UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering,
Mathematics & Computer Science

Automatic detection of outlandish signal behaviour in the spectrum of cellular networks

Wouter Bakker

M.Sc. Thesis
May 2019

Supervisors:
Prof. dr. ir. ing. F.B.J. Leferink
Dr. A. Alayón Glazunov
Ir. T.W.H. Fockens
Dr. ir. A.B.J. Kokkeler

Telecommunication Engineering Group
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Summary

Mobile telecommunications are becoming increasingly important for our society. It is widely used for applications like WhatsApp messaging and social media like Facebook or Instagram. Besides, mobile telecommunications enable vital societal functions and critical infrastructure. Examples of this are calling the emergency services at any time and controlling important infrastructures. A large part of the data in these services is transmitted over cellular networks run by network operators. The Dutch Radiocommunications Agency has a monitoring network to measure all the spectrum usage by these network operators.

In this thesis the data from this monitoring network is analysed, and a methodology is presented to automatically detect outlandish signal behaviour in the frequency bands reserved for cellular networks. Besides the detection of outlandish signals, e.g., like interference, the proposed algorithm is also able to detect other events in the spectrum of cellular networks. For example, events that can be detected are a power outage of a base station, power saving at night and changes in aggregated transmit power, e.g., due to deployment of new base stations. The outlandish signal variation detection is basically the detection of an ”unexpected” decrease or increase of the aggregated received power in the network at the measurement points.

To perform a structured analysis of the measurement data, a model of the measured signal is proposed. The model describes the relation between the measured signal strength, the wanted signal, the outlandish signals and the noise in the measurement. Furthermore, four cases considering various combinations of the wanted signal, the outlandish signals and the noise in the measurement.

The data set was gathered from 15 fixed measurement locations of the Radiocommunications Agency of the Netherlands and contains spectrum data between 20 MHz and 3 GHz. The proposed algorithm has four main steps that lead to the automated detection of outlandish signals in the frequency bands. First, for all 15 locations, noise floors need to be estimated to detect the different channels that are in use by different network operators. This is done by using the median forward consecutive mean excision algorithm (MED-FCME). Second, based on the noise floor threshold, channels in which network operators are active are detected. Then, statistics like the mean, median, minimum maximum and the CDF of all channels are computed. By also computing the variation of the multiple statistics, outlandish signals are detected. Finally, some of these outlandish signals are classified as events when the algorithm is able to connect certain characteristics to the outlandish signal.

The whole data collection and detection process is automated with MATLAB. The process is able to combine and automate the data collection, data processing, noise floor detection, event detection and the outlandish signals level detection. It is concluded that the detection process works and that it was possible to detect outlandish signals and events in the downlink channels. The measured values in the uplink bands are too close to the noise floor and were therefore not possible to do a useful analysis on. It was also concluded that the detection of channels only works if there were enough measured values that contain only noise. If the number of measured values that contain noise are relatively small compared to the total amount of measured values, the proposed algorithm is not able to correctly estimate the noise floor.

Recommendations are given to increase the success of the proposed algorithm for a next stage. For example, it is possible to detect narrow band signals, if the used measurement bandwidth is decreased. Also, placing measurement antennas inside crowded places like large train stations will increase received signal strength from mobile devices. This enables the possibility to analyse the uplink signals as well. The third recommendation tries to improve the channel detection algorithm by taking into account the local minimum, local maximum and the difference between these two. Finally, the algorithm can also be further improved by doing a deep-dive into improving the way events are currently classified.

Contents

List of acronyms

3GPP 3rd Generation Partnership Project

GSM Global System for Mobile Communications

GPRS General Packet Radio Service

EDGE Enhanced Data Rates for GSM Evolution

UMTS Universal Mobile Telecommunications System

HSDPA High-speed Downlink Packet Access

HSUPA High-Speed Uplink Packet Access

LTE Long Term Evolution

2G Second-generation Cellular Technology

3G Third-generation Cellular Technology

4G Fourth-generation Cellular Technology

FP false positive

TP true positive

FN false negative

TN true negative

FCME forward consecutive mean excision

MED-FCME median forward consecutive mean excision

SD standard deviation

CDF cumulative distribution function

AWGN additive white Gaussian noise

Chapter 1

Introduction

1.1 Motivation

Mobile telecommunications are becoming increasingly important for our society. It is widely used by the general public for applications like WhatsApp messaging and social media like Facebook or Instagram. Besides, mobile telecommunications enable vital societal functions and critical infrastructure. Examples of this are calling the emergency services at any time and controlling important infrastructures like bridges, locks and electronic road signs remotely. A large part of the data in these services is transmitted over cellular networks. In the Netherlands, Agentschap Telecom (i.e., the Radiocommunications Agency of the Netherlands) is responsible and accountable for monitoring and regulating these cellular networks. The monitoring is performed by a nation-wide measurement network. This network has 15 fixed measurement locations set up all over the country that are collecting data about the spectrum 24 hours a day. This measurement network is used to monitor the adequate use of the frequency spectrum in the Netherlands.

The measurement data is used on regular basis for all types of analysis. However, all data analyses are performed manually. In order to use this data more effectively and accurately, a systematic analysis approach needs to be developed where data processing is automated. In this thesis it is therefore proposed to automatically analyse the data from all 15 measurement locations. There is one other study, [1], where the same data used for this study is being evaluated. There it was found that the network of fixed locations can be used to analyse the spectrum usage of infrastructure-based applications like cellular, military, public safety (C2000) and broadcasting services. However, there was no automated analysis on the spectrum data, which would have enabled quicker and more accurate analysis.

1.2 Problem statement

The main research goal is to propose a methodology to automatically detect outlandish signal variations in the frequency bands reserved for cellular networks. The analysis shall consider the following:

1. Statistical characterisation and modelling of the noise floor.
2. Statistical characterisation and modelling of the received signal fluctuations over time for different detected channels.
3. Statistical characterisation and modelling of the following events: outage, power saving, change in used telecommunication standards and interference.

For the statistical characterisation and modelling, different time scales shall be considered, e.g., minute, hour, 12 hours, day, month and year.

1.3 Report organisation

The remainder of this report is organised as follows. In Chapter 2, the used telecommunication standards are explained. Chapter 3 explains the measurement done by the Radiocommunications Agency of the Netherlands. Chapter 4 describes the used models. Then, Chapter 5 considers the noise floor estimation. Chapter 6 outlines the outlandish signal detection algorithm based on Chapter 5. Chapter 7 discusses the automation of the detection process. The results are presented in Chapter 8. Finally, Chapter 9, presents the conclusions and recommendations.

Chapter 2

Telecommunication Spectrum Definitions

This chapter explains the basic information needed to understand the remainder of the report and states the most important definitions that will be used.

A mobile device, like a mobile phone can communicate with the base station of a cellular network operator. The communication from the base station to the mobile device is defined as the downlink. The communication from the mobile device back to the base station is defined as the uplink. These definitions of uplink and downlink are visualised in Fig. 2.1.

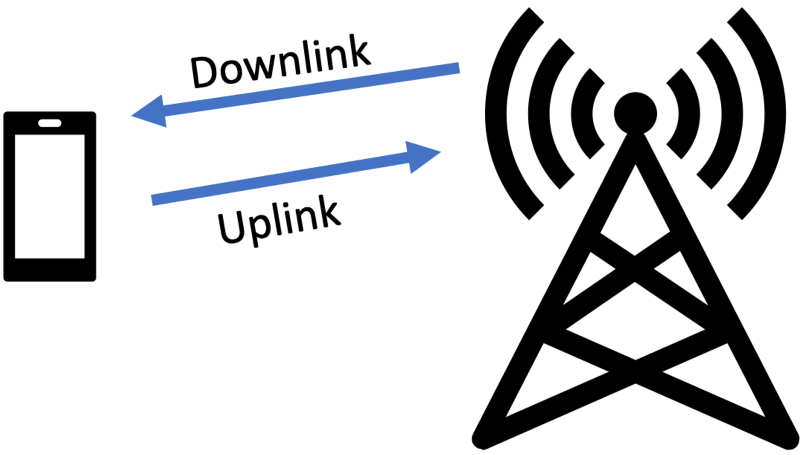

Figure 2.1: Uplink and Downlink

Not every mobile device uses the same technology for communicating with the base station at all times. By February 2019 the following telecommunication standards are used in the Netherlands:

- Second-generation Cellular Technology (2G):
  - Global System for Mobile Communications (GSM)
  - General Packet Radio Service (GPRS)
  - Enhanced Data Rates for GSM Evolution (EDGE)
- Third-generation Cellular Technology (3G):
  - Universal Mobile Telecommunications System (UMTS)
  - High-speed Downlink Packet Access (HSDPA)
  - High-Speed Uplink Packet Access (HSUPA)
- Fourth-generation Cellular Technology (4G):
  - Long Term Evolution (LTE)

Every standard has his own modulation and has his own way to give multiple mobile devices access to the same base station. This results from the fact that every standard has his own method of utilisation of the spectrum. Fig. 2.2 displays the measured field strength for a 5 MHz GSM channel, a 5 MHz UMTS channel and a 20 MHz LTE channel average over one day.

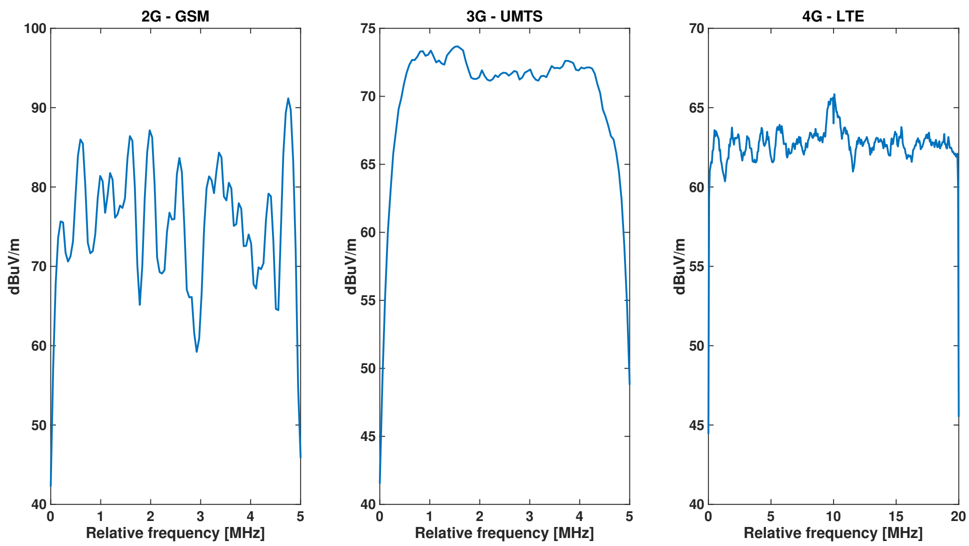

Figure 2.2: Spectrum occupation of 2G, 3G and 4G

To make sure that a mobile device can communicate with base stations worldwide, the 3rd Generation Partnership Project (3GPP) defines the standard frequency bands that can be used for cellular communication. Every defined band has a part of the spectrum that can be used for downlink communication and for most of the bands also have a part of the spectrum that can be used for uplink communication.

A government can licence parts of the 3GPP bands to cellular network operators. In the Netherlands, the bands that are licensed to cellular network operator are technology-neutral. Therefore, a network operator can choose freely between all of the telecommunication standards that he will use for his licensed part of the spectrum. This also means that a network operator can use a different standard at different locations. An operator can choose to switch between standards at any moment in time, for example when there is too much traffic for the used standard while a different standard is able to handle the traffic. The part of the spectrum that is licensed to one operator and uses only one telecommunication standard is defined as a channel. The space between two channels that is not used for any communication is defined as the guard band. The definition of band, guard band and a channel are visualised in Fig. 2.3.

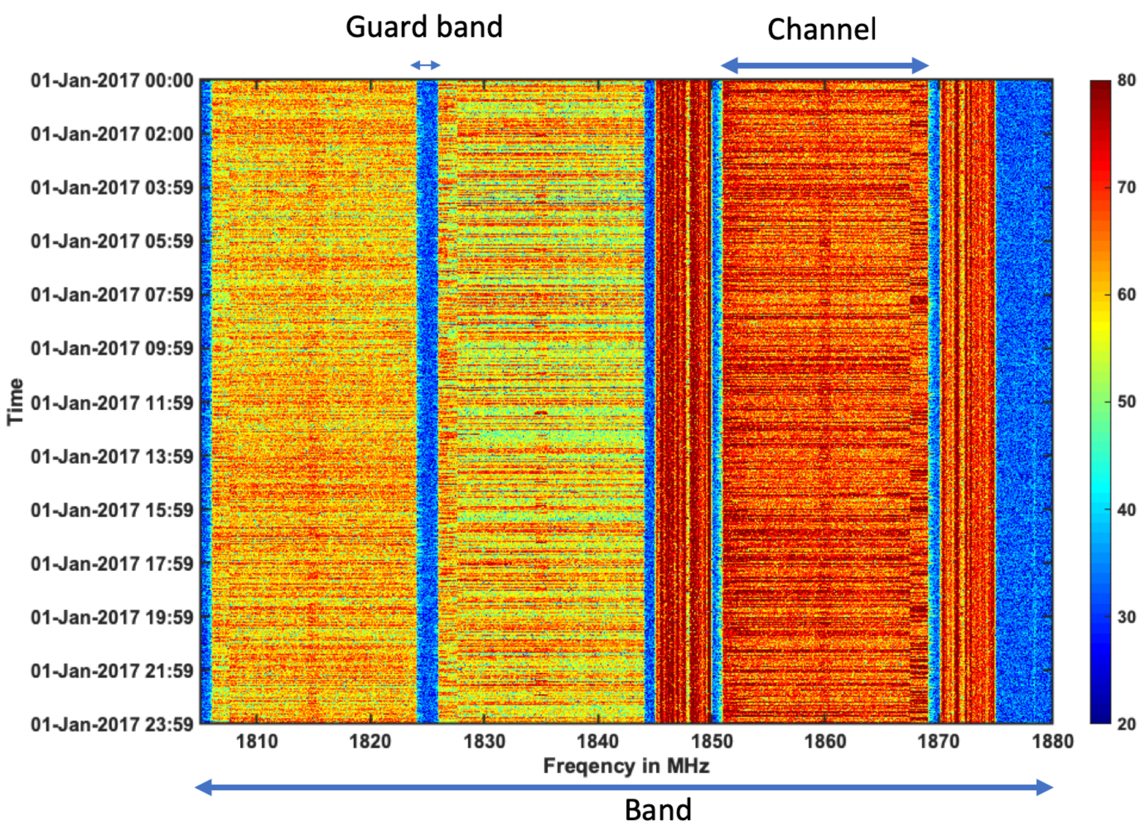

Figure 2.3: Example of spectrum measured field strength in dBuV/m

Chapter 3

Radiocommunications Agency Network Monitoring System

How the measurement is done by the Radiocommunications Agency of the Netherlands, how the data is saved on a central file server is explained in the first half of this chapter. The second half will comment on the noise and the missing data in the measurement.

