

**TOWARDS CLOUD-BASED CROWD-AUGMENTED SPECTRUM MAPPING  
FOR DYNAMIC SPECTRUM ACCESS**

**A Thesis by**

Prabhu Janakaraaj

Bachelor of Electronics and Communications Engineering, Anna University, 2011

Submitted to the Department of Electrical Engineering and Computer Science  
and the faculty of the Graduate School of  
Wichita State University  
in partial fulfillment of  
the requirements of the degree of  
Master of Science

December 2015

Copyright 2015 by Prabhu Janakaraj  
All Rights Reserved

**TOWARDS CLOUD-BASED CROWD-AUGMENTED SPECTRUM MAPPING  
FOR DYNAMIC SPECTRUM ACCESS**

The following faculty members have examined the final copy of this thesis for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Master of Science with a major in Computer Networking.

---

Pu Wang, Committee Chair

---

Zheng Chen, Committee Member

---

Jibo He, Committee Member

## ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Pu Wang, who made it possible for me to complete this thesis. His support and knowledge have guided me from the beginning to the end. I would also like to thank Dr. Zheng Chen and Dr. Jibo He for their kind help and for serving on my thesis committee.

## ABSTRACT

In recent years, applications using wireless technology has dramatically increased. This has led to the challenge of improving the spatio-temporal spectrum utilization. Dynamic spectrum access is a method of opportunistically accessing the spectrum. Identifying the unused spectrum is one of the key challenge for dynamic spectrum access. Regulatory bodies recommend the usage of centralized database to maintain the unused spectrum at various geographic locations. These databases are constructed with the help of propagation models. Propagation models are not accurate enough for efficient spectrum utilization. To encounter this problem, we propose a cloud-based and crowd-augmented spectrum mapping database by combining the manual spectrum measurements from various mobile devices. Then, we show how optimal sampling can help in achieving high resolution, high accurate spectrum mapping in minimal cost.

# TABLE OF CONTENTS

Chapter	Page
<b>1 Background</b>	<b>1</b>
1.1 Research Objective and Solution . . . . .	2
<b>2 TVWS Spectrum Modelling</b>	<b>4</b>
2.1 FCC Requirements . . . . .	4
2.2 Spectrum Sensing . . . . .	5
2.3 Pilot Signal Detection . . . . .	6
2.4 Mobile Spectrum Sensor . . . . .	7
<b>3 Cloud Augmented Spectrum Mapping</b>	<b>11</b>
3.1 Our System Model . . . . .	11
3.2 Variogram . . . . .	12
3.3 Semivariogram . . . . .	13
3.4 Variogram Fitting . . . . .	14
3.5 Interpolation . . . . .	16
3.6 Spatial Interpolation . . . . .	16
3.7 Local Interpolators . . . . .	17
3.8 Krigging . . . . .	17
3.9 Optimal sampling using spatially simulated annealing . . . . .	19
<b>REFERENCES</b>	<b>22</b>

# CHAPTER 1

## INTRODUCTION

### 1 Background

Growth in wireless communication technologies has led to rapid development of wireless devices which are also used in our daily life. Increase in wireless devices demands for more radio spectrum, which is a finite resource. Radio spectrum is categorized as licensed and unlicensed spectrum. Licensed spectrum meaning that the operator must obtain a license for operating the wireless device in a particular band of frequency whereas the unlicensed spectrum is designated as license exempt allowing the user to operate without any license from the regulatory agencies.

Numerous wireless devices are designed to operate in unlicensed band, so it is not suitable for a reliable communication which are delay sensitive. Meanwhile, most of the radio spectrum are given out for licenses in commercial use and it becomes a complicated process to assign the spectrum bands for new services or technologies. However, it has been identified that most of the licensed spectrum are underutilized or its utilization varies in time. Regulatory bodies adopted a new spectrum utilization scheme called Dynamic Spectrum Access, which allows the unlicensed users (*secondary users*) to occupy the locally unoccupied licensed spectrum without causing interference to the licensed users (*primary users*). Several studies have been done to identify the unused or under utilized licensed spectrum and found that the large portion of unused spectrum is available in TV band because of the move from Analog television broadcasting system to Digital Television broadcasting system. Study also showed that the spectrum availability in a location varies both in spatial and temporal, which is one of the promising feature for dynamic spectrum access.

Secondary devices should be environment aware, meaning that it should know the information priori about the spectrum occupancy in that location to prevent harmful interference to , the primary users. Two approaches exist for the spectrum awareness (1) Self Identification (2) Geo-location Database Approach. In Self Identification approach, the device itself identifies the spectrum occupancy on the surrounding environment and in Geo-location database approach the secondary device contacts the database to get the information on spectrum availability. Self Identification approach is expensive in terms of computation and energy consumption, since it involves performing signal processing operations.

Geo-location database must be approved by FCC before it is allowed for secondary device access. The database is constructed with the help of F-curve model defined by FCC, that uses TV transmitter location and characteristics to estimate the received signal strength at a particular location. These models are adapted for modeling the behavior of primary users under the conditions that antennas always radiate at maximum transmitting power and placed at fixed height. So it tend to be inaccurate on urban setting and also in places where there exist changes in environment. In urban areas database inaccuracy leads to loss of efficient spectrum usage missing large number of TVWS.

To overcome this issues, numerous research works are being done, in [3] the reliability of spectrum database has been verified by comparing the measurement augmented data constructed using spectrum measurements carried over by vehicle and public systems equipped with spectrum sensing devices. However, limitations of the movement of vehicles only to certain locations hinders the combined database performance. Also, the cost for the measurement will be relatively high to achieve high spatial resolution for covering the entire metropolitan area. In [8] measurement based statistical interpolation technique was studied for predicting the TV coverage estimation and the resulting comparison with database based models showed that over estimation of the database base models wastes large portion of the TV white space. However, the use of interpolation technique, which highly relies on correlation effect for estimation of TV coverage under perform when there is inadequate data to identify the correlation effect. The cost for obtaining enough data will be relatively more if aggressive data gathering is performed with multiple sensors placed in a various fixed locations.

## 1.1 Research Objective and Solution

Accuracy of the database can be improved by combining the model based data along with the empirical measurements. It is not possible to have the actual measurement at each and every location due to cost, geographic and access limitations. Taking the aforementioned issue as the key objective of our research, we show that the use of cloud based crowd-augmented spectrum mapping combined with geo-statistical techniques, can reduce the measurement cost by providing optimal sampling locations necessary to build the efficient spectrum database for secondary devices access.

