POLITECNICO DI MILANO

Scuola di Ingegneria Industriale e dell'Informazione

Corso di Studi in Ingegneria Matematica

Tesi di Laurea Magistrale

Geostatistics in Cognitive Radios

RELATORE: Prof. Piercesare Secchi

CORRELATORI: Prof. Visa Koivunen,
Dott. Ing. Jan Oksanen,
Dott. Ing. Sachin Chaudhari

Tesi di Laurea di:
Irene Rognoni
Matr. 782003

ANNO ACCADEMICO 2013/2014

Alla mia Famiglia,
per tutto.

ABSTRACT

Cognitive radios constitute a very interesting application, even though still at a conceptual stage, they have the potential of making a difference in how the radio spectrum is accessed and are therefore been defined as a *destructive, but not obtrusive technology* [29]. They represent a unique combination of multi-disciplinary and complex competences from fields such as communications, statistics, optimization, game theory, electronics and robotics

In this thesis state of the art technology and the effects it will cause once commercialized are presented, even though most of the work is focused on Spatial Statistics. The aim is to study techniques for interpolation, prediction and classification that could be useful in this peculiar field, that results to be determinant because of the computational limitations. These are mostly caused by the constraints on the implementation area and on batteries life due to a fully distributed sensing approach.

In the present work theoretical models of signal propagation are introduced and techniques for spectrum analysis are outlined, mostly focusing on the cyclostationary sensor that has been used in our measurement campaign around Helsinki (Finland). This allows to make distributional hypotheses on the measured signal, that are afterwards verified.

Numerous are the interpolation and prediction techniques used in this work, from the most classical ones (Voronoi, Thiessen, Natural Neighbour, Inverse Weighted Distance, Trend Surfaces, Splines), until Kriging, always paying particular attention to computational costs and algorithmic robustness.

Last but not least a Bayesian model has been fit not just for prediction itself, but to verify the correctness of theoretical models for signal transmission and consequently validate their usefulness in environments where not enough sample points are available to construct reliable prediction maps.

A brief parenthesis has been dedicated to sampling optimization and its benefit to the quality of obtained data.

Supervised classification techniques are developed, with very encouraging results, ideal in a completely decentralized approach to decisions.

The aim of acquiring occupation data from multiple bands almost in real time and consequently building a dynamic network cartography, predicting and revealing anomalies, granting quality and stability of transmissions without overloading the network is a fundamental issue for technological development; for this reason parsimonious, but effective and robust algorithms are needed.

SOMMARIO

Le radio cognitive costituiscono un campo applicativo molto interessante sebbene ancora ad uno stato concettuale, in quanto portano in sé il potenziale di fare la differenza nei metodi di accesso dello spettro radio e sono pertanto state definite una tecnologia *distruttiva, ma non invasiva* [29]. Esse rappresentano una combinazione unica di competenze davvero multidisciplinari e complesse da campi quali le comunicazioni, la statistica, l'ottimizzazione, la teoria dei giochi, l'elettronica e la robotica.

In questa tesi è presentato uno scorcio sullo stato dell'arte della tecnologia e sugli effetti che questa comporterà una volta commercializzata, anche se la maggior parte del lavoro si concentra nell'ambito della Statistica Spaziale. L'obiettivo è studiare tecniche di interpolazione, predizione e classificazione utili in questo particolare campo applicativo. L'applicazione comporta infatti notevoli limitazioni computazionali dovute ad una ridotta area di implementazione e alla durata delle batterie, soprattutto in un'ottica di analisi del campo completamente distribuita.

Nel presente lavoro vengono trattati i modelli teorici di propagazione del segnale e ci si sofferma sulle tecniche di analisi dello spettro, sui vari tipi di sensori a disposizione, in particolare su quello ciclostazionario utilizzato nelle campagne di misurazione nell'area di Helsinki (Finlandia). Questo permette di fare ipotesi distribuzionali sul segnale misurato che vengono puntualmente verificate.

Numerose sono le tecniche di interpolazione e predizione utilizzate, dalle più classiche (Voronoi, Thiessen, Natural Neighbours, Distance Weighted Interpolation, Trend Surfaces, Splines), fino al Kriging, sempre prestando attenzione al costo computazionale e alla robustezza algoritmica.

In ultimo un Modello Bayesiano è stato fittato, non a semplice fine predittivo, ma per verificare la correttezza di modelli teorici per la trasmissione e validarne l'utilizzo in ambienti in cui il numero di campionamenti non è sufficiente alla costruzione di mappe predittive.

Una breve parentesi è dedicata all'ottimizzazione del campionamento e ai suoi benefici in termini di qualità dei dati raccolti.

In ultimo sono sviluppate delle tecniche di classificazione supervisionata, con risultati decisamente incoraggianti, ideali nell'ipotesi di un approccio alle decisioni di tipo puramente decentralizzato.

L'obiettivo di acquisire dati di occupazione delle varie bande quasi in tempo reale e conseguentemente costruire una cartografia dinamica, che riveli eventuali anomalie, permetta di fare previsioni, garantendo la qualità e la stabilità delle trasmissioni senza oberare la rete è di fondamentale importanza per lo sviluppo tecnologico; per questo algoritmi parsimoniosi, ma al contempo efficaci e robusti sono necessari allo sviluppo tecnologico.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The present work has been carried out in the Signal Processing Department, that belongs to the Faculty of Electronics, Communication and Automation of the Aalto University, Helsinki, Finland.
First of all I would like to thank Professor Visa Koivunen for proposing this Master Thesis, that allowed me to work in a really awesome research group, for the encouragements and the guidance.
This work would not have seen the light without Prof. Piercesare Secchi, who gave me freedom and autonomy, of which I am really thankful.
I would like to express my sincere gratitude to my instructor Jan Oksanen, for being the super-nice and caring person he is and for accompanying me all the way through the project.
I really felt privileged for working in such a nice team, so thanks to Sachin Chaudhari, for his support and his kind words, thanks to Prof. Ollila, without whom this thesis would not exist, thanks to Jaakko Ojaniemi for being the best co-worker ever, thanks really to everybody: Semu, Markku, Joona, you made me feel at home at my work place.
Thanks to Mirja Lemetyinen and Niina Huovinen as well for taking care of all the administrative issues.
I am sincerely grateful to all the professors and doctoral students that shared a bit of their knowledge with me, specially to Prof. Guglielmi and Dott. Argiento.
I would like to thank my family as a whole, I know how hard it has been for you, thank you for supporting me, no matter what.
I would like to thank all my "colleagues", these years would have been a lot longer without all of you. Thanks to my friends in the village, I know you have always been there for me when I needed you.
Last, I would like to thank all my friends in Finland and abroad; thank you SIK and SIK International for giving me a glimpse of the Teekkari culture, I hope I can be a good ambassador of it.
This all turned out a lot more formal and long than I expected it to be, but I guess there are moments in which formality is important.
I do not really understand why I am so lucky, thank you all amazing people out there!

RINGRAZIAMENTI

Il presente lavoro è stato realizzato nel Dipartimento di Elaborazione Digitale del Segnale, che appartiene alla Facoltà di Elettronica, Comunicazioni e Automazione di Aalto University, Helsinki, Finlandia.
Prima di tutto vorrei ringraziare il Professor Visa Koivunen per avermi proposto questa Tesi, che mi ha permesso di lavorare in un meraviglioso gruppo di ricerca, per gli incoraggiamenti e la guida.
Questo lavoro non avrebbe visto la luce senza il Professor Piercesare Secchi, che mi ha dato libertà e autonomia, di cui sono realmente riconoscente.
Vorrei esprimere la mia sincera gratitudine al mio correlatore Jan Oksanen, per essere la persona meravigliosa che è e per avermi accompagnato lungo tutto il progetto.
Mi sento davvero privilegiata per avere lavorato in un gruppo così affiatato, quindi grazie a Sachin Chaudhari, per il suo supporto e le sue parole di incoraggiamento, grazie al Prof. Ollila, senza cui questa tesi non esisterebbe, grazie a Jaakko Ojaniemi per essere il miglior collaboratore possibile, grazie davvero a tutti: Semu, Markku, Joona, mi avete fatto sentire a casa.
Grazie inoltre a Mirja Lemetyinen e a Niina Huovinen per essersi prese cura della parte amministrativa.
Grazie a tutti i Professori e ai Dottorandi che mi hanno supportato al PoliMi, tra questi grazie soprattutto alla Prof.ssa Guglielmi e al Dott. Argiento.
Vorrei ringraziare la mia famiglia nel complesso, so quanto sia stato difficile per voi, grazie per il supporto incondizionato.
Grazie inoltre a tutti i miei colleghi, questi anni sarebbero stati molto più lunghi senza tutti voi.
Grazie alle storiche amiche del paese per esserci sempre state quando avevo bisogno.
In ultimo vorrei ringraziare tutti gli amici in Finlandia e all'estero; grazie a SIK e SIK International per avermi dato uno scorcio sulla cultura del Teekkari, spero di poterne essere una buona ambasciatrice.
Tutto questo è risultato più lungo e formale di quanto non avessi preventivato, ma immagino esistano momenti in cui la formalità sia necessaria.
Non mi capacito del fatto di essere così fortunata, grazie a tutti voi!

1 AN INTRODUCTION TO COGNITIVE RADIOS

The electromagnetic radio spectrum is a highly valued *natural resource*, the use of which, by transmitters and receivers, is licensed by governments [29].
Due to the huge growth of wireless technologies (in devices such as cellphones, TVs, garage door openers ...), a phenomenon known as *wastage of spectrum* [63] is becoming an important issue: some bands have become more and more crowded, even though some of the licensed ones are heavily underutilized, according to [29] only ten percent of the assigned electromagnetic radio spectrum is effectively in use.
Two main historical reasons can be identified: government policies and state of the art radio technology.

1.1 NEW POLICIES

Since the first regulations about Radio Frequency (RF) transmissions in the 20^{th} century a *static allotment policy*, granting exclusive use to licensees, was introduced. Artificial limits have caused massive inefficiency, instead of promoting harmony of the airwaves. Spectrum utilization is time and place dependant and fixed allocation prevents rarely used frequencies (those assigned to specific services, such as military and amateur radio) from being used by unlicensed users.
This extract is taken from a report prepared by the Spectrum-Policy Task Force, part of the Federal Communications Commission (FCC), as early as November 2002:

> There is some evidence indicating that the shortage of spectrum is often a *spectrum access problem*. That is, the spectrum resource is available, but its use is compartmented by traditional policies based on traditional technologies. New radio technologies may enable new techniques for access of spectrum and sharing of the spectrum resources that may create quantum increases in achievable utilization [20].

Regulatory bodies in the world have finally started considering allowing unlicensed users to transmit in licensed bands if no interference is caused; according to numerous experts, between whom Staple and Werbach in [63], spectrum normative reform should:

- reallocate spectrum from government and long standing users to new services,
- introduce the possibility of spectrum leases, granting the licensees the right to hire their spectrum share to third parties,
- allocate an amount of spectrum for unlicensed and shared services,
- allow the national authorities for spectrum usage to impose power limits with the aim of avoiding a cacophony of unlicensed devices.

1.2 ADVANCES IN TECHNOLOGY

Technology is evolving, nowadays technical limits that made *guard bands*[1] necessary have been overtaken and this has made regulations clearly outdated. Numerous are the improvements in technology that will or already make a difference, between these some are:

SHIFT FROM ANALOGICAL TO DIGITAL TRANSMISSIONS, for example TV transmissions have already switched both in Finland and in Italy, improving spectrum availability, since at least five TV shows can be broadcast on the same frequencies occupied by a single analogical channel and similarly digital cellular systems carry three times as many phone calls as their analogical predecessors, see e.g. [63].

NEW WIRELESS DEVICES TO SHARE THE SPECTRUM BETWEEN MULTIPLE SYSTEMS, high-power, undifferentiated narrow band transmissions are going to be substituted with modern, low power, coded wideband signals. Ultra-Wide Band (UWB) technology brings this idea to the acme, since transmissions are so low power that licensed services almost do not notice an increase in interference. In this type of technology safeguards policy is of course of major importance.

SWITCH FROM STATIC TO DYNAMIC PROTOCOL, that will allow cooperative systems, according to numerous experts, between whom Staple in [63], to cause the greatest technological gain in wireless capacity. The traditional Static Spectrum Access (SSA) will be substituted by the Dynamic Spectrum Access (DSA) protocol, therefore enabling every device to use some of the network capacity, but also to add capacity back, so that devices could be added to the network almost without an impact on interference. Cognitive Radio (CR) is the (theorical) implementation of this new generation of devices in the field of radio technology: the idea is to analyze nearby transmissions in order to identify currently unemployed bands and assign them on the fly to unserviced users.

Regulatory and technical developments suggest that an era of *spectrum abundance* is just behind the corner and the market of communication services will be reshaped.

1.3 SPECTRUM ACCESS MODELS

The typologies of spectrum access devised so far in literature can be classified into:

DYNAMIC EXCLUSIVE MODEL, in which frequency bands are reserved for the exclusive use of a particular service, but providers can share the spectrum.

OPEN SHARING MODEL, that allows all users to access the spectrum equally, given some constraints on the transmit signal. This approach is already in use in Industrial Scientific and Medical (ISM) radio bands[2].

HIERARCHICAL ACCESS MODEL, that establishes a priority structure between a primary system and all the unlicensed systems, which are bound to access the spectrum without causing harmful interference to the license owner, whose Quality of Service

1 Unused frequency range between bands, intended to ensure transmission without interference.

2 Portion of the spectrum reserved for purposes other than telecommunications, such as microwave ovens and short-range low power communications systems like cordless phones, garage openers, Bluetooth devices and wireless computer networks.

(QOS) is granted. In the following part of this work, if not differently stated, the hierarchical model will be the intended spectrum access model.

1.4 HIERARCHICAL ACCESS MODEL

This type of access can be further declined into different protocols, determining the pairwise behaviour of Primary User (PU) and Secondary User (SU):

SPECTRUM UNDERLAY, in which the SU keeps its transmit power at such a low level that the primary receiver sees only a slight increase in of its effective noise level, even though the transmit spectrum might overlap. A typical example is UWB communications: this technique has strict constraints on the transmitted power of the SUs, therefore being the most conservative approach.

SPECTRUM INTERWEAVE, in which the SU tries to identify and transmit at times, locations and frequencies where PUs are not active, the secondary system thus has to perform first spectrum sensing. It is noteworthy that, as well evidenced in [46], a secondary system can usually only sense the presence of a primary transmitter, not that of a receiver, possibly causing unwanted interference. For an insight refer to [25].

SPECTRUM OVERLAY, a rather different approach: the secondary system identifies the propagation channel from both primary and secondary transmitters to the primary receiver, assuring that an eventual secondary transmission would not reduce the quality of the primary one, even when they both occur in the same frequency band. There is no need of artificial limit to the power of the SU, but knowledge of the complex impulse response over the relevant transmit-receive paths (e.g. between primary and secondary transceivers) is required, thus the need for more computation power.

Please refer to [76] for a finer classification of spectrum accesses.
The last two approaches have in common the necessity to make reliable channel predictions and to understand what is the achievable accuracy, this will be further analysed in this thesis work.

1.5 SOFTWARE DEFINED RADIOS

Let's face the actors that will be fundamental in the new DSA environment.
A PU is the licence holder of the spectrum band, while the SU is one of the agents that are allowed to transmit, provided they do not cause any type of interference to the PU; the contract between the two types of agents will depend on regulation and on the chosen spectrum access model.
A Software Defined Radio (SDR), proposed by Mitola in 1991, is a

> Multiband radio that supports multiple air interfaces[3] and protocols and is reconfigurable through software run on Digital Signal Processor (DSP) or general purpose microprocessors[44].

A CR is considered as a goal towards which a SDR platform should evolve: a fully reconfigurable wireless transceiver which automatically adapts its communication parameters

3 Radio-based communication links between the mobile station and the active base station.

to network and user demands, exchanging information about the environment with the networks and other CRs. See, between others [76].

1.6 COGNITIVE RADIOS

The concept of CR has been described as:

> An intelligent wireless communication system that is *aware* of its environment and uses the methodology of understanding-by-building to learn from the environment and *adapt* to statistical variations in the input stimuli, with two primary objectives in mind:
>
> - highly reliable communication whenever and wherever needed,
> - efficient utilization of the radio spectrum, see for example [29].

Cognitive radios can be divided into two important categories, according to the quantity of parameters that can be adapted to the current environment:

FULLY CR (MITOLA RADIO), the very first idea of CR, it would allow the setting of numerous parameters on the fly: the modulation format, the access method and the coding; it is still mostly a theoretical model, see the original work: [45].

SPECTRUM-SENSING OR SPECTRUM-AGILE CR, only allows fine-tuning of center frequency, bandwidth and transmission time. Most of the research is now focusing in this category of CR, particularly in the TV bands, see [46].

Numerous are the tasks to be implemented to obtain the features needed for a cognitive radio, between these:

EFFICIENT SPECTRUM SENSING TECHNIQUES, to provide continuously reliable spectrum monitoring and detect unused spectrum with a short sensing time. The identification of ideal moments for transmission is performed in a passive manner at the receiving end of a cognitive radio link by listening to incoming electromagnetic waves in the local neighborhood of the receiver, determining whether the observed signal can be explained by noise alone or whether it indicates the presence of a primary signal.

POWER CONTROL, to manage different QOS requirements, setting the cut-off level in Signal Noise Ratio (SNR) for the protection of the PUs and sculpting the spectrum adaptively, satisfying a variety of technical and legal constraints,

FEEDBACK CHANNEL, connecting SUreceiver and transmitter and thereby allowing for the transmitter to adapt itself to the environment, it operates at a low bit rate compared to the forward link from the transmitter to the receiver.

These tasks form what Mitola in [45] has defined a *cognitive circle*, a closed-loop wireless system aiming at maintaining harmony between transmitters and receivers through a dynamic spectrum management. First the radio-scene analyzer detects moments for transmission in time and space, then the transmit power controller selects a modulation strategy that adapts to the radio environment, trying to assure reliable communication across the channel.
A theoretical multi-user communication problem is therefore outlined; two are the stringent limitations on network resources: constraints on maximum interference imposed by regulatory agencies and the availability of spectrum to communicate. This difficult prob-

lem has been tackled in literature in two ways:

- as a non-cooperative stochastic game involving multiple players, a similar problem as path-planning in robotics, for an introduction see [26], [54] presents an adaptation of the potential field method for CRs purposes, based on a routing cost metric.
  The convergence to the Nash equilibrium is quite fast, however, once this is reached, no user is permitted to change its power level in a unilateral manner. Unfortunately, because of noncooperativity and therefore "selfishness", the algorithm performance could be seriously compromised in certain situations, for this reason expansion of the algorithm structure to include a form of cognitive immunity against exploitation have been proposed.

- as an iterative water-filling algorithm, with the object of jointly maximizing data transmission rates, with the same constraints in produced interference. This algorithm is particularly well suited for the context, because a sub-optimal solution is reached in an autonomous manner, thereby avoiding explicit communication links among the multiple users and therefore simplifying the network design.

It has been shown in [30] that the solution of the iterative water-filling, although sub-optimal and of local nature, is equivalent to a Nash equilibrium.

1.7 SPECTRUM HOLES

One of the major aims of CRs is promoting an *efficient* use of the electromagnetic spectrum, this is achieved through an optimal management of *spectrum holes* in space and time in which a frequency resource results not in use. If the position of the PUs and the time of transmission were known *a priori*, spectrum holes would be known, unfortunately this is not always the case.
Every spectral hole is slotted in time, the cycle is visualized in Figure 1: when the PU stops transmitting and the channel becomes idle, the SU needs a period of time, called *sensing time*, to detect the spectrum hole; when this is found a new transmission is in act until the SU notices that the PU has started transmitting again and is creating interference. When the primary transmission is sensed the secondary one is not on the air until the PU leaves the channel empty once again and the cycle restarts. The necessity of minimizing the sensing times points to the importance of accurate sensing algorithms.

Figure 1: Main stages of the circular process that a spectrum hole faces in time.

Various typologies of spectrum holes exist, these are classified in [29] as:

TEMPORAL SPECTRUM HOLES, no signal is emitted in the band of interest for a fixed period of time.

FREQUENCY SPECTRUM HOLES, frequency bands in which a SU can transmit without interfering with any PUs across all frequencies. Problem is that transmissions could

potentially interfere with primary receivers without noticing. These holes, as temporal ones, are mostly researched for in an overlay approach.

SPATIAL SPECTRUM HOLES, created when the Primary Transmitter (PT) uses the frequency band only in a restricted space inside the coverage area, these holes are mostly taken advantage of in an underlay approach. They can be further divided into: *black*, *grey*, *white* spaces, depending on the severity of the perceived interference. Unlicensed users can transmit only if white and grey holes are detected and provided they do not interfere with PUs in the coverage area, for an insight see [64].

To minimize the harmful interference caused to PUs, the area of transmission can be divided in sub-areas, exemplified in Figure 2 with circles around the PU. Primary receivers are guaranteed an error free reception and a specific QOS, within the protection range denoted by r_prot, that defines the innermost circle, the *ptotection circle*. The radius r_{pcov} delimits the maximum distance from which a signal emitted by the PU can be properly decoded. r_n, *non talk radius*, defines the circle where SUs are not allowed to transmit, to assure an adequate quality of transmission inside the protection area. The circular crown, in red in the figure, is called *sacrificial zone*: inside it users can experience some interference when the secondary system is switched on. Together with the protection zone it forms he *coverage area*.

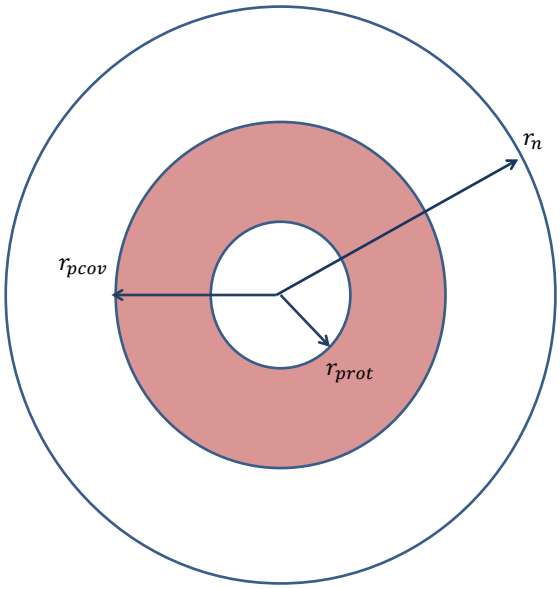

Figure 2: Exemplification of a transmission area.

2 PROPAGATION MODELS OF WIRELESS SIGNAL

A transmission strategy for the secondary system consists of important decisions such as: Which secondary user can transmit? On which frequency band? For how long? What is the maximum allowed power? What kind of data modulation is to be used? How much power the intended receiver gets and to how much interference is a "victim receiver" subject to? How often potential interference levels have to be estimated and, consequently, transmission strategies have to be updated?
A key ingredient in the answers to these questions is the use of a realistic propagation model.
A propagation model simplifies the physics behind electromagnetic propagation into a set of practical equations that can be then used to describe the spatial occupancy of a band, practically taking care of the channel estimation and of its time variations *a priori* from any eventual measurement data.

2.1 PATH LOSS

The field strength varies as a function of location, due to a wide number of factors: transmitter distance, refraction, diffraction, reflection and absorption, but also terrain contours, environment (urban or rural, vegetation and foliage), propagation medium (dry or moist air), height and location of antennas. All these are summarized in the concept of *path loss*: the ratio between a local spatial average of the received power averaged in time (hence the $<>$ symbol) and the transmitted one:

$$pl = Path\ Loss = \frac{< Received\ Power >}{Transmitted\ Power} = \frac{< p_r >}{p_t}. \tag{2.1}$$

The expression can be rewritten in decibels as:

$$PL = -10\ \log_{10}(pl) = -10\ \log_{10}\left(\frac{< p_r >}{p_t}\right) = P_t - P_r \tag{2.2}$$

where the acronyms in capital letters stand for the decibel transformation of the lowercase ones.
Given the path loss, the transmitted power and antenna gains at both transmitter and receiver end (G_t and G_r), it should be possible to know how much power is received on average from a particular user:

$$P_r = P_t + G_t + G_r - PL. \tag{2.3}$$

These are some of the simplest path loss models.

2.1.1 LOS Model

Used to model free space propagation, when there is a direct line of sight (LOS) from Transmitter (TX) to Receiver (RX) without obstructions, reflections or scatterings. It can be expressed as:

$$pl = \left(\frac{4\pi d}{\lambda}\right)^{\alpha} = \left(\frac{4\pi df}{c}\right)^{\alpha} \tag{2.4}$$

where α is an attenuation constant, also known as *path loss exponent*, whose values typically range from 2 to 4, depending on the application, even though $\alpha = 2$ is the most used value, d is the path length in kilometres, f is the frequency in Megahertz, c is the speed of light in vacuum in km/s and λ is the wavelength in km.

2.1.2 NLOS Model

Unfortunately radio paths do not always (or almost never) meet the LOS hypothesis, so it is necessary to introduce another model consequently called Non Line Of Sight (NLOS)

$$\frac{p_r}{p_t} = \left[a\left(\frac{d}{do}\right)^{-\alpha}\zeta s\right] \tag{2.5}$$

where a is a constant that depends on frequency and on antennas characteristics, d is the path length and d_0 is a reference distance, e.g. the *a priori* correlation distance, in consistent units, α is, as in (2.4), the path-loss exponent, ζ and s represent *small and large scale fading* respectively.
Let's assume that both antennas are omnidirectional and, as typical in literature, that the spatial average of ζ, $<\zeta>$ is unitary, according to (2.2) and (2.5):

$$\begin{aligned} PL &= -10\ \log_{10}\left(\frac{<p_r>}{p_t}\right) = -10\ \log_{10}\left[a\left(\frac{d}{do}\right)^{-\alpha}s\right] = \\ &= -10\log_{10} a + 10\alpha\log_{10}\left(\frac{d}{d_0}\right) - 10\log_{10} s = \\ &= \underbrace{A + 10\alpha\log_{10}\left(\frac{d}{d_0}\right)}_{Median\ Path\ Loss} + S. \end{aligned} \tag{2.6}$$

Where A and S are the negative dB values of a and s. Path loss is consequently decomposed into *Median Path Loss* and *Shadow Fading* (S).
Small-scale and large-scale fading play an important role in the design of CRs by introducing randomness into the received power: if fading did not exist the received signal could be computed deterministically from the geographical locations of the transceivers.

2.2 MEDIAN PATH LOSS MODEL

The median path is defined by pieces

$$\text{Median } PL = \begin{cases} \underbrace{20\log_{10} 4\pi d_0\lambda}_{A} + 10\alpha\log\frac{d}{d_0} & \text{if } d < d_0, \\ \mu_\alpha + f_\alpha & \text{otherwise} \end{cases} \tag{2.7}$$

where d_0 is the reference distance mentioned beforehand, α is typically assumed to be equal to 2 in the first equation, μ_α depends on transmitter antenna height (h) through some constants a, b, c, that vary according on terrain type, so that $\mu_\alpha = a - bh + \frac{c}{h}$ and f_α is a truncated Gaussian to prevent the path loss from being negative $f_\alpha \sim trunc\mathcal{N}(0, \sigma_\alpha)$, with σ_α that depends on the terrain type.

2.3 LARGE SCALE, SLOW FADING, SHADOWING

The factor s, known as *shadowing* or *slow fading*, reflects the existence of obstacles on the path, such as hills and buildings, that, because of reflection and refraction, cause many replicas of the same signal to come from different paths, adding one another in a constructive or destructive way, depending on the relative phase shift, that, in turn, depends on motion speed, frequency and relative path length.
Shadowing varies spatially over much longer distances (tens or hundreds of meters), than the small-scale fading, hence the name, it is typically normalized, so that its median value is one.
It has been shown in [28] that s has a log-normal distribution over space, consequently S can be modelled as a two-dimensional process over space, whose mean is the path loss in decibels; models for its second order statistics (i.e. the autocorrelation function over space) are far less documented, according to [46], a widely used model is Gudmundson's:

$$< S(x), S(x + \Delta x) >= \sigma^2 \exp\left(-\frac{|\Delta x|}{X_c}\right) \tag{2.8}$$

where X_c is the correlation distance of shadow fading, that, according to the same paper, typically varies from 10 m in the urban area to 500 m in the countryside, σ has been given values between 3 dB in indoor environment to 12 dB in outdoor areas, but is mostly fixed to 8 dB.

2.4 SMALL SCALE, FAST, MULTIPATH FADING

ζ, also known as *multipath* or *fast fading*, is experienced as fast changes within symbol duration, all the channels which have a *coherence time*[1] shorter than the *symbol time* are subject to this type of fading. The signal experiences frequency dispersion, because of these fast temporal changes and time selective fading because of Doppler effect.
The value of ζ varies spatially, i.e. if the receiver moves slightly, since the amplitude of the received signal is the sum of multiple vectors whose relative phases change with receiver position. Multipath fading causes therefore rapid fluctuations in signal level over distances comparable to a radio wavelength.
It is usually normalized so that the spatial average has a unitary value.

2.4.1 Rayleigh Fading

Considered as the "worst" type of fading because of its complexity, it is present when there is no direct LOS. The field is a sum of several components, each with a proper magnitude

1 In communications systems, a coherence time is the time duration over which a communication channel impulse response is considered to be not varying, e.g. in wireless communications systems, due to Doppler effects.

and phase, if all the components have a similar magnitude, the *Central Limit Theorem* asserts that their sum will resemble a complex *Gaussian* random variable over space, with a phase uniformly varying in $(-\pi, \pi]$ and a magnitude distributed as a *Rayleigh*.

2.4.2 Ricean Fading

If, on the other side, a dominant component exists, e.g. if there is a LOS path from TX to RX, the sum of magnitudes will have a *Ricean* probability density function, which is characterized by two components: the power of the main ray and the average of all the weak scattered signals. The ratio between these two quantities has been considered to be a lognormal random variable, whose mean is a linear function of distance from the transmitter, with coefficients that depend on the environment (indoor, rural, urban ...).

2.4.3 Other Distributions

Other distributions have been presented to describe channel fading, such as Nagakami and Weibull, they will not be described in detail in this work, see, for example [47].

2.4.4 Temporal and Angular Fading

Fading can also be a temporal process, if the receiver or the transmitter are moving or if external objects are moving, such as windblown leaves, limbs of trees, cars and people. Multipath components, coming from different angles, have different Doppler shifts, so the composite received signal has a distribution of power over frequency, called *Doppler Spectrum*.

2.4.5 Flat Fading

This type of fading, which is not frequency selective, only depends on the bandwidth of the transmitted signal and is characterized by the same interference both in gain and phase for the whole transmitted signal.

2.5 QUANTILE MODEL

This model uses a quantised version of the fading distribution, trying to model the fact that the primary transmitter does not trust completely the nominal model, this increases the model realism and allows to compute interesting quantities, such as the fear of harmful interference.

2.6 STANDARDIZED PROPAGATION MODELS

2.6.1 Okumura-Hata Model

This is one of the most used models in large coverage cells with distances up to 100 km, it presents a logarithmic dependance on distance, antenna height and frequency and

includes different approaches and propagation losses depending on the area (urban, sub-urban or open).
Extensions of the Okamura-Hata model have been presented for various reasons, e.g. to adapt it to European cities, which present buildings way shorter and in lower density than Tokio, between these two famous ones are the COST-231 Hata model and the ECC-33 Path Loss Model, that can be adapted even for suburban and open areas.

2.6.2 Stanford University Interim Model

Other famous model, divided in sub-models according to the terrain type (hilly or flat) and on the tree density (heavy to light), a further insight can be found in [23].

2.7 STANDARDIZED DOUBLE DIRECTION MODELS

Channel characterizations exist that take into account which specific antenna is present at each ending. To separate the effect of the antennas in the estimation of the propagation model, this has been divided into the estimation of the *radio channel*, that takes the antennas into account and the *propagation channel*, that does not. A list of the most famous models can be found in [23], while further details and other, more complex models can be found in [46].

2.8 PROPAGATION MODEL USED IN THIS THESIS WORK

In this thesis work the propagation model used in the Bayesian part is taken from the recommendation by the International Telecommunication Union, see [33].
The model is developed for frequencies from 30 MHz to 3 GHz and is a combination of several propagation models. A major benefit of the ITU-R P1546-4 model is that it allows for taking into account terrain elevation data (when available) in predicting the pathloss.

3 SENSING

Sensing algorithms can be coarsely divided into two types: *Single User Spectrum Sensing* and *Collaborative Spectrum Sensing*, the reason to introduce the second technique is that assuming the primary receivers to be totally passive, it is fundamental to know where they are located to avoid harmful interference. It has been shown e.g. in [23] that the easiest way of proceeding is not detecting the presence of primary signal individually in each secondary user, that is computationally a demanding task, but cooperatively detect the presence of the protection region, where the presence of primary users is probable, therefore avoiding the *hidden node problem*.

3.1 SINGLE USER SPECTRUM SENSING AS AN HYPOTHESIS TEST

In the following the presence of primary signal is presented as a binary hypothesis test performed by every secondary transmitter through local observations.
The main purpose of spectrum sensing is to reliably and quickly identify whether the primary signal is present or not. It is possible to formulate this problem as a binary hypothesis test.
The SU has to decide whether to accept the *null hypothesis* $\mathcal{H}_0$ (received signal is only noise) or reject it and choose the *alternative hypothesis* $\mathcal{H}_1$ (received signal contains a contribute from the PU, alongside with noise). Under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively, the received signal at the SU is:

$$x(t) = \begin{cases} n(t) & \mathcal{H}_0 \\ s(t) + n(t) & \mathcal{H}_1 \end{cases} \qquad t = 1 \dots N \tag{3.1}$$

where t is the discrete index time, N is the number of observations, $x(t)$ is the received signal at the SU, $s(t)$ is the signal received by the PU, that agglomerates the effects of the channel, $n(t)$ is Addictive White Gaussian Noise (AWGN). The usual way to perform a hypothesis test, is to formulate a test statistics Λ from the observation vector $\mathbf{x} = [x(1), \dots, x(N)]$ and choose a threshold μ, above which $\mathcal{H}_0$ is rejected:

$$\Lambda(\mathbf{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \mu. \tag{3.2}$$

The choice of the test statistics and the threshold depend on the decision making strategy and on the desired performance parameters. If the assumption of conditional independence can be made, the optimal simple hypothesis test statistics is the Likelihood Ratio Test (LRT):

$$\Lambda = \frac{\mathbb{P}(\mathbf{x} \mid \mathcal{H}_1)}{\mathbb{P}(\mathbf{x} \mid \mathcal{H}_0)} \overset{i.i.d.}{=} \frac{\prod_{t=1}^{N} \mathbb{P}(x(t) \mid \mathcal{H}_1)}{\prod_{t=1}^{N} \mathbb{P}(x(t) \mid \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \mu. \tag{3.3}$$

A hypothesis test is said *simple* if the distributions do not depend on hidden parameters and the the hypotheses specify the population distribution completely, otherwise is said to be *composite*. If the study case enters the first category then the test statistics can be transformed into:

$$\Lambda = \frac{\prod_{t=1}^{N} \int_{\Theta_1} \mathbb{P}(x(t) \mid \theta_1; H_1)\mathbb{P}(\theta_1)\, d\theta_1}{\prod_{t=1}^{N} \int_{\Theta_0} \mathbb{P}(x(t) \mid \theta_0; H_0)\mathbb{P}(\theta_0)\, d\theta_0} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \mu \tag{3.4}$$

where $\theta_i, i = 0, 1$ are unknown parameters. If the distributions have some unknown, but *deterministic* parameters that differ under the two hypotheses, an option consists in using the Maximum Likelihood (ML) estimator and substituting the results inside the LRT, the resulting detector is known as Generalized Likelihood Ratio Test (GLRT), that in the subcase of equal probability of the null and alternative hypothesis can be written as:

$$\Lambda = \frac{\prod_{t=1}^{N} \max\limits_{\theta_1 \in \Theta_1} \mathbb{P}(x(t) \mid \hat{\theta}_1; \mathcal{H}_1)}{\prod_{t=1}^{N} \max\limits_{\theta_0 \in \Theta_0} \mathbb{P}(x(t) \mid \hat{\theta}_0; \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \mu. \tag{3.5}$$

Test statistics can be computationally complicated to evaluate and require a perfect knowledge of the distribution of $\mathbf{x}$ under both alternative and null hypotheses, alongside with information about parameters, such as noise and signal variance and channel coefficients. For these reasons, sometimes, simpler test statistics, even though suboptimal could be a better option, together with non-parametric techniques. For a deeper insight in test statistics please refer to [11].