3.1 Measurement equipment and locations

The measurement locations are tactically placed across the Netherlands. All used measurement locations with their corresponding measurement numbers can be found in Table 3.1. In total the Radiocommunications Agency of the Netherlands has 15 fixed measurement locations, all of these locations are used for this research. A map of the fixed measurement locations in the Netherlands is showed in Fig. 3.1. Nine measurement locations are in an urban environment, four measurement locations are at the countryside, one is at the coast and one is at Schiphol Airport. The setup available at a measurement location is measuring 24/7 and saves the measurement data once a day on a central file server. The data set was gathered from 15 fixed measurement locations of the Radiocommunications Agency of the Netherlands.

Table 3.1: Measurement locations and environment classification

Location	Location name	Environment
1	Heerhugowaard	Urban
2	Wijdemeren	Countryside
3	Hoek van Holland	Coast
4	Breda	Urban
5	Axel	Countryside
6	’t Harde	Countryside
7	Eindhoven	Urban
8	Groningen	Urban
9	Hoogeveen	Urban
11	Nijmegen	Urban
12	Sittard	Urban
14	Leeuwarden	Countryside
54	Schiedam	Urban
55	Schiphol	Airport
72	Hengelo	Urban

Figure 3.1: Measurement locations

Every measurement setup consists of a wide band receiver, a wide band antenna and a PC which collects the measured data. An overview of the measurement setup is visualised in Figure 3.2. At 13 locations a Rohde and Schwarz ESMD wide band Monitoring Receiver is used. The other 2 locations have a Rohde and Schwarz EB500 Monitoring Receiver. All receivers have a measurement frequency range starting at 20 MHz going up to 3.6 GHz. The wideband antenna is a Rohde and Schwarz HK033 Coaxial Dipole. Rohde and Schwarz has specified the gain of the antenna between 80 MHz and 2GHz. The frequency response for this range can be found in Fig. 3.3. The typical horizontal radiation pattern specified by Rohde and Schwarz valid between 80 MHZ and 2 GHz is shown in Fig. 3.4. A desktop PC collects the measured field strength and sends the data to the central file server once a day. All the measurement values have the unit dBμV/m and are saved with a 1 dB accuracy.

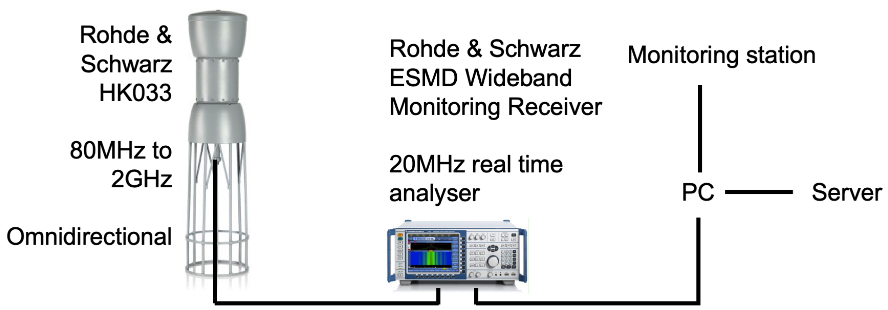

Figure 3.2: Overview of the measurement setup

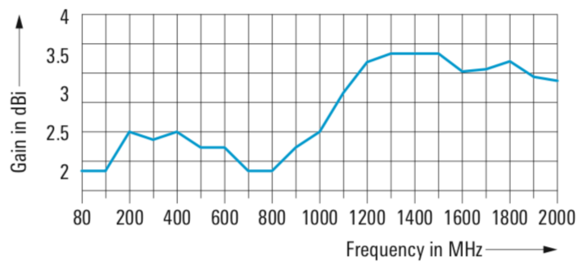

Figure 3.3: Antenna gain as a function of frequency (source: Rohde and Schwarz datasheet)

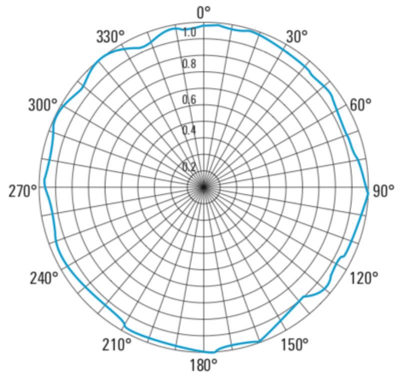

Figure 3.4: Horizontal radiation pattern of the antenna (source: Rohde and Schwarz datasheet)

3.2 Measurement data description

At each measurement location, spectrum data is collected between the 20 MHz and 3 GHz. This spectrum data contains the signal strength sampled every minute. The whole range is split up into 8 different measurement data sets. Each measurement data set has different measurement settings are used, e.g., measurements are taken with different measurement bandwidth. The specifications of the 8 different measurements can be found in Table 3.2.

In this report only the frequency bands operated by cellular networks are of interest. All the frequencies that are licensed to cellular network operators are listed in Table 3.3. Hence, it is concluded from Tables 3.2 and 3.3 that only measurement frequency bands 5, 7 and 8 are useful for the present study.

Table 3.2: Measured data sets

Measurement data set	Start frequency (MHz)	End frequency (MHz)	Measurement bandwidth (kHz)
1	20	87.5	5
2	87.5	108	50
3	108	137	2
4	137	470	5
5	470	863	200
6	863	870	5
7	870	1900	50
8	1900	3000	200

Table 3.3: Cellular bands in the Netherlands, source: antenneregister.nl (July 2018)

3GPP Band #	Downlink (MHz)	Uplink (MHz)	Bandwith (MHz)	Standard
20	796.0000	837.0000	10	4G
20	802.3025	843.3025	0.18	NB-IoT
20	806.0000	847.0000	10	4G
20	816.0000	857.0000	10	4G
8	930.0000	885.0000	10	2G
8	937.3000	892.3000	3	2G
8	942.2000	897.2000	5	3G
8	950.0000	905.0000	10	4G
8	954.7025	909.7025	0.18	NB-IoT
8	957.4000	912.4000	5	3G
3	1815.0000	1720.0000	20	4G
3	1835.0000	1740.0000	20	4G
3	1847.5000	1752.5000	5	2G
3	1860.0000	1765.0000	20	4G
3	1875.0000	1775.0000	10	2G
1	2112.8000	1922.8000	5	3G
1	2117.6000	1927.6000	5	3G
1	2122.4000	1932.4000	5	3G
1	2127.4000	1937.4000	5	3G
1	2132.2000	1942.2000	5	3G
1	2134.7000	1944.7000	5	4G
1	2137.2000	1947.2000	5	3G
1	2142.2000	1952.2000	5	3G
1	2142.2000	1952.2000	5	4G
1	2144.7000	1954.7000	5	4G
1	2147.2000	1957.2000	5	3G
1	2152.2000	1962.2000	5	3G
1	2157.2000	1967.2000	5	3G
1	2162.2000	1972.2000	5	3G
1	2167.2000	1977.2000	5	3G
38	2580.0000	NaN	20	4G
38	2605.0000	NaN	20	4G
7	2625.0000	2505.0000	10	4G
7	2640.0000	2520.0000	20	4G
7	2652.5000	2532.5000	5	4G
7	2660.0000	2540.0000	10	4G
7	2675.0000	2555.0000	20	4G

3.3 Noise in the measured signal

The measurement noise plays an important role in the present study. It depends on several factors that are explained in this section. The first factor is that the measured noise is correlated to the measurement bandwidth used during measurements. Indeed, it was noted that the noise floor in a measurement depends on the measurement bandwidth setting in the spectrum analyser. A higher measurement bandwidth results in a higher noise floor. This effect is clearly visible when comparing Fig. 3.5 and Fig. 3.6. Fig. 3.5 shows the daily average for a specific frequency band. The used measurement bandwidth is 5 kHz and the measurement is performed between 863 MHz and 870 MHz. The noise floor at 870 MHz is around 14 dBμV/m. Fig. 3.6 shows the measurement data of the same day. However, the used measurement bandwidth in this measurement was 50 kHz. The noise floor at 870 MHz for this measurement with exactly the same equipment for the same day is 24 dBμV/m, which gives a difference of 10 dB. The difference can be explained by calculating the theoretically minimum detected noise floor. The minimum detected noise floor is given by:

$$Noisefloor_{dBm} = 10\log_{10}(k \cdot T_0 \cdot 1000) + NF + 10\log_{10}(BW), \tag{3.1}$$

where k is the Boltzmann constant, T_0 the temperature in Kelvin, NF the receiver noise figure and BW the measurement bandwidth in Hz. Since only the measurement bandwidth differs between the two measurement data sets, the theoretical difference between the two measured noise floors is given by:

$$Noisefloor_{diff,dB} = 10\log_{10}\left(\frac{Bandwidth\ data\ set\ 7}{Bandwidth\ data\ set\ 6}\right) = 10\log_{10}\left(\frac{50}{5}\right) = 10dB \tag{3.2}$$

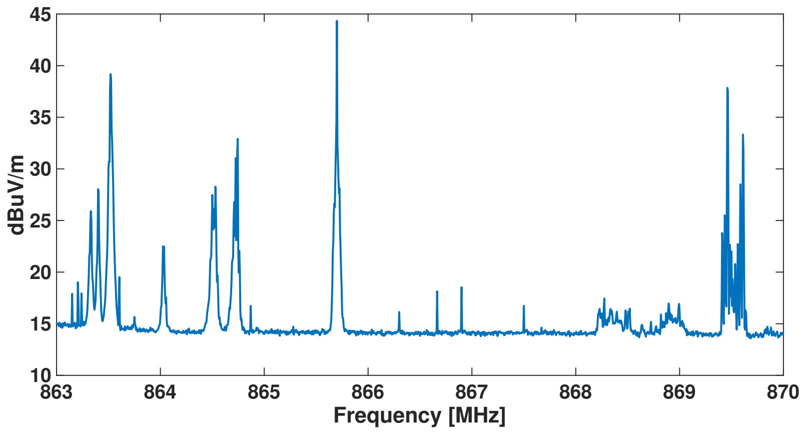

Figure 3.5: Measurement data set 6, used measurement bandwidth 5 kHz

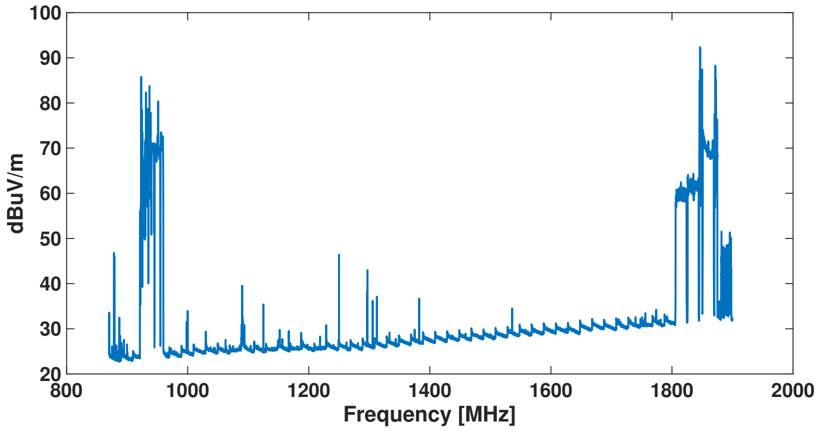

Figure 3.6: Measurement data set 7, used measurement bandwidth 50 kHz

The second factor that explains the noise has to do with the used antenna. The used Rohde and Schwarz antenna is specified between the 80 MHz and 2 GHz. So only for data between the 80 MHz and 2 GHz the gain and the radiation patterns of the antenna are given. However, the measurements are performed between 20 MHz and 3 GHz. For this reason, the measurement results between 20 MHz and 80 MHz cannot be used to draw conclusions. The same holds true for the measurement results between 2 GHz and 3 GHz. In this frequency range, there are three 3GPP bands, namely 1, 38 and 7. Therefore, this research will mainly focus on bands that are within the 80 MHz - 2 GHz range and are used for cellular networks in the Netherlands. This are 3GPP bands 20, 8 and 3 as listed in Table 3.3.

Another factor that impacts the measured noise is the height of the measurement antenna. Since the height is different at every measurement location, the length of the cable between the antenna and the spectrum analyser is not the same at every location. Consequently, this change in cable length can result in a different measured noise level at the spectrum analyser. Especially at higher frequencies, a longer cable will result in higher losses. This difference in noise is clearly visible by comparing Fig. 3.7 and Fig. 3.8. The noise floor at 1800 MHz for the measurement in Hoogeveen is around 36 dBμV/m, while the noise floor at the same frequency measured at Schiphol Airport is around 24 dBμ V/m. This a difference of 12 dB.

Another factor contributing to the noise is the way the measurement equipment works. The spectrum analyser is able to do a real time analysis of 20 MHz at once. To do a measurement with a larger range than 20 MHz, the spectrum analyser is doing a sweep over the whole range with a step of 20 MHz. A spike with an exponential decay every 20 MHz is clearly visible in Fig. 3.6 and 3.7. After communication with the Radiocommunications Agency of the Netherlands, it is verified that this spike with an exponential decay is the result of a calibration done by Rohde &

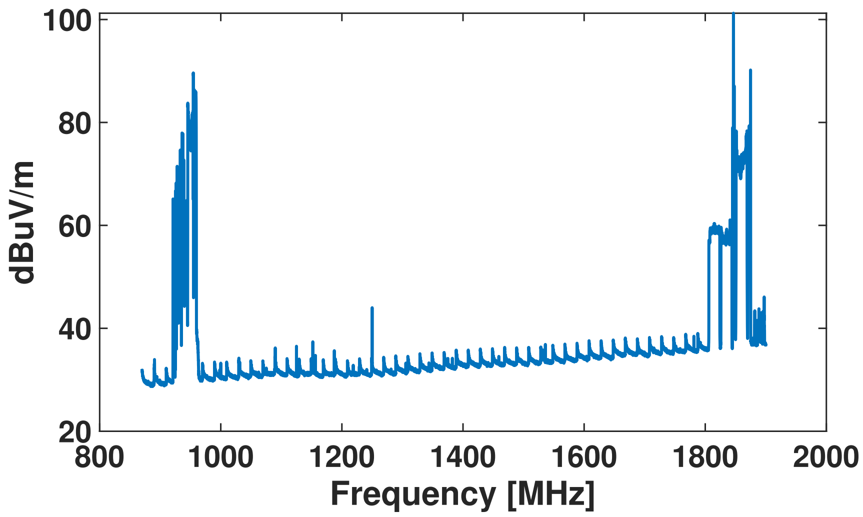

Figure 3.7: Measurement data set 7 measured at Hoogeveen

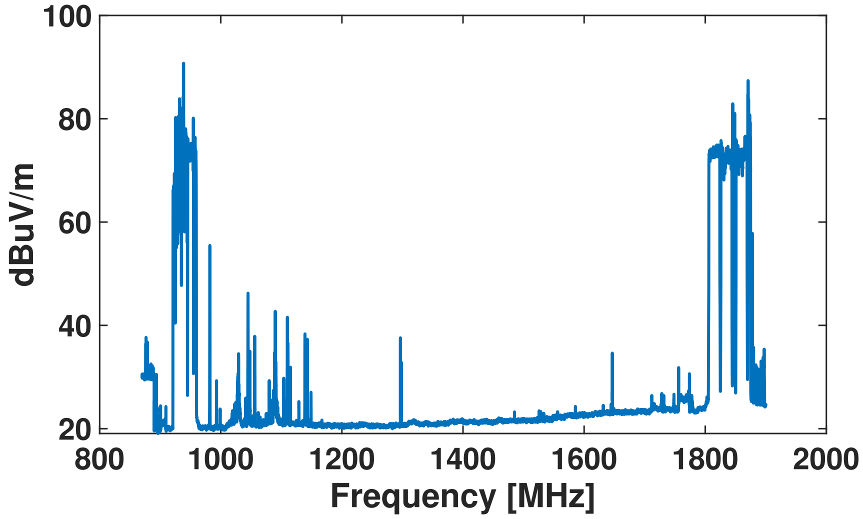

Figure 3.8: Measurement between 870 and 1900 MHz at Schiphol Airport

Schwarz. This calibration compensates for the filtering used in the spectrum analyser. However, as part of the study it was verified by the Radiocommunications Agency of the Netherlands that the measured signal levels are indeed correct for the whole measurement range.

