Flexible and Spectrum Aware Radio Access through Measurements and Modelling in Cognitive Radio Systems

FARAMIR

Document Number D3.3

Spectrum Use Models: Methodology and Measurement Results

Contractual date of delivery to the CEC:	31.12.2011
Actual date of delivery to the CEC:	31.12.2011
Project Number and Acronym:	248351 - FARAMIR
Editor:	Janne Riihijärvi (RWTH)
Authors:	Janne Riihijärvi (RWTH), Alexandros Palaios (RWTH), Jad Nasreddine (RWTH), Petri Mähönen (RWTH), Zaheer Abbas (RWTH), Mihajlo Pavloski (UKIM), Vladimir Atanasovski (UKIM), Liljana Gavrilovska (UKIM), Fernando Casadevall (UPC), Miguel López-Benítez (UPC), Antoine Dejonghe (IMEC), Berna Sayrac (FT), Sana Ben Jemaa (FT), Pascal Cordier (FT), Peter Scheele (BNetzA)
Participants:	RWTH, UKIM, IMEC, UPC, FT, TRL, BNetzA
Workpackage:	WP3
Security:	PU
Nature:	R
Version:	1.0
Total number of pages:	101

Abstract:

In this document we give an overview of the spectrum use measurements carried out in FARAMIR together with a summary of main findings and results. We also discuss the application of the measurement results in modeling of spectrum use in time, frequency and spatial domains. The data sets all the results have been derived from will be released towards the research community during the summer of 2012, enabling interested parties to verify and replicate the analysis carried out.

Keywords: Spectrum use measurements, two days of spectrum, spectrum use models

Document Revision History

Version	*Date*	*Author*	*Summary of main changes*
0.1	10.11.2011	Janne Riihijärvi (RWTH)	Document structure for circulation
0.3	17.12.2011	Janne Riihijärvi (RWTH)	Added partner contributions
0.5	21.12.2011	Janne Riihijärvi (RWTH)	Results from London and Two Days of Spectrum campaigns added
0.8	29.12.2011	Janne Riihijärvi (RWTH)	Integrated version finalized
1.0	30.12.2011	Janne Riihijärvi (RWTH)	Final version for European Commission
1.0f	31.12.2011	Petri Mähönen (RWTH)	Coordinator review and approval

Table of Contents

1 Introduction

The main objective of the FARAMIR project is to carry out research and to develop techniques for increasing the radio environmental and spectral awareness of future wireless systems. Improving the understanding on how radio environment is used at present by the existing technologies, how this usage can be characterized and modeled, and how the arising models can be used as a foundation for new technologies for improving spectral efficiency is a core part of this work. In order to accomplish this, the project has carried out, and will continue to do so, measurement campaigns on spectrum use in different European regions. The data gathered from these measurements forms a solid scientific basis for the characterization and modeling work, and will also form a valuable asset for the research community at large. The project is planning to release the gathered measurement data sets together with an extensive documentation on how the data was collected by the end of the project lifetime, namely summer of 2012.

In this document we provide an overview and status report on the measurement campaigns carried out in the project, and also discuss results and conclusions drawn from those data sets. We give both statistical characterizations of the raw data sets obtained from the measurements, especially focusing on large collaborative spectrum measurement campaigns carried out, and also discuss the development of spectrum use models based on the gathered data. For example, we show how spatial spectrum use models can be used for coverage prediction for large-scale wireless networks, and which kinds of challenges arise from the indoor propagation environment for radio resource management problems of femtocell networks. We believe these examples are of independent interest to the research community, and also highlight relevant and extremely challenging research problems to tackle in the future. The statistical analysis of the spectrum use in different size and comparison of thereof is still on-going beyond what is reported in this deliverable. Also due to the submission and to-be-submitted status of some papers, some results have not been reported in this public deliverable, but will be eventually made public when the peer-review process and our analysis situation allow it.

The rest of the document is structured as follows. In Section 2 we give an overview of the measurement campaigns conducted, as well as the equipment used by the different partners to carry out spectrum measurements. In Section 3 the different key modeling frameworks for spectrum use will be discussed, together with the main statistical characterizations for spectrum used related to these modeling frameworks. Example results from the European Days of Spectrum will be then discussed in Section 4, and results from further collaborative spectrum measurement campaigns will be given in Section 5. Finally, additional results from the work carried out at individual partner sites and smaller collaborations will be given in Section 6, before drawing the conclusions in Section 7.

2 Overview of Spectrum Measurement Campaigns

In this section we provide an overview of the different spectrum measurement campaigns carried out in the project. We begin by describing the major collaborative measurement efforts carried out by multiple partners in synchronized fashion, and then continue by briefly discussing additional measurements carried out at individual partner sites.

2.1 European Days of Spectrum

The largest collaborative measurement efforts carried out in FARAMIR have been the two European Days of Spectrum. These consisted of two measurement campaigns, each two days long, during which all the partners active in the spectrum measurements work carried out synchronized measurements of spectrum use at the partner sites and also in selected additional locations. First of these campaigns took place in 20-21 October, 2010, whereas the latter was organized 13-14 October 2011. Figure 1 illustrates the locations of the measurement sites. These included Aachen, Hannover, Krefeld and Konstanz in Germany, Maastricht in the Netherlands, Leuven in Belgium, Paris in France, Barcelona in Spain, Skopje in Republic of Macedonia, and Bristol in the United Kingdom.

Figure 1: Measurement locations for the two European Days of Spectrum campaigns.

The objectives of the two days campaigns were to gather measurement data on spectrum use over a wide band (from FM frequencies all the way up to 3 GHz), in a variety of different countries and environments. For the first of the two days whole 3 GHz bandwidth was swept continuously, whereas for the second day partners with slower spectrum analyzers focused on the cellular and ISM bands. Depending on the availability of the facilities at the different measurement sites, at some locations measurements were also continued over the weekend following the designated days of the measurements. In order to obtain comparable results, the measurement settings were carefully harmonized across all sites, and processing of the measurement data was carried out centrally with the assistance from the partners involved. Common data exchange format was also developed to facilitate this approach, and also the sharing of the data later one.

We describe the joint measurement settings used for the two days campaigns in Subsection 2.4 below, and discuss some of the key results and findings from measurements in Section 4.

2.2 Joint FARAMIR-ACROPOLIS Measurements in London

The two days campaigns provided an extensive collection of snapshots of spectrum use across Europe over the course of several days, but did not by design focus on the more detailed characterization of spectrum use *within* a given environment. There is very limited number of such detailed spatio-temporal measurements that have been carried out and reported on in the literature, with the earlier work carried out at RWTH Aachen University still being amongst the most extensive ones (see, for example, [22], [30] and [31]) documented in the research literature and with measurement data made openly available. In particular, to the best of our knowledge no such data set exists for any of the major population centres.

In order to fill in this gap in the understanding of spectrum use and availability of data, the FARAMIR project arranged in collaboration with the ACROPOLIS European Network of Excellence a week long measurement campaign in London. The campaign focused on sampling of spectrum use with the “blue box” mobile measurement platforms of RWTH over a variety of different regions within the greater London area, ranging from the shopping streets and touristic areas around Oxford Circus and Trafalgar square, to suburban locations at the edge of the city. The main focus in the campaign was on frequency bands used for cellular systems, and the 2.4 GHz ISM band was also covered. This focus was chosen first because of the relatively static nature of the spectrum use in other bands, and also in order to obtain higher temporal resolution for the measurements through the reduction of the number of sampled points in the frequency domain. In total over 150 individual locations were covered for periods ranging from tens of minutes to several hours, with an additional spectrum analyzer placed for an entire week on the roof of one of the buildings of King’s College London, the ACROPOLIS partner providing local organization and support for the measurements. Figure 2 shows as overview of the measurement locations within London.

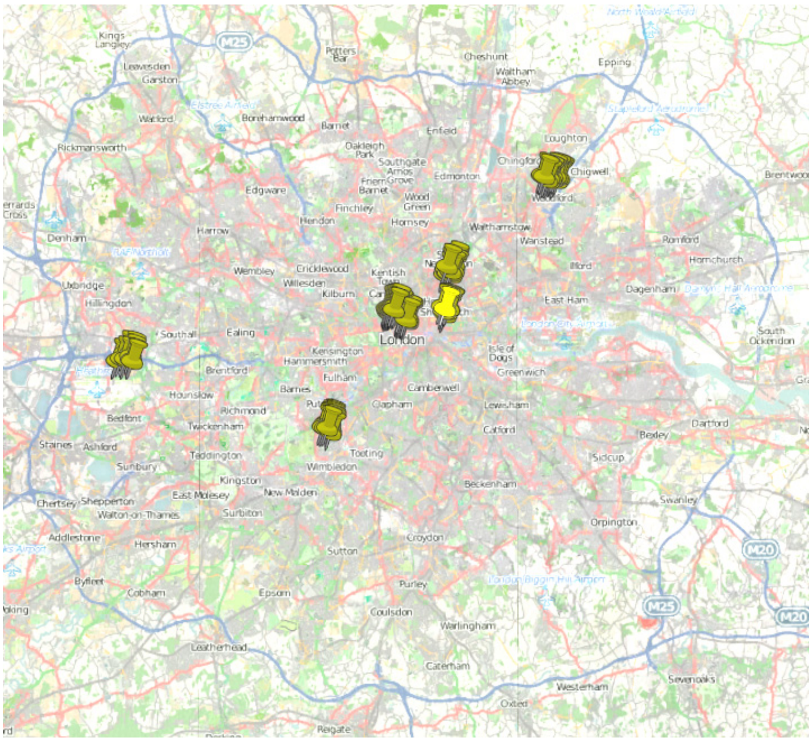


Figure 2: Measurement locations in London.

2.3 Additional Measurements Performed

In addition to the major collaborative campaigns discussed above, the project partners have carried out a large number of additional spectrum use measurements, and also measurements going beyond classical "passive" approaches, including measurements and modeling of the shadowing and propagation environment. Some of these have also been collaborative in nature. For example, FT and RWTH collaborated on organizing a major campaign on spatial spectrum use measurements in Paris during October 2010, and IMEC and RWTH have initiated a collaboration on comparing the spectrum measurement results obtained using high fidelity spectrum analyzers against those obtained with the FARAMIR spectrum sensing engine based on the Scaldio chip developed at IMEC. Also numerous individual measurement s have been carried out by the partners. For example, RWTH has initiated both spatial and temporal spectrum use measurements, as well as dedicated "active" measurements towards characterizing shadowing and propagation environments. UPC has carried out long term spectrum use measurements at several locations around Barcelona region, and also UKIM and TREL have worked extensively on spectrum measurements at and around their respective sites. We shall give selected examples on results from these additional spectrum measurement campaigns in Section 6 below.

It should be noted that FARAMIR has not carried out DVB-T specific measurements, including measurements based on DVB-T decoding and quality assessment. There are several reasons for this. First, the liaison between FARAMIR and QUASAR resulted in the understanding that the latter project is actively working on such measurements, and agreement to exchange data between the projects was achieved, making additional measurements on the FARAMIR side somewhat redundant (there are also additional projects such as COGEU in Europe carrying out DVB-T specific spectrum use measurements, albeit in more limited fashion especially geographically). There are also further scientific reasons for this decision. First, our measurements already inherently cover in most cases the whole spectrum at least until 3 GHz, so the DVB-T bands are naturally included. Further, the transmissions of DVB-T systems are very well understood in terms of tower locations and temporal behavior, so all the modeling questions of interest rapidly converge to issues related to propagation issues, role of interference, and to questions on demodulation quality. All of these require specialized measurement approaches, including the capability of demodulating DVB-T signals, leading into highly specific research questions. Given that there are other projects active in this space, and the limited resources FARAMIR has for measurement work, the project decided to focus on applications of data analysis techniques on DVB-T data sets gathered through liaisons rather than direct project measurement work.

2.4 Equipment and Setups Used

In this subsection we outline briefly the common settings and data formats utilized in the major joint measurement campaigns, after which an overview of the equipment and measurement setups used at different partner sites will be given.

2.4.1 Common Settings and Data Formats

As with any large measurement endeavor of collaborative nature, the major challenges are the acquisition of comparable measurement results, and the harmonization of the data formats used to ensure easy the problem-free exchange and processing of the data. In order to accomplish this especially for the large two days of spectrum campaigns, RWTH introduced a dedicated measurement coordinator to organize the process. This coordinator collaborated closely with each of the measurement teams, first gathering information on the available equipment, including spectrum analyzers, antennas, and cables, and then working out, based on the research questions that were deemed most important to the consortium, a measurement plan that would result in most comparable results possible. To further assure smooth data gathering and exchange process, trial runs were organized at each of the measurement sites, with data collected for joint analysis to rapidly find any potential sources of errors. The key settings that emerged from this process included the chosen resolution bandwidth (selected to be 100 kHz based on capabilities of the measurement devices, and the desired frequency and time resolutions), sweeping patterns (whether on the second measurement day of the two days campaign to focus on specific bands or continue wideband sweeps depending on the spectrum analyzer performance characteristics), and the detector type (chosen to be an RMS detector since focus on the measurements is on determination of the short term mean power spectral density values). Complete logbooks of the settings used together with the measured cable and amplifier characteristics will be included on the portal through which the measurement data sets will be made available to the research community.

In terms of data representation and exchange, a simple format consisting of a binary data file for storing the raw measurement results together with time stamps and basic diagnostic indicators (for, e.g., detecting overloads), and a text-based metadata file documenting the spectrum analyzer configuration and the arising frequency bins. The structure of the main part of the binary data file is illustrated in Figure 3. The dimensions of the matrix structure shown were computed based on the information stored in the textual metadata file, enabling the data to be read into any widely used computational environment or analysis toolkit.

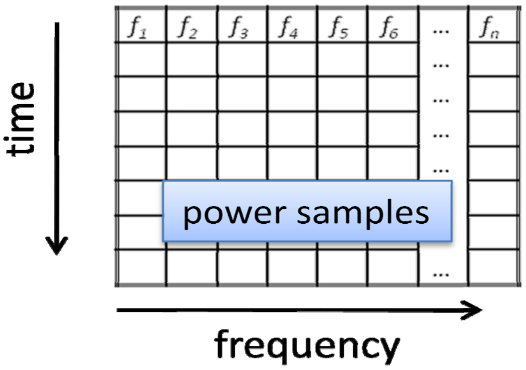


Figure 3: Representation and storage format for the spectrum measurements.

2.4.2 Measurement Equipment and Setup at RWTH

For high precision measurements RWTH has two types of measurement setups, one for static measurements carried out over long time periods, and another four spectrum analyzer setup for mobile measurements, including studies of spatial spectrum occupancy over large spatial regions and its correlation structure. We shall briefly describe both here, although the latter type setup was mainly used in the FARAMIR spectrum measurements.

The architecture of the static measurement setup is shown in Figure 4. It consists of an Agilent E4440 high-end spectrum analyzer connected to a laptop for data collection and control and to three different antennas together covering entire frequency band the spectrum analyzer is usable on. The setup uses antenna of type AOR DA-5000 to investigate the lowest frequency band between 20MHz and 1.52 GHz, a smaller discone antenna of type AOR DA-5000JA for next subband from 1.5GHz up to 3GHz, and, finally, a radom antenna of type Antennentechnik Bad Blankenburg AG KS 1-10 specified up to 10GHz for receiving the frequency range between 3GHz and 6 GHz. All antennas are vertically polarized, have an omnidirectional characteristic in the horizontal plane, and are slightly directive in the vertical plane. Whole measurement setup is enclosed into a weather-proof box that can be deployed in outdoor settings.

While providing high fidelity measurement of spectrum use over time, the setup described above is not particularly well suited for short-term measurement over different regions. The Agilent spectrum analyzer has fairly high power consumption, and the enclosure was developed for stability and high level of protection, making it rather bulky. Due to this separate branch of mobile spectrum analyzer setups, called *"blue boxes"* have been developed. Each blue box is equipped with Rohde & Schwarz FSL 6 Spectrum Analyzer, along with a laptop, a Garmin GPS18 LVC GPS receiver (two blue boxes are equipped with Trimble Mini-T GPS receiver instead) and Absorbent Glass Mat (AGM) lead-acid accumulator. Besides for estimation of the current geographical position of the blue box, GPS receiver is used also for time stamping and synchronization between the blue boxes.

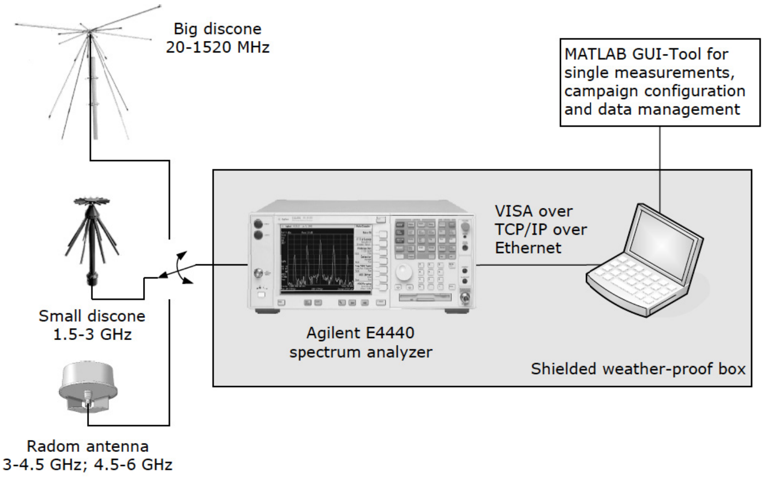


Figure 4: Schematic of the measurement setup for long-term measurements used at RWTH.

Figure 5: Two of the four blue boxes deployed at CeBIT.

The R&S FSL 6 Spectrum Analyzer was chosen due to its portable design. This means lower power consumption (only 30% compared to the Agilent E4440A which was used in permanent measurement setups), vibration resistant construction, and DC input for battery powering. It also includes NiMH accumulator which is, however, only suitable for measurements of around 90 minutes due to its lower capacity.

Each blue box can be equipped with three different antennas, corresponding to the ones used for static spectrum measurements as described above. Figure 5 shows two of the BlueBoxes used during a measurement campaign at the CeBIT.

The key properties of the two types of spectrum analysers used by RWTH are summarized in the Table 1below.

Table 1: Key properties of the spectrum analyzers used
by RWTH in spectrum use and propagation related measurements.

Model	Freq. Range	Freq. Accuracy	Phase noise, 1GHz (10kHz offset)	DANL at 1GHz w/ min.RBW	Third Order Intercept (TOI) at 1 GHz	RBW	Impedance	Power Consumption (no options/all options)
R&S FSL 6	9kHz-6GHz	1E-6	-103 dBc/Hz	-117 dBm	+18dBm	300Hz-10 MHz	50Ω	45W/65 W
Agilent E4440A	3Hz-26.5GHz	1E-7	-128 dBc/Hz	-160 dBm	+20 dBm	3Hz-10MHz	50Ω	260W/450W

In addition to the dedicated spectrum analyzers discussed above, there are numerous software defined radio platforms and other devices capable of lower fidelity measurements used by RWTH also for spectrum measurement purposes. These include

- WARP boards
- USRP and USRP2 devices
- CalRadios
- PCMCIA form factor spectrum analyzers for the ISM-bands

Due to their limited accuracy these have not been employed in the core measurement campaigns described here, but have been used, for example, to study especially the influence of propagation environment on measured spectrum use characteristics in the indoor environment.

2.4.3 Measurement Equipment and Setup at IMEC

For the spectrum use measurements reported on in this deliverable, IMEC also has a number of spectrum analyzers at the disposal of the project team. The key performance characteristics of these are listed in Table 2. During the two days of spectrum measurements, the R&S FSQ 26 spectrum analyzer was used, together with an DA 753 G antenna covering the entire targeted frequency range. The antenna was installed on the roof of the building, and in order to increase sensitivity, partially lost through the higher cable losses correspondingly induced, a low noise amplifier was also included into the measurement setup.

Table 2: Spectrum analyzers at IMEC used for measurements.

Model	Freq. Range	Resolution bandwidth
HP 8565	30 Hz – 50 GHz	1 Hz to 1 MHz in a 1, 3, 10 sequence and 2 MHz (3 MHz at –6 dB)
Rohde & Schwarz FSQ 26	20 Hz – 26.5 GHz	10 Hz to 50 MHz, FFT filter: 1 Hz to 30 kHz
Rohde & Schwarz FSU 67	20 Hz – 67 GHz	10 Hz to 20 MHz in 1/2/3/5 sequence, 50 MHz

2.4.4 Measurement Equipment and Setup at UKIM

UKIM's measurement setup is very similar to those of RWTH and IMEC, and comprises the following hardware components:

- a high precision spectrum analyzer specified up to 6 GHz,
- an omnidirectional antenna specified in the band 25 – 3000 MHz and
- a remote computer for data acquisition (Figure 2.2).

The antenna is connected to the spectrum analyzer through a low-loss cable. Data is gathered to the remote laptop through an Ethernet cable. The setup is placed in an urban outdoor location. The antenna is fixed at an altitude of 10-15 meters above the ground. The placement of the antenna is chosen far enough from any transmitters to avoid possible overloads of the spectrum analyzer. According to the common measurement setup above, measurements were performed during continuous 24 hours in a sweeping manner with a resolution bandwidth (RBW) of 100 kHz.

The signal analyzer chosen for the measurements is of type Anritsu MS2690A, enabling high speed and high accuracy measurements of transmission characteristics for various communication types. Like most of the equipment on the market today it works in sweeping manner and enables real-time monitoring on smaller spectrum bands only. Also, it enables simultaneous analysis in both frequency and time domains. Sensitivity, as the most important parameter of any spectrum measurement device, is -155 dBm (value normalized to 1 Hz) and puts the analyzer in some middle-class category of equipment.

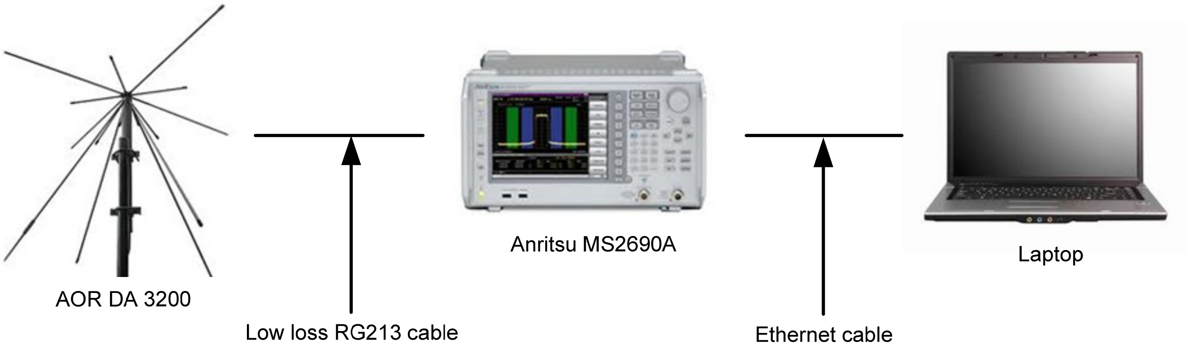


Figure 6: Measurement setup used by UKIM for the Two Days of Spectrum campaign.

2.4.5 Measurement Equipment and Setup at UPC

The measurement equipment relies on a spectrum analyzer setup where different external devices have been added in order to improve the detection capabilities of the system and hence obtain more accurate and reliable results. The design is composed of two broadband discone-type antennas covering the frequency range from 75 to 7075 MHz, a Single-Pole Double-Throw (SPDT) switch to select the desired antenna, several filters to remove undesired overloading (FM) and out-of-band signals, a low-noise pre-amplifier to enhance the overall sensitivity and thus the ability to detect weak signals, and a high performance spectrum analyzer to record the spectral activity. A detailed scheme with all the devices and their main technical characteristics is shown in Figure 7.

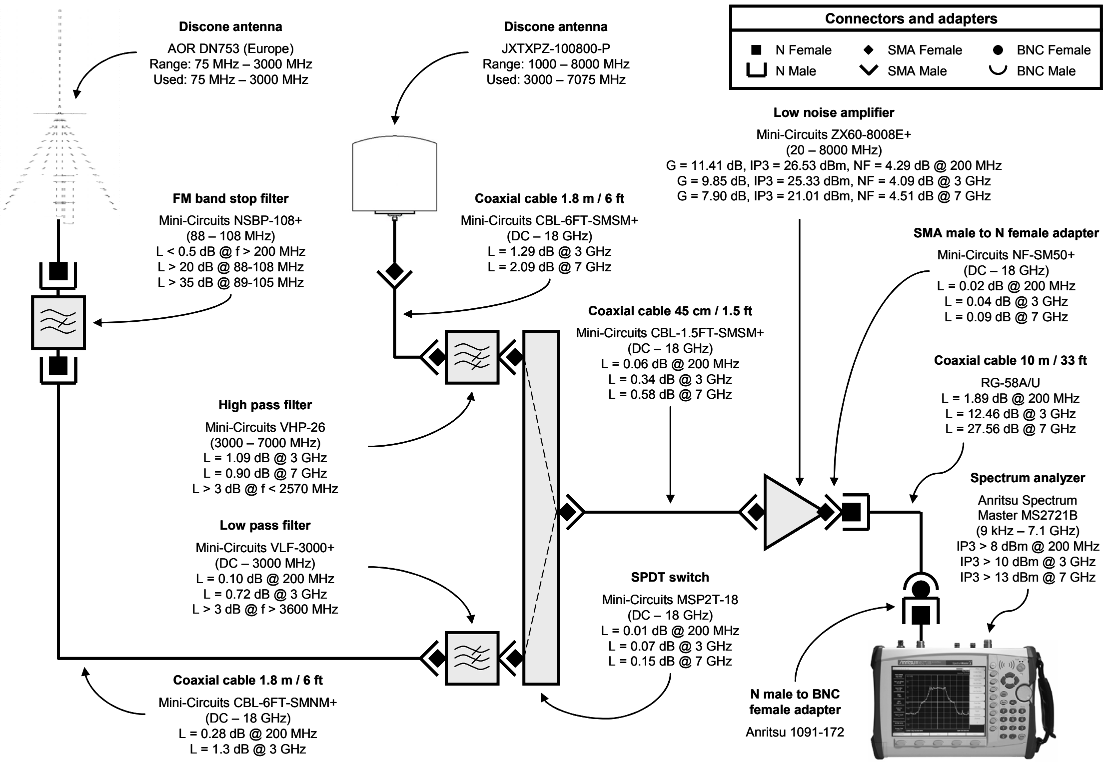


Figure 7: Measurement setup used at UPC.

Two wideband discone-type antennas are used to cover the frequency range from 75 to 7075 MHz. The first antenna (AOR DN753) is used between 75 and 3000 MHz, while the second antenna (A-INFO JXTXPZ-100800/P) is employed between 3000 and 7075 MHz. Discone antennas are wideband antennas with vertical polarization and omni-directional receiving pattern in the horizontal plane. Even though some transmitters are horizontally polarized, they usually are high-power stations (e.g., TV stations) that can be detected even with vertically polarized antennas. The exceptionally wideband coverage (allowing a reduced number of antennas in broadband spectrum studies) and the omni-directional feature (allowing the detection of primary signals coming for any directions) make discone antennas attractive in radio scanning and monitoring applications, and have been a preferred option for many past spectrum studies.

The desired antenna is selected by means of a Single Pole Double Throw (SPDT) switch (Mini-Circuits SPDT Switch MSP2T-18). An electromechanical switch has been selected because of itshigh isolation (90-100 dB) and low insertion loss (0.1-0.2 dB). When compared to other switch types, electromechanical switches in general provide slower switching times and shorter lifetimes. Nevertheless, this is not an inconvenient since antenna switching is always performed off-line by manually turning the switch on/off.

To remove undesired signals, three filters are included. A band stop filter blocks signals in the frequency range of Frequency Modulation (FM) broadcast stations (87.5-108 MHz). Usually, such stations are high power transmitters that may induce overload in the receiver thus degrading the receiver performance by an increased noise floor or by the presence of spurious signals, which inhibits the receiver's ability to detect the presence of weak signals. Since the FM band is of presumably low interest for secondary use due to its usually high transmission power and expected high occupancy rate, a FM band stop filter (Mini-Circuits NSBP-108+) has been employed in order to remove FM signals and avoid overload problems, improving the detection of weak signals at other frequencies. Low pass (Mini-Circuits VLF-3000+) and high pass (Mini-Circuits VHP-26) filters have been used to remove out-of-band signals and reduce the potential creation of inter-modulation products. Finally, to compensate for device and cable losses and increase the system sensitivity, a low-noise pre-amplifier has been included. The selected mid-gain amplifier (Mini-Circuits ZX60-8008E+) provides significant sensitivity improvements while guaranteeing the Spurious-Free Dynamic Range (SFDR) required by the measured signals.

An Anritsu Spectrum Master MS2721B high performance handheld spectrum analyzer is used to provide power spectrum measurements and record the spectral activity over the complete frequency range. This spectrum analyzer provides a measurement range from 9 kHz to 7.1 GHz, low noise level (Displayed Average Noise Level lower than -163 dBm typical in a 1 Hz resolution bandwidth at 1 GHz) and a built-in pre-amplifier (≈25 dB gain) that facilitate the detection of weak signals, fast sweep speed automatically adjusted, and the possibility to connect an external USB storage device to save measurements for later data post-processing. The spectrum analyzer provides a RJ45 Ethernet 10/100 Base T connection that enables controlling the instrument either directly (from a laptop or PC) or remotely (through a local area network or Internet). The spectrum analyzer also includes an external GPS sensor that allows including the coordinates of the current measurement location into the saved traces. Moreover the handheld, battery-operated design simplifies the displacement of the equipment to different measurement locations.

2.4.6 Measurement Equipment and Setup at BNetzA

As the national regulator also operating a number of measurement stations and setup, the BNetzA is naturally in possession of a large number of high end spectrum measurement devices. For the purposes of the two days of spectrum measurement campaigns, Hannover, Krefeld and Konstanz measurement locations utilized a Tektronix RSA 6100A real time spectrum analyzers, connected with high bandwidth antennae and low-loss cable. As an alternative to the Tektronix device, the Konstanz monitoring station also prepared an ESPI 7 type spectrum analyzer for the measurement campaign, with otherwise similar component setup compared to the Tektronix case.

3 Modeling Frameworks for Spectrum Use

In addition to gathering data on how spectrum is used in Europe, the various measurements have also targeted enabling *modeling* of spectrum use. A number of different modeling frameworks have been presented in the literature, each with different foreseen application domains. The key model types considered within the project can be roughly categorized as follows.

1. Models for large-scale characteristics of spectrum occupancy in terms of power spectral density over space, time and frequency. The modelling frameworks appropriate are (at least) random fields (for spatial and space-time behaviour) and after thresholding ON/OFF time series models (for time dependency).
2. Specific usage of spectrum for selected widespread wireless communications systems such as cellular systems. Especially the distribution and time behaviour of load for different basestations and for different systems is of interest. Key modelling tools are from the domain of real-valued time series analysis, hidden Semi-Markov models for session/user level behaviour, and spatial statistics linked with analysis based on proper covariates and trends such as the population distribution and the diurnal cycle.
3. Propagation environment for different frequency bands, both related to large-scale deterministic path loss models, and shadowing. Especially different types of correlations in shadowing are of interest, since they play a key role in performance characteristics of, for example, collaborative spectrum sensing approaches. Appropriate modeling frameworks include correlation functions for random fields and autoregressive spatial models.

In the following we shall give a brief overview of the two most widely applicable modeling frameworks used in the project, namely ON/OFF models for detector outputs and/or transmitter activity patterns, and spatial models for reasoning about the large-scale properties of spectrum use.

3.1 Time and Frequency Domain ON/OFF Models

We begin by briefly recalling the most common ON/OFF models used in the literature to model the behavior of transmitters or the output of a signal detector with binary outputs, often chosen in measurements to the be the energy detector. The term “ON/OFF” refers to the defining property of such models, namely that the *state space* of the models is binary one, that is, we can write the state space as $\mathrm{S} = \{s_0, s_1\}$, where the state s_0 state indicates that the channel is idle (i.e, “OFF”) and the s_1 state indicates that the channel is busy (i.e., “ON”). For dynamic spectrum access applications these often are taken to correspond to availability of the corresponding frequency band for secondary use.

3.1.1 Duty Cycle of the Model

In terms of characterising ON/OFF models, most fundamental property is the *duty cycle*, defined as the probability of the model being in the ON-state, or s_1 in the above notation. From a set of measurement data, assuming the energy detection model and measurements of power spectral density or received signal strength (RSS), we can determine the *empirical duty cycle* as follows. If an RSS sample at time t_i and frequency point f_i is denoted with *RSS*(t_i, f_i), the duty cycle per each frequency point f_i will simply be the empirical average

$$\Psi_{\mathrm{f_i}} = \frac{\sum_{t=1}^{N_i} \Omega(t_i, f_i)}{N_i}$$

where the binary labels

$$\Omega(t_i, f_i) = \begin{cases} 0, if\ RSS(t_i, f_i) < \gamma \\ 1, if\ RSS(t_i, f_i) > \gamma \end{cases}$$

represent the spectrum occupancy at time t_i and frequency point f_i. The parameter γ is the threshold used.

