Addis Ababa University

Addis Ababa Institute of Technology

School of Computer and Electrical Engineering

Analysis of Land Mobile Spectrum Occupancy using Spatial Interpolation Algorithms: For the case of Addis Ababa, Ethiopia.

By

Rufael Getachew

Advisor

Dr. Ing. Dereje Hailemariam

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 Telecommunication Engineering

September, 2021

Addis Ababa, Ethiopia

Addis Ababa University

Addis Ababa Institute of Technology

School of Computer and Electrical Engineering

Analysis of Land Mobile Spectrum Occupancy using Spatial Interpolation Algorithms: For the case of Addis Ababa, Ethiopia.

By: Rufael Getachew

Approval by Board of Examiners

______________________________ ____________

Dean, School of Electrical and Computer Engineering — Signature

Dr. –Ing. Dereje Hailemariam ____________

Advisor

______________________________ ____________

Internal Examiner — Signature

______________________________ ____________

External Examiner — Signature

Declaration

I declare that the work contained is my own, and has not been summited for a degree in any other university or professional qualification, and all sources of materials used for the thesis have been fully acknowledged.

Rufael Getachew ______________________

Name Signature

Place: Addis Ababa

Date of Submission: ______________________

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

Dr. –Ing. Dereje Hailemariam ______________________

Advisor's Name Signature

Abstract

An increase in the development of information industry and wireless communication services shows an increase demand for spectrum resource. Spectrum is a limited resource that needs to be managed properly to enable additional new services, without spectrum, none of todays' wireless communication will be true. The spectrum occupancy map can be a powerful tool for developing a better knowledge on the occupancy status of this scares resource.

Traditionally, spectrum measurement is based on drive test, which conveys geographically measuring different spectrum bands with moveable vehicles equipped with mobile spectrum measurement capability. Drive test cannot be conducted in all regions due to constraints such as, buildings, road conditions, and inefficient measurement tools. Therefore, the drive test is inefficient to get information on spectrum occupancy status and it cannot offer a complete and reliable occupancy information.

In this thesis, the performance of three spatial interpolation methods, namely Inverse Distance Weight (IDW), Ordinary Kriging (OK) and Natural Neighbor (NN) are evaluated to select which one is best method to develop spectrum occupancy map. The experimental analysis is performed on sample data, collected from TCI 700-series spectrum measurement unit, in the range of 137MHz-174MHz of VHF bands and 400MHz-470MHz of UHF bands for land mobiles in selected area of Addis Ababa. The collected sample data is given for K-means clustering algorithm to classify the occupancy status at different locations. The three interpolation methods are employed for only UHF bands.

The prediction performance of the three algorithms is evaluated through cross validation and Root Mean Square Error (RMSE). The obtained results showed that, OK with Gaussian model of semivariogram estimate with a prediction error of 0.807, while IDW and NN estimate values 2.801 and 2.119 respectively. This show OK is more accurate than IDW and NN.

Keywords*: Spatial Interpolation, Spectrum, Kriging, Inverse Distance Weight, Natural Neighbor, Land Mobile.*

Acknowledgement

First of all, I am thankful to God and Virgin Mary for giving me strength throughout my thesis work.

Secondly, I want to thank my family and Ethiopian Communication Authority for giving me the opportunity to attend the postgraduate program at Addis Ababa University. I would also like to thank the Addis Ababa Institute of Technology (AAiT) for providing such a favorable environment for the study.

I would like to give my special regards to my respected advisor Dr. –Ing. Dereje Hailemariam for his support in developing this thesis, advising and consulting. His dedication and willingness to discuss any new ideas are always inspiring me as a model to follow in my future work and study. Finally, I would like to give my special thanks to my friends for understanding my situation and giving me support, strength and encouragement on my working.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ECA	Ethiopian Communication Authority
IDW	Inverse Distance Weight
ITU	International Telecommunication Union
LM	Land Mobile
MCS	Mobile Crowd Sensing
MDT	Minimizing Drive Test
NFAT	National Frequency Allocation Table
NN	Nearest Neighbor
OK	Ordinary Kriging
REM	Radio Environment Map
RF	Radio Frequency
RBW	Resolution Bandwidth
RMSE	Root Mean Square Error
RSS	Received Signal Strength

Chapter I: Introduction

1.1 Overview

Spectrum is a limited resource that needs to be managed properly to enable new services and applications. One of the services provided by spectrum is wireless communication. Nowadays this communication service is fundamental for our daily activities; which may include mobile phone calls, internet service, broadcasting services, and other communication services that we can use on our mobile phones. The frequencies we use for wireless are the only portion of what is called the electromagnetic spectrum.

The continuous increase in wireless systems, attractive applications and, subscribers growth have created an increasing demand for this resource and motivated the search for new strategies for efficiently utilizing and taking the greatest advantage of the radio spectrum [1].

Spectrum occupancy measurement and evaluation, in modern radio frequency (RF) environments with increasing density of digital systems and frequency bands shared by different radio services, become a more and more complex and challenging task for monitoring services [2].

Despite the shortage of the radio spectrum, according to research reported in [3], measurements on radio spectrum usage have revealed an unoccupied bands of spectrum that belong to primary (licensed) networks. Prior knowledge about the occupancy of such bands and the expected achievable performance on those bands can help secondary (unlicensed) networks to devise effective strategies to improve utilization[3].

To deploy new and additional radio services, spectrum occupancy knowledge is mandatory. It is a bit complex job for regulators to perform an occupancy measurement, and this is due to additional interference signals, uncomfortable environment, and obstacles of buildings. Since the radio environment is unpredictable, due to the characteristics of the wireless environment we need to consider multiple parameters, to get a better insight and knowledge on spectrum occupancy status.

Unlike in our country, Ethiopia, the statistics of spectrum allocation around the world show that the radio spectrum has been allocated properly and the available spectrum for deploying new services is quite limited; spectrum usage, allocation, usage, management, and utilization on the other hand in our country is still a virgin area to explore.

To improve the spectrum occupancy knowledge and to provide well information of the spectrum resource, we can build a spectrum occupancy map to understand the occupancy status. The map can offer multi-domain environmental information, such as geographical information, available services on an area, locations of spectrum usage, occupancy status, and time of use of radio services.

In [4] spectrum occupancy map is suggested as a promising concept for storing radio environmental information that can be used to enhance radio resource management in wireless networks. In cellular networks, the radio environment map (REM) can be used to improve the resource utilization, or to minimize the operational costs by replacing or at least minimizing drive tests (MDT), for troubleshooting of interference for instance [4]. The concept of collecting geo-located information on the radio environment and developing an occupancy map using gathered information has also been investigated and developed further by many other research groups.

The study in [5] suggests developing REM for constructing a dynamic spectrum map for each frequency at each location of interest. The study also shows it is impractical to have measurements at each location in the radio environment operation area, the map is used as an advanced knowledge base to stores available multi-domain information on the entities in the spectrum resource.

To enable REM, different interpolation algorithms are suggested in different literatures. Among them, Kriging interpolation algorithm plays a vital role. The Kriging interpolation algorithm has been widely used in geostatistics principle for spatial interpolation for its accuracy. Kriging is an advanced geostatistical procedure that generates an estimated surface from a scattered set of points with z-values. Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. The Kriging tool fits a mathematical function to a specified number of points, or all points within a specified radius, to determine the output value for each location. Kriging is a multistep process; it includes exploratory statistical analysis of the data, variogram modeling, creating the surface, and (optionally) exploring a variance surface [5].

1.2 Statement of Problem

Spectrum occupancy knowledge is needed both in spatial and temporal domains. Traditionally, occupancy measurement is conducted through drive test in the operation environment. However, drive test is limited to certain places as it is resource (i.e., time, man-power and equipment) intensive; access to buildings and other places may not be possible; and spectrum users have a dynamic and random usage patterns that are not known. Hence, an alternative approach is to augment the drive test with interpolation techniques so as generate spectrum occupancy map in locations where no measurement data is taken. Multiple spatial interpolation techniques are available, and one of them is the Kriging interpolation algorithm. The algorithm has an advantage of measuring the error or uncertainty for each estimated surface.

In the Ethiopian context, the spectrum allocation is static; i.e., each licensed spectrum band is statically assigned to specific licensed service and its users and this allocation is not allowed to change. Moreover, given the monopoly nature of spectrum utilization by limited government-affiliated organizations and operator (in most frequency bands), knowledge of spectrum utilization by the regulating authority, both in space and time domains, is limited. Hence, this thesis intends to address this limitation by taking actual spectrum measurements on certain frequency bands, apply spatial interpolation and draw meaningful conclusions regarding the spectrum occupancy in spatial domain for the considered band.

1.3 Objective

1.3.1 General Objective

The main objective of this research is to develop a spectrum occupancy map using spatial interpolation algorithm for a better understanding of the spectrum occupancy status of Land mobiles frequency ranges for the selected area in Addis Ababa.

1.3.2 Specific Objectives

To achieve the above-mentioned general objective, the following specific objectives are set:

- Review of related papers, journals, and books;
- Collect data in the spatial domain;
- Study interpolation algorithms;
- To build a semivariogram model to study the relationship between variables in collected data;
- Study how collected data can be analyzed;
- To construct occupancy map using interpolation techniques through geo-located measurements collected by monitoring machine.

1.4 Literature review

For this research, various literatures related to developing spectrum map are reviewed. A study conducted in [5] proposes to apply Mobile Crowd sourcing (MCS) to collect the radio information in the area of 150m^2 in 2.4GHz band and propose to apply Kriging_interpolation algorithm to predict the spectrum values in transmission power (dBm) of unmeasured locations based on the measurement data collected at locations where measurement can be taken.

They used the Kriging interpolation algorithm to understand the spatiotemporal correlation between radio environment data, and the Kriging interpolation algorithm can be applied to infer the missing radio environment information data [5]. To analyze the distribution of sample points scattered in the sensing environment they propose to apply a variogram model to show the spatial structure and distribution characteristics of the variables. To verify the accuracy of the RME developed the authors compares Kriging, Inverse Distance Weight (IDW), and Natural Neighbor (NN) spatial interpolation algorithms using RMSE value, and among the three interpolation methods Kriging shows the least error (RMSE) value and it used to develop REM.

The paper presented in [6] has revealed that measurements on spectrum usage have shown there exists an unused spectrum bands of licensed users. Shepard's interpolation algorithm is used to estimate the spectrum usage distribution over the spatial domain; the authors use secondary users to sense the operating environment for bandwidth range between 100 kHz-2MHz for 50-1000 channels. The energy detection principle is used while sensing the environment to find the set of free channels at (x_t, y_t) coordinate; after the spectrum usage is known the authors attempt to evaluate the proposed method in three dimensions; which are channel, usage distribution, measure the correctness of the predicted channel occupancy vectors and predict key performance matrices (i.e. channel capacity, secondary network throughput, spectral efficiency, and bit error rate) for secondary users who will share the free bands[3][6]. They also develop a mathematical model to estimate the performance matrices like spectral efficiency and bit-error rate. And based on prior knowledge of spectral efficiency they also demonstrate how performance matrices like channel capacity, system throughput, and spectral efficiency.

In [7] *K*-means clustering algorithm is used to find appropriate locations for dedicated sensor deployment to improve the REM construction performance significantly by utilizing user distribution probabilities on the area for finding uncovered locations. In this study, the dedicated sensors send their RSS measurement data to REM manager to predict the real situation in terms of Received Signal Strength (RSS) in the study area. Here, the REM manager collects each nodes location data and RSS value from the sensor nodes in the form of (x, y, RSS) and construct radio interference map using this information. The authors use two techniques to develop REM, spatial statistics based (direct) and transmitter location determination based methods (indirect); and also they use Kriging and location estimation based (LIvE) REM construction techniques to develop accurate REM. The authors finally propose KARMA deployment algorithm for deploying new dedicated sensors to enhance the construction of radio map.

Research conducted in [8] exploits a crowd-augmented spectrum mapping scheme to develop spectrum mapping; in which the objective is to create a power spectrum density (PSD) map for an area of interest, where each location on the map is associated with its corresponding PSD value. The authors in this study present an iterative Bayesian decision framework to construct an accurate representation of the spectrum map of the sensed geographical area. In this study, Bayesian spatial prediction was utilized to address the limitations in a two-phase kriging scheme by implementing the optimal predictor, whose prediction error can be accurately assessed.

1.5 Methodology

The study approach began with a review of related literature on the construction of spectrum occupancy map, followed by an examination of a measurement dataset of Addis Ababa's selected area occupancy data.

The model selection is performed in spatial domain using spatial interpolation algorithm methods. The interpolation algorithm predicts values for the unknown locations based on values obtained in the neighboring locations. To study the spatial correlation between measured data points, semivariogram model is used. It is used in wide range of applications/areas like wireless signal coverage prediction, UMTS network coverage prediction, developing spectrum utilization map and developing spectrum occupancy map.

The tools used are; Quantum Geographic Information System (QGIS) and ArcMap to develop the spectrum map, interpolation algorithms to predict values for unmeasured locations, Excel for preprocessing, spectrum occupancy measurement device to capture the geographical locations of each spectrum data by longitude and latitude values and occupancy value.

1.6 Scope and Limitation

1.6.1 Scope of the Thesis

The scope of the research is developing spectrum occupancy map in spatial domain by using interpolation algorithm for land mobile frequency bands of 137 MHz - 174 MHz in VHF bands and 400 MHz - 470 MHz in UHF bands, in Addis Ababa for selected area.

1.6.2 Limitation

This research is limited to developing spectrum occupancy map in selected area in spatial domain only. Time domain is not incorporated in this thesis; due to lack of having additional spectrum measurement stations it is difficult to take measurement in as many different locations as possible. The research does not cover all land mobile frequency ranges. In addition, the age of the measurement device used to collect spectrum data is one challenge which makes it difficult to make sure that the collected data is accurate.

1.6.3 Contribution

This research deals about developing spectrum occupancy map in selected area in Addis Ababa. The main contribution of this research is to develop spectrum occupancy map for land mobile frequency bands in spatial domain. Based on the knowledge of spectrum occupancy is different locations spectrum usage status at various locations can be identified, unoccupied land mobile frequency bands can be identified and this will help the regulator to have better knowledge on occupancy status of land mobile frequency usage in the study area.

1.6.4 Thesis Layout

The research paper is organized as follows. Chapter two presents spectrum usage and regulation trend in Ethiopia's regulatory body, ECA and spectrum measurement activity. In Chapter three, the basis of interpolation algorithms are presented. Chapter four, the experimental analysis is presented, which describes the considered scenario, including data collection and parameters, as well as the proposed system model for prediction of unknown values through interpolation algorithms and evaluation. Chapter five, includes results, discussion, and analysis of developed occupancy map. Finally, conclusions and recommendations are included in Chapter six.

Chapter II: Spectrum and Spectrum Measurement System

2.1 Overview of Spectrum

Radio frequencies are a scarce natural resource, they must be used effectively and efficiently by all wireless users to minimize wastage of this resource. National Frequency Plan (NFP) is a key instrument in spectrum resource management providing information on which radio communications services are permitted in each frequency band.

Radio spectrum is the range of frequencies used for a wireless applications such as broadcast television and radio, cell phones, satellite radio and television, wireless computer network, Bluetooth, Global Positioning System (GPS), police dispatch, and countless other general and specialized applications that we use in every day activity. For the most part, it is unreal for these applications to utilize the same frequencies at the same time and location [9].

To help avoid wastage of this resource, the radio spectrum is curved up into different portions of frequencies, and each portion is allocated to one or more service providers that, generally speaking, may be able to co-exist with each other without interference [9].

2.2 Spectrum Measurement Systems

[

A variety of spectrum measuring systems has been proposed by various writers.

Research published in [10] presented the V-Scope spectrum measuring method, which uses spectrum sensors on public transportation to gather and report spectrum occupancy by sensing their received signal energy. They tested their technique on a single metro bus that traveled across a mid-sized US city, collecting data at over 1 million places over a 120-square-kilometer region for UHF television bands in between 512 MHz - 698 MHz. The authors were able to handle a variety of locations, however extra interference from the measurement environment was a big problem due to the bus's movement within the city.

