitable combination of transfer/activation functions for path loss predictions in a heterogeneous smart campus environment. On training dataset, the use of *logsig* and *tansig* activation functions at the hidden and output layers of the neural network produced the best path loss prediction accuracy with MAE, MSE, RMSE, SED, and R values of 4.437 dB, 33.928 dB, 5.822 dB, 5.822 dB, and 0.857 respectively as presented in Table 4.16. However, the ANN model requires relatively much training time (i.e. 127.401 seconds). The worst path loss prediction accuracy (MAE, MSE, RMSE, SED, and R values of 8.043 dB, 99.406 dB, 9.970 dB, 9.966 dB, and 0.474 respectively) was obtained when *purelin* and *logsig* activation functions were employed at the hidden and output layers respectively. On the other hand, ANN-based path loss models which utilized *purelin-logsig* and *purelin-tansig* combinations took lesser time (i.e. 3.326 and 1.555 seconds respectively) to train but the prediction accuracy of the two ANN-based path loss models were very low.

On testing dataset, the use of *logsig* and *tansig* activation functions at the hidden and output layers of the neural network produced the best model generalization with MAE, MSE, RMSE, SED, and R values of 4.421 dB, 33.992 dB, 5.828 dB, 5.828 dB, and 0.855 respectively as presented in Table 4.17. The worst model generalization (MAE, MSE, RMSE, SED, and R values of 8.006 dB, 98.022 dB, 9.901 dB, 9.897, and 0.475 respectively) was obtained when *purelin* and *logsig* activation functions were employed at the hidden and output layers respectively.

Table 4.16. Training Results of ANN-Based Path Loss Models with Varying Transfer Functions

Activation Function at the Hidden Layer	Activation Function at the Output Layer	Training Time (seconds)	MAE (dB)	MSE (dB)	RMSE (dB)	SED (dB)	R
Purelin	Purelin	0.713	7.932	98.407	9.920	9.920	0.481
Logsig	Purelin	117.709	4.615	36.772	6.042	6.042	0.844
Tansig	Purelin	107.118	4.631	37.245	6.095	6.095	0.842
Purelin	Logsig	3.326	8.043	99.406	9.970	9.966	0.474
Logsig	Logsig	105.118	5.574	60.150	7.539	7.232	0.723
Tansig	Logsig	115.303	5.831	66.808	7.868	7.684	0.725
Purelin	Tansig	1.555	7.898	97.924	9.896	9.896	0.485
Logsig	Tansig	127.401	4.437	33.928	5.822	5.822	0.857
Tansig	Tansig	110.060	4.581	36.537	6.029	6.029	0.845

Table 4.17: Testing Results of ANN-Based Path Loss Models with Varying Transfer Functions

Activation Function at the Hidden Layer	Activation Function at the Output Layer	MAE (dB)	MSE (dB)	RMSE (dB)	SED (dB)	R
Purelin	Purelin	7.904	97.129	9.855	9.856	0.481
Logsig	Purelin	4.599	36.852	6.049	6.049	0.841
Tansig	Purelin	4.625	37.320	6.101	6.102	0.839
Purelin	Logsig	8.006	98.022	9.901	9.897	0.475
Logsig	Logsig	5.532	59.500	7.496	7.193	0.723
Tansig	Logsig	5.793	66.216	7.828	7.648	0.723
Purelin	Tansig	7.869	96.685	9.833	9.834	0.485
Logsig	Tansig	4.421	33.992	5.828	5.828	0.855
Tansig	Tansig	4.576	36.748	6.047	6.048	0.842

4.8. ANN-Based Path Loss Models with Varying Training Algorithms

Nine ANN-based path loss models were developed to determine the best learning algorithm for path loss predictions in a heterogeneous smart campus environment. As shown in Table 4.18, training the ANN-based path loss model using Levenberg-Marquardt learning algorithm produced the best path loss prediction accuracy with MAE, MSE, RMSE, SED, and R values of 4.485 dB, 34.720 dB, 5.888 dB, 5.888 dB, and 0.854 respectively. However, Levenberg-Marquardt learning algorithm took relatively much time of 123.763 seconds to train the ANN-based path loss. The worst path loss prediction accuracy (MAE, MSE, RMSE, SED, and R values of 13.652 dB, 372.883 dB, 15.412 dB, 9.937 dB, and 0.566 respectively) was obtained when the ANN-based path loss model was trained based on Variable Learning Rate Backpropagation learning rule. Conversely, Variable Learning Rate Backpropagation learning algorithm required the least time of training.

On testing dataset, the ANN-based path loss model which employed Levenberg-Marquardt learning algorithm produced the best model generalization with MAE, MSE, RMSE, SED, and R values of 4.473 dB, 34.849 dB, 5.899 dB, 5.900 dB, and 0.851 respectively as presented in Table 4.19. The worst model generalization (MAE, MSE, RMSE, SED, and R values of 13.638 dB, 372.194 dB, 15.389 dB, 9.906 dB, and 0.564 respectively) was obtained when the ANN-based path loss model was trained based on Variable Learning Rate Backpropagation learning algorithm.

Table 4.18: Training Results of ANN-Based Path Loss Models with Varying Learning Rules

Training Algorithm	Training Time (seconds)	MAE (dB)	MSE (dB)	RMSE (dB)	SED (dB)	R
LM	123.763	4.485	34.720	5.888	5.888	0.854
BFG	45.165	6.165	96.598	7.825	7.492	0.793
RP	11.301	4.904	40.414	6.355	6.355	0.827
SCG	18.951	4.710	37.888	6.154	6.154	0.839
CGB	32.893	9.254	217.297	10.999	9.384	0.655
CGF	41.167	4.753	38.391	6.195	6.196	0.837
CGP	40.326	4.828	39.541	6.287	6.288	0.831
OSS	45.275	7.841	153.670	9.450	7.408	0.730
GDX	8.332	13.652	372.883	15.412	9.937	0.566

Table 4.19: Testing Results of ANN-Based Path Loss Models with Varying Learning Rules

Training Algorithm	MAE (dB)	MSE (dB)	RMSE (dB)	SED (dB)	R
LM	4.473	34.849	5.899	5.900	0.851
BFG	6.155	96.509	7.833	7.503	0.791
RP	4.881	40.299	6.346	6.346	0.825
SCG	4.682	37.748	6.143	6.143	0.837
CGB	9.226	216.792	10.993	9.376	0.653
CGF	4.729	38.339	6.191	6.192	0.835
CGP	4.808	39.500	6.284	6.285	0.829
OSS	7.813	153.166	9.421	7.384	0.729
GDX	13.638	372.194	15.389	9.906	0.564

4.9. ANN-Based Path Loss Models with Varying Number of Hidden Neuron

The number of neuron in the hidden layer was varied between one and fifty in order to determine the optimal number of hidden neuron for path loss predictions in a heterogeneous smart campus environment. It was observed that both prediction accuracy and generalization ability improved with increase in the number of hidden neuron as shown in Figures 4.29-4.31. Optimal path loss prediction accuracy and model generalization were achieved when the number of hidden neuron was set to 43. Further increase in the number of hidden neuron beyond 43 did not significantly increased the prediction accuracy and generalization ability of the ANN-based path loss model. The ANN-based path loss model with 43 hidden neurons produced the least prediction error with MAE, MSE, RMSE, SED, and R values of 3.002 dB, 16.871 dB, 4.107 dB, 4.108 dB, and 0.932 when evaluated using training dataset. On generalization ability, the identified ANN-based path loss model demonstrated the best generalization ability with MAE, MSE, RMSE, SED, and R values of 3.084 dB, 18.255 dB, 4.273 dB, 4.273 dB, and 0.925 when evaluated using testing dataset.