Noise classification

To test if the noise in the measurement is additive white Gaussian noise (AWGN) multiple verifications are done. First, it is visually verified that the measured noise is Gaussian distributed. To do that, 1000 noise samples in the guard between two channels at 1825 MHz measured in Heerhugowaard in 2017 are taken. With these

1000 noise samples a normalised CDF is made and plotted in Fig. 3.9. All the measured samples are normalised with equation 3.3. The estimated mean $\hat{\mu}$ was 31.2 dBμV/m and the estimated standard deviation $\hat{\sigma}$ was 2.1 dB.

$$Y_{i,normalised} = \frac{(Y_i - \hat{\mu})}{\hat{\sigma}} \tag{3.3}$$

The CDF of the normalised noise is compared with the standard normal CDF. As showed in Fig. 3.9 the standard normal CDF is within the 95% confidence bounds of the CDF of the normalised noise.

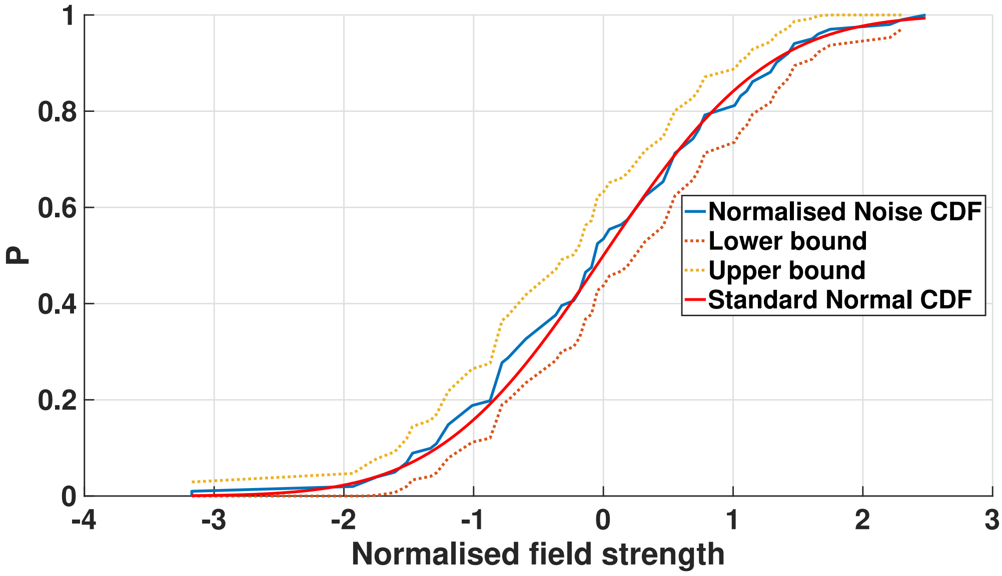

Figure 3.9: Normalised CDF of 100 sample measured noise between 1824.85 and 1825.10 MHZ at Heerhugowaard in 2017

Second, a χ^2test on the same 1000 noise samples is performed. This test shows with a significance level of 90.52% that the noise samples are Gaussian distributed. The χ^2 value of the test can be computed with:

$$\chi^2 = \sum_{i=-\infty}^{\infty} \frac{(O_i - E_i)^2}{E_i}, \tag{3.4}$$

where E_i is the expected number of samples with value i based on the standard normal distribution and O_i are the observed number of samples with value i.

Third, the correlation matrices of the guard band at 1825MHz is also computed. When the noise is AWGN all the measured samples that do no contain a wanted signal, should be independent and there should be no correlation between the different frequency components. The correlation coefficients used within the matrices are the Pearson correlation coefficients and can be computed with:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{3.5}$$

Where X and Y are the frequency components, cov the covariance and σ the standard deviation.

For the guard band around 1825 MHz measured in Heerhugowaard in 2017, the computed correlation matrix is displayed in Fig. 3.10. From this correlation matrix it is concluded that the maximum correlation component between two frequency components is 0.08. Where 0 means no correlation and ±1 means the strongest possible correlation [2]. This relative low correlation together with the visually verification and the χ^2test, the assumption is made that the noise is indeed AWGN. This assumption will be used in Chapter 5, where the noise floor is estimated.

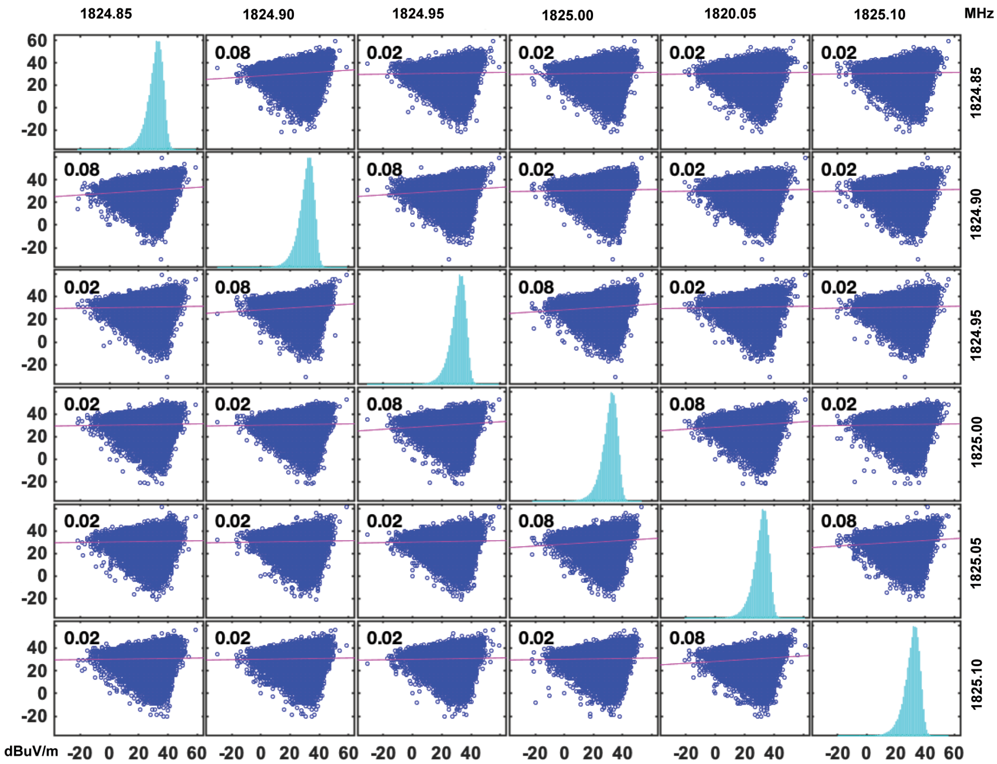

Figure 3.10: Correlation matrix guard band between 1824.85 and 1825.10 MHz at Heerhugowaard in 2017

3.4 Missing data

The measurements should take place 24/7. However, there is a relative high amount of data that is missing. It is assumed that this is mainly caused by two reasons:

1. When something breaks down at a measurement location there is some time needed to repair the measurement equipment.
2. Another function of this measurement equipment is to stream real time data from the spectrum analyser back to a central monitoring station. When somebody turns on this streaming function, the spectrum analyser is not able to do other measurements and will therefore cause missing data points.

Hence, in order to get an estimate of the number of useful data points in a measurement data set, the percentage of available measurement points over the period comprising 4 years from 2015 to 2018 is computed according to the equation

$$Percentage\ of\ useful\ data = \frac{Number\ of\ measured\ samples}{Maximum\ number\ of\ samples} \cdot 100\% \tag{3.6}$$

The percentage of useful data for each measurement is summarised in Table 3.4 for each measurement location. The overall percentage of useful data is about 85 %.

Table 3.4: Percentage of useful measured samples

Location	2015	2016	2017	2018
1	98.1%	98.4%	97.1%	96.8%
2	82.3%	95.5%	94.2%	86.5%
3	95.6%	92.0%	45.8%	89.4%
4	96.7%	97.5%	96.3%	92.3%
5	90.0%	94.6%	97.4%	96.0%
6	96.9%	97.9%	97.4%	89.5%
7	97.7%	94.3%	96.1%	84.3%
8	99.3%	97.9%	96.5%	95.3%
9	98.2%	95.6%	93.9%	93.3%
11	92.3%	91.4%	97.1%	93.9%
12	80.2%	97.9%	97.0%	88.2%
14	97.7%	89.1%	94.5%	91.1%
54	89.8%	78.0%	46.7%	84.2%
55	71.0%	81.4%	84.5%	79.6%
72	0.0%	61.3%	96.2%	85.6%
Average	**86.3%**	**86.9%**	**83.2%**	**84.1%**

Chapter 4

System Model

4.1 Signal + interference + noise model

In this section, a simple signal model describing the total measured signal strength Y at a single frequency within the considered channel is introduced.

$$Y_i = X_i + \Theta_i + Z_i \tag{4.1}$$

Where the measured channel Y represented by multiple samples Y_i with discrete time index i. The sum of the wanted signal X_i, the outlandish signal level variations Θ_i and the noise Z_i is defined as Y_i (4.1). The noise Z_i is assumed to be additive white Gaussian noise and can be modelled as a Gaussian distributed variable (4.2), with zero-mean and variance N (4.3). The distribution of the wanted signal and the outlandish signal level variations are unknown.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{4.2}$$

Where σ^2 is the variance, μ the mean, and x the random variable.

$$Z_i \sim \mathcal{N}(0, N) \tag{4.3}$$

Since (X_i),(Θ_i) and (Z_i) cannot be directly measured, all analysis must be done on the measured sample Y_i.

To analysis measured sample Y_i, four cases are defined.

Case 1: The measured signal is noise

$$Y_i = Z_i. \tag{4.4}$$

Case 2: The measured signal is the wanted signal plus noise

$$Y_i = X_i + Z_i. \tag{4.5}$$

Case 3: The measured signal is interference plus noise

$$Y_i = \Theta_i + Z_i. \tag{4.6}$$

Case 4: The measured signal contains the wanted signal, the interference signal and noise

$$Y_i = X_i + \Theta_i + Z_i. \tag{4.7}$$

4.2 Threshold model and derived signal detection

In this section the automatic detection of signals in the presence of noise is described. For this purpose, a threshold model is considered. A threshold T_h defines the signal level below for which the measured sample is assumed to contain only noise. Hence, according to the threshold model, it is assumed that the Y_i contains noise and a wanted signal if

$$Y_i > T_h \tag{4.8}$$

Otherwise if

$$Y_i \leq T_h \tag{4.9}$$

the measured signal strength contains noise only.

Based on the above model multiple types of errors can be identified. For example, if a sample Y_i only contains noise Z_i, but is larger than the threshold T_h, it is concluded the that the sample Y_i contains a wanted signal X_i, which is not true. This assumption results in a false positive error. In total 4 errors are defined and displayed in Table 4.1.

Table 4.1: Truth Table

	Sample is noise	Sample contains a wanted signal
$Y_i > T_h$	False positive (FP)	True positive (TP)
$Y_i \leq T_h$	True negative (TN)	False negative (FN)

For every threshold T_h a false positive (FP), true positive (TP), false negative (FN) and true negative (TN) rate can be computed. The amount of false positives dived by the total number of samples N_s is the false positive rate (P_{fp}). On the same way the true positive rate (P_{tp}), false negative rate (P_{fn}) and true negative rate (P_{tn}) can be computed.

$$P_{fp} = \frac{FP}{N_s} = \frac{FP}{FP + TN} \tag{4.10}$$

$$P_{tp} = \frac{TP}{N_s} = \frac{TP}{TP + FN} \tag{4.11}$$

$$P_{fn} = \frac{FN}{N_s} = \frac{FN}{TP + FN} \tag{4.12}$$

$$P_{tn} = \frac{TN}{N_s} = \frac{TN}{FP + TN} \tag{4.13}$$

Since the noise floor can be modelled as a Gaussian distributed variable, the noise and the wanted signal always have an overlapping part. Hence, theoretically under this assumption, there can never be a threshold with a 100% signal detection rate and a 0% false positive rate. A threshold is therefore always a trade-off between the true positive rate and the corresponding false positive rate. This is illustrated in Fig. 4.1. The figure shows two Gaussian distributed signals with arbitrarily chosen mean and variance. In most working communication systems, the noise is lower than the wanted signal, therefore the blue line in this example is illustrating noise. The red line is the wanted signal. The filled red part in fig 4.1 corresponds to the false positive rate and the blue part to the false negative rate. 4.1.

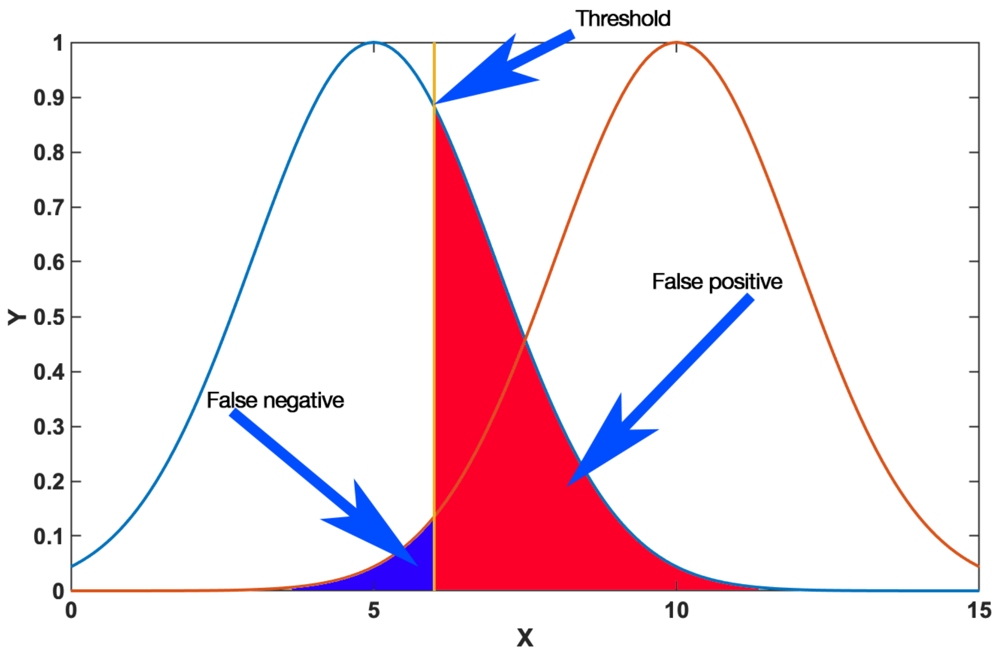

Figure 4.1: Example of a threshold with 2 Gaussian sources.

