



## THÈSE

*pour obtenir le grade de docteur délivré par :*

UNIVERSITÉ DE LILLE

Ecole doctorale : ENGSYS Sciences de l'ingénierie et des systèmes

Spécialité : Electronique, microélectronique, nanoélectronique et micro-ondes

Département Sciences et technologies de l'information et de la communication

Préparée à : Institut d'Électronique et de Microélectronique et de Nanotechnologie

INTITULÉE :

---

# **Caractérisation multidimensionnelle du canal de propagation et contribution de l'intelligence artificielle à la classification des scénarios de propagation pour les communications véhiculaires**

---

Réalisée par : **Nor El Islam Dahmouni**

Soutenue publiquement **le 21 Juin, 2024**

Directrice de thèse : **Mme. Martine Liénard, Professeur, Université de Lille**

Co-directeur : **M. Davy Gaillot, Maître de Conférences-HDR, Université de Lille**

Co-encadrant **M. Eric P. Simon, Maître de Conférences-HDR, Université de Lille**

### Membres du Jury :

M. Guillaume Andrieux, Professeur, Université de Nantes

**Rapporteur**

M. Guillaume Villemaud, Maître de Conférences HDR, INSA Lyon

**Rapporteur**

M. Pascal Pagani, Directeur de Recherche HDR, CEA-LETI Rennes (*Président*)

**Examineur**

Mme. Lina Mroueh, Professeur, ISEP Paris

**Examineur**



## Déclaration de l'auteur

Moi, Nor El Islam DAHMOUNI, déclare que cette thèse intitulée « Caractérisation multidimensionnelle du canal de propagation et contribution de l'intelligence artificielle à la classification des scénarios de propagation pour les communications véhiculaires » et le travail présenté dans celle-ci sont les miens. Je confirme que :

- Ce travail a été réalisé entièrement ou principalement pendant ma candidature à un diplôme de recherche dans cette université.
- Lorsque toute partie de cette thèse a été précédemment soumise pour un diplôme ou toute autre qualification dans cette université ou toute autre institution, cela a été clairement indiqué.
- Lorsque j'ai consulté les travaux publiés d'autres personnes, cela est toujours clairement attribué.
- Lorsque j'ai cité les travaux d'autres personnes, la source est toujours indiquée. À l'exception de telles citations, cette thèse est entièrement mon propre travail.
- J'ai reconnu toutes les principales sources d'aide.
- Lorsque la thèse est basée sur un travail que j'ai réalisé conjointement avec d'autres, j'ai clairement précisé ce qui a été fait par les autres et ce que j'ai contribué moi-même.

Signé :

---

Date : 21/06/2024

---



## *Remerciements*

Je souhaite exprimer ma profonde gratitude envers toutes les personnes qui ont contribué de manière significative à l'aboutissement de cette thèse.

Tout d'abord, je remercie chaleureusement Mme Martine Liénard, ma directrice de thèse, pour ses conseils éclairés, son soutien constant et son expertise précieuse qui ont grandement enrichi ce travail. Mes remerciements vont également à Davy Gaillot, mon co-directeur de thèse, ainsi qu'à Eric Simon, mon co-encadrant, pour leur collaboration étroite et leurs précieuses contributions tout au long de ce parcours de recherche.

En plus de mes superviseurs, je tiens à remercier les différents membres de mon jury de thèse, M. Guillaume Andrieux, professeur à l'Université de Nantes, M. Guillaume Villemaud, maître de conférences HDR à l'INSA Lyon, M. Pascal Pagani, directeur de recherche HDR au CEA-LETI Rennes et Mme Lina Mroueh, professeur à l'ISEP Paris. Merci d'avoir accepté d'examiner mon travail et pour vos précieuses observations et suggestions.

Je suis également reconnaissant envers toute l'équipe du laboratoire IEMN, groupe TELICE, pour son accueil chaleureux et son soutien logistique. Mes remerciements spéciaux vont à Pierre Laly et Gautier Delbarre pour leur assistance précieuse lors de mes mesures, ainsi qu'à Marwan Yusuf, ancien doctorant conjointement avec l'Université de Gand (Belgique), pour ses réunions fructueuses et les explications qu'il m'a fournies.

J'exprime ma profonde gratitude à ma femme Rayane pour son soutien inconditionnel, son amour et sa compréhension tout au long de cette aventure académique. Je remercie également mes parents, Fatiha et Saleh, ainsi que mes frères, Houssem et Bessam, pour leurs encouragements, qui ont été des sources de motivation essentielles. Leur soutien indéfectible m'a permis de persévérer dans les moments difficiles et de célébrer les succès. Ils ont toute ma reconnaissance pour leur patience, leurs encouragements et leur amour inconditionnel.

Enfin, gratitude à tous ceux qui, de près ou de loin, ont contribué à la réalisation de cette thèse. Votre soutien et votre engagement ont été d'une importance capitale pour moi, et je vous en suis sincèrement reconnaissant.

## Résumé

Au cours des dernières années, les réseaux mobiles ont considérablement amélioré leurs performances, notamment avec l'avènement de la 5G offrant des débits bien plus élevés que ceux de la 4G. Ces avancées sont attribuables aux progrès technologiques dans divers domaines tels que les systèmes antennaires intégrés et les composants électroniques alliant rapidité et faible consommation. La 5G inaugure une ère de services diversifiés incluant l'Internet des Objets (IoT) et les communications véhiculaires, de Véhicule à Infrastructure (V2I) et de Véhicule à Véhicule (V2V). Dans ce contexte, la 5G propose une classe de communication (URLLC) offrant une latence ultra-faible et une fiabilité accrue.

Pour accroître les performances de la 5G en termes d'efficacité spectrale, de nombre d'utilisateurs et de temps de latence, des recherches ont débuté en octobre 2022 dans le cadre de la Release 18 du 3GPP avec pour objectif d'explorer le potentiel que pourrait apporter l'Intelligence Artificielle (IA) et les modèles d'apprentissage de type Machine Learning (ML) dans l'interface radio. Ce projet vise ainsi à optimiser la technique d'estimation du canal et d'allocation des ressources, ainsi que la précision de positionnement. Le point commun à ces trois objectifs est le canal de propagation.

Cette thèse se focalise sur les liens V2I dans les milieux suburbains et dans les tunnels. L'originalité des travaux réside dans i) la caractérisation multidimensionnelle (temps-espace-fréquence-polarisation) du canal en milieu suburbain et tunnel grâce aux nombreuses campagnes de mesures effectuées avec le sondeur de canal MaMIMOSA, ii) l'analyse de la variabilité temporelle et spatiale du canal MIMO, iii) la création d'une base de données des paramètres du canal et iv) l'adaptation des modèles d'apprentissage automatique pour la détection des types de scénarios de propagation (LOS/NLOS). Notons que l'état de l'art ne concerne que les milieux indoor ou l'environnement urbain avec des liens satellitaires.

Ces travaux ont permis de constituer une base de données expérimentales obtenues à partir de l'exploitation de mesures faites en site suburbain et en tunnel, respectivement sur  $\sim 100k$  et  $\sim 200k$  matrices de canal V2I. Le modèle ML pour la classification des scénarios de propagation, soit en visibilité directe (LOS), soit sans visibilité directe (NLOS), affiche des performances de prédiction de 97,7%; le nombre de points en LOS et NLOS est équilibré pour alimenter le modèle.

## *Abstract*

Over the past few years, mobile networks have significantly improved their performance, especially with the advent of 5G offering much higher data rate than 4G. These advancements are attributable to technological progress in various areas such as integrated antenna systems and electronic components that combine high response time with low power consumption. 5G ushers in an era of diversified services including the Internet of Things (IoT) and vehicular communications, from Vehicle-to-Infrastructure (V2I) and Vehicle-to-Vehicle (V2V). In this context, 5G introduces a class of communication (URLLC) offering ultra-low latency and increased reliability.

To enhance the performance of 5G in terms of spectral efficiency, the number of users, and latency time, research began in October 2022 as part of the 3GPP Release 18 with the goal of exploring the potential that Artificial Intelligence (AI) and Machine Learning (ML) models could bring to the radio interface. This project aims to optimize the channel estimation technique and resource allocation, as well as positioning accuracy.

As part of these studies, this thesis focuses on V2I links in suburban environments and tunnels. Its objective is to create an experimental database of the most relevant parameters of the propagation channel, which will then allow for the training of ML models. On this point, the state of the art only concerns indoor environments or the urban environment with satellite links.

The originality of the work lies in : i) measurement campaigns carried out in realistic environments and scenarios, ii) the multidimensional characterization (time-space-frequency-polarization) of the propagation channel considering MIMO (Multiple Input Multiple Output-MIMO) antenna networks at the transmission and reception, iii) the analysis of the temporal and spatial variability of MIMO channels, and, iv) the adaptation of machine learning models for the classification of propagation scenarios.

These works have led to the creation of an experimental database obtained from the exploitation of measurements made in suburban sites and tunnels, and respectively on approximately 100k and 200k V2I channel matrices. The ML model for the classification of propagation scenarios, either in line of sight (LOS) or Non line of sight (NLOS), shows prediction performances of 97.7%; the number of points in LOS and NLOS is balanced to feed the model.



# Table des matières

<b>Déclaration de l’auteur</b>	<b>iii</b>
<b>Remerciements</b>	<b>v</b>
<b>Résumé</b>	<b>vi</b>
<b>Abstract</b>	<b>vii</b>
<b>Liste des publications</b>	<b>xxiv</b>
<b>Introduction générale</b>	<b>1</b>
<b>1 État de l’art</b>	<b>5</b>
1.1 Introduction . . . . .	5
1.2 Caractérisation multidimensionnelle du canal . . . . .	8
1.2.1 Description stochastique d’un canal de propagation SISO . . . . .	11
1.2.2 Estimation du temps de stationnarité d’un canal de propagation SISO . . . . .	11
1.2.3 Estimation de la stationnarité spatiale pour un canal SU-MIMO . . . . .	15
1.2.4 La consistance spatiale pour un canal MU-MIMO . . . . .	17
1.2.5 Estimation des paramètres de canal . . . . .	18
1.2.6 Applications de l’hypothèse WSSUS . . . . .	21
La simulation de canal et l’évaluation des performances . . . . .	21
Capacité Ergodique . . . . .	21
Degrés de diversité dans le canal . . . . .	22
1.3 Utilisation de l’intelligence artificielle dans l’estimation du canal et l’allocation de ressources . . . . .	23
1.3.1 L’apprentissage automatique (ML) et l’apprentissage profond (DL) . . . . .	23
1.3.2 Application de ML et DL à la classification LOS-NLOS . . . . .	24
Introduction . . . . .	24
État de l’art de la classification LOS-NLOS à l’aide de l’IA . . . . .	25
Conclusion . . . . .	30
1.3.3 Application de IA à Estimation du canal . . . . .	32
Introduction . . . . .	32
État de l’art de l’estimation du canal à l’aide de l’IA . . . . .	32

Conclusion	38
1.4 Conclusion du chapitre	38
<b>2 Sondage de canal MIMO</b>	<b>41</b>
2.1 Introduction	41
2.2 Architecture du sondeur	41
2.3 Les réseaux d'antennes	42
2.4 Description des campagnes de mesure et analyse bande étroite	45
2.4.1 Le tunnel Beveren à Anvers	45
2.4.2 Analyse bande étroite	45
Qualité du signal mesuré	45
Affaiblissement de propagation	46
Discrimination de polarisation croisée (XPD)	50
2.4.3 Environnement suburbain cité scientifique-université Lille	51
2.5 Conclusion	56
<b>3 Caractérisation multidimensionnelle du canal MIMO</b>	<b>59</b>
3.1 Introduction	59
3.2 Caractérisation de la variabilité temporelle en canal suburbain	59
3.2.1 Description des expérimentations	59
3.2.2 Évaluation du temps de stationnarité $T_s$	60
Calcul du GLSF discret	60
Estimation de $T_s$ à partir du CCF	61
Estimation de $T_s$ à partir de la colinéarité	61
Résultats et discussion	62
3.2.3 Évaluation des caractéristiques du canal radio large bande	63
Temps de cohérence et bande de cohérence	64
Les profils de puissance retard - Doppler	64
Degrés de diversité dans le canal	67
Coefficient $N_i$	68
Caractéristiques des directions de départ des rayons du réseau d'émission	69
3.3 Caractérisation de la variabilité temporelle du canal en tunnel	72
3.3.1 Évaluation du temps de stationnarité $T_s$	72
Résultats et discussions	72
Influence de la fréquence sur $T_s$	74
3.3.2 Évaluation des caractéristiques du canal radio large bande	75
Les profils de puissance retard-Doppler	76
L'étalement des retards et l'étalement Doppler	77
Corrélation entre les antennes	78
Relation entre le temps de cohérence $T_c$ et l'étalement Doppler $\sigma_v$	84
3.4 Conclusion	85

<b>4</b>	<b>Apport du ML sur la classification LOS NLOS</b>	<b>87</b>
4.1	Introduction	87
4.2	Processus de Classification LOS/NLOS	88
4.2.1	Collecte des données	88
4.2.2	Pré-traitement des données	88
4.2.3	Dérivation de caractéristiques	89
	Extraction de caractéristiques	89
	Normalisation des Caractéristiques	101
4.2.4	Sélection des Caractéristiques	102
	Sélection de caractéristiques par filtrage	103
	Sélection de caractéristiques par Enveloppement	105
	Sélection des caractéristiques - Méthode Hybride	105
4.3	Algorithmes utilisés pour la classification LOS-NLOS	107
	Algorithme kNN - k-Nearest Neighbors	107
	Algorithme SVM	109
	Les trois ensembles de données en apprentissage automatique	110
4.4	Entraîner et valider le modèle d'apprentissage	111
4.5	Conclusion	127
	<b>Conclusion générale et perspectives</b>	<b>129</b>
<b>A</b>	<b>Stationnarité d'un canal de propagation</b>	<b>133</b>
A.1	Stationnarité d'un canal SISO	133
	A.1.1 Description systémique d'un canal de propagation SISO	133
	A.1.2 Description stochastique d'un canal de propagation SISO	134
A.2	Illustration de la colinéarité	138
<b>B</b>	<b>Caractéristiques et Algorithmes d'apprentissage Automatique</b>	<b>139</b>
B.1	Calculs derrière l'extraction de Caractéristiques	139
	B.1.1 Nombre de pics par PDP	139
	B.1.2 Nombre et largeur médiane des évanouissements fréquents N.E.F et L.M.E.F	140
	B.1.3 Groupe de retards $g_D$	141
B.2	k-Nearest Neighbors (kNN)	141
	B.2.1 Différentes distances pour kNN	141
	B.2.2 Problème d'optimisation dans kNN	143
B.3	Machine à Vecteurs de Support (SVM)	143
	B.3.1 Description mathématique du SVM	143
	B.3.2 Problème d'optimisation dans SVM	143
	B.3.3 Concept d'hyperplans dans SVM	143
	B.3.4 Hyperparamètres dans SVM	144
B.4	Sélection des Caractéristiques	145
	B.4.1 t-Test	145



# Table des figures

1.1	Comparaison des technologies 2G, 3G, 4G et 5G . . . . .	6
1.2	Cas d'usage 5G . . . . .	6
1.3	Différents types de scénarios de déploiement des réseaux MIMO massifs. . . . .	10
1.4	Illustration des concepts de WSS-US, WSS-US local et Non-WSSUS appliqués à un canal de propagation SISO . . . . .	12
1.5	Illustration des GLSF dans le domaine temporel. . . . .	13
1.6	Distance de la matrice de corrélation spatiale MU-MIMO . . . . .	17
1.7	Région de cohérence et de stationnarité. . . . .	19
1.8	L'intelligence artificielle et ses sous-domaines . . . . .	23
1.9	Classification LOS NLOS à l'aide du ML/DL . . . . .	25
1.10	Apprentissage profond pour l'estimation du canal . . . . .	33
1.11	Étapes de l'algorithme CNN. . . . .	33
1.12	Étapes de la méthode proposée. . . . .	35
1.13	Étapes de la méthodes proposé LSTM MLP . . . . .	35
1.14	Structure de la méthode proposée . . . . .	37
2.1	Sondeur de canal : émetteur fixe (a) et récepteur mobile (b) . . . . .	42
2.2	Structure de la trame des séquences transmis . . . . .	42
2.3	F=1,35 GHz : polarisation V, H, D, A du patch (c), réseau d'émission à 4 ports de sortie par polarisation (a) et réseau de réception à 4 ports d'entrée par polarisation en tunnel (b). . . . .	43
2.4	5,89 GHz : réseau de 64 antennes illuminant le tunnel (a), la réception se faisant sur 8 antennes omnidirectionnelles dans le plan azimut, placées sur le toit du véhicule (b). . . . .	45
2.5	(a) Entrée du tunnel Beveren (b) Réseau d'émission à la sortie du tunnel à gauche . . . . .	46
2.6	Puissance moyenne relative du signal et puissance de bruit (Configuration 2) en fonction de la distance Tx-Rx. Polarisations VV et VH. F=1,35 GHz . . . . .	47
2.7	SNR en dB en fonction de la distance Tx-Rx. Polarisations VV et VH. F=1,35 GHz . . . . .	47
2.8	Pathloss parcours réel (Configuration 2) et théorique . . . . .	48
2.9	PL trafic important avec les configurations 1, 3 et 6, trafic important polarisations en VV et PL théorique tunnel vide . . . . .	48

2.10	PL Trafic important configuration 3 avec différentes polarisations et PL théorique tunnel vide . . . . .	49
2.11	ECDF du PL trafic important et faible avec différentes polarisations . . . . .	50
2.12	ECDF XPDs pour (a) un trafic faible et (b) pour un trafic important, avec les différentes polarisations . . . . .	51
2.13	Sondeur MaMIMOSA Tx et Rx, le réseau d'antennes massif en émission et un réseau ULA en réception. . . . .	52
2.14	Vue de dessus de la campagne de mesure pour la stationnarité dans un environnement suburbain . . . . .	52
2.15	SNR en fonction de l'antenne et la distance Tx-Rx (Configuration 30) . . . . .	53
2.16	Vue de dessus de l'environnement suburbain des campagnes de mesure . . . . .	54
2.17	Puissance signal et du bruit avec le SNR en fonction de la distance pour un scénario suburbain LOS et NLOS, configuration 27 . . . . .	55
2.18	Cartographie du SNR pour la première antenne de réception et 64 antennes d'émission à 40 m LOS (a) et 150 m NLOS (b) . . . . .	55
3.1	$T_s$ en fonction de la distance Tx-Rx obtenue par les méthodes CCF (bleu) et de colinéarité (rouge). La valeur médiane est représentée par une ligne continue et les 25ème/75ème percentiles par une ligne pointillée. . . . .	62
3.2	Corrélation spatiale du réseau URA (8,8) calculée à partir (a) des valeurs $T_s$ variant dans le temps et (b) des fonctions de transfert pour l'antenne #46 (flèche noire). . . . .	63
3.3	$\rho_V$ et de $\rho_H$ moyens en fonction de l'indice d'antenne calculés à partir de $T_s$ ( <b>bleu et jaune</b> ) et à partir des fonctions de transfert ( <b>rouge et violet</b> ). . . . .	64
3.4	Distribution spatiale en boxplot de $T_c$ (a) en bleu et $B_c$ (b) en rouge en fonction du temps. . . . .	65
3.5	PDP (a) et DPS (b) en fonction de la distance Tx-Rx. . . . .	66
3.6	Étalement des retards médian (bleu) et étalement de Doppler médian (rouge). . . . .	66
3.7	Régions identifiées sur le parcours. . . . .	67
3.8	Ordre de la diversité efficace en temporel ( $d_t$ ), fréquentiel ( $d_f$ ), en retard ( $d_\tau$ ) et en Doppler ( $d_\nu$ ) sur les 64 antennes en fonction de la distance . . . . .	68
3.9	Nombre de réalisations indépendantes i.i.d. offerts par le canal $N_i$ en fonction d'antenne et de la distance . . . . .	69
3.10	Direction de départ des rayons : Valeur moyenne . . . . .	70
3.11	Direction de départ : Étalement angulaire. . . . .	70
3.12	Temps de stationnarité en fonction de la distance Tx-Rx pour les 4 configurations de polarisation. . . . .	73

3.13 Temps de stationnarité en VV à 1,35 GHz et à 5,89 GHz configurations (2 et 7 trafic faible). . . . .	75
3.14 ECDF du $T_s$ en VV à 1,35 GHz et à 5,89 GHz. . . . .	75
3.15 Profil de puissance retards PDP en VV (a) et en HH (b) variant en fonction de la distance, moyennés sur toutes les antennes Tx, configuration (2) sans trafic. . . . .	76
3.16 Spectres de puissance Doppler (DSP) en VV (a) et en HH (b), configuration (2) en fonction de la distance, moyennés sur toutes les antennes Tx. . . . .	77
3.17 Étalement des retards et étalement de Doppler, configuration (2) à faible trafic . . . . .	78
3.18 Profil de puissance retard (PDP) normalisé en VV (a) et en HH (b) en fonction de la distance, moyenné sur toutes les antennes Tx, Configuration (2). . . . .	79
3.19 Profil de puissance Doppler normalisé en VV (a) et en HH (b) en fonction de la distance, moyenné sur toutes les antennes Tx, Configuration (2). . . . .	80
3.20 Coefficient de corrélation entre les antennes d'émission en VV expérimentale (configuration 6) et théorique avec les écarts $3\lambda/2$ et $3\lambda$ . . . . .	80
3.21 ECDF des coefficients de corrélation entre les antennes d'émission en VV expérimentale (configuration 6) et théorique avec les écarts $3\lambda/2$ et $3\lambda$ . . . . .	81
3.22 Coefficient de corrélation entre les antennes d'émission pour les 4 polarisations expérimentales (configuration 6) avec un écart de $3\lambda/2$ . . . . .	81
3.23 Coefficient de corrélation entre les antennes de la réception en VV expérimentale (Configuration 6) et théorique avec les écart $3\lambda/2$ et $3\lambda$ . . . . .	82
3.24 ECDF des coefficients de corrélation entre les antennes de la réception en VV expérimentale (Configuration 6) et théorique avec les écarts $3\lambda/2$ et $3\lambda$ . . . . .	82
3.25 Coefficient de corrélation entre les antennes de la réception pour les 4 polarisations expérimentales (configuration 6) avec un écart de $3\lambda/2$ . . . . .	83
3.26 Temps de cohérence ( $T_c$ ) avec autocorrélation et étalement Doppler en VV (à droite) associé à la distance de cohérence $D_c$ avec l'autocorrélation et l'étalement Doppler (à gauche) pour la configuration (2). . . . .	84
3.27 ECDF du facteur $k_{dopp}$ 5,89 GHz suburbain et 1,35 GHz. . . . .	85
4.1 Étalement des retards $\tau_{rms}$ pour les scénarios LOS/NLOS (a) et leur ECDF (b) . . . . .	90
4.2 Étalement des retards moyenne $\tau_{MOY}$ pour les scénarios LOS/NLOS (a) et leur ECDF (b) . . . . .	90
4.3 Identification des pics pour un PDP LOS (a) NLOS (b) . . . . .	91

4.4	Nombre de pics pour les scénarios LOS/NLOS (a) et leur ECDF (b) . . .	91
4.5	Puissance de signal pour les scénarios LOS/NLOS (a) et leur ECDF (b) .	92
4.6	Différence de puissance $d_p$ pour les scénarios LOS/NLOS (a) et leur ECDF (b) . . . . .	92
4.7	Moyenne (a), et écart type (b), asymétrie (c) et Kurtosis (d) . . . . .	93
4.8	Facteur de rice $K_{rice}$ pour les deux scénarios LOS/NLOS (a) et leur ECDF (b) . . . . .	94
4.9	Valeurs des pentes des PDP pour les deux scénarios LOS/NLOS (a) et leur ECDF (b) . . . . .	94
4.10	Bande de cohérence à 0.7 (a) et leur ECDF (b) pour les deux scénarios LOS/NLOS . . . . .	95
4.11	Évanouissements fréquentiels et leur largeur dans les fonctions de transfert pour les deux scénarios de propagation LOS/NLOS . . . . .	95
4.12	Nombre et largeur médiane des évanouissements fréquentiels pour les deux scénarios de propagation LOS/NLOS . . . . .	96
4.13	Retard de groupe pour un échantillon LOS (a) et NLOS (b) . . . . .	97
4.14	Moyenne de la valeur absolu $g_D$ LOS/NLOS (a) et leur ECDF (b) . . . . .	97
4.15	SNR en (dB) LOS/NLOS (a) et leur ECDF (b) . . . . .	98
4.16	Méthode de calcul des coefficients de corrélation $\rho_{Tx}$ en ULA V et ULA H et $\rho_{Rx}$ . . . . .	99
4.17	Coefficients de corrélation $\rho_{Rx}$ LOS/NLOS (a) et leur ECDF (b), $\rho_{Tx}$ ULA H LOS/NLOS (c) et leur ECDF (d) et $\rho_{Tx}$ ULA V LOS/NLOS (e) et leur ECDF (f) . . . . .	100
4.18	Histogramme des angles de départ en azimut (AoDa) avec leur étalement angulaire (b) et en élévation (AoDe) avec leur étalement angulaire (d) . . . . .	101
4.19	Représentation de la méthode Intégrée (Embedded) . . . . .	103
4.20	Représentation de la méthode de filtrage . . . . .	103
4.21	Représentation de l'approche par enveloppement (wrapper) . . . . .	105
4.22	Représentation de la nouvelle méthode Hybride . . . . .	106
4.23	Fonctionnement du kNN ( <a href="https://medium.com/@rndayala">https://medium.com/@rndayala</a> ) . . . . .	108
4.24	Fonctionnement du SVM ( <a href="https://medium.com/@analytisarora">analytisarora.com</a> ) . . . . .	109
4.25	(Crédit : © 2017, Haydar Ali Ismail, Medium.com) La projection des données dans une espace de plus grande dimension via une fonction noyau pour les séparer linéairement. . . . .	110
4.26	Division des données en ensembles d'entraînement, de validation et de test . . . . .	111
4.27	Validation croisée pour un $k_{fold} = 5$ . . . . .	112
4.28	Visualisation des caractéristiques . . . . .	112
4.29	Matrice de confusion . . . . .	114
4.30	Matrice de confusion pour les données de validation et de test . . . . .	114
4.31	Variance proportionnelle des caractéristiques . . . . .	118

4.32	Résultat de t-Test sur kNN (a) et SVM (b)	119
4.33	Optimisation des hyperparamètres pour kNN	122
4.34	Optimisation des hyperparamètres pour SVM	123
4.35	Matrice de confusion pour les données de test	124
A.1	Illustration des fonctions de Bello	133
A.2	Illustration des fonctions d'autocorrélation	135
A.3	Illustration des fonctions de corrélation pour un canal WSSUS	136
A.4	Relation entre les fonctions de système, les fonctions de corrélation et les paramètres du canal WSSUS	137
A.5	Calcul de la collinearité pour une trame de 5 seconde (a) sans seuil et (b) avec un seuil de 0.9	138
B.1	Exemple d'utilisation de la fonction <i>findpeaks</i> pour deux PDP en LOS et NLOS	139
B.2	Identification du nombre d'évanouissements et de leur largeur pour un scénario LOS (a) et un scénario NLOS (b)	140
B.3	Profils de puissance des fonctions de transfert pour un scénario LOS (a) et un scénario NLOS (b)	140



# Liste des tableaux

1.1	Cas d'usage V2X avec les types de cas d'utilisation 5G, les infrastructures, les technologies et les normes correspondantes [24-33]	7
1.2	Comparaison des Caractéristiques Techniques entre ITS-G5 et C-V2X dans les Communications Véhiculaires	7
1.3	Valeurs de $I$ et $J$ proposées dans la littérature	13
1.4	Synthèse sur la stationnarité. Pour les scénarios V2V, le Tx et le Rx se déplacent dans la même direction. Équations (1.14) et (1.15), [47] $L=25$ ; [71] $L=16$ ; [40] $L=$	16
1.5	Distance de corrélation pour la consistance spatiale pour les environnements suivants tels que définis dans le 3GPP - RMa : environnement rural, UMi : canyon urbain, UMa : Urbain et O2I : de l'extérieure à l'intérieur	18
1.6	Résultats des algorithmes.	29
1.7	Résumé des résultats des algorithmes pour la classification LOS/NLOS avec les précisions et les caractéristiques utilisées.	31
1.8	Résumé des performances des algorithmes d'apprentissage profond DL dans l'estimation du canal	39
2.1	Distance entre les antennes Tx et Rx à 5,89 GHz et 1,35 GHz	43
2.2	Tableau de synthèse des configurations utilisées. (a) Stationnarité, (b) Doppler, (c) Dépolarisation, (d) Corrélacion, (e) Pathloss.	44
2.3	Comparaison de l'influence du trafic à 1,35 GHz/5,89GHz sur le Pathloss $d_0=80$ m	49
2.4	Tableau de synthèse des configurations utilisées en suburbain	57
3.1	Configuration des paramètres de MaMIMOSA.	60
3.2	Quelques statistiques sur les différents ordres de diversité	68
3.3	$N_i$ , nombre de réalisations indépendantes i.i.d. offerts par le canal	69
3.4	Tableau de synthèse des paramètres du canal dans un environnement suburbain	71
3.5	Caractéristiques des directions de départ des rayons du réseau d'émission	71
3.6	Tableau des configurations et valeurs de $(I, J)$ associées	72
3.7	Statistiques du $T_s$ pour différentes polarisations et configurations	74

3.8	Coefficients de corrélation $\rho$ , configuration (2- trafic faible) en bleu, configuration (3- trafic important) en rouge . . . . .	74
3.9	Médiane de corrélation entre Tx et Rx avec différents espacements . . . . .	83
3.10	Tableau des valeurs pour différents scénarios . . . . .	86
4.1	Caractéristiques du canal candidates pour le ML . . . . .	102
4.2	Méthodes de normalisation, standardisation et mise à l'échelle pour chaque caractéristique $x$ [135][136]. . . . .	102
4.3	Trois sous-ensembles des caractéristiques de la variabilité du canal . . . . .	106
4.4	Performances de kNN et SVM sur les ensembles de validation et de test avec toutes les caractéristiques. . . . .	113
4.5	Hyperparamètres pour la classification binaire avec kNN et SVM. . . . .	115
4.6	Méthodes de sélection de caractéristiques . . . . .	116
4.7	Caractéristiques sélectionnées via la technique du seuil de variance avec les résultats de kNN et SVM. . . . .	117
4.8	Caractéristiques sélectionnées via la technique du t-Test avec les résultats de kNN et SVM . . . . .	119
4.9	Caractéristiques sélectionnées par la méthode de la sélection séquentielle pour kNN et SVM . . . . .	120
4.10	Caractéristiques sélectionnées par la méthode hybride pour kNN et SVM	121
4.11	Comparaison des valeurs de la fonction objectif observées et estimées pour les algorithmes kNN et SVM . . . . .	122
4.12	Hyperparamètres pour la classification binaire avec kNN et SVM. . . . .	123
4.13	Caractéristiques sélectionnées par la méthode hybride pour kNN et SVM après l'optimisation des hyperparamètres . . . . .	124
4.14	Comparaison des résultats avec la littérature. . . . .	126

# Liste des Abréviations

2G	Deuxième génération
3D	Trois Dimensions
3GPP	3rd Generation Partnership Project
4G	Quatrième génération
5G	Cinquième génération
6G	Sixième génération
A	Polarisation croisée $-45^\circ$
AoA	Angle of Arrival
AoDa	Angle of Departure azimuth
AoDe	Angle of Departure elevation
AoD	Angle of Departure
ANN	Artificial Neural Network
AS	Angular Spread
CCF	Channel Correlation Function
CDL	Clustered Delay Line
C-V2X	Cellular Vehicle-to-Everything
CSI	Channel State Information
CMD	Correlation Matrix Distance
CNN	Convolutional Neural Network
D	Polarisation croisée $+45^\circ$
DL	Deep Learning
DNN	Deep Neural Network
DSRC	Dedicated Short Range Communications
DMRS	Demodulation Reference Signal
DPS	Doppler Power Spectrum
DPSS	Discrete Prolate Spheroidal Sequences
DPA	Data Pilot Aided
ECDF	Empirical Cumulative Distribution Function
ETSI	European Telecommunications Standards Institute
EVM	Error Vector Magnitude
FFT	Fast Fourier Transform
FCNN	Fully Connected Neural Network
FN	False Negative
FP	False Positive
GMM	Gaussian Mixture Model

GP	Gaussian Process
GLSF	Generalized Local Scattering Function
GNSS	Global Navigation Satellite System
H	Polarisation horizontale
HAD	Hybrid Analog-Digital
HSR	High-Speed Rail
IA	Intelligence Artificielle
IEEE	Institute of Electrical and Electronics Engineers
IEMN	Institut d'Électronique, de Microélectronique et de Nanotechnologie
INTEC	Information Technology Department
IoT	Internet des objets
ITS	Intelligent Transportation System
ITS-G5	Intelligent Transportation Systems-G5
kNN	k-Nearest Neighbors
KPI	Key Performance Indicator
LSF	Local Scattering Function
LS	Least Squares
LSMMSE	Least Squares Minimum Mean Square Error
LSTM	Long Short-Term Memory
LTE	Long-Term Evolution
LTE-V2X	Long-Term Evolution Vehicle-to-Everything
L.M.E.F	Largeur Médiane d'Évanouissement Fréquentielle
mMIMO	Massive MIMO
ML	Machine Learning
mMTC	massive Machine-Type Communications
MIMO	Multiple-Input Multiple-Output
MLP	Multi-Layer Perceptron
MU-MIMO	Multi-User MIMO
mMTC	Massive Machine Type Communications
MSE	Mean Squared Error
N.E.F	Nombre d'évanouissement fréquentielle
NB	Naive Bayes
NN	Neural Network
NLOS	Non-Line of Sight
NMSE	Normalized Mean Square Error
NR	New Radio
NR-V2X	New Radio Vehicle-to-Everything
OFDM	Orthogonal Frequency Division Multiplexing
PAS	Power Angular Spectrum
PCA	Principal Component Analysis
PDP	Power Delay Profile
PL	Path Loss

PHY	Physical Layer
QoS	Quality of Service
RAN	Radio Access Network
RNN	Recurrent Neural Network
RMa	Rural Macro
RMa	Rural Macro
RF	Random Forest
RMS	Root Mean Square
RSS	Received Signal Strength
Rx	Receiver
SF	Shadow Fading
SBL	Sparse Bayesian Learning
SoN	Self-Organizing Network
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
SU-MIMO	Single User Multiple Input Multiple Output
SVM	Support Vector Machine
TDD	Time Division Duplex
TEB	Taux d'Erreur Binnaire
TEBL	Taux d'Erreur par Bloc
TNN	Tensor Neural Network
Tx	Transmission
UE	User Equipment
ULA	Uniform Linear Array
URA	Uniform Rectangular Array
URLLC	Ultra-Reliable Low Latency Communication
UWB	Ultra-Wideband
V	Polarisation verticale
V2G	Vehicle-to-Grid
V2I	Vehicle-to-Infrastructure
V2N	Vehicle-to-Network
V2P	Vehicle-to-Pedestrian
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-Everything
VoIP	Voix sur IP
WSSUS	Wide-Sense Stationary Uncorrelated Scattering
XPD	Cross-Polarization Discrimination
ZF	Zero Forcing

## Liste des publications

- **Conference** : Dahmouni, N. E. I., Mallik, M., Laly, P., Egea-Lopez, E., & Gaillot, D. P. (2024, March). A Comparative Analysis of Single-Input Single-Output Stationarity Time and Pathloss in Suburban Environments for Vehicle-to-Infrastructure Channel Using Veneris Ray-Tracing and Real Data. In The Thirteenth International Conference on Advances in Vehicular Systems, Technologies and Applications VEHICULAR 2024 (p. 6).
- **Journal** : Dahmouni, N. E. I., Laly, P., Yusuf, M., Delbarre, G., Liénard, M., Simon, E. P., & Gaillot, D. P. (2022). On the Stationarity Time of a Vehicle-to-Infrastructure Massive Radio Channel in a Line-of-Sight Suburban Environment. *Sensors*, 22(21), 8420.
- **Conference** : Gaillot, D. P., Laly, P., Dahmouni, N., Delbarre, G., Van den Bossche, M., Vermeeren, G., ... & Liénard, M. (2021, March). Measurement of the V2I Massive Radio Channel with the MaMIMOSA Sounder in a Suburban Environment. In 2021 15th European Conference on Antennas and Propagation (EuCAP) (pp. 1-4). IEEE.

## Articles en cours de rédaction

- **Journal** : Machine Learning-Based Classification of LOS and NLOS Scenarios for 5G Massive MIMO Vehicle-to-Infrastructure Communication.
- **Journal** : Comparative Analysis of Stationarity Time and Pathloss in Suburban Massive MIMO Channels for V2I Communication : Ray-Tracing-Real Data
- **Conférence** : Channel Matrix Distance-Based Estimation of Spatial Stationarity in Suburban Massive MIMO V2I Communication : A Comparative Analysis of URA and ULA Configurations

## Autres

- **Séminaire** : Présentation des résultats du sondeur de canal MaMIMOSA lors d'un séminaire du projet SMARTIES dans le cadre de ELSAT2020, 2021.
- **Doctoriales** : Prix de la communication pour un projet sur l'identification et le traitement ciblé et précis des nuisibles dans les vignes à l'aide de l'apprentissage profond (Deep Learning), Doctoriales Hauts-de-France, 2022.





# Introduction générale

Ces dernières années, les performances et en particulier les débits théoriques des réseaux mobiles ont connu un accroissement remarquable. En effet, si la 4G propose des débits maximums en liaison descendante de 300 Mb/s, la 5G a multiplié cette capacité par 7. Cette augmentation des performances entre 4G et 5G est attribuable, entre autres, aux avancées technologiques dans les domaines de l'électronique des processeurs de traitement du signal, des systèmes antennaires et composants pour les bandes en dessous (FR1) et au-delà (FR2) de 6 GHz. Ces progrès ne se limitent pas seulement à l'amélioration des capacités de transmission de données, mais marquent également le début d'une ère de services diversifiés, incluant des applications telles que l'Internet des Objets (IoT), la VoIP, la communication véhiculaire LTE-V2X, etc. Le lancement de la Release 16 rédigé par le 3GPP (3rd Generation Partnership Project) a marqué une étape importante, ouvrant la voie à la 5G New Radio (NR) pour les communications véhiculaires avec le New Radio C-V2X ou NR-V2X. Cette mise à jour a introduit de nouvelles fonctionnalités visant à prendre en charge des cas d'utilisation avancés tels que le Véhicule à Infrastructure (V2I), le Véhicule à Véhicule (V2V) et le Véhicule à Piéton (V2P). Ces améliorations comprennent des gains en fiabilité, une réduction de la latence, une précision accrue dans le positionnement et une adaptation à la mobilité à haute vitesse, des éléments cruciaux pour répondre aux contraintes des systèmes de transport intelligent (ITS). Dans ce contexte, la 5G propose, entre autres, une classe de communication URLLC (Ultra Reliable Low Latency Communication) permettant des temps de latence de l'ordre de la milliseconde et une fiabilité de communication accrue, particulièrement bien adaptée aux communications V2X (Vehicle-to-Everything). Par exemple, alors qu'en 4G, la garantie de fiabilité se limitait à vérifier que 256 bits pouvaient être transmis avec un taux de réussite de 99,9% en moins de 50 millisecondes, la 5G en mode URLLC réduit ce temps de latence à 3 millisecondes avec un taux de fiabilité de 99,9999%. Ces spécifications ont été intégrées dans la release 15 du 3GPP.

Pour accroître ces performances et réduire la complexité des algorithmes intégrés dans la couche physique (PHY) du réseau d'accès radio (RAN) de la 5G, des recherches développées dans le cadre de la release 18 du 3GPP [1][2] sont actuellement en cours pour analyser le potentiel que pourrait apporter l'intelligence artificielle (IA) et plus particulièrement les algorithmes d'apprentissage de type Machine Learning (ML) dans l'interface radio. Les 3 objectifs de ce projet incluent

l'amélioration de l'efficacité de l'estimation du canal (CSI) en fréquence avec une éventuelle prédiction en temps, l'optimisation de la stratégie d'allocation des ressources en temps-fréquence espace, en particulier la prédiction des allocations de faisceaux par utilisateur, et l'amélioration du positionnement direct par l'IA/ML. Ce projet envisage également la proposition de nouveaux indicateurs de performance (Key Performance Indicators - KPI) liés à la complexité algorithmique et au coût calculatoire, ainsi que le développement d'une méthodologie pour collecter des données et rendre disponible une base de données à la fois à la station de base et au terminal UE (User Equipment). Les conclusions de ce projet devront mettre en lumière les performances apportées par les prédictions de l'IA/ML par rapport aux systèmes actuels. Notons que le ML a récemment été introduit au niveau de la couche réseau pour assurer l'organisation autonome des réseaux (Self organized network-SON) [3] dont bénéficient les points d'accès Wifi, small cell ou encore hot spot en 4G/5G.

L'objectif de la thèse s'inscrit dans le cadre général de mise en place d'une base de données pertinente et du développement de modèles prédictifs d'estimation du type de scénario de propagation de type V2X. Il s'agit de briques de base susceptibles d'être utilisées dans les 3 axes du projet de recherche du 3GPP cité précédemment, le point commun à ces axes étant le canal de propagation.

Cette thèse s'articule ainsi autour de plusieurs objectifs : i) la caractérisation multidimensionnelle expérimentale du canal de propagation en lien avec les communications V2I, les environnements étudiés étant le milieu suburbain et le tunnel, ii) la création de la base de données expérimentale et iii) l'adaptation d'un modèle d'apprentissage machine (ML) pour la classification LOS/NLOS dans des scénarios de communication véhiculaire 5G de type V2I à partir de la base de données expérimentale. L'originalité des travaux concerne :

- La caractérisation expérimentale multidimensionnelle du canal de propagation MIMO (temps-fréquence-espace-polarisation) à l'aide d'un sondeur de canal temps-réel.
- L'analyse de la variabilité temporelle et spatiale de canaux de propagation MIMO en milieu suburbain et tunnel. La rapidité d'acquisition du sondeur a permis de développer une étude originale sur l'estimation des zones sur lesquelles les caractéristiques statistiques d'ordre 2 du canal demeurent invariantes en temps et fréquence. Dans la suite du document, ces zones sont nommées zones de stationnarité locale. Dans un contexte de communication MIMO avec des réseaux d'antennes à la station de base de grande taille en termes de nombre d'éléments, l'étude des zones de stationnarité est complétée par l'analyse de la variabilité spatiale sur le réseau. En caractérisant la

---

variabilité temporelle et spatiale des canaux de propagation MIMO dans des conditions de mobilité élevée et en identifiant les scénarios de propagation LOS/NLOS, il est possible de parvenir à la modélisation d'un canal de propagation et ultérieurement au développement de systèmes de communication véhiculaire plus résilients et performants.

- L'extraction de nouvelles caractéristiques pertinentes du canal dont le retard moyen du groupe, le nombre d'évanouissements fréquentiels et le nombre de pics par réponses impulsionnelles. À ces caractéristiques s'ajoutent celles déduites du MIMO telles que les corrélations entre les antennes d'émission et de réception.
- Développement d'une méthode de sélection hybride des caractéristiques du canal et adaptation d'un modèle d'apprentissage automatique pour la classification des scénarios V2I LOS/NLOS à partir de la base de données expérimentale.

Cette thèse s'articule autour de quatre chapitres. Le premier chapitre décrit l'état de l'art sur la caractérisation multidimensionnelle d'un canal de propagation en mobilité, ainsi que sur les différentes méthodes pour estimer les zones de stationnarité. De plus, une revue de l'état de l'art est réalisée sur l'adaptation des modèles de Machine Learning (ML) et de Deep Learning (DL) pour la classification des scénarios de propagation Line-of-Sight (LOS) et Non-Line-of-Sight (NLOS) et pour l'estimation du canal. Le deuxième chapitre se concentre sur l'architecture du sondeur utilisé lors des 3 campagnes de mesures avec différentes configurations et environnements, et présente quelques résultats saillants illustrant la qualité des données recueillies. Dans le troisième chapitre, les méthodes de la fonction de corrélation du canal (CCF) et de la colinéarité pour caractériser la variabilité temporelle sont appliquées aux données expérimentales en milieu suburbain et en tunnel. De plus, les caractéristiques spatiales et large bande sont analysées localement pour chaque région de stationnarité. L'ensemble des paramètres estimés alimentent une base de données pour le modèle d'apprentissage automatique pour la classification des scénarios de propagation LOS/NLOS dans les communications V2I tel que décrit dans le dernier chapitre.



# Chapitre 1

## État de l'art

### 1.1 Introduction

Les communications mobiles ont évolué de la 2G à la 5G, marquant des avancées majeures en termes de capacité, de couverture et de débits de transmission. Chaque génération, de la 2G introduisant le concept cellulaire et la transmission numérique à la 4G exploitant la LTE, a apporté des améliorations significatives, allant de l'efficacité spectrale accrue à des débits plus rapides et à une latence réduite, propulsant ainsi l'ère de la connectivité mobile (voir la figure 1.1) [4-8]. La 5G est considérée comme une avancée majeure pour la communication véhiculaire grâce à ses technologies clés qui améliorent les performances, telles que le massive MIMO [9-11], le beamforming [12-14] et les ondes millimétriques [15, 16]. Ces améliorations ont permis une extension des cas d'utilisation (voir la figure 1.2) au-delà du haut débit mobile initial, eMBB (Enhanced Mobile Broadband). Cela inclut les cas d'utilisation massifs de la communication machine, qui nécessitent des appareils à faible coût et une longue durée de vie de la batterie, mMTC (Massive Machine Type Communications), ainsi que des applications sensibles telles que l'automatisation industrielle et les systèmes de transport intelligents, qui exigent une fiabilité de communication élevée et une latence minimale. Ces applications sont souvent désignées sous le terme de communication à latence ultra-faible et de fiabilité ultra-élevée, URLLC (Ultra-Reliable Low Latency Communications), dans les normes internationales de télécommunications [17].

Divers cas d'utilisation pour les communications véhiculaires, telles que V2X, comprenant V2I (Vehicle-to-Infrastructure), V2P (Vehicle-to-Pedestrian), V2V (Vehicle-to-Vehicle), V2N (Vehicle-to-Network), et V2G (Vehicle-to-Grid), impliquent la distribution de différents indicateurs clés de performance (KPI) selon le type de communication véhiculaire et ses exigences. Les communications véhiculaires rentrent dans la couche des applications URLLC, notamment les systèmes de transport intelligents (ITS) qui nécessitent des systèmes de communication fiables, robustes, à faible latence et avec une qualité de service supérieure [18][19]. Ces exigences sont définies par les principaux groupes de normalisation, notamment le projet de partenariat de troisième génération (3GPP) pour le standard cellulaire-V2X

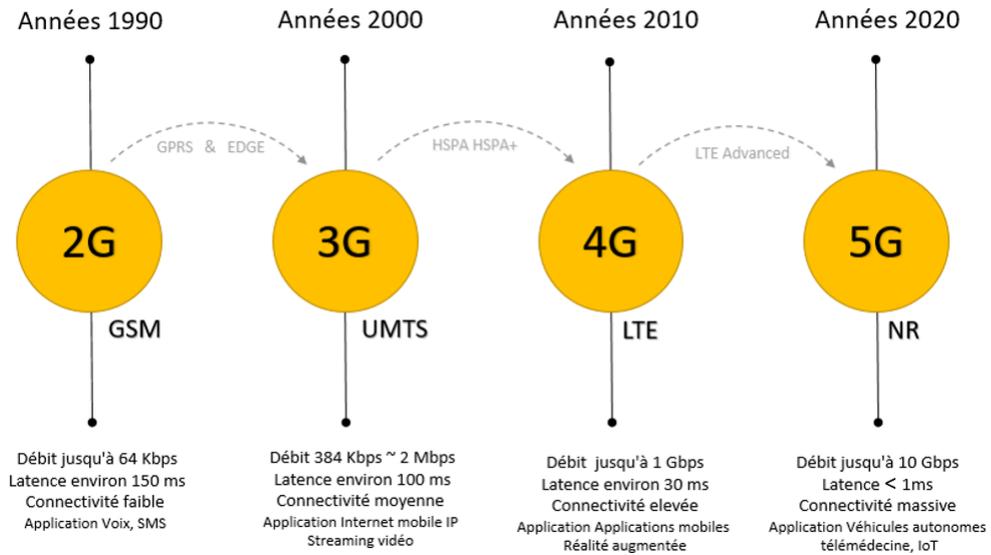


FIGURE 1.1 – Comparaison des technologies 2G, 3G, 4G et 5G

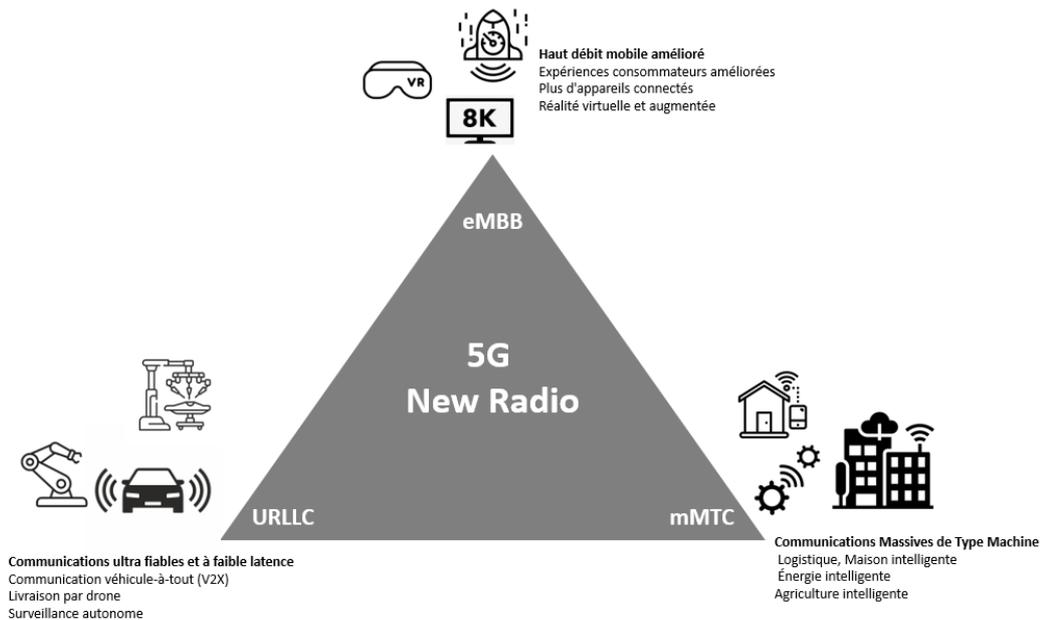


FIGURE 1.2 – Cas d'usage 5G

(C-V2X) dans le Release 14 pour les réseaux 4G LTE-V2X [20] et dans le Release 16 pour les réseaux 5G NR-V2X [21], ainsi que par l'Institut européen des normes de télécommunication (ETSI) pour l'ITS-G5 basée sur IEEE 802.11p en 2012 [22], avec une version améliorée en 2019, IEEE 802.11bd [23]. Les différents cas d'utilisation V2X et leurs KPI sont présentés dans le tableau 1.1, tandis qu'une comparaison des caractéristiques techniques entre ITS-G5 et C-V2X est fournie dans le tableau 1.2.

Catégorie V2X	Débit	Latence	Fiabilité	Usage 5G	Infrastructure	Technologie	Norme
Coopération véhiculaire (V2V)	10-100 Mbps	<1ms	99,999%	URLLC	Small cells, Edge Computing	5G NR	3GPP R15/R16
Le pelotonage de véhicules (V2V)	< 65 Mbps	<25ms	90-99,99%	URLCC	Cellulaire, Macro cells, Small cells	5G eV2X	3GPP R16
Conduite avancée (V2V/V2I)	< 50 Mbps	10-100 ms	90-99,999%	URLCC	Cellulaire, Macro cells, Small cells	5G eV2X	3GPP R16
Conduite à distance (V2V/V2I)	< 25 Mbps	<5ms	99,999%	URLCC	Cellulaire, Macro cells, Small cells	5G eV2X	3GPP R16
Infotrafic et navigation (V2I)	1-10 Mbps	10-100 ms	99,9%	eMBB/mMTC	Cellulaire, Macro cells, Small cells	5G NR	3GPP R15/R16
Gestion de flotte et logistique (V2I)	1-10 Mbps	10-100 ms	99,9%	mMTC	Cellulaire, NB-IoT, Small cells	5G NR	3GPP R15/R16
Services de divertissement (V2I)	10-100 Mbps	10-100 ms	99,9%	eMBB	Macro cells, Small cells, MIMO	5G NR	3GPP R15/R16
Services de sécurité et d'urgence (V2I)	10-100 Mbps	1-10 ms	99,999%	URLLC	Cellulaire, Small cells, Edge Computing	5G NR	3GPP R15/R16
Interaction avec les piétons (V2P)	Varie	< 10 ms	99,999%	URLLC	Cellulaire, Small cells, Edge Computing	5G NR	3GPP R15/R16
Communication avec le réseau électrique (V2G)	Varie	500ms-2s	99,9%	eMBB/mMTC	Smart Grid Infrastructure	5G NR	3GPP R15/R16
Communication directe (V2V/V2I)	1-10 Mbps	< 1ms	99,999%	URLLC	Small cells, MIMO, Edge Computing	NR-V2X	3GPP R16
Communication cellulaire (V2N)	1-10 Mbps	10-100 ms	99,9%	mMTC	Macro cells, Small cells	LTE-V2X	3GPP R14/R15
ITS-G5 (V2V/V2I)	3-27 Mbps	10-100 ms	99,9%	mMTC	WLAN ITS-G5	IEEE 802.11p	ETSI/IEEE 2010/2012
ITS-G5 (V2V/V2I)	6-54 Mbps	< 10 ms	99,9%	mMTC	WLAN ITS-G5	IEEE 802.11bd	ETSI/IEEE 2019

TABLEAU 1.1 – Cas d'usage V2X avec les types de cas d'utilisation 5G, les infrastructures, les technologies et les normes correspondantes [24-33]

Caractéristiques	ITS-G5	ITS-G5 (new)	LTE-V2X	NR-V2X
Technologie	WLAN	WLAN	Cellulaire LTE	Cellulaire NR
Bande de fréquences	5,9 GHz	5,9 GHz et 60 GHz	5,9 GHz et cellulaire	5,9 GHz et cellulaire
Bande passante	10 MHz à 20 MHz	20 MHz	20 MHz	10 MHz à 40 MHz
Standard	ETSI/IEEE 802.11-2010	ETSI/IEEE 802.11bd-2019	3GPP Release 14	3GPP Release 16
Latence	< 100 ms	< 10 ms	< 100 ms	< 1 ms (URLLC)
Fiabilité	99,9%	99,9%	<99,99%	< 99,999%
Débit	3 Mbit/s à 27 Mbit/s	6 Mbit/s à 54 Mbit/s	Dépend de la configuration	Dépend de la configuration
Portée	< 1 km	<2km	Longue portée	Longue portée
Interopérabilité	Établie	En cours	En cours	En cours

TABLEAU 1.2 – Comparaison des Caractéristiques Techniques entre ITS-G5 et C-V2X dans les Communications Véhiculaires

Cependant, en raison de la grande mobilité des émetteurs et récepteurs et de l'apparition sporadique des diffuseurs dans les systèmes de communication véhiculaire, il est difficile d'appliquer l'hypothèse de la stationnarité au sens large (WSS) et de diffusion non corrélée (US) qui suppose que la moyenne et l'autocorrélation du canal  $H(t, f)$  restent constantes dans le temps et la fréquence, les descriptions mathématiques étant présentées ultérieurement. Cette hypothèse permet de simplifier la conception de l'émetteur-récepteur en fournissant des modèles mathématiques simples pour le canal, autorisant l'utilisation de techniques d'estimation de canal ou de précodage moins complexes. Il est donc nécessaire de caractériser le canal de propagation et de définir les régions de stationnarité (généralement le temps de stationnarité) où l'hypothèse WSS-US est valide. En complément, l'intelligence artificielle (IA) peut jouer un rôle clé dans les systèmes de communication V2X basés sur la 5G. En utilisant des techniques d'apprentissage automatique (ML) et profond (DL), il serait possible d'exploiter les données du canal, comme les fonctions de transfert  $H(t, f)$  par exemple, et de développer des modèles d'IA pour améliorer les performances des systèmes de communications en termes de taux d'erreur par bit, d'efficacité spectrale et énergétique. Cela permet aussi de diminuer la complexité des algorithmes. Dans ce premier chapitre, la caractérisation multidimensionnelle d'un canal MIMO est abordée avec la description stochastique d'un canal de propagation et en estimant le temps de stationnarité et les paramètres du canal. Ensuite, un état de l'art de l'utilisation de l'IA dans l'estimation du canal et l'allocation des ressources avec différents modèles ML et DL est décrit. Enfin, une conclusion qui résume les principaux points et les perspectives pour le chapitre suivant est présentée.

## 1.2 Caractérisation multidimensionnelle du canal

Lors d'une communication sans fil, les signaux rayonnés par une antenne d'émission peuvent suivre différents chemins pour atteindre l'antenne de réception. En effet, au chemin direct s'il existe, peuvent s'ajouter des trajets plus longs liés aux différentes interactions du signal avec l'environnement telles que les réflexions, diffractions et/ou diffusions. Le canal de propagation, qui englobe toutes ces interactions entre les antennes d'émission et de réception, est souvent caractérisé par la réponse impulsionnelle  $h(\tau)$  ou la fonction de transfert  $H(f)$ , une transformée de Fourier (FFT) reliant ces deux fonctions. Dans le domaine temporel, la réponse du canal de propagation à une impulsion met en évidence ces différentes répliques du signal, dont les retards relatifs par rapport au premier trajet reçu sont liés aux longueurs des trajets. Leurs amplitudes respectives vont dépendre du nombre d'interactions subies, de la géométrie de l'environnement et du type d'interaction. C'est donc un filtre linéaire dont les coefficients peuvent varier plus ou moins rapidement au cours du temps; cette variation est liée à la mobilité de l'émetteur

et/ou du récepteur et/ou de l'environnement. La réponse impulsionnelle et la fonction de transfert deviennent  $h(t, \tau)$  et  $H(t, f)$ , respectivement. Cette notion de variabilité prend toute son importance dans le domaine des communications dans les systèmes de transport intelligents (ITS - Intelligent Transportation System) [34][35]. Dans ce contexte, une fiabilité de la communication peut être apportée par l'ajout d'un réseau d'antennes à la station de base et parfois au mobile. Cette solution, intégrée dans les derniers standards des réseaux mobiles, permet également d'augmenter le débit fourni par la station de base et réparti vers les utilisateurs connectés à celle-ci. Ainsi, la fonction de transfert du canal de propagation entre les antennes  $k$  d'émission et  $l$  de réception devient  $H(t, f)_{l,k}$ .

À la station de base et pour des fréquences inférieures à 6 GHz, les réseaux d'antennes peuvent atteindre 64 éléments répartis, par exemple, en 8 colonnes de 8 éléments. Si à chaque élément est associée une chaîne RF, le coût du système antennaire ainsi créé devient élevé. Dans certains environnements, comme le milieu rural, il est préférable de disposer d'un gain de réseau pour couvrir la limite de la cellule. Le réseau précédent est par exemple divisé en 8 sous-réseaux connectés, chacun à une chaîne radiofréquence dont le nombre est aussi réduit, et le gain de réseau de 9 dB permet d'augmenter la portée ou le débit [36].

L'article d'Ericsson [36] illustre, pour trois environnements, les structures antennaires les plus appropriées pour accroître la capacité. Sur la Figure 1.3, le réseau 64T64R est constitué de 64 antennes bipolarisées. Trente-deux sous-réseaux verticaux dans chaque polarisation nécessitent au total 64 chaînes radiofréquence à l'émission (64T) ainsi qu'à la réception (64R). Cette structure antennaire permet de tirer profit des étalements angulaires des rayons en azimut et en élévation pour des applications Multi-user MIMO (MU-MIMO) en beamforming 3D. Ce réseau est donc bien adapté au milieu urbain dense. Les distances à couvrir n'excèdent pas 500 m. Le canal étant riche en trajets multiples, la capacité offerte à la cellule par ce type de réseau est supérieure à celle que pourraient procurer les réseaux 32T ou 16T, mieux adaptés respectivement pour les milieux urbains et ruraux. Dans ces deux derniers scénarios, les sous-réseaux verticaux de 4 ou 8 antennes (B. et C. Figure 1.3) procurent un gain de réseau non négligeable qui permet d'étendre les limites de la cellule.