3.1.1 Neyman - Pearson Detector

It is used to perform a hypothesis test between two point hypotheses $\mathcal{H}_0 : \theta = \theta_0$ and $\mathcal{H}_1 : \theta = \theta_1$, that are deterministically true. It maximizes the probability of detection for a given constraint on the false alarm probability

$$\text{Probability of False Alarm (PFA)} = \mathbb{P}(\Lambda \geq \mu \mid \mathcal{H}_0) = \bar{\alpha} \tag{3.6}$$

where the statistics Λ is written in the aforementioned LRT form:

$$\Lambda = \frac{\mathbb{P}(\mathbf{x} \mid \mathcal{H}_1)}{\mathbb{P}(\mathbf{x} \mid \mathcal{H}_0)}. \tag{3.7}$$

3.1.2 Bayesian Detector

If θ is an unknown parameter, with a prior distribution π, $\hat{\theta} = \delta(\mathbf{x})$ is an estimator of θ, $L(\theta, \delta)$ is a *loss function*[1], then the Bayesian risk of δ is defined as:

$$\mathbb{E}_\pi\{L(\theta, \delta \mid \mathbf{x})\}. \tag{3.8}$$

The idea is to minimize the Bayesian risk, that depends on the cost assignments and on the prior probability of the two hypotheses, which, unfortunately, may not be completely known. Interestingly, although the deterministic and the Bayesian approaches are substantially different, the Bayesian test results to be a LRT. For more details please refer to [34].

1 Function that maps an event onto a real number, representing some "cost" associated with the event, an example can be the squared error.

3.1.3 Min - Max Detector

Robust detector that seems to be the better choice for heavy-tailed noise models and if some uncertainties are present in the observations, see e.g. [42].

3.1.4 Locally Optimum Detector

Optimal detector for weak signals, since it maximizes the slope of the detection probability where the signal strength tends to zero. For computation details see [12].

3.1.5 Performance Criteria

Sensing algorithms have to be compared and evaluated according to some performance criteria, that have to be carefully chosen between the following, according to the scenario.

TYPE I AND II ERROR (Also known as PFA and Probability of Miss Detection (PMD)). The first is defined as the probability that the PU is declared to transmit, even if this is not the case $\mathbb{P}(\Lambda(\mathbf{x}) > \mu \mid \mathcal{H}_0)$, controlling it is crucial for an efficient spectrum usage. The latter is the probability that the presence of a transmitting PU is not recognized $\mathbb{P}(\Lambda(\mathbf{x}) < \mu \mid \mathcal{H}_1)$. Minimizing it is analogous to maximizing the Probability of Detection (PD), defined as the probability of the complementary event $\mathbb{P}(\Lambda(\mathbf{x}) > \mu \mid \mathcal{H}_1)$. The PMD depends from the propagation environment through the SNR. Controlling the PMD is essential to keep the interference under a predetermined threshold.
The relationship between the two error types is contained in the Receiver Operating Characteristics (ROC) curve, that illustrates the performance of a binary classifier system while varying its discrimination threshold. The fraction of true positives out of the positives (True Positive Rate (TPR)), also known as *sensitivity* is plotted versus the fraction of false positives out of the negatives (False Positive Rate (FPR)), that corresponds to *1 - specificity*, at various threshold settings.
For an insight refer to [35].

DETECTION TIME The time variations of the channel response determine how often the potential interference levels have to be estimated and, thus, how often transmission strategies may have to be updated. Consequently, depending on the application, the detection time could be a concern, (e.g. if the signal is liable to very sudden bursts, as it has been shown experimentally in [38] with a Wireless Local Area Network (WLAN)). In this thesis this is fortunately not an issue, since the Digital Video Broadcasting - Terrestrial (DVB-T) signal is quite stable.

DETECTION RANGE Maximum distance from which the PU can be detected reliably, is an interesting index, once the many factors on which it depends (e.g. SNR, sensing time, propagation environment) are fixed. It has to be carefully fine-tuned because if it is too wide the detector will perceive also the PUs outside its interference range, and this will prevent it from transmitting on bands that could be considered idle.

PRIOR KNOWLEDGE OF PRIMARY SIGNAL AND NOISE DISTRIBUTION Specific statistical information about the noise and knowledge about the receiver or the transmitter (location, mobility, sensitivity ...) reflect on a better detector performance, but this type of information is not always available and will be determinant on the choice of the algorithm to implement.

ABILITY TO DETECT DIFFERENT WAVEFORMS Some detectors are targeted for a specific type of waveform, while others are not. This can be particularly valuable because it could allow to discriminate PU signals from SU ones and interference, once the waveform of the PU is known. As usually the context helps choosing the best detector.

BANDWIDTH Depending on the situation it could be needed to sense a very wide bandwidth or a particularly narrow one, this could lead to very different choices of detectors.

ROBUSTNESS AGAINST NON – IDEALITIES Channel effects, loss of synchronization, hardware issues and impractical hypotheses comport the degradation of received signal, compared to the theoretical one, robustness is a very valuable property in a detector.

COMPLEXITY Simple algorithms are energy efficient, if power consumption is an issue (e.g. in mobile detectors) the implemented complexity is important.

AREA OF THE IMPLEMENTATION Mobile detectors have to physically fit in very small spaces, if this is the case, the area that is needed for the implementation of the algorithms can become a criterion (an example is offered in [38], where the detector has to fit in a mobile phone case).

3.2 DETECTOR TYPES AND SENSING ALGORITHMS

The choice of a detection criterion depends on which objective function has to be maximized and which characteristics, between the ones listed above are considered significant. This is a short review of the most common ones.

3.2.1 Energy Detector

This detector measures the energy received in a predefined time interval and compares it to a threshold.
Following [68] and [14] let's consider a deterministic signal $x(t)$, the statistic used in the conventional test, in the usual form is given by:

$$\Lambda = \frac{2}{N_0} \sum_{t=1}^{N} |x(t)|^2 \tag{3.9}$$

where N_0 is the noise spectral density.
The test statistics behaves differently under the two testing hypotheses:

$$\Lambda \sim \begin{cases} \chi^2_{2N} & \mathcal{H}_0 \\ \chi^2_{2N}(2\gamma) & \mathcal{H}_1 \end{cases} \tag{3.10}$$

where $\gamma = \frac{\sum_{t=1}^{N} |s(t)|^2}{N_0}$, the ratio between the signal energy and the noise spectral density. Under the hypothesis of zero mean Gaussian noise of known power[2] the energy detector is the optimal detector, in reality it works well in more general cases, even though it might

2 In practice the noise power is not known and has to be estimated as well.

be suboptimal.
Flexibility is the biggest advantage of energy detectors, since they can be used for any type of signal, there is no need of assumptions, they are simple to implement and robust to changes in the PU signal, consequently when little is known about the signal form, and the designer is reluctant to make unwarranted assumptions, choosing this kind of technique is the standard decision.
On the other hand, they make it impossible to distinguish among different signals (PU, SU or interferences) and some information, which could be known about the signal and could improve detector performance and robustness, would remain unexploited.
CRs must be able to detect very weak PU signal, but accurate detection is impossible under a certain level of SNR, known as *SNR wall*, even with infinite measurement duration, the reason is that the noise statistics is not accurately known in low SNR environments. Since the threshold for the test statistics to distinguish between the central and not-central χ^2 depends directly on the noise estimation, it can be hard to maintain a low false alarm rate and there could be severe performance limitations, for a demonstration see [1].
For all these reasons energy detectors are best suited for coarse scanning of the spectrum, even though are by far the most used in practice [67].

3.2.2 CFAR detectors

The aim of these detectors is to set the threshold μ in an adaptive way, trying to maintain a predetermined PFA, without knowing the noise power. Since decision threshold is set independently from unknown signals, CFAR detectors do not suffer for the SNR wall phenomenon. The test statistics of a CFAR detector $\Lambda(\mathbf{x})$ is unaffected by scaling, this is of course a very useful property in practice, since commonly the noise power is not known.

3.2.3 Matched Filter Detectors

If the transmitted signal is known and if the noise is AWGN, matched filter detectors are CFAR detectors and consequently, in this particular case, are the optimal filter for PU detection. They are obtained correlating a known transmitted signal with the received signal, to attain the desired false alarm probability the magnitude of the filter output is compared with a threshold.
The relatively short sensing time and the easy implementation, straight consequence of the linearity, are the main advantages.
The main drawback is that only one type of PU signal can be detected at a time, leading to the need of a bank of matched filters and dedicated receivers in every SU receiver for all signal types. The needed *a priori* information (e.g. pilot function, modulation, pulse shape, packet format), is usually inferred from characteristics of the wireless signals, such as the pilot function, the preambles and the synchronization words.
Compared to energy detection, the hardware requests are more demanding and it is difficult to modify the detector if the system evolves and new waveforms are introduced, moreover perfect synchronization is required.

3.2.4 Feature Detectors

Feature detectors are characterized by a very little need of theoretical assumptions, by the possibility to operate reliably at very low SNR levels and to differentiate certain signal

types from others. This is due to the fact that signals used in practical communication systems contain distinctive statistical properties that can be exploited to estimate unknown parameters, such as the noise power, to circumvent the problem of the SNR wall and achieve a better performance.
Modern modulation and coding techniques have aided in the design of efficient spectrum sensing algorithms. For example pilot signals and the use of cyclic prefix in OFDM cause very distinctive signal features that can be exploited in the detector design. In CR prior information about the primary user waveforms typically exists e.g. some of the cyclic frequencies of the primary user signals, since regulatory bodies require the transmitted signal parameters to be disclosed and from this it is possible to calculate the cyclic frequencies of interest.
In the following, focus will be on the detection of primary user signal and the cyclic frequencies will be assumed to be known.

3.2.5 Detector based on Second Order Statistics

The easier way to check if a signal is only composed of noise or if some information is contained is to look at the second order statistics.
If enough samples are collected, knowing first and second moment should be enough, at least for standard practical accuracy. Moreover, since base-band communication signals are almost always zero-mean to minimize the emitted power, only looking at the second order statistics is typically sufficient, see e.g.[1].
These detectors can be used for any kind of signal with a Cyclic Prefix (CP), let's consider an OFDM signal, let N_d be the block size of the data and N_c the length of the CP, that is typically a repetition of the last transmitted samples and $N = K(N_c + N_d) + N_d$ be the total number of samples as in Figure 3.

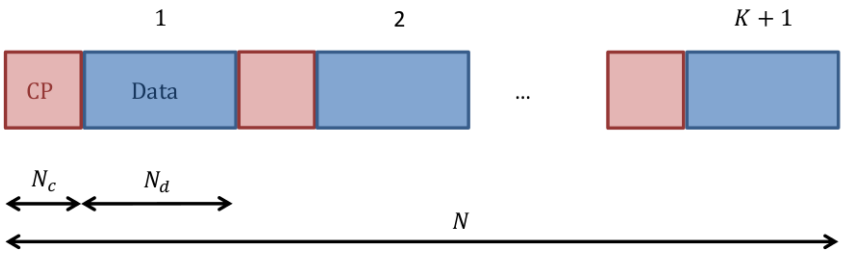

Figure 3: N samples of a received OFDM signal.

The Autocorrelation Function (ACF):

$$r_x[t, \tau] = \mathbb{E}[x[t]x^*[t + \tau]] \tag{3.11}$$

where t is the time index and τ the autocorrelation lag, is time-varying and non-zero because of the insertion of the CP that makes the signal non stationary.

Let's consider for simplicity $n(t)$ to be AWGN, thanks to this hypothesis for any $\tau \neq 0$ $r_n[t,\tau] = 0$, hence the received signal

$$x(t) = \begin{cases} n(t) & \mathcal{H}_0 \\ s(t) + n(t) & \mathcal{H}_1 \end{cases} \qquad t = 1 \dots N \tag{3.12}$$

has an ACF $r_x[t,\tau] = r_s[t,\tau]$ for any $\tau \neq 0$, $s(t) = hz(t) + i(t)$, where $z(t)$ is the original transmitted signal, h takes into account the medium and $i(t)$ is the interference.
A possible estimate of $r_x[t,\tau]$ is

$$\hat{r}_x[t] = x[t]x^*[t + N_d] \quad t = 1, \dots, K(N_c + N_d) \tag{3.13}$$

Tests for making decisions on the signal presence versus absence based on $\hat{r}_x[t]$ are various, they all start from the basic affirmation that the mean of $\hat{r}_x[t]$ is non-zero only for some t.
Chaudhari in [14] cites between the first statistical tests proposed:

$$\max_{\theta} \left| \sum_{t=\theta+1}^{\theta+N_c} \hat{r}_x[t] \right| \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \mu \tag{3.14}$$

where θ has to be interpreted as a synchronization mismatch, the drawback of this test is that both N_c and N_d need to be known, together with the fact that the threshold μ depends on the power of the noise and consequently this detector is susceptible to the SNR wall phenomenon and thus the test is not CFAR.
In [15] the test presented is:

$$\frac{\sum_{t=1}^{N-N_d} \mathrm{Re}(\hat{r}_x[t])}{\sum_{t=1}^{N} |x[t]|^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \mu. \tag{3.15}$$

The advantage of this test is that only N_d has to be known, this is useful if N_c is unknown or if CPs of different lengths are in use at the same time; in addiction it is a CFAR test, hence no knowledge of the noise power is required.

3.2.6 Detector based on Cyclostationarity

A subclass of feature detectors exploits cyclostationarity, characteristic of almost all artificial signals, due to modulation and the use of CPs, that cause the ACF to be periodic.
In the already introduced OFDM the ACF is periodic, with a fundamental period equal to the length of the OFDM symbol $N_c + N_d$.
Second order periodic signals can be modeled with second-order cyclostationary random processes.

Definition 1 (Cyclostationarity) *A signal $x[t]$ is second-order cyclo-stationary if its mean and autocorrelation are periodic in time.*

Hence it is possible to express the ACF as its Fourier series:

$$r_x[t,\tau] = \sum_{\alpha} R_x(\alpha,\tau) e^{2\pi j \alpha t} \tag{3.16}$$

where α are *cyclic frequencies*, which depend on type of modulation and on the length of the symbol T_b since $\alpha = \frac{k}{T_b}$ or $\alpha = \frac{k}{T_b} + 2f_c$, where f_c is the carrier frequency.
The Fourier coefficients $R_x(\alpha,\tau)$ depend on the time lag τ as:

$$R_x(\alpha,\tau) = \frac{1}{N} \sum_{t=0}^{N-1} r_x[t,\tau] e^{-2\pi j \alpha t}. \tag{3.17}$$

An alternative definition is therefore possible:

Definition 2 (Cyclostationarity II) *A process $x[t]$ is second-order cyclostationary when $\alpha \neq 0$ exists such that $R_x(\alpha, \tau) > 0$.*

Detection of cyclostationarity can be performed both in time and in frequency domain.
Cyclic autocorrelation R_x^α at cyclic frequencies α can be estimated as:

$$\hat{R}_x(\alpha, \tau) = \frac{1}{N} \sum_{t=0}^{N-1} x[t] x^*[t-\tau] e^{-2\pi j \alpha t} \tag{3.18}$$

$\hat{R}_x(\alpha, \tau)$ is an estimation of $R_x(\alpha, \tau)$, consequently it can be written as:

$$\hat{R}_x(\alpha, \tau) = R_x^\alpha + \epsilon(\alpha). \tag{3.19}$$

where $\epsilon(\alpha)$ is the estimation error.
Usually, in real datasets, $\hat{R}_x(\alpha, \tau)$ is rarely exactly null, therefore decisions have to be made about the existence of a cyclostationarity.
Let's formulate this problem as a classical binary hypothesis test, that has cyclostationarity in the alternative hypothesis, while the null hypothesis is that the signal is not cyclostationary.

$$\begin{cases} \mathcal{H}_0 & \hat{R}_x^\alpha = \epsilon(\alpha) \ \forall \ \alpha \in A \\ \mathcal{H}_1 & \exists \ \alpha \in A \colon R_x^\alpha \neq 0 \end{cases} \tag{3.20}$$

where the set A contains all the cyclic frequencies for a fixed value of τ, which are assumed to be known a priori:

$$A = \{\alpha \in [0, 2\pi) \colon R_x(\alpha, \tau) \neq 0\}. \tag{3.21}$$

It is possible to assume explicit knowledge of PUs cyclic frequencies, since regulatory bodies monitoring spectrum allocation require the transmitted signal parameters to be disclosed and from this it is possible to calculate the cyclic frequencies of interest.
Under $\mathcal{H}_0$ $R_x^\alpha = 0$ and therefore$\hat{R}_x^\alpha = \epsilon(\alpha)$, which is an asymptotically normal zero mean complex random variable: $\epsilon(\alpha) = X(\alpha) + jY(\alpha)$, where $X(\alpha)$ and $Y(\alpha)$ are real normal random variables, the covariance matrix is therefore:

$$\Sigma = \begin{bmatrix} \mathbb{E}[X^2] & \mathbb{E}[XY] \\ \mathbb{E}[XY] & \mathbb{E}[Y^2] \end{bmatrix} \tag{3.22}$$

whose estimate is:

$$\hat{\Sigma} = \begin{bmatrix} \frac{1}{N} \sum_{k=0}^{N-1} \mathrm{Re}\{\hat{R}_x^{\alpha_k}\}^2 & \frac{1}{N} \sum_{k=0}^{N-1} \mathrm{Re}\{\hat{R}_x^{\alpha_k}\} \mathrm{Im}\{\hat{R}_x^{\alpha_k}\} \\ \frac{1}{N} \sum_{k=0}^{N-1} \mathrm{Re}\{\hat{R}_x^{\alpha_k}\} \mathrm{Im}\{\hat{R}_x^{\alpha_k}\} & \frac{1}{N} \sum_{k=0}^{N-1} \mathrm{Im}\{\hat{R}_x^{\alpha_k}\}^2 \end{bmatrix}. \tag{3.23}$$

The distribution of $\hat{R}_x$ differs only in mean under the two hypotheses.
Testing for the presence of a particular α in A is equivalent to a binary classification problem and requires knowledge of the distribution of $\epsilon(\alpha)$, since the data distribution is unknown, the asymptotic properties of the cyclic covariance estimators are made use of to estimate its asymptotic distribution.
Under the assumption of samples of $x(t)$ being well separated in time and approximately independent, it is possible to prove that the estimator $\hat{R}_x(\alpha, \tau)$ is mean square consistent, i.e.:

$$\lim_{N \to \infty} \hat{R}_x(\alpha, \tau) = R_x(\alpha, \tau) \tag{3.24}$$

and, additionally, $\sqrt{N}[\hat{R}_x(\alpha,\tau) - R_x(\alpha,\tau)]$ is asymptotycally complex normal, for demonstration see [17].
It is therefore possible to show that

$$\lim_{N\to\infty} \sqrt{N}\epsilon(\alpha) \stackrel{\mathcal{D}}{=} \mathcal{N}(\underline{0}, \Sigma). \tag{3.25}$$

From now on Test Statistics will indicate

$$T = \hat{R}_x^{\alpha}\hat{\Sigma}^{-1}(\hat{R}_x^{\alpha})^T. \tag{3.26}$$

Under the null hypothesis ($\mathcal{H}_0$) it is asymptotically χ^2_2 distributed, once a value for τ is chosen:

$$\lim_{N\to\infty} T \stackrel{\mathcal{D}}{=} \chi^2_2. \tag{3.27}$$

It is therefore possible to derive a CFAR test for the presence of cyclostationary:

$$F_{\chi^2_2}(\mu) = 1 - p \tag{3.28}$$

where $F_{\chi^2_2}$ is the cumulative distribution of a χ^2_2 distribution and p is the desired false alarm rate, that implies the choice of the threshold μ, so that $p = \mathbb{P}(T > \mu|\mathcal{H}_0)$. Finally α is declared a cyclic frequency if $T > \mu$.
The test is consistent, asymptotically optimal in the generalized likelihood sense and does not require knowledge of data distribution, the normalization via the inverse of the covariance matrix, makes thresholding easier and standard because it is enough to look up the central χ^2 distribution.
It has to be noted that if a couple (α,τ) exists for which $R_x(\alpha,\tau) \neq 0$ then there is presence of second order ciclostationary, this does not imply, however, that $x(t)$ is cyclostationary.
Under the alternative hypothesis it results:

$$\begin{aligned}\lim_{N\to\infty} \sqrt{N}(\hat{R}_x(\alpha,\tau)\hat{\Sigma}^{-1}\hat{R}_x(\alpha,\tau)^T) - R_x(\alpha,\tau)\Sigma^{-1}R_x(\alpha,\tau)^T) \stackrel{\mathcal{D}}{=} \\ \mathcal{N}(0, 4R_x(\alpha,\tau)\Sigma^{-1}R_x(\alpha,\tau)^T).\end{aligned} \tag{3.29}$$

For large enough N it is possible to write the approximate distribution of T under $\mathcal{H}_1$ as:

$$T \sim \mathcal{N}(R_x(\alpha,\tau)\Sigma^{-1}R_x(\alpha,\tau)^T, \frac{4}{N}R_x(\alpha,\tau)\Sigma^{-1}R_x(\alpha,\tau)^T). \tag{3.30}$$

Once the threshold μ has been set it is possible to evaluate the probability of detection $p_D = \mathbb{P}(T > \mu|\mathcal{H}_1)$, using the distribution of T under $\mathcal{H}_1$, that can be approximated by substituting $R_x(\alpha,\tau)$ and Σ with their relative estimations.
The test can be modified to include multiple lag values or multiple cyclic frequencies, that, even though way more expensive, can be useful in the case of an exhaustive search for absence of cyclostationary, e.g. to validate that a channel is idle.
In the multiple lags case, let $\tau_1, \ldots, \tau_N$ be a fixed set of lags, defining

$$\hat{R}_x(\alpha) = \left[\mathrm{Re}\{\hat{R}_x(\alpha,\tau_1)\}, \ldots, \mathrm{Re}\{\hat{R}_x(\alpha,\tau_N)\}, \mathrm{Im}\{\hat{R}_x(\alpha,\tau_1)\}, \ldots, \mathrm{Im}\{\hat{R}_x(\alpha,\tau_N)\}\right]. \tag{3.31}$$

Under the null hypothesis, the test statistic T, defined as before, that is now a $2N \times 1$ vector, is asymptotically χ^2_{2N} distributed:

$$\lim_{N\to\infty} T \stackrel{\mathcal{D}}{=} \chi^2_{2N}. \tag{3.32}$$

The CFAR test therefore becomes:

$$F_{\chi^2_{2N}}(\mu) = 1 - p. \tag{3.33}$$

Finally α is declared a cyclic frequency if $\exists \tau_k,\ k \in 1, 2, \ldots, N \colon T(\alpha, \tau_k) > \mu$.
It is possible to derive the same test in a frequency domain using the asymptotic normality of the spectrum, instead than the one of the cyclic covariance, without requiring knowledge of the data distribution, for a deeper insight see [17].
Much of the work in developing a detector has been done in the framework of the standard IEEE 802.22, the first that takes into account CR technology. The authors of [67] find the standard a bit poor because it simplifies the sensing phase choosing to rule only the case in which detectors and Fusion Center (FC) (whose work is to fuse information from different detectors) are fixed in space and the only estimated signal is DVB-T, that is known for being particularly stable and is addressed also in the experimental part of this thesis.
Cyclostationary detection has attracted interest for robustness to noise uncertainty and for the capability of distinguishing systems with different cyclostationary features.
The first characteristic is particularly useful in interference limited communication channels, where it is hard to estimate the noise power reliably. As reported in [40], when this is known perfectly, the energy detector outperforms the cyclic one, but raising the noise uncertainty augments the performance gap in favor of the cyclostationary detector, since even augmenting the number of samples could not improve the energy detector performance because of the SNR wall. The latter characteristic, or the ability to disinguish systems with different cyclostationary features, allows to distinguish between signals emitted by PU, SU, interference or random noise, which usually does not have cyclostationarity properties.
Nevertheless some important issues arise.
First of all the theory has been developed in the case of samples of $x(t)$ *well separated in time*, so that they can be considered approximately independent, this is not true in real detectors, that need to estimate the field as fast as possible, to allow the starting of a new transmission. The violation of an hypothesis is translated in a degrade of detection performance.
An issue, presented both in [61] and [74] is sensibility to clock and frequency offsets. According to the authors, the detector could suffer from a *cyclic frequency mismatch* and consequently be unable to properly estimate the frequency where the signal has the cyclostationary feature. Even a very small mismatch error can cause a dramatic reduction of the detector performance and increasing the sample size does not directly imply an improvement of the detector performance.
Another problem presented in [61] is that, after a reconfiguration of the hardware, e.g. changing the carrier frequency or the decimation rate some transient peaks occur, so that additional quiet periods become necessary.
Numerous and very different are the proposals for improving cyclostationary detection.
The proposal of [74] and [40] to attenuate the problem of the frequency mismatch is to evaluate the cyclostationary auto correlation at multiple frequencies in an interval centered in α_0, to increase the robustness of the whole procedure. This technique can only be used if the primary signal has multiple strong cyclic frequencies, otherwise the performance could deteriorate because of the increased number of degrees of freedom.
A new approach based on *cyclostationary signatures* seems to improve the identification of a particular network embedding additional information concerning the bandwidth and the modulation scheme in the signal and therefore altering the characteristics of the signal to make it unique; an overview, together with some interesting results can be seen in [61].

3.2.7 Other Detectors

A wide range of detectors exist, which is not listed here, between these the most present in literature are wavelet detectors and covariance detectors, see eg [1].

3.3 COLLABORATIVE SENSING

Non cooperative spectrum sensing techniques experience issues when the primary signal is under heavy shadowing or multipath fading, that could lead to avoid problems such as the *hidden terminal problem*, due to unsensed primary receivers.
The consequent technique is cooperative sensing, where information from secondary cognitive users is shared, so to have a better primary users detection, without excessively long detection times. In practice the shadowing process may be correlated between secondary users, thus reducing the gain from collaboration and consequently emphasizing the importance of spatial diversity between them.
Users collaboration can be realized in different ways e.g. all users can monitor the band of interest or they can sense a portion of it and then share the information with the other users or with a central calculating center, called FC in literature.
Collaborative sensing can be divided into two major types: *centralized* and *decentralized* sensing.

3.3.1 Centralized Cooperative Sensing

All cognitive users send the gathered information to a FC, a cognitive radio user, that "fuses" the information and takes the final decision about the presence of primary signal. Two are the techniques that are at the moment the most researched about: *decision fusion* and *data fusion*.

Decision Fusion (Hard Combination)

In this type of cooperative sensing, all the single secondary users take a 1-0 decision about the presence of transmitting primary users. These decisions are conveyed to the FC that has the final word on the matter. The final decision can be taken with multiple rules, such as the *OR rule*, also known as *1-out-of-M rule*, that is very conservative and assumes the primary user is transmitting if at least a SU is saying so, another option is the *Majority Rule*, also known as *K-out-of-M rule*, and the *AND* rule, which is the *M-out-of-M rule*.

Data Fusion (Soft Combination)

The FC takes its decision by averaging the estimates of the primary signal coming from all the secondary users. This approach requires a higher bandwidth and overhead for data transmission to the FC, because of the higher complexity of the conveyed information. In a centralized radio-environment with fusion information a centralized database stores all what is known about the state of the radio environment. The information can include transmitters and receivers location, field strength estimates, results from field measurement campaigns, etc. This information is combined and processed to give a realistic estimate of the radio environment, whose accuracy ensures the protection of the primary network operation, while allowing the maximization of the secondary system throughput.
A censoring scheme, as the one presented in [40], in which only informative test statistics

are transmitted, seems to be the way to go to improve the technique feasibility.
This technique is the one which is the mostly thought about while developing algorithms in this thesis work.

3.3.2 Decentralized Cooperative Sensing

In this type of cooperation all users exchange information with each other to decide whether the PU is transmitting or not. In this case there is no need for a backbone infrastructure, because there is no FC, since each SU is operating as an independent FC by collecting sensing information shared by its neighbors.
For this reason in this framework a lot of theoretical effort has to be put into transmit power control, that is operated each SU; the solutions presented so far seem to focus on etiquette, protocol and cooperative, ad-hoc networks, whose form of collaborations can be variegated, e.g. only between directly connected nodes.
For a deeper insight into *Cooperative Sensing*, please refer to [14].
Numerous are the advantages presented by a fully decentralized control situation, e.g.Haykin in [29], under some reasonable assumptions, shows that the network can scale to almost arbitrary number of nodes; in the same paper some simulations results are also presented.

3.4 FUTURE OF SPECTRUM SENSING

At the moment database-centric collaborative approaches are still the core solutions in primary signal detection because of a multitude of still unresolved issues encountered by the decentralized collaborative approach, it still is not clear which will be the dominant path in the near future, but surely the need of a new standard is very present.

3.5 IMPLEMENTATION OF THE DETECTOR USED IN THIS THESIS WORK

The software and hardware used in the measurement campaigns present a cyclostationary detector, especially modified for these measurement campaigns, they have already been described in two papers [38] and [67], here a brief introduction is presented, please refer to these for further insights.

3.5.1 Measurement Setup and Hardware

The hardware used in the measurements consists of the detector connected via USB to a notebook equipped with a commercial external UHF antenna for the DVB-T signal, a GPS receiver to track the position of the measurements and a supplementary external battery, see in Figure 4. A Matlab Graphical User Interface was used to illustrate the results per every channel, for both the test statistics and the RSSI, allowing the operator to coarsely check the correctness of the measurement in almost real time, see Figure 5. The detection algorithm has been implemented in a field-programmable integrated circuit, see Figure 6 so that the measurement device is a mobile-scale implementation.

Figure 4: Measurement setup in the Helsinki campaigns, consisting of a waterproof case containing a portable computer connected to the detector via USB.

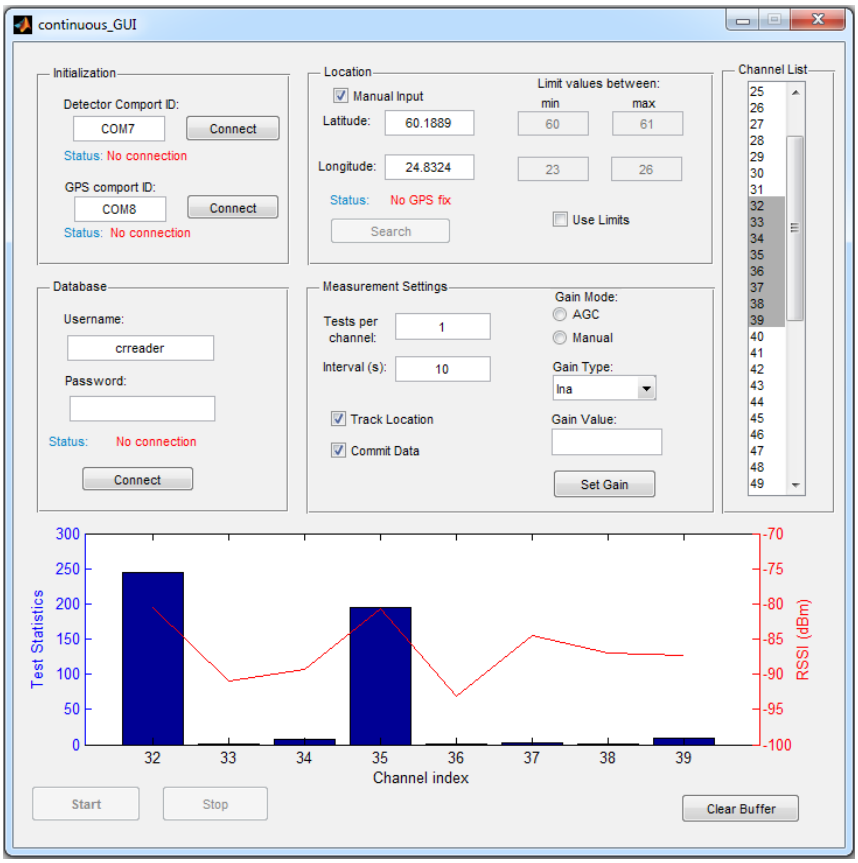

Figure 5: Screenshot of the Matlab GUI, on the left it is possible to see if the GPS has or not been locked, or enter the location manually, on the right it is possible to manually select which channels to scan. In the bottom graphic the histograms show the test statistics, while the line pictures the RSSI in the measurement location, varying the channel.

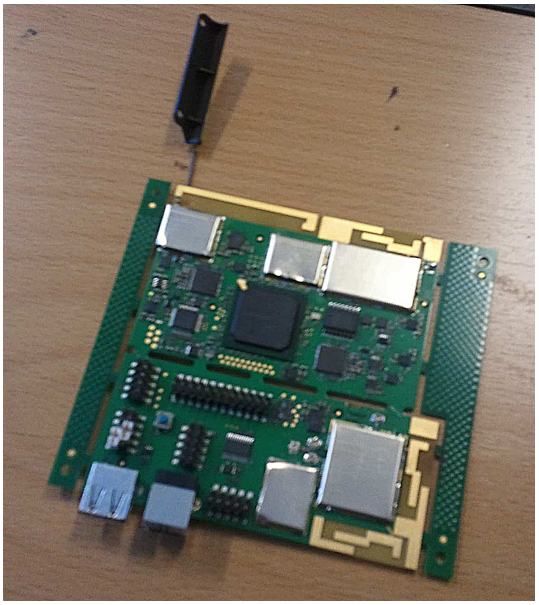

Figure 6: Detector board equipped with the UHF antenna.

3.5.2 The implementation

The implemented detector is a cyclostationary feature detector, that takes advantage of the cyclostationarity inhered from modulation and cyclic prefixes.
The detector algorithm is based on Fast Fourier Transform (FFT), for very pragmatical reasons: as clearly stated in [67] the device was built to support a later incorporation of an energy detector, that requires to work in the frequency domain. Less hardware would have been sufficient for an estimation of the covariance matrix in time domain.
The implementation consists of a single lag, single frequency detector, since it has been illustrated, see e.g. [75], that using only one or few cyclic frequencies is sufficient to get a satisfying performance, since it already outperforms the feature detector based on second order statistics.
The implementation is shown in Figure 7.

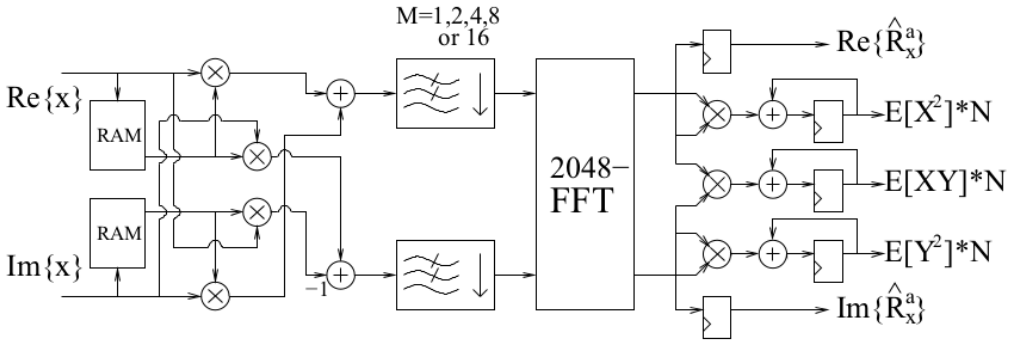

Figure 7: Block diagram of the detector, as reported in [67].