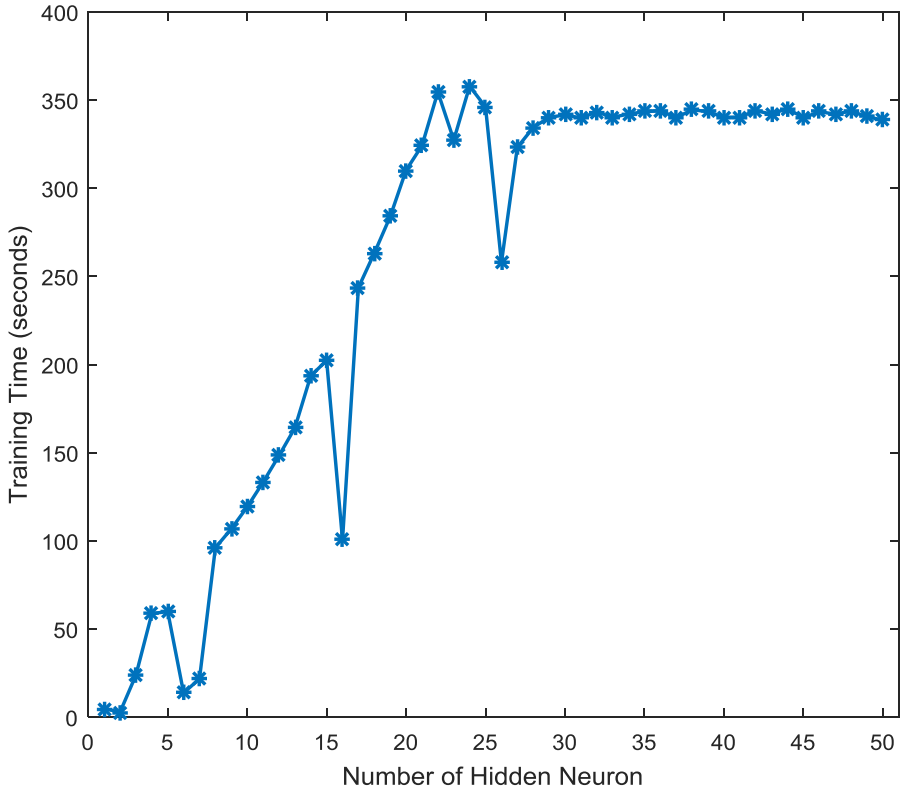

Figure 4.29. Training time versus number of hidden neuron

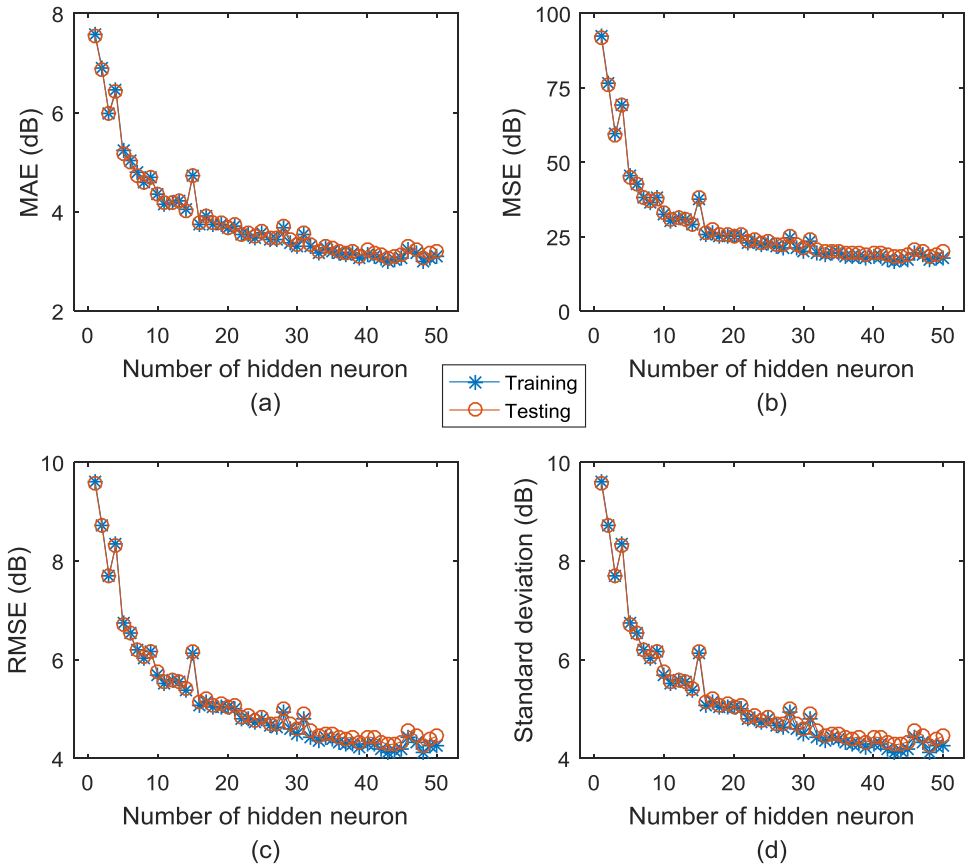

Figure 4.30. Performance evaluation of ANN-based path loss model

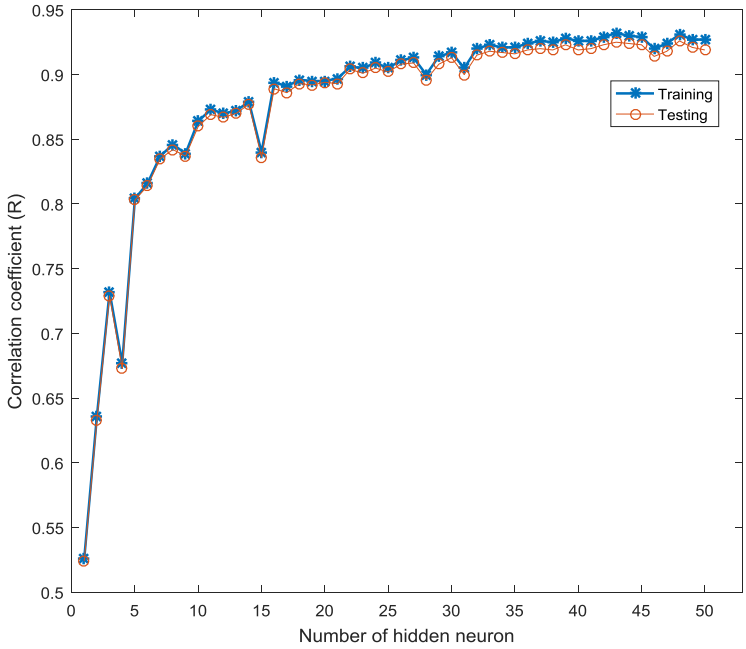

Figure 4.31. Correlation coefficient versus number of hidden neuron

Based on the results of the extensive experimentations that were performed earlier, the optimal ANN-based path loss model was identified. The network configuration of the optimal ANN-based path loss model is shown in Figure 4.29. The model has seven input nodes (each for distance, frequency, clutter height, elevation, altitude, latitude, and longitude respectively), single hidden layer with 43 neurons and *logsig* activation function, and a single output neuron (for path loss variable) with *tansig* activation function. The ANN-based path loss model was trained based on Levenberg-Marquardt learning algorithm and the model yielded prediction outputs with R values of 0.93104, 0.92667, and 0.92323 for training, validation, and testing. When the training dataset was used to evaluate the developed model, an overall R value of 0.92919 was obtained as shown Figure 4.30. MAE, MSE, RMSE, and SED of 3.108 dB, 19.023 dB, 4.340 dB, and 4.339 dB were obtained respectively when the developed model was tested with training dataset. Meanwhile, the developed model yielded good generalization (MAE, MSE, RMSE, and SED of 3.184 dB, 20.214 dB, 4.480 dB, and 4.479 dB respectively) when tested with testing dataset. The input weight, output weight, and bias matrices of the developed ANN-based path loss model are presented in Table 4.20.

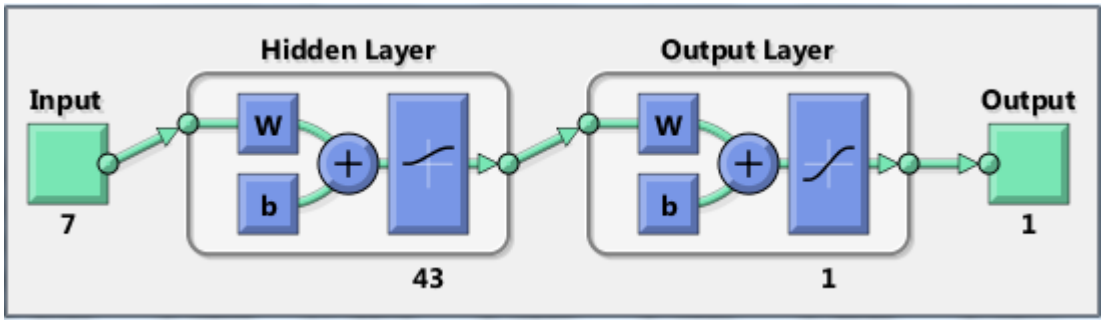

Figure 4.32. Final Training Result

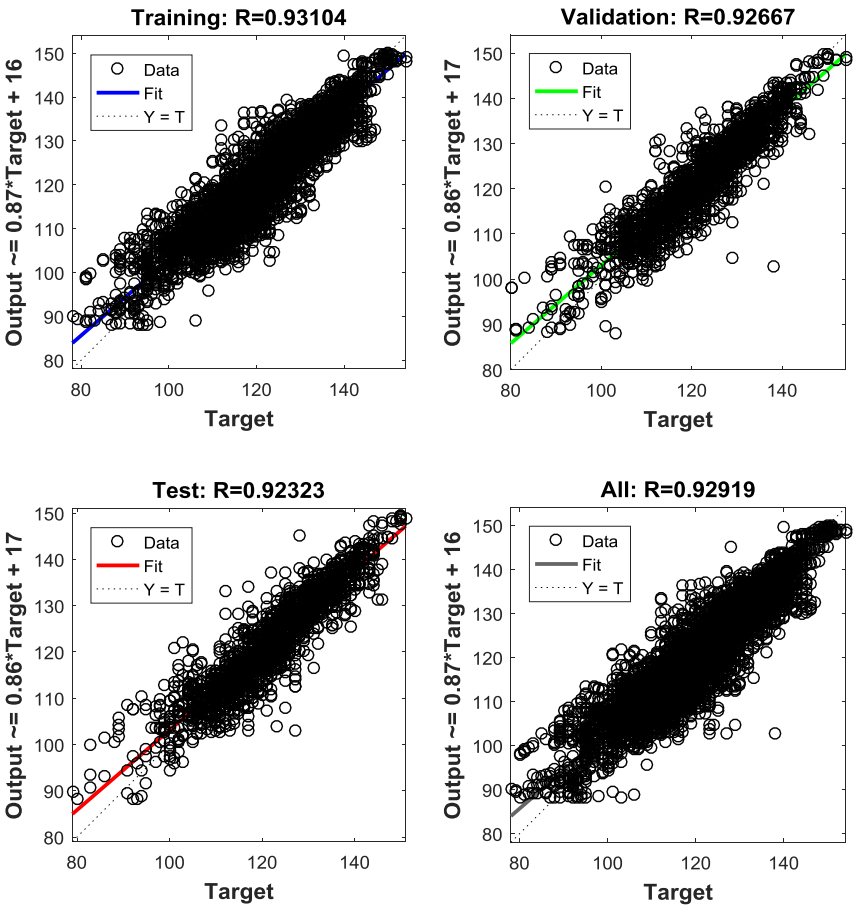

Figure 4.33. Performance evaluation of ANN training

Table 4.20. Input Weight, Output Weight, and Bias Matrices

Hidden Neurons	**Input Weights**							**Output Weight**	**Bias**
	Longitude	*Latitude*	*Elevation*	*Altitude*	*Frequency*	*Clutter Height*	*Distance*	*Path Loss*	
1	-1.1647	1.0144	1.2098	-5.6955	1.2353	-2.9329	6.8904	5.0271	4.6611
2	-6.5746	3.3664	-0.5617	4.4289	-0.7001	-3.8496	16.8524	-2.4043	13.3304
3	3.6368	4.3414	12.6045	-1.5757	-2.1277	-2.6229	-3.0796	-6.8717	14.9232
4	5.8190	3.9993	-1.2422	10.1673	5.9335	8.0071	11.9913	2.9233	11.0021
5	-21.3067	-11.6870	1.3688	9.7959	1.9902	-17.2661	2.9116	0.8231	19.8516
6	-2.5291	-1.7287	-6.0018	-7.4067	5.6940	6.2848	1.3888	5.7389	2.4994
7	-6.1390	-6.3188	4.2439	-5.4567	5.9467	6.8346	-2.4119	1.8685	6.7059
8	4.1317	-3.2941	-3.9665	-3.0668	-2.9972	7.3317	1.8664	3.4749	-1.1231
9	-5.5773	-3.9727	1.6345	-6.3527	2.6811	6.0845	-12.7139	-1.4983	-0.6685
10	8.1197	5.9916	0.2892	24.8493	-18.7230	-14.8042	-23.0398	-0.6706	14.7883
11	-4.0214	-0.3229	-0.0920	3.8044	-4.7344	0.4629	6.2667	-8.5678	-0.0709
12	-2.6721	0.5802	3.0978	-1.3779	21.0825	-15.0793	-9.8069	0.5265	-2.4126
13	-12.3700	-5.7421	-4.1809	7.8168	-2.1296	-14.4388	15.1632	1.1564	1.2257
14	-7.4359	-2.4303	5.7433	-0.1158	-0.6593	2.6127	-3.6018	4.3200	0.7931
15	2.9897	-7.8852	-19.1060	-2.3256	3.0125	-12.3885	10.4222	-1.9206	-11.1400
16	23.0873	0.6791	3.2318	4.4410	-13.5028	7.5166	2.8091	1.3288	-0.5127
17	0.7450	1.4475	3.0738	-10.2353	-1.7295	-5.9026	16.3209	1.9940	2.6844

18	11.8032	11.4836	-7.4273	-21.9591	-2.5165	-4.1633	-7.4386	-1.2352	-14.4191
19	5.7129	-4.3738	6.5946	4.7379	2.2395	-0.3414	-2.9928	6.3395	1.9817
20	-0.9791	3.6767	1.8288	-7.2161	-0.0804	0.0913	1.5361	-3.6730	-1.6694
21	1.0161	1.5787	3.5591	-4.1724	1.8664	9.0980	4.4899	4.8202	-0.2561
22	3.0899	-13.2627	2.9909	-1.2963	4.5792	1.4902	-3.4174	-5.2126	5.3247
23	2.3686	-1.4569	-0.8880	0.1387	2.6346	2.0665	-9.2243	-7.1451	2.9632
24	11.1791	0.4913	-1.7265	12.1077	-12.7133	-3.4139	-1.6619	-1.1898	12.4900
25	-2.3960	-0.0711	1.2742	2.9049	-1.6354	-4.9487	-2.3859	7.1991	2.2771
26	-41.8074	11.6425	-6.5461	-16.6022	0.2866	-3.9688	0.6798	0.6877	-0.1740
27	3.8466	11.7364	4.7355	5.2495	5.7146	-6.6175	-7.5288	2.7356	-1.2506
28	4.4717	4.9343	26.0678	5.5076	19.2066	-28.3311	24.1815	-0.6520	3.3912
29	4.1011	0.3194	0.1809	-5.5614	6.3900	0.3270	-10.4355	-8.1563	-0.3978
30	13.7775	-8.0055	0.2438	5.8668	5.4100	16.2921	-16.5347	-0.8656	2.0533
31	6.7051	3.4076	-0.4362	-4.5378	-1.6563	4.0729	-7.7680	4.0183	6.5374
32	-8.1386	2.2305	-12.6375	-8.5748	-25.6857	-6.7388	-16.1981	0.6241	-22.1927
33	3.0623	-0.7296	0.9199	-8.0281	1.6417	-5.4914	12.3783	-3.2761	2.3962
34	-7.8304	0.6727	-18.4581	3.3249	10.1082	-17.1042	-19.8296	1.6670	16.2587
35	-0.4189	-3.1391	4.4448	3.0473	-0.9398	2.1643	2.5948	-6.6269	6.8230
36	3.4371	-7.1415	3.8139	-1.9433	10.0976	-6.5080	13.6256	-1.7346	5.5802
37	3.8683	-7.2219	4.9167	-4.5420	6.4579	-5.1511	12.7581	2.8410	7.1137
38	14.2854	2.8307	4.2686	-19.2774	11.0680	5.3180	-15.0089	1.9353	2.9050
39	-4.8952	5.9181	2.5305	-1.9711	-0.3778	-0.8254	2.3223	-8.3000	-7.9063

40	-9.6751	-1.4825	0.4304	1.1692	1.1656	8.2648	9.4580	-2.8908	-6.4856
41	2.9944	-10.6110	-9.4666	-1.4103	3.8450	-0.1727	-5.8757	3.2299	-4.7854
42	-4.6585	-0.6060	1.4738	-5.1518	8.5773	-0.5781	12.0575	-4.2554	-9.4979
43	20.0528	1.8272	0.9092	7.0742	6.2465	15.2742	18.2011	-9.9760	21.5526

4.10. SVM-Based Model for Path Loss Predictions

The results of the attribute selection performed using 10-fold cross validation show that all the independent variables (i.e. longitude, longitude, latitude, elevation, altitude, frequency, clutter height, and distance) have equal influence on the dependent variable (i.e. path loss). Detailed information about these results is presented in Table 4.21. A SVM-based path loss model was developed using SMOreg regression algorithm. The model was trained using 10-fold cross validation technique instead of dataset splitting approach. The parameters of the developed SVM-based path loss model are provided in Table 4.22. The time taken to build the SVM-based path loss model and the number of kernel evaluations are 694.74 seconds and -2090011411 respectively. The developed SVM-based path loss model accepts normalized input data and the corresponding normalized path loss values may be obtained by Equation 4.1. The obtained path loss values may be denormalized using Equation 4.2. Therefore, the real path loss values in dB can be computed using Equation 4.3. The final complete mathematical representation of the SVM-based path loss model is given by Equation 4.4.