Another spectrum measuring approach is discussed in [5], which uses MCS to conduct spectrum measurements. In this approach, the authors organize mobile terminals to do the measurements, which are carried out by mobile users. Mobile users will acquire spectrum data by traveling around the designated areas with their phones. MCS has the benefit of being able to measure and gather data in tiny spaces, however the issue with this technology is that mobile devices are incapable of collecting huge amounts of data and have limited battery life.

Additional spectrum measurement technique is explained in [12], in which spectrum measuring Sensor Nodes (SN) are placed in distinct geographic locations of demand. The SNs use multidimensional views to gather and monitor electromagnetic data. The data is then instantly uploaded to the clouds through wireless or cable transmission channels. The three-way handshake protocol was used to communicate between clouds and SNs, which is based on request and response. SNs submit a request to the clouds to create a connection in the initial handshake, and the clouds react to SNs and establish a connection. In the second handshake, SNs request that particular data packets be downloaded in order to configure the state of the monitoring system. Clouds instantly reply to SNs and transfer data packets. Finally, in the third handshake, SNs request the connection to terminate from the clouds. The clouds react to it and the link will be terminated [11].

2.2.1 Spectrum Management and Monitoring in Ethiopia

The Spectrum management and monitoring department in ECA have two main bodies that are responsible for managing and monitoring the spectrum resource namely;

- *spectrum planning and licensing and*
- *spectrum monitoring*

The first division, spectrum planning, is the core task in spectrum management, which is used to show the road map of the current and future spectrum usages. It is also used to solve the problem of spectrum scarcity, minimize harmful interference, enhance spectrum efficiency, optimize spectrum usage, and increase the economic value of the spectrum. However, only frequency planning can't solve potential inefficient spectrum usage, proper frequency assignment and monitoring have to be maintained. Right frequency assignment for the right radio service is expected, according to ITU's recommendation. Frequency assignment in Ethiopia is based on traditional system, first come first served, which shows that dynamic spectrum access is not implemented yet.

Secondly, spectrum monitoring division has the responsibility of monitoring the radio spectrum. Spectrum monitoring consists of many activities such as measuring occupancy, identifying and geo-locating interference, checking spectrum users whether they are using radio spectrum according to the license agreement, and follow principles of RF systems or not. Additional functionalities of spectrum monitoring includes, ensuring EMC compliance, identifying unauthorized frequency users, and supporting administrative decisions by reporting radio spectrum monitoring data.

The two departments in collaboration perform many tasks regarding to spectrum management between them and some of the activities are listed below [9].

- ✓ Create policy for spectrum usage across the country.
- ✓ Bind national frequency resource and with the policy created.
- ✓ Allocate frequency according to NFAT.
- ✓ Develop policy that will help us to use international frequencies.
- ✓ Develop efficient spectrum usage strategies.
- ✓ Develop frequency tariff and formula for national frequency usage.
- ✓ Perform frequency monitoring to minimize the interference between operators and primary users.

- ✓ Provide license for spectrum owners.
- ✓ Claim bid for national frequency bands.
- ✓ Allocate signal and call sign for radio communications.
- ✓ Perform coordination on allocated frequency bands of satellite and radio communication channels.

2.2.2 Spectrum Monitoring Experience in Ethiopia

Spectrum monitoring is carried out in industrialized nations utilizing various measuring techniques. Air-based monitoring, which uses drone technology to collect data on spectrum occupancy, fixed type monitoring, which uses fixed sensors to collect data on spectrum occupancy, portable type monitoring, which uses portable measuring tools, and mobile station type monitoring, which uses vehicles equipped with spectrum measuring stations to drive across the city. Because of the availability of this measuring equipment's, measurements can be performed 24 hours a day, seven days a week, and throughout all spectrum bands at all places in need.

In comparison to industrialized countries, Ethiopia uses just one technology for spectrum measuring, which is a mobile station-based monitoring system. Due to its great size, this measuring technology cannot be used in all areas where measurements are required. The above-mentioned technologies for spectrum monitoring are not accessible in Ethiopia.

In Ethiopia, the use of radio spectrum is regulated by Ethiopia Communication Authority (ECA), which was established in 2019 as a regulatory authority body to regulate the communications services in Ethiopia.

However, due to a lack of measuring equipment and qualified man-power in our nation, spectrum measurements are rarely done on a regular basis. Measurements are carried out on a regular basis for the chosen region of interest. Spectrum monitoring is done for two reasons: one, to analyze the occupancy, and the other, to mitigate interference. The monitoring experience is restricted to only specific bands of spectrum due to a shortage of measurement equipment, which includes Radio Access Network (RAN) bands, broadcast bands, and land mobile bands.

Chapter III: Basis of Spatial Interpolation Techniques

3.1 General Introduction

Spatial interpolation technique is a method designed to predict the RSS value at unknown geographical points using known values at neighboring locations [12]. The estimation of values at unknown locations is based on a theoretical foundation of the Tobler's first law of geography. Which is, 'Everything is related to everything else, but near things are more related than distant things'. Meaning that adjacent measuring points are more similar than those measuring points that are far away from each other [13].

There are two main groupings of interpolation techniques: deterministic and geo-statistical. Both methods rely on the similarity of nearby sample points to create a surface. Deterministic techniques use mathematical functions that form weighted averages of nearby measured values to create the surface, while geo-statistical techniques use both mathematical and statistical methods [14].

Interpolation techniques include Ordinary Kriging, Inverse Distance Weight (IDW), Nearest Neighbor (NN), Spline, and Spline with Barriers. IDW, spline, and NN are classified as deterministic techniques, whereas OK is classified as a geostatistical interpolation approach. The next sections give basic information on the interpolation methods used in this study.

Different interpolation methods give their best results when the data to be used is normally distributed. So we need to check the statistical behavior of the data using data exploration techniques [15].

3.2 Exploratory Data Analysis

Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations. It can also be defined as, an approach using descriptive statistics and graphical tools to better understand data. It is used mainly to maximize insight into a dataset, detect outliers and anomalies, and test underlying assumptions. It is a robust first step before the application of other statistical methods[16].

The goal of exploratory data analysis (EDA) is to create a comprehensive description of the data as well as to discover general insights about the data [17]. Data exploration is the process, which is important to examine the data characteristics of data in different ways. It is an important step in spatial interpolation work to study the characteristics of the data to be used. The different techniques that can be used for such analysis are Histogram, Normal QQPlot, Trend Analysis, and Spatial Autocorrelation [17] [18].

3.2.1 Histogram

Histogram provides a univariate (one-variable) description of the data. The tool displays the frequency distribution for the dataset of interest and calculates summary statistics [17]. The frequency distribution is a bar graph, which shows how often observed values fall within certain intervals or classes. The tool offers descriptive statics such as median, mean, skewness, kurtosis, variance, standard deviation, 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 [15].

3.2.2 Spatial Autocorrelation

Spatial autocorrelation can help us to understand the degree to which one object is similar to other nearby objects. Moran's I (Index) measures spatial autocorrelation between data samples based on both feature locations and feature values simultaneously [19]. Before creating the interpolation model, we need to make sure that the data is spatially auto-correlated. To do so, I use ArcMap to perform a Global Moran's I test on the data to determine whether the data have 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 and no spatial autocorrelation is when Moran's I value is 0 [15].

3.3 Deterministic Interpolation Techniques

Deterministic interpolation techniques are based on the surrounding measured values or on specified mathematical formula that determines the smoothness of the resulting surface [17]. 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) [17]. Inverse Distance Weighted, Spline, Nearest Neighbor and radial basis functions are some types of deterministic interpolation techniques. For this thesis, Kriging is selected to predict occupancy status for unmeasured location due to its valid prediction and less error.

3.3.1 Natural Neighbor

Natural neighbor interpolation is a method of spatial interpolation, developed by Robin Sibson [20]. It is a simple method of multivariate interpolation in one or more dimensions. The method is based on Voronoi tessellation of a discrete set of spatial points. This has advantages over simpler methods of interpolation, such as nearest-neighbor interpolation, in that it provides a smoother approximation to the underlying "true" function [20][21].

For a given set of points in space, a Voronoi diagram is a decomposition of space into cells, one for each given point, so that anywhere in space, the closest given point is inside the cell. The Natural Neighbor method uses Thiessen or Voronoi polygons to define areas of influence around each input (i.e. measured) data point. Thiessen polygons have boundaries that define the area that is closest to each point relative to all other points. This property is mathematically defined by the perpendicular bisectors of the lines between all input data points. In this case, every un-sampled location is thus assigned the value of its nearest measurement point. [5][22].

The basic equation is:

$$G_{(x)} = \sum_{i=1}^{n} \omega_i(x) f(x_i) \qquad 3.1$$

Where, G(x) is the estimate at x, w_i are the weights and f (x_i) are the known data at (x_i). The weights, w_i, are calculated by finding how much of each of the surrounding areas is "stolen" when inserting x into the tessellation.

Sibson weights

$$w_{i}(x) = \frac{A(x_i)}{A(x)} \qquad 3.2$$

Where, $A(x)$ is the volume of the new cell centered in x, and $A(x_i)$ is the volume of the intersection between the new cell centered in x and the old cell centered in x_i.

Laplace weights

$$w_{i(x)} = \frac{\frac{l(x_i)}{d(x_i)}}{\sum_{k=1}^{n} \frac{l_{(x_K)}}{d(x_k)}} \qquad 3.3$$

Where, $l(x_i)$ is the measure of the interface between the cells linked to x and x_i in the Voronoi diagram and $d(x_i)$, the distance between x and x_i.

3.3.2 Inverse Distance Weight

Inverse distance weighting (IDW) is a deterministic type of interpolation technique for multivariate interpolation with a known scattered set of data points. It implements an assumption that things that are close to each other are more alike than those that are farther apart. To estimate the value for any unmeasured location, IDW will use the measured values surrounding the prediction location. 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 decreases with distance. The predictor of IDW is formed as a weighted sum of the data[17][18].

In [17] the general formula of IDW interpolation method is stated in Equation (3.4).

$$\hat{z}(s_0) = \sum_{I=1}^{N} \lambda_i z(s_i) \qquad 3.4$$

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

The weight of known points against the prediction locations will exponentially decrease as the increase in their distance[12]. The weight is calculated using Equation (3.5).

$$\lambda_i = \frac{d_{i_o}^{-P}}{\sum_{i=1}^{N} d_{i_o}^{-p}} \qquad 3.5$$

Where: p *is* the power exponent; d_{i0} is the distance between the prediction point s_0 and the known sample data point s_i.

According to equation 3.5 as the distance, d_{i0} between point's increases the weight decreases by a factor of p. 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 values. A lower power results in higher influence from distant points. 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[23][24].

3.4 Geostatistical Interpolation Technique

Unlike the deterministic interpolation technique, the Geostatistical interpolation technique estimates the value of unknown locations based on statistical models that include autocorrelation (statistical relationships among the measured points). Geostatistical techniques create surfaces that incorporate the statistical properties of the measured data[25]. Geostatistics includes a number of methods and techniques to analyze the variability of spatially distributed or spatially structured variables. 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 [17][18].

3.4.1 Kriging 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 or reserves from the samples taken over a mining field [25]. Kriging utilizes the variogram, which does not depend on the actual value of the variable (data), rather its spatial distribution and internal spatial structure [24]. It is a powerful interpolation technique that uses complex mathematical formulas to estimate values for unknown points based on the values obtained at known points. There are several types of Kriging: Ordinary, Universal, CoKriging, and Indicator Kriging[26]. For this thesis work the common type, Ordinary Kriging is selected.

3.4.1.1 Ordinary Kriging (OK)

Ordinary Kriging is similar to IDW in a way 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 [5][13][25].

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 data samples. Therefore, the predictor of OK is formed as a weighted sum of the data, as follows[18][25].

$$\hat{z} = \sum_{i=1}^{n} W_i z_i \qquad 3.6$$

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.

OK is same type of predictor like IDW. However, in IDW, the weight, λ_i merely depends on the distance to predict for unknown locations. In OK, the weight, W_i, depends not only on the distance, also on the spatial relationship among the measured values around the prediction location. 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[24][27].

$$\sum_{i=1}^{n} w_i = 1 \qquad 3.7$$

To estimation the values of unknown locations using Equation (3.6), OK follows the following procedures.

Procedure 1: Calculate the empirical semivariogram: OK quantifies the basic principles of Tobler's first law of geography as a 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 is represented by the empirical semivariogram model [15].

Procedure 2: Fit a model: a line that provides the best fit through the points in the empirical semivariogram cloud graph.

Procedure 3: 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 [15].

Procedure 4: Make a prediction: using weights of the measured values, which are found from the above step, determine a prediction for the unknown locations using Equation (3.6).

In general, geostatistical analysis of the above four steps can be summarized into two major steps: a) semivariogram (autocorrelation) analysis, b) Kriging estimation (prediction) and mapping [5][25].

3.4.2 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 [15][17][18].

Semivariograms' Spatial modeling begin with a graph of the empirical semivariogram computed as Equation (3.5). Empirical semivariogram provides information on the spatial autocorrelation of datasets.

3.4.2.1 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 [17]. 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., $Cov(z_i, z_j) \le Cov(z_i, z_k)$ if $h_{i,j} \le h_{i,k}$) [13][28]. The semivariogram is a plot of semivariance as a function of distances between the observations (h) and is denoted by $\gamma(h)$ [13].

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, we need to 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 can be computed using the following equation for all pairs of locations separated by distance h [5][29].

$$\gamma(h) = \frac{E\left[Z_{(x)} - z_{(x+h)}\right]^2}{2} \qquad 3.8$$

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; h is the distance between each pair of locations in the x y coordinate system. Euclidean distance is used to calculate the distance between two locations as shown in Equation (3.6)[18].

$$h_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \qquad 3.9$$

3.4.2.2 Binning the empirical Semivariogram

Sometime 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 [15]. Its computational formula is expressed in Equation (3.7).

$$\bar{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{Z_{(x_i)} - Z(x_i + h)\}^2 \qquad 3.10$$

Where: Z(xi) is the value at location xi , Z(xi + h) is the value at location xi +h , and N(h) is the number of pairs of points of observations of the values of attribute Z(xi), Z(xi+h) separated by distance h ; the plot of $\bar{\Upsilon}$(h) against h is known as the empirical semivariogram [5].

3.4.2.3 Fitting a model to 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 prediction 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)[15] [12][29].

Once each pair of locations are plotted after being binned, a model is fit through them. Range, sill, and nugget are commonly used parameters to describe the models as shown in Figure (3.1). The definition and explanation of the parameters are presented as follows.

The range (a) is the distance at which semivariance stops increasing. Range indicates the distance from which samples are spatially independent of each other.

The nugget is the value at which the semi-variogram (almost) intercepts the y-value.

The sill is the value at which the model first flattens out and is the maximum semivariance found between pairs of points.

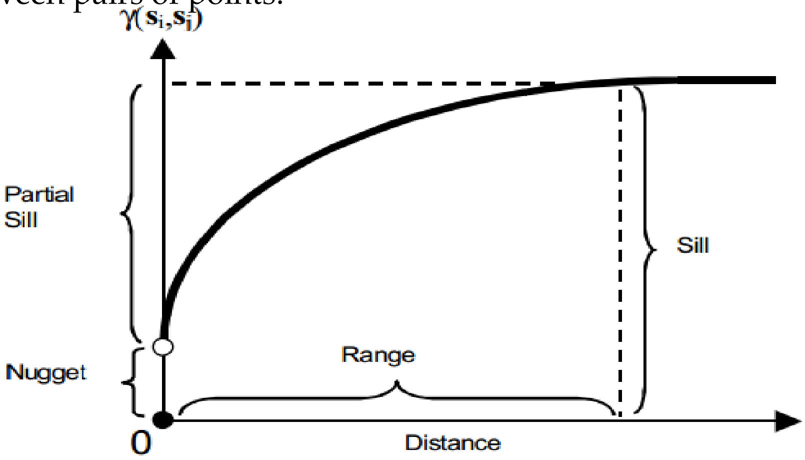


Figure 3- 1 Semivariogram parameters (Range, Sill, and Nugget)[19].