Compte tenu de la taille de ces nouveaux réseaux antennaires, la stratégie d'allocation d'antennes, leur taille et donc les algorithmes de traitement spatial qui en découlent doivent être optimisés. Dans le contexte de forte mobilité, cette optimisation dépendra de la variation spatio-temporelle des signaux vue du réseau d'émission de la station de base et vue des utilisateurs qui vont communiquer simultanément dans les mêmes ressources temps-fréquences. Tous ces aspects de variabilité spatio-temporelle sont quantifiés par des paramètres pertinents extraits de la caractérisation expérimentale multidimensionnelle du canal. L'étude se fait en

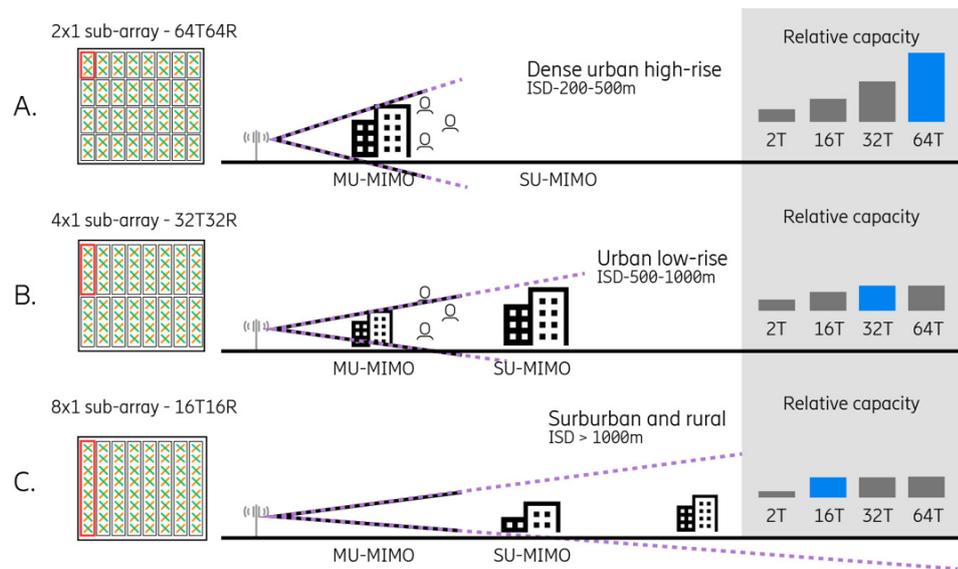


FIGURE 1.3 – Différents types de scénarios de déploiement des réseaux MIMO massifs.

deux étapes. Dans la première, il est indispensable d'estimer les intervalles de temps et fréquence pour lesquels les paramètres statistiques d'ordre 2 du canal ne varient pas. L'estimation du temps de stationnarité ( $T_s$ ) prend toute son importance dès lors qu'il s'agit de scénarios de communication à haute mobilité (V2X) pour lesquels la vitesse élevée des Rx et/ou Tx et/ou la présence de diffuseurs dynamiques entraînent la naissance et la disparition rapide de certains trajets multiples ou des effets de masque. Cette non-stationnarité peut apparaître également en fréquence. C'est la raison pour laquelle on parle par la suite de région de stationnarité définie par un intervalle de durée  $T_s$  et une bande de largeur  $F_s$ . Plusieurs méthodes adaptées aux canaux SISO sont suggérées et utilisées dans la littérature : la méthode basée sur i) la fonction de corrélation du canal (CCF) [35, 37-39], ii) la colinéarité [40-44] et iii) la divergence spectrale [45, 46]. Pour les canaux MIMO, cette région de stationnarité est déduite de la méthode *Correlation Matrix Distance* (CMD) [47-52]. Cette variabilité du canal doit être introduite dans les modèles de propagation pour optimiser les techniques de beamforming et d'allocation de ressources [53-57]. C'est pourquoi, récemment dans le modèle du 3GPP [58], la notion de *consistance spatiale* a été introduite pour caractériser la région dans laquelle les utilisateurs voient le même canal et dont les paramètres changent lentement au cours de leur déplacement. En particulier, la consistance spatiale fait référence à la manière dont les paramètres de canal de communication évoluent de manière continue lorsque le récepteur mobile se déplace. Cela signifie que deux liens avec des positions similaires ont des caractéristiques de canal similaires, et le canal devrait évoluer de manière régulière dans le temps et dans l'espace [59-63]. Le paragraphe suivant rappelle brièvement les fonctions descriptives du canal de propagation sur lesquelles s'appuieront les différentes méthodes d'estimation de la région de stationnarité décrites aux

paragraphe 1.2.2- 1.2.3. Le dernier paragraphe décrit les paramètres permettant de définir la consistance spatiale.

### 1.2.1 Description stochastique d'un canal de propagation SISO

Bello [64] a introduit quatre fonctions pour représenter le canal comme un quadripôle. Une description de ces fonctions est donnée en [annexe A](#). Pour un environnement et un scénario de communication donnés, une modélisation basée sur une description plus ou moins fine de la scène devient vite exhaustive. Une approche stochastique lève cette difficulté et permet alors de modéliser le canal à l'aide de variables aléatoires. Dans un contexte de mobilité, les paramètres de ces variables aléatoires vont évoluer au cours du temps. La fonction de corrélation  $R_H(t, t + \Delta t; f, f + \Delta f)$  permet de donner une représentation déterministe de ce processus stochastique. Elle se calcule comme suit :

$$R_H(t, t + \Delta t; f, f + \Delta f) = E[H(t, f).H^*(t + \Delta t, f + \Delta f)] \quad (1.1)$$

Si  $R_H(t, t + \Delta t; f, f + \Delta f)$  est indépendant de l'instant «  $t$  » à partir duquel le calcul est réalisé, alors le processus décrivant les évanouissements du signal, ne dépendant que de  $\Delta t$ , est stationnaire au sens large (WSS-Wide Sense Stationary). Si de plus,  $R_H$  est indépendante de «  $f$  », le canal est à diffuseurs non corrélés (US-Uncorrelated Scattering) [65][66]. (1.1) se simplifie et devient :

$$R_H(t, t + \Delta t; f, f + \Delta f) = R_H(\Delta t, \Delta f) \quad (1.2)$$

Avec l'hypothèse d'un canal stationnaire, la fonction de diffusion  $C(\tau, \nu)$  est introduite pour caractériser la distribution de puissance sur les trajets dans le domaine des retards ( $\tau$ ) et des fréquences Doppler ( $\nu$ ).  $C(\tau, \nu)$ , donnée par (1.3), est obtenue en appliquant une double transformée de Fourier.

$$C(\tau, \nu) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} R_H(\Delta t, \Delta f) e^{-j2\pi(\nu\Delta t - \tau\Delta f)} d\Delta f d\Delta t \quad (1.3)$$

Pour les scénarios V2X, l'hypothèse de WSS-US n'est pas valide sur tout le parcours de déplacement du mobile [67], [47], [68]. D'un point de vue développement de modèles de propagation et d'optimisation des schémas de transmission, il est important d'estimer les régions en temps et en fréquence sur lesquelles l'hypothèse WSSUS est valide.

### 1.2.2 Estimation du temps de stationnarité d'un canal de propagation SISO

#### — Estimation de la fonction de diffusion locale généralisée GLSF

Dans le cas de canaux non WSS-US, la fonction de diffusion (1.3) doit être associée à une région de stationnarité locale sur laquelle l'hypothèse WSS-US est valide. La

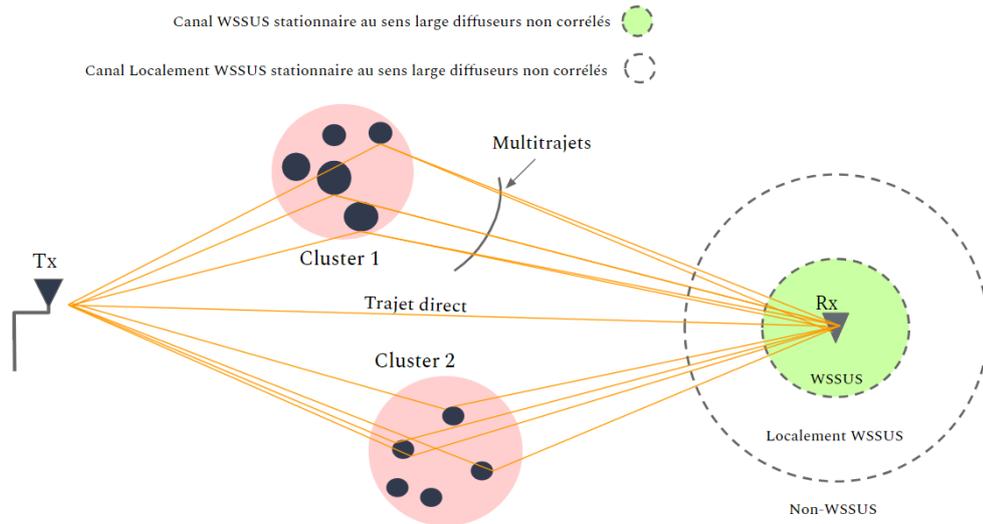


FIGURE 1.4 – Illustration des concepts de WSS-US, WSS-US local et Non-WSSUS appliqués à un canal de propagation SISO

fonction de diffusion locale (LSF), notée  $C(t, f; \nu, \tau)$ , est calculée de façon similaire à (1.3) mais en conservant les dépendances temporelle et fréquentielle de la fonction de corrélation  $R_H$ .  $C(t, f; \nu, \tau)$  est donnée par la relation :

$$C(t, f; \tau, \nu) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} R_H(t, f; \Delta t, \Delta f) e^{-j2\pi(\nu\Delta t - \tau\Delta f)} d\Delta f d\Delta t \quad (1.4)$$

Cependant, le calcul de  $C(t, f; \nu, \tau)$  présente des valeurs imaginaires ou réelles négatives, ce qui physiquement ne peut s'expliquer puisqu'il s'agit, pour les environnements WSS-US, d'un spectre en puissance. De plus, la prise en compte de l'ensemble de la fonction de corrélation peut engendrer des coûts de calcul importants. Dans son article, [69] contourne ce problème en introduisant la fonction de diffusion locale généralisée (GLSF) notée  $\hat{C}(t, f; \nu, \tau)$ . La GLSF est calculée en effectuant un fenêtrage local en temps et en fréquence de  $H(t, f)$  avec des fonctions déterministes orthogonales respectivement  $g_i(t)$  et  $g_j(f)$ . Il s'agit de séquences sphéroïdales étendues (DPSS) [69, 70]. La fonction de filtrage globale peut s'exprimer comme suit :

$$G_w(t, f) = g_i(t) \times g_j(f) \quad (1.5)$$

où  $i \in 1, \dots, I$ ,  $j \in 1, \dots, J$  et  $w \in 0, \dots, IJ - 1$  avec  $IJ$  le nombre de fenêtres choisies en temps et en fréquence. Dans [69], où  $I$  et  $J$  sont obtenus en minimisant l'erreur quadratique moyenne entre les GLSF  $\hat{C}(t, f; \nu, \tau)$  et le cas idéal LSF  $C(t, f; \nu, \tau)$ . Différentes valeurs de  $I$  et  $J$  variant entre 1 et 6 ont été publiées dans la littérature et sont regroupées dans le tableau 1.3.

La méthode pour estimer la GLSF consiste à diviser l'ensemble du parcours en  $N$  régions. La taille de ces régions est un compromis entre un nombre suffisant

Référence	Scénario	La valeur $I$	La valeur $J$
[69]	Suburbain	4	6
[41]	Urbain, Autoroute	5	1
[71]	Urban	2	2
[43]	Urban	4	2
[38]	Suburbain	3	3
[39]	Tunnel	3	3
[39]	Tunnel	2	2

TABLEAU 1.3 – Valeurs de  $I$  et  $J$  proposées dans la littérature

d'observations obtenues dans cette région et une taille suffisamment faible pour que l'hypothèse WSSUS soit satisfaite. La GLSF est donnée par l'expression [69] :

$$\hat{C}(t, f; \tau, \nu) = E \left\{ \sum_{w=0}^{IJ-1} \gamma_w |H^{(G_w)}(t, f; \tau, \nu)|^2 \right\} \quad (1.6)$$

$E$  représente la moyenne sur  $I \times J$  fenêtres orthogonales. Avec

$$H^{(G_w)}(t, f; \tau, \nu) = e^{j2\pi f \tau} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} H(t', f') \times G_w(t' - t, f' - f) e^{-j2\pi(\nu t' - \tau f')} dt' df' \quad , \quad (1.7)$$

et

$$\sum_{w=0}^{IJ-1} \gamma_w = 1 \quad (1.8)$$

Pour une région donnée, la GLSF peut être vue comme un spectrogramme multidimensionnel. La Figure 1.5 illustre schématiquement, dans le domaine temporel, le parcours comportant  $M$  observations du canal tous les  $T_{snapshot}$ . Les  $L$  points encadrés sur cette courbe correspondent à une des  $N$  régions sur laquelle le calcul de (1.6) est réalisé. Pour chaque intervalle, un spectre dans le domaine retard-Doppler est obtenu.

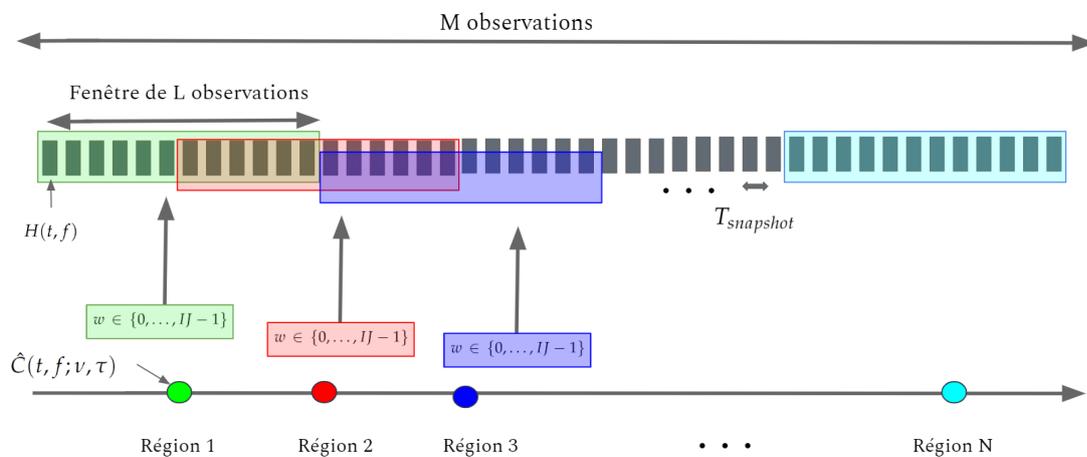


FIGURE 1.5 – Illustration des GLSF dans le domaine temporel.

[40] a montré que, pour les fréquences porteuses actuelles des réseaux mobiles, la bande de fréquence  $F_s$  sur laquelle l'hypothèse US est valide est supérieure à 150 MHz. Par conséquent, dans la suite, seul le temps de stationnarité est étudié.

— **Estimation du temps de stationnarité à partir de la fonction de corrélation du canal (CCF)**

Dans la littérature, il a été montré que le temps de stationnarité  $T_s$  peut être dérivé de la GLSF en utilisant trois approches. La première approche repose sur la fonction de corrélation du canal (CCF)  $R^G(\Delta\nu; \Delta f, \Delta t)$  1.9, qui est obtenue par la transformée de Fourier tridimensionnelle de la GLSF :

$$R^G(\Delta\nu; \Delta f, \Delta t) = \iiint \hat{C}(t; \tau, \nu) \times e^{-j2\pi(\nu\Delta t - \tau\Delta f - t\Delta\nu)} dt d\tau d\nu \quad (1.9)$$

Le temps de stationnarité  $T_s$  est déduit du moment Doppler d'ordre 1 du CCF  $\overline{\Delta\nu}$  (1.11), comme expliqué dans [69][37]. Le calcul du moment d'ordre 1 du CCF  $\overline{\Delta\nu}$  (1.10) quantifie l'intervalle de décalage Doppler  $\Delta\nu$  dans lequel une corrélation significative existe. Ce moment Doppler d'ordre 1 de CCF est calculé comme suit :

$$\overline{\Delta\nu} = \frac{1}{\|R^G\|_1} \iiint |\Delta\nu| R^G(\Delta\nu; \Delta f, \Delta t) d\Delta t d\Delta f d\Delta\nu \quad , \quad (1.10)$$

avec  $\|R^G\|_1$ , la première norme du CCF dans les trois dimensions. Par conséquent, le temps de stationnarité  $T_s$  est déduit de la relation suivante :

$$T_s \triangleq \frac{1}{\overline{\Delta\nu}} \quad (1.11)$$

— **Estimation de  $T_s$  à partir de la colinéarité des fonctions GLSF**

Cette deuxième approche, décrite dans [41], permet d'estimer  $T_s$  empiriquement en calculant la corrélation entre deux GLSF estimées à  $t_1$  et  $t_2$ .  $T_s$  est l'intervalle de temps pour lequel la fonction de corrélation calculée en (1.12) franchit un certain seuil  $\alpha$ , où 1 indique des fonctions en  $t_1$  et  $t_2$  identiques. La colinéarité  $d_{col}$  est calculée comme suit :

$$d_{col}[t_1, t_2] = \frac{\text{tr}\{\hat{C}(t_1; \tau, \nu)\hat{C}(t_2; \tau, \nu)\}}{\|\hat{C}(t_1; \tau, \nu)\|_F \|\hat{C}(t_2; \tau, \nu)\|_F} \quad , \quad (1.12)$$

où  $\text{tr}\{X\}$  et  $\|X\|_F$  représentent respectivement la trace et la norme de Frobenius de  $X$ . Comme la colinéarité est basée sur la GLSF, elle prend en compte les variations du spectre retard-Doppler au cours du temps. Le résultat de (1.12) est une matrice carrée.  $T_s$  correspond à l'intervalle de temps centré autour de la diagonale principale, il est défini dès que le coefficient de corrélation atteint la valeur  $\alpha$ . Différents seuils dans la littérature sont utilisés et varient entre 0.80 et 0.94 dans [38, 40, 41, 47, 71, 72].

— Estimation de la région de stationnarité en utilisant la divergence spectrale

En raison de la normalisation, la mesure de la colinéarité varie entre 0 et 1 et ne prend pas en compte la densité de puissance. La divergence spectrale est une autre métrique permettant d'évaluer la non-stationnarité. Elle a été introduite comme un outil de comparaison des densités spectrales de puissance [46]. La divergence spectrale est définie comme suit :

$$d_{div}[t_1, t_2] = \log \left( \frac{1}{|S|^2} \iint \frac{\hat{C}(t_1; \tau, \nu)}{\hat{C}(t_2; \tau, \nu)} d\nu d\tau \iint \frac{\hat{C}(t_2; \tau, \nu)}{\hat{C}(t_1; \tau, \nu)} d\nu d\tau \right) , \quad (1.13)$$

où  $S$  représente la région retard-Doppler  $(\nu, \tau)$  et  $|S|^2$  sa taille. On note que  $\hat{C}(t; \tau, \nu)$  est un spectre de puissance variant en temps, et  $d_{div}\{\hat{C}\}[t_1, t_2] \geq 0$  avec égalité si et seulement si  $\hat{C}(t_1; \tau, \nu) = \hat{C}(t_2; \tau, \nu)$ , ce qui est vérifié pour un canal de propagation WSS. Cependant, comme il s'agit d'une méthode non bornée, elle ne peut évaluer que qualitativement la nature non-WSS d'un canal.

### 1.2.3 Estimation de la stationnarité spatiale pour un canal SU-MIMO

$T_s$  peut également être évalué à partir de l'analyse de la variabilité spatiale sur le réseau d'antennes d'émission et/ou de réception. Pour un canal radio de propagation Single-User MIMO (SU-MIMO), qui varie dans le temps, la non-stationnarité du canal peut être caractérisée spatialement à l'émission ou à la réception en utilisant une approche empirique qui évalue la différence entre les matrices de corrélation CMD introduites dans [47, 51]. Cette dernière permet de surveiller les changements dans la structure spatiale des canaux MIMO non stationnaires, par le calcul de la similarité entre deux matrices de corrélation  $R$  à deux instants différents  $t_1$  et  $t_2$ , définie comme suit :

$$d(R_H^{Tx}(t_1), R_H^{Tx}(t_2)) = 1 - \frac{\text{tr}\{R_H^{Tx}(t_1)R_H^{Tx}(t_2)\}}{\|R_H^{Tx}(t_1)\|_F \|R_H^{Tx}(t_2)\|_F} , \quad (1.14)$$

et

$$d(R_H^{Rx}(t_1), R_H^{Rx}(t_2)) = 1 - \frac{\text{tr}\{R_H^{Rx}(t_1)R_H^{Rx}(t_2)\}}{\|R_H^{Rx}(t_1)\|_F \|R_H^{Rx}(t_2)\|_F} , \quad (1.15)$$

avec

$$R_H^{Tx}(t) = E[H(t)^T H(t)^*] , \quad R_H^{Rx}(t) = E[H(t)^H (t)^H], \quad (1.16)$$

en appliquant cette formulation aux matrices  $H$  expérimentales  $R_H^{Tx}$  et  $R_H^{Rx}$  deviennent :

$$R_H^{Tx}(iT_{\text{snapshot}}) = \frac{1}{L} \sum_{l=i}^{i+L-1} H(lT_{\text{snapshot}})^T H(lT_{\text{snapshot}})^* \quad (1.17)$$

$$R_H^{Rx}(iT_{snapshot}) = \frac{1}{L} \sum_{l=i}^{i+L-1} H(lT_{snapshot})H(lT_{snapshot})^H \quad (1.18)$$

dans laquelle  $R_H^{Tx}$  et  $R_H^{Rx}$  sont respectivement les matrices de covariance de dimension  $N_{Tx} \times N_{Tx}$  à l'émission et  $N_{Rx} \times N_{Rx}$  à la réception, déduites de la matrice canal  $H$  de dimension  $N_{Rx} \times N_{Tx}$  à l'instant  $t$ , et  $E$  la moyenne sur  $L$  échantillons dans le temps.  $L$  doit être suffisamment grand pour estimer avec précision les matrices de corrélation mais aussi suffisamment petit pour faire la moyenne des échantillons dans le temps avec les mêmes statistiques [47]. Une valeur de la CMD à 0 traduit une corrélation maximale. La CMD est désormais introduite dans le modèle de canaux MIMO du 3GPP [73, 74].

Référence	Méthode	Environnement	Vitesse Km/h	scénario	$T_s$ (ms)	$D_s$ (m)
[38]	CCF	Suburbain	40	V2I	568	6.3
[38]	Collinéarité	Suburbain	40	V2I	568	6.3
[40]	Collinéarité	Suburbain	40	V2V	1380	15.33
[75]	Collinéarité	Tunnel	90	V2V	1060	34.5
[39]	CCF	Tunnel	90	V2I	330	8.25
[40]	Collinéarité	Autoroute	90	V2V	2060	5.15
[71]	Collinéarité	Urbain	10	V2I	648	1.80
[71]	CMD	Urbain	10	V2I	11277	31.24
[47]	CMD*	Urbain	7.5	V2V	1440	3
[47]	CMD*	Autoroute	7.5	V2V	1790	3.72
[47]	CMD*	Suburbain	7.5	V2V	1723	3.58
[47]	CMD*	Campus	7.5	V2V	1684	3.5
[51]	CMD	Indoor	1.8	-	3000	1.5
[76]	CMD	Tunnel	72	V2I	175	3.5

TABLEAU 1.4 – Synthèse sur la stationnarité. Pour les scénarios V2V, le Tx et le Rx se déplacent dans la même direction. Équations (1.14) et (1.15), [47]  $L=25$ ; [71]  $L=16$ ; [40]  $L=$

Le tableau 1.4 rassemble les valeurs de la distance de stationnarité  $D_s$  déduite de  $T_s$  et publiées pour différents scénarios et environnements [77][78]. Les valeurs sont comprises entre 6,3 m et 8,5 m pour les scénarios V2I et entre 5,3 m et 34,4 m pour les scénarios V2V dans différents environnements. Dans le cas d'un SU-MIMO avec CMD, les valeurs trouvées sont comprises entre 3 m et 3,72 m. La formule de  $D_s$  est la suivante :

$$D_s = v \times T_s \times 10^{-3} \quad , \quad (1.19)$$

avec  $v$  représentant la vitesse relative entre l'émetteur et le récepteur en mètres par seconde  $ms^{-1}$ ,  $T_s$  est le temps de stationnarité en ms.

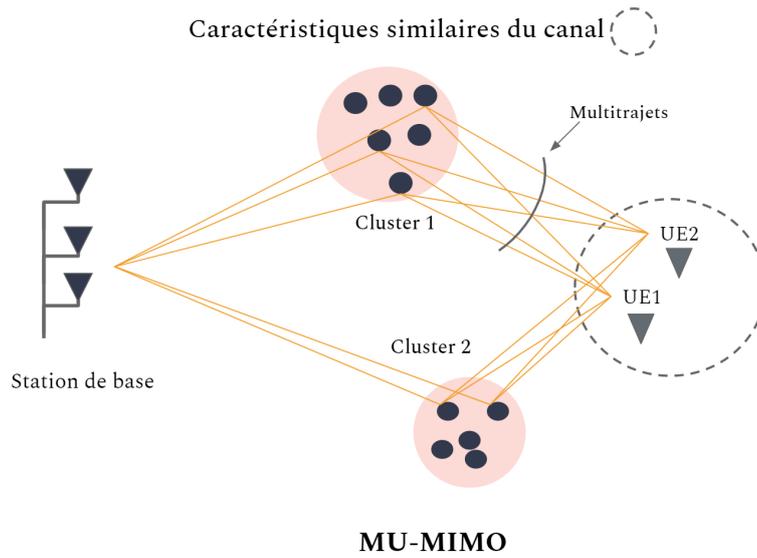


FIGURE 1.6 – Distance de la matrice de corrélation spatiale MU-MIMO

#### 1.2.4 La consistance spatiale pour un canal MU-MIMO

En 5G/6G, la stratégie d'allocation en temps-fréquence et spatiale des ressources vers les utilisateurs et la mise à jour des algorithmes de traitement spatial vont dépendre, entre autres, de la variabilité du canal. La consistance spatiale a été introduite récemment dans les modèles géométriques-stochastiques du 3GPP [58]. Elle représente la distance sur laquelle un utilisateur ou groupe d'utilisateurs très corrélés va se déplacer dans un environnement qui évolue peu. C'est-à-dire que les canaux sur cette zone sont spatialement corrélés, comme l'illustre la Figure 1.6 qui montre le principe de base. Cette consistance spatiale va dépendre des effets de masque locaux, de la visibilité directe ou non, des directions d'arrivée des rayons, etc. Cette consistance spatiale ne s'exprime donc pas en fonction d'un seul paramètre, comme le souligne [79]. Pour évaluer cette consistance spatiale, certains auteurs [80], [81] et [82] appliquent la CMD décrite précédemment pour suivre les changements dans la structure spatiale des canaux MIMO non stationnaires. Cette distance  $d_{CMD}$  est calculée en utilisant les matrices de covariance de deux UEs, désignés par  $R_1$  et  $R_2$ , comme suit :

$$d_{CMD}(R_1, R_2) = 1 - \frac{\text{tr}\{R_1^H R_2\}}{\|R_1\|_F \|R_2\|_F} \quad (1.20)$$

Le tableau 1.5 rassemble les valeurs publiées dans le 3GPP [58].  $d_{CMD}$  représente la distance de corrélation pour laquelle les utilisateurs reçoivent les mêmes caractéristiques de canal [58].  $d_{SF}$  est la distance de corrélation du Shadow Fading (SF) entre les différents utilisateurs. Les modèles SF sont utilisés pour calculer la variance du pathloss. SF est défini par une distribution log-normale dont l'écart type varie pour chaque scénario.

	RMa			UMi			UMa			Indoor	
	LOS	NLOS	O2I	LOS	NLOS	O2I	LOS	NLOS	O2I	LOS	NLOS
$d_{CMD}$ (m)	50	60	15	12	15	15	40	50	15	10	10
$d_{SF}$ (m)	37	120	10	13	7	35	37	50	7	10	6

TABLEAU 1.5 – Distance de corrélation pour la consistance spatiale pour les environnements suivants tels que définis dans le 3GPP - RMa : environnement rural, UMi : canyon urbain, UMa : Urbain et O2I : de l'extérieure à l'intérieur

## 1.2.5 Estimation des paramètres de canal

### — A partir de la GLSF

Si les conditions de stationnarité du canal sont valides, il est possible de déterminer, sur des intervalles de durée inférieure à  $T_s$ , à partir des GLSF, les principaux paramètres du canal de propagation, tels que :

#### L'étalement des retards

$$\sigma_\tau(t) = \sqrt{\frac{1}{\rho_H^2(t)} \int (\tau - \bar{\tau})^2 P_H(t; \tau) d\tau} \quad (1.21)$$

#### L'étalement Doppler

$$\sigma_\nu(t) = \sqrt{\frac{1}{\rho_H^2(t)} \int (\nu - \bar{\nu})^2 Q_H(t; \nu) d\nu} \quad , \quad (1.22)$$

avec  $P_H$  et  $Q_H$  respectivement les profils de puissance de retard et de Doppler local, définis par :

$$P_H(t; \tau) = \int \hat{C}(t; \tau, \nu) d\nu \quad , \quad (1.23)$$

$$Q_H(t; \nu) = \int \hat{C}(t; \tau, \nu) d\tau \quad , \quad (1.24)$$

et  $\rho_H^2(t) = \iint \hat{C}(t; \tau, \nu) d\tau d\nu$  est le gain du canal à l'instant  $t$ ,  $\bar{\tau}$  et  $\bar{\nu}$  étant les moments d'ordre 1 local des retards et des fréquences Doppler, respectivement. Dans une région de stationnarité, la bande de cohérence  $B_c$  et le temps de cohérence  $T_c$  du canal peuvent être déduits respectivement de l'étalement des retards  $\sigma_\tau$  et de l'étalement Doppler  $\sigma_\nu$  par des relations empiriques :

$$T_c \triangleq \frac{1}{\sigma_\nu} \quad , \quad B_c \triangleq \frac{1}{\sigma_\tau} \quad (1.25)$$

Mais ces formulations dépendent d'un facteur de proportionnalité à estimer. Une autre méthode (1.26), (1.27) consiste à fixer une valeur pour  $R_H(\Delta t, \Delta f)$ , généralement à 0,7 ou à 0,5, et d'en déduire les valeurs de  $T_c$  et  $B_c$  dans une région de stationnarité.

$$R_H(\Delta t_{\rho_t}, 0) = \rho_t \text{ Donc } T_c = \Delta t_{\rho_t} \text{ ,} \quad (1.26)$$

$$R_H(0, \Delta f_{\rho_f}) = \rho_f \text{ Donc } B_c = \Delta f_{\rho_f} \text{ ,} \quad (1.27)$$

avec  $\rho_t$  et  $\rho_f$  respectivement les seuils de corrélation désirés à atteindre en temps et

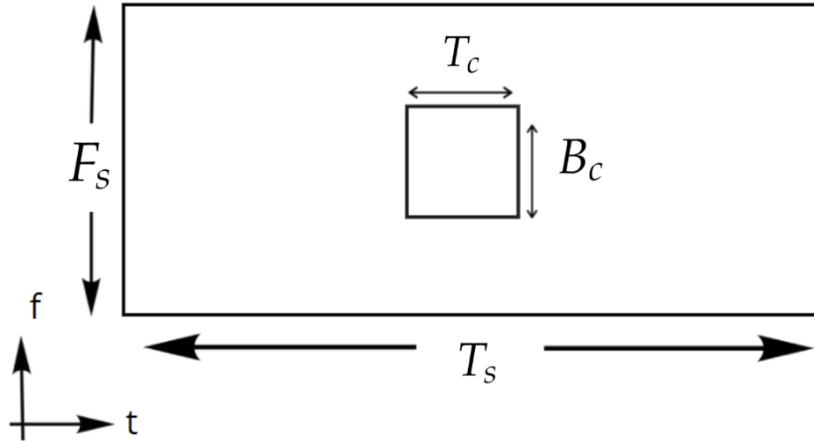


FIGURE 1.7 – Région de cohérence et de stationnarité.

en fréquence. En conclusion sur ces aspects de stationnarité et de cohérence, la Figure 1.7 illustre la région de cohérence qui est incluse dans la région de stationnarité.

#### — Paramètres large bande dérivés des réponses impulsionnelles $h(t, \tau)$

D'autres paramètres peuvent être déduits de la réponse impulsionnelle large bande  $h(t, \tau)$  pour un échantillon donné, parmi lesquels :

**La moyenne  $\mu_h$  et l'écart type  $\sigma_h$**  sur les retards  $\tau$  d'une réponse impulsionnelle  $h(t, \tau)$ , définis par :

$$\mu_h = E(|h(\tau)|) = \int \tau |h(\tau)| d\tau, \quad (1.28)$$

$$\sigma_h = \sqrt{\int (\tau - \mu_h)^2 |h(\tau)| d\tau} \quad (1.29)$$

**Le coefficient de variation  $A$**  est le rapport entre l'écart-type  $\sigma_h$  et la moyenne  $\mu_h$ . Il permet de comparer les dispersions des distributions statistiques des amplitudes de  $h(\tau)$  qui peuvent être très différentes.

$$A = \frac{\sigma_h}{\mu_h} \quad (1.30)$$

**Le coefficient d'asymétrie**  $\mathcal{S}$ , par définition, est un paramètre qui mesure le degré d'asymétrie de la distribution d'une variable aléatoire.

$$\mathcal{S} = \frac{E(|h(\tau)| - \mu_h)^3}{\sigma_h^3} \quad (1.31)$$

**Le coefficient d'aplatissement**  $\mathcal{K}$ , ou Kurtosis, est un paramètre qui mesure le degré d'aplatissement ou de rétrécissement de la distribution d'une variable aléatoire.

$$\mathcal{K} = \frac{E(|h(\tau)| - \mu_h)^4}{\sigma_h^4} \quad (1.32)$$

**Le facteur de Rice**  $K_{rice}$  est défini comme le rapport entre la puissance du trajet direct  $\tau_0$  et la puissance des trajets multiples.

$$K_{rice} = \frac{|h(\tau)|_{\tau=\tau_0}^2}{\sum |h(\tau)|_{\tau \neq \tau_0}^2}, \quad (1.33)$$

**Le retard moyen**  $\tau_{MOY}$  est la moyenne du retard du profil de retard de puissance  $|h(\tau)|^2$  du canal, qui est définie par :

$$\tau_{MOY} = \frac{\int \tau |h(\tau)|^2 d\tau}{\int |h(\tau)|^2 d\tau} \quad (1.34)$$

**L'étalement des retards**  $\tau_{RMS}$  est une mesure de la richesse des trajets multiples d'un canal de communication ; il est défini par :

$$\tau_{RMS} = \sqrt{\frac{\int (\tau - \tau_{MOY})^2 |h(\tau)|^2 d\tau}{\int |h(\tau)|^2 d\tau}} \quad (1.35)$$

**Le temps de montée**  $\Delta\tau_{rise}$  mesure l'intervalle de temps entre le premier trajet et le trajet multiple de plus forte amplitude, défini par :

$$\Delta\tau_{rise} = \operatorname{argmax}_{\tau} (|h(\tau)|) - \min(\tau_l) \quad (1.36)$$

Avec  $l$  est l'index des trajets multiple.

**La différence angulaire**  $\Delta\lambda$  mesure la différence entre les angles de départ en azimut  $\phi_D$  et les angles d'arrivée  $\phi_A$  du trajet multiple de plus forte amplitude  $l$ , défini par :

$$\Delta\lambda_{\tau_l} = |e^{j\phi_{D,\tau_l}} - e^{j\phi_{A,\tau_l}}| \quad (1.37)$$

**L'étalement angulaire**  $AS$  mesure les dispersions des angles de départ AoA ou d'arrivée AoD de toutes les composantes des trajets multiples. Il peut être calculé

comme suit :

$$AS = \sqrt{-2 \ln \left| \frac{\int e^{j\phi\tau} |h(\tau)|^2 d\tau}{\int |h(\tau)|^2 d\tau} \right|} \quad (1.38)$$

**Moyenne du spectre angulaire**  $\mu_\phi$ , la moyenne du spectre angulaire de puissance, calculée comme suit :

$$\mu_\phi = \arg \left\{ \int e^{j\phi\tau} |h(\tau)|^2 d\tau \right\} \quad (1.39)$$

Il est essentiel de distinguer entre les paramètres statistiques à large bande décrits précédemment et les paramètres statistiques à bande étroite décrits par la suite, qui sont des valeurs calculées à partir d'un signal reçu en bande étroite (RSS, RSSI, etc.). Ces paramètres sont estimés à partir d'un nombre d'échantillons distincts dans le temps, par opposition à ceux à large bande, avec la variable "t" représentant le temps. Par exemple, la moyenne  $\mu_t$ , l'écart type  $\sigma_t$ ,  $\mathcal{K}_t$ ,  $S_t$ , etc.

### 1.2.6 Applications de l'hypothèse WSSUS

Dans cette section, la pertinence de la caractérisation de la non-stationnarité est brièvement discutée en ce qui concerne certains aspects pratiques, comme suggéré dans [37], [38] et [77].

#### La simulation de canal et l'évaluation des performances

Les hypothèses associées au canal WSSUS ont permis de simplifier la conception, la simulation et l'évaluation de nombreux systèmes de communication. Les simulateurs d'un canal de propagation radio qui prennent en compte la non-stationnarité des canaux présentent un intérêt pratique considérable, notamment pour l'évaluation des performances (à long terme) des systèmes de communication. Les propriétés du canal de propagation à long terme sont évaluées et supposées stationnaires, tandis que les canaux pratiques, notamment dans les communications véhiculaires, ne satisfont pas à ces hypothèses, ce qui influe sur les performances de ces systèmes [37].

#### Capacité Ergodique

En plus d'améliorer significativement les performances du système, l'information sur la non-stationnarité du canal permet de comprendre le comportement à long terme du système. L'importance de ces informations dans la définition de la capacité du canal est prise en considération. La capacité ergodique  $C_{\text{erg}}$  du canal est souvent recherchée afin de révéler les propriétés à long terme d'un processus de fading arbitraire. Pour atteindre la capacité ergodique, une moyenne par codage sur de nombreuses réalisations indépendantes et identiquement distribuées (i.i.d.) d'un très long codebook gaussien est requise. En particulier, celui-ci doit être suffisamment long pour que le fading reflète sa nature ergodique, c'est-à-dire que le codage devrait

couvrir de nombreuses réalisations de fading i.i.d. [83]. Cela peut être formulé comme suit :

$$C_{\text{erg}} = E[B \log_2(1 + \gamma)] = \int B \log_2(1 + \gamma) P(\gamma) d\gamma, \quad (1.40)$$

avec  $\gamma$  représentant le rapport signal sur bruit instantané, et  $B$  la bande passante, où  $P(\gamma)$  est la distribution probabiliste de  $\gamma$ .

Les propriétés statistiques du canal pour les canaux doublement sélectifs demeurent constantes sur une région de  $T_s \times F_s$ . Si nous supposons que  $N_i$  réalisations du canal i.i.d. se produisent à chaque intervalle  $T_c$  dans le temps et  $B_c$  en fréquence, avec

$$N_i = \frac{T_s F_s}{T_c B_c}, \quad (1.41)$$

pour un canal WSSUS  $T_s F_s \rightarrow \infty$ , ce qui fait  $N_i \rightarrow \infty$ , le nombre de réalisations indépendantes est suffisamment grand et la capacité ergodique  $C_{\text{erg}}$  est atteinte [84, 85]. La capacité ergodique dépendante de  $N_i$  peut alors être donnée par :

$$C_{\text{erg}} = \lim_{N_i \rightarrow \infty} \sum_{N_i} B \log_2(1 + \gamma_i) P_{N_i}(\gamma_i). \quad (1.42)$$

### Degrés de diversité dans le canal

La stationnarité et les paramètres du canal extraits de cette caractérisation présentent l'avantage d'évaluer les gains offerts par les techniques de diversité temporelle  $d_t$ , fréquentielle  $d_f$ , dans le domaine des retards  $d_\tau$  et Doppler  $d_\nu$  [37, 38, 77]. Ces degrés de diversité dans le canal sont décrits comme suit :

$$d_t = \frac{T_s}{T_c}, \quad d_f = \frac{F_s}{B_c}, \quad (1.43)$$

$$d_\nu = \frac{\sigma_\nu}{\Delta\nu}, \quad d_\tau = \frac{\sigma_\tau}{\Delta\tau}, \quad (1.44)$$

avec  $\overline{\Delta\tau}$  représentant le retard d'ordre 1 du CCF qui quantifie l'intervalle de décalage en retard dans lequel une corrélation significative existe.  $\overline{\Delta\tau}$  est calculé dans la dimension du retard de la même manière que l'équation (1.10).

## 1.3 Utilisation de l'intelligence artificielle dans l'estimation du canal et l'allocation de ressources

### 1.3.1 L'apprentissage automatique (ML) et l'apprentissage profond (DL)

En tant que techniques clés pour l'intelligence artificielle, l'apprentissage automatique ou machine learning (ML) et l'apprentissage profond ou deep learning (DL) sont capables de résoudre des problèmes complexes ou difficiles à modéliser mathématiquement, et ce, sans les programmer et les coder explicitement. L'apprentissage automatique peut s'adapter automatiquement avec un minimum d'intervention humaine. Les données existantes sont exploitées pour trouver des modèles et des caractéristiques permettant de prédire ou d'anticiper des événements futurs. D'autre part, l'apprentissage profond est un sous-ensemble de l'apprentissage automatique qui utilise des réseaux neuronaux artificiels profonds pour imiter le processus d'apprentissage du cerveau humain.

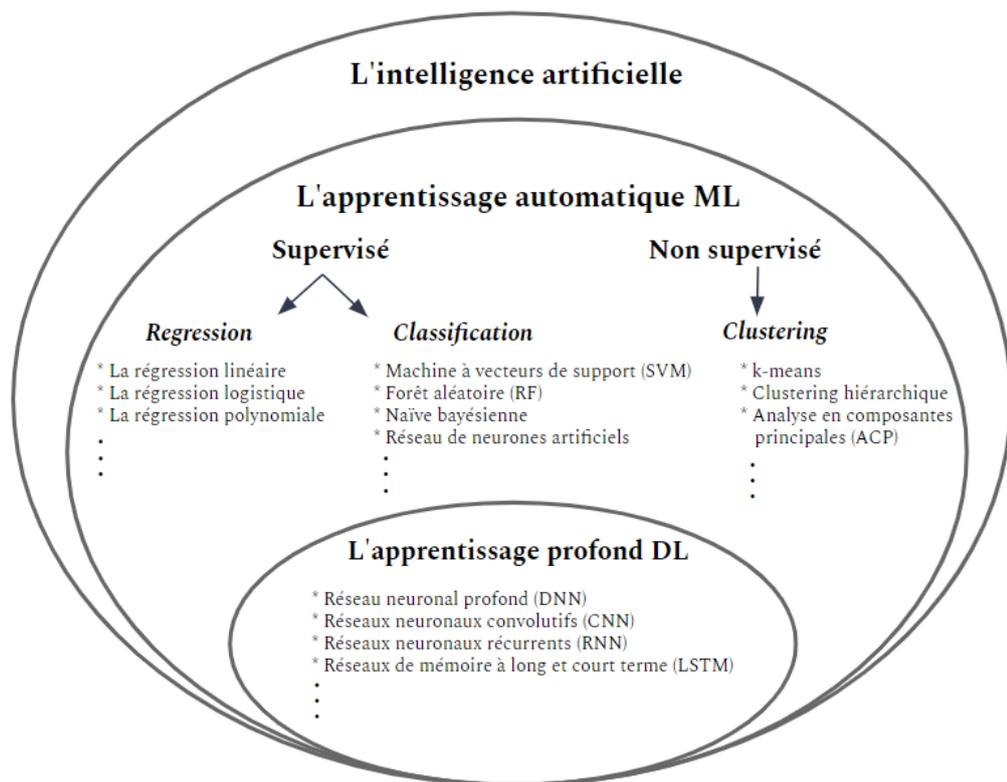


FIGURE 1.8 – L'intelligence artificielle et ses sous-domaines .

#### Principe de base de l'apprentissage automatique

Deux techniques sont utilisées dans l'apprentissage automatique, comme le montre la Figure 1.8 : i) l'apprentissage supervisé, qui entraîne un modèle sur des

données d'entrée et de sortie connues (données étiquetées) pour prédire les résultats futurs. Ce dernier est divisé en deux types : le premier est l'apprentissage par régression, dont la sortie est continue, parmi ses principaux algorithmes, la régression linéaire, logistique et polynomiale. Le second type est l'apprentissage par classification, dont le résultat est discret, certains des principaux algorithmes utilisés sont : la machine à vecteurs de support (SVM), la forêt aléatoire (RF), la classification naïve bayésienne (NB), le réseau de neurones artificiels (ANN), etc. ii) L'apprentissage non supervisé (données non étiquetées) consiste à trouver des modèles cachés ou des structures intrinsèques dans les données d'entrée en partitionnant les données (analyse de cluster), parmi ses algorithmes : k-means, clustering hiérarchique, analyse en composantes principales (PCA), etc. [86-90].

### **Principe de base de l'apprentissage profond**

Une classe d'algorithmes d'apprentissage automatique connue sous le nom d'apprentissage profond ou deep learning (DL) utilise une cascade de couches multiples d'unités de traitement non linéaires pour extraire et transformer automatiquement les caractéristiques. La sortie d'une couche sert d'entrée à la couche suivante. Actuellement, DL regroupe les méthodes les plus efficaces et performantes dans la communauté de l'intelligence artificielle [91, 92].

## **1.3.2 Application de ML et DL à la classification LOS-NLOS**

### **Introduction**

Pour certaines applications de la modélisation de propagation indoor et outdoor, l'estimation du canal et la localisation, entre autres, les systèmes de communications doivent être capables d'estimer la nature du scénario à partir des informations du canal remontées par les utilisateurs UE (User Equipment). Cependant, il est crucial de disposer d'une classification LOS/NLOS avec de faibles ressources de calcul et des méthodes de détection simples, afin de ne pas accroître les temps de latence et les surcharges des signaux [58][93]. Le problème de la classification LOS/NLOS a été abordé selon différentes approches qui dépendent du type de statistiques du canal utilisées, ainsi que des algorithmes d'apprentissage (ML ou DL) mis en œuvre. Il est essentiel de disposer d'une base de données d'entrées pour développer un modèle ML en extrayant les caractéristiques et en les utilisant pour la phase d'apprentissage et de validation, ou un modèle DL en injectant directement ces entrées pour l'apprentissage afin de réaliser l'identification LOS et NLOS. La Figure 1.9 illustre les mécanismes. Une fois le modèle validé, il peut être utilisé pour catégoriser des scénarios en utilisant des données réelles dans une application en temps réel.

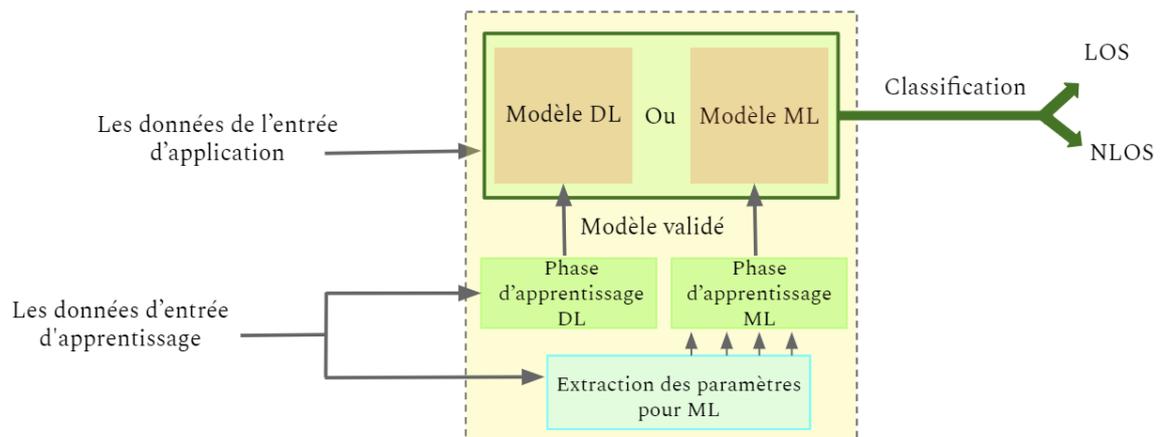


FIGURE 1.9 – Classification LOS NLOS à l'aide du ML/DL

### État de l'art de la classification LOS-NLOS à l'aide de l'IA

Les auteurs dans [94] ont utilisé une combinaison d'algorithmes ML supervisés SVM et le plus proche voisin (kNN), et non supervisés comprenant k-means et le modèle de mélange gaussien (GMM) pour aborder le problème d'identification LOS/NLOS pour les scénarios liés aux environnements ferroviaires à grande vitesse (HSR). Les caractéristiques utilisées incluent le pathloss  $K_{\text{rice}}$ ,  $\tau_{\text{rms}}$  et l'étalement angulaire  $AS$ . Les caractéristiques sont extraites des simulations de réponse impulsionnelle du canal (CIR). L'étude utilise deux ensembles de données : l'un généré par la plateforme QuaDRiGa, qui peut émuler la communication 5G sur plusieurs scénarios, et l'autre composé de données de canal HSR mesurées réellement contenant deux scénarios typiques. L'ensemble de données basé sur QuaDRiGa comprend 500 échantillons de données pour chaque scénario, tandis que l'ensemble de données mesurées réellement comprend 10 000 échantillons de données pour chacun des deux scénarios HSR. Les résultats de l'étude sont significatifs en démontrant l'efficacité des algorithmes ML dans l'identification des scénarios de propagation. Les approches kNN et SVM ont atteint une précision de plus de 90%, tandis que l'approche GMM a montré des résultats prometteurs avec des performances proches des algorithmes d'apprentissage supervisés. L'algorithme k-means a donné une précision d'environ 80%. Ces résultats suggèrent que les approches ML, en particulier kNN et GMM, peuvent être efficacement utilisées pour l'identification de scénarios de propagation à faible complexité et à haute précision.

Dans l'article [95], des techniques d'identification et de classification LOS/NLOS dans des scénarios indoor basées sur ML sont proposées, à partir des mesures de la puissance du signal reçu (RSS). Les algorithmes ont également été testés sur des modèles stochastiques et analytiques du canal. Les auteurs explorent plusieurs

caractéristiques des données RSS qui sont décrits ci-après, en utilisant deux algorithmes ML supervisés : machine à vecteurs de support (SVM), processus gaussien (GP). Après avoir collecté des données RSS à partir d'un réseau Wifi (3 360 000 échantillons RSS au total), chaque caractéristique est dérivée d'un ensemble de RSS entre 10 à 20 échantillons (en bande étroite) qui sont : L'écart type  $\sigma_t$ , La moyenne  $\mu_t$ , le coefficient d'aplatissement  $\mathcal{K}_t$ , le coefficient d'asymétrie  $S_t$ ,  $K_{rice}$  et la qualité d'ajustement  $\chi_t^2$ . Les résultats montrent que les techniques SVM et GP pour l'ensemble des caractéristiques  $\{\mu_t, k_{rice}, \chi_t^2\}$  (Cas statique) peuvent classifier les scénarios LOS ou NLOS avec une précision de l'ordre de 95%. Dans le cas dynamique, la précision de l'identification est réduite à 87% pour l'ensemble des caractéristiques  $\{\mu_t, \sigma_t, S_t\}$ . Les auteurs soulignent que la précision d'identification est affectée par la taille des données utilisées ainsi que de l'état dynamique ou statique de l'utilisateur.

Dans un autre papier [96], les auteurs proposent une identification du scénario LOS/NLOS en temps réel basée sur l'algorithme ANN pour les scénarios statiques à l'intérieur du bâtiment. À partir d'une base de données collectées sur l'état du canal CSI, plusieurs caractéristiques sont extraites tels que :  $\mu_t$ ,  $\sigma_t$ ,  $A$ ,  $S_t$ ,  $\mathcal{K}_t$ ,  $K_{rice}$ ,  $\sigma_t^2$  et  $\sigma_t^4$ . Le réseau neuronal artificiel est reconnu pour ses capacités d'approximation non linéaire, d'auto-organisation, de tolérance aux défaillances et d'auto-apprentissage. La plupart des modèles utilisent l'algorithme de rétro-propagation, qui est efficace pour classer les données. Il fonctionne en utilisant l'erreur de sortie pour estimer l'erreur de la couche précédente et en utilisant cette erreur de rétro-propagation pour estimer toutes les erreurs. Les résultats expérimentaux montrent que ANN avec rétro-propagation surpasse les approches existantes et que ANN atteint un taux de détection de 95% avec l'ensemble des caractéristiques  $\{\sigma_t, S_t, \mathcal{K}_t, \sigma_t^2, \sigma_t^4\}$  qui donnent la meilleure performance.

Le papier [97] traite de l'identification LOS/NLOS avec trois techniques ML différentes, SVM, forêt aléatoire (RF) et réseau neuronal artificiel (ANN) pour les systèmes MIMO  $8 \times 8$  dans les scénarios V2V à 5,9 GHz. L'analyse a été effectuée avec 28 876 réponses impulsionnelles du canal  $h(t, \tau)$ , entre des scénarios LOS et NLOS. Les caractéristiques d'apprentissage choisies sont : la puissance maximale reçue  $\max(|h|^2)$ ,  $\mathcal{K}$ ,  $S$ ,  $\Delta\tau_{rise}$ ,  $\tau_{RMS}$ ,  $K_{rice}$ ,  $AS_{AoA}$ ,  $AS_{AoD}$ ,  $\Delta\lambda$ ,  $PAS$  le spectre angulaire de la puissance et  $\Delta\varphi_D^{(t)}$ ,  $\Delta\varphi_A^{(t)}$  qui mesure la variation des AoD et AoA du trajet multiple ayant la plus forte amplitude dans une certaine fenêtre temporelle. L'auteur souligne que les résultats en termes d'identification sont plus sensibles aux choix des caractéristiques qu'à l'algorithme d'apprentissage utilisé. D'autre part, l'exploitation des caractéristiques angulaires a considérablement augmenté la précision de l'identification LOS/NLOS. L'identification LOS/NLOS basée sur l'algorithme RF donne les meilleures performances avec une précision de 94%, par rapport au SVM et à l'ANN qui donnent respectivement 90% et 92%.

Dans [98], les auteurs étudient la classification LOS/NLOS pour un système de communication IEEE 802.11ad à l'intérieur d'un avion, en appliquant l'algorithme RF. À l'aide de Matlab, un canal de propagation simulé à 60 GHz de fréquence porteuse et de 7 GHz de largeur de bande a été utilisé pour obtenir l'ensemble des données des réponses impulsionnelles. Les données collectées ont été divisées en 60% pour l'apprentissage et la validation et 40% pour les tests. Pour chaque réponse impulsionnelle, les caractéristiques sont :  $K_{rice}$ ,  $\tau_{RMS}$ ,  $\tau_{MOY}$ , le gain d'amplitude du canal  $G$ , les cinq premiers trajets identifiables  $P_5$ , le trajet dominant  $P_{dom}$  et la séquence de préambule reçue  $S_{pr}$ . Les auteurs testent les performances soit avec les caractéristiques issues de la réponse impulsionnelle, soit directement avec la séquence du préambule  $S_{pr}$  de type séquences de Golay (symboles bipolaires  $\pm 1$ ). Les résultats montrent que les caractéristiques  $\tau_{RMS}$ ,  $K_{rice}$  et  $\tau_{MOY}$  ne sont pas performantes, contrairement aux caractéristiques  $G$ ,  $P_5$  et  $P_{dom}$  qui donnent une précision d'estimation allant jusqu'à 95%. Mais si on utilise sans traitement particulier comme la caractéristique  $S_{pr}$ , ce qui évite les erreurs d'estimation du canal, les performances atteignent jusqu'à 90% pour un SNR faible de -10 dB.

Un autre article [99] a également utilisé l'algorithme RF pour étudier le problème d'identification LOS-NLOS, à l'aide des caractéristiques extraites des réponses impulsionnelles, d'un système de communication Indoor des bâtiments à une fréquence porteuse de 5,75 GHz et une largeur de bande de 25 MHz. SVM, Naïve Bayesian (NB) et RF sont les trois algorithmes ML utilisés. Pour les deux scénarios LOS et NLOS, les caractéristiques du canal ont été dérivées respectivement, de 6678 et 5352 estimations totales. Les caractéristiques suivantes ont été prises en compte :  $\mu$ ,  $\sigma$ ,  $\mathcal{S}$ ,  $\mathcal{K}$ . Après la phase d'apprentissage et de test, l'ensemble des caractéristiques  $\sigma, \mathcal{S}, \mathcal{K}$  avec l'algorithme RF donne les meilleures performances avec une précision d'identification de 97,3% et 95% pour les scénarios NLOS et LOS, respectivement. Les auteurs indiquent que l'algorithme RF est une approche plus efficace avec un temps de calcul et une complexité plus faibles par rapport aux algorithmes SVM et NB.

Dans cet article [100], une classification LOS-NLOS dans les systèmes de navigation par satellite (GNSS) est présentée en utilisant des algorithmes SVM et NN. Les auteurs se concentrent sur la corrélation du signal à la sortie GNSS, qui est une sortie fondamentale du traitement du signal GNSS, et proposent l'utilisation de cette sortie pour ce problème de classification LOS-NLOS. Basée sur des données expérimentales dans un environnement urbain, qui se compose des signaux de corrélation GNSS LOS et NLOS calculés à l'aide de 420 points de corrélation avec 3 caractéristiques : puissance du signal, les maxima locaux de la corrélation et la distribution du retard. Les résultats d'évaluation ont montré que l'algorithme NN surpassait SVM dans la détection NLOS. NN a atteint une précision de classification de 97,7% pour discriminer correctement les signaux NLOS. Cette précision indique

l'efficacité de cette approche basée sur ML proposée dans la classification LOS-NLOS dans les applications GNSS urbaines.

Dans ce papier [101], une technique ML supervisée utilise l'algorithme SVM pour la classification LOS/NLOS. En raison de la robustesse et de l'approche rigoureuse de la technique SVM, elle représente l'une des techniques de classification les plus populaires. Les données ont été collectées à l'aide d'un système ultra-large bande (UWB) pour un environnement à l'intérieur d'un laboratoire de recherche avec 1024 réponses impulsionnelles pour les deux scénarios. Après avoir obtenu cette base de données, les caractéristiques suivantes ont été extraites : le gain d'amplitude du canal  $G$ , le trajet dominant  $P_{dom}$ ,  $\Delta\tau_{rise}$ ,  $\tau_{MOY}$ ,  $\tau_{RMS}$ ,  $\mathcal{K}$ . Les résultats indiquent que l'ensemble des caractéristiques  $\{G, \Delta\tau_{rise}, \mathcal{K}\}$  fournit les meilleures performances avec une précision d'identification supérieure à 91%.

Dans cet article [102], un modèle de classification LOS-NLOS utilisant les signaux 5G dans les environnements aéroportuaires pour des scénarios intérieurs est présenté, en se basant sur les algorithmes de classification XGBoost et RF. Les données expérimentales et de simulation utilisées comprenaient trois types de modèles de canal 5G : Tapped Delay Line (TDL), Clustered Delay Line (CDL) et Winner II. Pour chaque modèle, 1000 échantillons aléatoires ont été générés, totalisant 5000 échantillons. 80% des données ont été utilisées pour l'entraînement et 20% pour les tests. Les résultats montrent que les méthodes de détection LOS basées sur l'apprentissage automatique surpassent significativement la détection basée sur des seuils définis par le modèle, avec jusqu'à 28% de meilleure performance. Plus précisément, les classificateurs XGBoost et SVM ont obtenu des précisions de détection élevées, XGBoost présentant le meilleur compromis entre précision et complexité. Les précisions de détection ont atteint jusqu'à 98% pour les canaux Winner II en simulation et 100% pour les données de mesure. Les canaux TDL sont notés comme étant les plus difficiles en termes de détection LOS, avec une précision de 96% et 100% respectivement pour les canaux CDL et Winner II. Cependant, l'algorithme RF n'a pas montré de promesses dans la détection LOS/NLOS en l'absence de grandes quantités de données d'entraînement.

Comme évoqué précédemment, les méthodes d'apprentissage supervisé nécessitent la disponibilité des données étiquetées. De plus, ces méthodes ne fournissent pas une classification générale pour de nombreux modèles et caractéristiques. Dans ce papier [103], une méthode non supervisée de classification sans données étiquetées a été proposée, appelée k-means. En utilisant un processus de sélection des caractéristiques du canal extrait de CIR tels que  $\mathcal{K}$ ,  $S$ , le trajet dominant  $P_{dom}$ , le gain d'amplitude du canal  $G$ ,  $\tau_{MOY}$ ,  $\tau_{RMS}$ . Il a été démontré que la stratégie k-means proposée surpasse les schémas non supervisés actuels évalués dans l'article en termes de précision de classification jusqu'à 95%.

Dans ce papier [104], les méthodes du DL pour la classification NLOS/LOS sont exploitées. (1) Le réseau neuronal convolutif (CNN), qui est un réseau neuronal artificiel utilisé pour l'analyse des pixels et le traitement et la reconnaissance des images. (2) L'algorithme mémoire à long terme (LSTM), une architecture de réseau neuronal récurrent (RNN) utilisée dans DL. Les LSTM ont des connexions de rétroaction ou (feedback), contrairement aux réseaux neuronaux à propagation directe. En employant les réponses impulsionnelles qui ont été utilisées directement comme entrée pour ces méthodes. Parmi les 35000 réponses impulsionnelles collectées en LOS et NLOS dans différents endroits intérieurs, 25000 échantillons de données ont été utilisés pour l'entraînement et la validation, le reste a été dédié aux données de test. La configuration du CNN avec deux couches de traitement convolutifs a atteint une précision optimale de 81 %. Pour la deuxième méthode, trois architectures ont été testées : LSTM, LSTM stacked, LSTM bi-directionnel. Le LSTM a obtenu la précision de classification la plus faible, soit 51 %. Alors qu'avec l'association du CNN, le CNN-LSTM stacked a obtenu une amélioration significative avec une précision de 82,14% comme le montre le tableau 1.6. Finalement, les auteurs ont démontré que le modèle combinant CNN et LSTM (CNN-LSTM) s'est révélé efficace pour la classification des scénarios LOS/NLOS et que l'utilisation conjointe de ces deux techniques a permis d'améliorer la précision de la classification par rapport à l'utilisation séparée de CNN ou de LSTM.