If such an energy based spectrum sensing technique is used to analyze spectrum occupancy, then there are two distinct classes of data, represented by the noise and the signal itself. As a result, setting the decision threshold between both classes is basically a single threshold classification process. There are several thresholding techniques for energy based spectrum sensing, primarily based on empirical data analysis [2, 3], histogram analysis [4], computation of receiver properties [5], thresholding among boundaries [4] etc. ITU recommends a fixed threshold above the ambient noise level [6].

Calculation of an appropriate decision threshold has a great impact on the estimation of the duty cycle. Figure 8 illustrates the duty cycle dependence on the decision threshold for different inspected bands. It is clear that a small variation of the decision threshold can produce high duty cycle variations.

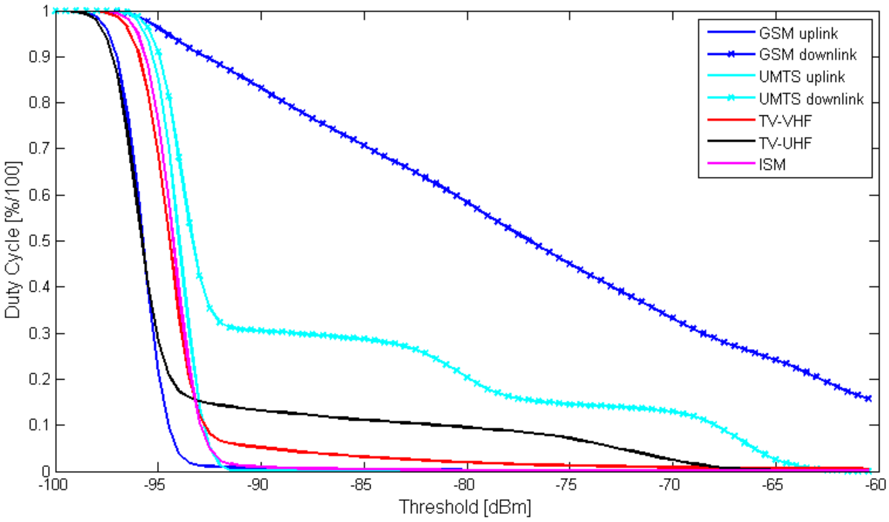


Figure 8: Duty cycle vs. selected threshold for different frequency bands

Generally, the decision threshold for a given measurement set can be chosen based on the histogram of collected measurement samples. Figure 9 depicts an example of measured data in an UMTS downlink channel. The histogram is bimodal, i.e. consists of two peaks representing the noise and the signal samples, respectively. It is evident that a decision threshold should be chosen in the local minimum between the two peaks.

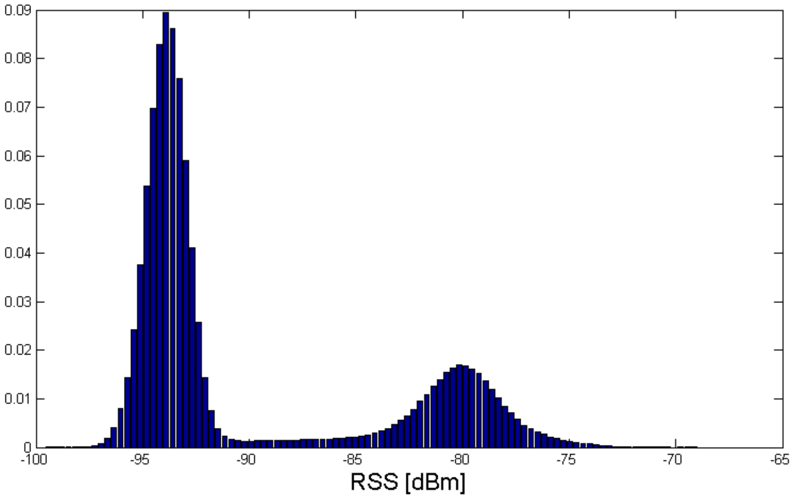


Figure 9: Histogram of measured signal strength in an UMTS downlink channel

Another approach is to measure only the noise samples in a given band and choose a threshold for a given probability of false alarm (P_{fa}). This technique is known as Complementary Cumulative Distribution Function (CCDF) technique. The calculation of the probability of false alarm threshold implies the estimation of a value that is being exceeded from certain percentage of the noise samples. Measurement of noise samples is performed by applying a matched load to the measurement device.

The following experiment illustrates several thresholding techniques and chooses the most accurate one [7]. For this purpose a signal with a known duty cycle is generated with a signal generator. The generated signal, depicted on Figure 10, is being transmitted 20% of time thus producing a duty cycle of 20%. Three different thresholding techniques are applied to the generated samples:

- CCDF based technique with $P_{fa} = 1\%$, $P_{fa} = 5\%$ and $P_{fa} = 10\%$,
- Histogram analysis technique,
- Otsu's algorithm (a thresholding technique typically used in image processing) [8].

Table 3 summarizes the calculated thresholds for the different cases. The technique which most accurately calculates the duty cycle of 20% is regarded as most accurate (in this case, CCDF with $P_{fa} = 5\%$).

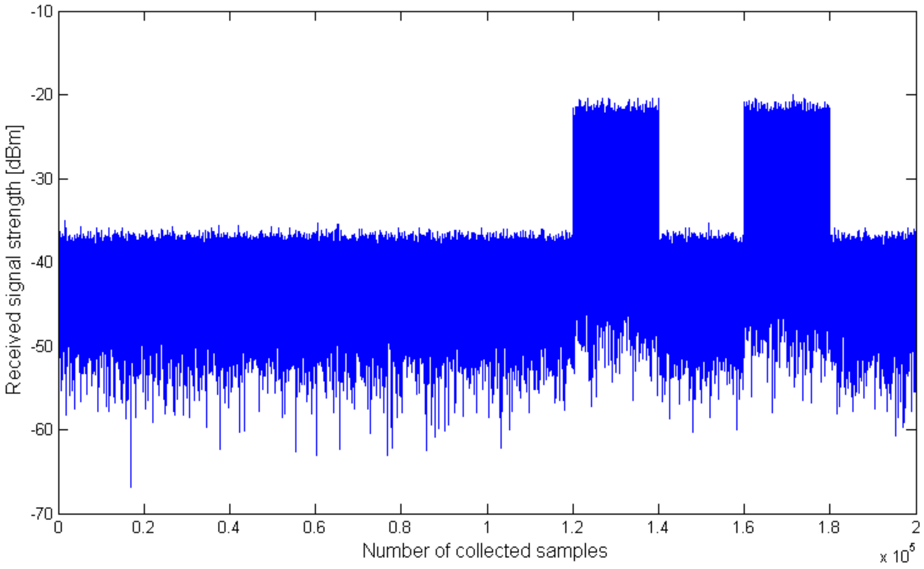


Figure 10: Generated signal with 20% duty cycle

Table 3: Calculated thresholds and corresponding duty cycles for the generated signal

Technique	CCDF, P_F= 10%	CCDF, P_F= 5%	CCDF, P_F= 1%	Histogram analysis	Image processing technique
Threshold [dBm]	-77.7	-76.7	-74.8	-75.5	-69.5
Duty cycle [%]	22.31	18.35	14.15	15.27	11.59
Error [% of duty cycle]	2.31	1.65	5.85	4.73	8.51

In this experiment, the CCDF based technique with P_{fa}= 5% resulted in the best performance in terms of the estimated duty cycle. For analysis of individual measurement campaigns this is precisely the technique most of the consortium partners have adopted.

3.1.2 Time Domain ON/OFF Models

The duty cycle characterization alone does not, of course, take differences in temporal behavior into account. For modeling the behavior of transmitters or detector outputs over time, several approaches have been applied in earlier work. Especially in theoretical work the most common choice is that of a two-state Markov chain, which can be defined either in terms of discrete time (appropriate for systems with discrete time slots) or in continuous time.

In the Discrete-Time Markov Chain (DTMC) model, the time index set is discrete (i.e., $t = t_k = kT_s$, where k is a non-negative integer representing the step number and T_s is the time period between consecutive transitions or state changes). According to this model, the channel remains in a certain state at each step, with the state changing randomly between steps. The behaviour of the channel is described by means of a set of transition probabilities between states as depicted in Figure 11. The transition probabilities can be expressed in matrix form as:

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

where p_{ij} represents the probability that the system transitions from state s_i to state s_j. In its simplest form, the transition matrix $\mathbf{P}$ may be assumed to be constant and independent of the time instant t, in which case the DTMC is said to be *stationary* or *time-homogeneous*.

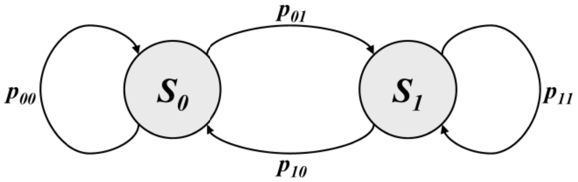


Figure 11: Discrete-Time Markov Chain (DTMC) model.

The duty cycle (DC) of a channel, henceforth denoted by Ψ, continuous to be defined as the probability that the channel is busy. The probabilities that the model of Figure 11 is in each of its states in the long term are given by [10]:

$$P(S = s_0) = \frac{p_{10}}{p_{01} + p_{10}} = 1 - \Psi$$
$$P(S = s_1) = \frac{p_{01}}{p_{01} + p_{10}} = \Psi$$

Thus, the DTMC model can be configured to reproduce any arbitrary duty cycle Ψ by selecting the transition probabilities as $p_{01} = p_{11} = \Psi$ and $p_{10} = p_{00} = 1 - \Psi$, which yields:

$$\mathbf{P} = \begin{bmatrix} 1-\Psi & \Psi \\ 1-\Psi & \Psi \end{bmatrix}$$

Nevertheless, reproducing not only the DC of a channel but also the lengths of the busy and idle periods, that is, the lengths of the periods the channel is in the ON and OFF states, is an important characteristic of a realistic time-dimension model of spectrum usage. In the case of the DTMC model, however, there is no means to account for this feature and, as such, the model is not able to reproduce, in general, this property of spectrum usage. The state holding times for the DTMC models are always geometrically distributed, and cannot especially model more heavy tailed holding times.

The key alternative to the DTMC approach is the two-state Continuous-Time Markov Chain (CTMC) model and especially its extensions, where the channel remains in one state for a random time period before switching to the other state. The state holding time or *sojourn time* is modelled in the CTMC as an exponentially distributed random variable. Although the CTMC model has been widely employed in the domain of DSA/CR research, earlier work also carried out by the FARAMIR consortium has observed based on field measurements that the state holding times are more adequately described by means of generalised Pareto [18], a mixture of uniform and generalised Pareto [19][20], hyper-Erlang [19][20], generalised Pareto and hyper-exponential [21] as well as geometric and log-normal [22] distributions. Based on these results, the typically most appropriate model for the time-domain behaviour is therefore the Continuous-Time Semi-Markov (CTSM) model, where the state holding times can follow any arbitrary distribution. Appropriate parameters for the aforementioned distributions have been derived from field measurements performed with high time-resolution ([19]-[21]) and low time-resolution [23] measurement equipment, and for various radio technologies of practical relevance. It is worth noting that the CTSM channel model provides an explicit means to characterize and reproduce the lengths of the busy and idle periods, which implicitly offers the possibility to reproduce any arbitrary DC by appropriately selecting the parameters of the sojourn time distributions.

Any duty cycle value can be obtained from a Semi-Markov model by selecting the parameters of the distribution in such a way that

$$\bar{\Psi} = \frac{\mathrm{E}\{T_1\}}{\mathrm{E}\{T_0\} + \mathrm{E}\{T_1\}}$$

where $\mathrm{E}\{T_0\}$ and $\mathrm{E}\{T_1\}$ represent the mean duration of idle and busy periods, respectively. The actual choice of distributions and their parameters is typically done using the empirical distributions of state holding times or successive ON/OFF periods. From these empirical distributions a list of candidate models is then compiled, and each of these is then fitted to the data. This is usually done using either the method of moments, or maximum likelihood techniques.

The *method of moments* is a technique for constructing estimators of the parameters that is based on matching the empirical moments with the modeled distribution moments. This method equates empirical moments to population (theoretical) ones. Table 4 summarizes the first and the second theoretical moments for different distributions commonly used for modeling the distributions of ON/OFF periods.

Table 4: A list of common distributions with corresponding PDFs, first and second theoretical moments for selected mathematical distributions

Distribution	**Equation**	**Input param.**	**Mean**	**Variance**
Normal	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ, σ^2	μ	σ^2
Gamma	$f(x; k, \theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)}\ for\ x \geq 0\ and\ k, \theta > 0$	k, θ	$k\theta$	$k\theta^2$
Weibull	$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, & x \geq 0 \\ 0, & x < 0 \end{cases}$	λ, k	$\lambda\Gamma(1+\frac{1}{k})$	$\lambda^2\Gamma\left(1+\frac{2}{k}\right) - \mu^2$
Log-normal	$f_X(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(lnx-\mu)^2}{2\sigma^2}}, \quad x > 0$	μ, σ	$e^{\mu+\sigma^2/2}$	$(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$
Pareto	$f_X(x) = \begin{cases} \alpha \frac{x_m^\alpha}{x^{\alpha+1}}, & x > x_m \\ 0, & x < x_m \end{cases}$	α	$\frac{\alpha x_m}{\alpha-1}$ for $\alpha > 1$	$\frac{x_m^2\alpha}{(1-\alpha)^2(\alpha-2)}$ for $\alpha > 2$
Beta	$f_X(x; \alpha, \beta) = \frac{1}{\mathrm{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$	α, β	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

The *method of maximum likelihood* offers a more principled approach for estimating parameters. The parameters which maximize the sample (empirical) likelihood function are known as the maximum likelihood estimates (MLE). The likelihood function is defined as

$$L(x_1, x_2, \ldots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

where $f(x_i, \theta)$ is the PDF of the empirical data and θ represents the estimated parameters. MLE consists of finding θ which maximizes $L(x_1, x_2, \ldots, x_n, \theta)$ or its logarithmic function.

Both the method of moments as well as the maximum likelihood fitting always yield the "best" parameter values for a given distribution. However, this does not, of course, guarantee that the fit is actually a good one. Further, especially the method of moments does not provide guidance on how to select the "best" distribution amongst a family of candidate distributions. A goodness of fit measure is typically therefore used for matching empirical samples with fitted ones by a theoretical model. Typical examples of goodness of fit metrics include mean deviation or the root mean squared error (RMSE). Also probabilistic goodness of fit metrics are often used. For instance, the Kolmogorov-Smirnov test is used to decide if a sample comes from a population with a specific distribution. It can be applied to both discrete (count) and continuous data. The test is based on a comparison between the empirical distribution function (ECDF) and the theoretical one of the inspected distributions. The test statistic of the test is:

$$D_n = \sup|F(x_i) - F_n(x_i)|, \qquad 1 \leq i \leq n$$

which is the upper extreme among absolute value differences between ECDF and theoretical CDF.

3.1.3 Frequency Domain ON/OFF Models

The most common approach for modeling the frequency domain or joint frequency-time domain behaviour of a system is to model the distribution of duty cycles over the specified frequency band. This distribution is often further modelled in two phases, first based on the (marginal) probability distribution of the DC values for channels within the same band, followed by the modeling DC clustering over frequency, i.e., the existence of groups of contiguous channels with similar DC values. Both aspects are discussed in detail in the following paragraphs.

3.1.3.1 Duty cycle distribution models

Assuming that the DSA/CR system operates over a set of C primary radio channels, denoted by $\Upsilon = \{\upsilon_1, \upsilon_2, \ldots, \upsilon_c, \ldots, \upsilon_C\}$, and given the set $\boldsymbol{\Psi} = \{\Psi_1, \Psi_2, \ldots, \Psi_c, \ldots, \Psi_C\}$, where Ψ_c is the DC of channel υ_c, the probability distribution of the elements of $\boldsymbol{\Psi}$ can accurately be fitted with the beta [15] and Kumaraswamy [16] distributions. The CDF for the former is given by:

$$F_B(x; \alpha, \beta) = I_x(\alpha, \beta) = \frac{B_x(\alpha, \beta)}{B(\alpha, \beta)}, \quad x \in (0,1)$$

where $\alpha > 0$ and $\beta > 0$ are shape parameters, $I_x(\alpha, \beta)$ is the regularised incomplete beta function, $B_x(\alpha, \beta)$ is the incomplete beta function given by:

$$B_x(\alpha, \beta) = \int_0^x z^{\alpha-1}(1-z)^{\beta-1}\,dz$$

and $B(\alpha, \beta)$ is the ordinary beta. The CDF for the latter is given by:

$$F_K(x; a, b) = 1 - \left(1 - x^a\right)^b, \quad x \in (0,1)$$

where $a > 0$ and $b > 0$ are shape parameters.

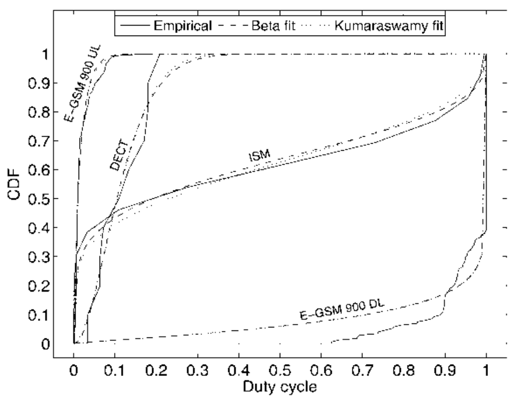


Figure 12: Empirical DC distributions and their corresponding beta and Kumaraswamy fits.

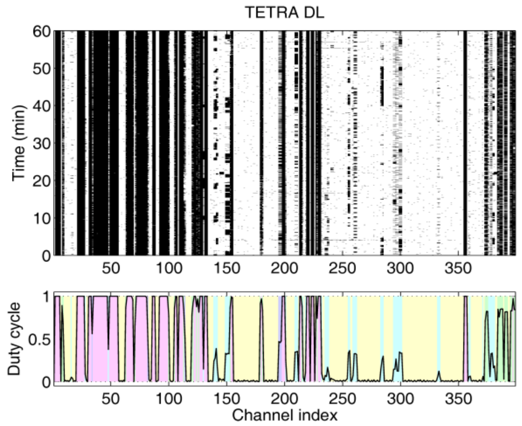


Figure 13: Example of empirical DC clustering (TETRA downlink band).

Figure 12 shows some examples of empirical DC distributions and their corresponding beta and Kumaraswamy fits. The selected bands represent examples for a wide variety of load levels, including very low (E-GSM 900 UL), low (DECT), medium (ISM) and very high (E-GSM 900 DL) average band DCs. From the figure, it can be appreciated that the beta and Kumaraswamy distributions are able to provide reasonably accurate fits in all cases.

Table 5 provides example values for the distribution parameters for these cases and selected additional ones, computed based on one of the measurement data sets gathered in the project by means of maximum likelihood estimation techniques, along with the average band DCs $\overline{\Psi} = \left(1/C\right)\sum_{c=1}^{C}\Psi_c$.

Table 5: Fitted values for DC distributions.

Measured Band	Average DC $\overline{\Psi}$	Beta		Kumaraswamy	
		α	β	a	b
Amateur	0.17	0.5796	2.8963	0.6311	2.5599
Paging	0.28	1.4867	3.9601	1.3449	4.2382
TETRA UL	0.03	0.7105	44.0554	0.7849	26.9302
TETRA DL	0.36	0.1840	0.2837	0.1389	0.4223
GSM900 UL	0.02	1.6044	116.6408	1.2690	208.5805
GSM900DL	0.96	0.9119	0.0778	0.8970	0.0786
DCS1800UL	0.02	0.2023	6.0738	0.2545	2.6118
DCS1800DL	0.44	0.4525	0.6118	0.4463	0.6846
DECT	0.12	2.3217	17.5170	1.7434	34.2432
ISM	0.42	0.2022	0.3418	0.1426	0.4155

3.1.3.2 Duty cycle clustering models

Channels with similar load/DC levels rarely occur alone, but in groups of a certain size. The existence of groups of contiguous channels with similar DC values can be analysed and modelled by defining a set of DC archetypes. As a reference, the following DC archetypes can be considered: very low $\Psi \in [0, 0.05]$, low $\Psi \in (0.05, 0.40]$, medium $\Psi \in (0.40, 0.60]$, high $\Psi \in (0.60, 0.95]$ and very high $\Psi \in (0.95, 1]$ levels. Based on these archetypes, Figure 13 shows an example of DC clustering in the TETRA downlink band (the upper graph shows the instantaneous spectrum occupancy for each channel for a time period of 60 minutes, where white/black points indicate idle/busy observations respectively, while the lower graph shows the channel DCs and their corresponding classification into the considered archetypes). As it can clearly be appreciated, channels with similar occupancy levels appear together in blocks of a certain size, i.e. the DC is clustered in the frequency domain. The number of contiguous channels per cluster (i.e., group of channels belonging to the same DC archetype, shown in different colours in the figure) can be modelled as a geometrically distributed random variable, whose CDF is given by [15]

$$F_{Geom}(k; p) = 1 - (1 - p)^k ,$$

where k represents the number of channels belonging to the same group (i.e., the cluster size) and $1/p$ $(0 \le p \le 1)$ represents its expected (mean) value, i.e. $\mathrm{E}\{k\} = 1/p$. The value of the parameter p can be set for a particular average number M of channels per cluster as $p = 1/M$. The relation $p \approx C \cdot 10^{-3}$, with C being the number of channels in the whole band, can be used as an empirical approximation as long as the resulting value satisfies $p \le 1$. An alternative empirical approximation is to draw the value of p from a uniform distribution within the interval $[0.1,\ 0.6]$.

3.2 Spatial Models

While most of the work in modeling spectrum use in the literature has focused on the time-frequency domain behavior, for a number of practical applications the *spatial* structure of spectrum use is crucial as well. Key example of inherently spatial property of spectrum use is the coverage area of a given transmitter, defined as the region in which the SINR for the corresponding receivers is above a chosen threshold. Spatially correlated shadow fading is another key example of an inherently spatial phenomenon. We give in this section a brief overview of the techniques that can be utilized for constructing spatial models suitable for dealing with such questions, and apply them on our measurement data sets in Section 6.

The basic concept throughout the following will be that of a *random field Z(x)*. Depending on the application, the values of *Z* will denote, for example, the shadow fading coefficient for a given transmitter, or spectrum occupancy in a given region. The common feature in all of these cases is that for any location *x*, the value of *Z*(*x*) will be a random variable, and that the values of *Z* at nearby locations will *not* in general be independent.

Constructing a model for a spatial phenomenon using the random field framework consists of selecting the *marginal distribution* of Z, and specifying the *correlation structure*. Marginal distribution means simply the probability distribution the values of Z in general follow, and is usually taken to be Gaussian. Given an empirical data set the marginal distribution of Z can also be determined using the techniques introduced above in the case of fitting sojourn time distributions to empirical data sets. The correlation structure will of the other hand dictate how dependent nearby values of Z will be of each other, and how that dependency changes with distances between locations considered. The most commonly used measure for correlation in this context is the *semivariogram*, defined by

$$\gamma(x,y) \equiv \frac{1}{2} E(Z(x) - Z(y))^2 .$$

Here "E" denotes the expectation operator, so the semivariogram measures the average difference squared between two locations. The values for nearby locations tend to be similar leading to low values of the semivariance, whereas values of Z at distant locations tend to become independent, causing the semivariogram to level off at long distances. The asymptotic value corresponds simply to the variance of the random field values.

In order to make this discussion more concrete, let us consider a specific example. Figure 14 shows an example realization of a random field, which we shall use throughout as a basis for our case study. The correlation structure of the example is selected to resemble mean PSD values measured on UMTS bands during one of our measurement campaigns, with the simulation area rescaled to the unit square. We clearly see the correlation of nearby values, and shall now focus on estimating and modeling it. In order to estimate the semivariogram for the given data set, we replace the mathematical expectation operator in the above definition by the simple empirical average given by $\hat{E}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$. That is, the estimate of the semivariogram value at given distance between *x* and *y* is found by finding all the pairs of data separated by that distance, and then averaging the squared different of the data values. Carrying out this computation[1] for a collection of different distances results in the *empirical semivariogram* for the data set.

---

[1] We use the geoR-package [1] for the "R" environment [2] throughout for the calculations (namely functions "variog" and "variofit"), although implementations of the necessary routines are available for other platforms as well.

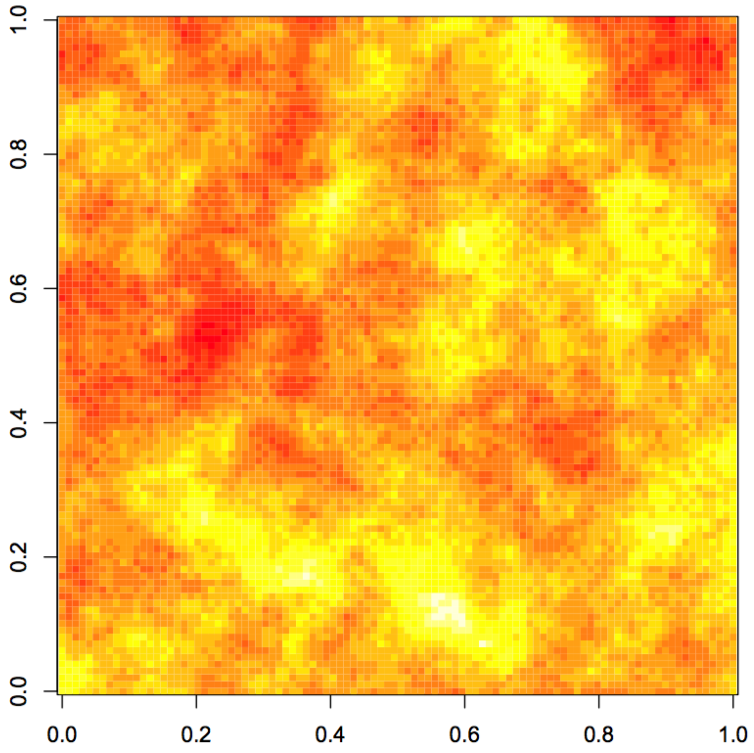


Figure 14: The example random field used in the case study.

For the data set depicted in Figure 14, the corresponding empirical semivariogram is illustrated in Figure 15. The low values of the semivariance as origin is approached indicate that nearby values in our data set are indeed correlated, and also that values at locations separated more than 0.2 units start to become uncorrelated, since the semivariogram levels off there.

The empirical semivariogram is useful graphical tool for understanding the correlation structure in data. However, in order to create a useful *model* for the data one more step is required, namely fitting of a *semivariogram model* to the empirical semivariogram. Semivariogram models are functions of distance with three or more parameters, such a

$$\gamma(r) = a + b(1 - \exp(-r/c)).$$

The model is fitted to the empirical semivariogram using a method such as least squares in order to determine the parameters *a*, *b* and *c*. Of these parameters *a*, called the *nugget*, measures short-scale variation in the data (such as induced by receiver noise). The *sill* parameter *b* then measures the overall variance of the data, while the *range* parameter *c* measures the distance at which correlations start to significantly decay.

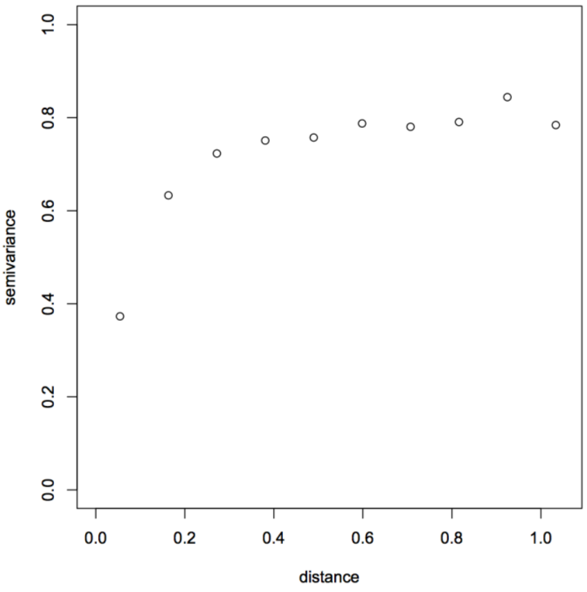


Figure 15: The empirical semivariogram for the random field given in Figure 1.

Carrying out the fitting procedure for the empirical semivariogram depicted in Figure 15 resulted in a fit with $a = 0$, $b = 0.7839$ and $c = 0.0997$. The corresponding graph of the model is shown in Figure 16, visually confirming the good quality of the fit.

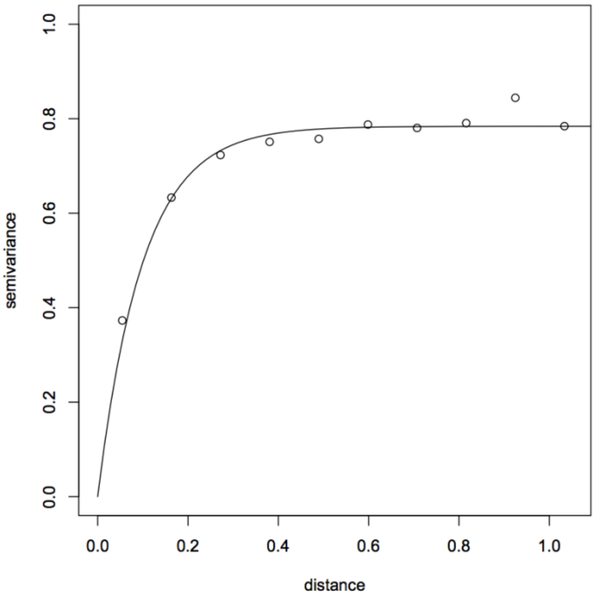


Figure 16: The fitted semivariogram model together with the empirical semivariogram values.

We shall now focus on a specific application of the random field techniques, namely on prediction of spectrum occupancy at locations for which no direct measurement data is available. This procedure is illustrated in Figure 17 below. The leftmost panel in the figure illustrates the "ground truth", that is, the *true* spectrum occupancy over space. From this a smaller sample shown in the middle panel is made. The goal is now to *predict* the values in areas "between" the different measurement locations. Example of an outcome of such a prediction process (called *kriging* in the spatial statistics literature) is shown on the panel on the right. Clearly the result is not completely accurate, but is nevertheless relatively good considering the small number of samples the prediction is made with. Techniques of this kind can be very powerful especially in coverage estimation problems, in which classical, for example path loss model based, techniques can result in significant errors.

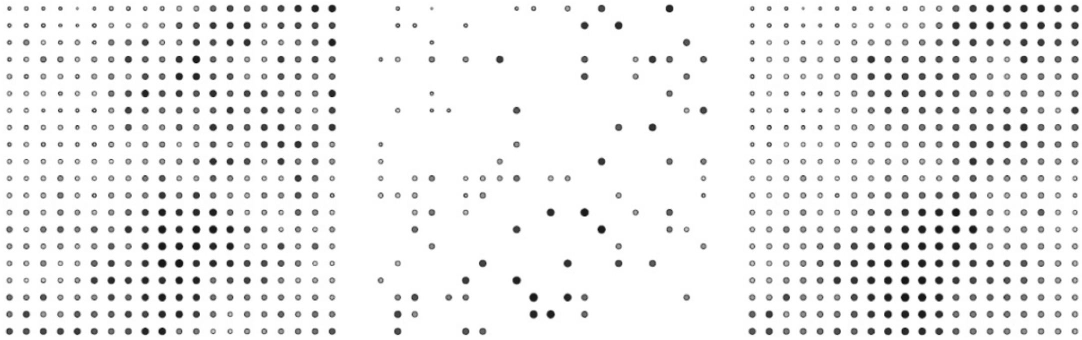


Figure 17: Illustration of prediction of spectrum occupancy over space.

We shall now outline in more detail how such predictions can be made based on the developed spatial models, and also how its accuracy can be estimated in terms of formal error bars at individual locations. The foundation for both is again the fitted semivariogram model. Intuitively one would expect the amount and range of correlation to be closely related to the accuracy of the prediction process. This intuition turns out to be exactly correct, as we shall soon observe.

Simplest predictor that can be written down for the value of *Z* at a location *y* for which no direct measurement data is available is the weighted linear combination of other values:

$$\hat{Z}(y) = \sum_i w_i(x_i) Z(x_i).$$

The problem is to now determine the weights *w* in manner that uses the information available in the semivariogram, and minimizes some type of prediction error. Typical conditions involve minimizing the *variance* of the estimator, that is, minimizing

$$\mathrm{var}(Z(x) - \hat{Z}(x))$$

subject to the condition

$$E(Z(x) - \hat{Z}(x)) = 0$$

stating that the estimator is *unbiased*. This results in the following equation for the weights:

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(x_1, x_1) & \cdots & \gamma(x_1, x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n, x_1) & \cdots & \gamma(x_n, x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(x_1, y) \\ \vdots \\ \gamma(x_n, y) \\ 1 \end{pmatrix}.$$

Here μ is a parameter allowing for the estimation of the prediction variance using the equation

$$\sigma^2(y) = \sum_i w_i(y)\gamma(x_i, y) + \mu.$$

From this, we see that the behavior of the semivariogram also plays a key role in the expected error of the obtained estimate.

4 Results from European Days of Spectrum

We shall now move on to discuss the results obtained from the various spectrum measurement campaigns. We begin with the Two Days of Spectrum in Europe, first commenting on the overall spectrum occupancy figures observed in the measurements. Following this, we move on to the cellular bands, and compare the usage of frequencies assigned for cellular systems across the measurement locations. We also look into the changes induced by the diurnal cycle, i.e., differences between daily and nightly usage. In order to deepen the understanding of the more detailed characteristics of spectrum use, we then present selected results also on raw power spectral density values observed during the measurements, and discuss how the selection of the threshold for ON/OFF statistics and models influences the results. Finally, selected modeling applications of the measurement data set are given.

