COLLEGE ED MATHEMATIQUES
DOCTORAL STIC
PAYS DE LA LOIRE

université angers

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ D'ANGERS

ÉCOLE DOCTORALE N° 641
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Informatique*

Par
Nadia GHERNAOUT

Optimisation de tournées de surveillance d'un réseau de transport de gaz à l'aide de l'apprentissage automatique

Thèse présentée et soutenue à Angers, le 27 février 2026
Unité de recherche : LARIS (Laboratoire Angevin en Ingénierie des Systèmes)

Rapporteurs avant soutenance :

Nacima LABADIE	Professeur, Université de Technologie de Troyes
Fabien LEHUEDE	Professeur, IMT-Atlantique

Composition du Jury :

Attention, en cas d'absence d'un des membres du Jury le jour de la soutenance, la composition du jury doit être revue pour s'assurer qu'elle est conforme et devra être répercutée sur la couverture de thèse

Président :	Jorge E. MENDOZA	Professeur titulaire, HEC Montréal
Examinateurs :	Jorge E. MENDOZA	Professeur titulaire, HEC Montréal
	Florent BRISSAUD	Ingénieur et chercheur, NaTran
Dir. de thèse :	Christelle GUERET	Professeur, Université d'Angers
Co-dir. de thèse :	David RIVREAU	Professeur, Université Catholique de l'Ouest
Co-enc. de thèse :	Martin COUSINEAU	Professeur agrégé, HEC Montréal

Remerciements

Ce manuscrit de thèse conclut plusieurs années de travail, rendues possibles grâce au soutien de nombreuses personnes que je souhaite remercier ici.

Tout d'abord, je tiens à remercier ma directrice de thèse, Christelle Guéret, mon co-directeur de thèse David Rivreau et co-encadrant Martin Cousineau. Merci pour votre encadrement tout au long de ces années, entre Angers et Montréal, pour votre disponibilité ainsi que pour vos précieux conseils, depuis la définition du sujet jusqu'à la soutenance. Votre bienveillance et votre patience ont été essentielles à l'aboutissement de cette thèse. Un merci particulier à Martin pour son accueil à Montréal : j'ai beaucoup apprécié mon stage de recherche là bas.

Je remercie également les membres du jury, Nacima Labadie, Fabien Lehuede, Jorge E. Mendoza et Florent Brissaud pour l'intérêt qu'ils ont porté à mon travail et pour le temps qu'ils ont consacré à sa lecture et son évaluation. J'ai également beaucoup apprécié les échanges lors de ma soutenance.

Je souhaite aussi remercier mes professeurs ainsi que mes collègues du Master Data Science, en particulier Destin et Elvina, avec qui j'ai eu beaucoup de plaisir à travailler, et qui m'ont encouragée à poursuivre en doctorat. Je leur souhaite plein d'épanouissement dans leurs parcours respectifs.

Mes remerciements s'adressent également à l'ensemble des membres du laboratoire LARIS pour leur accueil chaleureux et leur bienveillance. Je souhaite beaucoup de réussite aux doctorants dans la poursuite et l'aboutissement de leur thèse, en particulier à Clément et Loïc. Je remercie également Valentin pour son stage de fin d'études réalisé à nos côtés, et lui souhaite bon courage pour la suite de son parcours en thèse.

Je souhaite remercier l'ensemble des équipes de NaTran avec lesquelles j'ai eu le plaisir de collaborer depuis mon stage au sein de l'équipe Data Science « Korriban ». Je remercie tout particulièrement cette équipe pour son accueil et pour les nombreux échanges et

conseils qui ont marqué cette première expérience. Un grand merci à Marketa pour avoir initié ce sujet de thèse et pour ses précieux conseils lors des débuts de mon encadrement. Je remercie également Jasmin, Florence, Hamza, Céline, Qui, Idir et Grégory, ainsi que Jaime, qui a repris la suite de mon encadrement.

Ma thèse s'est ensuite déroulée au sein de NaTran R&I, dans le laboratoire SPESY. J'adresse mes remerciements à Leïla et Florent pour leur accueil et la confiance qu'ils m'ont accordée tout au long de mon doctorat. Je remercie également l'ensemble du laboratoire pour leur soutien, leur bonne humeur et les nombreux moments partagés qui ont rendu cette aventure encore plus chaleureuse : Corinne, Houda, Amel, Alexandre, Pascal, Sébastien, Mohamed, François, Alaa, Sara, Yanshu, Josué, Ambre, Raphaël, Vincent, Romuald, Clara, Romain et Jérôme. Enfin, un immense merci à mon fidèle binôme de thèse Théo, avec qui nous avons su nous épauler dans les moments de doute comme dans les réussites, rendant ces années de doctorat bien plus agréables.

J'aimerais également remercier les équipes métiers pour leur aide dans le cadrage du sujet, pour l'accès aux données, ainsi que pour le temps consacré à répondre à mes nombreuses questions sur les tournées de surveillance. Merci également pour leur suivi tout au long de la thèse, notamment lors des évaluations sur données réelles, et pour la confiance qu'ils m'ont accordée, en particulier Daouda, Michel et Pascale.

Je n'oublie pas toutes les personnes rencontrées en entreprise au fil de ces années : Carine, Laïdin, Myriem, Lê Dan, Hana, Shourya, Manon, Benjamin, Nikos, Tina, Camille, Blandine, Mathis, Hippolyte, Vincent, Nicolas, Patrice, Mathias, Luan, Théo, Arthur, Saïd, ainsi que tous ceux croisés au laboratoire, à l'UCO, au CIRRELT ou lors de conférences.

Je voudrais également remercier l'ANRT pour le financement de cette thèse, qui a rendu ce travail possible.

Enfin, mes dernières pensées vont à mes parents ainsi qu'à mes deux frères Yacine et Nabil pour leur soutien indéfectible. Merci de m'avoir accompagnée et encouragée tout au long de mes études.

A toutes et à tous, merci !

Table of Contents

Introduction **13**

1 Description du cas d'usage de NaTran **15**

1.1 Présentation de l'entreprise et ses ouvrages 16

1.1.1 Les infrastructures gazières 16

1.1.2 Organisation interne de l'entreprise 20

1.2 Contexte et réglementation 22

1.2.1 Réglementation de la surveillance et maintenance du réseau de transport de gaz 22

1.2.2 Études de dangers (EDD) 23

1.2.3 Outils de gestion pour la surveillance 26

1.2.4 La gestion des déclarations travaux tiers 26

1.3 Surveillance 27

1.3.1 La surveillance réglementaire 28

1.3.2 La surveillance industrielle 29

1.3.3 Modes de surveillance 34

1.3.4 Vue d'ensemble du processus de surveillance 35

1.3.5 Limites de la méthode actuelle et analyses préliminaires 35

1.4 Conclusion 39

2 Formalisation et modélisation du problème **41**

2.1 Présentation du problème 42

2.1.1 Le problème de surveillance périodique 42

2.1.2 Périmètre et hypothèses de modélisation 43

2.1.3 Notations 45

2.2 Modèle de tournées périodiques sur les arêtes 47
2.2.1 Variables de décision 47
2.2.2 Fonction Objectif 48
2.2.3 Contraintes 48
2.3 Conclusion 52

3 État de l'art **57**
3.1 Problème de tournées sur les arcs 58
3.1.1 Problèmes de tournées sur les arcs avec un véhicule 59
3.1.2 Problèmes de tournées sur les arcs multi-véhicules avec capacité . . 61
3.1.3 Problèmes de tournées sur les arcs avec périodicité 68
3.2 Problèmes de tournées avec plusieurs modes de transport 74
3.3 Apprentissage automatique pour la recherche opérationnelle 77
3.3.1 Définitions et méthodes d'apprentissage automatique 77
3.3.2 Apprentissage automatique et recherche opérationnelle 79
3.3.3 L'apprentissage automatique pour des problèmes de tournées 83
3.4 Estimation des coûts et caractéristiques d'instances 85
3.4.1 Estimation des coûts 85
3.4.2 *Features* pour la description d'une instance 89
3.5 Limites et perspectives de la littérature existante 91

4 Approche de résolution par décomposition **95**
4.1 Principe de l'approche 96
4.2 Module de planification 98
4.3 Module de tournées 102
4.3.1 Tournées au sol 102
4.3.2 Tournées aériennes 104
4.4 Conclusion 107

5 Méthodes de construction des tournées **109**
5.1 Récupération des données 110
5.1.1 Collecte des données sources 110

5.1.2 Construction du graphe et calcul des temps de parcours 112
5.2 Analyse topologique des réseaux 115
5.3 Stratégies de réduction de la taille des instances 118
5.4 Stratégies de résolution d'instances 119
5.4.1 Méthodes de construction de la tournée géante 120
5.4.2 Méthodes de résolution de multi-tournées 122
5.5 Évaluation et comparaison des méthodes de résolution 126
5.5.1 Protocole expérimental 126
5.5.2 Résultats du modèle exact (PLNE mono-véhicule) 127
5.5.3 Comparaison globale 128
5.6 Conclusion 130

6 Modèles prédictifs pour l'estimation des coûts 133
6.1 Élaboration d'une base d'apprentissage 135
6.1.1 Simulations de tournées 135
6.1.2 Base d'évaluation 140
6.1.3 Description des *features* 140
6.2 Modèle d'apprentissage 148
6.2.1 Les modèles d'apprentissage 148
6.2.2 Critères de sélection 151
6.2.3 Sélection de *features* 152
6.3 Application et résultats pour la surveillance automobile 156
6.3.1 Composition de la base d'apprentissage 156
6.3.2 Comparaison des modèles de régression 160
6.3.3 Sélection de features 163
6.3.4 Évaluations sur des instances réelles 166
6.3.5 Résultats multi véhicules 170
6.4 Conclusion 172

7 Heuristique pour la planification des arêtes 173
7.1 Motivations pour une heuristique de planification 174

7.2 Résolution de la planification par heuristique . . . 181
7.2.1 Principe de l'approche . . . 181
7.2.2 Ordre des fréquences et ordre des portions . . . 191
7.2.3 Ordre des secteurs pour la surveillance aérienne et bimodale . . . 198
7.2.4 Ordre des modes . . . 200
7.3 Conclusion . . . 202

8 Résultats de la méthode 203
8.1 Évaluation de l'approche itérative en 2023 . . . 204
8.1.1 Configuration de la méthode . . . 204
8.1.2 Résultats de l'approche itérative . . . 205
8.1.3 Comparaison avec les tournées de surveillance effectivement réalisées 210
8.2 Évaluation de l'approche itérative en 2025 . . . 217
8.2.1 Contexte opérationnel . . . 217
8.2.2 Périmètre considéré, analyse des données . . . 218
8.2.3 Résultats et comparaison . . . 219
8.3 Conclusion . . . 223

9 Conclusion et perspectives 225
9.1 Conclusion . . . 226
9.2 Perspectives . . . 229
9.2.1 Méthode de résolution . . . 229
9.2.2 Fiabilisation des données . . . 230
9.2.3 Perspectives industrielles . . . 232

A Annexes 235
A.1 Tables de résultats dans le cas multi-tournées pour la surveillance terrestre 235
A.2 Figures et tables pour la surveillance aérienne . . . 239

Bibliographie 243

Acronymes

AMF Arrêté Multi-Fluide. 22, 23

API *Application Programming Interface*. 113, 114

ARP *Arc Routing Problem*. 58–61, 65, 68, 74–76, 92, 110, 139, 140, 145

BIH *Best Insertion Heuristic*. 121, 129, 205, 227

CARP *Capacitated Arc Routing Problem*. 46, 61–67, 71, 91, 96

CARP-DD *Capacitated Arc Routing Problem with Deadheading Demands*. 65, 66, 102, 104, 134, 152, 170

CARP-IF *Capacitated Arc Routing Problem with Intermediate Facilities*. 67, 68, 134

CARP-RP *Capacitated Arc Routing Problem with Refill Points*. 67

CLARPIF *Arc Routing Problem with Intermediate Facilities under Capacity and Length Restriction*. 68

CPP *Chinese Postman Problem*. 60, 88

CPP-TW *Chinese Postman Problem with time windows*. 60

CVRP *Capacitated Vehicle Routing Problem*. 63, 86, 90

DREAL Direction Régionale de l'Environnement, de l'Aménagement et du Logement. 20, 23

ECARP *Extented Capacitated Arc Routing Problem*. 123

EDD Études De Dangers. 23, 26–28, 32

EGIG European Gas Pipeline Incident Data Group. 24

ELS Effets Létaux Significatifs. 24, 30

GESIP Groupe d'Étude de Sécurité des Industries Pétrolières. 22–24

HCPP *Hierarchical Chinese Postman Problem*. 60, 70

IRE Effets Irréversibles. 24, 30

LASSO *Least Absolute Shrinkage and Selection Operator*. 154

MAE *Mean Absolute Error*. 151, 152, 158, 159, 162, 163, 165, 167, 170, 171, 197, 208

MAPE *Mean Absolute Percentage Error*. 151, 158, 162, 167, 170, 171, 197, 208

MBCPP *Maximum Benefit Chinese Postman Problem*. 60

MCARP-TIF *Mixed Capacitated Arc Routing Problem with Time restrictions and Intermediate Facilities*. 68

MCPP *Mixed Chinese Postman Problem*. 60

MDCARP *Multi-Depots Capacitated Arc Routing Problem*. 67

MDHVCARP *Multi-Depot Heterogeneous-Vehicle Capacitated Arc Routing Problem*. 75

MDPOCARP *Multi-Depot Periodic Open Capacitated Arc Routing Problem*. 74

MRPP *Mixed Rural Postman Problem*. 61, 123

MSE *Mean Squared Error*. 163

NNH *Nearest Neighbor Heuristic*. 120, 121, 205, 227

OCARP *Open Capacitated Arc Routing Problem*. 66

PARP *Periodic Arc Routing Problem*. 68–70, 72, 73

PCARP *Periodic Capacitated Arc Routing Problem*. 70–74, 91, 96, 107

PCARP-IS *Periodic Capacitated Arc Routing Problem with Irregular Services*. 72, 73

PEL Premiers Effets Létaux. 24, 30

PLNE Programmation Linéaire en Nombre Entiers. 42, 47, 62, 66, 73, 98, 102, 104, 107, 110, 118, 120, 121, 127, 128, 130, 134, 139, 140, 163, 174, 180, 181, 226–228

POCARP *Periodic Open Capacitated Arc Routing Problem*. 73

PRPP *Periodic Rural Postman Problem*. 69, 70

PRPP-IS *Periodic Rural Postman Problem with Irregular Services*. 70

PSM Plan de Surveillance et Maintenance. 22, 23

PVRP *Periodic Vehicle Routing Problem*. 69

RMSE *Rooted Mean Squared Error*. 151, 158, 159, 162, 163, 165, 167, 170, 171, 197, 208

ROS Regroupements Opérationnels de Surveillance. 34–37, 215, 223

RPP *Rural Postman Problem*. 60, 61, 65, 67–69, 120–122, 127, 139, 163

SIG Système d'Information Géographique. 26, 111, 112, 141, 254

TCARP *Time Capacitated Arc Routing Problem*. 66, 110

TSP *Travelling Salesman Problem*. 60, 84, 86, 87, 89

VNS *Variable Neighborhood Search*. 68

VRP *Vehicle Routing Problem*. 63, 69, 83–87, 90

WPP *Windy Postman Problem*. 60

WRPP *Windy Rural Postman Problem*. 61

INTRODUCTION

Cette thèse de doctorat traite de la modélisation et de la résolution d'un problème d'optimisation de tournées de surveillance d'un réseau de transport de gaz en France. Elle a été réalisée dans le cadre d'une thèse CIFRE en collaboration avec l'entreprise NaTran et l'Université d'Angers.

NaTran est le principal gestionnaire du réseau de transport de gaz naturel en France. Ses infrastructures comprennent notamment un vaste ensemble de canalisations enterrées, qui assurent le transit du gaz des fournisseurs vers les réseaux de distribution ou clients raccordés au réseau. A des fins de sécurité et performance, le réseau de canalisations ainsi que son environnement font l'objet d'opérations régulières de maintenance et surveillance. Parmi elles, des tournées de surveillance de l'environnement réseau sont mises en place. L'administration française impose une surveillance stricte et régulière du réseau de transport de gaz. Ces inspections permettent de suivre l'évolution de l'environnement des canalisations et de détecter toute situation susceptible de compromettre l'intégrité du réseau et, par conséquent la sécurité. L'entreprise contrôle l'absence de travaux tiers non maîtrisés, mais aussi les mouvements de terrain ou encore les effondrements de berges. La surveillance des canalisations est assurée à l'aide de moyens aériens, routiers et terrestres. Afin d'adapter la fréquence d'inspection au niveau de risque associé à chaque canalisation et aux modes de surveillance disponibles, le réseau est subdivisé en portions de canalisations distinctes. Pour chacune d'elles, un niveau de risque est calculé, déterminant ainsi une fréquence minimale de surveillance à respecter dans l'année.

L'objectif principal de cette thèse de pouvoir proposer des solutions industriellement et opérationnellement exploitables pour le cas d'usage de NaTran, à savoir la résolution du problème de planification et de tournées de surveillance sur un horizon d'un an, en prenant en compte les niveaux de risques par segment de canalisation ainsi que les contraintes liées aux différents modes de transport. Ces travaux s'inscrivent dans un cadre plus large

d'actions dédiées à l'optimisation des déplacements des équipes terrain et de l'entreprise, à la réduction des coûts opérationnels, et à l'adoption d'une démarche plus respectueuse de l'environnement.

Ce problème s'inscrit dans la famille des problèmes de tournées de véhicules sur les arcs, qui ont fait l'objet de nombreuses études dans la littérature.

Cette thèse s'organise en huit chapitres, précédés par cette introduction et suivis d'une conclusion générale. Le Chapitre 1 est dédié à la présentation du contexte industriel dans lequel s'inscrivent nos travaux. Il décrit l'entreprise, les infrastructures, ainsi que les activités de surveillance du réseau et la réglementation les encadrant. Le Chapitre 2 présente le problème de tournées périodique sur les arcs avec une flotte multimodale ainsi que les notations associées. Le Chapitre 3 dresse l'état de l'art des problèmes de tournées sur les arcs mais aussi des méthodes d'apprentissage automatique pour les problèmes de tournées de véhicules, et permet de situer nos travaux dans la littérature existante. Le Chapitre 4 présente une approche itérative en deux phases pour résoudre le problème des tournées de surveillance, en décomposant celui-ci par niveau de décision : la construction des tournées hebdomadaires, au niveau opérationnel, et la planification des tournées sur l'horizon, au niveau tactique. Le Chapitre 5 détaille des méthodes exactes et approchées utilisées pour la construction de tournées. Le Chapitre 6 introduit les modèles prédictifs pour l'estimation des coûts de surveillance, fondés sur des simulations de tournées, et décrit les jeux de données et les variables explicatives utilisés. Le Chapitre 7 présente une heuristique pour le problème de planification des tournées ainsi que la répartition des modes et montre les résultats. Le Chapitre 8 détaille le résultat des expérimentations. Enfin la conclusion propose une synthèse des travaux réalisés et ouvre sur plusieurs pistes de recherche futures, qu'elles soient d'ordre industriel ou académique.

Chapitre 1

Description du cas d'usage de NaTran

Ce chapitre présente le contexte industriel de notre étude ainsi que les motivations qui ont conduit à la réalisation de ces travaux. Nous présentons l'entreprise NaTran, ses activités et ses ouvrages, puis plus particulièrement l'activité de surveillance du réseau, en passant par la réglementation qui l'encadre. Nous concluons ce chapitre en exposant les objectifs de nos travaux.

1.1 Présentation de l'entreprise et ses ouvrages

NaTran est un leader européen du transport de gaz et un expert mondial des systèmes gaziers. En France, l'entreprise exploite plus de 32 500 km de canalisations enterrées, pour transporter le gaz des fournisseurs vers les consommateurs raccordés à son réseau : gestionnaires des distributions publiques qui desservent les communes, centrales de production d'électricité et sites industriels. En tant qu'opérateur majeur de transport de gaz à haute pression, NaTran assure une mission de service public visant à garantir la continuité d'acheminement du gaz. L'entreprise s'engage par ailleurs résolument en faveur du développement des gaz renouvelables et de la transition énergétique dans les territoires.

1.1.1 Les infrastructures gazières

NaTran possède le plus vaste réseau de transport de gaz naturel en Europe, comprenant notamment des stations de compression, des postes de détente et des canalisations. L'entreprise assure la conception, la construction, l'exploitation, la maintenance et la surveillance de l'ensemble du réseau et de ses installations annexes.

Canalisations du réseau

Le réseau de transport de NaTran, représenté dans la Figure 1.1, comprend plus de 33000 kilomètres de canalisations (en France et en Allemagne) principalement en acier faisant circuler du gaz naturel à haute pression. Ce gaz est acheminé depuis les points d'interconnexion avec les pays adjacents, sous forme liquide via les terminaux méthaniers ou depuis des sites de production de biométhane. NaTran assure en plus le raccordement

et la livraison du gaz naturel auprès des réseaux de distribution mais aussi des clients industriels directement raccordés sur le réseau de transport.

Nous distinguons deux ensembles composant le réseau de transport : le réseau **principal** ou national est composé des éléments du réseau reliant les points d'interconnexion avec les réseaux de transport adjacents, les terminaux méthaniers et les stockages. Ce réseau est composé de canalisations de diamètre important, presque toujours supérieur ou égal à 600 mm et jusqu'à 1200 mm. Une part importante de ce réseau est maillée et constitue le « cœur du réseau ». Le gaz peut y circuler dans les deux sens, selon la configuration des entrées et sorties de gaz aux interconnexions. Les réseaux **régionaux** sont composés d'éléments du réseau permettant d'acheminer le gaz depuis le réseau principal jusqu'aux clients finaux ou jusqu'aux réseaux de distribution. Ces réseaux fonctionnent pour la plupart en antenne car le gaz y circule dans un seul sens depuis le réseau principal jusqu'au client final. Ces réseaux sont composés de canalisations de diamètre presque toujours inférieur à 600 mm.

Pour des raisons fonctionnelles, le réseau est analysé selon différentes échelles de maillage. Il fait ainsi l'objet de plusieurs découpages du réseau de canalisations. La Figure 1.2 illustre les trois niveaux de maillage retenus, notamment pour les activités de surveillance du réseau.

Tout d'abord, le **tronçon** est l'unité de base de description physique du réseau. Il porte la plupart des caractéristiques de la canalisation et en définit sa géométrie. Un **segment gaz** est une portion de canalisation entre deux points susceptibles d'avoir une action sur le flux du gaz. Ces points extrémités sont des brides, des robinets, des éléments de piquage, des fonds bombés, des plaques pleines, etc. Ils sont rattachés à des installations comprises dans des emprises. La **portion de surveillance**, spécifique à l'activité de surveillance, correspond à un redécoupage du segment gaz permettant de distinguer différentes parties de celui-ci nécessitant des modalités de surveillance adaptées, par exemple en fonction des conditions de visibilité. Les portions de surveillance ne correspondent pas nécessairement à un ensemble de tronçons précisément définis.

Les dimensions de ces trois éléments sont très variables. En moyenne, un segment est découpé en deux portions de surveillance et se compose d'une quarantaine de tronçons.

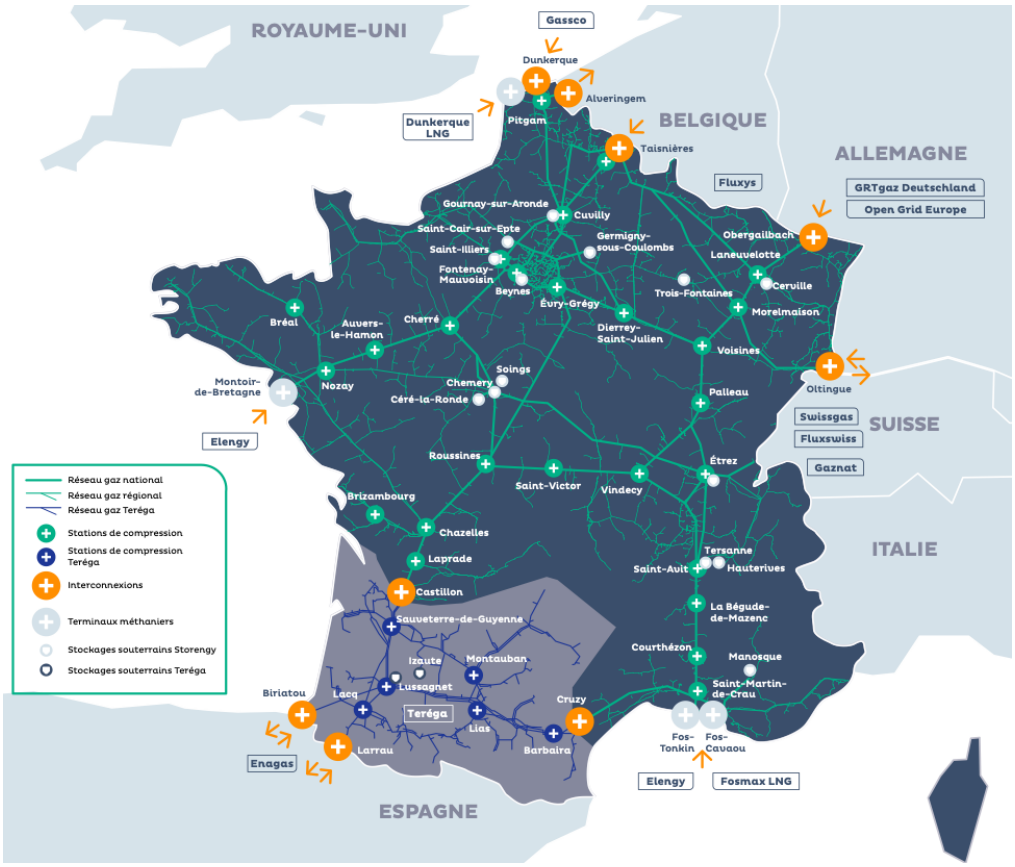


FIGURE 1.1 – Carte du réseau de transport NaTran avec stations de compression et points d'interconnexion (Source : *Rapport NaTran Plan Décennal*)

Installations annexes

En dehors des canalisations, NaTran opère sept types d'installations annexes qui permettent d'assurer la bonne circulation du gaz ainsi que la gestion des incidents et des opérations de maintenance.

Le gaz naturel transite à une vitesse moyenne de 30 km/h dans les canalisations, cependant il subit des pertes de charges à cause des frottements sur les parois des canalisations. Ainsi NaTran dispose de 26 **stations de compressions** réparties sur son réseau tous les 150 à 200 km afin d'augmenter régulièrement la pression du gaz .

Le réseau est équipé de plus de 700 **postes de pré-détente** servant à abaisser la pression entre deux sections du réseau de transport de Pression Maximum en Service (PMS) différentes.

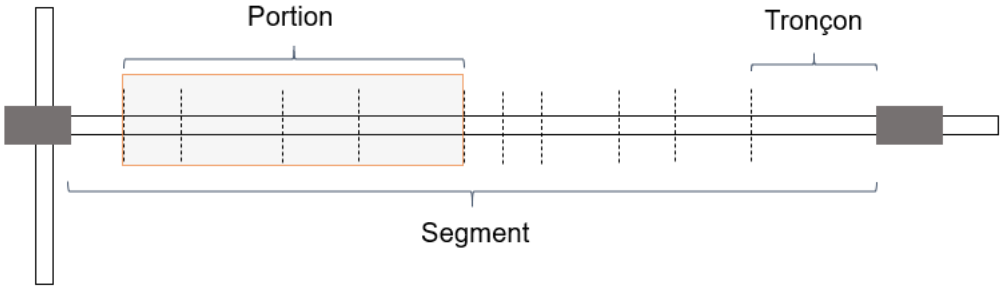


FIGURE 1.2 – Schéma de maillage du réseau de canalisations

Pour interrompre la circulation du gaz en cas d'incident ou opération de maintenance, des **postes de sectionnement** sont implantés à espace régulier sur les canalisations.

Afin de nettoyer l'intérieur des canalisations ou détecter d'éventuels défauts, des pistons instrumentés sont insérés dans les canalisations. Les **postes de coupure** (Figure 1.3a) permettent l'insertion et la réception de ces pistons de nettoyage ou d'inspection.

Par ailleurs, un nombre croissant de **postes d'injection de biométhane** [1] (Figure 1.3b) sont implantés sur le réseau. Ils permettent d'injecter directement dans les réseaux de transport ou de distribution du biométhane produit localement et à la bonne pression. Ces postes permettent également de vérifier la qualité du gaz.

L'essor du biométhane en France entraîne l'apparition de points de saturation, c'est-à-dire là où la production devient supérieure à la consommation locale. Des **postes de rebours** peuvent comprimer le gaz du réseau de distribution produit localement pour l'injecter à la bonne pression sur le réseau de transport pour qu'il puisse être envoyé dans des zones de consommation plus éloignées ou les stockages. Ce dispositif permet d'absorber les excédents de biométhane et rend ainsi le système gazier bidirectionnel.

Enfin des **postes de distribution publique** situés en aval du réseau de transport, permettent la livraison du gaz à un réseau de distribution publique. Ils se trouvent généralement à l'entrée des agglomérations. Le réseau est équipé de plus de 3400 postes de distribution.

1. Le biométhane est un gaz renouvelable produit localement à partir des déchets organiques. Il a les mêmes propriétés que le gaz naturel, ainsi il peut être injecté dans le réseau de transport ou de distribution.



(a) Poste de coupure



(b) Poste d'injection de biométhane

FIGURE 1.3 – Exemples d'installations annexes NaTran (Source : *Photothèque NaTran*)

1.1.2 Organisation interne de l'entreprise

Différentes directions [2] sont chargées de l'exploitation, l'entretien, le développement du réseau de transport ainsi que la livraison du gaz naturel aux clients industriels et distributions publiques :

La Direction des Opérations (DO) a pour rôle d'assurer en temps réel, la sécurité des personnes et des biens, la fiabilité de la disponibilité des ouvrages de NaTran, la continuité d'alimentation de l'ensemble des clients et le respect des exigences réglementaires et environnementales.

Quatre territoires ou pôles d'exploitation composent la DO. Ils sont chargés de l'exploitation, de la maintenance et de la surveillance des ouvrages sur leur territoire, ainsi que de la relation avec la Direction Régionale de l'Environnement, de l'Aménagement et du Logement (DREAL) :

— le **pôle d'exploitation Val de Seine (PEVS)** couvre les régions Île-de-France et Normandie. Bien qu'il s'agisse du territoire le plus petit en superficie, son environnement est très urbanisé et particulièrement dynamique ;

— le **pôle d'exploitation Nord Est (PENE)** regroupe des régions à forte activité

2. L'organisation décrite ici correspond à celle en vigueur avant la réorganisation de l'entreprise engagée en 2024.

industrielle telles que le Nord–Pas-de-Calais, la Picardie et l'Alsace ;

— le **pôle d'exploitation Centre Atlantique (PECA)**, comprenant la Bretagne et l'Ouest de la France, correspond majoritairement à une zone rurale ;

— le **pôle d'exploitation Rhône Méditerranée (PERM)** regroupe les régions du Sud-Est et se caractérise par un environnement marqué par le relief.

Chaque territoire est ensuite divisé en quatre départements réseau, puis en secteurs. En 2022, le réseau était divisé en 75 secteurs, chargés de l'exploitation d'un périmètre géographique du réseau de transport de gaz. Leurs principales missions consistent à garantir la sécurité, la fiabilité et la continuité du service, tout en veillant au respect des exigences environnementales. Les exploitants du secteur assurent la maintenance préventive des installations et la traçabilité des interventions, participent à la prévention des risques liés aux activités de tiers, et encadrent les travaux réalisés sur le réseau à travers la délivrance des autorisations de travail et la sécurisation des chantiers. Le secteur est également responsable de la surveillance du réseau et de son environnement, de la gestion de certaines activités internes selon l'organisation locale, et contribue à la permanence opérationnelle grâce à la participation aux astreintes d'intervention immédiate.

La Direction Technique (DT) est en charge de la politique industrielle de NaTran. Elle porte l'expertise technique, et la gestion des actifs industriels sur le plus long terme et la gestion des opérations de maintenances « spéciales ».

La DT est divisée en quatre pôles :

— le **Pôle Expertise** est chargé de la gestion optimale des actifs industriels, depuis leur phase de conception jusqu'à leur exploitation, en regroupant et en développant les compétences nécessaires. Les experts assurent un rôle de conseil, d'appui et d'expertise technique, tout en garantissant le suivi réglementaire et la veille technique.

— le **Pôle Intégrité** assure l'expertise opérationnelle nationale en matière de protection cathodique et d'inspection des ouvrages. Il intervient également dans la réparation des canalisations, la protection des ouvrages contre la corrosion, la construction des ouvrages, ainsi que dans l'expertise des dommages affectant les infrastructures.

— le **Pôle Intervention** est responsable des opérations en charge, des interventions et réparations des ouvrages, quatre départements intervention territoriaux le composent,

— le **Pôle M2TS (Machines Tournantes et Techniques Spéciales)** est en charge des activités de maintenance spécialisée sur les ouvrages de NaTran. Il assure la maintenance et le dépannage des appareils de mesure ainsi que des machines tournantes, et supervise les opérations réalisées par les prestataires de maintenance.

1.2 Contexte et réglementation

Le réseau de NaTran est soumis à de nombreux risques naturels (séismes, affaissements, inondations,...) ou liés à l'activité humaine avec les travaux tiers. Par conséquent, les activités de maintenance et surveillance sont régulées par plusieurs règles et prescriptions.

1.2.1 Réglementation de la surveillance et maintenance du réseau de transport de gaz

L'Arrêté du 5 Mars 2014 dit Arrêté Multi-Fluide (AMF) définit les modalités d'application des dispositions du code de l'environnement. Il établit un certain nombre de prescriptions minimales concernant la construction, l'exploitation, la maintenance et la surveillance d'un réseau de transport de gaz. Son objectif est de préserver la sécurité des personnes et des biens et d'assurer la protection de l'environnement. Il cite notamment les guides techniques du Groupe d'Étude de Sécurité des Industries Pétrolières (GESIP), ce groupe rassemble des industriels français et étrangers qui collaborent pour assurer la sécurité des personnes, des installations ainsi que le respect de l'environnement.

Chez NaTran, conformément à l'article 18 de l'AMF, les activités de surveillance et de maintenance des ouvrages du réseaux entrent dans le cadre du Plan de Surveillance et Maintenance (PSM). Il concerne les actes de surveillance dits industriels et non réglementaires et prévoit pour chaque type d'installations les opérations à réaliser ainsi que les fréquences de surveillance ou de maintenances associées. La mise en œuvre de ce plan est de la responsabilité de l'exploitant, il doit lui même organiser, effectuer ou faire effectuer et tracer les actes du plan de maintenance ainsi que leurs résultats. Son

programme périodique de surveillance et maintenance établit un examen complet des ouvrages et ne doit pas dépasser dix ans. Chaque pôle d'exploitation est muni d'un PSM. Sa mise en œuvre fait l'objet d'un compte rendu d'exploitation annuel (CREX) transmis à la DREAL.

1.2.2 Études de dangers (EDD)

Une Études De Dangers (EDD) est un document réglementaire destiné à identifier et analyser les risques associés à une installation industrielle, depuis sa conception jusqu'à son exploitation, afin de définir les mesures permettant de les prévenir ou d'en limiter les conséquences. Pour chaque scénario de danger, qu'il soit d'origine naturelle ou humaine, l'EDD intègre une analyse systématique du danger, une évaluation du risque fondée sur la probabilité d'occurrence et la gravité des effets, le retour d'expérience d'événements similaires sur le réseau NaTran, ainsi que la définition des mesures de maîtrise du risque applicables aux différentes phases du projet.

Ces études concernent à la fois les canalisations de gaz et les installations aériennes associées.

Analyses de risques Ces rapports répondent à des obligations réglementaires de l'AMF suivant la méthodologie du guide professionnel du GESIP et sont renouvelées tous les cinq ans afin de prendre en compte les potentielles évolutions de l'urbanisation ou du zonage sismique.

NaTran définit le **risque** comme *la possibilité de survenance d'un dommage résultant d'une exposition aux effets d'un phénomène dangereux. Dans le contexte propre à une canalisations de transport de gaz, le risque est, pour un accident donné, la combinaison de la* ***probabilité d'occurence*** *d'un évènement redouté/final considéré (incident ou accident) et la* ***gravité*** *de ses conséquences sur des éléments vulnérables.*

Ces études de risques reposent sur les analyses de bases de données recensant tous les incidents survenus sur l'ensemble des installations annexes du réseau de transport de gaz naturel depuis 1988 ainsi que toutes les fuites accidentelles survenues depuis 1970 sur les canalisations du réseau. De plus, NaTran alimente une base de données regroupant tous

incidents signalés par les transporteurs de fluides européens : l'European Gas Pipeline Incident Data Group (EGIG). Cette organisation établit des rapports d'analyses tous les trois ans qui permettent de dégager plus d'informations sur les fréquences d'incidents affectant les pipelines de gaz, mais aussi d'identifier les principales sources de dangers.

Trois scénarios de fuites représentatifs liés aux facteurs de risque sont définis dans les guides GESIP : la petite brèche (ouverture de 12 mm), la moyenne brèche (ouverture de 70 mm) ou la rupture totale de la canalisation.

Ensuite, trois bandes d'effets, exprimées en mètres, sont distinguées en fonction des conséquences potentielles sur les personnes. Il s'agit de la bande des Effets Létaux Significatifs (ELS), correspondant à la distance de part et d'autre de la canalisation dans laquelle le danger est très grave pour la vie humaine ; de la bande des Premiers Effets Létaux (PEL), définissant la zone où le danger est grave pour la vie humaine ; et enfin de la distance des Effets Irréversibles (IRE), correspondant à la zone dans laquelle le danger est considéré comme significatif pour la vie humaine.

La **gravité** dépend pour chacun de ces périmètre *du nombre de personnes susceptibles d'être présentes, en considérant que ces personnes peuvent évacuer les lieux ou se mettre à l'abri.*

Analyse des causes d'accidents Dans le rapport de 2020 de l'EGIG [3], portant sur la période 2010–2019, les agressions externes constituent la principale cause d'incidents sur les canalisations, représentant plus d'un quart des événements recensés. Elles sont suivies par la corrosion, les défauts de construction, les mouvements de terrain et d'autres causes diverses.

Le rapport 2024 de l'EGIG [4] , fondé sur l'analyse des incidents survenus entre 2013 et 2022, met en évidence, pour la première fois depuis la création de ces rapports, que la corrosion constitue désormais la principale cause de défaillance. Elle est suivie de près par les agressions externes, puis par les mouvements de terrain. Les défaillances liées à la corrosion se traduisent généralement par l'apparition de microfissures, tandis que celles résultant d'agressions externes peuvent engendrer des conséquences beaucoup plus sévères.

3. `https://www.egig.eu/reports/$60/$61`
4. `https://www.egig.eu/reports/$60/$178`

Les incidents dus aux agressions externes et aux mouvements de terrain sont en effet caractérisés par des conséquences potentiellement graves. Les rapports soulignent ainsi l'importance de mesures mises en œuvre par les exploitants de pipelines et les autorités pour prévenir ces événements. Entre 2013 et 2022, plus de 60 % des incidents ayant entraîné une perforation supérieure à 2 cm, sans toutefois atteindre le diamètre de la canalisation, sont dus à des agressions externes, contre près de 30 % liés aux mouvements de terrain. Ces derniers constituent par ailleurs la principale cause de rupture complète des canalisations, représentant plus de la moitié des cas, les interventions externes arrivant en seconde position.

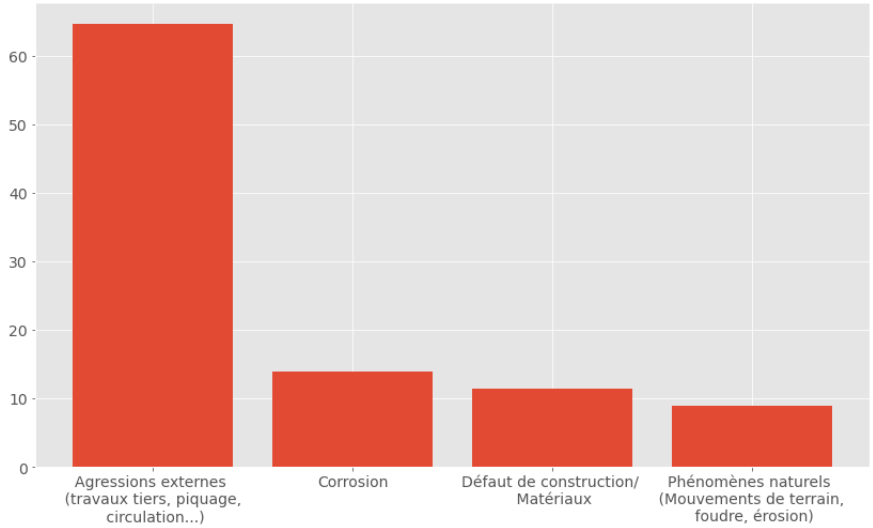


FIGURE 1.4 – Répartition des principaux facteurs de risques toutes tailles de brèches confondues sur le réseau NaTran de 1970 à 2016

Source : *Études De Dangers 2017* — `https://www.normandie.developpement-durable.gouv.fr/IMG/pdf/1-p7-1-edd_partie_generique.pdf`

La Figure 1.4 illustre la répartition des principaux facteurs de risque des incidents sur le réseau de transport de gaz en France entre 1970 et 2016. Pour les canalisations exploitées par NaTran, les dangers les plus significatifs proviennent des agressions par des engins de travaux publics, puis de la corrosion ou de fissuration, ainsi que des mouvements de terrain.

Les installations aériennes annexes, sont quant à elles peu soumises à des aléas liés à l'activité humaine.

Dispositions compensatoires et Programme Réglementaire de Traitement (PRT)

Dans certains cas les études de dangers peuvent donner lieu à la mise en place de mesures ou dispositions dites compensatoires. Ces mesures ont pour objectif de réduire la probabilité d'occurrence d'un danger et sont divisées en quatre catégories :

— l'information : sensibilisation des entreprises, collectivités et particuliers...

— l'exploitation avec la surveillance renforcée dite surveillance réglementaire et une baisse de la pression d'exploitation de la canalisation

— l'aménagement d'équipements supplémentaires (balise, dalle de protection, ...)

— et la construction d'infrastructures de protection autour des canalisations et postes

L'ensemble des mesures compensatoires déterminées par les EDD constituent un **Programme Réglementaire de Travaux (PRT)** puis un **Programme Opérationnel de Travaux (POT)**.

1.2.3 Outils de gestion pour la surveillance

Un Système d'Information Géographique (SIG) est un système informatique dédié à la gestion et à la visualisation de données géographiques géo-référencées. Dans un cadre réglementaire, il doit notamment intégrer le tracé des canalisations, la localisation des installations annexes, les zones d'effets issues des EDD, les données techniques et administratives, l'occupation des sols, les servitudes et l'historique des réparations.

Chez NaTran, le SIG *SWING* centralise ces informations et constitue un outil d'aide à la décision pour la maintenance, la surveillance du réseau, la gestion des incidents et le suivi des travaux tiers.

1.2.4 La gestion des déclarations travaux tiers

La réglementation française impose la déclaration préalable de tout projet de travaux susceptible d'affecter des réseaux. Le responsable de projet consulte d'abord le Guichet Unique national pour identifier la présence éventuelle d'ouvrages et, si nécessaire, transmet une Déclaration de Projet de Travaux (DT) aux exploitants concernés. L'exécutant des travaux réalise ensuite une Déclaration d'Intention de Commencement de Travaux (DICT),

parfois combinée avec la DT, afin de confirmer la localisation des réseaux et les mesures de sécurité.

Chez NaTran, les déclarations sont traitées par le pôle d'exploitation avant transmission aux secteurs. Conformément au Code de l'environnement, un rendez-vous sur site est organisé lors de la DICT avant tout démarrage des travaux, tandis que les réponses aux DT se font généralement par l'envoi de plans de localisation.

1.3 Surveillance

D'après des statistiques d'accidentologie, les canalisations représentent le moyen de transport le plus sûr pour les matières dangereuses tel que le gaz naturel avec un impact très limité sur le milieu naturel. Bien qu'ils soient rares, les accidents peuvent néanmoins être très graves. Ainsi, le tracé doit être surveillé périodiquement afin d'examiner l'environnement des ouvrages et vérifier leur situation et marquage.

Historiquement le réseau de transport de gaz naturel était surveillé indépendamment dans chaque secteur sans règle pré-définie. Mais depuis la mise en place de l'Arrêté Multi Fluide en 2006, NaTran a dû mettre en place une procédure avec des règles précises encadrant les activités de surveillance.

La surveillance terrain du réseau et de son environnement répondent à plusieurs objectifs :

— la détection de constructions en cours ou récemment achevées proches des infrastructures de NaTran,

— la détection de chantiers ou traces de chantiers publics ou privés,

— le contrôle du balisage (y compris les signalements renforcés issus des EDD),

— la recherche d'anomalies de couvertures (fonds de fossés, affaissements de terrain...)

— le contrôle de la servitude (accessibilité, entretien...)

L'acte de surveillance doit permettre de détecter le maximum d'anomalies sans pour autant exposer l'agent qui procède à ce constat visuel à un risque notamment lors de la surveillance routière ou pédestre.

Pendant cette phase de surveillance, en cas d'identification d'une fuite ou d'une situation dangereuse imminente pour l'intégrité de l'ouvrage, l'agent qui constatera la situation (ou l'agent alerté par le prestataire missionné par NaTran) prendra les premières mesures qu'il jugera nécessaires (arrêt des travaux tiers, mise en place d'un périmètre de sécurité, accueil des secours si besoin ...) en liaison avec le chef d'exploitation qui l'alertera.

Nous distinguons deux types de surveillance : la surveillance dite réglementaire, de celle industrielle.

1.3.1 La surveillance réglementaire

Elle découle des études de dangers, qui définissent une surveillance renforcée des ouvrages comme mesure compensatoire de sécurité. Comme indiqué dans la section précédente, les EDD peuvent en effet conduire à la mise en place de dispositions compensatoires, notamment sous la forme d'une surveillance dite réglementaire. Il s'agit d'une surveillance supplémentaire appliquée à certaines zones du réseau, appelées portions SRT, pour lesquelles les fréquences annuelles de surveillance sont définies en amont et ne peuvent être modifiées.

Certains points de l'ouvrage se distinguent du reste des tronçons enterrés et présentent des risques différents de ceux du tracé courant, ils sont appelés points singuliers.

Ces points singuliers peuvent être de natures différentes et leur surveillance est définie dans les EDD. Cette surveillance ne nécessite pas de moyen particulier ou supplémentaire par rapport aux surveillances industrielles.

Les **Zone d'instabilité de sol (ZIS)** correspondent à des zones à risque de mouvement de terrain ou d'érosion. Ces points sont surveillés lors de la surveillance générale du tracé. Ils peuvent être également visités suite à un mouvement de sol.

La surveillance des **traversées aériennes** (Figure 1.5a), également appelées zones de pose à l'air libre, est réalisée chaque année. Elle consiste dans un premier temps à examiner l'état général de la structure supportant la canalisation, puis dans un second temps à examiner l'état de la canalisation aérienne (défaut de revêtement voire métallique)

La surveillance des **traversées sous fluviales** (Figure 1.5b) consiste en un examen visuel annuel de tout tronçon de canalisation qui franchit un cours d'eau de surface (non

souterrain), comme un fleuve, une rivière ou un ruisseau, à l'exception des franchissements pour lesquels l'inspection pédestre est possible en période d'étiage.



(a) Traversée Aérienne du pont de la Drôme



(b) Traversée Sous Fluviale dans la Creuse

FIGURE 1.5 – Illustrations de points singuliers (Source : *Photothèque de NaTran*)

1.3.2 La surveillance industrielle

Elle résulte du PSM et est définie à l'aide d'une règle interne de NaTran présentée ci-dessous qui permet d'attribuer des fréquences de passages par tronçon de canalisation.

Pour planifier l'activité de surveillance de l'environnement des canalisations, une règle de hiérarchisation a été mise en place et définit quatre groupes dit « enjeux ». Ces enjeux sont liés à des fréquences minimales de passages par tronçons de canalisations.

1. Un premier groupe dit **"Enjeu limité"** correspond à des zones où la présence et l'activité humaine sont faibles. Un passage par an est prévu.
2. Un second groupe dit **"Enjeu normal"** se satisfait d'une surveillance courante. Il n'y a pas de risque particulier avec les travaux-tiers. Pour ces tronçons un passage par mois est préconisé, avec une tolérance de 2 visites en moins par an si les conditions météorologiques le justifient.
3. Un troisième groupe dit **"Enjeu Fort"**, justifie une surveillance supérieure ou égale à 2 fois par mois. De même il y a une tolérance de 4 visites en moins par an.
4. Un dernier groupe dit **"Enjeu Très Fort"** justifie d'une surveillance fortement renforcée égale à 1 fois par semaine avec une tolérance de 8 visites en moins par an.

Ces calculs sont mis à jour chaque année auprès de la DT et la Direction des Systèmes d'Information (DSI) de NaTran puis communiqués à la DO.

Plusieurs variables sont retenues pour l'algorithme de calcul des enjeux afin de définir le niveau de risque auquel un tronçon de canalisation est soumis. Un critère lié à la fréquence du risque est définit à partir de l'occurrence de travaux tiers à proximité du tronçon. Un critère de gravité est modélisé par le coefficient minimal de sécurité, et un critère de transit est donné avec le diamètre de la canalisation. Enfin la présence ou non de mesures compensatoires est prise en compte.

Le Coefficient Minimal de Sécurité (CMS) : Ce coefficient résulte des études de dangers et traduit la **gravité** d'un incident potentiel se produisant sur le tronçon en question. Une valeur A, B ou C est attribuée selon le nombre de victimes de différents scénarios d'endommagements de la canalisation : petite brèche, moyenne brèche ou rupture de la canalisation. Différentes bandes d'effets sont définies dépendant du diamètre nominal (DN) et de la Pression Maximale de Service d'Exploitation (PMS-E). Plus le diamètre et la pression augmentent plus les bandes d'effets sont larges : la bande des PEL, la bande des ELS et la bande des IRE.

Il existe un processus pour déterminer la valeur de CMS à attribuer au tronçon : il consiste à déplacer tous les cinq mètres un cercle dont le rayon est égale à la bande des PEL.

— Une valeur A correspond à un emplacement à faible présence humaine et est attribuée si moins de trente personnes résident dans le cercle et si la densité de population est inférieure à huit personnes par hectare

— Une valeur C correspond à un emplacement à présence humaine plus importante et est attribuée si plus de 299 personnes résident dans le cercle et si la densité de population est supérieure à 80 personnes par hectare.

— Une valeur B est attribuée dans les autres cas.

Le coefficient d'activité travaux tiers : Depuis le 1er juillet 2012, la réglementation impose à toute personne envisageant la réalisation de travaux, sur le domaine public

comme privé, de déposer une déclaration auprès du Guichet Unique. Plusieurs types de déclarations et de situations de travaux sont distingués : les déclarations d'intention de travaux (DT), les déclarations d'intention de commencement de travaux (DICT), les déclarations conjointes DT–DICT (DC), les travaux urgents (TU), ainsi que les chantiers en infraction découverts (CEI). Ces derniers regroupent les chantiers non déclarés ou ceux faisant l'objet de déclarations erronées, notamment lorsque les informations relatives à la localisation ou à la durée des travaux sont inexactes. Les données concernant les CEI ne sont pas totalement fiables car ce sont des données remontées en internes par les agents pendant la surveillance. Ainsi elles dépendent de la bonne saisie des informations mais aussi de la quantité de surveillance réalisée.

Le coefficient Activité Travaux Tiers a pour objectif de représenter la dynamique d'urbanisation d'une commune. Il permet de recenser les travaux et déclarations de travaux pour chaque commune. Ce coefficient représente la moyenne sur cinq ans de la densité de déclaration de travaux.

$$C_{\text{Activité TT}} = \frac{DT_c + DICT_c + DC_c + TU_c + CEI}{5 \times \text{Surface d'étude}}$$

L'indice c désigne les déclarations de travaux *concernées*, c'est-à-dire celles relatives aux canalisations de NaTran situées dans une bande de 50 mètres autour de la canalisation. La surface d'étude correspond, quant à elle, à la zone couverte par une bande de 250 mètres autour de la canalisation.

La valeur du coefficient est ensuite comparée à différents seuils S_0, S_1, S_{2a} et S_{2b} découlant d'une étude probabiliste. Cependant ces seuils sont revus à la hausse chaque année, afin de garder un prescrit de surveillance similaire dans le temps.

Le diamètre nominal de la canalisation : Le réseau de transport est composé d'artères régionales et nationales de diamètres différents allant de 80 mm à 1200 mm. Plus le diamètre de la canalisation est important plus elle supportera de plus grosses pressions d'exploitation. Le seuil pour ce paramètre est fixée à 400 mm car il équivaut au diamètre médian des canalisations et sépare les artères régionales et nationales

La présence de dispositions compensatoires : Certaines mesures compensatoires physiques issues des EDD, telles que des dalles de protection, ont pour objectif de réduire le risque associé aux travaux tiers, permettant ainsi de diminuer les enjeux. Dans le cadre de la règle de classification, nous considérons uniquement la présence ou l'absence de la disposition compensatoire appelée « Augmentation Travaux Tiers », qui protège spécifiquement les canalisations.

Règle de classification :

Dans la Table 1.1, on présente une matrice d'enjeux pour les tronçons sur lesquels il n'y a pas de mesure compensatoire. Alors que dans la Table 1.2, on présente une matrice d'enjeux pour les tronçons sur lesquels il n'y a une mesure compensatoire.

		Activité travaux tiers						
		$\leq S_0$		$< S_1$ et $> S_0$		$\geq S_1$		$> S_{2A}$ et $> S_{2B}$
		DN $\leq$ 400 mm	DN $>$ 400 mm	DN $\leq$ 400 mm	DN $>$ 400 mm	DN $\leq$ 400 mm	DN $>$ 400 mm	tous les DN
Coefficient Env. Humain	A	Limité	Limité	Normal	Normal	Normal	Fort	Très Fort
	B	Limité	Normal	Normal	Normal	Fort	Fort	Très Fort
	C	Normal	Normal	Normal	Fort	Fort	Très Fort	Très Fort

TABLE 1.1 – Règle de hiérarchisation pour les tronçons sans mesure compensatoire (Source : *Procédure interne à NaTran*)

Lissage des enjeux à la maille du segment gaz

Le calcul des enjeux de surveillance est réalisé à la maille la plus fine, celle du tronçon. Un lissage est ensuite effectué à la maille du segment selon la procédure suivante : tous les tronçons appartenant à un même segment sont d'abord sélectionnés, puis classés par ordre décroissant des fréquences des enjeux, de l'enjeu Très Fort à l'enjeu Limité. La somme

		Activité travaux tiers						
		$\leq S_0$		$< S_1$ et $> S_0$		$\geq S_1$		$> S_{2A}$ et $> S_{2B}$
		DN $\leq$ 400 mm	DN $>$ 400 mm	DN $\leq$ 400 mm	DN $>$ 400 mm	DN $\leq$ 400 mm	DN $>$ 400 mm	tous les DN
Coefficient Env. Humain	A	Limité	Limité	Limité	Normal	Normal	Normal	Fort
	B	Limité	Limité	Normal	Normal	Normal	Normal	Fort
	C	Limité	Normal	Normal	Normal	Normal	Fort	Fort

TABLE 1.2 – Règle de hiérarchisation pour les tronçons avec mesure compensatoire (Source : *Procédure interne à NaTran*)

cumulée des longueurs des tronçons ainsi classés est ensuite calculée, et dès que cette somme dépasse 40% de la longueur totale du segment, le premier enjeu correspondant est retenu. Cet enjeu est enfin attribué à l'ensemble du segment.

Ces enjeux de surveillance lissés sont les *enjeux calculés* et constituent ce que nous appelons le **prescrit**. C'est la quantité de surveillance que doivent au minimum effectuer les exploitants.

La révision des enjeux par les exploitants

Pour diverses raisons, les exploitants peuvent ajuster les enjeux de surveillance à la hausse ou à la baisse, à condition de sélectionner une justification.

La réduction de l'enjeu calculé peut être motivée par la fin de travaux tiers localisés, par un enjeu précédemment augmenté en raison de travaux NaTran, par la mise hors service définitive d'un ouvrage, ou encore par le parallélisme ou l'adjacence avec un ensemble surveillé présentant un enjeu inférieur.

À l'inverse, l'augmentation de l'enjeu calculé peut résulter d'un ajustement fondé sur l'expertise de secteur, le plus souvent lié à l'information de travaux futurs, ou par le parallélisme ou l'adjacence avec un ensemble surveillé présentant un enjeu supérieur.

Les exploitants ont également la possibilité de renseigner quatre autres types d'enjeux ou fréquences qui ne sont pas pris en compte par la règle de classification. Les deux premiers sont liés à la visibilité sur le terrain : un enjeu **Clos** concerne les parcelles closes et bâties, pour lesquelles une surveillance pédestre quinquennale est réalisée après accord du propriétaire. Un enjeu **Nvar** (non visible en surveillance aérienne ou routière) correspond à des zones qui ne peuvent pas être contrôlées par des moyens aériens ou automobiles et qui nécessitent, de ce fait, une visite pédestre annuelle, le niveau de risque étant considéré comme faible.

Les deux autres permettent d'ajuster la fréquence des visites pour combler l'écart entre les enjeux limités et normaux. Ainsi, certains enjeux limités peuvent bénéficier d'une fréquence **semestrielle**, correspondant à deux visites par an, tandis que d'autres peuvent être suivis selon une fréquence **trimestrielle**, soit quatre visites par an.

Ces modifications peuvent entraîner un redécoupage des segments, donnant lieu à la création de portions de surveillance industrielle (PSD).

Construction des tournées de surveillance Les responsables de secteurs organisent des Regroupements Opérationnels de Surveillance (ROS) qui sont les équivalents des tournées de surveillance, elles reflètent l'organisation opérationnelle de la surveillance par secteur. Ces ROS font le lien entre le prescrit et les moyens mis en place pour y répondre.

Ainsi les secteurs reçoivent les enjeux prescrits pour les portions de surveillance industrielle (PSD) et les portions de surveillance réglementaires (SRT). Un ROS regroupe plusieurs portions de surveillance, la fréquence de surveillance retenue pour l'ensemble du ROS correspond à l'enjeu le plus fort. Ensuite ils organisent un planning de tournée pour visiter ces ROS en indiquant les différents moyens de surveillance utilisés.

1.3.3 Modes de surveillance

La surveillance des canalisations est réalisée selon plusieurs modes de surveillance complémentaires. On distingue la surveillance terrestre, effectuée à pied ou en véhicule, de la surveillance aérienne, réalisée par avion ou hélicoptère. Le choix du mode de surveillance est laissé à l'exploitant local, en fonction de la configuration du réseau et des zones à

contrôler.

La surveillance aérienne couvre plus de 600 000 kilomètres par an. Les avions sont mobilisés une à deux fois par mois par portion pour le suivi des chantiers et de l'environnement, tandis que les hélicoptères interviennent une fois par an pour le contrôle du balisage. Le recours aux hélicoptères tend néanmoins à diminuer. Cette surveillance est externalisée et confiée à deux prestataires : AirLoisir pour les territoires PEVS, et RECTIMO Air Transports pour les territoires PECA, PENE et PERM. On considère qu'un avion est dédié à chaque territoire.

La surveillance routière complète la surveillance aérienne en permettant l'accès à des zones peu visibles ou inaccessibles par avion. La surveillance pédestre, quant à elle, est utilisée pour atteindre des zones non visibles par la route ou depuis les airs, ainsi que pour inspecter des points singuliers identifiés dans les EDD.

Chaque surveillance fait l'objet d'un rapport, analysé par l'exploitant, afin de déterminer les actions correctives éventuelles et leurs échéances.

1.3.4 Vue d'ensemble du processus de surveillance

Nous présentons dans la Figure 1.6 le processus de récupération et traitement des SRT et PSD pour la création des ROS, puis la comparaison avec les portions de surveillance réalisées (PSR).

La Figure 1.7 illustre le processus de traitement des enjeux, depuis leur calcul initial au niveau des tronçons jusqu'à leur intégration dans les ROS. Ces enjeux sont ensuite comparés aux enjeux réalisés, déterminés après un an en fonction du nombre effectif de surveillances effectuées.

1.3.5 Limites de la méthode actuelle et analyses préliminaires

L'objectif est d'optimiser la surveillance afin de réduire les déplacements et les coûts, tout en assurant un niveau de contrôle proportionné au risque encouru par chaque portion.

L'approche actuelle présente plusieurs limites, tant en matière de méthodologie et d'analyse des risques que dans l'organisation et la construction des tournées. Les pistes

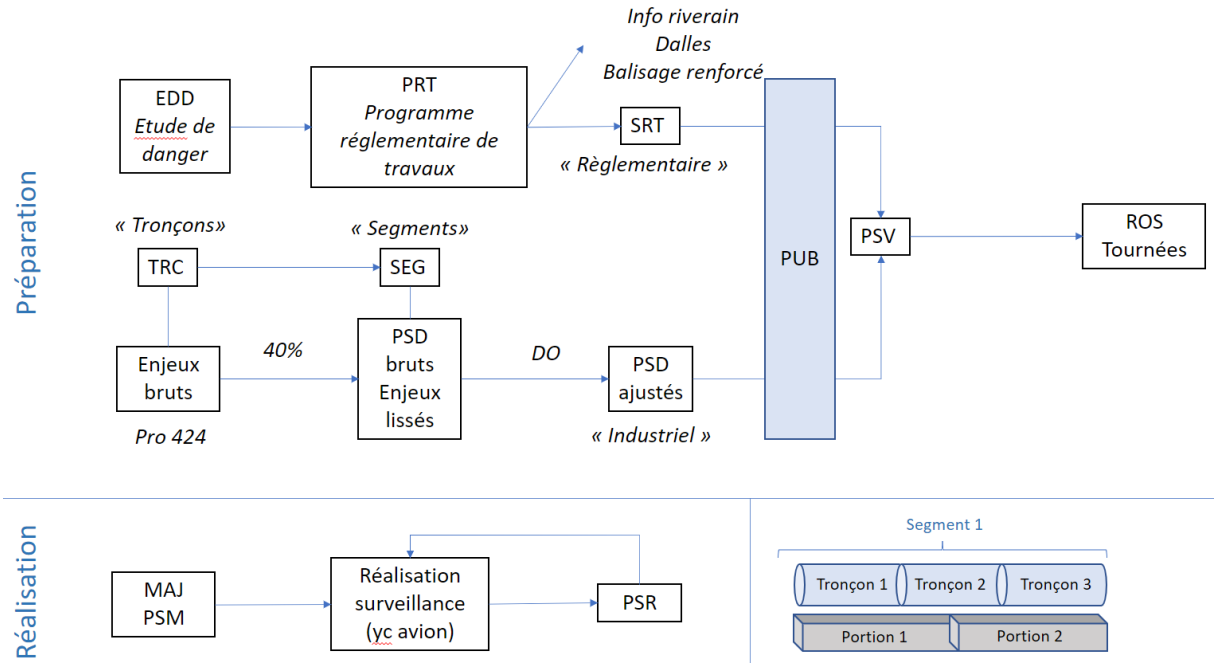


FIGURE 1.6 – Schéma du processus de construction des ROS (Source : *Document interne à NaTran*)

d'amélioration et les points d'attention identifiés sont présentés ci-dessous, à partir des entretiens menés auprès de la DT et de la DO.

