



HAL
open science

Modélisation des zones de réception de l'AIS en environnement maritime par réseaux de neurones en graphes

Ambroise Renaud

► To cite this version:

Ambroise Renaud. Modélisation des zones de réception de l'AIS en environnement maritime par réseaux de neurones en graphes. Risques. Université Paris sciences et lettres, 2025. Français. ⟨NNT : 2025UPSLM048⟩. ⟨tel-05524941⟩

HAL Id: tel-05524941

<https://pastel.hal.science/tel-05524941v1>

Submitted on 24 Feb 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Mines Paris - PSL

**Modélisation des zones de réception de l' AIS en
environnement maritime par réseaux de neurones
en graphes**

Soutenue par

Ambroise RENAUD

Le 28 novembre 2025

Ecole doctorale n° 621

**Ingénierie des Systèmes,
Matériaux, Mécanique,
Énergétique**

Spécialité

**Sciences et génie des
activités à risques**

Composition du jury :

John, PUENTES Professeur des universités, IMT Atlantique	<i>Président</i>
Abdel, BOUDRAA Professeur des universités, École Navale	<i>Rapporteur</i>
Nicolas, TRAVERS Full Professor, ESILV	<i>Rapporteur</i>
Anne-Laure, JOUSSELME Directrice de recherche HDR, CS GROUP	<i>Examinatrice</i>
Clément, IPHAR Ingénieur de recherche, UBO	<i>Examineur</i>
Aldo, NAPOLI Directeur de recherche, Mines Paris - PSL	<i>Directeur de thèse</i>

Remerciements

Mes premiers remerciements vont naturellement à mes encadrants de thèse, Aldo NAPOLI et Clément IPHAR. Je souhaite exprimer toute ma gratitude à Aldo pour sa patience, ses conseils avisés, son exigence scientifique et son accompagnement tout au long de ces années, ainsi que pour la confiance qu'il m'a accordée dès le début de ma thèse. Je remercie également Clément pour sa disponibilité et ses précieux retours, que ce soit dans les phases de conception, de prototypage ou de rédaction.

Je tiens à remercier très sincèrement les membres de mon jury de thèse, Abdel BOUDRAA, Nicolas TRAVERS, Anne-Laure JOUSSELME et John PUENTES, pour avoir accepté d'évaluer ce travail. Je les remercie également pour leur présence physique lors de la soutenance, qui a constitué un moment important de discussion et d'échange. J'adresse enfin un remerciement tout particulier à John PUENTES pour avoir accepté de présider le jury.

Je suis également reconnaissant aux membres des comités de suivi, Cyril RAY et Éric RIGAUD, pour leurs conseils lors des étapes clés de mon parcours doctoral.

Mes remerciements vont au CRC, à son directeur Franck GUARNIERI, ainsi qu'à l'ensemble du personnel administratif et technique, en particulier Brice CHAPPE, Samuel OLAMPI et Luca ISTRATE, pour leur aide précieuse dans la mise à disposition, la maintenance et la configuration du matériel informatique.

Je voudrais adresser une pensée particulière à ma famille, ma compagne et mes amis, dont le soutien moral, affectif et parfois matériel a été essentiel.

Enfin, je tiens à remercier toutes les personnes, rencontrées au fil de mes études et de mes stages, qui ont contribué de près ou de loin à nourrir ma curiosité scientifique.

Table des matières

Introduction.....	1
Contexte applicatif.....	1
La connaissance de la situation maritime	1
Le système d'identification automatique	2
Exploitation des données AIS.....	3
Anomalies dans le domaine maritime.....	4
Anomalies liées à la donnée.....	4
Problématique de recherche.....	6
Approches traditionnelles de modélisation de la réception de l'AIS.....	6
Apprentissage automatique et approches dirigées par les données pour la modélisation de la réception AIS	8
Défis liés aux données et problématique de recherche	10
Postulats, hypothèses et objectifs de recherche	11
Postulats de départ	11
Hypothèses de travail et objectif de la thèse.....	12
Organisation de la thèse.....	12
Chapitre 1 : Propagation des signaux AIS en environnement maritime.....	15
1.1 Phénomènes et paramètres de la propagation de l'AIS	15
1.1.1 Phénomènes de propagation	16
1.1.2 Typologies des phénomènes de propagation	23
1.1.3 Caractéristiques des systèmes de communication AIS.....	25
1.2 Modélisation physique de la propagation de l'AIS.....	26
1.2.1 Modèles physiques classiques de propagation.....	27
1.2.2 Application des modèles physiques à la propagation de l'AIS.....	34
1.2.3 Prise en compte des phénomènes de propagation dans les modèles.....	36
1.2.4 Les modèles ITU comme références.....	39
1.3 Conclusion	40
Chapitre 2 : Modélisation de la propagation des ondes radio par apprentissage automatique et profond.....	43
2.1 Apprentissage automatique pour la modélisation de la propagation	43

2.1.1	Types d'apprentissage.....	44
2.1.2	Défis liés aux données pour la modélisation dirigée par les données.....	48
2.1.3	Méthodes d'apprentissage automatique non profond	50
2.1.4	Limites de ces approches	51
2.2	Apprentissage profond pour la modélisation de la propagation	52
2.2.1	Principales architectures	53
2.2.2	Un pouvoir de représentation limité	53
2.3	Réseaux de neurones en graphes pour la modélisation de la propagation.....	54
2.3.1	Représentation des données sous forme de graphe.....	55
2.3.2	Principales architectures	58
2.4	Conclusion	65
Chapitre 3 : Méthodologie de modélisation des zones de réception de l'AIS à l'aide de réseaux de neurones en graphes et de données environnementales.....		67
3.1	Vue d'ensemble de la démarche	67
3.1.1	Démarche proposée.....	68
3.1.2	Schémas de la démarche	72
3.2	Architecture neuronale proposée	74
3.2.1	Architecture proposée	74
3.2.2	Variantes explorées.....	77
3.3	Données mobilisées et transformations	78
3.3.1	Sources de données	78
3.3.2	Données issues de mesures	79
3.3.3	Donnée contextuelles et environnementales.....	87
3.3.4	Construction du jeu de données	90
3.4	Construction de graphes et ingénierie des caractéristiques.....	92
3.4.1	Définition des nœuds	92
3.4.2	Construction des arêtes	92
3.4.3	Standardisation et transformations.....	95
3.5	Conclusion	95
Chapitre 4 : Evaluation expérimentale de la démarche proposée		97
4.1	Vue d'ensemble de la démarche expérimentale.....	97
4.2	Présentation des modèles de référence	98
4.2.1	Caractéristiques des modèles	98
4.2.2	Modèle physique : ITU-R P.2001	99
4.2.3	Modèle d'apprentissage automatique : XGBoost	102

4.2.4	Modèle à base de graphes : GCN et GAT.....	103
4.3	Implémentation logicielle	103
4.3.1	Implémentation des modèles.....	103
4.3.2	Implémentation du module de test.....	109
4.3.3	Implémentation du module de prédiction	111
4.3.4	Jeux de données et entraînement des modèles	112
4.3.5	Optimisation des hyperparamètres.....	117
4.4	Evaluation des performances	118
4.4.1	Présentation des métriques d'évaluation.....	118
4.4.2	Prédiction et visualisation	118
4.4.3	Résultats expérimentaux et comparaison avec les autres modèles	122
4.4.4	Evaluation par zones	122
4.5	Limites et discussion.....	126
4.5.1	Limites méthodologiques.....	126
4.5.2	Limites liées à la modélisation.....	126
4.5.3	Limites liées à l'évaluation	127
4.5.4	Contraintes liées aux ressources	127
4.5.5	Perspectives de modélisation et transfert de tâche.....	128
4.6	Conclusion	130
	Conclusion	133
	Bilan et conclusion générale	133
	Méthodes pour la prédiction des zones de réception de l'AIS.....	133
	Conception d'une chaîne de traitement pour la modélisation en graphes et l'apprentissage par GNN.....	134
	Prototypage	135
	Evaluation	136
	Perspectives.....	136
	Perspectives de modélisation	136
	Perspectives d'évaluation.....	138
	Perspectives d'application	139

Table des figures

Figure 1.1 : Vue schématique des phénomènes de propagation de l’AIS et importance de leurs effets sur la propagation (rouge : Négligeable, orange : Peu significatif, Vert : Influence).....	17
Figure 1.2 : Profils typiques d’indice de réfraction modifié : (a) infra-réfraction, (b) atmosphère standard, (c) super-réfraction, (d) conduit d’évaporation, (e) conduit de surface, (f) conduit basé en surface, (g) conduit élevé.....	20
Figure 1.3 : Schématisation géométrique d’une communication par diffraction au-dessus d’un obstacle isolé dans un modèle en lame de couteau.....	21
Figure 1.4 : Courbe représentative d’une atténuation de référence calculée par ITS ITM (Hufford et al. 1982)	34
Figure 2.1 : Les trois grands paradigmes d’apprentissage automatique, leurs entrées typiques et sorties.....	45
Figure 2.2 : Diagramme de flux de la représentation d’un graphe de propagation (Pedersen et al. 2012).....	56
Figure 2.3 : Schématisation de la méthode d’échantillonnage et d’agrégation de GraphSAGE.....	62
Figure 2.4 : Réseau JK à 3 couches. Chaque bloc de convolution représente une méthode arbitraire d’agrégation des voisins sur un graphe.....	63
Figure 3.1 : Etapes conceptuelles de la démarche proposée.....	73
Figure 3.2 : Cadre général de classification de graphes.....	74
Figure 3.3 : Architecture détaillée du GNN proposé pour la prédiction de la réception AIS.....	76
Figure 3.4 : Typologie des sources de données pour le jeu de données intégré de propagation AIS..	79
Figure 3.5 : Localisation géographique de la station de réception AIS du CRC Mines Paris - PSL à Sophia Antipolis sur le Campus Pierre Laffitte	80
Figure 3.6 : Schéma du système d’acquisition des données AIS du CRC.....	81
Figure 3.7 : Diagramme de directivité de l’antenne du CRC (reproduction autorisée par SIRIO).	81
Figure 3.8 : Courbe de gain de l’antenne du CRC (reproduction autorisée par SIRIO).....	81
Figure 3.9 : Architecture en médaillon et chaîne de traitement ETL pour le traitement des messages AIS. Elle traite les trames AIS brutes selon trois niveaux de raffinement : L1 (brutes), L2 (nettoyées, décodées et sous-échantillonnées) et L3 (validées, agrégées et filtrées).....	83
Figure 3.10 : Vue d’ensemble des chaînes de traitement ETL pour le traitement des données AIS. La première chaîne de traitement (haut) transforme les données brutes de la couche L1 vers la couche L2 à travers des étapes de nettoyage, décodage, filtrage et sous-échantillonnage. La seconde chaîne de traitement (bas) réalise les transformations avancées de la couche L2 vers la couche L3, incluant validation, réduction temporelle et filtrage, en préparation des tâches de modélisation	83
Figure 3.11 : Trame AIS en deux parties, au format 4.10, issue de la station du CRC.....	85
Figure 3.12 : Exemple de grille spatiale étiquetée pour un créneau horaire donné. Chaque cellule contient une étiquette : 1 (reçu localement), 0 (reçu uniquement via AISHub).....	91
Figure 3.13 : Comparaison entre différentes stratégies de modélisation de la propagation d’un émetteur (Tx) vers un récepteur fixe (Rx), en fonction des conditions de propagation (vert : bonnes, orange : moyennes, rouge : mauvaises, marron : effets du relief). À gauche : une propagation possible représentée par un vecteur direct naïf. À droite : des propagations complexes représentées par un graphe filtré, contraint par un cône directionnel réduisant la dispersion spatiale et mettant l’accent sur la sous-région la plus pertinente.....	94
Figure 3.14 : Chaîne de traitement pour la construction des graphes.....	95

Figure 4.1 : Vue d'ensemble de la chaîne de traitement pour la modélisation.	98
Figure 4.2 : Valeurs moyenne de N_0 données par (ITU-R 2019b) pour le mois de février.	100
Figure 4.3 : Grille spatiale étiquetée montrant les étiquettes 0 générées artificiellement Une grille de cellules de $0,25^\circ \times 4$ est surimprimée, assurant que chaque cellule contient au moins un point de données.	113
Figure 4.4 : Évolution de la perte d'entraînement en fonction du nombre d'itérations pour les modèles GCN, GAT et pour le GNN proposé (GraphSAGE avec Jumping Knowledge et agrégations multiples). Courbes lissées par moyenne mobile exponentielle pondérée par le temps.	115
Figure 4.5 : Répartition spatiale des points de données de test pour les étiquettes 1 (gauche) et 0 (droite).	116
Figure 4.6 : Répartition temporelles des points de données de test.	116
Figure 4.7 : Comparaison visuelle des prédictions du modèle pour différents scénarios de propagation. (a) correspond à une propagation à courte portée, avec certaines réceptions observées à l'arrière de la Corse. (b) illustre des cas de propagation avec réceptions situées plus à l'est, près des côtes italiennes. (c) et (d) correspondent à des scénarios de propagation à longue portée, les signaux émanant des côtes africaines, avec un faisceau plus étroit dans le cas de (d)	120
Figure 4.8 : Comparaison des prédictions de XGBoost, GAT et du GNN proposé sur les cas (a) et (c)	121
Figure 4.9 : Sous-régions maritimes spatialement cohérentes obtenues par un regroupement hiérarchique spatio-temporel des cellules de la grille ($k=6$).	123
Figure 4.10 : Cartes des faux positifs et des faux négatifs pour le modèle proposé.	125
Figure 4.11 : Cartes des faux positifs et des faux négatifs pour XGBoost.	126
Figure 4.12 : Courbes d'exactitude en fonction du nombre d'itérations d'apprentissage, sur le jeu de validation. Courbes lissées par moyenne mobile exponentielle pondérée par le temps.	129
Figure 4.13 : Prédictions pour le modèle proposé dans une tâche de classification multi-classe et entraîné par apprentissage par transfert, pour les cas (a) et (c)	130
Figure C.0.1 : 10 % des arêtes les plus importantes dans la prédiction d'un modèle attentionnel nœud émetteur mis en évidence.	138

Liste des tableaux

Tableau 1.1: Types de propagation en fonction des gradients de l'indice de réfraction	19
Tableau 1.2 : Phénomènes de propagation, effets sur la propagation de l'AIS, variables caractéristiques et milieux.....	25
Tableau 1.3 : Phénomènes de propagation et l'importance de leurs effets sur la propagation de l'AIS (rouge : Négligeable, orange : Peu significatif, Vert : Influence).....	25
Tableau 1.4 : Valeurs en Watt de la puissance en mode réduit et normal pour les dispositifs AIS de classes A et B.....	26
Tableau 1.5 : Classification des modèles physiques classiques de propagation selon leur catégorie.	28
Tableau 1.6 : Descriptions et recommandations ITU-R d'intérêts pour l'étude de la propagation de l'AIS	32
Tableau 1.7 : Modèles de propagations sur la plage de fréquence AIS, paramètres du système, paradigme et phénomènes pris en compte.	38
Tableau 1.8 : Limites des approches de modélisation de la propagation AIS.	39
Tableau 3.1 : Caractéristiques de l'antenne de la station CRC (documentation SIRIO).....	80
Tableau 3.2 : Pertes de signal dans le câble de la station CRC (documentation Messi & Paoloni). ...	81
Tableau 3.3 : Caractéristiques du récepteur AIS de la station du CRC (documentation Comar Systems).....	81
Tableau 4.1 : Résumé des modèles de référence et du modèle proposé pour la prédiction de la réception AIS.	99
Tableau 4.2 : Paramètres requis pour les prédictions à l'aide du modèle ITU-R P.2001, ainsi que les valeurs utilisées lors de sa mise en œuvre.	102
Tableau 4.3 : Entrées et sorties du module de construction des graphes.	105
Tableau 4.4 : Entrées et sorties du module de prétraitement des graphes.	107
Tableau 4.5 : Entrées et sorties du module d'entraînement, validation et de suivi.....	109
Tableau 4.6 : Entrées et sorties du module de test.	110
Tableau 4.7 : Entrées et sorties du module de prédiction.	111
Tableau 4.8 : Nombre de paramètres entraîables par modèles.	115
Tableau 4.9 : Espace de recherche des hyperparamètres exploré lors du réglage du modèle.	117
Tableau 4.10 : Résumé de la configuration d'apprentissage et des hyperparamètres utilisés pour l'optimisation du modèle.	117
Tableau 4.11 : Matrice de confusion pour le modèle GNN proposé sur le jeu de test. VN : Vrai Négatif, FP : Faux Positif, FN : Faux Négatif, VP : Vrai Positif.	118
Tableau 4.12 : Performances globales des modèles comparés sur le jeu de test (2602 échantillons).	122
Tableau 4.13 : Performances par classe et par cluster sur l'ensemble de test (meilleures valeurs en gras).....	124

Acronymes

ADS-B	Automatic Dependent Surveillance-Broadcast
ACP	Analyse en Composantes Principales
AIS	Automatic Identification System (Système d'Identification Automatique)
AISHub	Automatic Identification System Hub (plateforme collaborative de données AIS)
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CROSS	Centres Régionaux Opérationnels de Surveillance et de Sauvetage
CSV	Comma-Separated Values
DAM	Direction des Affaires Maritimes
DL	Deep Learning
ETL	Extract, Transform, Load
FFNN	Feed-Forward Neural Network
GAN	Generative Adversarial Network
GAT	Graph Attention Network
GCN	Graph Convolutional Network
GNN	Graph Neural Network
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GPU	Graphics Processing Unit
GraphSAGE	Graph Sample and AggreGatE
IA	Intelligence Artificielle
ITU-R	International Telecommunication Union - Radiocommunication Sector
JK	Jumping Knowledge
(Bi)-LSTM	(Bidirectional) Long Short-Term Memory
LoS	Line of Sight
MMSI	Maritime Mobile Service Identity
ML	Machine Learning
MLP	Multilayer Perceptron
MSA	Maritime Situational Awareness (Connaissance de la Situation Maritime)
NMEA	National Marine Electronics Association
NLoS	Non Line of Sight
OMI	Organisation Maritime Internationale
RL	Reinforcement Learning
RNN	Recurrent Neural Network
RSSI	Received Signal Strength Indicator
Rx	Receiver (Récepteur)
SIG	Système d'Information Géographique
SOLAS	Safety Of Life At Sea
SPATIONAV	Système de Surveillance Maritime Français
TL	Transfer Learning
Tx	Transmitter (Émetteur)
UDP	User Datagram Protocol
VHF	Very High Frequency
XGBoost	eXtreme Gradient Boosting

Introduction

Contexte applicatif

La connaissance de la situation maritime

La Connaissance de la situation maritime (MSA : Maritime Situational Awareness) est un enjeu primordial dans le domaine naval. Elle est définie par la (Commission of the European Communities 2009) comme étant une compréhension effective des activités associées au domaine maritime susceptibles d'avoir un impact sur la sécurité, la sûreté, l'économie ou l'environnement. Elle désigne plus largement la capacité à surveiller, comprendre et anticiper les événements se produisant dans l'environnement maritime. Cette capacité repose sur la collecte, l'intégration et l'analyse d'un ensemble hétérogène de données en temps réel, permettant aux autorités civiles et militaires, aux gestionnaires d'infrastructures portuaires ou encore aux opérateurs privés, de prendre des décisions éclairées en matière de navigation, de sécurité ou de gestion environnementale.

L'importance de la MSA s'est accrue au fil des dernières décennies, en réponse à l'augmentation significative du trafic maritime mondial, à la diversification des menaces (piraterie (Robitaille 2020), pêche illégale (Agnew et al. 2009 ; Vince et al. 2021), trafics illicites (Bruwer 2020), immigration clandestine (Koka et Veshi 2019) entre autres) et à la nécessité d'assurer la protection des ressources naturelles (United Nations, General Assembly 2017) et des infrastructures critiques en mer (Bueger et Liebetau 2023). La connaissance de la situation ne se limite pas à la simple détection des navires : elle implique également leur identification, classification, suivi, la prédiction des trajectoires, ainsi que la détection de comportements inhabituels ou suspects (Lee et al. 2019 ; Ribeiro et al. 2023). La MSA, en fournissant une vision globale et précise de l'espace maritime, constitue ainsi une condition préalable à la mise en œuvre de la sûreté maritime, laquelle s'intéresse plus particulièrement aux menaces intentionnelles et aux activités illicites.

La sûreté maritime concerne la protection contre les menaces intentionnelles d'origine humaine. Elle vise à prévenir et contrer les actes de piraterie, de terrorisme maritime, d'immigration clandestine, de trafics illicites, ainsi que la pêche illégale ou toute activité susceptible de porter atteinte à la souveraineté d'un État (Bueger 2015). Cette dimension implique une coordination étroite entre les acteurs navals et les services de l'État. Elle repose sur la surveillance, le contrôle, l'évaluation des risques et la capacité d'intervention. La sûreté maritime se situe donc dans une logique de défense et de prévention (République française, Premier ministre 2019).

La sécurité maritime se distingue de la sûreté en ce qu'elle concerne la prévention des accidents, incidents et catastrophes liées aux activités maritimes. Elle recouvre la sécurité des navires, des

équipages, des passagers et de l'environnement marin face aux aléas techniques, humains ou naturels (Formela et al. 2019). En France, elle s'applique notamment à la navigation commerciale, au transport maritime et à la plaisance. L'État exerce dans ce domaine des responsabilités en tant qu'État côtier, État du port et État du pavillon, assurées par la Direction des Affaires Maritimes (DAM) à travers la mise en œuvre des réglementations nationales et internationales (République française, Premier ministre 2019).

Pour garantir cette sûreté et cette sécurité, les Centres Régionaux Opérationnels de Surveillance et de Sauvetage (CROSS) disposent de différents outils de détection et de suivi (systèmes électro-optiques, radars, navires, drones). Parmi ceux-ci, le Système d'identification automatique (AIS : *Automatic Identification System*) constitue l'une des sources essentielles d'information (Direction interrégionale de la mer Manche Est – Mer du Nord 2024).

Le système d'identification automatique

L'AIS, a été introduit au début des années 2000 dans le cadre de la convention SOLAS (*Safety of Life at Sea*), sous l'impulsion de l'Organisation Maritime Internationale (OMI) dans le but d'améliorer la sécurité en mer, de prévenir les collisions et de renforcer la surveillance du trafic maritime. L'AIS est aujourd'hui un équipement obligatoire pour de nombreux navires, notamment ceux d'une jauge brute supérieure à 300 tonneaux engagés dans des voyages internationaux, ceux de plus de 500 tonneaux opérant hors du cadre international, ainsi que l'ensemble des navires à passagers, quelle que soit leur taille.

L'AIS fonctionne comme un système de diffusion, émettant à des intervalles, dépendant du statut de navigation, des messages contenant des informations essentielles : position GPS, cap, vitesse, identifiant unique du dispositif (MMSI : *Maritime Mobile Service Identity*), statut de navigation, dimensions du navire, port de destination, etc. Ces messages sont émis sur deux canaux à très hautes fréquences (VHF : *Very High Frequency*) dédiés (161,975 MHz et 162,025 MHz) en utilisant le protocole NMEA 0183 établi par la *National Marine Electronics Association*. Le système prend en charge jusqu'à 27 types de messages, regroupés en différentes catégories (ITU-R 2014).

Les émetteurs AIS se divisent en deux grandes classes :

- Classe A : obligatoire, elle est requise pour les navires commerciaux de grande taille ainsi que pour les navires à passagers. Ces équipements disposent d'une puissance d'émission élevée (jusqu'à 12,5 W).
- Classe B : volontaire, cette classe constitue une version allégée et moins coûteuse destinée aux petites embarcations pour lesquelles l'AIS n'est pas obligatoire. La puissance d'émission est réduite (jusqu'à 5 W).

Le rôle principal de l'AIS est de fournir une image partagée du trafic maritime aux navires eux-mêmes, aux stations côtières, et aux centres de surveillance. Il facilite la navigation en visibilité réduite, l'anticipation des croisements de routes, et l'assistance en cas de détresse. Mais ses usages se sont rapidement étendus, notamment grâce à l'accès croissant aux données en temps réel via des services commerciaux ou institutionnels. La réception des signaux AIS se fait par des stations terrestres, les navires, ou encore par des satellites depuis le développement de l'AIS spatial (Wahl et al. 2005 ; Carson-Jackson 2012).

Bien que conçu initialement comme un outil d'anticollision, le système AIS est devenu une source d'information critique pour l'observation globale de la situation maritime, grâce d'une part au développement de plateformes de visualisation en ligne en temps réel et de techniques de stockage collaboratif des données, et d'autre part à l'essor des constellations de satellites de petite taille permettant une couverture mondiale. Sa performance opérationnelle dépend cependant étroitement de la capacité réelle des récepteurs à capter les messages émis.

Exploitation des données AIS

Dans le cadre de la surveillance maritime, les données AIS émises par les navires sont aujourd'hui massivement exploitées. Elles permettent de construire une image dynamique du trafic, utilisable aussi bien en temps réel qu'en différé. Un large éventail d'acteurs s'en saisit : depuis les centres de surveillance maritime jusqu'aux plateformes collaboratives grand public telles que *MarineTraffic*¹, *VesselFinder*² ou *AISHub*³, qui collectent et diffusent les signaux reçus par des stations terrestres volontaires. Ces interfaces reposent sur des systèmes d'information géographique (SIG) pour représenter spatialement les positions, parfois enrichies de couches complémentaires (météorologiques, bathymétriques ou réglementaires) (Baylon et Santos 2013 ; Isbaex et al. 2025).

En France, la surveillance maritime repose sur des outils spécialisés, au premier rang desquels figure la plateforme nationale SPATIONAV. Considérée comme l'un des systèmes de surveillance maritime les plus complets actuellement en service, SPATIONAV V2 couvre l'ensemble du littoral métropolitain ainsi que plusieurs zones ultramarines. Il fédère et traite en temps réel les informations issues d'un large éventail de capteurs (radars côtiers, systèmes électro-optiques, stations AIS), mais aussi de données issues de capteurs embarqués (satellites, aériennes ou navales) (Vatin 2014 ; Airbus Defence and Space 2020). L'objectif est de fournir aux administrations concernées une vision globale et complète de la situation maritime, afin de renforcer la coordination des interventions et la prévention des menaces. Contrairement aux portails grand public de visualisation du trafic maritime, SPATIONAV se distingue par sa vocation opérationnelle : il permet non seulement de suivre les

¹ <https://www.marinetraffic.com/>

² <https://www.vesselfinder.com/>

³ <https://www.aishub.net/>

navires, mais aussi de détecter en temps réel des comportements suspects (République française 2012). Cette valorisation opérationnelle des données AIS ne doit cependant pas occulter une contrainte technique majeure : la réception des messages AIS dépend fortement de la couverture effective des stations terrestres ou satellitaires.

Une « zone de réception » définit l'enveloppe géographique dans laquelle un récepteur peut capter et décoder les signaux émis. Des fluctuations et des anomalies apparaissent aussi bien à l'intérieur qu'à l'extérieur de ces zones théoriques. En outre, malgré son déploiement massif, le système reste faiblement sécurisé. Enfin, comme le souligne (Iphar 2017), l'AIS demeure vulnérable à diverses menaces cyber, susceptibles de compromettre la connaissance de la situation maritime. L'exploitation des données AIS révèle ainsi des limites structurelles, tant du point de vue de la fiabilité technique que de la sécurité intrinsèque. Ces fragilités renforcent l'importance de la détection et de l'interprétation des anomalies maritimes, désormais au cœur des enjeux de surveillance et de sûreté en mer.

Anomalies dans le domaine maritime

Une anomalie représente une observation qui s'écarte tellement des autres observations au point de remettre en cause le mécanisme par lequel elle aurait été générée (Hawkins 1980 ; Aggarwal 2017). Dans le domaine maritime, les anomalies sont des événements rares qui diffèrent d'un comportement habituel. De telles observations peuvent par exemple être les prémisses d'une activité illicite ou d'une situation dangereuse. Les anomalies maritimes peuvent être regroupées en différentes catégories (Riveiro et al. 2018) : anomalies de position (déviations de la route, trajectoire anormale, retard d'arrivée, rapprochement d'un autre navire, entrée dans une zone restreinte (Lane et al. 2010 ; Wolsing et al. 2022)), anomalies contextuelles (comportement incohérent avec le contexte), anomalies cinématiques (comportement anormal lié aux paramètres de mouvement et aux manœuvres), anomalies complexes (anomalies dont la détection nécessite une fusion de plusieurs détecteurs) ou anomalies liées à la donnée. Dans cette dernière catégorie (Riveiro et al. 2018) incluent les anomalies liées à l'absence de données, qui apparaissent lorsqu'une station ne reçoit pas, ou reçoit de manière incohérente, des messages AIS en provenance d'une zone où des émissions sont normalement attendues. Ce type d'anomalie représente un point critique de la chaîne d'observation maritime, puisqu'une rupture ou une pollution du signal compromet directement la fiabilité de l'image du trafic.

Anomalies liées à la donnée

Une absence de données peut être due à différentes causes, combinées ou individuelles. Elle est donc, selon la classification de (Riveiro et al. 2018) à l'interface entre les anomalies complexes et les anomalies liées à la donnée. D'un point de vue technique, des défaillances matérielles, telles qu'un mauvais câblage, une mauvaise calibration, un défaut matériel ou une mauvaise position de l'antenne

émettrice à bord du navire peuvent conduire à une mauvaise propagation des messages AIS (Last et al. 2015 ; Laso et al. 2017). Des anomalies liées à la donnée peuvent également provenir de l'opérateur, de manière volontaire ou involontaire. Une mauvaise initialisation du système GNSS peut conduire à la transmission de positions erronées (Harati-Mokhtari et al. 2007). Une attribution incorrecte de l'identifiant unique MMSI du transpondeur AIS, lorsqu'elle entraîne une duplication avec un autre navire, peut générer des situations d'ubiquité et un partage de coordonnées géographiques (Mazzarella et al. 2013). Enfin, l'extinction volontaire du dispositif AIS (Mazzarella et al. 2017) ou son passage dans un mode de puissance réduit (Kelly 2022) constituent d'autres causes possibles d'anomalies liées à la donnée. Au-delà du caractère opérationnel, une anomalie peut également résulter de la falsification des coordonnées GPS, réalisée par des attaques d'usurpation (*spoofing*) ou de détournement (*hijacking*) du signal (Balduzzi et al. 2014 ; Iphar 2017 ; Iphar et al. 2020). (Riveiro et al. 2018) définissent cette dernière cause d'anomalie comme une anomalie complexe, mais elle peut néanmoins se traduire par des anomalies dans les données effectivement reçues.

Par ailleurs, l'AIS est sensible à diverses interférences pouvant affecter la réception des signaux. Celles-ci incluent les collisions de données (Smith et Nelson 2009), l'obstruction physique dues au relief ou à des obstacles (Salmon et al. 2016) ainsi que les conditions météorologiques défavorables (Bruin 2016). Ces interférences peuvent conduire à une mauvaise réception de l'AIS et donc à une absence de données.

En effet, la réception du signal AIS est conditionnée par les propriétés physiques de sa propagation en bande VHF. À cette fréquence, les conditions météorologiques, notamment les événements se produisant dans la troposphère ainsi que les variations de relief, exercent une influence directe sur la propagation du signal (Qinjiang et al. 1998 ; Tang, Cha, Wei, Tian, et al. 2019 ; Rautiainen et al. 2025). Parmi eux, on trouve les effets liés au terrain comme la diffraction (Mazzarella et al. 2016 ; Wang et Zhang 2017), mais aussi les conditions atmosphériques, responsables de phénomènes tels que la diffusion troposphérique, la réfraction, ou encore la propagation par conduits (*ducting*), induits par des variations de la réfractivité de l'air (Bean et Dutton 1966 ; Cornélis 2005 ; Mentès et Kaymaz 2007 ; Grandjean 2016 ; Zhang et al. 2016). La propagation par trajets multiples, due aux réflexions, en particulier les réflexions diffuses, joue également un rôle significatif dans la qualité de réception (Barrick 1971 ; Qinjiang et al. 1998 ; IEEE 2019).

Les phénomènes apparaissant avec une probabilité faible (Caillet 2015) ou dans une atmosphère non standard (IEEE 2019), sont regroupés sous le terme de « propagations anormales », et leur prise en compte est essentielle pour comprendre et déterminer avec précision les zones de réception de l'AIS.

Problématique de recherche

Approches traditionnelles de modélisation de la réception de l'AIS

Les limitations évoquées précédemment soulignent un enjeu fondamental pour la surveillance maritime : la nécessité de modéliser de manière précise la propagation des signaux AIS afin de mieux caractériser les zones de réception et d'améliorer l'interprétation des données reçues. En effet, l'aire de réception ne se réduit pas à un simple rayon théorique autour d'un récepteur, elle résulte d'un processus complexe influencé par la topographie, les conditions atmosphériques ainsi que les caractéristiques des systèmes émetteurs et récepteurs. La modélisation de la propagation des ondes radio est vaste, avec des contributions majeures dans des domaines tels que les télécommunications, le radar ou la radiodiffusion. Trois grandes catégories de modèles ont émergé, chacune offrant une approche spécifique et des compromis distincts.

Modélisations déterministes

Fondés sur la théorie électromagnétique, ces modèles visent à fournir des prédictions précises à partir de la résolution d'équations électromagnétiques. On y retrouve par exemple la méthode des équations paraboliques (Barrios 1994 ; Sirkova et al. 2006 ; Choi et al. 2020 ; Ozgun et al. 2020), les modèles en visibilité directe (LoS : *Line of Sight*) (Strohbehn 1968 ; Stephansen 1981), et la perte de propagation en espace libre (FSPL : *Free-Space Path Loss*) (Friis 1946 ; Saunders et Aragón-Zavala 2024 ; Rappaport 2024).

Dans le cadre de la réception AIS, (Kelly 2022) développe une méthode de détection de réduction volontaire de puissance d'émission, en s'appuyant sur une modélisation mathématique du niveau de puissance en réception (RSSI : *Received Signal Strength Indicator*) attendu en fonction de la distance. Leur approche compare les signaux reçus aux valeurs prédites par un modèle électromagnétique simplifié. Cette démarche s'inscrit pleinement dans la logique des modèles déterministes, en s'appuyant sur des lois de propagation bien définies pour inférer des comportements anormaux. Si les approches déterministes reposent sur des fondements physiques, leur application pratique reste souvent complexe et coûteuse. Pour pallier ces limites, la communauté scientifique s'est tournée vers des modèles empiriques et statistiques, construits à partir d'observations de terrain et adaptés à des contextes opérationnels variés.

Modélisations empiriques et statistiques

Les modèles empiriques sont construits à partir d'observations de terrain et reposent sur des relations statistiques établies à partir de campagnes de mesures. Ils permettent d'estimer les pertes de propagation ou la couverture en fonction de variables telles que la distance, la fréquence ou la hauteur d'antenne. Parmi les modèles classiques, on retrouve ITU-R P.1546 (ITU-R 2019d), Okumura (Okumura 1968 ; Hata 1980) ou encore Egli (Egli 1957).

Dans le contexte de la couverture AIS, (Hammond et Peters 2012) proposent une méthode statistique originale fondée sur un traitement bayésien des transmissions reçues. Leur approche consiste à estimer de manière probabiliste la position des messages manquants à partir des rapports effectivement reçus, et à inférer une carte de couverture sur une grille spatiale. L'algorithme repose sur un modèle binomial de réception, enrichi par des hypothèses sur la distribution des messages non observés. Testée sur des données simulées, la méthode montre une réduction de l'erreur quadratique moyenne par rapport à des approches antérieures. Une autre contribution notable est celle de (Salmon et al. 2016) définissant les « trous noirs » dans les flux de données maritimes. Cette approche permet d'identifier des zones où aucun signal AIS n'est reçu, ce qui peut signaler des lacunes dans la couverture. En combinant des données historiques et des flux en temps réel, le modèle offre une estimation dynamique des zones couvertes et non couvertes. Pour concilier précision théorique et applicabilité opérationnelle, des modèles semi-empiriques ont été développés, combinant bases physiques et ajustements empiriques.

Modélisations semi-empiriques

Les modèles semi-empiriques (ou semi-déterministes) combinent des fondements théoriques à des ajustements empiriques pour en faciliter l'usage opérationnel. On peut citer les modèles ITU-R P.452 (ITU-R 2023a), P.1812 (ITU-R 2023b), P.2001 (ITU-R 2023c), COST 231-Hata (European Commission et al. 1999) ou encore le modèle de Longley-Rice (ITM) (Longley et Rice 1968). Ces approches sont couramment utilisées pour estimer la perte de propagation en environnement terrestre, côtier ou maritime. Dans le contexte spécifique de la surveillance maritime, (Bruin 2016) propose une approche semi-empirique visant à modéliser l'impact des conditions atmosphériques sur la portée des signaux AIS et radar dans la zone économique exclusive néerlandaise. L'étude montre que la présence de *ducts* atmosphériques (*ducts* de surface, *ducts* élevés, *ducts* d'évaporation) c'est-à-dire des couches atmosphériques qui piègent les ondes radio et modifient leur propagation, entraîne des variations importantes de la couverture radio. L'auteur s'appuie sur des données météorologiques de prévision numérique pour produire des cartes de couverture journalières, démontrant ainsi le potentiel opérationnel d'un tel modèle pour la gestion dynamique de la surveillance maritime. Cette démarche illustre comment les effets environnementaux peuvent être intégrés à la modélisation de la propagation à l'aide d'entrées météorologiques, dans une logique à la fois empirique et opérationnelle.

Limites des approches traditionnelles

Les modèles physiques déterministes, bien qu'ils offrent une grande précision dans la modélisation de la propagation des ondes, souffrent d'une certaine rigidité, caractérisée par une forte dépendance aux données d'entrée, une rigidité paramétrique, un coût computationnel élevé pour des zones étendues et une difficulté à intégrer des phénomènes stochastiques tels que la variabilité météorologique ou les interférences. Celle-ci tient au fait qu'ils reposent sur des formulations

analytiques et des hypothèses idéalisées, qui les rendent peu flexibles face à la variabilité des environnements réels. Leur application dans des environnements spécifiques demeure souvent complexe, car ils requièrent un accès à des données fines dont la disponibilité et la précision sont limitées en pratique. À l'inverse, les approches empiriques ou semi-déterministes sont appréciées pour leur simplicité de mise en œuvre, leur nombre réduit de paramètres et leur coût à l'inférence. Cependant, les modèles existants restent généralement génériques ou développés pour un cas d'utilisation ou un site spécifique (Seretis et Sarris 2021 ; Vasudevan et Yuksel 2024 ; Bocus et Lodhi 2025). Ainsi, au vu des relations complexes entre paramètres topographiques, environnementaux et opérationnels, des spécificités propres à la propagation AIS et du caractère intrinsèquement dynamique des conditions atmosphériques, un verrou scientifique réside dans la conception d'un modèle capable de caractériser les zones de réception de l'AIS en tenant compte de la variabilité spatio-temporelle des phénomènes de propagation, d'intégrer des sources de données hétérogènes (mesures, données météorologiques, informations contextuelles) et de généraliser ses prédictions à différents environnements opérationnels.

Dans ce contexte, une approche dirigée par les données et mobilisant des méthodes d'apprentissage automatique apparaît comme une alternative pertinente aux modèles traditionnels. Elle permet en effet d'apprendre directement à partir d'observations, de tirer parti de la richesse des données disponibles et d'intégrer des signaux hétérogènes, tout en s'affranchissant d'une modélisation physique explicite.

Apprentissage automatique et approches dirigées par les données pour la modélisation de la réception AIS

Les approches dirigées par les données reposent sur l'exploitation de grands volumes d'observations empiriques pour construire des modèles prédictifs sans recourir directement à des équations physiques explicites. Dans le cadre de l'apprentissage automatique, cela implique l'usage d'algorithmes capables d'extraire des structures latentes, des corrélations ou des dynamiques à partir de jeux de données. Contrairement aux modèles théoriques classiques, ces approches se distinguent par leur capacité à s'adapter aux données disponibles et à intégrer des signaux bruités ou incomplets (C. Zhang et al. 2017). Ce paradigme est particulièrement pertinent dans le contexte de phénomènes partiellement observés ou difficiles à modéliser, comme les variations atmosphériques, les interférences radio ou les effets du relief sur la propagation du signal AIS (Rautiainen et al. 2025). Dans le cas de phénomènes partiellement observés, l'impact sur la réception est mesuré indirectement (par exemple via la présence et l'absence de messages AIS (Mazzarella et al. 2016) ou la mesure de la puissance reçue (Alincourt et al. 2016)), sans disposer d'observations directes et complètes de toutes les variables physiques sous-jacentes. Ce caractère partiellement observé des phénomènes rend difficile l'usage exclusif de modèles physiques, qui supposent une connaissance complète des variables

en jeu. À l'inverse, une approche dirigée par les données est bien adaptée : elle permet d'exploiter les observations disponibles, même indirectes ou bruitées, et d'apprendre des corrélations complexes entre conditions environnementales et performance de réception, sans nécessiter une modélisation explicite de l'ensemble des phénomènes physiques (Seretis et Sarris 2021). Le recours à des méthodes d'apprentissage supervisé ou non supervisé permet ainsi de capturer des comportements ou des régularités difficilement exprimables par des modèles physiques classiques.

Dans le cas de l'AIS, (Mazzarella et al. 2017) proposent une approche semi-empirique combinant des modèles classiques de propagation (FSPL, le modèle à deux rayons, ITU-R P.1546 (ITU-R 2019d)) avec un algorithme d'apprentissage automatique non supervisé. Le modèle repose sur un *One-Class* SVM entraîné à partir des écarts entre RSSI observés et théoriques, afin de détecter les comportements anormaux (désactivation de transpondeur, réduction de puissance). Cette approche illustre l'intérêt de coupler des fondements physiques avec une détection statistique dirigée par les données. Dans une logique similaire, (Singh et Heymann 2020) développent un système de détection d'anomalies fondé sur un réseau de neurones artificiels, entraîné à partir de données AIS réelles. Le modèle est conçu pour distinguer entre des interruptions naturelles et des désactivations intentionnelles. Leur approche s'appuie uniquement sur les informations dynamiques issues des messages AIS (position, vitesse, cap, distance au récepteur), sans recourir à une modélisation explicite du canal.

(Vanche et al. 2022) proposent une méthodologie de prédiction de la couverture d'une station AIS côtière fondée sur l'exploitation conjointe de cartes de prévision de propagation troposphérique et d'algorithmes d'apprentissage supervisé. Leur approche vise à dépasser les limites de la simple visibilité directe en intégrant des conditions atmosphériques affectant la propagation au-delà de l'horizon radioélectrique. Les auteurs utilisent des cartes issues de modèles météorologiques (NOAA/NCEP 2015) décrivant l'indice de réfraction atmosphérique, qu'ils transforment en indicateurs de qualité de propagation (Grandjean 2016). Ces variables, croisées avec les positions de navires, alimentent un modèle de classification supervisée basé sur XGBoost (*eXtreme Gradient Boosting*), entraîné sur six mois de données AIS collectées par une station expérimentale déployée par le CRC sur le Campus Pierre Laffitte de Mines Paris à Sophia Antipolis. Le modèle permet de prédire, pour chaque zone géographique, la probabilité de réception d'un message AIS. Les résultats montrent un taux de classification correcte de 77,1 % et une précision de 80,2 % sur la classe « réception », tout en produisant des cartes de couverture exploitables opérationnellement. Cette approche illustre le potentiel du couplage entre données atmosphériques prévisionnelles et méthodes d'apprentissage automatique pour améliorer la cartographie dynamique des zones de réception de l'AIS.

Une approche complémentaire est proposée par (Pohontu et al. 2024) qui introduisent un modèle d'apprentissage profond pour prédire la portée de détection de stations AIS à partir de données

météorologiques historiques. Leur méthodologie repose sur des réseaux neuronaux, un perceptron multicouche (MLP : *Multilayer Perceptron*) et un réseau récurrent à mémoire court et long terme (LSTM : *Long Short-Term Memory*), entraînés à partir de données collectées sur 72 heures, incluant à la fois les conditions atmosphériques (température, humidité, courant, vent) et des étiquettes liées à la réception du signal AIS. Ils montrent que les modèles LSTM atteignent des performances proches de 89 % de précision. L'intérêt de cette approche réside dans sa capacité à capturer des corrélations complexes entre conditions environnementales et performance radio.

Malgré la popularité croissante des approches dirigées par les données pour la modélisation de la réception AIS, peu de travaux intègrent des représentations géométriques capables de refléter les interactions spatiales non linéaires entre stations, navires et environnement. Or, dans le domaine des télécommunications, de telles représentations ont déjà été explorées, notamment à l'aide de réseaux de neurones en graphes (GNN : *Graph Neural Networks*) qui permettent de modéliser les relations entre émetteurs, récepteurs et environnement pour améliorer la prédiction de la couverture ou la gestion des interférences (S. He et al. 2021). Ce manque d'application au domaine de l'AIS souligne une opportunité de recherche pour adapter et formaliser une approche basée sur ces méthodes pour la modélisation des zones de réception de l'AIS. Si les travaux existants montrent le potentiel des approches dirigées par les données, ils révèlent également plusieurs défis méthodologiques et pratiques. Ces verrous doivent être levés pour concevoir des modèles robustes et généralisables.

Défis liés aux données et problématique de recherche

Si les approches dirigées par les données apparaissent particulièrement pertinentes pour la modélisation de la réception AIS, leur efficacité dépend fortement de la capacité à surmonter plusieurs défis inhérents aux données : qualité, structuration spatiale et hétérogénéité des sources.

Qualité et intégrité des données AIS

Plusieurs études soulignent que les données AIS brutes contiennent fréquemment du bruit, des erreurs et des lacunes, en raison par exemple d'informations statiques mal renseignées ou de défaillances d'équipement (Emmens et al. 2021 ; W. He et al. 2021). Un prétraitement rigoureux des données est donc indispensable, incluant des étapes de nettoyage, de filtrage, d'imputation et de validation croisée avec d'autres sources (D. Zhang et al. 2017 ; Hartawan et al. 2021 ; Zhang et Li 2022 ; Lv et al. 2023).

Complexité de la structuration spatiale

La propagation radioélectrique, notamment en environnement maritime, ne suit pas une géométrie simple, l'onde se propage sur un front d'onde ellipsoïde et non en suivant une ligne droite dans l'espace. Le trajet du signal est également affecté par des phénomènes non linéaires (Bocus et

Lodhi 2025). Des travaux récents en apprentissage automatique suggèrent qu'il faut recourir à des représentations non euclidiennes pour modéliser des données à structure complexe (Bronstein et al. 2017). Il devient alors nécessaire d'adopter des structures de données capables de modéliser des interactions complexes. Par exemple, la modélisation de la propagation radioélectrique peut tirer parti d'une représentation en graphe où les nœuds correspondent à des positions géographiques et les arêtes représentent les relations spatiales ou les trajets possibles entre ces positions (Liu et al. 2023 ; Bufort et al. 2023).

Hétérogénéité et fusion des sources de données

La modélisation de la réception AIS nécessite d'intégrer de multiples sources hétérogènes (données topographiques, conditions météorologiques, caractéristiques techniques des stations). Ces sources varient en résolution spatiale, temporelle et en précision, posant un défi de fusion multimodale. Des approches multimodales traitant simultanément des données d'origines variées se révèlent nécessaires pour exploiter pleinement ces informations (Koldasbayeva et al. 2024). En pratique, il faut également normaliser et aligner les données de différentes modalités pour atténuer les écarts d'échelle ou d'unité entre elles, cela peut passer par des techniques de normalisation qui réduisent les différences entre dimensions tout en préservant les caractéristiques propres à chaque source (Xiao et al. 2024).

Les approches dirigées par les données apparaissent alors comme particulièrement pertinentes pour modéliser la réception de l'AIS. À partir de ces constats, la problématique de recherche peut être formulée comme suit :

Comment modéliser de manière robuste, précise et généralisable les zones de réception de l'AIS en environnement maritime ?

Cette problématique implique de développer des modèles capables d'exploiter la richesse des données disponibles, de structurer l'espace d'étude et d'apprendre les dynamiques de propagation de manière inductive et généralisable. C'est dans cette perspective que s'inscrit cette thèse.

Postulats, hypothèses et objectifs de recherche

Postulats de départ

La présente recherche repose sur un ensemble de postulats liés à la nature des signaux AIS, à leur comportement en environnement maritime et aux possibilités offertes par l'apprentissage automatique pour en modéliser la propagation. Les postulats suivants motivent la démarche :

- Les messages AIS subissent des interférences dues à des facteurs environnementaux, ce qui entraîne une couverture spatialement hétérogène et fluctuante dans le temps.

- Les modèles physiques traditionnels atteignent des limites, rendant difficile l'intégration de données spécifiques ou multimodales, contraignant leur expressivité.

Hypothèses de travail et objectif de la thèse

Sur la base de ces postulats, plusieurs hypothèses sont formulées afin de guider la démarche :

- **Hypothèse 1** : Il est possible de développer un modèle de réception de l'AIS performant à partir de données empiriques, sans recourir à des modèles physiques explicites, en exploitant des descripteurs pertinents.
- **Hypothèse 2** : L'intégration multimodale de données hétérogènes permet d'améliorer la capacité prédictive d'un modèle de réception.
- **Hypothèse 3** : L'utilisation de structures de représentation adaptées, notamment les graphes intégrant les relations entre les entités, améliore la précision des prédictions de réception par rapport à des méthodes classiques.

L'objectif de cette recherche est de développer une méthodologie basée sur l'apprentissage profond, intégrant les conditions environnementales, afin de modéliser la réception de l'AIS et de prédire dynamiquement les zones de réception.

Organisation de la thèse

Cette thèse est organisée en quatre chapitres principaux :

- **Chapitre 1** : Ce chapitre s'intéresse à la propagation des signaux AIS en environnement maritime. Il présente les caractéristiques physiques de cette propagation et examine les modèles classiques ainsi que les paramètres qui conditionnent les divers mécanismes de propagation. Un état de l'art met en lumière les limites de ces approches.
- **Chapitre 2** : Ce chapitre explore via un état de l'art les approches d'apprentissage automatique et profond appliquées à la modélisation de la propagation des ondes sur des bandes semblables à celles des très hautes fréquences. Il introduit les fondements de l'apprentissage automatique (ML : *Machine Learning*), de l'apprentissage profond (DL : *Deep Learning*) et des réseaux de neurones en graphes, et discute leur pertinence pour traiter les données inhérentes au problème de la modélisation de la propagation de l'AIS.
- **Chapitre 3** : Ce chapitre constitue la proposition méthodologique de la thèse. Il décrit une approche dirigée par les données pour modéliser les zones de réception de l'AIS, en s'appuyant sur une représentation de l'environnement sous forme de graphes enrichis par des données environnementales et contextuelles. La chaîne de traitement complète y est présentée.

- **Chapitre 4** : Ce dernier chapitre est dédié à l'évaluation expérimentale de la méthodologie proposée. Il détaille l'implémentation technique ainsi que les contraintes rencontrées. Les performances du modèle sont analysées selon plusieurs critères, comparées à des approches classiques basées sur la physique et sur l'apprentissage automatique. L'inférence sur des cas réels permet de valider la pertinence de l'approche et d'identifier des axes d'amélioration.

Chapitre 1 : Propagation des signaux AIS en environnement maritime

L'AIS constitue aujourd'hui un outil incontournable pour la connaissance de la situation maritime en permettant notamment l'échange automatisé d'informations de positions. Toutefois, la fiabilité et la portée de ce système dépendent étroitement des conditions de propagation radioélectrique dans lesquelles les signaux sont émis et reçus. En environnement maritime, ces conditions résultent d'une interaction entre phénomènes atmosphériques, caractéristiques de la surface de la mer, topographie côtière et paramètres propres aux équipements AIS. Comprendre et modéliser la propagation de ces signaux constitue donc un préalable indispensable pour prédire les zones de réception de l'AIS.

Ce premier chapitre est consacré à l'étude de la propagation des signaux AIS en environnement maritime. Dans un premier temps, les phénomènes physiques fondamentaux, tels que la réfraction, la diffraction, la diffusion, l'absorption et la réflexion, sont décrits, ainsi que les paramètres systèmes (positionnement des antennes, puissance d'émission, gain et directivité) qui conditionnent la qualité de la liaison radio. Dans un second temps, les principaux modèles physiques de propagation, qu'ils soient analytiques, empiriques ou semi-déterministes, sont présentés et appliqués au cas de l'AIS. Cette revue intègre les modèles de référence proposés par l'Union Internationale des Télécommunications (ITU), largement utilisés pour l'évaluation et la prévision de la couverture radio.

L'objectif de ce chapitre est double : établir une compréhension des phénomènes qui dirigent la propagation des signaux AIS en extrayant leurs variables caractéristiques et fournir un cadre théorique de référence pour comparer et évaluer, dans les chapitres suivants, les approches fondées sur l'apprentissage automatique et profond.

1.1 Phénomènes et paramètres de la propagation de l'AIS

La compréhension de la propagation des signaux radioélectriques en environnement maritime requiert l'examen des phénomènes physiques élémentaires qui régissent la transmission des ondes électromagnétiques. Dans le cas de l'AIS, qui opère dans la bande VHF, ces phénomènes présentent des manifestations et des contraintes spécifiques, liées à la fréquence d'émission et aux conditions particulières de l'environnement maritime, ainsi que des paramètres propres aux systèmes de communication qui en conditionnent la portée. Ces phénomènes interviennent simultanément et avec des intensités variables, modulées par les caractéristiques atmosphériques et par l'environnement maritime. Leur influence se combine aux spécificités techniques des dispositifs AIS, qui déterminent la performance opérationnelle des communications.

L'objectif de cette sous-partie est d'analyser successivement ces phénomènes et paramètres, en explicitant leur origine physique, leurs conditions d'occurrence et leur rôle dans la propagation de l'AIS. Cette mise en perspective constitue une étape pour aborder la modélisation physique, car elle permet d'identifier les facteurs clés à intégrer, notamment lors de la définition des variables caractéristiques des phénomènes.

1.1.1 Phénomènes de propagation

La propagation des signaux AIS résulte d'une combinaison de phénomènes physiques et de paramètres propres aux systèmes de communication. Ces phénomènes : réflexion, réfraction, diffraction, diffusion, absorption, interviennent dans des proportions variables selon l'environnement et les conditions atmosphériques. Leur influence est modulée par les caractéristiques du système AIS lui-même, telles que la puissance d'émission, le gain et la directivité de l'antenne ou encore sa position par rapport à la surface de la mer. L'étude détaillée de ces phénomènes est essentielle pour :

- Caractériser les portées théoriques et pratiques dans des contextes variés, qu'ils soient en ligne de vue (LoS : *Line of Sight*) ou hors ligne de vue (NLoS : *Non Line of Sight*).
- Identifier les conditions favorisant ou dégradant la réception, qu'elles soient liées à la physique de l'onde ou aux caractéristiques du système (positionnement, gain, puissance d'émission).
- Fournir des bases à la modélisation et à la prévision de la réception de l'AIS.

La Figure 1.1 présente un schéma synthétisant les principaux phénomènes physiques susceptibles d'affecter la propagation des signaux AIS. Chaque phénomène y est positionné par rapport au milieu dans lequel il intervient. Ce schéma met en évidence les interactions possibles entre phénomènes, par exemple la combinaison d'une réflexion sur la mer et d'une propagation directe, ainsi que leur impact potentiel sur la portée de la liaison radio (propagation transhorizon etc.). La section qui suit décrit successivement ces phénomènes, en précisant leurs origines physiques, leurs conditions d'occurrence et leurs impacts potentiels sur les transmissions AIS, avant de présenter les paramètres systèmes qui influencent directement leurs manifestations.

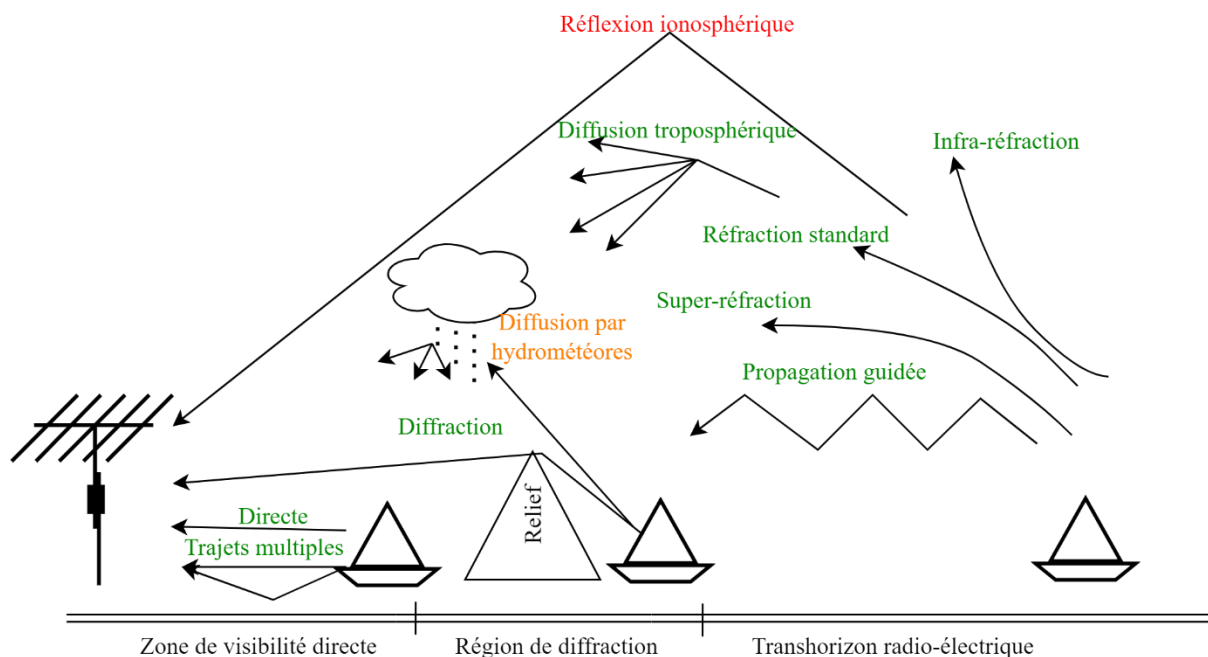


Figure 1.1 : Vue schématique des phénomènes de propagation de l'AIS et importance de leurs effets sur la propagation (rouge : Négligeable, orange : Peu significatif, Vert : Influence).

