

NON-STATIONARY SPATIAL SHADOWING MODELS FOR MICRO-CELLULAR RADIO LINKS

Julian David Villegas Gutierrez

*Thesis submitted in partial fulfillment of the requirements for
the Degree of Doctor in Science and Engineering*

October 2021

ICTEAM
Louvain School of Engineering
Université catholique de Louvain
Louvain-la-Neuve
Belgium

Thesis Committee:

Pr. Claude Oestges	Supervisor (Advisor)	UCLouvain/ICTEAM, Belgium
Pr. Jérôme Louveaux	Doctor	UCLouvain/ICTEAM, Belgium
Pr. Luc Vandendorpe	Doctor	UCLouvain/ICTEAM, Belgium

Non-stationary spatial shadowing models for micro-cellular radio links

by Julian David Villegas Gutierrez

© Julian David Villegas Gutierrez 2019

ICTEAM

Université catholique de Louvain

Place Sainte-Barbe, 2

1348 Louvain-la-Neuve

Belgium

This work was supported by the SWIPT PROJECT ARC.

Acknowledgments

I would like to thank the different people involved in the completion of my PhD at UCLouvain. To my colleagues, for their support and help that ease my stay at UCLouvain and Belgium. To my supervisor Claude Oestges for his guiding advices in wireless propagation modelling. To the friends I have made along my way in Belgium for their emotional support and reciprocity living in a different country and culture. To my family for whom I have already say what I wanted to say.

Contents

Acknowledgments	i
Table of Contents	iii
1 Introduction	1
2 Channel sounder and parameter estimation	9
2.1 Wireless channel as a linear system	9
2.2 Measurement setup	10
2.2.1 Channel sounder	10
2.2.2 Antenna arrays	13
2.3 Channel Parameter Estimation	16
2.3.1 Barlett's beamformer	20
2.3.2 Capon's spectrum	20
2.3.3 MUSIC	20
2.3.4 SAGE	21
2.4 Measurement scenario	24
2.4.1 SAGE estimates	25
2.5 Conclusion	32
3 Propagation Analysis	35
3.1 Introduction	35
3.2 Distance-based path loss	36
3.2.1 Are measurements of the same route coming from the same environment?	42
3.2.2 Degree of non-stationarity	45
3.3 Shadowing	48
3.4 LoS propagation	53
3.5 Wideband characterization	57
3.5.1 Stationary regions	57
3.5.2 Statistics of the PDP	71
3.5.3 Angular Statistics	76
3.6 Conclusion	83

4	Multi-Path Clustering	87
4.1	Introduction	87
4.2	Clustering formalism	88
4.2.1	K-means	92
4.2.2	Channel parameters physical mapping	95
4.2.3	K-Medoids	100
4.3	Hierarchical clustering	101
4.4	Conclusion	110
5	Spatial Loss Field for Outdoor Micro-cells	113
5.1	Introduction	113
5.2	Covariance estimation	114
5.3	Gaussian regression	119
5.4	Hyperparameters inference	123
5.5	Non-stationary kernels	124
5.5.1	Non-stationary gaussian kernel	128
5.5.2	Kroenecker inference	134
5.6	Shadowing correlation	135
5.7	Underlying Spatial Field	136
5.7.1	How to find $\mathbf{A}$?	139
5.7.2	Finding the underlying field	147
5.7.3	Regularization	148
5.8	Prediction	152
5.9	Conclusion	159
6	Conclusions and perspectives on extensions to this work	161

Introduction

| 1

Wireless communication systems rely on transmitting energy through the air. Without the proper characterization of such medium, source signals won't be capable of transmitting information. The study of the medium could deal with the properties of the composition of the air, although for non-satellite or land (terrestrial) communications this is not taken into account, due to the assumption that it does not constitute a major source of perturbation to the signals being treated. On the other hand, since any antenna and propagation theory is rooted at its fundamentals into Maxwell electromagnetic equations, electrical and magnetic properties are required to properly model the wireless channel. Effective material properties as the conductivity, permittivity and permeability are necessary to compute the electrical and magnetic fields.

However, to solve Maxwell equations for complex scenarios involving non-homogeneous (or non layered) medium with multiple transmitters, receivers and scatterers is a computationally cumbersome task. Besides so, depending upon the type of analysis to be done it is not always required to have a detailed knowledge about the electrical and magnetic fields at each space location and time instant.

For the case of cellular networks, Wireless Local Networks or wireless communication networks placed in urban or semi urban scenarios, most of the transmitted energy propagates through a medium composed mainly by the same components: the air, streets or roads, vegetation with different levels of leaf density, buildings, and a multiplicity of objects as light posts, stairs, etc. Thus, the medium could be considered homogeneous by regions. Or in other words, the permittivity, permeability and conductivity of the medium could be regarded as constant by region. For example, the space between objects, solely filled by air, could be modelled as homogeneous, since as said before, no big changes in the composition of the air are considered for terrestrial communications. This means that, instead of solving Maxwell equations in order to get the electrical and magnetic fields, and so, completely characterize the propagation across the medium, it is possible to get information about propagation of transmitted signals across the medium, by considering simpler models based on assumptions as the homogeneity of the medium assumption.

Radio propagation deals with radiation of electromagnetic signals whose highest frequencies are in the order of THz. From the full spectrum of ra-

radio communication systems, excluding radio stations, most of consumer used communications lie in the VHF, UHF and SHF frequency bands according to the International Telecommunication Union (ITU) division of the radio spectrum. That is to say, that the wavelength varies between 0.1 cm to 10 m. For this wavelength range, radiated energy from wireless communication networks will interact with objects larger than the size of the wavelength, allowing to use techniques as *ray tracing* that assume radio signals propagate as streams of rays, for which one can easily compute the path followed. If such an assumption and technique are valid, one gain knowledge about where is the radiated energy interacting with, at a very little computational and analytical expense.

The wireless radio communication chain is composed by a transmitter, the channel and a receiver. Both, the receiver and transmitter blocks include all the required blocks in order to send and receive radio signals. The physical entity after transmission and between reception, aka the wireless radio channel, serves as the medium for the signals to propagate. Estimation of the wireless radio channel is a cornerstone for many of the algorithms used to achieve higher capacity and higher speed data rates. With nowadays Multiple Input Multiple Output (MIMO) antenna systems, all channel information is comprised in the channel matrix $\mathbf{H} \in \mathbb{C}^{N_{RX} \times N_{TX}}$, where N_{RX} and N_{TX} are the number of antennas at the receiver and transmitter, respectively. Thus, channel estimation would require to estimate somehow such a matrix.

While estimation could be done by transmitting a signal of reference and other calibration processes, to model the wireless radio channel in such a way that no estimation is required is an advantage for signal processing algorithms requiring such an input. Since the environments through which radio signals propagate can be extremely different, an unique general model for the wireless radio channel is impossible. Rather than that, channel modelling efforts are put into different models capturing different propagation environments that are representative of the most important propagation environments commonly used. Such a modelling could be divided into multiple categories, each one characterized by a different methodology:

- Analytical channel models which explore how to use matrix theory in order to model a general channel matrix $\mathbf{H}$. The goal of this kind of models is to provide the more general analytical tools that could be used in any specific propagation environment analysis. For example, they provide propagation-focused interpretations of the spectral decomposition of the channel matrix, or how information of the eigenmodes of the matrix are related with propagation characteristics as the scattering richness of the channel, and so on.
- Empirical-based channel models, make use of extensive measurement

campaigns to produce models that fit such measurements. These models are very site-specific and are not necessarily generalizable for even similar environments but different measurement settings, as the central frequency of operation of the system, bandwidth of the system, etc.

- **Deterministic channel models.** Although, previous quoted models also assume the channel matrix is deterministic under some conditions for which such a model applies, under the category of deterministic channel models fall computational electromagnetics and ray tracing techniques, that provide a detailed characterization of the physical propagation of EM waves, whose level of detail depends on the level of detail of the geographical and geometrical characterization of the scenario/environment.
- **Stochastic channel models.** These models provide either a statistical characterization of the channel matrix $\mathbf{H}$, or of specific aspects of propagation as small-scale perturbations of the received signal, or of the geometrical characteristics of the environment as the placement of the scatterers and transceivers. These type of channel models could be seen as a middle point between the generality that analytical models want to achieve and the specific-site information provided by empirical-based channel models. Furthermore, stochastic channel models rely in information extracted from measurements as, for example, dispersion characteristics as the mean delay, angular spread, etc., capable of characterize multiple types of environments, and not only the specific environment where measurements took place.

However, wireless channel modelling doesn't end there. In a broader sense, wireless channel modelling deals not only with the MIMO channel matrix, but also with the information obtained from such a matrix. This is because there is much information contained in the MIMO channel matrix. With this in mind, the task of modelling the wireless channel can be stated in a two-fold general way: on one side, its purpose points to enhance the understanding of the physics of the communications systems involved, such that from such understanding better communication schemes can be designed; and, on the other hand, it tackles a more pragmatic purpose that is to evaluate the performance of a deployed technology in a realistic environment, or in other words, to measure the response of the deployed wireless technology under realistic scenarios for which a measurement planning stage is required. The final goal of any model developed for the wireless channel or signals extracted from it, is to be general, simple and accurate. However, the challenge in the modelling process is that there are trade-offs between those goals (i.e. simpler models often imply less accuracy).

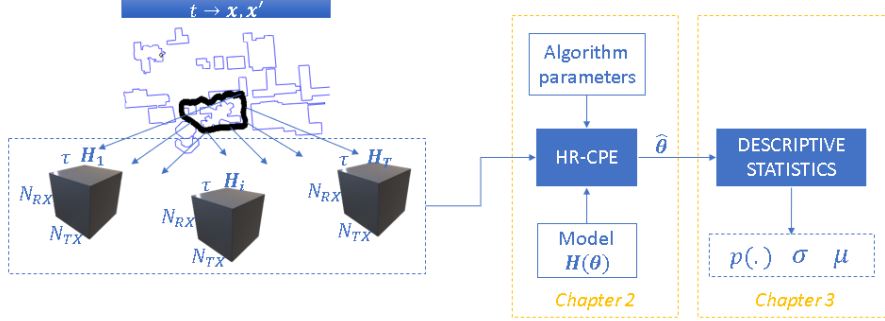


Figure 1.1: Work-flow followed through all this document. The first components of the work-flow are presented in this Figure.

The wireless channel modelling task produces models that, ideally, characterize signals describing physical properties of the channel (i.e. time-averaged received power), by using parameters that could be seen as high-level propagation characteristics of the environment. It is meant by high-level, parameters defined by the model that are not directly tied with a physical quantity, and so, imply a 'higher-level' of abstraction. Among the signals of interest in the wireless propagation community is the shadowing. It is typically believed that shadowing accounts for losses due to obstacles, although scarce models have been proposed that relate the shadowing signal with the geometry of the scenario and so, with the obstacles placement. Furthermore, shadowing has been widely considered as a random variable which ultimately does not expose any connection with propagation characteristics of the environment. The purpose of this thesis is to extend the available models for shadowing by relating it with propagation characteristics (i.e. propagation paths) that make more clear the link between shadowing and the losses due to obstacles that shadowing is supposed to account for. In order to tackle such goal, the work-flow followed all along this thesis is shown in Figures 1.1, 1.2, 1.3.

In Figure 1.1, the Channel Impulse Response (CIR) tensors are taken from measurements carried out at specific transmitter-receiver locations $\mathbf{x}, \mathbf{x}'$ or time snapshots t . Those CIR tensors feed a High-Resolution Channel Parameter Estimation (HR-CPE) algorithm whose goal is to obtain the set of parameters θ that characterize the model $H(\theta)$ for the wireless channel. The parameter estimates $\hat{\theta}$ are used to compute descriptive statistics and other propagation information that summarize characteristics of the outdoor scenario. Such information is useful to validate estimated parameters of models of large-scale propagation losses (as the path loss) where the environment is similar to the one considered in this thesis. Furthermore, such parameters might be used for simulation tools of the wireless channel.

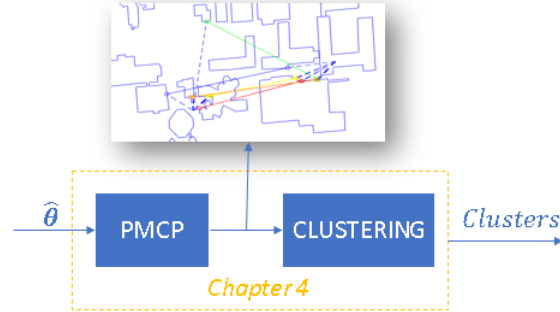


Figure 1.2: Work-flow followed through all this document. The second components of the work-flow are presented in this Figure.

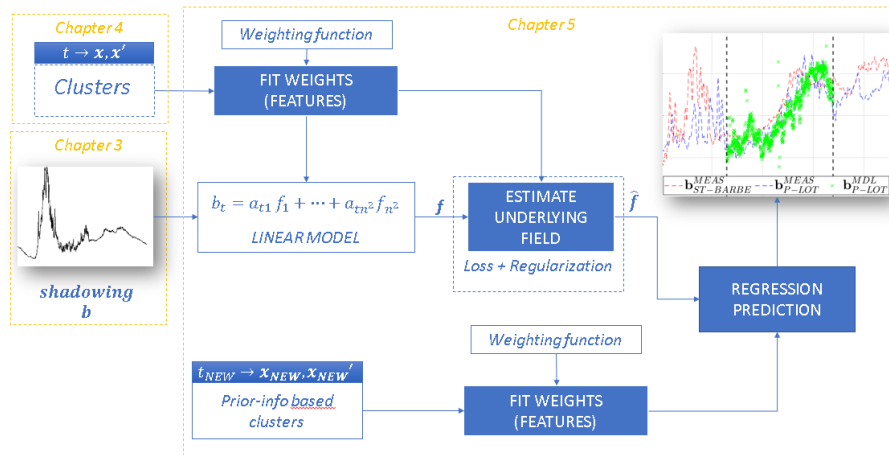


Figure 1.3: Work-flow followed through all this document. The third components of the work-flow are presented in this Figure.

After the previous steps, the parameter estimates $\hat{\theta}$ are clustered as shown in Figure 1.2. Such clustering task is divided into two steps: a first one where Physical Mapping of Channel Parameters (PMCP) is done, resulting in propagation paths; and a second one, where actual clustering is performed on the set of propagation paths. This step seeks to reduce the number of channel parameter estimates, such that instead of considering all the propagation paths, the centroids of the clusters of propagation paths are used for the following step. It is an intermediate step in the work-flow whose purpose is to obtain the centroids of the clusters of propagation paths, that are required for the shadowing model proposed in the next step.

Finally, Figure 1.3 shows the last part of the work-flow. Here, the clusters of propagation paths (their centroids) serve as an input for the computation

of the coefficients of the linear model proposed for the shadowing. Both, computation and meaning of these coefficients $a_{t1}, \dots, a_{tn^2}$, and what is called *weighting function* in Figure 1.2 are given in Chapter 5. The other component of the model, the underlying field $\mathbf{f}$ is estimated for the propagation paths computed from specific transmitter-receiver locations. More detailed explanation about the meaning of these underlying field idea is presented, as well, in Chapter 5. In order to make predictions for the shadowing at new locations $\mathbf{x}_{NEW}, \mathbf{x}'_{NEW}$, prior information is assumed for the computation of the propagation paths. In other words, no measurement information is used to compute the propagation paths of the new locations, and so, the shadowing predictions.

To conclude this chapter, the contributions of this thesis are summarized in what follows:

- Large scale parameter extraction is followed by an analysis of how such large scale parameters, as the path loss exponent, behave at different sections of the measurement routes under consideration in Chapter 3. Computed delay and directional mean and dispersion statistics characterize the type of propagation environment where measurements took place. Computed values could be used as reference values for propagation environments similar to the one described in Chapter 2
- Non-stationary regions for each measurement route and a new metric for the non-stationarity of the overall measurement route are found in Chapter 3. A discussion about the information this non-stationary regions provide about the mechanisms of propagation is also presented there.
- A new approach to map the channel parameter estimates to the physical scenario is provided in Chapter 4. This approach is followed by a new clustering problem which overcomes the interpretation problems of classical parameter space clustering.
- New model for the shadowing correlation is presented in Chapter 5. The theoretical framework for the model is presented as well as its challenges.
- New weighting function based on the clusters found in Chapter 4 is presented in Chapter 5. This weighting function extends weighting functions used in the literature of underlying field estimation (better known as Radio Tomography Imaging).
- Empirical characterization of the Active Paths Matrix introduced in Chapter 5. This matrix is itself a new tool of characterizing measure-

ments by providing a more direct level of information about the propagation paths and reflection bounces than the Channel Impulse Response measured matrices.

- New approach to characterize the shadowing correlation in a non-stationary way in Chapter 5. The goal of this type of approach is to be able to extend isotropic and anisotropic shadowing models usually available in the literature, such that shadowing correlations for specific locations are revealed.
- Based on the ideas of the underlying field and Active Paths Matrix, predictions for shadowing at locations close to the locations used as the reference to compute the underlying field are done in Chapter 5. The methodology used to predict the shadowing is an useful tool that might be tested over extensive measurement campaigns in different propagation environments.

Channel sounder and parameter estimation

2

This chapter introduces the channel models that will be considered in this thesis, the measurement equipment that will be used, and the algorithms that can be used to extract the channel model parameters from measurements. After describing the measurement campaign set-up (antennas choose, scenario, routes followed by the mobile terminal, etc.), channel parameter estimation algorithms are presented in order to show the parameters of the channel model, extracted from one of those algorithms. A first set of angular, delay and power metrics obtained from the channel parameter estimates is provided.

2.1 Wireless channel as a linear system

From a system point of view inputs passing through the wireless channel are transformed into outputs. Thus, the wireless channel acts as an operator. In its actual nature, this type of system is non-linear, time-variant and stochastic. However, under some circumstances it can be characterized as a Linear Time-Invariant (LTI) system. If transmitted signals are considered only in the time-domain, this relationship between an input and an output is written as

$$y(t) = x(t) * h(t) \quad (2.1)$$

or equivalently, in the frequency domain

$$Y(f) = X(f)H(f) \quad (2.2)$$

with operator $*$ denoting convolution, $h(t)$ the *impulse response* function, and $H(f)$ the transfer function. By considering small-enough time intervals, during which physical or statistical properties of the channel remain unchanged, the time invariant property of the system can be guaranteed. The randomness of the channel is essential to some analytical radio channel models, that assume the wireless channel fits into a number of categories for which random variables could be used to model it: rayleigh channel, rician channel, double rayleigh channel, among others. Furthermore, the randomness of the channel is the cornerstone of present day Multiple Input Multiple

Output (MIMO) systems, that try to exploit such a phenomena to increase throughput or capacity.

The linear assumption arises from the fact that in long-enough distances spherical waves could be seen locally as planar waves. This is true if the distance between the transmitter and the receiver is $d_{TX-RX} \gg \frac{2D^2}{\lambda}$, with D being the largest size in the antenna array. As a reference, at 3.8 GHz and a maximum $D = 10$ cm, the far field zone is reached after 2.5 m. Thus, the linear assumption arising from the planar wave assumption is valid if the mobile device (the transmitter) moves a small distance compared to the distance to reflectors or scatterers [Ekm02]. This classical assumption due to the seminar works by Clarke [Cla68] and Jakes [Jak74], has also a measurement-based explanation. Electrical field is measured at an antenna indirectly through the voltage and has a frequency, spatial and time dependence. For a single transmitted tone $\exp(j2\pi ft)$, the open circuit voltage at the measured antenna can be approximated by a space dependent function at the tone frequency $H(f, \mathbf{r})$ multiplied by the complex sinusoid $\exp(j2\pi ft)$. This is a traditional way to uncouple space-time dependence from frequency dependence for the electromagnetic fields [Bal12].

All modelling efforts try to effectively describe $H(f, \mathbf{r})$. Modern communications use transmission systems that use more complex signals than a single tone, but the strong assumption that the voltage at the receiving antenna can be regarded as in equations 2.1 or 2.2 remains almost unchanged. Once this assumption is accepted, the purpose of wireless channel modelling is then two-fold:

1. Design ways to measure the transfer function $H(f, \mathbf{r})$.
2. Propose mathematical models that fit a measured $H(f, \mathbf{r})$.

This chapter deals with each of this steps in sections 2.2 and 2.3.

2.2 Measurement setup

Analyses used throughout this document are based in a measurement campaign carried out in Louvain-la-Neuve, Belgium. Figure 2.7 shows a top view of the scenario and principal areas of interest.

2.2.1 Channel sounder

The transfer function $H(f, \mathbf{r})$ can be rewritten as $H(f, t)$, where the position vector $\mathbf{r}$ is parametrized as $\mathbf{r}(t) = [x(t), y(t), z(t)]$, with x , y , and z being the axes of the rectangular coordinate system in $\mathbb{R}^3$. The time parameter t is tied to position. By Fourier-transforming $H(f, \mathbf{r})$ with respect to the frequency

variable, the impulse response function $h(\tau, t)$ is get. The new variable τ also represents time, but at a different scale: within the planar wave assumption, it represents the time it takes a wave-front to reach the receiver antenna. Assuming air propagation, such that waves propagate close to light speed, the delays τ are less than $35 \mu\text{s}$ for travelled distances less than 10 km.

Systems where the available bandwidth is small-enough can't resolve wave-fronts arriving at different time delays τ . For such systems, called narrow-band, τ -dependency is dropped and were used at early measurement [Cla68] to sense the transfer function $H(\mathbf{r})$, by transmitting single tone Continuous Wave (CW) signals, due to their mathematical tractability. The same principle of operation is used by Vector Network Analysers (VNA), but instead of transmitting a single tone, they transmit a sweep of tones, sense the S-parameters of the network (more precisely the transmission S-parameter) and fourier-transform with respect to the frequency to get the impulse response function. An usual problem with this approach is that the measurement time due to the sweep is high, which in turn prevent this methodology to detect location $\mathbf{r}$ variations (or time t variations). Many of early measurement campaigns relied on the CW methodology [Tur+72], [OJ64], [CL75]. An alternative used to avoid such problem was to transmit short duration repetitive pulses [VR87]. However, high peak-to-mean power ratios at the transmitter result in higher undesirable power requirements.

On the other hand, spread spectrum systems (used at least back in the 80's [Dev86]), are based on a different approach. A carrier is modulated by a pseudo-random noise sequence, whose autocorrelation function is a triangular pulse. Due to this autocorrelation property, the correlation of the received signal with the sent code is a good approximation of the impulse response of the channel [Xia92]. The code is short compared to the rate of change of the radio channel, in order to achieve channel time-invariance during the measurement phase. This process is repeated each T_{IRF} seconds, so for $i = 1, 2, \dots, T$ snapshots [Dzi+99] of the channel are measured. The period impulse response time T_{IRF} has to be greater than the maximum difference between time delays (maximum excess delay), in order to avoid overlaps between consecutive transmitted signals.

The wireless channel measurements used along this document were taken using the Proposound channel sounder by Elektrobit. The architecture of the system is shown in Figure 2.1. System operates between carrier frequencies 3.4 to 4.2 GHz. After modulation with the pseudo-random sequence at the TIU block, the signal is up-converted in the TRU unit. Similar process is done at the receiving side: down-conversion in the RRU unit, and demodulation at the RIU unit. To be able to recover time-delay variable τ synchronization between transmitter and receiver ends is crucial. For indoor scenarios this synchronization can be done by connecting both ends with a wire, but for

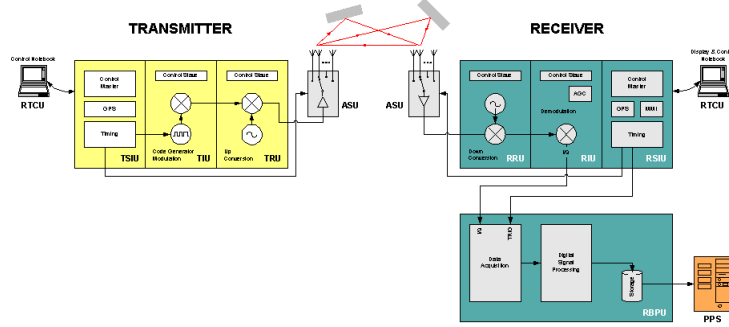


Figure 2.1: Propsound channel sounder system by Elektrobit.

outdoor measurements, where this is not possible, a calibration step is required for the high-precision clocks to be synchronized. Antenna arrays are incorporated into the architecture through use of switching systems at both the transmitter and receiver ends. This avoids the need for expensive and heavy Radio Frequency (RF) chains for each antenna element, besides the calibration difficulties. Instead of that, both transmitter and receiver switch through each of the antenna elements, following the timing concept exposed in Figure 2.2. It means that a complete channel impulse response (CIR) is obtained only after codes have been sent through each and all transmitter antennas and sensed through each and all receiver antennas. This means that the total time to get a CIR is above: $T_{IRF} \geq 2\tau_{MAX}N_{TX}N_{RX}$, with τ_{MAX} being the maximum excess delay, and N_{TX} , N_{RX} the number of transmitter and receiver antennas, respectively. The factor 2 comes from the propagation guard (PG) at the receiver side, used to avoid switching transients [Lan08].

Channel invariance remains for a short period of time. In order to account for changes in the channel, the period between modulated pseudo-random signals must be short enough to be able to capture such movements. Time-evolution of the wireless radio channel could be explained as coming from two main factors: transceivers and scatterers mobility. Considering one of the antennas to be mobile, with speed v_m , the minimum frequency f_{IRF} at which the channel has to be sampled is [Niz14]:

$$f_{IRF} = \frac{1}{T_{IRF}} \geq \frac{v_m + 2v_s}{\lambda} N_\lambda \quad (2.3)$$

with v_s the maximum speed of the scatterers, λ the wavelength, and N_λ the desired number of channels per wavelength.

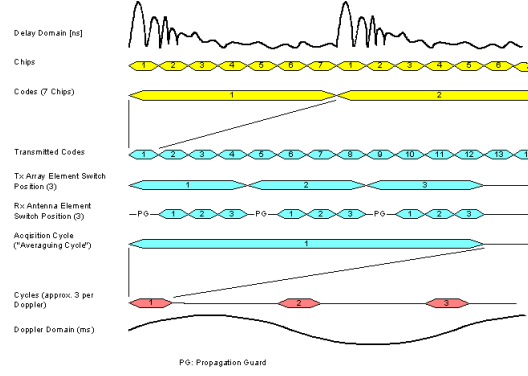


Figure 2.2: Propsound channel sounder timing concept.

2.2.2 Antenna arrays

The complex radiation pattern in the far field for an antenna is a function of the frequency f , azimuth angle θ , elevation angle ϕ and polarization. It is measured through a calibration process with proper propagation conditions. To reduce the interference of non desired energy and interactions with scatterers, the calibration measurement takes place in an anechoic chamber following this guidelines [Lan08]

1. Place the Antenna Under Test (AUT) at the origin of the reference system.
2. Place the Antenna of Reference (AOR) at a known and fixed distance r .
3. For each antenna array element and polarization, the calibration step provides a measurement y related with the radiation pattern b_{AUT} of the AUT through

$$y(f, \theta, \phi) = b_{AOR}(f)H(f)G_{SYS}(f)b_{AUT}(f, \theta, \phi) \quad (2.4)$$

with $H(f)$ being the channel transfer function, G_{SYS} the measurement system gain at measured frequency f , $b_{AOR}(f)$ the known reference antenna radiation pattern in its beam direction.

4. For an anechoic chamber conditions, it is reasonable to assume that only Line-of-Sight transmission occurs between both antennas, and so from Friis's equation in the frequency domain:

$$H(f) = \frac{\exp(-j2\pi f\tau(r))}{4\pi f\tau(r)} \quad (2.5)$$

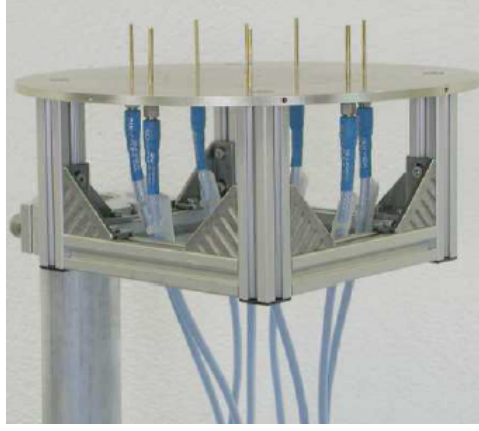


Figure 2.3: Photo of the UCA used.

with $\tau(r)$, the delay corresponding to the distance r between both antennas.

5. This means the measured radiation pattern is obtained from:

$$b_{AUT}(f, \theta, \phi) = \frac{y(f, \theta, \phi) 4\pi f \tau(r) \exp(j2\pi f \tau(r))}{b_{AOR}(f) G_{SYS}(f)} \quad (2.6)$$

Two antenna arrays were used: an Uniform Circular Array (UCA), shown in Figure 2.3, composed by 8 vertically polarized dipoles (7 around the circumference and 1 at the center) with a nominal 5 dBi gain per element; and a 2×4 dual-polarized planar array with orthogonal polarizations at $\mp \frac{\pi}{4}$ rads and array gain of 6 dBi approximately. Figures 2.5 and 2.4 show the measured radiation patterns of each antenna at the central frequency 3.8 GHz. The UCA was used as the mobile station (mobile terminal) and placed in a moving trolley at a 2 m height. On the other hand, the planar array was placed in the top of Maxwell and Mercator buildings at 15 - 25 m with respect to the lowest terrain of the considered scenario surrounding the Mercator building.

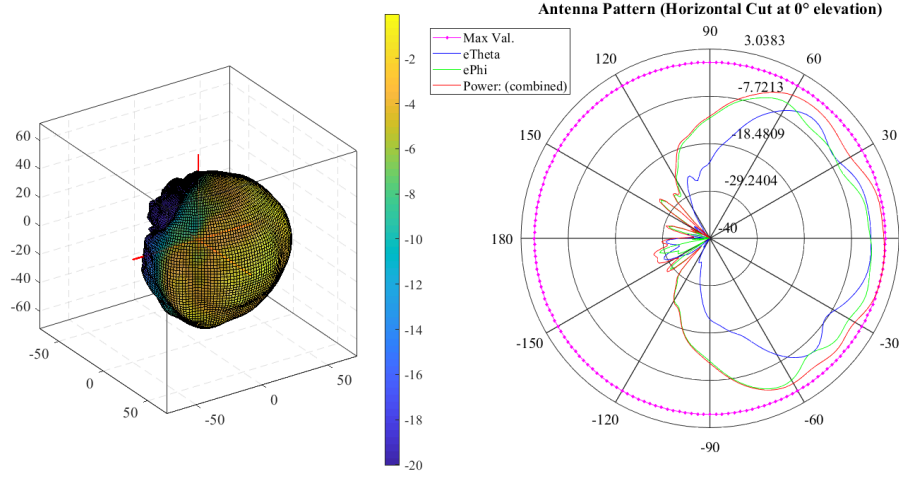


Figure 2.4: Calibrated radiation pattern of 1 antenna element of the double-polarized. Left panel: 3D radiation pattern in rectangular coordinates for the vertical polarization, with the colorbar representing normalized power in dB. Right panel: 2D measured radiation pattern for the azimuth (at elevation 0 rad). The curves eTheta and ePhi represent $\pm \frac{\pi}{4}$ rad orthogonal polarizations.

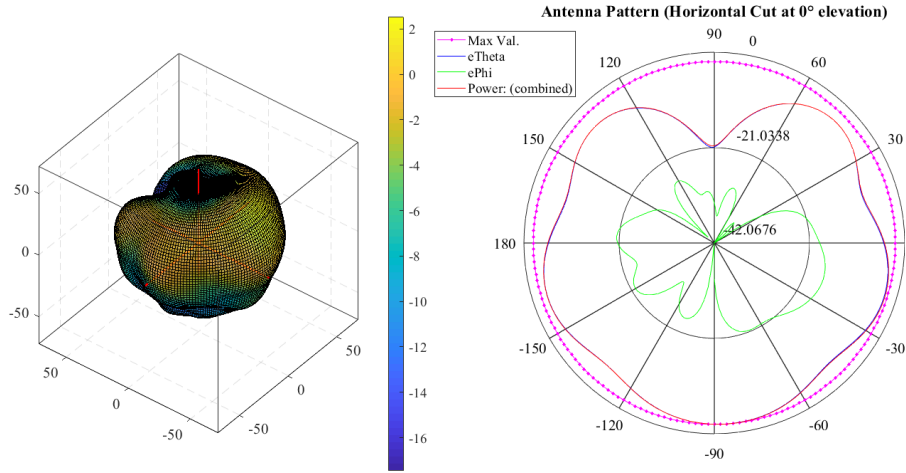


Figure 2.5: calibrated radiation pattern of 1 antenna element of the UCA. Left panel: 3D radiation pattern in rectangular coordinates for the vertical polarization, with the colorbar representing normalized power in dB. Right panel: 2D measured radiation pattern for the azimuth (at elevation 0 rad). The curve named eTheta represent the vertical polarization measurement, while the curve named ePhi represent the horizontal.

2.3 Channel Parameter Estimation

The wireless radio channel transfer function $H(\tau, t)$ can be modelled as a superposition of planar waves arriving from different directions and after travelling different time-delays. Let us consider baseband signals. For a narrowband signal $X(f)$ sent through 1 antenna and a received signal $Y(f)$, received at 1 antenna, the channel transfer function can be approximated by to be [Ric05]

$$\frac{Y(f)}{X(f)} = H(f) = \alpha G_{RX}(f) G_{TX}(f) \exp(j2\pi f \tau) \exp(-j2\pi f_c \tau) \quad (2.7)$$

where τ denotes the time-delay the signal takes to reach the receiver, f_c the central frequency, $G_{RX}(f)$ the receiver system transfer function, $G_{TX}(f)$ transfer function, and α the gain introduced by the physical wireless radio channel. Previous expression considers any of the receiver, transmitter or objects with which the energy interacts static. If any of these moves, a time-variant version of 2.7

$$H(f, t) = \alpha(t) G_{RX}(f) G_{TX}(f) \exp(j2\pi f \tau) \exp(-j2\pi f_c (\tau + \frac{v}{c} t)) \quad (2.8)$$

where c denotes the speed of light, and v the effective speed of all the moving objects with respect to the path through which energy propagates. All v, α, τ are assumed to not change within a short interval of observation $t \ll \frac{1}{vB}$, with B being the bandwidth of the system. Equation 2.8 is rewritten to introduce the Doppler-shift $f_D = -\frac{v}{c} f_c$, and to combine the channel complex gain with the phase shift $\exp(-j2\pi f_c \tau)$ into the new complex gain $\alpha'(t) = \alpha(t) \exp(-j2\pi f_c \tau)$

$$H(f, t) = \alpha'(t) G_{RX}(f) G_{TX}(f) \exp(j2\pi f \tau) \exp(-j2\pi f_D t) \quad (2.9)$$

It is well-known that energy transmitted doesn't reach the receiver as a superposition of planar waves. Electrical field is rather than a combination of discrete waves a continuum. Furthermore, diffuse scattering phenomena arising from non-smooth surfaces compared to the wavelength doesn't behave in such a way. Therefore, models for $H(f, t)$ usually split it into a Specular Component (SC) and a Dense Multipath Component (DMC)

$$H(f, t) = H_{SC}(f, t) + H_{DMC}(f, t) \quad (2.10)$$

where the variable H_{DMC} groups all the non-specular phenomena, and DMC is preponderant among them. It is assumed that such variable is dependent on frequency and time. In outdoor propagation usually the SC dominates the power received, since energy interacts with objects that are bigger than the wavelength. For SC propagation, if we drop time-invariance assumption for the channel complex gain, and gather system transfer functions into an effective gain $\alpha^{eff}(t) = \alpha'(t)G_{TX}(f)G_{RX}(f)$, we have that for L plane waves

$$H_{SC}(f, t) = \sum_{l=1}^L \alpha_l^{eff}(t) \exp(-j2\pi f \tau_l) \exp(-j2\pi f_{D,l}t) \quad (2.11)$$

The effective gain α^{eff} include system responses and is the actual gain that the channel sounder measures: transmitter and receiver systems are included or compensated in the measured radio channel gain.

So far, the antennas considered are omnidirectional. To include directivity, we multiply by the transmitter $b_{TX}(\theta_{TX}, \phi_{TX})$ and receiver $b_{RX}(\theta_{RX}, \phi_{RX})$ antenna radiation patterns. Angles θ_{RX}, ϕ_{RX} refer to the azimuth and elevation at which the incoming wave reaches the receiver, and are called Azimuth of Arrival (AoA) and Elevation of Arrival (EoA). Similarly, angles θ_{TX}, ϕ_{TX} refer to the azimuth and elevation at which most of the radiated energy departs from the transmitter, and are called Azimuth of Departure (AoD) and Elevation of Departure (EoD). Angular conventions are depicted in Figure 2.6.

$$H_{SC}(f, t) = \sum_{l=1}^L b_{TX}(\theta_{TX,l}, \phi_{TX,l}) b_{RX}(\theta_{RX,l}, \phi_{RX,l}) \alpha_l^{eff}(t) \exp(-j2\pi f \tau_l) \exp(-j2\pi f_{D,l}t) \quad (2.12)$$

Such a model, including discrete directions of departure and arrival is called a double-directional channel model [Ste98], [SMB01]. In the case of directional channel models, the channel is considered to behave in such a way that it itself is causing energy to depart and arrive at certain directions. This is a way of assuming the channel, at least locally, due to its geometry, terrain, materials of the Interacting Objects (IO), etc., drives propagation in certain ways, regardless of the antenna array. Such a channel is called a *propagation channel* to differentiate it from the *radio channel*. The latter concept assumes that antenna arrays drive energy into specific locations and so, for each antenna array energy is driven in a different way.

On the other hand, the following vectors comprising radiation patterns and complex gains for both polarizations are introduced

$$\mathbf{b}(\theta, \phi) = [b_1(\theta, \phi) \quad b_2(\theta, \phi)], \quad \boldsymbol{\alpha}^{eff}(t) = \begin{bmatrix} \alpha_{1,1}^{eff}(t) & \alpha_{1,2}^{eff}(t) \\ \alpha_{2,1}^{eff}(t) & \alpha_{2,2}^{eff}(t) \end{bmatrix} \quad (2.13)$$

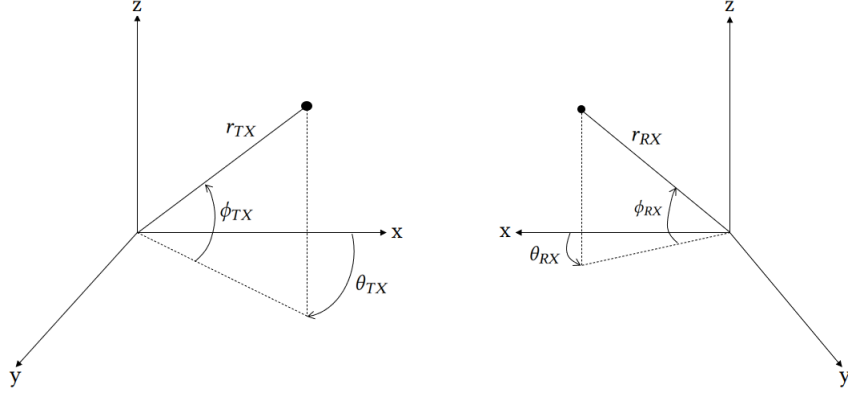


Figure 2.6: Rectangular reference coordinate system. Angular notation is depicted.

where $b_i(\theta, \phi)$ represents the radiation pattern for polarization i , and $\alpha_{i,j}^{eff}$ is the effective channel gain for transmitter polarization i and receiver polarization j . Then the SC transfer function becomes

$$H_{SC}(f, t) = \sum_{l=1}^L \mathbf{b}_{RX}(\theta_{RX,l}, \phi_{RX,l}) \alpha_l^{eff}(t) \mathbf{b}_{TX}^T(\theta_{TX,l}, \phi_{TX,l}) \exp(-j2\pi f \tau_l) \exp(-j2\pi f_D l t) \quad (2.14)$$

Considered systems so far are Single Input Single Output (SISO). To include multiple antennas in the receiver and the transmitter side (MIMO), radiation pattern matrices for each antenna array are introduced

$$\mathbf{c}_{RX}(\theta_{RX}, \phi_{RX}) = \begin{bmatrix} b_{RX,1,1}(\theta_{RX}, \phi_{RX}) & b_{RX,1,2}(\theta_{RX}, \phi_{RX}) \\ b_{RX,2,1}(\theta_{RX}, \phi_{RX}) & b_{RX,2,2}(\theta_{RX}, \phi_{RX}) \\ \vdots & \vdots \\ b_{RX,N_{RX},1}(\theta_{RX}, \phi_{RX}) & b_{RX,N_{RX},2}(\theta_{RX}, \phi_{RX}) \end{bmatrix} \quad (2.15)$$

$$\mathbf{c}_{TX}(\theta_{TX}, \phi_{TX}) = \begin{bmatrix} b_{TX,1,1}(\theta_{TX}, \phi_{TX}) & b_{TX,1,2}(\theta_{TX}, \phi_{TX}) \\ b_{TX,2,1}(\theta_{TX}, \phi_{TX}) & b_{TX,2,2}(\theta_{TX}, \phi_{TX}) \\ \vdots & \vdots \\ b_{TX,N_{TX},1}(\theta_{TX}, \phi_{TX}) & b_{TX,N_{TX},2}(\theta_{TX}, \phi_{TX}) \end{bmatrix} \quad (2.16)$$

where $b_{i,j,k}$ refers to $i = \{RX, TX\}$ antenna number $j = \{1, \dots, N_i\}$ with polarization $k = \{1, 2\}$. We define, as well, $\mathbf{c}_{i,j}$ with $i = \{RX, TX\}$ and $j = \{1, 2\}$ as the array response for polarization 1 or 2: first or second column of $\mathbf{c}_i$. For planar waves, this array responses are also known as *steering vectors*. Now equation 2.14 is rewritten as

$$\mathbf{H}_{SC}(f, t) = \sum_{l=1}^L \mathbf{c}_{RX}(\theta_{RX}, \phi_{RX}) \boldsymbol{\alpha}_l^{eff}(t) \mathbf{c}_{TX}^T(\theta_{TX,l}, \phi_{TX,l}) \exp(-j2\pi f \tau_l) \exp(-j2\pi f_{D,l} t) \quad (2.17)$$

with $\mathbf{H}_{SC} \in \mathbb{C}^{N_{RX} \times N_{TX}}$. The channel impulse response is obtained after applying an inverse Fourier transform with respect to frequency f to the transfer function

$$\mathbf{H}_{SC}(\tau, t) = \sum_{l=1}^L \mathbf{c}_{RX}(\theta_{RX,l}, \phi_{RX,l}) \boldsymbol{\alpha}_l^{eff}(t) \mathbf{c}_{TX}^T(\theta_{TX,l}, \phi_{TX,l}) \delta(\tau - \tau_l) \exp(-j2\pi f_{D,l} t) \quad (2.18)$$

with $\delta(\cdot)$ being the Dirac delta function. The purpose of any estimation algorithm would be to find the $6L$ - dimensional parameter vector (also called structural parameter vector) $\boldsymbol{\mu}_{SC} = [\tau_1, \dots, \tau_L, \theta_{TX,1}, \dots, \theta_{TX,L}, \theta_{RX,1}, \dots, \theta_{RX,L}, \phi_{TX,1}, \dots, \phi_{TX,L}, \phi_{RX,1}, \dots, \phi_{RX,L}, f_{D,1}, \dots, f_{D,L}]$ and the $4L$ -dimensional weights vector $\boldsymbol{\gamma}_{SC} = [\alpha_{1,1,1}^{eff}, \dots, \alpha_{L,1,1}^{eff}, \alpha_{1,1,2}^{eff}, \dots, \alpha_{L,1,2}^{eff}, \alpha_{1,2,1}^{eff}, \dots, \alpha_{L,2,1}^{eff}, \alpha_{1,2,2}^{eff}, \dots, \alpha_{L,2,2}^{eff}]$ that minimizes a given loss function relating the parameters and the channel model. Both vectors are found for each and all the instants of time t under consideration. Different methodologies are used to solve the problem. They were classified in [KV96] as static or dynamic, depending upon time-evolution is taken into account or not. Among the static methodologies there are the classical spectral methods (Barlett's and Capon's spectrum), Subspace-based methods (MUSIC, ESPRIT [Haa+95]) and Maximum Likelihood (ML) methods (SAGE [FH94], [Fle+99], [FJS02] [Wan+17], RIMAX [Ric05], [RLT03]).

In most of the methodologies mentioned a channel spatial covariance matrix is required. The channel covariance matrix exhibits correlations between different channels: that is to say, a propagation channel using a different pair of antenna array element at both ends. Different spatial correlations can be computed: at the transmitter, receiver or at both sides:

$$\mathbf{R}_{TX} = \mathbb{E}[\mathbf{H}^T (\mathbf{H}^T)^*], \quad \mathbf{R}_{RX} = \mathbb{E}[\mathbf{H} \mathbf{H}^*], \quad \mathbf{R}_{FULL} = \mathbb{E}[\text{vec}(\mathbf{H}) \text{vec}(\mathbf{H})^*] \quad (2.19)$$

here, $\mathbf{H} \in \mathbb{C}^{N_{RX} \times N_{TX}}$ is the wireless channel matrix, including all kinds of propagation phenomena, which is observed during measurements. Spatial

correlation matrices play an important role in signal wireless channel analysis and signal processing for radio channels. For instance, analytical channel models, deal with the MIMO channel matrix $\mathbf{H}$ as a pure mathematical entity by modelling $\mathbf{R}$.

2.3.1 Barlett's beamformer

This methodology maximizes the SNR if spatial white noise is considered in a given direction ϕ [Bar53], [Bar48]. Barlett's beamformer finds produces an Angular Power Spectrum (APS) for which the peaks are the AoA, AoD, EoA, EoD. For a double-directional channel, azimuth plane (all elevation angles are 0) and considering only one polarization at the transmitter and one polarization at the receiver, the Barlett's APS is

$$APS_B(\theta_{RX}, \theta_{TX}) = (\mathbf{b}_{RX}(\theta_{RX}) \otimes \mathbf{b}_{RX}^*(\theta_{RX}))^T \mathbf{R} (\mathbf{b}_{TX}(\theta_{TX}) \otimes \mathbf{b}_{TX}^*(\theta_{TX})) \quad (2.20)$$

where $\mathbf{R}$ is the full spatial correlation matrix in 2.19 and $(\cdot)^*$ is the transpose and conjugate operator.

2.3.2 Capon's spectrum

Also called the Minimum Variance Distortionless Response Filter (MDVR), Capon's spectrum reduces the sidelobes introduced by Barlett's spectrum, while reducing the width of the peaks in the APS. This is done by projecting the steering vectors into the inverse of the channel spatial covariance matrix. This is an improvement respect Barlett's spectrum, since closer sources than in Barlett's case are able to be detected: angular resolution is increased [CGK67], [Cap69]. For a double-directional channel, azimuth plane (all elevation angles are 0) and considering only one polarization at the transmitter and one polarization at the receiver, the Capon's APS is

$$APSC(\theta_{RX}, \theta_{TX}) = \frac{1}{(\mathbf{b}_{TX}(\theta_{TX}) \otimes \mathbf{b}_{RX}(\theta_{RX}))^* \mathbf{R}^{-1} (\mathbf{b}_{TX}(\theta_{TX}) \otimes \mathbf{b}_{RX}(\theta_{RX}))} \quad (2.21)$$

2.3.3 MUSIC

Estimated spatial channel correlation $\hat{\mathbf{R}}$ introduces noise into the actual spatial channel correlation

$$\hat{\mathbf{R}} = \mathbf{R} + \sigma^2 \mathbf{I} \quad (2.22)$$

where $\mathbf{I}$ is identity matrix of the same dimensions as $\mathbf{R}$, and σ^2 the noise power. The spectral decomposition of the estimated matrix can be split into 2 components:

$$\hat{\mathbf{R}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^* = \mathbf{U}_{SIG}((\mathbf{\Lambda}) + \sigma^2\mathbf{I})\mathbf{U}_{SIG}^* + \sigma^2\mathbf{U}_{NOISE}\mathbf{U}_{NOISE}^* \quad (2.23)$$

This splits the eigen-decomposition $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$ into a signal subspace (composed by eigenvectors $\mathbf{U}_{SIG}$) and a noise subspace (composed by eigenvectors $\mathbf{U}_{NOISE}$). Such a separation is possible if the rank of the signal subspace is known beforehand. In other words, if the number of paths are known beforehand. The Multiple Signal Classification (MUSIC) [Sch82], [Sch86] spectrum idea is that the steering vector of a pair of direction of arrival and departure contribute to the signal subspace, and so, is orthogonal to the noise subspace. This makes the denominator in 2.24 zero and the spectrum at such pair of directions infinite. Non-orthogonal separation of subspace makes the denominator close to zero and noticeable peaks in the MUSIC spectrum [Wei03]. For a double-directional channel, azimuth plane (all elevation angles are 0) and considering only one polarization at the transmitter and one polarization at the receiver, the MUSIC's APS is

$$P_{MUSIC}(\phi_{TX}, \phi_{RX}) = \frac{\|\mathbf{b}_{TX}(\phi_{TX})\|_2^2 \|\mathbf{b}_{RX}(\phi_{RX})\|_2^2}{(\mathbf{b}_{TX}(\theta_{TX}) \otimes \mathbf{b}_{RX}(\theta_{RX}))^* \mathbf{U}_{NOISE} \mathbf{U}_{NOISE}^* (\mathbf{a}_{TX}(\phi_{TX}) \otimes \mathbf{b}_{RX}(\phi_{RX}))} \quad (2.24)$$

2.3.4 SAGE

It is common to model $\mathbf{H}_{DMC}$ as a complex gaussian random variable with a given covariance. This assumes that for all non-specular phenomena $\mathbf{H}_{DMC} \sim \mathcal{CN}(0, \mathbf{R}_{N-SC})$. Let us define $\mathbf{h}_{SC} \in \mathbb{C}^{N_{RX} N_{TX} \times 1} = \text{vec}(\mathbf{H}_{SC})$ and $\mathbf{h} \in \mathbb{C}^{N_{RX} N_{TX} \times 1} = \text{vec}(\mathbf{H})$. Combining this with 2.10, the observed channel distributes as $\mathbf{h} \sim \mathcal{CN}(\mathbf{h}_{SC}, \mathbf{R}_{N-SC})$. Thus, the log-likelihood is proportional to

$$\log p(\mathbf{h} \mid \boldsymbol{\mu}_{SC}, \boldsymbol{\gamma}_{SC}, \mathbf{R}_{N-SC}) \propto -(\mathbf{h} - \mathbf{H}_{SC})^* \mathbf{R}_{N-SC} (\mathbf{h} - \mathbf{H}_{SC}) \quad (2.25)$$

Space Alternating Generalized Expectation-Maximization (SAGE) algorithm finds the parameter and weight vectors for which 2.25 is maximized. Due to the computationally-prohibitive grid search solution for a $6L$ -dimensional log-likelihood, for a given iteration SAGE fixes some of the parameters while maximizing the function with respect to the rest of parameters. Two steps are

involved in the process: an Expectation step and a Maximization step. During the Expectation step, a function for the expectation of the log-likelihood is create with the actual parameter estimates. During the Maximization step, the function created in the Expectation step is used to find the parameters that maximize the function, which become the new estimates [LN07]. The algorithm ensures that the successive log-likelihood estimates are non-decreasing.

Under the SAGE algorithm the model assumed for the channel is a superposition of L multipath components during the window of time under consideration as the one considered for the specular channel in (2.14). If the sent signal is $u(\tau)$, the received signal is

$$\mathbf{y}(\tau) = \sum_{l=1}^L \mathbf{s}(\tau, \boldsymbol{\mu}_{SC,l}) + \sqrt{\frac{N_0}{2}} \mathbf{N}(\tau) \quad (2.26)$$

where the signal per path is

$$\mathbf{s}(\tau, \boldsymbol{\mu}_{SC,l}) = \mathbf{c}_{RX}(\theta_{RX}, \phi_{RX}) \boldsymbol{\alpha}^{eff} \mathbf{c}_{TX}^T(\theta_{TX}, \phi_{TX}) u(\tau - \tau_l) \quad (2.27)$$

N_0 is the noise power, $\mathbf{N}$ is a standard complex gaussian vector of size $N_{RX}N_{TX} \times 1$, and $\boldsymbol{\mu}_{SC,l}$ is the parameter vector for propagation path l . An observable data defined as

$$\mathbf{x}_l = \mathbf{s}(\tau, \boldsymbol{\mu}_{SC,l}) + \varsigma_l \sqrt{\frac{N_0}{2}} \mathbf{N}(\tau) \quad (2.28)$$

is required for the expectation step. Here, ς_l are coefficients such that the vector $\{\varsigma_l \mathbf{N}(\tau)\}_{l=1}^L$ forms a decomposition of $\mathbf{N}(\tau)$ [Vin17]. The complete data $y(\tau)$ is the superposition of the observable data $\mathbf{y}(\tau) = \sum_{l=1}^L \mathbf{x}_l$. Authors in [FJS02], [Yin+03] derived the solution for the parameter vector $\boldsymbol{\mu}_{SC,l}$ which maximizes its log-likelihood given an estimate for the observable data $\mathbf{x}_l$. Such a solution finds the argument which maximizes the function z

$$z(\boldsymbol{\mu}_{SC,l}, \mathbf{x}_l) = \mathbf{f}^H \mathbf{D}^{-1} \mathbf{f} \quad (2.29)$$

with

$$\mathbf{f} = \begin{bmatrix} \mathbf{c}_{RX,1}^* \mathbf{X}_l \mathbf{c}_{TX,1} \\ \mathbf{c}_{RX,1}^* \mathbf{X}_l \mathbf{c}_{TX,2} \\ \mathbf{c}_{RX,2}^* \mathbf{X}_l \mathbf{c}_{TX,1} \\ \mathbf{c}_{RX,2}^* \mathbf{X}_l \mathbf{c}_{TX,2} \end{bmatrix} \quad (2.30)$$

and

$$\mathbf{D} = [\mathbf{c}_{RX}^* \mathbf{c}_{RX}] \otimes [\mathbf{c}_{TX}^* \mathbf{c}_{TX}] \quad (2.31)$$

All the dependences on the parameter vector $\mu_{SC,l}$ are omitted for better readability. In (2.30) the $N_{RX} \times N_{TX}$ matrix $\mathbf{X}_l$ has as its (i, j) entry the quantity

$$[\mathbf{X}_l]_{i,j} = \int_{T_0} u^*(t - \tau_l) x_{l,i,j} d\tau \quad (2.32)$$

where T_0 is the window of time under consideration or time snapshot, and $x_{l,i,j}$ corresponds to the entry of the observable data $\mathbf{x}_l$ when transmitting from antenna j and receiving at antenna i . Previous equation doesn't take into consideration the doppler effect. Once obtained the parameter vector $\hat{\mu}_{SC,l} = \arg \max z(\mu_{SC,l}, \mathbf{x}_l)$, the gain coefficients α^{eff} are obtained from

$$\alpha^{eff} = \frac{1}{P} \mathbf{D}^{-1} |_{\mu_{SC,l}=\hat{\mu}_{SC,l}} \mathbf{f} |_{\mu_{SC,l}=\hat{\mu}_{SC,l}} \quad (2.33)$$

with $P = \int_{T_0} u(\tau) d\tau$ as the transmitted power over the window T_0 . As said previously, SAGE obtains an estimate for the parameter vector by performing two steps. In the first step, an estimate for the observable data $\mathbf{x}_l(\tau, \mu_{SC,l})$ is obtained by subtracting the signals of all paths (except the one under consideration) to the complete signal $\mathbf{y}(\tau)$

$$\hat{\mathbf{x}}_l = \mathbf{y}(\tau) - \sum_{k=1, k \neq l} \mathbf{s}(\tau, \mu_{SC,k}) \quad (2.34)$$

In the second step the problem of maximizing z with respect to $\mu_{SC,l}$ (excluding the channel parameters $f_{D,l}$) is reduced to multiple maximizations problems:

- Given an initial estimate $\hat{\mu}_{SC,l}$, the time delay is obtained by fixing all other channel parameters and maximizing z with respect to τ : $\tau_l = \arg \max z(\tau_l, \mathbf{x}_l)$
- An estimate for $\theta_{RX,l}$ is obtained by fixing all the channel parameters, while using the recently delay estimate, and maximizing z with respect to $\theta_{RX,l}$: $\theta_{RX,l} = \arg \max z(\theta_{RX,l}, \mathbf{x}_l)$
- An estimate for $\theta_{TX,l}$ is obtained by fixing all the channel parameters, while using the recently delay estimate and $\theta_{RX,l}$, and maximizing z with respect to $\theta_{TX,l}$: $\theta_{TX,l} = \arg \max z(\theta_{TX,l}, \mathbf{x}_l)$
- An estimate for $\phi_{RX,l}$ is obtained by fixing all the channel parameters, while using the recently delay estimate, $\theta_{RX,l}$ and $\theta_{TX,l}$, and maximizing z with respect to $\phi_{RX,l}$: $\phi_{RX,l} = \arg \max z(\phi_{RX,l}, \mathbf{x}_l)$

- An estimate for $\phi_{TX,l}$ is obtained by fixing all the channel parameters, while using the recently delay estimate, $\theta_{RX,l}$, $\theta_{TX,l}$ and $\phi_{RX,l}$, and maximizing z with respect to $\phi_{TX,l}$: $\phi_{TX,l} = \arg \max z(\phi_{TX,l}, x_l)$

This two steps are called the SAGE iteration step. A SAGE iteration cycle performs the SAGE iteration step for all the assumed propagation paths L . The benefit of the SAGE approach is that, starting from an initial estimate for $\mu_{SC,l}$ it is guaranteed that sequence produced by evaluating the log-likelihood at each estimate μ_{SC} is non-decreasing. It is necessary thus, to obtain an initial estimate for the channel parameter vector. From [Yin+03], the first estimate $\mathbf{x}_l^0$ for $\mathbf{x}_l$ is

$$\mathbf{x}_l^0 = \mathbf{y}(\tau) - \sum_{k=1}^{l-1} \mathbf{s}(\tau, \mu_{SC,k}) \quad (2.35)$$

The first estimate τ_l^0 for the delay is the delay corresponding to the maximum of the power delay profile

$$\tau_l^0 = \arg \max \left\| \int_{T_0} u^*(\tau - \tau_l) \mathbf{x}_l^0 d\tau \right\|_2^2 \quad (2.36)$$

Both initial estimates are used to find the initial estimates $\theta_{TX,l}^0, \phi_{TX,l}^0$ for the directions of departure

$$\theta_{TX,l}^0, \phi_{TX,l}^0 = \arg \max \|\mathbf{X}_l(\tau_l^0, \mathbf{x}_l^0) \mathbf{c}_{TX}\|_2^2 \quad (2.37)$$

where $\mathbf{X}_l(\tau_l^0, \mathbf{x}_l^0)$ is computed by substituting τ_l and $x_{l,i,j}$ by τ_l^0 and the $N_{RX}(i-1) + N_{TX}$ entry of $\mathbf{x}_l^0$ in (2.32). Finally, the first estimate $\theta_{RX,l}^0, \phi_{RX,l}^0$ for the directions of arrival are found by substituting all computed initial estimates in z function (2.29) and maximizing it with respect to $\theta_{RX,l}, \phi_{RX,l}$

2.4 Measurement scenario

Measurement campaigns were carried out in Louvain-la-Neuve, Belgium. There were 2 static Base Stations (BS): one on top of Mercator and the other on top of Maxwell building. For each BS there were 3 routes considered: around Saint Barbe building (here called, Saint Barbe route), around Marie Curie and Pierre Curie buildings (here called, Marie Curie route), and around the Parking Lot below Saint Barbe building (here called, Parking Lot route). For each route, there was 1 or more measurements (trials) done.