Chapter 5

Noise floor estimation

As defined above a threshold T_h defines when a measured sample Y_i is detected as noise Z_i or as a wanted signal X_i. The noise floor estimation algorithms considered here are based on threshold level detection. Roughly, two main different methods can be defined: the fixed threshold method [3], and the dynamic threshold method [4]. Both methods will be explained below, but only the latter type will be used in this thesis.

Fixed threshold

A fixed threshold is a threshold which is, as the name suggests, fixed, i.e., not variable. This fixed threshold will therefore be the same for every measurement and will not change for different noise levels [5]. This threshold will not depend on the actual measured samples. The threshold must be defined before the actual measurement and can be based on the information found in the literature, a guess, or a single reference measurement. Using this method, can therefore result in unexpected high false positive or false negative rate. This is the reason why this method will not be used.

Dynamic threshold

A dynamic threshold depends on the actual measured samples and can therefore only be obtained after gathering the measurement data [4]. As found in [6], the estimated threshold will be more accurate if a dynamic threshold is used. Dynamic thresholds can be computed in multiple ways, two of them are, e.g., the forward consecutive mean excision (FCME) algorithm [7] and the adjusted median forward consecutive mean excision (MED-FCME) algorithm [8]. Both algorithms are described in detail below.

FCME

The FCME algorithm starts by computing a T_{CME} parameter. This parameter depends on the distribution of the noise and a given false positive rate P_{fp}. The false positive rate is defined in equation 4.10. When assuming that this is AWGN, the noise samples Z_i should follows a Gaussian distribution, as defined is equation 4.3, the energy of the noise samples $|Z_i|^2$ should follow the Chi-squared distribution with two degrees of freedom k and random variable x:

$$f(x;k) = \begin{cases} \dfrac{x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\Gamma\left(\dfrac{k}{2}\right)}, x > 0 \\ 0, \text{otherwise.} \end{cases} \tag{5.1}$$

The T_{CME} parameter needed for the FCME algorithm is given by [9]

$$T_{CME} = -ln(P_{fp}), \tag{5.2}$$

where ln is the natural logarithm.

Beside the computation of the T_{CME} parameter, the noise-only sample set needs to be determined. The noise-only sample set contains samples that do not contain any signal information. This starts with reordering samples $|Z_i|^2$ from small to large. Then, the first p % of the smallest samples are taken. It is assumed that this p % only contains noise. The number of samples in this set is denoted by q and the set itself is denoted as Q. As found in the literature, a commonly used value for p is 10 %. [8]. A higher value of p will result in a larger noise-only sample set and therefore there is a higher probability that the noise-only sample set contains a wanted signal. When there is a wanted signal in the noise-only sample set the estimated noise floor threshold will be to high. A lower p value will result in a longer computation time of the algorithm.

To find the definitive threshold, an iterative process will start. Every iteration has four steps. First, the mean of the noise-only sample set is calculated. Second, a temporary threshold is calculated by multiplying the mean with the T_{CME} parameter. Third, the calculated threshold is compared to the smallest value of the measured sample set Y that is not part of the noise-only sample set Q. The fourth step is the decision step. If the threshold is higher than the smallest value, the process will start again. The smallest value will then be part of the noise-only sample set, since it can be considered as noise. When the threshold is lower than the smallest value, the iterative process ends since the smallest value is considered a signal. This also means that the temporary threshold computed in step 1 will be become the threshold of this data set [9]. The flow diagram of the FCME algorithm described above is shown in Fig. 5.1.

MED-FCME

The FCME algorithm is according to [8] not robust against large outliers, since the influence of one large outlier will have a large impact on the mean. This can result in a higher threshold than expected. To improve the robustness against large outliers of the FCME algorithm, the calculation of the mean can be replaced by a calculation of the median. This algorithm is called the MED-FCME algorithm [5]. By taking the median, the influence of large outliers is much less than when taking the mean. Therefore, the expected range of a threshold computed with the MED-FCME algorithm is smaller than the expected range of a threshold computed with the FCME [5]. In [8], this difference was corroborated shown on the analysis of data gathered at the 2,4 GHz ISM band. In this thesis the MED-FCME algorithm is therefore used for the noise floor estimation.

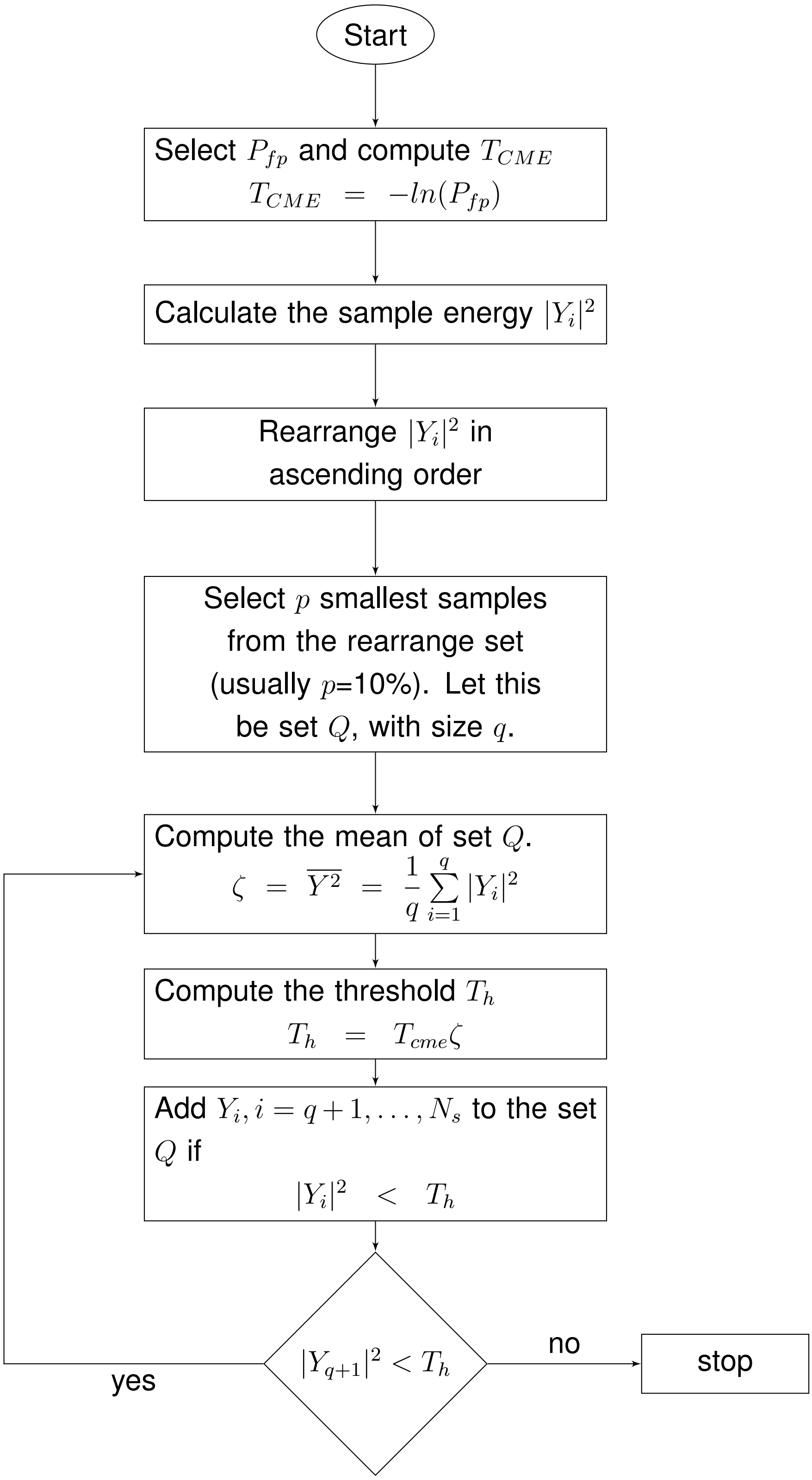

Figure 5.1: Flowdiagram of the FCME algorithm

Chapter 6

Outlandish signal level detection

This chapter describes how an outlandish signal level is detected or more specifically, this chapter focus on Cases 3 and 4 described in section 4.1. The first step is to compute the statistics of all measured samples (Y_i). The statistics that are calculated are the mean, median, standard deviation (SD), minimum, maximum, the dynamic range and the cumulative distribution function (CDF). How all these statistics are calculated is explained below. After the computation of the statistics the different statistics are compared and some of the significant differences are classified as an event.

6.1 Computing statistics

All the measured samples Y_i are given in $dB\mu V/m$. To compute some of the statistics of a channel, all measured samples Y_i should be first converted to linear scale Y_i. The relation between between y_i and Y_i is given by:

$$y_i = v_0 \cdot 10^{\frac{Y_i}{20}} \tag{6.1}$$

Where v_0 is the reference voltage, which is 1 μV.

$$v_0 = 1\mu V \stackrel{\text{def}}{=} 0dB\mu V \tag{6.2}$$

Mean

The mean of a data set Q with size q can be computed with the following function:

$$\hat{\mu}_y = \frac{1}{q}\sum_{i=1}^{q} y_i \tag{6.3}$$

Median

To compute the median all samples y_i should be reordered from small to large. This reordered is set is defined as set r, with size s. When s is even the median of r can be computed with:

$$r_{median} = \frac{r_{s/2} + y_{s/2+1}}{2} \tag{6.4}$$

When s is odd the median of r can be computed with:

$$r_{median} = r_{(s+1)/2} \tag{6.5}$$

Standard Deviation

In contrast to the mean and the median the standard deviation of the measurement data can be directly computed in dB-scale. Therefore, the unbiased standard deviation in dB-scale of a set with size q is given by:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{q} |Y_i - \hat{\mu}_Y|^2}{q-1}} \tag{6.6}$$

Minimum, maximum and dynamic range

For every detected channel, the minimum and maximum field strength are taken. The difference between this minimum and maximum is the dynamic range DR of a channel.

$$DR = max(Y) - min(Y) \tag{6.7}$$

Cumulative distribution

The cumulative distribution CDF_Y for the sample set Y, with probability $p_i = P(Y_i)$ is given by:

$$CDF_Y(Y_i) = P(Y \leq Y_i) = \sum_{-\infty}^{Y_i} P(Y = Y_i) = \sum_{-\infty}^{Y_i} p(Y_i) \tag{6.8}$$

For the automatic analysis the value $CDF_Y(T_h)$ will be used. This value should correspond to the probability that a channel does not contain a wanted signal. In a situation where all measured samples contain a wanted signal, this value should correspond to the chosen false positive rate together with the false negative rate.

If Y only contains a wanted signal:

$$CDF_Y(T_h) = P_{fp} + P_{fn} \tag{6.9}$$

6.2 Comparing channels

The computed statistics of the different channels are compared between the different locations and between different moments in time. How this is done, is explained in the two sections below.

6.2.1 Different time windows

To compare statistics of the same location for different moments in time, the derivative is used.

$$\dot{S} = \frac{S_t - S_{t-1}}{\Delta t} \tag{6.10}$$

Where S can be any of the statistics computed above for different moments in times as long as Δt is the same. The values that Δt will have in this analysis are: 1 minute, 12 hours, 1 day, 1 week, 1 month and 1 year.

For each of the statistics, the derivative can be calculated for different time frames. The result of each derivative gives the variation of the statistic. A relative high derivative means a change in the signal which needs further examination and is therefore indicated as outlandish signal. A relative low derivative is expected and therefore needs no further examination. What a relatively high or relative low value is, is defined by a threshold. This threshold can be computed with the same method that was used for the noise floor.

6.3 Significant event detection

Every significant variation is detected by a significant change in the derivative as explained above. Some of these significant variations are classified as an event. In total 4 different events are classified, namely: outage or maintenance, power saving, a change in telecommunication standard or interference.

6.3.1 Outage or Maintenance

A full outage of one base station is detected when the mean of a channel for 1 minute is below a threshold for the estimated noise floor. The threshold is computed with the MED-FCME algorithm. The length of an outage is the number of minutes where the mean is below this threshold.

6.3.2 Power saving

A power saving event is almost the same as an outage of a base station, since in both cases the base station is turned off. A power saving event can be discovered in the same way as a full outage of a base station can be discovered. However, power saving is an intended choice of a network operator and an outage of a base station most of the time is not. It is also expected that these power saving events will repeat every week and occur when there are less people using the mobile network. Therefore, an outage is classified as a power saving event if it is repeated on a weekly basis.

6.3.3 Change in telecommunication standard

Since the measurement setup only measures field strength and does not do any kind of signal demodulation, it is not possible to directly know the technology that an operator is using on a specific frequency band. However, there are a few properties that can be measured and that are correlated to a technology. The first property that can be measured is the channel width. Every technology has its own specification of the guard band. This guard band is a part of the spectrum that is not used between two channels. Therefore, a change in the measured channel width can be an indication for a technology change. A second property that can give an indication for a technology change is a significant change in statistics. Therefore all statistics that are described in section 6.1 are compared with the statistics that are computed for the previous day. When the statistics between two days show a significant difference, it possible that a network operator changed the technology used in that frequency band.

6.3.4 Interference

There are multiple ways to detect interference. One of them is to compare the channel widths with the expected channel widths. According to the current 3GPP standards, the smallest channel width that can be used for cellular communication is 1.4 MHz [10]. Taking the guard band into account, the maximum occupied bandwidth for a 1,4 MHz channel is 1.08 MHz. The maximum measurement bandwidth used in the measurement was 200 kHz. Therefore, the measurement setup will measure the channel width of a 1.08 MHz wide channel as 1MHz or as 1,2 MHz. Therefore, any detected channel with a channel width smaller than 1 MHz, is not expected to be a regular channel used for cellular networks and its therefore also identified as interference. This corresponds to Case 3 (4.6). The second method to detected interference is the same as for the technology change. If there is interference within a channel, the statistics of that specific channel will change. If this change is significant, it can be that there is in-channel inference.

Chapter 7

Automate detection process

Now the noise floor estimation, the outlandish signal level detection and event detection are put together in an automated process. This automation process is able to detect the defined cases in section 4.1. The details of the process are explained below and displayed in Fig. 7.1.