L'article [105] utilise des CNN pour améliorer la localisation et la classification des scénarios de communication Indoor, en se basant sur des données expérimentales brutes de CIR obtenues via des radios à bande ultra-large (UWB) et en comparant les performances avec des modèles ML. Une base de données comprenant des mesures dans divers environnements Indoor, avec 3000 mesures pour chaque condition de canal LoS et NLoS. Les résultats de classification montrent que les CNN surpassent les approches ML, avec une précision de détection allant jusqu'à 87,4%. Etendant ainsi leur application aux signaux 5G dans [106], avec une précision de classification comprise entre 91,5% et 93,3%.

Méthodes	Précision
LSTM	51,11 %
CNN+LSTM	81,56 %
CNN+ LSTM bi-directionnel	78,93 %
CNN+ LSTM stacked	82,14 %

TABLEAU 1.6 – Résultats des algorithmes.

Cependant, la plupart des études se concentrent sur les scénarios indoor, principalement pour la localisation des signaux UWB et WiFi. La communication

entre véhicules (V2V) n'a été prise en compte que dans une seule publication [97]. Aucun article n'aborde le scénario V2I, pourtant essentiel pour les communications véhiculaires 5G (ITS-G5 et C-V2X), afin de répondre aux exigences ITS.

### **Conclusion**

L'analyse approfondie de l'état de l'art concernant l'identification des scénarios de propagation LOS-NLOS par intelligence artificielle révèle l'utilisation variée d'algorithmes de machine learning (ML) et de deep learning (DL). Les algorithmes Random Forest (RF) et Support Vector Machine (SVM) sont fréquemment employés et affichent une bonne précision d'identification, atteignant jusqu'à 96%. Le SVM se distingue par son efficacité avec les données linéairement et non linéairement séparables, sa performance en classification binaire et sa capacité à gérer un grand nombre de caractéristiques. Toutefois, cette méthode est sensible au bruit et exige une normalisation des données. Le RF, quant à lui, est robuste face au bruit, performant avec des données déséquilibrées et capable de sélectionner automatiquement les caractéristiques pertinentes. Néanmoins, il peut être sujet à l'overfitting et sa complexité peut croître avec l'augmentation du nombre de classes. Les réseaux de neurones artificiels (ANN) sont aptes à traiter des données non linéaires mais requièrent un volume conséquent de données d'entraînement pour prévenir l'overfitting et sont complexes ainsi que difficiles à interpréter. Pour l'identification LOS/NLOS, l'utilisation des caractéristiques angulaires a significativement amélioré la précision. En somme, pour élaborer un modèle performant, il est crucial d'extraire adéquatement les caractéristiques et de collecter les données de manière appropriée, plutôt que de se focaliser exclusivement sur le choix de l'algorithme.

Algorithme	Caractéristique	Données	Scénario	Signal	Précision	Réf.
SVM, GP	$\mu_t, K_{\text{rice}}, \chi_t^2$	Expérimentale	Indoor	RSS	95% statique, 87% dynamique	[95]
kNN, SVM	$AS, K_{\text{rice}}, \tau_{\text{mis}}$	Expérimentale	Environnements ferroviaires à grande vitesse (HSR)	CIR	(kNN, SVM >90%)	[94]
k-means, GMM	$AS, K_{\text{rice}}, \tau_{\text{mis}}$	Expérimentale et Simulation	Environnements ferroviaires à grande vitesse (HSR)	CIR	(k-means 90%) (GMM > 89%)	[94]
RF, SVM, ANN	$\max( h_t^2 , \mathcal{K}, S, f_{\text{rise}}, \tau_{\text{RMS}}, AS_{\text{AoA}}, AS_{\text{AoD}}, \Delta\phi_D^{(t)}, \Delta\phi_A^{(t)}, K_{\text{rice}}, \Delta\lambda, PAS$	Expérimentale	Outdoor VZV	CIR	(RF 94%) (SVM 90%) (ANN 92%)	[97]
NN, SVM	Puissance du signal, Les maxima locaux de la corrélation, Distribution du retard	Expérimentale	Outdoor urbain	GNSS	(NN 97,7%) (SVM 95%)	[100]
XGBoost, SVM		Expérimentale	Indoor aéroport	CIR	(XGBoost 98%) (SVM 92%)	[102]
XGBoost, SVM		Simulation	Indoor aéroport	CIR	(XGBoost 100%) (SVM 92%)	[102]
ANN	$\sigma_t, s_t, \chi_t, \sigma_t^2, \sigma_t^4, K_{\text{rice}}$	Expérimentale	Indoor	CSI	95%	[96]
RF	$K_{\text{rice}}, G, P_5, P_{\text{pre}}, \tau_{\text{moy}}$	Simulation	Indoor	CIR	90%	[98]
RF, NB, SVM	$\sigma, \mu, S, \mathcal{K}$	Expérimentale	Indoor	CIR	(RF 96,34%) (NB 88,56%) (SVM 95,12%)	[99]
SVM	$f_{\text{rise}}, E, \mathcal{K}$	Expérimentale	Indoor	CIR	91%	[101]
k-means	$\mathcal{K}, S, P_{\text{max}}, E, \tau_{\text{MOY}}, \tau_{\text{RMS}}$	Expérimentale	Indoor	CIR	95%	[103]
LSTM	-	Expérimentale	Indoor	CIR	51,11%	[104]
CNN+LSTM	-	Expérimentale	Indoor	CIR	81,56%	[104]
CNN+LSTM bi-dir.	-	Expérimentale	Indoor	-	78,93%	[104]
CNN+LSTM stacked	-	Expérimentale	Indoor	CIR	82,14%	[104]
CNN	-	Expérimentale	Indoor	CIR	93,3%	[106]
MPL, SVM	Premier pic d'un signal RSS, $\tau_{\text{mis}}, \tau_{\text{MOY}}, \mathcal{K}, S$	Expérimentale	Indoor	CIR	(MPL 82,9%) (SVM 82,6)	[105]
CNN	-	Expérimentale	Indoor	CIR	87,4%	[105]

TABLEAU 1.7 – Résumé des résultats des algorithmes pour la classification LOS/NLOS avec les précisions et les caractéristiques utilisées.

### 1.3.3 Application de IA à Estimation du canal

#### Introduction

Dans un système de communication de multiplexage par répartition orthogonale de la fréquence (OFDM), le canal présenté souvent par sa fonction de transfert, est estimé à la réception par l'intermédiaire de pilotes insérées dans les données transmises. L'estimation de cette fonction de transfert  $\hat{H}(t, f)$  est obtenu grâce à plusieurs méthodes d'estimation de canal, notamment les moindres carrés (LS), ou encore les moindres carrés à erreur quadratique moyenne minimale (LSMMSE). Ces algorithmes supposent que la variation du canal entre le symbole pilote et les données soit linéaire, la non linéarité intervenant lors des variations rapides du canal. Dans ce contexte, les algorithmes DL ont été récemment exploités pour estimer le canal et ceci grâce à leurs performances et leur capacité à gérer de grandes quantités de données et en présence de forte mobilité.

#### État de l'art de l'estimation du canal à l'aide de l'IA

Cette revue [107] exploite les CIR à partir de données de simulation basées sur des modèles statistiques linéaires et non linéaires pour évaluer les performances de l'estimation de canal basée sur DL dans les systèmes de communication à entrée unique et à multiples sorties (SIMO) en utilisant des réseaux de neurones profonds avec la fonction d'activation unité linéaire rectifiée (ReLU). Une base de données de 20 000 échantillons pour l'entraînement et de 5 000 pour le test a été utilisée. Les auteurs présentent un estimateur DL basé sur des réseaux de neurones profonds entièrement connectés avec des fonctions d'activation ReLU. Comparée aux méthodes traditionnelles LS et LMMSE, l'estimation DL montre une performance supérieure. Pour les signaux linéaires, les estimateurs DL et LMMSE affichent une erreur quadratique moyenne (MSE) de  $10^{-2}$  à 20 dB SNR, surpassant LS. La performance de DL s'aligne presque avec LMMSE, indiquant une approximation étroite. L'erreur de DL diminue avec l'augmentation de la largeur du réseau ( $\tilde{d}$ ), soulignant l'importance de la dimension du réseau. Dans le cas non linéaire, DL égale MMSE avec un MSE de  $10^{-2}$  à 20 dB SNR, dépassant LMMSE qui est à  $10^{-1}$ . Cela démontre la capacité de DL à gérer les non-linéarités et à potentiellement égaler MMSE sous certaines conditions, tout en mettant en lumière les défis de la qualité des données d'entraînement et de la configuration du réseau.

Ce papier de [108] propose un algorithme d'estimation de canal basé sur un réseau neuronal convolutif (CNN), un type de réseau de DL pour des environnements de trains à grande vitesse. L'algorithme CNN a été entraîné en utilisant une base de données de simulation des fonctions de transfert  $H(t, f)$ . L'algorithme CNN classe les fonctions de transfert estimées en identifiant les différentes caractéristiques afin de choisir le canal le mieux estimé (canal correctement estimé) pour l'égalisation des signaux reçus (données) (figure 1.10).

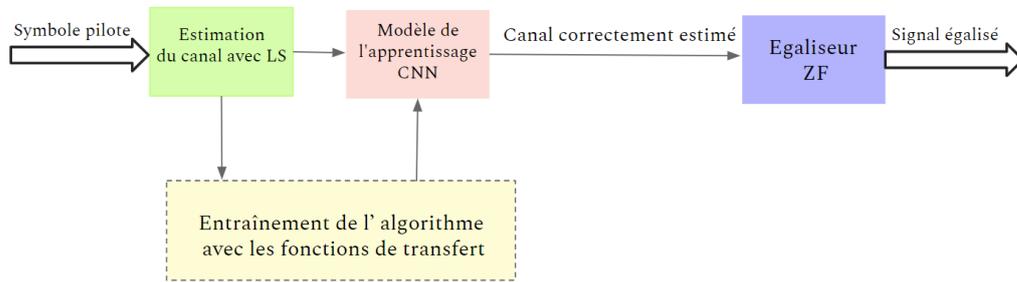


FIGURE 1.10 – Apprentissage profond pour l'estimation du canal

Afin de rendre les données de l'entrée  $H(t, f)$  compatibles avec le CNN, les parties réelles et imaginaires des fonctions de transfert ont été traitées séparément. Les principales couches d'un CNN pour cette estimation, comme le montre la Figure 1.11, comprennent : une couche d'entrée (1) utilisant les symboles pilotes, une couche de convolution (2) appliquant des filtres pour calculer les pondérations et le produit de l'entrée, une couche d'activation (3) calculant le biais des couches de convolution, une couche de pooling (4) pour réduire la complexité du réseau, et enfin, une sortie (5) fournissant un canal correctement estimé après entraînement, utilisé pour égaliser le signal. Pour illustrer la performance et l'efficacité du modèle DL proposé, la méthode ZF a été utilisée pour la comparaison. Les résultats de la simulation montrent que, pour un SNR de 15 dB, le modèle DL proposé avec 6 couches convolutives et 32 filtres permet de réduire le taux d'erreur binaire (TEB) de  $10 \times 10^{-2}$  avec la méthode ZF à  $6 \times 10^{-3}$  dans les mêmes conditions de canal.

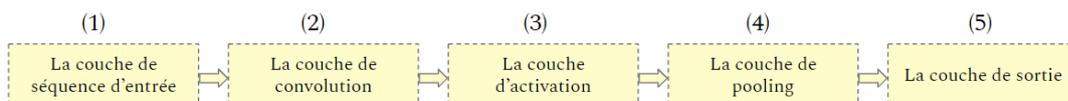


FIGURE 1.11 – Étapes de l'algorithme CNN.

Dans [109], les auteurs se concentrent sur l'intégration des méthodes de DL avec des approches statistiques traditionnelles LS et LMMSE pour améliorer la performance de l'estimation de canal dans les systèmes de communication sans fil basés sur des pilotes OFDM conformes aux normes 3GPP. L'algorithme DL utilisé est basé sur un réseau de neurones profonds interpolé augmenté de LS (LSiDNN), qui réduit considérablement le temps d'exécution et l'utilisation des ressources par rapport aux méthodes d'estimation de canal existantes. Basée sur la simulation des

réponses impulsionnelles du canal (CIR), une base de données pour l'entraînement des modèles DL comprend 10 000 trames OFDM pour chaque type de canal, avec une répartition de 80%/20% entre les ensembles d'entraînement et de validation, a été générée. Une base de données séparée de 1 000 trames OFDM est utilisée pour les tests. L'estimation de canal basée sur DL et LMMSE surpasse LS en termes de MSE et de TEB à divers SNR. L'algorithme LSiDNN proposé offre des réductions substantielles du temps d'exécution (88-90%) et de l'utilisation des ressources (38-85%) par rapport aux approches DL basées sur l'état de l'art étudiées. L'étude montre que LSiDNN offre un équilibre entre performance et complexité, ce qui en fait une solution prometteuse pour une estimation de canal efficace et performante dans les systèmes de communication sans fil.

Dans ce papier [110], un modèle de DL appelé ChanEstNet utilise un CNN pour extraire les vecteurs de caractéristiques et un LSTM pour l'estimation du canal. Le réseau ChanEstNet est divisé en deux phases : un apprentissage hors ligne pour l'apprentissage du modèle DL en utilisant une grande quantité de données recueillies issues du modèle de canal WINNER II, et une prédiction en ligne en utilisant comme entrée l'information d'état du canal (CSI). L'exploitation de l'ensemble des informations sur le canal lors de son apprentissage lui permet d'apprendre les caractéristiques du canal qui varient rapidement dans le temps et qui ne sont pas stationnaires. Pour la partie apprentissage, et comme le montre la Figure 1.12 (1), le réseau CNN est principalement utilisé pour extraire les valeurs des caractéristiques des séquences pilotes. (2) Un réseau 1D Maxpooling est principalement utilisé pour réduire le nombre de caractéristiques à estimer. (3) Le réseau LSTM est utilisé pour la prédiction. (4) En utilisant le Full-CNN pour réduire la sortie du LSTM. La partie de prétraitement est utilisée pour adapter le CSI, qui est un nombre complexe, au FCNN en entraînant les parties réelles et imaginaires séparément, puis en les concaténant à la fin pour avoir le canal estimé. Les résultats montrent que pour un rapport SNR de 10 dB, ChanEstNet permet d'obtenir un TEB de  $5 \times 10^{-3}$  alors que le TEB pour LS et LMMSE n'est que de  $8 \times 10^{-2}$  et  $10^{-2}$ , respectivement. Les auteurs font ressortir que la méthode proposée montre une amélioration significative des performances en ce qui concerne la faible complexité de calcul par rapport aux méthodes conventionnelles.

Avec un nombre faible de pilotes, les méthodes d'estimation du canal traditionnelles ne sont pas très efficaces dans des conditions réelles. De plus, il est difficile d'ajouter plus de pilotes à une trame tout en garantissant l'efficacité spectrale. Pour résoudre ce problème, de nombreux chercheurs utilisent la méthode DPA (Data Pilot Aided). Cette méthode utilise les sous-porteuses démodulées des symboles OFDM précédents pour estimer le canal pour le symbole OFDM actuel. Cependant, cette méthode n'est pas très efficace en présence de bruit et de variations rapides du canal. Pour répondre à ce défi, le papier [111] propose une estimation du

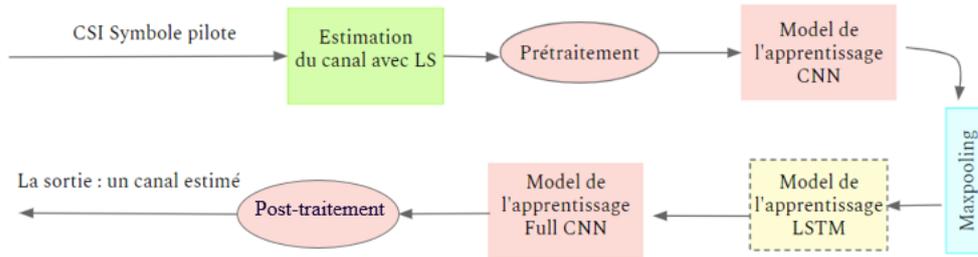


FIGURE 1.12 – Étapes de la méthode proposée.

canal basée sur un modèle DL qui combine la méthode DPA avec deux algorithmes : LSTM et un réseau de perception multicouche (MLP). Ce réseau neuronal suivra la variation du canal ainsi que l'atténuation du bruit. Trois étapes sont nécessaires pour l'estimation du canal à l'aide de ce modèle DL assisté par la procédure DPA comme le montre la Figure 1.13 : (1) l'estimation initiale du canal en obtenant le  $H$  estimé à l'aide de l'algorithme LS. (2) La compensation des erreurs en utilisant l'estimation initiale du canal dans le réseau LSTM-MLP, qui nous donne une nouvelle fonction de transfert  $\hat{H}$ . (3) La procédure DPA pour égaliser le signal transmis avec les nouvelles fonctions de transfert estimées  $\hat{H}$ . Les résultats de la simulation montrent, pour une vitesse de 150 km/h et un SNR de 25 dB, une baisse de TEB de  $5 \times 10^{-1}$  avec la méthode DPA à  $8 \times 10^{-2}$  avec le modèle proposé. Le modèle LSTM-MLP assisté par DPA est plus performant que la méthode DPA seule, en particulier dans des canaux à variation temporelle rapide.

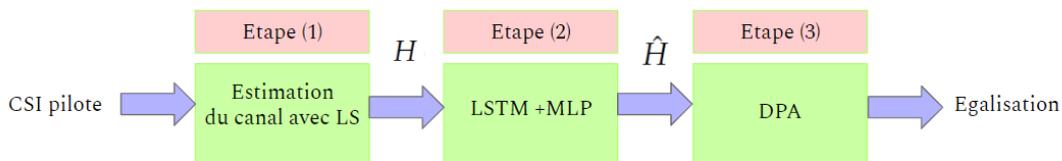


FIGURE 1.13 – Étapes de la méthodes proposé LSTM MLP

Dans [112], une approche d'estimation de canal basée sur le DL pour les systèmes d'architecture hybrides analogiques-numériques (HAD) à ondes millimétriques (mmWave) pour un réseau massif MIMO ( $32 \times 32$ ) à 4 GHz est présentée. Ce papier utilise un modèle de canal d'ondes millimétriques pour générer les matrices de canal  $H$  nécessaires à l'entraînement et à la validation du modèle. L'algorithme d'apprentissage bayésien parcimonieux (SBL) est déployé en un DNN pour mettre à jour les paramètres de variance gaussienne à chaque couche SBL, capturant ainsi

efficacement les structures complexes du canal. La matrice de mesure est également optimisée conjointement pour améliorer les performances. L'approche est étendue aux cas multi-blocs, exploitant la corrélation du canal dans le temps pour prédire de manière adaptative la matrice de mesure et faciliter la mise à jour des paramètres de variance. Les résultats de simulation montrent que les méthodes proposées surpassent significativement les approches existantes en termes de MSE normalisé (NMSE de  $10^{-2}$  à 10 dB de SNR) tout en réduisant la complexité, offrant une solution prometteuse pour l'estimation de canal dans les systèmes HAD MIMO massifs pratiques.

Dans [113], les auteurs étudient des méthodes d'estimation de canal basées sur le DL pour un système MIMO massif dans un environnement d'évanouissement de Rice. Ils développent un modèle de réseau neuronal optimisé qui conçoit intelligemment les pilotes du côté utilisateur avec un réseau neuronal (NN) à trois couches et estime efficacement les canaux en utilisant les pilotes améliorés avec un autre NN à trois couches du côté de la station de base. L'ensemble du processus est optimisé en utilisant une approche de bout en bout pour minimiser la fonction de coût MSE de l'estimation du canal. Ainsi, le réseau neuronal optimisé reconstruit le signal original à la station de base. Les résultats de la simulation confirment l'efficacité de la méthode proposée basée sur le DL pour les systèmes MIMO massifs dans les canaux d'environnements avec évanouissement de Rice et permettent d'augmenter le nombre d'antennes à l'émission jusqu'à 140 avec un MSE de 0.8.

Les auteurs de [114] ont étudié les performances d'un modèle DL basé sur le CNN pour l'estimation du canal dans les environnements de véhicules à haute mobilité utilisés dans la release 14 de 3GPP [20]. Ce modèle DL basé sur le CNN produit une nouvelle matrice d'estimation du canal  $\hat{H}$ , qui est ensuite appliquée au problème de l'égalisation. Les données sont générées à l'aide d'un simulateur avec une configuration de système C-V2X, avec un total de  $2 \times 10^4$  sous-trames générées; chaque sous-trame (1 ms) contient 14 symboles OFDM, dont 4 sont utilisés pour transporter les symboles DMRS (signal de référence de démodulation) et 10 pour transporter les données utilisateur. Les résultats du taux d'erreur bloc (TEBL) et de l'amplitude du vecteur d'erreur (EVM) montrent que l'architecture CNN à 4 couches convolutives proposée est plus performante que le schéma LS existant à des vitesses élevées. L'approche proposée permet, par exemple, pour une vitesse de 200 km/h et un SNR de 6 dB, de diminuer le TEBL et l'EVM respectivement à  $6 \times 10^{-2}$  et 3. Cela indique que l'architecture proposée offre la meilleure résistance aux changements rapides de canaux dans les scénarios de grande mobilité.

Un autre papier [115] propose une estimation du canal basée sur DL pour les

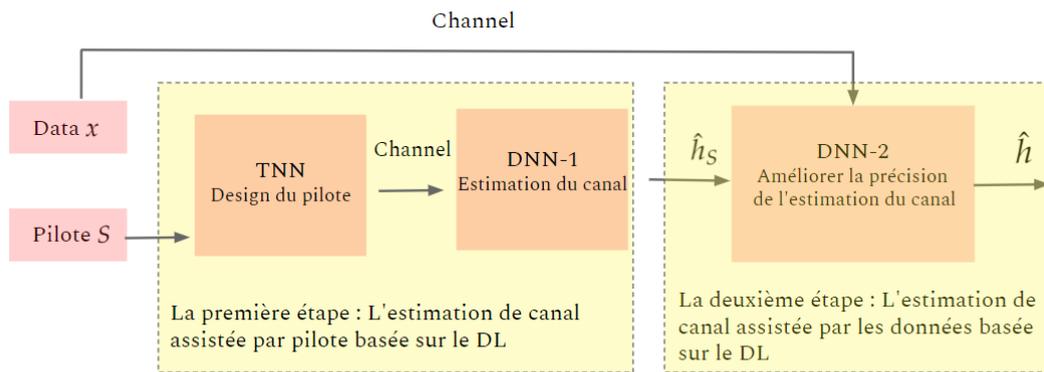


FIGURE 1.14 – Structure de la méthode proposée

systèmes massifs MIMO. La particularité de ce papier est l'utilisation d'un processus d'estimation en deux étapes. (1) La première étape est l'estimation de canal assistée par pilote basée sur un modèle DL en utilisant un réseau à deux couches (TNN) pour le design du pilote  $S$  et un réseau neuronal profond (DNN-1) pour l'estimation du canal  $\hat{h}_S$ , dans le but de minimiser l'erreur quadratique moyenne (MSE) de l'estimation du canal. (2) La deuxième étape est l'estimation de canal  $\hat{h}$  assistée par les données basée sur le DL pour améliorer la précision de l'estimation de canal en utilisant un autre DNN, appelé DNN-2 comme le montre la Figure 1.14. La phase d'apprentissage hors ligne a été effectuée sur la base de données de simulation (des pilotes  $S$ ). Les résultats montrent que la méthode proposée permet de diminuer le MSE de 0.73 pour la méthode classique à 0.42 pour la méthode proposée et augmenter l'efficacité spectrale de 22 bps/Hz pour la méthode classique à 33bps/Hz pour la méthode proposée. Le schéma d'estimation de canal proposé à l'aide de DL a prouvé des meilleures performances que la méthode classique d'estimation de canal.

Dans [116], les auteurs ont utilisé des modèles de deep learning (DL) et de machine learning (ML) pour développer une méthode d'estimation du canal duplex par séparation temporelle (TDD) pour les systèmes MIMO massifs. L'objectif est de réduire la surcharge de l'estimation du canal par les informations sur l'état du canal (CSI). Les auteurs emploient le DL pour concevoir et obtenir des pilotes CSI au lieu de recourir à l'estimation de canal conventionnelle. Ainsi, en diminuant la surcharge d'estimation du canal dans la structure de la trame, le schéma TDD basé sur ML proposé améliore les capacités de transmission des données. Deux architectures de DL ont été utilisées pour affiner la prédiction des CSI : (1) Un modèle CNN combiné à un prédicteur auto-régressif (CNN-AR) pour identifier l'évanouissement du canal. (2) Un modèle CNN combiné à un réseau neuronal récurrent (CNN-RNN) pour améliorer la prédiction des CSI. Les résultats démontrent que les modèles proposés (CNN-AR et CNN-RNN) permettent de réduire considérablement l'erreur quadratique moyenne normalisée (NMSE), par exemple, pour une fréquence Doppler

normalisée de 0.3, en passant de -5 dB à -10 dB, ce qui est plus performant qu'un simple modèle AR seul. Cela confirme que la méthode proposée basée sur le DL est plus efficace en termes de qualité d'estimation du canal que la méthode classique.

## Conclusion

En résumé, l'intégration des méthodes d'apprentissage profond (DL) avec des techniques d'estimation de canal traditionnelles, telles que LS et LMMSE, a démontré une optimisation significative des prédictions et une amélioration des performances du système en termes de TEB, MSE, TEBL, NMSE et EVM. Ces approches innovantes ont contribué à améliorer l'efficacité en réduisant la complexité de calcul par rapport aux méthodes conventionnelles, notamment dans les contextes de haute mobilité. Les algorithmes fréquemment employés incluent les réseaux de neurones convolutifs (CNN) et les réseaux de neurones à mémoire à long terme (LSTM), chacun présentant des avantages spécifiques. Les CNN excellent dans le traitement de grandes quantités de données et de modèles non linéaires, tandis que les LSTM sont optimisés pour gérer les séquences de données temporelles et mémoriser les informations sur le long terme. Néanmoins, ces modèles présentent également des défis, tels que leur complexité et le besoin substantiel de données d'entraînement. Un récapitulatif de l'état de l'art de l'estimation de canal à l'aide de l'IA est présenté dans le Tableau 4.1.

## 1.4 Conclusion du chapitre

La nouvelle génération de réseaux de communication mobile, la 5G, marque une révolution en facilitant le développement et l'intégration de systèmes de communication intelligents pour le secteur des transports intelligents (ITS), en particulier dans le domaine des communications véhiculaires. Face à la rapide variabilité du canal de propagation et aux exigences des ITS pour des systèmes fiables et robustes, la caractérisation multidimensionnelle et l'analyse de la variabilité du canal de propagation radio deviennent cruciales pour le développement de modèles ou la création de bases de données adaptées. En outre, l'incorporation de modèles d'intelligence artificielle (IA) dans la couche physique des systèmes 5G dédiés aux communications véhiculaires, notamment pour la classification des scénarios de propagation LOS/NLOS et l'estimation de canal, représente un défi majeur en raison de la variabilité multidimensionnelle (temps, fréquence, espace) du canal. Ce domaine de recherche reste encore peu exploré dans la littérature. La caractérisation multidimensionnelle du canal et l'intégration de l'IA sont essentielles pour concevoir des stratégies de transmission optimales, analyser la qualité des communications et simuler les performances du réseau, répondant ainsi aux exigences des ITS. Le chapitre suivant détaillera les approches de caractérisation d'un canal de propagation radio à l'aide d'un sondeur de canal Massive MIMO.

Algorithme	Type de Données	Type de canal	Scénario	Type de Signal	Performance	Ref.
DNN (ReLU)	Simulation Modèle linéaire Simulation Modèle non linéaire	SIMO	Général	CIR	MSE $10^{-2}$ à 20dB de SNR	[107]
CNN	Simulation	SISO	Environnements de trains à grande vitesse	$H(t, f)$	TEB : $6 \times 10^{-3}$ à 15dB de SNR	[108]
DNN (LSIDNN)	Simulation	SISO	Norme 3GPP	$H(t, f)$ pilotes OFDM	Réduction du temps d'exécution (88-90%) et de l'utilisation des ressources (38-85%)	[109]
CNN+LSTM	Simulation WINNER II	SISO	Scénario grande vitesse	$H(t, f)$ CSI	TEB : $5 \times 10^{-3}$ à 10dB de SNR	[110]
DPA à l'aide de LSTM+MPL	Simulation WINNER II	SISO	Suburbain V2X	$H(t, f)$	TEB : $8 \times 10^{-3}$ à 25dB de SNR	[111]
SBL+DNN	Simulation	Massive MIMO	5G hybride beamforming	matrice de canal Massive MIMO $H$	NMSE : $10^{-2}$ à 10dB de SNR	[112]
NN (ReLU)	Simulation	Massive MIMO	Évanouissement Rician	matrice de canal $H \in \mathbb{C}^{N_{Rx} \times N_{Tx}}$	40 antennes à l'émission avec un MSE de 0.3 à 20dB de SNR	[113]
CNN	Simulation	MIMO	C-V2X	matrice de canal $H \in \mathbb{C}^{N_{Rx} \times N_{Tx}}$	TEBL : $6 \times 10^{-2}$ et EVM : 3 à 6dB de SNR	[114]
TNN DNN	Simulation	Massive MIMO	Communication 5G	matrice de canal $H \in \mathbb{C}^{N_{Rx} \times N_{Tx}}$	MSE : 0.4 et 33bps/Hz d'efficacité spectrale pour 25 dB de SNR	[115]
CNN-AR	Simulation	Massive MIMO	Communication 5G	matrice de canal Massive MIMO $H$	NMSE de -18dB pour P=1	[116]

TABLEAU 1.8 – Résumé des performances des algorithmes d'apprentissage profond DL dans l'estimation du canal



## Chapitre 2

# Sondage de canal MIMO

### 2.1 Introduction

Un sondeur de canal MaMIMoSA a été développé dans le cadre de la thèse de [117] grâce à une collaboration étroite entre les universités de Lille/Laboratoire IEMN/Telice et Gand/Laboratoire INTEC. Il a été dimensionné pour permettre, entre autres, le sondage de canaux variant rapidement, typiquement pour des scénarios de communication sol-véhicule ou encore pour le sondage de canal MIMO ou MIMO distribué dédiés respectivement aux applications 5G et 6G. Pour ces dernières générations de réseaux massifs, une caractérisation multidimensionnelle (espace-temps-fréquence-polarisation) est nécessaire pour optimiser les modèles et/ou la structure des réseaux antennaires. Ce sondeur a été utilisé lors de campagnes de mesures en tunnel routier et en site suburbain (campus universitaire). Dans une première partie de ce chapitre, l'architecture simplifiée du sondeur est décrite, en mettant l'accent sur les différents modes du sondeur sélectionnés au regard des paramètres du canal à exploiter. Les campagnes de mesures en tunnel et en site suburbain sont détaillées respectivement dans les paragraphes 2 et 3, et elles sont illustrées par quelques résultats en bande étroite.

### 2.2 Architecture du sondeur

Le sondeur de canal est basé sur une architecture parallèle disposant ainsi de 8 ports de sortie à l'émission et 16 ports d'entrée à la réception. À titre d'exemple, une matrice de canal  $H_{16,8}(f)$  échantillonnée sur 818 points en fréquence dans 80 MHz de bande autour des fréquences 1,35 GHz ou 5,89 GHz est mesurée en 121,92  $\mu$ s. Un module de commutation d'antennes a été récemment ajouté à l'émission pour mesurer des matrices  $H_{16,64}(f)$  en 1 ms. Pour ne pas saturer la capacité de la mémoire tampon du sondeur, la périodicité d'envoi des signaux de sondage est ajustée en fonction de la vitesse de déplacement du récepteur et du nombre d'antennes de réception et d'émission. La trame des données transmises est constituée de blocs, chaque bloc comportant de 1 à 8 symboles OFDM de durée élémentaire  $T_B = 121,92 \mu$ s. Un symbole OFDM comporte 8192 sous-porteuses réparties sur 8 antennes, une sous-porteuse sur 8 étant affectée à une antenne. En mode commuté dédié au

Massive MIMO, un bloc contient au maximum 8 symboles OFDM dédiés chacun à un groupe de 8 antennes. Une trame de durée  $T_F$  est composée d'un préambule pour la synchronisation des signaux suivi de  $N_B$  blocs, le temps entre deux blocs successifs étant  $T_{IB}$ . La structure de la trame des signaux transmis en mode commuté est illustrée (Figure 2.2). Les applications envisagées imposent certaines contraintes sur le choix des paramètres de la trame et sur le nombre d'antennes à l'émission et à la réception. Par exemple, pour la mesure du temps de stationnarité,  $T_{IB}$  doit être choisi faible et  $T_F$  grand, limitant ainsi le nombre d'antennes à la réception. La puissance de sortie est de 1 Watts à 1,35 GHz et à 5,89 GHz.

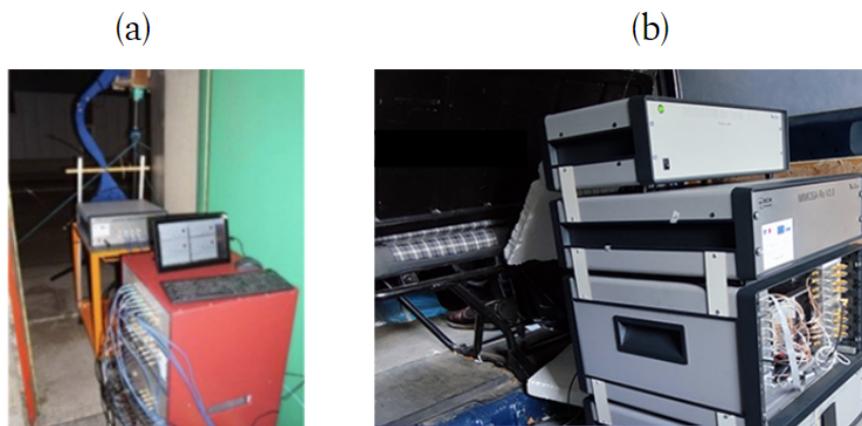


FIGURE 2.1 – Sondeur de canal : émetteur fixe (a) et récepteur mobile (b)

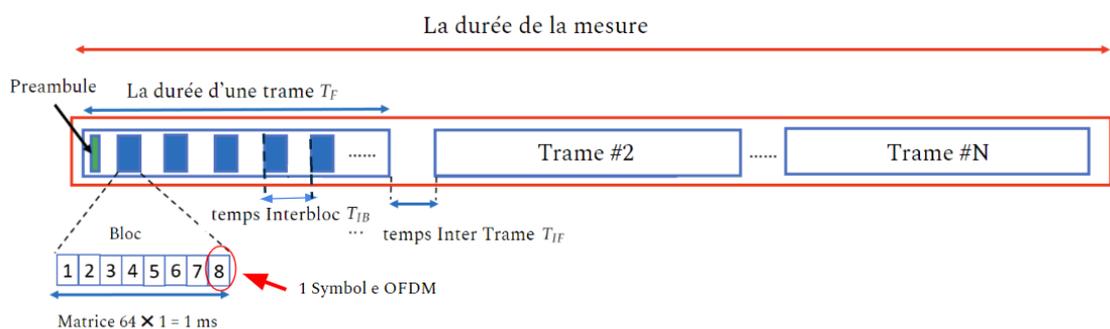


FIGURE 2.2 – Structure de la trame des séquences transmis

### 2.3 Les réseaux d'antennes

À 1,35 GHz, le réseau d'émission est un réseau ULA orienté suivant  $oz$  (Figure 2.3.(a)) dont les éléments sont des patchs bipolarisés. En fonction de leur orientation,

ils sont polarisés verticalement (V) ou horizontalement (H) ou à  $+45^\circ$  (ou D) et  $-45^\circ$  (ou A). L'ouverture des patchs est en moyenne de 60 degrés. Pour les expérimentations en tunnel, à l'émission les 8 éléments sont alignés verticalement. Deux symboles OFDM sont successivement transmis, respectivement sur le port 1 et 2 des patchs. Les polarisations des éléments de l'ULA sont donc successivement VDVDVDVD puis HAHAAAAA. À la réception, un réseau ULA identique au réseau d'émission est perpendiculaire à la direction de déplacement. Il est positionné à l'arrière d'une camionnette. La Figure 2.3 présente les 2 réseaux installés pour les mesures en tunnel. On définit la polarisation XY, avec X la polarisation de l'antenne Tx et Y la polarisation de l'antenne Rx. Il faut noter que grâce à la structure du réseau d'émission et de réception, 16 configurations de polarisation peuvent être étudiées simultanément. À 5,89 GHz, le réseau d'émission dont une photo est donnée dans Figure 2.4.(a) est un URA (10,10) dont les éléments sont des patchs polarisés verticalement. Seules les 64 antennes au centre du réseau sont connectées au sondeur. Le réseau de réception (la Figure 2.4.(b)) est un ULA dont les 8 éléments sont des antennes omnidirectionnelles.

	1,35 GHz	5,89 GHz
Tx	ULA : $dy = 3\lambda/2$	ULA : $dx=dy=\lambda/2$
Rx	ULA : $dx = 3\lambda/2$	ULA $dx = 1,85\lambda$

TABLEAU 2.1 – Distance entre les antennes Tx et Rx à 5,89 GHz et 1,35 GHz

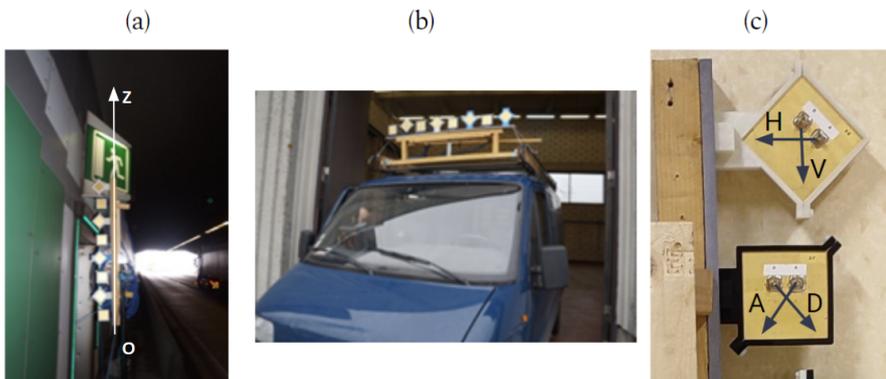


FIGURE 2.3 –  $F=1,35$  GHz : polarisation V, H, D, A du patch (c), réseau d'émission à 4 ports de sortie par polarisation (a) et réseau de réception à 4 ports d'entrée par polarisation en tunnel (b).

N° Configuration	$(N_{Rx} - N_{Tx})$	$F_c$ (GHz)	$T_b$ (ms)	$N_B$	$T_{IB}$ (ms)	N°de points par trame	$T_F$ (ms)	Application	Trafic
1	(1-1)	1,35	0,121	3075	0,975	3075	2999,23	(a),(b)	Important
2	(1-4)	1,35	0,243	256	1,95	256	499,4	(a),(b),(c),(d),(e)	Faible
3	(1-4)	1,35	0,243	256	1,95	256	499,4	(a),(b),(c),(d),(e)	Important
4	(4-4)	1,35	0,243	256	1,95	256	499,4	(c),(d),(e)	Faible
5	(4-4)	1,35	0,243	32	1,95	256	499,4	(c),(d),(e)	Important
6	(4-4)	1,35	0,243	32	15,6	32	499,4	(c),(d),(e)	Important
7	(1-8)	5,89	0,121	6150	0,48	6150	2999,32	(a),(b),(d),(e)	Vide
8	(1-8)	5,89	0,121	6150	0,48	6150	2999,32	(a),(b),(d),(e)	Important
9	(8-8)	5,89	0,121	128	3,9	128	499,4	(d),(e)	Moyen
10	(4-64)	5,89	0,975	32	15,6	32	499,4	(d),(e)	Trafic
11	(4-8)	5,89	0,121	1024	0,48	1024	499,4	(d),(e)	Moyen
12	(1-64)	5,89	0,975	512	0,975	512	499,4	(d),(e)	Fluide

TABLEAU 2.2 – Tableau de synthèse des configurations utilisées. (a) Stationnarité, (b) Doppler, (c) Dépolarisation, (d) Corrélation, (e) Pathloss.

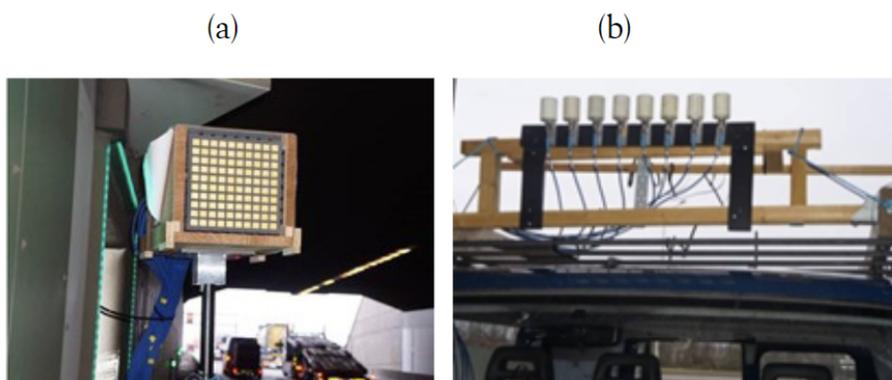


FIGURE 2.4 – 5,89 GHz : réseau de 64 antennes illuminant le tunnel (a), la réception se faisant sur 8 antennes omnidirectionnelles dans le plan azimut, placées sur le toit du véhicule (b).

## 2.4 Description des campagnes de mesure et analyse bande étroite

### 2.4.1 Le tunnel Beveren à Anvers

Dans le port d'Anvers, la voie rapide R2 passe sous un bras de l'Escaut par deux tunnels identiques de longueur 1 km, dédiés chacun à un sens de circulation. Dans chaque tunnel de dimensions transversales 10 m x 5 m (Figure 2.5.(a)), la circulation, principalement de camions, s'effectue sur deux voies. La Figure 2.5.(b) montre la position du réseau d'émission situé à la fin du tunnel à proximité de la voie de circulation de gauche. En fonction de l'état du trafic et des applications envisagées, la vitesse de déplacement de la camionnette varie de 60 km/h à 100 km/h. Le tableau 2.2 rassemble les configurations retenues pour la caractérisation du canal, le trafic est évalué visuellement à partir d'un enregistrement vidéo de chaque passage.

### 2.4.2 Analyse bande étroite

#### Qualité du signal mesuré

Avant toute analyse des matrices du canal, il est nécessaire de vérifier la qualité des signaux reçus en termes de SNR. À titre d'exemple, la configuration 2 est choisie pour illustrer la démarche suivie. Les fonctions de transfert ainsi que la puissance du bruit sont extraites des signaux mesurés [117]. La Figure 2.6 représente la puissance de bruit et la puissance moyenne relative du signal pour les polarisations VV et VH en fonction de la distance Tx-Rx. Avec une atténuation de 25 dB sur 900 m, la puissance moyenne décroît de manière similaire en VV et en VH. La puissance de bruit, de l'ordre de -75 dBm au début du tunnel, augmente jusqu'à -60 dBm à proximité de l'émetteur, ceci est dû au contrôle automatique de gain du système qui régule la puissance du signal à l'entrée du convertisseur analogique-numérique par l'ajout ou non d'atténuateurs.

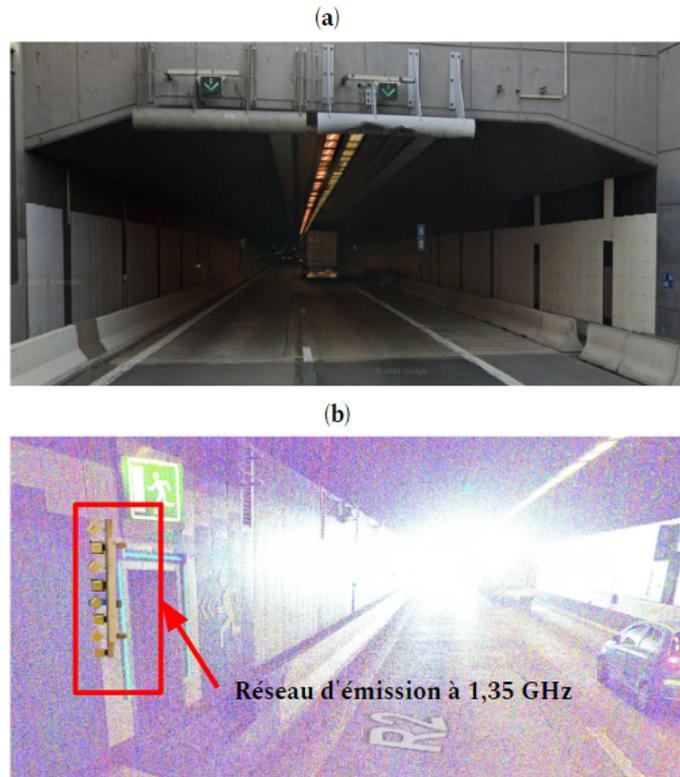


FIGURE 2.5 – (a) Entrée du tunnel Beveren (b) Réseau d'émission à la sortie du tunnel à gauche

Le SNR déduit de ces valeurs est tracé Figure 2.7 sur la totalité du parcours, le SNR en VV et VH varie respectivement de 28 dB à 15 dB et de 25 dB à 5 dB.

### Affaiblissement de propagation

Le modèle empirique d'atténuation de la puissance (PL) en fonction de la distance Tx-Rx est donné par l'expression :

$$PL(d)_{dB} = PL(d_0)_{dB} + 10n \log_{10} \left[ \frac{d}{d_0} \right] , \quad (2.1)$$

avec  $n$  l'exposant d'affaiblissement, qui vaut 2 en espace libre, et  $PL(d_0)_{dB}$ , l'atténuation mesurée à une distance de référence  $d_0=80$  m, distance correspondante à l'arrêt des mesures.

La courbe expérimentale de Pathloss (PL) de la Figure 2.8 a été obtenue à partir de la configuration 2 en polarisation VV. Cette courbe est comparée aux valeurs théoriques obtenues avec un modèle déterministe basé sur la théorie des images. Les valeurs de  $n$  expérimentale et théorique déduites par régression linéaire sont respectivement de 2,12 et 1,95. Les résultats obtenus dans cette étude sont en accord

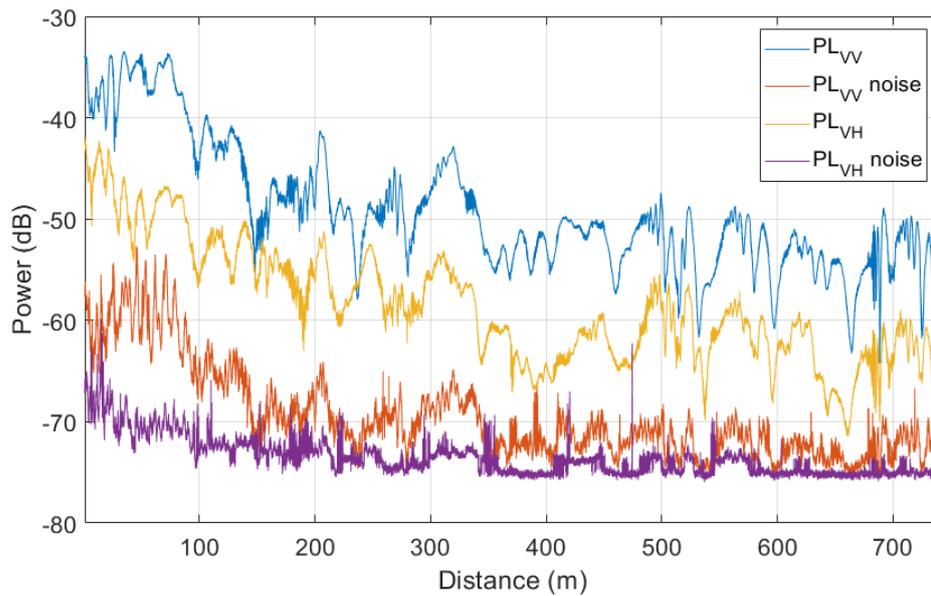


FIGURE 2.6 – Puissance moyenne relative du signal et puissance de bruit (Configuration 2) en fonction de la distance Tx-Rx. Polarisations VV et VH.  $F=1,35$  GHz

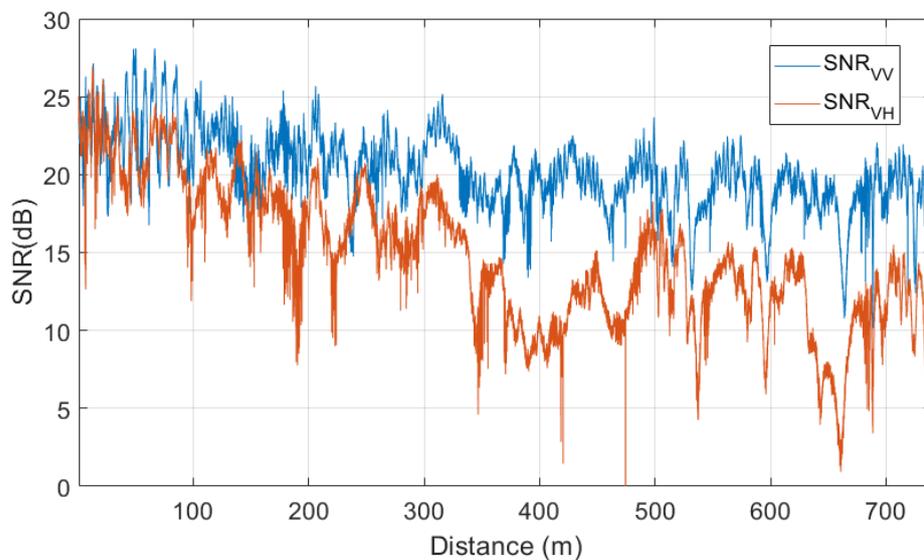


FIGURE 2.7 – SNR en dB en fonction de la distance Tx-Rx. Polarisations VV et VH.  $F=1,35$  GHz

avec les résultats de recherches antérieures portant sur des scénarios similaires avec une vue directe (LOS) [118-120].

Ces courbes à faible trafic sont comparées à celles des configurations 1, 3 et 6 obtenues à trafic important en polarisation VV. L'affaiblissement de propagation PL a été calculé comme indiqué sur la Figure 2.9 et a été comparé au modèle théorique précédemment mentionné. On peut remarquer qu'à une distance de  $d=d_0=80$  m, l'atténuation en présence de différentes densités de trafic est comprise entre 36 et 39

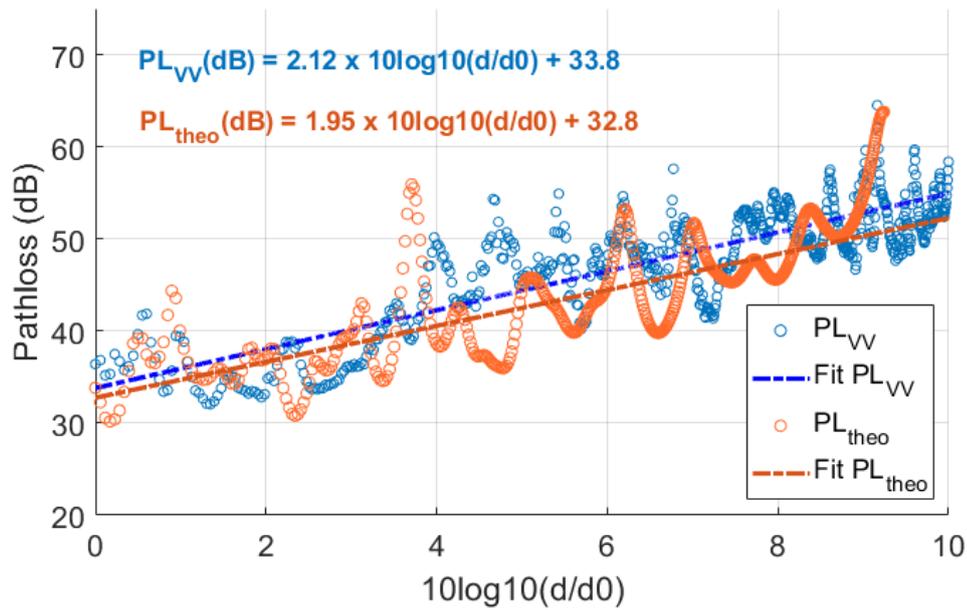


FIGURE 2.8 – Pathloss parcours réel (Configuration 2) et théorique

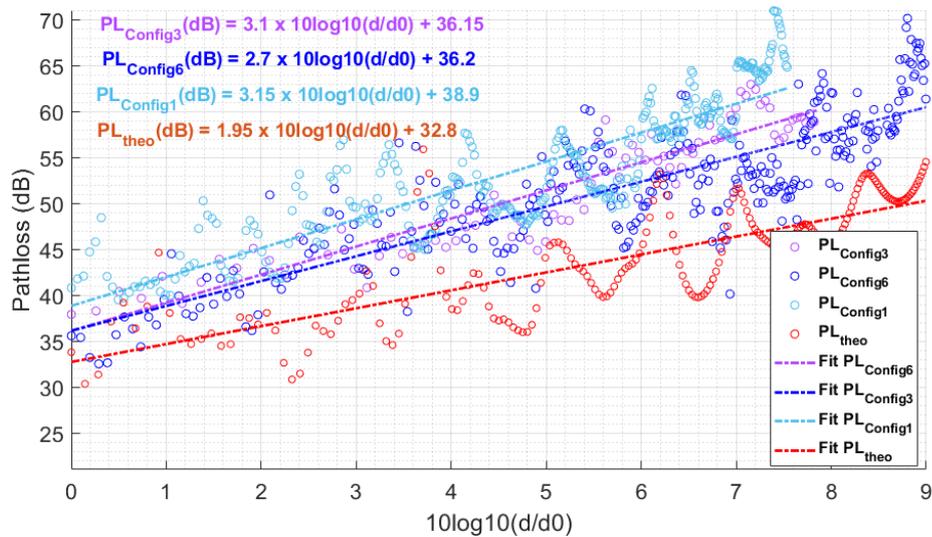


FIGURE 2.9 – PL trafic important avec les configurations 1, 3 et 6, trafic important polarisations en VV et PL théorique tunnel vide

dB pour les 3 configurations, soit plus élevée de 3 à 5 dB par rapport au modèle théorique. Il est également constaté que le facteur  $n$  est compris entre 2,70 et 3,15 pour les trois configurations, ce qui est supérieur au modèle théorique pour un tunnel vide où  $n$  vaut 1,95.

Les analyses de PL ont été effectuées pour un scénario de trafic important avec différentes configurations de polarisation en VV, HH, DD et AA (Configuration 3), et comparées au modèle théorique en polarisation VV d'un scénario de tunnel vide. Les résultats obtenus en présence de trafic (Figure 2.10) montrent que le facteur

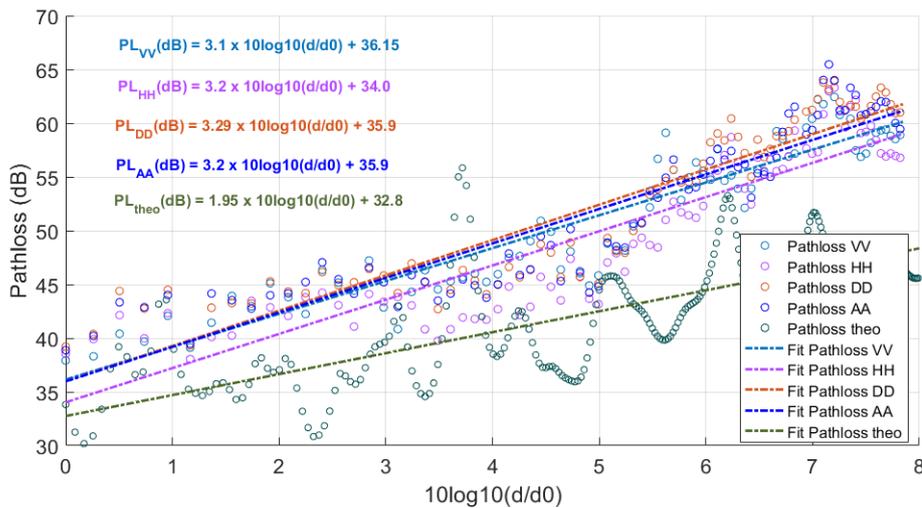


FIGURE 2.10 – PL Traffic important configuration 3 avec différentes polarisations et PL théorique tunnel vide

d'affaiblissement de propagation  $n$  en tunnel à trafic important pour les 4 polarisations est supérieur à celui du modèle théorique. Ceci montre l'impact de la densité du trafic sur l'affaiblissement du signal PL. Cependant, les valeurs de  $n$  pour les 4 polarisations présentent une variation minimale, allant de 3,1 à 3,3, tandis que  $n$  théorique est de 1,95. La variabilité du facteur  $n$  peut être due à plusieurs facteurs, notamment la structure du tunnel, la densité du trafic, les antennes utilisées pour chaque configuration et la position des antennes dans le tunnel.

Les fonctions de répartition empiriques (ECDF) du PL ont été analysées pour les quatre polarisations dans les deux scénarios de trafic : faible (configuration 2) et important (configuration 3). La Figure 2.11 montre que pour chaque polarisation, l'impact du trafic devient significatif à partir de la valeur médiane, sauf pour la polarisation HH. Cette différence de 3 à 6 dB observée après la valeur médiane correspond aux mesures effectuées entre 200 et 300 m, comme le montre la Figure 2.10.

Fréquence	Scénario de trafic	Atténuation à d0 (dB)	Facteur exponentiel (n)
1,35 GHz	Faible	33,8	2,12
	Important	36,15	3,1
5,89 GHz	Faible	47,9	1,72
	Important	41,66	2,48

TABLEAU 2.3 – Comparaison de l'influence du trafic à 1,35 GHz/5,89GHz sur le Pathloss  $d_0=80$  m

Les résultats de PL à 1,35 GHz ont été comparés à ceux de 5,89 GHz et sont présentés dans le tableau 2.3. Ce tableau donne les valeurs d'atténuation à  $d_0 = 80$  m et le facteur  $n$  pour les deux scénarios de trafic : faible (configuration 2), important (configuration 3) à 1,35 GHz et faible (configuration 7), important (configuration 8) à

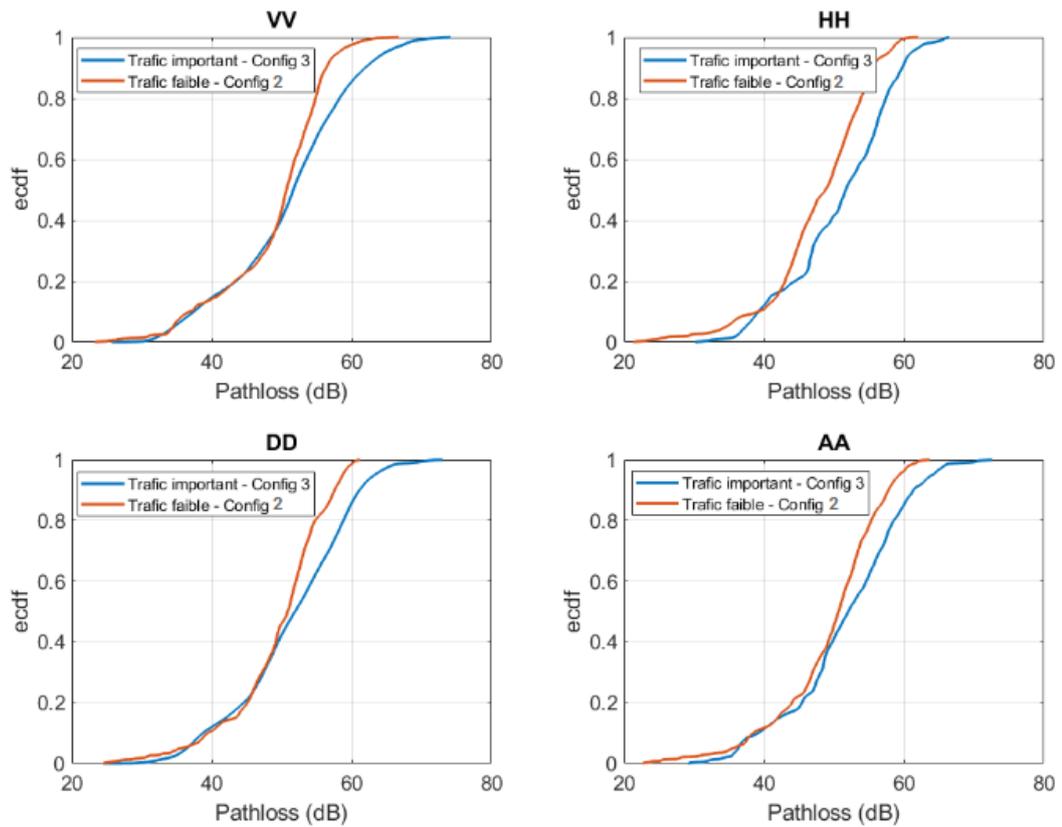


FIGURE 2.11 – ECDF du PL trafic important et faible avec différentes polarisations

5,89 GHz. On observe que l'atténuation à  $d_0$  est 15 dB plus importante à 5,89 GHz qu'à 1,35 GHz, ceci étant dû à la surface effective de l'antenne qui diminue lorsque la fréquence augmente. Le facteur d'atténuation exponentiel ( $n$ ), en trafic faible varie de 2,12 à 1,72 et en trafic important de 3,1 à 2,48 respectivement à 1,35 et 5,89 GHz. Toutefois, pour les deux fréquences, le facteur  $n$  est plus élevé dans un scénario de trafic important que dans un scénario de trafic faible.