### **Cloud Based Crowd Augmented Spectrum Mapping**

Cloud is evolving as the viable solution for scaling and high performance platform for computing and communication needs. In our research, we adopt such platform for spectrum sensing and spectrum mapping. Our spectrum sensing platform comprises of various mobile nodes which peri-



odically senses the surrounding TV channels and sends the spatial information such as the location coordinates, channel and PSD information to the computing cloud for mapping. Using spatial information, computing cloud will perform geo-statistical operations to identify the correlation of the data and does the interpolation for predicting PSD of unknown locations. Using this information a local spectrum database is continuously updated for serving the dynamic spectrum access users.

### **Variogram**

The variogram characterizes spatial continuity or roughness of a received power spectral density (PSD) over the location of interest. Variogram investigation comprises of the test variogram figured from the information and the variogram model fitted to the information. An experimental variogram describes the spatial autocorrelation structure of information, and it's accustomed represent the behaviour of a variable as a perform of distance. [5]. There are various variogram models and it is selected based on minimizing the error. We then ensure the estimation variance is positive and well-defined for all lag cases by fitting the experimental variogram to the predicted variogram model.

### **Interpolation**

Krigging is a interpolation technique used in general interpolation problem. It relies on the assumption that spectrum occupancy is a stationary random process [1]. The basic plan consists of generating best linear unbiased predictor of the unobserved values of the random field. The predictor is perfect as a result of it satisfies certain optimality condition which contains the interpolation error and it minimizes the Mean Square Error (MSE) [1]. Though there are number of interpolation techniques, we only consider Ordinary Krigging interpolant which consider that the mean  $\mu$  is constant unknown value.

### **Organization of Thesis**

The reminder of this thesis is structured as follows: Chapter 2 presents the TV White space modeling and mobile spectrum measurement device designed for our research. Then we detail the TV transmission detection methods complying to the FCC regulations. Chapter 3 details cloud based crowd augmented spectrum mapping with its system model, prediction of the unknown values using krigging interpolation technique from the known PSD's and optimal sampling location estimation using Spatially Simulated Annealing. Then we show the results using the heatmap and compares the result with the database administrators data. Chapter 4 concludes the thesis.

## CHAPTER 2

### TV WHITE SPACE MODELLING

## 2 TVWS Spectrum Modelling

Dynamic spectrum was proposed as the key idea for the increasing spectrum scarcity problem. Federal communication commission (FCC) and Electronic Communication Commission (ECC) has identified TV white space (TVWS) as the viable frequency band for allowing dynamic spectrum access. In accordance to the guidelines from FCC and ECC, dynamic spectrum users must obtain information about locally available spectrum from white space database administrators such as Google, Microsoft and Spectrum Bridge or it can be identified through self-sensing by following the sensing requirements defined by the regulators. A number of sensing techniques were studied in the IEEE 802.22 working group and summary of the techniques is provided in [include 802.22 doc]. In this chapter, we implemented pilot signal based spectrum sensing using mobile device with geo-location information. Using the sensing data we then constructed a local spectrum database.

### 2.1 FCC Requirements

There are three systems that must be protected from transmission of TV White Space Devices (1) ATSC Digital Television (2) NTSC Analog Television (3) Wireless Microphones. TV Band devices are classified into two types namely Fixed TV Band Device and Mobile TV Band device. Fixed TV Band device should operate by communicating its geo-location information to the white space spectrum database authorized by FCC, querying for the list of available channels for communication. Mobile TV Band devices must have the capability to sense the incumbent signals at the threshold of -114 dBm on an average over a 6 Mhz bandwidth for ATSC digital TV signals. Similarly for NTSC analog TV signals the devices must have the capability to sense the signals at the threshold of -114 dBm on an average over 100 khz bandwidth. A TV Band device can start operating on a TV channel if there is no signal detected at the specified threshold for 30 seconds. And also a continuous spectrum monitoring system must be employed on the TVBD to check for the presence of primary transmission at least once in a minute. If any transmission is detected, then the TVBD should suspend its operation within 2 seconds in order to avoid causing harmful interference to the primary user. Our research primarily focus on identifying the white space channels available for Mobile TV Band Devices.

## 2.2 Spectrum Sensing

There are different types of spectrum sensing architectures such as Centralized Spectrum Sensing, Distributed Spectrum Sensing and Mobile Spectrum Sensing.

(a) *Centralized Spectrum Sensing* : Centralized spectrum sensing scheme will have one spectrum sensor fixed at a particular location. This method is vulnerable to hidden terminal issues.

(b) *Distributed Spectrum Sensing*: Distributed spectrum sensing scheme utilizes multiple sensors spread across the area needs to be analyzed and this eliminates the vulnerability of hidden terminal as opposed to the previous method. But this method is expensive due to the number of sensor requirements.

(c) *Mobile Spectrum Sensing*: Mobile spectrum sensing composes of one or more spectrum sensors carried to multiple locations in vehicles or cart depending on the area to be covered. The spectrum sensing location can be identified with the help of Internal GPS on the spectrum sensor for later data analysis.

The above mentioned spectrum sensing architectures depends on the spectrum sensing schemes. The two popular spectrum sensing schemes used are Energy Based user detection and Cyclostationary feature detection. The goal of any spectrum sensing scheme is based on the binary hypotheses:

$$y(n) = \begin{cases} w[n] & : H_0 \\ hs[n] + w[n] & : H_1 \end{cases} \quad \text{for } n = 1, \dots, N \quad (1)$$

where  $y[n]$  is the received complex signal,  $s[n]$  is for transmitted signal from the licensed user,  $w[n]$  is the additive Gaussian white noise (AWGN),  $h$  is the complex gain of an ideal channel and  $n = 1, \dots, N$  is the discrete time of signal observation. Hypothesis  $H_0$  refers to the case that there is no signal transmission detected and  $H_1$  refers to the case that there is a signal transmission from the TV station. Summary of the two spectrum sensing schemes are as follows:

(a) *Energy Based*: This method calculates the power of the received signal in our frequency of interest and compares it with the noise variance. If the received power is greater than than the previously approximated power of noise then the scheme concludes as transmission is detected. On the other hand, if the computed noise power is close the noise level then it is depicted as no

transmission detected. The reliability of energy detection scheme strongly depends on the received power and on the accuracy of the approximated noise variance. It also important to consider the sensing time and the number of samples used for the calculation of the received power to improve the accuracy.