Due to the large number of measurement locations, as well as the vast amount of data collected at each of them, our aim here is not to give a comprehensive analysis of the data set. Instead, we focus on some of the key findings, and also lessons learned that we expect to be of interest to other research groups and projects and seek to carry out similar extensive measurement campaigns.

4.1 Overall Spectrum Use

Figure 18 shows the overall spectrum occupancy as measured at the Aachen measurement location during the first day of measurements for the 2011 campaign. For computation of the duty cycle an energy detection threshold of -93 dBm / 100 kHz was used. Unless mentioned otherwise this is also the value, chosen based on the sensitivities of the equipments used by the different partners, that will used in the subsequent results. The figure shows very typical spectrum usage behavior observed in the measurements, and also what would be expected based on earlier measurement campaigns and the regulatory environment. The highest occupancy figures can be seen to correspond to narrowband services used at low frequencies, terrestrial digital TV in the 470 MHz – 790 MHz UHF band, and cellular systems operating on the 900 MHz, 1800 MHz and 2.1 GHz bands. Beyond these, hardly any usage of spectrum can be seen at the Aachen location.

Figure 19 illustrates similar overall spectrum occupancy figure for the case of Leuven measurement location. For higher frequencies very similar behavior compared to the Aachen case can be observed. For frequencies below the DVB-T UHF band slightly higher usage can be seen, and due to differences in the channels assigned to local DVB-T transmitters, the UHF band usage itself is naturally slightly different. Nevertheless, no major changes compared to the Aachen case can be seen at this level.

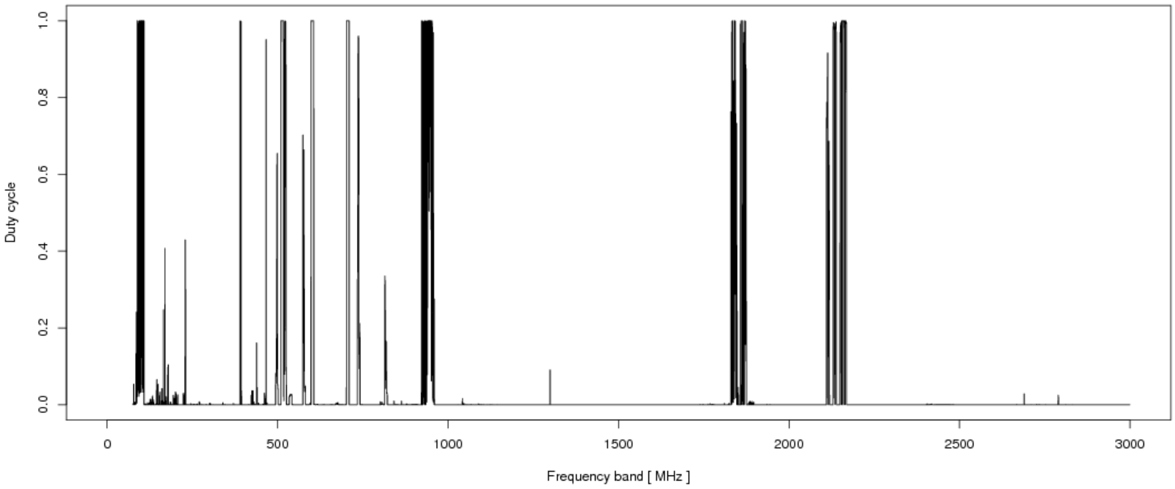


Figure 18: Overall spectrum utilization in Aachen for the first measurement day.

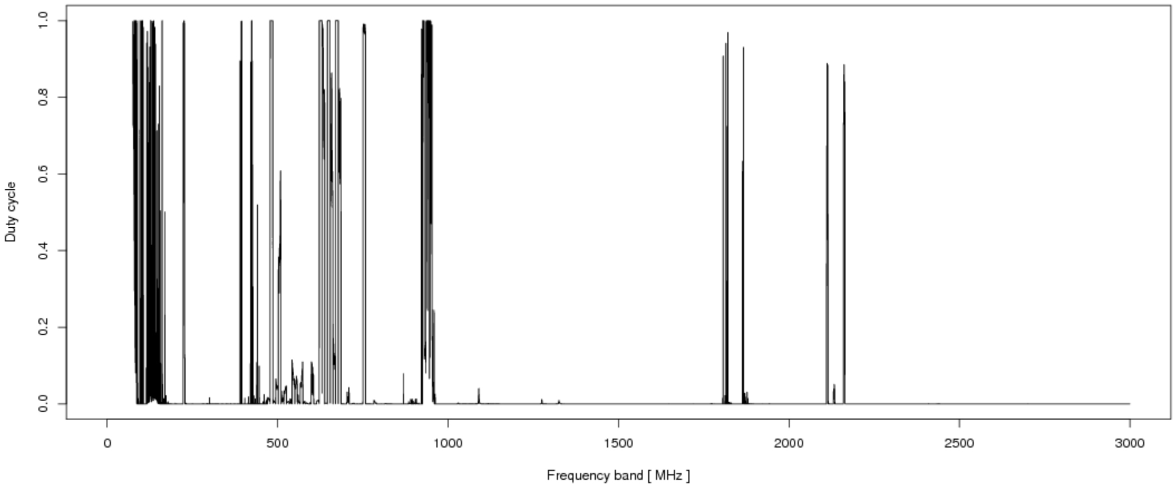


Figure 19: Overall spectrum utilization in Leuven for the first measurement day.

The situation at the Skopje location is, however, somewhat different. As can be seen from Figure 20, much higher usage of frequencies below 800 MHz was observed, and selected higher frequency bands that were almost completely unoccupied in the Aachen and Leuven cases have transmitters present in the Skopje data set. These results emphasize the heterogeneity of Europe also in terms of spectrum regulations and usage, and also highlight the significantly different spectral environments that cognitive radios and systems based on dynamic spectrum access principles must be able to cope with.

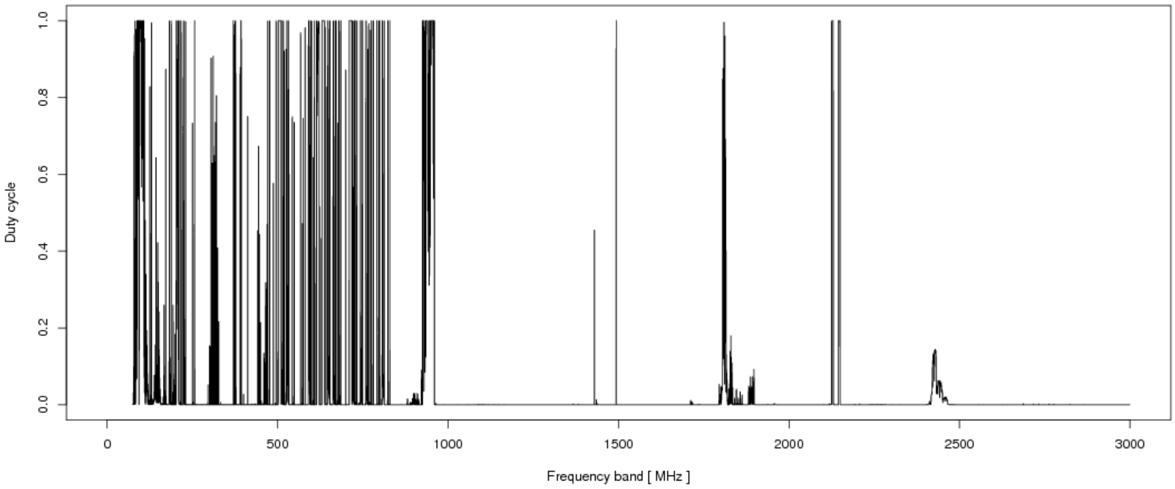


Figure 20: Overall spectrum utilization in Skopje for the first measurement day.

4.2 Spectrum Use in the Cellular Bands

We shall now move on to discuss spectrum use in the cellular bands. Most complicated usage patterns are expected in the 900 MHz GSM downlink band due to its universal availability as a service, combined with high degree of temporal dynamics and narrow bandwidths used at the physical layer. As can be seen from Figure 21, the results from the Aachen data set on the usage of this band are highly complex indeed. The overall spectrum occupancy is clearly very high, as can obviously be expected, but some small amount of unused spectral resources still remain (however, this might also be due to the threshold used, which is higher than the sensitivity of a typical mobile phone transceiver). Note also that the difference between daytime use and nighttime use can be seen, but is not particularly significant except for few very narrow bands around 945 MHz.

While in the Aachen data set some unused spectrum (with the caveats mentioned above) was observed in the 900 MHz band, in many other data sets almost nothing was shown to have been left unused. For example, as can be seen from Figure 22, showing the 900 MHz band occupancy for the Krefeld data set, the band is in practice completely occupied. Even at night time only small reduction in usage can be seen for the highest frequencies in this band. The situation in the Skopje data set is again somewhat different from the ones discussed above, as shown in Figure 23. We see that although the daytime occupancy is also very high, in the 940-948 MHz band the effects of the diurnal cycle are very clearly seen. This is presumably an outcome of the radio resource management strategy employed by a particular cellular operator.

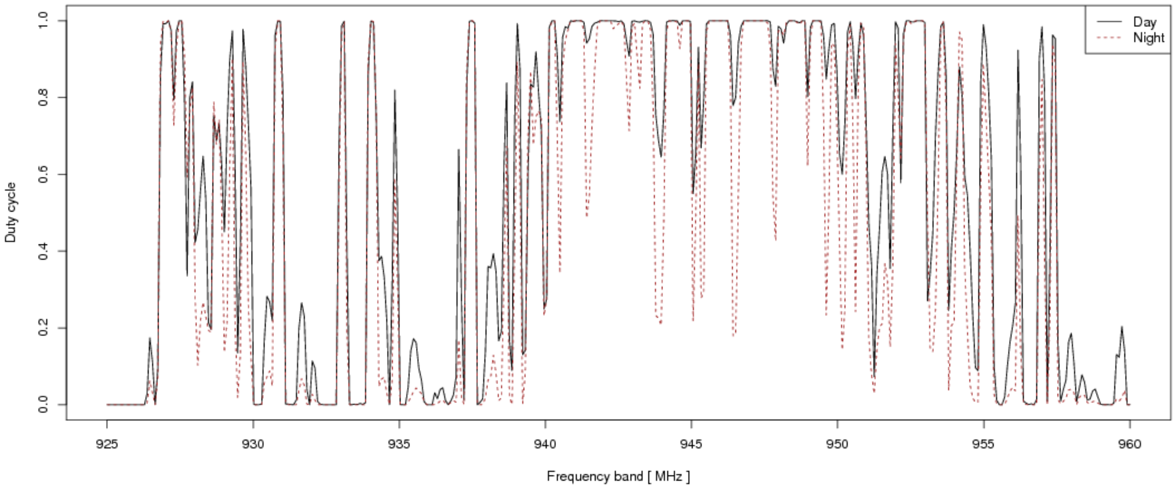


Figure 21: Usage of the 900 MHz GSM band in Aachen.

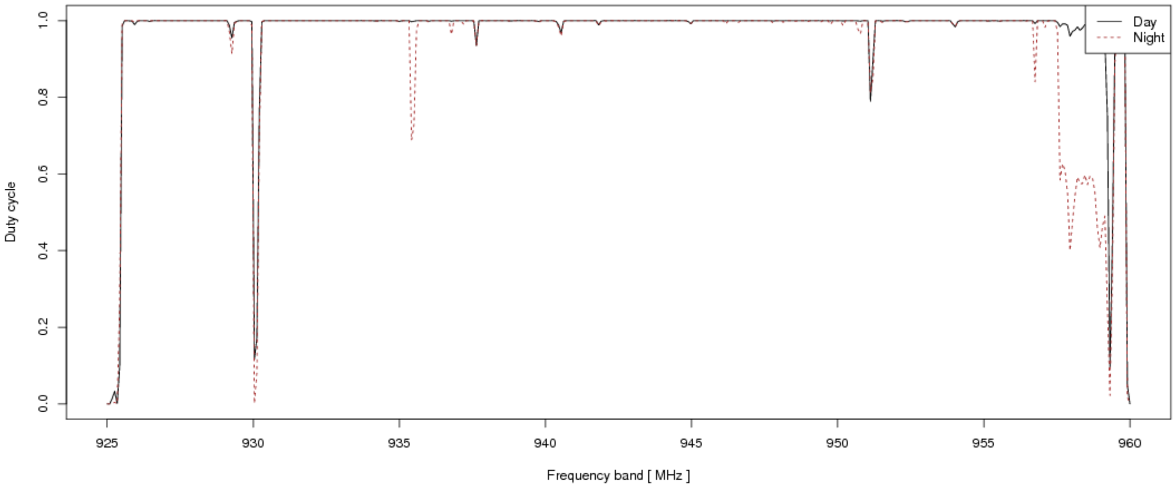


Figure 22: Usage of the 900 MHz GSM band in Krefeld.

Even though the 900 MHz band was very highly utilized at all measurement sites, interestingly the higher cellular bands at higher frequencies showed much greater degree of variability and unused frequencies. An example is given in Figure 24, contrasting the usage of the 2.1 GHz UMTS downlink band between Aachen and Skopje. In the Aachen region 8 channels each 5 MHz wide are clearly in use, whereas in Skopje only two can be observed. In both cases significant amount of this band is still unused. The results for the Aachen data set also provide an interesting example on how critical it is to interpret the duty cycle values correctly, as detector outputs or *perceived* spectrum use rather than actual activity patterns of transmitters. The two channels between 2110 and 2120 MHz show less than 100% duty cycle and even visible amount of variation between day and nighttime results, even though UMTS Node B is transmitting continuously due to the CDMA nature of the technology and the presence of the broadcast channel. Therefore all the seen effects of the diurnal cycle and lower duty cycle are cause either by environmental effects, or power control by the Node Bs around the measurement locations which cause the received signal strength to fluctuate around -93 dBm.

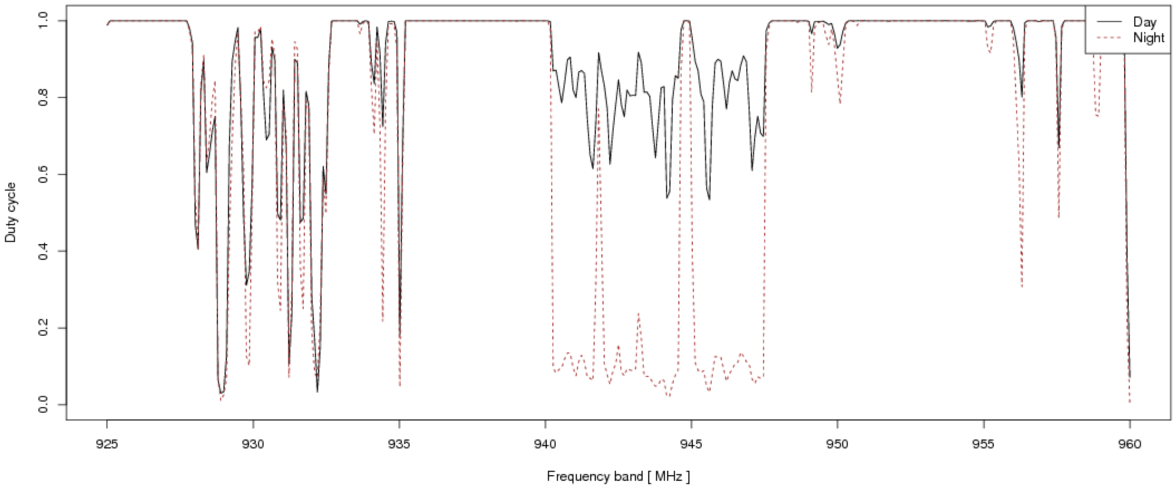


Figure 23: Usage of the 900 MHz GSM band in Skopje.

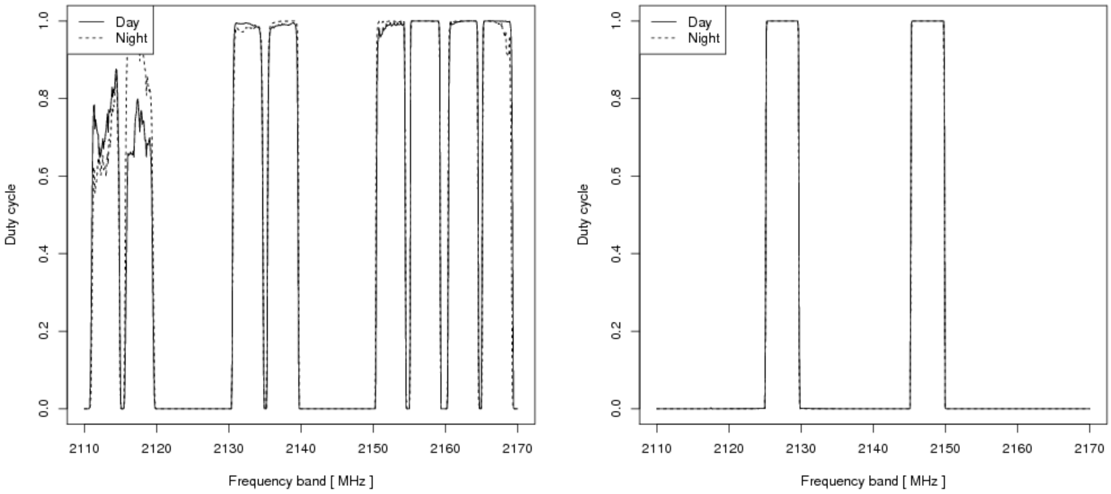


Figure 24: Usage of the 2.1 GHz UMTS downlink band in Aachen and Skopje.

4.3 Distributional Properties of Spectrum Use

Let us now consider in more detail the distribution of the power spectral density values in the measurement data, rather than the aggregate thresholded values the results given above have been based on. Figure 25 shows the estimated probability density functions (PDFs) for the power spectral densities (PSDs) measured at various sites for the DVB-T UHF and GSM900 downlink bands. Due to the lower overall usage of the DVB-T band the PDFs are rather flat with the exception of the prominent peak at low PSD values. The latter arises mainly from the noise and residual interference in the system. The minor modes seen at higher PSD levels then correspond to actually used channels, but since major part of the UHF band in unoccupied especially in the central European data sets, these are not as distinct as the peak due to the noise floor. The situation is very different for the highly occupied GSM900 downlink band. For most measurement locations the distributions are clearly multimodal, with the mode at lowest PSD values again corresponding to the noise floor, and other modes roughly corresponding to different base stations around be measurement location. From these results we can see how difficult it is to draw any conclusions on the usage of a band purely from aggregate statistics such as the PDFs, not to mention statistical characteristics of those.

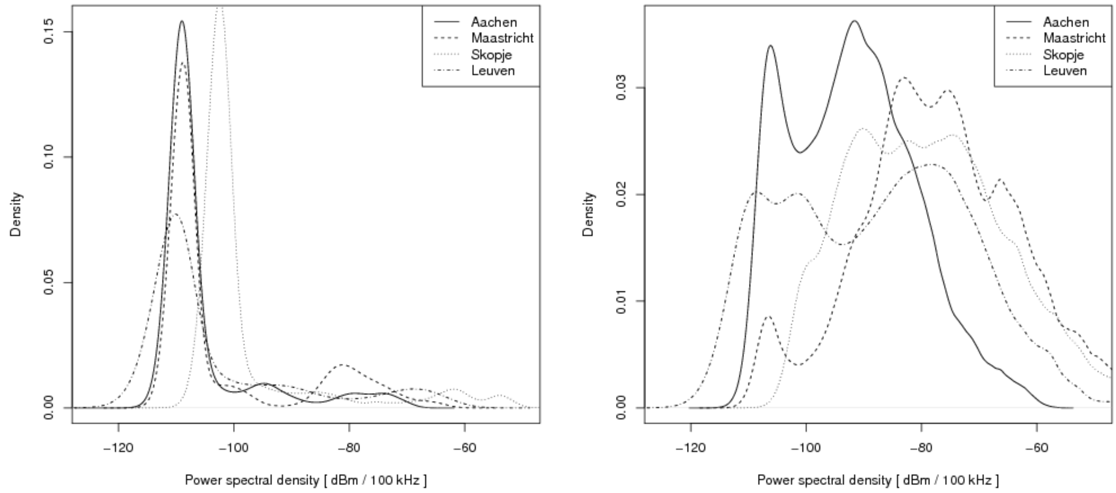


Figure 25: Probability density function estimates for the measured power spectral density values for the UHF DVB-T band and the GSM900 downlink band.

We have also carried out detailed studies of the distributional properties of, for example, the duty cycle over different frequency bands. Figure 26 shows the marginal distribution of the duty cycle for the GSM900 and GSM1800 downlink bands for a selection of measurement locations. Note that in both cases there are significant similarities in the shapes of the marginal distributions, although precise "location" of these distributions varies between measurement locations. It is precisely due to these similarities why the beta distribution fits discussed in Section 3 as means to model the distribution of duty cycles within given bands can yield such good results. Notice also how significantly lower the overall usage of the GSM1800 downlink band is opposed to the GSM900 case, further providing evidence of the potentially significant amount of unused spectrum within the cellular bands at higher frequencies.

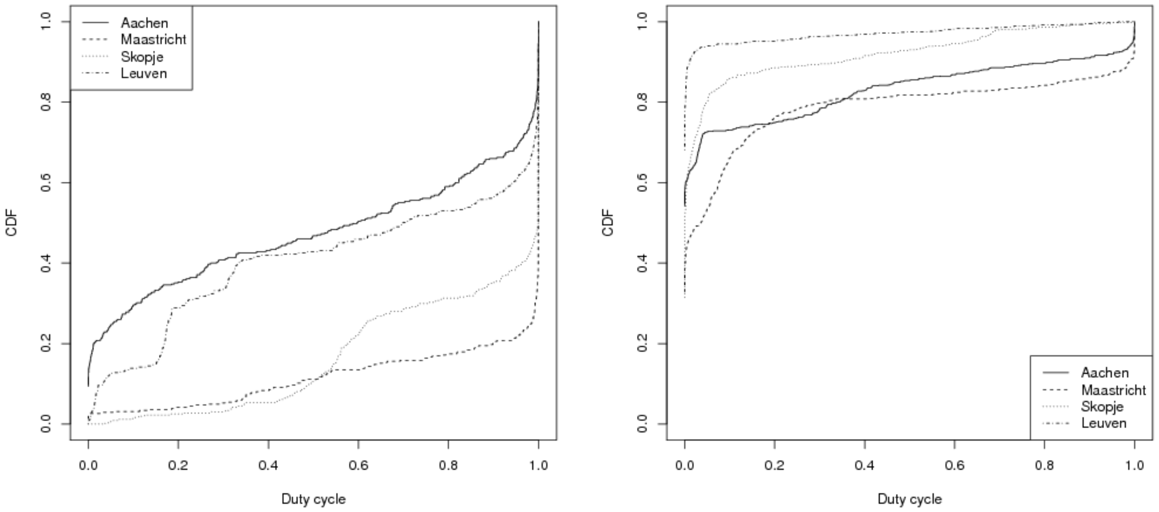


Figure 26: The cumulative distribution functions for the duty cycles computed over the GSM900 (left) and GSM1800 (right) downlink bands.

4.4 Influence of the Choice of Threshold

All the duty cycle related results given above were obtained by using a single threshold, namely -93 dBm. We shall now briefly discuss the impact of changing this threshold, and how such changes can influence the results inferred from the data sets.

First of all it is critical that the value chosen is low enough in order to be meaningful compared to the typical service characteristics of the technologies employed at given bands, but not so low as to close the upper tail of the noise and interference distribution of the measurement setup. We show in Figure 27 the overall spectrum usage plot for the Skopje data set, but this time using a threshold of -99 dBm which is close but not quite at the mean of the noise floor of the setup. Due to the impact of the noise, the frequency bands that were previously observed to be completely unused are now seen to show small amounts of usage, and even with relatively complex structure over the frequency domain. This shows both that awareness of the noise floor is important, but also, perhaps even more critically, that due to complexities of interference and non-linearities in the system, the typical assumption of white noise for unused bands is also not appropriate.

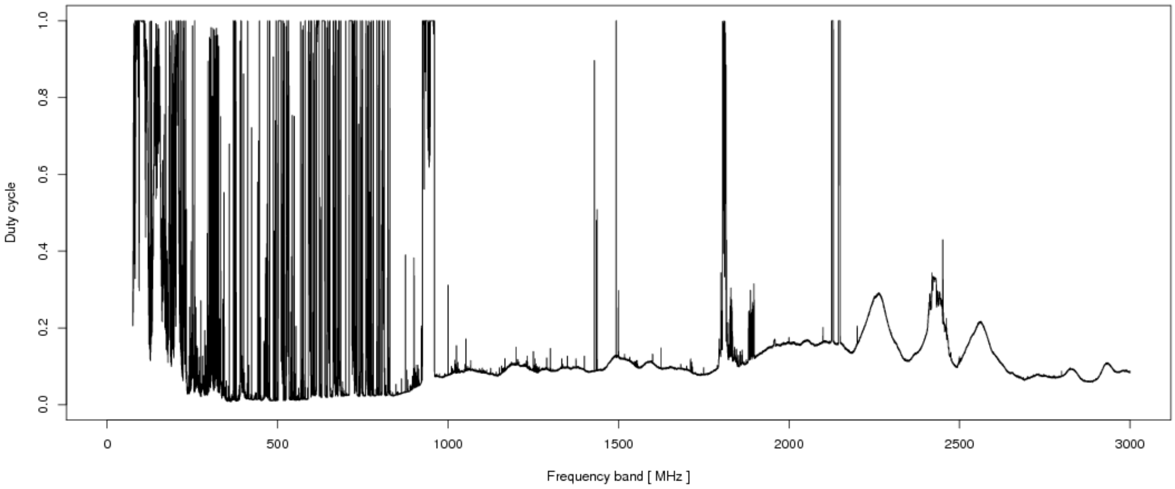


Figure 27: Overall spectrum utilization in Skopje for the
first measurement day computed using too low energy detection threshold.

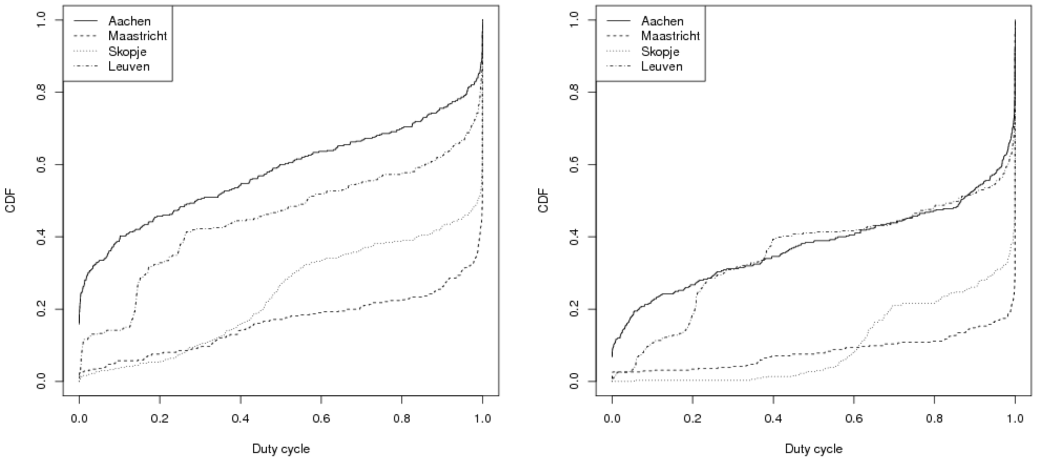


Figure 28: The cumulative distribution functions for the duty cycles
computed over the GSM900 downlink bands for thresholds of -90 dBm (left) and -96 dBm (right).

Figure 28 shows the CDFs of the distributions of duty cycles of the GSM900 downlink band analogously to Figure 20 above, but this time computed using two different threshold values. We see that the overall distribution experiences a shift due to the change in the threshold, but also that the overall shape of the distribution remains relatively unchanged. This indicates that for modeling purposes the influence of the aggregate duty cycle level on a band for a given application should be explored carefully, but also that the obtained models for the marginal distribution can be rather robust with respect to changes in other parameters of the fitted distribution.

4.5 Further Results and Models

We shall conclude our discussion of the two days measurements by giving selected examples of the models derived from the obtained data sets. Throughout this subsection the data set obtained at the Skopje measurement site will be used as a basis. We shall also proceed in a slightly more fine-grained manner compared to the above discussion, proceeding in per-channel fashion for technologies that have a well known channelization structure in the frequency domain.

The analysis carried out on each band and channel consists of the following steps:

- **Classification of measured data** - classification of the bands of interest that should be analysed through common characteristics, such as spectrum occupancy on the band;
- **Calculation of duty cycle** - calculation of the empirical duty cycle per frequency point for each band or channel;
- **Model choice** - choosing a mathematical model to represent the empirical duty cycle values;
- **Parameter estimation** - estimation of the parameters for the chosen mathematical model;
- **Goodness of fit** - matching the empirical values with fitted ones by a theoretical model.

4.5.1 Classification of measured data

Each of the inspected bands represents a collection of multiple communication channels. For the FDD based technologies (GSM 900, UMTS 1900), the uplink and the downlink bands are characterized with different spectrum use and should be analyzed separately. Furthermore, the UMTS 1900 band has only 2 active downlink and uplink channels as we also saw in Figure 24. Inspecting both downlink and uplink channels as a single band would also lead to wrong conclusions about the spectrum use. Therefore, each of the channels will be inspected separately in a per channel manner. Corresponding histograms of RSS samples for each UMTS channel are depicted on Figure 29. Bands used for broadcasting, such as the UHF and VHF TV bands comprise only of downlink traffic, and the analysis for them will be done in a per-band manner. Finally, the ISM band also does not have assigned frequencies for uplink and downlink transmission and its statistical usage will be inspected as a whole band.

Table 6 gives a summary of the inspected bands.

Table 6: Inspected bands for the detailed analysis of the Skopje data set.

Band	**Classification of measured data**	**Type of analysis**
GSM 900	GSM 900 uplink (880-915 MHz),	Per sub-band
	GSM 900 downlink (925-960 MHz)	Per sub-band
UMTS 1900	UMTS uplink channel 1 (1920-1935 MHz)	Per channel
	UMTS uplink channel 2 (1950-1965 MHz)	Per channel
	UMTS downlink channel 1 (2110-2125 MHz)	Per channel
	UMTS downlink channel 2 (2140-2155 MHz)	Per channel
UHF	UHF (470-582 MHz) full band	Per band
VHF	VHF (174-230 MHz) full band	Per band
ISM	ISM (2400-2500 MHz) full band	Per band

Figure 29 and Figure 30 depict the corresponding histogram of the RSS samples for all inspected bands in Table 6.