1.1.1.1 Réfraction

Le phénomène de réfraction correspond à une modification de la direction de propagation d'une onde lorsqu'elle traverse une interface (ou dioptre) entre deux milieux d'indices de réfraction différents. Ce changement de direction résulte de la variation de la vitesse de phase de l'onde dans chacun des milieux. Ce phénomène est régi par la loi de Snell-Descartes, qui s'écrit :

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

Où n_1, n_2 sont les indices de réfraction du premier et second milieu, θ_1, θ_2 l'angle d'incidence et de réfraction par rapport à la normale locale du dioptre dans le premier et second milieu.

1.1.1.1.1 Réfraction atmosphérique

Dans le contexte de la propagation des ondes radioélectriques dans l'atmosphère, cette loi s'applique de manière continue : l'indice de réfraction n n'est pas constant mais varie progressivement avec l'altitude, en fonction de paramètres tels que la pression, la température et l'humidité. Cette variation induit une courbure progressive du trajet des ondes, un phénomène souvent qualifié de réfraction atmosphérique. L'étude de l'indice de réfraction atmosphérique constitue la pierre angulaire de la modélisation de la propagation des ondes électromagnétiques dans les couches basses de l'atmosphère. Étroitement lié aux conditions météorologiques locales, cet indice permet d'expliquer une grande partie des communications qui dépassent l'horizon radioélectrique (Bean et Dutton 1966), c'est-à-dire l'ensemble des points où les trajets radioélectriques issus d'une source ponctuelle sont

tangents à la surface de la Terre, définissant la distance au-delà de laquelle la propagation directe des ondes n'est plus possible (IEEE 1993 ; ITU-R 2019a).

Sur le plan microscopique, une molécule dépourvue de moment dipolaire permanent, c'est-à-dire sans séparation fixe entre le centre des charges positives (noyaux) et le centre des charges négatives (électrons), peut néanmoins être polarisée sous l'effet d'un champ électrique. En partant de la formulation de la polarisation diélectrique, (Bean et Dutton 1966) montrent que, pour des fréquences inférieures à 100 GHz, et en tenant compte des composants polaires (la vapeur d'eau (H₂O) et l'ozone (O₃)) et non polaires (l'azote (N₂), l'oxygène (O₂) et le dioxyde de carbone (CO₂)) présents dans la troposphère, l'indice de réfraction de l'air n ou l'indice de réfraction atmosphérique N (coïndice N) peuvent s'exprimer comme :

$$N = (n - 1) \times 10^6 = \frac{K_1 P_d}{T} + \frac{K_2 e}{T} + \frac{K_3 e}{T^2} + \frac{K_4 P_c}{T}$$

Où K_1, K_2, K_3, K_4 sont des constantes empiriques, déterminées expérimentalement, P_d la pression de l'air sec en hPa, P_c la pression partielle du dioxyde de carbone en hPa, e la Pression partielle de la vapeur d'eau en hPa et T la température en K. Les valeurs habituellement retenues pour ces constantes ont été obtenues expérimentalement par (Smith et Weintraub 1953) et mènent à une expression simplifiée de l'indice de réfraction atmosphérique :

$$N = \frac{77,6}{T} \left(P + \frac{4810e}{T} \right)$$

Où P est la pression atmosphérique en hPa.

1.1.1.1.2 Propagation guidée et types de réfractions atmosphériques

La « propagation par conduits », également appelée « propagation guidée », désigne un phénomène dans lequel les variations verticales de l'indice de réfraction atmosphérique contraignent les ondes radioélectriques à suivre des trajectoires quasi-parallèles à la surface de la Terre. Ce phénomène, connu sous le nom de *ducting* en anglais, permet des communications à longue portée dans les bandes VHF et au-delà de l'horizon radioélectrique (Cornélis 2005 ; Grandjean 2016).

(Bean et Dutton 1966) rappellent la relation fondamentale qui relie la courbure effective du rayon radioélectrique à celle de la Terre. Cette relation s'écrit :

$$\frac{1}{a} + \frac{1}{n} \frac{dn}{dz} \cos(\theta) = \frac{1}{ka} \quad (1)$$

Où a est le rayon de la Terre en mètres, k un facteur, n l'indice de réfraction de l'air, θ l'angle d'incidence du rayon par rapport à la verticale et $\frac{dn}{dz}$ le gradient d'indice de réfraction. Dans cette équation, le terme $\frac{1}{a}$ représente la courbure terrestre, tandis que $\frac{1}{n} \frac{dn}{dz} \cos(\theta)$ modélise la courbure

induite par les variations verticales de l'indice de l'air. Le second membre, $\frac{1}{ka}$, correspond à la courbure effective du rayon radio, utilisée dans le modèle de Terre équivalente à rayon effectif $a_e = ka$.

Afin de simplifier l'interprétation de ces effets en environnement terrestre, la littérature introduit l'indice de réfraction modifié $M(z)$, défini par :

$$M(z) = N(z) + \frac{z}{a} 10^6$$

Où z est une altitude en mètres et $N(z) = (n - 1)10^6$ est l'indice de réfraction atmosphérique. L'indice $M(z)$ permet de transposer le problème dans un cadre géométrique plus intuitif, notamment en ramenant les effets de courbure à une atmosphère équivalente plane. En injectant $a = 6370$ km (rayon terrestre moyen) dans l'équation de l'indice de réfraction modifié, on obtient l'approximation classique :

$$M(z) = N(z) + 157z$$

Où z est exprimé en mètres. Cette formulation est largement utilisée pour caractériser les conditions de propagation dans les modèles opérationnels.

Selon le gradient vertical de $N(z)$ ou de $M(z)$, différents régimes de propagation peuvent être distingués. (Caillet 2015) fournit des valeurs de référence permettant de classer les types de réfraction en fonction de ces gradients. Ces classifications sont synthétisées dans le Tableau 1.1.

Type de propagation	$\frac{dn}{dz}$	$\frac{dM}{dz}$
Propagation guidée	< -157	< 0
Super-réfraction	$[-157; -79[$	$[0; 78[$
Réfraction normale	$[-79; 0[$	$[79; 157[$
Atmosphère standard	-39	118
Infra-réfraction	> 0	> 157

Tableau 1.1: Types de propagation en fonction des gradients de l'indice de réfraction

La présence de conduits de propagation (*ducts*), caractérisées par un gradient fortement négatif de $N(z)$, permet aux ondes de rester confinées entre deux altitudes. Ce confinement peut engendrer des portées de communication VHF dépassant plusieurs centaines de kilomètres, bien au-delà des limites théoriques imposées par l'horizon radioélectrique. L'indice de réfraction modifié est également utilisé pour tracer des profils verticaux permettant de caractériser les différents types de conduits de propagation. La Figure 1.2, inspirée de (Mentes et Kaymaz 2007 ; Zhang et al. 2016), illustre les profils typiques d'indice de réfraction modifié associés à diverses conditions atmosphériques. Dans cette figure, h désigne la hauteur totale du conduit, tandis que h_t correspond à la hauteur de la zone

d'emprisonnement, c'est-à-dire la couche dans laquelle l'indice de réfraction est suffisamment décroissant pour permettre à une onde de se propager parallèlement à la surface terrestre. Dans le cas des conduits de surface et des conduits d'évaporation (types (d) et (e)), l'égalité $h = h_t$ est vérifiée.

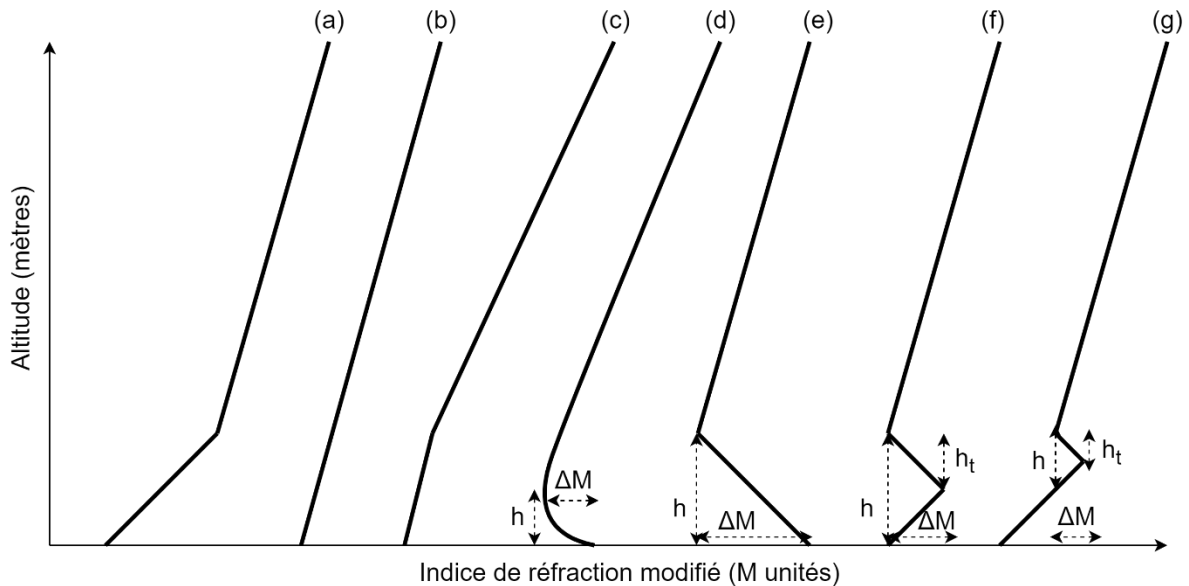


Figure 1.2 : Profils typiques d'indice de réfraction modifié : (a) infra-réfraction, (b) atmosphère standard, (c) super-réfraction, (d) conduit d'évaporation, (e) conduit de surface, (f) conduit basé en surface, (g) conduit élevé.

Selon (Caillet 2015), différentes conditions atmosphériques peuvent conduire à la formation de conduits, qu'ils soient de surface ou élevés. Les conduits de surface (types (d), (e) et (f)) sont généralement associés à des phénomènes d'advection, de quasi-advection, à l'évaporation, aux fronts atmosphériques ou encore au refroidissement radiatif du sol. Quant aux conduits élevés (type (g)), ils résultent principalement de l'advection, de la subsidence ou d'intrusions advectives. La plupart de ces phénomènes, à l'exception notable des intrusions advectives et du refroidissement du sol, peuvent survenir dans des environnements maritimes. Ils sont généralement provoqués par des gradients verticaux marqués de température et d'humidité, caractéristiques des conditions atmosphériques rencontrées en mer. Concernant les conduits de propagations, seuls les conduits de surface, les conduits basés en surface et les conduits élevés ont une influence significative sur la portée de l'AIS (Tang et al. 2018 ; Tang, Cha, Wei, et Tian 2019). En effet, l'impact du conduit d'évaporation n'est significatif qu'à partir de 2 GHz (Hitney et Vieth 1990).

1.1.1.2 Diffraction

La diffraction désigne le phénomène par lequel une onde électromagnétique est déviée de sa trajectoire initiale lorsqu'elle rencontre un obstacle ou franchit une ouverture dont les dimensions sont du même ordre de grandeur que sa longueur d'onde. La diffraction permet à l'onde de contourner partiellement l'obstacle, ce qui rend possible la réception même en l'absence de ligne de visée directe.

Dans le cadre des communications VHF, la prise en compte des effets de diffraction est essentielle pour estimer les pertes de liaison en présence de reliefs ou d'obstructions. Un outil classique utilisé à cette fin est l'ellipsoïde de Fresnel, qui définit le volume autour du trajet direct entre un émetteur et un récepteur à l'intérieur duquel les obstacles peuvent perturber significativement la propagation du signal. Le rayon du premier ellipsoïde de Fresnel s'exprime par :

$$r = \sqrt{\frac{\lambda d_1 d_2}{d_1 + d_2}}$$

Où r est le rayon de l'ellipsoïde, λ la longueur d'onde, d_1 la distance entre l'émetteur et l'obstacle, d_2 la distance entre l'obstacle et le récepteur en mètres.

En environnement maritime, la diffraction est particulièrement pertinente dans les situations de masquage partiel par un relief côtier, où elle permet de contourner ce relief (Qinjiang et al. 1998), comme schématisé en Figure 1.3 dans un modèle en lame de couteau, c'est-à-dire où l'obstacle est représenté de manière idéalisée par une arête de largeur négligeable (ITU-R 2019c).

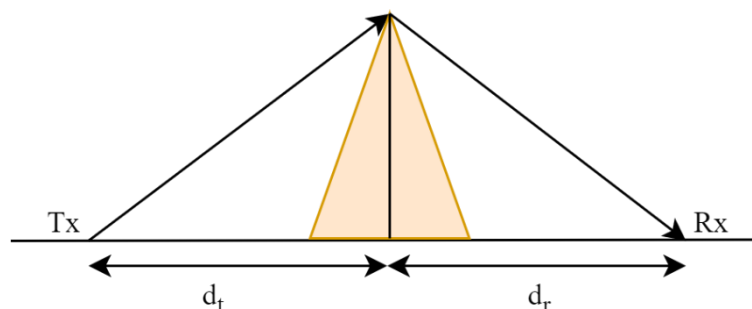


Figure 1.3 : Schématisation géométrique d'une communication par diffraction au-dessus d'un obstacle isolé dans un modèle en lame de couteau.

1.1.1.3 Diffusion

La diffusion désigne le phénomène par lequel une onde électromagnétique est partiellement déviée dans plusieurs directions lorsqu'elle interagit avec des irrégularités de taille comparable ou inférieure à sa longueur d'onde dans le milieu de propagation. Contrairement à la diffraction, qui concerne les obstacles nets, la diffusion résulte d'inhomogénéités ou de structures réparties dans le volume. En environnement troposphérique, il existe notamment :

- La diffusion troposphérique, causée par les fluctuations aléatoires de l'indice de réfraction dues à des turbulences de température et d'humidité (Section 1.1.1.1.1).
- La diffusion par hydrométéores, liée à la présence de particules comme la pluie, le brouillard ou la neige.

Pour cette dernière, dans le cas de l'AIS, le rapport longueur d'onde / diamètre de l'hydrométéore est si élevé que la taille de la particule est très inférieure à la longueur d'onde. Le régime de diffusion de Rayleigh s'applique alors, régime pour lequel l'efficacité de diffusion chute très rapidement avec la taille relative de la particule.

1.1.1.4 Absorption

L'absorption correspond à la conversion d'une partie de l'énergie électromagnétique en chaleur lors de la traversée de certains milieux. Ce phénomène est causé par l'interaction de l'onde avec les molécules du milieu, en particulier celles dotées d'un moment dipolaire. Dans la bande VHF, l'absorption atmosphérique reste relativement faible mais non négligeable. Elle est principalement liée à :

- L'absorption par la vapeur d'eau, dépendante de l'humidité relative.
- L'absorption par le dioxyde de carbone et le dioxygène, sensible à la pression et à la température.
- Dans une moindre mesure, les pertes par les précipitations.

L'absorption contribue à l'atténuation globale du signal sur de longues distances et peut être accentuée dans les couches denses et humides de la troposphère.

La recommandation ITU-R P.676 (ITU-R 2012) met en exergue le fait que les ondes de fréquences inférieures à la dizaine de GHz ne sont que très peu impactées par les gaz de l'atmosphère. Cela est confirmé par le spectre d'absorption de l'oxygène qui ne présente des raies qu'à partir de 50 GHz. De plus, la recommandation ITU-R P.527 (ITU-R 2021a) donne le diagramme de permittivité relative et de conductivité en fonction de la fréquence pour l'eau. Il peut en être conclu que l'absorption par l'eau contenue dans les hydrométéores est négligeable.

1.1.1.5 Réflexion

La réflexion est le phénomène par lequel une onde électromagnétique rebondit sur une interface séparant deux milieux aux propriétés électromagnétiques différentes. Elle obéit aux lois classiques de l'optique géométrique : l'angle d'incidence est égal à l'angle de réflexion, mesurés par rapport à la normale à la surface. En propagation VHF, les principales surfaces de réflexion sont :

- La surface de la mer ou d'un plan d'eau.
- Le sol, les reliefs ou les structures artificielles.
- L'ionosphère, bien que peu efficace pour les fréquences VHF.

L'analyse de (Wang et Zhang 2017) met en évidence que la propagation multi-trajets des signaux AIS, induite par les réflexions et diffractions sur la surface de la mer, constitue un facteur déterminant des performances du système. La réflexion sur la surface de l'eau engendre la formation de trajets multiples, susceptibles d'interférer de manière constructive ou destructive avec le signal direct, altérant ainsi la qualité de réception et la stabilité du lien. En milieu maritime, ces réflexions représentent un paramètre majeur à considérer, car elles sont à l'origine de variations significatives dans les transmissions. Les résultats de l'étude montrent que selon le mode de transmission (en LoS ou NLoS) et les conditions environnementales, les pertes de transmission liées aux signaux réfléchis peuvent être comparables à celles des signaux directs, intensifiant les effets de dégradation de la transmission VHF.

1.1.1.5.1 Réflexion ionosphérique

Des propagations à longue distance sont possibles pour les ondes de la bande métrique par réflexion sur l'ionosphère, notamment sur les couches sporadiques (notées E_s) (ITU-R 1994 ; Démoulin 2012). La couche E est une partie basse de l'ionosphère apparaissant à environ 100 km d'altitude. L'ionisation de cette couche par l'action des particules émises par le soleil ou des rayonnements X et γ va lui permettre de réfléchir les ondes (Castanet et Lassudrie-Duchesne 2008).

Néanmoins, seulement la propagation des ondes de la partie basse de la bande des ondes métriques est influencée par ce phénomène, ainsi il peut être négligé pour les fréquences de l'AIS (Bruin 2016 ; ITU-R 2023b). De plus, ce phénomène est présent sur un faible pourcentage de temps (ITU-R 2021b) pour des trajets de distances supérieures à 500 km. Dans certaines régions telles que l'équateur magnétique ou les zones polaires, il est présent plus fréquemment.

1.1.2 Typologies des phénomènes de propagation

À partir de cette revue des phénomènes de propagation, il est possible d'établir une typologie des principaux phénomènes de propagation dans les différents environnements traversés par les signaux AIS. Le Tableau 1.2 synthétise ces phénomènes, en précisant pour chaque milieu de propagation leurs effets sur la transmission AIS ainsi que les variables caractéristiques qui les influencent. Dans un système de communication, le terme Tx (*Transmitter*) désigne la voie d'émission, c'est-à-dire l'ensemble des éléments chargés de transmettre le signal, tandis que Rx (*Receiver*) désigne la voie de réception, chargée de capter et de traiter ce signal.

Le Tableau 1.3 complète cette analyse en offrant une liste de ces phénomènes selon leur caractère « normaux » ou « anormaux » selon la classification proposée par (Caillet 2015), ainsi que l'importance de leurs effets sur la propagation de l'AIS.

Milieu		Phénomène	Effets	Variables caractéristiques
Atmosphère	Troposphère	Réfraction	Augmentation de la portée du signal	Fréquence, Pression, Température, Humidité relative, Pression partielle de la vapeur d'eau
		Diffusion	Propagation au-delà de l'horizon radioélectrique	Fréquence, Pression, Température
	Ionosphère	Réflexion	Effet négligeable	Configuration de l'ionosphère, activité solaire
	Gaz	Absorption	Effet négligeable	Fréquence, Température, Pression
	Hydrométéores	Absorption	Effet négligeable	Fréquence, Température
		Diffusion	Contournement d'obstacles, effet global non significatif	Fréquence, Diamètres des hydrométéores
Terrain	Terrain	Diffraction	Contournement du relief et des obstacles	Fréquence, Hauteur du relief
		Réflexion	Changement de direction l'onde	Fréquence, Conductivité du matériau réfléchissant
	Mer	Diffraction	Changement de la direction de l'onde, faible augmentation de la portée	Fréquence, Hauteur des vagues, Période des vagues
		Réflexion/Réfraction	Changement du chemin de l'onde	Fréquence, Hauteur des vagues, Période des vagues, Indice de réfraction de l'eau salée
Système	Antenne	Transmission (Friis)	Augmentation ou diminution de la portée selon les paramètres	Fréquence, Puissance transmise, Gains/Directivité, Distance Tx-Rx
		Visibilité	Permet la propagation en visibilité directe	Hauteurs Tx/Rx, Distance Tx-Rx

Milieu		Phénomène	Effets	Variables caractéristiques
Système	Câble	Pertes en ligne	Perte lors du transport entre les antennes et les transpondeurs	Longueur du câble, pertes des câbles en dB m ⁻¹
	Emetteur mobile	Effet doppler	Effet négligeable	Fréquence, Vitesse de Tx

Tableau 1.2 : Phénomènes de propagation, effets sur la propagation de l'AIS, variables caractéristiques et milieux.

Phénomène		Milieu/Type	Importance
Propagations normales	Absorption	Gaz	
		Hydrométéores	
	Réflexion	Mer	
		Terrain	
	Diffraction	Mer	
		Terrain	
Diffusion	Hydrométéores		
	Troposphère		
Propagations anormales	Réfraction	Infra/Super réfraction	
		Conduits d'évaporation	
		Conduits de surface	
		Conduits basés en surface	
	Conduits élevés		
Réflexion	Ionosphère		

Tableau 1.3 : Phénomènes de propagation et l'importance de leurs effets sur la propagation de l'AIS (rouge : Négligeable, orange : Peu significatif, Vert : Influence).

Cette typologie constitue une base pour analyser comment les modèles existants incluent individuellement les différents phénomènes de propagation.

1.1.3 Caractéristiques des systèmes de communication AIS

Les performances de la transmission et de la réception des messages AIS dépendent fortement des caractéristiques des systèmes radio utilisés. Ces caractéristiques influencent directement la portée, la qualité du signal reçu et la sensibilité aux phénomènes de propagation. Trois paramètres principaux peuvent être identifiés : le positionnement des antennes, le gain et la directivité (ITU-R 2014 ; Last et al. 2015).

1.1.3.1 Positionnement des antennes

Le positionnement de l'antenne joue un rôle déterminant dans la couverture effective du système AIS. L'IMO propose des lignes de conduites (IMO 2003) sur le placement d'antennes dans le cadre des communications AIS. Elles stipulent notamment que l'antenne doit être placée à une hauteur d'au minimum deux mètres, loin de toute obstruction et potentielles sources d'interférences. Ainsi, la

hauteur minimale d'une antenne AIS peut être considérée comme étant 2 mètres. Quant à la hauteur maximale, elle est limitée par la hauteur de mât ou de la passerelle radar, soit un ordre de grandeur de la trentaine de mètres.

1.1.3.2 Gain de l'antenne

Le gain d'une antenne exprime sa capacité à concentrer l'énergie dans une direction donnée par rapport à une antenne isotrope. Un gain élevé permet d'améliorer la sensibilité à la réception ou la portée à l'émission. Pour une antenne AIS, le gain typique est de 2 à 5 dBi (Tang et al. 2018 ; Gómez-Guillamón Buendía et al. 2023) afin de conserver une couverture omnidirectionnelle mais peut atteindre 9 dBi pour les plus grands modèles.

La directivité d'une antenne caractérise l'aptitude du système à concentrer son rayonnement dans une direction privilégiée. Les antennes omnidirectionnelles, couramment utilisées dans les systèmes AIS (IMO 2003), émettent de manière uniforme dans le plan horizontal. Cependant, ce n'est pas toujours le cas pour les stations réceptrices côtières qui ont un intérêt à favoriser la réception suivant une certaine directivité.

1.1.3.3 Puissance d'émission

La puissance d'émission d'une antenne correspond à l'énergie rayonnée par le système d'émission (moins les pertes), généralement exprimée en watts (W) ou en décibels par rapport à 1 milliwatt (dBm). Dans les systèmes AIS, cette puissance conditionne directement la portée du signal. En pratique, l'AIS dispose deux niveaux de puissance : normal ou réduit. Le choix du niveau de puissance se fait en fonction du contexte, par exemple, un niveau de puissance réduit pourra être utilisé au port ou au mouillage, tandis qu'un niveau normal sera utilisé pendant la navigation. Le Tableau 1.4 présente ces différents niveaux de puissance en fonction de la classe du dispositif.

Classe	Réduit	Normal
A	1 W	12,5 W
B (SOTDMA)	1 W	5 W
B (CSTDMA)	1 W	2 W

Tableau 1.4 : Valeurs en Watt de la puissance en mode réduit et normal pour les dispositifs AIS de classes A et B.

1.2 Modélisation physique de la propagation de l'AIS

Au-delà de la description des phénomènes physiques, la modélisation de la propagation constitue un enjeu central pour quantifier et prédire la couverture effective des systèmes AIS. Historiquement développés pour les radiocommunications terrestres et maritimes, les modèles physiques offrent un cadre mathématique permettant d'estimer les pertes de transmission, d'évaluer

l'impact des conditions atmosphériques et de caractériser les effets de la topographie ou de la surface de la mer.

Cette sous-partie présente ainsi les modèles physiques classiques, leur application spécifique au cas de l'AIS, que ce soit pour un besoin d'évaluation ou de leur intégration en tant que sous modèles. Elle dresse également une typologie des phénomènes normaux et anormaux issues des phénomènes de propagation de base et de leur combinaison, avant d'expliquer leur prise en compte dans les modèles de référence, notamment ceux proposés par l'ITU.

1.2.1 Modèles physiques classiques de propagation

La modélisation physique de la propagation des signaux AIS repose sur des cadres théoriques initialement développés pour l'étude générale des ondes radioélectriques en milieu terrestre et maritime. Ces modèles permettent de quantifier les pertes de transmission et de caractériser les phénomènes de propagation en fonction de la géométrie, des propriétés du milieu et des conditions atmosphériques. Dans le contexte AIS, ils constituent une base de départ pour estimer la portée de détection, analyser l'influence des phénomènes de réfraction et de réflexion, et anticiper les effets imposés par la topographie ou l'environnement maritime.

Les modèles physiques classiques, présentés ci-après, peuvent être regroupés en trois catégories principales (Tableau 1.5) :

- Modèles géométriques : qui déterminent la portée à partir des contraintes de visibilité directe et des paramètres physiques des antennes.
- Modèles électromagnétiques en espace libre ou à trajets multiples : qui décrivent la perte de puissance en fonction de la distance et de la fréquence.
- Modèles empiriques ou semi-déterministes : qui combinent lois physiques et données issues de mesures pour intégrer des effets plus complexes comme la diffraction, la réfraction troposphérique ou les caractéristiques du terrain.

Modèle	Catégorie géométrique	Catégorie électromagnétique (espace libre / trajets multiples)	Catégorie empirique / semi-déterministe
Modèle géométrique de visibilité directe	✓		
Modèle 4/3	✓		
Équation de Friis (équation des télécommunications)		✓	
Modèle FSPL (Free-Space Path Loss)		✓	
Modèle d'Egli			✓
Modèle de réflexion sur le sol à deux rayons		✓	
Recommandations ITU-R			✓
Modèle ITM (Irregular Terrain Model)			✓

Tableau 1.5 : Classification des modèles physiques classiques de propagation selon leur catégorie.

1.2.1.1 Modèle géométrique de visibilité directe

La visibilité directe caractérise la propagation radioélectrique entre une source et un récepteur reliés par un trajet direct (IEEE 2019), sans obstacle significatif. Le chemin est suffisamment dégagé pour que les phénomènes de diffraction ou de réflexion aient une influence négligeable (ITU-R 2019a).

En appliquant le théorème de Pythagore à une sphère de rayon a pour une station émettrice à l'altitude z_t et une station réceptrice à l'altitude z_r , les distances à l'horizon s'expriment alors comme suit :

$$d_t = \sqrt{2az_t + z_t^2} \text{ et } d_r = \sqrt{2az_r + z_r^2} \quad (2)$$

Pour la Terre, il convient d'introduire le concept de rayon effectif ou équivalent, permettant de tenir compte de la réfraction atmosphérique. Il s'agit du rayon d'une Terre fictive, sphérique et sans atmosphère, choisi de sorte que les trajets radioélectriques puissent être considérés comme rectilignes, tout en conservant les mêmes altitudes et distances au sol que sur la Terre réelle entourée d'une atmosphère à gradient vertical constant de l'indice de réfraction (ITU-R 2019a).

Ce rayon équivalent est obtenu en appliquant au rayon réel de la Terre un facteur de correction k , appelé facteur k , qui permet de prendre en compte de manière simplifiée les effets de la réfraction

dans les calculs de propagation sphérique (IEEE 2019). Le rayon effectif de la Terre est alors $a_e = ka$, d'où :

$$d_{LoS} = d_t + d_r \approx \sqrt{2a_e z_t} + \sqrt{2a_e z_r} \quad (3)$$

1.2.1.2 Le modèle 4/3

Le modèle 4/3 correspond à la valeur de k établie par (Schelleng et al. 1933) dans l'expression du rayon effectif de la Terre dans le modèle de visibilité directe (Équation (3)). En appliquant l'approximation aux petits angles on remarque que l'Équation (1) permet d'obtenir la valeur :

$$k \approx \frac{1}{a} + \frac{1}{n} \frac{dn}{dz} \quad (4)$$

(Schelleng et al. 1933) injectent alors la valeur du gradient de n des premiers mètres de l'atmosphère standard $\frac{dn}{dz} = -\frac{1}{4a}$ dans l'Équation (4) pour obtenir $k = \frac{4}{3}$. Cette valeur permet de modéliser les effets du déclin en pression de l'atmosphère avec l'altitude, dans ses premiers mètres et dans le cas d'une atmosphère standard. En utilisant ce modèle, l'application numérique de l'Equation (2) donne :

$$d_{LoS} = 4,12 \cdot 10^3 (\sqrt{z_t} + \sqrt{z_r})$$

Cela correspond à une augmentation de la distance de visibilité directe d'environ 15 %.

1.2.1.3 Equation des télécommunications

Si l'onde est transmise sans subir de changement de direction ou d'intensité lié à son environnement, la puissance reçue est donnée par l'équation de *Friis* (équation des télécommunications) (Friis 1946) :

$$\frac{P_r}{P_t} = G_t G_r \left(\frac{\lambda}{4 \pi d_{rt}} \right)^2$$

Où P_r, P_t sont la puissance reçue et transmise en Watts, G_r, G_t sont les gains du récepteur et de l'émetteur, λ la longueur de l'onde transmise en mètres et d_{rt} la distance émetteur-récepteur en mètres.

1.2.1.4 Le modèle FSPL

FSPL (*Free-Space Path Loss*), ou « perte en espace libre » s'applique dans les transmissions électromagnétiques sans obstructions. C'est la perte entre deux radiateurs isotropes dans l'espace libre, exprimée sous forme de rapport de puissance. Il caractérise la propagation en espace libre, c'est-à-dire la propagation d'une onde dans un milieu idéal, homogène, sans obstacles, assimilé à un vide

parfait, caractérisé par les paramètres constitutifs du vide et considéré comme infini dans toutes les directions, dans lequel se propage une onde électromagnétique sans atténuation ni perturbation (IEEE 2019 ; ITU-R 2019a).

En fait, la perte en espace libre n'est pas due à la dissipation, mais au fait que la densité de flux de puissance diminue avec le carré de la distance de séparation (IEEE 1993). Son équation est dérivée de l'équation de (Friis 1946), (Rappaport 2024 ; Saunders et Aragón-Zavala 2024) donnent :

$$\text{FSPL} = \frac{P_t}{P_r} = \left(\frac{4\pi d_{rt} f}{c} \right)^2$$

Où f est la fréquence de l'onde en Hz, c la célérité de la lumière en $m s^{-2}$ et d_{rt} la distance entre l'émetteur et le récepteur.

Ou en dB :

$$\text{FSPL}_{\text{dB}} = 20 \log_{10} d_{rt} + 20 \log_{10} f + 20 \log_{10} \frac{4\pi}{c}$$

1.2.1.5 Modèle d'Egli

Le modèle d'Egli (Egli 1957) constitue l'un des modèles empiriques de propagation les plus connus et les plus largement utilisés pour la bande VHF et UHF. Il a été élaboré à partir d'une analyse statistique approfondie de données recueillies lors de nombreuses campagnes de mesures en environnement urbain et suburbain.

À partir de ces données expérimentales, Egli a établi des formules empiriques permettant de relier la perte de propagation aux caractéristiques physiques du canal radio, en particulier la hauteur des antennes, la distance de séparation entre l'émetteur et le récepteur, ainsi que la fréquence du signal. Ce modèle vise à prendre en compte l'impact de la variabilité du terrain et des obstacles urbains sur l'atténuation des ondes électromagnétiques, en comparaison avec les conditions idéales d'espace libre. L'expression de la perte selon ce modèle est :

$$\frac{P_r}{P_t} = G_t G_r \left(\frac{z_t z_r}{d_{rt}^2} \right)^2 \left(\frac{40}{f} \right)^2$$

Ce modèle est principalement applicable dans le cadre de communications point à point en ligne de vue, notamment pour les scénarios où l'émetteur est mobile et le récepteur fixe (Sharma et Singh 2010). Il est souvent employé pour modéliser les réseaux de radiodiffusion et de télécommunication dans des environnements urbains denses, car il permet de mieux estimer l'impact des bâtiments et de la topographie que le simple modèle en espace libre.

1.2.1.6 Modèle de réflexion sur le sol à deux rayons

Le modèle de réflexion sur le sol à deux rayons constitue un des modèles de propagation utilisés pour décrire la transmission d'ondes radioélectriques dans un environnement où le sol est considéré comme plan homogène. Ce modèle repose sur l'hypothèse que le signal transmis par l'antenne émettrice atteint l'antenne réceptrice en suivant deux chemins distincts :

- Le chemin direct, correspondant à la ligne de visée entre les deux antennes.
- Le chemin réfléchi, qui résulte de la réflexion de l'onde incidente sur la surface du sol avant d'atteindre le récepteur.

La combinaison de ces deux contributions engendre un phénomène d'interférence constructive ou destructive, selon la différence de phase entre le chemin direct et le chemin réfléchi. Cette différence dépend des distances parcourues, de la hauteur des antennes ainsi que des caractéristiques de la surface réfléchissante. La perte de propagation peut alors être déterminée en prenant simultanément en compte l'atténuation associée au chemin direct et celle liée au chemin réfléchi. Ainsi, l'expression mathématique proposée par. (Zhao et al. 2013) s'écrit :

$$L_{2-ray} = -10 \log_{10} \left[\left(\frac{\lambda}{4\pi d_{rt}} \right)^2 \left(\frac{2\pi z_r z_t}{\lambda d_{rt}} \right)^2 \right]$$

1.2.1.7 Recommandations ITU-R

L'Union Internationale des Télécommunications (ITU), notamment via le secteur des radiocommunications (R), fournit des recommandations sur l'étude des phénomènes, les techniques et les méthodes associées aux propagations (P) et aux prédictions de propagations des ondes radioélectriques.

Le Tableau 1.6 regroupe les recommandations applicables et d'intérêt dans le contexte des communications VHF et de l'étude de la propagation des messages AIS. Cette sélection est réalisée en ne considérant que les recommandations concernant : les très hautes fréquences, l'impact des conditions atmosphériques sur la propagation, l'impact de la topographie sur la propagation, les prédictions de propagation, les communications point à zone et fondées sur le trajet pour les très hautes fréquences.

Description	Recommandation
Choix d'indices pour les prévisions ionosphérique à long terme	P.371
Méthode de prévision pour évaluer les brouillages entre stations situées à la surface de la Terre à des fréquences supérieures à 100 MHz environ	P.452
Indice de réfraction radioélectrique : formules et données de réfractivité	P.453
Calcul de la propagation en espace libre	P.525
Propagation par diffraction	P.526
Données de propagation et méthodes de prévision nécessaires pour la conception de faisceaux hertziens à visibilité directe de Terre	P.530
Techniques de prévision de la propagation et données de propagation nécessaires pour la conception des faisceaux hertziens transhorizon	P.617
Affaiblissement dû aux gaz de l'atmosphère et effets associés	P.676
Données de propagation nécessaires pour la conception de systèmes de télécommunication mobiles maritimes Terre-espace	P.680
Effets de la réfraction troposphérique sur la propagation des ondes radioélectriques	P.834
Caractéristiques des précipitations pour la modélisation de la propagation	P.837
Modèle d'affaiblissement linéique dû à la pluie, destiné aux méthodes de prévision	P.838
Modèle d'altitude de pluie pour les méthodes de prévision	P.839
Affaiblissement dû aux nuages et au brouillard	P.840
Modélisation de la propagation des ondes radioélectriques : distributions de probabilité	P.1057
Bases de données topographiques numériques pour les études de propagation	P.1058
Propagation par trajets multiples et paramétrage de ses caractéristiques	P.1407
Méthode de prévision de la propagation point à zone pour les services de Terre entre 30 MHz et 4 000 MHz	P.1546
Méthode de prévision de la propagation fondée sur le trajet pour les services de Terre point à zone dans la gamme de fréquences comprises entre 30 MHz et 6 000 MHz	P.1812
Modèle général de large portée pour la propagation sur des trajets de Terre dans la gamme des fréquences comprises entre 30 MHz et 50 GHz	P.2001

Tableau 1.6 : Descriptions et recommandations ITU-R d'intérêts pour l'étude de la propagation de l' AIS

1.2.1.8 Modèle ITM

Le modèle ITM (*Irregular Terrain Model*) de l'ITS (*Institute for Telecommunication Sciences*) couvre les fréquences de 20 MHz à 20 GHz. C'est un modèle polyvalent permettant de résoudre divers problèmes d'ingénierie. Il se base sur la théorie électromagnétique et des analyses statistiques des caractéristiques du terrain ainsi que de mesures radio. Ce modèle permet de prédire l'atténuation médiane d'un signal radio en fonction de la distance et comment il varie dans le temps et dans l'espace. Il a été développé par Anita Longley et Phil Rice en 1968 (Longley et Rice 1968) puis adapté pour un usage par ordinateur (Hufford et al. 1982) dont l'algorithme est donné par (Hufford 1995).

Ce modèle fonctionne en deux modes de prédiction de propagation : « point-à-point » ou « zone », qui divergent par les paramètres d'entrée nécessaires. Le mode « point-à-point » nécessite des informations précises concernant le profil de terrain, tandis que le mode « zone » peut utiliser des informations empiriques. En fonction du mode, le modèle commence par effectuer des calculs préliminaires de nature géométrique. Par exemple, en mode zone l'horizon radioélectrique étant inconnu, il sera estimé. Le modèle calcule ensuite l'atténuation de référence en fonction de la distance. Le calcul de cette atténuation est divisé en trois régions : visibilité directe, diffraction et diffusion (Figure 1.4). Cette atténuation de référence peut être une mesure suffisante pour résoudre certains problèmes, mais le modèle offre la possibilité de calculer des statistiques supplémentaires d'atténuation liées à la variabilité des positions, du temps et de la situation.

Les auteurs précisent que ce modèle utilise à la fois des calculs théoriques sur la réflexion sur le sol, de réfraction atmosphérique, la diffraction et la diffusion troposphérique ainsi que des relations empiriques issues de mesures. Ce qui en fait un modèle semi-déterministe. Ils ajoutent que de par cette nature, les relations établies empiriquement influencent les prédictions du modèle. Dans ce modèle, les phénomènes à faible probabilité tels que les conduits de propagation sont pris en compte sous forme de constantes dépendantes du type de climat de propagation. Les mesures empiriques incluent évidemment de manière indirecte les propagations ionosphériques, la propagation par hydrométéores ou l'atténuation par des gaz atmosphériques mais les auteurs ne conseillent pas d'utiliser ce modèle pour prédire ou prendre en compte de tels phénomènes.

Plus récemment, une version incluant la recommandation ITU-R P.1546 nommé ITWOM (*Irregular Terrain With Obstructions Model*) a été développé par (Shumate 2010).

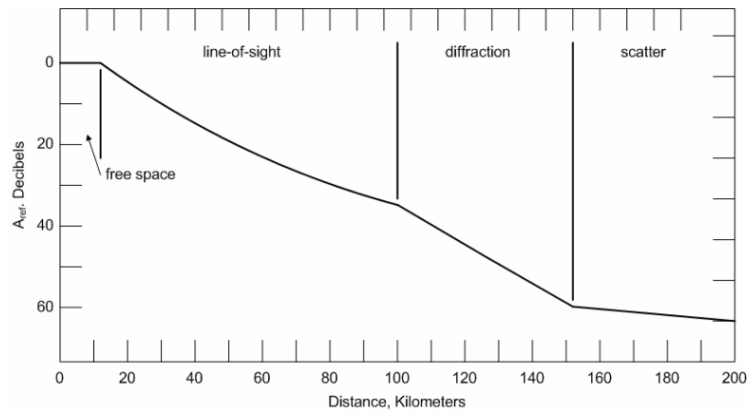


Figure 1.4 : Courbe représentative d'une atténuation de référence calculée par ITS ITM (Hufford et al. 1982)

1.2.2 Application des modèles physiques à la propagation de l'AIS

Ces modèles et paradigmes constituent un socle méthodologique, mais leur pertinence doit être évaluée au regard de situations réelles. C'est pourquoi plusieurs travaux ont cherché à les appliquer spécifiquement au cas de l'AIS, que ce soit dans une approche hybride ou à des fins d'évaluation. Dans un premier temps, la propagation des messages AIS peut être étudiée par une approche statistique. Dans (Salmon et al. 2016), les auteurs mettent en évidence l'absence de réception de messages AIS par certaines stations côtières pour des zones spécifiques. Dans ces régions, aucun signal n'est capté, ce qui les conduit à introduire la notion de « trous noirs » : des zones non couvertes pendant une période donnée. La détection de ces zones nécessite à la fois une analyse historique des données et l'intégration de flux en temps réel. Les auteurs soulignent également que la présence et l'ampleur des « trous noirs » varient dans le temps et sont fortement corrélées aux conditions atmosphériques.

La mesure de la puissance reçue par une station côtière, intégrée dans un modèle de propagation, permet également d'analyser la diffusion des signaux AIS. (Kelly 2022) compare la puissance reçue (RSSI) par une station réceptrice à la valeur théorique issue du modèle d'Egli. Il relie cette comparaison à la connaissance de la situation maritime, en détectant notamment les navires équipés d'AIS de classe A qui auraient volontairement réduit la puissance de leur dispositif afin de masquer leur position. (Mazzarella et al. 2017) combinent les deux approches précédentes dans un cadre de détection d'anomalies liées à l'extinction d'un dispositif AIS. Leur module de modélisation de la normalité s'appuie d'une part sur une approche statistique, en entraînant un algorithme d'apprentissage automatique à modéliser la zone de réception d'une station à partir de variables caractérisant les navires, de leurs coordonnées et des RSSI historiques. D'autre part, l'approche théorique y est intégrée en calculant, navire par navire, une distance critique de détection à l'aide du modèle ITU-R P.1546, de l'équation de Friis et de l'optique géométrique. Dans un second module, dédié à la détection d'anomalies, les risques sont évalués en comparant les RSSI observés aux valeurs critiques et aux données historiques, puis en appliquant un seuil de décision.

Ces approches modélisent donc les variations de réception à partir de mesures de puissance et de modèles théoriques, mais l'effet direct des conditions atmosphériques n'est pris en compte que via l'aspect statistique. De nombreux travaux se sont ainsi attachés à relier les phénomènes de propagation aux données AIS. (Tang et al. 2018) utilisent la différence entre le RSSI observé et la valeur théorique calculée à partir de l'équation de propagation pour définir une fonction objectif optimisée par algorithme à essaim de particules, afin d'estimer le profil du conduit atmosphérique responsable de l'écart observé. (Han et al. 2022) adoptent une approche similaire pour classifier le type de conduit à partir du profil d'indice de réfraction atmosphérique calculé. (Valčić et Brčić 2023) confirment que les propagations transhorizon AIS sont principalement dues aux variations de l'indice de réfraction, en analysant des profils générés par le logiciel AREPS (Patterson 2007). (L.-F. Huang et al. 2022) confirment empiriquement ce lien par l'utilisation de ballons-sondes et précisent que les longues distances sont associées aux conduits atmosphériques bas, tandis que les faibles extensions transhorizon proviennent des conduits d'évaporation, deux phénomènes renforcés par la faible hauteur des antennes AIS.

De nombreuses approches ont été développées pour modéliser la propagation AIS et l'associer à des problématiques de connaissance de la situation maritime, en particulier la détection d'extinctions volontaires de dispositifs. Toutefois, le développement de modèles intégrant simultanément les caractéristiques propres aux signaux AIS, la topographie et les paramètres atmosphériques demeure limité. (Bruin 2016) propose une méthode originale reposant sur une modélisation conjointe de l'atmosphère et de l'environnement maritime afin de mettre en place une approche semi-déterministe, dans laquelle le modèle AREPS (Patterson 2007) est intégré comme modèle de calcul de propagation. Cette intégration permet de tirer parti de la richesse fonctionnelle d'AREPS, c'est-à-dire une description fine des capteurs et cibles, l'ingestion de jeux de données atmosphériques étendus, l'estimation de la perte de propagation et cartographie de la couverture, tout en conservant un certain ancrage dans les données réelles. Néanmoins, l'utilisation d'AREPS présente des limites structurelles pour ce type d'application. En effet, chaque simulation est contrainte à un capteur unique et à une directivité unique, orienté depuis la position de l'antenne en ligne droite. Ce qui en pratique, engendre des discontinuités spatiales dans les cartes produites. Surtout, cette approche ne tire pas pleinement parti des positions AIS réelles : l'évaluation de la couverture est effectuée indépendamment de la distribution effective des navires observés.

La section suivante s'attache donc à quantifier plus précisément l'intégration de ces différents phénomènes dans les modèles physiques, afin d'évaluer leur contribution effective à la représentation de la propagation AIS/VHF.

1.2.3 Prise en compte des phénomènes de propagation dans les modèles

Le Tableau 1.7 décrit les phénomènes de propagation pris en compte dans les différents modèles de propagation vus dans l'état de l'art, en les regroupant par similarité ou par cas d'application. Leurs paradigmes sont définis comme :

- Empiriques : modèles dérivés directement de mesures.
- Semi-déterministes / Semi-empiriques / Stochastiques : modèles dérivés de mesures avec des ajustements théoriques/calculatoires spécifiques.
- Déterministes : modèles établis directement par équations (résolution d'équations optiques, électromagnétiques etc.) pour estimer une grandeur à une localisation particulière.

Les modèles généraux permettent d'estimer les pertes de propagation pour une grande variété de situation ou peuvent être intégrés en tant que sous-modèle au sein d'un modèle plus spécifique. Le modèle d'(Egli 1957), purement empirique est issu de campagne de mesures en VHF/UHF fournit une estimation simple et rapide de la perte sur le trajet en fonction de la fréquence, de la distance et de la hauteur d'antenne. Il reste limité quant à sa distance d'application. Le modèle ITM (Longley et Rice 1968), présenté en Section 1.2.1.8 est semi-déterministe et intègre le profil réel du terrain. Ce modèle est intégré au sein de (Zhao et al. 2013) pour les communications maritimes.

Les modèles de l'ITU sont normalisés pour assurer une cohérence internationale dans les prévisions de propagation et l'évaluation des interférences radioélectriques. L'ITU-R P.452 (ITU-R 2023a) est un modèle semi-déterministe conçu pour prédire la perte de propagation sur de longues distances. Il est largement utilisé pour l'étude des interférences entre services fixes et de radiocommunication. L'ITU-R P.1546 (ITU-R 2019d) inclus sous forme de courbes définies pour différentes fréquences entre 30 MHz et 4000 MHz, les prévisions de propagation pour différentes hauteurs d'antennes émettrices jusqu'à une distance de 1000 km. Ces courbes sont établies empiriquement et prennent en compte de nombreux types de terrain (mer froide, mer chaude, trajet maritime, trajet terrestre etc.). Dans ce modèle, des méthodes de correction sont proposées pour prendre en compte les phénomènes de diffraction et de diffusion. Les courbes de ce modèle ont été établies dans l'hypothèse où les effets ionosphériques sont négligeables. L'ITU-R P.1812 (ITU-R 2023b) prend en compte les mêmes phénomènes mais de manière semi-déterministe à l'aide d'équations en se basant sur l'analyse du trajet de l'onde. Dans ce modèle un paramètre est inclus et calculé pour quantifier l'impact des phénomènes liés aux conduits. L'ITU-R P.2001 (ITU-R 2023c) est un modèle semi-déterministe « tout-en-un » couvrant une large plage de fréquences et de distances. Il intègre dans un seul calcul les différents phénomènes de propagation : visibilité directe, diffraction, diffusion troposphérique, conduits, sub/super-réfraction et pertes dues aux hydrométéores. Les effets

anormaux y sont inclus de façon statistique à partir de données climatiques régionales, ce qui en fait un outil polyvalent pour la prévision de la propagation terrestre point à point.

Parmi l'ensemble des modèles existants, ceux développés par l'ITU suscitent un intérêt particulier : d'une part en tant qu'outils opérationnels polyvalents et d'autre part comme modèles de référence, grâce à la variété des phénomènes qu'ils intègrent ainsi qu'à l'étendue des plages fréquentielles et des distances couvertes. Afin de mettre en évidence les compromis associés aux différentes familles de modèles de propagation, le Tableau 1.8 synthétise les principales limites des approches empiriques, semi-empiriques et déterministes dans le contexte de la modélisation de la réception AIS.

Type	Modèle	Fréquence (MHz)	Distance (km)	Paradigme	Phénomènes normaux	Phénomènes anormaux
Modèles généraux	(Egli 1957)	[30; 3000]	LoS	Empirique	Visibilité directe, Variabilité terrain	
	(Longley et Rice 1968)	[20; 2000]	[1; 2000]	Semi-empirique	Visibilité directe, Diffraction topographique, Variabilité terrain, Diffusion troposphérique	Liés à la réfractivité troposphérique
	(Zhao et al. 2013)	1800	[50 ; 90]	Semi-empirique	Visibilité directe, Diffraction topographique, Variabilité terrain, Diffusion troposphérique , Propagation multi-trajets	Liés à la réfractivité troposphérique , Hydrométéores
Modèles ITU	ITU-R P.452 (ITU-R 2023a)	[100; 50000]	[1 ; 10000]	Semi-empirique	Visibilité directe, Diffraction topographique, Variabilité terrain, Diffusion troposphérique , Propagation multi-trajets	Propagations par conduits, Hydrométéores

Type	Modèle	Fréquence (MHz)	Distance (km)	Paradigme	Phénomènes normaux	Phénomènes anormaux
Modèles ITU	ITU-R P.1546 (ITU-R 2019d)	[30 ; 4000]	[1 ; 1000]	Empirique	Visibilité directe, Diffraction topographique, Variabilité terrain, Diffusion troposphérique , Propagation multi-trajets	Liés à la réfractivité troposphérique
	ITU-R P.1812 (ITU-R 2023b)	[30 ; 6000]	[0,25 ; 3000]	Semi-empirique	Visibilité directe, Diffraction topographique, Variabilité terrain, Diffusion troposphérique , Propagation multi-trajets	Propagations par conduits
	ITU-R P.2001 (ITU-R 2023c)	[30 ; 50000]	[3 ; 1000]	Semi-empirique	Visibilité directe, Diffraction topographique, Variabilité terrain, Diffusion troposphérique , Propagation multi-trajets	Propagations par conduits, Hydrométéores , Propagations ionosphériques
Modèles AIS	(Bruin 2016)	AIS	[0,5 ; 720]	Semi-empirique	Visibilité directe, Diffusion troposphérique	Propagations par conduits
	(Mazzarella et al. 2017)	AIS	[0 ; 900]	Semi-empirique	Visibilité directe, Diffusion troposphérique	
	(Kelly 2022)	AIS	LoS	Déterministe	Visibilité directe, Diffusion troposphérique	

Tableau 1.7 : Modèles de propagations sur la plage de fréquence AIS, paramètres du système, paradigme et phénomènes pris en compte.

Type de modèle	Caractéristiques	Limites principales
Empiriques (ex. Egli, ITU-R P.1546)	<ul style="list-style-type: none"> – Basés sur des campagnes de mesures et ajustements statistiques. – Simplicité de mise en œuvre, peu de paramètres requis. 	<ul style="list-style-type: none"> – Spécifiques à un site ou à un contexte (faible capacité de généralisation). – Ne prennent pas en compte la variabilité spatio-temporelle (atmosphère, météo). – Résultats moyens / statistiques : peu adaptés à la détection d'anomalies locales.
Semi-empiriques / semi-déterministes (ex. Longley-Rice, ITU-R P.452, P.1812, P.2001)	<ul style="list-style-type: none"> – Combinent lois physiques et ajustements empiriques. – Intègrent plusieurs phénomènes (diffraction, réfraction, diffusion...). 	<ul style="list-style-type: none"> – Dépendent d'entrées de qualité (profils de terrain, climat de propagation) pas toujours disponibles. – Précision limitée par les simplifications statistiques (effets de conduits, météo réduits à des moyennes). – Moins adaptés à un usage dynamique temps réel.
Déterministes (ex. équation de Friis, FSPL, modèles paraboliques, ray tracing)	<ul style="list-style-type: none"> – Fondés sur la résolution d'équations physiques. – Précision élevée dans des environnements contrôlés. 	<ul style="list-style-type: none"> – Très coûteux en calcul (difficile pour des zones étendues). – Forte dépendance aux données d'entrée (profils atmosphériques détaillés rarement disponibles). – Peu flexibles : difficile d'intégrer des phénomènes stochastiques (variabilité météo, interférences).

Tableau 1.8 : Limites des approches de modélisation de la propagation AIS.

1.2.4 Les modèles ITU comme références

Les modèles développés par l'ITU sont largement utilisés, d'une part ils peuvent être intégrés au sein d'un modèle plus spécifique, mais ils sont aussi utilisés pour évaluer les modèles développés par la communauté scientifique. De manière similaire, plusieurs travaux ont employé les recommandations P.452, P.1546, P.1812 et P.2001 pour évaluer des modèles empiriques ou d'apprentissage automatique appliqués aux bandes VHF, UHF et SHF. P.1546 est couramment utilisé

pour estimer la couverture de systèmes maritimes en VHF comme l'AIS (Mazzarella et al. 2017), P.452 pour modéliser et analyser les interférences entre services (Al-Jumaily et al. 2023), P.1812 pour les prévisions point à point prenant en compte le profil réel du terrain et les effets de conduits (Bocus et Lodhi 2025), et P.2001 comme référence polyvalente couvrant un large domaine fréquentiel et intégrant les phénomènes normaux et anormaux de propagation (Lee et al. 2023) .

Ces études confirment que, si P.452, P.1546 et P.1812 restent adaptés à des cas d'usage spécifiques, P.2001 présente l'avantage d'une couverture fréquentielle et de distance plus étendue, d'une modélisation unifiée de multiples phénomènes de propagation et d'une représentation statistique complète (0 à 100 % du temps), faisant ainsi de lui le modèle le plus polyvalent à mettre en œuvre pour la validation d'autres modèles.

1.3 Conclusion

Ce chapitre a permis d'examiner les phénomènes physiques et les caractéristiques du dispositif d'émission/réception qui gouvernent la propagation des signaux AIS en environnement maritime. Les principaux phénomènes : réfraction, diffraction, diffusion, absorption et réflexion, ainsi que les caractéristiques des antennes et des émetteurs ont été présentés, en soulignant leur rôle déterminant dans la portée et la qualité des transmissions. Cette analyse a montré que la variabilité de la propagation ne résulte pas d'un seul facteur, mais de l'interaction complexe entre phénomènes atmosphériques, conditions maritimes et paramètres techniques.

L'étude des modèles physiques a ensuite mis en évidence leur intérêt comme cadre théorique de référence. Les modèles géométriques et électromagnétiques fournissent des estimations de portée en conditions idéalisées, tandis que les modèles empiriques et semi-déterministes, dont ceux proposés par l'ITU, intègrent davantage de phénomènes et constituent des standards largement adoptés pour la prévision de la propagation radio. Toutefois, leur capacité à décrire fidèlement la réception effective de l'AIS demeure limitée. D'une part, ils reposent sur des hypothèses simplificatrices. D'autre part, ils peinent à capturer l'effet de l'exhaustivité des phénomènes de propagation à une granularité satisfaisante, par exemple, ils se basent sur des statistiques régionales des variations d'indices de réfraction plutôt que sur des données ou prévisions météorologiques instantanées.

Ces limites, présentées dans le Tableau 1.8, rendent difficile l'utilisation exclusive de modèles physiques pour la prédiction des zones de réception de l'AIS, en particulier lorsque l'on cherche à détecter des anomalies ou à caractériser des zones non couvertes. Dans ce contexte, l'exploitation directe des données AIS et environnementales s'impose comme une alternative pertinente. En adoptant une approche dirigée par les données, il devient possible de tirer parti du volume massif d'informations disponibles pour apprendre les relations complexes et non linéaires entre conditions réelles et propagation, sans recourir à une modélisation théorique ou simplificatrice.

Le chapitre suivant introduira les méthodes d'apprentissage automatique et profond appliquées à la modélisation de la propagation radio et comment ces approches peuvent dépasser les limites des approches classiques.