### Discrimination de polarisation croisée (XPD)

La discrimination de polarisation croisée (XPD) mesure l'interférence de dépolarisation en comparant le gain du signal (désiré) dans la polarisation souhaitée et le gain du signal (indésirable) dans la polarisation orthogonale. Un XPD élevé indique une faible dépolarisation, donc moins d'interférences. Pour un canal de transmission à une polarisation choisie, le XPD est donné par le rapport linéaire entre le gain moyen de la fonction de transfert  $H(f)$  de la composante copolaire (copolar) et le gain moyen de la composante croisée orthogonale (crosspolar).

$$XPD(dB) = 10 \log_{10} \left( \frac{E\{|H(f)^{copolar}|^2\}}{E\{|H(f)^{crosspolar}|^2\}} \right) \quad (2.2)$$

XPD peut être employé dans les canaux MIMO afin de quantifier l'efficacité de la diversité de polarisation. De plus, il peut fournir un degré de liberté supplémentaire (DoF) pour le multiplexage spatial des signaux.

La Figure 2.12 présente l'ECDF du XPD en dB pour les quatre polarisations (VV, HH, DD, AA) pour deux scénarios de trafic (faible Configuration 2 et important Configuration 3). Dans le scénario à trafic important, il est remarqué que les médianes des XPD en VV et HH sont respectivement à 10 dB et 11 dB, ce qui indique une faible dépolarisation contrairement aux polarisations en DD et AA dont les médianes sont proches de 0 dB. De même, dans le scénario à faible trafic, la polarisation HH présente une faible dépolarisation par rapport aux autres polarisations.

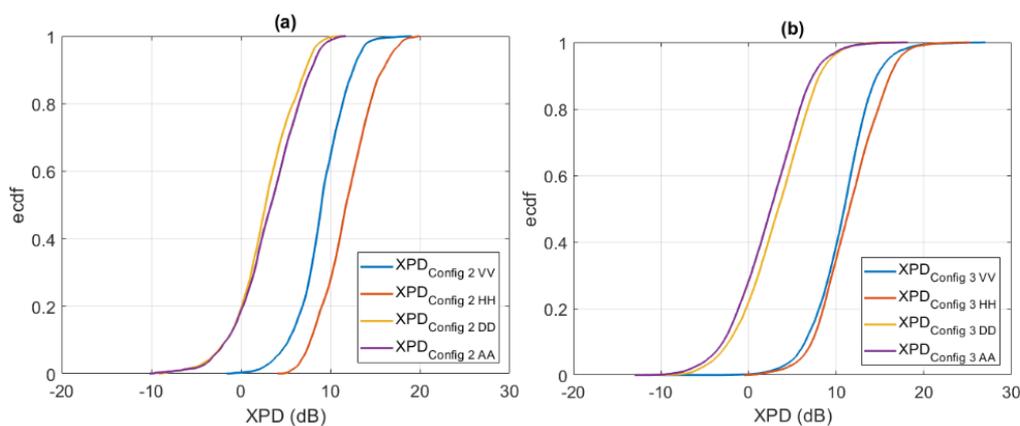


FIGURE 2.12 – ECDF XPDs pour (a) un trafic faible et (b) pour un trafic important, avec les différentes polarisations

### 2.4.3 Environnement suburbain cité scientifique-université Lille

Ces campagnes de mesures en milieu suburbain ont été menées sur différents secteurs du campus de l'Université de Lille, et dans diverses configurations présentées dans le tableau 2.4, à la fréquence de 5,89 GHz. L'environnement suburbain se compose de bâtiments de 3 à 4 étages, espacés de 5 à 10 mètres, et d'une végétation dense comprenant des arbres atteignant jusqu'à 30 mètres de haut. La circulation sur les routes est relativement faible, avec une vitesse maximale de 50 km/h.

L'émetteur (Tx) est un réseau carré massif de 64 antennes, installé à une hauteur de deux mètres au-dessus du sol. Les récepteurs (Rx) utilisent entre 1 et 8 antennes omnidirectionnelles, en fonction des objectifs de l'étude. Ces récepteurs sont placés sur le toit d'une camionnette, comme le montre la Figure 2.13.



FIGURE 2.13 – Sondeur MaMIMOSA Tx et Rx, le réseau d’antennes massif en émission et un réseau ULA en réception.



FIGURE 2.14 – Vue de dessus de la campagne de mesure pour la stationnarité dans un environnement suburbain

Pour évaluer la stationnarité spatiale, une seule antenne Rx a été utilisée. Dans cette configuration (configuration 30), la camionnette s’éloignait du Tx avec une vitesse de 40 km/h, comme illustré dans la Figure 2.14, de sorte que la visibilité directe (LOS) était toujours présente. La structure de la trame dans la Figure 2.2 est utilisée pour mesurer la fonction de transfert variant dans le temps  $H_{l,k}(t, f)$  du canal

radio. Une matrice mMIMO  $64 \times 1$  (un bloc) est mesurée en  $\sim 1$  ms et la durée de la trame  $T_F$  est de 5,504 s. L'étalement Doppler maximal est de  $\pm 250$  Hz (vitesse maximale  $\sim 46$  km/h) avec une résolution Doppler de 0,181 Hz. Cinq trames ont été mesurées lors de la campagne de mesure, donnant un temps de mesure total de 27,18 s # 300 mètres et un total de 13 760 blocs (c'est-à-dire, 880 640 fonctions de transfert).

Afin d'obtenir des valeurs fiables d'estimation de  $T_s$  pour l'ensemble du réseau, les SNR ont été calculés. Comme le montre la Figure 2.15, le niveau de SNR diminue de manière continue à mesure que la camionnette s'éloigne de l'émetteur (Tx), avec des valeurs allant de 40 à 10 dB. Le SNR, ramené à chaque élément du réseau, est spatialement uniforme, avec un écart type moyen de 3 dB et une médiane des écarts types de 1,7 dB. Cette valeur indique qu'aucun effet de masquage significatif ne s'est produit pendant le scénario de mesure. Une 2ème série de mesures est réalisée pour des applications de détection de scénario LOS et NLOS, 8 antennes omnidirectionnelles sont utilisées à la réception. Les vitesses de déplacement du Rx sont comprises entre 20 et 40 km/h. Le Tx a été installé dans trois endroits différents afin d'avoir une diversité d'environnements suburbains (Figure 2.16). La puissance transmise par chaque antenne a été maintenue à 25 dBm pendant toutes les campagnes de mesures.

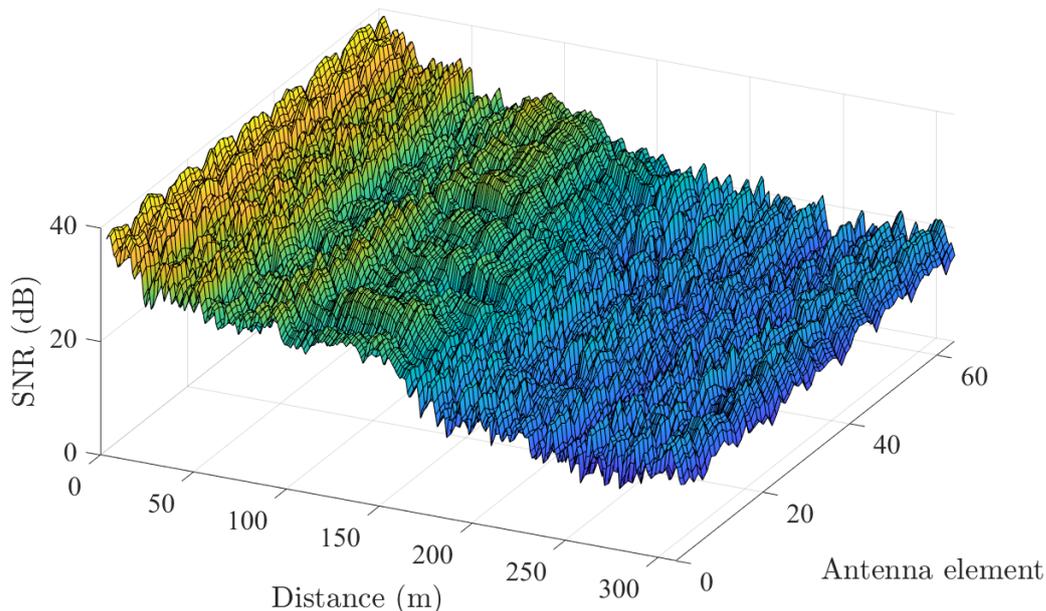


FIGURE 2.15 – SNR en fonction de l'antenne et la distance Tx-Rx (Configuration 30)

La Figure 2.17 (a) présente l'évolution de la puissance du signal et du bruit (en dB) en fonction de la distance Tx-Rx pour un scénario mixte estimé visuellement LOS-NLOS (configuration 27). Le SNR moyen est tracé (Figure 2.17 (b)). Le récepteur

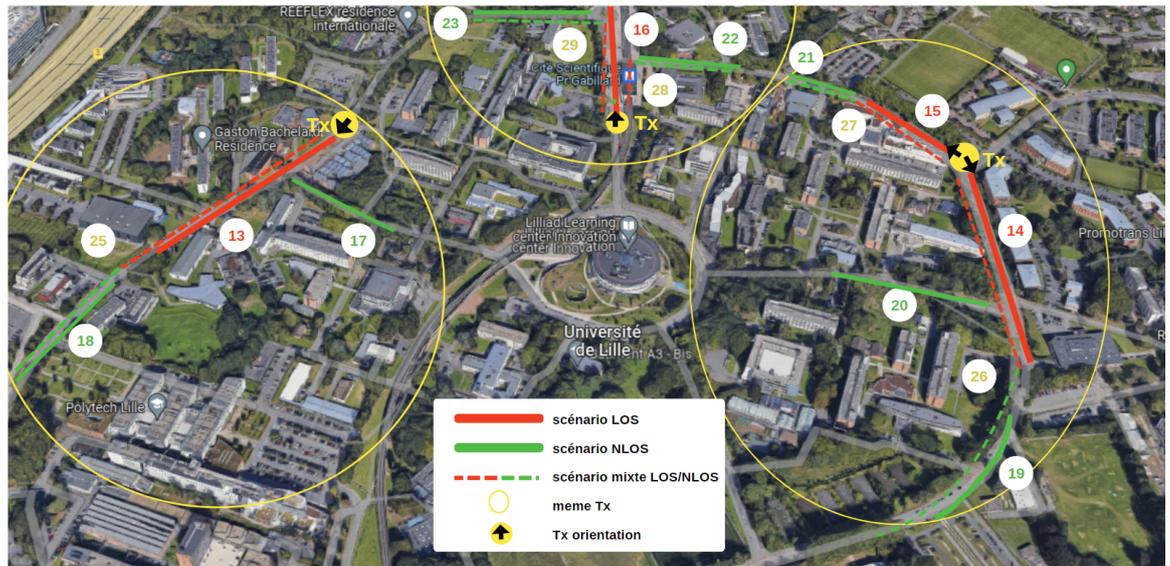


FIGURE 2.16 – Vue de dessus de l’environnement suburbain des campagnes de mesure

est en visibilité directe de l’émetteur, ce qui correspond à un scénario LOS, avec une puissance variant entre -50 dB et -70 dB. Une chute brutale de la puissance reçue est observée à 120 m, où le trajet direct est obstrué. Sur la fin du parcours, le scénario devient NLOS. Le bruit, quant à lui, varie de -70 dBm à moins de -90 dBm pour le cas LOS, et de moins de -90 dBm pour le cas NLOS. Sur les 120 premiers mètres associés au scénario LOS, le SNR moyen est de 17 dB, tandis que dans le cas NLOS, le SNR diminue de 15 dB à 5 dB. Les écarts types associés sur les 64 valeurs de SNR aux scénarios LOS et NLOS sont respectivement de 3 dB et 7 dB.

Sur la Figure 2.18, les cartographies couleur présentent le SNR calculé pour la première antenne de réception et les 64 antennes d’émission, respectivement, pour une distance de 40 mètres (scénario LOS) et 150 mètres (scénario NLOS). On observe d’importantes variations du SNR en NLOS par rapport au LOS.

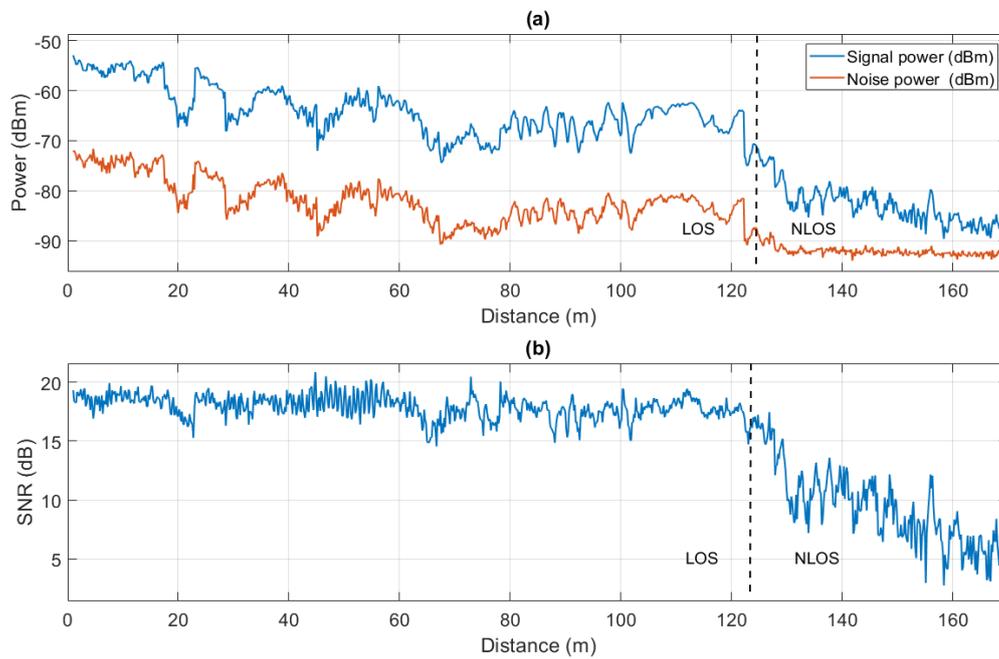


FIGURE 2.17 – Puissance signal et du bruit avec le SNR en fonction de la distance pour un scénario suburbain LOS et NLOS, configuration 27

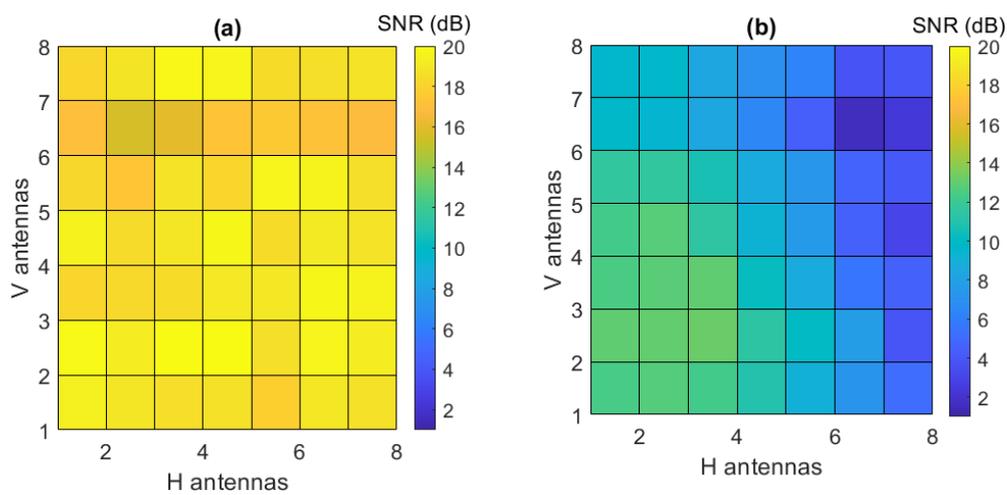


FIGURE 2.18 – Cartographie du SNR pour la première antenne de réception et 64 antennes d'émission à 40 m LOS (a) et 150 m NLOS (b)

## 2.5 Conclusion

Dans ce chapitre, l'architecture matérielle et logicielle du sondeur de canal, les réseaux d'antennes et les différentes configurations ont été décrites. Les campagnes de mesure en tunnel et en milieu suburbain ont été présentées. Les résultats et les analyses préliminaires en bande étroite en tunnel ont montré des exposants d'atténuation plus importants en présence de trafic. Ces exposants sont plus faibles à 5,89 GHz qu'à 1,35 GHz. L'analyse des mesures a révélé que ces exposants sont indépendants de la polarisation (verticale, horizontale, D/+45° et A/-45°).

En milieu suburbain, les mesures destinées à la détection des scénarios LOS/NLOS affichent des rapports signal/bruit (SNR) suffisants pour les analyses présentées au chapitre 4. Grâce à une qualité de signal satisfaisante, la caractérisation multidimensionnelle (temps, espace) du canal de propagation et l'identification du scénario de communication LOS/NLOS pourront être approfondies à l'aide de techniques d'apprentissage automatique, qui seront présentées dans les chapitres suivants.

Dans le cadre de cette thèse, ces campagnes de mesures en environnement suburbain ont permis d'apporter des contributions significatives, tant dans des revues spécialisées que lors de conférences (**Liste des publications**). Dans [121], une conférence a mis en avant la première campagne de sondage de canal V2I utilisant le sondeur de canal mMIMO en temps réel dans un environnement suburbain. En outre, un article [122] dans une revue scientifique a exploré les caractéristiques de non-stationnarité et les propriétés à grande échelle des canaux radio mMIMO V2I dans un contexte suburbain, soulignant l'importance de ces paramètres pour le développement de systèmes de communication fiables. De plus, dans [123], une étude comparative sur le temps de stationnarité et l'affaiblissement de parcours d'un canal de propagation SISO en environnements suburbains pour les communications véhicule-infrastructure a été réalisée. Cette étude s'appuie sur la méthode de traçage de rayons Veneris développée par le groupe GIRTEL de l'Université Polytechnique de Carthagène (UPCT), Espagne, ainsi que sur des données expérimentales issues de ces campagnes de mesures. .

N° Configuration	$(N_{R_x}, N_{T_x})$	$F_c$ (GHz)	$T_B$ (ms)	$T_{IB}$ (ms)	Nombre total de points	$T_F$ (ms)	Application	Scénario
13, 14, 15, 16	(8, 64)	5,89	0,975	15,6	10848	499,4	Classification	LOS
17, 18, 19, 20, 21, 22, 23	(8, 64)	5,89	0,975	15,6	10597	499,4	Classification	NLOS
25, 26, 27, 28, 29	(8, 64)	5,89	0,975	15,6	11200	499,4	Classification	Mixte
30	(1, 64)	5,89	0,975	1,975	13760	5504	Stationnarité	LOS

TABLEAU 2.4 – Tableau de synthèse des configurations utilisées en suburbain



## Chapitre 3

# Caractérisation multidimensionnelle du canal MIMO

### 3.1 Introduction

L'objectif de ce chapitre est d'analyser les résultats des mesures décrites dans le chapitre 2 en mettant l'accent sur la variabilité temporelle et spatiale d'un canal de propagation en milieu suburbain et en tunnel. Le temps de stationnarité  $T_s$  sera estimé en utilisant les méthodes décrites dans le chapitre 1. De plus, les caractéristiques spatiales et large bande seront analysées localement pour chaque région de stationnarité. Une étude sera d'abord entreprise sur un scénario d'environnement suburbain pour un canal MIMO massif de type  $1 \times 64$  à 5,89 GHz. Ensuite, des scénarios en tunnel aux deux fréquences (1,35 GHz et 5,89 GHz) pour différentes configurations de polarisation seront étudiés. Enfin, une synthèse de l'ensemble des résultats sera donnée en dernière partie.

### 3.2 Caractérisation de la variabilité temporelle en canal suburbain

Dans cette partie, une caractérisation de la variabilité temporelle en canal suburbain a été effectuée pour le scénario présenté au paragraphe 2.4.3 et présenté sur la Figure 2.14 (configuration 30).

#### 3.2.1 Description des expérimentations

Pour chaque liaison entre l'antenne d'émission  $k$  et l'antenne de réception  $l$ , la fonction de transfert du canal discrète peut être représentée comme suit :

$$H_{l,k}[m, q] = H_{l,k}(t_s m, f_s q), \quad (3.1)$$

où  $t_s = 1,9753$  ms est la périodicité des blocs et  $f_s = 97,79$  kHz est la résolution fréquentielle, avec  $m \in 0, \dots, M_t - 1$ ,  $q \in 0, \dots, M_f - 1$  et  $l \in 1, \dots, 64$ ,  $M_t$  étant le nombre de blocs avant resynchronisation. Le Tableau 3.1 résume la configuration des paramètres de MaMIMOSA utilisée pour ce scénario.

Paramètre	Configuration
Fréquence porteuse	5,89 GHz
Largeur de bande de mesure	80 MHz
Points de fréquence/Résolution en retard	818/12,5 ns
Puissance de transmission	0 dBm
Nombre d'antennes Tx/Rx	64/1
Durée de trame $T_F$	5,504 s (2752 blocs)
Durée de bloc (8 symboles OFDM)	975,36 $\mu$ s
Temps intertrame ( $T_{IF}$ )/interbloc ( $T_{IB}$ )	50 ms/1 ms
Nombre de blocs par trame $M_t$	2752

TABLEAU 3.1 – Configuration des paramètres de MaMIMOSA.

### 3.2.2 Évaluation du temps de stationnarité $T_s$

Comme mentionné dans la section 1.2.2, la fréquence de stationnarité minimale de 150 MHz publiée dans [40] pour des scénarios similaires est largement supérieure à la bande passante des mesures de 80 MHz. Par conséquent, le canal radio est considéré comme stationnaire en fréquence, les technologies ITS à 5,9 GHz occupant des bandes de moins de 20 MHz. Le temps de stationnarité  $T_s$  a été déterminé en utilisant les deux approches décrites dans le chapitre 2, le CCF (équation 1.11) et la colinéarité (équation 1.12). La version discrète de ces deux approches est décrite par la suite.

#### Calcul du GLSF discret

En se basant sur [37, 38, 40, 69], le processus d'évanouissement a été estimé comme étant localement stationnaire dans le temps sur une région de  $M_{ts} = 86$  blocs, soit une durée de 169,88 ms ( $M_{ts} \times t_s$ ). Pour chaque liaison  $(l, k)$ ,  $I$  et  $J$  fenêtres orthogonales multi-taper  $G_w$  respectivement dans le temps ( $I$ ) et la fréquence ( $J$ ) avec  $w \in \{0, \dots, IJ - 1\}$  sont choisies à partir de séquences sphéroïdales discrètes prolates (DPSS) pour obtenir plusieurs réalisations indépendantes du canal  $H_{l,k}$  [41, 69]. Le GLSF discret  $\hat{C}_{l,k}[k_{ts}, n, p]$  est calculé comme suit :

$$\hat{C}_{l,k}[k_{ts}, n, p] = \frac{1}{IJ} \sum_{w=0}^{IJ-1} |H_{l,k}^{(G_w)}[k_{ts}, n, p]|^2, \quad (3.2)$$

où  $n \in 0, \dots, M_f - 1$  représente l'indice de retard,  $p \in -M_{ts}/2, \dots, M_{ts}/2 - 1$  représente l'indice de la fréquence Doppler et  $k_{ts} \in 0, \dots, \frac{X - M_{ts}}{\Delta_t}$  est l'indice de temps pour chaque région, où  $X$  est le nombre total de blocs et  $\Delta_t$  désigne le décalage temporel entre les régions localement stationnaires consécutives. Ainsi, pour notre

application, la résolution Doppler locale est donnée par  $\nu_s = 1/(M_{ts}t_s) = 5,88$  Hz. La fonction multi-taper  $H^{(G_w)}$  est estimée comme suit :

$$H_{l,k}^{(G_w)}[k_{ts}, n, p] = \sum_{q'=-M_f/2}^{M_f/2-1} \sum_{m'=-M_{ts}/2}^{M_{ts}/2-1} G_w[m', q'] \times H_{l,k}[m' + \Delta_t k_{ts} + M_{ts}/2, q' + M_f/2] e^{-j2\pi(pm' - nq')},$$

où  $m'$  et  $q'$  représentent respectivement les indices de temps et de fréquence relatifs. La fenêtre  $G_w$  est localisée dans  $[-M_{ts}/2, M_{ts}/2 - 1] \times [-M_f/2, M_f/2 - 1]$ . Les valeurs  $I = J = 3$  utilisées par [38] pour un scénario similaire ont été choisies. Le CCF discret peut être obtenu à l'aide de transformée de Fourier discrète à trois dimensions (DTFT) du GLSF [37, 38] :

$$\hat{A}_{l,k}[\Delta m, \Delta q, r_{\Delta v}] = \mathcal{F}^3 \hat{C}_{l,k}[k_{ts}, n, p], \quad (3.3)$$

où  $\Delta m$ ,  $\Delta q$  et  $r_{\Delta v}$  représentent respectivement les indices de décalage temporel, de décalage fréquentiel et de décalage Doppler.

### Estimation de $T_s$ à partir du CCF

Comme il été mentionné dans le chapitre 1 (équation 1.10), l'estimation du temps de stationnarité  $T_s$  est déduite du moment Doppler d'ordre 1 du CCF comme suit [37, 38] :

$$\overline{(r_{\Delta v})_{l,k}} = \frac{1}{\|\hat{A}_{l,k}\|_1} \sum_{r_{\Delta v}} \sum_{\Delta q} \sum_{\Delta m} |r_{\Delta v} \Delta v| |\hat{A}_{l,k}[\Delta m, \Delta q, r_{\Delta v}]|, \quad (3.4)$$

où  $\Delta v$  est le décalage Doppler qui est égal à  $\frac{1}{N_r \times t_s \times M_{ts}/2}$ , avec  $N_r$  représentant le nombre de GLSF successives utilisées. Enfin, le temps de stationnarité  $T_s$  peut être calculé pour chaque canal d'indice antennaire  $(l, k)$  en utilisant la formule donnée dans [37] :

$$(T_s)_{l,k} = \frac{1}{\overline{(r_{\Delta v})_{l,k}}} \quad (3.5)$$

### Estimation de $T_s$ à partir de la colinéarité

Pour la deuxième approche, la colinéarité discrète  $d_{l,k}^{col}$  entre les GLSF échantillonnées à deux instants différents  $i_{ts}$  et  $j_{ts}$  est calculée pour chaque liaison  $l, k$  comme suit [40, 41] :

$$d_{l,k}^{col} \{ \hat{C}_{l,k} \} [i_{ts}, j_{ts}] = \frac{\text{tr} \{ \hat{C}_{l,k}[i_{ts}, n, p] \hat{C}_{l,k}[j_{ts}, n, p] \}}{\|\hat{C}_{l,k}[i_{ts}, n, p]\|_F \|\hat{C}_{l,k}[j_{ts}, n, p]\|_F} \quad (3.6)$$

$T_s$  est défini comme la région dans laquelle les valeurs de colinéarité dépassent un seuil prédéfini. Dans la littérature, une étude paramétrique a montré qu'un bon accord entre les deux méthodes d'estimation de  $T_s$  est obtenu pour un seuil qui varie entre 0,80 et 0,94 [38, 40, 41]. Pour notre cas particulier, un bon accord a été trouvé entre les 2 méthodes d'estimation de  $T_s$  pour un seuil de 0,85.

### Résultats et discussion

Les GLSF ont été calculées en se basant sur les fonctions de transfert mesurées. Pour le calcul de  $T_s$  à l'aide des deux méthodes décrites précédemment (1.2.2), 10 GLSF entrelacés correspondant à une durée de 934,34 ms sont considérés. Il convient de souligner que cette valeur a été délibérément choisie plus grande que le temps de stationnarité local de 567 ms rapporté dans [38] pour un scénario similaire.

La Figure 3.1 présente, en fonction de la distance Tx-Rx, les 25ème et 75ème percentiles ainsi que la médiane calculée à partir de la ECDF du réseau massif de 64 valeurs de  $T_s$ . Les résultats indiquent une valeur médiane moyenne de 568 ms pour l'approche basée sur le CCF et de 565 ms pour l'approche basée sur la colinéarité, suggérant que ces deux méthodes se comportent de manière similaire en moyenne.

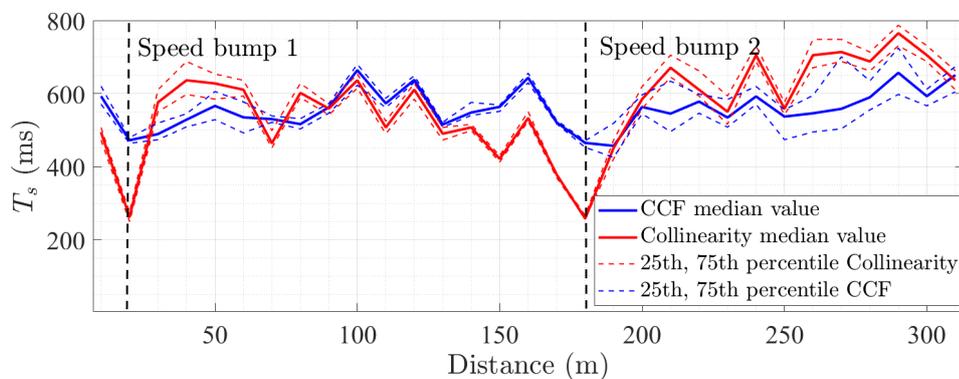


FIGURE 3.1 –  $T_s$  en fonction de la distance Tx-Rx obtenue par les méthodes CCF (bleu) et de colinéarité (rouge). La valeur médiane est représentée par une ligne continue et les 25ème/75ème percentiles par une ligne pointillée.

Toutefois, on observe deux instants spécifiques où les valeurs de  $T_s$  obtenues avec la méthode de la colinéarité sont nettement inférieures à celles calculées avec la méthode CCF. Cette différence s'explique par la présence de deux dos-d'âne (speed bumps) sur la route illustrée dans la Figure 2.14. En effet, la position de l'antenne Rx a subi un changement d'orientation brutal pendant une courte période d'environ 100 ms, valeur inférieure à la fenêtre de 934,34 ms de la méthode CCF. C'est la raison pour laquelle ce phénomène n'est pas observé avec la méthode CCF. Il est à noter que la méthode basée sur la colinéarité met en évidence ce changement rapide dans les statistiques du canal de propagation, contrairement à la méthode CCF. Cependant, il convient de souligner que la méthode de colinéarité est discrète et empirique, et que ses valeurs sont uniquement des multiples de  $\Delta_t$ , soit 84,92 ms. Cette restriction peut entraîner un biais par rapport à la valeur exacte de  $T_s$  décrite dans le premier chapitre. En revanche, le temps de stationnarité  $T_s$  basé sur le CCF est indépendant d'un paramètre arbitraire, offrant ainsi une caractérisation plus précise du canal.

Par ailleurs, une analyse approfondie de la distribution spatiale de  $T_s$  obtenue par la méthode CCF a été réalisée sur l'ensemble des canaux en évaluant ses propriétés

de corrélation. Pour chaque canal  $l, k$ , les valeurs de  $T_s$  sur l'ensemble du parcours ont été regroupées dans un vecteur. La corrélation entre un vecteur d'antenne de référence et tous les autres vecteurs d'antenne a été calculée. À titre d'exemple, la Figure 3.2.(a) présente la corrélation entre l'antenne n°46 (indiquée par la flèche noire) et les 63 autres antennes. Ces représentations en couleur montrent que  $T_s$  est fortement corrélé entre les antennes alignées verticalement par rapport à celles alignées horizontalement, avec des valeurs moyennes, respectivement de  $\rho_V = 0,95$  et  $\rho_H = 0,57$ . Les caractéristiques du canal radio sont donc similaires entre les antennes alignées verticalement (réseaux ULA verticaux ULA-V) pour lequel les diffuseurs dans le plan d'élévation sont rares contrairement au plan d'azimut. Ceci est également confirmé par la corrélation des fonctions de transfert entre les mêmes antennes, comme le montrent les résultats présentés dans la Figure 3.2.(b).

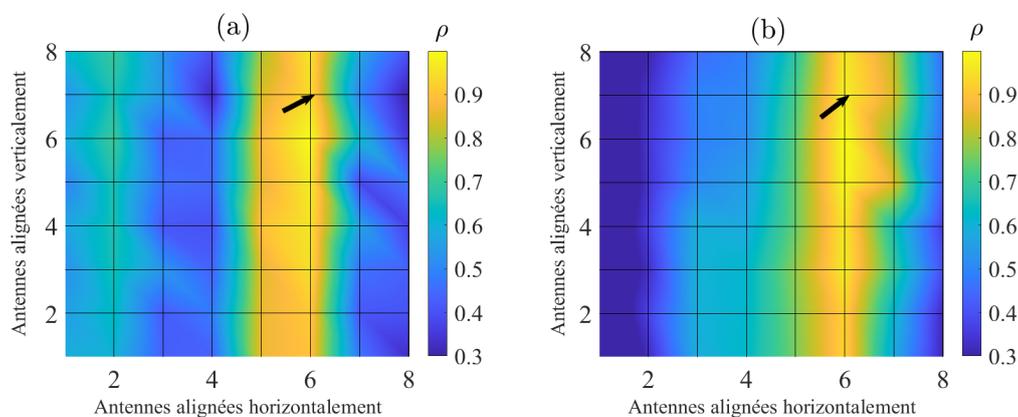


FIGURE 3.2 – Corrélation spatiale du réseau URA (8,8) calculée à partir (a) des valeurs  $T_s$  variant dans le temps et (b) des fonctions de transfert pour l'antenne #46 (flèche noire).

La Figure 3.3 présente la corrélation moyenne de  $T_s$  ainsi que celle des fonctions de transfert entre une antenne d'indice  $i$  et les autres antennes alignées horizontalement puis verticalement par rapport à l'antenne  $i$ .  $\rho_V$  est relativement constant pour toutes les antennes, tandis qu'une variation périodique de  $\rho_H$  est trouvée, car les antennes situées au centre du réseau ont des distances entre antennes plus faibles (corrélation élevée) que celles situées aux extrémités (corrélation faible). Les résultats obtenus sont cohérents avec ceux rapportés précédemment, comme cela a été démontré dans [124].

### 3.2.3 Évaluation des caractéristiques du canal radio large bande

Il est proposé ici d'étudier les caractéristiques du canal radio en termes de temps et de bande de cohérence, ainsi que d'étalement des retards et d'étalement angulaire. L'objectif de cette analyse est de corréler les propriétés de stationnarité du canal radio avec les propriétés spatiales des trajets multiples, qui varient dans le temps.

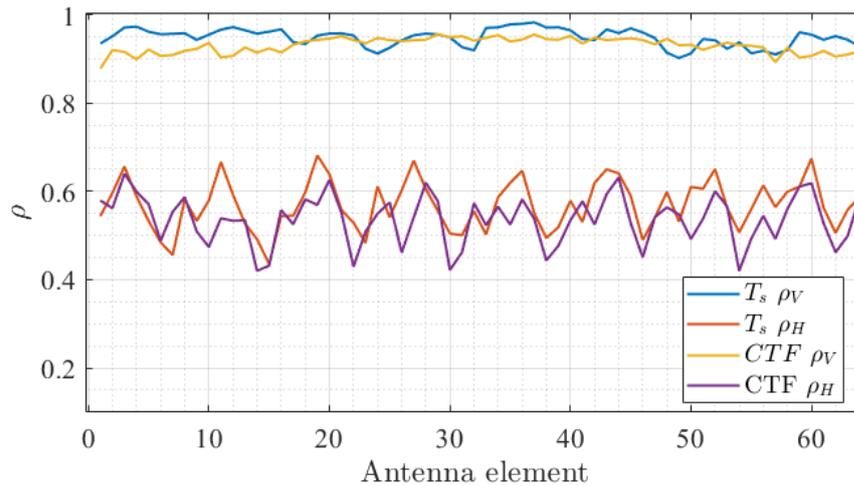


FIGURE 3.3 –  $\rho_V$  et de  $\rho_H$  moyens en fonction de l'indice d'antenne calculés à partir de  $T_s$  (bleu et jaune) et à partir des fonctions de transfert (rouge et violet).

### Temps de cohérence et bande de cohérence

La Figure 3.4 montre la distribution spatiale de  $T_c$  et  $B_c$  en fonction de la distance sous forme de boxplot; notez que chaque boxplot représente une moyenne de  $T_c$  et  $B_c$  pour chaque région de stationnarité. Ensuite les statistiques sont présentées en fonction des antennes. Dans un boxplot, la "boîte" représente l'intervalle interquartile, c'est-à-dire la plage entre le premier quartile 25% et le troisième quartile 75%, avec la médiane indiquée par une ligne à l'intérieur de la boîte. Les lignes pointillées s'étendent à partir de la boîte pour représenter la dispersion des données. Des points situés en dehors des moustaches peuvent indiquer des valeurs aberrantes.

À partir de la Figure 3.4, on note que les valeurs médianes de  $T_c$  et  $B_c$  sont comprises respectivement entre 2 ms et 100 ms et entre 10 MHz et 40 MHz, et sont significativement plus faibles que  $T_s$  et  $F_s$ . Cela confirme la propriété de "Doubly underspread" du scénario mesuré ( $B_c \ll F_s$  et  $T_c \ll T_s$ ), un concept introduit dans [69]. Le coefficient de corrélation de Pearson calculé entre  $T_c$  et  $B_c$  de 0,89 montre la forte corrélation entre ces deux paramètres. Une analyse fine des variations de  $B_c$  et  $T_c$  observées (Figure 3.4) est réalisée dans le paragraphe suivant avec les profils de puissance retard-Doppler.

### Les profils de puissance retard - Doppler

Les profils de retard de puissance (PDP) et les spectres de puissance Doppler (DPS) ont été calculés à partir du GLSF dans la région stationnaire locale et ont été moyennés sur l'ensemble des antennes émettrices. La forme discrète variant dans le temps de la

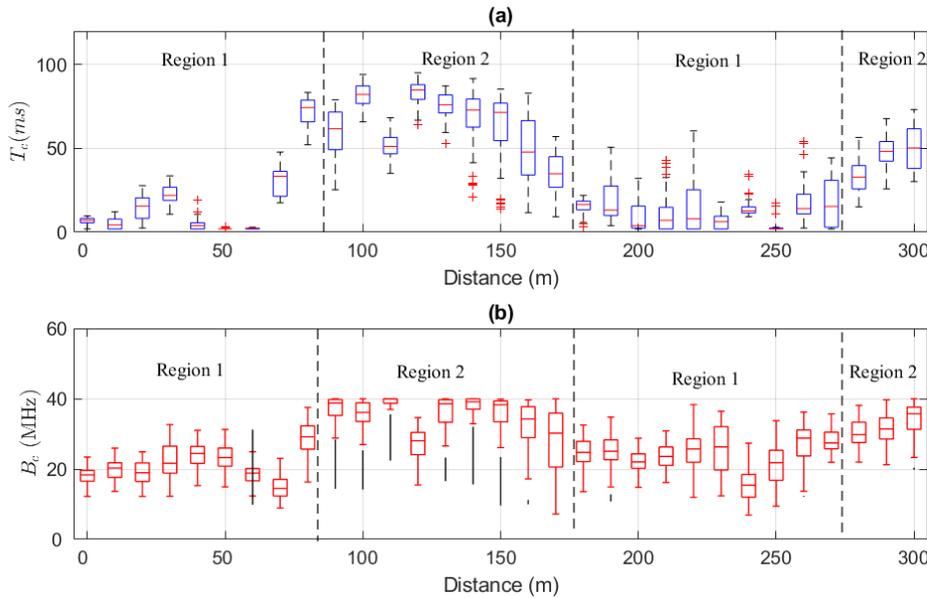


FIGURE 3.4 – Distribution spatiale en boxplot de  $T_c$  (a) en bleu et  $B_c$  (b) en rouge en fonction du temps.

PDP est calculée pour chaque liaison  $(l, k)$  comme suit [37, 38] :

$$PDP_{l,k}[k_{ts}, n] = \sum_p \hat{C}_{l,k}[k_{ts}, n, p], \quad (3.7)$$

alors que le DPS variant dans le temps est donné par [37, 38] :

$$DPS_{l,k}[k_{ts}, p] = \sum_n \hat{C}_{l,k}[k_{ts}, n, p] \quad (3.8)$$

Les PDP et les DPS sont représentés Figure 3.5 en fonction de la distance Tx-Rx. Sur le PDP, on observe que le retard du trajet direct augmente linéairement avec un retard maximal de  $0,98 \mu\text{s}$  lié à la longueur du chemin de propagation de 300 m, et un décalage Doppler négatif de  $-200$  Hz, le Rx s'éloignant du Tx à une vitesse constante de 40 km/h. Des trajets multiples peuvent être observés dans les 100 premiers mètres avec des fréquences Doppler négatives et positives. Ceci est dû à une forte diffusion des bâtiments pendant la première moitié du parcours. De plus, une variation de la fréquence Doppler du LOS est observée à l'emplacement des dos d'âne.

La richesse des trajets multiples est analysée à partir de l'étalement des retards  $\tau_{rms}$  et de l'étalement de Doppler  $\sigma_v$ , un seuil de -20 dB par rapport au maximum des fonctions est fixé (Figure 3.6). Deux régions présentant des caractéristiques différentes ont été identifiées et peuvent être physiquement liées à la structure géométrique du scénario étudié (voir la Figure 3.7). La première région (région 1) correspond aux premiers  $\sim 90$  mètres et entre  $\sim 170$  mètres et 270 mètres. La deuxième région (région 2) est située entre  $\sim 90$  mètres et 170 mètres et après 270 mètres. La région 1 dont les étalements des retards sont compris entre 10 ns (pure LOS) et 60 ns et les étalements Doppler entre 3 Hz à 120 Hz est riche en trajets

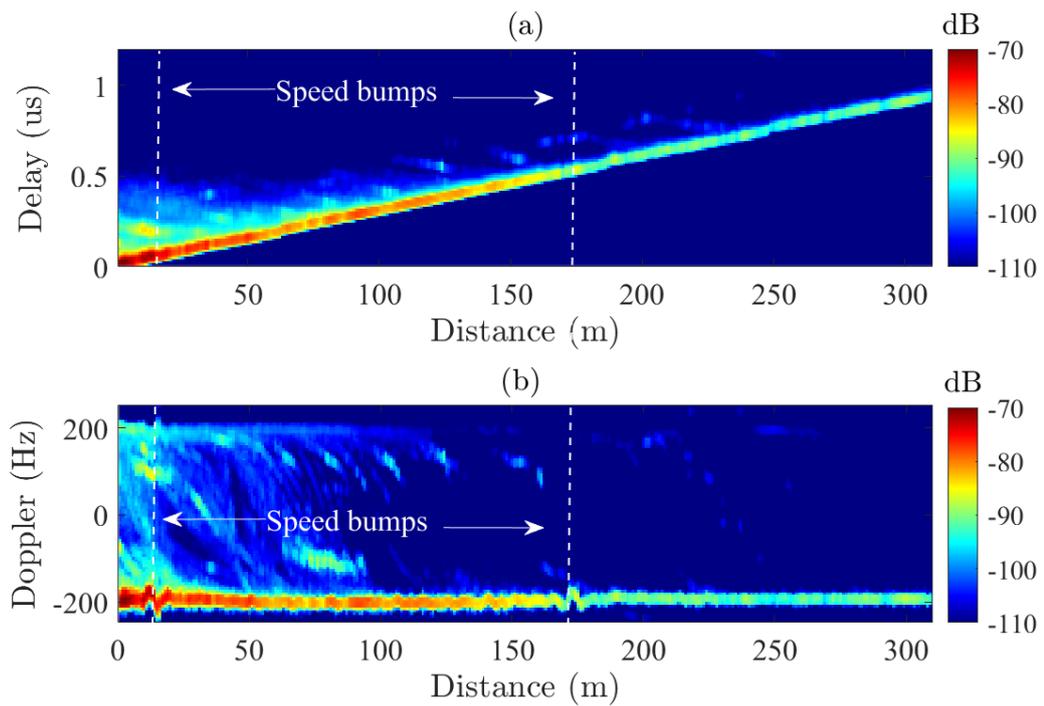


FIGURE 3.5 – PDP (a) et DPS (b) en fonction de la distance Tx-Rx.

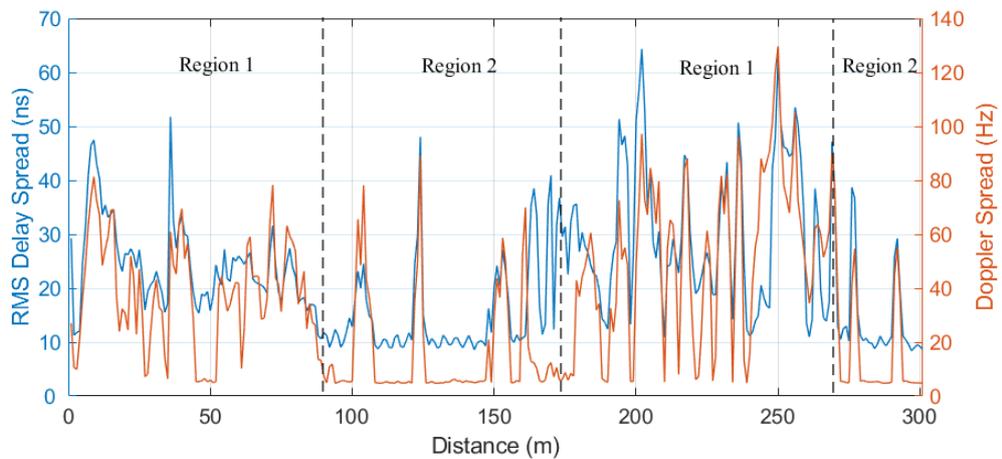


FIGURE 3.6 – Étalement des retards médian (bleu) et étalement de Doppler médian (rouge).

multiples. En revanche, la région 2 dépourvue de bâtiments à proximité du parcours présente des étalements de retards de l'ordre de 13 ns et des étalements de Doppler de 6 Hz. Comme pour  $T_c$  et  $B_c$ , les résultats suggèrent également un degré élevé de corrélation entre l'étalement des retards et l'étalement Doppler, avec un coefficient de corrélation de Pearson de 0,76.

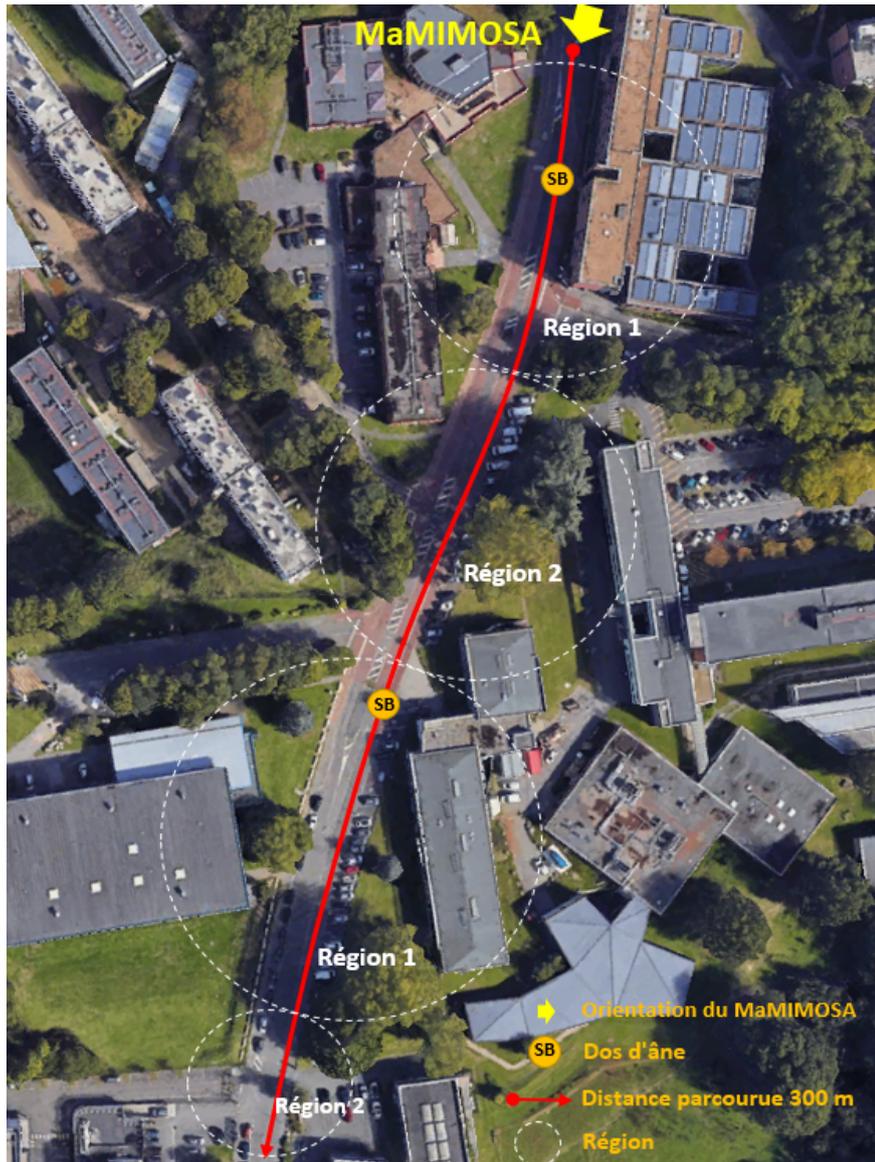


FIGURE 3.7 – Régions identifiées sur le parcours.

### Degrés de diversité dans le canal

En se basant sur les formules (1.43) et (1.44), le degré de diversité a été estimé pour un canal de propagation massive MIMO dans un environnement suburbain, prenant en compte les aspects temporels ( $d_t$ ), fréquentiels ( $d_f$ ), en retard ( $d_\tau$ ) et en Doppler ( $d_\nu$ ), sur les 64 antennes d'un réseau d'antennes URA  $8 \times 8$  comme le montre la Figure 3.8.

On observe que l'ordre de la diversité varie en fonction de la distance. On distingue deux régions l'une concentrant peu de trajets multiples, un étalement Doppler ainsi que des degrés de diversité faibles, contrairement aux régions riches en trajets multiples et à étalement Doppler élevés menant à des degrés de diversité élevé. Le Tableau 3.2 résume quelques valeurs statistiques moyennées sur les 64

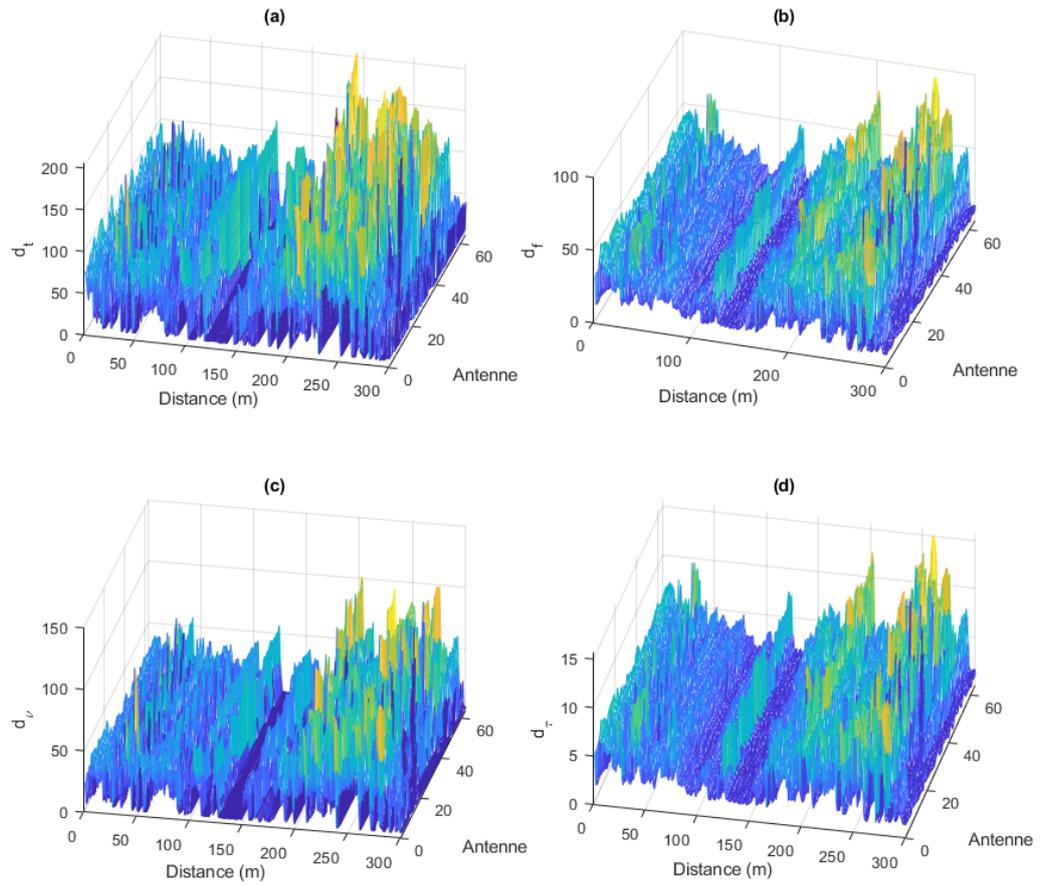


FIGURE 3.8 – Ordre de la diversité efficace en temporel ( $d_t$ ), fréquentiel ( $d_f$ ), en retard ( $d_\tau$ ) et en Doppler ( $d_v$ ) sur les 64 antennes en fonction de la distance

antennes. De plus, on remarque également que les degrés de diversité  $d_t$  et  $d_f$  sont respectivement proportionnels à  $d_v$  et  $d_\tau$ , avec seulement un facteur d'échelle.

	Max	Min	Médiane	Écart type
$d_t$	206	4	31	30
$d_v$	100	3	14	18
$d_\tau$	16	1	3	2
$d_f$	98	3	19	14

TABLEAU 3.2 – Quelques statistiques sur les différents ordres de diversité

### Coefficient $N_i$

Le nombre de réalisations indépendantes i.i.d.  $N_i$  offertes par le canal a été estimé en utilisant la formule (1.41) pour un canal massive MIMO dans un environnement

suburbain, pour chaque région de stationnarité et pour chaque antenne comme l'illustre la Figure 3.9.

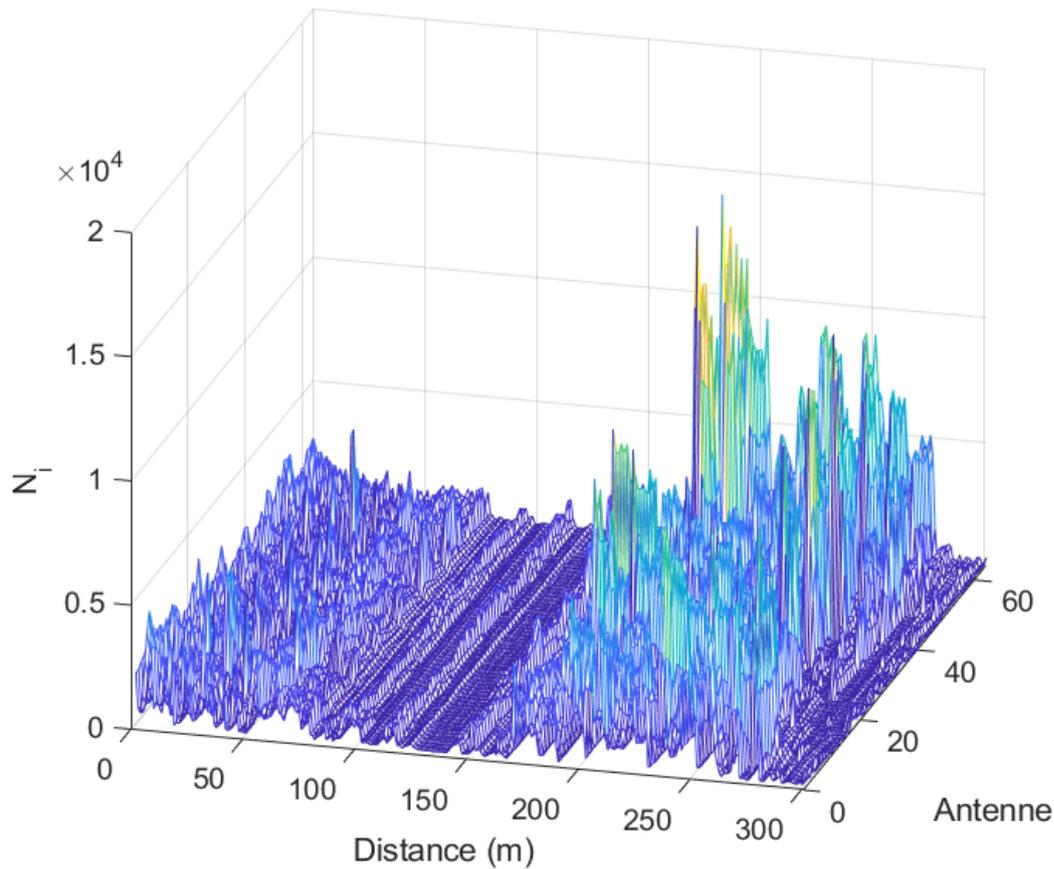


FIGURE 3.9 – Nombre de réalisations indépendantes i.i.d. offerts par le canal  $N_i$  en fonction d'antenne et de la distance

	WSSUS	Max.	Min.	Médiane	Écart type
$N_i$	$\infty$	16268	17	460	1429

TABLEAU 3.3 –  $N_i$ , nombre de réalisations indépendantes i.i.d. offerts par le canal

On observe que la valeur de  $N_i$  varie en fonction de la distance. Plus précisément, on remarque des valeurs faibles de  $N_i$  dans la région 2 et des valeurs importantes dans la région 1, surtout pour l'intervalle entre 200m et 270m, où les valeurs de  $N_i$  sont les plus élevées.

### Caractéristiques des directions de départ des rayons du réseau d'émission

En appliquant une méthode de formation de faisceau classique de type "Bartlett" [125], les angles de départ des rayons en azimut et en élévation sont extraits. Les

angles moyens et l'étalement angulaire sont tracés, respectivement sur la Figure 3.10 et la Figure 3.11. Un étalement angulaire moyen d'environ  $\sim 8^\circ$  et  $\sim 2^\circ$  respectivement, en azimut et en élévation, a été calculé.

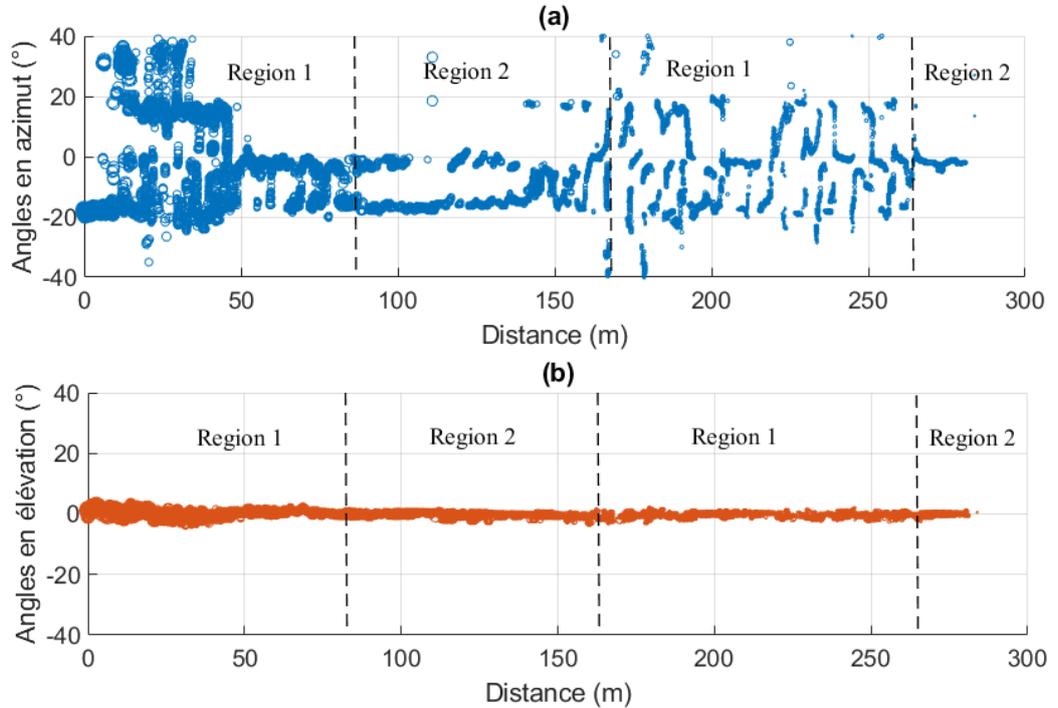


FIGURE 3.10 – Direction de départ des rayons : Valeur moyenne

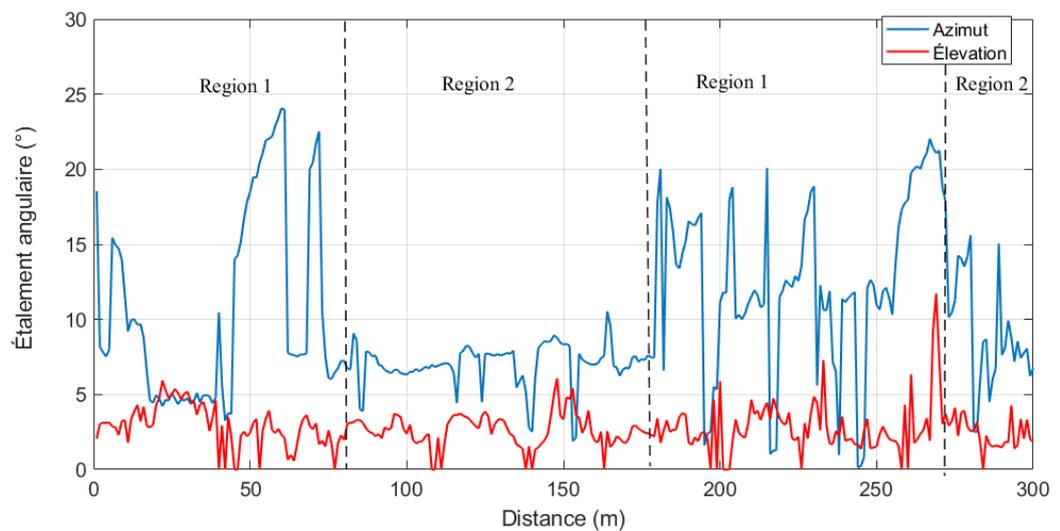


FIGURE 3.11 – Direction de départ : Étalement angulaire.