First step is a multiplication of the complex input with a delayed version of itself, this requires the four multipliers and a dual port Random Access Memory (RAM) to implement the lag, the size of which was determined by the expected maximum lag value.
At this point a resampling is operated by a Finite Impulse Response (FIR) filter, that supports multiple decimation ratios. *Decimation* is a technique used to improve the detector performance, since usually the information of the cyclic spectrum resides in the low cyclic frequencies, the hight-frequency part of the spectrum is of no importance, consequently the low frequencies can be resampled before applying the FFT. This increases the allowed detection time by a factor corresponding to the *decimation ratio* and consequently the probability of detection augments, as experimentally shown in [67]. Maximum detection ratio can be limited by both the cyclic frequencies of the signal in exam and by detection time constraints. Decimation allows to dynamically balance the detection time and therefore improve the performance, with a fixed size FFT. This is particularly useful in mobile detectors where the power consumption is certainly critical.
The decimator is followed by the FFT unit, chosen because of the low complexity that allows an implementation in a small area, with a low power consumption, important in a device that was built to be cheap and portable. The FFT is the unit with the major area and the most power consuming and it determines the minimum detection time as:

$$T_d = \frac{M \cdot N_{FFT}}{f_{s,in}} \tag{3.34}$$

where M is the decimation ratio, N_{FFT} is 2048, the default array size for the computation of the FFT and $f_{s,in}$ is the input sampling frequency. Therefore if, for example, the input is sampled at 20 MHz, varying M from 1 to 16 makes T_d range from 100 μs to 1.6 ms.
The elements of the covariance matrix are calculated by three multiply and accumulate blocks and an external microprocessor is added at the end of the process to operate the matrix inversion needed for the calculation of the test statistics.

3.5.3 Pros of the detector

Mobile detectors, like the one that was used in the experimental part, have two big advantages: being extremely cheap, compared to commercial spectrum analyzers and being easily transportable. In the not so far future, they could be implemented in mobile SUs, that could collaborate, sense local spectrum and provide a fast response to dynamic signal conditions.

3.5.4 Issues of the detector

As stated in [38] a miniaturized device encounters strong physical challenges:

BATTERY CAPACITY, that becomes more important when the number of channels to be monitored frequently increases,

TOTAL IMPLEMENTATION AREA, since as a constructive challenge the detector had to fit in the case of a N900 Nokia phone, and miniaturization is not straightforward in this kind of devices,

LOSSES IN THE ANTENNA, that is cheap and very portable, but whose gain is clearly not constant in a wide bandwidth, as it should,

FRONT END FILTER, known to cause increase unintentional false alarm rate, because of intermodulation, when there is a strong primary signal in the vicinity, see [38].

Others are causes of non ideality which are not strictly connected to miniaturization, but are still present in the dataset, such as:

NARROWBAND INTERFERENCE, as reported in [38] in few locations in Helsinki downtown a signal is detected from Estonia on channel 45, that is sometimes not perceived by the detector, because of the stronger signals on channel 44 and 46, that are transmitted by the much closer transmitter in Espoo.

DISTANCE BETWEEN SIMULATED AND SAMPLED DETECTION CURVES, issue that has already been presented in [67]: the probability of detection versus the SNR, while varying the decimation ratio at low levels of input signal does not match the theoretical curve, since the probability of detection tends to zero instead than to the theoretical false alarm rate. In the above cited paper this is explained by the fact that noise in practice is not AWGN and especially at low levels of signal it is dominated by interference coupled with the analog baseband. This issue does not translate into any real problems, because this happens at SNR values well below the noise floor.

As it is the device does not meet the sensing requirements imposed by the FCC, which are actually under questioning, because the current ruling does not include the possibility of using multiple small devices and such strict requirements could cause a lot of false alarms due to intermodulation.

4 MEASUREMENTS

As clearly stated in [46] channel models cannot completely describe the reality of propagation: typical or standardized values of channel parameters are helpful, but need to be compared with reality. They come without any form of metric to measure performance and are therefore inadequate for statistical prediction, therefore the aim of this thesis work is to adapt them to our aims.
The datasets that have been faced in this work are multiple, but they are all multi-channel measurements of the DVB-T signal, performed with similar set-ups, which have already been described in detail in chapter 3.

4.1 MEASUREMENT CAMPAIGNS

Locations where to sample depend on the particular wireless system, because of the coverage area of the single transmitter. In each location the number of samples depends mainly on the desired confidence level, that can be considered to be coarsely proportional to the standard deviation of the measurements.

4.1.1 First and Second Measurement Campaigns (MC1 and MC2)

The data of the first and second campaigns have been gathered in a moving car, equipped with the experimental set-up that has been described. On July 9^{th} 2013 the way from Helsinki to Porvoo and back has been covered, mostly to check the measurement set-up stability. These measurements are not included in the present analysis, since only one sensor was available and improvements in the detector had still to be implemented.
Doppler effect, caused by the relative motion of transmitter and receiver, had not been taken into account, even though it affects the performance of cyclic detectors because of the carrier frequency shifts and the subsequent changes in cyclic frequencies. For further details see [40], where the DVB-T signal is said to suffer more from this issue compared to other signal types, because of its longer symbol length. Another issue is caused by the possible coupling of the electromagnetic field between the field produced by the car engine and the received signal, that is not anyhow quantifiable.
In MC2 the route between Helsinki and Espoo and back has been covered. This time two sensors were used for measurements in order to validate the results. The results were noticeably more consistent compared to MC1. This was achieved by fixing some implementation bugs and by keeping speed under 90 km/h all the time, not to have a significant Doppler effect, that is known to be important only at high mobile speed, unrealistic in practical CR applications. Because of some issue related to ambiguity between time stamp and positioning registered by the GPS device, this campaign has not be used in the present work.

Figure 8: Sampled locations in MC3.

4.1.2 Third Measurement Campaign (MC3)

These measurements have been gathered in two different days: July, 30th and August, 23rd 2013 in Helsinki metropolitan area, locations an be seen in Figure 8. These data are different from the previous ones, even though the experimental set-up is the same, because they have been collected "statically". In this case six people, equipped with one sensor each, have moved in the city by foot and with public transportation, standing in each of the planned locations for a time sufficient to collect approximately 200 samples per location, and turning the sensor off while moving, to obtain results which are more statistically relevant.
Sensors were divided into few clusters about 100 m apart from each other, each of which composed by three sensors, separated about 10-20 meters. A map of the measurement locations can be visualized in Figure 8.
Table 1 keeps track of the sampled channels, their central frequency and their real occupational state (idle or occupied). Let us go a little more in detail about what is to be considered an idle band or an occupied band in a CR context. The idea is that SUs in Helsinki downtown area could transmit in an idle channel with a low power without interfering faraway TV transmission. However, if the SU would transmit with an extremely high power (e.g. hundreds of kilowatts) essentially all the TV channels would need to be considered as occupied, therefore channel occupancy is not a concept by itself, but a function of transmit power.

4.1.3 Fourth Measurement Campaign (MC4 and TSM1)

For this campaign a better experimental set - up was available thanks to the collaboration with Juha Kalliovaara from the University of Turku, for technical details refer to the manual [60]. In every location two sets of measurements have been collected, one per type of

sensor, in order to compare the results.
Per every location 200 samples were gathered with the Time Scale Modification (TSM) detector and 50 with the cyclostationary one; the difference is due to the fact that the sensing time is much shorter with the TSM detector, consequently in the same time interval it was possible to collect only a smaller number of samples with the cyclostationary detector.
The idea was to get as close as possible to the old measurements, in order to offer a referential measure to attest the validity of the previous campaign.

Channel Number	Frequency (MHz)	Occupied Channel	Transmitter Location
21	474	N	-
22	482	N	-
23	490	N	-
42	642	N	-
43	650	N	-
44	658	Y	Espoo
45	666	Y/N	Tallinn (Estonia)
46	674	Y	Espoo
47	682	N	-
48	690	N	-
49	698	N	-
50	706	N	-
51	714	N	-
52	722	N	-
53	730	Y	Espoo
54	738	N	-
55	746	N	-
56	754	N	-
57	762	N	-

Table 1: List of the sensed channels in MC3 and MC4. Channel 45 is somehow unique because it is received in some areas of Helsinki, but the signal comes from Tallinn (Estonia). The main transmitter that is considered in the work is visible in Figure 9.

For details about the first four measurement campaigns please refer to [41].

4.1.4 Fifth Measurement Campaign (MC5)

These measurements have been gathered in Turku, Finland, on the test DVB-T channel 38 with a dipole antenna and the TSM detector. The measurement set-up was put in a chart attached to a bicycle and a minimum of 1000 samples per location were collected. Together with these static measurements, moving measures between locations have been performed as well, since the sensor was not shut off during movements. In this campaign only the Received Signal Strength Indicator (RSSI) is available, not the test statistics. For further details on the fifth measurement campaign, please refer to: [51], [52] and [53].

Figure 9: The red dot indicates the transmitter position for channel 44, Helsinki peninsula is clearly visible in the South-East.

4.2 MAPPING

Shapefiles to allow the plotting (and cropping) of spatial interpolation and predictions maps through `ggmap` in R, were obtained by the the National Land Survey of Finland (NLS) and are freely and openly available online, see [50].

4.3 DATA DESCRIPTION

Data are composed of N (varying between the campaigns) observations of the following vector:

COORDINATES sampling locations. Data have been received in an unprojected format, projection is necessary to perform the variogram using a plane distance, instead than the Great Circle one. See the Appendix for clarification.

RSSI (Received Signal Strength Indicator), a measurement of the power level of the relative received signal strength in a wireless environment; the higher the RSSI, the stronger the signal. It is expressed in decibels [dB]. There are at least 50 (typically 200 or more) recordings per location in time.

TEST STATISTICS Test Statistics value from the cyclostationary detector. There are at least 50 (typically 200 or more) recordings per location in time.

TIME Time elapsed since the measurement started. The value is taken directly from the operating system.

DETECTOR Integer ranging from 1 to 6 that identifies which measurement set-up between those used in the campaign got that particular entry.

ELEVATION altitude of the measurement location from sea level, data have been extracted from Google Maps through Google Maps API V3 and Google Maps API Elevation Service through an interface available online [1]. It is expressed in meters.

SEA DISTANCE has been calculated through a minimum distance query between single points and a polyline interpolation of the seashore using GRASS GIS[2], a free and open source Geographic Information System (GIS) software suite used for geospatial data management and analysis. It is expressed in meters. The idea was to use this distance in the prediction of Chapter 45, see Table 1.

TRANSMITTER DISTANCE distance between the transmitter and the measurement points, calculated as a geodesic distance, along the surface. The method involves the solution of the geodetic inverse problem, using T. Vincenty's modification of Rainsford's method with Helmert's elliptical terms. To calculate this the function `geodDist` in the `oce` package has been used. The transmitter location (60°10′40.0″N 24°38′24.0″E) has been found on the DVB-T-map website[3], it is expressed in kilometers. As a double check this distance has been calculated using the Euclidean distance in the projected Universal Transverse Mercator (UTM) surface, giving an error of 3.5 meters in average, with a standard deviation of approximately 3.5 meters, that has been considered a very good approximation.

4.4 VISUAL ANALYSIS

Both test statistics and RSSI are good indexes to discern a channel occupancy, this can be coarsely visualized in Figure 10, 11 and 12. It will be shown that an extremely well performing classifier can be built on these two quantities.

1 http://www.daftlogic.com/sandbox-google-maps-find-altitude.htm
2 http://grass.osgeo.org/
3 http://www.dvbtmap.eu/mapglobal.html?tid=8228

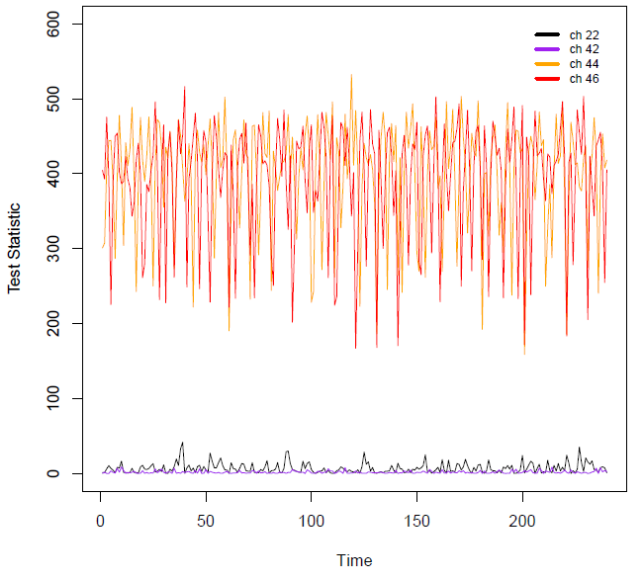

Figure 10: Comparison of test statistics in unoccupied (22,42) and occupied (44,46) channels in a random, but fixed location. The difference in signal amplitude is noteworthy between the two cases.

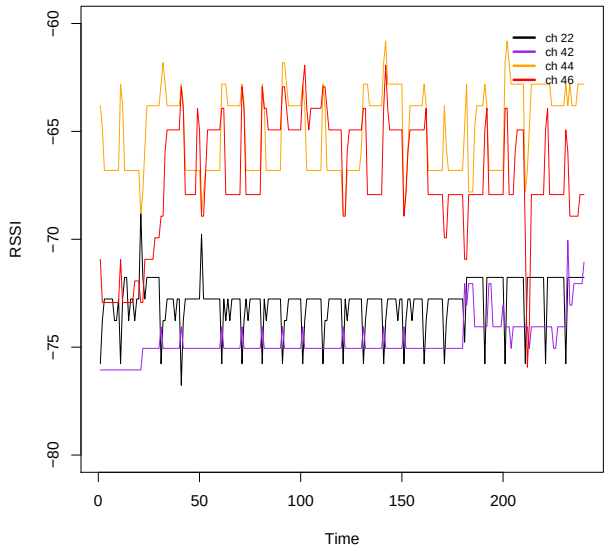

Figure 11: Comparison of test RSSIs in unoccupied (22,42) and occupied (44,46) channels in the same random, but fixed location. Even if smaller in percentage a clear difference than in the test statistics case, a clear difference is still present between the two channel occupations.

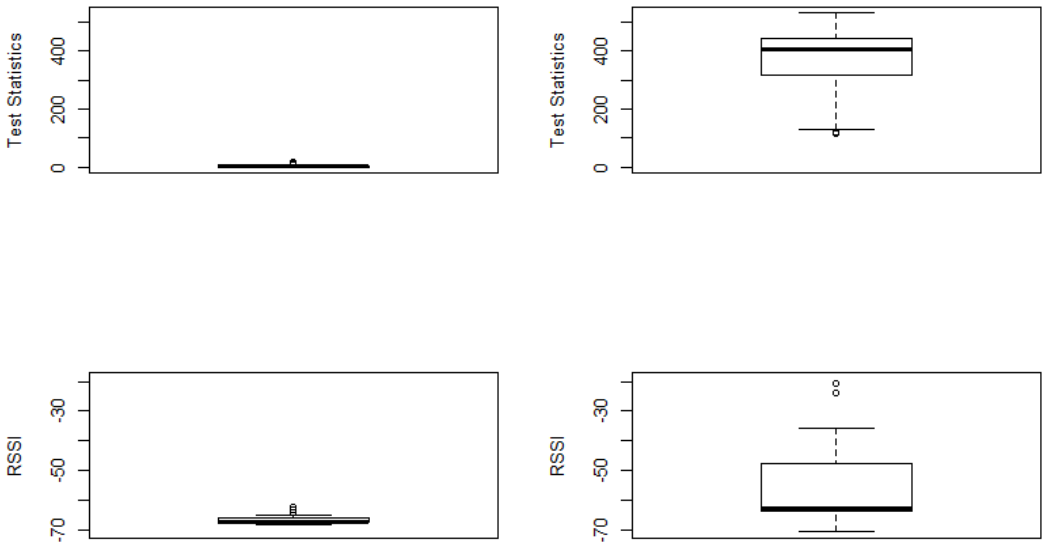

Figure 12: Boxplots for the comparison of an idle channel (22, left) and an occupied one (44, right). In the top plots the Test Statistics is used, in the bottom ones the RSSI is the chosen index.

4.5 DATA DISTRIBUTION

Let us start validating the results presented in the theoretical part for what concerns the distribution of the test statistic in the case of a cyclostationary detector.

4.5.1 Unoccupied Channel

We had foreseen that the distribution of the test statistic in an unoccupied channel, therefore under $\mathcal{H}_0$, had to be asymptotically χ^2_2 distributed, since there is only one fixed value for the lag. This can be easily verified, as can be seen in Figure 13, where a graphical analysis is performed comparing the empirical histogram with the theoretical one and drawing the quantile-quantile plot, that allows to compare the theoretical quantiles with the sampled ones. The result is quite satisfying, therefore the distributional assumption is validated.

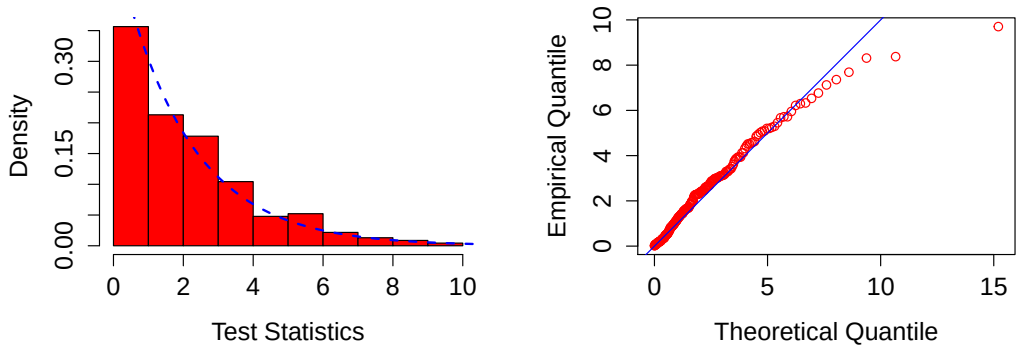

Figure 13: Histogram (left) and quantile quantile plot (right) comparing theoretical (blue line) and empirical (red) distribution of Test Statistics of idle channel 21 in a single random location.

4.5.2 Occupied Channel

The distribution of the Test Statistics in an occupied channel is theoretically asymptotically normal, in reality, the distribution does not result to be symmetric in every location, see Figure 14.

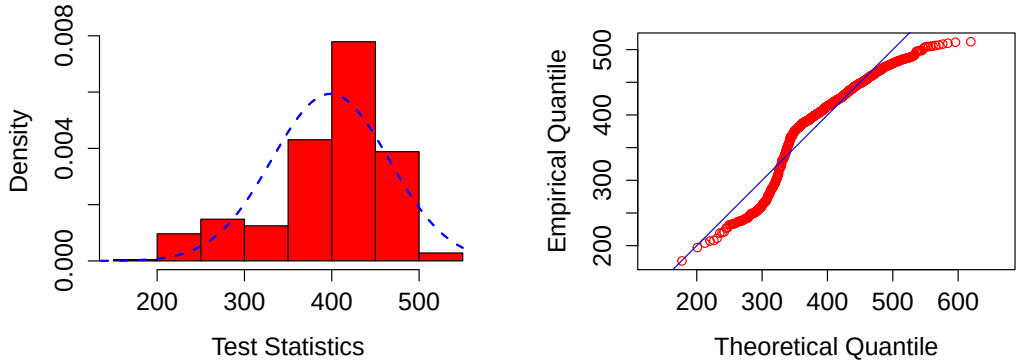

Figure 14: Histogram (left) and quantile quantile plot (right) comparing theoretical (blue) and empirical (red) distribution of Test Statistics of occupied channel 44 in a single random location.

Even if the qualitative analysis of the local distribution in the occupied channel in some locations does not exactly match the theoretical distribution and the Shapiro Test refuses the hypothesis of normality, the choice is to consider it valid in every location, for simplicity and because in the vast majority of cases it results to be verified.

5 DETERMINISTIC INTERPOLATION

Spatial spectrum models have as objects the generation of scenarios for evaluation of algorithms and protocols concerning the DSA.
Object of spectrum models is the generation of scenarios for evaluation of algorithms and protocols concerning the DSA.
As already suggested in [72], deterministic models do not fully consider the characteristics of a real network and therefore they are mostly useful in areas where no sampling data are available, while interpolation and prediction modeling are otherwise preferred.
The aim of this section is to briefly introduce the available interpolation techniques for geo-referentiated data. These are mostly historic models, but they will be shortly listed for the role they have played and still play in famous applications.
Let's start looking at those interpolators that are called *exact*, meaning that in the sampled points they exactly assume the sampled value.

5.1 THIESSEN OR VORONOI INTERPOLATION

The method, named after the Ukranian mathematician Georgy F. Voronoy, who defined and studied the general n-dimensional case in 1908, is tipically referred to as Thiessen tassellation in Geostatistics, in honour of the American meteorologist Alfred H. Thiessen.
The idea is straightforward: once the points where the variable of interest has been sampled are known, the measurement region is divided into subregions, each consisting of all the locations which are closer, in a pre-specified metric, to a particular sampled point. These are called *Voronoi cells* or *Thiessen polygons* and define the area of influence of a particular sampled point, see e.g. Figure 15.
More formally let (D, d) be a metric space, endowed with the distance function d and let $\vec{s}_i,\ i = 1, \ldots, N$ be the the measurement locations; the Voronoi cell V_i, associated with the site $\vec{s}_i$, is the set of all points in D whose distance from $\vec{s}_i$ is not greater than their distance to any of the other sites $\vec{s}_j, j \neq i$. In other words, if $d(\vec{s}, D) = \inf\{d(\vec{s}, \vec{a}) \mid \vec{a} \in A\}$ denotes the distance between the point $\vec{s}$ and the full set D, then

$$V_i = \{\vec{s} \in D \colon d(\vec{s}, \vec{s}_i) \leq d(\vec{s}, \vec{s}_j)\ \forall j = 1, \ldots, N,\ j \neq i\}. \tag{5.1}$$

The Voronoi diagram is simply the tuple of cells $V_i,\ i \in 1, \ldots, N$. In principle some of the sites can intersect and even coincide, but usually they are assumed to be disjoint. In addition, infinitely many sites are allowed in the definition, but again, usually only finitely many sites are considered.

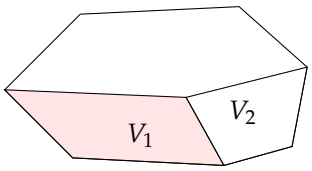

Figure 15: Example of Voronoi cells.

In applications the space is usually finite-dimensional Euclidean and all the measurement points differ, and therefore the Voronoi cells are convex polytopes, however, in general, they may not be convex or even connected.
The idea is to use this partition of the area for interpolation, setting the value of the field $Z(\vec{s})$ in an unsampled location $\vec{s}$ to be the value of the field in the only location in the area where a measurement has been done:

$$\forall\, \vec{s} \in V_i \;\hat{Z}(\vec{s}) := Z(\vec{s}_i). \tag{5.2}$$

5.1.1 Implementation

The implementation takes use of the R [58] package `deldir` to build the tassellation, then a value is associated to any tile, depending on the closest sampled value.
For details refer to the Listing chapter.

5.1.2 Results

From MC3, for an occupied channel (44) and an unoccupied one (22), for both RSSI and Test Statistics an interpolation has been conducted, taking for every sampling location the temporal median and from these the idea is to produce a interpolation map of all the Helsinki area.
In Figure 16 the tiles geometry is presented: it is clear that not all the polygons have the same area; polygons are smaller in the center compared to the docks because the sampled locations are closer. The decision has been to limit the polygons in a rectangular frame, to limit the external ones.

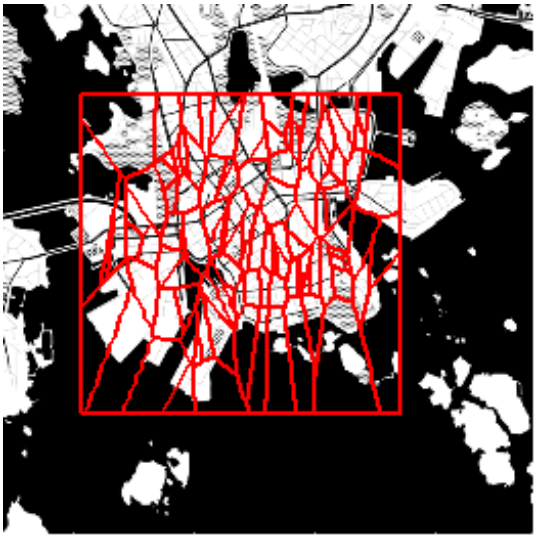

Figure 16: Tiles using a standard Thiessen interpolation.

Let us start comparing the results for the RSSI, between an idle (22) and occupied (44) channel in Figure 17: it is clear that there are a lot of discontinuities between the tiles and, even for two channels, whose occupancy is really different (see Figure 10), it happens that differences between tiles are bigger in absolute value compared to those between the channels themselves.
In Figure 18, instead, where the same operation has been conducted for the Test Statistics, instead than the RSSI, differences between tiles are still noticeable, but, since, as it has already been noticed, the test statistics presents a lot more difference in absolute value between channels that are occupied and those that are not, compared to RSSI, the interpolation looks more homogeneous between the tiles, and therefore, in a cognitive radio prospective, more useful.

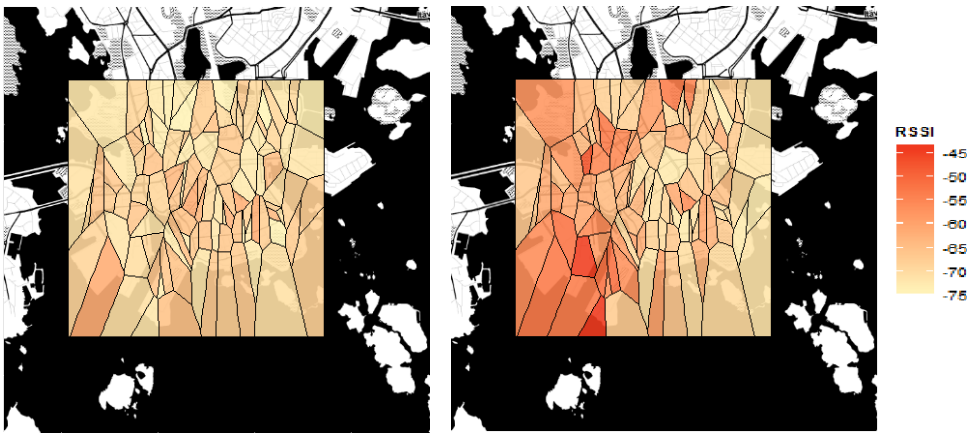

Figure 17: Results of the Thiessen interpolation of the RSSIs in an unoccupied channel (22, left) and occupied one (44, right). The transmitter for channel 44 is out of the drawn area, it is visible in Figure 9.

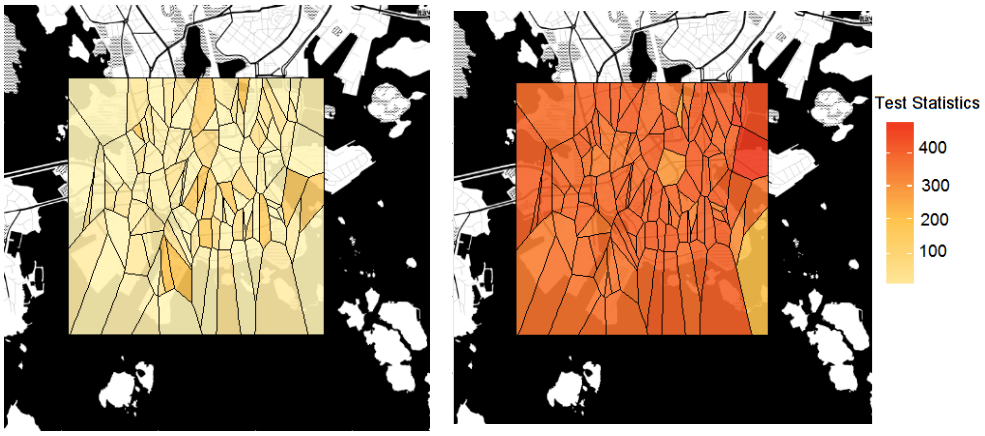

Figure 18: Results of the Thiessen interpolation of the Test Statistics in an unoccupied channel (22, left) and occupied one (44, right). The transmitter for channel 44 is out of the drawn area, it is visible in Figure 9.

5.1.3 Notes

Even though the technique is easy to use and implement, making it perfect in a static CR environment, it is not optimal by any means from the point of view of mobile users, since in that case the polygons would have to be re-computed continuously, making this approach, that already furnishes a coarse approximation, also computationally unfeasible. Let us make a brief summary of the technique characteristics, considering a CR prospective:

- All the predictions are based on the measurements in a single location, making the estimate very easily subject to errors and local bias, if the sensors are not all perfectly fine-tuned.
- The estimation is raw and "stepwise", since there is no continuity on the borders of the cells.
- There is no local error estimation, the only index is the local standard deviation, but that is valid only in the measurement location, not in the rest of the tile.

Consequently the technique could be a good choice in a static setting for its very low computational cost, but is really inadequate in the mobile environment.

5.2 WEIGHTED THIESSEN INTERPOLATION

This technique is the natural response to some of the critics moved to Thiessen's tassellation: it is still a simple and local interpolation, but the interpolated value depends now on three measurements, instead than only one.
Let $\vec{s}_0$ be the location where the prediction is to be calculated and $\vec{s}_i, i = 1, \dots N$ be the sampling locations, then

$$\hat{Z}(\vec{s}_0) = \sum \lambda_i Z(\vec{s}_i). \tag{5.3}$$

Where $\lambda_i = 0$ for all the vertices except the three closest ones and the non-zero λ_i are arbitrary and can depend, as an example, on the normalized distance from the new point.

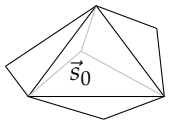

Figure 19: Example of triangulation.

5.2.1 Implementation

In this case the weights chosen for every $\vec{s}_0$ are the following:

$$\lambda_k := \begin{cases} \frac{1}{d(\vec{s}_0, \vec{s}_k)} & \text{if } \vec{s}_k \text{ is one of the 3 closest neighbours of } \vec{s}_0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

In our implementation the interpolation is done point per point in a regular 400×400 grid, in the rectangular area that can be seen in Figure 16. The implementation is similar to the standard Thiessen, apart for the weighting, see the Listing chapter.

5.2.2 Results

These are two estimations from the measurements coming from MC3, for an unoccupied (22) and an occupied (44) channel, using the Test Statistics.

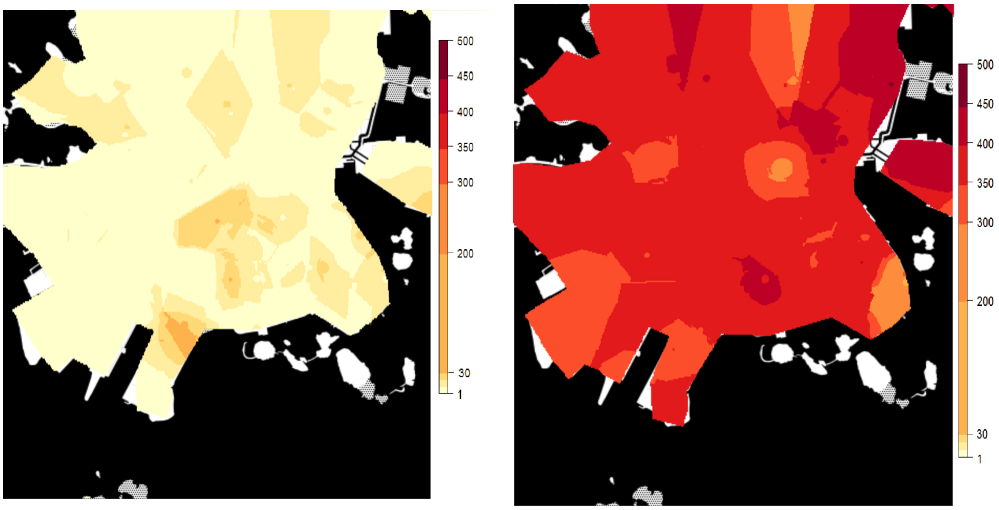

Figure 20: Results of the Weighted Thiessen interpolation of the Test Statistics in an unoccupied channel (22, left) and occupied one (44, right). The transmitter for channel 44 is out of the drawn area, it is visible in Figure 9.

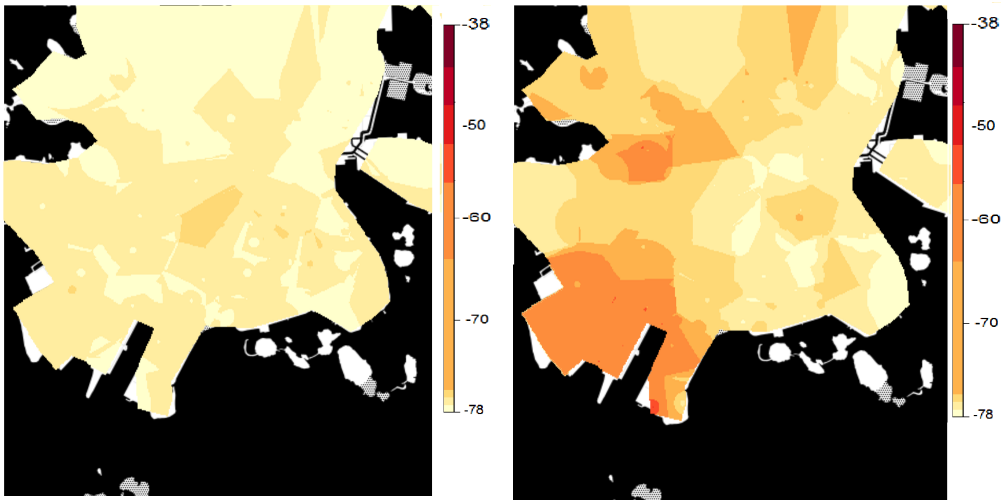

Figure 21: Results of the RSSI interpolation of the Test Statistics in an unoccupied channel (22, left) and occupied one (44, right). The transmitter for channel 44 is out of the drawn area, it is visible in Figure 9.

Qualitatively the results presented in Figures 20 and 21 seem correct, since values for idle channels are way lower than the ones for occupied ones and in these last it is possible to see how the strength of the signal diminishes from south-west radially (the transmitter is located in Espoo, see Figure 9). It is natural to compare the two previous interpolation techniques: to have a qualitative idea the effective difference between interpolation techniques we proceed first with boxplots, which probably make it easier for the comparison to take place.

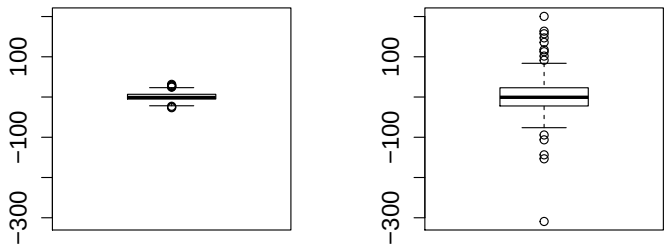

Figure 22: Boxplots of the differences in Test Statistics units between weighted and standard Thiessen interpolation in an unoccupied channel (22, left) and an occupied one (44, right).

To see if effectively, as the boxplots in Figure 22 seem to suggest, the differences are zero-mean distributed, therefore asserting that it is useless to utilize a more complicated and therefore computationally intensive interpolation technique, we proceed with a paired t-hypotheses test. This test is used because interpolation techniques are applied to the same

samples and consequently the results can be considered to be paired per location, moreover the true deviation is unknown, but the sample size is quite high.
Defining $\bar{d}$ the sample mean of the difference between the interpolated values of the test statistics across time in each grid point using the two techniques and s_d the standard deviation of d, it is possible to build a test statistics that behaves as a t of Student distribution with $n-1$ degrees of freedom:

$$T = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} \sim t_{n-1}. \tag{5.5}$$

The test can be stated as:

- If $\mathcal{H}_0$ cannot be refused the interpolation techniques are in average the same and therefore there is no need to complicate the interpolation technique.
- If, on the other side, $\mathcal{H}_1$ is selected, the interpolation techniques show a difference in mean and further reasoning is needed to decide which technique is to be chosen and why.

The results are the following:

- Unoccupied Channel 22

  $$T = 0.7985, \quad p - value = 0.4263. \tag{5.6}$$

  The sample mean of the difference is $\bar{d} = 0.89$ therefore o clearly belongs to the 95 percent confidence interval $IC_{0.95} = [-1.32, 3.10]$ and we cannot refuse the $\mathcal{H}_0$, for which the true difference has zero mean.

