

# Durham E-Theses

---

## *Measurement and modelling of spectrum occupancy*

Zhe Wang

### How to cite:

---

Wang, Zhe (2009) Measurement and modelling of spectrum occupancy. Doctoral thesis, Durham University.

### Use policy

---

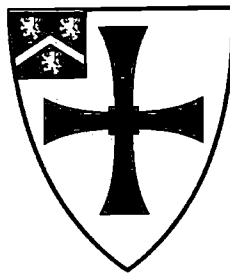
The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a <https://etheses.durham.ac.uk/id/eprint/2036/> is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

# Measurement and Modelling of Spectrum Occupancy



The copyright of this thesis rests with the author or the university to which it was submitted. No quotation from it, or information derived from it may be published without the prior written consent of the author or university, and any information derived from it should be acknowledged.

Zhe Wang

Ustinov College

University of Durham

A thesis submitted for the degree of

*Doctor of Philosophy*

2009

12 OCT 2009



To my parents and my aunt

## Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Sana Salous, for giving valuable guidance and suggestions for improving the work in this thesis. I would also like to record my gratitude to Stuart Feeney for sharing his RF knowledge, Roger Lewenz and Peter Baxendale for their contributions in the development of the data acquisition software.

I am indebted to Durham University and British Telecom for their funding of my studies.

## Abstract

Based on the conception of spectrum sharing, cognitive Radio as a promising technology for optimizing utilization of the radio spectrum has emerged to revolutionize the next generation wireless communications industry. In order to adopt this technology, the current spectrum allocation strategy has to be reformed and the real spectrum occupancy information has to be systemically investigated. To assess the feasibility of cognitive radio technology, the statistical information of the present spectral occupancy needs to be examined thoroughly, which forms the basis of the spectrum occupancy project.

We studied the 100–2500 MHz spectrum with the traditional radio monitoring systems whose technical details have been fully recorded in this thesis. In order to detect the frequency agile signals, a channel sounder, which is capable of scanning 300 MHz spectrum within 4 ms with multiple channel inputs, was used as a dedicated radio receiver in our measurements. The conclusion of the statistical information from the spectrum monitoring experiments shows that the spectrum occupancy range from 100–2500 MHz are low indeed in the measuring locations and period. The average occupancies for most bands are less than 20%. Especially, the average occupancies in the 1 GHz to 2.5 GHz spectrum are less than 5%.

Time series analysis was initially introduced in spectrum occupancy analysis as a tool to model spectrum occupancy variations with time. For instance, the time series Airline model fits well the GSM band occupancy data. In this thesis, generalized linear models were used as complementarity solutions to model occupancy data into other parameters such as signal amplitude. The validation of the direction of arrival algorithms (EM and SAGE) was verified with the anechoic chamber, by which we can determine the spectrum occupancy in space domain.

# Contents

<b>Nomenclature</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Radio Spectrum Occupancy Monitoring</b>	<b>5</b>
2.1 Considerations for Spectrum Monitoring . . . . .	5
2.2 Monitoring System and Site . . . . .	6
2.3 Monitoring Technology . . . . .	10
2.3.1 receiving antenna . . . . .	11
2.3.2 spectrum monitoring equipment . . . . .	13
2.3.2.1 spectrum analyzer . . . . .	13
2.3.2.2 channel sounder . . . . .	18
2.3.3 noise and system sensitivity . . . . .	20
2.3.4 distortion and dynamic range . . . . .	25
2.4 Data Acquisition . . . . .	30
2.4.1 data acquisition for spectrum analyzer . . . . .	30
2.4.2 data acquisition for channel sounder . . . . .	36
2.5 Summary . . . . .	43
<b>3 Parameter Estimations</b>	<b>45</b>
3.1 Direction of Arrival Estimation . . . . .	45
3.1.1 wave and sensor array . . . . .	46
3.1.1.1 wave propagation . . . . .	46
3.1.1.2 uniform linear array . . . . .	48
3.1.2 parametric data model . . . . .	52

## CONTENTS

---

3.2	EM and SAGE Algorithm . . . . .	57
3.2.1	maximum likelihood estimation . . . . .	57
3.2.2	expectation maximization and SAGE algorithms . . . . .	59
3.3	Algorithm Verifications . . . . .	67
3.4	Summary . . . . .	76
<b>4</b>	<b>Statistics and Modelling</b>	<b>77</b>
4.1	Occupancy analysis with descriptive Statistics . . . . .	77
4.2	Time Series Analysis . . . . .	95
4.2.1	model approaches of spectrum occupancy . . . . .	95
4.2.2	stationary time series . . . . .	98
4.2.3	ARMA model identification . . . . .	103
4.2.4	model estimation and diagnostics . . . . .	111
4.3	Generalized Linear Models . . . . .	115
4.3.1	linear models and generalization . . . . .	115
4.3.2	logistic regression model . . . . .	118
4.3.3	model analysis for GSM band occupancy data . . . . .	124
4.4	Summary . . . . .	129
<b>5</b>	<b>Conclusions</b>	<b>131</b>
	<b>References</b>	<b>139</b>

# List of Figures

2.1	Measurement location . . . . .	7
2.2	Monitoring system for 100-1500 MHz spectrum . . . . .	8
2.3	Monitoring system for 1500-2500 MHz spectrum . . . . .	9
2.4	Monitoring system for 2.4 GHz ISM band . . . . .	10
2.5	Return loss for TDJ-0825DSA log periodic antenna . . . . .	12
2.6	Block diagram of a spectrum analyzer [1] . . . . .	14
2.7	Trace point based on detector algorithm [1] . . . . .	17
2.8	Block diagram of the receiver of the channel sounder . . . . .	19
2.9	Linear chirp signal . . . . .	19
2.10	The noise signal with 10 kHz resolution bandwidth . . . . .	22
2.11	Cascade amplifier chain for noise figure calculations . . . . .	25
2.12	Second and third order distortion . . . . .	27
2.13	The insertion loss and return loss of the preselectors . . . . .	28
2.14	Dynamic range versus distortion [1] . . . . .	29
2.15	Data flow chart of the monitoring system . . . . .	31
2.16	The pin outs of the PC parallel port . . . . .	33
2.17	Communications between computer and spectrum analyzer . . . . .	35
2.18	Communications between computer and channel sounder . . . . .	36
2.19	The block diagram of the PCI-DAS4020/12 [14] . . . . .	37
2.20	The relation between SRF and clock . . . . .	39
2.21	Producer-consumer algorithm . . . . .	41
2.22	The flowchart of the data acquisition routine . . . . .	43
3.1	The linear aperture and its aperture smoothing function . . . . .	49
3.2	Linear array . . . . .	50

## LIST OF FIGURES

---

3.3	The grating lobe of the linear array . . . . .	51
3.4	Aliasing . . . . .	52
3.5	Calculation the steering vector of the uniform linear array . . . . .	56
3.6	Maximizing a function with lower bound [40] . . . . .	60
3.7	The relationship of the observed data and hidden-data . . . . .	66
3.8	The spectrum of the simulation signals . . . . .	68
3.9	The absorbers of the TDK anechoic chamber . . . . .	69
3.10	The configuration of DOA experiment . . . . .	70
3.11	The calibration data of the receiver . . . . .	71
3.12	The phase difference caused by cables . . . . .	73
3.13	The received waveforms when Tx 1 transmitted . . . . .	74
3.14	The receiving waveforms when the Tx 2 transmitted . . . . .	74
3.15	The receiving waveforms when both Tx1 and Tx 2 transmitted . . . . .	75
4.1	Box-and-whisker diagram . . . . .	78
4.2	Occupancy statistics of spectrum . . . . .	79
4.3	Occupancy statistics of Air Band . . . . .	81
4.4	Occupancy statistics of Mid Band . . . . .	82
4.5	Occupancy statistics of High Band . . . . .	83
4.6	Occupancy statistics of Band III . . . . .	84
4.7	Occupancy statistics of NATO Band . . . . .	85
4.8	Occupancy statistics of UHF 1 & 2 Band . . . . .	86
4.9	Occupancy statistics of TV, Band IV & V . . . . .	87
4.10	Occupancy statistics of GSM & mobile Band . . . . .	88
4.11	Occupancy statistics of L (lower) Band . . . . .	89
4.12	Occupancy statistics of L (upper) Band . . . . .	90
4.13	Occupancy statistics of S (lower) Band . . . . .	91
4.14	Snapshots of 2.4 GHz ISM band with Sounder . . . . .	92
4.15	Channel occupancy statistics in location 1 . . . . .	93
4.16	Channel occupancy statistics in location 2 . . . . .	93
4.17	Channel occupancy statistics in location 3 . . . . .	94
4.18	The azimuthal distributions of white band . . . . .	94
4.19	Occupancy of a single channel . . . . .	96

## LIST OF FIGURES

---

4.20	Time series of GSM band occupancy . . . . .	99
4.21	The trend and seasonal components of the GSM occupancy series	101
4.22	The difference series $\{\nabla_{24}X_t, t = 24, \dots, 168\}$ . . . . .	102
4.23	The difference series $\{\nabla\nabla_{24}X_t\}$ . . . . .	106
4.24	The <i>autocorrelation</i> function of the GSM band occupancy . . . . .	108
4.25	The <i>autocorrelation</i> function of the difference series $\{\nabla\nabla_{24}X_t\}$ . .	108
4.26	The <i>partial autocorrelation</i> function of the difference series $\{\nabla\nabla_{24}X_t\}$	110
4.27	The fitted series and observed occupancy series . . . . .	113
4.28	The standardized residuals after fitting ARIMA model . . . . .	114
4.29	The autocorrelation function of the residuals . . . . .	114
4.30	The flowchart of the modeling procedure . . . . .	115
4.31	Geometrical representation of estimation $\vec{\beta}$ [18] . . . . .	117
4.32	The <i>logistic transformation</i> . . . . .	120
4.33	The occupancy data by different thresholds . . . . .	124
4.34	The report for logit regression model . . . . .	126
4.35	The ANCOVA report for logit regression model . . . . .	127
4.36	The report for logit regression model . . . . .	128
4.37	Fitted series and observed occupancy series . . . . .	128
4.38	Model check for the logit model . . . . .	129

# Nomenclature

## Roman Symbols

$F$  complex function

$J$  joule

$K$  kelvin

## Greek Symbols

$\chi^2$  Chi-square distribution

$\kappa$  wavenumber

$\lambda$  wavelength

$\pi$   $\approx 3.14\dots$

$\omega$  angular frequency

## Superscripts

$j$  superscript index

$T$  transpose

## Subscripts

$0$  subscript index

## Other Symbols

## LIST OF FIGURES

---

$\hat{\alpha}$	slowness vector
dBc	Decibels relative to the carrier
dB	Decibel
dB <sub>i</sub>	Decibels gain of an antenna relative to the hypothetical isotropic antenna
dBm	Decibels relative to one milliwatt
$\equiv$	equivalent
$\triangleq$	define as
$j$	unit imaginary number $\sqrt{-1}$
$\hat{k}$	wavenumber vector

### Acronyms

*AIC* Akaike's Information Criterion

*AICC* Akaike's Information Corrected Criterion

*ANCOVA* (Analysis Of Covariance)

*ARIMA* Autoregressive Integrated Moving Average model

*SARIMA* Seasonal Autoregressive Integrated Moving Average model

*ARMA* Autoregressive Moving Average model

*CDMA* Code Division Multiple Access

*CR* Cognitive Radio

*DOA* Direction of Arrival

*EM* Expectation-Maximization algorithm

*GPIB* General Purpose Interface Bus

*GPS* Global Navigation Satellite System

## LIST OF FIGURES

---

<i>GSM</i>	Global System for Mobile Communications
<i>IF</i>	Intermediate Frequency
<i>IP</i>	Intermodulation Products
<i>ISM</i>	Industrial, Scientific and Medical
<i>LO</i>	Local Oscillator
<i>MIMO</i>	Multiple-Input and Multiple-Output
<i>MLE</i>	Maximum Likelihood Estimation
<i>OOP</i>	Object-Oriented Programming
<i>RBW</i>	Resolution Bandwidth
<i>RF</i>	Radio Frequency
<i>RFI</i>	Radio Frequency Interference
<i>RSS</i>	Residual Sum of Squares
<i>SAGE</i>	Space-Alternating Generalized Expectation-Maximization algorithm
<i>SHI</i>	Second Harmonic Distortion
<i>SRF</i>	Sweep Repetition Frequency
<i>ST</i>	Sweep Time
<i>THD</i>	Total Harmonic Distortion

# Chapter 1

## Introduction

Electromagnetic radiation is the propagation of energy that travels through space in the form of waves. It includes the visible light spectrum, as well as infrared, ultraviolet and x-rays. While the electromagnetic spectrum is theoretically boundless, the radio frequency spectrum which is the portion of the electromagnetic spectrum that holds for communications and other applications, while substantial, is finite. In practice, the properties of radio wave propagation and electronic equipment limit radio communications to frequencies between 9 kHz and 30 GHz. These properties also constrain particular types of communication systems to certain portions of the allocated spectrum, limiting the spectrum available for specific uses. Table 1.1 depicts some of the many uses of radio spectrum associated with various bands [44].

Two important characteristics of the spectrum are the propagation features and the amount of information which signals can carry. In general, signals sent using the higher frequencies have smaller propagation distances but a higher data carrying capacity. These propagation characteristics of the spectrum constrain the identified range of applications for which any particular band is suitable. A portion of spectrum range from 30 MHz to 3000 MHz is known to be suitable for a wide variety of services and is thus in great demand, which became the main investigation in our project.

Historically, in order to prevent interference among users of adjacent frequencies use of the radio spectrum has been highly regulated for reasons of defence

Name	Frequency	Some uses
Extremely low frequency	3 ~ 30 Hz	
Super low frequency	30 ~ 300 Hz	
Ultra low frequency	300 ~ 3000 Hz	
Very low frequency	3 ~ 30 kHz	Long range navigation
Low frequency	30 ~ 300 kHz	Aeronautical navigation
Medium frequency	300 ~ 3000 kHz	AM radio
High frequency	3 ~ 30 MHz	Global broadcast
Very high frequency	30 ~ 300 MHz	TV, Broadcast, Mobile
Ultra high frequency	300 ~ 3000 MHz	TV, satellite
Super high frequency	3 ~ 30 GHz	Satellite TV
Extremely high frequency	30 ~ 300 GHz	

Table 1.1: Radio frequency bands

and security. However, in some cases, this regulatory practices delayed the introduction of new technologies and growth of a variety of beneficial services, or artificially increased the cost [26]. During that same period of time, demands made on the use of spectrum itself have grown significantly. These two highlight the need for efficient use of all available spectrum in order to avoid scarcity and have brought policy-makers and regulators worldwide to focus on new spectrum regulations, which should allow improvements in cost, services and technologies to spread more readily to consumers and public services.

One radio communication system is more spectral efficient than another if it delivers the desired information using less of the spectrum resource. In this sense, spectrum is used inefficiently when systems are not packed together as tightly as possible in frequency bands or when excessive guard bands are used, or when portions of frequency bands are unused while other bands with similar physical characteristics are congested. The allocation of frequency bands, the development of channeling plans, and the assignment of frequencies to specific systems all affect spectrum efficiency.

In order to promote spectrum efficiency, two level efforts can be made for this objective. One is in the spectrum management level and another is in the

---

technique level. The general role of spectrum management is to ensure that no two users transmit on the same frequency at the same time or so close together that they interfere with each other. Management decisions can then be based on the relative spectrum efficiency of the various technologies and techniques. Spectrum measurements are critical to spectrum management in the development of new spectrum access technologies. Specifically, spectrum occupancy studies identify which spectrum bands have low or no active utilization and thus may be appropriate for spectrum sharing. They provide information on the signal characteristics within these bands, which is needed to design spectrum sharing algorithms.

The objective of this spectrum occupancy project was to perform a band by band measurement of the spectrum occupancy in time, frequency and space domains in the spectral region between 100 MHz to 2500 MHz [57]. The spectrum occupancy studies would identify the current spectral bands that have low or no active utilization, the seasonal variations in this usage and begin to identify the long term trends in the use of the various spectral bands. From these and future studies, candidate and non-candidate bands for spectrum reframing and sharing can be readily identified. These studies also provide information on the unique signal characteristics associated with each band and which are needed to properly design spectrum sharing algorithms. The main considerations of the spectrum occupancy project are listed as follow:

- enable a better understanding of how telecommunication systems use the allocated spectrum.
- provide timely information on variations in frequency-band usage, e.g., identify frequency bands becoming heavily used.
- assess the feasibility of promoting alternative types of services or systems that result in more effective and efficient use of the spectrum.

Techniques such as power control, different modulation schemes and multiplexed access in time, space and code space may be used to promote spectrum efficiency. The rapidly developing capabilities of digital electronics and signal processing are making practical many processes that were once only theoretically

---

possible. A Software Defined Radio (SDR) system is a radio communication system whose channel modulation waveforms and other communication components are defined and implemented in software rather than implemented in hardware. In the long term, software defined radio is expected by its proponents to become the dominant technology in radio communications.

Cognitive radio, which is introduced as the next step of software defined radio, is a paradigm for wireless communication in which either a network or a wireless node changes its transmission or reception parameters to communicate efficiently avoiding interference with licensed or unlicensed users. This alteration of parameters is based on the active monitoring of several factors in the external and internal radio environment, such as radio frequency spectrum, user behavior and network state [51]. The idea of cognitive radio was first presented officially in an article by Joseph Mitola III and Gerald Q. Maguire, Jr in 1999 [24]. It was thought of as an ideal goal towards which a software-defined radio platform should evolve: a fully reconfigurable wireless black-box that automatically changes its communication variables in response to network and user demands.

# Chapter 2

## Radio Spectrum Occupancy Monitoring

### 2.1 Considerations for Spectrum Monitoring

Cognitive radio technologies require reconsideration the current regulations and policies of spectrum management. Spectrum monitoring is a fundamental function to support spectrum management. While some crude information can be retrieved from spectrum licences, essential details, including the location of transmitters, transmitter output power, and antenna type, are often unknown. Additionally, licences do not specify how often the spectrum is being occupied. Furthermore, the local environment affects the propagation of radio waves. While this effect can be simulated, the results are hardly precision. Hence, to categorize spectrum usage, Spectrum monitoring are vastly preferable to theoretical analysis.

Spectrum measurements are critical to policy regulators and researchers in the development of new spectrum access technologies. Specifically, spectrum occupancy studies inspect what spectrum bands have low or no active utilization and thus may be appropriate for spectrum sharing. They also provide information on the signal characteristics within these bands, which is needed to design spectrum sharing algorithms. Referred to [26] the considerations of spectrum monitoring activities in our project include the following:

- that licensed user information from the frequency management databases only indicates that the use of the frequency is authorized. The number of assignments on a frequency does not give any actual use information of that particular frequency.
- that efficient Spectrum Management can only satisfactorily proceed if the monitoring information provides the radio spectrum regulators with adequate reliable information about the actual usage of the spectrum.
- that results of spectrum occupancy measurements will give information about the current use of frequencies to establish that the spectrum is being used efficiently and to assess the feasibility of the new technologies.

The overall goal of spectrum monitoring activities of our project is to depict the current level of spectrum usage in the range 100 to 2500 MHz and its implications for cognitive radio. Central objectives include the following:

- to provide information of spectrum efficiency for determining planned and actual frequency usage and occupancy, and for assessing the feasibility of spectrum sharing technique.
- to provide data for statistical modeling.

The measuring system should be chosen carefully to ensure capabilities exist with the spectrum management agency to effectively monitor and analyze the frequency bands.

## 2.2 Monitoring System and Site

A successful spectrum survey requires careful selection of a measurement site. The location chosen for measuring will affect measured spectrum occupancy. For example, measurements made in Durham are probably representative of many towns that have similar scale and do not have heavy military activity or maritime radio navigation etc. Generally, a site for spectrum monitoring requires [49]:

## 2.2 Monitoring System and Site

---

- “limited numbers of nearby transmitters to prevent intermodulation or saturation problems that can arise even though preselection and/or filtering is used for survey measurements
- limited man-made noise such as impulsive noise from automobile ignition systems and electrical machinery that can add to the received signals of interest.”

Figure 2.1 shows that the measurement locations for the spectrum occupancy project was the roof and inside of Engineering Building, Durham University.



Figure 2.1: Measurement location

Because of the complexity and sophistication of wireless communication technologies, it is an ever-increasing challenge to monitor the spectrum, particularly considering the rapid growth of wireless, satellite, and point-to-point communication devices. Key considerations in the design of spectrum monitoring systems include types of equipment, data rate and complexity of data capture and processing, degree of integration with software tools for analysis. While considered the limited project budget and the existing equipment, the author integrated independently 3 different spectrum monitoring systems for different frequency bands to satisfy the technique requirements.

The monitoring system for 100-1500 MHz spectrum in this project was configured as Figure 2.2. The system consisted of an omnidirectional Dressler ARA-1500 [55] active antenna range from 50 to 1500 MHz connected by a 6 m RF

## 2.2 Monitoring System and Site

cable to a diplexer RSM-2000 which allows the DC current for the preamplifier to be applied to the centre conductor of the RF coax, eliminating the need for an additional DC power feed conductor. The RSM-2000 also contains a 20 dB RF adjustable attenuator allowing received signals to be attenuated over the entire frequency range. In order to increase the dynamic range of the system, a highpass filter Mini-Circuits SHP-100 was connected to the output of the RSM-2000. A Jim M-75 low noise amplifier was inserted in front of HP 8560A spectrum analyzer to decrease its noise figure. The GPIB bus was used for logging the trace data onto the hard disk of a PC and for transferring control command sequences to the spectrum analyzer. Table 2.1 lists the configuration parameters of spectrum analyzer for scanning 100-1500 MHz. Justifications will be given in the next section.

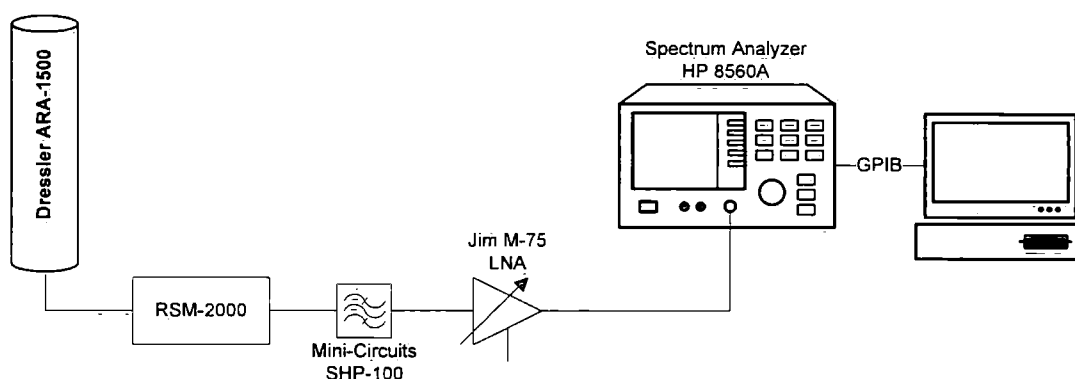


Figure 2.2: Monitoring system for 100-1500 MHz spectrum

The monitoring system for 1500-2500 MHz spectrum was configured as in Figure 2.3. Instead of an omnidirectional antenna, an antenna array consisted of 4 directional Log-periodic antennas which enable the detection of signal incident directions. An RF switch controlled by the PC parallel port was used to choose a given antenna. Table 2.2 lists the configuration parameters of spectrum analyzer for scanning 1500-2500 MHz. Justifications will be given in the next section.

The monitoring system for 2400-2483.5 MHz spectrum is shown in Figure 2.4. The solution of spectrum analyzer plus RF switch is incapable in catching fast frequency hopping signals which are adopted by many wireless applications using

## 2.2 Monitoring System and Site

Model	HP 8560A
Frequency span (MHz)/ sweep	6
Resolution bandwidth (kHz)	10
Sweep time (s)	6
Detection mode	Sample
RF attenuator (dB)	10
Reference level (dBm)	-20

Table 2.1: Configuration parameters for scanning 100-1500 MHz

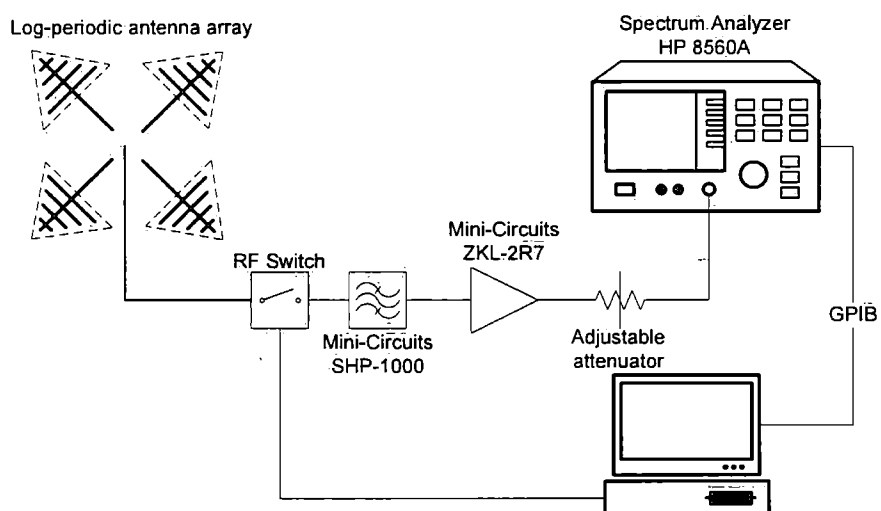


Figure 2.3: Monitoring system for 1500-2500 MHz spectrum

Model	HP 8560A
Frequency span (MHz)/ sweep	12
Resolution bandwidth (kHz)	10
Sweep time (s)	6
Detection mode	Sample
RF attenuator (dB)	10
Reference level (dBm)	-20

Table 2.2: Configuration parameters for scanning 1500-2500 MHz

this band. In order to maximize the signal catching probability, a channel sounder was deployed in monitoring the 2.4 GHz ISM band. The channel sounder [48] has the capability of scanning 8 channels simultaneously with a 300 MHz frequency span every 4 ms. An 8 element log-periodic antenna array was used as monitoring aeriels. The structure of the channel sounder will be illustrated in the next sector.

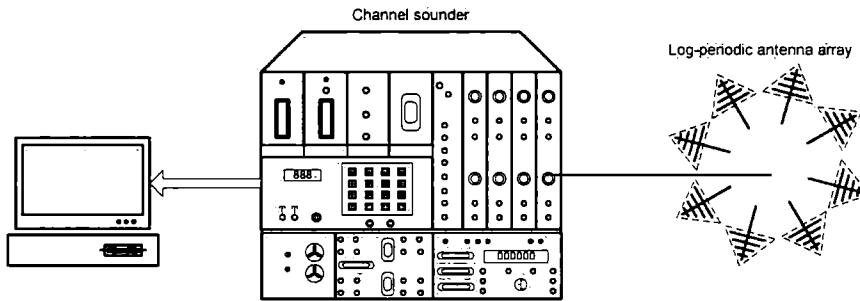


Figure 2.4: Monitoring system for 2.4 GHz ISM band

## 2.3 Monitoring Technology

The sophistication, complexity and prices of monitoring equipment have risen with the advent of spread spectrum and computer-based radio technologies like Cognitive Radio. Also, the approaches to monitoring and the architecture of the

spectrum monitoring system have a bearing on the types of systems needed and the configuration of operations and resources.

There are system architectural and functional aspects to designing spectrum monitoring systems. The basic types of monitoring equipment include radio receivers, spectrum analyzers, direction-finding equipment and antennas.

### 2.3.1 receiving antenna

An antenna converts electromagnetic waves into electrical currents and vice versa. Only receiving antennas have been employed in the spectrum monitoring measurements. A typical problem encountered with modern wideband monitoring is the choice of suitable antennas. In the spectrum occupancy project a wideband active antenna whose frequency range is 50-1500 MHz and wideband log periodic antennas whose frequency range are 800-2500 MHz have been deployed as receiving aeriels.

The usual choices for wideband receiving antenna are discone antennas and log-periodic antennas. High gain and directional antennas are more desirable in this project since the spatial spectrum occupancy is another subject which we want to survey in the frequency range 1500-2500 MHz, . Log periodic antenna TDJ-0825DSA manufactured by Kenbotong Communication Ltd was selected to cover the frequency range from 800 to 2500 MHz [34]. The specifications of the TDJ-0825DSA used in occupancy project is given in table 2.3 and the return loss figure is shown in Figure 2.5 measured with Agilent E5071A network analyzer.

Usually, passive antennas have low gain and take large physical spaces if they are a broadband design, especially in the lower bands because of the inverse relation between antenna and the size of an antenna. For wideband scanning measurement an active antenna is a good alternative, which has widely been used in Global Positioning System (GPS) and in wireless networks. An omnidirectional Dressler ARA-1500 active antenna was used to monitor 100-1500 MHz frequency spectrum. The specifications of the ARA-1500 [55] are given in table 2.4.

However, an active receiving antenna is capable of introducing some problems which will not happen with a passive antennas. Limitations of the active receiving

Model	TDJ-0825DSA
Frequency range (MHz)	800-2500
Gain (dBi)	8.2-9.5
Horizontal beamwidth	90°-70°
Vertical beamwidth	60°-55°
Front-to-back ratio (dB)	>18
VSWR	<1.5
Maximum power (W)	100
Input impedance ( $\Omega$ )	50

Table 2.3: Specifications for TDJ-0825DSA log periodic antenna

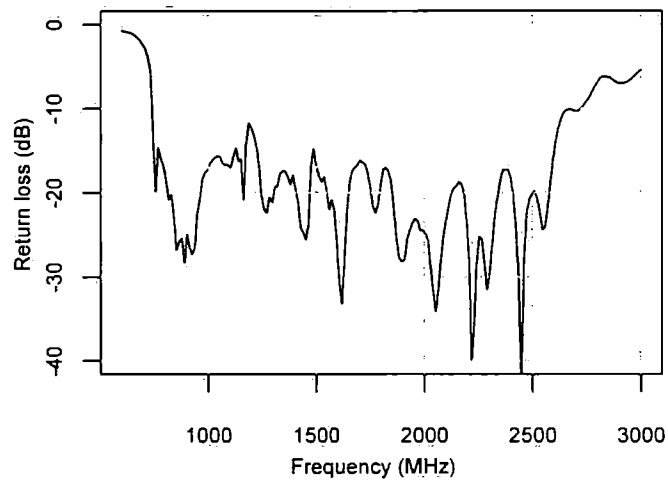


Figure 2.5: Return loss for TDJ-0825DSA log periodic antenna

Model	Dressler ARA-1500
Frequency range (MHz)	50-1500
Gain (dB)	11
Beamwidth	360°
Noise figure	3-4
Third order IP (dBm)	>21
Input impedance ( $\Omega$ )	50

Table 2.4: Specifications for Dressler ARA-1500 omnidirectional antenna

antenna include: 1. compromise of receiver dynamic range 2. cross modulation distortion 3. intermodulation distortion.

### 2.3.2 spectrum monitoring equipment

Spectrum monitoring equipment is an electronic device that must perform two basic functions [9]:

- “it must respond to and detect desired signals.
- it must not respond to, detect, or be adversely affected by undesired signals.”

The electrical signal can be sampled either in the time domain or in the frequency domain. Monitoring in time domain can be performed with radio measuring receiver, and in the frequency domain can be performed with spectrum analyzer. In our project, a spectrum analyzer has been employed since it can provide faster sweep speed than a measuring receiver.

#### 2.3.2.1 spectrum analyzer

A Hewlett-Packard 8560A spectrum analyzer was used as the equipment to provide spectral power measurement over the frequency range 100-2500 MHz. Figure 2.6 shows the simplified block diagram of the HP 8560A spectrum analyzer [1]. An input signal passes through an attenuator, then through a low-pass filter to

a mixer, where it mixes with a signal from the local oscillator (LO). Its output includes not only the two original signals, but also their harmonics and the sum and differences of the original frequencies and their harmonics since the mixer is a non-linear device. If any of the mixed signals falls within the passband of the intermediate-frequency (IF) filter, it is further amplified and perhaps compressed on a logarithmic scale. It is essentially rectified by the envelope detector and displayed. A ramp generator creates the horizontal movement across the display from left to right. The ramp also tunes the LO so that its frequency is in proportion to the ramp voltage.

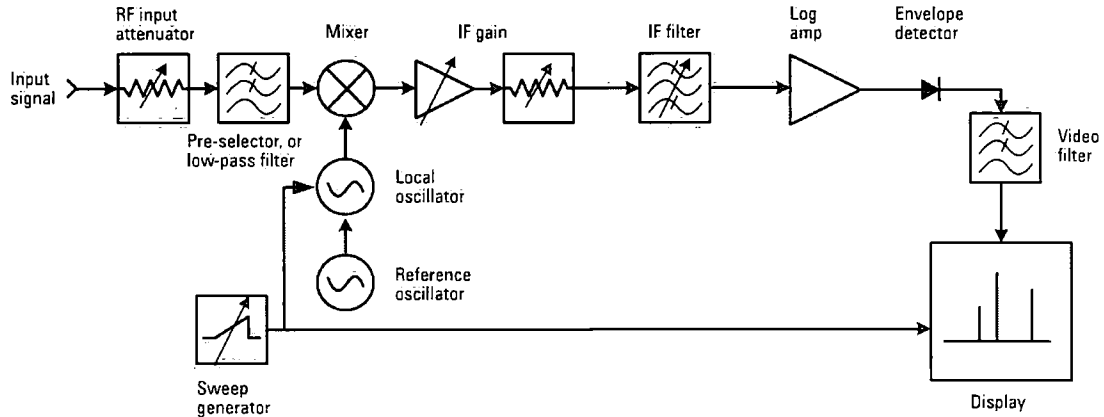


Figure 2.6: Block diagram of a spectrum analyzer [1]

The setting for resolution bandwidth, detector type, span, swept time, reference level, attenuation level, and data collection method were chosen with the intent of maximizing the probability of detection.

Since the power amplitudes of the strongest signals in our measurement location (referring Figure 2.1) are around  $-20$  dBm, the RF attenuator was set to 10 dB based on  $-20$  dBm reference level. The RF attenuator setting range of the HP 8450A spectrum analyzer is from 0 to 70 dB in steps of 10 dB. In order to prevent overload, gain compression and distortion, 10 dB attenuation is considered as a cautious and reasonable choice in the experimental radio environment.

The IF filter in Figure 2.6 determines the overall spectrum resolution. A narrow 10 kHz resolution bandwidth was employed to maximize the detection

of narrowband signals and to resolve the spectral content of wider bandwidth signals. However, if resolution was the only criterion in our experiment, the narrowest possible resolution IF filter will be chosen. But resolution affects sweep time which directly affects how long it takes to complete a measurement.

Because the IF filters are band-limited circuits they require finite times to charge and discharge. The time that a mixing product stays in the passband of the IF filter is directly proportional to its bandwidth and inversely proportional to the sweep in Hz per unit time [1], or:

$$\text{Time in passband} = \frac{RBW}{Span/ST} \quad (2.1)$$

where  $RBW$  is resolution bandwidth,  $ST$  is sweep time and  $Span$  is the bandwidth that the spectrum analyzer covers in one sweep. On the other hand, the rise time of a filter is inversely proportional to its bandwidth,

$$\text{Rise time} = \frac{k}{RBW} \quad (2.2)$$

where  $k$  is constant of proportionality and which value is in the 2 to 3 range for the synchronously-tuned, near-Gaussian filters used in many Agilent analyzers. If we make the rise time equal the time in passband and solve for sweep time, we have:

$$ST = \frac{k \cdot Span}{RBW^2} \quad (2.3)$$

One disadvantage of using a narrow bandwidth filter is the reduced ability to observe pulsed signals and hopping signals. In practice, the value  $k$  is around 3<sup>1</sup>. From (2.2), under the 10 kHz filter setting the measured power of pulses with duration of less than 0.3 ms is attenuated. Moreover, the re-visit time<sup>2</sup> is too long for certain signals such as hopping signals. On the other hand, if a wide bandwidth filter is chosen to improve the ability to observe pulsed signals and to decrease the re-visit time, the sensitivity and spectral content of spectrum analyzer will suffer correspondingly. For example, if the 100 kHz resolution bandwidth is chosen instead of 10 kHz, the noise floor will be 10 dB higher than before.

---

<sup>1</sup>When the span and RBW of HP 8560A have been set to 10 MHz and 10 kHz respectively, the readout of sweep time is 300 ms. From this the  $k$  value can be roughly estimated.

<sup>2</sup>the time taken to visit all the channels to be measured and return to the first channel

Spectrum analyzers have a defined number of discrete frequency buckets to save the results of a sweep; for the HP 8560A there are 601 fixed data points across the whole span. We had to decide what sample value should be displayed for each display data point. A bucket which is shown in Figure 2.7 is the interval between two consecutive trace points. Each bucket contains data from a span and time frame that is determined by:

$$\text{Frequency : bucket width} = \text{Span} / (\text{trace points} - 1)$$

$$\text{Time : bucket width} = \text{sweeptime} / (\text{trace points} - 1)$$

HP 8560A spectrum analyzer has four detector types:

- Sample detection
- Positive peak detection
- Negative peak detection
- Normal detection

The sample, positive peak, and negative peak are easily understood and visually represented in Figure 2.7. Positive and negative peak detection display the maximum and minimum values encountered in each bucket. Peak detection does not give a good representation of random noise because it only displays the maximum value in each bucket and ignores the true randomness of the noise [1]. Sample detection selects the data point as the instantaneous level at the center of each bucket. While the sample detection mode does a good job of indicating the randomness of noise, it is not a good mode for analyzing sinusoidal signals. To provide a better visual display of random noise than peak and yet avoid the missed-signal problem of the sample mode, the normal detection displays the maximum value of the odd-numbered data point and the minimum value of even-numbered data point encountered during its bucket.

In spectrum occupancy experiments, the spectrum analyzer was configured for sample detection mode followed by ‘RFI Measurement Protocol for Candidate SKA Sites’ by [2], which shares many common issues with our project. Although peak detection method has the ability to detect pulsed or intermittent signals,

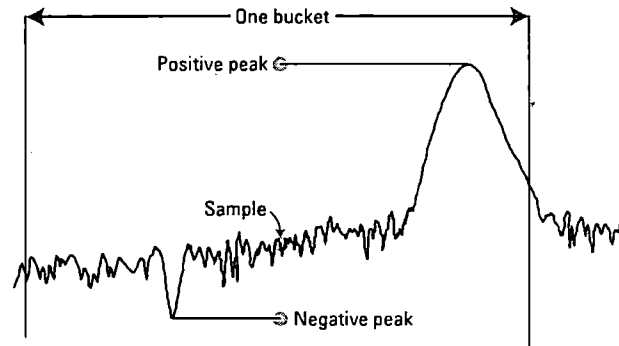


Figure 2.7: Trace point based on detector algorithm [1]

their presence can be overstated with this detection. Peak detection does some pre-selection of the data which makes it difficult to deduce the real rms power. Sample detection data, however, still contains all the statistical information. So for radiometric measurements comparing observed RFI rms power levels with clean parts of the spectrum band it is better to use sample detection.<sup>1</sup>

The sampling rates are different for various instruments, but greater accuracy can be obtained from decreasing the span and/or increasing the sweep time since the number of samples per bucket will increase in either case. For instance, if the span was set to 600 MHz spectrum, the spectrum analyzer will only be able to resolve measurements to a precision of 1 MHz. In order to obtain 10 kHz bandwidth resolution and to provide the maximize amount of valid data for data analysis, 6 MHz span was set to transfer the measurements of every 10 kHz resolution bandwidth. Abiding by [2], the dwell time<sup>2</sup> was set to 10 ms. Therefore, the 6 s sweep time is required for 6 MHz bandwidth.

Since the spectrum analyzer was not capable of detecting wideband pulses and hopping signals, our strategy is to optimize the sweep time to increase the detection accuracy and use dedicated equipment, which will be introduced in next section, to catch the fast signals.

<sup>1</sup>Acknowledge Professor Albert-Jan Boonstra, ASTRON for explaining this.

<sup>2</sup>the length of time that one slice of spectrum having width equal to the specified RBW is examined

Bandwidth (kHz)	Sweep time (ms)	Noise floor (dBm)
1000	50	-80
300	50	-84
100	50	-90
30	300	-94
10	300	-100

Table 2.5: Relationship between bandwidth and sweep time of HP 8560A

### 2.3.2.2 channel sounder

Modern RF signals such as frequency hopping signals change over time, often unpredictably. The traditional swept spectrum analyzers might provide snapshots of the signal in the frequency and modulation domains, but this is often not enough information to confidently describe the dynamic RF signals. For example, Table 2.5 [58] shows the relationship between *Bandwidth*, *Sweep time* and *Noise floor* of the HP 8560A spectrum analyzer in scanning the 2.4 GHz ISM band. Under the 1000 kHz *Bandwidth* setting, it takes 50 ms to scan the 83.5 MHz spectrum. This speed is rather slow compared with most hopping signals.