Chapitre 2 : Modélisation de la propagation des ondes radio par apprentissage automatique et profond

La modélisation de la propagation des ondes électromagnétiques est un défi central pour la modélisation des systèmes de télécommunication. Comme présenté dans le chapitre précédent, les modèles physiques classiques : déterministes, empiriques ou semi-empiriques fournissent des estimations dans des contextes idéalisés ou bien maîtrisés. Cependant, leur capacité à s'adapter aux environnements réels, hétérogènes et dynamiques demeure limitée (Seretis et Sarris 2021 ; Vasudevan et Yuksel 2024). Face à ces contraintes, une approche alternative a émergé grâce à l'abondance de données disponibles aujourd'hui et à l'essor de méthodes d'apprentissage automatique récentes, efficaces et adaptées : la modélisation dirigée par les données.

Ce chapitre propose d'explorer cette transition méthodologique majeure. Il présente dans un premier temps les fondements de l'apprentissage automatique appliqué à la modélisation de la propagation, en distinguant les grandes familles d'approches (supervisé, non supervisé, semi-supervisé, renforcement), puis en analysant les spécificités de la démarche dirigée par les données, ses exigences en matière de données et ses principales méthodes. Une deuxième partie est consacrée aux apports de l'apprentissage profond, qui a permis de dépasser certaines limites structurelles des approches classiques, en particulier pour la représentation des variables caractéristiques. Enfin, la dernière partie introduit les réseaux de neurones en graphes, une famille de modèles particulièrement adaptés aux environnements où les interactions topologiques entre entités jouent un rôle déterminant dans la propagation des signaux.

L'objectif de ce chapitre est d'exposer les fondements de l'apprentissage automatique appliqué à la modélisation de la propagation des ondes électromagnétiques, en mettant en avant les avancées de l'apprentissage profond et des réseaux de neurones en graphes pour ainsi introduire les spécificités des méthodes qui seront mobilisées dans les chapitres suivants.

2.1 Apprentissage automatique pour la modélisation de la propagation

Les modèles classiques de propagation fournissent des estimations basées sur des formules empiriques dérivées de campagnes de mesures spécifiques ou de formulations analytiques déterministes. Bien qu'utiles dans des contextes généraux, ces modèles montrent des limites dès lors que l'on souhaite représenter des environnements complexes ou spécifiques (Seretis et Sarris 2021 ; Chiroma et al. 2023).

L'apprentissage automatique constitue une alternative particulièrement adaptée à ce problème. En s'appuyant sur des données d'observation (issues de mesures de terrain, de simulations ou de capteurs), les modèles ML peuvent apprendre à prédire les caractéristiques du canal radio sans recourir à des hypothèses physiques. Ils permettent ainsi :

- De s'adapter automatiquement à de nouvelles configurations ou domaines par le concept de généralisabilité (Akrouit et al. 2023).
- De tenir compte de données multi-sources (Baltrušaitis et al. 2019).
- D'évoluer avec le temps, en réapprenant à partir de nouvelles données grâce aux approches de transfert d'apprentissage et d'ajustement (Wong et Michaels 2022).

De plus, la capacité des modèles issus de l'apprentissage automatique à modéliser des relations non linéaires permet d'obtenir une meilleure précision (Seretis et Sarris 2021 ; Vasudevan et Yuksel 2024 ; Bocus et Lodhi 2025). Toutefois, cette puissance nécessite un volume de données conséquent et une attention particulière portée à leur qualité (Vasudevan et Yuksel 2024). L'essor des techniques d'apprentissage automatique ouvre une nouvelle voie pour la modélisation des phénomènes de propagation radio. En s'affranchissant des limites rigides des approches analytiques, ces méthodes permettent d'élaborer des modèles plus souples et adaptatifs. L'enjeu devient alors non seulement de choisir les bons algorithmes, mais aussi de structurer et d'exploiter efficacement les données disponibles.

Cette section vise à poser les bases théoriques et pratiques de l'apprentissage automatique appliqué à la modélisation de la propagation radio. Elle expose les principaux paradigmes d'apprentissage, les principes de la modélisation dirigée par les données, les exigences en matière de données, ainsi que les méthodes classiques utilisées dans la littérature. Cette revue établit un socle pour comprendre les limites de ces méthodes et introduire dans les sections suivantes les avancées récentes issues de l'apprentissage profond.

2.1.1 Types d'apprentissage

Dans le domaine de la modélisation de la propagation radio par l'apprentissage automatique, le choix du paradigme d'apprentissage joue un rôle central. Il détermine la manière dont le modèle est entraîné à partir des données disponibles et conditionne les types de tâches qu'il peut accomplir. On distingue principalement trois grandes familles : l'apprentissage supervisé, l'apprentissage non supervisé, et l'apprentissage semi-supervisé. À cela s'ajoute, dans certains cas où entre en jeu un agent, l'apprentissage par renforcement.

La Figure 2.1 offre un aperçu de ces différents paradigmes.

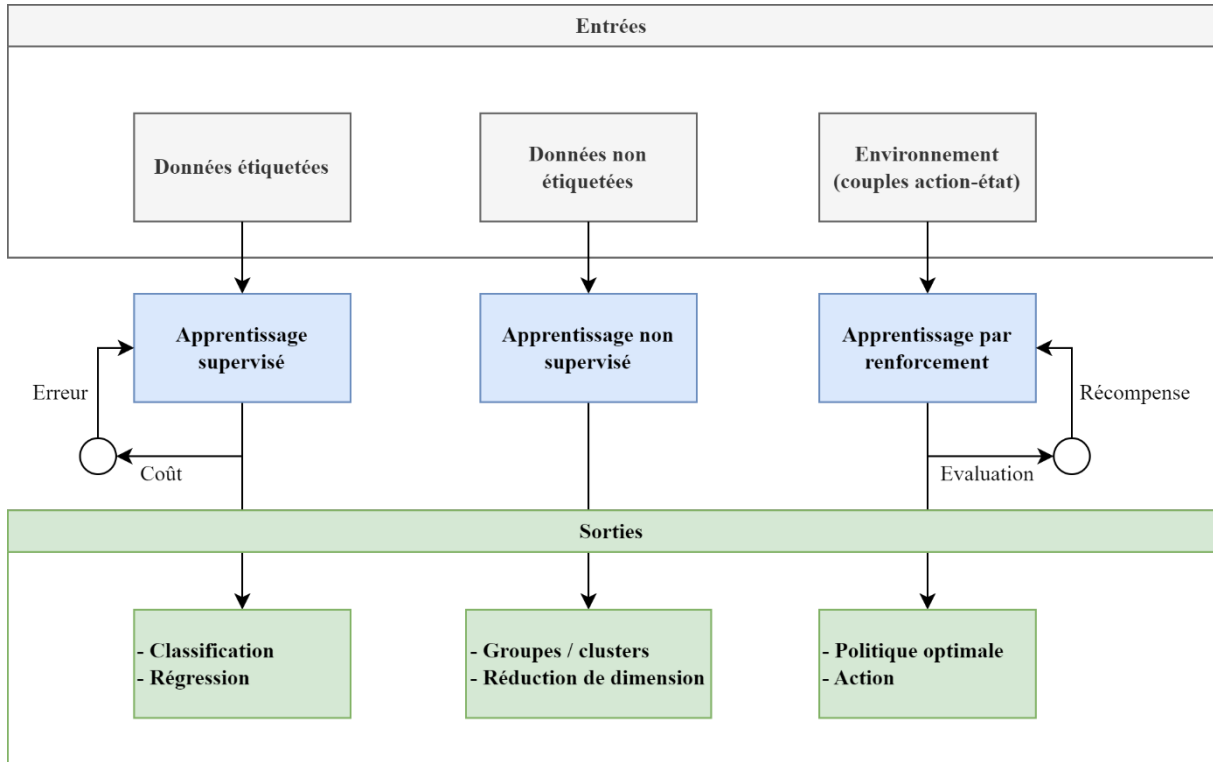


Figure 2.1 : Les trois grands paradigmes d'apprentissage automatique, leurs entrées typiques et sorties.

2.1.1.1 Apprentissage supervisé

L'apprentissage supervisé repose sur l'exploitation d'un ensemble de données étiquetées, où chaque vecteur d'entrée est associé à une sortie attendue. L'ensemble (ou jeu de données) d'apprentissage est alors défini comme suit :

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N, \quad x_i \in \mathbb{R}^d, y_i \in \mathcal{Y}$$

Où x_i représente un vecteur de caractéristiques, y_i la sortie correspondante et N le nombre d'échantillons. L'objectif est d'apprendre une fonction de mapping entre les entrées et les sorties :

$$f_\theta: \mathbb{R}^d \rightarrow \mathcal{Y}, \quad f_\theta(x_i) \approx y_i$$

de manière à généraliser correctement sur de nouveaux échantillons non observés lors de l'entraînement (Bishop 2006).

Ce type d'apprentissage est le plus fréquemment utilisé dans les travaux liés à la propagation radio, car il permet de prédire des variables mesurables sur le terrain, comme le niveau de signal reçu (Liu et al. 2022) ou la perte de propagation (Yan Zhang et al. 2019). Pour ce faire, le modèle est entraîné par minimisation d'une fonction de perte :

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^N l(f_\theta(x_i), y_i)$$

Où $l(\cdot)$ mesure l'écart entre la sortie prédite et la sortie réelle.

Parmi les algorithmes couramment mobilisés, figurent la régression linéaire, les forêts aléatoires, ainsi que les réseaux de neurones artificiels (ANN : *Artificial Neural Networks*). Ces méthodes diffèrent par leur capacité de modélisation et leur robustesse face au bruit, mais toutes dépendent fortement de la quantité et de la représentativité des données disponibles (Vasudevan et Yuksel 2024).

2.1.1.2 Apprentissage non supervisé

Dans l'apprentissage non supervisé, le modèle ne dispose pas de sorties cibles. L'ensemble de données est donc uniquement constitué des vecteurs d'entrée :

$$\mathcal{D} = \{(x_i)\}_{i=1}^N, \quad x_i \in \mathbb{R}^d$$

L'objectif est d'explorer les structures intrinsèques présentes dans ces données, en les regroupant ou en mettant en évidence des corrélations. De manière générale, il s'agit de construire une fonction de représentation :

$$g_\theta: \mathbb{R}^d \rightarrow \mathbb{R}^k$$

Permettant soit de réduire la dimensionnalité, soit de projeter les données dans un espace révélant leur organisation interne.

Ce paradigme est utilisé, par exemple, pour regrouper des zones géographiques afin d'entraîner des modèles spécifiques à chaque *cluster* (Moreta et al. 2019), détecter et supprimer des valeurs aberrantes dans des campagnes de mesures (Ye et Wang 2018), ou encore réduire la dimensionnalité des données afin de révéler des structures latentes (Wang et Ferrús 2021). Les méthodes emblématiques incluent le *clustering* (K-means (Lloyd 1982), SOM (Kohonen 1982)) et les techniques de réduction de dimensionnalité telles que l'analyse en composantes principales (ACP) (Pearson 1901). Ces approches constituent souvent une étape préalable à l'apprentissage supervisé ou semi-supervisé, en fournissant une meilleure structuration des données brutes et en réduisant la complexité du problème de modélisation (O'Shea et al. 2017).

2.1.1.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé se situe à l'interface entre l'apprentissage supervisé et non supervisé. Il exploite simultanément un petit ensemble de données étiquetées et un grand volume de données non étiquetées pour améliorer les performances du modèle (Chapelle et al. 2009). Cette approche conduit aux définitions suivantes :

$$\mathcal{D}_l = \{(x_i, y_i)\}_{i=1}^n, \quad \mathcal{D}_u = \{(x_j)\}_{j=n+1}^N, \quad n \ll N$$

Où \mathcal{D}_l représente l'ensemble des données étiquetées, \mathcal{D}_u celui des données non étiquetées, et N le nombre total d'échantillons disponibles. L'idée principale est que les données non étiquetées contiennent des informations utiles sur la structure sous-jacente de l'espace des variables, permettant de guider et de contraindre l'apprentissage. L'optimisation combine alors une perte supervisée (basée sur \mathcal{D}_l) et une perte non supervisée (basée sur \mathcal{D}_u) :

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{sup}}(\theta; \mathcal{D}_l) + \lambda \mathcal{L}_{\text{nsup}}(\theta; \mathcal{D}_u)$$

Où λ est un coefficient de pondération qui contrôle l'importance des données non étiquetées dans l'entraînement.

Dans le domaine de la propagation et plus largement des télécommunications, ce paradigme est particulièrement pertinent. En effet, la collecte et l'annotation de données étiquetées sont coûteuses et chronophages, tandis que les données brutes non étiquetées sont abondantes et relativement faciles à acquérir (O'Shea et al. 2017). Les travaux récents montrent plusieurs applications typiques : extension de modèles supervisés à de nouveaux environnements grâce à un pré-entraînement non supervisé (O'Shea et al. 2017 ; Zhang et Saad 2021 ; Berthelot et al. 2022), détection automatique de zones de mauvaise couverture avec peu de données annotées (Al-Ahmed et Shakir 2022), et régularisation par consistance afin d'exploiter au mieux les données non étiquetées (Wang et al. 2024).

Les méthodes mobilisées dans l'apprentissage semi-supervisé se regroupent en plusieurs grandes familles :

- Approches par self-training : où le modèle s'entraîne progressivement sur des pseudo-étiquettes générées à partir des données non annotées ou bruitées (Nguyen et al. 2020 ; Amini et al. 2025).
- Approches basées sur la consistance : qui imposent que les prédictions restent stables face à des perturbations des données d'entrée (Tarvainen et Valpola 2017).
- Méthodes de propagation d'étiquettes : qui diffusent l'information d'un petit ensemble étiqueté vers les échantillons non annotés dans un graphe de similarité (Iscen et al. 2019).
- Méthodes basées sur les graphes : qui exploitent la structure topologique des relations entre échantillons pour guider l'apprentissage (Kipf et Welling 2017).

L'apprentissage semi-supervisé permet ainsi de tirer parti de la masse de données disponibles à faible coût tout en minimisant l'effort d'annotation, ce qui en fait un levier particulièrement adapté aux problématiques de modélisation en télécommunications.

2.1.1.4 Apprentissage par renforcement

L'apprentissage par renforcement (RL : *Reinforcement Learning*) est une branche de l'apprentissage automatique où un agent apprend à prendre des décisions en interagissant avec un environnement. Contrairement à l'apprentissage supervisé, qui repose sur des données étiquetées, le RL repose sur un système de récompense : l'agent explore différentes actions et reçoit des retours selon leur impact, afin de maximiser une fonction de récompense cumulée (Sutton et Barto 2018).

Bien qu'il soit utilisé dans des domaines comme l'allocation des ressources et des canaux ou la réduction des interférences (Aldossari et Chen 2019 ; Seretis et Sarris 2021), ce paradigme est encore peu exploité dans le contexte de la modélisation de la propagation, mais il offre des perspectives intéressantes (Seretis et Sarris 2021).

2.1.2 Défis liés aux données pour la modélisation dirigée par les données

Une approche dirigée par les données (*data-driven*) repose sur l'exploitation directe des données d'observation pour construire un modèle prédictif, sans imposer explicitement de règles ou d'équations physiques préétablies. Contrairement aux approches traditionnelles fondées sur la physique (modèles déterministes ou semi-déterministes), la modélisation dirigée par les données ne cherche pas à décrire les lois de la propagation à partir de formules analytiques, mais à les inférer statistiquement à partir d'exemples empiriques.

Dans le contexte de la propagation radio, cela signifie par exemple prédire des grandeurs comme la puissance reçue ou les pertes de propagation à partir d'un grand volume de mesures issues du terrain, de simulations ou de sources ouvertes, sans chercher à formuler une équation explicite reliant tous ces facteurs. Cependant, cette approche est conditionnée par la capacité à relever plusieurs défis méthodologiques inhérents aux données utilisées. Les sections suivantes détaillent ces principaux enjeux.

2.1.2.1 Qualité, représentativité et volume des données

Les performances des modèles d'apprentissage automatique dépendent fortement de la qualité, de la quantité et de la représentativité des données utilisées pour leur entraînement (Gudivada et al. 2017 ; Seretis et Sarris 2021 ; Mohammed et al. 2025). Dans le contexte de la modélisation de la propagation radio, plusieurs exigences spécifiques doivent être prises en compte pour garantir la fiabilité des prédictions (Santra et al. 2025).

Un premier enjeu concerne la qualité et la précision des données. Celles-ci doivent refléter fidèlement les conditions réelles du canal radio à modéliser. Des erreurs de mesure, des données incomplètes ou bruitées peuvent fortement dégrader la performance des modèles (Gudivada et al. 2017 ; Mohammed et al. 2025). Un deuxième aspect réside dans l'homogénéité spatiale et la couverture des mesures. Les modèles d'apprentissage, notamment dans les tâches de prédiction de couverture ou de

perte de signal, nécessitent une couverture spatiale suffisamment équilibrée. Une répartition géographiquement déséquilibrée peut conduire à un biais qui limite la capacité prédictive du modèle (Zhang et al. 2025 ; Zeng et al. 2025). Un troisième défi porte sur le prétraitement des données (*preprocessing*). Avant d'être utilisées, les données nécessitent souvent un traitement préalable qui influence directement la robustesse et la performance du modèle final (Hastie et al. 2009a). Enfin, la question du biais et de la représentativité reste centrale. Un modèle entraîné uniquement sur certains environnements risque de produire des prédictions biaisées lorsqu'il est appliqué à d'autres contextes (Hastie et al. 2009b). De plus, certaines approches exigent des volumes massifs de données pour éviter le surapprentissage, ce qui souligne l'importance du volume et de la diversité des échantillons mobilisés (Yan Zhang et al. 2019).

2.1.2.2 Diversité et sources de données mobilisées

La modélisation de la propagation radio repose sur l'exploitation de sources de données multiples, provenant de mesures, de simulations ou d'informations environnementales. La diversité et la complémentarité de ces données jouent un rôle central dans la capacité des modèles à capturer la complexité des phénomènes radio (Vasudevan et Yuksel 2024).

Les données de mesure constituent une première source essentielle, puisqu'elles permettent d'enregistrer la puissance du signal reçu (Mazzarella et al. 2017 ; Belmonte-Hernández et al. 2020 ; Luo et Meratnia 2022 ; Bufort et al. 2023), la phase (Wang et al. 2015), les pertes de propagation (Faruk et al. 2021) ou les paramètres du canal (Li et al. 2017 ; Kram et al. 2019 ; Molina et al. 2025) dans des environnements réels.

Les données issues de simulations radio représentent un second apport. Les outils de simulation électromagnétique, tels que le *ray-tracing* (Yun et Iskander 2015 ; Marey et al. 2022 ; Bakirtzis et al. 2022 ; Bufort et al. 2023) ou les modèles physiques (Wen et al. 2019) permettent de générer des données synthétiques, utiles notamment dans des contextes où les mesures réelles sont coûteuses ou difficiles à obtenir.

Un troisième type de données regroupe les informations cartographiques et environnementales, telles que la topographie via les données SRTM (NASA JPL 2013), les bases BD ALTI⁴ ou BD TOPO⁵, la couverture du sol (Copernicus Climate Change Service 2019), les données météorologiques via ERA5 (C3S 2018a) ou encore les images satellites (Google Earth⁶, Landsat⁷, Sentinel-2⁸). Ces sources,

⁴ <https://geoservices.ign.fr/bdalti>

⁵ <https://geoservices.ign.fr/bdtopo>

⁶ <https://earth.google.com/>

⁷ <https://landsat.gsfc.nasa.gov/>

⁸ <https://dataspace.copernicus.eu/data-collections/copernicus-sentinel-data/sentinel-2>

disponibles via des bases publiques ou commerciales, enrichissent la modélisation en apportant des connaissances contextuelles.

Enfin, les données de réseau complètent cet éventail. Liées aux protocoles de communication utilisés, elles sont parfois accompagnées de métadonnées (horodatage, position, caractéristiques de l'émetteur/récepteur). Bien qu'elles reflètent des usages réels, elles peuvent souffrir de bruit, d'erreurs ou d'imprécision de localisation (Harati-Mokhtari et al. 2007 ; Iphar et al. 2020).

2.1.2.3 Fusion de données hétérogènes

Dans le domaine de la modélisation de la propagation radio, les données disponibles proviennent de sources variées, chacune offrant une perspective partielle sur le phénomène étudié. La fusion de données hétérogènes consiste à combiner ces sources complémentaires afin d'améliorer la précision des modèles prédictifs (Seretis et Sarris 2021). La fusion peut s'opérer à différents niveaux :

- Fusion au niveau des données brutes (*early fusion*) (Snoek et al. 2005) : les différentes sources sont concaténées ou alignées pour former un vecteur d'entrée unique. Cette approche nécessite que les données soient synchronisées et normalisées, ce qui pose des défis lorsque les résolutions spatiales ou temporelles diffèrent (Sotiroudis et al. 2021) .
- Fusion au niveau des décisions (*late fusion*) (Snoek et al. 2005) : plusieurs modèles sont entraînés séparément sur des sources spécifiques, puis leurs sorties sont combinées. Cette méthode est plus souple, mais peut perdre certaines interactions croisées entre les sources (Gadzicki et al. 2020).

La fusion de données hétérogènes soulève plusieurs difficultés techniques. Il s'agit notamment de l'alignement spatial et temporel des données (et le risque de dérive de horloges quand l'horodatage n'est pas réalisé par un système centralisé), des différences de granularité et de résolution, de la présence de bruit ou de valeurs manquantes, ou encore de l'hétérogénéité des formats. À cela s'ajoute la question du poids relatif des sources, puisque toutes n'ont pas la même fiabilité ni la même importance (Atrey et al. 2010 ; Baltrušaitis et al. 2019 ; Sotiroudis et al. 2021). La fusion de sources hétérogènes permet d'améliorer significativement la précision dans le domaines la propagation radio (Kumar Yadav et al. 2020 ; Sotiroudis et al. 2021) . Elle permet également d'enrichir les modèles, en apportant des informations qui complètent les données de mesures (Wen et al. 2019).

2.1.3 Méthodes d'apprentissage automatique non profond

De nombreuses études ont exploré l'application des techniques d'apprentissage automatique classiques pour la modélisation de la propagation radio (Seretis et Sarris 2021 ; Vasudevan et Yuksel 2024). Bien que ces méthodes précèdent l'essor des approches par apprentissage profond, elles ont

largement contribué à démontrer la faisabilité et l'intérêt de l'apprentissage statistique dans ce domaine. Elles sont notamment appréciées pour leur faible complexité computationnelle (Lim et al. 2000), leur interprétabilité et leur efficacité dans des contextes où les données disponibles sont limitées.

Les principales méthodes classiques mobilisées pour la modélisation de la propagation radio incluent la régression par vecteur de support (Drucker et al. 1996 ; Wen et al. 2019 ; Vergos et al. 2021 ; Elmezughi et al. 2022), les arbres de décision (Quinlan 1986 ; Elmezughi et al. 2022), les k plus proches voisins (Fix et Hodges 1989 ; Moreta et al. 2019), ainsi que les approches d'ensemble comme AdaBoost (Freund et Schapire 1997) ou XGBoost (Chen et Guestrin 2016) appliqués notamment dans (Vergos et al. 2021 ; Vanche et al. 2022).

2.1.4 Limites de ces approches

Malgré leur utilité démontrée dans de nombreux travaux en lien avec les télécommunications radio maritimes et le système AIS (Mazzarella et al. 2017 ; Zhou et al. 2019 ; Gao et Shi 2020 ; Vanche et al. 2022 ; Huang et al. 2024), les méthodes classiques d'apprentissage automatique présentent plusieurs limitations qui restreignent leur efficacité dans des scénarios complexes ou à grande échelle.

Tout d'abord, la performance de ces modèles dépend fortement de la qualité de la représentation des données en entrée. Ces méthodes ne sont généralement pas capables d'extraire automatiquement des représentations hiérarchiques ou abstraites à partir de données brutes. Elles requièrent donc un travail préalable d'ingénierie des caractéristiques (*feature engineering*), souvent coûteux et dépendant de l'expertise du domaine (Çayir et al. 2018).

Par ailleurs, la capacité de modélisation non linéaire des modèles classiques est limitée. Bien que certains algorithmes comme les SVM ou les arbres de décision permettent d'approcher des fonctions complexes, ils atteignent rapidement leurs limites lorsqu'il s'agit de capturer des interactions hautement non linéaires ou de traiter des structures de données complexes (Goodfellow et al. 2016).

Un autre frein important réside dans la difficulté d'adaptation de ces modèles à des données de grande dimension ou massives, typiques des applications modernes en télécommunications. Les performances de modèles comme k-NN peuvent se dégrader significativement en présence de données très volumineuses ou fortement bruitées (Frenay et Verleysen 2014 ; Halder et al. 2024).

Enfin, ces méthodes sont peu adaptées à l'apprentissage multimodal, où différentes sources de données doivent être combinées. Leur architecture rigide rend difficile l'intégration simultanée de plusieurs types de modalités de données (Baltrušaitis et al. 2019).

Ces limites ont progressivement conduit la communauté scientifique à se tourner vers des approches plus flexibles, capables de gérer de grandes quantités de données hétérogènes et de découvrir automatiquement des représentations optimales à différentes échelles. C'est dans ce contexte

que les techniques d'apprentissage profond ont connu un essor rapide, offrant des perspectives nouvelles pour la modélisation des phénomènes radio complexes.

2.2 Apprentissage profond pour la modélisation de la propagation

Au cours des dernières années, l'apprentissage profond (DL : *Deep Learning*) s'est imposé comme une extension naturelle du *machine learning* classique, grâce à sa capacité à extraire automatiquement des représentations non linéaires à partir de données brutes (LeCun et al. 2015 ; Goodfellow et al. 2016). Parmi les approches les plus emblématiques du DL, les réseaux de neurones artificiels se distinguent par leur aptitude à modéliser des relations complexes sans recourir à des hypothèses a priori sur les mécanismes sous-jacents à la propagation. Ils permettent également de traiter efficacement des données de haute dimension (Chiroma et al. 2023). De manière générale, un réseau de neurones à L couches définit une fonction paramétrée f_θ , où θ désigne l'ensemble des paramètres du modèle, telle que :

$$f_\theta(h^{(L)}) \text{ avec } h^{(l)} = \sigma(W^{(l)}h^{(l-1)} + b^{(l)}), \quad 1 \leq l \leq L$$

Où $h^{(0)} = x \in \mathbb{R}^d$ est un vecteur d'entrée, $W^{(l)}$ et $b^{(l)}$ sont respectivement la matrice de poids et le biais de la couche l , $\sigma(\cdot)$ est une fonction d'activation non linéaire, et $h^{(L)}$ correspond à la sortie du réseau.

Cette capacité est particulièrement précieuse dans le contexte de la propagation des ondes radio, où les interactions avec l'environnement sont souvent de nature fortement non linéaire. L'un des atouts majeurs de l'apprentissage profond réside dans sa faculté de *feature learning*, c'est-à-dire l'apprentissage automatique de représentations hiérarchiques des caractéristiques pertinentes. Ce mécanisme permet aux réseaux de neurones de capturer les facteurs déterminants dans les processus de propagation des ondes (Vasudevan et Yuksel 2024).

En outre, les réseaux neuronaux sont capables de traiter des données d'entrée beaucoup plus riches, comme :

- Des images satellites ou cartes *raster* (via des réseaux convolutifs) (Thrane et al. 2020 ; Hayashi et al. 2020).
- Des séries temporelles de mesures (avec des réseaux récurrents comme LSTM) (Elmezughi et al. 2022 ; Ji et al. 2024).
- Ou des représentations multimodales (Thrane et al. 2020).

Cette flexibilité permet une modélisation beaucoup plus complète, tout en restant strictement dirigée par les données. Le passage à l'apprentissage profond ouvre ainsi la voie à une nouvelle génération de modèles plus puissants.

Cette section dresse un panorama des principales approches d'apprentissage profond avant d'en analyser les limites.

2.2.1 Principales architectures

De nombreuses revues de la littérature ont mis en évidence l'intérêt croissant des techniques d'apprentissage profond pour la modélisation de phénomènes complexes, notamment dans le domaine de la propagation radio (Seretis et Sarris 2021 ; Mladenović et al. 2022 ; C. Huang et al. 2022 ; Chiroma et al. 2023 ; Vasudevan et Yuksel 2024). Ces méthodes se distinguent par leur capacité à extraire automatiquement des représentations pertinentes à partir de données hétérogènes et à modéliser des relations non linéaires souvent difficiles à capturer par des approches classiques.

Les principales architectures explorées incluent d'abord les perceptrons multicouches (MLP), appliqués à la prédiction de la perte de chemin ou des paramètres de canal (Isabona et al. 2022), ensuite les réseaux convolutifs (CNN), qui exploitent la structure spatiale des données et ont montré leur efficacité pour produire des cartes de propagation à partir de paramètres cartographiques (Imai et al. 2019 ; Zhang et al. 2020 ; Levie et al. 2021), ainsi que les réseaux récurrents (RNN) de type LSTM (Hochreiter et Schmidhuber 1997) ou l'unité récurrente à portes (GRU : *Gated Recurrent Unit*) (Cho et al. 2014), adaptés au traitement séquentiel pour modéliser des évolutions temporelles ou spatiales (Wang et al. 2020 ; L. Liu et al. 2021 ; Sasaki et al. 2021 ; Belmonte-Hernández et al. 2020 ; Pohontu et al. 2024). Des combinaisons multimodales de ces architectures, telles que CNN–RNN (Wang et al. 2020) ou CNN–MLP (Thrane et al. 2020), ont également été proposées pour intégrer simultanément cartes, séries temporelles et données vectorielles.

Par ailleurs, des approches non supervisées ou génératives, comme les auto-encodeurs et les réseaux antagonistes génératifs (GAN : *Generative Adversarial Networks*) (Goodfellow et al. 2020), sont utilisées pour la génération (Marey et al. 2022 ; Juang 2022) ou la reconstruction de cartes de propagation (Belmonte-Hernández et al. 2020), l'augmentation de résolution (Seretis et al. 2021) ou la création de données synthétiques (Li et al. 2021).

Malgré leurs capacités remarquables à modéliser des phénomènes complexes et à exploiter des données non structurées, les méthodes d'apprentissage profond classiques présentent plusieurs limitations.

2.2.2 Un pouvoir de représentation limité

Les modèles d'apprentissage profond classiques, tels que les CNN, sont conçus pour traiter des données euclidiennes, c'est-à-dire des données organisées sur des structures régulières comme des vecteurs (séries temporelles 1D), des images (matrices 2D de pixels) ou des volumes 3D. Cette structure impose une organisation régulière et ordonnée des données. Or, les phénomènes de propagation radio s'inscrivent dans des environnements complexes, où les relations spatiales ne

suivent pas une grille fixe (Wu et al. 2021 ; Luo et Meratnia 2022). Dans ces contextes, les données sont souvent définies sur des structures irrégulières, influencées par la géométrie des lieux ou les interactions entre entités. Les modèles euclidiens échouent à capturer ces relations.

Ces méthodes sont également limitées lorsqu'elles sont confrontées à des données partielles ou irrégulièrement échantillonnées. Les modèles classiques, notamment CNN ou MLP, exigent des entrées denses et homogènes pour apprendre efficacement. Cette dépendance à des données bien structurées rend leur généralisation difficile lorsque l'information est dispersée ou bruitée, ce qui est fréquent dans les systèmes réels (Redondi 2018 ; Zhou et al. 2020 ; Luo et Meratnia 2022).

Enfin, les architectures traditionnelles imposent une taille fixe pour les données d'entrée. Cela contraint fortement leur capacité à s'adapter à des structures dynamiques. Dans les réseaux sans fil, le nombre de points d'accès, la densité des nœuds ou la topologie du réseau peuvent varier dans le temps. Ces changements rendent inopérants les modèles qui exigent des dimensions constantes, comme c'est le cas pour la majorité des réseaux convolutifs (Zhou et al. 2020 ; Wu et al. 2021 ; S. He et al. 2021).

Ces limitations ont conduit à l'émergence de modèles conçus spécifiquement pour des données structurées en graphes. Les graphes permettent de représenter explicitement les entités du système (nœuds) et leurs relations (arêtes), qu'elles soient physiques, géométriques ou logiques. Les réseaux de neurones en graphes exploitent cette structure pour propager et agréger l'information de manière relationnelle, ce qui en fait une approche particulièrement bien adaptée aux défis posés par la modélisation de la propagation radio. Dans la section suivante, ces modèles, leurs avantages, et les méthodes spécifiques les plus pertinentes seront détaillées.

2.3 Réseaux de neurones en graphes pour la modélisation de la propagation

Les méthodes d'apprentissage profond classiques ont démontré leur efficacité pour extraire automatiquement des représentations non linéaires à partir de données structurées, telles que les séries temporelles ou les images. Toutefois, ces architectures restent fondées sur une représentation des données inadaptée à de nombreuses situations où les interactions entre entités ne suivent pas une grille régulière, mais une topologie complexe. Cette limitation est particulièrement marquée dans le domaine de la propagation radio, où les phénomènes physiques sont conditionnés par des relations entre entités hétérogènes.

Dans ce contexte, les réseaux de neurones en graphes (GNN) offrent un cadre formel et computationnel puissant. En s'appuyant sur une représentation explicite des données sous forme de graphes, ces modèles permettent de capturer les dépendances structurelles et les interactions au sein

du système. Cette section présente les fondements de la représentation en graphe appliquée à la propagation, les différents types de GNN utilisés dans la littérature, ainsi que leurs applications concrètes pour la prédiction de couverture radio ou de pertes de signal. Elle prépare ainsi l'introduction de l'architecture neuronale proposée au chapitre suivant.

2.3.1 Représentation des données sous forme de graphe

L'une des fondations de l'utilisation des réseaux de neurones en graphes pour modéliser la propagation des ondes repose sur la manière dont les données du problème sont représentées sous forme de graphe. À la différence des approches traditionnelles qui se basent sur des vecteurs d'attributs isolés, les GNN exploitent la structure relationnelle entre les entités du système, ce qui implique de formuler explicitement la connectivité du domaine étudié.

Dans ce cadre, un graphe est défini comme un ensemble de nœuds (ou sommets) et d'arêtes (ou liens), éventuellement orientés et/ou pondérés, auquel on peut associer des attributs. Cette structure permet de modéliser de manière naturelle les configurations spatiales, topologiques ou physiques du canal de propagation.

2.3.1.1 Graphes de propagation

Le graphe de propagation (Pedersen et al. 2012) est une modélisation mathématique adaptée à la représentation de la propagation des ondes dans les communications radio. Elle permet de représenter le flux de l'émetteur au récepteur par un graphe simple orienté, ainsi que son interaction avec les éléments diffuseurs. Ce graphe $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ est représenté à l'aide de ses nœuds \mathcal{V} et de ses arêtes \mathcal{E} . Par définition du graphe de propagation, l'ensemble des nœuds $\mathcal{V} = \mathcal{V}_t \cup \mathcal{V}_s \cup \mathcal{V}_r$ est composé de trois sous-ensembles : les nœuds correspondants à des émetteurs, récepteurs ou diffuseurs. Une arête $e = (v_1, v_2)$ appartient à \mathcal{E} si et seulement si le signal émis par le nœud v_1 peut se propager au nœud v_2 .

Le diagramme de graphe de propagation est un objet permettant de construire l'ensemble des graphes de propagation en schématisant les liens entre nœuds et matrices de transfert. La Figure 2.2 présente le diagramme de flux pour la construction d'un graphe de propagation. Aux différents nœuds sont respectivement décrits, le vecteur d'entrée représentant le signal \mathbf{X} , le signal à la sortie des diffuseurs \mathbf{Z} , le signal de sortie \mathbf{Y} . Les matrices de transfert associées aux différentes arêtes sont :

- **D** reliant \mathcal{V}_t à \mathcal{V}_r : la propagation directe.
- **R** reliant \mathcal{V}_s à \mathcal{V}_r : depuis un diffuseur vers un récepteur.
- **T** reliant \mathcal{V}_t à \mathcal{V}_s : depuis un émetteur vers un diffuseur.
- **B** reliant \mathcal{V}_s entre eux : la propagation entre diffuseurs.

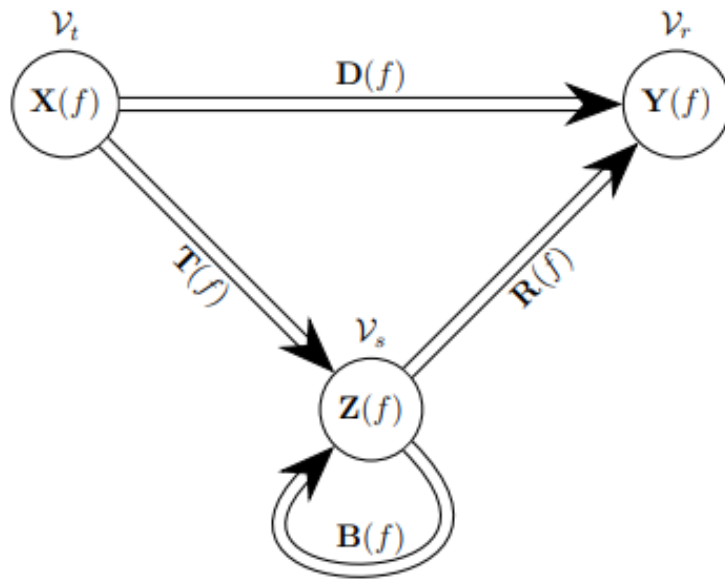


Figure 2.2 : Diagramme de flux de la représentation d'un graphe de propagation (Pedersen et al. 2012)

2.3.1.2 Définition des nœuds

Les nœuds du graphe représentent les entités spatiales ou physiques impliquées dans la propagation. Selon le cas d'usage, il peut s'agir de :

- Points d'émission.
- Points de réception.
- Éléments d'infrastructure ou d'environnement.
 - Cellules géographiques.

Chaque nœud peut être associé à un vecteur de caractéristiques décrivant ses propriétés.

2.3.1.3 Définition des arêtes

Les arêtes décrivent les relations entre les entités. Elles peuvent être définies selon plusieurs critères :

- Relation de visibilité (LoS/NLoS (Line of Sight / Non Line of Sight)) entre deux points (Zhang et al. 2024 ; Bufort et al. 2023).
- Proximité géographique (Zhang et al. 2024 ; Bufort et al. 2023).
- Connectivité électromagnétique (Suto et al. 2023).
- Similarité contextuelle (Liu et al. 2020 ; 2021).

Selon les besoins (S. He et al. 2021) , le graphe peut être :

- Statique (pour une zone fixe à un instant donné).
- Dynamique (évolutif dans le temps, en fonction du déplacement des utilisateurs, des changements d’environnement ou de configuration réseau).

2.3.1.4 Exemples de modélisation de la propagation radio fondée sur les graphes

La représentation des données ou de l’environnement sous forme de graphe est couramment utilisée afin d’obtenir un modèle propagation radio, capable d’intégrer les interactions complexes entre les éléments de l’environnement. Cette approche permet non seulement de capturer les phénomènes physiques tels que la diffraction, la réflexion ou la diffusion, mais également d’exploiter des techniques computationnelles efficaces, qu’elles soient déterministes ou dirigées par les données et l’apprentissage automatique, pour simuler ou prédire le comportement du signal dans des environnements complexes.

Dans (Liu et al. 2020) la modélisation de la propagation repose sur une représentation en graphe, conçue spécifiquement pour simuler les effets de diffraction multiple dans des environnements urbains. Chaque nœud du graphe correspond à un point de diffraction (dans ce contexte d’application, un bord de bâtiment), et les arêtes dirigées représentent les trajets potentiels du signal entre ces points, selon un critère de visibilité géométrique et physique. Cette structure permet de modéliser de manière déterministe les trajectoires diffractées à travers plusieurs obstacles, en appliquant le modèle en lame de couteau à chaque saut. L’approche est centrée sur une exploration systématique du graphe pour identifier les chemins de propagation pertinents sans recours à des techniques d’apprentissage. Une approche similaire (Liu et al. 2021) propose de simuler les phénomènes de propagation par chemins multiples en modélisant les surfaces réfléchissantes et diffusantes. Cela nécessite la définition de nouveaux types de nœuds et la modélisation sous forme de graphe de leurs interactions.

(Suto et al. 2023) introduisent le concept de *Propagation Graph Representation Learning*, basé sur l’apprentissage automatique et les graphes. Il vise à prédire l’atténuation du signal et l’effet de masquage dans les environnements urbains. Le processus de propagation est ainsi représenté comme un graphe où les nœuds sont les émetteurs, récepteurs et obstacles, et les arêtes modélisent les conditions de propagation entre ces nœuds. Chaque arête est associée à une probabilité de propagation, estimée à l’aide d’un réseau de neurones (FFNN), en fonction de la géométrie du site et des caractéristiques du signal. Cette approche permet de mieux capturer les phénomènes de réflexion, diffraction et occultation, tout en réduisant significativement le temps de calcul par rapport aux méthodes traditionnelles comme le *ray-tracing*.

Dans (Bufort et al. 2023) l’environnement de propagation est également représenté sous forme de graphe, cette fois construit à partir d’images géographiques transformées en structures de nœuds et d’arêtes. Les nœuds représentent des emplacements physiques (pixels), et les arêtes modélisent deux

types de relations : la proximité spatiale entre voisins immédiats, et des connexions orientées vers l’antenne simulant des trajectoires de type ray tracing. Ce graphe est ensuite exploité par un réseau de neurones en graphes, capable de faire circuler l’information à travers les nœuds selon ces deux types de liens. Grâce à cette architecture, le modèle apprend à prédire les niveaux de signal à partir de simples points de mesure et à générer des cartes de couverture complètes, tout en tenant compte des effets de l’environnement. L’approche se montre à la fois rapide et précise, surpassant les modèles physiques traditionnels dans des scénarios réels.

2.3.2 Principales architectures

Les réseaux de neurones en graphes (Scarselli et al. 2009) permettent de traiter efficacement les données structurées en graphes en combinant des mécanismes d’agrégation locale, de partage d’information entre nœuds, et de mise à jour adaptative des représentations. Plus formellement, étant donné un graphe $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ où \mathcal{V} est l’ensemble des nœuds et \mathcal{E} l’ensemble des arêtes, chaque nœud $v \in \mathcal{V}$ est associé à une représentation $h_v^{(k)}$ à la couche k . Le principe général d’un GNN repose sur le schéma de propagation suivant :

$$m_v^{(k)} = \text{AGG}^{(k)} \left(\left\{ M^{(k)} \left(h_v^{(k-1)}, h_u^{(k-1)}, e_{vu} \right) : u \in \mathcal{N}(v) \right\} \right)$$

$$h_v^{(k)} = U^{(k)} \left(h_v^{(k-1)}, m_v^{(k)} \right)$$

Où $\mathcal{N}(v)$ désigne l’ensemble des voisins du nœud v , $M^{(k)}(\cdot)$ est une fonction de message qui définit l’information transmise par un voisin u à v en prenant en compte la caractéristique de l’arête e_{vu} , $\text{AGG}^{(k)}(\cdot)$ est une fonction d’agrégation (somme, moyenne, max ou plus complexe), $U^{(k)}(\cdot)$ est une fonction de mise à jour, qui combine l’état précédent $h_v^{(k-1)}$ avec les messages agrégés $m_v^{(k)}$. Ainsi, les représentations des nœuds évoluent couche après couche en intégrant progressivement l’information issue de leur voisinage.

Plusieurs familles de modèles ont été développées et appliquées à la modélisation de la propagation radio, chacune répondant à des besoins spécifiques (S. He et al. 2021). Parmi la grande variété d’architectures de réseaux de neurones en graphes proposées dans la littérature, l’attention est portée ici sur celles ayant une pertinence directe pour les développements ultérieurs. Les GNN se répartissent généralement en quatre catégories principales, les GNN récurrents (RecGNN), les GNN convolutionnels (ConvGNNs), les autoencodeurs de graphes (GAEs) et les GNN spatio-temporels (STGNNs) (Wu et al. 2021) :

- RecGNNs : Modèles récurrents où les nœuds échangent des messages jusqu’à stabilité.

- ConvGNNs : Utilisent des convolutions pour agréger les infos des voisins et extraire des représentations hiérarchiques.
- GAEs : Encodent les graphes sans supervision pour des tâches de reconstruction ou génération.
- STGNNs : Capturent à la fois les dépendances spatiales et temporelles dans les graphes dynamiques.

Dans le reste de cette section, l'analyse porte plus précisément sur trois modèles issus de la famille des ConvGNNs : GCN, GraphSAGE et GAT. Ces modèles illustrent deux grandes approches de la convolution sur les graphes (Zhou et al. 2020) :

- GCN repose sur une convolution spectrale, formulée à partir des propriétés du Laplacien du graphe, défini comme la différence entre la matrice des degrés et la matrice d'adjacence, et qui encode la connectivité entre les nœuds.
- GraphSAGE et GAT, quant à eux, relèvent de la convolution spatiale, en se basant sur l'agrégation directe des informations issues des voisins à partir de la topologie du graphe.

Cette sélection permet d'explorer les principales stratégies d'apprentissage sur graphes en lien avec l'architecture neuronale proposée en Section 3.2.

2.3.2.1 Graph Convolutional Networks (GCN)

Les *Graph Convolutional Networks* (GCN) (Kipf et Welling 2017) ont été introduits comme une généralisation des convolutions classiques aux structures non euclidiennes. Leur formulation repose sur une approche spectrale, dans laquelle la convolution est définie à partir du spectre du Laplacien du graphe. Plus précisément, les filtres sont approximés par une expansion polynomiale des valeurs propres du Laplacien. Le modèle de (Kipf et Welling 2017) propose une simplification de cette formulation spectrale, aboutissant à la règle de mise à jour suivante :

$$H^{(k+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(k)} W^{(k)} \right)$$

Où $H^{(k)} \in \mathbb{R}^{|\mathcal{V}| \times d_k}$ regroupe les représentations des nœuds à la couche k , $W^{(k)}$ est la matrice de poids apprise à la couche k , $\tilde{A} = A + I$ est la matrice d'adjacence avec ajout d'auto-boucles, \tilde{D} est la matrice diagonale des degrés associée à \tilde{A} , $\sigma(\cdot)$ est une fonction d'activation non linéaire.

Dans le contexte de la propagation radio, les GCN peuvent être utilisés pour prédire la puissance reçue à un nœud donné à partir de l'information portée par les nœuds voisins. (Bufort et al. 2023) montrent par exemple qu'en exploitant les propriétés d'invariance topologique et la capacité des GCN à capturer des relations implicites entre variables, il est possible de représenter un environnement radio sous forme de graphe, où les nœuds correspondent à des emplacements et les arêtes traduisent

des relations spatiales ou issues de simulation (ici le *ray-tracing*). Leur approche, entraînée sur des mesures réelles, permet de générer des cartes de couverture radio avec une précision compétitive par rapport aux modèles physiques et tabulaires. De même (Zhang et al. 2024) modélisent le champ électromagnétique sous forme de graphes, où chaque nœud représente un point de l'espace associé à un vecteur de caractéristiques, tandis que les arêtes pondérées reflètent la corrélation spatiale, en particulier en conditions de visibilité directe. En formulant ainsi la tâche de cartographie radio comme un problème de régression semi-supervisée sur graphe, leur GCN exploite les mesures disponibles pour estimer les propriétés électromagnétiques aux emplacements non échantillonnés

2.3.2.2 Graph Attention Networks (GAT)

Les GAT (Veličković et al. 2018) introduisent un mécanisme d'attention permettant de pondérer différemment l'influence de chaque voisin dans l'agrégation des messages. Contrairement aux GCN, où tous les voisins ont une contribution uniformisée (après normalisation par les degrés), les GAT attribuent un poids appris à chaque arête, de manière à se concentrer sur les relations les plus pertinentes dans la structure du graphe. Le mécanisme de propagation peut être formalisé ainsi :

$$h_v^{k+1} = \sigma \left(\sum_{u \in \mathcal{N}(v)} \alpha_{vu}^{(k)} W^{(k)} h_u^{(k)} \right)$$

Avec les coefficients d'attention donnés par :

$$\alpha_{vu}^{(k)} = \frac{\exp \left(\text{LeakyReLU} \left(a^\top \left[W^{(k)} h_v^{(k)} \parallel W^{(k)} h_u^{(k)} \right] \right) \right)}{\sum_{j \in \mathcal{N}(v)} \exp \left(\text{LeakyReLU} \left(a^\top \left[W^{(k)} h_v^{(k)} \parallel W^{(k)} h_j^{(k)} \right] \right) \right)}$$

Où a est un vecteur de paramètre apprenant à scorer les interactions, \parallel l'opérateur de concaténation et $\alpha_{vu}^{(k)}$ est le poids d'attention attribué au voisin u et $\text{LeakyReLU} = \begin{cases} x & \text{si } x > 0 \\ \beta x & \text{si } x \leq 0 \end{cases}$ avec β un coefficient.

Dans le domaine des télécommunications sans fil, (Tekbıyık et al. 2021) utilisent les GAT pour l'estimation des canaux de communication dans des systèmes sans fil, notamment ceux envisagés pour la 6G. Dans cette étude, l'approche par GAT permet une meilleure estimation que celle donnée par la méthode des moindres carrés. Les auteurs soulignent la robustesse de cette architecture. Plus récemment, (Li et al. 2024) ont proposé RadioGAT, un cadre hybride combinant une modélisation basée sur la physique de la propagation radio et une approche d'apprentissage par GAT, spécifiquement conçu pour la reconstruction de cartes de propagation multi-bandes. Contrairement aux méthodes traditionnelles, souvent limitées par la disponibilité de données réelles ou la généralisation à partir de jeux de données simulés, RadioGAT repose sur une représentation par graphes dirigée par les données. Des résultats expérimentaux, obtenus à partir de simulations par ray tracing sur des

environnements urbains issus de cartes numériques, mettent en évidence la robustesse du modèle. Dans (Bufort et al. 2023) un GAT a été expérimenté comme GNN lors de la phase de conception du modèle, il permet d’obtenir des performances similaires à l’architecture finalement retenue dans le contexte de la génération de carte de propagation à partir des données représentées sous forme de graphes.

Bien que les GAT aient démontré leur efficacité aussi bien en transductif qu’en inductif, notamment en surpassant les approches de type GraphSAGE sur certains jeux de données (Veličković et al. 2018), leur dépendance à l’agrégation de l’ensemble des voisins peut constituer une limite pratique pour des graphes larges ou fortement connectés. En effet, l’absence de mécanisme d’échantillonnage induit un coût computationnel élevé. À l’inverse, l’approche inductive proposée par GraphSAGE (Hamilton et al. 2017) repose sur une stratégie d’échantillonnage contrôlé des voisins, garantissant une meilleure scalabilité pour l’apprentissage et l’inférence.

2.3.2.3 GraphSAGE (Graph Sample and aggreGatE)

L’architecture *Graph Sample and aggreGatE* (*GraphSAGE*) (Hamilton et al. 2017) repose sur l’apprentissage inductif et est particulièrement adaptée aux situations où l’entraînement doit permettre de généraliser à de nouveaux nœuds non observés. Elle applique un échantillonnage aléatoire d’un sous-ensemble de voisins à chaque couche, ce qui limite les coûts computationnels tout en préservant l’information locale essentielle. Les représentations sont ensuite mises à jour de manière itérative en combinant l’information du nœud courant avec celle de ses voisins, via une fonction d’agrégation. Trois types principaux d’agrégateurs sont proposés. L’agrégateur moyen effectue une moyenne élément par élément des représentations des voisins. Cette méthode est proche de la propagation utilisée dans GCN, mais adaptée au cadre inductif. L’agrégateur LSTM, plus expressif, traite les représentations des voisins via une cellule récurrente. Enfin, l’agrégateur *pooling* applique un réseau de neurones, suivi d’une opération de sous-échantillonnage (par exemple *max-pooling*), permettant d’extraire des caractéristiques non linéaires. Un schéma illustrant le processus de sampling et d’agrégation sur plusieurs couches est présenté en Figure 2.3.

Ainsi le mécanisme de propagation dans GraphSAGE repose sur deux étapes : (1) un échantillonnage aléatoire des voisins d’un nœud, et (2) une agrégation de leurs représentations. À chaque couche k , l’état du nœud v est mis à jour selon :

$$h_v^k = \sigma \left(W^{(k)} \cdot \left[h_v^{(k-1)} \parallel \text{AGG}^{(k)}(\{h_u^{k-1} : u \in \mathcal{N}(v)\}) \right] \right)$$

Avec $\text{AGG}^{(k)}(\cdot)$ la fonction d’agrégation (LSTM, moyen, *pooling*).

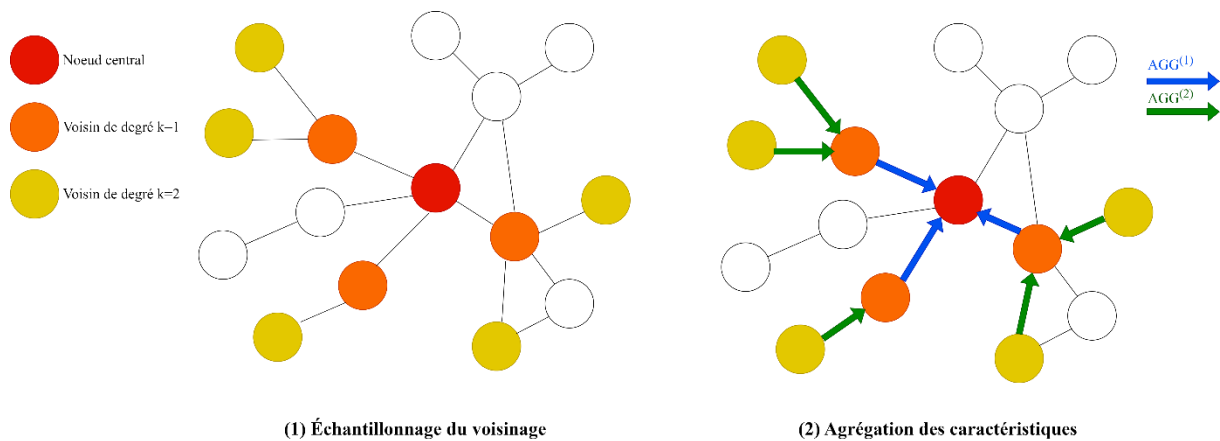


Figure 2.3 : Schématisation de la méthode d'échantillonnage et d'agrégation de GraphSAGE.

L'approche inductive portée par GraphSAGE trouve aujourd'hui des applications concrètes dans divers domaines liés aux réseaux sans fil, où les conditions dynamiques et l'incomplétude des données nécessitent une capacité de généralisation robuste. Par exemple, dans le contexte des systèmes « entrées multiples, sorties multiples » (MIMO), (Ranasinghe et al. 2021) exploitent GraphSAGE pour sélectionner les points d'accès de manière plus efficace que les méthodes traditionnelles. Leur méthode inductive permet d'éviter la mesure explicite de la puissance reçue de tous les points d'accès voisins, rendant l'algorithme plus scalable et pratique dans des environnements à grande échelle.

(Zhang et al. 2021) adoptent une autre stratégie dans des réseaux ultra-denses hétérogènes (HUDN) : en représentant les canaux radio comme des graphes hétérogènes, ils formulent une fonction de représentation optimale à apprendre via un GNN semi-supervisé issu de l'architecture GraphSAGE. Ils démontrent que l'apprentissage inductif permet une modélisation efficace des relations implicites entre utilisateurs et stations de base. Dans le cadre des systèmes de positionnement en intérieur, (Luo et Meratnia 2022) montrent que la géométrie implicite contenue dans les signaux RSSI peut être mieux exploitée via une modélisation en graphe, en particulier à travers une version de GraphSAGE utilisant un agrégateur par *pooling*. Cette approche surpasse les méthodes classiques (SVM, k-NN, CNN) en tenant compte de la structure topologique des signaux. Dans les réseaux maillés sans fil, (Lu et al. 2023) combinent GraphSAGE avec une recherche en largeur améliorée pour concevoir un algorithme de routage multi-trajets fiable. Ici encore, GraphSAGE sert à apprendre les performances des différents chemins disponibles via des agrégations inductives sur le graphe topologique du réseau, ce qui permet une adaptation efficace à des environnements dynamiques.

Ces travaux confirment l'intérêt de GraphSAGE, dans la modélisation des phénomènes de propagation radio, de connectivité et de traitement de signaux en milieux dynamiques et partiellement observés. La capacité de ces modèles à inférer à partir de sous-graphes locaux rend leur déploiement pertinent dans des systèmes réels où les conditions d'observation sont fragmentées, variables, voire bruitées.

2.3.2.4 Jumping Knowledge Networks

Contrairement à l'approche classique qui utilise la représentation issue de la dernière couche d'un GNN, la structure *Jumping Knowledge* (JK) (Xu et al. 2018) permet de fusionner les représentations produites à différentes couches d'un GNN, c'est-à-dire à différentes profondeurs d'agrégation. La Figure 2.4 schématise ce processus.

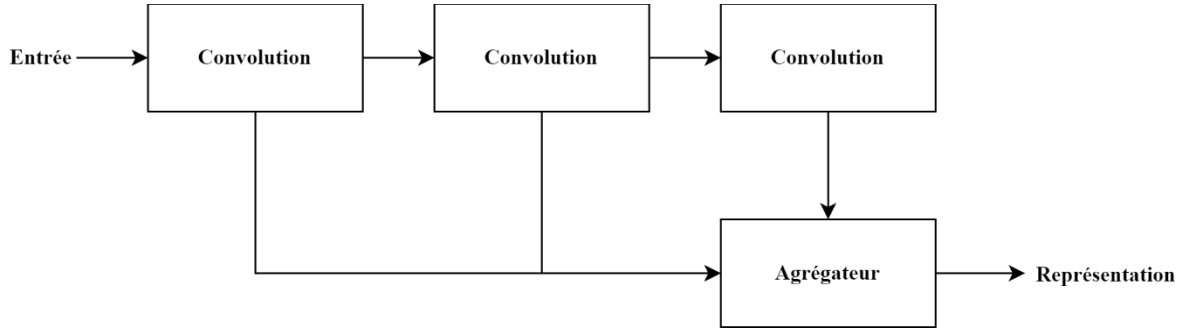


Figure 2.4 : Réseau JK à 3 couches. Chaque bloc de convolution représente une méthode arbitraire d'agrégation des voisins sur un graphe.