Table 4.21. Attribute selection using 10-fold cross validation

S|N	Attributes	Number of folds (%)
1	Longitude	10 (100 %)
2	Latitude	10 (100 %)
3	Elevation	10 (100 %)
4	Altitude	10 (100 %)
5	Frequency	10 (100 %)
6	Clutter height	10 (100 %)
7	Distance	10 (100 %)

Attribute Subset Evaluator: CFS Subset Evaluator
Search Method: Greedy Stepwise (forwards)
Selected Attributes: Longitude, longitude, latitude, elevation, altitude, frequency, clutter height, and distance

Table 4.22. SVM model parameters

S|N	Attributes	Weights
1	Constant	+0.4487
2	Longitude (*normalized*)	+0.1939
3	Latitude (*normalized*)	-0.2789
4	Elevation (*normalized*)	-0.2531
5	Altitude (normalized)	+0.4655
6	Frequency (*normalized*)	+0.1207
7	Clutter height (*normalized*)	-0.0149
8	Distance (*normalized*)	+0.1413

Number of kernel evaluations: -2090011411

Time taken to build model: 694.74 seconds

$$PL_{SVM,norm} = 0.4361 + (0.1939 \times longitude_{norm}) - (0.2789 \times latitude_{norm}) - (0.2531 \times elevation_{norm}) + (0.4655 \times altitude_{norm}) + (0.1207 \times frequency_{norm}) - (0.0149 \times clutter\ height_{norm}) + (0.1413 \times distance_{norm}) \quad (4.1)$$

$$PL_{SVM}\ (dB) = PL_{measured,min} + \left[\left(\frac{PL_{SVM,norm} - PL_{SVM,min}}{PL_{measured,max} - PL_{measured,min}}\right) \times \left(PL_{measured,max} - PL_{measured,min}\right)\right] \quad (4.2)$$

Where,

$PL_{measured,min} = 78\ dB$

$PL_{measured,min} = 154\ dB$

$PL_{SVM,min} = 0$

$PL_{SVM,max} = 1$

$$PL_{SVM}\ (dB) = 78 + (76 \times PL_{SVM,norm})$$

(4.3)

$$PL_{SVM,}(dB) = 111.14 + (14.7364 \times longitude_{norm}) - (21.1964 \times latitude_{norm}) - (19.2356 \times elevation_{norm}) + (35.378 \times altitude_{norm}) + (9.1732 \times frequency_{norm}) - (1.1324 \times clutter\ height_{norm}) + (10.7388 \times distance_{norm})$$

(4.4)

4.11. Results of Statistical Evaluation of Empirical, ANN, and SVM Path Loss Models

The prediction outputs of the developed ANN-based model, SVM-based model, and popular empirical models (i.e. Okumura-Hata, COST 231, ECC-33, and Egli) were compared to the measured path loss values in both training and testing datasets to evaluate the prediction accuracy and generalization ability of the path loss models. The outputs of the models for path loss predictions at 900, 1800, 2100 MHz relative to the measured path loss values in both training and testing datasets are graphically represented in Figures 4.34-4.39 respectively.

The results of the prediction accuracy evaluation of the models are presented in Table 4.23. ANN-based path loss model produced the least prediction error with MAE, MSE, RMSE, SED and R values of 3.108 dB, 19.023 dB, 4.340 dB, 4.339 dB, and 0.923 respectively when compared to the measured path loss values in training dataset. The prediction error produced by SVM-based path loss model (MAE, MSE, RMSE, SED and R values of 7.953 dB, 99.966 dB, 9.998 dB, 9.940 dB, and 0.478 respectively) is much relatively lower than those of all the empirical models. Egli model produced the highest prediction error with MAE, MSE, RMSE, SED and R values of 27.000 dB, 969.657 dB, 31.139 dB, 16.384 dB, and 0.266 respectively when compared to the measured path loss values in training dataset.

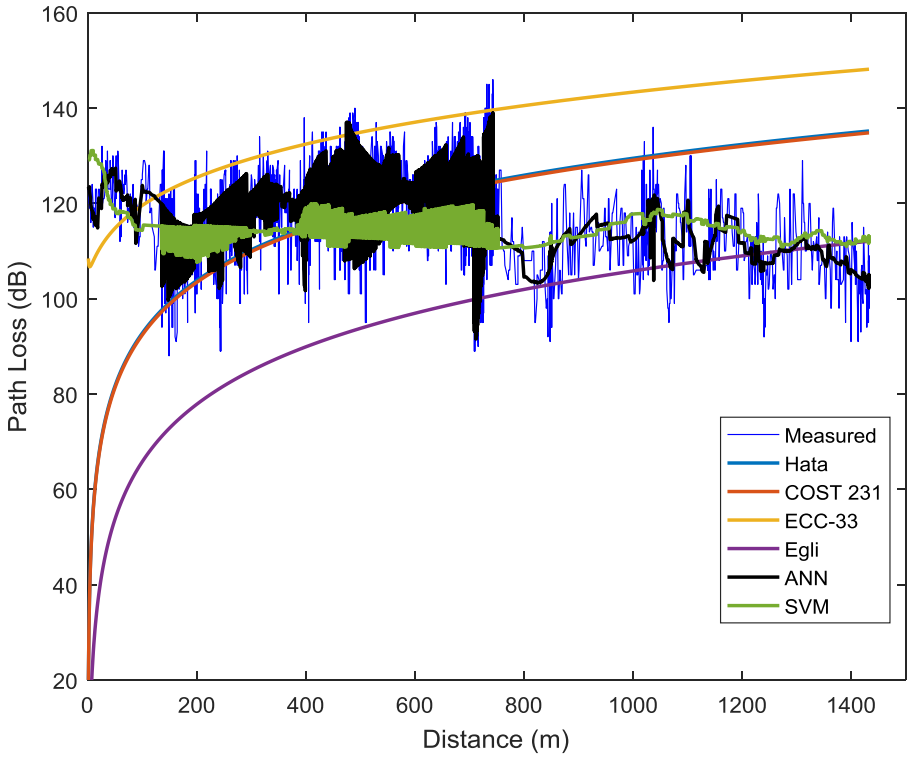

Figure 4.34. Training (900 MHz)

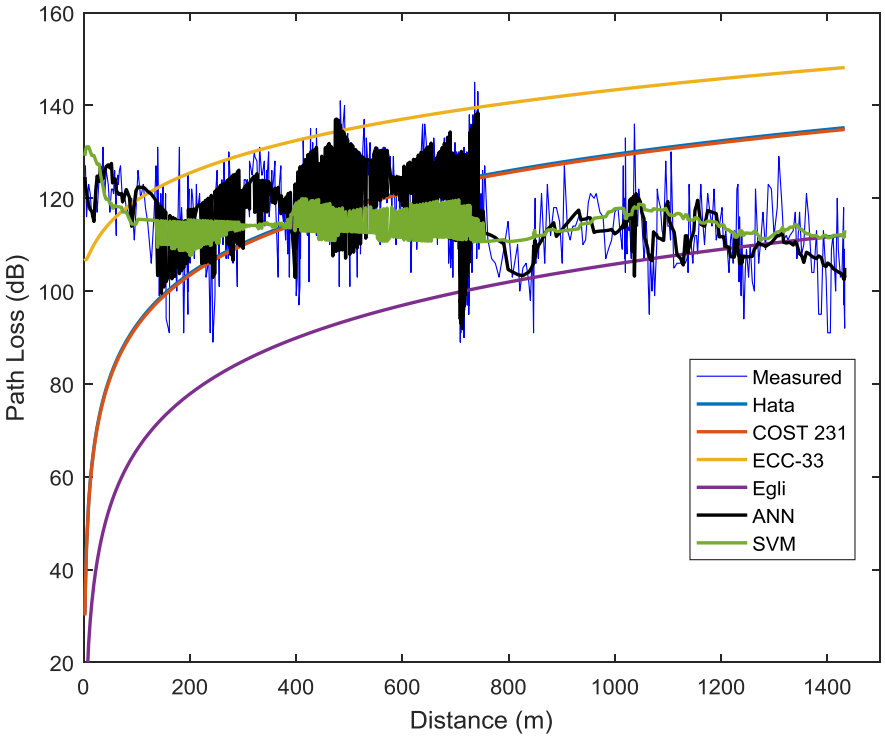

Figure 4.35. Testing (900 MHz)

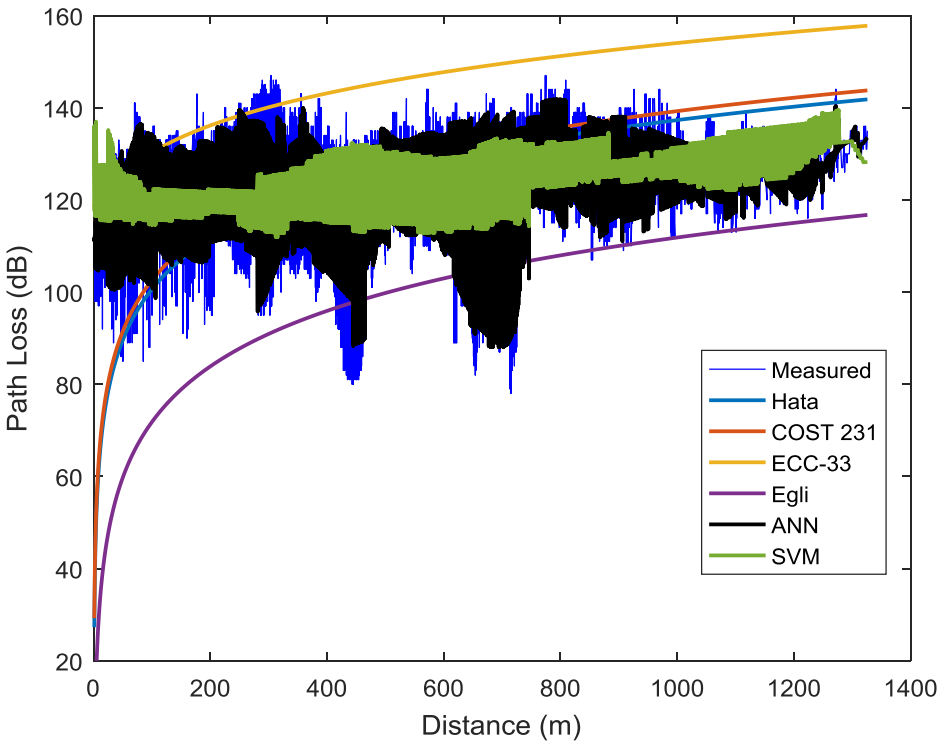

Figure 4.36. Training (1800 MHz)

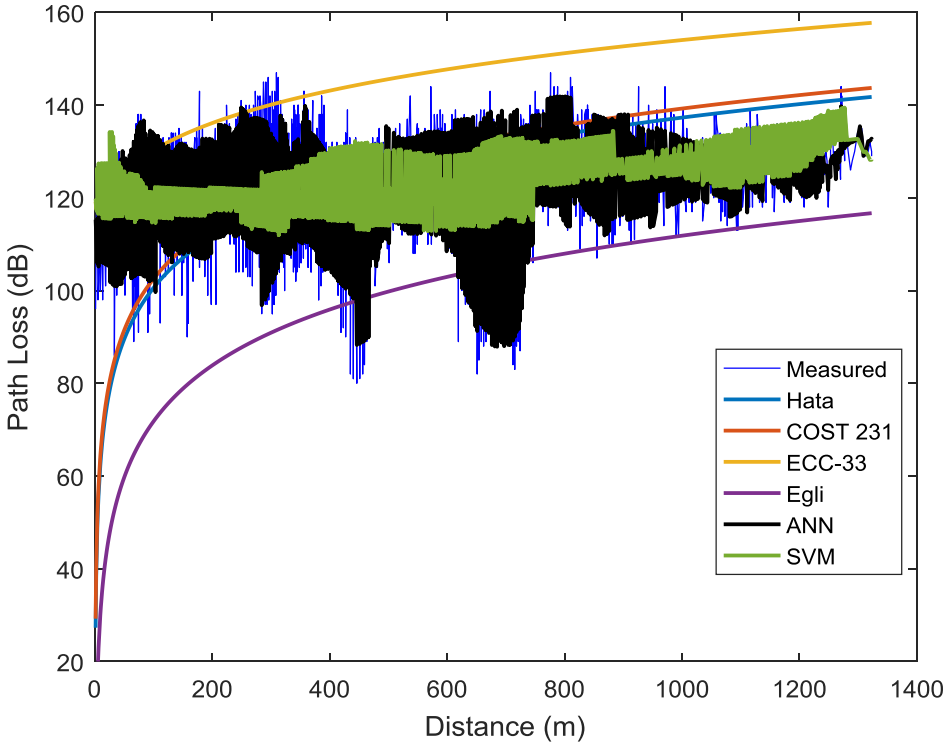

Figure 4.37. Testing (1800 MHz)