Information regarding the measurement campaign set-up is displayed in Table 2.1. The terrain is not regular: there are three noticeable ramps, besides the fact that Mercator building is at a lower terrain height than the rest of buildings highlighted in Figure 2.7. Transmitter antenna height is around 2

Table 2.1: Measurement setup

Code Length	5.11 μs
Samples per Code (N_r)	2044
CIR sampling frequency (f_{IRF})	30.2 Hz
Center frequency	3.7 GHz
Bandwidth	200 MHz

m. For the Maxwell antenna, its height is above all the buildings (at an approximate height of 15 m). For the Mercator antenna, the height is comparable to some of the buildings, since the building is at a lower terrain.

A spatial reference system (SRS) is composed by a coordinate system, a *datum* and a type of projection. *Datums* are used to model the curvature of the earth and the types of projections are used to convert from a 3D location in the earth to a 2D location. The Belgium Lambert 72 coordinate system (EPSG code 31370) is been used for the 2D projections. A number of GPS coordinates were available from the channel sounder for each measurement route. Since there were less GPS coordinates than IRF measurements, linear interpolation between available coordinates was used, assuming constant speed for the transmitter, in order to match the number of IRF for each route. Figure 2.7 show a top view of the scenario, and Figures 2.8 and 2.9 show the transmitter coordinates for all measurements. For some of the graphs, anomalies could be seen regarding transmitter movement (a trajectory might come in and out from a building contour). This is due to the GPS system used. We tried to be as loyal as possible to the GPS coordinates measured by the channel sounder system. All coordinates in an inner loop of a route were removed.

2.4.1 SAGE estimates

SAGE estimate parameters μ_{SC} and γ_{SC} are computed for the measurement (trial) 2 of Saint Barbe route with Maxwell BS and the only measurement (trial) taken for Saint Barbe route with Mercator BS. The maximum number of propagation paths per snapshot L were 20 for the Maxwell BS and 27 for the Mercator BS. Typical values for L range from 20 to 50 [FJS02], [Czi+08b], [Fle+02]. The value is chosen so that it can account for the most dominant paths in power. Although not done, model selection criteria as Akaike Information (AIC) or Bayesian Information (BIC) can be used to choose L , by, for example, penalizing models with high number of parameters, and so, high L . Elevation angles of arrival and departure are estimated, in order to have a good fit between the model and the measurements than, but only azimuthal propagation is considered throughout this document for posterior analysis. In other words, only estimated azimuths of arrival and departure are used

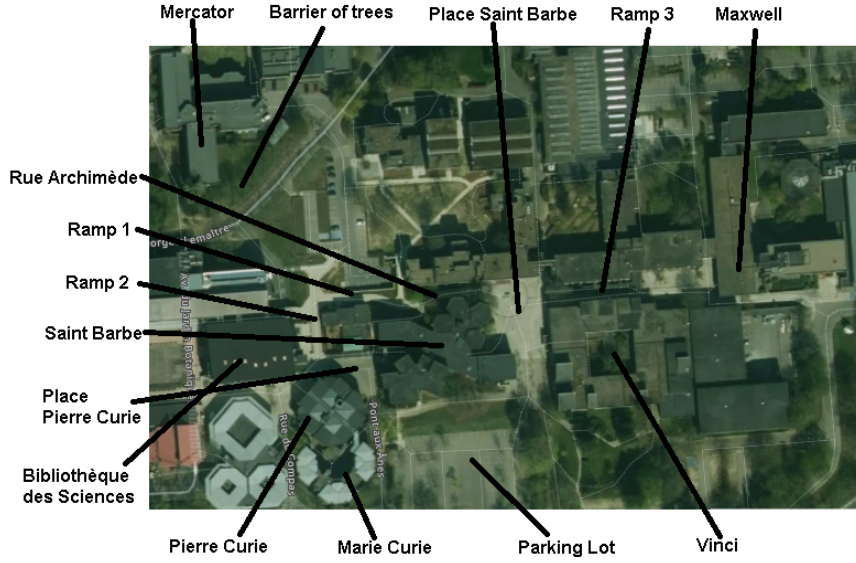


Figure 2.7: Top view of buildings surrounding transmitter routes.

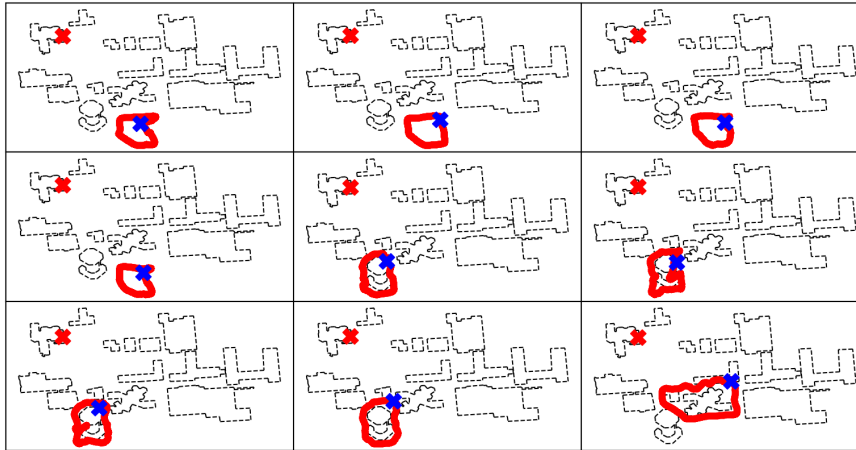


Figure 2.8: Top view of transmitter measurement routes for Mercator BS. The BS location, start of the route are indicated by the red cross and blue cross, respectively. By convention we designate graph in the top left as the graph (1,1). The convention used is: (row number, column number). Graph (1,1): Parking Lot route. Graph (1,2): Parking Lot route. Graph (1,3): Parking Lot route. Graph (2,1): Parking Lot route. Graph (2,2): Marie Curie route. Graph (2,3): Marie Curie route. Graph (3,1): Marie Curie route. Graph (3,2): Marie Curie route. Graph (3,3): Saint Barbe route.

for all the forthcoming analysis where angular information is required. The reference for the angles obtained from SAGE is set by considering the max-

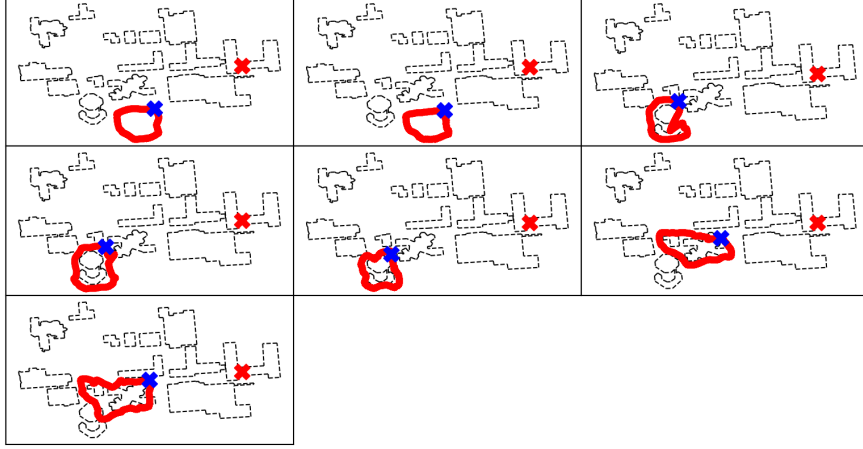


Figure 2.9: Top view of transmitter measurement routes for Maxwell BS. The BS location, start of the route are indicated by the red cross and blue cross, respectively. By convention we designate graph in the top left as the graph (1,1). The convention used is: (row number, column number). Graph (1,1): Parking Lot route. Graph (1,2): Parking Lot route. Graph (1,3): Marie Curie route. Graph (2,1): Marie Curie route. Graph (2,2): Marie Curie route. Graph (2,3): Saint Barbe route. Graph (3,1): Saint Barbe route.

imum power gain per snapshot. For a pair of transmitter-receiver locations, the line connecting them is considered the LoS. We denote the length of that line (its distance) by d_{LoS} . The absolute reference for the angles is the x-axis, longitude coordinate or equatorial line. The angle associated with the LoS is considered to be the AoD for which the maximum power gain is attained.

Figures 2.10 and 2.11 show the empirical Probability Density Functions (PDF) for all the AoA and AoD with each BS. Most probable AoAs are coming from a direction pointing to the center of the trajectory, while most probable AoD are pointing to the location of the BS. The size of the mean direction, called in circular statistics R or the mean vector length, is by itself a measure of circular dispersion. It can be seen that for both BS the AoA have a larger mean vector length, meaning that most of the AoA are coming from the same direction. This is expected since propagation arriving from directions opposite to the transmitter location is extremely unlikely due to losses by longer paths. On the other hand, for the AoD the mean vector length is lower, indicating angles departure from multiple directions with non-negligible probabilities. This is expected for the UCA, since its radiation pattern is nearly omnidirectional.

Instead of showing directly the distribution of the time delays obtained from SAGE, it is suitable to be able to compare those delays with a known delay or a delay that gives information about the scenario. For doing so we

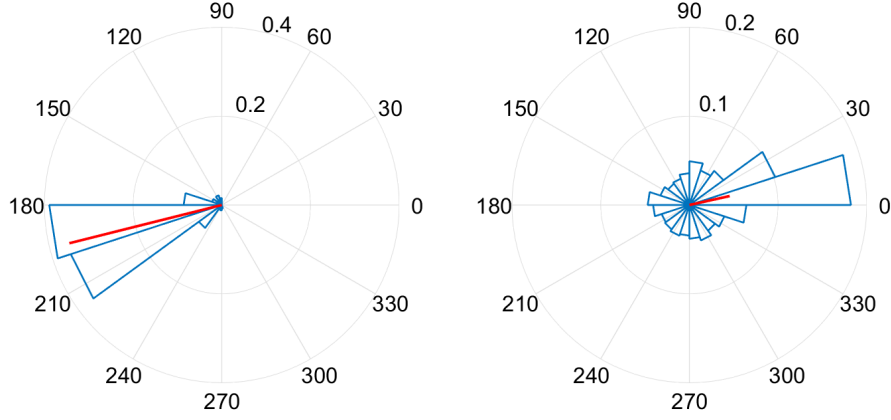


Figure 2.10: Empirical PDF of the azimuths of arrival and departure for the Saint Barbe route with Maxwell BS. Left panel: circular histogram for the AoA. Right panel: circular histogram for the AoD. The red solid line represents the mean direction. The histograms are built from all the time snapshots.

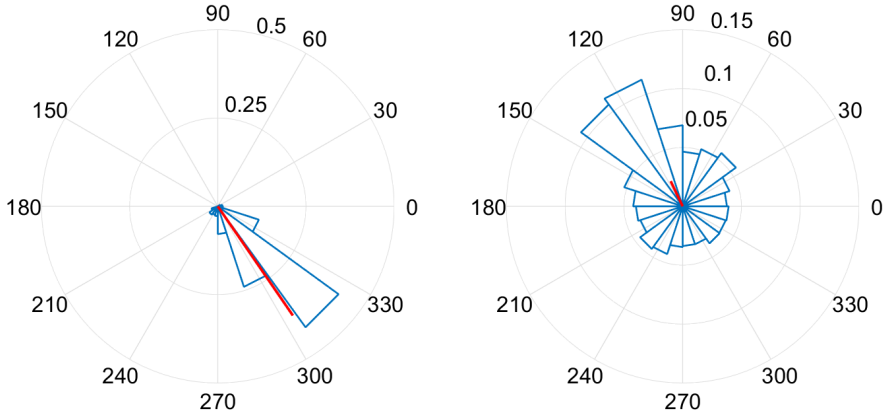


Figure 2.11: Empirical PDF of the azimuths of arrival and departure for the Saint Barbe route with Mercator BS. Left panel: circular histogram for the AoA. Right panel: circular histogram for the AoD. The red solid line represents the mean direction. The histograms are built from all the time snapshots.

define the parameter $N_{EXC-LoS}$ as

$$N_{EXC-LoS} = \frac{\tau_{SAGE}}{\frac{d_{LoS}}{c}} \quad (2.38)$$

where c is the speed of light and τ_{SAGE} are the time delays estimated by SAGE for each and all the time snapshots. This is giving an idea of how much

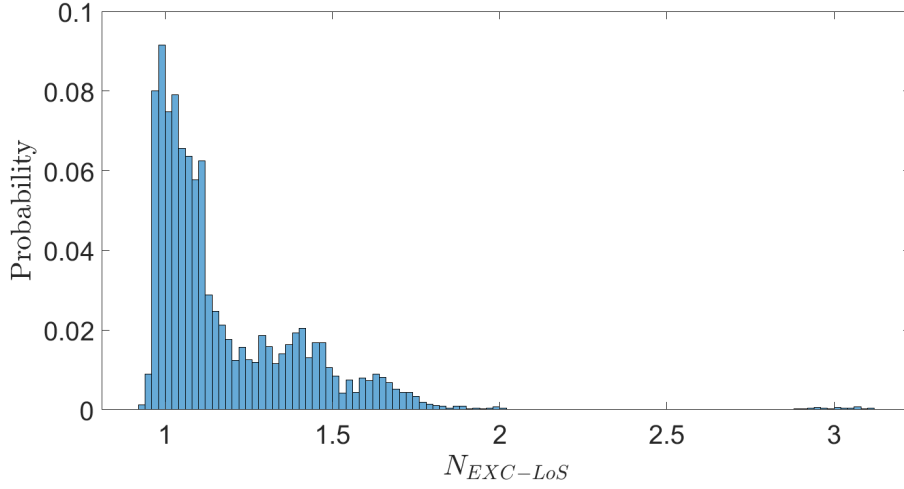


Figure 2.12: Empirical PDF of $N_{EXC-LoS}$ for the Saint Barbe route measurement with Maxwell BS.

longer are the propagation paths compared to the case when the propagation path is the LoS, and thus, how many reflections are occurring: the longer $N_{EXC-LoS}$, the longer the path, and the higher the probability there are multiple reflections. Figures 2.12 and 2.13 show the empirical PDFs of $N_{EXC-LoS}$ for each BS. It can be seen that most of the propagation paths for the Maxwell BS are similar in length to the LoS, indicating LoS is either a mechanism of propagation, or there are reflections that don't deviate too much from the LoS. This means there are no far IOs, since more than 99% of the propagation paths are shorter than twice the LoS of the corresponding pair of transmitter-receiver locations. On the other hand, for the Mercator BS, the propagation paths tend to be longer, and there are no LoS conditions. The mode of the propagation paths is around 2.8 times the LoS distance. Terrain height differences between transmitter and receiver locations, even if not taken into account since only azimuthal propagation is considered, have an impact on the length of the propagation paths. A closer look at Figure 2.12 shows $N_{EXC-LoS}$ values lower than 1. These values, indicating a lower estimated delay than the minimum possible delay, are an artefact caused by the fact that the transmitter and receiver coordinates provided by the GPS device are not always accurate.

It is possible to reconstruct the CIR from equation 2.17 by using the SAGE estimates. The CIR is averaged over all the MIMO channels, and the square of the magnitude computed, such that we have an instantaneous Power Delay Profile (PDP): $PDP(t, \tau) = \sum_{i,j} |[\mathbf{H}_{SC}(t, \tau)]_{i,j}|^2$. The PDP is a measure of the power per delay bin. Lets drop the SC index for better readability, and denote by $PDP_{TRUE}(t, \tau)$ and $PDP_{RECON}(t, \tau)$ the instantaneous PDP obtained by

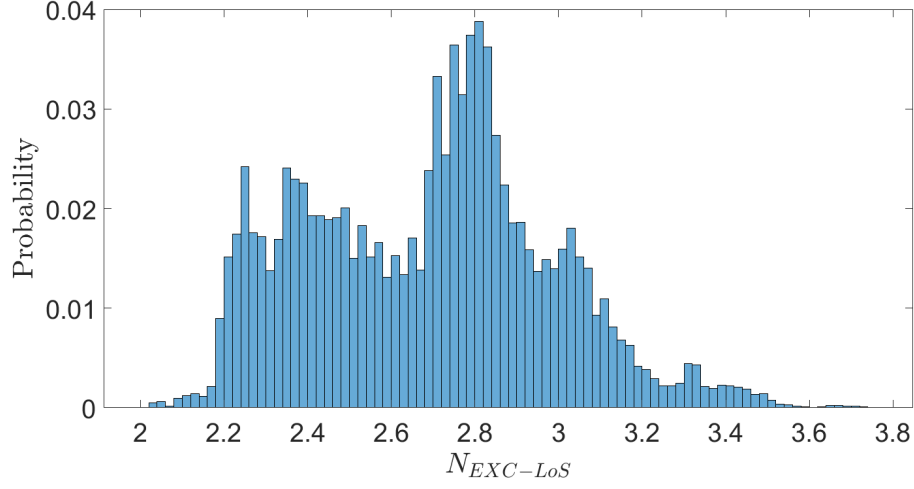


Figure 2.13: Empirical PDF of $N_{EXC-LoS}$ for the Saint Barbe route measurement with Mercator BS.

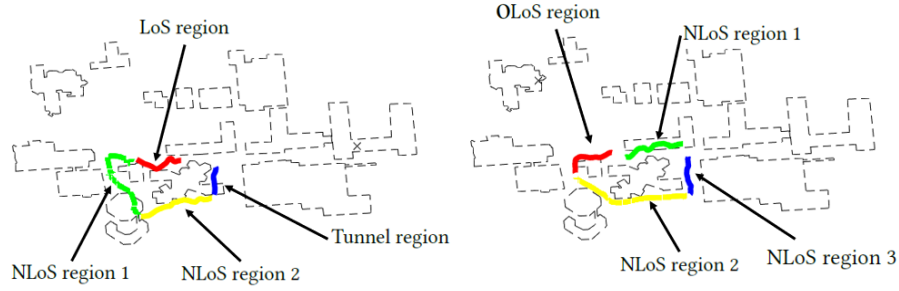


Figure 2.14: Regions of propagation for the Saint Barbe route measurements for which SAGE channel parameters were estimated. Left panel: Maxwell BS. Right panel: Mercator BS. Each propagation region is indicated in the figure and highlighted with a different color.

using the measured and the reconstructed CIR, respectively. The Saint Barbe route is divided into 4 propagation regions depicted in Figure 2.14. For each of these, the instantaneous PDP is averaged, producing an estimated PDP.

$$PDP(\tau) = \sum_{t \in \mathcal{PR}} PDP(t, \tau) \quad (2.39)$$

where the set $\mathcal{PR}$ contains the snapshots associated with each propagation region. Figures 2.15 and 2.16 show the reconstructed and true CIR, as well as the residual $R(\tau)$

$$R_0(\tau) = PDP_{TRUE}(\tau) - PDP_{RECON}(\tau), \quad R(\tau) = |R_0(\tau)| \quad (2.40)$$

Table 2.2: Power proportions for the propagation regions in Saint Barbe route.

Region	BS MXW		BS MKT	
	SC (%) ^a	DMC (%) ^b	SC (%)	DMC (%)
LoS/OLoS	99.5	0.5	70	30
NLoS 1	29	71	54	46
NLoS 2	20	80	53	47
Tunnel/NLoS 3	87	13	72	28

^a Specular power percentage of the total power P_{TRUE} .

^b DMC power percentage of the total power P_{TRUE} .

This residual indicates how much of the radiated power reaches the receiver through specular reflection or not. The bigger the residual, the lower the match between the SAGE estimates and the model, and so, the lower the impact of specular reflection in the radio channel. Furthermore, let's define the total, specular reflection and DMC power for a region as P_{TRUE} , P_{SC} and P_{DMC} in (2.41). The proportion between SC or DMC power and the total power is shown in Table 2.2 for both BS.

$$P_{TRUE} = \sum_{\tau} PDP_{TRUE}(\tau), \quad P_{SC} = \sum_{\tau} PDP_{RECON}(\tau), \quad P_{DMC} = \sum_{\tau} R_0(\tau) \quad (2.41)$$

For the Maxwell BS, the LoS and the tunnel regions the residual is low, while specular reflections account for 99% and 87% of the total power. This is expected for the LoS, but initially comes as a surprise for the tunnel region. However, it agrees with the assumption that within this zone multiple reflections occur similarly to those happening in a tunnel. The residual is high for both NLoS regions, as well as the DMC power proportion (between 70 - 80% of total power), indicating that rather than individual strong reflections from smooth or large physical objects (compared to the wavelength), power is arriving at the receiver after impinging in *many* small objects or rough surfaces with weak power.

For the Mercator BS, surprisingly, specular reflections account for 72% of the propagation in the NLoS 3 region. This doesn't mean that the power coming from such region is high, but rather, that there are few strong-enough paths accounting for the power received. For the regions NLoS 1 and NLoS 2, specular reflection also has half of the share of the total power. As it is expected, for the OLoS most of the power (70%) comes from specular reflections.

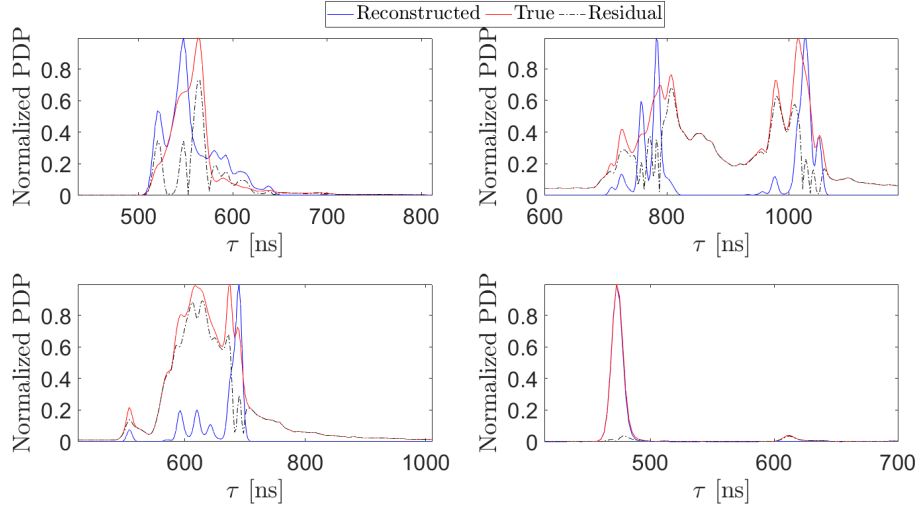


Figure 2.15: Comparison between measured and reconstructed PDP for the Saint Barbe propagation regions in Figure 2.14 with Maxwell BS. Up left: LoS region. Up right: NLoS 1 region. Bottom left: NLoS 2 region. Bottom right: tunnel region.

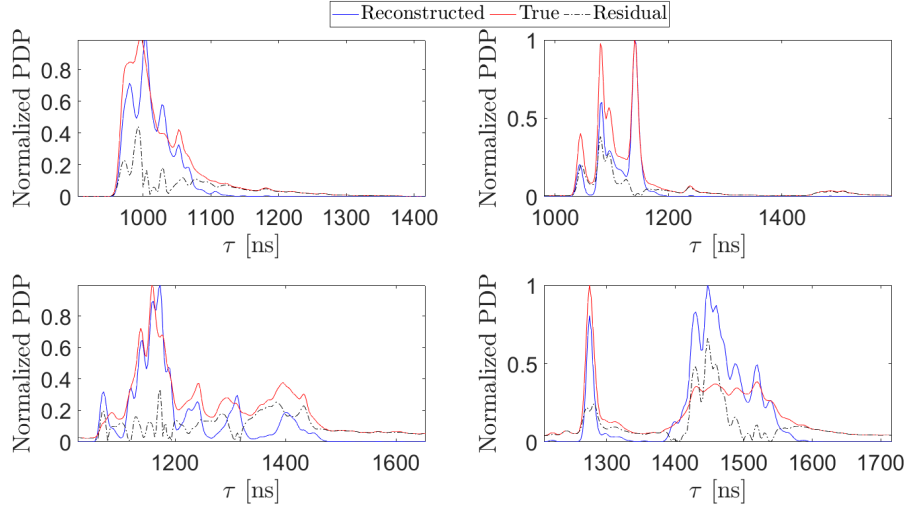


Figure 2.16: Comparison between measured and reconstructed PDP for the Saint Barbe propagation regions in Figure 2.14 with Mercator BS. Up left: OLoS region. Up right: NLoS 1 region. Bottom left: NLoS 2 region. Bottom right: NLoS 3 region.

2.5 Conclusion

After presenting the channel models with their parameters, probability distribution functions for the azimuths of arrival and departure was presented.

A delay metric was defined in order to show the distribution of the delay with respect to the delay associated to the LoS. Power proportions have been computed for each of the propagation regions introduced in this chapter and used in posterior chapters of this thesis. All these probability distribution results have been shown as a way to summarize the estimated channel parameters that will be used as an input of the clustering step in the work-flow presented in the Introduction chapter of this document.

Propagation Analysis

3

This chapter presents large-scale propagation analysis for the measurement campaign introduced in Chapter 2. Distance-based and shadowing losses are computed, as well as the parameters of the models that characterized both types of losses. Due to actual time-variant characteristics of the channel, regions where a criteria for stationary applies are found. A metric to measure the degree of non-stationarity of the received power is introduced as a way to provide a single-value measure of non-stationarity. The sizes and locations of the stationary regions founded are linked to the propagation mechanisms that are occurring at them. Angular statistics are computed for the estimated azimuths of arrival and departure. All these computations are done as separated components of an analysis that seeks to extract information about where and how propagated rays or plane waves interact with a scenario with similar characteristics to this scenario. In that sense, this chapter steps aside from the main work-flow presented in Chapter 1. Its purpose is to provide a measurement-based analysis of a real outdoor micro-cell measurement campaign that might be serve for comparison or validation with measurements with similar characteristics.

3.1 Introduction

Path losses account for power ratios between two points of a communication system. They are defined in general in (3.1).

$$L[dB] \doteq 10 \log \left(\frac{P_T}{P_R} \right) \quad (3.1)$$

Here, P_T and P_R represent the power transmitted and received, seen from 2 given points in a communication system. For instance, P_T might refer to the power transmitted *before* transmission line losses or after those losses, as well as before or after transmission losses and antennas gains, and so on. Equivalently, P_R represents the power available at reception at a given 'cut' of the communication system.

Path losses as defined in (3.1) account for all losses in the communication link. However, a typical practice in link budget analysis is to split them in three components: a distance-based loss, a short-term loss (usually referred

as small-scale fading or simply fading) and a medium-term loss (referred as shadowing). In spite of the confusion, the distance-based loss is also sometimes referred as path loss. Each of these components obey to a different type of phenomena. Distance-based losses account for energy dissipated with increasing distance even without the presence of obstacles. The short-term losses arise from the multipath behaviour of a communication link. Energy arrives at the receiver from different angles and at different times, due to the interaction of the transmitted energy with the medium. Thus, the behaviour of the signals arriving at the receiver is stochastic and produces wide signal variations within a distance of a wavelength. On the other hand, signal variations due to obstacles in between the antennas, as trees, buildings, or terrain, are accounted in the term called shadowing.

In the process of modelling losses the purpose is to state a functional dependency between the observed quantity (a particular loss) and a set of assumed independent variables and parameters: distance, frequency, propagation environment, antennas height, etc. The parameters allow the model to establish different functional dependencies between the variables and the losses, or allow the model to establish discontinuous ranges for the variables.

3.2 Distance-based path loss

Beyond free space losses, arising from electromagnetic theory, losses due to distance have been empirically modelled, at least since the 50's [Egl57]. Extensive campaign measurements carried out at different distances, and possibly different frequencies and terrains or building densities, indicate that distance losses can be modelled by a power law. The simplest power law model without any frequency, neither antenna heights dependency, is shown in (3.2)

$$L[dB] = 10 \log(K_1 r^n) = 10n \log r + K \quad (3.2)$$

where r is the distance, n is the path loss exponent (a parameter of the model) and K_1 can be seen as the propagation loss experienced at 1 m.

The path loss exponent seems to account for different types of propagation environments (factories, offices, urban macrocells, urban microcells, etc.). Although it is an empirical parameter and can't be interpreted strictly as a variable that takes continuous values, higher values of n tend to show highly obstructed propagation conditions. For those cases, for example, the intuitive explanation is that the received power drops by a higher factor in a highly obstructed scenario than in one with Line of Sight (LoS) conditions, for the same distance range. From a mathematical point of view it represents the linear slope of the relationship between the received power and the distance in the $\log_{10}$ scale.

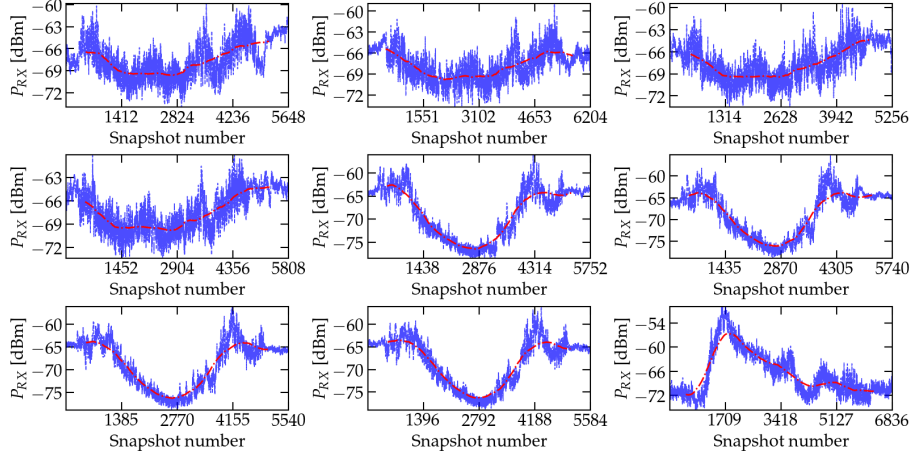


Figure 3.1: Received powers for all the measurements taken with Mercator BS. The corresponding routes are depicted in Figure 2.8 (Chapter 2). The dotted red lines represent the mean functions $m_f(t)$ after smoothing with $w = 1000$.

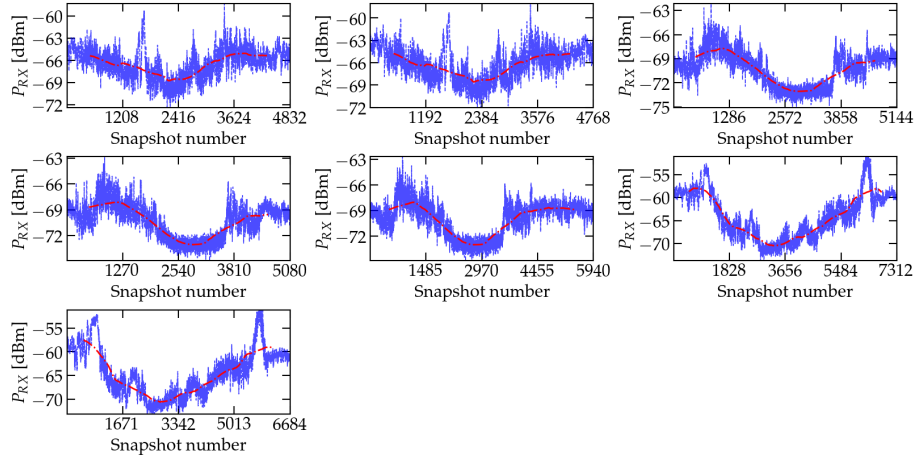


Figure 3.2: Received powers for all the measurements taken with Maxwell BS. The corresponding routes are depicted in Figure 2.9 (Chapter 2). The dotted red lines represent the mean functions $m_f(t)$ after smoothing with $w = 1000$.

Algorithm 1 shows the procedure to find out the path loss exponent. The procedure also produce the shadowing values (denoted in step 8 as X_{SH}) for which a separate analysis will be given in section 3.3.

Received power P_{RX} is computed from the channel impulse response. As explained in Chapter 2, there is an IRF $\mathbf{H}_t \in \mathbb{C}^{N_t \times N_{TX} N_{RX}}$ each $t = \frac{1}{f_{IRF}}$ seconds, meaning the channel is sensed at that rate. IRF entries for which the

square magnitude (power) is less than a noise threshold are removed. After that, all the delay contributions are added, in order to get a narrowband IRF, and all the channels magnitude are combined.

The smoothing window length w controls how much short-term fading is removed. Higher values of it can produce too much smoothing, which in turn might remove the medium-term losses or shadowing. Durgin proposed the concept of a *local area*, for which a terminal only experiences small-scale effects. That is to say, that for such zone, distance-based losses and obstruction losses (shadowing) are not relevant. More precisely, Durgin defines the zone as the largest volume enclosing a point, for which the radio channel can be modelled as a sum of homogeneous plane waves [Dur00]. He also proposes an upper limit for it:

$$\frac{L_A}{\lambda} < \frac{f_c}{B} \quad (3.3)$$

where L_A is the local area size, λ the wavelength, f_c the carrier frequency, and B the bandwidth. The smoothing (or moving average) step 2 in Algorithm 1 requires that the last $D-1$ samples of L_{SMOOTH} remain equal to the last $D-1$ samples of L . Equivalently, it is possible to make the first $D-1$ samples of L_{SMOOTH} equal to those of L , and the moving average will reach the end of the signal L . For the measurement campaign used throughout this document: $f_c = 3.7\text{GHz}$, $B = 200\text{MHz}$. So, $L_A < 18.5\lambda$.

Algorithm 1 Empirical extraction of distance-based loss and shadowing

Input: P_{TX} [dBm], $\mathbf{H}_t \in \mathbb{C}^{N_r \times N_{TX} \times N_{RX}}$, $t = 1, \dots, T$, smoothing window length w , noise threshold ξ

Output: n, Sh

- 1: Remove noise floor: $\forall i, j, t \mid |[\mathbf{H}]_{ijt}|^2 < \xi \rightarrow [\mathbf{H}]_{ij} = \text{NaN}$
 - 2: $[\mathbf{H}^{NARROW}]_{tj} = \sum_{i=1}^{N_r} [\mathbf{H}]_{ijt}$
 - 3: $P_{RX}[t] = 10 \log_{10} \left(\sum_{j=1}^{N_{RX} N_{TX}} |[\mathbf{H}^{NARROW}]_{tj}|^2 \right)$
 - 4: $L \text{ [dB]} = P_{TX} - P_{RX}$
 - 5: Smooth short-term losses: $L_{SMOOTH}[k] = \frac{1}{w} \sum_{t=k}^{k+w-1} L[t]$
 - 6: Sort L_{SMOOTH} by increasing r .
 - 7: Find $\hat{n}$: $\arg \min_n ||L_{SMOOTH} - 10n \log_{10} r||_2^2$
 - 8: Find X_{SH} : $L_{SMOOTH} - 10\hat{n} \log_{10} r$
-

Step 7 in Algorithm 1 fits a mathematical model to an empirical data input. The mathematical model considered in this case is solely based on the data, and no physical information is used. Such a model for the path loss is usually referred as a *floating-intercept* model (Equation 3.5), since the value of L at

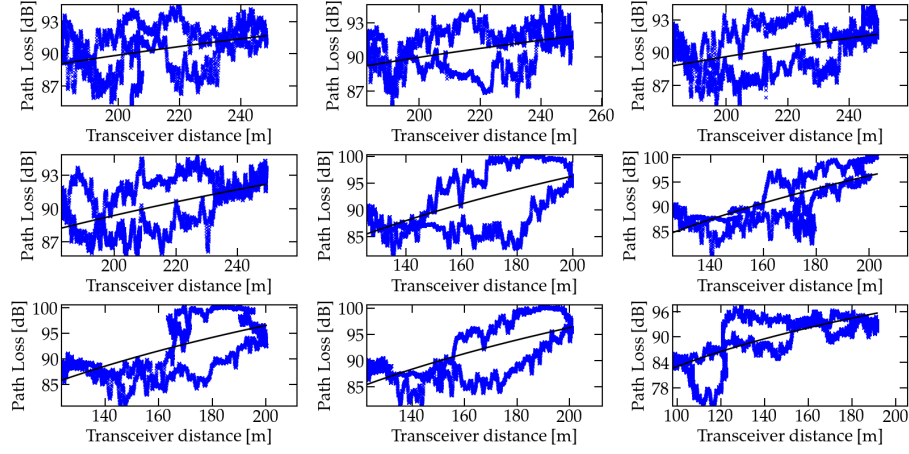


Figure 3.3: Path Loss for the received powers in Figure 3.1 (Mercator BS) with their linear fits. The linear fit uses the *floating-intercept* model.

the initial reference distance d_0 is provided by the fit. On the other hand, a *close-in* free space intercept (Equation 3.4), assumes L should have a specific value, based on physical considerations, for the reference distance d_0 . Such value is given by the Free Space Path Loss (FSPL).

$$L^{CI}[dB] = FSPL(d_0) + n^{CI} 10 \log_{10} \left(\frac{d}{d_0} \right) + X_{SH}, \quad FSPL(d_0) = \left(\frac{4\pi d_0}{\lambda} \right)^2 \quad (3.4)$$

$$L^{FI}[dB] = b + n^{FI} 10 \log_{10} \left(\frac{d}{d_0} \right) + X_{SH} \quad (3.5)$$

Path loss exponents estimated for each model, transmitter route and BS location are shown in Table 3.1. For the *close-in* model, all the linear regression fits are imposed to pass through the point $(1, \left(\frac{4\pi f_c}{c}\right)^2)$. Due to that, and to the fact that all observations are confined to a range of transmitter-receiver distances small (less than 50 % in some cases) compared to the full range of distances (around all the observations lie in a distance range of 100 m), the path loss exponents for this model have all a similar value. For instance, for the routes for which the range of distances is bigger (Marie Curie and Saint Barbe routes for the Mercator BS have observations within a distance range of 100 m) the path loss exponent increases its value, meaning that the lack of observations in the full range of distances or a lack of observations in a bigger range of distances might be hiding the true n^{CI} for this model. On the other hand, since the *floating-point* model doesn't impose an intercept for the

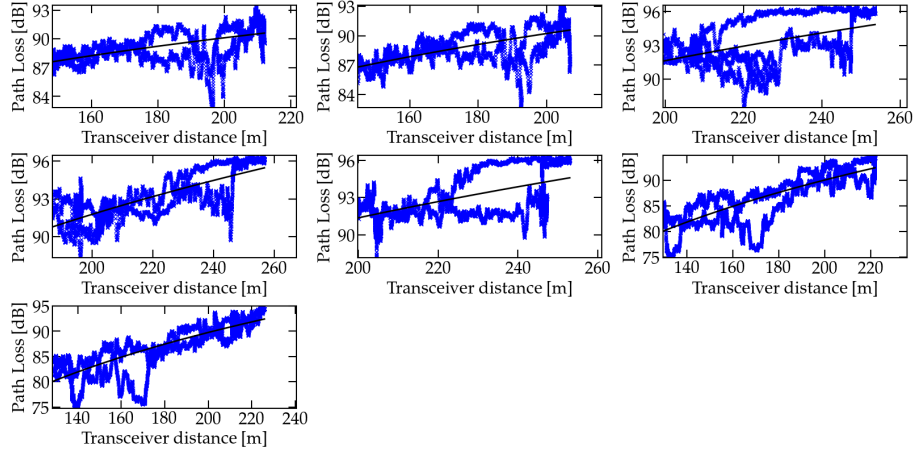


Figure 3.4: Path Loss for the received powers in Figure 3.2 (Maxwell BS) with their linear fits. The linear fit uses the *floating-intercept* model.

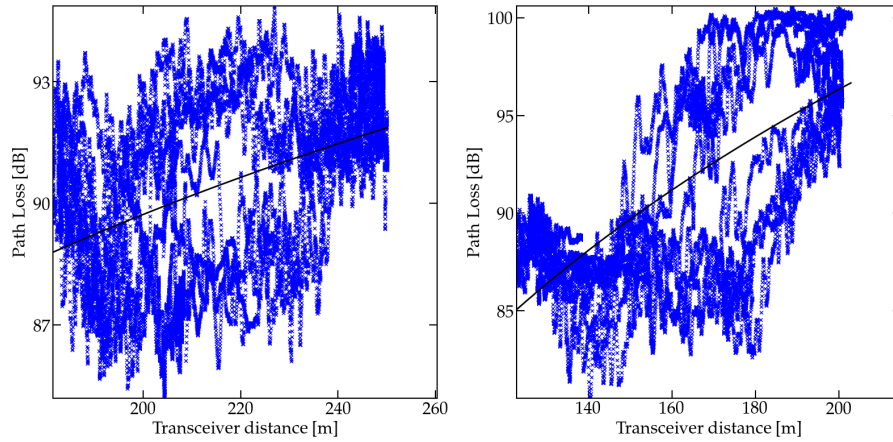


Figure 3.5: Path Loss for all the received power measurements with Mercator BS in Parking Lot (left) and Marie Curie (right). Results shown for the *floating-intercept* model.

equation of the line (b is a parameter to be estimated in 3.5), the situation described previously doesn't occur. Path loss exponents vary enough to indicate that each route have different propagation environment. However, the cost paid by the *floating-point* model is that the estimated intercept b might not have physical meaning. For instance, for some of the routes the estimated b indicates that the path loss is very small (negative in dB) or negligible, in contradiction with the $\approx 40\text{dB}$ free space loss at 3.7 GHz and 1m of distance.

Another bias of the linear regression fit should occur due to the lack of

Table 3.1: Path loss exponents and non-stationary measures.

BS	PL ^a	n^{FI}	n^{CI}	$n_{ALL-ROUTE}^{FI}$	$n_{ALL-ROUTE}^{CI}$	ROUGHNESS	$\sigma_{m_f(t)}$
MKT	1	1.93	2.08	2.21	2.08	5.8×10^{-7}	1.53
MKT	2	1.9	2.08			1.1×10^{-7}	1.42
MKT	3	2.15	2.07			4.8×10^{-7}	1.6
MKT	4	2.95	2.07			3.4×10^{-7}	1.86
MXW	1	2.44	2.09	2.18	2.09	4.6×10^{-7}	1.2
MXW	2	1.93	2.09			5×10^{-7}	1.17
BS	MC ^b	n^{FI}	n^{CI}	$n_{ALL-ROUTE}^{FI}$	$n_{ALL-ROUTE}^{CI}$	ROUGHNESS	$\sigma_{m_f(t)}$
MKT	1	5.4	2.24	5.32	2.24	1.2×10^{-6}	4.83
MKT	2	5.72	2.23			1.6×10^{-6}	4.57
MKT	3	5.17	2.25			8.9×10^{-7}	4.54
MKT	4	5.13	2.24			9.8×10^{-7}	4.67
MXW	1	3.12	2.18	3.22	2.18	3×10^{-7}	1.76
MXW	2	3.41	2.19			3.4×10^{-7}	1.64
MXW	3	3.17	2.17			2.6×10^{-7}	1.64
BS	SB ^c	n^{FI}	n^{CI}	$n_{ALL-ROUTE}^{FI}$	$n_{ALL-ROUTE}^{CI}$	ROUGHNESS	$\sigma_{m_f(t)}$
MKT	1	4.47	2.23	-	-	3.1×10^{-6}	4.64
MXW	1	5.29	2.01	5.16	2.01	1.23×10^{-6}	4.26
MXW	2	5.04	2.02			1.63×10^{-6}	4

^a Parking Lot route.^b Marie Curie route.^c Saint Barbe route.The mean function $\sigma_{m_f(t)}$ was computed using a window $w = 1000 \simeq 620\lambda$.

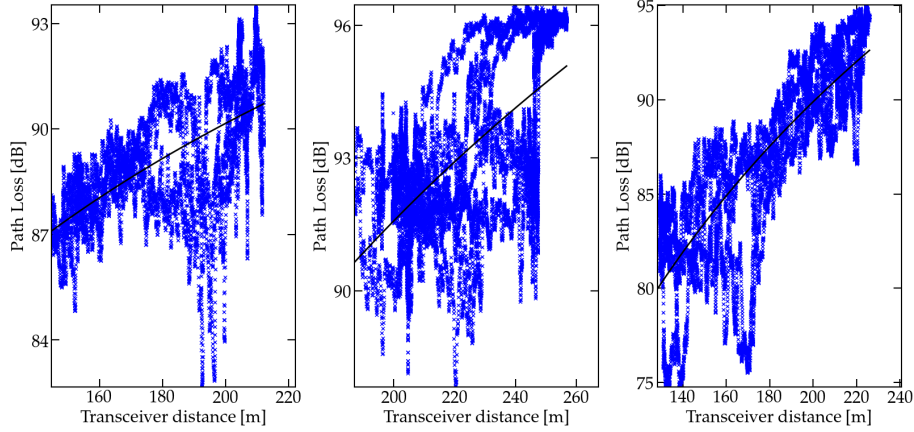


Figure 3.6: Path Loss for all the received power measurements with Maxwell BS in Parking Lot (left), Marie Curie (center) and Saint Barbe (right). Results shown for the *floating-intercept* model.

enough observations in the range of distances. In other words, observations at some distances might be biasing the model, since there are more available observations at those particular distances. To overcome this problem, the path loss for the *floating-point* model is again computed using all the observations of a given route in the same fit. Since all the measurements of the Parking Lot have the same two BS and follow approximately the same route it is reasonable to assume that they all belong to the same propagation environment. The same reasoning applies to the other routes. Figures 3.5 and 3.6 and column $n_{ALL-ROUTE}^{FI}$ of Table 3.1 show path loss computed following the previous approach. It becomes evident now, that more observations are available, and that for some of the routes the linear fit represent a clear linear tendency (see Parking Lot route with Mercator BS in the left panel of Figure 3.5). Besides, the new path loss exponent for a given route behave as the mean of all the n^{FI} computed for that route. Equivalently, the $n_{ALLROUTE}^{CI}$ is computed for all measurements per route and displayed in Table 3.1.

3.2.1 Are measurements of the same route coming from the same environment?

The assumption that all the measurements with the same BS and very similar transmitter routes come from the same propagation environment can be tested. In reality, at least minor changes in IRFs might occur due to the fact that even for the same route, each measurement is conducted at a different moment, for which the position of obstacles as people and vehicles have changed. In [Czi+08a], [Her+05] the correlation matrix distance (CMD)

measure 3.6 is used to compare the similarity between 2 matrices

$$c(\mathbf{A}, \mathbf{B}) = \frac{|tr(\mathbf{A}\mathbf{B}^T)|}{\|\mathbf{A}\|_F \|\mathbf{B}\|_F} \quad (3.6)$$

with $tr(\cdot)$ being the trace of a matrix and $\|\cdot\|_F$ the Frobenius norm. The CMD can be used by comparing channel matrices $\mathbf{H}_t^{MEAS-1}, \mathbf{H}_t^{MEAS-2}$ at time snapshot t for 2 measurements of the same route. However, since for even small changes in phase the complex gains in both matrices vary in a noticeable manner, the CMD will mostly have a low value when using instantaneous channel matrices. A solution to overcome such a situation is to average channel matrices over a time snapshot interval. Another approach would be to use instead, one of the spatial correlation matrices $\mathbf{R}_{TX} = \mathbb{E}[\mathbf{H}^T(\mathbf{H}^T)^*]$, $\mathbf{R}_{RX} = \mathbb{E}[\mathbf{H}\mathbf{H}^*]$ or $\mathbf{R}_{FULL} = \mathbb{E}[\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^*]$, to test if local propagation characteristics vary [Her+05]. In order to do so, we assume ergodicity, and so, replace the ensemble averages with time averages. A question must be arising at this point: why not, instead of assuming ergodicity, average over the available measurements of the same route, since they all constitute part of the same stochastic process? There are two answers. First: precisely, what we are trying to see is if all these measurements with the same BS and very similar routes, are referring to the same stochastic process, or in other words, to the same propagation environment. Second: there are not enough measurement to average over measurements.

For example, for the transmitter spatial correlation matrix $\mathbf{R}_{TX}$ we define disjoint 'snapshot regions' of size w_{ERGOD} over which we average

$$\mathbf{R}_{TX} = \mathbb{E}[\mathbf{H}^T(\mathbf{H}^T)^*] \Rightarrow \mathbf{R}_{TX,i} = \sum_{t=(i-1)*w_{ERGOD}+1}^{i*w_{ERGOD}} \mathbf{H}_t^T(\mathbf{H}_t^T)^*, i = 1, \dots, \lfloor \frac{N_T}{w_{ERGOD}} \rfloor \quad (3.7)$$

with N_T , the number of snapshots available in the measurement. The same procedure can be used for $\mathbf{R}_{RX}$ and $\mathbf{R}_{FULL}$. Figures 3.7 and 3.8 show the CMD between measurements of the same route, for each route and each BS. Regions of size $w_{ERGOD} = 150\lambda$ were used, such that there are enough IRFs available for the average in 3.7.

Both panels in Figure 3.7 show similar aspects. First, for the beginning and end of the Marie Curie and Parking Lot routes the CMD is low. It means that for this snapshot regions, the multiple measurements of each route have non-negligible differences in propagation conditions. Differences in the measurement routes are noticeable or there are changes in propagation conditions due to, for example, the presence of new obstacles. Second, the CMD remains high (bigger than 0.5) for more than half of the snapshot regions (see

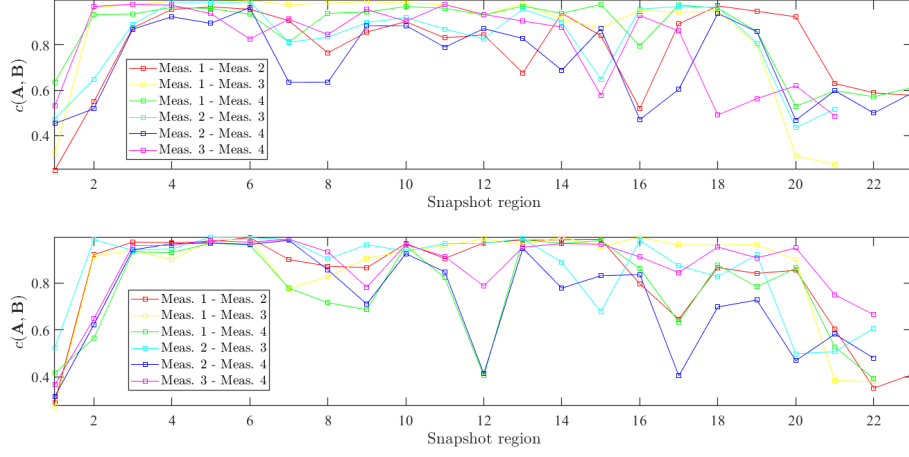


Figure 3.7: CMD with Mercator BS. Top panel: Parking Lot route measurements. Bottom panel: Marie Curie route measurements. R_{TX} matrices used for the CMD, with $w_{ERGOD} = 150\lambda$.

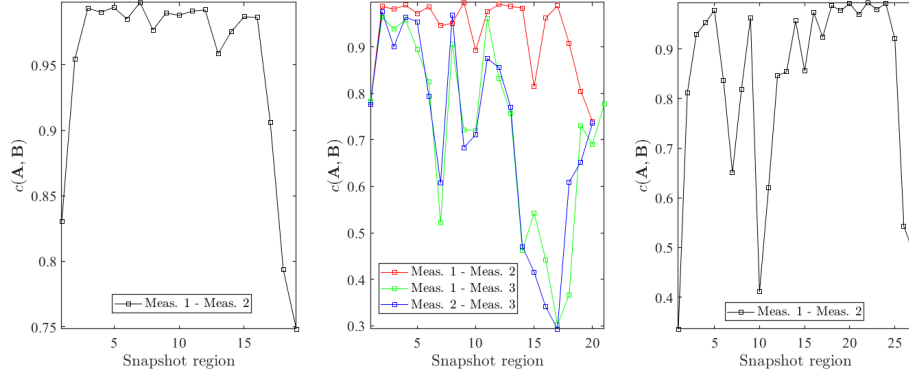


Figure 3.8: CMD with Maxwell BS. Left panel: Parking Lot route measurements. Central panel: Marie Curie route measurements. Right panel: Saint Barbe route measurements. R_{TX} matrices used for the CMD, with $w_{ERGOD} = 150\lambda$.

snapshot regions 2 to 11, on both panels). This means it could be said that for long segments of the route the measurements belong to the same environment, which intuitively was the conclusion. Third, the graphs show high variability at snapshot regions 14 to 20, specially for the Parking Lot case. The propagation conditions are different in these case, since in fact, for these snapshot regions the routes are very similar: the average difference between the transmitter locations for the Parking Lot route is less than 1 m.

On the other hand, in Figure 3.8 the measurements for Maxwell BS show different results. The available 2 measurements for the Parking Lot route are

quite similar. The CMD is always above 0.6, and for most of the time above 0.8. The 2 measurements available for the Saint Barbe route also exhibit a high similarity for most of the snapshot regions. However, it is evident that the similarity in the case of the Saint Barbe route is lower in average than in the Parking Lot route. This might be due to the mechanism of propagation mostly operating in the Parking Lot case. While there are no clear mechanisms of propagation for some snapshot regions of the Saint Barbe route, it is clear that most of the radio signals departure at Parking Lot and arrive at Maxwell BS through the rooftops of the near buildings. In the case of Marie Curie measurements, the first and second measurements have a high similarity along all the snapshot regions.

3.2.2 Degree of non-stationarity

Figures 3.1 and 3.2 show the received power for all the routes in Figures 2.9 and 2.8, and measurements taken with each BS. Small-scale fading is visible in the fast variation of the signals. They serve as a reference for the path losses shown in Figures 3.3 and 3.4. As each of the measurements was carried out separately, we plot the path loss L for each of them with its closest linear fit. These path losses and linear fits correspond to the complete set of received power samples in each route, thus, the path loss exponents correspond to the environment 'seen' by the links between the mobile terminal and each base station. That's the reason why, for example, for the Marie Curie route, the path loss exponents differ significantly for each BS. For each BS the environment through which most energy propagates differs.

For the Parking Lot routes all path loss exponents seem to have the same value. Both Mercator and Maxwell BS received powers show a high presence of small-scale fading and small changes in signal tendency. As the transmitter moves away from the local area, the signal doesn't drops or increases its value 'significantly'. From the propagation point of view, what is happening is that most of the energy comes from multiple paths, as there is no LoS between transmitter and receiver. It means that mostly diffraction is the mechanism of propagation. To measure what does it mean to have a 'significant' change in tendency, first, by comparison, the Marie Curie and Saint Barbe against the Parking Lot received power signals can be seen. Either in Mercator or Maxwell BS cases, it is evident the change in the tendency of the signals from Marie Curie and Saint Barbe, while for the Parking Lot route, the tendencies are less prominent. This is occurring since for both cases there are transmitter zones for which the mechanisms of propagation vary. For instance, for the Saint Barbe route with Maxwell BS there is a explicit LoS segment for which the received power increases.

Noticeable tendencies in a time series are indicative of a non-stationary

behaviour. A way of measuring the non-stationarity of a time series is proposed in [DN16]. The idea is to measure the 'roughness' of the underlying mean signal. There are different measures for the 'roughness' or 'wigglyness', but a natural measure would be proportional to the average of the instantaneous curvature [Ram04]. Since the curvature of $f(t) : [a, b] \mapsto \mathbb{R}$ at t is $\left(\frac{d^2 f(t)}{dt^2}\right)^2$, the roughness is:

$$\text{ROUGHNESS}[f] \doteq \int_a^b \left(\frac{d^2 f(t)}{dt^2}\right)^2 dt \quad (3.8)$$

To have a better look at the tendencies of $f(t)$, we estimate its mean function

$$m_f(t) \doteq \mathbb{E}[f(t)] \rightarrow \hat{m}_f(t) = \frac{1}{w} \sum_{n=t}^{t+w-1} f(n) \quad (3.9)$$

by smoothing $f(t)$ with a sufficiently large smoothing window size w (as in step 5 of Algorithm 1). However, the smoothed signals might not have second order derivatives. Moreover, no functional expression is available for the smoothed function, and so, derivatives have to be computed by numerical differentiation, introducing errors as a consequence of the approximation.

A way to overcome that is to fit a non-parametric function $g(t)$ to $f(t)$, by solving the minimization problem

$$\arg \min_{g(t)} \|f(t) - g(t)\|_2^2 + \lambda_R \text{ROUGHNESS}[g] \quad (3.10)$$

The first component of the sum is proportional to the mean squared error between $f(t)$ and $g(t)$. The second component is the roughness defined in 3.8, and the weighting factor λ_R controls how much importance has the error (how tied is $g(t)$ to $f(t)$) and the roughness: higher values mean the roughness has higher relevance, and so, the error less relevance in the solution.

A well-known solution to 3.10 is the spline of the form

$$g(t) = \sum_k^K c_k \phi_k(t) \quad (3.11)$$

with $\phi_k(t)$ as the cubic polynomial spline basis functions and c_k their associated coefficients. Conditions about continuity and the existence of second order derivatives are imposed on the basis functions for a set of discrete points called 'knots' $\{t_i\}_{i=1}^K$, $t_1 < t_2 < \dots < t_K$. For t that do not belong to the set of 'knots' the function values of $g(t)$ are interpolated.

The advantage gained through the use of splines is that now the roughness can be computed, since the basis functions $\phi_k(t)$ have analytical second order derivatives. Moreover, from a practical point of view, splines are readily-available in most of the software used to curve fitting as part of their core functionality, so there is no need to implement solvers for 3.10.