L'agrégateur peut être défini de trois façons différentes :

- Par concaténations des sorties de chaque couche du GNN

$$h_v^{\text{JK}} = h_v^{(1)} \parallel \dots \parallel h_v^{(L)}$$

- Par *max-pooling*

$$h_v^{\text{JK}} = \max(h_v^{(1)}, \dots, h_v^{(L)})$$

- Par somme pondérée, où les coefficients d'attention $\alpha_v^{(l)}$ sont obtenus par un réseau LSTM bidirectionnel

$$h_v^{\text{JK}} = \sum_{l=1}^L \alpha_v^{(l)} h_v^{(l)}$$

Où $h_v^{(l)}$ est l'embedding du nœud v à la couche l

Dans le contexte de l'allocation de ressource dans les réseaux sans fil, le modèle basé sur les GAT proposé par (Sun et al. 2025) utilise le mécanisme de JK afin d'améliorer la représentation au sein de leur GNN. Plus précisément ils utilisent un agrégateur par somme pondérée à l'aide d'un réseau LSTM. Ainsi l'algorithme proposé surpasse largement les méthodes de référence, avec des performances robustes et une forte capacité de généralisation.

2.3.2.5 Set2Set

Set2Set (Vinyals et al. 2016), proposée initialement pour les graphes chimiques, repose sur un mécanisme basé sur un mécanisme similaire à celui d'un LSTM attentionnel, permettant de générer une représentation globale du graphe à partir d'un ensemble de nœuds. Dans le cadre des GNN, l'avantage de cette méthode est d'apprendre à agréger des vecteurs dans un ordre invariant afin d'obtenir une représentation globale d'un graphe et d'améliorer les performances sur des tâches comme la classification ou la régression au niveau du graphe (Gilmer et al. 2017).

2.3.2.6 Agrégation par attention

L'agrégation par attention, notamment avec l'approche basée sur un mécanisme de *soft attention* proposée dans (Li et al. 2019), repose sur l'apprentissage de pondérations pour chaque vecteur de nœud afin de déterminer leur importance relative pour la tâche cible. Contrairement à une simple moyenne ou somme, cette méthode combine les vecteurs de nœuds de manière adaptative. Mathématiquement, la représentation agrégée d'un ensemble de nœuds h_v est donnée par :

$$r = \sum_{v \in \mathcal{V}} \text{softmax}(f(h_v)) \cdot g(h_v)$$

Où f et g sont des réseaux de neurones appliqués aux caractéristiques des nœuds. Le réseau f calcule les scores d'attention, tandis que h transforme les vecteurs de nœuds avant leur combinaison pondérée.

Dans le cadre des GNN, cette approche présente plusieurs avantages : elle permet d'apprendre quelles parties du graphe sont les plus pertinentes pour la tâche cible, tout en maintenant l'invariance à l'ordre des nœuds. Elle est particulièrement utile pour les tâches de classification ou régression au niveau du graphe, où certaines structures locales ou certains nœuds ont un impact plus fort sur la prédiction globale.

2.3.2.7 Agrégation multiple

Les architectures à multi-agrégation (Corso et al. 2020 ; Taylor et al. 2022) utilisent plusieurs fonctions d'agrégation de manière combinée, afin de capturer différents types d'interactions dans le graphe. La combinaison peut être réalisée de nombreuses manières : par somme, concaténation, moyenne, pondération par des coefficients d'attention parmi tant d'autres. Les nombreuses agrégations utilisées dans la littérature peuvent être catégorisées en deux groupes, celle basées sur des opérations simples ou statistiques, celles basées sur des méthodes apprenables :

- Les agrégations par opérations simples ou statistiques classiques : somme, moyenne, maximum, minimum, multiplication, variance, écart-type, médiane, quantile (Xu et al. 2019).
- Les agrégations par méthodes apprenables : agrégation par LSTM (Hamilton et al. 2017), agrégation par GRU (Buterez et al. 2022), agrégation par MLP (Buterez et al. 2022), Set2Set

(Vinyals et al. 2016), agrégation par mécanisme d'attention (Li et al. 2019), agrégation par *transformer* (Baek et al. 2021 ; Buterez et al. 2022 ; Li et al. 2023) entre autres.

2.4 Conclusion

La modélisation de la propagation des ondes a connu une évolution méthodologique marquée, passant des approches analytiques classiques aux méthodes d'apprentissage automatique, puis aux réseaux de neurones profonds et aux architectures basées sur les graphes. Cette progression a été motivée par la nécessité de capturer des phénomènes complexes, non linéaires et fortement dépendants de l'environnement, en particulier dans les zones hétérogènes.

Les approches d'apprentissage automatique traditionnelles, bien qu'efficaces dans certaines conditions, restent limitées par leur dépendance à la conception manuelle des variables, leur structure rigide et leur faible capacité à modéliser des interactions topologiques entre entités. Ces limitations sont accentuées dans les environnements où les phénomènes de réflexion, diffraction, obstruction requièrent une modélisation plus expressive.

L'introduction des réseaux neuronaux profonds a permis de dépasser certaines de ces contraintes, notamment par l'automatisation de l'extraction de caractéristiques, la prise en compte de données complexes, et l'amélioration des performances de généralisation. Toutefois, ces architectures demeurent centrées sur des données euclidiennes ou séquentielles, ce qui limite leur applicabilité lorsque les relations entre entités ne sont ni régulières, ni uniformes.

Les réseaux de neurones en graphes apportent une réponse formelle à cette problématique en permettant une représentation explicite des relations physiques, spatiales ou fonctionnelles entre les entités du système. Leur capacité à traiter des graphes, à intégrer des mécanismes d'agrégation et à généraliser à des structures nouvelles, en fait des outils particulièrement adaptés à la modélisation de la propagation VHF dans des environnements complexes et dynamiques.

Du fait de leur caractère général, les architectures issues des GNN offrent une grande flexibilité et une capacité d'adaptation à différents problèmes et domaines d'application. Leur efficacité repose sur la construction du graphe, le choix des mécanismes d'agrégation et leur ajustement aux spécificités du contexte étudié. Ces choix sont étroitement liés aux défis rencontrés dans les approches dirigées par les données, notamment en matière de qualité des données, de structuration spatiale et d'hétérogénéité des sources.

Le chapitre suivant est consacré à la conception, à la formalisation et à l'implémentation d'une architecture neuronale et d'une méthodologie dédiée à la génération de cartes de couverture AIS à partir de données massives, hétérogènes et partiellement observées. Y seront détaillés les choix

méthodologiques, les étapes de prétraitement, la définition du graphe de propagation ainsi que les justifications de l'architecture retenue.

Chapitre 3 : Méthodologie de modélisation des zones de réception de l’AIS à l’aide de réseaux de neurones en graphes et de données environnementales

Après avoir exposé dans le chapitre précédent les fondements des approches dirigées par les données et les apports récents des réseaux de neurones en graphes pour la modélisation de la propagation radio, ce chapitre présente la méthodologie développée pour la prédiction des zones de réception du signal AIS en environnement maritime (Renaud et al. 2025). L’objectif est d’établir une architecture neuronale exploitant à la fois les propriétés structurelles des graphes et les variables environnementales influençant la propagation radio.

La méthodologie proposée repose sur une représentation de l’espace sous forme de graphes orientés, où chaque instance correspond à un scénario localisé de propagation. Ces graphes intègrent des caractéristiques physiques, environnementales et géographiques, permettant une modélisation des conditions réelles de réception. Le modèle d’apprentissage développé s’appuie sur une architecture GNN inductive, enrichie par des mécanismes d’agrégation (*Jumping Knowledge*, multi-agrégation) et optimisée pour des prédictions binaires de réception.

Ce chapitre est structuré en plusieurs sections : une présentation générale de la démarche adoptée, la description détaillée de l’architecture neuronale proposée, la construction et le traitement du jeu de données utilisé, ainsi que la modélisation en graphes et l’ingénierie des caractéristiques associées.

3.1 Vue d’ensemble de la démarche

La modélisation de la réception AIS en tant que problème d’apprentissage supervisé nécessite une structuration et un étiquetage des données, ainsi qu’une représentation des phénomènes physiques sous-jacents. La démarche adoptée repose sur une approche modulaire et hiérarchique : elle combine des transformations de données à grande échelle, une formalisation des interactions spatiales sous forme de graphes, et la mise en œuvre d’un modèle neuronal de classification.

Cette section présente un aperçu global de la méthodologie, depuis la structuration des données jusqu’à l’inférence, en détaillant les différentes étapes successives : collecte, prétraitement, étiquetage, construction des graphes, extraction des caractéristiques et architecture du modèle. Ces étapes sont articulées dans une chaîne de traitement compatible avec des flux de données hétérogènes et massifs.

3.1.1 Démarche proposée

La démarche méthodologique retenue s'appuie sur trois éléments fondateurs :

- Des postulats selon lesquels la réception AIS présente une hétérogénéité spatio-temporelle induite par l'environnement et que les modèles purement physiques peinent à intégrer des sources multimodales à grande échelle.
- Des hypothèses de travail postulant qu'un modèle appris à partir de données empiriques, enrichi par une intégration multimodale et une représentation en graphes des interactions spatiales, peut dépasser les approches classiques.
- Un objectif opérationnel, qui est de développer une méthodologie d'apprentissage profond capable de prédire dynamiquement des zones de réception en tenant compte des conditions environnementales.

Concrètement, le Postulat 1 (influence environnementale et couverture fluctuante) motive l'intégration conjointe de variables météorologiques, environnementales et topographiques au-delà des messages AIS bruts (Section 3.3.3). Le Postulat 2 (limites des modèles physiques pour la fusion de sources) justifie une stratégie dirigée par les données, articulée autour d'une architecture en plusieurs couches assurant la traçabilité et l'alignement spatio-temporel des données (Sections 3.3.1-3.3.4).

Les Hypothèses 1-3 se traduisent méthodologiquement comme suit :

- H1 (apprendre sans modèles physiques explicites) : L'apprentissage repose sur une formulation supervisée binaire (« reçu / non reçu »). Les variables physiques sont intégrées à travers les interactions apprises par le réseau. Enfin, l'architecture GNN adoptée est inductive, ce qui lui permet de généraliser à de nouveaux sous-graphes (Section 3.2).
- H2 (intégration multimodale) : Cette hypothèse concerne la fusion de différentes sources au niveau des nœuds. Elle s'appuie sur des mécanismes de standardisation et d'alignement multi-résolution afin de combiner efficacement les informations (Section 3.3).
- H3 (représentations en graphes) : Les graphes considérés sont orientés de l'émetteur (Tx) vers le récepteur (Rx) et contraints par un cône directionnel. Des mécanismes multi-échelle, tels que *Jumping Knowledge*, ainsi que des approches multi-agrégation (moyenne, maximum, somme, attention, Set2Set) sont mobilisés afin de capturer simultanément les dépendances locales et globales (Section 3.2 et 3.4).

Cette articulation répond aux trois enjeux centraux : capter la structure spatiale et la directionnalité de la propagation, exploiter des données hétérogènes et massives de manière reproductible, produire des prédictions dynamiques et robustes.

Dans ce cadre, le choix de l'architecture d'apprentissage constitue un élément central de la méthodologie. Puisque les hypothèses de travail postulent que la structure spatiale et directionnelle du phénomène de propagation doit être représentée de manière explicite, il est nécessaire d'adopter un formalisme capable de traduire ces interactions locales et globales. C'est précisément ce rôle qu'assument les réseaux de neurones en graphes, qui apparaissent comme une réponse naturelle aux limites des approches classiques.

La section suivante développe ainsi la justification du recours aux GNN, en s'appuyant sur les acquis de l'état de l'art et sur leur adéquation avec les caractéristiques spécifiques du problème de réception AIS.

3.1.1.1 Recours aux GNN

L'état de l'art (Section 2.3) montre que les réseaux de neurones en graphes constituent une approche particulièrement adaptée dès lors qu'il s'agit de modéliser des phénomènes où les interactions spatiales et structurelles jouent un rôle déterminant. Contrairement aux modèles analytiques ou aux approches reposant sur des grilles régulières ou des réseaux convolutifs, qui imposent des structures fixes et isotropes (Imai et al. 2019 ; Zhang et al. 2020 ; Thrane et al. 2020 ; Levie et al. 2021), les GNN offrent la possibilité de représenter explicitement les relations locales et directionnelles entre entités (Zhou et al. 2020 ; Wu et al. 2021 ; S. He et al. 2021). Cette flexibilité est particulièrement pertinente pour la propagation radio, où la direction d'émission et les effets asymétriques de l'environnement jouent un rôle central.

Dans ce contexte, certaines architectures comme les GAT ont été explorées (Veličković et al. 2018), mais leurs mécanismes d'attention deviennent coûteux et instables face à de grandes volumétries ou à des graphes denses. À l'inverse, GraphSAGE adopte une stratégie d'échantillonnage et d'agrégation locale qui réduit le coût computationnel tout en maintenant la robustesse. Sa conception inductive permet de généraliser à de nouvelles instances sans dépendre de la structure globale du graphe, ce qui est déterminant dans des contextes dynamiques, partiellement observés et bruités (Hamilton et al. 2017 ; Ranasinghe et al. 2021 ; Luo et Meratnia 2022 ; Lu et al. 2023).

En somme, le recours à une architecture inductive de type GraphSAGE constitue une réponse directe aux limites des approches traditionnelles et garantit une meilleure adéquation entre la nature directionnelle et dynamique des phénomènes étudiés et les contraintes computationnelles liées à la volumétrie et à la connectivité des graphes. L'architecture retenue est présentée en Section 3.2.

3.1.1.2 Gestion de la volumétrie et hétérogénéité des données

La volumétrie élevée des messages AIS, qui peut atteindre plusieurs millions de trames par jour (Yang et al. 2019), impose la mise en place d'une organisation rigoureuse et modulaire du flux de données. Une structuration hiérarchisée en plusieurs couches successives permet de contrôler la

qualité, de documenter les transformations et de limiter l'impact des erreurs de traitement (Nambiar et Mundra 2022). À l'inverse, une gestion monolithique des données concentre l'ensemble des opérations dans une chaîne de traitement unique : toute erreur de décodage, d'alignement ou de filtrage se propage alors à l'ensemble de la chaîne, rendant leur identification et leur correction particulièrement coûteuses. Cette approche multicouche assure ainsi traçabilité, reproductibilité et robustesse, conditions indispensables pour exploiter des données massives et dynamiques. La description détaillée de la chaîne de traitement et d'ingestion est présentée en Section 3.3.2.

3.1.1.3 Choix des sources de données

Les données AIS locales fournissent les observations de référence, indispensables pour établir une vérité terrain de la réception effective. Toutefois, elles doivent être complétées par d'autres sources telles qu'AISHub⁹, un fournisseur de données populaire dans les applications AIS (Tu et al. 2018), qui permettent d'identifier les cas de non-réception une fois comparées aux données locales. Dans le cadre de l'apprentissage supervisé, cette combinaison est essentielle : elle assure une distribution équilibrée entre exemples positifs et négatifs, condition indispensable à l'entraînement d'un classifieur fiable (Bishop 2006).

Les variables environnementales, issues notamment des réanalyses météorologiques ERA5 (Hersbach et al. 2020) et de la topographie SRTM (NASA JPL 2013), sont mobilisées pour modéliser les conditions physiques de propagation. Elles permettent de tenir compte des phénomènes qui affectent la propagation radio. Pour les données ERA5 elles incluent les paramètres nécessaires (Nupen 1959) à la caractérisations de nombreux phénomènes de propagation (présentés en Section 1.1.2). Par exemple, ces données permettent de caractériser la réfractivité atmosphérique (ITU-R 2019b) et peuvent ainsi être mobilisées pour modéliser les phénomènes de propagation par conduit (Winning Jr. et al. 2024).

Les données SRTM sont quant à elles mobilisées pour la prise en compte de la topographie dans les phénomènes de propagation. Dans la littérature elles sont en effet associées à des modèles de terrain : par exemple, (James et al. 2018) les utilisent comme paramètres d'entrée d'un modèle ITM afin de prédire la portée de réception de l'AIS, tandis que (Song et al. 2020) les exploitent pour masquer les régions terrestres et ainsi supprimer des positions AIS improbables.

Du point de vue de l'apprentissage supervisé, ces variables jouent le rôle de caractéristiques ou variables explicatives (*features*) : elles fournissent au modèle des indicateurs permettant de relier des conditions physiques mesurables aux étiquettes de réception ou non-réception. Un avantage déterminant de ces bases de données réside dans leur couverture quasi globale et homogène : ERA5 fournit des réanalyses continues et cohérentes depuis 1940, avec une résolution horaire et spatiale

⁹ <https://www.aishub.net/>

suffisante pour la modélisation à l'échelle régionale ou mondiale (Hersbach et al. 2020). En effet, de nombreuses études confirment par ailleurs la qualité de ces données : elles reproduisent correctement les précipitations observées (Lavers et al. 2022), la vapeur d'eau (Yonglin Zhang et al. 2019) et divers paramètres troposphériques validés par radiosondes et GNSS (Guo et al. 2021 ; Huang et al. 2023). ERA5 restitue également fidèlement les principaux indicateurs convectifs comparés aux radiosondages dans une étude menée en Europe centrale (Varga et Breuer 2022). De son côté, SRTM offre une couverture quasi planétaire à une résolution de l'ordre de 30 à 90 m.

La Section 3.3 détaille les sources de données mobilisées et les variables caractéristiques sélectionnées.

3.1.1.4 Approche de fusion et choix des représentations

Le problème de prédiction de la réception AIS est formalisé sous forme de graphes orientés, afin de représenter explicitement la directionnalité entre l'émetteur et le récepteur, et de contraindre la propagation à des cônes spatiaux pertinents, définis comme des régions géométriques centrées sur l'axe émetteur-récepteur et limitées par une ouverture angulaire. Ce filtrage directionnel permet de concentrer le graphe sur les zones les plus susceptibles de contribuer à la propagation, tout en réduisant la complexité et l'influence des régions marginales. Cette formalisation s'appuie sur le constat que la propagation radio est un phénomène intrinsèquement directif et dépendant de la géométrie du milieu (Liu et al. 2020 ; 2021 ; 2023 ; Suto et al. 2023 ; Li et al. 2024 ; Bufort et al. 2023). Ces travaux démontrent qu'un graphe orienté permet de capter efficacement les interactions directionnelles, multi-sauts, et les effets de diffuseurs adjacents, qui agissent conjointement sur la réception effective du signal.

Chaque nœud combine un vecteur de caractéristiques environnementales (conditions atmosphériques ERA5), topographiques (SRTM) et issues des mesures AIS. Cette fusion au niveau des nœuds permet d'intégrer directement l'ensemble des descripteurs dans une représentation unifiée (Snoek et al. 2005). Plusieurs travaux démontrent l'intérêt de cette stratégie comparé à une fusion tardive, améliorant les performances et la robustesse au bruit (Gadzicki et al. 2020 ; Barnum et al. 2020 ; Sotiroudis et al. 2021). Pour l'apprentissage supervisé, ce choix est déterminant : il expose dès les premières couches du modèle les interactions croisées entre modalités, offrant au classifieur une information plus discriminante que si chaque modalité était traitée séparément (Gadzicki et al. 2020).

À l'instar des CNN, les modèles sur graphes peuvent souffrir de sur-lissage (Zhou et al. 2020), c'est-à-dire d'une homogénéisation progressive des représentations de nœuds au fil des couches, qui réduit leur pouvoir discriminant. Ce défi méthodologique a trouvé plusieurs réponses dans la littérature. L'utilisation de mécanismes multi-échelle tels que *Jumping Knowledge* permet d'agrèger les représentations issues de différentes profondeurs, combinant des représentations à différents

niveaux d'abstraction (Xu et al. 2018 ; Sun et al. 2025). La représentation globale du graphe est ensuite obtenue par une multi-agrégation combinant plusieurs stratégies (moyenne, max, somme, attention, Set2Set). Chaque fonction d'agrégation capture un type particulier de dépendance : distribution globale des caractéristiques, identification des éléments dominants, résumés structurels ou mise en avant des interactions principales par attention. Cette approche est cohérente avec les résultats de (Corso et al. 2020), qui démontrent que l'utilisation de plusieurs agrégateurs est nécessaire pour capturer toute la richesse statistique du graphe et améliorer significativement l'expressivité et la robustesse des GNN. Dans la même lignée (Tailor et al. 2022) exploitent explicitement des combinaisons multi-agrégateurs et montre que ces architectures surpassent même des modèles réputés plus expressifs, tout en restant plus efficaces en mémoire et en calcul. Dans un contexte supervisé, cette diversité de résumés statistiques permet d'éviter le biais induit par une stratégie unique et de fournir au classifieur final des représentations plus riches et plus stables.

En résumé, cette approche de fusion hiérarchique, intégrant données multimodales au niveau des nœuds et multi-agrégation au niveau du graphe, améliore la robustesse du modèle et optimise la qualité des représentations apprises, ce qui renforce la capacité du modèle supervisé à généraliser. Les stratégies d'ingénierie des graphes et de fusions retenues sont décrites en Section 3.4. La représentation du graphe et les stratégies d'agrégation sont décrites lors de la conception de l'architecture GNN en Section 3.2.

3.1.2 Schémas de la démarche

La démarche proposée peut être résumée sous la forme d'une chaîne de traitement hiérarchisée, où chaque étape correspond à une opération clé depuis la définition du cadre d'apprentissage jusqu'à la construction des graphes.

La première étape consiste à définir le type d'apprentissage retenu et l'architecture neuronale associée, en l'occurrence une classification de graphes supervisée fondée sur GraphSAGE, enrichie de mécanismes multi-échelle et de stratégies de multi-agrégation. Vient ensuite le choix des descripteurs et l'identification des sources de données pertinentes, combinant mesures AIS locales, données de réception AISHub et variables environnementales (ERA5, SRTM). Les données ainsi sélectionnées sont ingérées et structurées à travers un processus de nettoyage, d'alignement spatio-temporel et d'organisation en couches successives, garantissant la traçabilité des transformations. Elles sont ensuite fusionnées au niveau des nœuds par une stratégie d'*early fusion*, qui permet de construire des vecteurs unifiés intégrant simultanément variables AIS, environnementales et topographiques. Enfin, l'ensemble est transformé en graphes orientés où les cellules de grille jouent le rôle de nœuds, reliés par des arêtes orientées $Tx \rightarrow Rx$ et filtrés par un cône directionnel. Chaque graphe est associé à une étiquette binaire indiquant la réception ou non du signal AIS, formalisant ainsi le problème sous la forme d'une tâche supervisée de classification de graphes.

Afin de relier la justification théorique de la démarche et sa mise en œuvre effective, deux représentations complémentaires sont proposées :

- La Figure 3.1 illustre les étapes de la démarche méthodologique, depuis la définition du cadre d'apprentissage et l'identification des sources de données jusqu'à la formalisation en graphes orientés. Ce schéma met en avant les grandes étapes conceptuelles et leur articulation.
- La Figure 4.1 détaille la mise en œuvre pratique du modèle (Section 4.1), en décomposant la chaîne de traitement : construction et prétraitement des graphes, entraînement et validation du modèle neuronal, puis génération des prédictions et des cartes de réception.

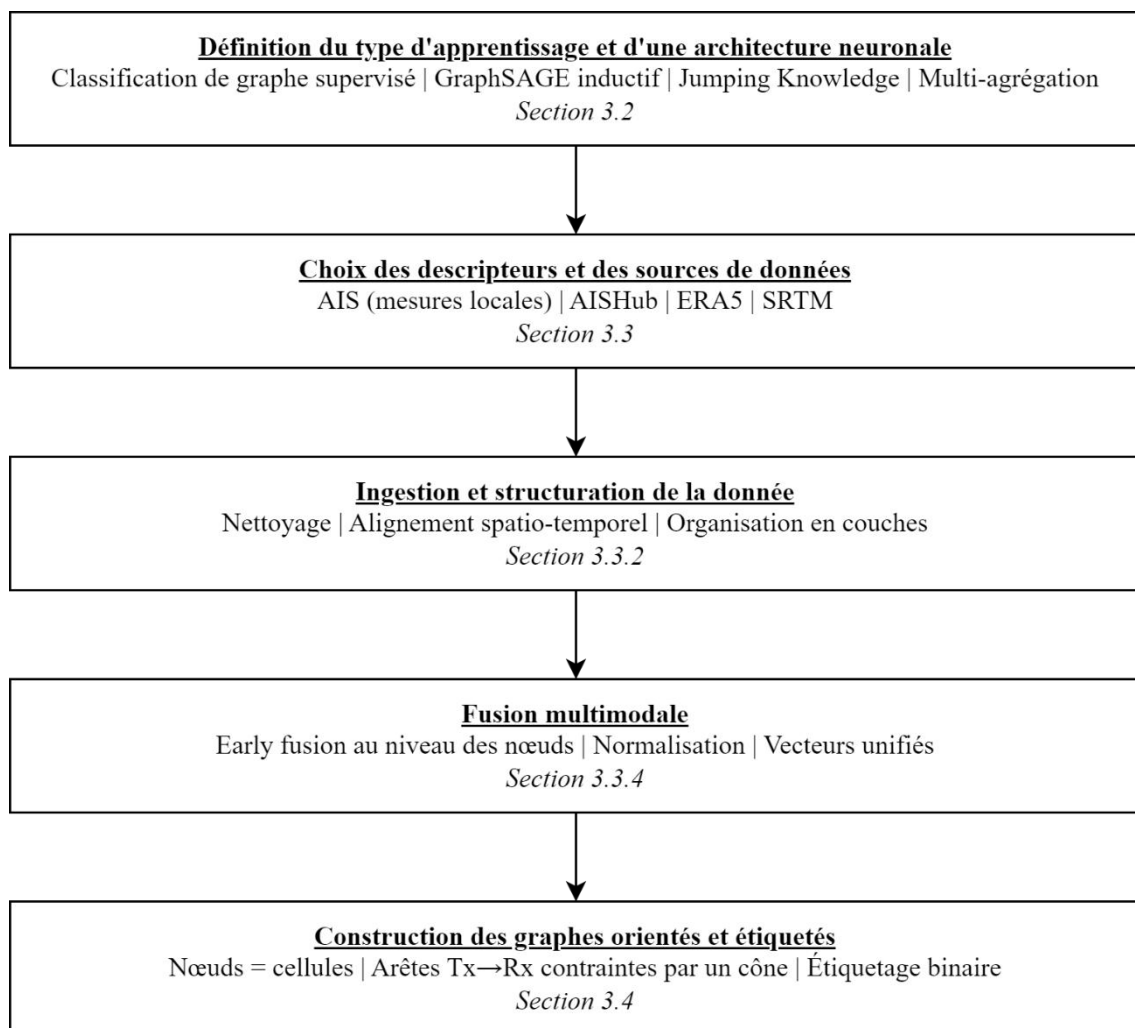


Figure 3.1 : Etapes conceptuelles de la démarche proposée.

Ces deux niveaux de représentation offrent une vision à la fois conceptuelle et opérationnelle de la démarche, assurant une cohérence entre les choix de conception et leur traduction en chaîne de traitement expérimentale.

3.2 Architecture neuronale proposée

L'architecture neuronale développée dans cette thèse repose sur une modélisation inductive de la propagation AIS via des réseaux de neurones en graphes. Ce choix méthodologique se justifie par la nature hautement structurée et localement directionnelle du phénomène étudié : la propagation du signal radio est à la fois contrainte par la topologie de l'environnement et influencée par des conditions météorologiques et géographiques variables.

L'approche adoptée est celle de la classification de graphes : chaque instance correspond à un sous-graphe orienté représentant un scénario local de propagation. Le modèle doit prédire si ce sous-graphe est associé à une réception effective du signal. L'architecture s'appuie sur GraphSAGE pour l'agrégation inductive, enrichie par des mécanismes *Jumping Knowledge* pour l'intégration multi-échelle des représentations, et une couche de *pooling* multi-stratégie pour capter simultanément les aspects locaux, globaux et attentionnels de la structure. Cette section décrit les choix architecturaux, les fonctions d'agrégation, les mécanismes d'apprentissage et les justifications sous-jacentes, en lien avec les propriétés physiques du problème.

3.2.1 Architecture proposée

Une approche par classification de graphes est adoptée, où chaque instance en entrée est un graphe orienté représentant un scénario localisé de propagation AIS. La tâche consiste à prédire une étiquette binaire indiquant si un sous-graphe donné correspond à une zone de réception positive du signal AIS. Le cadre général adopté pour ce problème de classification de graphes est présenté dans la Figure 3.2 (Wu et al. 2021).

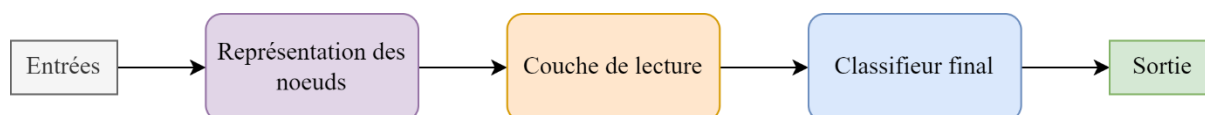


Figure 3.2 : Cadre général de classification de graphes.

L'architecture proposée est détaillée dans la Figure 3.3, elle repose sur GraphSAGE, une méthode d'agrégation de voisinage bien adaptée aux tâches d'apprentissage inductif (Hamilton et al. 2017 ; Zhou et al. 2020). GraphSAGE permet au modèle de généraliser à des topologies de graphes non vues en apprenant des fonctions qui agrègent et transforment l'information provenant des voisinages des nœuds. Contrairement aux méthodes transductives, GraphSAGE ne nécessite pas la structure complète du graphe lors de l'entraînement, ce qui correspond bien à ce cadre où les graphes sont générés dynamiquement pour différentes instances spatio-temporelles.

Pour améliorer l'expressivité, un mécanisme *Jumping Knowledge (JK)* reposant sur un schéma de sommation pondérée (Xu et al. 2018) est intégré. Plutôt que de s'appuyer uniquement sur la dernière couche du GNN, cette approche agrège les représentations de nœuds issues de toutes les couches

intermédiaires au moyen de poids d'attention obtenus grâce à un LSTM bidirectionnel (Bi-LSTM) (Hochreiter et Schmidhuber 1997) :

$$h_v^{\text{JK-LSTM}} = \sum_{t=1}^L \alpha_v^{(t)} h_v^{(t)}$$

Où $h_v^{(t)}$ désigne l'embedding du nœud v à la couche t , L le nombre total de couches, et les scores d'attention $\alpha_v^{(t)}$ sont appris de manière adaptative à chaque nœud par un Bi-LSTM. Cette conception permet au modèle d'ajuster dynamiquement son champ réceptif et d'intégrer des motifs spatiaux multi-échelles, ce qui est crucial compte tenu de la variabilité de la taille et de la topologie des graphes selon les instances. Sont ensuite appliqués des transformations non-linéaires grâce à la fonction d'activation $\text{ReLU}(x) = \max(0, x)$ très populaire en apprentissage profond (Agarap 2018).

Après obtention de la représentation des nœuds, une représentation au niveau du graphe est calculée (couche de lecture) via une couche à multi-agrégation. Plutôt que de s'appuyer sur une stratégie de *pooling* unique, les sorties de cinq fonctions d'agrégation distinctes sont concaténées :

- Agrégation par moyenne : pour capturer la distribution (Xu et al. 2019).
- Agrégation par maximum : pour identifier les éléments représentatifs (Xu et al. 2019).
- Agrégation par somme : pour apprendre les propriétés structurelles du graphe (Xu et al. 2019).
- Agrégation attentionnelle : pour pondérer la contribution des nœuds via un réseau de portes entraînable (Li et al. 2019).
- Agrégation Set2Set : pour modéliser les dépendances d'ordre supérieur via un mécanisme d'attention global récurrent (Vinyals et al. 2016).

Cette combinaison permet au modèle de préserver à la fois les caractéristiques locales et globales de la structure du graphe, augmentant la robustesse et la richesse de la représentation (Corso et al. 2020 ; Tailor et al. 2022).

Le vecteur de caractéristiques ainsi agrégé est ensuite transmis à un perceptron multicouche (MLP) à deux couches entièrement connectées (FC), qui joue le rôle de classifieur final. Le MLP applique une transformation non linéaire ReLU suivie d'une fonction d'activation sigmoïde $\sigma(x) = \frac{1}{1+e^{-x}}$ pour produire une probabilité $\hat{y} \in [0, 1]$, représentant la probabilité prédite de réception du signal AIS pour le sous-graphe en entrée.

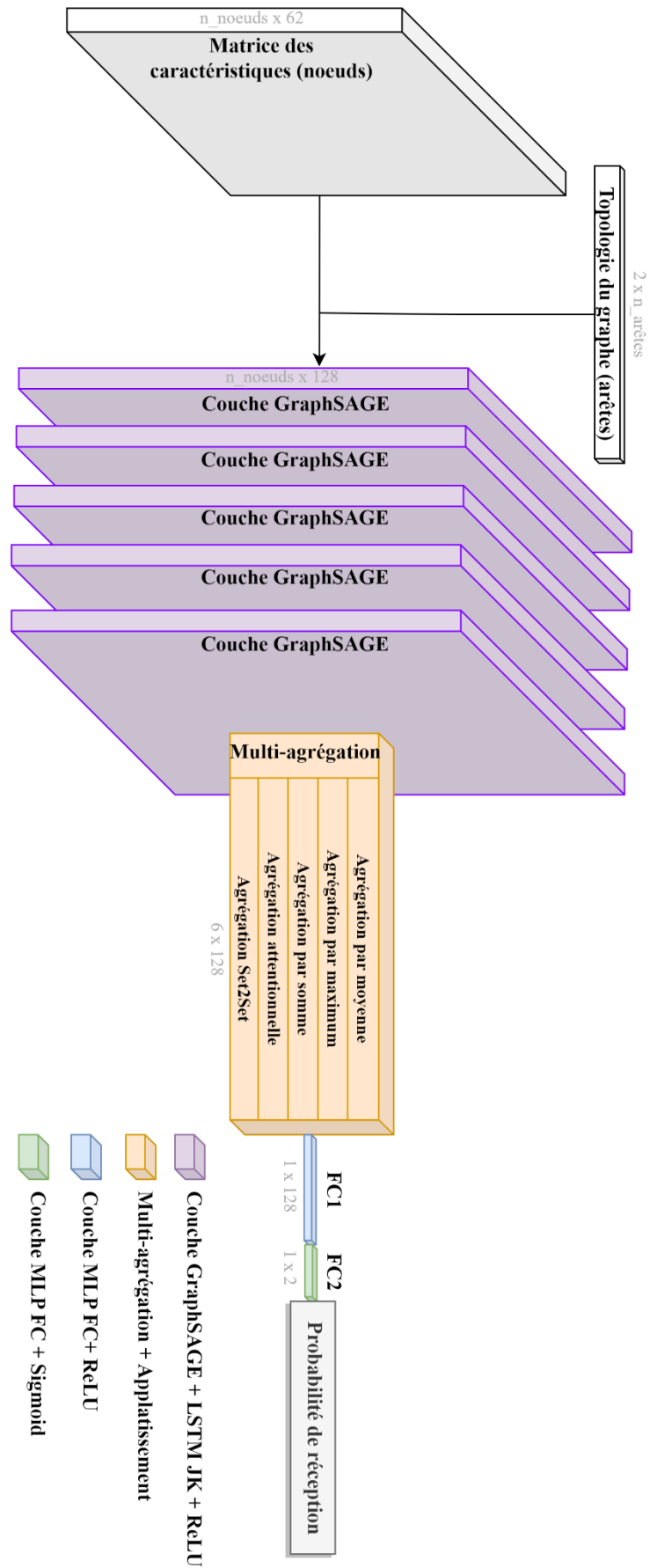


Figure 3.3 : Architecture détaillée du GNN proposé pour la prédiction de la réception AIS.

La conception de cette architecture GNN est non seulement guidée par les données, mais aussi inspirée par la nature physique de la propagation des ondes radio. Ensemble, ces composants forment une architecture neuronale qui non seulement apprend à partir des données mais reflète également des hypothèses physiques, améliorant ainsi l’alignement avec la connaissance du domaine :

- La structure du graphe modélise le domaine spatial à travers lequel les signaux AIS se propagent. Les nœuds correspondent à des unités géographiques discrétisées et les arêtes orientées représentent des chemins de propagation potentiels. L’orientation des arêtes Tx → Rx imite la direction physique d’émission du signal et tient compte de l’atténuation géométrique avec la distance et l’orientation.
- L’agrégation de voisinage dans GraphSAGE approxime la diffusion locale du front d’onde. En moyennant ou en résumant les caractéristiques des nœuds voisins, le GNN imite l’influence des facteurs environnementaux des régions adjacentes sur l’intensité du signal à un point donné (Bufort et al. 2023).
- L’extraction de sous-graphes en forme de cône (Section 3.4.2) est alignée avec le concept de diagrammes de rayonnement et de cônes de rayonnement en physique radio (Yun et Iskander 2015), limitant le domaine spatial aux zones les plus susceptibles d’affecter ou d’être affectées par la transmission. Ce biais géométrique améliore la pertinence des nœuds sélectionnés et réduit le bruit.
- L’agrégation attentionnelle permet au modèle de pondérer les zones spatiales les plus influentes (Veličković et al. 2018), ce qui revient à se focaliser sur les zones présentant des variables environnementales ou des conditions météorologiques plus favorables ou discriminantes. De manière similaire, le module Set2Set permet d’incorporer un contexte global, comparable à une propagation à longue distance modélisant des effets de plus hauts niveaux.

3.2.2 Variantes explorées

Plusieurs alternatives architecturales ont été envisagées mais n’ont pas été retenues dans le modèle final, en raison de limites théoriques et des résultats obtenus lors des expérimentations préliminaires. L’architecture GraphSAGE a été privilégiée par rapport aux GNN basés sur l’attention tels que GAT, en raison de sa scalabilité et de sa robustesse (Dwivedi et al. 2023). Dans le cadre du mécanisme *Jumping Knowledge*, différentes stratégies ont été expérimentées :

- JK-Moyen calcule une moyenne des représentations de nœuds issues des différentes couches :

$$h_v^{\text{JK-Moyen}} = \frac{1}{L} \sum_{t=1}^L h_v^{(t)}$$

Cette approche, bien que simple et efficace, tend à diluer les signaux provenant des couches profondes.

- JK-Concat empile les représentations de toutes les couches :

$$h_v^{\text{JK-Concat}} = h_v^{(1)} || \dots || h_v^{(L)}$$

Elle conserve la richesse des représentations multi-échelles, mais au prix d'une complexité accrue du modèle.

- JK-LSTM ajuste dynamiquement la combinaison des représentations en fonction du contexte structurel, ce qui permet de capturer les hiérarchies spatiales sans provoquer de sur-lissage excessif (Xu et al. 2018).

D'un point de vue empirique, JK-Moyen a montré une précision de validation instable, tandis que JK-Concat entraînait parfois un sur-paramétrage du modèle. JK-LSTM a offert un compromis satisfaisant entre expressivité et stabilité.

Enfin, l'utilisation d'une couche de sortie unique (moyenne) a été testée à la place d'une multi-agrégation. Bien que de tels designs soient moins coûteux en calcul, ils se sont révélés moins efficaces. La stratégie de multi-agrégation, en combinant des résumés statistiques et attentionnels complémentaires, a conduit à des performances de classification plus constantes. Finalement, ces choix architecturaux reflètent un compromis entre la complexité du modèle et sa capacité représentationnelle.

3.3 Données mobilisées et transformations

La performance des modèles d'apprentissage dirigés par les données repose directement sur la qualité, la diversité et la pertinence des jeux de données mobilisés. Dans ce contexte de modélisation de la réception AIS, il est essentiel d'intégrer simultanément des données issues de mesures de terrain (messages AIS bruts) et celles liés aux phénomènes sous-jacents, c'est à dire, des variables environnementales dynamiques et des descripteurs géographiques.

Cette section décrit l'ensemble des sources de données utilisées, les méthodes de prétraitement mises en œuvre, ainsi que l'architecture de traitement adoptée. L'organisation en couches permet d'assurer la traçabilité, la reproductibilité et la robustesse des chaînes de traitement. Les étapes de transformation, de filtrage et d'alignement temporel et spatial sont ensuite détaillées pour chaque source.

3.3.1 Sources de données

Afin de modéliser efficacement la propagation AIS, une variété de sources de données est mobilisée, ensemble, ces données constituent le jeu de données AIS intégré. Ces sources incluent à la fois des données de mesure brutes et des données contextuelles, chacune apportant une information unique et complémentaire pour la tâche de modélisation.

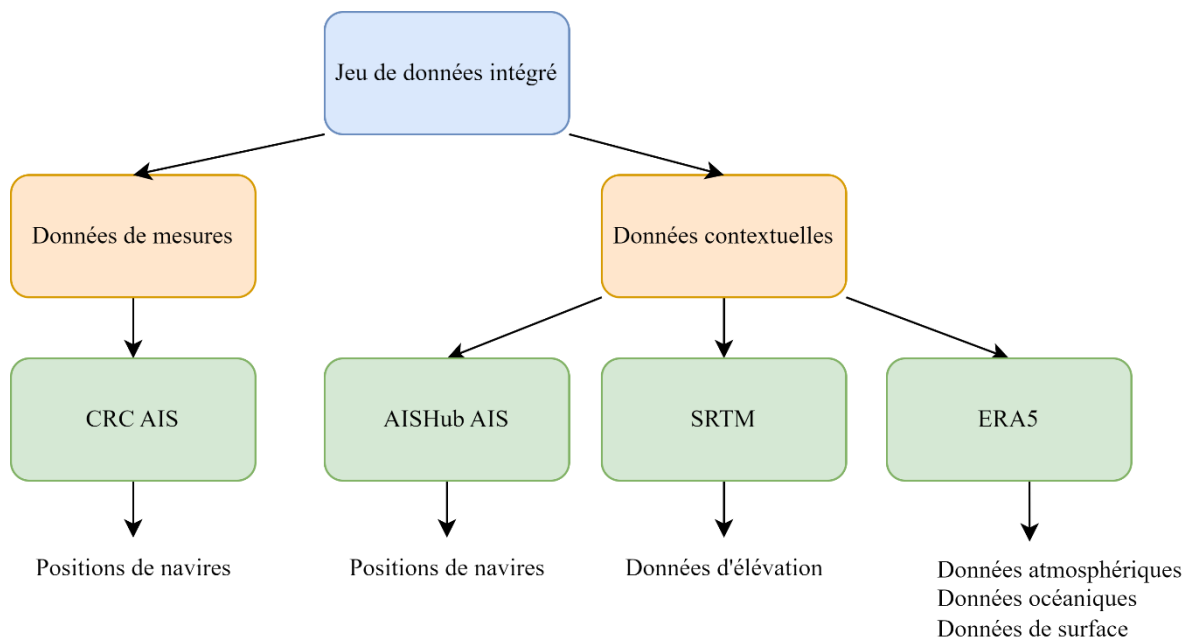


Figure 3.4 : Typologie des sources de données pour le jeu de données intégré de propagation AIS.

La Figure 3.4 présente un aperçu des données utilisées pour construire le jeu de données intégré de propagation AIS. La figure distingue les données de mesure, produites par le laboratoire CRC, des données contextuelles, obtenues à partir de sources externes, et détaille les origines et types de données extraites pour chacune. Les données de mesure comprennent des messages AIS tels que le MMSI, la position et l'horodatage (*timestamp*). Les données contextuelles incluent des sources telles que ERA5, AISHub et SRTM, fournissant des informations atmosphériques, des vagues océaniques, de la surface terrestre et de l'élévation, utilisées pour enrichir le processus de modélisation de la propagation.

3.3.2 Données issues de mesures

Depuis août 2018, le Centre de Recherche sur les Risques et les Crise (CRC) (Mines Paris–PSL – Campus Pierre Laffitte) (Figure 3.5) collecte des trames AIS au format NMEA transmises par des navires en mer Méditerranée. Ce jeu de données comprend jusqu'à un million de trames par jour. Pour cette étude, seuls les messages AIS de la classe A et de types 1, 2, 3, 18, 19 et 27 sont retenus, car ceux-ci correspondent à des rapports de position contenant des localisations géoréférencées des navires ainsi que des données dynamiques de navigation (ITU-R 2014).

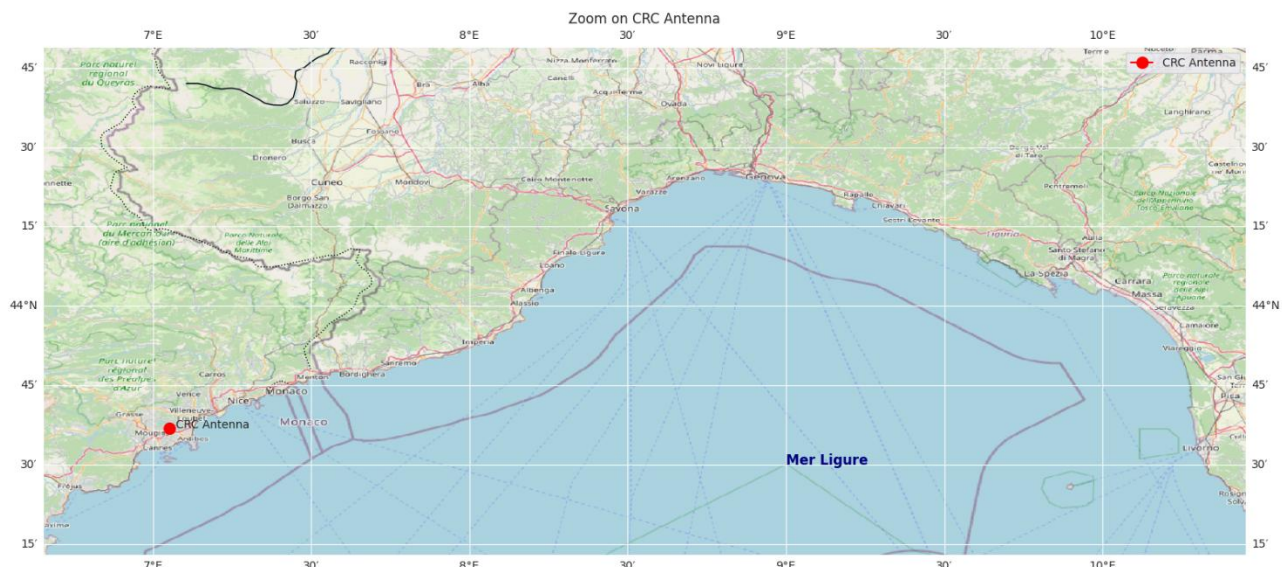


Figure 3.5 : Localisation géographique de la station de réception AIS du CRC Mines Paris - PSL à Sophia Antipolis sur le Campus Pierre Laffitte

L'acquisition de ces trames NMEA est assurée par un système composé d'une antenne *YAGI SIRIO WY 155-3N* installée à une altitude de 188 mètres au-dessus du niveau de la mer. L'antenne a un axe directionnel de 100 degrés et une ouverture de faisceau de 130 degrés. Elle est connectée, via un câble coaxial *Messi & Paoloni ULTRAFLEX 10*, à un récepteur AIS *Comar SLR350N*. La station de réception est géographiquement située à Sophia Antipolis, dans le sud-est de la France, à proximité de la côte méditerranéenne. Les spécifications de cette installation sont détaillées dans les Tableau 3.1, Tableau 3.2 et Tableau 3.3. Les Figure 3.7 et Figure 3.8 précisent la directivité de l'antenne ainsi que sa courbe de gain, valeurs extraites graphiquement et stockés dans un dictionnaire où chaque position angulaire est associée à une valeur de gain.

Une fois le système d'acquisition connecté au réseau local, les trames AIS sont horodatées et stockées. Les données brutes AIS collectées par la station CRC se composent ainsi de millions de messages par jour au format NMEA. Ce volume, ainsi que le besoin de contrôles de qualité et la multiplicité des transformations, motivent l'adoption d'une architecture de données robuste et modulaire.

Paramètre	Valeur
Type	Yagi à 3 éléments
Plage de fréquence ROS $\leq 1,5$	[155, 175] MHz
Largeur de faisceau horizontale à -3 dB	130°
Largeur de faisceau vertical à -3 dB	70°
Gain	Voir Figure 3.7 et Figure 3.8

Tableau 3.1 : Caractéristiques de l'antenne de la station CRC (documentation SIRIO).

Fréquence (MHz)	Atténuation (dB / 100 m)
144	7,40
200	8,90

Tableau 3.2 : Pertes de signal dans le câble de la station CRC (documentation Messi & Paoloni).

Paramètre	Valeur
Fréquences	161,975 MHz et 162,025 MHz
Sensibilité	-112 dBm

Tableau 3.3 : Caractéristiques du récepteur AIS de la station du CRC (documentation Comar Systems).

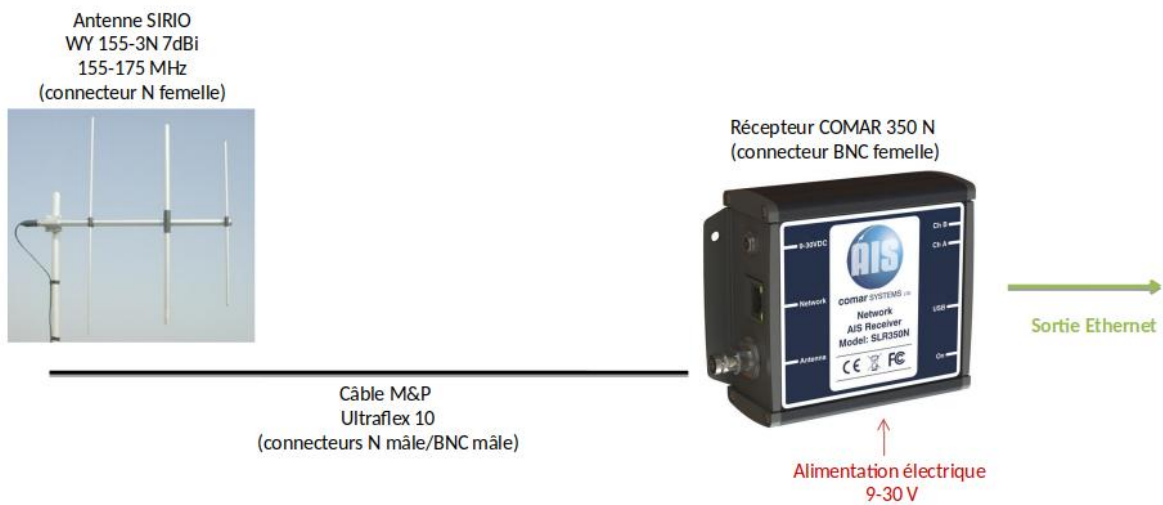


Figure 3.6 : Schéma du système d'acquisition des données AIS du CRC.

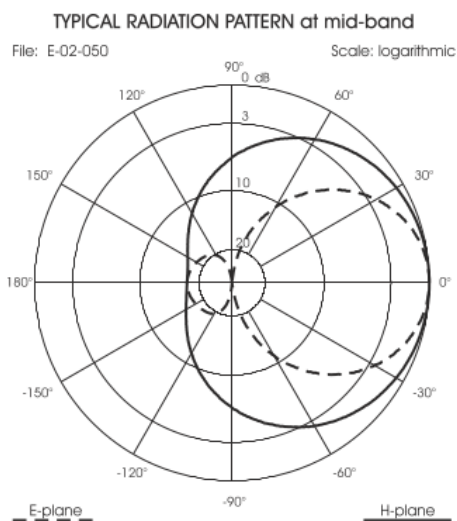


Figure 3.7 : Diagramme de directivité de l'antenne du CRC (reproduction autorisée par SIRIO).

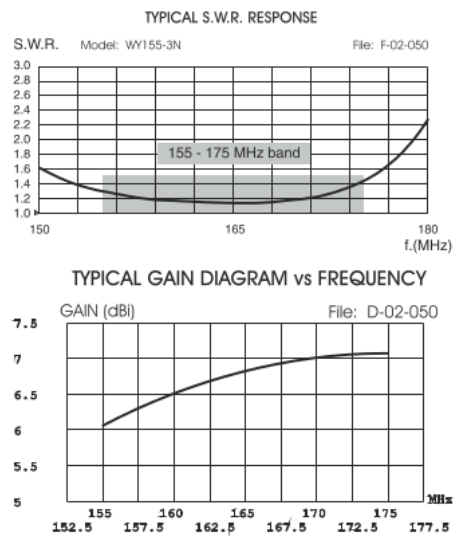


Figure 3.8 : Courbe de gain de l'antenne du CRC (reproduction autorisée par SIRIO).

Afin de garantir la fiabilité, la traçabilité et la réutilisabilité des données dans les différentes expériences, une architecture en plusieurs couches a été adoptée, un patron de conception de plus en plus utilisé dans les chaînes de traitement de données à grande échelle (Ravat et Zhao 2019 ; Nambiar et Mundra 2022). Le choix a été fait d'organiser cette architecture en trois couches, de manière analogue à l'architecture utilisée par certains services de stockage de données tel que Databricks¹⁰ (où cette architecture est nommée « architecture en médaillon »), chacune correspondant à un niveau de raffinement et de validation :

- Couche L1 : stocke les trames AIS brutes et non filtrées, telles que collectées par la station, sous forme de fichiers journaux, en conservant les informations originales et les horodatages.
- Couche L2 : contient des données structurées et interprétées, où les messages AIS sont décodés, sous-échantillonnés, filtrés par type et nettoyés (par exemple, suppression des trames incomplètes ou mal formées). Ces données sont stockées dans une base *TimescaleDB* optimisée pour les requêtes temporelles (« TimescaleDB : SQL made scalable for time-series data » 2017).
- Couche L3 : regroupe des jeux de données enrichis et validés, stockés sous forme de fichiers *CSV* mensuels, tels que des positions sous-échantillonnées et des données préfiltrées pour les tâches de modélisation.

Cette approche structurée permet la reproductibilité des chaînes de traitement de données, l'isolation des erreurs et un retraitement facilité, tout en maintenant une flexibilité pour d'éventuelles étapes d'enrichissement.

Pour gérer les transferts de données entre les couches, des chaînes de traitement ETL (*Extract, Transform, Load*) ont été mises en œuvre, comme décrit dans la Figure 3.9. Ces chaînes de traitement réalisent une série d'opérations, représentées de manière visuelle dans les diagrammes de flux de données (Coad et al. 1992) présentés en Figure 3.10.

¹⁰ <https://www.databricks.com/>

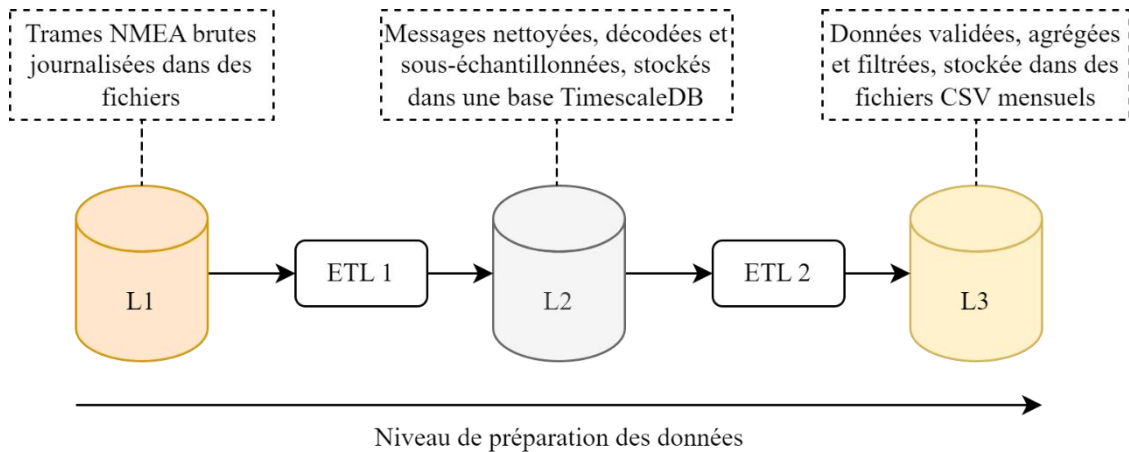


Figure 3.9 : Architecture en médaillon et chaîne de traitement ETL pour le traitement des messages AIS. Elle traite les trames AIS brutes selon trois niveaux de raffinement : L1 (brutes), L2 (nettoyées, décodées et sous-échantillonnées) et L3 (validées, agrégées et filtrées).

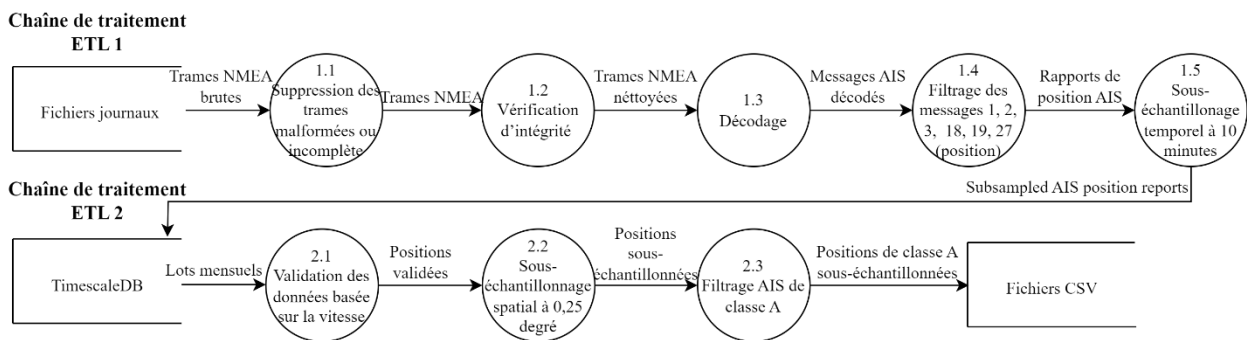


Figure 3.10 : Vue d'ensemble des chaînes de traitement ETL pour le traitement des données AIS. La première chaîne de traitement (haut) transforme les données brutes de la couche L1 vers la couche L2 à travers des étapes de nettoyage, décodage, filtrage et sous-échantillonnage. La seconde chaîne de traitement (bas) réalise les transformations avancées de la couche L2 vers la couche L3, incluant validation, réduction temporelle et filtrage, en préparation des tâches de modélisation.