La définition du prescrit et structure des ROS

On appelle **prescrit** le kilométrage de canalisations par an prévu pour la surveillance. Ce kilométrage est calculé en multipliant le nombre de passages prévu par an de la portion de surveillance par la taille de la portion.

On appelle **prévu ROS** le kilométrage de canalisations par an prévu pour la surveillance après construction des ROS. Ce kilométrage est calculé en multipliant le nombre de passages prévu par an du ROS (fréquence la plus élevée parmi les portions regroupées) par sa taille.

On appelle **réalisé** le kilométrage de canalisations par an réalisé pour la surveillance. Ce kilométrage est calculé en multipliant le nombre de passages réel par an du ROS par sa taille.

Chaque année la DT de NaTran analyse les évolutions et écarts de surveillance prescrite, prévue et réalisée. Il y a naturellement un écart entre le prescrit PSV et le prévu ROS

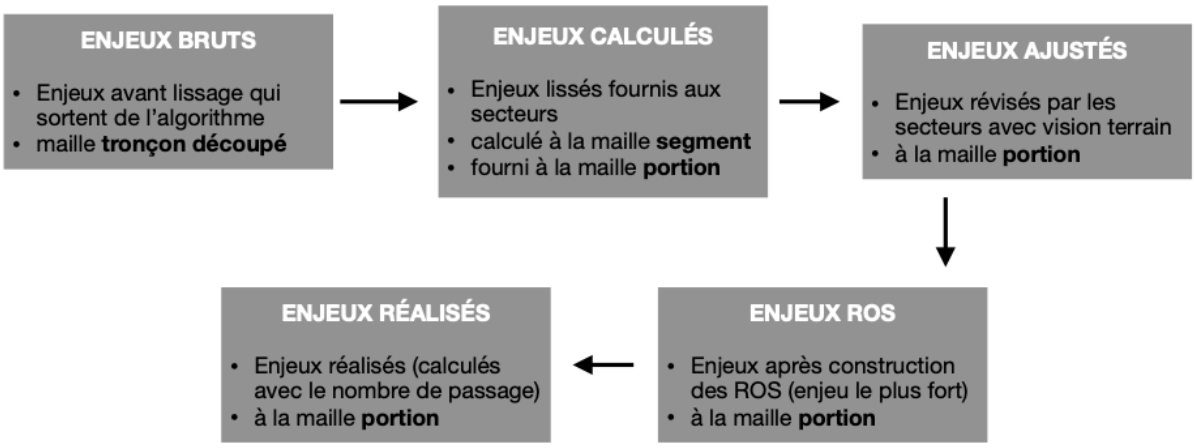


FIGURE 1.7 – Évolution des enjeux

car le prescrit est calculé à la maille portion et ne prend pas en compte les tournées de surveillance. Ainsi comme un ROS garde l'enjeu le plus fort parmi les enjeux des portions, il est normal d'avoir plus de surveillances réalisées que de surveillances prescrites. Cependant, la DT s'attache également à comprendre les écarts résultant d'une réévaluation des enjeux.

Dans la Figure 1.8 nous analysons les écarts entre prescrit et prévu ROS de 2018 à 2021. Des travaux entre la DT et DO ont permis de revoir la construction des ROS pour abaisser l'écart petit à petit entre le prévu et le prescrit.

Cependant, la seule reconstruction des ROS ne suffirait pas à optimiser pleinement la surveillance. En effet, leur caractère rigide impose qu'une portion de surveillance soit rattachée à un ROS unique, parcouru intégralement, à fréquence fixe et selon un mode prédéfini, ce qui limite la flexibilité de l'organisation.

De plus, les ROS ne constituent pas de véritables tournées opérationnelles, car ils regroupent des PSV sans intégrer systématiquement les trajets hors surveillance. Or, l'optimisation de la surveillance passe nécessairement par la prise en compte de l'ensemble des déplacements.

Enfin, les ROS sont construits manuellement par les secteurs et font l'objet d'une révision chaque année ce qui est chronophage pour l'ensemble des équipes concernées.

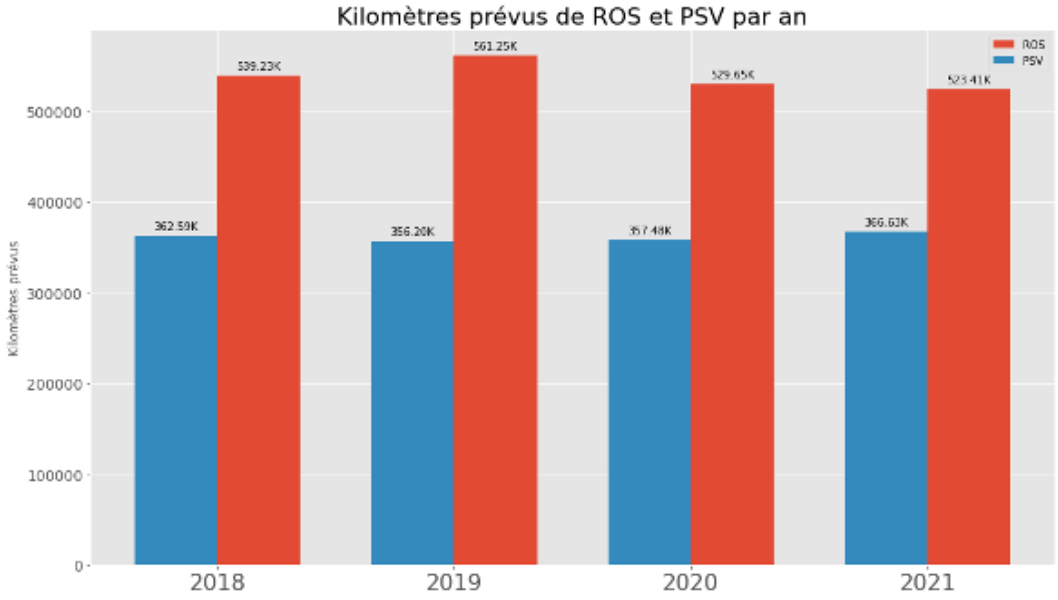


FIGURE 1.8 – Écarts par an entre le prescrit et le prévu ROS (Source : *Analyse NaTran)*

La variable liée à l'activité travaux tiers reflète peu le risque travaux

Les données actuelles sur les travaux tiers sont peu précises, car elles ne fournissent qu'un nombre de chantiers à l'échelle de la commune, sans géolocalisation exacte. Il serait nécessaire de disposer de la localisation précise des travaux et de considérer un périmètre autour des tronçons de canalisation afin de mieux refléter la dynamique d'urbanisation.

Par ailleurs, le critère repose sur une moyenne sur cinq ans, qui peut masquer les évolutions récentes de l'activité : une forte activité passée ne traduit pas nécessairement un dynamisme actuel des travaux tiers.

Certains enjeux ne sont pas pris en compte par l'algorithme

Comme indiqué dans la Section 1.3.2, certains enjeux ou fréquences peuvent être renseignés par les exploitants. Les premiers concernent le manque de visibilité du tracé pour les moyens de surveillance aériens et routiers, les seconds visent à réduire l'écart de surveillance entre les portions à enjeux limités et normaux.

Sans connaissance des zones de non-visibilité, l'algorithme surestime les fréquences de passage, obligeant les exploitants à ajuster manuellement les enjeux et à cartographier

ces zones. Cela explique en partie l'écart entre le prescrit et le réalisé, bien que ces zones soient aujourd'hui de mieux en mieux renseignées dans SWING.

Par ailleurs, l'introduction récente de fréquences intermédiaires (semestriels et trimestriels) pour l'enjeu limité permet de mieux graduer la surveillance entre les fréquences annuelle et mensuelle. Bien qu'ils ne soient pas générés par l'algorithme et non encore formalisés dans la procédure interne, ces enjeux peuvent être renseignés par les exploitants et devraient être intégrés dans la future règle des enjeux de la DT.

Mesurer l'efficacité de la surveillance

Les activités de surveillance visent à contrôler l'environnement du réseau NaTran de manière proportionnée au risque encouru par les canalisations.

Se pose dès lors la question de l'évaluation de leur efficacité afin d'adapter les fréquences aux endroits les plus exposés. Les chantiers en infraction, identifiés par les gestionnaires de réseau comme le principal risque, pourraient constituer un indicateur pertinent : les zones où ils sont les plus nombreux étant a priori les plus à risque. Toutefois, les données actuellement disponibles chez NaTran ne sont pas suffisamment fiables pour dégager des tendances claires. La détection des infractions n'est pas systématiquement remontée par tous les secteurs et dépend fortement de l'intensité de la surveillance, rendant la comparaison entre zones difficile.

Distinction entre surveillance industrielle et surveillance réglementaire

Les fréquences attribuées aux portions de surveillance industrielle se font indépendamment des préconisations relevant des études de dangers. La prise en compte de ces dernières avant le lissage à la maille segment serait envisageable.

1.4 Conclusion

L'objectif de cette thèse est de développer une méthode de planification et de construction de tournées de surveillance du réseau de transport de gaz en France intégrant plusieurs

modes de surveillance possibles. Ce travail s'inscrit à la fois aux niveaux tactique et opérationnel.

Les attentes en matière d'optimisation des activités de surveillance sont multiples. D'un point de vue managérial, les travaux visent à fournir aux responsables de secteur des outils d'aide à la construction et à la planification des tournées de surveillance. La méthode proposée devra par ailleurs conserver une certaine flexibilité, afin de s'adapter aux évolutions organisationnelles et opérationnelles, telles que la modification des périmètres de secteur ou des fréquences de surveillance. Ces enjeux s'inscrivent dans un contexte de transformation de l'organisation de NaTran, marqué par la réorganisation dite R24, qui conduit au regroupement des pôles d'exploitation et favorise l'harmonisation des pratiques.

Sur la base de ces considérations opérationnelles, nous précisons dans le chapitre suivant (Chapitre 2) la définition du périmètre traité et nous introduisons la modélisation mathématique associée.

Chapitre 2

Formalisation et modélisation du problème

Ce chapitre présente le problème d'optimisation de tournées de surveillance de NaTran. Dans ce problème, les surveillances d'arêtes doivent être réparties sur un ensemble de périodes de l'horizon de planification et associées à un mode de transport pour chacune d'elles. Pour chaque période et chaque mode, les tournées correspondantes sont ensuite déterminées de manière à minimiser le coût total. Dans une première partie, nous décrivons plus en détails le problème, ses hypothèses et introduisons les notations pour la modélisation. Dans une seconde partie, nous proposons une formulation mathématique du problème sous un modèle de Programmation Linéaire en Nombre Entiers (PLNE).

2.1 Présentation du problème

2.1.1 Le problème de surveillance périodique

Comme présenté dans le Chapitre 1, le réseau de transport de gaz de NaTran en France, est réparti sur quatre territoires. Chacun de ces territoires est également découpé en secteurs opérationnels comme illustré dans la Figure 2.1, ce qui conduit à des surveillances à différentes mailles. Par ailleurs, nous distinguons différents modes de surveillance aux pratiques et couvertures différentes. Enfin, des fréquences de surveillance sont allouées par portion selon le risque de dommages encouru par celle-ci.

Surveillance pédestre et automobile Les surveillances automobile et pédestre sont réalisées à la maille d'un secteur NaTran. Une équipe de techniciens est installée au sein de chaque secteur afin d'y réaliser les activités de surveillance et maintenance des installations rattachées à leur secteur. Ils sont notamment responsables de la surveillance des portions de canalisations appartenant à leur périmètre. Nous considérons qu'une seule voiture est allouée à chaque secteur pour la surveillance. Une semaine donnée, il y aura donc au plus 5 tournées effectuées par voiture dans chaque secteur (une par jour). De plus chaque tournée de surveillance débute et se termine à une des bases du secteur que l'on appellera dépôt. Certaines portions de canalisations, où la visibilité automobile est réduite, ne peuvent être inspectées qu'à pied. A cet effet, les exploitants garent leur véhicule à proximité, effectuent leur surveillance et reviennent à leur voiture pour continuer la tournée.

Surveillance aérienne La surveillance aérienne est réalisée à la maille d'un ou plusieurs territoires. En effet, jusqu'en 2024 deux prestataires de surveillance aérienne se partagent la surveillance aérienne sur l'ensemble du réseau NaTran. Chaque prestataire met un avion à disposition du (ou des) territoire(s) concerné(s). Ainsi un premier avion est alloué à la surveillance de la totalité du territoire Val de Seine par le prestataire AirLoisirs. Cet avion dispose d'un unique aérodrome où débuter et terminer ses tournées. Le second prestataire, RECTIMO Air Transports, alloue un avion pour la surveillance des trois territoires restants. Plusieurs aérodromes peuvent être utilisés pour le départ et la fin des tournées au sein d'une semaine. Toutefois, un aérodrome dit « principal », situé à Chambéry, sert de base : l'avion y débute ses opérations en début de semaine et y revient en fin de semaine. Contrairement aux tournées au sol, les tournées aériennes sont organisées par demi journées sur 3,5 jours dans la semaine soit 7 demi journées (tournées), cela permet en cas de manque de visibilité lié à la météo de reporter une tournée sur un autre jour.

Objectif L'objectif du problème est de définir des tournées de surveillance pour chaque semaine de l'année en tenant compte de la fréquence de passage préconisée de chaque portion mais aussi des différents modes de surveillance mis à disposition et des contraintes qui en résultent, tout en cherchant à minimiser les coûts globaux associés.

2.1.2 Périmètre et hypothèses de modélisation

Afin de mieux appréhender le problème nous prenons en considération un ensemble de remarques et hypothèses :

1. Les surveillances à pied sont considérées de façon transparente dans les tournées avec véhicules. En effet les arêtes du graphe sont valuées par des temps de parcours, ainsi les arêtes à parcourir à pied auront un temps de parcours plus élevé.
2. Nous n'allouons pas directement de temps de pause aux employés dans les tournées : la pause déjeuner est comprise dans la durée d'une tournée en véhicule, laissant la liberté aux exploitants de la planifier au moment et au lieu souhaités.

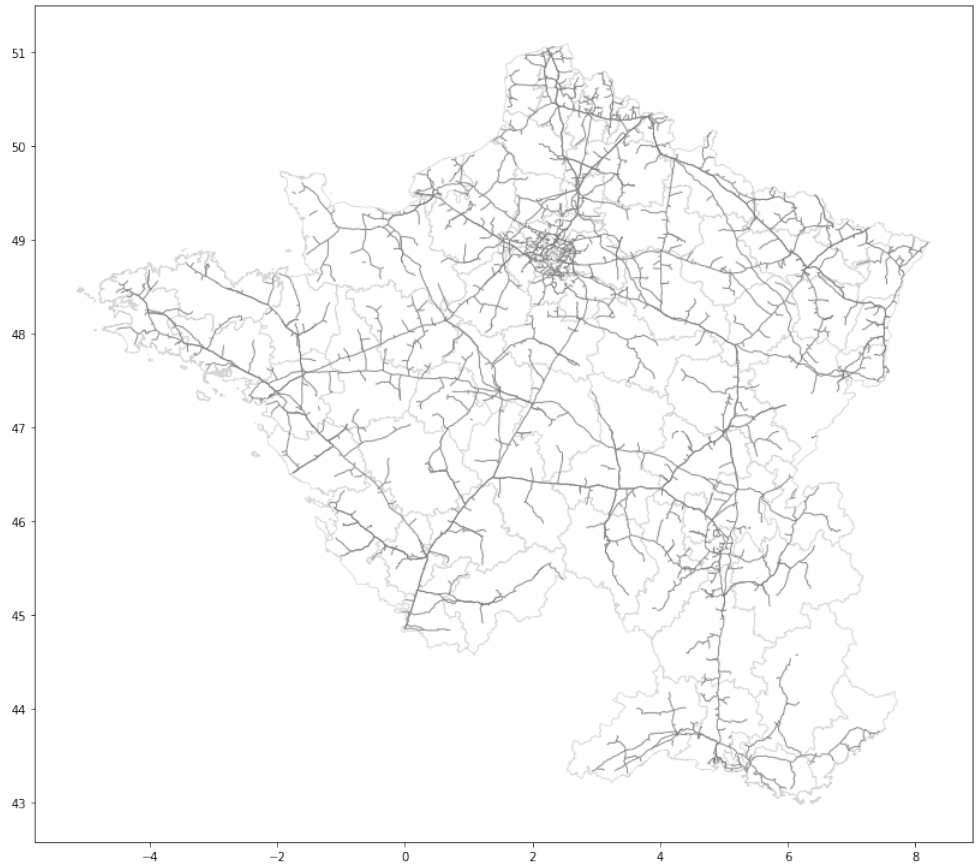


FIGURE 2.1 – Carte du réseau de transport NaTran

3. Lors de la surveillance des portions de canalisations, les voitures et avions circulent à une vitesse réduite par rapport à un simple passage.

4. Nous considérons qu'il peut y avoir un ou plusieurs dépôts au sol par secteur. Une tournée terrestre doit commencer et se terminer au même dépôt.

5. Les tournées d'une même semaine peuvent être effectuées dans n'importe quel ordre au sein de la semaine, à l'exception des tournées aériennes opérées par RECTIMO Air Transports, pour lesquelles la présence d'aérodromes intermédiaires impose le respect d'un ordre précis de passage.

2.1.3 Notations

Le problème tel que décrit précédemment correspond à un problème de tournées périodiques sur les arêtes. Il concerne un ensemble S de secteurs sur un horizon d'un an. On note $W = \{1, 2, \ldots, 52\}$ l'ensemble des semaines de l'horizon. Afin de ne pas alourdir les notations et puisque les périmètres couverts par les deux prestataires de surveillance aérienne sont distincts, c'est-à-dire que chaque portion de surveillance est prise en charge par un seul prestataire, les notations ainsi que le problème formulé s'appliquent séparément à chacun des périmètres. Nous introduisons un ensemble de notations inspirées de l'article Chu, Labadie et Prins 2004 et résumées dans la table 2.1 :

Description des notations relatives aux tournées routières Comme décrit précédemment les surveillances automobiles sont organisées à la maille d'un secteur. Pour chaque secteur $s \in S$ et pour chaque semaine w de l'horizon W, nous considérons un ensemble de tournées $K^v = \{1, 2, \cdots, 5\}$. L'objectif est de planifier et d'optimiser toute tournée $k \in K^v$ pour chaque semaine $w \in W$. Pour tout secteur $s \in S$ nous introduisons $X^v(s)$ l'ensemble des nœuds accessibles par voie terrestre, ainsi que l'ensemble des arêtes $E^v(s) = \{[i, j] : i, j \in X^v(s), i \leq j\}$, celles-ci étant pondérées par un temps de passage automobile t^v_{ij} en minutes pour toute arête $[i, j] \in E^v(s)$. Un sous-ensemble $R^v(s) \subset E^v(s)$ regroupe les arêtes du secteur $s \in S$ requérant une surveillance et pouvant être parcourues en voiture. Ces arêtes sont pondérées en plus par un temps de surveillance t'^v_{ij} en minutes supérieur au temps de passage t^v_{ij}. De plus, chaque secteur dispose d'un ensemble de dépôt(s) $D^v(s)$ où débutent et s'achèvent les tournées automobiles. Notons que la majorité des secteurs ne disposent que d'un dépôt unique. Par ailleurs une tournée automobile est limitée par un temps de δ^v minutes, et le coût horaire d'un déplacement d'un exploitant de secteur pour de la surveillance est de c^v euros.

Description des notations relatives aux tournées aériennes Pour chaque semaine w de l'horizon W un ensemble de tournées $K^p = \{1, 2, \cdots, 7\}$ est planifié. Nous notons X^p l'ensemble des nœuds accessibles en avion. Nous en déduisons l'ensemble des arêtes $E^p = \{[i, j] : i, j \in X^p, i \leq j\}$ qui sont pondérées par un temps de passage t^p_{ij} en minutes.

Les arêtes à surveiller et pouvant être parcourues en avion font partie du sous-ensemble $R^p \subset E^p$. En plus du temps de passage, ces arêtes $[i,j] \in R^p$ sont pondérées d'un temps de surveillance t'^p_{ij} en minutes. L'ensemble D^p regroupe les aérodromes servant de points de départ et d'arrivée des tournées aériennes. Pour le prestataire AirLoisir, un seul aérodrome est inclus dans D^p, tandis que pour RECTIMO Air Transports, plusieurs aérodromes intermédiaires sont inclus dans D^p, en plus de l'aérodrome principal $d^p_0 \subset D^p$. Ces tournées sont limitées par une durée de δ^p minutes, et le coût horaire d'un déplacement aérien est de c^p euros.

Notations générales À partir des ensembles et paramètres par mode de surveillance nous introduisons des ensembles et paramètres généraux. Nous notons $G = (X, E)$ le graphe constitué de l'ensemble des nœuds X accessibles par voiture ou avion ($X = \{\bigcup_{s\in S} X^v(s)\} \cup X^p$), et de l'ensemble des arêtes E qui peuvent être traversées par voiture ou avion ($E = \{\bigcup_{s\in S} E^v(s)\} \cup E^p$). De plus l'ensemble $D = \{\bigcup_{s\in S} D^v(s)\} \cup D^p$ avec $D \subset X$ correspond à tous les dépôts (bases secteurs ou aérodromes) où débutent et s'achèvent les tournées. Un sous-ensemble $R \subset E$ avec $R = \{\bigcup_{s\in S} R^v(s)\} \cup R^p$ désigne les arêtes où un acte de surveillance est requis. On peut noter que puisque certaines portions peuvent être surveillées par plusieurs modes de surveillances, ainsi, une même arête peut appartenir à la fois à $\bigcup_{s\in S} R^v(s)$ et à R^p. Une fréquence de passage f_{ij} est associée à chaque arête $[i,j] \in R$: elle correspond au nombre de visites requises de l'arête sur l'horizon W et est indépendante du mode de surveillance adopté. De plus, nous affectons à chaque arête $[i,j] \in R$ deux valeurs $\Delta_{\min}[i,j]$ et $\Delta_{\max}[i,j]$ correspondant aux délais minimal et maximal entre deux passages sur l'arête. Ces valeurs sont couplées à la fréquence de passage f_{ij}.

Ce problème peut être défini comme un problème périodique de tournées sur les arêtes avec contraintes de capacité temporelle et flotte multimodale. Il est NP-difficile car c'est une extension du *Capacitated Arc Routing Problem* (CARP), connu pour être NP-difficile (Golden et Wong 1981)

2.2 Modèle de tournées périodiques sur les arêtes

Une fois le problème décrit, et les notations introduites nous pouvons en proposer une formulation mathématique, sous la forme d'un PLNE. A cette fin, nous définissions dans cette section les variables de décisions, formulons la fonction objectif, puis précisons l'ensemble des contraintes associées.

2.2.1 Variables de décision

Afin de permettre la construction des tournées nous considérons les variables de décision suivantes. La variable binaire x_{ij}^{skw} (respectivement x_{ji}^{skw}) est relative à la surveillance par voie terrestre et est égale à 1 si une voiture traverse l'arête $[i,j] \in E^v$ de i vers j (respectivement de j vers i) lors de la tournée $k \in K^v$ de la semaine $w \in W$ du secteur $s \in S$, et 0 sinon. La variable binaire y_{ij}^{kw} (respectivement y_{ji}^{kw}) est relative à la surveillance aérienne et est égale à 1 si un avion traverse l'arête $[i,j] \in E^p$ de i vers j (respectivement de j vers i) lors de la tournée $k \in K^p$ de la semaine $w \in W$, et 0 sinon.

Dans un besoin de distinguer un simple passage sur une arête d'une opération de surveillance effectuée sur la même arête $[i,j] \in R$, nous introduisons ci-après trois variables supplémentaires liées à la surveillance. En effet, les temps de trajet, et par conséquent les coûts, diffèrent selon qu'il s'agisse d'une surveillance ou d'un simple passage sur une arête. Par ailleurs, afin d'assurer le respect des fréquences de surveillance pour chaque arête, il est important de dissocier les deux pratiques. Ainsi nous introduisons la variable binaire α_{ij}^{skw} relative à une surveillance en voiture : elle est égale à 1 si l'arête $[i,j] \in R$ est surveillée en voiture lors de la tournée $k \in K^v$ de la semaine $w \in W$ dans le secteur $s \in S$, et 0 sinon. De la même manière la variable binaire β_{ij}^{kw} est relative à une surveillance en avion : elle est égale à 1 si l'arête $[i,j] \in R$ est surveillée en avion lors de la tournée $k \in K^p$ de la semaine $w \in W$, et 0 sinon. Enfin la variable binaire λ_{ij}^{w} est égale à 1 quand une surveillance est effectuée sur l'arête $[i,j] \in R$ lors de la semaine $w \in W$, et 0 sinon. Une arête ne pouvant être surveillée plus d'une fois par semaine, la variable λ_{ij}^{w} est nécessairement inférieure ou égale 1. Cette variable est calculée de la façon suivante : $\lambda_{ij}^{w} = \sum_{s \in S} \sum_{k \in K^v} \alpha_{ij}^{skw} + \sum_{k \in K^p} \beta_{ij}^{kw}$.

Toutes les variables de décision sont résumées dans le tableau 2.2.

2.2.2 Fonction Objectif

L'objectif du problème est de minimiser le coût global de la surveillance du réseau de transport de gaz. Nous formulons notre fonction objectif de la manière suivante : dans un premier temps (2.1) nous formulons les coûts liés aux passages sur toutes les arêtes traversées d'une part par voie terrestre et d'autre part par voie aérienne. Dans un second temps (2.2) nous ajoutons les sur-coûts liés à la surveillance des arêtes $[i,j] \in R$ en faisant la différence entre les temps de parcours (t^v_{ij} et t^p_{ij}) et les temps de surveillance ($t^{v\,\prime}_{ij}$ et $t^{p\,\prime}_{ij}$).

$$\text{minimiser} \quad \sum_{s\in S}\sum_{k\in K^v}\sum_{w\in W}\sum_{[i,j]\in E^v(s)} \frac{c^v}{60} t^v_{ij}(x^{skw}_{ij} + x^{skw}_{ji}) + \sum_{k\in K^p}\sum_{w\in W}\sum_{[i,j]\in E^p} \frac{c^p}{60} t^p_{ij}(y^{kw}_{ij} + y^{kw}_{ji}) \tag{2.1}$$

$$+ \quad \sum_{s\in S}\sum_{k\in K^v}\sum_{w\in W}\sum_{[i,j]\in R^v(s)} \frac{c^v}{60}(t^{v\,\prime}_{ij} - t^v_{ij})\alpha^{skw}_{ij} + \sum_{k\in K^p}\sum_{w\in W}\sum_{[i,j]\in R^p} \frac{c^p}{60}(t^{p\,\prime}_{ij} - t^p_{ij})\beta^{kw}_{ij} \tag{2.2}$$

Cette fonction objectif est soumise à différentes contraintes liées à la construction des tournées, à la prise en compte de contraintes opérationnelles et au respect des fréquences établies.

2.2.3 Contraintes

Contraintes liées à la construction des tournées Tout d'abord, une tournée en voiture dans le secteur $s \in S$ doit débuter et finir à un même dépôt $d \in D_v(s)$ du secteur selon les contraintes 2.3 et 2.4. Les contraintes 2.5 assurent la continuité des tournées de voitures, elles garantissent notamment que le dépôt utilisé à l'arrivée est le même que le dépôt de départ. Lorsqu'une surveillance en voiture est effectuée sur l'arête $[i,j] \in R^v(s)$, il faut que celle-ci soit traversée en voiture selon les contraintes 2.6.

Les contraintes 2.7 permettent d'éliminer les sous-tours des tournées en voiture. Nous nous concentrons sur l'écriture de ces dernières contraintes inspirées de l'article Chu,

Labadie et Prins 2005 où les auteurs ont proposé une nouvelle formulation des contraintes de Benavent, Corberán et Sanchis 2000. Nous notons $Q \subseteq X^v(s) \setminus D^v(s)$ un sous-ensemble de nœuds accessibles en voiture excluant les dépôts. Pour chaque sous-ensemble Q, on considère deux nœuds $q_1 \in Q$ et $q_2 \in Q$ tels qu'un passage est requis pour l'arête $[q_1, q_2] \in R^v(s)$. Si la variable $\alpha^{skw}_{q_1 q_2} = 1$ alors la voiture de la tournée $k \in K^v$ doit traverser la coupe $(Q, X^v(s) \setminus Q)$ lors de la semaine $w \in W$ car l'arête $[q_1, q_2] \in Q$. Il faut qu'au moins une arête $[i, j] \in E^v(s)$ avec $i \in Q$ et $j \notin Q$ soit traversée si une arête du sous-ensemble Q est servie. Cette contrainte permet l'élimination de sous tours lorsque les coûts des arêtes sont non négatifs.

Une tournée en avion doit débuter et finir à un dépôt de l'ensemble D^p selon les contraintes 2.9 et 2.11. Par ailleurs, la première et la dernière tournée doivent commencer et finir au dépôt principal selon les contraintes 2.8 et 2.10. De plus, les contraintes 2.12 permettent que le dépôt d'une fin de tournée aérienne $k \in K^p$ soit le dépôt de départ de la prochaine tournée aérienne $k + 1 \in K^p$ de la même semaine w. Les contraintes 2.13 assurent la continuité des tournées aériennes. Les contraintes 2.14 assurent que l'arête $[i, j] \in R^p$ est traversée en avion si celle-ci est surveillée par le même mode de surveillance. Enfin les contraintes 2.15 éliminent les sous-tours des tournées en avion.

$$\sum_{\substack{j \in X^v(s) \\ d \in D^v(s)}} x^{skw}_{dj} = 1 \qquad \forall s \in S, \quad \forall k \in K^v, \quad \forall w \in W \tag{2.3}$$

$$\sum_{\substack{i \in X^v(s) \\ d \in D^v(s)}} x^{skw}_{id} = 1 \qquad \forall s \in S, \quad \forall k \in K^v, \quad \forall w \in W \tag{2.4}$$

$$\sum_{j \in X^v(s)} (x^{skw}_{ij} - x^{skw}_{ji}) = 0 \qquad \forall s \in S, \quad \forall k \in K^v, \quad \forall w \in W, \quad \forall i \in X^v(s) \tag{2.5}$$

$$x^{skw}_{ij} + x^{skw}_{ji} \geq \alpha^{skw}_{ij} \qquad \forall s \in S, \quad \forall k \in K^v, \quad \forall w \in W, \quad \forall [i, j] \in R^v(s) \tag{2.6}$$

$$\sum_{i\in Q}\sum_{j\notin Q}(x_{ij}^{skw}+x_{ji}^{skw})\geq\alpha_{q_1q_2}^{skw} \qquad \forall s\in S,\quad \forall k\in K^v,\quad \forall w\in W,\quad \forall Q\subseteq X^v(s)\setminus\{d^v(s)\},\quad \forall[q_1,q_2]\in R^v(s)\cap Q^2 \tag{2.7}$$

$$\sum_{j\in X^p} y_{d_0^p j}^{1w}=1 \qquad \forall w\in W \tag{2.8}$$

$$\sum_{\substack{j\in X^p\\ d\in D^p}} y_{dj}^{kw}=1 \qquad \forall k\in\{2,\cdots,|K^p|\}\quad \forall w\in W \tag{2.9}$$

$$\sum_{i\in X^p} y_{id_0^p}^{|K^p|w}=1 \qquad \forall w\in W \tag{2.10}$$

$$\sum_{\substack{i\in X^p\\ d\in D^p}} y_{id}^{kw}=1 \qquad \forall k\in\{1,\cdots|K^p|-1\},\quad \forall w\in W \tag{2.11}$$

$$\sum_{i\in X^p} y_{id}^{kw}=\sum_{i\in X^p} y_{di}^{k+1,w} \qquad \forall k\in\{1,\ldots,|K^p|-1\},\quad \forall w\in W,\quad \forall d\in D^p \tag{2.12}$$

$$\sum_{j\in X^p}(y_{ij}^{kw}-y_{ji}^{kw})=0 \qquad \forall k\in K^p,\quad \forall w\in W,\quad \forall i\in X^p \tag{2.13}$$

$$y_{ij}^{kw}+y_{ji}^{kw}\geq\beta_{ij}^{kw} \qquad \forall k\in K^p,\quad \forall w\in W,\quad \forall[i,j]\in R^p \tag{2.14}$$

$$\sum_{i\in Q}\sum_{j\notin Q}(y_{ij}^{kw}+y_{ji}^{kw})\geq\beta_{q_1q_2}^{kw} \qquad \forall k\in K^p,\quad \forall w\in W,\quad \forall Q\subseteq X^p\setminus D^p,\quad \forall(q_1,q_2)\in R^p\cap Q^2 \tag{2.15}$$

Contraintes opérationnelles La durée des tournées de surveillance est limitée. La durée d'une tournée en voiture ne dépasse pas un temps de δ^v minutes selon les contraintes 2.16 et respectivement une tournée en avion ne dépasse pas δ^p minutes selon les contraintes 2.17.

$$\sum_{[i,j]\in E^v(s)} t^v_{ij}(x^{skw}_{ij} + x^{skw}_{ji}) + \sum_{[i,j]\in R^v(s)} (t^{v\,\prime}_{ij} - t^v_{ij})\alpha^{skw}_{ij} \leq \delta^v \quad \forall s \in S, \quad \forall k \in K^v, \quad \forall w \in W \tag{2.16}$$

$$\sum_{[i,j]\in E^p} t^p_{ij}(y^{kw}_{ij} + y^{kw}_{ji}) + \sum_{[i,j]\in R^p} (t^{p\,\prime}_{ij} - t^p_{ij})\beta^{kw}_{ij} \leq \delta^p \quad \forall k \in K^p, \quad \forall w \in W \tag{2.17}$$

Contraintes de fréquences Nous traitons un problème de tournées sur les arêtes dit périodique. Les arêtes doivent être visitées un certain nombre de fois sur l'horizon donné W et de façon suffisamment espacée dans le temps.

Dans un premier temps afin de tenir compte de la périodicité des passages en avion ou en voiture sur une arête $[i,j] \in R$ nous introduisons la variable λ^w_{ij} à l'aide des contraintes 2.18. Cette variable est égale à la somme des variables α^{skw}_{ij} et β^{kw}_{ij} et est égale à 1 si l'arête $[i,j] \in R$ est surveillée lors d'une tournée de la semaine $w \in W$ sans considérer le mode de surveillance adopté.

Le respect des fréquences de passage f_{ij} sur l'horizon donné W est assuré par les contraintes 2.19. Les contraintes 2.20 et 2.21 imposent d'espacer les visites d'au moins $\Delta_{\min}[i,j]$ et d'au plus $\Delta_{\max}[i,j]$ semaines. Enfin pour la bonne construction de la variable λ^w_{ij}, les variables α^{skw}_{ij} et β^{kw}_{ij} ont été définies pour tout $[i,j] \in R$, par conséquent les contraintes 2.22 et 2.23 permettent de bien allouer la surveillance au sol pour tout $[i,j] \in R^v(s)$ pour un secteur $s \in S$, et la surveillance aérienne pour tout $[i,j] \in R^p$.

$$\sum_{s\in S}\sum_{k\in K^v} \alpha^{skw}_{ij} + \sum_{k\in K^p} \beta^{kw}_{ij} = \lambda^w_{ij} \quad \forall w \in W, \quad \forall [i,j] \in R \tag{2.18}$$

$$\sum_{w\in W} \lambda^w_{ij} = f_{ij} \quad \forall [i,j] \in R \tag{2.19}$$

$$\sum_{m=w}^{w+\Delta_{\min}[i,j]-1} \lambda_{ij}^{m \bmod 52+1} \leq 1 \qquad \forall w \in W, \quad \forall [i,j] \in R \tag{2.20}$$

$$\sum_{m=w}^{w+\Delta_{\max}[i,j]-1} \lambda_{ij}^{m \bmod 52+1} \geq 1 \qquad \forall w \in W, \quad \forall [i,j] \in R \tag{2.21}$$

$$\alpha_{ij}^{skw} = 0 \qquad \forall s \in S, \quad \forall k \in K^v, \quad \forall w \in W, \quad \forall [i,j] \in R \setminus R^v(s) \tag{2.22}$$

$$\beta_{ij}^{kw} = 0 \qquad \forall k \in K^p, \quad \forall w \in W, \quad \forall [i,j] \in R \setminus R^p \tag{2.23}$$

Domaine des variables Enfin les contraintes suivantes définissent le domaine des variables du problème

$$x_{ij}^{skw}, x_{ji}^{skw} \in \{0,1\} \qquad \forall s \in S, \quad \forall k \in K^v, \quad \forall w \in W, \quad \forall [i,j] \in E^v(s) \tag{2.24}$$

$$y_{ij}^{kw}, y_{ji}^{kw} \in \{0,1\} \qquad \forall k \in K^p, \quad \forall w \in W, \quad \forall [i,j] \in E^p \tag{2.25}$$

$$\alpha_{ij}^{skw} \in \{0,1\} \qquad \forall s \in S, \quad \forall k \in K^v, \quad \forall w \in W, \quad \forall [i,j] \in R \tag{2.26}$$

$$\beta_{ij}^{kw} \in \{0,1\} \qquad \forall k \in K^p, \quad \forall w \in W, \quad \forall [i,j] \in R \tag{2.27}$$

$$\lambda_{ij}^{w} \in \{0,1\} \qquad \forall w \in W, \quad \forall [i,j] \in R \tag{2.28}$$

2.3 Conclusion

Dans ce chapitre, nous avons formalisé le problème industriel lié à la surveillance du réseau de transport de NaTran en un modèle mathématique. Ce modèle est défini comme un problème périodique de tournées sur les arêtes avec capacité et flotte multimodale. Cette modélisation a permis de traduire les contraintes opérationnelles et les objectifs de planification en un cadre d'optimisation combinatoire, représentant la complexité du problème étudié. Toutefois, la taille et la complexité combinatoire du problème rendent le modèle complet intraitable en pratique par des méthodes de résolution exactes. Cela met en évidence les limites d'un modèle exact pour la résolution d'un tel problème et motive

l'utilisation de méthodes alternatives pour obtenir des solutions de bonne qualité dans des temps de calcul raisonnables. Nous présentons dans le prochain chapitre un état de l'art des modèles et méthodes proposés dans la littérature pour des problèmes similaires, situer notre problème dans le contexte des recherches existantes, et identifier des approches pertinentes pour le résoudre.

Ensembles	
S	ensemble des secteurs
$X^v(s)$	ensemble des nœuds du secteur $s \in S$ accessibles en voiture
$D^v(s) \subset X^v(s)$	ensemble des dépôts voiture du secteur $s \in S$
$E^v(s)$	ensemble des arêtes du secteur $s \in S$ accessibles en voiture
$R^v(s) \subset E^v(s)$	ensemble des arêtes accessibles en voiture à surveiller dans le secteur $s \in S$
K^v	ensemble des numéros de tournées effectuées en voiture dans une semaine
X^p	ensemble des nœuds du graphe G accessibles en avion
$D^p \subset X^p$	ensemble des dépôts du graphe G accessibles en avion
$d_0^p \subset D^p$	aérodrome principal
E^p	ensemble des arêtes du graphe G accessibles en avion
$R^p \subset E^p$	ensemble des arêtes du graphe G accessibles en avion à surveiller
K^p	ensemble des numéros de tournées effectuées en avion dans une semaine
W	ensemble des semaines de l'année
X	ensemble des nœuds du graphe G
E	ensemble des arêtes du graphe G
$D \subset X$	ensemble des dépôts du graphe G
$R \subset E$	ensemble des arêtes où une surveillance est requise

Paramètres	
t_{ij}^v	temps de passage sur l'arête $[i,j] \in E^v(s)$ en voiture ou à pied
t_{ij}^p	temps de passage sur l'arête $[i,j] \in E^p$ en avion
$t_{ij}^{v\,\prime}$	temps de surveillance de l'arête $[i,j] \in R^v$ en voiture ou à pied
$t_{ij}^{p\,\prime}$	temps de surveillance de l'arête $[i,j] \in R^p$ en avion
δ^v	durée maximale d'une tournée en voiture
δ^p	durée maximale d'une tournée en avion
f_{ij}	fréquence de passage requise par année sur l'arête $[i,j] \in R$
$\Delta_{\min}[i,j]$	délai minimum entre deux surveillances successives de l'arête $[i,j] \in R$
$\Delta_{\max}[i,j]$	délai maximum entre deux surveillances successives de l'arête $[i,j] \in R$
c^v	coût horaire pour les tournées en voiture
c^p	coût horaire pour les tournées en avion

TABLE 2.1 – Ensemble et paramètres pour le problème de tournées sur les arêtes

Variables	
$x_{ij}^{skw}, x_{ji}^{skw}$	variables binaires égales à 1 si et seulement si l'arête $[i,j] \in E^v(s)$ est parcourue de i vers j (respectivement de j vers i) en voiture ou à pied lors de la tournée voiture $k \in K^v$ de la semaine $w \in W$ dans le secteur $s \in S$
y_{ij}^{kw}, y_{ji}^{kw}	variables binaires égales à 1 si et seulement si l'arête $[i,j] \in E^p$ est parcourue de i vers j (respectivement de j vers i) en avion lors de la tournée avion $k \in K^p$ de la semaine $w \in W$
α_{ij}^{skw}	variable binaire égale à 1 si et seulement si l'arête $[i,j] \in R$ est surveillée en voiture lors de la tournée $k \in K^v$ de la semaine $w \in W$ dans le secteur $s \in S$
β_{ij}^{kw}	variable binaire égale à 1 si et seulement si l'arête $[i,j] \in R$ est surveillée en avion lors de la tournée $k \in K^p$ de la semaine $w \in W$
λ_{ij}^{w}	variable binaire égale à 1 si et seulement si l'arête $[i,j] \in R$ est surveillée en voiture ou en avion la semaine $w \in W$

TABLE 2.2 – Variables pour le problème de tournées sur les arêtes

Chapitre 3

État de l'art

Ce chapitre présente un état de l'art et situe nos travaux dans le cadre des recherches existantes. En théorie, les problèmes de tournées peuvent être séparés en deux catégories : les problèmes de tournées sur les nœuds où les clients sont représentés par des nœuds du graphe, et les problèmes de tournées sur les arcs où le service est réalisé sur les arcs ou arêtes du graphe. Comme cette thèse est consacrée à un problème de tournées sur des portions du réseau de transport de gaz, dans cet état de l'art nous allons nous concentrer sur ces problèmes de tournées sur les arcs. De plus, nous nous focalisons sur les problèmes déterministes et statiques. Dans une première partie (Section 3.1) nous nous intéressons aux différents problèmes de tournées sur les arcs ainsi qu'à leurs méthodes de résolution. La notion de planification, aspect important de notre problème, est également abordée dans cette section, où une part importante est consacrée aux problèmes périodiques de tournées sur les arcs. Nous abordons dans une deuxième partie (Section 3.2) d'autres variantes intéressantes des problèmes de tournées de véhicules, qui, dans la littérature, ont été principalement étudiés dans le cadre des problèmes de tournées sur les nœuds, mais n'ont pas ou peu, à notre connaissance, été explorés pour les problèmes de tournées sur les arcs. Dans une troisième partie (Section 3.3) nous présentons les méthodes d'apprentissage automatique et leur intégration dans des problèmes de recherche opérationnelle et plus particulièrement dans des problèmes de tournées de véhicules. Une quatrième partie (Section 3.4) est consacrée aux méthodes d'estimation de coûts pour les problèmes de tournées. Ce chapitre (Section 3.5) se conclut par l'identification des limites de la littérature existante en lien avec notre problématique.

3.1 Problème de tournées sur les arcs

Dans cette section, nous nous concentrons sur les problèmes de tournées sur les arcs (*Arc Routing Problem* (ARP)). Comme mentionné en introduction de ce chapitre, ces problèmes ont pour objectif de trouver une tournée optimale qui traverse les liens d'un graphe tout en respectant un ensemble de contraintes. De nombreuses situations réelles peuvent être modélisées sous forme de ARP tels que des problèmes de collecte de déchets, de sablage de routes en cas de verglas, de distribution de courriers, d'inspection d'infrastructures linéaires

comme les lignes à haute tension ou des canalisations dans notre cas. Plusieurs variantes ont été proposées dans la littérature, selon l'objectif à optimiser, les caractéristiques du graphe sous-jacent, les caractéristiques des véhicules, de la demande, ou encore du nombre de visites par lien. Dans la plupart des cas, les ARP à véhicule unique sont formulés sans contrainte de capacité, tandis que les ARP multi-véhicules incluent généralement une contrainte de capacité.

Dans cet état de l'art, le graphe sous-jacent des problèmes passés en revue est noté $G = (X, A, E)$ où X désigne l'ensemble des nœuds du graphe, A l'ensemble des arcs (i, j) et E l'ensemble des arêtes $[i, j]$. Les arêtes modélisent des liens non orientés comme des routes à double sens tandis que les arcs modélisent des liens orientés par exemple des routes à sens unique. Les graphes peuvent être mixtes car comportent à la fois des arêtes et des arcs. Par ailleurs, un graphe non dirigé est qualifié de venteux lorsque le coût de parcours de l'arête $[i, j]$ de i vers j n'est pas le même que celui de j vers i.

Ainsi les ARP peuvent être classifiés selon deux critères : le type de graphe sous-jacent, et le nombre de véhicules utilisés (un ou plusieurs), ce qui introduit ou non la prise en compte de contraintes de capacité. Dans cette partie, nous commençons par examiner les problèmes d'ARP impliquant un véhicule unique, sans prise en compte de la capacité, puis traitons les cas multi-véhicules. Enfin, nous consacrons une partie spécifique aux ARP périodiques. Au-delà des problèmes classiques, nous nous concentrons sur certaines variantes de l'ARP plus proches de notre problématique. Des revues recensant l'ensemble des variantes étudiées dans la littérature ont été proposées dans Corberán et Prins 2010, Corberán et Laporte 2015 et plus récemment dans Mourão et Pinto 2017 et Corberán, Eglese et al. 2021.

3.1.1 Problèmes de tournées sur les arcs avec un véhicule

Les ARP avec un seul véhicule consistent à déterminer une tournée de véhicule de coût minimal, sans contrainte de capacité, et qui traverse chaque lien requis au moins une fois. A l'exception du problème du postier chinois, tous les problèmes décrits sont NP-difficiles.

Chinese Postman Problem (CPP)

L'intérêt porté aux ARP commence par l'introduction du problème du postier chinois ou *Chinese Postman Problem* (CPP) en 1962. Ce problème est le problème d'optimisation de tournées sur les arcs le plus simple, qui constitue l'équivalent du problème de voyageur de commerce ou *Travelling Salesman Problem* (TSP) pour un problème de tournées sur les nœuds. Introduit par Mei-Ko 1962, ce problème modélise la situation d'un postier devant parcourir un ensemble de routes E qui lui sont assignées. L'objectif est de déterminer une tournée partant du bureau de poste, couvrant l'ensemble de ces routes au moins une fois puis revenant au point de départ en minimisant la distance de marche du postier. Le CPP dirigé ou non dirigé peut être résolu en un temps polynomial.

Cependant des extensions de ce problème comme le *Mixed Chinese Postman Problem* (MCPP) (Edmonds et Johnson 1973) contenant à la fois des arcs et des arêtes est NP-complet comme démontré par Papadimitriou 1976. Une généralisation de cette extension est le *Windy Postman Problem* (WPP) pour un graphe venteux. Win 1987 a démontré dans sa thèse que ce problème peut être résolu en temps polynomial lorsque le graphe d'entrée est pair (eulérien), en proposant un algorithme garantissant toujours une solution optimale dans ce cas.

De nombreuses autres variantes du CPP existent comme le *Hierarchical Chinese Postman Problem* (HCPP) défini dans Dror, Stern et Trudeau 1987 qui impose pour la tournée un ordre de précédence entre des classes d'arêtes, le *Maximum Benefit Chinese Postman Problem* (MBCPP) étudié par Pearn et K. Wang 2003, qui cherche une tournée traversant un sous-ensemble d'arêtes afin de maximiser un bénéfice, ou encore le *Chinese Postman Problem with time windows* (CPP-TW), étudié par Aminu et Eglese 2006, qui impose de respecter une fenêtre horaire pour la desserte de chaque arête.

Rural Postman Problem (RPP)

Le problème du postier rural ou *Rural Postman Problem* (RPP) a été introduit par Orloff 1974 pour étendre la définition du CPP pour des problèmes où seulement un sous ensemble d'arêtes appelées arêtes requises $R \subset E$ doivent être parcourus. Le problème a été démontré NP-complet par Lenstra et Kan 1976 dans le cas dirigé et non dirigé.

Dans Kang et Han 1998 les auteurs présentent un algorithme génétique ainsi qu'une nouvelle structure de chromosome pour le RPP. Les chromosomes sont construits à partir des nœuds du graphe hamiltonien obtenu après transformation du RPP. Cette approche permet de réduire la longueur des chromosomes par rapport aux structures existantes et, par conséquent, de diminuer la taille de l'espace de recherche.

Une variante intéressante du RPP est le *Windy Rural Postman Problem* (WRPP) introduite par Minieka 1979, pour illustrer le travail qu'un facteur devrait accomplir par journée de vent. Dans ce problème, pour chaque arête $[i, j] \in E$, le coût de traversée de i vers j peut être différent de celui de j vers i. Un graphe peut contenir à la fois des arcs requis et des arêtes requises : le problème de tournées associé est alors appelé *Mixed Rural Postman Problem* (MRPP). Dans Laporte 1997, les auteurs proposent une transformation du MRPP en problème du voyageur de commerce asymétrique (ATSP), permettant de résoudre des instances comportant jusqu'à 220 sommets et 660 liens (arêtes et arcs).

3.1.2 Problèmes de tournées sur les arcs multi-véhicules avec capacité

Les ARPs deviennent plus complexes lorsqu'une flotte de k véhicules est disponible pour desservir les liens requis $R \subset E$ du graphe G. Dans ce cadre, les véhicules, homogènes et basés en un dépôt commun, commencent puis terminent leur tournée à ce dépôt tout en parcourant l'ensemble des liens requis à coût minimal. Ces problèmes sont désignés sous le nom de *multiple* ou *k-vehicles* ARPs (k-ARPs).

Comme dans le cas à un seul véhicule, diverses variantes existent selon les contraintes imposées, la structure du graphe ou la fonction objectif retenue. La variante multi-véhicules la plus étudiée est le CARP. Dans ce problème les arêtes requises $[i, j] \in R$ ont une demande positive $q_{ij} \geq 0$ et chaque véhicule dispose d'une capacité limitée W. La grande majorité des k-ARPs sont formulés comme des variantes du CARP. Ces problèmes apparaissent dans de nombreuses applications réelles telles que la distribution de courrier, l'inspection de voies et pipelines, la collecte de déchets ménagers, etc.

Capacitated Arc Routing Problem **(CARP)**

Le CARP a été introduit dans Golden et Wong 1981. Il est défini pour un graphe non orienté. Il consiste à déterminer un ensemble de tournées de coût minimal permettant de parcourir l'ensemble des arêtes requises $R \subset E$, tout en satisfaisant la demande associée q_{ij} à chacune d'elles. Une tournée est un chemin fermé incluant le dépôt et traversant un ensemble d'arêtes, certaines sont requises, d'autres ne servent que de liaison (*deadheading*). La demande totale servie dans une tournée ne doit pas excéder la capacité maximale Q d'un véhicule, et l'on dispose de k véhicules pour assurer les services. Les auteurs ont proposé une formulation en PLNE pour ce problème, mais celle-ci génère un nombre de variables et de contraintes exponentiel. Ils ont également démontré que le problème est NP-difficile.

Malgré la complexité du CARP, on peut retrouver dans la littérature plusieurs méthodes exactes détaillées dans le chapitre de Belenguer, Benavent et Irnich 2015, tandis que des bornes inférieures sont recensées dans celui de Ahr et Reinelt 2015.

Comme indiqué précédemment, la première formulation de PLNE du CARP a été proposée par Golden et Wong 1981, mais elle génère un nombre exponentiel de variables et de contraintes, et la borne inférieure issue de sa relaxation continue est toujours nulle. Dans Benavent, Corberán et Sanchis 2000, les auteurs ont ensuite proposé une formulation dite *sparse*, qui utilise $O(K|R|)$ variables, ce qui reste acceptable pour de nombreuses applications pratiques sur des réseaux routiers clairsemés (où $|R| \leq |E| \approx 4|V|$). Cette modélisation repose sur des variables binaires x_{ij}^k, égales à 1 si le véhicule k dessert l'arête requise $[i, j]$, et sur des variables entières y_{ij}^k, correspondant au nombre de traversées de $[i, j]$ par le véhicule k sans service (*deadheading*). Les mêmes auteurs Belenguer et Benavent 2003 ont proposé un modèle simplifié en PLNE, appelé formulation *supersparse*. Bien qu'il ne soit pas valide pour le CARP, ce modèle permet d'obtenir de bonnes bornes inférieures. Il repose sur une unique variable entière x_{ij} par arête $[i, j]$, représentant le nombre de traversées sans service. Cependant sa compacité présente une limite : la solution ne précise pas quel véhicule emprunte quelle arête. Or, reconstruire les tournées à partir de cette solution est un problème NP-difficile, équivalent à un problème de *Bin Packing*. Néanmoins, ils proposent en plus un algorithme de coupes dérivé de cette formulation

qui fournit d'excellentes bornes inférieures, atteignant les meilleures performances des métaheuristiques sur des instances allant jusqu'à 50 sommets et 97 arêtes requises.

Une approche classique de résolution du CARP est de tirer parti des recherches développées pour le *Vehicle Routing Problem* (VRP) en le reformulant comme un problème de tournées sur les nœuds équivalent. Les performances obtenues dépendent de la transformation appliquée et sont pertinentes à condition que la dimension de l'instance n'augmente pas excessivement. Cette approche a d'abord été proposée par Pearn, Assad et Golden 1987, en transformant l'instance à $|R|$ arêtes requises en un graphe complet de VRP avec $3|R|+1$ nœuds. Chaque arête $[i,j]$ est remplacée par trois nœuds : les nœuds latéraux s_{ij} et s_{ji}, et le nœud central m_{ij}. Le VRP ainsi obtenu est défini sur l'ensemble de nœuds $N = \{1\} \bigcup \{s_{ij}, s_{ji}, m_{ij} | (i,j) \in E\}$ avec le nœud 1 correspondant au dépôt inchangé. Sachant que cette transformation équivaut à résoudre systématiquement une instance de grande taille de VRP, elle n'est pas vraiment utilisée en pratique. Une transformation du CARP en *Capacitated Vehicle Routing Problem* (CVRP) a également été proposée par Longo, De Aragao et Uchoa 2006. Dans celle-ci, chaque arête requise est remplacée par deux nœuds résultant en une instance de CVRP à $2|R|+1$ nœuds avec la contrainte que les deux nœuds issus de la même arête doivent être visités successivement. Notons que d'autres transformations en problèmes de tournées sur les nœuds ont été proposées dans la littérature notamment dans Foulds, Longo et Martins 2015.

Comme le CARP est NP-difficile plusieurs heuristiques constructives et métaheuristiques ont été développées et décrites dans la littérature. Une synthèse de ces approches figure dans le chapitre Prins 2015. La première heuristique proposée est l'algorithme *Construct Strike* introduit par Christofides 1973. Le graphe est d'abord rendu eulérien, mais les contraintes de capacité empêchent en général d'inclure toutes les tâches dans une seule tournée. L'heuristique extrait alors successivement des cycles réalisables, en supprimant leurs arêtes tout en conservant la connectivité (hors nœuds isolés). Quand cela n'est plus possible, des arêtes sont ajoutées pour rétablir des degrés pairs et éviter l'isolement du dépôt. Depuis, de nombreuses autres heuristiques ont été proposées pour le CARP comme l'algorithme *Path scanning* par Golden, DeArmon et Baker 1983, ou l'algorithme *Augmente-merge* par Golden et Wong 1981. Une autre classe d'heuristiques

dites *route first, cluster second* décompose le problème en deux étapes : dans un premier temps, un tour géant est construit, puis celui-ci est ensuite scindé en k tournées. Il a été adapté pour un CARP par Ulusoy 1985 puis amélioré par Prins, Labadie et Reghioui 2009.

Parmi les métaheuristiques, on distingue généralement les algorithmes à trajectoire unique (*single-solution based*), souvent appelés méthodes de recherche locale, qui travaillent à partir d'une seule solution et l'améliorent itérativement. La première métaheuristique présentée pour le CARP dans cette catégorie est le recuit simulé (ou *Simulated Annealing*, SA) développée par Eglese 1994 pour des problèmes de déneigement de routes. Elle a été suivie par plusieurs approches de Recherche Tabou introduites par Eglese et Li 1996 puis affinées par l'algorithme CARPET de Hertz, Laporte et Mittaz 2000. D'autres méthodes de voisinages adaptatifs ont été explorées, notamment le *Variable Neighborhood Search* (VNS) dont l'idée est de définir plusieurs voisinages afin d'explorer l'espace des solutions et réduire le risque d'être piégé dans un optimum local. Une variante nommée *Variable Neighborhood Descent* (VND) est présentée dans Hertz et Mittaz 2001. Celle-ci explore successivement plusieurs voisinages. A chaque itération, un voisinage est sélectionné et exploré à l'aide d'une recherche locale ; si une amélioration est trouvée, le processus recommence depuis le premier voisinage, sinon on passe au voisinage suivant et le processus s'arrête lorsque tous les voisinages ont été explorés sans qu'aucune amélioration ne soit obtenue. Une autre métaheuristique est le *Greedy Randomized Adaptative Search Procedure* (GRASP) appliqué au CARP par Prins et Wolfler Calvo 2005 où, à chaque itération, une solution initiale est générée à l'aide d'une heuristique gloutonne randomisée, puis améliorée grâce à une procédure de recherche locale. Enfin le *Guided Local Search* (GLS) introduit pour le CARP par Beullens et al. 2003 est une métaheuristique basée sur la recherche locale, qui modifie les coûts des composantes d'une solution afin d'échapper aux optima locaux, par exemple en pénalisant le plus long trajet à vide dans une tournée.

A côté de ces approches locales, d'autres métaheuristiques reposent sur des populations de solutions. Parmi elles, Greistorfer 2003 a proposé la *Tabu Scatter Search* (TSCS), combinant recherche tabou et recombinaison de solutions issues d'un pool initial : chaque amélioration y est ajoutée, puis jusqu'à cinq solutions sont combinées grâce à un opérateur

basé sur un problème de transport afin de générer de nouveaux points de départ. Dans la même logique hybride, Lacomme, Prins et Ramdane-Chérif 2001 ont introduit des algorithmes mémétiques, où des algorithmes génétiques sont enrichis par de la recherche locale appliquée aux solutions issues du croisement. Les chromosomes, codés comme des tournées géantes (RPP), sont optimisés avec la procédure de découpage de Ulusoy, leur valeur d'adaptation correspondant au coût de la solution CARP obtenue. Chaque descendant subit une recherche locale (relocation, échanges, 2-opt) avant d'être inséré dans la population. En 2004, les mêmes auteurs Lacomme, Prins et Ramdane-Cherif 2004 ont proposé une version plus performante intégrant des redémarrages rapides et une intensification augmentée de la recherche locale. Enfin, les algorithmes de colonies de fourmis ont également été adaptés au CARP par L. Santos, Coutinho-Rodrigues et Current 2010. Trois heuristiques ont été utilisées pour générer la population initiale et définir les niveaux de phéromones. A chaque itération, chaque fourmi construit un tour géant, décomposé ensuite en tournées réalisables grâce à l'heuristique de Ulusoy. Les solutions sont améliorées par une recherche locale (basée sur 12 types de mouvements) et remplacent, en cas de progrès, les moins performantes de la population, avant que les niveaux de phéromones ne soient mis à jour.

Extensions liées à la capacité et au temps (CARP-DD et TCARP)

Dans la formulation du CARP classique, seuls les arêtes requises sont associées à une demande, laquelle est déduite de la capacité du véhicule qui les parcourt ou les dessert. Cependant dans certains problèmes la capacité des véhicules n'est pas seulement contrainte par la demande des arêtes requises mais peut être impactée par des arêtes de liaisons (*deadheading arcs*), qui, bien que non requises, peuvent générer une consommation de capacité. Cette variante nommée *Capacitated Arc Routing Problem with Deadheading Demands* (CARP-DD) peut être utilisée pour modéliser des ARP où chaque tournée est soumise à une limite de temps ou de distance.

Dans Kirlik et Sipahioglu 2012, les auteurs proposent un premier modèle linéaire en nombre entiers qui permet de modéliser le problème et de résoudre de manière optimale de petites instances. Ils présentent également une heuristique constructive, dérivée et adaptée

d'une heuristique du CARP (Ulusoy 1985), afin de résoudre des instances de plus grandes tailles. De plus, deux procédures de post-optimisation sont introduites afin d'améliorer les solutions obtenues par l'heuristique. Dans Bartolini, Cordeau et Laporte 2013, les auteurs ont montré que les meilleures bornes inférieures du CARP sont faibles pour le CARP-DD. Ils introduisent donc une nouvelle famille d'inégalités pour renforcer ces bornes, puis développent un algorithme exact de type *branch-and-cut-and-price* qu'ils testent sur nombreuses instances dérivées des instances de référence.

Dans certaines variantes du CARP, il n'existe pas de contrainte de capacité liée aux arêtes requises, mais une limite de durée pour chaque tournée. Autrement dit, les capacités des véhicules et les coûts de traversée des arêtes sont exprimés en unités de temps. Ce problème est nommé le *Time Capacitated Arc Routing Problem* (TCARP). Son objectif est alors de minimiser le temps total nécessaire pour assurer le service des arêtes requises. Le TCARP peut être considéré comme un cas spécifique du CARP-DD.