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 [17][29]. Though the most frequently used models in geostatistical calculations are the spherical, the exponential and the Gaussian.

3.4.3 Spatial prediction using Kriging

After computing autocorrelation of 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 Equation (3.8) [18][25].

$$\begin{pmatrix} W1 \\ \cdot \\ \cdot \\ \cdot \\ Wn \\ \cdot \\ 1 \end{pmatrix} = \begin{pmatrix} \gamma(h_{1,1}) & \cdot\cdot\cdot & \gamma(h_{1,n}) & 1 \\ & \cdot & & \\ & \cdot & & \\ \gamma(h_{n,1}) & \cdot\cdot\cdot & \gamma(h_{n,n}) & 1 \\ & \cdot & & \\ & \cdot & & \\ 1 & & 1 & 0 \end{pmatrix} \begin{pmatrix} \gamma(h_{1,u}) \\ \cdot \\ \cdot \\ \gamma(h_{n,u}) \\ \cdot \\ \cdot \\ 1 \end{pmatrix} \qquad 3.11$$

Where: ϒ(hi, j) is a semivariogram which is a function of distance; *hi,j* is the distance between the known points i and *j*; *hi,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.,* $E(\hat{Z}_u - z_u) = 0$)[25][29].

Chapter IV: Experimental Analysis

In this chapter, the experimental process conducted to do this research will be discussed. The main topics that will be discussed are data collection procedures, data preprocessing and interpolation.

4.1 Data Collection

A geotagged spectrum occupancy data is needed to train the interpolation algorithms for occupancy estimation at unknown locations. So, the first task is collecting data using spectrum monitoring device in the working environment of land mobile frequency users. The data collected using TCI the frequency monitoring station is an occupancy data with geographical location. The key parameters obtained are GPS location coordinates, spectrum occupancy in percent (%), frequency range (Starting and ending), bandwidth resolution and duration of time.

Data is collected from the study are shown in Figure 4-1 below. Measurement was taken from 29 different places for 30 minutes starting from 9:00 AM-9:00 AM, due to the shortage of additional measurement devices, measurement is taken separately at each location, not simultaneously.

No.	Frequency Range (137 MHz – 174 MHz) (400 MHz – 470 MHz)	Location		Resolution Bandwidth (RBW)	Occupancy in (%)	Schedule time
		Longitude	Latitude			09:00:31 AM-09:30:00 AM
1.	137.000	9.01099	38.76118	25Khz	45	09:00:31 AM-09:30:00 AM
2.	137.025	9.01031	38.76844	25Khz	42	09:00:31 AM-09:30:00 AM
3.	137.050	9.01085	38.77469	25Khz	0	09:00:31 AM-09:30:00 AM
4.	137.075	9.01414	38.78141	25Khz	52	09:00:31 AM-09:30:00 AM
5.	137.100	9.01566	38.78694	25Khz	78	09:00:31 AM-09:30:00 AM
6.	400.000	9.01099	38.76118	25Khz	100	09:00:31 AM-09:30:00 AM
7.	400.025	9.01031	38.76844	25Khz	84	09:00:31 AM-09:30:00 AM
8.	400.05	9.01085	38.77469	25Khz	62	09:00:31 AM-09:30:00 AM

Table 4-1 Sample data collected from Spectrum monitoring station.

Figure 4-1: Study area Satellite Image.

Figure 4-2: Study area measurement points.

4.2 Data Preprocessing

Data preprocessing is a step that describes any type of process performed on raw data to prepare it for another procedure. This process can contain tasks like parameter selection, integration of data from multiple sources, classification and handling of missing data. In this research, only parameter selection and classification are used as data preprocessing. Details of tasks performed for the selected preprocessing modules will be explained below.

4.2.1 Parameter selection

The collected data have many parameters and contains information that is not significant for this thesis, it needs to be preprocessed before using it as an input for an interpolation algorithm. The data collected from the TCI frequency monitoring station (700 series) have more than 10 parameter features. Out of these features, four of them namely, Latitude, Longitude, Average occupancy (%), Bandwidth, and Center frequency are selected. The selection of parameters is based on referring to related works [5] [25] [29].

Geotagged data	
Selected Parameter	**Description**
Measurement Locations	Longitude and Latitude information
Occupancy	Average occupancy in each location

Table 4-2 Description of selected parameters.

No.	Frequency	Latitude	Longitude	Average Occupancy (%)
1.	137.000	9.01099	38.76118	45
2.	137.025	9.01031	38.76844	42
3.	137.050	9.01085	38.77469	0
4.	137.075	9.01414	38.78141	52
5.	137.100	9.01566	38.78694	78
6.	400.000	9.01099	38.76118	100
7.	400.025	9.01031	38.76844	84
8.	400.05	9.01085	38.77469	92

Table 4-3 Sample selected feature data.

4.2.2 Classification

For this research, the K-means clustering algorithm is used to classify the collected data samples into different classes according to their occupancy distribution with their respected occupancy rate. Classification is needed because occupancy among different locations varies, so to study the occupancy status among each center frequencies in the measurement area classification algorithm is used to group the data.

Data clustering techniques are descriptive data analysis techniques that can be applied to multivariate data set to uncover the structure present in the data. Clustering is an unsupervised form of classification, as the clusters are formed by evaluating similarities and dissimilarities of intrinsic characteristics between cases. It does not require a training dataset.

The two widely known types of clustering methods are partitioning clustering and hierarchical clustering. K-means clustering, where k represents the desired number of clusters, is a type of partitioning clustering. In k-means clustering, each cluster is defined by the centroid (or mean) of the data points in the cluster, grouping is based on iterative relocation of data points between clusters. The goal here is to produce groups of cases/ variables with a high degree of similarities with in each group and a low degree of similarities between groups [30][31].

For this research, the classification algorithm is applied on each occupancy data of the VHF land mobile band frequency range (137 MHz – 174 MHz) and UHF land mobile band frequency range (400 MHz – 470 MHz).

4.2.3 Test Point Selection

After parameter selection and classification tasks are completed, the collected data sets are portioned in to training and testing sets with 80% and 20% respectively. The 80% partition of the data is randomly selected with the probability of 0.8% and for testing with 0.2%. I select 24 measurement points for training and 5 points for testing the interpolation algorithm. The test points are used to validate the performance of interpolation algorithms that are used in this research.

4.3 Prediction Subsystem

This section describes the occupancy prediction task that will be performed by interpolation algorithms. After the data preprocessing and parameter selection task is completed, the next pace is to predict the value for the known locations using spatial interpolation algorithm to develop an occupancy map. Three main points are discussed in this section below: data exploration, development of occupancy prediction map and evaluation of the interpolation algorithm performance.

4.3.1 Data Exploration

Before using the interpolation algorithm, we have to check the data to have a normal distribution and need to be autocorrelated. 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 [15][17]. For this thesis, I explored the distribution and autocorrelation using geostatistical analyst tools in ArcGIS.

4.3.1.1 Normality Check

The histogram is used to check the normal distribution of the data. 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 spectrum occupancy data is close to normal as shown in Figure 4-3.

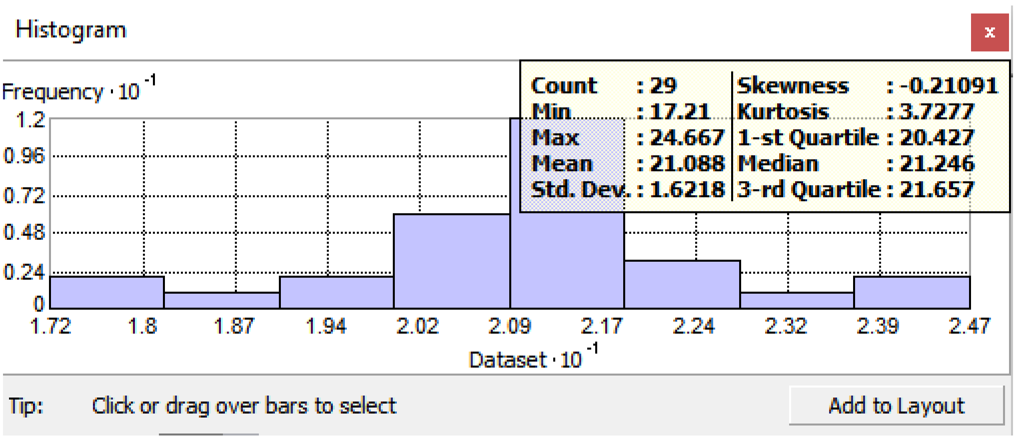


Figure 4-3: Histogram distribution.

4.3.1.2 Spatial Autocorrelation

To examine the spatial autocorrelation in the spectrum occupancy data, the Global Moran's I statistic test is used. Moran's Index is 0.56 with a Z-score of 3.693, indicating that the data is spatially autocorrelated, as illustrated in Figure 4-4.

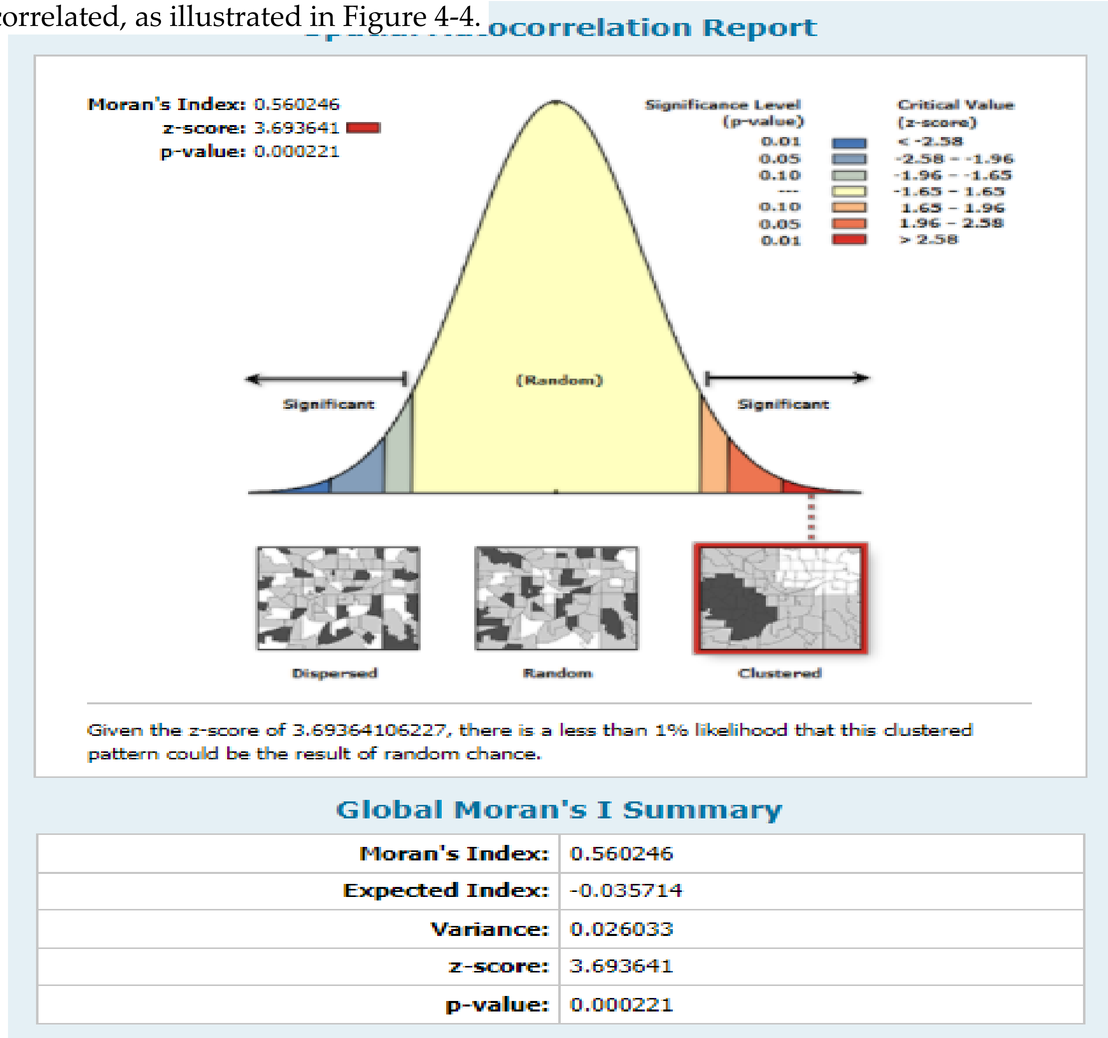


Figure 4-4: Spatial autocorrelation result for Spectrum occupancy data.

4.3.2 Spatial Interpolation Process

The spatial interpolation technique is used once the spectrum occupancy data has been investigated. In this work, three interpolation techniques are investigated: NN, IDW, and OK. Each algorithm uses various procedures to estimate the occupancy value for unknown locations. The details of each procedure as explained below.

4.3.2.1 IDW Method

IDW employs a straightforward distance-based method. The factors that define the distance (d) between measurement points for weight computation in Equation 3-2 are the power exponent (p), neighbor to include (NtoI), and include at least (IatL). The IDW method follows two procedures: parameter estimation and prediction using IDW.

Parameter Estimation using cross-validation

The Cross-validation approach is used to identify the best parameter values. Parameter combination that has the least RSME value is chosen to estimate the value of z(u).

Determine the optimum value of p

Minimizing the prediction error (RMSE), which is measured via cross-validation, yields the optimal P value. The optimal power P is the exponent of the distance, and the lowest RMSE value is calculated. Weights are proportional to the inverse distance raised to the power P, according to Equation 3-2. To achieve the best P-value, the P-value is modified from 1 to 3 without changing any other settings.

The influence of the power exponent P on the interpolation precision values is seen in Tables 4-4 and 4-5 below. The result indicates that when the P value increases from 1 to 3, the RMSE value also increases with it. When P is changed from 1 to 2, however, it has less of an impact on the RMSE value. As a result, 2 is picked as the best P value.

Power Exponent (p)	P=1	P=2	P=3
RMSE	2.12	2.08	2.24

Table 4-4: Prediction error of power exponents for UHF classification.

Power Exponent (p)	P=1	P=2	P=3
RMSE	7.37	7.29	8.2

Table 4-5: Prediction error of power exponents for VHF classification.

Determine the optimum number of Maximum and Minimum neighbors

P remained fixed while NtoI and IatL were adjusted from 10 to 20 and 2 to 10, respectively, to identify the optimum number of neighbors to include (NtoI) and include at least (IatL). Tables 4-6, 4-7, and 4-8 compare the prediction errors across different parameter combination values using the P value of 1, 2, and 3 correspondingly.

NtoI	NtoI=10			NtoI=15			NtoI=20		
IatL	2	5	10	2	5	10	2	5	10
RMSE	2.19	2.19	2.19	2.093	2.093	2.093	2.14	2.14	2.14

Table 4-6: Prediction error of NtoI and IatL with P=1 for UHF classification.

NtoI	NtoI=10			NtoI=15			NtoI=20		
IatL	2	5	10	2	5	10	2	5	10
RMSE	2.17	2.17	2.17	2.081	2.081	2.081	2.106	2.106	2.106

Table 4-7: Prediction error of NtoI and IatL with P=2 for UHF classification.

NtoI	NtoI=10			NtoI=15			NtoI=20		
IatL	2	5	10	2	5	10	2	5	10
RMSE	2.146	2.146	2.146	2.148	2.148	2.148	2.158	2.158	2.158

Table 4-8: Prediction error of NtoI and IatL with P=3 for UHF classification.