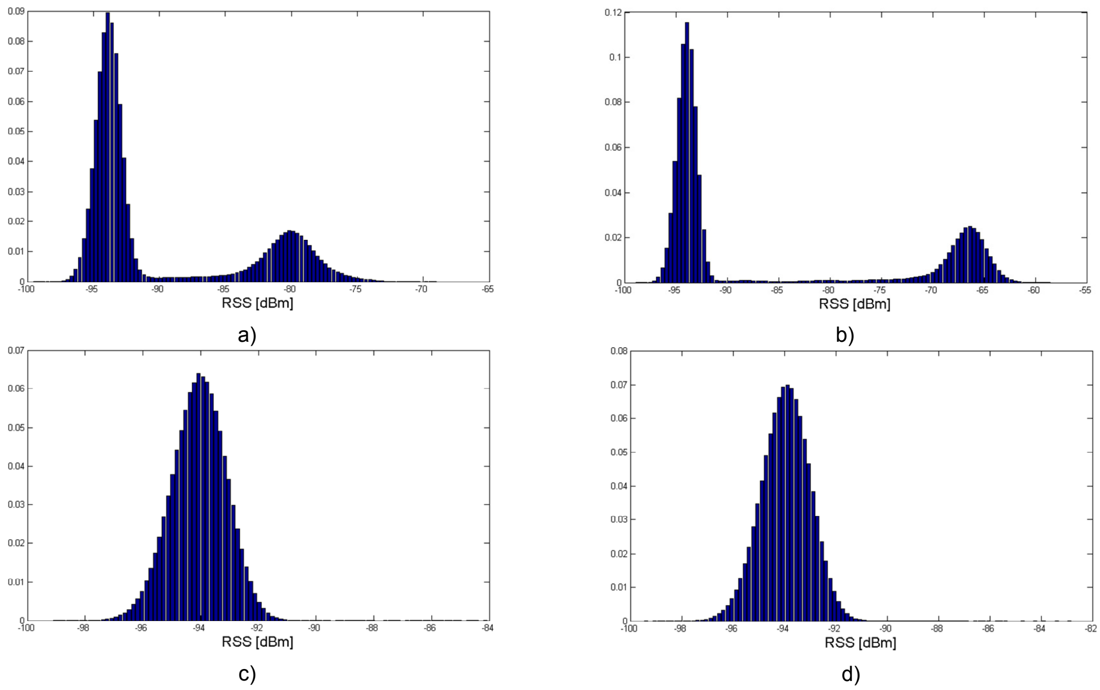
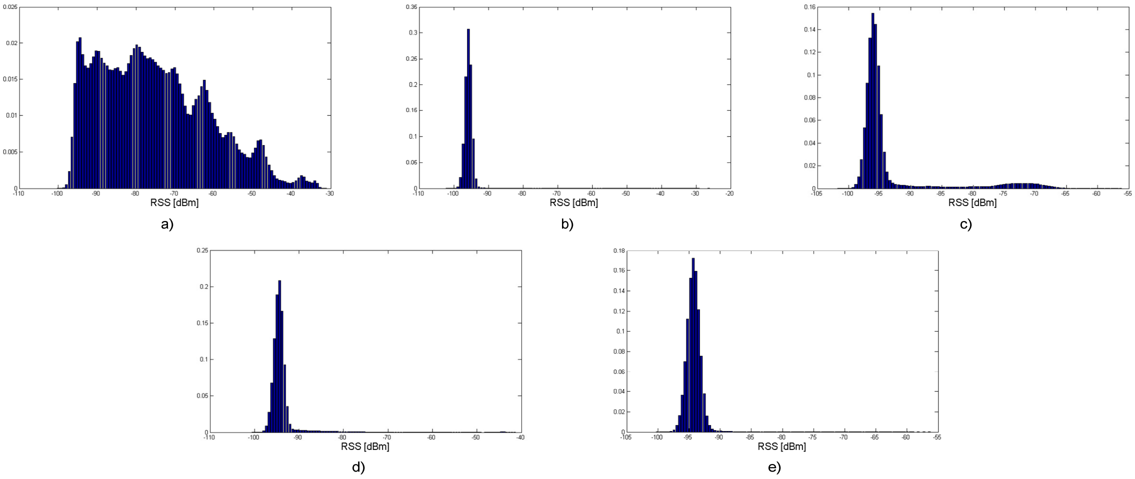


Figure 29: Histograms of RSS values for: a) UMTS downlink channel 1, b) UMTS downlink channel 2, c) UMTS uplink channel 1, d) UMTS uplink channel 2

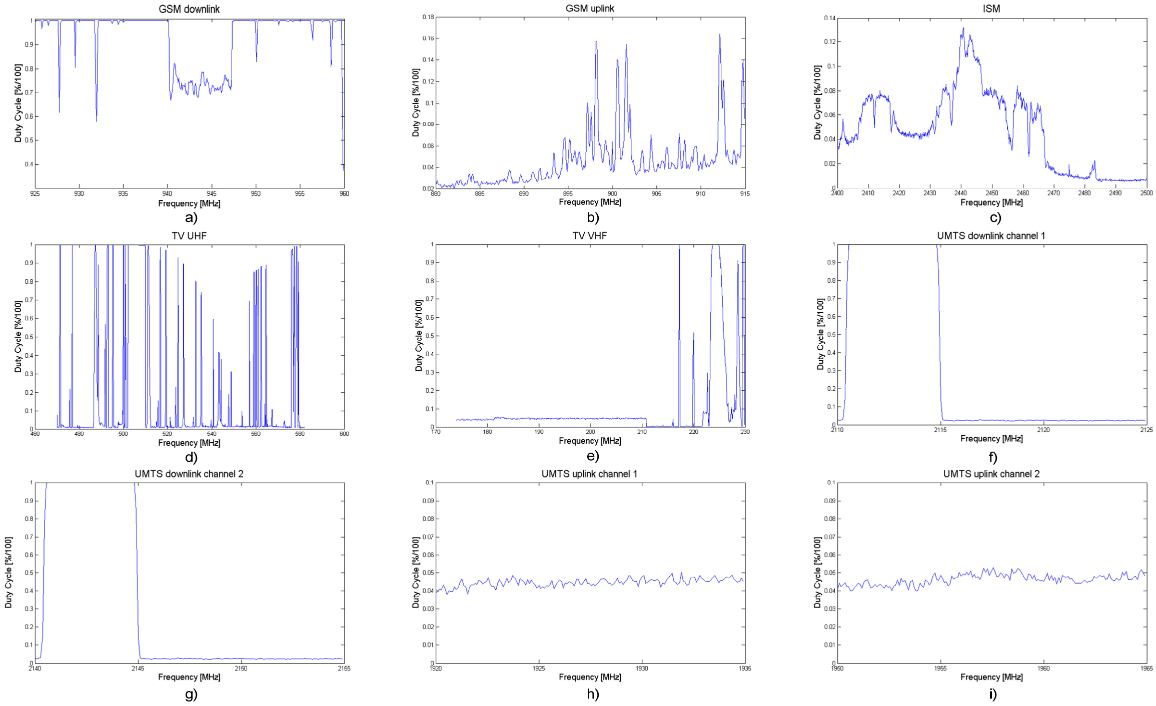


a)
b)
c)
d)
e)

Figure 30: Histograms of RSS values for:
a) GSM downlink, b) GSM uplink, c) TV - UHF, d) TV - VHF, e) ISM 2.4 GHz band.

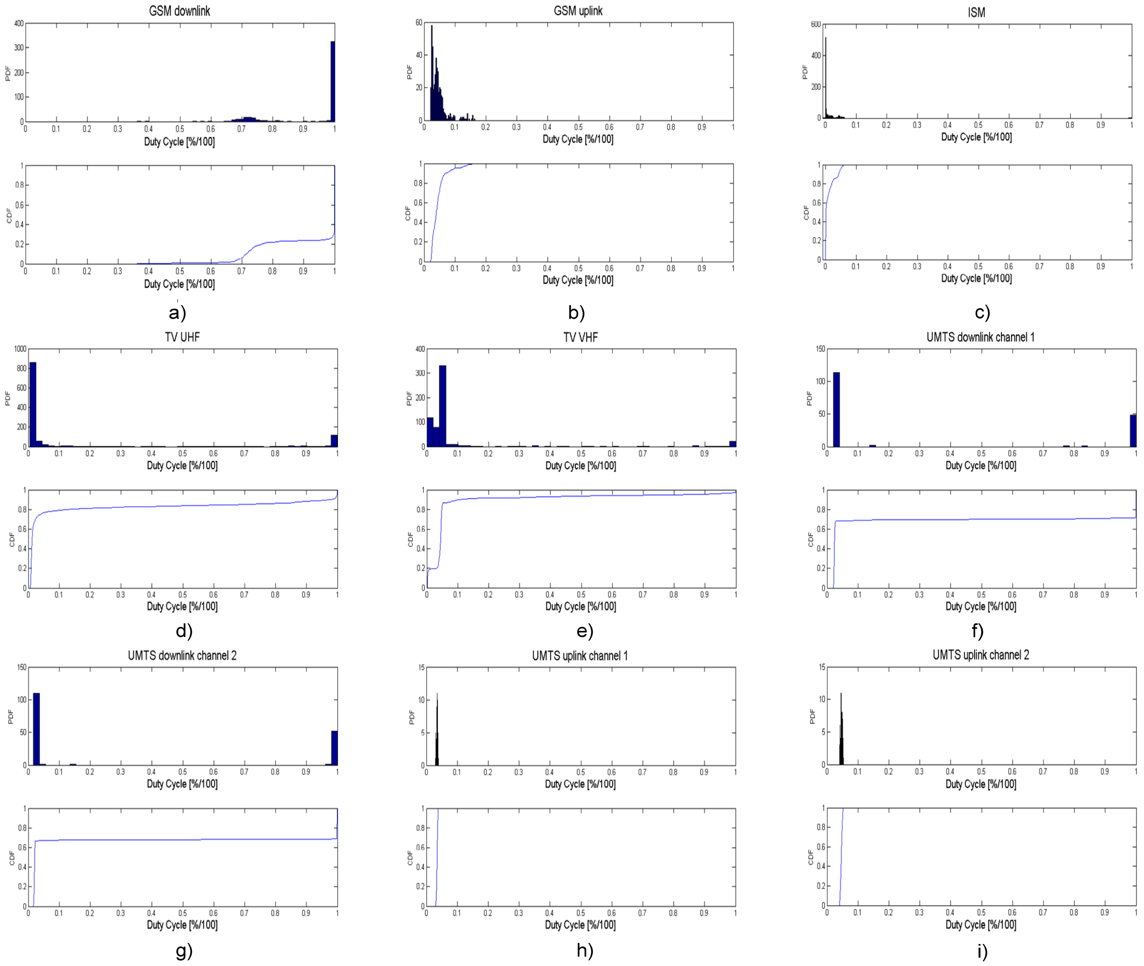
4.5.2 Calculation of duty cycle

For calculating the duty cycle the CCDF-based method using constant probability of false alarm was used, as introduced in Subsection 3.1.1. Figure 31 depicts the resulting duty cycles for each of the inspected bands, whereas Figure 32 depicts the empirical PDF and CDF of the duty cycle for each inspected band.



a)
b)
c)
d)
e)
f)
g)
h)
i)

Figure 31: Duty cycle per frequency for: a) GSM downlink band b) GSM uplink band c) ISM band d) UHF band e) VHF band f) UMTS downlink channel 1 g) UMTS downlink channel 2 h) UMTS uplink channel 1 i) UMTS uplink channel 2



a) b) c)

d) e) f)

g) h) i)

Figure 32: Empirical Probability Density Function (PDF) and Cumulative Distribution Function (CDF) for: a) GSM downlink band b) GSM uplink band c) ISM band d) UHF band e) VHF band f) UMTS downlink channel 1 g) UMTS downlink channel 2 h) UMTS uplink channel 1 i) UMTS uplink channel 2

These results can now be used further as a basis for deriving time and frequency domain models from the data set. By way of example, Figure 33 depicts a model fitting to the CDF of GSM 900 uplink duty cycle. Table 7 gives the estimated parameters of the fitted models using the MLE method.

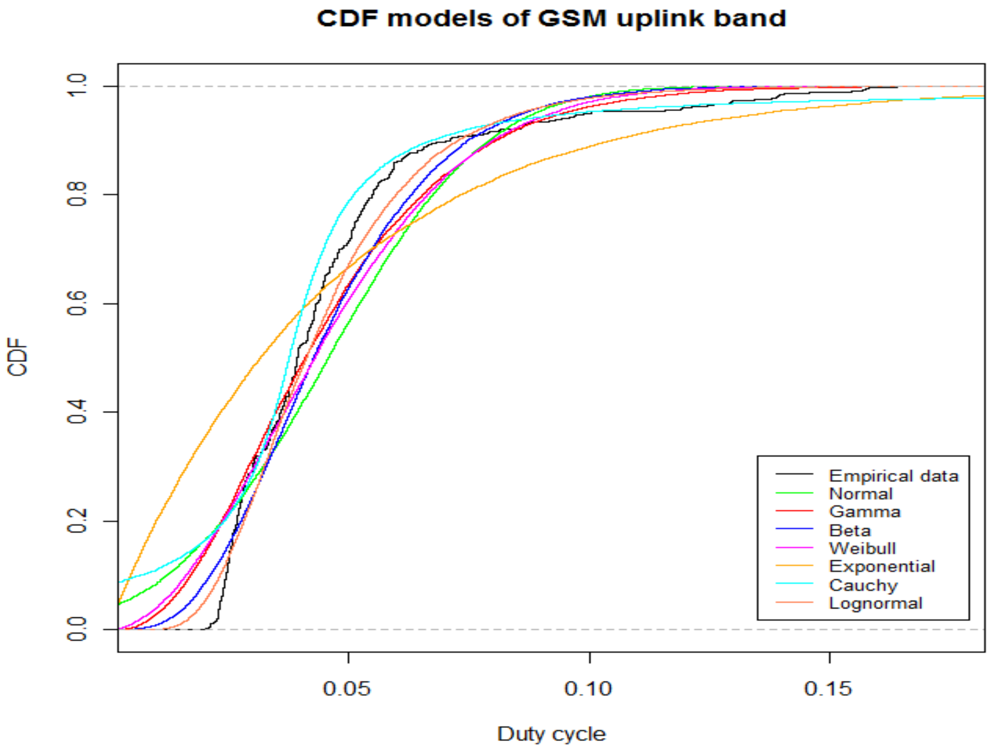


Figure 33: Model fitting to CDF of the duty cycle of the GSM 900 uplink.

Table 7: Estimated parameters for the empirical values of the GSM 900 uplink (-94.1 dBm threshold, Mean= 0.04581557, Variance=0.000674054)

Fitted distribution	**Parameter 1**	**Parameter 2**
Normal	0.04584261 (mean)	0.00067451 (variance)
Gamma	67.97017439 (rate)	3.11409232 (shape)
Beta	4.33681038 (shape 1)	90.06474288 (shape 2)
Weibull	1.91744360 (shape)	0.05201870 (scale)
Exponential	21.82664075 (rate)	N/A
Cauchy	0.03778847 (location)	0.00975010 (scale)
Lognormal	-3.19480934 (meanlog)	0.44166679 (sdlog)

Table 8 includes the Kolmogorov-Smirnov differences (*D*) for the models from Table 7. Based on these results, the duty cycle for the given GSM 900 uplink band can be best modelled with Lognormal function with parameters obtained with the MLE method.

Table 8: Kolmogorov – Smirnov goodness of fit values.

Fitted distribution	**Normal**	**Gamma**	**Beta**	**Weibull**	**Exponential**	**Cauchy**	**Lognormal**
Kolmogorov-Smirnov distance	0.168769	0.162095	0.104356	0.165438	0.371825	0.165844	0.071406

5 Results from the London Campaign

We shall now move on to discuss results from the London measurement campaign. Since the complexity of the resulting data set is even higher than for the two days campaigns, now covering over 150 locations and several disjoint frequency bands, we shall only give a small selection of results here. The complete data set will again be made publicly available in order to enable verification of results and more detailed analysis by the research community.

5.1 Measurement Approach and Regions Considered

Throughout the measurement campaign, the four RWTH "blue boxes" described in more detail in Sub-Section 2.4.2 have been employed. As discussed in the overview in Section 2, one of these was placed at the roof of the Strand building of the King's College London to carry out long term measurements akin to those carried out as part of the two days campaigns. The remaining three were moved into a number of different business and residential districts over the course of the campaign, which begun on Friday, 1st of July 2011, and lasted one week until the following Friday. These two days were reserved for transportation and setup, so dates of 2-7 July were used for the actual measurements.

Figure 34: Measurement locations around Oxford Circus.

First of the regions covered during the Saturday, 2nd of July 2011, was the Oxford Circus area, with the individual measurement locations illustrated in Figure 34 above. The measurements took place from roughly 11:00 until 17:00. The results obtained from these measurements should be representative of a heavily trafficked shopping and touristic area. Also during the measurements a parade took place through a region intersecting the Oxford Circus area, further increasing the density of pedestrians in the region.

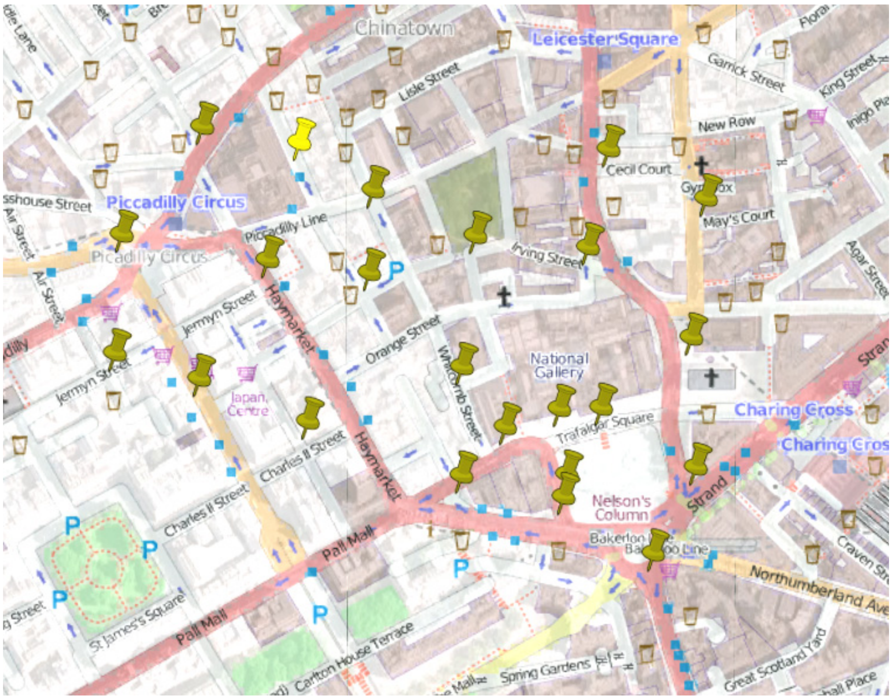


Figure 35: Measurement locations around Trafalgar Square.

For the evening of Saturday, from 18:00 until 21:00, following the measurements around Oxford Circus, the region around Trafalgar Square was covered. The individual measurement locations are shown in Figure 35. These results are also representative of densely populated touristic region, heavily used by the pedestrians also at this time. The area between Trafalgar Square, Piccadilly Circus and Leicester Square also has high concentration of restaurants, pubs and cafeterias, so usage of the ISM band might be expected to be heavier there compared to many of the other regions of London.

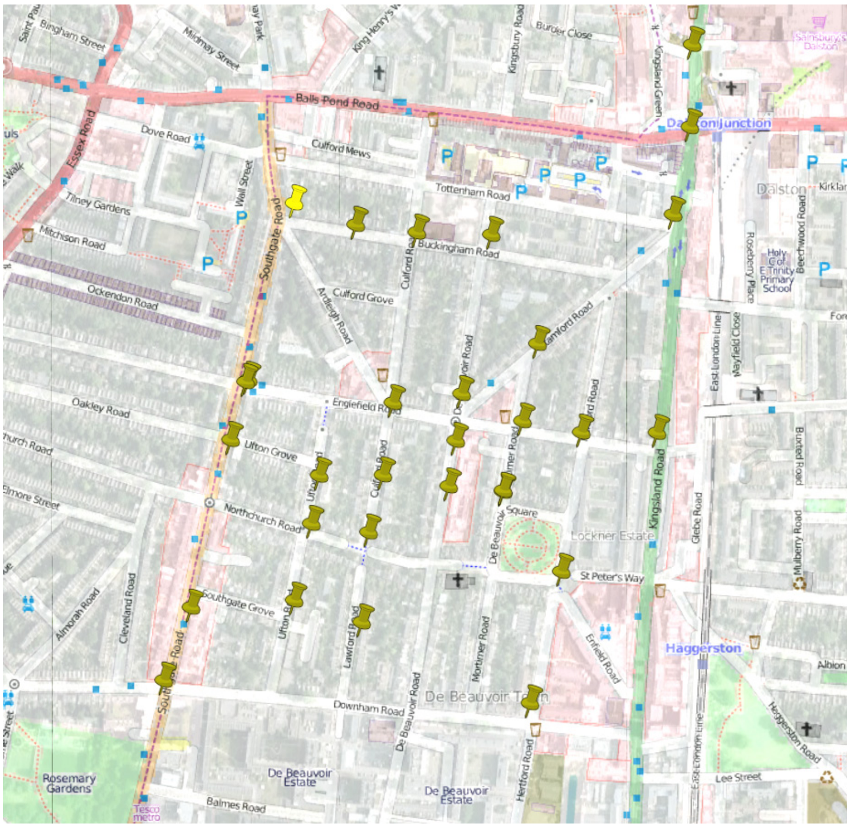


Figure 36: Measurement locations at a typical residential area.

We contrast in the following the results from these regions with those from more typical residential areas. Figure 36 illustrates locations of measurements carried out in a residential area between Islington and Hackney, in the northern region of the greater London area. This part of the measurement campaign took place on Monday, 4th of July, 2011, lasting from approximately 9:00 in the morning until 18:30 in the evening.

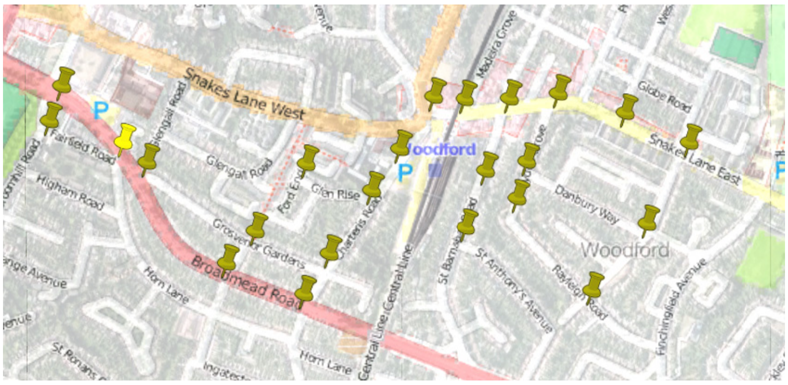


Figure 37: Measurement locations in Woodford,
a typical suburban region at the outskirts of London.

Finally, measurements in Woodford shown in Figure 37 are representative of results obtained in a suburban setting, with much lower population density compared to the residential area introduced above. The Woodford measurements took place on Wednesday, 6th of July, 2011, lasting from 11:45 until 16:00 in the afternoon. This part of the measurement campaign had somewhat shorter duration compared to the others due to the transit time needed to reach the measurement region from the location in downtown London in which the measurement equipment was stored overnight.

In addition to these four locations discussed in detail below, additional measurements were carried out in the regions around Wimbledon (Sunday, 3rd of July, corresponding to the men's final of the famous tennis tournament by the same name), in a typical business region surrounding the Liverpool Street railway station (Tuesday, 5th of July), and adjacent to the Heathrow airport (Thursday, 7th of July).

Finally, regarding the settings used in the measurements: the spectrum analyzers in the blue boxes were set to use a resolution bandwidth of 100 kHz in order to obtain sufficient frequency and time resolution especially the GSM-bands, and four distinct frequency bands were measured, covering the 900 MHz, 1800 MHz, 2.1 GHz cellular bands and the 2.4 GHz ISM band, respectively. The number of sweeps varied from one location to another, typical values ranging from 200-800, corresponding to overall measurement duration of 15-35 minutes per location. Average detector was used throughout. Times and locations of the individual measurement points were carefully logged into a measurement journal associated with each device, and for each measurement point a photograph was taken to assist in later geolocation of the point against high accuracy maps used for reference. This method was found yield better accuracy than GPS receivers used in some of the earlier measurement campaigns.

5.2 Usage of Cellular and ISM Bands at Selected Measurement Regions

We shall now give an overview of the results obtained from the London campaign, focusing on the four regions introduced above.

Figure 38 shows the distribution of the received power over the GSM900 downlink band for the Oxford Circus locations. We see that the shape of the distribution stays relatively constant from one location to another, while there are approximately 10 dB changes in the mean and median values from one location to another. The variability at individual locations is much higher than these changes between locations. This variability has multiple sources, ranging from the inherent differences in the received powers over different bands due to differences in distances to base stations and the used transmit powers, to fast fading caused especially by reflections from cars and other obstacles. There is also an element of heteroscedasticity in the data, meaning that the intra-site variability changes from one location to another, occasionally rather drastically at that (for example from location number 10 to location number 11).

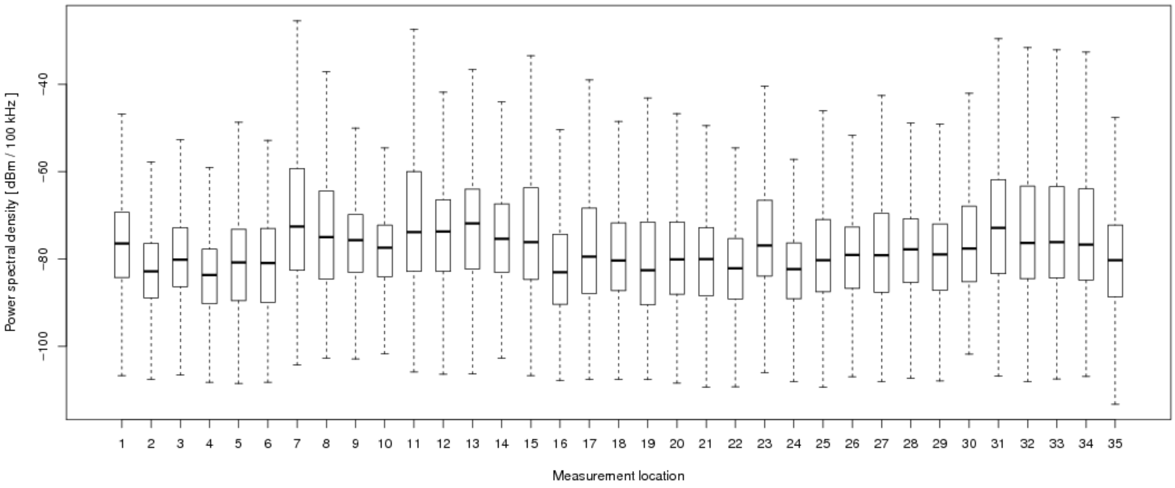


Figure 38: Distribution of received power on the
GSM900 downlink band over the Oxford Circus measurements.

These results can be contrasted against those shown in Figure 39, corresponding to measurements carried out in the residential area depicted in Figure 36. The variability both between the sites, as well as within an individual measurement location is significantly less in the residential area than in the Oxford Circus region. Finally, Figure 40 gives the results obtained in the suburban measurement campaign. Clearly the differences have reduced further still.

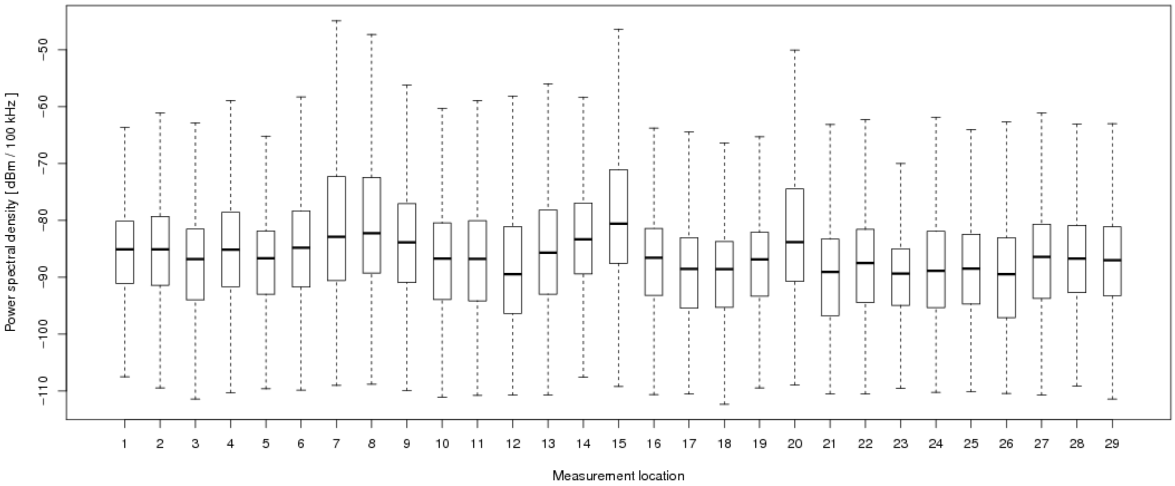


Figure 39: Distribution of received power on the GSM900 downlink band for the residential area.

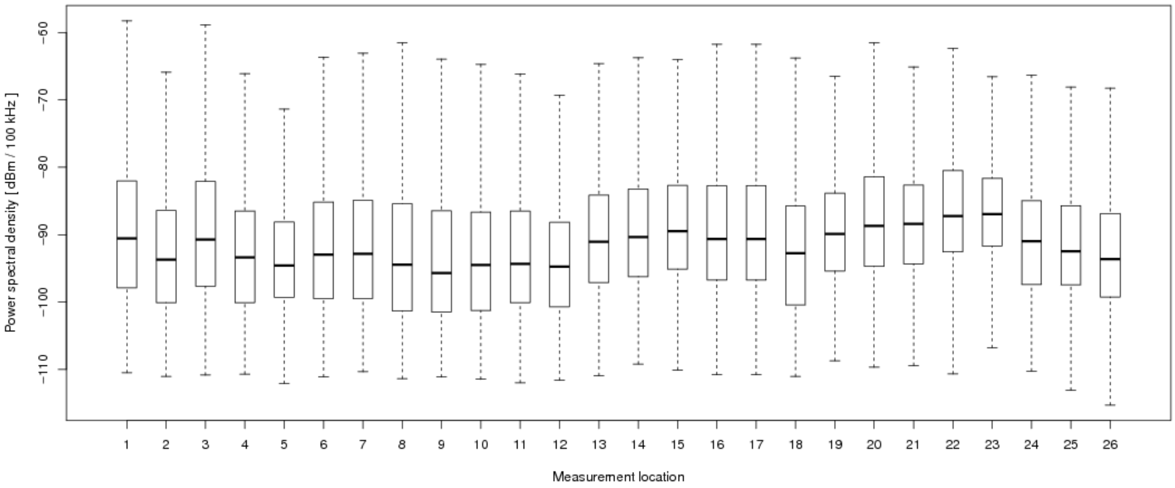


Figure 40: Distribution of received power on the GSM900 downlink band for the suburban region.

Similar trends can also be observed for the GSM900 uplink band, as can be seen by comparing Figure 41, giving the corresponding results for the Trafalgar Square measurements, against Figure 42, corresponding to the suburban case. The differences in variability are even higher than for the downlink case. This is easy to understand qualitatively, since in the regions densely utilized by pedestrians the likelihood of call being made during a measurement is much higher than for the suburban case. Note also that the average and median powers are extremely low in all cases. This is a natural consequence of the low transmit power and bursty nature of the uplink traffic (with few interesting exceptions given in the following subsection). We shall see in the following subsection that the latter plays a major role in high variability in the data as well.

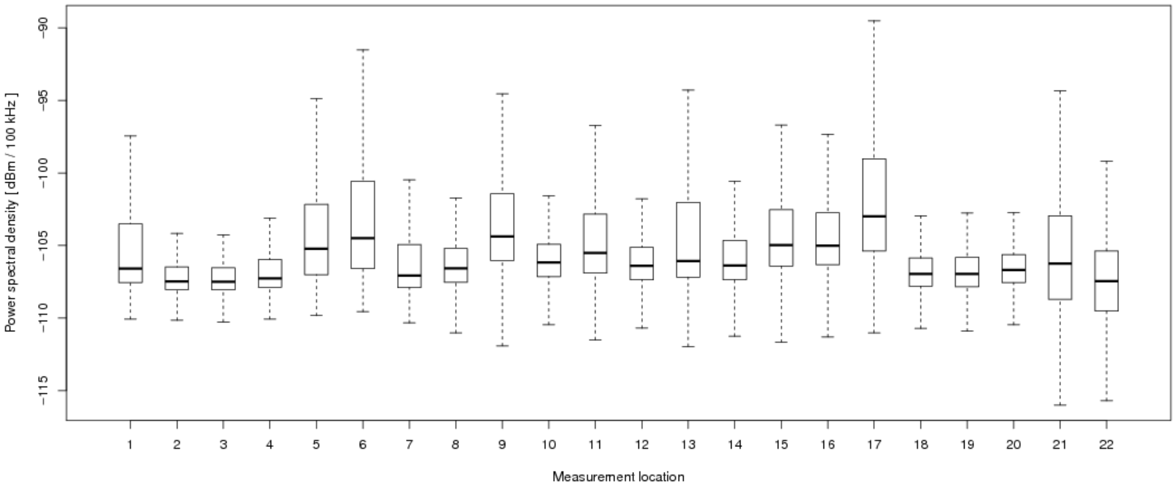


Figure 41: Distribution of received power on the
GSM900 uplink band over the Trafalgar Square measurements.

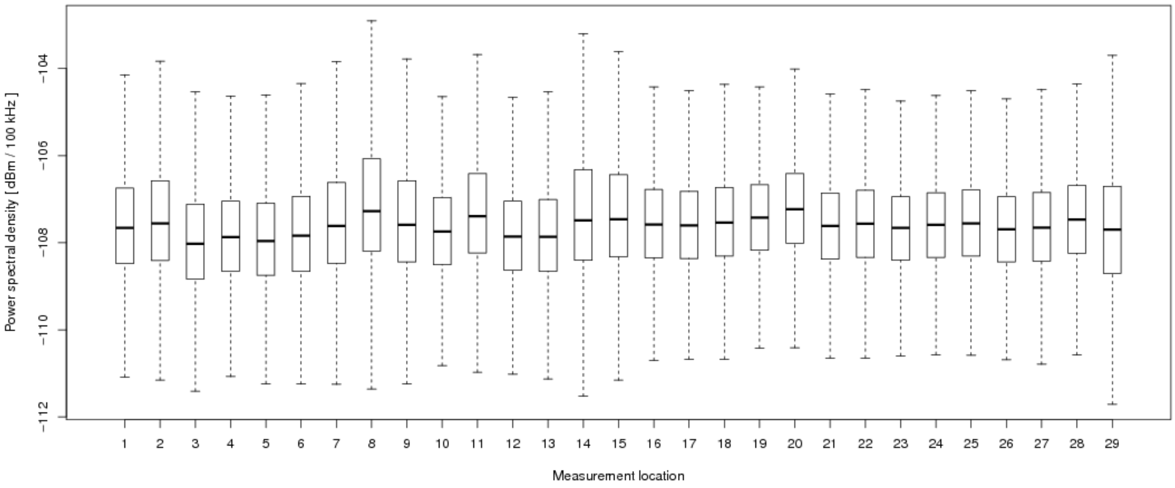


Figure 42: Distribution of received power on the GSM900 uplink band for the residential area.