(b) *Cyclostationary feature detection*: Cyclostationary feature detection identifies the presence of the signal by identifying unique set of features on the received signal. Although this method outperforms the energy detection scheme it is not heavily used because of the complexity in implementation and need for prior knowledge of the feature to be detected. The cyclostationary detector analyzes the Cyclic Autocorrelation Function of the received signal  $\alpha(k)$ :

$$R_\alpha(k, \tau) = \sum_{\alpha} R_\alpha^f(\tau) e^{2\phi j \alpha k} \quad (2)$$

where  $\tau$  is the lag related to the autocorrelation function,  $\alpha$  is the cyclic frequency and  $R_\alpha^f \tau$  is given by:

$$R_\alpha^f \tau = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} R_\alpha(k, \tau) e^{-2\phi j \alpha k} \quad (3)$$

Self-sensing of available TV White Space channels in Mobile TV Band devices is highly challenging due to high mobility, limited computational ability and battery constraints. Due to simplicity and less computational requirements our work adopted energy detection scheme in our spectrum sensing module.

TV signals presence is time limited and the time of TV transmission is known prior with the help of regulatory documents. Due to this nature it is not necessary to monitor the spectrum continuously to identify the TV signal transmission.

### 2.3 Pilot Signal Detection

Digital TV broadcasting in united states follow ATSC standard which is the scope of our research. ATSC signals contains various features that can be identified through feature detection scheme. The data stream of ATSC signal consists of a 511-symbol long PN sequence on its data stream for every 24.2 ms. ATSC standard uses the 8-level vestigial sideband modulation and this modulated signal is spread over the entire 6-MHz TV channel uniformly with a carrier pilot signal occupying 310 KHz from the lower edge of the channel frequency. Our spectrum sensor scheme follow the procedure as shown in figure[1] to detect the pilot signal.

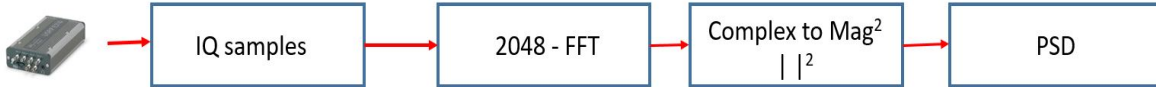


Figure 1: System model for spectrum sensing - Energy Detection

The figure [2] shows the pilot signal detection. It has been identified that pilot carrier contains 7% of the total signal power. The notion behind detecting the pilot signal is, it distinguishes the digital TV signal and inturn making the TV signal easy for detection.

TV signals are narrowband and it is much expensive to monitor the 6 MHz wide signal. So, it is much easier to identify the presence and absence of TV signal by only monitoring the received power of the pilot signal. Also ATSC pilot contributes 0.3 dB of the total signal power thereby making the difference between the measured carrier power level is  $10 * \log_{10}(1 - 10^{0.03}) = -11.45dBm$ . As fact that the pilot measured at -100 dBm will have a signal power of -88.55 dBm.

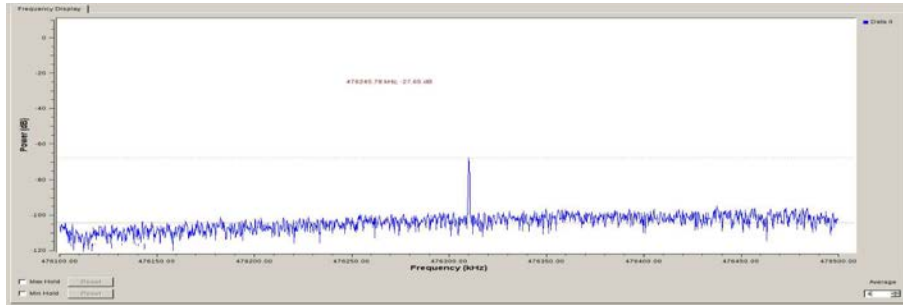


Figure 2: Pilot signal Detection - FFT Display

## 2.4 Mobile Spectrum Sensor

We constructed a mobile spectrum sensor with internal GPS for geo-tagging the location of measurement. Our mobile spectrum sensor consisted of USRP E-310 for computing and as the RF front-end for spectrum measurement. It also incorporated a 2.4 Ghz transceiver for wifi connectivity with on external battery for powering the device.

USRP E-310 is configured with Ubuntu as its base operating system and Gnuradio for RF signal processing. Android based Smart phone is connected to USRP E-310 using USB tethering mode for controlling the device. A python script was created to detect the pilot signal and computes its power spectral density (PSD). Calculated PSD is then saved on to a file with channel and geo-location information.



Figure 3: Mobile Spectrum Sensor, Wi-Fi Access point, Database Server

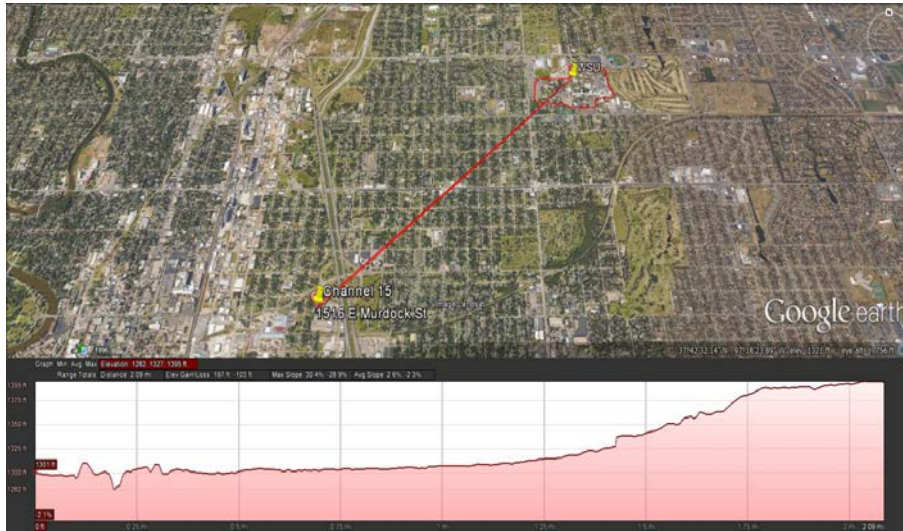


Figure 4: Location of TV transmitter considered for Study

This data then will be uploaded to our spectrum mapping server, when the spectrum sensing device is on close proximity to our research Wi-Fi network located in second floor of Jabara Hall where Electrical Engineering and Computer Science department office is located. Spectrum mapping server uses MySQL database to store the channel information and R statistical programming to interpret the data for producing the local spectrum map.