Dans Armas et al. 2019, les auteurs proposent une heuristique et une métaheuristique capables de résoudre des instances de grande taille. La première, TSHARP, est une heuristique basée sur la méthode des économies, capable de fournir des solutions de bonne qualité en temps réel. La seconde est une métaheuristique qui combine une version biaisée et randomisée de TSHARP avec un Recuit simulé (*Simulated Annealing*).

Extensions liées aux dépôts (OCARP, MDCARP et CARP-IF)

Dans le CARP classique, chaque tournée débute et finit au même dépôt. Or dans certains cas pratiques, la tournée ne finit pas nécessairement au même sommet de départ, lequel n'est d'ailleurs pas toujours un dépôt. Le *Open Capacitated Arc Routing Problem* (OCARP) n'est pas contraint de former un circuit fermé c'est à dire de commencer et finir une tournée au même nœud (souvent le dépôt). Une tournée débute à un nœud X_s, et se termine à un nœud X_t. Le CARP est un cas spécial du OCARP où $X_s = X_p = X_0$ avec X_0 le dépôt. Ce problème a été introduit dans Usberti, P. M. França et A. L. M. França 2011 où les auteurs proposent une formulation en PLNE du problème, une heuristique de type *reactive Path-Scanning*, guidée par une règle de sélection des arêtes basée sur le rapport coût/demande et des règles elliptiques. Cette dernière est comparée à d'autres

heuristiques de *Path-Scanning* pour le CARP décrites dans la littérature.

Le *Multi-Depots Capacitated Arc Routing Problem* (MDCARP) est une autre généralisation du CARP, où plusieurs dépôts sont disponibles et où les tournées peuvent débuter et se terminer dans n'importe lequel d'entre eux. L'intérêt pour l'étude de cette variante s'explique par ses nombreuses applications pratiques. L'utilisation de dépôts multiples permet de réduire les coûts sur de grandes zones de service. Ces dépôts correspondent généralement à des entrepôts, des points de ravitaillement ou des centres de distribution, comme c'est le cas dans le commerce en ligne. Parmi les applications présentées dans la littérature figurent la distribution de courrier, les patrouilles policières et le recyclage de déchets explosifs. Cette variante a été notamment étudiée dans Yu, Liao et Yang 2023, où les auteurs présentent des algorithmes exacts et des méthodes d'approximations pour le MDCARP et certaines de ses variantes.

Le *Capacitated Arc Routing Problem with Intermediate Facilities* (CARP-IF) a été introduit par Ghiani, Improta et Laporte 2001 et suppose que les véhicules peuvent s'arrêter tout au long de la tournée à des stations intermédiaires. Ce problème est aussi appelé *Capacitated Arc Routing Problem with Refill Points* (CARP-RP) notamment dans Amaya, Langevin et Trépanier 2007. Dans ce problème la tournée doit débuter et finir au même dépôt mais un véhicule peut se recharger ou remplir sa capacité dans des stations dédiées distinctes du dépôt. Ainsi la route finale est composée de plusieurs segment, le premier commençant au dépôt principal et s'arrêtant à une station intermédiaire. A l'inverse, le dernier segment commence à une station intermédiaire et se finit au dépôt principal. Les stations intermédiaires permettent au véhicule de se réapprovisionner en ressources dans le cadre de problèmes de livraison, par exemple pour des problèmes de déneigement de routes où les véhicules pourront recharger leur stock de sel dans ces stations. Elles peuvent également servir au déchargement des véhicules, comme c'est le cas pour la collecte des déchets, où les véhicules effectuent des visites intermédiaires dans des décharges ou des incinérateurs. Les auteurs proposent deux bornes inférieures : la première repose sur le RPP, et la seconde utilise une relaxation d'une formulation linéaire en nombres entiers du problème. Ils ont également développé deux bornes supérieures, basées sur les solutions du RPP et du CARP. Dans Polacek et al. 2008 pour résoudre

le CARP-IF, les auteurs utilisent une métaheuristique de *Variable Neighborhood Search* (VNS).

Ensuite Ghiani, Guerriero et al. 2004, se sont intéressés au CARP-IF avec des contraintes de longueur de tournée : le *Arc Routing Problem with Intermediate Facilities under Capacity and Length Restriction* (CLARPIF), dans laquelle plusieurs véhicules peuvent être utilisés et où la longueur de chaque itinéraire ne doit pas dépasser une valeur prédéfinie. Ils ont proposé une heuristique constructive fondée sur la découpe d'une tournée du RPP, ainsi que deux algorithmes de recherche tabou. Dans Ghiani, Laganà et al. 2010, les auteurs proposent une nouvelle méthode de colonies de fourmis pour résoudre le CLARPIF.

Willemse et Joubert 2016 ont introduit le *Mixed Capacitated Arc Routing Problem with Time restrictions and Intermediate Facilities* (MCARP-TIF), qui prend en compte un réseau routier mixte et impose que chaque tournée ne dépasse pas une durée maximale. Pour résoudre ce problème en quasi temps réel, ils ont étudié quatre heuristiques constructives adaptées au MCARP-TIF et ont comparé leurs performances.

3.1.3 Problèmes de tournées sur les arcs avec périodicité

Dans cette partie nous nous intéressons aux ARP périodiques ou *Periodic Arc Routing Problem* (PARP). La notion de périodicité apparaît naturellement dans de nombreuses applications telles que : l'inspection périodique des routes, la collecte de déchets ménagers ou encore le déneigement des rues. Dans ces situations, un arc ou une arête nécessite un certain nombre de visites au cours d'un horizon de temps donné.

Les problèmes de tournées périodiques interviennent à deux niveaux décisionnels : au niveau de décision tactique, qui consiste à planifier et à affecter les clients aux différentes périodes de l'horizon de planification, et au niveau opérationnel, où il s'agit de déterminer les tournées pour chaque période. La périodicité augmente considérablement la complexité des problèmes de tournées, car elle inclue à la fois l'assignation de tâches à des périodes et la détermination les tournées pour chaque période.

Les problèmes de tournées périodiques ont été introduits pour la première fois par Beltrami et Bodin 1974 dans des travaux portant sur le problème de collecte des

déchets. La version périodique du VRP, le *Periodic Vehicle Routing Problem* (PVRP), a été introduite en 1984 dans Christofides et Beasley 1984.

Dans Mourgaya et Vanderbeck 2006, les auteurs proposent de définir la périodicité des tournées de trois manières différentes. Ainsi la périodicité peut être exprimée en termes :

— d'espacements à respecter entre deux passages successifs. $\Delta_{\min}[i,j]$ et $\Delta_{\max}[i,j]$ représentent le nombre de semaines minimums et maximums entre deux passages consécutifs sur l'arête $[i,j] \in R$. Ainsi la visite de l'arête doit être comprise dans l'intervalle $[\Delta_{\min}[i,j], \Delta_{\max}[i,j]]$.

— de fréquences de visites f_{ij} qui indique le nombre de passages à effectuer sur l'arête $[i,j] \in R$ sur l'horizon de temps W. On peut les compléter par des contraintes d'espacements entre deux passages successifs.

— de scénarios de visites comprenant l'ensemble des semaines auxquelles un passage doit avoir lieu. Ils peuvent être fournis directement ou déduits des contraintes d'espacements et/ou de fréquences.

Les problèmes périodiques de tournées de véhicules ont été largement étudiés dans le cadre des VRP ; une synthèse de ces travaux est proposée dans Campbell et Wilson 2014. Nous présentons dans la suite les travaux concernant les PARP.

Problème de tournées sur les arcs périodique avec un véhicule

Le premier article concernant un problème de tournées sur des arcs périodique concerne le *Periodic Rural Postman Problem* (PRPP) proposé par Ghiani, Musmanno et al. 2005. Dans cette variante, chaque arête requise doit être visitée un certain nombre de fois dans un horizon de temps donné, avec un espacement minimal entre les visites. Le PRPP attribue à chaque lien requis une combinaison de périodes, puis résout un RPP distinct pour chaque période du planning. L'objectif est de minimiser la distance totale. Le PRPP peut être considéré comme la succession de plusieurs RPP, chacun correspondant à une période de l'horizon de planification. Ce modèle s'applique à des problèmes récurrents, par exemple la collecte des déchets chaque semaine. Comme le PRPP regroupe plusieurs RPP, il reste NP-difficile. L'heuristique proposée par les auteurs commence par attribuer

la même combinaison de périodes à toutes les arêtes partageant la même fréquence, puis applique une recherche locale pour réduire les coûts.

Dans Benavent, Corberán, Laganà et al. 2019, les auteurs introduisent le PRPP avec des services non réguliers ou *Periodic Rural Postman Problem with Irregular Services* (PRPP-IS) pour un graphe mixte. Cette variante intervient notamment dans des problèmes d'entretiens routiers ou de la surveillance de réseaux routiers où les activités sont généralement menées de façon périodique et où différentes routes peuvent avoir des besoins de service différents (nombre de visites, traversées, etc.) selon leur importance dans le réseau. Ils développent un algorithme de *branch-and-cut* pour résoudre ce problème de manière optimale, qui a été testé pour des instances jusqu'à 349 services. Il s'agit du premier modèle exact conçu pour un PARP. Ce problème est de nouveau étudié dans Benavent, Corberán, Laganà et al. 2023, où les auteurs proposent un algorithme à deux phases qui combine heuristiques et programmation mathématique. Lors de la première phase, deux heuristiques multi-démarrages distinctes sont employées pour générer des solutions réalisables au PRPP-IS. Ensuite, dans la seconde phase, certains segments des tournées produits précédemment (correspondant à plusieurs jours) sont regroupés au sein d'un programme linéaire binaire afin d'obtenir des solutions réalisables améliorées. Ils ont pu résoudre des instances de plus grandes tailles, jusqu'à 316 arêtes et 733 services, qui n'ont pas pu être résolues avec l'algorithme *branch-and-cut*.

Une variante périodique du HCPP a été proposée par Keskin et Triki 2022. Les auteurs y présentent une formulation mathématique en programmation linéaire en nombres entiers mixtes (MILP), ainsi qu'une heuristique hybride combinant un recuit simulé et une adaptation de l'algorithme en couches, destinée à traiter les instances de tailles moyennes et grandes pour le problème du HCPP périodique (P-HCPP).

Problème de tournées sur les arcs périodique avec capacité (PCARP)

Le problème de tournées sur les arcs périodique avec capacité a été défini par Lacomme, Prins et Ramdane-Chérif 2005 à travers l'exemple du problème de collecte des déchets. Les auteurs différencient le *Periodic Capacitated Arc Routing Problem* (PCARP) en deux classes : l'une où les demandes et coûts dépendent des périodes, comme des problèmes de

collecte de déchets, et l'autre où ils restent constants, pour des interventions préventives comme l'inspection de lignes électriques par exemple. Les auteurs décrivent un algorithme mémétique qui repose sur un opérateur de croisement, capable de modifier simultanément les décisions tactiques et opérationnelles. Ils ont confronté leur approche à une heuristique d'insertion ainsi qu'à une méthode en deux phases, et elle s'est avérée plus efficace sur les instances testées.

Chu, Labadie et Prins 2005 proposent une formulation mathématique du PCARP ainsi que trois heuristiques constructives : deux méthodes d'insertion et un algorithme en deux phases. La première heuristique d'insertion appelée *Decreasing Frequencies, Nearest insertion Heuristic* (DFNH), classe les tâches par fréquence décroissante, puis assigne chacune d'elles à la combinaison de périodes minimisant son coût d'insertion. La seconde heuristique *Best Insertion Heuristic* (BIH) ne suit aucun ordre particulier pour les tâches : pour chaque tâche à insérer, toutes les combinaisons de périodes sont évaluées, et celle présentant le coût d'insertion le plus faible est retenue. Enfin l'approche en deux phases nommée *Lower Bound Heuristic* (LBH) utilise dans sa première phase une borne inférieure pour regrouper les tâches par période. La seconde phase construit les tournées effectives en résolvant, pour chaque groupe, le CARP correspondant. Cette approche en deux phases fournit de meilleurs résultats que les deux heuristiques d'insertion sur les instances testées. Dans Chu, Labadie et Prins 2004 un programme linéaire ainsi que des bornes inférieures sont proposées. Ils ont aussi développé une méta heuristique *Scatter Search*. Dans Chu, Labadie et Prins 2006, l'objectif principal est la taille de la flotte de véhicules, et comme objectif secondaire le coût total des tournées. Deux algorithmes sont developpés : une heuristique de meilleure insertion ainsi qu'une méta heuristique *Scatter Search*. Cette dernière améliore les solution de l'heuristique de meilleure insertion.

Marzolf, Trépanier et Langevin 2006 étudient une variante du PCARP pour la surveillance du réseau routier. Un seul véhicule doit inspecter tous les segments d'un réseau routier sur un horizon de deux semaines avec des fréquences de surveillance différentes par segment, par exemple chaque autoroute doit être surveillée au moins une fois par week-end et au moins une fois dans la semaine de cinq jours, alors qu'une route nationale doit être visitée une fois tous les sept jours. La longueur de chaque tournée est limitée par

la durée d'une période de travail. Un véhicule peut aussi être amené pendant sa tournée à répondre à des interventions urgentes, puis reprendre sa tournée. Cela a un effet sur la durée de la tournée. Les auteurs présentent un outil d'aide à la décision pour les véhicules équipés de GPS, intégrant des heuristiques de planification ainsi que des procédures de re-planification à la fin de chaque journée.

Kansou et Yassine 2009 ont proposé une heuristique à colonies de fourmis, combinée à une procédure de meilleure insertion intégrant successivement une méthode de sélection de tâche puis une méthode de sélection de combinaison. Cette approche a surpassé les méthodes précédentes sur un ensemble d'instances tests.

Dans Mei, Tang et Yao 2011, les auteurs abordent le PCARP en poursuivant deux objectifs : minimiser le nombre de véhicules et réduire le coût total des tournées. Ils proposent un algorithme mémétique intégrant un nouveau schéma de représentation des solutions ainsi qu'un nouvel opérateur de croisement. L'algorithme intègre aussi une procédure de fusion des tournées, qui contribue à améliorer la qualité des résultats.

Ces mêmes objectifs et stratégies sont adoptés dans Yuzhou Zhang et al. 2017. Une étape de décomposition de tournées est intégrée à l'algorithme mémétique afin de l'améliorer. L'opérateur de croisement et une procédure de recherche locale sont appliqués, tandis que la phase d'initialisation de la population est optimisée afin d'intégrer à la fois des solutions issues d'heuristiques et des solutions générées aléatoirement.

Y. Chen et Hao 2018 abordent également le PCARP bi-objectif et proposent un algorithme hybride de recherche locale en deux phases (HLS). La première phase vise à minimiser le nombre de véhicules, tandis que la seconde optimise simultanément les deux objectifs en utilisant le nombre de véhicules obtenu comme borne supérieure. Évalué sur 63 instances, cet algorithme améliore 44 meilleures bornes connues pour le coût total et atteint toutes les valeurs optimales pour le nombre de véhicules.

Des variantes du PARP

Monroy, Amaya et Langevin 2013 ont introduit une extention du PCARP appelée *Periodic Capacitated Arc Routing Problem with Irregular Services* (PCARP-IS), qui prend également en compte les services dits irréguliers. La principale différence avec le PCARP

classique est que le nombre de visites à effectuer lors des sous périodes de l'horizon de planification varie selon les classes d'arcs ou arêtes requis. Les applications mentionnées sont celles des opérations de maintenance routière et de surveillance de réseau routier. Les auteurs proposent une formulation mathématique afin de résoudre de manière optimale de petites instances ainsi qu'une heuristique de type *Cluster First, Route Second* pour traiter des instances de plus grandes tailles. Plus récemment, Laganà et Paronuzzi 2025 ont défini le PCARP-IS pour un graphe mixte et proposent une nouvelle approche matheuristique, combinant une formulation en PLNE basée sur les tournées à une recherche locale à redémarrages multiples (*multistart local search*). Leur méthode intègre deux stratégies de recherche locale pour améliorer la qualité des solutions.

Une autre variante du PARP prend en compte des contraintes de stocks comme le proposent Riquelme-Rodríguez, Gamache et Langevin 2014. Dans ce modèle, les arcs du réseau sont assimilés à des clients qui consomment progressivement une ressource (par exemple de l'eau) jusqu'à ce qu'un véhicule vienne les réapprovisionner, ce qui définit un niveau de stock. Chaque intervention réinitialise ce niveau, ainsi des décisions de routage et de gestion des stocks doivent être prises simultanément. La quantité d'eau livrée peut être fixe ou variable, et la fréquence de service est planifiée sur un horizon temporel donné. Deux formulations mathématiques ont été apportées en fonction de l'application considérée. Ils permettent de résoudre des instances jusqu'à 55 arêtes et jusqu'à 45 périodes. Le temps de calcul est affecté par la taille du réseau mais aussi le nombre de périodes dans l'horizon. Pour améliorer les performances, dans un second article Riquelme-Rodríguez, Langevin et Gamache 2014, les mêmes auteurs proposent un algorithme ALNS qui combine dynamiquement plusieurs opérateurs de destruction et de réparation.

Une nouvelle variante du PCARP est proposée dans C. G. Santos et al. 2017. Elle permet de modéliser le suivi périodique des barrages, intégrant plusieurs tâches par arc avec des fréquences distinctes, une flotte hétérogène et la possibilité de multiples passages sur un même arc le même jour. Un modèle mathématique exact est formulé et validé sur des instances générées aléatoirement.

L'article de Kansou et Kanso 2022 s'intéresse au *Periodic Open Capacitated Arc Routing Problem* (POCARP), appliqué à la relève de compteurs sur un horizon multi-

périodes. Deux métaheuristiques hybrides sont proposées : un algorithme génétique et une colonie de fourmis, elles sont enrichies par une recherche locale (relocation, swap, 2-opt, changement de combinaison). L'objectif est de planifier les périodes de service et les tournées des véhicules pour minimiser la distance totale. Les résultats montrent que ces méthodes surpassent les heuristiques d'insertion adaptées au problème.

Le PCARP trouve une application et variante dans la planification des tournées d'inspection de voies ferrées, ce problème a été étudié dans une thèse Lannez 2010 puis dans un article Lannez et al. 2015 sous l'appellation *Rail Track Inspection Scheduling Problem* (RTISP). Le problème consiste à programmer et réaliser les tâches d'inspection sur un réseau ferroviaire tout en minimisant les distances parcourues à vide avec un véhicule d'inspection de voies. La capacité opérationnelle des véhicules est limitée à la fois par la capacité de leur réservoir d'eau et par la durée de leur poste de travail. Selon les auteurs, le problème est équivalent à un *Multi-Capacitated Arc Routing Problem with Time Windows, Refill Points and a Heterogeneous Fleet.* Les auteurs présentent plusieurs formulations en variables binaires du problème et proposent une heuristique combinant les décompositions de Benders et Dantzig-Wolfe pour résoudre ce problème complexe de tournées sur les arcs.

Dans Kanso 2020, les auteurs introduisent le *Multi-Depot Periodic Open Capacitated Arc Routing Problem* (MDPOCARP). Il consiste à déterminer, pour chaque période, des tournées optimales où chaque tournée débute dans un dépôt donné, dessert un ensemble d'arêtes requises et se termine sur la dernière d'entre elles, sans obligation de retour au dépôt. Une heuristique construit une solution initiale puis un algorithme hybride de colonie de fourmis est proposé. Il combine l'algorithme de colonie de fourmis avec une procédure d'insertion et un algorithme de recherche locale.

3.2 Problèmes de tournées avec plusieurs modes de transport

Cette section s'intéresse aux problèmes de tournées de véhicules avec plusieurs modes de transport qui ne relèvent pas directement de l'état de l'art des problèmes de tournées sur les arcs (ARP), mais qui se rapprochent des problématiques de NaTran. Différents

modes de transport peuvent être mobilisés pour la surveillance du réseau. Cependant au delà des différences de coûts ou de capacités, les modes se distinguent également par leurs dépôts et la couverture d'arêtes à servir. Chez NaTran, plusieurs modes sont alloués à la surveillance : les avions permettent de couvrir rapidement plusieurs portions dans divers secteurs alors que la voiture est associée à chaque secteur et doit emprunter le réseau routier pour les surveillances.

La plupart des problèmes de tournées avec capacités considèrent en effet des véhicules homogènes, de même type, disposant de capacités identiques, desservant les mêmes zones et clients et partant d'un même dépôt. Or, dans de nombreux contextes, plusieurs modes de transport coexistent, soit en raison de contraintes d'accessibilité ou de visibilité qui imposent leur combinaison, soit pour des considérations économiques, temporelles ou environnementales.

L'intégration des différents modes de transport peut se faire à deux niveaux décisionnels : au niveau stratégique, il s'agit de définir les types et modèles de véhicules à acquérir, ou encore les contrats de prestation à établir ; au niveau tactique, les modes étant fixés, la décision porte sur l'affectation des véhicules aux clients ou aux périodes. Nous nous concentrons sur les décisions tactiques.

Dans l'état de l'art des problèmes de tournées, différentes variantes peuvent faire référence à une l'utilisation de modes de transport multiples. Un premier cas concerne les flottes de véhicules hétérogènes ou flotte mixte : ces problèmes considèrent une flotte de véhicules, pouvant différer sur leur capacité et sur leur coûts. Chaque tournée est effectuée par un véhicule d'un seul mode. Cette variante de problèmes a déjà été étudiée pour des ARP comme dans Cao et al. 2021 avec le *Multi-Depot Heterogeneous-Vehicle Capacitated Arc Routing Problem* (MDHVCARP).

Une autre variante de problèmes avec plusieurs modes de transport est le problème de tournées multi-modales. Dans ce cas, une même tournée peut combiner plusieurs modes de transport. Cette approche apparaît notamment pour le transport de marchandises. Une marchandise entre le dépôt et son client pourra emprunter différents modes de transport. Cette logique découle en partie de l'augmentation des coûts liés au transport routier, qui incite à tirer parti d'autres alternatives. Cependant, ces dernières ne sont pas toujours

porte-à-porte et nécessitent donc d'être combinées entre elles.

Le recours à plusieurs modes de transport peut présenter des avantages économiques, un gain de temps ou encore un impact environnemental réduit. L'ajout et la synchronisation des différents modes ajoute naturellement de la complexité au problème de tournées, cela doit notamment souvent prendre en compte des contraintes de fenêtres temporelles. Une revue des problèmes avec plusieurs modes de transport est donnée dans Sun, Lang et D. Wang 2015 et Baykasoğlu et al. 2019. À notre connaissance, aucune variante avec plusieurs modes de transport n'a été étudiée pour les ARPs.

3.3 Apprentissage automatique pour la recherche opérationnelle

Dans cette section, nous donnons une description des différentes méthodes d'apprentissage automatique et de leur intégration dans les méthodes de recherche opérationnelle, et plus particulièrement dans des problèmes de tournées de véhicules. Nous examinons aussi les caractéristiques utilisées pour la description des instances de problèmes de tournées.

3.3.1 Définitions et méthodes d'apprentissage automatique

Nous commençons par donner quelques définitions relatives à l'apprentissage automatique. Pour cela nous adoptons les définitions proposées dans le livre de Hastie et al. 2009. En science des données, l'apprentissage consiste à utiliser des données pour entraîner un modèle capable de reconnaître des motifs, de faire des prédictions ou de prendre des décisions. Différentes approches d'apprentissage existent en fonction de la nature des données en entrée, ou du type de résultats attendus. La science de l'apprentissage joue un rôle central dans les domaines des statistiques, de l'exploration de données et de l'intelligence artificielle, tout en établissant des passerelles avec l'ingénierie et d'autres disciplines. L'apprentissage automatique travaille souvent à partir de données finies et bruitées. Il est particulièrement adapté aux signaux naturels pour lesquels aucune modélisation mathématique claire n'existe, car la distribution réelle des données n'est pas connue analytiquement, comme dans le traitement d'images, de textes, de voix, ou encore dans les systèmes de recommandation, les réseaux sociaux et les prévisions financières. Dans la plupart des cas, le problème d'apprentissage se formalise statistiquement et se résout à l'aide de méthodes d'optimisation mathématique. Dans le cadre de ce rapport comme en recherche opérationnelle, nous privilégierons l'utilisation du terme *features* ou caractéristiques pour désigner les variables du modèle d'apprentissage pour éviter toute confusion avec les variables utilisées dans les modèles d'optimisation.

Apprentissage supervisé

En apprentissage supervisé, l'objectif est de faire des prédictions à partir d'un ensemble de données étiquetées, c'est-à-dire où chaque exemple de données est associé à sa valeur à prédire. Le modèle apprend à faire des prédictions en se basant sur les correspondances entre les données d'entrée et leurs étiquettes. Ce type d'apprentissage se décline en plusieurs catégories :

— Classification binaire : consiste à attribuer une observation à l'une de deux classes possibles tels que la régression logistique,

— Classification multi-classe : extension de la classification binaire à plus de deux catégories,

— Régression : dans le cas où les étiquettes (ou variable cible) sont à valeurs réelles, on parle de modèle de régression.

Des exemples d'applications de l'apprentissage supervisé incluent la prévision de la demande énergétique (régression) et la détection de spams (classification binaire).

Apprentissage non supervisé

Contrairement à l'apprentissage supervisé, dans l'apprentissage non supervisé, les données ne sont pas étiquetées, et le modèle doit découvrir des structures, des schémas ou des regroupements intrinsèques dans les données.

— *Clustering* ou Segmentation consiste à regrouper des objets en un nombre limité de groupes : des clusters. Les objets ayant des caractéristiques similaires sont regroupés alors que les objets ayant des caractéristiques différentes sont séparés. Contrairement à la classification les clusters ne sont pas définis en fonction d'une unique variable, et le nombre de clusters n'est pas défini.

— Réduction de dimension : La réduction de dimensionnalité est l'ensemble des techniques réduisant le nombre de *features* dans les données d'apprentissage. Nous distinguons les méthodes de sélection de *features*, des méthodes d'extraction de nouvelles *features*. Réduire la dimension permet de mieux visualiser les données, réduire la complexité algorithmique mais aussi améliorer la qualité des modèles.

Ce type d'apprentissage est utilisé dans des applications telles que la détection d'anomalies ou la segmentation de clients dans le secteur marketing.

Apprentissage par renforcement

L'apprentissage par renforcement est une méthode qui consiste à apprendre par essais et erreurs pour améliorer progressivement les décisions. Ce type d'apprentissage implique un agent qui interagit avec un environnement à travers un processus de Markov et apprend à prendre des actions pour maximiser une récompense cumulative. L'agent apprend par essais et erreurs, en ajustant ses actions en fonction des récompenses reçues. L'objectif est d'entraîner l'agent à maximiser la somme des récompenses attendues. Ce type d'apprentissage trouve des applications en robotique, dans les jeux, ainsi qu'en traitement du langage naturel. Une présentation approfondie de l'apprentissage par renforcement est fournie dans l'ouvrage de Sutton, Barto et al. 1998.

Apprentissage profond

L'apprentissage profond est une branche de l'apprentissage automatique qui repose sur des réseaux de neurones artificiels composés de multiples couches, d'où le terme profond. Inspiré du fonctionnement du cerveau humain, ce système traite les données au travers d'unités interconnectées qui transforment progressivement l'information pour affiner les prédictions ou les classifications. Ces algorithmes riches en paramètres, demandent un nombre très important de données et des ressources computationnelles importantes afin d'être entraînés. Grâce à leur architecture, ils permettent à la machine d'apprendre directement à partir des données et trouvent des applications variées, allant de la reconnaissance d'images et la traduction automatique à la conduite autonome. Des informations détaillées sur l'apprentissage profond sont disponibles dans le livre de Goodfellow et al. 2016.

3.3.2 Apprentissage automatique et recherche opérationnelle

La recherche opérationnelle peut être considérée comme une branche de l'intelligence artificielle dans la mesure où les deux disciplines exploitent des données pour accompagner la prise de décision. La recherche opérationnelle et l'apprentissage automatique sont souvent

perçus comme des approches distinctes, parfois alternatives, de la prise de décision basée sur les données. De manière simplifiée, on peut considérer que l'apprentissage automatique est l'approche de référence pour l'analyse prédictive, tandis que la recherche opérationnelle constitue la principale méthode d'analyse prescriptive, orientée vers l'optimisation des décisions. L'analyse prédictive consiste à utiliser des données ainsi que des modèles statistiques pour identifier des tendances et prédire des résultats, pour comprendre ce qui est susceptible de se produire (Mishra et Silakari 2012). L'analyse prescriptive va plus loin en proposant, à partir de ces prévisions, les actions les plus efficaces pour atteindre les objectifs et améliorer la performance décisionnelle, à l'aide d'outils d'optimisation et de simulation (Lepenioti et al. 2020).

La recherche opérationnelle est utilisée de diverses manières pour concevoir et améliorer les méthodes d'apprentissage automatique, notamment à travers la résolution de problèmes d'optimisation complexes. Elle intervient particulièrement dans le traitement de problèmes d'optimisation combinatoire ou discrète, tels que la construction d'arbres de décision, la sélection de sous-ensembles de variables ou l'optimisation de l'architecture de réseaux neuronaux. Ainsi, la recherche opérationnelle permet d'améliorer l'efficacité et la performance des algorithmes d'apprentissage automatique.

Plus récemment, l'apprentissage automatique est de plus en plus utilisé dans le cadre de la recherche opérationnelle. Étant donné la difficulté intrinsèque des problèmes d'optimisation, les algorithmes les plus avancés reposent sur des heuristiques conçues manuellement pour prendre des décisions qui seraient autrement trop coûteuses à calculer. L'apprentissage automatique est ainsi de plus en plus exploité pour assister, voire améliorer, les différentes méthodes de résolution. Cette partie s'appuie principalement sur l'article de Bengio, Lodi et Prouvost 2021, qui étudie les différentes méthodes d'intégration de l'apprentissage machine aux problèmes d'optimisation combinatoire. Selon les auteurs, l'apprentissage est utilisé à des fins d'approximation et de découverte de nouvelles stratégies. En effet, dans certains travaux, le chercheur utilise ses connaissances théoriques et empiriques pour guider un algorithme d'optimisation combinatoire, dans ces cas il voudra utiliser l'apprentissage automatique pour réduire le coût computationnel. Dans d'autres cas, lorsque l'expertise est limitée, l'apprentissage automatique permet d'améliorer la prise de décision

via l'apprentissage par renforcement. Les auteurs décrivent trois manières d'intégrer les algorithmes d'apprentissage à la résolution d'un problème d'optimisation combinatoire, que nous détaillons ci dessous.

Approches d'apprentissage de bout en bout Une première approche pour exploiter l'apprentissage automatique dans un problème d'optimisation est d'entraîner le modèle d'apprentissage à directement fournir une solution à une instance d'optimisation donnée. Cette approche est nommée *end-to-end* et est illustrée par la Figure 3.1. Ce sont principalement l'apprentissage profond et l'apprentissage par renforcement qui seront exploités dans ce cas.

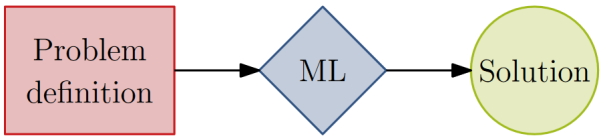


FIGURE 3.1 – Reproduite de BENGIO, LODI et PROUVOST 2021 *European Journal of Operational Research*, sous licence CC BY-NC-ND 4.0.

Dans KHALIL et al. 2017, les auteurs proposent d'automatiser la conception d'heuristiques pour des problèmes combinatoires NP-difficiles en combinant apprentissage par renforcement et *graph embedding*. Leur méthode apprend des politiques gloutonnes qui exploitent la structure récurrente des problèmes, et est efficace pour des problèmes sur graphes tels que le problème de couverture minimale, ou du problème de voyageur de commerce. Dans l'article de conférence KOOL, VAN HOOF et WELLING 2018, les auteurs utilisent l'apprentissage automatique pour apprendre des heuristiques efficaces pour le TSP et d'autres problèmes de tournées de véhicules. Leur modèle, basé sur des couches d'attention, est entraîné avec une méthode d'apprentissage par renforcement qui ajuste la politique de l'agent en favorisant les actions conduisant à de meilleures récompenses. Le modèle construit progressivement des solutions à partir des instances, et cette approche flexible permet d'obtenir des résultats proches de ceux d'heuristiques spécialisées sur plusieurs types de problèmes et avec un seul ensemble d'hyperparamètres.

L'apprentissage automatique en pré traitement Dans de nombreux cas, il est difficile de résoudre entièrement des problèmes d'optimisation à l'aide d'apprentissage automatique, en particulier lorsque le problème implique plusieurs niveaux de décision. Une autre utilisation de l'apprentissage automatique est en amont du problème d'optimisation, comme illustré dans la Figure 3.2. En effet les algorithmes d'optimisation ont parfois tout un ensemble de paramètres à régler qui influencent directement leurs performances.

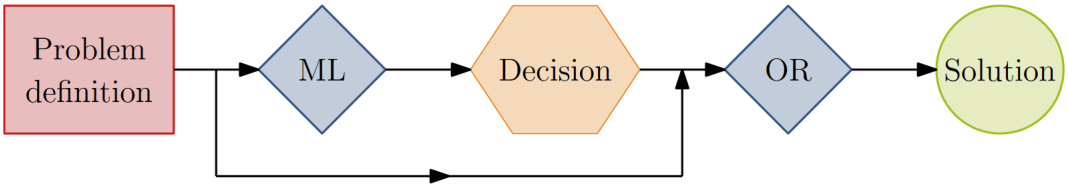


FIGURE 3.2 – Reproduite de BENGIO, LODI et PROUVOST 2021 *European Journal of Operational Research*, sous licence CC BY-NC-ND 4.0.

Par exemple, l'apprentissage automatique peut aider à configurer les paramètres d'un solveur MILP, à décider si une décomposition (KRUBER, LÜBBECKE et PARMENTIER 2017) ou une linéarisation (BONAMI, LODI et ZARPELLON 2018) sera bénéfique, ou encore à générer des solutions candidates dans des contextes appliqués comme la radiothérapie (MAHMOOD et al. 2018).

L'apprentissage automatique aux côtés de l'optimisation Pour généraliser le principe de la deuxième approche, les auteurs décrivent le cas où l'apprentissage automatique est utilisé en parallèle d'un algorithme d'optimisation, comme illustré dans la Figure 3.3. Ici c'est le même algorithme d'apprentissage automatique qui est utilisé un certain nombre de fois pour prendre les mêmes décisions locales au fil des itérations. L'intérêt est que ces décisions, bien qu'approximatives, n'affectent pas la validité de l'algorithme, mais peuvent considérablement accélérer sa convergence.

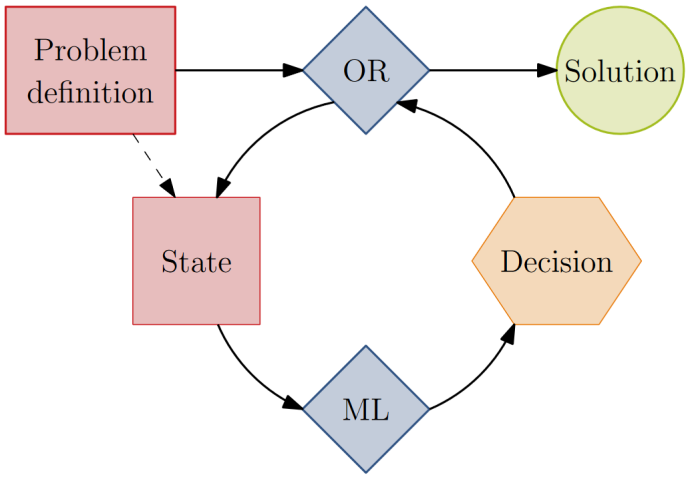


FIGURE 3.3 – Reproduite de BENGIO, LODI et PROUVOST 2021 *European Journal of Operational Research*, sous licence CC BY-NC-ND 4.0.

Par exemple, l'algorithme d'apprentissage automatique peut prendre des décisions concernant le choix des variables de branchements (LODI et ZARPELLON 2017), l'activation d'heuristiques primales (KHALIL et al. 2017) ou la sélection de coupes. Cette approche se veut hybride avec l'algorithme d'apprentissage automatique et l'algorithme d'optimisation qui interagissent en continu.

3.3.3 L'apprentissage automatique pour des problèmes de tournées

Dans cette partie nous nous concentrons sur les articles mettant en avant l'utilisation de l'apprentissage automatique pour des problèmes de tournées de véhicules et plus particulièrement pour des VRP. L'apprentissage automatique aident à affronter la complexité des problèmes de tournées de véhicules, mais aussi leurs incertitudes ou leur aspect dynamique.

Dans l'article BAI et al. 2023, les auteurs proposent une revue des méthodes hybrides qui associent des techniques analytiques à des outils d'apprentissage automatique pour traiter les problèmes de tournées de véhicules. Ils distinguent différentes méthodes d'intégration de l'apprentissage automatique : dans la modélisation du VRP (par exemple en estimant des paramètres ou des incertitudes), pour assister l'optimisation *offline* (par exemple avec des méthodes de décomposition ou de recherches de voisinages), et pour assister

l'optimisation *online* (comme le choix dynamique des heuristiques à appliquer). Ils concluent que l'apprentissage automatique peut jouer un rôle significatif, tant pour enrichir la modélisation des problèmes de VRP que pour améliorer les performances des algorithmes, aussi bien dans des contextes d'optimisation en ligne que hors ligne. Dans Shahbazian et al. 2024, les auteurs proposent aussi une revue des méthodes intégrant l'apprentissage automatique au problèmes de tournées mais mettent davantage l'accent sur les méthodes d'apprentissage par renforcement. Par exemple dans Furian et al. 2021, les auteurs proposent un cadre *branch-and-price* basé sur l'apprentissage, utilisé pour prédire la valeur des variables binaires dans la solution optimale et estimer les scores de branchement pour les variables fractionnaires. Ces prédictions sont intégrées pour la sélection de nœuds et de variables.

Une autre utilisation de l'apprentissage automatique pour la résolution de problèmes de tournées consiste à approximer les coûts d'une solution optimale. Dans Akkerman, Mes et Heijnen 2020, un modèle de régression linéaire permet d'approximer la distance d'un TSP ou VRP. Ce modèle est validé pour un problème de collecte de déchets. Le problème est formulé comme un *inventory routing problem* : il faut décider quels conteneurs vider quel jour et comment planifier les tournées des véhicules. L'approximation utilisée estime les coûts et niveaux de service des conteneurs sélectionnés à partir de leur localisation et remplissage.

Certaines études se concentrent sur la prédiction des durées sur des segments spécifiques d'un réseau routier, en tenant compte de la vitesse, ou encore du trafic. Par exemple, Yanru Zhang et Haghani 2015 utilise un algorithme de gradient boosting, qui combine plusieurs arbres simples pour obtenir une précision élevée tout en restant interprétable, dans le cadre de la prédiction des temps de trajet sur autoroute. Cet algorithme est particulièrement efficace pour gérer les imprévus liés aux incidents ou aux conditions météorologiques. Avec la croissance des données disponibles grâce aux capteurs, GPS et smartphones, l'algorithme de gradient boosting apparaît comme une méthode efficace pour exploiter les données massives et modéliser les relations non linéaires complexes dans la prédiction du trafic. Une revue de littérature concernant la prédiction de trafic à l'aide notamment d'apprentissage automatique figure dans Lana et al. 2018.

Pour conclure, il existe de nombreuses façons de combiner optimisation combinatoire et apprentissage automatique, selon les types de problèmes. L'apprentissage aide à accélérer la recherche de solutions, améliorer la qualité de celles ci, guider le choix des heuristiques, ajuster les paramètres des algorithmes ou encore faciliter la modélisation de problèmes complexes et dynamiques. Ces approches hybrides offrent un cadre flexible où les forces de l'optimisation exacte et les capacités de reconnaissance de motifs de l'apprentissage automatique se complètent, ouvrant la voie à des solutions plus efficaces et adaptées à des applications réelles.

3.4 Estimation des coûts et caractéristiques d'instances

Dans cette section, nous nous intéressons à l'estimation de coûts pour des solutions de problèmes de tournées. En effet, lorsque les problèmes étudiés comportent plusieurs niveaux de décisions, présentent un caractère dynamique ou incluent des incertitudes, il est particulièrement utile de pouvoir approximer rapidement le coût d'une solution. De plus certaines caractéristiques des instances à traiter peuvent être exploitées pour améliorer les approximations.

3.4.1 Estimation des coûts

De nombreux travaux se sont intéressés à l'approximation continue, notamment appliquée au TSP et au VRP. Ces approches permettent d'estimer rapidement les coûts des solutions de problèmes de tournées à partir des caractéristiques du problème. La plupart reposent sur l'approximation de distances selon une norme L_p tels que la distance euclidienne ou la distance de Manhattan, sur des instances où des points sont tirés aléatoirement dans une zone rectangulaire. Une revue de l'état de l'art concernant les modèles d'approximation continue est fournie dans Franceschetti, Jabali et Laporte 2017. Les premières recherches se concentrent sur des formules d'approximation dérivées analytiquement. L'une des premières approximations de distance est établie dans Beardwood,

Halton et Hammersley 1959 où les auteurs ont prouvé que la longueur du plus court chemin passant par n nœuds dans une région plane bornée de surface A est « presque toujours » asymptotiquement proportionnelle à $c\sqrt{nA}$ lorsque $n \to +\infty$, où c est une constante. Certaines extensions de cette approximation TSP, telles que celles de Chien 1992 et Christofides et Eilon 1969 prennent en compte de nouvelles informations telles que la forme et la taille du plus petit rectangle englobant tous les nœuds. Dans Kwon, Golden et Wasil 1995, les auteurs utilisent la régression linéaire et les réseaux neuronaux pour construire une formule d'approximation de la longueur du parcours TSP à travers 10 à 80 nœuds situés dans une région rectangulaire. Les caractéristiques utilisées sont le nombre de nœuds, l'aire du plus petit rectangle couvrant tous les nœuds et le rapport largeur/hauteur de la région de service. Cavdar et Sokol 2015 proposent un nouveau modèle d'estimation de la longueur des tournées basé sur la régression lorsque la distribution des nœuds n'est pas connue. Ils prennent en compte la dispersion des nœuds dans chaque dimension, ainsi que leur proximité et leur dispersion par rapport au centre de la zone.

Des approximations de distance sont également introduites pour le VRP. Webb 1968 est le premier à fournir une approximation de la longueur optimale du VRP à capacité limitée ou CVRP, en tenant compte de la distance entre les clients et le dépôt. Contrairement au TSP, des informations supplémentaires peuvent être prises en compte pour le VRP, comme le nombre de véhicules et leurs capacités respectives. Ainsi, Eilon et al. 1974 proposent de prendre en compte la forme et la surface de la zone de livraison, mais aussi la capacité de chaque véhicule en termes de nombre de clients pouvant être servis par un véhicule. Une approximation bien connue du coût du CVRP est l'approximation de Daganzo Daganzo 1984, où la longueur optimale est approximée par $2rnQ + 0,57\sqrt{nA}$, où Q représente le nombre maximal de clients par tournée, r la distance moyenne entre les clients et le dépôt, n le nombre de nœuds et A l'aire de la zone de service. Dans une étude complémentaire, les contraintes liées à la fenêtre temporelle sont également prises en compte par le même auteur Daganzo 1987. Les articles précédents se concentrent uniquement sur les approximations continues utilisant les distances de norme L_p (distances de Manhattan et euclidiennes). L'utilisation de ces distances rend les approximations moins adaptées aux cas réels, en

particulier dans les scénarios de déplacement urbain. Dans Merchán et Winkenbach 2019, les auteurs proposent un ajustement du modèle d'approximation continue pour le TSP de Beardwood, Halton et Hammersley 1959 en modifiant la constante c pour tenir compte de la sinuosité du réseau routier sous-jacent dans l'approximation. Ils émettent l'hypothèse que leur extension peut également s'appliquer à l'approximation de la distance VRP en adaptant la formulation de Daganzo 1984.

Les approches plus récentes ont adopté des méthodologies basées sur la régression pour estimer le coût de différentes solutions de problèmes de tournées. Dans Nicola, Vetschera et Dragomir 2019, les auteurs présentent des modèles de régression empiriques qui peuvent s'adapter à différents modèles et méthodes de résolution pour estimer la distance parcourue. Ils testent leur approche sur le TSP, le VRP à capacité limitée avec fenêtres temporelles (CVRPTW) et le problème de ramassage et de livraison multi-régions et multi-dépôts (MR-MDPDP). Ils utilisent des caractéristiques supplémentaires, notamment des mesures de dispersion, des mesures de fenêtres temporelles et des mesures géographiques. Plusieurs auteurs ont ensuite amélioré cette approche de régression linéaire en ajoutant de nouvelles caractéristiques, et plus précisément, des caractéristiques basées sur des heuristiques. Par exemple, Kou, Golden et Poikonen 2022 utilisent l'écart type des solutions heuristiques modifiées de Clarke et Wright pour développer un modèle de régression linéaire, afin de prédire les coûts optimaux des tournées TSP euclidiennes et non euclidiennes. Dans des travaux plus récents, Kou, Golden et Bertazzi 2024 les mêmes auteurs montrent comment l'ajout de la valeur minimale des solutions heuristiques modifiées de Clarke et Wright peut améliorer les approximations précédentes. De plus, Sinha Roy et al. 2021 proposent une approche par régression pour estimer la longueur optimale du parcours dans le problème du voyageur de commerce suffisamment proche (CETSP). En plus de plusieurs caractéristiques communes, ils utilisent le coût de la solution heuristique de recherche par voisinage variable de la zone de Steiner (SZVNS) comme nouvelle caractéristique du modèle de régression.

Des approches d'apprentissage automatique plus avancées ont été développées pour fournir des approximations rapides des coûts. Par exemple dans Akkerman, Mes et Heijnen 2020 et Akkerman et Mes 2022, les auteurs proposent de nouvelles *features*

pour leurs modèles de régression (régression linéaire, régression basée sur des arbres et modèles de réseaux neuronaux). Ensuite, ils sélectionnent les meilleures caractéristiques à l'aide des méthodes Elastic Net Regularization et Boruta-Shap (Keany 2020). Les modèles sélectionnés sont ensuite intégrés dans un cadre d'apprentissage adaptatif où ils sont utilisés pour soutenir une phase de sélection des clients. Enfin, un problème de routage de véhicules est résolu avec les clients sélectionnés, et le modèle de régression est ré-entraîné avec les nouvelles solutions afin de s'adapter à l'environnement changeant.

À notre connaissance, un seul article a abordé la question de l'approximation continue dans un problème de tournées sur les arcs. Dans Bostel et al. 2014, les auteurs proposent une formule d'approximation pour le CPP dirigé et non dirigé. Les arêtes sont réparties de manière aléatoire sur un carré unitaire et leur approximation est basée sur un modèle de régression utilisant uniquement le nombre de nœuds et d'arêtes.

Comme mentionné précédemment, les méthodes d'estimation des coûts se révèlent particulièrement pertinentes pour les problèmes d'optimisation comportant plusieurs niveaux de décision. En effet, les modèles d'approximation continue, plus rapides que les modèles discrets pour traiter de grandes instances, constituent un appui efficace aux décisions stratégiques et tactiques lorsqu'un choix de service est envisagé. Dans Francis et Smilowitz 2006, les auteurs introduisent un modèle d'approximation continue pour le problème de tournées de véhicules périodiques avec choix du service (PVRP-SC), une variante où la fréquence de visite des nœuds est une décision du modèle. Cette flexibilité peut améliorer l'efficacité des tournées et/ou la qualité du service. Le modèle d'approximation proposé soutient la planification stratégique et tactique des systèmes de distribution périodiques et fournit des indications pour élaborer des solutions au problème discret. Les modèles d'apprentissage automatique peuvent faire office de modèles d'approximations. Leur avantage est qu'ils peuvent s'adapter à des environnements changeants en étant ré-entraînés sur de nouvelles données. Dans Akkerman, Mes et Heijnen 2020 et Akkerman et Mes 2022, les auteurs intègrent un modèle d'approximation dans une approche en quatre phases combinant apprentissage *offline* et optimisation *online*. Dans la première phase, le modèle d'apprentissage est utilisé pour prendre des décisions de sélection de clients sur la base d'une estimation des coûts. La deuxième phase consiste à construire un

planning de tournées (*online*) pour desservir les clients sélectionnés. Les phases suivantes permettent de mettre à jour l'importance des anciennes données et de ré-entraîner le modèle d'approximation, qui sera à nouveau utilisé dans la première phase. Cette approche peut être appliquée à des environnements stables pour améliorer l'approximation, ou à des environnements dynamiques nécessitant des mises à jour régulières, par exemple lorsque la demande ou la zone géographique des clients change.

3.4.2 *Features* pour la description d'une instance

L'efficacité des approximations continues dépend fortement des caractéristiques ou des fonctionnalités utilisées pour décrire les instances du problème. Ces fonctionnalités vont des mesures de base, telles que le nombre de nœuds et les dimensions de la zone de service, à des mesures plus sophistiquées qui intègrent les relations spatiales et les informations heuristiques. Il est essentiel d'identifier les fonctionnalités appropriées pour les instances que nous souhaitons résoudre afin d'obtenir une approximation précise des solutions de routage. Elles peuvent dépendre du type de problème que nous traitons, des données disponibles ou même de la quantité que nous souhaitons approximer.

Dans la littérature, des études sur les caractéristiques ou *features* pour décrire des instances de tournées de véhicules sont proposées. Leur objectif ne consiste pas à fournir des estimations de coûts, mais à introduire des caractéristiques topologiques intéressantes qui pourraient aider à estimer des solutions de routage. Ces *features* vont des mesures de base, telles que le nombre de nœuds d'une instance et les dimensions de la zone de service, à des *features* plus sophistiquées qui intègrent les relations spatiales et des informations heuristiques. Ces caractéristiques sont essentielles pour décrire des instances de tournées de véhicules, pour approximer certaines valeurs à l'aide de méthodes d'apprentissage automatique ou encore aider à la sélection d'algorithmes de résolution. Elles peuvent dépendre du type de problème à traiter, des données disponibles ou même de la quantité que nous souhaitons approximer.

Dans Pihera et Musliu 2014, les auteurs identifient des *features* intéressantes pour le choix d'algorithme à utiliser pour résoudre un TSP. Ils classifient leurs *features* en trois catégories : des caractéristiques de recherche locale, des caractéristiques géométriques ou

encore des caractéristiques de plus proche voisin. Les caractéristiques de recherche locale sont dérivées de l'application de l'heuristique de Lin-Kernighan pour chaque instance, et prennent en compte à la fois son déroulement interne et ses résultats. La catégorie des *features* géométriques introduit de nouvelles caractéristiques fondées sur l'enveloppe convexe des sommets du graphe dans le plan euclidien. La catégorie de *features* de plus proches voisin nécessite une transformation du graphe de l'instance et permet de quantifier les relations locales entre les nœuds en se concentrant sur les k plus proches voisins de chaque point. Les *features* liées aux instances de TSP sont utilisées pour entraîner des modèles d'apprentissage automatique à prédire les performances de certains algorithmes d'heuristiques ou de métaheuristiques.

Dans Rasku, Kärkkäinen et Musliu 2016, les auteurs proposent une revue des *features* existantes pour la description d'instances de VRP. Ils introduisent également un ensemble de *features* pour améliorer la configuration de certaines algorithmes dans le choix des meilleurs paramètres ou encore le choix d'un algorithme plutôt qu'un autre pour résoudre le CVRP. Les *features* sont regroupées par type : les caractéristiques géométriques, les caractéristiques de plus proche voisin, des caractéristiques à propos de l'arbre recouvrant minimal, des caractéristiques sur la distribution des nœuds, etc. Leur objectif est d'identifier les types de problèmes et de mieux comprendre les propriétés des instances qui pourraient influencer leur solvabilité.

Arnold et Sörensen 2019 définit plusieurs *features* permettant de caractériser les solutions et les instances de VRP à l'aide d'exploration de données. De cette manière, ils sont capables de distinguer les bonnes solutions de VRP, de celles moins bonnes, à l'aide de caractéristiques comme la distance la plus longue entre deux clients connectés dans une tournée, ou encore la largeur moyenne d'une tournée. Ils concluent que les caractéristiques importantes permettant de distinguer des solutions de VRP quasi optimales de solutions non optimales sont notamment le nombre d'arêtes se croisant dans une tournée et leurs distances avec le dépôt.

3.5 Limites et perspectives de la littérature existante

La première partie de cet état de l'art a montré que de nombreux travaux portent sur des variantes du problème de tournées sur les arcs avec capacité (CARP) et, plus généralement, sur les problèmes de tournées sur les arcs. Le Tableau 3.1 en propose une synthèse comparative. Pour chaque problème, ce tableau précise le type de graphe considéré (U pour non orienté, D pour orienté, M pour mixte et W pour *windy*), le nombre de véhicules impliqués, ainsi que la présence éventuelle de contraintes de capacité, de modes multiples, de stations intermédiaires et de périodicité. Une colonne additionnelle mentionne d'autres caractéristiques spécifiques à certaines variantes.

La dernière ligne du tableau correspond au problème étudié dans ce manuscrit, noté *PCARP-MF* pour *Periodic Capacitated Arc Routing Problem with Multimodal Fleet*. Ce problème se distingue par la combinaison simultanée de contraintes de capacité, de périodicité, de stations intermédiaires et d'une flotte multimodale opérant sur des réseaux distincts. À notre connaissance, aucune des variantes recensées dans la littérature ne réunit l'ensemble de ces caractéristiques. Le problème périodique de tournées sur les arcs avec flotte multimodale, tel qu'il se pose dans le contexte de NaTran, constitue donc une contribution originale qui n'a pas encore été traitée.

Il ressort également de notre étude que, compte tenu de la complexité intrinsèque des problèmes PCARP, la grande majorité des méthodes proposées pour les résoudre sont des approches approchées. Ces méthodes se concentrent principalement sur le niveau opérationnel, à savoir la construction des tournées pour chaque période. Les décisions tactiques, telles que l'affectation des tâches aux périodes ou l'allocation des modes de transport, sont souvent traitées de manière séquentielle ou intégrées de façon simplifiée.

Dans le même temps, des approches récentes combinant recherche opérationnelle et apprentissage automatique ont démontré leur pertinence pour des problèmes de tournées de véhicules, comme l'attestent les revues de Bai et al. 2023 et Shahbazian et al. 2024. L'apprentissage automatique offre notamment une piste prometteuse pour estimer rapidement les coûts des solutions dans les problèmes d'optimisation comportant plusieurs niveaux de décision. Cependant, ces travaux portent quasi exclusivement sur des problèmes de tournées sur les nœuds. Les applications aux problèmes de tournées sur les arcs

demeurent très rares, et aucun modèle prédictif de coûts de tournées n'exploite, à notre connaissance, de caractéristiques (*features*) spécifiquement conçues pour les ARP.
Ces constats motivent l'étude approfondie du problème *PCARP-MF* rencontré chez NaTran, ainsi que le recours à une approche hybride combinant recherche opérationnelle et apprentissage automatique. Face à la complexité et à la dimension du problème, nous proposons dans les chapitres suivants une méthode de résolution structurée en deux niveaux. Le premier niveau, de nature tactique, est consacré à la planification des services et à l'allocation des modes de transport aux différentes périodes. Le second niveau, opérationnel, porte sur la construction effective des tournées de véhicules. Un modèle d'apprentissage automatique est développé pour assister la phase tactique en fournissant une approximation rapide des coûts de tournées, évitant ainsi un calcul systématique qui serait prohibitif en temps de calcul. Cette approche hybride, originale pour les problèmes de tournées sur les arcs, nécessite également l'introduction de nouvelles caractéristiques adaptées à la structure particulière des ARP.

Problème	Graphe	# véhicules	Capacité	Multi-modes	Stations intermédiaires	Périodicité	Autres	Références
CPP	U/D	1						Mei-Ko 1962
MCPP	M	1						Edmonds et Johnson 1973
WPP	W	1						Win 1987
HCPP	U	1						Dror, Stern et Trudeau 1987
MBCPP	U	1					profits maximaux	Pearn, Assad et Golden 1987
CPPTW	U	1					fenêtres temporelles	Aminu et Eglese 2006
RPP	U/D	1						Lenstra et Kan 1976
MRPP	M	1						Laporte 1997
WRPP	W	1						Minieka 1979
CARP	U	k	✓					Golden et Wong 1981
CARPDD	U	k	✓				capacité sur transit	Kirlik et Sipahioglu 2012
TCARP	U	k	✓				capacité = temps	Armas et al. 2019
OCARP	U	k	✓		✓			Usberti, P. M. França et A. L. M. França 2011
MDCARP	U	k	✓				multi dépôts	Yu, Liao et Yang 2023
CARPIF	U	k	✓		✓			Ghiani, Improta et Laporte 2001
CLARPIF	U	k	✓		✓			Ghiani, Guerriero et al. 2004
MCARPTIF	M	k	✓		✓			Willemse et Joubert 2016
MDHVCARP	M	k					flotte hétérogène	Cao et al. 2021
PRPP-IS	M	1				✓	services non réguliers	Benavent, Corberán, Laganà et al. 2019
PHCPP	D	1				✓		Keskin et Triki 2022
PCARP	M	k				✓		Lacomme, Prins et Ramdane-Chérif 2005
POCARP	U	k			✓	✓		Kansou et Kanso 2022
RTISP	D	k	✓		✓		fenêtres temporelles	Lannez et al. 2015 Lannez 2010
MDPOCARP	U	k			✓	✓		Kanso 2020
PCARP-MF	U	k	✓	✓	✓	✓	capacité sur transit	

Table 3.1 – Comparatifs des problèmes de tournées sur les arcs

Chapitre 4

Approche de résolution par décomposition

Comme indiqué dans le Chapitre 2, Le PCARP avec contrainte temporelle et flotte multimodale est NP-difficile car il est une extension du CARP qui est NP-difficle. Par conséquent une approche exacte ne peut pas résoudre ce problème en temps polynomial et risque d'être impraticable étant donnée la taille de notre problème. Dans ce chapitre, nous proposons donc une approche qui décompose le PCARP avec capacité temporelle et flotte multimodale en deux étapes. La première phase, dite de planification (niveau de décision tactique), consiste à allouer les portions aux semaines de l'année ainsi qu'à un mode de surveillance pour chaque semaine, la seconde phase, dite de routage (niveau de décision opérationnel), permet au sein de chaque semaine et pour chaque mode de surveillance de construire les tournées à réaliser. Cette structure s'appuie également sur des modèles prédictifs de coûts de tournées, basés sur des régressions linéaires, permettant d'évaluer rapidement les coûts de multiples combinaisons d'instances sans avoir à les résoudre explicitement, et ainsi de mieux répartir les portions de surveillance dans la phase de planification. L'approche est itérative afin d'améliorer la précision des modèles prédictifs au fur et à mesure des itérations, en fonction des résultats.

4.1 Principe de l'approche

Nous schématisons dans la Figure 4.1 le principe général de l'approche itérative en deux phases.

— **Étape 1 :** Nous fournissons en entrée l'ensemble des paramètres et des données répertoriés dans la Table 2.1, ainsi que les bases d'apprentissage initiales $\mathcal{B}_0^v = \{(\mathbf{X}_i^v, Y_i^v)\}_{i=1}^{N_0^v}$ et $\mathcal{B}_0^p = \{(\mathbf{X}_i^p, Y_i^p)\}_{i=1}^{N_0^p}$, correspondant respectivement aux tournées en voiture et en avion. Ces bases, composées de N_0^v et N_0^p observations $(\mathbf{X}_i^v, Y_i^v)$ et $(\mathbf{X}_i^p, Y_i^p)$, sont utilisées pour l'entraînement des modèles prédictifs associés. Les $\mathbf{X}_i^v$ et $\mathbf{X}_i^p$ représentent les vecteurs de caractéristiques décrivant chaque observation, tandis que les Y_i^v et Y_i^p correspondent au coût des tournées associées et constituent la valeur cible à prédire. Les coefficients de régression initiaux de ces modèles sont notés θ^v et θ^p pour les modes voiture et avion respectivement.

— **Étape 2 – Allocation des portions (*Module de Planification*) :** Le module de

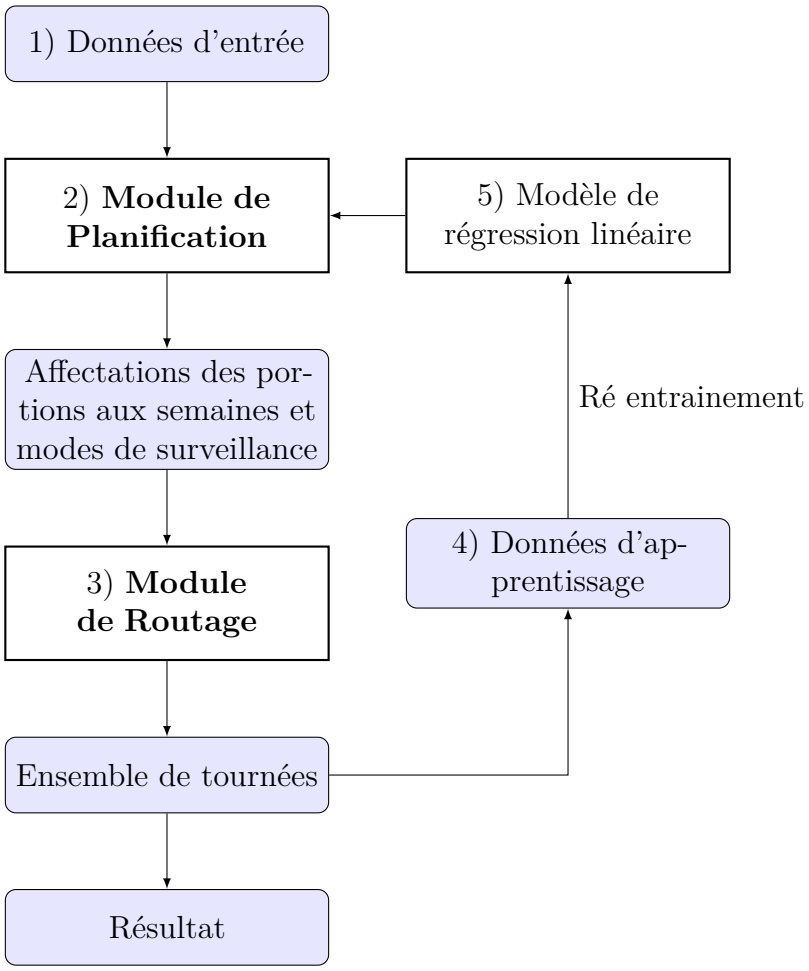


FIGURE 4.1 – Principe de l'approche par décomposition

planification alloue des portions aux semaines de l'année et aux modes de surveillance, en respectant les contraintes de planification et d'accessibilité des modes. Il s'appuie sur l'estimation des coûts des tournées fournie par les modèles de régression (θ^v et θ^p).