The channel sounder is a device originally designed for measuring the spatial and temporal characteristics of the multiple-input, multiple-output (MIMO) radio channel. A semi-sequential MIMO channel sounder system of Durham University, which can measure high resolution delay/Doppler information as well as spatial channel information, consists of transmitter and receiver. The receiver can be operated independently as a spectrum monitoring receiver which offers better performance in terms of fast scanning of wideband spectrum with multiple input channels.

Figure 2.8 shows the diagram of the receiver of the channel sounder [21; 33]. There are many similarities between the sounder system and a spectrum analyzer. The RF front end components have the same functions described in the preceding section. The local oscillator and sweep generator in spectrum analyzer have been replaced by a chirp signal generator. The synchronization unit of the sounder manages the flow of the data between the computer and the sounder, the signal

conditioning unit, RF switch and RF attenuators.

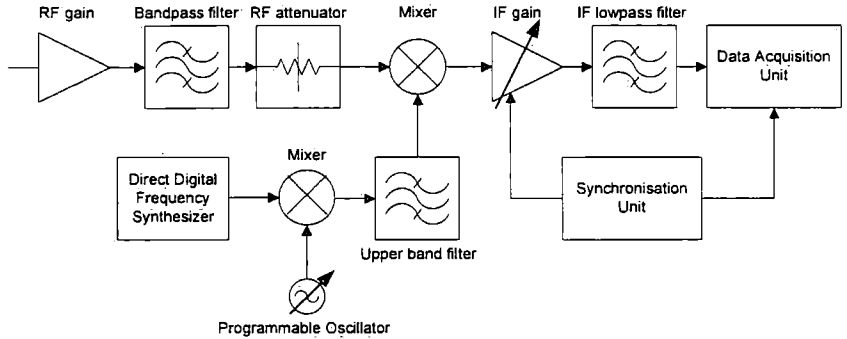


Figure 2.8: Block diagram of the receiver of the channel sounder

A chirp signal commonly used in sonar and radar is a signal in which the frequency increases or decreases with time. The chirp signal can be expressed as:

$$y(t) = \sin \left( 2\pi \left( f_0 + \frac{k}{2}t \right) t \right) \quad 0 \leq t \leq T$$

where  $f_0$  is the starting frequency (at time  $t = 0$ ),  $k$  is the rate of frequency increase and  $T$  is the sweep time. The waveform of a linear chirp signal is shown in Figure 2.9. The programmable direct digital frequency synthesizer can generate chirp signals whose bandwidth can be up to 300 MHz with 300 Hz sweep repetition frequency. This baseband chirp signal is modulated to 1.9 to 2.4 GHz passband signal by a programmable source. After filtering the undesired signals, this up-converted chirp signal is fed into the mixer's local oscillator port.

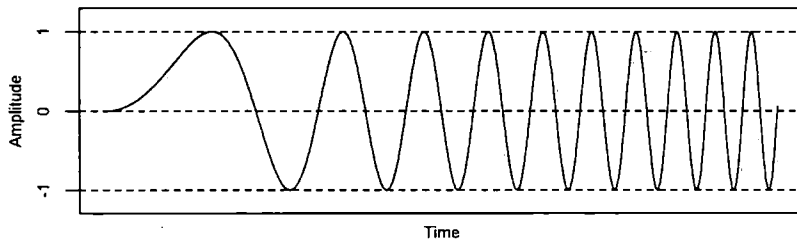


Figure 2.9: Linear chirp signal

The resultant IF signal must be optimized for the input range of the data acquisition unit so that a reasonable signal to noise ratio can be achieved [21; 33]. The signal conditioning unit amplifies the IF signal by up to around 60 dB by 2 stage amplification. One of two IF filters which have 165 kHz and 250 kHz bandwidths respectively can be chosen to remove any DC component and any unwanted higher frequencies.

The synchronization unit is the control unit of the channel sounder which contain a micro-controller and two field reprogrammable gate arrays. It communicates with the computer via 16 control commands shown in Table 2.11. The functions of these commands include:

- controlling the data flow between computer and sounder
- controlling the RF attenuator and IF gains
- generating different clock signals for the data acquisition unit

The details and protocols can be found in [21; 33].

### 2.3.3 noise and system sensitivity

Caused either by thermal noise and other electronic noise from the receiver input circuits or by interference from radiated electromagnetic noise picked up by the receiver's antenna, Radio noise in radio reception is the superposition of white noise and other disturbing influences on the signal. Knowledge of sources of noise may lead to methods by which it can be suppressed.

Noise comes from a number of different sources, but for the sake of this project they are divided into two categories: sources external to the receiver and sources internal to the receiver. Little can be done on external noise such as atmospheric, galactic and man made noise which fall within the receiver's passband [45].

Thermal noise in a receiver, also called white noise, is the electronic noise generated by the thermal agitation of the electrons inside an electrical conductor, which happens regardless of any applied voltage. A resistor generates noise power equal to [28]

$$P_n = kTB \text{ Watts}$$

Bandwidth	Noise power (dBm)	Notes
1 Hz	-174	
10 Hz	-164	
1000 Hz	-144	
10 kHz	-134	FM channel of 2-way radio
1 MHz	-114	
2 MHz	-111	Commercial GPS channel
6 MHz	-106	Analog television channel
20 MHz	-101	WLAN 802.11 channel

Table 2.6: Noise floors of different bandwidths

where  $P_n$  is noise power,  $k$  is Boltzmann's constant ( $1.38 \times 10^{-23} J/K$ ),  $T$  is temperature in degree Kelvin ( $K$ ) and  $B$  is the bandwidth in Hertz (Hz). Thermal noise for a resistor at room temperature ( $298K$ ) can be described in decibels relative to 1 milliwatt as:

$$P_{dBm} = -174 + 10 \log(B) \quad (2.4)$$

Equation (2.4) shows the total noise power is a function of the measurement bandwidth. For example, with 1 Hz bandwidth the noise floor will be -174 dBm. A signal must be above this noise level before it can be successfully detected. Table 2.6 shows the relation between bandwidth and corresponding noise floor.

The noise generated at the resistor can be transferred to the remaining circuit. When this noise passes through the first gain stage, the amplifier boosts the noise, and adds some of its own. The final noise level appeared in the monitor is referred to as *sensitivity*<sup>1</sup> [1]. Figure 2.10 shows the noise signal in the output of the HP 8560A spectrum analyzer with 10 kHz resolution bandwidth connected via a 50  $\Omega$  resistor to prevent receiving the signals from the air.

Noise factor is defined as the degradation of signal-to-noise ratio [9]:

$$F = \frac{S_i/N_i}{S_o/N_o}$$

---

<sup>1</sup>The term display average noise level is used in spectrum analyzer. Strictly, Sensitivity is a measure of the minimum signal level that yields a defined signal-to-noise ratio (SNR) or bit error rate (BER).

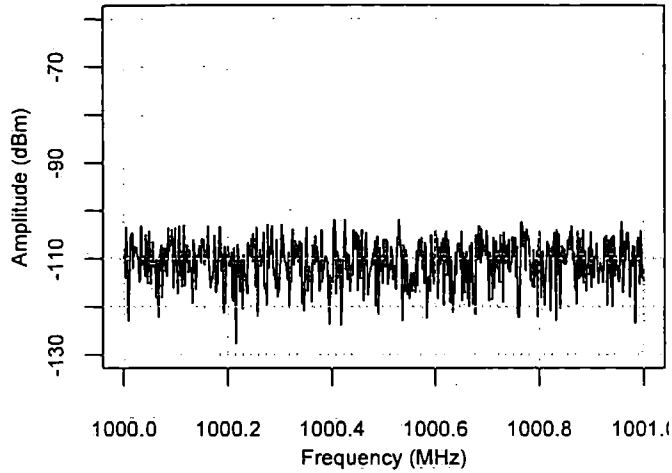


Figure 2.10: The noise signal with 10 kHz resolution bandwidth

where  $S_i$  and  $S_o$  are input and output signal power,  $N_i$  and  $N_o$  are input and output noise power respectively. If a receiver has unity gain, this definition can be simplified as:

$$F = \frac{N_o}{N_i}$$

The noise figure usually expressed in terms of dB of the noise factor:

$$NF = 10 \log(N_o) - 10 \log(N_i)$$

The noise figure of HP 8560A can be roughly estimated from the measured noise. From Figure 2.10, the maximum measured noise power is around  $-102^1$  dBm in 10 kHz resolution bandwidth. the noise figure can be derived:

$$\begin{aligned} NF &\approx [\text{measured noise}] - (-174 + 10\log 10000) \\ &= -102 + 174 - 40 \\ &= 32 \text{ dB} \end{aligned}$$

Noise figure is independent of resolution bandwidth of the receiver.

---

<sup>1</sup>This value is different with the value in Table 2.5 since there the resolution bandwidth is in uncoupled condition.

The 32 dB noise figure of HP 8560A means that a radio signal can be detected only if it is 32 dB above the noise power  $kTB$ . However, the amplitude of some kind of radio signals are below this threshold. For example, at the base station, 900 MHz GSM specification defines the minimum reception levels as -104 dBm for the base station [32]. So, there is not enough signal-to-noise ratio left for detecting the radio signals.

By introducing a *preamplifier* this problem can be solved. By placing an appropriate preamplifier in front of the spectrum analyzer, we can obtain a system (preamplifier + spectrum analyzer) noise figure that is lower than that of the spectrum analyzer alone. Rather than develop complex formulas to calculate the final system noise figure, [1] gives two criteria to estimate the system noise figure: criterion 1

$$\text{If } NF_{\text{pre}} + G_{\text{pre}} \geq NF_{\text{sa}} + 15 \text{ dB}$$

$$\text{Then } NF_{\text{sys}} = NF_{\text{pre}} - 2.5 \text{ dB}$$

and criterion 2

$$\text{If } NF_{\text{pre}} + G_{\text{pre}} \leq NF_{\text{sa}} - 10 \text{ dB}$$

$$\text{Then } NF_{\text{sys}} = NF_{\text{sa}} - G_{\text{pre}}$$

where  $NF_{\text{pre}}$  is the noise figure of a preamplifier,  $G_{\text{pre}}$  is the gain of a preamplifier in dB,  $NF_{\text{sa}}$  is the noise figure of a spectrum analyzer and  $NF_{\text{sys}}$  is the noise figure of the system.

However, the drawback to using a preamplifier is loss of measurement range. A preamplifier must meet criterion 2 to get better sensitivity without costing measurement range. That is, the sum of its gain and noise figure must be at least 10 dB less than the noise figure of the spectrum analyzer. In this case the displayed noise floor will not change noticeably, although we shift the whole measurement range down by the gain of the preamplifier, we end up with the same overall range that we started with.

In spectrum occupancy project, a Jim M-75 low noise wideband GaAs FET preamplifier was used in scanning 100-1500 MHz spectrum; Mini-Circuits ZKL-2R7 amplifier was used in scanning 1500-2500 MHz spectrum. The specifications of Jim M-75 preamplifier is in Table 2.7. Since there are strong signals in the low frequency spectrum, the gain of Jim M-75 amplifier was set to 10 dB to avoid

Model	Jim M-75 low noise preamplifier
Frequency range A band (MHz)	225-1500
Frequency range B band (MHz)	108-185
Frequency range C band (MHz)	24-2150
Gain (dB)	from -10 to 20 fully adjustable
Noise Figure (dB)	2
Impedance ( $\Omega$ )	50

Table 2.7: Specifications for Jim M-75 preamplifier

internal distortion. According to criterion 2, the noise figure of the preamplifier and spectrum analyzer is

$$NF_{\text{pre}} + G_{\text{pre}} = 12 \text{ dB} \leq NF_{\text{sa}} - 10 \text{ dB} = 20 \text{ dB}$$

$$NF_{\text{sys}} = NF_{\text{sa}} - G_{\text{pre}} = 20 \text{ dB}$$

The deployment of active antenna Dressler ARA-1500 complicates the final calculation of the measurement system since the Dressler ARA-1500 has its own noise figure. A noise signal is seen by any following amplifier as a valid input signal. Each stage in the cascade chain amplifies both signals and noise from the previous stages and also contributes some additional noise of its own. Thus, in a cascade amplifier chain in Figure 2.11 the final stage sees an input signal that consists of the original signal and noise amplifier by each successive stage, plus the noise contributed by earlier stages. The overall noise factor for a cascade amplifier can be calculated from *Friis' noise equation* [46]:

$$F_n = F_1 + \frac{F_2 - 1}{G_1} + \dots + \frac{F_N - 1}{G_1 G_2 \dots G_N} \quad (2.5)$$

where  $F_n$  is overall noise factor of N stages in cascade,  $F_1$  is the noise factor of stage 1,  $F_N$  is the noise factor of the nth stage,  $G_1$  is the gain of stage 1 and  $G_N$  is the gain of stage (n-1). From (2.5), the final noise factor can be calculated as:

$$F_n = F_1 + \frac{F_2 - 1}{G_1} = 10^{\frac{4}{10}} + \frac{10^{\frac{20}{10}} - 1}{11} = 12.41$$

The final noise figure is  $\log 12.41 = 10.84 \text{ dB}$ .

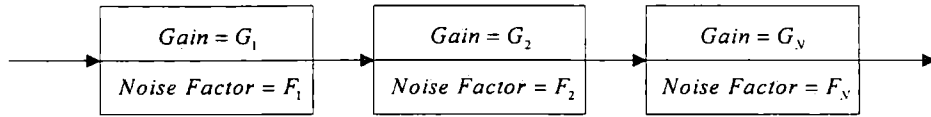


Figure 2.11: Cascade amplifier chain for noise figure calculations

Model	Mini-Circuits ZKL-2R7
Frequency range (MHz)	10-2700
Gain (dB)	24
Gain flatness	$\pm 0.7$
Noise Figure (dB)	5
IP3 (dBm)	30
1 dB compression (dBm)	13
Input no damage (dBm)	13

Table 2.8: Specifications for Mini-Circuits ZKL-2R7

The specifications of Mini-Circuits ZKL-2R7 amplifier is in Table 2.8. The noise figure of the preamplifier and spectrum analyzer can be calculated by criterion 2

$$NF_{pre} + G_{pre} = 29 \text{ dB} \leq NF_{sa} - 10 \text{ dB} = 20 \text{ dB}$$

$$NF_{sys} = NF_{sa} - G_{pre} = 6 \text{ dB}$$

This setting was used in monitoring 1500-2500 MHz spectrum with the high gain directional log periodic antennas.

### 2.3.4 distortion and dynamic range

The strength of GSM downlink signals near the base station is higher than  $-20$  dBm amplitude while the strength of satellite downlink signals is less than  $-120$  dBm amplitude. The implication that radio signals have dramatic dynamic range is that the receiver must have sufficient dynamic range to recover the entire signal. Otherwise, information may be lost and the output is distorted. Understanding the dynamic performance of the receiver requires knowledge of *harmonic distort-*

*tion* and *intermodulation products* (IP) and knowledge how they affect receiver operation.

Harmonic distortion can be defined as a single-tone distortion product caused by device non-linearity. Since the mixer diode of a spectrum analyzer is a non-linear device, it always generates distortion of its own. The current through an ideal diode [1] can be expressed as:

$$i = I_s (e^{qv/kT} - 1)$$

$I_s$  is the diode's saturation current,  $q$  is electron charge  $1.6 \times 10^{-19}$  coulombs,  $v$  is instantaneous voltage,  $k$  is Boltzmann's constant ( $1.38 \times 10^{-23}$  J/K) and  $T$  is temperature in degree Kelvin ( $K$ ). This expression can be expanded into a Taylor power series:

$$i = I_s (k_1 v + k_2 v^2 + k_3 v^3 + \dots)$$

where  $k_1 = q/kT$ ,  $k_2 = k_1^2/2!$  and  $k_3 = k_1^3/3!$ . For example, a mixer is fed into a local oscillator signal frequency at  $\omega_{LO}$  and a sinusoidal signal frequency at  $\omega_1$

$$v = V_{LO} \sin(\omega_{LO} t) + V_1 \sin(\omega_1 t)$$

We arrive at terms involving the input and local oscillator signals<sup>1</sup> as long as the input signal is in the linear range which is 15 to 20 dB below the level of LO:

$$k_2 V_{LO} V_1 \cos[(\omega_{LO} - \omega_1) t]$$

$$(3k_3/4) V_{LO} V_1^2 \sin[(\omega_{LO} - 2\omega_1) t]$$

$$(k_4/8) V_{LO} V_1^3 \sin[(\omega_{LO} - 3\omega_1) t]$$

...

The argument  $\omega_1$  of the sine represents the fundamental tone which is actually the frequency of the input signal. The argument  $2\omega_1$  of the sine represents the second harmonic of the input signal. The level of this second harmonic is a function of the square of the voltage of the fundamental,  $V_1^2$ . For every dB that we drop the level of the fundamental at the input mixer, the internally generated second

---

<sup>1</sup>The terms which frequencies are above LO are assumed to be filtered.

harmonic drops by 2 dB. The argument holds true for the  $n^{\text{th}}$  harmonic. For every dB that we drop the level the  $n^{\text{th}}$  harmonic drops by  $n$  dB.

Depressing the harmonic distortion can be addressed by two different approaches [9]. One way is reducing the distortion level back down under the noise floor by increasing the input RF attenuation referring Figure 2.12. Even a few dB of input attenuation is often enough to cause second and higher order<sup>1</sup> distortion to drop back into the noise floor, while affecting the desired signals only by a small amount. In the spectrum occupancy project, this approach was adopted by setting 10 dB input RF attenuation to reduce distortion.

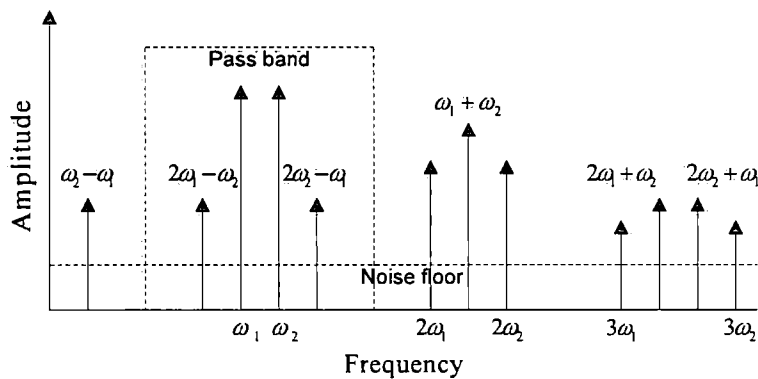


Figure 2.12: Second and third order distortion

Another approach that can help reduce distortion is to introduce the preselector. A preselector is an electronic device that is inserted between the antenna and the receiver, limiting the range of frequencies that can be applied to it. Tuning to the desired frequency keeps the preselector's narrow bandwidth centered at the operating frequency, rejecting or reducing out-of-band unwanted interference signals. From Figure 2.12, if we choose a proper passband of the preselector and proper span of spectrum analyzer, we can exclude distortion out of measurement range.

---

<sup>1</sup>Distortion is often described by its order. The order can be determined by noting the coefficient associated with the signal frequency or the exponent associated with the signal amplitude.

Two high pass filters were employed as preselectors in this project to provide better dynamic range and reject out-of-band signals. Mini-circuits SHP-100 high pass filter was used in the occupancy measurement range from 100-1500 MHz; mini-circuits SHP-1000 high pass filter was used in the occupancy measurement range from 1500-2500 MHz. The insertion loss and return loss are shown in Figure 2.13. In addition, the 6 MHz span was set, which is much narrower than the covered frequency range (100-2500 MHz). The narrow span excludes distortion generated by the mixer of the spectrum analyzer.

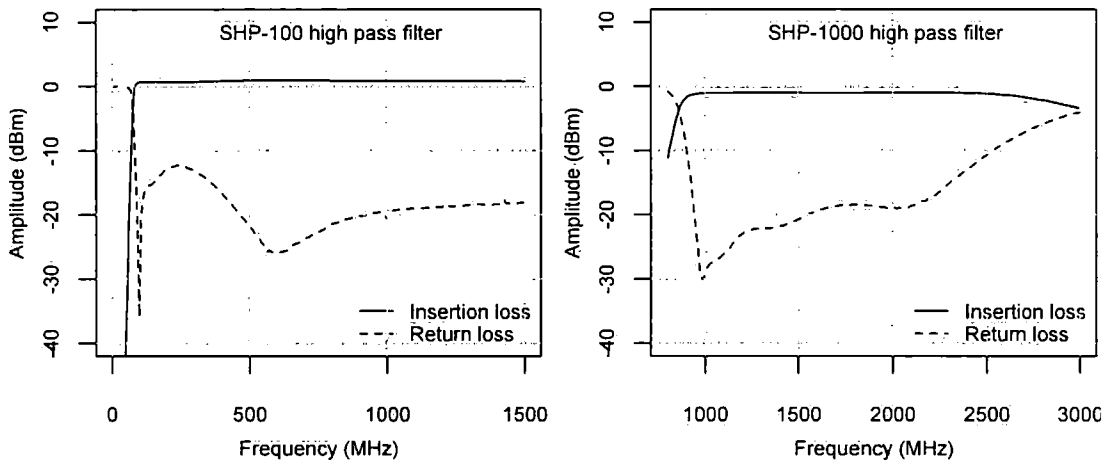


Figure 2.13: The insertion loss and return loss of the preselectors

Comparing harmonic distortion, it is difficult to get rid of intermodulation distortion since it often falls into the measurement band. For example, two sinusoidal signals are fed into a local oscillator:

$$v = V_{LO} \sin(\omega_{LO}t) + V_1 \sin(\omega_1t) + V_2 \sin(\omega_2t)$$

in addition to harmonic distortion, we get:

$$(k_4/8) V_{LO} V_1^2 V_2 \cos[\omega_{LO} - (2\omega_1 - \omega_2)] t$$

$$(k_4/8) V_{LO} V_1 V_2^2 \cos[\omega_{LO} - (2\omega_1 - \omega_2)] t$$

...

## 2.3 Monitoring Technology

From Figure 2.12, intermodulation distortion, the interaction of the two input signals with each other, is so close to the measuring signals that it falls into the measurement bandwidth. By increasing the input attenuation, the IP levels can be reduced back down under the noise floor. Third order intermodulation products are more important because they tend to reflect on receiver's dynamic range, as well as its ability to handle strong signals.

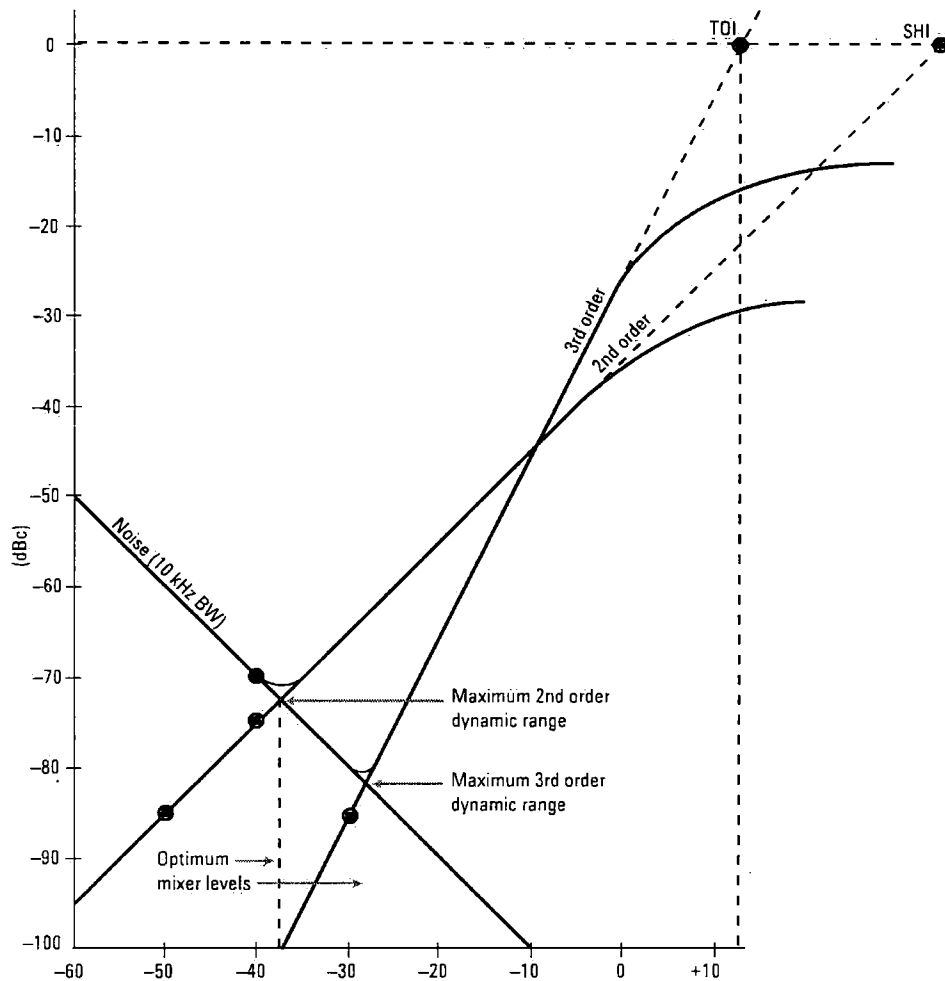


Figure 2.14: Dynamic range versus distortion [1]

With the help of Figure 2.14, we can determine how much power can reach the mixer which generates minimum distortion. The details can be found in [1].

In Figure 2.14, SHI is the second-order harmonic intercept and TOI is the third-order intercept to 0 dBc. This situation cannot be realized in practice because the mixer would be well into saturation. The 1-dB compression point is normally specified to prevent the saturation condition. Typically, this gain compression occurs at a mixer level in the range of  $-5$  to  $+5$  dBm. The levels of the strong signals are around  $-20$  dBm in Durham area. 20 dB gain may drive these signals to the 1-dB compression level.

## 2.4 Data Acquisition

Data acquisition is the sampling of the real world to generate data that can be manipulated by a computer. In this spectrum occupancy project, the data acquisition hardware included a General Purpose Interface Bus (GPIB) board for controlling the spectrum analyzer and PCI DAS-4020/12 data acquisition card for controlling channel sounder. The data acquisition softwares was programmed in C++ providing instructions to configure each receiver system, execute measurement routines, record measured data, and calibrate the system gains.

The data acquisition programming is a time-consuming work which took the author significant amount of research time on coding and testing the programmes. The data acquisition program for spectrum analyzer was completed by the author independently. Based on the works of Roger Lewenz and Peter Baxendale's, the author achieved the data acquisition program for the sounder system.

### 2.4.1 data acquisition for spectrum analyzer

The data flow diagram of the spectrum analyzer monitoring system is shown in Figure 2.15. The data acquisition and control program is basically three control subroutines that direct operation of multiple subroutine kernels that in turn control every function of the measurement system. RF front-end subroutine handles path selection via parallel port; GPS subroutine collects position data from GPS device via COM port; spectrum analyzer subroutine communicates between PC and spectrum analyzer and store trace data into hard disk via GPIB interface.

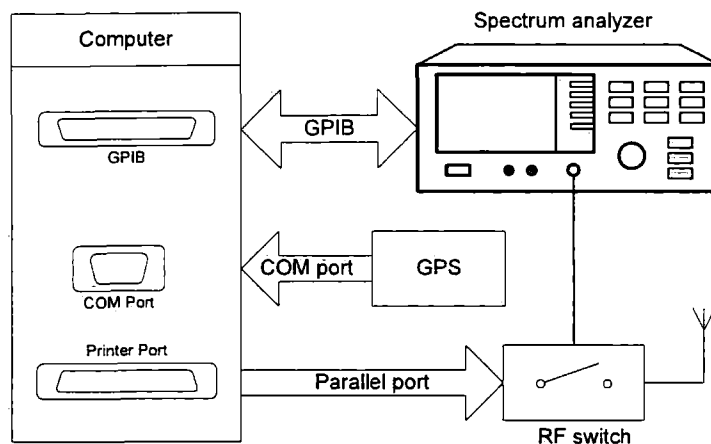


Figure 2.15: Data flow chart of the monitoring system

Object-oriented programming (OOP) is a programming language model organized around "objects" rather than "actions" and data rather than logic. OOP technology was adopted in data acquisition program with C++ [52]. Each subroutine was encapsulated into a class which is defined as abstract characteristics of an object, including the object's attributes, fields or properties and the object's behavior. Object-Oriented Programming has the following advantages over conventional approaches:

- OOP provides a clear modular structure for programs which makes it good for defining abstract data types where implementation details are hidden and the unit has a clearly defined interface.
- OOP makes it easy to maintain and modify existing code as new objects can be created with small differences to existing ones.
- OOP provides a good framework for code libraries where supplied software components can be easily adapted and modified by the programmer.

The GPS receiver sends the GPRMC sentence to computer via COM port. The GPRMC sentence, known as the "Recommended Minimum" sentence, is the most common sentence transmitted by GPS devices. This one sentence contains

nearly everything a GPS application needs: latitude, longitude, speed, bearing, satellite-derived time, fix status and magnetic variation. The format of the GPRMC is as follow:

*\$GPRMC, hhmmss.ss, A, llll.ll, a, yyyyyy.yy, a, x.x, x.x, ddmmyy, x.x, a, m \* hh*

Field	#
(1)	UTC time of fix
(2)	Data status
(3)	Latitude of fix
(4)	N or S of longitude
(5)	Longitude of fix
(6)	E or W of longitude
(7)	Speed over ground in knots
(8)	Track made good in degrees True
(9)	UTC date of fix
(10)	Magnetic variation degrees
(11)	E or W of magnetic variation
(12)	Mode indicator
(13)	Checksum

A COM port is a serial communication physical interface through which information transfers in or out one bit at a time. Many settings are required for serial connections used for asynchronous start-stop communication, to select speed, number of data bits per character, parity, and number of stop bits per character.

Serial ports use two-level signalling, so the data rate in bits per second is equal to the symbol rate in baud. The baud rate was configured to 4800 bits per second [4]. The number of data bits in each character was set to 8 which matches the size of a byte. 8 data bits are almost universally used in newer applications. Parity is a method of detecting some errors in transmission. An extra data bit is sent with each data character, arranged so that the number of 1 bits in each character, including the parity bit, is always odd or always even. If a byte is received with the wrong number of 1 bits, then it must have been corrupted. If parity is correct there may have been no errors or an even number of errors. In our application, the parity bit was set to 'None' without error detection handled by a communication protocol. Stop bits sent at the end of every character allow the receiving signal hardware to detect the end of a character and to synchronize with the character stream. GPS devices in this project use one stop bit.

Register	LPT1	LPT2
Data register	0x378	0x278
Control register	0x379	0x279
Status register	0x37A	0x27A

Table 2.9: Addresses of parallel ports' registers

Windows provides the application programming interface (API) of serial port, which supports requests for services by computer programs. APIs exempt programmers to write low level code to communicate with the serial port. In this project, serial port APIs are integrated into the class GPS.

The 8 way RF switch is controlled by parallel port of computer with TTL level. The parallel port is specifically designed to attach printers with a parallel port interface, but it can be used as a general input/output port for any device or application that matches its input/output capabilities. In this project, only data output function has been used to control RF switch.

The pin out of parallel port is show in Figure 2.16. The lines in DB25 connector are divided into data lines, control lines and status lines [3]. Lines are connected to data, control and status registers internally. So by manipulating these registers in the program, one can easily read or write to parallel port with programming languages. The register addresses are in Table 2.9.

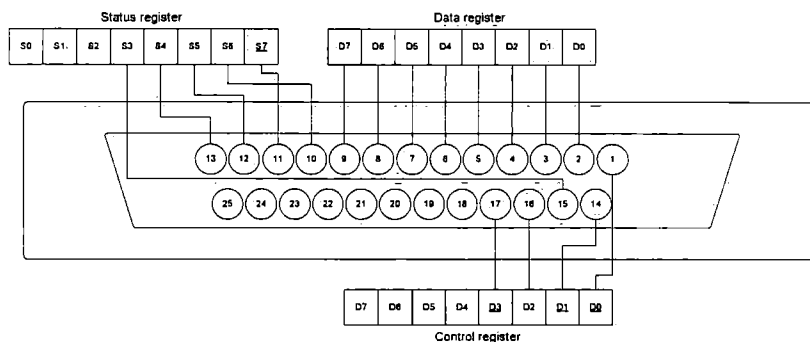


Figure 2.16: The pin outs of the PC parallel port

Direct parallel port controlling from application is not possible under Windows NT based operation systems like WIN2000, WINXP and to be able to control the parallel port directly you will need to write some kind of device driver to do this. To circumvent this situation, the dynamic-link library (DLL) <sup>1</sup> *Inpout32.dll* written by Jan Axelsons was introduced to talk to the parallel port [3].

IEEE-488 known as general purpose interface bus is a short range, digital communications bus specification that was used in the communication between the spectrum analyzer and computer. IEEE-488 allows up to 15 devices to share a single 8-bit parallel electrical bus by daisy chaining connections. The maximum transfer rate of the GPIB board is 300 kbyte/s in our system. The IEEE-488 bus employs 16 signal lines – eight bidirectional used for data transfer, three for handshake, and five for bus management – plus eight ground return lines.

There are 3 types of devices that can be connected to the IEEE-488 bus Listeners, Talkers, and Controllers [15]. A Talker sends data messages to one or more Listeners. The Controller manages the flow of information on the GPIB by sending commands to all devices. Devices can be Listeners, Talkers, and/or Controllers. A spectrum analyzer, for example, is a Talker and may be a Listener as well. It is possible to have several Controllers on the bus but only one may be active at any given time. The Active Controller may pass control to another controller which in turn can pass it back or onto another controller. A Listener is a device that can receive data from the bus when instructed by the controller and a Talker transmits data onto the bus when instructed. The Controller can set up a talker and a group of listeners so that it is possible to send data between groups of devices as well. Since only spectrum analyzer was attached to GPIB board in our application, a GPIB is always Controller-In-Charge or system controller.

The programming of the GPIB library interface is comprised of two different types of routines: higher level Device I/O routines which conceal most of the underlying complexity of GPIB operations and lower level Board I/O routines which take care of all of the low-level details involving GPIB messages and addressing. The board I/O routines are capable of coping with a single device

---

<sup>1</sup>Dynamic-link library is Microsoft's implementation of the shared library concept in the Microsoft Windows and OS/2 operating systems.

application. The communication dataflow between the controller (computer) and device (spectrum analyzer) is shown in Figure 2.17.

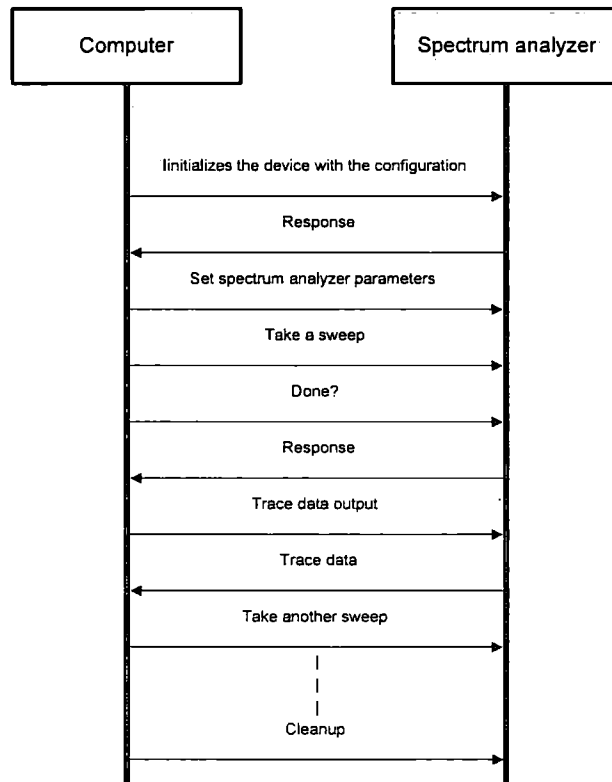


Figure 2.17: Communications between computer and spectrum analyzer

All devices and boards must be assigned unique primary addresses in the range from decimal 0 to 30. The default primary address of all GPIB boards is 0. The primary address of spectrum analyzer was set to 18 by software. In our simple application, the secondary address was set to 0, which is used for increasing the number of addresses that can be supported on a single bus. The timeout value is the approximate minimum length of time that I/O functions can take before a timeout occurs. Since the sweeping time of the spectrum analyzer may take tens seconds, the timeout value was set to 100 s. There are two methods for devices to indicate the last byte in a transfer of multiple bytes. The most common method is for the sending device to assert the EOI line before sending the last byte. This

method is adopted in our application. The second method involves sending an extra End-Of-String (EOS) character to indicate the end of data. A common EOS Byte is the linefeed character (10). Since most devices rely on the EOI line instead, EOS Byte can not be set.

### 2.4.2 data acquisition for channel sounder

The data acquisition system of the channel sounder is shown in Figure 2.18 and consists of 2 PCI-DAS4020/12 high-speed data acquisition boards manufactured by Measurement Computing Corporation and an additional logic board functioning as clock divider and adaptor. It was intended that two data acquisition boards should be used to provide eight channel of data acquisition simultaneously using the functions of the Universal Library<sup>TM</sup> which is the subroutine library that the engineer needs to write the programs for use with any of Measurement Computing's data acquisition and control boards.

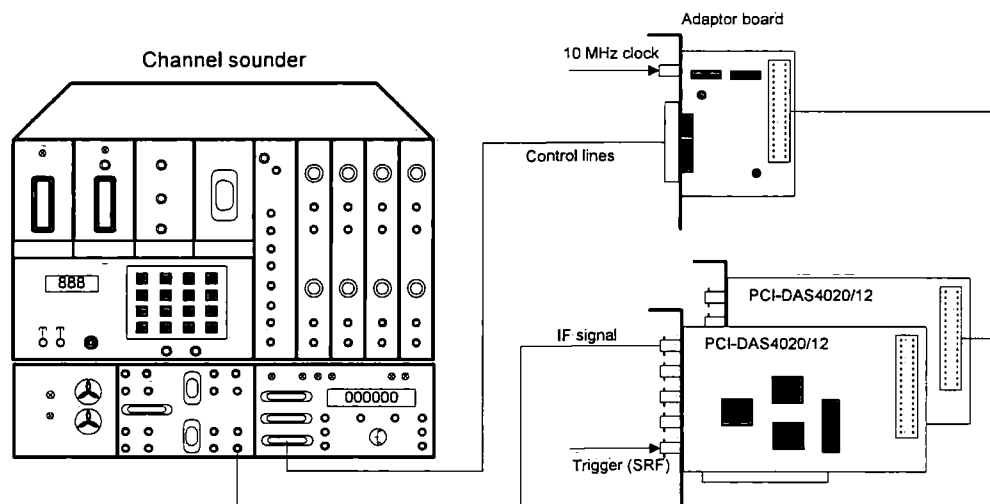


Figure 2.18: Communications between computer and channel sounder

The PCI-DAS4020/12 is a multi-function high-speed data acquisition I/O board that is designed for the PCI bus. It provides the following features [14]:

- 20 MHz sample rate

- Four high-speed 12-bit resolution analog input channels
- 24 digital I/O channels
- Software-selectable input ranges
- Analog and digital triggering

PCI-DAS4020/12 functions are illustrated in the block diagram shown in Figure 2.19. Four single-ended analog input channels connect from BNC connectors to individual amplifiers and then connect to dedicated ADC's. The Trig/Ext Clk BNC can be configured as A/D Start Trigger, A/D Stop Trigger or A/D Pacer Gate. The digital I/O is an 82C55 digital logic device connected with 40-pin connector [14].