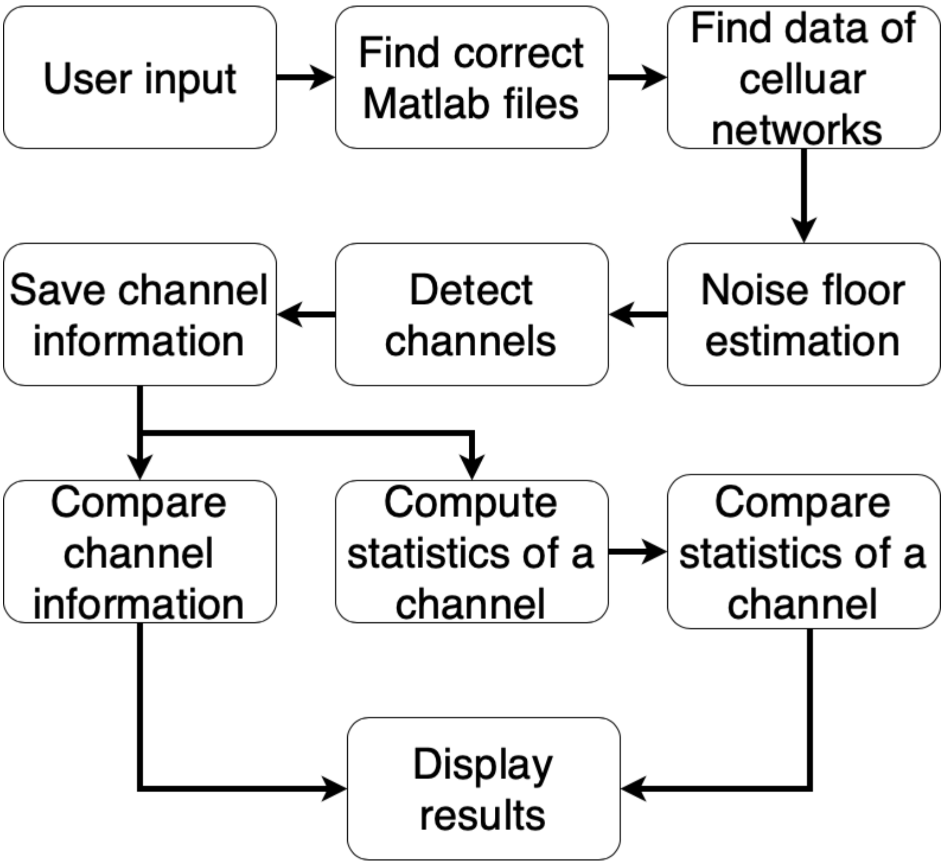

Figure 7.1: Flowchart of the automation process

7.1 Data preparation

All the measurement data is saved on a file server located at the Radiocommunications Agency. The measurement files on this server contain information about the whole measured spectrum. Since the spectrum needed for this research should be allocated for cellular networks, only a small part of the spectrum measured by the agency is needed. This needed part of the spectrum is taken out of the entire data set and stored separately. The search for the right measurement files is done with a MATLAB script. A detailed description of this script can be found in Appendix A.

7.2 Noise floor estimation

After the data is prepared, a threshold for the noise floor is estimated for every measurement location. This is done at every location for every 3GPP band individually. This is done by implementing the MED-FCME algorithm as explained in Chapter 5. The threshold estimated by the algorithm is saved and used for the channel detection algorithm as explained in the next section.

7.3 Cellular channel detection

The goal of the cellular channel detection algorithm is to find the parts of the spectrum where Case 2 (4.5), a wanted signal with noise, and case 4 (4.7), a wanted signal with an outlandish signal and noise are true. These are the cases where the measured sample contains a wanted signal. As defined in section 4.2 a part of the spectrum is assumed to contain a signal if a measured value is higher than a certain threshold (4.8). This situation corresponds to case 2 (4.5), case 3 (4.6), a outlandish signal with noise, and case 4 (4.7). The threshold used for this decision is the threshold estimated with the MED-FCME algorithm. The next step is to identify case 3 (4.6). As described in section 6.3.4 a detected channel with a channel width smaller than 1MHz cannot be a used for cellular communication and is identified as an outlandish signal, without a wanted signal. Which corresponds to case 3 (4.6). Therefore, the cellular channel detection algorithm computes the channel width of every part of the spectrum that is above the threshold. The parts that are above the threshold and have a channel width larger than 1MHz are used for the statistics computation explained below.

7.4 Outlandish signal level detection

For a channel, all detected channels with a channel width larger than 1 MHz the statistics that are described in section 6.1 are computed. When combining table 3.2 and table 3.3 it is found that all individual channels are larger than the measurement bandwidth used in the measurements. Therefore, all detected channel samples in both frequency and time can be represented by an array. The number of columns equals the number of time frequency samples of this matrix is given by:

$$N_f = \frac{Channel\ width}{Measurement\ Bandwidth} \tag{7.1}$$

The number of rows of the matrix equals the number of time samples $(N_t$ and is given by:

$$N_t = \frac{total\ measurement\ time}{sample\ time} \tag{7.2}$$

This array will be used to compute the statistics. The be sure that the edges of the channel do not have an impact on the statistics, only the middle 80% of the channel width will be used when computing the statistics. This allows for a guard of 10% at both sides of the channel.

The comparison of the channels is done using the derivative and difference method explained in section 6.2. After the outlandish signal level detection, the different events described in section 6.3 are detected.

Chapter 8

Results

8.1 Noise floor estimation

The noise floor is estimated for all locations and for all cellular network bands in table 3.3 at every day of the year. The corresponding threshold is computed with the FCME and MED-FCME algorithm. As expected, a threshold computed with the FCME algorithm has more fluctuations, than the threshold found with the MED-FCME algorithm. In Fig. 8.1 the computed thresholds for band 3 in 2017 are displayed. For some days there was no measurement data available, at those days there is no threshold detected. This explains the discontinuity in the graph. The thresholds detected with the MED-FCME algorithm are showing, except for day 151, a difference between the maximum and minimum value of 3dB. The threshold found with the FCME algorithm shows a difference of 32dB. The large deviation at day 151 can be explained by faulty equipment at a certain time for that day. At that day there are 44 minutes of data missing and for 1 minute all the 1501 measurement samples for band 3 at location 1 are 0.

For bands that are crowded and where the guard band between channels is small compared to the measurement bandwidth, both algorithms fail to estimate a noise floor. An example is 3GPP band 8.

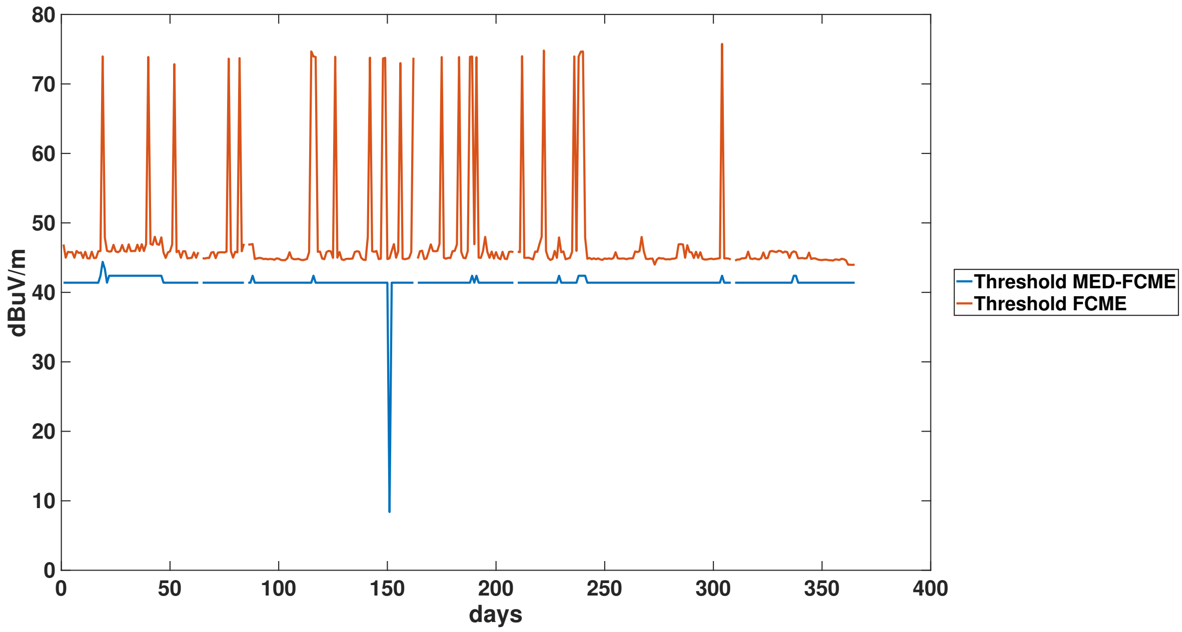

Figure 8.1: Noise floor threshold at Location 1 for Band 3 in 2017

8.1.1 Detected channels

The channel detection algorithm tries to detect channels at all cellular network bands displayed in Table 3.3 for every day of the year at every measurement location individually. In Fig. 8.2 are the detected channels at the measurement location in Heerhugowaard for the downlink of 3GPP band 3 in 2017. The noise floor threshold is found with the MED-FCME algorithm, which is plotted in Fig. 8.1. At most days in 2017, the algorithm detects 5 channels with a channel width greater than 1MHz. The estimated channel width of these 5 channels equals the expected bandwidth given in Table 3.3. Detected channels in band 3 that are smaller than 1MHz have been detected on 14 days of the 2017 yearly recorded data at the Heerhugowaard site. For the 5 days that there were no channels detected at all, there was no data available, i.e., no measurement samples were recorded during those days. In Table 8.1 are the number of days in 2017 where the amount of detected channel corresponds to the amount of licensed cellular channels displayed. Taking into account that there are 5 days in 2017 where there was no data available, the channel detection success rate was 100% for band 20 and 99.7% for band 3. For the other 4 bands the number of detected bands does not correspond to the know licensed cellular channels. For band 7 and band 38 this is due to the fact that, in Heerhugowaard, less channels are in use by network operators than the number of licensed channels. For band 1 and 8 this is due to the fact that guard band between (mainly) 3G channels is smaller compared to the used measurement bandwidth of the measurement. The number of detected channels is therefore lower than the actual active number of channels.

That the channel detection algorithm has difficulties with the detection of all the individual channels in 3GPP band 1 is also showed in Fig. 8.3. The number of detected channels should be 15, however as displayed, the maximum amount of detected is 10.

Table 8.1: Days in 2017 where the amount of detected channel corresponds to the amount of licensed cellular channels.

	band 1	band 3	band 7	band 8	band 20	band 38
location 1	0	359	0	5	360	3

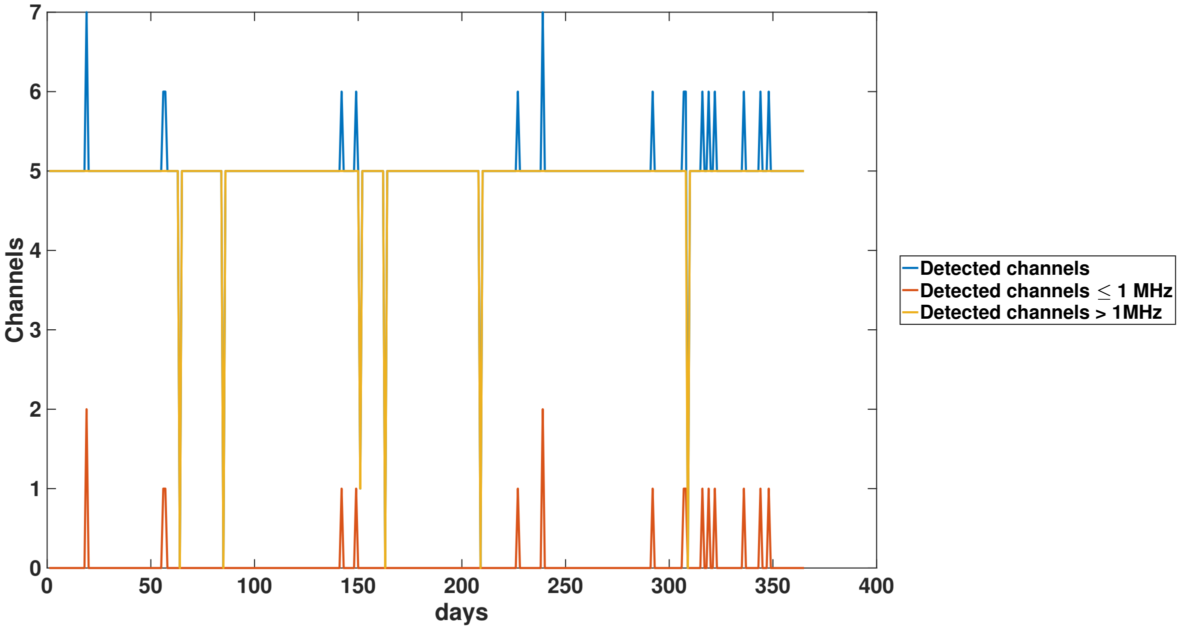

Figure 8.2: Detected channels in Heerhugowaard for Band 3 in 2017

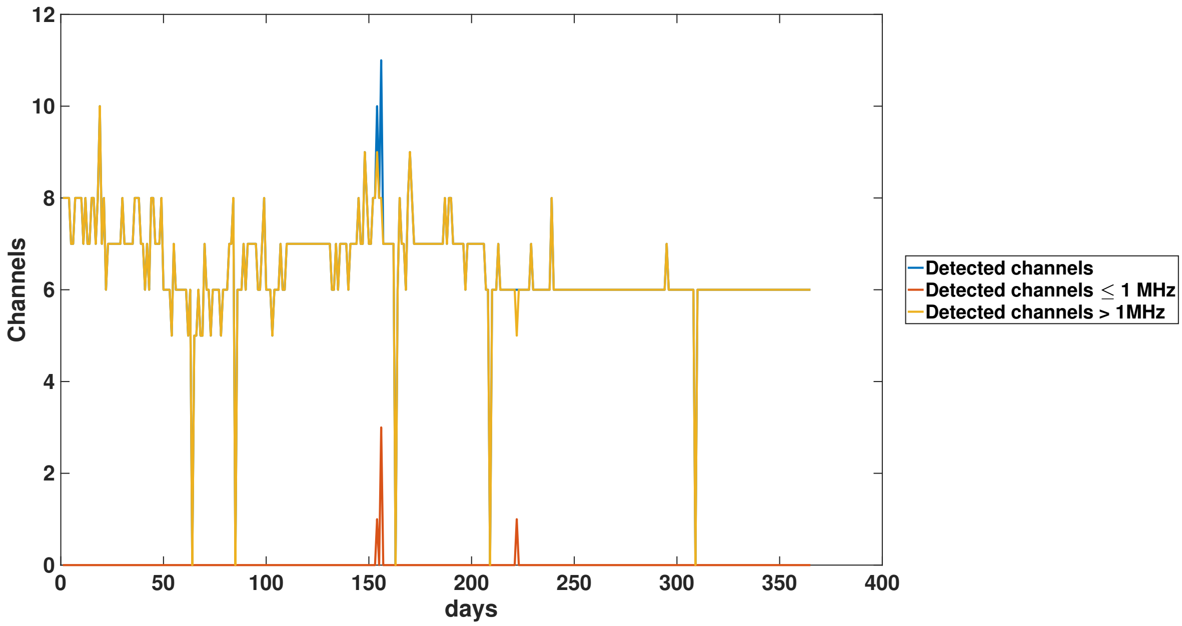

Figure 8.3: Detected channels in Heerhugowaard for Band 1 in 2017

The channel detection algorithm is not able to detect any uplink channel licensed to a network operator. This could be explained by the fact that mobile devices are sending with much less power than the base station of the mobile operator. Therefore, it is likely that the uplink signals are below the noise floor of the measurement setup and can therefore not be detected.

8.2 Statistics of detected channels

Statistics are computed for all detected channels at all different measurement locations. This results in a large amount of data. Therefore, only the data for band 20 at location 1 is plotted.