The variables P and NtoI were kept constant while determining the optimal value for IatL; as demonstrated in the above tables, when the IatL value changes, the RMSE value does not change. Because IatL has no effect on the interpolation procedure' accuracy, an arbitrary number between 2 and 10 is chosen. As a result, 5 is chosen as the optimal number for minimum neighbors (IatL). RMSE decreases as NtoI increases from 10 to 15 while keeping the variables P and IatL constant. However, as NtoI rises from 15 to 20, RMSE rises as well. As a result, 15 is picked as the optimal number for maximum neighbors (NtoI).

In summary, with an RMSE value of 2.081, the best values for IDW interpolation process parameters are P=2, NtoI=15, and IatL=5.

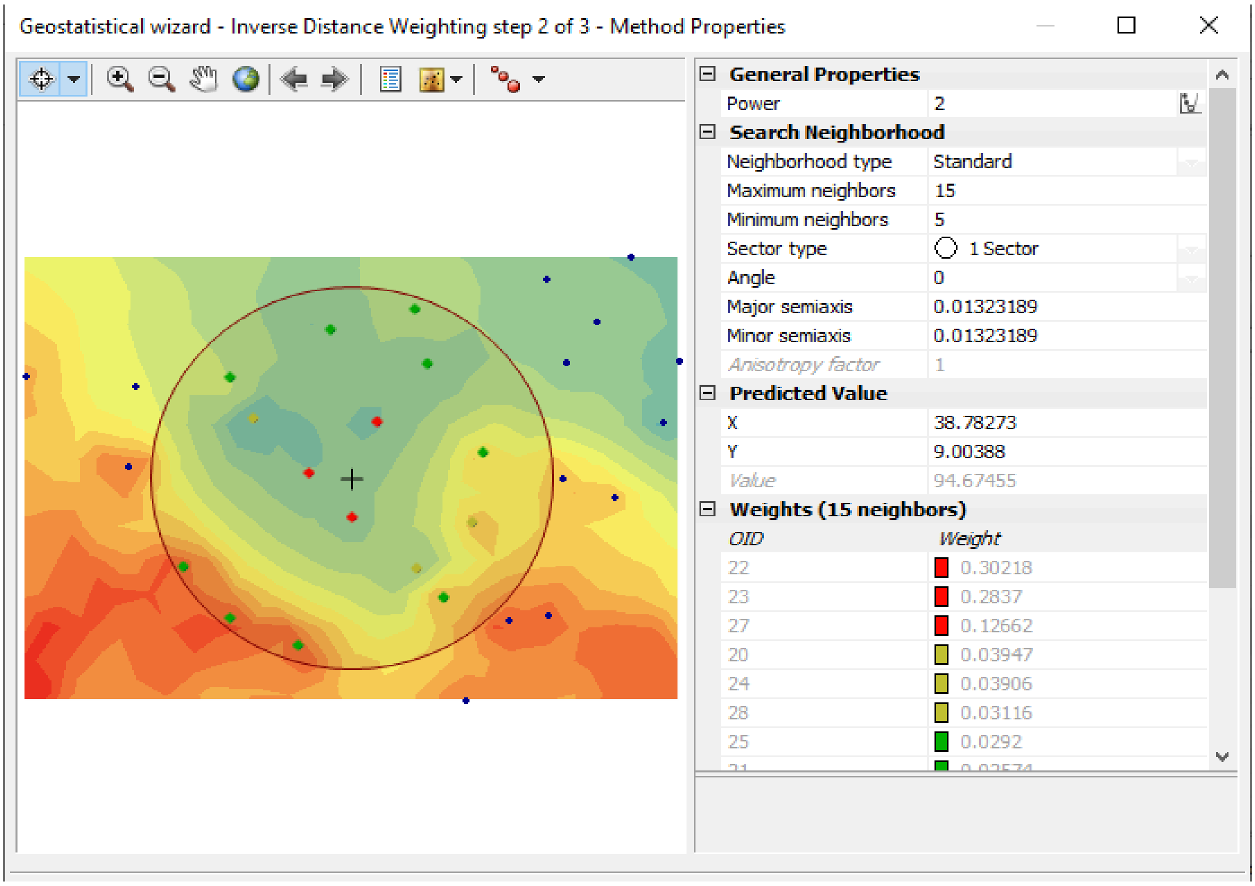


Figure 4-5: Weight of neighbors.

The prediction map is shown in Figure 4-6, which is a contour map of the spectrum occupancy. It displays the value in each region divided by contour lines that indicate the occupancy values, with each line passing through locations with the same occupancy value.

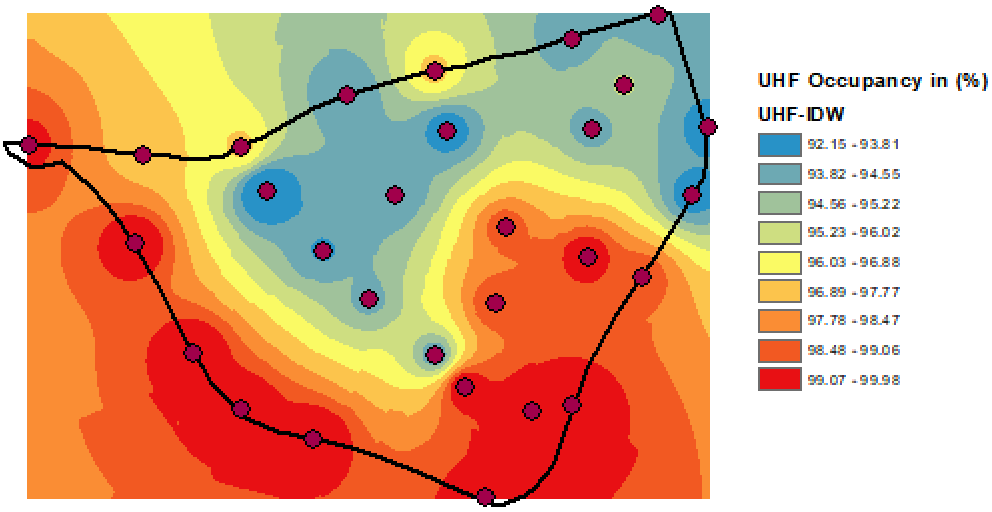
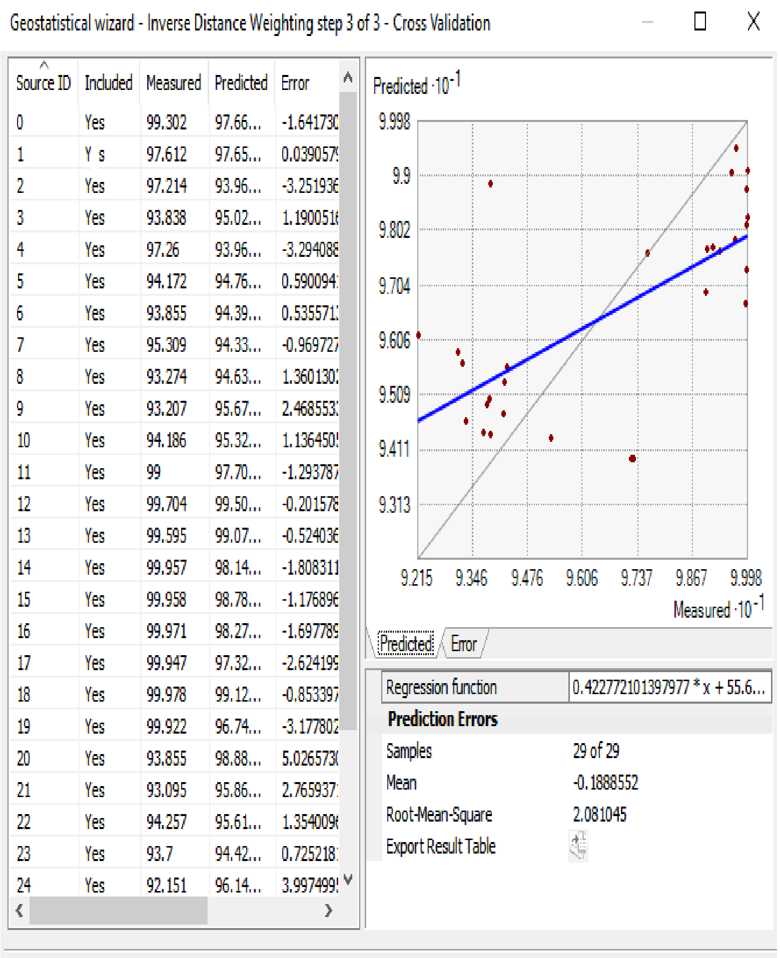


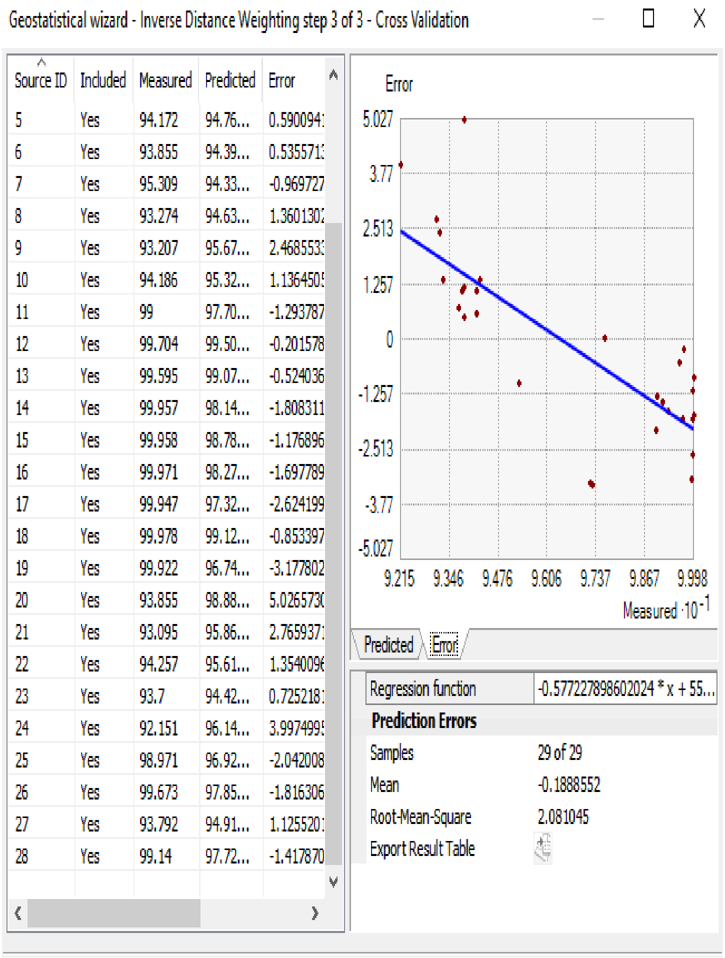
Figure 4-6: Prediction map by IDW-contour map.

Evaluating the accuracy of prediction

The cross-validation tool was used to assess the accuracy of the prediction map created by IDW. For the occupancy training dataset, the prediction error of IDW is provided in two plots: predicted and error plots, as illustrated in Figures 4-7-a and b, respectively. The measured vs. predicted plot shows the measured vs. projected value, whereas the IDW error plot shows the error vs. measured. IDW has an optimal error of prediction with a value of RMSE=2.081, as seen in the figure.



a) Predicted plot of IDW



b) Error plot of IDW

Figure 4-7: Error plot of IDW

4.3.2.2 OK Method

The OK method makes a prediction for locations in the study area based on the semivariogram model 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 [15].

Semivariogram computation

Modeling the measured occupancy data is the initial stage. Semivariogram/covariance modeling was used to model the occupancy data. As illustrated in Figure 4-8, the semivariogram fitting model was generated by graphing the empirical semivariogram versus distance. When the dataset is large, the number of pairings of locations increases quickly, as seen by the empirical semivariogram cloud.

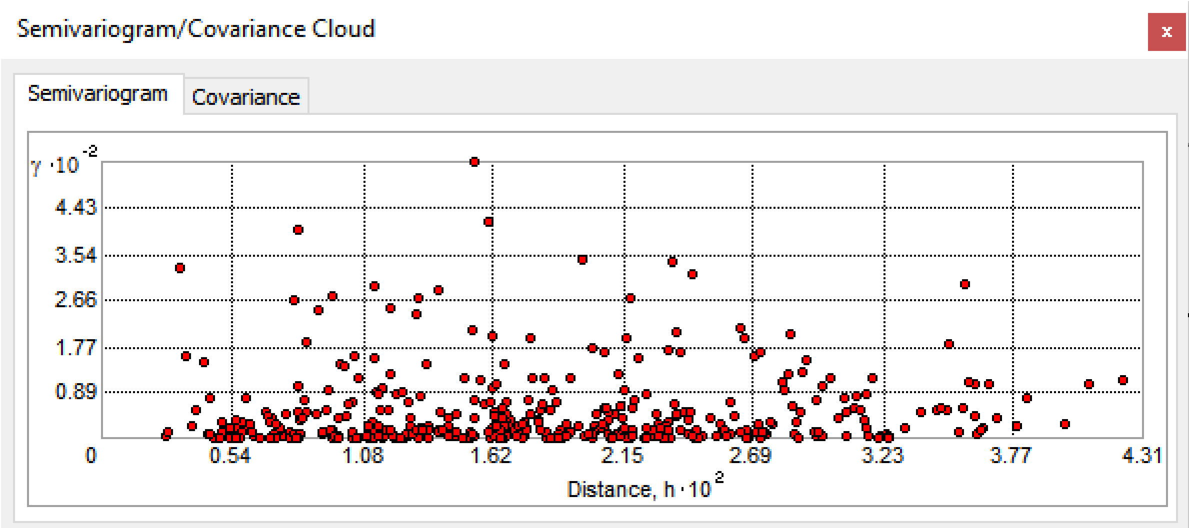


Figure 4-8: Empirical semivariogram cloud

The computation of the semivariogram fitting model gets harder when the dataset is big. As a consequence, the semivariogram cloud's points were grouped into distance classes ("bins") in order to determine the semivariogram fitting model for the target region [18]. Red dots in Figure 4-9 are the binned values. Figure 4-9 shows the binned data as red dots. Semivariogram/covariance points are grouped (binned) to create red dots. The average points are shown by blue crosses, which are produced by binning empirical semivariogram/covariance points that lie inside angular sectors. Binned points indicate local fluctuation in semivariogram/covariance values, whereas average values shows smooth semivariogram/covariance value variation [15].