3.3.2.1 Implémentation de la chaîne de traitement ETL 1

Les chaînes de traitement ont été implémentées en langages de programmation *Bash*, *C* et *Python*, afin de combiner l'efficacité du traitement bas niveau avec la flexibilité nécessaire selon la complexité des transformations et les contraintes de vitesse d'exécution, fortement dépendantes de la volumétrie des données. Les données intermédiaires sont systématiquement archivées sur une infrastructure de stockage dédiée, garantissant ainsi la traçabilité complète des transformations et facilitant la reprise éventuelle des traitements. L'ensemble des opérations est exécuté en mode multiprocesseur et par lots, de manière à optimiser l'utilisation des ressources CPU et à réduire les temps d'exécution.

La première chaîne de traitement a pour rôle de transformer les trames AIS brutes, stockées dans la couche de données brutes L1, en données structurées et nettoyées dans la couche de données nettoyées L2. Ce processus se déroule en plusieurs étapes successives.

3.3.2.1.1 Extraction des trames NMEA

L'interface réseau du dispositif AIS Comar assure le transfert des trames AIS par connexion UDP. Un serveur, implémenté en langage C, a été développé afin d'extraire du flux les trames complètes et correctement formatées. Ces trames sont ensuite journalisées à l'aide de l'utilitaire *logrotate*, de manière à générer pour chaque journée un fichier regroupant l'ensemble des trames AIS brutes collectées. Lors de ce processus, les trames sont converties au format NMEA 4.10, ce qui permet leur horodatage sous forme d'*epoch Unix*.

3.3.2.1.2 Vérification d'intégrité

Pour chaque trame reçue, une fonction de vérification d'intégrité est appliquée. Si cette vérification échoue, la trame est archivée et ne sera pas intégrée au jeu de données. Au niveau du format NMEA 0183 utilisé par l'AIS, chaque message se termine par **hh* où *hh* correspond à une somme de contrôle. Cette somme de contrôle est calculée à l'envoi en effectuant un XOR binaire de tous les codes ASCII des caractères situés entre le caractère initial et l'astérisque (exclus). La fonction de vérification d'intégrité implémente donc la formule :

$$\text{Checksum} = \bigoplus_{i=1}^n \text{ASCII}(c_i)$$

Où \bigoplus désigne l'opération XOR et c_i chaque caractère de la trame. Le résultat est affiché en hexadécimal sur deux chiffres. Si la valeur calculée ne correspond pas au checksum présent dans la trame, celle-ci est considérée comme corrompue et exclue du traitement.

3.3.2.1.3 Décodage des trames, filtrage et stockage

Le décodage des trames est réalisé en s'appuyant, d'une part, sur le protocole proposé par (ITU-R 2014) ainsi que par les éditeurs de *gpsd*¹¹ et, d'autre part, sur une version modifiée de la bibliothèque Python *libais*¹², adaptée afin de prendre en compte les spécificités des trames issues du CRC, en particulier le format NMEA 4.10. La modification de *libais* était nécessaire car les trames AIS collectées par le dispositif de réception du CRC et par le flux AISHub utilisent des extensions NMEA 4.10 qui ne sont pas entièrement gérées par la version originale de la bibliothèque. En particulier, les trames multiparties formatées en NMEA 4.10 par AISHub ne contiennent les informations de station (s) et d'horodatage (\c) que dans la première partie. Les parties suivantes ne comportent que la clé de groupage (\g) (Figure 3.11). La méthode originale de *libais*, qui regroupait les messages multiparties à partir du numéro de station et de l'ID de séquence AIS, ne pouvait donc pas fonctionner correctement, entraînant une perte d'information ou un assemblage erroné des trames.

¹¹ <https://gpsd.gitlab.io/gpsd/AIVDM.html>

¹² <https://github.com/schwehr/libais>

```
\g:1-2-22890,c:1659589714,s:2954*52\!AIVDM,2,1,7,B,55Uilt800000cQW?07Lnpbq@4dpm2222222220e3`;525wVN313  
\g:2-2-22890*46\!AIVDM,2,2,7,B,s0hCSQDp880,2*7E
```

Figure 3.11 : Trame AIS en deux parties, au format 4.10, issue de la station du CRC.

Pour pallier cette limite, *libais* a été modifiée¹³ pour introduire l'usage du champ `\g` (NMEA 4.10) et d'un identifiant de séquence afin de regrouper correctement les trames, tout en rendant la bibliothèque tolérante à l'absence des champs `\s` et `\c`. Ces changements garantissent un décodage robuste et complet des trames issues du CRC et du flux AISHub. Après recompilation, cette bibliothèque est utilisée pour les opérations suivantes :

1. Extraction, à partir des messages dits « dynamiques » (messages 1, 2, 3, 18, 19 et 27, à l'exclusion des messages 4 et 21 correspondants respectivement aux stations de base et aux aides à la navigation), des informations relatives à la latitude, la longitude, le MMSI, le numéro du message et la classe du navire.
2. Extraction, à partir des messages dits « statiques » (messages 5 et 24), du MMSI, du type et des dimensions du navire, du système de localisation employé (GPS, GLONASS, etc.), ainsi que du numéro du message.

Les messages dynamiques sont ensuite regroupés par MMSI et sous-échantillonnés à une résolution temporelle de dix minutes. Cette procédure permet de contenir la volumétrie tout en limitant l'impact sur l'exhaustivité des positions reçues. Les données résultantes sont stockées dans une base *PostgreSQL* intégrant l'extension *TimescaleDB*, avec indexation temporelle destinée à optimiser les requêtes nécessaires aux traitements ultérieurs. Les messages statiques sont enregistrés dans une table distincte de cette même base, à raison d'une ligne par jour et par MMSI. Les deux tables sont reliées par la clé MMSI, permettant leur croisement lors des analyses.

3.3.2.2 Implémentation de la chaîne de traitement ETL 2

La seconde chaîne de traitement a pour objectif de transformer les données issues de la couche L2 en jeux de données AIS prêts à être exploités pour la modélisation au sein de la couche de données agrégées L3. Cette étape met l'accent sur l'application de contrôles de cohérence, sur le filtrage en fonction des besoins expérimentaux, ainsi que sur la réduction temporelle et l'agrégation des enregistrements.

Le processus s'achève par la production d'exports mensuels au format CSV, assurant une exploitation aisée dans différents environnements d'analyse et garantissant une portabilité maximale. Contrairement à la chaîne de traitement L1–L2, dont la finalité principale réside dans le décodage et

¹³ <https://github.com/ambroisernd/libais>

le nettoyage, la chaîne de traitement L2–L3 privilégie la validation qualitative et la préparation des données en vue de leur utilisation finale.

3.3.2.2.1 Extraction de la base SQL

Pour chaque mois ciblé, cette étape de la chaîne de traitement extrait les enregistrements dynamiques issus de la table correspondante et leur associe, lorsque disponibles, les informations complémentaires provenant de la table des données statiques. La sélection temporelle s’appuie sur un fenêtrage mensuel implémenté par la fonction *time_bucket* de *TimescaleDB*, ce qui permet de restreindre la lecture aux données pertinentes et d’optimiser les accès disque grâce à l’utilisation des *hypertables*.

L’association entre enregistrements dynamiques et statiques est réalisée au moyen d’une jointure latérale (*LATERAL JOIN*). Pour chaque enregistrement dynamique du mois, le système recherche un enregistrement statique correspondant au même MMSI et relevant de la même période. Ce mécanisme conserve systématiquement l’ensemble des enregistrements dynamiques, y compris en l’absence de données statiques correspondantes, garantissant ainsi l’exhaustivité des jeux extraits.

Du point de vue des performances, l’implémentation repose sur plusieurs optimisations :

- Indexation temporelle sur les colonnes *ts* (horodatage UTC) pour exploiter le partitionnement natif de *TimescaleDB*.
- Indexation composée sur (*MMSI, ts*) dans la table statique pour accélérer les recherches par navire et par mois.
- Exécution par lots mensuels, afin de maîtriser la charge mémoire et de paralléliser les extractions.

Les données obtenues sont ensuite transmises aux étapes de validation et d’agrégation de la chaîne de traitement, soit en flux direct, soit via des fichiers intermédiaires.

3.3.2.2.2 Validation des positions

La validation des positions a pour objectif d’identifier et d’exclure les points géographiques présentant des incohérences cinématiques manifestes. Le principe retenu repose sur le calcul de la vitesse instantanée entre deux positions successives transmises par un même navire, identifié par son MMSI. Lorsqu’une vitesse excède un seuil prédéfini, la position correspondante est considérée comme incorrecte et étiquetée en conséquence.

Avant l’application du contrôle cinématique, un filtrage préalable est effectué afin de ne conserver que les coordonnées géographiques valides, à savoir une latitude comprise dans l’intervalle $[-90^\circ, 90^\circ]$ et une longitude comprise dans l’intervalle $[-180^\circ, 180^\circ]$. Les positions de chaque navire

sont ensuite ordonnées chronologiquement. La distance entre deux points successifs est calculée à l'aide d'une approximation équirectangulaire, méthode privilégiée en raison de sa rapidité d'exécution et de sa précision suffisante aux échelles considérées, plutôt que la formule exacte de Haversine. La vitesse est obtenue en divisant cette distance par l'intervalle temporel correspondant, puis exprimée en kilomètres par heure. La règle de validation appliquée est la suivante : le premier point de chaque trajectoire est systématiquement marqué comme valide et les points suivants sont considérés comme corrects si leur vitesse calculée reste inférieure ou égale au seuil fixé, et comme incorrects dans le cas contraire.

Le traitement est parallélisé par exploitation d'un nombre configurable de cœurs de processeur. Les données sont regroupées par MMSI afin que les calculs puissent être réalisés en mémoire, indépendamment pour chaque navire, ce qui favorise l'efficacité et permet une parallélisation optimisée tout en limitant les accès disque à une seule lecture et une seule écriture. Les résultats obtenus sont exportés au format CSV en vue de leur intégration dans les étapes ultérieures de la chaîne de traitement.

3.3.2.2.3 Sous-échantillonnage spatial et temporel

Le sous-échantillonnage spatial et temporel des données AIS a été implémenté en *Python* en s'appuyant sur les bibliothèques *pandas* et *numpy*. Les fichiers AIS bruts sont lus séquentiellement, en ne conservant que les colonnes nécessaires à la modélisation. Un filtre conditionnel est appliqué dès la lecture afin d'exclure les enregistrements invalides ou situés hors de la zone d'étude.

Les coordonnées géographiques sont ensuite projetées sur une grille régulière de résolution $0,25^\circ \times 0,25^\circ$, obtenue par arrondi de la latitude et de la longitude à la cellule la plus proche. Un champ étiquette est ajouté pour indiquer la présence (1) ou l'absence (0) d'un navire dans une cellule donnée, sur la base de l'identifiant de la station réceptrice (CRC ou AISHub). Les données sont ensuite regroupées par tranches horaires et, pour chaque combinaison {heure, cellule spatiale}, la valeur maximale de l'étiquette est retenue. Cette opération revient à indiquer si au moins un message a été reçu dans la maille au cours de l'heure considérée.

Le résultat produit est un tableau réduit, comprenant pour chaque maille horaire les coordonnées de la grille, l'horodatage associé et un indicateur binaire de présence. Cette procédure permet de conserver l'information spatio-temporelle essentielle à la modélisation tout en facilitant l'alignement avec les données environnementales.

3.3.3 Donnée contextuelles et environnementales

En complément des messages AIS, plusieurs sources de données environnementales sont intégrées afin de contextualiser les mouvements des navires et de soutenir la modélisation. Ces jeux

de données apportent des informations complémentaires telles que les conditions météorologiques et la topographie, qui influencent les conditions de réception du signal.

3.3.3.1 Données AISHub

Le jeu de données AIS local est complété avec des messages récupérés depuis la plateforme AISHub. Ces messages suivent le même standard d'encodage NMEA 4.10 et sont soumis aux mêmes étapes de prétraitement que les données locales. La chaîne de traitement inclut le décodage des trames brutes, le sous-échantillonnage et des routines de nettoyage pour éliminer les enregistrements mal formés ou incomplets via l'architecture en médaillon. Avant stockage, une étape de filtrage est appliquée pour exclure les messages en dehors de la zone d'intérêt, à savoir la mer Méditerranée. En termes de volume, le flux AISHub contribue à environ 40 millions de messages par jour pour cette région.

3.3.3.2 Données de réanalyse atmosphérique ERA5

Pour caractériser le contexte environnemental de la propagation des messages AIS, les jeux de données de réanalyse ERA5 fournis par le Copernicus Climate Change Service (C3S)(Hersbach et al. 2020) sont utilisés. Deux jeux de données complémentaires sont exploités :

- Les données ERA5 horaires sur un seul niveau, de 1940 à aujourd'hui (C3S 2018b).
- Les données ERA5 horaires sur niveaux de pression, de 1940 à aujourd'hui (C3S 2018a).

Ces jeux de données fournissent des estimations maillées des variables météorologiques, avec une résolution temporelle horaire et une résolution spatiale de $0,25^\circ \times 0,25^\circ$, permettant une modélisation fine des conditions atmosphériques affectant la propagation du signal AIS.

À partir des données sur un seul niveau, des variables de surface et proches de la surface pertinentes pour les conditions de *ducting* et la modélisation atmosphérique maritime ainsi que pour la propagation terrestre (Nupen 1959) sont extraites :

- Vent et température : vent zonal à 10 mètres, vent méridional à 10 mètres, température de l'air à 2 mètres, température du point de rosée à 2 mètres.
- Pression : pression au sol, pression au niveau moyen de la mer.
- Réfractivité et *ducting* : hauteur de la base du *duct*, gradient vertical moyen de réfractivité dans la couche de piégeage, gradient vertical minimal de réfractivité dans cette couche, hauteur de base et hauteur de sommet de la couche de piégeage.
- Indicateurs océaniques et de vagues : direction moyenne des vagues, période moyenne des vagues, hauteur maximale de vague individuelle, hauteur significative des vagues combinées (vent + houle).

- Précipitations et surface terrestre : précipitations totales, type de précipitation, température de surface de la mer.
- Couverture végétale et indice foliaire (LAI) : couverture végétale haute, couverture végétale basse, indice foliaire des végétaux hauts et bas, type de végétation haute et basse.

Les gradients verticaux de réfractivité importants peuvent induire des conduits de propagation, ils sont encapsulés par des variables ERA5 telles que la base de la couche de piégeage et les gradients de réfractivité (Cornélis 2005 ; Grandjean 2016). Ces variables, combinées aux profils verticaux de température et d'humidité, forment un ensemble de paramètres liés aux phénomènes de propagation troposphérique, incluant le *ducting* et la diffusion due aux variations d'indice de réfraction.

Un second groupe de variables est associé à l'atténuation induite par les hydrométéores, incluant les précipitations totales, le type de précipitations et la teneur en vapeur d'eau ou en nuages. Ces paramètres sont essentiels pour prendre en compte l'interférence du signal en conditions météorologiques dégradées (Mufti 2012 ; Caillet 2015).

Une troisième catégorie concerne les effets d'interaction avec la surface, notamment la couverture végétale, l'indice foliaire (LAI) et les types de terrain. Ces descripteurs de surface permettent de caractériser le bruit parasite (*clutter*), la diffraction près du sol et le blocage local du signal dans les zones côtières ou continentales (ITU-R 2019c).

Enfin, les variables océanographiques telles que la direction moyenne des vagues et la hauteur significative des vagues sont intégrées pour représenter les effets d'interaction avec la surface marine et de propagation multi-trajets. Ces phénomènes peuvent induire des interférences constructives ou destructives selon l'état de surface, et sont particulièrement pertinents pour la modélisation de la réflexion et de la diffusion du signal AIS en mer (Barrick 1971 ; Qinjiang et al. 1998 ; Wang et al. 2018).

Le jeu de données sur niveaux de pression est utilisé pour reconstruire des profils verticaux de l'atmosphère. La température et l'humidité relative aux niveaux de pression suivants (en hPa) sont extraites :

500, 550, 600, 650, 700, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000

Ces niveaux correspondent aux altitudes auxquelles les phénomènes de réfractions sont présents et exercent une influence sur la propagation VHF. Les variables de température et d'humidité relative sont critiques pour l'évaluation des *ducts* élevés et des profils de réfractivité de l'indice radio, notamment le gradient $\frac{dN}{dz}$ (Bean et Dutton 1966 ; ITU-R 2019b), permettant de classer les conditions de réfraction (Mentes et Kaymaz 2007 ; Caillet 2015 ; Zhang et al. 2016).

3.3.3.3 Données d'élévation SRTM

Les données topographiques ont été extraites du jeu de données *Shuttle Radar Topography Mission (SRTM)* (Earth Resources Observation and Science (EROS) Center 2017), qui fournit des mesures d'élévation quasi-globales. La résolution native du SRTM est d'environ 30 mètres. Afin d'assurer la compatibilité avec la résolution d'ERA5, un rééchantillonnage spatial vers une grille de $0,25^\circ \times 0,25^\circ$, a été effectué. Pour chaque cellule, des descripteurs statistiques incluant la moyenne, l'écart-type, le minimum et le maximum d'élévation ont été calculés. Pour gérer les données manquantes, le jeu *SRTM Void Filled* (EROS Center 2017) a été utilisé. Ces variables servent de descripteurs géographiques statiques et peuvent aider à expliquer les variations de réception des messages AIS à proximité des côtes et des zones montagneuses.

3.3.4 Construction du jeu de données

Afin de permettre l'apprentissage supervisé et la modélisation spatio-temporelle, les données AIS doivent être alignées avec les variables environnementales et structurées sur une grille de référence commune. Un jeu de données intégré est construit en combinant les mesures AIS et les descripteurs environnementaux à un pas de temps horaire, sur une grille spatiale de résolution $0,25^\circ \times 0,25^\circ$, cohérente avec les produits de réanalyse ERA5. Chaque cellule de la grille, à un instant donné, contient un vecteur de données composé de caractéristiques extraites de SRTM, de la station réceptrice du CRC et des variables environnementales.

3.3.4.1 Structure du vecteur de données

Pour chaque cellule spatiale et créneau horaire, un vecteur est construit avec les variables caractéristiques suivantes :

- Caractéristiques de la station de base : gain associé au diagramme de rayonnement de l'antenne (tel que dérivé des Figure 3.7 et Figure 3.8) et distance entre la station de base et chaque navire.
- Caractéristiques atmosphériques : variables ERA5 décrites dans la Section 3.3.3.
- Caractéristiques topographiques : statistiques issues des données SRTM, également décrites en Section 3.3.3.

Chaque vecteur est généré pour les cellules individuelles appartenant à un domaine spatial borné, couvrant la région nord-ouest de la Méditerranée, comme représenté en Figure 3.12. Le domaine est défini par les bornes de latitude et de longitude suivantes :

$$\begin{aligned} \text{LAT}_{\min} &= 36^\circ, & \text{LAT}_{\max} &= 45^\circ \\ \text{LON}_{\min} &= 3^\circ, & \text{LON}_{\max} &= 19^\circ \end{aligned}$$

3.3.4.2 Etiquetage supervisé

Pour chaque pas de temps horaire, une grille complète est générée sur la zone d'étude, une étiquette est attribuée à chaque cellule selon la présence et la détection effective de navires par la station locale :

- Une valeur de **1** indique qu'au moins un navire a été détecté dans la cellule et reçu avec succès par la station AIS locale.
- Une valeur de **0** indique qu'un navire était présent dans la cellule selon les données AISHub, mais qu'il n'a pas été reçu par la station locale.
- Une valeur de **-1** signifie qu'aucun navire n'a été détecté dans la cellule, ni par l'antenne locale, ni par les antennes AISHub, durant ce créneau horaire.

Cette stratégie d'étiquetage permet de définir une tâche d'apprentissage supervisé binaire (reçu / non reçu), en excluant les cellules masquées (-1) de l'ensemble d'apprentissage. Elle permet aux modèles d'apprendre les conditions de réception à partir des caractéristiques spatiales, temporelles et environnementales associées à chaque cellule étiquetée.

La Figure 3.12 illustre un exemple de grille spatiale étiquetée à un pas de temps horaire donné. Chaque cellule y est représentée par un point central coloré selon son étiquette : vert pour réception positive (étiquette 1), rouge pour absence de réception locale (étiquette 0). Les cellules avec l'étiquette -1 ne sont pas représentées.

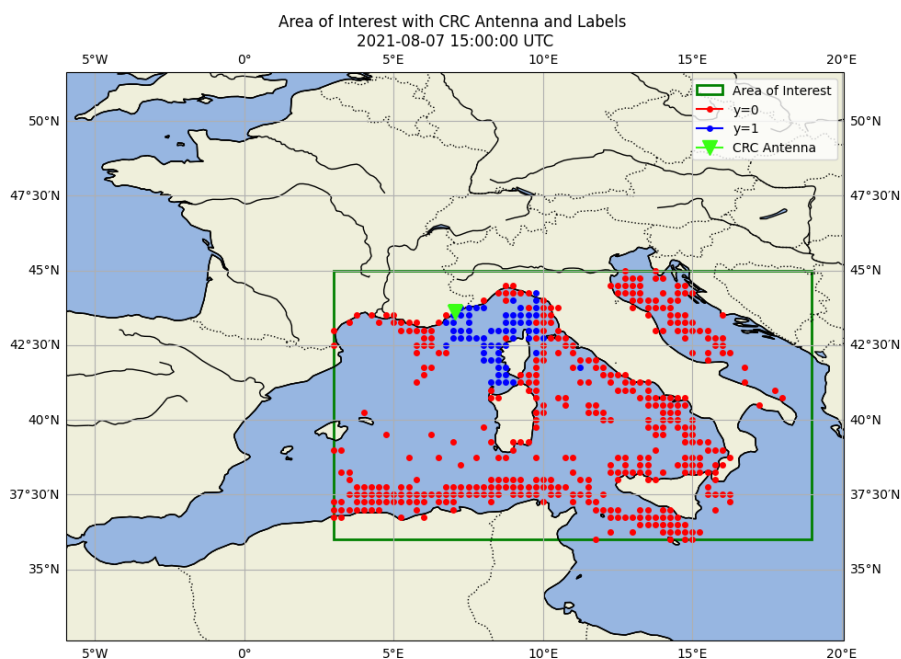


Figure 3.12 : Exemple de grille spatiale étiquetée pour un créneau horaire donné. Chaque cellule contient une étiquette : 1 (reçu localement), 0 (reçu uniquement via AISHub).

3.4 Construction de graphes et ingénierie des caractéristiques

Pour permettre au modèle de capter les dépendances structurelles et directionnelles qui régissent la propagation du signal AIS, l'espace de modélisation est transformé en graphes orientés. Ces graphes sont construits dynamiquement pour chaque créneau horaire, à partir des cellules spatiales de la grille et reflètent la transmission du signal à partir de points d'émission vers un récepteur fixe.

Chaque graphe encode une topologie de propagation localisée, contrainte par une structure conique orientée et chaque nœud porte un vecteur de caractéristiques environnementales et topographiques. Cette section détaille la stratégie de génération des graphes, la définition des nœuds et arêtes, l'encodage des relations directionnelles ainsi que les transformations appliquées aux variables en entrée pour garantir une homogénéité des représentations et une convergence stable du modèle.

3.4.1 Définition des nœuds

Chaque nœud du graphe correspond à une cellule de la grille spatiale horaire introduite en Section 3.3.4. Ces cellules sont définies à une résolution spatiale fixe ($0,25^\circ \times 0,25^\circ$) et répétées à chaque pas de temps horaire. Le vecteur de caractéristiques associé à chaque nœud comprend l'ensemble des descripteurs environnementaux extraits du vecteur de données précédemment décrit. Ces caractéristiques sont calculées par cellule et par créneau horaire, et sont supposées alignées et synchronisées avec les données de réception AIS (van der Werff et al. 2024).

Les graphes sont générés séparément pour chaque créneau horaire. Pour chaque grille temporelle, l'ensemble de nœuds associés est traité afin de construire plusieurs graphes orientés. Plus précisément, un graphe est généré pour chaque nœud dont l'étiquette est différente de -1 . Chaque nœud étiqueté (avec une valeur 0 ou 1) est considéré comme un transmetteur (Tx), et un graphe est construit pour modéliser la propagation du signal à partir de ce nœud vers un nœud récepteur (Rx) fixe représentant la station de base AIS (Pedersen et al. 2012).

3.4.2 Construction des arêtes

Pour construire ces graphes, une structure de message-passing basée sur la topologie locale de la grille est définie. À partir du nœud Tx, une connexion itérative est établie entre celui-ci et son voisinage de Moore (c'est à dire, les cellules situées à une distance de Chebyshev $\|\cdot\|_\infty = 1$). Ce processus se poursuit de manière récursive, en connectant les voisins à leurs propres voisins, jusqu'à atteindre le nœud Rx. Le graphe obtenu est orienté, avec des arêtes suivant la direction de propagation de Tx vers Rx.

Le cône directionnel est défini comme la région géométrique orientée le long de l'axe Tx–Rx, à l'intérieur de laquelle les trajets de propagation sont pris en compte. L'ouverture du cône α (un

hyperparamètre du modèle) correspond à l'angle total de ce cône directionnel. Seuls les nœuds situés à l'intérieur de ce cône sont conservés dans le graphe final. La procédure complète de génération de graphe est détaillée dans l'Algorithme 3.1.

Données : Grille G_t pour l'intervalle de temps t , étiquettes L , position du récepteur Rx , ouverture du cône α

Résultat : Ensemble de graphes $\{G_i\}$ associés à leurs étiquettes

pour tous les cellules $v_i \in G_t$ faire

- si $L[v_i] \in \{0, 1\}$ alors**
 - Définir v_i comme émetteur (Tx)
 - Initialiser un graphe G_i avec v_i comme nœud source
 - Initialiser une file $Q \leftarrow \{v_i\}$
 - tant que Q non vide et Rx non atteint faire**
 - Extraire v_j de Q
 - pour tous les voisins v_k de v_j dans le voisinage de Moore faire**
 - si v_k est dans le cône d'ouverture α centré sur Tx alors**
 - Ajouter une arête $v_j \rightarrow v_k$ à G_i
 - Ajouter v_k à Q si non déjà visité
 - Sauvegarder G_i avec l'étiquette $L[v_i]$

Algorithme 3.1 : Construction de graphes à partir de cellules étiquetées d'une grille

Cette procédure garantit que chaque graphe correspond à un scénario de propagation plausible, associé à une cellule de grille étiquetée. L'utilisation d'une structure orientée permet de capturer la nature directionnelle de la propagation des ondes radio. Le graphe résultant n'est cependant pas nécessairement acyclique, des arêtes bidirectionnelles pouvant apparaître selon la configuration spatiale. Le filtrage angulaire permet de contrôler la complexité du graphe et son orientation (Qiao et al. 2018).

La Figure 3.13 illustre l'évolution entre une approche naïve basée sur la grille et la stratégie de modélisation par graphe adoptée ici. Le panneau gauche de la Figure 3.13 montre le modèle de propagation en ligne droite exploré dans des travaux précédents (Vanche et al. 2022) où le chemin $Tx \rightarrow Rx$ est défini comme un vecteur traversant toutes les cellules de grille qu'il intersecte. Bien que simple à implémenter, cette méthode ne prend pas en compte les interactions spatiales locales ni la variabilité environnementale entourant l'axe principal de propagation.

Le panneau droit de la Figure 3.13 présente le modèle de propagation basé sur un graphe, où chaque nœud est initialement connecté à ses huit voisins afin de permettre une expansion récursive de Tx vers Rx . Une contrainte directionnelle sous forme d'un cône d'ouverture α centré sur l'axe $Tx \rightarrow$

Rx est introduite. Ce filtre géométrique concentre la structure du graphe sur les régions les plus susceptibles de contribuer à la propagation du signal, réduisant ainsi l'influence des zones marginales. L'angle α constitue un hyperparamètre réglable permettant de contrôler l'équilibre entre expressivité et sélectivité.

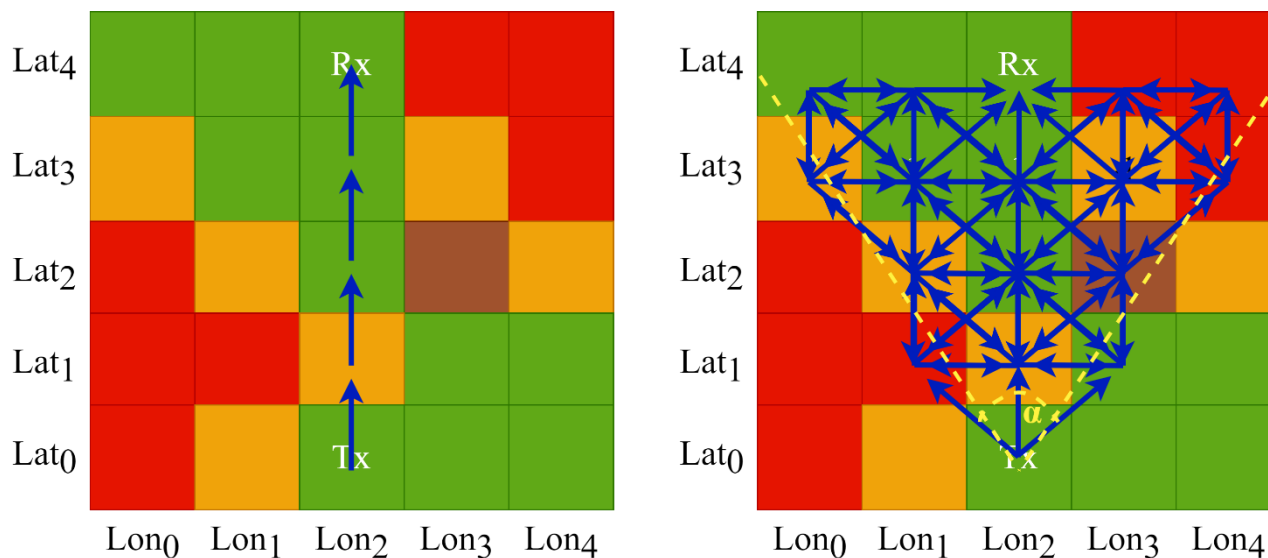


Figure 3.13 : Comparaison entre différentes stratégies de modélisation de la propagation d'un émetteur (Tx) vers un récepteur fixe (Rx), en fonction des conditions de propagation (vert : bonnes, orange : moyennes, rouge : mauvaises, marron : effets du relief). À gauche : une propagation possible représentée par un vecteur direct naïf. À droite : des propagations complexes représentées par un graphe filtré, contraint par un cône directionnel réduisant la dispersion spatiale et mettant l'accent sur la sous-région la plus pertinente.

La Figure 3.14 présente la chaîne de traitement pour la préparation du jeu de données et la construction des graphes. Les données brutes hétérogènes (CSV d'étiquettes, fichiers NetCDF ERA5 à un ou plusieurs niveaux, données topographiques) sont d'abord fusionnées. Ces données sont ensuite projetées sur des grilles spatiales horaires, où chaque cellule correspond à un nœud potentiel. Pour chaque grille horaire, des sous-graphes sont construits en centrant la propagation sur les scénarios d'émissions/réceptions. Ce processus aboutit à la génération de trois ensembles : les nœuds, les arêtes et les étiquettes.

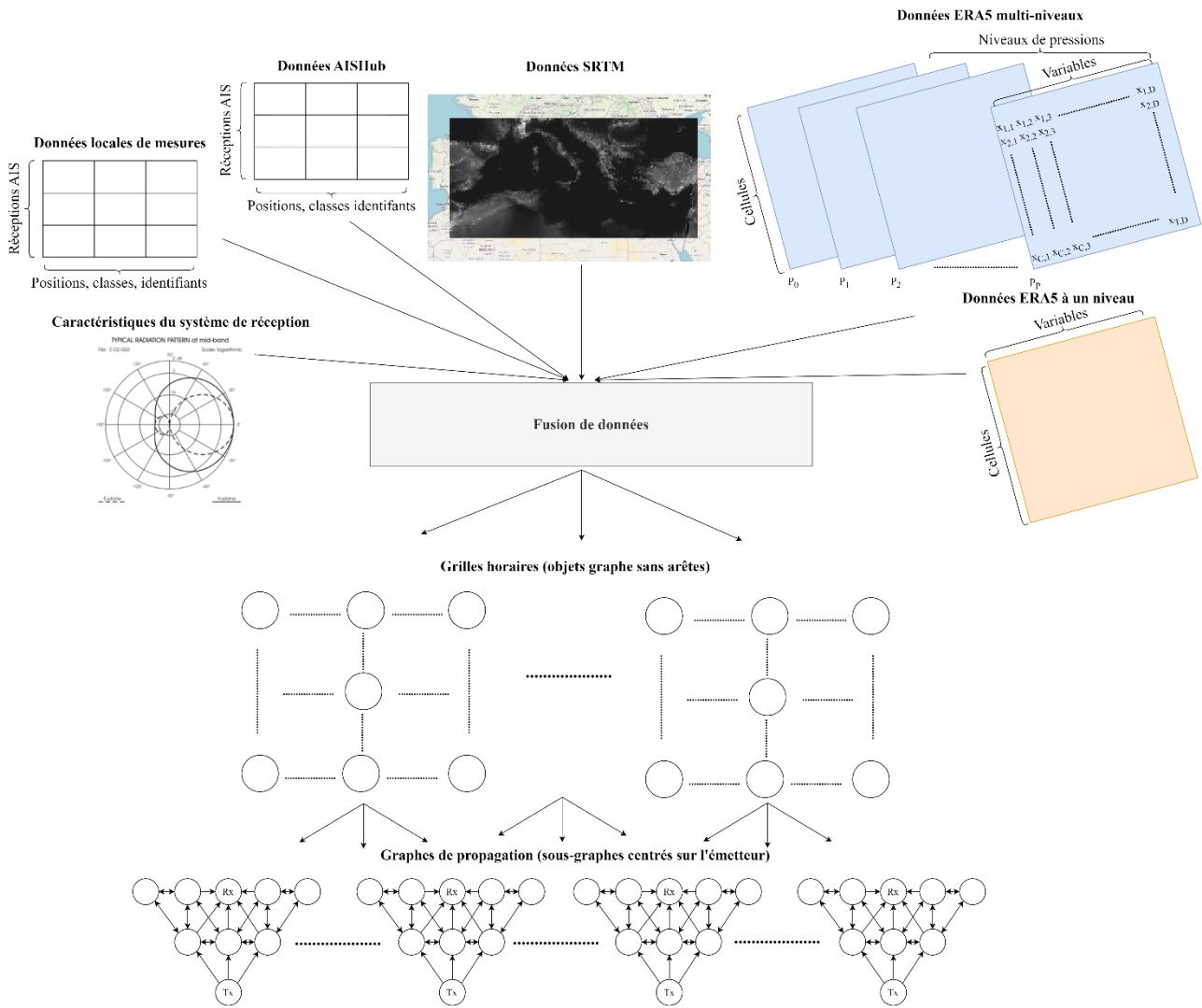


Figure 3.14 : Chaîne de traitement pour la construction des graphes

3.4.3 Standardisation et transformations

Afin de réduire l'hétérogénéité des ordres de grandeur entre les différentes caractéristiques et de faciliter la convergence lors de l'apprentissage, toutes les caractéristiques des nœuds sont standardisées. Une normalisation min-max est appliquée, qui ramène chaque variable dans l'intervalle $[0, 1]$ à partir des valeurs minimale et maximale observées sur l'ensemble du jeu de données. Cette méthode est particulièrement adaptée aux variables physiques naturellement bornées (Patro et Sahu 2015).

3.5 Conclusion

Ce chapitre a présenté l'ensemble de la méthodologie développée pour modéliser la réception AIS par apprentissage profond sur graphes. L'architecture proposée s'appuie sur une formalisation explicite des interactions spatiales entre cellules géographiques, combinée à l'exploitation de données environnementales et multi-sources. Le recours à GraphSAGE, enrichi par des mécanismes

d'agrégation avancés (*Jumping Knowledge*, attention, Set2Set), permet d'encoder des relations locales et globales de manière flexible et expressive.

Les choix de conception : structuration du jeu de données, architecture en médaillon, génération dynamique de graphes orientés, standardisation des variables sont motivés à la fois par les contraintes du domaine (variabilité des conditions de propagation, contraintes directionnelles, influence de l'environnement) et par les exigences du modèle (stabilité, expressivité, efficacité computationnelle). Cette méthodologie permet de transformer un problème physique complexe en une tâche formelle de classification de graphes, optimisable par des techniques d'apprentissage automatique.

Le chapitre suivant est consacré à l'évaluation expérimentale du modèle proposé. Il présentera les jeux de données utilisés pour l'entraînement et la validation, les métriques de performance, les résultats obtenus ainsi que leur analyse.

Chapitre 4 : Evaluation expérimentale de la démarche proposée

Après la formalisation de la démarche de modélisation de la réception de l'AIS ainsi que la conception d'une architecture neuronale à base de graphes destinée à la prédiction des zones de réception en environnement maritime, ce chapitre est consacré à l'évaluation expérimentale de l'approche proposée. L'objectif est de valider empiriquement la pertinence des choix méthodologiques présentés précédemment en les confrontant à des données réelles collectées sur une période de quatre années et couvrant un périmètre géographique représentatif.

L'évaluation s'appuie sur un protocole expérimental dans lequel les performances du modèle sont mesurées sur un jeu de test indépendant, au moyen de métriques standards utilisées dans les tâches de classification binaire. Ces performances sont comparées à celles de plusieurs modèles de référence : un modèle déterministe fondé sur la physique (ITU-R P.2001), un modèle d'apprentissage automatique de type XGBoost, ainsi que deux architectures de réseaux de neurones en graphes (GCN et GAT). Le chapitre décrit également l'implémentation logicielle de la chaîne de traitement, les ressources matérielles mobilisées, la construction des graphes, l'ingénierie des données et le paramétrage des expériences. Enfin, une analyse des résultats est menée, incluant la visualisation des prédictions obtenues et une discussion relative aux limites inhérentes à l'approche.

4.1 Vue d'ensemble de la démarche expérimentale

La chaîne de traitement complète du modèle est illustrée en Figure 4.1. Elle couvre l'ensemble du processus, depuis la construction des graphes et la normalisation des caractéristiques, jusqu'aux étapes d'entraînement, de validation, d'évaluation et de génération des prédictions. Chaque graphe est défini autour d'un nœud transmetteur étiqueté, les arêtes étant établies conformément à un cône directionnel tel que spécifié en Section 3.4.2. L'entraînement est formulé comme une tâche de classification binaire au niveau du graphe, et les performances sont évaluées à l'aide de métriques standards calculées sur un ensemble de graphes de test indépendants.

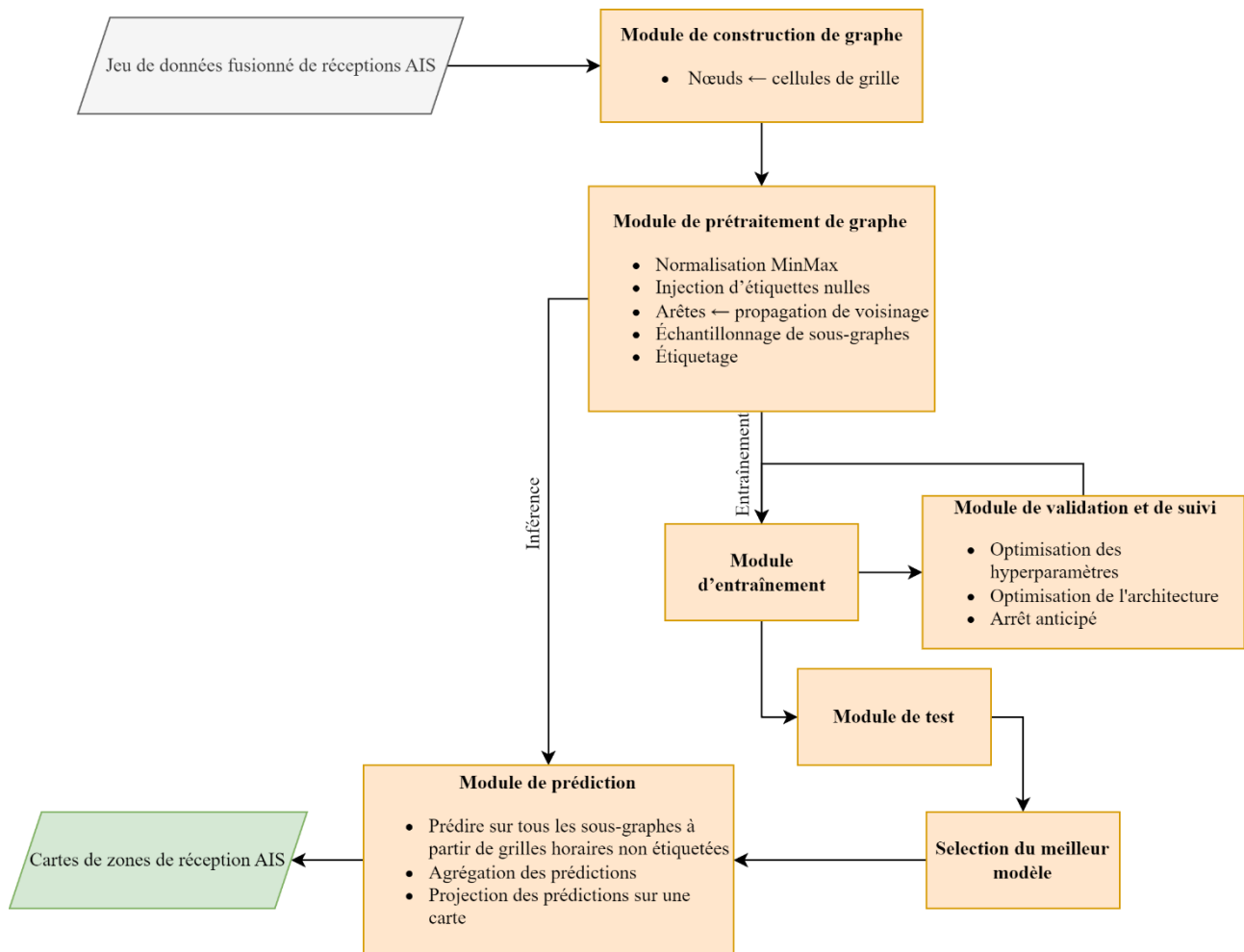


Figure 4.1 : Vue d'ensemble de la chaîne de traitement pour la modélisation.

4.2 Présentation des modèles de référence

Dans le but de situer les performances du modèle proposé dans un cadre comparatif robuste, plusieurs modèles de référence appartenant à différentes familles méthodologiques ont été retenus. Ces modèles se distinguent par la nature de leur représentation des données, leur aptitude à intégrer des relations spatiales ainsi que par leur niveau de complexité algorithmique.

Chacun repose sur des hypothèses distinctes concernant le processus de propagation radio et s'appuie sur des formats d'entrée spécifiques. La diversité ainsi introduite permet de confronter l'approche développée à des méthodes établies, tout en quantifiant l'apport de la représentation en graphes et des mécanismes d'agrégation neuronale. Cette section expose les fondements théoriques et pratiques de chacun de ces modèles, ainsi que les conditions expérimentales définies pour leur implémentation.

4.2.1 Caractéristiques des modèles

L'évaluation des performances du modèle à base de graphes est réalisée par comparaison avec un ensemble de modèles de référence appartenant à trois catégories : les modèles traditionnels fondés

sur la physique, les algorithmes classiques d'apprentissage automatique et d'autres architectures de réseaux de neurones en graphes. Chacun de ces modèles mobilise une représentation spécifique des données en entrée et présente des biais ainsi que des limitations propres.

Le Tableau 4.1 synthétise les principales caractéristiques des modèles considérés. Il met en évidence les différences relatives à la structure des données d'entrée, aux mécanismes de modélisation spatiale et à la nature de l'approche sous-jacente. Le modèle ITU-R P.2001 repose sur un formalisme semi-déterministe combinant physique et statistiques, tandis que XGBoost exploite des données tabulaires dans un cadre d'apprentissage automatique classique. Les architectures de type GCN et GAT utilisent une représentation en graphes, respectivement avec un schéma d'agrégation fixe et un mécanisme fondé sur l'attention. Enfin, le modèle proposé s'appuie également sur une représentation en graphes, mais introduit un mécanisme d'agrégation hiérarchique, permettant de capter de manière plus fine les dépendances spatiales.

Modèle	Structure d'entrée	Modélisation spatiale	Type
ITU-R P.2001	Voir Tableau 4.2 + Diagrammes dans (ITU-R 2023c)	Chemin de propagation analytique en ligne droite	Basé sur la physique et les statistiques
XGBoost	Données tabulaires	Vecteur direct en ligne droite	ML
GCN	Graphes	Agrégation fixe	GNN
GAT	Graphes	Agrégation basée sur l'attention	GNN
Modèle proposé	Graphes	Agrégation hiérarchique	GNN

Tableau 4.1 : Résumé des modèles de référence et du modèle proposé pour la prédiction de la réception AIS.

4.2.2 Modèle physique : ITU-R P.2001

Le modèle ITU-R P.2001 (ITU-R 2023c) constitue une référence déterministe fondée sur les principes physiques de la propagation des ondes radioélectriques. Il estime la perte de transmission entre un émetteur et un récepteur à partir de données environnementales détaillées, comprenant notamment les profils topographiques, la réfractivité atmosphérique, la nature des surfaces et les caractéristiques des antennes. Conçu pour des communications point-à-point à longue distance, ce modèle est largement mobilisé dans les applications d'ingénierie des télécommunications.

Dans ce cadre, un message est considéré comme reçu lorsque la puissance prédite au niveau du récepteur dépasse la sensibilité de l'équipement AIS du CRC, (-112 dBm). Bien que rigoureusement fondé sur la théorie électromagnétique, ce modèle demeure sensible aux incertitudes présentes dans

les variables d'entrée et engendre un coût de calcul élevé à l'inférence. Les paramètres pris en compte sont synthétisés dans le Tableau 4.2.

4.2.2.1 Paramétrage de la variable Tpc

Pour la variable Tpc définie comme « le pourcentage d'une année moyenne durant lequel l'affaiblissement de transmission de référence prévu n'est pas dépassé » (ITU-R 2023c). Elle correspond à la disponibilité des conditions de propagation dites normales. L'analyse statistique de trois années de données montre que, dans le modèle 4/3, les navires sont détectés au-delà de l'horizon radioélectrique en moyenne 45 % du temps (Tpc = 55 %).

Conformément à la recommandation (ITU-R 2022), Tpc peut être convertie en sa valeur pour le pire mois, notée Tpc_w en appliquant la relation :

$$Tpc_w = Q \cdot Tpc$$

Où le facteur de conversion Q dépend du paramètre N_s , l'indice de réfraction atmosphérique en surface. Les valeurs de N_s sont fournies par (ITU-R 2019b) à partir de l'indice de réfraction atmosphérique au niveau de la mer N_0 , disponible sous forme de cartes mensuelles (Figure 4.2).

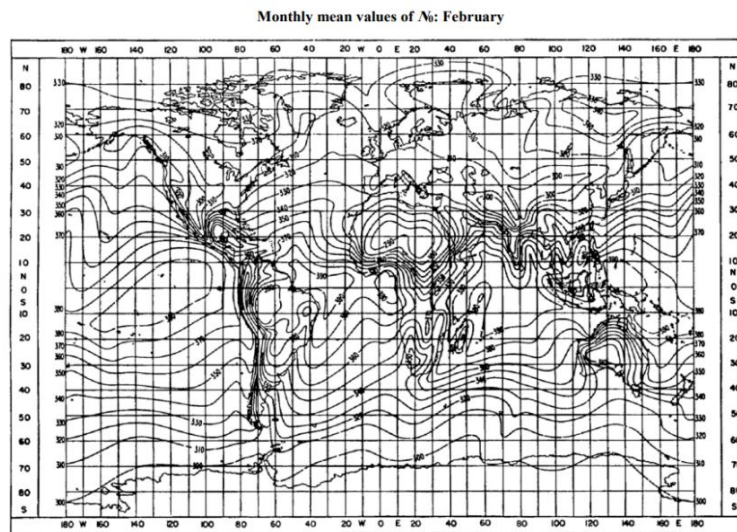


Figure 4.2 : Valeurs moyenne de N_0 données par (ITU-R 2019b) pour le mois de février..

La méthode repose sur le calcul du facteur Q à l'aide des paramètres suivants :

$$Q = 5,8 \exp\left(\frac{N_s}{75}\right) \text{ et } \beta = 0,13$$

permettant d'intégrer les phénomènes de propagation transhorizon au-dessus de la mer (ITU-R 2022). En utilisant des valeurs de Tpc se situant dans l'intervalle $Tpc > 30 \%$, il apparaît que $Tpc_w = 93,5 \%$, correspondant au mois de février où $N_s = 330$, ce résultat est fortement corrélé à $1 - p_{duct}$

où p_{duct} désigne la probabilité d'apparition d'un conduit de propagation observée statistiquement une étude climatologique en Méditerranée (von Engeln et Teixeira 2004).

4.2.2.2 Approche dynamique avec ERA5

Dans cette seconde méthode, N_s est calculé avec une résolution spatiale de $0,25^\circ$ et une résolution temporelle horaire à partir des variables atmosphériques ERA5 (température, pression et humidité relative) et injecté dans l'équation de Q . L'indice de réfraction de surface est alors évalué à partir de la relation :

$$N_0 = \frac{77,6}{T_0} \left(P_0 + \frac{4810e}{T_0} \right)$$

Où $P_0 = 1000$ hPa et T_0 la température à 1000 hPa et,

$$e = \frac{H_0 e_s}{100}, \quad e_s = 6,1094 \exp\left(\frac{17,625T}{243,04 + T}\right)$$

avec H_0 l'humidité relative à 1000 hPa et e_s la pression de vapeur saturante (Alduchov et Eskridge 1996), dont l'erreur relative est inférieure à 0,4 % de -40 à 50°C .

Cette méthode permet d'obtenir pour le mois de février $\text{Tpc}_w = 97,4$ %.

En complément, une seconde estimation de Tpc est réalisée à partir de la hauteur de base du *duct* : à l'échelle horaire, la proportion de temps pendant lequel cette hauteur dépasse un seuil fixé est utilisée comme indicateur du pourcentage de conditions de propagation anormales. Cette méthode exploite directement les phénomènes physiques responsables du *ducting* et sa corrélation avec $1 - p_{\text{duct}}$. Ainsi la valeur obtenue est en moyenne pour le mois de février de $1 - p_{\text{duct}} = 93,7$ %, à l'année cette valeur est de 53,50 %

La comparaison des différentes approches (méthode ITU basée sur N_0 , estimation dynamique via ERA5 à partir de N_0 calculé, et proxy basé sur la hauteur de base du *duct*) montre que les valeurs de Tpc_w obtenues sont similaires. Cette convergence constitue un argument en faveur de la validité du proxy utilisé basé sur ERA5 qui présente une résolution spatio-temporelle nettement plus fine que les données fournies par l'ITU.

Ainsi, l'exploitation directe des conditions de *ducting* issues d'ERA5 constitue une estimation physiquement fondée, robuste et cohérente avec les méthodes ITU-R, tout en offrant un gain significatif en termes de précision temporelle et spatiale.

Variable	Type	Unité	Description	Valeur
d	Vecteur	km	Distances du profil de terrain	Données SRTM
h	Vecteur	m	Hauteurs du profil de terrain	Données SRTM
z	Vecteur	-	Code de zone (1 : Mer, 3 : Côtier, 4 : Terres)	Données SRTM
GHz	Scalaire	GHz	Fréquence	0,162
Tpc	Scalaire	%	Pourcentage de temps pour lequel la perte n'est pas dépassée	-
Phire	Scalaire	deg	Longitude du récepteur	7,0
Phirn	Scalaire	deg	Latitude du récepteur	43,5
Phite	Scalaire	deg	Longitude de l'émetteur	-
Phitn	Scalaire	deg	Latitude de l'émetteur	-
Hrg	Scalaire	m	Hauteur de l'antenne réceptrice	188
Htg	Scalaire	m	Hauteur de l'antenne émettrice	15
Grg	Scalaire	dBi	Gain de l'antenne réceptrice (vers l'émetteur)	Voir Figure 3.7
Gtg	Scalaire	dBi	Gain de l'antenne émettrice (vers le récepteur)	5
FlagVP	Scalaire	-	Polarisation (1 : Verticale, 0 : Horizontale)	0

Tableau 4.2 : Paramètres requis pour les prédictions à l'aide du modèle ITU-R P.2001, ainsi que les valeurs utilisées lors de sa mise en œuvre.

4.2.3 Modèle d'apprentissage automatique : XGBoost

XGBoost (Chen et Guestrin 2016) correspond à une méthode d'ensemble fondée sur des arbres de décision, ayant démontré une efficacité notable dans un large éventail de tâches d'apprentissage supervisé à partir de données tabulaires. Ici, le modèle exploite en entrée un ensemble de variables environnementales et géographiques extraites pour chaque paire émetteur-récepteur (Tx-Rx). Ces variables sont agrégées sous la forme d'un vecteur, construit le long du chemin de propagation en ligne droite reliant les deux points. Bien que XGBoost présente des avantages en termes de rapidité et

d'interprétabilité, il ne permet pas de représenter ni de modéliser explicitement des structures de type graphe.

4.2.4 Modèle à base de graphes : GCN et GAT

GCN (Kipf et Welling 2017) constitue une architecture fondatrice de l'apprentissage sur graphes. Il repose sur un mécanisme d'agrégation des caractéristiques des nœuds à partir de leur voisinage local, effectué par une moyenne à poids fixes. Cette approche se révèle efficace pour des architectures peu profondes, mais elle est sujette au phénomène de sur-lissage lorsque la profondeur du réseau augmente (Li et al. 2018). Dans la présente implémentation, le GCN est configuré avec cinq couches de 128 neurones chacune, activées par des fonctions ReLU. Cette architecture sert de référence pour l'évaluation des performances des modèles à base de graphes.

L'architecture GAT (Veličković et al. 2018) constitue une extension du GCN en introduisant un mécanisme d'attention permettant d'apprendre des poids différenciés entre nœuds voisins lors de l'agrégation. Ce procédé offre au modèle la capacité de mettre l'accent sur les parties les plus informatives du graphe. Dans le contexte de la propagation AIS, il peut ainsi mettre en évidence les régions les plus influentes le long du chemin de propagation. Néanmoins, l'utilisation de ce mécanisme d'attention entraîne un coût computationnel accru, et les performances peuvent se dégrader en présence de graphes bruités (Veličković et al. 2018 ; Dwivedi et al. 2023). L'implémentation considérée repose sur cinq couches de 128 neurones et un schéma d'attention multi-têtes comprenant une tête par couche.

4.3 Implémentation logicielle

Cette section expose les aspects techniques relatifs à l'implémentation de la méthodologie expérimentale. Elle décrit l'implémentation des modèles, ainsi que les modules destinés à la génération et au prétraitement des graphes, et à la gestion des différentes étapes de l'apprentissage supervisé.

L'ensemble du processus a été conçu selon une approche modulaire et reproductible, de manière à assurer la traçabilité des transformations de données, la répliquabilité des expériences et l'optimisation de l'utilisation des ressources de calcul. Les paramètres d'entraînement, la stratégie de partitionnement des ensembles de données et les méthodes de réglage des hyperparamètres sont par la suite détaillées, avec une attention particulière portée aux mesures mises en œuvre pour limiter les déséquilibres de classes et garantir une couverture spatiale représentative.

4.3.1 Implémentation des modèles

Cette section décrit l'implémentation logicielle des différents modèles étudiés. Pour les modèles GNN, elle détaille l'implémentation de la chaîne de traitement pour la modélisation, c'est à dire, les modules de construction et de prétraitement des graphes, ainsi que les modules dédiés à l'entraînement, à la validation et au suivi des performances.

4.3.1.1 Implémentation du modèle physique

Le modèle déterministe basé sur la physique, ITU-R P.2001, a été implémenté en *Python* à l'aide du package *Py2001* version 4.0 (Stevanovic, s. d.). Cette implémentation constitue une traduction en *Python* de la version de référence officielle MATLAB/Octave de la Recommandation, telle qu'approuvée par le groupe de travail *ITU-R 3M* et publiée par le Groupe d'étude 3 sur le portail *ITU-R SG 3 Software, Data and Validation*¹⁴.

4.3.1.2 Implémentation du modèle ML

Le modèle XGBoost a été implémenté en *Python* en utilisant la bibliothèque *xgboost* version 3.0.3 (Chen et Guestrin 2016) et son entraînement a été orchestré avec *Scikit-learn* version 1.5.1 (Pedregosa et al. 2011), une librairie utilisée principalement pour l'entraînement et l'évaluation des modèles tabulaires.