8.2.1 Mean

Fig. 8.4 shows the daily mean (6.3) for all detected channels at band 20. For channels 2 and 3, the daily mean is relatively constant over 1 year (2017). On the other hand, for channel 1 the mean is steadily increasing over the year. The difference between the maximum computed mean and the minimum computed mean for channel 1 is 10dB, where this difference for channel 2 and 3 is only 3dB. Especially at the end of the year there is a large increase of the mean, this is probably the result of a new nearby base station that is taken into use.

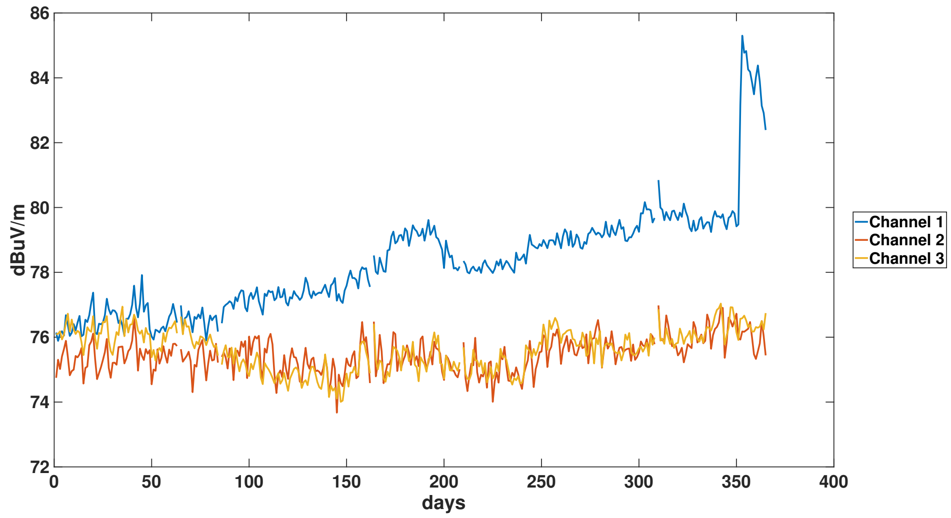

Figure 8.4: Mean for multiple found channels at location 1 for Band 20 in 2017

8.2.2 Median

Besides the mean, the median is computed. The result is plotted in figure 8.5. Also the median shows a increase for channel 1 over the year. The difference between the maximum computed median and the minimum computed median for channel 1 is 8dB, where this difference for channel 2 and 3 is only 3dB.

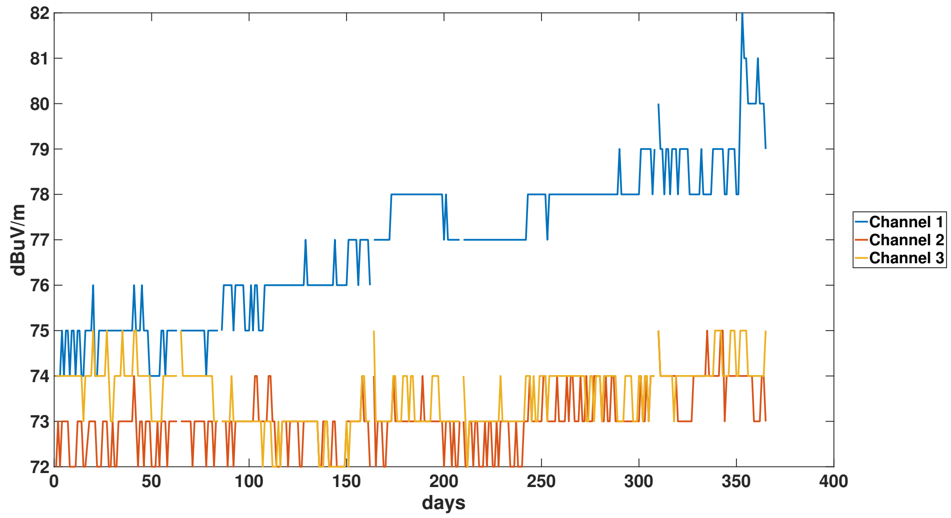

Figure 8.5: Median for multiple found channels at location 1 for Band 20 in 2017

8.2.3 Standard Deviation

The standard deviation is computed for these channels and plotted in figure8.6. Interesting to see, is that the standard deviation of channel 2 and 3 follows the same trend over a year, where channel 1 follows completely different trend.

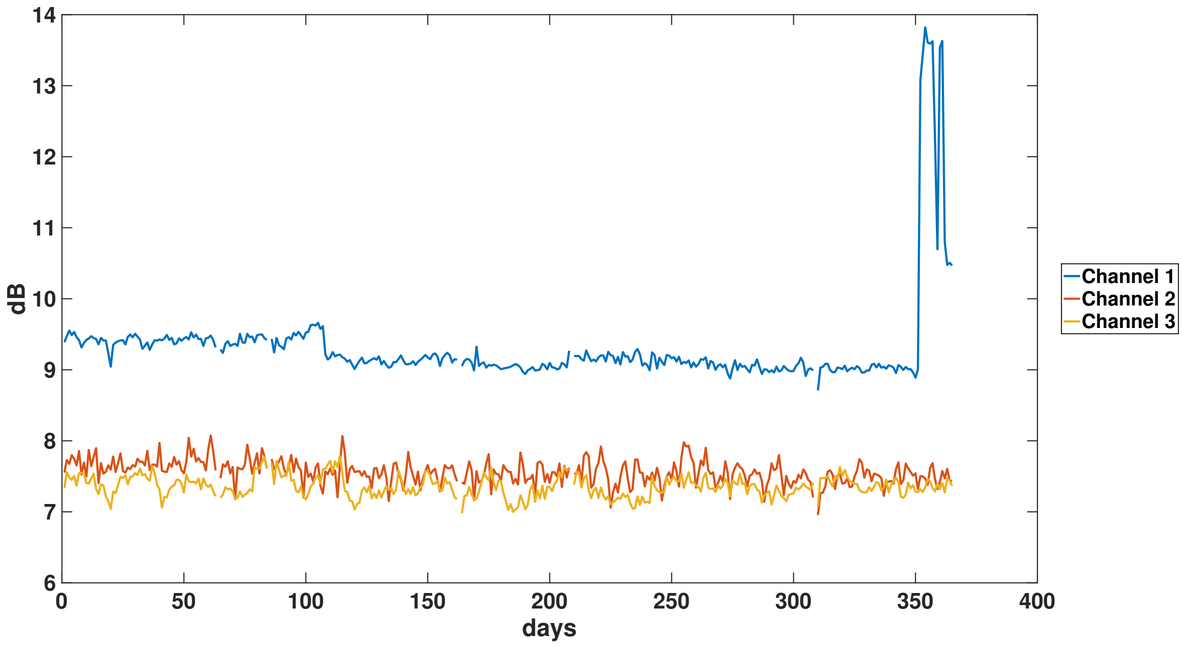

Figure 8.6: Standard deviation for multiple channels at location 1 for Band 20 in 2017

Minimum and Maximum

The minimum is plotted in figure 8.7 and the maximum is plotted in figure 8.8. Combing the maximum and minimum information with the information about the standard deviation, mean and the median, It is likely that the operators base station from channel 1 is closer to the measurement location than the base stations for channels 2 and 3. This was indeed the case, the base station of the operator that is controlling channel 1 is around 100 meter away from the measurement location. The operators of channel 2 and 3 have their equipment in the same transmission mast, which is around 500 meters away from the measurement location.

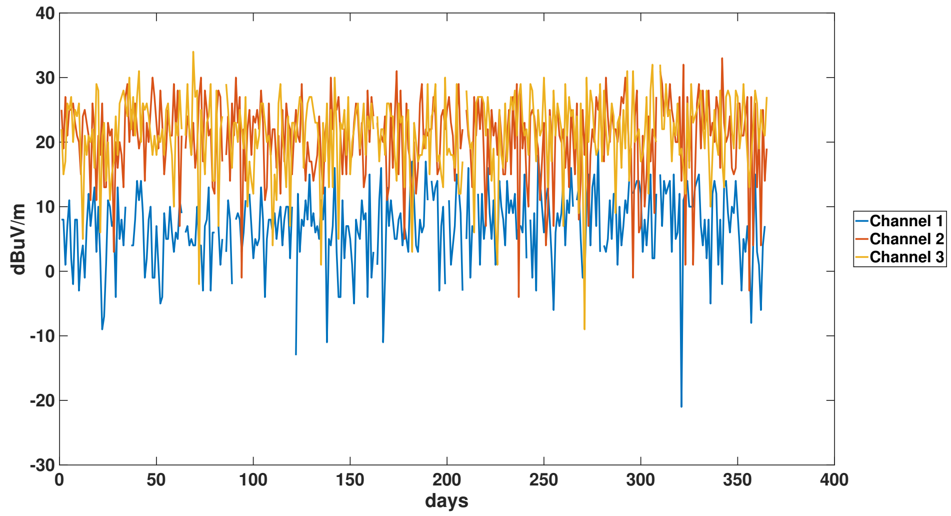

Figure 8.7: Minimum for multiple channels at location 1 for Band 20 in 2017

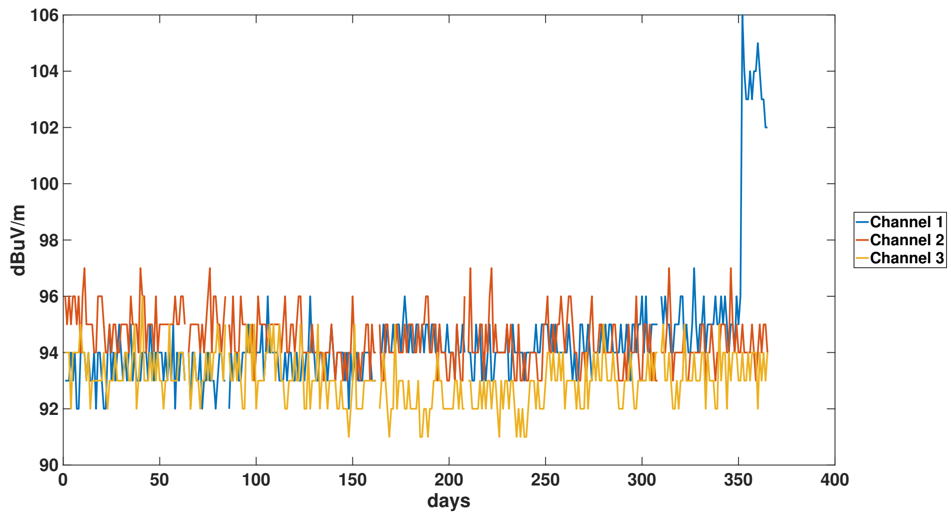

Figure 8.8: Maximum for multiple channels at location 1 for Band 20 in 2017

8.2.4 Cumulative distribution

The CDF for channel 1 is in figure 8.9 and for channel 2 in figure 8.10. All lines in these plots are corresponding to 1 day of data. In respectively in Fig. 8.11 and Fig. 8.12 is a time axes added to the plot. Especially in the plot of channel 1, it is clear that at the end of the year, the received signal strength is increased. The CDF for channel 2 is relative constant over 1 year of measurement data. These plots of the CDF can be used as a graphical tool to visually analyse the channel deviation over a period of time.