We conducted a measurement survey around the Wichita State University Main Campus. We choose channel 15 as our target because of the location of TV base station is in our close proximity.

Figure [4] shows the terrain elevation and location of TV transmitter which is operating at Channel 15 which is considered in our study.

The survey covered an area of 1500m over 1 month. The measurement area consists of large buildings and trees. Our measurement campaign followed the two important criteria for the purpose of modeling geo-statistical modeling[2]. (1)No two sample locations are too far apart and also the samples must cover the complete area (2)Samples must be collected at a variety of lags for the sufficient estimation of the relationship between variance and distance. Figure [5] shows the initial measurement locations planned.

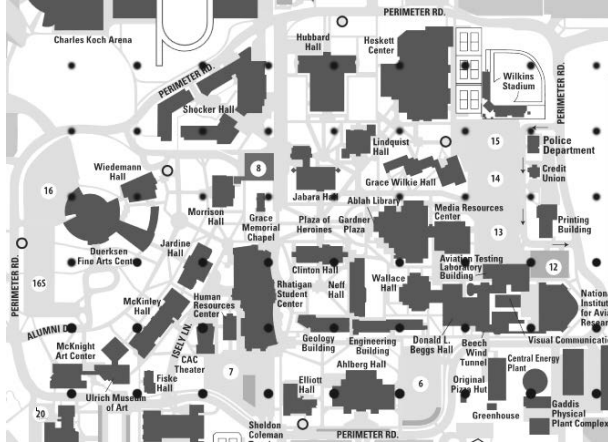


Figure 5: Campus Map with sensing locations in dotted spots - Planned

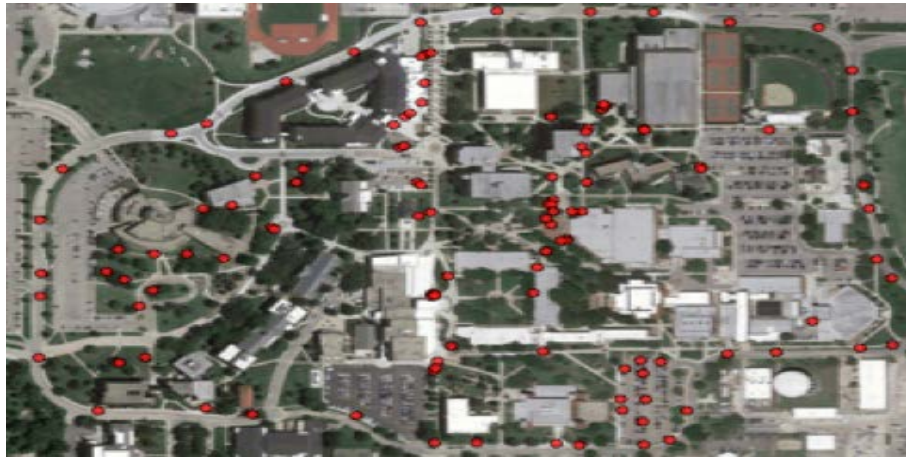


Figure 6: Campus Map with sensing locations in dotted spots - Carried Over

We adopted grid sampling design but due to the accessibility issues we are not able to follow the exact grid points for measurements on our region of interest.

For balancing the time taken to complete the measurement and to maintain the interpolation resolution, we initially planned for a grid spacing of 50 meters between the two measurement locations. Although, grid spacing between each points is kept as 50 meters approximately to balance the time taken for completing the measurement and interpolation resolution. Henceforth, some regions will have dense points. Figure [6] clearly depict the actual measurement locations carried over for this research.

## CHAPTER 3

### Cloud Augmented Spectrum Mapping

### 3 Cloud Augmented Spectrum Mapping

Optimized sampling is needed to reduce the cost of sampling and get adequate samples to obtain the accurate Radio Environment Map for the target area. In this chapter we describe our cloud augmented model for spectrum mapping. Then we show that optimal sampling design scheme outperforms non-optimized sampling design by using Spatially Simulated Annealing Scheme.

Spatial spectrum model is constructed using two approaches (1) Empirical model (2) Deterministic model. Empirical model relies on the measurements and thus reproduces the correlation characteristics of the data set. Number of measurements of measurements locations on the dataset will highly impact the accuracy of the semivariogram estimate. On the other hand, deterministic model relies on the system model of the TV transmitter and the radio environment. White space spectrum databases are constructed using deterministic model as directed by FCC based on F-curve model.

#### 3.1 Our System Model

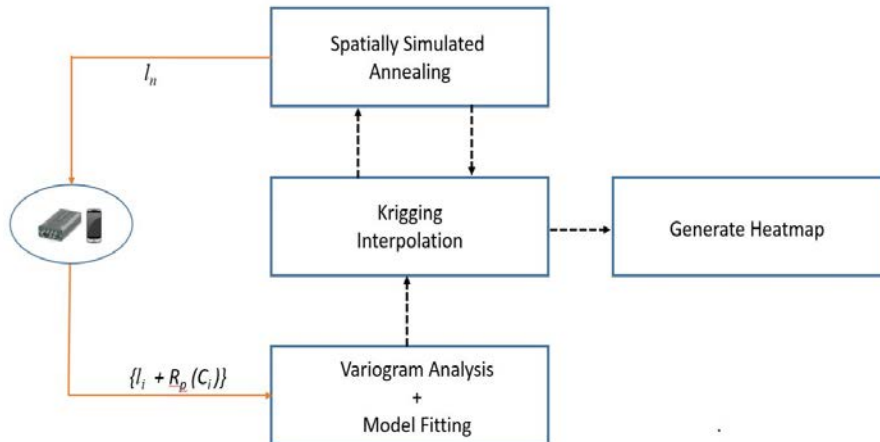


Figure 7: System model - Augmented Spectrum Mapping

Figure [7] shows the system model for augmented spectrum mapping. Our model considers the mobile spectrum sensors are cellular phones equipped with the additional hardware for sensing the incumbent signals in TV band. Every mobile sensor ( $ms_i$ ) will be identified using a unique ID by the



spectrum database server. Mobile sensors will periodically performs spectrum sensing operation as described in section [2.3]. The sensing information will be stored in a local storage on the mobile spectrum sensor. The stored information of the form:

$$\text{Location coordinates : } l_i$$

$$\text{PSD of the channel : } R_p(C_i)$$

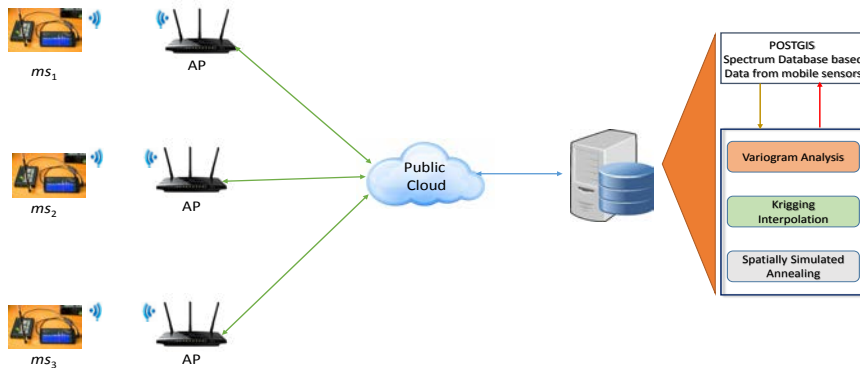


Figure 8: Cloud Based Measurement Augmented Spectrum Mapping

This information will be forwarded to the spectrum mapping server using the general internet infrastructure. Figure [8] shows our proposed cloud based measurement augmented spectrum mapping. Spectrum mapping server hence will receive data from various mobile sensors in the form of:

$$ms_i = \{l_i + R_p(C_i)\}$$

Spectrum mapping server has two roles (1) To store the received spectrum sensing information (2) Compute the spectrum mapping based on the sensing information using geo-statistical approach. We use PostgreSQL with POSTGIS extension for geographical objects as our database platform to store the sensed spectrum information. Then using R statistics tool for computing and graphics [6], spectrum mapping is performed following the geo-statistical approach.

### 3.2 Variogram

Variogram is the core of geostatistics which is a function that models the variance between two points in space as a function of the distance between them. In our work, we considered grid sampling in which the distance between the two measurement points is a fixed lag distance.

The theoretical variogram,  $\gamma$ , is typically written as a function of expected value of the squared distance between a given point value and a point some lag  $h$ :

$$\gamma(h) = \frac{1}{2}E[(Z(x+h) - Z(x))^2] \quad (4)$$

If the field is second order stationary then the covariance function is defined as:

$$C(h) = E[(Z(x) - \mu)(Z(x+h) - \mu)] = C(0) - \gamma(h) \quad (5)$$

However, second order stationary is probably not a safe assumption for the radio environment. So with some set of measurements, an empirical variogram can be defined as the sum of squared differences for each observed lag distance  $h_i$ :

$$\gamma'(h_i) = \frac{1}{2n} \sum_{j=1}^n (z(x_j + h_i) - z(x_j))^2 \quad (6)$$

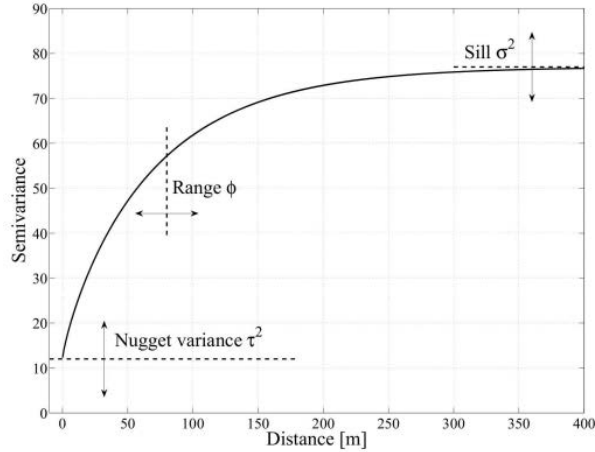


Figure 9: Variogram model parameters

Figure [9] shows the explanation of model parameters of variogram using Matern model as an example.

### 3.3 Semivariogram

If the spatial random field has constant mean  $\mu$ , this is equivalent to the expectation for the squared increment of the values between locations  $x$  and  $y$ :

$$2\gamma(x, y) = E[(Z(x) - Z(y))^2] \quad (7)$$

where  $\gamma(x, y)$  itself is called the semivariogram. For observations  $z_i, i = 1, 2, 3, 4, \dots, k$  at locations  $x_1, x_2, x_3, \dots, x_k$  the empirical variogram

$$\gamma(h) = \frac{1}{|N(h)|} \sum_{(i,j) \in N(h)} |z_i - z_j|^2 \quad (8)$$

where  $N(h)$  denotes the set of pair of observations  $i, j$  such that  $|x_i - x_j| = h$  and  $|N(h)|$  is the number of pairs in the set.

### 3.4 Variogram Fitting

Using some number of measurements to fit a variogram is a typical problem. There are a number of models that can be used for fitting such as circular, spherical, exponential, gaussian, stable, etc. These models influence the prediction of the unknown values, especially when there is a difference in shape of the curve near the origin. Gaussian and Exponential are the two common models formed as:

$$\gamma_{exp}(h) = \tau^2 + \sigma^2(1 - e^{-h/\phi}) \quad (9)$$

Figure [10] and Figure [11] shows the difference in two common models:

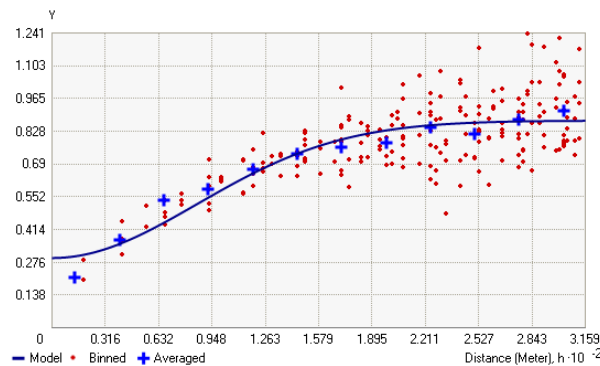


Figure 10: Gaussian Model

The steeper curve near the origin, the more closest neighbors will have on the prediction and this has less smooth surface output.

Figure [6] shows the effect of model parameters on the shape of the fitted variogram.  $\tau^2$  is known as the nugget variance and is used to model discontinuity around the origin by increasing or decreasing the starting threshold. This parameter helps to model the likelihood of the rare things in geological surveys.  $\sigma^2$  sets the maximum value of the semivariogram and is known as sill. High value of  $\sigma$  will increase the level at which the curve flattens out.