4.3.1.3 Implémentation des modèles GNN

Les GNN (GCN, GAT, GNN proposé) ont été implémentés en *Python* en utilisant *PyTorch Geometric* version 2.5.3 (Fey et Lenssen 2019) comme *framework* principal pour les couches de réseaux de neurones graphes et les opérations de *message passing*. *PyTorch* version 2.3.1 (Ansel et al. 2024) a été utilisé comme *backend* pour les calculs tensoriels, et les processus d'entraînement ont été gérés via *PyTorch Lightning* version 2.3.3 (Falcon et The PyTorch Lightning team 2019).

4.3.1.4 Implémentation de la chaîne de traitement pour la modélisation par GNN

4.3.1.4.1 Module de construction des graphes

Objectif fonctionnel

Ce module construit, pour chaque pas de temps horaire, les nœuds d'un graphe structurant les données issues :

- Des observations AIS (positions de navires, stations, classes).
- Des données météorologiques ERA5 (niveaux de pression et un niveau).
- Des statistiques topographiques SRTM.

Chaque nœud du graphe représente une cellule géographique, enrichie par des variables physiques et spatiales. L'étiquette binaire associée à chaque nœud indique la présence ou l'absence d'un signal reçu par la station du CRC. Le Tableau 4.3 présente les entrées et les sorties de ce module.

¹⁴ <https://www.itu.int/en/ITU-R/study-groups/rsg3/pages/iono-tropo-spheric.aspx>

Élément	Description
Entrées	<ul style="list-style-type: none"> – Fichiers AIS au format CSV – Données ERA5 au format NetCDF – Données SRTM prétraitées au format CSV
Sorties	<ul style="list-style-type: none"> – Un objet de type graphe contenant : – Les caractéristiques des nœuds – Les étiquettes binaires – Les coordonnées des nœuds – L’horodatage du graphe – Le nom des variables caractéristiques

Tableau 4.3 : Entrées et sorties du module de construction des graphes.

Déroulement et description interne

Le module de construction de graphe repose sur la classe *AISDatasetGrid*, héritant de l’interface *Dataset* (*torch_geometric.datasets*) fournie par la bibliothèque *PyTorch Geometric*. Ce choix d’implémentation permet d’assurer une compatibilité directe avec les bibliothèques de graphes neuronaux, tout en structurant le processus. Le traitement s’opère en plusieurs étapes :

1. Agrégation et préparation des données brutes

Les fichiers AIS, ERA5 (niveaux de pressions et surface) ainsi que les statistiques SRTM sont d’abord indexés puis chargés. Les traitements suivants sont réalisés :

- Les observations AIS sont regroupées par date et par heure, avec une discrétisation spatiale fixée à une résolution de $0,25^\circ \times 0,25^\circ$. Chaque cellule reçoit une étiquette binaire indiquant la présence ou l’absence de détection par la station CRC.
- Les données météorologiques ERA5 sont filtrées temporellement pour ne retenir que les pas horaires correspondant aux observations AIS.
- Les variables topographiques issues de SRTM (statistiques locales) sont fusionnées sur la grille spatiale.

2. Construction de la grille géographique

Une grille régulière est générée à partir des bornes géographiques définies ($Lat_{min}, Lat_{max}, Lon_{min}, Lon_{max}$) et de la résolution choisie. Chaque cellule de la grille devient un candidat pour la création d’un nœud de graphe.

3. Enrichissement des nœuds

Pour chaque cellule :

- Les variables météorologiques (pression, humidité relative, température, variables de surface) sont projetées sous forme de colonnes, les niveaux de pression étant réindexés en colonnes distinctes.
- Des variables dérivées sont calculées : distance au CRC et azimuth (comme clé du gain directif), permettant de contextualiser spatialement la propagation du signal.
- Les statistiques SRTM sont également ajoutées à ces variables caractéristiques.

4. Gestion des étiquettes

Chaque nœud se voit attribuer une étiquette initialisée à -1 . En cas de correspondance avec les observations AIS, l'étiquette est mise à jour à 1 ou à 0. Cette stratégie permet de conserver l'intégralité du maillage spatial, tout en différenciant les cellules contenant des observations des cellules purement contextuelles.

5. Encapsulation dans un objet graphe

L'ensemble des nœuds et de leurs caractéristiques est converti en un objet *torch_geometric.data.Data*, contenant : La matrice des caractéristiques X , le vecteur des étiquettes Y , les coordonnées spatiales, l'horodatage du graphe ainsi que la liste des noms de variables.

6. Sauvegarde et réutilisabilité

Chaque graphe est sérialisé au format binaire *.pt*, ce qui permet un accès direct et optimisé lors des étapes ultérieures de prétraitement et d'apprentissage. La gestion de la persistance inclut un mécanisme de vérification pour éviter les traitements redondants.

A ce stade, les graphes produits ne contiennent pas d'arêtes et ne correspondent pas à un scénario de propagation, mais à l'ensemble des observations sur l'heure. Ils se résument alors à des ensembles de nœuds. Ce choix est réalisé afin de réduire la redondance d'informations entre scénario de propagation et ainsi optimiser le stockage des graphes. Les arêtes associées à chaque propagation seront générées dynamiquement dans le prochain module.

4.3.1.4.2 Module de prétraitement des graphes

Objectif fonctionnel

Ce module prépare les graphes construits en entrée pour l'entraînement des modèles GNN. Il s'agit de générer, pour chaque pas de temps, un ou plusieurs sous-graphes dont le départ est le nœud d'émission du message AIS, avec :

- Un filtrage spatial pour éliminer les zones non pertinentes.
- Une sélection équilibrée des exemples selon les classes.
- Un calcul des arêtes du sous-graphe (cône de propagation).
- Une normalisation des caractéristiques.
- Une prise en compte du diagramme de directivité de l’antenne réceptrice.

Le Tableau 4.4 présente les entrées et les sorties de ce module

Élément	Description
Entrées	– Objet graphe produit par le module de construction de graphes
Sorties	<ul style="list-style-type: none"> – Lot de sous graphes comprenant : – Un vecteur de caractéristiques des nœuds (normalisé) : <ul style="list-style-type: none"> ○ Données ERA5 ○ Données SRTM ○ Données de directivité – Une liste des arêtes du sous-graphe – L’étiquette associée au nœud d’émission – Des métadonnées (horodatage, coordonnées des nœuds etc.)

Tableau 4.4 : Entrées et sorties du module de prétraitement des graphes.

Déroulement et description interne

Ce module assure la préparation des graphes initiaux afin de les transformer en sous-graphes exploitables par les modèles de type GNN. L’implémentation repose sur une combinaison de filtres spatiaux, de règles d’équilibrage des classes et de normalisation des caractéristiques. Le processus s’articule en plusieurs étapes successives :

1. Processus de génération d’étiquettes 0, décrit en Section 4.3.4
2. Sélection équilibrée des échantillons

Les indices correspondants aux classes 0 et 1 sont extraits du graphe. En cas de déséquilibre, un sous-échantillonnage aléatoire est appliqué de façon à obtenir une distribution équilibrée. Ce mécanisme garantit que les modèles entraînés ne privilégient pas systématiquement la classe majoritaire.

3. Construction des arêtes

Pour chaque nœud émetteur sélectionné, une fonction calcule la structure locale du sous-graphe. Elle implémente l’algorithme décrit en Section 3.4.2 afin de construire les arêtes de manière contrainte par

un cône de propagation délimité par un angle configurable. Cette modélisation explicite la directivité spatiale de la propagation radio, en intégrant la position relative du CRC et de l'émetteur AIS.

4. Prétraitement des variables caractéristiques

Une fonction standardise les vecteurs de variables caractéristique :

- Remplacement des valeurs manquantes.
- Sélections des variables pertinentes (définies dans un fichier de configuration).
- Transformation de l'azimut en une valeur reflétant le gain directif de l'antenne réceptrice, suivant une quantification angulaire de 30°.
- Une normalisation additionnelle peut être appliquée, ici *minmax* (Section 3.4.3). En pratique, cette normalisation est précalculée sur l'ensemble du jeu de données AIS de manière incrémentale, afin de prendre en compte la distribution réelle des données et ne pas surcharger la mémoire.

5. Génération de sous-graphes

En considérant les nœuds, leurs variables caractéristiques prétraitées, l'étiquette associée ainsi que les arêtes générées, des graphes sont construits. Plusieurs stratégies sont implémentées :

- Génération d'un seul sous-graphe par grille.
- Génération d'un lot de sous-graphe sur une même grille.
- Génération déterministe : assure une reproductibilité en mémorisant les indices sélectionnés dans un objet persistant pour chaque grille. Utilisé notamment en phase de test.

En somme, ce module opère comme une chaîne de transformation : il convertit des graphes représentant l'ensemble des observations à une heure donnée en sous-ensembles centrés sur une observation.

4.3.1.4.3 Module d'entraînement, validation et de suivi

Objectif fonctionnel

Ce module prend en charge l'entraînement du modèle GNN à partir des sous-graphes fournis, tout en assurant :

- Une évaluation périodique sur un ensemble de validation.
- Un suivi des métriques en temps réel (*monitoring*).
- Une gestion automatique des checkpoints.

- Une sélection du meilleur modèle basé sur les performances de validation.

Le Tableau 4.5 présente les entrées et les sorties de ce module.

Élément	Description
Entrées	<ul style="list-style-type: none"> – Jeux d’entraînement et de validation – Modèle GNN – Configuration des hyperparamètres d’entraînement
Sorties	<ul style="list-style-type: none"> – Modèle entraîné – Tableaux récapitulatifs des performances – Courbes d’évolution

Tableau 4.5 : Entrées et sorties du module d’entraînement, validation et de suivi.

Déroulement et description interne

Le module d’entraînement constitue le cœur de la boucle d’apprentissage supervisée. Il orchestre la préparation des données, la configuration des modèles GNN, l’optimisation et le suivi des performances au cours du temps.

Après un nettoyage de la mémoire GPU et le chargement de la configuration, le module charge les graphes produits par *AISDatasetGrid* directement en mémoire afin de réduire le temps d’accès. Le jeu complet est ensuite divisé de manière aléatoire, mais reproductible grâce à une graine fixée, en un ensemble d’entraînement et un ensemble de validation.

Les données sont chargées sous forme de mini-lots et prétraitées en parallèle de l’entraînement. Autrement dit, quand un mini-lot est utilisé pour entraîner le modèle GNN, le processeur génère en parallèle les sous-graphes associés au mini-lot suivant. L’entraînement est orchestré par un objet *Trainer* issue de la bibliothèque *Pytorch Lightning*, qui gère l’exécution sur multi-GPUs, le suivi des métriques en continu et la validation périodique. Un mécanisme de sauvegarde automatique conserve les meilleurs modèles en fonction de la perte de validation, garantissant que les expériences puissent être reprises ou comparées a posteriori. Le contenu du fichier de configuration est copié dans le répertoire de *logs* afin d’assurer la traçabilité complète des expériences. L’architecture finale du modèle est également sauvegardée pour documenter précisément la version du réseau entraîné.

4.3.2 Implémentation du module de test

Objectif fonctionnel

Ce module évalue les performances des modèles sur un jeu de test. Il fournit une mesure quantitative des performances, ainsi qu’une analyse qualitative des erreurs sous forme de cartes et diagrammes. Il permet également de comparer les différents paradigmes de modélisation sur un même jeu de données. Le Tableau 4.6 présente les entrées et les sorties de ce module.

Elément	Description
Entrées	<ul style="list-style-type: none"> – Jeu de test – Modèle entraîné – Configuration des hyperparamètres de test
Sorties	<ul style="list-style-type: none"> – Prédications et étiquettes – Rapport de performance – Matrice de confusion – Fichier CSV des erreurs – Cartes de densité géospatiale des erreurs – Distributions temporelles des erreurs

Tableau 4.6 : Entrées et sorties du module de test.

Déroulement et description interne

Ce module met en œuvre une évaluation reproductible des modèles sur un jeu de test indépendant. L’initialisation commence par la configuration déterministe du module prétraitement, qui fige l’indexation des nœuds émetteurs et garantit l’alignement strict entre échantillons, étiquettes et prédictions à chaque exécution.

L’architecture du module permet de comparer plusieurs paradigmes de modélisation dans un même cadre expérimental. Trois branches d’inférence coexistent et peuvent être activées par un sélecteur : des modèles physiques fondés sur les recommandations de l’UIT (implémentations ITU-R P.2001, ITU-R P.1812 et ITU-R P.452), un classifieur de type *gradient boosting* (XGBoost), et des GNN entraînés. Chaque branche produit un vecteur de prédictions, associé à un vecteur d’étiquettes. Le temps d’inférence global est mesuré au niveau du lot complet, fournissant un indicateur d’efficacité computationnelle complémentaire aux performances statistiques.

Les performances sont agrégées par une fonction qui calcule les métriques et la matrice de confusion. Dans le cas des modèles déterministes, un seuil explicite est appliqué pour convertir la sortie en décision binaire. Ce seuil correspond au critère opérationnel de réception et harmonise la comparaison entre paradigmes. Les métriques et la matrice de confusion sont sauvegardées sur disque dans un répertoire d’expérience dédié. Au-delà des métriques globales, le module peut être utilisé dans le cadre d’une analyse qualitative des prédictions. Il permet de générer des cartes de densité géospatiale des erreurs ou bien d’afficher la distribution temporelle de celles-ci.

4.3.3 Implémentation du module de prédiction

Objectif fonctionnel

Ce module applique un modèle entraîné (GNN, ML ou modèle semi-déterministe) à de nouvelles données spatio-temporelles pour générer des prédictions sur la réception de signal dans l'espace géographique étudié. Il prend en charge :

- La préparation des données de prédiction.
- L'exécution du modèle sélectionné en mode inférence.
- La sauvegarde des valeurs prédites.
- La visualisation cartographique des prédictions.

Le Tableau 4.7 présente les entrées et les sorties de ce module

Elément	Description
Entrées	<ul style="list-style-type: none">– Données pour la zone spatiale où effectuer les prédictions– Modèle entraîné– Configuration des hyperparamètres de prédiction
Sorties	<ul style="list-style-type: none">– Prédications au format CSV– Cartes de prédictions

Tableau 4.7 : Entrées et sorties du module de prédiction.

Déroulement et description interne

Ce module est consacré à l'application de modèles préalablement entraînés sur de nouvelles données, en vue de générer des prédictions sous forme de cartes de réception du signal AIS. Contrairement aux modules dédiés à l'entraînement et au test, son objectif principal n'est pas l'évaluation des performances globales, mais la production de sorties exploitables, à la fois sous forme de cartes et de fichiers, utilisables pour l'analyse opérationnelle et la visualisation géographique.

Son fonctionnement repose sur une organisation en deux couches. Dans un premier temps, une classe encapsule les graphes horaires construits à partir des données. Chaque nœud y est ensuite échantillonné afin de construire le sous-graphe qui lui est associé. Ces sous-graphes sont enregistrés en mémoire. Dans un second temps, le cœur du module applique le modèle sélectionné en mode inférence aux sous-graphes par mini-lots. Les prédictions obtenues sont sauvegardées au format CSV. Par ailleurs, une fonction de cartographie permet de générer une représentation spatiale des probabilités de réception prédites. Cette carte intègre un fond géographique (côtes, frontières, surfaces terrestres et maritimes), de manière à contextualiser la distribution spatiale des prédictions.

Ainsi, ce module constitue le point de jonction entre l'apprentissage supervisé et l'application opérationnelle. Il offre la possibilité d'examiner la cohérence spatiale des prédictions et de mettre en évidence les zones de réception et de non-réception.

4.3.4 Jeux de données et entraînement des modèles

Toutes les procédures d'entraînement et d'évaluation ont été réalisées sur un serveur de calcul haute performance équipé de deux GPU *NVIDIA Quadro RTX 8000* (2×46 Go de mémoire vidéo), d'un processeur *Intel Xeon Silver 4210R* (40 cœurs, fréquence de base de 2,40 GHz), de 376 Go de mémoire vive *DDR4* et de 11 To de stockage SSD avec une vitesse de lecture d'environ 7212 Mo/s. Chaque cycle complet d'entraînement du modèle GNN proposé a nécessité environ 100 heures, incluant la génération des graphes, la standardisation des caractéristiques et l'apprentissage avec arrêt anticipé. Les expériences d'inférence ont été menées sur une machine grand public équipée d'un GPU *NVIDIA RTX 3070* (8 Go), d'un processeur *Intel Core i7-10700K* (8 cœurs / 16 threads, 3,80 GHz) et de 32 Go de mémoire vive. Cette configuration a été utilisée pour évaluer la latence de prédiction et tester les performances dans des conditions plus représentatives d'un environnement de déploiement classique.

Le jeu de données couvre une période de quatre ans, allant de 2020 à 2023. La zone d'étude est géographiquement bornée par les limites de latitude et de longitude définies dans la Section 3.3.4. Les expérimentations ont été réalisées sur les navires d'AIS de classe A, notamment par leur puissance d'émission supérieure permettant une couverture spatiale étendue. Pour l'évaluation, le jeu de données a été partitionné en trois sous-ensembles disjoints : apprentissage (95 %), validation (5 %) et un jeu de test séparé. Ce dernier comprend 2602 échantillons, chacun étant extrait d'instances horaires distinctes et uniformément réparties sur l'ensemble du jeu de données. Cela correspond à 10 % des données avant la division entre apprentissage et validation. La distribution spatiale de l'ensemble de test n'a pas été artificiellement homogénéisée, elle reflète directement la distribution naturelle des données observée dans la zone d'étude. Cela garantit que les résultats de l'évaluation sont représentatifs des conditions opérationnelles réelles, plutôt que biaisés par un échantillonnage uniforme.

L'optimisation du modèle a été réalisée à l'aide de l'optimiseur Adam (Kingma et Ba 2015), avec une taille de lot fixée à 32, choisie afin de concilier stabilité de l'entraînement et efficacité computationnelle compte tenu des ressources de calcul disponibles, un arrêt anticipé basé sur les performances en validation, et une sauvegarde des poids correspondant aux trois plus faibles pertes de validation sont configurés. Le taux d'apprentissage a été fixé à 1×10^{-3} . Afin d'atténuer le déséquilibre de classes au sein de chaque grille horaire, chaque échantillon a été tiré avec une probabilité uniforme entre les classes présentes dans la grille correspondante. La configuration complète de l'apprentissage est résumée dans le Tableau 4.10. Il optimise la fonction de perte d'entropie croisée binaire :

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i))$$

Où $\hat{y}_i \in (0,1)$ la probabilité prédite pour l'échantillon i , et $y_i \in \{0,1\}$ l'étiquette binaire.

Pour traiter le déséquilibre régional dans la couverture AIS, notamment dans les zones présentant une faible densité de réception en raison de l'absence de récepteurs AISHub, une stratégie de génération de données visant à assurer une couverture spatiale plus uniforme a donc été mise en œuvre. Plus précisément, il a été attribué de manière artificielle l'étiquette 0 (non-réception) à un petit nombre de cellules de grille choisies aléatoirement dans les zones sous-représentées (Figure 4.3). Cette approche vise à limiter le surapprentissage sur les zones fortement couvertes et à améliorer la généralisation aux régions moins surveillées. Cette stratégie est conçue comme un cadre général pour parcourir systématiquement le domaine spatial par blocs de taille fixe ($0,25^\circ \times 4$). À l'intérieur de chaque bloc, si la réception est déjà représentée (étiquette 1) ou si la non-réception est déjà représentée (étiquette 0), le bloc est laissé tel quel. Ce n'est que dans le cas où toutes les cellules sont non étiquetées (étiquette -1) que l'algorithme assigne aléatoirement une cellule comme non-réception (0). De cette manière, la méthode garantit au moins un point de données par bloc sans gonfler artificiellement les régions denses, réduisant ainsi le risque de surapprentissage dans les zones fortement couvertes et améliorant la généralisation aux régions faiblement couvertes.

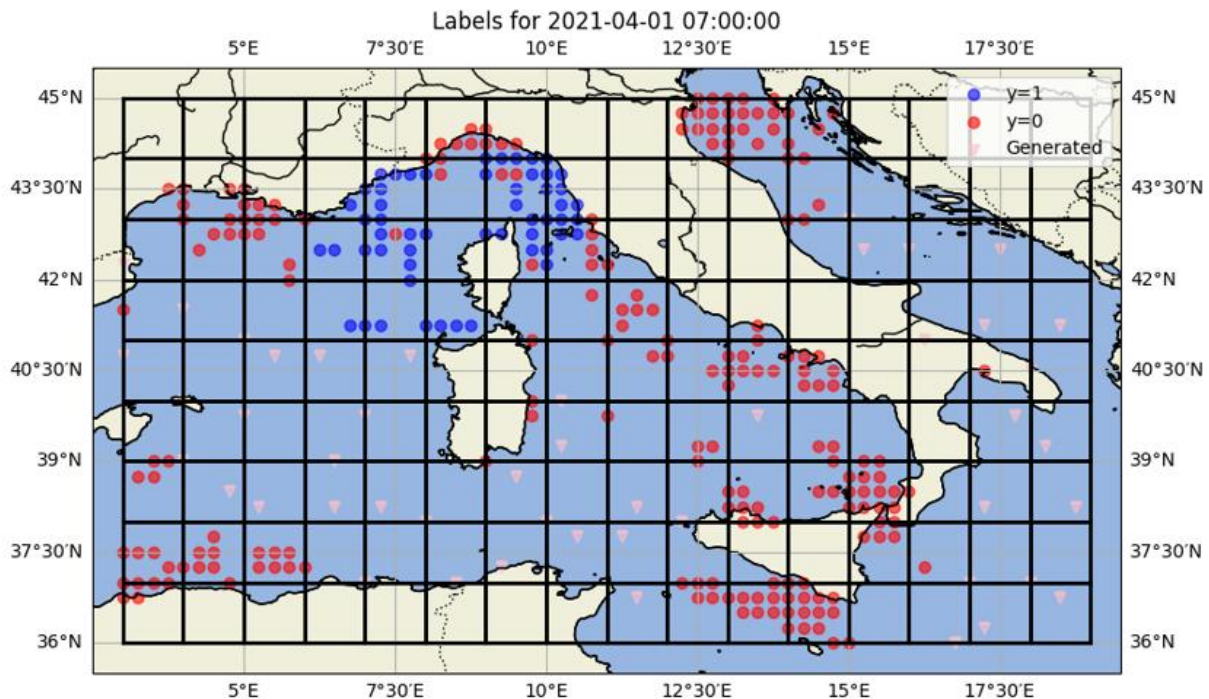


Figure 4.3 : Grille spatiale étiquetée montrant les étiquettes 0 générées artificiellement. Une grille de cellules de $0,25^\circ \times 4$ est surimprimée, assurant que chaque cellule contient au moins un point de données.

La procédure de génération parcourt le domaine spatial complet à l'aide de blocs de taille fixe et examine la distribution des étiquettes dans chaque bloc. Si un bloc ne contient que des cellules non étiquetées (-1), une cellule est sélectionnée aléatoirement et étiquetée comme 0. La procédure est détaillée dans l'Algorithme 4.1.

Données : D (coordonnées, étiquettes y), bornes (LAT_{\min}, LAT_{\max}), (LON_{\min}, LON_{\max}), résolution r , taille de bloc N

Résultat : D modifié

pour tous les blocs B de D de taille $N \times N$ faire

- ┌ Extraire les étiquettes ℓ_B de toutes les cellules de B
- ┌ **si** $1 \in \ell_B$ **et** ($0 \in \ell_B$ **ou** $-1 \in \ell_B$) **alors**
- ┌ $\quad\quad\quad$ // La réception est déjà modélisée
- ┌ **continuer**
- ┌ **sinon si** $0 \in \ell_B$ **alors**
- ┌ $\quad\quad\quad$ // La non-réception est déjà modélisée
- ┌ **continuer**
- ┌ **sinon si toutes les valeurs de ℓ_B valent -1 alors**
- ┌ Sélectionner aléatoirement un index $i \in B$ et poser $y_i \leftarrow 0$

retourner D modifié

Algorithme 4.1 : Génération d'étiquettes 0 pour l'équilibrage spatial.

L'évolution des pertes d'entraînement pour les trois architectures est présentée à la Figure 4.4. Le modèle GCN montre une décroissance rapide au début de l'optimisation, mais converge vers un plateau relativement élevé. Le GAT, bien qu'expressif, présente une dynamique plus irrégulière et une variance plus marquée, traduisant une stabilité d'entraînement moindre. En revanche, le modèle proposé, basé sur GraphSAGE avec agrégations multiples et mécanisme *Jumping Knowledge*, atteint systématiquement des valeurs de perte inférieures à celles obtenues par GCN et GAT, et ce, dès les premières itérations.



Figure 4.4 : Évolution de la perte d’entraînement en fonction du nombre d’itérations pour les modèles GCN, GAT et pour le GNN proposé (GraphSAGE avec *Jumping Knowledge* et agrégations multiples). Courbes lissées par moyenne mobile exponentielle pondérée par le temps.

Modèle	Nombre de paramètres entraînaibles
GCN	1253072
GAT	1254332
GNN Proposé	1579917

Tableau 4.8 : Nombre de paramètres entraînaibles par modèles.

L’analyse du nombre de paramètres (Tableau 4.8) montre que le modèle proposé, atteint environ 1,58 million de paramètres, contre 1,25 million pour GAT et GCN. Ce résultat peut sembler contre-intuitif dans la mesure où GraphSAGE, dans sa version de base, est généralement considéré comme moins paramétrique que des architectures attentionnelles. Toutefois, l’ajout des agrégations multiples, qui concatènent différentes représentations issues de fonctions d’agrégation distinctes, ainsi que l’intégration du mécanisme *Jumping Knowledge*, qui combine les sorties de plusieurs couches, entraînent une augmentation notable de la dimension des représentations intermédiaires et donc du nombre total de paramètres.

Néanmoins, cette complexité accrue ne relève pas d’un simple surdimensionnement du modèle, mais bien d’un enrichissement de la structuration des représentations apprises. Les résultats expérimentaux indiquent que le modèle proposé ne se contente pas d’exploiter un plus grand nombre de paramètres : il présente également une convergence plus stable et atteint des pertes d’entraînement inférieures aux autres architectures. Cela suggère que les gains observés proviennent principalement de la capacité du modèle à combiner efficacement des informations multi-niveaux et multi-facettes au sein du graphe, plutôt que d’une augmentation brute de la capacité paramétrique.

Avant d’analyser les performances du modèle proposé, il est nécessaire de présenter brièvement le jeu de test sur lequel l’évaluation a été conduite. Celui-ci se compose de 2602 échantillons

indépendants, répartis temporellement afin de refléter la diversité des conditions de propagation rencontrées dans le bassin méditerranéen.

Sur le plan spatial, la Figure 4.5 illustre la répartition des points de données annotés respectivement avec l'étiquette réception (1, gauche) et non-réception (0, droite). Une concentration notable des réceptions dans le nord du bassin, en particulier en mer Ligure est observée, tandis que les cas de non-réception sont davantage dispersés et couvrent une zone géographique plus étendue, notamment autour de la Sicile et du sud de l'Italie. Elle présente la répartition agrégée des échantillons avec les pourcentages relatifs par maille spatiale ($1^\circ \times 1^\circ$). Cette représentation met en évidence une forte hétérogénéité géographique, avec des zones densément peuplées en observations contrastant avec de larges régions à faible densité. D'un point de vue temporel, la Figure 4.6 donne la distribution mensuelle des échantillons. Les données couvrent une période de juin 2020 à novembre 2023, avec une répartition relativement homogène dans le temps. Cette diversité temporelle garantit que le modèle est évalué dans des conditions saisonnières variées.

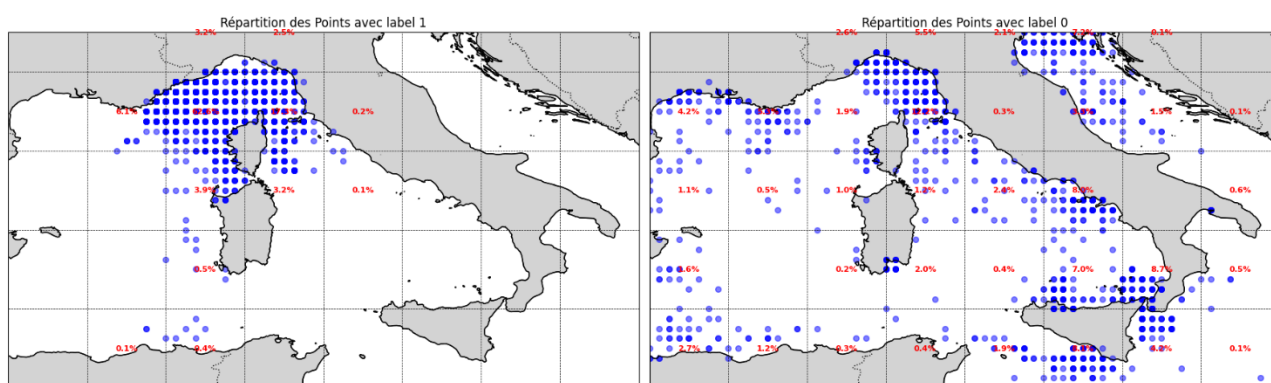


Figure 4.5 : Répartition spatiale des points de données de test pour les étiquettes 1 (gauche) et 0 (droite).

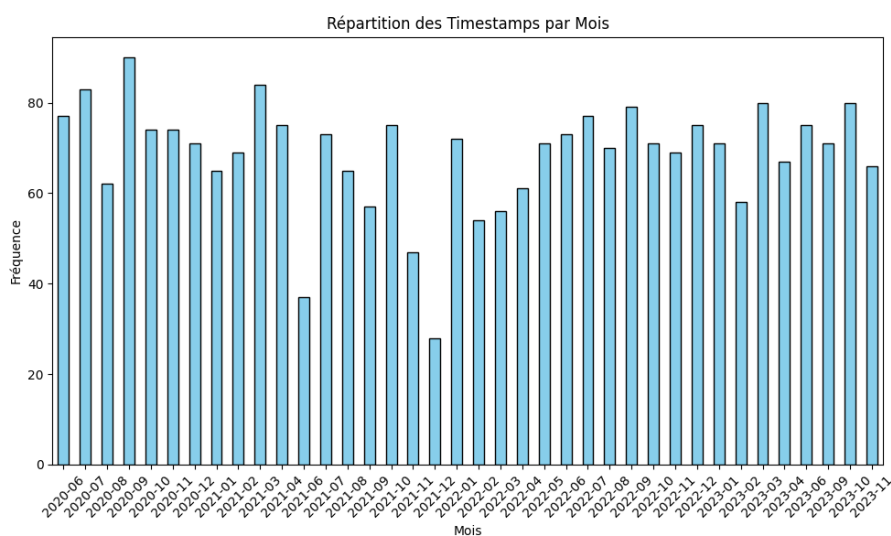


Figure 4.6 : Répartition temporelle des points de données de test.

4.3.5 Optimisation des hyperparamètres

La sélection des hyperparamètres a été réalisée à l'aide d'une recherche en grille sur l'ensemble de validation. Les paramètres explorés incluaient le nombre de dimensions cachées, le nombre de couches GNN et le taux de *dropout*. En pratique, le choix des hyperparamètres finaux a été guidé par de courtes simulations sur quelques *epochs*, au cours desquelles la convergence, la vitesse de convergence et l'exactitude sur le jeu de validation ont été systématiquement suivis. Cette procédure a permis d'identifier la configuration offrant le meilleur compromis entre stabilité et performance prédictive. Le Tableau 4.9 résume l'espace de recherche des hyperparamètres ainsi que les valeurs finales retenues pour le modèle.

Hyperparamètre	Plage testée	Valeur finale
Taux d'apprentissage	$\{1 \times 10^{-2}, 1 \times 10^{-3}, 1 \times 10^{-4}, 1 \times 10^{-5}\}$	1×10^{-3}
Dimensions cachées	{64, 128, 256, 512}	128
Nombre de couches GNN	{3,5,7}	5
Nombre de couches MLP	{2,4}	2
Ouverture du cône directionnel	{15, 30, 60, 90, 120}	30
Taille de bloc (génération d'étiquettes)	{ \emptyset , 3,4}	4

Tableau 4.9 : Espace de recherche des hyperparamètres exploré lors du réglage du modèle.

Composant	Configuration
Optimiseur	Adam
Taux d'apprentissage	1×10^{-3}
Taille du lot	32
Nombre de couches GNN	5
Nombre de couches MLP	2
Dimensions cachées	128
Fonction d'activation	ReLU
Dernière activation	Sigmoïde
Fonction de perte	Entropie croisée binaire
Ouverture du cône directionnel	30
Taille des blocs de génération d'étiquettes	4

Tableau 4.10 : Résumé de la configuration d'apprentissage et des hyperparamètres utilisés pour l'optimisation du modèle.

4.4 Evaluation des performances

L'évaluation des performances vise à quantifier l'efficacité du modèle proposé sur un jeu de test indépendant, en le comparant aux modèles de référence selon des critères standards. Pour cela, plusieurs métriques de classification sont mobilisées : exactitude, rappel, précision, F1-score.

Cette section présente d'abord les résultats numériques obtenus par chaque modèle, avant d'analyser en détail les performances du modèle proposé à travers une matrice de confusion et des visualisations géospatiales. Enfin, le modèle est comparé à XGBoost dans le cadre d'une évaluation par zones.

4.4.1 Présentation des métriques d'évaluation

Les métriques classiques suivantes pour la classification sont utilisées afin d'évaluer les modèles : exactitude (*Accuracy*), rappel (*Recall*), précision (*Precision*) et F1-score. En complément, le temps d'inférence sur le jeu de test est inclus comme indicateur de l'efficacité computationnelle. Ces métriques offrent des perspectives complémentaires : l'exactitude mesure la justesse globale des prédictions, le rappel met l'accent sur la capacité du modèle à détecter les réceptions effectives, tandis que la précision évalue la tolérance aux fausses alertes. Les résultats reportés correspondent à des moyennes pondérées.

4.4.2 Prédiction et visualisation

Pour analyser le comportement du modèle GNN proposé, la matrice de confusion est dressée dans le Tableau 4.11. Sur les 2602 échantillons du jeu de test, le modèle identifie correctement 1484 cas de réception (vrais positifs) et 1006 cas de non-réception (vrais négatifs). Seuls 25 cas de réception sont manqués (faux négatifs), et 87 cas sont incorrectement prédits comme des réceptions (faux positifs). Cela correspond à un taux de vrais positifs élevé et un taux de faux positifs relativement faible, en cohérence avec le F1-score du modèle.

Réel	Prédit	0 (Non-réception)	1 (Réception)
0 (Non-réception)		1006 [VN]	87 [FP]
1 (Réception)		25 [FN]	1484 [VP]

Tableau 4.11 : Matrice de confusion pour le modèle GNN proposé sur le jeu de test. VN : Vrai Négatif, FP : Faux Positif, FN : Faux Négatif, VP : Vrai Positif.

La Figure 4.7 illustre des prédictions typiques réalisées par le modèle proposé pour quatre scénarios de propagation contrastés. Les cartes de droite représentent les étiquettes réelles, tandis que celles de gauche montrent les prédictions du modèle. Les cartes (a) et (b) correspondent à des cas de réception à courte portée, où l'empreinte spatiale du signal est compacte et proche de la station de base. En revanche, les cartes (c) et (d) illustrent des scénarios à longue portée, dans lesquels le modèle

identifie une zone de réception étendue, cohérente avec des effets atmosphériques tels que la propagation par conduits.

Pour générer ces cartes de prédiction, le modèle est appliqué à tous les graphes construits à partir de la grille spatiale sous-jacente. Chaque graphe correspond à une cellule candidate de transmission, et l'étiquette prédite est utilisée pour reconstruire la zone de réception spatiale en agrégeant les sorties sur l'ensemble de la grille.

En complément, la Figure 4.8 compare les prédictions des modèles GAT et XGBoost avec celles du modèle proposé, pour les scénarios **(a)** et **(c)** :

– Cas **(a)** : propagation à courte portée (29 septembre 2020, 23h00)

Dans ce scénario, le modèle proposé restitue fidèlement la zone de réception compacte observée au nord de la Corse, en limitant les erreurs de surextension. Le modèle GAT identifie également la zone principale de réception, mais ses prédictions apparaissent plus diffuses, traduisant une moindre capacité à localiser finement la couverture. Quant à XGBoost, il fait apparaître des artefacts ponctuels, probablement dus à la représentation de ses entrées sous forme de vecteurs.

– Cas **(c)** : propagation à longue portée (17 juillet 2022, 14h00)

Ce scénario illustre un cas de propagation étendue, probablement lié à des phénomènes atmosphériques. Le modèle proposé capture correctement la zone continue de réception s'étendant vers le sud, avec une structure spatiale cohérente et réaliste. GAT reproduit partiellement ce comportement avec une zone plus fine au sud mais surestime au nord-ouest où les prédictions de réceptions s'étendent jusque dans les terres françaises. XGBoost présente une zone de réception très cohérente avec les étiquettes mais inhomogène faisant apparaître des discontinuités et des prédictions isolées, réduisant la lisibilité globale, ce comportement peut être caractéristique d'un surapprentissage.

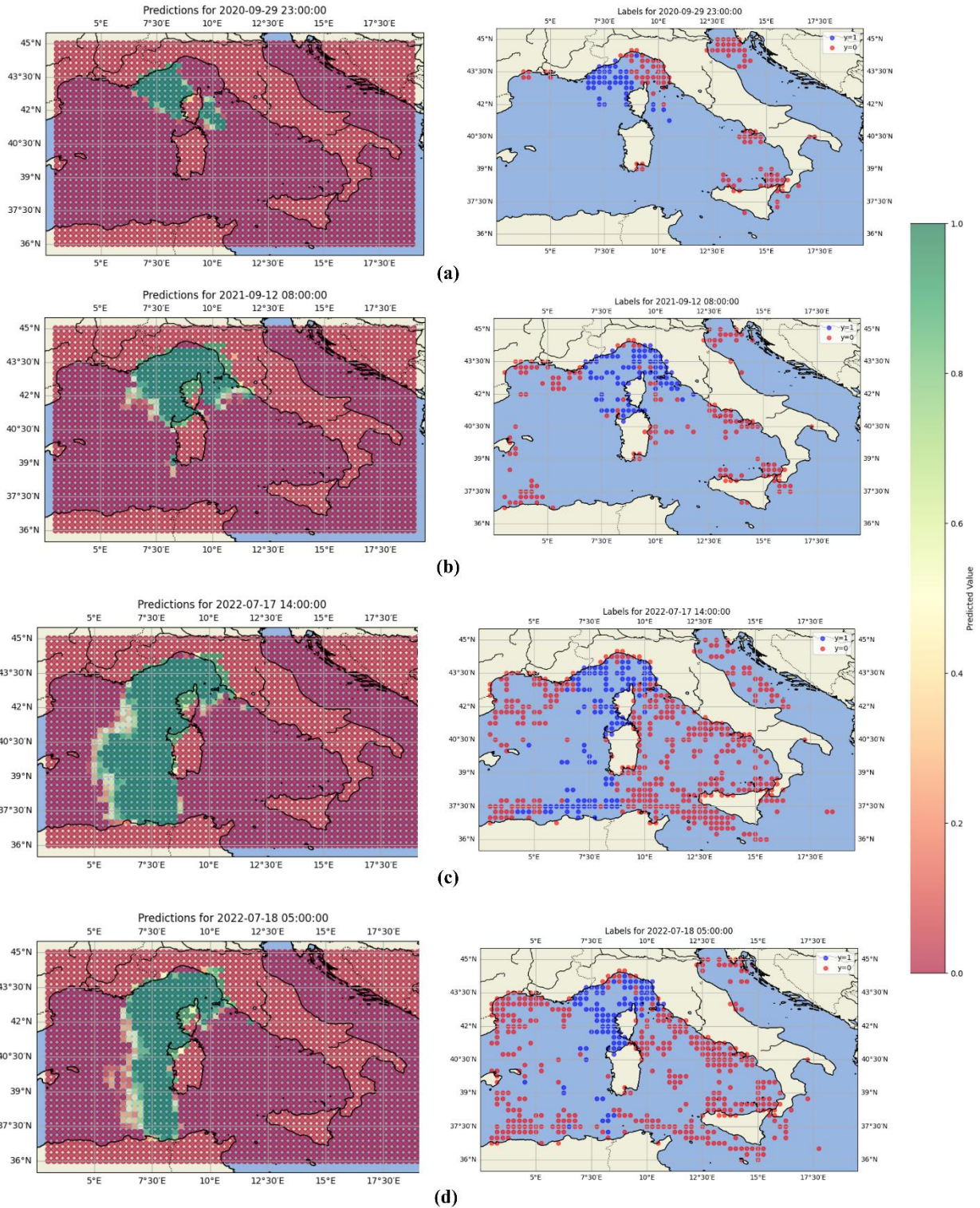


Figure 4.7 : Comparaison visuelle des prédictions du modèle pour différents scénarios de propagation. (a) correspond à une propagation à courte portée, avec certaines réceptions observées à l'arrière de la Corse. (b) illustre des cas de propagation avec réceptions situées plus à l'est, près des côtes italiennes. (c) et (d) correspondent à des scénarios de propagation à longue portée, les signaux émanant des côtes africaines, avec un faisceau plus étroit dans le cas de (d).

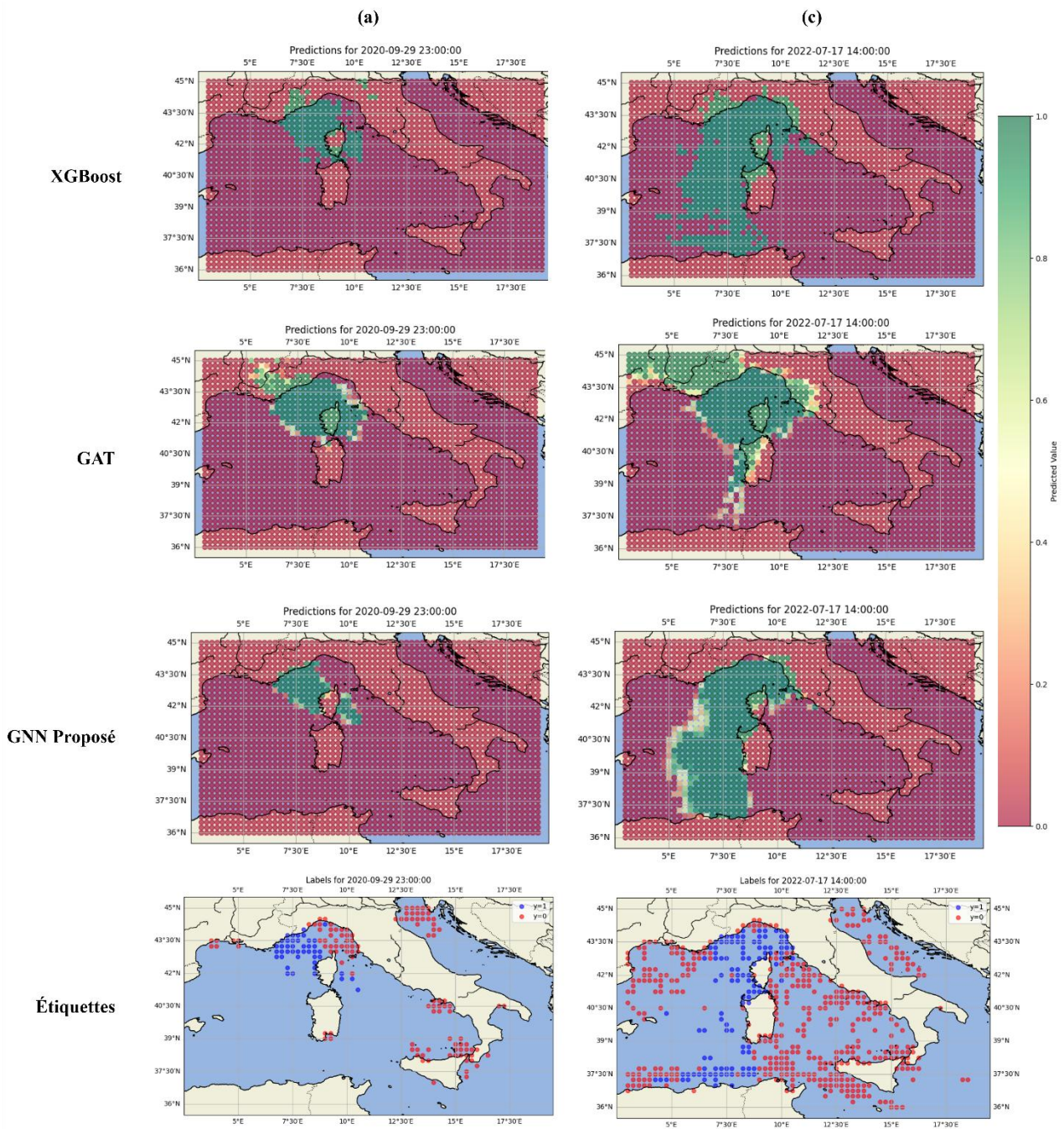


Figure 4.8 : Comparaison des prédictions de XGBoost, GAT et du GNN proposé sur les cas (a) et (c).

4.4.3 Résultats expérimentaux et comparaison avec les autres modèles

Le Tableau 4.12 résume les performances de l'ensemble des modèles évalués sur le jeu de test. Le modèle GNN proposé obtient les meilleurs résultats sur toutes les métriques d'évaluation : exactitude, rappel, précision et F1-score. Avec un F1-score de 0,957, il montre une forte capacité à classer correctement les zones de réception et de non-réception, surpassant le second meilleur modèle (GAT) de plus de 2,5 points de pourcentage. En termes de temps d'inférence, le modèle GCN est légèrement le plus rapide (54,73 s), suivi de près par GAT et XGBoost. Le modèle proposé effectue l'inférence sur les 2602 échantillons du jeu de test en 57,73 secondes, ce qui reste compétitif et adapté à un usage opérationnel.

Ces résultats confirment que l'approche fondée sur les graphes offre un compromis favorable entre performance prédictive et efficacité computationnelle, ce qui la rend pertinente pour des applications en temps réel ou quasi-réel. En pratique, sur une machine grand public (décrite en Section 4.3.4), le modèle a permis de prédire l'ensemble d'une zone de réception AIS (composée de 2405 classifications de graphes) sous une condition environnementale donnée, en moins de 10 secondes.

Modèle	Exactitude	Rappel	Précision	F1-Score	Temps de test (s)
ITU-R P.2001	0,826	0,826	0,839	0,827	60,45
XGBoost	0,916	0,916	0,916	0,916	56,93
GCN	0,918	0,918	0,920	0,918	54,73
GAT	0,932	0,932	0,932	0,932	55,71
GNN Proposé	0,957	0,957	0,958	0,957	57,73

Tableau 4.12 : Performances globales des modèles comparés sur le jeu de test (2602 échantillons).

4.4.4 Evaluation par zones

L'analyse précédente a montré que le modèle proposé surpasse les méthodes de référence lorsqu'il est évalué globalement sur l'ensemble du jeu de test. Cependant, une telle évaluation agrégée peut masquer d'importantes disparités locales liées à l'hétérogénéité géographique et temporelle des conditions de propagation. Pour mieux comprendre les points forts et les limites des modèles dans ces environnements contrastés, il est nécessaire de réaliser une évaluation par zones. L'objectif est d'identifier les régions où les modèles conservent une robustesse élevée, et celles où leurs performances se dégradent, en analysant plus finement la répartition spatiale et temporelle des erreurs.

À cette fin, une approche de *clustering* spatio-temporel a été développée, permettant de regrouper les cellules de la grille en ensembles homogènes selon leurs dynamiques de réception. L'étude des performances par *clusters* offre ainsi une vision plus nuancée du comportement des modèles, révélant des disparités régionales qui complètent utilement les métriques globales présentées précédemment.

4.4.4.1 Méthodologie de construction des zones

Afin de raffiner l'analyse des performances et d'identifier d'éventuelles disparités géographiques ou temporelles, une approche de *clustering* spatio-temporel a été mise en œuvre. Chaque cellule de la grille spatiale est représentée par un vecteur décrivant l'évolution de la présence binaire (0 : non-réception, 1 : réception) au cours des trois années couvertes par le jeu de données. À partir de cette représentation, une matrice de connectivité est construite en considérant le voisinage spatial de chaque cellule (8 directions cardinales et diagonales). Cette contrainte spatiale permet de regrouper des cellules contiguës dans un même *cluster*. Un algorithme de *clustering* hiérarchique avec liaison de Ward (Müllner 2011) est ensuite appliqué, en intégrant la matrice de connectivité. Cette méthode favorise la formation de *clusters* spatialement compacts et homogènes vis-à-vis des variations temporelles de présence.

La Figure 4.9 illustre la répartition spatiale des cinq *clusters* ainsi obtenus. Des zones cohérentes géographiquement se distinguent, couvrant par exemple la mer Ligure, le bassin Tyrrhénien ou encore la façade sud du bassin méditerranéen. Les *clusters* sont de populations hétérogènes : Cluster 0 (Bleu, 606 points de données du jeu de test), Cluster 1 (Vert, 385 points de données du jeu de test), Cluster 2 (Jaune, 332 points de données du jeu de test), Cluster 3 (Rouge, 542 points de données du jeu de test), Cluster 4 (Rose, 298 points de données du jeu de test) et Cluster 5 (Orange, 439 points de données du jeu de test). Cette distribution met en évidence des sous-régions maritimes densément ou faiblement représentées, ce qui est important pour interpréter les métriques de performance présentées ci-dessous.

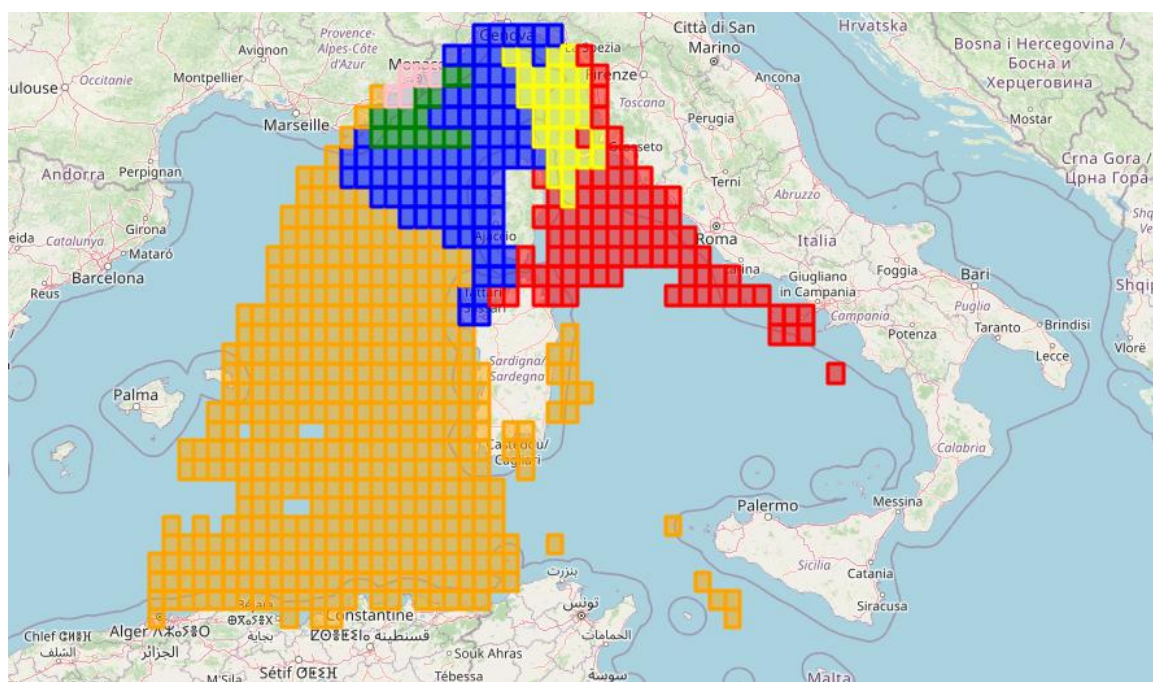


Figure 4.9 : Sous-régions maritimes spatialement cohérentes obtenues par un regroupement hiérarchique spatio-temporel des cellules de la grille ($k=6$).

4.4.4.2 Résultats par zones

Les métriques classiques de classification (précision, rappel, F1-score) ont été calculées séparément pour chaque *cluster* et pour chaque modèle évalué. Cette analyse permet de mettre en évidence les forces et faiblesses relatives des approches selon les zones.

Cluster	Modèle	Précision		Rappel		F1-score	
		0	1	0	1	0	1
Cluster 0 (Bleu)	GNN Proposé	0,853	0,950	0,777	0,970	0,813	0,960
	XGBoost	0,866	0,922	0,634	0,978	0,732	0,949
Cluster 1 (Vert)	GNN Proposé	1,000	1,000	1,000	1,000	1,000	1,000
	XGBoost	1,000	1,000	1,000	1,000	1,000	1,000
Cluster 2 (Jaune)	GNN Proposé	0,839	0,904	0,839	0,904	0,839	0,904
	XGBoost	0,845	0,838	0,702	0,923	0,767	0,879
Cluster 3 (Rouge)	GNN Proposé	0,989	0,871	0,974	0,942	0,981	0,905
	XGBoost	0,978	0,792	0,956	0,884	0,967	0,835
Cluster 4 (Rose)	GNN Proposé	1,000	1,000	1,000	1,000	1,000	1,000
	XGBoost	1,000	1,000	1,000	1,000	1,000	1,000
Cluster 5 (Orange)	GNN Proposé	0,968	0,879	0,990	0,691	0,979	0,773
	XGBoost	0,968	0,829	0,985	0,691	0,976	0,753

Tableau 4.13 : Performances par classe et par cluster sur l'ensemble de test (meilleures valeurs en gras).

Les résultats, comparés entre XGBoost et le modèle proposé sont résumés dans le Tableau 4.13 : Performances par classe et par cluster sur l'ensemble de test (meilleures valeurs en gras). Le choix d'opposer le modèle proposé à XGBoost, plutôt qu'aux autres GNN (GCN, GAT) ou encore au modèle ITU-R, repose sur plusieurs considérations. D'une part, XGBoost constitue une référence largement adoptée car son implémentation est quasi instantanée et sa diffusion particulièrement large dans la communauté scientifique et industrielle. D'autre part, XGBoost se distingue par des performances ici comparables à celles des modèles GCN et GAT. Enfin, XGBoost surpasse le modèle ITU-R. Cette comparaison à son sens dans la mesure où elle permet de mettre en exergue que les gains obtenus proviennent bien de l'exploitation de la structure relationnelle offerte par les graphes.

Les *clusters* les plus contrastés peuvent être mis en évidence comme suit :

- Dans le golfe de Gênes (Cluster 2 : Jaune), le GNN proposé a maintenu des performances équilibrées, avec une précision et un rappel de 0,904 pour la classe « réception » (1). En revanche, XGBoost présentait un déséquilibre : bien que son rappel pour la classe 1 ait été légèrement supérieur (0,923), il souffrait d'une détection médiocre de la classe 0 (rappel =

0,702), ce qui conduisait à des F1-scores globalement plus faibles. Cela indique que le modèle proposé fournit des prédictions plus stables entre les deux classes dans cette région.

- Dans le bassin Tyrrhénien (Cluster 3 : Rouge), les deux modèles ont atteint de hautes performances, mais le GNN proposé s’est démarqué sur la classe 1 avec un rappel de 0,942 et un F1-score de 0,905, contre 0,884 et 0,835 pour XGBoost. Cela confirme sa capacité de détection renforcée pour la classe « réception » dans cette zone à forte densité d’observations.
- Dans toutes les autres régions, les deux modèles ont montré des performances comparables, avec des scores parfaits dans les *clusters* 1 et 4, et seulement des différences marginales ailleurs.

Cette analyse par *clusters* révèle que les performances globales (Tableau 4.12) masquent des disparités locales. Dans ces deux *clusters* le modèle proposé démontre une meilleure stabilité que XGBoost, en réduisant le compromis entre détection de réceptions et limitation des fausses alertes. En affichant les faux positifs et faux négatifs dans les zones correspondantes (Figure 4.10 et Figure 4.11), celles-ci coïncident avec les régions où les deux modèles présentent les performances les plus faibles, comme déjà mis en évidence par l’analyse par *clusters* (notamment pour les *clusters* 2 et 3). Toutefois, bien que les deux modèles soient affectés, le modèle proposé conserve un avantage : il limite davantage l’étendue spatiale des faux positifs et réduit le nombre de faux négatifs, traduisant une meilleure capacité à distinguer les réceptions effectives.

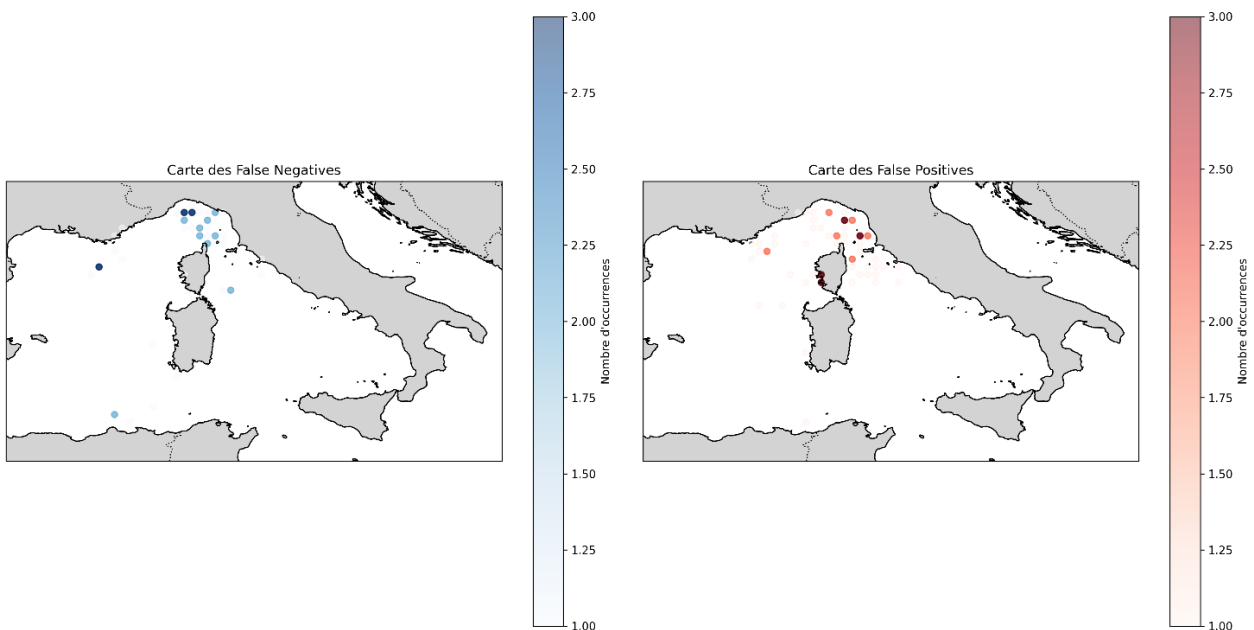


Figure 4.10 : Cartes des faux positifs et des faux négatifs pour le modèle proposé.