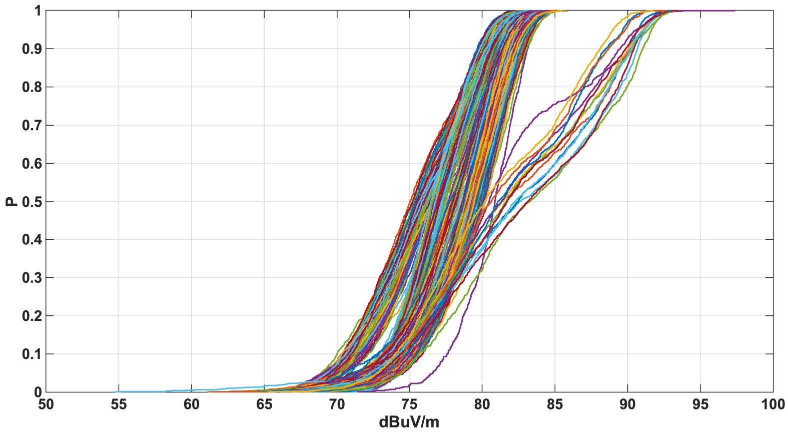

Figure 8.9: CDF for channel 1 at location 1 for Band 20 in 2017

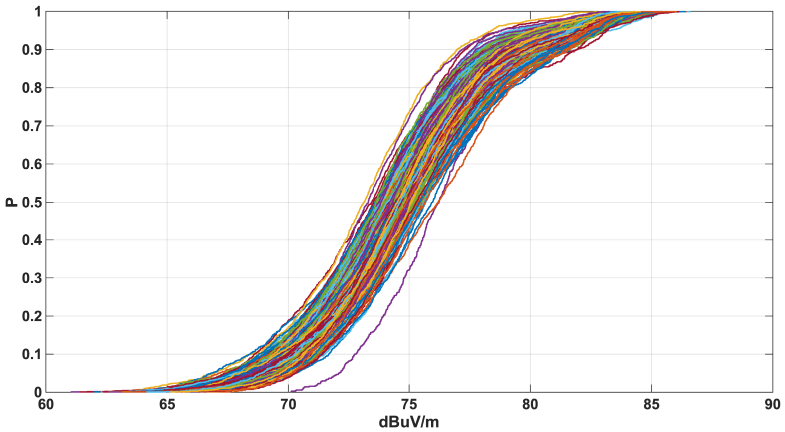

Figure 8.10: CDF for channel 2 at location 1 for Band 20 in 2017

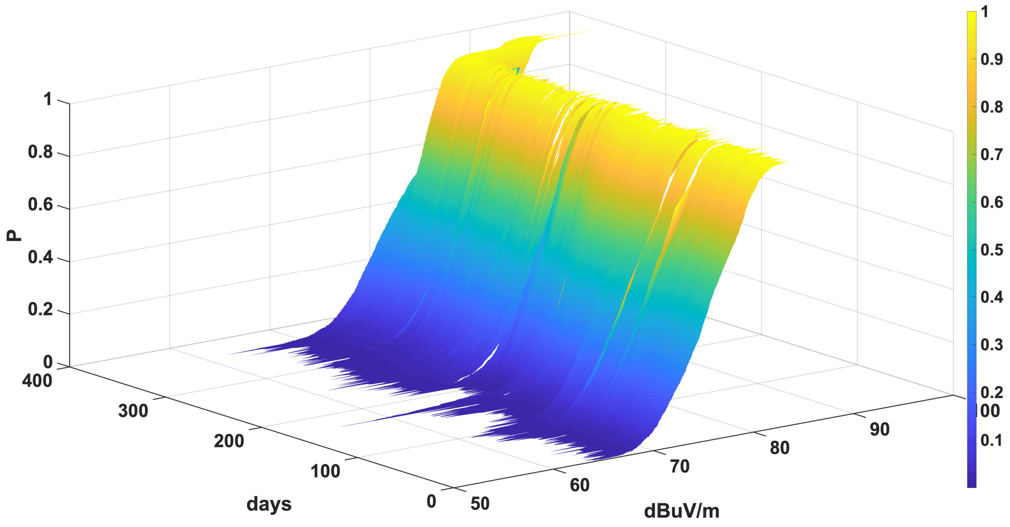

Figure 8.11: CDF for channel 1 at location 1 for Band 20 in 2017

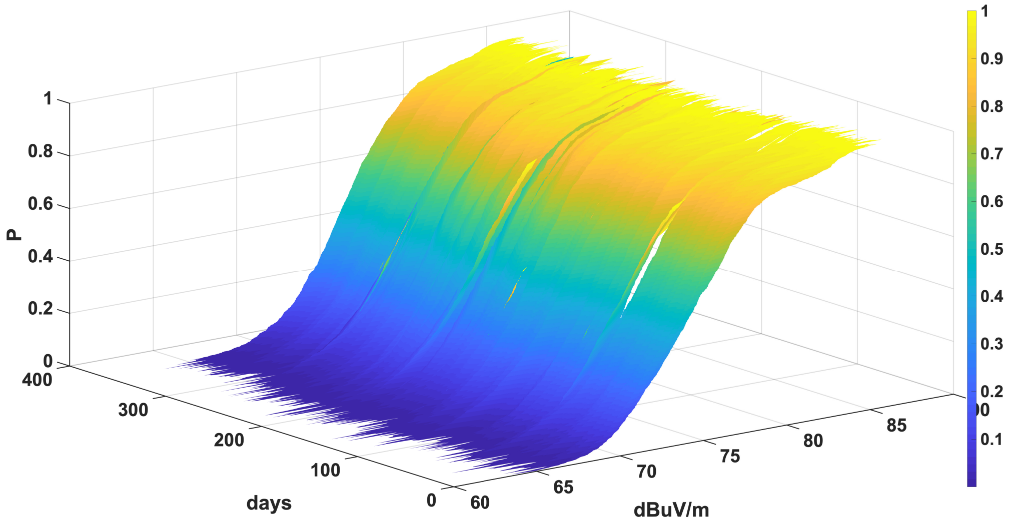

Figure 8.12: CDF for channel 2 at location 1 for Band 20 in 2017

8.3 Outlandish signal level detection

As explained in section 6.2.1, the derivative of the computed statistics is used to compare channels between different moment in time. The derivative is computed for different values of Δt. The measured values for the day/night comparison are taken between 5:00 and 17:00 CET. In Table 8.2 is the absolute average difference between 2 moments in time from the mean. This shows that the average daily deviation for these channels is between 0.25 dB and 0.37 dB under normal circumstances. It also shows that the average difference between day and night is between 5.2 dB and 6.5 dB. This difference is also visible in Fig. 8.13.

Table 8.2: Average absolute difference of the mean of multiple channels in band 20 in 2017

	1 minute	5:00 and 17:00	1 day
Channel 1	3.7 dB	5.2 dB	0.25 dB
Channel 2	4.1 dB	6.4 dB	0.37 dB
Channel 3	5.0 dB	6.5 dB	0.27 dB

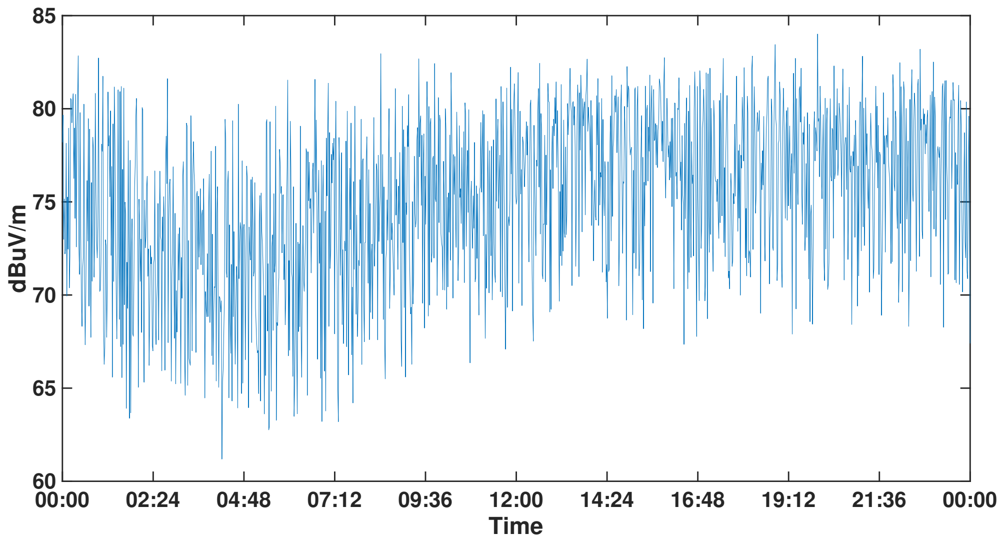

Figure 8.13: Measured values at 2 January 2017 for channel 3 in band 20 at location 1

Besides the day night variation, the detection process is also able to detect significant difference over time. As shown in Figures 8.4, 8.5, 8.6 and 8.8. A significant change around the end of the year 2017 in channel 1 is clearly visible. The derivative of the computed mean, median, standard deviation and the maximum show a significant peak between day 351 and day 352. The computed derivative in displayed in Fig. 8.14. This is an indication for an outlandish signal or a significant change in the network of a mobile network operator. The same picture can be made for the other statistics.

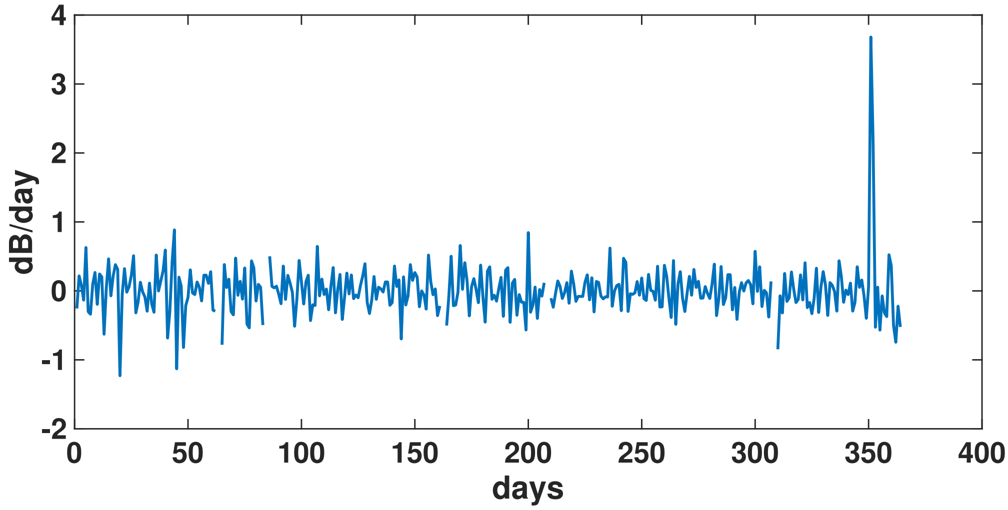

Figure 8.14: Derivative of the computed daily mean in 2017 for channel 1 in band 20 at location 1

8.4 Detecting events

Outage or maintenance

The algorithm is able to detect multiple outages of different base stations. For example, on a day in 2017, an outage of multiple minutes was detected. In Fig. 8.15 the measured strength of multiple channels displayed. Please note that due to the privacy of a network operator, the frequency and location information is removed from this figure. In Fig. 8.15 an outage around 1:00 is clearly visible. The algorithm found this outage by looking at the mean of individually channels, as described in section 6.3.1. However, by visually looking at the individual 3D CDF plot for the different channels. The outage is clearly visible and can be easily identified by a human being. This 3D CDF is displayed in Fig. 8.16.

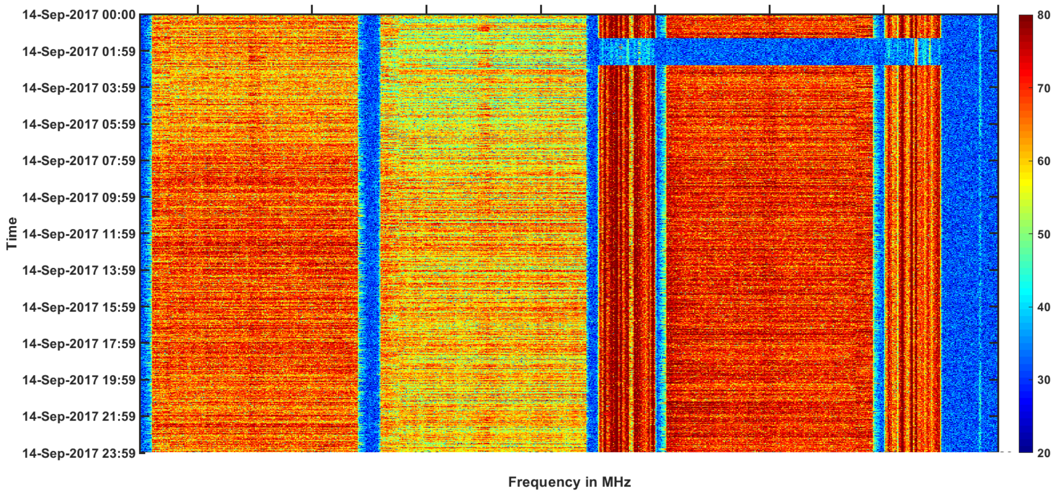

Figure 8.15: Measured signal strength at 14 September 2017

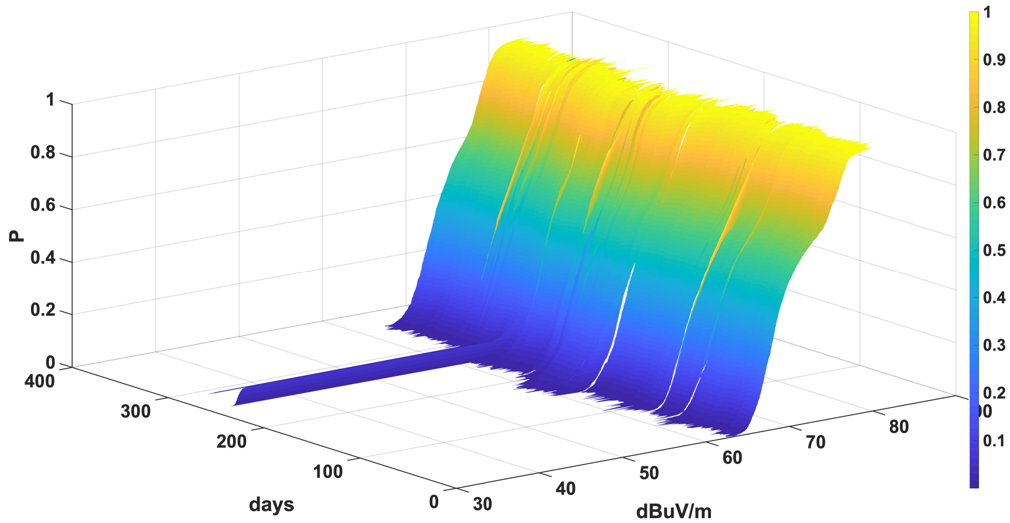

Figure 8.16: 3D CDF for a channel with an outage

Power saving

Especially at higher frequencies, power saving events are detected by the algorithm. This can be explained due to the fact that in general that the path loss between transmitter and receiver will increase with frequency. Therefore, base stations with higher frequencies can be used to increase the capacity of a network operator locally, where a base station that is using lower frequencies, can be used to increase the coverage in an area. At night when a network operator does not need a high capacity, an operator can turn off the base that is using the higher frequencies, while remaining the coverage with base station that is using the lower frequencies. A power saving event is detected using the method described in section 6.3.2. Power saving events can be visualised in a CDF. In Fig. 8.17 the CDF's of 1 year of a channel where saving was detected is shown. Here the threshold found with the MED-FCME algorithm was 52.4dB. At that 52.4dB the corresponding P value is around 0.25, which corresponds to 6 hours of power saving a day. Looking into the raw measurement data, this was indeed the case. The base station was off between 0:00 and 6:00 daily.

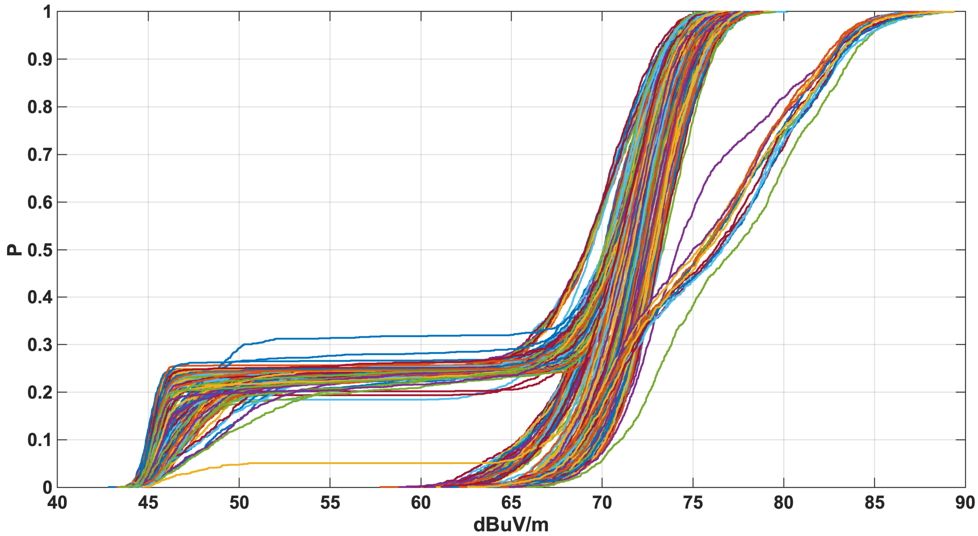

Figure 8.17: CDF for a channel were power saving is detected

Change in technology

A change in technology that is detected is the introduction of NB-IoT. NB-IoT can be deployed in the guard band of a normal LTE channel [11], it will therefore extend the used channel width of a LTE channel with 180 kHz. This extension of the channel width can be detected. In Fig. 8.18 is the detected channel width of a LTE channel displayed. The frequency step used in this measurement was 50 kHz and the detected difference was between 150 en 250 kHz.

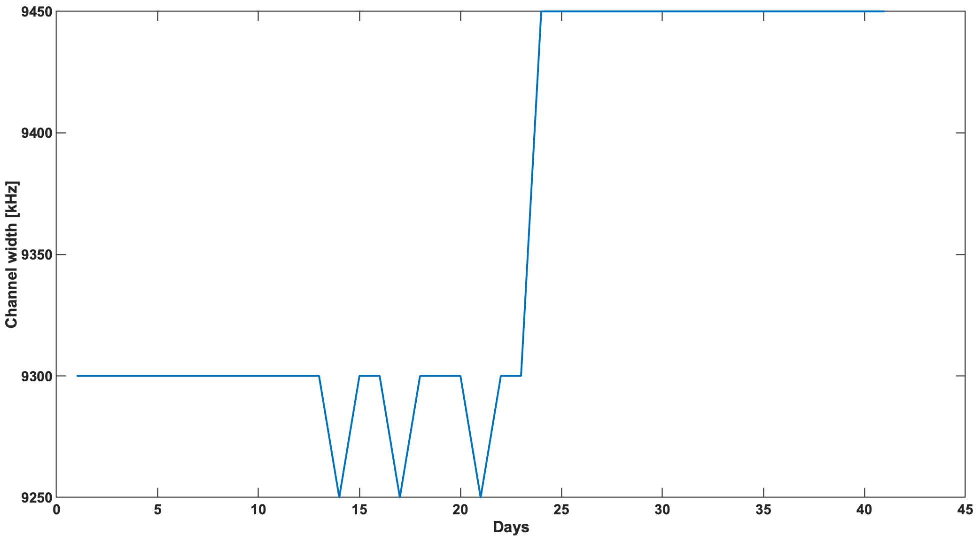

Figure 8.18: Detected channel width of a LTE channel

Interference

As showed in Fig., 8.2 there are channels with a detected channel width smaller than 1MHz. As described in section 6.3.4 channels with a channel width smaller than 1 MHz are not likely to be a channel used by a network operator and are identified as interference. One example of an interference is displayed in Fig. 8.19, this interference is detected by the algorithm. The same holds for the very narrow band signal at the right of figure 8.15.