5.3 Time-Frequency Behavior at Individual Locations

In order to gain more detailed understanding on the behavior of a particular frequency band during the measurements, we shall study here the full time-frequency characteristics for an individual measurement location. We have chosen at random one of the Oxford Circus cites to be used as a case study for this subsection. We begin our analysis with the 2.1 GHz UMTS downlink band, since this should have the most straightforward behavior of all cellular bands due to the “always on” nature of the UMTS Node Bs. Figure 43 shows the time-frequency plot or *spectrogram* for this band at our chosen measurement location. We see that the behavior is indeed completely as expected, with basically constant power received over each of the 5 MHz UMTS channels. Small variations within the band are due to combined effects of fast fading and changes in transmit power due to power control. We can also see that the 2.1 GHz UMTS downlink band is much more heavily occupied in London compared to many of the sites at which the two days measurements took place. This is, of course, very natural due to the high population density and large overall customer base in the area for cellular operators.

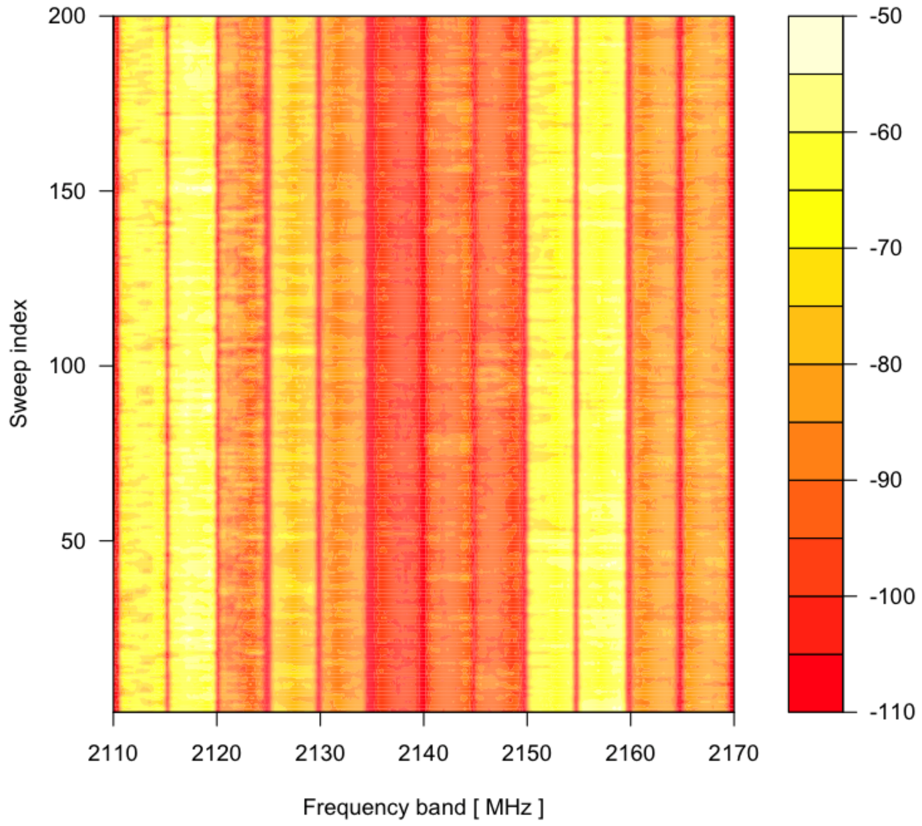


Figure 43: Spectrogram plot for an individual measurement location near Oxford Circus for the 2.1 GHz UMTS downlink band.

Let us now move on to the 900 MHz GSM downlink band, for which the spectrogram is shown in Figure 44. The TDMA nature of GSM is clearly visible in most of the spectrogram area, with individual unused timeslots resulting in dark red points interleaved between time-frequency regions with high received powers corresponding to ongoing transmissions. Overall the utilization of the band is clearly very high. The possibly surprising feature in the spectrogram can be seen in the frequency band between 930 and 935 MHz, where a continuous transmission is clearly taking place. This corresponds in fact to one of the first deployments of UMTS in the 900 MHz band.

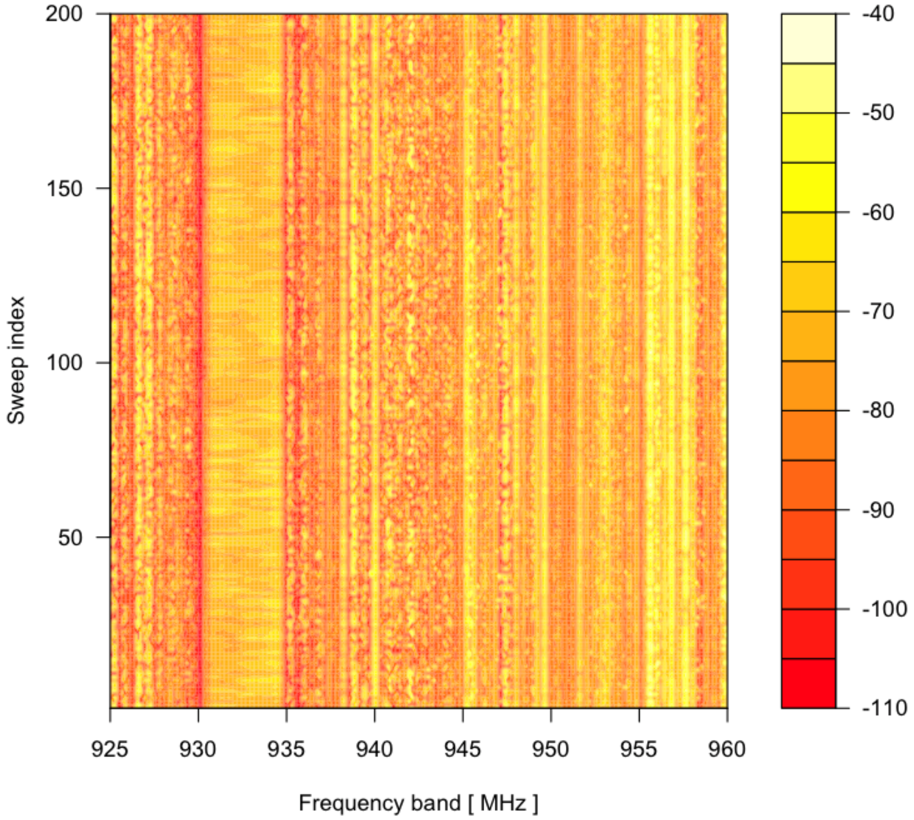


Figure 44: Spectrogram plot for an individual measurement location near Oxford Circus for the GSM900 downlink band.

The spectrogram for the uplink band of GSM900 shown in Figure 45 shares some of the characteristics seen in the downlink case. However, the level of burstiness is much higher, since how the TDMA nature is combined with the randomness induced by the call behavior and mobility of the users. Interestingly the usage of UMTS in this band is also clearly visible. The high variability of received powers encountered already in Figure 41 can also be clearly seen visually. However, we also see that the distributional plots used in the previous subsection do not capture very well just how heavy-tailed the distribution of the received power is in the uplink case. Only utilizing statistics that capture the extreme values suffice in such situations to reveal such behavior.

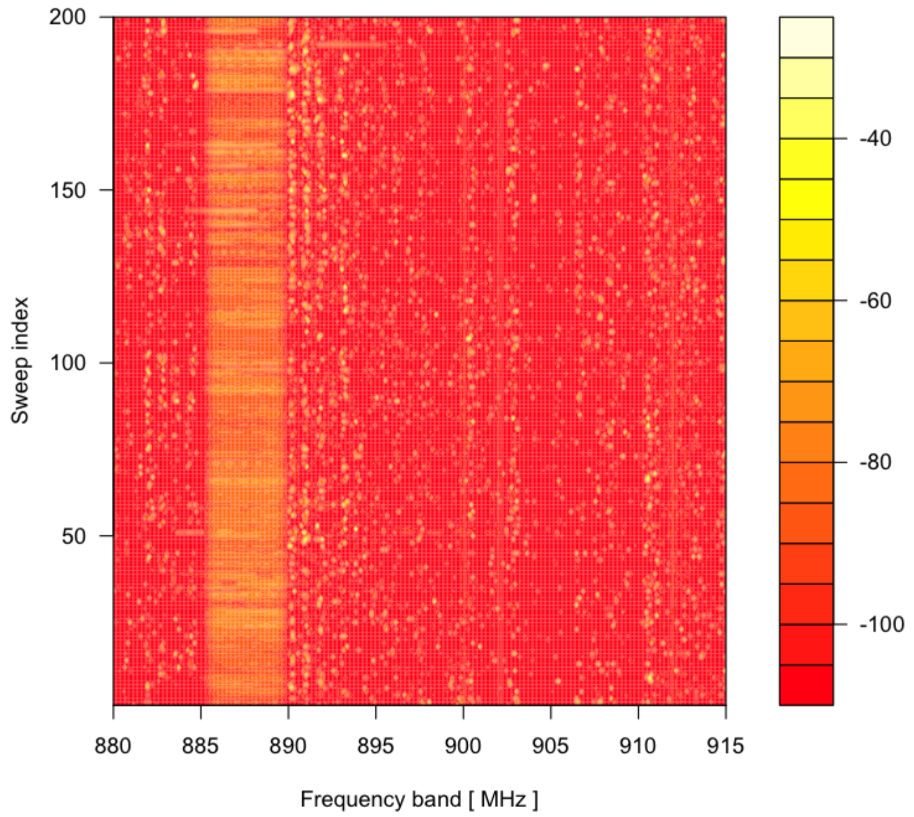


Figure 45: Spectrogram plot for an individual measurement location near Oxford Circus for the GSM900 uplink band.

We conclude this case study with the spectrogram of the 2.4 GHz ISM band shown in Figure 46. Clearly the usage of the band is again highly bursty, with the exception of few almost continuously transmitting narrowband sources around 2450 MHz. The three most commonly used Wi-Fi channels can also be seen clearly. Note that due to the sweeping operation of the spectrum analyzer and the very short duration of a typical Wi-Fi frame, individual frames are not seen as the whole 22 MHz channel as being occupied at the same time. Due to such effects, wideband transmitters can easily be mistaken as larger number of narrowband ones.

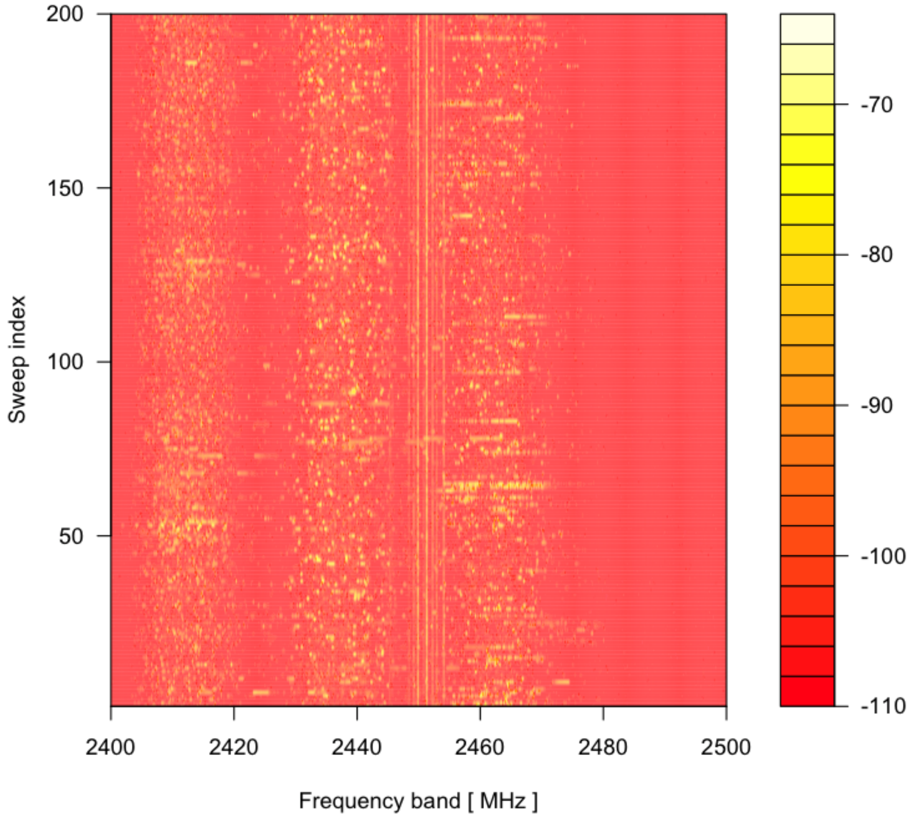


Figure 46: Spectrogram plot for an individual measurement location near Oxford Circus for the 2.4 GHz ISM band.

6 Additional Models and Results

In this section we present selected additional outcomes from our measurements and modeling work. As discussed in the previous sections as well, this is not intended to be a comprehensive overview due to the large amount of efforts ongoing and already finalized. Instead, we aim at to highlight novel contributions and results that we believe to also have practical value for applications, or that provide intuition and understanding on the interpretation of spectrum measurement results.

We begin by discussing the application of the spatial modeling techniques introduced in Subsection 3.2 on spatial measurement data sets, and make the link to coverage prediction and estimation for wireless communication systems, such as cellular networks. We then extend the duty cycle based modeling framework introduced for time and frequency domain analysis in Subsection 3.1 to address some of the shortcomings of the classical stationary Markov and Semi-Markov models used until now. We also make the link to the spatial case by discussing the interplay of propagation and detection, and its influence on the perceived duty cycle. Finally, we discuss measurements and modeling related to indoor propagation environment from radio resource management point of view, as opposed to the signal processing viewpoint that is highly prevalent in the literature.

6.1 Spatial Spectrum Use Models and Prediction Accuracy

Our objective in this subsection is to show through a detailed case study how the spatial modeling techniques from Subsection 3.2 can be applied to spatial measurement data sets, and how the resulting models can be applied for, e.g., coverage prediction of wireless communication systems. The consortium has gathered sizable spatial measurement data sets, primarily using the "blue box" mobile measurement platforms of RWTH. For the present case study we utilize a data set collected at downtown Aachen during the summer of 2010. The data set consist of approximately 140 measurement locations, selected to form a dense grid at the downtown of Aachen, and then to become slightly more sparse towards the outer regions of the town. Within the measurement region there is a large number of Wi-Fi access points and cellular base stations belonging to the four major operators in Germany. Measurement settings chosen are analogous to those used for the London measurements discussed above. In particular, resolution bandwidth of 100 kHz was used with minimal sweeptime over four selected frequency bands covering the cellular services and the 2.4 GHz ISM band. At each measurement location 200 sweeps were recorded on each of the bands using the average detector.

Figure 47 shows an overview of the obtained data set for selected frequencies. As we target coverage prediction types of applications in the following, all the data process and analysis was carried out on per-operator basis, as opposed for the whole band simultaneously as was done in earlier studies. The figure shows the distribution of the measurement locations, graphical illustration of the mean received power values at those locations, together with the marginal distribution of the arising data set.

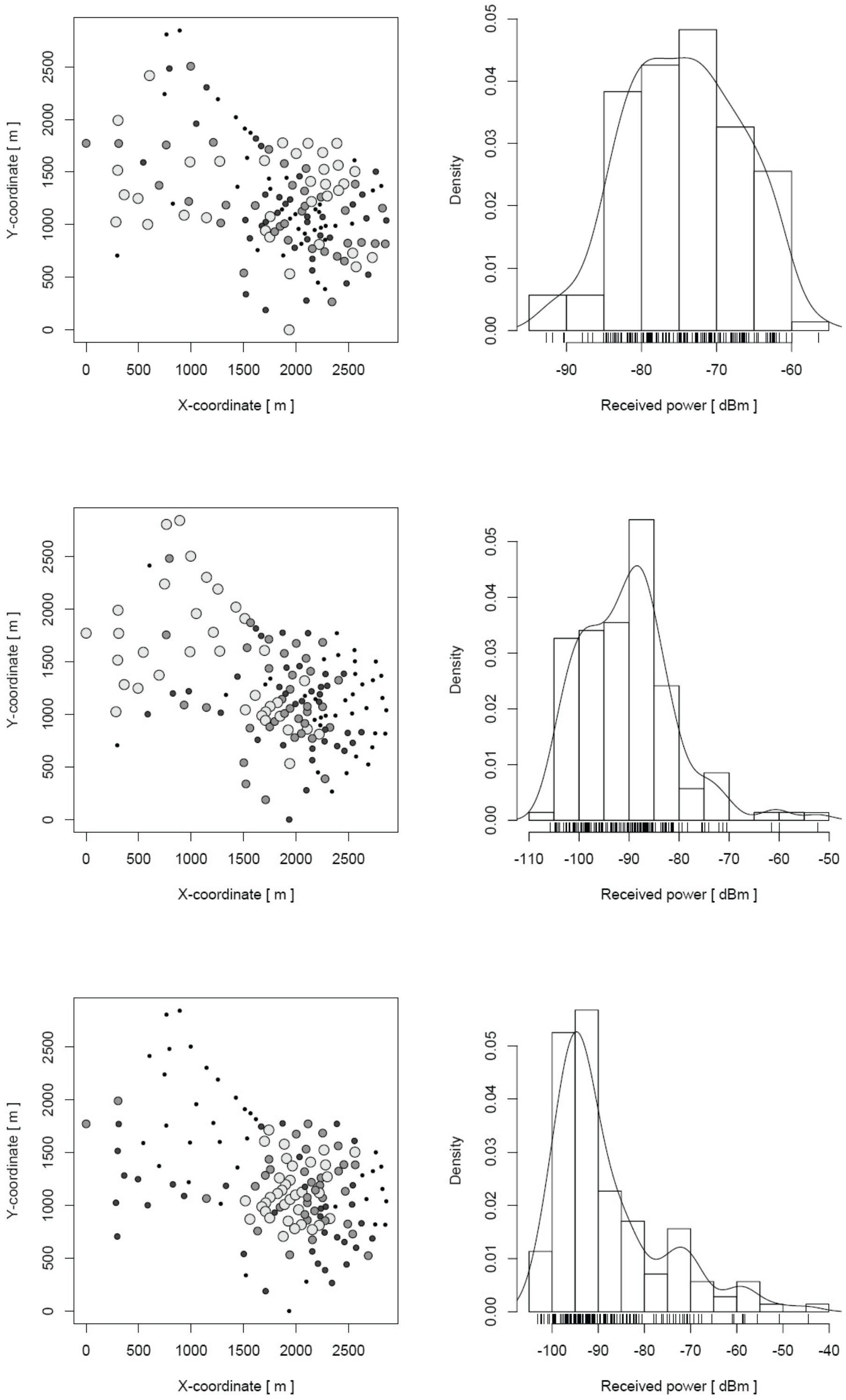


Figure 47: Results from spatial spectrum use measurements for the E-Plus GSM1800 network (top), T-Mobile GSM900 network (middle) and the O2 GSM900 network (bottom). Leftmost panels illustrate the spatial structure of the data sets while the rightmost panels show the marginal distributions of the data.

Based on these results, models for the correlations in the data were developed using the semivariogram fitting approach given in Subsection 3.2. The selection of the considered semivariogram models was done based on our earlier experiences documented in [31], wherein the definitions of those models can be found as well.

Table 9 summarizes the obtained semivariogram models, giving the form of the semivariogram that resulted in the best fit according to the weighted least squares criterion, as well as the model parameters and the estimated effective correlation range. In few instances (namely for the E-Plus GSM 900 and Vodafone GSM 1800, and to lesser extent for O2 GSM 900) the estimated effective ranges were very large, in fact much larger than the measurement region in itself. This indicates that for corresponding networks larger measurement region should be chosen in order to obtain reliable parameter estimates. The corresponding models are illustrated in Figure 48.

Table 9: Semivariogram model parameters for the four spectrum analyzer measurement campaigns for different cellular operators in downtown Aachen.

Network	Variogram model	Nugget	Sill	Range	κ	Eff. range	Quality of fit
DB GSM 900	Exponential	15.72	82.69	648.61	0.50	1943.05	9.3×10^4
E-Plus GSM 900	Matern	10.36	449.35	1710.09	2.12	9414.83	1.1×10^4
O2 GSM 900	Matern	6.87	1031.08	108.17	111.34	3959.80	6.7×10^5
T-Mobile GSM 900 (1)	Spherical	0.63	114.59	331.29	1.00	331.29	7.1×10^5
T-Mobile GSM 900 (2)	Spherical	0.00	107.66	332.16	1.00	332.16	4.3×10^5
T-Mobile GSM 900 (3)	Cauchy	26.02	74.62	5259.55	50.07	1306.01	1.6×10^5
Vodafone GSM 900 (1)	Spherical	0.00	165.60	389.25	1.00	389.25	6.7×10^6
Vodafone GSM 900 (2)	Spherical	12.68	82.78	255.90	1.00	255.90	1.1×10^5
Vodafone GSM 900 (3)	Cauchy	0.00	124.20	166.40	1.00	725.33	2.1×10^5
E-Plus GSM 1800	Spherical	9.88	44.64	411.88	1.00	411.88	7.1×10^4
O2 GSM 1800	Cauchy	11.15	58.49	233.30	1.00	1016.93	2.6×10^5
T-Mobile GSM 1800	Spherical	6.11	23.75	347.62	1.00	347.62	3.2×10^5
Vodafone GSM 1800	Matern	0.00	2354.15	231.75	108.26	8366.53	1.6×10^5

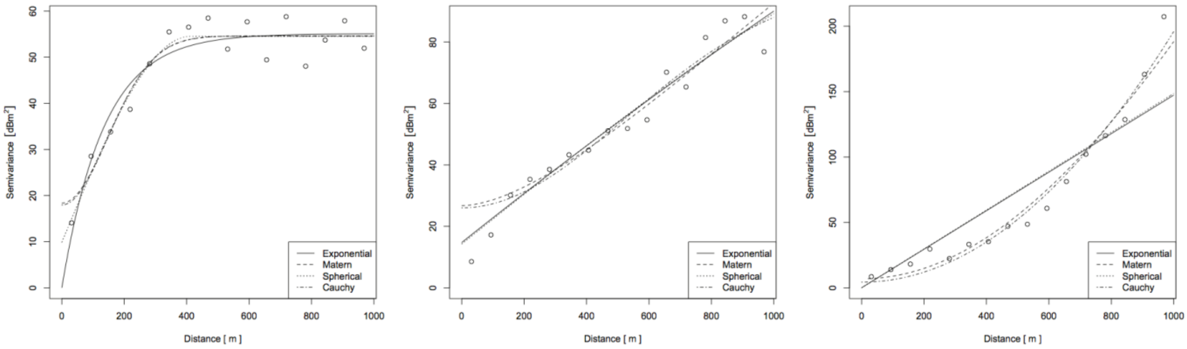


Figure 48: Semivariogram fits for the E-Plus (left), T-Mobile (middle) and O2 (right) data sets.

We shall now apply the developed spatial correlation models to *prediction* of spectrum use over space towards regions in which measurement data is not directly available. For the kriging estimator this consists of, as shown above, solving for the weights *w* in the expression

$$\hat{Z}(y) = \sum\nolimits_i w_i(x_i) Z(x_i).$$

from the matrix equation

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(x_1,x_1) & \cdots & \gamma(x_1,x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n,x_1) & \cdots & \gamma(x_n,x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(x_1,y) \\ \vdots \\ \gamma(x_n,y) \\ 1 \end{pmatrix}$$

using the semivariogram models given in Table 9. Recall also that the corresponding theoretical estimator variance is then given by

$$\sigma^2(y) = \sum\nolimits_i w_i(y) \gamma(x_i, y) + \mu.$$

Figure 49 illustrates this process for the E-Plus GSM1800 spatial data set analyzed in the previous chapter. Using the fitted semivariogram model kriging interpolation was applied to the data set, resulting in the estimate of the mean PSD surface shown on the left panel of the figure. As per usual properties of the kriging estimator, the interpolated surface is faithful to the data in the sense that each data point is also point of the interpolated surface, and at regions far enough of any data point the interpolated surface plateaus towards the mean of the estimates. Another useful property of the kriging estimator is that the variance or standard deviation of the estimates is available theoretically. These are shown in the right panel of the figure. Near the locations at which measurements were made the standard deviation approaches zero, whereas far away it asymptotically approaches the standard deviation of the underlying random field. Both of these are, of course, very natural properties for any estimator operating on the raw measurement data directly. More accurate predictions could potentially be made using techniques utilizing further information about the nature of the problem, such as compressive sensing motivated techniques. Some of these were introduced in earlier deliverables in the project, especially in D4.2, and found to yield promising performance based on simulated data sets. Such approaches are naturally interesting, but also do often result in more modeling parameters, the estimation errors of which can propagate in unpredictable manner through the data processing chain. Further work is clearly called for to better understand the robustness and performance tradeoffs in these two types of estimators.

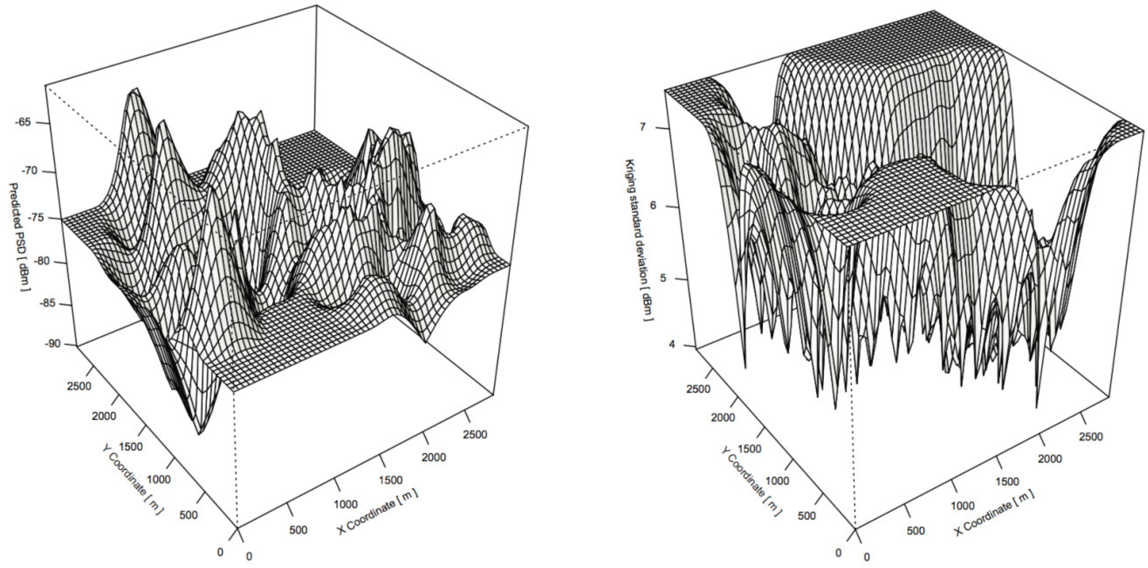


Figure 49: Kriged estimates of the mean PSD field (left) and the standard deviations of the estimator (right) for the E-Plus GSM1800 network in downtown Aachen.

Another question of interest is how to characterize the performance of such estimators. The theoretical results on estimator variance do give an indication of the expected performance, but are ultimately derived under assumptions that might not be valid for the data set they are applied on. For example, we have seen in the previous chapter that the mean PSD fields are rarely completely Gaussian, commonly having a heavier upper tail. A number of estimator types can be evaluated using cross-validation statistics, or “leave one out” approach. This is based on estimating the value of the field at a location where a measurement is available, *without* using that particular measurement point in the estimation process (otherwise for, e.g., the kriging estimator would always have zero error). The process is then repeated for all the measurement points, and estimation errors are put together either into a (marginal) empirical probability distribution, or treated as samples of another associated random field. Results for this process in the case of the kriging estimator are shown in Figure 50 for the E-Plus data set. The marginal distribution of the estimation errors follows well the expected normal distribution, as also more carefully shown in figure 10.9. However, as can be seen from the spatial structure of the estimation error the use of the marginal distribution alone might result in too pessimistic estimates, as much of the contribution at the tails is induced by points at the edge of the measurement areas, which would be expected to be difficult to estimate accurately in any case.

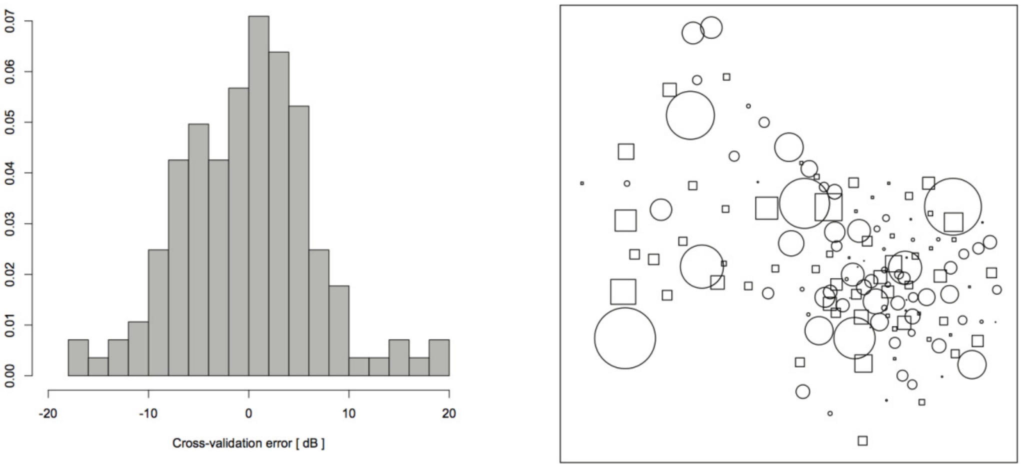


Figure 50: The marginal (left) and spatial (right) distributions of the "leave one out" cross-validation errors for the E-Plus GSM1800 data set.

6.2 Extended Duty Cycle Models

In this section we introduce extensions to the classical duty cycle based models presented in Section 3. We first consider different approaches towards introducing non-stationarities into Markov and Semi-Markov models as means to account for changes in the perceived spectrum use caused by, for example, the diurnal cycle. We then discuss how to extend the Markov and Semi-Markov modeling frameworks to introduce correlations between successive ON and OFF period lengths, a phenomenon observed in [22] and with first models introduced in [25]. Finally, we study the interplay between the duty cycle based models and spatial models by analyzing the influence of the propagation environment on the perceived spectrum occupancy.

6.2.1 Non-stationary DTMC models

For channels with varying load patterns the DC changes over time, meaning that the probabilities of the transition matrix $\mathbf{P}$ are also time-variant. In such a case, a *non-stationary* or *time-inhomogeneous* DTMC needs to be considered, with a time-dependent transition matrix:

$$\mathbf{P}(t) = \begin{bmatrix} 1-\Psi(t) & \Psi(t) \\ 1-\Psi(t) & \Psi(t) \end{bmatrix}$$

where $t = t_k = kT_s$. In the stationary case Ψ represents a constant parameter. However, in the non-stationary case of the equation above, Ψ(t) represents a time-dependent function that needs to be characterised in order to characterise the complete DTMC channel model in the time domain [11]. The measurement results discussed above indicate the existence of two well-defined types of channel load variation patterns, namely patterns with an important and remarkably predominant deterministic component (such as arising from the diurnal patter) and patterns where the carried load appears to vary following a random behaviour (for example, as induced by local temporal fluctuations in the network load).

Below time-dependent duty cycle models for Ψ(t), following deterministic and stochastic approaches are presented.

Deterministic duty cycle models

In certain cases, the load variation patterns of primary radio channels are characterised by a predominant deterministic component arising from social behaviour and common habits. These examples correspond to cellular mobile communication systems, namely the Global System for Mobile communications (GSM) operating in the 900-MHz band and its counterpart in the 1.8-GHz band, the Digital Cellular System (DCS). Similar patterns can also be present in other cellular technologies such as the TErrestrial Trunked Radio (TETRA) system.

The load variation pattern of a cellular mobile communication system can be described by means of Auto-Regressive Integrated Moving Average (ARIMA) time series models [12]. This section presents an alternative modelling approach based on the observation that the time evolution of Ψ(t) overtime periods of certain length exhibits a clear and predominant deterministic component. In particular, these techniques are appropriate for scenarios for which the variation pattern of Ψ(t) is periodic with a period of one day (24 hours) and a slightly different shape between weekdays and weekends due to the lower traffic load normally associated with weekends. Two different shapes for Ψ(t) can be identified. The first shape type is normally present in channels with low/medium loads (average DCs), while the second shape type is more commonly observed in channels with medium/high loads as in the example of [13].

For channels with low/medium load the shape of Ψ(t) can be approximated by the summation of M bell-shaped exponential terms centred at time instants τ_m, with amplitudes A_m and standard deviation σ_m as in

$$\Psi(t) \approx \Psi_{min} + \sum_{m=0}^{M-1} A_m e^{-\left(\frac{t-\tau_m}{\sigma_m}\right)^2}, \quad 0 \le t \le T$$

where $\Psi_{min} = \min\{\Psi(t)\}$ and T is the time interval over which Ψ(t) is periodic (i.e., one day).

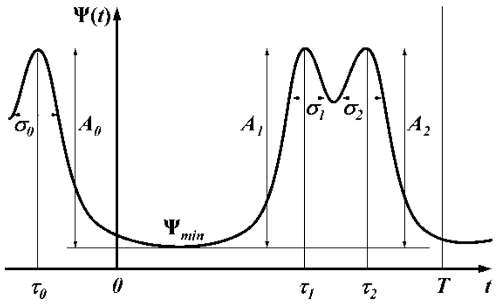


Figure 51: Parameters of the deterministic duty cycle model for low/medium loads.