For a given set of points $(t_i, f(t_i))$ and solution $g(t)$ of the form in 3.11, 3.8 can be rewritten as:

$$\text{ROUGHNESS} \doteq \mathbf{c}^T \mathbf{B} \mathbf{c} \quad (3.12)$$

with

$$\mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_K \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \int \frac{d^2 \phi_1(t)}{dt^2} \frac{d^2 \phi_1(t)}{dt^2} dt & \dots & \int \frac{d^2 \phi_1(t)}{dt^2} \frac{d^2 \phi_K(t)}{dt^2} dt \\ \vdots & \ddots & \vdots \\ \int \frac{d^2 \phi_K(t)}{dt^2} \frac{d^2 \phi_1(t)}{dt^2} dt & \dots & \int \frac{d^2 \phi_K(t)}{dt^2} \frac{d^2 \phi_K(t)}{dt^2} dt \end{bmatrix} \quad (3.13)$$

A lower value for the roughness means the time series is stationary, as in the case of a constant function for which the roughness measure is zero. Alternatively, the standard deviation $\sigma_{m_f(t)}$ of the mean function is used to measure in average how much the mean function varies from its mean value: that is to say, how much the mean function differs from a constant function. Roughness and $\sigma_{m_f(t)}$ are given in Table 3.1. A look of both values confirms what our intuition was saying before: Parking Lot routes, with any of the BS, have the lowest values of roughness and $\sigma_{m_f(t)}$ of all the routes, thus, indicating a more stationary behaviour overall each route. On the other hand, for Mercator BS, both Marie Curie and Saint Barbe routes seem to exhibit a similar degree of non-stationary. For instance, both routes have segments of snapshots for which there might be LoS or at least OLoS conditions, as there are no big obstacles obstructing the direct path between the transmitter and the BS.

In the case of Maxwell BS, the stationarity measures are showing something slightly different. Now, the Marie Curie routes are more stationary, due to the NLoS conditions mostly taking place along the route, causing that multiple paths arrive to the BS, for most of the transmitter route. Changes in tendency are not as pronounced as in the case of Mercator BS.

3.3 Shadowing

Shadowing variable X_{SH} is computed from step 8 in Algorithm 1. It can be understood as the remaining signal after removing small-scale fading and subtracting the distance-based path loss. Due to its stochastic nature it has been modelled as a random variable. The widely accepted [LR92], [GGF94] assumption that shadowing distributes according to a Gaussian distribution in the dB scale (or lognormal in the linear scale 3.14), arises from what follows. From now and on, if nothing specified X_{Sh} refers to the shadowing in the dB scale.

$$X_{Sh}[dB] \sim \mathcal{N}(\mu_N, \sigma_N^2) \Rightarrow X_{Sh}[LIN. SCALE] = 10^{\frac{X_{Sh}[dB]}{10}} \sim p_{X_{Sh}}(x) \quad (3.14)$$

$$p_{X_{Sh}[LIN. SCALE]}(x) = \frac{10}{\log_e(10)\sqrt{2\pi}\sigma_N} \exp\left(-\frac{(10\log_{10}(x) - \mu_N)^2}{2\sigma_N^2}\right)$$

$$\forall x \geq 0$$

It is called shadowing the local variation of received power, assumed to be a consequence of obstacles. If along a path each obstacle independently attenuates the signal by an amount A_i , the total attenuation due to all the obstacles in the path is $\prod_i A_i$ or $\sum_i A_i [dB]$ in the dB scale. It is easy to notice that as the number of obstacles grows, no matter which probability distribution we use to model the individual attenuations they make, the total attenuation must distribute as a Gaussian due to the Central Limit Theorem.

In reality the individual attenuations of the obstacles are not independent. There might exist correlations between some obstacles that don't allow to express the total attenuation as a product between all the obstacles. Besides that, not all the losses due to the obstacles have the same contribution in the total loss: an example are the losses close to the mobile antenna, which might have a bigger impact. Moreover, losses due to diffraction do not add up in such a way [SAZ07]. While the shadowing-gaussian assumption is well established, there are other models for it also [AK99], [Sha08].

In Figures 3.9 and 3.11 the empirical distributions of the shadowing for all the measurements of each route and BS are shown. For each of the shadowing signals computed, two statistical tests are conducted in which we test if there is enough statistical evidence to reject or no the null hypothesis H_0 . First, we test if the shadowing empirical distribution follows a gaussian distribution (3.15). Second, we test if it follows a given probability distribution $p_{X_{Sh}}(x)$ (3.16) from the following set of test distributions: alpha, beta, gaussian, expo-

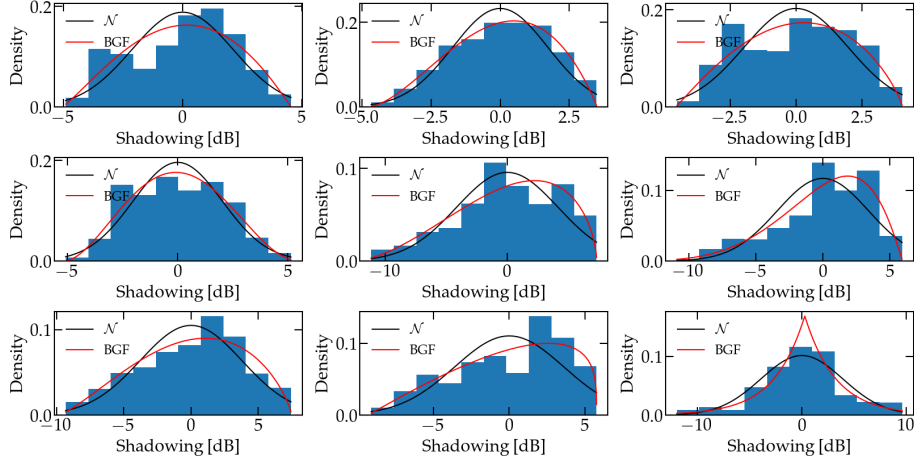


Figure 3.9: Empirical distribution of shadowing for all the measurements with Mercator BS. The solid black line represents a gaussian distribution fit with parameters shown in Table 3.2. The solid red line represents the BGF distribution with the parameters shown in Table 3.3.

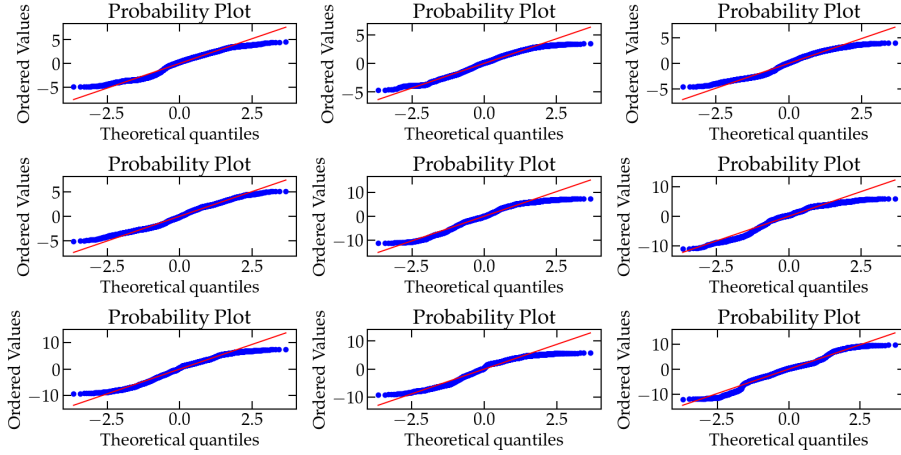


Figure 3.10: Q-Q plots: shadowing empirical distribution against gaussian distribution for all measurements with Mercator BS.

nential, gamma, laplace, nakagami, rayleigh, rice, uniform, gaussian hypergeometric. Each of the distributions have at least location (LOC parameter controlling the median location) and scale (SCA parameter controlling the spread), and some might have shape (SHAPE parameter controlling shape aspects of the distribution) parameters as well.

$$H_0 : X_{Sh} \sim \mathcal{N}(\mu_N, \sigma_N^2), \quad H_1 : X_{Sh} \not\sim \mathcal{N}(\mu, \sigma^2) \quad (3.15)$$

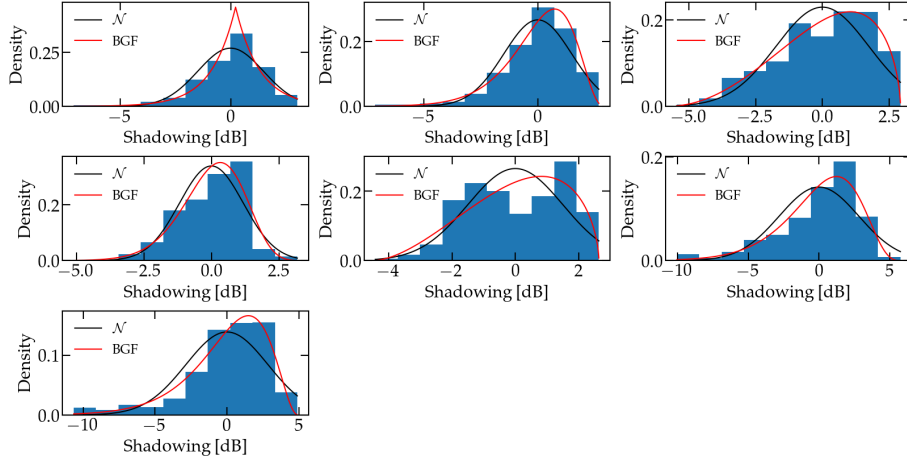


Figure 3.11: Empirical distribution of shadowing for all the measurements with Maxwell BS. The solid black line represents a gaussian distribution fit with parameters shown in Table 3.2. The solid red line represents the BGF distribution with the parameters shown in Table 3.3.

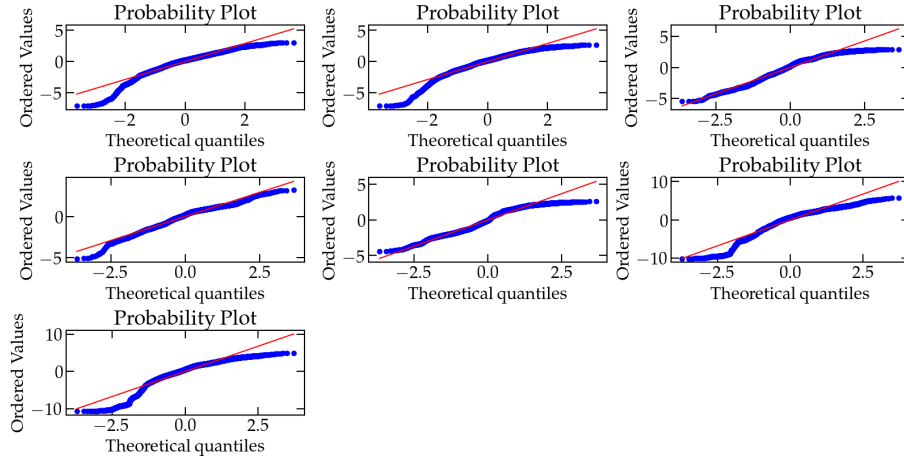


Figure 3.12: Q-Q plots: shadowing empirical distribution against gaussian distribution for all measurements with Maxwell BS.

$$H_0 : X_{Sh} \sim p_{X_{Sh}}(x)(LOC, SCA, SHAPE), \quad H_1 : X_{Sh} \not\sim p_{X_{Sh}}(x) \quad (3.16)$$

For the second test, it is kept only the test probability distribution providing the Best Goodness of Fit (BGF) to the empirical distribution: the one with the lowest test statistic value. The Kolmogorov-Smirnov statistical test

Table 3.2: Gaussian distribution fit for shadowing.

BS	PL	$\mathcal{N}$ Stat. ^a	$\mu_{\mathcal{N}}$ ^b	$\sigma_{\mathcal{N}}$ ^c	CV ^d	$\mathcal{N}$ Stat.	$\mu_{\mathcal{N}}$	$\sigma_{\mathcal{N}}$
MKT	1	0.067	0	2.12	0.018	0.054	0	1.96
MKT	2	0.041	0	1.72	0.017			
MKT	3	0.071	0	1.97	0.018			
MKT	4	0.059	0	2.03	0.017			
MXW	1	0.073	0	1.49	0.020	0.072	0	1.49
MXW	2	0.083	0	1.48	0.019			
BS	MC	$\mathcal{N}$ Stat.	$\mu_{\mathcal{N}}$	$\sigma_{\mathcal{N}}$	CV	$\mathcal{N}$ Stat.	$\mu_{\mathcal{N}}$	$\sigma_{\mathcal{N}}$
MKT	1	0.064	0	4.16	0.018	0.069	0	3.76
MKT	2	0.085	0	3.41	0.018			
MKT	3	0.083	0	3.78	0.018			
MKT	4	0.118	0	3.61	0.018			
MXW	1	0.092	0	1.74	0.019	0.092	0	1.5
MXW	2	0.094	0	1.18	0.019			
MXW	3	0.094	0	1.5	0.017			
BS	SB	$\mathcal{N}$ Stat.	$\mu_{\mathcal{N}}$	$\sigma_{\mathcal{N}}$	CV	$\mathcal{N}$ Stat.	$\mu_{\mathcal{N}}$	$\sigma_{\mathcal{N}}$
MKT	1	0.065	0	3.92	0.016	-	-	-
MXW	1	0.11	0	2.8	0.016	0.091	0	2.84
MXW	2	0.089	0	2.85	0.016			

^a Kolmogorov-Smirnov statistic value for the gaussian distribution fit.

^b Mean of the gaussian distribution fit.

^c Standard deviation of the gaussian distribution fit.

^d Critical value to reject null hypothesis: $\sqrt{\frac{-\log(\alpha * 0.5)}{n}}$, with $\alpha = 0.05$ and n as the number of samples.

is used. The statistic computed by the test (3.17) is compared to a critical value that depends on the size of the sample and the desired level of significance. In 3.17, $F_{EMP}(x)$ is the empirical cumulative distribution, while $F_{X_{Sh}}(x)$ is the cumulative test distribution.

$$S = \sup_x |F_{X_{Sh}}(x) - F_{EMP}(x)| \quad (3.17)$$

Table 3.3: Best Goodness of Fit (BGF) distribution for shadowing.

BS	PL	Distr.	S ^a	Loc. ^b	Sca. ^c	Shape
MKT	1	GHG ^d	0.054	-4.96	9.58	$a = 1.85, b = 2.7, c = 1.76, z = -0.79$
MKT	2	Beta	0.018	-4.8	8.3	$a = 2.69, b = 1.96$
MKT	3	GHG	0.046	-4.65	8.67	$a = 2.67, b = 1.68, c = 1.64, z = 2.13$
MKT	4	Beta	0.04	-5.18	10.75	$a = 2.95, b = 3.16$
MXW	1	GHG	0.028	-166.7	169.6	$a = 167.7, b = 3.9, c = -116.6, z = 0.8$
MXW	2	GHG	0.031	-122.3	126	$a = 164.7, b = 6.9, c = 4.6, z = -0.97$
BS	MC	Distr.	S	Loc.	Sca.	Shape
MKT	1	Beta	0.045	-12.7	20	$a = 2.77, b = 1.6$
MKT	2	Beta	0.043	-17.9	23.8	$a = 6.31, b = 2.1$
MKT	3	GHG	0.043	-9.47	17	$a = 1.92, b = 2, c = 1.6, z = -0.57$
MKT	4	GHG	0.056	-9.2	15	$a = 1.72, b = 1.76, c = 1, z = -0.89$
MXW	1	GHG	0.039	-6.64	9.58	$a = 4, b = 1.29, c = -0.45, z = -1$
MXW	2	Beta	0.068	-4358.1	4362.4	$a = 13261.8, b = 12.9$
MXW	3	GHG	0.055	-4.49	7.13	$a = 2.79, b = 1.58, c = 0.26, z = 0.9$
BS	SB	Distr.	S	Loc.	Sca.	Shape
MKT	1	Laplace	0.056	0.28	2.94	-
MXW	1	GHG	0.033	-16.2	22.7	$a = 1.63, b = 12.33, c = 21.87, z = -0.82$
MXW	2	GHG	0.028	-12.5	18.29	$a = 1.15, b = 10.43, c = 17.67, z = -0.83$

^a Kolmogorov-Smirnov statistic value for the BGF distribution.^b Location statistical parameter of the distribution.^c Scale statistical parameter of the distribution.^d Gaussian Hyper-Geometric distribution.

Results for test 3.15 are comprised in Table 3.2. Given a level of significance of 0.05, all the statistic values (column $\mathcal{N}$ Stat. in Table 3.2) are above the critical value, and so the null hypothesis must be rejected. Strictly speaking, since H_0 is rejected there is no statistical evidence to say shadowing distributes as a gaussian. Most of the empirical distributions have non-symmetrical shapes, meaning the shadowing values after and before the mode do not have the same probability of occurrence. This is expected to happen, since the route is imposing specific propagation mechanisms, that is to say: incoming signals do not combine constructively or destructively in the same amount for all the route. The QQ-plots in Figures 3.10 and 3.12 are showing the same information in another way: it is at the extreme values of the distribution where the empirical distribution separates from a gaussian. Best values for $\mu_{\mathcal{N}}$ and $\sigma_{\mathcal{N}}^2$ are estimated and shown in Table 3.2, while the closest gaussian PDF to the empirical distribution is shown in the figures.

Test results for 3.16 are comprised in Table 3.3. Again, except for one of the measurements, none of the test distributions 'pass' the Kolmogorov-Smirnov test: the null hypothesis is rejected in all, but one, cases. Moreover, the test statistic (S column in Table 3.3) is in some cases not much better than the test statistic for the gaussian test. That can be appreciated in some of the graphs in Figures 3.9, 3.11: both distributions fitted are very close.

As it was stated in section 3.2, a bias in the path loss is occurring due to the lack of enough observations in the range of distances. To overcome the problem, all the measurements for a given route and BS were used as if they were coming from the same environment: the same route and BS, with the same propagation conditions. Figures 3.13 and 3.14 show the shadowing empirical distributions for this approach. It can be seen that in fact, the gaussian fit is closer to the empirical distribution, and that the BGF fit is also close to the gaussian fit, meaning there is no need to use another probability distribution rather than the gaussian, at the expense of higher analytical complexity (all the BGF distributions are either Gaussian Hyper-Geometric or Beta distributions).

3.4 LoS propagation

There exists LoS propagation for the Saint Barbe route with Maxwell BS. This is noticeable in the first peak of Figure 3.2 (Graphs (2,3) and (3,1)). The corresponding coordinates for the received power samples around the first peak are shown in Figure 3.15. Table 3.4 summarizes Kolmogorov-Smirnov tests for normality and BGF distribution for the shadowing in coordinates corresponding to the LoS received power conditions. Figure 3.16 shows the empirical and fitted distributions. For both measurements the path loss exponent is below 2, as expected, since not only LoS conditions are occurring, but there exists also a

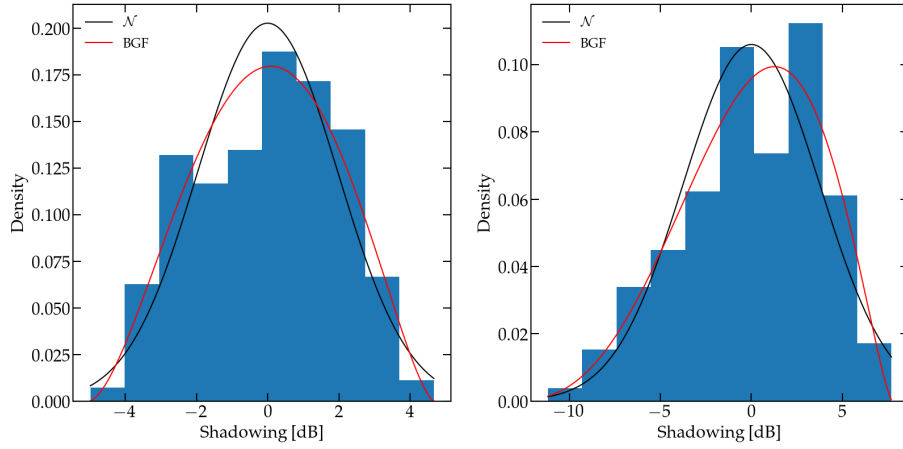


Figure 3.13: Empirical distribution of shadowing with Mercator BS for all Parking Lot (left), Marie Curie (center) and Saint Barbe (right) routes. The solid black line represents a gaussian distribution fit with parameters shown in Table 3.2. The solid red line represents the BGF distribution with the parameters shown in Table 3.3.

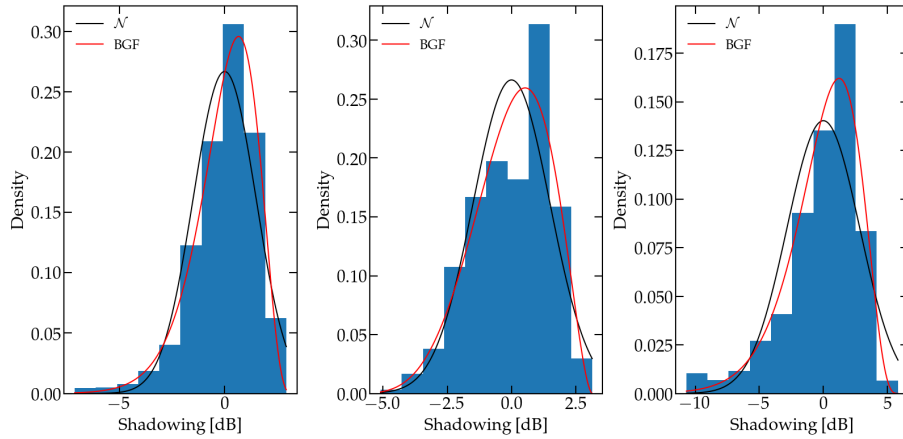


Figure 3.14: Empirical distribution of shadowing with Maxwell BS for all Parking Lot (left), Marie Curie (center) and Saint Barbe (right) routes. The solid black line represents a gaussian distribution fit with parameters shown in Table 3.2. The solid red line represents the BGF distribution with the parameters shown in Table 3.3.

'tunneling' effect due to the proximity of Saint Barbe and Réaumur buildings.

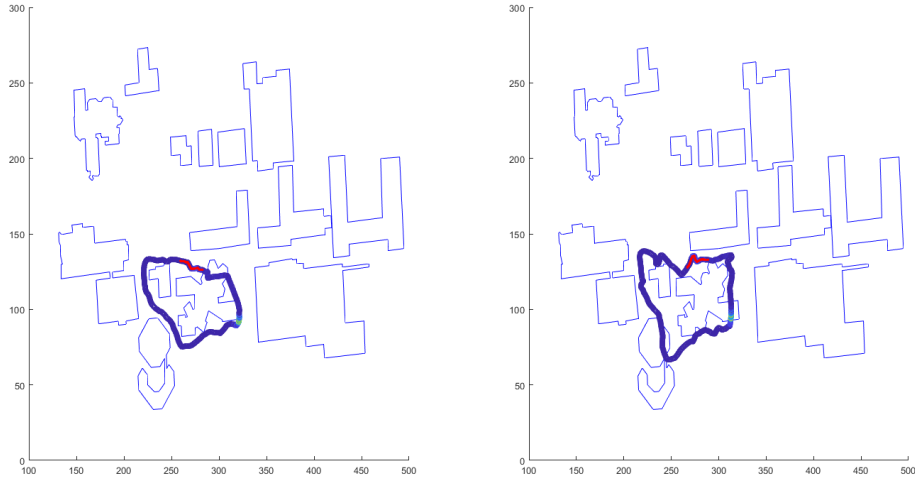


Figure 3.15: Saint Barbe routes with Maxwell BS. LoS points are highlighted in red

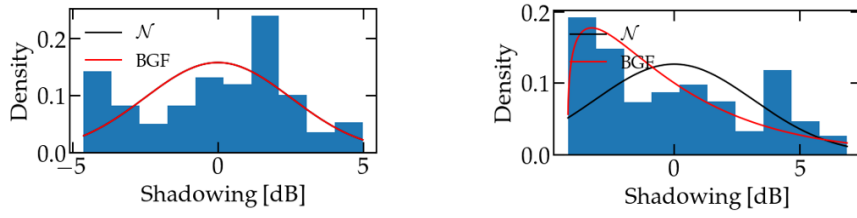


Figure 3.16: Distribution fits for the shadowing for the coordinates corresponding to LoS conditions in Figure 3.15.

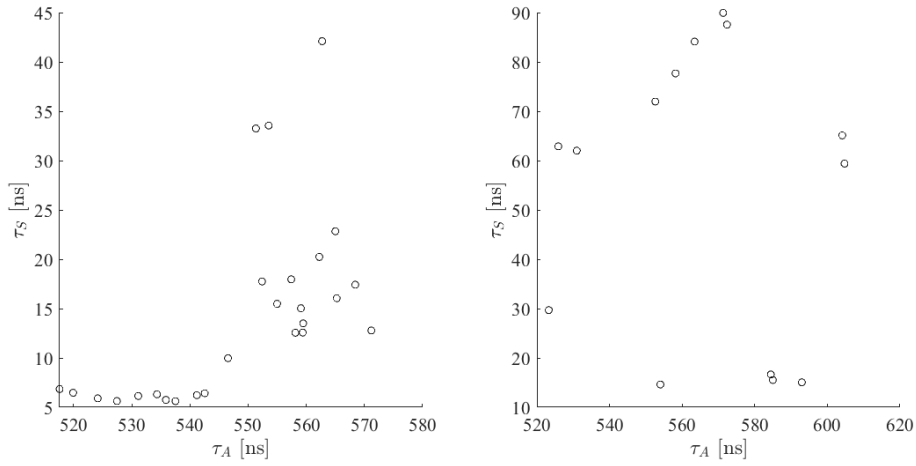


Figure 3.17: Delay spreads and mean delays for the stationary regions corresponding to the coordinates in LoS conditions shown in Figure 3.15 for the Saint Barbe route and Maxwell BS.

Table 3.4: Path loss exponent, delay statistics and shadowing Kolmogorov-Smirnov tests for LoS conditions in Saint Barbe route and Maxwell BS.

Parameter	Saint Barbe 1	Saint Barbe 2
n^{CI}	1.75	1.71
$\mathcal{N}$ KS-Stat.	0.1	0.14
$\mu_{\mathcal{N}}$	0	0
$\sigma_{\mathcal{N}}$	2.52	3.14
BGF PDF	Gaussian	Gamma
BGF KS-Stat.	0.1	0.09
CV	0.051	0.056
Loc.	0	-4.22
Scale	2.52	3.28
Shape	-	1.28
Median τ_S	12.7	62.4
Median τ_A	553	567
Std. Dev. τ_S	9.7	29.1
Std. Dev. τ_A	15.7	27

3.5 Wideband characterization

Multipath propagation introduces delay dispersion. Signals arrive at the receiver at different time delays. However, it is intuitively reasonable to think that not all the signals propagated through the different paths arrive with the same power, and that those coming from shorter paths will have more power. At a first glance there is no intuitive idea about how is distributed the power of the arriving signals on their times of arrival. Precisely because of that, one way of analysing the specificity of the environment under consideration is to compute a profile of the distribution of power over all the times of arrival. Such a profile is called a Power Delay Profile (PDP). For time-variant signals measured $\mathbf{H}_t(\tau) \in \mathbb{C}^{N_{RX} \times N_{TX}}$, an instantaneous (3.18, left panel in Figure 3.18) and an average PDP (3.19, right panel in Figure 3.18) could be computed:

$$PDP_t(\tau) = \frac{1}{N_{RX}N_{TX}} \sum_{i=1}^{N_{RX}} \sum_{j=1}^{N_{TX}} |[\mathbf{H}_t(\tau)]_{ij}|^2 \quad (3.18)$$

$$PDP_{\mathcal{SR}}(\tau) = \frac{1}{|\mathcal{SR}|N_{RX}N_{TX}} \sum_{t \in \mathcal{SR}} \sum_{i=1}^{N_{RX}} \sum_{j=1}^{N_{TX}} |[\mathbf{H}_t(\tau)]_{ij}|^2 \quad (3.19)$$

$\mathcal{SR}$ denotes a region of time (a set of consecutive snapshots) over which to average the PDP. More on the choice of this regions is presented in section 3.5.1. The number of snapshots $|\mathcal{SR}|$ considered in the average PDP plays an important role in the reduction of noise. If AWGN is considered with zero mean and variance σ_{AWGN}^2 , its variance will be reduced by a factor of $\frac{1}{\sqrt{N_T}}$, which in turn gives an enhancement of $20 \log_{10}(N_T)$ in the SNR. As in the case of the smoothing step to remove the small-scale fading, the N_T needs to be carefully chosen. Ideally, what is desired is to have maximum noise reduction while being sure about the coherency of the signals being averaged. In other words, it is necessary to know the stationarity regions, or the time regions (equivalently, the space regions) over which incoming signals remain coherent.

Before continuing a side note must be said to clarify concepts. In section 3.2.2 a similar issue arise when we try to measure the degree of non-stationarity of the received power signals. In this case, however, we are interested in looking for regions, that is to say, sets of snapshots for which we can assume stationarity assumption holds.

3.5.1 Stationary regions

A stationarity region $\mathcal{SR}$ is a region for which 1) the mean of the stochastic process do not vary, and 2) the covariance function doesn't depend on the

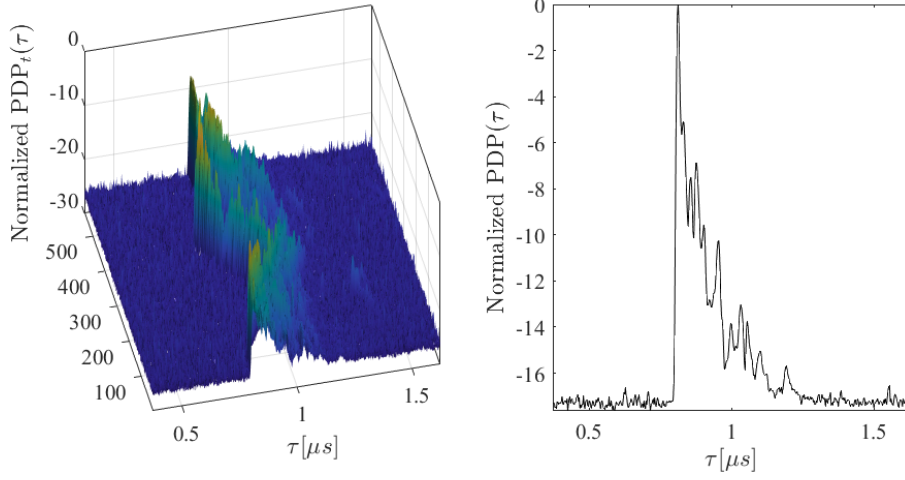


Figure 3.18: Left panel: an example of an instantaneous PDP. Right panel: example of an averaged PDP over 20 time snapshots ($\approx 12\lambda$). Both graphs correspond to real measured IRFs for the Parking Lot route and the Mercator BS. For the left panel graph, the time snapshots were sub-sampled at a third of the original rate for displaying purposes. Only relevant delay taps are shown.

specific time instants within the region, but rather, on the difference between such instants. It is thus a stationarity in the *wide-sense*.

$$\mathcal{SR} = \{t \mid \mathbb{E}[X(t)] = cte\} \quad (3.20)$$

$$\mathcal{SR} = \{t_1, t_2 \mid \text{COV}[t_1, t_2] = \text{COV}[t_2 - t_1, 0]\} \quad (3.21)$$

$$\text{COV}[t_1, t_2] = \mathbb{E}[(X(t_1) - \mathbb{E}[X(t_1)])(X(t_2) - \mathbb{E}[X(t_2)])] \quad (3.22)$$

where $\text{COV}[t_1, t_2]$ is the auto-covariance function. Then, for the instantaneous $\text{PDP}_t(\tau)$, $t = 1, \dots, T$ (with T the total number of available time-snapshots), we are given the problem of splitting the initial set of time-snapshots ($\mathcal{T}_0 = \{1, 2, \dots, T\}$) into disjoint sets $\mathcal{SR}_1, \mathcal{SR}_2, \dots, \mathcal{SR}_{N_{SR}}$ over which conditions 3.20 and 3.21 are fulfilled. Neither the number of regions or the size of each region is known.

Different approaches have been proposed to solve the problem. However, all of them rely on looking for similarities of a '*summary function*' capable of describe the channel under analysis, at different space locations or time samples. In [Pai+08], the authors use the *local scattering function* defined by Matz [Mat05], [Mat03] as the '*summary function*'. In [Ren+10], the spatial

Algorithm 2 Methodology to find out stationary regions.

Input: $\text{PDP}_t(\tau) \forall t \in \mathcal{T}_0, \mathcal{T}_0, n_{DISC}, \beta$ **Output:** $\mathcal{SR}_1, \dots, \mathcal{SR}_{N_{SR}}$

```

1:  $\alpha \leftarrow \mathcal{T}_0, k \leftarrow 1$ 
2: Choose the median of  $\alpha$  as the pivot  $p$ 
3: Split  $\alpha$  into 2 sets: the set left to  $p$  ( $\mathcal{L}$ ) and the one right to  $p$  ( $\mathcal{R}$ )
4: Test stationarity of  $\text{PDP}_t(\tau)$  for the time-snapshots in  $\mathcal{L}$ 
5: if  $\text{PDP}_t(\tau)$  in  $\mathcal{L} \neq$  stationary then
6:   Remove the last  $n_{DISC}$  elements of  $\mathcal{L}$ 
7:   Go to step 4
8: else
9:    $\alpha \leftarrow \alpha - \mathcal{L}$  ▷ Difference set operator
10:  if  $|\alpha| < \beta$  then
11:    return  $\mathcal{SR}_1, \dots, \mathcal{SR}_{N_{SR}}$ 
12:  else
13:     $\mathcal{SR}_k \leftarrow \mathcal{L}$ 
14:     $k \leftarrow k + 1$ 
15:    Go to step 2

```

correlation matrices are used, and in [Geh+01] the PDP. The idea behind the algorithms previously cited is to compare sliding portions of the 'summary function' to check when the similarity passes or undergoes a given threshold.

With an exploratory intention, we propose a different methodology that is explained in Algorithm 2 and then is contrasted with Algorithm 5 based on [Ren+10]. The methodology that we proposed, tries to exploit the fact that each stationary region can be seen as a cluster whose members are 'summary functions' at different time samples. The benefit gained through this approach is that simple clustering algorithms are already optimally deployed in all the main programming languages, and so, computing time for the stationary regions could be reduced.

Key remarks to better understand Algorithm 2 are enumerated below:

1. The inputs of the algorithm are: the initial set of time-snapshots $\mathcal{T}_0$, the number of samples to discard n_{DISC} , and the minimum set size β .
2. The underlying idea is to split a set of PDPs into 2, until the leftmost group is said to be stationary. When stationary condition is met, save the stationary region and do the process again with the remaining set of PDPs.
3. The pivot point p that splits the set under consideration at a given iteration of the algorithm, is chosen to be the median but other choices

are available. Non-equal splitting at the beginning of an iteration is possible, if for instance a priori information is known about stationary regions.

4. The stationary test for the PDP must reveal if the signal $\text{PDP}_t(\tau)$ is stationary over the snapshots in $\mathcal{L}$. If the available $|\mathcal{L}|$ PDP signals are considered as function realizations of the stochastic process, then for each delay tap τ we have a random variable X_τ . If the size of the set $|\mathcal{L}|$ is big enough, sample statistics as the mean and covariance can be computed for each random variable $X_{\tau_1}, \dots, X_{\tau_{N_\tau}}$, with N_τ being the number of delay taps. However, such naive approach will give us the averaged PDP over the region $\mathcal{L}$, for which we are testing stationarity! Moreover, we know the mean will vary: the mean of each X_{τ_i} is nothing but the PDP over $\mathcal{L}$, which we know beforehand is not a constant signal, but a 'peaky' signal.

Different approaches might be used to test stationarity. The underlying idea is to test how dissimilar are the instantaneous $\text{PDP}_{t_i}(\tau)$ for the set of time snapshots $t_1, \dots, t_{|\mathcal{L}|}$ in $\mathcal{L}$. Algorithm 3 provides a way of testing such a situation. The idea is that for a given set of snapshots $\mathcal{L}$, we cluster the PDP signals of those snapshots into 2 clusters. If cluster members and centroids are very closed in some distance sense, then both clusters are actually one cluster, meaning the signals can be thought to belong to the same stationary region. There is no need to partition the signals in more than 2 clusters, since it is clear that if all observations are close for 2 clusters, they will be closed for more than 2 clusters. The proper choice of the distance is important in general. In this case, since the signals all vary within the same scale and units, and even more, their shapes are similar, RMSE is used. Thus, the measure d_{CROSS} in Algorithm 3 is giving a sense of how close are the two clusters. It does so by computing the 'cross distances' between members of one cluster and the centroid of the other cluster.

A question that may arise is: why not cluster from the beginning the whole set of PDPs, and take the indexes of the clusters as the stationary regions? For doing so, algorithms like kmeans could not be used, or they have to be modified, since the resultant groups do not guarantee their time-snapshot members are consecutive. Approaches like the proposed by Hallac *et al.* [Hal+17] allow the time-snapshots members of a cluster to be consecutive. In Chapter 4, more discussion about clustering algorithms is provided.

5. It is assumed that reducing the size of the set $\mathcal{L}$ will achieve stationarity. In other words, the smaller the size of the set under test, the higher the

Algorithm 3 Test stationarity through clustering.**Input:** $\text{PDP}_t(\tau)$, $\mathcal{L}$, THR**Output:** True or False

-
- 1: Gather each PDP_t as a row of a matrix $\mathbf{X}$
 - 2: Standardize (i.e. minimax) $\mathbf{X}$
 - 3: Cluster $\mathbf{X}$ through kmeans, assuming there are two clusters
 - 4: Group observations of cluster 1 in $\mathcal{G}_1$
 - 5: Group observations of cluster 2 in $\mathcal{G}_2$
 - 6: Add all 'cross' distances between $\mathcal{G}_1$ and centroid 2 and save them in $d_{\text{CROSS},1}$
 - 7: Add all 'cross' distances between $\mathcal{G}_2$ and centroid 1 and save them in $d_{\text{CROSS},2}$
 - 8: $d_{\text{CROSS}} \leftarrow \frac{d_{\text{CROSS},1} + d_{\text{CROSS},2}}{|\mathcal{G}_1| + |\mathcal{G}_2|}$
 - 9: **if** $d_{\text{CROSS}} < \text{THR}$ **then**
 - 10: return True
 - 11: **else**
 - 12: return False
-

probability of PDP being stationary within its snapshot members.

6. The parameter n_{DISC} controls how fast stationarity is checked in the set under test. It is thus a kind of *learning rate*. Lower values slower down the computing time of the sets, but higher values can make smaller stationary regions than what they are. An adaptive learning rate would be desirable, however, no analytical expression for this learning rate in this context is developed or available. If the algorithm that test stationarity provides a measure for such stationarity or similarity of the PDPs in $\mathcal{L}$, then n_{DISC} can be increased each time such a measure changes very slowly, and vice-versa.
7. Parameter β is controlling the minimum set size in which test stationarity for the PDP. It implicitly means, in this methodology, that we don't know if the last β time-snapshot belong to a stationary region or not. Or even, if they belong to multiple stationary regions. This last affirmation, however, is more unlikely in the case when β is chosen to be the equivalent of samples of the local area defined in section 3.2.

Algorithm 3 does not guarantee that reducing the size of $\mathcal{L}$ the value of d_{CROSS} will reduce monotonically. In fact, the value of d_{CROSS} could vary in a semi-random fashion when the starting set $\mathcal{L}$ is big enough. This is due to the fact that within the initial set there might be actually multiple stationary regions. In other words, as the set is big, stationary conditions might

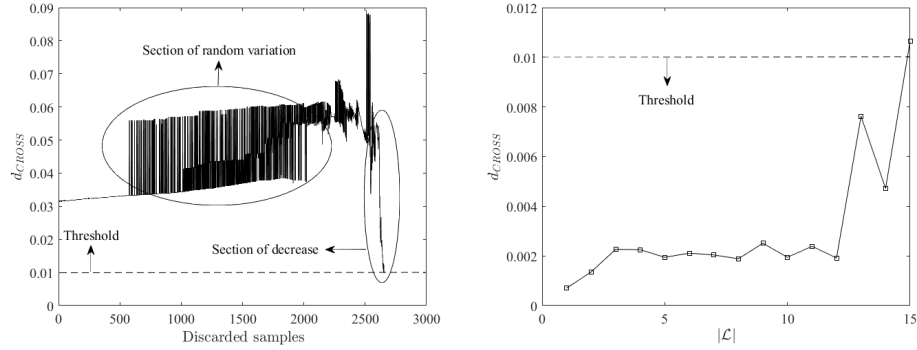


Figure 3.19: Left panel: example of variation of d_{CROSS} for one iteration of Algorithm 2, until its value is below the desired threshold. Right figure: example of variation of d_{CROSS} for one iteration of Algorithm 4, until the threshold value is surpassed. For both cases, during each iteration, one different stationary region is found.

change multiple times. Left panel in Figure 3.19 exemplifies this behaviour for an initial set size $|\mathcal{L}| = 2659$. Initial values of d_{CROSS} are random, while as the number of samples discarded grows (and so, $|\mathcal{L}|$ decreases) d_{CROSS} shows a slight tendency to decrease as well, until it reaches the region where definitively start to decrease. This behaviour is making computations in Algorithm 2 very slow. During each iteration, the measure d_{CROSS} used against the threshold is fluctuating, due to the big size of the initial groups chosen in step 3. So, we need to wait until set sizes are small enough, before actually d_{CROSS} starts to decrease. In order to avoid such a situation, the algorithm is revised and a 'reverse' process is followed: starting from the smallest set size possible (2), the size is increased until a threshold is surpassed. The procedure is explained in Algorithm 4. Right panel of Figure 3.19 shows this behaviour. Now d_{CROSS} has a clear increasing tendency and so, the set size of the stationary region under consideration is achieved faster, when d_{CROSS} surpass the threshold.

Figures 3.20, 3.21, 3.22, 3.23, 3.24 and 3.25 show the stationary regions found using Algorithm 4 for 3 different threshold levels. The regions exhibit different sizes and locations when the threshold value is changed. Since we rely on kmeans, and its initialization scheme places randomly the first centroids, stationary regions are different each time they are computed using the same inputs for the algorithm. To reduce the impact of this situation, multiple initializations are used, and the solution with the lowest within-cluster-sum of point-to-centroid distances is chosen.

Notice that most of the stationary regions of all this figures are within the range of $2 - 10 \lambda$, except for 1 or 2 regions considerably bigger than the average size. Consider the mentioned figures by threshold level. Two things

Algorithm 4 Revised version of Algorithm 2.

Input: $\text{PDP}_t(\tau) \forall t \in \mathcal{T}_0, \mathcal{T}_0, n_{DISC}, \beta$ **Output:** $\mathcal{SR}_1, \dots, \mathcal{SR}_{N_{SR}}$

```

1:  $\alpha \leftarrow \mathcal{T}_0, k \leftarrow 1$ 
2:  $\mathcal{L} \leftarrow [\alpha]_{1:2}$ 
3: Test stationarity of  $\text{PDP}_t(\tau)$  for the time-snapshots in  $\mathcal{L}$ 
4: if  $\text{PDP}_t(\tau)$  in  $\mathcal{L} \neq \text{stationary}$  then
5:   Increase  $\mathcal{L}$  by the next  $n_{DISC}$  elements of  $\alpha$ 
6:   Go to step 3
7: else
8:    $\alpha \leftarrow \alpha - \mathcal{L}$  ▷ Difference set operator
9:   if  $|\alpha| < \beta$  then
10:    return  $\mathcal{SR}_1, \dots, \mathcal{SR}_{N_{SR}}$ 
11:  else
12:     $\mathcal{SR}_k \leftarrow \mathcal{L}$ 
13:     $k \leftarrow k + 1$ 
14:    Go to step 2

```

can be noticed when the threshold level is increased. First, some of the small regions increase their size proportionally to the threshold increment. This size increment is reasonable since we are allowing more dissimilar PDPs to merge. Second, some big regions behave in an unexpected manner as the threshold increases. It can be seen, for example, in the Parking Lot routes for the Mercator BS: for the first increment the 2 bigger regions merge, and for the second increment they split again. Under the assumption that d_{CROSS} measures stationarity correctly, one might expect that increasing the threshold level implies small regions will merge. How many would merge can't be known beforehand, but it is intuitively clear that they will merge. On the other hand, it is not expected that with increasing threshold level, the bigger stationary regions split into smaller regions. This merging-splitting situation for some big regions, besides the appearance of stationary regions extremely big (as in one of the Saint Barbe routes with Maxwell BS and threshold level 0.035) are showing a flaw in the algorithm.

In spite of being easy to implement, Algorithm 4 might detect false stationary regions (improbable long stationary regions). An alternative to it can be implemented by using the concepts explained in section 3.2.1 and based on [Ren+10]. Spatial correlation matrices provide information about the channel impulse response and can be used as '*summary functions*'. We can assume that for a window w of time snapshots ergodicity holds, and estimate the full spatial correlation matrix for such w . Then, compare 2 spatial correlation matrices estimated from consecutive time windows, until the CMD between

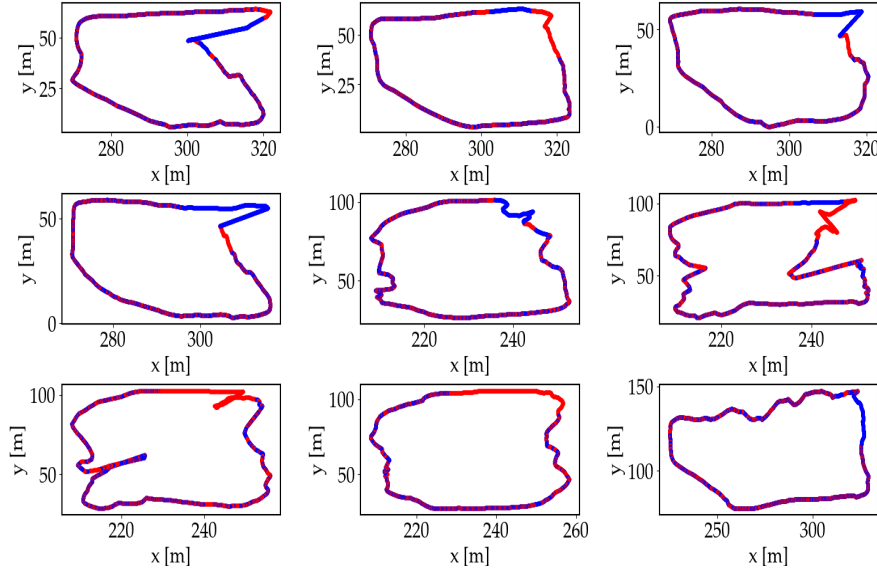


Figure 3.20: Stationary regions found using Algorithm 4 for all the measurements and routes with Mercator BS. Threshold level is set to 0.025. The 2 colors used (blue and red) don't have by themselves any meaning, but are used to differentiate one region from the next one.

those matrices undergoes a given threshold. Figure 3.27 illustrates the time sliding concept, and Algorithm 5 describes all the computing steps involved.

New founded stationary regions using this algorithm are shown in Figures 3.29 and 3.28 for a CMD of 0.7. False big stationary regions are not detected any more, and some previously found regions (highlighted with black ellipses in the figures) remain the same. For the Mercator BS, all the Parking Lot routes have their biggest stationary regions located in the right corner of the route. As this region is located behind the Saint Barbe building, it is indicating that propagation there is mostly occurring through the rooftop of Saint Barbe building, as otherwise, losses due to the Saint Barbe building attenuate completely the radio waves. The Saint Barbe and Marie Curie routes show large stationary regions for zones where there are either LoS conditions or OLoS due to presence of trees between Mercator BS and the transmitter. Since the measurements were taken at the end of september, foliage of the trees was still an attenuating object.

On the other hand, for the Maxwell BS, LoS propagating conditions in Saint Barbe route are the responsible for most of the highlighted regions for that route. For the region highlighted at the bottom right of the Saint Barbe routes, it is probable that radio waves are bouncing repeated times between Saint Barbe and Vinci buildings, creating a similar effect of what is observed

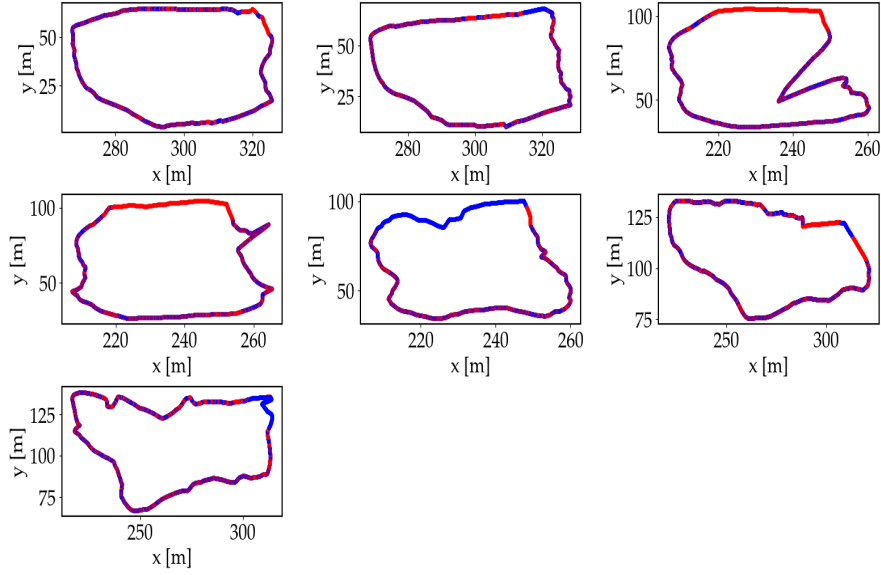


Figure 3.21: Stationary regions found using Algorithm 4 for all the measurements and routes with Maxwell BS. Threshold level is set to 0.025. The 2 colors used (blue and red) don't have by themselves any meaning, but are used to differentiate one region from the next one.

in a tunnel; it is thus a stationary region over which the mechanism of propagation remain the same. This assumption is checked in Figure 3.26. Path Loss estimation methodology of section 3.2 is applied to the red zone highlighted in the right panel in the figure. The *closed-in* linear model fit to the path loss comprising the received power measured for both red zones is shown in the left panel. The PLE is 2.02, indicating a clear 'tunnel' effect, as its value is close to free space PLE. For the Marie Curie routes, it is more challenging to know which propagation mechanisms are producing the longer stationary regions highlighted by the ellipses. There might be some rooftop diffraction through Saint Barbe and close buildings (Pierre Curie, Louvain-la-Neuve's Museum), combined with mutiple bounces between the same buildings and Rue Archimède.

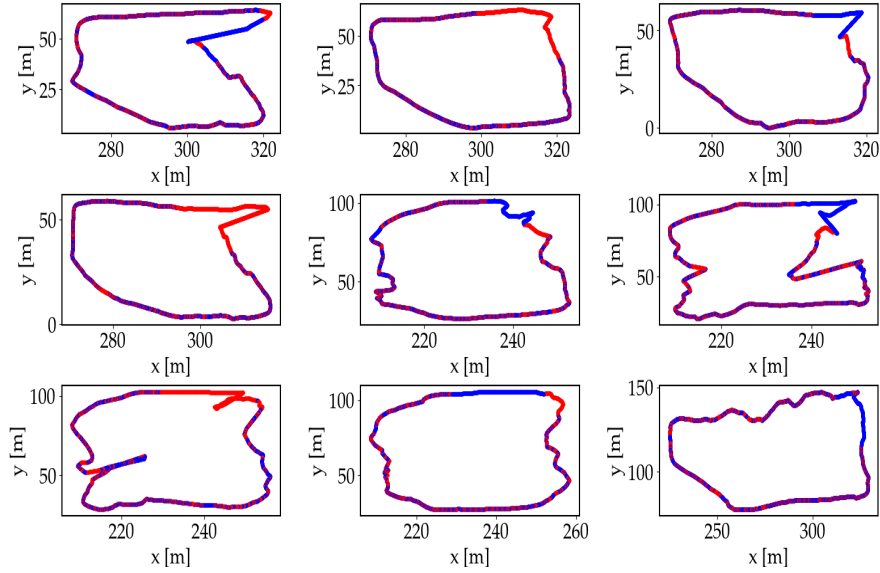


Figure 3.22: Stationary regions found using Algorithm 4 for all the measurements and routes with Mercator BS. Threshold level is set to 0.03. The 2 colors used (blue and red) don't have by themselves any meaning, but are used to differentiate one region from the next one.

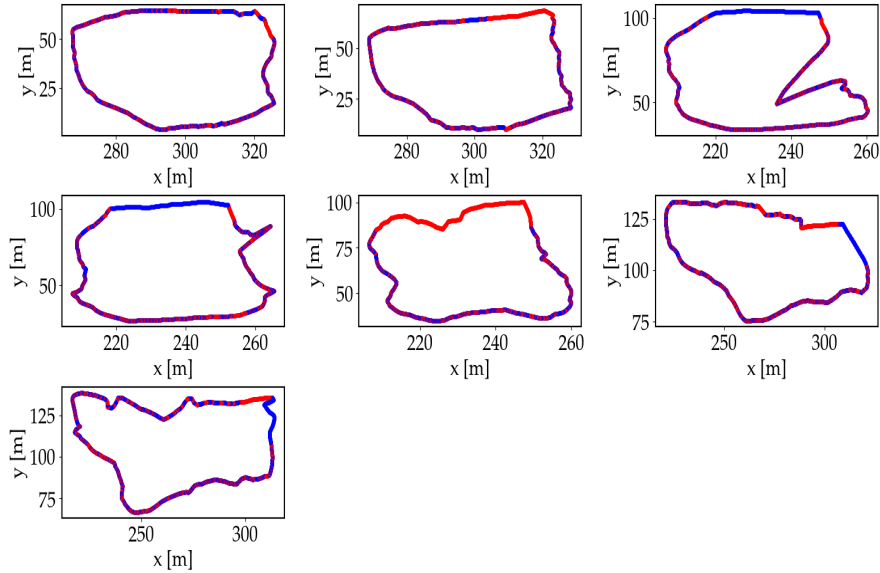


Figure 3.23: Stationary regions found using Algorithm 4 for all the measurements and routes with Maxwell BS. Threshold level is set to 0.03. The 2 colors used (blue and red) don't have by themselves any meaning, but are used to differentiate one region from the next one.

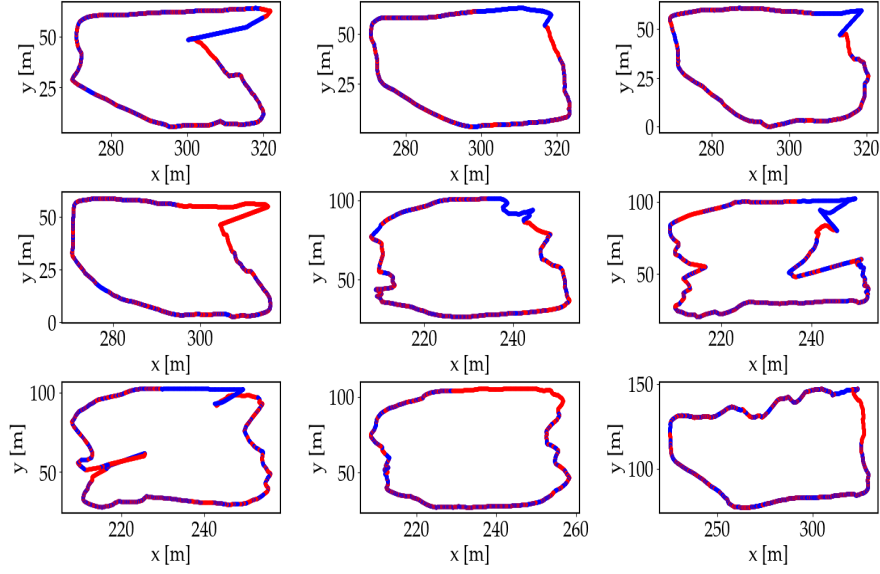


Figure 3.24: Stationary regions found using Algorithm 4 for all the measurements and routes with Mercator BS. Threshold level is set to 0.035. The 2 colors used (blue and red) don't have by themselves any meaning, but are used to differentiate one region from the next one.

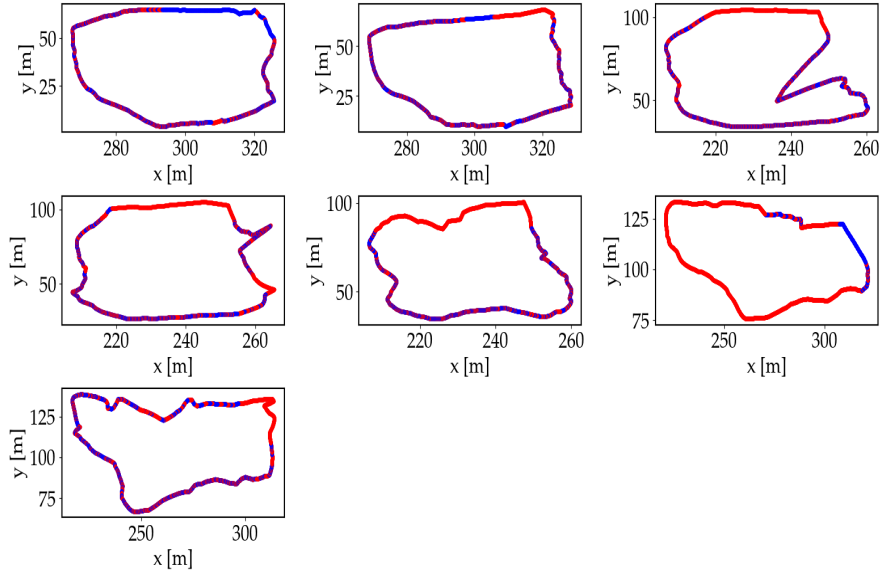


Figure 3.25: Stationary regions found using Algorithm 4 for all the measurements and routes with Maxwell BS. Threshold level is set to 0.035. The 2 colors used (blue and red) don't have by themselves any meaning, but are used to differentiate one region from the next one.

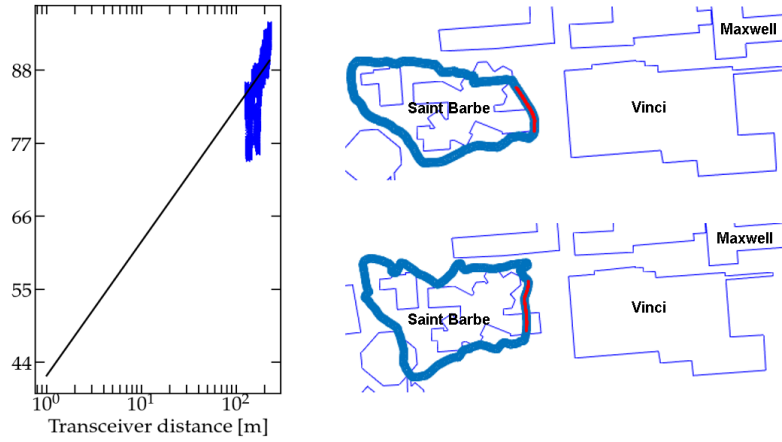


Figure 3.26: Left panel: Linear regression fit for the path loss of the zone highlighted on the right. Right panel: the 2 Saint Barbe routes used with Maxwell BS shown in blue. The red zone is the region over which the 'tunnel' effect is occurring.