- Occupied Channel 44

  $$T = 0.3583, \quad p - value = 0.7208. \tag{5.7}$$

  The sample mean of the difference: $is\bar{d} = 2.20$ therefore o clearly belongs to the 95 percent confidence interval $IC_{0.95} = [-9.96, 14.36]$ and we cannot refuse the $\mathcal{H}_0$, for which the true difference in mean is zero-mean.

All this applies to the interpolation using the RSSI, for brevity only the test results are reported here:

- Unoccupied Channel 22 This is the only case in which $\mathcal{H}_0$ should be refused, since

  $$T = 3.9775, \quad p - value = 0.0001263. \tag{5.8}$$

  The sample mean of the difference is $\bar{d} = 1.25$ and therefore o does not belong to the 95 percent confidence interval $IC_{0.95} = [0.63, 1.88]$. In this case there is statistical significance to assert that the true difference has not zero mean, but been the interval so close to the origin, it does not seem to make sense to use a much more complex technique.

- Occupied Channel 44

  $$T = -1.4469, \quad p - value = 0.1508. \tag{5.9}$$

  The sample mean of the difference is $\bar{d} = -1.14$ therefore o clearly belongs to the 95 percent confidence interval $IC_{0.95} = [-2.71, 0.42]$ and we cannot refuse the $\mathcal{H}_0$, for which the true difference has zero mean.

5.2.3 Notes

Unfortunately the technique still suffers of some of the issues that characterized "pure" Thiessen's technique:

- There is no error estimate.
- The prediction surface is now continuous, but the gradient changes abruptly on the border of the area.
- No weighting is a priori better than another, but the results that can be obtained are quite different, depending on the chosen weights.
- In our implementation the interpolation is done point per point in a regular 400×400 grid, the computational cost is way higher than in the standard Thiessen interpolation because for any location the distance from all the sampling location has to be built, while before just one Voronoi tessellation was enough before.

A higher computational cost, together with a non existing improvement in mean, as shown by the t-tests, make this weighted interpolation not a better candidate as a future interpolation model in a cognitive radio environment, mostly from a mobile technology point of view, where battery life is a concern.

5.3 NATURAL NEIGHBOUR INTERPOLATION

This technique due to Sibson is built on top of the Thiessen interpolation technique, with the intent of having a smoother interpolation.
For every unsampled $\vec{s}_0$, the Thiessen tile to which it belongs is built, see Figure 24; Sibson refers to the vertices as *natural neighbours*, hence the technique denomination. Let us define A_i as the area of the original tile corresponding to the i^{th} vertex that now belongs to the new tile, in Figure 23 they are visualized for $i = 1, \ldots, 4$.
For every location in which to interpolate a full set of N weights is calculated, one for every sampled location, proportional to the percentage of the original tile incorporated in the new one:

$$\lambda_{0,i} := \begin{cases} \frac{A_i}{\sum_{k=1}^{N} A_k} & \text{if } \vec{s}_i \text{ is a neighbouring location,} \\ 0 & \text{otherwise.} \end{cases} \tag{5.10}$$

The interpolated field in the new location is therefore

$$\hat{Z}(\vec{s}_0) = \sum \lambda_i Z(\vec{s}_i). \tag{5.11}$$

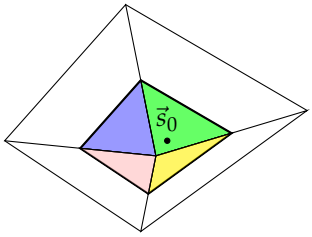

Figure 23: Example of natural neighbour interpolation. The area with the ticker border is the new tile, this intersects four pre-existing tiles. The intersecting $A_i, i = 1, \ldots, 4$ are colored differently.

Figure 24: Evolution of the Thiessen tassellation, while adding the evidenced location.

5.3.1 Implementation and Results

Unfortunately the technique has been proven by Laslett to be unacceptable with noisy data, see [71] and another big issue is that it is not possible to extrapolate with it, we opt for a linear extrapolation outside the convex hull, with disastrous results, see Figure 25. The listing can be visualized in the Listing chapter, but the idea is to build a new tassellation for every location in the 400×400 grid, calculate the intersections with the pre-existing grid and consequently calculate the weights as aforementioned.

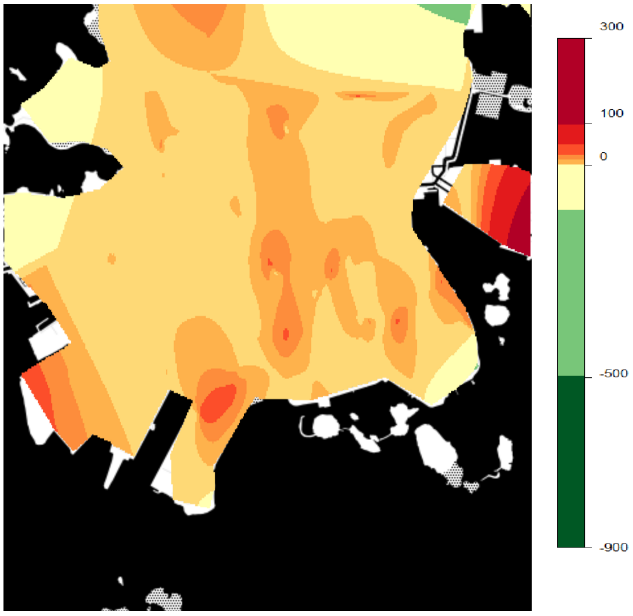

Figure 25: Natural Neighbours interpolation of Test Statistics, it is clear that outside the sampling locations hull the result is absolutely meaningless, since the one considered is an idle channel and interpolated values are very low or very high, considering test statistics varies typically in $[0, 40]$ in these situations, however always remains positive.

5.4 INVERSE DISTANCE WEIGHTING

This technique is still quite in use between soil scientists; the idea is to weight the measurement samples depending on the distance between the desired point and the sampling locations, so that:

$$\hat{Z}(\vec{s}_0) = \sum_{i=1}^{N} \lambda_i Z(\vec{s}_i) \tag{5.12}$$

where the weights $\lambda_i, i = 1 \dots N$ are proportional to the power of the distance between the measurement point and the unsampled location: $\lambda_i = \frac{1}{d(\vec{s},\vec{s}_i)^\beta}$. A value for β needs to be chosen, the most popular one is $\beta = 2$, because it allows a local prediction, since the distance goes to zero quite fast. This interpolation technique has the advantage of been continuous and a lot smoother than the preceding ones, on the other side the weighting function is arbitrary and there is no account for the directional configuration of the sampling design, that in cases where anisotropy is strong, can be a big issue.
In case $\beta = 1$ it is very similar to the Thiessen interpolation shown beforehand, with the difference that here all samples are used, not only the three closest ones.

5.4.1 Implementation

The implementation takes advantage of the `gstat` package for R, the chosen grid is the same 400×400 points, $\beta = 2$ is the chosen power for a theoretically smoother result.

5.4.2 Results

The results, shown in Figure 26 and 27 seem to confirm the quality of the implemented interpolation for $\beta = 1$.

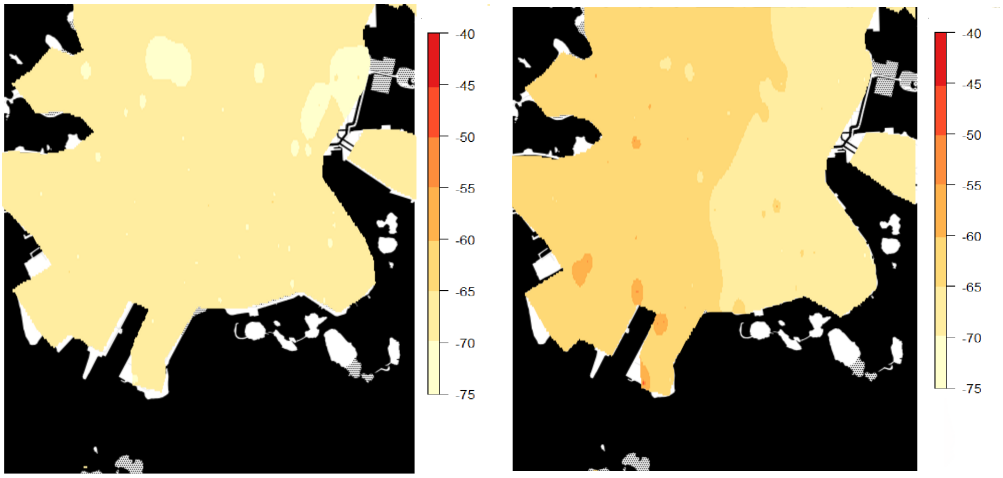

Figure 26: RSSIs in unoccupied (22, left) and occupied (44, right) channels, using $\beta = 2$. The transmitter for channel 44 is out of the drawn area, it is visible in Figure 9.

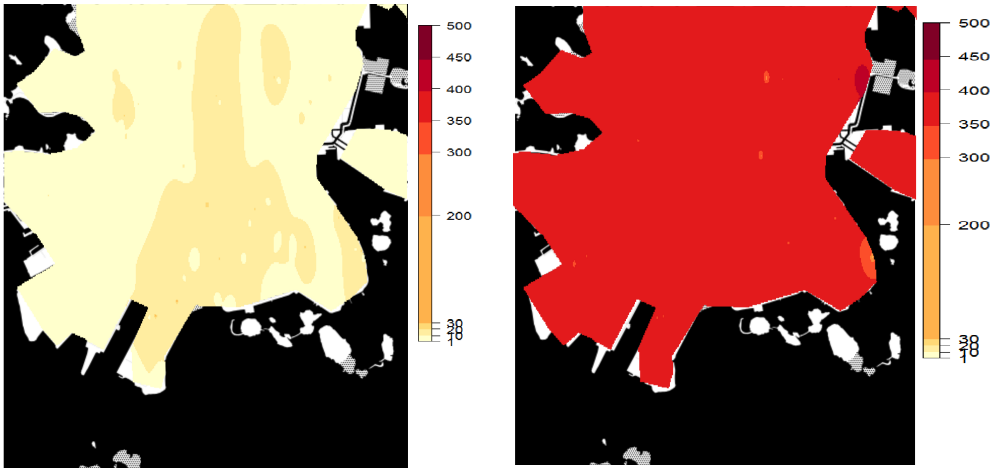

Figure 27: Test Statistics in unoccupied (22, left) and occupied (44, right) channels, using $\beta = 1$. The transmitter for channel 44 is out of the drawn area, it is visible in Figure 9.

In Figure 28 the interpolation is presented in the case of $\beta = 2$, the one that is considered to be almost the standard for IDW technique. It is to be noticed how much the only degree of freedom impacts the result.

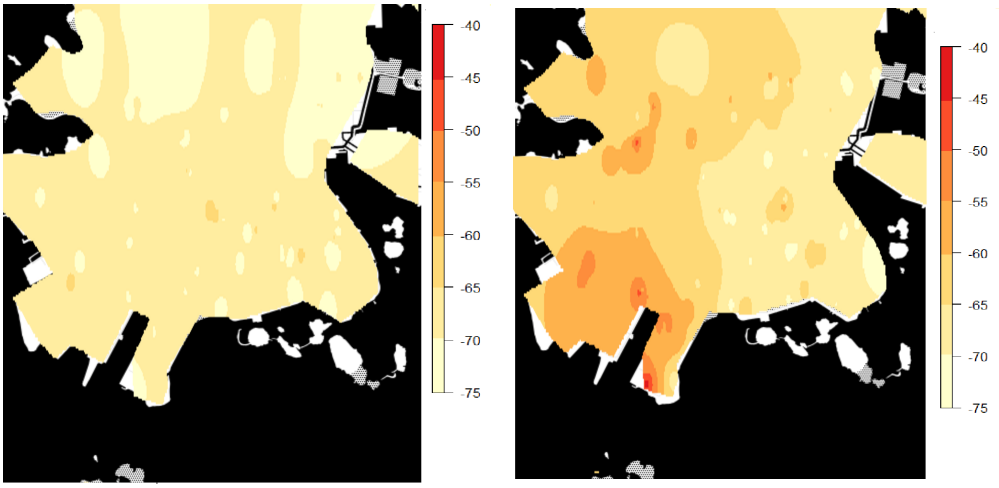

Figure 28: RSSI in unoccupied (22, left) and occupied (44, right) channels, using $\beta = 2$.

5.4.3 Notes

Excluding the natural neighbors interpolation, that has been proven to be not adapt for our aims, we opt for comparing IDW interpolation with the weighted Thiessen tessellation. Boxplots of the differences can be seen in Figures 29 and 30.

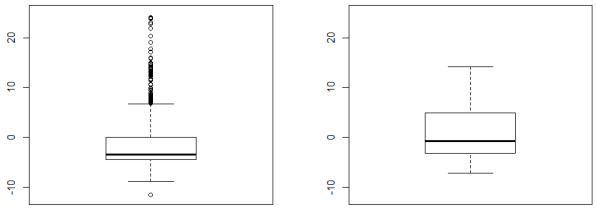

Figure 29: Boxplots of the difference of Test Statistics using IDW and Thiessen weghted tessellation in unoccupied (22, left) and occupied (44, right) channels. The transmitter for channel 44 is out of the drawn area, it is visible in Figure 9.

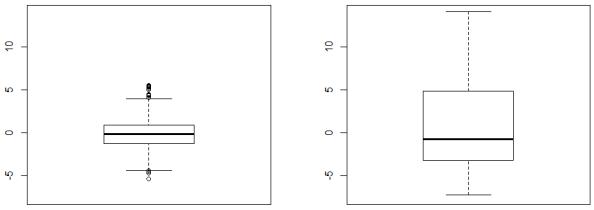

Figure 30: Boxplots of the difference of RSSI using IDW and Thiessen weghted tessellation in unoccupied (22, left) and occupied (44, right) channels.

Paired t-test have been effectuated and in all the cases the null hypothesis has been refused, the interpolation using the IDW is negatively biased, since it tends to smooth the values while interpolating them.

5.5 TREND SURFACES

This is an old technique that is basically a multiple regression, where the only allowed predictors are the coordinates,

$$Z(\vec{s}) = f(\vec{s}) + \epsilon \tag{5.13}$$

where the predictor can be a plane:

$$f(\vec{s}) = b_0 + b_1\vec{s}_1 + b_2\vec{s}_2 \tag{5.14}$$

or a higher order polynomial, such as a quadratic

$$f(\vec{s}) = b_0 + b_1 s_1 + b_2 s_2 + b_3 s_1^2 + b_4 s_2^2 + b_5 s_1 s_2 \tag{5.15}$$

where $\vec{s} = (s_1, s_2)$.
The interpolation is therefore transformed in the solution of a linear system:

$$\vec{b} = (X^T X)^{-1} X^T \vec{c} \tag{5.16}$$

where, in the case of a planar interpolation, given the sampled location $\vec{s}_i = (s_{i,1}, s_{i,2})$, $\forall i = 1, \ldots, N$ and a location where to sample $\vec{s}_0 = (s_{0,1}, s_{0,2})$, the system can be written as:

$$X = \begin{bmatrix} 1 & s_{1,1} & s_{1,2} \\ \vdots & \vdots & \vdots \\ 1 & s_{N,1} & s_{N,2} \end{bmatrix} \tag{5.17}$$

$$\vec{c} = \begin{bmatrix} 1 & s_{0,1} & s_{0,2} \end{bmatrix} \tag{5.18}$$

and finally

$$\hat{Z}(\vec{s}_0) = \vec{s}_0^T \vec{b}. \tag{5.19}$$

If the interpolating polynomial has a higher degree, the eventual columns are easily added to the X matrix.

The technique in not optimal in any sense, but is for sure way computationally more convenient than kriging and is therefore to be considered in settings such as mobile cognitive radios.
Even though it has been shown by Mofflat et al. that trend surfaces can interpolate well long range trend, and could therefore be ideal in our situation, they present some issues, such as:

- The necessity for high order polynomials to properly fit the surface that comports matrix instability and over-fitting problems.
- A global and non-local fitting, that does not guarantee a high quality of the interpolation locally, but could be enough to determine if a channel is or not idle.
- The technique leaves autocorrelated residuals, somehow not using all the information available in the sample.

5.5.1 Implementation

We have taken advantage of the `spatial` package in R, but an implementation "by hand" is surely possible and not time consuming, even though, potentially, maybe not optimal from a computing power point of view.

5.5.2 Results

First, we start comparing the interpolation of the Test Statistics using two widely different degrees of the interpolating polynomials.

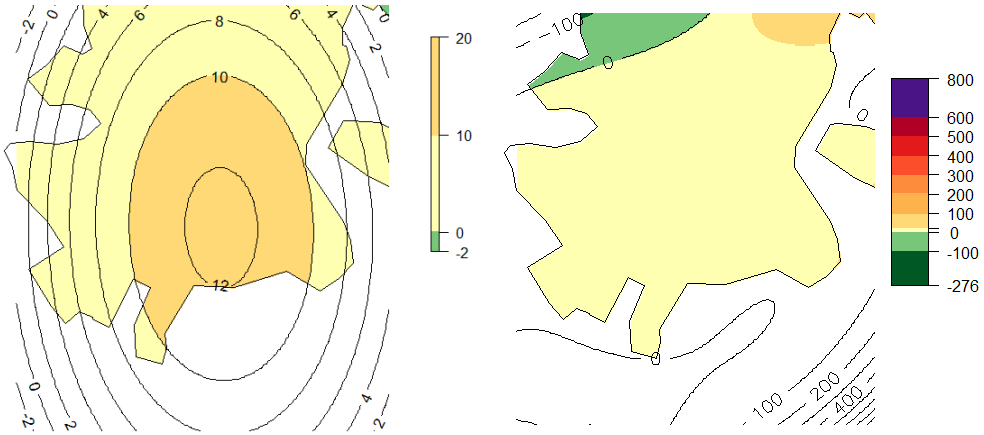

Figure 31: Interpolation of the Test Statistics in unoccupied channel (22) with a 2 (left) and a 5 degree polynomial interpolant (right). We preferred not to report the map in the background, because we thought the contour lines were more revealing in the case and they were more clearly visualized in the present fashion.

The difference between the two interpolations shown in Figure 31 is very wide, the problem is that in our case the sampling locations are very numerous, but we cannot assume

that in a real case, as many as a hundred sampling locations will be available. For this reason a high-degree polynomial interpolant would be troublesome if not in a computational point of view surely for a robustness of the whole algorithm. An option could be to set the interpolant degree and therefore the quality of the results iteratively depending on the number of sampled locations, but being the results so far apart, we are not sure this is the way to go.
It can be seen in Figure 31 in the higher order polynomial interpolation (right), that far from the sampling locations, e.g. in the South-Eastern corner values are already growing uncontrollably because of the non-robustness of the method, on the other side, the shape of the level-line is way more similar to other more computationally intensive techniques, therefore, once again, this technique does not suit extrapolating and can therefore be problematic in a cognitive radio environment.
Nonetheless, because of the simplicity and the low computational cost, we proceed in Figure 32 in plotting Test Statistics and RSSI with a quadratic interpolant; the graphs show another clear issue of trend surfaces: outliers cause a high impact on them. This is clear if comparing the two pictures, because they show a complete different pattern in the data: it is clear that this technique, which is not suitable for numerous static measurements will surely fail in a much less controlled system.

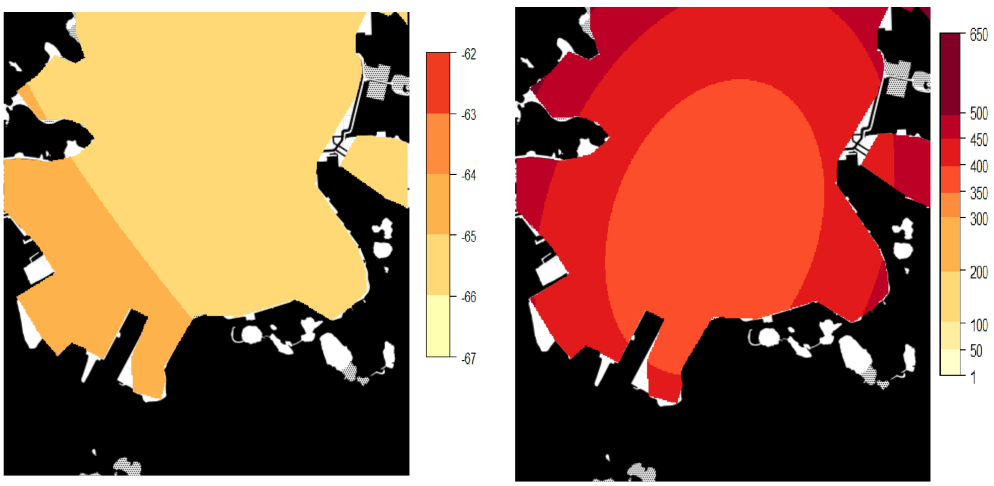

Figure 32: Interpolation of the RSSI (left) and of the Test Statistics (right) in an occupied channel (44) with a 2 degree polynomial interpolant. In the left plot the level curves are ellipsoidal as in the right one, they are simply bigger, the reason is that the decision has been to keep constant the colour palette between figures, to ease the comprehension, even if in few cases, as in this the image is maybe not optimal for the intent from other points of view.

The interpolation of the Test Statistics looks incorrect in Figure 32, since as it has been said the power should decrease radially from South-West, see Figure 9. To validate this thought quantitatively we compare the trend surface estimation with the interpolation that had been obtained with the IDW, to see if the interpolated field can be considered to have the same mean. As usual, we proceed with a paired t-test, that, non surprisingly, refuses the hypothesis of same mean for the two interpolations:

$$T = 218.4159, \quad p - value < 2.2 \times 10^{-16}. \tag{5.20}$$

The sample mean of the difference is $\bar{d} = 0.89$ and clearly 0 does not belong to the 95 percent confidence interval $IC_{0.95} = [17.33, 17.65]$, and therefore $\mathcal{H}_0$ must be refused. This is not surprising since the IDW is known to smooth the interpolation results, while the trend statistics has been shown to emphasize the outlier behaviour.

5.6 SPLINES

Splines are local polynomials, that allow continuous interpolation, derivable up to $p-1$ order, where p is the degree of the polynomials, which is usually $p = 3$.
The locations where the different polynomials join are called *knots*.
Splines are subdivided into two main groups: *Interpolating Splines* which are forced to pass for the sampled data values, and *Smoothing Splines* fitted by Least Squares (LS), on which we will focus on.

5.6.1 Implementation and Results

We have taken advantage of the `fields` package for R. In Figure 33 some of these prediction maps are shown, they result to be more realistic and more robust than trend surfaces.

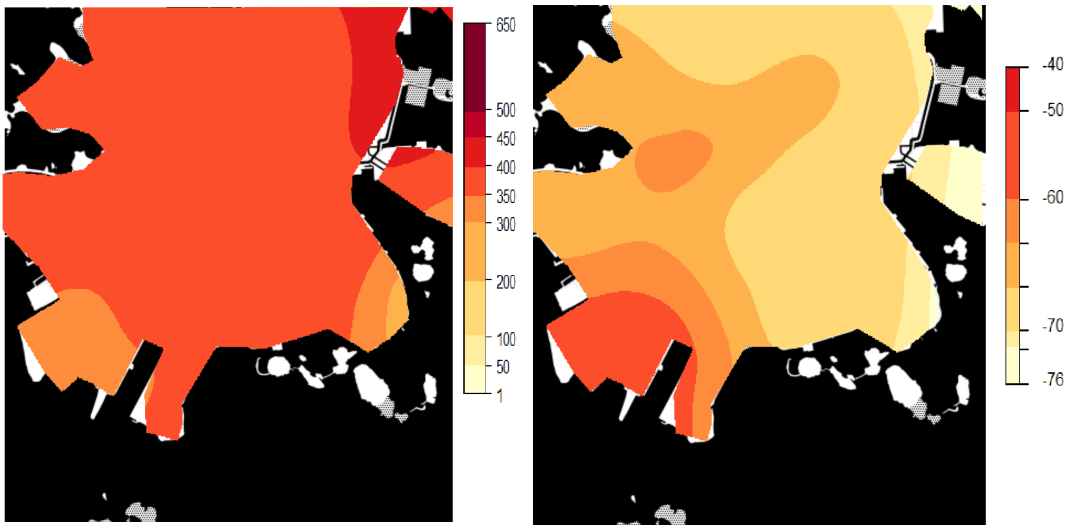

Figure 33: Prediction of Test Statistics (left) and the RSSI (right) in an occupied field (44). The strong area in the North-Eastern corner is difficult to explain, since it is not present using the RSSI. A cluster of sensor is strongly influencing the interpolation, as was happening with the Trend Surfaces in the occupied case.

As a first comment we state that this method is way less sensible to outliers presence, even though still being of relatively easy implementation . To quantify this a simple leave-one-out crossvalidation has been implemented to compare the result to trend surfaces, which have been said to be particularly sensible to outliers, see Table 2.

Interpolation Technique	Mean Squared Deviation	Relative Error
Trend Surfaces	56.99	15%
Thin Plate Splines	7.46	< 2%

Table 2: Comparison between trend surfaces and thin plate splines for the interpolation of Test Statistic in occupied channel 44.

Splines really seem to be the way to go, they give a prediction, not just an interpolation of the field and have been shown to perform quite well. But, since the signal is spatially correlated, this can be exploited for interpolation and prediction, at the cost of much heavier computations, but with numerous vantages, as we will show.

5.7 COMPARISON BETWEEN DIFFERENT NON-STATISTICAL TECHNIQUES

In Table 3 a comparison between different interpolation techniques is performed, the number of star is accounted to each method for how well it performs. Results are of course valid only referring to this specific data-set, but we still consider it interesting and highly informative.

Interpolation Technique	Computational Complexity	Ability to Extrapolate	Robustness MSE (1CV)
Thiessen	*		**
Weighted Thiessen	**		**
Natural Neighbor	***		*
Inverse Distance	**	**	***
Trend Surfaces	*	**	
Splines	**	***	***

Table 3: Comparison between deterministic interpolation techniques according to different factors, stars are assigned in a number (0 to 4) that varies depending on the aforementioned quantitative results.

6 GEOSTATISTICS

Extracting patterns from spatial data is a complicated matter because of the high complexity of spatial data types, spatial relationships and spatial autocorrelations.
According to Shekhar in [62], modeling can be done in three possible ways:

POINT PROCESS model for the spatial distribution of points in a point pattern, where the location of the next random process is not known (e.g. crime spots),

LATTICE model for gridded space in a spatial framework, where a countable collection of spatial sites relates to itself via neighborhood relationships, often modeled through a *contiguity matrix*, based on distance and different forms of connectivity, once a definition of spatial neighbourhood has been agreed upon, on which the robustness of the methods depends,

GEOSTATISTICS deals with analysis of spatial continuity and weak stationarity, it provides a set of statistic tools, such as Kriging, to the interpolation of attributes at unsampled locations.

One of the fundamental assumptions in traditional statistics is that data samples are independently generated; this is generally false when talking about spatial data, see Tobler's first law of geography: *Everything is related to everything else, but near things are more related than distant things*[65]; this basically states that spatial correlation is typically strong and positive between data at nearby spatial locations. Spatial hetereogeneity, spatial nonstationarity, or the variability of observed processes over space, are another important issues.
The aim of Geostatistics is to find appropriate models for the description of the spatial reality of a phenomenon, to estimate the relative parameters and to infer the values of the random field at unobserved locations.
The theory was developed by the French mathematician Georges François Paul Marie Matheron (1930, 2000) based on the Master Thesis of the South African mining engineer Danie Gerhadus Krige (1919, 2013), after whom Kriging is named. The original problem was that

> Mining leases for deep gold mines were [granted] on the basis of very few boreholes and without proper scientific analysis of the scant data. This situation posed very hight risks to profitability but there seemed to be no real alternative.[70]

The strength of Geostatistics, compared to more classical statistical tools is that it models trend (*large scale*) and spatial variability (*small scale*) at the same time.

6.1 THE MODEL

The aims of a stochastic model are multiple, at the very least it should be used to summarize data and predict unobserved field values, it may (or may not) have a causative dynamic that explains the phenomenon itself.

The first step is to create the random process that best describes the set of experimental observed data.
Let $\vec{s}$, the location vector, vary continuously in D, fixed subset of $\mathbb{R}^d$, $d \geq 2$, so that for every measurement at least longitude and latitude are recorded and let $\vec{Z}(\vec{s}) : \vec{s} \in D$ be a random field, a spatial process of which $\vec{z}(\vec{s}) : \vec{s} \in D$ is a realization.
In the space there are N realizations of the correlated random variables $\vec{Z}(\vec{s}_1), \vec{Z}(\vec{s}_2), \cdots, \vec{Z}(\vec{s}_N)$, but only one realization per location could be available, consequently it is impossible to determine any statistical parameter of the individual variable $\vec{Z}(\vec{s}_i)$. In our case more measurements per location are available, but we opt for using the temporal median, for which robustness and local distributional hypotheses have already been verified. It is possible to consider spatio-temporal data $\vec{Z}(\vec{s}, t)$ but in this work the focus will be on data aggregated over time, that can be considered to be gathered in a single time instant.
In the mono-dimensional case $\vec{Z}$ is not a vector, but a scalar and the model for the random field is:

$$Z(\vec{s}) = m(\vec{s}) + \delta(\vec{s}) \quad \vec{s} \in D. \tag{6.1}$$

The *drift* or *trend* $m(\vec{s})$ is supposed to be deterministic, eventually dependent on external factors (*external drift*), while $\delta(\vec{s})$ is a stochastic term that defines the characteristics of the random field $Z(\vec{s})$, which is defined by the joint probability function, also known as *finite - dimensional law*:

$$F(z_1, \ldots, z_N) = \mathbb{P}(Z(\vec{s}_1) \leq z_1, \ldots, Z(\vec{s}_N) \leq z_N) \quad z_1, \ldots z_N \in \mathbb{R}. \tag{6.2}$$

6.2 STATIONARITY CONDITIONS

Because only one measurement per location is available, assuming various degrees of stationarity is necessary to infer some statistic properties.

Definition 3 (Strictly Stationarity) *$\vec{Z}(\vec{s}), \vec{s} \in D$ is said* Strictly *or* Strongly Stationary *if* $\forall \vec{h} \in D$, $\forall \{\vec{s}_k\}_{k=1}^N \in D$, $\vec{Z}(\vec{s}) = \left(Z(\vec{s}_1), \ldots, Z(\vec{s}_N)\right)$ *and* $\vec{Z}(\vec{s}+\vec{h}) = \left(Z(\vec{s}_1+\vec{h}), \ldots, Z(\vec{s}_N+\vec{h})\right)$ *have the same joint distribution.*

This assumption is very strong and difficult to verify, other weaker and more practical conditions follow.

Definition 4 (Weakly or Second Order Stationarity) *$Z(\vec{s}), \vec{s} \in D$ is said* Weakly *or* Second Order Stationary *if*

- $\forall\, \vec{s} \in D \qquad \mathbb{E}[Z(\vec{s})] = m$
- $\forall\, \vec{s}_i, \vec{s}_j \in D \quad \mathrm{Cov}(Z(\vec{s}_i), Z(\vec{s}_j)) = \mathbb{E}[(Z(\vec{s}_i) - m)(Z(\vec{s}_j) - m)]$

The first equation in the definition implies that the first moment of Z is stationary and consequently all random variables have the same mean m, that can be estimated e.g. by the arithmetic average of the sampled values.
At the same time the second equation implies that the covariance function between two random locations solely depends on the vectorial distance that separates them.
Strong stationarity implies the weak one, the reverse holds true only for Gaussian processes.

Definition 5 (Intrinsic Stationarity) $Z(\vec{s}), \vec{s} \in D$ *is said* Intrinsic Stationary *if, given that* $\mathbb{E}[Z(\vec{s}+\vec{h}) - Z(\vec{s})] = 0$ *it results that*

$$\mathbb{E}[Z(\vec{s}+\vec{h}) - Z(\vec{s})]^2 = Var(Z(\vec{s}+\vec{h}) - Z(\vec{s})). \tag{6.3}$$

The definition of Intrinsic Stationarity defines the moments of the difference $Z(\vec{s}+\vec{h}) - Z(\vec{s})$, but says nothing about the joint distribution of $Z(\vec{s})$ and thus provides no likelihood.

Definition 6 (Ergodicity) $Z(\vec{s}), \vec{s} \in D$ *is said to be* Ergodic *if the covariance vanishes when the distance between sampling points increases* $\lim_{\|h\| \to +\infty} Cov(Z(\vec{s}) - Z(\vec{s}-\vec{h})) = 0$.

Let's introduce two functions to describe the degree of spatial dependence of a stochastic field $Z(\vec{s})$.

Definition 7 (Covariance Function or Covariogram) *The* Covariance Function *or* Covariogram *is defined as:*

$$C(\vec{s}_i, \vec{s}_j) = \text{Cov}(Z(\vec{s}_i), Z(\vec{s}_j)) \; \forall \; (\vec{s}_i, \; \vec{s}_j) \; \in \; D. \tag{6.4}$$

The property of *validity* asserts that the covariogram is *positive definite*:

$$\forall \lambda_i, \lambda_j \in \mathbb{R}, \; \forall \; \vec{s}_i, \vec{s}_j \in D \; \sum_i \sum_j \lambda_i \lambda_j C(\vec{s}_i, \vec{s}_j) \geq 0. \tag{6.5}$$

Definition 8 (Variogram and Semivariogram) *The* Variogram *is defined as the variance of the difference between field values at two locations, under hypotheses of intrinsic stationarity.*

$$\gamma(\vec{s}_i, \vec{s}_j) = Var(Z(\vec{s}_i) - Z(\vec{s}_j)) = \mathbb{E}\left[\mid (Z(\vec{s}_i) - \mu(\vec{s}_i)) - (Z(\vec{s}_i) - \mu(\vec{s}_j)) \mid^2 \right]. \tag{6.6}$$

The prefix Semi *comes from the fact that the semivariogram is defined as half of the variogram, in this work these two terms will be treated as synonyms.*

The variogram has a series of interesting properties:

- It is non-negative: $\gamma(z(\vec{s}_i), z(\vec{s}_j)) \geq 0$,
- $\gamma(z(\vec{s}_i), z(\vec{s}_i)) = 0$,
- It is an even function $\gamma(z(\vec{s}_i), z(\vec{s}_j)) = \gamma(z(\vec{s}_j), z(\vec{s}_i))$,
- It might be non continuous only at the origin,
- It is conditionally negative (property of *validity*):

$$\forall w_i \colon \sum_{i=1}^{N} w_i = 0 \; \sum_{i=1}^{N} \sum_{j=1}^{N} w_i \gamma(\vec{s}_i, \vec{s}_j) w_j \leq 0. \tag{6.7}$$

It has to be noted that luckily the property of *validity* is invariant for scalar products and summation, so that, if $\gamma(h)$ is valid then $c\gamma(h)$ is valid as well for every $c \in \mathbb{R}$. Moreover if both $\gamma_1(h)$ and $\gamma_2(h)$ are valid so is $\gamma_1(h) + \gamma_2(h)$.
The choice of using the variogram instead than the covariogram is due to theorical results

that can be found in Cressie's work [16]. Basically if the process is second order stationary the simple relationship that links covariogram and variogram is:

$$\gamma(\vec{h}) = C(\vec{0}) - C(\vec{h}). \tag{6.8}$$

Weak stationarity implies the intrinsic one, for this reason spatial data structure is typically analyzed through the variogram and not the covariogram, since in cases where C is not defined $\hat{C}$ is *de facto* estimating a non existent parameter (see Cressie for a much deeper and exemplified dissertation).