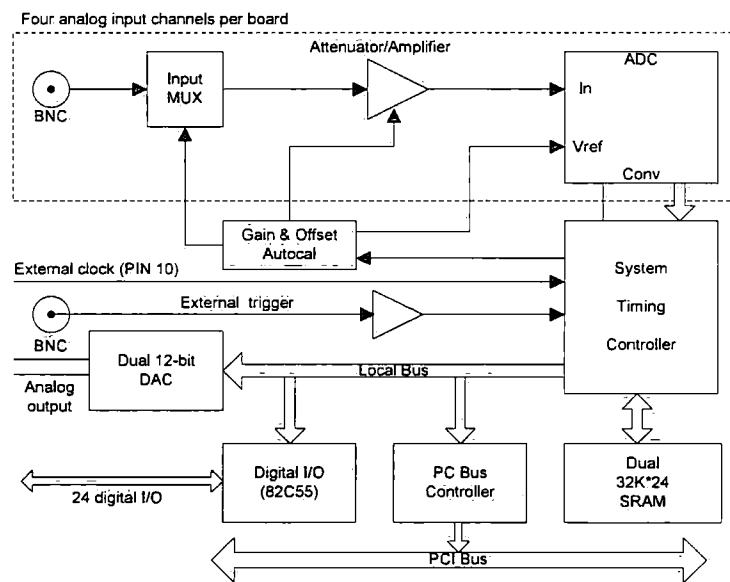


Figure 2.19: The block diagram of the PCI-DAS4020/12 [14]

24 digital I/O channels have been configured as 2 banks of 8 channels, 2 banks of 4 channels. Each bank can be programmed as input or output. Port A was configured as a data port of 8 channels for exchanging data and address between channel sounder and computer. The lower 4 channels of Port C were configured as output port for controlling the sounder via DB25 control register. The signal

## 2.4 Data Acquisition

Pin	Map	Signal name	Pin	Map	Signal name	Pin	Map	Signal name
1		Int in	15		Port B 2	29	24	GND
2		+5 V	16	16	Port C 2	30	5	Port A 3
3		Int enable	17		Port B 1	31		N/C
4		GND	18	18	Port C 1	32	4	Port A 2
5		Port B 7	19		Port B 0	33	25	GND
6		Port C 7	20	1	Port C 0	34	3	Port A 1
7		Port B 6	21	22	GND	35		+5 V
8		Port C 6	22	9	Port A 7	36	2	Port A 0
9		Port B 5	23		N/C	37		GND
10		PC 5 (T/C)	24	8	Port A 6	38		N/C
11		Port B 4	25	23	GND	39		N/C
12		Port C 4	26	7	Port A 5	40		N/C
13		Port B 3	27		N/C			
14	17	Port C 3	28	6	Port A 4			

Table 2.10: Signal mapping on the 40-pin connector and the DB25

mapping between 40-pin connector on the data acquisition card and the DB25 connector to the sounder is shown in Table 2.10. Pin 2 of the 40 pin connector was used to provide +5 V power for the adaptor board in Figure 2.18. Pin 10 on the 40-pin connector was used as external clock input connected to the output of the clock divider in the adaptor board.

As shown in Figure 2.19 the Trig/Ext Clk BNC was configured as an external trigger triggered by the *Start of Sweep* signal from the channel sounder. The analog BNC input range for each channel is software configured to  $\pm 5$  V in our application. At this setting, the 12-bits resolution can reach to  $10/2^{12} = 2.44$  mV.

Besides the function of converting the 40-pin connector to DB25 connector, the adaptor board has a clock divider to convert 10 MHz signal to divide-by-1 and divide-by-2 signals. Although the PCI-DAS4020/12 is capable of 10 MHz sampling rate, the huge amount of data collected at this rate is difficult to process by data processing software such as MATLAB. For example, if each sample has 2

bytes storage, there is around 19 MBytes<sup>1</sup> for 2 boards at 10 MHz sampling rate for one second. The reasonable solution for this is using divide-by-2 to 5 MHz clock to downsample the signal.

In channel sounding applications, the PCI-DAS4020/12 should sample the channels at the external clock and the external trigger signal to keep synchronization with the sounder. Under the EXTCLOCK and EXTTRIGGER option, the clock signal presented to pin 10 is divided by 4 in four channel mode. This limitation of PCI-DAS4020/12 means that the channels have been sampled sequentially. On the other hand, it is necessary for applications such as channel sounding that each *Start of Sweep* which contain a number of clocks starts at the same time. Otherwise, the pseudo doppler shift will be introduced by [21; 33]. All of these require that the clock number in one sweep should be an integer times 4. Figure 2.20 shows the relation between SRF and clock, where clock A is feasible for our application. Fortunately, the clock number per sweep is 40024 in the sounder system, which means that the base clock and divide-by-2 clock are still an integer times of 4.

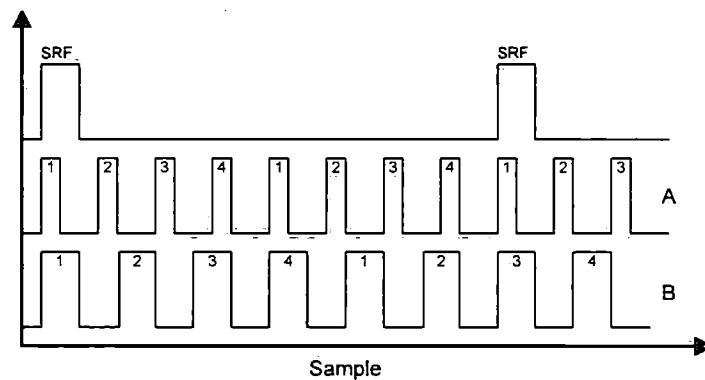


Figure 2.20: The relation between SRF and clock

In high speed sample rate application, in order to achieve the maximum sample rate under some conditions, a contiguous area of memory must be set up using *InstaCal*. *InstaCal* is a comprehensive software program that manages everything about data acquisition hardware. The contiguous memory allocated in

---

<sup>1</sup>1 MBytes = 1024\*1024 Bytes

*InstaCal* is an area of contiguous memory which is set aside for use by Universal Library <sup>TM</sup> scan functions that require an allocation of memory. Since each sample needs 2 byte memory for storage, the maximum capacity of the contiguous memory is 256 MBytes for 134, 217, 728 samples. This is hard coded [14] and it could be a restriction of Windows. Therefore, if we only increase the capacity of memory, there will not be significant improvements. Furthermore, the virtual memory technique of Windows' makes data acquisition more difficult. Since the contiguous memory allocated in *InstaCal* is the total memory that is available to all boards installed, in our application, 2 PCI-DAS4020/12 cards shared the total contiguous memory and each can get the maximum 128 MBytes memory.

256 MBytes contiguous memory is sufficient for most applications, But to certain application, it restricts the long time data acquisition. To remove the physical 256 MBytes memory limitations, higher specification hardware and advanced algorithm have to be found to solve this problem. A new methodology was adopted in our data acquisition which transfers the acquired data into hard disk in real time. In order to achieve this objective, a 7200 rpm Serial Advanced Technology Attachment (SATA)<sup>1</sup> hard Disk was deployed to increase physical throughput which can transfer data at a rate of 1.5 gigabits per second.

The Producer-consumer algorithm in Figure 2.21 was introduced to solve high speed transfer and multi-process synchronization problems. The algorithm describes two processes, the producer and the consumer, who share a common, fixed-size buffer. The producer's job is to generate a piece of data, put it into the buffer and start again. At the same time the consumer is consuming the data (i.e. removing it from the buffer) one piece at a time. The algorithm is to make sure that the producer won't try to add data into the buffer if it's full and that the consumer won't try to remove data from an empty buffer. The solution for the consumer is to go to sleep if the buffer is empty. The next time the producer puts data into the buffer, it wakes up the sleeping consumer. In the same way the solution for the producer is to go to sleep if the buffer is full. The next time the

---

<sup>1</sup>Serial Advanced Technology Attachment is a computer bus primarily designed for transfer of data between a computer and mass storage devices such as hard disk drives and optical drives.

consumer removes an item from the buffer, it wakes up the producer who starts to fill the buffer again.

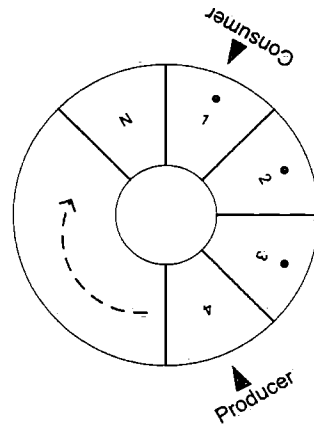


Figure 2.21: Producer-consumer algorithm

In the case of our application, the 128 MBytes contiguous memory for each board was configured as a circular buffer referring to Figure 2.21. The data collecting process is modeled as a *Producer* who scans a range of A/D channels and stores the data into the circular buffer in order. The *Producer* algorithm is as follow:

```

Function Producer();
if (position of Producer + 1 < position of Consumer) %N then
|   Producer waiting;
else
|   (position of Producer ++) %N;
end
    
```

where % stands for modulus. The circular buffer is conceptual memory by which the data are being written in circular way. The data transferring process is modeled as a *Consumer* who transfers a batch of data from the circular buffer

into the hard disk. The *Consumer* algorithm is as follows:

```

Function Consumer();
if (position of Producer = position of Consumer) %N then
  | Consumer waiting;
else
  | (position of Consumer ++) %N;
end

```

In our application, we can not control the speed of the *Producer* and can only control the speed of the *Consumer*. An *overwritten* might happen occasionally depending on Windows' schedule. Because of this reason, when an *overwritten* has been detected, an Error will be reported and the data acquisition program will be suspended. Another important issue is the size of the block memory in Figure 2.21. The efficiency of transferring data from the contiguous memory to disk decides the size of the chunk. Too small size costs the heavy signalling traffic in PCI bus and diminishes the throughput; too big chunk size will cause the *overwritten*. An empirical value 2 MBytes was chosen which caused the minimum overwritten rate.

In practice, the situation is complicated when 2 boards work simultaneously to sample 8 channels. These boards must work on *background* mode. If the *background* option is not used then the control will not return to the program until all of the requested data has been collected and returned to the buffer. When the *background* option is used, control will return immediately to the next line in the program and the data collection from the A/D into the buffer will continue in the background. After the program gets control, it generates a *Consumer* process to transfer the acquired data into disk. In the 2 board situation, attention should be paid on synchronicity between two boards since each board may work properly. The flowchart of the data acquisition routine is shown in Figure 2.22.

Communications between computer and channel sounder are via the PCI-DAS4020/12 digital I/O port. The signal path is shown in Figure 2.18. At the channel sounder end, the data, address and command lines are shown in Figure 2.16. At the computer end, the Port A of 40-pin digital I/O port was configured as data and address line and the lower 4 channels of Port C was configured as control lines. Since pin 1, 14 and 17 of DB25 have inverted logic, the commands

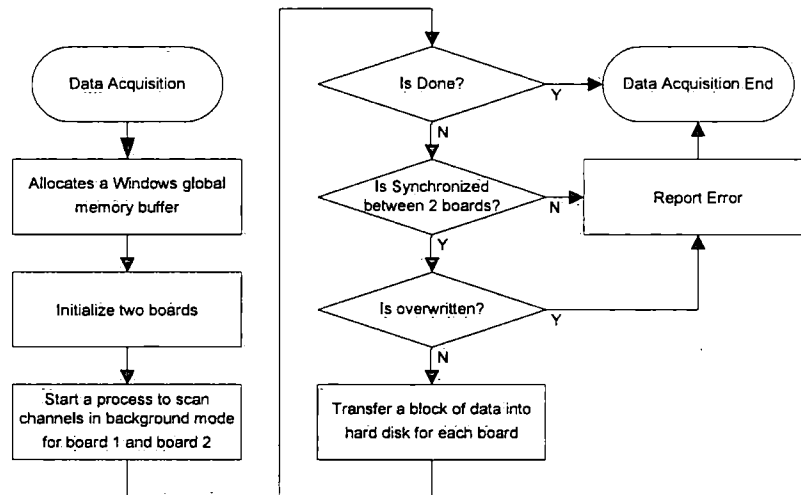


Figure 2.22: The flowchart of the data acquisition routine

must be coded correctly before sending to the sounder as shown in Table 2.11. The details can be referred to in [21; 33].

## 2.5 Summary

This chapter focuses on the engineering aspects of the spectrum occupancy project. We start with considerations for spectrum monitoring in view of the spectrum management. Then, we go into technique details of the spectrum monitoring. Justifications were given for reasons why the specific kinds of antennas and radio receivers and their parameter settings were chosen. We discussed the preamplifier technique regarding the system sensitivity and the filtering technique regarding the radio distortion. In addition, the data acquisition systems and the specific data acquisition algorithm were introduced in our project to accelerate data transfer.

Code	Value	Signal name	Code	Value	Signal name
0000 <sub>(0)</sub>	1101 <sub>(11)</sub>	INT0	1000 <sub>(8)</sub>	0011 <sub>(3)</sub>	Rx0
0001 <sub>(1)</sub>	1010 <sub>(10)</sub>	INT1	1001 <sub>(9)</sub>	0010 <sub>(2)</sub>	Rx1
0010 <sub>(2)</sub>	1001 <sub>(9)</sub>	SCWS	1010 <sub>(10)</sub>	0001 <sub>(1)</sub>	Rx2
0011 <sub>(3)</sub>	1000 <sub>(8)</sub>	S0	1011 <sub>(11)</sub>	0000 <sub>(0)</sub>	START
0100 <sub>(4)</sub>	1111 <sub>(15)</sub>	S1	1100 <sub>(12)</sub>	0111 <sub>(7)</sub>	N0
0101 <sub>(5)</sub>	1110 <sub>(14)</sub>	S2	1101 <sub>(13)</sub>	0110 <sub>(6)</sub>	N1
0110 <sub>(6)</sub>	1101 <sub>(13)</sub>	S3	1110 <sub>(14)</sub>	0101 <sub>(5)</sub>	N2
0111 <sub>(7)</sub>	1100 <sub>(12)</sub>	W	1111 <sub>(15)</sub>	0100 <sub>(4)</sub>	NOP

Table 2.11: Control commands for channel sounder

# Chapter 3

## Parameter Estimations

### 3.1 Direction of Arrival Estimation

Cognitive radio is an intelligent radio system, which is capable of being sense of the communication environment and consequently changes its transmission or reception parameters. Based on the spectrum share conception, it is essential for cognitive radio to recognize spectrum occupancy in particular frequency band, a particular time slot and a particular direction to avoid interferences. Spectrum occupancy in the frequency and time domains will be discuss in the next chapter. This chapter focus on the detection techniques for radio arriving directions, which might be helpful to understand the spectrum occupancy in the space domain.

Radio direction of arrival estimation is the technique for determining the direction of a radio transmission. Radio direction finding is used by spectrum engineers to locate the source of radio frequency interference. With triangulation techniques radio direction finding can even determine the location of a radio transmission.

There are two common technical approaches to radio direction of arrival estimation. The first approach involves the use of directional antennas which are designed to be more sensitive to signals received in some directions rather than in others. As the antenna is turned in various directions, a signal being received will either increase or decrease in strength. All other things being equal, the direction in which the signal is strongest is the likely direction in which the radio transmitter is located. The movement of the antenna and the determination of the peak

signal strength can be made by a human operator or can be done automatically by electronics.

The second approach exploits the effects of phase shift. A fixed antenna array is deployed in a precise geometric pattern. By computing the amount of phase shift present on the signal from antenna to antenna, a direction to the signal source can be computed. This approach will be discussed in detail in this chapter.

### 3.1.1 wave and sensor array

#### 3.1.1.1 wave propagation

In array signal processing, signals propagating from the source to the receiving array are functions of position as well as time and have properties governed by the laws of the wave equation (3.1) [46].

$$\nabla^2 E - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} E = 0 \quad (3.1)$$

The wave equation is a second-order linear partial differential equation that describes the propagation of EM waves, sound and water waves as well. In the equation (3.1)  $E$  is the electric field and  $\nabla^2$  is the Laplace operator

$$\nabla^2 E = \frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2}$$

A solution can be found to the wave equation as [27]

$$s(\vec{x}, t) = A \exp \left\{ j \left( \omega t - \vec{k} \cdot \vec{x} \right) \right\} \quad (3.2)$$

where  $\vec{x}$  is the position vector  $[x \ y \ z]^T$  and  $\vec{k}$  is the *wavenumber* vector  $[k_x \ k_y \ k_z]$ , such that  $k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c^2}$ . The equation (3.2) represents a monochromatic plane wave solution because the value  $s(\vec{x}, t)$  is the same at all points lying in a plane given by  $k_x x + k_y y + k_z z = C$ , where  $C$  is a constant. In the physics of electromagnetic wave propagation, a plane wave is a constant frequency wave whose wavefronts are infinite parallel planes of constant amplitude normal to the propagation direction. In practice, the plane wave can be approximated in the far-field region of the radio source.

### 3.1 Direction of Arrival Estimation

---

The wavenumber is defined as  $\vec{k} = 2\pi/\vec{\lambda}$ , where  $\vec{\lambda}$  is the wavelength vector. The direction of the *wavenumber* vector  $\vec{k}$  represents the direction of propagation of the plane wave; the magnitude  $|\vec{k}|$  represents the number of cycles in radians per meter of length that the monochromatic plane wave exhibits in the direction of propagation. The *wavenumber* vector  $\vec{k}$  can be considered as a spatial frequency just as  $\omega$  is a temporal frequency. Instead of a scale value, the spatial frequency content of waves must be represented as a three dimensional vector. For simplicity, we define another vector  $\vec{\alpha} = \vec{k}/\omega$ , called the *slowness* vector, whose magnitude equals to the reciprocal of the propagation speed.

Reference [27] gives the summary of the relations which governs the propagation of plane waves

Propagation of plane wave	$s(t - \vec{\alpha} \cdot \vec{x})$	(3.3)
Propagation of sinusoidal plane wave	$\sin(\omega t - \vec{k} \cdot \vec{x})$	
Slowness vector	$\vec{\alpha} = \vec{k}/\omega,  \vec{\alpha}  = 1/c$	
Wavenumber vector	$\vec{k} = \omega\vec{\alpha},  \vec{k}  = 2\pi/\lambda$	
Frequency and wavelength	$c = \lambda \cdot \omega/2\pi$	

The Fourier technique has proved to be an enormously useful tool for simplifying and analyzing DOA algorithms. The concept of Fourier analysis can be straightforwardly extended to multidimensional signals. The transform and the inverse transform between *Position-Time* space and *Wavenumber-Frequency* space are defined in (3.4) [27].

$$S(\vec{k}, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(\vec{x}, t) \exp\{-j(\omega t - \vec{k} \cdot \vec{x})\} d\vec{x} dt$$

$$s(\vec{x}, t) = \frac{1}{(2\pi)^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(\vec{k}, \omega) \exp\{j(\omega t - \vec{k} \cdot \vec{x})\} d\vec{k} d\omega$$
(3.4)

For example, when *space-time* signal is a propagating wave,  $s(\vec{k}, \omega) = s(t - \vec{\alpha}_0 \cdot \vec{x})$ , the *Wavenumber-Frequency* spectrum is

$$S(\vec{k}, \omega) = S(\omega) \delta(\vec{k} - \omega\vec{\alpha}_0)$$
(3.5)

where  $\delta(\cdot)$  is an impulse function. Thus, this waveform contains energy only along the direction  $\vec{k} - \omega\vec{\alpha}_0$  in *Wavenumber-Frequency* space. The *Wavenumber-*

### 3.1 Direction of Arrival Estimation

---

*Frequency* spectrum of the monochromatic plane wave signal (3.2) can be expressed as

$$S(\vec{k}, \omega) = A \cdot \delta(\vec{k} - k_0) \delta(\omega - \omega_0)$$

Then, the monochromatic plane wave signal corresponds to a single point in *Wavenumber-Frequency* space, which dictates the temporal frequency and the spatial wavelength.

#### 3.1.1.2 uniform linear array

Any *transducer*, which interact with propagating wave to produce an electrical signal, has finite physical dimension. The effects of a finite aperture on a space-time signal will be illustrated in one dimension aperture shown in Figure 3.1. First, we define the aperture function [56] as

$$w(\vec{x}) = \begin{cases} 1 & : |\vec{x}| \leq D/2 \\ 0 & : else \end{cases}$$

where  $D$  is the length of sensor. This aperture function acts like a window function through which the wave has been observed. Through this finite aperture, the observed waveform can be expressed as

$$z(\vec{x}, t) = w(\vec{x}) \cdot s(\vec{x}, t)$$

By the convolution theorem, we obtain a transform in *Wavenumber-Frequency* space

$$Z(\vec{k}, \omega) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} W(\vec{k} - \vec{l}) S(\vec{l}, \omega) d\vec{l} \quad (3.6)$$

where

$$W(\vec{k}) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} w(\vec{x}) \exp\{j\vec{k} \cdot \vec{x}\} d\vec{x} = \frac{\sin k_x D/2}{k_x/2}$$

and  $S(\vec{l}, \omega)$  is the *Wavenumber-Frequency* spectrum of the propagating wave. This convolution means that the waveform's spectrum is smoothed by the kernel function  $W(\vec{k})$  once we observe it from a finite aperture.

For a propagating plane wave, by substituting (3.4) into (3.6) we can obtain

$$Z(\vec{k}, \omega) = S(\omega) W(\vec{k} - \omega \vec{\alpha}_0)$$

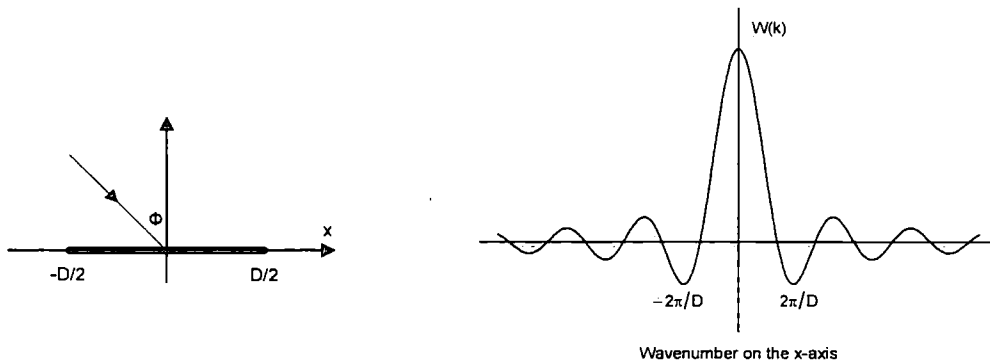


Figure 3.1: The linear aperture and its aperture smoothing function

Along the direction  $\vec{k} = \omega \vec{\alpha}_0$  in *Wavenumber-Frequency* space, the output spectrum equals the signal spectrum multiplied by a constant  $W(0)$ , which suggests that all information concerning the propagating signal is present in the aperture's output.

Figure 3.1 shows that the spatial extent of an aperture determines the resolution. A large spatial extent has narrower aperture smoothing function, which leads to minimal spectral smoothing. The larger the extent, the more focused the aperture can be on a specific direction. The Rayleigh criterion [30] states that two plain waves with different propagating directions are resolved if the mainlobe peak of one aperture smoothing function replica falls on the zero of the other's replica. So, the resolution equals the smallest wavenumber that produces a zero aperture smoothing function. The resolution is  $k_x = 2\pi/D$  for the linear aperture.

The linear aperture was introduced to illustrate the spatial smoothing effect of finite dimension. In practice, instead of a continuous aperture an antenna array is employed in discrete spatial sampling, which is rather similar with discrete signal processing in time domain. A 4 element linear array which is relatively simple to analyze and permits the use of fast algorithms, is shown in Figure 3.2, where the spatial interval between elements is  $d$ .

In addition to the spatial smoothing effect of the spatial window, discrete spatial sampling can introduce aliasing. For mathematical simplicity, define a set

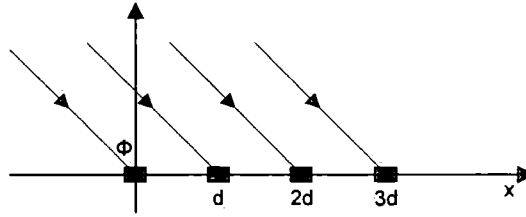


Figure 3.2: Linear array

of signals  $\{\hat{z}_m(t) = w_m s(t), m \in \mathbb{Z}\}$ , where  $\mathbb{Z}$  is *Integers* and

$$w_m = \begin{cases} 1 & : \text{at sensor's location} \\ 0 & : \text{elsewhere} \end{cases}$$

This notation conveniently allows to incorporate sensor weights into the equation for advanced algorithms, which are not discussed in this thesis. Then, the *Wavenumber-Frequency* representation (3.4) is modified as [27]

$$\hat{Z}(k, \omega) = \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \hat{z}_m(t) \exp\{-j(\omega t - kmd)\} dt$$

This *Wavenumber-Frequency* spectrum is to be a circular convolution between the *Wavenumber-Frequency* spectrum of the propagating wave  $S(k, \omega)$  and the Fourier transform of the discrete aperture function  $w_m$

$$\hat{Z}(k, \omega) = \frac{d}{2\pi} \int_{-\pi/d}^{\pi/d} S(l, \omega) W(k - l) dl$$

The *array pattern* is defined as

$$W(k) = \sum_m w_m \exp\{jkmd\}$$

and in the case of the uniform linear array which equals [27]

$$W(k) = \frac{\sin \frac{kMd}{2}}{\sin \frac{kd}{2}} \tag{3.7}$$

where  $M$  is the number of sensors shown in Figure 3.3.

Figure (3.7) shows that, unlike the case of a continuous linear aperture, the *array pattern* of the linear array is a periodic function of  $k$  with the period  $2\pi/d$ .

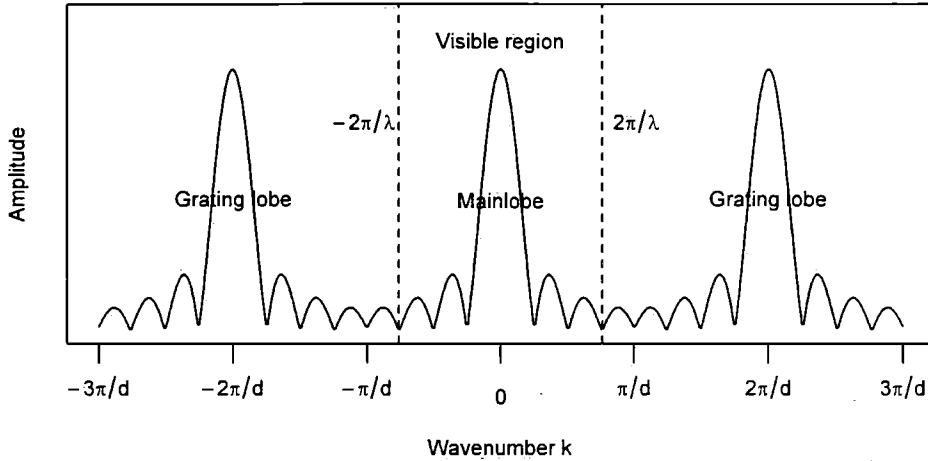


Figure 3.3: The grating lobe of the linear array

Each period of  $W(k)$  consists of a mainlobe which has largest amplitude and a number of sidelobes. The first zero of  $W(k)$  occurs when the argument of the numerator's sine function equals  $k = 2\pi/Md$ . The mainlobe width is  $k = 4\pi/Md$  which is closely correlated with the resolution of the array. The mainlobe width decreases as the number of sensors increases or the sensor interval  $d$  increases. Either action serves to increase the array's resolution.

A grating lobe is defined by [38] as an unwanted peak value in the discrete array shown in Figures 3.3. Signals propagating from directions corresponding to spatial frequencies at which grating lobes occur would be indistinguishable from signals propagating from the main lobe. The *visible region* is the range of real angles of incidence wave for given wavelength shown in Figure 3.3. For example, consider the case of a linear array on which a monochromatic plane wave with wavelength  $\lambda$  is impinging at incidence angle  $\phi$  shown in Figure 3.2. Because  $k_x = 2\pi \sin \phi / \lambda$  and  $|\sin \phi| \leq 1$ , the *visible region* is the region where  $k_x$  takes on real values only between  $\pm 2\pi/\lambda$ . Because  $W(k)$  in (3.7) is periodic with period  $2\pi/d$ , inadequate spatial sampling causes grating lobes entering the *Visible region*. If  $\lambda \leq 2d$ , the visible region occupies more than one period of the *array-pattern*  $W(k)$ ; indicating that spatial aliasing has occurred.

The Nyquist sampling [27] theorem in time domain can be extended to space

## 3.1 Direction of Arrival Estimation

domain without any difficulty. It states that if a continuous variable signal is bandlimited to frequencies below  $k_0$ , then it can be periodically sampled without loss of information so long as the sampling period  $d \leq \pi/k_0$ . Figure 3.4 illustrates the Nyquist sampling theorem in the wavenumber domain. Figure 3.4(a) represents a bandlimited *Wavenumber-Frequency* Fourier transform where the sample interval in space satisfies  $d \leq \pi/k_0$ . In the wavenumber domain, the transform reserves the whole information of the original waveform. Consequently, it can be recovered from the finite linear array with an ideal lowpass filter. If the Nyquist sampling theorem does not hold as shown in Figure 3.4(b), the wavenumber information distorted by each other referred to as *aliasing* and the original information is no longer reconstructed.

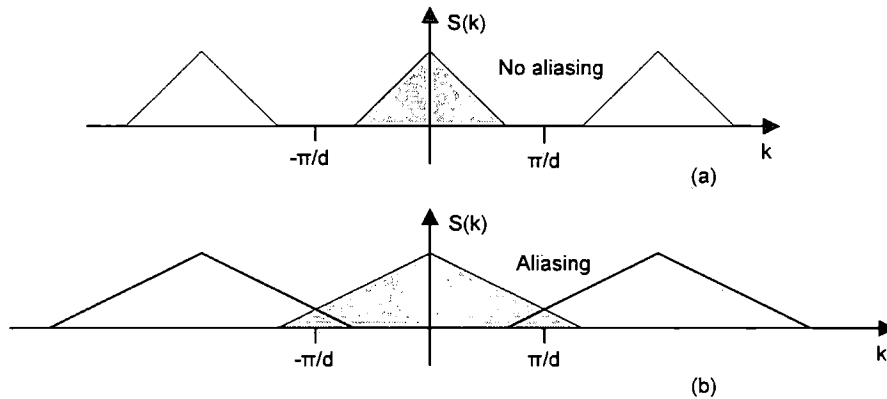


Figure 3.4: Aliasing

### 3.1.2 parametric data model

Most modern signal processing approaches are model based, in the sense that they rely on certain assumptions made on the observed data. The monochromatic plane waves discussed in the previous section are deterministic so that each value of the signal is fixed and can be determined by Maxwell's equations as a function of space and time. On the other hand, since radio noise cannot be characterized by a simple, well defined mathematical equation. Their future values cannot be predicted exactly. Rather, theory of statistics and random processes

made up of the mathematical foundation of modeling random signals [23]. The deterministic model was chosen for our experiments. In this model the receiving signals are assumed to be the deterministic signals plus random radio noise. Although deterministic, signals are unknown and quantities such as waveform and parameters must be estimated.

Radio noise does play an important role in this models in determining the characteristics of propagating waveforms. Usually, we adopt the random noise as 0-mean circular symmetric<sup>1</sup> complex Gaussian random vector with unknown covariance, and successive samples are independent but share a common density. Let  $n(\vec{x}, t)$  denote a random field and its probability density function is  $p_{n(\vec{x}, t)}(\gamma)$ , the expectation, correlation function and covariance function of a random field are defined as [29]

$$E[n(\vec{x}, t)] \triangleq \int \gamma \cdot p_{n(\vec{x}, t)}(\gamma) d\gamma = \mu_n$$

$$R_{n_0, n_1}(\vec{x}_0, \vec{x}_1, t_0, t_1) \triangleq \int \int \gamma \lambda \cdot p_{n_0, n_1}(\gamma, \lambda) d\gamma d\lambda$$

$$C_{n_0, n_1}(\vec{x}_0, \vec{x}_1, t_0, t_1) \triangleq \int \int (\gamma - \mu_{n_0})(\lambda - \mu_{n_1}) \cdot p_{n_0, n_1}(\gamma, \lambda) d\gamma d\lambda$$

An important case of random fields is the stationary random field where correlation function does not vary with absolute position and time. The correlation function can be represented as [23]

$$R_{n_0, n_1}(\vec{\chi}, \tau) = E[n_0(\vec{x}_0, t_0) n_1(\vec{x}_0 + \vec{\chi}, t_0 + \tau)] \quad (3.8)$$

where  $\vec{\chi} = \vec{x}_1 - \vec{x}_0$  denotes space difference and  $\vec{\tau}$  denotes time lag. Equation (3.8) shows that the correlation function of a stationary field depends only on the space and time difference, never on absolute position and time. For array signal processing, stationarity means that the noise portion of the observations recorded at each sensor have identical statistical characteristics and any cross correlations between sensor outputs depends only on the physical distance separating the sensors.

---

<sup>1</sup>A complex-valued random variable  $Z = X + jY$  is a circular symmetric complex Gaussian variable, or it follows complex Gaussian distribution, if its real and imaginary parts,  $X$  and  $Y$ , are jointly Gaussian, independent, and they have the same variance of  $\sigma^2$ .

### 3.1 Direction of Arrival Estimation

---

The power spectral density function of a stationary random field is defined as the Fourier transform of the correlation function by

$$S_n(\vec{k}, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_n(\vec{\chi}, \tau) \exp\{-j(\omega\tau - \vec{k} \cdot \vec{\chi})\} d\vec{\chi} d\tau$$

The power spectral density function of a *spherically isotropic noise* field would have the form [27]

$$S_n(\vec{k}, \omega) = G(\omega) \delta(|\vec{k}| - \omega/c)$$

$G(\omega)$  represents the distribution of power with respect to temporal frequency. The  $\delta(\cdot)$  represents the propagation by relating the magnitude of the wavenumber vector and temporal frequency. The correlation function  $R_n(\vec{\chi}, \tau)$  of *spherically isotropic noise* can be evaluated by (3.4) and obtained as

$$R_n(\vec{\chi}, \tau) = 0, \quad |\vec{\chi}| = n \frac{\lambda}{2}, \quad n \neq 0$$

Thus, in the monochromatic plane wave circumstance, sampling at spatial intervals separated by  $\lambda_0/2$  would yield uncorrelated noise components. Especially, when *spherically isotropic noise* is Gaussian, such spatial sampling interval would yield statistically independent waveforms [27].

The signal model which we adopt is that observations  $\{y_m(\vec{x}_m, t)\}_{m=1}^M$ , where  $M$  is the number of sensors, made at sensor outputs consist of an additive combination of signals and noise. The continuous time response of the  $i^{th}$  sensor is  $y_i(\vec{x}_i, t)$  and the  $M \times 1$  vector of such responses is denoted

$$\vec{y}(t) = [y_1(\vec{x}_1, t) \ y_2(\vec{x}_2, t) \ \cdots \ y_M(\vec{x}_M, t)]^T$$

where  $T$  denotes transpose.  $\vec{y}(t)$  is composed of the superposition of the responses to  $P$  signals with coherent wavefronts, plus white noise. Further define  $a_i(\theta_l)$  as the gain of  $i^{th}$  sensor to a signal from direction  $\theta_l$ . This implies that the response of the  $i^{th}$  sensor is [31]

$$y_i(\vec{x}_i, t) = \sum_{l=1}^P a_i(\theta_l) s_l(\vec{x}_i, t) + n_i(t) \quad i = 1, 2, \dots, M \quad (3.9)$$

The basic narrow band assumption is that signal amplitude does not change appreciably during the period of time required to transit the array, or put in

another way, at any instant of time the signal is perfectly coherent along the array [31]. For  $B \ll c/|\vec{x}|$ , where  $B$  is the bandwidth of  $s(t)$ , equation (??) can be written as

$$s(t - \vec{\alpha} \cdot \vec{x}) \exp \{j\omega(t - \vec{\alpha} \cdot \vec{x})\} = s(t) \exp \{j\omega(t - \vec{\alpha} \cdot \vec{x})\}$$

If we define  $\tau_i(\theta_l) = \vec{\alpha}_l \cdot \vec{x}$  the time delay experienced by signal  $l$  arriving from direction  $\theta_l$  as measured at sensor  $i$ . Equation (3.9) can be expressed as an analytic signal<sup>1</sup> as [31].

$$y_i(\vec{x}_i, t) = \sum_{l=1}^P a_l(\theta_l) s_l(t) \exp \{-j\omega\tau_i(\theta_l)\} + n_i(t) \quad i = 1, 2, \dots, M$$

This equation can be made more compact with matrix notation by defining

$$\vec{\theta} = [\theta_1 \ \theta_2 \ \dots \ \theta_P]_{P \times 1}^T$$

$$\vec{s}(t) = [s_1(t) \ s_2(t) \ \dots \ s_P(t)]_{P \times 1}^T$$

$$\vec{d}(\theta) = [a_1(\theta) e^{-j\omega\tau_1(\theta)} \ a_2(\theta) e^{-j\omega\tau_2(\theta)} \ \dots \ a_M(\theta) e^{-j\omega\tau_M(\theta)}]_{M \times 1}^T$$

and

$$D(\vec{\theta}) = [d(\theta_1) \ d(\theta_2) \ \dots \ d(\theta_P)]_{M \times P}$$

$\vec{d}(\theta)$  is referred to as a steering vector, and the set of all such steering vectors, parameterized by  $\theta$ , is called the *array manifold*. It is assumed that  $\vec{d}(\theta)$  is a fixed and known function for all  $\theta$ .

The basic parametric data model for DOA can be expressed as

$$\vec{y}(t) = D(\vec{\theta})\vec{s}(t) + \vec{n}(t) \tag{3.10}$$

where  $\vec{n}(t) = [n_1(t) \ n_2(t) \ \dots \ n_M(t)]_{M \times 1}^T$ . Until now, for simplifying conception, we took the time variable as a continuous variable. In practice, we sample sequences of sensor outputs simultaneously and with a fixed time interval. The data consists of a set of discrete samples  $\{\vec{y}(n); 1 \leq n \leq N\}$ . These vector samples are referred to as *snapshots*.

---

<sup>1</sup>In practice the signal is usually down-converted to baseband before sampling. So we drop the carrier term  $\exp \{j\omega t\}$  for convenience.

### 3.1 Direction of Arrival Estimation

---

Without loss of generality, take a uniform linear array shown in Figure 3.5 as an example to derive the steering vector. The direction perpendicular to the array is called the *broadside direction* or simply the broadside of the array. All DOA's will be measured with respect to this direction, assuming a far field plane wave has incidence angle  $\theta$  with the broadside.

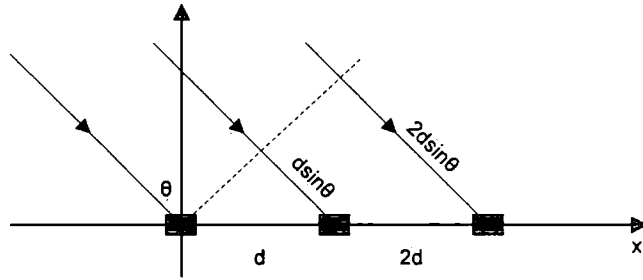


Figure 3.5: Calculation the steering vector of the uniform linear array

Under the narrow band assumption, DOA algorithms use the phase information present in the array sensors to estimate the incidence angle. We define the sensor at the origin as the reference sensor. The wavefront presented in sensor 2 has to travel an extra distance of  $d \sin \theta$  as compared to the signal incident on sensor 1 and an extra distance of  $2d \sin \theta$  for traveling to sensor 3. In other words, the signal incident on sensor 2 is a time-delayed version of the wavefront incident on sensor 1 with the delay being  $d \sin \theta / c$ , where  $c$  is the speed of light [19].

$$\begin{aligned} \vec{d}(\theta) &= [a_1(\theta) e^0 \quad a_2(\theta) e^{-j\frac{\omega}{c} d \sin \theta} \dots a_M(\theta) e^{-j\frac{\omega}{c} (M-1) d \sin \theta}] \\ &= [a_1(\theta) \quad a_2(\theta) e^{-jk_x d \sin \theta} \dots a_M(\theta) e^{-jk_x (M-1) d \sin \theta}] \end{aligned}$$

Many array processing algorithms depend on the measured field via the correlation function computed from the sensor outputs. It needs a *spatiotemporal correlation matrix* to completely depict the space and time relation which has  $(MN)^2$  values and each entry is defined as in (3.8), where  $M$  denotes the number of sensors and  $N$  denotes the sample number. If we assume that each *snapshot* is statistically independent and noise is *isotropic white noise*, only the *spatial covariance matrix* can present the whole correlation between sensor signals. The

*spatial covariance matrix* is defined as

$$E \{ \vec{y}(t) \vec{y}^H(t) \} = D(\vec{\theta}) E \{ \vec{s}(t) \vec{s}^H(t) \} D^H(\vec{\theta}) + E \{ \vec{n}(t) \vec{n}^H(t) \}$$

where  $H$  is the *Hermitian transpose*.  $E \{ \vec{s}(t) \vec{s}^H(t) \} = B$  is the source covariance matrix and  $E \{ \vec{n}(t) \vec{n}^H(t) \} = \sigma^2 I$  is the noise covariance matrix, where  $I$  is the identity matrix, .

## 3.2 EM and SAGE Algorithm

### 3.2.1 maximum likelihood estimation

Maximum likelihood estimation (MLE) [47] is a statistical method used for fitting a mathematical model to data. Modeling real world data by estimating maximum likelihood offers a way of tuning the free parameters of the model to provide a good fit.

In order to introduce the maximum likelihood estimation, we shall first define the *likelihood function*. The *likelihood function*  $l(\theta; x_1, x_2, \dots, x_N)$  of  $N$  random variables  $X_1, X_2, \dots, X_N$  is defined to be the joint density of the  $N$  random variables  $f_{X_1, X_2, \dots, X_N}(x_1, x_2, \dots, x_N)$ , which is considered to be a function of  $\theta$ . In particular, if  $X_1, X_2, \dots, X_N$  is a independent sample from the density  $f(x; \theta)$ , then the likelihood function is  $f(x_1, \theta) f(x_2, \theta) \dots f(x_N, \theta)$ .