Une valeur moyenne d'environ  $\sim 0^\circ$  en l'élévation (Figure 3.10.(a)) correspondant à une visibilité directe LOS entre Tx et Rx. En azimut, une valeur moyenne entre  $-20^\circ$  et  $40^\circ$  (Figure 3.10.(b)) est obtenue. Néanmoins, les valeurs en azimut présentent une plus grande déviation dans la région 1 par rapport à la région 2, en raison d'un

environnement de diffusion plus riche. Enfin, la corrélation spatiale pour ULA-V, qui a été rapportée précédemment, peut être corroborée par le fait que les valeurs d'étalement sont plus faibles en élévation qu'en azimut. Dans ce type d'environnement, les sous-réseaux ULA-H permettent de mieux mettre en évidence les trajets multiples que les réseaux ULA-V.

Les Tableaux de synthèse 3.4 et 3.5 résument les paramètres du canal dans un environnement suburbain.

Paramètre	Médiane	Écart-type
$T_s$	565 ms	58 ms
$F_s$	150 MHz	–
$T_c$	20 ms	30 ms
$B_c$	34 MHz	13 MHz
$\tau_{RMS}$	20 ns	12 ns
$\sigma_v$	30 Hz	19 Hz
$\rho_H$	0,64	0,02
$\rho_V$	0,96	0,01

TABLEAU 3.4 – Tableau de synthèse des paramètres du canal dans un environnement suburbain

Paramètre	Azimut	Élévation
Angle de départ moyen	-20° à 40°	Environ 0°
Étalement angulaire médiane	8°	2°

TABLEAU 3.5 – Caractéristiques des directions de départ des rayons du réseau d'émission

### 3.3 Caractérisation de la variabilité temporelle du canal en tunnel

Dans cette partie, l'attention est portée sur l'étude de la variabilité temporelle du canal en tunnel en se basant sur les campagnes de mesure décrites dans le paragraphe 2.4. La caractérisation est réalisée en se basant sur les configurations (2) et (3) pour la fréquence de 1,35 GHz, ainsi que sur les configurations (7) et (8) pour la fréquence de 5,89 GHz, décrites dans le Tableau 2.2 avec une variation de la vitesse du véhicule entre 90 km/h et 100 km/h. Les détails concernant les structures des trames et les différentes configurations utilisées pour mesurer les fonctions de transfert  $H_{l,k}(t, f)$  sont exposés en détail dans le Tableau 2.2.

#### 3.3.1 Évaluation du temps de stationnarité $T_s$

L'estimation des fonctions GLSF est réalisée dans un premier temps. Ensuite, l'évaluation du temps de stationnarité  $T_s$  est effectuée en utilisant les méthodes du CCF et de la colinéarité, à une fréquence de 1,35 GHz, pour différentes configurations de polarisation : VV, HH, DD et AA. Dans le cadre de cette étude, les valeurs en VV à 1,35 GHz sont comparées à celles obtenues avec un réseau VV à 5,89 GHz.

Le Tableau 3.6 présente les différentes valeurs de  $I$  et  $J$  optimisées pour chaque configuration, ainsi que les valeurs de  $M_{ts}$  pour obtenir une durée de fenêtres de 128 ms environ.

Configuration	$M_{ts}$	$M_{ts} \times t_s$	$(I, J)$
(2)	66 blocs	128,7 ms	(2,2)
(3)	66 blocs	128,7 ms	(2,2)
(7)	266 blocs	127,7 ms	(3,3)
(8)	266 blocs	127,7 ms	(3,3)

TABLEAU 3.6 – Tableau des configurations et valeurs de  $(I, J)$  associées

#### Résultats et discussions

Le temps de stationnarité  $T_s$  a été évalué en utilisant 20 GLSF entrelacés d'une durée totale de 1351 ms. Cette durée a été choisie volontairement pour être plus grande que le temps de stationnarité moyen de 428 ms rapporté dans [39] pour un scénario similaire. Pour simplifier la présentation, l'étude suivante est basée sur la méthode CCF. Un tableau récapitulatif incluant les résultats obtenus par les deux méthodes est proposé en conclusion.

La Figure 3.12 présente l'évolution du temps de stationnarité  $T_s$  pour différentes polarisations (VV, HH, DD et AA) de la configuration (2) à trafic faible, en fonction de la distance. On observe une augmentation linéaire du temps de stationnarité  $T_s$  pour

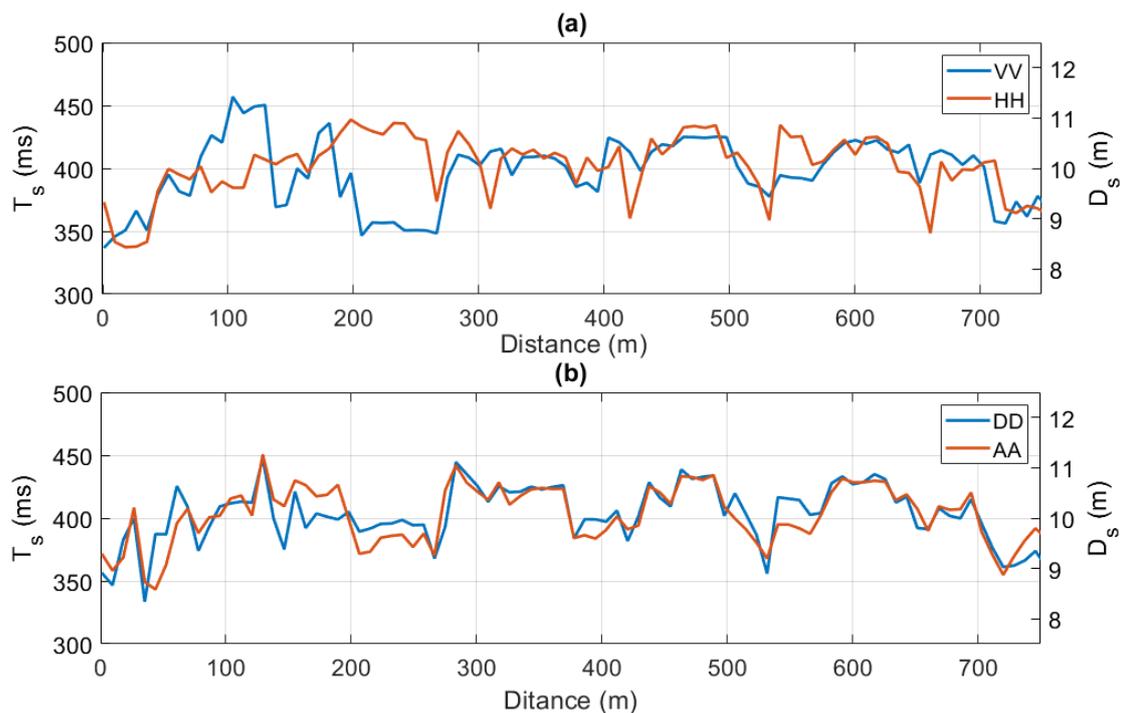


FIGURE 3.12 – Temps de stationnarité en fonction de la distance Tx-Rx pour les 4 configurations de polarisation.

les quatre polarisations, passant de 350 ms à 430 ms sur les premiers 100 mètres. Ensuite, pour le reste du trajet jusqu'à la fin du parcours à 740 m du Tx, le temps de stationnarité varie entre 300 ms et 500 ms.

Le modèle de propagation basé sur la théorie des rayons [126] en tunnel vide peut aider à l'interprétation des résultats. Ainsi, lorsque le récepteur est situé à très courte distance de l'émetteur, le modèle montre un étalement des directions d'arrivée des rayons, la majorité de l'énergie incidente se propageant sous un écart angulaire inférieur à 40 degrés par rapport à l'axe du tunnel. Ces rayons, ayant parfois subi de nombreuses réflexions, vont donc s'additionner de façon constructive ou destructive, engendrant des variations rapides du champ électrique reçu. Si la distance augmente, les rayons contribuant de façon significative au signal sont évidemment ceux qui présentent un angle d'incidence de plus en plus faible sur les parois, leur mise en opposition de phase n'intervenant que sur des périodes spatiales plus importantes [127]. Bien entendu, le trafic routier et les équipements de signalisation engendrent l'apparition de trajets multiples supplémentaires que le modèle ne fait pas apparaître.

Le Tableau 3.7 présente une vue d'ensemble des statistiques de  $T_s$  pour les différentes configurations et polarisations. On remarque que, dans la configuration (3) avec un trafic important, les valeurs médianes et minimales du  $T_s$  sont légèrement supérieures à celles de la configuration (2) avec un faible trafic. Les valeurs médianes à trafic faible et important sont en moyenne respectivement de 405 ms et 415 ms pour

Configuration	Statistique	VV	HH	DD	AA	VV <sub>theo</sub>
(2)	$T_s$ médiane (ms)	406	407	403	406	442
	$T_s$ écart type (ms)	27	24	23	22	33
	$T_s$ min (ms)	336	337	333	343	288
(3)	$T_s$ médiane (ms)	411	412	418	416	442
	$T_s$ écart type (ms)	23	24	23	24	33
	$T_s$ min (ms)	344	341	340	337	288

TABLEAU 3.7 – Statistiques du  $T_s$  pour différentes polarisations et configurations

les 4 polarisations, l'écart type de  $T_s$  étant de l'ordre de 6%.

$\rho$	VV	HH	DD	AA
VV	1	0,14	0,67	0,77
HH	0,43	1	0,73	0,50
DD	0,52	0,74	1	0,83
AA	0,63	0,72	0,91	1

TABLEAU 3.8 – Coefficients de corrélation  $\rho$ , configuration (2- trafic faible) en bleu, configuration (3- trafic important) en rouge

Le Tableau 3.8 présente les coefficients de corrélation ( $\rho$ ) entre les temps de stationnarité ( $T_s$ ) pour différentes polarisations : VV, HH, DD et AA, pour les configurations (2) à faible trafic en bleu et la configuration (3) à trafic important en rouge. Pour la configuration (2) à faible trafic, on observe une faible corrélation entre  $T_s$  en VV et HH avec une valeur de 0,14. Pour les autres polarisations DD et AA, on observe une corrélation importante avec des valeurs respectivement de 0,61 et 0,67. En revanche, on constate une corrélation supérieure à 0,5 entre les valeurs de  $T_s$  en HH, DD et AA, avec une forte corrélation DD/HH et DD/AA respectivement de 0,73 à 0,83, ce qui indique un comportement statistique de 2ème ordre similaire. Pour la configuration (3) à trafic important, on remarque des corrélations légèrement plus élevées que pour la configuration (2) à faible trafic, sauf pour VV/DD et VV/AA qui diminuent un peu. Le coefficient de corrélation le plus élevé  $\rho=0,91$  est observé entre les polarisations DD et AA. On souligne que la présence de trafic n'a pas un impact significatif sur les statistiques du temps de stationnarité en tunnel.

### Influence de la fréquence sur $T_s$

La Figure 3.13 illustre la variation du temps de stationnarité  $T_s$  en fonction de la distance Tx-Rx pour la configuration (2) en VV à 1,35 GHz et pour la configuration (7) en VV à 5,89 GHz. On observe des temps de stationnarité à 1,35 GHz supérieurs à ceux obtenus à 5,89 GHz, avec une médiane déduite de l'ECDF, de la Figure 3.14, respectivement de 406 ms et 332 ms. Notons que les valeurs de  $T_s$  pour les deux fréquences diffèrent d'environ 70 ms, avec des valeurs minimales de l'ordre de 270 ms et 336 respectivement à 5,89 GHz et 1,35 GHz, comme le montre la Figure 3.14. Il convient de noter que le type d'antennes utilisées et la distance entre les éléments du

réseau d'émission peuvent également influencer cette différence.

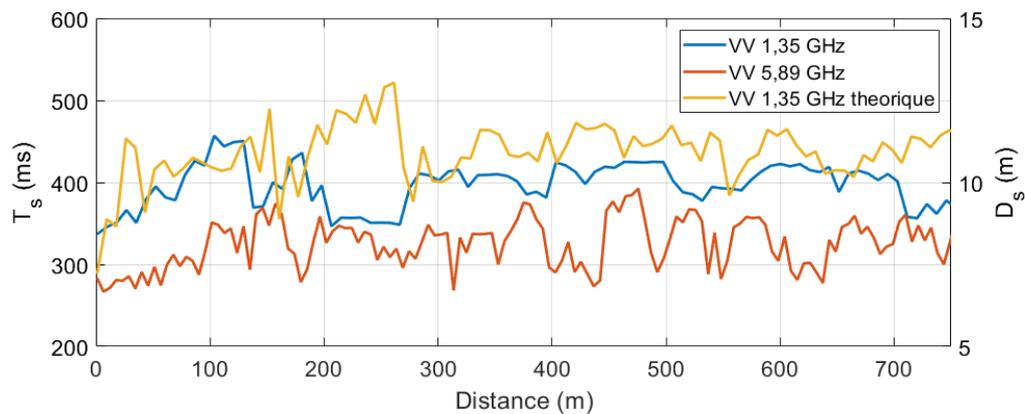


FIGURE 3.13 – Temps de stationnarité en VV à 1,35 GHz et à 5,89 GHz configurations (2 et 7 trafic faible).

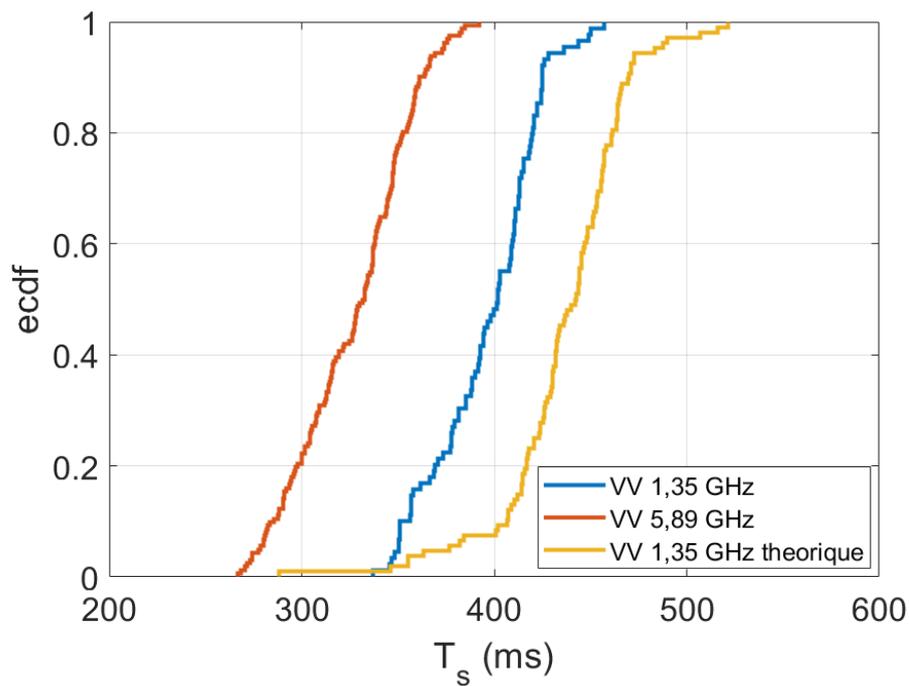


FIGURE 3.14 – ECDF du  $T_s$  en VV à 1,35 GHz et à 5,89 GHz.

### 3.3.2 Évaluation des caractéristiques du canal radio large bande

Dans cette partie, les caractéristiques du canal large bande sont estimées et présentées.

### Les profils de puissance retard-Doppler

Les Figures 3.15 et 3.16 représentent respectivement les PDPs et les DPS en fonction de la distance Tx-Rx, pour les polarisations : (a) VV et (b) HH. L'échelle des couleurs est associée à la puissance (dB). Les représentations en VV et HH (Figure

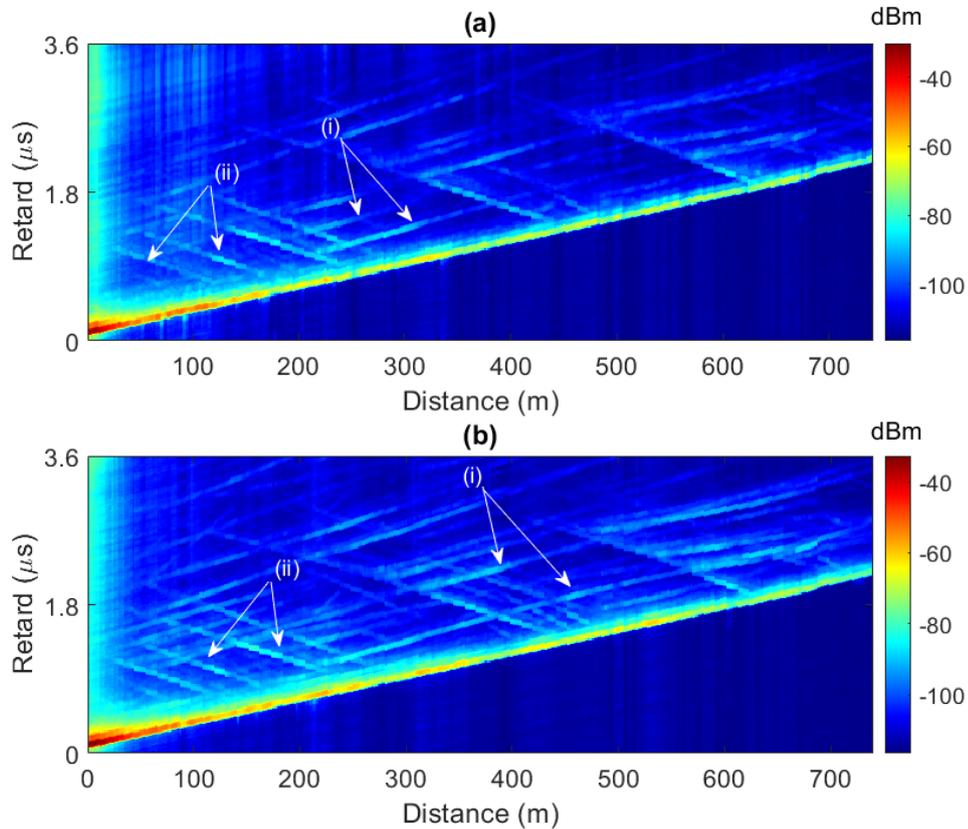


FIGURE 3.15 – Profil de puissance retards PDP en VV (a) et en HH (b) variant en fonction de la distance, moyennés sur toutes les antennes Tx, configuration (2) sans trafic.

3.15) mettent en évidence le trajet direct (LOS), dont le temps de propagation augmente proportionnellement avec la distance Tx-Rx. À proximité du LOS, on observe une grille de réflexions le long du parcours, avec une présence plus marquée en HH qu'en à VV. Comme le montre la figure, on distingue des réflexions parallèles à la composante principale du LOS (i), causées par des véhicules circulant devant le récepteur (Rx) à une vitesse similaire. Des réflexions opposées à la composante principale du LOS (ii) sont également présentes, dues aux objets métalliques fixes dans le tunnel, tels que les portes et les armoires. De plus dans la Figure 3.16, un décalage Doppler positif de 112 Hz indique que le récepteur (Rx) se rapproche de l'émetteur (Tx) à une vitesse constante de 90 km/h. On observe également des fréquences Doppler à 50 Hz et -50 Hz associées aux trajets multiples indiqués précédemment.

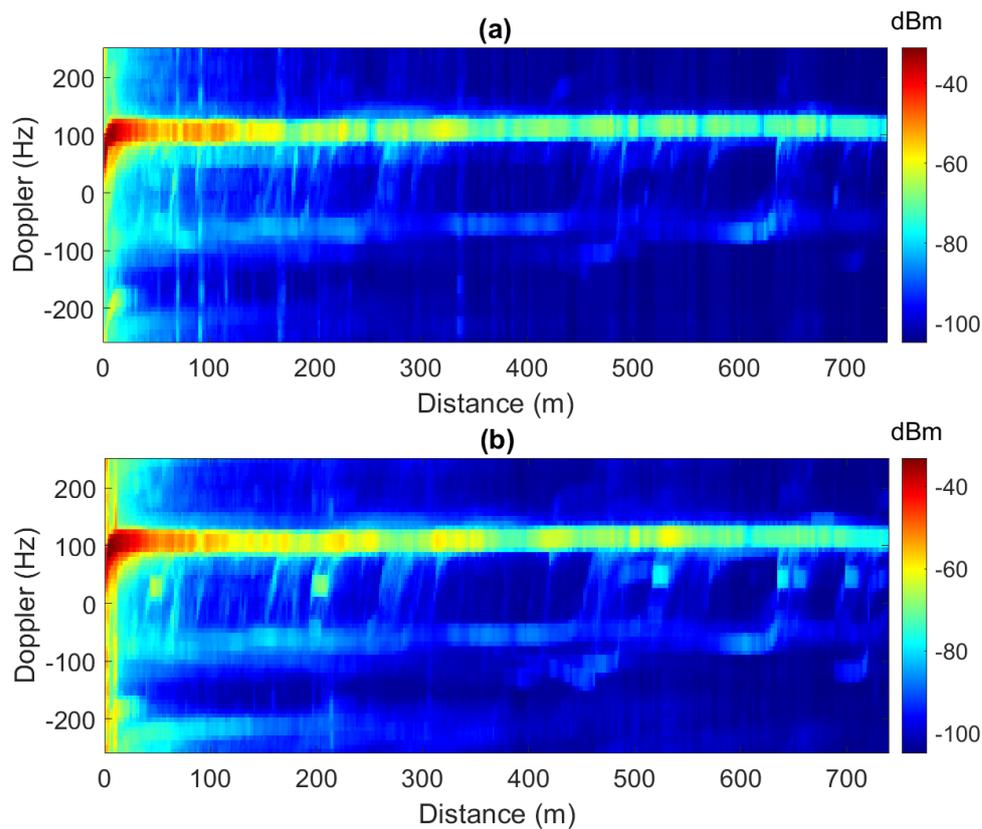


FIGURE 3.16 – Spectres de puissance Doppler (DSP) en VV (a) et en HH (b), configuration (2) en fonction de la distance, moyennés sur toutes les antennes Tx.

### L'étalement des retards et l'étalement Doppler

Pour un canal dont le temps de stationnarité varie, une évaluation précise des paramètres du canal à large bande peut être réalisée dans chaque région de stationnarité. La Figure 3.17 montre la variation de l'étalement des retards et de l'étalement Doppler en fonction de la distance Tx-Rx pour les deux polarisations (VV, HH) de la configuration (2). Pour simplifier la présentation, l'analyse porte sur les polarisations VV et HH. Un seuil de 20 dB par rapport à la puissance maximale a été choisi pour chaque point de distance. Au cours des premiers 150 mètres, on observe un étalement de retard de 10 ns associé à un étalement de Doppler de 15 Hz pour les deux polarisations.

Ces faibles valeurs sont dues à la composante LOS, dont l'écart entre la puissance du trajet direct est parfois 20 dB supérieur à celle des trajets multiples. Entre 150 et 300 mètres, une augmentation de l'étalement des retards pouvant atteindre 100 ns et un étalement Doppler maximum de 30 Hz sont observés en raison des réflexions provenant de la structure du tunnel et des objets métalliques présents sur les murs, comme indiqué respectivement dans les Figures 3.18 et 3.19. Ensuite, entre 350 et 550

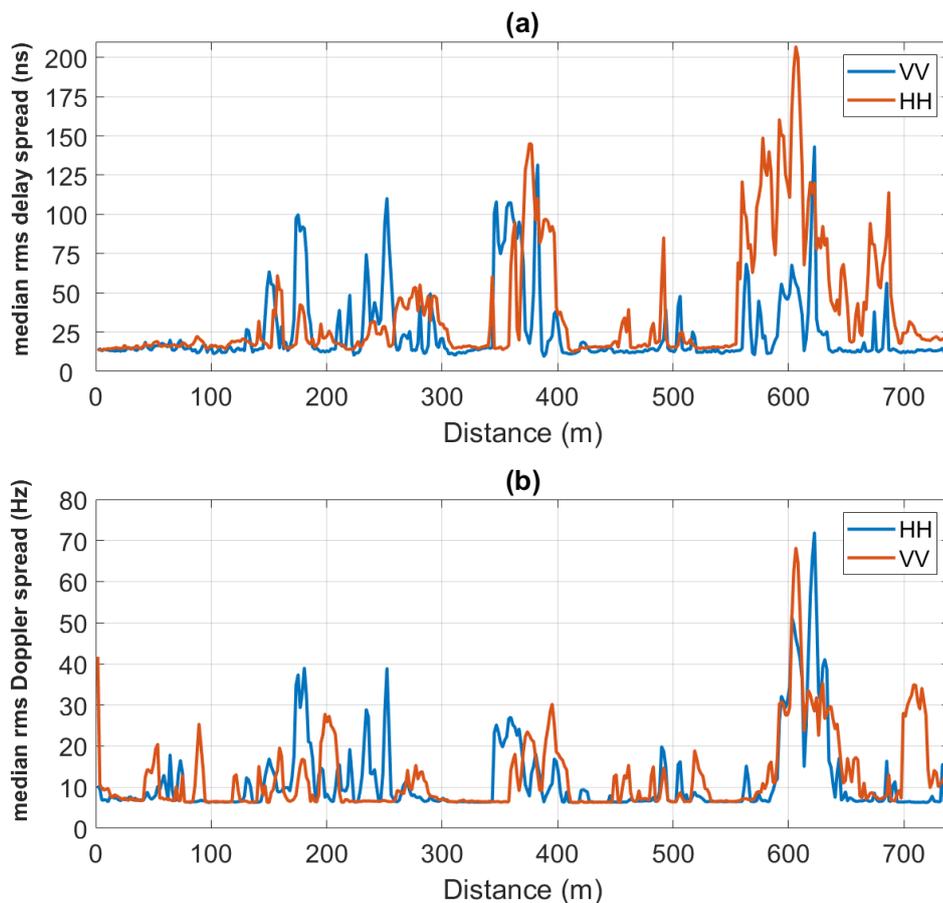


FIGURE 3.17 – Étalement des retards et étalement de Doppler, configuration (2) à faible trafic

mètres, on observe une forte augmentation de l'étalement de retard, atteignant environ 150 ns pour la polarisation VV et environ 170 ns pour la polarisation HH. L'étalement Doppler reste à environ 30 Hz. Entre 550 et 700 mètres, les valeurs d'étalement des retards et d'étalement Doppler augmentent encore, atteignant 200 ns et 70 Hz, respectivement. Ceci est dû à la faible amplitude du premier pic du PDP, résultant de la mise en opposition de phase du rayon direct et d'un rayon se propageant sous une incidence rasante. La différence entre les temps de propagation de ces deux rayons étant inférieure à 12,5 ns, la puissance des trajets multiples devient prépondérante.

### Corrélation entre les antennes

Pour la configuration (6), les coefficients de corrélation de Pearson ( $\rho$ ) entre les antennes d'émission (ULA-Vertical) ainsi qu'entre les antennes de réception (ULA-Horizontal) pour différents espacements spatiaux, ont été représentés pour les

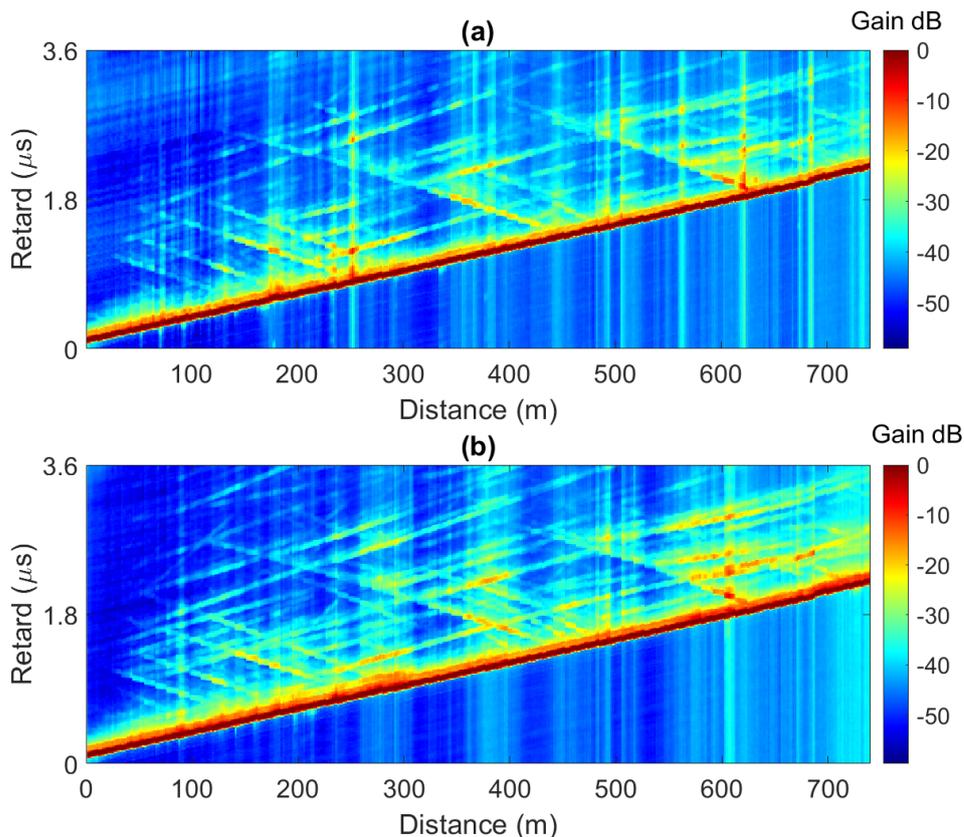


FIGURE 3.18 – Profil de puissance retard (PDP) normalisé en VV (a) et en HH (b) en fonction de la distance, moyenné sur toutes les antennes Tx, Configuration (2).

quatre polarisations à  $3\lambda/2$  et à  $3\lambda$ .

Les coefficients de corrélation  $\rho_{VV}$  pour la polarisation VV, tant expérimentale que théorique, à l'émission pour les espacements  $3\lambda/2$  et  $3\lambda$ , sont présentés dans la Figure 3.20. On remarque que les valeurs issues du modèle théorique présentent les coefficients de corrélation les plus élevés en fonction de la distance, étant donné que le modèle ne prend pas en considération tous les objets présents dans un vrai tunnel. Une médiane environ 0,97 pour  $3\lambda/2$  et  $3\lambda$  (Figure 3.21). En revanche, les coefficients de corrélation expérimentaux sont inférieurs à ceux du modèle théorique, avec des médianes de 0,89 et 0,90 respectivement pour  $3\lambda/2$  et  $3\lambda$  (Figure 3.21). La corrélation à l'émission pour les quatre polarisations, avec un espacement de  $3\lambda/2$ , est présentée dans la Figure 3.22. On observe que la corrélation  $\rho_{VV}$  est la plus élevée, avec une médiane de 0,89. Les polarisations HH, DD et AA affichent un comportement similaire, avec des médianes respectives de 0,83, 0,84 et 0,85.

À la réception, les coefficients de corrélation  $\rho_{VV}$ , tant expérimentaux que théoriques, pour les espacements  $3\lambda/2$  et  $3\lambda$ , sont présentés dans la Figure 3.23. On

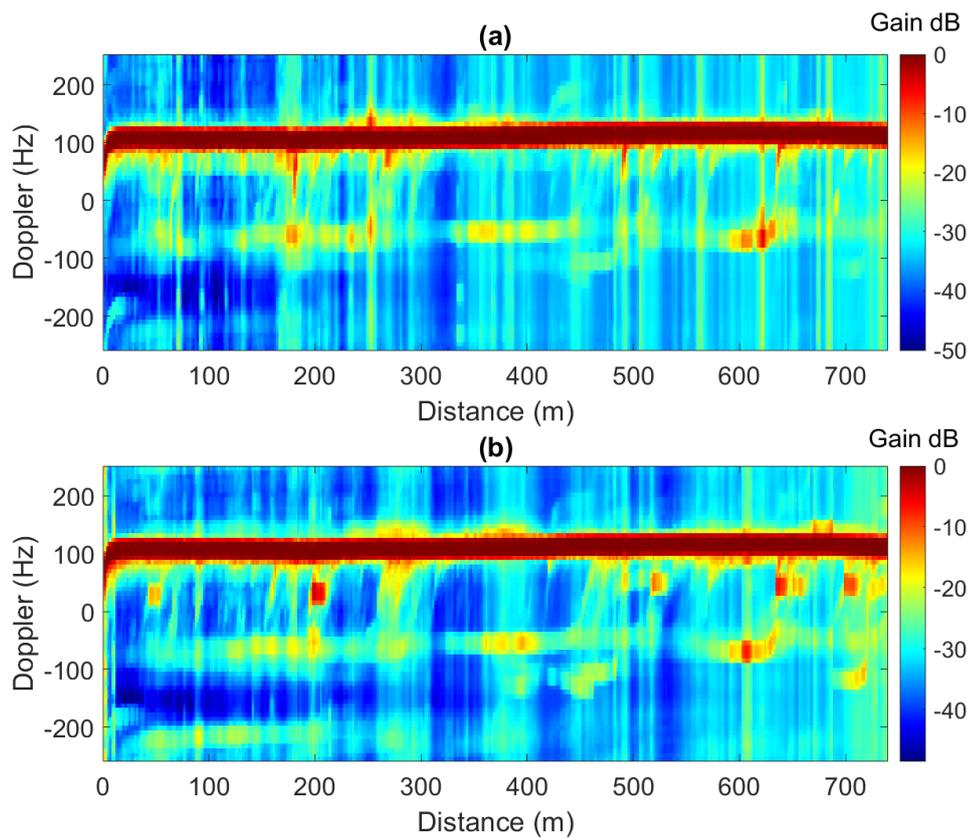


FIGURE 3.19 – Profil de puissance Doppler normalisé en VV (a) et en HH (b) en fonction de la distance, moyenné sur toutes les antennes Tx, Configuration (2).

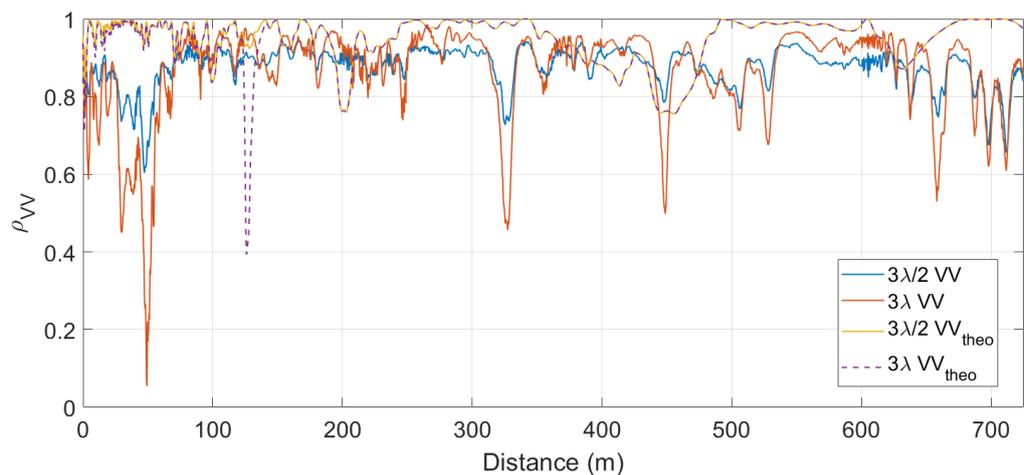


FIGURE 3.20 – Coefficient de corrélation entre les antennes d'émission en VV expérimentale (configuration 6) et théorique avec les écarts  $3\lambda/2$  et  $3\lambda$ .

observe que les valeurs sont faibles par rapport à la corrélation à l'émission pour les deux espacements  $3\lambda/2$  et  $3\lambda$ , avec des valeurs médianes respectivement de 0,78 et

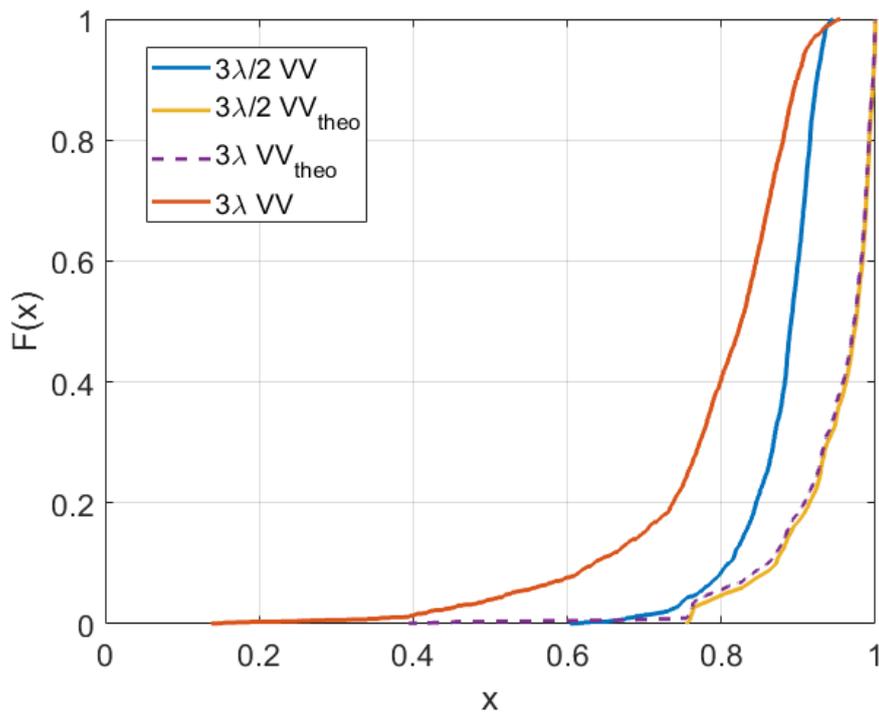


FIGURE 3.21 – ECDF des coefficients de corrélation entre les antennes d'émission en VV expérimentale (configuration 6) et théorique avec les écarts  $3\lambda/2$  et  $3\lambda$ .

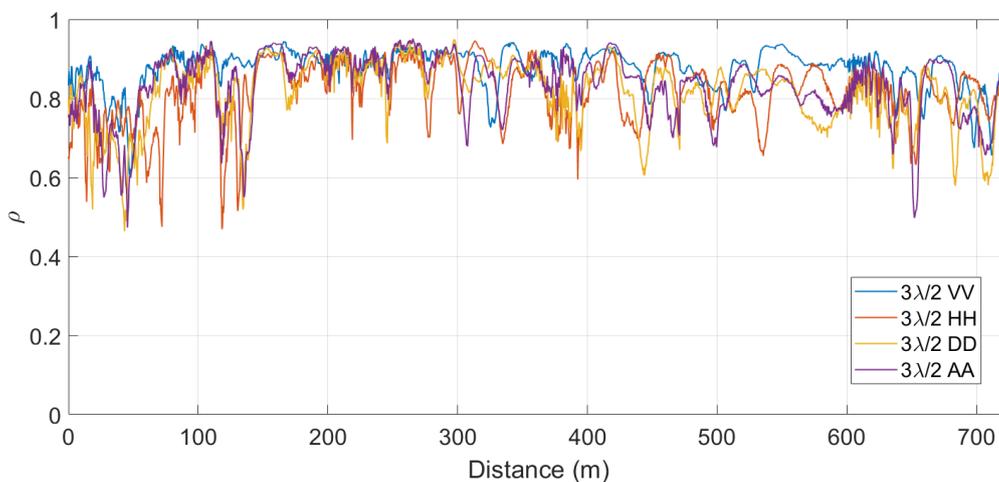


FIGURE 3.22 – Coefficient de corrélation entre les antennes d'émission pour les 4 polarisations expérimentales (configuration 6) avec un écart de  $3\lambda/2$ .

0,68. Pour le modèle théorique, les valeurs médianes sont de 0,96 et 0,60 (Figure 3.24).

Pour toutes les polarisations expérimentales à la réception, on observe sur la Figure 3.25 une augmentation moyenne du coefficient de corrélation d'environ 0,5 à 0,7 sur les 100 premiers mètres, suivie d'un rapide affaiblissement entre 0,2 et 0,9 pour le modèle théorique sur les premiers 100 mètres. Contrairement à l'émission, le

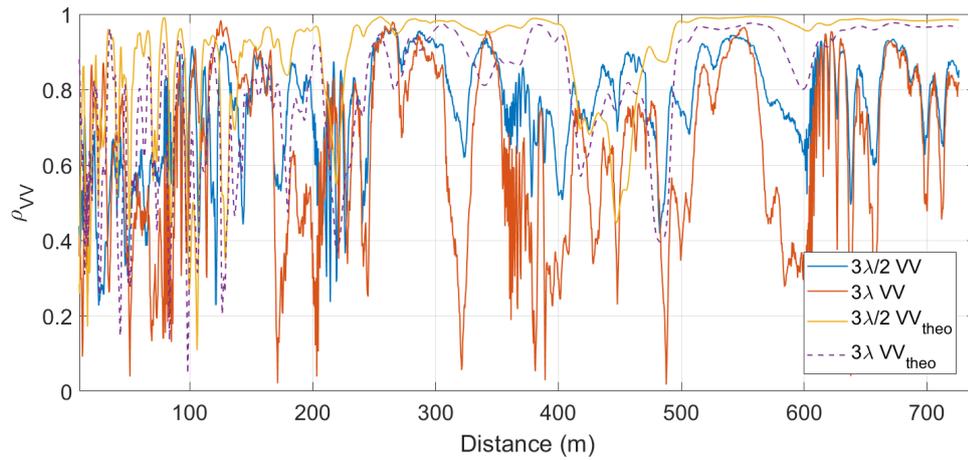


FIGURE 3.23 – Coefficient de corrélation entre les antennes de la réception en VV expérimentale (Configuration 6) et théorique avec les écart  $3\lambda/2$  et  $3\lambda$ .

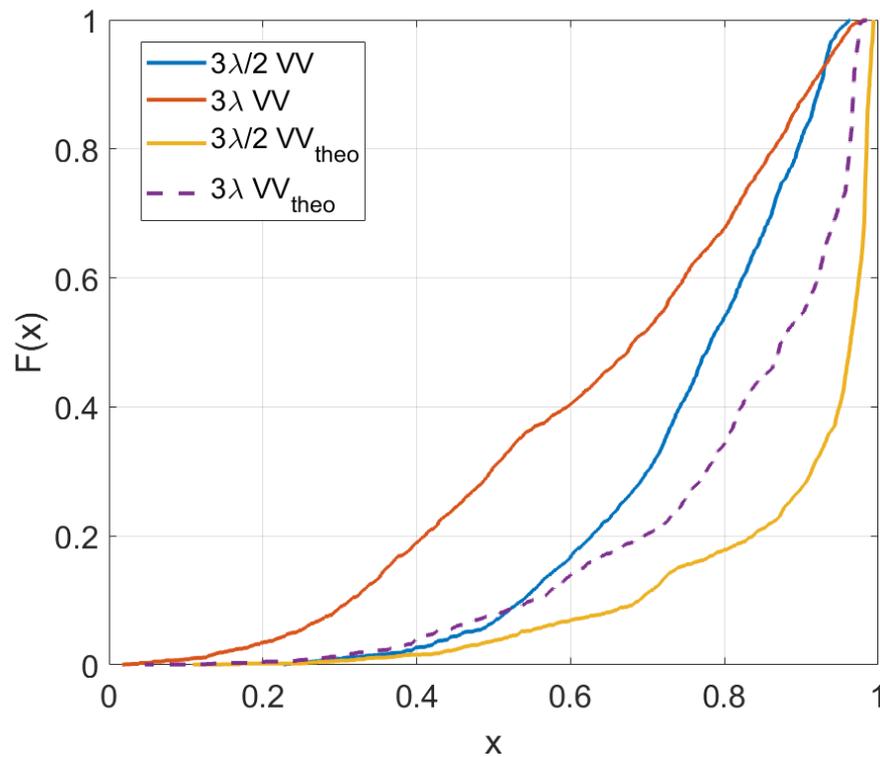


FIGURE 3.24 – ECDF des coefficients de corrélation entre les antennes de la réception en VV expérimentale (Configuration 6) et théorique avec les écarts  $3\lambda/2$  et  $3\lambda$ .

passage de l'espacement  $3\lambda/2$  à  $3\lambda$  n'a pas d'effet significatif sur les coefficients de corrélation, en particulier pour la polarisation VV. Les valeurs théoriques de corrélation sont sensiblement identiques à l'émission et à la réception. En résumé, le Tableau 3.9 présente les valeurs médianes des coefficients de corrélation pour les quatre polarisations à l'émission et à la réception.

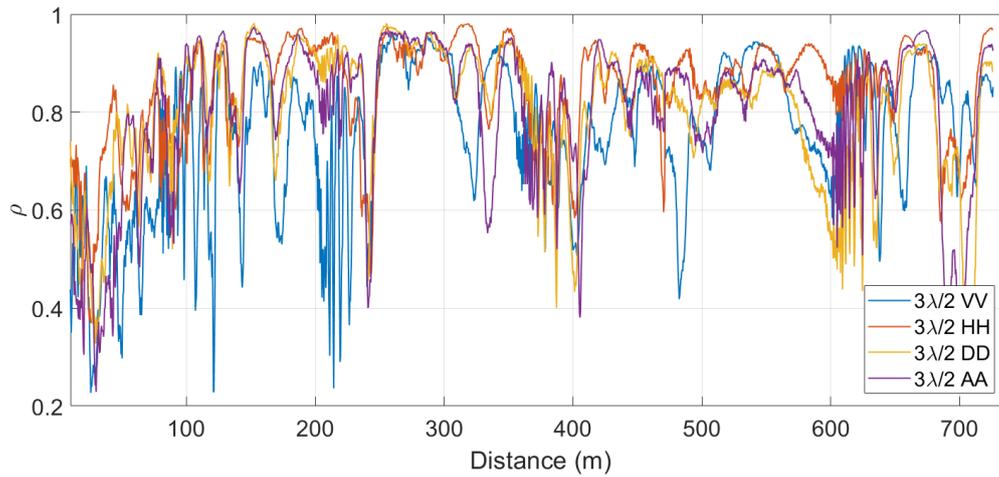


FIGURE 3.25 – Coefficient de corrélation entre les antennes de la réception pour les 4 polarisations expérimentales (configuration 6) avec un écart de  $3\lambda/2$ .

	Espacement	$\rho_{VV}$	$\rho_{HH}$	$\rho_{DD}$	$\rho_{AA}$	$\rho_{VV_{theo}}$
Tx(ULA-V)	$3\lambda/2$	0,89	0,83	0,84	0,85	0,97
	$3\lambda$	0,82	0,73	0,78	0,76	0,97
Rx(ULA-H)	$3\lambda/2$	0,78	0,88	0,84	0,85	0,96
	$3\lambda$	0,68	0,72	0,70	0,70	0,60

TABLEAU 3.9 – Médiane de corrélation entre Tx et Rx avec différents espacements

### Relation entre le temps de cohérence $T_c$ et l'étalement Doppler $\sigma_v$

$T_c$  et  $\sigma_v$  ont été estimés localement en utilisant des fenêtres de taille 450 ms autour de  $T_s$ . Cette estimation a été réalisée en balayant l'ensemble du parcours de la configuration (2) et en faisant glisser ces fenêtres avec un décalage de  $T_s/2$ . Pour chaque fenêtre, le temps de cohérence  $T_c$  a été estimé pour les 4 polarisations en utilisant les deux méthodes. La relation entre  $T_c$  et  $\sigma_v$  est :

$$T_c \triangleq \frac{1}{k_{dopp} \times \sigma_v} \quad , \quad (3.9)$$

où  $k_{dopp}$  dans notre étude est un facteur empirique déterminé statistiquement comme l'inverse de la médiane des produits entre  $T_c$  basé sur l'autocorrélation à un seuil de 0,7 et l'étalement Doppler  $\sigma_v$ , le long du tunnel. Le temps de cohérence  $T_c$  déduit de la première méthode pour un seuil de 0,7 et le temps de cohérence Doppler avec un coefficient  $k_{dopp}$  de 4,03 sont représentés respectivement en bleu et en rouge sur l'axe de gauche de la Figure 3.26, associé à la distance de cohérence  $D_c$  sur l'axe de droite.

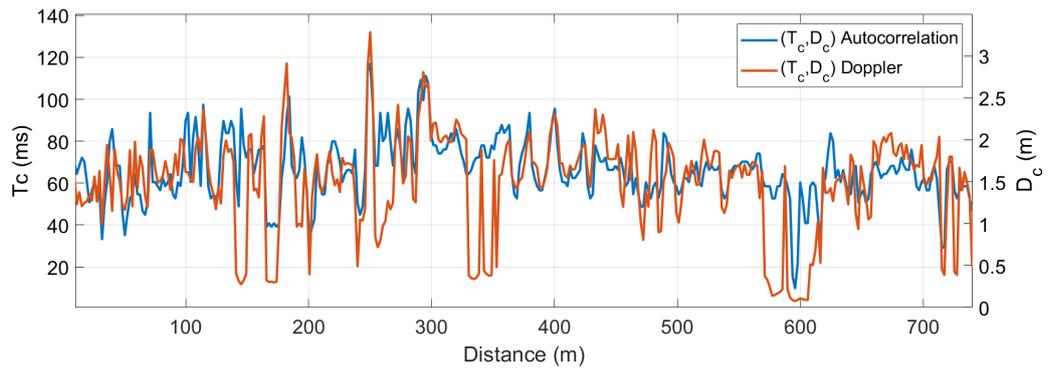


FIGURE 3.26 – Temps de cohérence ( $T_c$ ) avec autocorrélation et étalement Doppler en VV (à droite) associé à la distance de cohérence  $D_c$  avec l'autocorrélation et l'étalement Doppler (à gauche) pour la configuration (2).

On observe que la valeur estimée du facteur  $k_{dopp}$  permet d'obtenir un temps de cohérence Doppler proche de celui de l'autocorrélation, avec un coefficient de corrélation de Pearson d'environ 0,65. Les valeurs de  $T_c$  varient entre 10 ms et 110 ms, avec une médiane de 66 ms. De plus, les valeurs de la distance de cohérence  $D_c$  varient entre 0,3 m et 3 m, avec une médiane de 1,65 m.

Finalement, la Figure 3.27 présente l'ECDF des valeurs du facteur  $k_{dopp}$  pour un environnement en tunnel à 1,35 GHz et 5,89 GHz, pour un environnement suburbain à 5,89 GHz. Dans [128], la valeur estimée de  $k_{dopp}$  est 2,36 en milieu urbain. Cette valeur de 2,36 est tracée dans la Figure 3.27 (en se basant sur la fréquence Doppler max  $f_m$ ).

On observe que pour les fréquences de 1,35 GHz et 5,89 GHz, en tunnel et en suburbain, les valeurs de  $k_{dopp}$  varient de 0,2 à 6. De plus, les médianes de  $k_{dopp}$  sont de 3,86 à 4,03 respectivement pour 5,89 GHz et 1,35 GHz en tunnel et 1,80 en suburbain à 5,89 GHz. Les valeurs médianes de  $k_{dopp}$  à 5,89 GHz et à 1,35 GHz sont entre la valeur mentionnée dans [128]. Cela suggère que le facteur  $k_{dopp}$  peut varier en fonction de l'environnement et de la fréquence, ce qui souligne l'importance de l'estimation locale du temps de cohérence en utilisant des paramètres spécifiques à chaque scénario. Le

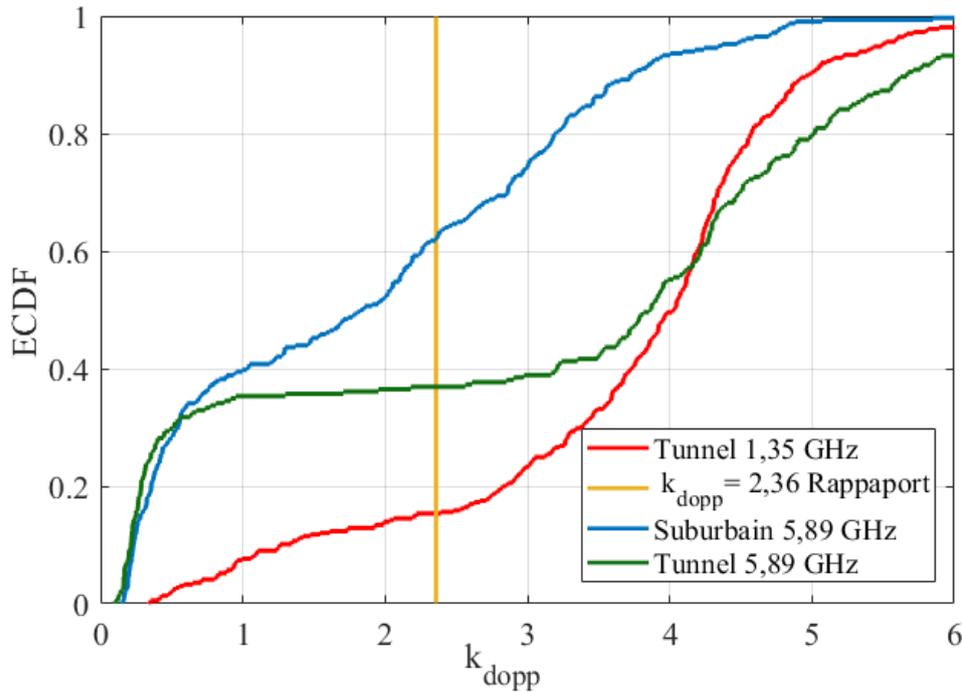


FIGURE 3.27 – ECDF du facteur  $k_{dopp}$  5,89 GHz suburbain et 1,35 GHz.

Tableau 3.10 rassemble les paramètres statistiques déduits de l'étude en tunnel et en milieu suburbain.

### 3.4 Conclusion

Dans ce chapitre, une analyse détaillée de la variabilité temporelle et spatiale d'un canal de propagation MIMO a été menée, en se basant sur les résultats de mesure décrits au chapitre 2. En milieu suburbain, le temps de stationnarité médian  $T_s$  localement estimé est de 565 ms, avec une corrélation spatiale élevée à 0,9 pour les réseaux ULA verticaux et 0,6 pour les réseaux ULA horizontaux. En tunnel, une médiane du temps de stationnarité a été estimée à 406 ms (polarisation VV) et 332 ms (polarisation HH) respectivement à 1,35 GHz et 5,89 GHz. Une analyse des caractéristiques large bande a également été présentée pour les deux environnements. La base de données en environnement suburbain, comportant  $\sim 33000$  matrices étiquetées LOS ou NLOS servira au développement du modèle d'apprentissage décrit dans le chapitre suivant.

Scénario	Fréquence	Vitesse	Médiane $\tau_{rms}$	Médiane $\sigma_{rms}$	Médiane $B_c$	Médiane $D_c$ (m)	Médiane $D_s$ (m)	Médiane $k_{dopp}$	Méthode
Tunnel (2)	1,35	90	14 ns	6,66 Hz	39 MHz	1,65	10,15	4,03	CCF
							11,1		Colinéarité
Tunnel (3)	1,35	90	15 ns	9 Hz	39 MHz	0,7	10,3	4,03	CCF
							11,05		Colinéarité
Tunnel (7)	5,89	90	14 ns	6,6 Hz	23 MHz	1,04	8,3	3,86	CCF
							9,4		Colinéarité
Tunnel (8)	5,89	90	120 ns	11 Hz	20 MHz	0,7	7,4	3,86	CCF
							7,9		Colinéarité
Suburbain (30)	5,89	40	20 ns	30 Hz	34 MHz	0,22	6,31	1,8	CCF
							6,27		Colinéarité

TABLEAU 3.10 – Tableau des valeurs pour différents scénarios

## Chapitre 4

# Apport du ML sur la classification LOS NLOS

### 4.1 Introduction

L'objectif de ce chapitre est d'étudier l'apport du ML à la classification des scénarios de propagation LOS et NLOS dans un environnement suburbain en développant un modèle ML de classification robuste, entraîné avec des données expérimentales étiquetées. L'utilisation du ML pour classifier le LOS/NLOS permettra d'exploiter les caractéristiques disponibles du canal de propagation qui reflètent les comportements de propagation physique réelle, créant ainsi des modèles précis et adaptés aux conditions spécifiques du canal de communication. Ces modèles seront ensuite utilisés par les stations de base pour prendre des décisions éclairées concernant le choix des techniques spatiales appropriées à appliquer aux signaux émis, contribuant ainsi à améliorer la qualité et la fiabilité des systèmes de communication mobile.

La littérature examinée dans le chapitre précédent a abordé la détection et la classification des scénarios de propagation LOS/NLOS, en s'appuyant sur des données extraites d'expérimentations ou de simulations. Ces études ont intégré l'utilisation de techniques d'intelligence artificielle ML et DL. Toutefois, la majorité de ces recherches se sont concentrées sur des scénarios en environnement indoor ou pour des communications V2V à partir de signaux satellites (GNSS).

Dans ce chapitre, la classification LOS/NLOS pour un canal de propagation Massive MIMO en scénario V2I et en environnement suburbain à 5,9 GHz est abordée. En adaptant de nouvelles caractéristiques et une méthode hybride de sélection des caractéristiques issues de la base de données expérimentales, un modèle robuste et générique a été construit.

Ce chapitre est organisé comme suit : une première partie aborde le processus de classification LOS/NLOS dans lequel la collecte et le traitement des données sont détaillés, dans une deuxième partie, les différentes méthodes de sélection des

caractéristiques dont une approche originale dite hybride, sont décrites, la troisième partie présente les algorithmes utilisés dont les performances en terme de pourcentage de réussite de la classification LOS/NLOS sont comparées. Ce chapitre se termine sur une comparaison entre les performances de ce modèle et celles de l'état de l'art.

## 4.2 Processus de Classification LOS/NLOS

Pour aborder un problème de classification en ML, il est impératif de suivre un processus de traitement comprenant plusieurs étapes appelé "*Workflow*". Cela commence par la collecte des données suivi d'un pré-traitement des données impliquant le nettoyage, l'élimination des valeurs aberrantes et la normalisation pour faciliter la convergence des algorithmes d'apprentissage. Vient ensuite une étape cruciale qui consiste à dériver et sélectionner des caractéristiques significatives à partir des données prétraitées. Ces éléments exploitables sont par la suite exploités par le modèle d'apprentissage automatique [129, 130]. Les détails spécifiques de ces étapes sont exposés plus en détail dans la suite de ce chapitre.

### 4.2.1 Collecte des données

Plusieurs campagnes de mesures ont été menées en milieu suburbain, comme mentionné dans le chapitre 2. Ces campagnes ont englobé plusieurs scénarios LOS, NLOS et mixtes (voir la Figure 3.7), fournissant ainsi une base de données essentielle composée d'environ 33 000 matrices de canal, comme le montre le Tableau 2.4. Il est essentiel de noter que parmi les données collectées, des valeurs aberrantes sont inévitablement présentes. Par conséquent, un pré-traitement est nécessaire pour une exploitation optimale des données, comme cela sera décrit par la suite.

### 4.2.2 Pré-traitement des données

Le nettoyage des données, également connu sous le nom de pré-traitement des données ou "*Data Cleaning*", joue un rôle essentiel dans le développement d'un modèle de ML et l'analyse des données. L'objectif principal du nettoyage des données est d'assurer la précision, la fiabilité et l'adéquation des données pour l'entraînement et le test des algorithmes de ML. Cela implique l'organisation des données, la correction et l'élimination des incohérences et des inexactitudes au sein d'une base de données, afin de les préparer à une utilisation efficace dans les modèles de ML [131-134]. Dans cette optique, il est impératif de supprimer les données manquantes, telles que des trames corrompues ou des transmissions non enregistrées. Par conséquent, seules les fonctions de transfert  $H(t, f)$  avec SNR supérieur à 3 dB doivent être prises en compte.

Enfin, nous pouvons être confrontés au défi de traiter des caractéristiques qui présentent des échelles différentes comme les étalements des retards mesurés en nanosecondes et la bande de cohérence mesurée en MHz. Cette disparité rend impossible les comparaisons directes affectant ainsi la performance des algorithmes utilisés, en particulier pour les algorithmes basés sur la mesure de distances (tels que k-Nearest Neighbors (kNN), le clustering par K-means ou l'analyse en composantes principales - PCA). Pour résoudre ce problème, il est essentiel que nos données soient mises à la même échelle ou qu'elles soient normalisées [135, 136]. Les méthodes de normalisations utilisées sont décrites dans la suite de ce chapitre.

### 4.2.3 Dérivation de caractéristiques

En ML, la dérivation des caractéristiques, également connue sous le nom d'ingénierie des caractéristiques, occupe une place centrale et essentielle dans le développement de modèles ML efficaces et robustes. Elle consiste à transformer des données brutes en informations adaptées au traitement par les algorithmes de ML. La complexité des données relatives au canal de propagation nécessite une caractérisation approfondie et multidimensionnelle en termes de temps, de fréquence et d'espace, etc.

Ces caractéristiques, qui comprennent des aspects tels que les trajets multiples, l'atténuation, les angles de départ, ainsi que les effets de l'environnement sur le canal de propagation, doivent être interprétées correctement pour une analyse efficace. Une étude statistique réalisée par un statisticien ne suffit pas; une interprétation physique par un expert et la corrélation entre différents paramètres sont nécessaires pour une compréhension complète et pour développer des modèles ML précis. Cette étape fondamentale et cruciale de dérivation de caractéristiques en ML nécessite une expertise dans la caractérisation d'un canal de propagation.