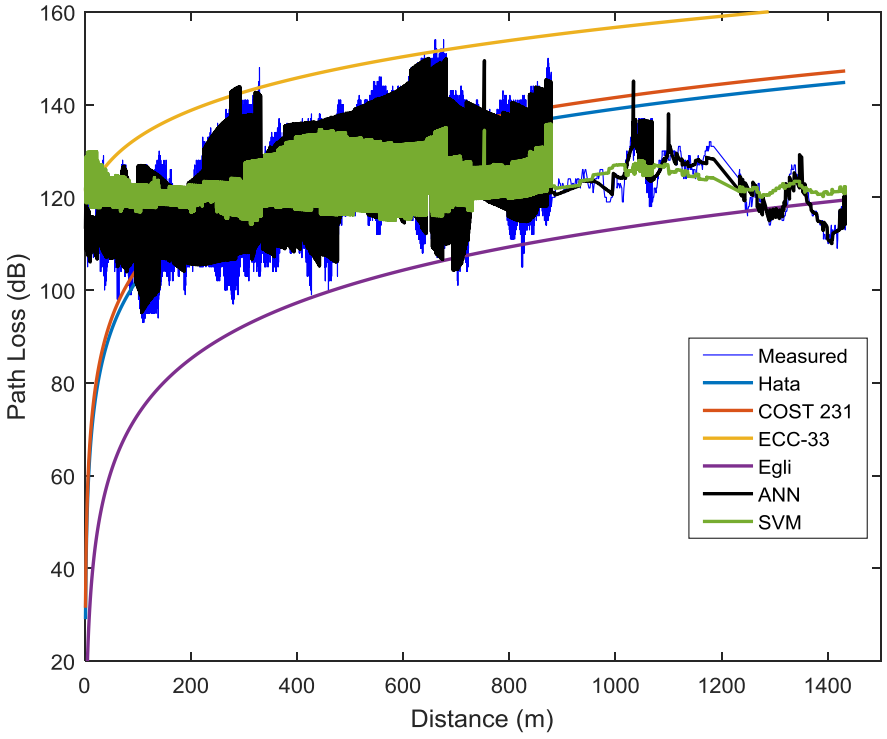

Figure 4.38. Training (2100 MHz)

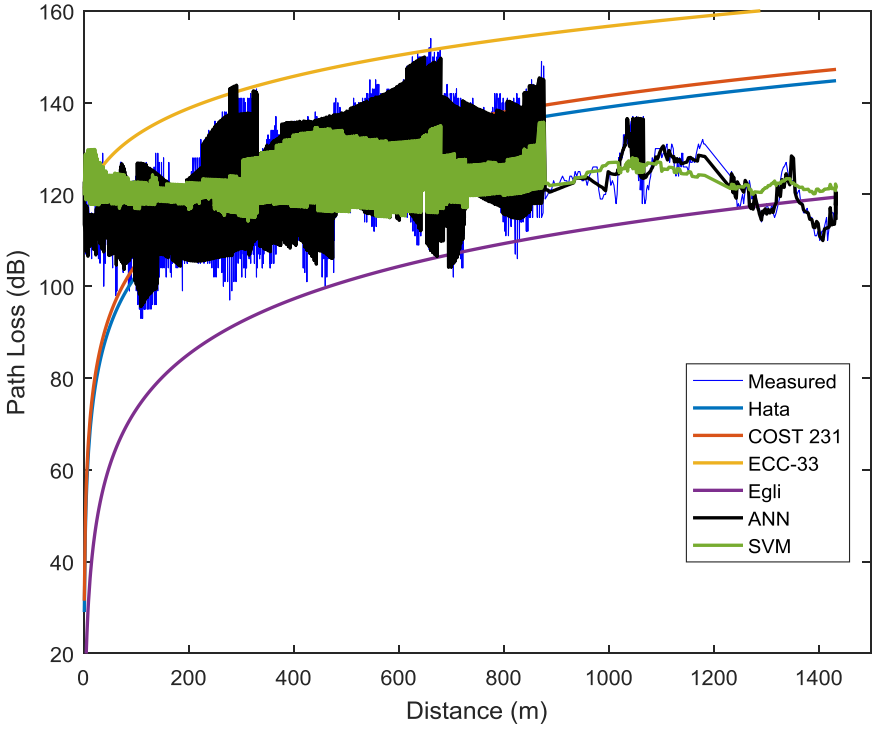

Figure 4.39. Testing (2100 MHz)

On generalization ability, the ANN-based path loss model yielded the best generalization with MAE, MSE, RMSE, SED and R values of 3.184 dB, 20.214 dB, 4.480 dB, 4.479 dB, and 0.917

respectively when compared to the measured path loss values in testing dataset. The generalization ability demonstrated by SVM-based path loss model (MAE, MSE, RMSE, SED and R values of 7.933 dB, 98.773 dB, 9.938 dB, 9.878 dB, and 0.478 respectively) is much relatively better than those of all the empirical models. Egli model demonstrated the least generalization ability with MAE, MSE, RMSE, SED and R values of 27.044 dB, 974.318 dB, 31.214 dB, 16.429 dB, and 0.266 respectively when compared to the measured path loss values in testing dataset. The regression plots showing the relationships between the predicted outputs and the measured path loss values in testing dataset are depicted in Figures 4.40-4.45.

Table 4.23. Final Training Results

	MAE (dB)	**MSE (dB)**	**RMSE (dB)**	**SED (dB)**	**R**
Okumura-Hata	11.510	236.930	15.393	15.391	0.279
COST 231	11.778	241.055	15.526	15.374	0.290
ECC-33	21.884	609.750	24.693	11.948	0.320
Egli	27.000	969.657	31.139	16.384	0.266
ANN	3.108	19.023	4.340	4.339	0.923
SVM	7.953	99.966	9.998	9.940	0.478

Table 4.24. Final Testing Results

	MAE (dB)	**MSE (dB)**	**RMSE (dB)**	**SED (dB)**	**R**
Okumura-Hata	11.507	237.888	15.424	15.424	0.279
COST 231	11.765	241.847	15.551	15.409	0.290
ECC-33	21.831	607.061	24.639	11.896	0.320
Egli	27.044	974.318	31.214	16.429	0.266
ANN	3.184	20.214	4.480	4.479	0.917
SVM	7.933	98.773	9.938	9.878	0.478