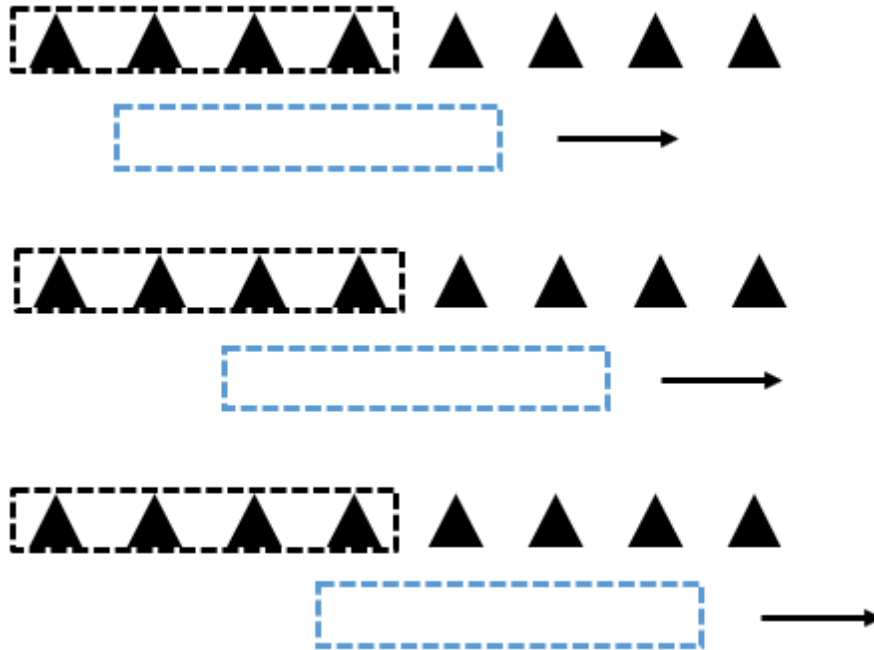


Figure 3.27: Schematic representation on the time intervals used to compute spatial correlation matrices in Algorithm 5. Each rectangle represents the time window over which spatial correlation matrices are estimated. The time window for one of the 2 spatial correlation matrices slides to the right each iteration of the Algorithm 5. The triangle represents channel impulse response matrices H_t .

Algorithm 5 Methodology to compute stationary regions from spatial correlation matrix.

Input: $\mathbf{H}_t \in \mathbb{C}^{8 \times 8}$, $t \in \mathcal{T}_0$, w , THR, n_{STEP} , $\mathcal{T}_0$

Output: $\mathcal{SR}_1, \mathcal{SR}_2, \dots, \mathcal{SR}_{N_{SR}}$

```

1:  $t_1 \leftarrow 1, k \leftarrow 1$ 
2:  $t_2 \leftarrow t_1 + n_{STEP}$ 
3: if  $t_2 + w - 1 < |\mathcal{T}_0|$  then
4:   Get  $\mathcal{T}_1 = \{t_1, t_1 + 1, \dots, t_1 + w - 1\}$ 
5:   Get  $\mathcal{T}_2 = \{t_2, t_2 + 1, \dots, t_2 + w - 1\}$ 
6:   Compute:  $\hat{\mathbf{R}}_1 = \frac{1}{w} \sum_{t \in \mathcal{T}_1} \text{vec}(\mathbf{H}_t) \text{vec}(\mathbf{H}_t)^H$ 
7:   Compute:  $\hat{\mathbf{R}}_2 = \frac{1}{w} \sum_{t \in \mathcal{T}_2} \text{vec}(\mathbf{H}_t) \text{vec}(\mathbf{H}_t)^H$ 
8:    $\alpha \leftarrow c(\hat{\mathbf{R}}_1, \hat{\mathbf{R}}_2)$  (Equation 3.6)
9:   if  $\alpha < \text{THR}$  then
10:     $\mathcal{SR}_k \leftarrow \{t_1, t_1 + 1, \dots, t_2\}$ 
11:     $t_1 \leftarrow t_2$ 
12:     $k \leftarrow k + 1$ 
13:    Go to step 2
14:   else
15:     $t_2 \leftarrow t_2 + n_{STEP}$ 
16:    Go to step 4
17: else
18:   return  $\mathcal{SR}_1, \dots, \mathcal{SR}_{N_{SR}}$ 

```

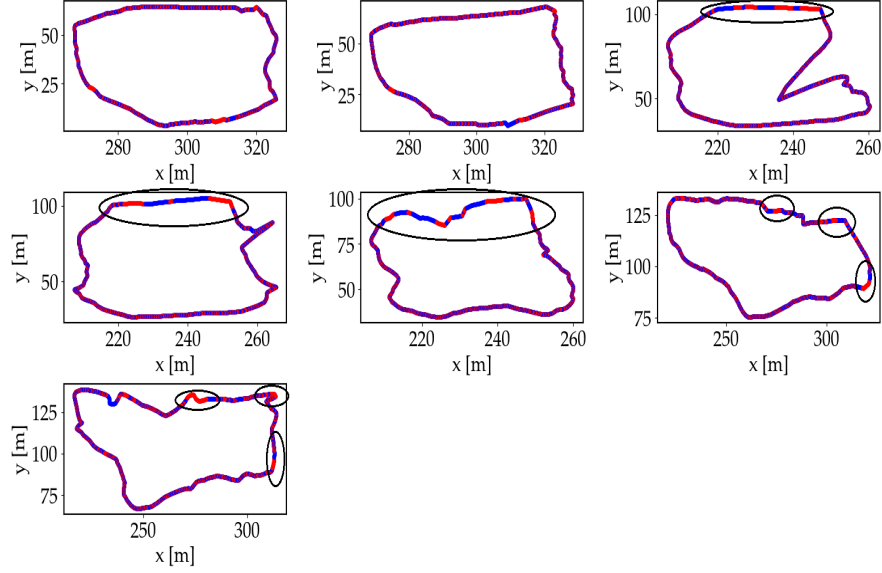


Figure 3.28: Stationary regions found using Algorithm 5 for all the measurements and routes with Maxwell BS. Threshold level for the CMD is set to 0.7. The 2 colors used (blue and red) don't have by themselves any meaning, but are used to differentiate one region from the next one.

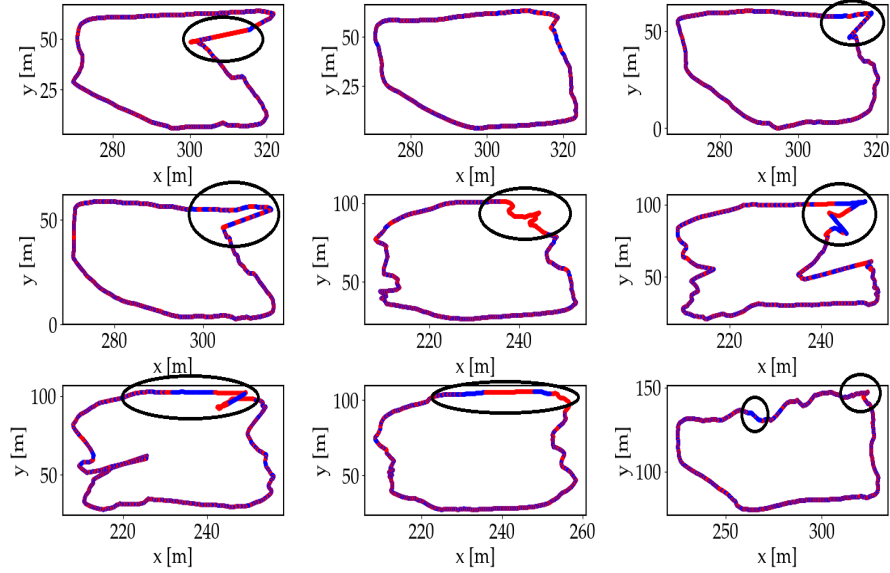


Figure 3.29: Stationary regions found using Algorithm 5 for all the measurements and routes with Mercator BS. Threshold level for the CMD to 0.7. The 2 colors used (blue and red) don't have by themselves any meaning, but are used to differentiate one region from the next one.

3.5.2 Statistics of the PDP

The stochastic nature of the PDP make necessary the use of summary statistics to characterize its random behaviour in long runs of time, or over many experiments. Due to their mathematical simplicity and enough-information provided, the first two moments are typically used to characterize the environment under consideration. From the PDP, the maximum excess delay is computed as the difference between the maximum (τ_{MAX}) and minimum delay (τ_{MIN}) for which the PDP signal remains above certain threshold. It defines the interval

$$\mathcal{T}_{EXC} = [\tau_{MIN}, \tau_{MAX}] \quad (3.23)$$

In other words, it is the time-delay interval during which signals with significant energy arrive. The interval is used in the computations of the first (3.24) and second moments (3.25) of the PDP signal.

$$\tau_A = \frac{\int_{\mathcal{T}_{EXC}} \tau PDP(\tau) d\tau}{\int_{\mathcal{T}_{EXC}} PDP(\tau) d\tau} \quad (3.24)$$

$$\tau_S = \frac{\int_{\mathcal{T}_{EXC}} (\tau - \tau_m)^2 PDP(\tau) d\tau}{\int_{\mathcal{T}_{EXC}} PDP(\tau) d\tau} \quad (3.25)$$

From both moments, the second central moment τ_s , better know as the *delay spread*, has a special relevance, not only as a summarizing parameter but because it plays a role in error probability computations due to delay dispersion (Chapter 12 in [Mol12]). Notice that the PDP used in 3.24 and 3.25 is an average PDP over which stationarity is guaranteed. Such stationarity is provided by the stationary regions $\mathcal{SR}_1, \dots, \mathcal{SR}_{N_{SR}}$ computed in section 3.5.1.

Figure 3.30 shows a scatter plot with the τ_A and τ_S computed for all the measurements carried out. Each data point in both graphs represents a pair (τ_A, τ_S) computed using the PDP of a stationary region of a particular BS and route. Algorithm 5 was used to compute the stationary regions. For the Mercator BS (left panel) most of the paths propagate with an average path duration close to either $1\mu s$ or $2\mu s$, and delay spreads below 100 ns and around $1.5\mu s$, respectively. For the Saint Barbe route it is clear that most of the propagation is occurring in NLoS conditions as the distance between the transmitter and receiver ranges between 350 - 650 ns, approximately, and mean delays for the stationary regions are above this interval.

The ratio

$$\frac{\tau_S}{\tau_A} \quad (3.26)$$

measures how much paths deviate from the average path length, which in turn could be seen as a measure of much propagation paths, (and thus, propagation mechanisms) is changing. The reason is that although paths with similar delays *don't* necessary arise from the same mechanism of propagation, however, for NLoS conditions, since radio signals are highly attenuated due to buildings (and obstacles in general), signals arriving with strong enough power must exhibit similar paths, and thus similar delays. For the Saint Barbe route this ratio never surpasses 60%, nearly for a third of the regions stays below 10%, and for the rest stays close to 60%, indicating that although most of the propagation conditions are NLoS, there are some stationary regions with very similar mean delays, mostly probable due to the same path lengths arising from OLoS. For the Parking Lot routes the ratio ranges between 60% and 70% for most of the stationary regions (more than 60% of the stationary regions) and below 10% for 20% of the stationary regions. This is consistent with the fact that this route is on average the farthest route to Mercator BS, and also the more obstructed by buildings, having as a consequence, NLoS conditions for all the stationary regions. For the Marie Curie routes, the ratio behaviour is similar to the Saint Barbe route, as a consequence of two things: first, Marie Curie is on average a bit more far away to Mercator BS (122 - 200 m) than Saint Barbe route (97 - 191 m), but not as much as Parking Lot (180 - 250 m); second, both routes share a zone for which there are OLoS conditions.

What comes as a surprise are the higher mean delays for Saint Barbe route compared to those for the other routes, if it is recalled that this route is the closest on average to the Mercator BS. The high delays are coming for the stationary regions in Place Saint Barbe, the corridor SB-PL and in most of the Rue Archimede (before Place EPL). Those regions are clearly in NLoS conditions, and so high mean delays are expected. However, the Parking Lot routes have on average the lowest mean delays, in spite of being the farthest route to Mercator BS. No simple explanation is available for this. The received powers in Figure 3.1, for the mentioned Saint Barbe stationary regions, showed lower values (around -72 dB) than the mean power for the Parking Lot routes (around -67 dB). This is consistent with the delay average values, since for longer delays (longer path lengths) losses are higher. It is quite clear that energy for the Parking Lot route is propagating through the rooftop of the Saint Barbe route, and since its location is towards the direction of the Mercator BS, this might explain somehow the lower delays: while for the Saint Barbe route, excluding the OLoS zone in Place EPL, most of the radio signals are following non-straight paths, where combined propagation mechanisms as

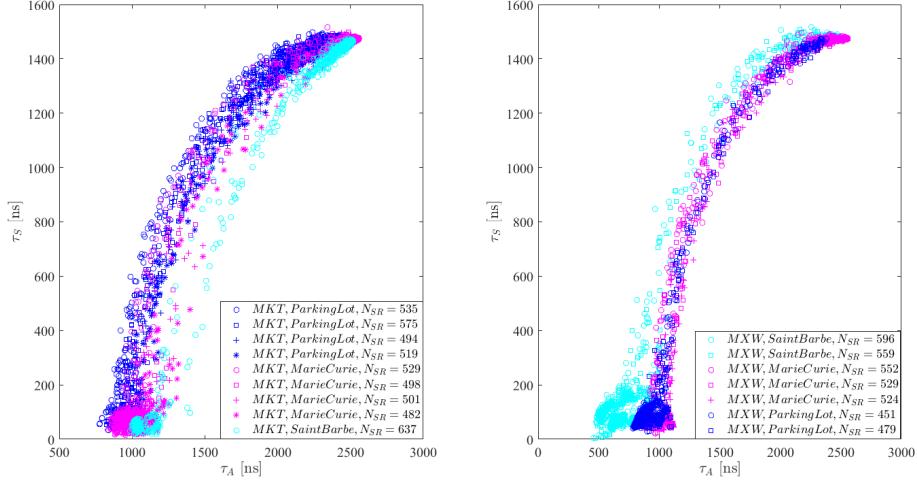


Figure 3.30: Scatter plot of mean delays and delay spreads for the stationary regions computed from Algorithm 5 with CMD = 0.7. Left panel: routes with Mercator BS. Right panel: routes with Maxwell BS.

diffraction and reflection are occurring.

On the other hand, for the Maxwell BS (right panel in figure 3.30) mean delays behave in an expected manner. The Saint Barbe routes have the lowest mean delays due to their close proximity to the Maxwell BS, and to the fact that there are multiple stationary regions with LoS or OLoS conditions, as well as the 'tunneling' effect happening in Place Saint Barbe between Vinci and Saint Barbe buildings. Marie Curie and Parking Lot have very similar pairs (τ_A , τ_S), indicating mechanisms of propagation might be similar (diffraction through rooftops and bounces arriving through Place Saint Barbe).

Tables 3.5, 3.6 and Figures 3.33, 3.34, 3.31 and 3.32 summarize statistics of the delay spreads and mean delays for the stationary regions of all routes and BS using Algorithm 5. The CDF figures and the Kolmogorov-Smirnov tests show that the probability distributions fitted do not match the empirical distribution of the delays. This is because the delay spreads and mean delays are mostly grouped into two clusters (noticeable from Figure 3.30), meaning that the empirical PDF has 2 peaks, shape that none of the distributions can follow.

Table 3.5: Delay spread statistics for the stationary regions in Figures 3.29 and 3.28 with CMD = 0.7.

PL	BS	Med. ^a	Std. ^b	BGF	KS-S.	CV	Loc.	Sca.	Shape
1	MKT	1347	591	Beta	0.16	0.058	33.4	1464	$a = 0.42, b = 0.3$
2	MKT	1312	585	Beta	0.17	0.056	41.9	1456	$a = 0.46, b = 0.34$
3	MKT	1300	587	Gaussian	0.24	0.06	937	587	-
4	MKT	1353	592	Gaussian	0.25	0.059	959	592	-
1	MXW	84	441	Beta	0.26	0.062	29.8	21812	$a = 0.86, b = 116.4$
2	MXW	82	450	Gamma	0.26	0.063	28.3	318	0.63
MC	BS	Med.	Std.	BGF	KS-S	CV.	Loc.	Sca.	Shape
1	MKT	1467	623	Gaussian	0.34	0.058	1024	623	-
2	MKT	1461	646	Beta	0.27	0.06	-188	1686	$a = 1, b = 0.37$
3	MKT	1455	637	Beta	0.28	0.06	-161.4	1657	$a = 1, b = 0.4$
4	MKT	1455	637	Beta	0.29	0.06	-157	1648	$a = 1, b = 0.4$
1	MXW	1465	600	Gamma	0.3	0.057	-8265	41.2	225.1
2	MXW	1464	571	Beta	0.3	0.059	-211	1697	$a = 0.77, b = 0.33$
3	MXW	1464	617	Gamma	0.32	0.06	-8483	41.4	229
SB	BS	Med.	Std.	BGF	KS-S	CV.	Loc.	Sca.	Shape
1	MKT	1382	609	Gaussian	0.3	0.05	976	609	-
1	MXW	125	609	Gamma	0.25	0.055	4.5	796	0.66
2	MXW	131	621	Gamma	0.24	0.057	4.5	805	0.67

^a Sample median [ns].^b Sample standard deviation [ns].

Table 3.6: Mean delay statistics for the stationary regions in Figures 3.29 and 3.28 with $\text{CMD} = 0.7$.

PL	BS	Med. ^a	Std. ^b	BGF	KS-S.	CV	Loc.	Sca.	Shape
1	MKT	1855	621	Gamma	0.17	0.058	-10036	33.4	350
2	MKT	1882	598	Beta	0.14	0.056	822	1637	$a = 0.8, b = 0.77$
3	MKT	1857	579	Gamma	0.16	0.06	-9563	30.3	372
4	MKT	2016	575	Gaussian	0.18	0.059	1781	575	-
1	MXW	912	399	Gamma	0.14	0.062	800	370	0.72
2	MXW	897	401	Gamma	0.15	0.063	784	293	0.91
MC	BS	Med.	Std.	BGF	KS-S	CV	Loc.	Sca.	Shape
1	MKT	2467	721	Beta	0.26	0.058	783	1774	$a = 1.42, b = 0.48$
2	MKT	2447	715	Beta	0.29	0.06	800	1758	$a = 1.5, b = 0.66$
3	MKT	2442	688	Gaussian	0.28	0.06	1926	688	-
4	MKT	2451	660	Beta	0.27	0.06	990	1567	$a = 1.12, b = 0.46$
1	MXW	2322	683	Beta	0.2	0.057	806	1748	$a = 0.89, b = 0.39$
2	MXW	2363	653	Gamma	0.23	0.059	-7308	49	190
3	MXW	2368	671	Gaussian	0.24	0.06	1937	671	-
SB	BS	Med.	Std.	BGF	KS-S	CV	Loc.	Sca.	Shape
1	MKT	2312	597	Beta	0.23	0.053	987	1565	$a = 0.58, b = 0.32$
1	MXW	760	728	Gamma	0.14	0.055	457	783	0.88
2	MXW	782	737	Gamma	0.14	0.057	473	800	0.92

^a Sample median [ns].^b Sample standard deviation [ns].

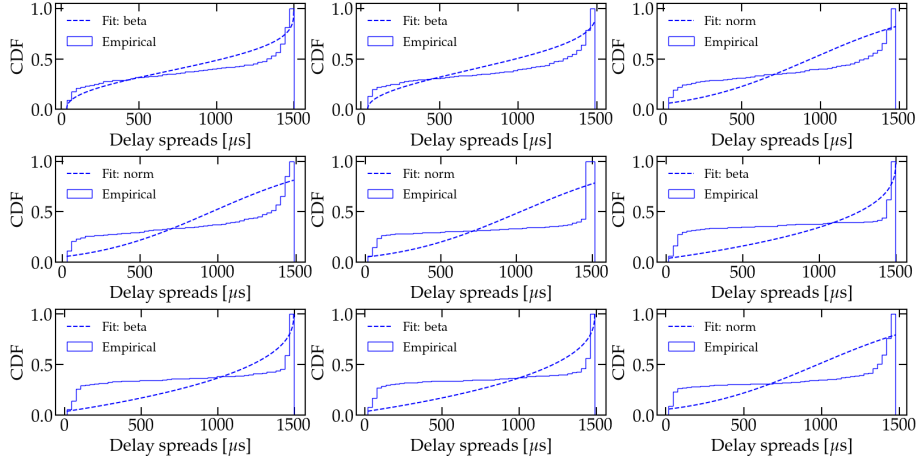


Figure 3.31: Empirical CDF of the delay spreads for all the measurements with Mercator BS. The solid black line represents the BGF distribution from the Kolmogorov-Smirnov test with the parameters shown in Table 3.5. Stationary regions computed using Algorithm 5 with $\text{CMD} = 0.7$.

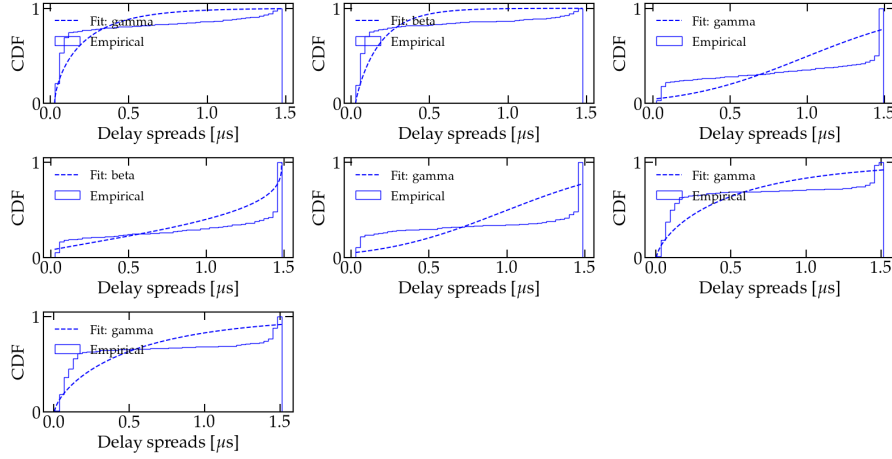


Figure 3.32: Empirical CDF of the delay spreads for all the measurements with Maxwell BS. The solid black line represents the BGF distribution from the Kolmogorov-Smirnov test with the parameters shown in Table 3.5. Stationary regions computed using Algorithm 5 with $\text{CMD} = 0.7$.

3.5.3 Angular Statistics

Circular statistics take into consideration the fact that the dataset is composed by angles. There are different measures for central tendency and for dispersion well established [Ber09]. Measures of central tendency are the mean angle vector (3.27), mean angle (3.28), among others. Measures of dispersion

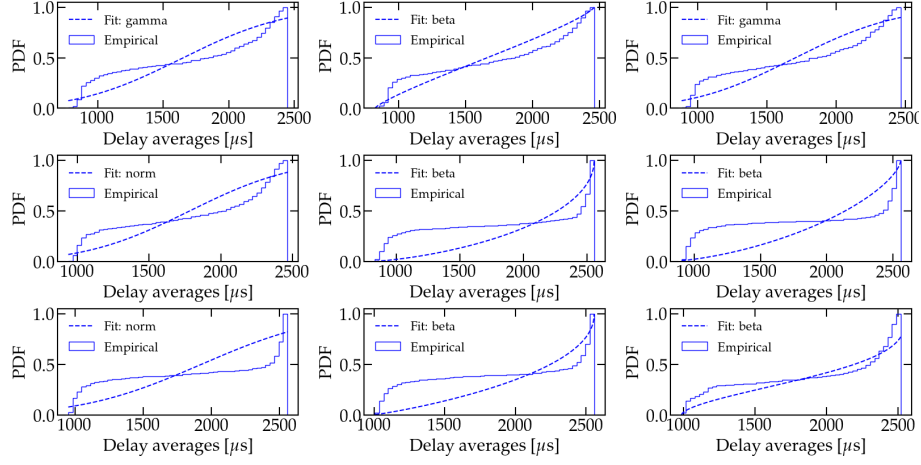


Figure 3.33: Empirical CDF of the mean delays for all the measurements with Mercator BS. The solid black line represents the BGF distribution from the Kolmogorov-Smirnov test with the parameters shown in Table 3.6. Stationary regions computed using Algorithm 5 with CMD = 0.7.

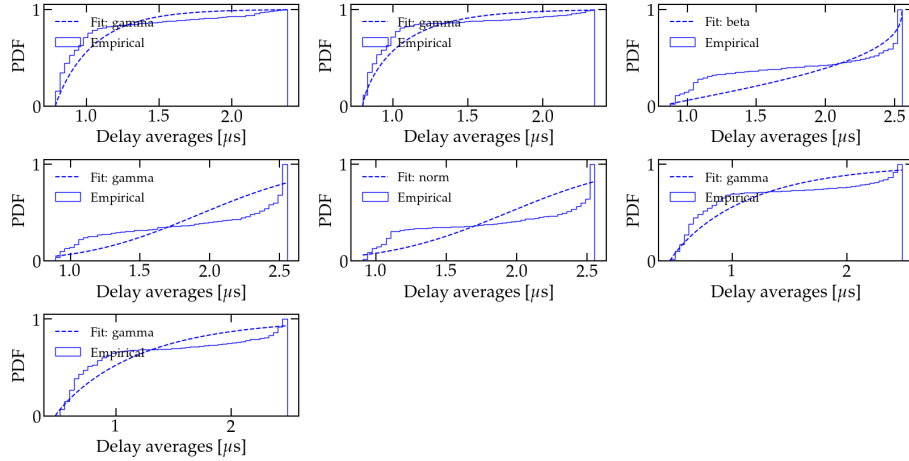


Figure 3.34: Empirical CDF of the mean delays for all the measurements with Maxwell BS. The solid black line represents the BGF distribution from the Kolmogorov-Smirnov test with the parameters shown in Table 3.6. Stationary regions computed using Algorithm 5 with CMD = 0.7.

are closely related with the length of the mean angle vector: the variance (3.29) and multiple definitions for the standard deviation (3.30).

$$\theta_A \in \mathbb{C} = \sum_l w_l \exp(i\theta_l), \quad R_\theta \in [0, 1] = |\theta_A| \quad (3.27)$$

$$\theta_A = \arctan \left(\frac{\text{Imag} \{ \theta_A \}}{\text{Re} \{ \theta_A \}} \right) \quad (3.28)$$

$$S_\theta \in [0, 1] = 1 - R \quad (3.29)$$

$$s_\theta^0 \in [0, \sqrt{2}] = \sqrt{2(1 - R)}, \quad s_\theta^1 \in [0, \infty] = \sqrt{-2 \log R} \quad (3.30)$$

with weights w_l , $\sum_l w_l = 1$. Within the context of radio propagation, Fleury introduced angular measures of dispersion slightly different from the ones in (3.30) [Fle00]. For azimuth propagation, the root second central moment is computed as in (3.31). Both, the length of the mean angle vector and Fleury's angle spread are related by $\theta_S = \sqrt{1 - R^2}$.

$$\theta_S \in [0, 1] = \sqrt{\sum_l |\exp(i\theta_l) - \theta_A|^2 w_l} \quad (3.31)$$

where the weights w_l refer to the Power Angular Profile, or the estimated power gains from SAGE algorithm. Figures 3.35 and 3.36 show the mean angles and angle spreads for the AoAs and AoDs of each stationary region of Saint Barbe route with Maxwell and Mercator BS. The AoDs angle spreads are higher since propagation paths originate from many directions at the transmitter. Mean angles for both the AoAs and AoDs concentrate around the direction pointing to the center of the trajectory or to the BS. There is a clear cluster for the AoDs of the Mercator BS. Since the spreads are high (nearly 1), this means propagation paths departure from many directions, and thus the mean will remain the same. The spreads and mean angles distribution is showing the same information of what was shown in the empirical distribution plots for the estimated AoAs and AoDs in Chapter 2. More useful information might be obtained by looking for statistical moments on specific regions of each route.

With this in mind, Saint Barbe routes for each BS are partitioned in 4 sections depicted in Figure 2.14. For each of this propagation regions the empirical PDF is plotted for the AoAs in Figures 3.37 and 3.38. The highest peak in each graph corresponds to the direction pointing to the center of the propagation region. However, for the Mercator BS the NLoS propagation regions behave in a well defined manner: there is a high peak pointing towards the center of the propagation region, but there is another non-negligible peak, from which radio signals are coming. The second peak is not pointing to the center of the trajectory, but rather than that to a zone where most of the propagation is coming. For NLoS region 1 both peaks point to the Réaumur

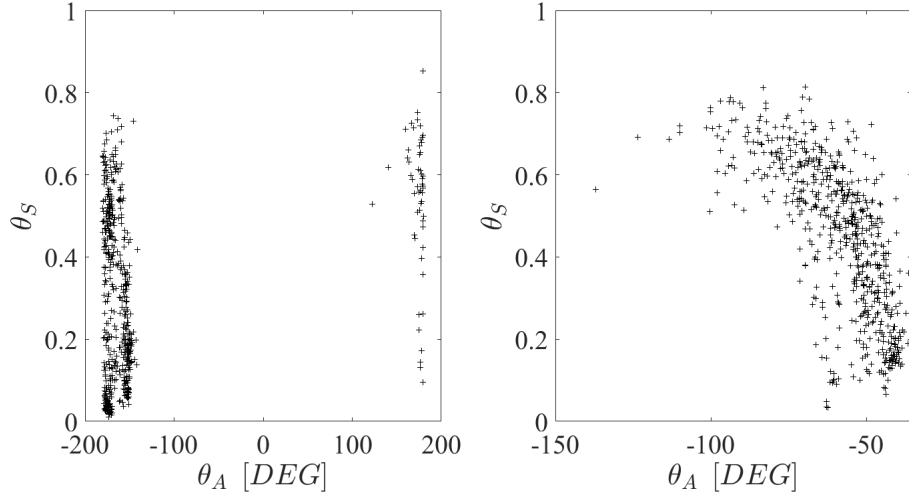


Figure 3.35: Scatter plot of θ_A and θ_S of the AoAs for the stationary regions computed from Algorithm 5 with $\text{CMD} = 0.7$. Left panel: Saint Barbe route with Maxwell BS. Right panel: Saint Barbe route with Mercator BS.

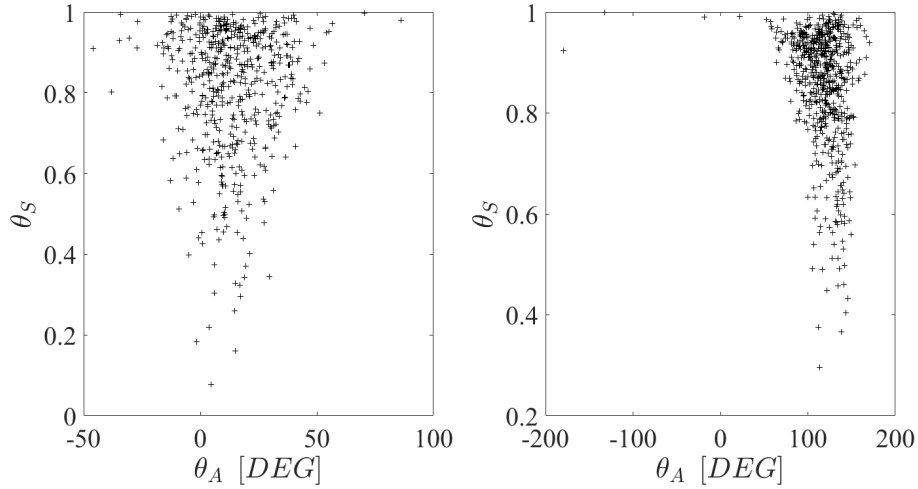


Figure 3.36: Scatter plot of θ_A and θ_S of the AoDs for the stationary regions computed from Algorithm 5 with $\text{CMD} = 0.7$. Left panel: Saint Barbe route with Maxwell BS. Right panel: Saint Barbe route with Mercator BS.

building, which might be indicative of rooftop propagation. For NLoS region 2, the second largest peak points to the Place Pierre Curie, which means propagation might be possible due to a combination of multiple reflections between the Pierre Curie and Bibliothèque des Sciences buildings and propagation through the rooftop of Saint Barbe building. This is precisely what

we encounter when the stationary regions were computed. Finally, the second peak of the NLoS region 3 points close to Réaumur and Vinci buildings. Here, it's hard to say what is happening. The propagation region considered to have a 'tunnel' effect for the Maxwell BS, has a corridor between Saint Barbe and Vinci buildings, as well as the corridor between Place Saint Barbe and Maxwell building. This is not the case for the Mercator BS, and so, no 'tunnel' effect might be occurring, and the second peak in the AoAs distribution could come from many IOs.

Since the distribution of the AoAs is multimodal it doesn't make sense to fit a unimodal distribution, as for example, the Von Mises distribution. The Von Mises probability distribution is defined as

$$p_{VM}(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta - \mu)) \quad (3.32)$$

where I_0 is the modified Bessel function of order zero.

On the other hand, for the Maxwell BS it doesn't exist a second peak in the empirical distribution, but rather than that, propagation comes with high probability from a direction pointing to the center of Saint Barbe route, and with lesser and lesser probability at angles away from such main peak. The MLE parameter estimations for the Von Mises PDF (3.32) are shown in Table 3.7. All the μ estimates point to the same direction, that is close to the center of the Saint Barbe route. The parameter κ that measures how concentrated is the distribution around the mean, is similar in all the regions (as also can be seen from the empirical histograms in 3.37) but NLoS 2 region. This is a consequence of the high probability peak around $-\pi$ rads, indicating a strong mechanism of propagation: in other words, while for the NLoS 1 region angles arrive from fewer more directions than the one pointing to the center of Saint Barbe route, for NLoS 2 region this doesn't occur. This is indicative that for NLoS 2 region the mechanisms of propagation are less diverse, as if the radio signals have found a strong mechanism of propagation through which more energy could arrived at the receiver. This, in combination with the fact that the transmitter is closer to the receiver explains why the received power increases by 5 dBm when passing from NLoS 1 to NLoS 2 region. In the context of this scenario, since the angular spread vary within a margin of 0.2 (Table 3.7), in relative terms, the propagation paths are 50% more concentrated around the mean angle in NLoS 2 with respect to the NLoS 1 region.

In Table 3.7 are also shown the mean angles θ_A and the angular spreads θ_S as defined by Fleury. The reference for all the angular values is the horizontal axis of the scenario (the longitude coordinate). Even though the Von Mises fits doesn't match all the data, the MLE estimate for μ is almost the same as the sample mean angle θ_A . Moreover, the spreads behaviour for the NLoS regions is similar to the behaviour of the κ estimated: lower spread (or

more concentration) for the NLoS 2 than for NLoS 1. It is expected that the tunnel region has a higher AoA spread since radio signals are bouncing in multiple parts and adding coherently at the receiver. Precisely, once assumed propagation during the Tunnel region behaves similarly to the propagation in a tunnel, the mechanism by how radio signals can reach longer distances than what might be expected, is due to the constructive superposition of radio signals bouncing at both sides of the 'corridor' or 'tunnel' at different angles. On the other hand, at a first glance, the high angular spread for the LoS is a surprise. One would expect a lower spread, assuming that there is only one source of propagation, precisely, the LoS. However, since the transmitter array radiates through different directions, radio signals will interact with the media through different mechanism of propagation, i.e. bouncing in the buildings. The difference is that the power of the radio signals coming from the LoS is noticeable higher than the power of signals arriving separately from other mechanisms. A look at the angular power distribution (Figure 3.37) for the Saint Barbe route with Maxwell BS confirms this. The reference line for the LoS graph is nearly the maximum power achieved at both NLoS propagation regions. The directions circled are those for which the power is above such reference power. For each of these highlighted directions the power is between 2 to nearly 4 times higher than the maximum power in the NLoS regions.

Table 3.7: Von Mises MLE estimated parameters for AoAs of Maxwell BS, and mean angles and angle spreads for each propagation region of both BS.

BS	Propagation region	μ	κ	θ_A	θ_S
MXW	LoS	-174	5	-175	0.63
MXW	NLoS 1	-167	6.8	-167	0.52
MXW	NLoS 2	-155	14.6	-157	0.43
MXW	Tunnel	-172	4	-176	0.62
BS	Propagation region	-	-	θ_A	θ_S
MKT	OLoS	-	-	-74	0.67
MKT	NLoS 1	-	-	-55	0.5
MKT	NLoS 2	-	-	-63	0.56
MKT	NLoS 3	-	-	-49	0.44

Meanwhile, the distribution of the AoDs for the Maxwell BS (Figure 3.40) is expected. A clear peak pointing to the BS, and an almost uniform probability of radiating in any other angle. This is confirmed by the high spreads θ_S and low κ parameter for the Von Mises distribution fit in Table 3.8. However, there is a remark to say. For the LoS and NLoS regions the peak in the empirical PDF is big, indicating there is a high probability that radio signals departure from the angle pointing to the BS (the LoS angle). This is not the

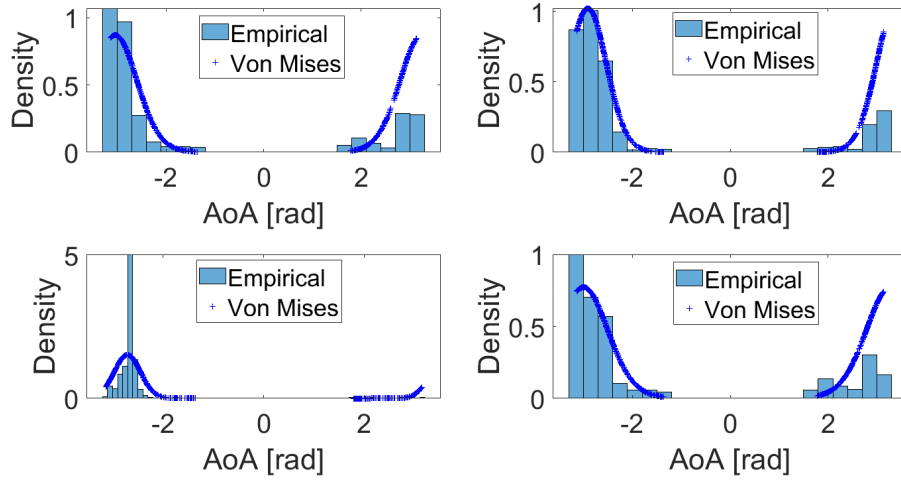


Figure 3.37: Empirical PDF for the AoAs of each propagation region in 2.14 with Maxwell BS. Top left: LoS region. Top right: NLoS region 1. Bottom left: NLoS region 2. Bottom right: tunnel region.

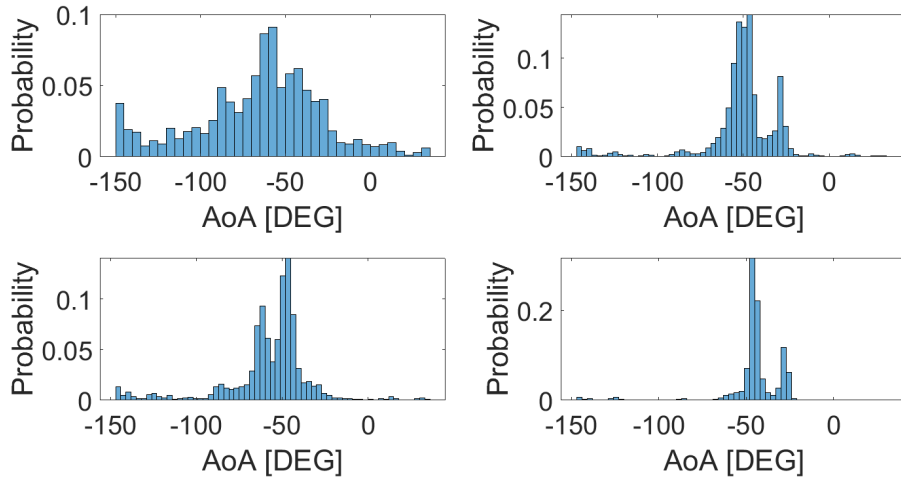


Figure 3.38: Empirical PDF for the AoAs of each propagation region in 2.14 with Mercator BS. Top left: OLoS region. Top right: NLoS region 1. Bottom left: NLoS region 2. Bottom right: NLoS region 3.

case for the tunnel region: it is slightly more probable that radio signals departure from the LoS angle, but this probability is not as high as in the other regions. In fact, the Von Mises κ parameter is the lowest for this region, indicating more spread of the distribution. This means relevant signals departure from many directions at an equal probability, and a close look at the power angular profile (Figure 3.42) shows that they also departure at a very similar

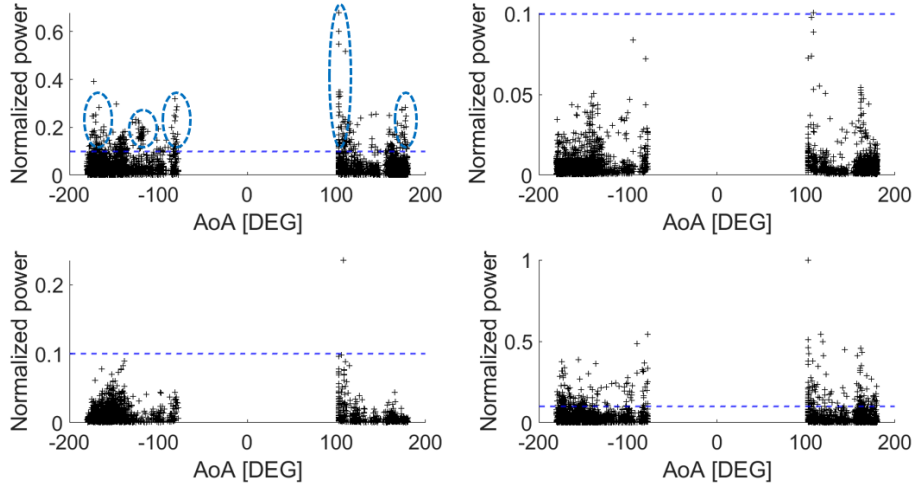


Figure 3.39: Normalized power of propagation paths against the AoAs for the propagation regions in Saint Barbe route with Maxwell BS. Upper left: LoS region. Upper right: NLoS 1 region. Bottom left: NLoS 2 region. Bottom right: tunnel region.

power. If we know that for this region radiated signals from almost all directions have the same relevance (same power and probability of occurrence), how it would be possible for the signals to point, for example in a direction opposite to the BS, to reach the BS? Again, this could be explained partially by the idea of a tunnel effect, since those signals would bounce repeatedly between Saint Barbe and Vinci buildings, before finding their way to Maxwell BS through the rooftop of Vinci building, for example, or through any other way.

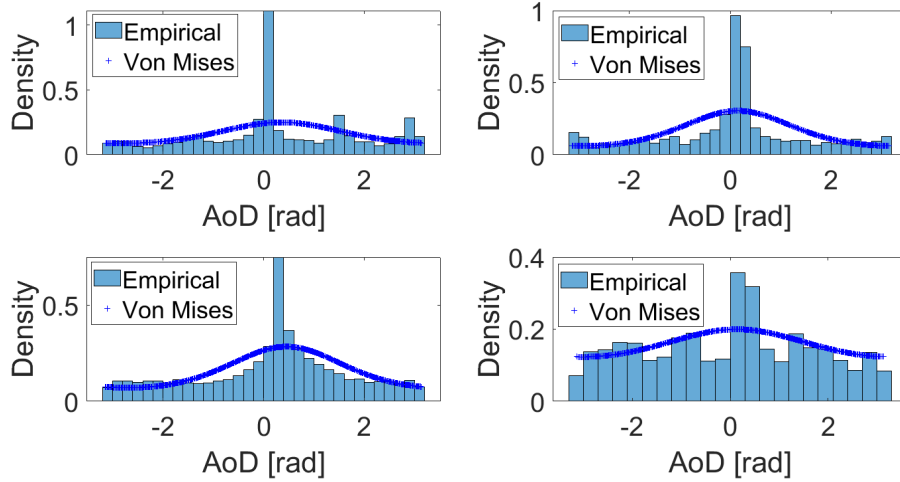
A look at Figure 3.41, shows that what has been said for the tunnel propagation region with Maxwell BS, doesn't hold for the NLoS 3 region with Mercator BS: there is a clear dominant and highly probable AoD. This is confirmed by the κ parameter of the Von Mises fit (Table 3.8): the κ parameter of the Von Mises fit for the NLoS 3 region is the highest, indicating lower spread and higher concentration around μ than for the other regions. Moreover, the mean power of the propagation paths for this region is half the mean power of the propagation paths for the NLoS 2 region (Figure 3.43); also, the mean received power for the NLoS region is the lowest of all the regions (around -70 dBm), 3 dBm less than the next one which is NLoS 2 region with -67 dBm.

3.6 Conclusion

Different propagation mechanisms arise when the link between a transmitter and a receiver is obstructed. On the other side, for regions of the trajectory

Table 3.8: MLE Von Mises parameter estimation and angular statistics for the AoDs of Saint Barbe route per propagation region.

BS	Propagation region	μ	κ	θ_A	θ_S
MXW	LoS	20	0.52	14.3	0.92
MXW	NLoS 1	8.5	0.8	9.8	0.9
MWX	NLoS 2	26	0.7	23	0.88
MWX	Tunnel	7.8	0.24	18	0.95
BS	Propagation region	μ	κ	θ_A	θ_S
MKT	OLoS	87.8	0.4	103.8	0.93
MKT	NLoS 1	144	0.48	136	0.95
MKT	NLoS 2	113	0.39	105	0.92
MKT	NLoS 3	119	0.91	128	0.81

**Figure 3.40: Empirical PDF for the AoDs of each propagation region in 2.14 with Maxwell BS. Top left: LoS region. Top right: NLoS region 1. Bottom left: NLoS region 2. Bottom right: tunnel region.**

of the mobile terminal, where there are LoS conditions, measurements show that the spatial correlation matrix remains highly correlated. It is thus indicative that there are channel characteristics that remain unchanged through the region. As expected, for such LoS regions the power of the dominant path is various times bigger than the average power of the remaining paths, as seen from the power angular profiles presented in the chapter. Besides LoS, tunneling or street canyon effects can make that the power at the receiver coming from such a phenomena increase with respect to the case where there are pure NLoS conditions with no such effect. In this case, rays coming from different directions arrive with significant power at the receiver, contrary to the LoS

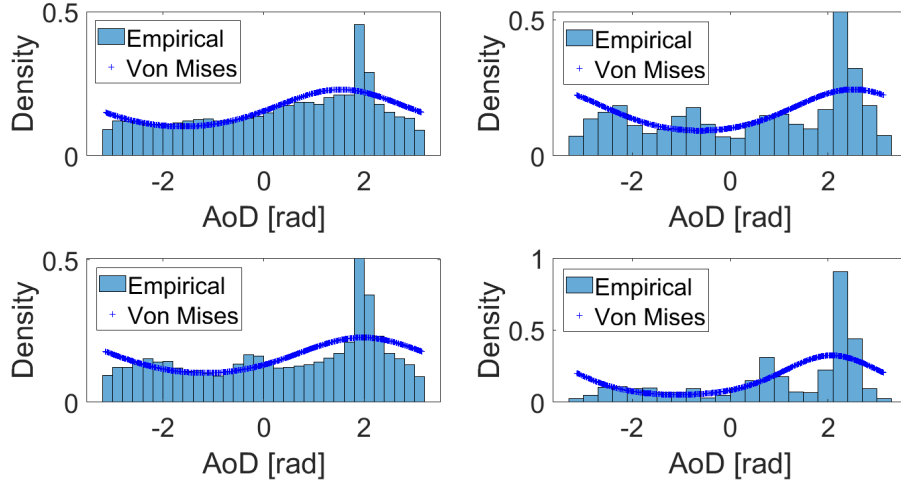


Figure 3.41: Empirical PDF for the AoDs of each propagation region in 2.14 with Mercator BS. Top left: OLoS region. Top right: NLoS region 1. Bottom left: NLoS region 2. Bottom right: NLoS region 3.

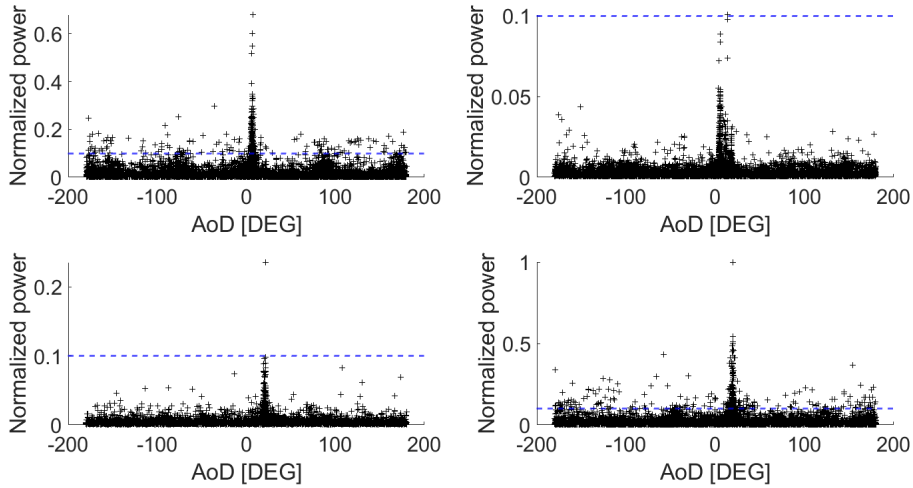


Figure 3.42: Normalized power of propagation paths against the AoDs for the propagation regions in Saint Barbe route with Maxwell BS. Upper left: LoS region. Upper right: NLoS 1 region. Bottom left: NLoS 2 region. Bottom right: tunnel region.

case. It has been seen that angular and power information are strongly related with the placement of obstacles between the transmitter and the receiver, as was intuitively thought. For medium size building scenario, with the static antenna base station above most of the buildings, rooftop propagation is one of the main mechanisms of propagation. Combined mechanisms of propaga-

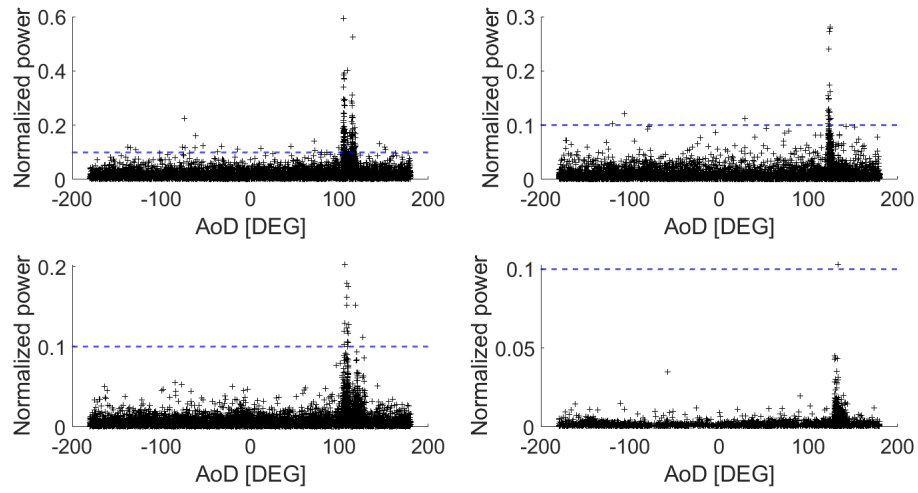


Figure 3.43: Normalized power of propagation paths against the AoDs for the propagation regions in Saint Barbe route with Mercator BS. Upper left: LoS region. Upper right: NLoS 1 region. Bottom left: NLoS 2 region. Bottom right: NLoS region 3.

tion arise in NLoS conditions, where there are no strong coherent reflectors where rays easily reach their way to the base station.

Multi-Path Clustering

4

This chapter presents the *clustering* block of the work-flow introduced in Chapter 1. The clustering task is described formally, as well as different methodologies to tackle it. The inconvenience of traditional clustering methods for channel parameter estimates is explained, along with an alternative to those methods based on mapping channel parameter estimates to the physical scenario before clustering the resultant mapped propagation paths. Although examples of found clusters are provided, since it is not possible to visualize all them, metrics like power per snapshot, cluster lifetime and number of clusters per snapshot, seek to summarize information about all found clusters. Due to the lack of a ground-truth partition for the observations (the propagation paths), this information is descriptive of how well the partition is done for all the clusters that can't be displayed. A brief discussion about the challenges in using indexes that measure how well is a partition based on a given criteria is provided.

4.1 Introduction

Given a channel model, channel parameter estimation finds a set of parameters that minimizes a loss function and best fit the model assumed. Due to the time-variant nature of the channel both sets μ_{SC} and γ_{SC} are dependent of time. Due to the movement of the transceivers and scatterers, radiated energy might propagate through the physical medium in different ways: assuming waves behave as rays, the trajectory followed by the ray could change completely from a time instant to another, giving rise to variations in the channel parameter estimated. A ray is an effective way to look at the continuum of energy received. Waves sharing similar properties will be seen by the receiver system as an unique wave, since any system is bandwidth and aperture limited. Thus, channel parameters, rather than properties of a specular propagation path, are properties of a cluster of specular propagation paths lying within the bandwidth and angular resolution of the system.

In a similar way, channel parameter estimation algorithms as SAGE produce estimates that maximize (possibly locally and not-globally) a log-likelihood function, without any clustering process involved. In other words, estimated parameters might be *similar*, in some sense, such that from a propagation or

physical perspective, they belong to the same specular path. In general, if a full model for the measured channel matrix $\mathbf{H}$ is available (as in RIMAX algorithm), estimation of the parameters of that model might produce *similar* parameters, in some specific sense, that physically correspond to an unique mechanism of propagation. In the case of specular propagation paths, similar channel parameters in the angular domain, for instance, might be indicative of the same Interacting Object (IO), for instance a window, section of a building, which could be identified in the scenario for further analysis.

Multi-path clustering deals with the problem of grouping channel parameter estimates. This clusters of parameters serve as inputs for wireless channel models as the COST2100 model, used to simulate scenarios related to the one for which the clustering was done. Besides so, clustering reduces the number of observations used in any data analysis chain.

In a broader sense, the purpose of clustering information about the wireless propagation channel stays within the purpose of any exploratory data analysis: summarize main characteristics of the dataset under examination. In this sense, it goes in a different direction than Computational Electromagnetics: while for the latter one it is desired to model precisely the interaction of the Electric and Magnetic fields with the environment and objects composing it at any time instant and space location, for the former one, it is desired few information that can, up to some point, encompass many features or characteristics of the wireless channel.

Clustering of wireless channel information is related with *change point* detection theory, as the latter looks for points in the function under consideration for which the statistics of the probability distribution of the function change, thus, segmenting (grouping function values) the function. An example of this is the Saleh-Valenzuela model. The Power Delay Profile (PDP) is assumed to be composed by exponentially-decaying clusters of powers at certain delays. Those delays are the change points for which the function under consideration (the PDP) abruptly changes.

4.2 Clustering formalism

For a given time t , high-resolution parameter estimation provides a set of channel parameters $\mu_{SC,l}$ and $\gamma_{SC,l}$, where l represents the propagation path l , and there are L propagation paths in time snapshot considered. For convenience, from now and on, we omit the subscript SC . Propagation with similar parameters are clustered. The space where each observation, each propagation path parameters, belongs, is usually called the parameter space. The parameter space is composed by dimensions with different units, for instance: angular and time dimensions. A pre-processing stage is required for many clustering algorithms, in which given the mean and standard deviation of the

population, an observation x is standardized, as in 4.1 or 4.2.

$$x \in [0, 1] \leftarrow \frac{x - \min(x)}{\max(x) - \min(x)}, \quad x \in \mathbb{R} \quad (4.1)$$

$$x \leftarrow \frac{x - \mu_x}{\sigma_x} \quad (4.2)$$

with μ_x and σ_x being the sample mean and standard deviation of observations x . Alternatively, it is possible to compensate for this difference in units along dimensions, directly through the distance measure that compares two observations. The clustering procedure can be stated as follows:

Given a set of observations $\mathcal{X} = \{\mu_1, \mu_2, \dots, \mu_L\}$, a k -clusterization or k -partition $\Gamma_k : \mathcal{X} \mapsto \mathcal{M}_k$ is a mapping that assigns to each member of the set of propagation path parameters $\mathcal{X}$ a number in the set $\mathcal{M}_k = \{1, 2, \dots, k\}$.

All propagation path parameters assigned with the same number in $\mathcal{M}_k$ belong to the same *cluster*, denoted here by C . The center-of-gravity or *centroid* of a *cluster* can be defined as the simultaneously closest-point to each and all the observations of the cluster

$$c(C) = \frac{1}{|C|} \sum_{\mu_i \in C} \mu_i \quad (4.3)$$

A weighted-centroid wc weights each member of the cluster by the amount w as in 4.4. This might be emphasizing that some members of the cluster have higher relevance in some sense. For instance, if the power γ of each propagation path is used as a weight, centroids will tend to be closer to propagation paths with higher power. In this way, the direction of propagation is driven by propagation paths that contribute more to the received power.

$$wc(C) = \frac{\sum_{\mu_i \in C} w_i \mu_i}{\sum w} \quad (4.4)$$

The purpose of any *clusterization* is to minimize a distance measure between the members of the same cluster, while maximizing the distance between the centroids of different clusters. This distance measure is responsible for introducing a sense of similarity between observations. In other words, such a distance states in which sense two observations are similar.