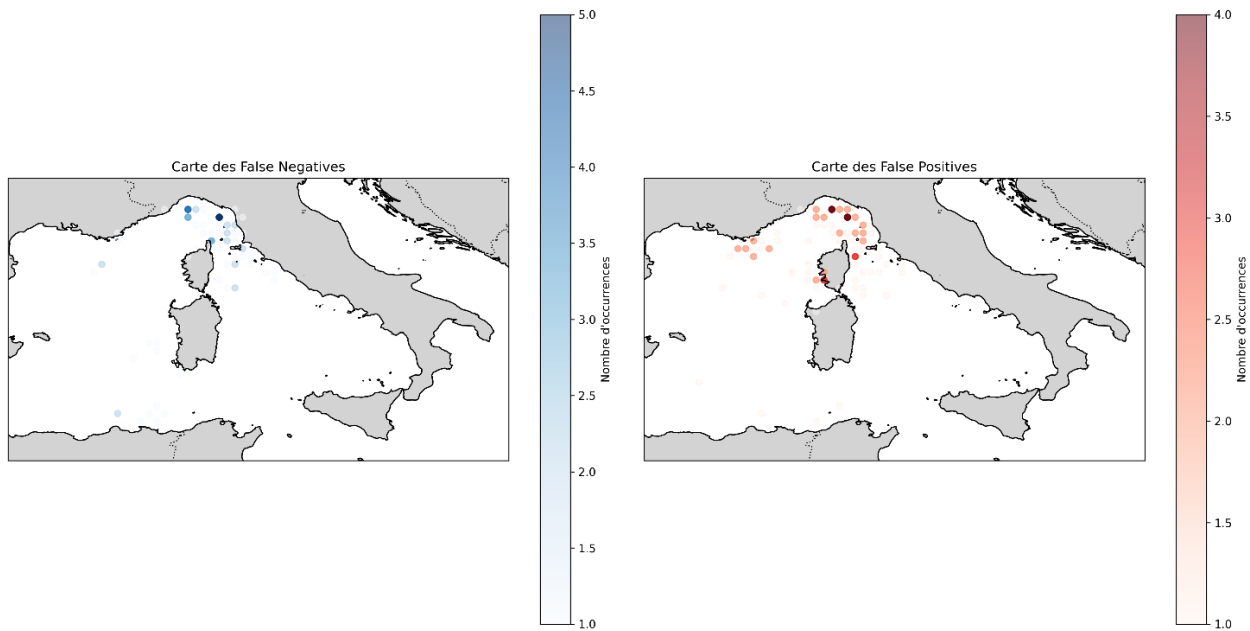


Figure 4.11 : Cartes des faux positifs et des faux négatifs pour XGBoost.

4.5 Limites et discussion

L'analyse des résultats expérimentaux met en lumière les points forts du modèle proposé, mais aussi un certain nombre de limitations méthodologiques, structurelles et pratiques. Cette section propose une discussion critique sur les hypothèses formulées lors de la conception du modèle, les simplifications introduites dans la représentation des phénomènes de propagation, ainsi que les contraintes inhérentes aux données disponibles.

Cette réflexion ouvre la voie à des pistes d'amélioration, tant sur le plan de la modélisation que sur celui des données mobilisées.

4.5.1 Limites méthodologiques

Une première limite provient des hypothèses formulées lors du prétraitement des données et de la construction des graphes. Le modèle repose sur une grille spatiale fixe de résolution $0,25^\circ \times 0,25^\circ$, qui peut s'avérer trop grossière pour capturer les variations fines du relief ou des propriétés atmosphériques influençant la propagation du signal. De plus, le cône directionnel utilisé pour restreindre le voisinage des graphes constitue un hyperparamètre, qui ne reflète pas nécessairement toute la complexité des dynamiques de propagation réelles. Bien que ce filtrage améliore les performances computationnelles et s'aligne sur certaines heuristiques physiques, il introduit une frontière artificielle susceptible d'exclure des interactions spatiales pertinentes.

4.5.2 Limites liées à la modélisation

Du point de vue de la modélisation, l'architecture proposée permet de capturer la structure spatiale, mais ne prend pas explicitement en compte les dynamiques temporelles. Les graphes sont en

effet construits de manière indépendante pour chaque créneau horaire, sans considération pour la continuité ni pour l'évolution des conditions environnementales au fil du temps. Cette caractéristique peut limiter la capacité du modèle à identifier des motifs de propagation persistants. Par ailleurs, si l'agrégation multiple accroît le pouvoir expressif de l'architecture, elle induit également une augmentation du nombre de paramètres à optimiser, avec les implications associées en termes de complexité d'entraînement et de risque de surapprentissage.

4.5.3 Limites liées à l'évaluation

L'évaluation a été conduite sur un jeu de données géographiquement restreint. Bien que l'approche adoptée repose sur un apprentissage inductif de représentations sur graphes, la capacité du modèle à généraliser vers des régions maritimes totalement absentes de l'entraînement demeure à confirmer. Certains modèles de référence, tels que l'ITU-R P.2001, présentent par ailleurs une dépendance marquée à la qualité des données mobilisées, lesquelles ne sont pas toujours disponibles ni homogènes en termes de précision selon les zones étudiées.

4.5.4 Contraintes liées aux ressources

Plusieurs contraintes pratiques ont restreint le périmètre de l'étude. En premier lieu, la collecte et l'alignement des données environnementales, en particulier celles issues des réanalyses météorologiques, nécessitent un volume de stockage important ainsi qu'un temps de prétraitement conséquent. En second lieu, la construction de graphes à haute résolution, de même que l'entraînement des modèles de type GNN, s'avèrent coûteux en ressources de calcul, ce qui limite le nombre de variantes architecturales et de répétitions expérimentales envisageables.

Les données de réception AIS présentent un caractère intrinsèquement bruité : la variabilité du comportement de transmission des navires, les incohérences matérielles ou encore les défauts d'alignement des antennes peuvent entraîner des pertes de signal indépendantes des conditions réelles de propagation, bien que l'approche de sous-échantillonnage à 10 minutes puis l'agrégation sur une grille horaire lisse ces effets entre les différents émetteurs. En outre, les étiquettes utilisées comme vérité terrain pour l'apprentissage sont également dérivées des réceptions AIS. Dans les régions caractérisées par un trafic maritime limité ou par une couverture réduite du réseau AISHub, l'absence de réception peut de la même manière résulter d'un déficit d'émetteurs plutôt que d'une atténuation effective du signal. Bien que la procédure de génération d'étiquettes proposé mitige ces effets à l'entraînement, cette ambiguïté peut néanmoins réduire la fiabilité de l'étiquetage et limiter la qualité des exemples de non-réception utilisés. En effet, davantage de vraies étiquettes 0, en particulier dans des zones où le trafic est structurellement absent (par exemple des zones réglementées ou interdites à la navigation), permettraient de mieux caractériser les situations de non-réception réelles et d'améliorer la robustesse du modèle.

4.5.5 Perspectives de modélisation et transfert de tâche

Afin d'évaluer qualitativement la transférabilité du modèle proposé vers une nouvelle tâche et un autre dispositif expérimental, une campagne d'acquisition complémentaire a été menée sur deux mois de données (juillet et août 2025). La méthodologie de prétraitement et de préparation des données reste strictement identique à celle décrite précédemment, mais sans recourir aux données AISHub. La sensibilité de la mesure diffère toutefois de la campagne initiale, car le dispositif a été modifié :

- Le boîtier Comar a été remplacé par un Raspberry Pi.
- Un filtre passe-bande a été introduit.
- La réception s'appuie sur une clé RTL-SDR.

Ce changement de sensibilité constitue une opportunité pour tester la robustesse et l'adaptabilité du modèle proposé à travers une approche de transfert d'apprentissage. Deux stratégies ont été comparées :

- Classification multi-classe : entraînement d'un classifieur sur les deux mois de données RSSI, où les valeurs de RSSI sont discrétisées en catégories de niveau de signal.
- Classification multi-classe avec transfert d'apprentissage (TL) : réutilisation du modèle pré-entraîné, puis spécialisation (*fine-tuning*) sur la tâche de classification multi-classe.

La prédiction exacte du RSSI n'est pas toujours nécessaire dans une perspective opérationnelle. Pour faciliter l'exploitation des cartes de réception, il peut être suffisant de disposer d'une indication qualitative de la qualité de réception. Cette reformulation en classification permet d'obtenir une sortie directement interprétable en termes de performance radio, tout en réduisant la sensibilité du modèle aux fluctuations locales des mesures continues. De plus elle permet de directement optimiser la fonction de perte d'entropie croisée multi-classes :

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^N \log(p_{i,y_i})$$

Où $p_{i,c} = \frac{\exp(x_{i,c})}{\sum_{k=1}^C \exp(x_{i,k})}$ la probabilité prédite pour la classe c par une fonction softmax avec $x_i \in \mathbb{R}^C$ les logits du modèle pour l'échantillon i , C le nombre de classes et $y_i \in \{1, \dots, C\}$ la vraie classe de l'échantillon i .

A partir des données continues, le RSSI est discrétisé en quatre intervalles de puissance, définis sur la base de seuils déterminés statistiquement (quantiles) sur l'ensemble des données disponibles :

- Faible (*Weak*, -50 à -30 dBm) : signal faible, proche du seuil de sensibilité.

- Modéré (*Moderate*, -30 à -23 dBm) : signal intermédiaire.
- Bon (*Fair*, -23 à -13 dBm) : signal satisfaisant.
- Excellent (≥ -13 dBm) : signal fort, garantissant une réception optimale.

La Figure 4.12 présente la comparaison des performances en termes d'exactitude sur l'ensemble de validation. Elle démontre que l'approche avec transfert d'apprentissage converge plus rapidement et atteint un niveau d'erreur inférieur à l'entraînement direct, confirmant l'intérêt de capitaliser sur les représentations déjà apprises. Enfin, bien que la sensibilité du dispositif de réception ait changé (et donc la portée effective), deux prédictions ont été réalisées par le modèle entraîné par TL sur deux scénarios représentatifs de la propagation, à savoir les cas **(a)** et **(c)** déjà introduits dans la Section 4.4.2. Ces cartes sont présentées en Figure 4.13. Ces résultats qualitatifs illustrent la capacité du modèle transféré à capturer correctement la morphologie spatiale des zones de réception, même dans des conditions matérielles différentes.

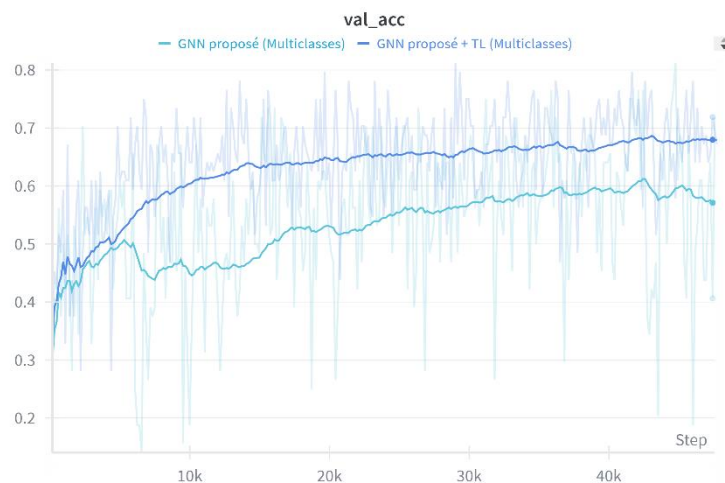


Figure 4.12 : Courbes d'exactitude en fonction du nombre d'itérations d'apprentissage, sur le jeu de validation. Courbes lissées par moyenne mobile exponentielle pondérée par le temps.

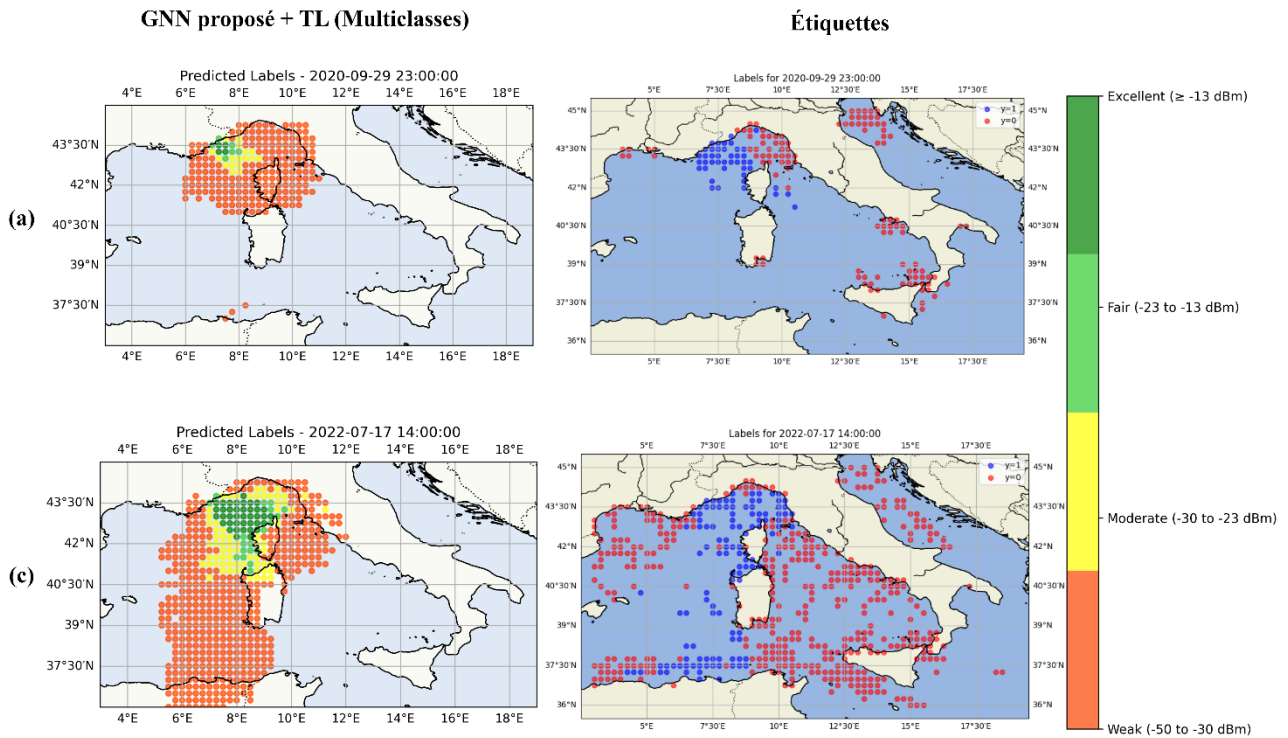


Figure 4.13 : Prédictions pour le modèle proposé dans une tâche de classification multi-classe et entraîné par apprentissage par transfert, pour les cas (a) et (c).

4.6 Conclusion

Ce chapitre a présenté une évaluation de l’approche proposée pour la modélisation de la réception de l’AIS, en la confrontant à des modèles de référence issus de paradigmes différents. Les résultats obtenus montrent que l’architecture neuronale fondée sur les graphes atteint des performances supérieures sur l’ensemble des métriques de classification, tout en maintenant une latence de prédiction adaptée à des applications en temps réel ou quasi-réel. La capacité du modèle à capter les relations spatiales locales, renforcée par l’agrégation hiérarchique et le mécanisme *Jumping Knowledge*, se traduit par une détection plus précise et plus robuste des zones de réception et de non-réception.

L’analyse qualitative des prédictions illustre que le modèle proposé restitue fidèlement les structures spatiales observées, qu’il s’agisse de scénarios de courte portée ou de phénomènes de propagation étendue liés à des conditions atmosphériques particulières. Comparativement, les modèles de type GAT et XGBoost montrent des performances intéressantes mais plus limitées : le premier souffre d’une tendance à la diffusion spatiale, tandis que le second présente des discontinuités et des artefacts ponctuels liés à sa représentation tabulaire. Les évaluations par zones, basées sur un *clustering* spatio-temporel, confirment la supériorité du modèle proposé en termes de robustesse régionale, avec une réduction notable des faux positifs et des faux négatifs dans les régions les plus complexes.

Cette évaluation, met en évidence la pertinence du cadre méthodologique proposé. Toutefois, plusieurs limites subsistent, notamment la granularité spatiale fixe des graphes, l'absence de modélisation explicite des dynamiques temporelles, la dépendance à des données bruitées et le coût computationnel associé aux graphes.

En complément de cette analyse, une campagne d'acquisition indépendante sur deux mois de données RSSI a permis d'ouvrir des perspectives sur la transférabilité des représentations apprises vers un autre contexte de mesure. Cette campagne s'appuie sur un dispositif matériel distinct, caractérisé par une sensibilité différente et donc une portée effective réduite. Dans ce contexte, l'apprentissage par transfert s'est révélé particulièrement efficace : la réutilisation des représentations apprises sur les données AIS a permis d'accélérer la convergence et d'augmenter l'exactitude par rapport à un entraînement direct. Les prédictions réalisées sur des scénarios représentatifs confirment la capacité du modèle transféré à reproduire fidèlement la morphologie spatiale des zones de réception, malgré le changement d'instrumentation et de conditions de mesure.

Ces constats ouvrent des perspectives claires pour la suite des travaux. L'intégration d'informations temporelles, l'adaptation du modèle à d'autres zones maritimes constituent autant de pistes d'amélioration. Ainsi, le présent chapitre valide expérimentalement la démarche proposée et souligne à la fois son potentiel opérationnel et les axes de recherche à poursuivre, qui seront discutés de manière plus générale dans le chapitre suivant.

Conclusion

Bilan et conclusion générale

Cette thèse a pour objectif de développer une méthodologie d'apprentissage profond fondée sur les réseaux de neurones en graphes, afin de modéliser les zones de réception du système AIS en environnement maritime tout en tenant compte des conditions environnementales. L'enjeu central consistait à dépasser les limites des modèles traditionnels, physiques et empiriques, dont la rigidité et la faible capacité de généralisation rendent difficile la caractérisation dynamique de la propagation dans le contexte de l'AIS. L'intérêt de ce travail s'explique d'abord par la dimension opérationnelle du problème : la couverture effective de l'AIS conditionne la qualité de la surveillance maritime, la sûreté des espaces maritimes et la détection des comportements anormaux des navires. Or, cette couverture est soumise à de fortes fluctuations dues à des facteurs topographiques, atmosphériques et techniques, qu'aucun modèle physique ne permet de représenter de manière suffisamment souple et intégrée. D'où l'importance de concevoir une approche capable d'intégrer des données hétérogènes, allant des observations AIS locales aux bases environnementales globales (réanalyses ERA5, topographie SRTM), en passant par les caractéristiques des systèmes de réception. Dans ce cadre, les deux publications réalisées durant cette thèse ont joué un rôle structurant : la première, présentée dans (Vanche et al. 2022), propose une approche initiale fondée sur une seule modalité de données troposphériques et sur un modèle tabulaire ; la seconde, publiée dans (Renaud et al. 2025), étend ce travail en intégrant plusieurs modalités de données et en introduisant une modélisation par graphes. Ces deux contributions ont permis d'établir progressivement la pertinence et la robustesse de la démarche proposée.

C'est dans ce contexte que le recours aux GNN s'est imposé comme choix méthodologique. En effet, la propagation radio se déroule dans un espace caractérisé par des interactions locales complexes et des asymétries fortes liées à la topographie et aux conditions météorologiques. Les graphes offrent un cadre opportun pour formaliser ces relations, tandis que les GNN permettent de capturer simultanément les structures locales et globales, d'intégrer des modalités variées de données et de généraliser à de nouveaux contextes de propagation. L'architecture proposée, reposant sur GraphSAGE enrichi de mécanismes multi-agrégation et *Jumping Knowledge*, illustre cette ambition de traduire les contraintes physiques du phénomène en un cadre d'apprentissage inductif robuste.

Méthodes pour la prédiction des zones de réception de l'AIS

L'étude des approches classiques de modélisation de la propagation radio appliquées à l'AIS met en lumière trois grandes familles de méthodes, chacune présentant des avantages et des limites marqués. Les modèles déterministes, fondés sur la résolution d'équations physiques et

électromagnétiques, permettent d'obtenir des prédictions précises dans des environnements contrôlés, mais se révèlent coûteux en calcul et peu flexibles lorsqu'il s'agit de couvrir de vastes zones maritimes ou de prendre en compte la variabilité atmosphérique. À l'opposé, les modèles empiriques et statistiques, élaborés à partir de campagnes de mesures, se distinguent par leur simplicité d'application et leur faible besoin en paramètres, mais leur portée reste limitée à des contextes spécifiques et ils peinent à capturer la variabilité spatio-temporelle des phénomènes de propagation. Entre ces deux extrêmes, les modèles semi-empiriques combinent lois physiques et ajustements statistiques, intégrant de plus nombreux mécanismes, mais demeurent tributaires de données d'entrée peu adaptées à une exploitation opérationnelle dynamique.

Dans un second temps, des approches d'apprentissage automatique ont été explorées, mobilisant notamment des algorithmes tels que les arbres de décision ou les méthodes d'ensemble comme XGBoost. Celles-ci ont montré leur intérêt pour modéliser la réception AIS, en particulier lorsqu'elles intègrent des variables environnementales issues de modèles météorologiques ou de données de terrain. Toutefois, elles demeurent limitées par leur dépendance à l'ingénierie manuelle des variables caractéristiques, leur difficulté à représenter des interactions fortement non linéaires et leur faible aptitude à intégrer des sources de données multiples et hétérogènes. Ces limites ont conduit à l'exploration d'approches en apprentissage profond, et plus spécifiquement des réseaux de neurones en graphes (GNN). Ceux-ci offrent un cadre computationnel apte à représenter explicitement les relations spatiales et directionnelles qui gouvernent la propagation AIS en environnement maritime.

Conception d'une chaîne de traitement pour la modélisation en graphes et l'apprentissage par GNN

La conception de la méthodologie repose sur le développement d'une chaîne de traitement complète, destinée à transformer les observations brutes en représentations exploitables par des GNN. Un processus de construction de graphes est mis en place : chaque nœud correspond à une cellule géographique de la grille spatiale et porte un vecteur de caractéristiques environnementales et contextuelles, tandis que les arêtes orientées traduisent les relations spatiales et directionnelles de propagation du signal. Les graphes sont générés dynamiquement pour chaque pas de temps, afin de capturer la variabilité spatio-temporelle des conditions de réception. Ces graphes intègrent des données hétérogènes issues de multiples sources : les observations AIS, enrichies de variables météorologiques extraites d'ERA5 (conditions atmosphériques, profils verticaux, paramètres de surface), ainsi que de descripteurs topographiques provenant des données SRTM. L'ensemble de ces informations permet de caractériser l'environnement de propagation, incluant à la fois les effets atmosphériques, maritimes et géographiques.

L'architecture neuronale proposée repose sur GraphSAGE, assurant la capacité de généraliser à des graphes non vus de manière scalable. Afin d'accroître la richesse des représentations, elle est

enrichie par un module *Jumping Knowledge* permettant la fusion multi-échelle des représentations de nœuds, et par des mécanismes d'attention et de multi-agrégation capturant à la fois les interactions locales et les dépendances globales. Enfin, la chaîne de traitement est conçue de manière modulaire, de sorte à permettre l'exploration de variantes architecturales et l'intégration de nouvelles sources de données. Cette modularité garantit à la fois la flexibilité de la méthodologie et son évolutivité.

Prototypage

L'implémentation expérimentale de la méthodologie a nécessité la mise en place de chaînes de traitement ETL, conçus pour nettoyer, transformer et enrichir les données AIS brutes. Ces chaînes de traitement, structurées en plusieurs couches (L1 : données brutes, L2 : données nettoyées et structurées, L3 : jeux validés et agrégés), assurent la reproductibilité, la traçabilité et la flexibilité du processus. Elles prennent en charge des opérations variées, allant du décodage et du filtrage à la validation cinématique et à l'alignement spatio-temporel, de sorte à préparer des jeux cohérents et exploitables pour la modélisation. L'entraînement des modèles a été conduit selon une stratégie rigoureuse de partitionnement des données, distinguant des ensembles dédiés à l'apprentissage, à la validation et au test. Cette organisation permet de contrôler la généralisation des modèles, d'optimiser les hyperparamètres et d'éviter tout biais lié à la distribution spatiale ou temporelle des échantillons. Le protocole inclut des mécanismes de suivi en temps réel, de sauvegarde automatique des modèles les plus performants et de gestion des déséquilibres de classes, afin de garantir la robustesse des résultats.

Les traitements sur des données à forte volumétrie ont reposés sur une infrastructure matérielle et logicielle adaptée. Les expériences principales ont été réalisées sur un serveur haute performance, doté de processeurs multicœurs et de plusieurs GPU de grande capacité, permettant d'entraîner des modèles profonds sur plusieurs années de données AIS. Une machine grand public a ensuite été mobilisée pour l'évaluation et l'expérimentation dans des conditions de déploiement réalistes. L'ensemble de la chaîne de traitement logiciel s'appuie sur un cadre modulaire, intégrant des bibliothèques récentes pour l'orchestration des phases d'entraînement et d'inférence. Enfin, un cadre logiciel comparatif a été implémenté pour situer les performances du modèle proposé par rapport à plusieurs références. Ce banc d'évaluation inclut le modèle semi-déterministe ITU-R P.2001, représentatif des approches physiques ; XGBoost, représentatif des algorithmes classiques d'apprentissage automatique ; ainsi que deux architectures de réseaux de neurones en graphes (GCN et GAT), mobilisant respectivement une agrégation fixe et une agrégation attentionnelle. La diversité de ces modèles de référence permet de comparer la capacité du modèle développé à capturer les relations spatiales et à dépasser les limites des approches existantes.

Evaluation

L'évaluation des performances a été conduite sur un jeu de test indépendant, mobilisant les métriques classiques de classification : exactitude, rappel, précision, F1-score. Le modèle proposé obtient des résultats supérieurs à ceux des approches de référence tout en conservant un temps d'inférence compatible avec une utilisation opérationnelle. Ces performances témoignent de sa capacité à identifier de manière fiable les zones de réception et de non-réception, en limitant les faux positifs et faux négatifs. Au-delà des résultats globaux, une validation régionale a été mise en œuvre afin de tester la robustesse du modèle face à l'hétérogénéité géographique et temporelle des conditions de propagation. Des *clusters* spatio-temporels, construits à partir de la dynamique des réceptions, ont permis de révéler des disparités locales : le modèle conserve une bonne stabilité dans des zones comme le golfe de Gênes ou le bassin Tyrrhénien, où il surpasse XGBoost en réduisant le compromis entre détection de réceptions et limitation des fausses alertes.

Ensuite, une discussion critique a permis de souligner plusieurs limites méthodologiques. La qualité des données AIS constitue une première contrainte : les réceptions utilisées comme vérité terrain sont susceptibles d'être affectées par des erreurs ou par les lacunes de couverture du réseau AISHub, qui peut introduire une ambiguïté dans l'étiquetage. À cela s'ajoutent l'hétérogénéité et la résolution des données environnementales mobilisées (topographie, réanalyses météorologiques), susceptibles de biaiser la représentation des conditions de propagation. Enfin, la complexité computationnelle de l'approche, liée à la construction des graphes et à l'entraînement de modèles profonds, limite le nombre de variantes explorées. En complément de cette analyse, une campagne d'acquisition indépendante sur deux mois de données RSSI a permis d'ouvrir des perspectives de modélisation et de transfert de tâche.

Perspectives

À l'issue de ce travail, plusieurs perspectives se dessinent. Elles concernent d'abord la modélisation, avec l'intégration de nouvelles données et l'exploration de variantes architecturales. Ensuite l'évaluation, par la mise à l'épreuve du modèle dans des contextes diversifiés et notamment face à des données volontairement falsifiées. Enfin, les applications qui ouvrent la voie à une exploitation opérationnelle et à une possible extension vers d'autres systèmes de surveillance.

Perspectives de modélisation

Plusieurs axes d'amélioration et de prolongement méthodologique peuvent être envisagés à l'issue de ce travail. Le premier concerne l'intégration de nouvelles données. La reconstruction de trajectoires (Iphar et al. 2019) constitue une piste pour générer artificiellement des observations de non-réception, en considérant les périodes durant lesquelles un navire n'apparaît pas dans le réseau de réception. Cette stratégie permettrait de renforcer l'homogénéité spatiale du jeu d'apprentissage. Par

ailleurs, la validation croisée des positions reçues et des non-réceptions par des données issues du SAT-AIS (Carson-Jackson 2012) offrirait un moyen de consolider la fiabilité des étiquettes utilisées. Enfin, l'intégration des caractéristiques des antennes et des directivités d'autres stations du réseau AISHub ouvrirait la voie à l'entraînement de modèles contextualisés, capables de prédire la couverture du point de vue de stations différentes.

Une autre perspective de modélisation réside dans l'évolution de la tâche prédictive elle-même. Dans cette thèse, l'approche retenue est une classification binaire (réception / non-réception). Bien que suffisante pour caractériser des zones de réceptions, plusieurs arguments peuvent motiver un passage à une tâche de régression ou de classification multi-classe. Dans le cas de la régression, la prédiction directe de la valeur du RSSI permettrait de capturer de manière continue les variations du niveau de signal. Cette approche nécessite toutefois une vigilance particulière face aux situations de non-réception, qui introduisent un seuil de sensibilité et posent la problématique de la censure des données (Gustafson et al. 2015). À l'inverse, la classification multi-classe, fondée sur une discrétisation du RSSI en intervalles qualitatifs (par exemple : faible, modéré, bon, excellent), fournit une sortie plus immédiatement interprétable d'un point de vue opérationnel. Cette reformulation permet non seulement de réduire la sensibilité du modèle aux fluctuations locales des mesures continues, mais aussi de s'affranchir des contraintes liées à la censure.

Un autre axe concerne le développement d'un modèle interprétable et explicable. Si l'interprétabilité désigne la capacité d'un modèle à être compris dans son fonctionnement intrinsèque, l'explicabilité vise à fournir des justifications locales ou globales a posteriori des décisions produites (Kim 2015 ; Doshi-Velez et Kim 2018 ; Miller 2019). Dans cette perspective, l'exploitation des coefficients d'attention issus de l'architecture attentionnelle constitue une première voie d'interprétation. L'extraction de ces coefficients d'attention permet de visualiser et de pondérer l'importance relative des arêtes dans la prédiction, en mettant en évidence les relations jugées les plus déterminantes par le modèle (Veličković et al. 2018). La Figure C.0.1 illustre cette approche : pour un modèle attentionnel donné, les 10 % des arêtes les plus influentes sont représentées et colorées en fonction de leur poids dans la décision. Contrairement aux données tabulaires ou aux images, les méthodes d'explicabilité usuelles ne peuvent pas toujours être directement appliquées aux graphes, en raison de leur structure relationnelle complexe. Cette analyse peut néanmoins être complétée par des approches d'explicabilité fondées sur l'importance des variables, comme celle proposée dans (Ying et al. 2019). Celles-ci permettent d'identifier les facteurs environnementaux les plus contributifs et de consolider les résultats obtenus par l'analyse des arêtes en un cadre explicatif cohérent.

L'explicabilité des modèles ouvre des perspectives scientifiques pour améliorer la compréhension physique des mécanismes sous-jacents. En particulier, l'utilisation conjointe de l'attention et des méthodes d'explicabilité pourrait faciliter une discrétisation explicite des

phénomènes de propagation. Une telle approche offrirait la possibilité d’associer chaque prédiction à un scénario physique interprétable. Parallèlement, l’exploration d’approches de fusion au niveau des décisions (*late fusion* (Snoek et al. 2005)) représente une piste complémentaire. Elle consisterait à entraîner des sous-modèles spécialisés sur des modalités distinctes (par exemple descripteurs météorologiques différents, représentatifs de phénomènes de propagation distincts) puis à agréger leurs prédictions. Outre l’amélioration potentielle des performances, cette stratégie offrirait une meilleure lisibilité quant au rôle joué par chaque phénomène de propagation dans la prédiction.

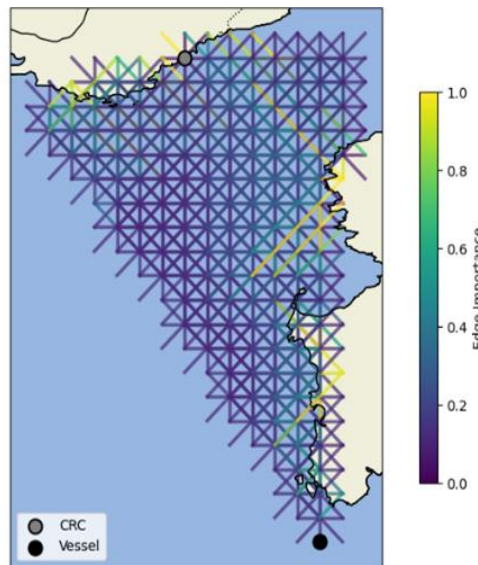


Figure C.0.1 : 10 % des arêtes les plus importantes dans la prédiction d’un modèle attentionnel nœud émetteur mis en évidence.

Enfin, l’extension vers des variantes architecturales temporelles (Yu et al. 2017 ; Zhao et al. 2020 ; Ni et al. 2022 ; Alourani et al. 2023 ; Jiang et al. 2024) constitue un axe particulièrement prometteur. Cette approche permettrait de modéliser explicitement l’évolution des conditions atmosphériques et de dégager des représentations de plus haut niveau dans les situations ambiguës ou difficilement discriminables avec des données instantanées.

Perspectives d’évaluation

Les perspectives d’évaluation ouvrent plusieurs pistes destinées à renforcer la robustesse et la généralisabilité du modèle. Une première étape consisterait à étendre la validation à des zones maritimes différentes de celles utilisées dans le cadre expérimental initial. Cette diversification permettrait de vérifier la capacité du modèle à généraliser vers des contextes géographiques variés, caractérisés par des conditions topographiques, climatiques et opérationnelles distinctes.

Un deuxième axe concerne l’évaluation de la robustesse face à des données imparfaites. Dans un contexte opérationnel, les données AIS peuvent être bruitées, incomplètes ou intentionnellement falsifiées (Laso et al. 2017 ; Iphar et al. 2020). Tester la résilience du modèle face à ces perturbations

constitue un enjeu crucial, tant pour la validation scientifique que pour les applications pratiques de sûreté et de sécurité maritime. Enfin, toujours dans une optique de généralisation, une perspective majeure réside dans l'exploration de la transférabilité du modèle vers d'autres domaines et bandes de fréquences. Le cas du trafic aérien, avec le système ADS-B (Schäfer et al. 2014), offre un terrain d'expérimentation pertinent : la proximité conceptuelle et fréquentielle (1090 MHz) entre AIS et ADS-B ouvre la possibilité d'adapter la méthodologie développée afin de prédire la couverture radio dans le domaine aérien. Plus largement, ce type de démarche pourrait être transposé à d'autres systèmes de communication et de surveillance fonctionnant sur des fréquences voisines.

Perspectives d'application

Les perspectives d'application de ce travail s'inscrivent dans le cadre plus large de l'amélioration de la connaissance de la situation maritime. En produisant une cartographie dynamique des zones de réception de l'AIS, le modèle proposé contribue à affiner la vision globale de l'espace maritime et à renforcer la capacité d'anticipation et de réaction des autorités civiles et militaire. Une telle modélisation ouvre également la voie à la détection de comportements suspects (Riveiro et al. 2018 ; Wolsing et al. 2022 ; Ribeiro et al. 2023). L'analyse des zones de réception peut permettre en effet d'identifier les falsifications ou manipulations des signaux AIS, les inadéquations entre la trajectoire et le contexte de propagation, ainsi que les interruptions volontaires du dispositif AIS. Ces capacités de détection sont cruciales pour prévenir les menaces liées à la piraterie, aux trafics illicites ou encore à la pêche illégale.

Il convient de souligner que les applications directes de cette thèse se situent au cœur de la sûreté et de la sécurité maritime. Dans la logique de la sûreté, la modélisation des zones de réception de l'AIS participe à la prévention des actes intentionnels portant atteinte à la souveraineté et aux intérêts stratégiques des États. Du point de vue de la sécurité, elle permet de mieux prévenir les accidents et incidents liés à une mauvaise compréhension de la couverture radio, et d'assurer ainsi une meilleure protection des navires, des équipages et de l'environnement marin.

Bibliographie

- Agarap, Abien Fred. 2018. « Deep Learning using Rectified Linear Units (ReLU) ». Version 2. Prépublication, arXiv. <https://doi.org/10.48550/ARXIV.1803.08375>.
- Aggarwal, Charu C. 2017. « An Introduction to Outlier Analysis ». In *Outlier Analysis*. Springer International Publishing. https://doi.org/10.1007/978-3-319-47578-3_1.
- Agnew, David J., John Pearce, Ganapathiraju Pramod, et al. 2009. « Estimating the Worldwide Extent of Illegal Fishing ». *PLoS ONE* 4 (2): e4570. <https://doi.org/10.1371/journal.pone.0004570>.
- Airbus Defence and Space. 2020. « La France compte à nouveau sur Airbus pour la maintenance de son système de surveillance côtière — SPATIONAV ». juillet 8. https://www.airbus.com/sites/g/files/jlcpta136/files/9b0b53c52d65a634e109de5517b26279_FR-Press-Release-Connected-Intelligence-Spationav-Maintenance-Contract-Airbus.pdf.
- Akrout, Mohamed, Amal Feriani, Faouzi Bellili, Amine Mezghani, et Ekram Hossain. 2023. « Domain Generalization in Machine Learning Models for Wireless Communications: Concepts, State-of-the-Art, and Open Issues ». *IEEE Communications Surveys & Tutorials* 25 (4): 3014-37. <https://doi.org/10.1109/COMST.2023.3326399>.
- Al-Ahmed, Shahriar Abdullah, et Muhammad Zeeshan Shakir. 2022. « Semi-supervised learning-based coverage hole detection in cellular networks ». *ITU Journal on Future and Evolving Technologies* 3 (3): 623-31. <https://doi.org/10.52953/tlfd1744>.
- Aldossari, Saud Mobark, et Kwang-Cheng Chen. 2019. « Machine Learning for Wireless Communication Channel Modeling: An Overview ». *Wireless Personal Communications* 106 (1): 41-70. <https://doi.org/10.1007/s11277-019-06275-4>.
- Alduchov, Oleg A., et Robert E. Eskridge. 1996. « Improved Magnus Form Approximation of Saturation Vapor Pressure ». *Journal of Applied Meteorology* 35 (4): 601-9. [https://doi.org/10.1175/1520-0450\(1996\)035%253C0601:imfaos%253E2.0.co;2](https://doi.org/10.1175/1520-0450(1996)035%253C0601:imfaos%253E2.0.co;2).
- Alincourt, Erwan, Cyril Ray, Pierre-Michel Ricordel, Delphine Dare-Emzivat, et Abdel Boudraa. 2016. « Methodology for AIS signature identification through magnitude and temporal characterization ». *OCEANS 2016 - Shanghai*, 1-6. <https://doi.org/10.1109/OCEANSAP.2016.7485420>.
- Al-Jumaily, Abdulmajeed, Aduwati Sali, Víctor P. Gil Jiménez, et al. 2023. « Evaluation of 5G and Fixed-Satellite Service Earth Station (FSS-ES) Downlink Interference Based on Artificial Neural Network Learning Models (ANN-LMS) ». *Sensors (Basel, Switzerland)* 23 (13): 6175. <https://doi.org/10.3390/s23136175>.
- Alourani, Abdullah, Farzeen Ashfaq, N. Z. Jhanjhi, et Navid Ali Khan. 2023. « BiLSTM- and GNN-Based Spatiotemporal Traffic Flow Forecasting with Correlated Weather Data ». *Journal of Advanced Transportation* 2023 (1): 8962283. <https://doi.org/10.1155/2023/8962283>.
- Amini, Massih-Reza, Vasilii Feofanov, Loïc Pauletto, Liès Hadjadj, Émilie Devijver, et Yury Maximov. 2025. « Self-training: A survey ». *Neurocomputing* 616 (février): 128904. <https://doi.org/10.1016/j.neucom.2024.128904>.

- Ansel, Jason, Edward Yang, Horace He, et al. 2024. « PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation ». *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*, avril. <https://doi.org/10.1145/3620665.3640366>.
- Atrey, Pradeep K., M. Anwar Hossain, Abdulmotaleb El Saddik, et Mohan S. Kankanhalli. 2010. « Multimodal Fusion for Multimedia Analysis: A Survey ». *Multimedia Systems* 16 (6): 345-79. <https://doi.org/10.1007/s00530-010-0182-0>.
- Baek, Jinheon, Minki Kang, et Sung Ju Hwang. 2021. « Accurate Learning of Graph Representations with Graph Multiset Pooling ». *International Conference on Learning Representations*. <https://openreview.net/forum?id=JHcqXGaqiGn>.
- Bakirtzis, Stefanos, Kehai Qiu, Jie Zhang, et Ian Wassell. 2022. « DeepRay: Deep Learning Meets Ray-Tracing ». *2022 16th European Conference on Antennas and Propagation (EuCAP)*, 1-5. <https://doi.org/10.23919/EuCAP53622.2022.9769203>.
- Balduzzi, Marco, Alessandro Pasta, et Kyle Wilhoit. 2014. « A Security Evaluation of AIS Automated Identification System ». *Proceedings of the 30th Annual Computer Security Applications Conference* (New York, NY, USA), ACSAC '14, 436-45. <https://doi.org/10.1145/2664243.2664257>.
- Baltrušaitis, Tadas, Chaitanya Ahuja, et Louis-Philippe Morency. 2019. « Multimodal Machine Learning: A Survey and Taxonomy ». *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41 (2): 423-43. <https://doi.org/10.1109/TPAMI.2018.2798607>.
- Barnum, George, Sabera J. Talukder, et Yisong Yue. 2020. « On the Benefits of Early Fusion in Multimodal Representation Learning ». *NeurIPS 2020 Workshop SVRHM*. <https://openreview.net/forum?id=FMJ5e0IoFFk>.
- Barrick, Donald E. 1971. « Theory of HF and VHF Propagation Across the Rough Sea, 2, Application to HF and VHF Propagation Above the Sea ». *Radio Science* 6 (5): 527-33. <https://doi.org/10.1029/RS006i005p00527>.
- Barrios, A.E. 1994. « A terrain parabolic equation model for propagation in the troposphere ». *IEEE Transactions on Antennas and Propagation* 42 (1): 90-98. <https://doi.org/10.1109/8.272306>.
- Baylon, Angelica, et Eduardo Ma Santos. 2013. « Introducing GIS to TransNav and its Extensive Maritime Application: An Innovative Tool for Intelligent Decision Making? ». *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation* 7 (4): 557-66. <https://doi.org/10.12716/1001.07.04.11>.
- Bean, Bradford R, et EJ Dutton. 1966. *Radio meteorology*. Vol. 92. Superintendent of Documents, US Government Print. Office.
- Belmonte-Hernández, Alberto, Gustavo Hernández-Peñaloza, David Martín Gutiérrez, et Federico Álvarez. 2020. « Recurrent Model for Wireless Indoor Tracking and Positioning Recovering Using Generative Networks ». *IEEE Sensors Journal* 20 (6): 3356-65. <https://doi.org/10.1109/JSEN.2019.2958201>.
- Berthelot, David, Rebecca Roelofs, Kihyuk Sohn, Nicholas Carlini, et Alexey Kurakin. 2022. « AdaMatch: A Unified Approach to Semi-Supervised Learning and Domain Adaptation ».

International Conference on Learning Representations.
<https://openreview.net/forum?id=Q5uh1Nvv5dm>.

- Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag.
- Bocus, Mohammud Z., et Afzal Lodhi. 2025. « Application of Machine Learning for Radiowave Propagation Modeling Below 6 GHz ». *IEEE Access* 13: 9755-65. <https://doi.org/10.1109/ACCESS.2025.3526329>.
- Bronstein, Michael M., Joan Bruna, Yann LeCun, Arthur Szlam, et Pierre Vandergheynst. 2017. « Geometric Deep Learning: Going beyond Euclidean data ». *IEEE Signal Processing Magazine* 34 (4): 18-42. <https://doi.org/10.1109/MSP.2017.2693418>.
- Bruin, E. R. 2016. *On Propagation Effects in Maritime Situation Awareness: Modelling the Impact of North Sea Weather Conditions on the Performance of AIS and Coastal Radar Systems*. <https://studenttheses.uu.nl/handle/20.500.12932/21952>.
- Bruwer, Carina. 2020. « Smuggling and Trafficking of Illicit Goods by Sea ». In *Global Challenges in Maritime Security: An Introduction*, édité par Lisa Otto. Springer International Publishing. https://doi.org/10.1007/978-3-030-34630-0_4.
- Bueger, Christian. 2015. « What is maritime security? » *Marine Policy* 53: 159-64. <https://doi.org/10.1016/j.marpol.2014.12.005>.
- Bueger, Christian, et Tobias Liebetau. 2023. « Critical maritime infrastructure protection: What's the trouble? » *Marine Policy* 155: 105772. <https://doi.org/10.1016/j.marpol.2023.105772>.
- Bufort, Adrien, Laurent Lebecq, et Stefan Cathabard. 2023. « Data-Driven Radio Propagation Modeling using Graph Neural Networks ». Prépublication, novembre 7. <https://doi.org/10.36227/techrxiv.22820273.v2>.
- Buterez, David, Jon Paul Janet, Steven J. Kiddle, Dino Oglic, et Pietro Liò. 2022. « Graph Neural Networks with Adaptive Readouts ». *Advances in Neural Information Processing Systems* 35 (décembre): 19746-58.
- C3S. 2018a. « ERA5 hourly data on pressure levels from 1940 to present ». Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://doi.org/10.24381/CDS.BD0915C6>.
- C3S. 2018b. « ERA5 hourly data on single levels from 1940 to present ». Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://doi.org/10.24381/CDS.ADBB2D47>.
- Caillet, T. 2015. *Rapport ANFR sur les phénomènes de propagation anormale : propagation par conduit et réflexion/réfraction sur les couches élevées de l'atmosphère*. ANFR(15)01.
- Carson-Jackson, J. 2012. « Satellite AIS – Developing Technology or Existing Capability? » *The Journal of Navigation* 65 (2): 303-21. <https://doi.org/10.1017/S037346331100066X>.
- Castanet, Laurent, et Patrick Lassudrie-Duchesne. 2008. *Propagation des ondes radioélectriques à travers l'atmosphère*. http://olivier.granier.free.fr/Seq16/res/ondes-plasma_atm.pdf.

- Çayir, Aykut, Işil Yenidoğan, et Hasan Dağ. 2018. « Feature Extraction Based on Deep Learning for Some Traditional Machine Learning Methods ». *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, septembre, 494-97. <https://doi.org/10.1109/UBMK.2018.8566383>.
- Chapelle, O., B. Scholkopf, et A. Zien Eds. 2009. « Semi-Supervised Learning (Chapelle, O. et al., Eds.; 2006) [Book reviews] ». *IEEE Transactions on Neural Networks* 20 (3): 542-542. <https://doi.org/10.1109/TNN.2009.2015974>.
- Chen, Tianqi, et Carlos Guestrin. 2016. « XGBoost: A Scalable Tree Boosting System ». *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, août 13, 785-94. <https://doi.org/10.1145/2939672.2939785>.
- Chiroma, Haruna, Ponman Nickolas, Nasir Faruk, et al. 2023. « Large scale survey for radio propagation in developing machine learning model for path losses in communication systems ». *Scientific African* 19 (mars): e01550. <https://doi.org/10.1016/j.sciaf.2023.e01550>.
- Cho, Kyunghyun, Bart van Merriënboer, Caglar Gulcehre, et al. 2014. « Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation ». In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, édité par Alessandro Moschitti, Bo Pang, et Walter Daelemans. Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1179>.
- Choi, Seongro, Jun Heo, Changseong Kim, Sungsik Wang, Hosung Choo, et Yong Bae Park. 2020. « Prediction of Electromagnetic Wave Propagation in Troposphere Using Parabolic Equation and Two-Dimensional Refractivity ». *Journal of Electrical Engineering & Technology* 15 (3): 1287-92. <https://doi.org/10.1007/s42835-020-00395-9>.
- Coad, Peter, Edward Yourdon, et others. 1992. *Object-oriented analysis*. Vol. 2. Yourdon press New York.
- Commission of the European Communities. 2009. *Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions: Towards the integration of maritime surveillance: A common information sharing environment for the EU maritime domain*. COM(2009)538 final. European Commission. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0538:FIN:EN:PDF>.
- Copernicus Climate Change Service. 2019. « Land cover classification gridded maps from 1992 to present derived from satellite observations ». ECMWF. <https://doi.org/10.24381/CDS.006F2C9A>.
- Cornélis, Pierre. 2005. « Chapitre 7: La propagation ». In *Cours de Radioamateur en Vue de L'obtention de La Licence Complète*. ON7PC.
- Corso, Gabriele, Luca Cavalleri, Dominique Beaini, Pietro Liò, et Petar Velickovic. 2020. « Principal neighbourhood aggregation for graph nets ». *Proceedings of the 34th International Conference on Neural Information Processing Systems (Red Hook, NY, USA), NIPS '20*, décembre 6, 13260-71.
- Démoulin, Bernard. 2012. « « Quelques aspects de la propagation des ondes radioélectriques » ». *Territoire en mouvement Revue de géographie et aménagement* 12: 30-45.

- Direction interrégionale de la mer Manche Est – Mer du Nord. 2024. « CROSS Gris-Nez. Bilan d'activité 2024 ». https://www.cross-grisnez.mer.gouv.fr/IMG/pdf/bilan_2024_cross_gris-nez.pdf.
- Doshi-Velez, Finale, et Been Kim. 2018. « Considerations for Evaluation and Generalization in Interpretable Machine Learning ». In *Explainable and Interpretable Models in Computer Vision and Machine Learning*, édité par Hugo Jair Escalante, Sergio Escalera, Isabelle Guyon, et al. Springer International Publishing. https://doi.org/10.1007/978-3-319-98131-4_1.
- Drucker, Harris, Chris J. C. Burges, Linda Kaufman, Alex Smola, et Vladimir Vapnik. 1996. « Support vector regression machines ». *Proceedings of the 10th International Conference on Neural Information Processing Systems* (Cambridge, MA, USA), NIPS'96, décembre 3, 155-61.
- Dwivedi, Vijay Prakash, Chaitanya K. Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, et Xavier Bresson. 2023. « Benchmarking Graph Neural Networks ». *Journal of Machine Learning Research* 24 (43): 1-48.
- Earth Resources Observation and Science (EROS) Center. 2017. « Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global ». U.S. Geological Survey. Tiff. <https://doi.org/10.5066/F7PR7TFT>.
- Egli, John J. 1957. « Radio Propagation above 40 MC over Irregular Terrain ». *Proceedings of the IRE* 45 (10): 1383-91. <https://doi.org/10.1109/JRPROC.1957.278224>.
- Elmezughi, Mohamed K., Omran Salih, Thomas J. Afullo, et Kevin J. Duffy. 2022. « Comparative Analysis of Major Machine-Learning-Based Path Loss Models for Enclosed Indoor Channels ». *Sensors* 22 (13): 13. <https://doi.org/10.3390/s22134967>.
- Emmens, Ties, Chintan Amrit, Asad Abdi, et Mayukh Ghosh. 2021. « The Promises and Perils of Automatic Identification System Data ». *Expert Systems with Applications* 178 (avril): 114975. <https://doi.org/10.1016/j.eswa.2021.114975>.
- Engeln, Axel von, et João Teixeira. 2004. « A ducting climatology derived from the European Centre for Medium-Range Weather Forecasts global analysis fields ». *Journal of Geophysical Research: Atmospheres* 109 (D18). <https://doi.org/10.1029/2003JD004380>.
- EROS Center. 2017. « Shuttle Radar Topography Mission (SRTM) Void Filled ». U.S. Geological Survey. Tiff,Other. <https://doi.org/10.5066/F7F76B1X>.
- European Commission, Directorate-General for the Information Society, et Media. 1999. *COST Action 231 : Digital mobile radio towards future generation systems: Final Report*. Publications Office.
- Falcon, William et The PyTorch Lightning team. 2019. *PyTorch Lightning*. Version 1.4. Released mars. <https://doi.org/10.5281/zenodo.3828935>.
- Faruk, Nasir, I.Y. Abdulrasheed, N.T. Surajudeen-Bakinde, et al. 2021. « Large-scale radio propagation path loss measurements and predictions in the VHF and UHF bands ». *Heliyon* 7 (6). <https://doi.org/10.1016/j.heliyon.2021.e07298>.
- Fey, Matthias, et Jan Eric Lenssen. 2019. *Fast Graph Representation Learning with PyTorch Geometric*. Released mai. https://github.com/pyg-team/pytorch_geometric.

- Fix, Evelyn, et J. L. Hodges. 1989. « Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties ». *International Statistical Review / Revue Internationale de Statistique* 57 (3): 238. <https://doi.org/10.2307/1403797>.
- Formela, Kamil, Tomasz Neumann, et Adam Weintrit. 2019. « Overview of Definitions of Maritime Safety, Safety at Sea, Navigational Safety and Safety in General ». *TransNav, International Journal on Marine Navigation and Safety of Sea Transportation* 13 (2): 285-90. <https://doi.org/10.12716/1001.13.02.03>.
- Frenay, Benoit, et Michel Verleysen. 2014. « Classification in the Presence of Label Noise: A Survey ». *IEEE Transactions on Neural Networks and Learning Systems* 25 (5): 845-69. <https://doi.org/10.1109/TNNLS.2013.2292894>.
- Freund, Yoav, et Robert E Schapire. 1997. « A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting ». *Journal of Computer and System Sciences* 55 (1): 119-39. <https://doi.org/10.1006/jcss.1997.1504>.
- Friis, H.T. 1946. « A Note on a Simple Transmission Formula ». *Proceedings of the IRE* 34 (5): 254-56. <https://doi.org/10.1109/JRPROC.1946.234568>.
- Gadzicki, Konrad, Razieh Khamsehashari, et Christoph Zetsche. 2020. « Early vs Late Fusion in Multimodal Convolutional Neural Networks ». *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, juillet, 1-6. <https://doi.org/10.23919/FUSION45008.2020.9190246>.
- Gao, Miao, et Guo-You Shi. 2020. « Ship-Handling Behavior Pattern Recognition Using AIS Sub-Trajectory Clustering Analysis Based on the T-SNE and Spectral Clustering Algorithms ». *Ocean Engineering* 205 (juin): 106919. <https://doi.org/10.1016/j.oceaneng.2020.106919>.
- Gilmer, Justin, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, et George E. Dahl. 2017. « Neural message passing for Quantum chemistry ». *Proceedings of the 34th International Conference on Machine Learning - Volume 70 (Sydney, NSW, Australia), ICML'17, août 6, 1263-72*.
- Gómez-Guillamón Buendía, Victoria, Salvatore Liberto, George Goussetis, et Nelson J. G. Fonseca. 2023. « Review of Antenna Technologies for Very High Frequency Data Exchange Systems ». *International Journal of Satellite Communications and Networking* 41 (2): 122-33. <https://doi.org/10.1002/sat.1428>.
- Goodfellow, Ian, Yoshua Bengio, et Aaron Courville. 2016. *Deep Learning*. MIT Press.
- Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, et al. 2020. « Generative adversarial networks ». *Commun. ACM* 63 (11): 139-44. <https://doi.org/10.1145/3422622>.
- Grandjean, Pascal. 2016. « Prévisions de l'indice de réfraction dans les couches basses de la troposphère ». *Radio Ref*, n° 896 (mars): 21-25.
- Gudivada, V., A. Apon, et Junhua Ding. 2017. « Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations ». <https://www.semanticscholar.org/paper/Data-Quality-Considerations-for-Big-Data-and-Going-Gudivada-Apon/625a9e9822603b79f754c4ce044760f7363b5eb6>.

