Coverage Prediction Based on Spatial Interpolation Techniques: The Case of UMTS Network in Addis Ababa, Ethiopia

By

Zeneb Kassaw

Advisor: Dr. –Eng. Yihenew Wondie

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of Masters of Science in Telecom Network Engineering

Addis Ababa University

Addis Ababa Institute of Technology

January, 2020
Addis Ababa, Ethiopia

Addis Ababa University

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Coverage Prediction Based on Spatial Interpolation Techniques: The Case of UMTS Network Addis Ababa, Ethiopia

By

Zeneb Kassaw

Approval by Board of Examiners

_____	_____	_____
Chair Person, School Graduate Committee	Date	Signature
Dr. Yihenew Wondie	_____	_____
Advisor	Date	Signature
_____	_____	_____
Examiner	Date	Signature
_____	_____	_____
Examiner	Date	Signature

Declaration

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

Zeneb Kassaw Kindie _____

Name Signature

Place: Addis Ababa

Date of Submission: _____

This thesis has been submitted for examination with my approval as an advisor.

Dr. -Eng. Yihenew Wondie _____

Advisor Signature

Abstract

Cellular network coverage prediction is a cornerstone of mobile network operators and service providers in order to provide good services to users. Without coverage provisioning, it is meaningless to talk about service or Quality of service (QoS) provisioning. Coverage planning is a complex task for operators during deploying Radio Access Technology (RAT). This is because it needs to consider multiple and correlated network parameters as well as environmental conditions that are out of their control. It is impossible to completely avoid the existence of coverage holes in cellular networks during the planning phase. Therefore, coverage prediction processes are usually required during the operational phase.

Traditionally, the cellular coverage estimation performed through drive tests, which consist of geographically measuring different network coverage metrics with a motor vehicle equipped with mobile radio measurement facilities. The collected coverage measurements through drive test are accurate but limited to roads and other regions accessible by motor vehicles. Drive tests cannot be conducted in the whole region of the network due to many obstacles such as buildings, lakes, and vegetation. Therefore, the drive test is quite inefficient means to solve the coverage problems and cannot offer a complete and reliable picture of the network situation.

In this thesis, the performance of two spatial interpolation methods namely, Inverse Distance Weight (IDW) and Ordinary Kriging (OK) were evaluated to select which method is best for Universal Mobile Telecommunication System (UMTS) network coverage prediction using the Common Pilot Channel Received Signal Code Power (CPICH RSCP) collected from drive test. The experimental analysis was performed on a sample data collected from drive test UMTS network in Addis Ababa Ethiopia. Two general interpolation methods were employed with different parameters. The first method is IDW with various powers and number of neighbors and the second method is OK with Gaussian, Spherical and Exponential semivariogram models with different numbers of neighbors. The performance of estimation those algorithms were evaluated through the cross-validation, coefficient of determination (R^2), Mean absolute error (MAE) and Root Mean Square Error (RMSE).

The test results showed the two coverage prediction methods are able to predict coverage. However, based on the Exponential model of semivariogram with an optimal number of neighbors the OK method estimated with an error of prediction 4.84 RMSE whereas the IDW estimated 5.33 RMSE with a percentage difference of 17%. This shows that OK is more accurate than IDW. The OK method can infer the missing RSCP data and generates a more accurate coverage map than the IDW algorithm. This could probably be OK was able to take into account the spatial structure of data. Therefore, this thesis proposes the OK method as the optimal interpolation model to build a radio coverage map for cellular coverage prediction and hole detection purposes.

Keywords*:* *Spatial interpolation; UMTS coverage prediction; Ordinary Kriging; IDW; coverage map; spatial variation and semivariogram*.

Acknowledgment

First and foremost, I would like to thank God Almighty for giving me the strength, knowledge, ability and opportunity to undertake this research study and to persevere and complete it. Without his blessings, this achievement would not have been possible.

I wish to express my deepest gratitude to my advisor Dr. -Eng. Yihenew Wondie for his dedicated support and guidance. He continuously provided encouragement was always willing and enthusiastic to assist in any way he could throughout the research thesis. I would also like to thank my evaluators Dereje Hailemariam (PhD) and Surafel Lemma (PhD) for the feedbacks during the thesis progress presentations.

Finally, I must express my very profound gratitude to my parents and to my husband Dani for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Table of Content

List of Figures

List of Tables

List of Acronyms

2G	2nd Generations
3GPP	3rd Generation Partnership Project
AUC	Authentication Centre
BTS	Base transceiver station
BSSAP	Base Station System Application Part
CAPEX	Capital Expenditure
CN	Core Network
Ec/No	Received energy per chip divided by the power density in the band
FDD	Frequency Division Duplex
GERAN	GSM/EDGE radio access technologies
GGSN	Gateway GPRS Support Node
GMSC	Gateway MSC
HSPA	High Speed Packet Access
KPI	Key Performance Indicator
IDW	Inverse Distance Weight
MDT	Minimization of Drive Test
MNO	Mobile Network Operator
OK	Ordinary Kriging
OPEX	Operational Expenditures
PC	Personal Computer
QoE	Quality of Experience
QoS	Quality of Service
RAB	Radio Access Bearers.
RF	Radio Frequency
RAN	Radio Access Network
RAT	Radio Access Technology
REM	Radio Environment Map
RL	Radio Link

RMSE	Root Mean Square Error
MAE	Mean Absolute Error
RRC	Radio Resource Control
RRM	Radio Resource Management
RSCP	Received Signal Code Power
RSSI	Received Signal Strength Indicator
SIR	Signal to Interference Ratio
TDD	Time division duplex
UE	User Equipment
UMTS	Universal Mobile Telecommunication System
USIM	UMTS Service Identity Module
UTRAN	UMTS Terrestrial Access Network.
WCDMA	Wideband Code Division Multiple Access

1 Introduction

Mobile communication is a rapidly growing technology to fulfill the increased demand for subscribers. The increasing demand for different services such as data and voice to meet this demand the network operator deploys new cellular network. To deploy a new Radio Access Technology (RAT), network coverage planning is a mandatory process. It is a complex task for operators since operators use sophisticated planning tools to estimate coverage. They need to consider multiple and correlated network parameters as well as environmental conditions that are out of their control [3] [2]. Despite the great effort usually spent during the planning and deployment phase, the existence of coverage holes is almost impossible to avoid because of the complexity of network configuration, unforeseen changes in the propagation environment or equipment failures [4]. Deploying a new RAT causes network complexity in terms of a number of monitored parameters and parallel operation of 2nd Generation (2G), 3th Generation (3G) and 4th Generation (4G) networks [5]. Those networks still suffer from a number of issues relating to network coverage. These include cell coverage limitations, area-specific network traffic, and air interface interference [6]. Therefore, completely avoiding the existence of coverage problems in cellular networks during the planning phase is almost impossible. Not only before but also after the deployment of network infrastructure, the level of network coverage provided to various parts of the region under consideration is measured on a regular basis. Furthermore, the coverage area is not always steady, as it is also influenced by the time varying band and channel characteristics as well as the instantaneous capacity demand in the cell, particularly in the case of 3G access [7]. Therefore, coverage optimization processes are usually required during the operational phase. Where it is first necessary to detect the coverage problems and then deploy a solution, which remedies or removes the coverage problem in the uncovered zones. The deployed solution must be cost-efficient and at the same time effective, that is, providing full coverage without creating excessive pollution to the already covered neighboring areas. In order to achieve such a solution, precise information on the locations and shapes of the coverage problems are needed. Obtaining this information is called *coverage prediction* [4]. Coverage prediction is important to determine any coverage related problems produced due to either the planning phase or construction of new buildings, highways or

changes in customer residential preferences. The coverage holes thus diagnosed are dealt with either by changing antenna tilt, its height, power level or deployment of new base stations etc. [8].

The major problem that needs to be addressed by the operators in this respect is the coverage hole in 3G connectivity experienced by mobile users. Such hole is experienced possibly due to loss of signal power or lack of coverage. Quality of Service (QoS) such as: data rate, transmission delay, packet loss, and jitter depends upon the distribution of signal strength over a targeted area of service. So that coverage affects the QoS of an operator. A wireless network should provide 100% coverage with signal strength above a minimum threshold value overall its target area. To ensure an acceptable level of QoS for the users of a fixed infrastructure wireless network designers rely on on-site survey techniques and/or signal propagation models [9]. It also affects the quality of Experience (QoE) such as low data rate connections, drop call and handover to other RAT thus reducing the revenues from data services and cause imperceptions regarding service quality on the customer side. It is therefore of utmost importance for operators to deal with identify geographical regions with coverage problems through coverage prediction in cellular network connectivity [7].

This study has been done based on the ethio telecom data, which is the sole telecom service provider in Ethiopia. ethio telecom provides cellular network services through all cellular networks (i.e., 2G, 3G and 4G). Most of the services are carried by the (3G) technology. As [10] explained around eighty-three percent of the data traffic in the capital, Addis Ababa, is carried by UMTS. This is the main reason for this research to select the UMTS network for coverage prediction analysis.

So far, cellular coverage prediction is performed through drive tests, which consist of geographically measuring different network coverage metrics with a motor vehicle equipped with mobile radio measurement facilities. The collected network measurements need to be processed by radio experts for network coverage optimization, e.g. by tuning network parameters such as transmission power, antenna orientations and tilts, etc. [4] [3]. Drive test has its own advantage and disadvantage in terms of prediction of coverage for operators. The advantage of drive tests generates a tremendous amount of data to be processed, allowing the operators to get realistic

network information close to the actual user experience. This is very useful and desired information by operators.

With the processed drive test measurements, the operator can have a realistic picture of the network in terms of coverage and service quality [4]. However, drive tests are quite an inefficient means to solve the coverage problems because they provide an *incomplete* picture of the network because they are limited to roads and other regions accessible by motor vehicles. Drive tests are inefficient and very expensive solutions but it gives the operators to have realistic information on the "ground-truth". Therefore, it is mandatory for the operators to make the most of the information collected through drive tests, and to minimize the use of them. To this end, the cellular network-related standardization body, the 3rd Generation Partnership Project (3GPP), has included a feature on the Minimization of Drive Tests (MDT) since Release 9 [2] [4]. Due to the cost reduction promised by MDT, it is seen as one of the highest-priority features of next-generation cellular networks, and therefore a considerable amount of effort is put forward for its standardization, product development, and marketing. The basic concept of MDT is that the User Equipment's (UEs) report their geo-located measurements to the network upon operator request. The mobile network operators (MNO) including ethio telecom also use drive testing and walk testing approaches to obtain real network parameter measurements. These measurements are very accurate but are limited to publicly accessible areas [6]. This means that only a limited amount of measurement data is available for analysis. This method has such limitations, therefore, selecting the best method to estimate the spatial variations of coverage distribution of a UMTS network has an important strategic significance for reasonable coverage optimization and sustainable development of coverage around the service area [11].

In this thesis, a comparative study of spatial interpolation techniques has been done, the one which has the least root mean square error (RMSE) method is selected among the different spatial interpolation techniques to estimate the value of RSCP for the unknown location in the UMTS network [9]. Moreover, to create an accurate and complete network coverage map of a target service area from a limited number of RSCP measurements. The main target of this research is reducing the limitation of not obtaining measurement value due to the inaccessible of drive test by using spatial interpolation techniques. Since interpolation is done using a few measurement points

it reduces the cost of fuel for accessing large areas for data collection to estimation coverage of UMTS network using a drive test.

In addition, the proposed approach can be used to estimate the network coverage where samples could not be taken due to inaccessibility and create coverage map which is used for coverage hole detection.

The motivation of this thesis work is to propose a mobile coverage prediction model using a spatial interpolation method to create a coverage map based on measurement statistical data collected from drive test of UMTS. The thesis work also focuses on the performance evaluation of two spatial interpolation techniques to estimate the RSCP value where there is no enough measurement report to create a smooth surface. Such methods are Inverse Distance Weight (IDW) and Ordinary Kriging (OK). Finally, the coverage map of the UMTS network for the selected area is created by the selected interpolation techniques. This approach uses data collected from drive test consist of radio link quality parameters, mainly CPICH RSCP which includes location information. These parameters have been selected for this thesis work since they have a direct impact on network coverage [6] [12].

1.1 Statement of Problem

Coverage prediction is needed both in the deployment and operation of cellular networks. Traditionally, cellular coverage estimation is performed through drive tests. Drive test cannot be conducted in the whole region of the network because of many obstacles such as buildings, lakes, and vegetation [12]. Therefore, the drive test is quite inefficient means to solve the coverage problems and cannot offer a complete and reliable picture of the network situation.

The proposed approach aims to reduce the limitation of not obtaining measurement value due to the inaccessible of drive test by using spatial interpolation techniques. So that complete network coverage related information is available by estimating RSCP value at points where no measurement data is available using samples taken at other points through drive test.

1.2 Objective

1.2.1 General Objective

- The main objective of this thesis is to build radio coverage map for the purpose of UMTS network coverage prediction and hole detection using spatial interpolation techniques such as IDW and OK methods, using georeferenced RSCP measurement data collected from drive test and network topology information taken from ethio telecom Addis Ababa, Ethiopia.

1.2.2 Specific Objectives

- To grasp the global trend of coverage prediction through literature review;
- To explore the statistical property of the collected RSCP data;
- To build a semivariogram model to study the relationship of variables in space;
- To select fitting models;
- To estimate UMTS network coverage in the selected area;
- To evaluate different interpolation techniques (OK, IDW) ;
- To construct coverage map using spatial interpolation techniques through geo-located measurements collected from drive test data.

1.3 Literature Review

For this research, various literature on the coverage prediction for mobile networks were reviewed. The reviewed literature was presented in two sections based on the input data for coverage analysis and the techniques to predict coverage in cellular networks.

Based on Input data

This section presents a brief survey of literature focusing on the input data source used for prediction of coverage in the cellular networks

A study in [7], propose a 3G coverage analysis method that makes use of ”big data” processing schemes and the vast amounts of *network data logged* in mobile operators. They identify service discontinuity area using the network logs for mobile uses, corresponding to both their idle and connected mode signaling data using the Hadoop platform, to come up with dynamic 3G service discontinuity maps. In the proposed scheme, the logged data consists of the Base Station System Application Part (BSSAP) mobility and radio resource management messages between the BSS and MSC nodes of the operator network that are processed to identify coverage discontinuity. The 3G discontinuity is experienced when the user is forced to handover from 3G connectivity down to 2G, possibly due to loss of signal quality. More specifically, mobility and radio resource management signaling data (BSSAP messages) collected by the probes attached to the relevant network switches are used. Through such data, the inter radio access technology handovers (IRAT HOs) from 3G to 2G for mobile users are mapped to geographical locations of the users to identify the 3G service discontinuity areas. Experimental results show that the service discontinuity areas achieved through this approach match the drive test results. The limitation of the paper was it did not show how geographical locations of the users' data obtained to identify the 3G service discontinuity areas.

In [13], proposed an approach to detect cells with coverage holes and diagnose their type and severity. This is achieved through the use of both traditional statistics indicators (obtained from the Operation, Administration and Maintenance (OAM) system) and the user information obtained from the mobile traces gathered both in LTE and in the co-existing underlying radio access technology(uRAT). The main difference of the proposed solution of this paper with the approaches available in the other literature is that the presented diagnosis system does not only detect cells with coverage holes but also diagnoses their type and severity.

[14], this paper proposed a deterministic propagation model for coverage hole detection CHD by utilizing the radio link failure report, which can significantly reduce the costly drive test. They proposed coverage hole detection (CHD) under the condition of no prior terrain information and no spatial correlation. They also only compare the proposed algorithm to a heuristic CHD approach. They proposed algorithm which is termed a Piecewise Linear model-based CHD (PL-

CHD) in a subsequent paragraph, gives the best efforts in predicting radio map and terrain information around the coverage hole, where the (partial) realistic terrain information.

However, the author of this paper replaces the drive test by radio link failure with the radio propagation model to reconstruct the radio map of the target cell. It is a mechanism originated in feeding back handover-issue, where the mobile device, which loses connectivity with BS, is specified to send RLF report to the newly connected BS. They use their own CHD algorithm, which is termed a Piecewise Linear model-based CHD (PL-CHD) in a subsequent paragraph. Then, a simplified version of PL-CHD will be given by using a heuristic approach to reduce the calculation in estimating terrain information.

Based on Methods

Several studies have used Kriging for coverage prediction in wireless networks:

In[8] uses crowdsourced measurement data along with spatial interpolation techniques to generate coverage maps for cellular coverage prediction with less expense and labor. The Author also empirically studies the accuracy of a wide range of spatial interpolation techniques, including various forms of Kriging, in different scenarios that capture the unique characteristics of crowdsourced measurements (inaccurate locations, sparse and non-uniform measurements, etc.). The results indicate that Ordinary Kriging is a robust technique overall, across all scenarios.

In [15] use Kriging Method for power levels analysis level radiated by a Base Station (also called Node B) in WCDMA, through a group of samples of this power level, measured at different positions and distances. These samples were obtained using an spectrum analyzer, which will allow to have georeferenced measurements, to implement the interpolation process and generate coverage maps, making possible to know the power level distribution and therefore understand the behavior and performance of the Node B. The author also uses different amount of the point data to generate reliable continuous coverage maps through a process of interpolation, which predicts the power values in, places where there is not information based on adjacent values.

[16] in this paper, coverage estimation based on either drive tests or data obtained through the MDT approach can be improved using Bayesian spatial statistics. This work is an extension of [2]

and they showed that, using realistic coverage data that spatial prediction techniques can significantly increase the likelihood of discovering coverage holes for a given number of measurements conducted while retaining low probability of false alarms.

A study in [9] proposed the kriging method to predict network coverage in wireless networks. The proposed approach aimed to reduce the cost of active site surveys by estimating network coverage at points where no site survey data is available. The author also compared the proposed approach with a radial basis function artificial neural network using several problems. However, the main objective was to compare the performances of the ANN and kriging approaches on simulated data and demonstrate the advantages and disadvantages of each approach. The authors used two statistics: the coefficient of determination (R2) and the p-value of a paired t-test to compare the ANN and kriging approaches.

In [1] used Kriging to estimate a coverage area in Melbourne, FL. The author used drive test data to estimate the coverage area. The authors reported that It was difficult to collect the Receive Signal Level (RSL) for an entire coverage area using drive test because of many obstacles, such as buildings, lakes, and vegetation so that the estimation of the coverage area is essential for locations for which it is difficult to measure the RSL. They found the estimated coverage area could help to reduce the consumption of time and money. The author proposed the spatial interpolation technique called Kriging.

1.4 Methodology

To have efficient coverage prediction using spatial interpolation techniques this thesis follows the following processes.