The *likelihood function*  $l(\theta; x_1, x_2, \dots, x_N)$  gives the *likelihood* that the random variables assume a particular value  $x_1, x_2, \dots, x_N$ . The *likelihood* is the value of a density function; so for discrete random variables it is a probability. In the *estimation* circumstance, we want to know what value of  $\hat{\theta}$  is the largest *likelihood* from the set of observed data  $x_1, x_2, \dots, x_N$ .

We formalize the definition of a maximum likelihood estimator [23]. Let  $l(\theta; x_1, x_2, \dots, x_N)$  be the *likelihood function* for the random variables  $X_1, X_2, \dots, X_N$ . If  $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_N)$  is the value of  $\theta$  which maximizes  $l(\theta; x_1, x_2, \dots, x_N)$ , then  $\hat{\theta}$  is the maximum likelihood estimation of  $\theta$ . Since many of the density functions are exponential in nature, it is therefore easier to compute the MLE of a *likelihood function* by finding the maximum of the natural log of  $l(\theta; x_1, x_2, \dots, x_N)$ , known

as the *log likelihood*:

$$\bar{L}(\theta; x_1, x_2, \dots, x_N) = \log(l(\theta; x_1, x_2, \dots, x_N))$$

due to the monotonicity of the log function.

In DOA applications, if we assume that each snapshot is statistically independent and noise is Gaussian noise, the *likelihood function* for model (3.9) can be written as [27]

$$l(\vec{\theta}, \vec{s}; Y) = \sigma^{-NM} (\det I)^{-N} \cdot \exp \left\{ - \sum_{n=1}^N \left[ \vec{y}(n) - D(\vec{\theta})\vec{s}(n) \right]^* \cdot I^{-1} \cdot \left[ \vec{y}(n) - D(\vec{\theta})\vec{s}(n) \right] \right\}$$

where  $*$  is the conjugate transpose. The *log likelihood* function, with unnecessary constant terms discarded, is

$$L(\vec{\theta}, \vec{s}; Y) = - \sum_{n=1}^N \left[ \vec{y}(n) - D(\vec{\theta})\vec{s}(n) \right]^* \left[ \vec{y}(n) - D(\vec{\theta})\vec{s}(n) \right] \quad (3.11)$$

MLE maximizes (3.11) with respect to  $\vec{\theta}$  and  $\vec{s}(n)$ . Inspection of (3.11) reveals that this is a nonlinear least squares problem in the sense that the squared error of the residuals is minimized.

For any fixed  $\theta$  one can find the values for  $\vec{s}(n)$  which maximize (3.11). Of course, this must be right at the joint maximum. For given  $\hat{\theta}$ , it is straightforward to show that the maximizing  $\hat{s}(n)$  is [47]

$$\hat{s}(n) = \left[ D^*(\hat{\theta})D(\hat{\theta}) \right]^{-1} D^*(\hat{\theta})\vec{y}(n) \quad (3.12)$$

By substituting (3.12) into (3.11) we can eliminate  $\vec{s}$  from the optimization problem. Defining the sample correlation matrix of the data as

$$C_y = \frac{1}{N} \sum_{n=1}^N \vec{y}(n)\vec{y}^*(n)$$

the maximizing angle vector is given by

$$\hat{\theta} = \arg \max_{\vec{\theta}} \text{tr} \left[ D(\vec{\theta}) \left[ D^*(\vec{\theta})D(\vec{\theta}) \right]^{-1} D^*(\vec{\theta})C_y \right] \quad (3.13)$$

where  $tr[\cdot]$  denotes the matrix trace operation.

Beyond the simplification of (3.13), normally the MLE is a difficult optimization problem and does not yield an analytical solution. Thus we are led to consider iterative methods for the solution. There is an enormous literature on methods for solving multidimensional optimization problems. The expectation-maximization algorithm and Space-alternating generalized expectation-maximization algorithm were chosen to solve the computation problems.

### 3.2.2 expectation maximization and SAGE algorithms

The iterative algorithms are a procedure for maximizing the *log likelihood function*  $L(\theta) = \log l(Y|\theta)$ , where  $\theta$  is unknown parameter vector and  $Y$  is the data vector. Assume that after the  $n^{th}$  iteration the current estimate for  $\theta$  is given by  $\theta_n$ . Since the objective is to maximize  $L(\theta)$ , we wish to compute an updated estimate  $\theta$  such that

$$L(\theta) > L(\theta_n)$$

Equivalently we want to maximize the difference,

$$L(\theta) - L(\theta_n) = \log l(Y|\theta) - \log l(Y|\theta_n) \quad (3.14)$$

The expectation-maximization (EM) algorithm [16] [37] is an iterative optimization technique specifically designed for statistical models. It uses a different strategy rather than gradient descent or Newton's method and provides faster convergence sometimes depending on conditions. Figure 3.6 [40] shows the difference between EM and gradient descent algorithms. In order to find the maximum point, the gradient descent algorithm starting from the current guess makes a linear approximation to the maximum-likelihood function, then takes some step uphill. Since there is no knowledge in advance how good the linear approximation is and consequently how big of a step it is.

EM instead makes a local approximation that is a lower bound to the maximum-likelihood function. In principle, the lower bound can have any functional form. Choosing the new estimate to maximize the lower bound will always be an improvement over the previous guess, unless the gradient was zero there. So the

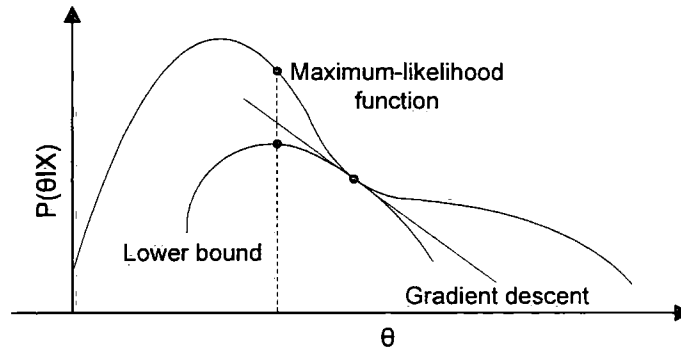


Figure 3.6: Maximizing a function with lower bound [40]

idea is to alternate between computing a lower bound *E-step* and maximizing the bound *M-step*, until a point of zero gradient is reached [40].

In statistics literature, the term *incomplete data* in its general form implies the existence of two sample space  $Y$  and  $X$  and many-one mapping from  $X$  to  $Y$ . The data set  $x$  in  $X$  is not observed directly, but only indirectly through  $Y$ . More specifically, we assume there is a mapping  $x \rightarrow y(x)$  from  $X$  to  $Y$ , and that  $x$  is known only to lie in  $X(y)$ , the subset of  $X$  determined by the equation  $y = y(x)$ , where  $y$  is the observed data. We refer to  $x$  as the complete data even though in certain examples  $x$  includes what are traditionally called parameters [16].

We assume a set of sampling densities  $f(x|\phi)$  depending on parameters  $\phi$  and derive its corresponding set of sampling densities  $g(y|\phi)$ . The complete-data specification  $f(x|\phi)$  is related to the incomplete-data specification  $g(y|\phi)$  by

$$g(y|\phi) = \int_{x(y)} f(x|\phi) dx$$

The EM algorithm is directed at finding a value of  $\phi$  which maximizes  $g(y|\phi)$  given an observed  $y$  by using of the associated family  $f(x|\phi)$ . There are many possible complete-data specifications  $f(x|\phi)$  that will generate  $g(y|\phi)$  given the incomplete-data specification  $g(y|\phi)$ .

Before giving a description of the general EM algorithm, we first introduce the *Jensen's inequality*. For a *convex function*<sup>1</sup>  $f(\cdot)$  defined on a space  $\Omega$ , if a

<sup>1</sup>Let  $f(\cdot)$  be a real function defined on an interval  $I$ ,  $f(\cdot)$  is said to be convex on  $I$  if  $\forall x_1, x_2 \in I, \lambda \in [0, 1], f(\lambda x_1 + (1 - \lambda) x_2) \leq \lambda f(x_1) + (1 - \lambda) f(x_2)$

function  $p(\cdot)$  is a probability density function  $\int_{\Omega} p(x) dx = 1$  and another function  $h(\cdot)$ , *Jensen's inequality* states as: [10]

$$f\left(\int_{\Omega} p(x) \cdot h(x) dx\right) \leq \int_{\Omega} p(x) \cdot f(h(x)) dx$$

*Jensen's inequality* extends the theorem that the arithmetic mean is greater than or equal to the geometric mean.

In the DOA framework, the introduction of the *complete data* is making the maximum likelihood estimation of  $\theta$  tractable. In this case, it is assumed that knowledge of the *complete data* will make the maximization of the likelihood function easier. Denote the *complete data* vector by  $X$  and a given realization by  $x$ . The total probability  $l(Y|\theta)$  can be written in terms of the *complete data*  $x$  as,

$$l(Y|\theta) = \int_{\Omega} l(Y|x, \theta) l(x|\theta) dx$$

Equation (3.14) can be rewritten as [40]

$$\begin{aligned} L(\theta) - L(\theta_n) &= \log \int_{\Omega} l(Y|x, \theta) l(x|\theta) dx - \log l(Y|\theta_n) \\ &= \log \int_{\Omega} l(Y|x, \theta) l(x|\theta) \cdot \frac{l(x|Y, \theta_n)}{l(x|Y, \theta_n)} dx - \log l(Y|\theta_n) \\ &= \log \int_{\Omega} l(x|Y, \theta_n) \cdot \frac{l(Y|x, \theta) l(x|\theta)}{l(x|Y, \theta_n)} dx - \log l(Y|\theta_n) \end{aligned}$$

Since  $l(x|Y, \theta_n)$  is a probability measure and  $\int_{\Omega} l(x|Y, \theta_n) dx = 1$ , *Jensen's inequality* can be applied

$$\begin{aligned} L(\theta) - L(\theta_n) &\geq \int_{\Omega} l(x|Y, \theta_n) \cdot \log \left( \frac{l(Y|x, \theta) l(x|\theta)}{l(x|Y, \theta_n)} \right) dx - \log l(Y|\theta_n) \\ &= \int_{\Omega} l(x|Y, \theta_n) \cdot \log \left( \frac{l(Y|x, \theta) l(x|\theta)}{l(x|Y, \theta_n) l(Y|\theta_n)} \right) dx \\ &\triangleq \Delta(\theta|\theta_n) \end{aligned}$$

We made use of the fact that  $\int_{\Omega} l(x|Y, \theta_n) dx = 1$  so that

$$\log l(Y|\theta_n) = \int_{\Omega} l(x|Y, \theta_n) \log l(Y|\theta_n) dx$$

which allows the term  $\log l(Y|\theta_n)$  to be brought into the integral.

By rewriting  $L(\theta) \geq L(\theta_n) + \Delta(\theta|\theta_n)$  and for convenience define

$$Q(\theta|\theta_n) \equiv L(\theta_n) + \Delta(\theta|\theta_n)$$

so that we can rewrite the relationship as  $L(\theta) \geq Q(\theta|\theta_n)$ . We have now a function  $Q(\theta|\theta_n)$ , which is lower bound of the likelihood function  $L(\theta)$ . Applied Bayes' theorem  $p(A|B) = \frac{P(B|A)P(A)}{P(B)}$ ,

$$\begin{aligned} Q(\theta_n|\theta_n) &= L(\theta_n) + \Delta(\theta_n|\theta_n) \\ &= L(\theta_n) + \int_{\Omega} l(x|Y, \theta_n) \cdot \log \left( \frac{l(Y|x, \theta_n) l(x|\theta_n)}{l(x|Y, \theta_n) l(Y|\theta_n)} \right) dx \\ &= L(\theta_n) + \int_{\Omega} l(x|Y, \theta_n) \cdot \log \left( \frac{l(Y, x|\theta_n)}{l(Y, x|\theta_n)} \right) dx \\ &= L(\theta_n) \end{aligned}$$

so for  $\theta = \theta_n$  the lower bound is tangent with the likelihood function shown in Figure 3.6 [40].

Our objective is to choose a value of  $\theta$  so that  $L(\theta)$  is maximized. On the other hand,  $Q(\theta|\theta_n)$  is the lower bound of the  $L(\theta)$  and that the value of the function  $L(\theta)$  is equal with the value of the function  $Q(\theta|\theta_n)$  at the current estimate for  $\theta = \theta_n$ . Therefore, any  $\theta$  which increases  $Q(\theta|\theta_n)$  in turn increases  $L(\theta)$ . In order to achieve the greatest possible increase in the value of  $L(\theta)$ , the EM algorithm selects  $\theta_{n+1}$  such that  $Q(\theta|\theta_n)$  is maximized. Formally,

$$\theta_{n+1} = \arg \max_{\theta} \{Q(\theta|\theta_n)\}$$

where  $\arg \max_{\theta}$  stands for the argument of the maximum, that is to say, the value  $\theta$  of the given argument for which the value of the given expression attains its maximum value.

$$\theta_{n+1} = \arg \max_{\theta} \left\{ L(\theta_n) + \int_{\Omega} l(x|Y, \theta_n) \cdot \log \left( \frac{l(Y|x, \theta) l(x|\theta)}{l(x|Y, \theta_n) l(Y|\theta_n)} \right) dx \right\}$$

Now drop terms which do not vary with  $\theta$  we get

$$\begin{aligned}
 \theta_{n+1} &= \arg \max_{\theta} \left\{ \int_{\Omega} l(x|Y, \theta_n) \cdot \log (l(Y|x, \theta) l(x|\theta)) dx \right\} \\
 &= \arg \max_{\theta} \left\{ \int_{\Omega} l(x|Y, \theta_n) \cdot \log \left( \frac{l(Y, x, \theta)}{l(x, \theta)} \cdot \frac{l(x, \theta)}{l(\theta)} \right) dx \right\} \\
 &= \arg \max_{\theta} \left\{ \int_{\Omega} l(x|Y, \theta_n) \cdot \log (l(Y, x|\theta)) dx \right\} \\
 &= \arg \max_{\theta} \left\{ E_{X|Y, \theta} \{ \log l(Y, x|\theta) \} \right\}
 \end{aligned}$$

In the direction of arrival applications, define the complete data to be the set of  $N$  samples of  $P$  independent complex Gaussian vector  $\{\bar{x}_l(n); 1 \leq l \leq P, 1 \leq n \leq N\}$ , where  $P$  denotes the number of sources and  $N$  denotes the number of sample points. The  $l^{\text{th}}$  vector  $x_l(n)$  has mean  $d(\theta_l) s_l(n)$  and has identical covariance  $(1/M)I$ , where  $I$  is an identity matrix [39]. Introducing the complete data as if one could somehow observe each of the incident plane waves separately, then estimation of their angles-of-arrival would be straightforward. The incomplete data is simply the set of observations themselves, and the many-to-one function which maps the complete data to incomplete data is

$$\vec{y}(n) = \sum_{l=1}^P \bar{x}_l(n)$$

The complete data log likelihood function becomes

$$Q(X|\vec{\theta}) = - \sum_{n=1}^N \sum_{l=1}^P \left| \bar{x}_l(n) - \vec{d}(\theta_l) s_l(n) \right|^2 \tag{3.15}$$

Applying the EM algorithm to the DOA problem is defined such that at iteration  $n + 1$

- *E-step*: expectation step involves the conditional expectation of the log likelihood function of the complete-data, denoted as  $Q(\theta|\theta_n)$ , where the condition is with respect to the incomplete-data and previous estimate  $\theta_n$ .

- *M-step*: maximization step yields the estimate of the angles  $\theta_{n+1}$  by maximizing  $Q(\theta|\theta_n)$  with respect to  $\theta$ .

Since  $x_l$  is a sufficient statistic<sup>1</sup> for  $\vec{d}(\theta_l) s_l$ , the *E-step* can be reduced to finding its conditional expectation  $E\{x_l(n) | \theta_l^n\}$ . [39] and [12] showed that this conditional mean is

$$E\{\bar{x}_l(n) | \theta_l^n\} \triangleq \bar{x}_l^n(n) = \vec{d}(\theta_l^n) \bar{s}_l^n(n) + \frac{1}{P} \left[ \bar{y}(n) - D(\bar{\theta}^n) \bar{s}^n(n) \right] \quad (3.16)$$

The conditional mean is the known signal component estimated from the previous iteration, plus an equal share of that component of the observation vector which is orthogonal to signal subspace defined by  $D(\bar{\theta})$ .

For the *M-step*, define

$$C_{x_l}^n = \frac{1}{N} \sum_{n=1}^N \bar{x}_l^n(n) \bar{x}_l^n(n)^*$$

where  $*$  is complex conjugate. Maximization of (3.15) with respect to  $\theta_l$  and waveform estimation  $s_l(n)$ , with  $x_l(n)$  replaced by  $\bar{x}_l^n(n)$ , yields the following equations which the estimates must satisfy [39]:

$$\theta_l^{n+1} = \arg \max_{\theta_l} \frac{\vec{d}^*(\theta_l) C_{x_l}^n \vec{d}(\theta_l)}{\left| \vec{d}(\theta_l) \right|^2} \quad (3.17)$$

$$s_l^{n+1} = \frac{\vec{d}^*(\theta_l^{n+1}) \bar{x}_l^{n+1}(n)}{\left| \vec{d}(\theta_l) \right|^2} \quad (3.18)$$

Equations (3.16)–(3.18) define the basic EM algorithm. Equations (3.17) requires a one dimensional search. The quadratic form to be maximized in (3.17) is precisely the form of the projection matrix of (3.13), except that (3.17) only involves a single column of the original matrix  $D(\bar{\theta})$ .

EM algorithms are most useful when the M-step is easier than maximizing the original likelihood. The simultaneous update all parameters in a classical EM algorithm necessitates overly informative complete-data spaces, which in turn leads

---

<sup>1</sup>Intuitively, a sufficient statistic captures all information in the data that is relevant to guessing the values of the unobservable parameters, or more generally, to guessing the underlying probability distribution from which the data were drawn.

to slow convergence. The convergence rate of an EM algorithm is inversely related to the Fisher information of its complete-data space. Improved convergence rate can be achieved by introducing less informative complete-data spaces.

Space-alternating generalized expectation-maximization algorithm (SAGE) was proposed to solve the problems where one can sequentially update small groups of the elements of the parameter vector with EM algorithm, rather than using one large complete-data space [20]. The SAGE algorithm is a twofold extension of the EM algorithm. First, each iteration of the SAGE algorithm is an EM iteration to re-estimate a subset of the components of parameters  $\theta$  while keeping the estimates of the other components fixed. Second, the notion of complete data is generalized in the sense that the mapping from the complete-data space to the incomplete-data space may be random rather than deterministic in EM algorithm.

We first define the *index set*  $S$  which is a subset of the set  $\{1, 2, \dots, P\}$ , for example  $\{1\}$ ,  $\{1, 2\}$ , where  $P$  is the dimension of the parameter space. The set  $\bar{S}$  denotes the complement of  $S$  intersected with  $\{1, 2, \dots, P\}$ . The difference is that the complete-data is not based on the whole parameter space but the sub-space  $\theta_S$  of the whole parameters, which is dependent on *index set*  $S$  [20]. Just as for an EM algorithm, we replace the maximization of *likelihood function*  $L(\theta_S)$  over  $\theta_S$  with the maximization of low bound function  $Q(\theta_S)$ , whose computation will be less than the former.

Then, we introduce the *admissible hidden-data space* of a random vector  $X^S$  with probability density function  $f(x; \theta)$  with respect to  $\theta_S$  for another density function  $f(y; \theta)$ . The joint density of  $X^S$  and  $Y$  satisfies

$$f(y, x; \theta) = f(y|x; \theta_{\bar{S}})f(x; \theta)$$

and the conditional distribution  $f(y|x; \theta_{\bar{S}})$  is independent of  $\theta_S$ . In other words,  $X^S$  must be a complete-data space for  $\theta_S$  given that  $\theta_{\bar{S}}$  is known. It can be thought of  $Y$  as the output of a channel that may depend on  $\theta_{\bar{S}}$  but not on  $\theta_S$ , as illustrated in Figure 3.7 [20]. The conditional expectation of the log-likelihood of the *hidden-data* of  $X^S$  can be expressed as

$$Q(\theta_S) = E \{ \log f(X^S; \theta_S | Y; \theta_S, \theta_{\bar{S}}) \}$$

$$= \int_{\Omega^S} \log f(x; \theta_S, \theta_{\bar{S}}) f(x|Y; \theta_S, \theta_{\bar{S}}) dx$$

which is an essential ingredient of any SAGE algorithm.

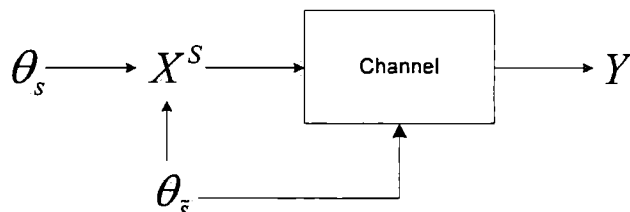


Figure 3.7: The relationship of the observed data and hidden-data

Let  $\theta^0$  be an initial parameter estimate. A generic SAGE algorithm produces a sequence of estimates  $\{\theta^i\}_{i=0}^{\infty}$  via the following recursion [12]:

```

Function SAGE();
for i ← 1 to ∞ do
    choose an indexset S = Si;
    choose an admissible hidden-data space XSi for θSi;
    E-step: compute Q(θSi);
    M-step: θSii+1 = arg maxθSi Q(θSi), θ $\bar{S}^i$ i+1 = θ $\bar{S}^i$ i;
    optional: repeat E-step and M-step;
end
    
```

In DOA applications, the parameter space  $\vec{\theta}$  has been divided into  $P$  subsets  $\{\vec{\theta}_l = (\theta_l, \vec{s}_l) : 1 \leq l \leq P\}$  with  $\vec{s}_l = \{s_l(1), s_l(2), \dots, s_l(N)\}$ . Each subset corresponds to one source signal. The *admissible hidden-data space* associated with this parameter subset  $\{\vec{x}_l(n); 1 \leq l \leq P, 1 \leq n \leq N\}$

$$\vec{x}_l(n) = \vec{d}(\theta_l) s_l(n) + \vec{n}(n)$$

consists of the signal transmitted by the  $l^{\text{th}}$  source and the total noise contained in array output. Since the noise component  $\vec{n}(n)$  is completely incorporated in each iteration, the *admissible hidden-data space* becomes less informative than

that in EM algorithm. [11] showed that the *admissible hidden-data space* can lead to a faster convergence than the EM algorithm.

In DOA, *index sets*  $S^i, i = 0, 1, \dots$  have been chosen as

$$S^i = \{1 + (i \bmod P)\}$$

By this, the parameter subset associated with this *index sets* is updated by maximizing the conditional expectation of likelihood of the *admissible hidden-data*  $X_l$  in each iteration. Every  $P$  iterations in the SAGE algorithm consist of a circle, in which the parameter subsets  $\{\vec{\theta}_l = (\theta_l, s_l) : 1 \leq l \leq P\}$  are updated sequentially.

In the SAGE algorithm, *E-step* calculates

$$E\{x_{S^i}(n)|Y, \theta_{S^{i-1}}\} \triangleq \vec{x}_{S^i}(n) = \vec{d}(\theta_{S^{i-1}})\vec{s}_{S^i}(n) + \vec{y}(n) - D(\theta_{S^{i-1}})\vec{s}_{S^{i-1}}(n)$$

where  $S^i = \{1 + (i \bmod P)\}$ . *M-step* calculates [39]

$$C_{x_{S^i}} = \frac{1}{N} \sum_{n=1}^N \vec{x}_{S^i}(n)\vec{x}_{S^i}^*(n)$$

$$\theta_{S^i} = \arg \max_{\theta} \frac{\vec{d}^*(\theta)C_{x_{S^i}}\vec{d}(\theta)}{|\vec{d}(\theta)|^2}$$

$$s_{S^i}(n) = \frac{\vec{d}^*(\theta_{S^i})\vec{x}_{S^i}(n)}{|\vec{d}(\theta_{S^i})|^2}$$

After each *E-step* and *M-step*, update the complete-data space

$$\theta_{\bar{S}^i} = \theta_{\bar{S}^{i-1}}$$

### 3.3 Algorithm Verifications

The verifications of the SAGE start with the Matlab simulations. The source signals composed of two sine waves whose normalized angular frequencies are  $2\pi/8$  and  $2\pi/5$  and the signal-to-noise ratio are less 5 dB shown in Figure 3.8. These two sources located at  $13^\circ$  and  $-13^\circ$  of the broadside direction of a linear phased array which is composed of 4 omnidirectional antennas. Since the nature

### 3.3 Algorithm Verifications

Transmitters	Transmitted angles	Estimation of incident angles
$Tx 1$	$13^\circ$	$13^\circ$
$Tx 2$	$-13^\circ$	$-12^\circ$
$Tx 1$ and $Tx 2$	$13^\circ$ and $-13^\circ$	$12^\circ$ and $-15^\circ$

Table 3.1: Estimation of incident angles in simulation

of the SAGE algorithm is statistical signal processing which estimates parameters with statistics and probability information of measured data, estimated errors are inevitable. The estimated angles of the incident signals by the SAGE algorithm are shown in Table 3.1 started with the initial values  $10^\circ$  and  $30^\circ$ . However, the estimating errors can be reduced by increasing the signal-to-noise ratio and increasing the number of the antennas. The next step in verifying the algorithm is to set up the phased array and transmitting sources physically in the restrained environment.

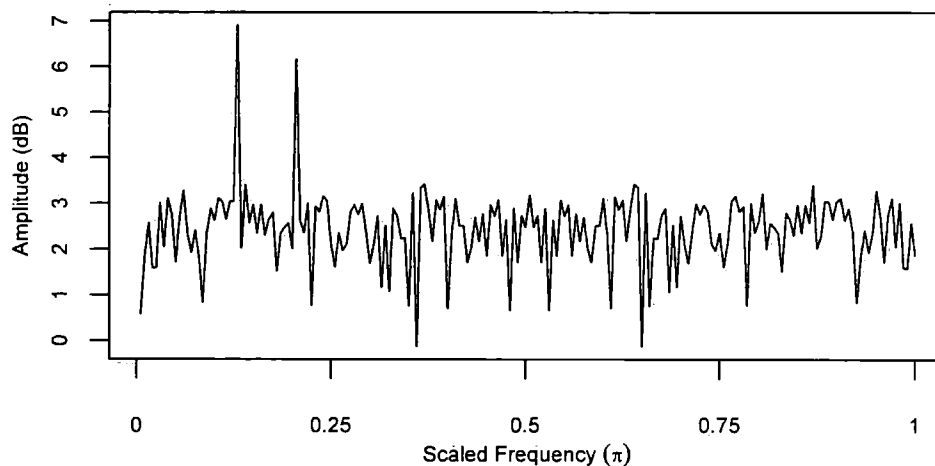


Figure 3.8: The spectrum of the simulation signals

There are anomalies of radio propagation which at ground level can affect these techniques. Common potential problems include reflections or multi-path loss. In a multi-path situation, the radio signal may be arriving at the antenna or antennas from multiple directions, perhaps because the signal is reflecting off

### 3.3 Algorithm Verifications

nearby buildings, hills, or metal structures such as fences. The strongest signal may, in fact, be coming from a reflection rather than the direct path, especially if the direct path includes terrain features that might attenuate the signal. This can result in false directional readings.

The TDK compact fully anechoic chamber  $CAC - S^{TM}$  which is designed to suppress the electromagnetic waves is a test facility to perform radiated EMC measurements between 26 MHz and 18 GHz [25]. The radiation absorbent material shown in Figure 3.9 is designed and shaped to absorb incident RF radiation, as effectively as possible, from as many incident directions as possible. TDK ferrite tiles X-131 are mounted onto dielectric panels to cover the complete frequency range from 26 MHz to 1000 MHz. Another kind of radiation absorbent material is the TDK IP-045C pyramid shaped, resistive absorbers in selected areas of ceiling and walls extended the frequency range up to 18 GHz. White end caps complete the absorbers and greatly improve the illumination levels inside the chamber. For example, the reflectivity loss at 500 MHz frequency is more than 25 dB [53].

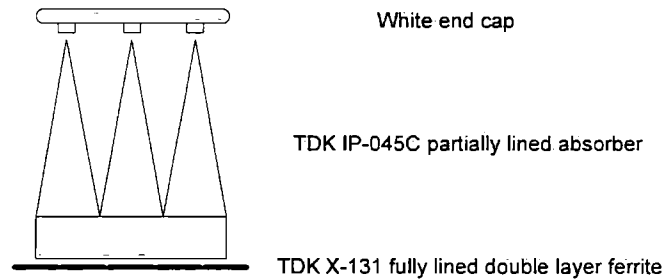


Figure 3.9: The absorbers of the TDK anechoic chamber

The direction of signal arrival experiment conducted in the anechoic chamber was constructed as in Figure 3.10. In the experiments, a 4 element linear array was deployed as the receiving array to estimate the arrival directions of 1 and 2 signal sources. A *Marconi 2020* RF signal generator was used to transmit 500 MHz sine wave. When 2 signals were transmitted, the 500 MHz signal generated by *Marconi 2020* was split by 2 way splitter and fed into the antennas. The reason why only one signal generator was used is that the different phase jitters in each

signal generator caused the unstable situation in the receiver end and resulted in false directional findings.

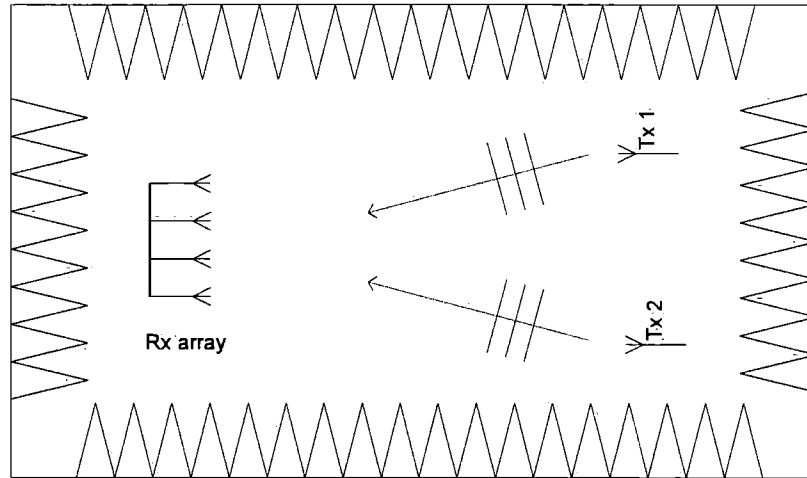


Figure 3.10: The configuration of DOA experiment

In this experiment, a Agilent oscilloscope *Infiniium* 54845A was deployed as 4 channel receiver, whose specifications are [54]:

- 1.5 GHz bandwidth
- 4 input channels
- 8 GSa/s sample rates (2 channel mode)
- 4 GSa/s sample rates (4 channel mode)

4 GSa/s sample rate was chosen to prevent aliasing and to increase the DOA resolution. Each channel has a maximum memory depth<sup>1</sup> of 32 K sample points. *Infiniium* 54845A can sample as fast as 8 GSa/s. However, under this mode, *Infiniium* adds the acquisition resources of channel 2 to channel 1 and channel 4 to channel 3. The memory depth was chosen 512 points which is enough for SAGE algorithm.

<sup>1</sup>Memory depth control determines the number of waveform data points stored for all channel. When you select a small memory depth value and a fast sampling rate, you will have a very fast display update rate but a small amount of the waveform data.

Instead of the equivalent time model, real time mode [54] was chosen which acquires all of the waveform samples during one trigger event. In real time model, the sample rate should be at least 4 times the highest frequency component of the waveform; otherwise, it is possible for reconstructed waveform to be distorted or aliased. A  $\text{Sin}(x)/x$  interpolation filter was used to improve the reconstruction of the measurement by adding data points between the acquired data points.

Since most DOA algorithms rely on the phase information of incident waveforms, phase shifting plays a crucial role in DOA estimation. This requires that each channel in the receiver should be sampled at the same time or at exact known delays which can be compensated in the data processing. In *Infiniium* 54845A calibration, a signal generated by *Marconi* 2020 was split into 4 equal power signals. Then these signals were fed into the input channels of the *Infiniium* 54845A with the same length cables. Figure 3.11 shows the result of back-to-back calibration. Although the second channel has short delay, most algorithms can tolerate this delay.

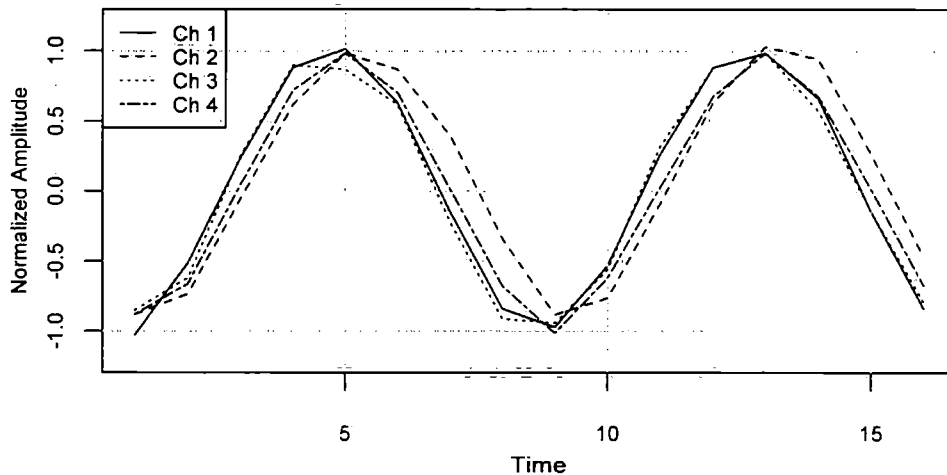


Figure 3.11: The calibration data of the receiver

The sensitivity of the *Infiniium* 54845A is about -50 dBm. The maximum output of the *Marconi* 2020 is 12 dBm and the RF signal will be separated into two. At the receiving array end, the incident power is about -45 dBm. So there is 5 dB signal-to-noise ratio.

To simplify the experiment environment, the omnidirectional antennas were adopted as transmitting and receiving antennas. The receiving array was configured as a half wavelength uniform linear array which permits the use of fast algorithm, such as the fast fourier transform, to compute spatial spectra and array output signals. The array consists of 4 elements placed in a straight line with a uniform distance. Since 500 MHz frequency signal was used in this experiment, the distance between adjacent antennas equals half wavelength  $\lambda/2 = (C/f)/2 \approx 30 \text{ cm}$ , where  $C$  is the speed of light.

In most DOA algorithms, the receiving array is assumed to be in the far field of the transmitting antennas. This means that the distance of the array from the source is much greater than the dimension of the source antenna. The far-field region is commonly taken to exist at distances greater than  $D^2/\lambda$  from the source,  $\lambda$  being the wavelength and  $D$  being the overall dimension of the transmitting antenna. The dimensions of the TDK  $CAC - S^{TM}$  anechoic chamber are 7.5 m long, 4.0 m wide and 3.0 m high [25]. The distance between transmitting antennas and receiving array is 5 m, which can be considered within the far field of the source antennas.

Figure 3.12 shows the phase difference caused by the cables which connected the transmitter and antennas. From Figure 3.12, there is one point delay between two waveforms. Since the sample rate is 8 times faster than the frequency of the waveform, this correspond to  $45^\circ$  phase difference. This phase difference will not affect the accuracy of the DOA algorithms since most of them use only relative phase difference. The phase difference information can be useful as a reference in simulation by which the real receiving waveforms can be compared.

Since it is difficult to measure the exact physical angle of the incident waveform, in this experiment, we fixed the position of two transmitting antennas as in Figure 3.10 and rotated the receiving array to estimate the relative angle. In the receiving array position, the angle between 2 transmitting antennas is  $26^\circ 5'$  measured with a laser angle measurer. Another issue is that, in order to simplify the representation of the algorithm, most DOA algorithms are written with *analytic representation*. However, in practice, the *Infinium* 54845A only acquires data in *real part*. One solution is to modify the DOA algorithms to *real value*

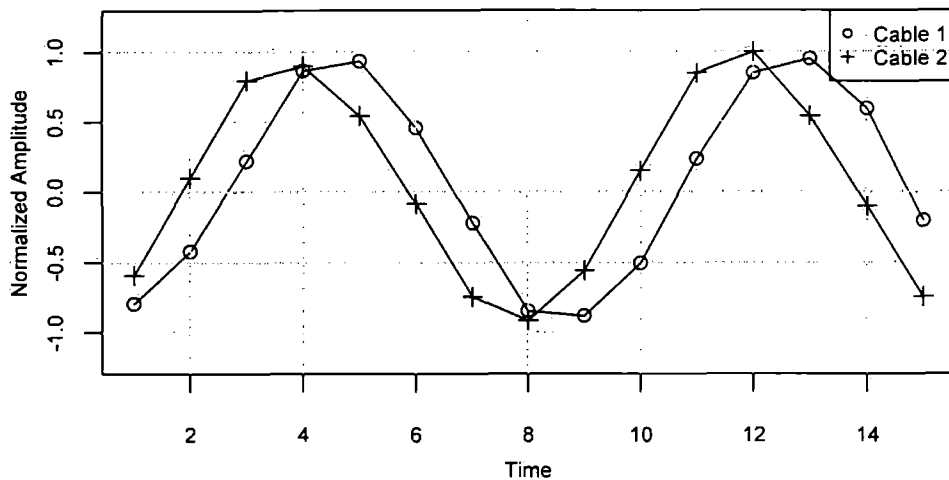


Figure 3.12: The phase difference caused by cables

format; another is to use *Hilbert transform* to convert the *real series* to *analytic signal*, which was adopted by our experiment.

Before transmitting and receiving signals, we fed the calibrated data in Figure 3.11 into the SAGE algorithm. The expected value  $0^\circ$  was obtained which partly proved the correctness of the algorithm. Next we fed the data which were generated with the calibrated data by introducing a point delay. Since the sample rate is 8 times as fast as the frequency of the 500 MHz signal, a point delay corresponds to  $45^\circ$  phase delay and to  $14^\circ 5'$  incident angle. The SAGE algorithm estimates successfully  $14^\circ$ .

In experiment 1, the receiving array was placed at a perpendicular angle with the long axis of the chamber shown in Figure 3.10. So, the receiving angles are roughly  $-13^\circ$  and  $13^\circ$  respectively. At first, only a source signal  $Tx1$  was transmitted and the received waveforms are shown in Figure 3.13. Then  $Tx1$  was disconnected and another source signal  $Tx2$  was transmitted and the received waveforms are shown in Figure 3.14. Theoretically only phase differences exist between equal power receiving waveforms. However, from Figure 3.13 and 3.14, there are amplitude changes of receiving signals because of unknown factors. At the third step, two sources were transmitted and the received waveforms are shown in Figure 3.15.