For two observations $\mathbf{x}, \mathbf{y}$ of the set $\mathcal{X}$, a distance metric $d : \mathcal{X} \times \mathcal{X} \mapsto [0, \infty)$ is a mapping that obeys the following properties:

1. *Identity of indiscernibles*: given $\mathbf{x} = \mathbf{y}$, then $d(\mathbf{x}, \mathbf{y}) = 0$.
2. *Symmetry*: $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$

Table 4.1: Common used distance measures

Name	Equation
Euclidean	$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y})}$
Cosine	$d(\mathbf{x}, \mathbf{y}) = 1 - \frac{\mathbf{x}\mathbf{y}^T}{\ \mathbf{x}\ _2 \ \mathbf{y}\ _2}$
Correlation	$d(\mathbf{x}, \mathbf{y}) = 1 - \frac{(\mathbf{x} - \mu_{\mathbf{x}})(\mathbf{y} - \mu_{\mathbf{y}})^T}{\ \mathbf{x} - \mu_{\mathbf{x}}\ _2 \ \mathbf{y} - \mu_{\mathbf{y}}\ _2}$

3. *Triangle inequality*: for $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X$, $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$

If there is a function that satisfies the first two properties but not necessarily the triangle inequality we refer to it as a distance measure. Common distance measures used for clustering algorithms are shown in Table 4.1. There $\|\cdot\|_2 = \frac{1}{N} \sum_{i=1}^N [\mathbf{x}]_i^2$ is the euclidean norm, $\mathbf{x} \in \mathbb{R}^N$ and $[\cdot]_i$ represents the i -th position of vector $\mathbf{x}$. These distances assume so far the entries of the vector $\mathbf{x}$ are non-periodic. However, this is not the case for angular entries, which are periodic by nature. In such a case the distance measure $d : \mathbb{S}^2 \times \mathbb{S}^2 \mapsto [0, 2]$ in 4.5 is used

$$d_{\Omega}(\Omega_1, \Omega_2) = \left\| \begin{bmatrix} \cos(\theta_1) \sin(\phi_1) \\ \sin(\theta_1) \sin(\phi_1) \\ \sin(\phi_1) \end{bmatrix} - \begin{bmatrix} \cos(\theta_2) \sin(\phi_2) \\ \sin(\theta_2) \sin(\phi_2) \\ \sin(\phi_2) \end{bmatrix} \right\| \quad (4.5)$$

which is nothing but the conversion from spherical coordinates to Cartesian coordinates, assuming all points lie in the unit sphere $\mathbb{S}^2$, $\Omega = [\theta, \phi]$, and ϕ is the elevation angle measured with respect to the azimuth plane. To compare delays, it is enough to use the 1-dimensional euclidean distance

$$d_{\tau}(\tau_1, \tau_2) = |\tau_1 - \tau_2| \quad (4.6)$$

Furthermore, the purpose of standardizations like 4.1 and 4.2 is to provide an hyper-dimensional space like a sphere where clustering algorithms like k -means perform better. Thus, standardizations for 4.5 and 4.6 are introduced in the well-known Multi-Path Component Distance (MCD) [Ste+02], [Czi+05], [Czi+06a], to make each and all entries of the observation $\boldsymbol{\mu} = [\tau, \Omega_{RX}, \Omega_{TX}]^T$ to lie in the interval $[0, 1]$

$$d_{MCD}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \sqrt{\frac{1}{4}d_{\Omega}^2(\Omega_{TX,1}, \Omega_{TX,2}) + \frac{1}{4}d_{\Omega}^2(\Omega_{RX,1}, \Omega_{RX,2}) + \left(\frac{1}{\delta\tau} \frac{\sigma_{\tau}}{\delta\tau} d_{\tau}(\tau_1, \tau_2)\right)^2} \quad (4.7)$$

with $\delta\tau = \max_{i,j} |\tau_i - \tau_j|$ and σ_τ the sample standard deviation of all time-delays considered. The *clusterization* Γ_k requires the number of clusters k to be known beforehand. This is the case for algorithms like k -means. However, when the number of clusters is not known in advance, multiple *clusterizations* $\Gamma_2, \Gamma_3, \dots, \Gamma_M$ must be computed and compared under a given criterion to decide which is the best one under such criterion. In general, any criterion would try to relate an inter-cluster and intra-cluster distance measures. If the inter-cluster distance measure is denoted by d^α and the intra-cluster distance measure by d^β , with $\cdot$ indicating that the distance measure could be any of the Table 4.1 or the d_{MCD} in 4.7, then, any criterion looks for a function $f(d^\alpha, d^\beta)$ such that:

1. A higher value $f(d^\alpha, d^\beta)$ represents a better clusterization (it is equivalent to make the function have a higher value when the clusterization is worse).
2. If the amount by which d^α increases is bigger than the amount by which d^β increases, then $f(d^\alpha, d^\beta)$ increases.
3. If the amount by which d^α decreases is bigger than the amount by which d^β decreases, then $f(d^\alpha, d^\beta)$ decreases.
4. $f(d^\alpha, d^\beta)$ increases when d^α increases **and** d^β decreases.
5. $f(d^\alpha, d^\beta)$ remains the same if the amount by which d^α decreases is the same by which d^α decreases, or, if the amount by which d^α increases is the same by which d^α increases.

It is not trivial to choose such a function or criterion. Well-known criteria like Davies-Bouldin Index [DB79] (4.8) or Dunn Index [Dun74] (4.9) have been available since the 70's, and there is no index which out-performs all the other for any dataset [KH10]. These indexes are considered sometimes internal measures, because they take into account properties of the cluster themselves. However, there can be defined external measures or indexes, to define which clusterization is better, as Information-theoretic measures: Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i,j, i \neq j} \left(\frac{d^\beta(C_i) + d^\beta(C_j)}{d^\alpha(C_i, C_j)} \right) \quad (4.8)$$

$$DU = \frac{\min_{i,j} d^\beta(C_i, C_j)}{\max_i d^\beta(C_i)} \quad (4.9)$$

Table 4.2: Examples of inter and intra-cluster measures for a cluster or pair of clusters

Specific d^β	Specific d^α
$\frac{1}{ C } \sum_{a \in C} d.(a, c(C)^a)$	$d.(c(C_1), c(C_1))$
$\max_{a, b \in C} d.(a, b)$	$\max_{a \in C_1, b \in C_2} d.(a, b)$
$\min_{a, b \in C} d.(a, b)$	$\min_{a \in C_1, b \in C_2} d.(a, b)$
$\max_{a \in C} d.(a, c(C))$	$\max_{a \in C_1} d.(a, c(C_2))$
$\min_{a \in C} d.(a, c(C))$	$\min_{a \in C_1} d.(a, c(C_2))$

^a $c(C)$: Centroid of cluster C as in 4.3.

where $d^\beta(C)$ is the intra-cluster distance measure for cluster C and $d^\alpha(C_i, C_j)$ is the inter-cluster distance measure for clusters C_i and C_j . Both distance measures, the inter-cluster and intra-cluster can be defined in various ways. Table 4.2 shows some of the common used.

Moreover, distance measures can be defined in terms of how well they measured a specific property of the clusters under analysis. For example, they can be used for measuring distance from centroids to centroids, or they can be used to measure distance from centroids to element of centroids.

4.2.1 K-means

Unsupervised learning algorithms are used to cluster datasets without providing labels for it. Within this algorithms, k -means [Mac+67] has gained attention due to its simplicity and its relative good performance. Brute force approaches are not feasible when the size of the dataset increases. Given a dataset, the objective function of k -means in 4.10 is to find a partition of k clusters such that sum of the distances between the centroids of each cluster and all the members of the cluster is minimized. Let us call $\mathbf{C} = \{C_1, C_2, \dots, C_k\}$ the set of all the clusters. Each cluster C is composed itself by a set of observations $\mu_{i_1}, \mu_{i_2}, \dots, \mu_{i_{|C|}}$, where $\mathcal{I}(C) = \{i_1, i_2, \dots, i_{|C|}\}$ is the index set, $i_j \in \mathcal{M}_k$ and $\mathcal{M}_k = \cup_{i=1}^k \mathcal{I}(C_i)$.

$$\arg \min_{\mathbf{C}} \sum_{i=1}^k \sum_{\mu \in C_i} \|\mu - c(C_i)\|_2 \quad (4.10)$$

Algorithm 6 shows the 3 steps involved in k -means. Different ways of initializing the algorithm have been proposed and it is known that the performance of the algorithm is dependent on this procedure [CKV13]. To decrease the impact of the initialization step, multiple runs of the algorithm are used for the same number of clusters. Depending on the size of the dataset and the

Algorithm 6 K-means algorithm.**Input:** $\{\mu_1, \mu_2, \dots, \mu_L\}, k$ **Output:** $C, c(C)$

- 1: Initialization: from the initial set of observations, select randomly k centroids.
- 2: Assignment: assign each observation to the closest centroid based on a given distance measure (i.e. euclidean).
- 3: Update: for a given cluster, update the centroid by using 4.3.
- 4: **if** stop condition **then**
- 5: Return the centroids and the clusters.
- 6: **else**
- 7: Go to step 2.

time-computing restrictions this might be possible up to some degree. Multiple stopcondition can be used to detect the end of the algorithm: when the new centroids have moved less than a given tolerance respect to the old ones; the change in the objective function is below a given tolerance.

Other versions of k -means have been proposed. For instance, within the context of wireless propagation channel analysis, the KPowerMeans algorithm was proposed in [Czi+06b] to include the powers associated to each propagation path (γ) such that the centroids are weighted as in 4.4, and the partitioning is driven by bigger power centroids. The structure of Algorithm 6 remains the same, but now, during the update step, the centroids are computed according to 4.4.

On the other hand, k -medoids [KR90], [PJ09] algorithm substitutes the notion of the centroid, by replacing the mean involved in its computation by the concept of the *medoid*. While in k -means the centroid doesn't belong to the input matrix of observations (X), in k -medoids it does. This is not only suitable for categorical datasets, but also to datasets the centroid is meaningless. Originally called Partitioning Around Medoids (PAM) the algorithm steps slightly differ from the ones in k -means. During the update step, the *medoid* is found as the cluster member that minimizes the total distance between itself and each cluster member.

Finally, the Fuzzy C-Means [Dun73] provides *soft*-partitioning, meaning each observation belongs to each and all the clusters up to some degree. In this case, an output matrix P_{FZZY} of size $L \times k$ is produced (4.11), where an entry (i, j) represents the how much observation i belongs to cluster j . This is particularly suitable for the case when observations can belong to multiple clusters due to use of a specific distance measure, and human visual inspection needs to be used to decide the clusters. Figure 4.1 shows a comparison between a partitions made by KPowerMeans, Fuzzy C-means and k -means.

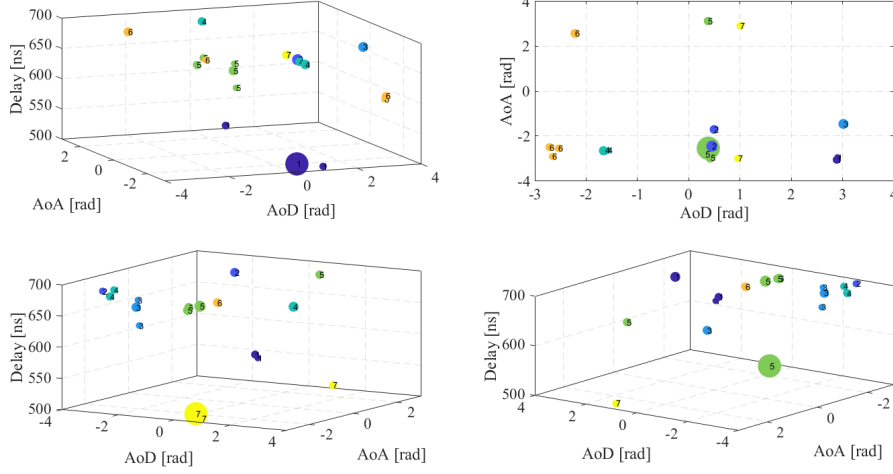


Figure 4.1: Partitioning of the first time snapshot of the Saint Barbe route with Maxwell BS. Left-upper panel: KPowerMeans. Right-upper panel: k -means without power weighting. Left-bottom panel: Fuzzy C-Means, where the maximum value of each column of matrix P_{FZZY} is chosen. Right-bottom panel: Fuzzy C-Means, where the second maximum value of each column of matrix P_{FZZY} is chosen. All panels show the same observations from different perspectives in order to have a better insight into their locations in the parameter space. Each cluster is identified redundantly by a color and a number. The same number of clusters (7) is used for all the panels. The size of the observation is associated with its power.

In Fuzzy C-means clusters an observation might have similar probabilities of belonging to different clusters. For instance, the observation with the highest power (biggest in the picture) belongs to different clusters in the bottom panels. In the left-bottom panel the observation has a probability of 0.374 of belonging to the 4th cluster; while on the right-bottom panel, the same observation has a probability of 0.246 of belonging to the 1st cluster.

$$\mathbf{P}_{FZZY} = \begin{bmatrix} p(\mu_1 \in C_1) & p(\mu_2 \in C_1) & \dots & p(\mu_L \in C_1) \\ p(\mu_1 \in C_2) & p(\mu_2 \in C_2) & \dots & p(\mu_L \in C_2) \\ \vdots & \vdots & \ddots & \vdots \\ p(\mu_1 \in C_k) & p(\mu_2 \in C_k) & \dots & p(\mu_L \in C_k) \end{bmatrix} \quad (4.11)$$

where $p(\mu_i \in C_j)$ represents the probability that observation i belongs to cluster j , and $\sum_{j=1}^k p(\mu_i \in C_j) = 1, \forall i \in 1, \dots, L$.

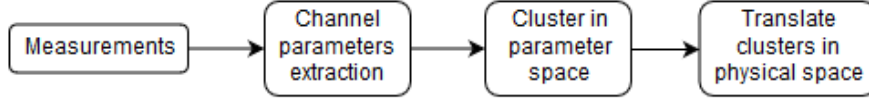


Figure 4.2: Common approach to cluster channel parameters in literature

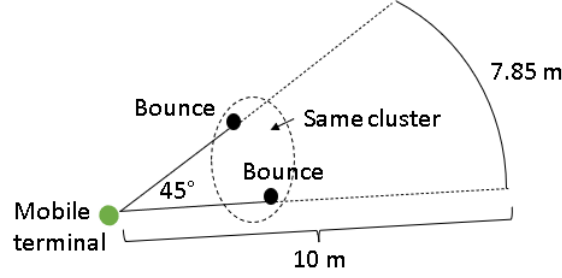


Figure 4.3: Physical mapping of a cluster of azimuths of departure (bounces). When the cluster is badly done, the physical mapping of the channel parameter (in this case, the azimuth of departure) led to big distance between mapped parameter in the real scenario.

4.2.2 Channel parameters physical mapping

Channel models as COST2100 rely on grouping sets of Multipath Components (MPCs) with similar properties. It has been shown experimentally [Liu+12], that MPCs tend to appear in clusters.

In order to identify clusters physically, a common approach is to follow the procedure given by Figure 4.2.

In this case, there is a need to translate clusters in the parameter space into clusters in the physical space, in order to find out the real-world scatterers. Such a translation is performed in [Pou+11] and [Zhu14], through the use of a measurement-based ray tracer. However, this translation is not trivial and it can lead to a great mismatch between parameter-based clusters and their physical counterpart, as the angular spreads of parameter-based clusters can allow to have different members of the same cluster, located in different regions in space. This is shown in Figure 4.3. The observations belong to the same cluster, despite the fact that the angular difference between them is around 45° . This means that after 10 m, both observations are separated by a distance of 7.85 m. As a result of this propagation paths that are said to belong to the same cluster could touch be interacting with objects that are enough-separated in the physical scenario.

Besides this, MCD measure (4.7) does not give a direct relationship with the real distance between scatterers in the environment. Therefore, under the situation in which the KPowerMeans algorithm does not perform well,

it is not evident how to interpret the closeness given by the MCD between datapoints in the parameter space. To overcome this situation, the clustering algorithm is only performed after channel parameters are related with the physical scenario as in [Han+17]. The consequence of this is that observations are no longer clustered in a parameter space, but instead in the physical space with coordinates in meters.

A given observation μ represents a propagation path. The propagation path starts its trajectory at the transmitter location and moves in a direction given by Ω_{TX} ; it will interact with the scenario in an unknown way, until it reaches the receiver from direction Ω_{RX} after a time τ . Power of the path does not intervene in the physical mapping. In the parameter-clustering approach, the power of the path was used to drive centroids close to bigger-power observations. In this approach, the clustering is solely based on the physical proximity of the propagation paths, thus, if two paths with high power are close-enough they will belong to the same cluster.

In what follows, we consider only 2D propagation (elevation angles are not taken into account). For a given snapshot t , each propagation path l is mapped into the scenario by means of the following considerations:

- The reference (0 rad) for angles $\theta_{RX,l}$ and $\theta_{TX,l}$ is set to the x-axis (longitude coordinate). For doing so, it is assumed that the maximum power is coming from the virtual line connecting the RX with the TX; putted in other words, the LoS without the presence of the buildings.
- Only the buildings may cause reflection. As shown in Figure 4.4, the first bounce is found by extending a line with angle $\theta_{RX,l}$ between the TX and the contours of the buildings. The first building the line hits is the first bounce of the path.
- The last bounce is found in a similar way. Extending a line with angle $\theta_{TX,l}$ between the RX and the contours of the buildings. The first building the line hits is the last bounce of the path.

From those considerations, the path l is composed by the three line segments shown in Figure 4.4:

$$\begin{aligned}\mathcal{L}_{l,t}^1 &= \{\mathbf{x}_t + p(\mathbf{p}_{l,t} - \mathbf{x}_t) \mid p \in [0, 1]\} \\ \mathcal{L}_{l,t}^2 &= \{\mathbf{p}_{l,t} + p(\mathbf{q}_{l,t} - \mathbf{p}_{l,t}) \mid p \in [0, 1]\} \\ \mathcal{L}_{l,t}^3 &= \{\mathbf{q}_{l,t} + p(\mathbf{r} - \mathbf{q}_{l,t}) \mid p \in [0, 1]\}\end{aligned}\tag{4.12}$$

We account for the delay in the cost function of the minimization problem.

From the previous considerations, we have a basic idea about the detailed propagation paths: there is no knowledge about the real number of bounces

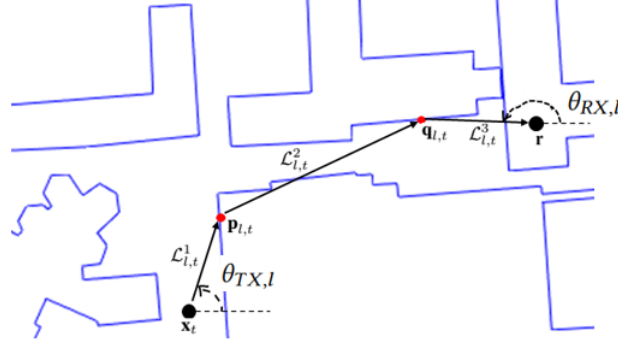


Figure 4.4: Physical model used. One propagation path l of time snapshot t depicted.

and the exact angles at which the waves reflect, transmit or diffract once arriving at the buildings. However, up to this point we are not interested in a more detailed knowledge about the propagation paths. We know that regarding the channel parameters mapping, the only thing that can add complexity to our model is the delay τ_l . We consider that what happens in between the first bounce $\mathbf{p}_l$ and the last one $\mathbf{q}_l$ can be characterized only by an effective path length η_l :

$$\eta_l = c\tau_l - |\mathcal{L}_l^1| - |\mathcal{L}_l^3|, \quad (4.13)$$

where for path l , $|\mathcal{L}_l^1|$ and $|\mathcal{L}_l^3|$ represent the distance between transmitter and first bounce, and the distance between last bounce and receiver, respectively. The line segment $\mathcal{L}_{l,t}^2$ in Figure (4.4) is a representation of the "effective" path with propagation of length η_l .

Clustering of first and last intercept observations, require some changes to the clustering introduced in section 4.2. From the channel parameters mapping, we end up with bounces $\mathbf{p}_{l,t}$ and $\mathbf{q}_{l,t}$ where for each time-snapshot t the number of paths $l = 1, \dots, L_t$ can vary. Each bounce is representing a 2D coordinate (latitude and longitude). As for each $\mathbf{p}_{l,t}$ there is a corresponding $\mathbf{q}_{l,t}$ in the same path, we consider them as an unique coordinate $\mathbf{z}_{l,t} \in \mathbb{R}^4 = [\mathbf{p}_{l,t}, \mathbf{q}_{l,t}]$. As shown in Figure 4.4, the path l starts at location $\mathbf{x}_t$ of the transmitter (the location of the transmitter at time-snapshot t), then hits $\mathbf{p}_{l,t}$, follows an unknown trajectory whose effective distance is η_l , then hits $\mathbf{q}_{l,t}$, and finally reaches receiver $\mathbf{r}$.

Let $\mathcal{Q}$ be a set comprising all $\mathbf{z}$ to be clustered. The set $\mathcal{Q}$ could be formed by all the $\{\mathbf{z}_{l,t}\}_{l=1}^L$ of a time-snapshot t . It could be formed by the $\{\mathbf{z}_{l,t}\}_{l=1,t \in \mathcal{T}}^L$ of the time-snapshots in the set $\mathcal{T}$. For example, the set $\mathcal{T}$ could be formed by consecutive time-snapshots belonging to a stationary region, or non-consecutive

Table 4.3: Special snapshot sets $\mathcal{Q}$

Notation	Description
$\mathcal{Q}_{LoS}$	Composed by all the propagation paths $\mathbf{z}_{l,t}$ of the snapshots for which there are LoS conditions.
$\mathcal{Q}_1$	Composed by all the propagation paths $\mathbf{z}_{l,t}$ of <i>one</i> time-snapshot.
$\mathcal{Q}_{SC-ALL}$	Composed by all the propagation paths $\mathbf{z}_{l,t}$ of all the time-snapshot for the route under consideration and a given BS.

snapshots that have similar propagation paths, among others. Table 4.3 show special sets $\mathcal{Q}$ that are of interest for posterior analyses.

A *clusterization* Γ^k is a mapping $\Gamma^k : \mathcal{Q} \rightarrow \mathcal{M}$ that assigns to each $\mathbf{z} \in \mathcal{Q}$ a number in the set $\mathcal{M} = \{1, 2, \dots, k\}$, with k unknown beforehand. Let us call *cluster groups* the sets composed by $\mathbf{z}$ with the same assignment $i \in \mathcal{M}$. We denote them by C_i , and it is evident that:

$$\begin{aligned} \forall i \quad C_i &\subset \mathcal{Q}, \quad i = 1, \dots, k \\ \cup_i C_i &= \mathcal{Q}, \\ \cap_i C_i &= \emptyset. \end{aligned} \tag{4.14}$$

Recall that for different clusterizations we end up with different groups C_i . Therefore, it might be more accurate to denote the cluster groups by C_i^k ; however, we drop the k index for convenience.

Over each $\mathbf{z}$ a "disk" of radii ϵ can be defined as $Dz(\epsilon) = \{\mathbf{x} \in \mathcal{R}^4 \mid d(\mathbf{z}, \mathbf{x}) \leq \epsilon\}$, where the function $d(\cdot, \cdot)$ is a distance measure defined in (4.19). We define as well, the intersection set $\mathcal{I}_i(\epsilon)$ of all the disks $Dz(\epsilon)$ of all the $\mathbf{z}$ in a *cluster group* i : $\mathcal{I}_i(\epsilon) = \cap_{j=1}^{|C_i|} Dz_j(\epsilon)$, $\forall \mathbf{z}_j \in C_i$.

Inter-cluster ξ and intra-cluster β measures are introduced in equations (4.15) and (4.16), in order to quantify the goodness of the mapping Γ .

$$\beta = \max_{C_i} \max_{\mathbf{u}, \mathbf{v} \in C_i} d(\mathbf{u}, \mathbf{v}) = \max_{C_i} \beta_i \tag{4.15}$$

$$\xi = \min_{\mathbf{u} \in \mathcal{I}_i(\beta_i), \mathbf{v} \in \mathcal{I}_j(\beta_j), i \neq j} d(\mathbf{u}, \mathbf{v}) \tag{4.16}$$

The definition of disks $Dz(\epsilon)$ must be made beforehand. The intersection sets $\mathcal{I}_i(\epsilon)$ might be complicated to compute, or even to define analytically. Therefore, an approximation for (4.16) is done by considering $\mathcal{I}_i(\beta_i)$ as the centroid $\mathbf{y}_i$ of the cluster C_i : $\mathcal{I}_i(\beta_i) \simeq \frac{1}{|C_i|} \sum_{\mathbf{z} \in C_i} \mathbf{z} \triangleq \mathbf{y}_i$. Equation (4.16) is replaced by equation (4.17).

$$\xi \simeq \min_{i \neq j} d(\mathbf{y}_i, \mathbf{y}_j) \quad (4.17)$$

Any Γ^k must try to maximize ξ and minimize β simultaneously. It is therefore a multi-objective optimization problem. We convert it to a single-objective problem, by linearly combining both requirements in (4.18):

$$\min_k \quad \beta^k - \xi^k + \lambda_\tau c \sum_i \sum_{\tau_1, \tau_2 \in C_i} |\tau_1 - \tau_2| \quad (4.18)$$

where c is the speed of light and β^k, ξ^k account for different clusterizations. The third component in 4.18 are the delays τ . The controllable setting λ is used to weight how much the delays matter into the clusterization. For more on this refer to [GVO19].

The distance measure used in equations (4.16) and (4.15) is defined in (4.19). Given two points $\mathbf{u}, \mathbf{v} \in \mathcal{Q}$:

$$d_{IO}(\mathbf{u}, \mathbf{v}) = \frac{1}{N} \sum_{i=0}^N \left\| [\mathbf{u}]_{1:2} + \frac{i}{N} ([\mathbf{u}]_{3:4} - [\mathbf{u}]_{1:2}) - \left([\mathbf{v}]_{1:2} + \frac{i}{N} ([\mathbf{v}]_{3:4} - [\mathbf{v}]_{1:2}) \right) \right\|_2 \quad (4.19)$$

where $\|\cdot\|_2$ is the euclidean norm in $\mathbb{R}^2$, the operator $[\mathbf{x}]_{i:j}$ represents the coordinates i to j of $\mathbf{x}$, and N is the number of divisions in the line segments associated with $\mathbf{u}$ and $\mathbf{v}$ ($\mathcal{L}_{l,t}^2$ in Figure 4.4). Those line segments are built, by just extending a line from $\mathbf{p}_{l,t}$ to $\mathbf{q}_{l,t}$ for each $\mathbf{z}$. Intuitively, what the distance measure in (4.19) is doing is to compute the degree of "match" between the line segments associated with $\mathbf{u}$ and $\mathbf{v}$. The defined measure still can be used if instead of using line segments we use piecewise line segments. That would be the case if we have information about the intermediate bounces. It is clear that as $N \rightarrow \infty$ the distance d_{IO} computes the area of the polygon formed by the vertices of $\mathbf{u}$ and $\mathbf{v}$, if there is no crossing between the line segments.

However, the distance defined in the way in (4.19) fails to detect observations $\mathbf{z}_1$ and $\mathbf{z}_2$ that might be clustered in the same group. It happens because the distance is not taking into account that the line segments in comparison can cross. For example, the left and middle graphs in Figure 4.5 are two pairs of different observations that must give the same distance value (they correspond to equivalent situations according to our criteria for defining a cluster). However, the initial distance h in the left graph will decrease - due to the crossing between line segments -, whereas it will not in the middle

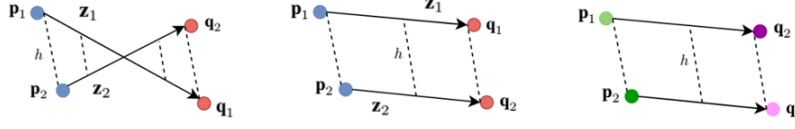


Figure 4.5: Left and middle figures represent two situations for which (4.19) must be equal. The figure in the right represents the corrected version of the figure in the left. Arrows represent the direction followed by the path. z_1, z_2 represent different paths.

graph, giving rise to a different distance computation for each case. To overcome this, we introduce a correction represented in the graph at the right of Figure 4.5. There, the new equivalent line segments are composed by the $q_{l,t}$ and $p_{l,t}$ of both initial line segments.

4.2.3 K-Medoids

Once the channel parameters are mapped to the scenario, clustering procedure starts. The clustering algorithm discussed in Section 4.2.1 performs grouping in the channel parameter space. Through the use of distance measure (4.19) and the observations $\mathbf{z}$ as defined in Section 4.2.2, k -means can be applied to the set of observations. However, some challenges arise, due to the nature of the observations being clustered. Figure 4.6 shows a problem faced when using k -means. Two propagation paths that depart from the same transmitter location and hit different buildings, might be grouped in the same cluster due to their proximity to one centroid at a given iteration of k -means. If the purpose of the grouping in the physical space is to identify IOs, the propagation paths in the figure refer to different IOs, since the region composed by the two buildings and the air, is non-homogeneous: its permittivity, conductivity and permeability change.

K -medoids algorithm avoids this situation by forcing the *medoid* (counterpart of the centroid) to be one of the observations. However, k -medoids operates in the same manner as k -means: it is a distance-based algorithm and doesn't differentiate between different surfaces, buildings or materials. It means, that for propagation paths z_1, z_2 bouncing in different buildings they might be clustered together, if the distance between them is comparable with the intra-cluster distances of other clusters.

If Dunn's Index (4.9) is used, Figures 4.7 and 4.8 show examples of the trade-off between using either k -means or k -medoids algorithms. On one side, k -means tends more than k -medoids to group propagation paths that are dissimilar, or separated belonging to different buildings. On the other side, k -medoids tends to produce more singleton clusters than k -means (Figure 4.8).

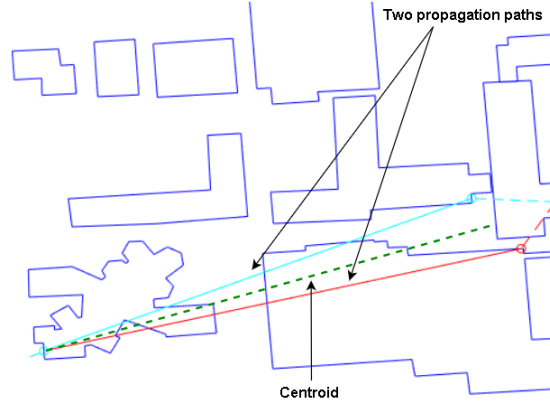


Figure 4.6: k -means centroid problem for the propagation paths mapped. The red and blue trajectories are two different propagation paths. For a given threshold distance, this two propagation paths might be grouped in the same cluster, even though they hit different buildings.

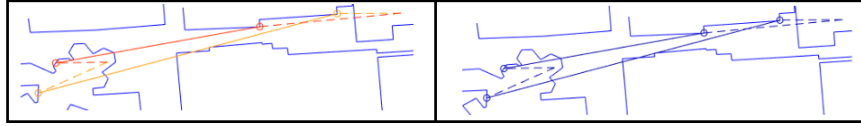


Figure 4.7: Accuracy example for k -medoids and k -means. Left panel: partition made by k -medoids. Right panel: partition made by k -means.

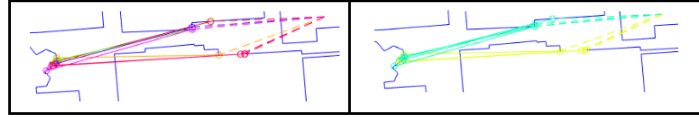


Figure 4.8: Cluster detection example for k -medoids and k -means. Left panel: partition made by k -medoids. Right panel: partition made by k -means.

This is also a consequence of the index used: for different clusterizations, Dunn's Index tend to be better as the number of clusters increase [VGVO19]. This is also a consequence of the input dataset to the clustering algorithms, since the mapping of the channel parameters produce propagation paths that are distant from each other: in other words, there are multiple clusters with only one propagation path.

4.3 Hierarchical clustering

When the set to cluster is Q_{SC-ALL} all the time snapshots of a given BS are used. This means, in principle, that cluster members can be from non-consecutive

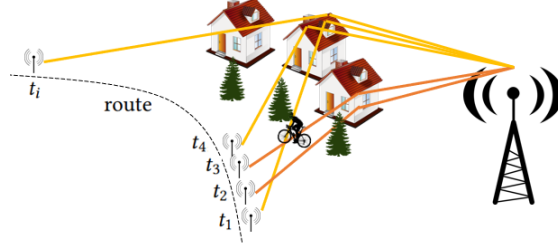


Figure 4.9: Example of clusters of channel parameters from non-necessarily-consecutive time-snapshots. During snapshot t_1 there is strong rooftop propagation. Due to the cyclist and trees, in snapshots t_2 and t_3 strong propagation primarily comes from a side of the house. At t_4 rooftop return to be the main source of propagation.

time-snapshots. This is an advantage in order to find sections of the route where mechanisms of propagation remain unchanged, since channel conditions can change from snapshot to snapshot. For the period t_{IRF} during which a CIR is produced, the channel is assumed to be invariant. Channel parameters as AoA, AoD, EoA and EoD can change quickly, from snapshot to snapshot. This means that by only clustering time-consecutive channel parameters we might lose the picture of how the CIR is behaving similar in average over regions of snapshots. This is another way to think about the stationary regions found in Chapter 2: although channel parameters don't remain the same necessarily between consecutive snapshots, they behave similarly over non-consecutive set of time-snapshots. In other terms, for instance in Figure 4.9, the AoA and AoD might look different in t_1 than in t_2 , but might become very similar again in t_4 . If we found similarities in this channel parameters for two time snapshots separated by a time less than the coherence time of the channel or the local region of stationarity, this reinforces the idea that some conservation of propagation mechanisms is occurring. On the other hand, channel parameter similarities for time snapshots separated by a time longer than the local stationary region, are showing representative IO of the whole measurement (as during t_i in Figure 4.9).

For measurements with around 5000 snapshots and a maximum of 20 paths per snapshot, the number of observations is in the order of 100000. For this dataset, without any knowledge of the number of clusters, k -means is not a feasible option, since thousands of clusterizations would be required to be compared. Moreover, memory becomes a constraint, since a distance matrix of 100000-by-100000 is also required to be computed. Thus, alternatives to this algorithms must be considered. There is no *a priori* knowledge regarding the number of clusters. Algorithms that do not require this as an input parameter are, therefore, suitable. Furthermore, in Section 4.2.3 the choice of

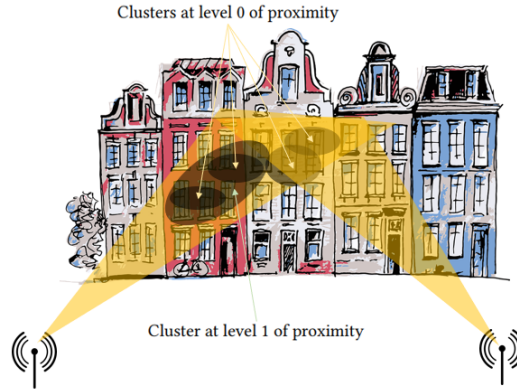


Figure 4.10: Illustration of the levels of proximity for clusters in AHC. The black regions represent clusters of IOs. The levels of proximity refer to the encounters in the hierarchical tree in the AHC algorithm.

Dunn's Index show some problems with clustering algorithms like k -means and k -medoids. Instead of looking for a better index, within the universe of indexes available, an approach that don't rely on a *known* number of clusters is followed.

For small and medium data sets, Agglomerative Hierarchical Clustering (AHC) provides an easy-to-interpret way to partition data. Similarity between each and all pair of observations is required as a first step. After doing so, close observations are linked or grouped iteratively until there is only one group comprising all the observations. At each level or iteration, the degree of similarity between observations decreases. This produces a tree structure whose bottom branches are most similar, and whose upper branches less.

After mapping the channel parameters to the physical environment, the clusters obtained (called here IO), behave under AHC as in Figure 4.10. There will be IO well separated by the geometry of the scenario: for instance, groups of IOs in different buildings, streets, and so on. However, there will be also the case, that big objects as buildings, contain multiple IOs: the radiated energy illuminates a whole region of space (in the case of Figure 4.10, the façade of the residential houses) and the size of the IO depends on the threshold imposed to the AHC algorithm. This approach is valid until the size of the clusters is comparable to the wavelength, since at such moment, ray tracing approximations are no longer valid. For 3.7 GHz the wavelength is approximately 8 cm, and so, IOs should be bigger than around a meter.

An example of the well known dendrogram is shown in Figure 4.11. Each and all propagation paths of the 2 snapshots are shown in the y-axis labels. All the the propagation paths are grouped according to their proximity, by using the distance measure d_{IO} in (4.19). The level at which the clusters are de-

cided is called a cut in the dendrogram and represents the maximum distance allowed between cluster members. Each color represent a different cluster, except for black branches in the graph that represent clusters that might appear after the cut. Moving to the left, the clusters are bigger in size, meaning more distant observations are grouped together. At the level of most closeness groups of propagation paths tend to be from the same snapshot. Propagation paths from different snapshots tend to be grouped as the cutting distance is higher. However, for those propagation paths of different snapshots being clustered at the most close level, it means the IO remains as a strong source of reflection. If propagation paths from consecutive snapshots belong to the same cluster C , the bounces $\mathbf{p}, \mathbf{q}$ are strong reflection points that last for that number of consecutive snapshots. This number is called the *lifetime* of the cluster. It gives an idea of how fast or slow the scenario is changing. In a controlled experiment with no scatterer, neither transceiver movement, and no changes in the propagation environment, the bounces will remain the same endlessly, and so, the lifetime would be infinity.

For the Saint Barbe route there are 4 sections of interest shown in Figure 4.12: a LoS region, a region for which a tunnelling effect is occurring (see Chapter 2), and two regions of NLoS conditions with highly probable rooftop propagation. For them the empirical distribution of clusters lifetime is shown in Figure 4.13. An exponential distribution as in (4.20) is fitted through Maximum Likelihood Estimation (MLE). It is expected that the probability behaves as a monotonically decreasing function, since channel characteristics are changing. The exponential behaviour of the empirical distributions shows channel propagation paths are changing fast. Furthermore, the *rate parameter* λ of the exponential distribution behaves in an expected manner: it has a low value for the LoS and tunnel region, meaning the clusters lifetime lasts more, since there is a strong coherence in the channel and power is arriving from clear propagation mechanisms; it has higher value for both NLoS regions, for which bounces are changing faster. Even if rooftop propagation is the dominant mechanism of propagation, for NLoS regions the IO locations vary quicker, as the received power might be a result of multiple propagation paths with low power, rather than power-strong and dominant paths as in LoS.

$$p_{LT}(x; \lambda) = \lambda \exp(-\lambda x) \quad (4.20)$$

Rate parameter is higher for the NLoS 1 than for the NLoS 2, due to Ramps 1 and 2, since rooftop propagation might occur for both regions. In the corner of region NLoS 1 there is no LoS, but there are buildings at both sides of Rue Archimède, thus creating a tunnelling mechanism similar to the one in the

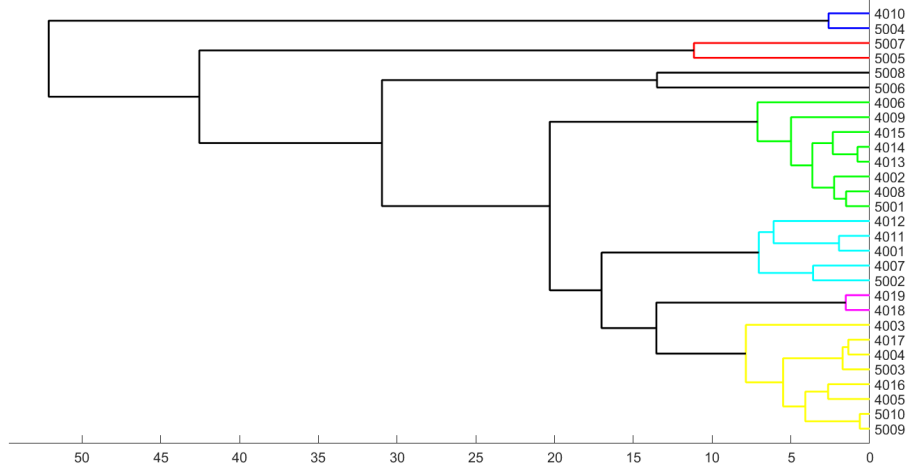


Figure 4.11: Dendrogram for the propagation paths in snapshots 4 and 5 of the second measurement of Saint Barbe route with Maxwell BS. The x-axis represent the distance d_{IO} value. The y-axis are the labels for the propagation paths, codified as follows: thousand times the snapshot number plus the number of the propagation path for that snapshot.

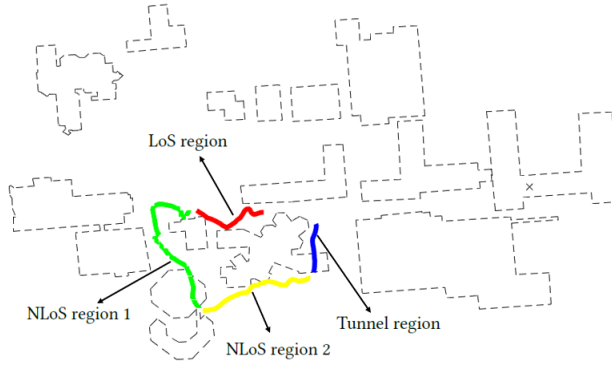


Figure 4.12: Four regions of interest in measurement 2 of Saint Barbe route with Maxwell BS.

Tunnel region. It isn't as strong as the one in the Tunnel region, since the corner of region NLoS 1 is farther to the receiver. It is also expected that the LoS region has the lowest λ and that the clusters last for more time.

By using the dendrogram, clusterizations are produced due to the selection of the distance cut or distance threshold, instead of the number of clusters. Since a distance threshold provides information closely related to the physical location of the IOs, multiple clusterizations obtained when the distance threshold is increased are easily explained. Figures 4.14, 4.15, 4.16 and 4.17 show 4 consecutive clusters for the LoS, NLoS 1, NLoS 2 and Tunnel

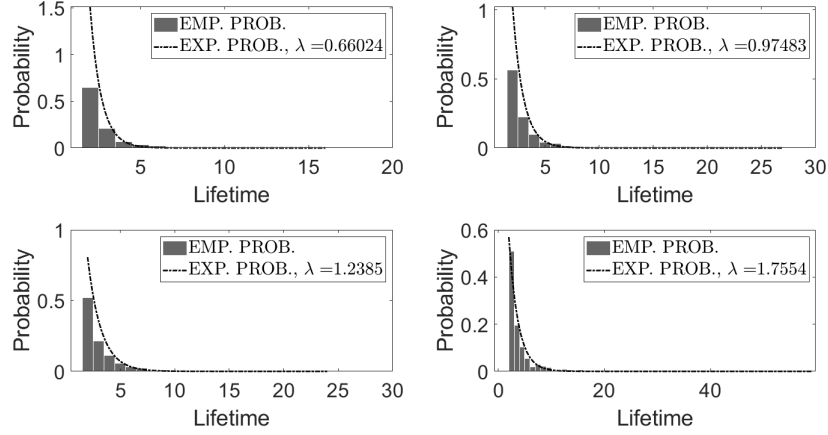


Figure 4.13: Cluster lifetime probability for the second measurement in Saint Barbe route with Maxwell BS. LoS region is the upper left graph. Tunnel region corresponds to the upper right graph. NLoS 1 region is at the bottom left panel. NLoS 2 region is at the bottom right panel.

regions respectively. The number of clusters is reduced since there are less ways of propagating to the Maxwell BS. The average number of clusters for the LoS region is around 5, while for the NLoS region 1 is 3, and for the NLoS region 2 is 4, and for the Tunnel region is 6. This average number of clusters changes depending upon where each region starts. However, the tendency does not change, as revealed in the left panel of Figure 4.18: on average there are more clusters for the LoS and Tunnel regions, than for the NLoS regions. However, it is not expected to see slightly more clusters on average for the NLoS region 2 than for the NLoS region 1. Even more, the number of non-singleton clusters (clusters with more than one observation) remains slightly higher for the NLoS region 2 than for the NLoS region 1, as shown in the right panel of Figure 4.18. This might be indicative of what we have said previously: as propagation seems to strongly depend on rooftop propagation for the NLoS region 2, thus propagating through similar propagation paths, and giving similar IOs and clusters.

From the channel parameter estimates, the power per propagation path α is available. The power per cluster $PW(C)$ would be simply

$$PW(C) = \sum_{k \in C} \alpha_k \quad (4.21)$$

and the power per snapshot

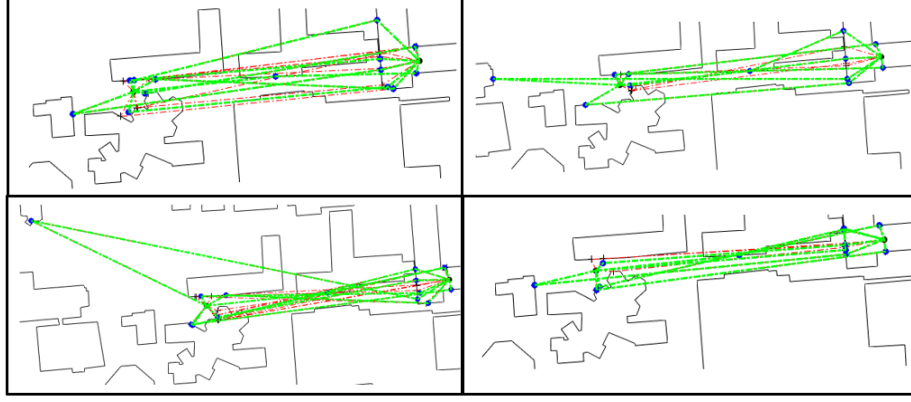


Figure 4.14: Four time snapshot with their clusters for the LoS region. Time snapshots 776 to 779. Each blue filled circle pair (p, q) represent a centroid of the cluster. The green dotted lines are the propagation paths for the IOs corresponding to those centroids. Black crosses represent IOs and red dotted lines the propagation paths for the corresponding IOs. Distance threshold for d_{IO} is 10 m.

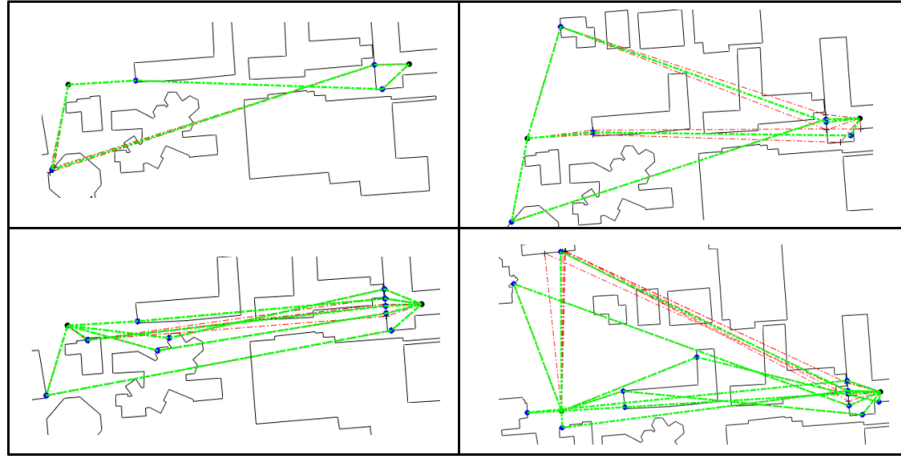


Figure 4.15: Four time snapshot with their clusters for NLoS 1 region. Time snapshots 2163 to 2166. Each blue filled circle pair (p, q) represent a centroid of the cluster. The green dotted lines are the propagation paths for the IOs corresponding to those centroids. Black crosses represent IOs and red dotted lines the propagation paths for the corresponding IOs. Distance threshold for d_{IO} is 10 m.

$$PW_t = \sum_{l=1}^{L_t} \alpha_l \quad (4.22)$$

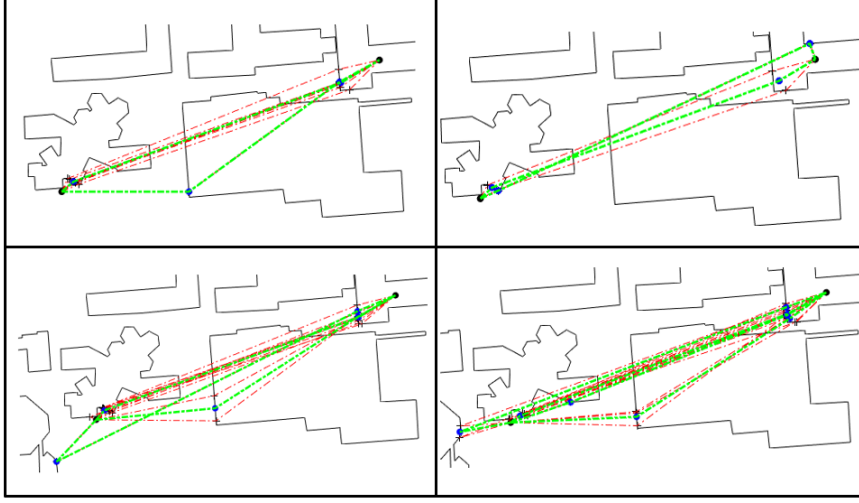


Figure 4.16: Four time snapshot with their clusters for NLoS 2 region. Time snapshots 4781 to 4784. Each blue filled circle pair (p, q) represent a centroid of the cluster. The green dotted lines are the propagation paths for the IOs corresponding to those centroids. Black crosses represent IOs and red dotted lines the propagation paths for the corresponding IOs. Distance threshold for d_{IO} is 10 m.

where L_t is the number of propagation paths per snapshot. Figure 4.19 shows the power per snapshot for the Saint Barbe route with Maxwell BS. The power profile seen is similar to the profile of the received power shown in Figure of Chapter 1. Differences arise due to non-modelled $\mathbf{H}_{DMC}$ channel matrix. After the mobile device turns in the Saint Barbe route (from Ramp 1 to Ramp 2, Figure of Chapter 2), the accumulated power of all propagation paths, that is to say, the power per snapshot, decreases and is lower than in NLoS region 2. If propagation is strongly occurring through one mechanism or reduced number of mechanisms of propagation, it is expected to have low variance in the angles of arrival.

In other words, lower variance of the AoA implies energy is arriving from the same direction, and although this doesn't imply by itself that the propagation paths are similar, we have seen from the Figure 4.18 that the number of non-singleton clusters is lower for NLoS region 1, meaning more observations are grouped into a cluster, and thus, our assumption that propagation paths are similar is well supported. This assumption is confirmed by Figure 4.20. There, the circular standard deviation $\sigma_{\theta_{RX}}$ of the AoA is shown as function of the time snapshots. The highlighted region corresponds to NLoS region 1, and is the region for which the standard deviation is the lowest.

As the distance threshold d_{IO} is increased, it is expected that cluster groups

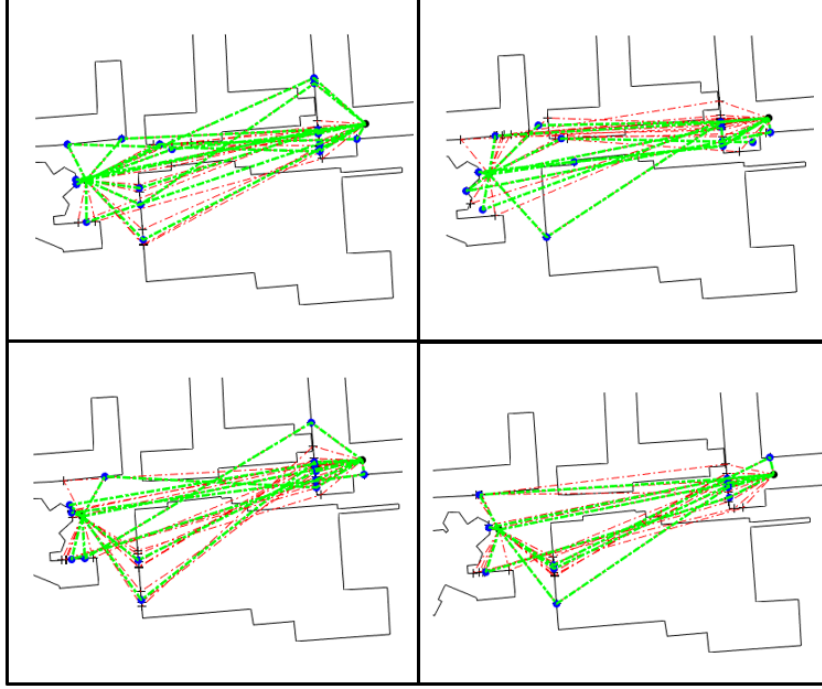


Figure 4.17: Four time snapshot with their clusters for Tunnel region. Time snapshots 6364 to 6367. Each blue filled circle pair (p, q) represent a centroid of the cluster. The green dotted lines are the propagation paths for the IOs corresponding to those centroids. Black crosses represent IOs and red dotted lines the propagation paths for the corresponding IOs. Distance threshold for d_{IO} is 10 m.

have bigger sizes. However, it doesn't mean that the clusters found remain the same, meaning have the same propagation paths. To see how the distance threshold changes the cluster composition let us define a set of distance thresholds $d_{IO,i} = 9 + i$, $i = 1, 11$, for which the clusters within a snapshot are computed. We denote the cluster i from the time snapshot t , with the distance threshold d by $C_{i,t,d}$. Each centroid $c(C_{i,t,d})$ of a snapshot is compared with the centroids from the same snapshot with a different threshold. The mean distance of between centroids of the same snapshot with different thresholds, for all the snapshots is computed. The interval $\delta_i = d_{IO,i+1} - d_{IO,i} = 1$ always remain constant. However, it is expected that for higher distance thresholds the difference between the partition with a given distance threshold and the partition with the distance threshold used immediately before decreases. Figure 4.21 confirms this for both BS measurements. This means that at some distance threshold the partition produced doesn't change too much compared to the partition produced with a distance threshold lower.

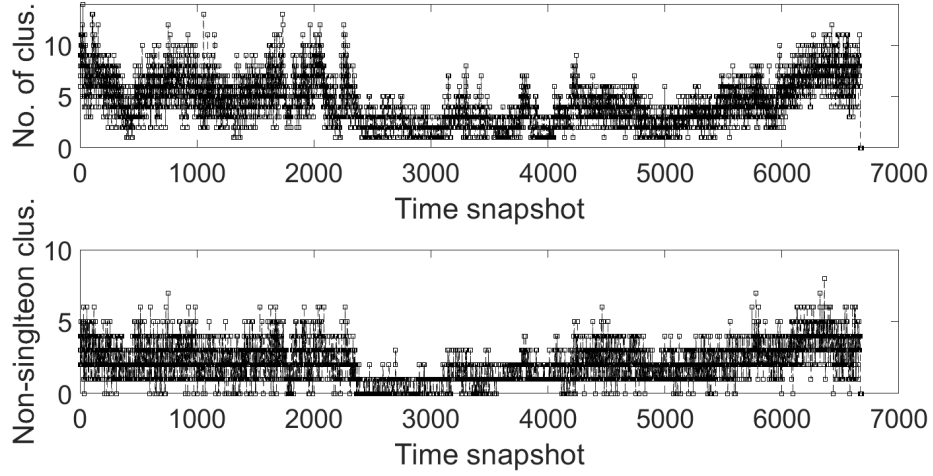


Figure 4.18: Upper panel: number of clusters for the second measurement of the Saint Barbe route with Maxwell BS. Bottom panel: number of non-singleton clusters for the second measurement of the Saint Barbe route with Maxwell BS.

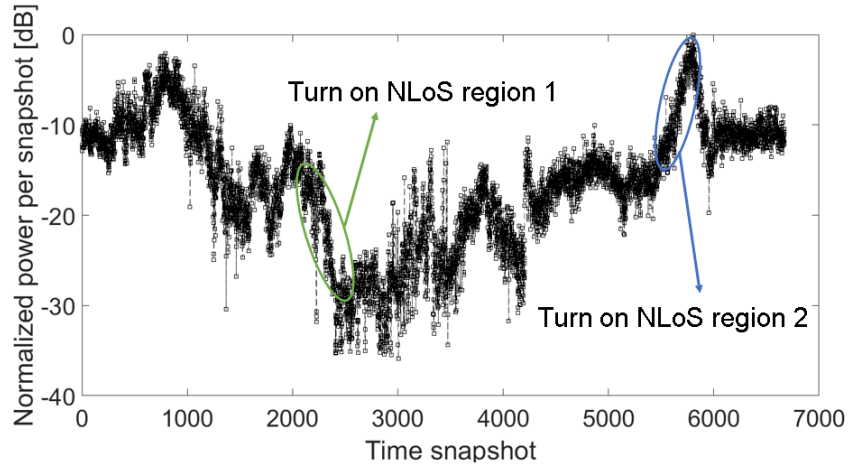


Figure 4.19: Power per snapshot as the sum of the power per propagation path for the measurement 2 of the Saint Barbe route with Maxwell BS.

4.4 Conclusion

Challenges of clustering of channel parameters have been exposed in this chapter. Due to the lack of a ground-truth partition for the estimated channel parameters, the only ultimate truthful methodology to decide whether a partition of the channel parameters is correct is expert visual inspection. This is due to the fact that no index measuring the goodness of a partition seems

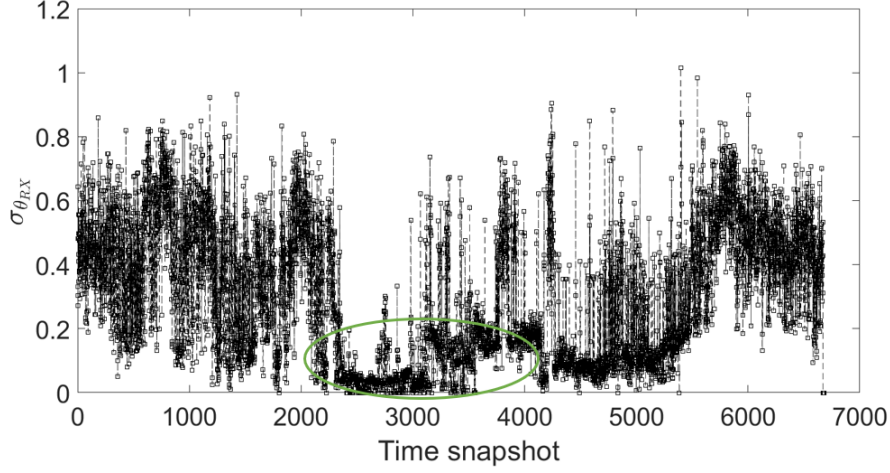


Figure 4.20: Circular standard deviation for the AoA (θ_{RX}) for the measurement 2 of the Saint barbe route with Maxwell BS.

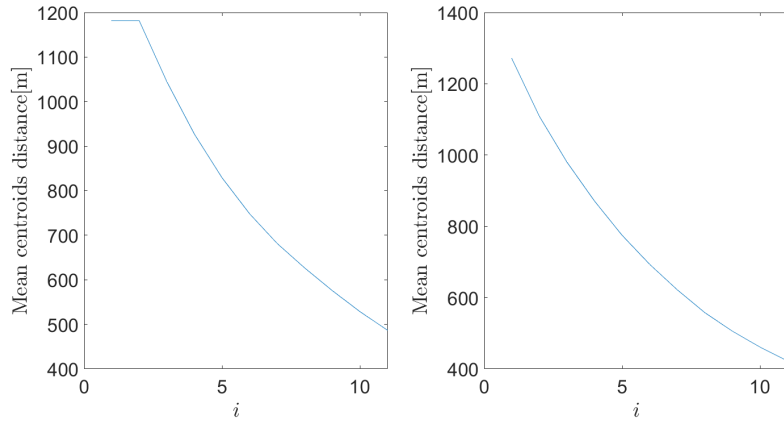


Figure 4.21: Mean centroids distance when increasing the distance threshold. Left panel: Maxwell BS. Right panel: Mercator BS. The interval $d_{IO,i+1} - d_i = 1$ remains the same for each i . The x axis indicate the partitions compared: for $i = 1$ partitions with $d_{IO} = 10$ and $d_{IO} = 11$ are compared; for $i = 2$ partitions with $d_{IO} = 11$ and $d_{IO} = 12$, and so on.

to provide definitive results in accuracy, when compared to visual inspection. Besides so, traditional clustering methods as KPowerMeans introduce parameters that strongly rely on heuristic methods. Moreover, methods like KPowerMeans have been traditionally applied directly to the channel parameters, so the clustering is done in a parameter space for which no intuition is provided. This, along with the fact that the most accurate partition ac-

complished for a set of channel parameters relies on visual inspection, make channel parameter clustering in the parameter space even harder. To tackle that, an improvement in intuition compared to previous approaches was proposed. Before clustering, channel parameters are physically mapped to the scenario in order to identify propagation paths. Clustering is done in a conventional euclidean space for which we have intuition. A defined distance measure capturing the dissimilarity of propagation paths (piecewise line segments) allows easier comparisons between observations than the MCD distance traditionally used in KPowerMeans. Hierarchical clustering proves to be the method that do not rely in any other parameter rather than the threshold cut of the dendrogram, which is strongly tied with the intuitive distance measure defined.