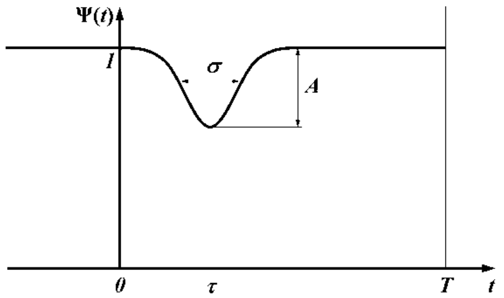


Figure 52: Parameters of the deterministic duty cycle model for medium/high loads.

The analysis of empirical data indicates that Ψ(t) can accurately be described by means of $M=3$ terms with τ_1 and τ_2 corresponding to *busy hours* and $\tau_0=\tau_2-T$ as illustrated in Figure 51. Moreover, the approximations $A_0=A_1=A_2=A$ and $\sigma_0=\sigma_1=\sigma_2=\sigma$ are acceptable without incurring in excessive errors, which simplifies the model:

$$\Psi(t)\approx\Psi_{min}+A\sum_{m=0}^{M-1}e^{-\left(\frac{t-\tau_m}{\sigma}\right)^2},\quad 0\le t\le T$$

Notice that A determines the average value of $\Psi(t)$ in the time interval $[0,T]$, denoted by $\overline{\Psi}$, and the DC model can therefore be expressed (see [11]) as

$$\Psi(t)\approx\Psi_{min}+\frac{2T\left(\overline{\Psi}-\Psi_{min}\right)}{\sigma\sqrt{\pi}}\cdot\frac{f_{\mathrm{exp}}^{l/m}(t,\tau_m,\sigma)}{f_{\mathrm{erf}}^{l/m}(T,\tau_m,\sigma)}$$

where $\overline{\Psi}\ge\Psi_{min}$ and:

$$f_{\mathrm{exp}}^{l/m}(t,\tau_m,\sigma)=\sum_{m=0}^{M-1}e^{-\left(\frac{t-\tau_m}{\sigma}\right)^2}$$
$$f_{\mathrm{erf}}^{l/m}(T,\tau_m,\sigma)=\sum_{m=0}^{M-1}\left[\mathrm{erf}\left(\frac{\tau_m}{\sigma}\right)+\mathrm{erf}\left(\frac{T-\tau_m}{\sigma}\right)\right]$$

For channels with medium/high load (see Figure 52), the shape of $\Psi(t)$ can be approximated by an expression based on a single bell-shaped exponential term centred at time instant τ, with amplitude A and standard deviation σ:

$$\Psi(t)\approx1-Ae^{-\left(\frac{t-\tau}{\sigma}\right)^2},\quad 0\le t\le T$$

where T is the time interval over which $\Psi(t)$ is periodic (i.e., one day).

As in the previous case, A determines the average value of $\Psi(t)$ in the time interval $[0,T]$ and the DC model can therefore be expressed, [11], as:

$$\Psi(t)\approx1-\frac{2T\left(1-\overline{\Psi}\right)}{\sigma\sqrt{\pi}}\cdot\frac{f_{\mathrm{exp}}^{m/h}(t,\tau,\sigma)}{f_{\mathrm{erf}}^{m/h}(T,\tau,\sigma)}$$

where:

$$f_{\exp}^{m/h}(t,\tau,\sigma) = e^{-\left(\frac{t-\tau}{\sigma}\right)^2}$$

$$f_{\text{erf}}^{m/h}(T,\tau,\sigma) = \text{erf}\left(\frac{\tau}{\sigma}\right) + \text{erf}\left(\frac{T-\tau}{\sigma}\right)$$

The DC ranges within which each model is valid depends on the particular set of values selected for the configuration parameters. As a rough approximation, the DC model for low/medium loads can be valid for average DC values up to $\overline{\Psi} \approx 0.70$, while the DC model for medium/high loads can be valid down to $\overline{\Psi} \approx 0.45$. Any set of values for the model parameters can be valid as long as $\Psi(t)$ is confined within the interval $[0,1]$.

The capability of the non-stationary DTMC model along with the deterministic DC models to reproduce the statistical distributions of busy and idle periods in real channels is illustrated in Figure 53 and Figure 54, including the distributions of busy and idle periods obtained by means of simulation with the non-stationary DTMC channel model. As it can be appreciated, the deterministic DC models are able to closely follow and reproduce the deterministic component of Ψ(t) in the time domain and, as a result, the overall model is able to reproduce not only the mean DC of the channel, $\overline{\Psi}$, but also the statistical properties of busy and idle periods, which is not the case of the stationary DTMC model.

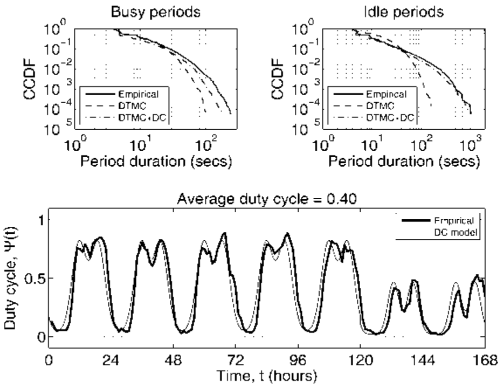


Figure 53: Empirical and DTMC-simulated distributions of busy and idle periods along with DC time evolution for a DCS 1800 downlink channel.

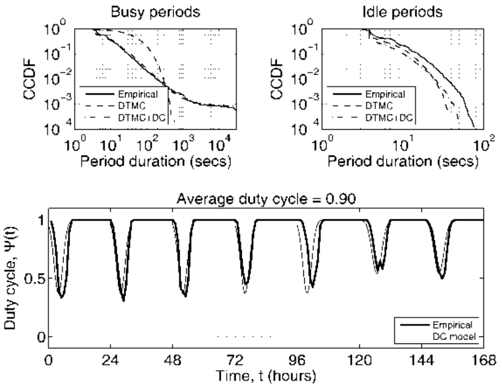


Figure 54: Empirical and DTMC-simulated distributions of busy and idle periods along with DC time evolution for a DCS 1800 downlink channel.

The presented DC models are envisaged to reproduce the deterministic pattern normally observed in cellular mobile communication systems such as E-GSM 900 and DCS 1800, which may also be present in other systems. Nevertheless, this does not imply that the model is always applicable to such type of systems. If the system is studied over relatively short time periods (e.g., a few hours), social behaviour and external events, which may not be easily predicted, may have a significant short-term impact on the channel usage. This may cause the deterministic component of Ψ(t) to loss importance with respect to the random component and, as a result, the occupancy of a channel may experience high and unpredictable variations (e.g., see [14]). In such a case, deterministic DC models may be no longer valid and stochastic modelling approaches may constitute a more appropriate alternative.

Stochastic duty cycle models

The traffic load experienced in a radio channel is the consequence of a significant number of random factors such as the number of incoming and outgoing users, the resource management policies employed in the system, and so forth. Therefore, the channel usage level, represented by means of Ψ(t), is itself a random variable. As such, Ψ(t) can be described and characterised by its probability density function (PDF). Recall from Section 3 that the PDFs of Ψ(t) in real systems over a given band can often be accurately fitted with the beta distribution [15] and the Kumaraswamy distribution [16]. The PDF for the former is given by:

$$f_x^B(x;\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in (0,1)$$

where $\alpha > 0$ and $\beta > 0$ are shape parameters and $B(\alpha,\beta)$ is the beta function:

$$B(\alpha,\beta) = \int_0^1 z^{\alpha-1}(1-z)^{\beta-1} dz$$

while the PDF for the latter is given by:

$$f_x^K(x;a,b) = abx^{a-1}(1-x^a)^{b-1}, \quad x \in (0,1)$$

where $a > 0$ and $b > 0$ are shape parameters.

The beta distribution is a well-known and widely used distribution that can be found in many popular software simulation packages, thus facilitating the implementation of the stochastic DC model in simulation tools. However, it might present some difficulties in analytical studies due to the complex expression of its PDF. The Kumaraswamy distribution is similar to the beta distribution, but much simpler to use in analytical studies due to the simpler closed form of its PDF [17]. Therefore, while the former may be more appropriate for simulations, the latter may be more convenient for analytical studies.

Both distributions can be configured to reproduce any arbitrary mean DC, $\overline{\Psi}$, by properly selecting the distribution's parameters. In particular, the mean values of the beta and Kumaraswamy distributions are related with their shape parameters as [15], [17]:

$$\overline{\Psi} = \begin{cases} \dfrac{\alpha}{\alpha+\beta} & \text{for beta distribution} \\ bB\left(1+\dfrac{1}{a}, b\right) & \text{for Kumaraswamy distribution} \end{cases}$$

with $B(\alpha,\beta)$ again being the Beta function. Notice that the above equation can be satisfied for a given $\overline{\Psi}$ with different combinations of shape parameters α,β and a,b. The particular selection of the shape parameters determines the shape of the distributions as well as the resulting channel occupancy pattern in the time domain. The possible PDF shapes can be classified into six elemental archetypes, each with a characteristic time-domain pattern. Each archetype is defined by its load level (L: low, M: medium, and H: high) as well as its load pattern (type I: very bursty, and type II: moderately bursty, but not constant). The ranges of shape parameters for each archetype are related to the corresponding time-domain patterns as follows:

- Case L.I ($\alpha < 1$, $\beta \geq 1$): The channel is used (Ψ(t)>0) sporadically and remains unused ($\Psi(t) \approx 0$) most of the time.
- Case L.II ($1 < \alpha < \beta$): The channel is used (Ψ(t)>0) regularly by traffic with low activity factors.
- Case M.I ($\alpha < 1$, $\beta < 1$): The channel is subject to an intermittent use, where high-load periods are followed by low-load periods in a similar proportion.
- Case M.II ($\alpha > 1$, $\beta > 1$, $\alpha \sim \beta$): The channel usage level oscillates weakly around the average level.
- Case H.I ($\alpha \geq 1$, $\beta < 1$): The channel is used (Ψ(t)≈1) most of the time, with some periods of lower occupancy levels (Ψ(t)<1).
- Case H.II ($\alpha > \beta > 1$): The channel is not fully used (Ψ(t)<1) but subject to a constant, intensive usage.

The range of values indicated for the parameters of the beta distribution is also valid for the Kumaraswamy distribution by replacing α with a and β with b in type-I cases. In type-II cases, the resulting Kumaraswamy distribution is more difficult to control since the same constraints on a and b may hold for various load levels. Figure 55 and Figure 56 show some examples of the shape of the distributions and the resulting channel occupancy patterns in the time domain for channels with medium loads. Based on the above archetypes and the corresponding range of shape parameters, the parameters of the models can be configured in order to reproduce not only arbitrary mean load levels but also a wide range of occupancy patterns.

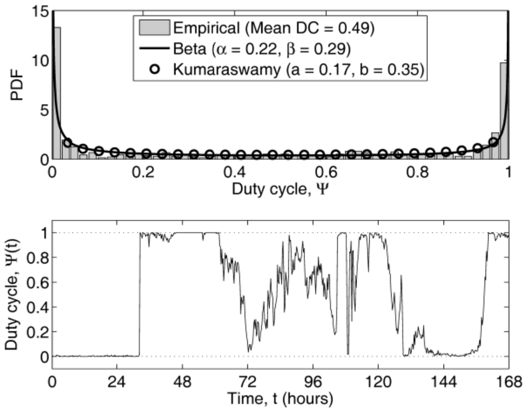


Figure 55: Stochastic DC models: case M.I.

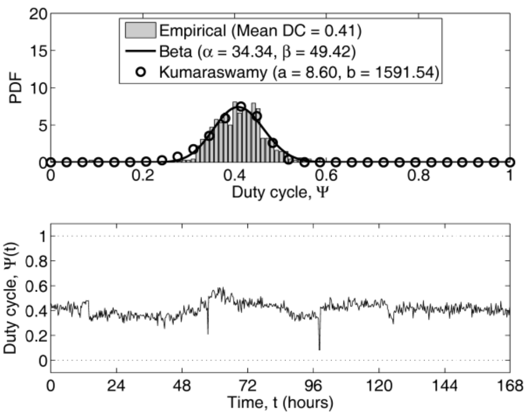


Figure 56: Stochastic DC models: case M.II.

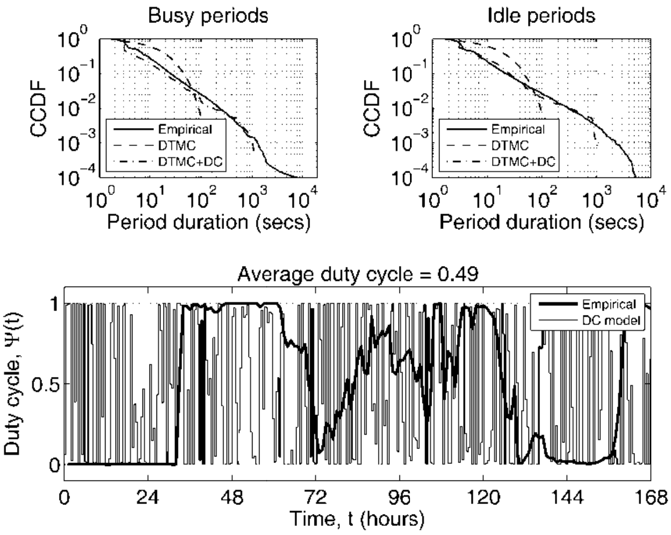


Figure 57: Empirical and DTMC-simulated distributions of busy and idle periods along with DC time evolution for a TETRA downlink channel.

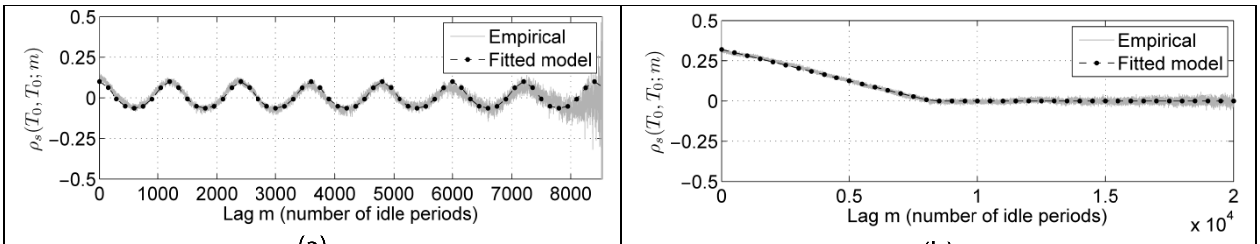
The capability of the non-stationary DTMC model along with the stochastic DC models to reproduce the statistical distributions of busy and idle periods in real channels is illustrated in Figure 57, including the distributions of busy and idle periods obtained by means of simulation with the non-stationary DTMC channel model. As it can be appreciated, the sequence of $\Psi(t)$ values generated with the stochastic DC model does not follow the empirical $\Psi(t)$ values of the channel in the time domain. However, it is important to note that the stochastic approach is not aimed at reproducing the time evolution of a particular realisation of the stochastic process $\Psi(t)$, but the statistical properties thereof. The results shown in Figure 57 indicate that this is a valid and accurate modelling approach for channels with random load variation patterns.

When implemented in simulation tools, the non-stationary DTMC with stochastic DC models may not lead to accurate results if some observations are not carefully taken into account. In particular, the DTMC has to be iterated for a sufficient number of times, N , before updating the transition matrix $\mathbf{P}(t)$ according to the stochastic DC model. During such amount of iterations, the transition probabilities of the DTMC must remain unaltered. After N iterations, a new value of $\Psi(t)$ can be generated from a beta or Kumaraswamy distribution, and used to update the transition matrix $\mathbf{P}(t)$ for the next N iterations. If the transition matrix is updated excessively fast (e.g., every iteration) the overall model may not be able to accurately reproduce the lengths of busy and idle periods.

In summary, the non-stationary DTMC model along with the presented deterministic and stochastic DC models is able to accurately reproduce not only the mean occupancy level but also the statistical properties of busy and idle periods observed in real channels with varying load patterns.

6.2.2 Time-correlation models

The DTMC (with appropriate DC models) and CTSMC channel models are capable to capture and reproduce the mean channel occupancy and the statistical distributions of busy and idle periods observed in real channels. Nevertheless, previous studies [22], [23] have indicated that in some cases the lengths of the busy and idle periods can be correlated, a feature that the DTMC and CTSMC models cannot reproduce. Although in general high correlation levels are not observed in practice, an accurate and realistic model of spectrum usage should take this feature into account. Experimental studies [23] have shown that the lengths of busy and idle periods exhibit negative correlation coefficients, meaning that when the length of a busy period increases, the length of the next idle period tends to decrease and vice versa. This can be explained by the fact that when the channel load increases, then the fraction of time that it remains in use increases and, as a result, the duration of busy periods increases while idle periods become shorter. On the other hand, the opposite behaviour is observed when the channel load decreases (i.e., the length of busy periods decreases and idle periods become longer). The correlation between the sequence of periods of the same type (either busy or idle) of a channel and a shifted version of itself (i.e., the autocorrelation) has experimentally been observed to exhibit two different behaviours, namely one periodic and another non-periodic. This is illustrated in Figure 58 which shows some examples of the autocorrelation function of idle periods as a function of the lag number, m, based on the Spearman's rank correlation coefficient, [24], which is denoted as $\rho_s (T_0, T_0; m)$.



(a) (b)

Figure 58: Autocorrelation function of busy or idle periods: (a) periodic pattern, (b) non-periodic pattern.

For channels with periodic autocorrelation functions (see Figure 58a) with period M, the correlation coefficient can be expressed as the summation of two bell-shaped exponential terms centred at lags $m=1$ and $m=M+1$, with amplitudes A and widths σ:

$$\rho_s(T_i,T_i;m)=\begin{cases}1, & m=0\\ \rho_s^{min}+Ae^{-\left(\frac{m-1}{\sigma}\right)^2}+Ae^{-\left(\frac{m-M-1}{\sigma}\right)^2}, & 1\le m\le M\end{cases}$$

where ρ_s^{min} is the minimum correlation. Based on field measurements, it has been empirically been observed that $\rho_s^{min}\approx -0.1$, in most cases, $A\in[0.2,0.5]$, M is equal to the average number of lags equivalent to 24 hours, and $\sigma\approx M/4$. For channels with non-periodic autocorrelation functions (see Figure 58b), the correlation coefficient takes its maximum value ρ_s^{max} at $m=1$ and decreases linearly with m until $m=M$, beyond which the correlation is approximately zero. This behaviour can adequately be modelled by:

$$\rho_s(T_i,T_i;m)=\begin{cases}1, & m=0\\ \rho_s^{max}\left(\dfrac{M-m}{M-1}\right), & 1\le m\le M\\ 0, & m>M\end{cases}$$

In this case, it has been observed that $\rho_s^{max}\in[0.1,0.4]$ and $M\in[2000,8000]$.

As it can clearly be appreciated in Figure 58, the models proposed are able to accurately describe the time-domain autocorrelation properties of spectrum usage empirically observed in real systems.

To reproduce time correlations, a simulation approach based on the aggregation and superposition of the realizations of several CTSMC processes has been proposed in [25]. The main limitation of such proposal is that the resulting correlations depend on the number of aggregated processes as well as their distributions, and the distribution parameters as well as the number of processes to be aggregated in order to reproduce a particular correlation level cannot easily be determined, making necessary the use of simulations to this end, which complicates the configuration of the model and hence its application. An alternative approach based on random variable generation principles has been proposed in [23]. Such another proposal requires as input information the particular distributions of idle and busy periods to be reproduced, the desired correlation coefficient between busy/idle periods as well as the desired autocorrelation function (periodic or non-periodic) of idle or busy periods. The algorithm outputs sequences of busy/idle period durations satisfying the desired specifications.

6.2.3 Spatial Duty Cycle Models

The spatial models presented in this section are envisaged to describe the average level of occupancy (expressed in terms of the DC) that would be perceived by DSA/CR users at various geographical locations based on the knowledge of some simple primary signal parameters. Moreover, an extension is proposed in order to characterise not only the average occupancy perception but also the simultaneous observations of various DSA/CR users on the spectrum occupancy pattern of the same transmitter. The interest of employing the DC lies in its ability to summarise the overall spectrum occupancy within a certain time interval and frequency range in a single numerical value.

Before going into details on the models, it is important to make a clear distinction between the Activity Factor (AF) of a primary transmitter in a certain channel and the DC perceived by secondary DSA/CR terminals in that channel. The AF of a primary transmitter represents the fraction of time that the transmitter is active (i.e., transmitting in the channel). A DSA/CR terminal in an arbitrary location with good radio propagation conditions with respect to the primary transmitter would observe the channel as busy whenever the primary transmitter is active. However, at other locations where the propagation conditions are not so favourable, the primary signal might not be detected. In such a case, the level of spectrum activity perceived by the DSA/CR terminal in terms of the DC would be lower than the actual AF of the primary transmitter. While the AF is unique for a given transmitter, the DC perceived at different locations may be different. Since the propagation conditions strongly vary with the geographical location, the perceived DC will vary over space accordingly. The models discussed in this section describe the spatial distribution of the DC as a function of the radio propagation conditions.

The occupancy state of a channel as perceived by a DSA/CR terminal depends on the employed spectrum sensing method [27]. Due to its simplicity, wide range of application and relevance, energy detection (ED) has been a preferred choice for DSA/CR. According to ED, a DSA/CR terminal measures the power received in a certain frequency band over a predefined time period T, which can be expressed as:

$$P_R = \frac{1}{T}\int_{-T/2}^{+T/2} P_R(t)\,dt$$

where $P_R(t)$ is the instantaneous power received by the DSA/CR terminal (including noise) and P_R is the average power over the sensing period T. The average power P_R is compared with a predefined threshold in order to decide on the primary channel state: if P_R is above the threshold the channel is declared to be busy; otherwise, it is assumed to be idle. The perceived spectrum occupancy at a particular location therefore depends on the statistics of the received average power, P_R. Note that the instantaneous power $P_R(t)$ is a stochastic process that can be thought of as a non-countable infinity of iid random variables, one for each time instant. Since P_R is obtained as the average of an infinite number of random variables, the central limit theorem can therefore be employed to approximate the PDF of P_R as a normal distribution, regardless of the real distribution of $P_R(t)$ [28].

Let us denote the distribution of the noise power as $f_N(P_N) \sim N(\mu_N, \sigma_N^2)$ and the distribution of the signal power (received in the presence of noise) as $f_S(P_S) \sim N(\mu_S, \sigma_S^2)$. According to this formulation, μ_N represents the noise floor of the DSA/CR receiver and σ_N denotes the standard deviation of the noise powers P_N experienced at various sensing events. The primary power P_S received in the presence of noise is characterised by an average value μ_s that depends on the transmission power and radio propagation conditions and a standard deviation σ_s that is additionally affected by the noise of the DSA/CR receiver.

If the sensed channel is idle, the PDF of the observed average power, $f_R(P_R)$, will be that of the noise, $f_N(P_N)$. In such a case, the probability that the observed power is above the threshold (i.e., the perceived DC) is given by (see Figure 59):

$$\Psi_{idle} = \int_{\lambda}^{\infty} f_R(P_R) dP_R = \int_{\lambda}^{\infty} f_N(P_N) dP_N = P_{fa}$$

where it has been assumed that the decision threshold λ is set so as to meet a specified target probability of false alarm P_{fa}.

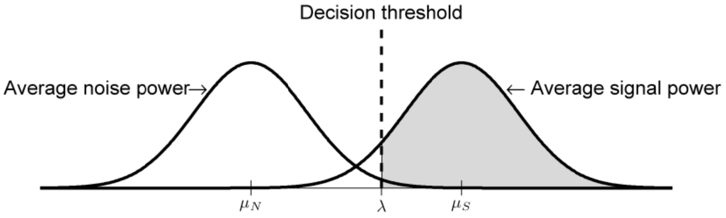


Figure 59: Model considered to compute the DC (shaded area).

On the other hand, if the channel is busy when it is sensed, the PDF of the observed average power, $f_R(P_R)$, will be that of the received signal, $f_S(P_S)$. Assuming an ideal situation where there is no noise, the DC perceived by the DSA/CR user would be given by:

$$\begin{aligned}\Psi_{busy}^{ideal} &= \int_{\lambda}^{\infty} f_R(P_R) dP_R = \int_{\lambda}^{\infty} f_S(P_S) dP_S \\ &= \frac{1}{\sqrt{2\pi}\sigma_S} \int_{\lambda}^{\infty} e^{-\frac{1}{2}\left(\frac{P_S-\mu_S}{\sigma_S}\right)^2} dP_S = Q\left(\frac{\lambda-\mu_S}{\sigma_S}\right)\end{aligned}$$

where $Q(\cdot)$ represents the Gaussian *Q*-function. Notice that the above equation indicates that the perceived occupancy in terms of the DC would tend to zero as the received power decreases (i.e., $\mu_S \rightarrow -\infty$). However, if the received signal power is below the receiver's noise, this situation would be equivalent to an idle channel where the receiver observes only noise. In such a case, the perceived DC should be P_{fa}. A more realistic model for $f_R(P_R)$ when the channel is busy, taking into account the presence of noise, would be $f_R(P_R) = M\left\{f_N(P_N), f_S(P_S)\right\}$, where $M\left\{\cdot\right\}$ denotes a *realisation-wise maximum operator* defined as follows. If $A = \{x_{a_1}, x_{a_2}, \ldots, x_{a_n}, \ldots, x_{a_N}\}$ and $B = \{x_{b_1}, x_{b_2}, \ldots, x_{b_n}, \ldots, x_{b_N}\}$ represent two sets of N random numbers (realisations) drawn from distributions $f_a(x_a)$ and $f_b(x_b)$, respectively, then $f_c(x_c) = M\{f_a(x_a), f_b(x_b)\}$ represents the distribution of the elements of the set $C = \{x_{c_1}, x_{c_2}, \ldots, x_{c_n}, \ldots, x_{c_N}\}$, where $x_{c_n} = \max\{x_{a_n}, x_{b_n}\}$ for $n = 1, 2, \ldots, N$, when N tends towards infinity. Notice that this operator reproduces the effect of the noise floor on the observed power (i.e., the DSA/CR user observes the received signal power when it is above the noise floor or the noise power otherwise). Therefore, this definition of $f_R(P_R)$ provides a more realistic model for the average power P_R actually observed by the DSA/CR receiver. Based on this model, the DC perceived by the DSA/CR user when the channel is busy will then be given by:

$$\begin{aligned}\Psi_{busy}^{real} &= \int_\lambda^\infty f_R(P_R)dP_R = \int_\lambda^\infty M\left\{f_N(P_N), f_S(P_S)\right\}dP_R \\ &= \max\left\{\int_\lambda^\infty f_N(P_N)dP_N, \int_\lambda^\infty f_S(P_S)dP_S\right\} = \max\left\{P_{fa}, Q\left(\frac{\lambda-\mu_S}{\sigma_S}\right)\right\}\end{aligned} \tag{1}$$

As it can be appreciated, this model rightly predicts that the perceived activity level is never lower than the target P_{fa}. The average DC perceived by the DSA/CR user will depend on the transmission power of the primary transmitter and its particular activity pattern.

Constant-power continuous transmitters

If the primary transmitter is always active, (e.g., TV and audio broadcasting services), the probability that the received average power P_R is above the decision threshold λ and the DSA/CR user observes the channel as busy is given by equation (1).

Assuming that the decision threshold is set to meet a certain P_{fa}:

$$P_{fa} = \int_\lambda^\infty f_N(P_N)dP_N = \frac{1}{\sqrt{2\pi}\sigma_N}\int_\lambda^\infty e^{-\frac{1}{2}\left(\frac{P_N-\mu_N}{\sigma_N}\right)^2} dP_N = Q\left(\frac{\lambda-\mu_N}{\sigma_N}\right) \tag{2}$$

Solving in equation (2) for λ yields the decision threshold

$$\lambda = Q^{-1}(P_{fa})\sigma_N + \mu_N \tag{3}$$

where $Q^{-1}(\cdot)$ denotes the inverse of $Q(\cdot)$. Substituting equation (3) into equation (1) finally yields the DC model:

$$\Psi = \max\left\{P_{fa}, Q\left(\frac{Q^{-1}(P_{fa})\sigma_N - \Gamma}{\sigma_S}\right)\right\} \tag{4}$$

where $\Gamma = \mu_S - \mu_N$ represents the average SNR expressed in decibels, while σ_S and σ_N are the standard deviation of the signal and noise average powers also in decibels.

Constant-power discontinuous transmitters

If the primary transmitter is characterised by a certain AF, denoted as $0 < \alpha < 1$, the PDF of the received average power $f_R(P_R)$ will be that of the primary signal (with noise) $M\{f_N(P_N), f_S(P_S)\}$ whenever the transmitter is active (which will occur with probability α) or noise $f_N(P_N)$ otherwise. Hence:

$$f_R(P_R) = (1-\alpha)f_N(P_N) + \alpha M\{f_N(P_N), f_S(P_S)\}$$

and the resulting expression for the DC becomes:

$$\begin{aligned}\Psi &= \int_\lambda^\infty f_R(P_R)dP_R \\ &= (1-\alpha)\int_\lambda^\infty f_N(P_N)dP_N + \alpha\int_\lambda^\infty M\{f_N(P_N), f_S(P_S)\}dP_R \\ &= (1-\alpha)\int_\lambda^\infty f_N(P_N)dP_N + \alpha\max\left\{\int_\lambda^\infty f_N(P_N)dP_N, \int_\lambda^\infty f_S(P_S)dP_S\right\} \\ &= (1-\alpha)P_{fa} + \alpha\max\left\{P_{fa}, Q\left(\frac{Q^{-1}(P_{fa})\sigma_N - \Gamma}{\sigma_S}\right)\right\}\end{aligned}$$

Notice that equation (4) is a particular case of equation (5) with $\alpha = 1$.

Variable-power discontinuous transmitters

The previous model could be extended for the case of variable-power transmitters. In this case, the average transmission power is not constant but characterised by a certain PDF. To simplify the model, let's assume that the variability of the transmission power can adequately be described by a discrete set of K average transmission power levels, instead of a continuous PDF. This assumption not only simplifies the analytical expressions of the model, but also enables the application of the model to the case in which a channel is time-shared by K transmitters with different power levels as it may be the case of various TDMA-based systems such as GSM/DCS, TETRA, etc. The model can embrace the cases of a single variable-power transmitter with K transmission power levels and K constant-power transmitters time-sharing the channel. In both cases, the problem reduces to the possibility of observing K different average transmission powers in the channel.

Let's denote as $f_{S_k}(P_{S_k})$, with mean μ_{S_k} and standard deviation σ_{S_k}, the PDF of the received average power at certain location when the *k*-th transmission power level is present in the channel ($k = 1, 2, \ldots, K$). In general it can be assumed that $\mu_{S_p} \neq \mu_{S_q}$ and $\sigma_{S_p} \neq \sigma_{S_q}$ for $p \neq q$. Let's define an AF α_k for each transmission power representing the probability that the *k*-th transmission power level is present in the channel. In the case of a single-transmitter with K transmission power levels, only one out of the K power levels can be selected at a time. Moreover, in the case of K transmitters time-sharing the channel it is reasonable to assume that there exists some MAC mechanism so that when one primary transmitter accesses the channel the rest of potential primary transmitters remain inactive. In both cases, the K average power levels are mutually exclusive events. Hence:

$$\sum_{k=1}^{K} \alpha_k \leq 1 \tag{6}$$

where the equality holds when the channel is always busy.

The left-hand side of equation (6) represents the probability that any of the K transmitters is active, i.e. the probability that the channel is busy, and its complementary probability $1 - \sum_{k=1}^{K} \alpha_k$ is the probability that the channel is idle. The PDF of the received average power $f_R(P_R)$ will be that of the *k*-th primary signal (with noise) $M\left\{ f_N(P_N), f_{S_k}(P_{S_k}) \right\}$ whenever the *k*-th transmission power is active (which will occur with probability α_k) or it will be noise $f_N(P_N)$ otherwise. Hence:

$$f_R(P_R) = \left(1 - \sum_{k=1}^{K} \alpha_k \right) f_N(P_N) + \sum_{k=1}^{K} \alpha_k M\left\{ f_N(P_N), f_{S_k}(P_{S_k}) \right\}$$

and the resulting expression for the DC becomes

$$
\begin{aligned}
\Psi &= \int_{\lambda}^{\infty} f_R(P_R)dP_R \\
&= \left(1-\sum_{k=1}^{K}\alpha_k\right)\int_{\lambda}^{\infty} f_N(P_N)dP_N + \sum_{k=1}^{K}\alpha_k \int_{\lambda}^{\infty} M\left\{f_N(P_N), f_{S_k}(P_{S_k})\right\}dP_R \\
&= \left(1-\sum_{k=1}^{K}\alpha_k\right)\int_{\lambda}^{\infty} f_N(P_N)dP_N + \sum_{k=1}^{K}\alpha_k \max\left\{\int_{\lambda}^{\infty} f_N(P_N)dP_N, \int_{\lambda}^{\infty} f_{S_k}(P_{S_k})dP_{S_k}\right\} \\
&= \left(1-\sum_{k=1}^{K}\alpha_k\right)P_{fa} + \sum_{k=1}^{K}\alpha_k \max\left\{P_{fa}, Q\left(\frac{Q^{-1}(P_{fa})\sigma_N - \Gamma_k}{\sigma_{S_k}}\right)\right\}
\end{aligned}
$$

where $\Gamma_k = \mu_{S_k} - \mu_N$ is the SNR resulting from the *k*-th average transmission power level expressed in decibels.