#### Extraction de caractéristiques

Certaines caractéristiques des fonctions de transfert mesurées  $H(f)$  en scénario LOS et NLOS sont présentées suivant des échelles linéaires ou logarithmiques, telles que la puissance du signal, le SNR, le facteur de Rice  $K_{\text{rice}}$ , etc. Cependant, ces caractéristiques sont initialement stockées sous forme linéaire et seront ensuite normalisées à l'aide d'une méthode adaptée à leur distribution.

### Étalement des retards $\tau_{rms}$ et retard moyen $\tau_{MOY}$

La Figure 4.1 (a) présente les étalements de retards de l'ensemble des scénarios LOS en bleu et NLOS en rouge suivant une échelle linéaire puis sous forme d'ECDF (Figure 4.1 (b)). Ces résultats font nettement apparaître des étalements de retards médians de 12 ns et 68 ns respectivement en LOS et NLOS. On peut remarquer un comportement similaire pour l'étalement des retards moyens  $\tau_{MOY}$  Figure 4.2 avec des médianes LOS et NLOS respectivement de 15 ns et 77 ns.

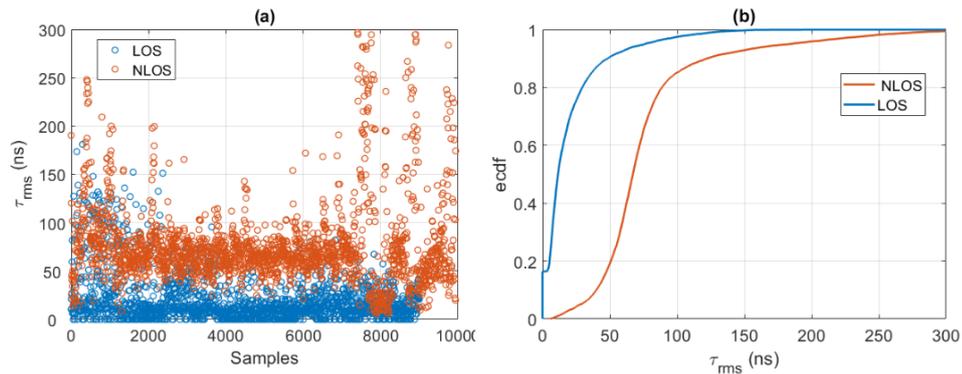


FIGURE 4.1 – Étalement des retards  $\tau_{rms}$  pour les scénarios LOS/NLOS (a) et leur ECDF (b)

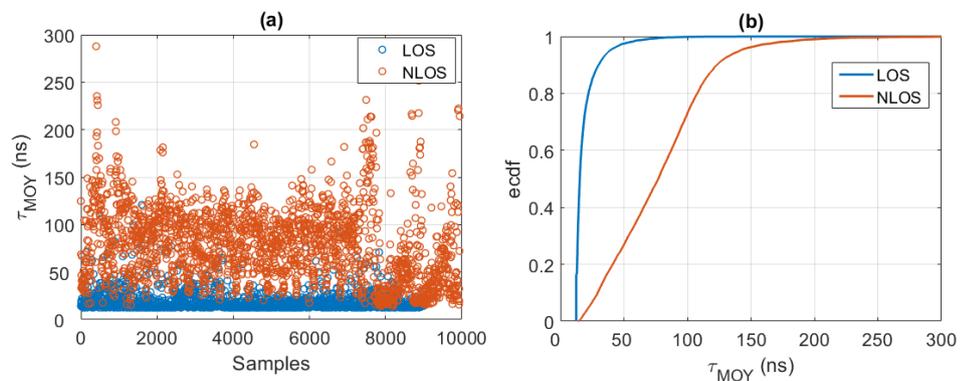


FIGURE 4.2 – Étalement des retards moyenne  $\tau_{MOY}$  pour les scénarios LOS/NLOS (a) et leur ECDF (b)

### Nombre de pics par PDP

Cette caractéristique repose sur le calcul du nombre de pics dépassant un seuil de -20 dB par rapport au maximum du PDP comme illustré dans la Figure 4.3. Chaque pic peut correspondre à un trajet spéculaire ou à la somme vectorielle de trajets multiples ayant des retards similaires. Plus de détails sont fournis dans l'annexe B.

Il est observé sur la Figure 4.4 (a), représentant le nombre de pics en LOS et NLOS, que le nombre de pics est concentré entre 0 et 10, avec une médiane à 5 pics pour un

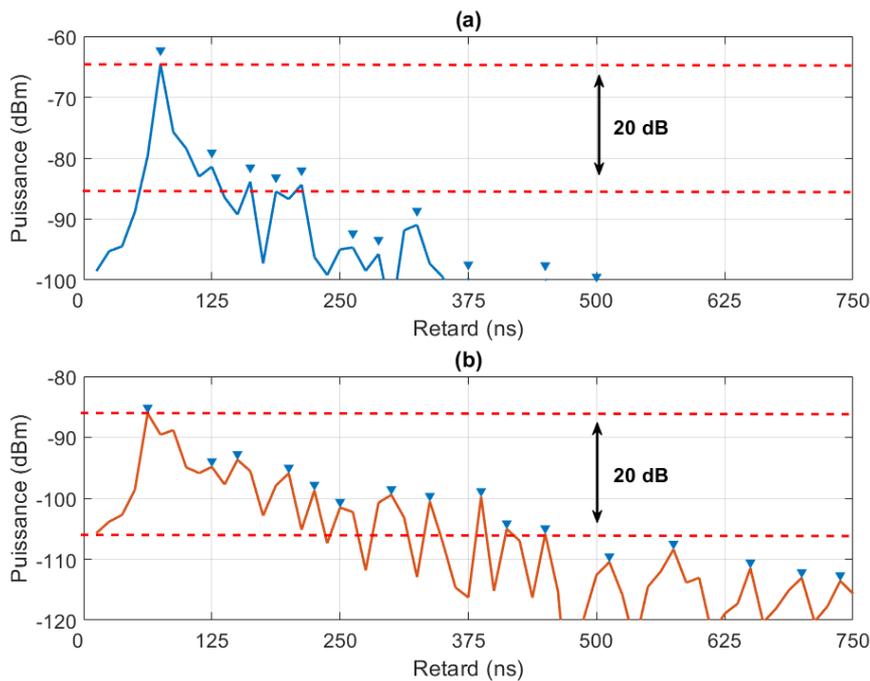


FIGURE 4.3 – Identification des pics pour un PDP LOS (a) NLOS (b)

scénario LOS et entre 10 et 40 avec une médiane à 20 pics pour un scénario NLOS (voir la Figure 4.4 (b)). Le nombre de pics par PDP peut constituer une caractéristique importante dans l'estimation des scénarios de propagation LOS/NLOS.

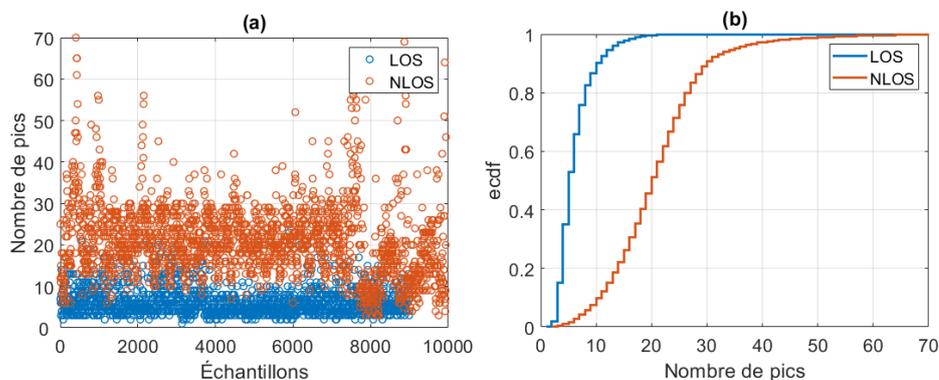


FIGURE 4.4 – Nombre de pics pour les scénarios LOS/NLOS (a) et leur ECDF (b)

### Puissance maximale d'un PDP $P_{\max}$

La puissance maximale  $P_{\max}$  de chaque PDP a été estimée. En général, pour le même canal, la puissance maximale d'un PDP en LOS est plus élevée que la puissance maximale d'un PDP en NLOS. Par conséquent, la puissance du signal peut être utilisée comme une caractéristique pour l'identification LOS/NLOS.

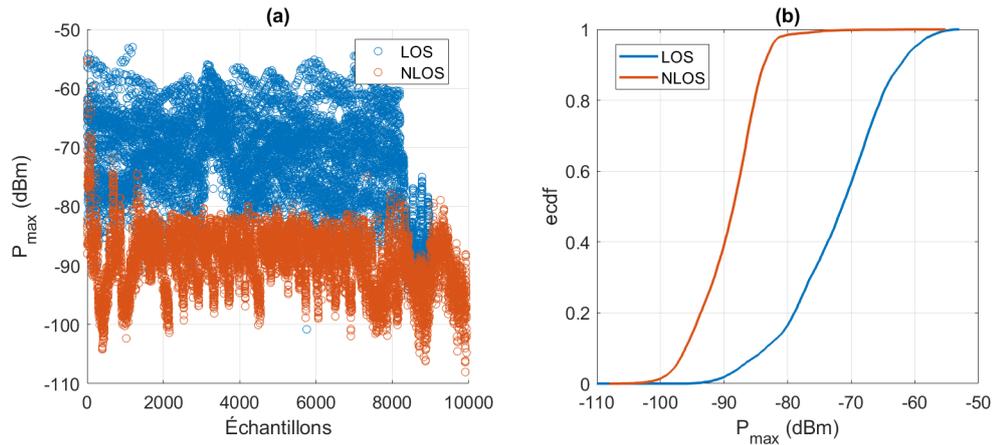


FIGURE 4.5 – Puissance de signal pour les scénarios LOS/NLOS (a) et leur ECDF (b)

La Figure 4.5 représente  $P_{\max}$  calculée pour les deux scénarios de notre base de données (a) avec leur ECDF (b). On observe un écart entre les puissances LOS/NLOS avec des médianes calculées en linéaire, respectivement de -72 dBm et -90 dBm.

#### Différence de puissance $d_p$ dB

Cette caractéristique consiste à calculer la différence en dB entre les deux trajets les plus puissants. La Figure 4.6.(a) représente les valeurs de  $d_p$  en dB pour les deux scénarios. Les médianes de ces écarts calculées en LOS et NLOS (calculées linéairement et converties en dB) sont respectivement de 15 dB et 5 dB. L'étape décrite par la suite de la sélection des caractéristiques nous permettra de déterminer si cette caractéristique est utile ou non.

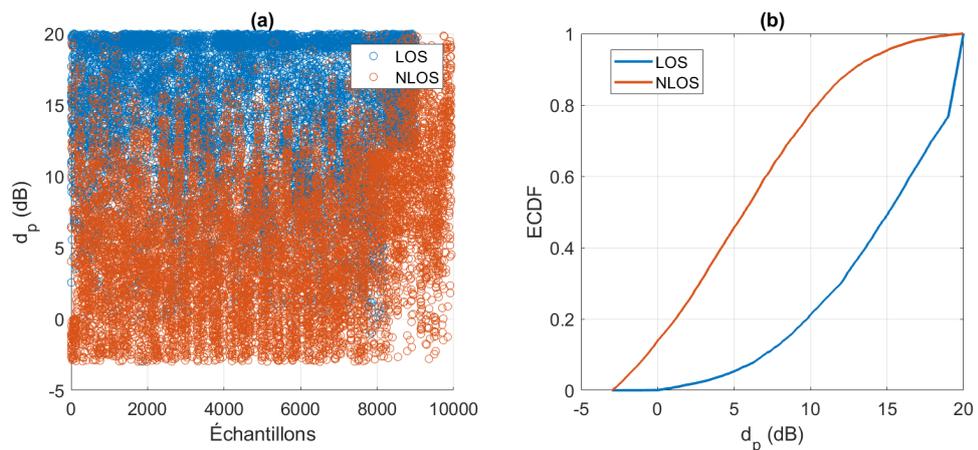


FIGURE 4.6 – Différence de puissance  $d_p$  pour les scénarios LOS/NLOS (a) et leur ECDF (b)

### Caractéristiques statistiques du PDP

Les statistiques d'ordre 1 à 4 des PDP normalisés (chaque PDP est normalisé par rapport à son maximum), telles que la moyenne, l'écart type, le coefficient de l'asymétrie (Skewness) et le coefficient d'aplatissement (Kurtosis) sont calculés .

Comme le montrent les distributions statistiques figurant dans la Figure 4.7, des variations significatives sont observées entre les scénarios de propagation LOS (en bleu) et NLOS (en rouge). On constate que les distributions de moyennes et d'écart-types sont plus étalées en NLOS qu'en LOS. De plus, les valeurs de l'asymétrie et d'aplatissement sont moins élevées en NLOS qu'en LOS, ce qui peut être exploité comme caractéristiques de classification.

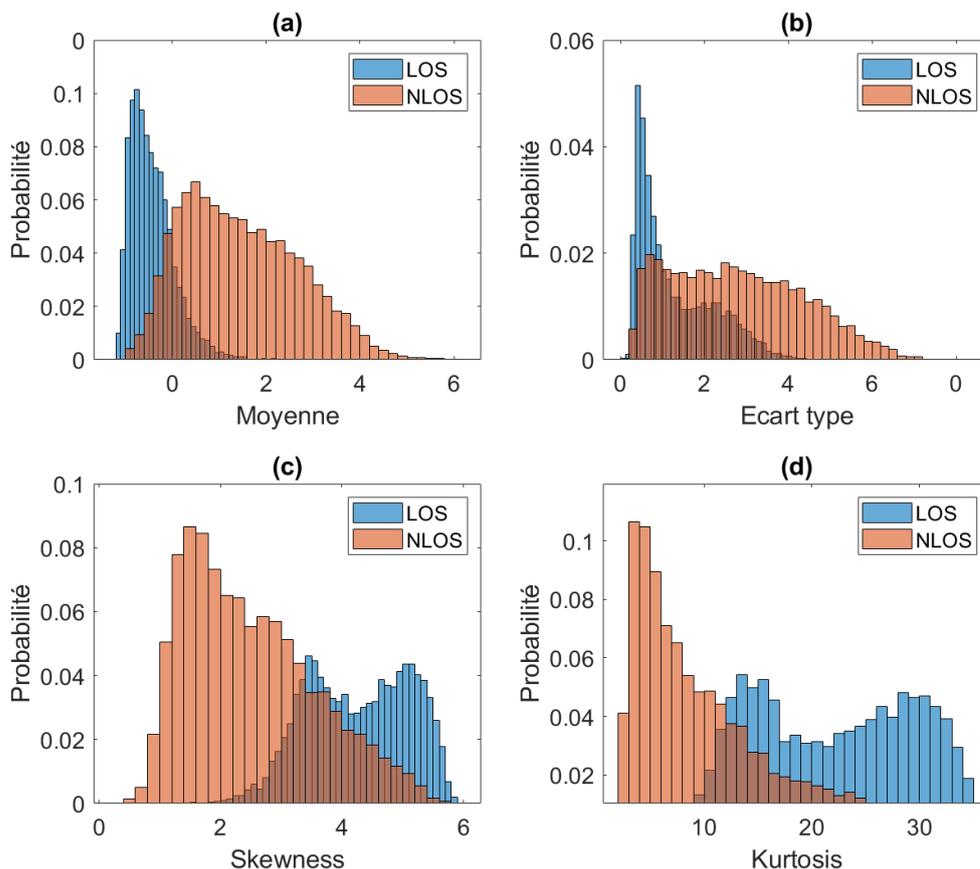


FIGURE 4.7 – Moyenne (a), et écart type (b), asymétrie (c) et Kurtosis (d)

### Facteur de rice $K_{rice}$

Le facteur de Rice  $K_{rice}$  peut être une caractéristique importante pour la classification des scénarios de propagation LOS/NLOS. Comme le montre la Figure

4.8, les ECDF associées à chaque scénario ne se croisent pas et présentent un écart de 15 dB, les facteurs de Rice médians pour le LOS et NLOS étant respectivement de 0 dB et 15 dB. Il existe une corrélation entre l'amplitude du facteur de Rice  $K_{\text{rice}}$  et la présence des conditions LOS et NLOS.

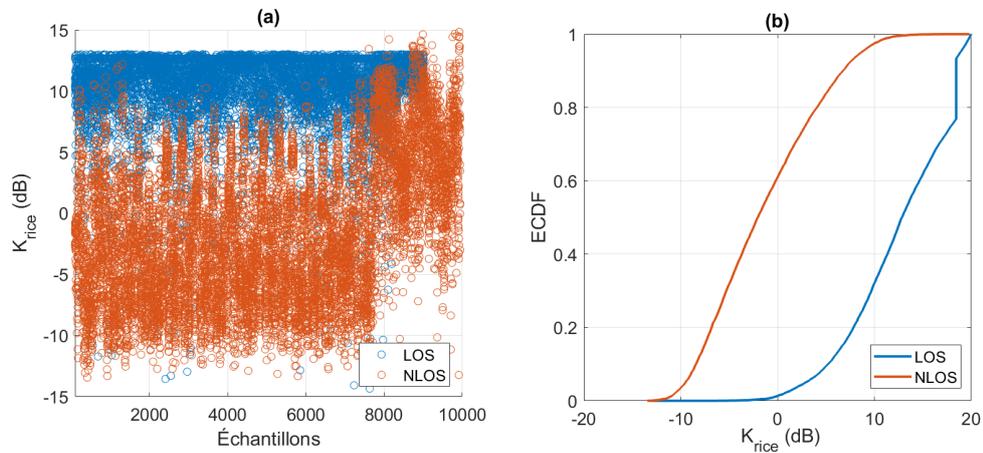


FIGURE 4.8 – Facteur de rice  $K_{\text{rice}}$  pour les deux scénarios LOS/NLOS (a) et leur ECDF (b)

### Pente du PDP

La pente de chaque PDP est déterminée par la valeur du coefficient directeur d'une fonction linéaire. Elle est calculée en sélectionnant la zone linéaire de la PDP, excluant ainsi les trajets multiples forts. Cette valeur est obtenue par ajustement linéaire des points de puissance situés dans un intervalle défini entre  $P_{\text{max}}$  et  $P_{\text{max}} - 20$  dB. Ces pentes ont été calculées pour l'ensemble des PDP dans les deux scénarios LOS et NLOS. Dans la Figure 4.9.(a), on observe les valeurs des pentes en fonction des échantillons LOS/NLOS. Il est constaté que les pentes en LOS sont plus faibles qu'en NLOS, avec des valeurs médianes respectives de -3 et -0,5 la Figure 4.9.(b).

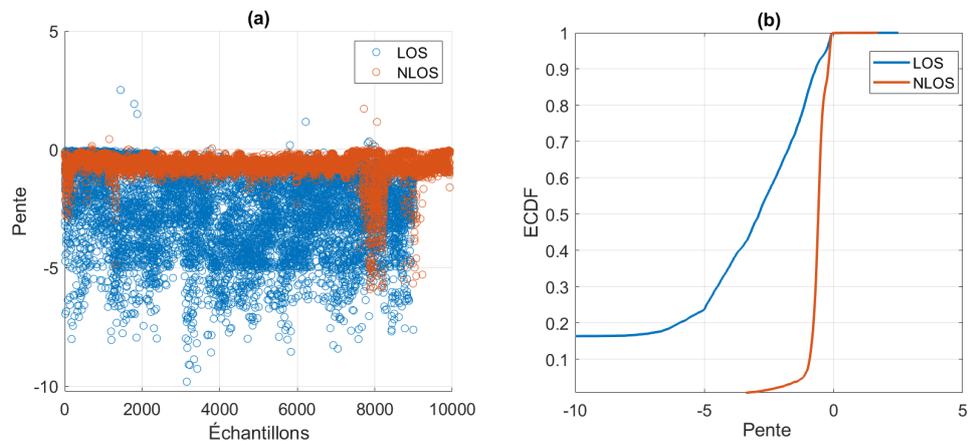


FIGURE 4.9 – Valeurs des pentes des PDP pour les deux scénarios LOS/NLOS (a) et leur ECDF (b)

### Bande de cohérence $B_c$

La Figure 4.10.(a) montre les bandes de cohérence à 0,7 des fonctions de transfert. On constate que la valeur médiane de  $B_c$  est pour les scénarios LOS et NLOS respectivement de 9 MHz et 1,5 MHz. En conclusion, la bande de cohérence apparaît comme une caractéristique pertinente pour notre modèle ML.

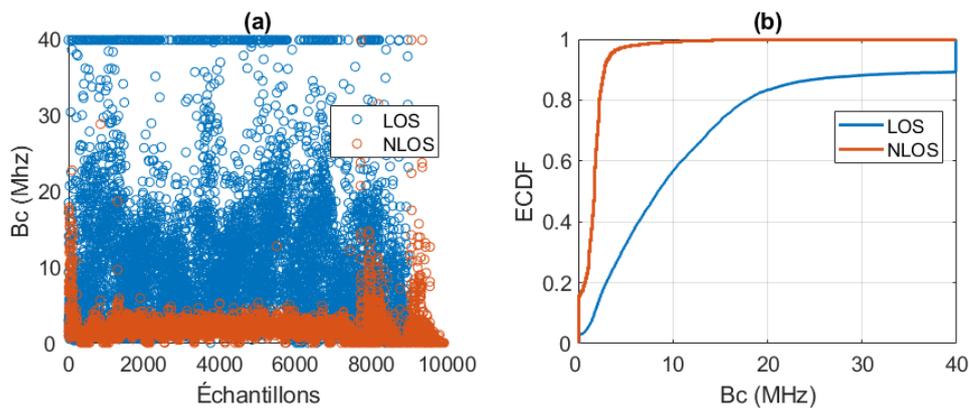


FIGURE 4.10 – Bande de cohérence à 0.7 (a) et leur ECDF (b) pour les deux scénarios LOS/NLOS

### Nombre et largeur médiane des évanouissements fréquentiels

La Figure 4.11 illustre deux fonctions de transfert en LOS et NLOS. Ces courbes révèlent un nombre d'évanouissements plus important en NLOS qu'en LOS. Au contraire, leur largeur (flèches à deux têtes) est plus importante en LOS qu'en NLOS.

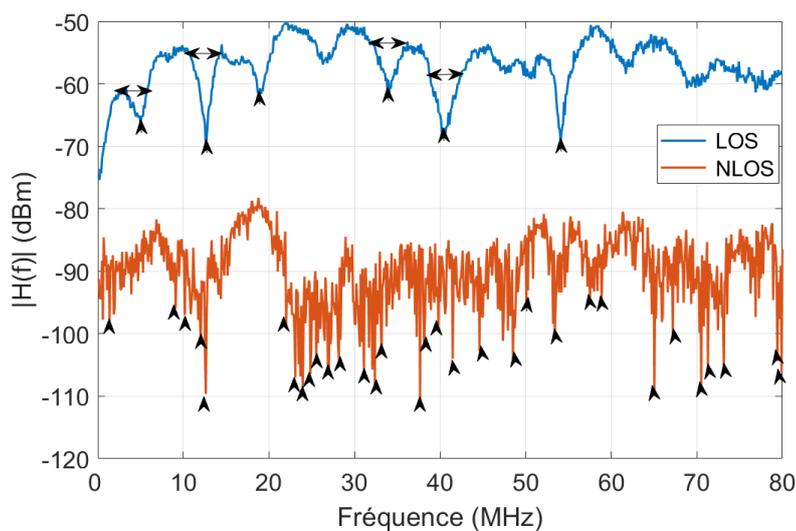


FIGURE 4.11 – Évanouissements fréquentiels et leur largeur dans les fonctions de transfert pour les deux scénarios de propagation LOS/NLOS

On observe dans la Figure 4.12 qu'il y a un écart entre les deux scénarios de propagation LOS/NLOS en termes de nombre et de largeur médiane des évanouissements fréquentiels. L'analyse statistique montre un nombre médian d'évanouissements de 61 et 181 respectivement en LOS et NLOS, les médianes des largeurs fréquentielles associées étant de 195 kHz et 136 kHz. Cette observation nous mène à considérer le nombre d'évanouissements fréquentiels et la bande de fréquence associée à ces évanouissements comme des caractéristiques pertinentes. Plus de détails sont décrits dans l'annexe B.

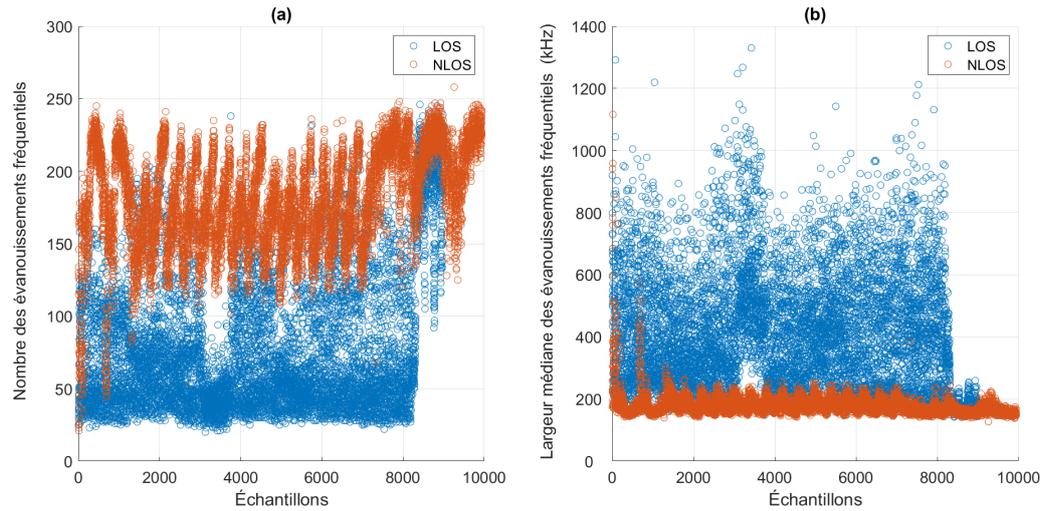


FIGURE 4.12 – Nombre et largeur médiane des évanouissements fréquentiels pour les deux scénarios de propagation LOS/NLOS

### Retard de groupe $g_D$

Un canal de propagation sélectif en fréquence modifie non seulement l'amplitude du signal, mais également sa phase. Sur ce dernier point, le retard de groupe  $g_D$  pour un échantillon, défini comme la dérivée de la phase de la fonction de transfert du canal en fonction de la fréquence  $\Phi_H = \arg(H(f))$ , s'exprime par la relation suivante.

$$g_D = -\frac{1}{2\pi} \frac{d\Phi_H}{df} \quad (4.1)$$

La Figure 4.13 montre le retard de groupe  $g_D$  en fonction de la fréquence pour deux fonctions de transfert en LOS (a) et NLOS (b). On observe que le retard de groupe dans le scénario NLOS présente beaucoup plus de variations et de fluctuations par rapport au scénario LOS. Pour donner une meilleure description quantitative du retard de groupe, plusieurs paramètres statistiques ont été calculés, notamment la moyenne et la variance de  $|g_D|$ .

La Figure 4.14.(a) présente la moyenne de  $|g_D|$ , et montre qu'il existe une différence significative entre les scénarios LOS et NLOS. En effet, 90% des valeurs se situent

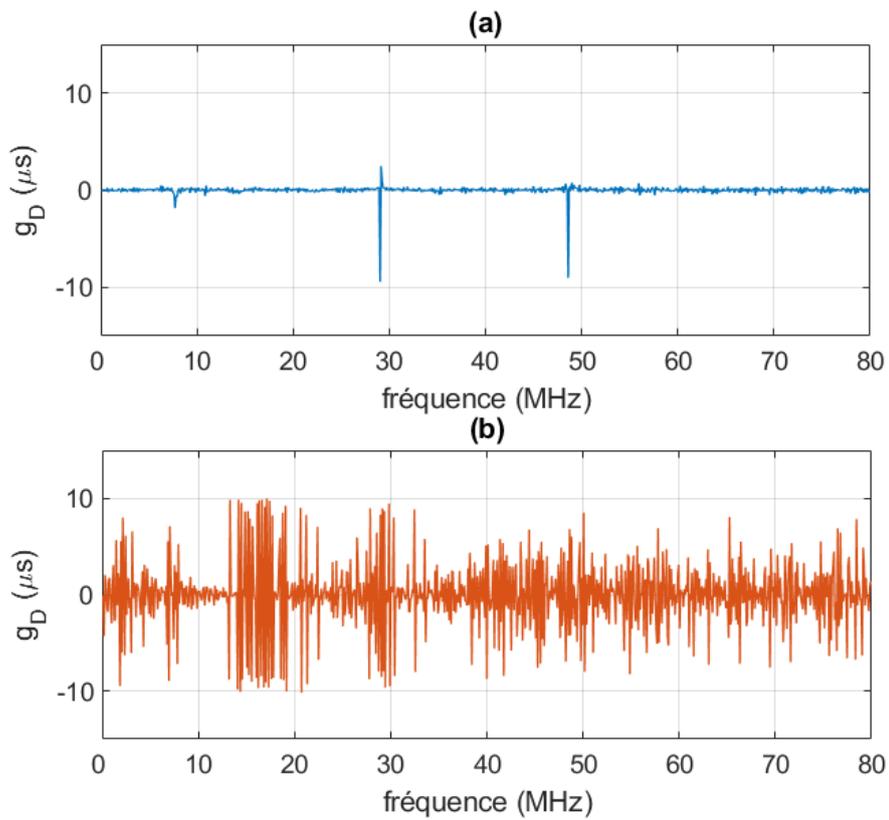


FIGURE 4.13 – Retard de groupe pour un échantillon LOS (a) et NLOS (b)

respectivement entre  $0,1 \mu\text{s}$  et  $1 \mu\text{s}$  pour le LOS, et entre  $0,2 \mu\text{s}$  et  $2,3 \mu\text{s}$  pour le NLOS, avec des valeurs médianes respectives de  $0,2 \mu\text{s}$  et  $1,2 \mu\text{s}$  (voir Figure 4.14.(b)). La méthode de calcul de cette dérivée est expliquée dans l'annexe B.

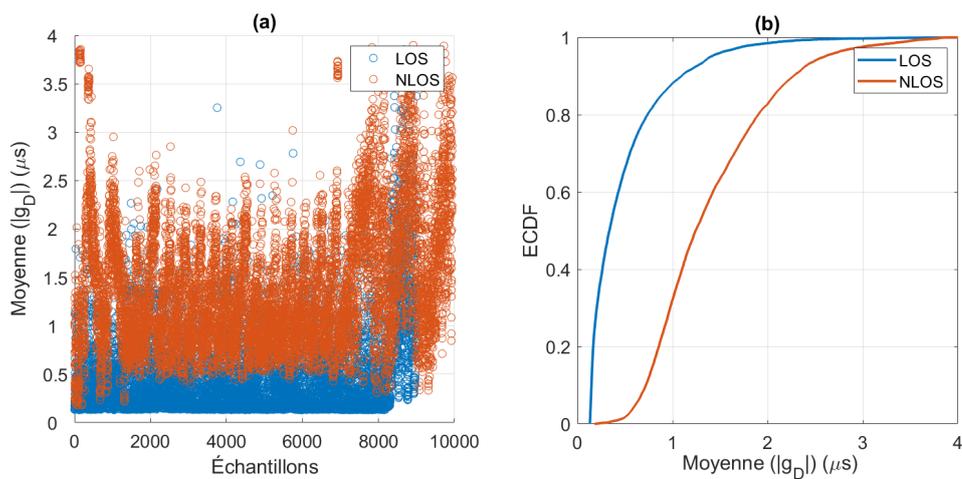


FIGURE 4.14 – Moyenne de la valeur absolu  $g_D$  LOS/NLOS (a) et leur ECDF (b)

### Rapport signal sur bruit SNR (dB)

Le SNR est plus élevé dans un canal LOS, comme dans le cas d'un canal de Rice, que dans un NLOS, en particulier pour un canal riche aux multitrajets de Rayleigh.

La Figure 4.15.(a) montre que plus de 90% des valeurs de SNR en (dB) pour un scénario LOS sont supérieures à celles de NLOS, avec des valeurs médianes respectives de 18 dB et 9 dB (Figure 4.15.(b)). L'ajout de cette caractéristique comme entrée du modèle semble pertinente.

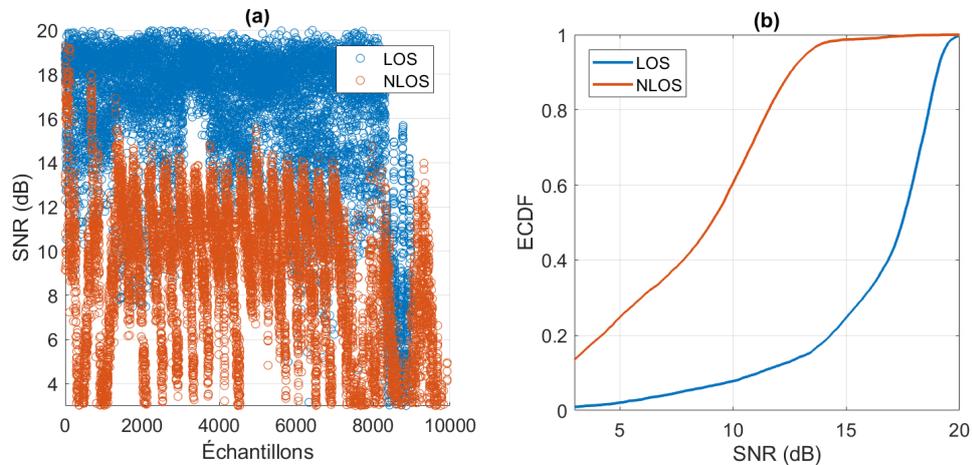


FIGURE 4.15 – SNR en (dB) LOS/NLOS (a) et leur ECDF (b)

### — Caractéristiques MIMO

Dans cette partie, les caractéristiques du canal MIMO prises en compte sont la corrélation entre les antennes d'émission  $\rho_{Tx}$  et de réception  $\rho_{Rx}$ , ainsi que les angles de départ (AoD) du trajet fort en azimuth et en élévation et les étalements angulaires.

#### Corrélation entre les antennes d'émission $\rho_{Tx}$ et de réception $\rho_{Rx}$

Les coefficients  $\rho_{Rx}$  entre les 8 antennes de réception ont été calculés en moyennant les coefficients de corrélation entre les antennes espacées d'une antenne, c'est-à-dire (1 et 3), (2 et 4), (3 et 5), (4 et 6), (5 et 7) et finalement (6 et 8) (Figure 4.16.(b)). De la même manière, les coefficients de corrélation à l'émission  $\rho_{Tx}$  ont été calculés séparément sur ULA-V et sur ULA-H (Figure 4.16.(a)).

Les coefficients de corrélation  $\rho_{Rx}$  et  $\rho_{Tx}$  pour ULA-H et ULA-V sont tracés respectivement dans la Figure 4.17 pour les scénarios LOS et NLOS en fonction des échantillons et de leurs ECDF. On observe que  $\rho_{Rx}$  et  $\rho_{Tx}$  ULA-H ont des comportements similaires, où la majorité des points LOS présentent des coefficients plus élevés que NLOS, avec des médianes respectives de 0,76 et 0,29. Environ 10%

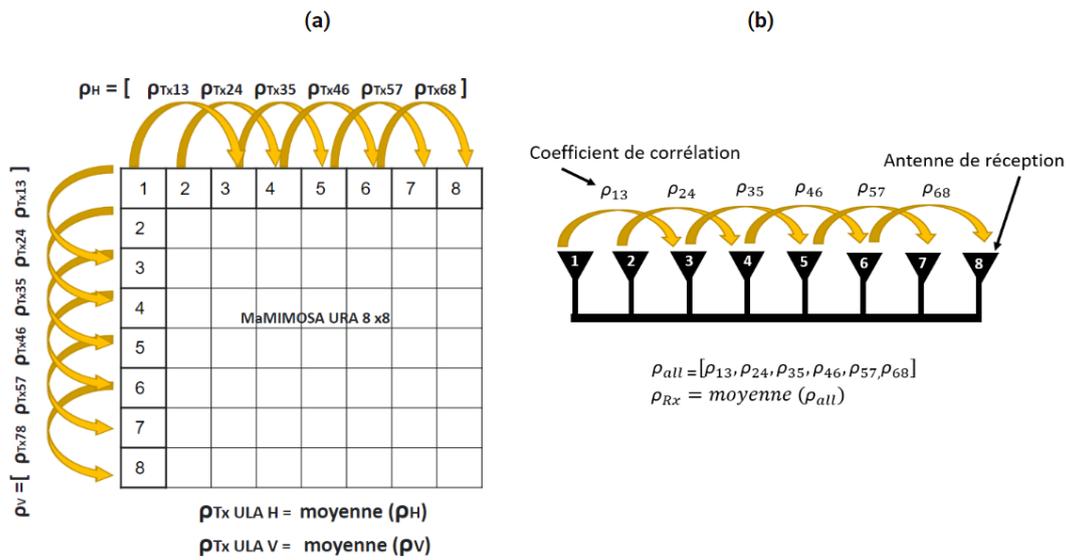


FIGURE 4.16 – Méthode de calcul des coefficients de corrélation  $\rho_{Tx}$  en ULA V et ULA H et  $\rho_{Rx}$

des coefficients de corrélation pour le scénario LOS sont inférieurs à 0,5, ce qui est dû à des points de scénario LOS éloignés de l'émetteur Tx. D'autre part,  $\rho_{Tx}$  ULA-V montre des corrélations plus élevées, avec des médianes de 0,90 et 0,72 respectivement pour LOS et NLOS.

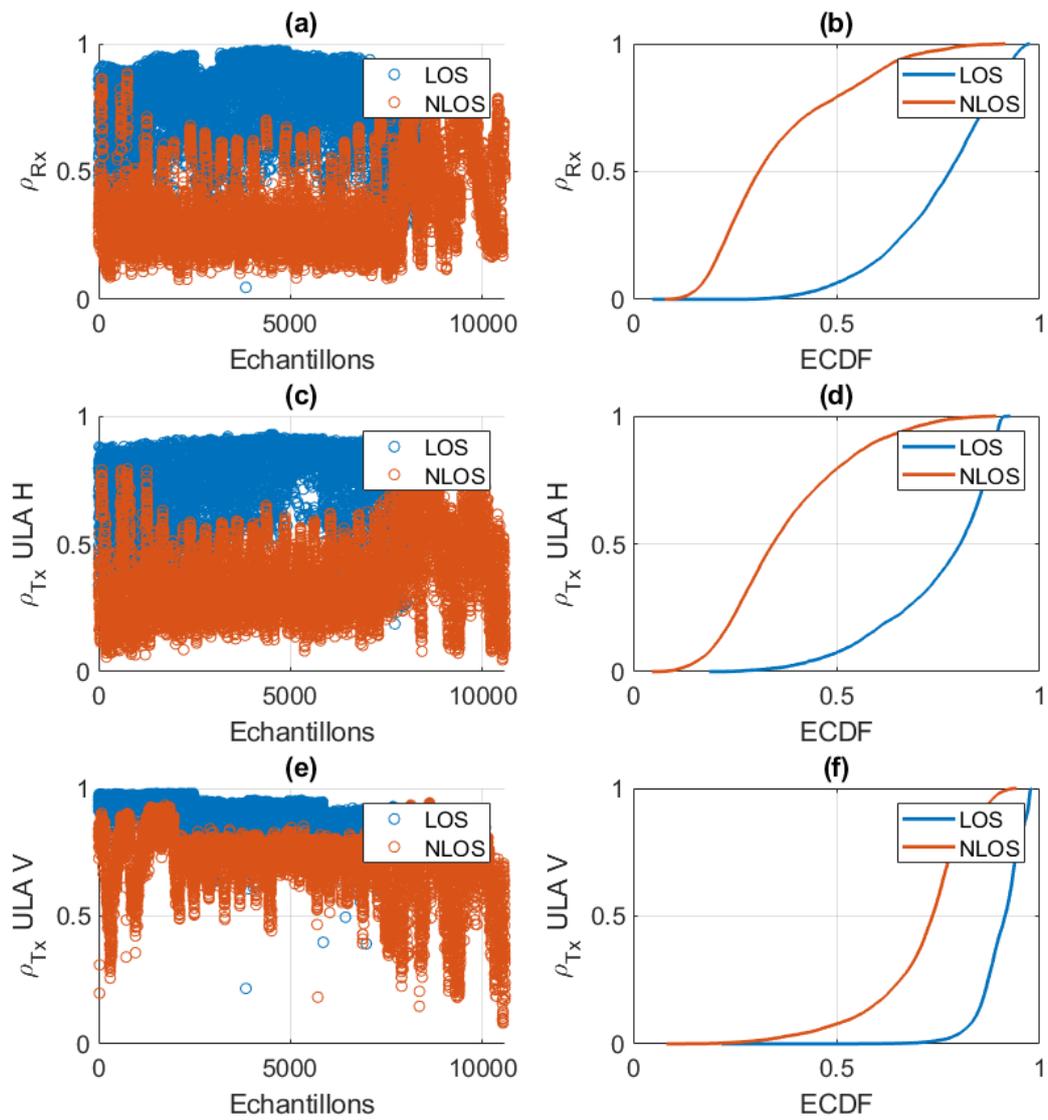


FIGURE 4.17 – Coefficients de corrélation  $\rho_{Rx}$  LOS/NLOS (a) et leur ECDF (b),  $\rho_{Tx}$  ULA H LOS/NLOS (c) et leur ECDF (d) et  $\rho_{Tx}$  ULA V LOS/NLOS (e) et leur ECDF (f)

### Étalement Angulaire

En utilisant la technique de beamforming de Bartlett [137] et les formules du 3GPP pour calculer l'étalement angulaire (équation (1.38)), les angles de départ (AoD) en azimut et en élévation du trajet de plus forte amplitude. (Figure 4.18).

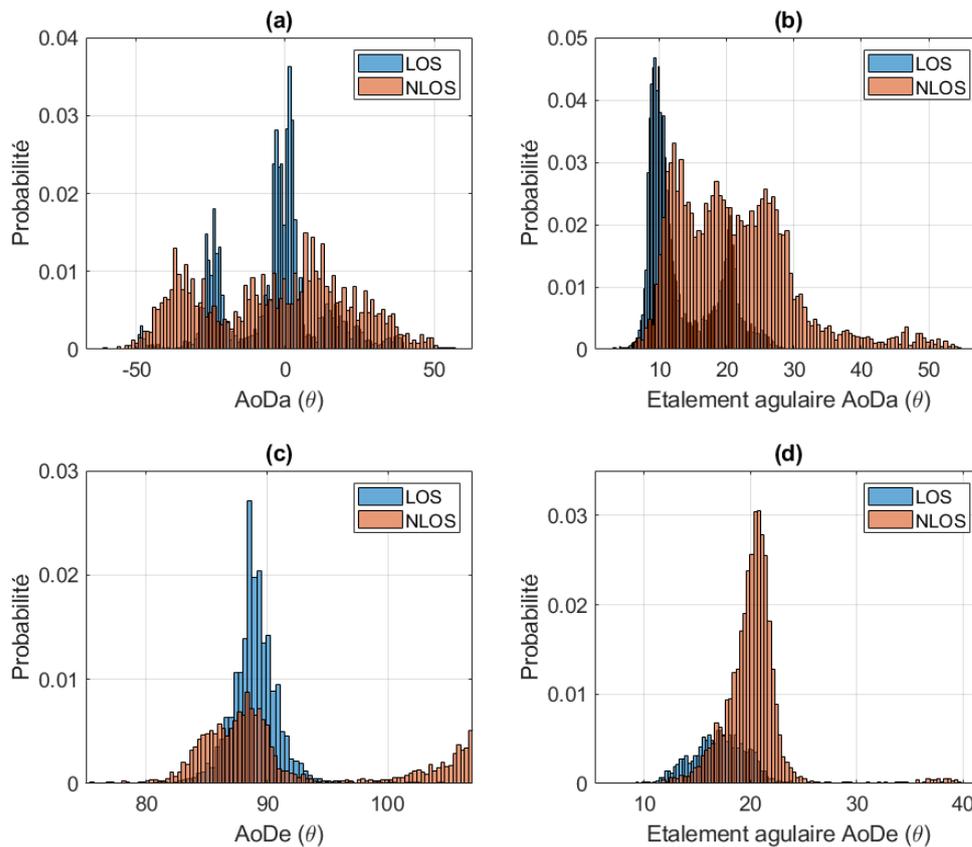


FIGURE 4.18 – Histogramme des angles de départ en azimut (AoDa) avec leur étalement angulaire (b) et en élévation (AoDe) avec leur étalement angulaire (d)

Le Tableau 4.1 présente une synthèse des caractéristiques du canal candidates pour le ML.

### Normalisation des Caractéristiques

Chaque caractéristique a été normalisée avant d'être combinée et comparée dans la prochaine étape de sélection des caractéristiques. Cela garantit que toutes les caractéristiques sont considérées comme étant d'égale importance, quelles que soient leurs échelles. Le Tableau 4.2 résume les types de mise à l'échelle utilisés.

	Caractéristique
À partir des PDP	$\tau_{rms}$ $\tau_{MOY}$ Nombre de pics $d_p$ Moyenne $P_{max}$ Écart type Skewness (S) Kurtosis (K) $K_{rice}$ Pente SNR
À partir de $H(f)$	$Bc$ Nombre des évanouissements fréquentiels (N.E.F) Largeur médiane des évanouissements fréquentiels (L.M.E.F) Moyenne $ g_D $ AoD Élévation (AoDe) AoD Azimut (AoDa) Étalement angulaire AoDe (E.A AoDe) Étalement angulaire AoDa (E.A AoDa) $\rho_{Rx}$ $\rho_{Tx}$ ULA H ( $\rho_{TxH}$ ) $\rho_{Tx}$ ULA V ( $\rho_{TxV}$ )

TABLEAU 4.1 – Caractéristiques du canal candidates pour le ML

Méthode	Formule	Description	Utilisation
Z-Score	$\frac{x - \mu}{\sigma}$	Met à l'échelle les données pour qu'elles aient une moyenne de 0 et un écart-type de 1	Données suivant une distribution gaussienne (normale)
Range	$\frac{x - x_{\min}}{x_{\max} - x_{\min}}$	Met les données à l'échelle dans la plage [0, 1].	Données avec des distributions uniformes ou non spécifiées
Robust Scaling	$\frac{x - \text{médiane}(x)}{\text{mad}(x)}$	Met les données à l'échelle de manière à avoir une médiane de 0 et un écart absolu médian de 1	Données avec des valeurs aberrantes ou des distributions non gaussiennes

TABLEAU 4.2 – Méthodes de normalisation, standardisation et mise à l'échelle pour chaque caractéristique  $x$  [135][136].

#### 4.2.4 Sélection des Caractéristiques

Pour élaborer un modèle de classification robuste LOS/NLOS, la sélection des caractéristiques est une étape cruciale. Le but de cette sélection est d'améliorer le modèle en ne conservant que les caractéristiques apportant de la valeur, tout en éliminant les autres. Le modèle résultant sera également plus simple à interpréter et plus efficace du point de vue du calcul. Les approches de sélection de caractéristiques

sont généralement regroupées en trois catégories principales : par filtrage (voir la Figure 4.20), par enveloppement (wrapper) (voir la Figure 4.21), et intégrées (embedded) (voir la Figure 4.19).

Par la suite, une nouvelle méthode hybride est introduite, nous permettant de sélectionner les caractéristiques en nous basant à la fois sur des méthodes conventionnelles et sur l'expertise dans la caractérisation du canal de propagation (Figure 4.22).

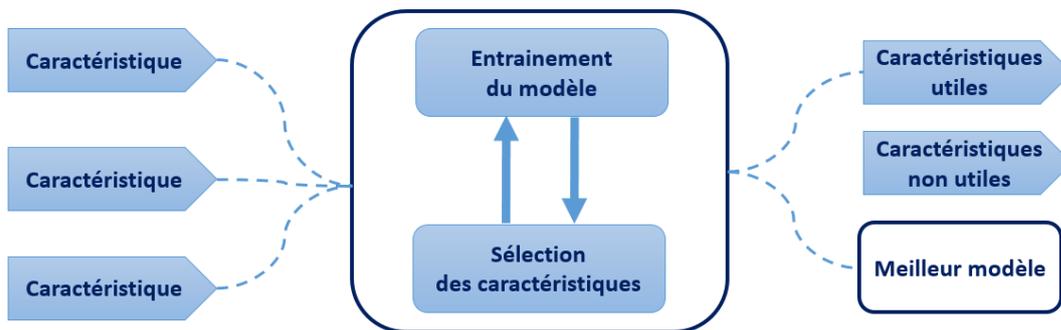


FIGURE 4.19 – Représentation de la méthode Intégrée (Embedded)

### Sélection de caractéristiques par filtrage

Pour cette approche (Figure 4.20), deux techniques sont utilisées : le seuil de variance et le t-Test, décrits par la suite.

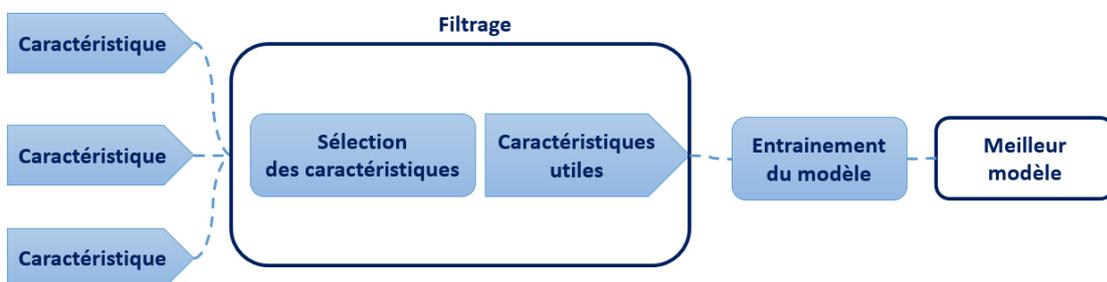


FIGURE 4.20 – Représentation de la méthode de filtrage

— **Seuil de variance :**

Une technique utilisée pour la sélection de caractéristiques dans l'apprentissage automatique repose sur leur variance proportionnelle. La variance proportionnelle d'une caractéristique spécifique est le résultat obtenu en calculant la variance de cette caractéristique individuelle et en divisant cette valeur par la somme totale des variances de toutes les caractéristiques. Cette mesure quantifie la contribution relative d'une caractéristique à la variabilité totale des données. Une caractéristique présentant une variance proportionnelle élevée aura un impact plus important sur la variabilité globale. Pour calculer la variance proportionnelle  $V_p$  pour un ensemble de caractéristiques, la formule suivante :

$$V_p = \frac{V}{\sum_{i=1}^N V_i} ,$$

avec  $V$ , le vecteur contenant les variances de chaque caractéristique normalisée  $V = \{V_1, V_2, V_3, \dots, V_N\}$ , où  $N$  est le nombre de caractéristiques.  $V_p$  est le vecteur de variances proportionnelles, où chaque élément est la variance d'une caractéristique divisée par la somme des variances de toutes les caractéristiques. Cela donne un vecteur où chaque composante représente la proportion de la variance totale attribuée à la caractéristique correspondante.

Un seuil est utilisé pour déterminer quelles caractéristiques conserver, distinguant celles ayant une variabilité significative de celles présentant une variance moindre. Un seuil bien choisi doit séparer efficacement les caractéristiques à haute variance des caractéristiques à faible variance qui peuvent être rejetées. L'élimination excessive de la variance peut avoir un impact négatif sur les performances du modèle. Une technique pour choisir ce seuil consiste à effectuer des tests itératifs sur différents seuils et à évaluer des modèles de ML. Cette technique est particulièrement utile pour les ensembles de données comportant un grand nombre de caractéristiques, car elle peut contribuer à améliorer les performances du modèle en éliminant les caractéristiques non pertinentes ou redondantes [138, 139].

— **t-Test :**

Le test t est une méthode statistique utilisée en classification en ML pour déterminer la signification des caractéristiques individuelles dans la différenciation entre les classes. Si une différence significative de moyennes est présente, cela peut guider la sélection des caractéristiques. Il est particulièrement utile lorsqu'il s'agit de caractéristiques numériques et d'une variable de réponse catégorielle. Le test t évalue si les moyennes de deux groupes sont statistiquement différentes l'une de l'autre. Dans le contexte de la sélection de caractéristiques pour le ML, les caractéristiques présentant des moyennes distinctes entre deux groupes peuvent contribuer de

manière significative à la performance du modèle [140-143].

Pour ce faire, la bibliothèque "scikit-learn feature selection" en Python a été utilisée (voir l'annexe B). Cette approche s'avère particulièrement bénéfique dans le contexte d'ensembles de données complexes, permettant ainsi d'identifier les caractéristiques ayant un impact significatif sur la prédiction en se concentrant sur les variations statistiquement importantes entre les groupes.

### Sélection de caractéristiques par Enveloppement

Pour notre problème de classification LOS/NLOS, l'approche par enveloppement (Figure 4.21) avec la méthode de la sélection ascendante des caractéristiques (forward feature selection (FFS)) est recommandée pour un problème de classification [144] [145].

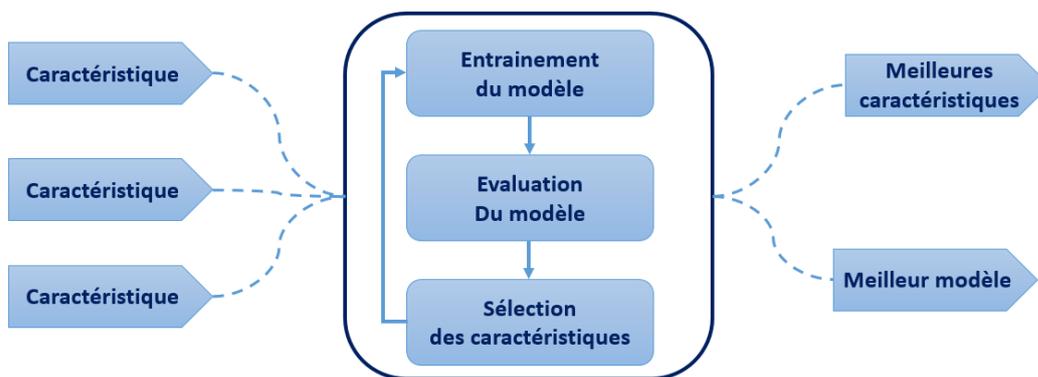


FIGURE 4.21 – Représentation de l'approche par enveloppement (wrapper)

#### — Sélection ascendante des caractéristiques :

Un processus itératif vise à identifier les caractéristiques les plus importantes d'un ensemble de données en ajoutant progressivement des caractéristiques, afin d'optimiser les performances du modèle. Par conséquent, la sélection des caractéristiques est effectuée en même temps que l'entraînement du modèle, qui est décrit par la suite.

### Sélection des caractéristiques - Méthode Hybride

Pour cette nouvelle méthode (Figure 4.22), les caractéristiques sont regroupées en trois catégories distinctes (Tableau 4.3). Premièrement, celles qui décrivent la variabilité du gain  $|H(f)|^2$  de la fonction de transfert. Deuxièmement, celles qui

décrivent la variabilité des PDP  $|h(\tau)|^2$ . Enfin, celles qui décrivent la variabilité de la réponse complexe  $h(\tau)$  ou  $H(f)$ .

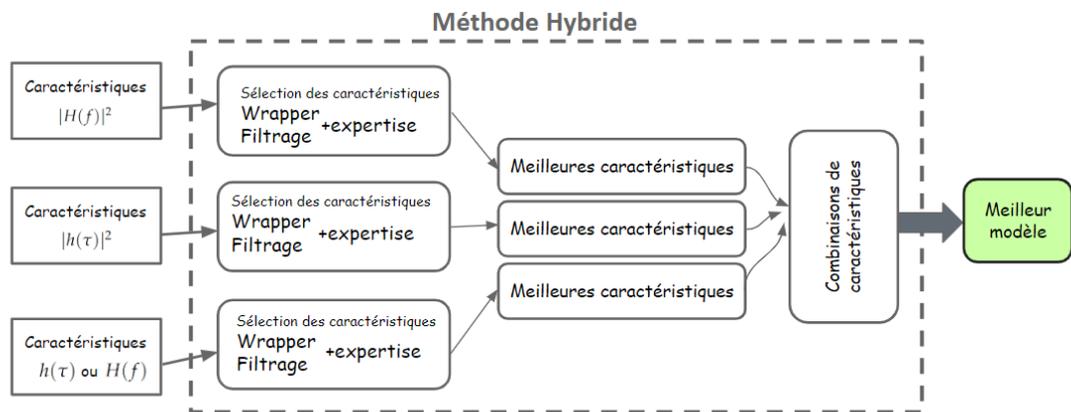


FIGURE 4.22 – Représentation de la nouvelle méthode Hybride

Les résultats obtenus grâce aux deux techniques précédemment décrites ont été utilisés et combinés à la méthode hybride pour affiner la sélection des caractéristiques appropriées en vue de la classification LOS et NLOS. Cette analyse prend en compte des informations sur le canal dans les domaines des retards, des fréquences et de la phase. Ce dernier, se rapportant uniquement aux informations sur la phase, inclut les informations angulaires, ainsi que les corrélations à l'émission dans contexte de communication MIMO.

$ h(\tau) ^2$	$ H(f) ^2$	$h(\tau)$ ou $H(f)$ complexe
Nombre de pics	N.E.F	$Moyenne(g_D)$
$\tau_{rms}$	L.M.E.F	<b>Caractéristiques MIMO :</b>
$\tau_{MOY}$	$B_c$	
$P_{max}$		
$d_p$		
SNR		
Pente		
$K_{rice}$		
Kurtosis		
Skewness		
Moyenne, Écart type		
		AoDe
		AoDa
		E.A AoDe
		E.A AoDa
		$\rho_{Rx}$
		$\rho_{TxH}$
		$\rho_{TxV}$

TABEAU 4.3 – Trois sous-ensembles des caractéristiques de la variabilité du canal

### 4.3 Algorithmes utilisés pour la classification LOS-NLOS

Il existe des dizaines d'algorithmes du ML supervisés et non supervisés, chacun ayant une façon différente d'aborder l'apprentissage. Le choix de l'algorithme dépend de la nature et de la taille des données, du type d'application et de l'utilisation de la mémoire, de la vitesse et de la précision de l'apprentissage. Dans ce chapitre, on limitera la description aux deux algorithmes utilisés pour notre modèle de classification binaire de l'apprentissage supervisé : kNN (k-Nearest Neighbors) et le SVM (Support Vector Machine).

Le kNN est apprécié pour sa simplicité et son interprétabilité, ce qui le rend potentiellement bien adapté aux scénarios de propagation LOS et NLOS dans les réseaux 5G, où une classification rapide et efficace basée sur la proximité est cruciale [146]. D'un autre côté, le SVM excelle dans la séparation de données complexes et non linéaires, offrant une précision élevée et une robustesse, même avec des ensembles de données limités. Cette capacité de généralisation supérieure le rend éventuellement fort à-propos pour la classification précise des scénarios de propagation LOS et NLOS, incluant des caractéristiques et des données non linéaires [147].

#### Algorithme kNN - k-Nearest Neighbors

L'algorithme de classification kNN a été largement utilisé dans les applications de la science des données et du ML en raison de sa simplicité de fonctionnement et de ses performances. Le kNN est basé sur le calcul simple des k distances les plus proches du point à prédire, par rapport à ses voisins, ainsi que le principe de majorité.

L'idée clé de la méthode kNN est de prédire un point de données de test (point vert) par la règle de la majorité entre les deux classes (Figure 4.23 : classe 1 rectangle bleu et classe 2 triangle rouge) selon le nombre de k voisins choisis, c'est-à-dire, que le point de données de test est prédit avec la classe principale de ses k points voisins de données d'entraînement les plus similaires dans l'espace des caractéristiques. Dans l'exemple illustré dans la Figure 4.23, si  $k=1$  la donnée de test est considérée comme étant un carré, mais si  $k=3$  la donnée de test est considérée comme étant un triangle. Il suffit maintenant de choisir la valeur du k voisins, en fonction de l'erreur de validation du modèle. Dans la suite, la façon dont on choisit et valide cet hyperparamètre k est décrite.

**Fonctionnement du kNN :** La plupart des classificateurs kNN utilisent des distances euclidiennes simples pour mesurer les écarts entre des exemples représentés sous forme d'entrées vectorielles [148, 149]. Il s'agit d'un algorithme

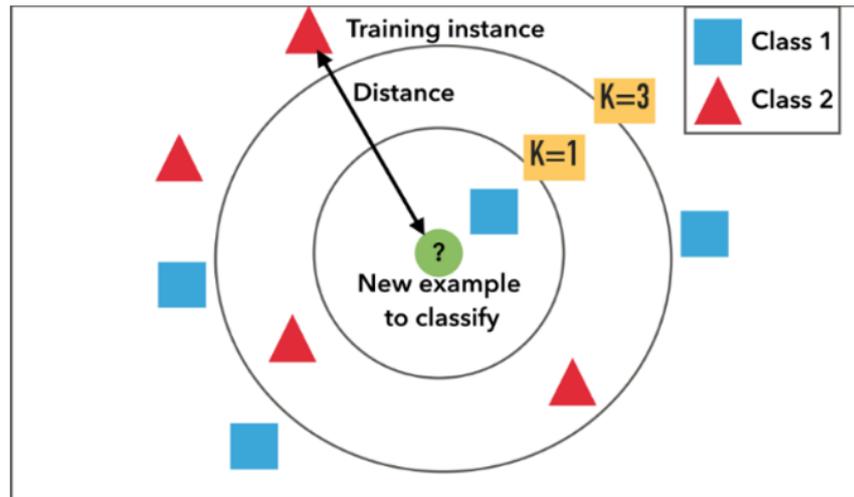


FIGURE 4.23 – Fonctionnement du kNN  
(<https://medium.com/@rndayala>)

d'apprentissage non paramétrique et basé sur des instances.