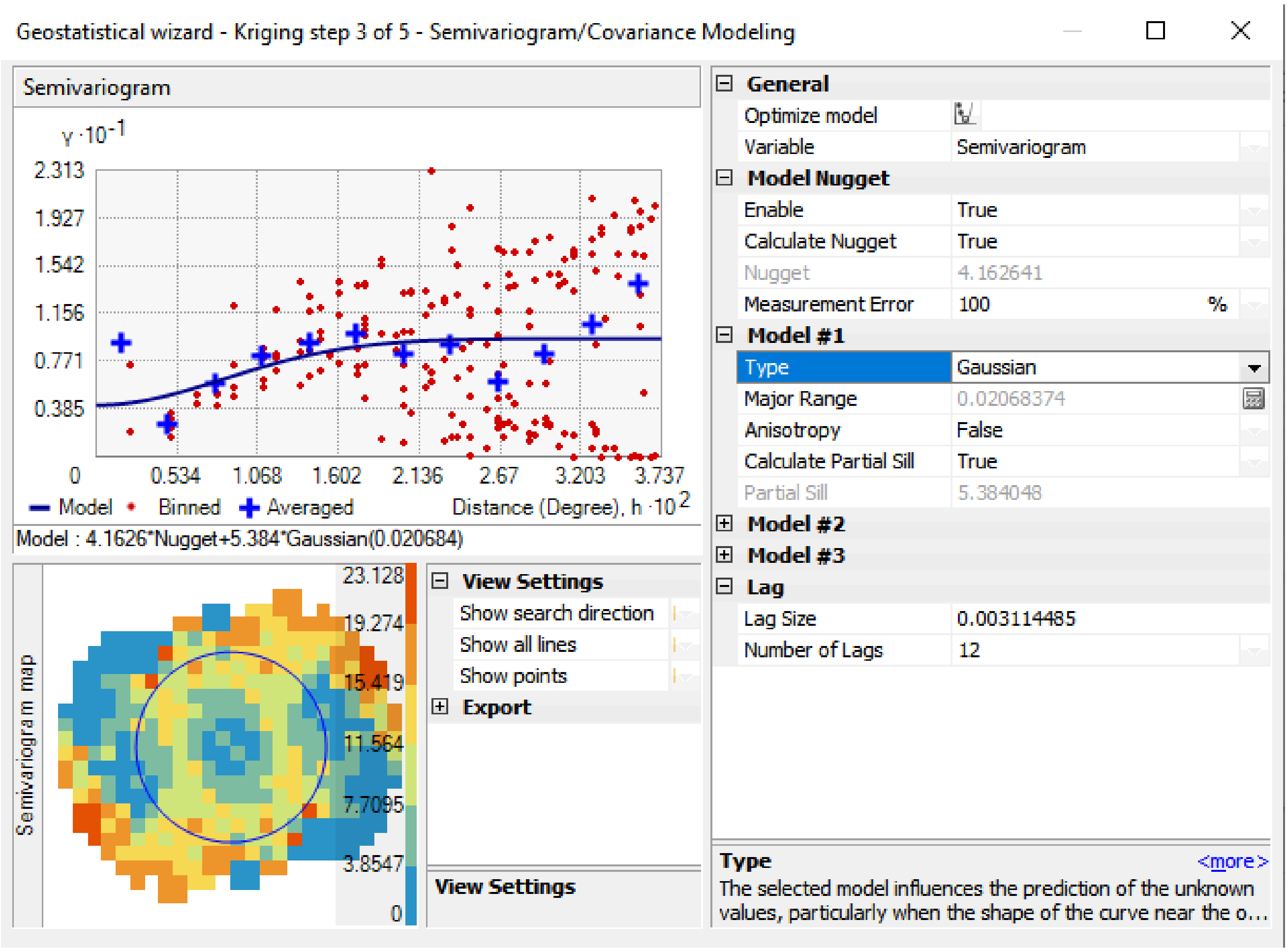


Figure 4-9: Exponential semivariogram model for the selected area.

Select the fitting semivariogram model

Cross-validation is used to find the optimal parameter value. For the interpolation procedure, the parameters with the lowest RMSE value are chosen. Computing the RMSE value for the selected models yielded the best model that suited the points in the semivariogram cloud, as shown in Table 4-9 below. The RMSE values of the three chosen models are compared.

Model	Nugget	Partial Sill	Sill	Range	RMSE
Gaussian	4.162	5.384	9.546	0.02	0.798
Spherical	1.972	7.279	9.251	0.019	0.851
Exponential	1.015	9.304	10.319	0.028	0.871

Table 4-9: Prediction Error of Semivariogram models

The RMSE of the Gaussian model is lower than the RMSE of the Spherical and Gaussian models, as seen in Table 4.9. As a result, given occupancy data, the Gaussian model is chosen as the best-fitted model. A model should be fitted to the observed data points once the empirical variogram has been calculated. The Gaussian fitting variogram model was chosen as the best suited model for the measurements of the occupancy data as shown in Figure 4.14, based on the parameter selections listed in Table 4.9. The values of parameters such as nugget, range, and partial sill are determined by the semivariogram model used.

Determine the optimum number of Maximum and Minimum neighbors.

The next procedure is to determine 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 [17]. Include at least (IatL) is the minimum number of neighbors surrounding the estimated point u, whereas neighbor to include (NtoI) is the maximum number of neighbors. The parameters with the lowest RMSE are chosen.

NtoI	**NtoI=10**			**NtoI=15**			**NtoI=20**		
IatL	2	5	10	2	5	10	2	5	10
RMSE	0.807	0.807	0.807	0.803	0.803	0.803	0.801	0.801	0.801

Table 4-10: Kriging Prediction Error

Keeping the number of NtoI constant while selecting the optimal number of IatL; the RMSE value did not change when the number of IatL was increased from 2 to 10, as shown in Table

4.10. As a result, any value in the middle can be picked, and 5 is chosen as the best number for IatL. By raising the number of NtoI from 10 to 20, the optimum number of NtoI was found, and the RMSE value decreased.

As a result, the Gasussian semivariogram model with NtoI = 20 and IatL = 5 is chosen as the parameter combination for OK interpolation.

OK uses Equation 3.10 to estimate the weights of zu's neighbors after determining the values of the parameters. The weights for the observed locations were calculated using the fitted semivariogram model and the data configuration inside the designated neighbor. Figure 4-10 shows the weights of all neighbor points as well as the value of zu.

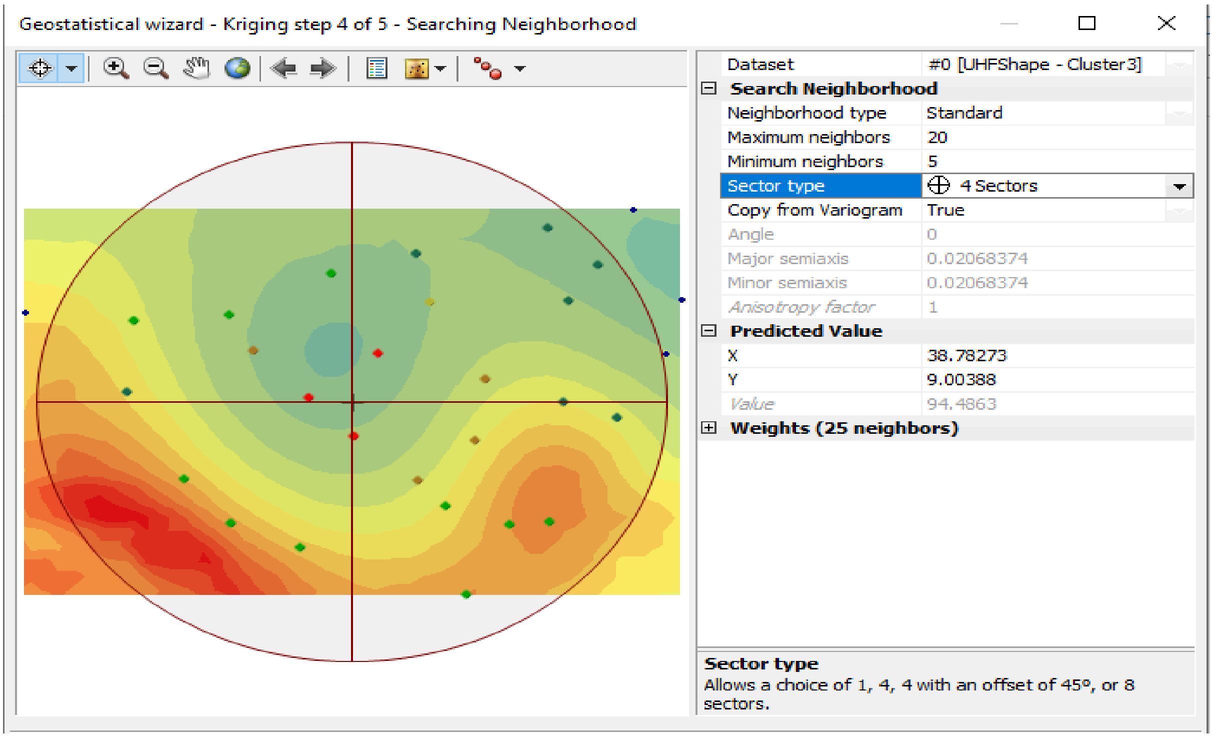


Figure 4-10: Weight of neighbors-Kriging.

The weights and values of the known measured points are used to estimate the unknown value, zu, for the prediction location u. To construct a model of the continuous surface, this process is repeated for each spatial point.

Evaluating the accuracy of prediction

The accuracy of the prediction map is evaluated as the final stage in Ordinary kriging. The cross-validation tool was used to assess the accuracy of the prediction. As illustrated in Figure 4-11, the projected plot of OK shows the measured vs. predicted value. OK has an optimal error of prediction, as shown in the figure, with an RMSE value of 0.801.

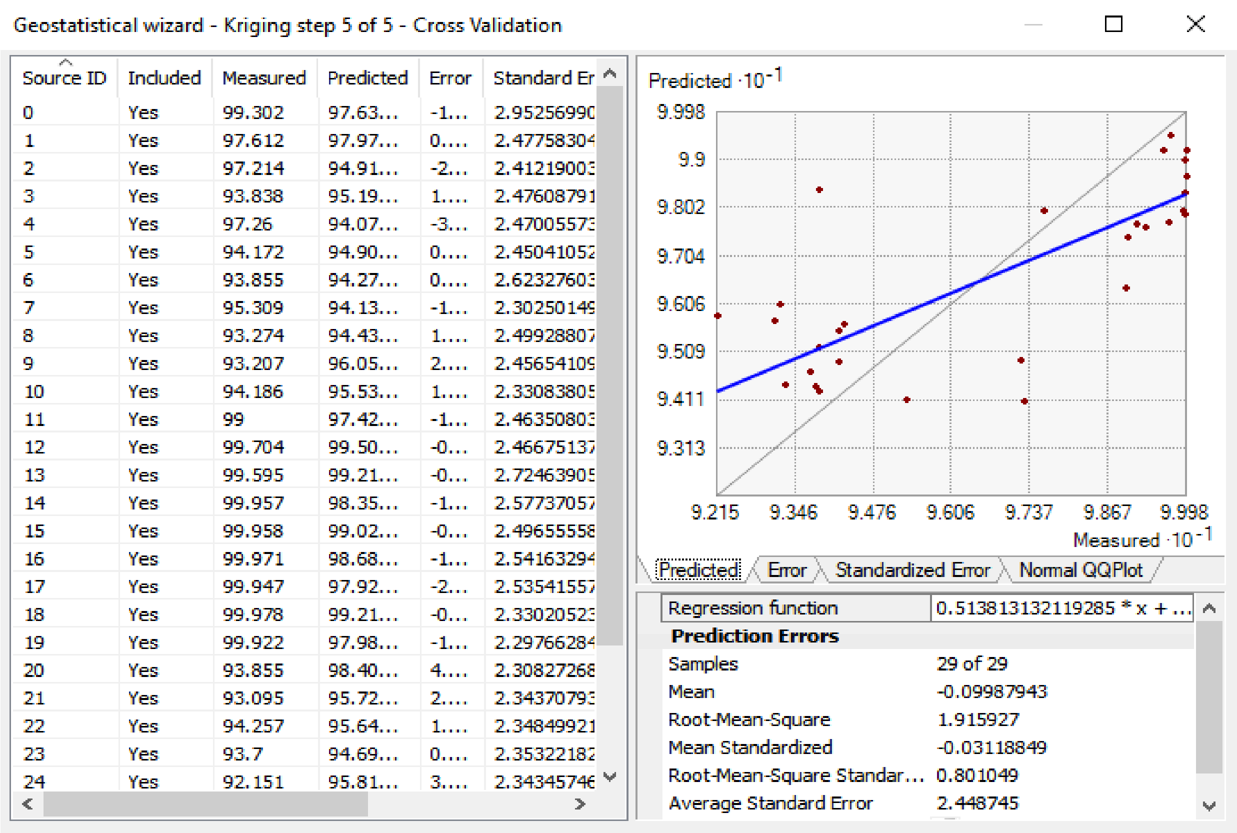


Figure 4-11: Predicted vs measured plot of OK.

4.4 Performance Evaluation of Different Interpolation Techniques

4.4.1 Cross-validation

To determine which model makes the best accurate predictions, cross-validation is performed. The RMSE of OK 0.807, IDW 2.081, and NN 2.11 was determined by cross-validation. Figures 4-12 and 4-13 illustrate the RMSE comparison between OK and IDW and OK and NN, respectively.

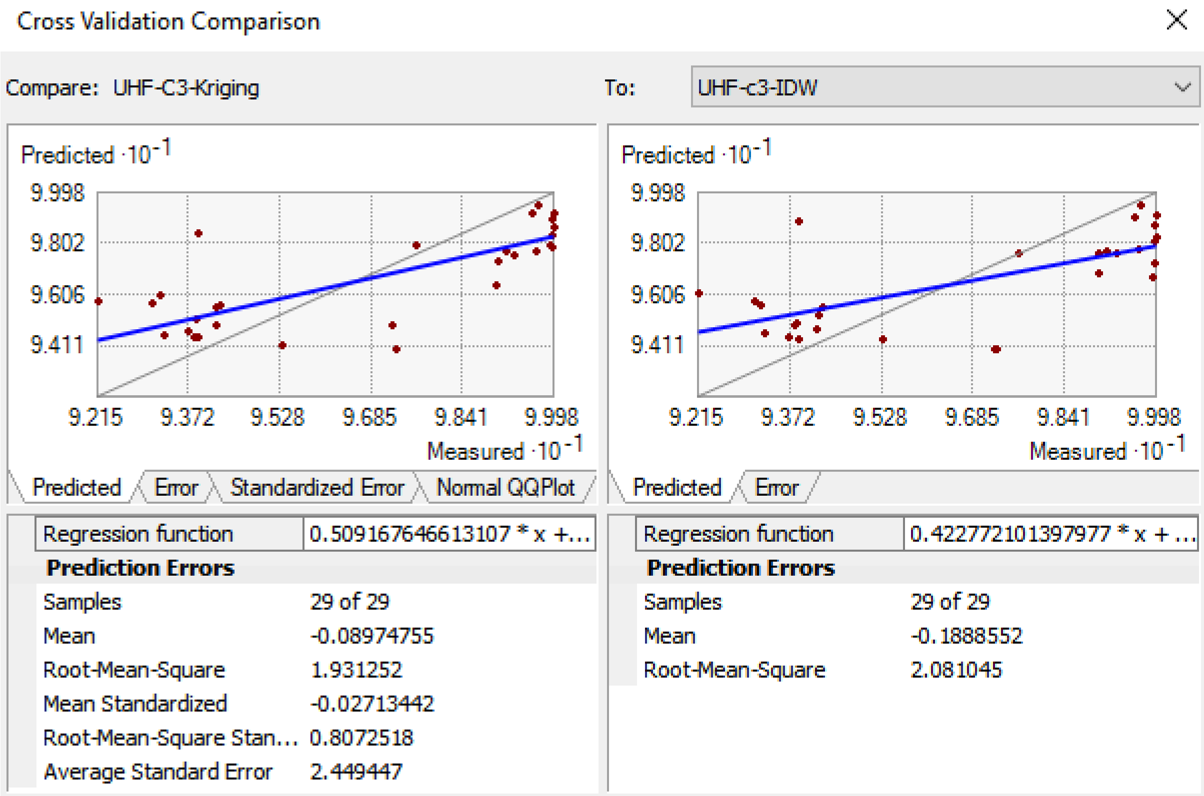


Figure 4-12: Cross-validation of OK and IDW

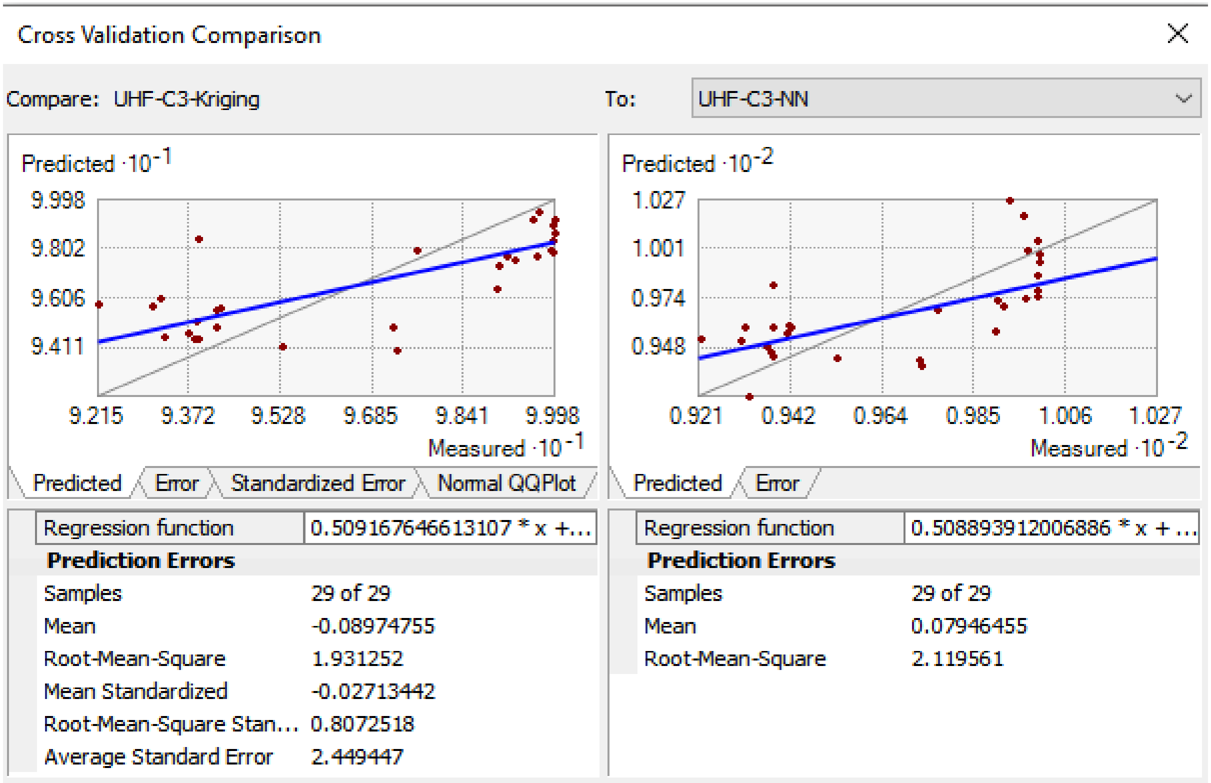


Figure 4-13: Cross-validation of OK and NN.