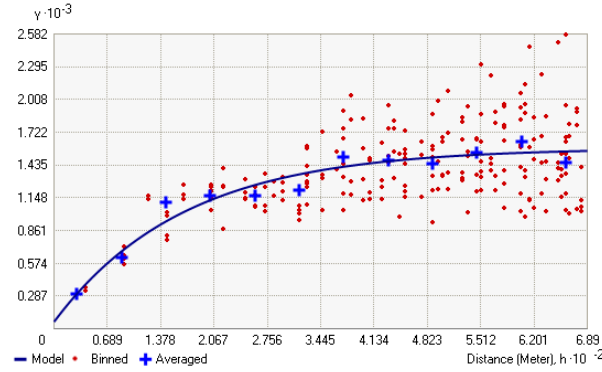


Figure 11: Exponential Model

The parameter  $\phi$  acts as a scale which will have the impact on overall shape of the curve.

Also it determines the rate at which the variance is expected to appear as a function of distance between the points. Variogram fitting can be performed with various methods. Historically, a fitted variogram can be derived using the method of moments. Interestingly, Maximum Likelihood Estimator (MLE) and Weighted Least Squares (WLS) methods are the more commonly used methods due to its substantial success.

Using this approach we constructed our empirical model using variogram and fitted with the variogram obtained from the empirical measurements. WLS determines the parameter set that fits the empirical semivariance.

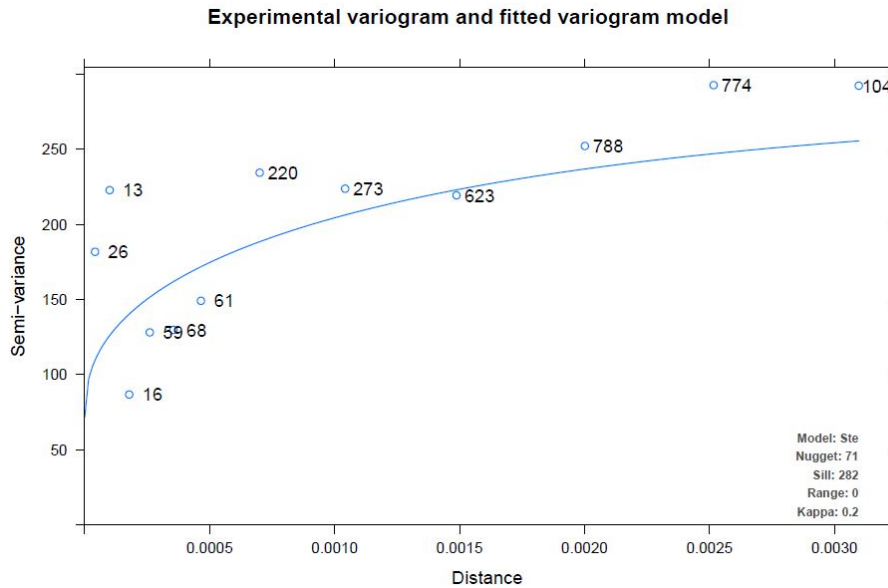


Figure 12: Variogram Fitting for the Measured Data

We considered powered exponential model due to the nature of the data that received signal between two points vary exponentially. Figure [12] shows the semivariogram constructed using the observed PSD values at various location of our target study area.

### 3.5 Interpolation

Interpolation is a technique used to estimate the value of a variable at an unmeasured location from the observed values at surrounding location. This technique is heavily used in Geo-Statistics for various purposes. Estimation is achieved through various interpolation algorithms which estimates the value at a given location as a weighted sum of data values at surrounding locations. Various interpolation methods have been studied in context to RF mapping [7].

### 3.6 Spatial Interpolation

Interpolation methods can be divided into two main types local and global. Local interpolators use only information in the vicinity of the point being estimated and where as the global interpolators use all the available information to provide estimates for the points with unknown values. For analyzing the residuals it is essential to remove the effects of major trends before using local interpolations with the help of global interpolators. Krigging is a particular type of local interpolation with more advanced geostatistical techniques.

Interpolators can also be classified as exact or inexact. With exact interpolation the predicted values at the points for which the data values are known will be known values whereas for inexact interpolation methods removes this constraint meaning that the observed values and the interpolated values at a given point are not necessarily same.

Interpolation methods can be either stochastic or deterministic. Stochastic method provide probabilistic estimates where as deterministic method provides no indication of extent to which there will be errors. It is common to justify the more sample points we have the better and also it is good to have most wide spread sample points on the study area. There are different forms of data points placement or collection such as regular, random, cluster, transect, contour and stratified random.

(a) *Regular*: It guarantees that the sample points are spread but it can result in biases if the attribute to be mapped has continuous fluctuations.

(b) *Random*: The samples points locations are selected using random number but its not guaranteed that the points will be located covering the whole area under analysis leaving behind some uncovered places.

(c) *Stratified Random*: In this the sample points are regular spaced lattice which allocates the points

randomly. But these forms are not designed to minimize the prediction error.

### 3.7 Local Interpolators

Local interpolators attempt to estimate the values for unknown points using the known values for neighbor points. The idea is to identify the lattice of data points for which the values are to be estimated. For every single point the search area is defined around it and the sample points are identified within that area. After that a mathematical function is used to model the variation between the data points and the value for the data point is estimated using the function. The resulting output may vary by the factors like the search area size and the applied mathematical function.

*Weighted Moving Average:* Weighted Moving Average is one of the local interpolation technique which is local, exact and deterministic in nature. It is given by:

$$z'(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (10)$$

where  $z(x_i)$  are the data values for  $n$  points  $(x_1, x_2, \dots, x_n)$  within the radius of the search area and  $\lambda_i$  are the weights that are applied to the data values at each point. The weights are usually some function of the distance between the point for which the estimate is being made and the sample points. Inverse Distance Weighting (IDW) is the commonly applied function. In this case the about equation will become:

$$z'(x_0) = \sum_{i=1}^n \frac{d_{ij}^{-r}}{\sum_{i=1}^n d_{ij}^{-r}} . z(x_i) \quad (11)$$

where  $j$  is the point whose value being interpolated and  $d_{ij}$  is the distance between the sample point  $i$  and  $j$ . Arbitrary value is identified by  $r$  which can be selected by the investigator. If  $r$  is set equal to 1, then it becomes a simple linear interpolator. The value  $r$  is frequently set to 2 where the influence of each sample point is in proportion to the square root of its distance from the point to be interpolated. The problem with this technique is that data points tend to produce 'duck-egg' patters. This issue can be solved by increasing the search radius having undesirable effect of over-smoothing other parts of the surface.