— Non paramétrique :

Il ne fait aucune hypothèse sur la distribution des données sur laquelle il repose. Il est utile car les données pratiques disponibles dans le monde réel n'obéissent pas souvent à des hypothèses théoriques. Ainsi, kNN est privilégié pour sa simplicité, son efficacité, son faible temps de calcul et sa facilité d'interprétation des résultats.

— Basé sur l'instance :

"Instance" est un anglicisme, qui signifie "cas" ou "exemple". Notre algorithme n'apprend pas explicitement le modèle; il essaie plutôt de mémoriser des exemples d'entraînement utilisés comme "connaissance" pour la phase de prédiction.

Habituellement, la distance euclidienne est utilisée comme mesure de la distance, mais d'autres métriques sont parfois plus pertinentes, telles que la distance de Minkowski, Manhattan, la similarité cosinus, etc. Ensuite, le point est assigné à la classe parmi ses  $k$  voisins les plus proches (où  $k$  est un entier positif). La distance euclidienne peut être expliquée comme suit :

$$\text{Distance Euclidienne} = \sqrt{\sum_{i=1}^M (x_i - y_i)^2} ,$$

avec  $M$  le nombre des caractéristiques (dimensions) utilisées pour l'apprentissage.

La distance est calculée entre la donnée de test  $x_i$  et chaque observation d'entraînement  $y_i$ , puis on prend les  $k$  distances voisines les plus proches du point de

test. Nous appellerons "k" les points de l'ensemble des données les plus proches de x. k est généralement impair pour éviter les situations d'égalité. L'hyperparamètre k est déduit par une évaluation des erreurs des données d'entraînement et de validation en fonction de k. Plus de détails sur l'algorithme sont présentés dans l'annexe B.

### Algorithme SVM

Les machines à vecteurs de support (SVM) sont des algorithmes d'apprentissage automatique supervisés qui peuvent être utilisés pour résoudre une variété de problèmes, tels que la classification, la régression et la détection d'anomalies. Les SVM sont appréciés pour leurs solides fondements théoriques, leur grande adaptabilité et leur facilité d'utilisation.

**Fonctionnement du SVM :** Le concept fondamental des SVM réside dans la séparation des données en classes au moyen d'une frontière ou séparateur, tout maximisant la distance entre les groupes de données et cette frontière, appelée "marge" (voir la Figure 4.24) [150]. Plus de détails sur l'algorithme sont présentés dans l'annexe B.

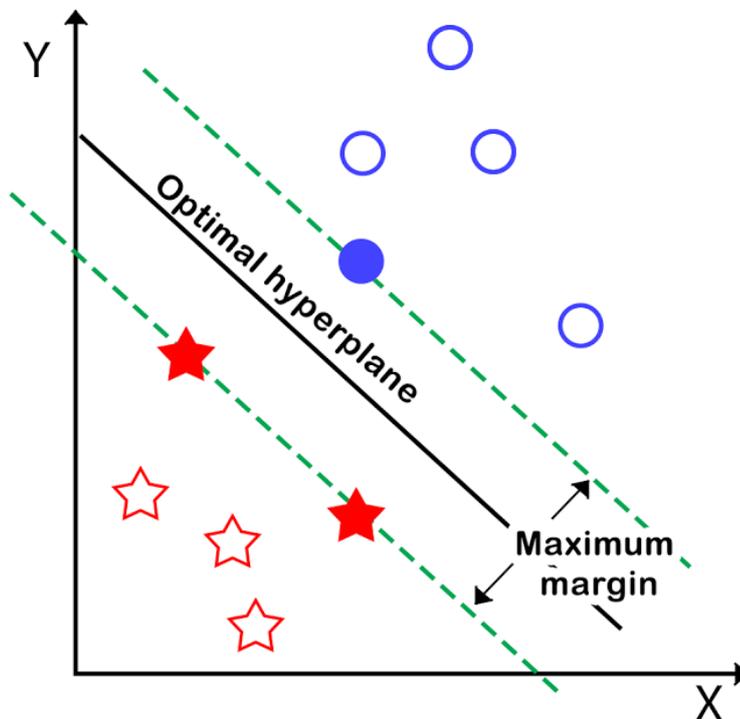


FIGURE 4.24 – Fonctionnement du SVM (analyticsarora.com)

Dans les machines à vecteurs de support (SVM), lorsque les données ne sont pas linéairement séparables, une idée clé est de transformer l'espace de représentation des données d'entrée en un espace de dimension supérieure (Figure 4.25), où il est

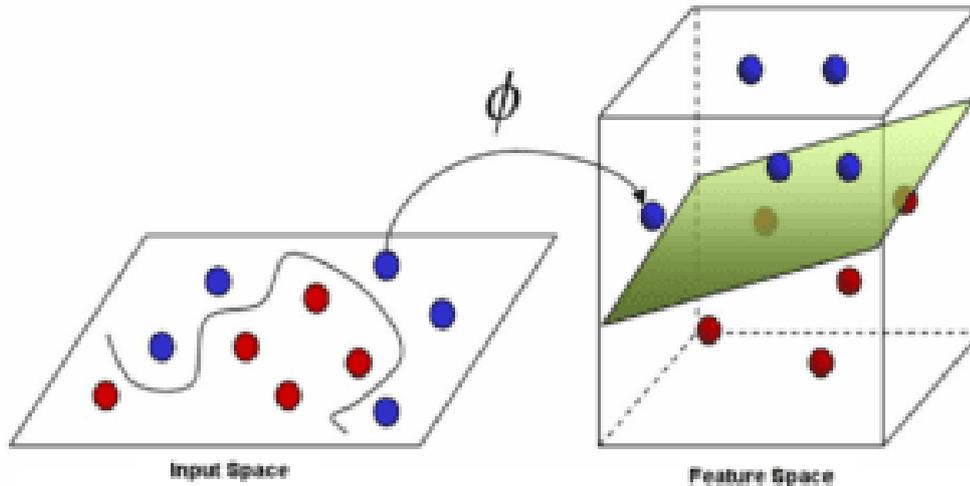


FIGURE 4.25 – (Crédit : © 2017, Haydar Ali Ismail, Medium.com) La projection des données dans une espace de plus grande dimension via une fonction noyau pour les séparer linéairement.

probable qu’une séparation entre les classes existe. Ceci est réalisé grâce à une fonction noyau (kernel function) afin de générer des hyperplans. Ce passage à des dimensions plus élevées (hyperplan lorsque les dimensions sont supérieures à 4) permet de traiter des données qui ne peuvent être linéairement séparées dans l’espace d’origine. En utilisant l’optimisation convexe (Annexe B) pour résoudre des problèmes d’optimisation [151]. L’application «*Classifier Learner*» ainsi les fonction «*fitcsvm*» de MATLAB effectue ce calcul.

### Les trois ensembles de données en apprentissage automatique

En apprentissage automatique, trois ensembles distincts sont utilisés : l’ensemble d’entraînement, l’ensemble de validation et l’ensemble de test (Figure 4.26). Ces ensembles sont employés respectivement pour entraîner, affiner et évaluer les performances d’un modèle [152].

**L’ensemble d’entraînement :** Constitue la base de données utilisée pour ajuster les paramètres du modèle tout au long du processus d’apprentissage. Il revêt une importance particulière dans ce contexte.

**L’ensemble de validation :** Représentent un sous-ensemble utilisé pour affiner les paramètres ou hyperparamètres du modèle. Elles permettent d’évaluer la qualité de l’apprentissage et de l’adaptation du modèle, facilitant ainsi les ajustements et les optimisations nécessaires.

**L’ensemble de test :** Est une base de données distinct permettant une évaluation impartiale de l’adéquation d’un modèle. Il est utilisé pour l’évaluation finale des

performances du modèle après son entraînement et son ajustement.

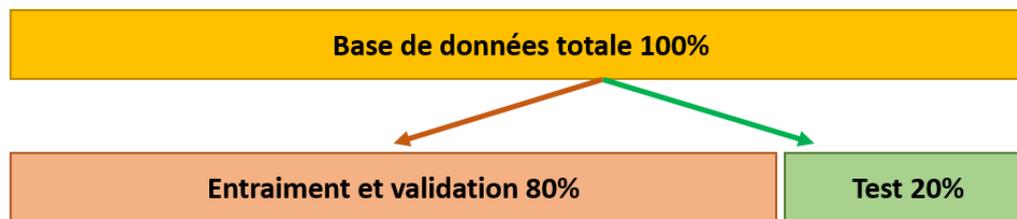


FIGURE 4.26 – Division des données en ensembles d'entraînement, de validation et de test

La principale distinction entre les données de validation et de test réside dans le fait que les données de validation servent à apporter des ajustements au modèle, tandis que les données de test sont utilisées pour fournir une évaluation finale des performances du modèle. Il est crucial que les ensembles de validation et de test suivent la même distribution de probabilités que l'ensemble d'apprentissage afin d'éviter le surajustement (overfitting) [153]. Pour notre étude, 80 % des données sont utilisées pour l'entraînement et la validation en appliquant le principe de la validation croisée décrit ultérieurement, tandis que 20 % des données sont réservées pour les tests.

#### 4.4 Entraîner et valider le modèle d'apprentissage

Matlab et Python (Jupyter Notebook) sont utilisées pour entraîner, paramétrer, évaluer et valider les modèles d'apprentissage automatique pour notre problème de classification binaire LOS/NLOS.

La validation croisée est une méthode indispensable pour évaluer précisément la performance d'un modèle, prévenir le surapprentissage (overfitting) et comparer les performances de différents modèles. Dans notre cas, la validation croisée de type  $k_{fold} = 10$  a été choisie. L'ensemble de données est divisé en  $k_{folds}$  sous-ensembles de taille égale. Le modèle est entraîné sur  $k-1$  sous-ensembles et testé sur le dernier sous-ensemble. Cette procédure est répétée  $k$  fois, chaque sous-ensemble étant utilisé une fois comme ensemble de test. Un fold est choisi pour la validation tandis que les autres sont utilisés pour l'entraînement du modèle, comme le montre la Figure 4.27 pour un  $k_{fold} = 10$ .

MathWorks a développé une application avec une interface interactive qui facilite la visualisation des caractéristiques (voir Figure 4.28) et le travail avec les données,

depuis le chargement des données jusqu'à l'optimisation et la validation du modèle.

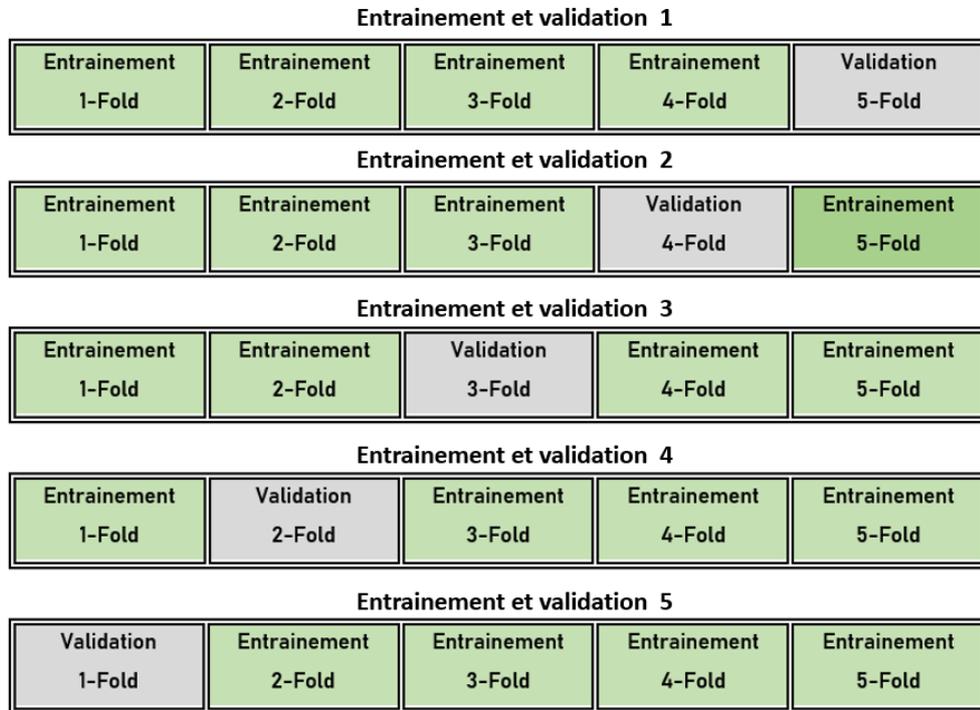


FIGURE 4.27 – Validation croisée pour un  $k_{fold} = 5$

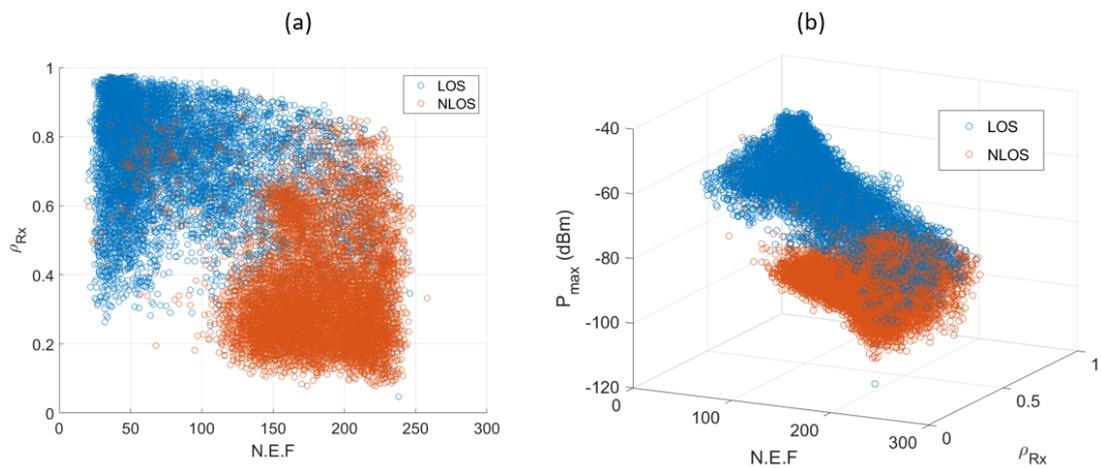


FIGURE 4.28 – Visualisation des caractéristiques

### Résultats kNN et SVM avec toutes les caractéristiques

Les deux modèles kNN et SVM ont été entraînés en utilisant l'ensemble complet des caractéristiques que nous avons identifiées précédemment (soit 23 caractéristiques au total). Cette étape initiale est cruciale pour établir une base de comparaison avant et après la sélection des caractéristiques, ainsi que l'optimisation des hyperparamètres. L'utilisation de l'ensemble complet des caractéristiques permet une évaluation globale des performances des modèles avant toute modification. Cela nous offre une référence pour mesurer l'impact de la sélection ultérieure de caractéristiques et de l'optimisation des hyperparamètres sur les résultats.

Les résultats de précision des deux algorithmes sur l'ensemble de validation et de test sont présentés dans le Tableau 4.4. Initialement, avec 23 caractéristiques, les modèles avaient une précision élevée sur l'ensemble de validation, soit 97,5% pour kNN et 97% pour SVM, respectivement. Sur l'ensemble de test, la précision était légèrement inférieure, atteignant respectivement 96,8% et 96,7% pour kNN et SVM.

	Précision du kNN	Précision du SVM
Données de Validation	97,2%	96,5%
Données de Test	96,8%	96,7%

TABLEAU 4.4 – Performances de kNN et SVM sur les ensembles de validation et de test avec toutes les caractéristiques.

La matrice de confusion (voir Figure 4.29) est un outil essentiel dans l'évaluation des performances d'un modèle de classification. Elle est divisée en quatre parties, représentant les résultats suivants :

- *TP* (Vrais Positifs) représente les échantillons LOS correctement prédites.
- *TN* (Vrais Négatifs) représente les échantillons NLOS correctement prédites.
- *FN* (Faux Négatifs) représente les échantillons LOS prédites incorrectement.
- *FP* (Faux Positifs) représente les échantillons NLOS prédites incorrectement.

En examinant la matrice de confusion dans la Figure 4.30, on peut observer les performances du modèle par classe. On constate que la classe LOS est légèrement moins bien classée que la classe NLOS pour les deux algorithmes, à la fois sur les ensembles de validation et de test. En effet, le taux de faux négatifs (FN) est légèrement plus élevé. Cela signifie que l'algorithme est plus susceptible de classer incorrectement un échantillon LOS comme NLOS que l'inverse.

Classe réelle	LOS	TP	FN
	NLOS	FP	TN
		LOS	NLOS
		Classe prédite	

FIGURE 4.29 – Matrice de confusion

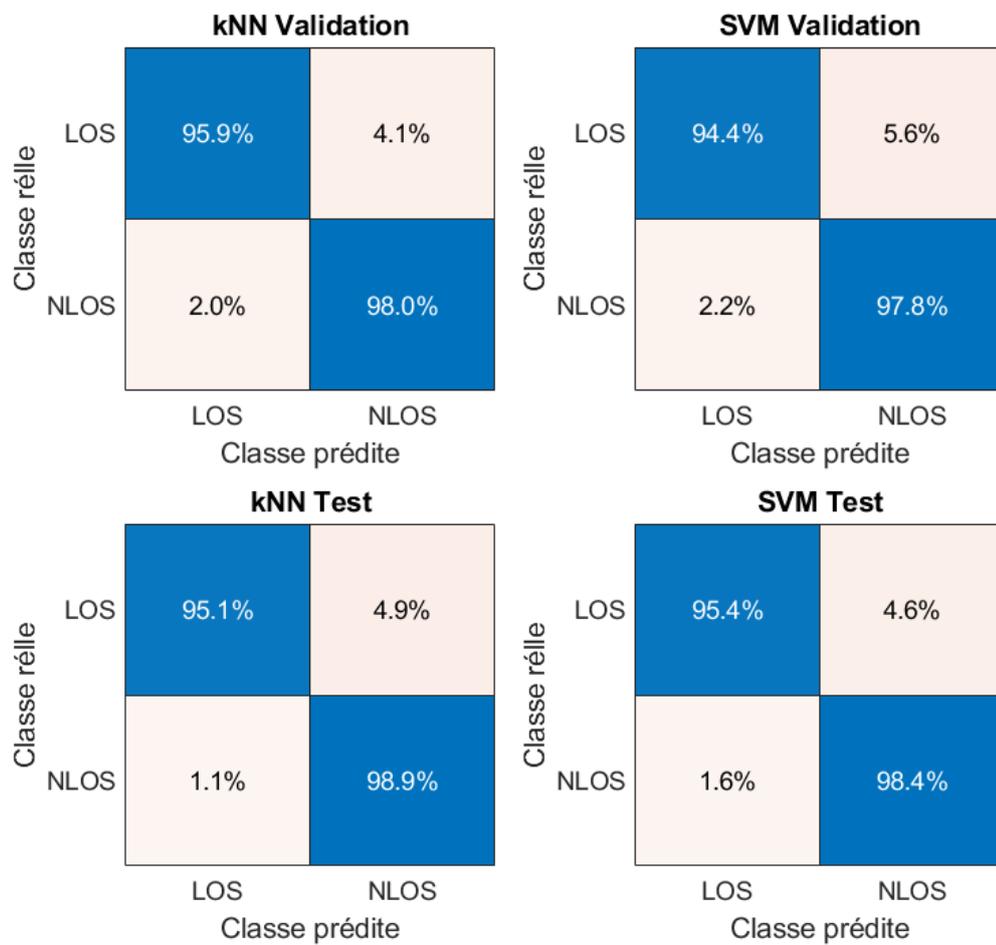


FIGURE 4.30 – Matrice de confusion pour les données de validation et de test

Bien que les modèles SVM et kNN aient produit de bons résultats, il est crucial de considérer la réduction de la complexité par une sélection adéquate des caractéristiques. L'élimination des caractéristiques peu informatives ou redondantes améliore non seulement l'interprétabilité du modèle mais aussi réduit les temps de calcul, un avantage notable dans les contextes de communications véhiculaires. Pour poursuivre l'amélioration de ces performances, nous envisageons la sélection des caractéristiques (voir Tableau 4.6) et l'optimisation des hyperparamètres des modèles (voir Tableau 4.5).

Algorithme	Hyperparamètre	Description
kNN	$k$	Nombre de voisins à considérer lors de la classification.
kNN	Distance	Métrique de distance utilisée pour mesurer la proximité entre les points.
SVM	C (BoxConstraint)	Paramètre de régularisation qui contrôle le compromis entre la classification correcte des exemples d'entraînement et la maximisation de la marge.
SVM	Fonction de noyau (KernelFunction)	Est utilisée pour transformer l'espace des caractéristiques, par exemple, linéaire, polynomial, gaussien (RBF).
SVM	Échelle de noyau (KernelScale)	Est un paramètre de mise à l'échelle utilisé pour normaliser les données avant l'évaluation de la matrice de Gram appropriée.

TABLEAU 4.5 – Hyperparamètres pour la classification binaire avec kNN et SVM.

Méthode	Description	Avantages	Inconvénients
Filtrage	Les caractéristiques sont sélectionnées en fonction de mesures statistiques telles que le seuil de variance et le t-Test (bibliothèque SelectKBest en Python) pour évaluer l'importance de chaque caractéristique ou groupe de caractéristiques.	Rapide et simple à mettre en œuvre.	Peut négliger les interactions entre les caractéristiques.
Enveloppe	Un modèle de classification est utilisé pour évaluer les caractéristiques une par une, par exemple, Recursive Feature Elimination (RFE) et la sélection ascendante des caractéristiques. Les caractéristiques qui contribuent le plus à la performance du modèle sont sélectionnées.	Peut prendre en compte les interactions entre les caractéristiques.	Peut être plus lent et complexe que le filtrage. Non-garantie de convergence. Risque de surajustement. Coût computationnel élevé
Hybride	Combine les deux approches précédentes et exploite l'expertise dans la caractérisation d'un canal de propagation.	Combine les avantages des deux approches précédentes en ayant une vision d'expertise pour d'autres caractéristiques afin de trouver la meilleure combinaison.	Cela nécessite une expertise dans le domaine combinée avec les méthodes conventionnelles, ce qui pourrait être complexe à mettre en œuvre. Nombre d'itération important

TABLEAU 4.6 – Méthodes de sélection de caractéristiques

### Sélection de caractéristiques par filtrage

Après la normalisation de toutes les caractéristiques estimées pour la classification LOS/NLOS, la comparaison de leurs variances revêt une importance particulière. Deux techniques ont été utilisées à cet effet : (i) le seuil de variance et (ii) le t-Test.

#### — Seuil de variance

La Figure 4.31.(a) présente la variance proportionnelle des caractéristiques, ordonnées de la plus forte à la plus faible. La Figure 4.31.(b) présente la précision des algorithmes sur les données de validation et de test en accumulant chaque fois une caractéristique.

Un seuil de 6% a été choisi après un processus itératif commençant par la première caractéristique ayant la plus grande proportion de variance, et en ajoutant une caractéristique à chaque étape jusqu'au 23ème. Ensuite, le seuil est sélectionné car les modèles montrent une valeur de précision élevée avec un minimum de caractéristiques pour les deux modèles kNN et SVM, comme le montre la Figure 4.31.(b).

Un seuil de > 6% a été choisi, représentant les cinq premières caractéristiques (voir Tableau 4.7), où les performances des algorithmes ne varient pas trop après l'ajout des caractéristiques supplémentaires.

On constate que la caractéristique angulaire AoDa présente de faibles performances avec une précision de classification de 40% pour kNN et 60% pour SVM. En ajoutant le nombre d'évanouissements fréquentiels (N.E.F), les performances augmentent à 91% 93%, respectivement pour kNN et SVM. Cela démontre l'efficacité de la caractéristique N.E.F pour optimiser les deux modèles de classification.

Caractéristiques	> 6% sans redondance
AoDa	✓
N.E.F	✓
$\rho_{TxH}$	✓
$\rho_{Rx}$	✓
Kurtosis	✓
kNN Validation	97,3%
kNN Test	96,5%
SVM Validation	96,5%
SVM Test	96,4%

TABLEAU 4.7 – Caractéristiques sélectionnées via la technique du seuil de variance avec les résultats de kNN et SVM.

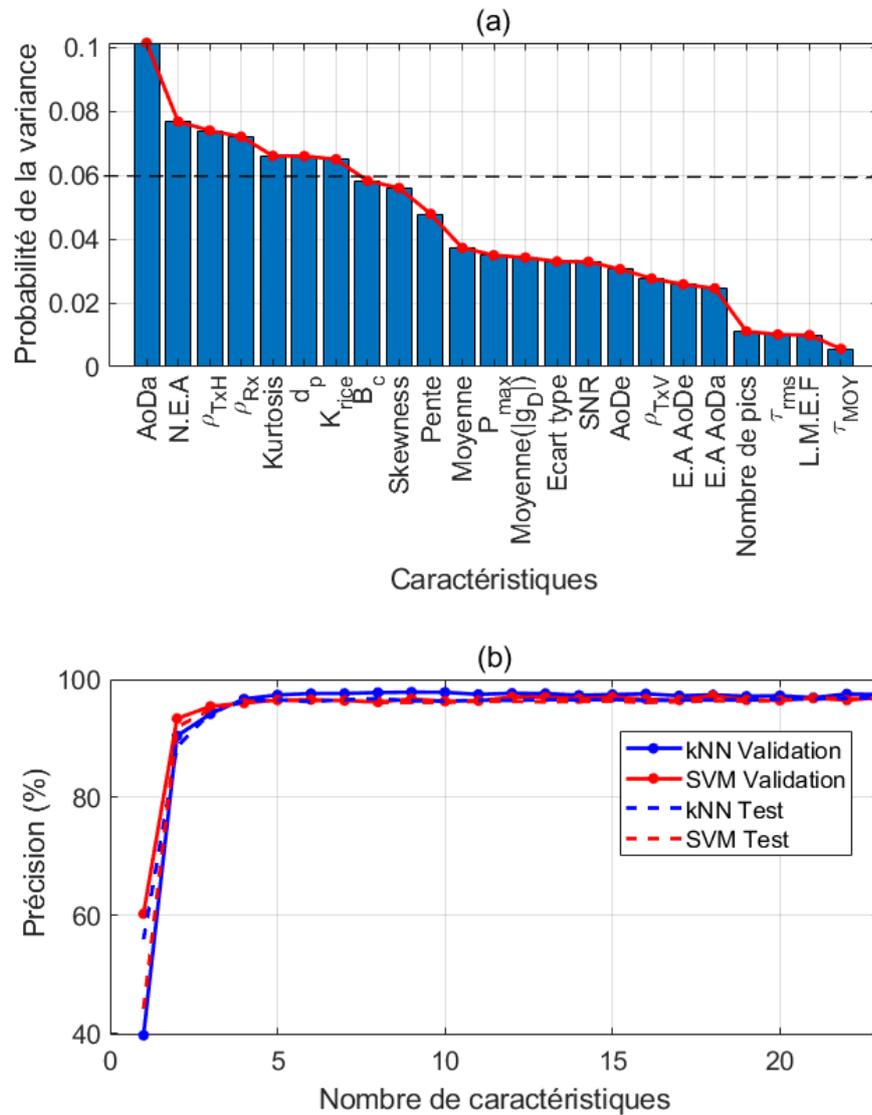


FIGURE 4.31 – Variance proportionnelle des caractéristiques

L'avantage de cette approche réside dans sa rapidité de sélection, bien que son inconvénient réside dans son absence de prise en compte de la relation entre les caractéristiques. Cependant, il demeure une mesure pertinente pour évaluer la variance de chaque caractéristique.

#### — t-Test

La méthode du t-Test est utilisée pour la sélection des caractéristiques, appliquée à l'aide de la bibliothèque Python scikit-learn (voir l'Annexe A). Après des tests itératifs, 11 caractéristiques ont été sélectionnées pour kNN et 9 pour SVM, comme illustré dans la Figure 4.32. L'ordre des caractéristiques sélectionnées par le t-Test

après ces tests itératifs est présenté dans le Tableau 4.8. De plus, on constate l'efficacité de la caractéristique "Nombre de pics" qui seule atteint une précision de classification supérieure à 93% pour les deux algorithmes.

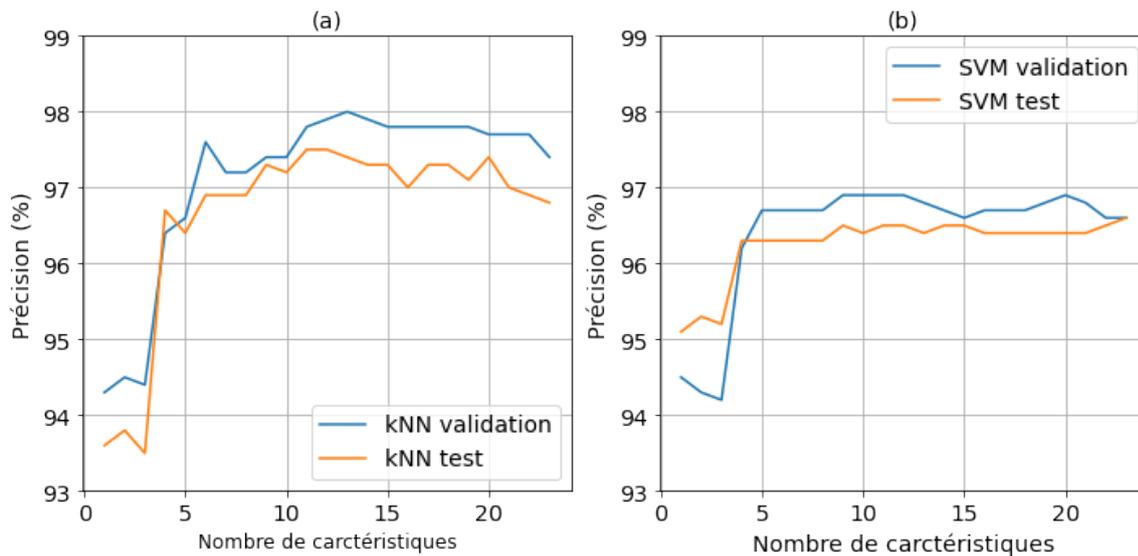


FIGURE 4.32 – Résultat de t-Test sur kNN (a) et SVM (b)

Caractéristiques	Top t-Test kNN	Top t-Test SVM
Nombre de pics	✓	-
$P_{max}$	✓	✓
N.E.F	✓	✓
L.M.E.F	✓	✓
Moyenne	✓	-
$K_{rice}$	✓	✓
Moyenne $ g_D $	✓	✓
SNR	✓	✓
$\rho_{Rx}$	✓	✓
$\rho_{TxH}$	✓	✓
$\rho_{TxV}$	✓	✓
kNN Validation	97,7%	-
kNN Test	97,5%	-
SVM Validation	-	96,9%
SVM Test	-	96,5%

TABEAU 4.8 – Caractéristiques sélectionnées via la technique du t-Test avec les résultats de kNN et SVM

### Sélection de caractéristiques par la méthode de l'enveloppe (Wrapper)

Dans cette section, la méthode de sélection séquentielle des caractéristiques a été exploitée pour affiner la pertinence des variables dans le cadre de la modélisation. Plus précisément, la technique de sélection séquentielle des caractéristiques vers l'avant a été employée, mettant en œuvre la fonction *sequentialfs* fournie par la Toolbox Statistics and Machine Learning de MATLAB.

#### — Sélection ascendante des caractéristiques

En utilisant les fonctions *fitcknn* et *fitcsvm* de la Toolbox Statistics and Machine Learning de MATLAB, les modèles kNN et SVM ont été créés. Une validation croisée avec une partition *k*-folds égale à 10 a été effectuée pour garantir une répartition équilibrée des classes. Ensuite, la fonction *sequentialfs* a été appliquée pour sélectionner les caractéristiques. Cependant, cette méthode peut être chronophage en termes de temps de calcul. Les résultats de cette sélection sont présentés dans le Tableau 4.9, montrant que pour kNN, 10 caractéristiques ont été sélectionnées avec une précision de prédiction de 97,5% et 97,3% respectivement pour les données de validation et de test, tandis que pour SVM, 7 caractéristiques ont été sélectionnées avec une précision de prédiction de 96,6% pour les données de test et de validation.

Caractéristiques	sequentialfs kNN	sequentialfs SVM
$\tau_{rms}$	✓	✓
$\tau_{MOY}$	✓	-
Nombre de pics	-	✓
$P_{max}$	✓	✓
L.M.E.F	✓	-
Moyenne	✓	-
SNR	✓	-
$\rho_{Rx}$	✓	✓
$\rho_{Tx_H}$	✓	✓
$\rho_{Tx_V}$	✓	-
AoDa	✓	-
Skewness	-	✓
Kurtosis	-	✓
kNN Validation	97,5%	-
kNN Test	97,3%	-
SVM Validation	-	96,6%
SVM Test	-	96,6%

TABLEAU 4.9 – Caractéristiques sélectionnées par la méthode de la sélection séquentielle pour kNN et SVM

### Sélection de caractéristiques par la méthode hybride

À travers plusieurs combinaisons de caractéristiques, en prenant en compte les résultats des deux méthodes précédentes, un ensemble de caractéristiques mettant en évidence la variabilité dans le domaine fréquentiel, les retards, ainsi que la variabilité de la réponse complexe avec les informations angulaires et de corrélation du réseau mMIMO, a été sélectionné. Ces caractéristiques ont montré sont présentées dans le Tableau 4.10.

Avec la méthode de sélection hybride, des précisions de prédiction supérieures sont obtenues, avec seulement 6 caractéristiques pour les deux approches. Notamment, pour kNN, une précision de 98% est atteinte sur les données de validation et de 97,5% sur les données de test. Cependant, la méthode hybride nous permet d'évaluer de manière approfondie la contribution des caractéristiques, en se basant sur l'expertise de la caractérisation d'un canal de propagation combinée avec des approches conventionnelles.

Caractéristiques	Hybride kNN	Hybride SVM
Moyenne( $ g_D $ )	-	✓
Nombre de pics	✓	-
N.E.F	✓	✓
$\tau_{rms}$	✓	✓
AoDa	✓	✓
$\rho_{Rx}$	✓	✓
$\rho_{TxH}$	✓	✓
kNN Validation	98%	-
kNN Test	97,5%	-
SVM Validation	-	96,6%
SVM Test	-	96,5%

TABLEAU 4.10 – Caractéristiques sélectionnées par la méthode hybride pour kNN et SVM

### Optimisation des hyperparamètres

Après validation de la sélection des caractéristiques obtenue par la méthode hybride, l'optimisation des hyperparamètres s'avère essentielle. En utilisant l'option *OptimizeHyperparameters* réglée sur *auto*, les caractéristiques ont été ajustées de manière itérative via les fonctions *fitcknn* et *fitsvm* pour les algorithmes SVM et kNN. Cette stratégie a permis de modifier les hyperparamètres, notamment en optimisant la métrique de distance et le nombre de voisins  $k$  pour le kNN, ainsi que le choix du noyau et la valeur du paramètre de régularisation  $C$  pour le SVM.

Pour le modèle kNN, et après plus que 1000 itérations par fonction d'évaluation de comparaison et d'optimisation (illustrées dans la Figure 4.33 (a)), l'utilisation de la métrique de distance de *cityblocks* avec un nombre de voisins fixé à 3 fournit les résultats les plus prometteurs (Modèle minimum réalisable). Cette observation est confirmée par la Figure 4.33.(b), qui représente "Min Observed Objective" (en bleu), la meilleure performance observée du modèle à chaque étape de l'optimisation des hyperparamètres, et la courbe "Estimated Min Objective" (en vert), qui indique l'estimation de la performance optimale que le modèle prévoit de pouvoir atteindre. La convergence des objectifs observés et estimés indique une optimisation réussie (les valeurs présentées dans le Tableau 4.11).

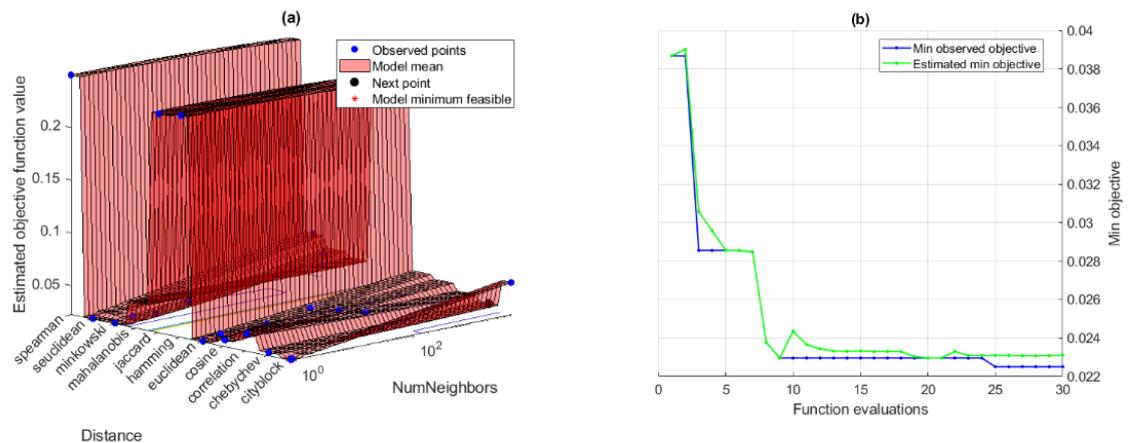


FIGURE 4.33 – Optimisation des hyperparamètres pour kNN

De même pour l'algorithme SVM, on observe dans la Figure 4.34.(a) et 4.34.(b) que le facteur  $C$  de 0,003 et un "KernelScale" de 0,08 présentent des performances optimales du modèle après 30 itérations, avec des valeurs optimales de la meilleure estimation observée et estimée présentées dans le Tableau 4.11. Tous les hyperparamètres optimisés pour kNN et SVM sont présentés dans le Tableau 4.12.

Algorithme	Valeur Objectif Observée	Valeur Objectif Estimée
kNN	0,022	0,023
SVM	0,032	0,033

TABLEAU 4.11 – Comparaison des valeurs de la fonction objectif observées et estimées pour les algorithmes kNN et SVM

Après l'entraînement des modèles kNN et SVM avec les nouvelles caractéristiques et les nouveaux hyperparamètres, une optimisation a été réalisée dans les deux cas. Pour kNN et SVM seulement 6 caractéristiques hybrides ont été retenues, avec une précision respectivement de 98,1% et 96,7% sur les données de validation et de 97,7% et 96,7% sur les données de test. Les détails sont présentés dans le Tableau 4.13.

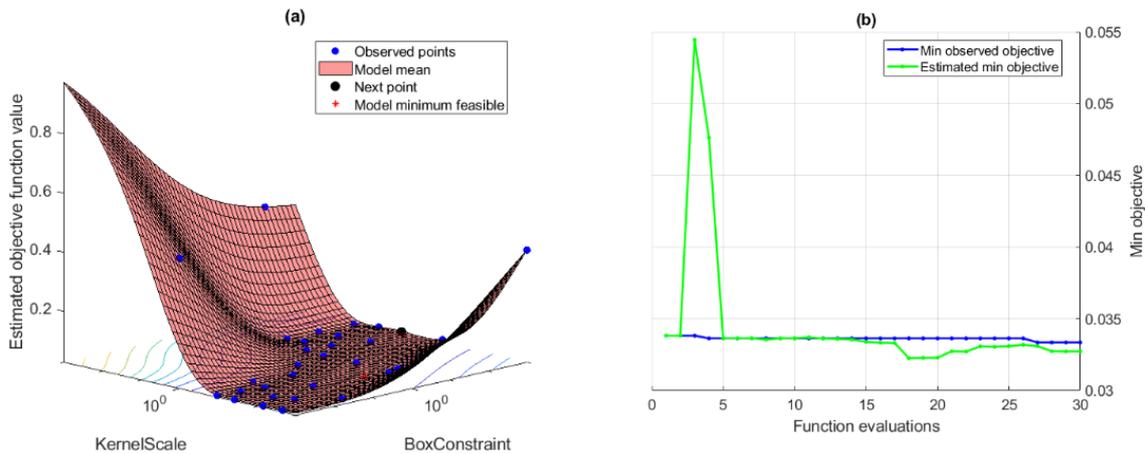


FIGURE 4.34 – Optimisation des hyperparamètres pour SVM

Algorithme	Hyperparamètre	Valeur
kNN	$k$	3
	Distance	minkowski
	Distance Weight	Inverse
SVM	$C$ (BoxConstraint)	0,0034
	Kernel Function	Polynomiale de degré 2
	Kernel Scale	0,08

TABLEAU 4.12 – Hyperparamètres pour la classification binaire avec kNN et SVM.

Les matrices de confusion (Figure 4.35) montrent que la réduction des caractéristiques de 23 à 6 avec la méthode hybride a considérablement amélioré les performances des modèles. Cette approche hybride, combinant l'expertise dans la caractérisation d'un canal de propagation avec des méthodes conventionnelles, permet de réduire la complexité du modèle tout en maintenant des performances comparables, avec seulement 6 caractéristiques au lieu des 23 initiales.

Pour kNN, une amélioration significative est observée dans les données de test, avec des taux de faux positifs (FP) et de faux négatifs (FN) réduits respectivement à 1% et 2,6%. Pour SVM, une légère amélioration sur les FN de 4,5%, mais une dégradation sur FP à 2% est acceptable compte tenu de la réduction des caractéristiques.

Les PDP des échantillons FP et FN ont été analysées : pour les FN, une diminution de la puissance du signal est souvent constatée dans les cas de LOS éloignés, les rapprochant de ceux de NLOS. En revanche, pour les FP, une puissance de signal élevée est observée en raison d'un trajet de forte puissance, les rendant similaires aux cas de LOS.

Caractéristiques	Hybride kNN	Hybride SVM
Moyenne( $ g_D $ )	-	✓
Nombre de pics	✓	-
N.E.F	✓	✓
$\tau_{rms}$	✓	✓
AoDa	✓	✓
$\rho_{Rx}$	✓	✓
$\rho_{TxH}$	✓	✓
kNN Validation	98,1%	-
kNN Test	97,7%	-
SVM Validation	-	96,7%
SVM Test	-	96,7%

TABLEAU 4.13 – Caractéristiques sélectionnées par la méthode hybride pour kNN et SVM après l'optimisation des hyperparamètres

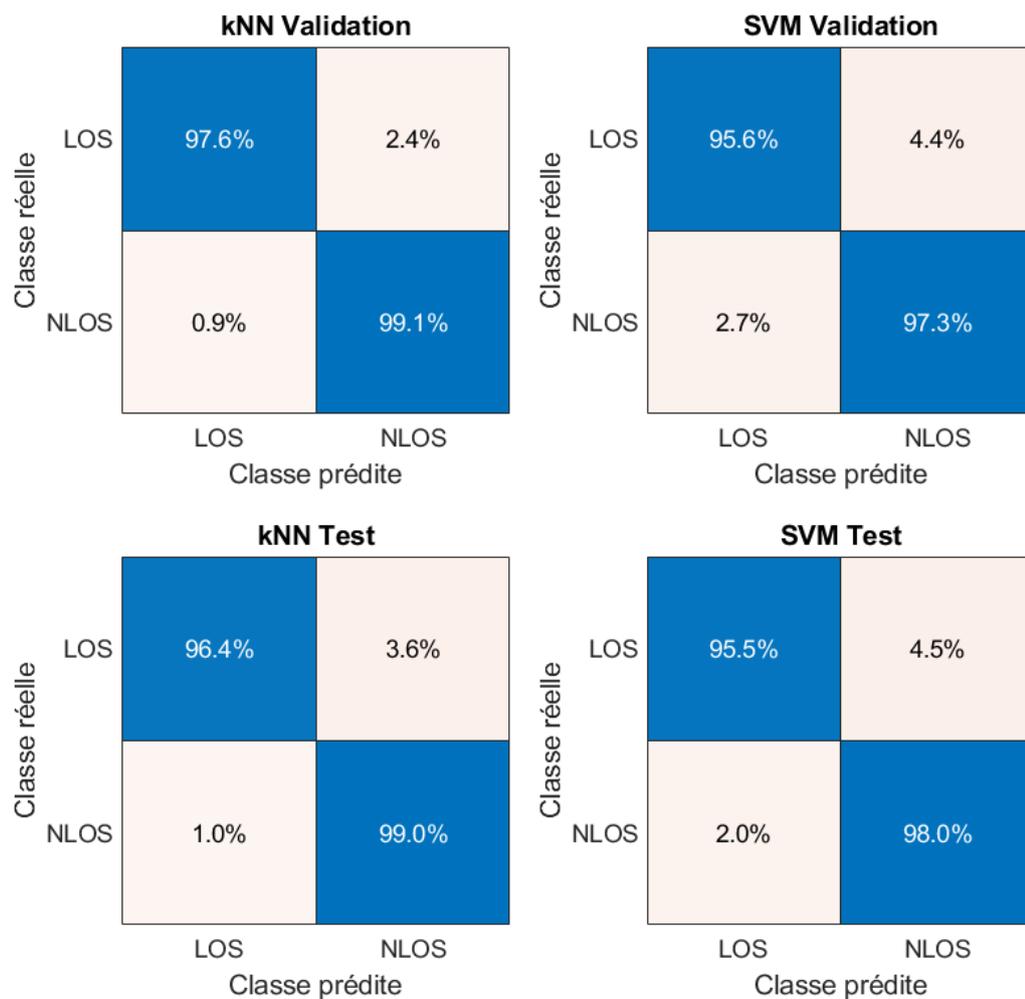


FIGURE 4.35 – Matrice de confusion pour les données de test

Enfin, le kNN se démarque comme le meilleur modèle pour la classification des scénarios LOS et NLOS, ne nécessitant que 6 caractéristiques pour atteindre une précision de 97,7% sur de nouvelles données de test. De plus, les performances de nos deux modèles kNN et SVM dépassent la majorité des performances présentées dans la littérature, comme indiqué dans le Tableau 4.14.

Bien que la comparaison ne soit pas basée sur des études menées dans des environnements, scénarios ou technologies identiques, à l'exception de l'article de référence [97], elle offre néanmoins une évaluation générale de l'intégration du ML dans la classification des scénarios de propagation, indépendamment du contexte spécifique.

		Littérature		
Notre approche	Réf.	Scénario	Base de données	Algo. et précision
	[95]	Indoor	Expérimentale	SVM, GP 95% statique, 87% dynamique
	[94]	Environnements ferroviaires (HSR)	Expérimentale	(kNN, SVM >90%)
	[94]	Environnements ferroviaires (HSR)	Expérimentale et Simulation	(k-moyennes 90%) (GMM > 89%)
V2I Environnement suburbain	[97]	Outdoor V2V MIMO	Expérimentale	(RF 94%) (SVM 90%) (ANIN 92%)
Canal de propagation mMIMO	[100]	Outdoor urbain GNSS	Expérimentale	(NIN 97,7%) (SVM 95%)
Base de données 33000	[102]	5G Indoor aéroport	Expérimentale	(XGBoost 98%) (SVM 92%)
SVM : 96,7%	[102]	5G Indoor aéroport	Simulation	(XGBoost 100%) (SVM 92%)
kNN : 97,7%	[98]	Indoor	Expérimentale	RF 90%
	[99]	Indoor	Expérimentale	(RF 96,34%) (NB 88,56%) (SVM 95,12%)
	[96]	Indoor	Expérimentale	ANIN 95%
	[106]	Indoor	Expérimentale	CNN 93,3%
	[105]	Indoor	Expérimentale	CNN 87,4%
	[105]	Indoor	Expérimentale	(MPL 82,9%) (SVM 82,6)
	[103]	Indoor	Expérimentale	K-moyennes 95%
	[101]	Indoor	Expérimentale	SVM 91%
	[104]	Indoor	Expérimentale	CNN+LSTM stacked 82,14%

TABLEAU 4.14 – Comparaison des résultats avec la littérature.

## 4.5 Conclusion

Ce chapitre présente une étude sur la classification des scénarios de propagation LOS et NLOS d'un canal mMIMO appliquée aux communications véhiculaires V2I à l'aide du ML. Cette étude est une première dans la littérature. Les données collectées dans divers environnements suburbains sur le campus de l'Université de Lille ont été utilisées pour l'entraînement, la validation et le test des modèles de ML. Un prétraitement des données a été réalisé en amont pour garantir la robustesse et la fiabilité du modèle. Au total, 23 caractéristiques ont été calculées à partir de la réponse complexe du canal  $H(f)$ , incluant de nouvelles caractéristiques telles que les Nombres des Évanouissements Fréquentiels (N.E.F), le retard moyen du groupe ( $\text{Moyenne}(|g_D|)$ ), ainsi que les caractéristiques MIMO et massive MIMO telles que la corrélation entre les antennes d'émission ( $\rho_{Tx_H}$  et  $\rho_{Tx_V}$ ) et de réception ( $\rho_{Rx}$ ). Deux algorithmes, kNN et SVM, ont été employés. Une approche hybride de sélection de caractéristiques a été adoptée, combinant l'expertise du domaine avec plusieurs techniques de sélection de caractéristiques conventionnelles.

Après optimisation des hyperparamètres, les résultats montrent que cette approche hybride atteint une précision de 97,7% et 96,7% respectivement pour kNN et SVM, en utilisant seulement 6 caractéristiques : Nombre de pics, N.E.F,  $\tau_{rms}$ , AoDa,  $\rho_{Rx}$ ,  $\rho_{Tx_H}$  pour kNN et  $\text{Moyenne}(|g_D|)$  au lieu de Nombre de pics pour SVM. Ces résultats démontrent l'efficacité des nouvelles caractéristiques ainsi que les propriétés mMIMO. De plus, notre modèle optimal utilise un nombre minimal de caractéristiques, surpassant la majorité des modèles de la littérature et développé dans des conditions réalistes et diversifiées, avec une interopérabilité supérieure.

Les nouvelles caractéristiques extraites, telles que les N.E.F, le nombre de pics et la Moyenne du Retard de Groupe, ainsi que les caractéristiques mMIMO, se révèlent être de bonnes caractéristiques pour nos modèles. L'adoption de la méthode de sélection de caractéristiques hybride nous a permis de déterminer le modèle optimal pour notre problème de classification des scénarios de propagation LOS-NLOS.



# Conclusion générale et perspectives

Ce travail s'est articulé autour de la caractérisation expérimentale multidimensionnelle d'un canal de propagation MIMO ainsi que, de la classification des scénarios de communication en visibilité directe (LOS) et non visibilité directe (NLOS) à l'aide de l'intelligence artificielle (IA) par le biais de l'apprentissage automatique (ML) dédié aux communications véhiculaires. Cette étude était la première à porter sur l'analyse temporelle et spatiale du temps de stationnarité  $T_s$  d'un canal de propagation massif MIMO pour chaque liaison  $k$  d'émission et  $l$  de réception. Ainsi, sur la dérivation de nouvelles caractéristiques et l'adoption d'une méthode de sélection de caractéristiques hybride adaptée au domaine, pour la classification des scénarios de propagation LOS et NLOS.

Dans le premier chapitre, un état de l'art sur la caractérisation multidimensionnelle d'un canal de propagation non-WSSUS de type SISO, SU-MIMO et MU-MIMO a été exposé. Ainsi, les méthodes d'estimation du temps de stationnarité  $T_s$  basé sur différentes méthodes ont été abordées. Une évaluation des paramètres du canal large bande à partir des fonctions généralisées et des réponses impulsionnelles ont été examinées. Puis, pour une deuxième partie, un état de l'art sur les apports de l'IA dans l'estimation du canal et l'allocation de ressources ont été décrits. Ces apports étudiés se déroulant en deux points essentiels : (i) La classification des scénarios de propagation LOS et NLOS à l'aide de l'apprentissage automatique et l'apprentissage profond (ii) L'apport de l'IA sur l'estimation et la prédiction d'un canal de propagation.

Dans le deuxième chapitre, l'architecture et les réseaux d'antennes du sondeur de canal MaMIMOSA, utilisée pour la création de la base de données expérimentale ont été présentés. Les différentes campagnes de mesure réalisées pour cette étude (en tunnel et en milieu suburbain) sont également décrites. Une analyse en bande étroite des différentes configurations antennaires et polarisations est menée, portant sur des aspects tels que la qualité du signal (SNR), l'affaiblissement de propagation qui a été comparé à un modèle théorique, la qualité et la discrimination de la polarisation croisée (XPD). Cette analyse a conforté la qualité des matrices de canal obtenues lors des campagnes de mesure, nous permettant ainsi d'approfondir l'analyse et la caractérisation multidimensionnelle d'un canal de propagation ainsi que l'identification des scénarios de propagation LOS et NLOS pour les communications

véhiculaires de type V2I.

Par la suite, une caractérisation de la variabilité temporelle et spatiale d'un canal de propagation MIMO a été réalisée dans un environnement suburbain et milieu tunnel, détaillée dans le chapitre 3 en se basant sur les mesures obtenues dans le chapitre 2. Les paramètres de configuration, la discrétisation de la fonction de transfert en temps et en fréquence, ainsi que les calculs du GSLF discret, et les méthodes utilisées telles que la CCF et la colinéarité pour estimer le temps de stationnarité  $T_s$ , sont présentés. Dans un environnement suburbain pour un canal de propagation MIMO massif de 64Tx réseau carré de  $8 \times 8$  et 1Rx, le temps de stationnarité  $T_s$  a été estimé pour chaque liaison  $(k, l)$  de réception en utilisant les méthodes CCF et la colinéarité. La valeur médiane moyenne de  $T_s$  était de 568 ms et 565 ms, correspondant respectivement à une distance de stationnarité  $D_s$  de **6,31 m** et **6,27 m**, par la méthode CCF et la colinéarité, respectivement, suggérant que ces deux méthodes se comportent de manière similaire en moyenne. Par ailleurs, l'analyse spatiale du  $T_s$  sur les 64 antennes par la méthode CCF montre qu'il est fortement corrélé entre les antennes alignées verticalement (ULA-V) par rapport à celles alignées horizontalement (ULA-H), avec des valeurs moyennes de coefficients de corrélation de Pearson respectivement de  $\rho_V=0,95$  et  $\rho_H=0,57$ . Approximativement les mêmes coefficients de corrélation sont trouvés entre les fonctions de transfert. Cela confirme que les caractéristiques du canal radio sont similaires en ULA-V qu'en ULA-H, où les diffuseurs dans le plan d'élévation sont rares, contrairement au plan d'azimut. De plus, la corrélation spatiale des fonctions de transfert entre les antennes peut entraîner une corrélation du temps de stationnarité  $T_s$ . D'autre part, dans un scénario en tunnel et un canal de propagation MIMO avec deux fréquences de 1,35 GHz et 5,89 GHz, un temps de stationnarité moyen  $T_s$  a été respectivement estimé à 406 ms (**10,15m**) et 332 ms (**8,3m**). Ces valeurs sont approximativement similaires aux valeurs trouvées pour un modèle théorique pour le même scénario dans un tunnel vide. Les caractéristiques du canal radio large bande ont également été estimées, telles que l'étalement des retards, l'étalement de Doppler, le temps de cohérence  $T_c$ , la bande de cohérence  $B_c$  et les directions de départ des rayons du réseau d'émission et leur étalement angulaire. De cette étude, les points les plus importants extraits sont :

- Le temps de stationnarité  $T_s$  est influencé par l'environnement et la fréquence. Le canal est plus stationnaire en tunnel  $D_s=8,3$  m que en environnement suburbain  $D_s=6,3$  m à 5,89 GHz. En tunnel le canal est plus stationnaire à 1,35 GHz  $D_s=10,15$  m que 5,89 GHz  $D_s=8,3$  m.
- Pour un réseau massive MIMO, le temps de stationnarité  $T_s$  est fortement corrélé entre les antennes alignées verticalement ( $\rho_V=0,95$ ) par rapport à celles alignées horizontalement ( $\rho_H=0,57$ ).

- En tunnel, l'état du trafic n'a pas une influence significative sur les valeurs du temps de stationnarité  $T_s$ . De plus, la corrélation du temps de stationnarité pour les polarisations à +45 et -45 respectivement DD et AA est la plus élevée, avec 0,83 pour un trafic faible et 0,91 pour un trafic important.
- Les caractéristiques du canal large bande montrent une cohérence avec ce qui est représenté dans les profils de puissance de retard et de Doppler. De plus, une corrélation entre le comportement de l'étalement des retards et l'étalement Doppler est observée, avec un coefficient de corrélation de Pearson de 0,72, ainsi que le temps de cohérence et la bande de cohérence.

Dans le dernier chapitre, l'apport de l'IA dans la couche physique des réseaux 5G a été étudié pour une application qui concerne la classification des scénarios de propagation LOS et NLOS à l'aide du ML. L'originalité de ce travail repose sur l'extraction et l'utilisation de nouvelles caractéristiques, ainsi que sur l'adoption d'une nouvelle méthode hybride de sélection spécifiquement dédiée au domaine, qui s'est avérée plus efficace que les méthodes traditionnelles en termes de sélection des caractéristiques. La littérature sur ce sujet traite des environnements indoor pour les technologies UWB et Wifi. La détection des scénarios LOS/NLOS pour communications véhiculaires de type V2I, est notre connaissance à l'état de l'art. Deux algorithmes ont été utilisés : SVM et kNN, avec un choix final de 6 caractéristiques qui ont été obtenues grâce à notre méthode hybride qui a assuré la contribution de chaque dimension du canal à la classification LOS-NLOS, c'est-à-dire le retard, la fréquence, la phase et l'espace. Les caractéristiques choisies sont les suivantes :

- Pour kNN : { N.E.F,  $\rho_{TxH}$ ,  $\rho_{Rx}$ ,  $\tau_{rms}$ , AoD azimut, Nombre de pics }
- Pour SVM : { N.E.F,  $\rho_{TxH}$ ,  $\rho_{Rx}$ ,  $\tau_{rms}$ , AoD azimut, Moyenne( $|g_D|$ ) }

Après l'optimisation des hyperparamètres, une précision de classification de 97,7% et 96,7% respectivement pour kNN et SVM a été obtenue sur les données de test. Quatre caractéristiques sont utilisées pour la première fois dans notre étude : le nombre de pics par PDP, le nombre d'évanouissements fréquentiels (N.E.F), la moyenne du retard de groupe (Moyenne( $|g_D|$ )) et la corrélation entre les antennes d'émission ( $\rho_{TxH}$ ) et de réception ( $\rho_{Rx}$ ). De plus, cette étude met en évidence **l'importance des caractéristiques mMIMO, qui représentent 50% des caractéristiques choisies**, comprenant les angles de départ en azimut (AoDa) du trajet à forte amplitude, ainsi que les corrélations entre les antennes à l'émission et à la réception.

Les résultats de cette thèse ouvrent la voie à de futures investigations, notamment en ce qui concerne une comparaison entre les données de mesure réelles et les

simulateurs déterministes tels que Veneris, développé par le Groupe d'Ingénierie des Réseaux de Télécommunications (GIRTEL) de l'Université polytechnique de Carthagène (UPCT), Espagne, ainsi que d'autres simulateurs existants, dans les configurations de réseau massive MIMO. Cette comparaison est une extension de notre papier [123], et vise à évaluer le comportement spatial statistique du temps de stationnarité  $T_s$  à travers le réseau massive et les propriétés angulaires des composants de trajet multiple (MPC), telles que les valeurs et l'étalement angulaire des angles d'arrivée (AoA) et des angles de départ (AoD).

Il serait également intéressant d'appliquer la méthodologie aux résultats obtenus en tunnel pour prédire le canal de propagation vu du réseau massif d'antennes à partir d'une estimation parcimonieuse du canal et de la connaissance du type de scénario.

Dans le cadre de l'introduction des algorithmes d'apprentissage dans la couche PHY des réseaux, il serait intéressant d'utiliser les modèles de deep learning basés sur les séries temporelles de type LSTM. Les notions de temps de stationnarité et de temps de cohérence pourraient alors être introduites. Rappelons, que dans notre étude, les canaux présentant des  $SNR < 3$  dB ont été supprimés, ce qui introduit des ruptures dans les parcours utilisés pour alimenter la base de données. De plus, la détection des scénarios intermédiaires, OLOS (obstructed LOS) pourrait permettre d'améliorer les performances. Ces modèles devraient à terme prédire la remise à jour des allocations de beam, l'estimation du canal participant ainsi à l'optimisation de la stratégie d'allocation des ressources. Il serait également intéressant d'appliquer la méthodologie aux résultats obtenus en tunnel pour prédire le canal de propagation vu du réseau massif d'antennes à partir d'une estimation parcimonieuse du canal et de la connaissance du type de scénario.

## Annexe A

# Stationnarité d'un canal de propagation

## A.1 Stationnarité d'un canal SISO

### A.1.1 Description systémique d'un canal de propagation SISO

Pour caractériser de manière équivalente les canaux à large bande, on utilise un système linéaire variant dans le temps (LTV) qui dépend de deux variables : le temps et le retard. Cette dépendance est définie par la réponse impulsionnelle du canal passe-bas variant dans le temps, notée  $h(t, \tau)$  [154], et est associée à ses représentations dans les domaines du Doppler et des fréquences. Pour un canal de propagation SISO, Bello a introduit quatre fonctions qui représentent un canal de propagation avec différentes dimensions (Figure A.1) [68].