4.4.2 Statistical Method

Test points are used to compare the findings obtained. Each map's prediction values for the test points are retrieved, and the difference between them and the actual value is calculated. To compare the accuracy of each interpolation method, the RMSE value is utilized. Table 4.11 shows that OK with a prediction error of 0.807 RMSE is more accurate than IDW 2.081 and NN 2.119.

Interpolation Method	OK	IDW	NN
RMSE value	0.807	2.081	2.119

Table 4-11: Prediction Error of Interpolation methods.

Chapter V: Result and Discussion

5.1 Result

The primary goal of this work is to create a spectrum occupancy map in the space domain, using Addis Ababa as a case study. The occupancy prediction maps generated by the chosen interpolation method, as well as their evaluation outcomes, are presented in this chapter. The root-mean-square error is computed for each interpolation technique to pick the optimum model to build an occupancy map. Different interpolation methods, IDW, OK, and NN, are used to establish the spectrum occupancy prediction model.

As shown in Table 5.1 below, the RMSE of IDW with parameters P=2, NtoI=9 and IatL=4, 2.801, OK with Gaussian semivariogram model, NtoI=11 and IatL=4, 0.807 and NN 2.119. This shows that OK is more preferable that IDW and NN for occupancy prediction.

Parameter Values	Interpolation	RMSE
P=2, NtoI=9 and IatL=4	IDW	2.801
Gaussian model, NtoI=11 and IatL=4	OK	0.807
NtoI=9 and IatL=4	NN	2.119

Table 5.1: The statistic errors in the process of occupancy interpolation.

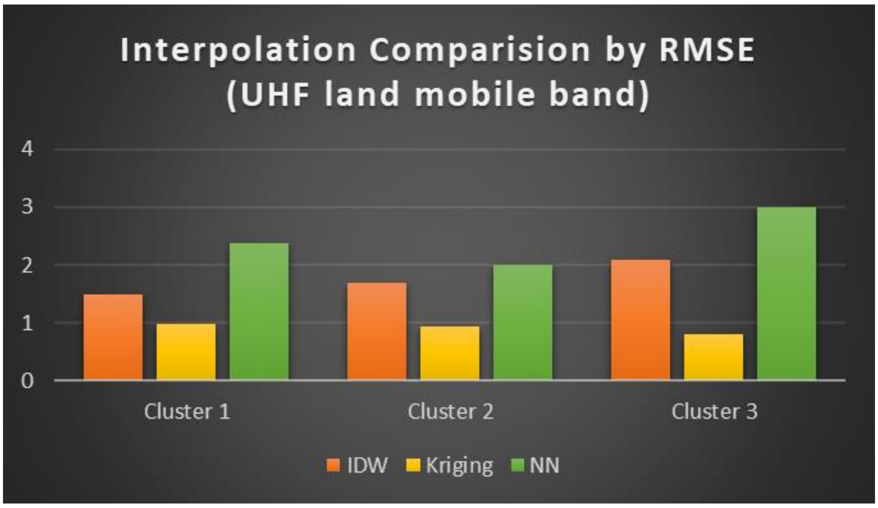


Figure 5-1: Prediction error for IDW, OK and NN in UHF bands.

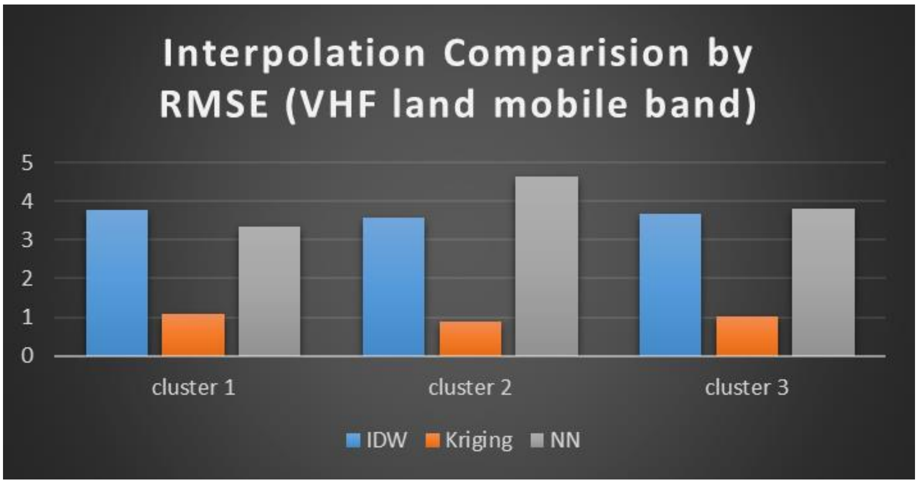


Figure 5-2: Prediction error for IDW, OK and NN in VHF bands.

The kriging spatial interpolation technique is used to produce a spectrum occupancy prediction map for the specified area based on occupancy data gathered from spectrum measurement stations, as illustrated in Figure 5-3 below. The prediction map's red color shows areas with greater occupancy rates, while the lighter colors suggest areas with lower occupancy rates.

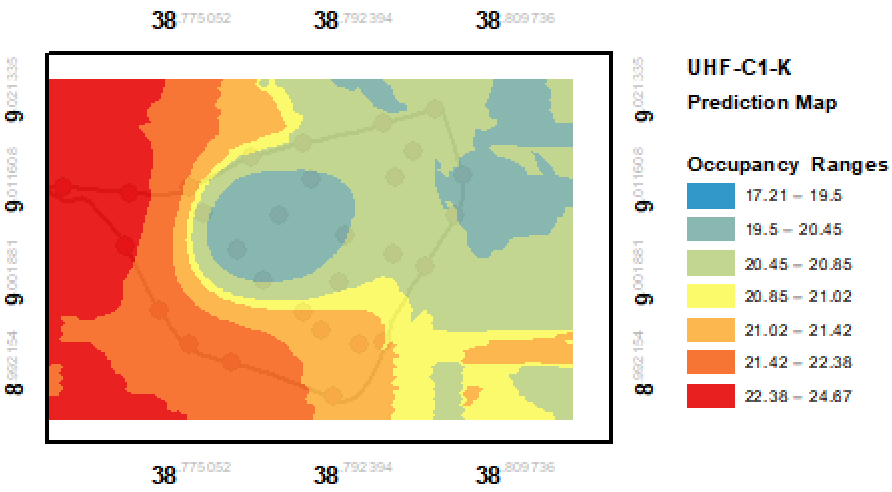


Figure 5-3: Contour map for UHF band of cluster1.

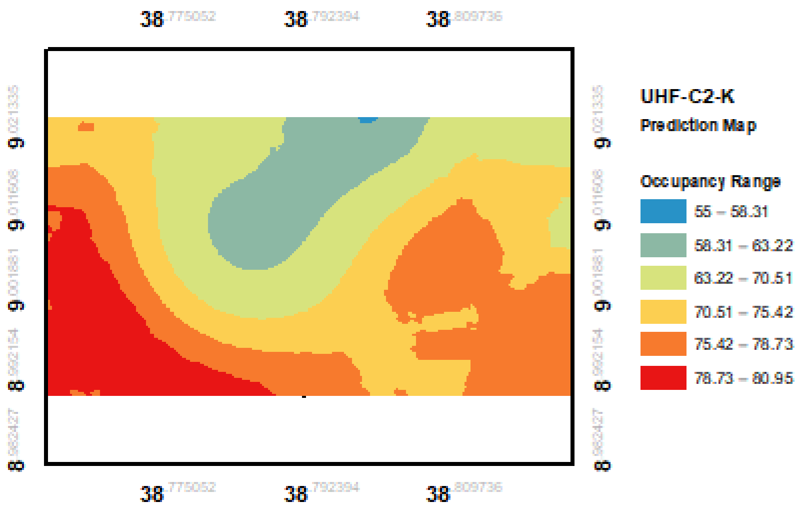


Figure 5-4: Contour map for UHF band of cluster2.

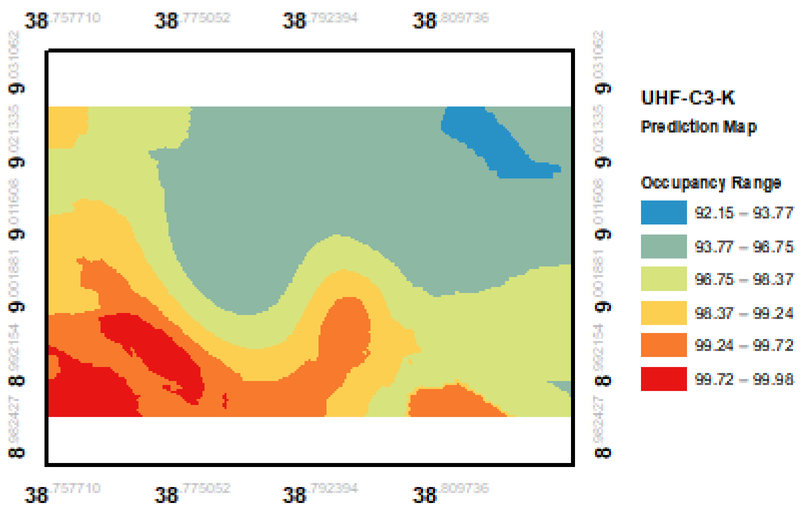


Figure 5-5: Contour map for UHF band of cluster3.

Another benefit of OK is that it generates a standard error map, which depicts the level of uncertainty associated with the projected values. Areas that are close to where the measurement is taken have less prediction errors than those that are far away, as illustrated in Figure 5-6 below. The light color implies a low prediction error, whereas the dark red color suggests a large forecast error.

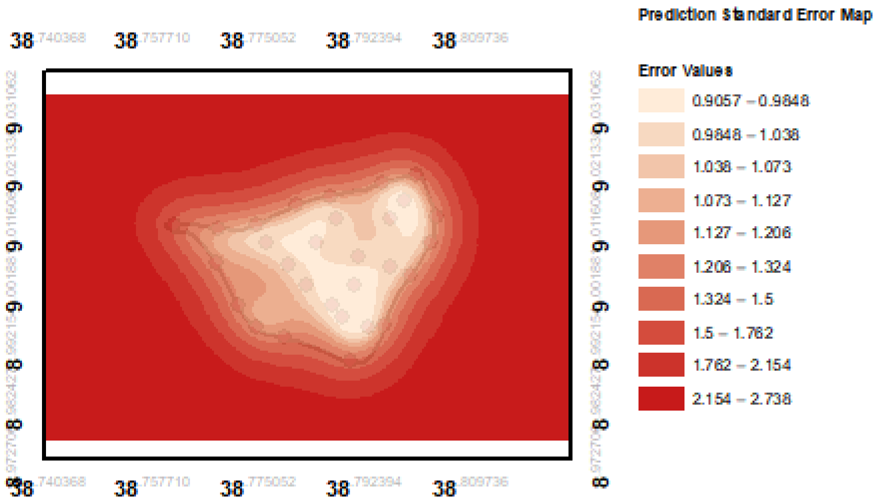


Figure 5-6: OK standard Error Map.

5.2 Discussion

In general, the interpolated surfaces created with the OK interpolation technique show that the OK performs better than the other two interpolation methods utilized in this thesis. However, the primary goal of this research is to create a spectrum occupancy map and compare the results of various interpolation methods using statistical criteria. As a result, the statistical approach RMSE would be used to analyze the interpolated.

The OK technique is 28.75 percent more accurate than IDW and 38 percent more accurate than NN, according to the findings, with an RMSE score of 0.807. The OK technique takes into account the impacts of the distance parameter as well as the semivariogram model as a parameter. As a result, the kriging model's interpolation impact is superior than that of IDW and NN. The OK interpolation is more approximate to the occupancy prediction data in terms of interpolation outcomes. OK's map appears to be smooth, indicating that the interpolation result is good. Based on the research above, the OK interpolation approach is determined to be the best interpolation model for this type of data.

As a result, based on the gathered sample data from the spectrum measurement station, the OK method with gaussian semivariogram model with maximum neighbor 11 and minimum neighbor 4 variables can be selected as an ideal model for spectrum occupancy map prediction.

Chapter VI: Conclusion and Future Work

6.1 Conclusion

In this thesis work, an optimized and accurate spatial interpolation techniques for spectrum occupancy map development for the case of land mobiles in the range of 137 MHz-174 MHz for VHF bands and 400 MHz-470 MHz for UHF bands is proposed by combining IDW, OK and NN interpolation algorithms. Spectrum occupancy data was collected using TCI 700-series spectrum measurement station from the selected study area in Addis Ababa city. After data was collected K-means clustering algorithm is used to identify the occupancy distribution status in both VHF and UHF bands.

The purpose of this thesis to develop a spectrum occupancy to maximize the know how knowledge on spectrum occupancy status of land mobiles ranges using interpolation algorithms, which creates occupancy map, using certain measurement points and reduce the time consuming and labor intensive data collection approach.

The result obtained shows that, all the proposed occupancy prediction algorithms in this paper are cable to produce occupancy prediction map. However, based on the Gaussian model of semivariogram with an optimal number of neighbors OK used is 28.75% more accurate that IDW and 38% more accurate than that of NN with acceptable error of 0.807 RMSE value. OK can predict the missing occupancy data for those locations in which measurement was not taken and it generates accurate predicted information for unmeasured locations than IDW and NN.

6.2 Future Work

The study considers the city of Addis Ababa's' selected area for the research. However, a similar approach can be followed for different coverage areas, even as big as the whole country. In this research only space domain is incorporated, it would be better if time domain is considered to study the spatio-temporal occupancy status. Only three spatial interpolation algorithm are discussed here; but there are other interpolation methods that are not implemented here, such as spline, and Global polynomial interpolation from deterministic types and Universal Kriging and CoKriging from geostatistical interpolation types, could be applied in the future studies for occupancy map development.

References

[1] M. Riahi Manesh, S. Subramaniam, H. Reyes, eta N. Kaabouch, «Real-time spectrum occupancy monitoring using a probabilistic model», *Comput. Networks*, libk. 124, or. 87–96, 2017, doi: 10.1016/j.comnet.2017.06.003.