6.3 ISOTROPY AND ANISOTROPY

If the process is *isotropic*, the phenomenon does not change in space depending on the direction, but only on the distance between the samplings. Covariogram and variogram can consequently be represented as a function only of the scalar distance $h = \mid z(\vec{s}_i) - z(\vec{s}_j) \mid$ it results:

$$\gamma(Z(\vec{s}_i), Z(\vec{s}_j)) = \gamma(Z(\vec{s}_i), Z(\vec{s}_i + \vec{h})) = \gamma(\vec{h}) = \gamma(h) \tag{6.9}$$

same applies to the covariogram:

$$C(Z(\vec{s}_i), Z(\vec{s}_j)) = C(Z(\vec{s}_i), Z(\vec{s}_i + \vec{h})) = C(\vec{h}) = C(h). \tag{6.10}$$

Anisotropy, on the other side, is a property of phenomenona that vary in magnitude differently in different directions. Assessing anisotropy involves the use of directional variograms, numerous authors state that, if no preferential direction is apparent *a priori*, the choice of these directions is absolutely arbitrary, typical in literature is the use of four angle classes $(0, 45°, 90°, 135°, 180°)$, with a tolerance in each class of $22.5°$, even though for what concerns the tolerance, the choice depends on the expected anisotropy and on the data density. For us a favourite direction exists, but being the transmitter-Helsinki direction almost in one of the coordinate direction (see Figure 9), the typical choice has been maintained, see Figure 34.
More types of anisotropy exist from a theoretical point of view, whether the sill, the range and/or the nugget vary depending on direction. A *caveat* is present in much literature (Cressie [16], Banerjee [2], Dauphin [18]) warning against reading too much from directional variograms, especially if very few samples are available.
Famous examples of anisotropy are:

GEOMETRIC ANISOTROPY, characterized by the fact that the covariate space can be transformed (through rotation and stretching) into an isotropic space.

ZONAL ANISOTROPY, characterized by a different asymptote of the variogram depending on binning direction, in this case the variogram has to be treated as $\gamma(\vec{h})$, no simplifications are available and this, of course, complicates the problem from both a theoretic and a pragmatic point of view. Typically in practice it is treated with a nested variogram model, together with a change of coordinates.

After analyzing numerous empirical variograms, not noticing a clear different pattern in any of the directions, see Figure 34, the option has been to maintain the hypothesis of isotropy that we know does not have a strong theoretical basis but allows to simplify the calculations and it is the almost only one that has been widely researched about in literature.

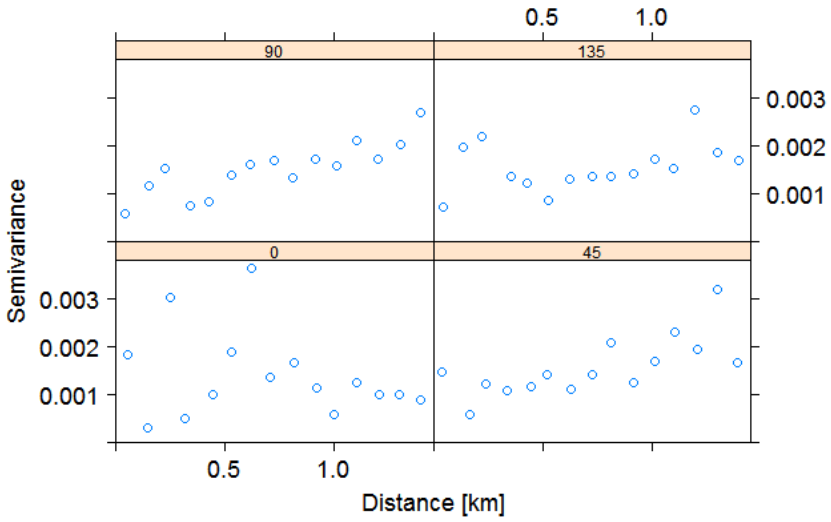

Figure 34: Example of directional variograms, to be read clockwise from bottom left. The sampled angles are $(0, 45°, 90°, 135°, 180°)$, with a tolerance in each class of 22.5°. The image, as previously stated, does not show any clear preferential direction.

6.4 ESTIMATION OF THE VARIOGRAM

The estimation of the variogram is done in two phases:

- First an empirical estimator is calculated, for which, there are no *a priori* reasons for which regularity conditions should hold and the empirical variogram has to be *valid*.
- A fit of the empirical variogram is then performed, a theoretical model is chosen, whose parameters are optimal according to some criteria (e.g. LS, Weighted Least Squares (WLS), or, if a distributional model for the data is available, ML or Residual Maximum Likelihood (REML), ...). This been said, traditionally the model choice has been done by "eye fitting", mostly because *the difference between models is never very neat and the questions "how can we choose?" and "can data really distinguish between models?" have never had a real answer*, see Banerjee [2].

  This fit is typically parametric, so that the model has already been proven to be valid, meaning that the property of conditional negativity is known to hold.

Let us go through the principal estimators of the variogram:
The classical estimator of the variogram is the Method of Moments (MOM) estimator, used by Matheron at first, it is unbiased but has very poor statistical properties, for example it is sensitive to outliers, see [24] between the others.
It is calculated at every couple of points $(\vec{s}_i, \vec{s}_j)$ as:

$$2\hat{\gamma}(\vec{s}_i, \vec{s}_j) = \{Z(\vec{s}_i) - Z(\vec{s}_j)\}^2. \tag{6.11}$$

This estimator is sufficient if the data are on a grid, but since this is not usually the case and not for every distance a couple of points exist, other estimators have been proposed.

The *Variogram cloud* is obtained considering all the points at the same time; since all the points are plotted together it is difficult to interpret and is difficult to determine if the skewness is due to atypical information or is derived from the natural skewness in the data. Even though it has been suggested by some authorities (e.g. Cressie) it usually does not reveal any clear pattern, Baneree refers to the idea of fitting a curve to it as a "folly" since *we have a weak signal and a great deal of noise*. Delfiner suggests to use the cloud mostly to know how many couples exist per every distance, saying clearly that a distribution resembling a triangle is typically of a stationary variogram.
The *binned variogram* is an aleatory vector defined for a limited number of distance classes $\left(\hat{\gamma}(t_1), \ldots, \hat{\gamma}(t_N)\right)$ in each of those an average is calculated:

$$\hat{\gamma}(t_k) = \frac{1}{2\left|N(t_k)\right|} \sum_{(\vec{s}_i, \vec{s}_j) \in N(t_k)} [Y(\vec{s}_i) - Y(\vec{s}_j)]^2, \; k = 1, \ldots, K \tag{6.12}$$

where $N(t_k) = \{(i,j) \colon \left\|\vec{s}_i - \vec{s}_j\right\| \leq t_k, \geq t_{k-1}\}$ and $\left|N(t_k)\right|$ indicates its cardinality. The choice of the right number of bins and samples for each bin is clearly reminiscent of the choice of the number of classes in a histogram and is, as in that case, fundamental. A trade-off between *overfitting* and *underfitting* has to be found, since the result does not have to be too noisy, but needs to show all the peculiarities of the process. Two options are therefore envisioned:

- Semivariogram can be applied to the residuals after fitting a linear model,
- Semivariogram can be applied to the data itself. This is possible if theoretically a constant mean seems appropriate and therefore the hypothesis of intrinsic stationarity of the data, necessary for drawing a variogram is met.

A more robust estimator, called consequently *robustified variogram* is presented in Cressie [16].

6.4.1 Parameters for Empirical Variogram Estimation

As previously stated, the usual models are parametric, for this reason it is important to know which are the characteristics of a variogram. Numerous are the available models, each with different characteristics of stationarity and regularity: Linear, Spherical, Exponential, Power Exponential, Matèrn et cetera.
Once the class of the semivariogram has been chosen, parameters have to be fine-tuned according to the dataset.

NUGGET Typically indicated with τ^2 it is defined as $\lim_{h \to 0} \gamma(h)$, while, by definition $\gamma(0) = 0$, since the variogram is a pair function, null in the origin, it does not have to be continuous there, when this happens it is typically said that there is no *nugget effect*. The nugget can be extrapolated *a posteriori*, after the fit, it is fixed to zero only if there are strong theoretical reasons to do it. A discontinuity in the origin means that the process is not continuous in quadratic mean and therefore is considered to be highly irregular, due to microscale variations or measurement errors. If, on the other side, the variogram is linear in the origin (e.g. like in linear or spheric models without nugget effect), the process is continuous, but not differentiable, while differentiability occurs if the model is parabolic in the origin, and therefore very regular and unlikely to be seen in practice.

SILL If the random field is stationary and ergodic, the limit corresponds to the variance of the field $\lim_{h\to\infty}\gamma(\vec{h}) = Var(Z(\vec{s}))$, if it is finite the process is weakly stationary. This has brought to the idea of estimating the sill with the sample variance, but this has been questioned by numerous authors, between whom Barnes in [3].

PARTIAL SILL Indicated with σ^2. It is calculated as the difference between the sill and the nugget, as a result $\lim_{h\to\infty}\gamma(h) = \sigma^2 + \tau^2$.

RANGE If the range exists it indicates the distance from the origin at which the semivariogram levels and the asymptote is reached: $\gamma(R) = \tau^2 + \sigma^2$.

EFFECTIVE (OR PRACTICAL) RANGE This is used when the range is not finite. It identifies the radius of influence of the process, the distance until which samples are spatially correlated. It is conventionally defined as the distance when the semivariance first reaches 95% of the sill:

$$R = \min \quad \{h \mid \gamma(h) = 0.95(\sigma^2 + \tau^2)\}. \tag{6.13}$$

DECAY PARAMETER Inverse of the range parameter, indicated with ϕ.

6.4.2 Theorical Models for Semivariogram Fitting

Under the hypothesis of stationarity it is possible to divide the variograms in two major classes:

- Variograms that reach an asymptote, that exemplify a correlation that ceases at some maximum distance.
- Variograms without an asymptote, that exemplify a correlation that has no limit in distance, typically unrealistic.

Many models for variogram fitting are available, all underlying different hypotheses and different meanings. In the following definitions τ^2 is the nugget, R is the range, σ^2 the partial sill and h the lag distance.
In the following the indicator function $\mathbb{1}^{(h)}_{(a,b)}$ should be read as:

$$\mathbb{1}^{(h)}_{(a,b)} := \begin{cases} 0 & h \leq a, \text{or } h \geq b, \\ 1 & \text{otherwise.} \end{cases} . \tag{6.14}$$

LINEAR MODEL $\gamma(h) = \{\tau^2 + \sigma^2 h\}\mathbb{1}^{(h)}_{(0,\infty)}$. For $\sigma^2 = 0$ the result is pure nugget, white noise. A non-continuity point is admitted in the origin.

SPHERICAL MODEL $\gamma(h) = (\tau^2 + \sigma^2[\frac{3}{2R}h\ -\frac{1}{2}(\frac{h}{R})^3])\mathbb{1}^{(h)}_{(0,R)} + (\tau^2 + \sigma^2)\mathbb{1}^{(h)}_{[R,\infty)}$, spatial dependence levels at a certain distance.

EXPONENTIAL MODEL $\gamma(h) = \{\tau^2 + \sigma^2(1 - \exp\{-\frac{h}{R}\})\}\mathbb{1}^{(h)}_{(0,\infty)}$, spatial dependence decreases exponentially, therefore only a practical range is defined.

GAUSSIAN MODEL $\gamma(h) = \{\tau^2 + \sigma^2(1 - \exp\{-\frac{h^2}{R^2}\})\}\mathbb{1}^{(h)}_{(0,\infty)}$, a very regular model, difficult to see in reality.

The choice between these models depends greatly on the assumed hypotheses, mostly for what concerns the behaviour of the variogram in the origin. This is usually chosen as linear, a quadratic trend would mean assuming very strong hypotheses on the regularity and smoothness of the data.

6.5 KRIGING

The difference between Kriging and the aforementioned interpolation methods is that before the attention was mostly focusing on the interpolating functions, for which a model had been fixed, Kriging, instead, focuses on modeling the phenomenon starting from the data, their covariance structure and from this build a Best Linear Unbiased Predictor (BLUP), as linear combination of the data.
Prevision is a common problem in Geostatistics: given N observations $Z(\vec{s}_1), \dots, Z(\vec{s}_N)$ in a domain D how should the field be estimated at a site $\vec{s}_0$ where it has not been observed? What is the best predictor of $Z(\vec{s}_0)$ given $\vec{z} = [Z(\vec{s}_1), \dots, Z(\vec{s}_N)]^T$?

$$\hat{Z}(\vec{s}_0)|\vec{z} = m(\vec{s}_0) + \sum \lambda_i [Z(\vec{s}_i) - m(\vec{s}_i)] \tag{6.15}$$

where the optimal weights $\vec{\lambda}$ are so that

$$\vec{\lambda} = \text{argmin}_{\vec{\lambda} \in \mathbb{R}^N : \, \mathbb{E}[(\hat{Z}(\vec{s}_0)] = Z(\vec{s}_0)} Var \left[\sum \lambda_i Z(\vec{s}_i) - Z(\vec{s}_0) \right]. \tag{6.16}$$

This constrained optimization problem can be solved using Lagrange multipliers and the resulting $\vec{\lambda}$ is a function of $\gamma(h)$. With this approach there is no need for distributional hypotheses, but to have an estimate of $\gamma(h)$ it is necessary to assume intrinsic stationarity of the phenomenon. Unfortunately, not always this hypothesis holds true, e.g. if the mean cannot be considered to be constant. In that case the idea is to consider the signal as a sum of a drift (e.g.modelled wih first or second order polynomials) and a stationary residual $\delta(\vec{s})$ formed by two terms: a random autocorrelated component, that has to be captured with Kriging and the real error term, which should not be correlated. A more general theory has been suggested, that of Intrinsic Random Function of k^{th} order (IRF-k), this will not be further analyzed here.
We will go through the most important classes of Kriging, for a clear and easy introduction please refer to [13].

6.5.1 Simple Kriging

Simple Kriging assumes that the first moment is known and constant, .i.e., there is no trend: $\mu(\vec{s}) = m = \mathbb{E}[Z(\vec{s})]$, known and constant.
Once the model parameters have been estimated it is possible to evaluate the field in any point with a Best Linear Unbiased Estimator (BLUE) by solving a set of simultaneous equations. The unknown value of the field in the new point is proportional through some weights $\lambda_i, i = 1 \dots N$ to the value of the field in the neighboring measurement points $\vec{s}_i, i = 1 \dots N$. The unknown weights $\lambda_i, i = 1 \dots N$ are found after calculating the known vector whose term is $\gamma_{i,0} = \gamma(\vec{s}_i, \vec{s}_0) \; \forall i = 1, \dots, N$.
Unbiasdness is not imposed since it is granted by the fact that the mean of the field is

supposed to be known.
This brings to the following system of equations:

$$\begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1N} & \vdots \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{N1} & \cdots & \gamma_{NN} & \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \vdots \\ \gamma_{N0} \end{bmatrix}. \tag{6.17}$$

The calculation of the weights is quite demanding because it requires a matrix inversion, whose size depends on the number of sampled locations, for every point in which a prediction is needed, therefore building a prediction map is becomes complicated in an environment where battery life is a issue. On the other side Kriging allows to easily compute of a measure of reliability, namely:

$$s^2 = \sum_{i=1}^{N} \lambda_i \gamma_{i,0}. \tag{6.18}$$

This is very useful and gives a clear representation of the errors that have been committed. Maps of Kriging variance are a very useful tool and very used in practice.

6.5.2 Ordinary Kriging

Ordinary Kriging is very similar to Simple Kriging with the exception that in Ordinary Kriging the first moment is assumed to be an unknown constant $\mu(\vec{s}) = m$. Ordinary Kriging considers a model with no covariates (with the exception of the location ones, it is basically a pure spatial model). Once the model parameters have been estimated it is possible to evaluate the field in any point with a BLUE by solving a set of simultaneous equations very similar to the one presented in the case of Simple Kriging, with the only difference that in this case a Lagrange multiplier μ is added to grant the unbiasness of the BLUE estimator. This can be seen in the following system:

$$\begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1N} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{N1} & \cdots & \gamma_{NN} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \vdots \\ \gamma_{N0} \\ 1 \end{bmatrix} \tag{6.19}$$

where the last line corresponds to a normalization of the weights:

$$\sum_{i=1}^{N} \lambda_i = 1. \tag{6.20}$$

It is still possible to calculate a Kriging variance punctually as an index of the local reliability of the estimate:

$$s^2 = \sum_{i=1}^{N} \lambda_i \gamma_{i,0} + \mu. \tag{6.21}$$

6.5.3 Universal Kriging

Universal Kriging is used to model non stationary phenomena, or to account for a lack of homogeneity in the data, see e.g. [16].

The drift functional form has to be modeled and sometimes it is the mere result of a trade-off between model fit and parsimony, as well stated in [43], consequently typically polynomials models (e.g. first or second order) are chosen.
In formulas the trend is modeled as a weighted sum of known functions f_l:

$$\mathbb{E}[Z(\vec{s})] = m(\vec{s}) = \sum_{l=0}^{L} \beta_l f_l(\vec{s}) \tag{6.22}$$

unless the physics of the phenomenon indicates the shape of the function f_l, their expression is arbitrary, most often polynomials are retained almost only for the computational complexity and analytical convenience, with the convention that $f_0(\vec{s}) = 0, \forall \vec{s}$, β_l are the relative coefficients. It has to be noted that the choice of the functions $f_l(\vec{s})$ determines completely the value of the estimate beyond the data range, where no data exist, this makes extrapolating a hazardous endeavor. In this case Kriging is performed on the residuals after the trend has been removed.
The unknown value of the field in the new point is proportional through some weights $\lambda_i, i = 1 \dots N$ to the value of the field in the neighboring measurement points $\vec{s}_i, i = 1 \dots N$, in this case the Lagrange multiplier μ is added to grant the unbiasness of the BLUE estimator.
The estimation consists in solving the following system of equations for every $\vec{s}_0$ in which the author is interested.

$$\begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1N} & f_1(\vec{s}_1) & \cdots & f_L(\vec{s}_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{N1} & \cdots & \gamma_{NN} & f_1(\vec{s}_N) & \cdots & f_L(\vec{s}_N) \\ 1 & \dots & 1 & 0 & \ddots & 0 \\ f_1(\vec{s}_1) & \dots & f_1(\vec{s}_N) & 0 & \ddots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_L(\vec{s}_1) & \dots & f_L(\vec{s}_N) & 0 & \ddots & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ \mu_0 \\ \vdots \\ \mu_L \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \vdots \\ \gamma_{N0} \\ 1 \\ f_1(\vec{s}_0) \\ \vdots \\ f_L(\vec{s}_0) \end{bmatrix} \tag{6.23}$$

Where the last lines correspond to a normalization of the weights:

$$\sum_{i=1}^{N} \lambda_i f_l(\vec{s}_i) = f_l(\vec{s}_0) \quad \forall l = 1 \dots L. \tag{6.24}$$

The calculation of the weights is quite demanding, but on the other side it allows to easily compute of a measure of reliability:

$$s^2 = \sum_{i=1}^{N} \lambda_i \gamma_{i,0} + \sum_{l=0}^{L} \mu_l f_l(\vec{s}_0), \quad f_0(\vec{s}) = 0 \ \forall \vec{s} \in D. \tag{6.25}$$

It is easy to see that Ordinary Kriging is a very special case of Kriging, with $\vec{\mu} = \vec{\beta}$ and the matrix being simply an identity matrix.
The prediction problem therefore becomes

$$\min_{f(\vec{z})} \mathbb{E}[(Z(\vec{s}_0) - f(\vec{s}))^2 | \vec{z}]. \tag{6.26}$$

After some calculation the optimal predictor, that minimizes the error in the conditional expectation of $Z(\vec{s}_0)$, is the posterior mean of $Z(\vec{s}_0)$, that is known to be the minimizer of the squared error loss function.

Suppose that spatial data in different locations $\vec{s}_1, \ldots, \vec{s}_N$ are modeled as if they were generated by the random process

$$Z(\vec{s}) = \sum_{i=1}^{L} \{\beta_l x_l(\vec{s})\} + \delta(\vec{s}) \quad \vec{s} \in D \subset \mathbb{R}^d \tag{6.27}$$

where $x_l, l = 1 \ldots L$ is a collection of non random explanatory variables that may or may not depend on spatial location and can be collected in the matrix X, while δ is a zero-mean, finite variance error process that may or may not be spatially correlated.
The model can be written in matrix notation as:

$$\vec{Z} = X\vec{\beta} + \vec{\delta} \tag{6.28}$$

where $\vec{Z} = (Z(\vec{s}_1), \ldots Z(\vec{s}_N))^T$, $\vec{\beta} = (\beta_1, \ldots \beta_L)^T$, parameters that account for large scale variations and $\vec{\delta} = (\delta(\vec{s}_1), \ldots \delta(\vec{s}_N))^T$.
If $\Sigma = \text{var}(\vec{\delta})$ was known the BLUE for $\vec{\beta}$ would be the Generalized Least Squares (GLS) estimator:

$$\hat{\vec{\beta}}_{BLUE} = \hat{\vec{\beta}}_{GLS} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\vec{Z} \tag{6.29}$$

since it minimizes

$$(\vec{Z} - X\vec{\beta})^T\Sigma^{-1}(\vec{Z} - X\vec{\beta}) \tag{6.30}$$

$$\mathbb{E}[Z(\vec{s}_0)|\vec{z}]. \tag{6.31}$$

In the wildly unrealistic hypothesis that the population parameters $\left(\beta, \sigma^2, \phi, \tau^2\right)$ were known the estimator would be:

$$\begin{aligned} \mathbb{E}[Z(\vec{s}_0)|\vec{z}] &= \vec{x}_0^T\vec{\beta} + \vec{\gamma}^T\Sigma^{-1}(\vec{z} - X\vec{\beta}) \\ Var[Z(\vec{s}_0)|\vec{z}] &= \sigma^2 + \tau^2 - \gamma^T\Sigma^{-1}\vec{\gamma}. \end{aligned} \tag{6.32}$$

Unfortunately parameters need to be estimated from the data, therefore

$$\hat{f}(\vec{z}) = \vec{x}_0^T\hat{\vec{\beta}} + \hat{\vec{\gamma}}^T\hat{\Sigma}^{-1}(\vec{y} - X\hat{\vec{\beta}}) \tag{6.33}$$

where $\hat{\gamma}$ is an estimate of the variogram, $\hat{\vec{\beta}} = (X^T\hat{\Sigma}^{-1}X)^{-1}X^T\hat{\Sigma}^{-1}\vec{z}$ and $\hat{\Sigma}$ is the sample variance. We can consider $\vec{x}_0 = x(\vec{s}_0)$ to be always observable, because in our case, being the transmitter distance it can always be calculated, if it is not the case, an estimate for it will have to be computed as well.
Predicting not in a new location, but in an already sampled one, would give the sampled value if there is no nugget effect, but this will not be true in general. If we want to predict the field not in very few locations, but over a finite grid of sites, a *prediction surface* has to be built, it will be smoother than the interpolation surface.
In reality Σ is unknown, consequently a procedure is needed for the joint estimation of $\vec{\beta}$ and Σ. A possibility is to use a Ordinary Least Squares (OLS) estimator:

$$\hat{\vec{\beta}}_{OLS} = (X^TX)^{-1}(X^T\vec{Z}) \tag{6.34}$$

which is highly inefficient when Σ is not diagonal, that is a typical situation in a spatial context. Cressie in [16] suggests the use of Extended Generalized Least Squares (EGLS), when a parametric estimation of Σ is at hand.
A surface plot of the standard error is useful, standard errors are supposed to be higher where there are less data. Once the model parameters have been estimated it is possible to evaluate the field in any point with a BLUE by solving a set of simultaneous equations.

6.5.4 Other Kriging Models

Indicator Kriging is used when data are binary, e.g. a point is a forest or not, it assumes $\mu(\vec{s}) = m(\vec{s})$, unknown and constant.
Disjunctive Kriging is a nonlinear generalization of Kriging, *lognormal Kriging* interpolates positive data by means of logarithms, *cokriging* is performed when other variables, correlated with the primary one are available and are believed to improve the prediction of the primary variable.

6.5.5 Crossvalidation in Kriging

Choosing the best model for interpolation is not easy, one classic method is computing Residual Sum of Squares (RSS) via 1-fold cross-validation: for every point in the dataset and for every interpolation model considered:

- Remove 1 points from the data set.
- Use the remaining points to estimate the value of the removed point.
- Finally, per every interpolation model calculate:

$$\text{RSS} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(z_i - \hat{z}_i)^2}. \tag{6.35}$$

At the end choose the model that minimizes the RSS (Residual Sum of Squares), or the one that offers the better trade-off between complexity and performance.
Variations of this method include choosing more than one point at once, n-fold crossvalidation or, if the number of collected data is hight, dividing the dataset in two groups, using one for identification and the other for validation of the model. See [5] for details.

6.5.6 Kriging in Cognitive Radios

Kriging has already been used in a CR context to obtain a *Spectrum Cartography*, a local spectrum map with the exchange of information between terminals, to locally optimize the communications, bypassing or cooperating with the Base Station (BS).

In [59] Kriging has been used to model the average Power Spectral Density (PSD) from simulated Universal Mobile Telecommunications System (UMTS) downlink data, that present numerous BSs, whose position can be seen in the krigged estimation. To fit the semivariogram a Matèrn model has been used.
In [51] Kriging is used as a geostatistical approach to minimize the errors due to limited geographical information in available propagation model, whose resolution is typically around hundreds of square meters, insufficient from a CR point of view. According to the authors field measurements are resource consuming and systematic sampling does not capture all the characteristics of the field. For these reasons they make use of Kriging not just for interpolating and predicting in an optic of transmission, but, instead, as a way to optimize the location where to sample, so to minimize the Universal Kriging variance, which is to be considered a more significant index than the simple minimum distance between sampling locations, that would lead to a homogeneous sampling in space, while, instead more samples are needed where the gradient of the field is bigger. Using Universal

Kriging Variance as an index is the same as using different metric, that takes care of the field covariance matrix.
In [72] the main aim of using Kriging is producing spatial spectrum models useful to generate simulation scenarios for evaluation of DSA protocols. The authors also state that it could be useful to create runtime models of occupancy to be used where few or none data are available and this is the use we mostly look forward to meet in this thesis work.
In [31] and [9] Ordinary Kriging has been chosen as prediction technique, even though assumptions of constant mean are not fulfilled, and it is found to outperform classical interpolation techniques. Classical interpolation methods have the heavy drawback of not granting the possibility to interpolate points outside the convex hull, while, paying particular attention, with Kriging it is possible to extrapolate. In the same paper two different implementations have been carried out:

- Centralized Kriging (CK), where all the nodes communicate the data to a central unit,
- Distributed Kriging (DK), where the domain is divided in a fixed number of clusters (of fixed size that depends on sensor capability to exchange information) or adapting to improve field estimation.

In each of these implementations an estimation of the variogram and a Kriging prediction are performed.
DK option is to be preferred to ensure robustness to node fails and scalability to network size and it is ideal if the user is only interested in local prediction, it verifies the hypotheses of constant mean better because the only variation is in shadowing which has zero mean. On the other side a distributed implementation is based on a reduced number of samples, this bring to loss of information both about spatial dependencies at long distance and about some classes while constructing the empirical variogram. The last problem can be tackled with an additional regression to be performed in each node.

6.6 RESULTS

We proceed with the prediction modelling that have been introduced, measurements of the RSSI from channel 44 in MC3 will be our reference, because of its clear pattern in space, see Figure 35.

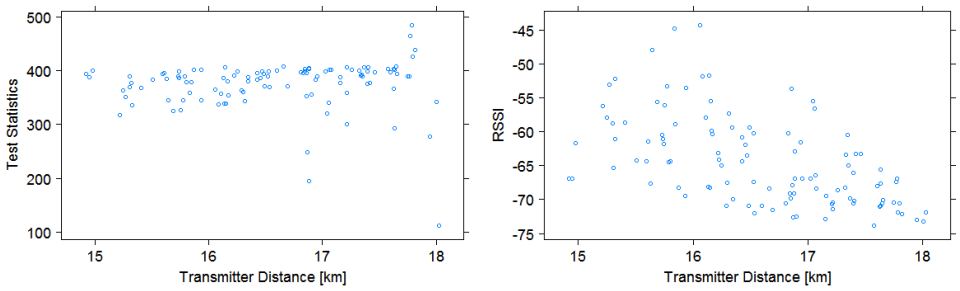

Figure 35: Maps of the Test Statistics (left) and RSSI (right) versus the distance in kilometers from the transmitter. In both images the points indicate the temporal median in a specific location and refer to channel 44. While the RSSI map clearly indicates the existence of a trend, this is not true for the Test Statistics. This could be a good reason for assuming a constant mean in space for the Test Statistics but not for the RSSI.

At first we have to check that the process is really spatially correlated, to see this, we compare the actual sample variogram, still non-fitted, and therefore non regular, with a random reassembling of the measurements to spatial locations in the dataset, see Figure 36.

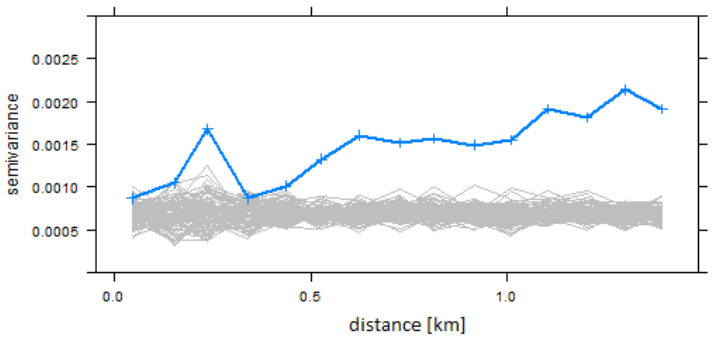

Figure 36: Comparison between the experimental variogram (blue line) and other variograms, calculated randomly reassigning the RSSI field to the sampling locations (grey lines).

As it can be seen in Figure 36, the hypothesis of absence of spatial correlation seems unlikely, since the blue line is quite apart from the sheaf of grey lines, basically stating that that the particular configuration of couples (location, RSSI) that is verified in reality is apart from all the other possible reassemblings. Theoretically, this makes perfect sense, being the signal all from the same transmitter, it is normal to assume a spatial correlation between the samples. We can therefore proceed with Kriging.

6.6.1 Variogram Fitting

As it has been said in the theoretical part of this work the RSSI should follow the path loss in average while fluctuations are mostly due to shadowing.
The extent of the *radius of spatial coherence*, that is to say, the distance until which observations appear to be correlated is clearly visible looking at empirical semivariograms, see

e.g. Figure 37. It typically is between 500 meters and 1 kilometer. This distance range is typical of the DVB-T signal in the far field, as can be seen in [52], and this is our case since a channel is considered to be busy if it is occupied by a signal that comes from 15 to 18 kilometers far away, see Figure 35.

6.6.2 Simple Kriging

As it has been said in the theoretical part simple Kriging is the easiest form of Kriging: the mean of the field is considered to be constant and known, in the implementation we assume it to coincide with the sample mean of the field.
The variogram of the absolute value of the logarithm of the RSSI has been approximated with a Spherical variogram, which is a quite regular one and linear in the origin, with a low nugget. The fit can be visualized in Figure 37. The choice of the family is due to multiple reasons: we needed a variogram with a low number of parameters, so that it could be fit even with a small number of samples, that presented a finite range, because of the physics of the phenomenon.

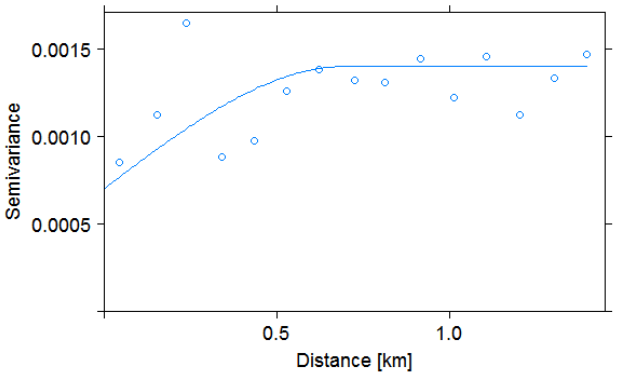

Figure 37: Result of the experimental variogram fit (blue circles) with the theoretical, spherical one (blue line).

This fit allows to produce the Kriging estimation visible in Figure 38.

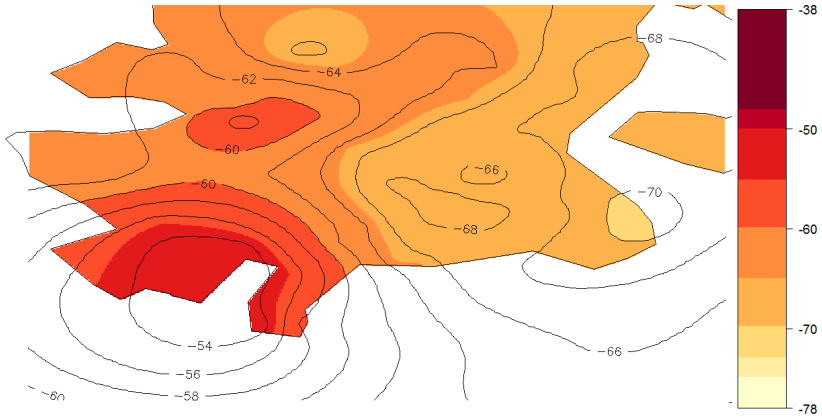

Figure 38: Map of prediction of the RSSI in occupied channel 44, using Simple Kriging.

To check if this method is already improving the estimation, we operate a 1-fold cross-validation, whose result can be seen in Figure 39.

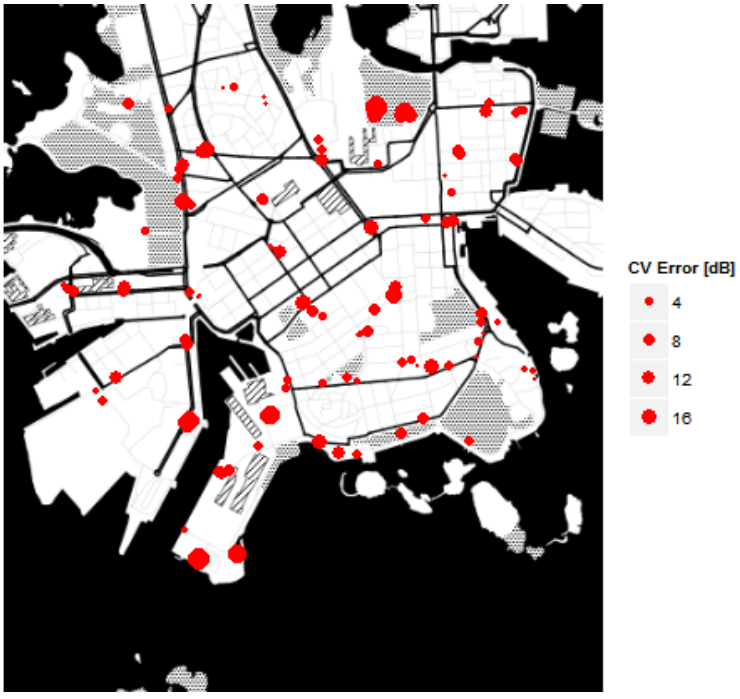

Figure 39: 1-fold crossvalidation using Simple Kriging. The bubbles indicate the difference in absolute value between the local prediction in a sampling location if the dataset had not contained that measurement. It is possible to visualize that most of the errors are concentrated on the borders of the sampling area, where fewer samples are available.

The total squared error is of approximately $5,64$ dB, that compared with a sample mean of approximately -64.38 dB implies a relative error, approximately of 8.6%. The accuracy

level is definitely sufficient for this type of application, because, luckily, as we have seen difference in RSSI between unoccupied and occupied channel is way above this quantity.

6.6.3 Ordinary Kriging

Ordinary Kriging has the same theoretical hypotheses than simple one, except that an equation is added to the system, to insure unbiaseness, because the real field mean is supposed to be unknown.
The map of the krigged surface is visible in Figure 40.

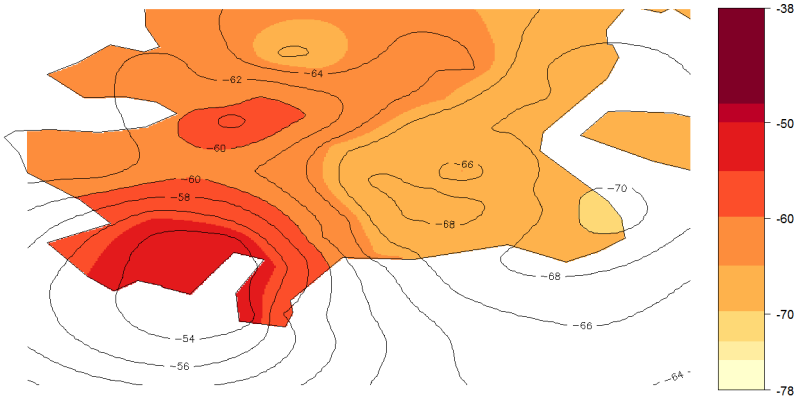

Figure 40: Map of prediction of the RSSI in occupied channel 44, using Ordinary Kriging.