Literature review: Kinds of literature were reviewed to identify and select which measurement report parameter affects coverage in the cellular network, explore different spatial interpolation techniques to estimate coverage, and explore different data collection methods.

Data Collection: The data used for this study consist of 3G received pilot powers with geographic location. The geo-located Received Signal Code Power (RSCP) measurements were collected from the drive test.

Coverage Map Generation/development: The coverage prediction was performed based on collected RSCP value using different spatial interpolation techniques. Geostatistical analyst extension of ArcMap was used for spatial data exploration and for interpolation analysis.

Accuracy Assessment of different interpolation techniques: Accuracy assessment of different interpolation techniques was performed using cross-validation and statistical methods.

All analysis was accomplished by using the Geostatistical Analyst and Spatial Analysis tool in ArcGIS 10.3 and excel.

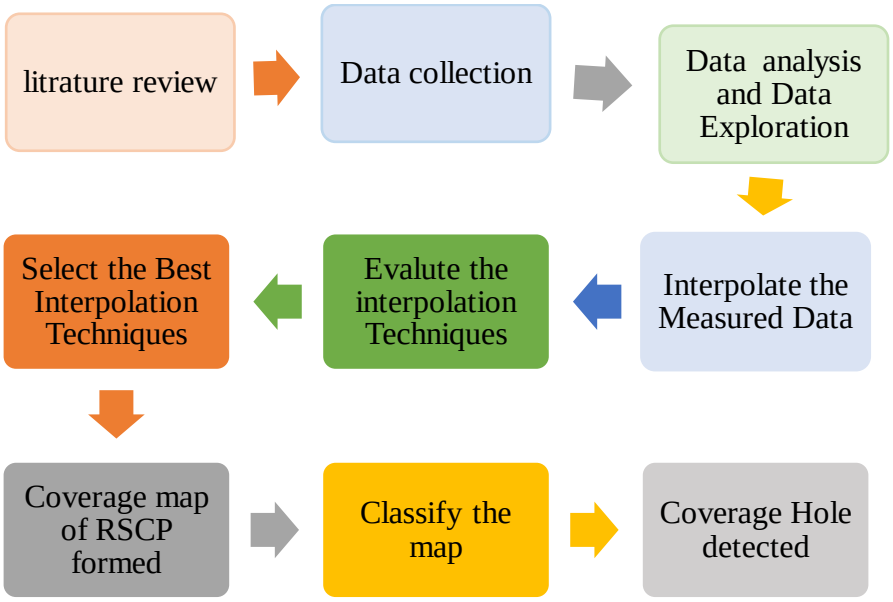


Figure 1.1: Methodology Used

1.5 Scope

This research proposes Ordinary Kriging is the best model to estimate the received signal code power in cellular networks by comparing two spatial interpolation algorithms. As the research is a case study, only considers coverage prediction for Addis Ababa city's UMTS network only by

considering RSCP data collected using a drive test for the selected area, which is 4km by 3km. It only considers 26 UMTS Node-B sites.

1.6 Contribution of the Thesis

This thesis proposes to apply the OK interpolation algorithm to infer the uncollected coverage information (RSCP) being not accessed by the drive test . This approach helps the operator to see the overall network coverage for the UMTS cellular network and shows the pinpoint where coverage hole (poor coverage) happens.

Ultimately mobile network coverage prediction is important for network optimization because it indicates network coverage gaps and areas with poor serviceability as well as indicates interference between different sites since this research considers more than one site. Generally, this thesis important for operators to reduce their Capital expenditures (CAPEX) and Operating Expenses (OPEX) costs by estimating coverage using sample measured points from the drive test. Using such kind of data for coverage analysis helps the operator (ethio telecom) to save time and cost from collecting field measurement data through the drive test.

1.7 Thesis Structure

The work of this thesis is organized into six chapters. Chapter 1 introduces the thesis work, namely its motivations, objectives, methodology, scope, and limitations. Chapter 2 is the technical background of the thesis. It consists of an overview of UMTS cellular network architecture and some cellular network measurement parameters related to coverage analysis. Then in Chapter 3, different spatial estimation (interpolation) approach, including the necessary mathematical foundations presented. In Chapter 4, the experimental analysis of this thesis is presented. Which, describes the considered scenario, including data collection and related parameters, as well as the proposed system model for coverage prediction through interpolation algorithms and evaluation. Chapter 5 presented the results of the paper, the coverage map using the selected prediction method and the performance of the different spatial interpolation process and the selected prediction method are presented. Finally, Chapter 6 conclusions and future work are drawn.

2 Background

This chapter provides the state of the art and the theoretical knowledge that supports the development of this thesis. It highlights some mobile communication's basic concepts. A brief introduction of the UMTS network, along with a description of its network architecture is given. Later, the main characteristics 3G RAN (UTRAN) radio interface is introduced. At the end of the chapter, the coverage affecting parameters that are discussed throughout this thesis is presented.

2.1 UMTS Cellular Network

Universal Mobile Telecommunication System (UMTS) is the third generation telecommunications system based on WCDMA. WCDMA (Wideband Code Division Multiple Access) is the radio interface for UMTS. It will be the successor of GSM. UMTS will offer more high-speed services to mobile computer users no matter where they are located in the world [17].

2.2 UMTS System Architecture

This subsection gives an overview of the UMTS architecture, including an introduction to the logical network elements and the interfaces. The UMTS consists of a number of logical network elements that each has a defined functionality. Functionally, the network elements are grouped into the Universal Terrestrial Radio Access Network (UTRAN) that handles all radio-related functionality, and the Core Network (CN), which is responsible for switching and routing calls and data connections to external networks. To complete the system, the User Equipment (UE) that interfaces with the user and the radio interface is defined [18].

The empirical architecture of UMTS is described by the figure below. It consists of three major parts [18]: UE, UTRAN and CN.

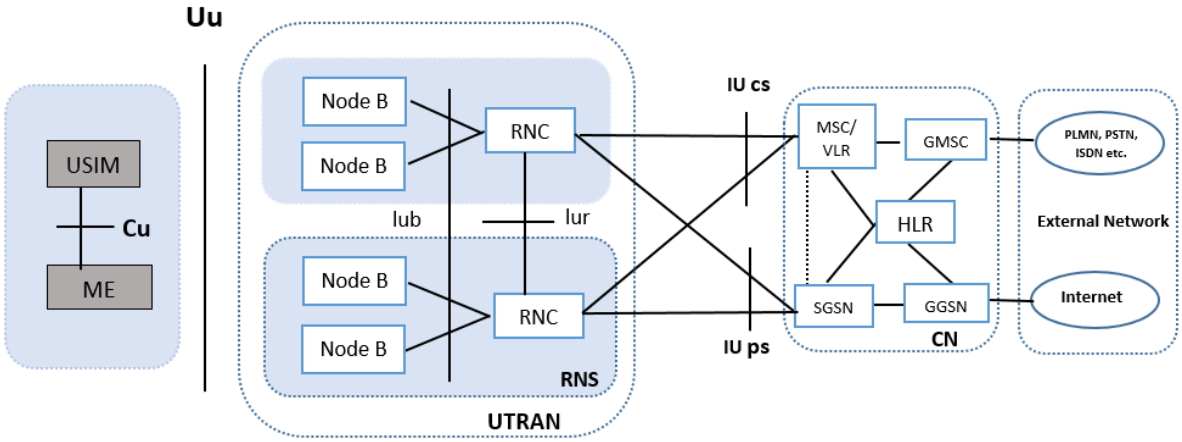


Figure 2.1: UMTS architecture [17].

2.2.1 User Equipment (UE)

The UE consists of two parts Mobile Equipment (ME) and UMTS Subscriber Identity Module (USIM) . ME is the radio terminal used for radio communication over the Uu interface. The second part of UE is USIM is a smart card that contains user identity information and encryption keys [18].

2.2.2 UMTS Terrestrial Radio Access Network (UTRAN)

UTRAN is responsible for the entire radio interface. UTRAN further contains Radio Network Subsystems (RNS). RNS is connected with CN via IU interfaces and UTRAN connects UE via Uu interface [18].

UTRAN also consists of two distinct elements [17]:

Node B converts the data flow between the Iub and Uu interfaces. It also participates in radio resource management (RRM). (*Note*: the term 'Node B' from the corresponding 3GPP specifications is used throughout this chapter. The more generic term 'base station' used.

The *Radio Network Controller (RNC*) owns and controls the radio resources in its domain (the Node'Bs connected to it). The RNC is the service access point for all services that UTRAN provides the CN, e.g. management of connections to the UE.

2.2.3 **Core Network (CN)**

The CN accommodates the remaining network elements needed for switching/ routing and subscriber control. The core network mainly deals with functionalities that are not directly related to radio access technology. Core network connected to UTRAN through the Iu interface as shown in figure 2.2. UMTS core network is mainly divided into two domains: Circuit Switched Domain and Packet Switched Domain [19]

Circuit Switched Domain

This domain mainly deals with the circuit-switched traffic which requires dedicated network resources and interconnection to the external circuit-switched networks. Circuit Switched domain connected to UTRAN through the Iu-CS interface. The main elements of the Circuit Switched Domain are Mobile service Switching Centre (MSC), Home Location Register (HLR), Visitor Location Register (VLR) and Gateway MSC (GMSC).

Packet Switched Domain

This domain mainly deals with the packet data traffic and connect mobile network with external packet-switched networks. Packet Switched Domain connected to UTRAN through Iu-PS interface. The main elements of this domain are Serving GPRS Support Node (SGSN), Gateway GPRS Support Node (GGSN), and GPRS Register (GR). Some elements are common in both domains. These elements are HLR, VLR, Equipment Identity Register (EIR) and Authentication Centre (AuC) [18].

UMTS Interfaces

Four new interfaces are defines in UMTS that are as follows: Uu, Iub, Iur, and Iu [18].

Iu Interface: It connects radio network controllers (RNCs) with core network nodes.

Iur Interface: The Iur interface has no equivalent in the GSM system. It connects the two RNCs in the UTRAN.

Iub Interface: As mentioned earlier that lub is the logical interface, which connects a Node B with RNC.

Uu Interface: provides interconnection between the user terminal and RNC through node B.

2.3 Cellular Radio Measurements

The use of radio waves for transmitting information is the base of all radio communications [20]. In general, all radio-related measurements are well defined and described in 3GPP specification [21]. Therefore, this section will provide a summary combined with information on message flows and some measurement result examples.

Physical Layer – Measurements (FDD) defines two major groups: UE measurement abilities and UTRAN measurement abilities. Looking at protocol functions it emerges that all reports related to UE measurement abilities are provided by the RRC protocol while Node B Application Part(NBAP) and Iub FP are used for reporting values of UTRAN measurement abilities. A closer insight to NBAP reveals that there are two different categories defined: dedicated measurement and common measurement procedures. While common measurement procedures are related to cells, dedicated procedures are related to dedicated physical channels connecting UEs and networks [19].

UE measurement reports that indicate coverage level are [21]:

CPICH RSCP

The "Received Signal Code Power" (RSCP) is the collected RF energy after the correlation / descrambling process, usually given in dBm. Because this process already "filters out" the signal RSCP is the received power on the Common Pilot Channel (CPICH) measured over the 5 MHz 3G carrier bandwidth [2]. The reference point for the RSCP is the antenna connector at the UE

[21]. CPICH RSCP also can be used in the open-loop power control for path loss calculations, higher values can be measured especially close to the base station.

CPICH Ec/I0

The received energy per chip divided by the power density in the band. The Ec/No is identical to RSCP/RSSI. Measurement shall be performed on the CPICH. The reference point for Ec/No is the antenna connector at the UE[21]. CPICH_EcNo is the most important measurement in WCDMA from the radio planning view. This measurement shows how strong the code power is relative to the power of the wideband channel RSSI. However, in a UMTS network, the UE normally receives signals from multiple base stations, all transmitting on the same frequency. Therefore, it is possible that even at a location close to a base station, with a high RSCP, no login is possible, due to high interference levels from a second nearby base station. This effect is called “pilot pollution” and network planners try to avoid too close spacing of base stations to minimize regions where it can occur [22].

RSSI

Received Signal Strength Indicator (RSSI) is the received wideband power in the downlink. This index shows how strong the wideband (5MHz). The value of the RSSI can be easily calculated from CPICH_RSCP and the CPICH_EcNo value using the following formula:

RSSI = CPICH_RSCP - CPICH_EcNo

Where: CPICH_RSCP and RSSI are in dBm, while CPICH_EcNo is in dB.

2.4 Coverage in UMTS

In order to perform the UMTS coverage prediction process, the actual network status and measurements need to be captured. In practice, the network operator can collect diverse information about the network status from different sources such as [23]: Network counters, Measurement reports and Drive tests.

Network counters: These are measurements taken at Node-Bs and/or Radio Network Controllers (RNC), such as load levels, number of users in soft/softer handover, etc. that are transferred to the O&M module.

Measurement reports: These are measurements taken by the mobile terminals while in an active connection and transmitted live to the RNCs (e.g., received power of the pilot channel, pilot channel energy per chip over total received power density for the serving and neighboring cells, etc.). The primary usage of these reports is as input for Radio Resource Management (RRM) algorithms, such as a handover algorithm. In order to use these measurements for coverage and optimization operator need additional software to able to store and forward them to the O&M.

Drive tests: These are measurements carried out by one or several specialized terminals able to record a number of parameters while following a trajectory in the field. They are equipped with a Global Positioning System (GPS) so that the position where each measurement has been taken is also recorded.

3 Spatial Interpolation Techniques

In this chapter, the main concepts that are important to construct this thesis are discussed. Such concepts are spatial interpolation and data exploration methods explained. A brief description of different types of spatial interpolation techniques with their mathematics also discussed. Finally, evaluation methods of different interpolation methods are followed.

The spatial interpolation is designed to estimate the value of RSCP, SNR in the wireless network ,or any other parameters at a given geographical location based on the observations at neighboring locations [24]. The estimation of unknown values based on a theoretical foundation is the Tobler's first law of geography. Which is 'Everything is related to everything else, but near things are more related than distant things'. That means the adjacent measuring points are more similar than those measuring points that are far away from each other [25] [8]. There are two main groupings of interpolation techniques: deterministic and geostatistical. All methods rely on the similarity of nearby sample points to create the surface. Deterministic techniques use mathematical functions that form weighted averages of nearby measured values to create the surface, while geostatistical techniques use both mathematical and statistical methods [26]. The different interpolation methods give the best results when the data is normally distributed. So we to have check the statistical behaver of the data by using data exploration techniques.

3.1 Exploratory Data Analysis (EDA)

The goal of exploratory data analysis (EDA) is to create a comprehensive description of the data as well as discover general insights about the data. Data exploration is the process, which is important to examine the data characteristics in different ways. It is an important step in spatial interpolation work. The different techniques for such analysis are Histogram, Normal QQPlot, Trend Analysis, Spatial autocorrelation [24] [27].

3.1.1 Histogram

The histogram tool provides a univariate (one-variable) description of the data. The tool displays the frequency distribution for the dataset of interest and calculates summary statistics [24]. The frequency distribution is a bar graph, which describes how often observed values fall within certain intervals or classes. The histogram tool offers descriptive statics such as variance, standard deviation, median, mean, skewness, kurtosis, coefficient of variation, minimum and maximum values. *Coefficient of skewness*, which is a measure of the symmetry of a distribution. And *kurtosis* is based on the size of the tails of the distribution and provides a measure of how likely the distribution will produce outliers.

3.1.2 Normal QQ Plot

QQ plots are graphs on which quantiles from two distributions are plotted relative to each other. General QQ plots are used to assess the similarity of the distributions of the two datasets. Points on the normal QQ plots provide an indication of the univariate normality of the dataset. If the data is normally distributed, the points will fall on a 45-degree reference line, and if the data is not normally distributed, the points will deviate from the reference line [24]. If the data is not Normal distributed apply log transformation techniques to transform into normal distribution.

3.1.3 Trend Analysis

The trend analysis helps to identify the trends in the input dataset. The tool provides a three dimensional perspective of the data. The locations of the sample points are plotted on the x, y plane. The value is given by the height of a stick in the z dimension above each sample point. A unique feature of the trend analysis tool is that the values are then projected on to the x, z plane, and y, z plane as scatterplots. This can be thought of as sideways views through the three-dimensional data. Polynomials are then fitted through the scatter plots on the projected plane. An additional feature is that data can be rotated to isolate directional trends.

3.1.4 **Spatial Autocorrelation**

Spatial autocorrelation helps understand the degree to which one object is similar to other nearby objects**.** Moran's I (Index) measures spatial autocorrelation. Before creating the Kriging model, make sure the data was spatially autocorrelated. To do so, we used ArcMap to perform *a Global Moran's I* test on the data to determine whether the data exhibits spatial autocorrelation. Moran's I can be classified as positive, negative and no spatial auto-correlation. Positive spatial autocorrelation is when similar values cluster together on a map. Negative spatial autocorrelation is when dissimilar values cluster together on a map.

3.2 Deterministic Interpolation Techniques

Deterministic interpolation techniques based on the surrounding measured values or on specified mathematical formulas that determine the smoothness of the resulting surface [24]. Deterministic interpolation techniques create surfaces from measured points, based on either the extent of similarity (e.g., Inverse Distance Weighted) or the degree of smoothing (e.g. Spline, radial basis functions) [24]. Inverse Distance Weighted, Spline, and radial basis functions are some types of deterministic interpolation techniques. For this thesis, Inverse Distance Weighted is selected to estimate coverage signal power in the UMTS network.

Inverse Distance Weighted (IDW)

Inverse Distance Weighted interpolation explicitly implements the assumption that things that are close to one another are more alike than those that are farther apart [24]. To predict a value for any unmeasured location, IDW will use the measured values surrounding the prediction location. Those measured values closest to the prediction location will have more influence on the predicted value than those farther away. Thus, IDW assumes that each measured point has a local influence that diminishes with distance. The predictor of IDW is formed as a weighted sum of the data

The general formula of IDW predictor is [24]:

$$\hat{Z}(s_o) = \sum_{I=1}^{N} \lambda_i Z(s_i) \quad (3.1)$$

Where: $\hat{Z}(s_o)$ is the value which is predicted for a location s_o; N is the number of measured sample points surrounding the predicted location; λ_i are the weights assigned to each measured point; $Z(s_i)$ is the known value at the location s_i.

The weight of the standard known point against the predicted point will exponentially decrease with the increase in their distance [28]. The weight is determined according to Equations (3.2):

$$\lambda_i = \frac{d_{io}^{-p}}{\sum_{i=1}^{N} d_{io}^{-p}} \quad (3.1)$$

Where: p *is* the power exponent; d_{i0} is the distance between the prediction point s_o and the known sample point s_i.