[2] Radiocommunication Sector of ITU, «Spectrum occupancy measurements and evaluation», libk. 2256, or. 53, 2016.

[3] S. Debroy, S. Bhattacharjee, eta M. Chatterjee, «Spectrum Map and Its Application in Resource Management in Cognitive Radio Networks», *IEEE Trans. Cogn. Commun. Netw.*, libk. 1, zenb. 4, or. 406–419, 2015, doi: 10.1109/TCCN.2016.2517001.

[4] S. Grimoud, B. Sayrac, S. Ben Jemaa, eta E. Moulines, «An algorithm for fast REM construction», *Proc. 2011 6th Int. ICST Conf. Cogn. Radio Oriented Wirel. Networks Commun. CROWNCOM 2011*, or. 251–255, 2011, doi: 10.4108/icst.crowncom.2011.245788.

[5] Z. Han, J. Liao, Q. Qi, H. Sun, eta J. Wang, «Radio environment map construction by kriging algorithm based on mobile crowd sensing», *Wirel. Commun. Mob. Comput.*, libk. 2019, 2019, doi: 10.1155/2019/4064201.

[6] S. Debroy, S. Bhattacharjee, eta M. Chatterjee, «Spectrum Map: Toward Predicting the Spatial Distribution of Spectrum Usage in CRNs», zenb. February 2015, 2012, doi: 10.4108/icst.crowncom.2011.245923.

[7] H. B. Yilmaz, C. B. Chae, eta T. Tugcu, «Sensor placement algorithm for radio environment map construction in cognitive radio networks», *IEEE Wirel. Commun. Netw. Conf. WCNC*, zenb. January, or. 2096–2101, 2014, doi: 10.1109/WCNC.2014.6952633.

[8] A. Rabanimotlagh, P. Janakaraj, eta P. Wang, «Optimal crowd-augmented spectrum mapping via an iterative Bayesian decision framework», *Ad Hoc Networks*, libk. 105, 2020, doi: 10.1016/j.adhoc.2020.102163.

[9] P. Lange, «Ethiopian National Frequency Allocation Table».

[10] T. Zhang, N. Leng, eta S. Banerjee, «P17-Zhang», *Acm Mobicom*, or. 17–28, 2014.

[11] R. Li eta J. Li, «A novel clouds based spectrum monitoring approach for future monitoring network», *2014 2nd Int. Conf. Syst. Informatics, ICSAI 2014*, zenb. Icsai, or. 520–524, 2015, doi: 10.1109/ICSAI.2014.7009342.

[12] A. Kumar, S. Maroju, eta A. Bhat, «Application of ArcGIS geostatistical analyst for interpolating environmental data from observations», *Environ. Prog.*, libk. 26, zenb. 3, or. 220–225, 2007, doi: 10.1002/ep.10223.

[13] W. Xu, Y. Zou, G. Zhang, eta M. Linderman, «A comparison among spatial interpolation techniques for daily rainfall data in Sichuan Province, China», *Int. J. Climatol.*, libk. 35, zenb. 10, or. 2898–2907, 2015, doi: 10.1002/joc.4180.

[14] S. B. Babu, «Comparative Study on the Spatial Interpolation Techniques in GIS», *Int. J. Sci. Eng. Res.*, libk. 7, zenb. 2, or. 550–554, 2016, [Sarean]. Available at: https://www.ijser.org/researchpaper/Comparative-Study-on-the-Spatial-Interpolation-Techniques-in-GIS.pdf.

[15] Z. Kassaw, «Coverage Prediction Based on Spatial Interpolation Techniques : The Case of UMTS Network in Addis Ababa , Ethiopia», 2020.

[16] E. Camizuli eta E. J. Carranza, «Exploratory Data Analysis (EDA)», *Encycl. Archaeol. Sci.*, zenb. 3, or. 1–7, 2018, doi: 10.1002/9781119188230.saseas0271.

[17] K. Johnston, J. M. Ver Hoef, K. Krivoruchko, eta N. Lucas, «Using ArcGIS geostatistical analyst», *Analysis*, libk. 300, zenb. January 2004, or. 300, 2001, [Sarean]. Available at: http://direitosminerarios.com/pdf/ESRI - Using ArcGIS Geostatistical Analyst.pdf.

[18] K. Krivoruchko eta N. Lucas, «Using ArcGIS geostatistical analyst», zenb. January 2004, 2014.

[19] J. McCoy, K. Johnston, S. Kopp, B. Borup, J. Willison, eta B. Payne, «Using ArGis Spatial Analyst», *Esri*, or. 238, 2002, [Sarean]. Available at: http://downloads.esri.com/support/documentation/ao_/776Using_Spatial_Analyst.pdf.

[20] R. C. J. W. Sibson, «A brief description of natural neighbor interpolation», 1981, or. 21–36.

[21] «COMPARISON BETWEEN NATURAL NEIGHBOR INTERPOLATION METHOD».

[22] R. Olivier eta C. Hanqiang, «Nearest Neighbor Value Interpolation», *Int. J. Adv. Comput. Sci. Appl.*, libk. 3, zenb. 4, or. 1–6, 2012, doi: 10.14569/ijacsa.2012.030405.

[23] C. Y. Wu, J. Mossa, L. Mao, eta M. Almulla, «Comparison of different spatial interpolation methods for historical hydrographic data of the lowermost Mississippi River», *Ann. GIS*, libk. 25, zenb. 2, or. 133–151, 2019, doi: 10.1080/19475683.2019.1588781.

[24] C. A. Rishikeshan, S. K. Katiyar, eta 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*, or. 666–671, 2014, doi: 10.1109/CICN.2014.148.

[25] A. Konak, P. O. Box, eta T. Road, «A kriging approach to predicting coverage in wireless networks», libk. 3, zenb. 2, 2009.

[26] Y. Xiao *et al.*, «Geostatistical interpolation model selection based on ArcGIS and spatio-temporal variability analysis of groundwater level in piedmont plains, northwest China», *Springerplus*, libk. 5, zenb. 1, 2016, doi: 10.1186/s40064-016-2073-0.

[27] M. Suchański, P. Kaniewski, J. Romanik, E. Golan, eta K. Zubel, «Radio environment maps for military cognitive networks: density of small-scale sensor network vs. map quality», *Eurasip J. Wirel. Commun. Netw.*, libk. 2020, zenb. 1, or. 195–207, 2020, doi: 10.1186/s13638-020-01803-4.

[28] S. Wang, G. H. Huang, Q. G. Lin, Z. Li, H. Zhang, eta Y. R. Fan, «Comparison of

interpolation methods for estimating spatial distribution of precipitation in Ontario, Canada», *Int. J. Climatol.*, libk. 34, zenb. 14, or. 3745–3751, 2014, doi: 10.1002/joc.3941.

[29] «Introduction to Geostatistics ardossy».

[30] H. Zhou, *Learn Data Mining Through Excel*. 2020.

[31] L. Morissette eta S. Chartier, «The k-means clustering technique: General considerations and implementation in Mathematica», *Tutor. Quant. Methods Psychol.*, libk. 9, zenb. 1, or. 15–24, 2013, doi: 10.20982/tqmp.09.1.p015.

Appendix

Analysis of Land Mobile Spectrum Occupancy using Spatial Interpolation Algorithms.

Dereje Hailemariam (Phd) derejehmr@gmail.com Rufael Getachew rufaelgetachew@gmail.com

School of Electrical and Computer Engineering,

Addis Ababa Institute of Technology, Addis Ababa University.

Abstract

An increase in the development of information industry and wireless communication services shows an increase demand for spectrum resource. Spectrum is a limited resource that needs to be managed properly to enable additional new services, without spectrum, none of todays' wireless communication will be true. The spectrum occupancy map can be a powerful tool for developing a better knowledge on the occupancy status of this scares resource. Spatial interpolation techniques are commonly employed for creating continuous data (raster data) from a distributed set of data points over a geographical region. In this paper, the comparison between three spatial interpolation techniques Ordinary Kriging (OK), Natural Neighboring (NN) and Inverse Distance Weighting (IDW)) is done. The goal is to determine which method creates the best spectrum occupancy map for measured occupancy data in Addis Ababa, Ethiopia. These maps show spatial variation in the spectrum usage. OK method results smoother map and lesser error than IDW and NN.

Keywords: *Spatial Interpolation, Ordinary Kriging, Inverse Distance Weight, Natural Neighbor.*

1. Introduction

Spectrum is a limited resource that needs to be managed properly to enable new services and applications. One of the services provided by spectrum is wireless communication. Nowadays this communication service is fundamental for our daily activities; which may include mobile phone calls, internet service, broadcasting services, and other communication services that we can use on our mobile phones.

The continuous increase in wireless systems, attractive applications and, subscribers growth have created an increasing demand for this resource and motivated the search for new strategies for efficiently utilizing and taking the greatest advantage of the radio spectrum [1].

According to research reported in [3], measurements on radio spectrum usage have revealed an unoccupied bands of spectrum that belong to primary (licensed) networks. Prior knowledge about the occupancy of such bands and the expected achievable performance on those bands can help secondary (unlicensed) networks to devise effective strategies to improve utilization[3].

To deploy new and additional radio services, spectrum occupancy knowledge is mandatory. To improve the spectrum occupancy knowledge we can build a spectrum occupancy map to understand the occupancy status. The map can offer multi-domain environmental information, such as geographical information, available services on an area, locations of spectrum usage, occupancy status, and time of use of radio services.

In [4] spectrum occupancy map is suggested as a promising concept for storing radio environmental information that can be used to enhance radio resource management in wireless networks. In cellular networks, the radio environment map (REM) can be used to improve the resource utilization, or to minimize the operational costs by replacing or at least minimizing drive tests (MDT), for troubleshooting of interference for instance [4]. The concept of collecting geo-located information on the radio environment and developing an occupancy map using gathered information has also been investigated and developed further by many other research groups.

2. Related Work

A study conducted in [5] proposes to apply Mobile Crowd sourcing (MCS) to collect the radio information in the area of 150m^2 in 2.4GHz band and propose to apply Kriging interpolation algorithm to predict the spectrum values in transmission power (dBm) of unmeasured locations based on the measurement data collected at locations where measurement can be taken. They used the Kriging interpolation algorithm to understand the spatiotemporal correlation between radio environment data, and the Kriging interpolation algorithm can be applied to infer the missing radio environment information data [5].

Research conducted in [8] exploits a crowd-augmented spectrum mapping scheme to develop spectrum mapping; in which the objective is to create a power spectrum density (PSD) map for an area of interest, where each location on the map is associated with its corresponding PSD value. The authors in this study present an iterative Bayesian decision framework to construct an accurate representation of the spectrum map of the sensed geographical area. In this study, Bayesian spatial prediction was utilized to address the limitations in a two-phase kriging scheme by implementing the optimal predictor, whose prediction error can be accurately assessed.

3. Description of the Study Area

A geotagged spectrum occupancy data is needed to train the interpolation algorithms for occupancy estimation at unknown locations. Data is collected using TCI 700 series frequency monitoring station and an occupancy data with geographical location is obtained in the range of VHF land mobile bands between 137MHz to 174MHz and UHF land mobile bands of 400MHz to 470MHz. The key parameters obtained are GPS location coordinates, spectrum occupancy in percent (%), frequency range (Starting and ending), bandwidth resolution and duration of time.

Figure 1: Study Area

No.	Frequency Range (137 MHz – 174 MHz) (400 MHz – 470 MHz)	Location		Resolution Bandwidth (RBW)	Occupancy in (%)	Schedule time
		Longitude	Latitude			09:00:31 AM- 09:30:00 AM
1.	137.000	9.01099	38.76118	25Khz	45	09:00:31 AM- 09:30:00 AM
2.	137.025	9.01031	38.76844	25Khz	42	09:00:31 AM- 09:30:00 AM
3.	137.050	9.01085	38.77469	25Khz	0	09:00:31 AM- 09:30:00 AM
4	400.000	9.01099	38.76118	25Khz	100	09:00:31 AM- 09:30:00
5.	400.025	9.01031	38.76844	25Khz	84	09:00:31 AM- 09:30:00
6.	400.05	9.01085	38.77469	25Khz	62	09:00:31 AM- 09:30:00 AM

Table 1: Sample data collected

4. Methodology

Spatial interpolation technique is a method designed to predict the RSS value at unknown geographical points using known values at neighboring locations [12]. The estimation of values at unknown locations is based on a theoretical foundation of the Tobler's first law of geography. Which is, 'Everything is related to everything else, but near things are more related than distant things'. Meaning that adjacent measuring points are more similar than those measuring points that are far away from each other [13].

Three methods are examined in this study: IDW, OK and NN, which represent two kinds of interpolation methods – deterministic and geostatistical methods.

The experiment method includes four steps: (1) determining the parameters of IDW such as the Neighbors to Include and the Include at Least of IDW method, and NN (2) determining the semivariogram model and the parameters of OK and CK methods, (3) choosing appropriate training and testing data ratio and (4) comparing the three methods for the occupancy interpolation results of the RMSE value.

4.1 IDW Method

IDW is a deterministic method for spatial interpolation based on similarity or smoothness within a research area. It employs a straightforward distance-based method. The assigned values to unknown points are calculated as a weighted average of the values available at the known points. The weight increases as the distance decreases from the known points. The general equation can be expressed as

$$\hat{z}(s_0) = \sum_{I=1}^{N} \lambda_i z(s_i) \quad \textbf{(1)}$$

Where: $\hat{Z}$ *(so)* is the value which is predicted for a location *so;* *N* is the number of measured sample points surrounding the predicted location; λ_i are the weights assigned to each measured point; Z(si) is the known value at the location s_i. The weight of known points against the prediction locations will exponentially decrease as the increase in their distance[12]. The weight is calculated using Equation (2).

$$\lambda_i = \frac{d_{i_o}^{-P}}{\sum_{i=1}^{N} d_{i_o}^{-P}} \quad \textbf{(2)}$$

Where: p *is* the power exponent; *di*0 is the distance between the prediction point s_0 and the known sample data point s_i.

The power exponent *p* controls the influence of the distance among sample points on the interpolation results. Greater *p* values means that the assigned values of the unknown points are more affected by their closest known points. d_{i0} is the distance between the prediction point s_0 and the known sample point s_i. The weight of the standard known point against the predicted point will exponentially decrease with the increase in their distance.

Power Exponent	P=1	P=2	P=3
RMSE	2.12	2.08	2.24

Table 2: Interpolation precision of different power exponents.

4.2 OK Method

OK is an interpolation technique based on the methods of geostatistics. It utilizes the variogram, which does not depend on the actual value of the variable (data), rather its spatial distribution and internal spatial structure [24]. It is a powerful interpolation technique that uses complex mathematical formulas to estimate values for unknown points based on the values obtained at known points.

Ordinary Kriging is similar to IDW in a way 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 [5][13][25]. The calculation equation of the OK method is expressed as Equation (3):

$$\hat{z} = \sum_{i=1}^{n} W_i z_i \qquad (3)$$

4.2.1 Semivariogram Modeling

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 [15][17][18].

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 [17]. According to the geostatistics, as the distance *hi,j* between two points *i* and *j* increases, the correlation between them is expected to decrease (i.e., Cov(zi, zj) ≤ Cov(zi, zk) if hi,j ≤ hi,k) [13][28]. The semivariogram is a plot of semivariance as a function of distances between the observations (h) and is denoted by γ(h) [13].

$$\gamma(h) = \frac{E\left[Z_{(x)} - z_{(x+h)}\right]^2}{2} \qquad (4)$$

Where: γ(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; h is the distance between each pair of locations in the x y coordinate system. Euclidean distance is used to calculate the distance between two locations as shown in Equation (5) [18].

$$h_{i,j} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \qquad (5)$$

After computing autocorrelation of the data and using the spatial information in the data to compute distances, now a prediction will be performed using the data with the fitted model. Weights are calculated using the following simultaneous Equation (6) [18][25].

$$\begin{pmatrix} W1 \\ \vdots \\ Wn \\ 1 \end{pmatrix} = \begin{pmatrix} \gamma(h_{1,1}) & \cdots & \gamma(h_{1,n}) & 1 \\ \vdots & & \vdots & \vdots \\ \gamma(h_{n,1}) & \cdots & \gamma(h_{n,n}) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \gamma(h_{1,u}) \\ \vdots \\ \gamma(h_{n,u}) \\ 1 \end{pmatrix} \qquad (6)$$

4.3 Natural Neighbor

Natural neighbor interpolation is a method of spatial interpolation method of multivariate interpolation in one or more dimensions. The method is based on Voronoi tessellation of a discrete set of spatial points. This has advantages over simpler methods of interpolation, such as nearest-neighbor interpolation, in that it provides a smoother approximation to the underlying "true" function [20][21].