It results very alike the simple implementation, as the crossvalidation error will show, basically this means that the mean of the field can be considered to be constant and equal to the sample mean.
To check if this method is improving the estimation, we operate a 1-fold crossvalidation, whose result can be seen in Figure 41.

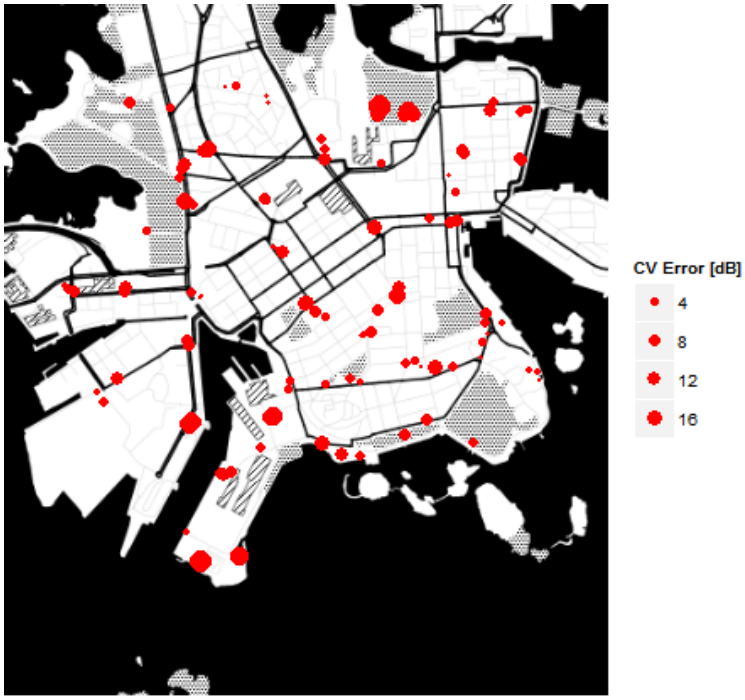

Figure 41: 1-fold crossvalidation using Ordinary Kriging. It has to be noted that the error is larger at the boundaries of the sampled area, wherre less samples are available, see Figure 8.

The total squared error is of approximately $5,54$ dB, that compared with a sample mean of approximately -64.38 dB implies a relative error approximately of 8.6%. The situation is very similar to the previous case, but in this case no false hypotheses have been assumed, since we are not stating that the real mean of the field is known.

6.6.4 Universal Kriging

It has been shown that the mean of the RSSI field cannot be considered to be constant in space, at least from a first graphical analysis, see Figure 35. For this reason we opt for a universal Kriging, hoping that the use of geographical coordinates and the distance from the transmitter as a covariate will help the prediction.
We opt for adding the covariates one at a time, to see if there are any improvements in the prediction, as in the previous cases we consider crossvalidation as a metric to evaluate the performance.

Universal Kriging, only coordinates

At first we use location coordinates as the only covariate, the results are not much better in term of crossvalidation error compared to what was obtained with ordinary Kriging. The variogram fit can be visualized in Figure 42, the consequent estimation map in Figure 43.

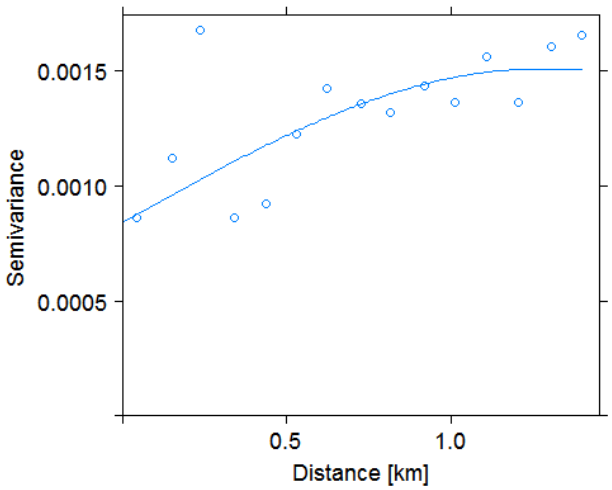

Figure 42: Result of the experimental variogram fit (blue circles) with the theoretical, spherical one (blue line).

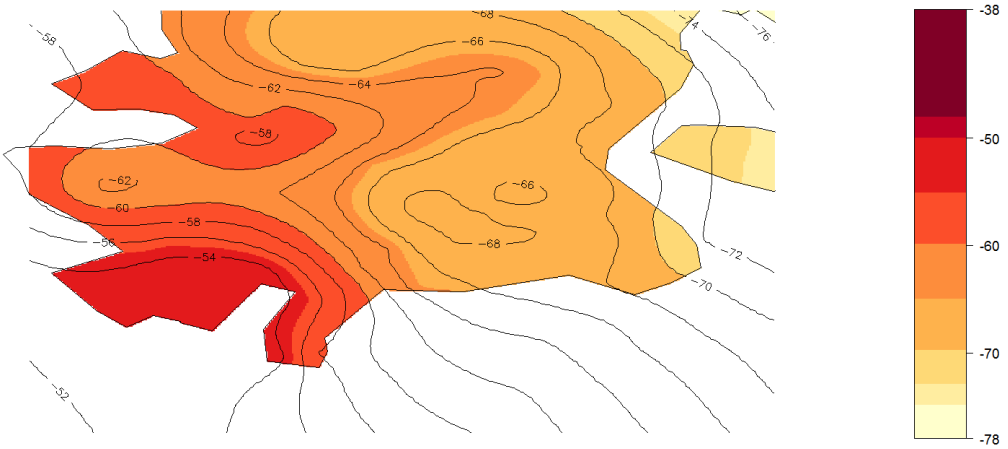

Figure 43: Map of prediction of the RSSI field in occupied channel 44, using Universal Kriging, with only location as covariates.

To check if this method is already improving the estimation, we operate a 1-fold cross-validation, whose result can be seen in Figure 44.

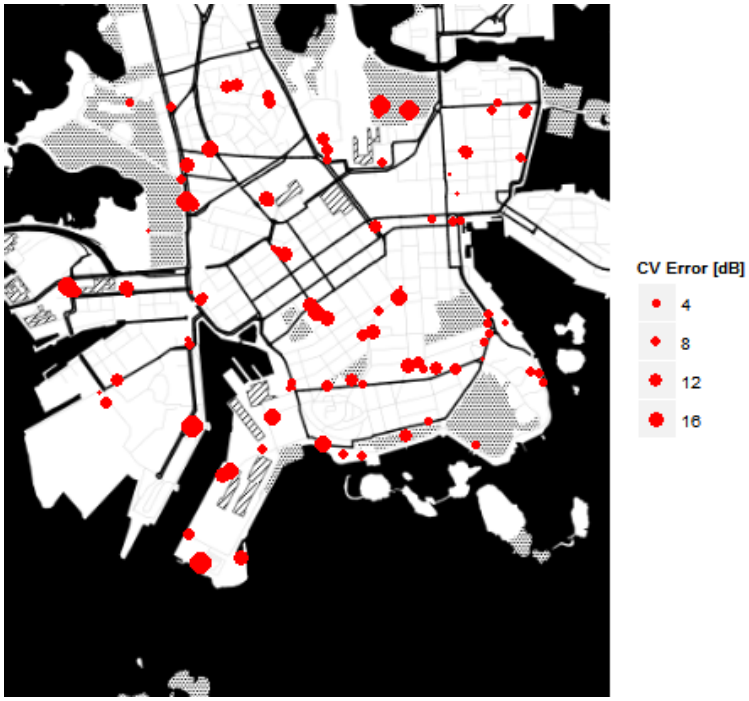

Figure 44: 1-fold crossvalidation using Universal Kriging, with only coordinates as covariates.

The solution is not very different from the preceding ones, with the difference in the South - Eastern area. This is the reason why the Kriging maps have not been put on a map, otherwise the prediction out of the window, that is basically pure extrapolation would have been cut.
In [36] Journel and Rossi show why the prediction is not strongly influenced by the existence of a trend in the model, they demonstrate that trend matters only in extrapolation situations and simpler Ordinary Kriging is the preferred option. They advocate that the choice of a particular trend model is most often arbitrary and for this reason it does not affect the final estimation of the total variability. They show that if the location where the prediction is desired is surrounded by samples in all the directions at a shorter distance than the range of the fitted variogram the estimates are virtually identical, however, at locations outside the sampling window this does not hold true and the estimations could be sensibly different and this is clearly visible from our Kriging map predictions.
Quantitatively the relative error is still approximately 8% of the sampled field if we make an average of the estimations through crossvalidation.

Universal Kriging, considering the distance from the transmitter

We use the distance from the transmitter as a covariate, alternative to the couple of longitude and latitude, basically changing the "origin of the geographical plane", to see if this impacts anyhow the estimation, this of course should not be the case. This option would be chosen only if a very notable benefit would emerge from the crossvalidation results, in terms of standard error, because calculating all the distances could be a non negligible computational cost if not only a very local prediction is looked for, but a prediction map, as the one in Figure 46 is desired and we hope this will not be the case otherwise the moel

would not be robust to a change of coordinates.
The variogram fit can be visualized in Figure 45 and the prediction map in Figure 46.

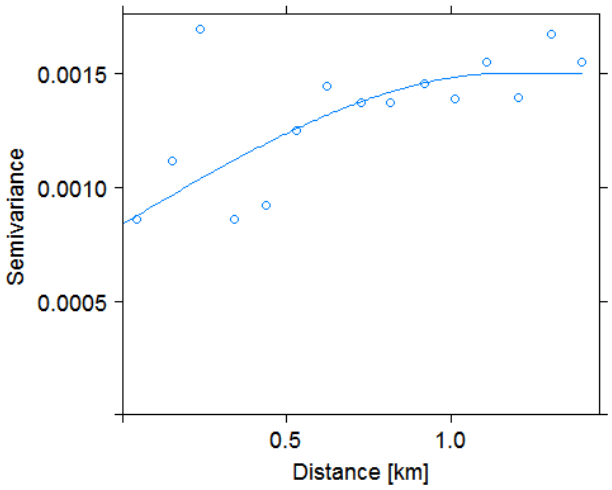

Figure 45: Result of the experimental variogram fit (blue circles) with the theoretical, spherical one (blue line).

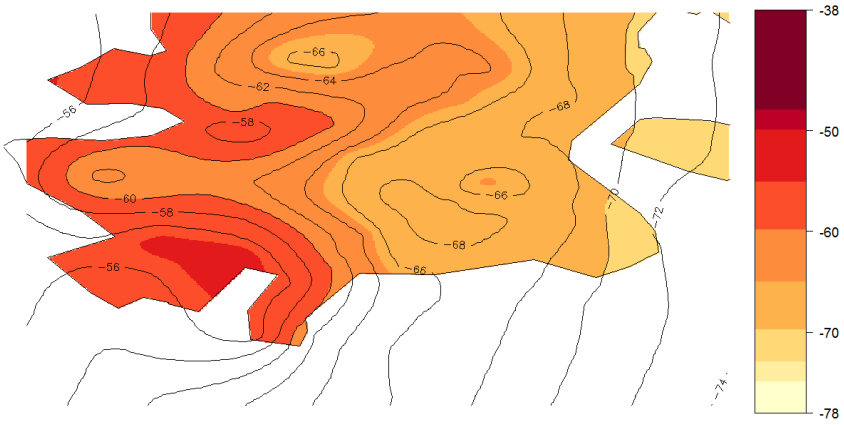

Figure 46: Map of prediction of the RSSI field in occupied channel 44, using Universal Kriging, with only distance from the transmitter as covariate.

To check if this method is already improving the estimation, we operate a 1-fold cross-validation, whose result can be seen in Figure 47.

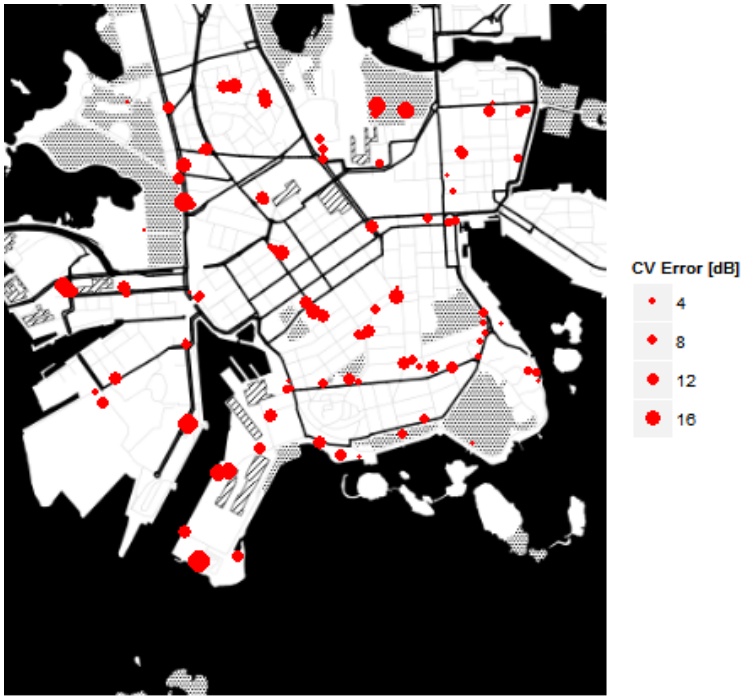

Figure 47: 1-fold crossvalidation using Universal Kriging with only distance from the transmitter distance as covariate.

6.6.5 Comparisons between Kriging Techniques

Since the performance of the two Universal Kriging is comparable the option is to choose the one with the coordinates as predictors, since it is much less requiring computationally, because no distances from the transmitter need to be calculated. Let us focus on the comparison between the various techniques, to do so and to verify Journel and Rossi statement, we plot the boxplots to compare the crossvalidation error, so to choose one of the techniques, see Figure 48.

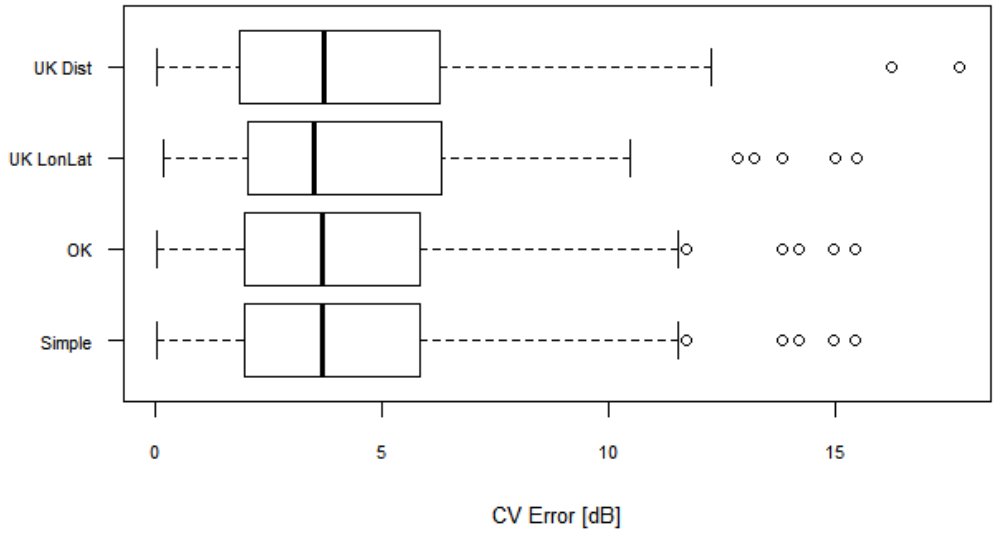

Figure 48: Boxplots to compare the 1-fold crossvalidation errors using different Kriging techniques, with or without the existence of a trend. From top to bottom Universal Kriging with distance as a caovariate, Universal Kriging with geographical coordinates, Ordinary Kriging and Simple Kriging. As in the usual boxplots, the thicker lines indicates the median of the distribution, that results to be approximately 4 dB, a very good result for our objectives.

Looking at Figure 48 our option is to choose Ordinary Kriging, proved that no extrapolation is needed, otherwise the more complex Universal Kriging with the coordinates as covariates will be the way to go.

6.6.6 Conclusions

Kriging is theoretically more complex than the preceding options and way more computationally expensive, but on the other side, it offers notorious and visible improvement in the results. Question is if such an expensive tool (in computational terms) could be an option in this type of implementation or if worse performing, but easier to implement algorithms are sufficient. In scenarios, such as in DVB-T, where the field is typically almost constant in occupation for long periods of time, we think the use Kriging is justified, because the computation can be effectuated at long intervals in time one from the other, for the low risk of interference.

A major benefit of Kriging in addition to better accuracy and ability to extrapolate is that it also provides variance estimates for the interpolated values that indicate the level of confidence in the prediction. These can be extremely useful to guide the sensing in the CR network: SUs in areas where there is a large interpolation variance may be asked to sense the spectrum more often than those where the confidence of a correct prediction is high. Similarly the variance estimates could be also used in scheduling the SUs when they want to access the spectrum. In a highly dynamic radio environment Kriging might be

an overkill, since the field would need to be updated all the time. In those scenarios the non-statistical methods may be more feasible. Also in a highly dense SU network the non-statistical methods are probably giving a low enough relative error in the interpolation with much lower computational cost than Kriging, however, without any insight into the interpolation variance, therefore the only way to check the correctnes of the prediction is to look at crossvalidation errors.
Another important factor that has not been taken into account yet is the fact that we still do not know which path CR technology will take, whether SUs will be able to share information with their closest neighbors or whether a FCwill collect every SUs data. It is clear that the information sharing will definitely affect the choice of the best technique to use.

7 BAYESIAN ESTIMATION

In this section we provide a Bayesian estimation and a prediction of the RSSI measurements of Channel 44 from MC3 in the Helsinki area, in two main ways:

- An empirical Bayes approach is followed: a linear model is fit to measurements of channel 44 in MC3 data. A variogram is subsequently built on the residual of the model. The parameters of this (nugget, partial sill, range) are subsequently used to compute the prior distribution of the Bayesian method.
- The second model uses a "truly Bayesian" prior, whose parameters are initialized by a physical model addressed in the ITU-R (International Telecommunication Union) document "Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 3 000 MHz". A simulation from this model has been built in Matlab and the output has been analyzed in the same fashion as in the empirical Bayes approach.

Before adventuring in the estimation part, a remark of the spatial process theory is necessary, we limit to a uni-dimensional response function, as in the data.

7.1 GAUSSIAN PROCESSES

When we write $Z(\vec{s} : \vec{s} \in D)$ we are envisioning a spatial process indexed by $\vec{s}$. The process of interest $Z(\vec{s})$ can be modeled as a Gaussian process:

$$\vec{Z}|(\vec{\mu}, \vec{\theta}) = \mathcal{N}(\vec{\mu}, \Sigma(\vec{\theta})). \tag{7.1}$$

Where $\vec{\theta} = (\tau^2, \sigma^2, R)$, respectively are the already introduced nugget, sill and range. The sampled values of Z correspond to $Z(\vec{s}_i),\ i = 1 \dots N$. As a matter of fact, we do this through specification of arbitrary finite dimensional distributions, i.e. for an arbitrary number and choice of locations. Consistency of such specifications in terms of ensuring a unique joint distribution will rarely hold and will be difficult to establish. Therefore we avoid such technical concerns by confining ourselves to Gaussian processes or the mixtures of such processes, so that the only required theoretical element is a valid correlation function. In practice to clarify the inference setting we will only observe $Z(\vec{s})$ at a finite set of locations: $\vec{s_1}, \vec{s}_2, \dots, \vec{s}_N$. Based upon $Z(\vec{s}_i), i = 1, , \dots, N$ we seek to infer mean, variability and association structure of the process and predict $Z(\vec{s})$ at arbitrary unobserved locations.
Restriction to Gaussian processes leads to several advantages:

- Very convenient distribution theory: joint marginal and conditional distributions are all immediately obtained from standard theory once the mean and covariance structure have been specified.
- A Gaussian process assumption for spatial random effects introduced in the second stage of the hierarchical Bayesian model is very natural.

- Strong stationarity is equivalent to weak stationarity.
- It is difficult to criticize Gaussian assumption: in the absence of replication $\vec{Z} = (Z(\vec{s}_1), Z(\vec{s}_2), \ldots, Z(\vec{s}_N))$ is a single realization of an N-dimensional distribution, with a sample size of one, *no* multivariate distributional specifications (such as the Gaussian one) can be criticized.

Even in the small class of isotropic processes very numerous famous models do exist. Some examples are the Exponential, the Power Exponential, the Gaussian, the Matérn model, etcetera. The choice between those should depend on the expected regularity of the signal (is it *strictly stationary*, *weakly stationary*, *intrinsicly stationary* or only *ergodic*?).
In this work we opt for an easy scheme for the theoretical variogram, where easy stands for a variogram model with a low number of parameters that needs to be fitted, for computational reasons, for the low number of available data and for consistency with the previous frequentist analysis.

7.2 BAYESIAN MODELS

7.2.1 First Bayesian Model

The result of the the assumption of the process to be Gaussian is:

$$\begin{aligned} \vec{Z}(\vec{s}) &= \vec{\mu}(\vec{s}) + \vec{\epsilon}(\vec{s}) = X(\vec{s})\vec{\beta} + \vec{\epsilon}(\vec{s}) \\ \vec{\epsilon}(\vec{s}) &\sim \mathcal{N}(\vec{0}, \Sigma) \\ \Sigma &= \sigma^2 H(R) + (\tau^2 I) \\ H(R)_{i,j} &= \rho(R; d_{i,j}). \end{aligned} \tag{7.2}$$

Where the nugget effect $\tau^2 I$ is not mandatory, but has the positive results of making the model more stable, see Banerjee in [2] for a reference. Moreover it allows the residuals to be not spatially continuous ($\epsilon(\vec{s} + \vec{h}) - \epsilon(\vec{s}) \nrightarrow 0$ as $\vec{h} \to 0$), not because the spatial process is not smooth, but because some additional variability, associated with $Z(\vec{s})$, is envisioned. This can be viewed as measurement noise or replication noise due to repetition of measurements at the same location $\vec{s}$ but also to a *microscale variability*, at distances smaller than the smallest interlocation distance between samples ($d_{i,j} = d(\vec{s}_i, \vec{s}_j)$).
A stationary model is yield while specifying a correlation function dependent only on the separation between samples.
As in the frequentist part in every location we scorporate the signal as:

$$Z(\vec{s}) = \mu(\vec{s}) + \epsilon(\vec{s}) = \mu(\vec{s}) + w(\vec{s}) + n(\vec{s}) \tag{7.3}$$

where the mean structure is the result of a linear model

$$\mu(\vec{s}) = X(\vec{s})\vec{\beta}. \tag{7.4}$$

The residual is partitioned into two pieces: one, $w(\vec{s})$, assumed to be a realization of a zero centered, stationary, Gaussian, spatial process, capturing residual spatial association, that has to be taken into account while modeling the partial sill and the range parameters and $n(\vec{s})$, n the other side, the uncorrelated pure error term, a non-spatial residual, used to

model the nugget effect.
Following this thought, $n(\vec{s})$ could be seen as a spatial process as well, with very rapid decay and a very small range. Cressie suggests to partition $n(\vec{s})$ itself in two pieces, one for taking care of spatial variability, the other for pure error; Banerjee states that very little is typically known about microscale variation, so that, even though been perfectly meaningful in theory it becomes difficult to put this idea into practice, in this work Banerjee's option has been selected.
Since the basic Gaussian isotropic models are a special case of the general linear model, the problem boils down to the definition of the $\Sigma(\theta)$, an option is to choose

$$\Sigma = \sigma^2 H(R) + \tau^2 I \tag{7.5}$$

where H is the correlation matrix

$$H_{i,j} = \rho(\vec{s}_i - \vec{s}_j) = \rho(R, d_{i,j}) \tag{7.6}$$

where ρ is a valid correlation function indexed by R.
The Bayesian Model can then be stated as:

$$\begin{aligned} Z|\vec{\theta} &\sim \mathcal{N}(X\vec{\beta}, \sigma^2 H(R) + \tau^2 I) \\ \pi(\vec{\theta}) &= \pi(\vec{\beta})\pi(\sigma^2)\pi(\tau^2)\pi(R) \\ \pi(\vec{\beta}) &\sim \text{Normal} \\ \pi(\sigma^2) &\sim \text{Inverse Gamma} \\ \pi(\tau^2) &\sim \text{Inverse Gamma} \end{aligned} \tag{7.7}$$

where $\pi(\cdot)$ denotes the prior distribution. $\pi(R)$ depends on the choice of the correlation function.
Unfortunately proper but vague priors lead to improper posteriors or to proper ones, that could be computationally indistinguishable.
According to Banerjee independent priors are typically chosen for the different parameters, so that the joint prior distribution appropriate prior distribution $\pi(\vec{\theta})$ may be simply expressed as the product of the single priors: $\pi(\vec{\theta}) = \pi(\vec{\beta})\pi(\sigma^2)\pi(\tau^2)\pi(R)$. Typical choices have already been listed.

Since inference *a posteriori* will be on each parameter independently from the others, one of the objectives will be marginal posterior distributions, that will not typically be in closed form, so in general we can only say:

$$\begin{aligned} \pi(\vec{\beta}|\vec{z}) &= \int\int\int \pi(\vec{\beta}, \sigma^2, \tau^2, R|\vec{z})\, d\sigma^2\, d\tau^2\, dR \\ &\propto \pi(\vec{\beta}) \int\int\int f(\vec{z}|\vec{\theta})\pi(\sigma^2)\pi(\tau^2)\pi(R)\, d\sigma^2\, d\tau^2\, dR. \end{aligned} \tag{7.8}$$

7.2.2 Hierarchical Model

It is possible to build a hierarchical model augmenting the parameter space from $\vec{\theta}$ to $(\vec{\theta}, \vec{W})$, where $\vec{W} = \left(w(\vec{s}_1, \ldots, \vec{s}_N)^T\right)$ *de facto* increasing the parameter's space dimension by N values.

$$\begin{aligned} Z|(\vec{\theta}, \vec{W}) &\sim \mathcal{N}(X\vec{\beta} + \vec{W}, \tau^2 I) \\ \vec{W}|(\sigma^2, R) &\sim \mathcal{N}(\vec{0}, \sigma^2 H(R)) \\ \pi(\vec{\theta}) &= \pi(\vec{\beta})\pi(\sigma^2)\pi(\tau^2)\pi(R) \\ \pi(\vec{\beta}) &\sim \text{Normal} \\ \pi(\sigma^2) &\sim \text{Inverse Gamma} \\ \pi(\tau^2) &\sim \text{Inverse Gamma} \\ \pi(R) &\sim \text{Uniform.} \end{aligned} \tag{7.9}$$

Fitting the model as $f(\vec{z}|\vec{\theta})\pi(\vec{\theta})$ or as $f(\vec{z}|\vec{\theta}, \vec{W})\pi(\vec{W}|\vec{\theta})\pi(\theta)$ holds the same posterior $\pi(\vec{\theta}|\vec{z})$. Of course it makes sense to do as much marginalization as possible before implementing an MCMC algorithm, that can consequently be in the smallest possible dimension and therefore faster to explore. Notwithstanding this Banerjee affirms that determinant and inversion calculation will be better behaved in the marginal, than in the conditional model form, in this optic, the nugget has also the effect of facilitating the inversion process since $\sigma^2 H(R) + \tau^2 I$ is further from been singular than $\sigma^2 H(R)$ when sample locations draw closer.

7.3 EMPIRICAL BAYES

We opt for treating the data in a pure spatial model, forgetting about the time dimension. As it had already been shown in Figure 35 the RSSI is a function of the distance from the transmitter, even though the pattern is less clear than expected. This attempt is here called *empirical Bayes* because the prior parameters are chosen according to the measurements, that are also used to fit the parameters of the likelihood, therefore they are used twice and this typically is a deprecated behavior in Bayesian analysis, even though quite common.

7.3.1 Linear Model

As a first step we fit the linear model, using all the available data as input variables, it has to be noted that in reality the response variable is the negative logarithm of the RSSI. In Table 4 it is possible to see the output of the fitting; it is clear that al the variables are significant except for the Elevation. This was not expected, because elevation typically has a major role in propagation models, but this is not the case. Possible reasons are:

- A non sufficient accuracy of the elevation data due to a too large pace of the altitude grid,
- The change in altitude in the area is not so relevant when compared to the height of the nearby building in the Helsinki downtown where the measurements where conducted: an average 3-4 floors construction can be about 12 meters and the 75%

percentile of the terrain elevation is about 11 meters, therefore it can be that the shadowing due to the buildings, that is not considered in the model fit is actually stronger than the elevation difference.

	Estimate [1]	Std. Error[2]	t-value [3]	p-value [4]
Intercept	-497	127	-3.9	0.000164
Elevation	0.5	0.5	1.024	0.3
Transmiter Distance	-0.57	0.176	-3.26	0.0015
Longitude	33.43	9.70	3.4	0.0008
Latitude	-5.39	1.98	-2.718	0.0077

Table 4: Complete linear model fit.

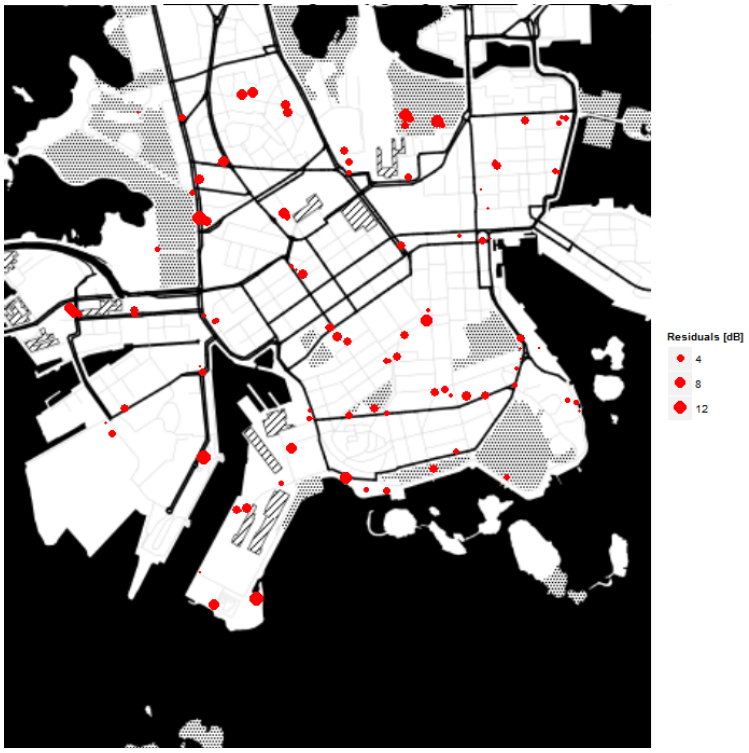

Figure 49: Bubble plot of the residuals, after they have been back transformed to the original units, as usual, these are bigger where less samples have been taken.

Residuals are shown in their geographic location in Figure 49, before going any further we check their distribution; the result is noteworthy, since they appear to be normal, as can be seen in Figure 50.

1 Value of $\hat{\beta}_i$, the punctual estimate of the linear coefficient.
2 Measure of the variability in the estimate for the coefficient.
3 Value of the Test Statistics to be compared with the quantile of the t of Student.
4 p-value of the t test.

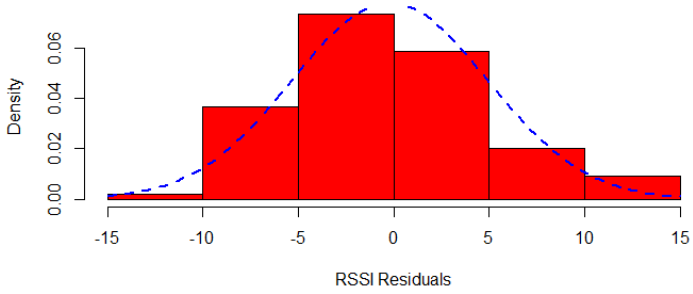

Figure 50: Histogram of the residuals, after they have been back transformed to the original units, a normal density lin has been added for a qualitative comparison.

To prove the graphical fit we use a Shapiro Test, that does not refuse the hypothesis of normality:

$$W = 0.9855, \qquad p - value = 0.2897. \tag{7.10}$$

Empirical Variogram

After looking at Table 4 we opt for a smaller model, a subset of the present one, that does not have the Elevation as a covariate. Fitting parameters and residuals are almost identical to the previous model and therefore are not reported.
At this point the variogram of the residuals is drawn. As it has been anticipated in the theoretical part, a binned estimation of the variogram is done, two options have been considered: the traditional binning and the Cressie's regularized version, see [16], that is supposed to be more robust. As it can be seen in Figure 51, the fit is not so different, therefore the classical estimation is used. The fact that different techniques produce similar results is good news a validation of the goodness and robustness of the chosen path.

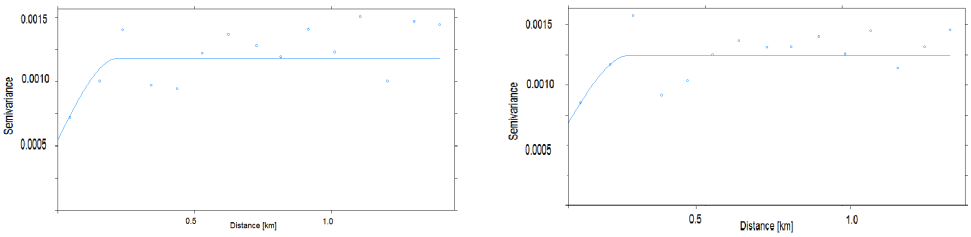

Figure 51: Comparison of Cressie's (left) and classical (right) Empirical Variogram.

As it has been announced what is left to do is estimate the easiest model that fits the variogram. The parameters estimation is done through the option `optim` in `variofit` in the `gstat` library using the R statistical environment. There is not a lot of difference between the various models, the choice is for the spherical model, since it is easy, with a low number of parameters, presents a finite range, that has theoretical reasons to exist and shows a nice regularity, but not an excessive one, since it is only linear in the origin. It has to be noted

that various techniques for optimization have been used (WLS, OLS, `optim`, `nlm`), but they all return similar results. These techniques are very different e.g., while `wls` furnishes a point estimate, the optimization routine `optim` is able to infer which one is the model in the determined class (in this case the spherical one), that best fits the data, given a value for τ^2 and σ^2 which are estimated "by eye" from the empirical variogram at the beginning of the process.

Gaussian Model

Until now we have fit a model without specifying any type of distributional hypothesis. Now we use the `linkfit` command to estimate the coefficients of the quota and the covariate relative to the transmitter distance. The resulting fit is:

$$\hat{\beta}_0 = -0.7638 \; \hat{\beta}_1 = -5.277. \tag{7.11}$$

7.3.2 Fixed Effects Model

We first plot the traceplots for the chains of the variogram parameters: it looks like the chains have reached convergence.
To decide whether the Markov chain has reached its stationary and the desired posterior distribution, to set the burn-in and the total number of iterations, we look at traceplots.
These are diagnostic tools designed to verify a necessary but not sufficient condition for convergence but there are no conclusive tests that can tell whether and when the Markov chain has converged to its stationary distribution.
It is important to check the convergence of all parameters, and not just those of interest, before proceeding to make any inference because certain parameters can appear to have very good convergence behavior, but that could be misleading due to the slow convergence of other parameters.
Trace plots of samples versus the simulation index can be very useful in assessing convergence. The trace can tell if the chain has not yet converged to its stationary distribution (are a longer burn in or more iterations needed?), but a trace can also tell you whether the chain is mixing well. A chain might have reached stationarity if the distribution of points is not changing as the chain progresses. The aspects of stationarity that are most recognizable from a trace plot are a relatively constant mean and variance. A chain that mixes well traverses its posterior space rapidly, and it can jump from one remote region of the posterior to another in relatively few steps. In Figure 52 it is possible to see the results of the trace plots, which are very thick, meaning that the chain has likely reached stationarity, the densities on the right look like normal distribution and this means that the space is well explored, there are no holes nor locations which are visited too many times.