6.2.3.1 Model applicability: Statistical prediction approach

To demonstrate the applicability of the proposed Spatial Duty Cycle model consider the scenario consisting of a generic urban environment where buildings of height h_r(m) are deployed following a uniform layout with inter-building separation b(m) and street width w(m) as shown in Figure 60. The area under study comprises a grid of 5×5 buildings. A TV transmitter (K = 1 and $\alpha_k = 1$) with height h_b(m) and operating frequency f = 800 MHz is located d(km) apart (taking as a reference the geometrical centre of the area under study). Within this scenario, CR nodes with antenna height h_m = 2 m may be located at the centre of building rooftops (height is $h_r + h_m$), at the ground level between buildings (height is h_m), or inside buildings at any floor (height is $n \cdot h + h_m$, with n being the floor number and h = 3 m/floor). The considered locations represent various cases of practical interest and embrace a wide range of receiving conditions and levels of radio propagation blocking, ranging from direct line-of-sight at rooftops to severely blocked and faded signals at the ground level and inside buildings. CR nodes intend to opportunistically access the spectrum band used by the primary transmitter, and to this end they sense the spectrum. The aim is to determine the spectrum occupancy that would be perceived by CR nodes within this scenario.

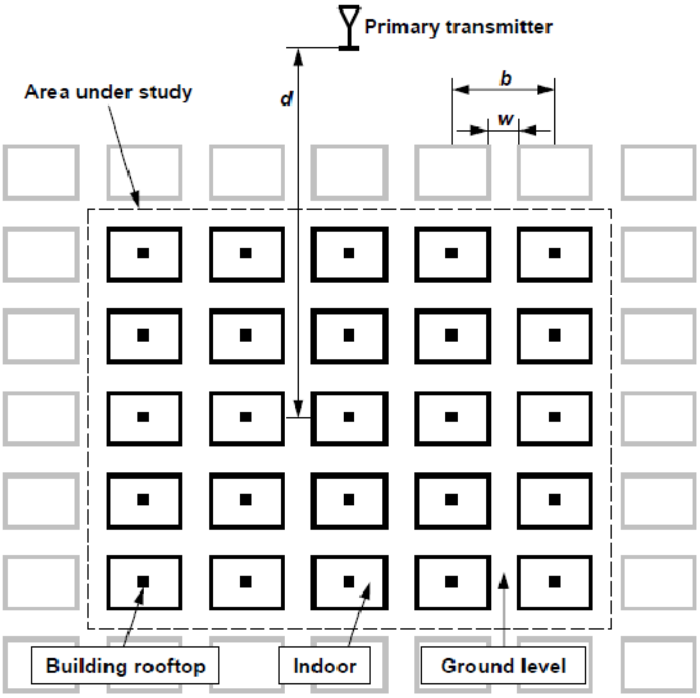


Figure 60: Scenario under study.

To compute the Duty Cycle at the CR locations consider the statistical prediction approach illustrated in Figure 61, which makes use of radio propagation models to compute the signal to noise ratio at the different locations of the CR nodes.

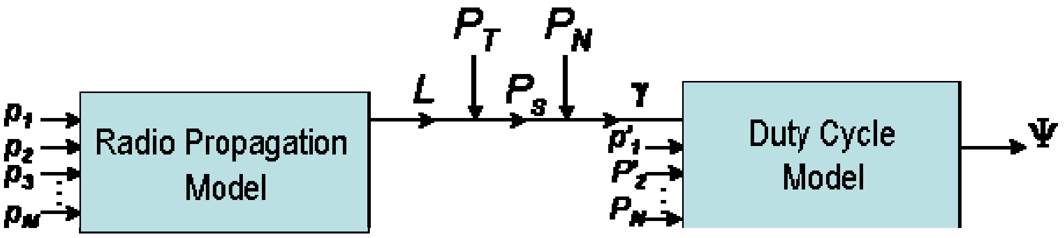


Figure 61: Model for statistical prediction approach.

As a first step, the radio propagation model is used to estimate, based on a set of input parameters $p = (p_1, p_2, \ldots, p_M)$ such as operating frequency, distance, etc., the radio propagation loss L between the primary transmitter and the considered CR receiver that is sensing the spectrum. Notice that the propagation model is not constrained by an specific proposed model, allowing the selection of the propagation model that best fits the scenario under study. Next, based on the primary transmission power P_T and the computed losses L, the primary power P_S, observed by CR nodes at various locations, is obtained. Then, the received P_S values are translated to SNR values γ

by making use of the noise power P_N at the CR receiver. The resulting SNR values are then fed, along with an additional set of input parameters $p' = (p'_1, p'_2, \ldots, p'_N)$, to the DC model, which outputs an estimation Ψ of the DC that would be perceived by the CR nodes at the considered geographical locations. The perceived DC, Ψ, is computed as derived above.

Notice that, in this case, the set of parameter p′ are: the K transmission power levels that may be present in the channel, the activity factor (α_k) of the k-th power level, the target probability of false alarm (P_{fa}) considered for fixing the energy decision threshold (λ), and the standard deviation of the k-th signal (σ_{Sk}) and noise (σ_N) associated to the sensed signal and noise power levels at the CR node respectively. Such values (σ_{Sk} and σ_N) can be estimated or obtained experimentally using, for instance, a spectrum analyser. Table 10 shows some examples of σ_{Sk} and σ_N for various bands, which have been derived from the measurements performed in [28].

Table 10: Experimental values of σ_{Sk} and σ_N.

Band	Bandwidth	Sensing Time (ms)	σ_{Sk}	σ_N (dB)
TV	8.0 MHz	200	0.5252	0.1679
UMTS	5.0 MHz	125	0.4138	0.2093
DAB-T	1.7 MHz	42,5	0.8298	0.3640
GSM/DCS	200 KHz	5	1.6421	0.8921
TETRA	25 KHz	0,625	2.0469	1.3624

Figure 62 shows the primary signal power P_S (dBm) received at various locations within the area under study. The highest power level is observed at rooftops, as expected. It is interesting to note that the signal strength at that locations is slightly above the sensitivity of conventional TV receivers (≈ −85dBm), meaning that the area under study corresponds to the border of the coverage area intended for the primary TV transmitter. That is, primary users may be present within the area under study. The results of Figure 62 also indicate that the primary signal power received at the ground level and inside buildings suffers important attenuations with respect to rooftops, in the order of 10 and 20 dB respectively. This suggests that CR nodes within the same geographical area may experience quite dissimilar perceptions depending on their particular locations and propagation conditions.

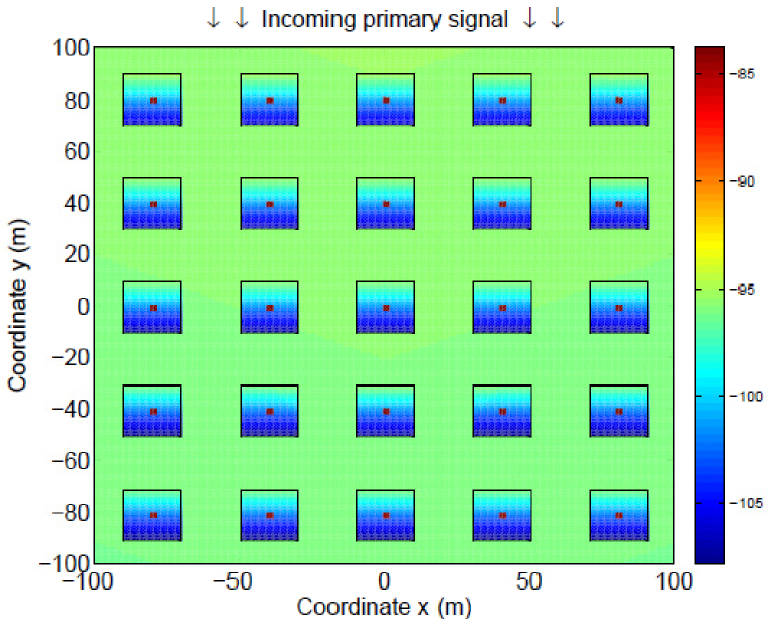


Figure 62: Primary signal power received at various locations within the area under study.

To estimate the spectrum occupancy, within the considered scenario, perceived by the CR nodes first of all, the simple approach to sense the received primary signal and compare it with a given threshold is used. The decision threshold, $\lambda \approx -96$dBm, is selected to provide a probability of false alarm $P_{fa} = 1\%$. Figure 63 shows the obtained results, where the white and gray colours indicate that the primary signal is declared to be present or absent, respectively, at the corresponding location. At rooftops, where the highest power levels are observed, the primary signal is always detected according to this prediction approach. On the other hand, the signal appears to always be undetected in all indoor locations, where the lowest power levels are observed (excepting some buildings where the signal can be detected in indoor locations very near to the externally illuminated wall). At the ground level, the considered prediction approach indicates that the primary signal may be detected depending on the CR node's location. These results indicate that the considered simple approach is able to predict that, in the border of the primary coverage area (i.e., under low SNR conditions), there may be situations where primary users are present but secondary CR nodes may not detect their transmissions, which would lead to potential situations of harmful interference. It is important to note, however, that the resulting characterization of the perceived spectrum occupancy is not only excessively simple but also unrealistic. In fact, Figure 63 shows, at the ground level, that this prediction approach indicates the existence of a hard limit such that the primary signal is always detected at locations slightly above, but never detected at locations slightly below.

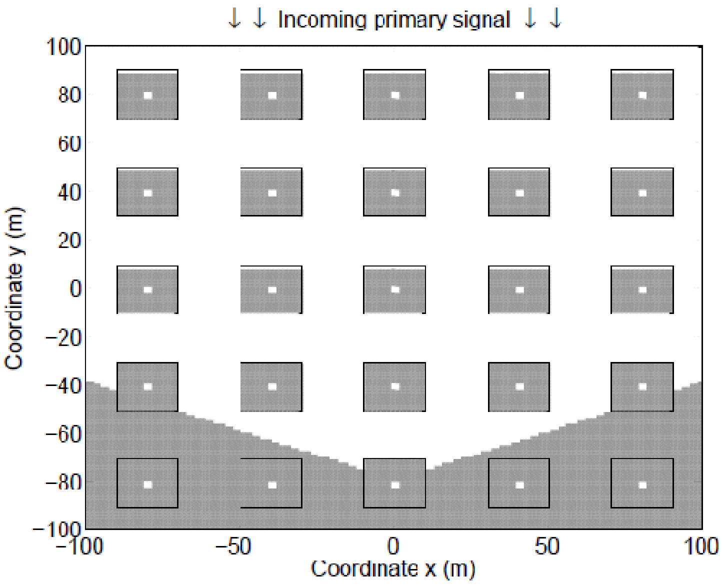


Figure 63: Binary spectrum occupancy pattern perceived at various locations within the area under study.

However, when applying the proposed Statistical Prediction Approach the spectrum occupancy characterization shown in Figure 64 result. Comparing Figure 63 with Figure 64, it can be observed that the two considered prediction approaches agree for locations at building rooftops and indoor environments, which can be considered as extreme cases in the sense that the highest and lowest primary signal powers are observed, respectively, at such locations. Thus, at rooftops (highest primary power), the simple approach predicts that the sensed spectrum is always observed as busy while the proposed approach indicates that this occurs with probability equal to one. At indoor locations (lowest primary power), the simple approach indicates that spectrum is always observed as idle, which is also corroborated by the proposed approach with a probability of observing the spectrum as busy approximately equal to zero. The differences between both approaches are observed in the spectrum occupancy perception predicted at the ground level. While the simple approach provides a simple binary characterization with a hard borderline, the proposed approach provides a more sophisticated characterization by means of the probability that the spectrum is observed as busy, which increases progressively as the considered location approaches the primary transmitter (without observing any abrupt transitions).

It is also interesting to note that, in some locations where the simple approach predicts the channel as always idle (busy), the proposed statistical prediction approach indicates that the probability of observing the channel as busy is low (high), but not equal to zero (one). This example shows how the proposed statistical prediction approach is able to provide a sophisticated and realistic characterization of the perceived spectrum occupancy as a function of the considered propagation scenario.

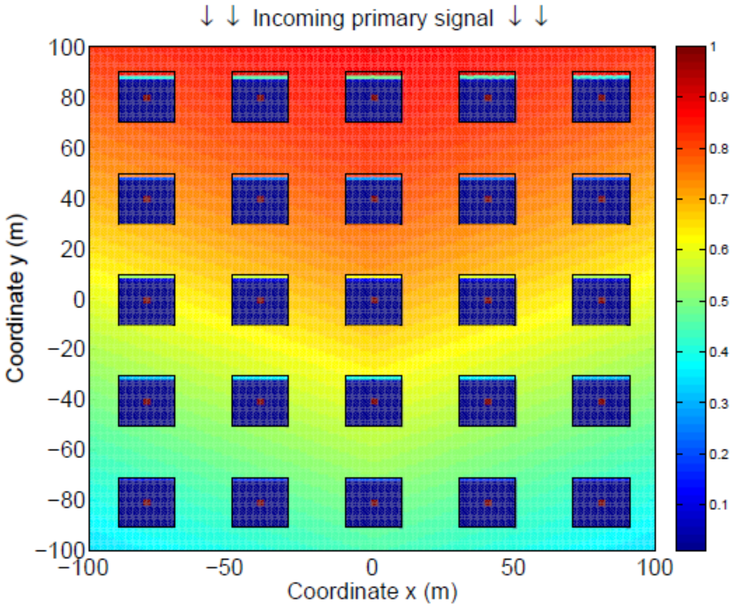


Figure 64: Probabilistic spectrum occupancy pattern perceived at various locations within the area under study.

The proposed approach can also be used to study the impact of certain scenario and propagation parameters on the user's perception. Figure 65, Figure 66 and Figure 67 show some examples.

Figure 65 evaluates the impact of the transmitter and receiver antenna heights when the CR node is at the ground level. As expected, the detection performance improves as antenna heights increase, which results from the reduction of the radio propagation blocking caused by buildings. However, there are other interesting observations that can be inferred from the prediction provided by the proposed approach, and that would not have been possible with the simple approach considered as a reference. For example, Figure 65 indicates that the primary signal would not be detected for $h_b < 42$m, but it would always be detected for $h_b > 46$m, regardless of h_m. To guarantee that the primary signal is detected at the ground level with a probability of 0.9, the primary antenna height should be around 10m above the rooftop level (i.e., $h_b \approx h_r + 10$m) for the selected configuration parameters. For fixed h_b, the detection performance can be improved by increasing h_m. In this sense, it is interesting to note that Figure 65 indicates that the CR node's antenna height, h_m, should be increased about 1m for Medium-Small Cities (MSC) with respect to Large Cities (LC) in order to obtain the same detection performance. This can be explained by the fact that the COST231 Walfisch-Ikegami model employed to obtain Figure 65 includes an additional attenuation factor for the higher amount of vegetation usually present in MSCs.

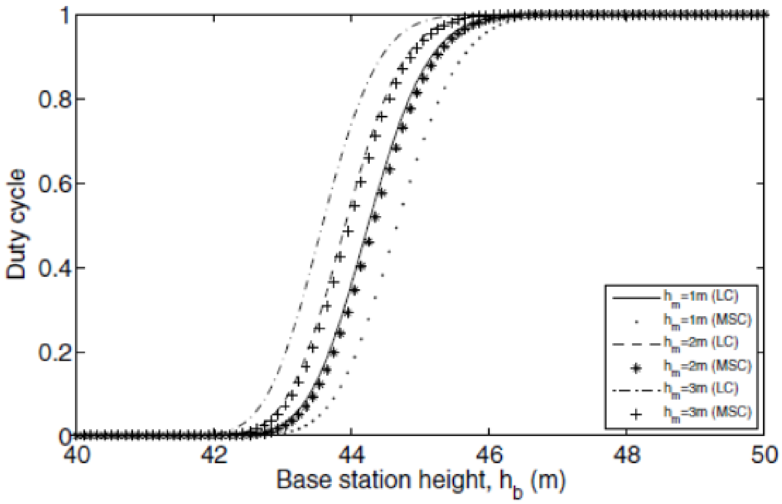


Figure 65: Impact of the transmitter and receiver antenna heights for CR nodes at the ground level

Figure 66 shows the impact of building height and street width on the perceived spectrum occupancy for CR nodes at the ground level. Again, the detection performance improves as the radio propagation blocking becomes less significant, which in this case occurs for smaller buildings and wider streets. For example, for a fixed building height (e.g., h_r =40m), the primary signal might be detected with probability close to one in wide streets (w = 25m) but it might be completely undetected in narrower streets (w = 15m). In general, the perceived spectral activity is higher in open areas than in narrow streets between buildings. It is worth noting that the trend shown in Figure 66 has also been observed empirically in the experimental study carried out in [29].

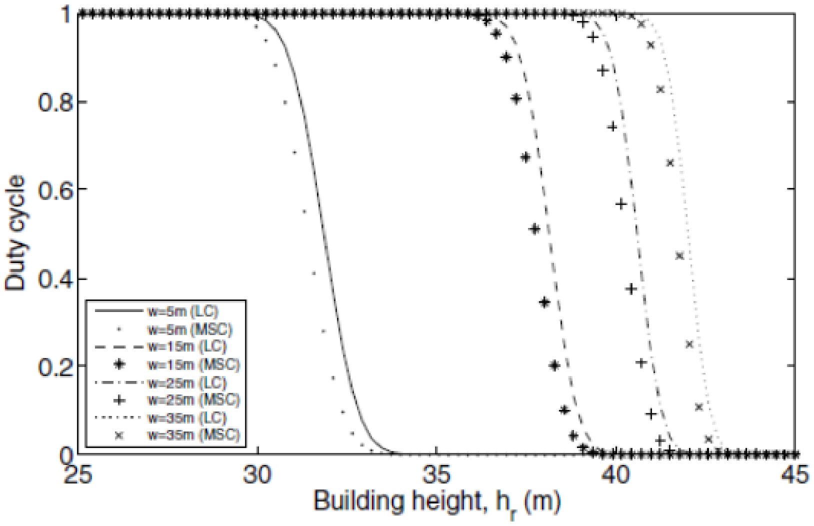


Figure 66: Impact of building height and street width on the perceived spectrum occupancy for CR nodes at the ground level.

Finally, Figure 67 explores the spectrum occupancy perceived in indoor environments as a function of the distance from the externally illuminated wall l and floor number n. Compared to the user's perception at the ground level, the received primary signal strength in the indoor environment decreases faster with the distance. In fact, the signal would be detected close to the wall, but would be missed a few meters apart. The exact point where the signal begins to be undetected depends on the considered floor. As it can be observed, the perceived spectral activity is more significant in higher floors, which again can be related to a lower radio propagation blocking.

These few examples illustrate how the proposed approach can be employed not only to provide a statistical prediction of the spectral activity perceived by CR nodes, but also to quantify and analyse the impact of certain particular scenario and propagation parameters on the user's perception, which constitutes an important aspect in the design and dimensioning of DSA/CR systems in real deployments.

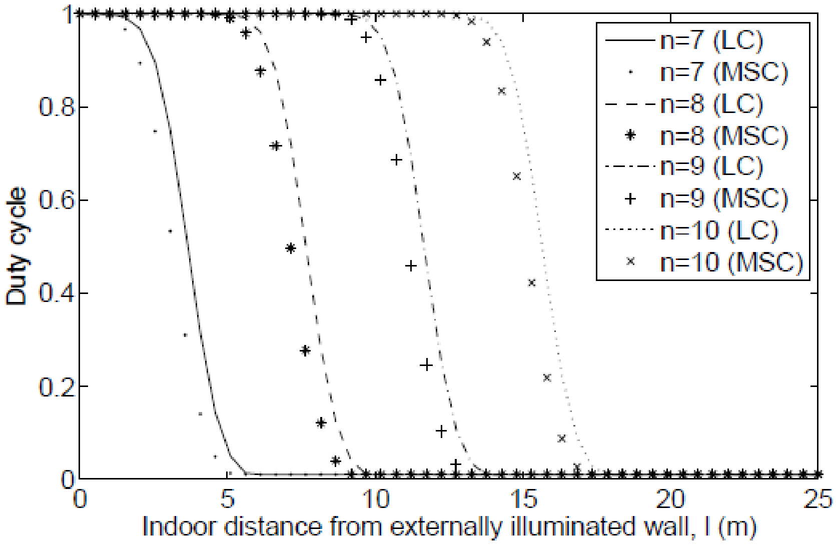


Figure 67: Perceived spectrum occupancy at indoor locations as a function of the distance from the externally illuminated wall *l* and floor number *n.*

6.3 Measurements and Models on Long-Term Indoor Propagation

The success of wireless communications has lead to fast increase in the demand for wireless services and data rates, and thus to the saturation of the spectrum. This has made the management of radio resources a very complicated problem that cannot be dealt with using traditional Radio Resource Management (RRM) techniques. Hence, more advanced paradigms and concepts, such as Dynamic Spectrum Access (DSA) and femtocells have been developed to increase the efficiency of spectrum use [32], [33]. However, these concepts require a better knowledge of the radio environment, especially radio propagation losses since they are very sensitive to interference.

Propagation models were the subject of a vast amount of research work during the last few decades [32]-[41]. However most of the developed models are either for outdoor or for very short-time scale applications (i.e., in the order of a symbol or at maximum few frames). Short-time models are suitable and have contributed to the evolution of signal processing techniques and RRM techniques operating on such small time scales such as fast power control. However, they are generally inappropriate for most RRM and planning techniques, especially those based on statistical models such spatial DSA techniques and interference management in femtocell networks. This is due to the fact that these techniques require thorough knowledge on propagation losses not only between transmitter and its receiver but also between the former and possible interference victims. These losses cannot be reported very fast since there are no direct connections between the interferer and the interfered node. More importantly, cellular operators are more interested in proactive techniques that prevent destructive interference situations than techniques that are triggered by such events.

In order to understand more the behaviour of indoor long-term propagation models, we present in this section the results of our extensive measurements in indoor environment. From these results we derive interesting conclusions on the dynamic characteristics of indoor propagation environment and we draw guidelines on how we develop statistical models that can be used by RRM techniques. The designed models are necessary especially for simulation models used to evaluate wireless network performance.

6.3.1 Problem Formulation

Most of the research works in RRM techniques uses a specific static propagation model, including the distribution characteristics of fast and slow fading. These models are normally represented by the sum of time-independent, slow, and fast fading factors. Practically all papers on RRM and planning techniques use predefined fading distributions that do not change with time. However, measurement campaigns and theoretical work have shown that this is not true, especially for indoor environment where fast and slow fading distributions are time dependent [34]-[36]. This is because the number of obstacles affecting the fast and slow fading in indoor environment is very low. Therefore, any changes in the positions of the obstacles will not only change the instantaneous value of the fading but also its distribution. For instance, moving the position of one chair, opening or closing a door can change both multipath fading and slow fading losses. This is not the case of outdoor environment where the large number of obstacles and higher variations in the lengths of the paths of individual signal components lead to more stable distributions.

In addition, a key difference between indoor and outdoor propagation is that in outdoor environment, radio wave propagation is fairly predictable; by using database including topographical and building information, we can efficiently determine the range and shape of a cell for a typical base station and also can characterize the channel environment with certain renown models like Okumura-Hata, COST 231-Hata and COST 231-Walfish-Ikegami [39][41]. However, modelling indoor RF propagation is much more challenging and requires more complex tools. In general, these tools generate deterministic losses between two points based on furniture and human positions. Due to the complexity of these tools, most of RRM designers use more simple statistical propagation models similar to the ones used in outdoor [42], which may generate wrong understanding of the environment and thus produce miss-leading results.

Our objective in this section is, therefore, to understand the long-term behavior of indoor propagation losses, in particular slow and fast fading. Moreover, we provide some guidelines on how to develop dynamic path loss models. Specifically, we study the characteristics of the path loss in power domain and not in signal amplitude as typically done in propagation and signal processing papers. We have chosen this approach since most of RRM techniques use signal powers and not signal amplitude. It should be noted that this work is not intending to characterize the short-term propagation losses that have been extensively studied in the literature.

In order to facilitate the study we consider, in a similar way to most propagation studies, that the path loss G is divided into three components, namely the distance-dependent component Δ(d), the shadow factor or slow fading χ(t) and the fast fading φ(t), where d is the distance between the two transceivers of the link and t is the time index (It should be noted here is that the range of the time t is normally considered to be higher than the coherence time of the channel). Without loss of generality we assume that antenna gains are included in the static distance-dependent component. Thus, the received power R(t) can be written as a function of these components, the noise power P_N and the transmit power P(t) in the logarithmic domain as

$$R(t,d) = P(t) - \left[\Delta(d) + \chi(t) + \phi(t)\right] + P_N \tag{7}$$

Therefore, we can write

$$G(t,d) = \Delta(d) + \chi(t) + \phi(t) \tag{8}$$

6.3.2 Experimental Setup and Results

In order to understand the behaviour of indoor radio environment, we conducted a measurement campaign using the following hardware and software:

- Agilent E4438C signal generator [43] was used to generate an un-modulated signal with the bandwidth of 30 kHz centred at 5.2 GHz. This centre frequency was chosen to avoid interference from uncontrolled equipments such as microwave ovens and to have minimum impact of WiFi networks since this frequency is not used in the buildings where the measurements are conducted. The power of the transmit signal was 30 dBm.
- Rohde & Schwarz FSL series spectrum analyzer was used for measuring the power of the received signal [44]. Table 11 shows the typical settings used in our campaign.
- A custom-made random antenna [45] was used throughout the campaign.
- Distributed spectrum system Graphical User Interface (GUI) was used to facilitate the collection of data from the spectrum analyzer. This GUI was developed in MATLAB to operate on a notebook connected to the spectrum analyzer. It allows easy controlling of all the settings of the spectrum analyzer [46].

- Rohde & Schwarz Tracerecorder application was used for storing measurement data from the spectrum analyzer. It was directly installed on the spectrum analyzer and used to store the measurement files in the hard disk of spectrum analyzer [44]. The advantage of using the tracerecorder application over the MATLAB GUI is the short delay that is incurred in storing measurement files directly on the spectrum analyzer. This means that measurements with higher sampling rates than the MATLAB program can be conducted using this software. However, this software may not be used for long-time measurements due to small storage memory of spectrum analyzer.

Table 11: Basic settings of spectrum analyzer.

Span	Sweeptime	Measurement points	Resolution	Preamplifier
30 kHz	250/4 ms	501	10 kHz	off

The measurements were conducted in three different rooms in the Institute for Networked Systems of the RWTH Aachen University (Figure 68). The characteristics of the rooms are collected in Table 12. The building the rooms are a part of is a typical office building, with both soft partitioning and supporting walls. In addition, the Foyer has an entrance to two corridors that are used frequently by people to move between the offices. Hence the indoor environment in this location is changing frequently due to movement of the personnel.

Table 12: Description of the measurement environment.

Location	Furniture and Characteristics
Student office	Office furniture with four desktop computers, large glass windows and a glass door
Laboratory	Office furniture, electronic equipments placed in cartons along with six desktop computers, large glass windows with a wooden
Lecture room	Wooden chairs and a tables large glass windows and a glass door
Foyer	Large sofa where people are sitting during the lunch break, glass window to outdoor, and three glass doors for the three offices in

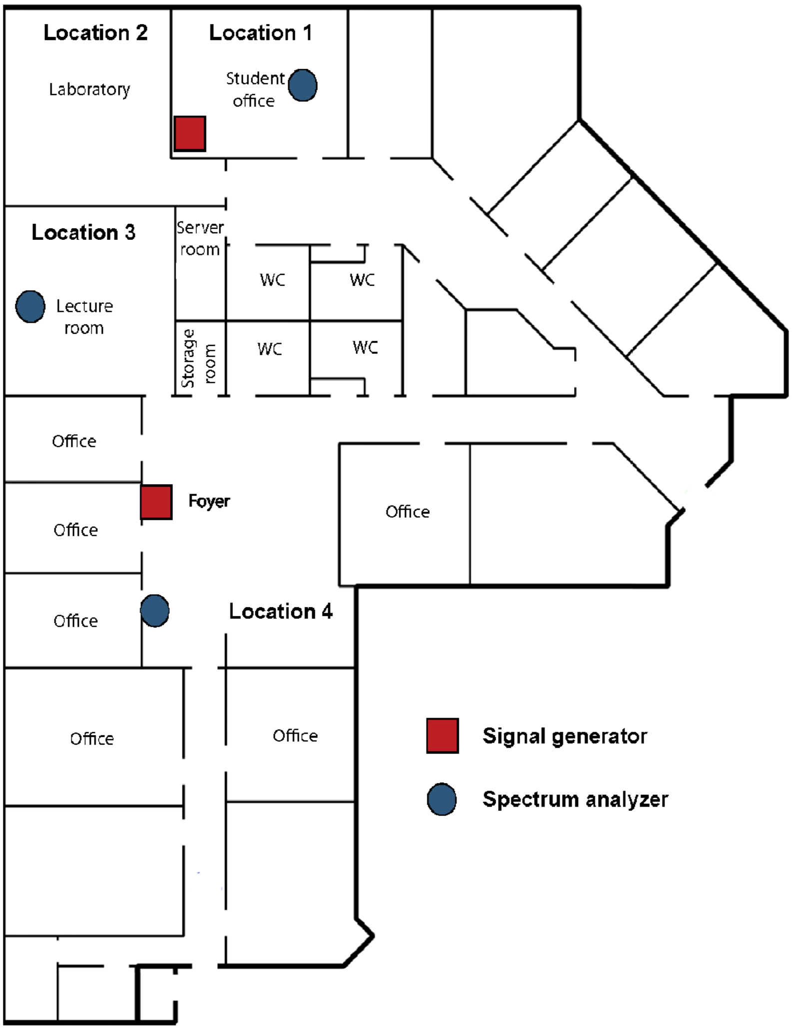


Figure 68: Map of the institute where the measurements were conducted.

We first evaluated the noise power in different locations and time periods when no signal was transmitted, which was always lower than -100 dBm. This value is very low compared to the received powers that generally are in the range of -30 dBm to -70 dBm for 90 % of the time in all scenarios. This implies that even if all of the measured signal was not noise, the interference component from other systems would thus not contaminate the measurements.

We have collected measurement data for over 109 hours including four full days of measurements of 24 hours each. The measurements are made in a real environment — without planned movements of the people — in order to understand the real behaviour of the RF channels. Figure 69 shows path loss levels obtained from measurements with Line of Sight (LOS) and Non Line of Sight (NLOS) propagation conditions over one day. As it can be seen from the figure, the path loss may have different periods with different behaviours. This can be represented by a mixture of distributions each valid for a period of time. Hence, each measurement is decomposed into a sequence of 1 minute blocks or observations. The one minute period is chosen since most of the studies in the literature [34]-[36] consider this time scale. Each observation is thereafter treated as a time series and characterized. In particular, the path loss $G(d,t)$ is extracted using (7) and its marginal probability density function is estimated using Kernel Density Estimation (KDE) with a Gaussian kernel [47]. Furthermore, we applied a maximum likelihood estimator over each observation to estimate the characterization parameters corresponding to different well known distributions, namely Gaussian, Laplace, Nakagami, Weibull, Rician and Rayleigh.

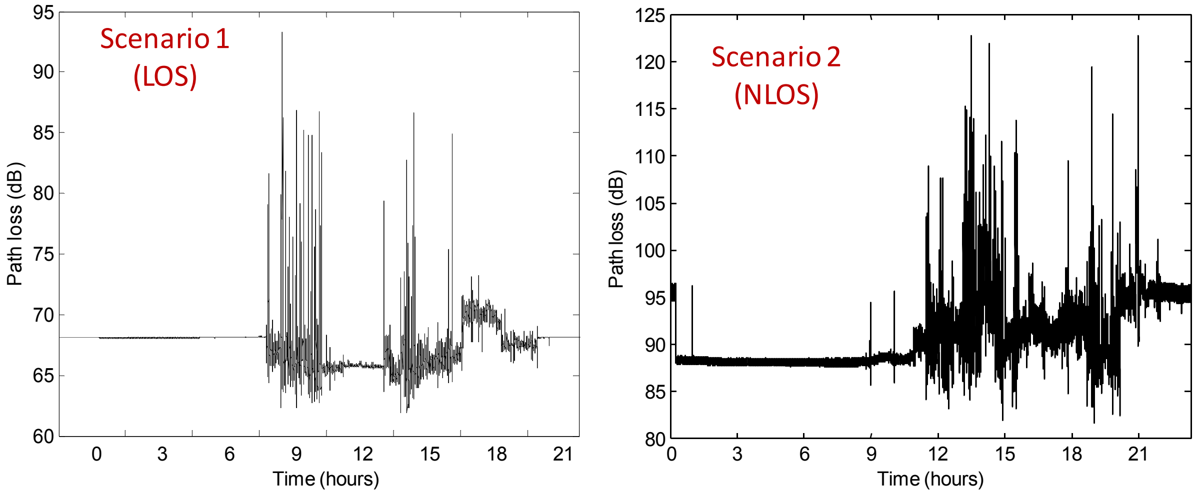


Figure 69: Path loss levels obtained from 24 hours measurements.