— **Étape 3 – Construction des tournées (*Module de Routage*) :** Le module de routage construit les tournées réelles pour chaque secteur, semaine et mode de surveillance, à partir des portions allouées. Il produit en sortie un ensemble de tournées ainsi que leurs coûts et modes utilisés, notés $\mathcal{T}^v_{w,s}$ pour la voiture et $\mathcal{T}^p_w$ pour l'avion.

— **Étape 4 – Mise à jour des bases d'apprentissage :** Les tournées obtenues sont

utilisées pour enrichir les bases d'apprentissage initiales, donnant les nouvelles bases $\mathcal{B}_i^v$ et $\mathcal{B}_i^p$ pour l'itération i.

— **Étape 5 – Ré-entraînement des modèles de régression :** Les modèles de régression sont ré-entraînés sur les nouvelles bases d'apprentissage selon différentes stratégies de mise à jour (par exemple, pondération plus forte des données récentes). Les coefficients θ^v et θ^p sont mis à jour à l'issue de cette étape.

Nous itérons entre les étapes 2 à 5 jusqu'à ce que les performances restent stables, ce que nous pouvons déterminer avec différents critères d'arrêts, comme un nombre fixe d'itérations, des performances qui ne s'améliorent plus significativement... À la fin, l'approche renvoie l'ensemble des tournées terrestres par secteur s et semaine w : $\mathcal{T}_{w,s}^v$, et l'ensemble des tournées aériennes par semaine w : $\mathcal{T}_w^p$

L'Algorithme 1 illustre le fonctionnement de l'approche itérative en deux phases.

Remarques additionnelles Il convient de noter que le *Module de Planification* est **multimodal**, dans la mesure où certaines portions requises peuvent être surveillées à la fois en voiture et en avion, tandis que dans le *Module de Routage*, la flotte est considérée comme **unimodale**, puisque ce module construit les tournées une fois que les arêtes ont été assignées à un mode.

Les sections suivantes décrivent le modèle de PLNE associé au problème de décision de chacun des modules.

4.2 Module de planification

Le module de planification fait partie du niveau de décision tactique. Il consiste à allouer des portions de surveillance aux différentes semaines de l'année et à un mode de surveillance pour chacune de ces semaines, en fonction des contraintes de planification ainsi que des contraintes de modes de surveillance. Dans cette section, nous commençons par présenter des notations, puis formulons le problème de planification sous forme d'un PLNE.

Nous introduisons les notations relatives à la phase de planification dans la Table 4.1. Les

variables de décision sont dans la Table 4.2. La planification s'organise sur l'ensemble des secteurs S couverts par un même prestataire de surveillance aérienne. Un ensemble d'arêtes R nécessite une surveillance, parmi lesquelles on distingue les arêtes $R^v(s)$ accessibles par voie terrestre pour un secteur s, et les arêtes R^p accessibles par voie aérienne. La phase de planification consiste à proposer des instances de tournées pour chaque semaine w de l'horizon de planification W, tout en respectant, pour chaque arête requise $[i, j] \in R$, sa fréquence de passage requise f_{ij} ainsi que les délais minimums $\Delta_{\min}[i, j]$ et maximums $\Delta_{\max}[i, j]$ entre deux surveillances successives. Le modèle de planification vise à minimiser un coût estimé, calculé en fonction des variables de décisions et des vecteurs de coefficients des modèles prédictifs au sol θ^v et en avion θ^p.

Ensembles	
S	ensemble des secteurs
$R^v(s)$	ensemble des arêtes à surveiller et accessibles en voiture dans le secteur $s \in S$
θ^v	vecteur des coefficients du modèle de régression linéaire pour les tournées au sol
R^p	ensemble des arêtes à surveiller accessibles en avion
θ^p	vecteur des coefficients du modèle de régression linéaire pour les tournées aériennes
W	ensemble des semaines de l'année
R	ensemble des arêtes où une surveillance est requise
Paramètres	
f_{ij}	fréquence de passage requise par année sur l'arête $[i, j] \in R$
$\Delta_{\min}[i, j]$	délai minimal entre deux surveillances successives de l'arête $[i, j] \in R$
$\Delta_{\max}[i, j]$	délai maximal entre deux surveillances successives de l'arête $[i, j] \in R$

TABLE 4.1 – Données d'entrée pour le module de planification

Variables	
x_{ij}^{skw}	variables binaires égales à 1 si et seulement si l'arête $[i, j] \in E^v(s)$ est surveillée de i vers j en voiture ou à pied lors de la tournée voiture $k \in K^v$ de la semaine $w \in W$ dans le secteur $s \in S$
y_{ij}^w	variables binaires égales à 1 si et seulement si l'arête $[i, j] \in E^p$ est surveillée lors de la semaine $w \in W$
λ_{ij}^w	variable binaire égale à 1 si et seulement si l'arête $[i, j] \in R$ est surveillée en voiture ou en avion la semaine $w \in W$

TABLE 4.2 – Variables pour le problème de tournées sur les arêtes

Fonction objectif

$$\text{minimiser} \qquad \sum_{s \in S} \sum_{w \in W} \hat{c}^{v,sw} + \sum_{w \in W} \hat{c}^{p,w} \tag{4.1}$$

La fonction objectif du module de planification 4.1 vise à minimiser la somme des coûts estimés $\hat{c}^{v,sw}$ des tournées terrestres par secteur s par semaine w, ainsi que la somme des coûts estimés $\hat{c}^{p,w}$ des tournées aériennes par semaine w.

Pour estimer ces coûts, nous nous basons sur l'article de BERGMAN et al. 2022 qui propose une méthode appelée JANOS, entre modélisation prédictive et analyse prescriptive. Ce *framework* permet d'incorporer directement des modèles d'apprentissage automatique dans un cadre d'optimisation. Les auteurs font notamment la distinction entre *features* prédites dépendantes des variables de décision du modèle et *features* fixes indépendantes des variables de décision.

Ces prédictions de coûts $\hat{c}^{v,sw}$ et $\hat{c}^{p,w}$, fournies par des modèles de régression linéaire, s'écrivent donc comme une combinaison linéaire de caractéristiques. Ces caractéristiques sont soit calculées à l'avance (« fixes ») soit dépendantes des variables de décision du modèle.

Le modèle d'estimation de coût pour les tournées terrestres $\hat{c}^{v,sw}$ (Équation 4.2)s'écrit comme la fonction f des variables de décision x_{ij}^{sw} du secteur s pour la semaine w, où f peut aussi être représentée comme une combinaison linéaire des coefficients de régression et des caractéristiques :

$$f\left(x_{ij}^{sw}\right) = \theta_0^v + \sum_{k=1}^{m_v} \theta_k^v X_{i,k}^v$$

Le modèle d'estimation de coût pour les tournées aériennes $\hat{c}^{p,w}$ (Équation 4.3) s'écrit comme la fonction g des variables de décision y_{ij}^{w} pour la semaine w, où g peut aussi être représentée comme une combinaison linéaire des coefficients de régression et des caractéristiques :

$$g\left(y_{ij}^{w}\right) = \theta_0^p + \sum_{k=1}^{m_p} \theta_k^p X_{i,k}^p$$

Dans ces expressions, $\boldsymbol{\theta}^v = (\theta_0^v, \theta_1^v, \ldots, \theta_{K_v}^v)$ et $\boldsymbol{\theta}^p = (\theta_0^p, \theta_1^p, \ldots, \theta_{K_p}^p)$ représentent les vecteurs de coefficients de régression, tandis que $\mathbf{X}_i^v = (X_{i,1}^v, \ldots, X_{i,K_v}^v)$ et $\mathbf{X}_i^p = (X_{i,1}^p, \ldots, X_{i,K_p}^p)$ sont les vecteurs de caractéristiques (*features*) associés à chaque observation.

$$\hat{c}^{v,sw} = f\left(x_{ij}^{sw}\right) \qquad \forall s \in S, \quad \forall w \in W \tag{4.2}$$

$$\hat{c}^{p,w} = g\left(y_{ij}^{w}\right) \qquad \forall w \in W \tag{4.3}$$

Contraintes pour le modèle de régression : Les contraintes principales auxquelles est soumis le modèle sont des contraintes de planification.

$$\sum_{s \in S} x_{ij}^{sw} + y_{ij}^{w} = \lambda_{ij}^{w} \qquad \forall w \in W, \quad \forall [i,j] \in R \tag{4.4}$$

$$\sum_{w \in W} \lambda_{ij}^{w} = f_{ij} \qquad \forall [i,j] \in R \tag{4.5}$$

$$\sum_{m=w-1}^{w+\Delta_{\min}[i,j]-2} \lambda_{ij}^{m \bmod 52+1} \leq 1 \qquad \forall w = 1,2,\ldots,52, \quad \forall [i,j] \in R \tag{4.6}$$

$$\sum_{m=w-1}^{w+\Delta_{\max}[i,j]-2} \lambda_{ij}^{m \bmod 52+1} \geq 1 \qquad \forall w = 1,2,\ldots,52, \quad \forall [i,j] \in R \tag{4.7}$$

$$x_{ij}^{sw} = 0 \qquad \forall s \in S, \quad \forall w \in W, \quad \forall [i,j] \in R \setminus R^{v}(s) \tag{4.8}$$

$$y_{ij}^{w} = 0 \qquad \forall w \in W, \quad \forall [i,j] \in R \setminus R^{p} \tag{4.9}$$

Les contraintes 4.4 permettent de faire le lien entre les surveillances aériennes et les surveillances terrestres. Les contraintes 4.5 permettent de respecter le nombre de surveillances par arête $[i,j] \in R$. Les contraintes 4.6 et 4.7 permettent respectivement de respecter le nombre de semaines minimal et maximal entre deux visites consécutives. Les contraintes 4.8 garantissent que le mode de surveillance terrestre n'est appliqué qu'aux arêtes accessibles par voie terrestre. Les contraintes 4.9 imposent que le mode de surveillance aérien soit uniquement affecté aux arêtes accessibles en vol.

Domaine des variables Les domaines des variables de décision sont donnés ci-dessous :

$$x_{ij}^{sw} \in \{0,1\} \qquad \forall s \in S, \quad \forall w \in W, \quad \forall [i,j] \in R \tag{4.10}$$

$$y_{ij}^{w} \in \{0,1\} \qquad \forall w \in W, \quad \forall [i,j] \in R \tag{4.11}$$

$$\lambda_{ij}^{w} \in \{0,1\} \qquad \forall w \in W, \quad \forall [i,j] \in R \tag{4.12}$$

La variable x_{ij}^{sw} est égale à 1 si et seulement si la surveillance de l'arête $[i,j] \in R$ est

planifiée en voiture pour le secteur s au cours de la semaine w, et 0 sinon. La variable y_{ij}^w est égale à 1 si et seulement si la surveillance de l'arête $[i,j] \in R$ est planifiée en avion au cours de la semaine w, et 0 sinon. La variable λ_{ij}^w est égale à 1 si et seulement si la surveillance de l'arête $[i,j] \in R$ est planifiée pour la semaine w, et 0 sinon.

4.3 Module de tournées

Le module de tournée est constitué de deux modèles de PLNE : l'un pour le mode de surveillance terrestre, l'autre pour le mode de surveillance aérienne. Les données d'entrée du module sont un ensemble de portions à visiter correspondant à une semaine donnée et un mode donné. Le module renvoie alors la ou les tournée(s) correspondante(s).

4.3.1 Tournées au sol

Nous présentons ci-dessous le modèle concernant les tournées terrestres. C'est un modèle PLNE pour résoudre le problème de CARP-DD ou CARP-DD avec des multiples dépôts.

Fonction objectif L'objectif est de minimiser les coûts selon la fonction objectif 4.13, qui calcule d'un côté les coûts liés au passage et de l'autre les surcoûts liés à la surveillance.

$$\text{minimiser} \quad \sum_{k \in K^v} \sum_{[i,j] \in E^v(s)} \frac{c^v}{60} t_{ij}^v (x_{ij}^k + x_{ji}^k) + \sum_{k \in K^v} \sum_{[i,j] \in R^v(s)} \frac{c^v}{60} (t_{ij}^{v\,\prime} - t_{ij}^v) \alpha_{ij}^k \tag{4.13}$$

Contraintes au sol

$$\sum_{\substack{j \in X_s^v \\ d \in D^v(s)}} x_{dj}^k = 1 \qquad \forall k \in K^v \tag{4.14}$$

$$\sum_{\substack{i \in X_s^v \\ d \in D^v(s)}} x_{id}^k = 1 \qquad \forall k \in K^v \tag{4.15}$$

$$\sum_{j \in X_s^v} (x_{ij}^k - x_{ji}^k) = 0 \qquad \forall k \in K^v, \quad \forall i \in X_s^v \tag{4.16}$$

$$\sum_{[i,j] \in E^v(s)} t_{ij}^v (x_{ij}^k + x_{ji}^k) + \sum_{[i,j] \in R^v(s)} (t_{ij}^{v\,\prime} - t_{ij}^v) \alpha_{ij}^k \leq \delta^v \qquad \forall k \in K^v \tag{4.17}$$

$$\sum_{k \in K^v} \alpha_{ij}^k = 1 \qquad \forall [i,j] \in R^v(s) \tag{4.18}$$

$$x_{ij}^k + x_{ji}^k \geq \alpha_{ij}^k \qquad \forall k \in K^v, \quad \forall [i,j] \in R^v(s) \tag{4.19}$$

$$\sum_{i \in Q} \sum_{j \notin Q} (x_{ij}^k + x_{ji}^k) \geq \alpha_{q_1 q_2}^k \qquad \begin{array}{r} \forall k \in K^v, \quad \forall Q \subseteq X_s^v \setminus D^v(s) \\ \forall (q_1, q_2) \in R^v(s) \cap Q^2 \end{array} \tag{4.20}$$

Chaque tournée doit débuter et finir à un dépôt noté $d \in D_v^s$ selon les contraintes 4.14 et 4.15. Les contraintes 4.16 assurent la continuité des tournées. Une tournée au sol est limitée par un temps de δ^v minutes selon les contraintes 4.17. Les contraintes 4.18 permettent d'effectuer de la surveillance sur les arêtes requises. Les contraintes 4.19 imposent que toute arête surveillée doit être traversée. Enfin les contraintes 4.20 permettent d'éliminer les sous-tours.

Domaine des variables Les variables sont définies à l'aide des contraintes suivantes

$$x_{ij}^k,\ x_{ji}^k \in \{0,1\} \qquad \forall k \in K^v, \quad \forall [i,j] \in E^v \tag{4.21}$$

$$\alpha_{ij}^k \in \{0,1\} \qquad \forall k \in K^v, \quad \forall [i,j] \in R^v \tag{4.22}$$

La variable x_{ij}^k (respectivement x_{ji}^k) est égale à 1, si l'arête $[i,j] \in E^v$ est parcourue en voiture de i vers j (respectivement de j vers i) lors de la k-ième tournée, et 0 sinon. La variable α_{ij}^k est égale à 1, si l'arête $[i,j] \in R^v$ est surveillée en voiture lors de la k-ième tournée, et 0 sinon.

4.3.2 Tournées aériennes

Afin de résoudre le problème de tournées aériennes, nous considérons une version du CARP-DD intégrant des stations intermédiaires, que nous formulons sous la forme d'un modèle de PLNE.

Fonction objectif L'objectif est de minimiser les coûts selon la fonction objectif 4.23, qui calcule d'une part les coûts liés au passage et de l'autre les surcoûts liés à la surveillance.

$$\text{minimiser} \sum_{k \in K^p} \sum_{[i,j] \in E^p} \frac{c^p}{60} t_{ij}^p (y_{ij}^k + y_{ji}^k) + \sum_{k \in K^p} \sum_{[i,j] \in R^p} \frac{c^p}{60} (t_{ij}^{p\,\prime} - t_{ij}^p) \beta_{ij}^k \tag{4.23}$$

Contraintes aériennes

$$\sum_{j \in X^p} y_{d_0^p j}^1 = 1 \tag{4.24}$$

$$\sum_{\substack{j \in X^p \\ d \in D^p}} y_{dj}^k = 1 \qquad \forall k \in \{2, \cdots, |K^p|\} \tag{4.25}$$

$$\sum_{i \in X^p} y_{i d_0^p}^{|K^p|} = 1 \tag{4.26}$$

$$\sum_{\substack{i \in X^p \\ d \in D^p}} y_{id}^k = 1 \qquad \forall k \in \{1, \cdots, |K^p| - 1\} \tag{4.27}$$

$$\sum_{i \in X^p} y_{id}^k = \sum_{i \in X^p} y_{di}^{k+1} \qquad \forall k \in \{1, \ldots, |K^p| - 1\}, \quad \forall d \in D^p \tag{4.28}$$

$$\sum_{j \in X^p} (y_{ij}^k - y_{ji}^k) = 0 \qquad \forall k \in K^p, \quad \forall i \in X^p \tag{4.29}$$

$$\sum_{[i,j] \in E^p} t_{ij}^p (y_{ij}^k + y_{ji}^k) + \sum_{[i,j] \in R^p} (t_{ij}^{p\,\prime} - t_{ij}^p) \beta_{ij}^k \leq \delta^p \qquad \forall k \in K^p \tag{4.30}$$

$$\sum_{k \in K^p} \beta_{ij}^k = 1 \qquad \forall [i,j] \in R^p \tag{4.31}$$

$$y_{ij}^k + y_{ji}^k \geq \beta_{ij}^k \qquad \forall k \in K^p, \quad \forall [i,j] \in R^p \tag{4.32}$$

$$\sum_{i \in Q} \sum_{j \notin Q} (y_{ij}^k + y_{ji}^k) \geq \beta_{q_1 q_2}^w \qquad \begin{array}{r} \forall k \in K^p, \\ \forall Q \subseteq X^p \setminus D^p, \quad \forall [q_1, q_2] \in R^p \cap Q^2 \end{array} \tag{4.33}$$

Chaque tournée aérienne commence et se termine dans un dépôt $d \in D^p$ conformément aux contraintes 4.25 et 4.27. En particulier, les premières et dernières tournées débutent et s'achèvent au dépôt principal, comme l'imposent les contraintes 4.24 et 4.26. De plus, le dépôt d'une fin de tournée sera le dépôt de départ de la tournée suivante selon les contraintes 4.28. Les tournées aériennes sont limitées par une durée de δ^p minutes selon les contraintes 4.30. Une surveillance est requise pour les arêtes $[i,j] \in R^p$ via les contraintes 4.31, et les contraintes 4.32 imposent que toute arête surveillée doit être traversée.

Domaine des variables

$$y_{ij}^k,\ y_{ji}^k \in \{0,1\} \qquad \forall k \in K^p, \quad \forall [i,j] \in E^p \tag{4.34}$$

$$\beta_{ij}^k \in \{0,1\} \qquad \forall k \in K^p, \quad \forall [i,j] \in R^p \tag{4.35}$$

La variable y_{ij}^k (respectivement y_{ji}^k) est égale à 1, si l'arête $[i,j] \in E^p$ est parcourue en avion de i vers j (respectivement de j vers i) lors de la k-ième tournée, et 0 sinon. La variable β_{ij}^k est égale à 1, si l'arête $[i,j] \in R^p$ est surveillée en avion lors de la k-ième tournée, et 0 sinon.

Algorithm 1 Procédure d'apprentissage itératif pour la planification multimodale

Entrées :

Ensembles et paramètres du problème
$\mathcal{B}_0^v$ et $\mathcal{B}_0^p$: bases d'apprentissage initiales pour la voiture et l'avion
θ^v et θ^p : coefficients de régression initiaux des modèles prédictifs
W : ensemble des semaines
S : ensemble des secteurs
Critère de convergence ϵ ou nombre maximal d'itérations $T_{\max}$

Sortie :

$\mathcal{T}$: ensemble des tournées planifiées par semaine, secteur et mode

Initialiser $t \leftarrow 0$ et $convergence \leftarrow false$
Tant que not $convergence$ **et** $t < T_{\max}$ **faire**
  $t \leftarrow t + 1$
  Planifier les tournées (w, s, m) via le *Module de Planification* utilisant les modèles prédictifs avec coefficients (θ^v, θ^p)
  Pour chaque semaine $w \in W$ **faire**
    Résoudre les tournées en avion pour la semaine w via le *Module de Routage* et mettre à jour $\mathcal{T}_w^p$
    Pour chaque secteur $s \in S$ **faire**
      Résoudre les tournées en voiture pour le secteur s de la semaine w via le *Module de Routage* et mettre à jour $\mathcal{T}_{ws}^v$
    Fin Pour
  Fin Pour
  Mettre à jour les bases d'apprentissage avec $\mathcal{T}_w^p$

$$\mathcal{B}_t^p \leftarrow \text{mise à jour de } \mathcal{B}_{t-1}^p \text{ avec nouvelles observations}$$

  Mettre à jour les bases d'apprentissage avec $\mathcal{T}_{ws}^v$

$$\mathcal{B}_t^v \leftarrow \text{mise à jour de } \mathcal{B}_{t-1}^v \text{ avec nouvelles observations}$$

  Réentraîner les modèles prédictifs avec les nouvelles bases d'apprentissage et mettre à jour les coefficients (θ^v, θ^p)
  Vérifier la convergence selon le critère ϵ (stabilité des performances)
Fin Tant que
retourner $\mathcal{T} = \{\mathcal{T}_{ws}^v, \mathcal{T}_w^p\}$, ensemble des tournées planifiées

4.4 Conclusion

Dans ce chapitre, nous avons proposé une approche bimodale en deux phases pour résoudre le PCARP avec capacité temporelle et flotte multimodale. Cette approche contient une phase de planification et allocation de modes et une phase de routage, pour lesquelles nous avons présenté des modèles de PLNE. Cependant, pour des raisons de temps d'exécution, il parait raisonnable de résoudre ces problèmes à l'aide de méthodes approchées. Nous proposons de telles méthodes pour la phase de routage et la phase de planification dans les chapitres 5 et 7 respectivement. Le chapitre 6, quant à lui, décrit une méthode d'approximation des durées des tournées permettant d'estimer rapidement le coût de l'affectation d'un ensemble de portions de surveillance à une période donnée, afin de guider la phase de planification.

Chapitre 5

Méthodes de construction des tournées

Dans ce chapitre, nous présentons différentes stratégies pour construire les tournées, correspondant au *Module de Routage* de l'approche itérative décrite dans le Chapitre 4. En pratique, la résolution exacte des instances de tournées à l'aide du modèle PLNE détaillé dans ce même chapitre peut s'avérer complexe en raison de leur nature combinatoire. De plus, le *Module de Routage* doit être exécuté de manière répétée (jusqu'à plusieurs milliers de fois par itération), ce qui impose le recours à des méthodes rapides, telles que des heuristiques, afin d'obtenir des solutions de bonne qualité dans des délais raisonnables.

Dans un premier temps, nous décrivons le processus de préparation des données, incluant la collecte des données sources et la construction des graphes, étape indispensable à la résolution des tournées. Une analyse topologique de ces graphes est ensuite réalisée afin de mieux comprendre leurs propriétés structurelles et d'orienter les choix de modélisation. Nous présentons ensuite les méthodes de réduction de la taille des instances, avant de détailler les différentes stratégies de résolution, exactes ou approchées. Nous ne nous préoccupons pas de contraintes de planification ici, l'objectif du *Module de Routage* étant de traiter des ARP issus de la phase de planification et d'assignation aux modes, tels que le TCARP, TCARP *with multiple depots*, et le TCARP *with intermediate facilities*. Enfin, la dernière section évalue et compare les performances des méthodes proposées.

Ce chapitre vise donc à identifier les compromis les plus pertinents entre précision et efficacité computationnelle, dans l'objectif de l'insérer dans l'approche globale.

5.1 Récupération des données

Dans cette section, nous détaillons notre processus de récupération de données, en commençant par la collecte des données sources puis le calcul des temps de passage sur les arêtes nécessaires à la construction des graphes complets. Nous présentons par la suite une analyse topologique générale des réseaux.

5.1.1 Collecte des données sources

Dans cette partie, nous décrivons le processus de récupération des données sources nécessaires à retrouver les ensembles indiqués dans le Chapitre 2 constituant le graphe de

notre problème. Ces données incluent des nœuds, correspondant aux extrémités de portions ou dépôts, ainsi que des arêtes reliant toutes les paires de nœuds possibles, correspondant à des portions ou non.

Le réseau de surveillance ainsi que l'ensemble des informations géographiques de NaTran sont conservées dans le SIG. Cependant le réseau de surveillance et l'ensemble des portions qui le composent ne sont pas structurés en graphe dans le SIG. Les différentes portions sont représentées comme des lignes mais ne sont pas topologiquement structurées, c'est-à-dire que les relations spatiales entre les lignes ne sont pas explicitement définies.

Pour récupérer les localisations des portions et les représenter en graphe, nous effectuons un géo-traitement afin d'obtenir pour chaque portion $[i, j] \in R$, les coordonnées de latitude et de longitude des deux extrémités, ainsi que la longueur réelle de la portion. Chaque portion est également associée à un secteur et un territoire d'appartenance. Nous avons également connaissance à l'aide du SIG des modes de surveillance préférentiels et/ou disponibles par portion. En dernier lieu, nous obtenons un fichier dans lequel chaque ligne correspond à un sommet, incluant ses coordonnées, la liste des portions auxquelles il est affilié, la liste des modes de surveillance possibles, ainsi que son secteur et son territoire.

De plus, nous récupérons dans le SIG les localisations des bases et sites pour chaque secteur. La majorité des secteurs ne disposent que d'une base unique, tandis que d'autres en possèdent deux ou trois. Les prestataires de surveillance aérienne ont également fourni la liste des aérodromes utilisés ainsi que leurs localisations. Le prestataire RECTIMO Air Transports opère depuis un aérodrome principal à partir duquel les avions sont déployés en début de semaine et où ils doivent revenir en fin de semaine. Au cours de la semaine, plusieurs autres aérodromes en France sont utilisés entre les différentes tournées. Le prestataire AIRLOISIR intervenant sur le territoire Val de Seine, opère depuis un aérodrome unique comme point de départ et d'arrivée des tournées.

L'ensemble des données sources récupérées est illustré dans la Figure 5.1.

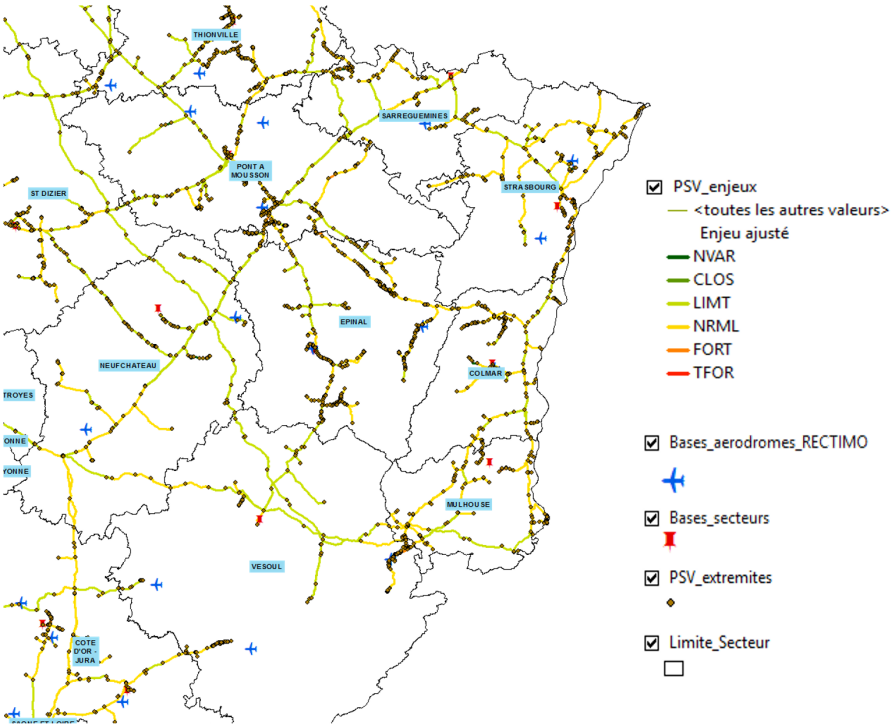


FIGURE 5.1 – Vue partielle du réseau NaTran présentant l'ensemble des données sources recueillies (Source : SIG de NaTran)

5.1.2 Construction du graphe et calcul des temps de parcours

Récupération des temps de passages et surveillance

Une fois les ensembles de nœuds et arêtes à surveiller retrouvés, nous souhaitons récupérer les différents paramètres du problème notamment ceux concernant les temps de passage t^v_{ij} et t^p_{ij} pour toute arête $[i,j] \in E$ et de surveillance $t^v_{ij}{}'$ et $t^p_{ij}{}'$ pour chaque arête $[i,j] \in R$.

Dans le cas des tournées aériennes nous considérons que les distances entre chaque nœud sont calculées à vol d'oiseau à l'aide de la distance géodésique d en kilomètres entre deux points, définie par :

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right)$$

avec r le rayon de la Terre (6 371 km), ϕ_1 et ϕ_2 les latitudes, et λ_1 et λ_2 les longitudes des deux points exprimées en radians.

Ces approximations restent proches des techniques empruntées par les prestataires de surveillance aérienne. Nous ne considérons pas les zones interdites de survol, les pilotes effectuent des détours qui seraient négligeables par rapport aux approximations faites. Par la suite, les distances sont ramenées à des temps de parcours et surveillance en fonction des vitesses de l'avion. Selon les données fournies par les prestataires de surveillance aérienne, les avions ont une vitesse de 230 km/h lorsqu'ils sont en phase de transit (volent « à vide », sans surveillance), et une vitesse de 130 km/h en cas de surveillance.

Concernant les tournées au sol, il est indispensable de prendre en compte l'utilisation du réseau routier pour déterminer les temps de passage. En effet, pour se déplacer d'un sommet à l'autre en voiture, il n'est pas réaliste de considérer les temps à vol d'oiseau. Il est nécessaire de tenir compte des vitesses spécifiques au réseau routier et des routes disponibles. De plus, toutes les portions de canalisations ne sont pas directement reliées au réseau routier. C'est pourquoi nous utilisons un service d'interface de programmation d'application ou *Application Programming Interface* (API) fourni par l'Institut national de l'information géographique et forestière (IGN) [1], qui permet de calculer un itinéraire en fonction des coordonnées géographiques (latitude et longitude) de deux ou plusieurs points. Avant le calcul, ces points sont d'abord automatiquement projetés sur le réseau routier par l'API. Il est également possible d'intégrer divers paramètres et contraintes, tels que le type de parcours ou le véhicule utilisé. L'API intègre des algorithmes de routage avancés pour déterminer les trajets optimaux entre les points spécifiés. Dans notre cas nous nous intéressons particulièrement à quatre résultats : le temps de parcours et le kilométrage du plus court chemin ainsi que le temps de parcours et le kilométrage du chemin le plus rapide. Pour les trajets de « transit » sur des arêtes ne nécessitant pas de surveillance, tels que les déplacements entre deux PSD ou entre une PSD et un dépôt, nous privilégierons le parcours le plus rapide. En revanche, pour les arêtes nécessitant une surveillance, nous opterons pour le trajet le plus court. Ces trajets sont généralement plus proches des canalisations et empruntent des routes où la vitesse est plus réduite. À noter

1. Pour plus d'information, consultez la documentation de l'API Calcul d'itinéraire de l'IGN via ce lien

que nous récupérons pour chaque secteur les temps pour chaque paire de sommets formant une arête, qu'elle soit requise ou non. Pour un secteur à n sommets, il faut récupérer les temps et distances pour $\frac{n(n-1)}{2}$ arêtes du graphe complet. Ce traitement peut prendre du temps pour les secteurs avec un grand nombre de sommets. C'est pour cela que nous avons parallélisé cette récupération de données.

Traitement des données

Une fois les matrices de distances et temps au sol récupérés, nous procédons à des ajustements pour corriger les erreurs détectées ou pour effectuer des ajustements nécessaires. Les traitements spécifiés ci-dessous ne concernent que les données pour les tournées au sol.

L'API utilisant des algorithmes distincts pour le trajet le plus court et le plus rapide, il peut arriver que le premier ait un temps de parcours inférieur au second. Cette incohérence engendrerait des coûts négatifs dans la fonction objectif. Pour y remédier, nous appliquons une règle de pré-traitement : si le temps du trajet court est inférieur à celui du trajet rapide, nous ajustons ce dernier à la valeur minimale, garantissant ainsi la cohérence des données. De plus la vitesse en voiture étant réduite lors de la surveillance nous rajoutons un facteur de 20 % au temps de parcours de portions de surveillance, conformément à ce qui est constaté en pratique.

Comme indiqué précédemment, avant de calculer les trajets les plus rapides ou les plus courts, l'API projette les sommets renseignés sur le réseau routier. Il peut arriver que ces sommets soient projetés sur le même point ce qui donne des temps et kilométrages égaux à zéro. Nous avons alors renseigné la taille de la portions comme distance et donné un temps minimal.

Enfin, l'API n'a pas pu déterminer d'itinéraires pour une vingtaine d'arêtes au total. Nous avons donc renseigné à la main les distances et temps de parcours manquants à l'aide de Google Maps.

5.2 Analyse topologique des réseaux

Dans cette section, nous analysons les données géographiques liées au réseau de transport de gaz opéré par NaTran. Toutes les informations fournies dans cette section correspondent aux données de l'année 2023. Le réseau comprend 32 634 kilomètres de réseau en France et est divisé en 20 325 PSD.

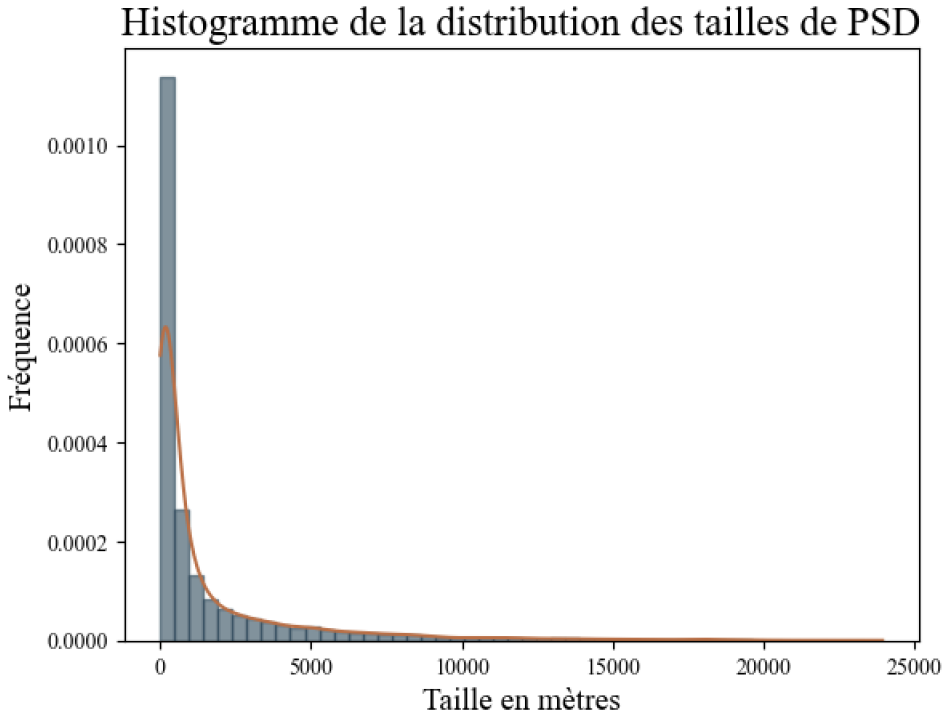


FIGURE 5.2 – Distribution des tailles de PSD en mètres

La Figure 5.2 présente la distribution des tailles en mètres de PSD. La plus petite PSD mesure 25 centimètres et la plus longue 23,94 kilomètres. La taille moyenne des PSD est de 1 600 mètres avec une médiane de 366 mètres. Cette distribution met en évidence une forte variabilité dans les tailles des PSD, allant de quelques dizaines de centimètres à plusieurs dizaines de kilomètres. L'écart important entre la moyenne et la médiane montre une distribution asymétrique, influencée par quelques très longues PSD.

De plus nous avons des données assez hétérogènes entre les différents secteurs.

La Figure 5.3 présente le réseau de transport de gaz et la répartition des PSD dans le secteur de Yonne. L'aire du secteur est de 8 707 km^2 et le périmètre de 846 kilomètres.

C'est le secteur avec le moins d'arêtes à surveiller (82 portions). La moyenne des tailles de portions est de 3 437 mètres, et la médiane de 2 033 mètres.

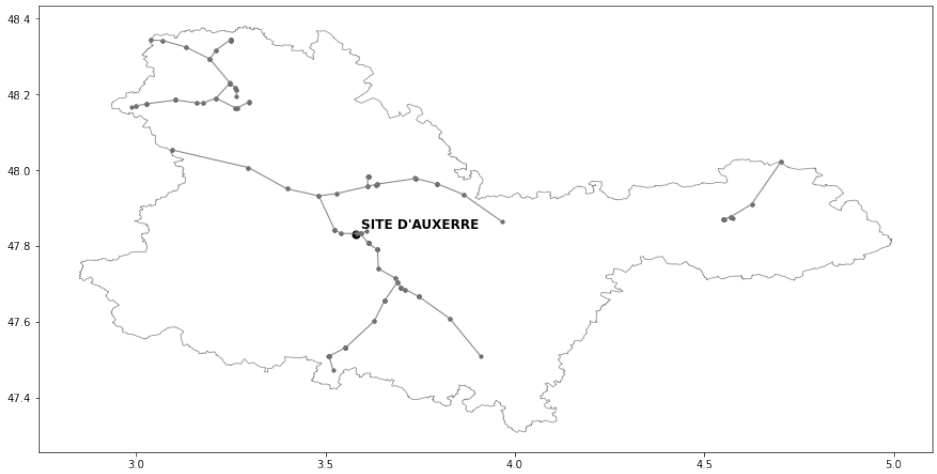


FIGURE 5.3 – Carte du secteur de YONNE (82 portions)

Dans la Figure 5.4 est représenté le secteur de Carvin. L'aire du secteur est de 2 043 km^2 et le périmètre de 411 kilomètres. Il représente le secteur avec le plus d'arêtes à surveiller (806). Cela est dû notamment à la taille plus petite des arêtes. La moyenne des tailles de portions est de 485 mètres, et la médiane de 192 mètres.

Nous pouvons également retrouver des secteurs contenant plusieurs dépôts au sol, c'est le cas du secteur de Basse Normandie présentant 3 dépôts pour 235 portions et dont les réseau est présenté dans la Figure 5.5. L'aire du secteur est de 17 601 km^2 et le périmètre de 1 388 kilomètres. La moyenne des tailles de portions est de 4 046 mètres, et la médiane de 2 138 mètres.

L'analyse topologique révèle que les portions de canalisation sont fortement interconnectées, conférant au réseau des arêtes requises une structure globalement connexe. De plus, il est fréquent que deux arêtes partagent une extrémité commune ou des extrémités géographiquement très proches. Ces caractéristiques structurelles seront exploitées dans la section suivante pour réduire la taille des instances.

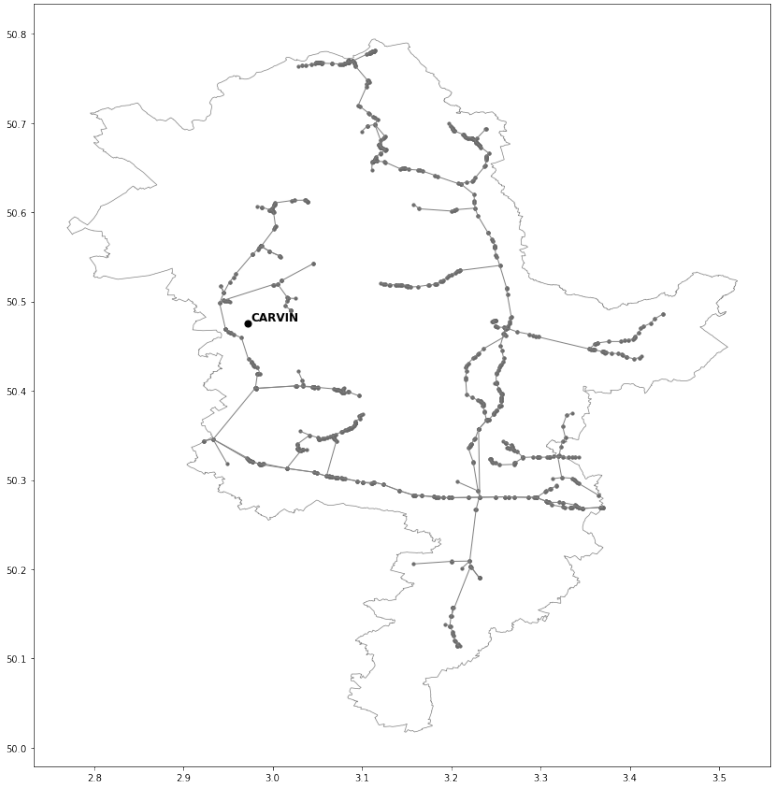


FIGURE 5.4 – Carte du secteur de CARVIN (806 portions)

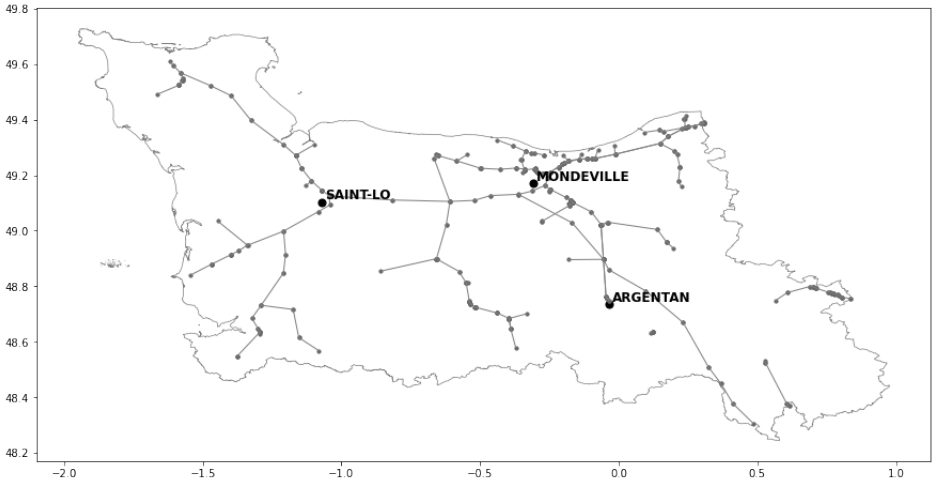


FIGURE 5.5 – Carte du secteur de BASSE NORMANDIE (235 portions et 3 dépôts)

5.3 Stratégies de réduction de la taille des instances

Cette section est consacrée à la gestion d'instances de grandes tailles pour la résolution de tournées où le modèle PLNE pour la résolution de tournées montre ses limites. En particulier pour les tournées aériennes servant un grand nombre d'arêtes. Pour pallier ces difficultés, nous proposons des techniques de réduction de taille d'instance, exploitant à la fois les propriétés structurelles du graphe sous-jacent et les caractéristiques spatiales des données, comme la connectivité des arêtes et la distribution géographique des sommets. Nous rappelons que les instances considérées correspondent à des instances à résoudre au niveau du *Module de Routage* ; par conséquent, il n'est pas nécessaire de considérer les fréquences de chaque portion.

Réduction du nombre de nœuds

Une première approche pour la réduction de la taille d'une instance est de réduire le nombre de nœuds dans cette même instance. En effet, certaines arêtes sont quasi connexes car elles ont chacune une de leurs deux extrémités qui sont très proches géographiquement. Bien que ces extrémités ne partagent pas les mêmes coordonnées GPS elles sont en réalité très proches et peuvent être considérées comme les mêmes sommets. L'idée est donc de réduire le nombre de nœuds de l'instance en regroupant les nœuds proches tout en conservant la structure topologique du graphe.

Nous identifions dans un premier temps toutes les paires de nœuds proches, c'est à dire se trouvant dans un même rayon qu'on peut faire varier (paramétrable de 10 à 100 mètres). Nous agrégeons ensuite ces sommets en un sommet unique. Nous éditons les arêtes requises en fonction de ces nœuds. Cette procédure permet notamment d'augmenter la connectivité des arêtes requises, ce qui peut améliorer l'efficacité des algorithmes de résolution.

Réduction du nombre d'arêtes

Une deuxième approche pour la réduction de la taille d'une instance est de diminuer le nombre d'arêtes requises dans cette même instance.

Dans une même instance de résolution, il est fréquent de retrouver plusieurs portions

de surveillance qui se succèdent. Notre idée est de les fusionner lorsque cela est possible, afin de former des « méga portions ». Nous considérons que deux portions peuvent être combinées ensemble si elles partagent exactement un sommet en commun. Ce sommet sera alors supprimé pour combiner les deux portions connexes.

Ainsi nous commençons par identifier tous les nœuds de degré 2 car ils peuvent être potentiellement supprimés. Itérativement nous supprimons un nœud de degré 2 et éditons la liste des portions requises jusqu'à ce qu'il n'y ait plus de sommet de degré 2. Chaque suppression de nœuds donne une fusion en une nouvelle arête requise, dont la durée est égale à la somme des temps des arêtes requises initiales qui la composent.

À noter que l'on peut également spécifier une distance maximale d'arêtes combinées, afin de limiter la taille des méga-portions. Toutefois, dans notre cas, cette contrainte n'a pas été activée, car nous spécifions une valeur seuil suffisamment élevée qui n'a pas eu d'impact significatif sur les résultats.

Pour réduire encore davantage le nombre d'arêtes, il est possible de les regrouper en composantes connexes formées par les arêtes requises. Ainsi chaque composante connexe peut être représentée comme une arête agrégée. Cependant cette méthode est plus approximative car elle nécessite de spécifier les nœuds d'extrémités de la portion agrégée.

Combinaison des deux réductions Pour maximiser l'efficacité de la réduction de la taille d'instances, nous pouvons combiner les deux méthodes de réductions mentionnées précédemment. Il est préférable de commencer par la réduction du nombre de nœuds, qui augmente la connectivité des arêtes requises. En effet, cela permet d'augmenter le nombre de sommets de degré 2 pour la seconde phase de réduction du nombre d'arêtes. L'enchaînement de ces deux méthodes permet de simplifier la structure de l'instance, tout en préservant les caractéristiques essentielles du problème.

5.4 Stratégies de résolution d'instances

Dans cette section, nous présentons les différentes stratégies utilisées pour résoudre les instances de tournées. Ces différents problèmes de tournées sur les arcs découlent du *Module de Routage* et gardent une complexité combinatoire, ce qui impose le recours à

différentes stratégies de résolution. Nous adoptons une stratégie globale de type *Route First, Cluster Second*, qui consiste à générer d'abord une tournée géante sans prendre en compte les contraintes de capacité, puis à la découper en sous-tournées en fonction des durées. Pour ce faire, nous mettons en œuvre plusieurs méthodes pour construire la tournée géante, puis différentes méthodes de découpage en sous-tournées en fonction de la nature du problème traité. L'objectif est de disposer de méthodes de résolution rapides, car le *Module de Routage* peut nécessiter la résolution de potentiellement des milliers d'instances de tournées par itération.

5.4.1 Méthodes de construction de la tournée géante

Nous abordons ici les méthodes de résolution, exactes ou approchées, pour générer une tournée géante. Pour cela on suppose, qu'on a un véhicule unique sans contrainte de capacité.

Résolution par PLNE

Pour résoudre exactement le problème de tournées à un véhicule, nous adaptons le modèle PLNE de tournées au sol présenté dans le Chapitre 4, pour un seul véhicule (donc en posant $k = 1$) et en relaxant les contraintes 4.17 de limite de temps. Le problème ainsi obtenu correspond à un RPP. Dans ce modèle, nous relaxons les contraintes d'élimination des sous-tours, dont le nombre croît fortement avec la taille des instances. Ces contraintes sont ensuite ajoutées progressivement, au fur et à mesure que les solutions obtenues révèlent des sous-tours à connecter à la tournée principale. Bien que la méthode exacte nous permette de résoudre des instances contenant jusqu'à quelques centaines d'arêtes requises, le RPP reste NP-difficile et la résolution de ce problème peut être coûteuse. Dans ce cas, pouvoir recourir à des méthodes approchées peut s'avérer utile.

Heuristique du plus proche voisin

L'heuristique du plus proche voisin ou *Nearest Neighbor Heuristic* (NNH) est une méthode constructive simple pour générer rapidement des tournées. Elle consiste à partir d'une arête requise initiale (choisie au hasard ou selon un critère) puis, à chaque étape, à

se déplacer vers le nœud de l'arête requise non encore visitée la plus proche, c'est-à-dire celle qui minimise la distance depuis la position actuelle. Une fois cette arête intégrée dans la tournée, on met à jour la position et on répète le processus jusqu'à ce que toutes les arêtes requises soient couvertes. Cette heuristique est rapide et facile à mettre en œuvre, mais elle se base uniquement sur des décisions locales, ce qui peut conduire à des tournées plus longues que celles obtenues avec des méthodes plus sophistiquées.

L'heuristique des k plus proches voisins (k-*NNH*) généralise l'approche classique du plus proche voisin. À chaque étape de la construction, au lieu de sélectionner systématiquement l'arête requise non visitée la plus proche, l'algorithme considère les k arêtes candidates les plus proches et en choisit une aléatoirement. Cette randomisation introduit une diversification qui permet d'explorer différentes solutions.

Heuristique de meilleure insertion

L'heuristique de meilleure insertion ou *Best Insertion Heuristic* (BIH) pour le RPP consiste à construire progressivement une tournée couvrant toutes les arêtes requises en insérant, à chaque étape, l'arête requise non encore visitée à l'endroit où son ajout augmente le moins le coût total. On part d'une tournée initiale contenant une ou plusieurs arête(s) requise(s), puis pour chaque arête requise restante on évalue tous les points d'insertion possibles dans la tournée actuelle : chaque insertion entraîne un surcoût composé du chemin vers l'arête, de la traversée de l'arête elle-même, puis du chemin de retour vers la tournée. Nous retenons l'emplacement, et le sens qui minimisent le surcoût. Ce processus est répété jusqu'à ce que toutes les arêtes requises soient intégrées, fournissant ainsi une solution rapide et généralement de bonne qualité, bien que non optimale. Comme pour NNH, la version k-BIH généralise la version classique de BIH en choisissant aléatoirement à chaque itération une position parmi les k de meilleur coût d'insertion. L'heuristique peut aussi être exécutée à partir d'une tournée initiale incluant 25 % d'arêtes choisies aléatoirement et obtenue par la résolution du modèle de PLNE. Avec un pourcentage de 25 % d'arêtes, les plus petites instances possèdent des tournées initiales composées de deux arêtes, et les plus grandes instances étudiées ici (200 arêtes) en comptent 50.

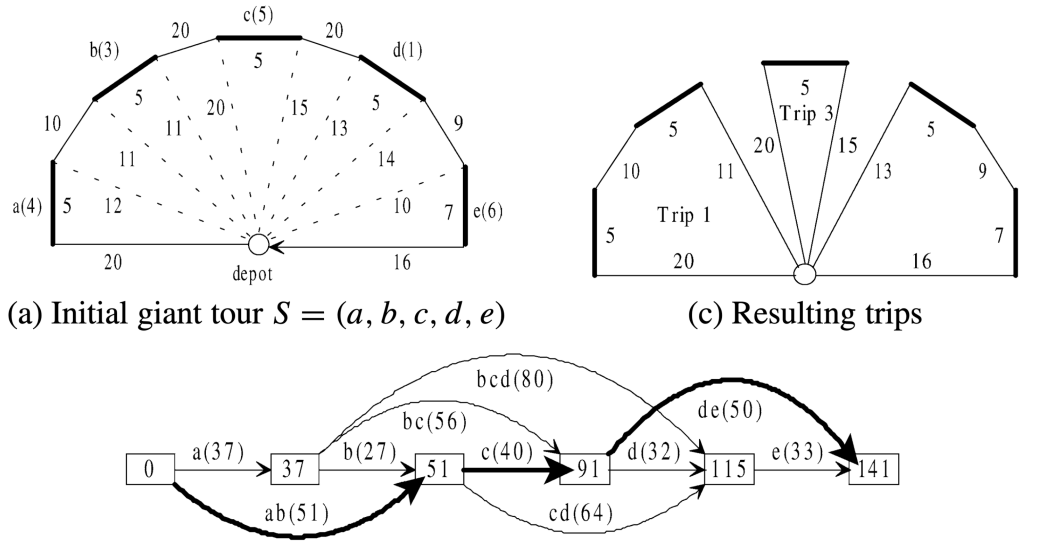
5.4.2 Méthodes de résolution de multi-tournées

Quelle que soit la méthode de résolution utilisée pour construire la tournée géante, celle-ci est ensuite découpée en plusieurs sous-tournées en fonction de la capacité considérée, ici le temps. Selon le problème étudié (tournées terrestres et tournées aériennes du prestataire AirLoisirs, ou tournées aériennes de RECTIMO Air Transports), on aura recours soit à la méthode *Split*, soit à l'ajout de stations intermédiaires au sein de la tournée géante.

Découpage en sous-tournées avec l'approche *Split*

La procédure Split, introduite par Ulusoy 1985, consiste à découper une tournée géante en plusieurs tournées faisables. De manière générale, cette approche construit un graphe auxiliaire orienté dans lequel chaque arête requise de la tournée géante est représentée par un nœud. Chaque arête de ce graphe correspond alors à une séquence réalisable d'arêtes requises dans la tournée géante (autrement dit, chaque arête du graphe auxiliaire représente une route RPP réalisable). Le plus court chemin reliant le dépôt à la dernière arête requise de la tournée géante fournit ainsi un ensemble de routes faisables qui couvrent l'ensemble des arêtes requises.

La Figure 5.6 illustre la procédure de découpe de tournée, appelée *Split*, pour une tournée géante $S = (a, b, c, d, e)$ avec les demandes entre parenthèses et les coûts de « transit », en supposant que la capacité $W = 9$. *Split* construit un graphe auxiliaire H avec $t + 1$ nœuds indexés à partir de 0. Chaque sous-séquence $(S_i, \cdots, S_j)$ correspondant à une tournée réalisable est modélisée par un arc $(i - 1, j)$, pondéré par le coût du trajet. Le chemin le plus court entre le nœud 0 et le nœud t dans H (en gras) indique la découpe optimale : 3 sous-tournées et un coût total de 141.



(a) Initial giant tour $S = (a, b, c, d, e)$ (c) Resulting trips

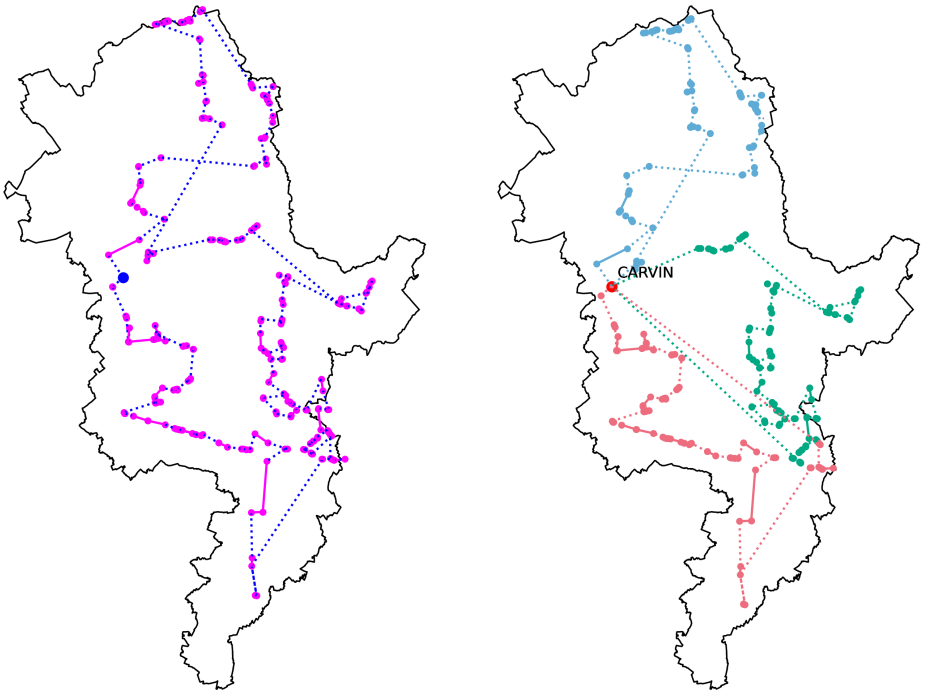
(b) Auxiliary graph and shortest path (labels in each node)

FIGURE 5.6 – Principe de l'algorithme *Split* (de LACOMME, PRINS et RAMDANE-CHERIF 2004 avec l'autorisation de Springer Nature.)

Une extension a été proposée par LACOMME, PRINS et RAMDANE-CHERIF 2004 pour résoudre le *Extented Capacitated Arc Routing Problem* (ECARP). Cette version traite notamment un multigraphe mixte et impose une limite de temps par tournée. Ils résolvent la tournée géante, un MRPP, à l'aide d'une heuristique : *Extended Path-Scanning* (EPS). Ils proposent alors une autre version de la méthode *Split*. La deuxième version de *Split*, intitulée « *Split Procedure, Version Minimizing Makespan Subject to a Limited Fleet* », est spécialement conçue pour minimiser le coût tout en respectant une contrainte de flotte limitée. Cette nouvelle version est par ailleurs adaptée pour les instances multi-dépôts. Cette version est notamment utilisée pour des problèmes de collecte des déchets. On dispose d'une flotte de véhicules limitée : on peut en utiliser plusieurs, mais jamais plus que la taille de la flotte. Tous les coûts sont des temps et l'objectif est de minimiser le temps de traitement (durée du trajet la plus longue). L'algorithme cherche un chemin particulier dans un graphe auxiliaire construit pour représenter toutes les possibilités de regroupement d'arcs. À chaque itération, il calcule les meilleurs enchaînements d'arcs possibles pour un nombre donné de tournées, en mettant à jour des tableaux qui stockent les meilleurs temps trouvés et le nombre de tournées nécessaires. Il s'arrête soit lorsque les valeurs ne changent

plus, soit lorsque le nombre maximal de véhicules a été atteint. Si l'algorithme trouve une valeur infinie pour le dernier nœud, cela signifie que la solution testée n'est pas réalisable. Sinon, il fournit le temps minimal obtenu et le nombre de tournées réellement utilisées.

Par exemple, la Figure 5.7 illustre la comparaison entre une tournée réalisée en une tournée géante et une répartition sur trois véhicules pour 155 portions choisies aléatoirement dans le secteur de Carvin. La Figure 5.7a présente la tournée unique, dont la durée est de 1 008 minutes, tandis que la Figure 5.7b montre le découpage obtenu avec *Split* en trois tournées totalisant 1 100 minutes.



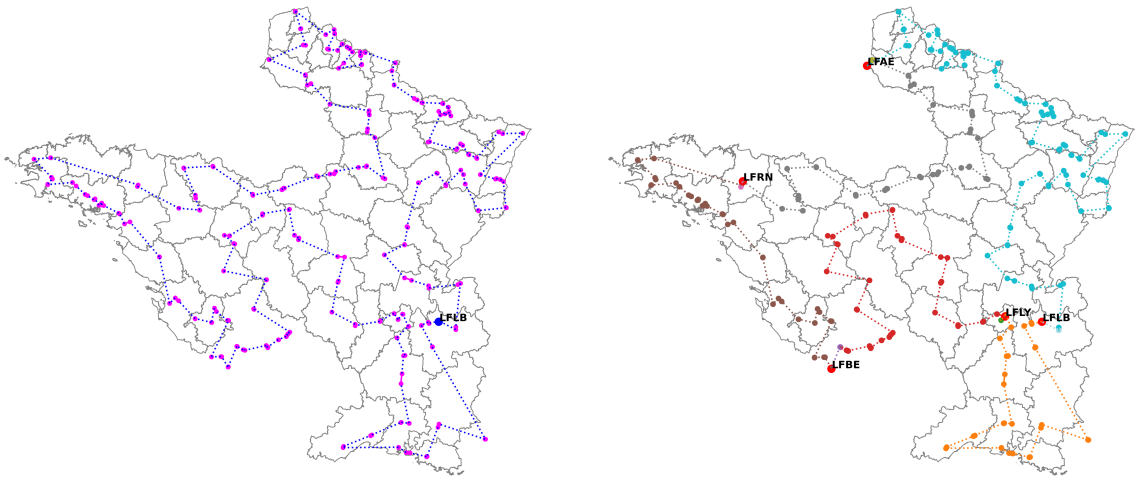
(a) Mono tournée (1 008 minutes) (b) Multi tournées (1 100 minutes)

FIGURE 5.7 – Mono tournée vs multi tournée (155 portions)

Ajout des stations intermédiaires

Pour la surveillance aérienne prestée au groupe RECTIMO Air Transports, plusieurs aérodromes sont disponibles tout au long de la semaine. Cependant, la première tournée de la semaine doit partir de l'aérodrome de Chambéry, situé dans le sud-est de la France, et la

dernière tournée doit également s'y terminer. Comme pour les autres modes de surveillance, nous commençons par générer une tournée géante ayant comme point de départ et d'arrivée cet aérodrome principal. Ensuite, en fonction de la durée totale de la tournée géante, nous estimons le nombre minimal d'aérodromes intermédiaires à insérer dans la solution. Nous parcourons ensuite la tournée dans l'ordre et identifions la première position à partir de laquelle la limite de temps d'une sous-tournée est dépassée. À partir de ce point, nous évaluons l'insertion d'un aérodrome dans les l dernières positions précédant le dépassement (avec $\ell = 5$ dans nos expérimentations). Nous choisissons ensuite l'aérodrome intermédiaire et la position d'insertion qui minimisent conjointement le temps additionnel ajouté à la tournée au total. Nous répétons ce processus à partir du dernier aérodrome inséré jusqu'à ce que les temps de chaque sous-tournées respectent la limite de durée en veillant également à ce que le temps de la dernière sous-tournée respecte cette contrainte. Un exemple de découpage de tournée aérienne est donné dans la Figure 5.8.