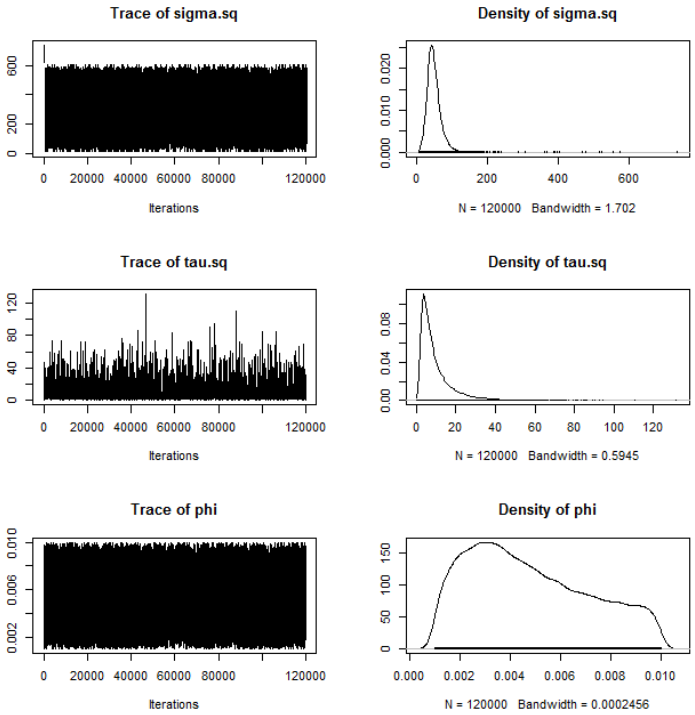

Figure 52: Traceplots and densities of the variables explored by the chains. From top to bottom the analyzed variables are σ^2, τ^2 and ϕ, which is the inverse ratio of the range R.

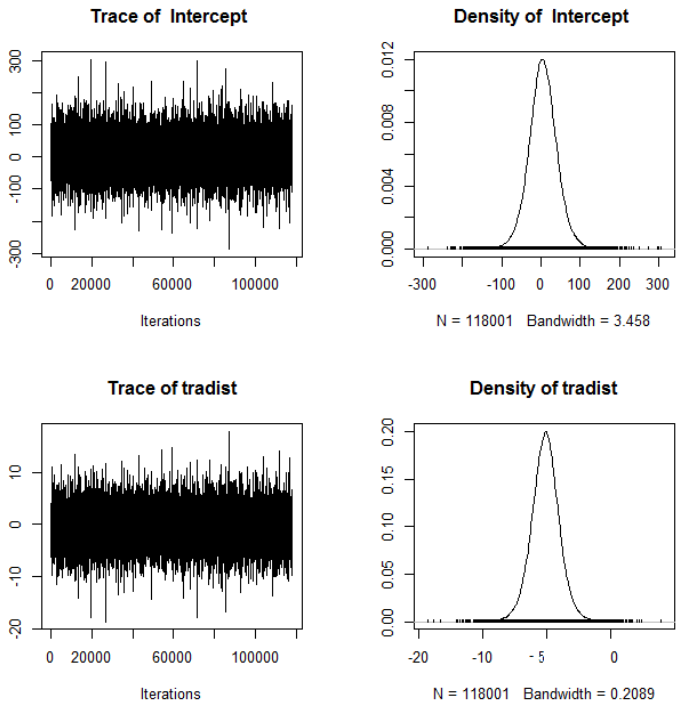

Figure 53: Traceplot and density estimator of the covariates coefficients.

As seen before, the covariate results significant, since its estimation is very far from 0 and the errors (both the naive and the time series one) are small in proportion. The difference between the two error types is that Naive Std. Error is the real standard error of the simulations, while the Time Series Std. Error is an estimation of the real Std. Error, keeping into account that realizations are not really *i.i.d.*. In this case the errors almost coincide, an index that the chains are mixing very well.

	Estimate	Std. Error	Naive Std. Error	Time-Series Std. Error
Intercept	-0.27	36.9	0.107	0.107
Transmiter Distance	-5.27	2.226	0.006	0.006

Table 5: Fitting Summaries.

	2.5%	25%	50%	75%	97.5%
Intercept	-96.3	-17.442	-0.27	27.749	79.329
Transmiter Distance	-7.7	-6.6	-5.27	-4.09	-0.13

Table 6: Percentiles of each variable. The transmitter Distance 95% interval does not contain the 0, therefore is considered a significant variable.

7.3.3 Prediction

Prediction therefore can be comparable with universal kriging one for the type of covariates.

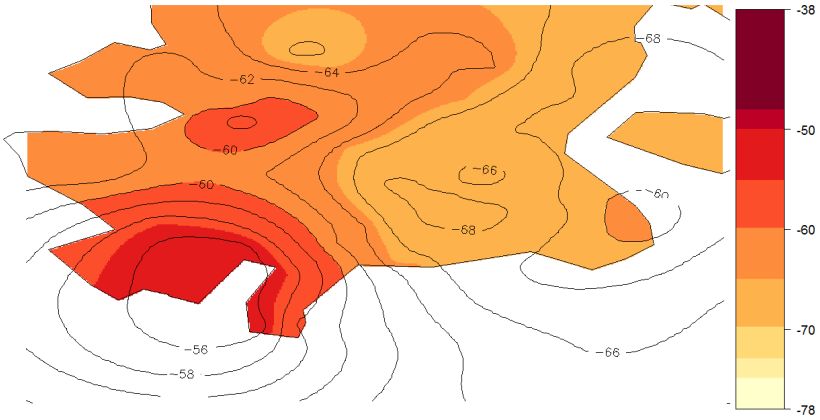

Figure 54: Prediction with Empirical Bayes approach.

7.4 TRULY BAYESIAN PRIOR

A "truly Bayesian prior" can be obtained fitting the parameters from a physical model. The idea is not to just obtain another estimation of the field, but mostly to validate what has been done so far with what the typical models used in literature.
The decision to use simulated data to fit the prior

This does not have any corresponding in literature to the best of our knowledge.
We proceeded simulating the field using the ITU-R recommendation [33] for predicting the path loss taking into account the terrain elevation, from the Espoo TV transmitter to the Helsinki downtown area where the measurements were made; the result can be seen in Figure 55.

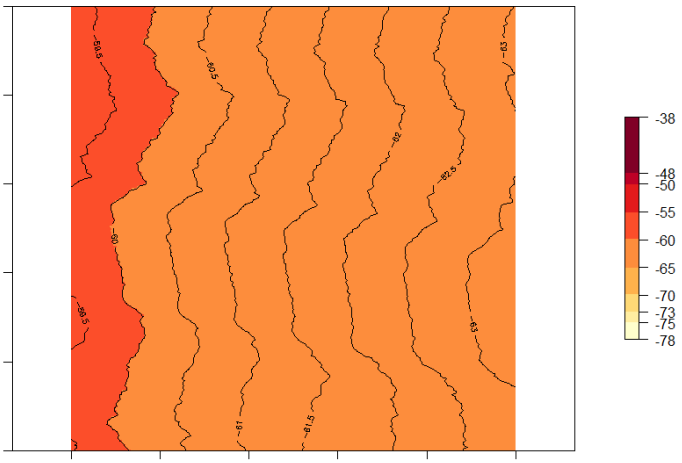

Figure 55: RSSI of thesimulated field according to the ITU Recommendation, see [33].

7.4.1 Variogram Estimation

The estimation of the variogram leads to a Spherical Variogram can be visualized in Figure 56.

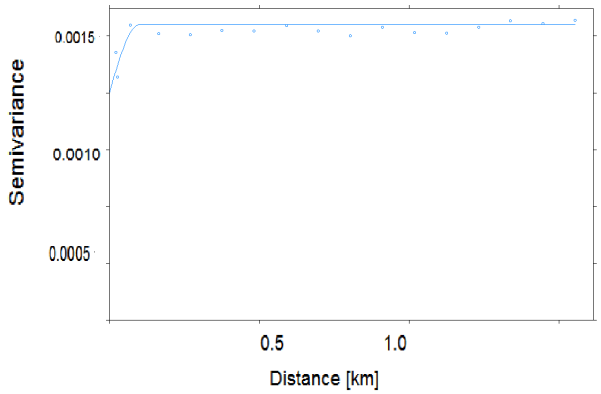

Figure 56: Fitting of the variogram.

Using similar priors as the ones that have already been used for the empirical Bayesian approach, choosing all the parameters interpolating from the variogram presented in Figure 56 and using reasonable variances (considering the sample mean) the estimation visible in Figure 57, which is surely encouraging, since it is perfectly in line with the results of the prediction models.

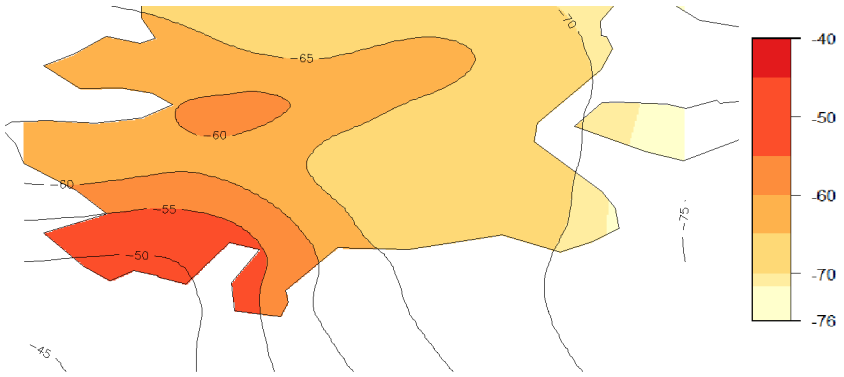

Figure 57: Prediction using the purely Bayesian approach.

Improvements could be done, between these:

- The physical model, that performs a good estimation in average, does not fit any type of covariance function, therefore it is probably not idyllic for our purposes.
- We have chosen to undersample the simulated field for computational reasons,. Other models could be used to get better results, here the point was to validate the model and the results, not just the prediction.

7.5 VANTAGES AND DRAWBACKS OF BAYESIAN MODELING

7.5.1 Vantages of Bayesian Modelling

- There is a critical difference between the one-dimensional time series data and the two-dimensional spatial data: *full-order* versus *partial order*. In one dimension only one "type" of asymptotics theory exists: for time that goes to infinite, in a multi-dimensional setting, though, spatial process data presents a multi-faceted situation:

  INFILL ASYMPTOTICS envisions a fixed region with more points filling the domain, that leads to a more accurate knowledge of the spatial field as the sample distance goes to zero. Being fixed the maximum distance between samples in the measurement field in a frequentist approach it is impossible to learn from association at an increasing distance. In this example of experimental design, the classical results, e.g. the Law of Great Numbers is not easily applicable.

  INCREASING DOMAIN ASYMPTOTICS The sampled points get further from each other as the sampling number increases, with a minimum relative distance.

  NEARLY INFILL ASYMPTOTICS The domain in infinite, but the number of sampled points increases in every sub-region of the domain.

  In the application we are envisioning the asymptotics we will focus on, at least at the beginning, will surely be of infill type, since a control area will be of interest, while of course, it will increase, when more devices will move in an outer area.
  The complicated definition of asymptotics in a frequentist domain is clearly encouraging for using a Bayesian framework for interference, since we do not need to rely

on any asymptotic theory, but rather obtain exact interference given whatever data have been observed.

- The fit of the valid variogram can be done in multiple ways: typical ways include LS, WLS, if a distributional model for the data is available, ML or REML techniques are available. In the classical, frequentist framework a lot of energy is devoted to the determination of the optimal estimates, however in the Bayesian setting, according to Banerjee, the estimation of the prior parameters is not such a big issue because the choice of the hyperparameters of the prior only weakly influence the point or interval estimate calculated on the posterior distributions. In this optic the parameters of the model (nugget, sill, range) are fitted, once the covariance model has been chosen, with a function like `variofit`, as in the frequentist approach, that allows to choose between different function for the optimization process, the results of which are then the end compared, but are far less influential than in the previous case.

7.5.2 Cons of Bayesian Modelling

Likelihood evaluation requires a computation of a quadratic form involving Σ^{-1} and its determinant. With spatial random effects this evaluation is deferred to the second stage of the model, but it is still present. With an increasing number of locations, such computation becomes very expensive and the result can also be unstable, therefore repeated computation, e.g. for simulation-based model fitting, can be very slow, perhaps unfeasible. It is a typical "big n" problem.
Approaches for treating this problem exist and are very different:

SUBSAMPLING, though it may be unattractive to ignore some of the available data, being the data interdependent could mean that an increment in sample size does not imply a direct increase in precision, on the contrary, adding a location very close to one that already exists, will probably increase the knowledge about the noise or measurement error component, but will not add much to the inference.

LOCAL KRIGING Making kriging "more local", the radius of the neighbourhood from which extracting samples for prediction can be modulated and made smaller when the sample density increases, this looks to me as the most natural solution to the problem.

PROXIMITY MATRIX More common for points in a lattice a proximity matrix can also be used for non lattice geo-referentiated data. The idea is to use a threshold and group all the data that are closer than that distance. This eases the problem computationally but of course there is a loss of information that depends on the size of the matrix cell.

SPECTRAL METHODS Working in the frequency domain implies building a periodogram and using Whittle approximation of the likelihood in the spectral domain to approximate the likelihood in the original domain, this implies no need for matrix inversion and therefore a faster computation, considering that going back to the original domain should be straightforward. The needed computational step include a discretization to implement an FFT and a certain arbitrariness for the choice of the low frequencies to be discarded before calculating the periodogram. Experience suggests that Whittle technique works well in the center, but it is not satisfactory in the tails, therefore variance estimation is rather poor.

8 OPTIMIZATION OF SAMPLE PATTERNS

So far we have talked about propagation models, but numerous are the variables to be considered, even before a measurement is actually planned. An important one is where to localize the measurement. Field measurements are resource consuming, therefore optimizing the sampling scheme could bring noteworthy economy in measurement costs. Moreover systematic sampling can sometimes not capture all the large-scale fading characteristics e.g. due to irregular terrain and therefore a cheaper sampling campaign could also mean an improvement in the precision of local field strength estimation.
The idea is to choose the optimal sampling pattern, so to minimize the average Kriging prediction error variance.
As clearly stated by Groenigen in [27], the only factors influencing Kriging variance are the variogram itself, the number of observations and the location where to predict. For this reason if it was possible to assume the variogram to be known in the first place, it would be possible to calculate the Kriging variance before the sampling actually takes place.
In a context like CR this is a very powerful assertion, and Spatial Simulated Annealing can be quite a powerful tool.
Unfortunately there is a problem: as suggested by [10], Kriging variance is optimum only for Kriging without covariates, if some sort of Universal Kriging is implemented the authors suggest an optimization of the sampling both in the feature and in the geographic space.
A compromise between spreading in the two spaces (geographical and feature) has to be found.
This is possible if prior knowledge of trend and covariance function is available, in the case of CR, after the calibration of a propagation model through a first sampling phase we can assume the covariance matrix is maintained.
Spatial Simulated Annealing is an iterative search algorithm for finding the global minimum of an objective function. In this case the object is to optimize the spatial sampling, with a constraint on the maximum number of samples and locations in which to sample.
Starting with a random sampling scheme S_0 the following one is obtained randomly perturbing the previous one. At this point the new scheme S_1 is accepted with probability one if the objective function Φ evaluated on the scheme improves, while it is accepted occasionally in the opposite case, to avoid local minima.
Let us see this in formulas:

$$\begin{cases} P(S_i \to S_{i+1}) = 1 & \Phi(S_{i+1}) \leq \Phi(S_i) \\ P(S_i \to S_{i+1}) = \exp\{\frac{\Phi(S_i) - \Phi(S_i+1)}{T}\} & \Phi(S_{i+1}) > \Phi(S_i) \end{cases} \tag{8.1}$$

T is a control parameter known as system temperature.
In [51] two objective functions have been considered:

- the Mean Shortest Distance (MMSD), whose idea is to minimize the distance between a random chosen point and the closest sampling location. This criterion aims at even spreading the sampling points on the sampling grid, see the original paper [69].

- the Mean Universal Kriging Variance (MUKV) that can be written as:

$$\Phi_{MUKV} = \frac{1}{N} \sum_{i=1}^{N} \sigma^2 \tag{8.2}$$

  where σ^2, the Kriging variance, depends both on kriging error variance and the prediction error variance of the trend.

Spatial Simulated Annealing is useful for designing field measurement campaigns and for applications in sensing that require real-time knowledge of radio environment. It is unfortunately very power consuming and therefore in CR field it will probably be an interesting technique at the moment, while the environments are being setting up and not in a long time optic. We think that better and cheaper solutions, have to be found, mostly in a distributed environment, that will have to be able to change with very short notice.

9 CLASSIFICATIONS

So far we have provided techniques for estimation of Test Statistics and RSSI, but what clearly matters to the final cognitive radio user is if the channel is or not idle.

9.1 SUPERVISED CLASSIFICATION

The aim of clustering is to label some measurements, so to group them, in a number, g of groups. Let us assume to group all the samples in a matrix $\mathbf{X} \in \mathbb{R}^{N \times p}$, where N is the number of samples and p is the dimension of the predictors space. In the present work p will always be equal to 2, since the space we are analyzing is the one defined by Test Statistics and RSSI. Given the label $L \in \{1, 2, \dots g\}$ we assume that $X|L = i \sim F_i$, and we define $p_i = \mathbb{P}(L = i)$, so that $\sum_{i=1}^{g} p_i = 1$. Finally let us define the misclassification costs $c(i, j) \geq 0, \forall (i, j) \in \{1, 2, \dots g\} \times \{1, 2, \dots g\}$ and correct classification induces no cost, so that $c(i, i) = 0 \; \forall i \in \{1, 2, \dots g\}$.
In this thesis work the classifiers are *dichotomic*, meaning that $g = 2$ and $L \in \{0, 1\}$, the channel can be idle or occupied, in the following the 0 label will refer to the idle channel, while 1 to the occupied one. The classifier can be identified by the partition of $\mathbb{R}^{g=2}$ in $\{R_0, R_1\}$, so that $R_0 \cap R_1 = \emptyset$ and $R_0 \cup R_1 = \mathbb{R}^2$.
It is therefore possible to define the Expected Cost of Misclassification (ECM) as the product between the probability a posteriori of wrongly labeling and therefore classifying, weighted by the misclassification cost:

$$\text{ECM} = c(1|0)p_0 \int_{R_1} f_0(\underline{x})d\underline{x} + c(0|1)p_1 \int_{R_0} f_1(\underline{x})d\underline{x} \tag{9.1}$$

where $f_i(\vec{x})$ is the probability density function of $X_{L=i}, \forall i \in 0, 1$.
De facto we are integrating in the two existing regions of the space ((R_0, R_1) has already been said to be a partition of $\mathbb{R}^2$) the a posteriori probability of misclassifying, weighting by the cost vector.
It is possible to define the optimal classifier the one that minimizes the expected misclassification cost.
From this it is possible to define the optimal partition as

$$R_0^* = \{\underline{x} \in \mathbb{R}^p \colon c(0|1)p_1 f_1(\underline{x})\} = \min\{c(0|1)p_1 f_1(\underline{x});\; c(1|0)p_0 f_0(\underline{x})\} \tag{9.2}$$

and equivalently

$$R_1^* = \{\underline{x} \in \mathbb{R}^p \colon c(1|0)p_0 f_0(\underline{x})\} = \min\{c(1|0)p_0 f_0(\underline{x});\; c(0|1)p_1 f_1(\underline{x})\}. \tag{9.3}$$

This is a meaningful result, since it states that for every single sample $\underline{x} \in \mathbb{R}^p$, the optimal region is the one that minimizes its expected misclassification cost.
Misclassification costs here will be considered to be symmetric ($c(0|1) = c(1|0)$), but in a real cognitive radio environment the choice will be to protect the PUs as much as possible, therefore $c(1|0) \gg c(0|1)$, it will be more costly to consider an occupied channel to be idle

than viceversa. Costs will be thought from a conservative viewpoint but transmission will have to be granted once in a while, for this reason an apposite study is needed to properly define the ratio between the costs.

9.2 QUADRATIC CLASSIFIER

A closed form of the ECM is needed for pragmatical reasons, let's therefore add two hypotheses that are not necessarily satisfied by the sample, but can help to tackle the problem more efficiently.
Let us assume that:

- $X|L = i \sim \mathcal{N}_p(\underline{\mu_i}, \Sigma_i)$
- $c(i,k)$ is constant $\forall i \neq k$.

Then the optimal classifier is quadratic, see [35], meaning that the line that divides the two tiles, also known as *discriminant function*, is quadratic. For this reason this technique is also known as Quadratic Discriminant Analysis (QDA).

9.3 LINEAR CLASSIFIER

Let us assume moreover that

- Prior probabilities are the same for every label: $p = p_i, \forall i$, in our case a priori it would mean that the probability of a channel of being idle or occupied is the same.
- Covariances are all the same: $\Sigma = \Sigma_i, \forall i$. This hypothesis, for what stated in the theoretic part concerning cyclostationary detectors is surely not valid in our setting, since we know that $\Sigma_1 \gg \Sigma_0$.

In this very specific setting the optimal classifier is linear, basically the R_i are single tiles in a Voronoi tassellation of the predictor space, using Mahalanobis distance and the classification technique is consequently called Linear Discriminant Analysis (LDA).
In a dichotomic setting if the discrimination function is linear it is also the direction of maximum discrimination, along which difference between groups is maximum compared to within groups one.
Fischer has shown that this classifier is robust for non normal distribution.

9.4 APPLICATION

The number of parameters introduced so far is quite high, densities f_0 and f_1 are unknown and priori probabilities p_0 and p_1 need to be estimated as well. To do this a *training set* is used, that is made up of already labeled samples. In the following all Σ_i will be considered equal, even though, as we have said this is not realistic. Since the single Σ_i would be estimated with sample deviations S_i, $\hat{\Sigma} = S_{\text{pooled}} = \frac{1}{n-g}\sum_{i=0}^{g-1} S_i(n_i - 1) = \frac{N-1}{2(N-1)}\sum_{i=0}^{1} S_i = \frac{S_0+S_1}{2}$, the "arithmetic mean" of the sample variances in the two cases of occupancy, that in our case have the same numerosity since the number of locations is constant through the

channels. p_i is estimated by $\hat{p}_i = \frac{n_i}{n}$, where n_i is the number of occupied (n_1)(or not (n_0)) channels out of the total number of samples. Since our test sample has only two channels and both have the same numerosity $n_i = n_0 = n_1 = N$, the total number of samples for each channel and therefore $\hat{p}_i = 0.5 i = 0, 1$. This assumptions is equivalent to state that there is no have previous information on the occupancy of a new channel, somehow making the model valid in all settings and not for a singular sampling campaign.
To estimate a classifier the Actual Error Rate (AER), should be used:

$$\text{AER} = p_0 \int_{R_1} f_0(\underline{x}) d\underline{x} + p_1 \int_{R_0} f_1(\underline{x}) d\underline{x}. \tag{9.4}$$

This is not possible to calculate because of the numerous unknowns $(p_0, p_1, f_0, f_1, R_0, R_1)$. A "dangerous" option is to use the Apparent Error Rate (APER): defining $n_i = n_{i,c} + n_{i,m}$, basically dividing the number of samples in the training set that can are correctly classified (c) and those that are not (n) per label, so that for example $n_{0,m}$ is the number of samples that are misclassified as idle, therefore:

$$\begin{aligned} \text{APER} =& \frac{n_{0,m} + n_{1,m}}{n_0 + n_1} = \frac{n_0}{n_0 + n_1} \frac{n_{1,m}}{n_0} + \frac{n_1}{n_0 + n_1} \frac{n_{0,m}}{n_1} = \\ =& \hat{p}_0 \int_{R_1} \widehat{f_0(\underline{x}) d\underline{x}} + \hat{p}_1 \int_{R_0} \widehat{f_1(\underline{x}) d\underline{x}}. \end{aligned} \tag{9.5}$$

The technique is dangerous because the idea is to build a classifier on already labeled data and then predict the same data through the classifier. Inevitably the classifier is adapted to the training set, therefore the estimation error would probably go underestimated and the risk of overfitting is quite present. This brings to a tradeoff problem: how to size training and test set, so to have a good classifier, but also a good estimation of the classification error?
A valid alternative to the division of the samples in training and test set is to operate crossvalidation, therefore computing the classifier on $N - x$ samples and use it to predict the occupancy in the remaining x locations. In this work only 1-fold ($x = 1$) crossvalidation has been computed, so that there was no need of checking if a classifier obtained using a subset of the training test was very different from the others, since the cardinality of the "changing test set" was varying.

9.5 RESULTS

We use the measurements from MC3. All the samples from channels 22 and 44, that are known to be respectively an idle and an occupied channel, will be our training set, while the test set in the following will be composed of channels 42, 44 and 46.
For both classifiers (linear and quadratic) we will calculate the classifier and use it to re-predict the channel occupancy on the training set, therefore calculating APER to estimate AER and then we will compare this result with the AER calculated through crossvalidation both on the training and on the test set.

9.5.1 Linear Classifier

The first technique we use is a linear classifier, with Test Statistics and RSSI as the only covariates, even though the hypotheses are not to be considered satisfied, since the covari-

ance matrices are very different in the two cases: in the idle channel the sample covariance matrix is

$$S_0 = \begin{bmatrix} 94.38 & 3.56 \\ 3.57 & 9.90 \end{bmatrix} \tag{9.6}$$

while in the occupied channel it amounts to:

$$S_1 = \begin{bmatrix} 2048.75 & -43.69 \\ -43.68 & 43.16 \end{bmatrix}. \tag{9.7}$$

Difference between sample variance is easy to see even in Figure 58, since the occupied channel has a much wider cluster in this space, but is especially visible in Figure 12.

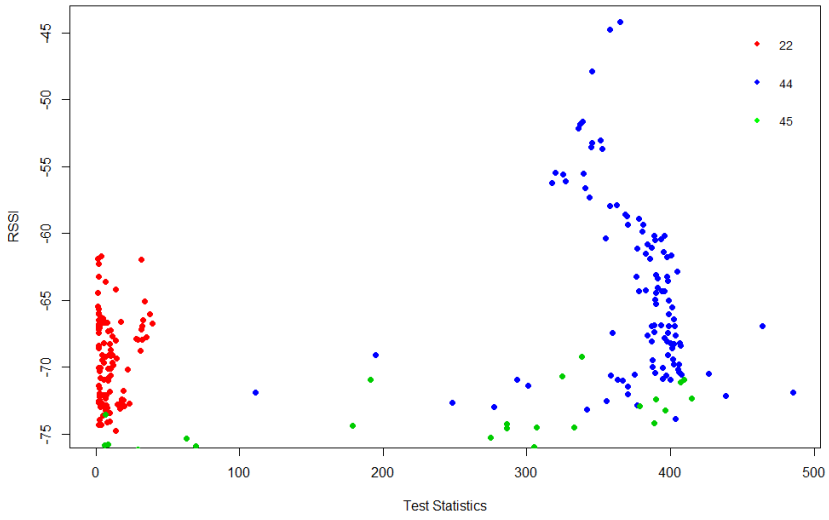

Figure 58: Visualisation of channels 22, 44, 45 in the plane (Test Statistics, RSSI). Channels 22 and 44 are respectively unoccupied and occupied in all locations and they form two clear clusters, while channel 45, whose occupancy is not constant through the locations, since the signal comes from Estonia and is therefore more easily perceived along the coast, is somehow between the two clusters, showing even very high values of Test Statistics and is reported here as a comparison.

Fortunately, though, the Mahalanobis distance between the clusters will be shown to be high enough to differentiate the clusters.
Using $p = (p_0, p_1) = (0.5, 0.5)$, the linear discriminant is visible in Figure 59.
This gives origin to a very good confusion matrix, that shows the number of misclassified samples per label:

	$\hat{L} = 0$	$\hat{L} = 1$
$L = 0$	109	0
$L = 1$	2	107

it is therefore possible to calculate the APER as the percentage of misclassified channel locations out of the total.

The obtained APER is approximately 0.00917, less than 1% of the total locations is misclassified. To be sure of the result, we opt for calculating the AER through 1-fold crossvalidation, that results into exactly the same confusion matrix. This is because in the measurement data that has been used the two groups have exactly the same number of samples, therefore $\hat{p}_i$ is a good estimation for $p_i, i = 0, 1$. The AER for 1-fold misclassification coincides with APER.

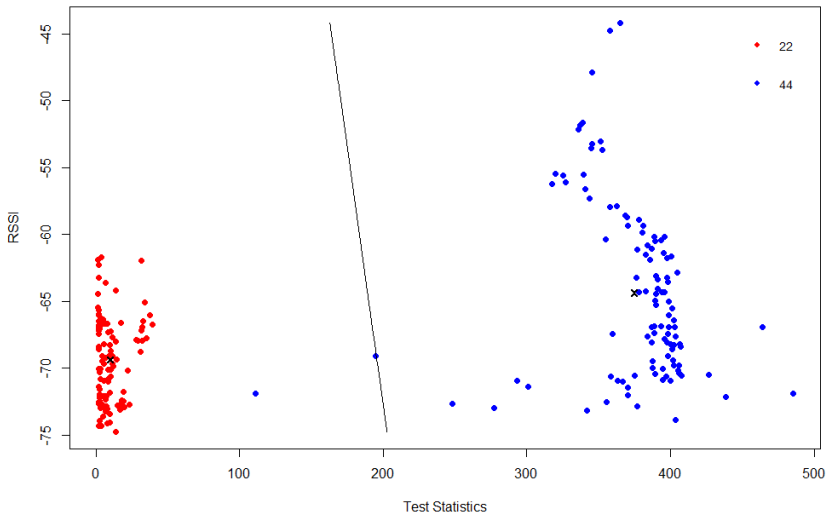

Figure 59: Linear classifier visualized on the test set. It is possible to see the two locations that are incorrectly labeled.

9.5.2 Quadratic Classifier

This classifier is adopted here, even if, as before, the hypothesis are not completely satisfied, since none of the tested $X|L = i$ has returned a p-value of the Shapiro Test that could have made us convinced not to refuse the Null Hypothesis of Normality. Notwithstanding this a confusion matrix is built and the result is even better than before, as a matter of fact no samples are misclassified:

	$\hat{L} = 0$	$\hat{L} = 1$
$L = 0$	109	0
$L = 1$	0	109

The APER is therefore exactly null, as the AER built with 1-fold crossvalidation. The classifier can be seen in Figure 60.

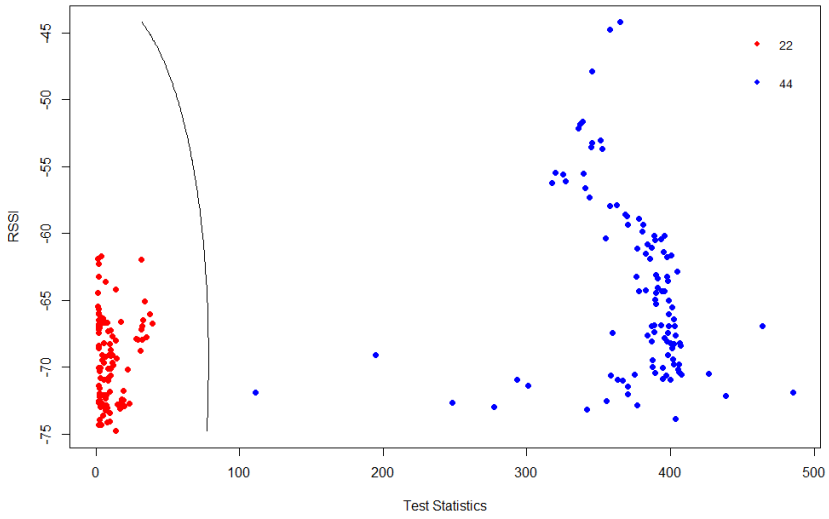

Figure 60: Quadratic classifier on the test set. No samples are misclassified.

9.5.3 Comparison between Linear and Quadratic Classifiers

It is natural to ask ourselves which of the two classifiers does a better job, since the computational cost is almost the same. To make this comparison we use other three channels: 42, 45, 46, the first is known to be idle, the last to be occupied and the middle one is the one that is perceived from Tallinn only in some locations, mostly close to the seashore, see Table 1.
We use the classifier that has been built on the channels 22 and 44, without adding any information, the result is to be seen in Figure 61: it is clear that the quadratic classifier is more conservative than the linear one, since it asserts the channel is occupied in more locations, *de facto* protecting primary receivers. While the two channels which have a strong character (idle or occupied) are almost everywhere correctly classified, the differences in channel 45 amounts to 25% of the total sampled locations.

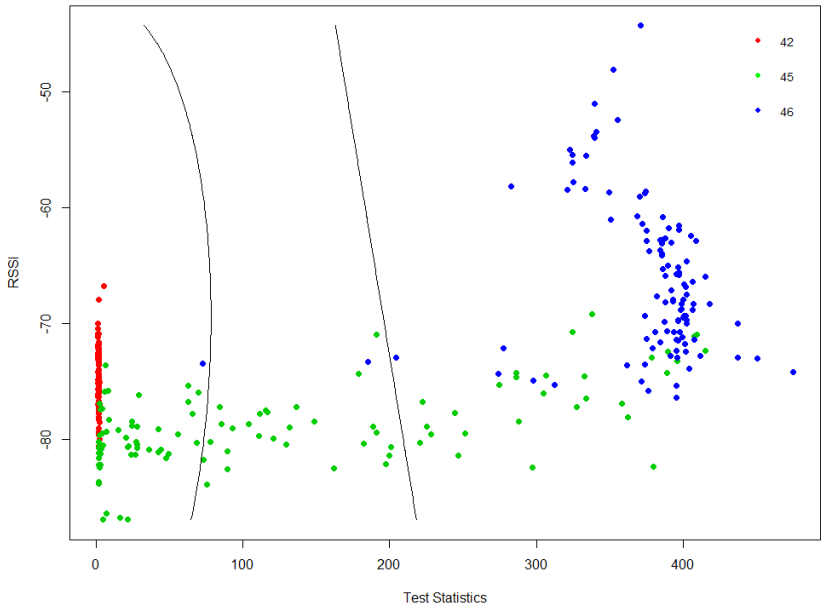

Figure 61: Comparison between linear and quadratic classifiers in three new channel, one busy (blue, 46), one idle (red, 42) and one that is location dependant (green, 45).

We proceed by visualizing in Figure 62 the positions that are considered to be idle or occupied for channel 45 by the two different classifiers. It is clear that, as already said, the quadratic classifier is more conservative, the choice between the two depends on the situation, in this case the Estonian channel does not have to be protected, but this certainly depends on location, time etc., that have to be implemented in the classifier.

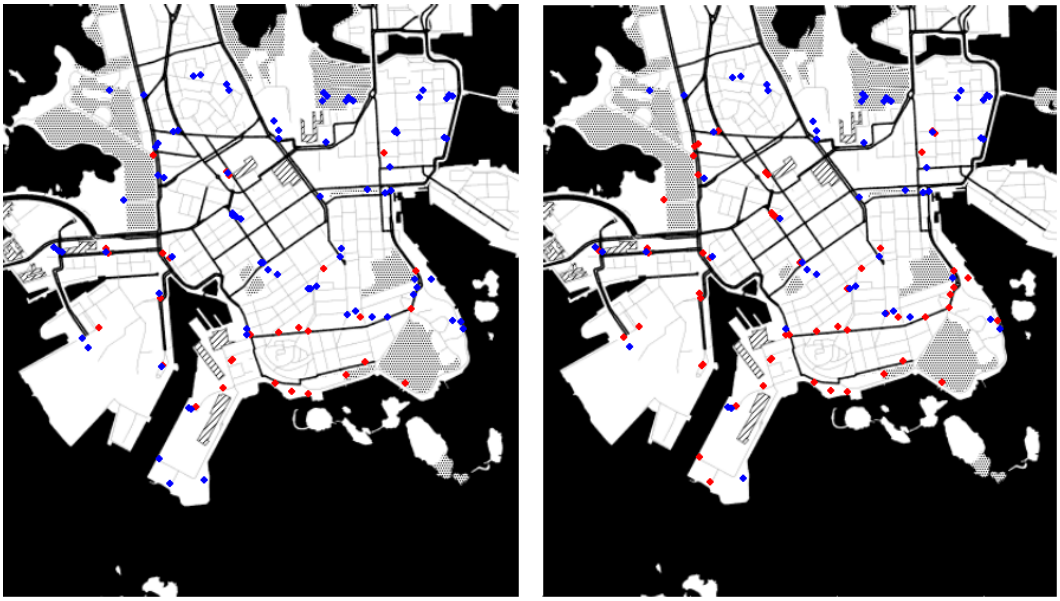

Figure 62: Comparison between linear classifier (left) and quadratic one (right). In red are locations where the channel 45 is considered to be busy by the classifier, blue where it is not.