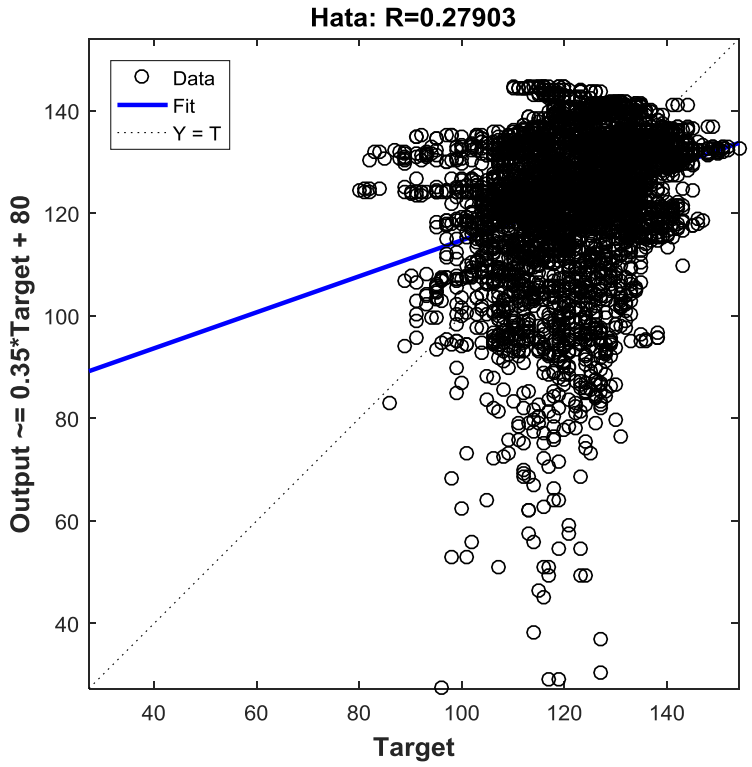

Figure 4.40. Testing results for Hata model

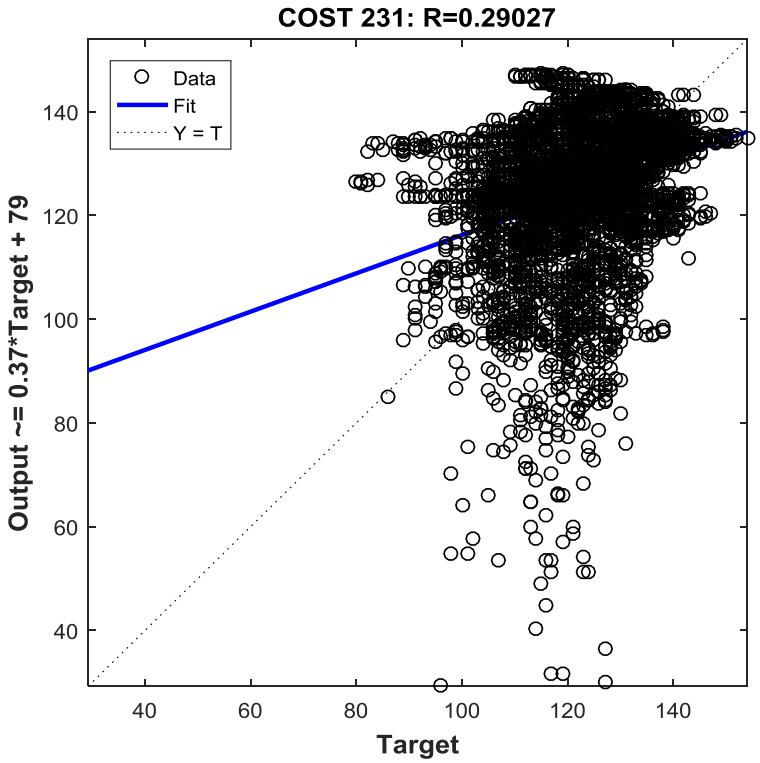

Figure 4.41. Testing results for COST 231 model

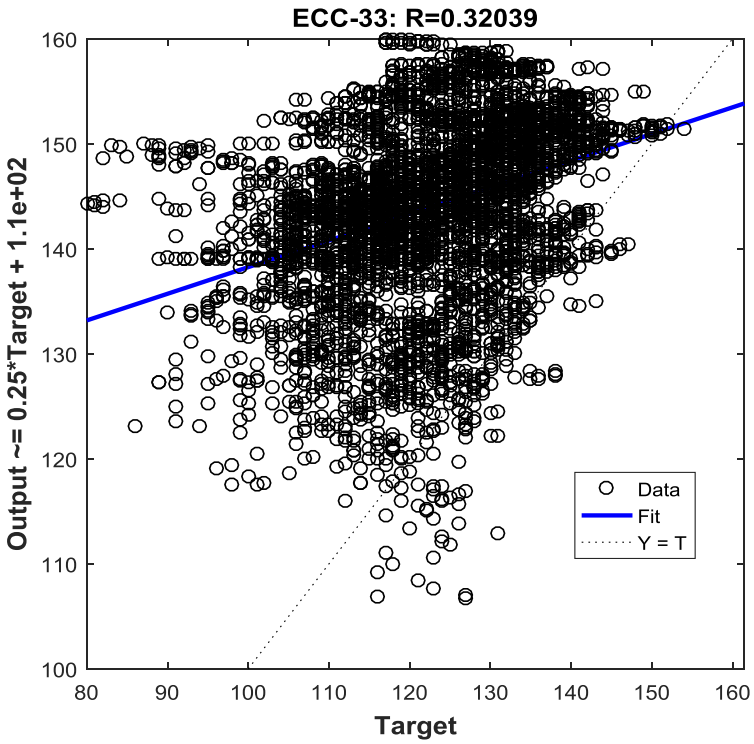

Figure 4.42. Testing results for ECC-33 model

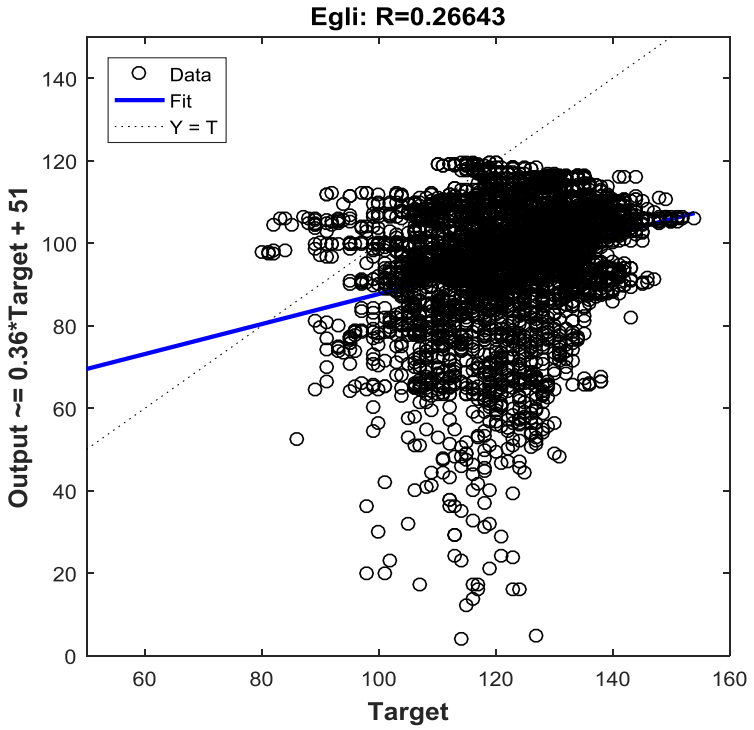

Figure 4.43. Testing results for Egli model

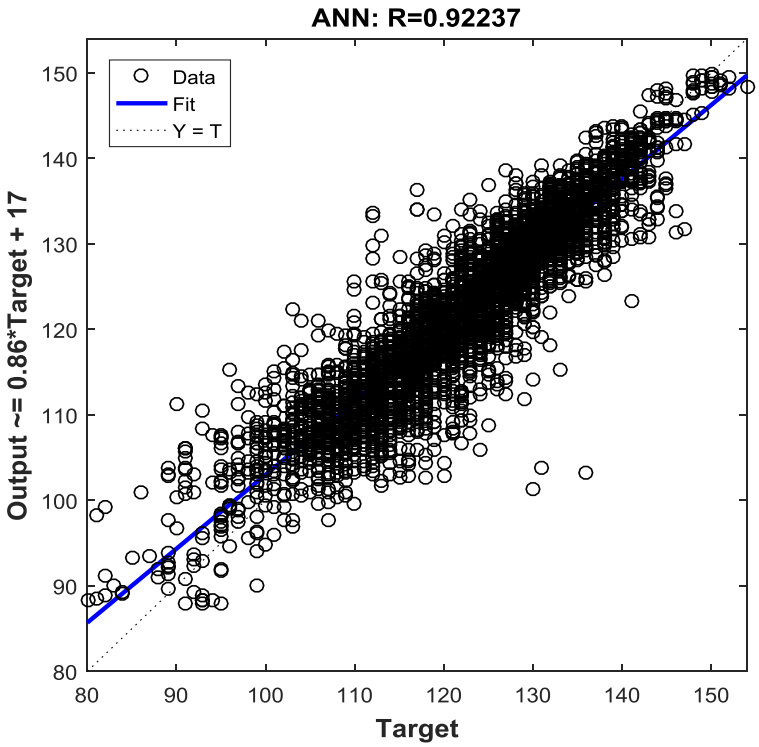

Figure 4.44. Testing results for ANN model

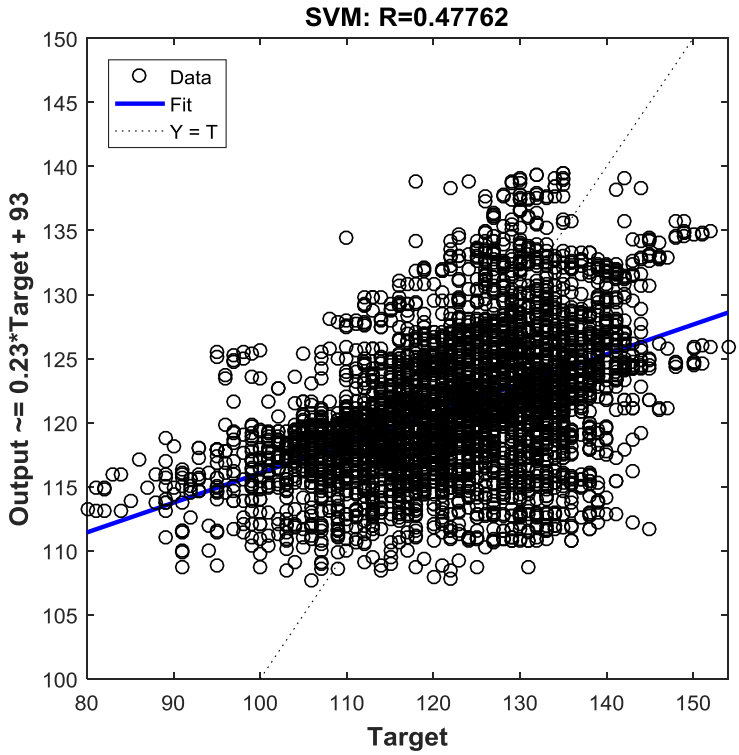

Figure 4.45. Testing Results for SVM model

Prediction outputs of the models were further subjected to statistical tests to establish whether the differences in the prediction outputs and the measured path loss values are statistically significant or not. Table 2.25 shows that there are significant differences ($p \leq 0.05$) in the prediction outputs and the measured path loss values in training dataset. Similarly, Table 2.26 shows that there are significant differences ($p \leq 0.05$) in the prediction outputs and the measured path loss values in testing dataset. The results of multiple comparison post-hoc tests performed on both training and testing datasets are presented in Table 4.25 and Table 4.26 respectively. On training dataset, the prediction outputs of COST 231, ECC-33, Egli, and SVM-based path loss models significantly differ from the measured path loss values. Meanwhile, the prediction outputs of Okumura-Hata and ANN-based path loss models (p-values of 0.9745 and 1.0000 respectively) do not significantly differ from the measured path loss values in training dataset. On testing dataset, the prediction outputs of COST 231, ECC-33, Egli, and SVM-based path loss models significantly differ from the measured path loss values. However, the prediction outputs of Okumura-Hata and ANN-based path loss models (p-values of 0.4472 and 1.0000 respectively) do not significantly differ from the measured path loss values in testing dataset.

Table 4.25. ANOVA results of path loss predictions using training data

Source of Variation	**Sum of Squares**	**Degree of Freedom**	**Mean Squares**	**F Statistic**	**Prob>F**
Columns	1.6581×10^7	6	2.763×10^6	1.975×10^4	0.0000
Error	1.3848×10^7	98987	139.895		
Total	3.0429×10^7	98993			

Table 4.26. ANOVA results of path loss predictions using testing data

Source of Variation	**Sum of Squares**	**Degree of Freedom**	**Mean Squares**	**F Statistic**	**Prob>F**
Columns	5.5330×10^6	6	9.2216×10^5	6.543×10^3	0.0000
Error	4.6499×10^6	32991	140.9		
Total	1.0183×10^7	32997			

Table 4.27. Multiple comparison post-hoc test results of training data

Groups Compared		Lower limits for 95% confidence intervals	Mean Difference	Upper limits for 95% confidence intervals	*p*-value
Measured	Hata	-0.9380	-0.2170	0.5039	0.9745
Measured	COST 231	-2.8329	-2.1119	-1.3909	3.71×10^{-8}
Measured	ECC-33	-22.2982	-21.5772	-20.8562	3.71×10^{-8}
Measured	Egli	25.8206	26.5416	27.2625	3.71×10^{-8}
Measured	ANN	-0.6980	0.0230	0.7440	1.0000
Measured	SVM	0.3791	1.1001	1.8210	0.0001
Hata	COST 231	-2.6159	-1.8949	-1.1739	3.71×10^{-8}
Hata	ECC-33	-22.0812	-21.3602	-20.6392	3.71×10^{-8}
Hata	Egli	26.0376	26.7586	27.4796	3.71×10^{-8}
Hata	ANN	-0.4809	0.2400	0.9610	0.9581
Hata	SVM	0.5961	1.3171	2.0381	1.53×10^{-6}
COST 231	ECC-33	-20.1863	-19.4653	-18.7443	3.71×10^{-8}
COST 231	Egli	27.9325	28.6535	29.3744	3.71×10^{-8}
COST 231	ANN	1.4139	2.1349	2.8559	3.71×10^{-8}
COST 231	SVM	2.4910	3.2120	3.9329	3.71×10^{-8}
ECC-33	Egli	47.3978	48.1188	48.8398	3.71×10^{-8}
ECC-33	ANN	20.8793	21.6002	22.3212	3.71×10^{-8}
ECC-33	SVM	21.9563	22.6773	23.3983	3.71×10^{-8}
Egli	ANN	-27.2395	-26.5186	-25.7976	3.71×10^{-8}
Egli	SVM	-26.1625	-25.4415	-24.7205	3.71×10^{-8}
ANN	SVM	0.3561	1.0771	1.7980	0.0002

Table 4.28. Multiple comparison post-hoc test results of testing data

Groups Compared		Lower limits for 95% confidence intervals	Mean Difference	Upper limits for 95% confidence intervals	*p*-value
Measured	Hata	-0.6889	-0.2742	0.1405	0.4472
Measured	COST 231	-2.5833	-2.1686	-1.7539	3.71×10^{-8}
Measured	ECC-33	-22.0250	-21.6103	-21.1956	3.71×10^{-8}
Measured	Egli	26.0662	26.4809	26.8956	3.71×10^{-8}
Measured	ANN	-0.4358	-0.0211	0.3936	1.0000
Measured	SVM	0.6643	1.0790	1.4937	3.71×10^{-8}
Hata	COST 231	-2.3091	-1.8944	-1.4797	3.71×10^{-8}
Hata	ECC-33	-21.7507	-21.3360	-20.9213	3.71×10^{-8}
Hata	Egli	26.3404	26.7551	27.1698	3.71×10^{-8}
Hata	ANN	-0.1616	0.2531	0.6678	0.5483
Hata	SVM	0.9385	1.3532	1.7679	1.53×10^{-6}
COST 231	ECC-33	-19.8563	-19.4416	-19.0269	3.71×10^{-8}
COST 231	Egli	28.2348	28.6495	29.0642	3.71×10^{-8}
COST 231	ANN	1.7328	2.1475	2.5622	3.71×10^{-8}
COST 231	SVM	2.8329	3.2476	3.6623	3.71×10^{-8}
ECC-33	Egli	47.6764	48.0911	48.5058	3.71×10^{-8}
ECC-33	ANN	21.1745	21.5892	22.0039	3.71×10^{-8}
ECC-33	SVM	22.2745	22.6892	23.1039	3.71×10^{-8}
Egli	ANN	-26.9167	-26.5020	-26.0873	3.71×10^{-8}
Egli	SVM	-25.8166	-25.4019	-24.9872	3.71×10^{-8}
ANN	SVM	0.6854	1.1001	1.5148	3.71×10^{-8}

CHAPTER FIVE

DISCUSSIONS

5.1. Introduction

In this research project, an optimal path loss prediction model was developed for heterogeneous radio network planning, deployment, and optimization in a smart campus propagation environment based on ANN and SVM learning algorithms. Radio signal measurements were conducted to obtain the strengths of radio signals received at varying separation distances between fourteen base station transmitters and two mobile receivers. Geographic and network information (i.e. longitude, latitude, elevation, altitude, frequency, clutter height, and RSS) recorded were stored in a local database. These data were further processed to remove duplicates and extraneous data points. Machine learning-based path loss models were developed using ANN and SVM techniques. Finally, prediction accuracy and generalization ability of popular empirical models (Hata, COST 231, ECC-33, and Egli), ANN-based models, and SVM models were evaluated using MAE, MSE, RMSE, SED, R, ANOVA, and multiple comparison post-hoc test.

5.2. Optimum Input Parameters for ANN-Based Path Loss Model Development

The choice of the right kind of input parameters for the design and development of machine learning-based path loss model often determines the prediction accuracy and generalization ability of the developed model. The solution to this optimization problem is critical to the realization of an efficient and reliable machine learning-based path loss model (Sotirios P Sotiroudis, Sotirios K Goudos, Konstantinos A Gotsis, Katherine Siakavara, & John N Sahalos, 2013).

In this work, six different ANN-based path loss models were developed to determine the optimum kinds and number of input parameters that will guarantee sufficiently high prediction accuracy and good generalization ability. The prediction accuracy and generalization ability of ANN-based path loss model increased as the number of input variables (i.e. longitude, latitude, elevation, altitude, frequency, clutter height, and distance) increased from one to seven. Conversely, the time required to train the ANN-based path loss models increased as the number of input variables increased from one to seven.

Ostlin et al. (2010) explained that the magnitude of path loss in a typical rural macrocell propagation environment can be correctly determined by the following parameters: (a) the separation distance between BTS antenna and Mobile Station (MS) antenna; (b) the height of the BTS antenna (Ostlin, Suzuki, & Zepernick, 2005); (c) the terrain profile of the space between BTS antenna and MS antenna i.e. terrain clearance angle, terrain usage, and the type of vegetation (Ostlin, Zepernick, & Suzuki, 2003); and (d) the vegetation density in the surroundings of the MS antenna (Östlin, Suzuki, & Zepernick, 2008). In a similar work, Eichie, Oyedum, Ajewole, and Aibinu (2017a) identified relative humidity, temperature, and dew point as appropriate inputs for the prediction of RSS using ANN. S. Sotiroudis et al. (2013) developed an optimal ANN model for path loss predictions in urban environments by supplying "map-specific information", "measurement-specific information", Cartesian coordinates of the BTS and the MS, and the separation distance between BTS and MS as input variables.

Ileana Popescu et al. (2005) developed some ANN models with different combinations of the information about the propagation environment as input parameters. These input information include separation distance between BTS and MS, street width, building height, separation distance between surrounding buildings, the location of BTS antenna relative to the rooftop, and street orientation (I. Popescu, Nafornita, Constantinou, Kanatas, & Moraitis, 2001). The ANN model with six input parameters produced the best prediction accuracy and generalization (Ileana Popescu et al., 2002). The feature vector of the SVM-based path loss model developed by Timoteo, Cunha, and Cavalcanti (2014) consists of separation distance between BTS and MS, terrain elevation, horizontal and vertical angles, latitude, longitude, as well as the horizontal and vertical attenuation of the BTS antenna.

5.3. Effect of Input Data Normalization on Model Performance

In this work, extensive experimentations were performed to validate the effect of input data normalization on path loss prediction accuracy and generalization ability. It was observed that input data normalization significantly improved path loss prediction accuracy and it equally produced better generalization. However, the input data normalization process further increased the training time in the development of the ANN-based models.

5.4. Optimal Transfer Function for ANN-Based Path Loss Model

Nine ANN-based path loss models were developed to determine the most suitable combination of transfer/activation functions for heterogeneous network planning and optimization in a smart campus propagation environment. The use of logsig and tansig activation functions at the

hidden and output layers of the neural network produced the best prediction accuracy and generalization ability. However, the ANN model requires relatively much training time.