### 3.8 Kriging

Kriging is a form of local interpolation developed by french geostatistician called Georges Matheron and a African miner called D.G Krigie. Kriging is one of the methods which is local, exact

and stochastic. Kriging process begins with recognition that spatial variation of any continuous attribute is too complex to be modeled by a simple and smooth mathematical function. Hence it is modeled as a random field. It is assumed that the spatial variation of any random variable can be expressed as the sum of following components:

- Constant mean or trend for a structural component
- Random and spatially correlated component called as a regionalized variable
- Spatially uncorrelated residual component

The random variable  $Z$  at point  $x$  is:

$$Z(x) = m(x) + \epsilon'(x) + \epsilon'' \quad (12)$$

where  $m(x)$  is structural function indicating the structural component,  $\epsilon'(x)$  is stochastic but spatially autocorrelated residual from  $m(x)$  and  $\epsilon''$  is the random noise having a normal distribution with a mean of 0 and variance  $\sigma^2$ .

”Ordinary” Kriging is a technique for interpolation which predicts the unknown value at a new location ( $Z(x')$ ) from the weighted known values at neighboring locations ( $x_i$ ):

$$Z_K(x') = \sum_{i=0}^n w_i Z(x_i) \quad (13)$$

and to determine the optimal weights ( $w$ ), we must minimize the estimation variance  $\sigma_E^2$ :

$$\sigma_E^2 = E[(Z_k(x') - Z(x'))^2] \quad (14)$$

with

$$\sigma_E^2 = -\gamma(x' - x') - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma(x_i - x_j) + 2 \sum_{i=1}^n w_i \gamma(x_i - x') \quad (15)$$

which leads to the following equation:

$$\begin{pmatrix} \gamma(x_1 - x_1) & \cdots & \gamma(x_i - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{pmatrix} \quad (16)$$

where  $\mu$  is called the Lagrange parameter. This interpolation is “exact”, meaning that  $Z_k(x') = Z(x)$  if  $x = x'$ . Using this approach received PSD at unknown location is calculated for radio environment mapping. Although it is quite efficient to do krigging at the regular grid, our main challenge was to estimate the PSD at an irregular grid which is the case of our study area.

In order to do krigging on the observed data, we need to specify the grid resolution and border of our study area as the function. The border of the study area is obtained using Google Earth KML export facility and the KML is converted to a Shape file which then imported into R for performing krigging operation. The grid resolution can be varied considering interpolation resolution.

Krigging predicts the PSD at unknown locations assuming that there is no noise on the input data. Figure [13] shows the Krigging Prediction and Standard error of the interpolated data.

### 3.9 Optimal sampling using spatially simulated annealing

Krigging process heavily depends on the measured values to predict the values at known locations. Sampling pattern or locations where needs to be measured is a key issue for the producing efficient interpolation results. In order to solve this issue, we use Spatially Simulated Annealing (SSA) which is an eminent method to solve optimization problems in geo-sciences and also we show that it improves the result of krigging interpolation

SSA uses previous sampling designs and random search technique to solve spatial optimization problems. Measurement locations are repeatedly moved around and optimized by minimizing the mean universal krigging variance (MUKV).

The SSA algorithm considers a set of spatial sampling schemes consisting of  $N$  observations ( $E^n$ ), with an objective function  $\phi(\cdot)E^n \rightarrow \mathfrak{R}^+$  which should be minimized. Optimization process begins with random sampling scheme  $E_0 \in E^n$ . Further, intermediate scheme ( $E_{i+1}$ ) is obtained from the random perturbation sequence  $E_0$  with the motion of random observation  $E_0$ , with a randomly chosen vector  $h_x$ . The transition probability  $P(E_i \rightarrow E_{i+1})$ , that  $E_{i+1}$  is accepted, is determined by the metropolis criterion:

$$P\{E_i \rightarrow E_{i+1}\} = \begin{cases} 1, & \text{if } \phi(E_{i+1}) \leq \phi(E_i) \\ e^{-\frac{\phi(E_i) - \phi(E_{i+1})}{t}}, & \text{if } \phi(E_{i+1}) > \phi(E_i) \end{cases} \quad (17)$$

where  $t$  is a positive control parameter, which decreases with the optimization progress. If  $E_{i+1}$  is accepted, it will be fixed as the starting point for the next scheme  $E_{i+2}$ ; and the process continues until the disturbances in the samples no longer significantly reduce the objective function  $\phi(\cdot)$



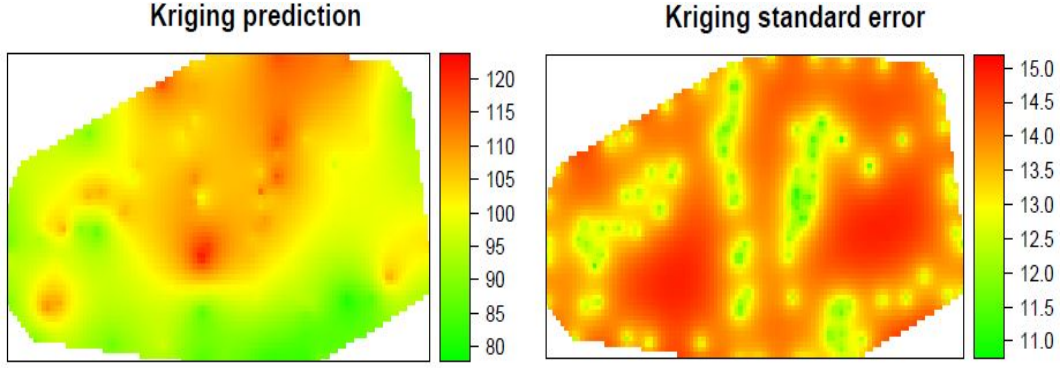


Figure 13: Kriging Prediction and Standard Error

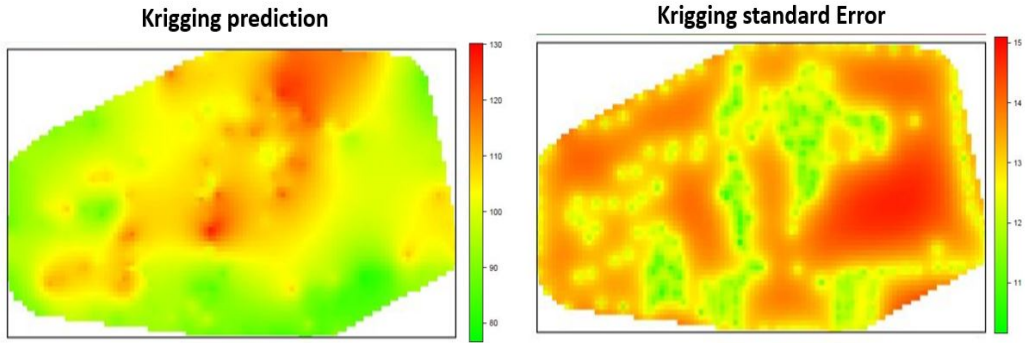


Figure 14: Kriging Prediction and Standard Error - SSA Optimized

With this process the associated sampling scheme is safely stabilized in an optimal configuration for global sampling. SSA benefits from certain criteria of optimization, including minimization of kriging variance. The objective or goal of this criteria is to minimize the kriging variance for the complete study area followed by