As shown in Equation (3.2), as the distance d_{io} becomes larger, the weight reduced by a factor of *p*. where p controls the influence of the distance among sample points on the interpolation results. The power variable decides how the surrounding points affect the estimated value. A lower power results in higher influence from distant points [28]. Therefore IDW is an interpolation method that takes the distance between a predicted point and sample points as the weight. The sample points are weighted by the inverse of their distance to the predicted point [29].

3.3 Geostatistical Interpolation Techniques

Geostatistical interpolation techniques estimate the unknown based on statistical models that include autocorrelation (statistical relationships among the measured points) [24]. As their name implies, geostatistical techniques create surfaces incorporating the statistical properties of the measured data. Geostatistics includes a number of methods and techniques to analyze the variability of spatially distributed or spatially structured regionalized variables [30]. Because geostatistics is based on statistics, these techniques produce not only prediction surfaces but also

error or uncertainty surfaces, giving an indication of how good the predictions are [24]. Kriging is a statistical version of interpolation.

Kriging is an interpolation technique based on the methods of geostatistics. Kriging was developed by Krige (1951), from whom the name kriging was derived, and Matheron (1963) to accurately predict ore reserves from the samples taken over a mining field [9]. Kriging utilizes the variogram, which does not depend on the actual value of the variable (data), rather its spatial distribution and internal spatial structure [29]. Kriging is a powerful type of spatial interpolation that uses complex mathematical formulas to estimate values at unknown points based on the values at known points. There are several different types of Kriging: Ordinary, Universal, CoKriging, and Indicator Kriging [31] [11]. For this thesis the most common type, Ordinary Kriging was selected.

Ordinary Kriging (OK)

OK is similar to IDW in that it weights the surrounding measured values to derive a prediction for each location. However, the weights are based not only on the distance between the measured points and the prediction location but also on the overall spatial arrangement among the measured points. To use the spatial arrangement in the weights, the spatial autocorrelation must be quantified [24] [7]. That means in OK, the information on spatial locations allows us to compute distances between observations and to model autocorrelation as a function of distance. In OK, an unknown value $\hat{Z}_u$ at point u is estimated as a weighted linear combination of n known samples. Therefore, the predictor of OK is formed as a weighted sum of the data, as follows [24] [9]:

$$\hat{Z}_u = \sum_{i=1}^{n} W_i Z_i \qquad (3.2)$$

Where: Z_i is the measured value at the i_{th} location; $\hat{Z}_u$ is predicted value for location *u;* W_i is an unknown weight for the measured value at the i_{th} location; u is the prediction location; n is the number of measured values.

This is the same type of predictor as for IDW interpolation [24]. However, in IDW, the weight, λ_i, depends solely on the distance to the prediction location. In OK, the weight, W_i, depends on the

distance to the prediction location, and the spatial relationships among the measured values around the prediction location. The estimation $\hat{Z}_u$ may be unbiased by choosing the weights W_i. Predicting numerous locations may give some values which could be below the real values and some above, so the sum of weights W_i must be equal to 1 in order to guarantee that the prediction of the unknown measurement is unbiased [1] [29].

$$\sum_{i=1}^{n} W_i = 1 \qquad (3.3)$$

The most significant task associated with the implementation of Equation (3.4) is the determination of the appropriate weights. In the OK estimation of the unknown value using Equation (3.4) follow the following procedures [24].

Calculate the empirical semivariogram: OK quantified the basic principles of Tobler's first law of geography here as spatial autocorrelation. The empirical semivariogram is a means to explore this relationship. Pairs that are close in distance should have a smaller measurement difference than those farther away from one another. This assumption represented by the empirical semivariogram.

Fit a model: a line that provides the best fit through the points in the empirical semivariogram cloud graph. Find a line such that the (weighted) squared difference between each point and the line is as small as possible. This is referred to as the (weighted) least-squares fit. This line is considered a model that quantifies the spatial autocorrelation in the data.

Create the matrices to determine the kriging weight. The equations for OK are contained in matrices and vectors that depend on the spatial autocorrelation among the measured sample locations and prediction location. The autocorrelation values come from the semivariogram model described above. The matrices and vectors determine the kriging weights that are assigned to each measured value.

Make a prediction: using weights for the measured values, which are found from the above step, determine a prediction for the location with the unknown value using Equation (3.4).

In general, any geostatistical analysis of the above four steps summarized into two major steps: a) semivariogram (autocorrelation) analysis, b) Kriging estimation (prediction) and mapping [1] [9] [30] [32]. This means that geostatistics uses the data twice: first to estimate the spatial autocorrelation and second to make the predictions [24]. More generally, that can be divided into four steps to complete geostatistical interpolation analysis and mapping of regionalized variables.

3.3.1 Semivariogram Modeling

Semivariogram modeling is a key step between spatial description and spatial prediction. Before creating semivariogram modeling, we have to talk about semivariogram. The semivariogram is a measure of how similar are the points in the space when they are farther apart [15]. In geostatistics examined and quantified, the autocorrelation is called spatial modeling, also known as structural analysis or variography [24]. Spatial modeling of the semivariogram, begin with a graph of the empirical semivariogram, computed as Equation (3.6).The empirical semivariogram provides information on the spatial autocorrelation of datasets.

Creating the empirical semivariogram

Semivariogram is a graph that allows analyzes the spatial behavior of a variable on a defined area, resulting in the influence of data at different distances [24]. According to the geostatistics, as the distance $h_{i,j}$ between two points i and j increases, the correlation between them is expected to decrease (i.e., $Cov(z_i, z_j) \leq Cov(z_i, z_k)$ if $h_{i,j} \leq h_{i,k}$) [33]. The semivariogram is a plot of semivariance as a function of distances between the observations (h) and is denoted by $\gamma(h)$ [25]. Assume that the value of the sample data point x in the studying area is Z(x) and the sample data value of point $x+h$ is Z(x +h). To create an empirical semivariogram, determine the squared difference between the values for all pairs of locations. When these are plotted, with half the squared difference on the y-axis and the distance that separates the locations on the x-axis, it is called the semivariogram cloud. Semivariogram, computed with the following equation for all pairs of locations separated by distance h [1] [24].

$$\gamma_{(h)} = \frac{E\left[Z_{(x)} - Z_{(X+h)}\right]^2}{2} \quad (3.4)$$

Where: $\gamma(h)$ is the semivariogram functions ; h the distance between x and x+h; $Z_{(x)}$ and $Z_{(x+h)}$ are the values of the random variable Z of interest at locations (x) and (x + h). E [ ] is the statistical expectation operator.

Where h is the distance between each pair of locations in the xy coordinate system. Euclidean distance is used to calculate the distance between two locations as shown in Equation (3.6) [1] [25].

$$h_{i,j} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \quad (3.5)$$

Binning the empirical Semivariogram

When datasets are large, the number of pairs of locations will grow quickly so that pair of points in the semivariogram cloud can be difficult to interpret due to the complete number of point pairs (for example, 465 point pairs created from just 50 sample points, and this just for 1/3 of the maximum distance lag.). A common approach to resolving this issue is to use binning. Binning is the process of reducing the number of points in the empirical semivariogram, by grouping the pairs of locations based on their distance from one another [24]. The pair of points into intervals called lags and to summarize the points within each interval and the plot is referred to as the *empirical variogram* plot [24]. The empirical semivariogram is a graph of the averaged semivariogram values on the y-axis and the distance (or lag) on the x-axis.

Some time the averaged semivariogram is called an experimental semivariogram. Which contains a set of discrete data points of distance and semivariance. It provides useful information for interpolation, optimizing sampling and determining spatial patterns. It is clear that the number of semivariogram samples is proportional to the available pairs of locations. Its computational formula as follows.

$$\bar{\Upsilon}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{Z(x_i) - Z(x_i + h)\}^2 \qquad (3.6)$$

Where: $Z(x_i)$ is the value at location x_i , $Z(x_i + h)$ is the value at location x_i +h , and N(h) is the number of pairs of points of observations of the values of attribute $Z(x_i)$, $Z(x_i+h)$ separated by distance h ; the plot of $\bar{\Upsilon}(h)$ against h is known as the *empirical semivariogram* [1].

Fitting a model to the empirical semivariogram:

The next step is to fit a model to the points forming the empirical semivariogram. The empirical semivariogram provides information on the spatial autocorrelation of datasets. However, it does not provide information for all possible directions and distances. For this reason, and to ensure that Kriging predictions have positive Kriging variances, it is necessary to fit a model that is, a continuous function or curve to the empirical semivariogram. Abstractly, this is similar to regression analysis, in which a continuous line or curve is fitted to the data points. The fitting model is obtained by plotting the empirical semivariance $\gamma(h)$ versus the lag distance (separation distance of the pairs) [23]. As previously discussed, the semivariogram depicts the spatial autocorrelation of the measured sample points. Once each pair of locations is plotted after being binned, a model is fit through them. Range, sill, and nugget are commonly used parameters to describe those models as shown in Figure (3.1). The definition and detail explanation of the parameters as follows.

The range (a) is the distance at which semivariance stops increasing. Therefore, range indicates the distance from which samples are spatially independent of each other. The *nugget* (C0) is the variance not explained by the model and is calculated as the intercept with the Y-axis. Also known as the error of variance, because the variance of two points separated by zero meters should be zero. That is why this variance is usually indicating a smaller scale variability of the sampled. The maximum semivariance found between pairs of points is known as *Sill*.

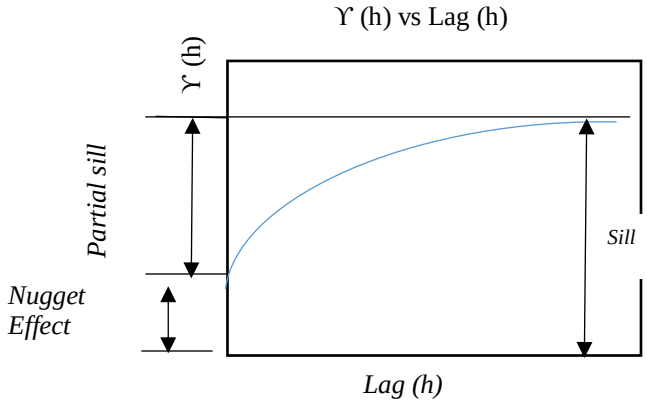


Figure 3.1: Illustration of semivariogram parameters (Range, Sill, and Nugget) [1]

Different types of semivariogram models

The Geostatistical Analyst provides the following functions to choose from to model the empirical semivariogram: Circular, Spherical, Tetraspherical, Pentaspherical, Exponential, Gaussian, Rational Quadratic, Hole Effect, K-Bessel, J-Bessel, and Stable. The selected model influences the prediction of the unknown values. That are used in the Kriging method to help interpolate data accurately [1]. The main models which use in geostatistical calculations are the spherical, the exponential and the gaussian. The equations bellow, C_0, C_1, and a represent the nugget effect, the structural variance and the range, respectively.

For the Spherical model, usually symbolized as Sph (C_0, C_1, a), the equation is:

$$\gamma(h) = C_0 + C_1\left(1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right), h \leq a \qquad (3.7)$$

$$\gamma(h) = C_0 + C_1, h > a$$

The Exponential model, symbolized as Exp (C_0, C_1, a), the equation is:

$$\gamma(h) = C_0 + C_1\left(1 - EXP\left(\frac{-3h}{a}\right)\right) \quad (3.8)$$

The Gaussian model, symbolized as Gau(C_0, C_1, a), the equation is:

$$\gamma(h) = C_0 + C_1\left(1 - EXP\left(-3\left(\frac{h}{a}\right)^2\right)\right) \quad (3.9)$$

3.3.2 Spatial prediction using Kriging

After computing the dependence or autocorrelation the data (see the above section) and using the spatial information in the data to compute distances and model the spatial autocorrelation, now make a prediction using the data with the fitted model. Weights are calculated using the following simultaneous equations [9] [33]:

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \lambda \end{pmatrix} = \begin{pmatrix} \gamma(h_{1,1}) & \cdots & \gamma(h_{1,n}) & 1 \\ \vdots & & \ddots \ \vdots & 1 \\ \gamma(h_{n,1}) & \cdots & \gamma(h_{n,n}) & 1 \\ 1 & & \cdots \ 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(h_{1,u}) \\ \vdots \\ \gamma(h_{n,u}) \\ 1 \end{pmatrix} \quad (3.10)$$

Where**:** $\Upsilon(h_{i,j})$ is a semivariogram which is a function of distance; $h_{i,j}$ is the distance between the known points i and *j*; $h_{i,u}$ is the distance between known points i and unknown points u and λ is the Lagrange multiplier to minimize the kriging error.

Ordinary kriging assumes that the mean is constant in the local neighborhood of each estimation point. As a result, the expected value of estimation error at an unknown point μ is zero (*i.e.,* $E(\hat{Z}_u - z_u) = 0$). The weights determined by equation (3.9) are called optimal since they minimize the variance of estimation error(i.e., $Var(\hat{Z}_u - z_u)$) [9] [33]. Where $\hat{Z}_u$ is the true value and z_u is the predicted value.

3.4 Accuracy Assessment Interpolation Methods

In this section, a method of different prediction accuracy assessments of various interpolation techniques discussed. There are two common validation methods that have been widely suggested for evaluating the accuracy of interpolation methods. The first method is a cross-validation method, namely as a 'leave-one-out technique'. The other validation method is called the split-sample method.

3.4.1 Cross-validation method

Cross-validation is a statistical technique of evaluating and comparing the performance of different interpolation methods [8] [11] [34]. This method involves omitting a single sample value from the entire dataset and performing an interpolation procedure by using the remaining dataset.

After the interpolation process, the difference between the actual and predicted values of the omitted point is then calculated. This procedure is repeated for n times, where n equal to the number of all samples.

3.4.2 Split-Sample method

In this method, the entire raw dataset is divided into two subsets: the training dataset and the test datasets. The training dataset is used to produce the interpolated surface, and the performance of each interpolation method is evaluated by comparing the difference between the predicted and observed values from the test dataset. Several statistical measurements are most widely used to evaluate the overall performance of interpolation methods, including root-mean-square error (RMSE), mean absolute error (MAE) and coefficient of determinations [28].

The following mathematical equations show how to calculate these statistical measurements:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(Z_i - Z(x)_i)^2}{N}} \quad (3.11)$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|Z_i - Z(x)_i|}{n} \quad (3.12)$$

Where: Zi is the actually observed value of at sampling point i (i=1, 2, 3…n); $Z(x)_i$ is the estimated value and n is the number of sample values for testing. The value of MAE tells whether or not there is a systematic error in interpolation results, while RSME indicates the deviation between interpolated values and the actual values [25].

The coefficient of determination (R^2) is used to measure the correlation between the predicted value and the measured value [11].

$$R^2 = \frac{[\sum_{i=1}^{n}(Z_i - Z_{ave})\,(P_i - P_{ave})]^2}{\sum_{i=1}^{n}(Z_i - Z_{ave})^2 \sum_{i=1}^{n}(P_i - P_{ave})^2} \quad (3.13)$$

Where: Z_i and P_i are actually estimated value and measured value at sampling point i respectively; Z_{ave} is the averaged estimated value, P_{ave} is the averaged measured value; and n is the number of values used for estimation.

4 Experimental Analysis

In this chapter, the overall workflow (experimental process) followed for conducting this thesis is presented. The system model, which describes this study as shown in Figure 4.1. It consists of three main modules: Data collection, data preprocessing and prediction based on spatial interpolation techniques. Details of tasks done under these modules are described separately in the next subsection. Especially the interpolation processes are discussed in detail.

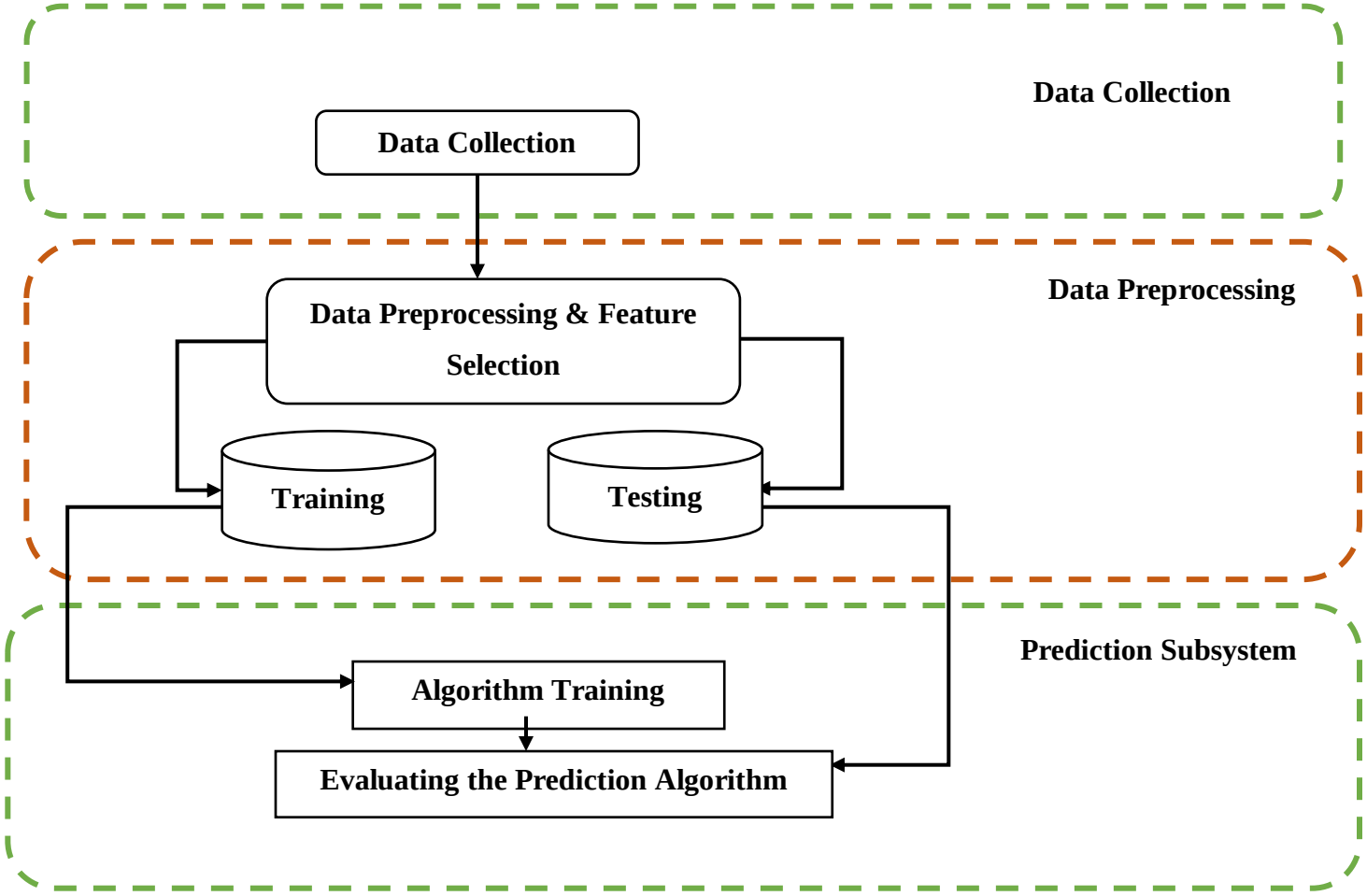


Figure 4.1: Overall Experimental Process [35]

4.1 Data Collection and Experimental Scenario

4.1.1 Data Collection

Hence, a georeferenced CPICH RSCP dataset is needed to train the spatial interpolation algorithms for coverage estimation. The first task done is to collect such data from drive test and site information collected from the ethio telecom UMTS mobile network. The data collected from the driving test is log file and analyzed using (analyzer tool). After analyzing the data collected from the drive test, the following network quality parameters and indicators can be obtained including GPS location coordinates. Radio link quality parameters, mainly CPICH RSCP and CPICH Ec/Io.