- Guo, Lijie, Liangke Huang, Junyu Li, et al. 2021. « A Comprehensive Evaluation of Key Tropospheric Parameters from ERA5 and MERRA-2 Reanalysis Products Using Radiosonde Data and GNSS Measurements ». *Remote Sensing* 13 (15). <https://doi.org/10.3390/rs13153008>.
- Gustafson, Carl, Taimoor Abbas, David Bolin, et Fredrik Tufvesson. 2015. « Statistical Modeling and Estimation of Censored Pathloss Data ». *IEEE Wireless Communications Letters* 4 (5): 569-72. <https://doi.org/10.1109/LWC.2015.2463274>.
- Halder, Rajib Kumar, Mohammed Nasir Uddin, Md. Ashraf Uddin, Sunil Aryal, et Ansam Khraisat. 2024. « Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications ». *Journal of Big Data* 11 (1): 113. <https://doi.org/10.1186/s40537-024-00973-y>.
- Hamilton, Will, Zhitao Ying, et Jure Leskovec. 2017. « Inductive Representation Learning on Large Graphs ». In *Advances in Neural Information Processing Systems*, édité par I. Guyon, U. Von Luxburg, S. Bengio, et al., vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/5dd9db5e033da9c6fb5ba83c7a7eb ea9-Paper.pdf.
- Hammond, T. R., et D. J. Peters. 2012. « Estimating AIS Coverage from Received Transmissions ». *The Journal of Navigation* 65 (3): 409-25. <https://doi.org/10.1017/S0373463312000057>.
- Han, Jie, Jiaji Wu, Lijun Zhang, et al. 2022. « A Classifying-Inversion Method of Offshore Atmospheric Duct Parameters Using AIS Data Based on Artificial Intelligence ». *Remote Sensing* 14 (13): 3197. <https://doi.org/10.3390/rs14133197>.
- Harati-Mokhtari, Abbas, Alan Wall, Philip Brooks, et Jin Wang. 2007. « Automatic Identification System (AIS): Data Reliability and Human Error Implications ». *The Journal of Navigation* 60 (3): 373-89. <https://doi.org/10.1017/S0373463307004298>.
- Hartawan, I Putu Noven, I Made Oka Widyantara, A. A. I. N. E. Karyawati, Ngurah Indra Er, Ketut Buda Artana, et Nyoman Putra Sastra. 2021. « AIS Data Pre-Processing for Trajectory Clustering Data Preparation ». *2021 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES)*, 1-5. <https://doi.org/10.1109/ICARES53960.2021.9665187>.
- Hastie, Trevor, Robert Tibshirani, et Jerome Friedman. 2009a. « Basis Expansions and Regularization ». In *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, édité par Trevor Hastie, Robert Tibshirani, et Jerome Friedman. Springer. https://doi.org/10.1007/978-0-387-84858-7_5.
- Hastie, Trevor, Robert Tibshirani, et Jerome Friedman. 2009b. « Model Assessment and Selection ». In *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, édité par Trevor Hastie, Robert Tibshirani, et Jerome Friedman. Springer. https://doi.org/10.1007/978-0-387-84858-7_7.
- Hata, M. 1980. « Empirical formula for propagation loss in land mobile radio services ». *IEEE Transactions on Vehicular Technology* 29 (3): 317-25. <https://doi.org/10.1109/TVT.1980.23859>.
- Hawkins, D. M. 1980. « Introduction ». In *Identification of Outliers*. Springer Netherlands. https://doi.org/10.1007/978-94-015-3994-4_1.

- Hayashi, Takahiro, Tatsuya Nagao, et Satoshi Ito. 2020. « A study on the variety and size of input data for radio propagation prediction using a deep neural network ». *2020 14th European Conference on Antennas and Propagation (EuCAP)*, mars, 1-5. <https://doi.org/10.23919/EuCAP48036.2020.9135876>.
- He, Shiwen, Shaowen Xiong, Yeyu Ou, et al. 2021. « An Overview on the Application of Graph Neural Networks in Wireless Networks ». *IEEE Open Journal of the Communications Society* 2: 2547-65. <https://doi.org/10.1109/OJCOMS.2021.3128637>.
- He, Wei, Jinyu Lei, Xiumin Chu, Shuo Xie, Cheng Zhong, et Zhixiong Li. 2021. « A Visual Analysis Approach to Understand and Explore Quality Problems of AIS Data ». *Journal of Marine Science and Engineering* 9 (2). <https://doi.org/10.3390/jmse9020198>.
- Hersbach, Hans, Bill Bell, Paul Berrisford, et al. 2020. « The ERA5 global reanalysis ». *Quarterly Journal of the Royal Meteorological Society* 146 (730): 1999-2049. <https://doi.org/10.1002/qj.3803>.
- Hitney, H., et R. Vieth. 1990. « Statistical assessment of evaporation duct propagation ». *IEEE Transactions on Antennas and Propagation* 38 (6): 794-99. <https://doi.org/10.1109/8.55574>.
- Hochreiter, Sepp, et Jürgen Schmidhuber. 1997. « Long Short-Term Memory ». *Neural Computation* 9 (8): 1735-80. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Huang, Chen, Ruisi He, Bo Ai, et al. 2022. « Artificial Intelligence Enabled Radio Propagation for Communications—Part II: Scenario Identification and Channel Modeling ». *IEEE Transactions on Antennas and Propagation* 70 (6): 3955-69. <https://doi.org/10.1109/TAP.2022.3149665>.
- Huang, I.-Lun, Man-Chun Lee, Chung-Yuan Nieh, et Juan-Chen Huang. 2024. « Ship Classification Based on AIS Data and Machine Learning Methods ». *Electronics* 13 (1): 98. <https://doi.org/10.3390/electronics13010098>.
- Huang, Liangke, Xiaoyang Fang, Tengxu Zhang, Haoyu Wang, Lei Cui, et Lilong Liu. 2023. « Evaluation of surface temperature and pressure derived from MERRA-2 and ERA5 reanalysis datasets and their applications in hourly GNSS precipitable water vapor retrieval over China ». *Geodesy and Geodynamics* 14 (2): 111-20. <https://doi.org/10.1016/j.geog.2022.08.006>.
- Huang, Li-Feng, Cheng-Guo Liu, Hong-Guang Wang, et al. 2022. « Experimental Analysis of Atmospheric Ducts and Navigation Radar Over-the-Horizon Detection ». *Remote Sensing* 14 (11): 2588. <https://doi.org/10.3390/rs14112588>.
- Hufford, G. A. 1995. « The ITS Irregular Terrain Model ». *version 1. 2 2.* (2).
- Hufford, G. A., A. G. Longley, et W. A. Kissick. 1982. *A guide to the use of the ITS Irregular Terrain Model in the area prediction mode*. NTIA Report 82-100.
- IEEE. 1993. « IEEE Standard Definitions of Terms for Antennas ». *IEEE Std 145-1993*, 1-32. <https://doi.org/10.1109/IEEESTD.1993.119664>.
- IEEE. 2019. « IEEE Standard Definitions of Terms for Radio Wave Propagation ». *IEEE Std 211-2018 (Revision of IEEE Std 211-1997)*, 1-57. <https://doi.org/10.1109/IEEESTD.2019.8657413>.

- Imai, T., K. Kitao, et M. Inomata. 2019. « Radio Propagation Prediction Model Using Convolutional Neural Networks by Deep Learning ». *2019 13th European Conference on Antennas and Propagation (EuCAP)*, 1-5.
- IMO. 2003. *Guidelines for the Installation of a Shipborne Automatic Identification System (AIS)*. [Online]. <http://www.imo.org/ourwork/safety/navigation/documents/227.pdf>.
- Iphar, Clément. 2017. « Formalisation of a data analysis environment based on anomaly detection for risk assessment : Application to Maritime Domain Awareness ». n° 2017PSLEM041. Theses, Université Paris sciences et lettres. <https://pastel.hal.science/tel-01783958>.
- Iphar, Clément, Anne-Laure Jousset, et Cyril Ray. 2019. « Pseudo-synthetic datasets in support to maritime surveillance algorithms assessment ». *proceedings of the VERITA Workshop, 19ieme Journées Francophones Extraction et Gestion des Connaissances (EGC)*.
- Iphar, Clément, Cyril Ray, et Aldo Napoli. 2020. « Data Integrity Assessment for Maritime Anomaly Detection ». *Expert Systems with Applications* 147 (juin): 113219. <https://doi.org/10.1016/j.eswa.2020.113219>.
- Isabona, Joseph, Agbotiname Lucky Imoize, Stephen Ojo, et al. 2022. « Development of a Multilayer Perceptron Neural Network for Optimal Predictive Modeling in Urban Microcellular Radio Environments ». *Applied Sciences* 12 (11): 5713. <https://doi.org/10.3390/app12115713>.
- Isbaex, Crismeire, Francisco dos Reis Fernandes Costa, et Teresa Batista. 2025. « Application of GIS in the Maritime-Port Sector: A Systematic Review ». *Sustainability* 17 (8). <https://doi.org/10.3390/su17083386>.
- Iscen, Ahmet, Giorgos Toliass, Yannis Avrithis, et Ondrej Chum. 2019. « Label Propagation for Deep Semi-Supervised Learning ». 5070-79. https://openaccess.thecvf.com/content_CVPR_2019/html/Iscen_Label_Propagation_for_Deep_Semi-Supervised_Learning_CVPR_2019_paper.html.
- ITU-R. 1994. *Recommendation ITU-R P.844-1: Ionospheric factors affecting frequency sharing in the VHF and UHF bands (30 MHz–3 GHz)*. Recommendation P.844-1. International Telecommunication Union.
- ITU-R. 2012. *Recommendation ITU-R P.676-9, Attenuation by atmospheric gases*. P Series Radiowave propagation. International Telecommunication Union.
- ITU-R. 2014. *Recommendation ITU-R M.1371-5: Technical characteristics for an automatic identification system using time-division multiple access in the VHF maritime mobile frequency band*. Recommendation M.1371-5. International Telecommunication Union.
- ITU-R. 2019a. *Recommendation ITU-R P.310-10: Definitions of terms relating to propagation in non-ionized media*. Recommendation P.310-10. International Telecommunication Union.
- ITU-R. 2019b. *Recommendation ITU-R P.453: The radio refractive index: its formula and refractivity data*. Recommendation P.453. International Telecommunication Union.
- ITU-R. 2019c. *Recommendation ITU-R P.526-15: Propagation by diffraction*. Recommendation P.526-15. International Telecommunication Union.

- ITU-R. 2019d. *Recommendation ITU-R P.1546-6: Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 4 000 MHz*. Recommendation P.1546-6. International Telecommunication Union.
- ITU-R. 2021a. *Recommendation ITU-R P.527-6: Electrical characteristics of the surface of the Earth*. Recommendation P.527-6. International Telecommunication Union.
- ITU-R. 2021b. *Recommendation ITU-R P.534-6: Method for calculating sporadic-E field strength*. Recommendation P.534-6. International Telecommunication Union.
- ITU-R. 2022. *Recommendation ITU-R P.841-7: Conversion of annual statistics to worst-month statistics*. Recommendation P.841-7. International Telecommunication Union.
- ITU-R. 2023a. *Recommendation ITU-R P.452-18: Prediction procedure for the evaluation of interference between stations on the surface of the Earth at frequencies above about 100 MHz*. Recommendation P.452-18. International Telecommunication Union.
- ITU-R. 2023b. *Recommendation ITU-R P.1812-7: A path-specific propagation prediction method for point-to-area terrestrial services in the frequency range 30 MHz to 6 GHz*. Recommendation P.1812-7. International Telecommunication Union.
- ITU-R. 2023c. *Recommendation ITU-R P.2001-5: A general purpose wide-range terrestrial propagation model in the frequency range 30 MHz to 50 GHz*. Recommendation P.2001-5. International Telecommunication Union.
- James, Mark, Tania Mendo, Esther L. Jones, Kyla Orr, Ali McKnight, et John Thompson. 2018. « AIS data to inform small scale fisheries management and marine spatial planning ». *Marine Policy* 91: 113-21. <https://doi.org/10.1016/j.marpol.2018.02.012>.
- Ji, Hanjie, Lixin Guo, Jinpeng Zhang, et al. 2024. « IPILT–OHPL: An Over-the-Horizon Propagation Loss Prediction Model Established by Incorporating Prior Information Into the LSTM–Transformer Structure ». *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17: 10067-82. <https://doi.org/10.1109/JSTARS.2024.3395630>.
- Jiang, Junhao, Yi Zuo, Yang Xiao, Wenjun Zhang, et Tieshan Li. 2024. « STMGF-Net: A Spatiotemporal Multi-Graph Fusion Network for Vessel Trajectory Forecasting in Intelligent Maritime Navigation ». *IEEE Transactions on Intelligent Transportation Systems* 25 (12): 21367-79. <https://doi.org/10.1109/TITS.2024.3465234>.
- Juang, Rong-Terng. 2022. « Deep Learning-Based Path Loss Model in Urban Environments Using Image-to-Image Translation ». *IEEE Transactions on Antennas and Propagation* 70 (12): 12081-91. <https://doi.org/10.1109/TAP.2022.3209229>.
- Kelly, Peter. 2022. « A novel technique to identify AIS transmissions from vessels which attempt to obscure their position by switching their AIS transponder from normal transmit power mode to low transmit power mode ». *Expert Systems with Applications* 202 (septembre): 117205. <https://doi.org/10.1016/j.eswa.2022.117205>.
- Kim, Been. 2015. « Interactive and interpretable machine learning models for human machine collaboration ». <https://api.semanticscholar.org/CorpusID:110496388>.

- Kingma, Diederik P., et Jimmy Ba. 2015. « Adam: A Method for Stochastic Optimization ». *ICLR (Poster)*. <http://arxiv.org/abs/1412.6980>.
- Kipf, Thomas N., et Max Welling. 2017. « Semi-Supervised Classification with Graph Convolutional Networks ». *International Conference on Learning Representations*. <https://openreview.net/forum?id=SJU4ayYgl>.
- Kohonen, Teuvo. 1982. « Self-Organized Formation of Topologically Correct Feature Maps ». *Biological Cybernetics* 43 (1): 59-69. <https://doi.org/10.1007/BF00337288>.
- Koka, Enkelejda, et Denard Veshi. 2019. « Irregular Migration by Sea: Interception and Rescue Interventions in Light of International Law and the EU Sea Borders Regulation ». *European Journal of Migration and Law* (Leiden, The Netherlands) 21 (1): 26-52. <https://doi.org/10.1163/15718166-12340041>.
- Koldasbayeva, Diana, Polina Tregubova, Mikhail Gasanov, Alexey Zaytsev, Anna Petrovskaja, et Evgeny Burnaev. 2024. « Challenges in data-driven geospatial modeling for environmental research and practice ». *Nature Communications* 15 (1). <https://doi.org/10.1038/s41467-024-55240-8>.
- Kram, Sebastian, Maximilian Stahlke, Tobias Feigl, Jochen Seitz, et Jörn Thielecke. 2019. « UWB Channel Impulse Responses for Positioning in Complex Environments: A Detailed Feature Analysis ». *Sensors* 19 (24). <https://doi.org/10.3390/s19245547>.
- Kumar Yadav, Rohan, Bimal Bhattarai, Lei Jiao, Morten Goodwin, et Ole-Christoffer Granmo. 2020. « Indoor Space Classification Using Cascaded LSTM ». *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, novembre, 1110-14. <https://doi.org/10.1109/ICIEA48937.2020.9248347>.
- Lane, Richard O., David. A. Nevell, Steven D. Hayward, et Thomas W. Beaney. 2010. « Maritime anomaly detection and threat assessment ». *2010 13th International Conference on Information Fusion*, 1-8. <https://doi.org/10.1109/ICIF.2010.5711998>.
- Laso, Pedro Merino, David Brosset, et John Puentes. 2017. « Analysis of quality measurements to categorize anomalies in sensor systems ». *2017 Computing Conference*, 1330-38. <https://doi.org/10.1109/SAI.2017.8252263>.
- Last, Philipp, Martin Hering-Bertram, et Lars Linsen. 2015. « How automatic identification system (AIS) antenna setup affects AIS signal quality ». *Ocean Engineering* 100: 83-89. <https://doi.org/10.1016/j.oceaneng.2015.03.017>.
- Lavers, David A., Adrian Simmons, Freja Vamborg, et Mark J. Rodwell. 2022. « An evaluation of ERA5 precipitation for climate monitoring ». *Quarterly Journal of the Royal Meteorological Society* 148 (748): 3152-65. <https://doi.org/10.1002/qj.4351>.
- LeCun, Yann, Yoshua Bengio, et Geoffrey Hinton. 2015. « Deep learning ». *Nature* 521 (7553): 436-44. <https://doi.org/10.1038/nature14539>.
- Lee, Eunsu, Amit Mokashi, Sang Moon, et GeunSub Kim. 2019. « The Maturity of Automatic Identification Systems (AIS) and Its Implications for Innovation ». *Journal of Marine Science and Engineering* 7 (août): 287. <https://doi.org/10.3390/jmse7090287>.

- Lee, In-Seok, Jung-Hoon Noh, et Seong-Jun Oh. 2023. « A Survey and analysis on a troposcatter propagation model based on ITU-R recommendations ». *ICT Express* 9 (3): 507-16. <https://doi.org/10.1016/j.ict.2022.09.009>.
- Levie, Ron, Çağkan Yapar, Gitta Kutyniok, et Giuseppe Caire. 2021. « RadioUNet: Fast Radio Map Estimation With Convolutional Neural Networks ». *IEEE Transactions on Wireless Communications* 20 (6): 4001-15. <https://doi.org/10.1109/TWC.2021.3054977>.
- Li, Ce, Rongpei Hong, Xovee Xu, Goce Trajcevski, et Fan Zhou. 2023. « Simplifying Temporal Heterogeneous Network for Continuous-Time Link Prediction ». *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, octobre 21, 1288-97. <https://doi.org/10.1145/3583780.3615059>.
- Li, Qimai, Zhichao Han, et Xiao-ming Wu. 2018. « Deeper Insights Into Graph Convolutional Networks for Semi-Supervised Learning ». *Proceedings of the AAAI Conference on Artificial Intelligence* 32 (1). <https://doi.org/10.1609/aaai.v32i1.11604>.
- Li, Qiyue, Heng Qu, Zhi Liu, et al. 2021. « AF-DCGAN: Amplitude Feature Deep Convolutional GAN for Fingerprint Construction in Indoor Localization Systems ». *IEEE Transactions on Emerging Topics in Computational Intelligence* 5 (3): 468-80. <https://doi.org/10.1109/TETCI.2019.2948058>.
- Li, Xiaohui, Xiong Cai, Yongqiang Hei, et Ruiyang Yuan. 2017. « NLOS Identification and Mitigation Based on Channel State Information for Indoor WiFi Localisation ». *IET Communications* 11 (4): 531-37. <https://doi.org/10.1049/iet-com.2016.0562>.
- Li, Xiaojie, Songyang Zhang, Hang Li, et al. 2024. « RadioGAT: A Joint Model-Based and Data-Driven Framework for Multi-Band Radiomap Reconstruction via Graph Attention Networks ». *IEEE Transactions on Wireless Communications* 23 (11): 17777-92. <https://doi.org/10.1109/TWC.2024.3457157>.
- Li, Yujia, Chenjie Gu, Thomas Dullien, Oriol Vinyals, et Pushmeet Kohli. 2019. « Graph Matching Networks for Learning the Similarity of Graph Structured Objects ». *Proceedings of the 36th International Conference on Machine Learning*, mai 24, 3835-45. <https://proceedings.mlr.press/v97/li19d.html>.
- Lim, Tjen-Sien, Wei-Yin Loh, et Yu-Shan Shih. 2000. « A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms ». *Machine Learning* 40 (3): 203-28. <https://doi.org/10.1023/A:1007608224229>.
- Liu, Li, Yuguang Yao, Zhichao Cao, et Mi Zhang. 2021. « DeepLoRa: Learning Accurate Path Loss Model for Long Distance Links in LPWAN ». *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, mai, 1-10. <https://doi.org/10.1109/INFOCOM42981.2021.9488784>.
- Liu, Sen, Teruo Onishi, Masao Taki, et Soichi Watanabe. 2023. « A Generalizable Indoor Propagation Model Based on Graph Neural Networks ». *IEEE Transactions on Antennas and Propagation* 71 (7): 6098-110. <https://doi.org/10.1109/TAP.2023.3281061>.
- Liu, Yaxi, Jiangbo Dong, Wei Huangfu, Jiangchuan Liu, et Keping Long. 2022. « 3.5GHz Outdoor Radio Signal Strength Prediction with Machine Learning based on Low-cost Geographic

- Features ». *IEEE Transactions on Antennas and Propagation*, 1-1. <https://doi.org/10.1109/TAP.2022.3175214>.
- Liu, Xuefeng Yin, Xiaokang Ye, Yongyu He, et Juyul Lee. 2021. « Embedded Propagation Graph Model for Reflection and Scattering and Its Millimeter-Wave Measurement-Based Evaluation ». *IEEE Open Journal of Antennas and Propagation* 2: 191-202. <https://doi.org/10.1109/OJAP.2021.3051478>.
- Liu, Yuan, Xuefeng Yin, Juyul Lee, et Meisong Tong. 2020. « A Graph-based Simulation Method for Propagation Channels with Multiple-knife-edge Diffraction ». *2020 IEEE International Conference on Computational Electromagnetics (ICCEM)*, 270-72. <https://doi.org/10.1109/ICCEM47450.2020.9219449>.
- Lloyd, S. 1982. « Least squares quantization in PCM ». *IEEE Transactions on Information Theory* 28 (2): 129-37. <https://doi.org/10.1109/TIT.1982.1056489>.
- Longley, Anita G., et Philip L. Rice. 1968. « PREDICTION OF TROPOSPHERIC RADIO TRANSMISSION LOSS OVER IRREGULAR TERRAIN. A COMPUTER METHOD-1968 ». <https://api.semanticscholar.org/CorpusID:129698815>.
- Lu, Pan, Chuanfang Jing, et Xiaorong Zhu. 2023. « GraphSAGE-Based Multi-Path Reliable Routing Algorithm for Wireless Mesh Networks ». *Processes* 11 (4): 4. <https://doi.org/10.3390/pr11041255>.
- Luo, Xuanshu, et Nirvana Meratnia. 2022. « A Geometric Deep Learning Framework for Accurate Indoor Localization ». *2022 IEEE 12th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, septembre, 1-8. <https://doi.org/10.1109/IPIN54987.2022.9918118>.
- Lv, Taizhi, Peiyi Tang, et Juan Zhang. 2023. « A Real-Time AIS Data Cleaning and Indicator Analysis Algorithm Based on Stream Computing ». *Scientific Programming* 2023 (1): 8345603. <https://doi.org/10.1155/2023/8345603>.
- Marey, Ahmed, Mustafa Bal, Hasan F. Ates, et Bahadir K. Gunturk. 2022. « PL-GAN: Path Loss Prediction Using Generative Adversarial Networks ». *IEEE Access* 10: 90474-80. <https://doi.org/10.1109/ACCESS.2022.3201643>.
- Mazzarella, Fabio, Alfredo Alessandrini, Harm Greidanus, et al. 2013. « Data Fusion for Wide-Area Maritime Surveillance ». juin.
- Mazzarella, Fabio, Michele Vespe, Alfredo Alessandrini, Dario Tarchi, Giuseppe Aulicino, et Antonio Vollero. 2017. « A novel anomaly detection approach to identify intentional AIS on-off switching ». *Expert Systems with Applications* 78: 110-23. <https://doi.org/10.1016/j.eswa.2017.02.011>.
- Mazzarella, Fabio, Michele Vespe, Dario Tarchi, Giuseppe Aulicino, et Antonio Vollero. 2016. « AIS reception characterisation for AIS on/off anomaly detection ». *2016 19th International Conference on Information Fusion (FUSION)*, 1867-73.
- Mentes, Sibel, et Zerefsan Kaymaz. 2007. « Investigation of Surface Duct Conditions over Istanbul, Turkey ». *Journal of Applied Meteorology and Climatology* (Boston MA, USA) 46 (3): 318-37. <https://doi.org/10.1175/JAM2452.1>.

- Miller, Tim. 2019. « Explanation in artificial intelligence: Insights from the social sciences ». *Artificial Intelligence* 267: 1-38. <https://doi.org/10.1016/j.artint.2018.07.007>.
- Mladenović, Jelena, Aleksandar Nešković, et Nataša Nešković. 2022. « An Overview of Propagation Models Based on Deep Learning Techniques ». *International Journal of Electrical Engineering and Computing* 6 (1): 1. <https://doi.org/10.7251/10.7251/IJEEC2206018M>.
- Mohammed, Sedir, Lukas Budach, Moritz Feuerpfel, et al. 2025. « The Effects of Data Quality on Machine Learning Performance on Tabular Data ». *Information Systems* 132 (juillet): 102549. <https://doi.org/10.1016/j.is.2025.102549>.
- Molina, Maria Camila, Iness Ahriz, Lounis Zerioul, et Michel Terré. 2025. « Multi-Task Learning for Joint Indoor Localization and Blind Channel Estimation in OFDM Systems ». *Sensors (Basel, Switzerland)* 25 (13): 4095. <https://doi.org/10.3390/s25134095>.
- Moreta, Carla E. Garcia, Mario R. Camana Acosta, et Insoo Koo. 2019. « Prediction of Digital Terrestrial Television Coverage Using Machine Learning Regression ». *IEEE Transactions on Broadcasting* 65 (4): 702-12. <https://doi.org/10.1109/TBC.2019.2901409>.
- Mufti, Naveed. 2012. « Investigation into the Effects of the Troposphere on VHF and UHF Radio Propagation and Interference between Co-frequency Fixed Links ». PhD Thesis, University of Leicester.
- Müllner, Daniel. 2011. « Modern hierarchical, agglomerative clustering algorithms ». arXiv. <https://doi.org/10.48550/ARXIV.1109.2378>.
- Nambiar, Athira, et Divyansh Mundra. 2022. « An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management ». *Big Data and Cognitive Computing* 6 (4): 132. <https://doi.org/10.3390/bdcc6040132>.
- NASA JPL. 2013. « NASA Shuttle Radar Topography Mission Global 1 arc second ». NASA Land Processes Distributed Active Archive Center. <https://doi.org/10.5067/MEASURES/SRTM/SRTMGL1.003>.
- Nguyen, Duc Tam, Chaithanya Kumar Mummadi, Thi Phuong Nhung Ngo, Thi Hoai Phuong Nguyen, Laura Beggel, et Thomas Brox. 2020. « SELF: Learning to Filter Noisy Labels with Self-Ensembling ». *International Conference on Learning Representations*. <https://openreview.net/forum?id=HkgsPhNYPS>.
- Ni, Qingjian, Yuhui Wang, et Yifei Fang. 2022. « GE-STDGN: a novel spatio-temporal weather prediction model based on graph evolution ». *Applied Intelligence* 52 (7): 7638-52. <https://doi.org/10.1007/s10489-021-02824-2>.
- NOAA/NCEP. 2015. « NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive ». Avec University Corporation For Atmospheric Research (UCAR):NSF National Center for Atmospheric Research (NSF NCAR):Computational and Information Systems Laboratory (CISL):Information Services Division (ISD):Data Engineering and Curation Section (DECS). NSF National Center for Atmospheric Research. WMO GRIB2, 574.507 Tbytes. <https://doi.org/10.5065/D65D8PWK>.

- Nupen, Wilhelm. 1959. *Checklist of references to literature on tropospheric propagation of UHF, VHF, and SHF radio waves (1929-1959): supplement #1 to NBS report 6001*. National Bureau of Standards. <https://doi.org/10.6028/nbs.rpt.6065>.
- Okumura, Y. 1968. « Field strength and its variability in VHF and UHF land-mobile radio service ». *Rev. Electr. Commun. Lab.* 16: 825-73.
- O'Shea, Timothy J., Nathan West, Matthew Vondal, et T. Charles Clancy. 2017. « Semi-supervised radio signal identification ». *2017 19th International Conference on Advanced Communication Technology (ICACT)*, février, 33-38. <https://doi.org/10.23919/ICACT.2017.7890052>.
- Ozgun, Ozlem, Volkan Sahin, Muhsin Eren Erguden, et al. 2020. « PETOOL v2.0: Parabolic Equation Toolbox with evaporation duct models and real environment data ». *Computer Physics Communications* 256: 107454. <https://doi.org/10.1016/j.cpc.2020.107454>.
- Patro, S. Gopal Krishna, et Kishore Kumar Sahu. 2015. « Normalization: A Preprocessing Stage ». arXiv. <https://doi.org/10.48550/ARXIV.1503.06462>.
- Patterson, W. L. 2007. « Advanced Refractive Effects Prediction System (AREPS) ». *2007 IEEE Radar Conference*, 891-95. <https://doi.org/10.1109/RADAR.2007.374337>.
- Pearson, Karl. 1901. « LIII. On lines and planes of closest fit to systems of points in space ». *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2 (11): 559-72. <https://doi.org/10.1080/14786440109462720>.
- Pedersen, Troels, Gerhard Steinbock, et Bernard H. Fleury. 2012. « Modeling of Reverberant Radio Channels Using Propagation Graphs ». *IEEE Transactions on Antennas and Propagation* 60 (12): 5978-88. <https://doi.org/10.1109/TAP.2012.2214192>.
- Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, et al. 2011. « Scikit-learn: Machine Learning in Python ». *Journal of Machine Learning Research* 12: 2825-30.
- Pohontu, Alexandru, Andra-Teodora Nedelcu, Nicolae-Silviu Popa, et Constantin Vertan. 2024. « Predicting Detection Range of Maritime Sensors: a Meteorological Data-Driven Approach ». *2024 16th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 1-6. <https://doi.org/10.1109/ECAI61503.2024.10607396>.
- Qiao, Lishan, Limei Zhang, Songcan Chen, et Dinggang Shen. 2018. « Data-driven graph construction and graph learning: A review ». *Neurocomputing* 312: 336-51. <https://doi.org/10.1016/j.neucom.2018.05.084>.
- Qinjiang, Rao, Gong Zhonglin, et Xu Chenghe. 1998. « Diffraction loss for radiowave propagation over an island in the presence of rough sea surface ». *ICMMT'98. 1998 International Conference on Microwave and Millimeter Wave Technology. Proceedings (Cat. No.98EX106)*, 1024-27. <https://doi.org/10.1109/ICMMT.1998.768464>.
- Quinlan, J. R. 1986. « Induction of decision trees ». *Machine Learning* 1 (1): 81-106. <https://doi.org/10.1007/bf00116251>.
- Ranasinghe, Vismika, Nandana Rajatheva, et Matti Latva-aho. 2021. « Graph Neural Network Based Access Point Selection for Cell-Free Massive MIMO Systems ». *2021 IEEE Global*

Communications Conference (GLOBECOM), décembre, 01-06.
<https://doi.org/10.1109/GLOBECOM46510.2021.9685221>.

- Rappaport, Theodore S. 2024. *Wireless Communications: Principles and Practice*. Cambridge University Press. <https://doi.org/10.1017/9781009489843>.
- Rautiainen, L., M. Johansson, M. Lensu, et al. 2025. « Studying anomalous propagation over marine areas using an experimental AIS receiver set-up ». *EGUsphere* 2025: 1-30. <https://doi.org/10.5194/egusphere-2025-1790>.
- Ravat, Franck, et Yan Zhao. 2019. « Data Lakes: Trends and Perspectives ». In *Database and Expert Systems Applications*, édité par Sven Hartmann, Josef Küng, Sharma Chakravarthy, Gabriele Anderst-Kotsis, A Min Tjoa, et Ismail Khalil, vol. 11706. Lecture Notes in Computer Science. Springer International Publishing. https://doi.org/10.1007/978-3-030-27615-7_23.
- Redondi, Alessandro Enrico Cesare. 2018. « Radio Map Interpolation Using Graph Signal Processing ». *IEEE Communications Letters* 22 (1): 153-56. <https://doi.org/10.1109/LCOMM.2017.2762318>.
- Renaud, Ambroise, Clément Iphar, et Aldo Napoli. 2025. « Attention-Aware Graph Neural Network Modeling for AIS Reception Area Prediction ». *Sensors* 25 (19). <https://doi.org/10.3390/s25196259>.
- République française. 2012. « Arrêté du 8 mars 2012 portant création par l'état-major de la marine d'un traitement automatisé de données à caractère personnel relatif à la surveillance des espaces maritimes ». <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000025545297/>.
- République française, Premier ministre. 2019. « Stratégie nationale de sûreté des espaces maritimes ». https://www.info.gouv.fr/upload/media/organization/0001/01/sites_default_files_contenu_pie ce-jointe_2019_12_snssem_2019_finale.pdf.
- Ribeiro, Claudio V., Aline Paes, et Daniel de Oliveira. 2023. « AIS-based maritime anomaly traffic detection: A review ». *Expert Systems with Applications* 231: 120561. <https://doi.org/10.1016/j.eswa.2023.120561>.
- Riveiro, Maria, Giuliana Pallotta, et Michele Vespe. 2018. « Maritime anomaly detection: A review ». *WIREs Data Mining and Knowledge Discovery* 8 (5): e1266. <https://doi.org/10.1002/widm.1266>.
- Robitaille, Marie-Claire. 2020. « Maritime Piracy and International Trade ». *Defence and Peace Economics* 31 (8): 957-74. <https://doi.org/10.1080/10242694.2019.1627511>.
- Salmon, Loic, Cyril Ray, et Christophe Claramunt. 2016. « Continuous detection of black holes for moving objects at sea ». octobre, 1-10. <https://doi.org/10.1145/3003421.3003423>.
- Santra, Avik, Pu Wang, George Shaker, et al. 2025. « Machine Learning-Powered Radio Frequency Sensing: A Review ». *IEEE Sensors Journal* 25 (13): 23164-83. <https://doi.org/10.1109/JSEN.2025.3547673>.
- Sasaki, Motoharu, Nobuaki Kuno, Toshiro Nakahira, Minoru Inomata, Wataru Yamada, et Takatsune Moriyama. 2021. « Deep Learning Based Channel Prediction at 2–26 GHz Band using Long

- Short-Term Memory Network ». *2021 15th European Conference on Antennas and Propagation (EuCAP)*, mars, 1-5. <https://doi.org/10.23919/EuCAP51087.2021.9411206>.
- Saunders, Simon R., et Alejandro A. Aragón-Zavala. 2024. *Antennas and Propagation for Wireless Communication Systems*. 3^e éd. John Wiley & Sons.
- Scarselli, Franco, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, et Gabriele Monfardini. 2009. « The Graph Neural Network Model ». *IEEE Transactions on Neural Networks* 20 (1): 61-80. <https://doi.org/10.1109/TNN.2008.2005605>.
- Schäfer, Matthias, Martin Strohmeier, Vincent Lenders, Ivan Martinovic, et Matthias Wilhelm. 2014. « Bringing up OpenSky: A large-scale ADS-B sensor network for research ». *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, 83-94. <https://doi.org/10.1109/IPSN.2014.6846743>.
- Schelleng, J.C., C.R. Burrows, et E.B. Ferrell. 1933. « Ultra-Short-Wave Propagation ». *Proceedings of the Institute of Radio Engineers* 21 (3): 427-63. <https://doi.org/10.1109/JRPROC.1933.227639>.
- Seretis, Aristeidis, Takahiro Hashimoto, et Costas D. Sarris. 2021. « A Generative Adversarial Network Approach for Indoor Propagation Modeling with Ray-Tracing ». *2021 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI)*, décembre, 657-58. <https://doi.org/10.1109/APS/URSI47566.2021.9704264>.
- Seretis, Aristeidis, et Costas D. Sarris. 2021. « An Overview of Machine Learning Techniques for Radiowave Propagation Modeling ». *IEEE Transactions on Antennas and Propagation*, 1-1. <https://doi.org/10.1109/TAP.2021.3098616>.
- Sharma, Purnima, et R K Singh. 2010. « Comparative Analysis of Propagation Path loss Models with Field Measured Data ». *International Journal of Engineering Science and Technology* 2 (juin).
- Shumate, Sidney E. 2010. « Longley-Rice and ITU-P.1546 Combined: A New International Terrain-Specific Propagation Model ». *2010 IEEE 72nd Vehicular Technology Conference - Fall*, 1-5. <https://doi.org/10.1109/VETEFCF.2010.5594342>.
- Singh, Sandeep Kumar, et Frank Heymann. 2020. « Machine Learning-Assisted Anomaly Detection in Maritime Navigation using AIS Data ». *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 832-38. <https://doi.org/10.1109/PLANS46316.2020.9109806>.
- Sirkova, I., et M. Mikhalev and. 2006. « Parabolic Wave Equation Method Applied to the Tropospheric Ducting Propagation Problem: A Survey ». *Electromagnetics* 26 (2): 155-73. <https://doi.org/10.1080/02726340500486484>.
- Smith, D. C., et D. J. Nelson. 2009. « A comparison of two methods for demodulating a target AIS signal through a collision with an interfering AIS signal ». In *Mathematics for Signal and Information Processing*, édité par Franklin T. Luk, Mark S. Schmalz, Gerhard X. Ritter, Junior Barrera, et Jaakko T. Astola, vol. 7444. SPIE. <https://doi.org/10.1117/12.825383>.
- Smith, Ernest K., et Stanley Weintraub. 1953. « The Constants in the Equation for Atmospheric Refractive Index at Radio Frequencies ». *Proceedings of the IRE* 41 (8): 1035-37. <https://doi.org/10.1109/JRPROC.1953.274297>.

- Snoek, Cees G. M., Marcel Worring, et Arnold W. M. Smeulders. 2005. « Early versus late fusion in semantic video analysis ». *Proceedings of the 13th annual ACM international conference on Multimedia* (New York, NY, USA), MULTIMEDIA '05, novembre 6, 399-402. <https://doi.org/10.1145/1101149.1101236>.
- Song, Juyoung, Duk-jin Kim, et Ki-mook Kang. 2020. « Automated Procurement of Training Data for Machine Learning Algorithm on Ship Detection Using AIS Information ». *Remote Sensing* 12 (9). <https://doi.org/10.3390/rs12091443>.
- Sotiroidis, Sotirios P., Panagiotis Sarigiannidis, Sotirios K. Goudos, et Katherine Siakavara. 2021. « Fusing Diverse Input Modalities for Path Loss Prediction: A Deep Learning Approach ». *IEEE Access* 9: 30441-51. <https://doi.org/10.1109/ACCESS.2021.3059589>.
- Stephansen, Erik T. 1981. « Clear-air propagation on line-of-sight radio paths: A review ». *Radio Science* 16 (5): 609-29. <https://doi.org/10.1029/RS016i005p00609>.
- Stevanovic, Ivica. s. d. *Python Implementation of Recommendation ITU-R P.2001*. Version 4.0. <https://github.com/eeveetza/Py2001>.
- Strohbehm, J.W. 1968. « Line-of-sight wave propagation through the turbulent atmosphere ». *Proceedings of the IEEE* 56 (8): 1301-18. <https://doi.org/10.1109/PROC.1968.6572>.
- Sun, Qiushi, Zhou Fang, Yin Li, et Ovanes Petrosian. 2025. « Jumping Knowledge Graph Attention Network for Resource Allocation in Wireless Cellular System ». *Scientific Reports* 15 (1): 17459. <https://doi.org/10.1038/s41598-025-00603-4>.
- Suto, Katsuya, Shinsuke Bannai, Koya Sato, et Takeo Fujii. 2023. « Propagation Graph Representation Learning and Its Implementation in Direct Path Representation ». *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, 1-6. <https://doi.org/10.1109/WCNC55385.2023.10118812>.
- Sutton, Richard S., et Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. Second. The MIT Press. <http://incompleteideas.net/book/the-book-2nd.html>.
- Tailor, Shyam A., Felix Opolka, Pietro Lio, et Nicholas Donald Lane. 2022. « Adaptive Filters for Low-Latency and Memory-Efficient Graph Neural Networks ». *International Conference on Learning Representations*. https://openreview.net/forum?id=hl9ePdHO4_s.
- Tang, Wenlong, Hao Cha, Min Wei, et Bin Tian. 2019. « The effect of atmospheric ducts on the propagation of AIS signals ». *Australian Journal of Electrical and Electronics Engineering* 16 (2): 111-16. <https://doi.org/10.1080/1448837X.2019.1622491>.
- Tang, Wenlong, Hao Cha, Min Wei, Bin Tian, et Yuefang Li. 2018. « A Study on the Propagation Characteristics of AIS Signals in the Evaporation Duct Environment ». *2018 International Applied Computational Electromagnetics Society Symposium - China (ACES)*, juillet, 1-2. <https://doi.org/10.23919/ACCESS.2018.8669309>.
- Tang, Wenlong, Hao Cha, Min Wei, Bin Tian, et Xichuang Ren. 2019. « Atmospheric refractivity estimation from AIS signal power using the quantum-behaved particle swarm optimization algorithm ». *Open Geosciences* 11 (1): 542-48. <https://doi.org/doi:10.1515/geo-2019-0044>.

- Tarvainen, Antti, et Harri Valpola. 2017. « Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results ». *Advances in Neural Information Processing Systems* 30. <https://proceedings.neurips.cc/paper/2017/hash/68053af2923e00204c3ca7c6a3150cf7-Abstract.html>.
- Tekbıyık, Kürşat, Güneş Karabulut Kurt, Chongwen Huang, Ali Rıza Ekti, et Halim Yanikomeroglu. 2021. « Channel Estimation for Full-Duplex RIS-assisted HAPS Backhauling with Graph Attention Networks ». *ICC 2021 - IEEE International Conference on Communications*, juin, 1-6. <https://doi.org/10.1109/ICC42927.2021.9500697>.
- Thrane, Jakob, Darko Zibar, et Henrik Lehrmann Christiansen. 2020. « Model-Aided Deep Learning Method for Path Loss Prediction in Mobile Communication Systems at 2.6 GHz ». *IEEE Access* 8: 7925-36. <https://doi.org/10.1109/ACCESS.2020.2964103>.
- « TimescaleDB: SQL made scalable for time-series data ». 2017. <https://api.semanticscholar.org/CorpusID:34446750>.
- Tu, Enmei, Guanghao Zhang, Lily Rachmawati, Eshan Rajabally, et Guang-Bin Huang. 2018. « Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey From Data to Methodology ». *IEEE Transactions on Intelligent Transportation Systems* 19 (5): 1559-82. <https://doi.org/10.1109/TITS.2017.2724551>.
- United Nations, General Assembly. 2017. *Oceans and the law of the sea: Report of the Secretary-General. Addendum. A/72/70/Add.1. United Nations*. <https://digitallibrary.un.org/record/1305829?ln=fr&v=pdf>.
- Valčić, Sanjin, et David Brčić. 2023. « On Detection of Anomalous VHF Propagation over the Adriatic Sea Utilising a Software-Defined Automatic Identification System Receiver ». *Journal of Marine Science and Engineering* 11 (6). <https://doi.org/10.3390/jmse11061170>.
- Vanche, Z., A. Renaud, et A. Napoli. 2022. « Predicting AIS reception using tropospheric propagation forecast and machine learning ». *2022 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)*, 70-71. <https://doi.org/10.23919/USNC-URSI52669.2022.9887465>.
- Varga, Ákos János, et Hajnalka Breuer. 2022. « Evaluation of convective parameters derived from pressure level and native ERA5 data and different resolution WRF climate simulations over Central Europe ». *Climate Dynamics* 58 (5): 1569-85. <https://doi.org/10.1007/s00382-021-05979-3>.
- Vasudevan, Manjuladevi, et Murat Yuksel. 2024. « Machine Learning for Radio Propagation Modeling: A Comprehensive Survey ». *IEEE Open Journal of the Communications Society* 5: 5123-53. <https://doi.org/10.1109/OJCOMS.2024.3446457>.
- Vatin, Gabriel. 2014. « Formalisation d'un environnement d'aide à l'analyse géovisuelle ». Theses, Ecole Nationale Supérieure des Mines de Paris ; CRC - Centre de recherche sur les Risques et les Crises. <https://theses.hal.science/tel-01253688>.
- Veličković, Petar, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, et Yoshua Bengio. 2018. « Graph Attention Networks ». *International Conference on Learning Representations*. <https://openreview.net/forum?id=rJXMpikCZ>.

- Vergos, George, Sotirios P. Sotiroidis, Georgia Athanasiadou, George V. Tsoulos, et Sotirios K. Goudos. 2021. « Comparing Machine Learning Methods for Air-to-Ground Path Loss Prediction ». *2021 10th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, juillet, 1-4. <https://doi.org/10.1109/MOCAST52088.2021.9493374>.
- Vince, Joanna, Britta Denise Hardesty, et Chris Wilcox. 2021. « Progress and challenges in eliminating illegal fishing ». *Fish and Fisheries* 22 (3): 518-31. <https://doi.org/10.1111/faf.12532>.
- Vinyals, Oriol, Samy Bengio, et Manjunath Kudlur. 2016. « Order Matters: Sequence to sequence for sets ». *ICLR (Poster)*. <http://arxiv.org/abs/1511.06391>.
- Wahl, Terje, Gudrun K. Høye, Aleksander Lyngvi, et Bjørn T. Narheim. 2005. « New possible roles of small satellites in maritime surveillance ». *Acta Astronautica* 56 (1): 273-77. <https://doi.org/10.1016/j.actaastro.2004.09.025>.
- Wang, Rongrong, Haiyong Luo, Qu Wang, Zhaohui Li, Fang Zhao, et Jingyu Huang. 2020. « A Spatial–Temporal Positioning Algorithm Using Residual Network and LSTM ». *IEEE Transactions on Instrumentation and Measurement* 69 (11): 9251-61. <https://doi.org/10.1109/TIM.2020.2998645>.
- Wang, Shaoxuan, et Ramon Ferrús. 2021. « Extracting Cell Patterns From High-Dimensional Radio Network Performance Datasets Using Self-Organizing Maps and K-Means Clustering ». *IEEE Access* 9: 42045-58. <https://doi.org/10.1109/ACCESS.2021.3065820>.
- Wang, Weidong, Cheng Luo, Jiancheng An, Lu Gan, Hongshu Liao, et Chau Yuen. 2024. « Semisupervised RF Fingerprinting With Consistency-Based Regularization ». *IEEE Internet of Things Journal* 11 (5): 8624-36. <https://doi.org/10.1109/JIOT.2023.3320975>.
- Wang, Xiaoye, et Shufang Zhang. 2017. « Evaluation of multipath signal loss for AIS signals transmitted on the sea surface ». *Ocean Engineering* 146: 9-20. <https://doi.org/10.1016/j.oceaneng.2017.09.022>.
- Wang, Xiaoye, Shufang Zhang, et Xiaowen Sun. 2018. « Research on the Additional Secondary Phase Factor for Automatic Identification System Signals Transmitted over a Rough Sea Surface ». *Sensors* 18 (2): 617. <https://doi.org/10.3390/s18020617>.
- Wang, Xuyu, Lingjun Gao, et Shiwen Mao. 2015. « PhaseFi: Phase Fingerprinting for Indoor Localization with a Deep Learning Approach ». *2015 IEEE Global Communications Conference (GLOBECOM)*, 1-6. <https://doi.org/10.1109/GLOCOM.2015.7417517>.
- Wen, Jinxiao, Yan Zhang, Guanshu Yang, Zunwen He, et Wancheng Zhang. 2019. « Path Loss Prediction Based on Machine Learning Methods for Aircraft Cabin Environments ». *IEEE Access* 7: 159251-61. <https://doi.org/10.1109/ACCESS.2019.2950634>.
- Werff, SE van der, Mark van Koningsveld, et Fedor Baart. 2024. *Vessel Behaviour under Varying Environmental Conditions in Coastal Areas*.
- Winning Jr., T. E., F. Xie, et K. J. Nelson. 2024. « Assessing the ducting phenomenon and its potential impact on Global Navigation Satellite System (GNSS) radio occultation refractivity retrievals over the northeast Pacific Ocean using radiosondes and global reanalysis ». *Atmospheric Measurement Techniques* 17 (23): 6851-63. <https://doi.org/10.5194/amt-17-6851-2024>.

- Wolsing, Konrad, Linus Roepert, Jan Bauer, et Klaus Wehrle. 2022. « Anomaly Detection in Maritime AIS Tracks: A Review of Recent Approaches ». *Journal of Marine Science and Engineering* 10 (1). <https://doi.org/10.3390/jmse10010112>.
- Wong, Lauren J., et Alan J. Michaels. 2022. « Transfer Learning for Radio Frequency Machine Learning: A Taxonomy and Survey ». *Sensors* 22 (4): 1416. <https://doi.org/10.3390/s22041416>.
- Wu, Zonghan, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, et Philip S. Yu. 2021. « A Comprehensive Survey on Graph Neural Networks ». *IEEE Transactions on Neural Networks and Learning Systems* 32 (1): 4-24. <https://doi.org/10.1109/TNNLS.2020.2978386>.
- Xiao, Ye, Yupeng Hu, Jizhao Liu, Yi Xiao, et Qianzhen Liu. 2024. « An Adaptive Multimodal Data Vessel Trajectory Prediction Model Based on a Satellite Automatic Identification System and Environmental Data ». *Journal of Marine Science and Engineering* 12 (3): 513. <https://doi.org/10.3390/jmse12030513>.
- Xu, Keyulu, Weihua Hu, Jure Leskovec, et Stefanie Jegelka. 2019. « How Powerful are Graph Neural Networks? » *International Conference on Learning Representations*. <https://openreview.net/forum?id=ryGs6iA5Km>.
- Xu, Keyulu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi, et Stefanie Jegelka. 2018. « Representation Learning on Graphs with Jumping Knowledge Networks ». *Proceedings of the 35th International Conference on Machine Learning*, juillet 3, 5453-62. <https://proceedings.mlr.press/v80/xu18c.html>.
- Yang, Dong, Lingxiao Wu, Shuaian Wang, Haiying Jia, et Kevin X. Li. 2019. « How Big Data Enriches Maritime Research – a Critical Review of Automatic Identification System (AIS) Data Applications ». *Transport Reviews* 39 (6): 755-73. <https://doi.org/10.1080/01441647.2019.1649315>.
- Ye, Yanzhen, et Bang Wang. 2018. « RMapCS: Radio Map Construction From Crowdsourced Samples for Indoor Localization ». *IEEE Access* 6: 24224-38. <https://doi.org/10.1109/ACCESS.2018.2830415>.
- Ying, Zhitao, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, et Jure Leskovec. 2019. « GNNExplainer: Generating Explanations for Graph Neural Networks ». In *Advances in Neural Information Processing Systems*, édité par H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, et R. Garnett, vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2019/file/d80b7040b773199015de6d3b4293c8ff-Paper.pdf.
- Yu, Bing, Haoteng Yin, et Zhanxing Zhu. 2017. *Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting*. <https://doi.org/10.48550/ARXIV.1709.04875>.
- Yun, Zhengqing, et Magdy F. Iskander. 2015. « Ray Tracing for Radio Propagation Modeling: Principles and Applications ». *IEEE Access* 3: 1089-100. <https://doi.org/10.1109/ACCESS.2015.2453991>.
- Zeng, Sheng, Yuhang Ji, Fan Rong, Lei Yang, Liping Yan, et Xiang Zhao. 2025. « Meta-PL: Path Loss Prediction of LTE Networks at Sparse Measurement Areas Using Meta-Learning ». *IEEE*

Antennas and Wireless Propagation Letters 24 (5): 1045-49.
<https://doi.org/10.1109/LAWP.2024.3525093>.

- Zhang, Chiyuan, Samy Bengio, Moritz Hardt, Benjamin Recht, et Oriol Vinyals. 2017. « Understanding deep learning requires rethinking generalization ». *International Conference on Learning Representations*. <https://openreview.net/forum?id=Sy8gdB9xx>.
- Zhang, Daiyong, Jia Li, Qing Wu, Xinglong Liu, Xiumin Chu, et Wei He. 2017. « Enhance the AIS data availability by screening and interpolation ». *2017 4th International Conference on Transportation Information and Safety (ICTIS)*, 981-86.
<https://doi.org/10.1109/ICTIS.2017.8047888>.
- Zhang, Han, Yu Han, Lingxin Meng, Guan Gui, Wei Xiang, et Yun Lin. 2025. « MFFGCN: Multimodal Feature Fusion Graph Convolution Network for Radio Map Estimation With Uneven Spatial Sampling ». *IEEE Transactions on Mobile Computing*, 1-17.
<https://doi.org/10.1109/TMC.2025.3590335>.
- Zhang, Qianqian, et Walid Saad. 2021. « Semi-Supervised Learning for Channel Charting-Aided IoT Localization in Millimeter Wave Networks ». *2021 IEEE Global Communications Conference (GLOBECOM)*, décembre, 1-6. <https://doi.org/10.1109/GLOBECOM46510.2021.9685865>.
- Zhang, Shoubin, Zhimeng Li, Hongjun Wang, Zhexian Shen, et Chao Chang. 2024. « Radio Environment Map Construction Based on Graph Convolutional Network ». *2024 IEEE 24th International Conference on Communication Technology (ICCT)*, octobre, 912-16.
<https://doi.org/10.1109/ICCT62411.2024.10946585>.
- Zhang, Xiangyu, Zhengming Zhang, et Luxi Yang. 2021. « Joint User Association and Power Allocation in Heterogeneous Ultra Dense Network via Semi-Supervised Representation Learning ». arXiv:2103.15367. Prépublication, arXiv, mars 29.
<https://doi.org/10.48550/arXiv.2103.15367>.
- Zhang, Xin, Xiujun Shu, Bingwen Zhang, Jie Ren, Lizhou Zhou, et Xin Chen. 2020. « Cellular Network Radio Propagation Modeling with Deep Convolutional Neural Networks ». *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (New York, NY, USA), KDD '20*, août 20, 2378-86.
<https://doi.org/10.1145/3394486.3403287>.
- Zhang, Yan, Jinxiao Wen, Guanshu Yang, Zunwen He, et Jing Wang. 2019. « Path Loss Prediction Based on Machine Learning: Principle, Method, and Data Expansion ». *Applied Sciences* 9 (9): 9. <https://doi.org/10.3390/app9091908>.
- Zhang, Yonglin, Changsheng Cai, Biyan Chen, et Wujiao Dai. 2019. « Consistency evaluation of precipitable water vapor derived from ERA5, ERA-Interim, GNSS, and radiosondes over China ». *Radio Science* 54 (7): 561-71. <https://doi.org/10.1029/2018RS006789>.
- Zhang, Yuanqiang, et Weifeng Li. 2022. « Dynamic Maritime Traffic Pattern Recognition with Online Cleaning, Compression, Partition, and Clustering of AIS Data ». *Sensors* 22 (16).
<https://doi.org/10.3390/s22166307>.
- Zhang, Zhihua, Zheng Sheng, Hanqing Shi, et Zhiqiang Fan. 2016. « Inversion for Refractivity Parameters Using a Dynamic Adaptive Cuckoo Search with Crossover Operator Algorithm ».

Computational Intelligence and Neuroscience 2016 (avril): 3208724.
<https://doi.org/10.1155/2016/3208724>.

Zhao, Ling, Yujiao Song, Chao Zhang, et al. 2020. « T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction ». *IEEE Transactions on Intelligent Transportation Systems* 21 (9): 3848-58. <https://doi.org/10.1109/TITS.2019.2935152>.

Zhao, Yuwei, Jia Ren, et Xun Chi. 2013. « Maritime Mobile Channel Transmission Model Based on ITM ». avril, 378-83. <https://doi.org/10.2991/3ca-13.2013.93>.

Zhou, Jie, Ganqu Cui, Shengding Hu, et al. 2020. « Graph neural networks: A review of methods and applications ». *AI Open* 1: 57-81. <https://doi.org/10.1016/j.aiopen.2021.01.001>.

Zhou, Yang, Winnie Daamen, Tiedo Vellinga, et Serge P. Hoogendoorn. 2019. « Ship Classification Based on Ship Behavior Clustering from AIS Data ». *Ocean Engineering* 175 (mars): 176-87. <https://doi.org/10.1016/j.oceaneng.2019.02.005>.

RÉSUMÉ

La connaissance de la situation maritime, essentielle à la sécurité en mer, s'appuie sur le Système d'Identification Automatique (AIS) dont la fiabilité est conditionnée par la propagation radio dépendante de facteurs environnementaux dont les fluctuations entraînent des zones de réception hétérogènes. Les limites des approches classiques de modélisation de la propagation en milieu maritime face à la complexité et la variabilité des phénomènes de propagation radio conduisent cette thèse à proposer une méthodologie alternative.

Cette thèse présente une approche fondée sur les graphes et dirigée par les données pour modéliser et prédire les zones de réception de l'AIS à partir d'entrées environnementales et géographiques. L'approche développée construit des graphes de propagation sur des données maillées horaires et utilise une architecture profonde combinant GraphSAGE, un mécanisme de Jumping Knowledge basé sur un réseau LSTM et une couche de lecture multi-agrégation. Le modèle est entraîné sur des données AIS réelles enrichies par des informations topographiques et météorologiques, ce qui lui permet de représenter à la fois des schémas de réception locaux et à longue portée. Les évaluations expérimentales montrent qu'il obtient des F1-scores et une précision supérieurs à ceux des méthodes ITU-R P.2001 et XGBoost. Les résultats suggèrent que ce modèle peut être appliqué à la surveillance maritime et contribuer à la modélisation des phénomènes de propagation.

MOTS CLÉS

Prédiction de réception AIS, Réseaux de neurones en graphes, Propagation des ondes radio, Modélisation dirigée par les données, Propagation VHF

ABSTRACT

Maritime situational awareness, essential for safety at sea, relies on the Automatic Identification System (AIS), whose reliability is conditioned by radio propagation. This propagation depends on environmental factors, whose fluctuations lead to heterogeneous reception zones. The limitations of classical propagation modeling approaches in maritime environments, when faced with the complexity and variability of radio propagation phenomena, motivate this thesis to propose an alternative methodology.

This thesis proposes a graph-based, data-driven methodology to model and predict AIS (Automatic Identification System) reception areas from environmental and geographic inputs. The developed framework constructs spatial propagation graphs on hourly gridded data and employs a deep architecture combining GraphSAGE, LSTM-based Jumping Knowledge and a multi-aggregation readout layer. Trained on real-world AIS data enriched with terrain and meteorological features, the model captures both local and long-range reception patterns. It outperforms classical baselines, including ITU-R P.2001 and XGBoost, in F1-score and accuracy. This work illustrates the value of deep learning for maritime situational awareness and highlights the potential of data-driven approaches in modeling propagation phenomena.

KEYWORDS

AIS reception prediction, Graph neural networks, Radiowave propagation, Data-driven modeling, VHF signal propagation