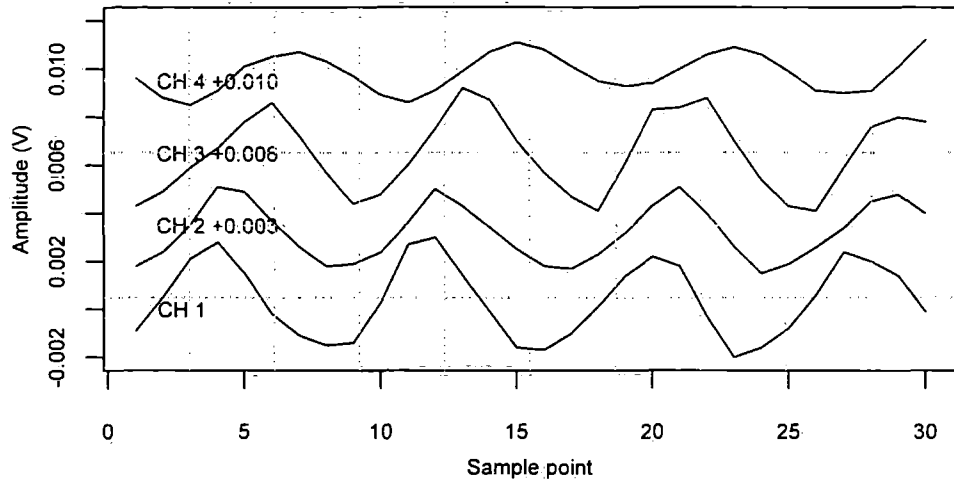


Figure 3.13: The received waveforms when Tx 1 transmitted

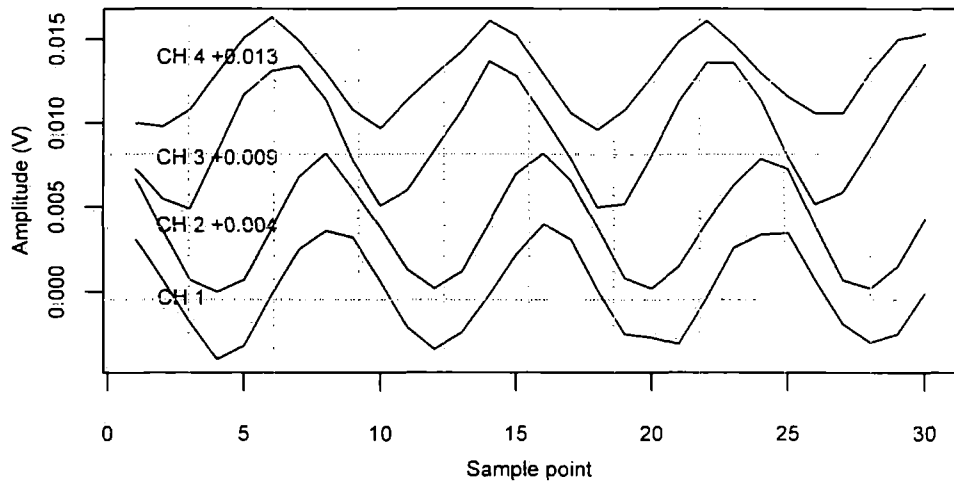


Figure 3.14: The receiving waveforms when the Tx 2 transmitted

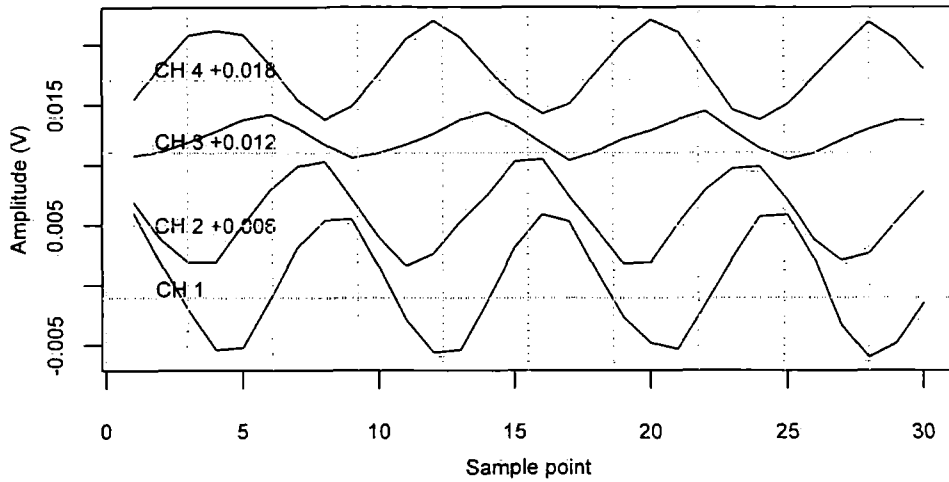


Figure 3.15: The receiving waveforms when both Tx1 and Tx 2 transmitted

Transmitters	Estimation of incident angles
<i>Tx 1</i>	13°
<i>Tx 2</i>	-12°
<i>Tx 1</i> and <i>Tx 2</i>	11° and -19°

Table 3.2: Estimation of incident angles in experiment 1

We use the SAGE algorithm to estimate the incident angles and verify its correction. Estimations are listed in Table 3.2. In the signal radio source scenario, estimations are more accurate than the counterpart estimations of two source signals. The SAGE algorithm gave estimated directions 11° and -19° for *Tx 1* and *Tx 2*. The 30° difference between two estimated sources has around 4° deviation with physical angle 26°5'. In experiment 2, the receiving array was clockwise rotated 15° - 25°; and in experiment 3, the receiving array was anti-clockwise rotated 15° - 25°. Then we transmitted the same signals as described in experiment 1 and the results are shown in Table 3.3 and Table 3.4.

From the results, the SAGE algorithm roughly estimated the directions of incident signals in the controlled environment. Although it is impossible to make the exact estimations in statistical signal processing, we can approximate the

Transmitters	Estimation of incident angles
<i>Tx 1</i>	$-5^\circ$
<i>Tx 2</i>	$-29^\circ$
<i>Tx 1</i> and <i>Tx 2</i>	$-1^\circ$ and $-26^\circ$

Table 3.3: Estimation of incident angles in experiment 2

Transmitters	Estimation of incident angles
<i>Tx 1</i>	$31^\circ$
<i>Tx 2</i>	$1^\circ$
<i>Tx 1</i> and <i>Tx 2</i>	$32^\circ$ and $4^\circ$

Table 3.4: Estimation of incident angles in the experiment 3

correct estimation in many ways. For example, in DOA experiments the accuracy can be improved by adopting more element antennas in the receiving array or increasing the searching steps in the SAGE algorithm.

### 3.4 Summary

The direction estimation of the incident radio wave is a technique to determine directions of radio waves, therefore to infer the spectrum occupancy spatially. We introduced the linear array as an example to illustrate the principles of the phased array and corresponding array signal processing. From a statistical signal processing point of view, the method of maximum likelihood estimation is considered to be more robust and yields estimators with good statistical properties. SAGE and EM algorithms are used to solve the intractable computation issues of the MLE. In our project, simulations and experiments were conducted to verify the validity of the algorithms.

# Chapter 4

## Statistics and Modelling

### 4.1 Occupancy analysis with descriptive Statistics

Before going into the actual statistical modeling and data analysis, it is often helpful to make some simple characterization of the data in terms of summary statistics and graphics.

The spectrum measurements contained in this thesis can only be used to assess the feasibility of using alternate services or systems under restricted conditions. Extrapolation of data in this thesis to general spectrum occupancy for spectrum sharing requires consideration of additional factors. These include spectrum management regulations, types of missions performed in the bands and new spectrum requirements in the development and procurement stages. Also, measurement area, measurement site, and measurement system parameters should be considered.

Highly dynamic bands where occupancy changes rapidly include those used by mobile radios (land, marine, and airborne) and airborne radars. These bands should be assigned a high priority and measured often during a spectrum survey in order to maximize opportunities for signal detection. Bands that are not very dynamic in their occupancy such as those occupied by commercial radio and television signals or fixed emitters such as air traffic control radars need not be observed as often, because the same basic occupancy profile will be generated



## 4.1 Occupancy analysis with descriptive Statistics

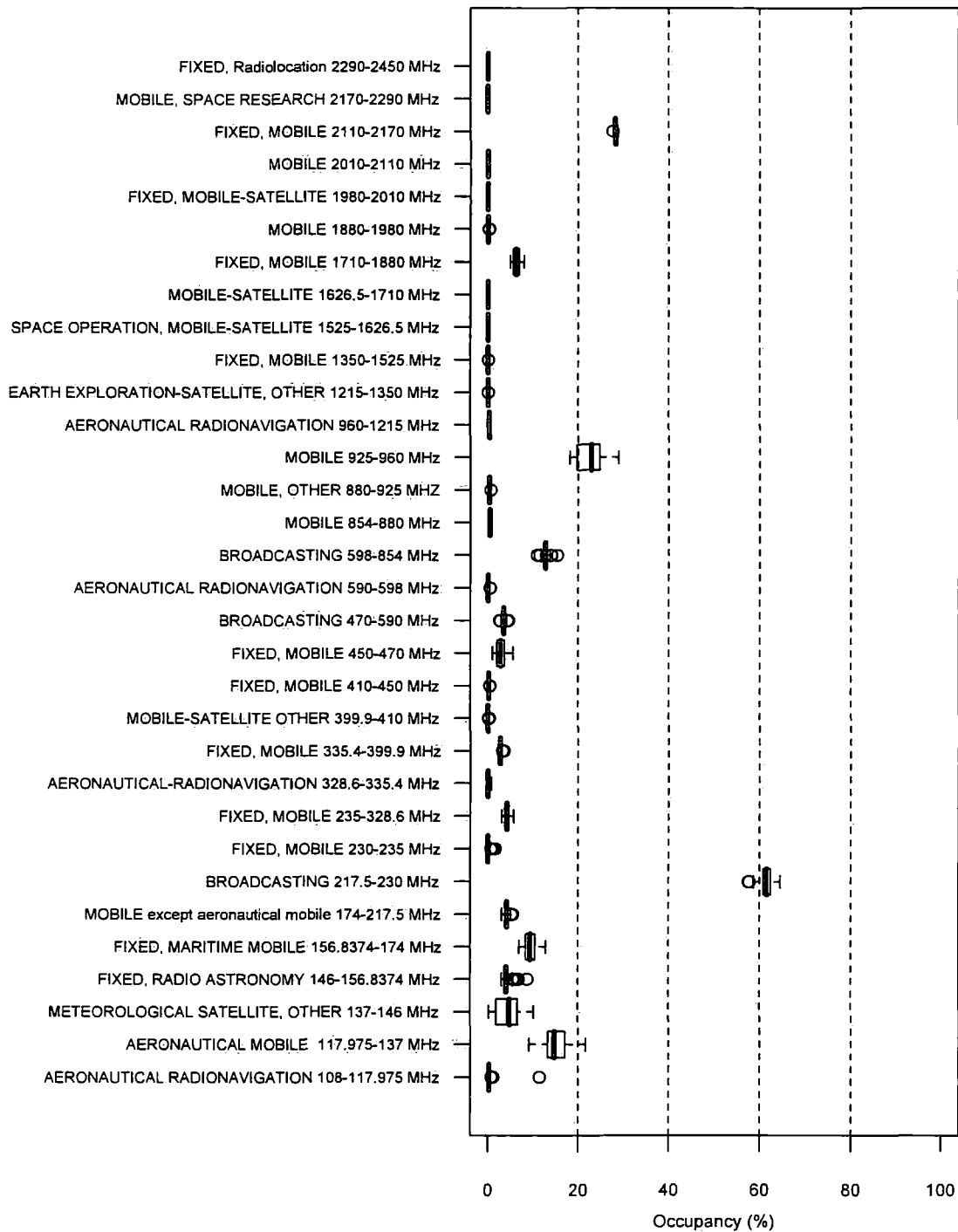


Figure 4.2: Occupancy statistics of spectrum

## 4.1 Occupancy analysis with descriptive Statistics

---

single transmitter which is always on and which experiences no fading. At the other extreme, a large difference between the mean and maximum measurements suggests intermittent use of the channel.

The middle panel shows the band occupancy in the *time domain* with thresholds -95 dBm and -100 dBm during the measurement period of 27/06/07 – 03/07/07. For example, the occupancy rate of the Air band shown in Figure 4.3 was calculated in each time point in a given threshold -95 dBm and -100 dBm respectively. Total 168 time points for each hour of 7-day were plotted in this panel.

The bottom panel shows the statistical distributions of the *white spectrum*. *White spectrum* can be defined as the continuous idle spectrum in a given bandwidth and in a given threshold which can be used for communications. For example, in Figure 4.9, for -100 dBm threshold we can find about 60% white band with 1000 kHz bandwidth distributed in the total spectrum, if we divide the TV band (470 - 854 MHz) into 384 sub-bands with 1000 kHz bandwidth.

The statistics table of each figure shows the minimum and maximum occupancy rates, 1<sup>st</sup>, 3<sup>rd</sup> quantile values, and mean and median<sup>1</sup> values.

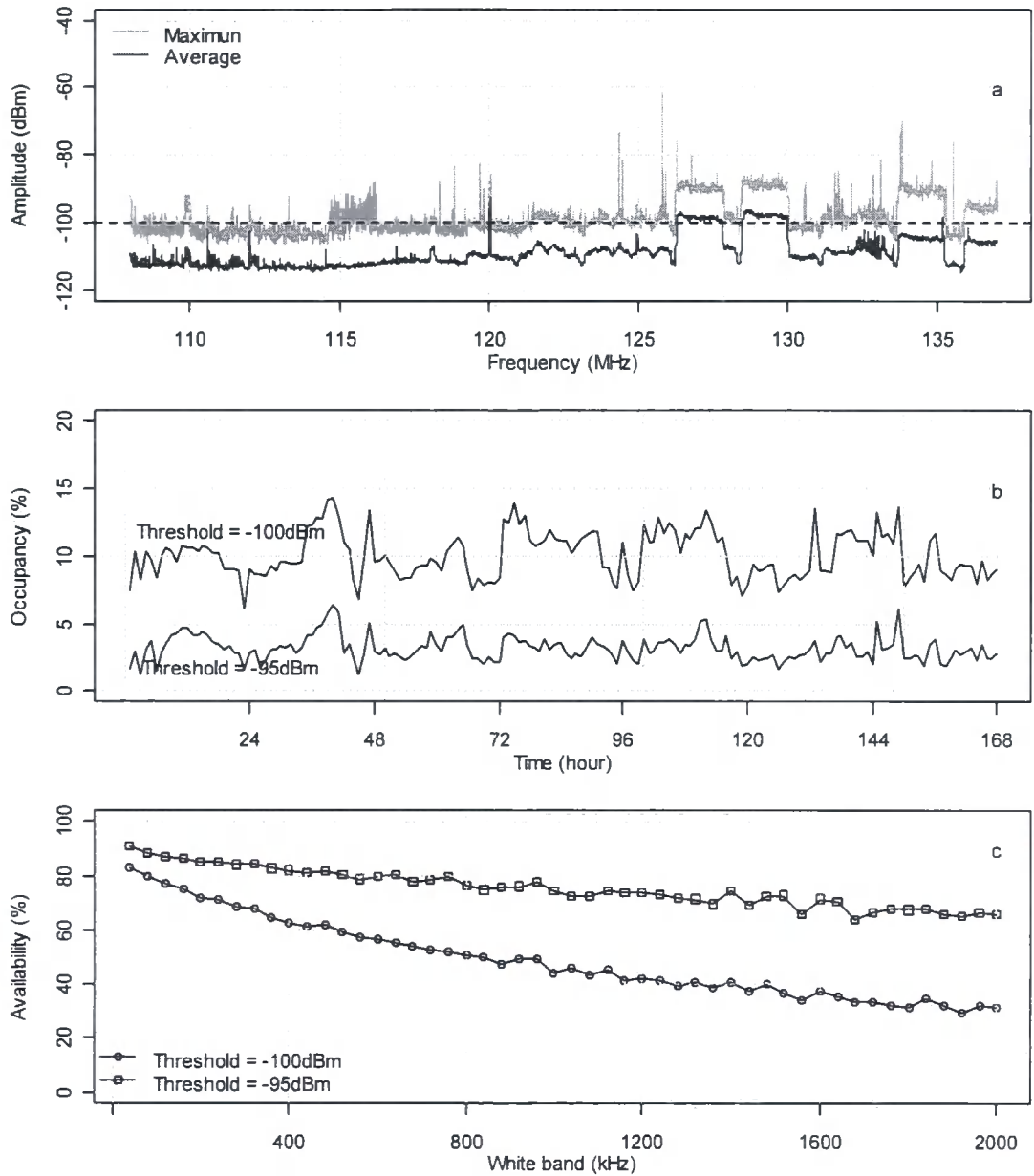
Figure 4.14 shows a snapshot of the occupancy data with Sounder by which we monitored the 2.4 GHz spectrum. Figures 4.15 to 4.17 show the occupancy statistics in the 2.4 GHz ISM band in the *space domain* with boxplot. The data were taken in the thermo lab, in the school entrance and the corridor of the Engineering School, which were labeled as Locations 1-3 respectively. The data in Figures 4.15 to 4.17 shows the ISM band occupancy for very short period – one second with 250 sweeps. The statistics shows that although the average occupancies are less than 30%, the distributions of the data in different locations and different channels are positively skewed<sup>2</sup>. The positive skewness of the data indicates the characterization of the transmitting signals is wideband in the 2.4 GHz ISM band while low average occupancies may be caused by fast moving blue tooth signals.

---

<sup>1</sup>a median is described as the number separating the higher half of a sample, a population, or a probability distribution, from the lower half.

<sup>2</sup>Positive skewness indicates a distribution with an asymmetric tail extending towards more positive values.

## 4.1 Occupancy analysis with descriptive Statistics

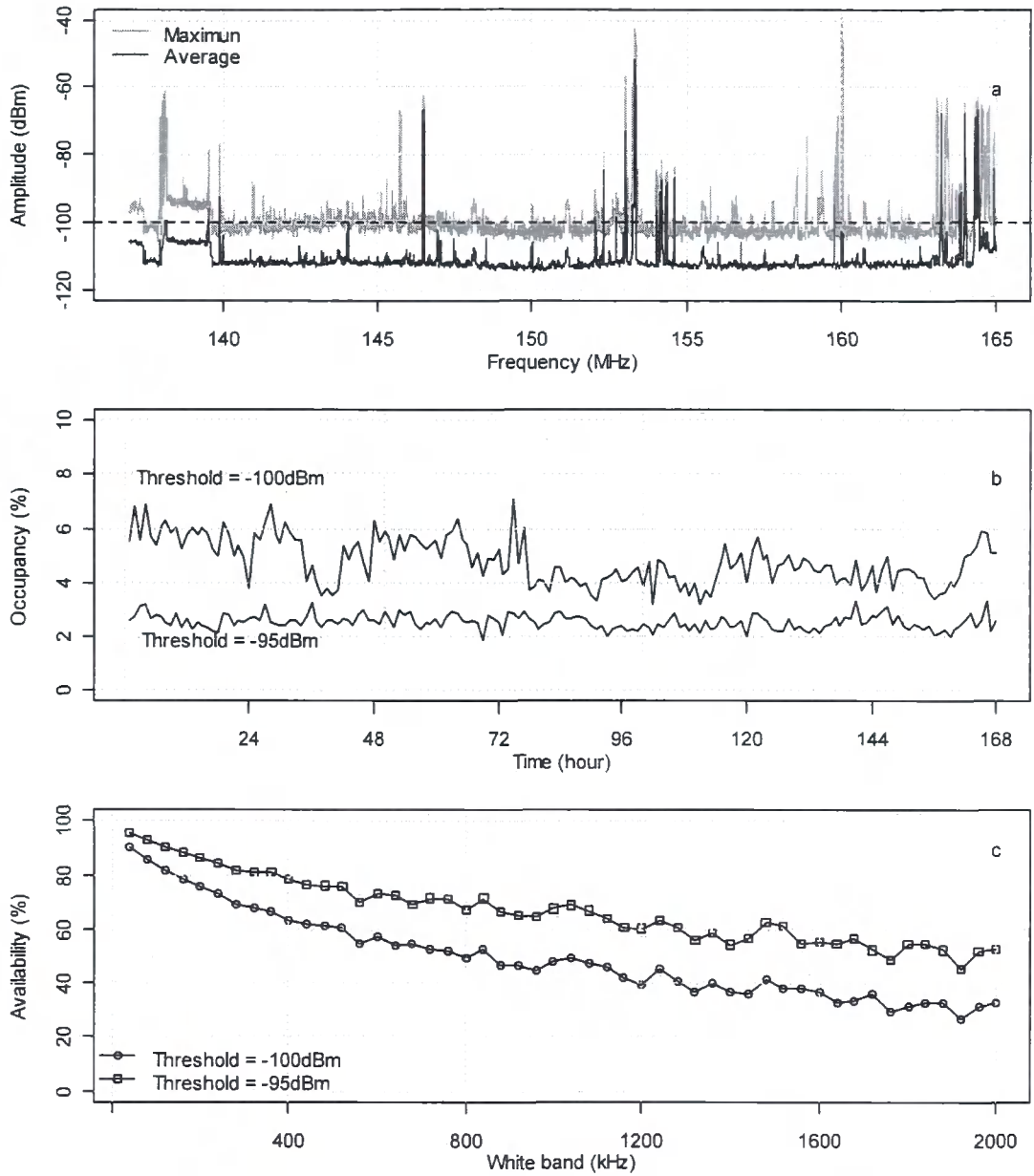


Air Band 108-137 MHz

	-105 dBm	-95 dBm
Min	6.06	1.20
1 <sup>st</sup> Quantile	8.81	2.58
Median	9.76	3.03
Mean	10.06	3.16
3 <sup>rd</sup> Quantile	11.19	3.67
Max	14.28	6.33

Figure 4.3: Occupancy statistics of Air Band

## 4.1 Occupancy analysis with descriptive Statistics

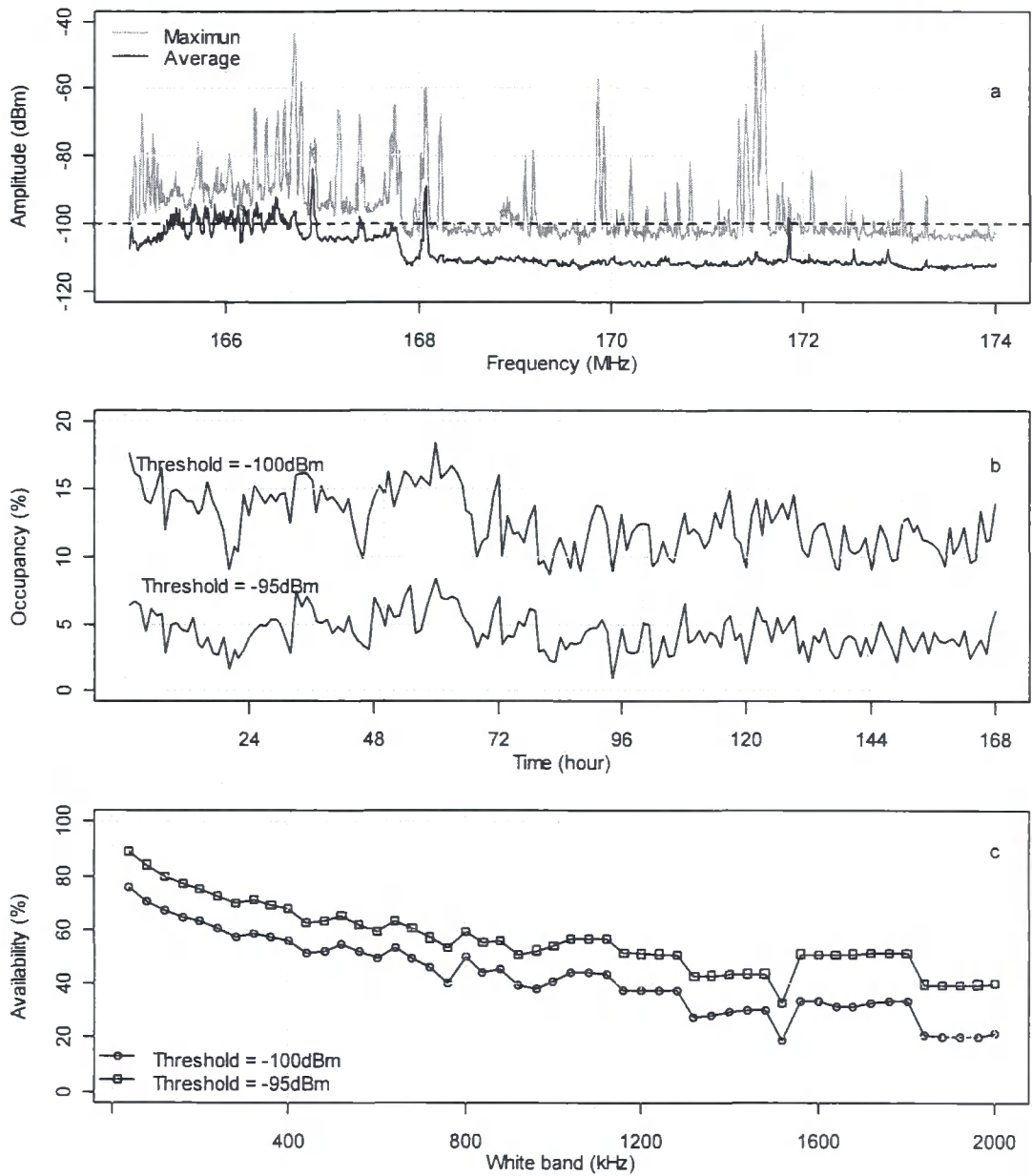


Mid Band 137-165 MHz

	-105 dBm	-95 dBm
Min	3.24	1.93
1 <sup>st</sup> Quantile	4.13	2.35
Median	4.81	2.53
Mean	4.84	2.52
3 <sup>rd</sup> Quantile	5.54	2.71
Max	7.09	3.32

Figure 4.4: Occupancy statistics of Mid Band

## 4.1 Occupancy analysis with descriptive Statistics

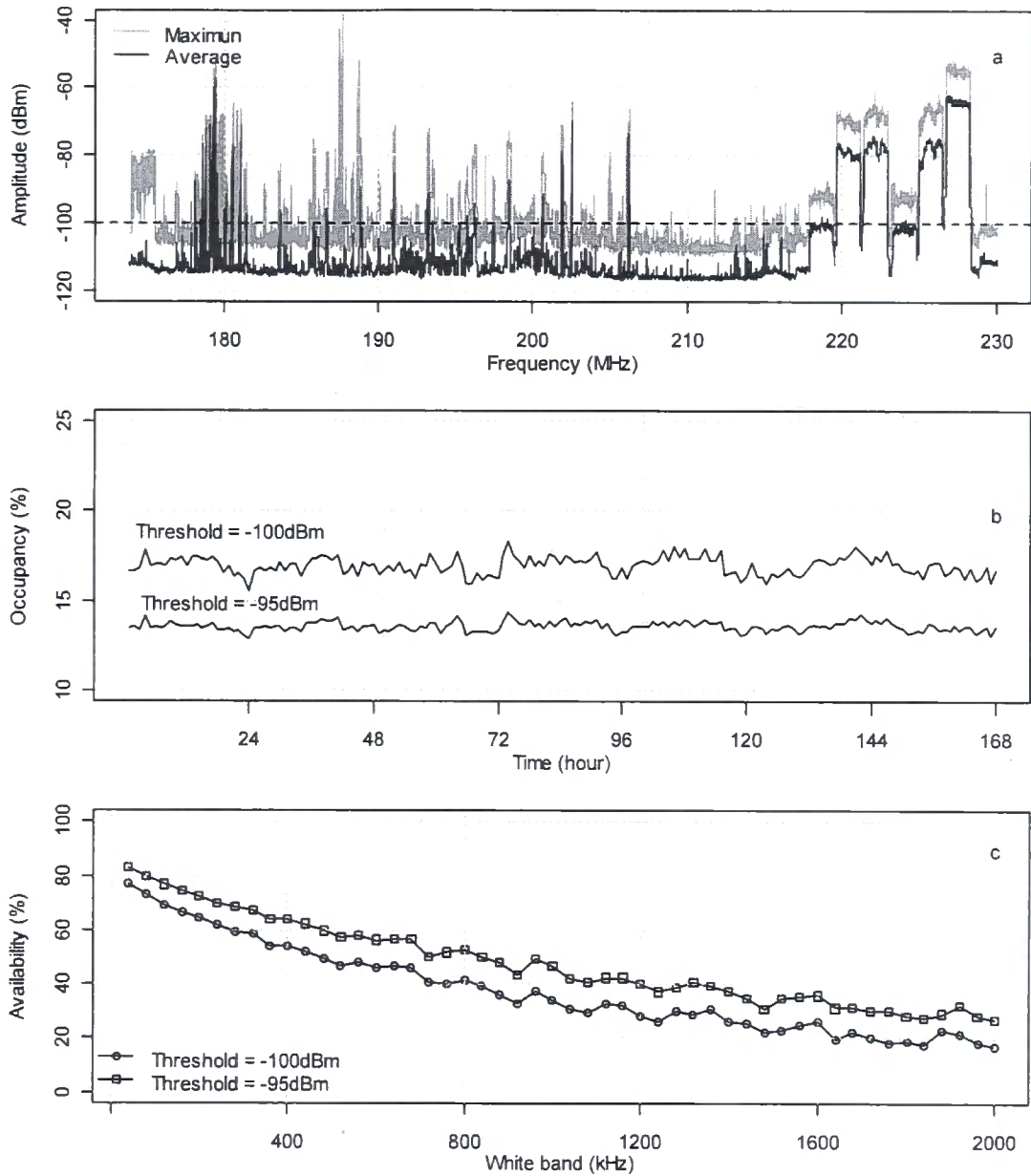


High Band 165-174 MHz

	-105 dBm	-95 dBm
Min	8.53	0.99
1 <sup>st</sup> Quantile	10.74	3.30
Median	12.29	4.26
Mean	12.51	4.35
3 <sup>rd</sup> Quantile	14.29	5.21
Max	17.94	8.31

Figure 4.5: Occupancy statistics of High Band

## 4.1 Occupancy analysis with descriptive Statistics

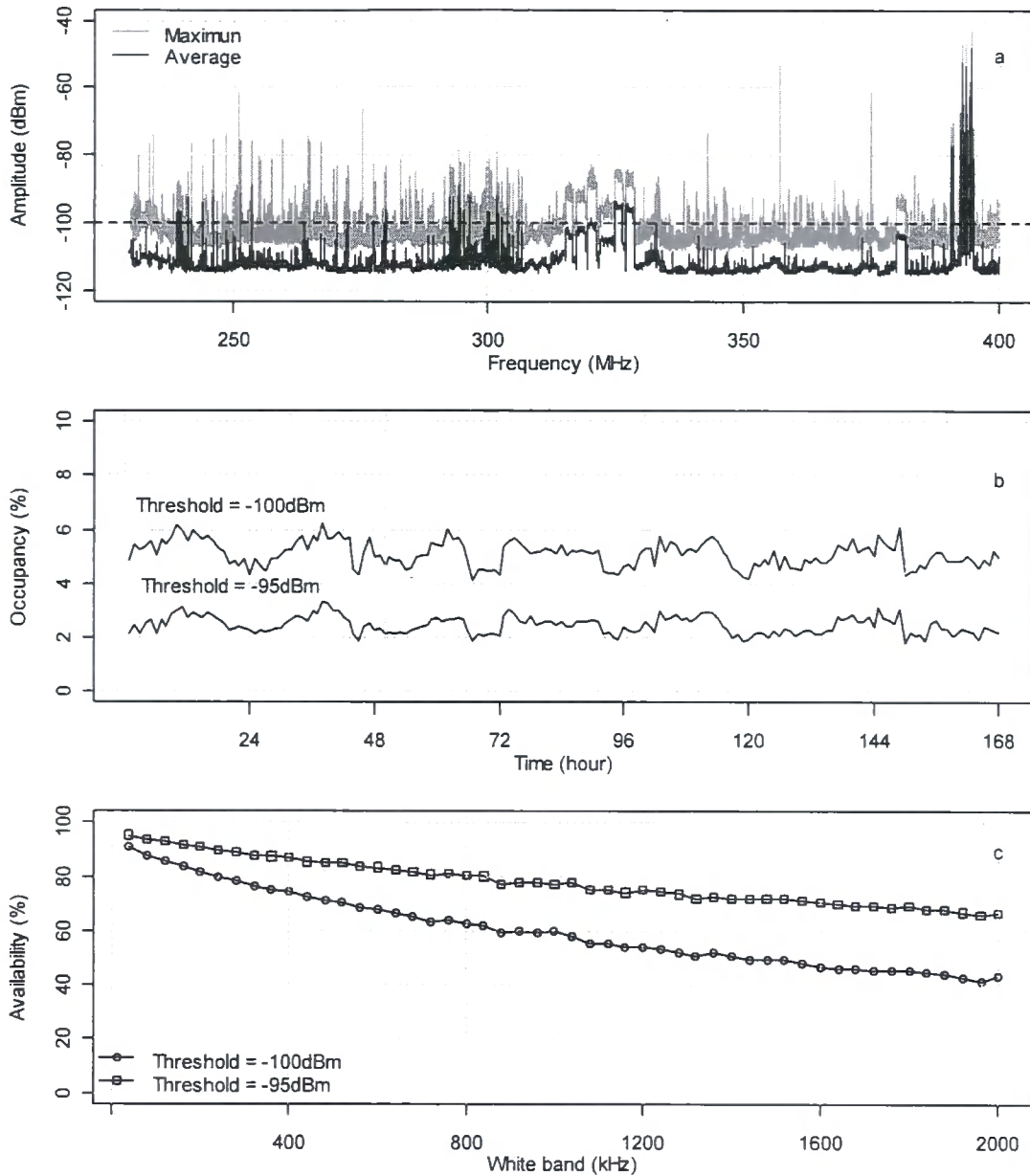


Band III 174-230 MHz

	-105 dBm	-95 dBm
Min	15.61	12.91
1 <sup>st</sup> Quantile	16.61	13.42
Median	16.99	13.60
Mean	16.97	13.60
3 <sup>rd</sup> Quantile	17.30	13.77
Max	18.36	14.37

Figure 4.6: Occupancy statistics of Band III

## 4.1 Occupancy analysis with descriptive Statistics

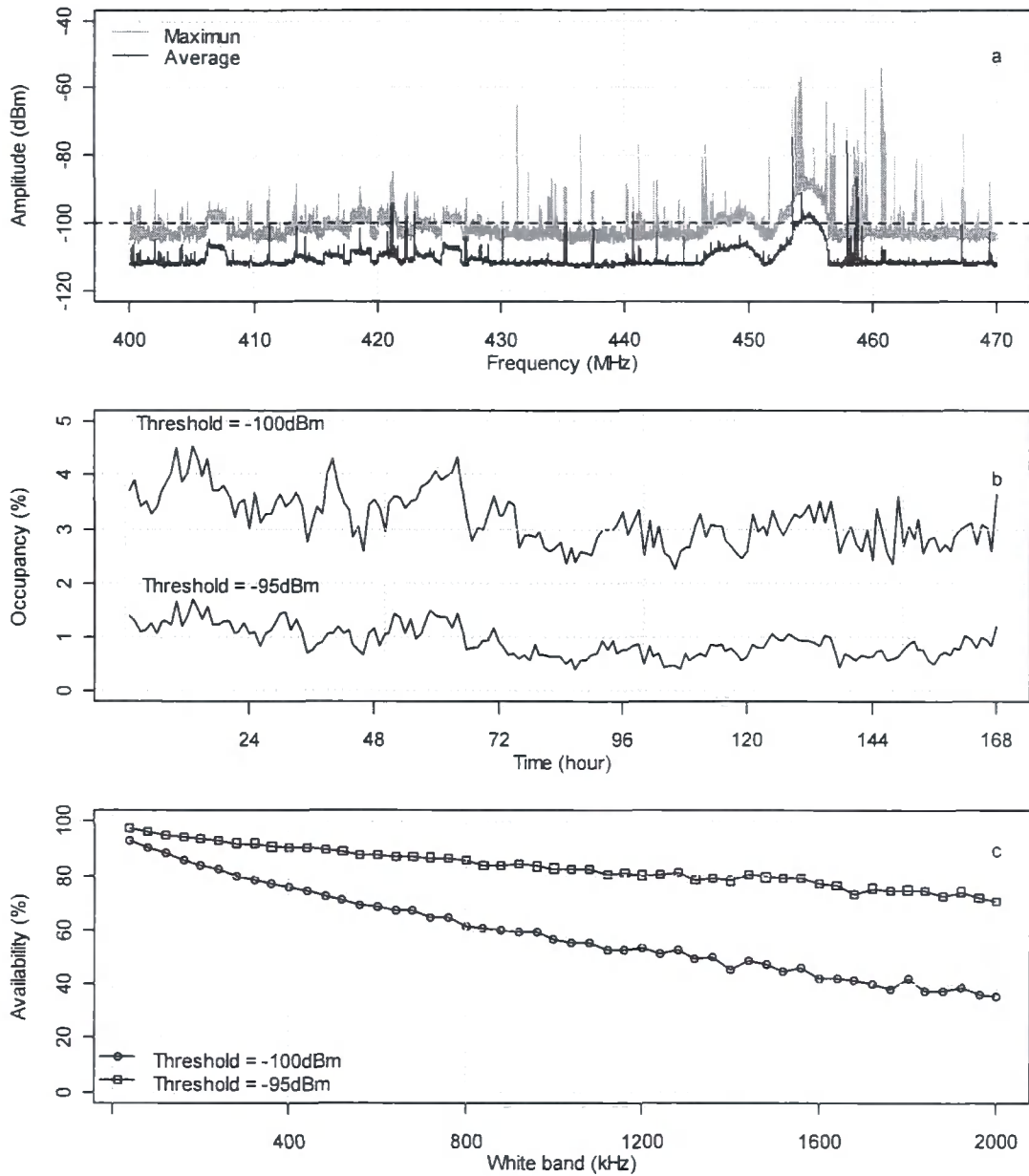


NATO Band 230-400 MHz

	-105 dBm	-95 dBm
Min	4.13	1.84
1 <sup>st</sup> Quantile	4.77	2.20
Median	5.14	2.41
Mean	5.10	2.46
3 <sup>rd</sup> Quantile	5.44	2.68
Max	6.19	3.29

Figure 4.7: Occupancy statistics of NATO Band

## 4.1 Occupancy analysis with descriptive Statistics

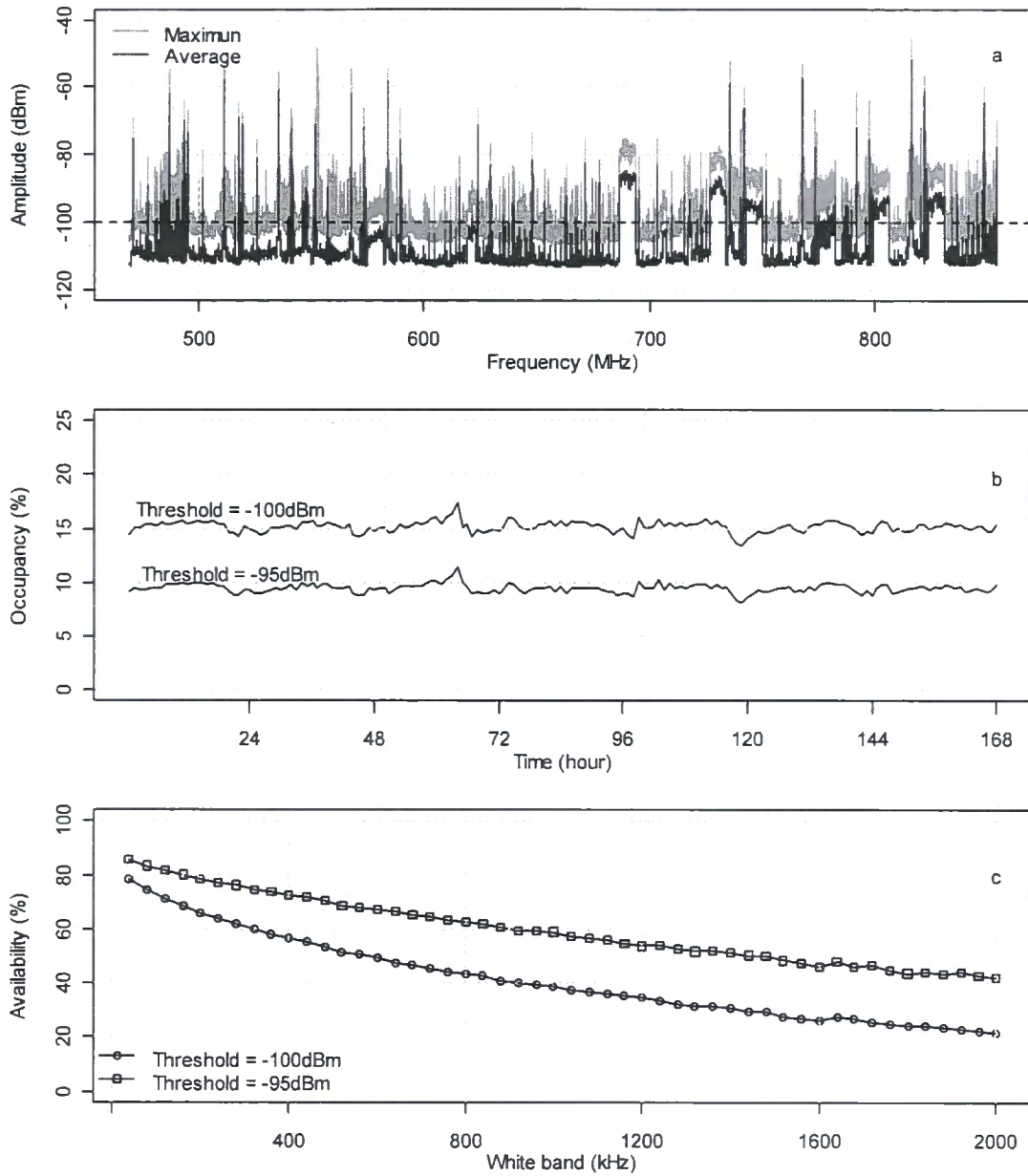


UHF 1 & 2 Band 400-470 MHz

	-105 dBm	-95 dBm
Min	2.25	0.40
1 <sup>st</sup> Quantile	2.81	0.70
Median	3.12	0.90
Mean	3.18	0.93
3 <sup>rd</sup> Quantile	3.51	1.11
Max	4.54	1.70