Table 4.1 shows sample data collected from drive test. It consists of the following parameter measurement. Where RSCP is the signal level at specific location in longitude and latitude, which indicates coverage; Serving Sector ID is the cell id of Node-B; Ec/No is the signal quality indicator. Table 4.2 shows sample site information collected from ethio telecom Operation and Maintenance Center (OMC).

Time	Distance	Longitude	Latitude	Ec/No	RSCP	Sector ID
Apr/06/19 02:08:09	19.30283	38.73808	8.99513	-7.45	-64.55	53273
Apr/06/19 02:08:10	25.300488	38.7379	8.99508	-13.3	-80.75	53273
Apr/06/19 02:08:11	34.981277	38.73785	8.99508	-13.6	-74.7	53273
Apr/06/19 02:08:12	44.950878	38.73774	8.99509	-8.55	-75	53273
Apr/06/19 02:08:13	55.104805	38.73765	8.99512	-13.7	-77.2	53273
Apr/06/19 02:08:14	65.567406	38.73752	8.99514	-13.6	-79	53273

Table 4.1: Sample data collected from drive test

Site_ID	Cell_Id	NodeB_name	Latitude	Longitude
111403	52743	CAAZ_Temama Buliding	9.005336	38.746
111403	52742	CAAZ_Temama Buliding	9.005336	38.746
112091	55523	CAAZ_(in front of Sudan Embassy)	9.00885	38.74434
112091	55522	CAAZ_(in front of Sudan Embassy)	9.00885	38.74434
112091	55521	CAAZ_(in front of Sudan Embassy)	9.00885	38.74434

Table 4.2: Site location data

4.1.2 **Experimental Real Network Scenario**

The ethio telecom UMTS network consists of 7500 number of Node-Bs throughout the country. Among them, 744 Node-Bs that incorporate 7159 cells have been deployed in Addis Ababa where the capital city of Ethiopia [36]. It is better to select a few UMTS network deployment sites and perform the cell coverage prediction by using the interpolation method. So, for this thesis analysis purpose, a 4km by 3km area selected from the central part of Addis Ababa. For this study, 26 Node B sites were nominated to create a real network scenario for the experiment, which is accessed by the drive test. The selected area satellite view in Figure 4.2. Totally we obtained 4427 coverage related datasets with longitude and latitude from drive test as shown in Figure 4.3. The color shows the level of received signal power. Thus, across the route on the map, the line shows the route has a specific color for each measurement point.

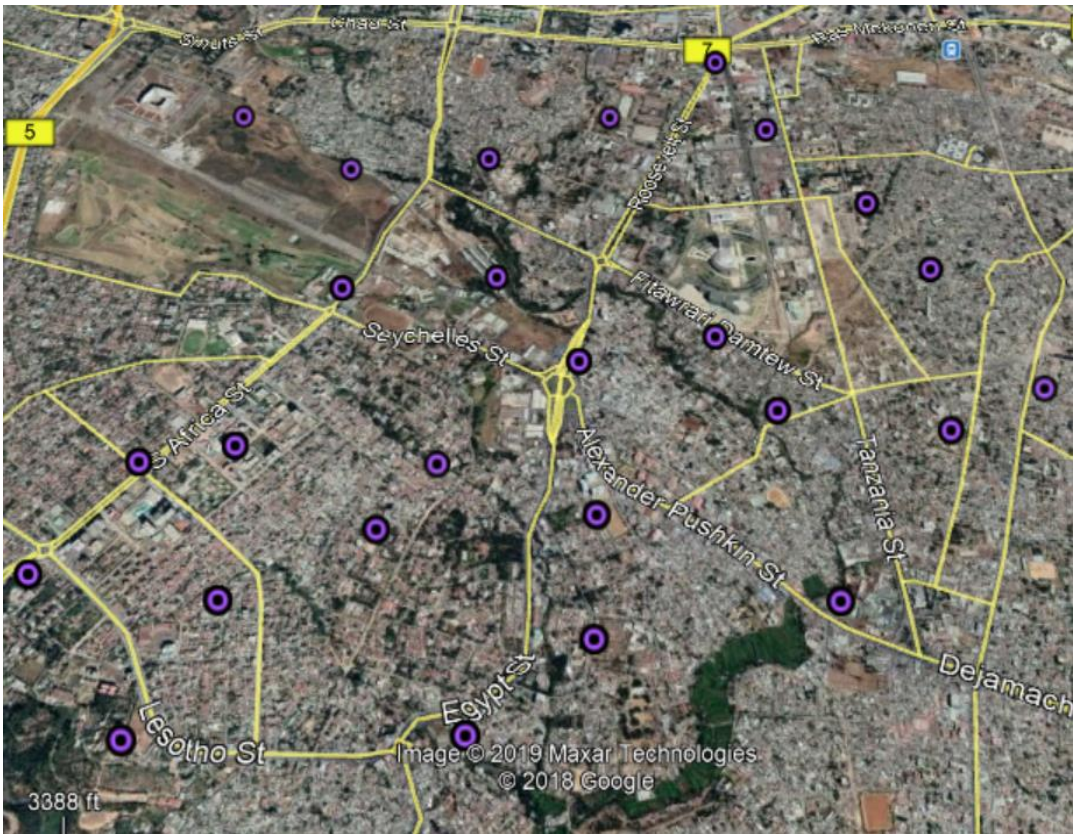


Figure 4.2: Satellite image of the selected sites

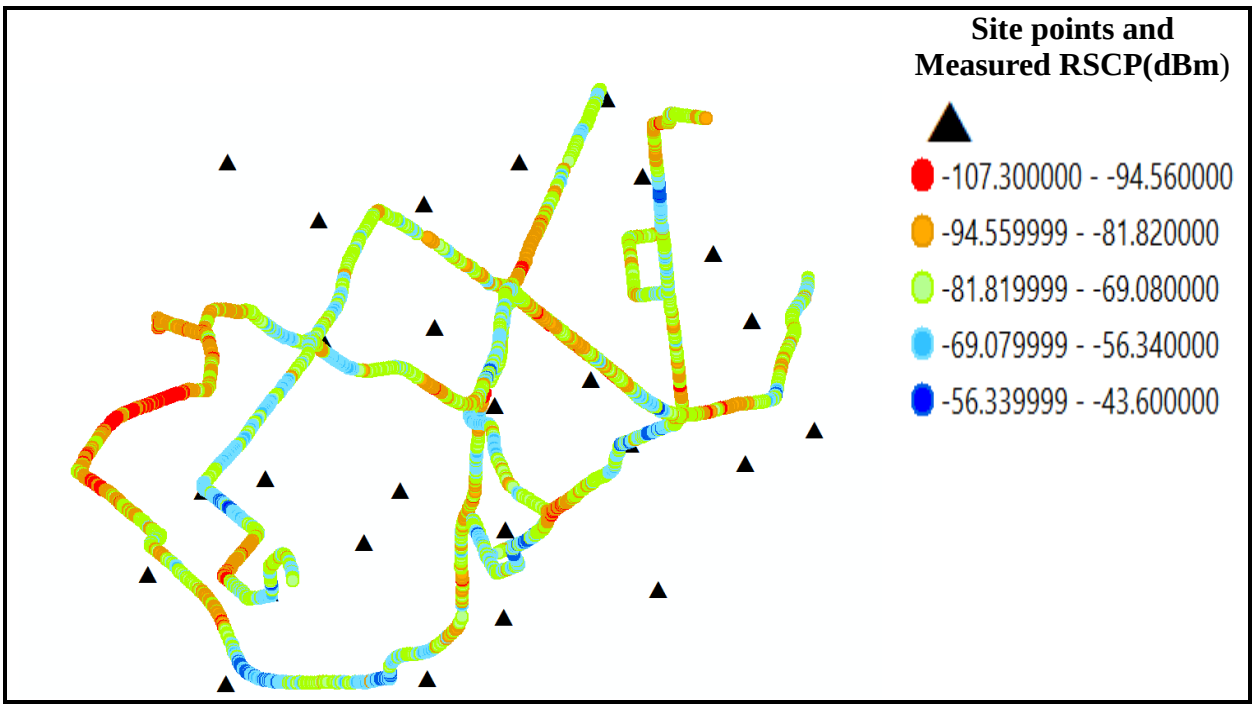


Figure 4.3: RSCP measurements in the study area

4.2 Data Preprocessing

Data preprocessing is the process of consisting of tasks, such as handling of missing data, and integration of multiple data sources, attribute/feature selection. In this research, handling noisy data (outliers) and the integration of multiple data sources are not done as part of data preprocessing. However, attribute/feature selection for coverage prediction and handling of missing data are performed as a data preprocessing task. Details of tasks done under this module; data cleaning, manual attribute selection, and training and testing point separation are discussed below.

4.2.1 Feature selection

Since the data has many parameters and contains information that is not significant for this thesis purpose, it needs to be preprocessed before using it as an input to a spatial interpolation algorithm. The data collected from the drive test consists of more than 8 parameter features. Out of these

features, only four of them namely Latitude, Longtidu, CPICH RSCP, and serving cell Id have been selected for spatial interpolation techniques.

The selection of these attributes is based on assessing related works[1] [3] [14] [16] [32]. The RSCP data collected from the drive test is normally geotagged (data per location) which contains information tabulated in Table 4.4. The thesis considers the coverage prediction per location. Short description of the selected attributes is given below:

Geotagged Data	
Selected Attributes	**Description**
Measurement location	Longitude and Latitude information
RSCP per location	Received signal code power per each location

Table 4.3: Description of the attributes

Longitude	**Latitude**	**RSCP**	**Serving SectorID**
38.74404	9.00929	-76.2	55521
38.74404	9.00928	-71.25	55521
38.74402	9.00925	-75.27	36276
38.744	9.00921	-74.2	36276
38.74397	9.0092	-75.35	36276
38.74395	9.00917	-77.5	36276

Table 4.4: Sample Selected featured Data

4.2.2 Data cleaning

Data collected from the drive test have missing value represented by value of zero as shown in Figure 4.4. The figure shows the scatter plot of RSCP vs longitude. So it is important removing this data from the collected data in order to improve the performance of interpolation algorithms. After data cleaning, we get 4405 dataset samples as shown in Figure 4.5. The color shows the value of the received signal power in each area.

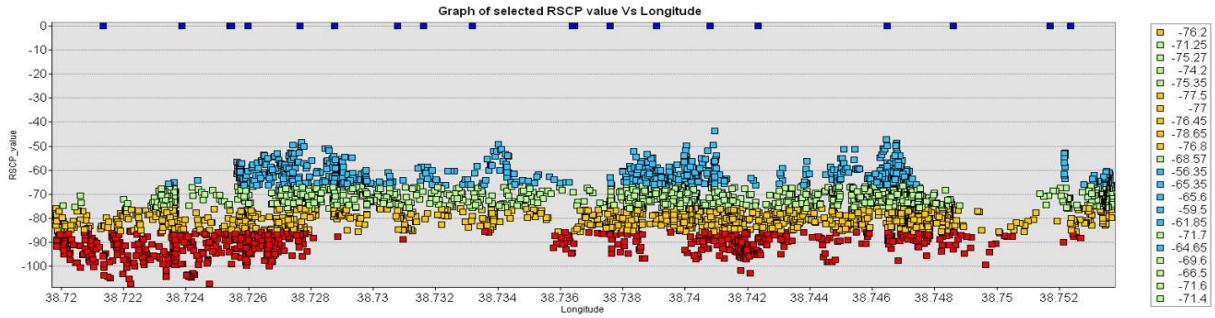


Figure 4.4: Collected RSCP datasets before cleaning

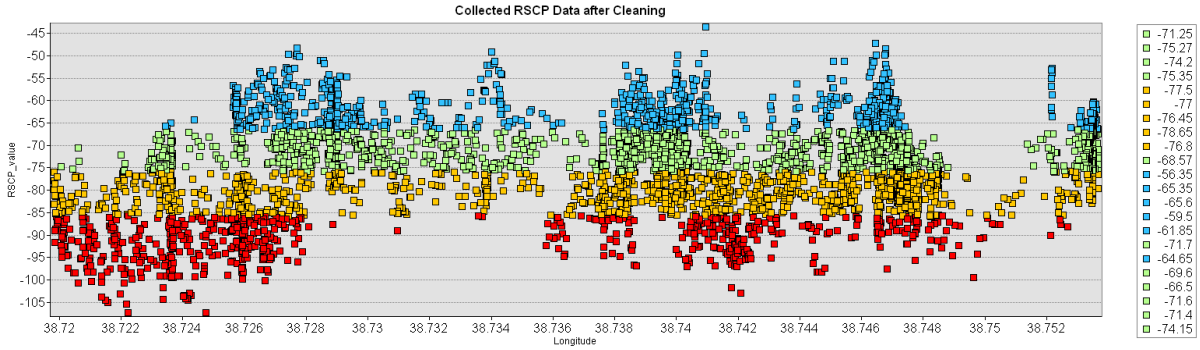


Figure 4.5: Cleaned Data Scatter plot

4.2.3 **Test Points Selection**

After data cleaning and feature selection preprocessing completed, now the collected datasets are partitioned into training and testing sets with a ratio of 75% to 25% respectively. The 75% partition means that a point is randomly selected for training with a probability of 0.75 and for testing with 0.25 [9]. We selected 1101 points as test points and other 3320 points were used for training the prediction algorithms. Test points are used to be able to validate the interpolation techniques performance.

4.3 Prediction Subsystem

This is the main part of cellular coverage prediction work and it is done based on spatial interpolation algorithm. Once the data preprocessing and feature selection task is completed, the next step is to estimate the unknown by creating a surface (map) by using the spatial interpolation

algorithms and select a prediction model. The process of estimating the unknown measurement from the known we called it prediction process. This subsection has three main parts: Data exploration, creating coverage prediction map (interpolation techniques process) and finally evaluating the interpolation algorithm performance. Two spatial interpolation algorithms; IDW and OK have been selected for this thesis to compare which one is best for wireless coverage prediction.

4.3.1 Data Exploration

Before using the interpolation algorithm, we have to check the data to have a normal distribution, needs to be autocorrelated, and cannot have any trends. Since some spatial interpolation modeling works best when the distribution of the data is close to normal, it is necessary to check for normality before performing interpolation [24]. For this thesis, we explored the distribution and the trend of the data using Geostatistical Analyst tools in ArcGIS. The autocorrelation of the data explored by Moran's I (Index).

Normality check

We used the Histogram and Normal QQPlot graph to check the data is normal distributed or not. The shape of the histogram looks bell-shaped, the skewness of the data close to zero and the mean and median values are very close. This indicates the distribution of RSCP data is close to normal as shown in Figure 4.6(a). The Normal QQPlot is created by plotting data values versus the value of a standard normal. Since the plot of the data points is close to a straight line, the data is nornaly distributed as shown in Figure 4.6(b). Both results showed that the data is normally distributed.

(a) **(b)**

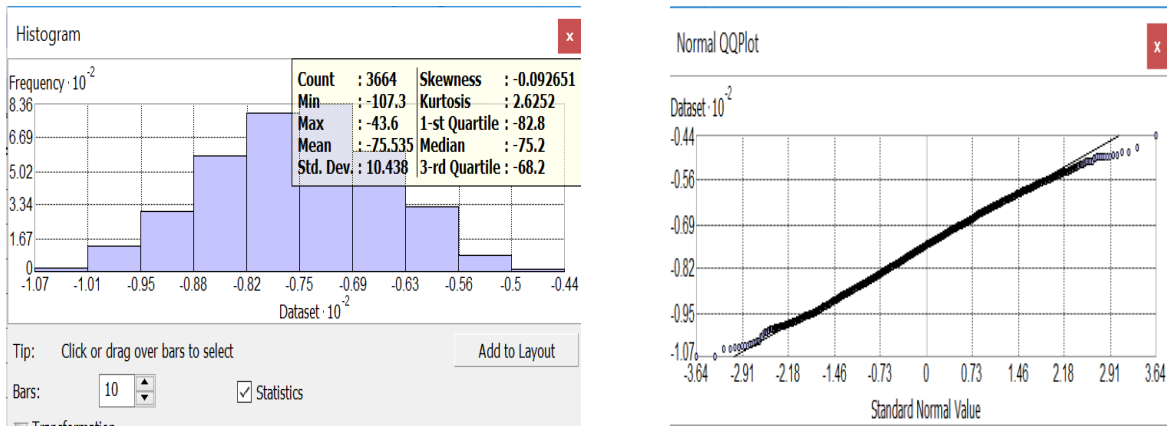


Figure 4.6: (a) Histogram distribution and (b) Normal QQPlot of RSCP data

Spatial Autocorrelation

The Global Moran's I statistic test is conducted to check the spatial autocorrelation of the data. The result of the spatial autocorrelation report shows that Moran's Index : 0.515 and Z-score is 101.2, since Moran's Index greater than zero shows that the data is spatially autocorrelated as shown in Figure 4.7.