In order to find the distribution that fits best with a given time block, we apply a chi-square goodness for the above-mentioned distributions. In addition, we used Matlab *probplot* built-in function [48] that creates a probability fitting plot for the particular distribution along a straight line specified by input arguments of the function. This test is a visual one and is used to ensure that the selected distribution is good enough to represent the collected data. The visual inspection is necessary for the case of observations when the empirical PDF is multimodal (i.e., has multiple peaks) due to change in the average received signal. The distribution with the best fit according the chi-square test is chosen for each block except when the visual test shows irregularity. In this case no distribution will be chosen.

Different scenarios with LOS and NLOS propagation conditions are considered. In addition, two sampling rates were used to study the impact of these factors on the distributions.

6.3.2.1 Scenario 1: LOS Conditions

In Scenario 1, measurements were taken for 24 hours when the spectrum analyzer and the signal generator were at location 1 of Figure 68, reflecting LOS conditions. The measurements were taken on a weekday when 3 to 4 students were either sitting or moving in and out of the rooms. In this scenario we used the Matlab program and a sampling rate of 4 samples/s. Figure 69 illustrates the measured path losses.

The measurement periods was divided into 24 periods of 1-hour duration where 60 observations are collected for each period (i.e., 1 observation per minute). From the obtained measurements, we first noticed that the Rayleigh and the Log-normal distributions show a very poor fit to the observations, which is intuitively clear since we are considering the value of the power in the logarithmic scale. Moreover, the collected data did not follow any of the considered probability distributions in some observations and can be represented only by multimodal probability distribution. In Figure 70, we show the percentage of best resulting fits of several well known distributions. The figure shows that the normal distribution has the highest percentage of best resulting fits during the time period from 22:00 hours in the night to 09:00 hours in the morning with a very small percentage of cases where the best fit was achieved with the Laplace distribution. The mean and the variance of the normal distribution do not change significantly and have values around 58.14 dB and 0.000493 dB, respectively. The occurrence of normal distribution corresponds to periods of no significant changes in the indoor environment due to no or low movement in the surroundings. In this case, the transmitted signal is affected by the White Gaussian noise only.

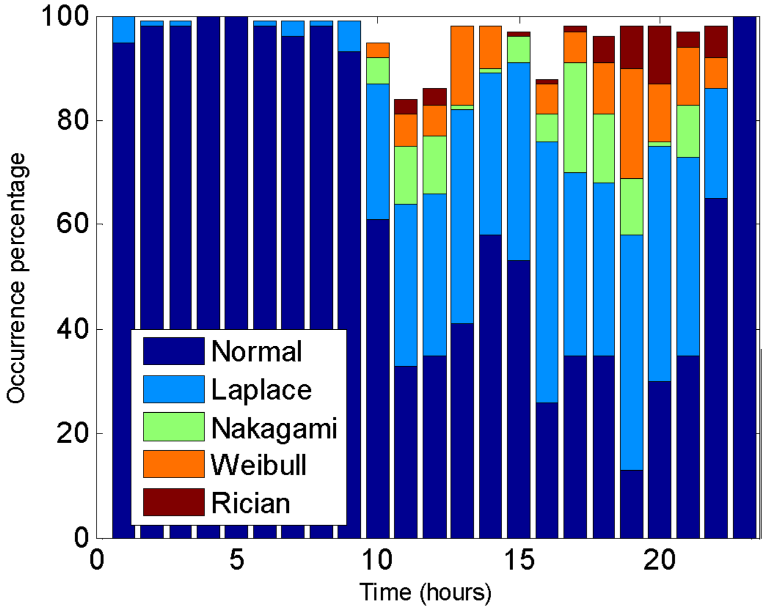


Figure 70: Percentage of best resulting fits of well known distributions for each hour period in scenario 1.

On the other hand, the time period from 09:00 hours onwards manifests a significant increase in the occurrence of the Laplace, Weibull, Rician and Nakagami distributions. This result is due to the increase in the movement around the transmitter and the receiver and is related to the start of normal office working time. Laplace distribution shows the best fit for around 30 % of the observations. Similarly, Weibull, Rice and Nakagami distributions show relatively high percentage compared to the previous period. It should be noted that the normal distribution shows the best fit in many observations where no movement occurs near to the transmitter and the receiver. This fits in part with the results found in [36]. However, we observed a wider variety of distributions than the authors of the reference.

6.3.2.2 Scenario 2: NLOS Conditions

This scenario is similar to scenario 1 except that the spectrum analyzer is now in location 3. The transceivers are separated by two walls: a drywall of width 13 cm and concrete wall of thickness more than 13 cm (Figure 68), representing a NLOS scenario. The measurements were taken on a weekday when 3-4 people are present around the transmitter with 2-3 people occasionally getting in and out of location 3. The presence of personnel inside the laboratory (location 2) also affects the RF propagation in this scenario as location 2 is between location 1 and location 3. Figure 69 illustrates the collected path losses. As shown in the figure, the average level of the received power is 20 dB less than in scenario 1 due to wall penetration losses and higher separation distances. Moreover, the variation due to the noise power is higher (period between 22:00 and 9:00) than in scenario 1 due to the lower received power. Figure 71 shows similar results as scenario 1, except that we noticed an increase in the Laplace and Nakagami distributions with Laplace distribution being the best fit for most of the observations matching approximately 35 % of the cases.

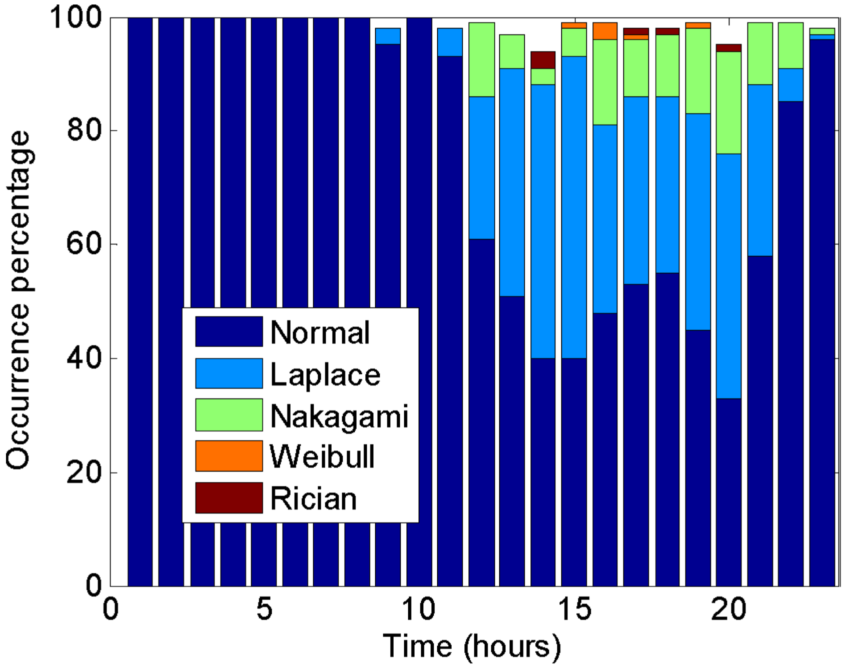


Figure 71: Percentage of best resulting fits of well known distributions for each hour period in scenario 2.

6.3.2.3 Scenario 3: Impact of Sampling Rate

The objective of this scenario is to study the impact of sampling rate on the probability distribution. This scenario represents a LOS environment where the transmitter and the receiver are placed in the same room, i.e. location 4 (see Figure 68). Measurements were taken on weekdays with 5 to 8 people sitting or moving around the transmitter and the receiver premises.

In addition, we used Rohde & Schwarz Tracerecorder application with a sampling rate of 250 samples/s. The duration of the measurements was 5 min. The results are summarized in Table 13. As shown in the table, Laplace distribution occurs most of the time with normal, Weibull, Rician distributions also contributing significantly, which is very comparable to scenario 1 during work time.

Furthermore, we conducted similar measurements during night time when the institute was empty and we found that, in contrary to scenario 1 where normal distribution was found to fit the best this type of scenarios, Weibull distribution was the best fit in this case. This result show the impact of the sampling rate on deciding which is the best probability distribution that fits a given environment.

Table 13: Results from scenario 3.

Distribution Model	**% of cases**	**Parameters**
Laplace	49 %	μ = 70.3, σ= 0.4
Nakagami	25 %	m = 7300, Ω = 905:9
Normal	10 %	μ = 70.3, σ = 0.0852
Rician	9%	K = 42.2 dB
Weibeull	3 %	a = 31.9 , b = 76.9

6.3.3 Qualitative Analysis and Guidelines

We have also conducted a qualitative analysis of the results to better understand the conditions the various distributions arise in. This analysis has shown that Laplace distribution is the best fit for the observations with single major fade occurring due to a movement of a single person near the measurement equipment. Normal distribution is found to be the best fit for the periods of low activity around the transmitter and the receiver during the weekdays and weekends (i.e. idle periods) when measured at a sampling period of 250 ms. However, as the sampling period is reduced to 4 ms, Weibull distribution shows the best fit for idle periods. Rician and Nakagami distributions have shown the best fit for the situations when there is a constant movement around the transmitter and the receiver giving rise to frequent fades in the received power.

We also studied the approach of characterizing the path loss between two fixed nodes with static distributions. Hence, we have considered the selected example depicted in Figure 72. The figure shows the level of the path loss measured for 5 minutes in location 4 with a sampling rate of 250 samples/s. Our objective is to find if, even in this short time, the path loss can be represented with one distribution. In the following we consider the average value and the variance of 2500 samples (i.e., 10 s of measurements). The average value over time is the combination of the distance dependent loss and the slow fading. Since we have a constant distance between the two transceivers, the distance-dependent loss will be constant and any change in the average value is *equal* to the change in the slow fading. In addition, the changes in the variance over time *reflect* the changes, at least, in the parameters of the fast fading distribution. Therefore, the average and the standard deviation are plotted in this figure. As can be seen, the changes in slow fading level can reach 5 dB, which is very high especially in indoor environment. In addition, the changes in the variance can reach 8 dB. Similar observations were also seen in the other measurements. By combining these observations with the results of the previous sections, we can conclude that *characterizing the medium to long term path loss in indoor environment with static distributions of fast and slow fading is an unrealistic assumption*. This observation has been also noted by other researchers [36]. However, they considered only 1 min measurement results and showed that the fast fading amplitude can be modelled with a two state Markov chain. The two states correspond in their model to Rayleigh and Rician distributions. However, from our results we can see that more states, reflecting more distributions, are needed. In addition, we also highlighted the fast changes in the slow fading. It should be also noted that even in some periods of the busy time, the path loss can experience very small changes due to the background noise. This phenomenon appears very frequently and thus, at least the idle state with either normal or Weibull distribution should be added to the two-state model of [36].

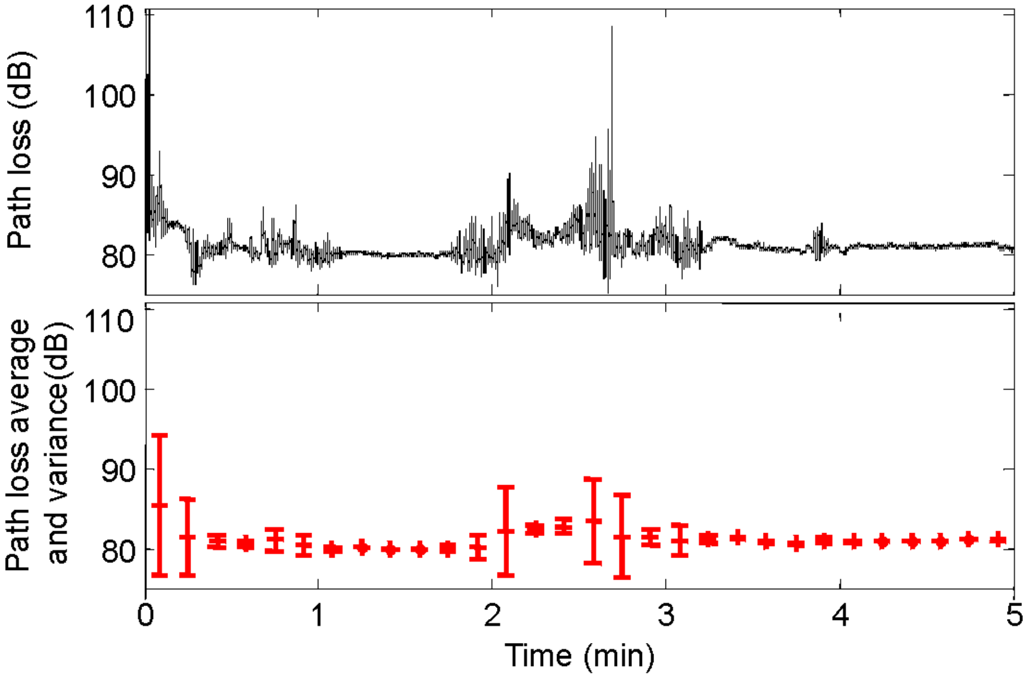


Figure 72: Illustrative example of 5 minutes campaign of path loss together with the average and variance.

In addition, the results have shown that the sampling rate has an impact on the distribution that best fit the path loss. This is very important especially when considering the time scale in the design of a radio resource management technique or when studying its performance in a simulation.

Therefore, we advocate the use of a model where we have two classes of time periods: idle and busy periods.

The idle period refers to the time period when nobody is in the studied indoor environment for a long time (i.e., hours). This corresponds for instance to either weekends or late night in work environment. In this time period, the signal from the transmitter to the receiver is mainly impaired by the additive white Gaussian noise and the path loss can be modelled with a normal distribution. However, when the sampling period of the measurement is low (i.e., few ms), Weibull distribution shows the best fit for the idle periods.

The busy period refers to the time period when there is a human activity in the indoor environment. During this period, the indoor RF channel varies due to the movement of the people which results in the multipath fading in the received signal. This fading can be fast or slow depending on the variations of the received signal. The fading in the received signal including the constant path loss can be best characterized with Laplace, Weibull and Nakagami distributions for both the LOS and NLOS conditions. This period can be modelled in three different ways:

- Use of hybrid model: The multi-model distribution can be characterized with multiple standard distribution models using a combination of different standard distribution functions (i.e., normal, Weibull, etc.). An example of such kind of model for the received electric field strength is given in [40].

- Use of multiple-state model: The multiple-peak PDF can also be described with Markov multi-state model as in [36]. The states in this model will be standard distribution functions. The model will switch between the different states depending on the transition probabilities between the states. These transition probabilities will depend on the indoor RF channel conditions. Figure 73 shows a candidate model for path loss during the busy period. The number of states, their characteristics and transition probabilities P_{tj} will depend on the specific time of the day and the amount of movement in the surroundings of the transmitter and the receiver. Instances of this model can be learned dynamically by the wireless network using similar measurements as we have carried out. However, it is an open question whether the model would be Markov (exponential state holding times), or something more general of the Semi-Markov variety. As we have discussed above, the behaviour of the channel was observed to be highly bursty, with stable periods of almost constant path loss followed by periods of high variability. One of the challenges this imposes on any state space model is the determination of the transition times between the states. It is not clear that the common Markov assumption, leading to exponential state dwell times is appropriate here. Preliminary studies on our data sets indeed indicate that more complex models might be needed, such as, Semi-Markov models with more heavy-tailed dwell time distributions. Figure 74 illustrates this by showing the standard deviation of the path loss, computed for one second intervals over the course of slightly over two hours. The "spikes" in the figure correspond to time instances in which the measured path loss changes drastically due to environmental dynamics. Figure 75 shows the cumulative distribution function of the time intervals between the spikes, computed using a threshold of 2 dB for the standard deviation of the path loss. We see that for this particular example, which is quite typical in our data set, the values are highly variable, and would be very poorly approximated by an exponential distribution.

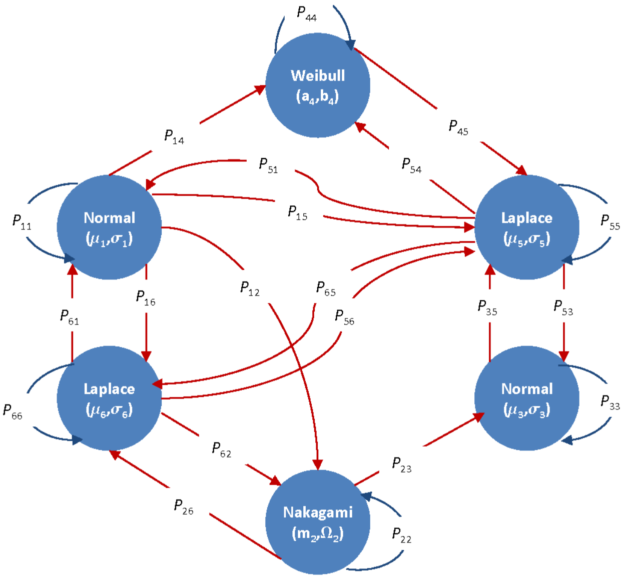


Figure 73: An example of the multiple-state Markov model. Depending on the environment some distributions may appear in several states with different parameters, and transitions between different states may or may not occur.

- Find a temporal distribution for the slow fading: In this case, the fast fading will be separated from the slow fading and therefore the number of states of the Markov chain or the component of the hybrid model can be reduced. However, in this case there is a need to find well fitting temporal distribution for the slow fading also.

In the first two options, the slow and fast fading are combined while in the third, a distribution for each type of fading can be found. We argue that the first two options combining the two types of fading are better for indoor due to the drawback of one more distribution being needed for the third one. The advantage of the hybrid model are (1) it is easier to design and can be used, to some extent, in a universal way like outdoor propagation models, and (2) easier to use in both theoretical or simulation-based performance studies. However, this model does not capture the specific characteristics of each environment and does not reflect the changes in the time domain.

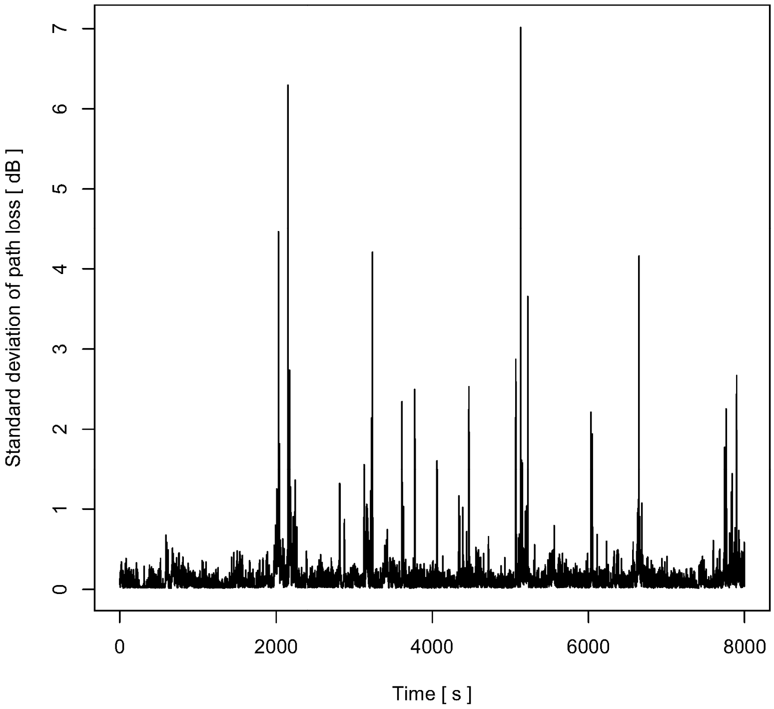


Figure 74: The standard deviation of the path loss, computed for one second intervals over the course of slightly over two hours from scenario 4.

6.3.4 Conclusions

In this section we have shown the results obtained from extensive measurement in indoor environments. The measurements were conducted in order to understand the long-term behaviour of indoor propagation models that are used in most of RRM techniques. The measurement results have shown that the path loss in indoor environment cannot be modelled with static distribution. This makes building indoor REMs much more challenging than generating large-scale outdoor radio environmental maps. It has been a long time known that generating any statistically accurate general indoor models is impossible. However, our early results show that on-site measurements that are tied in together with on-line modeling capabilities of REMs could produce usable and reasonably accurate models on the environment which can be used to network management and even real-time radio resource management. Therefore we advocate that there is a need to further develop either dynamic propagation models with several states representing the different distributions or a hybrid model including several distributions. The results have also shown that different distribution should be used depending on the sampling rate, which is directly connected to the time scale of the used RRM techniques.

One of the key applications of the models developed form these measurements would be model-based optimization and control of radio resources. Since the results show that the distributions of path losses seem to follow a stochastic process with distinct changes but also periods of stability, we believe approaches based on stochastic filtering techniques can be used to classify prevailing propagation conditions, and also detect changes in those. The estimate of currently valid distributions can, for example, directly be applied for interference minimization. However, more work is clearly needed to quantify the prediction accuracy that can be achieved with such filtering approaches, as well as the achievable gains in the overall system capacity.

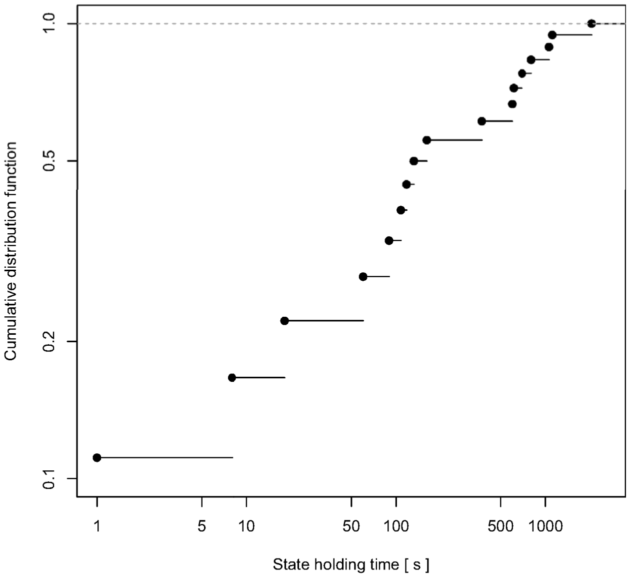


Figure 75: The cumulative distribution function of the time intervals between the spikes in the standard deviations of the path loss depicted in Figure 74.

7 Conclusions

In this document we provided an overview and status report on the extensive spectrum measurement campaigns carried out in the project. We also gave and discussed first results and conclusions drawn from those data sets, especially in relation to the large collaborative campaigns carried out both within the consortium, and together with other projects. In addition to statistical characterization of spectrum use across Europe, we discussed our work towards developing new spectrum use models based on the gathered data sets. As case studies we showed new approaches for coverage prediction in wireless systems based on spatial spectrum use models, discussed the interplay between spatial and time domain models and the impact of long-term variations of spectrum use on those, and also discussed research challenges involving indoor propagation models for femtocells. We are currently integrating selected models that have emerged as outcomes of this work into the overall REM data model used also in the prototype, and study their performance and applications as part of the field trial work that will take place during the remaining half a year of the project lifetime.

The measurements and modeling done as a part of the FARAMIR project shows that the field is finally rapidly reaching a level of maturity where measurement campaigns and networks can be deployed in a more systematic manner than previously, and rather accurate models can be generated semi-automatically based on the results. These methods developed and measurements done alone have high scientific and regulatory policy relevance. As the methodology is now maturing and REM-type of systems are shown to have a high potential for more efficient radio resource management, there is an urgent need to make final R&D steps to bring the technology to its full real-world fruition and deployment. The measurements reported in the present document are still based on relatively low density spatial sampling and hand-picked locations. It is clear that the effect of spatial sampling and a sort of crowd-sourcing should be considered to understand better the opportunities and limitations of different spectrum measurement campaigns.

The FARAMIR project is also firmly committed to giving open access to the gathered measurement data for the research community. Currently the project is planning to release the gathered measurement data sets together with an extensive documentation by summer of 2012 through a dedicated web portal accessible from the main project website. The data sets provided will, of course, include the outcomes of the major collaborative campaigns discussed here, but will also include several of the data sets gathered as parts of smaller, focused measurements by individual partners. We believe this collection of measurement data will be a highly valuable asset for the larger scientific community, enabling both the validation and verification of earlier results, as well as discovery and development of new ones.

References

[1] M. Buddhikot, "Understanding Dynamic Spectrum Access: Models, Taxonomy and Challenges", Proceedings of IEEE DySPAN 2007, Dublin, Ireland, April 17-21, 2007.

[2] M.Wellens and P.Mähönen, "Lessons Learned from an Extensive Spectrum Occupancy Measurement Campaign and a Stochastic Duty Cycle Model," *Springer Mobile Networks and Applications*, 2009.

[3] M. Islam, C. Koh, S. Oh, X. Qing, Y. Lai, C. Wang, Y.C. Liang, B. Toh, F. Chin, G. Tan and W. Toh, "Spectrum Survey in Singapore: Occupancy Measurements and Analyses," *Proc. Of International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, Singapore, May 2008.

[4] W. K. Pratt*, Digital Image Processing*, John Wiley & Sons, 2001.

[5] A. J. Petrin, "Maximizing the Utility of Radio Spectrum: Broadband Spectrum Measurements and Occupancy Model for Use by Cognitive Radio," *PhD thesis*, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA, August 2005.

[6] Radio communication Bureau, "Handbook on Spectrum Monitoring," International Telecommunication Union (ITU), 2002.

[7] M. Pavloski, V. Atanasovski, L. Gavrilovska, "Estimation of decision threshold in energy based spectrum sensing*," 10th International Conference ETAI*, September 2011.

[8] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems Man and Cybernetics*, vol. SMC - 9, pp. 62–66, January 1979.

[9] C. Ghosh, S. Pagadarai, D.P. Agrawal, A.M. Wyglinski, A.M., "A framework for statistical wireless spectrum occupancy modelling", IEEE. Trans. On Wireless Communications, Vol. 9, nº 1, pp.38-55, January 2010.

[10] O. C. Ibe, Markov processes for stochastic modelling, Academic Press, 2009.

[11] M. López-Benítez, F. Casadevall, "Empirical time-dimension model of spectrum use based on discrete-time Markov chain with deterministic and stochastic duty cycle models," IEEE Transactions on Vehicular Technology, vol. 60, no. 6, pp. 2519-2533, Jul. 2011.

[12] Z. Wang and S. Salous, "Spectrum occupancy statistics and time series models for cognitive radio," Journal of Signal Processing Systems, vol. 62, no. 2, pp. 145–155, Feb. 2011.

[13] D. Chen, S. Yin, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: a large scale spectrum measurement study," in Proceedings of the 15th ACM Annual International Conference on Mobile Computing and Networking (MobiCom 2009), Sep. 2009, pp. 13–24.

[14] V. Blaschke, H. Jaekel, T. Renk, C. Kloeck, and F. K. Jondral, "Occupation measurements supporting dynamic spectrum allocation for cognitive radio design," in Proceedings of the 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom 2007), Aug. 2007, pp. 50–57.

[15] A. Papoulis and S. U. Pillai, Probability, random variables, and stochastic processes, 4th ed. Boston: McGraw-Hill, 2002.

[16] P. Kumaraswamy, "A generalized probability density function for double-bounded random processes," Journal of Hydrology, vol. 46, no. 1-2, pp. 79–88, Mar. 1980.

[17] M. C. Jones, "Kumaraswamy's distribution: A beta-type distribution with some tractability advantages," Statistical Methodology, vol. 6, no. 1, pp. 70–81, Jan. 2009.

[18] S. Geirhofer, L. Tong, and B. M. Sadler, "A measurement-based model for dynamic spectrum access in WLAN channels," in Proceedings of the IEEE Military Communications Conference (MILCOM 2006), Oct. 2006, pp. 1–7.

[19] S. Geirhofer, L. Tong, and B. M. Sadler, "Dynamic spectrum access in WLAN channels: Empirical model and its stochastic analysis," in Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum (TAPAS 2006), Aug. 2006, pp. 1–10.

[20] S. Geirhofer, L. Tong, and B. M. Sadler, "Dynamic spectrum access in the time domain: Modeling and exploiting white space," IEEE Communications Magazine, vol. 45, no. 5, pp. 66–72, May 2007.

[21] L. Stabellini, "Quantifying and modeling spectrum opportunities in a real wireless environment," in Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC 2010), Apr. 2010, pp. 1–6.

[22] M. Wellens, J. Riihijärvi, and P. Mähönen, "Empirical time and frequency domain models of spectrum use," Physical Communication, vol. 2, no. 1-2, pp. 10–32, Mar. 2009.

[23] M. López-Benítez, F. Casadevall, "Modeling and simulation of time-correlation properties of spectrum use in cognitive radio," Proceedings of the 6th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom 2011), Jun. 2011, pp. 1-5.

[24] L. Devroye, "Non-uniform random variate generation, "Springer-Verlag, 1986

[25] M. Wellens, J. Riihijärvi, and P. Mähönen, "Modelling primary system activity in dynamic spectrum access networks by aggregated ON/OFF-process", in Proceedings 4th IEEE Workshop on Networking Techs. For Software Defined Radio Networks (SDR 2009), pp. 1-6, June 2009

[26] M. López-Benítez, F. Casavadell, D. López-Pérez and A. V. Vasilakos, "Modeling and simulation of joint time-frequency properties of spectrum usage in cognitive radio," in Proceedings of the 4th International Conference on Cognitive Radio and Advanced Spectrum Management (CogART 2011), Invited Paper, Barcelona, Spain, October 2011, pp. 1-5.

[27] T. Yücek, H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications", IEEE Communication Surveys and Tutorials , vol 11, nº 1 ,pp 116-130, First Quarter 2009.

[28] M. López-Benítez, F. Casadevall, "Spatial duty cycle model for cognitive radio," in Proceedings of the 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2010), Sep. 2010, pp. 1631-1636.

[29] M. López-Benítez, F. Casadevall, "On the spectrum occupancy perception of cognitive radio terminals in realistic scenarios," in Proceedings of 2nd IAPR International Workshop on Cognitive Information Processing (CIP 2010), June. 2010, pp. 1-6.

[30] M. Wellens, J. Riihijärvi, M. Gordziel, P. Mähönen, "Evaluation of Cooperative Spectrum Sensing based on Large Scale Measurements", Proc. of IEEE DySPAN 2008, Chicago, USA, October 2008.

[31] M. Wellens, J. Riihijärvi, and P. Mähönen, "Spatial Statistics and Models of Spectrum Use", Elsevier Computer Communications, vol. 32, no. 18, pp. 1998-2011, December 2009.

[32] Molisch, L. Greenstein, and M. Shafi, "Propagation issues for cognitive radio," Proceedings of the IEEE, vol. 97, no. 5, pp. 787 –804, May 2009.

[33]C. Phillips, D. Sicker, and D. Grunwald, "Bounding the error of path loss models," 2011 IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), May 2011, pp. 71 –82.

[34]R. J. Bultitude, "Measurement, characterization and modeling of indoor 800/900 mhz radio channels for digital communications," IEEE Communications Magazine, vol. 25, no. 6, pp. 5 – 12, June 1987.

[35]E.Walker, H.-J. Zepernick, and T.Wysocki, "Fading measurements at 2.4 ghz for the indoor radio propagation channel," in the International Zurich Seminar on Broadband Communications, 1998, pp. 171 –176.

[36]J. A. Roberts and J. R. Abeysinghe, "A two-state Rician model for predicting indoor wireless communication performance," Journal of Applied Mechanics, 1955.

[37]H. Hashemi, "The indoor radio propagation channel," Proceedings of the IEEE, vol. 81, no. 7, pp. 943 –968, July 1993.

[38]L. Maciel, H. Bertoni, and H. Xia, "Unified approach to prediction of propagation over buildings for all ranges of base station antenna height," IEEE transactions on vehicular technology, vol. 42, no. 1, pp. 41–45, 1993.

[39]A. Medeisis and A. Kajackas, "On the use of the universal okumura-hata propagation prediction model in rural areas," in IEEE 51st Vehicular Technology Conference (VTC-Spring 2000), May 2000, pp. 450 – 453.

[40]J. H. Tarng, W.-S. Liu, Y.-F. Huang, and J.-M. Huang, "A novel and efficient hybrid model of radio multipath fading channels in indoor environments," IEEE transactions on Antennas and Propagation, vol. 51, no. 9, March 2003.

[41]R. Vaughan and J. B. Andersen, "Channels, propagation and antennas for mobile communications," IEE press, 2003.

[42]Alcatel-Lucent, picoChip Designs and Vodafone, "Simulation assumptions and parameters for FDD HENB RF requirements," R4-092042, 3GPP TSG RAN WG4 (Radio) Meeting # 51, San Francisco, 2009.

[43]Agilent Technologies, "Agilent signal generators." [Online]. Available: www.home.agilent.com

[44]Rohde and Schwarz, "R and S FSL spectrum analyzer." [Online]. Available: www2.rohdeschwarz.com

[45]Antennen technik Bad Blankenburg. [Online]. Available: http://www.antenne-ag.de/

[46]M. Gordziel, "The distributed spectrum sensing GUI," in Evaluation of Time-Synchronized Spatially Distributed Radio Spectrum Occupancy Measurements. M.S. thesis, Department of Wireless Networks RWTH Aachen University, December 2008, pp. 95–96.

[47]A. W. Bowman and A. Azzalini, Applied Smoothing Techniques for Data Analysis. Oxford Science Publications, 1997.

[48]MATLAB, "Matlab probplot function." [Online]. Available: http://www.mathworks.com