Figure 4.8: Occupancy statistics of UHF 1 & 2 Band

## 4.1 Occupancy analysis with descriptive Statistics

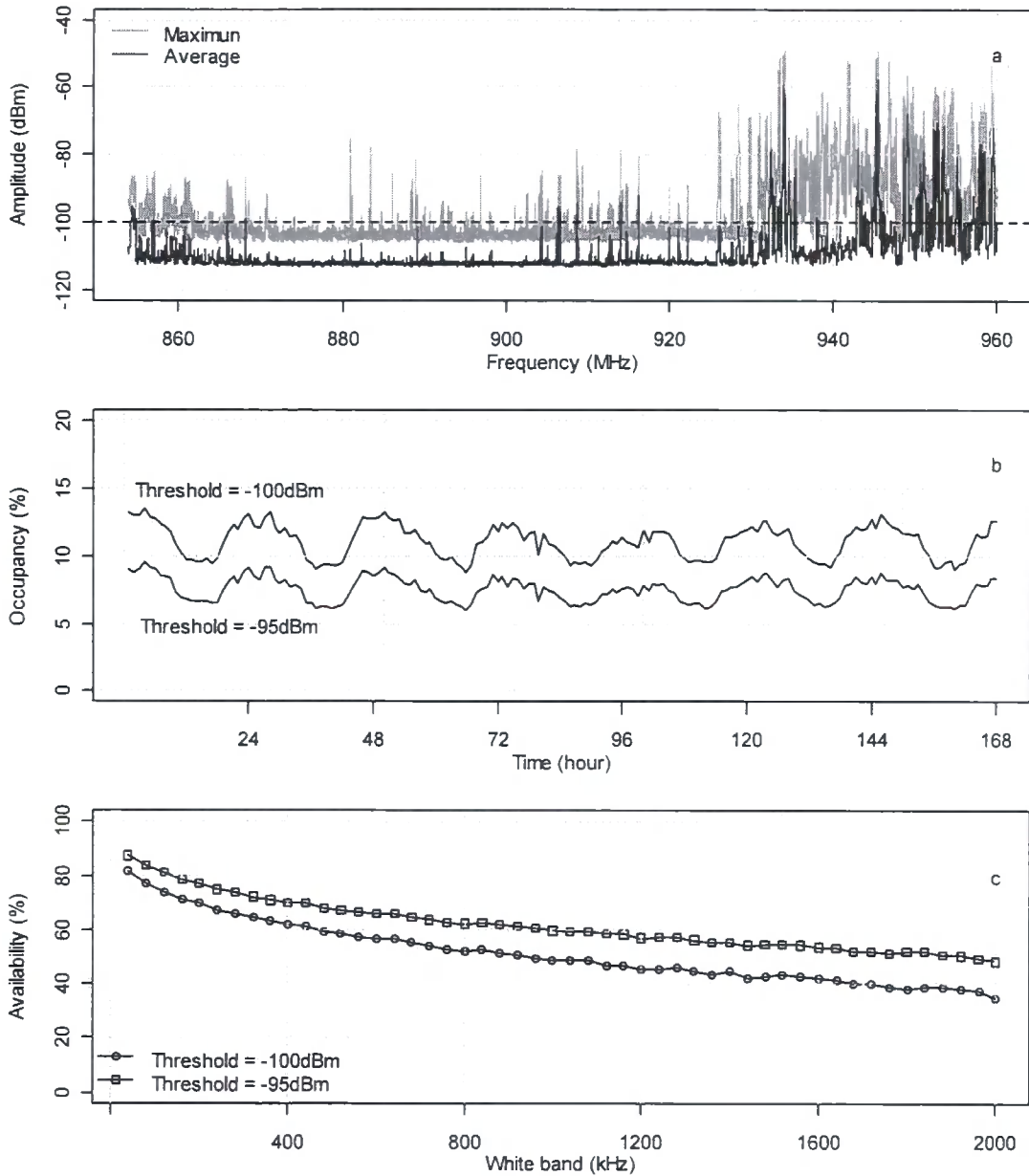


TV, Band IV & V 470-854 MHz

	-105 dBm	-95 dBm
Min	13.56	8.19
1 <sup>st</sup> Quantile	14.95	9.26
Median	15.34	9.48
Mean	15.26	9.49
3 <sup>rd</sup> Quantile	15.57	9.74
Max	17.51	11.58

Figure 4.9: Occupancy statistics of TV, Band IV & V

## 4.1 Occupancy analysis with descriptive Statistics

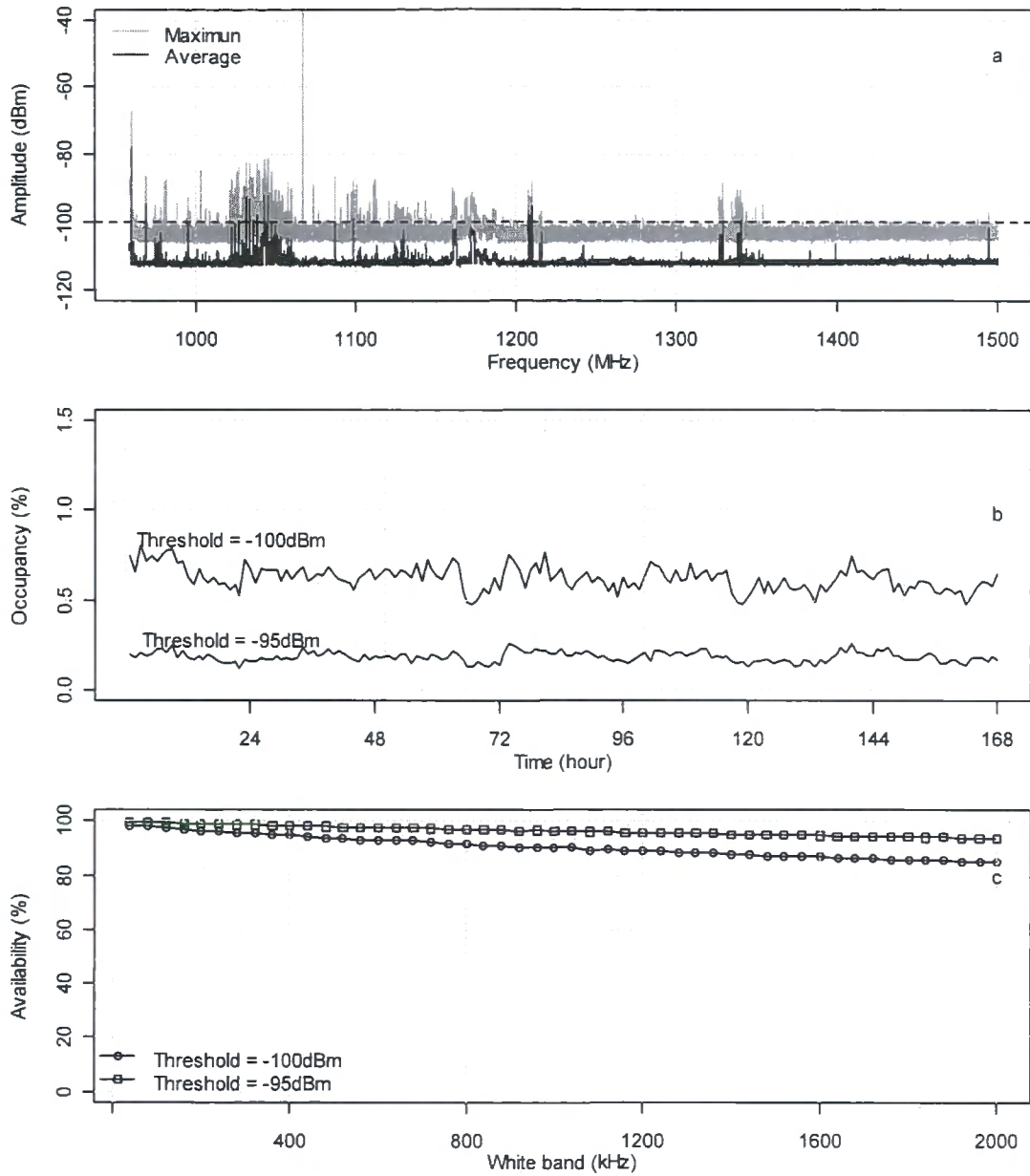


GSM & mobile 854-960 MHz

	-105 dBm	-95 dBm
Min	8.78	6.15
1 <sup>st</sup> Quantile	9.74	6.72
Median	11.29	7.76
Mean	11.1	7.68
3 <sup>rd</sup> Quantile	12.1	8.37
Max	13.65	9.80

Figure 4.10: Occupancy statistics of GSM & mobile Band

## 4.1 Occupancy analysis with descriptive Statistics

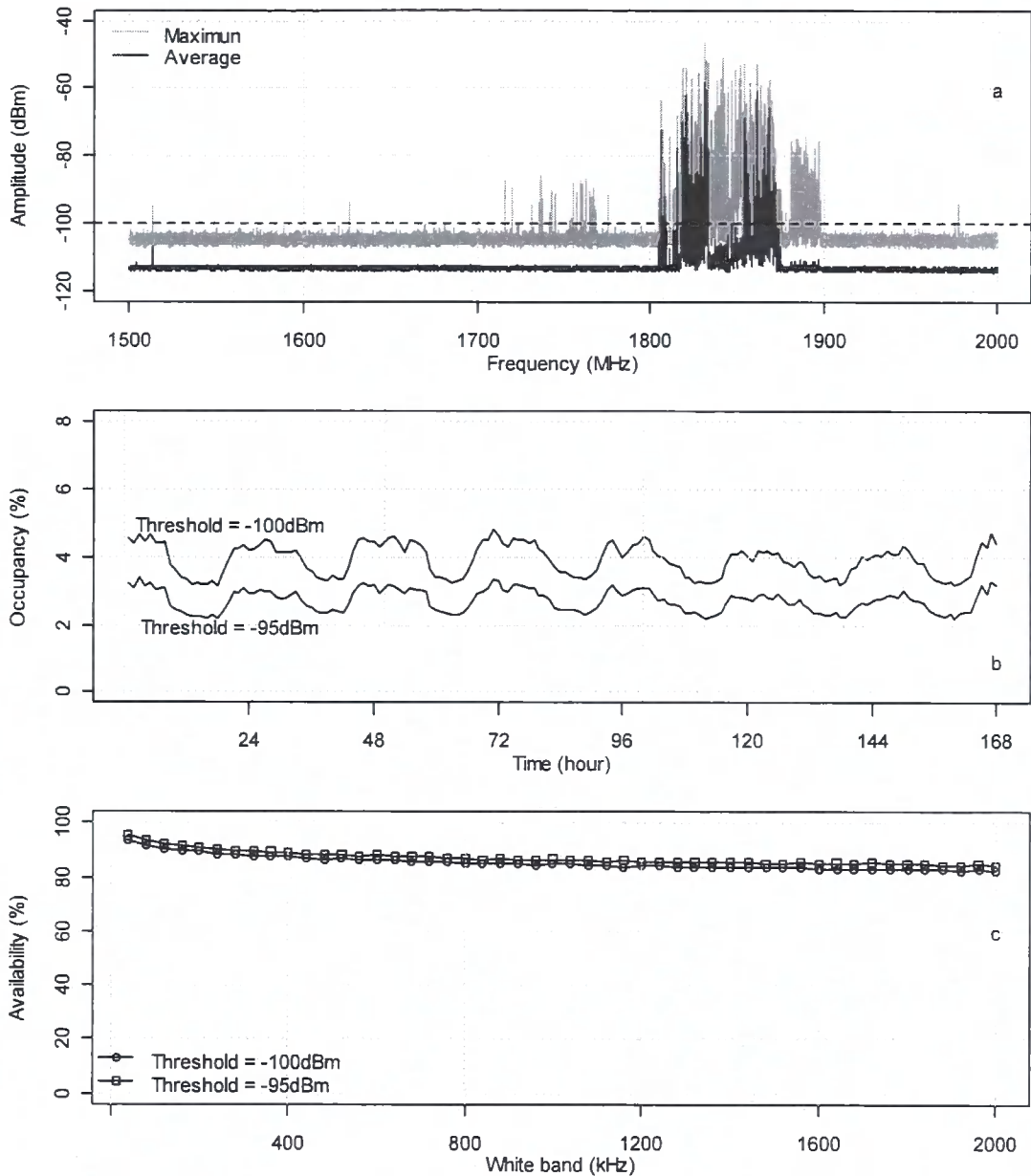


L Band (lower) 960-1500 MHz

	-105 dBm	-95 dBm
Min	0.40	0.06
1 <sup>st</sup> Quantile	0.52	0.11
Median	0.56	0.13
Mean	0.56	0.14
3 <sup>rd</sup> Quantile	0.59	0.16
Max	0.74	0.21

Figure 4.11: Occupancy statistics of L (lower) Band

## 4.1 Occupancy analysis with descriptive Statistics

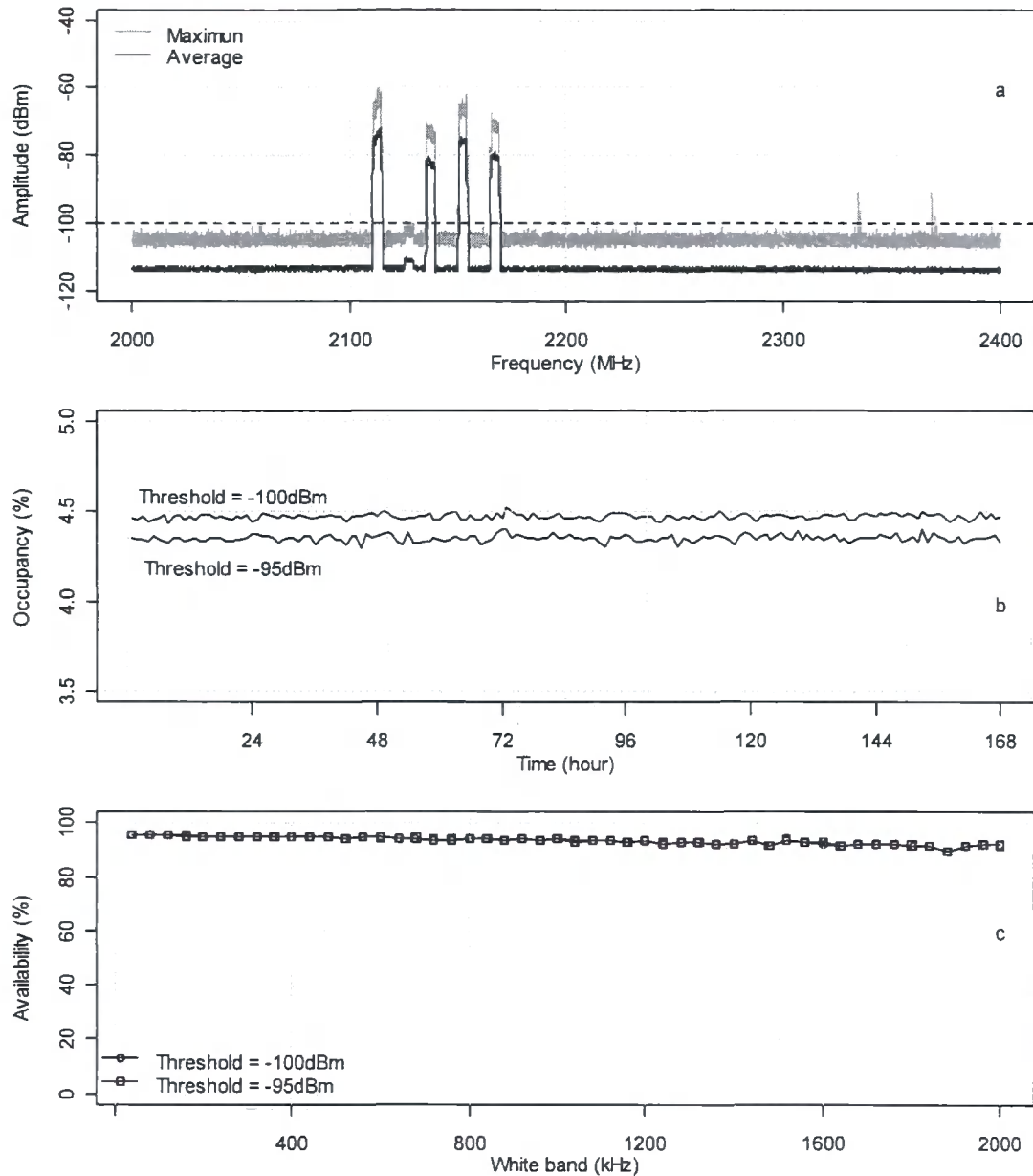


L Band (upper) 1500-2000 MHz

	-105 dBm	-95 dBm
Min	3.12	2.18
1 <sup>st</sup> Quantile	3.43	2.40
Median	3.99	2.75
Mean	3.92	2.72
3 <sup>rd</sup> Quantile	4.33	2.98
Max	4.82	3.34

Figure 4.12: Occupancy statistics of L (upper) Band

## 4.1 Occupancy analysis with descriptive Statistics



S Band (lower) 2000-2400 MHz

	-105 dBm	-95 dBm
Min	4.43	4.30
1 <sup>st</sup> Quantile	4.46	4.34
Median	4.48	4.35
Mean	4.47	4.35
3 <sup>rd</sup> Quantile	4.49	4.36
Max	4.52	4.41

Figure 4.13: Occupancy statistics of S (lower) Band

## 4.1 Occupancy analysis with descriptive Statistics

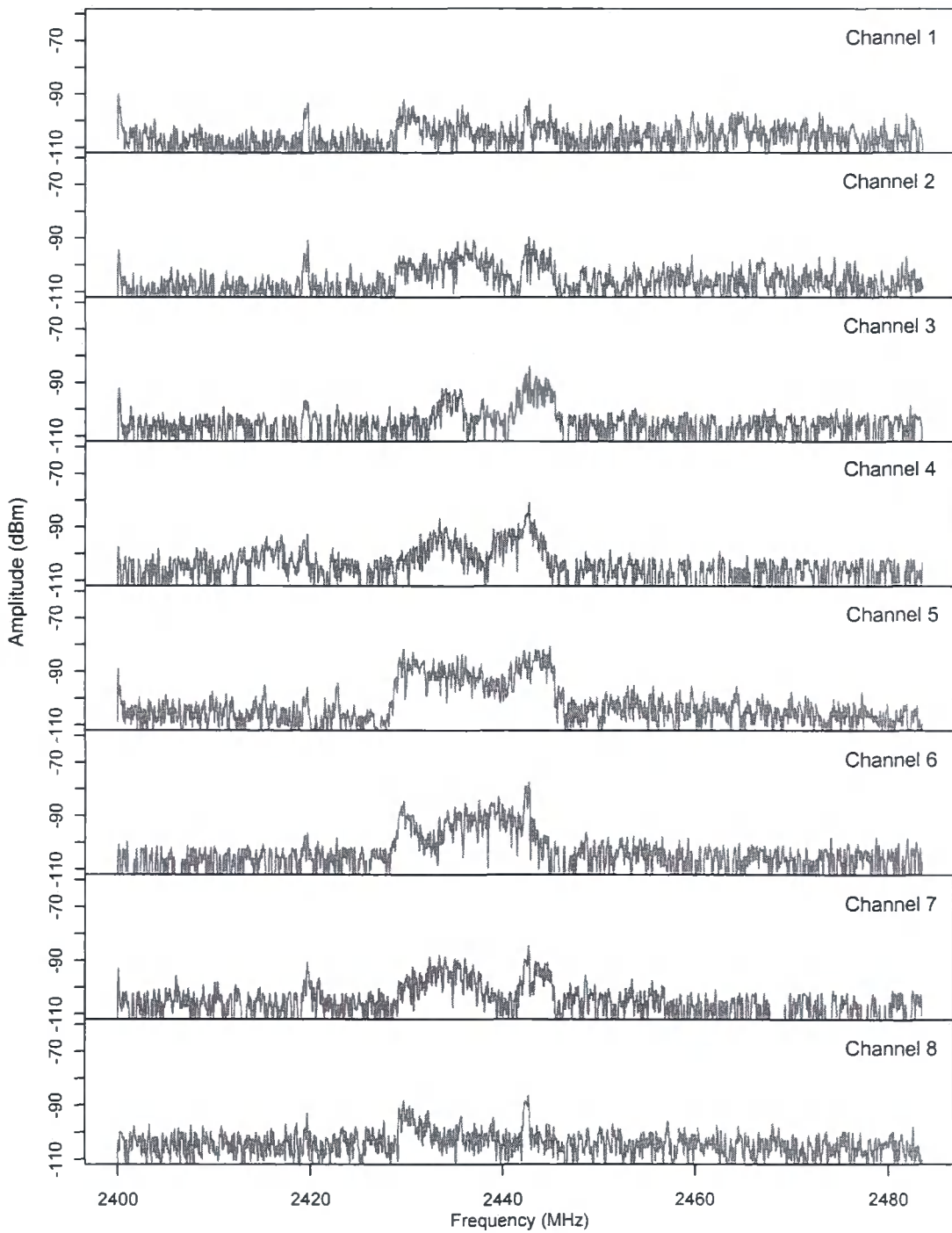


Figure 4.14: Snapshots of 2.4 GHz ISM band with Sounder

## 4.1 Occupancy analysis with descriptive Statistics

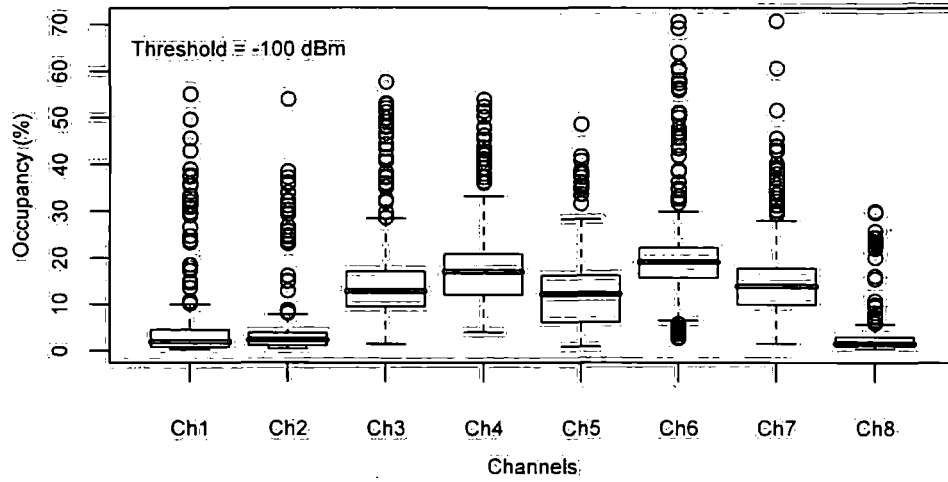


Figure 4.15: Channel occupancy statistics in location 1

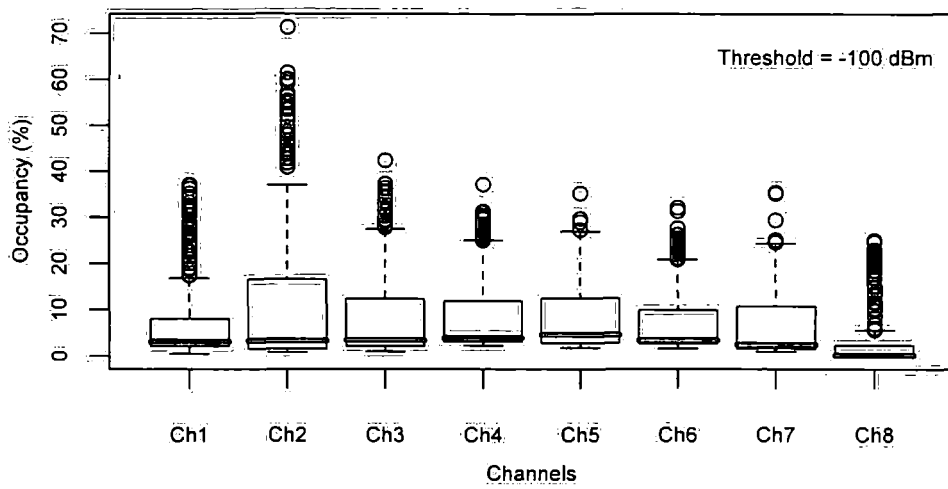


Figure 4.16: Channel occupancy statistics in location 2

## 4.1 Occupancy analysis with descriptive Statistics

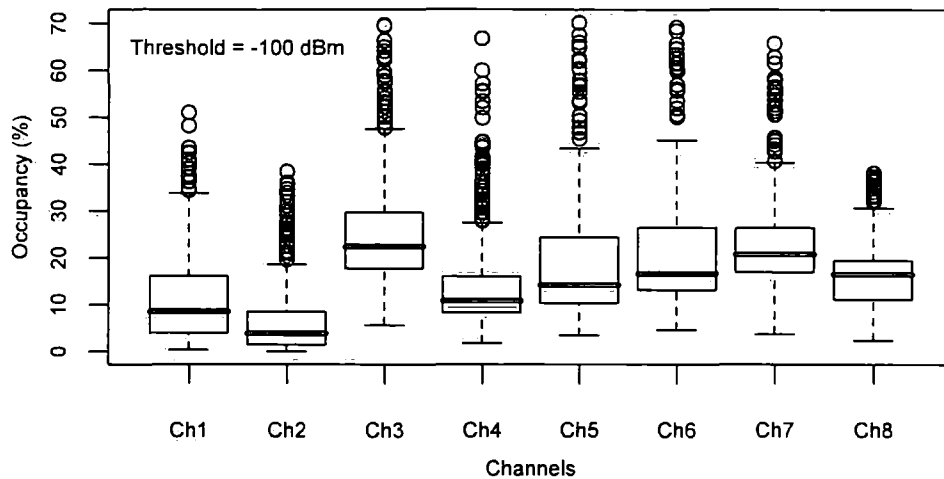


Figure 4.17: Channel occupancy statistics in location 3

The occupancy data range from 1500-2500 MHz were acquired with the multiple directional antennas. For example Figure 4.18 shows the white band distributions of the 1800-1900 MHz spectrum with different directions. The occupancy differences between the different azimuths suggest that the directional occupancy is an important parameter for cognitive radio.

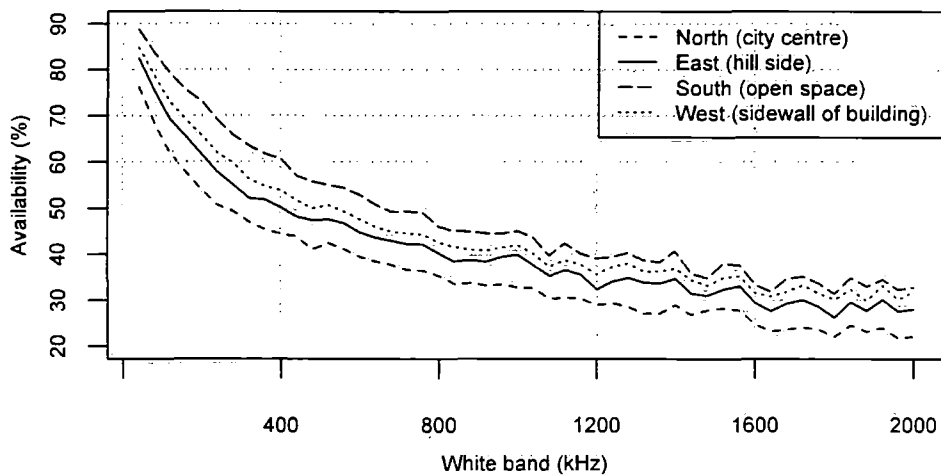


Figure 4.18: The azimuthal distributions of white band

The statistics presented in these figures shows that the spectrum occupancies are really sparse in the measuring locations and period. The average occupancies for most bands are less than 20%. Except the GSM and CDMA communication bands, the average occupancies in the 1 GHz to 2.5 GHz spectrum are less than 5%. The good propagation characteristics of this range of spectrum, in terms of propagating distance and data rates, make it an excellent candidate for cognitive radio technology. The data shows that, without additional spectrum, there are a great amount of spectrum resources for accommodating the cognitive radio systems if the current communication regulations could be changed.

## 4.2 Time Series Analysis

### 4.2.1 model approaches of spectrum occupancy

Statistical information on spectrum occupancy using graphics and tables provides general information by which the spectrum regulators can understand and plan channel and band usage as well as confirm the effectiveness of current planning and authorization activities.

However, static statistics intended to quantify the performance of a particular band and a particular measuring period usually cannot be extended directly to others. An alternative is to develop statistical models for more general estimation and prediction. This is the main motivation for modeling spectrum occupancy. In [50] the mathematical definition of spectrum occupancy on one variable occupancy model was given. The model has been extended in [22] to a three variable model with Markov processes. A generalized linear model to regress the congestion value on power amplitude, the number of sunspots and time in the HF band was introduced in [43]. We try to improve the model with the time series approach in this project.

Now, we define the frequency band and channels in a given band. A given frequency band is a set of frequencies which is set out for specific radio services. For example, the frequency band 87.5 ~ 108 MHz is for FM sound broadcasting and 470 ~ 590 MHz is for television broadcasting. The physical communication channel is a channel whose bandwidth is defined by the communication regulations

for transmitting and receiving messages. For instance, in the TV broadcasting band each channel occupies 8 MHz bandwidth. However, by this definition the bandwidths of the channel vary with different bands, which costs extra overhead in our application. In our project, the channel is defined by the bandwidth of the IF filter in the receiver. For example, if the bandwidth of the IF filter is set to 10 kHz, the total number of channels in the TV broadcasting band 470 ~ 590 MHz is  $(590000-470000)/10=12000$ .

We first define the occupancy of a selected channel as a two stage random process  $X(t)$  [50]. The first state  $X(t) = 1$  is labeled *occupied* and is defined as the event that, during an observation, the signal strength at a monitor receiver is above a given threshold; the complementary event  $X(t) = 0$  is that the signal strength is below this threshold. Figure 4.19 shows the occupancy of a given channel, at 2.44 GHz frequency with 3 kHz bandwidth, as a binary random process. Because the state of the channel is random, its state at any given measurement time cannot be predicted. However, its state can be described in terms of a probability law.

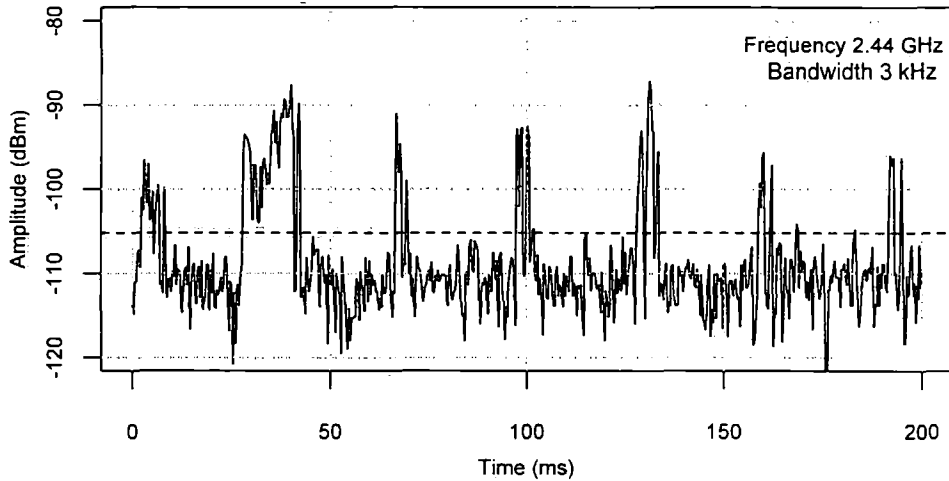


Figure 4.19: Occupancy of a single channel

Let us briefly restate the binomial and Bernoulli trials which form the basis of mathematical occupancy analysis. At a given time  $t_i$ , each sample of a given

channel can be regarded as a random variable  $X(t_i)$ . Define  $p[X(t_i) = 1] = p$ , terming  $p$  the probability of success. We can write

$$P[X(t_i) = x_i] = p^{x_i}(1 - p)^{1-x_i}, \quad x_i = 0, 1$$

where  $x_i$  is the observed binary value of the random variable  $X(t_i)$  [13].

In order to estimate the occupancy value of  $p$  for a given channel, a number of samples  $X(t_i)$   $i = 1, 2, \dots, N$  which can be assumed to have the same underlying success probability are taken. Let  $y$  be the total number of samples whose detected power exceeds the given threshold in the batch of  $N$  samples, and assume that the samples are independent of one another. The value of  $y$  can be regarded as the observed value of a random variable  $Y$ , associated with the total number of successes out of  $N$ . In view of the coding used for the values taken by each of the  $N$  random variables,  $X(t_1), X(t_2), \dots, X(t_N)$ , it follows that  $y = x_1 + x_2 + \dots + x_N$ , the sum of the grouped  $N$  binary observations, is the observed value of the random variable  $Y = X(t_1) + X(t_2) + \dots + X(t_N)$ . By probability theory, the probability of  $y$  in  $N$  observations is given by [41]

$$p[Y = y] = C_N^y p^y (1 - p)^{N-y} \text{ for } y = 0, 1, \dots, N$$

where

$$C_N^y = \frac{N!}{y!(N - y)!}$$

The random variable  $Y$  is said to have a binomial distribution. Therefore, we frame our problem to the case of estimating the probability of success in Bernoulli trials.

A straightforward estimate of the occupancy rate  $p$  is  $\hat{p} = y/N$ . As the number of samples  $N$  increases the estimated occupancy rate  $\hat{p}$  should be more reliable. An informative summary of the extent to which  $\hat{p}$  is reliable as an estimate of the true occupancy rate  $p$  is an interval estimate, termed a *confidence interval*. A *confidence interval* gives an estimated range of values which is likely to include an unknown population parameter. A *confidence interval* is associate with a particular *confidence level*, usually 0.90, 0.95 or 0.99.

Similarly, the occupancy of a selected band at a given time  $t$  can be defined as a random variable, which takes values between 0 and 1, that groups each channel

in the given band. The band occupancy estimate  $\hat{p}$  equals

$$\hat{p} = \frac{y}{N}$$

where  $y$  denotes the number of channels whose signal strength exceeds a given threshold and  $N$  denotes the total channel number in the band. Hence, the theory of Bernoulli trial can be used to analyze the band occupancy. Usually, people have more interest in the occupancy variations over a period of time rather than at specific times. A straightforward approach for this problem is time series analysis.

### 4.2.2 stationary time series

A time series is a sequence of observations which are often spaced at uniform time intervals. Time series analysis attempts to understand and identify the underlying context of the data points and to make forecasts.

The purpose of time series analysis is to draw inferences from such series. So, we can infer the general occupancy situation without monitoring the spectrum. Before doing this, it is necessary to set up a hypothetical probability model to represent the occupancy data. Having chosen a model, estimated parameters and checked for goodness of fit to the data it then becomes possible to use the fitted model to enhance the understanding of the mechanism generating the series. Before introducing the ideas of time series analysis, we outline general approaches to time series modeling [7].

- Plot the series and examine the main feature of the graph, checking in particular
  1. a trend
  2. a seasonal component
  3. any apparent sharp changes in behaviour
  4. any outlying observations
- Remove the trend and seasonal components to get stationary residuals. To achieve this goal it may sometimes be necessary to apply a preliminary transformation to the data

- Choose a model to fit the data
- Forecast the future

A time series model will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values in a series for a given time will be expressed as deriving in some way from past values. *Autocorrelation* and *partial autocorrelation* play important roles in time series analysis which partly and fully specified time series model appearance.

Now, we take the occupancy of the GSM down-link band range from 925–960 MHz as an example to introduce the time series analysis of occupancy data. Data were collected from 14:00 27/06/2007 to 13:00 03/07/2007 using a 10 kHz bandwidth with 1 hour interval with the spectrum analyzer. The total 35 MHz spectrum can be divided into 3500 channels with 10 kHz bandwidth. If we set the threshold as -100 dBm, the grouped channel occupancy data of the GSM down-link band is shown in Figure 4.20. The data show that low occupancy occurs

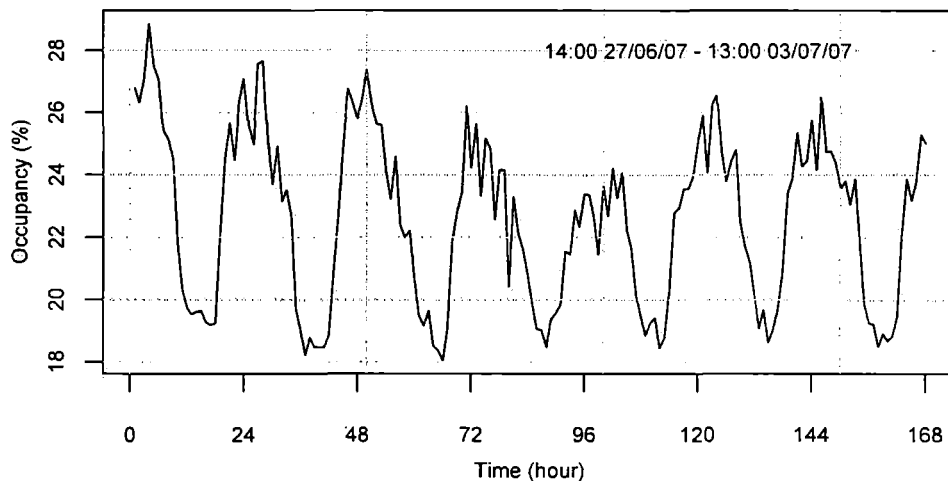


Figure 4.20: Time series of GSM band occupancy

on weekends. The 24 hour seasonal pattern is caused by the peak traffic of each

day in 10:00-18:00 local time; the flat off-peak occupancy is caused by signalling traffic [59; 60; 61].

An important and frequently used notion is stationarity of a time series. Loosely speaking, a time series  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  is said to be stationary if it has statistical properties similar to those of the time shifted series  $\{X_{t+h}, t = 0, \pm 1, \pm 2, \dots\}$ , for each integer  $h$ . Restricting attention to those properties which depend only on the first and second order statistics of  $\{X_t\}$ .

Let  $\{X_t\}$  be a time series with  $EX_t^2 \leq \infty$ . The *mean* function of  $\{X_t\}$  is

$$\mu_X(t) = E\{X_t\}$$

The *autocovariance* function of  $\{X_t\}$  is

$$\gamma_X(s, t) \triangleq \text{cov}(X_t, X_s) = E\{(X_t - \mu_X(t))(X_s - \mu_X(s))\} \quad (4.1)$$

for all integer  $s$  and  $t$ . A process  $\{X_t\}$  is called *wide-sense stationarity*<sup>1</sup> if

1.  $\mu_X(t)$  is independent of  $t$ .
2.  $\gamma_X(s, t)$  is independent of  $s$  and  $t$  and only is dependent of  $(s - t)$ .

So, to a stationary time series, the *autocovariance* can be expressed as one variable function  $\gamma_X(h)$  and the *autocorrelation* function of  $\{X_t\}$  is defined as

$$\rho_X(h) \triangleq \text{cor}(X_{t+h}, X_t) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

Inspecting all the occupancy data range from 100–2500 MHz, most of the data have stationary occupancy and some have a seasonal component with 24 hour period. The data may be represented as a realization of the classical decomposition model [6]

$$X_t = m_t + s_t + Z_t \quad (4.2)$$

where  $m_t$  is a slowly changing function known as a trend component,  $s_t$  is a function with known period  $d$  referred to as a seasonal component and  $Z_t$  is random noise which is stationary. Figure 4.21 shows the trend and seasonal

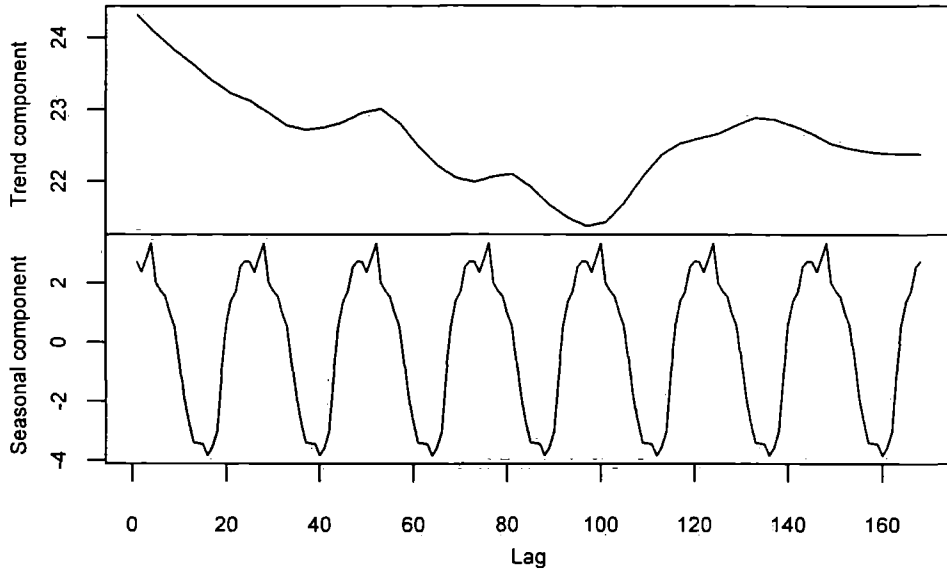


Figure 4.21: The trend and seasonal components of the GSM occupancy series

components of the GSM occupancy series. The data are obtained by the locally weighted scatterplot smoothing approach which details can be found in [35].