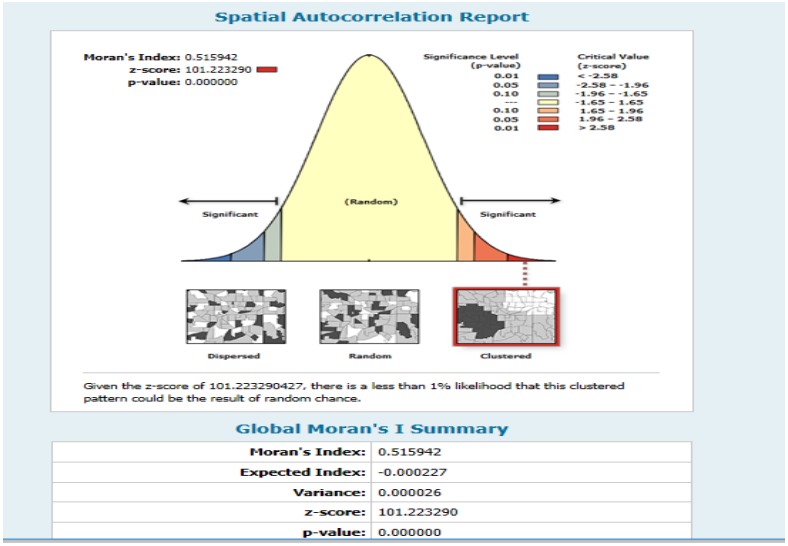


Figure 4.7: Spatial autocorrelation result of RSCP data

Trend Analysis

Trend analysis is conducted to check the data have trend or not. The result shows that the polynomials (blue and green lines) drawn through the projected points in x z and y z plan are flat this indicates the data don't have trend as shown in Figure 4.8. There is no trend when the lines are flat.

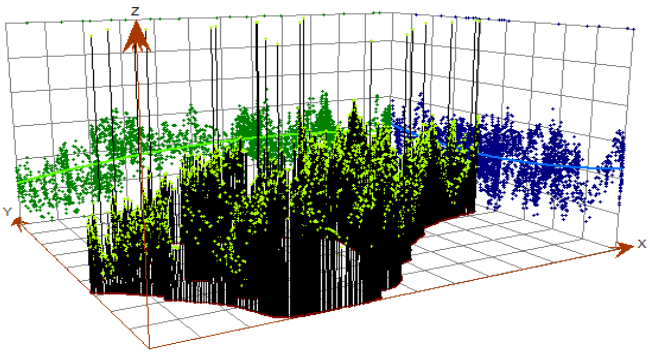


Figure 4.8: Trend Analysis of RSCP data

4.3.2 Interpolation Process

After the RSCP data explored, the spatial interpolation process followed. Two methods are examined in this study: IDW and OK. As shown in Figure 4.9 each algorithm follows different procedures to predict UMTS network coverage using RSCP measured data. The details of each procedure as explained below.

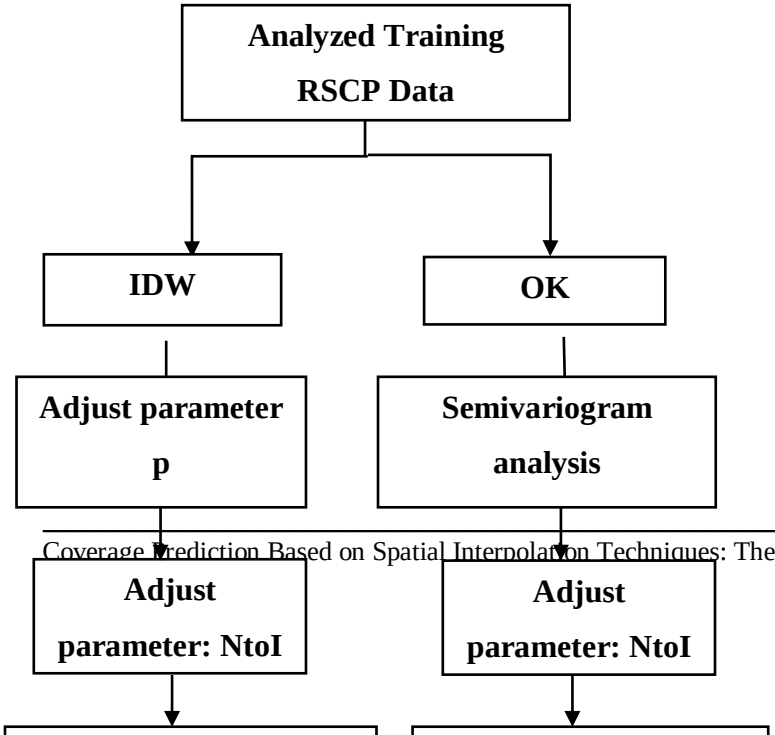


Figure 4.9: Workflow of Interpolation Process [34]

Let *S* be a set of known RSCP measurement points collected from drive test; and for each point *i* ∈ *S* , let *zi* and (*xi*, *yi*) denote the measured value and the *xy*-coordinates of the point, respectively. The problem is to estimate the value of RSCP at a point *u* where no measurement was taken. The procedures based on the OK and IDW to create the UMTS network coverage map by estimating the value of RSCP at an unknown point u as explained in the flow chart Figure 4.9.

4.3.3 IDW Method

IDW uses a simple algorithm based on distance. Power exponent (p) and Neighbor to include (NtoI) and Include at least (IatL) are parameters that determine the distance(d) for weight calculation in Equation (3.2). The IDW method followed the following procedures: parameter estimation and prediction using IDW.

Parameter Estimation using cross-validation

Optimum parameter value determined by using cross-validation. Parameter combination that has the least RSME to estimate z(u) is chosen.

Determine the optimum value of p

Optimal *P* value is determined by minimizing prediction error (RMSE) which is calculated from cross-validation.The smallest RMSE is determined as the optimal power *P* is the exponent of the distance. According to Equation (3.2) weights are proportional to the inverse distance raised to the power *P*. First *P* is changed from 1 to 3 without changing the other parameters to get optimum *P* value. Table 4.5 shows the effect of *P* on the interpolation precision. The result shows *P* changes from 1 to 3 and RMSE also increased accordingly. RMSE increased greatly when *P* increases from 2 to 3. However, when *P* changes from 1 to 2, it has less influence on RMSE. Therefore, the *P* value of 2 was chosen as the optimum value.

Power exponent(p)	**P = 1**	**P = 2**	**P = 3**
RMSE	4.972	5.029	5.251

Table 4.5: Prediction error of power exponents .

Determine the optimum number of Maximum and Minimum neighbors

To determine the optimum number of NtoI and IatL, *P* remained constant and NtoI changed from 10 to 20 and IatL from 2 to 10. The prediction errors among different parameter values are compared in Tables 4.6, 4.7 and 4.8 with the P, equal to 1, 2 and 3 respectively.

NtoI	NtoI = 10			NtoI = 15			NtoI=20		
IatL	2	5	10	2	5	10	2	5	10
RMSE	4.972	4.972	4.972	5.022	5.022	5.022	5.078	5.078	5.078

Table 4.6: Prediction error of NtoI and IatL with P=1

NtoI	NtoI = 10			NtoI = 15			NtoI=20		
IatL	2	5	10	2	5	10	2	5	10
RMSE	5.04	5.04	5.04	5.029	5.029	5.029	5.078	5.078	5.078

Table 4.7: Pridiction error of NtoI and IatL with P=2

NtoI	NtoI = 10			NtoI = 15			NtoI=20		
IatL	2	5	10	2	5	10	2	5	10
RMSE	5.25	5.25	5.25	5.243	5.243	5.243	5.243	5.243	5.243

Table 4.8 Prediction error of NtoI and IatL with P=3

To determine the optimum value of IatL, the variables P and NtoI remained constant. When IatL increased from 5 to 10, RMSE was not changed as shown in the above tables. Since IatL has no impact on the accuracy of interpolation, an arbitrary value from 2 to 10 is chosen. Therefore, the optimum number of minimum neighbors (IatL) is chosen which is 5.

Second, keeping the variables P and IatL as constants, RMSE decreases with NtoI increasing from 10 to 15. But RMSE increasing with NtoI increasing from 15 to 20. Thus, 15 are chosen as the optimum number of maximum neighbors(NtoI).

In summary, the optimum values of IDW interpolation method parameters are P=2, NtoI=15, and IatL=5 with the RMSE=5.025.

Prediction using IDW

After the value of the parameters is determined, the weights of the neighbors of z_u are calculated using Equation (3.2). The unknown value of RSCP (z_u) estimated using Equation(3.1). The weights of all neighbor points and the value of z_u are shown in Figure 4.10.

The prediction surface (map) is created by repeating this for many prediction locations and mapping the results as shown in Figure 4.11.

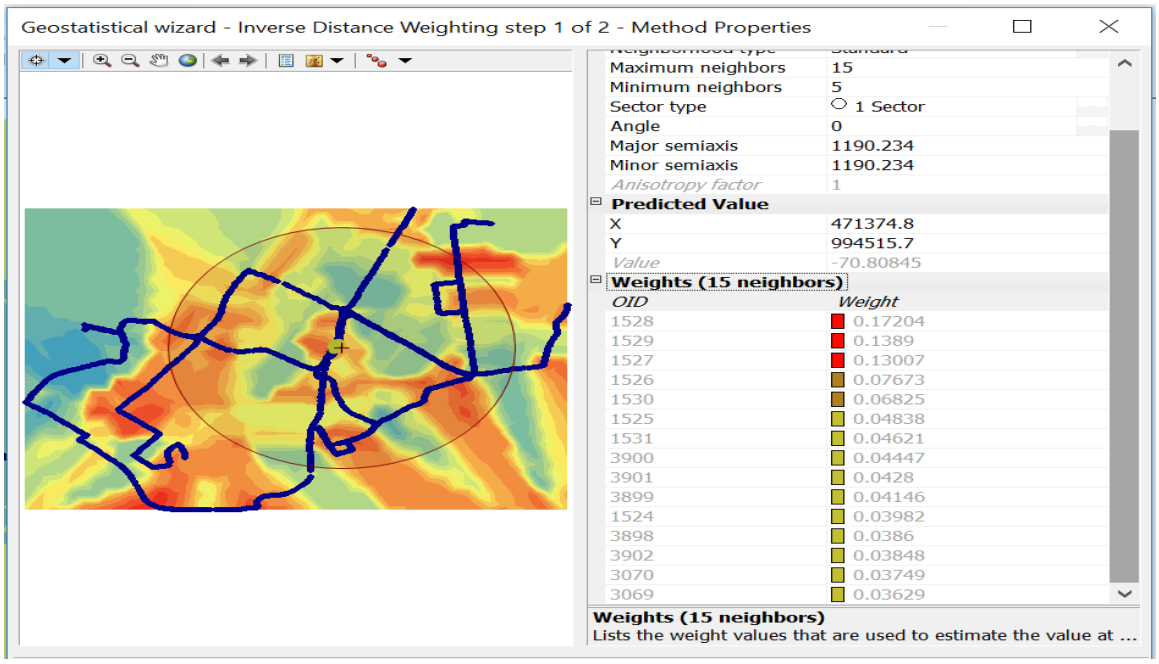
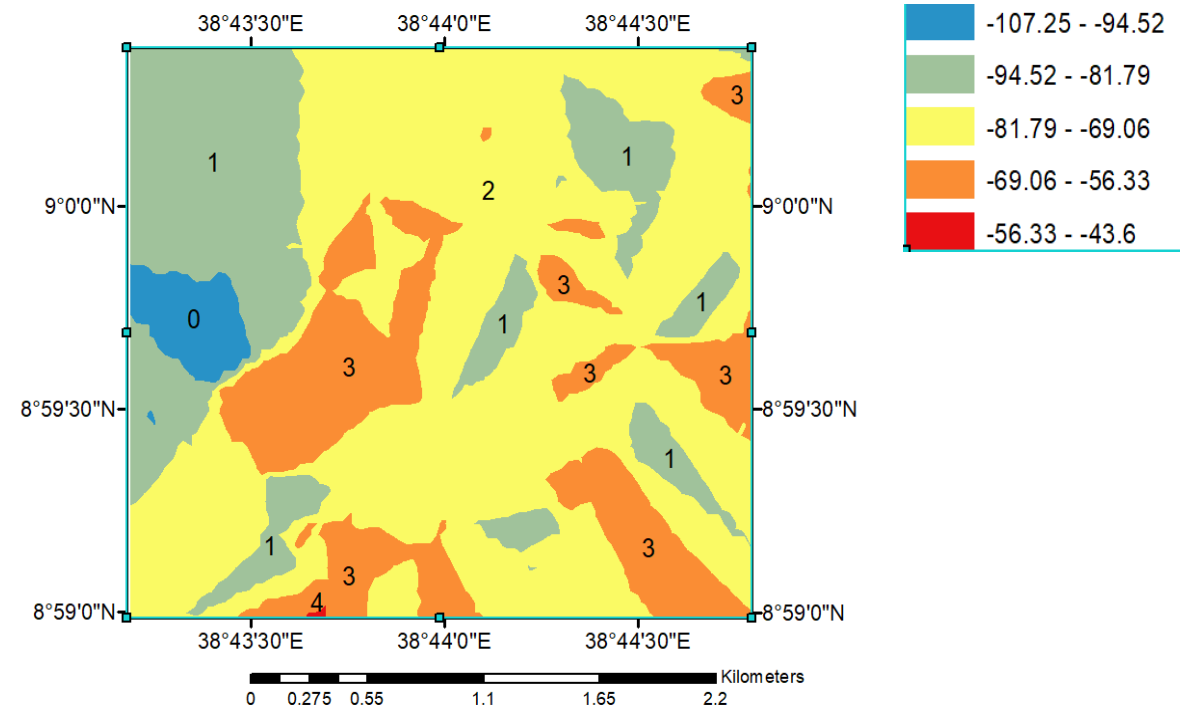


Figure 4.10: Weight of neighbors

The prediction map is presented in two formats: Contour map and Raster Map. Figure 4.11 (a) represents the contour map of the UMTS network coverage. It reveals the value in each area separated by means of contour lines that represent RSCP values where each line passes through points of the same RSCP value. Figure 4.11(b) shows the Raster map with the grids of equal-sized, square grid cells (sometimes termed pixels) with a size of 50m by 50m. Each grid contains a specific value of RSCP.

(a) Contour map **RSCP value(dBm)**



(b) Raster map

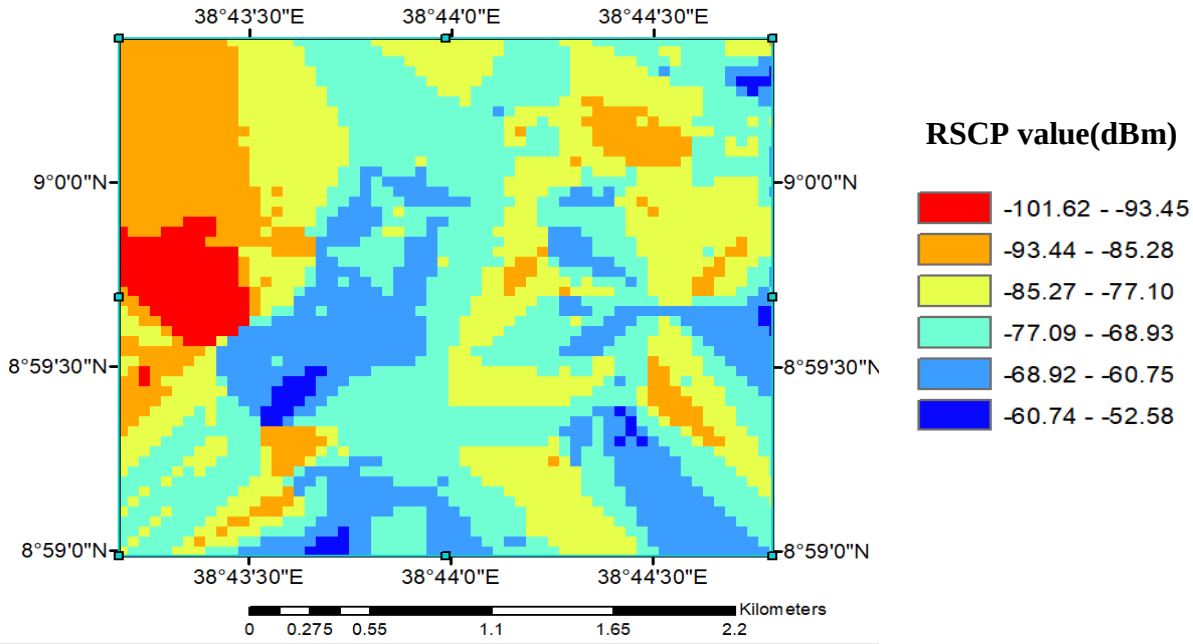


Figure 4.11: Prediction map by IDW (a) Contour (b) Raster

The distribution of RSCP for the selected study area by using the IDW technique. The distribution of RSCP value was classified into six classes by using equal intervals classification method. The result of prediction revealed UMTS network in the study area has good coverage (RSCP>-95) except zone zero which is represented in blue color in the contour map.

Evaluating the accuracy of prediction

The accuracy of prediction was evaluated by using the cross-validation tool. The prediction error of IDW is presented in two plots: predicted, and error plot for the RSCP training dataset as shown in Figure 4.12 (a)and (b) respectively. The predicted plot displays measured vs. predicted value and the error plot of IDW displays error vs. measured. As shown in figure IDW has an optimum error of prediction with the value of RMSE=5.02.

a) The Predicted plot of IDW **b) Error plot of IDW**

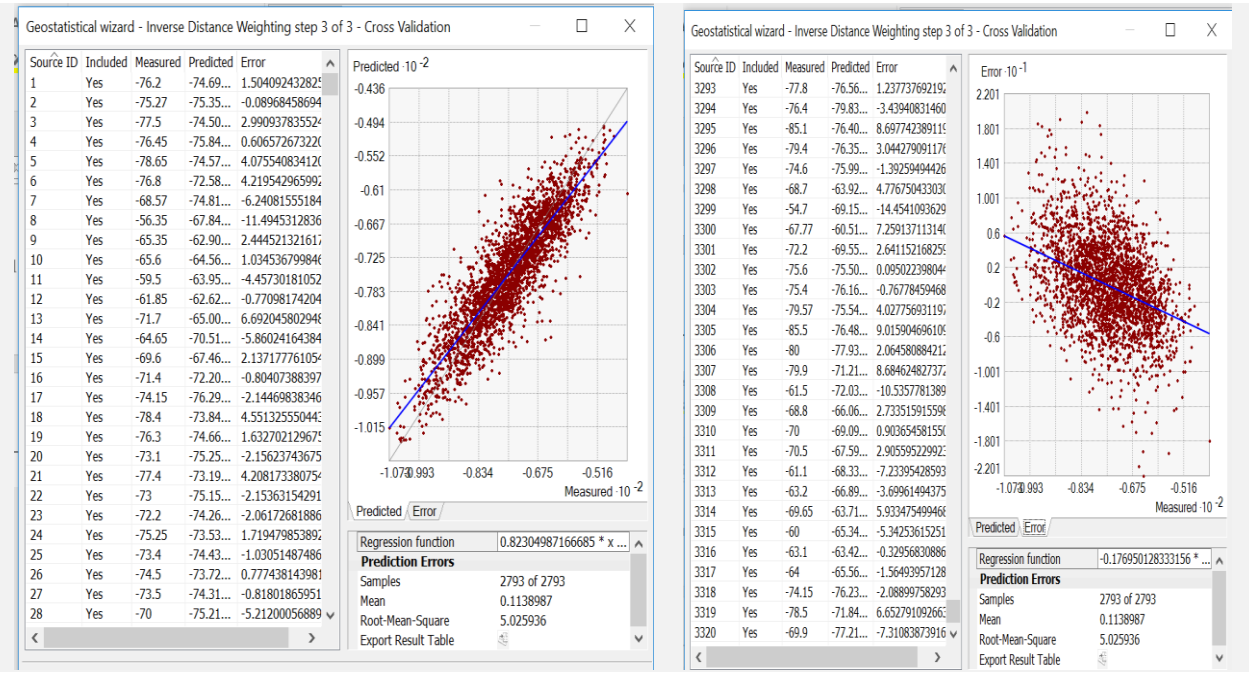


Figure 4.12: Error plot of IDW

4.3.4 **OK Method**

The OK method makes a prediction for locations in the study area based on the *semivariogram* and the *spatial arrangement of measured values* that are nearby and finally creates a continuous surface or map of the phenomenon.