5.5. Optimal Learning Algorithm for ANN-Based Path Loss Model

Nine ANN-based path loss models were developed to determine the best learning algorithm for path loss predictions in a heterogeneous smart campus environment. Training the ANN-based path loss model using Levenberg-Marquardt learning algorithm produced the best prediction accuracy and generalization ability. Conversely, Variable Learning Rate Backpropagation learning algorithm required the least time of training.

5.6. Optimal Number of Hidden Neurons for ANN-Based Path Loss Model

Kecman (2001) proposed that the number of hidden neurons may be set either as 75-100 % of the input neurons or as square root of the product of input neurons and output neurons. Zineb and Ayadi (2016) reported that the model performance does not change significantly when the number of hidden neurons was varied in accordance to the recommendations by Kecman (2001). In this work, the number of neuron in the hidden layer was varied between one and fifty in order to determine the optimal number of hidden neuron for path loss predictions in a heterogeneous smart campus environment. It was observed that both prediction accuracy and generalization ability improved with increase in the number of hidden neuron. Optimal path loss prediction accuracy and model generalization were achieved when the number of hidden neuron was set to 43. Further increase in the number of hidden neuron beyond 43 did not significantly increased the prediction accuracy and generalization ability of the ANN-based path loss model.

5.7. Optimal Multi-Frequency Machine Learning-Based Path Loss Model

The optimal ANN-based path loss model was identified based on the results of the extensive experimentations that were performed earlier. The ANN-based path loss model has seven input nodes (each for distance, frequency, clutter height, elevation, altitude, latitude, and longitude respectively), single hidden layer with 43 neurons and logsig activation function, and a single output neuron (for path loss variable) with tansig activation function. The ANN-based path loss model was trained based on Levenberg-Marquardt learning algorithm and the model yielded the overall best prediction accuracy and generalization ability.

The prediction outputs of the developed ANN-based model, SVM-based model, and popular empirical models (i.e. Okumura-Hata, COST 231, ECC-33, and Egli) were compared to the

measured path loss values in both training and testing datasets to evaluate the prediction accuracy and generalization ability of the path loss models. ANN-based path loss model produced the least prediction error with MAE, MSE, RMSE, SED and R values of 3.108 dB, 19.023 dB, 4.340 dB, 4.339 dB, and 0.923 respectively when compared to the measured path loss values in training dataset. The prediction error produced by SVM-based path loss model (MAE, MSE, RMSE, SED and R values of 7.953 dB, 99.966 dB, 9.998 dB, 9.940 dB, and 0.478 respectively) is much relatively lower than those of all the empirical models. Egli model produced the highest prediction error with MAE, MSE, RMSE, SED and R values of 27.000 dB, 969.657 dB, 31.139 dB, 16.384 dB, and 0.266 respectively when compared to the measured path loss values in training dataset.

Finally, the prediction outputs of the models were further subjected to statistical tests to establish whether the differences in the prediction outputs and the measured path loss values are statistically significant or not. The results obtained show that there are significant differences ($p \leq 0.05$) in the prediction outputs and the measured path loss values in both training and testing datasets. On training dataset, the prediction outputs of COST 231, ECC-33, Egli, and SVM-based path loss models significantly differ from the measured path loss values. Meanwhile, the prediction outputs of Okumura-Hata and ANN-based path loss models (p-values of 0.9745 and 1.0000 respectively) do not significantly differ from the measured path loss values in training dataset. On testing dataset, the prediction outputs of COST 231, ECC-33, Egli, and SVM-based path loss models significantly differ from the measured path loss values. However, the prediction outputs of Okumura-Hata and ANN-based path loss models (p-values of 0.4472 and 1.0000 respectively) do not significantly differ from the measured path loss values in testing dataset. In essence, ANN-based path loss model was found to be the optimal model for heterogeneous radio network planning, deployment, and optimization in a smart campus propagation environment.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

A simple but accurate multi-frequency path loss model is a necessary tool for heterogeneous radio network planning and optimization towards achieving a smart campus. Although deterministic models produce better performance in path loss predictions than empirical models, their use often require detailed knowledge about the propagation environment and more computational resources are needed to process these information. On the other hand, empirical models are easy to implement with less computational requirements in terms of time and cost. Empirical models are not as accurate as deterministic models because they do not effectively account for the unique geographical configurations of the propagation environment. Meanwhile, the reliability of the radio access network depends on the accuracy of the propagation model employed. Hence, the need for significant improvement in the prediction accuracy of empirical models while maintaining model simplicity and ease of use. The learning ability in artificial intelligence may be exploited to reduce computational complexity and to improve prediction accuracy.

In this research project, an optimal heterogeneous model was developed for path loss predictions in a typical university campus propagation environment using machine learning approach. First, extensive field measurement campaigns were conducted to obtain RSS values and path loss values at varying longitude, latitude, altitude, elevation, clutter height, distance, and available radio frequencies (900, 1800, and 2100 MHz) within the campus of Covenant University, Ota, Nigeria. Secondly, ANN-based path loss models were developed for heterogeneous network planning and optimization by: changing the transfer functions at the hidden and output layers, and the learning algorithm employed; varying the number of input variable requirement, and the number of hidden neuron; and determining the effect of input data normalization on prediction accuracy and generalization ability. Thirdly, SVM-based path loss model was developed for heterogeneous radio network planning and optimization using SMOreg algorithm. Finally, the prediction accuracy and generalization ability of Hata, COST 231, ECC-33, Egli, ANN-based, and SVM-based path loss models was evaluated such that the most suitable path loss prediction model for heterogeneous radio network planning, deployment, and optimization in a smart campus propagation environment was identified based

on seven statistical metrics and tests namely: MAE, MSE, RMSE, SED, R, ANOVA, and multiple comparison post-hoc test.

The summary of the main findings of this study is as follows:

(a) The prediction accuracy and generalization ability of ANN-based path loss model improved significantly as more input variables were added.

(b) Input data normalization significantly improved the performance of ANN-based path loss model but this process also increased the training time.

(c) The choice of *logsig* and *tansig* transfer functions at the hidden and output layers of the neural network produced the best model prediction accuracy and generalization ability but the use of these functions relatively took more training time when compared to other combinations of transfer functions.

(d) Training the ANN-based path loss model using LM learning algorithm produced the best prediction accuracy and generalization but it consumed more time in training. Conversely, the use of GDX learning algorithm offered a faster training process but the model performance was poor.

(e) It was observed that the prediction accuracy and generalization ability of ANN-based path loss model improved significantly as the number of hidden neuron increased. Further increase of the number of hidden neurons beyond 43 did not yield any significant improvement in the model performance.

The findings of this study showed that the prediction accuracy and generalization ability of the ANN-based model, which has seven input nodes (distance, frequency, clutter height, elevation, altitude, latitude, and longitude), single hidden layer with 43 neurons and *logsig* activation function, and a single output neuron (for path loss variable) with *tansig* activation function, was found to be the best when compared to the prediction outputs of SVM-based model, and popular empirical models (i.e. Okumura-Hata, COST 231, ECC-33, and Egli). The ANN-based path loss model was trained based on LM learning algorithm. The prediction outputs of the ANN-based path loss model has the lowest Root Mean Square Error (RMSE) of 4.480 dB, Standard Error Deviation (SED) of 4.479 dB, and the highest value of correlation coefficient (R) of 0.917, relative to the measured path loss values. This finding was further validated by the results of Analysis of Variance (ANOVA) and multiple comparison post-hoc tests.

In conclusion, ANN-based path loss model was found to be the optimal model for heterogeneous radio network planning, deployment, and optimization in a smart campus propagation environment.

6.2. Recommendations for Future Work

The following recommendations are made for future work:

(a) In order to improve the reliability of the radio signal measurement process, a Continuous Wave (CW) measurement procedure can be employed instead of the drive test approach. Control of radio network parameters is limited in drive test approach because it involves the use of commercial (already deployed) BTS. CW approach gives researchers adequate control over the choice and selection of radio network parameters since independent transmitter(s) and receiver(s) are used. Likewise, the scope of this work can be extended to include current and emerging wireless technologies such as LTE, LTE-Advanced, and 5G radio networks.

(b) The training speed, accuracy, and generalization ability of path loss predictions may be further improved using other machine learning techniques such as Extreme Learning Machine (ELM) and Adaptive Neuro-Fuzzy Inference System (ANFIS).

(c) Different optimization algorithms such as Genetic Algorithm (GA) and adaptive DE algorithms may also be explored to further improve the computational cost efficiency in path loss predictions.

(d) The developed optimal machine learning-based path loss model can be implemented in the form of user-friendly mobile and/or web applications for the use of radio engineers in Nigerian telecommunication industry.

6.3. Contribution to Knowledge

The findings of this research study contributed to the body of scientific knowledge in the following ways:

a) A robust radio signal measurement and local terrain profile datasets that represents a typical smart campus propagation environment in Nigeria was acquired. These datasets are vital for research reproducibility in the fields of wireless communications and machine learning towards the planning and deployment of optimal heterogeneous radio networks in university campuses (Popoola, Atayero, Arausi, & Matthews, 2018;

Popoola, Atayero, & Faruk, 2018; Popoola, Atayero, Faruk, & Badejo, 2018; Popoola, Atayero, & Popoola, 2018).

b) An optimal neural network configuration was proposed for path loss predictions in a heterogeneous radio propagation environment of a typical Nigerian smart campus (Popoola, Adetiba, Atayero, Faruk, & Calafate, 2018; Popoola, Misra, & Atayero, 2018).

6.4. Scholarly Publications

6.4.1. Selected Peer-Reviewed Journal Papers

Popoola, S. I., Misra, S., & Atayero, A. A. (**2018a**). Outdoor path loss predictions based on extreme learning machine. *Wireless Personal Communications*, *99*(1), 441-460. [**Springer Publishing;** Indexed in **Scopus** and **ISI Web of Science**].

Popoola, S. I., Adetiba, E., Atayero, A. A., Faruk, N., & Calafate, C. T. (**2018b**). Optimal model for path loss predictions using feed-forward neural networks. *Cogent Engineering*, *5*(1), 1444345. [**Taylor & Francis Publishing;** Indexed in **Scopus** and **ISI Web of Science**].

Surajudeen-Bakinde, N. T., Faruk, N., Salman, M., **Popoola**, S., Oloyede, A., & Olawoyin, L. A. (**2018c**). ON ADAPTIVE NEURO-FUZZY MODEL FOR PATH LOSS PREDICTION IN THE VHF BAND. *International Telecommunication Union Journal on ICT Discoveries*, Special Issue No.1, 1-8. [**Published by ITU**]

Popoola, S. I., Atayero, A. A., & Popoola, O. A. (**2018d**). Comparative assessment of data obtained using empirical models for path loss predictions in a university campus environment. *Data in Brief, 18*, 380-393. doi:https://doi.org/10.1016/j.dib.2018.03.040 [**Elsevier Publishing;** Indexed in **Scopus**].

Popoola, S. I., Atayero, A. A., Arausi, O. D., & Matthews, V. O. (**2018e**). Path loss dataset for modeling radio wave propagation in smart campus environment. *Data in Brief*, *17*, 1062-1073. [**Elsevier Publishing;** Indexed in **Scopus**].

Popoola, S. I., Atayero, A. A., & Faruk, N. (**2018f**). Received Signal Strength and Local Terrain Profile Data for Radio Network Planning and Optimization at GSM

Frequency Bands. *Data in Brief, 16*, 972-981. [**Elsevier Publishing;** Indexed in **Scopus**].

Popoola, S. I., Atayero, A. A., Faruk, N., & Badejo, J. A. (**2018g**). Data on the Key Performance Indicators for Quality of Service of GSM Networks in Nigeria. *Data in Brief, 16*, 914-928. [**Elsevier Publishing;** Indexed in **Scopus**].

Agboje, O. E., **Popoola**, S. I., & Atayero, A. A. (**2017**). LTE-Advanced for Rapid Mobile Broadband Penetration in Developing Countries. *International Journal of Applied Engineering Research, 12*(18), 7863-7872. [Indexed in **Scopus**]

6.4.2. Selected Peer-Reviewed Papers in Conference Proceedings

Popoola, S. I., Atayero, A. A., Faruk, N., Calafate, C. T., Adetiba, E., & Matthews, V. O. (**2017a**). *Calibrating the Standard Path Loss Model for Urban Environments using Field Measurements and Geospatial Data.* Paper presented at the Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering 2017, 5-7 July, 2017, London, U.K., pp513-518. [Indexed in **Scopus**].

Popoola, S. I., Badejo, J. A., Ojewande, S. O., & Atayero, A. (**2017b**). Statistical evaluation of quality of service offered by GSM network operators in Nigeria. *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science 2017, 25-27 October, 2017, San Francisco, USA*, 69-73. [Indexed in **ISI Web of Science**].

Matthews, V. O., Osuoyah, Q., **Popoola**, S. I., Adetiba, E., & Atayero, A. A. (**2017c**). C-BRIG: a network architecture for real-time information exchange in smart and connected campuses. *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering 2017, 5-7 July, 2017, London, U.K., 398-401*. [Indexed in **Scopus**].

Popoola, S. I., Atayero, A. A., Faruk, N., Calafate, C. T., Olawoyin, L. A., & Matthews, V. O. (**2017d**, **July**). Standard propagation model tuning for path loss predictions in built-up environments. In *International Conference on Computational Science and Its Applications* (pp. 363-375). Springer, Cham. [Indexed in **Scopus**].