Our aim is to estimate and extract deterministic components  $m_t$  and  $s_t$  in the hope that the residual or noise component  $Z_t$  will turn out to be a stationary time series. We can then use the theory of such process to find a satisfactory probabilistic model for the process  $Z_t$ , to analyze its properties, and to use it in conjunction with  $m_t$  and  $s_t$  for purposes of prediction and description of  $\{X_t\}$ .

If seasonality is absent in a time series, the trend component can be removed by the difference operators, developed extensively by Box and Jenkins [5]. We define the difference operator  $\nabla$  by

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$$

where  $B$  is the *backward shift operator*,

$$BX_t = X_{t-1} \tag{4.3}$$

<sup>1</sup>Strict stationarity of a time series  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  is defined by the condition that  $(X_1, X_2, \dots, X_n) \doteq (X_{1+h}, X_{2+h}, \dots, X_{n+h})$  that is it has the same joint distributions for all integer values of  $h$  and  $n > 0$ .



The power of the operators  $B$  and  $\nabla$  is defined at  $B^i(X_t) = X_{t-i}$  and  $\nabla^i(X_t) = \nabla(\nabla^{i-1}(X_t))$ ,  $i \geq 1$  with  $\nabla^0(X_t) = X_t$ . Polynomials in  $B$  and  $\nabla$  are manipulated in precisely the same way as polynomial functions of a real variable. The lag- $d$  difference operator  $\nabla_d$  is defined as [6]

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t$$

This operator should not be confused with the operator  $\nabla^d = (1 - B)^d$ .

Applying the operator  $\nabla_d$  to the model (4.2)  $X_t = m_t + s_t + Z_t$ , where  $s_t$  has period  $d$ , we obtain

$$\nabla_d X_t = m_t - m_{t-d} + Z_t - Z_{t-d}$$

which gives a decomposition of the difference  $\nabla_d X_t$  into a trend component ( $m_t - m_{t-d}$ ) and a noise term ( $Z_t - Z_{t-d}$ ). The trend  $m_t - m_{t-d}$  can be eliminated using the methods already described. Figure 4.22 shows the result of applying the operator  $\nabla_{24}$  to the GSM band occupancy series shown in Figure 4.20. The seasonal component evident in Figure 4.20 is absent from the series  $\{\nabla_{24} X_t, t = 24, \dots, 168\}$ .

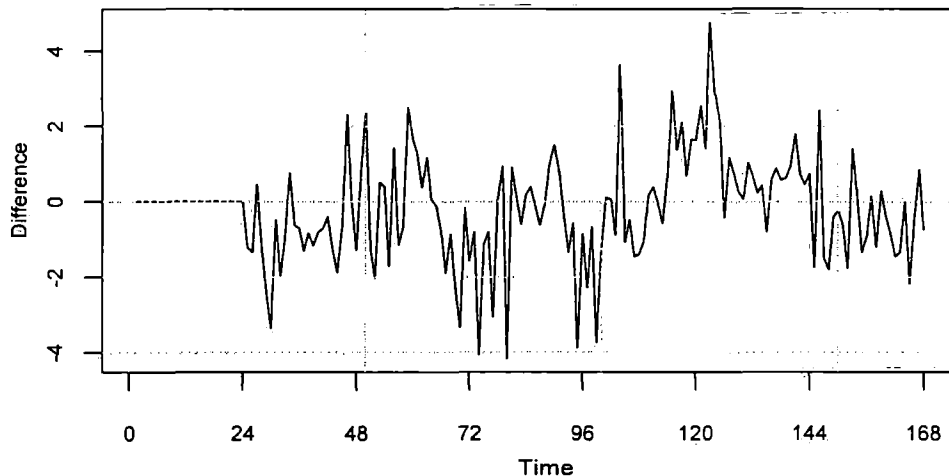


Figure 4.22: The difference series  $\{\nabla_{24} X_t, t = 24, \dots, 168\}$

### 4.2.3 ARMA model identification

In statistics, autoregressive moving average (ARMA) models, sometimes called Box-Jenkins models, play an important role in time series analysis. Before defining ARMA models, we first introduce the white noise series. A process  $\{Z_t\}$  is said to be white noise [6] with zero mean and  $\sigma^2$  variance, written as

$$\{Z_t\} \sim WN(0, \sigma^2)$$

if and only if  $\{Z_t\}$  has zero mean and has covariance function (4.1) given by

$$\gamma(h) = \begin{cases} \sigma^2, & h = 0 \\ 0, & h \neq 0 \end{cases}$$

The ARMA( $p, q$ ) is defined by [6] as if a process  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  is stationary and if for every  $t$ ,

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (4.4)$$

where  $\{Z_t\} \sim WN(0, \sigma^2)$ . Equation (4.4) can be written symbolically in the more compact form

$$\phi(B)X_t = \theta(B)Z_t \quad t = 0, \pm 1, \pm 2, \dots$$

where  $\phi$  and  $\theta$  are the  $p^{\text{th}}$  and  $q^{\text{th}}$  degree polynomials

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p$$

and

$$\theta(x) = 1 + \theta_1 x + \theta_2 x^2 + \dots + \theta_q x^q$$

and  $B$  is the the *backward shift* operator defined in (4.3). The polynomials  $\phi$  and  $\theta$  will be referred to as the *autoregressive* and *moving average* polynomials respectively.

If  $\phi(x) \equiv 1$  then (4.4) becomes  $X_t = \theta(B)Z_t$  and the process is said to be a *moving average* process MA( $q$ ) of order  $q$ . The solution  $\{X_t\}$  is a stationary process since

$$E\{X_t\} = \sum_{i=0}^q \theta_i E\{Z_{t-i}\} = 0$$

and

$$\text{cov}(X_{t+h}, X_t) = \begin{cases} \sigma^2 \sum_{i=0}^{q-|h|} \theta_i \theta_{i+|h|}, & |h| \leq q \\ 0, & |h| > q \end{cases}$$

On the other hand, if  $\theta(x) \equiv 1$  then (4.4) becomes  $\phi(B)X_t = Z_t$  and the process is said to be an *autoregressive* process of order  $p$  AR( $p$ ).

An ARMA( $p, q$ ) process  $\{X_t\}$  is *causal*, or a *causal* function of  $\{Z_t\}$ , if the polynomials  $\phi(\cdot)$  and  $\theta(\cdot)$  have no common zeros and there exist constants  $\{\Psi_i\}$  such that  $\sum_{i=0}^{\infty} |\Psi_i| < \infty$  and

$$X_t = \sum_{i=0}^{\infty} \Psi_i Z_{t-i} \quad t = 0, \pm 1, \pm 2, \dots$$

*Causality* is equivalent to the condition

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p \neq 0$$

for all  $|x| \leq 1$ . Another terminology which will be introduced is *invertibility*. An ARMA( $p, q$ ) process  $\{X_t\}$  is *invertible* if polynomials  $\phi(\cdot)$  and  $\theta(\cdot)$  have no common zeros and there exist constants  $\{\pi_i\}$  such that  $\sum_{i=0}^{\infty} |\pi_i| < \infty$  and

$$Z_t = \sum_{i=0}^{\infty} \pi_i X_{t-i} \quad t = 0, \pm 1, \pm 2, \dots$$

*invertibility* is equivalent to the condition [5]

$$\theta(x) = 1 + \theta_1 x + \theta_2 x^2 + \dots + \theta_q x^q \neq 0$$

for all  $|x| \leq 1$ . *Causality* and *invertibility* are properties not of  $\{X_t\}$  alone, but rather of the relationship between the two processes  $\{X_t\}$  and  $\{Z_t\}$  in the defining ARMA( $p, q$ ) (4.4).

We have already discussed the importance of the class of ARMA( $p, q$ ) models for representing stationary series. However, many empirical time series behave as though they have no fixed mean, such as the GSM occupancy time series. Even so, they exhibit homogeneity in the sense that apart from the local level, one part of the series behaves much like any other part. Models that describe such homogeneous non-stationary behavior can be obtained by applying some suitable

difference of the process to be stationary. We now consider a generalization of the ARMA( $p, q$ ) processes, which reduces to ARMA( $p, q$ ) models after limited differencing.

If  $d$  is a non-negative integer, then  $\{X_t\}$  is said to be an *autoregressive integrated moving average* ARIMA( $p, d, q$ ) process if  $Y_t \equiv \nabla^d X_t = (1 - B)^d X_t$  is a causal ARMA( $p, q$ ) process [6]. This definition means that  $\{X_t\}$  satisfies a difference equation of the form

$$\phi^*(B)X_t \equiv \phi(B)(1 - B)^d X_t = \theta(B)Z_t \quad Z_t \sim WN(0, \sigma^2) \quad (4.5)$$

where  $\phi(\cdot)$  and  $\theta(\cdot)$  are polynomials of degrees  $p$  and  $q$ , respectively, and  $\phi(x) \neq 0$  for  $|x| \leq 1$ . The polynomial  $\phi^*(x)$  has a zero of order  $d$  at  $x = 1$ . The process  $\{X_t\}$  is stationary if and only if  $d = 0$ , in which case it reduces to an ARMA( $p, q$ ) process.

For example, after eliminating the seasonal component of the original series, the data in Figure 4.22 still have non-stationary components. The correlation in the difference series shows that the data points above and below zero are clustered. If we apply the first order difference operator  $\nabla^1 = (1 - B)^1$  to  $\{\nabla_{24} X_t, t = 24, \dots, 168\}$  and plot the resulting difference series  $\{\nabla \nabla_{24} X_t\}$ , we obtain the graph shown in Figure 4.23<sup>1</sup>. The difference series  $\{\nabla \nabla_{24} X_t\}$  can be analyzed by the stationary ARMA( $p, q$ ) theories.

In practice it may not be reasonable to assume that the seasonality component repeats itself precisely in the same way cycle after cycle. Seasonal ARIMA models allow for randomness in the seasonal pattern from one cycle to the next [6]. Suppose we tabulate the GSM occupancy data as in Table 4.1. Each column in this table may itself be viewed as a realization of a time series. Suppose that each one of these twenty four time series is generated by the same ARIMA( $P, Q$ ) model, or more specifically that the series corresponding to the  $i^{th}$  day,  $X_{i+24t}$ ,  $t = 0, 1, \dots, n - 1$ , satisfies a difference equation of the form,

$$\begin{aligned} X_{i+24t} - \Phi_1 X_{i+24(t-1)} - \dots - \Phi_P X_{i+24(t-P)} \\ = U_{i+24t} + \Theta_1 X_{i+24(t-1)} + \dots + \Theta_Q X_{i+24(t-Q)} \end{aligned}$$

<sup>1</sup>There are still seasonal components in the difference series  $\{\nabla \nabla_{24} X_t\}$ . However, over difference will cost the model explicitness.

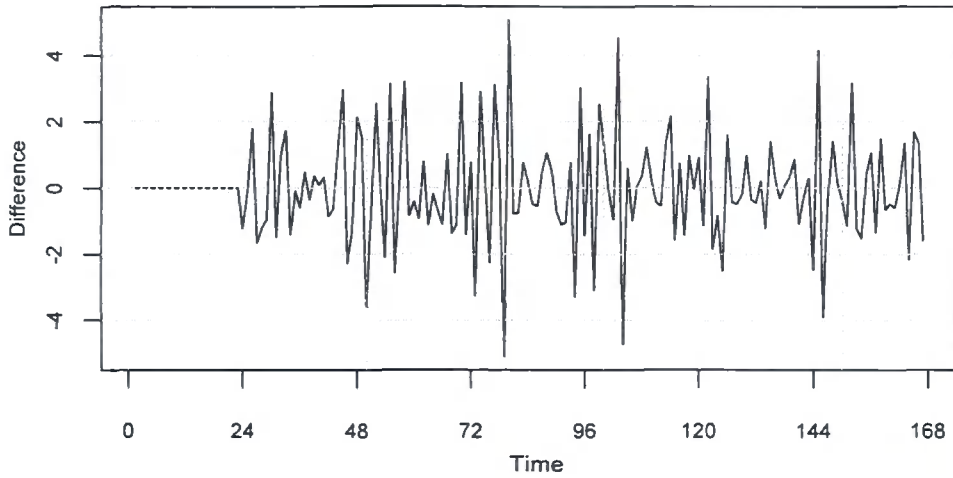


Figure 4.23: The difference series  $\{\nabla\nabla_{24}X_t\}$

Day	1	2	...	24
1	$X_1$	$X_2$	...	$X_{24}$
2	$X_{25}$	$X_{26}$	...	$X_{48}$
3	$X_{49}$	$X_{50}$	...	$X_{72}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	$X_{1+24(n-1)}$	$X_{2+24(n-1)}$	...	$X_{24+24(n-1)}$

Table 4.1: Tabulate the seasonal GSM occupancy date

for each  $i = 1, 2, \dots, 24$ , where  $\{U_{i+24t}\} \sim WN(0, \sigma_U^2)$ . Because of the interaction between the two models describing the between-day and the between-hour dependence structure, the seasonal ARIMA model can be quite complicated [5].

We have already seen in the GSM occupancy data how differencing the series at lag  $s = 24$  is a convenient way of eliminating a seasonal component of period  $s$ . If we fit an ARMA( $p, q$ ) model  $\phi(B)Y_t = \theta(B)Z_t$  to the differenced series  $Y_t = \nabla_s X_t = (1 - B^s)X_t$ , then the model for the original series is  $\phi(B)(1 - B^s)X_t = \theta(B)Z_t$ . This is a special case of the general seasonal ARIMA (SARIMA) model defined as follows. If  $d$  and  $D$  are nonnegative integers, then  $\{X_t\}$  is a seasonal ARIMA( $p, d, q$ ) $\times$ ( $P, D, Q$ ) $_s$  process with period  $s$  if the differenced series

$Y_t = \nabla^d \nabla_s^D X_t = (1 - B)^d (1 - B^s)^D X_t$  is a causal ARMA( $p, q$ ) process defined by

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t \quad Z_t \sim WN(0, \sigma^2) \quad (4.6)$$

where  $\phi(x) = 1 - \phi_1 x - \dots - \phi_p x^p$ ,  $\Phi(x) = 1 - \Phi_1 x - \dots - \Phi_P x^P$ ,  $\theta(x) = 1 - \theta_1 x - \dots - \theta_q x^q$  and  $\Theta(x) = 1 - \Theta_1 x - \dots - \Theta_Q x^Q$  [7].

If the parameters in the seasonal ARIMA model were known in advance this would be a straightforward application of the estimation techniques such as Yule-Walker algorithm and the innovations algorithm [7]. However, this is usually not the case, so appropriate values of the parameters should be identified. Model identification methods are rough procedures applied to a set of data to indicate the kind of representational model that is worthy of further research. The specific aim of the model identification is to obtain some idea of the values of  $p, d, q, P, D, Q$  and  $s$  needed in the seasonal ARIMA model and to obtain initial estimates for the parameters. Graphical methods are particularly useful in the model identification stage.

Seasonality and trend are usually detected by inspecting the graph of the series. However, they are also characterized by *autocorrelation* functions that are slowly decaying and nearly periodic [8]. Figure 4.24 shows the periodic *autocorrelation* function of the GSM band occupancy with 95% confidence limits<sup>1</sup>. It turns out that the tendency for the *autocorrelation* function does not die out quickly which might suggest non-stationarity, as Figure 4.20 shows the seasonality. So the difference operation should be applied to the series. After eliminating the seasonal component by applying the  $\nabla_{24}$  operator and applying  $\nabla$  on the differenced series, Figure 4.25 shows the *autocorrelation* function of the difference series  $\{\nabla \nabla_{24} X_t\}$ , where the tendency drops fairly rapidly. In Figure 4.25 the strong negative autocorrelation coefficient at the seasonal period 24 suggests to add a seasonal MA term  $Q = 1$  to the model. Generally in fitting seasonal time series models, it is recommended not to mix seasonal AR and seasonal MA terms in the same model, and avoid using more than one of either kind.

<sup>1</sup> Under the null hypothesis: 'The *autocorrelation* value at lag  $k$  has no statistical significance departed from zero at 95% confidence level.' The definition and descriptions of the statistical tests can be found in [41].

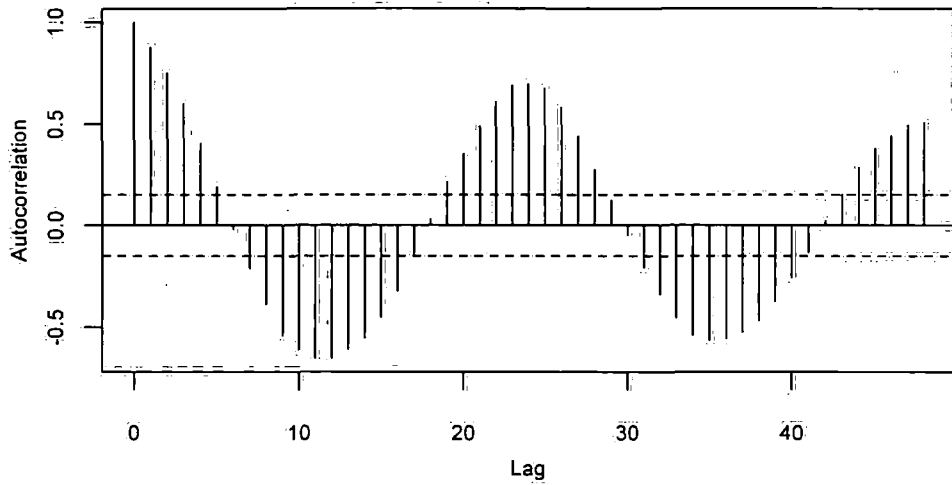


Figure 4.24: The *autocorrelation* function of the GSM band occupancy

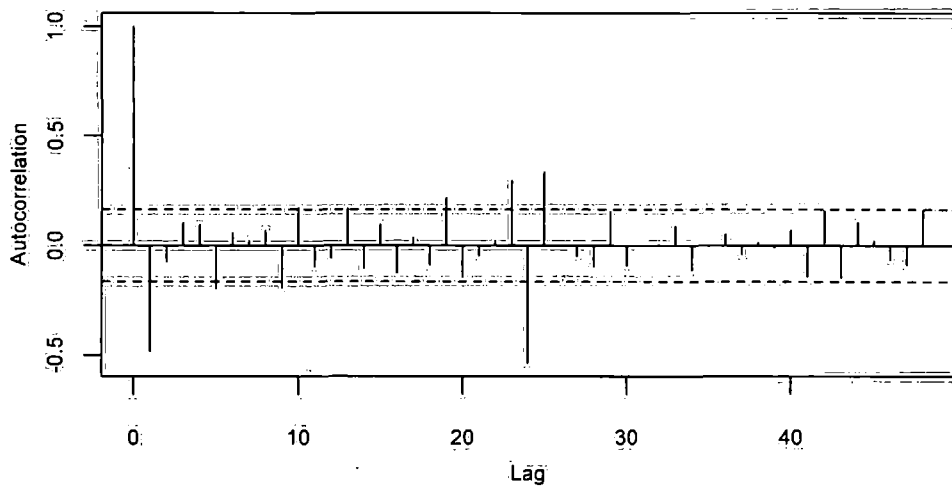


Figure 4.25: The *autocorrelation* function of the difference series  $\{\nabla\nabla_{24}X_t\}$

The *partial autocorrelation* function, like the *autocorrelation* function, delivers vital information regarding the dependence structure of a stationary process. A *partial autocorrelation* coefficient of order  $k$  measures the strength of correlation among pairs of entries in the time series while removing the effects of all autocorrelations below order  $k$ . The *partial autocorrelation*  $\alpha(k)$  at lag  $k$  may be regarded as the correlation between  $X_1$  and  $X_{k+1}$ , adjusted for the intervening observations  $X_2, \dots, X_k$ . *Partial autocorrelations* are useful in identifying the order of an autoregressive model while *autocorrelations* are useful in finding the order of a moving average model. For example, the *partial autocorrelation* of an AR(p) process is zero after lag  $p+1$ . The *partial autocorrelation*  $\alpha(\cdot)$  of a stationary time series is defined by [5]

$$\alpha(1) = \text{cor}(X_2, X_1) = \rho(1)$$

and

$$\alpha(k) = \text{cor}(X_{k+1} - P_{sp\{1, X_2, \dots, X_k\}} X_{k+1}, X_1 - P_{sp\{1, X_2, \dots, X_k\}} X_1) \quad k \geq 2$$

where  $P_{sp\{1, X_2, \dots, X_k\}} X_{k+1}$  is defined as the projection  $P_{sp\{1, X_2, \dots, X_k\}}(X) = \sum_{i=0}^k \alpha_i X_i$ . Figure 4.26 shows the *partial autocorrelation* function of the difference series  $\{\nabla \nabla_{24} X_i\}$ .

After a time series has been stationarized by differencing, the next step in fitting the model is to determine whether AR or MA terms are needed to correct any autocorrelation that remains in the differenced series. By looking at the *autocorrelation* function and *partial autocorrelation* function plots of the differenced series, one can roughly identify the numbers of AR and/or MA terms that are needed.

If the *partial autocorrelation* function displays a sharp cutoff while the *autocorrelation* decays more slowly, we say that the stationarized series displays an *AR signature*, meaning that the autocorrelation pattern can be explained more easily by adding AR terms than by adding MA terms. In principle, any autocorrelation pattern can be removed from a stationarized series by adding enough autoregressive terms to the forecasting equation, and the *partial autocorrelation* function tells us how many such terms are likely be needed. However, this is not

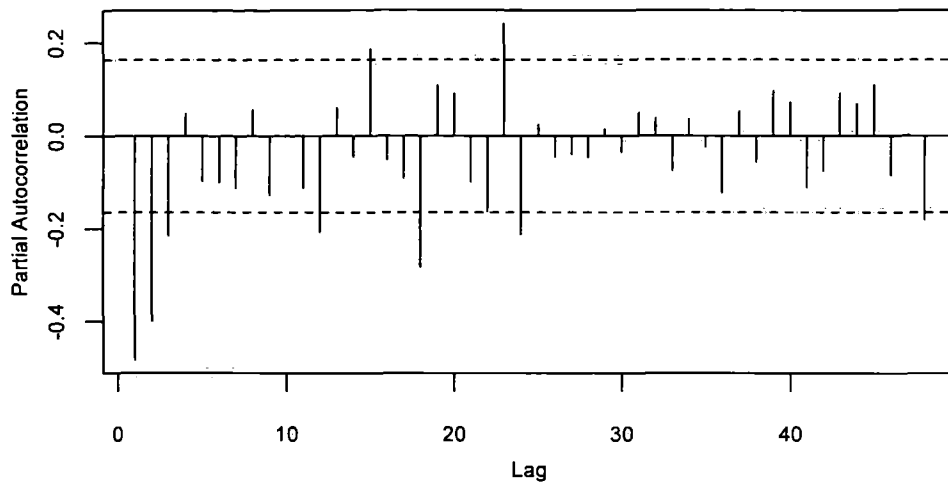


Figure 4.26: The *partial autocorrelation* function of the difference series  $\{\nabla\nabla_{24}X_t\}$

always the simplest way to explain a given pattern of autocorrelation. Sometimes it is more efficient to add MA terms instead. The *autocorrelation* function plays the same role for MA terms that the *partial autocorrelation* function plays for AR terms—that is, the *autocorrelation* function tells us how many MA terms are likely to be needed to remove the remaining autocorrelation from the differenced series. If the autocorrelation is significant at lag  $k$  but not at any higher lags, or if the *autocorrelation* function cuts off at lag  $k$ , this indicates that exactly  $k$  MA terms should be used in the forecasting equation. In the latter case, we say that the stationarized series displays a *MA signature*, meaning that the autocorrelation pattern can be explained more easily by adding MA terms than by adding AR terms [42].

Inspecting Figures 4.25–4.26, the *autocorrelation* function of the differenced GSM occupancy series displays a sharp cutoff and the lag-1 autocorrelation is negative, which suggests adding a MA term to the model. Meanwhile, since the lag at which the *autocorrelation* function cuts off is the indicated number of MA terms, the ARIMA(0, 1, 1) was adopted of the de-seasonal series. The signature of seasonal *autocorrelation* function and seasonal *partial autocorrelation* function behavior are similar to the signature of *autocorrelation* function and

seasonal *partial autocorrelation* function behavior in ARMA models, except that the pattern appears across multiples of lag  $s=24$  in the *autocorrelation* function and *partial autocorrelation* function. From figures 4.25–4.26, we can identify  $P = 0$  and  $Q = 1$ . So, the GSM occupancy series can be identified as a seasonal ARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>24</sub> model.

#### 4.2.4 model estimation and diagnostics

The identification process having led to a tentative formulation for a model, the parameters  $\vec{\phi} = (\phi_1, \dots, \phi_p)$ ,  $\vec{\theta} = (\theta_1, \dots, \theta_q)$  and the noise variance  $\sigma^2$  need to be estimated efficiently. After the parameters have been estimated, the fitted model will be subjected to the diagnostic check and test of goodness of fit.

The estimation of ARMA parameters in practice is not straightforward. Since it is not our intention to go too far into the details of estimation algorithms, only some important issues and problems are considered. Fortunately, there are many computer algorithms and computer programs existing for ARMA estimation. In our project, R language [35] has been chosen to analyze the statistical data.

To estimate the initial values there are a number of alternative preliminary estimation procedures. For general cases, good estimators of  $\vec{\phi}$  and  $\vec{\theta}$  can be obtained by imagining the data to be observations of a Gaussian time series and maximizing the likelihood with respect to the  $p + q + 1$  parameters  $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$  and  $\sigma^2$ . Suppose that  $\{X_t\}$  is a Gaussian time series with zero mean and *autocorrelation* function  $\gamma_X(i, j) = E\{X_i X_j\}$  defined by (4.1). The *likelihood* of  $\{X_t\}$  is

$$L(\vec{\phi}, \vec{\theta}, \sigma^2) = (2\pi)^{-n/2} (\det \Gamma_n)^{-1/2} \exp\left(-\frac{1}{2} \vec{X}' \Gamma_n \vec{X}\right) \quad (4.7)$$

where  $\Gamma_n$  denotes the covariance matrix,  $\Gamma_n = E\{\vec{X}_n \vec{X}_n'\} = [\gamma(i, j)]_{i,j=1}^n$

There is no advantage from a forecasting point of view to choose  $p$  and  $q$  arbitrarily large. Fitting a very high order model will generally result in a small estimated variance, but when using this over fitted model for forecasting, the mean squared error will depend not only on the white noise variance of the fitted model but also on errors arising from estimation of the parameters of the model [5]. For this reason we need to introduce a penalty factor to discourage the fitting of models with too many parameters. Although the final selection of the

p	d	q	P	D	Q	AICC
0	1	1	0	1	0	1.655
0	1	1	0	1	1	0.996
0	1	1	1	1	1	1.013
1	1	1	0	1	0	1.666
1	1	1	0	1	1	1.002

Table 4.2: The AICC values of different models

model depends on a variety of goodness criteria, it can be systematized to a large degree by the use of the Akaike's information criterion (AIC) and its bias-corrected Akaike's Information Corrected Criterion (AICC). The AICC Criterion [6] states that we choose  $p, q, \vec{\phi}_p$  and  $\vec{\theta}_q$  to minimize

$$AICC = -2 \ln L(\vec{\phi}, \vec{\theta}, \sigma^2) + 2n \frac{p + q + 1}{n - p - q - 2}$$

where  $L(\vec{\phi}, \vec{\theta}, \sigma^2)$  is defined in (4.7).

Model identification and model estimation can be merged as a procedure with computation software. In practice, we enumerated reasonable combinations of different parameters  $p, d, q$  and  $P, D, Q$  and calculated them. By the AICC value of each combination, we chose the correct model for our data. For instance, Table 4.2 lists part of parameter enumeration and their AICC values. From these, the final seasonal ARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>24</sub> (4.8) model has been chosen as our fitting model.

$$(1 - B)(1 - B^{24})X_t = (1 - 0.56B)(1 - 0.98B^{24})Z_t \quad Z_t \sim WN(0, 0.97) \quad (4.8)$$

Seasonal ARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>s</sub>, referred to the *airline model*, is probably the most commonly used seasonal ARIMA model.

Typically, the goodness of fit of a statistical model to a set of data is judged by comparing the observed values with the corresponding predicted values obtained from the fitted model. Figure 4.27 shows the fitted series and the observed GSM occupancy time series.

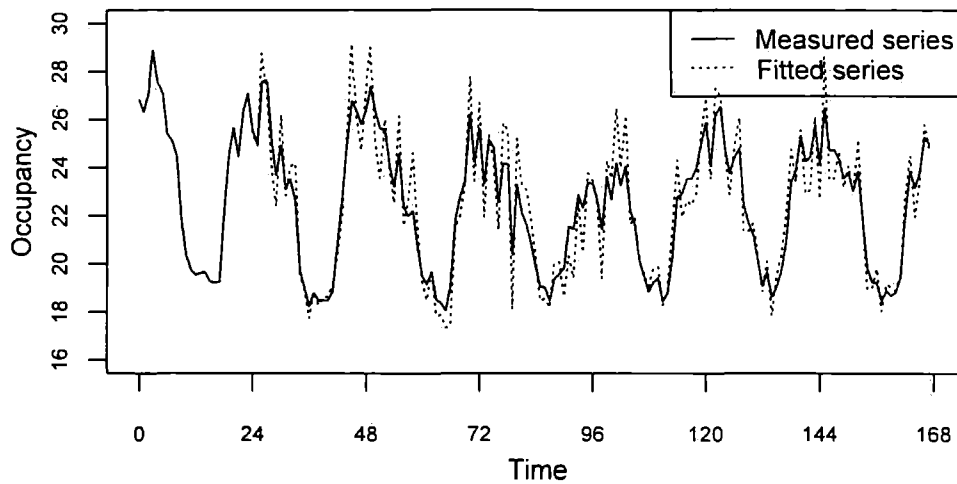


Figure 4.27: The fitted series and observed occupancy series

If we assume that the  $ARIMA(0, 1, 1) \times (0, 1, 1)_{24}$  is the true process generating the occupancy series  $\{X_t\}$ , then the *residuals* defined as [13]

$$\hat{W}_t = \frac{X_t - \hat{X}_t(\hat{\phi}, \hat{\theta})}{\sqrt{r_{t-1}(\hat{\phi}, \hat{\theta})}} \quad t = 1, \dots, n$$

where  $r_{t-1}(\hat{\phi}, \hat{\theta}) = E\{X_{t-1} - \hat{X}_{t-1}(\hat{\phi}, \hat{\theta})\}^2$ , should be a white noise series  $WN(0, \hat{\sigma}^2)$ . The *standardized residuals*  $\hat{R}_t$  are obtained by dividing  $\hat{W}_t$  by the estimated white noise standard deviation  $\hat{R}_t = \hat{W}_t / \hat{\sigma}$ . Figure 4.28 shows the *standardized residuals* of occupancy series after fitting with the ARIMA model.

The diagnostic check is based on the expected properties of *standardized residuals* under the assumption which has independent zero mean and variance one distribution. Figure 4.28 gives no indication of a nonzero mean or nonconstant variance, so on this basis there is no reason to doubt the compatibility of *standardized residuals* with unit variance white noise. The *sample autocorrelation* function also gives the fitness of the data model. Since for large sample  $n$  the *sample autocorrelations* of white noise are approximately with distribution  $N(0, 1/n)$ . We can therefore test whether or not the observed residuals are consistent with independent white noise by examining the *sample autocorrelations* of the residuals. As indicated the estimated residuals will not have precisely normal distribution.

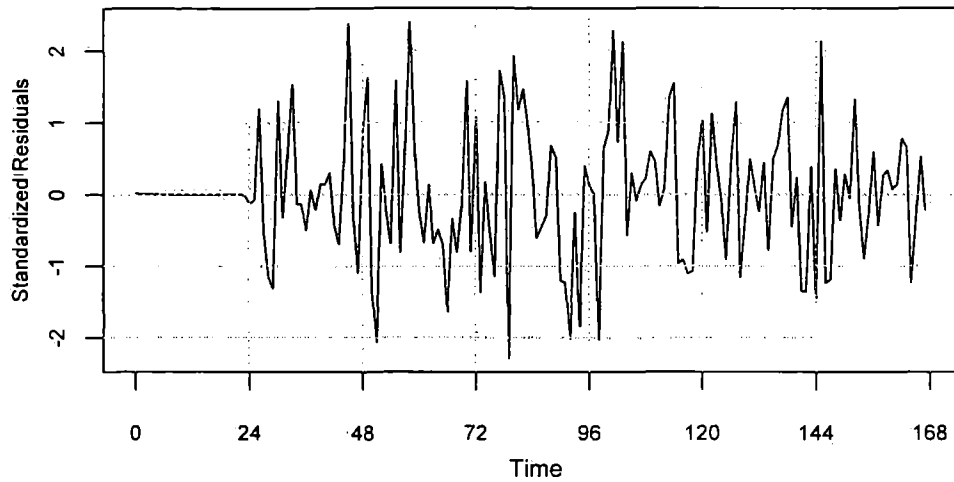


Figure 4.28: The standardized residuals after fitting ARIMA model

Figure 4.29 shows the *sample autocorrelation* function with bounds  $\pm 1.96/\sqrt{n}$ , where  $n = 168$ . Three of these fall outside the bounds. The modeling processing is summarized in Figure 4.30.

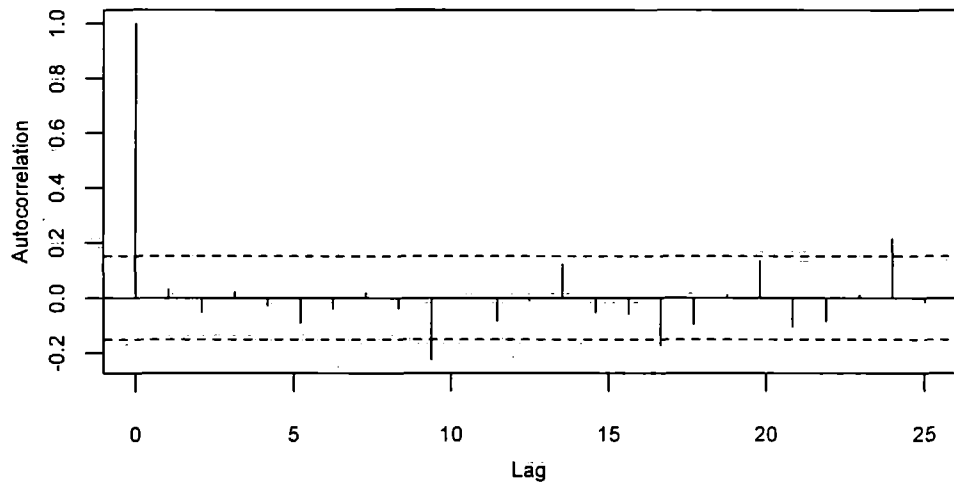


Figure 4.29: The autocorrelation function of the residuals

In this section, we introduced how the spectrum occupancy data can be analyzed by time series analysis approaches. The reason why we presented the GSM

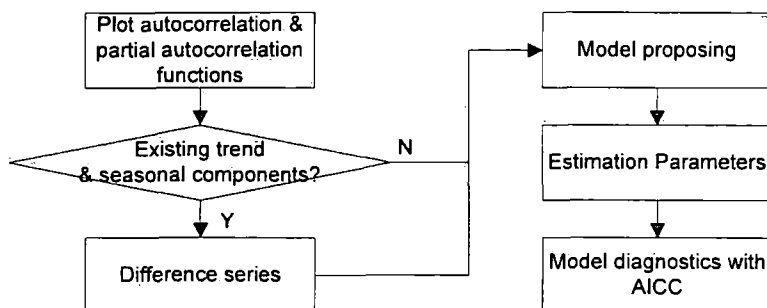


Figure 4.30: The flowchart of the modeling procedure

down-link band occupancy data as an example to illustrate is that it consists of typical season and trend components. Most occupancy data in our measurement bands, we noticed, are stationary which do not vary with time as TV band occupancy data shown in Figure 4.9. With this simpler scenario, the season ARIMA model can be degraded to simple ARMA model, or even to AR model. For example, the TV band occupancy data in Figure 4.9. can be modeled as AR(1) model as

$$(1 - 0.63B)X_t = Z_t \quad Z_t \sim WN(0, 0.1)$$

## 4.3 Generalized Linear Models

### 4.3.1 linear models and generalization

Time series analysis is a good tool for modeling variation in time domain. However, it fails to incorporate other important parameters, such as power threshold, into the models. The Laycock-Gott occupancy model developed by [43] regresses the occupancy data to some parameters with generalized linear model method. As time series analysis, regression analysis explains data structure with infinite variables with different viewpoints. For instance, in our project, the occupancy data can be regressed into *threshold* and *time*, which takes a general form  $y = f(s, t)$ , where  $y$  is occupancy rate,  $s$  is the threshold of the power amplifier and  $t$  is a time variable. Regression analysis has several possible objectives including [18]:

- A general description of data structure

- Prediction of future observations
- Assessment of the effect of, or relationship between, explanatory variables and the response

Linear regression analysis is used for explaining or modeling the relationship between a single variable  $y$ , called the *response* variable; and one or more *predictor* or *explanatory* variables,  $X_1, \dots, X_p$ . When  $p = 1$ , it is called simple regression and when  $p > 1$  it is called multiple regression. Generally, linear models take the form

$$y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon \quad (4.9)$$

where  $\beta_i$  are unknown parameters and  $\varepsilon$  is an observational error. In a linear model the parameters enter linearly and the *predictors* themselves do not have to be linear. For example,  $y = \beta_0 + \beta_1 \lg(X_1) + \beta_2 X_2^2 + \varepsilon$  is a linear model. Although, the term *linear* is often used as almost a synonym for simplicity. This gives the impression that linear models can only handle simple data structure. This is far from the truth — linear models can easily be expanded and modified to handle complex data structures. Nonlinear models are rarely absolutely necessary and most often arise from a theory about the relationships between the variables, rather than an empirical investigation [18].

If we take  $n$  measurements, We can write the model given in (4.9) as

$$y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{ip} X_{ip} + \varepsilon_i \quad i = 1, \dots, n$$

More convenient and compact with a matrix representation, the regression equation is written as

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon}$$

where  $\vec{y} = (y_1, \dots, y_n)^T$ ,  $\vec{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$ ,  $\vec{\beta} = (\beta_1, \dots, \beta_p)^T$  and

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix}$$

The regression model,  $\vec{y} = X\vec{\beta} + \vec{\varepsilon}$ , partitions the response into a systematic component  $X\vec{\beta}$  and a random component  $\varepsilon$ . We would like to choose  $\vec{\beta}$  so that

the systematic part explains as much of the *response* as possible. Geometrically speaking, the *response* lies in  $n$ -dimensional space, while  $\vec{\beta}$  is in a  $p$ -dimensional space. modeling data is to find  $\hat{\beta}$  so that  $X\hat{\beta}$  is as close to  $y$  as possible, which is represented geometrically in Figure 4.31 [18].  $\hat{\beta}$  is the best estimate of  $\beta$  within the model space and the *response* predicted by the model is  $\hat{y} = X\hat{\beta}$ . The difference between the actual *response* and the predicted *response* is denoted by  $\hat{\epsilon}$  which is called *residuals*. Thus if our model is successful, the structure in the data should be expressed in the  $p$  dimension, leaving just random variations in the residuals which lie in an  $n - p$  dimensional space.