The OK method has followed the four procedures to estimate the unknown z_u at u location: Semivariogram computation, searching neighbors around the predicted point, calculate the weight, and create a coverage map.

Semivariogram computation

The first step is modeling the measured RSCP data. The RSCP data was modeled using semivariogram/covariance modeling. The semivariogram fitting model was calculated by plotting the empirical semivariogram versus the distance as shown in Figure 4.13. The empirical semivariogram cloud shows that the number of pairs of locations will rise quickly when the dataset becomes large.

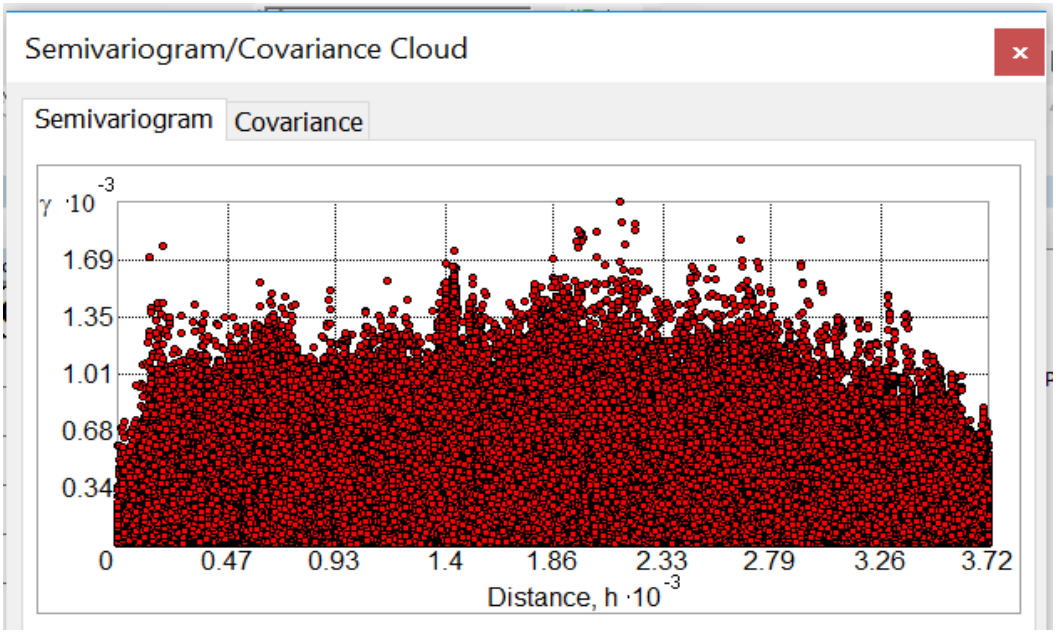


Figure 4.13: Empirical semivariogram cloud

When the dataset is large, the calculation of the semivariogram fitting model becomes difficult. As a result, the points of the semivariogram cloud were gathered into classes of distances ("bins") in order to find the semivariogram fitting model for the targeted area. Red dots in Figure 4.14 are the binned values. Red dots are generated by grouping (binning) semivariogram/covariance points. Blue crosses represent the average points, and these are generated by binning empirical semivariogram/covariance points that fall within angular sectors. Binned points show local variation in the semivariogram/covariance values, and smooth semivariogram/covariance value variation is shown by average values.

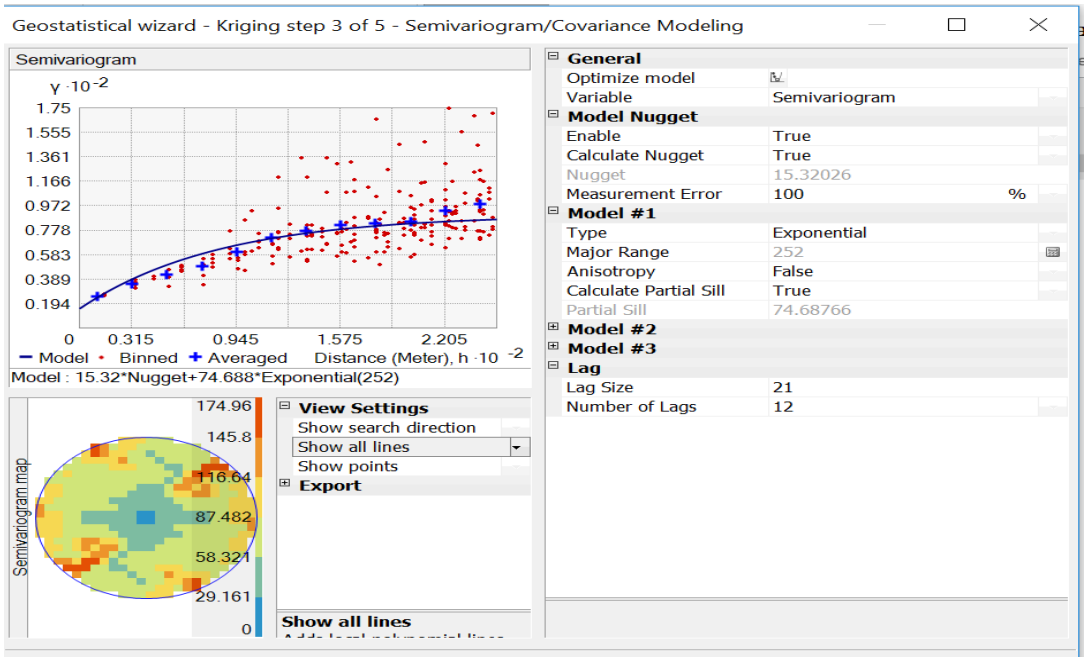


Figure 4.14: Exponential variogram model for the selected area

Select the fitting semivariogram model

The appropriate parameter value was determined by cross-validation. Parameters that have the least RMSE were selected.

The best model that fitted the points in the semivariogram cloud was determined by computing RMSE of candidate models as shown in Table 4.9. The three candidate models were compared with RMSE. The best-fitted model was selected that has the least RMSE.

Model	Nugget	Partial Sill	Nugget/Sill	Sill	Range	RMSE
Guasian	20.8	73.03	0.223	93.8	216.9	5.28
Spherical	20.3	74.9	0.213	95.2	252	4.91
Exponential	15.3	74.6	0.160	89.9	252	4.83

Table 4.9: Prediction Error of Semivariogram models

As shown in Table 4.9 the RMSE of the Exponential model (4.83) is less than that of the Spherical model (4.9) and the Gaussian model (5.28). Therefore, the Exponential model was selected as the best-fitted model for RSCP data.

After calculating the empirical variogram for the measured data points, a model should be fitted to the points. The Exponential fitting variogram model was selected based on the parameter selection as mentioned in Table 4.9 as the best fitted model for the measurements of RSCP data as shown in Figure 4.14. The selected semivariogram model determines the values of the parameters such as nugget, range, partial sill. The spatial relation of the data represented by the parameters nugget, range, and partial sill. This semivariogram model shows the values of the parameters such as: nugget = 15.3, range 252, partial sill = 74.6, Lag Size=21 and number of lags=12. Since the samples are random the average distance between neighboring samples used as Lag distance. Average distance determined by the Average Nearest Neighbor tool and result shows 21 meters is the mean distance between the points. The parameters for the models fitted to the semivariograms of RSCP data have some indication of the measured data. The range is that pairs of points that are greater apart are not spatially correlated. The value of the range is the 252-meter distance after which data are no longer correlated which means that signal strength values are correlated for short distance because the considered study area is an urban environment, where buildings obstruct signal propagation. In a continuous variable, we would expect that the nugget effect will be zero because at distance zero the values will be the same. However, in this case, the nugget effect has a value of 15 which shows there is some measurement error during data collection through the drive test. Therefore, semivariogram values are calculated using the selected exponential model using Equation (3.9).

Determine the optimum number of Maximum and Minimum neighbors

The second procedure is determining the optimum number of neighbor points (maximum and minimum) by using the searching neighborhood. The neighborhood search size defines the neighborhood shape and the number of points within the neighborhood that will be used in the prediction of an unmeasured location. Neighbor to include (NtoI) is the maximum number of neighbors and include at least (IatL) is the minimum number of neighbors around the estimated point *u*. The appropriate parameter value was determined by cross-validation. Parameters that have the least RMSE was selected.

NtoI	NtoI = 10			NtoI = 15			NtoI = 20		
IatL value	2	5	10	2	5	10	2	5	10
RMSE	4.837	4.837	4.837	4.838	4.838	4.838	4.838	4.838	4.838

Table 4.10: Prediction Error neighbors

To determine the optimum number of IatL, keep the number of NtoI unchanged. When the number of IatL was increasing from 2 to 10, the error of prediction (RMSE) was not changed as shown in Table 4.10. Therefore any number out of the compared was selected which is 5. The optimum number of NtoI was determined by increasing the number of NtoI from 10 to 20 and error of prediction. RMSE increases when NtoI increases from 10 to 15. However, it remained constant when NtoI increased from 15 to 20. The number of NtoI 10 was selected as the optimum number since it has the least prediction error as shown in Table 4.10.

Therefore, the parameter combination of OK interpolation is the exponential semivariogram model, NtoI is 10 and IatL is 5 are selected.

Prediction using OK

After the value of the parameters is determined, OK determined the weights of the neighbors of z_u using Equation(3.6). The weights for the measured locations were determined using the data configuration within the specified neighborhood, in conjunction with the fitted semivariogram model. The weights of all neighbor points and the value of z_u are shown in Figure 4.15.

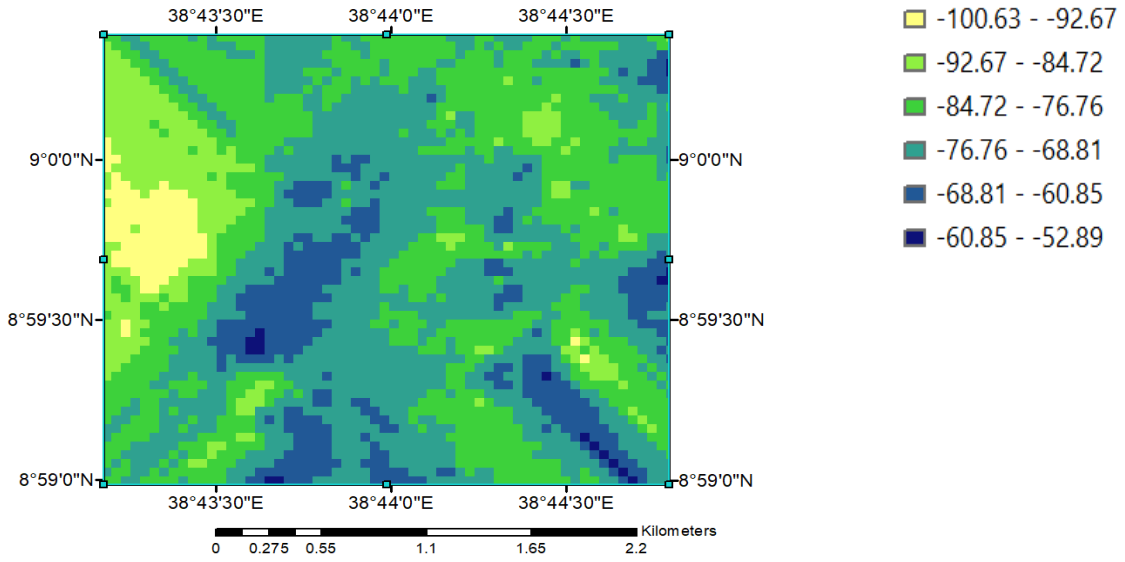
Figure 4.15: Searching neighborhood dialog box

The unknown value (z_u) for the prediction location *u* was predicted from the weights and the values of the known measured value using OK interpolation techniques. This process is performed for each spatial location to create a model of the continuous surface.

The prediction map is presented in two formats: Raster Map and Contour map as shown in Figure 4.16 (a) and (b) respectively.

a) Raster Map

RSCP value



b) Contour map

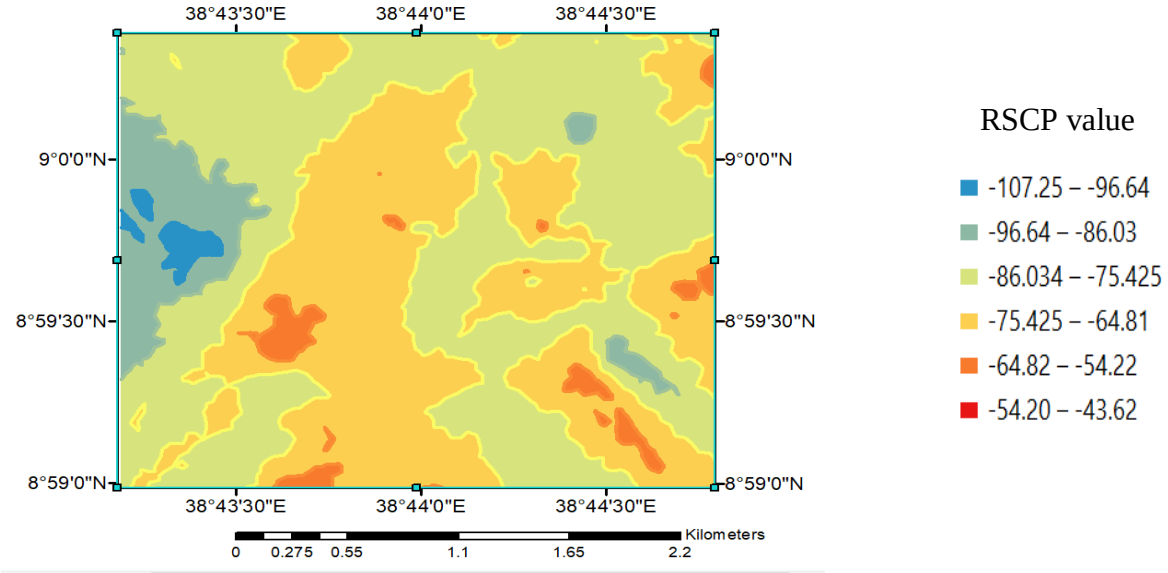


Figure 4.16: Coverage Prediction Map by OK (a) Raster (b) Contour

Evaluating the accuracy of prediction

The final step in Ordinary kriging is evaluating the accuracy of prediction. The accuracy of prediction was evaluated by using the cross-validation tool. The predicted plot OK displays

measured vs. predicted value as shown in Figure 4.17. As shown in figure OK has an optimum error of prediction with the value of RMSE=4.83.

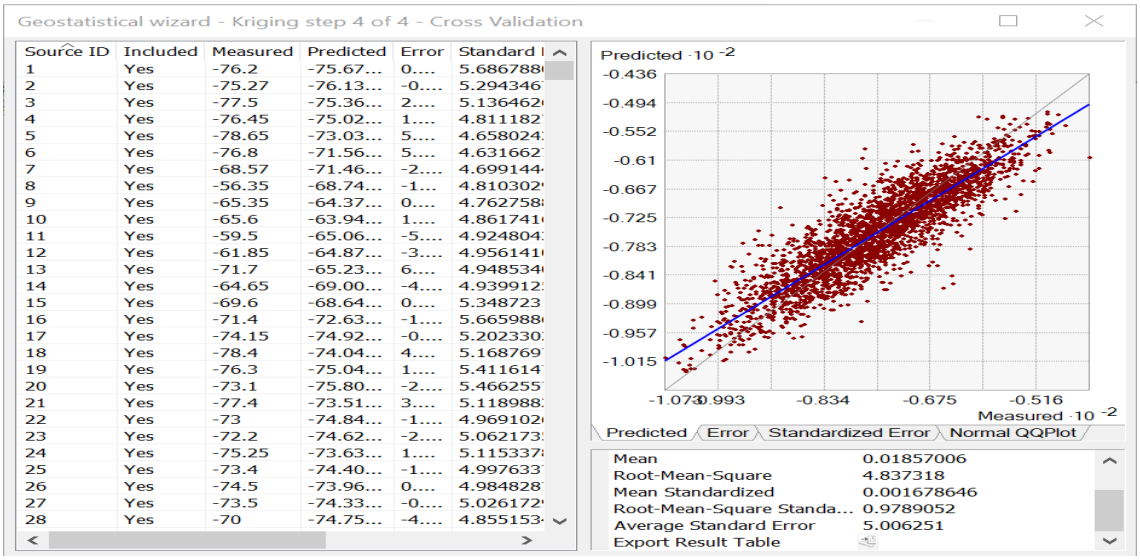


Figure 4.17: Predicted vs measured plot of OK

4.4 Performance Evaluation of Different Interpolation Techniques

- Cross-validation

The objective of cross-validation is used to select which model provides the most accurate predictions. The cross-validation comparison result shows that the prediction error of OK is 0.018, 4.837 whereas the prediction error of IDW is 0.11389, 5.025 mean error and RMSE respectively as shown in Figure 4.18.