As a side note to take into account for large datasets, it must be said that hierarchical clustering is not the most suitable method for such datasets. However, if sub-sampling the dataset do not implies losing doppler information, or information about the time-variance of the channel, hierarchical clustering could be still applied. In cases where subsampling is not possible, a two-phase methodology can be used: a first phase of clustering the large data into a predefined number of clusters in the order of hundreds, by using for example, k-means; and a second phase of using hierarchical clustering on the clusters previously found in the first phase.

Spatial Loss Field for Outdoor Micro-cells

5

This chapter presents the linear model used to predict shadowing. The two components of the model, coefficients and explanatory variables/covariates are introduced, and the way how to compute them is also explained. This chapter represents the last step in the work-flow introduced in Chapter 1. Previous chapters provide the inputs for the methodology introduced in this chapter to develop a model for shadowing, and ultimately, predict it. Theoretical framework for gaussian processes is presented in order to model the explanatory variables, called in this chapter, the underlying spatial loss field. To estimate such underlying spatial loss field, it is assumed it behaves as an isotropic gaussian random field. Once the underlying field is estimated, a non-stationary gaussian random field is fit to it, in order to model the shadowing correlation in a non-stationary way. The estimation of the underlying field is crucial for the accuracy of the predictions. A discussion about how to estimate the field and its challenges are also presented in this chapter. Finally, the chapter ends by predicting the shadowing from the model, at transmitter-receiver locations not used in the estimation of the explanatory variables.

5.1 Introduction

Losses due to small-scale fading are extensively modelled through the use of Rayleigh and Rician among other distributions. Less attention is pay to shadowing, despite accounting for variations of power over bigger local zones. It is widely assumed that shadowing distribution follows a log-normal distribution, with zero mean and a log-normal variance that depends on the environment of propagation. However, such approximation lacks of detail about the correlation of the received power at different locations. If the local variation around the mean tendency of the path loss is telling us the appearance and disappearance of obstacles, then a better modelling of shadowing can provide a better insight into scenario locations where received power is highly correlated. From the perspective of the operator of a network (i.e. cellular network) this is desirable, since it tells locations of high and low loss, which in turn help to design better power allocation schemes.

Shadowing losses have been considered to be correlated in an iso-tropical

way since Gudmunson paper [Gud91]. Non-isotropical models were proposed in [SYT10], assuming the correlation is still stationary, but anisotropical: depends on the angles of the transceiver positions as well. Anisotropical models improve shadowing correlation models, since they introduce what we can call within this context a sense of correlation by directivity. In other words, anisotropical correlated shadowing is able to account for the presence of obstructions, of a given size. For bigger objects for which correlations, although improbable, might occur, anisotropical models can't capture the correlation between links obstructed by the bigger object. This is due to the fact that, anisotropical models are still stationary, that is to say, that the correlation is not a function of locations, but rather than that a function of the difference of locations. This means that they don't take into account range.

These situations are shown schematically in Figure 5.1. According to the isotropical shadowing correlation model, any pair of links between 1 to 7, should exhibit the same correlation, since the distance separation between the links is the same. It can be seen, however, that there is an object obstructing link 4, which implies that the correlation between link 4 and any link between 1 to 4 (excluding itself) should be different than the correlation between any link from 1 to 4 (excluding link 4). Obstruction losses are higher for link 4 than for links 1 - 3. It can be seen also that links 1 - 4 are grouped in the same sector, while links 5 - 7 are grouped in another sector. All the black filled circles (representing a transceiver) of links 1 - 7 are at the same distance to the blue filled circle (representing another transceiver). It is clear that the correlation between any pair of links in the sector comprising links 1-4 is on average different than the correlation between any pair of links in the sector comprising links 5-7: there is a higher density of obstacles in the latter sector. An anisotropical shadowing correlation model will account for these differences. Lastly, consider links 8 - 10. They are in the same sector of links 5 - 7, but at a larger distance with respect to the transceiver identified by the black filled circle. For an anisotropical model the correlations between any pair of links 8 - 10 is the same as the correlations between any pair of links 5 - 7. We know, however, that it shouldn't be the case, since on one side the density of obstacles is different for each region comprising mentioned links, and on the other side, the distance with respect to the transceiver at the center changes as well. In such a case, shadowing correlations should be non-stationary: the correlation depend on the specific locations of the transceivers, and thus, of the links.

5.2 Covariance estimation

For two given random variables X and Y the covariance of the two variables is given by

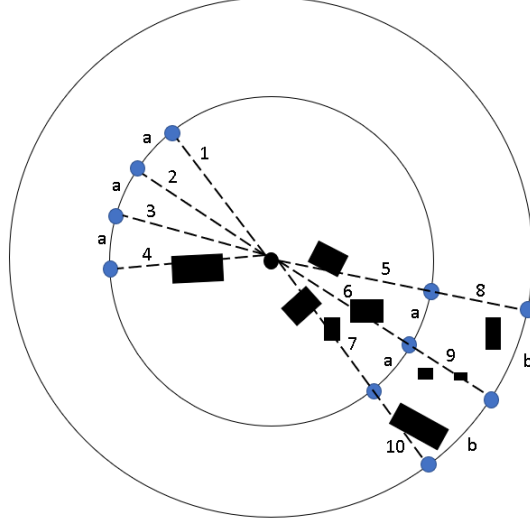


Figure 5.1: Schematic representation of isotropical, anistropical and non-stationary correlation of shadowing. The squares represent obstacles and the circles represent transceivers. A link is composed by any of the 10 blue filled circles and the black filled circle. The blue filled circles in a given black circumference are at the same distance to the black filled circle. Blue filled circles in the same circumference are separated by a distance a or b .

$$\text{COV}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \quad (5.1)$$

where $\mathbb{E}[X]$ is the first moment of the variable or the mean. The covariance provides an insight into the degree of dependence of any of the two variables with respect to the other one. However, covariance doesn't provide an absolute reference for the degree of dependence between the two variables, and for such a reason, it is not straightforward to interpret it in absolute terms. For doing so, instead of the covariance, correlation is introduced as a normalized covariance

$$\rho_{X,Y} = \text{CORR}[X, Y] = \frac{\text{COV}}{\mathbb{E}[(X - \mathbb{E}[X])^2]\mathbb{E}[(Y - \mathbb{E}[Y])^2]} \quad (5.2)$$

In a similar way, covariance and correlation can be defined for random fields. A random field is a stochastic process for which the domain has dimensionality one or higher. It is a common practice to call *spatial process* a random field for which its variable lies in a subset $\mathcal{A}$ of either $\mathbb{R}^2$ or $\mathbb{R}^3$. Let us denote the spatial process by $y(\mathbf{x})$, with $\mathbf{x} \in \mathcal{A}$. The spatial process is second-order stationary if

1. The mean of the process is constant: $\mathbb{E}[y(\mathbf{x})] = \mathbb{E}[y(\mathbf{x} + \mathbf{h})] = cte$.

2. The covariance of the process depends only on the difference on the locations $\mathbf{x}$: $\text{COV}[y(\mathbf{x}), y(\mathbf{x} + \mathbf{h})] = \text{COV}[y(\mathbf{0}), y(\mathbf{h})]$.

If furthermore, the covariance only depends on the distance between locations

$$\blacksquare \text{COV}[y(\mathbf{x}_1), y(\mathbf{x}_2)] \doteq C(\mathbf{x}_1, \mathbf{x}_2) = C(\|\mathbf{h}\|_2)$$

it is an isotropical process. Here $C(\cdot, \cdot)$ is called the covariance function and we use the notation $\mathbf{h} = \mathbf{x}_1 - \mathbf{x}_2$ to denote the difference between two locations. The efforts of modelling a spatial process are put into the proper modelling of the covariance function, since this function captures the dependence structure of the process and how it changes over space [RC15b].

Stationary covariance functions are also called *kernels*. Although the name *kernel* has multiple definitions within different domains, kernels understood as mappings of datasets into a higher dimensional feature space [Gen02], exhibit the positive definite property that is required for any covariance function

$$\sum_i \sum_j \lambda_i \lambda_j k(\mathbf{x}_i, \mathbf{x}_j) = \sum_i \sum_j \lambda_i \lambda_j C(\mathbf{x}_i, \mathbf{x}_j) \geq 0 \quad (5.3)$$

where i and j vary over all finite sequence of points that partition the space $\mathcal{A}$, and $k(\cdot, \cdot)$ denotes the kernel function. Many types of kernels (or covariance functions) can be built from the set of properties list as follows:

- For 2 kernels k_1, k_2 and positive real numbers a, b , $k(\mathbf{x}_1, \mathbf{x}_2) = ak_1(\mathbf{x}_1, \mathbf{x}_2) + bk_2(\mathbf{x}_1, \mathbf{x}_2)$ is a valid kernel.
- For 2 kernels k_1, k_2 , the multiplication of them is also a valid kernel: $k(\mathbf{x}_1, \mathbf{x}_2) = k_1(\mathbf{x}_1, \mathbf{x}_2)k_2(\mathbf{x}_1, \mathbf{x}_2)$.
- For a kernel k_1 and positive real numbers $a_1, \dots, a_n$, the function $k(\mathbf{x}_1, \mathbf{x}_2) = a_1k_1(\mathbf{x}_1, \mathbf{x}_2) + a_2k_2^2(\mathbf{x}_1, \mathbf{x}_2) + \dots + a_nk_1^n(\mathbf{x}_1, \mathbf{x}_2)$ is also a valid kernel.
- For a kernel k_1 , by taking the limit of the series expansion of the exponential function it can be proved as well that $k(\mathbf{x}_1, \mathbf{x}_2) = \exp(k_1(\mathbf{x}_1, \mathbf{x}_2))$.
- For a function $g(\mathbf{x})$ defined over $\mathcal{A}$, the function $k(\mathbf{x}_1, \mathbf{x}_2) = g(\mathbf{x}_1)g(\mathbf{x}_2)$ is also a valid kernel.
- If f is a $\mathbb{R}^P$ -valued function on $\mathcal{A}$ and k_1 is a kernel on $\mathbb{R}^P \times \mathbb{R}^P$, then $k(\mathbf{x}_1, \mathbf{x}_2) = k_1(f(\mathbf{x}_1), f(\mathbf{x}_2))$ is also a valid kernel

It is possible to built many of the stationary kernel by using their spectral representation introduced by Bochner theorem [Boc55]. This is because to

built any valid kernel it has to be positive semi-definite, and Bochner proved that any stationary kernel $k(\mathbf{h})$, $\mathbf{h} \in \mathbb{R}^p$ is positive definite if and only if it can be written as

$$k(\mathbf{h}) = \int_{\mathbb{R}^p} \cos(\mathbf{w}^T \mathbf{h}) F(d\mathbf{w}) \quad (5.4)$$

with F being a positive finite measure and the previous integral, written in Lebesgue notation. This equation is nothing but the Fourier transform, hence the spectral representation, of the distribution F . More properly, $\frac{F}{k(\mathbf{0})}$ is called the spectral distribution function. For kernels with isotropy, it is possible to derive the kernel from (5.5) by choosing a proper function F [Yag57]: nondecreasing and bounded

$$k(\|\mathbf{h}\|) = \int_0^\infty Z(w\|\mathbf{h}\|) F(dw) \quad (5.5)$$

with

$$Z_v(d) = \left(\frac{2}{d}\right)^{\frac{v-2}{2}} \Gamma\left(\frac{d}{2}\right) J_{\frac{v-2}{2}}(d) \quad (5.6)$$

where $\Gamma(\cdot)$ is the gamma function and $J_v(d)$ is the Bessel function of the first kind of order v . The function $Z_v(d)$ forms a basis for functions in $\mathbb{R}^p$. From their bounded nondecreasing spectral representation, many well-known kernels can be defined. Some of the most used are summarized in Table 5.1 and shown in Figure 5.2. Some of the kernels in Table 5.1 are particular cases of the Matérn family of functions. For instance, when $\nu = 0.5$ the Matérn kernel is the exponential kernel, and as $\nu \rightarrow \infty$ the Matérn converges to the Gaussian kernel.

The choice of the kernel depends on the dataset: different datasets require different kernels, so that the kernel can capture the structure of the data. For instance, datasets that have periodical structure would require a kernel that accounts for such periodicity: the correlations between points separated by certain distance factors, might exhibit some correlation higher. This is the case of the wave kernel k_W , for which correlations peak appear each $\frac{\pi}{\theta}$. From a computationally perspective, for example, compactly supported kernels as the circular kernel, are an advantage due to the fact that they produce sparse covariance matrices. For a large dataset prohibitive computations of covariance matrices could be solved through the use of a compactly supported kernel.

On the hand, each of the kernels shown in Table 5.1 are monotonic, meaning they decrease as the distance $h = \|\mathbf{h}\|$ decreases. Non-monotical kernels do not necessarily decrease with increasing distance, and are very useful for

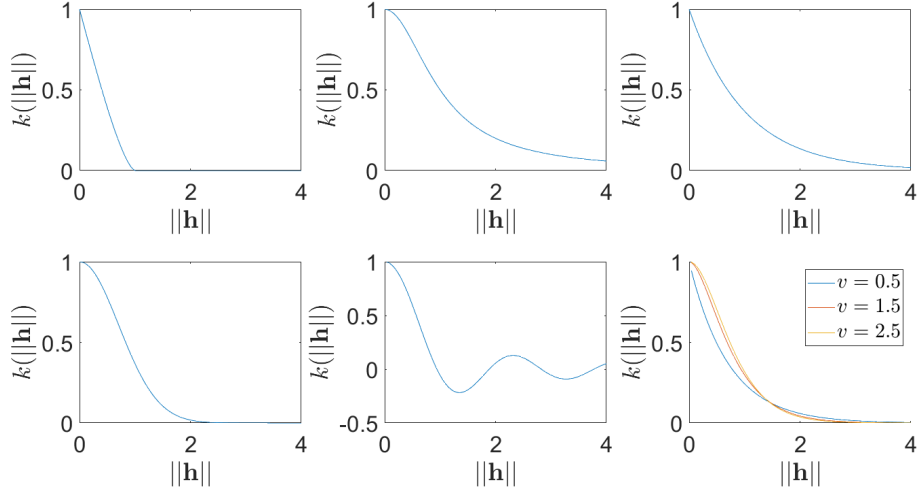


Figure 5.2: Examples of stationary isotropic kernels. Up left: circular kernel with $\theta = 1$. Up middle: rational quadratic with $\theta = 1$. Up right: exponential with $\theta = 1$. Bottom left: gaussian with $\theta = 1$. Bottom middle: wave with $\theta = 0.3$. Bottom right: Matérn with different parameter ν values.

Table 5.1: Common used stationary kernels

Name	Equation
Circular: $k_C(h; \theta)$	$\frac{2}{\pi} \arccos(h/\theta) - \frac{2}{\pi} \frac{h}{\theta} \sqrt{1 - \frac{h^2}{\theta^2}}$
Rational quadratic: $k_{RQ}(h; \theta)$	$1 - \frac{h^2}{h^2 + \theta}$
Exponential: $k_E(h; \sigma, l)$	$\sigma^2 \exp\left(-\frac{h}{l}\right)$
Gaussian: $k_{SE}(h; \sigma, l)$	$\sigma^2 \exp\left(-\frac{h^2}{l}\right)$
Wave: $k_W(h; \theta)$	$\frac{\theta}{h} \sin\left(\frac{h}{\theta}\right)$
Matérn: $k_M(h; \nu)$	$\frac{2^{1-\nu}}{\Gamma(\nu)} (2\sqrt{\nu}h)^\nu H_\nu(2\sqrt{\nu}h)^a$

^a H_ν denotes the modified Bessel function of the second kind of order ν .

modelling stationary spatial processes. In such cases, any non-periodical spatial process could be approximated by a combination of periodical basis functions with random amplitudes a_r, b_r and deterministic frequencies $\mathbf{w}_r$

$$f(\mathbf{x}) = \sum_{r=1}^m a_r \cos(2\pi \mathbf{w}_r^T \mathbf{x}) + b_r \sin(2\pi \mathbf{w}_r^T \mathbf{x}) \quad (5.7)$$

If the spatial process is non-periodic, by a suitable choice of the frequencies $\mathbf{w}_r$, the period of the process could be made arbitrarily long, such that the periodicity is in practical terms removed [Gen02]. Other non-monotonic kernel approaches have been presented in [SD16], [WA13].

5.3 Gaussian regression

Let $\mathbf{y} = y_1, \dots, y_n$ be a n -dimensional random vector following a multivariate gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance $\mathbf{K}$

$$p(\mathbf{y}) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}) = (2\pi^n |\mathbf{K}|)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{y} - \boldsymbol{\mu})\right) \quad (5.8)$$

A Gaussian Process (GP) is in a sense a generalization of the random vector. Loosely mathematical speaking, a GP is an infinite dimensional process, that is to say, defined for each and all locations $\mathbf{x}$ in the subset $\mathcal{A}$ of $\mathbb{R}^p$, for which any sampling of the process at a finite sequence of locations $\mathcal{X} = \mathbf{x}_1, \dots, \mathbf{x}_n$ is a multivariate gaussian random vector. It is important to clarify, that locations refer to points in a p -dimensional space. GP are useful because they allow to generalize the concepts of linear regression, by using a non-parametric and probabilistic Bayesian approach to the regression problem.

The regression problem, through the use of GP, consist in finding the parameters for which an unknown *latent* function explains a set of observed outputs. It is thus an inferential task. Given a *training* data set $\mathcal{D} = \{\mathbf{x}_i, y_i \mid i = 1, \dots, n\}$ with inputs $\mathbf{x}_i$ and outputs y_i , the model

$$y_i = f(\mathbf{x}_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_{NOI}^2) \quad (5.9)$$

assumes observed outputs are produced by the *latent* function f which is corrupted by additive gaussian noise with zero mean and variance σ_{NOI}^2 . GP regression consists in assuming a *prior* gaussian probability density for the latent function evaluated at the input locations

$$\mathbf{f} = \begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_n) \end{bmatrix}, \rightarrow p(\mathbf{f}|\mathcal{X}) \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_f) \quad (5.10)$$

where the covariance (sometimes refer as Gram) matrix $\mathbf{K}_f$ is defined as $[\mathbf{K}_f]_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{E}[f(\mathbf{x}_i)f(\mathbf{x}_j)]$. The combination between the latent function and the assumed noise distribution gives rise to the likelihood function. This likelihood as will be seen is gaussian for the model in (5.9), but

in general doesn't have to be gaussian. It will depend on assumptions about both the *latent* function and the noise probability distribution.

As the covariance matrix $\mathbf{K}_f$ is produced by the kernel function chosen to be suitable to fit to the training dataset, we know that there are some *hyperparameters* that might be known beforehand to be able to compute such matrix. If the *hyperparameters* are not known beforehand, which is the usual case, they have to be inferred from the gaussian process model and the training dataset. On the other hand, the purpose of any regression problem is to fit the model to the dataset, once estimated or given the *hyperparameters*, in order to make predictions for a *test* set $\mathcal{X}^* = \{\mathbf{x}_i^*\}_{i=1}^k$ of new input locations.

Assuming a gaussian prior for the *latent* function at the test locations $\mathbf{f}^* = f(\mathcal{X})$, allows to write jointly

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}^* \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \mathbf{K}(\mathcal{X}, \mathcal{X}) + \sigma_{NOI}^2 \mathbf{I}_{\mathcal{X}} & \mathbf{K}(\mathcal{X}, \mathcal{X}^*) \\ \mathbf{K}(\mathcal{X}^*, \mathcal{X}) & \mathbf{K}(\mathcal{X}^*, \mathcal{X}^*) \end{bmatrix} \right) \quad (5.11)$$

where $\mathbf{I}_{\mathcal{X}}$ stands for an identity matrix of dimension $n \times n$, and $\mathbf{K}(\mathcal{X}, \mathcal{X})$, $\mathbf{K}(\mathcal{X}, \mathcal{X}^*)$, $\mathbf{K}(\mathcal{X}^*, \mathcal{X})$ and $\mathbf{K}(\mathcal{X}^*, \mathcal{X}^*)$ stands for the evaluation of the kernel function solely at the training set, at the training and test sets, at the test and training sets, and solely at the test set. It is possible, due to the properties of the gaussian probability, to condition the function values at test locations $\mathbf{f}^*$ on the observed outputs, giving thus a predictive distribution from which precisely predictions can be done

$$\mathbf{f}^* | \mathbf{y}, \mathcal{X}, \mathcal{X}^* \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{f}^* | \mathbf{y}}, \mathbf{K}_{\mathbf{f}^* | \mathbf{y}}) \quad (5.12)$$

with predictive mean

$$\boldsymbol{\mu}_{\mathbf{f}^* | \mathbf{y}} = \mathbf{K}(\mathcal{X}^*, \mathcal{X}) [\mathbf{K}(\mathcal{X}, \mathcal{X}) + \sigma_{NOI}^2 \mathbf{I}_{\mathcal{X}}]^{-1} \mathbf{y} \quad (5.13)$$

and predictive covariance

$$\mathbf{K}_{\mathbf{f}^* | \mathbf{y}} = \mathbf{K}(\mathcal{X}^*, \mathcal{X}^*) - \mathbf{K}(\mathcal{X}^*, \mathcal{X}) [\mathbf{K}(\mathcal{X}, \mathcal{X}) + \sigma_{NOI}^2 \mathbf{I}_{\mathcal{X}}]^{-1} \mathbf{K}(\mathcal{X}, \mathcal{X}^*) \quad (5.14)$$

In order to estimate the *hyperparameters* the marginal likelihood or *evidence* of the data is introduced

$$p(\mathbf{y} | \mathcal{X}, \theta) = \int p(\mathbf{y} | \mathbf{f}, \mathcal{X}, \theta) p(\mathbf{f} | \mathcal{X}, \theta) d\mathbf{f} \quad (5.15)$$

where $p(\mathbf{y} | \mathbf{f}, \mathcal{X}, \theta)$ is the likelihood, $p(\mathbf{f} | \mathcal{X}, \theta)$ is the prior assumed for the latent function, and θ comprises all the *hyperparameters* involved in model: the kernel *hyperparameters* and the noise variance. Both components of the integral are known. The likelihood is a gaussian $\mathbf{y} | \mathbf{f}, \mathcal{X}, \theta \sim \mathcal{N}(\mathbf{f}, \sigma_{NOI}^2 \mathbf{I}_{\mathcal{X}})$

as well as the prior $\mathbf{f}|\mathcal{X}, \theta \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_f)$, and there is no loss of generality by assuming a mean $\mathbf{0}$ for the latent function. From the properties of the multiplication of gaussians, and after taking the natural logarithm, equation (5.15) transforms into

$$\log(\mathbf{y}|\mathcal{X}, \theta) = -\frac{1}{2}\mathbf{y}^T (\mathbf{K}_f + \sigma_{NOI}^2 \mathbf{I}_{\mathcal{X}})^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K}_f + \sigma_{NOI}^2 \mathbf{I}_{\mathcal{X}}| - \frac{n}{2} \log(2\pi) \quad (5.16)$$

where $|\cdot|$ is the determinant, and (5.16) was already clear from (5.9) and (5.10) giving $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_f + \sigma_{NOI}^2 \mathbf{I}_{\mathcal{X}})$.

An example of the results of such regression process is shown in Figure 5.3. The observed $\mathbf{y}$ received power for the Saint Barbe route and Mercator BS is used, after some averaging to remove part of the noise. Its mean is subtracted to preserve the zero mean gaussian assumption. By assuming the stationary gaussian kernel $k_{SE}(\cdot)$ in Table 5.1, the mean of the predictive distribution $\mu_{f^*|y}$ is found. One hundred uniformly distributed training points $\mathcal{D} = \{x_i, y_i\}_{i=1}^{100}$ were used at training step to fit the regression model to the observed data. The predictions were made at test time points, for which we know the true value (represented by the red crosses in the Figure).

It can be seen that the predictive mean function behaves in a smooth way. This is because we are using a gaussian kernel, on one side, and also because the kernel is stationary. As can be seen, the true received power at the test points varies in different manners. In other words: the signal is varying fastly, slowly and in between at the same time. For such a signal, the stationary gaussian kernel can't account for all this changes. In fact, the length-scale parameter l of the gaussian kernel which controls the distance $h = \|\mathbf{h}\| = \|x_i - x_j\| = |x_i - x_j|$ at which the correlation between signal values decays significantly, is a constant. This means that correlations between signal values at farther points are less and less probable. Moreover, stronger correlations at some locations might be better modelled by other a different value of length-scale parameter, but since it remains constant those correlations are not able to be modelled as strong as they are. The Gram matrix $\mathbf{K}_y$ with the kernel entries $k_{SE}(t_1, t_2)$ shows the isotropic behaviour. The highest dependence is encountered for the diagonal $t_1 = t_2$; covariance values decrease as we move away from the diagonal.

In principle, any physical signal is non-stationary due to the stochastic behaviour imposed by noise phenomena. However, for a given scale of accuracy and also for a given range of point locations x we might assume a signal behaves as a stationary signal. For such regions, the stationary kernel models well enough the signal and there is no need of further non-stationary modelling. Furthermore, accuracy is increased as the number of training points and how well spread are they in the input space increase. This can be seen in

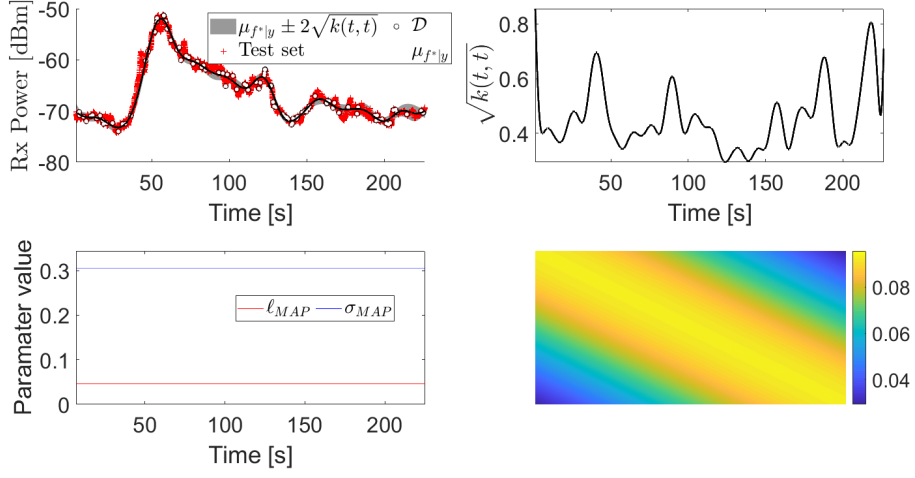


Figure 5.3: Gaussian regression with isotropic kernel for the received power over the Saint Barbe route and Mercator BS. Top left: training samples $\mathcal{D}$ and predicted received power at test locations. Top right: predictive standard deviation $\sqrt{k(t, t)}$. Bottom left: MAP estimates for the parameters l, σ^2 of the $k_{SE}(\cdot)$ kernel. Bottom right: covariance matrix $\mathbf{K}_y$ corresponding to the stationary kernel function $k(t_1, t_2) = k(|t_1 - t_2|)$.

the graph of the predictive standard deviation $\sqrt{k(t, t)}$: minimum values are reached when more training points well spread were available.

Anisotropical versions of the kernel functions in Table 5.1 could be made by substituting the euclidean distance h by the Mahalanobis distance

$$h = \|\mathbf{x}_i - \mathbf{x}_j\| \rightarrow h_M = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j) \quad (5.17)$$

where $\mathbf{M} \in \mathbb{R}^{p \times p}$ is matrix whose diagonal entries control the relevance of the correspondent dimension [VW99]. This entries have the meaning of the inverse of the *length-scale* parameter: a small *length-scale* highlights the relevance of the dimension tied to the length-scale parameter. Moreover, the matrix $\mathbf{M}$ can be decomposed as [WR06]

$$\mathbf{M} = \mathbf{\Lambda} \mathbf{\Lambda}^T + \mathbf{\Psi} \quad (5.18)$$

where $\mathbf{\Lambda} \in \mathbb{R}^{p \times k}$ is a matrix whose columns define k directions of relevance, and $\mathbf{\Psi}$ is the diagonal version of matrix $\mathbf{M}$ controlling each dimension relevance. This low-rank-plus-diagonal factorization of $\mathbf{M}$ requires $p + 2 * k$ (p for $\mathbf{\Psi}$, and the magnitude and angle for each direction vector in $\mathbf{\Lambda}$) parameters, in contrast to the p^2 initial parameters required for $\mathbf{M}$. For high-dimensional input datasets this boils down computational load.

5.4 Hyperparameters inference

Inference of the hyperparameters θ can be done by Maximum A Posteriori (MAP) estimation of the *latent posterior* $p(\theta|\mathbf{y})$, where θ gathers the parameters of the kernel function and variance noise of the model. From Bayes theorem the *latent posterior* could be estimated by

$$p(\theta|\mathbf{y}, \mathcal{X}) \propto p(\mathbf{y}|\theta, \mathcal{X})p(\theta|\mathcal{X}) \quad (5.19)$$

where the proportionality symbol means the normalization constant $p(\mathbf{y}|\mathcal{X})$ is not taken into account. The purpose of the MAP procedure is to find

$$\theta_{MAP} = \arg \max_{\theta} p(\mathbf{y}|\theta, \mathcal{X})p(\theta|\mathcal{X}) \quad (5.20)$$

or equivalently the logarithm of the posterior

$$\theta_{MAP} = \arg \max_{\theta} \log(p(\mathbf{y}|\theta, \mathcal{X})) + \log(p(\theta, \mathcal{X})) = \arg \max_{\theta} L(\theta) \quad (5.21)$$

Since θ comprises all the parameters to be estimated, it represents a joint density probability function. The prior we impose to θ could be set from some knowledge about the hyperparameters or no knowledge at all. If all the parameters are considered to be independent, the prior for them could be rewritten simply as $p(\theta|\mathcal{X}) = \prod_i p(\theta_i|\mathcal{X})$, where θ_i represents *each possible* parameter. We drop the conditioning on $\mathcal{X}$ for simplicity in notation, since its understood the input dataset is taken into consideration. The likelihood of the data $p(\mathbf{y}|\theta, \mathcal{X})$ is

$$L(\theta) = \log(\mathcal{N}(\mathbf{0}, \mathbf{K}_y)) + \sum_i \log(p(\theta_i)) = -\frac{1}{2} \log(2\pi)^p - \frac{1}{2} \log |\mathbf{K}_y| \quad (5.22)$$

$$-\frac{1}{2} \mathbf{y}^T \mathbf{K}_y^{-1} \mathbf{y} + \sum_i \log(p(\theta_i))$$

where $\mathbf{K}_y = \mathbf{K}_f + \sigma_{NOI}^2 \mathbf{I}_X$. To find the maximum of $L(\theta)$, its gradient is required

$$\nabla L(\theta) = \begin{bmatrix} \frac{\partial L}{\partial \theta_1} \\ \frac{\partial L}{\partial \theta_2} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial L}{\partial \theta_m} \end{bmatrix} \quad (5.23)$$

where m is the number of parameters. After such computation, typically an iterative optimization algorithm (i.e. BFGS, gradient descent, ...) providing local or global maximum is required, since an analytical solution is not always feasible.

5.5 Non-stationary kernels

Anisotropic kernels as the one defined by matrices in 5.18 are still stationary. It means that still smooth interpolation for the training dataset is used, making it hard to detect any periodic or long-range correlation between samples in it [RHK17a]. Literature about non-stationary kernel functions is broad. Different methodologies have been derived to tackle it. According to the classification done in [Gel+10], four methodologies are predominant

1. Smoothing and weighted-average methods.
2. Basis function methods.
3. Deformations of input space.
4. Spatially-varying parameters or process convolutions.

The first methodology divides the input space $\mathcal{A}$ into M disjoint regions $\mathcal{S}_m$, $m = 1, \dots, M$ such that $\mathcal{A} = \bigcup_{m=1}^M \mathcal{S}_m$. After so, separately, gaussian regression is applied for each of these subregions (as in section 5.3) using as the training set of each region, the points located within the region ($\mathbf{x}_i \in \mathcal{S}_m$) and as a kernel function, one of the stationary kernels in Table 5.1. To ensure the global spatial process behaves in a non-stationary way, different approaches are followed. In [FS01], given local spatial process $y_1(\mathcal{S}_1), \dots, y_M(\mathcal{S}_M)$, the global process is built from

$$y(\mathbf{x}), \mathbf{x} \in \mathcal{A} = \sum_{m=1}^M w_m(\mathbf{x}) y_m(\mathcal{S}_m) \quad (5.24)$$

with $w_m(\mathbf{x})$ as a weighting function taking into account the distance between the point $\mathbf{x}$ and the center of region $\mathcal{S}_m$. In this way the non-stationarity is built from simpler to estimate problems, but the definition of the regions $\mathcal{S}_m$ and the number of them is not trivial. Approaches might consider model selection criteria as the Bayesian Information Criterion (BIC), but defining the shapes is not a trivial task. The bigger the regions and the irregular the shapes, the less meaningful is the use of the weighting criteria to impose non-stationarity into the spatial process.

The second methodology is based on the Karhunen-Loève decomposition of a zero mean spatial process

$$y(\mathbf{s}) = \sum_{l=1}^{\infty} \lambda_l W_l \phi_l(\mathbf{x}) \quad (5.25)$$

where λ_l are weighing factors, W_l standardized and uncorrelated random variables, and ϕ_l orthogonal functions. Making use of this decomposition, the covariance function (or kernel function) of the process could be written as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sum_{l=1}^{\infty} \lambda_l \phi_l(\mathbf{x}_i) \phi_l(\mathbf{x}_j) \quad (5.26)$$

In this sense the λ and the orthogonal functions ϕ are the eigenvalues and eigenfunctions of the Fredholm integral equation $\int_{\mathcal{A}} k(\mathbf{x}_i, \mathbf{x}_j) \phi_l(\mathbf{x}_i) d\mathbf{x}_i = \lambda_l \phi_l(\mathbf{x}_j)$. Errors due to required truncation of (5.26) appeared, but their variance can be minimized if the eigenfunctions are the exact solutions to the Fredholm integral equation [Wik10].

Solutions to the Fredholm equation might be computationally expensive. To estimate the eigenvalues and eigenfunctions an empirical approach could be used by computing, if possible, the sample covariance matrix $\hat{\mathbf{K}}_y$. Then, from the spectral decomposition of the sample covariance matrix $\hat{\mathbf{K}}_y = \mathbf{V} \mathbf{D} \mathbf{V}^T$, the eigenvalues are available from the diagonal matrix $\mathbf{D}$, and the eigenfunctions $\phi_l(\mathbf{x})$ are available from the columns of the $\mathbf{V}$ matrix.

The third methodology is based on the idea that locations in the input space could be mapped to a deformed space where stationary holds [SG92]. The mapping $f : \mathcal{A} \rightarrow \mathcal{DS}$, where $\mathcal{DS}$ is the deformed space and the mapping is non-linear and should be ideally one-to-one. Once the transformation is done, the non-stationary kernel function k_{NS} is computed from

$$k_{NS}(\mathbf{x}_i, \mathbf{x}_j) = k_S(\|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|) \quad (5.27)$$

with $k_S(\cdot)$ is any valid isotropical kernel function. The core of this approach is the mapping function f . To estimate it, the original paper by Sampson and Guttorp [SG92] locations in $\mathcal{A}$ are mapped to $\mathcal{DS}$ by scaling each dimension, such that inter point distances in the deformed space represent sample spatial dispersions. After so, spline interpolation is used to fill in the mapping for all the deformed space.

This approach isn't capable of quantifying the uncertainty arising from estimation of f . Due to that and to the common situation of estimating non-one-to-one mappings, Bayesian approaches were proposed, by for example, imposing prior distributions for the transformed locations such that non-smooth and folding mappings are penalized [DSG01]. More approaches were

proposed in [SO03], [AS08]. However, most of this bayesian approaches require highly complex Markov Monte Carlo (MCMC) algorithms difficult to use to fit the model [Ris16].

Lastly, process convolutions or spatially-varying parameters methodology, uses a constructive specification of a gaussian process in order to introduce non-stationary behaviour. As an alternative to the Karhunen-Loève expansion, a stochastic process $y(\mathbf{x})$ could be written as the convolution between a kernel function and another stochastic process W

$$y(\mathbf{x}) = \int_{\mathcal{A}} k_{\mathbf{x}}(\mathbf{u}) dW(\mathbf{u}) \quad (5.28)$$

In this context, the function $k_{\mathbf{x}}(\cdot)$ is not a kernel in the sense has been used previously. It is a spatially varying kernel function centred at $\mathbf{x}$. Besides so, it has to fulfil the following two conditions: $\int_{\mathbb{R}^p} k(\mathbf{u}) d\mathbf{u} < \infty$ and $\int_{\mathbb{R}^p} k(\mathbf{u})^2 d\mathbf{u} < \infty$, but there is no positive definite condition, since by definition such a kernel function is a mapping of the type $k : \mathbb{R}^p \rightarrow \mathbb{R}$, while the covariance (kernel) function is a mapping of the type $k : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$. If W is a Brownian motion, the process $y(\mathbf{s})$ in (5.28) defines a non-stationary GP with zero mean [RC15a].

The advantage of expressing the spatial process as in (5.28) is that it allows to fully express the covariance function through the use of such a less restrictive kernel

$$\text{COV}[y(\mathbf{x}_i), y(\mathbf{x}_j)] = \int_{\mathcal{A}} k_{\mathbf{x}_i}(\mathbf{u}) k_{\mathbf{x}_j}(\mathbf{u}) d\mathbf{u} \quad (5.29)$$

Then, two ways of taking advantage of such characterization of the spatial process $y(\mathbf{x})$ are of particular interest. In [Hig98], (5.28) is rewritten as $y(\mathbf{x}) = \int_{\mathcal{A}} k(\mathbf{s} - \mathbf{u}) V(\mathbf{u}) d\mathbf{u}$ with $V(\cdot)$ as Gaussian white noise. Then, the spatial process is approximated by partitioning the space $\mathcal{A}$ into a grid of locations $\mathcal{A}_g = \{\mathbf{u}_1, \dots, \mathbf{u}_L\}$

$$y(\mathbf{x}) = \sum_{l=1}^L k_{\mathbf{x}}(\mathbf{x} - \mathbf{u}_l) V(\mathbf{u}_l) \quad (5.30)$$

The choice of the spatially-varying kernels $k_{\mathbf{x}}(\cdot)$ is the core of the procedure. The underlying idea is to compute such a local kernel by making use of the observed data within a neighbourhood around $\mathbf{x}$. In [Hig98] it is chosen to be a p -dimensional gaussian centred at $\mathbf{x}$ with covariance matrix $\Sigma(\mathbf{x})$. In order to reduce the computational complexity of estimating $k_{\mathbf{x}}(\cdot)$ for all the input space, some basis kernels are used. Again, basis kernels have a different meaning than the one used for the second methodology. Here, since the purpose of such basis kernels is to reduce computational load, few of them are

chosen at evenly selected locations $\mathbf{b}_1, \dots, \mathbf{b}_M$, with $M \ll L$. From such basis kernels, the spatially-varying kernels are computed as weighted-average versions of each and all the basis kernels

$$k_{\mathbf{x}}(\mathbf{u}) = \sum_{m=1}^M w_m(\mathbf{x}) K_m(\mathbf{u}) \quad (5.31)$$

with $K_m(\cdot)$ the basis kernels centred at $\mathbf{b}_m$, and weights

$$w_m(\mathbf{x}) \propto \exp\left(\frac{-\|\mathbf{x} - \mathbf{b}_m\|^2}{2}\right), \quad \forall \mathbf{x} \quad \sum_m w_m(\mathbf{x}) = 1$$

as gaussian functions smoothly weighting farther distances between the location and the basis kernel center as less relevant. The basis kernels are estimated from fitting an anisotropic Gaussian variogram model to the data within a specified radius R around $\mathbf{x}$. The variogram is a well-know function in the context of geostatistics (where the spatial process typically lies in $\mathbb{R}^2$) that measure spatial dependency for a stochastic process. To estimate it, a sample version of it is available from the values of the stochastic process $y_i = y(\mathbf{x}_i)$ at locations $\mathbf{x}_1, \dots, \mathbf{x}_n$.

Algorithm 7 Isotropic variogram computation

Input: Training set $\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_n \mid y_1, \dots, y_n\}$, Number of bins: N_b

Output: $\beta(h \pm \delta)$

- 1: Find $d_{MAX} = \max_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|$, $d_{MIN} = \min_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|$
 - 2: Set the bin width $\delta = \frac{d_{MAX} - d_{MIN}}{N_b}$
 - 3: Find the bin edges $BE_1, \dots, BE$ from: $BE_i = d_{MIN} + i\delta, i = 0, \dots, N_b$
 - 4: For each bin, find the set of locations $\mathcal{NB}(h \pm \delta) = \{(i, j) \mid \|\mathbf{x}_i - \mathbf{x}_j\| = h \pm \delta\}, i, j = 1, \dots, n$
 - 5: For each bin compute $\beta(h \pm \delta) = \frac{1}{|\mathcal{NB}(h \pm \delta)|} \sum_{(i,j) \in \mathcal{NB}} \|\mathbf{x}_i - \mathbf{x}_j\|^2$
-

Higdon et al. [Hig98] further specified the variogram model to be fitted to the empirical variogram

$$\frac{1}{2}\beta(\mathbf{x} - \mathbf{x}') = \sigma_\epsilon^2 + \gamma \left(1 - \exp(-a\mathbf{r}^T \mathbf{r})\right) \quad (5.32)$$

with σ_ϵ^2 as the nugget variance, γ, a parameters of the model, and $\mathbf{r}$ specifying the anisotropic parameters in $\mathbb{R}^2$

$$\mathbf{r} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{bmatrix} \begin{bmatrix} [\mathbf{x}]_1 - [\mathbf{x}']_1 \\ [\mathbf{x}]_2 - [\mathbf{x}']_2 \end{bmatrix} \quad (5.33)$$

and where the parameters σ_1, σ_2, ϕ constitute the local anisotropic parameters for the gaussian basis kernel K_m , whose covariance matrix is given by

$$\Sigma(\mathbf{x})^{-1/2} = \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{bmatrix} \begin{bmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{bmatrix} \quad (5.34)$$

To summarize, as a first step, a selected number M of basis kernels K_m is assumed based on the expertise or any prior information (i.e. locations for which the assumed random field tend to vary more). Secondly, the empirical variogram is found for the the training dataset $\mathcal{D}$. Thirdly, an anisotropic variogram model like the one in (5.32) is fitted to the data, to extract the parameters σ_1, σ_2, ϕ of the spatially-varying covariance matrices $\Sigma(\mathbf{x} = \mathbf{b}_m)$ of the basis kernels K_m . Then, varying kernels $k_{\mathbf{x}}(\cdot)$ are found using 5.31. Lastly, the observed data is reconstructed from (5.30).

Another path could be taken. Equation (5.29) served as the basis for the works of [HSK99], [Pac03] and later [Ste05], whom by assuming a multivariate Gaussian for the kernel functions $k_{\mathbf{x}}(\cdot)$ end-up with a highly flexible non-stationary covariance function

$$k_{NS}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_{\mathbf{x}_i} \sigma(\mathbf{x}_i) \frac{|\Sigma(\mathbf{x}_i)|^{1/4} |\Sigma(\mathbf{x}_j)|^{1/4}}{|\frac{\Sigma(\mathbf{x}_i) + \Sigma(\mathbf{x}_j)}{2}|^{1/2}} g(\sqrt{Q(\mathbf{x}_i, \mathbf{x}_j)}) \quad (5.35)$$

with the scaled distance function $Q(\cdot, \cdot)$ defined as

$$Q(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \left(\frac{\Sigma(\mathbf{x}_i) + \Sigma(\mathbf{x}_j)}{2} \right)^{-1} (\mathbf{x}_i - \mathbf{x}_j) \quad (5.36)$$

and where $g(\cdot)$ is any valid isotropic kernel (covariance) function. The non-stationary covariance in (5.35) has been widely used in [AS11], [KN12], [Kat13], [RC15c], [LS19], due to its high flexibility to model local variances and covariances as spatially-varying. In particular, when setting those parameters to be constant, the resultant kernel is a stationary kernel function from the set of kernel functions in Table 5.1.

5.5.1 Non-stationary gaussian kernel

A non-stationary version of the gaussian kernel $k_{SE}(\cdot, \cdot)$ (Table 5.1) was provided by Gibbs in [Gib98] for 1D input locations $\mathcal{X} \in \mathbb{R} = \{x_1, \dots, x_n\}$, $\mathcal{X}^* \in \mathbb{R} = \{x_1^*, \dots, x_n^*\}$. Such a kernel, shown in (5.37), has spatially-varying variance $\sigma(\cdot)$ and length-scale $l(\cdot)$.

$$k_{NS-SE}(x_i, x_j) = \sigma(x_i) \sigma(x_j) \sqrt{\frac{2l(x_i)l(x_j)}{l(x_i)^2 + l(x_j)^2}} \exp\left(-\frac{(x_i - x_j)^2}{l(x_i)^2 + l(x_j)^2}\right) \quad (5.37)$$

The gradient (5.23) of the *latent* posterior (5.22) must be computed in order to find MAP estimates for the parameters $\sigma(\cdot), l(\cdot)$ of the kernel k_{NS-SE} . As a first step, prior probability distributions $\boldsymbol{\sigma} \doteq [\sigma(x_1), \dots, \sigma(x_n)]^T \sim p(\boldsymbol{\sigma})$, $\mathbf{l} \doteq [l(x_1), \dots, l(x_n)]^T \sim p(\mathbf{l})$ must be assumed for the parameters of the kernel. Multivariate gaussian priors could be assumed for $\boldsymbol{\sigma}$ and $\mathbf{l}$. Since the parameters $l(x), \sigma(x)$ vary over the continuous space $\mathbb{R}$, once the multivariate gaussian assumption for the finite realizations is imposed, a GP as in (5.38) could be assumed for $l(x), \sigma(x)$ [Hei+16]. Moreover, by doing so, it is guaranteed that both spatially-varying parameters vary smoothly over the domain under consideration.

$$\begin{aligned} \log(l(x)) &\doteq \tilde{l}(x) \sim GP(\mu_{\tilde{l}}, k_{SE, \tilde{l}}(x_i, x_j)) \\ \log(\sigma(x)) &\doteq \tilde{\sigma}(x) \sim GP(\mu_{\tilde{\sigma}}, k_{SE, \tilde{\sigma}}(x_i, x_j)) \end{aligned} \quad (5.38)$$

In (5.38), the logarithm of each of the parameters is modelled as a GP, to restrict the parameters $\tilde{l}, \tilde{\sigma}$ to be positive. The kernels functions are chosen to be stationary. Each stationary kernel $k_{SE, l}(\cdot, \cdot), k_{SE, \sigma}(\cdot, \cdot)$ is a gaussian kernel with its own variance α_z and length-scale β_z parameters:

$$k_{SE, z}(x_i, x_j) = \alpha_z \exp\left(-\frac{(x_i - x_j)^2}{2\beta_z}\right) \quad (5.39)$$

where $z = \{\tilde{l}, \tilde{\sigma}\}$ refer to the spatially-varying length-scale $\tilde{l}(\cdot)$ or the spatially-varying variance $\tilde{\sigma}(\cdot)$. We define the vectors $\tilde{\boldsymbol{\sigma}} = [\tilde{\sigma}(x_1), \dots, \tilde{\sigma}(x_n)]^T$, $\tilde{\mathbf{l}} = [\tilde{l}(x_1), \dots, \tilde{l}(x_n)]^T$. MAP estimation objective is to produce an estimate for $\tilde{\boldsymbol{\sigma}}$ and $\tilde{\mathbf{l}}$, which then is used to find $\boldsymbol{\sigma}, \mathbf{l}$. With the prior probabilities for the spatially-varying parameters defined, the gradient (5.23) entries can be now properly stated

$$\begin{aligned} \frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i} &= -\frac{1}{2} \frac{\partial \log |\mathbf{K}_y|}{\partial \theta_i} - \frac{1}{2} \frac{\partial \mathbf{y}^T \mathbf{K}_y^{-1} \mathbf{y}}{\partial \theta_i} \\ &\quad - \frac{1}{2} \frac{\partial \log |\mathbf{K}_{\tilde{\sigma}}|}{\partial \theta_i} - \frac{1}{2} \frac{\partial (\boldsymbol{\sigma} - \mu_{\sigma})^T \mathbf{K}_{\tilde{\sigma}}^{-1} (\tilde{\boldsymbol{\sigma}} - \mu_{\tilde{\sigma}})}{\partial \theta_i} \\ &\quad - \frac{1}{2} \frac{\partial \log |\mathbf{K}_{\tilde{l}}|}{\partial \theta_i} - \frac{1}{2} \frac{\partial (\tilde{\mathbf{l}} - \mu_{\tilde{l}})^T \mathbf{K}_{\tilde{l}}^{-1} (\tilde{\mathbf{l}} - \mu_{\tilde{l}})}{\partial \theta_i} \end{aligned} \quad (5.40)$$

where $[\mathbf{K}_{\tilde{l}}]_{i,j} = k_{SE, \tilde{l}}(x_i, x_j)$, $[\mathbf{K}_{\tilde{\sigma}}]_{i,j} = k_{SE, \tilde{\sigma}}(x_i, x_j)$, θ_i is any among $\{\tilde{\sigma}(x_1), \dots, \tilde{\sigma}(x_n), \tilde{l}(x_1), \dots, \tilde{l}(x_n)\}$, and $\mu_{\tilde{l}}, \mu_{\tilde{\sigma}}$ are constant for any value of x .

Using properties of matrices and random vectors jointly distributed as multivariate gaussians ([PP+08], [WR06]) it is possible to write (5.40) as

$$\begin{aligned}
\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i} &= -\frac{1}{2} \text{tr} \left(\mathbf{K}_y^{-1} \frac{\partial \mathbf{K}_y}{\partial \theta_i} \right) - \frac{1}{2} \text{tr} \left(\mathbf{K}_{\tilde{\mathbf{l}}}^{-1} \frac{\partial \mathbf{K}_{\tilde{\mathbf{l}}}}{\partial \theta_i} \right) - \frac{1}{2} \text{tr} \left(\mathbf{K}_{\tilde{\boldsymbol{\sigma}}}^{-1} \frac{\partial \mathbf{K}_{\tilde{\boldsymbol{\sigma}}}}{\partial \theta_i} \right) \\
&\quad - [\mathbf{K}_{\tilde{\mathbf{l}}}^{-1} (\tilde{\mathbf{l}} - \boldsymbol{\mu}_{\tilde{\mathbf{l}}})]_i - [\mathbf{K}_{\tilde{\boldsymbol{\sigma}}}^{-1} (\tilde{\boldsymbol{\sigma}} - \boldsymbol{\mu}_{\tilde{\boldsymbol{\sigma}}})]_i \\
&= -\frac{1}{2} \text{tr} \left(\frac{\partial (\mathbf{y}^T \mathbf{K}_y^{-1} \mathbf{y})}{\partial \mathbf{K}_y} \frac{\partial \mathbf{K}_y}{\partial \theta_i} \right) - [\mathbf{K}_{\tilde{\mathbf{l}}}^{-1} (\tilde{\mathbf{l}} - \boldsymbol{\mu}_{\tilde{\mathbf{l}}})]_i - [\mathbf{K}_{\tilde{\boldsymbol{\sigma}}}^{-1} (\tilde{\boldsymbol{\sigma}} - \boldsymbol{\mu}_{\tilde{\boldsymbol{\sigma}}})]_i \\
&= \frac{1}{2} \text{tr} \left((\boldsymbol{\alpha} \boldsymbol{\alpha}^T) \frac{\partial \mathbf{K}_y}{\partial \theta_i} \right) - [\mathbf{K}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})]_i
\end{aligned} \tag{5.41}$$

where the derivative of a matrix is defined as

$$\frac{\partial \mathbf{K}_y}{\partial \theta_i} = \begin{bmatrix} \frac{\partial [\mathbf{K}_y]_{1,1}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{K}_y]_{1,n}}{\partial \theta_i} \\ \frac{\partial [\mathbf{K}_y]_{2,1}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{K}_y]_{2,n}}{\partial \theta_i} \\ \vdots & \cdots & \vdots \\ \frac{\partial [\mathbf{K}_y]_{n,1}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{K}_y]_{n,n}}{\partial \theta_i} \end{bmatrix} \tag{5.42}$$

and where $\text{tr}(\cdot)$ denotes the trace of a matrix, $\boldsymbol{\alpha} = \mathbf{K}_y^{-1} \mathbf{y}$, $\boldsymbol{\theta}$ represents either $\tilde{\boldsymbol{\sigma}}$ or $\tilde{\mathbf{l}}$, and $[\cdot]_i$ represents the i -position of the vector enclosed in the brackets. Since the derivatives are taken with respect to the θ_i variable, the index i of such vector enclosed in brackets is associated to the index i of either the vector $\tilde{\mathbf{l}}$ or the vector $\tilde{\boldsymbol{\sigma}}$, with respect of which the derivation is done. Note also, that (5.41) applies to non-stationary kernels having more parameters than the variance ($\boldsymbol{\sigma}$) and length-scale ($\mathbf{l}$), if those parameters distribute as well as GP. Furthermore, even though a particular non-stationary kernel (equation 5.37) was introduced at the beginning of this section, in the derivation of (5.41) no particular structure was assumed for the matrix $\mathbf{K}_y$ besides that it has some parameters θ_i , that could be any type of spatially-varying parameter of the kernel function such that at finite locations the random vector of the parameter evaluated at such locations is a gaussian random vector.

Indistinct of the non-stationary kernel, the derivative $\frac{\partial \mathbf{K}_y}{\partial \theta_i}$ needs to be computed. It is easy to note that this matrix is sparse and has a particular structure. The (i, j) entry of $\mathbf{K}_y$ is nothing but the kernel function $k(x_i, x_j)$.

Let us remember the notation we have used since the definition of the gradient (5.23): θ_k , $k = 1, \dots, m$, denotes any of the m spatially-varying and gaussian distributed parameters of the kernel function (in the case of the kernel 5.37, any entry of $\tilde{\sigma}$ or $\tilde{\mathbf{l}}$). That is to say, that if the number of parameters is d (2 in the case of the kernel in 5.37), then $m = nd$.

To simplify notation and be more clear, let us redefine the convention used for θ_i to understand the structure of $\mathbf{K}_y$. Lets say now that θ represents the parameter symbol (i.e. $\tilde{\sigma}$ or $\tilde{\mathbf{l}}$), and the subscript k represents the location $\theta_k = \theta(x_k)$. In this case $k = 1, \dots, m$ and the derivative with respect to θ_i does not loss its meaning: a derivative with respect to one of all the existing parameters (what changes in this redefined notation is only that we specify which parameter with θ , and which location with i). Thus, implicitly we have that $k(x_i, x_j) = k(x_i, x_j, \theta_i, \theta_j)$. This means that if we derive with respect to parameter θ at location x_k , only the k -th column and the k -th row of $\frac{\partial \mathbf{K}_y}{\partial \theta_i}$ are non-zero

$$\left[\frac{\partial \mathbf{K}_y}{\partial \theta_k} \right]_{i,j} = \begin{cases} 0 & i \neq k, j \neq k \\ \frac{\partial k(x_i, x_j, \theta_i, \theta_j)}{\partial \theta_i} & i = k, j = 1, \dots, n \\ \frac{\partial k(x_i, x_j, \theta_i, \theta_j)}{\partial \theta_j} & j = k, i = 1, \dots, n \end{cases} \quad (5.43)$$

The derivative (5.41) is computed for $i = 1, \dots, n$. A MAP estimate $\theta_{MAP} = \arg \max_{\theta} L(\theta)$ of the parameter $\theta = \{\tilde{\mathbf{l}}, \tilde{\sigma}\}$ is computed for all the training locations.