For a given set of points in space, a Voronoi diagram is a decomposition of space into cells, one for each given point, so that anywhere in space, the closest given point is inside the cell.

Thiessen polygons have boundaries that define the area that is closest to each point relative to all other points. This property is mathematically defined by the perpendicular bisectors of the lines between all input data points. In this case, every un-sampled location is thus assigned the value of its nearest measurement point. [5][22].

The basic equation is:

$$G_{(x)} = \sum_{i=1}^{n} \omega_i(x) f(x_i) \qquad \textbf{(7)}$$

Where, G(x) is the estimate at x, w_i are the weights and f (xi) are the known data at (xi). The weights, w_i, are calculated by finding how much of each of the surrounding areas is "stolen" when inserting x into the tessellation [5][22].

Sibson weights is given by:

$$w_{i(x)} = \frac{A(x_i)}{A(x)} \qquad \textbf{(8)}$$

Where, *A(x)* is the volume of the new cell centered in *x*, and $A(x_i)$ is the volume of the intersection between the new cell centered in *x* and the old cell centered in x_i.

Laplace weights is given by:

$$w_{i(x)} = \frac{\frac{l(x_i)}{d(x_i)}}{\sum_{k=1}^{n} \frac{l_{(x_K)}}{d(x_k)}} \qquad \textbf{(9)}$$

Where, $l(x_i)$ is the measure of the interface between the cells linked to *x* and x_i in the Voronoi diagram and $d(x_i)$, the distance between *x* and x_i [5][22].

4.4 Cross Validation

Cross-validation is used to evaluate the accuracy of the three interpolation methods. First, we can assume that the value of a station is unknown and is estimated by the value of its neighboring stations, then calculate the deviation between actual observed values and estimated values. Root-mean-square error (RMSE) are adopted to assess the accuracy of the interpolation

Where *Zi* is the actually observed value of Station I; *Zî* is the estimated value and *n* is the number of stations for testing. RSME indicates the deviation between interpolated values and the actual values.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(Z_i - \hat{Z}_i\right)^2}{n}} \qquad \textbf{(10)}$$

NtoI	NtoI=10			NtoI=15			NtoI=20		
IatL	2	5	10	2	5	10	2	5	10
RMSE	2.19	2.19	2.19	2.093	2.093	2.093	2.14	2.14	2.14

Table 3: *P* = 1 and accuracy comparison between different NtoI and different IatL.

NtoI	NtoI=10			NtoI=15			NtoI=20		
IatL	2	5	10	2	5	10	2	5	10
RMSE	2.17	2.17	2.17	2.081	2.081	2.081	2.106	2.106	2.106

Table 4: *P* = 2 and accuracy comparison between different NtoI and different IatL.

NtoI	NtoI=10			NtoI=15			NtoI=20		
IatL	2	5	10	2	5	10	2	5	10
RMSE	2.146	2.146	2.146	2.148	2.148	2.148	2.158	2.158	2.158

Table 5: *P* = 3 and accuracy comparison between different NtoI and different IatL.

For the three interpolation methods discussed above, the search radius is the most important factor that influences the accuracy as it limits the number of input sample points that are used to calculate each interpolating unit. The two parameters that will influence the search radius are: Neighbors to Include (hereinafter referred to as NtoI) and Include at Least (hereinafter referred to as IatL). Known points far away from the unknown points are meaningless for prediction, so NtoI determines the number of interpolation known points to include for analysis, whereas IatL is the minimum known points to include and should be smaller than NtoI.

Model	Nugget	Partial Sill	Sill	Range	RMSE
Gaussian	4.162	5.384	9.546	0.02	0.798
Spherical	1.972	7.279	9.251	0.019	0.851
Exponential	1.015	9.304	10.319	0.028	0.871

Table 5: OK interpolation errors of the three functions

5. Results and discussion

The interpolation of spectrum occupancy at unknown points was estimated using different techniques available in ArcGIS software. The interpolation is carried out using IDW, OK and NN. The performance analysis of these interpolation techniques were performed based on the Root Mean Square Error (RMSE) produced by each interpolation algorithms.

Parameter values	Interpolation	RMSE
P=2, NtoI=9 and IatL=4	IDW	2.801
Gaussian model, NtoI=11 and IatL=4	OK	0.807
NtoI=9 and IatL=4	NN	2.119

Table 6: Statistical errors in the process of occupancy interpolation.

From the result, it can be seen that IDW and NN produces the largest error of 2.801 and 2.119 respectively. The Geo-statistical Kriging method produces less RMSE value of 0.807. The interpolated surface produced by the Kriging method is smooth and realistic compared to the surface created by IDW and NN. So based on the RMSE value produced and the interpolated surface created, it can be concluded that Kriging is the best suitable approach to develop smooth spectrum occupancy map for the study area.

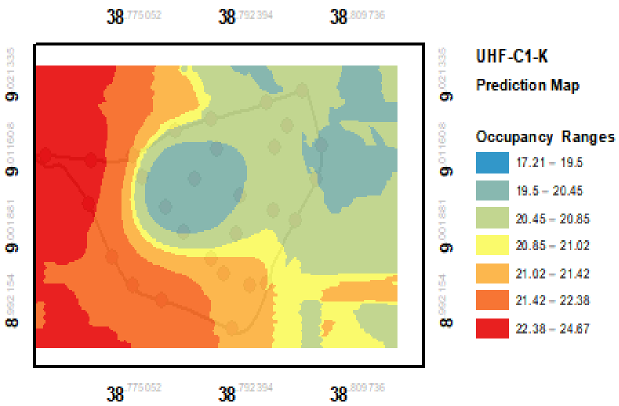


Figure 2: Contour map for UHF band

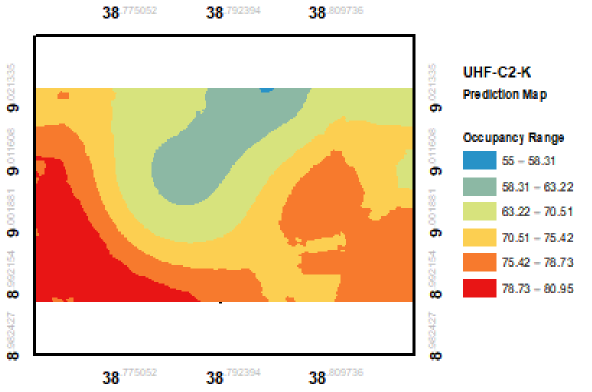


Figure 3: Contour map for UHF band

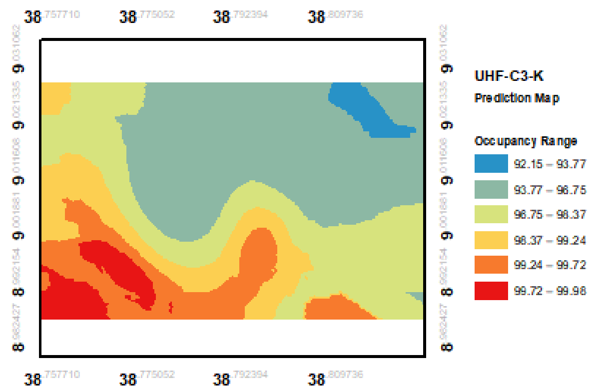


Figure 4: Contour map for UHF band

6. Conclusion

In this thesis work, an optimized and accurate spatial interpolation techniques for spectrum occupancy map development for the case of land mobiles in the range of 137 MHz-174 MHz for VHF bands and 400 MHz-470 MHz for UHF bands is proposed by combining IDW, OK and NN interpolation algorithms. Spectrum occupancy data was collected using TCI 700-series spectrum measurement station from the selected study area in Addis Ababa city.

The result obtained shows that, all the proposed occupancy prediction algorithms in this paper are cable to produce occupancy prediction map.

However, based on the Gaussian model of semivariogram with an optimal number of neighbors OK used is 28.75% more accurate that IDW and 38% more accurate than that of NN with acceptable error of 0.807 RMSE value. OK can predict the missing occupancy data for those locations in which measurement was not taken and it generates accurate predicted information for unmeasured locations than IDW and NN.

REFERENCES

[1] M. Riahi Manesh, S. Subramaniam, H. Reyes, eta N. Kaabouch, «Real-time spectrum occupancy monitoring using a probabilistic model», *Comput. Networks*, libk. 124, or. 87–96, 2017, doi: 10.1016/j.comnet.2017.06.003.

[2] Radiocommunication Sector of ITU, «Spectrum occupancy measurements and evaluation», libk. 2256, or. 53, 2016.

[3] S. Debroy, S. Bhattacharjee, eta M. Chatterjee, «Spectrum Map and Its Application in Resource Management in Cognitive Radio Networks», *IEEE Trans. Cogn. Commun. Netw.*, libk. 1, zenb. 4, or. 406–419, 2015, doi: 10.1109/TCCN.2016.2517001.

[4] S. Grimoud, B. Sayrac, S. Ben Jemaa, eta E. Moulines, «An algorithm for fast REM construction», *Proc. 2011 6th Int. ICST Conf. Cogn. Radio Oriented Wirel. Networks Commun. CROWNCOM 2011*, or. 251–255, 2011, doi: 10.4108/icst.crowncom.2011.245788.

[5] Z. Han, J. Liao, Q. Qi, H. Sun, eta J. Wang, «Radio environment map construction by kriging algorithm based on mobile crowd sensing», *Wirel. Commun. Mob. Comput.*, libk. 2019, 2019, doi: 10.1155/2019/4064201.

[6] S. Debroy, S. Bhattacharjee, eta M. Chatterjee, «Spectrum Map: Toward Predicting the Spatial Distribution of Spectrum Usage in CRNs», zenb. February 2015, 2012, doi: 10.4108/icst.crowncom.2011.245923.

[7] H. B. Yilmaz, C. B. Chae, eta T. Tugcu, «Sensor placement algorithm for radio environment map construction in cognitive radio networks», *IEEE Wirel. Commun. Netw. Conf. WCNC*, zenb. January, or. 2096–2101, 2014, doi: 10.1109/WCNC.2014.6952633.

[8] A. Rabanimotlagh, P. Janakaraj, eta P. Wang, «Optimal crowd-augmented spectrum mapping via an iterative Bayesian decision framework», *Ad Hoc Networks*, libk. 105, 2020, doi: 10.1016/j.adhoc.2020.102163.

[9] P. Lange, «Ethiopian National Frequency Allocation Table».

[10] T. Zhang, N. Leng, eta S. Banerjee, «P17-Zhang», *Acm Mobicom*, or. 17–28, 2014.

[11] R. Li eta J. Li, «A novel clouds based spectrum monitoring approach for future monitoring network», *2014 2nd Int. Conf. Syst. Informatics, ICSAI 2014*, zenb. Icsai, or. 520–524, 2015, doi:

10.1109/ICSAI.2014.7009342.

[12] A. Kumar, S. Maroju, eta A. Bhat, «Application of ArcGIS geostatistical analyst for interpolating environmental data from observations», *Environ. Prog.*, libk. 26, zenb. 3, or. 220–225, 2007, doi: 10.1002/ep.10223.

[13] W. Xu, Y. Zou, G. Zhang, eta M. Linderman, «A comparison among spatial interpolation techniques for daily rainfall data in Sichuan Province, China», *Int. J. Climatol.*, libk. 35, zenb. 10, or. 2898–2907, 2015, doi: 10.1002/joc.4180.

[14] S. B. Babu, «Comparative Study on the Spatial Interpolation Techniques in GIS», *Int. J. Sci. Eng. Res.*, libk. 7, zenb. 2, or. 550–554, 2016, [Sarean]. Available at: https://www.ijser.org/researchpaper/Comparative-Study-on-the-Spatial-Interpolation-Techniques-in-GIS.pdf.

[15] Z. Kassaw, «Coverage Prediction Based on Spatial Interpolation Techniques : The Case of UMTS Network in Addis Ababa , Ethiopia», 2020.

[16] E. Camizuli eta E. J. Carranza, «Exploratory Data Analysis (EDA)», *Encycl. Archaeol. Sci.*, zenb. 3, or. 1–7, 2018, doi: 10.1002/9781119188230.saseas0271.

[17] K. Johnston, J. M. Ver Hoef, K. Krivoruchko, eta N. Lucas, «Using ArcGIS geostatistical analyst», *Analysis*, libk. 300, zenb. January 2004, or. 300, 2001, [Sarean]. Available at: http://direitosminerarios.com/pdf/ESRI - Using ArcGIS Geostatistical Analyst.pdf.

[18] K. Krivoruchko eta N. Lucas, «Using ArcGIS geostatistical analyst», zenb. January 2004, 2014.

[19] J. McCoy, K. Johnston, S. Kopp, B. Borup, J. Willison, eta B. Payne, «Using ArGis Spatial Analyst», *Esri*, or. 238, 2002, [Sarean]. Available at: http://downloads.esri.com/support/documentation/ao_/776Using_Spatial_Analyst.pdf.

[20] R. C. J. W. Sibson, «A brief description of natural neighbor interpolation», 1981, or. 21–36.

[21] «COMPARISON BETWEEN NATURAL NEIGHBOR INTERPOLATION METHOD».

[22] R. Olivier eta C. Hanqiang, «Nearest Neighbor Value Interpolation», *Int. J. Adv. Comput. Sci. Appl.*, libk. 3, zenb. 4, or. 1–6, 2012, doi: 10.14569/ijacsa.2012.030405.

[23] C. Y. Wu, J. Mossa, L. Mao, eta M. Almulla, «Comparison of different spatial interpolation methods for historical hydrographic data of the lowermost Mississippi River», *Ann. GIS*, libk. 25, zenb. 2, or. 133–151, 2019, doi: 10.1080/19475683.2019.1588781.

[24] C. A. Rishikeshan, S. K. Katiyar, eta 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*, or. 666–671, 2014, doi: 10.1109/CICN.2014.148.

[25] A. Konak, P. O. Box, eta T. Road, «A kriging approach to predicting coverage in wireless networks», libk. 3, zenb. 2, 2009.

[26] Y. Xiao *et al.*, «Geostatistical interpolation model selection based on ArcGIS and spatio-temporal variability analysis of groundwater level in piedmont plains, northwest China», *Springerplus*, libk. 5, zenb. 1, 2016, doi: 10.1186/s40064-016-2073-0.

[27] M. Suchański, P. Kaniewski, J. Romanik, E. Golan, eta K. Zubel, «Radio environment maps for military cognitive networks: density of small-scale sensor network vs. map quality», *Eurasip J. Wirel. Commun. Netw.*, libk. 2020, zenb. 1, or. 195–207, 2020, doi: 10.1186/s13638-020-01803-4.

[28] S. Wang, G. H. Huang, Q. G. Lin, Z. Li, H. Zhang, eta Y. R. Fan, «Comparison of interpolation methods for estimating spatial distribution of precipitation in Ontario, Canada», *Int. J. Climatol.*, libk. 34, zenb. 14, or. 3745–3751, 2014, doi: 10.1002/joc.3941.

[29] «Introduction to Geostatistics ardossy».

[30] H. Zhou, *Learn Data Mining Through Excel*. 2020.

[31] L. Morissette eta S. Chartier, «The k-means clustering technique: General considerations and implementation in Mathematica», *Tutor. Quant. Methods Psychol.*, libk. 9, zenb. 1, or. 15–24, 2013, doi: 10.20982/tqmp.09.1.p015.