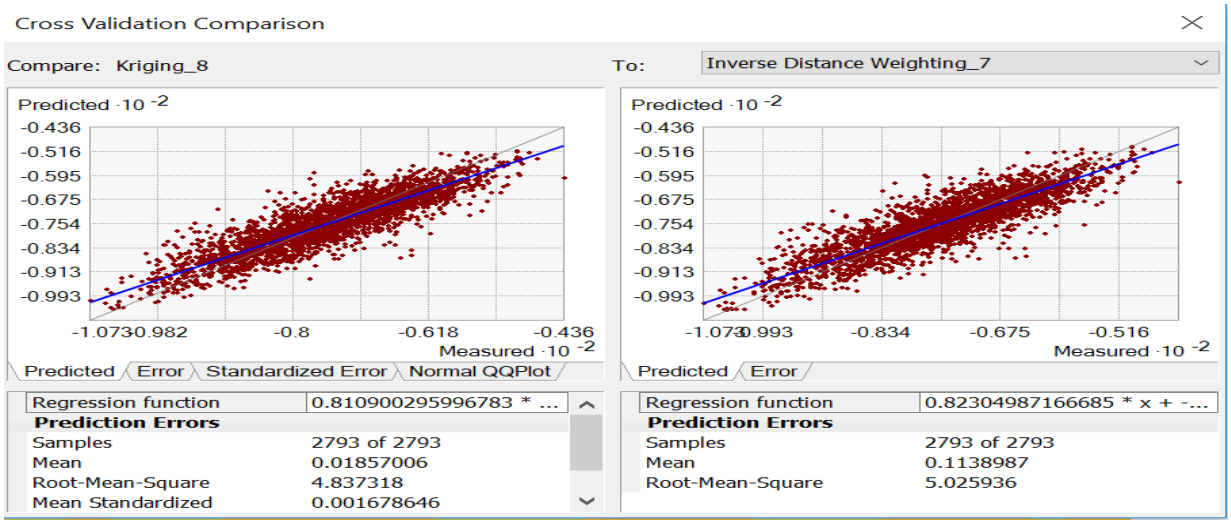


Figure 4.18: Cross-validation of OK and IDW

- Statistical Method

For the result comparison, we used test points. Prediction values for the test points are extracted from each map and the difference with actual value is computed. The RMSE and MAE are used to compare the accuracy of prediction of IDW and OK interpolation methods. As shown in Table 4.12 OK with error of prediction 4.84 RMSE is more accurate than IDW (5.33RMSE).

Interpolation	RMSE	MAE
OK	4.846124	3.729479
IDW	5.336532	4.076508

Table 4.11: Prediction Errors

5 Result and Discussion

5.1 Result

This chapter consists of the coverage estimation maps created by the selected interpolation method with a different grid of resolution as well as their evaluation results.

a) Performance evaluation

Evaluation of different interpolation models IDW and OK are applied to establish the coverage prediction model (RSCP level), root-mean-square errors, coefficient of determinations and absolute errors are calculated for each method to select the optimal model.

As shown in Table 5.1, RMSE of OK with Exponential semivariogram model, neighbors to Include(NtoI) =10, and include at least(IatL) =5 is 4.84 whereas RMSE of IDW with P=2, NtoI=15, and IatL=5 is 5.3. This shows that OK more preferable than IDW for cellular wireless coverage prediction. The percentage difference of errors of the two algorithms was 17%. Figure 5.1 shows the prediction error of the two methods in terms of RMSE and MAE in graph.

Parameter Values	Interpolation	RMSE	MAE
Exponential model, NtoI=10 and IatL=5	OK	4.846	3.729479
P=2,NtoI=15 and IatL=5	IDW	5.336	4.076508

Table 5.1: The statistic errors in the process of RSCP interpolation

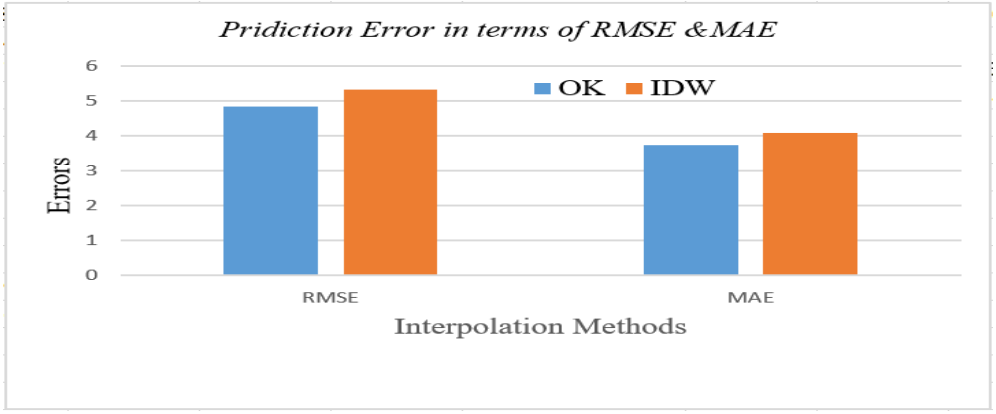


Figure 5.1: IDW and OK Prediction Error

A comparison of prediction performance of interpolation methods using the test dataset in R^2 presented as follows. As shown in Figure 5.2 both methods abled to predict the unknown measurement goodly since the value of R^2 greater than 0.5. But R^2 of OK is greater than that of IDW.

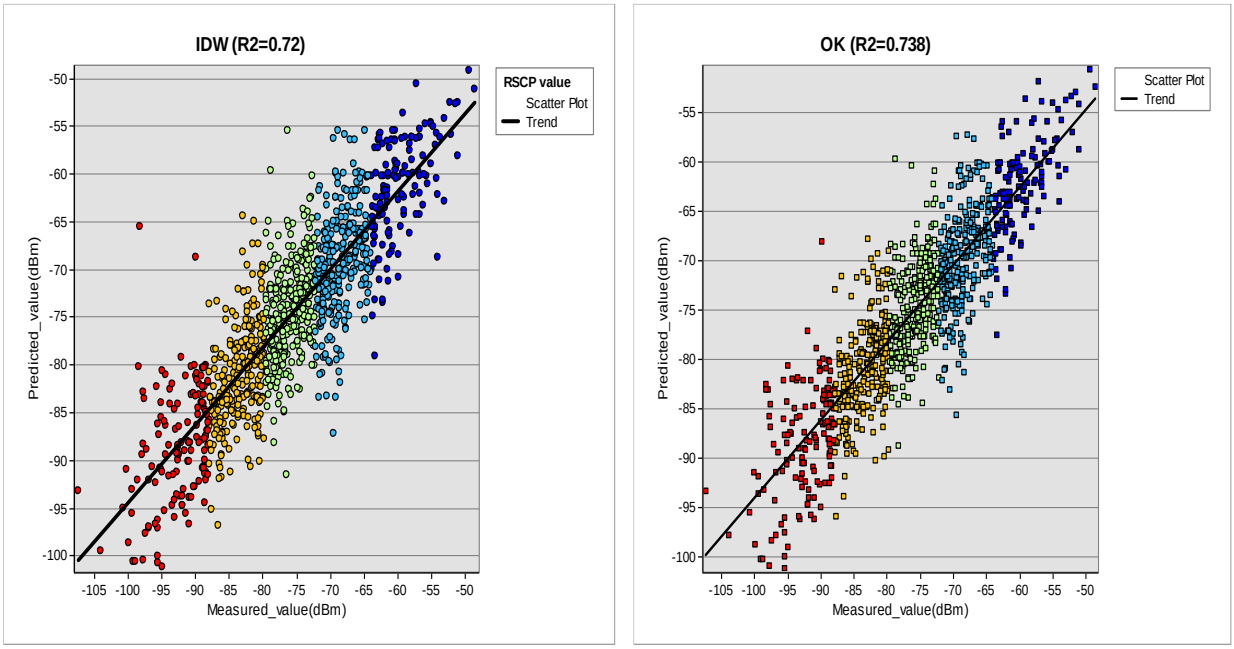


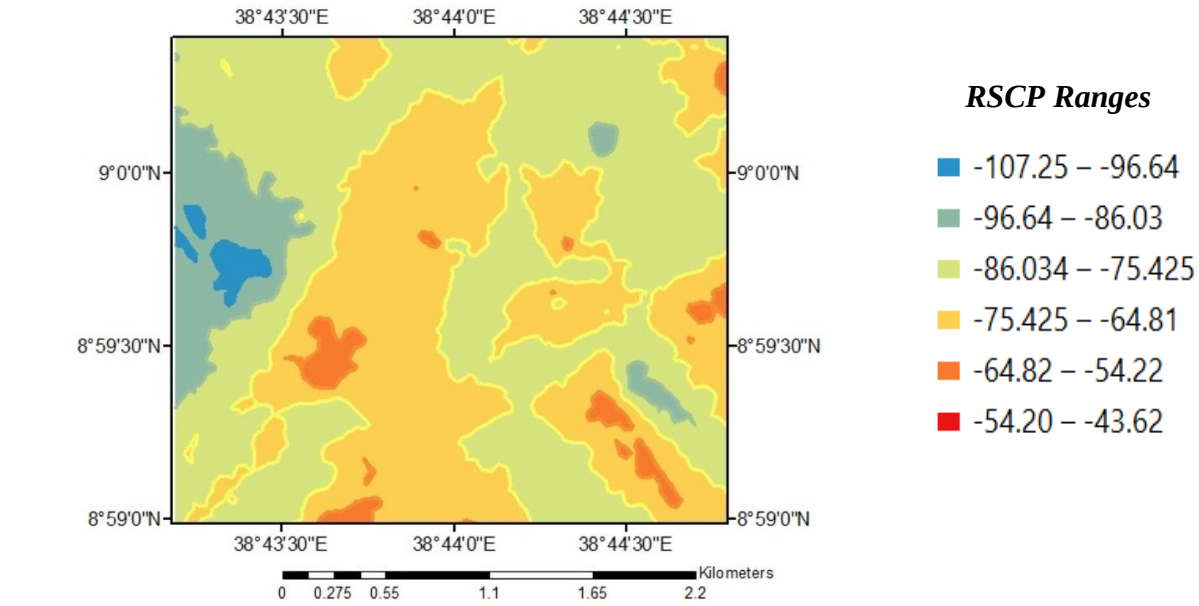
Figure 5.2: Accuracy of prediction using R^2

b) Prediction Map

UMTS coverage map for the target area based on the RSCP which is collected from drive test data was created using spatial interpolation algorithm OK, as shown in Figure 5.3. The lighter color of the coverage map indicates that the signal power of the UMTS network is greater, whereas the darker the coverage map color, the smaller the UMTS network power. The map was classified based on its RSCP values into six zones. The map shows that most of the coverage analysis area (91.08%) has good coverage except Zone zero. Zone zero has poor coverage with the value between -107.25 and -96.525. According to ethio telecom, regions which have RSCP value less

than -95dBm considered as poor coverage area. The map presented into two forms in contour and raster map.

a) *Contour Map*



b) Raster Map

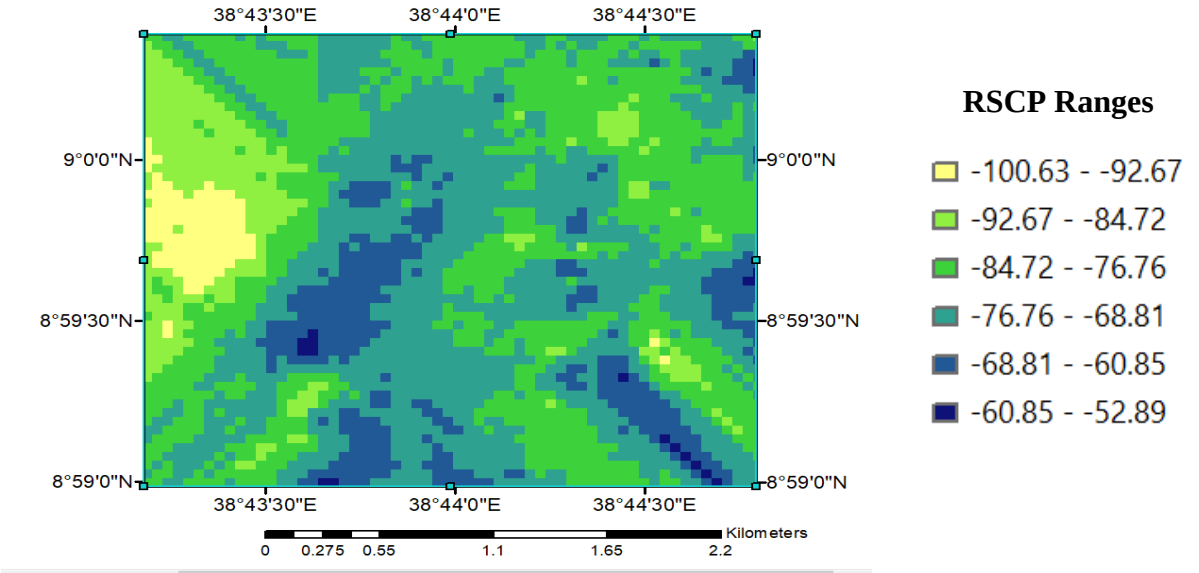


Figure 5.3: Coverage Map by OK (a) contour (b) Raster

As a summary, the whole area of study divided into six class of coverage ranges as presented in Table 5.2.

Shape	Zones	RSCP Value Min	RSCP Value Max	Shape Length(m)	Shape Area(m^2)	Percentage of coverage in the area
Polygon	0	-107.25	-96.525	1627.153638	66439.44228	0.60%
Polygon	1	-96.525	-85.8	8382.035462	585338.2794	5.29%
Polygon	2	-85.8	-75.075	48797.60537	5449534.952	49.29%
Polygon	3	-75.075	-64.35	50691.7097	4620077.746	41.79%
Polygon	4	-64.35	-53.625	9891.373721	329133.8976	2.98%
Polygon	5	-53.625	-42.9	293.20337	6071.812833	0.05%

Table 5.2: Range of RSCP for each area

The other advantage of OK provides a standard error map as shown in Figure 5.3. That shows the uncertainty related to the predicted values. Areas, which are near to the measurement RSCP values, have fewer errors of prediction than far distance. The light color indicates less error of prediction while the dark red color indicates a high error of prediction.

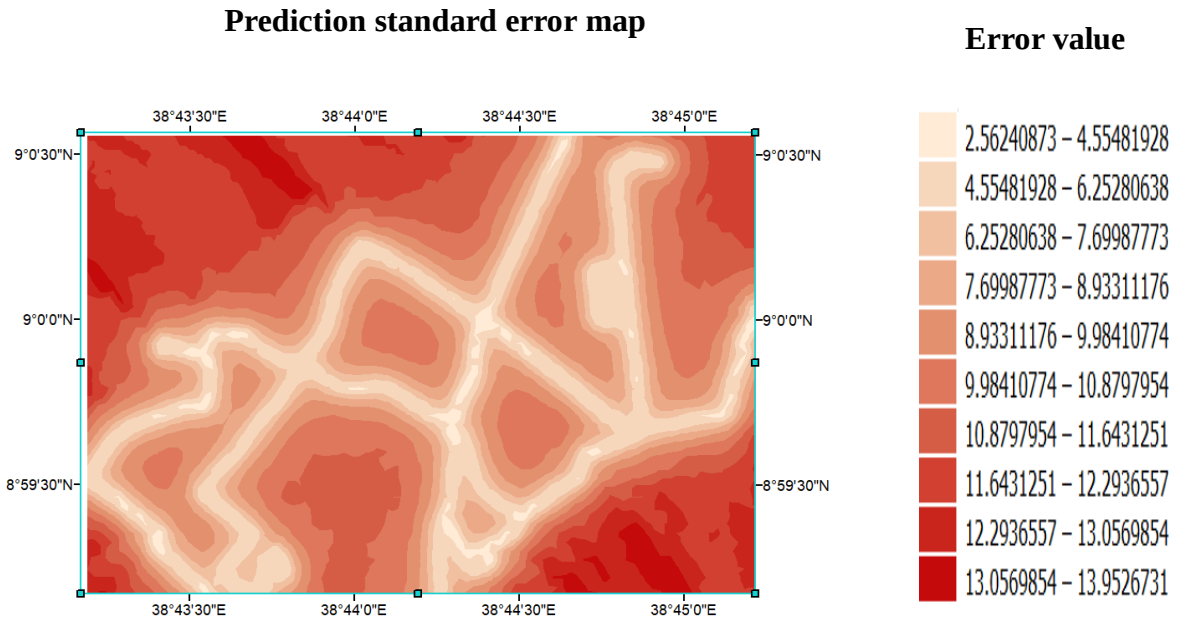
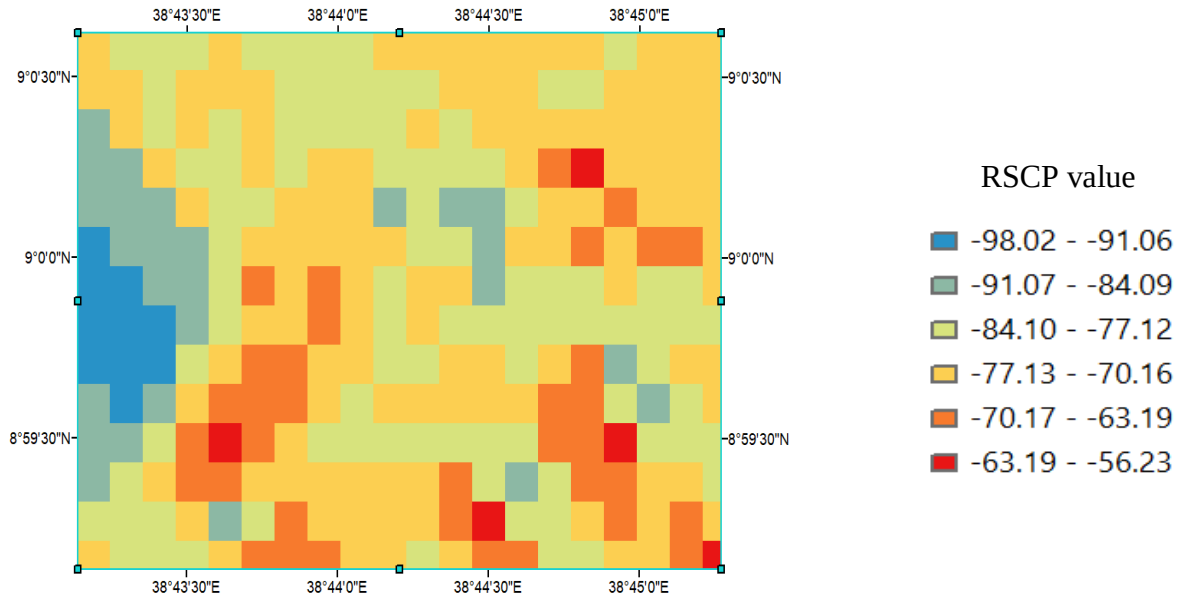