Predictive distribution mean and covariance presented in (5.13 - 5.14) computes the matrices $\mathbf{K}(\mathcal{X}, \mathcal{X}^*)$, $\mathbf{K}(\mathcal{X}^*, \mathcal{X}^*)$ using the MAP estimates of the parameters of the kernel, since such parameters do not vary over space. However, for non-stationary kernels with spatially-varying parameters, at new locations such parameters are not expected to remain similar to the MAP estimated values obtained using the training locations. It is thus necessary to find a posterior distribution of the spatially-varying parameters at new locations given the MAP parameters at the training locations. Fortunately, since the spatially-varying parameter $\theta = \{\tilde{\sigma}, \tilde{\mathbf{l}}\}$ is modelled as a GP, its posterior distribution $p(\theta^* | \tilde{\theta}_{MAP})$ follows the same structure of (5.12 - 5.14)

$$\begin{aligned} p(\theta^* | \theta_{MAP}) &\sim \mathcal{N}(\mu_{*|MAP}, \Sigma_{*|MAP}) \\ \mu_{*|MAP} &= \mathbf{K}_{\theta}(\mathcal{X}^*, \mathcal{X}) \mathbf{K}_{\theta}(\mathcal{X}, \mathcal{X})^{-1} (\theta - \mu_{\theta}) \\ \Sigma_{*|MAP} &= \mathbf{K}_{\theta}(\mathcal{X}^*, \mathcal{X}^*) - \mathbf{K}_{\theta}(\mathcal{X}^*, \mathcal{X}) \mathbf{K}_{\theta}(\mathcal{X}, \mathcal{X})^{-1} \mathbf{K}_{\theta}(\mathcal{X}, \mathcal{X}^*) \end{aligned} \quad (5.44)$$

This distribution gives an insight into which would be the values of the spatially-varying parameters at the test locations. Using this information, Heinonen *et al.* in [Hei+16] based on [GWB97] redefined the predictive distribution in 5.12 for the non-stationary case with spatially-varying parameters in the following manner

$$p(\mathbf{f}^*|\mathbf{y}, \tilde{\mathbf{l}}_{MAP}, \tilde{\boldsymbol{\sigma}}_{MAP}) = \int \int p(\mathbf{f}^*|\tilde{\mathbf{l}}_{MAP}, \tilde{\mathbf{l}}^*, \tilde{\boldsymbol{\sigma}}_{MAP}, \tilde{\boldsymbol{\sigma}}^*, \mathbf{y}) p(\tilde{\mathbf{l}}^*|\tilde{\mathbf{l}}_{MAP})p(\tilde{\boldsymbol{\sigma}}^*|\tilde{\boldsymbol{\sigma}}_{MAP})d\tilde{\mathbf{l}}^*d\tilde{\boldsymbol{\sigma}}^* \quad (5.45)$$

The distribution $p(\mathbf{f}^*|\mathbf{y}, \tilde{\mathbf{l}}_{MAP}, \tilde{\boldsymbol{\sigma}}_{MAP})$ is the posterior predictive distribution of the latent function, and its notation indicate that besides the observations $\mathbf{y}$, what must be given are the MAP estimates. Thus, implicitly it depends on the parameters posteriors $p(\tilde{\mathbf{l}}^*|\tilde{\mathbf{l}}_{MAP})$, $p(\tilde{\boldsymbol{\sigma}}^*|\tilde{\boldsymbol{\sigma}}_{MAP})$, as can be noted in the integrand of (5.45). On the other hand, the distribution $p(\mathbf{f}^*|\tilde{\mathbf{l}}_{MAP}, \tilde{\mathbf{l}}^*, \tilde{\boldsymbol{\sigma}}_{MAP}, \tilde{\boldsymbol{\sigma}}^*, \mathbf{y})$ has complete knowledge of the values of the parameters at the test locations. This distribution, as all the distributions in (5.45), is a gaussian, whose mean and covariance are not available because $\boldsymbol{\theta}^*$ is not known. However, we can sample values of $\boldsymbol{\theta}^*$ from the parameter distributions. Thus, for n_s k -dimensional ($|\mathcal{X}^*| = k$) samples of each parameter posterior, $\tilde{\mathbf{l}}_1^*, \tilde{\mathbf{l}}_2^*, \dots, \tilde{\mathbf{l}}_{n_s}^*$, $\tilde{\boldsymbol{\sigma}}_1^*, \tilde{\boldsymbol{\sigma}}_2^*, \dots, \tilde{\boldsymbol{\sigma}}_{n_s}^*$, equation (5.45) could be approximated by

$$p(\mathbf{f}^*|\mathbf{y}, \tilde{\mathbf{l}}_{MAP}, \tilde{\boldsymbol{\sigma}}_{MAP}) \simeq \frac{1}{n_s} \sum p(\mathbf{f}^*|\mathbf{y}, \tilde{\mathbf{l}}_{MAP}, \tilde{\boldsymbol{\sigma}}_{MAP}, \tilde{\mathbf{l}}_j^*, \tilde{\boldsymbol{\sigma}}_j^*) \quad (5.46)$$

where $p(\mathbf{f}^*|\mathbf{y}, \tilde{\mathbf{l}}_{MAP}, \tilde{\boldsymbol{\sigma}}_{MAP}, \tilde{\mathbf{l}}_j^*, \tilde{\boldsymbol{\sigma}}_j^*) \sim \mathcal{N}(\boldsymbol{\mu}_{MAP,j}^*, \boldsymbol{\Sigma}_{MAP,j}^*)$ with mean

$$\boldsymbol{\mu}_{j,MAP}^* = \mathbf{K}(\mathcal{X}^*|\boldsymbol{\theta}^*, \mathcal{X}|\boldsymbol{\theta}_{MAP})(\mathbf{K}(\mathcal{X}|\boldsymbol{\theta}_{MAP}, \mathcal{X}|\boldsymbol{\theta}_{MAP}) + \sigma_{NOI,MAP}^2 \mathbf{I}_{\mathcal{X}})^{-1} \mathbf{y} \quad (5.47)$$

and covariance

$$\boldsymbol{\Sigma}_{j,MAP}^* = \mathbf{K}(\mathcal{X}^*|\boldsymbol{\theta}_j^*, \mathcal{X}^*|\boldsymbol{\theta}_j^*) - \mathbf{K}(\mathcal{X}^*|\boldsymbol{\theta}_j^*, \mathcal{X}|\boldsymbol{\theta}_{MAP}) (\mathbf{K}(\mathcal{X}|\boldsymbol{\theta}_{MAP}, \mathcal{X}|\boldsymbol{\theta}_{MAP}) + \sigma_{NOI,MAP}^2 \mathbf{I}_{\mathcal{X}})^{-1} \mathbf{K}(\mathcal{X}|\boldsymbol{\theta}_{MAP}, \mathcal{X}^*|\boldsymbol{\theta}_j^*) \quad (5.48)$$

In (5.47) and (5.48) the matrix $\mathbf{K}(\mathcal{X}|\boldsymbol{\theta}_{MAP}, \mathcal{X}^*|\boldsymbol{\theta}^*)$ is built by evaluating the kernel function k_{NS-SE} at the training set $\mathcal{X}$ using the MAP estimates of its parameters, and at the test set $\mathcal{X}^*$ using the posterior parameter samples

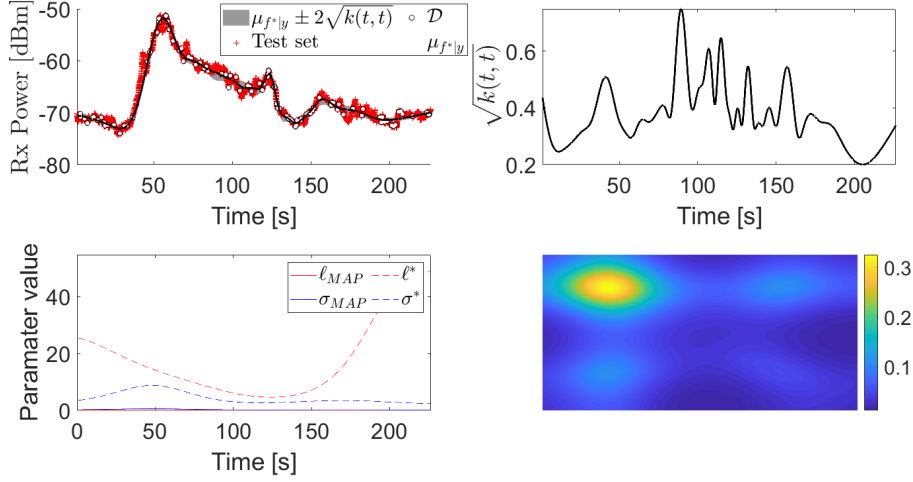


Figure 5.4: Function proportional to the received power for the Saint Barbe route with Mercator BS. The non-stationary kernel defined in (5.37) is used. Top: predictive mean, true received power values at test locations, and training points. Bottom: predictive standard deviation at each distance.

θ_j^* . The choice of the number of samples n_s might depend on the accuracy of the approximation, but also on how many training locations were used, or better say, how well spread over the input space were those training locations $\mathcal{X}$. If the number of training locations is dense enough [Hei+16], there is no need to use n_s posterior samples. Instead of that, only 1 sample is required: the mean value of the parameter posterior distribution $\mu_{*|MAP}$ (5.44).

Predictions using the non-stationary kernel in (5.37) for the smoothed version of the received power signal used in Figure 5.3 are shown in Figure 5.4. It is easy to note that, the predictive standard deviation ($\sqrt{k(x, x)}$) is reduced when the non-stationary kernel is used, meaning the non-stationary model that best fits the training data has a lower variability around the mean than the stationary model. This doesn't necessarily mean that the Mean Squared Error (MSE), or any similar error, between the predictive mean and the true function value is lower for the non-stationary kernel than for the stationary kernel. Since an iterative algorithm is used to find the local or global maxima of the MAP optimization problem (5.21) initial values for the parameters of the function being maximized have to be given. It is not a trivial choice to select this values. Besides, the MAP parameter estimates might change for different initial values. To overcome this problem, multiple initializations need to be done, and the one with the highest $L(\theta)$ is selected. This happens because starting at a different initial value for the parameter, the iterative algorithm might converge fast to a closer local maximum, which is a lower peak of $L(\theta)$ than the peak to which the algorithm converged in a previous initialization.

Since MSE for the stationary and non-stationary kernel are within the same order of magnitude and, moreover, not very different, one might ask: which is the benefit of using a non-stationary kernel if its analytical complexity is higher than its stationary counterpart? The answer is two-fold. First, the spatially-varying parameters provide more information about the correlation between observations $\mathbf{y}$ in focused regions of the space. With a stationary kernel these local correlations are not modelled; thus, the correlation between two observations only depends on the distance (isotropic kernel) or the distance and angle (anisotropic kernel) between them. The covariance matrix $\mathbf{K}_y$ in Figure 5.4 for the non-stationary kernel shows this local correlations. Highest values of correlation are not necessarily at the diagonal, and there exists correlations between distant values, because the length-scale parameter is no longer a constant, allowing to control fast vanishing correlations in a region, while in other region correlations vanish slower.

5.5.2 Kroenecker inference

So far, the non-stationary kernel considered in section 5.5 deal with 1D input space $x \in \mathbb{R}$. Higher-dimensional kernels with certain structure have been proposed [Saa12], [WN15]. If the input space $\mathcal{D}$ is sampled in a regular grid way (or not regular [Fla+15]) a Kroenecker structure could be built for the kernel with higher-dimensional input locations.

Let us assume we have an input set

$$\begin{aligned} \mathcal{X} = (x_{1,1}, \dots, x_{1,p}), (x_{2,1}, \dots, x_{2,p}), \dots, (x_{n,1}, x_{n,p}) \mid \\ x_{i,k} = x_{1,k} + (i-1)\delta_k, \quad i = 1, \dots, n, \quad k = 1, \dots, p \end{aligned} \quad (5.49)$$

where each input location $\mathbf{x} \in \mathbb{R}^p$, the notation $x_{i,k} = [\mathbf{x}_i]_k$ means "dimension k of the input location i ", and δ_k is the distance in dimension k (in this case regular) between input locations. In this case, we are considering a regular grid structure with n points in each k dimension.

Let us denote, as in previous sections, the training set as $\mathcal{D} = \{\mathcal{X}, \mathbf{y}\}$. For such a structure in the input space a kernel with kroenecker structure for the latent function can be built as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \prod_{k=1}^p k([\mathbf{x}_i]_k, [\mathbf{x}_j]_k), \quad i, j = 1, \dots, n^p \quad (5.50)$$

The matrix equivalent of (5.50) for the latent function is

$$\mathbf{K} \in \mathbb{R}^{n^p \times n^p} = \bigotimes_{k=1}^p \mathbf{K}_k \quad (5.51)$$

where $\otimes$ denotes a Kroenecker product, the subscript f in $\mathbf{K}_f$ is dropped for convenience in the notation, and where $\mathbf{K}_k \in \mathbb{R}^{n \times n}$ denotes the non-stationary covariance matrix for dimension k

$$[\mathbf{K}_k]_{i,j} = k([\mathbf{x}_i]_k, [\mathbf{x}_j]_k), \quad i, j = 1, \dots, n \quad (5.52)$$

Since in (5.52) the input locations $[\mathbf{x}_i]_k, [\mathbf{x}_j]_k$ are 1D, the non-stationary kernel used in section 5.5 can be used. However, these equations do not apply only for such kernel, but instead, for any higher-dimensional input space with the regular structure previously stated.

Each spatially-varying parameter θ of the kernel function (5.50) has np values (n at each dimension): $\theta(x_{1,1}), \dots, \theta(x_{n,1}), \dots, \theta(x_{1,p}), \dots, \theta(x_{n,p})$. Let us denote $\theta_{i,k} = \theta(x_{i,k})$. Inference of each of these parameters is done following using (5.41). Since now the matrix $\mathbf{K}$ has a Kroenecker structure, the derivative $\frac{\partial \mathbf{K}}{\partial \theta_{i,k}}$ has also a Kroenecker structure [RHK17b]

$$\frac{\partial \mathbf{K}}{\partial \theta_{i,k}} = \mathbf{K}_1 \otimes \mathbf{K}_2 \otimes \dots \frac{\partial \mathbf{K}_k}{\partial \theta_{i,k}} \otimes \dots \mathbf{K}_p \quad (5.53)$$

Due to the Kroenecker structure of $\mathbf{K}$ the factor $\boldsymbol{\alpha} = (\mathbf{K} + \sigma_{NOI}^2 \mathbf{I})^{-1} \mathbf{y}$ in 5.41 could be also expressed in a structured way

$$(\mathbf{K} + \sigma_{NOI}^2 \mathbf{I})^{-1} \mathbf{y} = \mathbf{Q}(\mathbf{V} + \sigma_{NOI}^2 \mathbf{I})^{-1} \mathbf{Q}^T \mathbf{y} = \left(\bigotimes_{k=1}^p \mathbf{Q}_k \right) (\mathbf{V} + \sigma_{NOI}^2 \mathbf{I})^{-1} \left(\bigotimes_{k=1}^p \mathbf{Q}_k^T \right) \mathbf{y} \quad (5.54)$$

where it is assumed the matrix $\mathbf{K}$ has the following eigen-decomposition

$$\mathbf{K} = \mathbf{Q} \mathbf{V} \mathbf{Q}^T, \quad \mathbf{Q} = \bigotimes_{k=1}^p \mathbf{Q}_k, \quad \mathbf{V} = \bigotimes_{k=1}^p \mathbf{V}_k \quad (5.55)$$

with $\mathbf{V}$ a diagonal matrix containing the eigenvalues, and $\mathbf{Q}$ a matrix whose columns correspond to the eigen-vectors of $\mathbf{K}$. Equivalently, in (5.55) each covariance of dimension k the following eigendecomposition

$$\mathbf{K}_k = \mathbf{Q}_k \mathbf{V}_k \mathbf{Q}_k^T \quad (5.56)$$

This structure makes easier de computation of $\boldsymbol{\alpha}$ since it is required now to invert a diagonal matrix.

5.6 Shadowing correlation

Let us denote by $\mathbf{b} = [b_1, \dots, b_n]$ the vector comprising the shadowing at time instants $t_1, \dots, t_n$. For each of these time instants, the mobile station is located

at positions $\mathbf{x}_1, \dots, \mathbf{x}_n$, while the base station is static. We call *link*, the abstract entity comprising the mobile station location, the medium through which radio signals propagate and the base station location. We are interested in the shadowing correlation $\rho_b(\cdot, \cdot)$ between two different links. If the base station location $\mathbf{r}$ is also moving, this correlation is in general a function of 4 locations $\rho_b(\mathbf{x}_i, \mathbf{r}_i, \mathbf{x}_j, \mathbf{r}_j)$. With this notation we are emphasizing that the correlation is among links.

Since the base station is static for the measurements used throughout this document, a link is fully characterized by the mobile station location. Thus, correlation between links could be denoted by $\rho_b(\mathbf{x}_i, \mathbf{x}_j)$ where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the mobile station locations. A common model for such correlation is the isotropic Gudmunson model $\rho_b = \sigma^2 \exp\left(-\frac{h^2}{l}\right)$, which considers that correlation depends on the distance h between locations rather than on specific locations. Such model is tested for the Saint Barbe route at both Maxwell and Mercator BS. Isotropic computed shadowing and its Gudmunson model fit are shown in Figure 5.5 and Table 5.2. For both BS, the Gudmunson performs well up to a distance d_{DCORR} , called *de-correlation* distance. For the decorrelation distance of the Maxwell BS (around 12 m) the Gudmunson model matches almost exactly the experimental shadowing correlation. Beyond this distance, isotropic correlation models are not able to capture long-distance correlations, possibly due to LoS conditions. Some of the peaks at distances beyond the decorrelation distance, in the experimental correlation graph, are also a consequence of NLoS conditions: that is to say, the average received signal value is low and remains constant.

Table 5.2: Parameters isotropic shadowing correlation model.

BS	σ	l	d_{DCORR} [m]
MXW	0.99	37.53	~ 12
MKT	0.97	216	~ 20

5.7 Underlying Spatial Field

Shadowing correlation might behave in a non-stationary way. In other words, ρ_b depends on mobile station locations. Since correlation is related to covariance, non-stationary covariance could be modelled through the use of kernel functions with spatially-varying kernel parameters. If such a situation is possible, instead of the isotropic Gudmunson model (aka the $k_{SE}(\cdot, \cdot)$ kernel), the non-stationary covariance produced by the kernel $k_{NS-SE}(\cdot, \cdot)$ will be inferred, with its specific parameter values. However, the set of mobile station locations $\mathcal{X}$ is highly correlated and do not distribute randomly over the whole

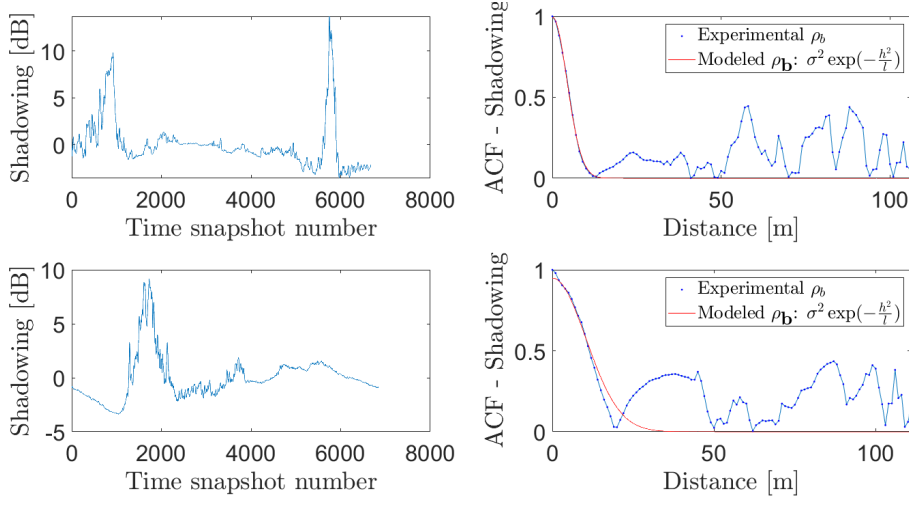


Figure 5.5: Experimental and modelled shadowing correlation ρ_b for the Saint Barbe route with Maxwell and Mercator BS. Top left: shadowing vs time snapshot at Saint Barbe route with Maxwell BS. Top right: isotropic shadowing correlation at Saint Barbe with Maxwell BS. Bottom left: shadowing vs time snapshot at Saint Barbe route with Mercator BS. Bottom right: isotropic shadowing correlation at Saint Barbe with Mercator BS.

space: the mobile station moves around a building following a 'circular' trajectory and measurements were not made inside buildings. Thus, inference of the parameters of the kernel will fail to produce reliable results, since the space-dependency of such parameters could not be infer, due to the lack of observations accounting for such spatial-dependency.

For that reason, it is required to follow a different approach. Rather than using the GP model in (5.9) for the shadowing, we assume the shadowing is caused by a linear transformation of an underlying spatial field. What we call an underlying spatial field, denoted by f , is a conventional scalar field: $f: \mathbb{R}^p \rightarrow \mathbb{R}$, for which we know its value at some locations $\mathbf{x}$. For the case of the measurements under analysis $p = 2$. To see which is the relevance of such approach, let us denote by $\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_{n^p})]^T$. A linear transformation is represented by a matrix $\mathbf{A} \in \mathbb{R}^{n \times q}$ with q a parameter to be defined. Thus, as previously said, we assume the shadowing $\mathbf{b}$ could be written as

$$\mathbf{b} = \mathbf{A}\mathbf{f} \quad (5.57)$$

We haven't provided yet any clue about where do we get the underlying field $\mathbf{f}$. However, for the moment, let us say that we can define $\mathbf{f}$ over a representative set of points $\mathcal{X}$ of the space $\mathcal{A}$. We can choose $\mathcal{X}$ to be formed by a regular grid $x_{i,k} = x_{1,k} + (i-1)\delta_k$, $i = 1, \dots, n$, $k = 1, \dots, p$ as it was used in

section 5.5.2. Now, since $\mathbf{f}$ is defined over a representative set of locations of the space $\mathcal{A}$, we can perform GP regression on it, by assuming a GP model for it

$$\mathbf{f} = \mathbf{f}^{GP} + \boldsymbol{\epsilon} \quad (5.58)$$

where $\mathbf{f}^{GP}$ is the latent gaussian process, and $\boldsymbol{\epsilon} = [\epsilon_1, \dots, \epsilon_{np}]$ with ϵ_i as in (5.9). In order to this approach be valid, the matrix $\mathbf{A}$ must be known. If that is the case, assuming the matrix $\mathbf{A}$ is deterministic, the shadowing vector $\mathbf{b}$ will be a multivariate gaussian random vector. It could be proved that the linear transformation $\mathbf{b} = \mathbf{A}\mathbf{f}$ of a multivariate gaussian distributed vector $\mathbf{f}$ is as well a gaussian with mean and covariance given by

$$\mathbf{b} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\mu}_f, \mathbf{A}\mathbf{K}_f\mathbf{A}^T) \quad (5.59)$$

where here $\mathbf{K}_f$ is the covariance matrix for the underlying field $\mathbf{f}^{GP}$ in (5.58). In other words, the covariance matrix for the shadowing $\mathbf{K}_b = \mathbf{A}\mathbf{K}_f\mathbf{A}^T$ can be computed from the known deterministic matrix $\mathbf{A}$ and from the inferred covariance matrix of the underlying field $\mathbf{K}_f$. Note that if $\mathbf{K}_f$ is non-stationary, in general, the covariance $\mathbf{K}_b$ is also. Furthermore, recall that there was no way to empirically compute the shadowing correlation as a function of locations $\rho_b(\mathbf{x}_i, \mathbf{x}_j)$. This is because we only have one sample of the shadowing at each location $\mathbf{x}$, and the empirical correlation $\rho_b(\mathbf{x}_i, \mathbf{x}_j)$ requires a set of L samples of the shadowing $\mathbf{b}(\mathbf{x}) = [b_1(\mathbf{x}), \dots, b_L(\mathbf{x})]$ per each location $\mathbf{x}$

$$\rho_b(\mathbf{x}_i, \mathbf{x}_j) = \rho(\mathbf{b}(\mathbf{x}_i), \mathbf{b}(\mathbf{x}_j))$$

(5.60)

$$= \frac{\sum_{l=1}^L \left(b_l(\mathbf{x}_i) - \frac{1}{L} \sum_{u=1}^L b_u(\mathbf{x}_i) \right) \left(b_l(\mathbf{x}_j) - \frac{1}{L} \sum_{u=1}^L b_u(\mathbf{x}_j) \right)}{\sqrt{\sum_{l=1}^L \left(b_l(\mathbf{x}_i) - \frac{1}{L} \sum_{u=1}^L b_u(\mathbf{x}_i) \right)^2} \sqrt{\sum_{l=1}^L \left(b_l(\mathbf{x}_j) - \frac{1}{L} \sum_{u=1}^L b_u(\mathbf{x}_j) \right)^2}}$$

However, note that by using 5.59, the shadowing covariance, and so, its correlation too, is computed in an indirect way. To summarize: this approach relies in 3 assumptions: there exists a matrix $\mathbf{A}$ for which the transformation (5.57) is valid; such a matrix is deterministic; and finally, it is possible to compute the underlying field vector $\mathbf{f}$ for which the GP regression model is applied.

5.7.1 How to find $\mathbf{A}$?

In what follows we assume the matrix $\mathbf{A}$ is deterministic. If that is not the case, equation (5.59) do not hold.

It is possible to build matrix $\mathbf{A}$, by first looking at a given entry of $\mathbf{b}$. From (5.57), the t -th entry of $\mathbf{b}$, $[\mathbf{b}]_t$ is expressed as

$$[\mathbf{b}]_t b_t = \sum_m [\mathbf{A}]_{t,m} [\mathbf{f}]_m \quad (5.61)$$

Assume that in the regular grid in (5.49) the step δ_k in each dimension is infinitesimal, so that the observed $\mathbf{f}$ is actually the continuous scalar underlying field f . In such case, the sum in (5.61) becomes the integral

$$b_t = \int_{\mathcal{A}} w_t(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \quad (5.62)$$

where $w_t(\mathbf{x})$ is the continuous function equivalent to the weighting done by the row t of the matrix $\mathbf{A}$. The idea of expressing the shadowing as the linear system in (5.57) and its extension to the continuous domain in (5.62), was originally presented in [AP09]. There, an underlying spatial loss field (SLF) is used to model the attenuation caused by obstacles. Shadowing is considered an isotropical gaussian random field whose values are produced from the SLF

$$b(\mathbf{x}, \mathbf{x}') \doteq b_t \in \mathbb{R} = \frac{1}{\sqrt{\|\mathbf{x} - \mathbf{x}'\|_2}} \int_{\mathbf{x}}^{\mathbf{x}'} f(\hat{\mathbf{x}}) d\hat{\mathbf{x}} \quad (5.63)$$

where as denoted previously, b is the shadowing, t the time snapshot corresponding to the locations $(\mathbf{x}, \mathbf{x}')$, and $\hat{\mathbf{x}}$ denotes a point in $\mathcal{A}$, or to put it in other words, a grid point in the continuous grid. The square root factor is used so the variance of the shadowing in link $(\mathbf{x}, \hat{\mathbf{x}})$ is approximately constant.

A generalization of (5.63) was proposed in [Ham+13] and [RLG18], and shown in (5.64)

$$b_t = \int_{\mathcal{A}} w(\mathbf{x}, \mathbf{x}', \hat{\mathbf{x}}) f(\hat{\mathbf{x}}) d\hat{\mathbf{x}} \quad (5.64)$$

The only difference between equations (5.64) and (5.61) is the notation. The weighting function w plays the same role as the one played by $w_t(\mathbf{x})$. In the notation used in (5.64), $w(\mathbf{x}, \mathbf{x}', \hat{\mathbf{x}})$ depends on the mobile station and base station locations, and can be seen as a function that weights each point $\hat{\mathbf{x}}$ in the space $\mathcal{A}$ depending on the location of both the mobile station and the base station. For example, the model in (5.63) has the following correspondent weighting function

$$w(\mathbf{x}, \mathbf{x}', \hat{\mathbf{x}}) = \frac{1}{\sqrt{\|\mathbf{x} - \mathbf{x}'\|_2}} \int_{\mathbf{x}}^{\mathbf{x}'} \delta(\mathbf{x}, \mathbf{x}') d\hat{\mathbf{x}} \quad (5.65)$$

Thus, as said before, w can be seen as a 'selection' function, choosing which infinitesimal pixels are active (have non-zero weight). If the continuous space $\mathcal{A}$ is replaced by a discrete regular grid using (5.49), equation (5.57) can be seen as the discrete counterpart of (5.64), applied to each shadowing value b_t

$$\begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_T \end{pmatrix} = \begin{pmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,N_x N_y} \\ w_{2,1} & w_{2,2} & \dots & w_{2,N_x N_y} \\ \dots & \dots & \dots & \dots \\ w_{T,1} & w_{T,2} & \dots & w_{T,N_x N_y} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \dots \\ f_{N_x N_y} \end{pmatrix} \rightarrow \mathbf{b} = \mathbf{A}\mathbf{f} \quad (5.66)$$

Here, the set of transceiver pairs is $\{(\mathbf{x}, \mathbf{x}')_i\}_{i=1}^T$, N_x is the number of points of the grid in the first dimension and N_y is the number of points of the grid in the second dimension. The underlying field values at each grid point $\mathbf{x}$ are stacked in the column vector $\mathbf{f}$. Its entries are $f_i = f(\hat{\mathbf{x}}_i)$, with i being a linear index mapping the indexes in each dimension of the coordinate $\mathbf{x}$ to the natural numbers $1, 2, 3, \dots, N_x N_y$. To simplify notation, we define $w_{ij} = w((\mathbf{x}, \mathbf{x}')_i, \hat{\mathbf{x}}_j)$ with $i = 1, \dots, T$ and $j = 1, \dots, N_x N_y$. We call matrix $\mathbf{A}$ the Active Paths Matrix (APM). From now and on, we assume the same number of grid points in each dimension: $N_x = N_y = n$. For that case: $\mathbf{f} \in \mathbb{R}^{n^2}$.

The question of how to define the weights w still stands up. First thing to be aware of, is that for each observation b_t there is an associated $w_t(\mathbf{x})$. In other words, to better understand the meaning of w , it is useful to think about it as a function defined for each sample t or pair of transmitter-receiver positions at t . The turning point in order to determine w is to think the shadowing value b_t as the contribution of the underlying field over the link between the mobile station and the base station. That is to say: for a given time snapshot t the weighting function assigns a non-zero value r to the grid points that are in the LoS between the mobile station and the base station, or close to it, depending upon the model used for w .

To define the weighting value r is not a trivial task. Different models have been proposed in the literature. Figure 5.6 shows 3 models for the weights $w_t(\mathbf{x})$. On the left graph, the model used in Equation 5.65 is shown. This model is called the NeSh model [PA08]. It gives a non-zero weight only to the grid points in the LoS between the mobile station and the base station. Oppose to the LoS assumption in the NeSh, the other 2 graphs in Figure 5.6, the Normalized Ellipse Model (NEM) and the Inverse Area Ellipse Model (IAEM), assume that there is a beam of rays departing and a beam of rays arriving,

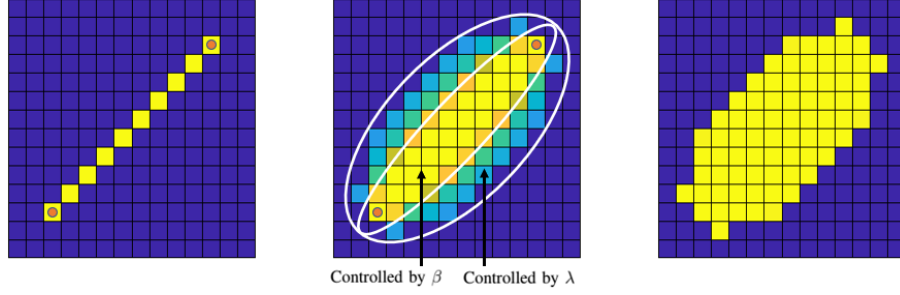


Figure 5.6: Different weighting functions. The plots correspond to the weights for the case of 20×20 pixels. The dark blue represents zero valued entries. The red circles represent the transceiver locations. Grid points are located at the center of the dark blue squares. Left: NeSh model. Center: IAEM. Right: NEM.

such that propagation mechanisms occur within ellipses enclosing the mobile station and base station. In other words, NEM [Ham+13] illustrated in the right of Figure 5.6, assigns the same non-zero value to all the grid points within an ellipse whose focii are the transceivers. The semi-minor axis of the ellipse is determined by the wavelength, which guarantees that the ellipsoid can be thought as the first Fresnel Zone, thus, giving a physical meaning to the weighting model. The last weighting model is displayed at the center of the figure: IAEM model uses the same ellipse explained previously, but allows grid points inside the ellipse to have different values depending upon their distance with respect to the LoS: the farther away from the link, the less value they have.

The value both models (NEM and IAEM) assign to a grid point $\hat{\mathbf{x}}$ is determined by equation 5.67 [RLG18]. For a given value of t

$$w_t(\zeta_1, \zeta_2) = \begin{cases} 0 & \zeta_2 < \zeta_1 + \frac{\lambda}{2} \\ \min \left\{ \Omega(\zeta_1, \zeta_2), \Omega(\zeta_1, \sqrt{\zeta_1^2 + \beta^2}) \right\} & \zeta_2 \geq \zeta_1 + \frac{\lambda}{2} \end{cases} \quad (5.67)$$

with $\Omega(\zeta_1, \zeta_2) = 4/(\pi\zeta_2\sqrt{\zeta_2^2 - \zeta_1^2})$, $\zeta_1 \doteq \zeta_1(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|_2$ and $\zeta_2 \doteq \zeta_2(\mathbf{x}, \mathbf{x}', \hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|_2 + \|\mathbf{x}' - \hat{\mathbf{x}}\|_2$. For $\beta = 0$, we have the NeSh model. For $\beta = \sqrt{\zeta_1\lambda}$, we have the NEM. Values of β between 0 and $\sqrt{\zeta_1\lambda}$, define two regions: an outer ellipse controlled by λ and an inner one defined by β . Weights in the inner ellipse have the same value, outside the inner but inside the outer behave like the IAEM, and outside the bigger ellipse have zero value (see Figure 5.6).

Equation 5.67 defines both the IAEM and NEM models through parameter

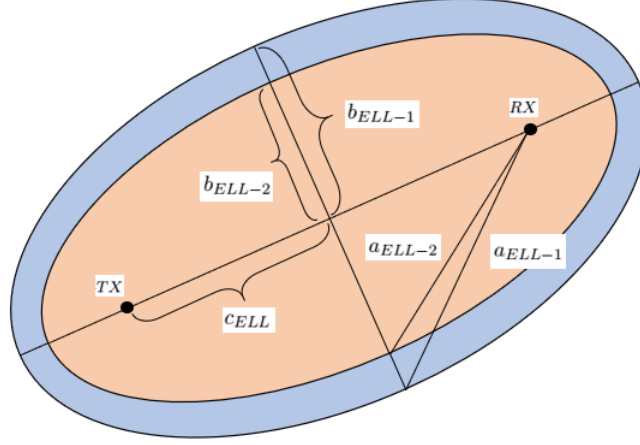


Figure 5.7: Ellipses defined by λ and β . Pixels in the red zone conform to the NEM. Pixels in the blue zone conform to the IAEM.

β . Figure 5.7 helps to understand it better. The ellipse of reference is set-up by the transceivers as focii, and a semi-minor axis $b_{ELL-1}(\lambda)$ that is controlled by λ . With these 2 things provided, the ellipse is defined. Parameter β can be seen as a parameter that controls the semi-minor axis b_{ELL-2} of another concentric ellipse whose focii are also the transceivers. There is a value β_{LIM} for which $b_{ELL-1} = b_{ELL-2}$, that is to say, the 2 ellipses coincide. For values $\beta > \beta_{LIM}$ all the pixels within the ellipse given by b_{ELL-1} have the same value and outside a value of zero. In this case, we are in the NEM model. For values $\beta < \beta_{LIM}$, we have the case shown in Figure 5.7. All pixels within the inner ellipse (controlled by β) have the same value. Pixels between the two ellipses will have a value defined by $\Omega(\zeta_1, \zeta_2) = \frac{4}{\pi \zeta_2 \sqrt{\zeta_2^2 - \zeta_1^2}}$.

We can find b_{ELL-1} by varying λ , and numerically setting β in such a way that we encounter the value β_{LIM} for which the 2 ellipses coincide (and so $b_{ELL-1} = \beta_{LIM}$). Figure 5.8 shows the numerical dependence between λ and β_{LIM} . It is clear that as the grid becomes continuous ($n \rightarrow \infty$) there is a strong dependence. If, for the maximum value of n , we fit a power model $\beta_{LIM}(\lambda) = p_1 \lambda^{p_2}$ with parameters p_1 and p_2 we have:

$$\beta_{LIM} = \sqrt{2c_{ELL}\lambda} = \sqrt{\zeta_1}\lambda/2 \quad (5.68)$$

where the $1/2$ comes from the fact that we are taking the distance to the center and not to the other extreme of the semi-minor axis.

Other approaches, as the one in [RLG18], learn the weights (and the underlying field) from the measurements $\mathbf{b}$, by making use, for example, of

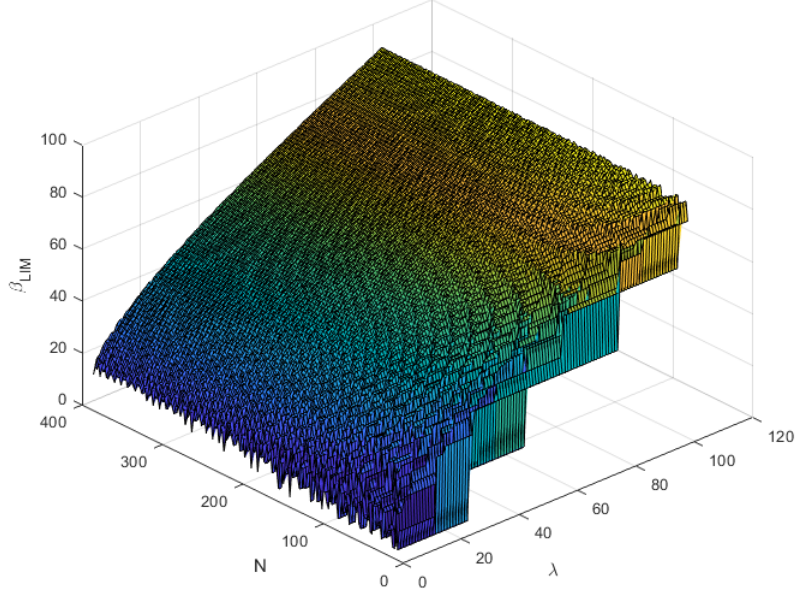


Figure 5.8: Parameter β_{LIM} as a function of λ and n .

kernel-based learning methodologies [RLG18]. Although such weighting learning methods have an undeniable advantage over the weighting models used in (5.67) due to the lack of heuristic parameters to set up (as β), it is not easy to physically interpret the learned weights. Models like those in (5.67), have a physical meaning since they can be seen straightly as selecting the physical locations through which the underlying field is going to be added (5.64). If the underlying field is an attenuation field, the role played by the weighting function in (5.64) is to make the total attenuation for a given time t the result of the sum of the attenuations over the path selected by that weighting function.

The three models presented, consider propagation paths as in LoS or Obstructed LoS (OLOs) conditions. Only grid points close to the LoS will have an associated non-zero weight value. Furthermore, for GHz frequencies the wavelength is in the order of centimeters while the distance between the focii of the ellipse (the transceivers) is in the order of tens of meters, which causes that the semi-minor axis of the ellipse will be very narrow. This means that the ellipses for both NEM and IAEM models are so narrow, that instead of selecting the grid points in an ellipse, the weighting function will select only the grid points in the LoS, as is done for the NeSh model.

In order to overcome such problem and in order to extend the weighting function for more than one NLoS path, we propose another model for the

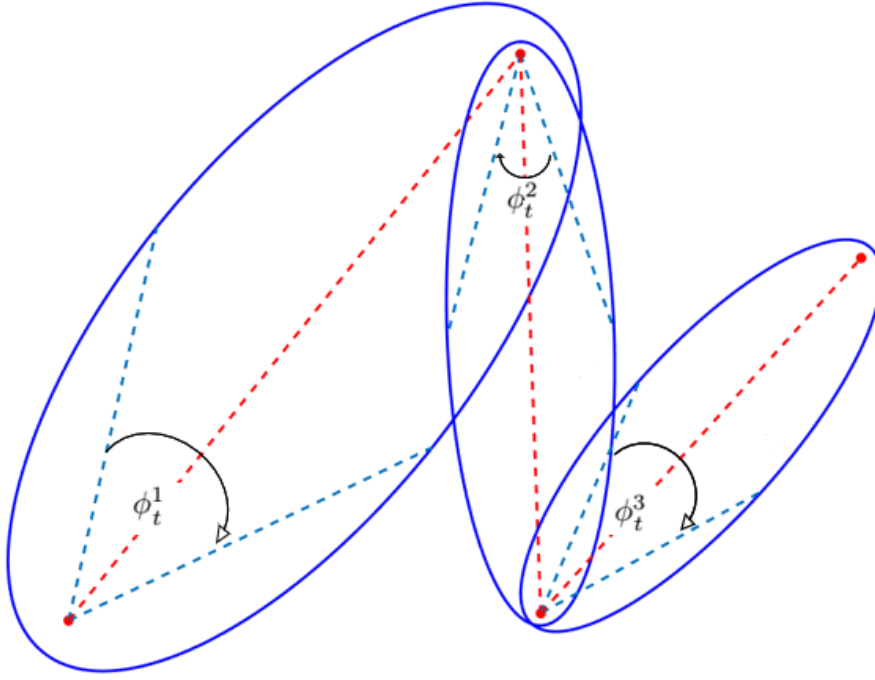


Figure 5.9: Proposed weighting model. The red points are the transceivers, first and last bounces of the propagation path. The latter is illustrated by the red dashed line. The blue dashed lines represent the 'beamwidth' of the launched and reflected rays, and control with the semi-minor axis of each ellipse.

weighting function. We make use of the propagation paths found through procedures in Chapters 2 and 4, to produce weights by including multiple paths. The semi-minor axis of the ellipses, previously linked to the wavelength, have a different interpretation in our model. Figure 5.9 shows the idea. For a link between transceiver and base station at snapshot t , three ellipses are used. The first one, with focii at mobile station and first bounce. The second one, with focii at first bounce and last bounce; and the third one, with focii at last bounce and base station.

Clusters found in Chapter 4 are used to reduce the number of observations, by considering centroids of clusters of propagation paths, rather than each individual propagation path. The focii of the ellipse that are bounces, correspond to centroids of the cluster bounces. For the first ellipse, the angle between the focii (the direction of the propagation path), with respect to horizontal axis is the mean angle of departure of the cluster. For the third ellipse, the angle between the focii (with respect to horizontal axis) is the mean angle of arrival of the cluster.

Angle ϕ_t^1 in Figure 5.9, correspond to the angular spread of the cluster (in rads) in the departure side. Angle ϕ_t^3 in the figure, correspond to the angular spread (in rads) of the cluster in the arrival side. Angle ϕ_t^2 is user-defined and controls how much the rays spread after the first bounce. For C_t number of clusters in snapshot t equation 5.64 is rewritten as

$$b_t \simeq \sum_{c=1}^{C_t} \sum_{\hat{\mathbf{x}}_i \in \mathcal{A}} w(\mathbf{x}_c, \mathbf{x}'_c, \hat{\mathbf{x}}_i) f(\hat{\mathbf{x}}_i) \quad (5.69)$$

With this in mind, for a time snapshot t the values $w_{t,i}$, $i = 1, \dots, n$ in (5.66) are determined as follows:

- Consider the propagation paths $\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_{L_t}$ that were extracted using the procedures explained in Chapters 2 and 4, with L_t being the number of propagation paths in time snapshot t .
- Build a regular grid by using 5.49.
- Take each grid point $\mathbf{x}_i$ and check if any of the propagation paths $\mathcal{L}_1, \dots, \mathcal{L}_{L_t}$ touches it. If that is the case assign a value r to $w_t(\mathbf{x}_i)$. Do this step for each and all of the propagation paths. If the grid point $\mathbf{x}_i$ is touched by more than one propagation path, add the value r the number of times the grid point is touched. If no propagation path touches the grid point $\mathbf{x}_i$, the value of $w_t(\mathbf{x}_i)$ is zero.
- Repeat the previous step for each and all the grid points $\mathbf{x}_1, \dots, \mathbf{x}_{n^2}$.

Figure 5.10 show the average weights assigned to each grid point (with $n = 100$) for the propagation regions defined in Chapter 2. The average weights are simply computed by

$$w_{AVG}(\mathbf{x}) = \sum_{t \in \mathcal{P}} w_t(\mathbf{x}) \quad (5.70)$$

where $\mathcal{P}$ is the propagation region and $\mathbf{x} \in \mathcal{A}$ are the grid points. This average $w_{AVG}(\mathbf{x})$ shows a heat map of where are occurring more bounces in the scenario $\mathcal{A}$. Thus, it serves as a summary tool to identify locations with more reflections for a given propagation region. However, it could be seen in Figure 5.10, that it seems as if solely few points are identified as the main reflection points and routes of propagation. That is not the case. What happens in Figure 5.10 is that this few points are more two order of magnitudes (100 times) more probable of occurring than the rest of the actual bounces. In order to cope for that, this highly probable bounces over each propagation region are removed in Figure 5.11.

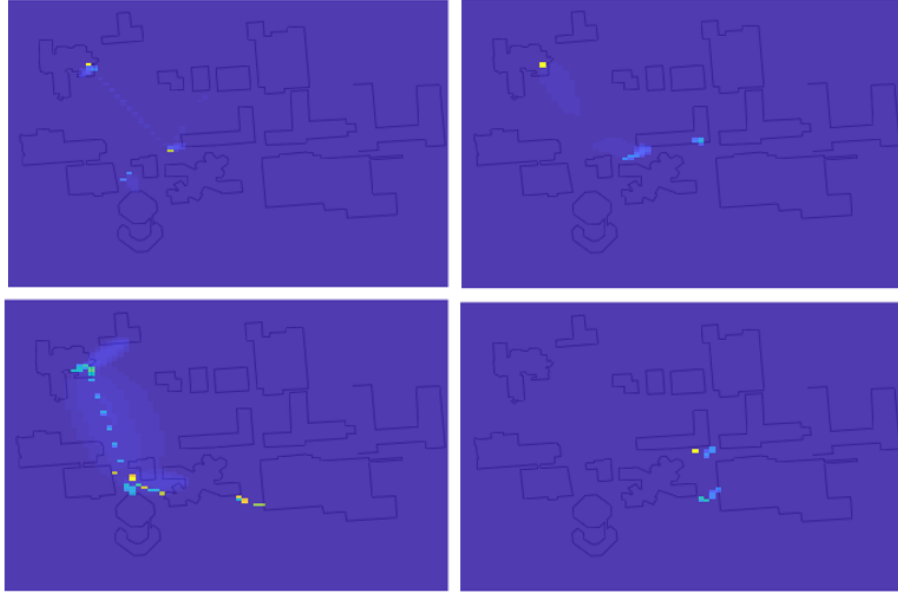


Figure 5.10: Averaged APM for each of the propagation regions defined in Figure 2.14 for the Saint Barbe route with Mercator BS. Grid points with weights equal to zero are painted in darker blue. Lighter colors towards yellow represent a higher weight assigned to the grid point. The value of the weight is not as relevant as the relative value between different weights, so for that reason no color scale is shown. The regular grid uses $n = 100$. Top left: OLoS region. Top right: NLoS 1 region. Bottom left: NLoS 2 region. Bottom right: NLoS 3 region.

From Figure 5.11 it is clear for the OLoS region that the LoS mechanism is prevailing. Outside the LoS, propagation paths cover equally zones close the the LoS between the mobile station and the base station (keep in mind that for a propagation region the mobile station have different locations, so the LoS direction is changing for such propagation region). This is the case, as well, for the NLoS 1 region. For NLoS 2 and NLoS 3 regions of higher transit of radio signals are identified in yellow.

It should be said however, that the value of the weights is not related with the power of the propagation paths, which means that by including such power these heat maps (or more directly, the weights) would change, since some zones that have high transit of radio signals in Figure 5.11 could be less probable to occur in terms of power, due to, for example, higher distance from the bounce with respect to the base station.

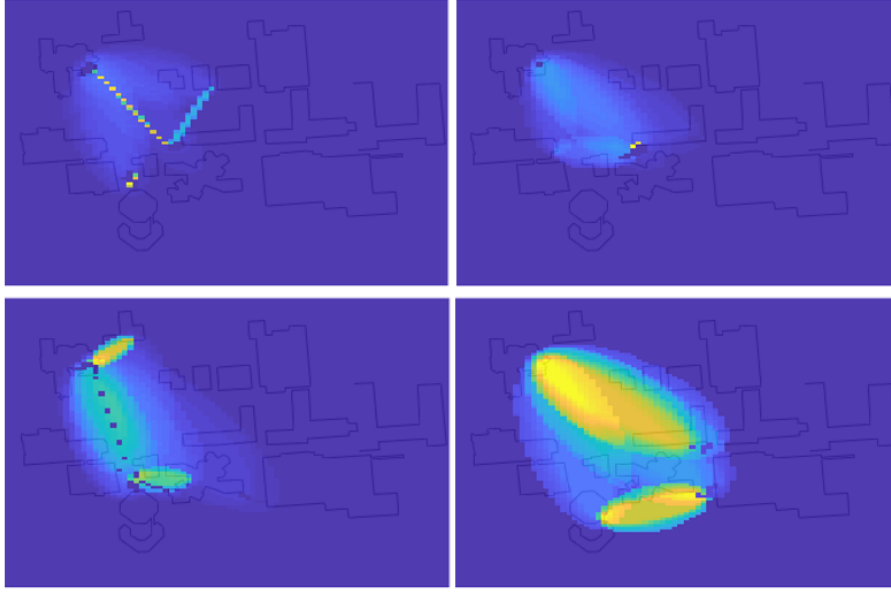


Figure 5.11: Averaged APM for each of the propagation regions defined in Figure 2.14 for the Saint Barbe route with Mercator BS. Weights with more than 100 times the value of the minimum non-zero weight are assigned a zero value, to illustrate more easily the values of all the other weights. Grid points with weights equal to zero are painted in darker blue. Lighter colors towards yellow represent a higher weight assigned to the grid point. The value of the weight is not as relevant as the relative value between different weights, so for that reason no color scale is shown. The regular grid uses $n = 100$. Top left: OLoS region. Top right: NLoS 1 region. Bottom left: NLoS 2 region. Bottom right: NLoS 3 region.

5.7.2 Finding the underlying field

Once the APM is computed, the underlying field $\mathbf{f}$ is found by solving the linear system in (5.66). As the APM is in general non-square, the linear system could have no solution or infinite solutions, depending upon the size and rank of $\mathbf{A}$. If $\mathbf{A}$ is full rank, the solution to (5.66) is already available from $\mathbf{f} = \mathbf{A}^\dagger \mathbf{b}$, where $\mathbf{A}^\dagger$ is either the usual inverse (for the case where APM is square) or the Moore-Penrose pseudo-inverse.

In general, $\mathbf{A}$ is rank deficient. Moreover, due to the fact that there are some regions of $\mathcal{A}$ with dense propagation (meaning a high number of propagation paths are passing through the zone), and others with no propagation or lower propagation paths concentration, the APM will be sparse and very ill-conditioned. In other words, this means that there is not enough information about the underlying field over $\mathcal{A}$. This is expected, since as said before, noticeable propagation doesn't occur everywhere in $\mathcal{A}$. One possibility is to

reduce the size of the space $\mathcal{A}$. As a trade-off, by reducing the size of $\mathcal{A}$, less grid points are considered, and thus, we are losing information about less probable propagation paths passing through those grid points no longer considered.

For solving the linear system for $\mathbf{f}$, it is needed to impose restrictions on the solution, in order to reduce the infinite space of solutions to one with some desirable properties. This conditioning on the characteristics of the field $\mathbf{f}$ is accounted by the term $R(\mathbf{f})$ in the minimization problem (5.71)

$$\mathbf{f}^* = \arg \min_{\mathbf{f}} \|\mathbf{A}\mathbf{f} - \mathbf{b}\|_2^2 + \lambda_R R(\mathbf{f}) \quad (5.71)$$

The minimization problem has 2 terms: a minimum least squares problem looking to minimize the error, and a regularization term given by the function $R(\mathbf{s})$, looking to restrict the space of infinite solutions. The constant λ_R decides how much the solution follows the condition imposed by the regularization term.

5.7.3 Regularization

The regularization term $R(\mathbf{f})$ in (5.71) accounts for desired properties of the solution $\mathbf{f}^*$. Different regularization terms can be considered. In [LKG17], a sparse and low rank approach is used. The solution vector $\mathbf{f}^*$ is decomposed into two terms, namely $\mathbf{L}$ and $\mathbf{E}$: the former, a low rank matrix and the latter, a sparse matrix. Low-rank condition is achieved by using the nuclear norm $\|\cdot\|_*$, and sparsity by using l_1 norm $\|\cdot\|_1$.

$$R(\mathbf{f}) = \lambda_1 \|\mathbf{L}\|_* + \lambda_2 \|\mathbf{E}\|_1 \quad (5.72)$$

Such a decomposition is physically justified by thinking about the low-rankness as a consequence of regular placement of building and walls, as in the indoor and urban cases; on the other hand, the sparsity term account for the presence of outliers that don't fit into the low-rank model.

Tikhonov regularization (also known as ridge regression) in [WPV09] and [RLG18], forces the solution to follow the properties in (5.73) given by equivalent matrices $\mathbf{Q}$ or $\mathbf{G}$, respectively.

$$R_{TIK}(\mathbf{s}) = \|\mathbf{Q}\mathbf{s}\|_2^2 = (\mathbf{Q}\mathbf{s})^T \mathbf{Q}\mathbf{s} = \mathbf{s}^T \mathbf{Q}^T \mathbf{Q}\mathbf{s} = \mathbf{s}^T \mathbf{G}\mathbf{s} \quad (5.73)$$

The solution to (5.71) using (5.73) as the regularizer term is found by setting the derivative of $\mathbf{s}$ to zero and solve for $\mathbf{s}$:

$$\begin{aligned}\mathbf{s}_{TIK} &= (\mathbf{A}^T \mathbf{A} + \lambda_R \mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{A}^T \mathbf{b} \\ &= (\mathbf{A}^T \mathbf{A} + \lambda_R \mathbf{G})^{-1} \mathbf{A}^T \mathbf{b}\end{aligned}\quad (5.74)$$

Common choices for the matrix $\mathbf{Q}$ start with the minimum norm regularizer $R_{TIK} = \|\mathbf{s}\|_2^2$ (by choosing $\mathbf{Q} = \mathbf{I}_{N_x \times N_y}$), which promotes smoothness by forcing the energy to spread over more entries of $\mathbf{s}$ [RLG18]. One can generate $\mathbf{Q}$ assuming that $\mathbf{f}$ is an isotropical gaussian random field [AP09]. In such a case, the covariance matrix $\mathbf{G}^{-1}$ entries correspond to the kernel function evaluated at locations $\mathbf{x}_i$ and $\mathbf{x}_j$. For example, for the case of the gaussian kernel in Table 5.1

$$[\mathbf{C}]_{ij} = [\mathbf{G}^{-1}]_{ij} = \sigma_f \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^p}{l_f}\right) \quad (5.75)$$

with $p = \{1, 2\}$, σ_f is regarded as the scale, and l_f as the length-scale. The minimization problem (5.71), has a bayesian interpretation [WP10]. If one considers the prior $\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ and the data likelihood $\mathbf{b}|\mathbf{f} \sim \mathcal{N}(\mathbf{A}\mathbf{f}, \mathbf{I}_{n^2})$, the $\mathbf{f}^*$ is the MAP estimate.

Another regularization choice is the Truncated Singular Value Decomposition (TSVD). It takes the biggest $K < \min(T, N_x N_y)$ eigenvalues from the singular value decomposition of matrix $\mathbf{A}$. Given the Singular Value Decomposition (SVD) $\mathbf{A} \in \mathbb{R}^{T \times N_x N_y} = \mathbf{U} \mathbf{D} \mathbf{V}^T$, with

$$\begin{aligned}\mathbf{U} &\in \mathbb{R}^{T \times T} = [\mathbf{u}_1, \dots, \mathbf{u}_K, \dots, \mathbf{u}_T] \\ \mathbf{V} &\in \mathbb{R}^{N_x N_y \times N_x N_y} = [\mathbf{v}_1, \dots, \mathbf{v}_K, \dots, \mathbf{v}_{N_x N_y}] \\ \mathbf{D} &\in \mathbb{R}^{T \times N_x N_y} = \text{diag}(\sigma_1, \dots, \sigma_K, \dots, \sigma_{\min(T, N_x N_y)})\end{aligned}\quad (5.76)$$

where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_K \geq \dots \geq \sigma_{\min(T, N_x N_y)}$ and $\text{diag}(a, b, c)$ stands for a matrix with non zero entries a, b, c in its principal diagonal. The solution is built from (5.76) as

$$\mathbf{s}_{TSVD} = \sum_{i=1}^K \frac{1}{\sigma_i} \mathbf{u}_i^T \mathbf{b} \mathbf{v}_i \quad (5.77)$$

Figure 5.12 shows the estimated underlying field by using ridge regression, while Figure 5.13 shows the reconstructed shadowing signal. An isotropic gaussian kernel is used to produce the covariance matrix $\mathbf{G}^{-1}$ in the regularization term (5.73). The regularization factor λ_R is set to 0.1. The choice of the

parameters for the gaussian kernel and the regularization is based on heuristics. The accuracy of the reconstructed signal $\mathbf{A}\mathbf{f}^*$, or better said, the Mean Absolute Error (MAE) $\frac{1}{T} = \sum_i^T [\mathbf{b}]_i - [\mathbf{A}\mathbf{f}^*]_i$ shows little dependence on any of the gaussian kernel parameters, neither on the regularization factor. The MAE is only reduced by increasing the number of grid points. The regularization factor controls the smoothness of the field $\mathbf{f}^*$. The bigger the value, the bigger the smoothness for the case of the isotropic gaussian kernel.

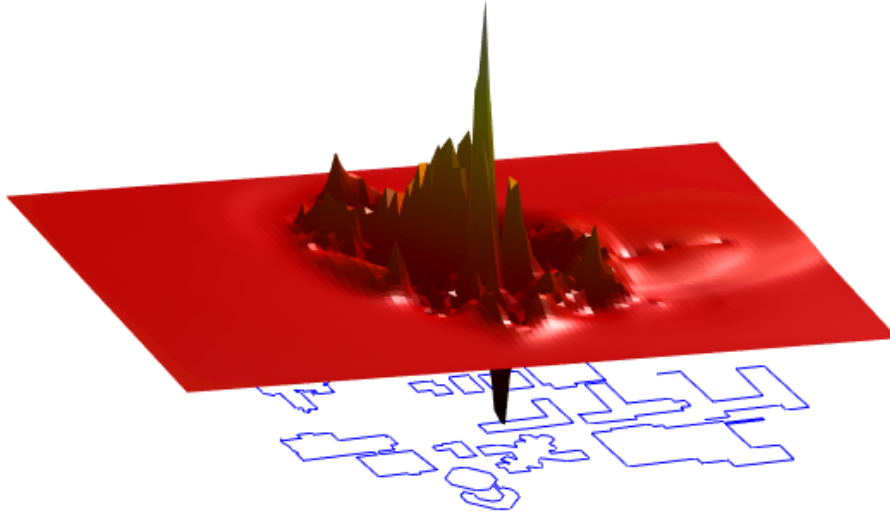


Figure 5.12: Estimated underlying field for the Saint Barbe route with Mercator BS. This field corresponds to a ridge regularization where the C matrix is produced from a isotropic gaussian kernel $k_{SE}(\cdot)$ with $\sigma_f = 1$ and $l_f = 5$.

The covariance of the shadowing $\mathbf{b}$, assuming a deterministic APM is given in (5.59). An immediate question arise from the procedure to find the underlying field $\mathbf{f}$: if the matrix $\mathbf{K}_f$ in (5.59) is determined from gaussian regression from training samples of $\mathbf{f}$, isn't it clear that we are imposing an stationary covariance matrix for $\mathbf{f}$ when selecting $\mathbf{G}$ to be isotropic in order to solve (5.71)? The answer depends on the regularization factor λ_R . If λ_R is smaller compared to 1, so that the regularization term has relevance in the solution than the MLS term, the actual covariance matrix of $\mathbf{f}$ could be non-stationary.