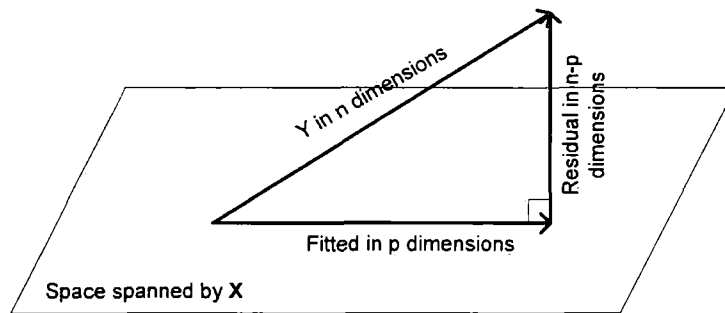


Figure 4.31: Geometrical representation of estimation  $\vec{\beta}$  [18]

Two of the most commonly used approaches to statistical estimation of parameters are the method of maximum likelihood and the method of least squares. Least squares corresponds to the maximum likelihood criterion if the experimental errors have a normal distribution. The maximum likelihood estimation will be discussed in the last chapter. The least squares estimate of  $\beta$ , called  $\hat{\beta}$  minimizes

$$\sum_{i=1}^n \epsilon_i^2 = \vec{\epsilon}^T \vec{\epsilon} = (y - X\vec{\beta})^T (y - X\vec{\beta})$$

Differentiating with respect to  $\beta$  and setting to zero, we find that  $\hat{\beta}$  satisfies the *normal equations*  $X^T X \hat{\beta} = X^T y$ . Provide for  $X^T X$  is invertible,  $\hat{\beta} = (X^T X)^{-1} X^T y$  and

$$\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y$$

Defining the *hat-matrix*  $H = X(X^T X)^{-1} X^T$ , which is the orthogonal projection of  $y$  onto the space spanned by  $X$ ,  $\hat{y} = Hy$ . The residuals are  $\bar{\epsilon} = y - \hat{y} = (I - H)y$ . The *residual sum of squares* (RSS) is  $\bar{\epsilon}^T \bar{\epsilon} = y^T (I - H)y$  [62]. Assuming  $\text{var } \epsilon = \sigma^2 I$ ,  $\hat{\beta}$  is unbiased and has variance  $(X^T X)^{-1} \sigma^2$ . We find that  $E \bar{\epsilon}^T \bar{\epsilon} = \sigma^2 (n - p)$ , which suggests the estimator

$$\hat{\sigma}^2 = \frac{\bar{\epsilon}^T \bar{\epsilon}}{(n - p)} = \frac{RSS}{(n - p)}$$

where  $n - p$  is the *degree of freedom* of the model.

For a generalized linear model we consider a parameter set  $\beta_1, \dots, \beta_p$  such that a linear combination of the  $\beta$ 's is equal to some function of the expected value  $\mu_i \triangleq EY_i$  of the *response variable*  $Y_i$

$$g(\mu_i) = \bar{x}_i \vec{\beta}$$

where  $g$  is a monotone, differentiable function called the *link function*;  $\bar{x}_i = (1, x_1, \dots, x_p)$ ; and the distribution of each  $Y_i$  is of *exponential family of distributions*  $f(y; \theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)]$  [36]. For example, the binomial distribution function

$$f(y; \theta) = C_n^y \theta^y (1 - \theta)^{n-y}$$

can be rewritten as

$$f(y; \theta) = \exp[y \log \theta - y \log(1 - \theta) + n \log(1 - \theta) + \log C_n^y]$$

where  $a(y) = y$ ,  $b(\theta) = \log \frac{\theta}{1-\theta}$ ,  $c(\theta) = n \log(1 - \theta)$  and  $d(y) = \log C_n^y$ .

### 4.3.2 logistic regression model

For binomial occupancy data, the number of channels  $Y_i$  at a given  $i^{\text{th}}$  time,  $i = 1, 2, \dots, n$ , whose signal amplitude is above a certain threshold, is a binomial distribution with parameters  $N$  and success probability  $p_i$ , where  $N$  is the total number of channels in a given band. Rather than directly modeling the dependence of  $EY_i$  on *explanatory* variables, it is more natural to explore how the occupancy rate  $p_i = EY_i/N$  can be described by observing *explanatory* variables [13].

A model often include terms corresponding to qualitative variables known as *factors*, which can take a limited set of values known as the *level* of the factor [13]. For example, while the *power amplitude* of the radio signal is a continuous variable range from minus infinity to infinity, the threshold of the occupancy level is a qualitative variable which takes a limited factor such as  $\leq -105$ ,  $\leq -100$ ,  $\leq -95$  and  $\leq -90$  dBm. For qualitative explanatory variables there is a parameter to represent each level of a factor. The corresponding elements of  $X$  are chosen to exclude or include the appropriate parameters for each observation. They are called *dummy* variables or *indicator* variables if only zeros and ones are used for  $X$ .

It is often necessary to include mixed terms in a model which corresponds to the situation where the coefficient of an *explanatory* variable varies according to the levels of some factor. Models that contain terms that are combinations of factors and variables are encountered when comparing regression relationships between groups of individuals. To include such a term, indicator variables  $T_j$  for the factor  $T$  are defined, and each of these is then multiplied by the value of  $X$ . The explanatory variables formed from the products,  $T_j X$ , have coefficients  $\beta_j$ . Taking the occupancy data as an illustration, we want to know the relationship between occupancy rate  $Y$  and time  $t$  under 4 different thresholds. At each threshold, we took 168 observations. This is achieved by defining 4 indicator variables,  $T_{105}$ ,  $T_{100}$ ,  $T_{95}$  and  $T_{90}$  as in Table 4.3.

One approach to model the binomial occupancy data is to adopt the linear model

$$p_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \quad (4.10)$$

and apply the method of least squares to obtain the parameters. For binomial occupancy data, there are a number of disadvantages to this approach [13]. One is that since the assumption of a normally distributed *response* variable cannot be made, the theory associated with fitting linear models to normal data is no longer valid. When the channel number  $N$  is reasonably large, this will not be a severe restriction, in view of the fact that the binomial distribution tends to normality for large sample size.

Another difficult is that there is the risk that the linear model given in (4.10) may give estimated values of occupancy which lie outside the range 0–1. The

### 4.3 Generalized Linear Models

Occupancy Y	T105	T100	T95	T90	T105-t	T100-t	T95-t	T90-t
$y_1$	1	0	0	0	$t_1$	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_{168}$	1	0	0	0	$t_{168}$	0	0	0
$y_{169}$	0	1	0	0	0	$t_1$	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_{336}$	0	0	0	0	0	$t_{168}$	0	0
$y_{337}$	0	0	1	0	0	0	$t_1$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_{672}$	0	0	0	1	0	0	0	$t_{168}$

Table 4.3: Data set with *dummy* variables

*logistic transformation* or *logit* of a success probability  $\theta$  is  $\log(\theta/(1 - \theta))$ , which is written as  $\text{logit}(\theta)$  [36]. Figure 4.32 shows the *logistic transformation*. The *logistic transformation* is essentially linear between  $\theta = 0.2$  and  $\theta = 0.8$ , but outside this range it becomes markedly nonlinear. As  $\theta \rightarrow 0$ ,  $\text{logit}(\theta) \rightarrow -\infty$ ; as  $\theta \rightarrow 1$ ,  $\text{logit}(\theta) \rightarrow \infty$ . Since a significant fraction of the experimental occupancy values close to 0, by *logistic transformation* this risk can be overcome.

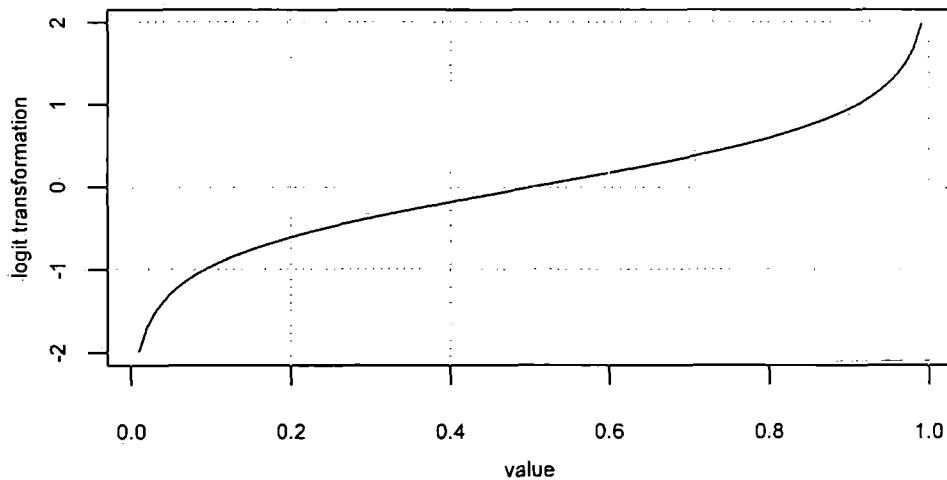


Figure 4.32: The *logistic transformation*

While the least squares method is usually used in fitting linear regression models, the method of maximum likelihood is the most widely used general method of estimation. The advantage of this method of estimation is its general applications. However, since exact analytic solutions of the resulting estimators are often difficult to obtain, iteration methods are widely used. With modern statistics softwares such as R, the parameter estimation is simplified greatly and researchers can focus on the model specifications.

Following the estimation of the  $\beta$  parameters in a logistic linear model, the next step is to make inference with testing hypotheses and obtaining confidence intervals, which need the knowledge of data distributions. In general, the problem of finding exact distributions is intractable and we rely instead on large sample asymptotic results. The rigorous development of these results requires careful attention to various regularity conditions. For independent observations from distributions which belong to the exponential family the necessary conditions are indeed satisfied. From the standard error of  $\vec{\beta}$ ,  $100(1 - \alpha)\%$  confidence limits for the corresponding true value,  $\hat{\beta} \pm z_{\alpha/2} se(\hat{\beta})$ , where  $z_{\alpha/2}$  is the upper  $\alpha/2$  point of the standard normal distribution and  $se(\hat{\beta})$  is the standard error of an estimate [17]. These interval estimates throw light on the likely range of values of the parameters.

One common question in inference is can one particular *predictor* be dropped from the model. That corresponds to the null hypothesis<sup>1</sup> which would be  $H_0 : \beta_i = 0$ . For example, if we assume that the dependence of occupancy on time *predictor*  $t$  in logistic regression model is the same in the different thresholds, there are only differences in the *intersection*. In other words, in the occupancy model only one *predictor*  $t$  is necessary and other cross products can be dropped. These statistics are often referred to as  $t$ -values, but in the analysis of binomial data they are generally taken to a normal distribution rather than a  $t$ -distribution and so are more appropriately referred to as  $z$ -values [13]. More specifically, under the null hypothesis that  $\beta = 0$ , the ratio  $\hat{\beta}/se(\hat{\beta})$  has a standard normal distribution. For example, Figure 4.34 shows the output of  $z$ -test for logistic regression model.

---

<sup>1</sup>In statistics, a null hypothesis  $H_0$  is a hypothesis set up to be nullified or refuted in order to support an alternative hypothesis.

After fitting a model to a set of data, it is natural to enquire about the extent to which the fitted values of the response variable under the model compare with the observed values. This aspect of the adequacy of a model is widely referred to as *goodness of fit*. There are a number of summary statistics that measure the discrepancy between observed binomial proportions and fitted proportions under an assumed model. Of these, the most widely used is based on the likelihood function for the assumed model. When the unknown parameters are set equal to their maximum likelihood estimations, the values of the likelihood can therefore be used to summarize the extent to which the sample data are fitted by the current model. This is the maximized likelihood under current model, denoted  $\hat{L}_c$ . The statistics can not be used on their own to assess the lack of fit of the current model since they are not independent of the number of observations [18]. It is therefore necessary to compare the current model with an alternative baseline model for the same data. The model is termed the *full* or *saturated* model denoted  $\hat{L}_f$  which is defined as follow:

- the *full* model is a generalized linear model using the same distribution as the current model
- the number of parameters in the *full* model is equal to the total number of observations.

The *full* model is not useful on its own since it does not provide a simpler summary of the data than the individual observations themselves. However, by comparing  $\hat{L}_f$  with  $\hat{L}_c$ , the extent to which the current model adequately represents the data can be judged.

To compare  $\hat{L}_f$  with  $\hat{L}_c$ , we define

$$D = -2 \log(\hat{L}_c / \hat{L}_f) = -2(\log \hat{L}_c - \log \hat{L}_f)$$

Large values of  $D$  indicate that the current model is a poor one; small values indicate that  $\hat{L}_c$  is similar with  $\hat{L}_f$  and therefore the current model is a good one. The statistics of  $D$  measures the extent to which the current model deviates from the *full* model and is termed the *deviance*.

In order to assess the extent to which an adopted model fits a set of binomial data, the distribution of the *deviance* is needed under the assumption that the

model is correct. Since the *deviance* is the likelihood ratio statistics for comparing a current model with the *full* model, the null distribution of the *deviance* follows directly from a result associated with likelihood ratio testing. It can be stated as asymptotically  $\chi^2$  with  $n - p$  distribution, where  $n$  is the number of binomial observations and  $p$  is the number of unknown parameters included in the current logistic model [17].

The main use of *deviance* is in comparing alternative linear logistic models for binomial data, although it can be useful as a measure of model adequacy in some circumstance. When one model contains terms that are additional to those in another, the two models are said to be nested. The difference in *deviance* between two nested models measures the extent to which the additional terms improve the fit of the model to the *response* variable.

To compare nested models for binomial data, no exact distribution theory is available. However, since the *deviance* for each model has an approximately  $\chi^2$  distribution, the difference between two *deviances* will also be approximately distributed<sup>1</sup> as  $\chi^2$ . ANCOVA (analysis of covariance) table is often used to test nested model. This term is used for mixed models in which some of the *explanatory* variables are dummy variables representing factor levels and others are continuous measurements.

Once a model has been fitted to the observed values of a binomial response variable, it is essential to check that the fitted model is actually valid. A thorough examination of the extent to which the fitted model provides an appropriate description of the observed data is a vital aspect of the modeling process. Measures of agreement between an observation on a response variable and the corresponding fitted value are known as the *residuals*. These quantities and summary statistics derived from them can provide much information about the adequacy of a fitted model. *Deviance residual*

$$d_i = \text{sgn}(y_i - \hat{y}_i) \sqrt{2y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + a(n_i - y_i) \log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)}$$

---

<sup>1</sup>Probability theory states that if  $D_1$  and  $D_2$  have  $\chi^2$  distributions with  $n_1$  and  $n_2$  degree of freedom respectively, where  $n_1 > n_2$ , and they are independent then their difference also has a  $\chi^2$  distribution with  $n_1 - n_2$  degree of freedom.

is a widely used residual definition among several different definitions [13]. For large sample, the *deviance residual* have approximately  $N(0, 1)$  distribution. The diagnostics that are derived from the Normal theory, such as Q-Q plots, need to be treated with some care as for small counts the approximate distributions can deviate from normality even when the model is perfectly adequate.

### 4.3.3 model analysis for GSM band occupancy data

Now, we want to regress the GSM data shown in Figure 4.20 to explanatory variables *time* and *power amplitude*. Physically, the *power amplitude* variable is a continuous variable. In our project, for computation issue we quantize it to 4 qualitative variables  $s_i$   $i = 1, 2, 3, 4$  corresponding to 4 different thresholds -105, -100, -95 and -90 dBm respectively. The data series are shown in Figure 4.33 which share similar patterns.

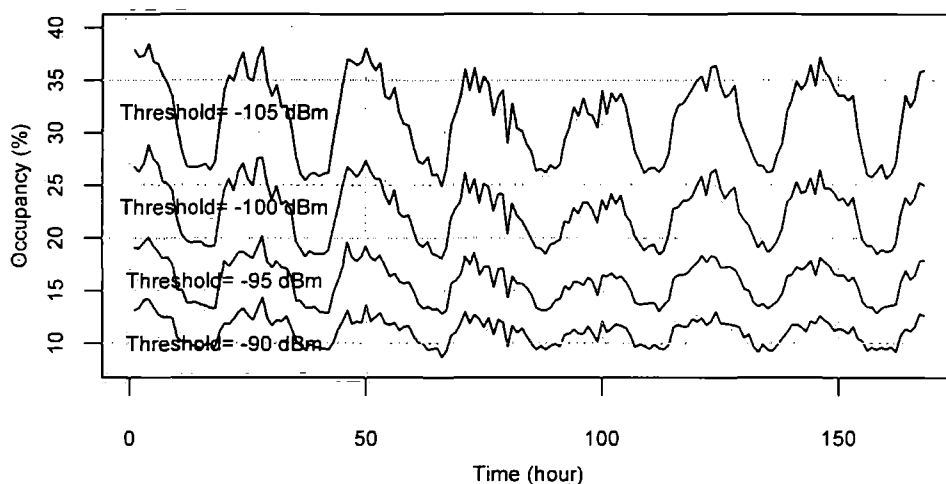


Figure 4.33: The occupancy data by different thresholds

Statistical modeling involves three steps: (1) specifying models; (2) estimating parameters; (3) making inferences – that is, testing hypotheses, obtaining confidence intervals and assessing the goodness of fit of models.

The first step is to postulate a model. As the previous analysis, the data in Figure 4.23 have trend and seasonal components. Linear relationship  $\logit y_i =$

$\beta_{i0} + \beta_{i1}t$   $i = 1, 2, 3, 4$  is not enough to correlate the *response* occupancy rate and *predictor* measurement time. We add a *predictor*  $t^2$  to fit the trend component and two *predictors*  $\sin(2\pi t/24)$  and  $\cos(2\pi t/24)$  to fit the periodic component. So, the first generalized linear model is expressed as

$$\text{logit } y_i = \beta_{i0} + \beta_{i1}t + \beta_{i2}t^2 + \beta_{i3} \sin\left(\frac{2\pi}{24}t\right) + \beta_{i4} \cos\left(\frac{2\pi}{24}t\right) \quad i = 1, 2, 3, 4 \quad (4.11)$$

In this model, each threshold has complete independent parameter set of *predictors*. The report generated by  $\hat{R}$  is shown in Figure 4.34.

The report starts off with a repeat of the model specification where it has an independent parameter set for each threshold. The *deviance residuals* summarize the distribution of residuals. The *coefficient* table is of primary interest. We get estimates of the regression coefficients, standard errors of the same, and tests for whether each regression coefficient can be assumed to be zero. *Residual deviance* corresponds to the *residual sum of squares* in ordinary linear regression analysis which is used to estimate the standard deviation about the regression line. In a binomial model, however, the standard deviation of the observations is known, and we can therefore use the deviance in a test for model specification. The Akaike information criterion (AIC) is a measure of goodness of fit which takes the number of fitted parameters into account as mentioned in the previous section. The null deviance is the deviance of a model that contains only the intercept. The difference between *Null deviance* and *Residual deviance*, here  $94265 - 1659 = 92606$ , which can be used for a joint test of whether any effects are present in the model.

ANCOVA table is usually specified in terms of parameters which are readily interpretable as effects due to factor levels and interactions [35]. The ANCOVA table of logit model is shown in Figure 4.35. Notice that the *Deviance* column gives differences between models as variables are added to the model in turn. The deviance are approximately  $\chi^2$ -distributed with the stated degree of freedom. Since  $p$  values of *thres* :  $t$  and *thresh* :  $t^2$  variables are not significant, they may be removed from the model. Actually, there is no more information in the deviance table 4.35 than in the *coefficient* table 4.34. In table 4.34, we notice that the  $p$  values of *thresX* :  $t$  and *thresX* :  $t^2$  are not significant either, which suggests those variables can be removed from the model.

### 4.3 Generalized Linear Models

```
glm(formula = occ ~ thres * (t + t2 + sin.t + cos.t), family = binomial)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-6.694843	-0.985693	0.008685	0.959203	4.761398

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-1.157e+00	9.338e-03	-123.863	< 2e-16	***
thresT105	4.479e-01	1.260e-02	35.539	< 2e-16	***
thresT90	-8.452e-01	1.545e-02	-54.700	< 2e-16	***
thresT95	-4.276e-01	1.414e-02	-30.251	< 2e-16	***
t	-2.247e-03	2.564e-04	-8.766	< 2e-16	***
t2	1.143e-05	1.475e-06	7.752	9.06e-15	***
sin.t	1.226e-01	4.449e-03	27.549	< 2e-16	***
cos.t	1.506e-01	4.424e-03	34.043	< 2e-16	***
thresT105:t	1.797e-04	3.456e-04	0.520	0.60314	
thresT90:t	3.872e-04	4.252e-04	0.911	0.36238	
thresT95:t	2.130e-05	3.886e-04	0.055	0.95629	
thresT105:t2	-6.112e-07	1.986e-06	-0.308	0.75834	
thresT90:t2	-3.080e-06	2.449e-06	-1.258	0.20853	
thresT95:t2	-6.096e-07	2.237e-06	-0.273	0.78523	
thresT105:sin.t	1.830e-02	5.986e-03	3.057	0.00223	**
thresT90:sin.t	-2.388e-02	7.394e-03	-3.229	0.00124	**
thresT95:sin.t	-1.353e-02	6.755e-03	-2.003	0.04522	*
thresT105:cos.t	2.465e-02	5.953e-03	4.142	3.45e-05	***
thresT90:cos.t	-4.211e-02	7.349e-03	-5.730	1.00e-08	***
thresT95:cos.t	-1.777e-02	6.716e-03	-2.646	0.00814	**

===

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 94265 on 671 degrees of freedom  
Residual deviance: 1659 on 652 degrees of freedom  
AIC: 7137.5

Number of Fisher Scoring iterations: 3

Figure 4.34: The report for logit regression model

### 4.3 Generalized Linear Models

```

Model: binomial, link: logit

Response: occ

Terms added sequentially (first to last)

          Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                671      94265
thres              3    85268      668      8997      0
t                  1      219      667      8777 1.321e-49
t2                 1      225      666      8553 9.350e-51
sin.t              1    2705      665      5848      0
cos.t              1    4038      664      1809      0
thres:t            3         3      661      1807 4.278e-01
thres:t2           3         2      658      1805      1
thres:sin.t        3         4      655      1760 1.400e-09
thres:cos.t        3        10      652      1659 7.663e-22

```

Figure 4.35: The ANCOVA report for logit regression model

So, the model of (4.11) is modified as

$$\text{logit } y_i = \beta_{i0} + \beta_{1t} + \beta_{2t^2} + \beta_{i3} \sin\left(\frac{2\pi}{24}t\right) + \beta_{i4} \cos\left(\frac{2\pi}{24}t\right) \quad i = 1, 2, 3, 4 \quad (4.12)$$

where each threshold has common parameters for *predictor*  $t$  and *predictor*  $t^2$ . A brief report of the new model is shown in Figure 4.36. The *degree of freedom* increase from 652 to 658. Although the *Residual Deviance* increases from 1659 to 1666, it can be easily explained by the randomness of data with increased *degree of freedom*. From the AIC criterion, the new model is better than the former. Figure 4.36 shows the fitted series and the observed GSM occupancy time series.

As usual, *diagnostics* may be based on formal statistical tests, but more frequently involve a less formal evaluation of a graphical representation of certain statistics. The residuals obtained after fitting a linear logistic model to an observed set of data form the basis of a large number of diagnostic techniques for assessing model adequacy. The residuals can be assessed for fitness using a  $Q-Q$  plot [35]. This compares the residuals to “ideal” normal observations. We plot the sorted residuals against  $\Phi^{-1}\left(\frac{i}{n+1}\right)$  for  $i = 1, \dots, n$  where  $\Phi(\cdot)$  is the normal distribution function and  $n$  is the sample number. The straight line in Figure

### 4.3 Generalized Linear Models

```
Call: glm(formula = occ ~ t + t2 + thres * (sin.t + cos.t), family = binomia
```

Coefficients:

(Intercept)	t	t2	thresT105
-1.159e+00	-2.120e-03	1.061e-05	4.573e-01
thresT90	thresT95	sin.t	cos.t
-8.417e-01	-4.316e-01	1.225e-01	1.507e-01
thresT105:sin.t	thresT90:sin.t	thresT95:sin.t	thresT105:cos.t
1.770e-02	-2.286e-02	-1.288e-02	2.470e-02
thresT90:cos.t	thresT95:cos.t		
-4.242e-02	-1.788e-02		

Degrees of Freedom: 671 Total (i.e. Null); 658 Residual

Null Deviance: 94260

Residual Deviance: 1666 AIC: 7132

Figure 4.36: The report for logit regression model

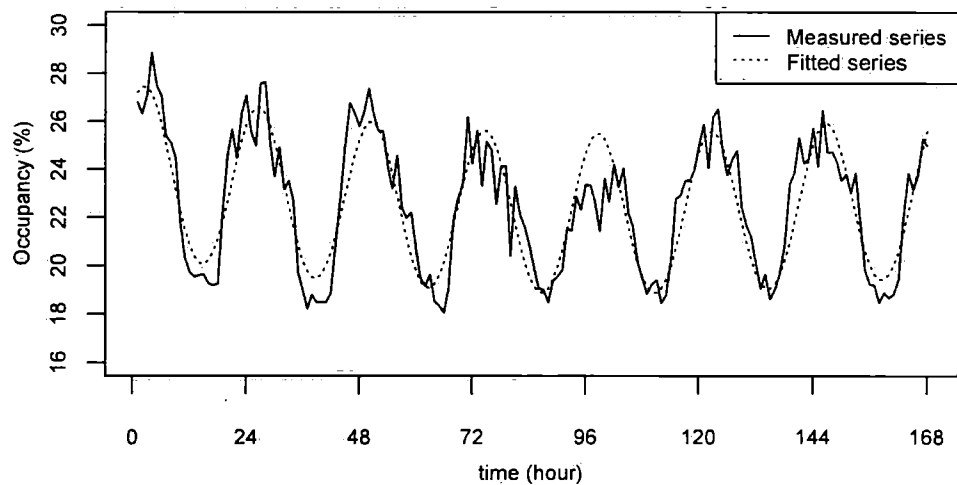


Figure 4.37: Fitted series and observed occupancy series

4.38 is a line joining the first and third quartiles<sup>1</sup>. Normal residuals should follow this line approximately. Here, the residuals look normal.

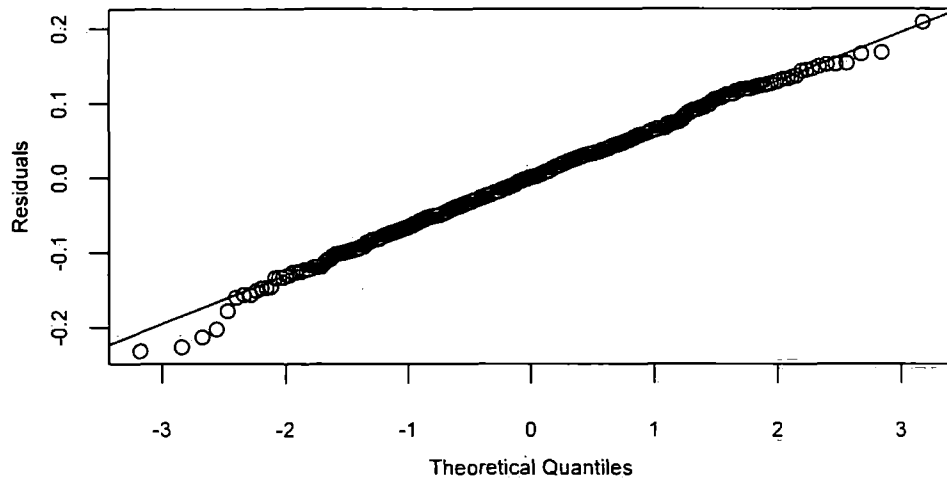


Figure 4.38: Model check for the logit model

Although the  $Q - Q$  plot was originally designed to examine whether the residuals can be taken to be normal distribution, in logit modeling it is more useful for diagnosing inadequacy and revealing the presence of outliers. There are no significant outliers in Figure 4.38.

## 4.4 Summary

With graphics and tables we presented the statistical information of the spectrum occupancy range from 100–2500 MHz in frequency, time and space domains. The low spectrum occupancy statistics shows that there is significant spectrum for cognitive radio to provide service, if the current telecommunication policies could be modified. Model approaches can be used to describe and predict the spectrum occupancy in a large scale. If the cognitive radio base station monitors

<sup>1</sup>The quantile function is the inverse of the cumulative distribution function. The  $p$ -quantile is the value with the property that there is probability  $p$  of getting a value less than or equal to it. The quartile is 25-quantile.

the occupancy of a given spectrum, the predicted occupancy messages can be broadcasted to cognitive radio devices, which can save a great deal of scanning time for each device. In this chapter, we introduced the methodologies for time series analysis and generalized linear model and modeled the GSM occupancy data with these two approaches, which fit the measured data well.

# Chapter 5

## Conclusions

The primary goal of this project is to survey the spectrum occupancy in frequency, time and space (direction) fields for the the feasibility of cognitive radio technology. While presenting the spectrum occupancy information with statistics, we have focused on modeling approaches.

First, we briefly summarize the main chapters in this thesis. In Chapter 2, from the point of view of engineering we described the spectrum monitoring systems employed in our project and corresponding techniques to detect radio signals and to diminish distortions. The details on control and data acquisition programs are also presented in this chapter to show the technique for fast transfer of data in a long duty cycle.

In Chapter 3, we focused on the verification of the high resolution DOA algorithms – EM and SAGE. After we achieved relatively accurate estimates with a few degree errors in Matlab simulations, the verifying experiments were conducted in the restrained environment – anechoic chamber. The results of the experiments showed the validation of the algorithms even with simple linear phased array, though there are still convergence issues.

In Chapter 4, we profiled the spectrum occupancy data range from 100 – 2500 MHz with figures and tables. Considering the limitations of static statistics, we introduced model approaches for description and prediction. Time series analysis was initially introduced in spectrum occupancy analysis while generalized linear models were used as complementarity solutions to model occupancy data into other parameters. Both of them achieved reasonable results.

---

Since the rapid development of wireless technologies, traditional spectrum monitoring equipment need to be upgraded to catch advanced and sophisticated radio signals. In the spectrum occupancy project the channel sounder, which can scan 300 MHz spectrum with 4 ms and have 8 input channels for spatial occupancy monitoring, was deployed to detect the fast moving wireless signals. For extensive studies, more dedicated equipment and advanced scanning schemes need to be developed to determine signal parameters.

There are generally positive findings in this project with respect to the prospects for cognitive radio. Statistics show that the spectral occupancies are sparse indeed. Occupancy rates of most bands in VHF and UHF are less than 10% overall and the distribution of the white band indicates that the bands are capable of running wideband wireless communications. Of course, the current static spectrum allocation policies and spectrum management strategies have to be modified to motivate applications of the cognitive radio technology. While this study is to identify the low utilizations of bands, long term studies are crucial in developing spectrum sharing technologies and for spectrum management.

However, detailed measurements and analysis intended to quantify the performance of a particular band and a particular measuring period usually cannot be extended directly to others. Model approaches try to describe and summarize occupancy data with the different perspectives. Time series models were initially introduced in this project to describe and predict spectrum occupancy variations with time. With this methodology, the cognitive radio base stations or servers monitor the spectrum and the predicted occupancy messages can be broadcast to cognitive radio devices, which can save a great deal of scanning time for each device. With the GSM occupancy data as an example, we introduced another modeling method – Generalized linear models, by which the occupancy data can be modeled with other parameters such as amplitudes and locations, etc. So, spectrum occupancies can be inferred with similar power thresholds and similar radio environments.

In order to improve the spectral efficiency sectorization technique is widely adopted by many wireless communications. Spectrum occupancy in spatial (directional) field becomes an important parameter for future cognitive radio technology. There are a couple of high resolution direction of arrival algorithms such

---

as MUSIC, ESPRIT and EM. This thesis focused on verifications of the EM algorithm and the SAGE algorithm in simulations and in anechoic chamber. The results basically showed that this technique can be used in the real world.

Above all, radio spectrum occupancies in time, frequency and space domains observed in this project are sparse indeed. This suggests that cognitive radio technologies have great prospects in the future wireless communication infrastructure, if current telecommunication policies and regulations are modified.

## References

- [1] AGILENT. *Agilent Spectrum Analysis Basics*. Agilent Technologies, California, 2005. vi, 13, 14, 15, 16, 17, 21, 23, 26, 29
- [2] R. AMBROSINI. RFI Measurement Protocol for Candidate SKA Sites. Recommendation for radio frequency interference measurements, 2003. 16, 17
- [3] JAN AXELSON. *Parallel Port Complete: Programming, Interfacing and Using the PC'S Parallel Printer Port*. Lakeview Research, Madison, 1997. 33, 34
- [4] JAN AXELSON. *Serial Port Complete: COM Ports, USB Virtual COM Ports, and Ports for Embedded Systems*. Lakeview Research, 2007. 32
- [5] GEORGE BOX, GWILYM JENKINS, AND GREGORY REINSEL. *Time Series Analysis: Forecasting and Control*. Prentice Hall, 1994. 101, 104, 106, 109, 111
- [6] PETER BROCKWELL AND RICHARD DAVIS. *Time Series: Theory and Methods*. Springer, 1998. 100, 102, 103, 105, 112
- [7] PETER BROCKWELL AND RICHARD DAVIS. *Introduction to Time Series and Forecasting*. Springer, 2003. 98, 107
- [8] PETER J. BURMAN. Issues Involved with the Seasonal Adjustment of Economic Time Series: Comment. *Journal of Business & Economic Statistics*, 2(4):325–327, 1984. 107

## REFERENCES

---

- [9] JOSEPH CARR. *Practical Radio Frequency Test and Measurement*. Newnes, Oxford, 1999. 13, 21, 27
- [10] DAVID CHANDLE. *Introduction to Modern Statistical Mechanics*. Oxford, 1987. 61
- [11] PEI-JUNG CHUNG AND JOHANN F. BOHME. DOA Estimation Using Fast EM and SAGE Algorithms. *Elsevier on Signal Processing*, **82**:1753–1762, 2001. 67
- [12] PEI-JUNG CHUNG AND JOHANN F. BOHME. Recursive EM and SAGE-Inspired Algorithms with Application to DOA Estimation. *IEEE Trans. on Signal Processing*, **53**(8):2664–2676, 2005. 64, 66
- [13] DAVID COLLETT. *Modelling Binary Data*. Chapman and Hall, 2002. 97, 113, 118, 119, 121, 124
- [14] MEASUREMENT COMPUTING CORPORATION. PCI-DAS4020/12 Analog and Digital I/O Board User Manual. Data Sheet, 2006. vi, 36, 37, 40
- [15] NATIONAL INSTRUMENTS CORPORATION. NI-488.2 User Manual. User Manual, 2002. 34
- [16] A. DEMPSTER, N. LAIRD, AND D. RUBIN. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**(1):1–38, 1977. 59, 60
- [17] ANNETTE DOBSON. *An Introduction to Generalized Linear Models*. Chapman & Hall, 2001. 121, 123
- [18] JULIAN J. FARAWAY. *Linear Models with R*. Chapman & Hall, 2004. viii, 115, 116, 117, 122
- [19] MEIR FEDER AND EHUD WEINSTEIN. Parameter Estimation of Superimposed Signals Using the EM Algorithm. *IEEE Trans. on Acoustics, Speech and Signal Processing*, **34**(4):477–489, 1988. 56

## REFERENCES

---

- [20] JEFFREY A. FESSLER AND ALFRED O. HERO. Space-Alternating Generalized Expectation-Maximization Algorithm. *IEEE Trans. on Signal Processing*, **42**(10):2664–2677, 1994. 65
- [21] P. FILIPPIDIS. *Multi-channel sounder for directional measurement*. PhD thesis, University of Manchester Institute of Science and Technology, 2002. 18, 20, 39, 43
- [22] A.J. GIBSON AND L. ARNETT. Statistical modelling of spectrum occupancy. *IEE Electronics Letters*, **29**(25):2175–2176, 1993. 95
- [23] MONSON H. HAYS. *Statistical Digital Signal Processing and Modeling*. Wiley, 1996. 53, 57
- [24] J. MITOLA III AND JR. G.Q. MAGUIRE. Cognitive radio: making software radios more personal. *IEEE Wireless Communications*, **6**, No.2:13–18, 1999. 4
- [25] TDK RF SOLUTIONS INC. Compact Fully Anechoic Chamber CAC-S<sup>TM</sup>. Tutorial, 2004. 69, 72
- [26] ITU. *ICT regulation toolkit*. ITU. 2, 5
- [27] DON JOHNSON. *Array Signal Processing: Concepts and Techniques*. Prentice Hall, New Jersey, 1993. 46, 47, 50, 51, 54, 58
- [28] J. B. JOHNSON. Thermal Agitation of Electricity in Conductors. *Physics Review*, **32**:97–109, 1928. 20
- [29] STEVEN M. KAY. *Fundamentals of Statistical Signal Processing*. Prentice Hall PTR, 1998. 53
- [30] JOHN D. KRAUS. *Antennas For All Applications*. McGraw-Hill Science, 2001. 49
- [31] HAMID KRIM AND MATS VIBERG. Two decades of array signal processing research: the parametric approach. *Signal Processing Magazine, IEEE*, **13**(4):67–94, 1996. 54, 55

## REFERENCES

---

- [32] JUKKA LEMPIAINEN. *Radio Interface System Planning for GSM/GPRS/UMTS*. Springer, 2001. 23
- [33] R. LEWENZ. *Temporal variation in the radio environment*. PhD thesis, Durham University, 2005. 18, 20, 39, 43
- [34] KENBOTONG COMMUNICATION LTD. Log periodic antenna TDJ-0825DSA data sheet. Data sheet. 11
- [35] JOHN MAINDONALD AND JOHN BRAUN. *Data Analysis and Graphics Using R: An Example-based Approach*. Cambridge University Press, 2006. 78, 101, 111, 125, 127
- [36] P. MCCULLAGH AND J. NELDER. *Generalized Linear Models*. Chapman & Hall, 1989. 118, 120
- [37] G. MCLACHLAN AND T. KRISHNAN. *The EM algorithm and extensions*. John Wiley & Sons, 1997. 59
- [38] HAMISH MEIKLE. *Modern Radar Systems*. Artech House Publishers, 2001. 51
- [39] MICHAEL I. MILLER AND DANIEL R. FUHRMANN. Maximum-Likelihood Narrow-Band Direction Finding and the EM Algorithm. *IEEE Trans. on Acoustics, Speech and Signal Processing*, **38**(9):1560–1577, 1990. 63, 64, 67
- [40] T. MINKA. Expectation-Maximization as lower bound maximization. Tutorial, 1998. vii, 59, 60, 61, 62
- [41] ALEXANDER MOOD AND FRANKLIN GRAYBILL. *Introduction to the theory of statistics*. McGraw-Hill, 1974. 97, 107
- [42] ROBERT NAU. *Economic and Environmental Risk and Uncertainty: New Models and Methods*. Springer, 1997. 110
- [43] C.A. PANTJARIOS, J.A. WYLIE, J. BROWN, G.F. GOTT, AND P.J. LAYCOCK. Extended UK models for high frequency spectral occupancy. *IEE Proc. Communications*, **145**(3):168–174, 1998. 95, 115

## REFERENCES

---

- [44] CHARLES PAPAS. *Theory of Electromagnetic Wave Propagation*. Dover Publications, 1989. 1
- [45] J.D. PARSONS. *The Mobile Radio Propagation Channel*. Wiley, 2000. 20
- [46] DAVIS M. POZER. *Microwave Engineering*. Wiley, 2004. 24, 46
- [47] IAN B. RHODES. A Tutorial Introduction to Estimation and Filtering. *IEEE Trans. on Automatic Control*, **16**(6):688–706, 1971. 57, 58
- [48] S. SALOUS. Digital techniques for mobile radio chirp sounders. *IEEE Proceedings on Communications*, **145**(3):191–196, 1998. 10
- [49] F. SANDERS. Broadband spectrum survey at San Diego, California. Report, Department of Commerce, U.S., 1996. 6
- [50] A.D. SPAULDING AND G.H. HAGN. On the definition and estimation of spectrum occupancy. *IEEE Transactions on Electromagnetic Compatibility*, **EMC-19**:269–280, 1977. 95, 96
- [51] G. STAPLE AND K. WERBACH. The end of spectrum scarcity. *Spectrum, IEEE*, **41**(3):48–52, 2004. 4
- [52] BJARNE STROUSTRUP. *The C++ Programming Language: Special Edition*. Addison-Wesley Professional, Boston, 2000. 31
- [53] RF SOLUTIONS TDK. Data Sheet for Compact Fully Anechoic Chamber. Data sheet, 2004. 69
- [54] AGILENT TECHNOLOGIES. Agilent Technologies Infiniium 54810/15/20/25/35/45A Oscilloscopes. Data Sheet, 2000. 70, 71
- [55] BROADBAND POWER TECHNOLOGY. Dressler ARA-1500 data sheet. Data sheet. 7, 11
- [56] HARRY L. VAN TREES. *Optimum Array Processing*. Wiley-Interscience, 2002. 48

## REFERENCES

---

- [57] ZHE WANG AND SANA SALOUS. Spectrum occupancy analysis for cognitive radio. London Communications Symposium, University College London, 2006. 3
- [58] ZHE WANG AND SANA SALOUS. Spectrum occupancy analysis for VHF and UHF band. 19th UK URSI Colloquium, the Rutherford Appleton Laboratory, 2006. 18
- [59] ZHE WANG AND SANA SALOUS. ARIMA model of spectrum occupancy. The IET Seminar on Cognitive Radio and Software Defined Radio, The IET, 2008. 100
- [60] ZHE WANG AND SANA SALOUS. Time series modelling of spectrum occupancy for cognitive radio. URSI GENERAL ASSEMBLY, Chicago, U.S.A, 2008. 100
- [61] ZHE WANG AND SANA SALOUS. Spectrum occupancy statistics and time series models. *Springer Journal of signal processing systems*, 2009. 100
- [62] SANFORD WEISBERG. *Applied Linear regression*. Wiley-Interscience, 2005. 118