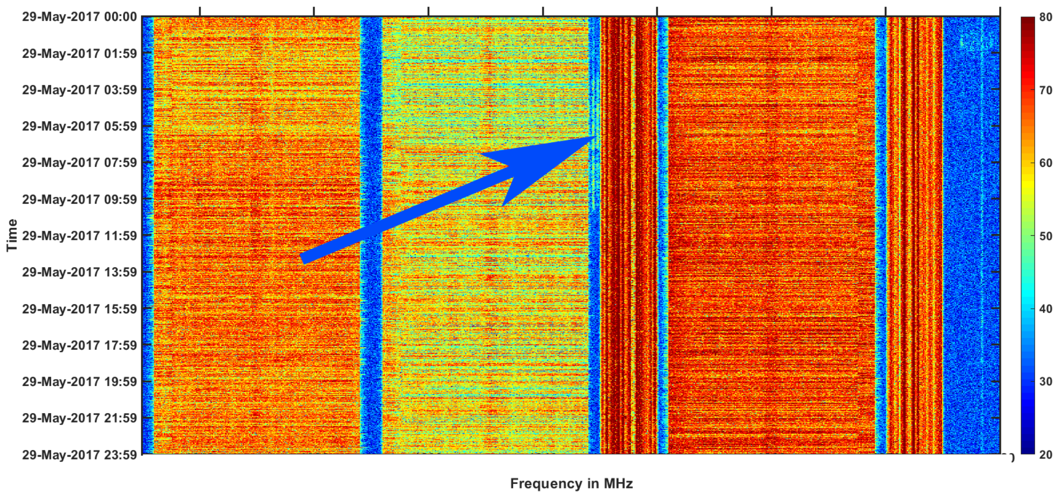

Figure 8.19: Detected narrow band channel

Chapter 9

Conclusions and recommendations

After carefully evaluating all results from the research, there are some conclusions and recommendations. First, the conclusions are presented. The recommendations for future work are elaborated in the second section.

9.1 Conclusions

The main research goal was to propose a methodology to automatically detect outlandish signals in the frequency bands reserved to cellular networks. Therefore, the following four analyses were done.

1. Statistical characterisation and modelling of the noise floor
2. Statistical characterisation and modelling of the received signal fluctuations over time for different detected channels.
3. Statistical characterisation and modelling of the following events: outage, power saving, change in used telecommunication standards and interference.
4. Automatic detection of the signal fluctuations and events.

In this report it is shown that the noise floor for cellular bands measured by the Radiocommunications Agency of the Netherlands can be estimated with the MED-FCME algorithm, if at least 5% of the measured values contain only noise. If the noise floor is estimated for every day in 2017, the maximum difference between the estimated noise floor levels is 3 dB.

Based on this noise floor threshold, it is possible to detect channels in the downlink to calculate the statistics for detecting outlandish signals as long as the guard band between two channels is relatively large compared to the measurement bandwidth. If the guard band between multiple channels is too small compared to the measurement bandwidth, the detection process detects multiple channels as one channel. It is not possible to detect any of the uplink channels.

Multiple events can be detected by the detection algorithm. Outages, base station maintenance and power savings at night can be detected. Furthermore, the algorithm is able to detect a change in the telecommunication standard used by a network operator. Additionally, multiple types of interference are detected, as well as the significant difference between the measured values at night compared to the measured values in the afternoon.

It can thus be concluded that it is possible to automatically detect outlandish signals with the current measurement setup of the Radiocommunications Agency of the Netherlands, as proved in this research. However, there are some improvement suggestions that may result in better detection of outlandish signals. These are stated in the following section.

9.2 Recommendations

In total, four recommendations have been made for future research. Two about the measurement data and two about the detection algorithm.

9.2.1 Measurement data

The first recommendation is related to the used measurement bandwidth in the measurements. The measurement bandwidth is not the same for all the measurements that are taken. This results in a different detection correctness for different frequency bands. For example, it is not possible to detect a NB-IoT signal between 791 and 821MHz, since the used measurement bandwidth is higher than the bandwidth used by the NB-IoT signal. The recommendation therefore is to find the minimal measurement bandwidth required to detect all telecommunication standards used by a network operator.

The second recommendation is about the detection of an uplink signal. With the current measurement data, it is not possible to detect an uplink signal. This detection can be improved by decreasing the distance between the measurement antenna and the mobile devices. Decreasing the distance between the mobile device and the measurement antenna will decrease the path loss between them. This will result in a higher received signal by the measurement antenna. Therefore, it is easier for the detection algorithm to detect the signals sent from a mobile device. This decrease in distance can be established, for example, by placing a measurement antenna inside crowded places, like the Amsterdam Arena or large train stations.

9.2.2 Detection algorithm

Further work can also be done by improving the detection algorithm. The current channel detection algorithm has difficulties with detecting channels that are close to each other. This detection can be potentially improved by looking at local minimum and local maximum and the difference between these. Another way can be by finding the correlation coefficient between different frequencies components. Frequency components that have a strong correlation are likely correspond to the same channel.

The last recommendation is about the event detection. With the current detection process it is possible to detect significant changes in the spectrum, however it is not always possible to relate these significant changes to events that are happening in the real world. It would be a relevant improvement if the algorithm is able to better classify multiple events by intertwining measurement data with the known events that have happened.

Bibliography

[1] R. Schiphorst and C. H. Slump, "Evaluation of spectrum occupancy in amsterdam using mobile monitoring vehicles," in *2010 IEEE 71st Vehicular Technology Conference*, May 2010, pp. 1–5.

[2] J. R. aylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. Sausalito, CA: University Science Books, 1997.

[3] M. Lopez-Benitez, A. Umbert, and F. Casadevall, "Evaluation of spectrum occupancy in spain for cognitive radio applications," in *VTC Spring 2009 - IEEE 69th Vehicular Technology Conference*, April 2009, pp. 1–5.

[4] ITU-T, "Spectrum occupancy measurements and evaluation," International Telecommunication Union, Recommendation SM.2256-1, 3 2016.

[5] J.-W. Lee, J.-H. Kim, H.-J. Oh, and S.-H. Hwang, "Energy detector using adaptive-fixed thresholds in cognitive radio systems," in *2008 14th Asia-Pacific Conference on Communications*, Oct 2008, pp. 1–4.

[6] A. E. Omer, "Review of spectrum sensing techniques in cognitive radio networks," in *2015 International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE)*, Sep. 2015, pp. 439–446.

[7] H. Saarnisaari, P. Henttu, and M. Juntti, "Iterative multidimensional impulse detectors for communications based on the classical diagnostic methods," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 395–398, March 2005.

[8] K. U. J.-P. M. Janne J. LEHTOMKI, Risto VUOHTONIEMI, "Energy detection based estimation of channel occupancy rate with adaptive noise estimation," *IEICE Transactions on Communications*, vol. E95.B, no. 4, pp. 1076–1084, 2012.

[9] J. Vartiainen, "Concentrated signal extraction using consecutive mean excision algorithms," Ph.D. dissertation, University of Oulu, 2010.

[10] 3GPP, “Standards for lte,” 3GPP, specifications Release15, Jul. 2018.

[11] 3GPP, “Standards for the internet-of-things,” 3GPP, Presentation 2016-11, Nov. 2016.

Appendix A

Data preparation in MATLAB

All the measurement data is saved on a file server. The directory structure of this file server is displayed in figure A.1.

```
2018
├─01
│ ├─Location-1
│ │ ├─Measurement-file-1.mat
│ │ ├─Measurement-file-2.mat
│ │ ├─Measurement-file-3.mat
│ │ ├─...
│ │ └─Measurement-file-N.mat
│ ├─Location-2
│ ├─Location-3
│ ├─...
│ └─Location-N
├─02
├─03
├─...
└─12
```

Figure A.1: Data structure on the file server

The measurement files located in the measurement-location-directory (Location-1, Location-2, etc..) are the measurement files that contain information about the whole spectrum that is measured by the radio communication agency. Since the spectrum needed for this research should be allocated for cellular networks, only a small part of the spectrum measured by the agency is needed. This needed part of the spectrum is taken out of the entire data set and stored separately. The searching for the right measurement files is done with a MATLAB script in a few steps. Step 1 starts with getting input from the end user. The MATLAB script will get the following input from the end user:

1. The year.
2. Measurement location numbers, as found in table 3.1.
3. Cellular network band numbers, as found in table 3.3.

Second, the script finds the directory where the data for the input year is located. In this directory there are new directories for every month. All the stored month-directories are saved in a variable. For every stored month-directory a new iterative process will start. This iterative process will find the directories that are located in the month-directory. For every measurement location there is a directory in this month-directory. The script checks if the name of this measurement-location-directory corresponds to the measurement number given by the end user. If the name matches, the script starts a new iterative process for every match. This process looks at all the files inside the measurement-location-directory and checks if the measurement file contains data for bands given by the end user. If that is the case, it will take this data and save it in a separate file, otherwise it continues with the next measurement file. The saved data is ordered, per day, per band and per location band. The whole process is displayed in figure A.2.

When the script is done, all the data must be saved in a logical way. This process is visualised in figure A.3 and described in the following paragraph. To do that, all the data is saved in a MATLAB 1x1 struct. A struct is a datatype that MATLAB uses to store a list of variables. In this data struct is a new data struct for every day. In these day data structs are new stucts, one for every band. These band structs hold the basic details about the saved band, like uplink and downlink frequencies. Besides the basic details there is again a new struct, this one is for the locations. In the location struct are new structs, one for every location. Since the up- and down-link are analysed separately they are both stored in a separate struct. In the up- and down-link structs is the actual data. Now there is 1 MATLAB file that holds the data about the spectrum allocated for the given cellular networks bands for 1 given year at a given measurement location.

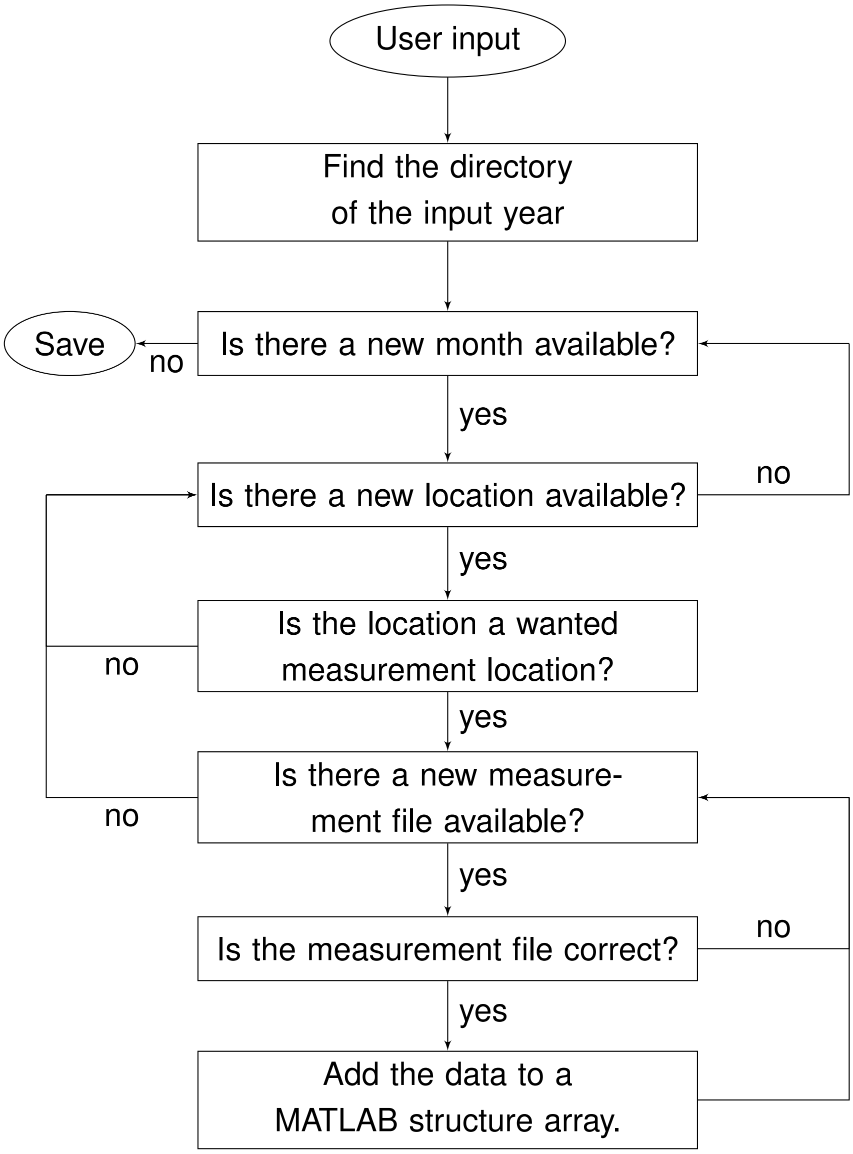

Figure A.2: Flowdiagram of the MATLAB scripts searching for the wanted data

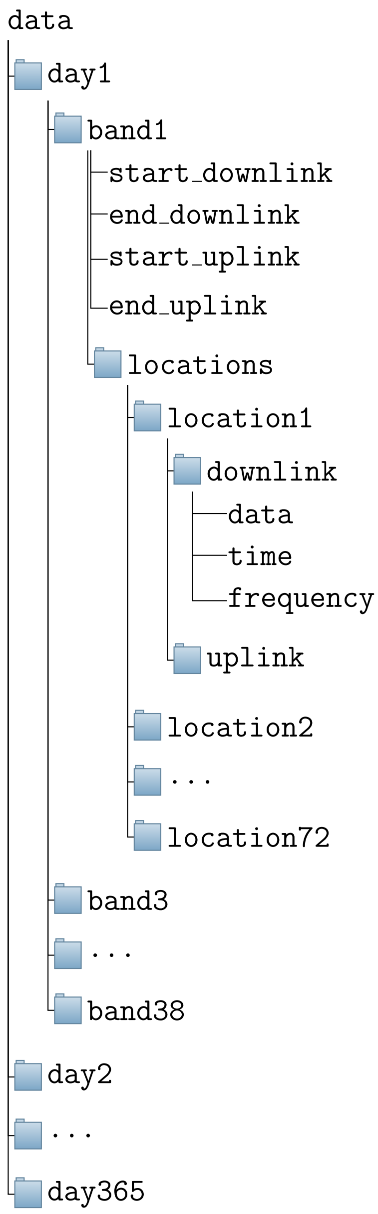

Figure A.3: Data struct for cellular data of 1 year