(a) Mono tournée (1 607 minutes) (b) Multi tournées (1 624 minutes)

FIGURE 5.8 – Mono tournée vs multi tournée en avion (150 portions)

5.5 Évaluation et comparaison des méthodes de résolution

Dans cette section, nous évaluons et comparons les différentes méthodes de résolution du *Module de Routage*, sur un ensemble varié d'instances de tournées. Nous présentons tout d'abord le protocole expérimental, puis les résultats obtenus avec le modèle exact, suivis d'une comparaison globale de l'ensemble des méthodes.

5.5.1 Protocole expérimental

Description du jeu d'instances Nous allons décrire les méthodes de simulation d'instances de calcul dans le Chapitre 6. Nous distinguons d'une part les instances formées de portions géographiquement regroupées, qualifiées de *clusterisées*, et d'autre part celles construites à partir de portions choisies aléatoirement. Pour chacune de ces deux catégories, 2 760 instances ont été générées, contenant entre 5 et 200 portions pour chaque secteur.

Environnement de calcul L'ensemble des expérimentations a été réalisé sur une machine virtuelle Amazon WorkSpace équipée d'un processeur *Intel Xeon Platinum 8259CL* à 2.50 GHz, comportant 4 cœurs physiques et 8 processeurs logiques, et disposant de 32 Go de mémoire vive. Le système d'exploitation utilisé est *Windows Server 2016*.

Les modèles de programmation linéaire ont été implémentés en `Python 3.9.21` à l'aide de la bibliothèque `docplex 2.30`, et résolus avec le solveur `CPLEX 22.0.1`.

Définitions des indicateurs Afin d'évaluer et de comparer les différentes méthodes de résolution proposées, nous définissons plusieurs indicateurs de performance permettant de mesurer à la fois la qualité des solutions obtenues et l'efficacité computationnelle des approches.

Le *temps de résolution* correspond à la durée totale nécessaire pour obtenir une solution faisable ou optimale à une instance donnée. Il est mesuré en secondes à partir du lancement de l'algorithme jusqu'à son arrêt, que celui-ci provienne d'un critère d'optimalité atteint ou d'une limite de temps imposée.

L'*écart à l'optimalité* permet de mesurer la différence entre la valeur de la solution trouvée par une méthode approchée et la solution optimale (ou la meilleure borne connue). Il est défini par :

$$\text{Gap}(\%) = \frac{z_{\text{méthode}} - z_{\text{opt}}}{z_{\text{opt}}} \times 100,$$

où $z_{\text{méthode}}$ désigne le coût de la solution fournie par la méthode considérée, et z_{opt} la valeur optimale ou la meilleure borne connue. Un écart nul indique une solution optimale.

5.5.2 Résultats du modèle exact (PLNE mono-véhicule)

Dans cette partie nous analysons les temps de résolution du RPP à l'aide du modèle PLNE. Nous comparons aussi les temps entre les instances générées à partir de portions regroupées et celles générées avec une sélection aléatoire des portions du secteur.

La Figure 5.9 présente les temps de résolution, en secondes, pour des instances de tournées avec des portions choisies aléatoirement, en fonction du nombre d'arêtes requises qu'elles contiennent. Comme précédemment, nous remarquons que les temps de résolution augmentent avec le nombre de portions à surveiller. Toutes les instances contenant jusqu'à 60 portions sont résolues en moins d'une minute. Le temps de résolution maximal était de 1 929 secondes pour une instance à 200 portions.

La Figure 5.10 présente les temps de résolution, en secondes, pour des instances de tournées générées à partir de clusters de portions, en fonction du nombre d'arêtes requises qu'elles contiennent. On observe que les temps de résolution augmentent avec le nombre de portions à surveiller. Toutes les instances contenant jusqu'à 75 portions sont résolues en moins d'une minute. Le temps de résolution maximal est de 683 secondes pour une instance à 165 portions

Dans la Figure 5.11, nous comparons les temps de résolution des instances en fonction de leur méthode de simulation. Pour les deux méthodes de simulation nous avons les mêmes nombres de portions à inclure. Nous remarquons que les temps de résolution sont globalement plus élevés pour les instances générées aléatoirement, avec une moyenne de temps de résolution de 104 secondes, alors que les instances générées par cluster ont un temps de résolution moyen de 51 secondes. Cela s'explique notamment par la structure

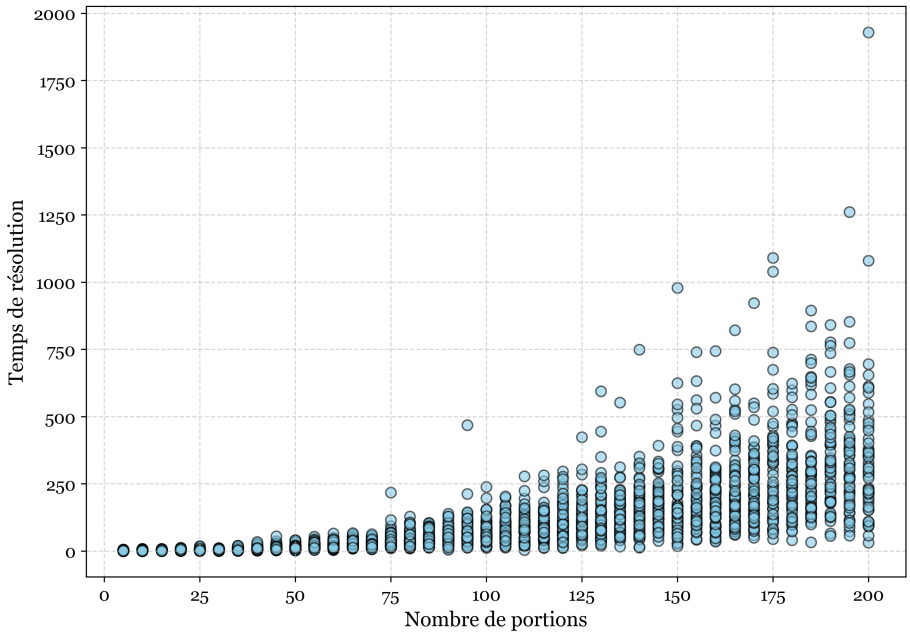


FIGURE 5.9 – Temps de résolution en fonction de la taille des instances générées aléatoirement

plus connectée des graphes issus des clusters de portions, comparée à celle des instances générées aléatoirement.

5.5.3 Comparaison globale

Dans la Table 5.1, nous comparons les temps de résolution et les écarts à l'optimalité (*Gap*) des différentes méthodes approchées par rapport à la méthode exacte du PLNE pour des mono-tournées en voiture. Pour l'heuristique k-NNH, nous l'exécutons avec des valeurs de k variant de 1 à 3, en réalisant 10 lancements pour chaque valeur, puis nous conservons la meilleure solution obtenue. Pour l'heuristique k-BIH, nous procédons également à 10 lancements pour chaque valeur de k et retenons la meilleure solution.

La méthode exacte nécessite en moyenne 77,88 secondes pour résoudre les instances étudiées. Après réduction de leur taille, le temps moyen tombe à 37,83 secondes, soit un gain de 51,43 % avec un écart à l'optimalité moyen de 1,18 % ce qui confirme l'efficacité de cette méthode. L'écart-type de 1,72 % montre une performance stable entre les différentes instances. L'heuristique k-NNH est la plus rapide avec un temps moyen de 0,16 secondes,

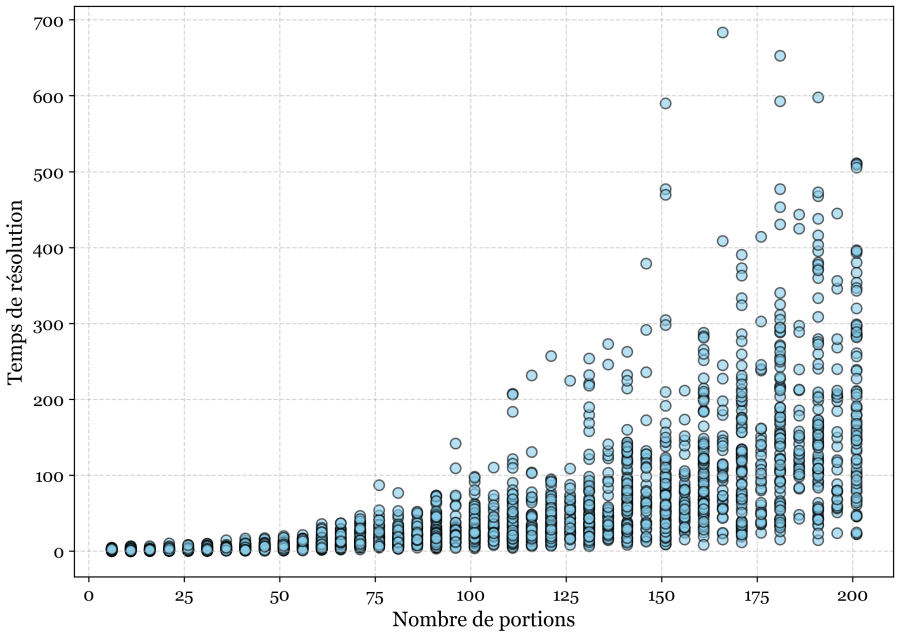


FIGURE 5.10 – Temps de résolution en fonction de la taille des instances *clusterisées*

réduit à 0,11 secondes lorsque nous appliquons les méthodes de réduction de taille d'instances. Elle peut en plus être traitée parallèlement. Cependant, cette heuristique présente un écart moyen à l'optimalité compris entre 11,68 % et 11,86 % avec un écart type entre 6,27 et 6,37 %. L'heuristique k-BIH présente un écart à l'optimalité inférieur à k-NNH (5,76 %) et est plus stable car l'écart type moyen de l'écart à l'optimalité est plus faible (3,52 %) avec un temps moyen de 6,8 secondes. Si l'on initialise k-BIH avec un tour initial incluant 25 % des arêtes, l'heuristique obtient de meilleurs résultats, avec un écart moyen à l'optimalité de 5,66 %, un écart-type moyen de 3,33 % et un temps de résolution moyen de 7,19 secondes.

Dans la Table 5.2, nous présentons les temps moyens de résolution ainsi que les écarts moyens à l'optimalité, en distinguant les différentes tailles d'instances. Ces comparaisons servent d'appui pour l'approche itérative : elles permettent d'adapter la méthode de résolution des tournées en fonction de la taille des instances, mais également en fonction des compromis acceptables entre temps de calcul et qualité des solutions. Nous remarquons que les gains moyens augmentent avec la taille des instances, car le temps de calcul de la méthode exacte augmente beaucoup plus rapidement que celui des heuristiques. Par ailleurs,

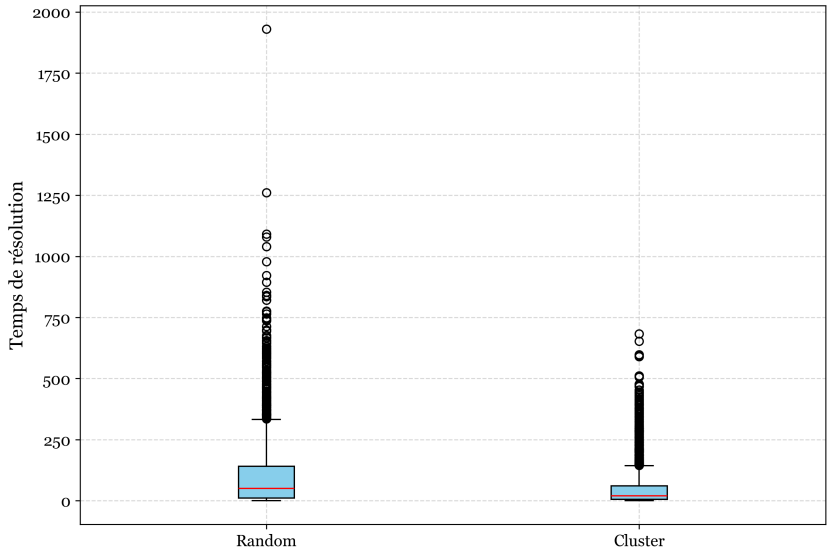


FIGURE 5.11 – *Box plot* comparatif des temps de résolution entre les instances simulées aléatoirement et celle simulées par proximité

TABLE 5.1 – Statistiques agrégées sur 5 520 instances

Méthode	**Gap moyen (%)**	**Gap std (%)**	**Gap max (%)**	**Temps moyen (s)**	**Gain moyen (%)**
PLNE	0,00	0,00	0,00	77,88	–
Réduction + PLNE	1,18	1,72	18,82	37,83	51,43
k-NNH	11,86	6,27	34,48	0,16	99,79
Réduction + k-NNH	6,37	11,68	32,37	0,11	99,86
k-BIH	5,76	3,52	35,57	6,80	91,27
Tour initial + k-BIH	5,66	3,33	35,36	7,19	90,77

lorsque la taille des instances augmente, la résolution du PLNE devient particulièrement coûteuse, cela rend le recours à des heuristiques encore plus intéressant.

5.6 Conclusion

Dans ce chapitre, nous avons présenté l'ensemble du processus, depuis la collecte des données jusqu'à leur préparation pour la construction des matrices de distances et de durées. Nous avons décrit le traitement de ces données ainsi que l'exploitation de la structure du graphe pour en réduire la taille des instances, notamment par le regroupement de nœuds ou

d'arêtes requises. Nous avons également proposé plusieurs méthodes, exactes ou approchées, permettant de résoudre dans un premier temps une tournée géante avant de la décomposer en sous-tournées. Le recours à des méthodes approchées s'avère particulièrement nécessaire pour les instances de grande taille ou lorsque des résultats rapides sont requis. C'est notamment le cas des tournées aériennes, dont les instances peuvent comporter plusieurs milliers d'arêtes requises. Mais c'est aussi essentiel au sein de la méthode itérative en deux phases qui nécessite la résolution de milliers de tournées à chaque itération. Pour le *Module de Routage*, nous disposons de plusieurs méthodes de résolution de tournées. Dans le cadre de l'approche globale, le choix de la méthode sera adapté en fonction du temps disponible et de la taille des instances.

TABLE 5.2 – Évolution des performances selon la taille des instances

Taille	**Méthode**	**Gap moyen (%)**	**Temps moyen (s)**	**Gain moyen (%)**
5–50	PLNE	0,00	4,59	–
	Réduction + PLNE	0,55	4,09	10,89
	k-NNH	7,80	0,02	99,56
	Réduction + k-NNH	7,46	0,02	99,56
	k-BIH	1,84	0,30	93,46
	Tour initial + k-BIH	2,38	1,83	60,13
50–100	PLNE	0,00	28,90	–
	Réduction + PLNE	1,11	18,27	36,78
	k-NNH	12,25	0,08	99,72
	Réduction + k-NNH	12,16	0,06	99,79
	k-BIH	5,30	1,90	93,43
	Tour initial + k-BIH	5,34	3,15	89,10
100-150	PLNE	0,00	87,81	–
	Réduction + PLNE	1,41	46,29	47,28
	k-NNH	13,52	0,19	99,78
	Réduction + k-NNH	13,45	0,14	99,40
	k-BIH	7,41	7,30	91,69
	Tour initial + k-BIH	7,00	7,28	91,71
150+	PLNE	0,00	205,68	–
	Réduction + PLNE	2,46	67,82	67,02
	k-NNH	14,05	0,36	99,83
	Réduction + k-NNH	13,85	0,24	99,88
	k-BIH	8,82	19,40	90,57
	Tour initial + k-BIH	8,21	17,77	91,36

Chapitre 6

Modèles prédictifs pour l'estimation des coûts

Dans ce chapitre, nous décrivons notre méthode d'approximation de durées de tournées. L'objectif de ces méthodes est de prédire, très rapidement, pour une instance donnée, le coût de la tournée optimale couvrant cette instance mais sans la résoudre explicitement. Ces modèles sont particulièrement pertinents dans des problèmes bi-niveaux, comme le nôtre, pour faire l'interaction entre les niveaux tactique et opérationnel. L'objectif d'un tel modèle dans nos travaux, est d'assister la phase de planification, en estimant rapidement le coût d'une instance donnée par rapport à d'autres, pour mieux allouer les portions de surveillance aux différentes périodes de l'horizon de planification. Si le module de planification est écrit comme un PLNE, il est nécessaire de pouvoir formuler le modèle d'approximation sous une forme linéaire pour l'intégrer. Dans notre problème, le coût dépend à la fois de la durée et du coût horaire associé à chaque mode ; nous choisissons donc de concentrer la prédiction sur la durée de la tournée. Tout au long de ce chapitre, les notions de *coût* et de *durée* seront considérées comme équivalentes et utilisées de manière interchangeable.

Les méthodes d'apprentissage automatique en particulier les modèles régressifs sont des bons outils pour prédire rapidement pour une observation donnée, une variable quantitative en fonction d'un ensemble de caractéristiques descriptives. L'idée consiste à disposer d'un ensemble d'instances de tournées déjà résolues, avec leurs coûts associés, afin d'entraîner des modèles capables d'apprendre à prédire ces coûts à l'aide d'un ensemble de caractéristiques associées. Notre approche, dans ce chapitre, inclut la génération de tournées fictives pour constituer une base d'apprentissage au modèle, la création de caractéristiques ou *features* pour la description des instances générées ainsi que la construction des modèles prédictifs, tout en prenant en compte les situations où ceux-ci doivent être formulés linéairement pour leur intégration au PLNE. Ces modèles prédictifs ont pour objectif d'estimer rapidement la durée d'une tournée afin de guider le module de planification à allouer des portions aux semaines de l'horizon et aux différents modes de transport. Nous élaborons trois modèles différents : le premier pour le CARP-DD et CARP-DD avec multiples dépôts lié aux tournées de véhicules par secteur, un second pour un CARP-DD lié aux tournées aériennes réalisées par un type de prestataire de surveillance aérienne, un dernier pour un CARP-IF pour un second type de prestataire.

6.1 Élaboration d'une base d'apprentissage

La première étape du processus d'apprentissage consiste à construire une base d'observations labellisées sur laquelle le modèle prédictif pourra s'entraîner. Elle est constituée d'observations représentant différentes instances de tournées, chacune représentée par un ensemble de *features* (ou caractéristiques) qui servent à décrire leurs propriétés spécifiques, et labellisée par un temps de parcours obtenu pour l'instance résolue. Le choix des *features* est important car ceux ci doivent permettre au modèle prédictif de capturer des relations sous-jacentes et donc de fournir des prédictions fiables du temps de parcours. Cette étape nécessite dans notre cas, la simulation d'instances de tournées mais aussi leurs résolutions pour récupérer la *feature* à expliquer : le temps de parcours. Nous devons aussi calculer ou concevoir des *features* explicatives permettant de décrire ces instances. Cette étape de conception de base d'apprentissage est importante pour pouvoir entraîner des modèles avec une bonne capacité de généralisation sur de nouvelles données.

6.1.1 Simulations de tournées

Comme expliqué précédemment, pour que le modèle d'apprentissage apprenne la relation entre des *features* explicatives d'entrée et une *feature* cible, il est nécessaire de générer des tournées qui constitueront les observations de la base d'apprentissage de notre modèle prédictif. La construction de notre ensemble de données d'apprentissage nécessite des instances étiquetées, chacune comprenant une à plusieurs centaines de portions à surveiller avec le coût ou la durée de la tournée associée. Nous devons établir quelle stratégie de génération d'instances adopter pour avoir un modèle qui se généralise bien, mais aussi la taille de la base d'apprentissage idéale pour entraîner le modèle.

La généralisation d'un modèle d'apprentissage c'est sa capacité à s'adapter à des nouvelles données, c'est-à-dire qui n'ont pas été observées lors de la phase d'entraînement. C'est un critère essentiel de performance, car un modèle efficace ne doit pas uniquement reproduire les exemples appris, mais également capturer les relations sous-jacentes afin de s'adapter à de nouvelles situations. Certains phénomènes peuvent compromettre cette généralisation comme le surapprentissage (*overfitting*), qui se produit lorsque le modèle

d'apprentissage automatique apprend trop précisément les particularités des données d'entraînement, au point de capturer les bruits et les variations aléatoires plutôt que les relations sous-jacentes. Il fournit alors des prédictions précises pour les données d'entraînement mais pas pour les nouvelles données. A l'inverse, le sous-apprentissage (*underfitting*) se produit lorsqu'un modèle n'est pas suffisamment complexe pour capturer les relations dans les données d'entraînement. Cela entraîne une mauvaise performance aussi bien sur les données d'entraînement que sur les données de test, car le modèle d'apprentissage est trop généraliste et incapable de fournir des prédictions précises.

Ainsi, afin d'assurer une bonne généralisation du modèle d'apprentissage, il est important de disposer d'un ensemble d'instances variées dans la base d'apprentissage. Cela permet au modèle de non seulement prédire correctement les coûts de nouvelles instances qui seraient proches des données de la base d'apprentissage, mais aussi de nouvelles instances plus éloignées. Il est également important que le modèle sache aussi bien prédire le coût des instances (regroupement de portions) menant à une bonne solution, que le coût des instances qui conduisent à une mauvaise solution. En effet, la finalité de ce modèle est d'assister la construction d'instances, constituées de portions judicieusement regroupées, de manière à produire des tournées efficaces et à coût minimal.

Ainsi afin d'avoir une bonne variété d'instances, nous les générons à l'aide de différentes stratégies de sélection de portions, mais aussi de différentes tailles. Pour cela nous sélectionnons un ensemble de portions à visiter, et résolvons la tournée associée à l'aide du modèle décrit dans le *Module de Tournées*.

Portions choisies aléatoirement Une première méthode consiste à sélectionner les portions aléatoirement parmi celles du secteur dans le cas de tournées terrestres, ou parmi celles du territoire dans le cas de tournées aériennes. La Figure 6.1, présente un exemple d'instance avec 10 portions choisies aléatoirement, ainsi que la tournée optimale qui en résulte. Bien que nous souhaitons que le module de planification évite de proposer de telles instances, qui résultent en des tournées peu optimales, il est important que le modèle de régression s'entraîne sur ces tournées afin d'assurer une capacité de généralisation à l'ensemble des instances possibles.

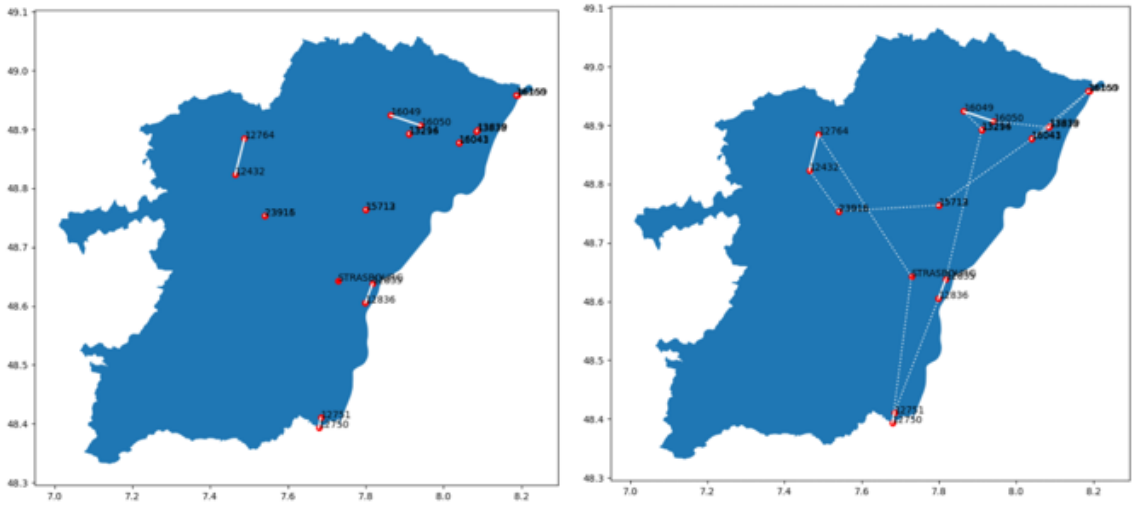


FIGURE 6.1 – À gauche, 10 PSV sélectionnées dans le secteur de Strasbourg et à droite la tournée qui en résulte

Portions proches Une deuxième méthode de sélection de portions consiste à choisir des portions qui sont proches géographiquement. Pour cela nous utilisons la méthode des k plus proches voisins. Nous commençons par sélectionner une portion aléatoirement puis itérativement on ajoute une portion parmi les k plus proches jusqu'à atteindre le nombre de portions souhaité. Cette méthode permet d'obtenir un cluster de portions proches qui devraient résulter en des tournées de plus faible coûts. Dans la Figure 6.2, nous présentons l'exemple d'une instance avec 20 portions sélectionnées avec la méthode des k plus proche voisins, et la tournée qui en résulte.

Nous pouvons étendre cette méthode à la sélection de portions à travers un certain nombre m de clusters. Chaque cluster aura un nombre de portions équitable et permet d'avoir des instances encore variées. Dans les figures suivantes, nous générons des instances à 20 portions dans le secteur de Vannes pour chaque méthode de sélection. Dans la Figure 6.3a les 20 portions sont choisies aléatoirement. Dans la Figure 6.3b les 20 portions sont choisies dans un cluster. Dans la Figure 6.3c les 20 portions sont choisies dans deux clusters.

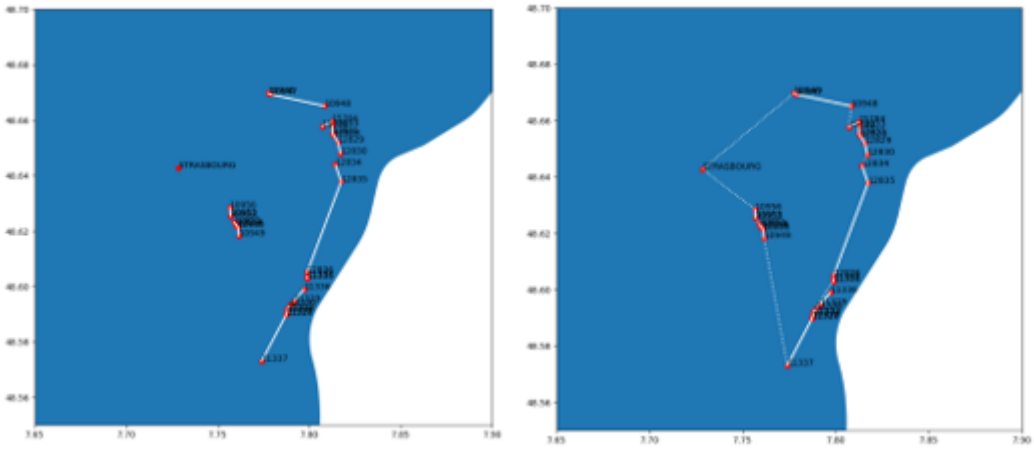
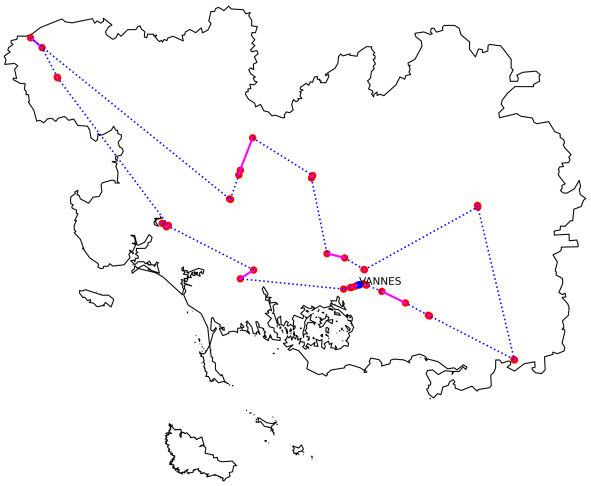
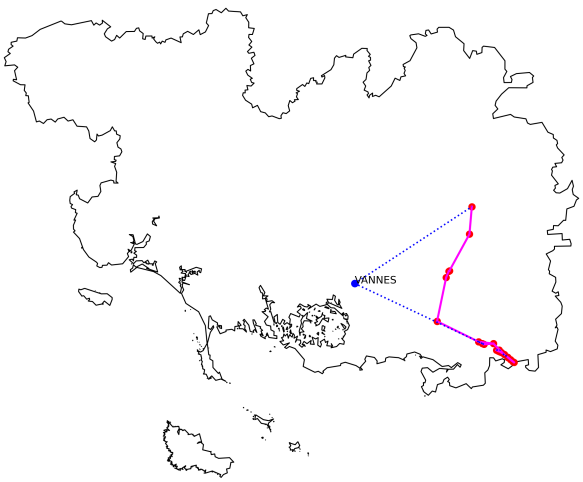


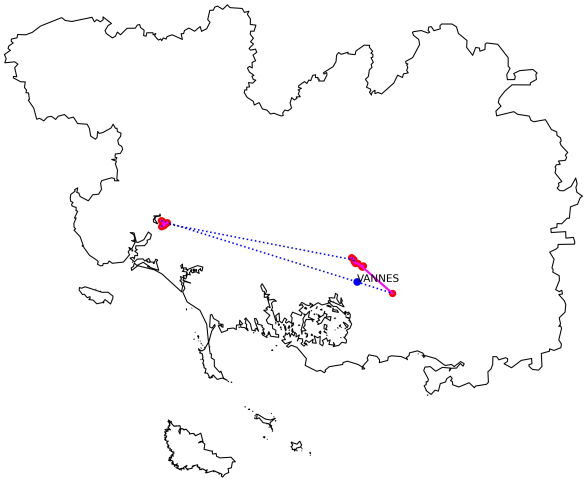
FIGURE 6.2 – À gauche, des 20 PSV proches sélectionnées dans le secteur de Strasbourg et à droite la tournée qui en résulte



(a) Tournée à 20 portions choisies aléatoirement



(b) Tournée à 20 portions à un cluster

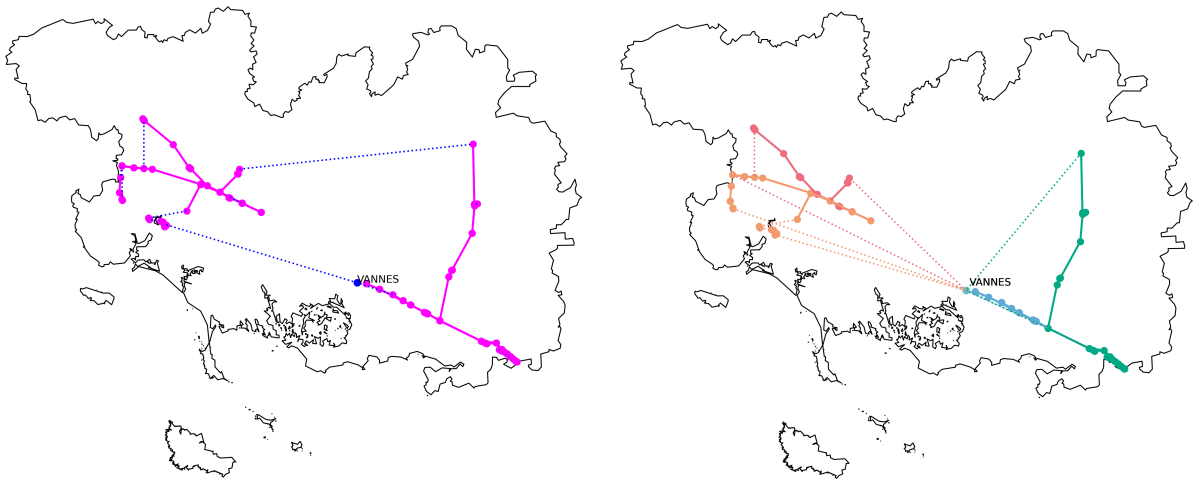


(c) Tournée à 20 portions via 2 clusters

FIGURE 6.3 – Tournées générées à 20 portions sélectionnées selon différentes stratégies

Enfin pour chaque modèle deux coûts sont disponibles. Le premier pour des ARP avec un véhicule unique et le second pour des ARP multi véhicules. Pour le cas avec véhicule unique équivalent au RPP, nous obtenons des solutions optimales à l'aide du modèle PLNE présenté dans le Chapitre 4, avec $k = 1$, et en excluant les contraintes de limites de durées qui ne s'applique pas pour les RPP. Pour le cas impliquant plusieurs véhicules, nous utilisons les tournées précédemment obtenues avec un véhicule unique qui constitue notre tour géant, puis nous appliquons la méthode *Split* de Lacomme, Prins et Ramdane-Cherif 2004 pour le séparer en maximum k tournées distinctes, chacune avec une durée inférieure à δ^v minutes (6 heures dans notre cas). Bien que la méthode *Split* ne garantisse pas l'optimalité, cette méthode produit des solutions quasi optimales.

Dans la Figure 6.4, nous exposons un exemple de deux solutions pour une même instance de 95 portions dans le secteur de Vannes. La première (Figure 6.4a) correspond à une solution à un véhicule, alors que la seconde (Figure 6.4b) est le résultat après l'utilisation de l'algorithme *Split* qui correspond à une solution à 4 tournées.



(a) Solution pour une tournée à un seul véhicule pour 95 portions

(b) Solution multi-véhicule pour 95 portions (4 tournées)

Figure 6.4 – Comparaison entre une tournée mono-véhicule et tournées à 4 véhicules comprenant 95 portions

6.1.2 Base d'évaluation

Dans un processus d'apprentissage, il est nécessaire d'avoir des données sur lesquelles tester les modèles prédictifs et évaluer leurs performances sur des exemples que le modèle n'a a priori jamais vus. L'ensemble de données tests sur lequel nous évaluons nos modèles est un ensemble d'instances réelles de tournées de surveillance de NaTran datant de 2023 contenant plus de 4 282 tournées pour le problème à véhicule unique et 2 138 solutions à véhicules multiples. Pour les instances à véhicule unique, nous récupérons les affectations quotidiennes par secteur et les résolvons à l'aide de notre modèle PLNE. Cela donne plus de 4 200 tournées réelles, soit 92 % des instances réelles, les 82 restantes étant irréalisables.

6.1.3 Description des *features*

Afin que le modèle prédictif sache prédire le coût d'instances de tournées, il faut que des caractéristiques ou *features* accompagnent les différentes instances. Nous nous intéressons aux *features* qui pourraient aider le modèle de prévision à estimer précisément le coût d'une tournée. Pour cela, nous nous sommes penchés sur les caractéristiques spatiales pour tenter d'en tirer des indicateurs intéressants. Le processus de conception des caractéristiques nécessite de la créativité et de l'exploration afin d'identifier les caractéristiques susceptibles d'influencer la durée d'une solution optimale de ARP. Les différentes caractéristiques que nous avons étudiées sont répertoriées dans les tables 6.1 à 6.8. Chaque ligne représente un groupe de caractéristiques. La première colonne indique l'identifiant (ID) du groupe de caractéristiques, la deuxième colonne décrit le groupe de caractéristiques. La troisième colonne indique le nombre de caractéristiques incluses dans le groupe. La dernière colonne indique les références à partir desquelles ces caractéristiques sont dérivées, et rien le cas échéant.

Les articles qui sont concernés sont listés ci-dessous. Nous leur avons attribué un numéro pour raison de gain de place dans les tableaux :

[1] Akkerman et Mes 2022

[2] Bostel et al. 2014

[3] Nicola, Vetschera et Dragomir 2019

[4] Pihera et Musliu 2014

[5] Rasku, Kärkkäinen et Musliu 2016

Dans ce qui suit, le terme « statistiques » désigne dans cet ordre, les caractéristiques qui calculent la somme, la moyenne, la médiane, le maximum, le minimum, la variance, l'étendue et l'intervalle interquartile.

Caractéristiques dimensionnelles Les caractéristiques dimensionnelles répertoriées dans la Table 6.1 reflètent la dimension ou taille générale du problème à travers trois indicateurs : le nombre de portions requises (DM1), le nombre total de nœuds du réseau, y compris les dépôts (DM2), et le nombre de dépôts disponibles pour les véhicules (DM3).

Table 6.1 – *Features* liées à la dimension

ID	Description	Nombre	Références
DM1	Nombre d'arêtes requises (portions)	1	[2]
DM2	Nombre total de nœuds	1	[1,3,4,5]
DM3	Nombre de dépôts	1	-

Caractéristiques liées à la taille des portions sélectionnées Comme indiqué au Chapitre 5, la taille des portions évolue de manière significative non seulement d'un secteur à l'autre, mais aussi d'une portion à l'autre. Ainsi nous introduisons, un ensemble de variables en rapport avec les tailles des portions. La Table 6.2 caractérise la taille des arêtes requises à travers trois ensembles de caractéristiques statistiques. Les *features* P1-P8 expriment les caractéristiques statistiques de longueur des portions de canalisations issues des données du SIG. Les *features* P9-P16 expriment les caractéristiques statistiques des distances routières des portions, en se basant sur la distance en kilomètres requise pour parcourir une portion d'une extrémité à l'autre à travers le réseau routier sous-jacent. Les *features* P17-P24 expriment les caractéristiques statistiques des durées routières des portions, en utilisant la durée en minutes pour parcourir une portion d'une extrémité à l'autre en utilisant le réseau routier sous-jacent.

TABLE 6.2 – *Features* relatives à la taille de portions

ID	Description	Nombre	Références
P1-P8	Longueurs des portions (km)	8	-
P9-P16	Distance routière de surveillance des portions (km)	8	-
P17-P24	Durée routière des portions (min)	8	-

Caractéristiques liées à l'éloignement des portions avec le(s) dépôt(s) La Table 6.3 caractérise l'accessibilité des arêtes requises à partir du ou des dépôt(s) disponible(s) à l'aide de mesures basées à la fois sur la distance et le temps. Pour les secteurs comportant plusieurs dépôts, les *features* DP1-DP16 calculent des statistiques sur la distance moyenne et le temps de trajet entre chaque portion et tous les dépôts, fournissant ainsi une mesure générale de l'accessibilité des dépôts. Les *features* DP17-DP32 fournissent des statistiques calculées sur le dépôt le plus proche de chaque portion, reflétant le point de départ le plus efficace pour la surveillance. Pour les secteurs disposant d'un seul dépôt, les deux perspectives donnent des valeurs identiques (DP1-DP16 correspondent à DP17-DP32), tandis que pour les secteurs disposant de plusieurs dépôts, elles fournissent des informations complémentaires sur le positionnement des dépôts par rapport aux arêtes requises.

Pour le problème des tournées aériennes sur le territoire géré par le prestataire RECTIMO Air Transports, un ensemble d'aérodromes est disponible tout au long de la semaine et peut être considéré comme un réseau de stations intermédiaires. Toutefois, un aérodrome principal sert de point de départ des tournées en début de semaine et de point d'arrivée en fin de semaine. Ainsi, les *features* DP1 à DP16 se rapportent à l'aérodrome principal, tandis que les *features* DP17 à DP32 se réfèrent à l'aérodrome ou la station intermédiaire la plus proche de la portion requise considérée.

TABLE 6.3 – Caractéristiques d'éloignement par rapport au(x) dépôt(s)

ID	Description	Nombre	Référence
DP1-DP8	Distance moyenne aux dépôts (km)	8	[1,3,5]
DP9-DP16	Durée moyenne aux dépôts (min)	8	-
DP17-DP24	Distance au dépôt le plus proche (km)	8	-
DP25-DP32	Durée au dépôt le plus proche (min)	8	-

Caractéristiques de dispersion La Table 6.4 présente les caractéristiques qui reflètent la répartition spatiale des arêtes requises dans le réseau considéré. Les *features* D1-D8 (respectivement D9-D16) expriment des statistiques sur les distances entre les portions (c'est-à-dire entre les extrémités des portions) en kilomètres (respectivement en durée en minutes) à l'aide du réseau routier. Les *features* D17-D24 (respectivement D25-D32) fournissent des statistiques sur les distances, en kilomètres (respectivement en durée de trajet), entre une portion et sa portion voisine la plus proche (c'est-à-dire entre les extrémités les plus proches des deux portions) en utilisant le réseau routier. Les *features* D33-D40 (respectivement D41-D48) fournissent des statistiques sur la distance en kilomètres (respectivement la durée en minutes) à l'aide du réseau routier entre une portion et la portion la plus éloignée.

TABLE 6.4 – Features de dispersion

ID	Description	Nombre	Références
D1-D8	Distances entre chaque paire de portions (km)	8	[1,3,5]
D9-D16	Durées entre chaque paire de portions (min)	8	-
D17-D24	Distance à la portion la plus proche (km)	8	[3,5]
D25-D32	Durée à la portion la plus proche (min)	8	-
D33-D40	Distance à la portion la plus éloignée (km)	8	[3]
D41-D48	Durée à la portion la plus éloignée (min)	8	-

Caractéristiques géométriques La Table 6.5 décrit la distribution spatiale des arêtes requises à l'aide de diverses mesures géométriques. Ces mesures peuvent être classées comme suit. Tout d'abord, G1 et G2 représentent respectivement le périmètre et l'aire de l'enveloppe convexe englobant les arêtes requises. De même, G3 et G4 représentent le périmètre et l'aire du rectangle minimal englobant les arêtes requises. Ces mesures fournissent des informations sur la forme générale et l'étendue des arêtes requises. La distribution géographique est également donnée par les statistiques sur les étendues de coordonnées G5-G10 (latitudes et longitudes). Ensuite, afin de faciliter les comparaisons entre différentes secteurs, G11-G14 fournissent des valeurs mises à l'échelle des étendues de coordonnées en divisant chaque valeur par la moyenne de toutes les valeurs du secteur. Enfin, G15 est le produit de G9 et G10 et fournit une mesure unique de la dispersion géographique.

TABLE 6.5 – Caractéristiques géométriques

ID	Description	Nombre	Références
G1-G2	Enveloppe convexe (périmètre, aire)	2	[1,4,5]
G3-G4	Rectangle englobant (périmètre, aire)	2	[1,4,5]
G5-G6	Étendue de latitude	2	-
G7-G8	Étendue de longitude	2	-
G9	Variance de latitude	1	-
G10	Variance de longitude	1	-
G11-G12	Extrémités de latitudes normalisées	2	-
G13-G14	Extrémités de longitudes normalisées	2	-
G15	Produit des variances des coordonnées	1	[3]

Caractéristiques de distances aux points centraux La Table 6.6 présente les caractéristiques basées sur deux points de référence centraux et leurs relations avec les arêtes requises et les dépôts. Toutes les *features* de ce groupe utilisent la distance euclidienne en kilomètres. Ces points de référence sont le centroïde géométrique, calculé comme la moyenne des coordonnées des nœuds (à l'exception des dépôts), et le point médian, défini

comme le milieu du rectangle englobant tous les nœuds (à l'exception des dépôts). Les caractéristiques C1-C16 capturent les statistiques sur la distribution des distances entre les arêtes requises et le centroïde et le point médian. La caractéristique C17 est la distance euclidienne entre le centroïde et le point médian. La caractéristique C18 exprime la distance euclidienne entre le dépôt et le point médian. La caractéristique C19 exprime la distance euclidienne entre le dépôt et le centroïde. Ces caractéristiques fournissent un aperçu de la répartition spatiale des arêtes requises par rapport à ces points centraux.

TABLE 6.6 – Distances des portions au centroïde et point médian

ID	**Description**	**Nombre**	**Références**
C1-C8	Distance des portions au centroïde	8	[1,3,5]
C9-C16	Distance des portions au point médian	8	[1,5]
C17	Distance du centroïde au point médian	1	-
C18	Distance du centroïde au dépôt	1	[1]
C19	Distance du point médian au dépôt	1	[1,5]

Caractéristiques de connectivité des arêtes Comme nous travaillons sur des ARP, nous pouvons analyser la connectivité globale des arêtes requises dans le graphe. Nous introduisons de nouvelles caractéristiques concernant la connectivité des portions dans la Table 6.7. Nous introduisons d'abord des caractéristiques statistiques sur les degrés des nœuds. Notez que nous ne tenons pas compte du degré minimum, car il est toujours égal à 1. Nous introduisons également une caractéristique de densité du réseau qui correspond au nombre d'arêtes requises par rapport au nombre total d'arêtes dans le secteur pour les problèmes de tournées pour les voitures. Enfin, les composantes connexes correspondent au nombre de groupes d'arêtes (les arêtes sont considérées comme regroupées si elles partagent au moins un nœud en commun). Ces caractéristiques structurelles complètent les mesures géométriques et spatiales en capturant les aspects topologiques du réseau de surveillance.

TABLE 6.7 – Caractéristiques de connectivité des arêtes

ID	Description	Nombre	Références
CN1-CN7	Degrés des nœuds	7	-
CN8	Densité du réseau (rapport arêtes requises / arêtes totales)	1	-
CN9	Composantes connexes	1	-

Caractéristiques liées aux secteurs Enfin, pour rappel le réseau de transport de gaz en France est divisé en secteurs opérationnels, chacun contenant plusieurs tronçons de canalisations. Ces secteurs peuvent être très différents les uns des autres, nous introduisons dans la Table 6.8 une nouvelle classe de caractéristiques liées aux secteurs qui expriment des informations générales sur la géométrie du secteur, mais aussi sur les portions qu'il contient. Ces caractéristiques restent constantes pour un secteur donné et ne dépendent pas des portions de surveillance requises. Elles peuvent donc être pré-calculées et utilisées dans chaque approximation continue sans avoir besoin d'être recalculées. A noter que ces *features* ne concernent que les tournées au sol.

TABLE 6.8 – Caractéristiques liées aux secteurs

ID	Description	Nombre	Références
S1	Nombre de portions dans le secteur	1	-
S2	Nombre de nœuds dans le secteur	1	-
S3-S10	Longueurs des portions du secteur (km)	8	-
S11-S18	Durées routières des portions (min)	8	-
S19-S26	Distances des portions du secteur au centroïde (km)	8	-
S27-S34	Distances des portions du secteur au point médian (km)	8	-
S35	Distances entre le centroïde et le point médian du secteur	1	-
S36-S37	Étendue en longitude du secteur (min, max)	2	-
S38-S39	Étendue en latitude du secteur(min, max)	2	-
S40	Variance en longitude des nœuds du secteur	1	-
S41	Variance en latitude des nœuds du secteur	1	-
S42-48	Degrés des nœuds du secteur	7	-
S49-S50	Enveloppe convexe des nœuds du secteur (périmètre, aire)	2	-
S51-S52	Rectangle englobant les nœuds du secteur (périmètre, aire)	2	-
S53	Nombre de composantes connexes dans le secteur	1	-

Caractéristiques exclusives aux tournées aériennes multi dépôts Concernant les tournées aériennes gérées par le prestataire de surveillance aérienne RECTIMO Air Transports, plusieurs aérodromes sont mis à disposition sur le territoire. Ainsi nous introduisons dans la Table 6.9 des *features* exclusives à ce problème relatives à la proximité avec ces aérodromes. Pour chaque arête requise, nous considérons son aérodrome le plus proche. Ainsi IF1 correspond au nombre d'aérodromes proches distincts. IF2 exprime le rapport entre IF1 et le nombre total d'aérodromes. Enfin, IF3 et IF4 représentent

respectivement le périmètre et l'aire de l'enveloppe convexe formée par l'ensemble des stations intermédiaires proches.

TABLE 6.9 – Caractéristiques liées aux stations intermédiaires

ID	Description	Nombre	Références
IF1	Nombre de stations intermédiaires proches uniques	1	-
IF2	Densité des stations intermédiaires proches (rapport stations proches/ stations totales)	1	-
IF3-IF4	Enveloppe convexe formée par les stations intermédiaires (périmètre, aire)	2	-

6.2 Modèle d'apprentissage

Une fois que la base d'apprentissage est construite avec pour chaque observation un ensemble de *features* explicatives et une durée de tournée optimale à prédire, il devient possible d'entraîner des modèles capables d'identifier les relations sous-jacentes entre ces *features* et la durée optimale de la tournée.

6.2.1 Les modèles d'apprentissage

Nous nous focalisons sur des modèles d'apprentissage de régression car la variable à prédire est quantitative : c'est la durée d'une tournée optimale. Des descriptions plus détaillées des différents modèles sont disponibles dans HASTIE et al. 2009.

Modèle de régression linéaire multivariée

La régression linéaire multivariée est un modèle statistique permettant de modéliser la relation entre une variable à expliquer Y (ici la durée d'une tournée optimale) et p caractéristiques explicatives $X_1, \cdots, X_p$ indépendantes. Cette méthode permet également de prédire la valeur de la variable Y qui s'écrit alors comme une combinaison linéaire des caractéristiques explicatives pondérées par un ensemble de coefficients $\theta = (\theta_0, \theta_1, \cdots, \theta_p)$

Ces coefficients mesurent la contribution de chaque caractéristique X_i à la prédiction de la variable Y sous la forme :

$$Y = \theta_0 + \sum_{i=1}^{p} \theta_i X_i + \epsilon$$

avec ϵ une variable aléatoire centrée indépendante de X_i. L'objectif est de minimiser la moyenne de l'erreur quadratique et de déterminer l'estimateur θ^* :

$$\theta^* = \underset{\theta \in \mathbb{R}^{p+1}}{\arg\min}\, \mathbb{E}[(Y - X^\intercal \theta)^2]$$

Modèle d'ensemble par forêt aléatoire

La forêt aléatoire est un modèle d'apprentissage supervisé basé sur un ensemble d'arbres de décision. Elle permet de modéliser la relation entre une variable à expliquer Y (ici la durée d'une tournée optimale) et un ensemble de p *features* explicatives $X_1, \cdots, X_p$.

Le principe de la forêt aléatoire repose sur la combinaison de plusieurs arbres de décision, chacun étant construit à partir d'un échantillon aléatoire des données et d'un sous-ensemble aléatoire des *features* explicatives. Cette approche introduit de la diversité entre les arbres et permet de réduire la variance du modèle tout en limitant le surapprentissage.

Dans le cas d'un problème de régression, la prédiction finale $\widehat{Y}$ d'une forêt aléatoire correspond à la moyenne des prédictions individuelles $\widehat{Y_b}$ issues de chacun des B arbres de la forêt :

$$\widehat{Y} = \frac{1}{B} \sum_{b=1}^{B} \widehat{Y_b}$$

où chaque arbre b est construit à partir d'un *bootstrap* de l'échantillon d'apprentissage et d'un sous-ensemble aléatoire de variables à chaque nœud.

L'objectif de la forêt aléatoire est d'obtenir une estimation robuste et généralisable de la variable Y, en exploitant la puissance de l'agrégation (« *bagging* ») pour diminuer la variance du modèle par rapport à un arbre de décision unique.

Les hyper-paramètres essentiels incluent notamment : le nombre d'arbres B dans la forêt, la profondeur maximale des arbres, et le nombre de *features* considérées à chaque

division.

Modèle d'ensemble par gradient boosting (XGBoost)

Le modèle XGBoost (*Extreme Gradient Boosting*) est une méthode d'apprentissage supervisé fondée sur le principe du *boosting* de gradient (Brownlee 2016). Il permet de modéliser la relation entre une variable à expliquer Y (ici la durée d'une tournée optimale) et un ensemble de p *features* explicatives $X_1, \cdots, X_p$.

Le principe du *boosting* consiste à construire une séquence d'arbres de décision de manière itérative, où chaque nouvel arbre est entraîné pour corriger les erreurs commises par la somme des arbres précédents. Ainsi, contrairement à la forêt aléatoire qui agrège des arbres indépendants, le *boosting* crée un modèle additif dans lequel chaque arbre affine progressivement la prédiction globale.

La prédiction finale $\hat{Y}$ d'un modèle XGBoost peut s'écrire comme la somme pondérée des prédictions des M arbres construits :

$$\widehat{Y} = \sum_{m=1}^{M} f_m(X), \qquad f_m \in \mathcal{F}$$

où chaque f_m représente un arbre de décision appartenant à l'ensemble des fonctions $\mathcal{F}$ possibles. Les arbres sont ajoutés séquentiellement de manière à minimiser une fonction de perte $\mathcal{L}$ pénalisée par un terme de régularisation Ω qui contrôle la complexité du modèle :

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{m=1}^{M} \Omega(f_m)$$

La régularisation introduite dans XGBoost permet de limiter le surapprentissage tout en améliorant la capacité de généralisation du modèle.

Les principaux hyper-paramètres à régler incluent : le nombre d'arbres M, le taux d'apprentissage (*learning rate*), la profondeur maximale des arbres, et les coefficients de régularisation (λ, α).

XGBoost par rapport à d'autres modèles ensemblistes d'arbres se distingue par son efficacité computationnelle, sa robustesse face aux données bruitées et sa capacité à capturer des relations complexes et non linéaires entre les *features* explicatives et la variable cible.

6.2.2 Critères de sélection

Afin d'évaluer la performance prédictive des modèles, de les comparer et de sélectionner les meilleurs hyper-paramètres, il est nécessaire d'utiliser différents indicateurs de performance.

Erreur absolue moyenne en pourcentage (MAPE)

$$\text{MAPE} = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{A_k - F_k}{A_k}\right| \times 100$$

La *Mean Absolute Percentage Error* (MAPE) est exprimée en pourcentage et mesure la moyenne des pourcentages d'erreur absolue pour chaque observation. Plus la MAPE est faible, plus le modèle de prévision est précis. L'avantage de cet indicateur est qu'il permet de comparer plusieurs modèles entre eux même quand les unités sont différentes. Son inconvénient est notamment sa sensibilité aux valeurs aberrantes et son incapacité à gérer les valeurs nulles ou proches de zéro.

Racine de l'erreur quadratique moyenne (RMSE)

$$\text{RMSE} = \frac{1}{n}\sqrt{\sum_{k=1}^{n}(F_k - A_k)^2}$$

La *Rooted Mean Squared Error* (RMSE) mesure l'écart moyen entre les valeurs réelles A_k et les valeurs prédites F_k, en prenant en compte les erreurs sous forme de carrés pour attribuer plus de poids aux erreurs importantes. La RMSE est exprimée dans les mêmes unités que les données de base, ce qui la rend facile à interpréter. Plus la RMSE est faible, plus le modèle de prévision est précis.

Erreur moyenne absolue (MAE) L'erreur moyenne absolue ou *Mean Absolute Error* (MAE) mesure la moyenne des erreurs absolues entre les valeurs réelles A_k et les valeurs

prédites F_k, sans tenir compte du signe de l'erreur.

$$\text{MAE} = \frac{1}{n}\sum_{k=1}^{n}(F_k - A_k)$$

La MAE est exprimée dans les mêmes unités que les données de base, ce qui la rend plus facile à interpréter. Plus la MAE est faible, moins le modèle de prévision est biaisé.

Coefficient de corrélation par rang de Spearman Contrairement aux métriques d'erreur traditionnelles mentionnées ci-dessus, le coefficient de corrélation par rang de Spearman mesure la cohérence dans l'ordre des valeurs prédites et réelles (Sedgwick 2014), plutôt que de se concentrer sur l'amplitude des erreurs de prédiction. Le coefficient se calcule comme ceci :

$$r_s = 1 - \frac{6 - \sum_{m=1}^{n} d_m^2}{n(n^2-1)} \tag{6.1}$$

où $d_m = R[\widehat{y_m}] - R[y_m]$ représente la différence entre les rangs prédits et réels pour chaque observation. Le coefficient varie de -1 à 1, les valeurs proches de ± 1 indiquant des relations monotones plus fortes. Cette métrique est particulièrement utile dans les contextes de tournées et de planification où le maintien de l'ordre relatif des options est crucial. Par exemple, dans un scénario de planification mensuelle où des portions doivent être réparties sur les semaines afin de minimiser les durées (CARP-DD), une corrélation de Spearman parfaite ($r_s = 1$) entre les coûts prédits et réels augmente la probabilité que les allocations hebdomadaires basées sur les prédictions correspondent à celles basées sur les coûts réels, améliorant ainsi les décisions de planification malgré d'éventuelles différences dans les valeurs absolues.

6.2.3 Sélection de *features*

Lors de la construction et de l'élaboration des *features*, il est fréquent que certaines d'entre elles soient similaires, voire redondantes. Elles transmettent ainsi une information équivalente au modèle, ce qui ne contribue pas à améliorer sa performance. Cela peut même la détériorer en augmentant la complexité du modèle, en introduisant de la multicolinéarité, et en favorisant le surapprentissage. Dans le cadre de nos travaux, il est aussi important de

garder un nombre de features limité, car celles-ci doivent être calculées rapidement pour chaque instance testée ; et le module de planification évalue un grand nombre d'instances possibles.

Afin d'éviter ces problèmes, la sélection des *features* à utiliser dans le modèle constitue une étape importante dans la modélisation du problème d'approximation des coûts. En effet, elle permet de réduire la dimension du problème, de limiter le risque de surapprentissage, d'améliorer l'interprétabilité du modèle, et de le rendre plus général et plus efficace.

Nous exposons différentes méthodes de sélection de *features*, dont une analyse de corrélations de caractéristiques mais aussi des modèles de régularisation, puis une méthode gloutonne de sélection itérative de caractéristiques.

Corrélation des *features* Un premier outil d'études de relations entre *features* est l'usage de la matrice de corrélations de features. Des caractéristiques fortement corrélées dans un modèle apportent une information qui est redondante et peuvent ainsi nuire à la qualité de l'apprentissage en rendant les coefficients instables et l'interprétation des résultats plus difficile. Cela peut aussi conduire à un surapprentissage et à une moins bonne généralisation sur de nouvelles données.

La matrice des corrélations (exemple dans la Figure 6.5) offre une bonne première représentation des potentielles corrélations entre variables. Les pixels en bleu indiquent une corrélation positive alors que ceux en rouge décrivent une corrélation négative.

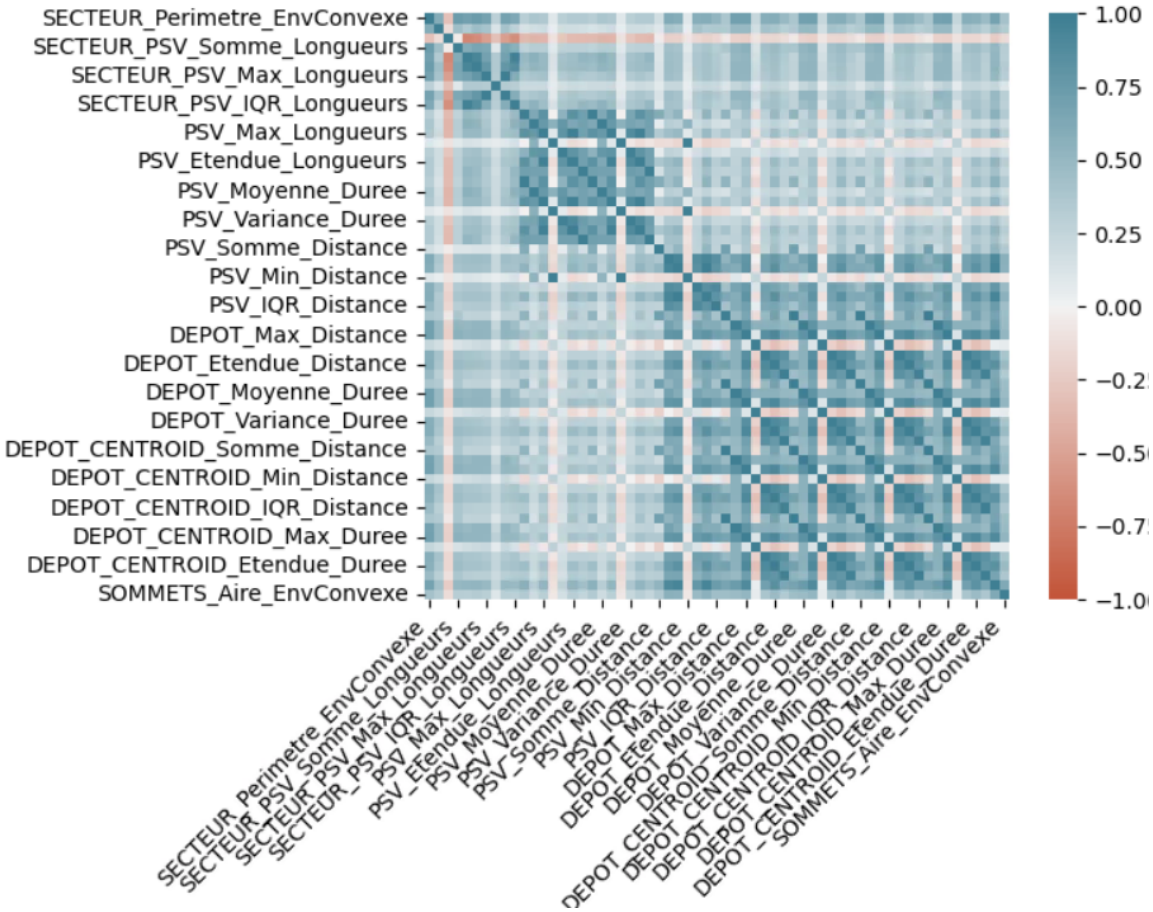


FIGURE 6.5 – Matrice de corrélation des variables

Modèles de régression pénalisée

Une alternative à la régression multivariée standard est la régression pénalisée HASTIE et al. 2009. Ces méthodes permettent d'éviter le surapprentissage en ajoutant un terme de pénalité (ou de régularisation) au modèle de régression qui diminuera certaines valeurs des coefficients de régression vers zéro. Cela permet aux caractéristiques les moins contributives d'avoir un coefficient proche ou égal à zéro afin de conserver un petit nombre de caractéristiques sur lesquelles un modèle fiable peut être construit. Il existe différents modèles de régression pénalisée en fonction du terme de pénalité ajouté.

La régression dite *Least Absolute Shrinkage and Selection Operator* (LASSO), également connue sous le nom de régularisation L1, pénalise la valeur absolue des coefficients de régression. Le terme de pénalité force certains coefficients à être exactement égaux à zéro.

La nouvelle expression à minimiser est la suivante :

$$\min_{\theta_0,\theta_1,\cdots,\theta_n} \left\{ \sum_{i=1}^{m} (Y_i - (\theta_0 + \theta_1 X_{i1} + \theta_2 X_{i2} + \cdots + \theta_n X_{in}))^2 + \lambda_1 \sum_{j=1}^{n} |\theta_j| \right\}$$

avec λ_1 le paramètre de régularisation qui contrôle la force de la pénalité appliquée aux valeurs absolues des coefficients de régression.