9.5.4 Univariate Classifier

Noticing that the discriminating functions divide the space "almost vertically" the question is if a univariate classifier, with only the Test Statistics as predictor. This idea is looked after in Figure 63.

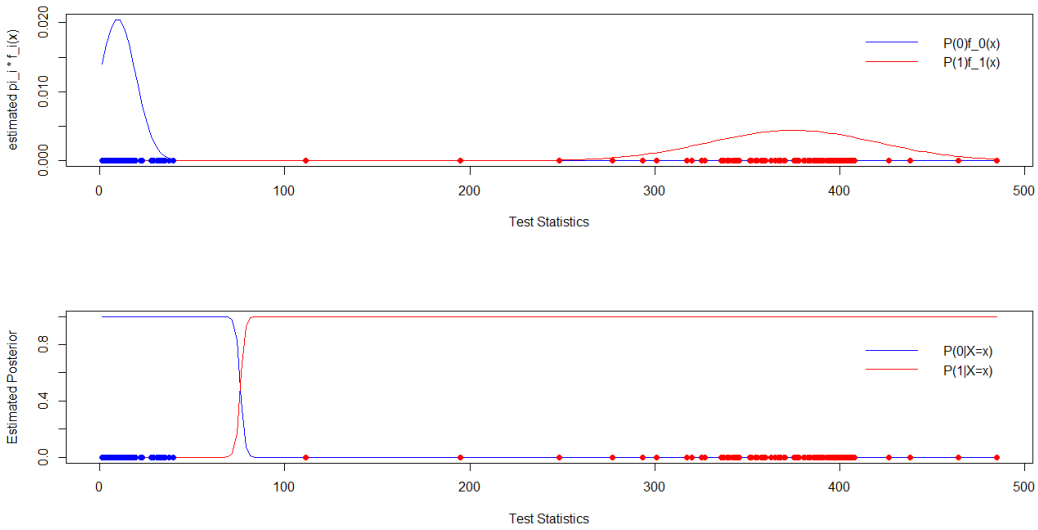

Figure 63: In the top figure the density a priori is multiplied by p. These are not densities, therefore they do not integrate to 1, but to p, they are a mixture of densities. In the bottom. All densities $f_i, i = 0, 1$ are represented as normal, centered in the sampled mean, with a variance equal to the sample one. The lines represent the theoretic curves, the dots the actual sampled points, which colour is the real label, not the colour of the classification. In the bottom plot the threshold is clearly individuated by the intersection between the curves.

A univariate classifier requires even less computational power, since it boils down to the comparison of the Test Statistics with a threshold, that is visible in the bottom graph in Figure 63. An option could be to use a *sequential classifier*, that takes care of classifying as idle the samples where the Test Statistics is way below the threshold and declare busy those where the value is way above it, while in the grey band around it a bivariate classifier (linear or quadratic) could be implemented. This would be a cheaper option in computational terms, however to decide how to shape the grey zone we think that some labeled samples from Channel 45 are needed before taking any decision in this regard.

9.6 CONCLUSIONS

Classifiers have been shown to perform extremely well and are very easy to implement, they surely will have an impact in real time decisions about transmission, where maps of channels occupancy are not needed, but fast decisions have to be taken, caring mostly about consumption of the mobile decision unit.
More complex, maybe Bayesian classifiers could be built, as an example adding the channel name as a factor, so that channels which are usually occupied would be more likely to be signaled as busy and viceversa, but we really do not see the reason to do so, because of the extremely good performance of lower power consuming classifiers.
It must be remembered that this type of analysis does is not optimal under any type of condition, they are simply geometrical tools to divide the space. Their utility is mostly

due to the fact that no assumptions have to be made in general more than the ones that we have made, this of course, implies that they have very few optimality properties, as already stated. Nonetheless we believe that very low computational power and very good classifications are the required characteristics of a classifier and therefore we strongly push towards adopting these easier tools.
The option of using only the Test Statistics as a classifier is intriguing, since it would half the state space and classification would simply translate in the comparison with a threshold and almost vertical lines in Figure 61 strongly point towards this decision. Nonetheless, we think that labeled samples from Channel 45 are needed before taking any decision in this regard.

10 CONCLUSIONS

Cognitive radios have a huge potential in shaping the future of wireless communications: they hold the promise of a new frontier with dynamic coordination for spectrum-sharing process for an improved spectrum utilization under constantly changing conditions.
In this thesis work applications of Spatial Statistics in the area of wireless communications have shown how to get a detailed understanding of the surrounding radio environment with low complexity models to guarantee a manageable overhead.
After implementing numerous interpolation techniques, whose vantages and drawbacks have been evidenced, and suggesting prediction models optimized for a context where computational cost is an issue, spatial classification models have been provided, and have been shown to perform extremely well.
The purpose has been to obtain full yet succinct representation and prediction of network state metrics and accurate identification of anomalies from possibly partial and corrupted measurement data.
As it has already been evidenced the aim of the work has been to compare a high number of existing techniques while looking for one that performs well enough for our scopes and is not too heavy computationally. In the case of primary signals that are almost static for long periods, such as the DVB-T that has been taken into account in this work, the best option in terms of correctness is to use the Ordinary Kriging if there is no need of extrapolation out of the measurements hull, otherwise the more expensive Universal Kriging is the only option to take care of the spatial trend. In the case of signals which, on the other side, are not to be considered constant in time and therefore cause the maps to have to be recomputed more often, Kriging is probably not an option, or at least, has to be computed only once in a while, keeping the best performing technique, splines interpolation, as the main option. Interpolation is clearly not a synonym for prediction, there is no account for the variance of the punctual value, but it can be used as a threshold for deciding whether or not transmitting, until a prediction is computed.
The Bayesian part of this work is to be intended from another point of view, mostly as a way to check whether the model presented in the standard [33] could be used in locations where no measurements are present. Another purpose of this model is using the posterior to predict few observations if these are needed since the computation of the posterior is a long process, but can be done almost automatically once the prior parameters have been fixed and mostly it does not need to be performed every time an interpolation is performed but much more rarely; how rarely clearly depends on the environment and the signal characteristics.
Looking forward in time there is an unceasing demand for continuous, innovative and large-scale distributed algorithms, complemented by collaborative and adaptive monitoring platforms to accomplish the objectives of network management and control.
This will have enormous impact on the design of critical cognition infrastructure to sense, learn, and adapt to the environment where networks operate, with a minimum need for human intervention and a high resilience, robustness to missing and corrupted network data and to possibly malicious attacks.
Dynamic network cartography and statistical modeling will allow to

- unveil, map and manage anomalies that could cause congestion and limit the quality of service,
- predict channel occupancy to build a transmission plan eventually in a fully distributed fashion

A fully distributed approach is of course to be preferred for scalability and robustness considerations, even if it is way more expensive to build and more difficult to manage.
The huge quantity of challenging problems gives an idea of why CRs are such an interesting object of study, the evolution of which is probably to be cognitive networking between clusters of cognitive radios.
According to [59] until now most cooperative sensing approaches have used a rather simple one-dimensional correlation models, with only one PU. No model is available that considers dependence on frequency in the case of multiple PUs. This shows how the path to walk is still long and full of difficulties, mostly because of the lack of theoretical framework for realistic scenarios that would allow to generate credible simulations. There is a strong need of efficient statistics to allow synthesis, sharing and comparison of results and a huge necessity to develop models to capture spatial correlation and shadowing, together with spectrum occupancy; we hope to have done a small contribution in this direction. Robust statistical models are not only of theoretical importance but will have a practical impact since they can be used to plan spectrum utilization strategies and algorithms and to dimension distributed spectrum sensing networks.
Numerous are the possible enhancements in the near future: cooperative sensing could be improved by selecting the less correlated sensors and spatial models could be used as basis for further discussion and development of spectrum regulations.
The time-aspect has not been researched about extensively enough, investigations have to be made on how to optimize the update rate for the parameters computations and to fuse only *differential information*, not to augment uselessly the communication overhead.
The future will see extensive measurement campaigns, with appropriate measurement grids and improved theoretical models, together with investigations of primary users behavior in a long time span, to finally reach a full time-space model for spectrum use.

11 LISTINGS

11.1 THIESSEN POLYGONS

```
voronoipolygons <- function(x) {

  if (.hasSlot(x, 'coords')) {
    crds <- x@coords
  } else crds <- x
  z <- deldir(crds[,1], crds[,2])
  w <- tile.list(z)
  polys <- vector(mode='list', length=length(w))
  for (i in seq(along=polys)) {
    pcrds <- cbind(w[[i]]$x, w[[i]]$y)
    pcrds <- rbind(pcrds, pcrds[1,])
    polys[[i]] <- Polygons(list(Polygon(pcrds)), ID=as.character(i))
  }
  SP <- SpatialPolygons(polys)
  voronoi <- SpatialPolygonsDataFrame(SP, data=data.frame(x=crds[,1], y=crds[,2], row.names=sapply(slot(SP,'
       polygons'),function(x) slot(x, 'ID'))))
}
```

11.2 TRIANGOLATION

```
#read data and find unique locations, build grid in which to estimate
bbox<-c(24.90, 60.14, 24.97, 60.18)
predgrid <- expand.grid(lon=seq(from=bbox[1], to=bbox[3], length.out=400),
                        lat=seq(from=bbox[2], to=bbox[4], length.out=400))
N=400*400
for(i in 1:N){
  for (j in 1:109){#calculate distances
    temp[i,j]=geodDist(locations[j,1], locations[j,2], as.numeric(locdf[i,1]), as.numeric(locdf[i,2]), alongPath
         =FALSE)}
  stima[i]<-which.min(temp[i,])
  stima2[i]<-which.min(temp[i,-stima[i]])
  if (stima[i]<=stima2[i]) {stima2[i]=stima2[i]+1}
  stima3[i]<-which.min(temp[i,c(-stima[i], -stima2[i])])
  if (stima[i]<=stima3[i])
  {stima3[i]=stima3[i]+1}
  if (stima2[i]<=stima3[i])
  {stima3[i]=stima3[i]+1}
  #calculate and normalize weights
  lam1[i]<-1/temp[i, stima[i]]
  lam2[i]<-1/temp[i, stima2[i]]
  lam3[i]<-1/temp[i, stima3[i]]
  lamdef1[i]<-lam1[i]/(lam1[i]+lam2[i]+lam3[i])
  lamdef2[i]<-lam2[i]/(lam1[i]+lam2[i]+lam3[i])
  lamdef3[i]<-lam3[i]/(lam1[i]+lam2[i]+lam3[i])
#save output
  locdf[i,3]<-lamdef1[i]*media.stat[stima[i]]+lamdef2[i]*media.stat[stima2[i]]+lamdef3[i]*media.stat[stima3[i]]
}
```

11.3 NATURAL NEIGHBORS INTERPOLATION

```
locations<- #unique sampling locations
v2 <- voronoipolygons(locations)
proj4string(v2) <- CRS("+proj=longlat +ellps=WGS84 +no_defs+ellps=WGS84 +towgs84=0,0,0 ")
N=100
loc<-as.matrix(cbind(predgrid[,1:2]))
int<-matrix(nrow=N, ncol=109)
for (i in 1:N){#add one location of the grid to the dataframe and calculate the new tessellation
  loc.temp<-rbind(locations, loc[i,])
  vor.temp<-voronoipolygons(loc.temp)
  proj4string(vor.temp) <- CRS("+proj=longlat +ellps=WGS84 +no_defs+ellps=WGS84 +towgs84=0,0,0 ")
  for (j in 1:109){
  #calculate the interceptions
    s<-gIntersection(SpatialPolygons(vor.temp@polygons[110]), SpatialPolygons(vor.temp@polygons[i]))
    if(is.null(s)) {
      int[i,j]=0
    } else {
#calculate the interception areas
gArea(gIntersection(SpatialPolygons(vor.temp@polygons[i]), SpatialPolygons(vor.temp@polygons[overlaid.poly])))
     }}}
```

11.4 TREND SURFACES

```
data22 <- surf.ls(2, locations[,1], locations[,2],data22[,3])
trsurf2 <- trmat(obj=data22, 24.90, 24.97, 60.14, 60.18, n=400)
eqscplot(trsurf2, type = "n")
contour(trsurf2, add = TRUE)
points(data22)
```

11.5 KRIGING

```
tr<-c(60+10/60+40/3600,24+38/60+24/3600)#transmitter position
dist[1:109]<-geodDist(tr[2], tr[1], data[1:109,1], data[1:109,2])
data$rssi = log10(data$rssi*-1)#*-1 to change the sign
#transform data, build a grid, for every node calculate distance from the transmitter
vt <- variogram(rssi ~ (dist+ lon +lat), data)
mr_dist = vgm(psill=0.0007, "Sph", range = 0.7, nugget=0.0007)
plot(vt, mr_dist)
vrfit_dist = fit.variogram(vt,mr_dist)
y_dist <- krige(rssi~dist+lon+lat, data, predgrid, vrfit_dist)
y_dist$var1.pred=-1*10^y_dist$var1.pred
y_dist.pred.ras<-as.data.frame(y_dist["var1.pred"])
coordinates(y_dist.pred.ras) <- ~ x + y
gridded(y_dist.pred.ras) <- TRUE
y_dist.pred.ras <- raster(y_dist.pred.ras)
#cv
x <- krige.cv(rssi~(dist+lon+lat), data)
x$var1.pred=-1*10^x$var1.pred
x$observed=-1*10^x$observed
x$residual=(+x$observed-x$var1.pred)
bubble(x, "residual", main = "")
mean(x$residual)
round(sqrt(mean(x$residual^2)), 2)
round(sqrt(mean(x$residual^2)), 2)/mean(x$observed)
```

11.6 BAYESIAN ESTIMATOR

```
n.samples <- 120000
n.burn <- 2000
thin <- 2
starting <- list("phi"=3/effective.range,
                 "sigma.sq"=partial.sill,
                 "tau.sq"=nugget)

tuning <- list("phi"=1/effective.range,
               "sigma.sq"=sig,
               "tau.sq"=tau)

prior <- list("phi.Unif"=c(a,b),
              "sigma.sq.IG"=c(c,d),
              "tau.sq.IG"=c(e,f), "beta.Flat")

fit.test.3 <- spLM( rssi ~ dist , data=data, coords=coords,
                    starting = starting ,
                    tuning=tuning,
                    priors=prior,
                    cov.model="exponential",
                    n.samples=n.samples,
                    verbose=TRUE, n.report=10)

fit.test.3$p.theta.samples
str(fit.test.3)
m.1 <- spRecover(fit.test.3, start=n.burn, verbose=FALSE)
round(summary(m.1$p.theta.recover.samples)$quantiles[,c(3,1,5)],2)
round(summary(m.1$p.beta.recover.samples)$quantiles[,c(3,1,5)],2)
m.1.w.summary <- summary(mcmc(t(m.1$p.w.recover.samples)))$quantiles[,c(3,1,5)]
```

11.7 PLOTTING SPATIAL DATA

```
#inport the shapefile from kartat.kapsi.fi
Finland<-readOGR(".","KorkeusAlue")
#project it
Finland<-spTransform(Finland, CRS("+proj=longlat +ellps=WGS84 +no_defs+ellps=WGS84 +towgs84=0,0,0 "))
#choose an area
ext <- extent(24.9, 24.97, 60.14, 60.18)
#clip the shapefile to the extent of interest
clipe <- as(ext, "SpatialPolygons")
proj4string(clipe) <- CRS(proj4string(Finland))
cropd <- SpatialPolygonsDataFrame(clipe, data.frame(x = 1), match.ID = FALSE)
int_shp <- gIntersection(Finland, cropd)
loc.vec <- c(left = 24.90, bottom = 60.14, right = 24.97, top = 60.18)
#get the background map from GoogleMaps through Stamen
peninsula <- get_map(location = loc.vec, zoom = 13, maptype = "toner", source = "stamen")
#crop the estimation to the extent of interest
stat.sub <- crop(stat.ras, extent(int_shp))
stat.sub <- mask(stat.sub, int_shp)
#plot
ggmap(peninsula, extent="device")+
  inset_raster(stat.sub, 24.90, 24.97, 60.14, 60.18)
```

APPENDIX A: PROJECTIONS

A map projection is a systematic representation of all or part of the Earth surface on a plane. The existence of a totally distortion-free flat map is precluded by Gauss' Theorema Eggregium in differential geometry, therefore the cartographer has to choose which are the characteristics that must be shown more accurately on the map, since there is not a *best* projection, but for each application there are likely to be several appropriate projections.
Since the Earth cannot be flattened onto a plane without distortion, the general strategy for map projection is to use an intermediate surface that can be flattened, known as *developable surface*, on which the sphere is first projected and only later is laid out on a plane. The two most common developable surfaces are the cone and the cylinder; different orientation of these surfaces lead to different classes of map projections.
Let's therefore consider the geographical coordinate system (λ, ϕ) for longitude and latitude and construct a polar or Cartesian coordinate system (x, y) according to the needs, frequently referred to as *eastings* and *northings* so that:

$$x = f(\lambda, \phi),\ y = g(\lambda, \phi).$$

The infinitesimal patches on the sphere (or ellipsoid, or geoid) will be approximated by the *wrapping* or tangential plane, a set of partial differential equations will be derived, whose solution will yield f and g, once appropriate initial conditions are set.
Let's therefore consider a patch formed by the infinitesimal quadrilater $ABCD$, given by the vertices

$$A = (\lambda, \phi),\ B = (\lambda, \phi + d\phi),\ C = (\lambda + d\lambda, \phi),\ D = (\lambda + d\lambda, \phi + d\phi).$$

This will be projected into the new patch $A'B'C'D'$.
Mainly two types of maps will be analyzed: *equal area* and *conformal* (locally shape-preserving) maps.

EQUAL AREA PROJECTIONS

Given R, the average Earth radius, as stated in the projection name, it is necessary to equate the area of the patches $ABCD$ and $A'B'C'D'$, after some calculation an undetermined system in the form of

$$(\frac{\partial f}{\partial \lambda}\frac{\partial g}{\partial \phi} - \frac{\partial f}{\partial \phi}\frac{\partial g}{\partial \lambda}) = R^2 \cos(\phi)$$

is obtained and further conditions need to be imposed, *de facto* choosing one of the many possible projections. Examples are the *sinusoidal projection*, characterized by $\frac{\partial g}{\partial \phi} = R$, or the *Lambert cylindrical projection* given by $f(\lambda, \phi) = R$ and $g(\lambda, \phi) = R\sin(\phi)$.

CONFORMAL PROJECTIONS

Conformal projections are angle preserving: $\angle(AC, AB) = \angle(A'C', A'B')$, this leads to:

$$\frac{\partial f}{\partial \lambda}\frac{\partial f}{\partial \phi} + \frac{\partial g}{\partial \lambda}\frac{\partial g}{\partial \phi} = 0.$$

Numerous are the projections that belong to this class, a classical example is *Mercator Projection*, a cylindrical projection, that exists in both spherical and ellipsoidal versions, in which case non linear anisotropic scaling is employed to ensure that the map is conformal. Mathematically it is obtained letting

$$\frac{\partial g}{\partial \phi} = R \sec(\phi).$$

The *Transverse Mercator* is a widely used adaptation of this standard projection, known for delivering high accuracy in zones with an extension of few degrees in longitude, since the central meridian can be chosen at will.
The UTM uses a 2-dimensional Cartesian projected coordinate system to identify locations on the surface of the Earth, but is itself not a single map projection; the system instead divides the Earth into sixty zones, each a six-degree band of longitude, and uses a secant transverse Mercator projection in each zone. The World Geodesic System (WGS84) ellipsoid is now generally used to model the Earth in the UTM coordinate system, even though, for different geographic regions, other datum systems can be used.

DISTANCE CALCULATION ON THE EARTH SURFACE

The most common approach in Spatial Statistics is to model spatial dependence between variables as a function of distance; for data sets covering relatively small spatial domains, ordinary Euclidean distance is enough, however for larger domain, the Earth curvature must be accounted for. A fundamental concept is that of *geodesic distance*, the shortest distance between two points on the surface, also known as *great circle distance* when referred to the Earth surface.
In this work data have been conveyed in an unprojected format, a subsequent projection has been operated, so that Euclidean distance and not the Great Circle one could be used in the computations. This decision has been taken because empirical variograms are highly dependent on the definition of distance, see e.g. [19], especially the fifth chapter, where this issue is treated in detail. The choice of a Cartesian-based UTM system is due to practicality, to the popularity of this standard and to the ease of interpretation of the results. The projection has been obtained through the use of the `rgdal` package in R.

LIBRARIES

List of the R libraries used in this thesis work and reference to publications:

DELDIR For Delauney tassellation, see [66]

FIELDS To plot objects of type MBAsurf and perform splines interpolation, see [49]

GEOR To calculate and plot the empirical variogram, see [57]

GGMAP For spatial visualization with Google Maps and Open Street Maps, see [37]

GSTAT To calculate and plot the empirical variogram, see [55]

MAPTOOLS For reading and handling spatial objects, see [7]

MBA To create a surface approximation from scattered data using B-splines, see [21]

OCE To calculate the geodetic distance, see [39]

PLYR To split-apply-combine patterns, see [73]

RASTER For raster type objects and plotting, see [32]

RCOLORBREWER For the colour palettes, see [48]

RGDAL To project the data, see [6]

RGEOS To project the data, see [8]

R.MATLAB To read MAT files, see [4]

SP For manipulating spatial data, see [56]

SPBAYES To fit a Bayesian spatial model, see [22]

Please give a look to the reference section for better specification.

ABBREVIATIONS

ACF	Autocorrelation Function
AER	Actual Error Rate
APER	Apparent Error Rate
AWGN	Addictive White Gaussian Noise
BLUE	Best Linear Unbiased Estimator
BLUP	Best Linear Unbiased Predictor
BS	Base Station
CFAR	Constant False Alarm Rate
CK	Centralized Kriging
CP	Cyclic Prefix
CR	Cognitive Radio
DK	Distributed Kriging
DSA	Dynamic Spectrum Access
DSP	Digital Signal Processor
DVB-T	Digital Video Broadcasting - Terrestrial
ECM	Expected Cost of Misclassification
EGLS	Extended Generalized Least Squares
FC	Fusion Center
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FPR	False Positive Rate
GLRT	Generalized Likelihood Ratio Test
GLS	Generalized Least Squares
GPS	Global Positioning System
IDW	Inverse distance weighting
IRF-k	Intrinsic Random Function of k^{th} order
ISM	Industrial Scientific and Medical
LDA	Linear Discriminant Analysis
LOS	Line-Of-Sight
LRT	Likelihood Ratio Test
LS	Least Squares
ML	Maximum Likelihood
MMSD	Mean Shortest Distance
MOM	Method of Moments
MUKV	Mean Universal Kriging Variance
NLOS	Non Line-Of-Sight
NLS	National Land Survey of Finland
OFDM	Orthogonal Frequency Division Multiplexing
OLS	Ordinary Least Squares
PD	Probability of Detection
PFA	Probability of False Alarm
PMD	Probability of Miss Detection
PSD	Power Spectral Density
PT	Primary Transmitter
PU	Primary User
QDA	Quadratic Discriminant Analysis
QOS	Quality of Service
RAM	Random Access Memory
REML	Residual Maximum Likelihood
RF	Radio Frequency
ROC	Receiver Operating Characteristics
RSS	Residual Sum of Squares

RSSI	Received Signal Strength Indicator
RX	Receiver
SDR	Software Defined Radio
SNR	Signal Noise Ratio
SSA	Static Spectrum Access
SU	Secondary User
TPR	True Positive Rate
TSM	Time Scale Modification
TX	Transmitter
UHF	Ultra Hight Frequency
UMTS	Universal Mobile Telecommunications System
UTM	Universal Transverse Mercator
UWB	Ultra-Wide Band
WLAN	Wireless Local Area Network
WLS	Weighted Least Squares
WGS84	World Geodesic System

BIBLIOGRAPHY

[1] Erik Axell, Geert Leus, Erik G Larsson, and H Vincent Poor. Spectrum sensing for cognitive radio: State-of-the-art and recent advances. *Signal Processing Magazine, IEEE*, 29(3):101–116, 2012.

[2] Sudipto Banerjee, Bradley P Carlin, and Alan E Gelfand. *Hierarchical modeling and analysis for spatial data*. Crc Press, 2004.

[3] Randal J Barnes. The variogram sill and the sample variance. *Mathematical Geology*, 23(4):673–678, 1991.

[4] Henrik Bengtsson. *R.matlab: Read and write of MAT files together with R-to-MATLAB connectivity*, 2013. R package version 2.1.0.

[5] Sergio Bittanti. *Teoria della Predizione e del Filtraggio*. Pitagora, 1993.

[6] Roger Bivand, Tim Keitt, and Barry Rowlingson. *rgdal: Bindings for the Geospatial Data Abstraction Library*, 2013. R package version 0.8-14.

[7] Roger Bivand and Nicholas Lewin-Koh. *maptools: Tools for reading and handling spatial objects*, 2014. R package version 0.8-29.

[8] Roger Bivand and Colin Rundel. *rgeos: Interface to Geometry Engine - Open Source (GEOS)*, 2014. R package version 0.3-5.

[9] Gabriele Boccolini, Gustavo Hernandez-Penaloza, and Baltasar Beferull-Lozano. Wireless sensor network for spectrum cartography based on kriging interpolation. In *IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 1565–1570. IEEE, 2012.

[10] Dick J Brus and Gerard Heuvelink. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma*, 138(1):86–95, 2007.

[11] George Casella and Roger L Berger. *Statistical inference*, volume 70. Duxbury Press Belmont, CA, 1990.

[12] Prakash R Chakravarthi. *High-level Adaptive Signal Processing Architecture with Applications to Radar Non-Gaussian Clutter: The Problem of Weak Signal Detection*. DTIC Document, 1995.

[13] Kang-Tsung Chang. An introduction to Kriging. Material for the course Introduction to Geographic Information Systems, Western Michigan University.

[14] Sachin Chaudhari. *Spectrum sensing for cognitive radios: Algorithms, performance, and limitations*. PhD thesis, Aalto University, 2012.

[15] Sachin Chaudhari, Visa Koivunen, and H Vincent Poor. Autocorrelation-based decentralized sequential detection of ofdm signals in cognitive radios. *IEEE Transactions on Signal Processing*, 57(7):2690–2700, 2009.

[16] Noel Cressie. *Statistics for spatial data*. John Wiley and Sons, Inc, 1993.

[17] Amod V Dandawaté and Georgios B Giannakis. Statistical tests for presence of cyclostationarity. *Signal Processing, IEEE Transactions on*, 42(9):2355–2369, 1994.

[18] Pierre Delfiner et al. *Geostatistics: modeling spatial uncertainty*, volume 497. John Wiley and Sons, Inc, 2009.

[19] Peter J. Diggle and Paolo J. Ribeiro Jr. *Model Based Geostatistics*, 2007. R package version 1.7-4.

[20] FCC. Spectrum policy task force report, FCC 02-155. Technical report, FCC, 2002.

[21] Andrew O. Finley and Sudipto Banerjee. *MBA: Multilevel B-spline Approximation*, 2010. R package version 0.0-7.

[22] Andrew O. Finley and Sudipto Banerjee. *spBayes: Univariate and Multivariate Spatial-temporal Modeling*, 2013. R package version 0.3-8.

[23] Ivan Pardina Garcia. Interference management in cognitive radio systems. Master's thesis, Aalto University, School of Science and Technology, 2010.

[24] Marc G Genton. Highly robust variogram estimation. *Mathematical Geology*, 30(2):213–221, 1998.

[25] Andrea Goldsmith, Syed Ali Jafar, Ivana Maric, and Sudhir Srinivasa. Breaking spectrum gridlock with cognitive radios: An information theoretic perspective. *Proceedings of the IEEE*, 97(5):894–914, 2009.

[26] Michael A Goodrich. Potential fields tutorial. •, 2002. Class Notes, University of California, Irvine.

[27] Jan Willem Van Groenigen. The influence of variogram parameters on optimal sampling schemes for mapping by kriging. *Geoderma*, 97(3):223–236, 2000.

[28] Masaharu Hata. Empirical formula for propagation loss in land mobile radio services. *IEEE Transactions on Vehicular Technology*, 29(3):317–325, 1980.

[29] Simon Haykin. Cognitive radio: brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23(2):201–220, 2005.

[30] Simon Haykin. Cognitive dynamic systems. *Proceedings-IEEE*, 94(11):1910, 2006.

[31] Gustavo Hernández-Penaloza and Baltasar Beferull-Lozano. Field estimation in wireless sensor networks using distributed kriging. In *Communications (ICC), 2012 IEEE International Conference on*, pages 724–729. IEEE, 2012.

[32] Robert J. Hijmans. *raster: raster: Geographic data analysis and modeling*, 2014. R package version 2.2-31.

[33] ITU-R, Radiocommunication Sector of International Telecommunication Union. *Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 3 000 MHz*, October 2009. Recommendation ITU-R P.1546-4.

[34] Simon Jackman. *Bayesian analysis for the social sciences*, volume 846. Wiley, 2009.

[35] Richard Arnold Johnson and Dean W Wichern. *Applied multivariate statistical analysis*. Prentice hall Upper Saddle River, NJ, 5th edition, 2002.

[36] A. G. Journel and M. E. Rossi. When do we need a trend model in kriging? *Mathematical Geology*, 21(7), 1989.

[37] David Kahle and Hadley Wickham. *ggmap: A package for spatial visualization with Google Maps and OpenStreetMap*, 2013. R package version 2.3.

[38] Sami Kallioinen, M Vaarakangas, Ping Hui, Jani Ollikainen, Ilari Teikari, A Parssinen, Vesa Turunen, Marko Kosunen, and J Ryynanen. Multi-mode, multi-band spectrum sensor for cognitive radios embedded to a mobile phone. In *Sixth International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications*, pages 236–240. IEEE, 2011.

[39] Dan Kelley. *oce: Analysis of Oceanographic data*, 2014. R package version 0.9-13.

[40] Jarmo Lundén, Visa Koivunen, Anu Huttunen, and H Vincent Poor. Collaborative cyclostationary spectrum sensing for cognitive radio systems. *IEEE Transactions on Signal Processing*, 57(11):4182–4195, 2009.

[41] Semu Mäkinen and Markus Laatta. ENCOR2 Project Report. Technical report, Aalto University, August 2013.

[42] Robert N McDonough. *Detection of signals in noise*. Academic Press, 1995.

[43] Alessandra Menafoglio. Geostatistica per Elementi di uno Spazio di Hilbert: Teoria e Applicazioni ai Dati Funzionali. Master's thesis, Politecnico di Milano, 2011.

[44] Joseph Mitola III. *Software Radios: Wireless Architecture for the 21st Century*. Wiley-Interscience, 2000.

[45] Joseph Mitola III and Gerald Q Maguire Jr. Cognitive radio: making software radios more personal. *Personal Communications, IEEE*, 6(4):13–18, 1999.

[46] Andreas F Molisch, Larry J Greenstein, and Mansoor Shafi. Propagation issues for cognitive radio. *Proceedings of the IEEE*, 97(5):787–804, 2009.

[47] Srinivas Nallagonda, Sudheer Suraparaju, Sanjay Dhar Roy, and Sumit Kundu. Performance of energy detection based spectrum sensing in fading channels. In 2nd *International Conference on Computer and Communication Technology, 2011*, pages 575–580. IEEE, 2011.

[48] Erich Neuwirth. *RColorBrewer: ColorBrewer palettes*, 2011. R package version 1.0-5.

[49] Douglas Nychka, Reinhard Furrer, and Stephan Sain. *fields: Tools for spatial data*, 2013. R package version 6.9.1.

[50] National Land Survey of Finland. Coastline shapefiles. Online, kartat.kapsi.fi, July 2012.

[51] Jaakko Ojaniemi, Juha Kalliovara, Ahmad Alan, Jussi Poikonen, and Risto Wichman. Optimal field measurement design for radio environment mapping. Unpublished Results.

[52] Jaakko Ojaniemi, Juha Kalliovara, Jussi Poikonen, and Risto Wichman. A practical method for combining multivariate data in radio environment mapping. Unpublished Results.

[53] Jaakko Ojaniemi, Juha Kalliovara, Jussi Poikonen, Risto Wichman, and Jarkko Paavola. Resampling white space geolocation database content based on field measurement statistics. Unpublished Results.

[54] Jan Oksanen. A centralized routing and scheduling policy for cognitive radio users. Unpublished Results.

[55] Edzer J. Pebesma. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30:683–691, 2004.

[56] Edzer J. Pebesma and Roger S. Bivand. Classes and methods for spatial data in R. *R News*, 5(2):9–13, November 2005.

[57] Diggle P.J. and Ribeiro Jr P.J. *geoR: a package for geostatistical analysis R-NEWS*, 2001. R package version 1.7-4.

[58] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[59] Janne Riihijarvi, Petri Mahonen, Matthias Wellens, and Martin Gordziel. Characterization and modelling of spectrum for dynamic spectrum access with spatial statistics and random fields. In *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1–6. IEEE, 2008.

[60] Rohde & Schwartz. *Operating Manual DVB-T/H Diversity Test Receiver R & S ®TSM-DVB*.

[61] Johannes Schmitz, Milan Zivkovic, and Rudolf Mathar. Extended cyclostationary signatures for OFDM in the presence of hardware imperfections. In *7th Karlsruhe Workshop on Software Radios*, pages 124–130. Karlsruhe Institute of Technology, 2012.

[62] Shashi Shekhar, Michael R Evans, James M Kang, and Pradeep Mohan. Identifying patterns in spatial information: A survey of methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(3):193–214, 2011.

[63] Gregory Staple and Kevin Werbach. The end of spectrum scarcity. *Spectrum, IEEE*, 41(3):48–52, 2004.

[64] Rahul Tandra, Anant Sahai, and Shridhar M Mishra. What is a spectrum hole and what does it take to recognize one? *Proceedings of the IEEE*, 97(5):824–848, 2009.

[65] Waldo R. Tobler. A Computer Movie Simulating Human Growth in the Detroit Region. *Economic Geography*, 46:234–240, 1970.

[66] Rolf Turner. *deldir: Delaunay Triangulation and Dirichlet (Voronoi) Tessellation.*, 2014. R package version 0.1-5.

[67] Vesa Turunen, Marko Kosunen, Anu Huttunen, Sami Kallioinen, Petri Ikonen, A Parssinen, and J Ryynanen. Implementation of cyclostationary feature detector for cognitive radios. In *4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, pages 1–4. IEEE, 2009.

[68] Harry Urkowitz. Energy detection of unknown deterministic signals. *Proceedings of the IEEE*, 55(4):523–531, 1967.

[69] Jan Willem Van Groenigen, W Siderius, and A Stein. Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma*, 87(3):239–259, 1999.

[70] V.V.A.A. Obituary: Danie Krige, South Africa's giant of geostatistics. The Northern Miner, April 2013.

[71] Richard Webster and Margaret A Oliver. *Geostatistics for environmental scientists*. John Wiley & Sons, 2007.

[72] Matthias Wellens, Janne Riihijärvi, and Petri Mähönen. Spatial statistics and models of spectrum use. *Computer Communications*, 32(18):1998–2011, 2009.

[73] Hadley Wickham. The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29, 2011.

[74] Yonghong Zeng and Ying-Chang Liang. Robustness of the cyclostationary detection to cyclic frequency mismatch. In *IEEE 21st International Symposium onPersonal Indoor and Mobile Radio Communications (PIMRC)*, pages 2704–2709. IEEE, 2010.

[75] Yonghong Zeng, Ying-Chang Liang, Anh Tuan Hoang, and Rui Zhang. A review on spectrum sensing for cognitive radio: challenges and solutions. *EURASIP Journal on Advances in Signal Processing*, 2010:2, 2010.

[76] Qing Zhao and Brian M Sadler. A survey of dynamic spectrum access. *Signal Processing Magazine, IEEE*, 24(3):79–89, 2007.