$$\phi_{KO} = \frac{1}{A} \int \sigma_{KO}^2(x_0|E).dx_0 \quad (18)$$

where  $E$  is the sampling scheme;  $A$  is the area size;  $x_0$  the vector location to be estimated; and  $\sigma_{KO}^2$  is the kriging variance.

In our study we applied SSA to find out the locations we need to do sampling to produce the optimized kriging results. This is achieved by passing the interpolation values, variogram fit model and target area as the input parameters for the SSA function. As a result SSA functions tells the geo-coordinates where it needs to be measurement for interpolation.

Figure [14] shows the resulting Kriging Prediction and Standard error after applying SSA optimization.

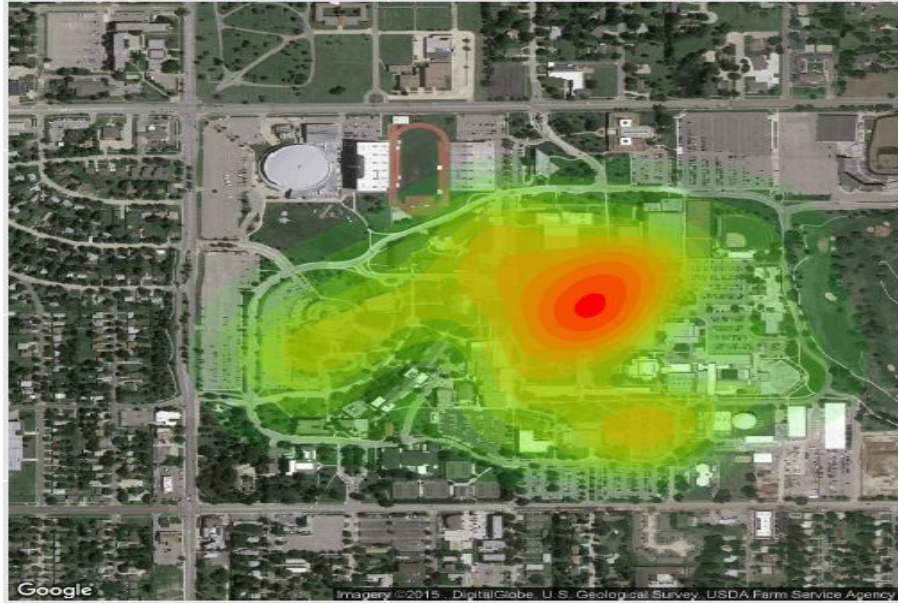


Figure 15: Heat map of channel 15 on our Campus

All Geo-coordinates provided by the SSA functions cannot be used as additional measurement locations because it doesn't take into account of the geographical access limitations since our target areas had many fields which includes buildings, play ground areas and some inaccessible places. Hence nearby coordinate location measurement were used. Figure [15] shows the resulting heatmap of channel 15 in our campus project using google maps [4].

## CHAPTER 4

### CONCLUSION

In this thesis, we quantified the availability of the white space channels in our campus using crowd sourcing methodology. Using one of the quantified channel PSD value at various locations, we analyzed the correlation of the received PSD using our cloud based spectrum mapping model . After the correlation analysis, we used statistical analytics techniques of geo-science to solve the problem of predicting the received PSD values at unknown location based on krigging interpolation and SSA for identifying the measurement locations.

In our future study, we would like to explore more advanced statistical analysis techniques which not only considers noise less data but also includes terrain and noisy data as the source of interpolation. With this information we would like to build a combined database incorporating the empirical measurement and the model based information provided by one of the White Space Database providers.

## REFERENCES

## REFERENCES

- [1] M. Angjelicinoski, V. Atanasovski, and L. Gavrilovska. Comparative analysis of spatial interpolation methods for creating radio environment maps. In *Telecommunications Forum (TELFOR), 2011 19th*, pages 334–337, Nov 2011. doi: 10.1109/TELFOR.2011.6143557.
- [2] R. Balamurthi, H. Joshi, Cong Nguyen, A.K. Sadek, S.J. Shellhammer, and Cong Shen. A tv white space spectrum sensing prototype. In *New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium on*, pages 297–307, May 2011. doi: 10.1109/DYSPAN.2011.5936218.
- [3] Ayon Chakraborty and Samir R. Das. Measurement-augmented spectrum databases for white space spectrum. In *Proceedings of the 10th ACM International on Conference on Emerging Networking Experiments and Technologies, CoNEXT '14*, pages 67–74, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-3279-8. doi: 10.1145/2674005.2675001. URL <http://doi.acm.org/10.1145/2674005.2675001>.
- [4] Google Inc. Google earth, . URL <http://google.com/earth> [cited 10 October 2015].
- [5] J. Ojaniemi, J. Kalliovaara, J. Poikonen, and R. Wichman. A practical method for combining multivariate data in radio environment mapping. In *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, pages 729–733, Sept 2013. doi: 10.1109/PIMRC.2013.6666232.
- [6] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- [7] G. Vanhoy, H. Volos, C.E.C. Bastidas, and T. Bose. A spatial interpolation method for radio frequency maps based on the discrete cosine transform. In *Military Communications Conference, MILCOM 2013 - 2013 IEEE*, pages 1045–1050, Nov 2013. doi: 10.1109/MILCOM.2013.181.
- [8] Xuhang Ying, Chang Wook Kim, and S. Roy. Revisiting tv coverage estimation with measurement-based statistical interpolation. In *Communication Systems and Networks (COMSNETS), 2015 7th International Conference on*, pages 1–8, Jan 2015. doi: 10.1109/COMSNETS.2015.7098692.