La régression RIDGE, également connue sous le nom de régularisation L2, réduit certains coefficients proches de zéro avec une pénalité proportionnelle à la somme des valeurs au carré des coefficients. Cette régularisation est efficace pour traiter la multicolinéarité. L'objectif est de minimiser l'équation suivante :

$$\min_{\theta_0,\theta_1,\cdots,\theta_n} \left\{ \sum_{i=1}^{m} (Y_i - (\theta_0 + \theta_1 X_{i1} + \theta_2 X_{i2} + \cdots + \theta_n X_{in}))^2 + \lambda_2 \sum_{j=1}^{n} \theta_j^2 \right\}$$

avec λ_2 le paramètre de régularisation qui contrôle la force de la pénalité appliquée aux valeurs au carré des coefficients de régression.

Elastic Net combine les propriétés des régressions LASSO (régularisation L1) et Ridge (régularisation L2). Il peut sélectionner certaines caractéristiques comme le LASSO et traiter la multicolinéarité comme le RIDGE.

Sélection séquentielle de *features*

Parmi les différentes stratégies de sélection de *features*, certaines reposent sur une approche itérative et progressive, permettant de construire progressivement de manière gloutonne un sous-ensemble de *features* à partir des données disponibles.

La sélection séquentielle des caractéristiques permet de construire, étape par étape, un sous-ensemble de variables efficace pour la modélisation, bien qu'il ne soit pas nécessairement optimal (Hastie et al. 2009). Cette approche évalue itérativement les caractéristiques en fonction de leur contribution prédictive au modèle. Le processus de sélection dite ascendante commence avec un ensemble de caractéristiques vide et ajoute progressivement les caractéristiques les plus utiles. Ce processus se poursuit jusqu'à ce que les modifications supplémentaires n'apportent plus d'améliorations significatives aux performances du

modèle. En procédant de manière ascendante, nous conservons un ensemble de caractéristiques minimal mais efficace, garantissant que seules les caractéristiques ayant une valeur prédictive importante sont incluses.

Construction d'un modèle naïf Un modèle d'apprentissage naïf correspond à une modélisation simple pour résoudre le problème de prévision de coût donné. Il sert de modèle de comparaison pour notre modèle de régression linéaire. Nous supposons que si les performances du modèle élaboré sont moins bonnes ou équivalentes aux performances du modèle naïf, alors le modèle n'est pas assez performant voire inutile. Dans notre cas, nous calculons la moyenne des coûts par secteur, à partir de notre base d'apprentissage.

6.3 Application et résultats pour la surveillance automobile

Dans cette partie nous appliquons la méthode de conception de la base d'apprentissage ainsi que la conception du modèle prédictif décrite dans les précédentes sections. L'objectif est d'obtenir la meilleure composition et taille de base d'apprentissage pour l'apprentissage du modèle prédictif. Ensuite, nous comparons différents modèles de régression avec le réglage de leurs hyper-paramètres et gardons les *features* importantes pour le modèle. Par ailleurs, nous distinguons les résultats pour les tournées mono véhicule (sans contraintes de capacité) et les résultats pour les tournées multi-véhicules (avec contraintes de capacité).

6.3.1 Composition de la base d'apprentissage

Nous générons des instances de surveillance pour chacun des 75 secteurs, chacune comprenant entre 10 et 200 portions à surveiller. Comme décrit dans la section précédente, nous adoptons différentes stratégies de sélection d'arêtes requises pour garantir que notre ensemble d'observations couvre à la fois des scénarios de surveillance géographiquement dispersés et concentrés.

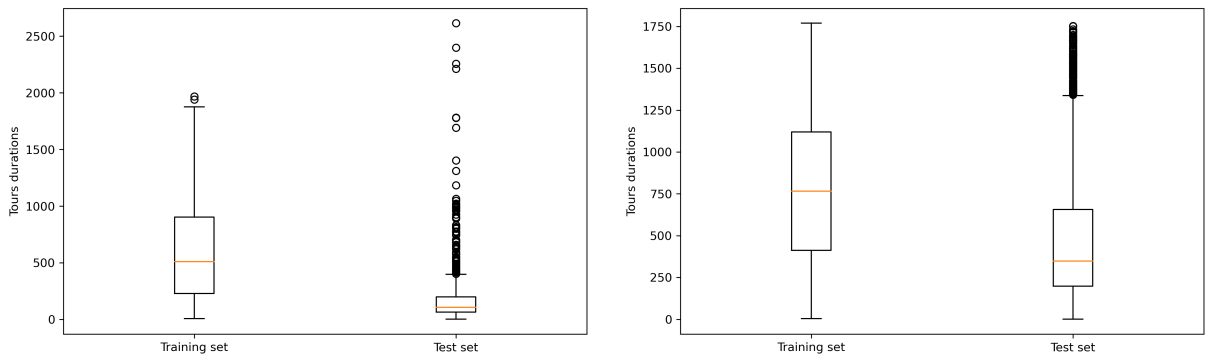
Pour la surveillance au sol, nous avons réussi, à l'aide du modèle PLNE, à générer et résoudre 4 249 instances avec des portions sélectionnées aléatoirement et 6 219 instances

avec des portions regroupées en un ou deux cluster(s). Ces instances sont résolues optimalement pour le cas mono-véhicule. Pour le cas multi-véhicules, la méthode Split trouve des solutions réalisables pour 94,5 % des instances générées aléatoirement soit 4 033 instances et 96,5 % des instances générées par clusters soit 6 009 instances réparties entre une et cinq tournées, comme indiqué dans la Table 6.10.

TABLE 6.10 – Distribution des instances en fonction du nombre de tournées

Nombre de tournées	**Instances aléatoires**	**Instances clusterisées**
1	1 095	2 801
2	803	1 174
3	784	846
4	708	659
5	643	529
Total	4 033	6 009

Nous comparons dans la Figure 6.6 la distribution des coûts entre la base d'apprentissage et la base test, pour les solution mono-véhicules et multi véhicules. On peut remarquer que plus de la moitié des instances test contiennent moins de 35 portions. Comme le montre la Figure 6.6a, ces instances de test ont tendance à avoir des durées plus courtes que nos instances d'entraînement, ce qui reflète une proportion plus élevée de petites tournées.



(a) Distribution des durées de tournées à véhicule unique

(b) Distribution des durées de tournées multi-véhicules

FIGURE 6.6 – Distribution des durées de tournées mono et multi véhicules

Dans un premier temps, une fois les instances de tournées résolues, nous analysons comment la composition des données d'entraînement affecte les performances d'un modèle

de régression.

Pour évaluer les différentes configurations et modèles on divise indépendamment les instances générées aléatoirement et par clusters en ensembles d'apprentissage et de validation selon un ratio de 75-25 %. Nous créons une base d'apprentissage dite mixte qui comprend de manière égale des instances générées aléatoirement et des instances générées par clusters. Elle est aussi divisée en ensembles d'apprentissage et de validation selon le même ratio.

La Table 6.11 présente les mesures de performance (MAE, RMSE, MAPE) des modèles de régression linéaire entraînés sur différentes compositions d'ensembles d'observations : portions sélectionnées aléatoirement uniquement, portions sélectionnées par clusters uniquement et mixtes (portions sélectionnées aléatoirement et via cluster). Les résultats montrent que la construction de l'ensemble de données d'apprentissage a un fort impact sur les performances du modèle de régression linéaire.

Les modèles entraînés selon une stratégie de sélection unique ne fonctionnent bien que sur des instances similaires. Le modèle entraîné sur des instances générées aléatoirement atteint ses meilleures performances sur des instances générées de la même manière (MAE :18,7 minutes ; RMSE : 24,9 minutes ; MAPE : 4,3 %), mais ses performances se détériorent considérablement sur les instances de portions regroupées. De même, le modèle formé par des instances de portions regroupées excelle sur les mêmes types d'instances (MAE : 14,4 minutes ; RMSE : 19,5 minutes ; MAPE : 5,2 %), mais obtient de mauvais résultats sur des instances générées aléatoirement. En revanche, le modèle qui s'est entraîné sur les deux types d'observations affiche des performances robustes pour tous les types d'instances, avec des valeurs MAE (18,4-22,3 minutes), RMSE (24,9-29,4 minutes) et MAPE (5-7,4 %) constantes. Compte tenu de ces résultats et de l'importance de la robustesse et des capacités de généralisation du modèle prédictif dans notre problème, nous adoptons l'ensemble de données mixtes comme ensemble d'entraînement pour la suite de cette partie.

Ces résultats montrent l'importance de constituer un ensemble de données d'apprentissage composé d'instances variées afin de construire un modèle d'apprentissage beaucoup plus robuste. En effet, il est important de pouvoir prédire avec précision la durée des

TABLE 6.11 – Performance d'un modèle de régression linéaire en fonction de la base d'apprentissage pour les problèmes mono-véhicules

	Inst. aléatoires			**Inst. Clusterisées**			**Inst. mixtes**		
Base d'apprentissage	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**
Inst. Aléa.	**18,7**	**24,9**	**4,3 %**	51,6	87	18,5 %	34,2	48,2	10,7 %
Inst. Cluster.	66.1	105,7	15,1 %	**14,4**	**19,5**	**5.2 %**	39,7	75,7	10 %
Inst. Mixtes	22,3	29,4	5 %	18,4	24,9	7,4 %	**20,2**	**26,9**	**6 %**

tournées « bonnes », mais aussi celle des tournées « moins bonnes ».

Maintenant, pour évaluer l'impact de la taille de la base d'apprentissage sur les performances du modèle prédictif, nous avons entraîné un modèle de régression linéaire sur différentes bases d'apprentissage dont les tailles varient de 10 à 4 500 instances de l'ensemble d'instances d'apprentissage mixte précédent. La Figure 6.7 montre l'évolution de la MAE et de la RMSE en fonction du nombre d'instances incluses dans la base d'apprentissage. Pour chaque taille de base d'apprentissage i, nous sélectionnons 100 sous-bases d'apprentissage différentes en choisissant aléatoirement i instances sur lesquelles entraîner les modèles prédictifs. Ces modèles sont testés sur le même ensemble de test mixte. Nous représentons les intervalles de confiance à 95 % et les courbes médianes pour la RMSE et la MAE.

La figure montre que l'augmentation de la taille de la base d'apprentissage au-delà de 2 000 instances ne réduit pas de manière significative la MAE médiane (environ 27 minutes) ou la RMSE (environ 36 minutes). Si un minimum de 1 500 instances semble nécessaire pour stabiliser l'erreur, dépasser 2 000 instances n'apporte que des avantages limités. Compte tenu du coût computationnel lié à la génération de données d'entraînement à l'aide de modèles d'optimisation, nous privilégions l'efficacité et utilisons un ensemble d'entraînement mixte de 2 000 instances pour la suite de cette étude (1 000 instances

aléatoires et 1 000 instances regroupées).

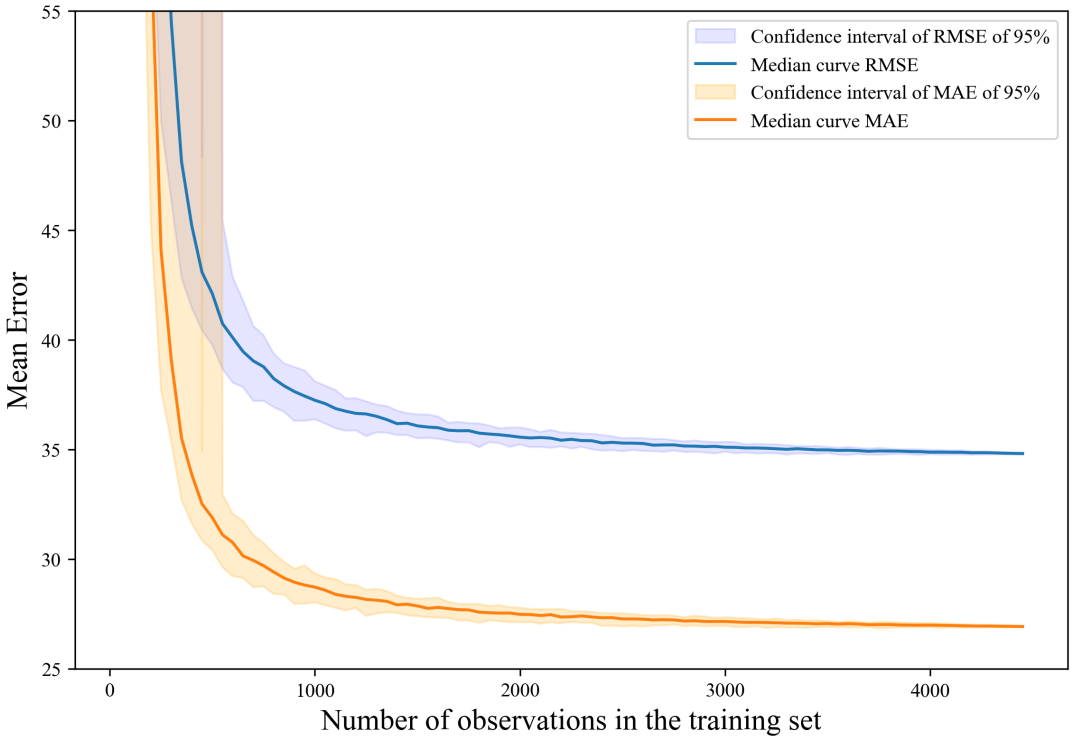


FIGURE 6.7 – MAE et RMSE du modèle de régression linéaire en fonction de la taille de la base d'apprentissage pour un véhicule

6.3.2 Comparaison des modèles de régression

Après avoir construit une base d'apprentissage, nous pouvons tester et comparer les performances des différents modèles de régression décrits dans la section précédente. Ces modèles sont testés sur le même ensemble de validation que précédemment, qui est composé d'un mélange égal d'instances générées aléatoirement et par clusters.

Nous avons d'abord sélectionné les meilleurs hyper-paramètres de chaque modèle prédictif à l'aide de la méthode de validation croisée à 10 échantillons ou *10-fold cross-validation*. Cette méthode consiste à diviser l'ensemble des données de la base d'apprentissage en 10 sous-ensembles (ou *folds*). Le modèle est entraîné sur 9 de ces sous-ensembles et testé sur le 10ème, et ce processus est répété 10 fois pour que chaque sous-ensemble serve une fois de jeu de test, ce qui permet d'évaluer la performance du modèle de manière plus fiable et

stable.

Tout d'abord, comme les termes de pénalisation dans les modèles de régression pénalisés sont sensibles à l'échelle des *features*, nous les normalisons en les ramenant à une moyenne de 0 et à une variance unitaire, avant la phase d'entraînement pour les modèles Elastic Net, Ridge et Lasso. Les meilleurs hyper-paramètres sont présentés dans la Table 6.12 pour les modèles de régression pénalisés Pour ces trois modèles de régression pénalisés, nous ajustons deux hyper-paramètres communs : L'hyper-paramètre `alpha` contrôle la force de la régularisation et `max_iter` contrôle le nombre maximal d'itérations. Le modèle Elastic Net utilise un hyper-paramètre supplémentaire, le `l1_ratio`, pour ajuster l'équilibre entre la régularisation L1 (Lasso) et L2 (Ridge).

Les meilleurs hyper-paramètres pour les modèles d'arbres sont répertoriés dans la Table 6.13. Nous testons plusieurs hyper-paramètres tels que `n_estimators` (le nombre d'arbres de décision à inclure) et `max_depth` (la profondeur maximale des arbres). Le modèle de Random Forest dispose d'un hyper-paramètre supplémentaire, `max_features`, qui spécifie la fraction de caractéristiques à choisir à chaque division dans un arbre donné. Le modèle XGBoost utilise deux hyper-paramètres supplémentaires : `learning_rate`, qui contrôle la taille du pas à chaque itération de *boosting*, et le ratio `subsample` des instances d'apprentissage.

TABLE 6.12 – Hyper-paramètres évalués pour les modèles pénalisés dans le cas mono-véhicule

Modèle	**Hyper-paramètre**	**Valeurs testées**	**Meilleure valeur**
Elastic Net	`alpha`	logspace(-5, 2, 8)	0,001
	`l1_ratio`	{0,1 ; 0,3 ; 0,5 ; 0,7 ; 0,9}	0,9
	`max_iter`	{1 000, 2 000, 5 000, 10 000}	10 000
Lasso	`alpha`	logspace(-5, 2, 8)	0.01
	`max_iter`	{1 000, 2 000, 5 000, 10 000}	10 000
Ridge	`alpha`	logspace(-5, 2, 8)	0.1
	`max_iter`	{1 000, 2 000, 5 000, 10 000}	1 000

Nous gardons les meilleurs hyper-paramètres pour chaque type de modèle et allons maintenant comparer leurs performances. La Table 6.14 présente les indicateurs de perfor-

TABLE 6.13 – Hyper-paramètres évalués pour les modèles basés sur les arbres dans le cas mono-véhicule

Modèle	**Hyper-paramètre**	**Valeurs testées**	**Meilleure valeur**
Random Forest	`n_estimators`	{50, 100, 200, 300, 500}	300
	`max_depth`	{20, 30, 40, 50}	30
	`max_features`	{'auto', 'sqrt', 'log2'}	'sqrt'
XGBoost	`n_estimators`	{100, 200, 300, 500}	500
	`max_depth`	{3, 5, 7}	3
	`learning_rate`	{0,01 ; 0,1 ; 0,2}	.1
	`subsample`	{0,8 ; 0,9 ; 1}	.8

mance pour chaque modèle de régression. Les résultats seront examinés en deux groupes : les modèles linéaires (régression linéaire, Elastic Net, Ridge et Lasso) et les modèles non linéaires (Random Forest et XGBoost)

TABLE 6.14 – Comparaison des performances des modèles de régression pour le cas d'un mono véhicule

Modèles	***Features* sélectionnées**	**MAE**	**RMSE**	**MAPE**
Régression Linéaire	195	20,67	27,47	5,94 %
Elastic Net	190	20,70	**27,45**	5,98 %
Ridge	195	20,67	27,47	5,94 %
Lasso	156	20,67	27,48	5,94 %
Random Forest	193	33,03	46,63	7,76 %
XGBoost	195	**20,51**	27,53	**5,16** %

Les différents modèles linéaires affichent des performances très simimaires, avec une MAPE compris entre 5,94 % et 5,98 %, une MAE comprise entre 20,67 et 20,7 minutes et une RMSE comprise entre 27,45 et 27,48 minutes. La régression Lasso sélectionne 156 caractéristiques parmi les 195 disponibles, tandis que Elastic Net en sélectionne 168. XGBoost obtient les meilleures performances avec une MAPE de 5,16 %, un MAE de 20,51 minutes et un RMSE de 27,53 minutes. Bien que les valeurs MAE et RMSE soient similaires à celles des régressions linéaires, XGBoost est capable d'obtenir une MAPE inférieure en effectuant de meilleures prédictions pour les instances de plus courte durée. Random Forest affiche les pires performances pour tous les indicateurs, avec une MAPE

de 7,76 %, une MAE de 33,03 minutes et une RMSE de 46,63 minutes.

Nous sélectionnons le modèle de régression linéaire pour la suite de cette étude en nous basant sur plusieurs considérations. Les modèles linéaires démontrent de solides performances dans la prédiction de la durée des solutions RPP, tout en restant simples et interprétables. De plus, un modèle de régression linéaire est l'un des modèles prédictifs les plus faciles à intégrer dans un PLNE.

6.3.3 Sélection de features

Nous avons apporté plus de 195 *features* pour la surveillance terrestre. Comme indiqué dans la section précédente, certaines features apportent de l'information redondante et peuvent détériorer les performances du modèle ou sa capacité de généralisation à de nouvelles données. Ainsi il faut sélectionner les *features* les plus importantes.

La sélection des *features* permet d'identifier les caractéristiques les plus pertinentes, tout en réduisant la complexité du modèle. Nous utilisons la sélection séquentielle dite ascendante décrite à la section 6.2.3 pour déterminer le meilleur ensemble de caractéristiques pour le modèle de régression linéaire. La Figure 6.8 montre qu'après que le modèle de régression linéaire a pris en compte environ dix caractéristiques, les taux de diminution de la MAE et de la RMSE deviennent moins prononcés, tendant vers la stabilité. Avec 12 caractéristiques, la MAE oscille autour de 30 minutes et la RMSE autour de 40 minutes, ce qui indique des rendements décroissants avec des caractéristiques supplémentaires.

Pour identifier les caractéristiques les plus pertinentes, nous avons utilisé une sélection séquentielle des caractéristiques afin de minimiser l'erreur quadratique moyenne (*Mean Squared Error* (MSE)), en extrayant à la fois les caractéristiques les plus contributives globalement et les caractéristiques les plus contributives qui sont linéaires par rapport aux variables de décision (c'est-à-dire les caractéristiques linéaires). La Table 6.15 répertorie les 12 caractéristiques les plus contributives pour le modèle linéaire, ainsi que les 12 caractéristiques pouvant s'écrire linéairement les plus contributives.

Nous commençons par examiner les caractéristiques sélectionnées parmi toutes les caractéristiques disponibles. Parmi les caractéristiques sélectionnées, le modèle attribue la plus grande importance à une caractéristique liée aux dimensions des portions : la

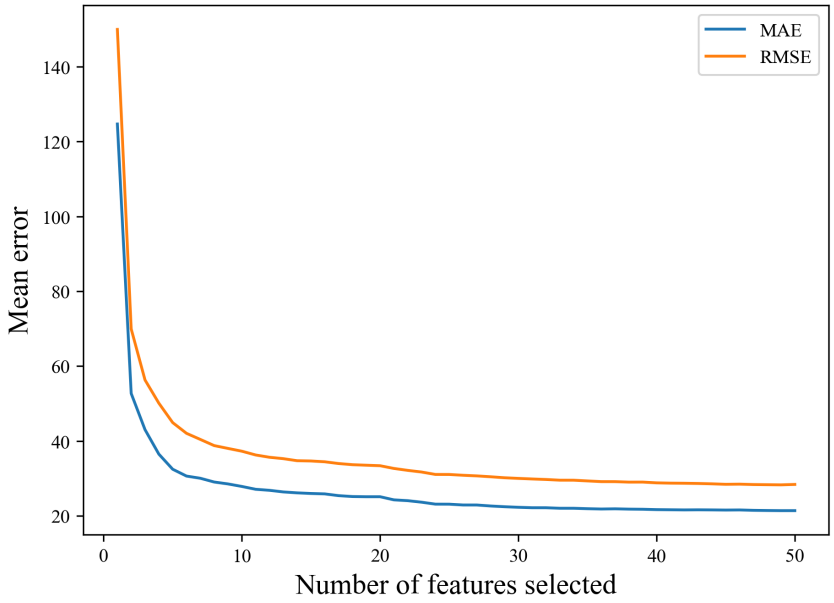


FIGURE 6.8 – MAE et RMSE en fonction du nombre de *features* sélectionnées dans le cas mono-véhicule

somme des durées des portions sélectionnées (P17). Il met également en évidence les caractéristiques liées aux portions et à leurs portions voisines les plus proches, notamment la somme des durées entre les portions et leurs portions voisines les plus proches (D25) et, en 5e position, la durée moyenne entre les portions et leurs voisins les plus proches (D26). Le modèle identifie en outre comme importantes les caractéristiques liées à la distance par rapport au centroïde, c'est-à-dire la somme des distances euclidiennes entre les portions et leur centroïde (C1) et la distance médiane entre les portions et leur centroïde (C3), qui occupe la dernière place (12e). Il met également l'accent sur les caractéristiques liées à la distance par rapport au(x) dépôt(s), telles que la durée maximale entre les portions et le(s) dépôt(s) (DP12) qui reflète la durée entre le(s) dépôt(s) et la portion la plus éloignée et la variance des distances entre les portions et le(s) dépôt(s) (DP6), classée 10e. De plus, le modèle met en évidence l'aire du rectangle englobant toutes les portions requises et tous les dépôts (G4). Parmi les autres caractéristiques clés figurent les caractéristiques liées aux dimensions, telles que le nombre de composants connectés (CN9), le nombre de nœuds (DM2) et le degré moyen des nœuds d'un secteur (S42). Enfin, le modèle identifie la durée médiane entre les portions (D10) comme une caractéristique importante.

Comme expliqué dans la section précédente, les approximations continues sont parfois destinées à être intégrées dans un programme linéaire. Pour cette raison, nous effectuons une autre analyse en utilisant uniquement les caractéristiques qui peuvent être exprimées de manière linéaire par rapport aux variables de décision dans un modèle de programmation linéaire. Toutes les caractéristiques qui calculent la somme, le maximum, le minimum ou la plage peuvent être exprimées de manière linéaire, à l'exception de celles qui calculent les distances ou les durées requises entre les arêtes, ainsi que les caractéristiques impliquant le centroïde et le point médian, car leurs coordonnées ne peuvent pas être calculées de manière linéaire. De plus, toutes les caractéristiques sectorielles de la Table 6.8 sont considérées comme « linéaires » car elles sont fournies à l'avance et sont indépendantes de la sélection des portions. Dans la Figure 6.9, nous observons que, comme dans la Figure 6.8, le RMSE et le MAE se stabilisent après l'ajout d'environ 10 *features* au modèle de régression linéaire, le MAE se stabilisant autour de 39 minutes et le RMSE autour de 50 minutes.

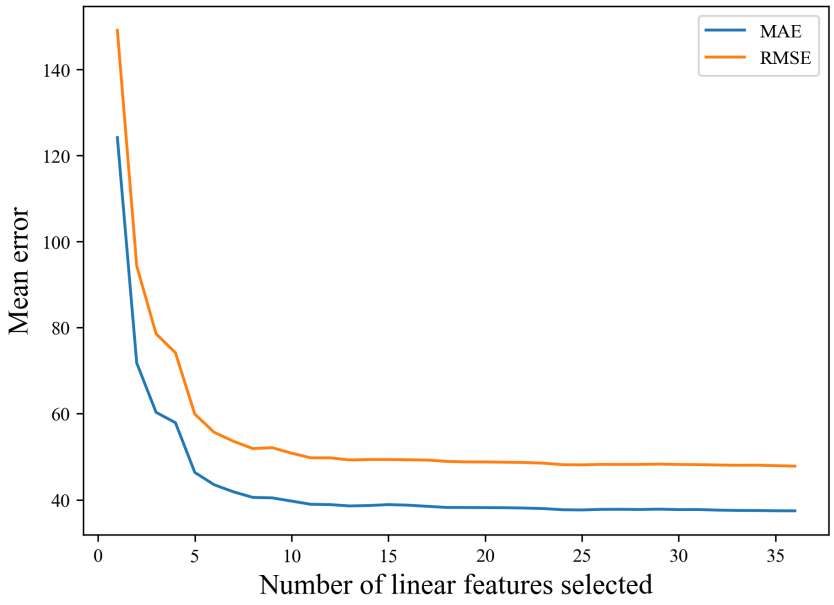


FIGURE 6.9 – MAE et RMSE en fonction du nombre de caractéristiques linéaires sélectionnées pour le cas d'un seul véhicule

Nous examinons les 12 caractéristiques linéarisables sélectionnées. La première caractéristique sélectionnée est également la somme des durées des portions (P17), car il s'agit d'une caractéristique linéaire. Le modèle met ensuite en évidence les plages des

caractéristiques de longitude et de latitude (G7 et G5), qui sont des caractéristiques géographiques importantes représentant la répartition des portions sur la carte. Viennent ensuite les caractéristiques liées à la dimension, telles que le nombre de nœuds (DM1) et de portions (DM2). Bien que ces caractéristiques puissent sembler redondantes ensemble, elles fournissent des informations complémentaires au modèle, liées au degré des nœuds et à la connectivité des arêtes requises. Par exemple, un certain nombre de portions peuvent avoir moins de nœuds si les arêtes requises sont liées et partagent des nœuds communs. Plusieurs caractéristiques liées au secteur sont également sélectionnées, notamment le degré moyen des nœuds du secteur (S42) et la superficie du rectangle englobant toutes les portions du secteur (S52). De plus, la somme des durées des portions du secteur (S11) et la somme des distances entre toutes les portions du secteur et le centroïde (S19) sont incluses. La somme des longueurs des parties (P1) est également sélectionnée. Enfin, deux caractéristiques liées aux dépôts sont sélectionnées : la somme des distances par rapport au(x) dépôt(s) (DP1) et la distance minimale par rapport au(x) dépôt(s) (DP5), classées respectivement 8e et 11e.

En résumé, les deux ensembles de caractéristiques sélectionnées contiennent certains éléments communs. Cependant, dans le premier sous-ensemble de caractéristiques sélectionnées, le modèle intègre davantage de caractéristiques de dispersion, telles que les distances et les durées entre les portions. Comme celles-ci ne peuvent être exprimées de manière linéaire, l'ensemble de caractéristiques linéaires comprend des caractéristiques géographiques supplémentaires qui reflètent la dispersion spatiale des portions. En outre, davantage de caractéristiques sectorielles sont utilisées dans l'ensemble de caractéristiques linéaires.

6.3.4 Évaluations sur des instances réelles

Nous évaluons maintenant nos modèles sur les données réelles de NaTran, en tenant compte à la fois de la précision des prédictions et de la qualité des décisions à l'aide du coefficient de corrélation de Spearman. Les résultats sont présentés dans la Table 6.16. Pour établir une base de comparaison, nous utilisons le modèle naïf décrit dans la section 6.2.3. Nous comparons également nos modèles à ceux qui utilisent uniquement les caractéristiques

issues de la littérature, comme décrit dans la section 6.1.3.

Sans surprise, le modèle naïf affiche de mauvaises performances avec une MAE de 322,3 minutes, une RMSE de 400,4 minutes et une MAPE de 271 %. Le coefficient de corrélation de Spearman est faible, à 0,28, ce qui indique que les rangs entre les valeurs prédites et les valeurs réelles ne sont pas bien alignés. Cela signifie que ce modèle est non seulement inefficace pour faire des prédictions précises, mais qu'il conduit également à de mauvaises décisions, car l'ordre de classement des durées n'est pas préservé.

Lorsque toutes les caractéristiques sont utilisées, le modèle de régression linéaire (LR) atteint une MAE de 54,1 minutes, une RMSE de 324,0 minutes et une MAPE de 18,3 %. Ces erreurs sont supérieures aux résultats de la phase de validation (voir Table 6.11) et peuvent être dues à un surajustement. Dans l'ensemble, nous obtenons de meilleurs résultats en utilisant les 12 caractéristiques sélectionnées dans le modèle de régression linéaire. Nous obtenons une MAE de 22,3 minutes, une RMSE de 30,6 minutes et une MAPE de 12 %. Ainsi, l'utilisation de 12 caractéristiques permet d'obtenir de meilleures performances, offrant à la fois une précision de prédiction améliorée et une plus grande stabilité. Par rapport au modèle utilisant les 12 caractéristiques sélectionnées dans la littérature, la MAE est réduite de 71, la RMSE de 75,8 et la MAPE de 15,5 points de pourcentage.

La sélection des seules caractéristiques linéaires augmente la MAE globale de 10 minutes par rapport au meilleur modèle, qui utilise 12 caractéristiques (linéaires ou non), avec une MAE de 37 minutes, une RMSE de 48,0 minutes et une MAPE de 20 %. Par rapport au modèle utilisant des caractéristiques linéaires issues de la littérature, notre modèle permet de réduire la MAE de 68 %, la RMSE de 71 % et la MAPE de 26,2 points de pourcentage.

En ce qui concerne l'évaluation des erreurs de décision, le coefficient de corrélation de Spearman est très proche de 1 pour nos modèles de régression linéaire, avec un coefficient supérieur à 0,97. Cela indique que, malgré les fluctuations des erreurs de prédiction, l'ordre de classement entre les durées prédites et réelles est préservé. Par conséquent, les erreurs de prédiction n'ont pas d'impact significatif sur la prise de décision.

TABLE 6.15 – *Features* principales influençant le modèle de régression linéaire dans le scénario mono véhicules

	***Features* sélectionnées**		***Features* linéaires sélectionnées**	
Rang	**ID**	**Nom**	**ID**	**Nom**
1	P17	Somme de la durée des portions	P17	Somme de la durée des portions
2	D25	Somme des durées entre les portions et leur portion la plus proche	G7	Étendue de longitude
3	C1	Somme des distances au centroïde	G5	Étendue de latitude
4	DP12	Durée maximale jusqu'aux dépôts	DM2	Nombre de nœuds
5	D26	Durée moyenne entre les portions et leur portion la plus proche	DM1	Nombre de portions
6	G4	Aire du rectangle englobant	S42	Degré moyen des nœuds du secteur
7	CN9	Nombre de composantes connexes	P1	Somme des longueurs des portions
8	DM2	Nombre de nœuds	DP1	Somme des distances jusqu'aux dépôts
9	S42	Degré moyen des nœuds du secteur	S52	Aire du rectangle englobant du secteur
10	DP6	Variance des distances par rapport aux dépôts	S11	Somme des durées des portions du secteur
11	D10	Durée médiane entre les portions	DP5	Distance minimale entre le dépôt et les portions
12	C3	Distance médiane au centroïde	S19	Somme des distances entre les portions de secteur et le centroïde

TABLE 6.16 – Résultats du modèle sur l'ensemble de test pour le cas d'un seul véhicule

Modèles	**MAE**	**RMSE**	**MAPE**	**Corr.**
Modèle naïf	322,3	400,4	271,0 %	0,28
Modèle de RL avec des *features* de la littérature				
Toutes les *features*	70,0	115,9	25,8 %	0,94
12 *features* sélectionnées	77,0	126,3	27,5 %	0,95
8 *features* linéaires	115,7	165,5	46,2 %	0,90
Modèle de RL				
Toutes les *features*	54,1	324,0	18,3 %	0,98
12 *features* sélectionnées	**22,3**	**30,6**	**12 %**	**0,99**
12 *features* linéaires sélectionnées	37,0	48,0	20,0 %	0,97

6.3.5 Résultats multi véhicules

Cette section étend notre analyse au cas de plusieurs véhicules (CARP-DD et Multi-Depot CARP-DD), en suivant une méthodologie similaire à celle utilisée pour le cas d'un seul véhicule, tout en tenant compte de la complexité accrue. Nous commençons par comparer les performances des modèles de régression linéaire entraînés sur des instances groupées par rapport à des instances aléatoires, puis nous déterminons la taille minimale de l'ensemble d'entraînement et comparons différents modèles de régression. Enfin, nous sélectionnons les meilleurs ensembles de caractéristiques et évaluons le modèle final sur des instances multi-tour réelles provenant de NaTran.

Comme dans le cas d'un seul véhicule, la Table 6.17 présente une analyse de l'impact de la composition de l'ensemble d'apprentissage sur différents ensembles de validation. Nous arrivons à la même conclusion : il est essentiel d'inclure à la fois des instances aléatoires et des instances regroupées dans l'ensemble de données d'apprentissage. Avec un ensemble d'apprentissage mixte, nous obtenons une MAE de 28,6 minutes, une RMSE de 40,4 minutes et une MAPE de 8,1 %. En moyenne, les erreurs sont plus élevées que celles du cas à véhicule unique, mais les modèles restent performants.

TABLE 6.17 – Performance de la régression linéaire sur le jeu d'entraînement (cas multi-véhicules)

	Inst. Aléatoires			**Inst. Clusterisées**			**Inst. Mixtes**		
Jeu d'ent.	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**
Inst. Aléa.	**26,7**	**36,3**	**5,7 %**	50,1	68,5	16,9 %	38,1	54,3	11,5 %
Inst. Cluster.	124	154,7	24,6 %	**23,1**	**34,4**	**8,1 %**	76,5	112	16,3 %
Inst. Mixtes	30,1	41,1	6,2 %	27	38,9	9,9 %	**28,6**	**40,4**	**8,1 %**

Dans cette section, nous fournissons également une analyse des performances en fonction du nombre de véhicules utilisés, ou, en d'autres termes, du nombre de tournées. Afin d'analyser l'impact de la taille de la flotte, nous construisons un ensemble de données d'apprentissage équilibré pour chaque valeur de taille de flotte (1 à 5), garantissant ainsi une représentation complète. De plus, comme le montre la Figure 6.10, qui illustre la relation entre la taille de l'ensemble d'apprentissage et les performances du modèle, nous avons veillé à ce que l'ensemble d'apprentissage contienne au moins 3 000 instances, avec

300 instances pour chaque nombre de tournées et chaque méthode de sélection de portion. Notons que l'ensemble d'apprentissage pour le problème multi-véhicules nécessite plus d'instances (3 000) que le problème à véhicule unique (2 000).

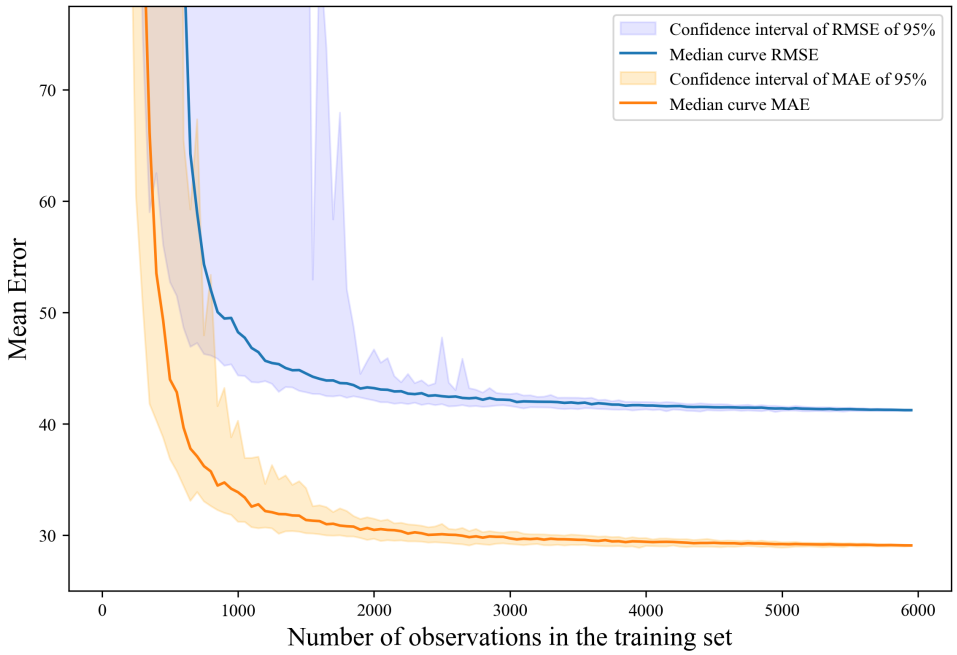


FIGURE 6.10 – MAE et RMSE en fonction de la taille de la base d'apprentissage pour le cas multi-véhicules

La Table A.1 présente les performances du modèle de régression linéaire entraîné sur un ensemble d'apprentissage contenant un nombre de tournées réparti de manière uniforme. Ce modèle est testé sur 5 ensembles de validation (1 à 5 véhicules) de 300 instances chacun. Nous notons que les valeurs MAE et RMSE augmentent avec le nombre de tournées, tandis que la valeur MAPE diminue. Les valeurs MAE et RMSE les plus faibles sont observées pour les solutions à tournée unique, respectivement 23,5 minutes et 31,0 minutes, tandis que la valeur MAPE est la plus élevée avec une valeur de 18,0 %. La MAE augmente de 23,5 minutes à 52,9 minutes pour les solutions à cinq tournées, de même que la RMSE augmente de 31,0 à 103,7 minutes. La MAPE diminue de 18 % pour les solutions à tournée unique à 4,0 % pour les solutions à cinq tournées. Nous notons que le MAPE pour les prévisions de tournées à véhicule unique est plus élevé car la durée totale des tournées est inférieure à celle des autres ensembles. Lorsqu'il est entraîné sur des instances à véhicules

multiples, le modèle testé sur des tournées à véhicule unique est moins performant que lorsqu'il est entraîné sur des instances à véhicule unique. De plus, le modèle entraîné sur des instances à véhicules multiples affiche de moins bonnes performances sur les tournées à véhicule unique par rapport au modèle entraîné spécifiquement sur des instances à véhicule unique.

L'ensemble des tableaux complémentaires liés à l'élaboration du modèle de prédiction de durée de tournée dans le cas multi-véhicules est disponible en annexe A.

Extension aux tournées aériennes Nous avons conduit des analyses analogues pour la prédiction des durées de tournées de surveillance aérienne. Les tournées ont été simulées selon la même méthodologie que précédemment, à partir d'instances générées suivant différentes stratégies et en nombre suffisant pour constituer la base d'apprentissage. Les caractéristiques (*features*) essentielles à l'apprentissage des modèles ont ensuite été sélectionnées, tant pour les cas de tournée aérienne unique que de tournées aériennes multiples. L'ensemble des résultats est présenté en annexe.

6.4 Conclusion

Dans ce chapitre, nous avons présenté une méthodologie fondée sur l'apprentissage automatique visant à prédire le coût d'une ou de plusieurs tournées, qu'elles soient terrestres ou aériennes. Cette approche repose sur la conception de différentes caractéristiques (features), ainsi que sur des stratégies de sélection de portions permettant la simulation de tournées et d'instances variées. Ces choix méthodologiques ont permis de constituer une base d'apprentissage représentative, favorisant une bonne capacité de généralisation des modèles. Après validation, les modèles ont été évalués sur de nouvelles données correspondant à des tournées réelles réalisées en entreprise, et ont montré des performances satisfaisantes. Nous avons également étudié le cas de caractéristiques pouvant s'exprimer de manière linéaire par rapport aux variables de décision, ouvrant ainsi la possibilité de leur intégration directe dans des modèles de programmation linéaire.

Chapitre 7

Heuristique pour la planification des arêtes

Dans ce chapitre, nous nous concentrons sur le *Module de Planification* de l'approche en deux phases. Son objectif est d'allouer, pour chaque semaine de l'horizon annuel des portions du réseau à surveiller, tout en respectant les fréquences de surveillance ainsi que les modes de surveillance possibles.

La fonction objectif de ce module s'appuie sur l'estimation de la durée d'une tournée, fournie par les modèles prédictifs développés dans le Chapitre 6, et exprimés à l'aide des différentes *features* retenues. Nous avons d'abord formulé ce problème comme un PLNE dans le Chapitre 4, ce qui nécessite d'utiliser un modèle prédictif exprimé avec des *features* qui peuvent s'écrire linéairement. Nous montrons dans la section 7.1, la formulation des *features* pour l'intégration dans le PLNE.

Cependant, le modèle PLNE écrit pour le module de planification montre ses limites lorsque la taille des instances augmente ou lorsque plusieurs modes de surveillance sont considérés simultanément. C'est pourquoi, nous introduisons dans la Section 7.2 une nouvelle heuristique permettant d'affecter successivement les arêtes requises aux semaines de l'horizon annuel, tout en respectant les contraintes de fréquences ainsi que les modes de surveillance possibles. L'heuristique étant gloutonne, nous évaluons et identifions différentes stratégies pour déterminer un ordre de traitement des arêtes et modes. Enfin, nous présentons les résultats de l'approche itérative.

7.1 Motivations pour une heuristique de planification

Dans cette section, nous montrons comment nous pouvons résoudre le problème de planification en le formulant sous la forme d'un PLNE. L'objectif de ce modèle est de minimiser la durée totale des tournées. Ainsi, la fonction objectif correspond à la somme des durées estimées par les modèles prédictifs présentés dans le Chapitre 6, lesquels s'appuient sur différentes *features*. Dans ce même chapitre, nous avons distingué les *features* pouvant s'écrire linéairement par rapport aux variables de décisions d'un programme linéaire, de celles qui présentent une forme plus générale.

Nous reprenons donc le PLNE présenté dans la section 4.2 pour le module de planification, et nous y intégrons les *features* du modèle d'apprentissage linéarisables. Pour cela

nous nous aidons des méthodes de transformation citées dans ASGHARI et al. 2022.

Les *features* pour prédire la durée de multi-tournées terrestres sont donnés par la Table A.7. Tandis que les *features* pour prédire la durée de multi-tournées aériennes sont donnés par la Table A.10.

Ajout de variables de décision intermédiaires Certaines des *features* retenues ne sont pas relatives aux arêtes actives mais plutôt aux nœuds de ces arêtes. Cela nécessite donc l'ajout de variables de décision supplémentaires, pour formuler les contraintes supplémentaires relatives à l'expression de ces *features*.

On note χ_i^{sw} la variable binaire égale à 1 si le sommet i est surveillé en voiture dans le secteur s la semaine w, et 0 sinon. Les contraintes suivantes (de 7.1 à 7.3) permettent de déduire la valeur de la variable χ_i^{sw} en fonction de la variable binaire x_{ij}^{sw} relative à la surveillance de l'arête $[i,j]$ en voiture dans le secteur s la semaine w.

$$x_{ij}^{sw} \leq \chi_i^{sw} \qquad \begin{array}{r} \forall s \in S, \quad \forall w \in W, \quad \forall i \in X^v(s) \\ \forall j \in X^v(s), \exists [i,j] \in R^v(s) \end{array} \tag{7.1}$$

$$x_{ji}^{sw} \leq \chi_i^{sw} \qquad \begin{array}{r} \forall s \in S, \quad \forall w \in W, \quad \forall i \in X^v(s) \\ \forall j \in X^v(s), \exists [j,i] \in R^v(s) \end{array} \tag{7.2}$$

$$\sum_{\substack{j \in X^v(s) \\ [i,j] \in R^v(s)}}^{n} x_{ij}^{sw} + \sum_{\substack{j \in X^v(s) \\ [j,i] \in R^v(s)}}^{n} x_{ji}^{sw} \geq \chi_i^{sw} \qquad \forall s \in S, \quad \forall w \in W, \quad \forall i \in X^v(s) \tag{7.3}$$

On note ψ_i^w la variable binaire égale à 1 si le sommet i est surveillé en avion la semaine w, et 0 sinon. Les contraintes suivantes permettent de déduire la valeur de la variable ψ_i^w en fonction de la variable binaire y_{ij}^w relative à la surveillance de l'arête $[i,j]$ en avion la semaine w.

$$y_{ij}^w \leq \psi_i^w \qquad \forall w \in W, \quad \forall i \in X^p \quad \forall j > i \tag{7.4}$$

$$y_{ji}^w \leq \psi_i^w \qquad \forall w \in W, \quad \forall i \in X^p, \quad \forall j < i \tag{7.5}$$

$$\sum_{\substack{j \\ i<j}}^{n} y_{ij}^{w} + \sum_{\substack{j \\ i>j}}^{n} y_{ji}^{w} \geq \psi_i^w \qquad \forall w \in W, \quad \forall i \in X^p \tag{7.6}$$

Le domaine des nouvelles variables est donné par :

$$\chi_i^{sw} \in \{0,1\} \qquad \forall s \in S, \quad \forall w \in W, \quad \forall i \in X^v(s) \tag{7.7}$$

$$\psi_i^w \in \{0,1\} \qquad \forall w \in W, \quad \forall i \in X^p \tag{7.8}$$

Écriture des *features* pour la voiture Les premières formulations permettent d'exprimer les *features* G5 et G7, correspondant respectivement aux étendues de latitudes et de longitudes. Il est donc nécessaire, dans un premier temps, de déterminer les valeurs maximales et minimales de latitude et de longitude. Ainsi, on note $z^{sw}_{\text{lat, min}}$ la latitude minimale par semaine et par secteur pour les tournées en voiture et $z^{sw}_{\text{lat, max}}$ la latitude maximale par semaine et par secteur pour les tournées en voiture :

$$z^{sw}_{\text{lat, min}} \in \mathbb{R} \qquad \forall s \in S, \quad \forall w \in W \tag{7.9}$$

$$z^{sw}_{\text{lat, max}} \in \mathbb{R} \qquad \forall s \in S, \quad \forall w \in W \tag{7.10}$$

Certaines features nécessitent l'écriture de contraintes supplémentaires, notamment pour exprimer des opérateurs comme le maximum, le minimum... On introduit dans le modèle les contraintes ci dessous pour exprimer la *feature* d'étendue de latitude (G5). Ces contraintes permettent de linéariser les opérateurs maximum et minimum quand on a un problème de minimisation, et sont inspirées de l'article Asghari et al. 2022.

$$z^{sw}_{\text{lat, max}} \geq \text{lat}_s(i)\chi_i^{sw} \qquad \forall s \in S, \quad \forall w \in W, \quad \forall i \in X^v(s) \tag{7.11}$$

$$z^{sw}_{\text{lat, min}} \leq \text{lat}_s(i)\chi_i^{sw} + L_{lat}(1-\chi_i^{sw}) \qquad \forall s \in S, \quad \forall w \in W, \quad \forall i \in X^v(s) \tag{7.12}$$

$$z^{sw}_{\text{lat, min}} \leq \sum_{i \in X^v(s)} \text{lat}_s(i)\chi_i^{sw} \qquad \forall s \in S, \quad \forall w \in W \tag{7.13}$$

avec $L_{lat} = \max_{i \in X^v(s)} \text{lat}_s(i)$ et $lat_s(i)$ est la latitude du nœud i situé dans le secteur

s.

Si le coefficient de régression lié à cette *feature* est négatif nous ajoutons les contraintes suivantes :

$$\zeta^{sw}_{\text{i,lat, min}} \in \{0,1\} \qquad \forall i \in X^v(s), \quad \forall s \in S, \quad \forall w \in W \quad (7.14)$$

$$\sum_{i \in X^v(s)} \zeta^{sw}_{\text{i, lat, min}} \leq n-1 \qquad \forall s \in S, \quad \forall w \in W \quad (7.15)$$

$$z^{sw}_{\text{lat, min}} \geq \text{lat}_s(i)\chi^{sw}_i - L_{lat}\ \zeta^{sw}_{\text{i, lat, min}} \qquad \forall s \in S, \quad \forall w \in W, \quad \forall i \in X^v(s) \quad (7.16)$$

$$\zeta^{sw}_{\text{i, lat, max}} \in \{0,1\} \qquad \forall i \in X^v(s), \quad \forall s \in S, \quad \forall w \in W \quad (7.17)$$

$$\sum_{i \in X^v(s)} \zeta^{sw}_{\text{i, lat, max}} \leq n-1 \qquad \forall s \in S, \quad \forall w \in W \quad (7.18)$$

$$z^{sw}_{\text{lat, max}} \leq \text{lat}_s(i)\chi^{sw}_i + L_{lat}\ \zeta^{sw}_{\text{i, lat, max}} \qquad \forall s \in S, \quad \forall w \in W, \quad \forall i \in X^v(s) \quad (7.19)$$

De la même manière, on note respectivement $z^{sw}_{\text{lon, min}}$ et $z^{sw}_{\text{lon, max}}$ la longitude minimale et la longitude maximale par semaine w et par secteur s pour les tournées en voiture :

$$z^{sw}_{\text{lon, min}} \in \mathbb{R} \qquad \forall s \in S, \quad \forall w \in W \quad (7.20)$$

$$z^{sw}_{\text{lon, max}} \in \mathbb{R} \qquad \forall s \in S, \quad \forall w \in W \quad (7.21)$$

On introduit dans le modèle les contraintes ci-dessous pour exprimer la *feature* d'étendue de longitude (G7) :

$$z^{sw}_{\text{lon, max}} \geq \text{lon}_s(i)\chi^{sw}_i \qquad \forall s \in S, \quad \forall w \in W, \quad \forall i \in X^v(s) \quad (7.22)$$

$$z^{sw}_{\text{lon, min}} \leq \text{lon}_s(i)\chi^{sw}_i + L_{lon}(1-\chi^{sw}_i) \qquad \forall s \in S, \quad \forall w \in W, \quad \forall i \in X^v(s) \quad (7.23)$$

$$z^{sw}_{\text{lon, min}} \leq \sum_{i \in X^v(s)} \text{lon}_s(i)\chi^{sw}_i \qquad \forall s \in S, \quad \forall w \in W \quad (7.24)$$

avec $L_{lon} = \max_{i \in X^v(s)} lon_s(i)$ et $lon_s(i)$ est la longitude du nœud i situé dans le secteur s.

Si le coefficient de régression lié à cette *feature* est négatif nous ajoutons les contraintes suivantes :

$$\zeta^{sw}_{\text{i,lon, min}} \in \{0,1\} \qquad \forall i \in X^v(s), \quad \forall s \in S, \quad \forall w \in W \quad (7.25)$$

$$\sum_{i \in X^v(s)} \zeta^{sw}_{\text{i, lon, min}} \leq n-1 \qquad \forall s \in S, \quad \forall w \in W \quad (7.26)$$

$$z^{sw}_{\text{lon, min}} \geq \text{lon}_s(i)\chi^{sw}_i - L_{lon}\ \zeta^{sw}_{\text{i, lon, min}} \quad \forall s \in S, \quad \forall w \in W, \quad \forall i \in X^v(s) \quad (7.27)$$

$$\zeta^{sw}_{\text{i, lon, max}} \in \{0,1\} \qquad \forall s \in S, \quad \forall w \in W \quad (7.28)$$

$$\sum_{i \in X^v(s)} \zeta^{sw}_{\text{i, lon, max}} \leq n-1 \qquad \forall s \in S, \quad \forall w \in W \quad (7.29)$$

$$z^{sw}_{\text{lon, max}} \leq \text{lon}_s(i)\chi^{sw}_i + L_{lon}\ \zeta^{sw}_{\text{i, lon, max}} \quad \forall s \in S, \quad \forall w \in W, \quad \forall i \in X^v(s) \quad (7.30)$$

Nous introduisons de façon analogue les mêmes *features* : $z^{w}_{\text{lon, min}}$, $z^{w}_{\text{lon, max}}$, $z^{w}_{\text{lat, min}}$ et $z^{w}_{\text{lat, max}}$, pour les tournées aériennes, cette fois sans tenir compte du secteur s. De la même manière, les contraintes associées aux autres *features* de minimum et de maximum s'écrivent selon le même principe, mais nous ne les détaillerons pas ici.

Équation de l'estimation des coûts de tournées automobiles Ainsi les coûts estimés des tournées en voiture s'écrivent :

$$\begin{aligned}\hat{c}^{v,sw} = {} & \theta^v_0 + \theta^v_1 \sum_{[i,j] \in R^v(s)} {t^v_{ij}}' x^{sw}_{ij} + \theta^v_2 (z^{sw}_{\text{lon, max}} - z^{sw}_{\text{lon, min}}) + \theta^v_3 (z^{sw}_{\text{lat, max}} - z^{sw}_{\text{lat, min}}) \\ & + \theta^v_4 \sum_{i \in X^v(s)} \chi^{sw}_i + \theta^v_5 \sum_{[i,j] \in R^v(s)} x^{sw}_{ij} + \theta^v_6 \sum_{[i,j] \in R_v(s)} d^c_{ij} x^{sw}_{ij} + \theta^v_7\ \deg_{\text{mean}}(s) \\ & + \theta^v_8 \sum_{[i,j] \in R^v(s)} {\delta^v_{ij}}' x^{sw}_{ij} + \theta^v_9\ \sigma^2_{\text{lon}}(s) + \theta^v_{10}\ t^v_{\text{max},d} + \theta^v_{11}\ (\delta^v_{\text{max},d} - \delta^v_{\text{min},d}) \\ & + \theta^v_{12}\ {\delta_{\text{min}}}'(s) \qquad \forall s \in S \quad \forall w \in W\end{aligned} \quad (7.31)$$

avec

— d^c_{ij} : représente le dépôt le plus proche de l'arête $[i,j]$

— $\delta^v_{\text{max},d} = \max_{i \in X, d \in D^v(s)} \delta^v_{di}$ la distance maximale entre les portions et les dépôts

— $t^v_{\text{max},d} = \max_{\substack{i \in X^v(s) \\ d \in D^v(s)}} \chi^{sw}_i \delta^v_{di}$ représente le temps maximal entre une portion et les dépôts

— $t^v_{\min,d} = \min_{\substack{i \in X^v(s) \\ d \in D^v(s)}} \chi^{sw}_i \delta^v_{di}$ représente le temps minimal entre une portion et un dépôt

— $\delta^v_{ij}{}'$ la longueur de l'arête $[i,j]$

— $\delta_{\min}{}'(s) = \min_{[i,j] \in R^v(s)} \delta^v_{ij}$ la longueur minimale des portions du secteur

— $\deg_{\text{mean}}(s)$ représente le degré moyen pour un secteur s

— $\sigma^2_{\text{lon}}(s)$ est la variance des longitudes du secteur s

Équation de l'estimation des coûts de tournées avion Ainsi les coûts estimés des tournées en avion s'écrivent :

$$\begin{aligned}\hat{c}^{p,w} = {} & \theta^p_0 + \theta^p_1 (z^w_{\text{lon, max}} - z^w_{\text{lon, min}}) + \theta^p_2 \sum_{i \in X^p} \psi^w_i + \theta^p_3 (z^w_{\text{lat, max}} - z^w_{\text{lat, min}}) \\ & + \theta^p_4 \sum_{[i,j] \in R^p} t^p_{ij}{}' y^w_{ij} + \theta^p_5 \sum_{[i,j] \in R^p} y^w_{ij} + \theta^p_6 \deg_{\text{max,i}} + \theta^p_7 \, t^p_{\text{max},d^p_0} \\ & + \theta^p_8 (\delta^p_{\text{max}}{}' - \delta^p_{\text{min}}{}') + \theta^p_9 \, t^p_{\text{min},d^p_0} + \theta^p_{10} \, \delta^p_{\text{min}}{}' + \theta^p_{11} (t^p_{\text{max},d^p_0} - t^p_{\text{min},d^p_0}) \\ & + \theta^p_{12} t^p_{\text{min}}{}' \qquad \forall w \in W\end{aligned} \tag{7.32}$$

avec

— $\deg_{\text{max,i}}$ le degré maximal entre les nœuds $i \in R^p$

— t^p_{max,d^p_0} et t^p_{min,d^p_0} respectivement les durées maximales et minimales entre les portions et le dépôt principal $d_0 \in D^p$

— $\delta^p_{\text{max}}{}'$ et $\delta^p_{\text{min}}{}'$ respectivement les longueurs maximales et minimales parmi les arêtes requises $[i,j] \in R^p$

— $t^p_{\text{min}}{}'$ la durée minimale de parcours d'une portion $[i,j] \in R^p$

Ajustements du code Le temps constituait un véritable enjeu lors du développement du code. Nous avons donc ajusté certaines *features*, leur écriture ou les données pré-calculées. Nous avons effectué plusieurs tests afin d'identifier les *features* qui rendent la résolution du problème la plus difficile. Par exemple, les *features* liées au secteur (S1–S53) sont peu « coûteuses » à intégrer : leurs valeurs sont pré-calculées et elles ne dépendent que du secteur considéré, ce qui a un impact presque négligeable sur le temps de résolution.

Par exemple écrire un opérateur de maximum générait bien plus de contraintes et variables supplémentaires que l'opérateur de minimum. Il en va de même pour les opérateurs de minimum quand le coefficient de régression associé est négatif.

Les portions de surveillance étant fixes dans l'horizon en cours considéré, les distances sont pré-calculées et nous pouvons fournir à l'avance pour chaque arête son dépôt le plus proche, ce qui réduit considérablement la complexité.

Nous fournissons une solution initiale réalisable au solveur CPLEX pour accélérer la résolution du problème. Pour cela, nous résolvons un problème de planification avec les contraintes de fréquences mais sans fonction objectif.

Résultats La résolution pratique des problèmes à l'aide du PLNE s'avère difficile. La figure ci-dessous illustre, pour chaque itération, la comparaison entre les durées prédites et les durées calculées pour un secteur donné, en l'occurrence le secteur de Vannes, comprenant 100 portions requises.

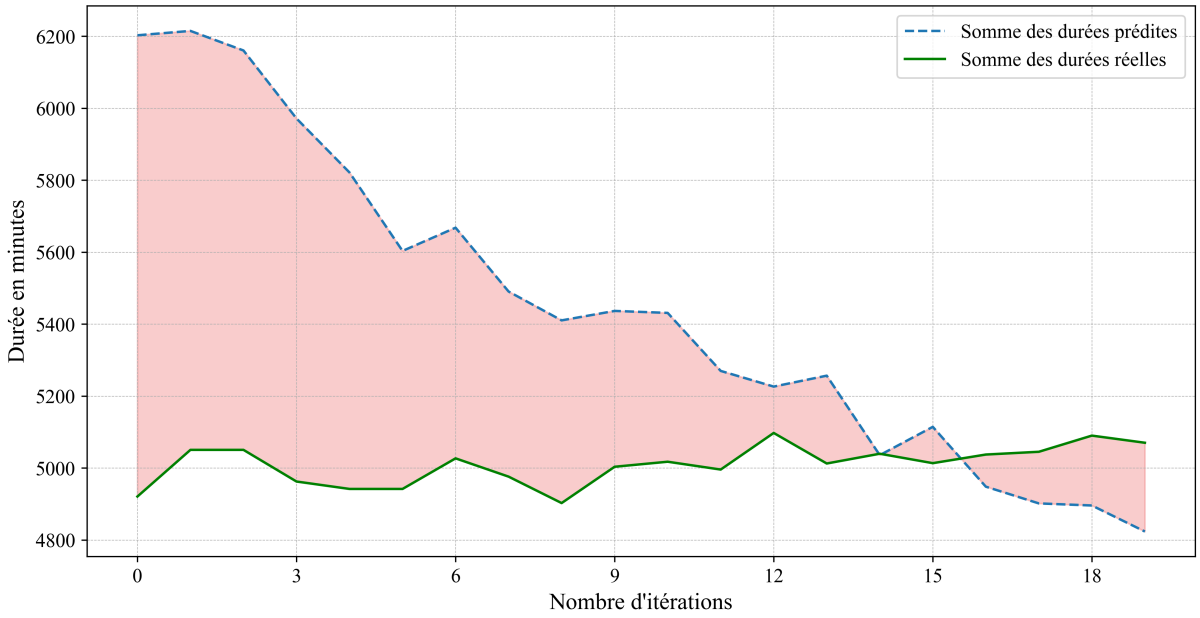


FIGURE 7.1 – Durées des tournées par an et par secteur en fonction de la méthode de tri

Le temps maximal de résolution du modèle de planification en PLNE a été fixé à 1 500 secondes par itération lors de l'utilisation du solveur CPLEX. Ce paramétrage permet d'atteindre, en moyenne, un écart à l'optimalité de 2,3 %.

Nous remarquons que la somme des durées prédites diminue au fur et à mesure des itérations, pour se rapprocher de la somme des durées réelles, ce qui indique que le processus itératif permet d'améliorer la précision du modèle d'apprentissage. En revanche, la somme des durées calculées demeure relativement stable d'une itération à l'autre et ne présente pas d'amélioration significative par rapport à la première itération.

Néanmoins, la formulation en PLNE montre ses limites pour résoudre le module de planification sur l'ensemble des données, particulièrement lorsque deux modes sont considérés simultanément, la taille des variables ainsi que des contraintes augmente considérablement et rend la création du modèle et sa résolution impraticables. Cette flexibilité modale étant pourtant indispensable, nous nous orientons vers une approche heuristique pour traiter le *Module de Planification*. Cette méthode est détaillée ci-après.