Gaussian regression is performed for a training sample of the estimated underlying field $\mathbf{f}^*$. $\frac{n}{2}$ equally separated samples from the grid points are chosen as the training locations $\mathcal{X}$. The remaining grid points are the test locations. Figure 5.14 shows the true underlying field at only the test locations and the prediction from the gaussian regression at the test locations. Although,

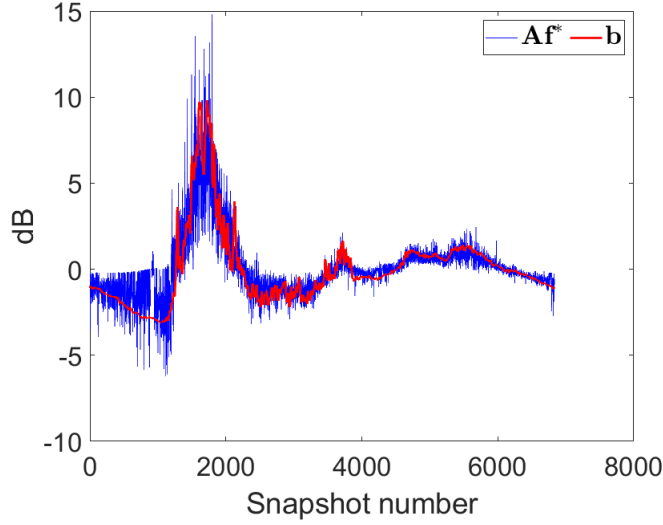


Figure 5.13: Shadowing signal for the Saint Barbe route with Mercator BS, and the reconstructed signal $\mathbf{Af}^*$.

the actual values of both fields are different, after normalizing both the true underlying field values and their predictions, the accuracy of the prediction is better appreciated. All the estimated $\sigma(\mathbf{x}_{i,k})$ are equal to 1, which was the value used for the isotropic covariance matrix $\mathbf{G}^{-1}$ parameter σ_f . On the other hand, estimated lengthscales $l_{i,k}$ are shown in Figure 5.15. Spatially-varying lengthscales in both dimensions have near constant values in the range $[0.2, 0.25]$, for almost all the grid points.

Finally, the correlation matrix of $\mathbf{b}_t$ is computed from (5.59), by dividing the covariance matrix $\mathbf{K}_b = \mathbf{AK}_f\mathbf{A}^T$ of $\mathbf{b}$ over the spatially-varying underlying field variances

$$\mathbf{R}_b = \text{diag}(\mathbf{K}_f)^{-\frac{1}{2}} \mathbf{K}_f \text{diag}(\mathbf{K}_f)^{-\frac{1}{2}} \quad (5.78)$$

where $\text{diag}(\mathbf{K}_f)$ is the matrix of the diagonal elements of $\mathbf{K}_f$ (the spatially-varying field variances). Figure 5.16 shows the computed correlation matrix for the shadowing. A close look at the graph reveals the matrix is not completely symmetric. It is positive definite, but possible due to round trip errors is not completely symmetric. We say is not 'completely symmetric' since the Frobenius norm $\|\mathbf{K}_b - \mathbf{K}_b^T\|_F \sim 10^{-5}$. High correlations are found for multiple regions, some of which are enclosed by ellipses in the figure. There is an unexpected high correlation for the shadowing samples near the peak and those at the end of the route. This is indicative that the matrix correlation $\mathbf{K}_b$ found might be not reliable, due to bad fit done by the gaussian regression.

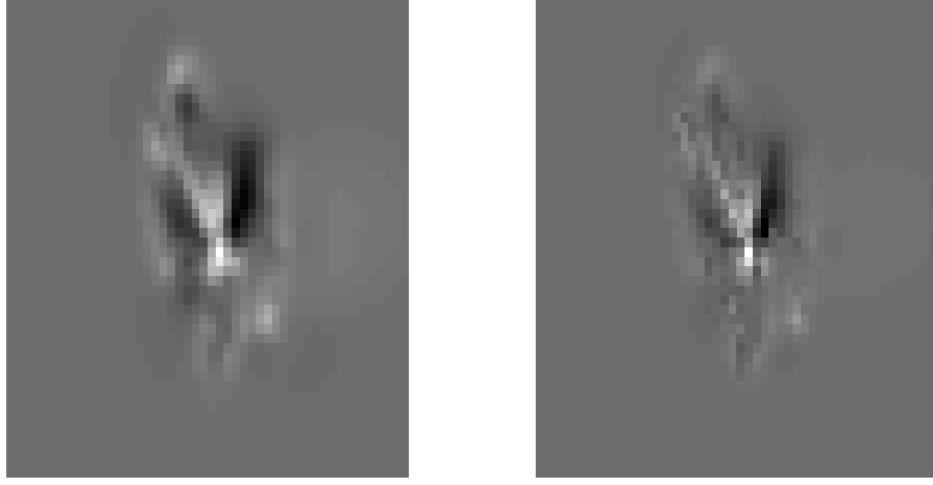


Figure 5.14: Left: true underlying field at test locations after min-max normalization. Right: predicted underlying field at test locations after min-max normalization.

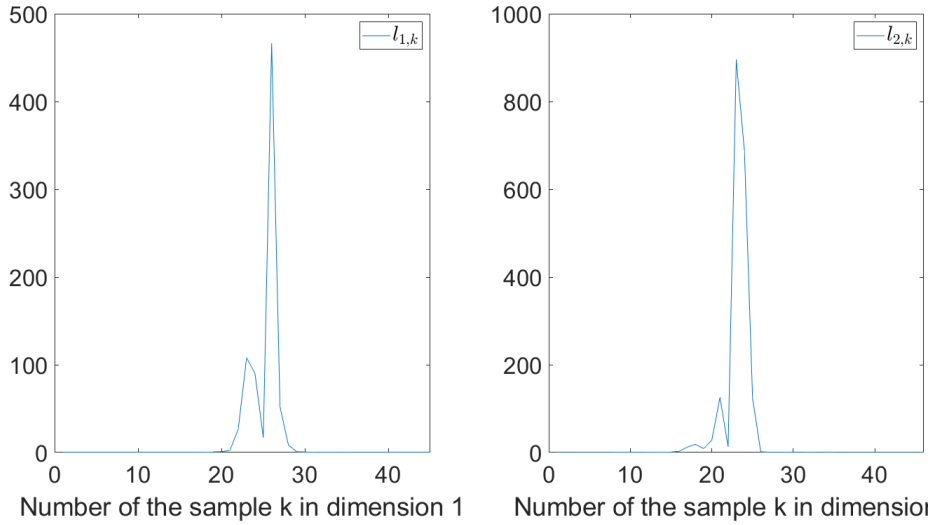


Figure 5.15: Estimated lengthscales for the non-stationary gaussian kernel fitted through gaussian regression. Left: estimated lengthscales in the first dimension ($l_{i,1}$). Right: estimated lengthscales in the second dimension ($l_{i,2}$).

5.8 Prediction

It was said in section 5.7 that the matrix $\mathbf{A}$ is deterministic. However, if such a matrix, instead of being a constant is regarded as a realization of a stochastic

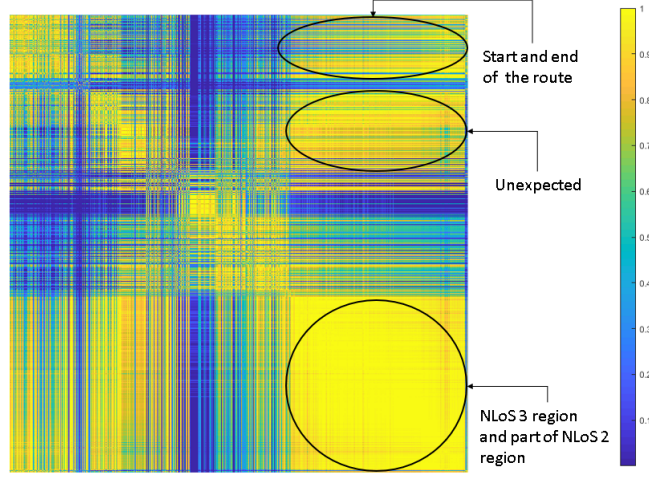


Figure 5.16: Correlation matrix $\mathbf{K}_b \in \mathbb{R}^{6838 \times 6838}$ for the shadowing $\mathbf{b} \in \mathbb{R}^{6838}$ of the Saint Barbe route with Mercator BS. The number of samples of the shadowing signal is 6838.

process, it could be used to model propagation environments with similar characteristics to the one considered along this thesis. More precisely, if there is a model for the APM that can produce a new matrix $\mathbf{A}_{\mathcal{X}^*}^{MDL}$ at any locations in $\mathcal{X}^*$, given that the underlying field $\mathbf{f}$ is estimated by solving (5.71), predictions for the signal $\mathbf{b}$ could be found at new locations by computing

$$\mathbf{b}_{\mathcal{X}^*} \sim \mathbf{A}_{\mathcal{X}^*}^{MDL} \mathbf{f}^* \quad (5.79)$$

where $\mathcal{X}^*$ is a new set of locations where the shadowing signal $\mathbf{b}$ want to be computed. The success of such predictions will depend mainly on how well the estimated underlying field $\mathbf{f}^*$ actually represents the true underlying field over $\mathcal{A}$, and how well is the model of the APM. A more realistic computation of such field would require to place multiple base stations and mobile stations in the perimeters of the area $\mathcal{A}$, so that the links between mobile stations and base stations traverse the space under consideration as is the case of what is done in Figure 1 in [WPV09]. However, for the case being considered (one base station at a given location and mobile station at the Saint Barbe route), the links between known transmitter-receiver locations (for which the channel have been sensed) also traverse regions of space for which measurements are not available. This is shown in Figure 5.17. This situation suggest that the estimated underlying field could provide representative information of the true underlying field at least for these locations. Even more, if the non-stationary model for the gaussian underlying field is substituted by an isotropic model, we can still use the same $\mathbf{f}^*$ and (5.79) for new locations $\mathcal{X}^*$)

outside of the region traversed by the links, but a distance with respect to these region much smaller than the decorrelation distance of the covariance function of the isotropic model.

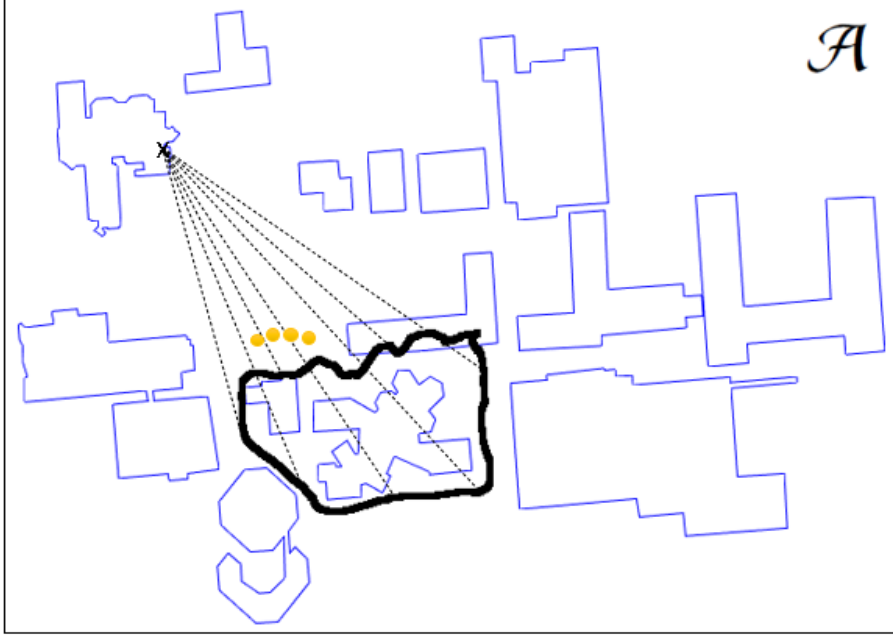


Figure 5.17: Locations where no measurement was performed but that lie into the region traversed by the links between the mobile station positions and the base station. The dashed lines represent the direct link. The yellow circles represent the new locations $\mathcal{X}^*$ for which there are no measurements.

Besides a good estimate for the underlying field, it is required that the model $\mathbf{A}_{\mathcal{X}^*}^{MDL}$ approximates $\mathbf{A}$, in order to produce a good estimate for $\mathbf{b}_{\mathcal{X}^*}$. For instance, let's call $\mathbf{A}_{\mathcal{X}}^{MDL}$ the matrix produced by the model at the measured locations $\mathcal{X}$. In such a case the measured shadowing is $\mathbf{b}$. It is clear from (5.79), that if $\mathbf{A}_{\mathcal{X}}^{MDL} \sim \mathbf{A}$, then $\mathbf{b}_{\mathcal{X}} \sim \mathbf{b}$. Since the predicted shadowing at the known measured locations $\mathbf{b}_{\mathcal{X}}$ should be equal to the measured shadowing at those locations $\mathbf{b}$, the matrix $\mathbf{A}_{\mathcal{X}^*}^{MDL}$ has to be very similar to $\mathbf{A}$. This is a challenge due to the fact that $\mathbf{A}$ is very ill-conditioned, and so by definition of an ill-conditioned matrix, any small change $\delta\mathbf{A}$ in the entries of $\mathbf{A}$ produce a big change $\delta\mathbf{b}$ in $\mathbf{b}$: $(\mathbf{A} + \delta\mathbf{A}) = \mathbf{b} + \delta\mathbf{b}$.

If $\mathbf{A}$ is given, an approximation of it is already known as the low rank approximation matrix $\hat{\mathbf{A}}$ that solves the minimization problem (5.80)

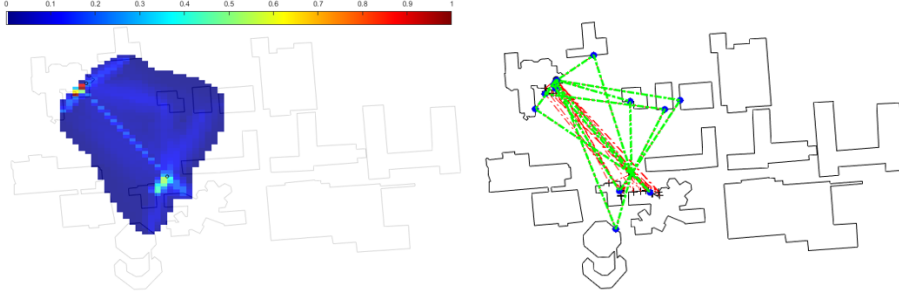


Figure 5.18: Left panel: Normalized weights for a given snapshot, for the Saint Barbe route. Right panel: Propagation paths and centroids for the same snapshot. Transceivers are represented with black dots. Dashed red lines are propagation paths. Black crosses represent propagation paths first and last bounces. Blue dots are centroid first and last bounces. Thick green line represent the propagation path of the centroid. There are 24 propagation paths and 9 clusters for the snapshot. Six of the clusters are singleton, one has 2 members, another 7 and another 9.

$$\begin{aligned} \min & ||\mathbf{A} - \hat{\mathbf{A}}||_F \\ \text{s.t. } & \text{RANK}(\mathbf{A}) \leq k \end{aligned} \quad (5.80)$$

Due to the Eckart-Young-Mirsky Theorem the solution 5.80 is built as follows. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{U} \in \mathbb{R}^{m \times r}$, $\mathbf{V} \in \mathbb{R}^{n \times r}$. Given the SVD of $\mathbf{A}$, $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^H$ with $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$, we have that:

$$\hat{\mathbf{A}} = \sum_{i=1}^k \lambda_i \mathbf{U}_{[:,i]} \mathbf{V}_{[:,i]}^H \quad (5.81)$$

However, the real challenge lies in modelling $\mathbf{A}$ in such a way that, at the *same time*, it closely approximates $\mathbf{A}$ at the known locations, and is general-enough to predict matrix $\mathbf{A}_{\mathcal{X}^*}^{MDL}$ at any new locations $\mathcal{X}^*$. From the previous discussion about the underlying field, those locations would ideally have to lie in the region traversed by the links between the mobile station and the base station, or close enough to this region. If that is not the case, the accuracy of the prediction of the shadowing at new locations, using the estimated underlying field at the known locations, will depend on the new locations. More specifically, it will depend on the degree of *alignment* with respect to the base station, between the route used to estimate the underlying field and the route used to predict the shadowing. It is meant by *alignment*, the angular difference between the line segment going from the center of each route to the base station. The lower the angular difference the higher the alignment,

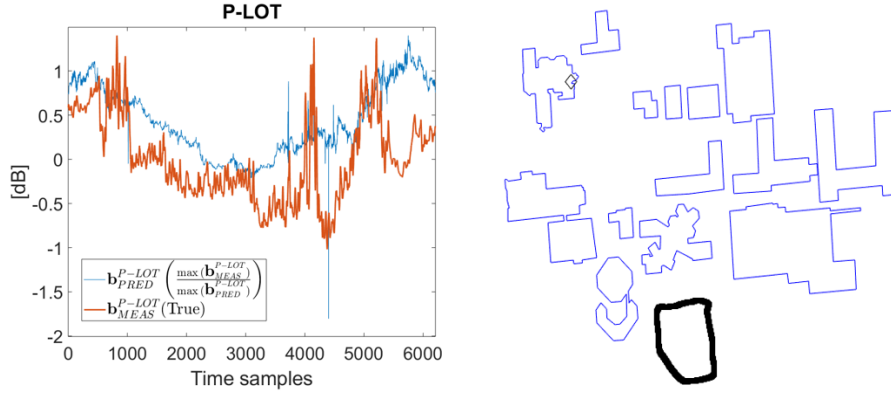


Figure 5.19: Shadowing prediction for Parking Lot mobile route. The underlying field was estimated from the Saint Barbe route with $n^1 = n^2 = 70$. Left panel: shadowing prediction and ground-truth value for the Parking lot route shown in the right panel of this figure. Right panel: Parking Lot route.

and the better the shadowing prediction. This can be seen in Figure 5.19 and 5.20. The underlying field was estimated using the same parameters used in the field shown in Figure 5.12, but the number of input locations in each dimension is reduced.

The figures show the shadowing predictions for 2 routes. The shadowing predicted for the Parking Lot mobile route has a 0.8 correlation with respect to the ground truth shadowing for the same route. On the other side, the shadowing predicted for the Marie Curie mobile route, does not exhibit a high correlation (the correlation is 0.54). The alignment between Saint Barbe and Parking Lot (around 10°) is higher than the one between Saint Barbe and Marie Curie (around 26°), which explains the better prediction for the Parking Lot route. However, for the highlighted magenta region of the Marie Curie route, the predicted shadowing follows the trend of the ground-truth signal.

It is convenient to have a closer look at the Parking Lot shadowing predictions, for locations of the mobile station that are close in distance and orientation to locations in the Saint Barbe route. As a convention, all symbols associated with Saint Barbe route are denoted as $\{\}_{ST-BARBE}^{MEAS}$, to stress the fact that they use information extracted from measurements. On the other hand, all symbols associated with the Parking Lot route are denoted by $\{\}_{P-LOT}^{MDL}$, to stress that some model is used as prior information instead of measurements.

For all the ellipses in singleton clusters and for the second ellipses of all clusters, the semi-minor axis is set to 40% the distance between one of the foci and the center of the ellipse. It means we assume energy spreads with an angle $\phi_t^2 \simeq 43.6^\circ$ from first bounce towards last bounce, for the case of the second ellipses. This choice is heuristically based. For a given time snapshot,

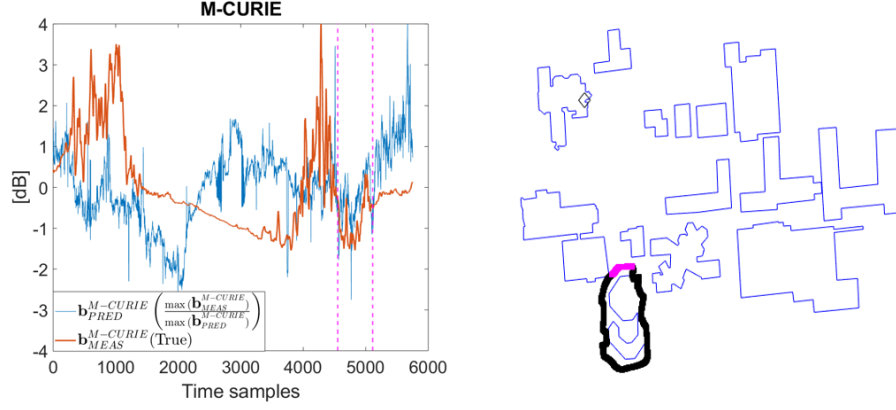


Figure 5.20: Shadowing prediction for Marie Curie mobile route. The underlying field was estimated from the Saint Barbe route with $n^1 = n^2 = 70$. Left panel: shadowing prediction and ground-truth value for the route. Right panel: Marie Curie mobile route.

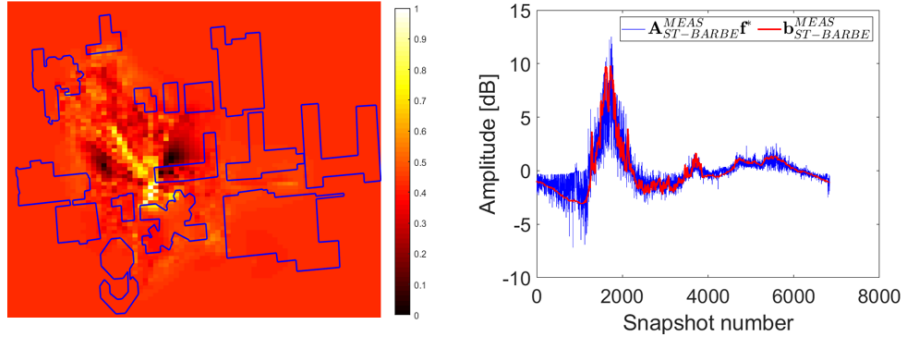


Figure 5.21: Left panel: Estimated SLF $\mathbf{f}^*$, using Saint Barbe mobile station route. Right panel: Reconstructed $\mathbf{A}\mathbf{f}^*$ and measured shadowing from Saint Barbe mobile station route.

the computed weights and corresponding clusters of propagation paths are shown in Figure 5.18 to illustrate a row of the APM. Tikhonov regularization is used, assuming a stationary gaussian random field with covariance matrix produced by the kernel in 5.75.

Figure 5.21 shows the estimated SLF $\mathbf{f}^*$ and the reconstructed shadowing $\mathbf{A}\mathbf{f}^*$, when $\delta_f = 2$, $\sigma_f = 30$ and $\lambda_R = 0.1$ for the Saint Barbe route. The ground truth shadowing shown in red is computed from the measured channel impulse response, by filtering fading with a moving average window of size 10λ . Then, path loss trend with distance is found and removed as well. The higher the regularization constant the smoother is the SLF and the higher the MSE.

Shadowing prediction was done for a part of the Parking Lot route with

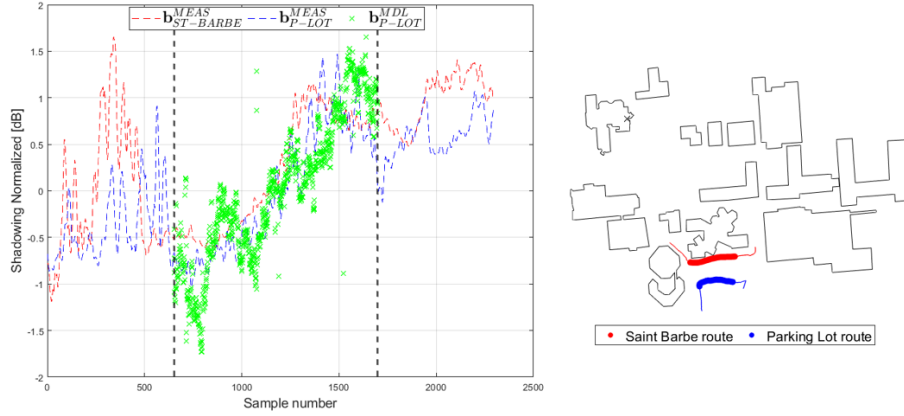


Figure 5.22: Left panel: shadowing predicted and from measurements. $b_{ST-BARBE}^{MEAS}$ is the shadowing from measurements for the Saint Barbe route, b_{P-LOT}^{MEAS} is the shadowing from measurements for the Parking Lot route, and b_{P-LOT}^{MDL} is the predicted shadowing for the Parking Lot route. Right panel: mobile station routes considered. The thicker points correspond to the snapshot zone in between the dashed black lines in the left panel.

Mercator BS. This part of the route was selected since it is close to the Saint Barbe route, for which the APM was computed. The weights for the new locations are produced by assuming some prior information for the angles of arrival and departure, instead of finding them from measurements through SAGE. The model used for the APM $\mathbf{A}_{MDL}^{P-LOT}$ at the Parking Lot consists in what follows. We assume that for the mobile station side, angles departure within an interval of $[-\pi, \pi]$ rads, with the LoS as the angular reference. For the base station side, we assume angles arrive within an interval of $[-70^\circ, 70^\circ]$, again, with the LoS as the angular reference. The choices are justified by the beamwidths of the mobile station and base station antennas. First and last bounces in the buildings are found from the prior angles of departure and arrival. Then, artificial propagation paths are found. From such paths is possible to compute the entries of $\mathbf{A}_{MDL}^{P-LOT}$ using the proposed model in 5.9. The predicted shadowing for the part of the Parking Lot route then becomes: $\mathbf{b}_{P-LOT}^{MDL} = \mathbf{A}_{P-LOT}^{MDL} \mathbf{f}^*$. For computing the new weights, the semi-minor axis of all second ellipses and all ellipses for singleton clusters is set to 70% the distance between one of the foci and the center of the ellipse. The choice of the semi-minor axis for the model and measured cases, is justified as it provides the highest correlation between predicted and measured shadowing. Furthermore, a wider angle ϕ_i^2 choice for the prediction is justified as no information of the propagation paths is known beforehand.

Predicted shadowing and ground truth shadowing for mobile station trajectories at Saint Barbe and Parking Lot are shown in Figure 5.22. The mean

value of all shadowing signals was subtracted to their signals. In addition, the predicted shadowing $\mathbf{b}_{P-LOT}^{MDL}$ was scaled by $\max(\mathbf{b}_{ST-BARBE}^{MEAS})/\max(\mathbf{b}_{P-LOT}^{MDL})$. This scaling is justified as the absolute value of the weights is not known, but rather than that, their relative value. This means that the entries of $\mathbf{A}_{P-LOT}^{MDL}$ can be in a different scale than the entries of $\mathbf{A}_{ST-BARBE}^{MDL}$, and so the same occurs to the entries of $\mathbf{b}_{P-LOT}^{MDL}$. Both mobile station trajectories shown in the right panel are very similar. For instance, there is a high correlation ($\rho(\mathbf{b}_{ST-BARBE}^{MEAS}, \mathbf{b}_{P-LOT}^{MEAS}) = 0.916$) between the shadowing extracted from measurements for both routes. This high correlation is due to the average short distance without obstacles between the routes and their alignment.

Pearson Correlation coefficient between the predicted shadowing at Parking Lot and the shadowing from measurements at Saint Barbe is $\rho(\mathbf{b}_{M-CURIE}^{MEAS}, \mathbf{b}_{P-LOT}^{MDL}) = 0.834$ (for the interval in between the black dashed lines in the left panel in the figure). This high correlation is showing that the prediction captures the trend of the shadowing, even though no Parking Lot measurement was taken into account for the SLF computation.

5.9 Conclusion

The linear model used for shadowing is capable of explaining the relationship between propagation features as the clusters of propagation paths, introduced in previous chapters, and a mathematical entity, called here the underlying spatial loss field, accounting for the density of losses across the scenario under consideration. This relationship is a linear one, but more important, through this model, the common believe that shadowing accounts for losses due to obstacles is justified in a more specific way. Losses arise from the contribution of spatial intrinsic losses, accounted by the underlying spatial loss field, through the clusters of propagation paths were it is believed (based on measurements) most dominant power transits. The accuracy of the prediction strongly relies on the estimation of the underlying field. However, it has been shown that predictions can be follow the trend of the ground-truth signal even when the underlying field was not properly estimated, or in better words, even when the estimated field is actually not underlying/intrinsic. A condition for estimating such underlying field was presented and has to be considered at the measurement planning step. When the estimated field is not truly an underlying field and it has been computed using only one base station, the accuracy of the shadowing prediction for new transmitter-receiver locations depends of the alignment (with respect to the base station) between the transmitter-receiver locations used to estimate spatial loss field, and the new locations used for prediction.

Conclusions and perspectives on extensions to this work

6

Wireless radio channels are considered LTI systems under the assumption that the channel remains stationary within a local zone or time window. Such an assumption is essential to any measurement procedure of the channel impulse response matrix. However, the fact that true wireless radio channels are in fact non-stationary is itself a reason to analyse which are the sources of non-stationarity and to which extent it is possible to assume a LTI system where the input and the output of such system are related by linear operators as the Fourier transform.

From the perspective of the network operator it might be useful to have ready-to-go or simple analytical tools that are able to provide a first insight into propagation of radio signals in the environment. Large scale parameters as the path loss exponent and the shadowing variance, for the case of log-normal shadowing, are an example of such type of measures. However, with the deployment of more complex communication systems, more rapidly evolving channels, and more densely connected networks as in the case of 5G technologies, large-scale parameters are not enough to characterize received signals correlated in space. This might be useful for adaptive control of the power, since in principle, less correlated locations might require more power than higher correlated locations. Thus, if the network operator has to decide in which direction beamforming is done, or where to point the bigger lobes of the radiated energy, bigger correlated regions require less attention than isolated regions, if the purpose is to serve all users equally.

The plane wave assumption under which the MIMO channel matrix is approximated by a model characterized by a set of delays, angles, power factors and Doppler shifts accounts for a percentage of the total power that depends on the type of propagation region: LoS, OLoS, NLoS. For LoS propagation regions, almost all the average received power over such a region is accounted by the plane wave assumption model. Despite in a lower degree, also for the OLoS propagation region, 70% of the average received power was encountered to be accounted by the plane waves model. NLoS propagation regions are seen to rely on more mechanisms of propagation than specular reflec-

tion, for instance diffuse scattering, and so for such a case, up to 50 % of the average received power of the region is explained by the model. For propagation regions with strong NLoS conditions, less percentage of the received power is accounted by the model. For such regions, other models that account for the dense multipath components are required, and high resolution parameter estimation algorithms as RIMAX can be used to find out the parameters of the model. Since in the received power (and so the shadowing) signal all the propagation phenomena are taken into account, predictive models (as presented in Chapter 5) for such a signal relying on propagation paths computed from the estimated parameters of the plane wave model, lack of enough information.

Non-stationary measures of the received power as the roughness parameter provide one-value information about the time evolution of the channel. Although not explored in this thesis, the roughness parameter could be used to categorize different types of time-varying channels, in addition to Doppler information. If there are no available channel matrices, only received power information is required or any signal related to it as the RSSI. In addition to the non-stationarity measure, non-stationary regions have been computed as the regions where the instantaneous PDP is averaged. No unique methodology exists to find out such regions, but it has been seen that not enough methodologies are available to compute such regions. For the measurement campaign considered, discussion about the link between such regions and the mechanisms of propagation. However, extensions to this work are possible. It is important to distinguish two things in order to understand which are the challenges of finding such regions. On one side, stationary regions are computed as the regions required to average the instantaneous PDP, and obtain a distribution of power of the delay domain. For those types of regions it is required to look the time evolution of the instantaneous PDP. However, in a more broad sense, it is also of interest to look for stationary regions where the channel from a directional perspective is not changing. Or in other words, to look for regions where the mechanisms of propagation (reflectors, diffraction points) remain the same. As those mechanisms are not accounted by the PDP, different stationary regions from the ones computed by looking at the instantaneous PDP might arise.

Multipath clustering has been widely considered in the parameter space, although intuition about the MCD measure is not straightforward. Due to this, clusters of estimated channel parameters for the plane wave model, might be badly clusterized. Furthermore, due to clusters in which the angular difference between members of the cluster might be non-negligible, and to the fact that with distance this angular differences imply bigger differences, mapping of the clusters to the physical space produce non-desirable results. This mapping is done as a simple tool to identify main reflectors over the sce-

nario, which are specific objects: a part of a building, a light post, a window, etc. However, when mapping parameter-space clusters, wide apart objects are considered part of the same cluster: for instance, two regions of two different separated buildings might be considered belonging to the same cluster, although they are clearly separated reflectors.

Clustering algorithms as k -means relying on providing the number of clusters required metrics to compare different clusterizations or partitions. This metrics are not suitable for the dataset considered: propagation paths. For such a reason, hierarchical clustering is used, despite its computation speed is lower for bigger datasets. The benefit gained through its use is that the distance measured used could be understood as an average distance between propagation paths, and it controls the threshold required by hierarchical clustering to cut the dendrogram and produce a partition. As a complement of hierarchical clustering, fuzzy clustering was also introduced. Although, it requires the number of clusters, fuzzy clustering provides a probability approach to identify clusters. Further analysis into that direction could led to probability distribution of propagation paths over space, or at least some regions of space: if the number of estimated paths is big enough, and the propagation paths are spread over different regions of space rather than concentrating in one region (as in the LoS case), the probability that a propagation path is close to the computed centroid by the algorithm, could be seen as the probability that the propagation path is nearby a region enclosing the centroid. As said before, if there are enough propagation paths, a probability distribution for each propagation path could be fit. This distribution refers to the probability of being part of the cluster identified by the centroid of the cluster, which in turn is tied to the location of the centroid in space.

Shadowing values at a given pair of transmitter-receiver locations were modelled as the superposed contribution of an underlying field over all propagation paths between the transmitter and receiver locations. If such underlying field is actually an underlying field, that is to say, if it comprises propagation characteristics, shadowing values at any pair of transceiver-receiver coordinates could be computed. However, such a situation is only possible if the links between the transmitter and receiver locations traverse the whole space $\mathcal{A}$ or a representative section of it. This was not the case of the measurements considered in this document. For those measurements, links between transmitter and receiver locations only traverse a subset of $\mathcal{A}$. Thus, the underlying field is rather than underlying, very specific to the measurements carried. For outdoor micro-cell scenarios, existing cellular infrastructure could be used to compute the underlying field. If the receiver stations are spread over $\mathcal{A}$, a realistic assumption for cellular networks, and users spread over the cell of the receiver station, a realistic underlying field computation could be done. This would require estimation of the channel parameters, and

ultimately propagation paths, as a step to be performed offline.

Non-stationary correlation for the shadowing was computed. Since it relies on the assumption that the APM is deterministic, the correlation found could not be regarded as model of correlation extensible to other channel conditions. Rather than that, the non-stationary correlation assumes all the stochastic behaviour arises from the underlying field, which in turn requires good estimate of such field and a good gaussian regression fit. Suitable measurements should be used to validate the non-stationary correlation model. Coming back to the cellular network case, simultaneous measurements of the received power time evolution (and so, the shadowing) for multiple users spread over the cell, are a way to compute the empirical shadowing correlation between links at the different locations of the users. Such empirical correlation is a direct way to validate the shadowing correlation computed from the approach proposed.

Although shadowing prediction for new places is constrained by the degree of correlation among the reference trajectory (Saint Barbe route) and the model trajectory (part of the Parking Lot route) it provides an insight into the tendency of shadowing. The prediction must not be regarded as a low-error (in the sense of Mean Squared Error or Mean Absolute Error) prediction, but instead as capable of follow the tendency of the real signal. A situation that remains challenging and that defines the quality of the prediction is the ability to synthesize matrix $\mathbf{A}$ from a model that doesn't depend on the measurements, or that captures the structure of many measurements for a given environment type (i.e. semi-urban microcell). Here, by assuming omnidirectional angles of arrival to synthesize $\mathbf{A}$ at the new locations (Parking Lot route), a high correlation between predicted and ground truth signal is achieved. However, NLoS propagation conditions must probably drop omnidirectional assumption, resulting in low fitness between the real $\mathbf{A}$ and the one synthesize from a model. On the other hand, as no measurement information is assumed available for the prediction stage, the model used to synthesize $\mathbf{A}$ must assume the more general prior information (omnidirectional angles of departure), giving arise to a trade-off between prediction accuracy and prediction need for measurements.

References Chapter Channel sounder and Parameter Estimation

- [Bal12] C. A. Balanis. *Advanced engineering electromagnetics*. John Wiley & Sons, 2012.
- [Bar48] M. S. Bartlett. “Smoothing periodograms from time-series with continuous spectra”. In: *Nature* 161.4096 (1948), pp. 686–687.
- [Bar53] M Bartlett. “The statistical approach to the analysis of time-series”. In: *Transactions of the IRE Professional Group on Information Theory* 1.1 (1953), pp. 81–101.
- [Cap69] J. Capon. “High-resolution frequency-wavenumber spectrum analysis”. In: *Proceedings of the IEEE* 57.8 (1969), pp. 1408–1418.
- [CGK67] J Capon, R. Greenfield, and R. J. Kolker. “Multidimensional maximum-likelihood processing of a large aperture seismic array”. In: *Proceedings of the IEEE* 55.2 (1967), pp. 192–211.
- [CL75] D. Cox and R Leck. “Correlation bandwidth and delay spread multipath propagation statistics for 910-MHz urban mobile radio channels”. In: *IEEE Transactions on Communications* 23.11 (1975), pp. 1271–1280.
- [Cla68] R. H. Clarke. “A statistical theory of mobile-radio reception”. In: *Bell system technical journal* 47.6 (1968), pp. 957–1000.
- [Czi+08b] N. Czink, G. Del Galdo, X. Yin, E. Bonek, and J. Ylitalo. “A novel environment characterization metric for clustered MIMO channels”. In: *Wireless Personal Communications* 46.1 (2008), pp. 83–98.
- [Dev86] D Devasirvatham. “Time delay spread and signal level measurements of 850 MHz radio waves in building environments”. In: *IEEE Transactions on antennas and propagation* 34.11 (1986), pp. 1300–1305.

- [Dzi+99] Z. Dziong, M Krishnan, S Kumar, and S. Nanda. "Statistical" snapshot" for multi-cell CDMA system capacity analysis". In: *WCNC. 1999 IEEE Wireless Communications and Networking Conference (Cat. No. 99TH8466)*. Vol. 3. IEEE. 1999, pp. 1234–1237.
- [Ekm02] T. Ekman. "Prediction of mobile radio channels: Modeling and design". PhD thesis. Institutionen för materialvetenskap, 2002.
- [FH94] J. A. Fessler and A. O. Hero. "Space-alternating generalized expectation-maximization algorithm". In: *IEEE Transactions on signal processing* 42.10 (1994), pp. 2664–2677.
- [FJS02] B. H. Fleury, P. Jourdan, and A. Stucki. "High-resolution channel parameter estimation for MIMO applications using the SAGE algorithm". In: *2002 International Zurich Seminar on Broadband Communications Access-Transmission-Networking (Cat. No. 02TH8599)*. IEEE. 2002, pp. 30–30.
- [Fle+02] B. H. Fleury, X. Yin, K. G. Rohbrandt, P. Jourdan, and A. Stucki. "High-resolution bidirection estimation based on the SAGE algorithm: Experience gathered from field experiments". In: *Proc. XXVIIth General Assembly of the Int. Union of Radio Scientists (URSI)* 2127 (2002).
- [Fle+99] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. I. Pedersen. "Channel parameter estimation in mobile radio environments using the SAGE algorithm". In: *IEEE Journal on selected areas in communications* 17.3 (1999), pp. 434–450.
- [Haa+95] M. Haardt, M. D. Zoltowski, C. P. Mathews, and J Nossek. "2D unitary ESPRIT for efficient 2D parameter estimation". In: *1995 International Conference on Acoustics, Speech, and Signal Processing*. Vol. 3. IEEE. 1995, pp. 2096–2099.
- [Jak74] W. C. Jakes. "Multipath interference". In: (1974).
- [KV96] H. Krim and M. Viberg. "Two decades of array signal processing research: the parametric approach". In: *IEEE signal processing magazine* 13.4 (1996), pp. 67–94.
- [Lan08] M. Landmann. *Limitations of experimental channel characterisation*. Citeseer, 2008.
- [LN07] L. Liu and D. K. Nielsen. "Parametric Characterization and Estimation of Dispersive Multi-Path Components with SAGE in Radio Propagation Channel". MA thesis. 2007.
- [Niz14] K. Nizabat. "Multi-Link Channel Modeling and Interference Characterization for Beyond 4G Networks". PhD thesis. École Polytechnique de Louvain, Université Catholique de Louvain, 2014.

- [OJ64] J. F. Ossanna Jr. "A model for mobile radio fading due to building reflections: Theoretical and experimental fading waveform power spectra". In: *Bell System Technical Journal* 43.6 (1964), pp. 2935–2971.
- [Ric05] A. Richter. "Estimation of radio channel parameters: Models and algorithms". PhD thesis. TU Ilmenau, 2005.
- [RLT03] A. Richter, M. Landmann, and R. Thoma. "Maximum likelihood channel parameter estimation from multidimensional channel sounding measurements". In: *The 57th IEEE Semiannual Vehicular Technology Conference, 2003. VTC 2003-Spring*. Vol. 2. IEEE. 2003, pp. 1056–1060.
- [Sch82] R. O. Schmidt. *A signal subspace approach to multiple emitter location and spectral estimation*. Stanford University, 1982.
- [Sch86] R. Schmidt. "Multiple emitter location and signal parameter estimation". In: *IEEE transactions on antennas and propagation* 34.3 (1986), pp. 276–280.
- [SMB01] M. Steinbauer, A. F. Molisch, and E. Bonek. "The double-directional radio channel". In: *IEEE Antennas and propagation Magazine* 43.4 (2001), pp. 51–63.
- [Ste98] M. Steinbauer. "A comprehensive transmission and channel model for directional radio channels". In: *EURO-COST, TD (98) 27* (1998).
- [Tur+72] G. L. Turin, F. D. Clapp, T. L. Johnston, S. B. Fine, and D. Lavry. "A statistical model of urban multipath propagation". In: *IEEE Transactions on Vehicular Technology* 21.1 (1972), pp. 1–9.
- [Vin17] E. Vinogradov. "Multi-dimensional radio channel models for indoor distributed communications". PhD thesis. École Polytechnique de Louvain, Université Catholique de Louvain, 2017.
- [VR87] J. Van Rees. "Measurements of the wide-band radio channel characteristics for rural, residential, and suburban areas". In: *IEEE Transactions on Vehicular Technology* 36.1 (1987), pp. 2–6.
- [Xia92] N. Xiang. "Using M-sequences for determining the impulse responses of LTI-systems". In: *Signal Processing* 28.2 (1992), pp. 139–152.
- [Yin+03] X. Yin, B. Fleury, P. Jourdan, and A. Stucki. "Polarization estimation of individual propagation paths using the SAGE algorithm". In: *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003*. Vol. 2. 2003, 1795–1799 vol.2. DOI: 10.1109/PIMRC.2003.1260424.

References Chapter

Propagation Analysis

- [AK99] A. Abdi and M. Kaveh. “On the utility of gamma PDF in modeling shadow fading (slow fading)”. In: *1999 IEEE 49th Vehicular Technology Conference (Cat. No. 99CH36363)*. Vol. 3. IEEE. 1999, pp. 2308–2312.
- [Ber09] P. Berens. “CircStat: a MATLAB toolbox for circular statistics”. In: *Journal of statistical software* 31.1 (2009), pp. 1–21.
- [Czi+08a] N. Czink, B. Bandemer, G. Vazquez-vilar, L. Jalloul, and A. Paulraj. *Can Multi-User MIMO Measurements Be Done Using a Single Channel Sounder?* 2008.
- [DN16] S. Das and G. P. Nason. “Measuring the degree of non-stationarity of a time series”. In: *Stat* 5.1 (2016), pp. 295–305. doi: <https://doi.org/10.1002/sta4.125>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sta4.125>.
- [Dur00] G. D. Durgin. “Theory of stochastic local area channel modeling for wireless communications”. PhD thesis. Virginia Tech, Nov. 2000.
- [Egl57] J. J. Egli. “Radio Propagation above 40 MC over Irregular Terrain”. In: *Proceedings of the IRE* 45.10 (1957), pp. 1383–1391. doi: [10.1109/JRPROC.1957.278224](https://doi.org/10.1109/JRPROC.1957.278224).
- [Fle00] B. Fleury. “First- and second-order characterization of direction dispersion and space selectivity in the radio channel”. In: *IEEE Transactions on Information Theory* 46.6 (2000), pp. 2027–2044. doi: [10.1109/18.868476](https://doi.org/10.1109/18.868476).
- [Geh+01] A. Gehring, M. Steinbauer, I. Gaspard, and M. Grigat. “Empirical channel stationarity in urban environments”. In: *Proceedings of the 4th European Personal Mobile Communications Conference (EPMCC’01)*. 2001.

- [GGF94] A. J. Goldsmith, L. J. Greenstein, and G. J. Foschini. “Error statistics of real-time power measurements in cellular channels with multipath and shadowing”. In: *IEEE Transactions on Vehicular Technology* 43.3 (1994), pp. 439–446.
- [Hal+17] D. Hallac, S. Vare, S. Boyd, and J. Leskovec. “Toeplitz inverse covariance-based clustering of multivariate time series data”. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017, pp. 215–223.
- [Her+05] M. Herdin, N. Czink, H. Ozcelik, and E. Bonek. “Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels”. In: *2005 IEEE 61st Vehicular Technology Conference*. Vol. 1. 2005, 136–140 Vol. 1. DOI: 10 . 1109 /VETECS . 2005 . 1543265.
- [LR92] J. Liberti and T. Rappaport. “Statistics of shadowing in indoor radio channels at 900 and 1900 MHz”. In: *MILCOM*. IEEE. 1992, pp. 45–6.
- [Mol12] A. F. Molisch. *Wireless communications*. Vol. 34. John Wiley & Sons, 2012.
- [Pai+08] A. Paier, T. Zemen, L. Bernado, G. Matz, J. Karedal, N. Czink, C. Dumard, F. Tufvesson, A. F. Molisch, and C. F. Mecklenbrauker. “Non-WSSUS vehicular channel characterization in highway and urban scenarios at 5.2 GHz using the local scattering function”. In: *2008 International ITG Workshop on Smart Antennas*. IEEE. 2008, pp. 9–15.
- [Ram04] J. O. Ramsay. “Functional data analysis”. In: *Encyclopedia of Statistical Sciences* 4 (2004).
- [Ren+10] O. Renaudin, V.-M. Kolmonen, P. Vainikainen, and C. Oestges. “Impact of correlation matrix estimation accuracy on the computation of stationarity intervals”. In: *Proceedings of the Fourth European Conference on Antennas and Propagation*. IEEE. 2010, pp. 1–5.
- [SAZ07] S. R. Saunders and A. Aragón-Zavala. *Antennas and propagation for wireless communication systems*. John Wiley & Sons, 2007.
- [Sha08] P. Shankar. “Macrodiversity and microdiversity in correlated shadowed fading channels”. In: *IEEE Transactions on Vehicular Technology* 58.2 (2008), pp. 727–732.

References Clustering Multiple Paths

- [CKV13] M. E. Celebi, H. A. Kingravi, and P. A. Vela. “A comparative study of efficient initialization methods for the k-means clustering algorithm”. In: *Expert systems with applications* 40.1 (2013), pp. 200–210.
- [Czi+05] N. Czink, P. Cera, J. Salo, E. Bonek, J. Nuutinen, and J. Ylitalo. “Automatic Clustering of MIMO Channel Parameters using the Multi-Path Component Distance Measure”. In: *International Symposium on Wireless Personal Multimedia Communications*. 2005.
- [Czi+06a] N. Czink, P. Cera, J. Salo, E. Bonek, J. Nuutinen, and J. Ylitalo. “Improving clustering performance using multipath component distance”. In: *Electronics Letters* 42 (2006), pp. 33–35.
- [Czi+06b] N. Czink, P. Cera, J. Salo, E. Bonek, J.-P. Nuutinen, and J. Ylitalo. “A framework for automatic clustering of parametric MIMO channel data including path powers”. In: *IEEE Vehicular Technology Conference*. IEEE. 2006, pp. 1–5.
- [DB79] D. L. Davies and D. W. Bouldin. “A cluster separation measure”. In: *IEEE transactions on pattern analysis and machine intelligence* 2 (1979), pp. 224–227.
- [Dun73] J. C. Dunn. “A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters”. In: (1973).
- [Dun74] J. C. Dunn. “Well-separated clusters and optimal fuzzy partitions”. In: *Journal of cybernetics* 4.1 (1974), pp. 95–104.
- [GVO19] J. D. V. Gutierrez, D. Pimienta-del Valle, and C. Oestges. “Physical Identification of Common Clusters for Outdoor Microcells”. In: *2019 13th European Conference on Antennas and Propagation (EuCAP)*. IEEE. 2019, pp. 1–5.
- [Han+17] P. Hanpinitasak, K. Saito, J.-i. Takada, M. Kim, and L. Materum. “Multipath clustering and cluster tracking for geometry-based stochastic channel modeling”. In: *IEEE Transactions on Antennas and Propagation* 65.11 (2017), pp. 6015–6028.

- [KH10] K. Kryszczuk and P. Hurley. “Estimation of the number of clusters using multiple clustering validity indices”. In: *International workshop on multiple classifier systems*. Springer. 2010, pp. 114–123.
- [KR90] L. Kaufman and P. J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 1990.
- [Liu+12] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. De Doncker. “The COST 2100 MIMO channel model”. In: *IEEE Wireless Communications* 19.6 (2012), pp. 92–99.
- [Mac+67] J. MacQueen et al. “Some methods for classification and analysis of multivariate observations”. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 14. Oakland, CA, USA. 1967, pp. 281–297.
- [PJ09] H.-S. Park and C.-H. Jun. “A simple and fast algorithm for K-medoids clustering”. In: *Expert systems with applications* 36.2 (2009), pp. 3336–3341.
- [Ste+02] M. Steinbauer, H. Ozcelik, H. Hofstetter, C. F. Mecklenbrauker, and E. Bonek. “How to quantify multipath separation”. In: *IEICE transactions on electronics* 85.3 (2002), pp. 552–557.
- [VGVO19] J. D. Villegas Gutierrez, D. Pimienta-del Valle, and C. Oestges. “Physical Identification of Common Clusters for Outdoor Micro-cells”. In: *2019 13th European Conference on Antennas and Propagation (EuCAP)*. 2019, pp. 1–5.
- [Wei03] W. Weichselberger. *Spatial structure of multiple antenna radio channels: a signal processing viewpoint*. na, 2003.
- [Zhu14] M. Zhu. *Geometry-based radio channel characterization and modeling: Parameterization, implementation and validation*. Lund University, 2014.

References Spatial Loss Field

- [AP09] P. Agrawal and N. Patwari. “Correlated link shadow fading in multi-hop wireless networks”. In: *IEEE Transactions on Wireless Communications* 8.8 (2009), pp. 4024–4036.
- [AS08] E. B. Anderes and M. L. Stein. “Estimating deformations of isotropic Gaussian random fields on the plane”. In: *The Annals of Statistics* 36.2 (2008), pp. 719–741.
- [AS11] E. B. Anderes and M. L. Stein. “Local likelihood estimation for nonstationary random fields”. In: *Journal of Multivariate Analysis* 102.3 (2011), pp. 506–520.
- [Boc55] S. Bochner. *Harmonic analysis and the theory of probability*. University of California press, 1955.
- [DSG01] D. Damian, P. D. Sampson, and P. Guttorp. “Bayesian estimation of semi-parametric non-stationary spatial covariance structures”. In: *Environmetrics: The official journal of the International Environmetrics Society* 12.2 (2001), pp. 161–178.
- [Fla+15] S. Flaxman, A. Wilson, D. Neill, H. Nickisch, and A. Smola. “Fast Kronecker inference in Gaussian processes with non-Gaussian likelihoods”. In: *International Conference on Machine Learning*. PMLR, 2015, pp. 607–616.
- [FS01] M. Fuentes and R. L. Smith. *A new class of nonstationary spatial models*. Tech. rep. North Carolina State University. Dept. of Statistics, 2001.
- [Gel+10] A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes. *Handbook of spatial statistics*. CRC press, 2010.
- [Gen02] M. G. Genton. “Classes of Kernels for Machine Learning: A Statistics Perspective”. In: *J. Mach. Learn. Res.* 2 (Mar. 2002), 299–312. ISSN: 1532-4435.
- [Gib98] M. N. Gibbs. “Bayesian Gaussian processes for regression and classification”. PhD thesis. Citeseer, 1998.
- [Gud91] M. Gudmundson. “Correlation model for shadow fading in mobile radio systems”. In: *Electronics letters* 27.23 (1991), pp. 2145–2146.

- [GWB97] P. W. Goldberg, C. K. Williams, and C. M. Bishop. “Regression with input-dependent noise: A Gaussian process treatment”. In: *Advances in neural information processing systems* 10 (1997), pp. 493–499.
- [Ham+13] B. R. Hamilton, X. Ma, R. J. Baxley, and S. M. Matechik. “Propagation modeling for radio frequency tomography in wireless networks”. In: *IEEE Journal of Selected Topics in Signal Processing* 8.1 (2013), pp. 55–65.
- [Hei+16] M. Heinonen, H. Mannerström, J. Rousu, S. Kaski, and H. Lähdesmäki. “Non-stationary gaussian process regression with hamiltonian monte carlo”. In: *Artificial Intelligence and Statistics*. PMLR. 2016, pp. 732–740.
- [Hig98] D. Higdon. “A process-convolution approach to modelling temperatures in the North Atlantic Ocean”. In: *Environmental and Ecological Statistics* 5.2 (1998), pp. 173–190.
- [HSK99] D. Higdon, J. Swall, and J. Kern. “Non-stationary spatial modeling”. In: *Bayesian statistics* 6.1 (1999), pp. 761–768.
- [Kat13] M. Katzfuss. “Bayesian nonstationary spatial modeling for very large datasets”. In: *Environmetrics* 24.3 (2013), pp. 189–200.
- [KN12] W. Kleiber and D. Nychka. “Nonstationary modeling for multivariate spatial processes”. In: *Journal of Multivariate Analysis* 112 (2012), pp. 76–91.
- [LKG17] D. Lee, S.-J. Kim, and G. B. Giannakis. “Channel gain cartography for cognitive radios leveraging low rank and sparsity”. In: *IEEE Transactions on Wireless Communications* 16.9 (2017), pp. 5953–5966.
- [LS19] Y. Li and Y. Sun. “Efficient estimation of nonstationary spatial covariance functions with application to high-resolution climate model emulation”. In: *Statistica Sinica* 29.3 (2019), pp. 1209–1231.
- [PA08] N. Patwari and P. Agrawal. “NeSh: A joint shadowing model for links in a multi-hop network”. In: *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. 2008, pp. 2873–2876.
- [Pac03] C. J. Paciorek. “Nonstationary Gaussian processes for regression and spatial modelling”. PhD thesis. Carnegie Mellon University, 2003.
- [PP+08] K. B. Petersen, M. S. Pedersen, et al. “The matrix cookbook”. In: *Technical University of Denmark* 7.15 (2008), p. 510.

- [RC15a] M. D. Risser and C. A. Calder. “Local likelihood estimation for covariance functions with spatially-varying parameters: the convoSPAT package for R”. In: *arXiv preprint arXiv:1507.08613* (2015).
- [RC15b] M. D. Risser and C. A. Calder. “Regression-based covariance functions for nonstationary spatial modeling”. In: *Environmetrics* 26.4 (2015), pp. 284–297.
- [RC15c] M. D. Risser and C. A. Calder. “Regression-based covariance functions for nonstationary spatial modeling”. In: *Environmetrics* 26.4 (2015), pp. 284–297.
- [RHK17a] S. Remes, M. Heinonen, and S. Kaski. “Non-stationary spectral kernels”. In: *arXiv preprint arXiv:1705.08736* (2017).
- [RHK17b] S. Remes, M. Heinonen, and S. Kaski. “Non-stationary spectral kernels”. In: *arXiv preprint arXiv:1705.08736* (2017).
- [Ris16] M. D. Risser. “Nonstationary spatial modeling, with emphasis on process convolution and covariate-driven approaches”. In: *arXiv preprint arXiv:1610.02447* (2016).
- [RLG18] D. Romero, D. Lee, and G. B. Giannakis. “Blind radio tomography”. In: *IEEE Transactions on Signal Processing* 66.8 (2018), pp. 2055–2069.
- [Saa12] Y. Saatçi. “Scalable inference for structured Gaussian process models”. PhD thesis. Citeseer, 2012.
- [SD16] A. Sinha and J. C. Duchi. “Learning kernels with random features”. In: *Advances in Neural Information Processing Systems* 29 (2016), pp. 1298–1306.
- [SG92] P. D. Sampson and P. Guttorp. “Nonparametric estimation of nonstationary spatial covariance structure”. In: *Journal of the American Statistical Association* 87.417 (1992), pp. 108–119.
- [SO03] A. M. Schmidt and A. O’Hagan. “Bayesian inference for nonstationary spatial covariance structure via spatial deformations”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65.3 (2003), pp. 743–758.
- [Ste05] M. L. Stein. “Nonstationary spatial covariance functions”. In: *Unpublished technical report* (2005).
- [SYT10] S. S. Szyszkowicz, H. Yanikomeroglu, and J. S. Thompson. “On the feasibility of wireless shadowing correlation models”. In: *IEEE Transactions on Vehicular Technology* 59.9 (2010), pp. 4222–4236.

- [VW99] F. Vivarelli and C. K. Williams. “Discovering hidden features with Gaussian processes regression”. In: *Advances in Neural Information Processing Systems*. 1999, pp. 613–619.
- [WA13] A. Wilson and R. Adams. “Gaussian process kernels for pattern discovery and extrapolation”. In: *International conference on machine learning*. PMLR. 2013, pp. 1067–1075.
- [Wik10] C. K. Wikle. “Low-rank representations for spatial processes”. In: *Handbook of spatial statistics* 107 (2010), p. 118.
- [WN15] A. Wilson and H. Nickisch. “Kernel interpolation for scalable structured Gaussian processes (KISS-GP)”. In: *International Conference on Machine Learning*. PMLR. 2015, pp. 1775–1784.
- [WP10] J. Wilson and N. Patwari. “Radio tomographic imaging with wireless networks”. In: *IEEE Transactions on Mobile Computing* 9.5 (2010), pp. 621–632.
- [WPV09] J. Wilson, N. Patwari, and F. G. Vasquez. “Regularization methods for radio tomographic imaging”. In: *2009 Virginia Tech Symposium on Wireless Personal Communications*. Citeseer. 2009.
- [WR06] C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*. Vol. 2. 3. MIT press Cambridge, MA, 2006.
- [Yag57] A. M. Yaglom. “Some classes of random fields in n-dimensional space, related to stationary random processes”. In: *Theory of Probability & Its Applications* 2.3 (1957), pp. 273–320.