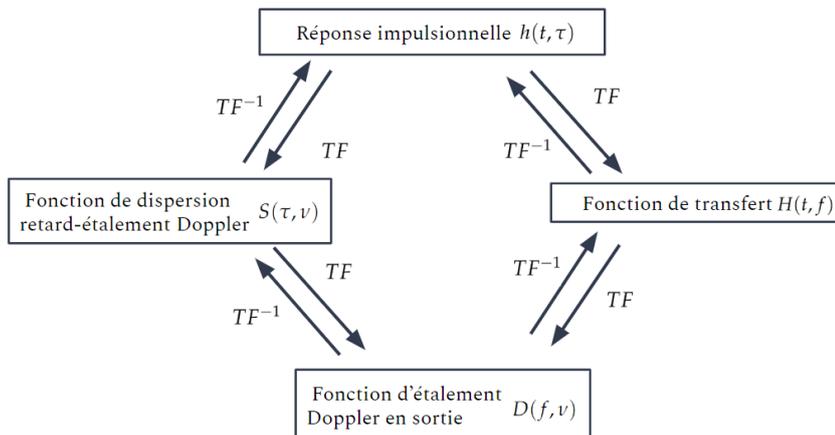


FIGURE A.1 – Illustration des fonctions de Bello

La réponse impulsionnelle du canal,  $h(t, \tau)$ , peut être représentée comme suit :

$$h(t, \tau) = \sum_{i=0}^{\infty} a_i(t) e^{j\phi_i(t)} \delta(t - \tau_i(t)) \quad (\text{A.1})$$

où  $a_i(t)$  est l'amplitude variant en fonction du temps, associée aux retards  $\tau_i$ , et  $\phi_i$  est la phase.

La description d'un canal de propagation variant dans le domaine fréquentiel au fil du temps peut être représentée par une fonction de transfert  $H(t, f)$ , où  $f$  est la fréquence dans une bande en Hertz. Cette fonction permet de suivre la description fréquentielle (la sélectivité en fréquence) et est obtenue par la transformation de Fourier par rapport aux retards  $\tau$  :

$$H(t, f) = \int_{-\infty}^{+\infty} h(t, \tau) e^{-j2\pi f\tau} d\tau \quad (\text{A.2})$$

De même,  $S(\nu, \tau)$  est la fonction d'étalement de dispersion Doppler qui décrit l'étalement d'un signal d'entrée dans le domaine des retards et de Doppler, avec  $\nu$  représentant le décalage Doppler en Hz et  $\tau$  le retard, obtenue par une transformation de Fourier de la réponse impulsionnelle dans le temps  $t$  :

$$S(\tau, \nu) = \int_{-\infty}^{+\infty} h(t, \tau) e^{-j2\pi\nu t} dt \quad (\text{A.3})$$

Enfin,  $D(\nu, f)$  est la fonction d'étalement Doppler obtenue comme suit :

$$D(f, \nu) = \int_{-\infty}^{+\infty} H(t, f) e^{-j2\pi\nu t} dt \quad (\text{A.4})$$

Cette formulation permet de caractériser de manière déterministe un canal de propagation linéaire variant dans le temps, en prenant en compte à la fois les variations temporelles et fréquentielles, ainsi que les retards et les effets Doppler.

### A.1.2 Description stochastique d'un canal de propagation SISO

En général, la réponse impulsionnelle d'un canal variant dans le temps est aléatoire en raison du comportement aléatoire des amplitudes, des phases et des retards. Dans ce cas, seule une approche stochastique permet de caractériser le canal, car une caractérisation déterministe complète du canal est irréalisable. Même si une telle caractérisation était possible, elle ne s'appliquerait qu'à un environnement spécifique, la rendant non interopérable [68].

Tout cela motive une caractérisation stochastique. Les fonctions d'autocorrélation sont reliées les unes aux autres par des transformées de Fourier doubles, comme le montre la Figure A.2.

$$R_h(t, t'; \tau, \tau') = E[h(t, \tau) \times h^*(t', \tau')] \quad (\text{A.5})$$

$$R_H(t, t'; f, f') = E[H(t, f) \times H^*(t', f')] \quad (\text{A.6})$$

$$R_s(\tau, \tau'; \nu, \nu') = E[S(\tau, \nu) \times S^*(\tau', \nu')] \quad (\text{A.7})$$

$$R_D(f, f'; \nu, \nu') = E[D(f, \nu) \times D^*(f', \nu')] \quad (\text{A.8})$$

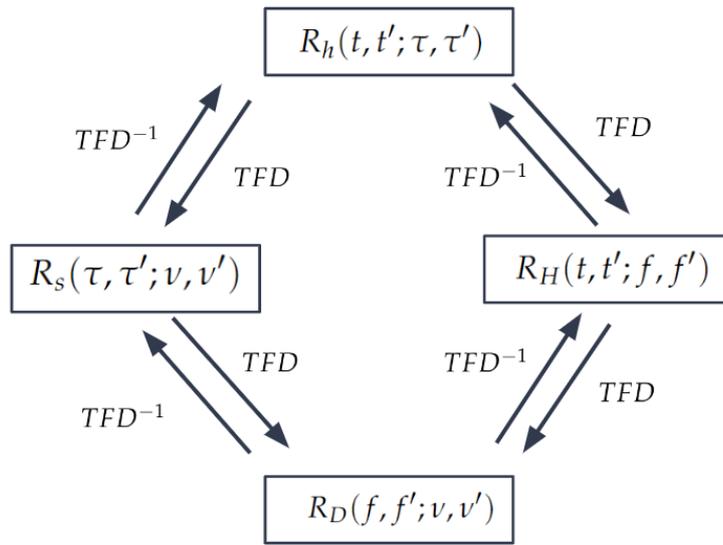


FIGURE A.2 – Illustration des fonctions d'autocorrélation

Ces fonctions de corrélation, dépendant de quatre variables, peuvent être simplifiées en introduisant des hypothèses de stationnarité du canal. Pour un canal stationnaire en temps, et donc stationnaire au sens large (WSS), cela signifie que les fonctions de corrélation ne varient pas sur un intervalle de temps  $\Delta t = t' - t$ . En parallèle, la Figure A.2 illustre les fonctions de corrélation. De plus, l'hypothèse d'un canal stationnaire en fréquence ou de diffuseurs non corrélés (US) suppose que les signaux arrivant avec différents retards ne sont pas corrélés. Par dualité en fréquence, cela signifie que les propriétés du canal de deuxième ordre en fréquence ne dépendent pas de la fréquence, mais seulement du changement d'écart  $\Delta f = f' - f$  [68].

Pour un canal présentant une stationnarité au sens large dans la variable temporelle (canal WSS), la description statistique sera indépendante du temps absolu. De plus, en appliquant la transformée de Fourier à la fonction d'autocorrélation d'un canal WSS, on obtient une décorrélation dans le domaine de Doppler. Par conséquent, les fonctions d'autocorrélation seront les suivantes :

$$R_h(t, t'; \tau, \tau')_{WSS} = R_h(\Delta t; \tau, \tau') \quad (\text{A.9})$$

$$R_H(t, t'; f, f')_{WSS} = R_H(\Delta t; \Delta f) \quad (\text{A.10})$$

$$R_s(\tau, \tau'; \nu, \nu')_{WSS} = C(\tau, \tau'; \nu) \delta(\nu' - \nu) \quad (\text{A.11})$$

$$R_D(f, f'; \nu, \nu')_{WSS} = P_D(f, f'; \nu) \delta(\nu' - \nu) \quad (\text{A.12})$$

Maintenant, pour un canal avec des diffuseurs non corrélés (US), autrement dit, un canal stationnaire en fréquence, les fonctions d'autocorrélation deviennent :

$$R_h(t, t'; \tau, \tau')_{US} = P_h(t, t'; \tau) \delta(\tau, \tau') \quad (\text{A.13})$$

$$R_H(t, t'; f, f')_{US} = R_H(t, t'; \Delta f) \quad (\text{A.14})$$

$$R_s(\tau, \tau'; \nu, \nu')_{US} = C(\tau; \nu, \nu') \delta(\tau, \tau') \quad (\text{A.15})$$

$$R_D(f, f'; \nu, \nu')_{US} = R_D(\Delta f; \nu, \nu') \quad (\text{A.16})$$

Finalemnt, la combinaison des hypothèses WSS et US définit un canal WSSUS, où les fonctions d'autocorrélation sont représentées comme suit :

$$R_h(t, t'; \tau, \tau')_{WSSUS} = P_h(\Delta t, \tau) \delta(\tau' - \tau) \quad (\text{A.17})$$

$$R_H(t, t'; f, f')_{WSSUS} = R_H(\Delta t, \Delta f) \quad (\text{A.18})$$

$$R_s(\tau, \tau'; \nu, \nu')_{WSSUS} = C(\tau, \nu) \delta(\tau' - \tau) \delta(\nu' - \nu) \quad (\text{A.19})$$

$$R_D(f, f'; \nu, \nu')_{WSSUS} = P_D(\Delta f, \nu) \delta(\nu' - \nu) \quad (\text{A.20})$$

Avec l'hypothèse de WSSUS, les fonctions d'autocorrélation sont maintenant simplifiées en des fonctions de corrélation de deux variables appelées fonctions de corrélation :

- $P_h(\Delta t, \tau)$  : densité spectrale de puissance croisée en retard
- $R_H(\Delta t, \Delta f)$  : fonction de corrélation temps-fréquence
- $C(\tau, \nu)$  : fonction de diffusion
- $P_D(\Delta f, \nu)$  : densité spectrale de puissance croisée en Doppler

La relation entre ces fonctions de corrélation est présentée dans la Figure A.3.

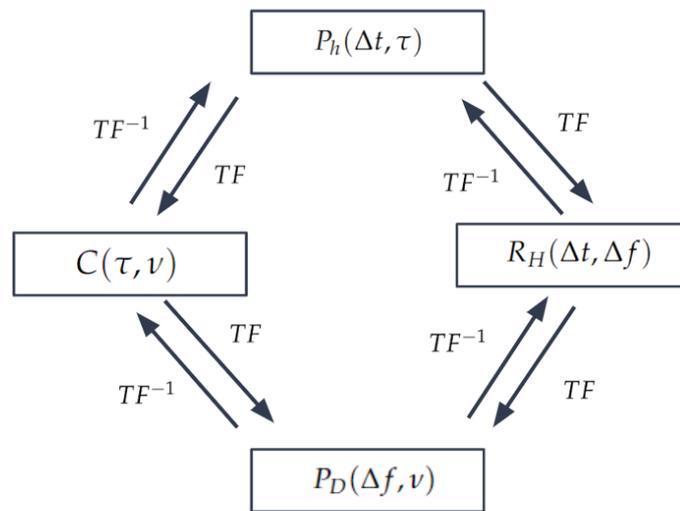


FIGURE A.3 – Illustration des fonctions de corrélation pour un canal WSSUS

À partir de ces fonctions de corrélation pour un canal WSSUS, les paramètres du canal peuvent être calculés, comme illustré dans la figure A.4.

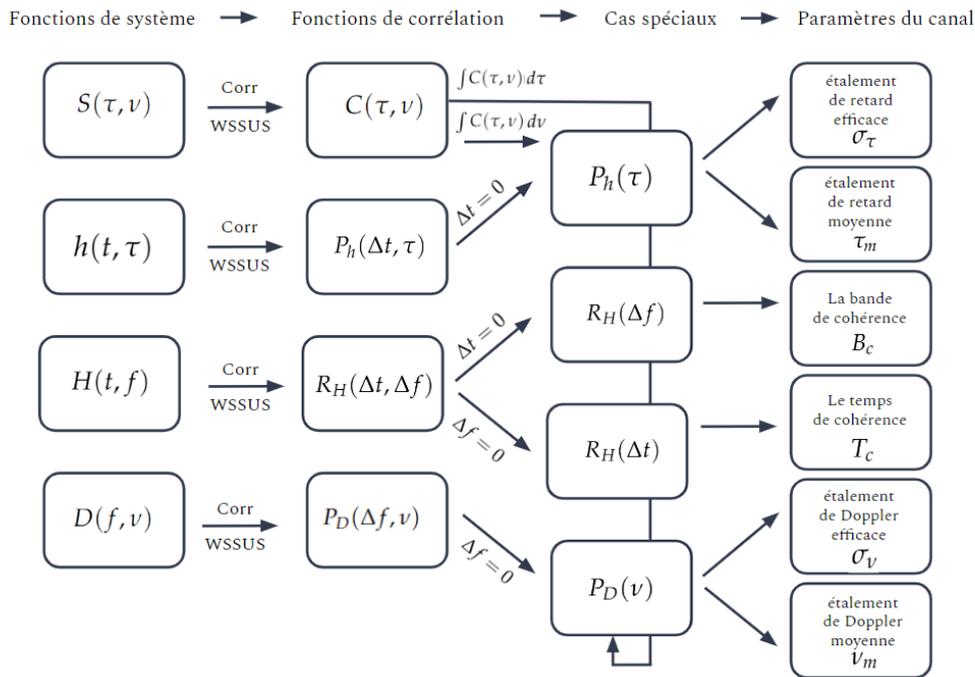


FIGURE A.4 – Relation entre les fonctions de système, les fonctions de corrélation et les paramètres du canal WSSUS

## A.2 Illustration de la colinéarité

Comme mentionné dans le chapitre 1, la colinéarité est une approche qui permet de quantifier la corrélation entre les GLSF pour deux instants  $t_1$  et  $t_2$ . Ensuite, le temps de stationnarité correspond à l'intervalle de temps centré autour de la diagonale, comme le montre la figure A.5. Il est défini dès que le coefficient de corrélation atteint un seuil choisi [68].

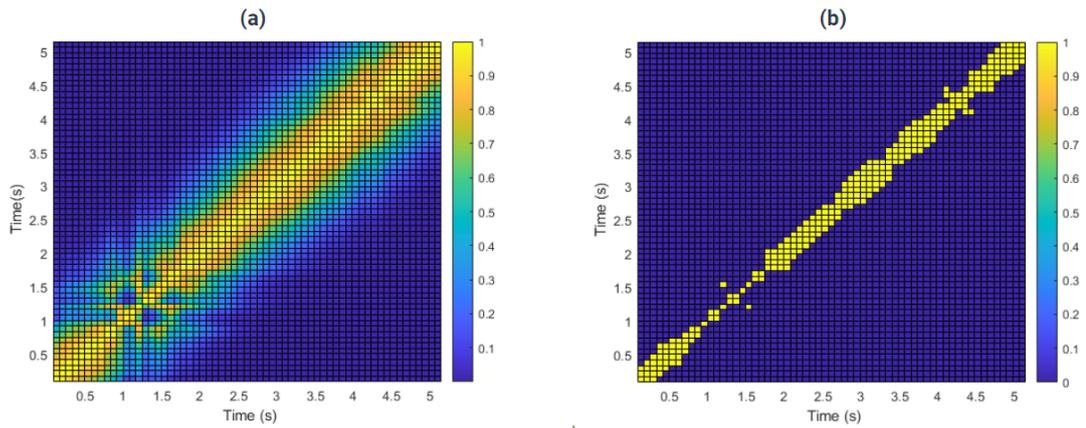


FIGURE A.5 – Calcul de la collinearité pour une trame de 5 seconde  
(a) sans seuil et (b) avec un seuil de 0.9

## Annexe B

# Caractéristiques et Algorithmes d'apprentissage Automatique

## B.1 Calculs derrière l'extraction de Caractéristiques

### B.1.1 Nombre de pics par PDP

Pour cette caractéristique, le nombre de pics par PDP a été calculé en utilisant la fonction *findpeaks* de MATLAB. Cette fonction est employée pour identifier les maxima locaux (ou "pics") pour chaque PDP, ensuite les points qui sont supérieurs de 20 dB à la valeur maximale sont retenus. Cette fonction est particulièrement utile dans l'analyse de données pour détecter des caractéristiques ou des événements significatifs dans des signaux ou des séries temporelles.

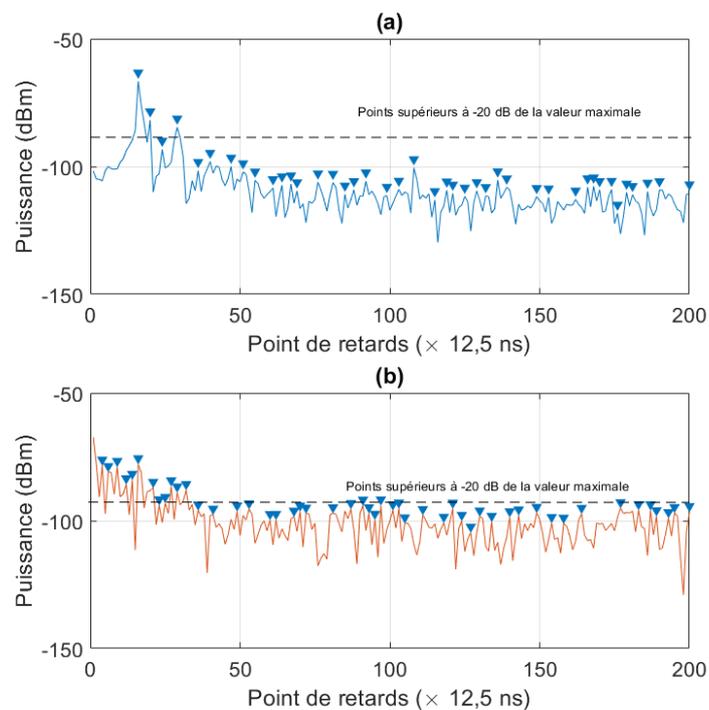


FIGURE B.1 – Exemple d'utilisation de la fonction *findpeaks* pour deux PDP en LOS et NLOS

L'option *MinPeakProminence* est un paramètre qui a été utilisé avec *findpeaks* pour spécifier la proéminence minimale d'un pic pour qu'il soit considéré comme tel ; une valeur de 3 dB a été choisie. La proéminence d'un pic est une mesure de la différence de hauteur entre le pic et son environnement, en tenant compte de la hauteur et de la largeur du pic.

### B.1.2 Nombre et largeur médiane des évanouissements fréquentiels N.E.F et L.M.E.F

Le nombre et la largeur des évanouissements fréquentiels ont été calculés pour chaque profil de puissance de la fonction de transfert  $H(f)$  (un bloc) en utilisant la même fonction *findpeaks*. De plus, la largeur de chaque pic a été spécifiée à la sortie de cette fonction. L'idée est en fait de multiplier le module de puissance de chaque fonction de transfert par -1 afin de pouvoir identifier les pics et leur largeur. Une proéminence de 3 dB a été utilisée, comme le montre le script ci-dessus.

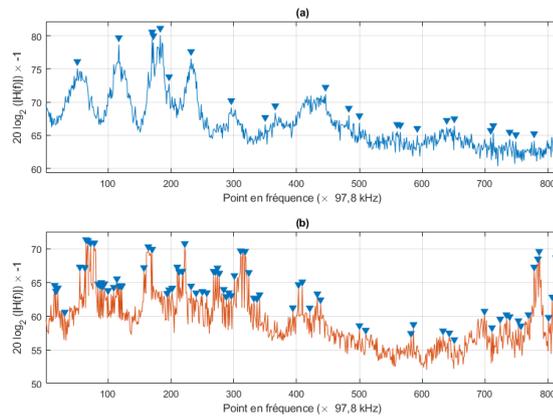


FIGURE B.2 – Identification du nombre d'évanouissements et de leur largeur pour un scénario LOS (a) et un scénario NLOS (b)

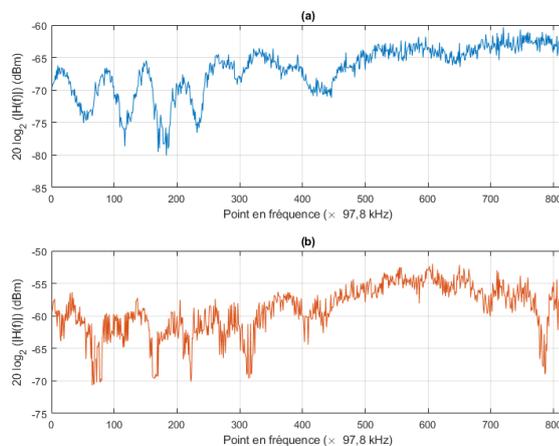


FIGURE B.3 – Profils de puissance des fonctions de transfert pour un scénario LOS (a) et un scénario NLOS (b)

### B.1.3 Retard de groupe $g_D$

Le script MATLAB pour la caractéristique *Moyenne*( $|g_D|$ ) défini comme la moyenne de la valeur absolue de la dérivée de la phase de la fonction de transfert du canal en fonction de la fréquence  $\Phi_H = \arg(H(f))$ , s'exprime par la relation :

$$g_D = -\frac{1}{2\pi} \frac{d\Phi_H}{df}$$

La dérivée de la phase est calculée en utilisant la méthode de différence avant. Cela consiste à prendre la différence entre les valeurs de phase adjacentes et à la diviser par la différence entre les fréquences correspondantes. Mathématiquement, cela peut être exprimé comme :

$$\Delta\phi = \phi_{i+1} - \phi_i$$

$$\Delta f = f_{i+1} - f_i$$

Enfin, nous divisons la différence de phase par la différence de fréquence correspondante pour obtenir une estimation du taux de variation instantané de la phase par rapport à la fréquence :

$$\frac{\Delta\phi}{\Delta f}$$

En utilisant cette méthode, nous obtenons une approximation de la dérivée de la phase par rapport à la fréquence. Cette approximation est basée sur les taux de variation instantanés entre les valeurs de phase adjacentes. Plus les différences de fréquence entre les points sont petites, plus cette approximation sera précise.

```

1 Phase = angle(squeeze(H_LOS(bloc, :))); % 1 x 818 angles
2 Df = - diff(Phase) ./ (diff(freq)*2*pi);
3 Df_final = [Df, Df(end)];
4 df_mean(bloc) = mean(abs(Df_final)); % Calcul de la moyenne

```

LISTING B.1 – Script Matlab pour la dérivée de la phase

## B.2 *k*-Nearest Neighbors (kNN)

### B.2.1 Différentes distances pour kNN

L'algorithme kNN fonctionne sur le principe de base des métriques de distance. Pour classer un nouveau point, les distances de ce nouveau point à tous les autres points dans l'ensemble de données sont calculées. Les *k* points les plus proches (ou voisins) sont identifiés, et le nouveau point est attribué à la classe qui est la plus fréquente parmi ces *k* voisins. La distance peut être calculée de plusieurs manières, la

plus courante étant la distance euclidienne présentée dans le chapitre 4. D'autres métriques de distance incluent :

#### **Distance de Manhattan (L1) :**

La distance de Manhattan, également connue sous le nom de norme L1, est utilisée pour mesurer la distance entre deux points dans un espace géométrique en suivant uniquement des trajets parallèles aux axes du système de coordonnées. Elle est définie comme la somme des différences absolues de leurs coordonnées cartésiennes.

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

#### **Distance de Minkowski :**

La distance de Minkowski est une généralisation des distances euclidiennes et de Manhattan. Elle est définie par la formule suivante, où  $p$  est un paramètre qui détermine la forme de la distance. Pour  $p = 2$ , elle devient la distance euclidienne, et pour  $p = 1$ , la distance de Manhattan.

$$d(x, y) = \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}$$

#### **Distance de Hamming :**

La distance de Hamming est utilisée pour comparer deux chaînes de même longueur en comptant le nombre de positions à laquelle les caractères correspondants sont différents. Elle est particulièrement utile pour les données catégorielles.

$$d(x, y) = \frac{1}{n} \sum_{i=1}^n \begin{cases} 1, & \text{si } x_i \neq y_i \\ 0, & \text{sinon} \end{cases}$$

#### **Distance de Chebyshev :**

Utile pour les scénarios où la distance est définie par l'écart le plus significatif entre deux points dans n'importe quelle dimension.

$$d(x, y) = \max_i |x_i - y_i|$$

Ces formules fournissent un cadre mathématique pour calculer la distance entre les points dans différents contextes, ce qui est essentiel pour l'application de l'algorithme kNN dans la classification ou la régression.

### B.2.2 Problème d'optimisation dans kNN

Il n'y a pas de problème d'optimisation explicite résolu pendant l'entraînement d'un algorithme kNN puisqu'il s'agit d'une méthode non paramétrique. Cependant, choisir la bonne valeur de  $k$  et la métrique de distance constitue l'aspect de réglage des hyperparamètres du kNN. La valeur optimale de  $k$  est généralement déterminée par validation croisée.

## B.3 Machine à Vecteurs de Support (SVM)

La machine à vecteurs de support (SVM) est un modèle d'apprentissage supervisé utilisé pour la classification et l'analyse de régression. Il est basé sur le concept de trouver un hyperplan qui divise au mieux un ensemble de données en classes.

### B.3.1 Description mathématique du SVM

L'objectif de l'algorithme SVM est de trouver l'hyperplan de séparation optimal qui maximise la marge entre les différentes classes. L'hyperplan est défini comme :

$$\mathbf{w} \cdot \mathbf{x} - b = 0 \quad ,$$

où  $\mathbf{w}$  est le vecteur normal à l'hyperplan,  $\mathbf{x}$  est le vecteur de caractéristiques, et  $b$  est le terme de biais.

### B.3.2 Problème d'optimisation dans SVM

Le problème d'optimisation pour le SVM est de minimiser :

$$\frac{1}{2} \|\mathbf{w}\|^2$$

sous les contraintes :

$$y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1, \quad \forall i \quad ,$$

où  $y_i$  est la classe LOS ou NLOS dans notre cas de la caractéristique  $\mathbf{x}_i$  et  $i$  indice des échantillons sur les points de données.

### B.3.3 Concept d'hyperplans dans SVM

Dans SVM, un hyperplan est un sous-espace affine plat de dimension  $n - 1$  dans l'espace de caractéristiques  $n$ -dimensionnel qui sépare les points de données en différentes classes. Le meilleur hyperplan est celui qui représente la plus grande séparation, ou marge, entre les deux classes. Si les données ne sont pas linéairement séparables, des fonctions de noyau (*Kernel Function*) sont utilisées pour mapper les données dans un espace de dimension supérieure où un hyperplan séparateur peut être trouvé.

### B.3.4 Hyperparamètres dans SVM

Les hyperparamètres courants dans les machines à vecteurs de support (SVM) comprennent :

**La contrainte de la boîte (C)** dans les machines à vecteurs de support (SVM) contrôle le compromis entre trouver un hyperplan de séparation de marge maximale et minimiser l'erreur de classification. Une valeur élevée de C signifie que l'erreur de classification est plus pénalisée. En d'autres termes, à mesure que la valeur de C augmente, le modèle SVM tente de minimiser l'erreur d'entraînement en permettant une marge plus petite, ce qui peut conduire à un surajustement. D'autre part, une valeur plus faible de C permet une marge plus large, ce qui peut conduire à une meilleure généralisation mais avec un risque d'augmentation de l'erreur d'entraînement. Autrement dit, Le paramètre C permet d'introduire une marge souple, permettant à certains points de se situer du mauvais côté de la marge ou même de l'hyperplan de décision pour gérer les données bruitées et éviter le sur-ajustement. Une marge dure, en revanche, n'autorise aucune violation de marge.

**Les fonctions de noyau (Kernel Function)** permettent à SVM de traiter des données non linéaires en mappant implicitement les entrées dans des espaces de caractéristiques de haute dimension où elles peuvent être séparées linéairement. Les fonctions de noyau les plus utilisées sont :

— **Noyau Linéaire**

Le noyau linéaire est défini comme suit :

$$K(x, y) = x^T y$$

Le noyau linéaire est utilisé lorsque les données sont linéairement séparables, c'est-à-dire que les classes peuvent être séparées par une seule ligne ou un hyperplan.

— **Noyau Polynomial**

Le noyau polynomial est défini comme suit :

$$K(x, y) = (x^T y + c)^d ,$$

où  $c$  est une constante et  $d$  est le degré du polynôme. Le noyau polynomial est utilisé pour transformer les données dans un espace de dimension supérieure, ce qui peut aider à séparer les classes qui ne sont pas linéairement séparables dans l'espace d'origine.

### — Noyau à Fonction de Base Radiale (RBF)

Le noyau à fonction de base radiale est défini comme suit :

$$K(x, y) = \exp(-\gamma \|x - y\|^2) \quad ,$$

où  $\gamma$  est un paramètre positif. Le noyau RBF est utilisé pour capturer les relations non linéaires entre les données et est particulièrement efficace lorsque les frontières de décision entre les classes sont complexes et non linéaires.

## B.4 Sélection des Caractéristiques

### B.4.1 t-Test

Le t-Test est une méthode statistique couramment utilisée dans le domaine de l'apprentissage automatique pour évaluer si la différence entre les moyennes de deux groupes est statistiquement significative. Dans le contexte de la sélection des caractéristiques, un test t peut être employé pour comparer les moyennes des caractéristiques entre deux groupes (par exemple, les caractéristiques pertinentes par rapport aux non pertinentes) afin d'identifier celles qui ont un impact significatif sur la prédiction du modèle. Pour effectuer ces calculs, un script et une bibliothèque sous Python ont été utilisés, comme illustré dans le script ci-dessous [B.2](#). Ce script est exécuté après l'entraînement des modèles kNN et SVM.

```

1 # T-test
2 # Import libraries for feature selection
3 from sklearn.feature_selection import SelectKBest
4 # Import libraries for the (T-test ANOVA) F-statistic
5 from sklearn.feature_selection import f_classif
6 f1_score_list_svm = []
7 f1_score_list_knn= []
8 f1_score_list_svm_test = []
9 f1_score_list_knn_test= []
10 for k in range(1, 24): # Iterate through features
11     selector_svm = SelectKBest(f_classif, k=k)
12     selector_svm.fit(X_train_v2, y_train_v2)
13     selector_knn = SelectKBest(f_classif, k=k)
14     selector_knn.fit(X_train_v2, y_train_v2)
15     # Transform Training and Testing Data for SVM and KNN:
16     sel_X_train_v2_svm = selector_svm.transform(X_train_v2)
17     sel_X_test_v2_svm = selector_svm.transform(X_test_v2)
18     sel_X_train_v2_knn = selector_knn.transform(X_train_v2)
19     sel_X_test_v2_knn = selector_knn.transform(X_test_v2)
20     sel_X_test_final_knn = selector_knn.transform(X_test_final)
21     sel_X_test_final_svm = selector_svm.transform(X_test_final)
22     # Fit SVM and KNN Classifiers:
23     svm_model.fit(sel_X_train_v2_svm, y_train_v2)
24     knn.fit(sel_X_train_v2_knn, y_train_v2)
25     # Predictions on the validation Data for SVM and KNN:
26     kbest_preds_svm = svm_model.predict(sel_X_test_v2_svm)
27     kbest_preds_knn = knn.predict(sel_X_test_v2_knn)
28     kbest_preds_svm_test = svm_model.predict(sel_X_test_final_svm)
29     kbest_preds_knn_test = knn.predict(sel_X_test_final_knn)
30     # F1 Scores for SVM and KNN and Store them:
31     f1_svm = f1_score(y_test_v2, kbest_preds_svm, average='weighted')
32     f1_knn = f1_score(y_test_v2, kbest_preds_knn, average='weighted')
33     f1_svm_test = f1_score(y_test_final, kbest_preds_svm_test, average='weighted')
34     f1_knn_test = f1_score(y_test_final, kbest_preds_knn_test, average='weighted')
35     f1_score_list_svm.append(f1_svm)
36     f1_score_list_knn.append(f1_knn)
37     f1_score_list_svm_test.append(f1_svm_test)
38     f1_score_list_knn_test.append(f1_knn_test)
39
40 plt.plot(range(1, 25), f1_score_list_svm, marker='o', label='SVM Train')
41 plt.plot(range(1, 25), f1_score_list_knn, marker='o', label='kNN Train')
42 plt.plot(range(1, 25), f1_score_list_svm_test, marker='o', label='SVM Test')
43 plt.plot(range(1, 25), f1_score_list_knn_test, marker='o', label='kNN Test')
44 plt.title('SVM & kNN Models with Various Features Selected')
45 plt.xlabel('Number of Features Selected')
46 plt.ylabel('F1 Score')
47 plt.legend()
48 plt.show()

```

LISTING B.2 – Script Python pour la méthode t-Test

# Bibliographie

1. 3GPP Technologies : AI/ML in NR <https://www.3gpp.org/technologies/ai-ml-nr>. Accessed : 2024. 2022.
2. 3GPP Specification Details <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3983>. Accessed : 2024. 2024.
3. CORNING. *Corning In-Building Networks* <https://www.corning.com/in-building-networks/cala/en/home/applications/cellular-solutions/small-cell/technology/son-auto-configuration.html>. Accessed : 2024. 2020.
4. SHARMA, P. Evolution of mobile wireless communication networks-1G to 5G as well as future prospective of next generation communication network. *International Journal of Computer Science and Mobile Computing* **2**, 47-53 (2013).
5. MISHRA, A. R. *Fundamentals of cellular network planning and optimisation : 2G/2.5G/3G... evolution to 4G* (John Wiley & Sons, 2004).
6. HALONEN, T., ROMERO, J. & MELERO, J. *GSM, GPRS and EDGE performance : evolution towards 3G/UMTS* (John Wiley & Sons, 2004).
7. HEINE, G. & SAGKOB, H. *GPRS : gateway to third generation mobile networks* (Artech House, 2003).
8. MEHTA, H., PATEL, D., JOSHI, B. & MODI, H. 0G to 5G mobile technology : A survey. *J. Basic Appl. Eng. Res* **5**, 56-60 (2014).
9. LU, L., LI, G. Y., SWINDLEHURST, A. L., ASHIKHMIN, A. & ZHANG, R. An overview of massive MIMO : Benefits and challenges. *IEEE journal of selected topics in signal processing* **8**, 742-758 (2014).
10. MARZETTA, T. L. Massive MIMO : an introduction. *Bell Labs Technical Journal* **20**, 11-22 (2015).
11. LARSSON, E. G., EDFORS, O., TUFVESSON, F. & MARZETTA, T. L. Massive MIMO for next generation wireless systems. *IEEE communications magazine* **52**, 186-195 (2014).
12. AHMED, I. *et al.* A survey on hybrid beamforming techniques in 5G : Architecture and system model perspectives. *IEEE Communications Surveys & Tutorials* **20**, 3060-3097 (2018).

13. ALI, E., ISMAIL, M., NORDIN, R. & ABDULAH, N. F. Beamforming techniques for massive MIMO systems in 5G : overview, classification, and trends for future research. *Frontiers of Information Technology & Electronic Engineering* **18**, 753-772 (2017).
14. ZHANG, J., YU, X. & LETAIEF, K. B. Hybrid beamforming for 5G and beyond millimeter-wave systems : A holistic view. *IEEE Open Journal of the Communications Society* **1**, 77-91 (2019).
15. RAPPAPORT, T. S. *et al.* Millimeter wave mobile communications for 5G cellular : It will work! *IEEE access* **1**, 335-349 (2013).
16. NIU, Y., LI, Y., JIN, D., SU, L. & VASILAKOS, A. V. A survey of millimeter wave communications (mmWave) for 5G : opportunities and challenges. *Wireless networks* **21**, 2657-2676 (2015).
17. LEI, W. *et al.* *5G system design* (Springer, 2021).
18. GOHAR, A. & NENCIONI, G. The role of 5G technologies in a smart city : The case for intelligent transportation system. *Sustainability* **13**, 5188 (2021).
19. GUEVARA, L. & AUAT CHEEIN, F. The role of 5G technologies : Challenges in smart cities and intelligent transportation systems. *Sustainability* **12**, 6469 (2020).
20. LTE, E. *Service requirements for V2X services (3GPP TS 22.185 version 14.3. 0 Release 14)* 2017.
21. ETSI, T. 123 501 V16. 6.0 (Oct. 2020). *Technical Specification G 5*.
22. DRAFT, E. EN 302 663 V1. 2.0 (2012-11) Intelligent Transport Systems (ITS). *Access layer specification for Intelligent Transport Systems operating in the 5GHz frequency band,(01.09. 2019.)*
23. ETSI. *Intelligent Transportation Systems (ITS); Vehicular communications; Basic Set of Applications; Part 2 : Specification of Cooperative Awareness (CA) applications* ETSI Technical Report TR 103 766. Version V1.1.1 (2021). <https://www.etsi.org/standard/103766>.
24. YACHEUR, B. Y., AHMED, T. & MOSBAH, M. *Implementation and assessment of IEEE 802.11 BD for improved road safety* in 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC) (2021), 1-6.
25. UZAIR, M. Vehicular Wireless Communication Standards : Challenges and Comparison. *International journal of electrical and computer engineering systems* **13**, 379-397 (2022).
26. GARCIA, M. H. C. *et al.* A tutorial on 5G NR V2X communications. *IEEE Communications Surveys & Tutorials* **23**, 1972-2026 (2021).
27. MOLINARO, A. *et al.* 5G for V2X Communications. *5G Italy White eBook* (2019).

28. ANWAR, W., FRANCHI, N. & FETTWEIS, G. *Physical layer evaluation of V2X communications technologies : 5G NR-V2X, LTE-V2X, IEEE 802.11 bd, and IEEE 802.11 p in 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)* (2019), 1-7.
29. RIBOUH, S. *Identification de l'environnement basée sur l'estimation de canal et génération de clés de sécurité pour les communications véhiculaires* thèse de doct. (Université Polytechnique Hauts-de-France, 2020).
30. ZEINALI, M., BAYRAM, I. S. & THOMPSON, J. *Performance assessment of UK's cellular network for vehicle to grid energy trading : opportunities for 5G and beyond in 2020 IEEE International Conference on Communications Workshops (ICC Workshops)* (2020), 1-6.
31. ALALEWI, A., DAYOUB, I. & CHERKAOUI, S. On 5G-V2X use cases and enabling technologies : A comprehensive survey. *Ieee Access* **9**, 107710-107737 (2021).
32. 3GPP. *V2X Services based on NR; User Equipment (UE) radio transmission and reception* <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3615>. Specification Number : 38.886, Release 16, Status : Under change control, Creation Date : 2019-04-01, Last Update : 2020-07-07, Primary responsible group : RAN 4. 2020.
33. MIKAMI, M. *et al. Field experimental evaluation on latency and reliability performance of 5G NR V2V direct communication in real express highway environment in 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)* (2020), 1-5.
34. GHAZAL, A., WANG, C.-X., AI, B., YUAN, D. & HAAS, H. A nonstationary wideband MIMO channel model for high-mobility intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems* **16**, 885-897 (2014).
35. BERNADÓ, L., ZEMEN, T., TUFVESSON, F., MOLISCH, A. F. & MECKLENBRÄUKER, C. F. Time-and frequency-varying  $K$ -factor of non-stationary vehicular channels for safety-relevant scenarios. *IEEE transactions on intelligent transportation systems* **16**, 1007-1017 (2014).
36. Von BUTOVITSCH, P *et al. Ericsson White Paper—Advanced Antenna Systems for 5G Networks. 2020*
37. MATZ, G. On non-WSSUS wireless fading channels. *IEEE transactions on wireless communications* **4**, 2465-2478 (2005).
38. YUSUF, M. *et al. Stationarity analysis of V2I radio channel in a suburban environment. IEEE Transactions on Vehicular Technology* **68**, 11532-11542 (2019).
39. YUSUF, M. *et al. Experimental study on the impact of antenna characteristics on non-stationary V2I channel parameters in tunnels. IEEE Transactions on Vehicular Technology* **69**, 12396-12407 (2020).

40. BERNADÓ, L., ZEMEN, T., TUFVESSON, F., MOLISCH, A. F. & MECKLENBRÄUKER, C. F. *The (in-) validity of the WSSUS assumption in vehicular radio channels* in 2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications-(PIMRC) (2012), 1757-1762.
41. PAIER, A. *et al.* *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 (2008), 9-15.
42. RADOVIC, D. *et al.* *Stationarity Evaluation of High-mobility sub-6 GHz and mmWave non-WSSUS Channels* in 2023 XXXVth General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS) (2023), 1-4.
43. RASHDAN, I., SAND, S., JIANG, S., WANG, W. & CAIRE, G. *Non-Stationarity Analysis of Vehicle-to-Vulnerable Road Users Channel in Critical Scenarios* in 2023 17th European Conference on Antennas and Propagation (EuCAP) (2023), 1-5.
44. UNTERHUBER, P., WALTER, M., FIEBIG, U.-C. & KÜRNER, T. *Stochastic channel parameters for train-to-train communications*. *IEEE Open Journal of Antennas and Propagation* **2**, 778-792 (2021).
45. BERNADÓ, L. *et al.* *Non-WSSUS vehicular channel characterization at 5.2 GHz-spectral divergence and time-variant coherence parameters*. *Dipl.-Ing. Dr. techn. Thomas Zemen*, 45 (2008).
46. MATZ, G. *Characterization and analysis of doubly dispersive MIMO channels* in 2006 Fortieth Asilomar Conference on Signals, Systems and Computers (2006), 946-950.
47. RENAUDIN, O., KOLMONEN, V.-M., VAINIKAINEN, P. & OESTGES, C. *Non-stationary narrowband MIMO inter-vehicle channel characterization in the 5-GHz band*. *IEEE transactions on vehicular technology* **59**, 2007-2015 (2010).
48. ZOU, Z., CAREEM, M., DUTTA, A. & THAWDAR, N. *Joint spatio-temporal precoding for practical non-stationary wireless channels*. *IEEE Transactions on Communications* **71**, 2396-2409 (2023).
49. ZHANG, Y. *et al.* *Non-Stationarity Characterization and Geometry-Cluster-Based Stochastic Model for High-Speed Train Radio Channels*. *IEEE Transactions on Intelligent Transportation Systems* (2023).
50. PRATSCHNER, S. *et al.* *Measured user correlation in outdoor-to-indoor massive MIMO scenarios*. *IEEE Access* **8**, 178269-178282 (2020).
51. HERDIN, M., CZINK, N., OZCELIK, H. & BONEK, E. *Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels* in 2005 IEEE 61st Vehicular Technology Conference **1** (2005), 136-140.

52. RODRÍGUEZ-CORBO, F. A., AZPILICUETA, L., CELAYA-ECHARRI, M., SHUBAIR, R. & FALCONE, F. MmWave Channel Stationarity Analysis of V2X Communications in an Urban Environment. *IEEE Antennas and Wireless Propagation Letters* (2023).
53. YUAN, Y., WANG, C.-X., HE, Y., ALWAKEEL, M. M. *et al.* 3D wideband non-stationary geometry-based stochastic models for non-isotropic MIMO vehicle-to-vehicle channels. *IEEE Transactions on Wireless Communications* **14**, 6883-6895 (2015).
54. LI, Y., AI, B., CHENG, X., LIN, S., ZHONG, Z. *et al.* A TDL based non-WSSUS vehicle-to-vehicle channel model. *International journal of antennas and propagation* **2013** (2013).
55. HE, R., BAI, F., MAO, G., HÄRRI, J. & KYÖSTI, P. Guest editorial 5G wireless communications with high mobility. *IEEE Journal on Selected Areas in Communications* **38**, 2717-2722 (2020).
56. PEI, H., CHEN, X., ZHANG, M. & ZHANG, A. *Over-the-air testing of 5G millimeter-wave system with adaptive beamforming in 2019 IEEE MTT-S International Wireless Symposium (IWS)* (2019), 1-3.
57. CHENG, X., HUANG, Z. & CHEN, S. Vehicular communication channel measurement, modelling, and application for beyond 5G and 6G. *IET Communications* **14**, 3303-3311 (2020).
58. ETSI, T. 138 901 v16. 1.0, "5G. Study on channel model for frequencies from 0.5 to 100.
59. CHENG, X., HUANG, Z. & BAI, L. Channel nonstationarity and consistency for beyond 5G and 6G : A survey. *IEEE Communications Surveys & Tutorials* **24**, 1634-1669 (2022).
60. HUANG, H., WANG, K., WU, P., ZHANG, J. & XIA, M. Flexible 3GPP MIMO Channel Modeling and Calibration With Spatial Consistency. *IEEE Access* (2023).
61. ZHENNIAN, S. *et al.* The new architecture with time-spatial consistency for 5G networks. *china communications* **13**, 68-79 (2016).
62. YU, G., TIAN, L., ZHANG, J. & WANG, C. *The method to implement 5G channel model with spatial consistency in 2018 IEEE/CIC International Conference on Communications in China (ICCC)* (2018), 736-740.
63. BOBAN, M., GONG, X., MANOLAKIS, K. & XU, W. *Towards Spatially and Temporally Consistent Channel Modeling for Reliable V2X Communication in WSA 2017; 21th International ITG Workshop on Smart Antennas* (2017), 1-5.
64. BELLO, P. Characterization of randomly time-variant linear channels. *IEEE transactions on Communications Systems* **11**, 360-393 (1963).

65. HLAWATSCH, F. & MATZ, G. *Wireless communications over rapidly time-varying channels* (Academic press, 2011).
66. WALTER, M. *Scattering in non-stationary mobile-to-mobile communications channels* (Herbert Utz Verlag, 2016).
67. CHENG, X., HUANG, Z. & BAI, L. *Channel Characterization and Modeling for Vehicular Communications* (Springer Nature, 2023).
68. PARDONCHE, J.-F. Systèmes de transmission sans fil multi-émetteurs, multi-récepteurs pour des applications transports. etude des modèles de canal de propagation. *Université des sciences et technologies de Lille* (2004).
69. MATZ, G. *Doubly underspread non-WSSUS channels : Analysis and estimation of channel statistics* in 2003 4th IEEE Workshop on Signal Processing Advances in Wireless Communications-SPAWC 2003 (IEEE Cat. No. 03EX689) (2003), 190-194.
70. SLEPIAN, D. Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V : The discrete case. *Bell System Technical Journal* **57**, 1371-1430 (1978).
71. ISPAS, A., ASCHEID, G., SCHNEIDER, C. & THOMA, R. *Analysis of local quasi-stationarity regions in an urban macrocell scenario* in 2010 IEEE 71st Vehicular Technology Conference (2010), 1-5.
72. CZANIERA, D. *et al. Investigation on stationarity of v2v channels in a highway scenario* in 2019 13th European Conference on Antennas and Propagation (EuCAP) (2019), 1-5.
73. ETSI, T. 125 996 V6. 1.0 (2003-09). *Universal Mobile Telecommunications System (UMTS); Spatial channel model for Multiple Input Multiple Output (MIMO) simulations* (2003).
74. XIAO, H., BURR, A. G. & SONG, L. *A time-variant wideband spatial channel model based on the 3GPP model* in IEEE Vehicular Technology Conference (2006), 1-5.
75. BERNADÓ, L. *et al. In-tunnel vehicular radio channel characterization* in 2011 IEEE 73rd Vehicular Technology Conference (VTC Spring) (2011), 1-5.
76. EGEA-LOPEZ, E., MOLINA-GARCIA-PARDO, J.-M., LIENARD, M., LALY, P. & DEGAUQUE, P. *Numerical and Experimental Analysis of Massive MIMO Channel Characteristics in a Rectangular Highway Tunnel at 5.9 GHz* in 2023 27th International Conference on Circuits, Systems, Communications and Computers (CSCC) (2023), 277-282.
77. CHUDE-OKONKWO, U. A., NGAH, R. & ABD RAHMAN, T. Time-scale domain characterization of non-WSSUS wideband channels. *EURASIP Journal on Advances in Signal Processing* **2011**, 1-20 (2011).
78. ALDAYEL, O., BENGTSSON, M. & ALSHEBEILI, S. A. *Evaluation of MIMO channel non-stationarity* in 21st European Signal Processing Conference (EUSIPCO 2013) (2013), 1-5.

79. JU, S. & RAPPAPORT, T. S. *140 GHz urban microcell propagation measurements for spatial consistency modeling* in *ICC 2021-IEEE International Conference on Communications* (2021), 1-6.
80. DAI, S. & KURRAS, M. *Spatial Consistency Evaluation Based on Massive SIMO Measurements* in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)* (2019), 1-6.
81. KURRAS, M., DAI, S., JAECKEL, S. & THIELE, L. *Evaluation of the spatial consistency feature in the 3GPP geometry-based stochastic channel model* in *2019 IEEE Wireless Communications and Networking Conference (WCNC)* (2019), 1-6.
82. AL-TOUS, H., KAZEMI, P. & TIRKKONEN, O. *Coordinated uplink precoding for spatially consistent mmWave channel covariance measurements* in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)* (2020), 1-5.
83. BIGLIERI, E., PROAKIS, J. & SHAMAI, S. *Fading channels : Information-theoretic and communications aspects. IEEE transactions on information theory* **44**, 2619-2692 (1998).
84. CHUDE-OKONKWO, U. A., NGAH, R. & ABD RAHMAN, T. *Non-WSSUS analysis of measurement data for mobile communication channel* in *The 17th Asia Pacific Conference on Communications* (2011), 682-687.
85. SAYEED, A. M. & AAZHANG, B. *Joint multipath-Doppler diversity in mobile wireless communications. IEEE transactions on communications* **47**, 123-132 (1999).
86. DAS, K. & BEHERA, R. N. *A survey on machine learning : concept, algorithms and applications. International Journal of Innovative Research in Computer and Communication Engineering* **5**, 1301-1309 (2017).
87. BURKART, N. & HUBER, M. F. *A survey on the explainability of supervised machine learning. Journal of Artificial Intelligence Research* **70**, 245-317 (2021).
88. QIU, J., WU, Q., DING, G., XU, Y. & FENG, S. *A survey of machine learning for big data processing. EURASIP Journal on Advances in Signal Processing* **2016**, 1-16 (2016).
89. GORI, M., BETTI, A. & MELACCI, S. *Machine Learning : A constraint-based approach* (Elsevier, 2023).
90. CHONG, E. K., LU, W.-S. & ZAK, S. H. *An Introduction to Optimization : With Applications to Machine Learning* (John Wiley & Sons, 2023).
91. ZHANG, A., LIPTON, Z. C., LI, M. & SMOLA, A. J. *Dive into deep learning* (Cambridge University Press, 2023).
92. MOHAMMED, A. & KORA, R. *A comprehensive review on ensemble deep learning : Opportunities and challenges. Journal of King Saud University-Computer and Information Sciences* (2023).

93. XIROUCHAKIS, I. Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations A Ray Tracing Simulator Based on 3GPP TR 25.996 v. 6.1.0. *Physics Department, University of Athens* (2008).
94. ZHANG, J. *et al.* Wireless channel propagation scenarios identification : A perspective of machine learning. *IEEE Access* **8**, 47797-47806 (2020).
95. XIAO, Z. *et al.* Non-line-of-sight identification and mitigation using received signal strength. *IEEE Transactions on Wireless Communications* **14**, 1689-1702 (2014).
96. XIAO, F., GUO, Z., ZHU, H., XIE, X. & WANG, R. *AmpN : Real-time LOS/NLOS identification with WiFi in 2017 IEEE International Conference on Communications (ICC)* (2017), 1-7.
97. HUANG, C. *et al.* Machine learning-enabled LOS/NLOS identification for MIMO systems in dynamic environments. *IEEE Transactions on Wireless Communications* **19**, 3643-3657 (2020).
98. KURNIAWAN, E., ZHIWEI, L. & SUN, S. *Machine learning-based channel classification and its application to IEEE 802.11 ad communications in GLOBECOM 2017-2017 IEEE Global Communications Conference* (2017), 1-6.
99. RAMADAN, M., SARK, V., GUTIERREZ, J. & GRASS, E. *NLOS identification for indoor localization using random forest algorithm in WSA 2018 ; 22nd International ITG Workshop on Smart Antennas* (2018), 1-5.
100. SUZUKI, T. & AMANO, Y. NLOS multipath classification of GNSS signal correlation output using machine learning. *Sensors* **21**, 2503 (2021).
101. MARANO, S., GIFFORD, W. M., WYMEERSCH, H. & WIN, M. Z. NLOS identification and mitigation for localization based on UWB experimental data. *IEEE Journal on selected areas in communications* **28**, 1026-1035 (2010).
102. JAYAWARDANA, P. A. N., OBAID, H., YESILYURT, T., TAN, B. & LOHAN, E. S. Machine-Learning-Based LOS Detection for 5G Signals with Applications in Airport Environments. *Sensors* **23**, 1470 (2023).
103. KIRMAZ, A., MICHALOPOULOS, D. S., BALAN, I. & GERSTACKER, W. LOS/NLOS Classification Using Scenario-Dependent Unsupervised Machine Learning, 1134-1140 (2021).
104. JIANG, C. *et al.* UWB NLOS/LOS classification using deep learning method. *IEEE Communications Letters* **24**, 2226-2230 (2020).
105. BREGAR, K. & MOHORČIČ, M. Improving indoor localization using convolutional neural networks on computationally restricted devices. *IEEE Access* **6**, 17429-17441 (2018).
106. SOSNIN, S., LOMAYEV, A. & KHORYAEV, A. *NLOS links detection algorithm for improved 5G NR indoor positioning in 2021 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)* (2021), 1-6.

107. HU, Q., GAO, F., ZHANG, H., JIN, S. & LI, G. Y. Deep learning for channel estimation : Interpretation, performance, and comparison. *IEEE Transactions on Wireless Communications* **20**, 2398-2412 (2020).
108. SIRIWANITPONG, A., BOONSRIMUANG, P., MORI, K. & BOONSRIMUANG, P. *A Deep Learning-Based Channel Estimation for High-Speed Train Environments in 2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) (2022)*, 1-4.
109. SHARMA, A., HAQ, S. A. U. & DARAK, S. J. Low complexity deep learning augmented wireless channel estimation for pilot-based ofdm on zynq system on chip. *IEEE Transactions on Circuits and Systems I : Regular Papers* (2024).
110. LIAO, Y., HUA, Y., DAI, X., YAO, H. & YANG, X. *ChanEstNet : A deep learning based channel estimation for high-speed scenarios in ICC 2019-2019 IEEE international conference on communications (ICC) (2019)*, 1-6.
111. PAN, J. *et al.* Channel estimation based on deep learning in vehicle-to-everything environments. *IEEE Communications Letters* **25**, 1891-1895 (2021).
112. GAO, J., ZHONG, C., LI, G. Y., SORIAGA, J. B. & BEHBOODI, A. Deep learning-based channel estimation for wideband hybrid mmWave massive MIMO. *IEEE Transactions on Communications* (2023).
113. RAHMAN, M. H., SHAHJALAL, M., ALI, M. O., YOON, S. & JANG, Y. M. *Deep Learning Based Pilot Assisted Channel Estimation for Rician Fading Massive MIMO Uplink Communication System in 2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN) (2021)*, 470-472.
114. SATTIRAJU, R., WEINAND, A. & SCHOTTEN, H. D. *Channel estimation in C-V2X using deep learning in 2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS) (2019)*, 1-5.
115. CHUN, C.-J., KANG, J.-M. & KIM, I.-M. Deep learning-based channel estimation for massive MIMO systems. *IEEE Wireless Communications Letters* **8**, 1228-1231 (2019).
116. YUAN, J., NGO, H. Q. & MATTHAIYOU, M. *Machine learning-based channel estimation in massive MIMO with channel aging in 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) (2019)*, 1-5.
117. LALY, P. *Sondeur de canal MIMO temps réel et applications* thèse de doct. (Lille 1, 2016).
118. CHOI, M.-S., KIM, D.-Y., JO, H.-S., YOON, J.-G. & PARK, H.-K. Path-loss characteristics in subway tunnels at 2.65 GHz. *Microwave and Optical Technology Letters* **48**, 383-386 (2006).

119. QIAN, J., WU, Y., SALEEM, A. & ZHENG, G. Path loss model for 3.5 GHz and 5.6 GHz bands in cascaded tunnel environments. *Sensors* **22**, 4524 (2022).
120. GUAN, K. *et al.* Measurements and analysis of large-scale fading characteristics in curved subway tunnels at 920 MHz, 2400 MHz, and 5705 MHz. *IEEE Transactions on Intelligent Transportation Systems* **16**, 2393-2405 (2015).
121. GAILLOT, D. *et al.* Measurement of the V2I massive radio channel with the MaMIMOSA sounder in a suburban environment in 2021 15th European Conference on Antennas and Propagation (EuCAP) (2021), 1-4.
122. DAHMOUNI, N. E. I. *et al.* On the Stationarity Time of a Vehicle-to-Infrastructure Massive Radio Channel in a Line-of-Sight Suburban Environment. *Sensors* **22**, 8420 (2022).
123. DAHMOUNI, N. E. I., MALLIK, M., LALY, P., EGEEA-LOPEZ, E. & GAILLOT, D. P. A Comparative Analysis of Single-Input Single-Output Stationarity Time and Pathloss in Suburban Environments for Vehicle-to-Infrastructure Channel Using Veneris Ray-Tracing and Real Data in The Thirteenth International Conference on Advances in Vehicular Systems, Technologies and Applications VEHICULAR 2024 (2024), 6.
124. GAO, X., EDFORS, O., RUSEK, F. & TUFVESSON, F. Massive MIMO performance evaluation based on measured propagation data. *IEEE Transactions on Wireless Communications* **14**, 3899-3911 (2015).
125. SCHARF, L. L. *Statistical signal processing : applications to beamforming, detection, and estimation in Medical Imaging 2004 : Ultrasonic Imaging and Signal Processing* **5373** (2004), 1-17.
126. MARIAGE, P., LIÉNARD, M. & DEGAUQUE, P. Theoretical and experimental approach of the propagation of high frequency waves in road tunnels. *IEEE Transactions on Antennas and Propagation* **42**, 75-81 (1994).
127. LIENARD, M. & DEGAUQUE, P. Propagation in wide tunnels at 2 GHz : A statistical analysis. *IEEE Transactions on Vehicular Technology* **47**, 1322-1328 (1998).
128. RAPPAPORT, T. S. *Wireless communications : Principles and practice, 2/E* (Pearson Education India, 2010).
129. WANG, M., CUI, Y., WANG, X., XIAO, S. & JIANG, J. Machine learning for networking : Workflow, advances and opportunities. *Ieee Network* **32**, 92-99 (2017).
130. VALENTIN, S. *et al.* Designing optimal behavioral experiments using machine learning. *Elife* **13**, e86224 (2024).
131. ZELAYA, C. V. G. Towards explaining the effects of data preprocessing on machine learning in 2019 IEEE 35th international conference on data engineering (ICDE) (2019), 2086-2090.

132. ILYAS, I. F. & REKATSINAS, T. Machine Learning and Data Cleaning : Which Serves the Other? *ACM Journal of Data and Information Quality (JDIQ)* **14**, 1-11 (2022).
133. TAE, K. H., ROH, Y., OH, Y. H., KIM, H. & WHANG, S. E. *Data cleaning for accurate, fair, and robust models : A big data-AI integration approach* in *Proceedings of the 3rd International Workshop on Data Management for End-to-End Machine Learning* (2019), 1-4.
134. CHU, X., ILYAS, I. F., KRISHNAN, S. & WANG, J. *Data cleaning : Overview and emerging challenges* in *Proceedings of the 2016 international conference on management of data* (2016), 2201-2206.
135. RAJU, V. G., LAKSHMI, K. P., JAIN, V. M., KALIDINDI, A. & PADMA, V. *Study the influence of normalization/transformation process on the accuracy of supervised classification* in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)* (2020), 729-735.
136. SINGH, D. & SINGH, B. Investigating the impact of data normalization on classification performance. *Applied Soft Computing* **97**, 105524 (2020).
137. MARINO, C. S. & CHAU, P. M. *Computationally efficient beamforming on real experimental data* in *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004.* **1** (2004), 275-279.
138. DHANYA, R, PAUL, I. R., AKULA, S. S., SIVAKUMAR, M. & NAIR, J. J. F-test feature selection in Stacking ensemble model for breast cancer prediction. *Procedia Computer Science* **171**, 1561-1570 (2020).
139. CASO, G., DE NARDIS, L. & GAVRILOVSKA, L. *Cognitive Radio-Oriented Wireless Networks* (Springer, 2021).
140. WANG, D., ZHANG, H., LIU, R., LV, W. & WANG, D. t-Test feature selection approach based on term frequency for text categorization. *Pattern Recognition Letters* **45**, 1-10 (2014).
141. CHAVES, R *et al.* SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting. *Neuroscience letters* **461**, 293-297 (2009).
142. CHU, C. *et al.* Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage* **60**, 59-70 (2012).
143. ABDULMOHSIN, H. A., ABDULWAHAB, H. B. & HOSSEN, A. M. J. A. A New Hybrid Feature Selection Method Using T-test and Fitness Function. *Computers, Materials & Continua* **68** (2021).
144. LUO, L. *et al.* Methods of forward feature selection based on the aggregation of classifiers generated by single attribute. *Computers in biology and medicine* **41**, 435-441 (2011).

145. ZOLLANVARI, A. in *Machine Learning with Python : Theory and Implementation* 283-302 (Springer, 2023).
146. ZAIDI, S. A. R. Nearest neighbour methods and their applications in design of 5G & beyond wireless networks. *ICT Express* 7, 414-420 (2021).
147. TABAA, M, DIOU, C, EL AROUSSI, M, CHOURI, B & DANDACHE, A. *LOS and NLOS identification based on UWB stable distribution in 2013 25th international conference on microelectronics (ICM)* (2013), 1-4.
148. GUO, G., WANG, H., BELL, D., BI, Y. & GREER, K. *KNN model-based approach in classification in On The Move to Meaningful Internet Systems 2003 : CoopIS, DOA, and ODBASE : OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings* (2003), 986-996.
149. PRASATH, V. *et al.* Distance and similarity measures effect on the performance of K-Nearest Neighbor classifier–A review. *arXiv preprint arXiv :1708.04321* (2017).
150. De l'auteur (s'il est DISPONIBLE), N. *Support Vector Machine (SVM)* <https://dataanalyticspost.com/Lexique/svm/>.
151. DENG, N., TIAN, Y. & ZHANG, C. *Support vector machines : optimization based theory, algorithms, and extensions* (CRC press, 2012).
152. LEE, S. B., GUI, X., MANQUEN, M. & HAMILTON, E. R. *Use of training, validation, and test sets for developing automated classifiers in quantitative ethnography in Advances in Quantitative Ethnography : First International Conference, ICQE 2019, Madison, WI, USA, October 20–22, 2019, Proceedings 1* (2019), 117-127.
153. RASCHKA, S. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv :1811.12808* (2018).
154. MOLISCH, A. F. *Wireless communications* (John Wiley & Sons, 2012).