7.2 Résolution de la planification par heuristique

Comme expliqué précédemment nous ne parvenons pas à résoudre le *Module de Planification* avec un modèle exact. Dans cette section, nous proposons de résoudre ce problème à l'aide d'une heuristique. Nous gardons l'idée du modèle de régression pour l'estimation rapide de la durée d'une tournée, mais cette fois-ci nous l'intégrons à une heuristique. Les arêtes sont traitées progressivement selon un certain ordre défini, et attribuées à chaque période de temps ainsi qu'à un mode. Dans les heuristiques gloutonnes, le résultat peut varier considérablement selon l'ordre dans lequel les arêtes sont traitées. Nous étudions donc différents ordres de traitement et analysons leur impact sur les résultats.

7.2.1 Principe de l'approche

L'approche intègre progressivement les arêtes au planning, ainsi qu'à un mode de surveillance. Pour chaque arête, il faut déterminer la meilleure combinaison valide, c'est-à-dire celle qui minimise le surcoût ajouté.

Un planning annuel $\mathcal{T}^v_{s,w}$ est initialisé pour chaque secteur $s \in S$, avec une liste vide pour chacune des semaines w de l'année. Un planning annuel $\mathcal{T}^p_w$ est également initialisé pour les tournées aériennes d'un prestataire de surveillance aérienne. Les zones couvertes par chaque

prestataire étant distinctes, ils seront traités indépendamment. Les entrées de l'algorithme comportent l'ensemble des portions dont la surveillance est à planifier. Chaque portion est associée à un ou plusieurs modes ainsi qu'à un secteur d'appartenance. Elles possèdent également un enjeu de surveillance, lui-même lié à une fréquence de passage annuelle, ainsi qu'à un nombre minimal et maximal de semaines entre deux visites consécutives. À partir de la fréquence de surveillance ainsi que des délais minimum et maximum entre deux visites, nous avons établi pour chaque enjeu de surveillance un ensemble de combinaisons possibles de semaines. Formellement, une combinaison est définie par un vecteur de taille 52 pour chaque semaine w de l'horizon W, avec une valeur égale à 1 quand l'arête est affectée à cette semaine, et 0 sinon. Ainsi pour chaque arête l'objectif est d'identifier la ou l'une des combinaisons qui minimise le coût de son insertion à l'aide du modèle prédictif.

Rappelons que les modes de surveillance disponibles varient selon les portions. Certaines ne peuvent être surveillées que par un seul mode et sont donc traitées séparément, tandis que d'autres peuvent l'être selon deux modes possibles. Nous présentons ci-après trois algorithmes : deux dédiés aux cas uni-modaux (terrestre et aérien) et un pour le cas bi-modal. Ces algorithmes seront ensuite exécutés selon un ordre spécifique que nous définirons.

Planification uni-modale dédiée au mode terrestre

L'Algorithme 2 traite de la planification des portions dont la surveillance ne peut être effectuée que par voie terrestre. Cet algorithme est exécuté pour un secteur s donné et prend en entrée les arêtes requises $[i, j] \in R^v(s) \setminus R^p$ appartenant au secteur s ainsi que leurs fréquences de surveillance associées. Le planning annuel $\mathcal{T}^v_{s,w}$ peut être soit vide, soit déjà partiellement pré-rempli, si les portions bimodales ont été traitées au préalable. Les portions sont ensuite regroupées par fréquences, puis classées au sein de chaque fréquence selon un second opérateur de tri que nous définirons par la suite. À l'aide du modèle prédictif, nous estimons pour chaque semaine w le surcoût monétaire associé à l'ajout de la portion considérée. Le vecteur des surcoûts par semaine est mis en regard avec la matrice des combinaisons possibles. Une combinaison est alors sélectionnée parmi les k combinaisons avec le plus petit surcoût pour l'arête requise considérée.

Algorithm 2 Planification uni-modale en voiture par secteur

Initialisation :
Pour chaque secteur $s \in S$, initialiser le planning annuel $\mathcal{T}^v_{s,w}$ avec une liste vide pour chaque semaine $w \in W$ ou récupérer le planning en cours.
Récupérer l'ensemble des arêtes $[i, j] \in R^v(s) \setminus R^p$ appartenant au secteur s ainsi que leurs fréquences de passage associée.

Pour chaque fréquence unique f **faire**
    Récupérer l'ensemble des combinaisons de semaines admissibles pour la fréquence f, compte tenu des contraintes $\Delta_{\min}$ et $\Delta_{\max}$.
    Extraire l'ensemble des arêtes associées à la fréquence f.
    Trier ces arêtes dans une liste L_f selon l'opérateur de tri choisi.
    Pour chaque arête $[i, j]$ dans L_f **faire**
        Estimer, pour chaque semaine $w \in W$, le surcoût lié à l'ajout de l'arête $[i, j]$ à l'aide du modèle d'estimation (régression linéaire).
        Associer ces estimations aux différentes combinaisons de semaines admissibles.
        Sélectionner la combinaison de semaines minimisant le coût.
        Ajouter l'arête $[i, j]$ au planning annuel $\mathcal{T}^v_{s,w}$ selon la combinaison optimale retenue.
    Fin Pour
Fin Pour

retourner planning annuel $\mathcal{T}^v_{s,w}$

Planification uni-modale dédiée au mode aérien

L'Algorithme 3 traite de la planification des portions dont la surveillance ne peut être effectuée que par voie aérienne. Les problèmes étant indépendants entre les différentes zones gérées par les prestataires de surveillance aérienne, les formulations ici ne prennent en compte qu'un prestataire avionneur. L'algorithme prend en entrée les arêtes requises $[i, j] \in R^p \setminus \bigcup_{s \in S} R^v(s)$ ainsi que leurs fréquences de surveillance associées f_{ij}. Le planning annuel $\mathcal{T}_w^p$ peut être soit vide, soit déjà partiellement pré-rempli, si les portions bimodales ont été traitées au préalable. De la même façon que l'Algorithme 2, les arêtes requises sont d'abord classées par fréquence, puis pour chaque fréquence, celles partageant une même valeur sont triées selon un second opérateur de tri. Ensuite, de manière itérative, le modèle prédictif évalue le surcoût induit par l'ajout de l'arête considérée à chaque semaine. Cette estimation est confrontée à la matrice des combinaisons de semaines possibles. La (ou l'une) combinaison(s) minimisant le surcoût global est alors sélectionnée, et l'arête requise est affectée aux semaines correspondantes.

Planification bi-modale des arêtes

L'Algorithme 4 permet de planifier les arêtes requises $[i, j] \in R^p \cap (\bigcup_{s \in S} R^v(s))$ pouvant être surveillées par plusieurs modes. Outre le choix de la meilleure combinaison de semaines, l'algorithme est aussi chargé de choisir le mode qui engendre le plus petit détour. L'algorithme récupère initialise les plannings $\mathcal{T}_{sw}^v$ et $\mathcal{T}_w^p$, ou récupère ceux déjà existants si la planification terrestre, aérienne, ou les deux ont déjà été effectuées. Ainsi, chaque semaine, l'heuristique évalue le coût de l'ajout de l'arête en voiture à l'aide du modèle de prédiction pour la voiture, et le coût de l'ajout de l'arête en avion via le modèle de prédiction pour l'avion, avant de déterminer le mode de transport préférentiel pour cette semaine. Ainsi, de la même manière que pour les deux algorithmes précédents, nous commençons par identifier les fréquences uniques, puis, pour chaque fréquence, nous trions les arêtes correspondantes selon un second critère. Pour chaque arête, nous estimons les surcoûts liés à son ajout à une semaine pour chacun des deux modes de surveillance. Pour chaque semaine, nous conservons le mode présentant le surcoût le plus faible. Ensuite, comme précédemment, nous associons ces estimations aux différentes combinaisons de

Algorithm 3 Planification uni-modale en avion

Initialisation :
Initialiser le planning annuel $\mathcal{T}_w^p$ avec une liste vide pour chaque semaine $w \in W$, ou récupérer le planning en cours.
Récupérer l'ensemble des arêtes $[i, j] \in R^p \setminus \bigcup_{s \in S} R^v(s)$ ainsi que leurs fréquences de passage associées.

Pour chaque fréquence unique f **faire**
    Récupérer l'ensemble des combinaisons de semaines admissibles pour la fréquence f, compte tenu des contraintes $\Delta_{\min}$ et $\Delta_{\max}$.
    Extraire l'ensemble des arêtes associées à la fréquence f.
    Trier ces arêtes dans une liste L_f selon l'opérateur de tri choisi.
    Pour chaque arête $[i, j]$ dans L_f **faire**
        Estimer, pour chaque semaine $w \in W$, le surcoût lié à l'ajout de l'arête $[i, j]$ à l'aide du modèle d'estimation (régression linéaire).
        Associer ces estimations aux différentes combinaisons de semaines admissibles.
        Sélectionner la combinaison de semaines minimisant le coût.
        Ajouter l'arête $[i, j]$ au planning annuel $\mathcal{T}_w^p$ selon la combinaison optimale retenue.
    Fin Pour
Fin Pour

retourner planning annuel $\mathcal{T}_w^p$

semaines et sélectionnons celle qui minimise le coût total.

Algorithm 4 Planification Bi modale

Initialisation :
Initialiser ou récupérer le planning annuel $\mathcal{T}_w^p$ pour le prestataire de surveillance aérienne considéré
Initialiser ou récupérer les plannings annuels $\mathcal{T}_{s,w}^v$ pour les tournées terrestres par secteur s
Récupérer l'ensemble des arêtes $[i,j] \in R^p \cap (\bigcup_{s \in S} R^v(s))$
Pour chaque fréquence unique f **faire**
    Récupérer les combinaisons possibles de semaines pour cette fréquence
    Pour chaque secteur s de S **faire**
        Extraire l'ensemble des arêtes du secteur s associées à la fréquence f.
        Trier ces arêtes dans une liste L_f selon l'opérateur de tri choisi.
        Pour chaque arête $[i,j]$ de L_f **faire**
            Pour chaque mode possible **faire**
                Estimer, pour chaque semaine $w \in W$, le surcoût lié à l'ajout de l'arête $[i,j]$ à l'aide du modèle d'estimation (régression linéaire).
            Fin Pour
            Pour chaque semaine conserver le mode qui minimise le surcoût
            Associer ces estimations aux différentes combinaisons de semaines
            Sélectionner la combinaison de semaines minimisant le coût.
            Ajouter l'arête $[i,j]$ au planning annuel $\mathcal{T}_{s,w}^v$ ou $\mathcal{T}_w^p$ selon la combinaison optimale retenue, et le mode pour chaque semaine.
        Fin Pour
    Fin Pour
Fin Pour

retourner le planning annuel voiture $\mathcal{T}_{s,w}^v$ et le planning annuel avion $\mathcal{T}_w^p$

Procédure illustrée sur un cas représentatif uni-modal Nous illustrons la méthode à l'aide d'un exemple simple comportant peu de combinaisons pour une surveillance uni-modale.

Nous souhaitons insérer une arête $[i,j]$ avec un enjeu correspondant à une visite toutes les 2 semaines. La fréquence est donc $f = 26$ par an, et avec des délais minimums et maximums de 2 et 3 semaines entre chaque visite. Deux combinaisons de visites sont

possibles : la première correspond aux visites sur les semaines paires, la seconde aux visites sur les semaines impaires :

$$\text{combinaison } 1 = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & \dots & 0 & 1 \end{pmatrix} \tag{7.33}$$

$$\text{combinaison } 2 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & \dots & 1 & 0 \end{pmatrix} \tag{7.34}$$

Il convient de noter que pour d'autres ensembles de fréquences et de contraintes de délais minimum et maximum, le nombre de combinaisons possibles peut devenir très élevé. Par exemple, pour une fréquence de 24 avec des délais de 2 à 3 semaines, 23 023 combinaisons sont possibles ; pour une fréquence de 1, ce nombre tombe à 52 ; et pour une fréquence de 12 avec des délais compris entre 3 et 5 semaines, le nombre de combinaisons atteint 122 694.

Nous regroupons l'ensemble des combinaisons possibles dans une matrice de combinaisons M_{comb} de taille $n \times 52$ avec n le nombre de combinaisons (ici égal à deux). Pour notre exemple :

$$M_{\text{comb}} = \begin{pmatrix} 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 1 \end{pmatrix} \tag{7.35}$$

Nous récupérons les vecteurs de coûts prédits $\hat{\mathbf{c}} \in \mathbb{R}^{52}$ pour chaque semaine $w \in W$ avant l'ajout de l'arête :

$$\hat{\mathbf{c}} = \begin{pmatrix} \widehat{c_1} \\ \widehat{c_2} \\ \vdots \\ \widehat{c_{52}} \end{pmatrix} \tag{7.36}$$

Ensuite nous ajoutons l'arête $[i, j]$ aux instances de chaque semaine, et à l'aide du modèle prédictif, on estime le coût de tournées de ces instances pour chaque semaine $w \in W$. On obtient alors un vecteur $\hat{\mathbf{c}}'$ de taille 52 avec les prédictions de coûts pour chaque semaine :

$$\hat{\mathbf{c}}' = \begin{pmatrix} \widehat{c_1}' \\ \widehat{c_2}' \\ \vdots \\ \widehat{c'_{52}} \end{pmatrix} \tag{7.37}$$

Nous calculons l'écart Δ entre les coûts des instances avant et après ajout l'arête $[i, j]$:

$$\Delta = \begin{pmatrix} \max(\widehat{c}_1, \widehat{c}'_1) - \widehat{c}_1 \\ \max(\widehat{c}_2, \widehat{c}'_2) - \widehat{c}_2 \\ \vdots \\ \max(\widehat{c}_{52}, \widehat{c}'_{52}) - \widehat{c}_{52} \end{pmatrix}$$

Le recours à la fonction max permet de neutraliser les écarts négatifs. Il serait en effet incohérent que le coût total diminue suite à l'ajout d'une nouvelle portion. Grâce à cette formulation, Δ ne reflète que l'augmentation réelle du coût générée sur la semaine.

Le produit de la matrice des combinaisons M_{comb} et du vecteur Δ permet d'obtenir un vecteur de taille n. Le surcoût total associé à chaque combinaison $k = 1, \ldots, n$ est alors :

$$\Delta^{(k)} = \sum_{w=1}^{52} M_{\text{comb}}^{(k,w)} \, \Delta_w$$

ou, de manière matricielle :

$$\mathbf{\Delta}_{\text{comb}} = M_{\text{comb}} \cdot \Delta = \begin{pmatrix} \Delta^{(1)} \\ \Delta^{(2)} \\ \vdots \\ \Delta^{(n)} \end{pmatrix}$$

Le vecteur $\mathbf{\Delta}_{\text{comb}}$ contient le surcoût estimé pour chaque combinaison après l'insertion de l'arête $[i, j]$. L'arête $[i, j]$ est affectée à la combinaison k^* telle que :

$$k^* = \arg\min_k \Delta^{(k)}$$

Enfin, nous mettons bien à jour le vecteur de coûts, en ajoutant les surcoûts des

semaines sélectionnées :

$$\hat{\mathbf{c}} \leftarrow \hat{\mathbf{c}} + \Delta^{(k^*)}$$

Et l'arête $[i, j]$ sera bien ajoutée aux instances des semaines sélectionnées.

Procédure illustrée sur un cas représentatif multi-modal Nous souhaitons insérer une arête $[i, j]$ appartenant à un secteur s avec un enjeu correspondant à une visite toutes les 2 semaines, soit une fréquence $f = 26$ visites par an, comme précédemment. Contrairement au cas uni-modal, cette arête peut être surveillée selon deux modes : par avion ou par voiture.

La matrice de combinaisons M_{comb} reste inchangée.

Nous considérons deux vecteurs de coûts initiaux : $\widehat{\mathbf{c}}_p \in \mathbb{R}^{52}$ correspondant aux coûts prédits hebdomadaires des tournées aériennes, et $\widehat{\mathbf{c}}_{v,s} \in \mathbb{R}^{52}$ correspondant aux coûts prédits hebdomadaires des tournées terrestres pour le secteur s :

$$\widehat{\mathbf{c}}_p = \begin{pmatrix} \widehat{c}_{p,1} \\ \widehat{c}_{p,2} \\ \vdots \\ \widehat{c}_{p,52} \end{pmatrix}, \qquad \widehat{\mathbf{c}}_{v,s} = \begin{pmatrix} \widehat{c}_{v,s,1} \\ \widehat{c}_{v,s,2} \\ \vdots \\ \widehat{c}_{v,s,52} \end{pmatrix}$$

Après ajout de l'arête $[i, j]$ aux instances de chaque semaine pour chacun des deux modes, les modèles prédictifs permettent d'estimer les coûts hebdomadaires suivants :

$$\hat{\mathbf{c}}'_p = \begin{pmatrix} \hat{c}_{p,1} \\ \hat{c}_{p,2} \\ \vdots \\ \hat{c}_{p,52} \end{pmatrix}, \qquad \hat{\mathbf{c}}'_{v,s} = \begin{pmatrix} \hat{c}_{v,s,1} \\ \hat{c}_{v,s,2} \\ \vdots \\ \hat{c}_{v,s,52} \end{pmatrix}$$

Les surcoûts hebdomadaires associés à chaque mode de surveillance sont alors définis par les expressions suivantes ; pour le mode terrestre, ils sont en outre calculés pour chaque secteur s :

$$\Delta_p = \max(\widehat{c}_p, \widehat{c}'_p) - \hat{c}_p, \qquad \Delta_{v,s} = \max(\widehat{c}_{v,s}, \widehat{c}'_v) - \hat{c}_{v,s}$$

avec, pour tout $w \in W$:

$$\Delta_{p,w} = \max(\widehat{c}_{p,w}, \widehat{c}'_{p,w}) - \hat{c}_{p,w}, \qquad \Delta_{v,s,w} = \max(\widehat{c}_{v,s,w}, \widehat{c}'_{v,w}) - \hat{c}_{v,s,w}$$

Pour chaque semaine w, et chaque secteur s, nous retenons le mode de surveillance induisant le surcoût minimal. On introduit pour cela les fonctions de sélection suivantes :

$$\delta_p(s,w) = \begin{cases} 1 & \text{si } \Delta_{p,w} \leq \Delta_{v,s,w}, \\ 0 & \text{sinon,} \end{cases} \qquad \delta_v(s,w) = 1 - \delta_p(s,w)$$

Le surcoût hebdomadaire effectif est alors donné par :

$$\Delta_{s,w} = \delta_p(s,w)\,\Delta_{p,w} + \delta_v(s,w)\,\Delta_{v,s,w}$$

Les surcoûts hebdomadaires sont ensuite projetés sur les différentes combinaisons possibles. Le surcoût total associé à chaque combinaison $k = 1, \ldots, n$ est :

$$\Delta^{(k)} = \sum_{w=1}^{52} M_{\text{comb}}^{(k,w)}\,\Delta_w$$

ou, sous forme matricielle :

$$\boldsymbol{\Delta}_{\text{comb}} = M_{\text{comb}} \cdot \Delta$$

La combinaison optimale est alors définie par :

$$k^* = \arg\min_k \Delta^{(k)}$$

Enfin, les vecteurs de coûts sont mis à jour séparément pour chaque mode, en fonction des semaines sélectionnées dans la combinaison k^* :

$$c_{p,w} \leftarrow c_{p,w} + \delta_p(w)\,M_{\text{comb}}^{(k^*,w)}\,\Delta_{p,w}$$

$$c_{v,s,w} \leftarrow c_{v,s,w} + \delta_v(s,w)\, M_{\text{comb}}^{(k^*,w)}\, \Delta_{v,s,w}$$

Enfin, l'arête $[i,j]$ est intégrée aux instances hebdomadaires correspondant à la combinaison k^*, en respectant, pour chaque semaine, le mode de surveillance induisant le surcoût minimal.

Nous effectuons l'évaluation de notre heuristique de planification sur le territoire du Val de Seine (PEVS), où opère l'avionneur Air Loisir. Il s'agit d'un territoire relativement urbanisé, où l'activité de surveillance est importante, tant par voie terrestre qu'aérienne. Cette heuristique étant gloutonne, nous devons définir l'ordre de traitement des portions de surveillance, des modes, et des fréquences. En effet, l'ordre dans lequel nous traitons puis insérons les portions a un impact sur la qualité de la solution finale.

Dans la suite de cette section, nous ferons référence à la durée ou au coût prédit, correspondant aux valeurs estimées par les modèles de régression, ainsi qu'à la durée ou au coût calculé, correspondant aux valeurs obtenues après résolution des modèles de tournées.

7.2.2 Ordre des fréquences et ordre des portions

L'ordre de traitement des arêtes joue un rôle essentiel. Un premier critère, déjà exploité dans certaines heuristiques de planification, repose sur la fréquence de passage requise pour chaque portion. Nous pouvons considérer les fréquences par ordre croissant ou décroissant. Dans notre problème elles varient de 1 à 52, correspondant respectivement à un passage annuel et à un passage hebdomadaire. Les enjeux de surveillance possibles, et donc les fréquences de passage associées, sont en nombre limité : nous avons plus de 20 000 portions pour seulement cinq types d'enjeux. Par conséquent, de nombreuses arêtes se voient attribuer les mêmes fréquences de passage. Ainsi, un tri basé uniquement sur ces fréquences ne suffit pas. C'est pourquoi nous étudions des opérateurs de tri secondaires permettant de départager l'ordre de traitement des arêtes présentant une même fréquence. Dans tous les cas, la fréquence reste utilisée comme critère primaire, ce qui permet d'éviter de devoir ré-énumérer toutes les combinaisons possibles. Les opérateurs proposés sont les suivants :

— **proximité par rapport au dépôt** : tri des arêtes en fonction de leur proximité avec le dépôt principal ;

— **éloignement par rapport au dépôt** : tri des arêtes en fonction de leur distance au dépôt principal ;

— **taille croissante** : tri des arêtes selon leur longueur, par ordre croissant ;

— **taille décroissante** : tri des arêtes selon leur longueur, par ordre décroissant.

Pour les opérateurs basés sur la proximité ou l'éloignement par rapport au dépôt, lorsque plusieurs dépôts sont présents dans le secteur considéré, la distance retenue est la distance moyenne aux dépôts.

La première méthode testée consistait à trier les fréquences par ordre décroissant (donc en commençant par les arêtes nécessitant une visite hebdomadaire) puis, au sein de chaque fréquence, à ordonner les arêtes en fonction de leur proximité au dépôt. L'analyse des solutions de planification obtenues avec cette approche a toutefois mis en évidence, pour certains secteurs (notamment LAMORLAYE, PONTOISE et PROVENCE), une forte dispersion des arêtes requises au sein de chaque semaine.

Afin d'évaluer l'impact du mode de tri sur la planification, différentes stratégies de tri des arêtes ont alors été étudiées. Les secteurs présentant les résultats les moins satisfaisants ont en commun la présence d'arêtes devant être surveillées chaque semaine (enjeu Très Fort). L'analyse s'est donc concentrée sur ces secteurs, tout en conservant un secteur « témoin », PALAISEAU. Dans ce cadre, l'étude porte non seulement sur le tri des arêtes partageant une même fréquence, mais également sur l'ordre de traitement des fréquences elles-mêmes. Les 15 stratégies de tri suivantes sont ainsi considérées :

— **Tri 1** : Fréquences décroissantes, Arêtes triées aléatoirement

— **Tri 2** : Fréquences décroissantes mais avec enjeu Très Fort en dernier, Arêtes triées aléatoirement

— **Tri 3** : Fréquences croissantes, Arêtes triées aléatoirement

— **Tri 4** : Fréquences décroissantes, Éloignement dépôt

— **Tri 5** : Fréquences décroissantes mais avec enjeu Très Fort en dernier, Éloignement dépôt

— **Tri 6** : Fréquences croissantes, Éloignement dépôt

— **Tri 7** : Fréquences décroissantes, Proximité dépôt

— **Tri 8** : Fréquences décroissantes mais avec enjeu Très Fort en dernier, Proximité dépôt

— **Tri 9** : Fréquences croissantes, Proximité dépôt

— **Tri 10** : Fréquences décroissantes, Taille croissante

— **Tri 11** : Fréquences décroissantes mais avec enjeu Très Fort en dernier, Taille croissante

— **Tri 12** : Fréquences croissantes, Taille croissante

— **Tri 13** : Fréquences décroissantes, Taille décroissantes

— **Tri 14** : Fréquences décroissantes mais avec enjeu Très Fort en dernier, Taille décroissante,

— **Tri 15** : Fréquences croissantes, Taille décroissante

Il convient de noter que, pour les secteurs ne comportant pas de portions à enjeu « Très Fort » (comme le secteur Palaiseau), certains tris sont équivalents : 1 et 2, 4 et 5, 7 et 8, 10 et 11, ainsi que 13 et 14.

La Figure 7.2 présente, pour les quatre secteurs considérés, la somme annuelle des durées des tournées en fonction des différentes stratégies de tri des arêtes utilisées lors de la planification.

Pour le secteur de Palaiseau, qui ne comporte pas de portions associées à un enjeu « Très Fort », les durées annuelles restent relativement proches quelle que soit la stratégie de tri considérée. En revanche, pour les trois autres secteurs, qui incluent des arêtes à enjeu « Très Fort », cinq opérateurs de tri parmi les quinze évalués conduisent à des durées sensiblement plus élevées que celles obtenues avec les autres méthodes. Ces cinq opérateurs ont en commun le fait de trier en priorité les portions selon un ordre décroissant de leurs fréquences de passage, et donc de commencer par placer les portions avec des enjeux « Très Fort ».

La Table 7.1 présente les différentes méthodes de tri classées selon les sommes de durées annuelles calculées ou prédites associées pour les 4 secteurs étudiés. Pour chaque tri, nous

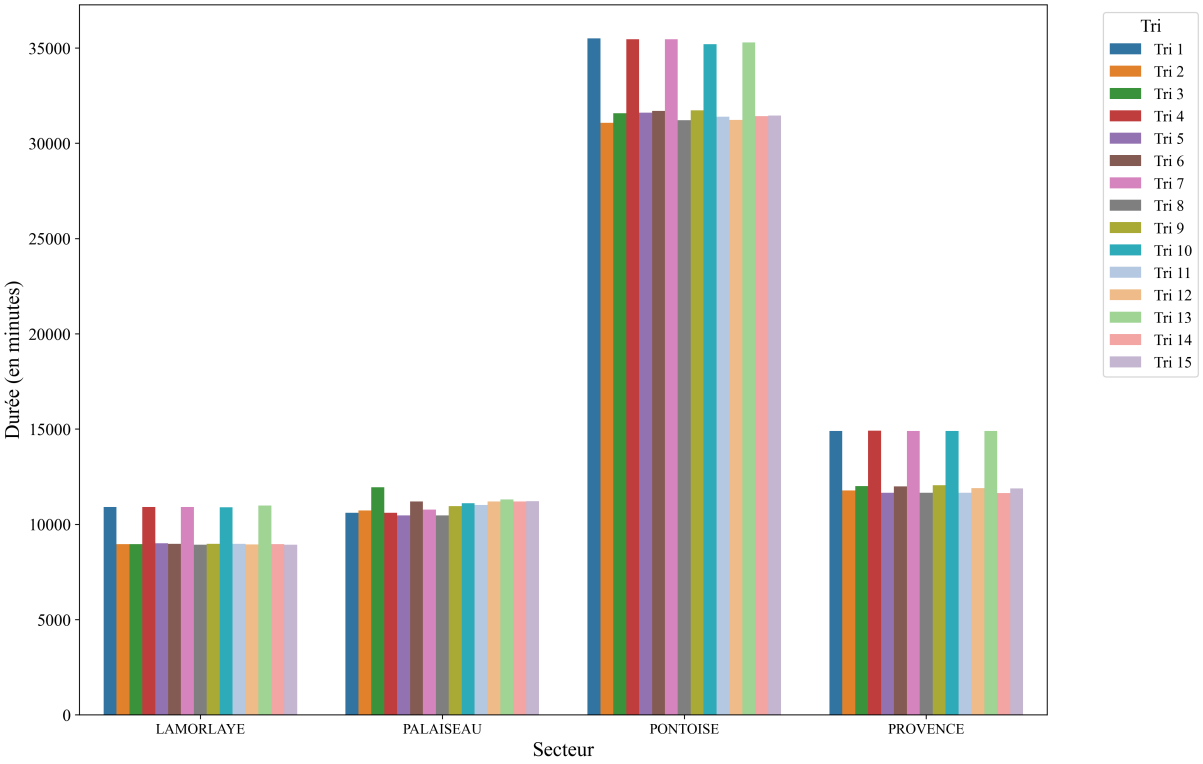


FIGURE 7.2 – Durées des tournées par an et par secteur en fonction de la méthode de tri

calculons également l'écart relatif en pourcentage par rapport à la durée calculée obtenue avec le meilleur tri.

On observe tout d'abord que l'ordre obtenu à partir des durées calculées diffère légèrement de celui fondé sur les durées prédites, certaines stratégies de tri étant inversées (écrites en rouge). Par ailleurs, les stratégies conduisant aux durées calculées les plus courtes sont celles qui traitent les arêtes selon un ordre de fréquence décroissant, en accordant ensuite la priorité aux arêtes associées à un enjeu « Très Fort ». Ces tris présentent un écart de durée inférieur à 2 % par rapport au meilleur tri (Tri 8). Les tris suivants reposent sur un traitement des portions selon une fréquence croissante et conduisent à des durées comprises entre 2 et 3,4 % au-dessus de la durée calculée du meilleur tri. Enfin, les durées calculées les plus élevées sont observées pour les stratégies reposant sur un tri strictement décroissant des fréquences avec une durée totale plus de 14 % plus élevée que la durée du meilleur tri. Ainsi, la stratégie présentant les durées totales calculées et prédites les plus faibles correspond au Tri 8 : les arêtes sont d'abord ordonnées selon des fréquences

décroissantes, mais avec les arêtes à enjeu « Très Fort » traitées en dernier, puis classées selon une proximité croissante au dépôt.

TABLE 7.1 – Ordre des Tris par somme des durées prédites

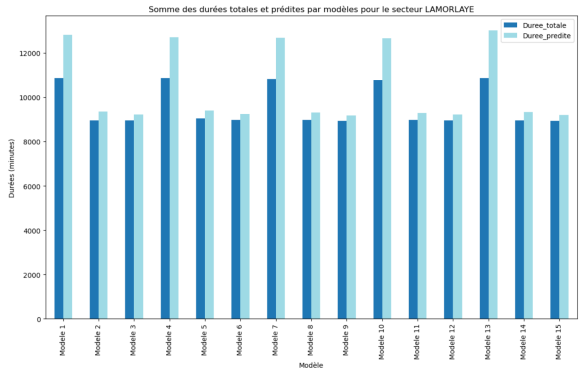
	Tris et durées calculées			**Tris et durées prédites**		
Rank	**Tri**	**Durée**	**Ecart**	**Tri**	**Durée**	**Ecart**
1	Tri 8	62293.56	0%	Tri 8	64569.43	0%
2	Tri 2	62525.33	0.37%	Tri 2	64902.52	0.51%
3	Tri 5	62885.6	0.95%	Tri 5	65210.32	0.99%
4	Tri 11	63189.2	1.43%	Tri 11	65264.66	1.07%
5	Tri 14	63520.09	1.95%	Tri 12	65414.19	1.3%
6	Tri 12	63551.81	2%	Tri 14	65531.9	1.48%
7	Tri 9	64033.56	2.75%	Tri 9	65657.4	1.67%
8	Tri 15	64128.05	2.9%	Tri 6	65684.31	1.71%
9	Tri 6	64304.36	3.18 %	Tri 15	65766.63	1.84%
10	Tri 3	64436.98	3.38%	Tri 3	66326.15	2.68%
11	Tri 4	71798.55	14.18%	Tri 4	75974.90	16.23%
12	Tri 1	71919.87	14.34%	Tri 7	76009.70	16.28%
10	Tri 7	72011.17	14.47%	Tri 1	76042.63	16.32%
11	Tri 10	72240.84	14.79%	Tri 10	76320.55	16.68%
12	Tri 13	72733.08	15.46%	Tri 13	76995.28	17.56%

Nous poursuivons à présent par une analyse sectorielle afin de vérifier si ces conclusions se vérifient indépendamment du secteur considéré. La Figure 7.3 présente les sommes de durées calculées et prédites pour chaque secteur et chaque stratégie de tri. Le modèle prédictif a tendance a surestimer les durées des tournées, excepté pour le secteur PROVENCE où on a une sous estimation par rapport aux durées calculées. On distingue bien que pour les 3 secteurs avec de la surveillance hebdomadaire, les stratégies de tris avec une fréquence décroissante stricte montrent des solutions moins intéressantes que les autres stratégies. Les résultats sont un peu plus variables pour le secteur de Palaiseau qui ne présente pas de surveillance hebdomadaire.

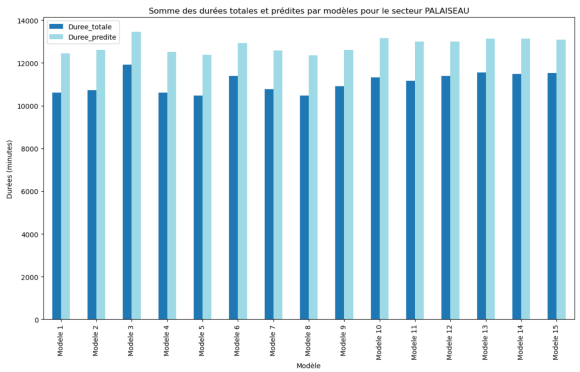
Les six premiers rangs des méthodes de tri par secteur, selon les durées calculées, sont présentés dans la Table 7.2. Le Tri 8 reste globalement bien positionné pour l'ensemble des secteurs étudiés. Bien qu'il ne figure pas parmi les six premiers pour le secteur de LAMORLAYE, la Figure 7.3a met en évidence que les Tris 2, 3, 5, 6, 8, 9, 11, 12, 14 et 15

conduisent, pour ce secteur, à des durées calculées et prédites très proches. Dans ce cas, le choix de l'opérateur de tri secondaire a un impact limité sur les performances, à l'exception des stratégies fondées sur un tri strictement décroissant des fréquences, lesquelles entraînent des durées nettement plus élevées.

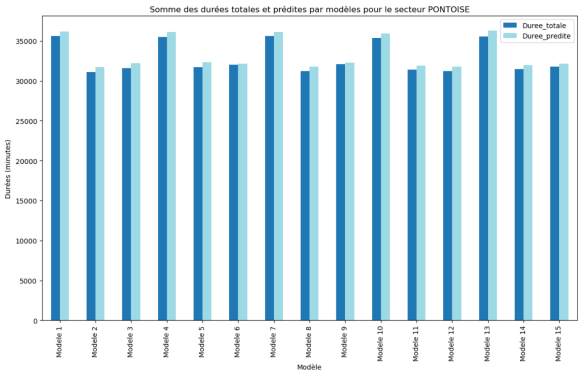
Finalement, nous retenons le Tri 8, qui affiche les durées les plus faibles pour le secteur de PALAISEAU (ne contenant pas de portions avec surveillance hebdomadaire) tout en étant bien classé pour les autres secteurs où la surveillance hebdomadaire est présente.



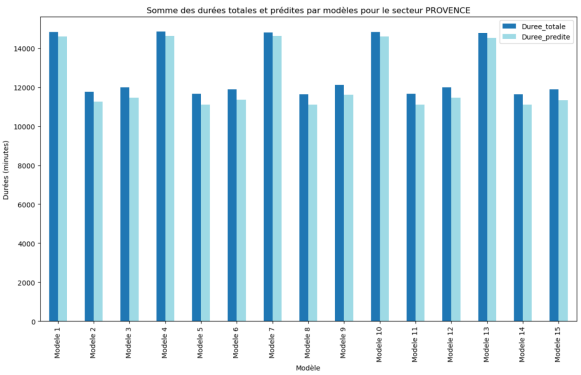
(a) LAMORLAYE



(b) PALAISEAU



(c) PONTOISE



(d) PROVENCE

FIGURE 7.3 – Comparaison des durées totales et prédites en fonction des méthodes d'ordre de traitement des arêtes pour les quatre secteurs étudiés.

La Table 7.3 présente les performances du modèle prédictif pour chaque secteur lors de cette première itération de l'approche à deux phases. Conformément à la méthode, une fois la planification réalisée, nous lançons le *Module de Routage* afin d'obtenir les durées dites calculées. Nous évaluons les écarts entre les durées prédites et les durées calculées pour chaque semaine. On constate que le coefficient de corrélation de Spearman reste toujours

TABLE 7.2 – Ordre des modèles par somme des durées totales et par secteur

	PALAISEAU		**PROVENCE**		**PONTOISE**		**LAMORLAYE**	
Rang	**Tri**	**Durée**	**Tri**	**Durée**	**Tri**	**Durée**	**Tri**	**Durée**
1	Tri 8	10469.1	Tri 14	11642.6	Tri 2	31065.6	Tri 9	8932.2
2	Tri 5	10469.7	Tri 8	11646.1	Tri 8	31204.3	Tri 15	8934.3
3	Tri 1	10601.4	Tri 11	11649.6	Tri 12	31220.3	Tri 12	8947.7
4	Tri 4	10604.1	Tri 5	11655.7	Tri 11	31397.9	Tri 3	8955.1
5	Tri 2	10722.7	Tri 2	11774.3	Tri 14	31429.2	Tri 2	8962.7
6	Tri 7	10771.3	Tri 15	11880.3	Tri 3	31570.9	Tri 14	8962.8

supérieur à 0,97, bien que le modèle prédictif puisse présenter des erreurs moyennes. Les erreurs moyennes sont assez variables d'un secteur à l'autre avec une MAE allant de 11,4 à 50,9 minutes, une RMSE allant de 13,2 à 62,3 minutes et une MAPE allant de 3,8 à 34,5 %. De plus, ce sont des résultats pour une première itération, sans modèle ré-entraîné.

TABLE 7.3 – Performance du modèle prédictif sur différents secteurs

Secteur	**MAE**	**RMSE**	**MAPE**	**Corr.**
4 secteurs confondus	24.6	35.2	14.7%	0.985
Pontoise	13.5	18.9	3.8%	0.992
Lamorlaye	26.8	30.8	15.7%	0.974
Palaiseau	50.9	62.3	34.5%	0.995
Provence	11.4	13.2	7.9%	0.988

Approche avec de l'aléa Afin de rendre l'heuristique moins gloutonne, nous avons testé différentes méthodes de choix d'une combinaison de planning pour chaque arête en rajoutant de l'aléa. Nous avons ainsi gardé la logique du Tri 8 qui permet de traiter les arêtes par rapport à leur proximité au dépôt d'abord. La première variation consistait à sélectionner, pour chaque arête, une combinaison de semaines parmi celles dont le coût se situe à 1 %, 2 % ou 5 % du coût de la meilleure combinaison. Cette approche n'a pas été retenue, car au-delà d'un certain nombre d'arêtes, la différence de coût devenait trop faible et toutes les arêtes restantes entraient dans l'intervalle considéré.

Nous avons donc testé une autre variation consistant à choisir une arête parmi les 2, 3 ou 5 arêtes les plus proches du dépôt. Bien que donnant de meilleurs résultats que la première variation, ces tris restent moins efficaces que la stratégie du Tri 8.

Tri par composante connexe Nous avons remarqué que pour certains secteurs, les résultats pouvaient varier en fonction des exécutions lancées. C'était notamment le cas du secteur de PONTOISE. En effet bien qu'on utilise deux opérateurs de tri différents pour déterminer l'ordre de traitement des arêtes requises, il arrive qu'on obtienne des ordres différents en fonction des lancements.

Pour stabiliser les solutions tout en accélérant le processus, il est également possible d'insérer les arêtes par composantes connexes. En effet, l'ensemble des portions présentant la même fréquence étant souvent relativement connecté, il est alors possible d'intégrer directement un groupe de portions dans le planning.

Ainsi, nous avons exécuté 100 fois l'heuristique de planification pour ce secteur : la moitié des simulations utilise les portions de surveillance individuellement, tandis que l'autre moitié se base sur les composantes connexes de ces portions. Les résultats, présentés dans la Figure 7.4, montrent que la distribution des durées (en minutes) est légèrement plus étendue lorsque l'on considère les arêtes isolément, alors que les prédictions sont plus stables avec les composantes connexes. De plus, les durées prédites en utilisant les composantes connexes sont en moyenne plus faibles que celles obtenues avec les arêtes seules, avec un écart d'environ 0,15 % dans cet exemple. Même si cette baisse n'est pas très importante, l'insertion des composantes connexes plutôt que des arêtes permet d'accélérer le processus en réduisant considérablement le nombre d'itérations. Selon la connectivité des arêtes, l'utilisation des composantes peut diviser jusqu'à trois fois le nombre d'itérations nécessaires.

7.2.3 Ordre des secteurs pour la surveillance aérienne et bimodale

Dans les points précédents, nous nous sommes concentrés sur la surveillance terrestre par secteur et avons identifié les meilleurs ordres d'insertion des portions de surveillance.

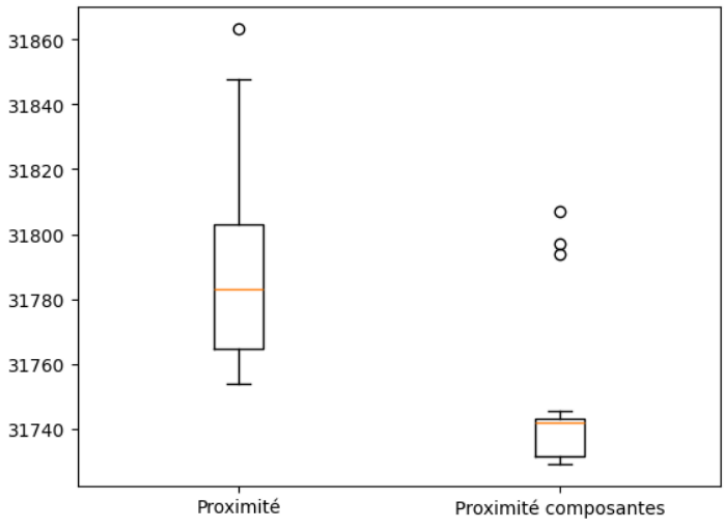


FIGURE 7.4 – Comparaison des durées prédites pour le secteur de PONTOISE en triant par arêtes ou par composantes

Cette problématique se pose également lorsqu'il s'agit de traiter les arêtes pouvant être surveillées par voie aérienne, que ce soit exclusivement ou de manière bimodale. Nous ne comparons plus les durées comme dans le cas unimodal mais les coûts monétaires. En effet l'objectif global est d'optimiser les coûts liés aux tournées de surveillance, or une heure de survol en avion n'a pas le même coût qu'une heure de parcours en voiture. Dans ce cas, nous fixons conventionnellement le coût d'une heure de survol à 35 € et celui d'une heure de déplacement en voiture à 15 €. Ces valeurs ne correspondent pas aux coûts réels, lesquels ne sont pas disponibles, seules les données de coût par kilomètre surveillé étant connues. L'objectif est uniquement de respecter un rapport de coût, le mode aérien étant environ 2,3 fois plus cher que le mode terrestre.

Dans ce cadre, nous nous intéressons plus particulièrement à l'ordre de traitement des secteurs. Nous choisissons de procéder secteur par secteur, car chaque secteur nécessite la récupération de ses matrices de distances et de durées pour la surveillance terrestre. Récupérer toutes les matrices simultanément pourrait alourdir le processus, tandis que les extraire à chaque traitement de portion, selon le secteur concerné, rallongerait le temps de calcul.

Dans la Figure 7.5, nous comparons différents tris pour le traitement des arêtes bimodales, de manière similaire à ce qui a été fait pour les portions avec surveillance

terrestre. Ces tris concernent l'ordre de traitement des secteurs, l'ordre au sein de chaque secteur, ainsi que l'ordre global : trier les portions individuellement ou par composantes connexes de portions.

Pour ordonner les secteurs, nous prenons en compte leur proximité avec l'aérodrome principal de l'avionneur. Ainsi, nous calculons la distance de chaque dépôt de secteur à cet aérodrome, puis classons les secteurs par ordre croissant de distance.

Nous pouvons tout d'abord confirmer que traiter arête par arête montre de moins bons résultats que traiter par composante connexe. Enfin l'ordre des secteurs a une importance dans l'ordre de traitement des arêtes bimodales.

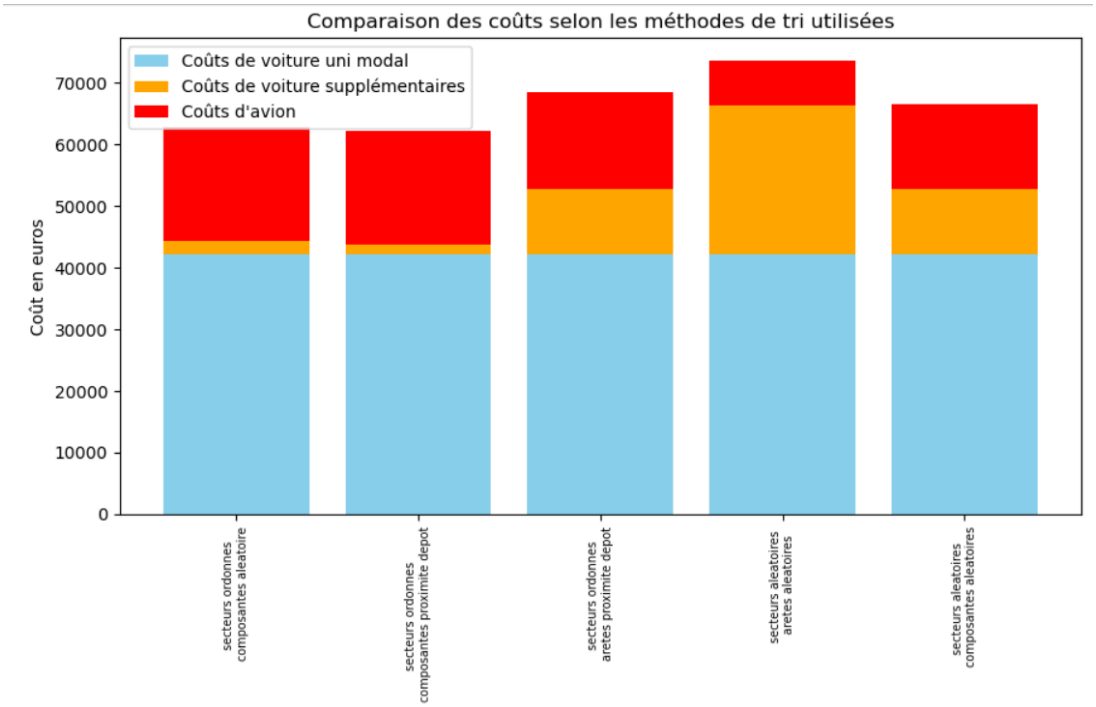


FIGURE 7.5 – Comparaison des durées prédites pour AIR LOISIR

7.2.4 Ordre des modes

Nous distinguons trois types de portions de surveillance : celles qui ne peuvent être surveillées que par voie aérienne, celles qui ne peuvent l'être que par voie terrestre, et celles pouvant être surveillées selon les deux modes. Il est donc nécessaire de définir l'ordre dans lequel ces modes de surveillance doivent être traités. Par conséquent, la planification doit

être effectuée successivement pour les modes terrestre, aérien, puis pour le cas bi-modal combinant surveillance terrestre et aérienne.

Cinq configurations possibles d'ordre de traitement des modes ont donc été testées :

— **Ordre 1 :** Aérien unimodal – Bimodal – Terrestre unimodal

— **Ordre 2 :** Unimodal (Aérien et Terrestre) – Bimodal

— **Ordre 3 :** Bimodal – Unimodal (Aérien – Terrestre)

— **Ordre 4 :** Terrestre unimodal – Bimodal – Aérien unimodal

— **Ordre 5 :** Terrestre à enjeu « Très Fort » – Bimodal – Unimodal (Aérien – Terrestre)

Dans la Figure 7.6 nous représentons les coûts pour une itération selon différents ordres de traitement des modes dans l'approche.

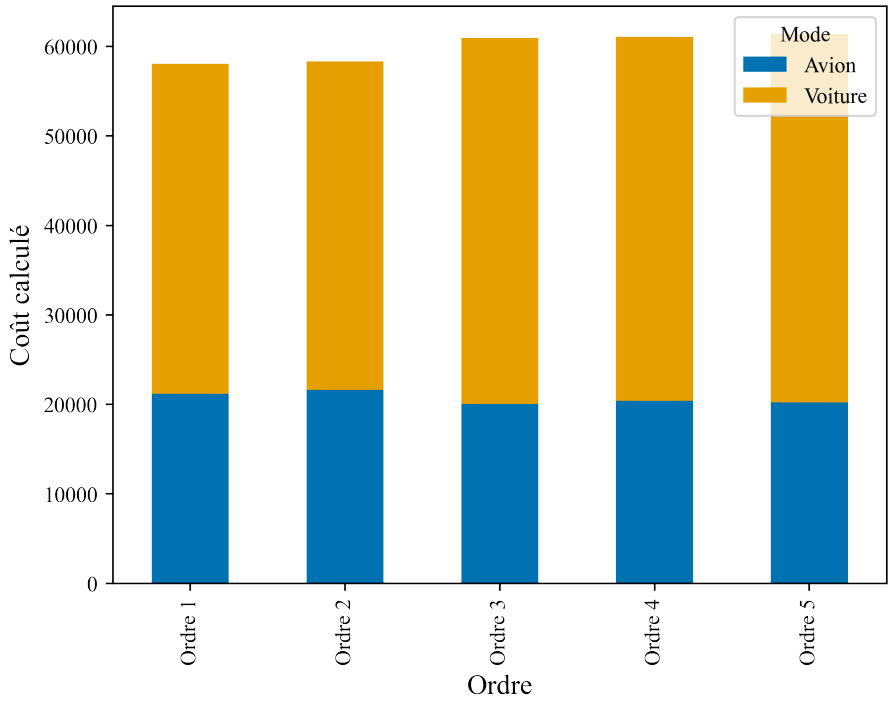


FIGURE 7.6 – Comparaison des coûts selon les ordres de modes empruntés

Les deux premiers ordres de traitement des modes correspondent aux coûts les plus faibles et ont en commun de commencer par la surveillance aérienne. Toutefois, le premier ordre se distingue par le coût minimal qu'il permet d'atteindre. Pour une première itération, l'ordre de traitement des modes qui minimise le coût consiste à traiter en priorité les

portions ne pouvant être surveillées qu'en avion, suivies des arêtes pouvant être surveillées par les deux modes, et enfin des portions exclusivement terrestres. De plus, les tournées terrestres, organisées par secteur, fonctionnent indépendamment les unes des autres. Si nous débutons par ce mode seul, les plannings risquent de se désynchroniser d'un secteur à l'autre. Cela pourrait ensuite affecter les tournées aériennes, qui, elles, sont multi-secteurs.

7.3 Conclusion

Dans ce chapitre, nous avons proposé une approche pour la résolution du Module de Planification. La taille et la complexité de notre problème, notamment dues à la présence de multiples modes, ne permettaient pas de résoudre la phase de planification à l'aide d'un modèle linéaire. Nous avons donc proposé une heuristique visant à traiter successivement les arêtes requises afin de les intégrer dans les plannings correspondants. Le caractère glouton de cette méthode nécessitait l'identification de différentes stratégies d'ordonnancement des portions, puis la sélection de la plus performante. Nous présentons dans le chapitre suivant le résultats des expérimentations de la méthode itérative globale.

Chapitre 8

Résultats de la méthode

Dans ce chapitre, nous présentons les résultats de l'approche itérative appliquée à deux jeux de données : le premier, correspondant à l'année 2023, sur lequel la méthode a été initialement mise en place, et le second, relatif au premier semestre 2025 des activités de surveillance de NaTran. Cette analyse nous permet également de comparer les résultats obtenus avec les tournées effectivement réalisées. Le problème et les pratiques de surveillance ayant légèrement évolué depuis 2024, nous avons souhaité tester notre méthode sur des données plus récentes afin de la valider avant de procéder à des expérimentations en conditions réelles sur différentes zones du territoire.

8.1 Évaluation de l'approche itérative en 2023

8.1.1 Configuration de la méthode

Dans le Chapitre 7 nous avons proposé une heuristique pour le planification des tournées de surveillance tout en identifiant une bonne combinaison de tris et d'ordre de modes définie pour l'ordre de traitement des arêtes requises. Nous avons choisi d'appliquer et d'évaluer notre approche sur le territoire couvert par le prestataire de surveillance aérienne AirLoisir. Ce territoire comprend à la fois des secteurs variés et des zones très urbanisées, où la surveillance automobile est particulièrement importante.

En 2023, le territoire couvert par AirLoisir comprend 6 018 portions de surveillance : 371 peuvent être surveillées à la fois par voie routière et aérienne (bi-modales), 2 992 sont accessibles uniquement par voie terrestre, et 2 655 uniquement par voie aérienne. Concernant les fréquences de surveillance :

— « Enjeu Limité » : 496 portions nécessitent une surveillance annuelle, 160 une surveillance trimestrielle et 20 une surveillance semestrielle,

— « Enjeu Normal » : 2 260 portions requièrent une surveillance mensuelle,

— « Enjeu Fort » : 2 565 portions demandent une surveillance bi-mensuelle,

— « Enjeu Très Fort » : 112 portions font l'objet d'une surveillance hebdomadaire.

— « Enjeu Clos » : 495 portions font l'objet de surveillance de parcelles closes.

Dans les analyses effectuées dans le Chapitre 7 nous nous focalisions sur une seule itération, or les modèles prédictifs ont été entraînés sur des données de simulation, donc leur précision initiale n'est pas toujours optimale. L'avantage d'une approche itérative est de pouvoir mettre à jour les bases d'apprentissage des modèles prédictifs, permettant ainsi d'améliorer leur précision à mesure que de nouvelles données sont intégrées. Nous décidons que les itérations s'arrêtent quand on améliore la solution de moins de 1% entre deux itérations.

Pour la résolution des tournées terrestres, nous utilisons l'heuristique k-BIH. Afin de garantir la cohérence, les instances de la base d'apprentissage du modèle prédictif des coûts de tournées terrestres comprennent également des tournées résolues à l'aide de cette même heuristique. De manière similaire, pour les tournées aériennes, nous appliquons d'abord la méthode permettant de réduire la taille de l'instance, suivie de l'heuristique des k-NNH. Par souci de cohérence, les instances de la base d'apprentissage du modèle prédictif des coûts de tournées aériennes incluent également des tournées résolues selon cette même approche. Enfin, lors des différentes itérations, nous considérons le modèle mono-tournée et utilisons donc les modèles prédictifs correspondants. Cela s'explique par plusieurs raisons : la résolution en multi-tournées peut être coûteuse, surtout pour les grandes instances ; les modèles prédictifs de coûts donnent de meilleurs résultats dans le cadre mono-tournée ; et, en pratique, la capacité totale considérée n'est pas atteinte à l'échelle hebdomadaire, même lorsque l'ensemble des portions est surveillé. À l'issue des itérations, les mono-tournées hebdomadaires seront alors divisées en multi-tournées. Par ailleurs, nous avons conservé les coûts horaires présentés dans le chapitre précédent, à savoir 15€ pour les tournées terrestres et 35 € pour les tournées aériennes. Bien que ces valeurs ne soient pas représentatives des coûts réels, elles permettent néanmoins de distinguer clairement les deux modes de surveillance.

8.1.2 Résultats de l'approche itérative

La Figure 8.1 présente les résultats obtenus à l'issue de l'approche itérative en deux phases. Les barres bleues représentent la somme des coûts calculés (à l'issue de la phase de routage), tandis que les barres rouges correspondent à la somme des coûts prédits par

les modèles de régression, lors de la phase de planification. Les sous-barres en bleu clair et bleu foncé indiquent respectivement les coûts calculés associés à la surveillance terrestre (automobile) et à la surveillance aérienne. De la même manière, les sous-barres en rouge clair et rouge foncé représentent les coûts prédits par le modèle de régression pour la surveillance terrestre et la surveillance aérienne. Nous remarquons déjà que l'écart entre la somme des coûts prédits et calculés diminue au fur et à mesure des itérations. Par ailleurs, la somme totale des coûts calculés diminue également à mesure que les itérations avancent.

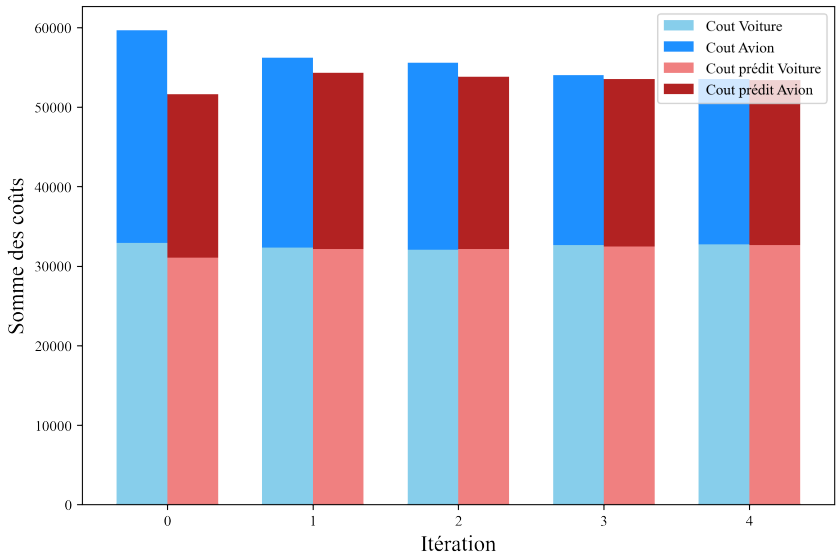


FIGURE 8.1 – Évolution des coûts calculés et estimés au cours des itérations pour chaque mode

Dans la Table 8.1 nous présentons la somme des coûts prédits et calculés pour chaque itération. La dernière colonne « Écart » montre en pourcentage l'écart entre les coûts calculés de la i-ème itération par rapport à la première. Le coût total calculé diminue au fur et à mesure des itérations pour passer de 59 674,10 € à 53 550,47 €. L'approche effectue au total 5 itérations car l'amélioration entre la 4-ième et la 5-ième itération est de 0,86 %. Entre la première et la dernière itération les coûts totaux diminuent de 10,26 %.

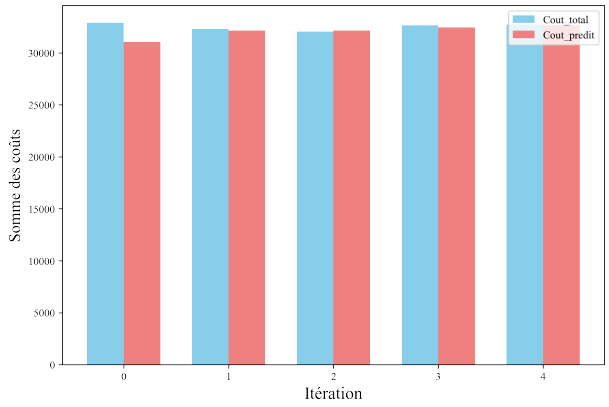
Nous pouvons analyser les méthodes par mode de surveillance.

Dans la Figure 8.2a, on observe que la somme des coûts prédits et calculés pour la surveillance terrestre augmente au fil des itérations, en raison du nombre croissant de portions de surveillance affectées à la voiture plutôt qu'à l'avion. La Figure 8.2b montre

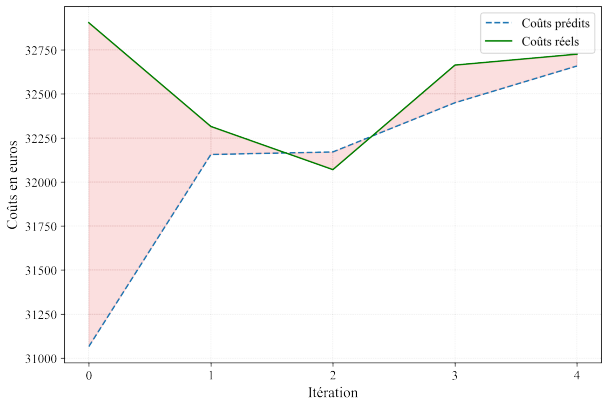
TABLE 8.1 – Coûts prédits et calculés et par itération

Itération	Prédit	Calculé	Écart
1	51 634,35	59 674,10	0%
2	54 353,07	56 248,56	-5,74%
3	53 839,59	55 607,47	-6,82%
4	53 516,63	54 014,36	-9,48%
5	53 414,72	53 550,47	-10,26%

quant à elle que l'écart entre la somme des coûts prédits et celle des coûts calculés se réduit progressivement au cours des itérations. Dès la première itération, cet écart est de 5,92 %, ce qui reste relativement faible. Il diminue ensuite pour atteindre seulement 0,21 % à la dernière itération.



(a) Évolution itérative des coûts réels et prédits des tournées

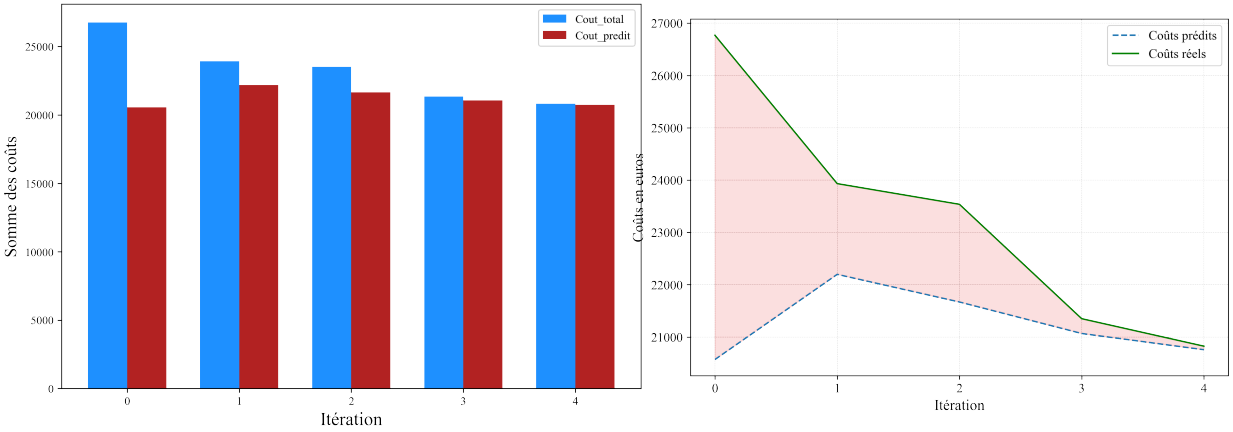


(b) Évolution itérative de l'écart entre coûts prédits et réels

FIGURE 8.2 – Évolution des coûts réels et prédits pour la surveillance terrestre

La Figure 8.3a montre que la somme des coûts calculés et prédits pour la surveillance aérienne décroît au fil des itérations, tout comme l'écart entre ces deux valeurs (Figure 8.3b). Le modèle se révèle moins précis lors de la première itération pour les tournées aériennes contrairement aux tournées terrestres, avec un écart de 30,15 % entre la somme des coûts calculés et la somme des coûts prédits. Toutefois, cet écart relatif se réduit progressivement au cours des itérations pour atteindre 0,33 % à la dernière itération.