Abdulrasheed, I. Y., Faruk, N., Surajudeen-Bakinde, N. T., Olawoyin, L. A., Oloyede, A. A., & **Popoola**, S. I. (**2017e**, **November**). Kriging based model for path loss prediction

in the VHF band. In *Proceedings of the 3rd International Conference on Electro-Technology for National Development, Federal University of Technology, Owerri (FUTO), Imo State, Nigeria* (pp. 173-176). IEEE. [Indexed in **IEEE Xplore** and **ISI Web of Science**].

Faruk, N., Imam-Fulani, Y., Sikiru, I. A., **Popoola**, S. I., Oloyede, A. A., Olawoyin, L. A., ... & Sowande, A. O. (**2017f, November**). Spatial variability analysis of duty cycle in GSM band. In *Proceedings of the 3rd International Conference on Electro-Technology for National Development, Federal University of Technology, Owerri (FUTO), Imo State, Nigeria* (pp. 163-166). IEEE. [Indexed in **IEEE Xplore** and **ISI Web of Science**].

Sikiru, I. A., Faruk, N., **Popoola**, S. I., Imam-Fulani, Y., Oloyede, A. A., Olawoyin, L. A., & Surajudeen-Bakinde, N. T. (**2017g, November**). Effects of detection threshold and frame size on duty cycle in GSM bands. In *Proceedings of the 3rd International Conference on Electro-Technology for National Development, Federal University of Technology, Owerri (FUTO), Imo State, Nigeria* (pp. 7-10). IEEE. [Indexed in **IEEE Xplore** and **ISI Web of Science**].

Salman, M. A., **Popoola**, S. I., Faruk, N., Surajudeen-Bakinde, N. T., Oloyede, A. A., & Olawoyin, L. A. (**2017h, October**). Adaptive Neuro-Fuzzy model for path loss prediction in the VHF band. In *Computing Networking and Informatics (ICCNI), 2017 International Conference on* (pp. 1-6). IEEE. [Indexed in **IEEE Xplore** and **ISI Web of Science**].

REFERENCES

Abhayawardhana, V., Wassell, I., Crosby, D., Sellars, M., & Brown, M. (2005). *Comparison of empirical propagation path loss models for fixed wireless access systems.* Paper presented at the Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st.

Action, C. (1999). *231,"Digital mobile radio towards future generation systems*. Retrieved from

Adamkó, A. (2017). *Building Smart University Using Innovative Technology and Architecture.* Paper presented at the International Conference on Smart Education and Smart E-Learning.

Adeyemo, Z. K., Ogunremi, O. K., & Ojedokun, I. A. (2016). Optimization of okumura-hata model for long term evolution network deployment in lagos, nigeria. *International Journal on Communications Antenna and Propagation, 6*(3), 146-152. doi:10.15866/irecap.v6i3.9012

Ahlbom, A., Green, A., Kheifets, L., Savitz, D., Swerdlow, A., & Epidemiology, I. S. C. o. (2004). Epidemiology of health effects of radiofrequency exposure. *Environmental Health Perspectives, 112*(17), 1741.

Akhoondzadeh-Asl, L., & Noori, N. (2007). *Modification and tuning of the universal Okumura-Hata model for radio wave propagation predictions*.

Al Salameh, M. S., & Al-Zu'bi, M. M. (2015). *Prediction of radiowave propagation for wireless cellular networks in Jordan.* Paper presented at the Knowledge and Smart Technology (KST), 2015 7th International Conference on.

Alamoud, M. A., & Schütz, W. (2012). *Okumura-Hata model tuning for TETRA mobile radio networks in Saudi Arabia*.

Angeles, J. C. D., & Dadios, E. P. (2015). *Neural network-based path loss prediction for digital TV macrocells.* Paper presented at the Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2015 International Conference on.

Aragon-Zavala, A. (2008). *Antennas and propagation for wireless communication systems*: John Wiley & Sons.

Axtell, R. (2000). Why agents?: on the varied motivations for agent computing in the social sciences.

Ayadi, M., Zineb, A. B., & Tabbane, S. (2017). A UHF Path Loss Model Using Learning Machine for Heterogeneous Networks. *IEEE Transactions on Antennas and Propagation*.

Bakken, J. P., Uskov, V. L., Kuppili, S. V., Uskov, A. V., Golla, N., & Rayala, N. (2017). *Smart University: software systems for students with disabilities.* Paper presented at the International Conference on Smart Education and Smart E-Learning.

Begovic, P., Behlilovic, N., & Avdic, E. (2012). *Applicability evaluation of Okumura, Ericsson 9999 and winner propagation models for coverage planning in 3.5 GHZ WiMAX systems*.

Benmus, T. A., Abboud, R., & Shatter, M. K. (2015). *Neural network approach to model the propagation path loss for great Tripoli area at 900, 1800, and 2100 MHz bands.* Paper presented at the Sciences and Techniques of Automatic Control and Computer Engineering (STA), 2015 16th International Conference on.

Benmus, T. A., Abboud, R., & Shatter, M. K. (2016). *Neural network approach to model the propagation path loss for great Tripoli area at 900, 1800, and 2100 MHz bands*.

Bhuvaneshwari, A., Hemalatha, R., & Satyasavithri, T. (2016). *Performance evaluation of Dynamic Neural Networks for mobile radio path loss prediction.* Paper presented at the Electrical, Computer and Electronics Engineering (UPCON), 2016 IEEE Uttar Pradesh Section International Conference on.

Bhuvaneshwari, A., Hemalatha, R., & Satyasavithri, T. (2017). *Performance evaluation of Dynamic Neural Networks for mobile radio path loss prediction*.

Bracewell, R. a., Budden, K., Ratcliffe, J., Straker, T., & Weekes, K. (1951). The ionospheric propagation of low-and very-low-frequency radio waves over distances less than 1000 km. *Proceedings of the IEE-Part III: Radio and Communication Engineering, 98*(53), 221-236.

Budden, K. G. (2009). Radio waves in the ionosphere. *Radio Waves in the Ionosphere, by KG Budden, Cambridge, UK: Cambridge University Press, 2009*.

Cerri, G., Cinalli, M., Michetti, F., & Russo, P. (2004). Feed forward neural networks for path loss prediction in urban environment. *IEEE Transactions on Antennas and Propagation, 52*(11), 3137-3139.

Chernov, L. A., & Silverman, R. A. (1960). *Wave propagation in a random medium*: McGraw-Hill New York.

Cisco. (2017). Visual Networking Index: Global mobile data traffic forecast update, 2016-2021.

Clarke, R. (1968). A statistical theory of mobile-radio reception. *Bell Labs Technical Journal, 47*(6), 957-1000.

Cota, N., Serrador, A., Vieira, P., Beire, A. R., & Rodrigues, A. (2013). *On the use of Okumura-Hat a propagation model on railway communications*.

Cruz, J. C. D., & Caluyo, F. S. (2014). *Heuristic modelling of path loss inside residences at 677MHz.* Paper presented at the Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2014 International Conference on.

Delos Angeles, J. C., & Dadios, E. P. (2016). *Neural network-based path loss prediction for digital TV macrocells*.

Eichie, J. O., Oyedum, O. D., Ajewole, M. O., & Aibinu, A. M. (2017a). Artificial Neural Network model for the determination of GSM Rxlevel from atmospheric parameters. *Engineering Science and Technology, an International Journal, 20*(2), 795-804.

Eichie, J. O., Oyedum, O. D., Ajewole, M. O., & Aibinu, A. M. (2017b). Comparative Analysis of Basic Models and Artificial Neural Network Based Model for Path Loss Prediction. *Progress In Electromagnetics Research M, 61*, 133-146.

Erceg, V. (1999). Urban transmission loss models for mobile radio in the 900 and 1800 MHz bands. *IEEE Journal on Selected Areas in Communications*, 1205-1211.

Erceg, V., Greenstein, L. J., Tjandra, S. Y., Parkoff, S. R., Gupta, A., Kulic, B., . . . Bianchi, R. (1999). An empirically based path loss model for wireless channels in suburban environments. *IEEE Journal on Selected Areas in Communications, 17*(7), 1205-1211.

Farhoud, M., El-Keyi, A., & Sultan, A. (2013). *Empirical correction of the Okumura-Hata model for the 900 MHz band in Egypt*.

Faruk, N., Adediran, Y. A., & Ayeni, A. A. (2013). Characterization of Propagation Path Loss at VHF and UHF bands for Ilorin City, Nigeria. *Nigerian journal of Technology (NIJOTECH), University of Nnsuka, 32*(2).

Faruk, N., Adediran, Y. A., & Ayeni, A. A. (2013). *Error bounds of empirical path loss models at VHF/UHF bands in Kwara State, Nigeria.* Paper presented at the EUROCON, 2013 IEEE.

Faruk, N., Ayeni, A., & Adediran, Y. A. (2013). On the study of empirical path loss models for accurate prediction of TV signal for secondary users. *Progress In Electromagnetics Research B, 49*, 155-176.

Fernández Anitzine, I., Romo Argota, J. A., & Fontán, F. P. (2012). Influence of Training Set Selection in Artificial Neural Network-Based Propagation Path Loss Predictions. *International Journal of Antennas and Propagation, 2012*.

Ferreira, G. P., Matos, L. J., & Silva, J. M. (2016). Improvement of Outdoor Signal Strength Prediction in UHF Band by Artificial Neural Network. *IEEE Transactions on Antennas and Propagation, 64*(12), 5404-5410.

Friis, H. T. (1946). A note on a simple transmission formula. *Proceedings of the IRE, 34*(5), 254-256.

Ghassemzadeh, S. S., Jana, R., Rice, C. W., Turin, W., & Tarokh, V. (2002). *A statistical path loss model for in-home UWB channels.* Paper presented at the Ultra Wideband Systems and Technologies, 2002. Digest of Papers. 2002 IEEE Conference on.

Glassner, A. S. (1989). *An introduction to ray tracing*: Elsevier.

Goldsmith, A. (2005). *Wireless communications*: Cambridge university press.

Gschwendtner, B., & Landstorfer, F. (1996). *Adaptive propagation modelling based on neural network techniques.* Paper presented at the Vehicular Technology Conference, 1996. Mobile Technology for the Human Race., IEEE 46th.

Har, D., Watson, A. M., & Chadney, A. G. (1999). Comment on diffraction loss of rooftop-to-street in COST 231-Walfisch-Ikegami model. *IEEE transactions on Vehicular Technology, 48*(5), 1451-1452.

Hata, M. (1980). Empirical formula for propagation loss in land mobile radio services. *IEEE transactions on Vehicular Technology, 29*(3), 317-325.

Haykin, S. S. (1988). *Digital communications*: Wiley New York.

Hufford, G. A. (1952). An integral equation approach to the problem of wave propagation over an irregular surface. *Quarterly of Applied Mathematics, 9*(4), 391-404.

Huschka, T. (1994). *Ray tracing models for indoor environments and their computational complexity.* Paper presented at the Personal, Indoor and Mobile Radio Communications, 1994. Wireless Networks-Catching the Mobile Future., 5th IEEE International Symposium on.

Ibhaze, A. E., Ajose, S. O., Atayero, A. A.-A., & Idachaba, F. E. (2016). *Developing smart cities through optimal wireless mobile network.* Paper presented at the Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech), IEEE International Conference on.

Isabona, J., Konyeha, C., Chinule, C., & Isaiah, G. (2013). Radio Field Strength Propagation Data and Pathloss calculation Methods in UMTS Network. *Advances in Physics Theories and Applications, 21*, 54-68.

ITU. (2008). International Telecommunication Union. *Radiocommunication Sector. ITU-R. Geneva*.

Jeruchim, M. C., Balaban, P., & Shanmugan, K. S. (2006). *Simulation of communication systems: modeling, methodology and techniques*: Springer Science & Business Media.

Kalakh, M., Kandil, N., & Hakem, N. (2012). *Neural networks moDel of an uwb channel path loss in a mine environment.* Paper presented at the Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th.

Kecman, V. (2001). *Learning and soft computing: support vector machines, neural networks, and fuzzy logic models*: MIT press.

Kelso, J. M. (1964). *Radio ray propagation in the ionosphere*: McGraw-Hill.

Laiho, J., Wacker, A., & Novosad, T. (2006). *Radio network planning and optimisation for UMTS*: John Wiley & Sons.

Lee, W. C. (1985). Estimate of local average power of a mobile radio signal. *IEEE transactions on Vehicular Technology, 34*(1), 22-27.

Lempiäinen, J., & Manninen, M. (2003). UMTS radio network planning, optimization and QoS management. *Dodrecht Kluwer Academic Publishers*.

Luebbers, R. (1984). Propagation prediction for hilly terrain using GTD wedge diffraction. *IEEE Transactions on Antennas and Propagation, 32*(9), 951-955.

Mao, G., Anderson, B. D., & Fidan, B. (2007). Path loss exponent estimation for wireless sensor network localization. *Computer Networks, 51*(10), 2467-2483.

Mardeni, R., & Pey, L. Y. (2010). The optimization of Okumura's model for Code Division Multiple Access (CDMA) system in Malaysia. *European Journal of Scientific Research, 45*(4), 508-528.