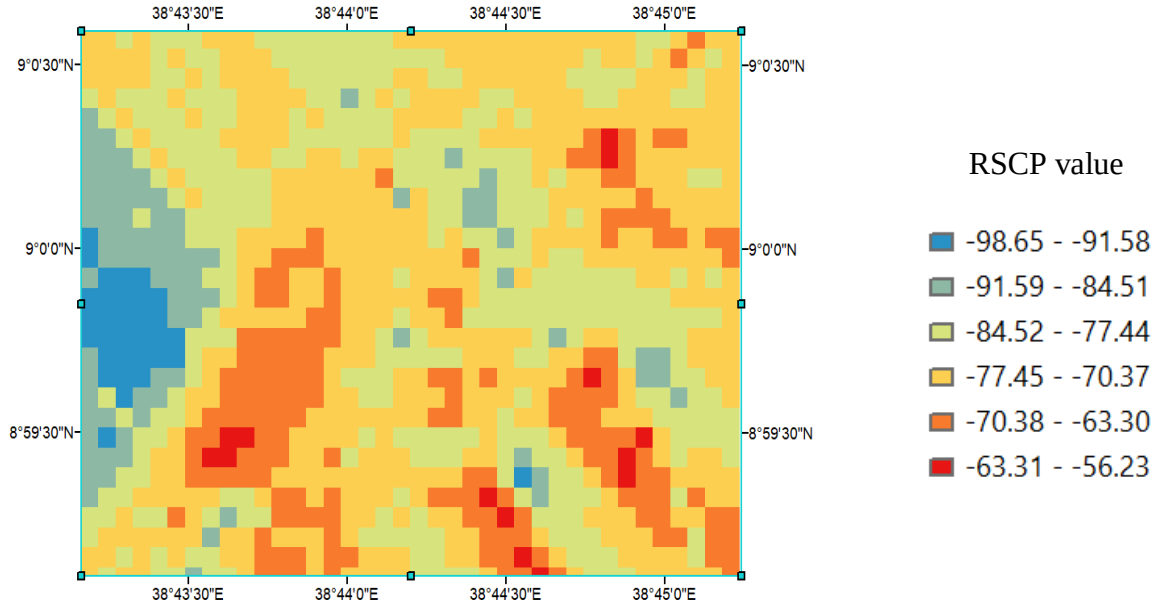
Figure 5.4: OK Standard Error Map

The interpolation map of the coverage area for 200x200, 100x100, 50x50, and 25x25 meters grid of resolution are shown in Figure 5.4(a), (b), (c), (d) respectively. The interpolation of the coverage area of 25x25 meters resolution gave more details about RSCP, and it is smooth as shown in Figure 5.4(d).

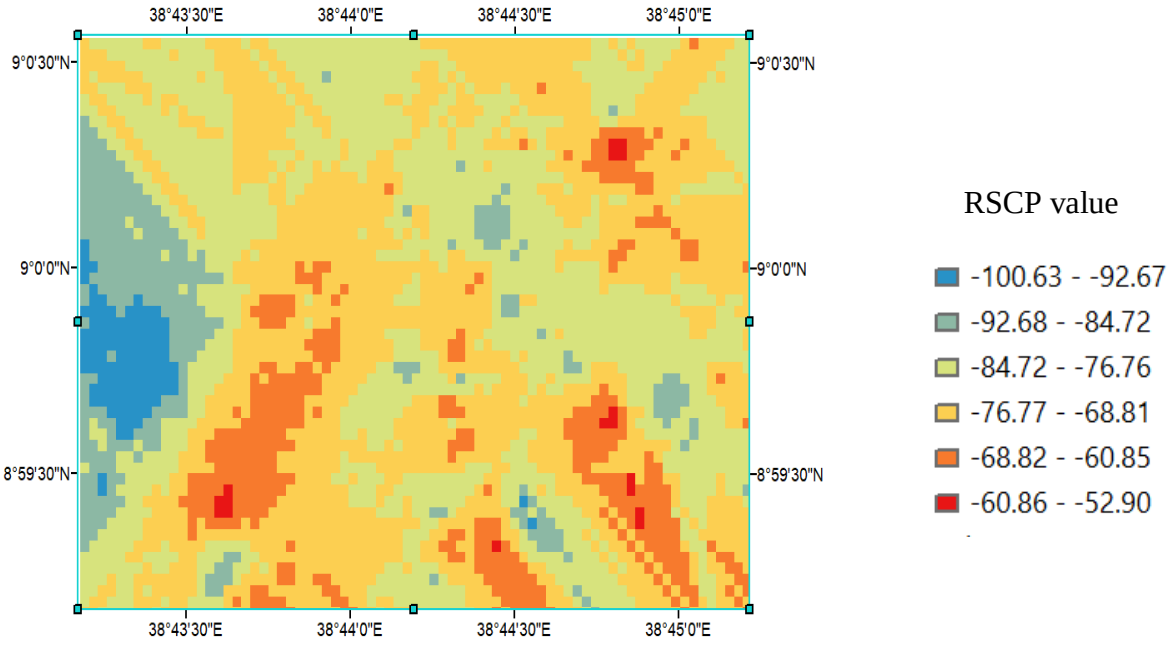
Coverage Map of(200x200)



Coverage Map of(100x100)



Coverage Map of(50x50)



Coverage Map of(25x25)

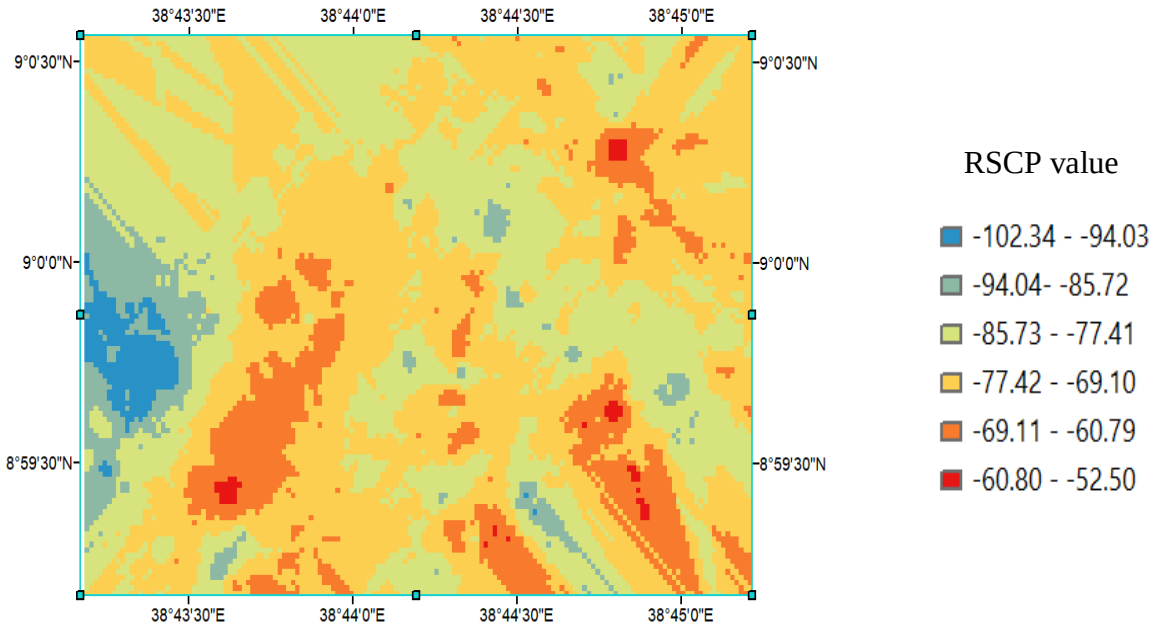


Figure 5.5: Coverage Map of the study area in different resolution

5.2 Discussion

Generally, these interpolated surfaces indicate that the OK interpolation method has better performances. However, the main purpose of this study is to assess the performances of different interpolators by using statistical criteria. Hence, the interpolated surfaces would be further assessed by integrating the following statistical methods RMSE, MAE and R^2. Experiment results show that the OK method used here is 17% more accurate than IDW and has an error of 4.84 RMSE. As the unbiased optimal interpolation method, the OK method combines the effects of distance parameters and semivariogram model parameters, it represents the spatially continuous and irregular change of variables. Therefore, the interpolation effect of the Kriging model is better than IDW in theory. From the perspective of interpolation results, the OK interpolation is more approximate to the cellular coverage prediction data (RSCP). The map of OK looks smooth, which means the interpolation effect is good. Based on the above analysis, OK interpolation method is evaluated as the optimal interpolation model for such kind of data. Therefore, our result shows that the OK algorithm with an exponential semivariogram model with 10 and 5 maximum and minimum neighbor variables can be selected as an optimal model for UMTS mobile network coverage prediction based on the collected sample data from the drive test.

6 Conclusion and Future Work

6.1 Conclusion

In this thesis work, an optimized and accurate spatial interpolation technique for UMTS networks coverage prediction has been proposed by comparing IDW and OK spatial interpolation techniques. CPICH RSCP statistical data collected from the drive test have been considered for the prediction. The purpose of this thesis is maximizing coverage prediction, of drive test data by using interpolation techniques, which create coverage maps, using limited measurement points and to reduce the cost of time-consuming and labor-intensive data collection. The result shows that both the proposed coverage prediction algorithms are able to predict coverage. However, based on the Exponential model for semivariogram with an optimal number of neighbors the OK methods used here are 17% more accurate than IDW and have an acceptable error that is 4.84 RMSE. OK interpolation algorithm could estimate the missing RSCP data and generate more accurate coverage of information data than the IDW algorithm. OK shows a good result when the coverage area was divided into different bin sizes that are 200x200, 100x100, 50x50, and 25x25 meters of resolution. This result is probably due to the ability of Kriging to take into account the spatial structure of data.

6.2 Future Work

The future works proposed from this research include:

This study considered only two spatial interpolation algorithm but there are other interpolation methods that are not used such as spline, Global polynomial interpolation from deterministic methods and Universal Kriging and CoKriging from geostatistical interpolation techniques could be applied in the future studies for cellular network coverage prediction.

The accuracy of OK for coverage prediction can be improved in the future by considering regularly spaced and uniformly distribute RSCP data collection since this thesis uses data collected

from distant roads using drive test data. So that it is possible to minimize the cost of purchasing expensive tools for analyzing coverage related parameters.

7 Reference

[1] A. Basere and I. Kostanic, "Cell coverage area estimation from receive signal level (RSL) measurements," *Proceedings of the World Congress on Engineering and Computer Science (WCECS)*, vol. 1, pp. 73–79, 2016.

[2] A. Galindo-Serrano, B. Sayrac, S. Ben Jemaa, J. Riihijarvi, and P. Mahonen, "Harvesting MDT data: Radio environment maps for coverage analysis in cellular networks," *8th Int. Conf. Cogn. Radio Oriented Wirel. Networks Commun.*, pp. 37–42, 2013.

[3] A. Galindo-serrano, B. Sayrac, and S. Ben Jemaa, "Automated Coverage Hole Detection for Cellular Networks Using Radio Environment Maps," *9th International Workshop on Wireless Network Measurements 2013,* pp. 108–113, 2013.

[4] B. Sayrac, A. Galindo-serrano, S. Ben Jemaa, and J. Riihijärvi, "Bayesian spatial interpolation as an emerging cognitive radio application for coverage analysis in," *Trans. Emerging Tel. Tech*, pp. 636–648, 2013.

[5] M. Sousa, A. Martins, and P. Vieira, "Self-Diagnosing Low Coverage and High Interference in 3G / 4G Radio Access Networks based on Automatic RF Measurement Extraction," *In Proceedings of the 13th International Joint Conference on e-Business and Telecommunications (ICETE 2016)*, vol. 6, pp. 31–39, 2016.

[6] Jaymin D. Mankowitz and Andrew J. Paverd, "Mobile Device-Based Cellular Network Coverage Analysis Using Crowd Sourcing."

[7] Omer¨ Faruk Celebi, Omer¨ Faruk Kurt and Engin Zeydan, "On Use of Big Data for Enhancing Network Coverage Analysis." in *Proc. ICT*, Casablanca, Morocco, pp. 1_5, May 2013.

[8] Massimiliano Molinari, Mah-Rukh Fiday, Mahesh K. Marinay and Antonio Pescape, “Spatial interpolation based Cellular Coverage Prediction with Crowdsourced Measurements,” *Proceedings of the 2015 ACM SIGCOMM* ., pp. 33–38, Aug. 2015.

[9] A. Konak, “A kriging approach to predicting coverage in wireless networks" *Int. J. Mobile Network Design and Innovation,* Vol. 3, No. 2, pp.65–71, January 2009.

[10] Y. Abera, “Spatiotemporal Mobile Data Traffic Modeling : The Case of UMTS Network in Addis Ababa," *MSc. dissertation, Dept. Elect. and Comp. Eng, Addis Ababa Univ., Addis Ababa,* 2018.

[11] Yong Xiao, Xiaomin Gu, Jingli Shao and Qiulan Zhang, “Geostatistical interpolation model selection based on ArcGIS and spatio - temporal variability analysis of groundwater level in piedmont plains , northwest China,” *Springerplus,* 2016.

[12] S. Kolyaie and M. Yaghooti, “Evaluation of Geostatistical Analysis Capability in Wireless Signal Propagation Modeling,” Proceedings of the 11th International Conference on GeoComputation, pp. 76–83, 2011.

[13] A. Gómez-andrades, R. Barco, and I. Serrano, “Open Access A method of assessment of LTE coverage holes,” *EURASIP J. Wirel. Commun. Netw.*, 2016.

[14] H. Liang and L. Chen, “Coverage Hole Detection in Cellular Networks with Deterministic Propagation Model.” *IEEE*, 2016.

[15] J. Ponce-Rojas, S. Vidal-Beltrán, J. L. López-Bonilla, and M. Jimenez-Licea, “Krige Method Application for the Coverage Analysis of a Node-B in a WCDMA Network,” *Int. J. Commun. Netw. Syst. Sci.*, vol. 04, no. 03, pp. 180–188, 2011.

[16] B. Sayrac, J. Riihijärvi, P. Mähönen, S. Ben Jemaa, E. Moulines, and S. Grimoud, “Improving coverage estimation for cellular networks with spatial Bayesian prediction based on measurements,” *CellNet’12 - Proc. ACM Work. Cell. Networks Oper. Challenges, Futur. Des.*, pp. 43–48, 2012.

[17] H. Harri and T. Antti, *WCDMA for UMTS – HSPA evolution and LTE*, 5th Edition, John Wiley & Sons Ltd, Finland, 2010.

[18] M. Abdur and R. Haider, "Radio Resource Management In 3G UMTS Networks," *MSc. dissertation, Blekinge Institute of Technology*, 2007.

[19] R. Kreher, *UMTS Performance Measurement A Practical Guide to KPIs for the UTRAN Environment*, John Wiley & Sons Ltd, Chichester, England, 2006.

[20] A. Catarina, G. Gomes "Interference Detection and Reduction in 3G / 4G Wireless Access Network," 19*th IEEE Mediterranean Electrotechnical Conference (MELECON), 2018.*

[21] 3GPP TS 25.215, version 11.0.0, "Universal Mobile Telecommunications System (UMTS); Physical layer; Measurements (FDD)" Release 11, pp. 0–24, 2012.

[22] Electronic Communications and Committee (ECC), *ECC report on UMTS Coverage Measurements*, no.103, May 2007.

[23] O. Sallent, J. Pérez-Romero, J. Sánchez-González, and R. Agust, "A Roadmap from UMTS Optimization to LTE Self-Optimization," *IEEE Communications Magazine*, pp. 172–182, 2011.

[24] K. Johnston, J. M. Ver Hoef, K. Krivoruchko, and N. Lucas, *ArcGIS Geostastistic_Analysyst by ESRI*, 2001.

[25] W. Xu, Y. Zou, and M. Linderman, "A comparison among spatial interpolation techniques for daily rainfall data in Sichuan Province , China," vol. 2907, no. October 2014, pp. 2898–2907, 2015.

[26] B. S. Babu, "Comparative Study on the Spatial Interpolation Techniques in GIS," vol. 7, no. 2, pp. 550–554, 2016.

[27] J. McCoy, K. Johnston, S. Kopp, B. Borup, J.Willison, B. Payne, "ArcGIS 9 *Using ArcGIS Spatial Analyst*," ESRI, 2002.

[28] C. Y. Wu, J. Mossa, L. Mao, and M. Almulla, “Comparison of different spatial interpolation methods for historical hydrographic data of the lowermost Mississippi River,” *Ann. GIS*, vol. 25, no. 2, pp. 133–151, 2019.

[29] C. A. Rishikeshan, S. K. Katiyar, and V. N. Vishnu Mahesh, “Detailed evaluation of dem interpolation methods in GIS using DGPS data,” *Proc. - 2014 6th Int. Conf. Comput. Intell. Commun. Networks, CICN 2014*, pp. 666–671, 2014.

[30] B. A. Mert and A. Dag, “A Computer Program for Practical Semivariogram Modeling and Ordinary Kriging: A Case Study of Porosity Distribution in an Oil Field,” *Open Geosci.*, vol. 9, no. 1, pp. 663–674, 2018.

[32] Z. Han, J. Liao, Q. Qi, H. Sun, and J. Wang, “Radio Environment Map Construction by Kriging Algorithm Based on Mobile Crowd Sensing,” *Hindawi Wireless Communications and Mobile Computing*, 2019.

[33] A. Konak, “Estimating path loss in wireless local area networks using ordinary kriging,” *Proc. - Winter Simul. Conf.*, pp. 2888–2896, 2010.

[34] S. Wang, G. H. Huang, Q. G. Lin, Z. Li, H. Zhang, and Y. R. Fan, “Comparison of interpolation methods for estimating spatial distribution of precipitation in Ontario , Canada,” *Int. J. Climatol.*, vol. 34, pp. 3745–3751, Feb. 2014.

[35] T. Hailu, “Network Traffic Classification Using Machine Learning : A Step Towards Over-the-Top Bypass Fraud Detection,” *MSc. dissertation, Dept. Elect. and Comp. Eng, Addis Ababa Univ., Addis Ababa,* 2018.

[36] S. Bekele, “Cell Outage Detection Through Density-based Local Outlier Data Mining Approach : In case of Ethio,” *MSc. dissertation, Dept. Elect. and Comp. Eng, Addis Ababa Univ., Addis Ababa,* November 2018.