McLeod, A., Bai, S., & Meyer, J. GPU Accelerated Radio Path Loss Estimation with Neural Networks.

McLeod, A., Bai, S., & Meyer, J. (2013). Graphical Processing Unit accelerated radio path-loss estimation with neural networks. *The Journal of Defense Modeling and Simulation, 10*(2), 117-130.

Medeisis, A., & Kajackas, A. (2000). *On the use of the universal Okumura-Hata propagation prediction model in rural areas*, Piscataway, NJ, United States.

Milanovic, J., Rimac-Drlje, S., & Bejuk, K. (2007). *Comparison of propagation models accuracy for WiMAX on 3.5 GHz.* Paper presented at the Electronics, Circuits and Systems, 2007. ICECS 2007. 14th IEEE International Conference on.

Mitra, A., & Reddy, B. (1987). *Handbook on Radio propagation for tropical and subtropical countries*.

Mohtashami, V., & Shishegar, A. (2012). Modified wavefront decomposition method for fast and accurate ray-tracing simulation. *IET microwaves, antennas & propagation, 6*(3), 295-304.

Mukubwa, E. (2006). A statistical per cell model tuning approaches for cellular networks.

Mukubwa, E., Karien, A., & Chatelain, D. (2006). *A cell focused approach to propagation prediction modelling.* Paper presented at the Satnac Conference, Western Capa South Africa.

Nadir, Z., & Ahmad, M. I. (2010). *Pathloss determination using Okumura-Hata model and cubic regression for missing data for Oman*.

Nawrocki, M., Aghvami, H., & Dohler, M. (2006). *Understanding UMTS radio network modelling, planning and automated optimisation: theory and practice*: john Wiley & sons.

Nimavat, V. D., & Kulkarni, G. (2012). *Simulation and performance evaluation of GSM propagation channel under the urban, suburban and rural environments.* Paper presented at the Communication, Information & Computing Technology (ICCICT), 2012 International Conference on.

Ogawa, N., & Shimizu, A. (2017). *Building a Smarter College: Best Educational Practices and Faculty Development.* Paper presented at the International Conference on Smart Education and Smart E-Learning.

Oseni, O. F., Popoola, S. I., Abolade, R. O., & Adegbola, O. A. (2014a). Comparative analysis of received signal strength prediction models for radio network planning of GSM 900

MHz in Ilorin, Nigeria. *International Journal of Innovative Technology and Exploring Engineering, 4*(3), 45-50.

Oseni, O. F., Popoola, S. I., Abolade, R. O., & Adegbola, O. A. (2014b). Comparative analysis of received signal strength prediction models for radio network planning of GSM 900 MHz in Ilorin, Nigeria. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 4*(3), 45-50.

Oseni, O. F., Popoola, S. I., Enumah, H., & Gordian, A. (2014). Radio Frequency Optimization of Mobile Networks in Abeokuta, Nigeria for Improved Quality of Service. *International Journal of Research in Engineering and Technology, 3*(8), 174-180.

Östlin, E., Suzuki, H., & Zepernick, H.-J. (2008). Evaluation of the propagation model recommendation ITU-R P. 1546 for mobile services in rural Australia. *IEEE transactions on Vehicular Technology, 57*(1), 38-51.

Ostlin, E., Suzuki, H., & Zepernick, H. (2005). *Evaluation of a new effective antenna height definition in ITU-R recommendation P. 1546-1*. Paper presented at the Communications, 2005 Asia-Pacific Conference on.

Ostlin, E., Zepernick, H.-J., & Suzuki, H. (2010). Macrocell path-loss prediction using artificial neural networks. *IEEE transactions on Vehicular Technology, 59*(6), 2735-2747.

Ostlin, E., Zepernick, H., & Suzuki, H. (2003). *Evaluation of the new semi-terrain based propagation model Recommendation ITU-R P. 1546.* Paper presented at the Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th.

Parsons, J. D. (2000). *The mobile radio propagation channel*: Wiley.

Phillips, C., Sicker, D., & Grunwald, D. (2013). A survey of wireless path loss prediction and coverage mapping methods. *IEEE Communications Surveys & Tutorials, 15*(1), 255-270.

Popescu, I., Kanatas, A., Constantinou, P., & Nafornită, I. (2002). Applications of general regression neural networks for path loss prediction. *Neural Networks, 1*, 3.

Popescu, I., Nafomita, I., Constantinou, P., Kanatas, A., & Moraitis, N. (2001). *Neural networks applications for the prediction of propagation path loss in urban environments.* Paper presented at the Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd.

Popescu, I., Nafornita, I., & Constantinou, P. (2005). *Comparison of neural network models for path loss prediction.* Paper presented at the Wireless And Mobile Computing, Networking And Communications, 2005.(WiMob'2005), IEEE International Conference on.

Popescu, I., Nafornita, I., Constantinou, P., Kanatas, A., & Moraitis, N. (2001). Neural networks applications for the prediction of propagation path loss in urban environments. *IEEE Vehicular Technology Conference, 1*(53ND), 387-391. doi:10.1109/VETECS.2001.944870

Popoola, S., & Oseni, O. (2014). Empirical Path Loss Models for GSM Network Deployment in Makurdi. *Nigeria*, 85-86.

Popoola, S. I., Adetiba, E., Atayero, A. A., Faruk, N., & Calafate, C. T. (2018). Optimal model for path loss predictions using feed-forward neural networks. *Cogent Engineering, 5*(1), 1444345.

Popoola, S. I., Atayero, A. A., Arausi, O. D., & Matthews, V. O. (2018). Path loss dataset for modeling radio wave propagation in smart campus environment. *Data in Brief, 17*, 1062-1073. doi:https://doi.org/10.1016/j.dib.2018.02.026

Popoola, S. I., Atayero, A. A., & Faruk, N. (2018). Received Signal Strength and Local Terrain Profile Data for Radio Network Planning and Optimization at GSM Frequency Bands. *Data in Brief, 16*, 972-981.

Popoola, S. I., Atayero, A. A., Faruk, N., & Badejo, J. A. (2018). Data on the Key Performance Indicators for Quality of Service of GSM Networks in Nigeria. *Data in Brief, 16*, 914-928.

Popoola, S. I., Atayero, A. A., Faruk, N., Calafate, C. T., Adetiba, E., & Matthews, V. O. (2017). *Calibrating the Standard Path Loss Model for Urban Environments using Field Measurements and Geospatial Data.* Paper presented at the Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering 2017, 5-7 July, 2017, London, U.K., pp513-518.

Popoola, S. I., Atayero, A. A., Faruk, N., Calafate, C. T., Olawoyin, L. A., & Matthews, V. O. (2017). *Standard Propagation Model Tuning for Path Loss Predictions in Built-Up Environments.* Paper presented at the International Conference on Computational Science and Its Applications.

Popoola, S. I., Atayero, A. A., & Popoola, O. A. (2018). Comparative assessment of data obtained using empirical models for path loss predictions in a university campus environment. *Data in Brief, 18*, 380-393. doi:https://doi.org/10.1016/j.dib.2018.03.040

Popoola, S. I., Badejo, J. A., Ojewande, S. O., & Atayero, A. (2017). Statistical evaluation of quality of service offered by GSM network operators in Nigeria. *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science 2017, 25-27 October, 2017, San Francisco, USA*, 69-73.

Popoola, S. I., Misra, S., & Atayero, A. A. (2018). Outdoor path loss predictions based on extreme learning machine. *Wireless Personal Communications, 99*(1), 441-460.

Popoola, S. I., & Oseni, O. F. (2014a). Empirical Path Loss Models for GSM Network Deployment in Makurdi, Nigeria. *International Refereed Journal of Engineering and Science, 3*(6), 85-94.

Popoola, S. I., & Oseni, O. F. (2014b). Performance Evaluation of Radio Propagation Models on GSM Network in Urban Area of Lagos, Nigeria. *International Journal of Scientific & Engineering Research, 5*(6), 1212-1217.

Proakis, J. G., & Salehi, M. (2007). *Fundamentals of communication systems*: Pearson Education India.

Proakis, J. G., Salehi, M., Zhou, N., & Li, X. (1994). *Communication systems engineering* (Vol. 2): Prentice Hall New Jersey.

Rappaport, T. S. (1996). *Wireless communications: principles and practice* (Vol. 2): prentice hall PTR New Jersey.

Rappaport, T. S., & Sandhu, S. (1994). Radio-wave propagation for emerging wireless personal-communication systems. *Ieee Antennas and Propagation Magazine, 36*(5), 14-24.

Rath, H. K., Verma, S., Simha, A., & Karandikar, A. (2016). *Path Loss model for Indian terrain-empirical approach.* Paper presented at the Communication (NCC), 2016 Twenty Second National Conference on.

Repacholi, M. H. (1998). Low-level exposure to radiofrequency electromagnetic fields: Health effects and research needs. *Bioelectromagnetics, 19*(1), 1-19.

Repacholi, M. H. (2001). Health risks from the use of mobile phones. *Toxicology letters, 120*(1-3), 323-331.

Salman, M. A., Popoola, S. I., Faruk, N., Surajudeen-Bakinde, N., Oloyede, A. A., & Olawoyin, L. A. (2017). *Adaptive Neuro-Fuzzy model for path loss prediction in the VHF band.* Paper presented at the Computing Networking and Informatics (ICCNI), 2017 International Conference on.

Schaubach, K. R., Davis, N., & Rappaport, T. S. (1992). *A ray tracing method for predicting path loss and delay spread in microcellular environments.* Paper presented at the Vehicular Technology Conference, 1992, IEEE 42nd.

Schelkunoff, S. A., & Friis, H. T. (1952). *Antennas: theory and practice* (Vol. 639): Wiley New York.

Schneider, P., Lambrecht, F., & Baier, A. (1996). *Enhancement of the Okumura-Hata propagation model using detailed morphological and building data*, Piscataway, NJ, United States.

Seybold, J. S. (2005). *Introduction to RF propagation*: John Wiley & Sons.

Simon, M. K., & Alouini, M.-S. (2005). *Digital communication over fading channels* (Vol. 95): John Wiley & Sons.

Sklar, B. (2001). *Digital communications* (Vol. 2): Prentice Hall Upper Saddle River.

Sotiroudis, S., Goudos, S., Gotsis, K., Siakavara, K., & Sahalos, J. (2013). *Modeling by optimal Artificial Neural Networks the prediction of propagation path loss in urban environments.* Paper presented at the Antennas and Propagation in Wireless Communications (APWC), 2013 IEEE-APS Topical Conference on.

Sotiroudis, S. P., Goudos, S. K., Gotsis, K. A., Siakavara, K., & Sahalos, J. N. (2013). Application of a composite differential evolution algorithm in optimal neural network design for propagation path-loss prediction in mobile communication systems. *IEEE Antennas and Wireless Propagation Letters, 12*, 364-367.

Sotiroudis, S. P., Goudos, S. K., Gotsis, K. A., Siakavara, K., & Sahalos, J. N. (2013). *Optimal Artificial Neural Network design for propagation path-loss prediction using adaptive evolutionary algorithms*.

Sotiroudis, S. P., & Siakavara, K. (2015). Mobile radio propagation path loss prediction using artificial neural networks with optimal input information for urban environments. *AEU-International Journal of Electronics and Communications, 69*(10), 1453-1463.

Sotiroudis, S. P., & Siakavara, K. (2015). Mobile radio propagation path loss prediction using Artificial Neural Networks with optimal input information for urban environments. *AEU - International Journal of Electronics and Communications, 69*(10), 1453-1463. doi:10.1016/j.aeue.2015.06.014

Subrt, L., & Pechac, P. (2011). Advanced 3D indoor propagation model: calibration and implementation. *Eurasip Journal on Wireless Communications and Networking, 2011*(1), 180.

Timoteo, R. D., Cunha, D. C., & Cavalcanti, G. D. (2014). A proposal for path loss prediction in urban environments using support vector regression.

Turin, W., Jana, R., Martin, C., & Winters, J. (2001). *Modeling wireless channel fading.* Paper presented at the Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th.

Uskov, V. L., Bakken, J. P., Karri, S., Uskov, A. V., Heinemann, C., & Rachakonda, R. (2017). *Smart University: Conceptual Modeling and Systems' Design.* Paper presented at the International Conference on Smart Education and Smart E-Learning.

Verma, R., & Saini, G. (2016). *Statistical tuning of Cost-231 Hata model at 1.8 GHz over dense urban areas of Ghaziabad*.

Walfisch, J., & Bertoni, H. L. (1988). A theoretical model of UHF propagation in urban environments. *IEEE Transactions on Antennas and Propagation, 36*(12), 1788-1796.

Zaarour, N., Affes, S., Kandil, N., & Hakem, N. (2015). *Comparative Study on a 60 GHz Path Loss Channel Modeling in a Mine Environment Using Neural Networks*.

Zelley, C. A., & Constantinou, C. C. (1999). A three-dimensional parabolic equation applied to VHF/UHF propagation over irregular terrain. *IEEE Transactions on Antennas and Propagation, 47*(10), 1586-1596.

Zineb, A. B., & Ayadi, M. (2016). A Multi-wall and Multi-frequency Indoor Path Loss Prediction Model Using Artificial Neural Networks. *Arabian Journal for Science and Engineering, 41*(3), 987-996. doi:10.1007/s13369-015-1949-6