Tant pour la prédiction des coûts des tournées terrestres que pour celles des tournées aériennes, l'écart entre la somme des coûts prédits et la somme des coûts réels diminue



(a) Évolution itérative des coûts réels et prédits des tournées

(b) Évolution itérative de l'écart entre coûts prédits et réels

FIGURE 8.3 – Évolution des coûts réels et prédits pour la surveillance aérienne

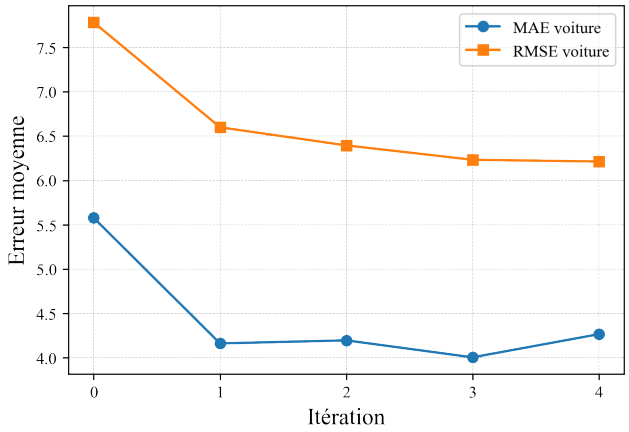
au fil des itérations. Il convient désormais de vérifier si les modèles prédictifs commettent effectivement moins d'erreurs à mesure que les itérations progressent et que les bases d'apprentissage s'enrichissent. Ainsi, nous calculons à chaque itération les erreurs de prédictions avec la MAE, RMSE et MAPE pour chaque instance de tournée de l'itération.

La Figure 8.4 présente l'évolution des MAE et RMSE par mode. Plus spécifiquement, la Figure 8.4a détaille l'évolution des erreurs pour les tournées terrestres. Avant la première itération, la MAE et la RMSE s'élèvent respectivement à 5,57 € et 7,78 €. Après cette itération, elles se stabilisent entre 4 et 4,26 € pour la MAE, et entre 6,21 et 6,6 € pour la RMSE. Pour les tournées aériennes (Figure 8.4b), la baisse de l'erreur moyenne est plus flagrante : dès la première itération, la MAE chute de 122 à 36,5 € et la RMSE de 132,49 à 46,94 €. À la dernière itération, la MAE atteint 12,36 € et la RMSE 14,44 €.

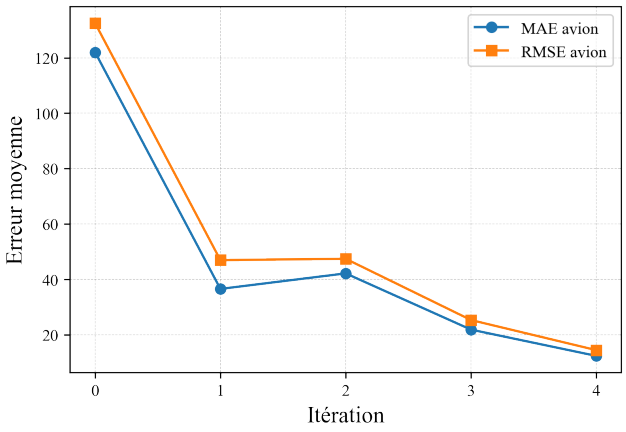
La Figure 8.5 présente l'évolution des MAPE par mode. Les tournées terrestres (en rouge) présentent d'abord une erreur moyenne de 32,07 % à la première itération, elle atteint 14,41 % à la dernière itération.

Les tournées aériennes (en bleu) présentent d'abord une erreur moyenne de 48,92 % à la première itération, elle atteint 8,01 % à la dernière itération.

Ainsi, les modèles prédictifs s'améliorent effectivement au fil du temps. Dans le cas du modèle terrestre, l'amélioration est particulièrement marquée dès la seconde itération,



(a) MAE et RMSE du modèle prédictif pour les tournées terrestres



(b) MAE et RMSE du modèle prédictif pour les tournées aériennes

FIGURE 8.4 – Évolution itérative des erreurs MAE et RMSE pour les modèles prédictifs de coûts par mode de transport

avec une stabilisation plus rapide de l'erreur. En revanche, pour les tournées aériennes, on observe une amélioration marquée dès la seconde itération, suivie d'une diminution de l'erreur plus progressive par la suite.

Dans la Figure 8.6, nous présentons les résultats de la planification des tournées aériennes à l'issue de la dernière itération pour les 6 premières semaines de l'année. Les portions sont représentées selon leur enjeu ou leur fréquence de surveillance. Les portions à enjeu « Très fort » sont affichées en rouge, celles à enjeu « Fort » en orange, les portions à enjeu normal en jaune, celles à enjeu limité en vert, les portions à surveillance trimestrielle en bleu et celles à surveillance semestrielle en bleu clair.

Nous remarquons que la méthode a tendance à sélectionner des portions proches au sein de chaque semaine. Aussi, les profils se répètent par exemple toutes les 4 semaines : comme les semaines 1 (Figure 8.6a) et 5 (Figure 8.6e) ; voire toutes les deux semaines : comme les semaines 2 (Figure 8.6b), 4 (Figure 8.6d) et 6 (Figure 8.6f).

La Figure 8.7 présente les cartes de surveillance du secteur de BRETIGNY-SUR-ORGE pour les six premières semaines. Ce secteur a été choisi pour exemple car il comprend des portions à surveillance hebdomadaire, des portions bimodales ainsi que des portions uni-modales. Les tournées suivent un motif identique qui s'alterne entre semaines paires

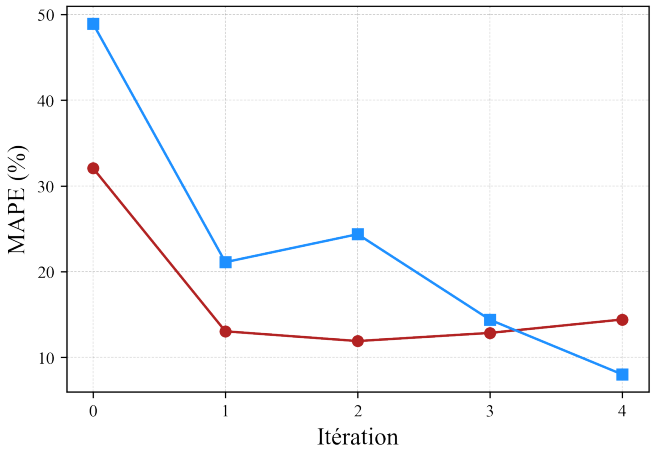


FIGURE 8.5 – Évolution des coûts réels et estimés au cours des itérations pour chaque mode : terrestre (en rouge) et aérien (en bleu)

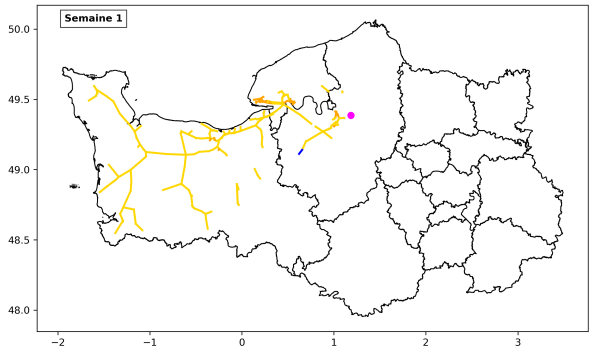
et semaines impaires sur l'ensemble de l'horizon étudié. Lors des semaines impaires (Figures 8.7a, 8.7c et 8.7e), les portions bimodales sont surveillées par voie aérienne. À l'inverse, lors des semaines paires (Figures 8.7b, 8.7d et 8.7f), les portions nécessitant deux inspections mensuelles sont visitées en voiture, car elles ne peuvent pas être survolées.

Cela montre que l'approche permet de construire correctement les tournées en tenant compte des contraintes liées aux modes de surveillance et à la planification, tout en allouant les portions aux modes appropriés afin d'obtenir des tournées de coût minimal.

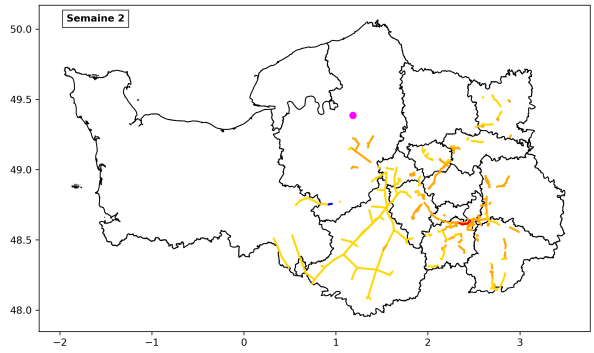
La Table 8.2 présente les temps de calcul par algorithme pour l'approche itérative. En moyenne, chaque itération nécessite 88,23 minutes. Pour un total de 5 itérations, le temps cumulé atteint 441,17 minutes, soit environ 7 heures et 21 minutes. À ce temps s'ajoutent le passage en multi-tournées : 1 heure 25 minutes 32 secondes pour les tournées terrestres et 2 heures 12 minutes et 40 secondes pour les tournées aériennes.

8.1.3 Comparaison avec les tournées de surveillance effectivement réalisées

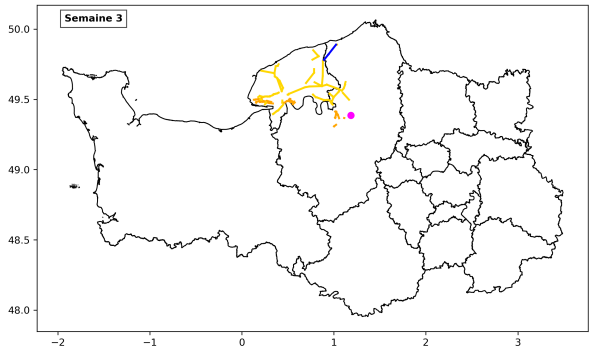
Le problème périodique de tournées sur les arêtes avec capacité et flotte multi-modale étant nouveau, aucune méthode existante de la littérature ne permet actuellement de le résoudre. C'est pour cela que nous comparons les résultats avec les tournées réalisées par



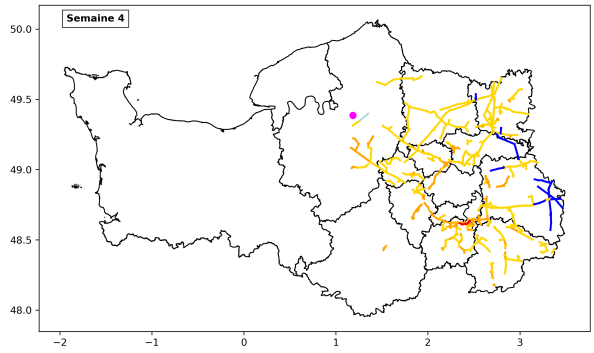
(a) Semaine 1 : Portions sélectionnées pour surveillance aérienne



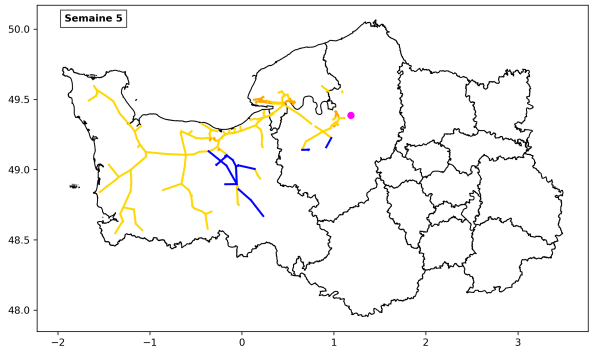
(b) Semaine 2 : Portions sélectionnées pour surveillance aérienne



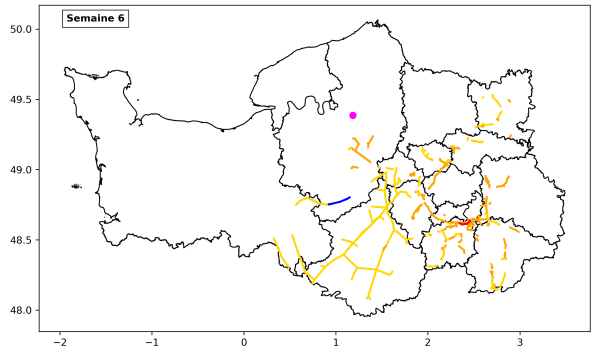
(c) Semaine 3 : Portions sélectionnées pour surveillance aérienne



(d) Semaine 4 : Portions sélectionnées pour surveillance aérienne

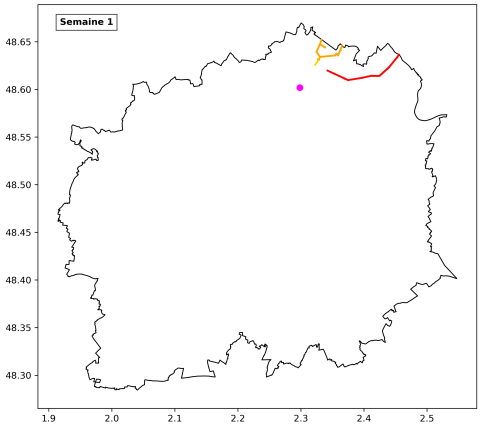


(e) Semaine 5 : Portions sélectionnées pour surveillance aérienne

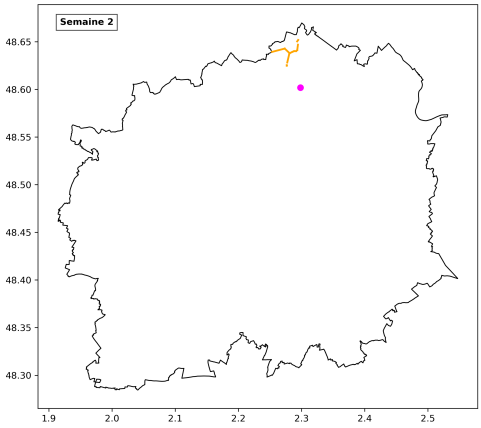


(f) Semaine 6 : Portions sélectionnées pour surveillance aérienne

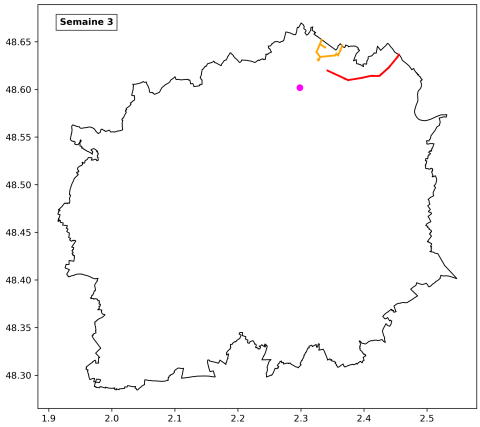
FIGURE 8.6 – Portions sélectionnées par semaine pour la surveillance aérienne



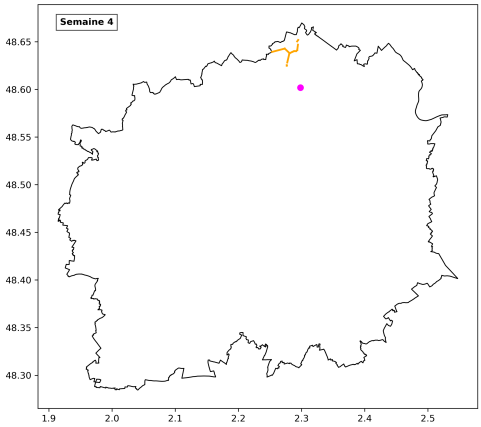
(a) Semaine 1 : Portions sélectionnées pour surveillance terrestre



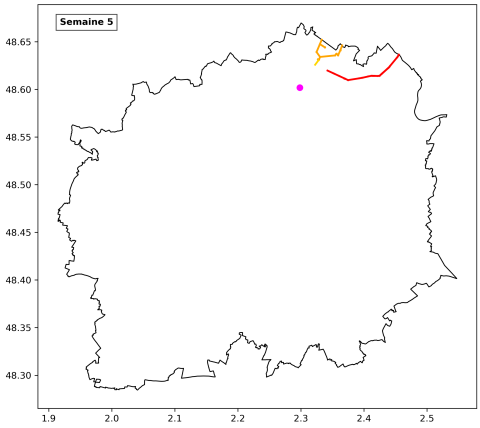
(b) Semaine 2 : Portions sélectionnées pour surveillance terrestre



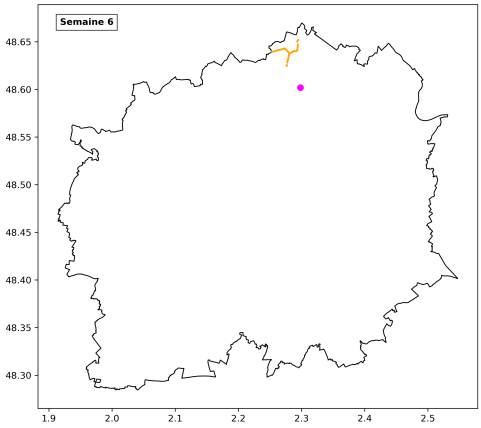
(c) Semaine 3 : Portions sélectionnées pour surveillance terrestre



(d) Semaine 4 : Portions sélectionnées pour surveillance terrestre



(e) Semaine 5 : Portions sélectionnées pour surveillance terrestre



(f) Semaine 6 : Portions sélectionnées pour surveillance terrestre

FIGURE 8.7 – Portions sélectionnées par semaine pour la surveillance aérienne

	Temps par itération (en minutes)		**Temps pour toutes les itérations (en minutes)**
	Temps moyen	**Temps std**	
Planification aérienne	43,27	23,01	216,36
Planification bimodale	26,98	4,80	134,88
Planification terrestre	5,78	1,49	28,92
Tournées aériennes	6,362	1,42	31,81
Tournées terrestres	5,42	0,67	27,1
Ré-entraînement	0,42	0,02	2,1
Total	88,23	29,49	441,17

TABLE 8.2 – Temps de calcul moyen par algorithme

les exploitants.

Nous avons pu récupérer les traces GPS des tournées terrestres et aériennes réalisées en 2023. Ces données permettent d'identifier, pour chaque jour, les portions surveillées ainsi que le mode de surveillance utilisé. Afin d'assurer une comparaison cohérente et de reconstruire correctement les tournées, nous appliquons les mêmes méthodes d'optimisation que celles utilisées par le module de routage. Nous parlerons alors de durées, et kilométrages « réalisés ».

Dans la Table 8.3, nous présentons les durées, et kilométrages calculés à l'aide de notre approche et ceux réalisés, par mode. Nous y reportons la durée totale des tournées, le kilométrage total parcouru et le kilométrage de surveillance, correspondant aux distances sur lesquelles une surveillance est effectivement réalisée le long des portions. Le mode hélicoptère est également renseigné, car il a été utilisé pour la surveillance de quelques portions spécifiques. Toutefois, son usage est voué à disparaître et il ne fait donc pas partie de notre approche. Nous calculons une diminution de 36,75 % de la durée totale des tournées, et de 33,44% du kilométrage total de tournées par rapport à ce qui a été réalisé en 2023.

Le nombre d'heures allouées aux tournées aériennes diminue de 6,23 %, tandis que celui des tournées terrestres recule de 42,05 %. On observe ainsi une réduction nettement plus marquée du temps consacré à la surveillance terrestre, au profit d'un report vers la surveillance aérienne.

	Durée (en h)		**Kilométrage Tournée**		**Kilométrage Surveillance**	
	Réalisé	**Calculé**	**Réalisé**	**Calculé**	**Réalisé**	**Calculé**
Avion	859,52	807,81	133 859,13	123 516,76	76 107,51	52 737,72
Terrestre	4 713,82	2 731,74	181 047,77	85 693,53	91 694,47	46 179,81
Hélicoptère	2	0	123,785	0	2,5	0
Total	5 596,32	3 539,54	314 302,215	209 210,29	167 801,98	98 917,53

TABLE 8.3 – Comparaison des résultats par mode pour la surveillance en 2023

En 2023, la surveillance aérienne représentait 15,39 % du temps alloué à la surveillance, 42,39 % du kilométrage parcouru, et 45,36 % du kilométrage de portions parcouru. Avec notre méthode la surveillance aérienne représente 22,82 % du temps alloué à la surveillance, 59,04 % du kilométrage parcouru, et 53,32 % du kilométrage de portions parcouru. Ainsi on remarque une augmentation de la part de la surveillance aérienne dans la surveillance globale, et une bonne diminution de la surveillance terrestre.

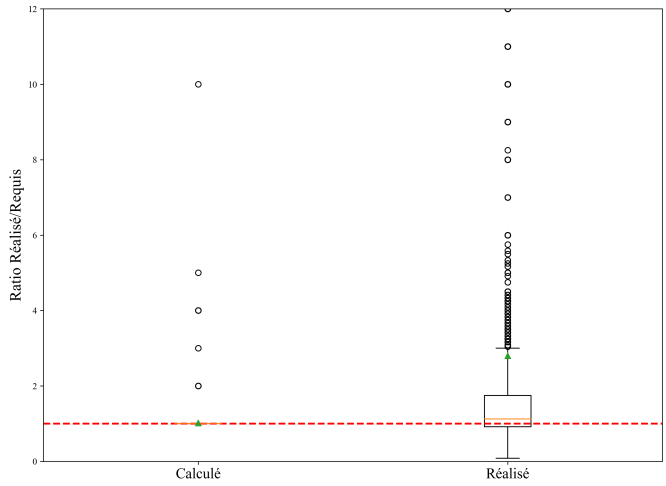
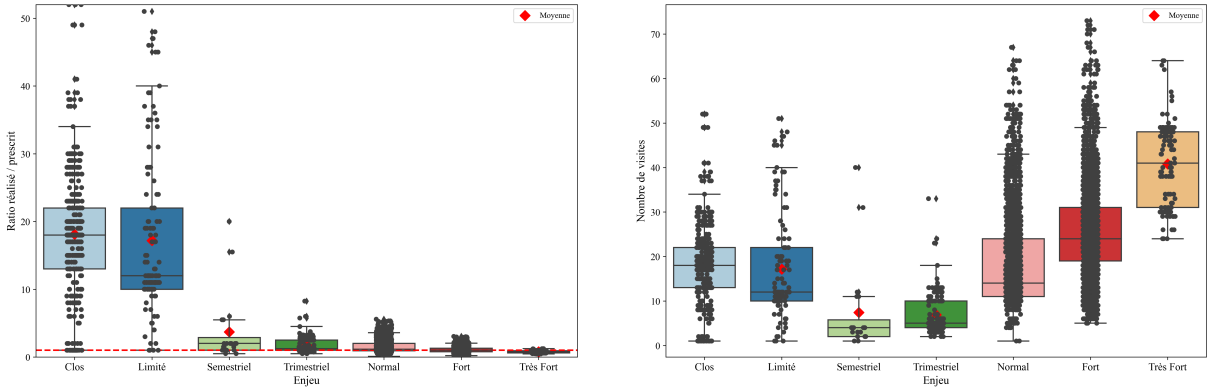


FIGURE 8.8 – Distribution des rapports entre le nombre de surveillances réalisées et requises en 2023

Dans la Figure 8.8, nous présentons le rapport entre le nombre de surveillances réalisées et le nombre de surveillances requises, comparant notre méthode aux tournées effectuées en 2023. Pour améliorer la lisibilité, l'échelle de la figure se concentre sur la majorité des valeurs ; les valeurs extrêmes ne sont donc pas visibles à cette échelle, mais sont bien présentes dans les données. En moyenne, dans les tournées réalisées par NaTran, une

portion est surveillée 2,8 fois plus que le nombre requis, contre seulement 1,02 fois avec notre approche. Afin d'expliquer ces fréquences élevées, nous étudions plus en détail la solution de NaTran, en analysant la répartition des passages par portion et par enjeu de surveillance.



(a) Distribution des ratios

(b) Distribution du nombre de surveillances

FIGURE 8.9 – Analyse des surveillances par enjeu en 2023 : rapports et nombres de passages réalisés comparés aux prescriptions par enjeu de surveillance

La Figure 8.9 présente la distribution des rations entre le nombre de passages réalisés dans la solution de NaTran et le nombre requis (Figure 8.9a), ainsi que les distributions du nombre de surveillances annuelles des portions, regroupées par enjeu de surveillance (Figure 8.9b). Sur la figure 8.9a, nous voyons que, comme attendu, les enjeux nécessitant le moins de surveillances présentent les distributions les plus étendues : en moyenne, les portions à « Enjeu Limité » (surveillance annuelle) sont surveillées 17,18 fois plus que nécessaire, celles à « Enjeu Clos » 18,12 fois, les surveillances semestrielles 3,7 fois, les surveillances trimestrielles 1,72 fois, les portions à « Enjeu Normal » 1,57 fois, celles à « Enjeu Fort » 1,07 fois, et enfin celles à « Enjeu Très Fort » seulement 0,79 fois, soit en dessous du requis.

Cette augmentation du nombre de surveillances par portion de canalisation par rapport aux prescriptions s'explique par la structure des ROS, qui regroupe, pour rappel, des portions indépendamment de leur fréquence afin de les surveiller périodiquement avec les mêmes modes.

Dans la Figure 8.9b on observe que le nombre de visites peut varier significativement

d'une portion à l'autre au sein d'un même enjeu. Étonnamment, en moyenne, les portions à « Enjeu Limité » et « Enjeu Clos » génèrent globalement plus de passages que celles soumises à des surveillances semestrielles ou trimestrielles.

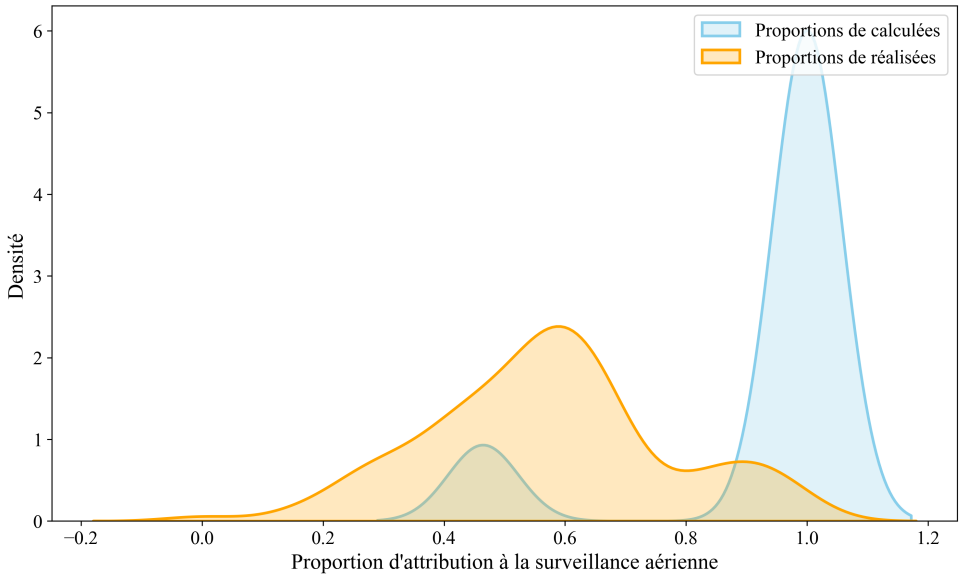
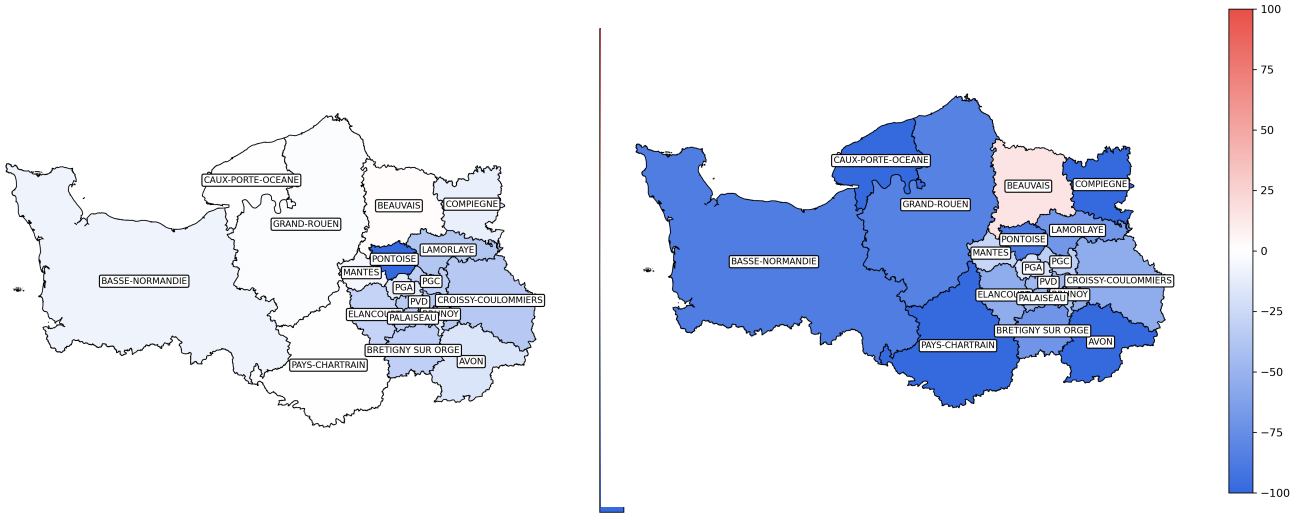


FIGURE 8.10 – Distribution de la proportion de passages attribués aux tournées aériennes par portion bimodale, comparant les valeurs calculées et réalisées en 2023

La Figure 8.10 montre la distribution de la proportion de passages attribués à la surveillance aérienne pour les portions bimodales. Pour chaque portion pouvant être surveillée par plusieurs modes, nous avons déterminé les modes utilisés à chaque passage, et calculé la part affectée à la surveillance aérienne. La courbe bleue correspond à la répartition calculée avec notre méthode, tandis que la courbe orange reflète la répartition réalisée. Pour les tournées de surveillance calculées, 86,39 % des portions sont surveillées exclusivement par voie aérienne. Pour 12,43 % des portions, la répartition des passages correspond à 6 passages aériens sur 13, le reste étant effectué au sol, tandis que 1,18 % des portions présentent une répartition équitable entre surveillance aérienne et terrestre. Ainsi, avec notre méthode, la surveillance aérienne représente en moyenne 92,72 % des passages par portion bimodale. À l'inverse, la part de surveillance réalisée par avion dans les tournées effectivement réalisées est plus variable selon les portions. En moyenne, l'aérien y représente 57 % des passages par portion bimodale.

Dans la Figure 8.11, nous présentons une analyse sectorielle de l'écart de durée allouée

à la surveillance terrestre entre les valeurs réalisées et nos résultats. Dans la Figure 8.11a, on observe que le secteur de PONTOISE enregistre la plus forte diminution, avec une baisse de 633,41 heures allouées aux modes terrestres. À l'inverse, le secteur de BEAUVAIS est le seul à connaître une augmentation, mais modérée, de 14,27 heures. En termes d'écart en pourcentage (Figure 8.11b), certains secteurs n'effectuent plus de surveillance terrestre et affichent ainsi une baisse de 100 % de leurs heures. C'est le cas des secteurs CAUX PORTE OCEANE et AVON.



(a) Écart de durées de surveillance (en h)

(b) Écart de pourcentages de durées de surveillance

FIGURE 8.11 – Analyse des écarts de surveillance entre les tournées réalisées et les tournées calculées pour 2023

8.2 Évaluation de l'approche itérative en 2025

Dans cette courte section, nous évaluons notre approche sur un nouveau jeu de données correspondant au problème de tournées de surveillance en 2025.

8.2.1 Contexte opérationnel

Une réorganisation importante a été menée au sein de l'entreprise en 2024. Celle-ci a conduit à une refonte de la structure des directions ainsi qu'à un redécoupage sectoriel,

incluant des fusions de secteurs. Les entités opérationnelles sont désormais organisées en équipes d'exploitation, au nombre de 47. Cette réorganisation a également introduit des évolutions dans les modes d'organisation, notamment dans un objectif d'harmonisation des pratiques entre les équipes.

Concernant les activités de surveillance, un nouveau prestataire de surveillance aérienne, nommé Sourse, est venu s'ajouter aux deux prestataires historiques pour couvrir une partie du nord de la France. Ses avions sont équipés de solutions d'intelligence artificielle embarquée destinées à la détection d'engins, tels que des pelleteuses. À ce stade, cet outil de détection est encore en cours de fiabilisation par l'entreprise.

Par ailleurs, le déploiement de la surveillance satellitaire permet de réduire le recours à la surveillance automobile dans les zones urbaines les plus denses, où la fréquence de surveillance est hebdomadaire à enjeux « Très Fort ».

8.2.2 Périmètre considéré, analyse des données

En 2025, le réseau de transport est divisé en 21 820 portions de surveillances dont 619 qui sont réglementaires réparties sur 47 équipes d'exploitation. L'équipe d'exploitation Aquitaine Sud est celle qui compte le plus grand nombre de portions, avec 967 portions de surveillance, tandis que l'équipe la moins dotée est CAZ qui en recense 160. En moyenne, chaque équipe d'exploitation est responsable de 471 portions de surveillance.

Pour notre étude, nous considérons à nouveau le territoire opéré par AirLoisir, qui comprend 14 équipes d'exploitation en charge de la surveillance de 7 381 portions. Parmi celles-ci, 3 562 sont surveillées uniquement par voie aérienne, 1 846 uniquement par voie routière, et 1 973 en mode bimodal. On constate que le nombre de portions bimodales a augmenté, ce qui correspond à l'introduction récente du satellite comme mode de surveillance. En effet, 1 503 de ces portions combinent désormais le satellite et la surveillance routière. Les images satellite permettent de couvrir un large périmètre, notamment dans la région parisienne où la surveillance aérienne est impossible. Comme ces images sont prises à des semaines fixes, nous avons simplement adapté les combinaisons de semaines possibles pour les portions concernées. Étant donné que cela ne génère pas de déplacement, le coût reste identique entre le réalisé et le calculé. C'est pourquoi nous n'affichons pas les coûts

liés à la surveillance satellitaire : il n'y a pas d'optimisation possible et cela n'affecte pas la comparaison. Concernant les fréquences de surveillance nous retrouvons des portions à :

— « Enjeu Limité » : 234 portions nécessitent une surveillance annuelle, 194 une surveillance trimestrielle et 19 une surveillance semestrielle,

— « Enjeu Clos » : 468 portions nécessitent une surveillance annuelle,

— « Enjeu Normal » : 3 035 portions requièrent une surveillance mensuelle,

— « Enjeu Fort » : 2 756 portions demandent une surveillance bi-mensuelle,

— « Enjeu Très Fort » : 144 portions font l'objet d'une surveillance hebdomadaire.

Dans nos calculs précédents, nous avons utilisé des coûts indicatifs, non représentatifs des coûts réels, définis uniquement à des fins de comparaison entre la surveillance terrestre et aérienne, ces derniers n'étant pas disponibles. À partir des tournées précédemment réalisées par les exploitants et les différents prestataires, ainsi que des coûts par kilomètre de canalisation surveillée fournis par l'entreprise, nous avons estimé de nouveaux coûts horaires, a priori plus proches de la réalité. Ceux-ci s'élèvent à 255 € par heure pour la surveillance terrestre et 675 € par heure pour la surveillance aérienne. Ces valeurs restent des estimations, mais elles permettent surtout à notre approche de répartir correctement les portions entre les différents modes de surveillance.

8.2.3 Résultats et comparaison

La Figure 8.12 présente l'évolution de la somme des coûts calculés et estimés au fur et à mesure des itérations. La première itération permet une réduction de plus de 45 % des coûts calculés, après quoi ceux-ci se stabilisent au cours des itérations suivante. Nous constatons également que l'écart entre la somme des durées prédites et calculées diminue dès la première itération, puis se stabilise en conservant un certain décalage. Cela peut s'expliquer notamment par le fait que la base d'apprentissage contient des observations issues d'instances correspondant aux données antérieures à la réorganisation de 2024.

Nous avons récupéré les portions de surveillance réalisées (PSR) du premier semestre 2025. Les tournées ont ensuite été reconstruites et résolues pour chaque journée du planning réalisé, en respectant systématiquement les modes de surveillance effectivement utilisés.

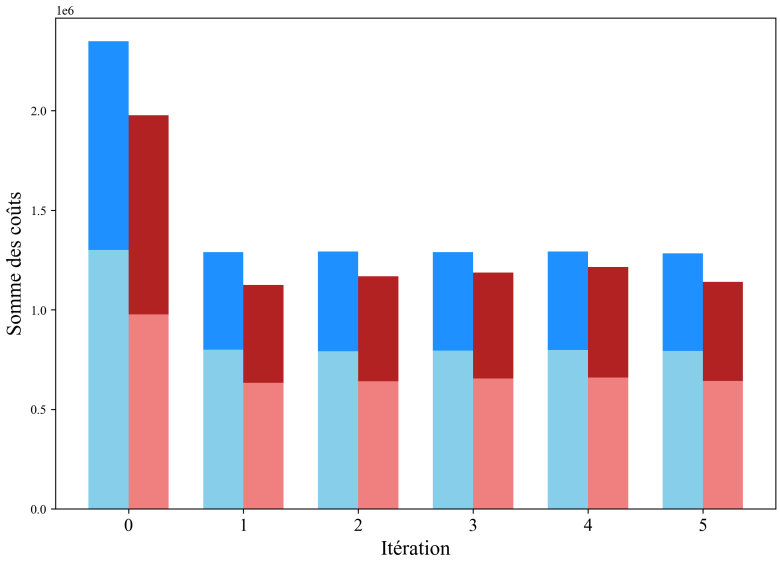


FIGURE 8.12 – Évolution des coûts calculés et estimés au cours des itérations pour chaque mode

	Durée (en h)		**Kilométrage Tournée**		**Kilométrage Surveillance**	
	Réalisé	**Calculé**	**Réalisé**	**Calculé**	**Réalisé**	**Calculé**
Avion	734,03	533,19	118 406,13	88 773,68	58 406,26	41 485,83
Terrestre	2 719,82	1 849,16	92 331,36	59 863,88	50 913.77	27 989,17
Hélicoptère	3,4	0	780,08	0	0,03	0
Total	3 457,25	2 382,35	211 517,57	148 637,56	109 320,06	69 764,54

TABLE 8.4 – Comparaison des résultats par mode pour la surveillance en 2025

La Table 8.4 présente les durées et le kilométrages des tournées qui ont été réalisées par NaTran au premier semestre 2025 vs les tournées calculées à l'aide de notre approche. Cela représente une baisse du nombre d'heures de 31,09% (27,6 % pour les tournées aériennes et 32,01 % pour les tournées terrestres), une baisse du nombre de kilomètres parcourus de 29,72% et une baisse du kilométrage de portions parcourus de 36,18%. Contrairement aux résultats obtenus pour l'année 2023, on n'observe pas de différence marquée entre la diminution de la surveillance terrestre et celle de la surveillance aérienne. En 2025, la surveillance aérienne représentait 21,33 % du temps alloué à la surveillance, 56,35 % du kilométrage parcouru, et 53,43 % du kilométrage de portions parcouru. Avec notre méthode la surveillance aérienne représente 22,38 % du temps alloué à la surveillance,

59,72 % du kilométrage parcouru, et 59,47 % du kilométrage de portions parcouru. Ainsi on remarque une légère hausse de la part de la surveillance aérienne dans la surveillance globale, mais qui reste plus modérée qu'en 2023.

La Figure 8.13 illustre la distribution de la proportion de passages aériens pour chaque portion de surveillance bimodale. La courbe bleue correspond à la répartition obtenue avec notre méthode, alors que la courbe orange représente la répartition effectivement réalisée. En moyenne, dans la solution de NaTran, 45,3 % des passages de surveillance pour les arêtes bimodales sont réalisés en avion. Avec notre méthode 19,9 % des passages de surveillance pour les arêtes bimodales sont attribués à la surveillance aérienne. Nous augmentons la surveillance routière pour 7 équipes d'exploitation parmi les 15.

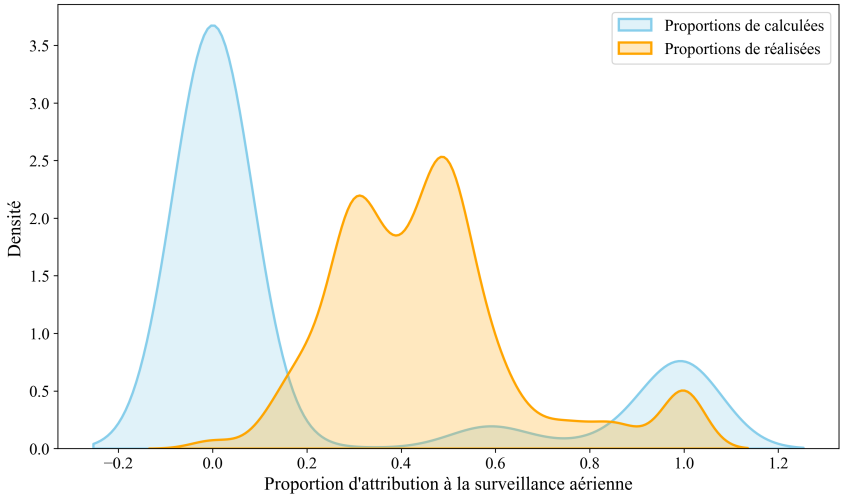
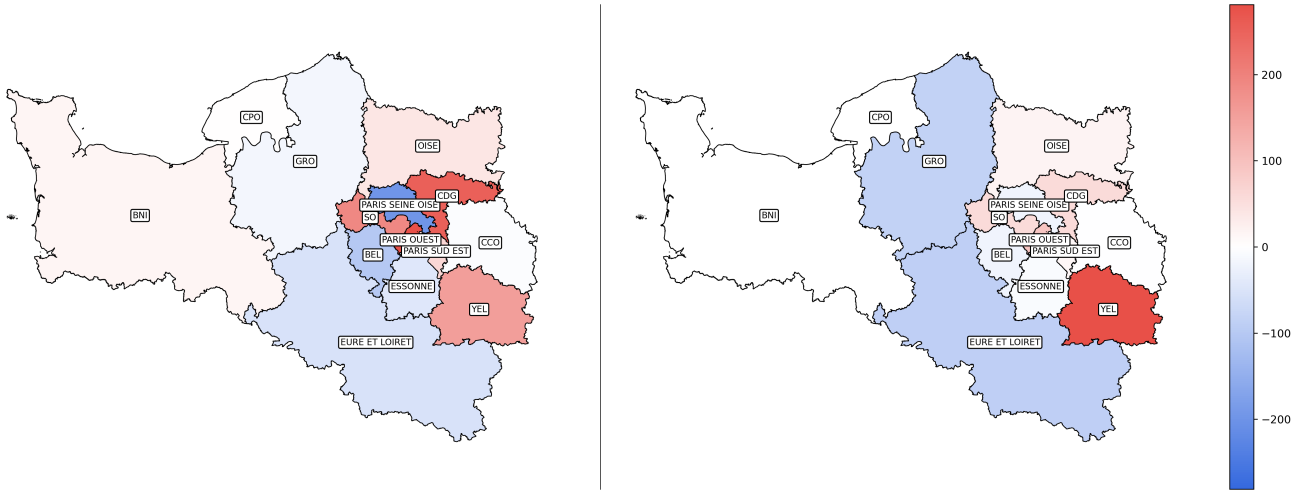


FIGURE 8.13 – Distribution de la proportion de fois où une portion est attribuée aux tournées aériennes

La Figure 8.14 présente une analyse sectorielle des écarts entre les temps de surveillance réalisés et ceux calculés par notre méthode. Contrairement aux résultats obtenus pour l'année 2023, on observe en 2025 un équilibre global entre les hausses et les baisses des heures de surveillance terrestre allouées par notre approche par rapport au réalisé. En termes de volumes horaires (Figure 8.14a), l'équipe d'exploitation PARIS OUEST enregistre une augmentation de 175,27 heures de surveillance terrestre, tandis que l'équipe PARIS SEINE OISE connaît une diminution de 123,53 heures. Exprimés en pourcentage (Figure 8.14b), les écarts apparaissent plus contrastés : l'équipe d'exploitation de YEL

voit son temps de surveillance terrestre multiplié par 2,81, alors que l'équipe EURE ET LOIRET enregistre une baisse de 89,57 % du temps alloué à ce mode de surveillance.



(a) Écart de durées de surveillance (en h)

(b) Écarts de pourcentages de durées de surveillance

FIGURE 8.14 – Analyse des écarts de surveillance entre les tournées réalisées et les tournées calculées pour 2025

Les résultats obtenus en 2025 mettent en évidence la bonne capacité de généralisation de la méthode que nous proposons. En effet, bien que la base d'apprentissage initiale repose sur les données du découpage des portions et secteurs de 2023, la méthode présente de bonnes performances, puis continue de s'améliorer au fil des itérations successives. Ces observations suggèrent qu'il n'est pas nécessaire de reconstruire systématiquement la base d'apprentissage à chaque nouvelle campagne, la méthode étant capable de s'adapter efficacement à de nouvelles données.

8.3 Conclusion

Dans ce chapitre nous avons présenté l'application de l'ensemble de l'approche itérative sur deux ensemble de données. Le premier concerne la surveillance pour l'année 2023 alors que le second concerne l'année 2025.

L'approche itérative présente un intérêt majeur, dans la mesure où les résultats s'améliorent au fil des itérations : nous avons ainsi observé une réduction de plus de 10 % du coût total des tournées entre la première et la cinquième itération en 2023 et de plus de 45 % en 2025. Cette différence pourrait s'expliquer par le fait que la base d'apprentissage initiale repose sur des données anciennes par rapport à l'année d'évaluation 2025, offrant ainsi une marge d'amélioration plus importante au fil des itérations. Par ailleurs, les modèles prédictifs gagnent en précision à chaque phase de réapprentissage.

La méthode actuellement utilisée par l'entreprise, qui repose sur le regroupement des portions de surveillance en ROS, entraîne une sur-surveillance importante, avec des fréquences par portion largement supérieures à celles préconisées avec les enjeux. Notre approche, en revanche, introduit davantage de flexibilité dans la planification des tournées et dans le choix des modes de surveillance lorsque cela est possible, permettant ainsi de réduire le temps total de surveillance de plus de 30 % tant en 2023 qu'en 2025.

La part de surveillance routière réalisée au premier semestre 2025 par rapport à la surveillance totale a nettement diminué par rapport à 2023, notamment en raison du recours à la surveillance satellitaire. En 2023, notre approche attribuait une part beaucoup plus importante de portions à la surveillance aérienne, ce qui se traduisait par un recours à l'aérien sensiblement supérieur au réalisé. En revanche, en 2025, cette répartition tend à s'équilibrer, les portions étant désormais attribuées de manière plus homogène entre surveillance aérienne et routière.

Chapitre 9

Conclusion et perspectives

9.1 Conclusion

Dans cette thèse, nous avons introduit un nouveau problème de tournées sur les arcs, inspiré des problématiques de surveillance des réseaux de transport de gaz, et plus particulièrement de celles rencontrées par l'entreprise NaTran.

Dans le Chapitre 1, nous avons présenté le contexte de l'entreprise ainsi que des activités de surveillance du réseau. L'entreprise exploite en France plus de 32 000 kilomètres de canalisations, qui doivent faire l'objet d'une surveillance quotidienne afin de garantir la sécurité et la performance du réseau. Cette surveillance vise à détecter tout élément susceptible de porter atteinte à l'intégrité du réseau, notamment la présence d'engins ou de traces de chantiers, ainsi que des modifications de l'environnement telles que des changements de balisage ou des mouvements de terrain. Pour réaliser ces opérations de surveillance, différents moyens sont mobilisés. La surveillance peut être aérienne, confiée à des prestataires spécialisés, ou terrestre, assurée en interne au sein de chaque secteur d'exploitation. Par ailleurs, selon le niveau de risque associé à chaque portion de canalisation, des fréquences de surveillance différenciées sont définies. La méthode actuellement utilisée par l'entreprise, qui repose sur des regroupements manuels de portions à surveiller, est à la fois chronophage et peu flexible. Elle conduit à des tournées rigides avec souvent un mode de surveillance unique et génère également de la sur-surveillance, dûe à la fréquence maximale retenue parmi les portions composant chaque regroupement. Il existe ainsi une volonté de réduire non seulement les déplacements, mais plus globalement les coûts liés à la surveillance du réseau, qui représentent une part importante des dépenses de l'entreprise et ont également un impact significatif sur le bilan carbone.

Dans le Chapitre 2, nous avons introduit les notations et présenté le nouveau problème sous la forme d'un PLNE : un problème périodique de tournées sur les arcs avec une flotte multimodale. Ce modèle exact permet de représenter de manière rigoureuse l'ensemble des contraintes et des objectifs liés à la planification des tournées. Cependant, en raison de la complexité intrinsèque du problème et de la taille des instances considérées, sa résolution directe s'est révélée impraticable, rendant nécessaire le recours à des méthodes approchées et heuristiques pour obtenir des solutions réalisables en temps raisonnable.

Dans le Chapitre 3, nous avons présenté un état de l'art des problèmes et des méthodes

de résolution des tournées sur les arcs, exactes ou approchées, ainsi que des travaux portant sur l'utilisation de l'apprentissage automatique dans les problèmes de tournées. Le problème étudié dans cette thèse est nouveau et, à notre connaissance, n'a jamais été abordé dans la littérature. Par ailleurs, compte tenu de sa complexité, le recours à des modèles d'apprentissage automatique pour assister les modèles d'optimisation, ainsi que la décomposition de la résolution par niveaux de décision, apparaissent particulièrement pertinents.

Ainsi, en cohérence avec la formulation du problème et l'état de l'art, nous avons proposé dans le Chapitre 4 une approche globale itérative en deux phases afin de résoudre ce problème à deux niveaux de décision. Un premier niveau dit *Module de Planification* s'occupe de la planification des portions de surveillance ainsi que de l'attribution d'un mode de surveillance. Le second est le *Module de Routage* permettant de résoudre les tournées de surveillance. Un modèle d'apprentissage automatique assiste la phase de planification afin de prédire le coût des tournées en fonction des arêtes requises qui les composent. Nous avons formulé pour chaque module différents modèles de PLNE.

En pratique, les modèles exacts sous forme de PLNE, utilisés à chaque niveau de décision, peuvent rapidement devenir impraticables dès que la taille ou la complexité des problèmes augmente, c'est-à-dire lorsqu'il s'agit de résoudre des instances de tournées comportant plusieurs véhicules ou des centaines de portions à surveiller, ou impliquant plusieurs modes de surveillance simultanément pour la phase de planification.

Ainsi, dans le Chapitre 5, nous nous plaçons au niveau du *Module de Routage*, nous présentons différentes stratégies pour réduire la taille des instances de tournées à résoudre, en tirant parti de la structure du maillage du transport de gaz, ainsi que diverses méthodes approchées basées sur des heuristiques pour résoudre les tournées. Il est également prévu de résoudre d'abord une tournée dite « géante », puis de la décomposer en plusieurs sous-tournées à l'aide de la méthode *Split*, ou par insertion de stations intermédiaires dans le cas des tournées aériennes effectuées par le prestataire RECTIMO Air Transports. Les tournées « géantes » peuvent être résolues soit à l'aide du modèle exact sous forme de PLNE, soit à l'aide d'heuristiques (k-BIH ou k-NNH), en fonction du nombre de portions à couvrir.

Le Chapitre 6 présente une approche complète pour la conception de modèles prédictifs visant à estimer la durée d'une tournée en fonction du mode de surveillance et des portions à couvrir. Pour ce faire, un ensemble d'observations a été généré, correspondant à des instances de tournées résolues et associées à leurs durées. Des *features* ont été conçues pour décrire ces instances, en s'appuyant à la fois sur des caractéristiques proposées dans la littérature et sur des descripteurs spécifiques au problème étudié. Différentes stratégies de génération d'instances ont été mises en œuvre afin de constituer des bases d'apprentissage suffisamment complètes, permettant d'entraîner des modèles capables de bien se généraliser. Nous avons également identifié la taille minimale de la base d'apprentissage nécessaire à la stabilisation des erreurs. Plusieurs modèles de régression issus de l'apprentissage automatique, ainsi que différents jeux d'hyper-paramètres, ont été évalués. La régression linéaire a finalement été retenue en raison de ses bonnes performances et de sa simplicité d'intégration dans un modèle de programmation linéaire. Cette approche a été testée à la fois sur des mono-tournées et des multi-tournées, et les *features* les plus contributrices ont été sélectionnées. La phase d'évaluation montre toutefois une moins bonne précision par rapport à celle de la phase de validation, due au fait que la distribution des données évaluées diffère des tournées simulées. Des phases de réapprentissage, permises par l'approche itérative permettent d'améliorer les performances des modèles.

Le Chapitre 7 détaille les modèles du *Module de Planification*. Nous y présentons tout d'abord la formulation des modèles de PLNE, construits à partir de *features* s'exprimant linéairement par rapport aux variables de décision et des coefficients issus des modèles de régression retenus. Cette approche montre toutefois rapidement ses limites lorsque l'on considère des secteurs comportant un grand nombre de portions ou lorsque les deux modes de surveillance sont pris en compte simultanément. Afin de pallier ces limitations, nous avons proposé une heuristique consistant à affecter progressivement les portions à différentes combinaisons de semaines du planning de surveillance, en fonction de leur fréquence requise et du ou des mode(s) de surveillance possibles, tout en s'appuyant sur les prédictions de durée fournies par le modèle de régression sélectionné. Plusieurs stratégies de tri des arêtes requises ont également été étudiées, reposant notamment sur les fréquences de surveillance, la proximité au dépôt, ainsi que sur les modes ou les secteurs concernés.

Enfin, le Chapitre 8 présente les résultats de l'approche itérative en deux phases et les compare aux tournées réalisées en 2023. Des mesures et une réorganisation de l'entreprise ainsi que des équipes ayant été mises en place à partir de 2024, nous avons également évalué notre approche sur les données du premier semestre 2025. L'approche itérative a montré de l'intérêt apportant un grande amélioration dès la première itération. Les tests et les comparaisons avec le réalisé montrent que nous améliorons de plus de près de 30 % les durées et kilométrages de surveillance.

Nous avons ainsi proposé une méthodologie pour résoudre le problème des tournées de surveillance chez NaTran. La complexité et la taille du problème nous ont conduits à développer plusieurs stratégies pour traiter les différentes phases de l'approche itérative, rendant ainsi la méthode modulable et adaptable. Par ailleurs, différentes perspectives et travaux sont en cours ou en réflexion afin de prolonger ce travail, dans le but d'optimiser davantage l'activité de surveillance et, plus globalement, les déplacements au sein de l'entreprise.

9.2 Perspectives

Dans cette section, nous proposons différentes perspectives à donner à nos travaux afin de les fiabiliser ou les développer.

9.2.1 Méthode de résolution

Comme l'heuristique employée est gloutonne, des mouvements locaux peuvent être introduits pour améliorer la solution, soit par intensification (en affinant une solution déjà bonne), soit par exploration (en parcourant de nouvelles zones de l'espace des solutions). Ces mouvements peuvent se produire d'une semaine à l'autre tout en conservant la validité de la solution, ou entre combinaisons de semaines valides. Ils peuvent également intervenir entre différents modes de surveillance pour les portions où une surveillance bimodale est possible. Enfin, ces deux types de mouvements peuvent être combinés pour déplacer une ou plusieurs arêtes requises d'une semaine à l'autre et d'un mode à l'autre.

Lors du réapprentissage des modèles prédictifs, les nouvelles observations peuvent

être très similaires entre elles, d'une itération à l'autre, et se voir en outre attribuer des poids élevés. Afin de préserver la capacité de généralisation des modèles, il est possible d'enrichir les bases d'apprentissage en ajoutant des observations issues de tournées simulées aléatoirement, ce qui permet de limiter le surapprentissage. Il est également envisageable d'introduire de légères perturbations sur les solutions obtenues afin de diversifier davantage les données d'apprentissage. Nous pouvons également envisager une re-sélection des *features* pertinentes à chaque itération.

De plus, il serait pertinent de développer une autre méthode de résolution, par exemple fondée sur une approche matheuristique ou sur des techniques de génération de colonnes, afin de disposer de comparaisons supplémentaires pour évaluer la qualité et la robustesse des résultats obtenus avec notre approche.

9.2.2 Fiabilisation des données

Dans cette section, nous présentons plusieurs pistes permettant d'améliorer la fiabilité des résultats de notre méthode, notamment par la consolidation des données d'entrée.

Fiabilisation de la construction du graphe Certains segments ou portions de canalisations, dits « parallèles », sont très proches les uns des autres et sont implantés parallèlement, par exemple de part et d'autre d'une même rue. Dans la pratique, un passage en voiture ou en avion permet de couvrir simultanément ces deux portions. La méthode de réduction de la taille des instances, basée sur la proximité des nœuds, permet déjà dans certains cas de considérer les deux portions comme une seule, mais ce n'est pas systématique. Une autre approche envisageable pour les tournées terrestres consisterait à projeter les portions de surveillance sur le réseau routier. Lorsque plusieurs portions sont projetées sur un même tronçon routier, elles pourraient être regroupées et représentées par une seule portion, en conservant celle présentant l'enjeu le plus élevé.

Fiabilisation des trajets sur les portions surveillées Il existe un besoin de fiabilisation des données liées aux itinéraires de surveillance des portions de canalisation. Dans le cadre de cette thèse, les extrémités des portions sont considérées comme les

sommets de départ et d'arrivée des trajets pour chaque portion. Or, certaines portions pouvant s'étendre sur plusieurs kilomètres, l'itinéraire reliant uniquement leurs extrémités ne garantit pas nécessairement un suivi tout au long de la canalisation. Afin de s'assurer que les trajets récupérés couvrent, en termes de visibilité, les portions à surveiller, il est nécessaire de prendre en compte l'intersection entre les itinéraires de surveillance et les zones de visibilité automobile. Cela suppose toutefois de disposer de données correctement référencées concernant ces zones de visibilité. Plusieurs pistes peuvent être envisagées. Un prétraitement des données pourrait être réalisé avant l'appel à l'algorithme de calcul d'itinéraires, en projetant, au sein du SIG, les extrémités des portions sur le réseau routier tout en tenant compte des zones de visibilité automobile. Une autre approche consisterait, en particulier pour les portions de grande longueur, à les subdiviser en plusieurs sommets intermédiaires, qui seraient ensuite fournis à l'API comme points de passage obligatoires afin de mieux contraindre l'itinéraire.

Fiabilisation des modes possibles pour chaque portion Dans notre approche, il est nécessaire de définir, pour chaque portion de canalisation, les modes de surveillance possibles, c'est-à-dire les modes permettant d'assurer une visibilité suffisante sur l'ensemble de la portion. Ces modes admissibles sont actuellement identifiés à partir des modes de surveillance effectivement utilisés les années précédentes. Cette méthode conduit toutefois à un nombre limité de portions pour lesquelles une surveillance bimodale est envisageable. En effet, de nombreuses portions sont systématiquement affectées à une surveillance aérienne, sans que l'on sache réellement si d'autres modes de surveillance pourraient être utilisés. Plus généralement, il reste difficile de déterminer avec précision, pour chaque portion, l'ensemble des modes de surveillance possibles. Cette difficulté s'explique notamment par le fait que les zones de visibilité automobile ne sont pas renseignées de manière homogène par l'ensemble des secteurs et nécessitent des mises à jour régulières. Une connaissance de la couverture par mode des différentes portions de surveillance permettrait une plus grande flexibilité à la méthode.

9.2.3 Perspectives industrielles

Différentes perspectives industrielles globales sur l'activité de surveillance sont envisagées, notamment par la mutualisation des opérations et l'intégration de nouveaux modes de surveillance.

Intégration de nouveaux modes de surveillance Le satellite est un mode de surveillance du réseau de transport de gaz relativement récent, principalement utilisé dans les secteurs les plus urbanisés. Il fournit des images à haute résolution, récupérées périodiquement, permettant la détection d'objets tels que des pelleteuses ou des traces de travaux. Cette acquisition et analyse d'images engendre un coût, et le satellite est généralement utilisé en complément de la surveillance terrestre. Cela peut tout à fait être intégré à notre approche actuelle, comme nous l'avons montré dans les analyses relatives à la surveillance en 2025, en imposant, pour les portions concernées, des semaines de passage du satellite.

En complément des satellites, la société NaTran étudie aussi l'utilisation d'autres modes de surveillance fiables et moins coûteux pour l'environnement, tels que les drones ou les ballons dirigeables. Ces solutions impliquent toutefois des contraintes supplémentaires, notamment l'obtention d'autorisations pour le déploiement des drones ou la dépendance aux conditions météorologiques pour les ballons dirigeables.

Mutualisation d'activités Dans un objectif de réduction des déplacements, il est aussi possible de tirer parti d'autres activités réalisées par les exploitant générant déjà des déplacements. C'est notamment le cas des opérations de maintenance sur les installations annexes du réseau, ainsi que des rendez-vous dits Travaux Tiers, qui consistent à se rendre sur le terrain auprès des entreprises ayant déclaré un commencement de travaux afin de leur transmettre les règles à respecter pour creuser en toute sécurité à proximité des canalisations. Il s'agit de chercher à mutualiser ces différentes activités en tirant parti des déplacements déjà effectués pour des opérations de maintenance sur les postes ou pour des rendez vous afin d'assurer la surveillance des portions de réseau situées à proximité.

Des tournées de recherche systématique de fuites sont également menées afin d'identifier,

de localiser et de qualifier toute source d'émission de gaz avant qu'elle ne représente un risque pour les personnes, les biens ou l'environnement. Ces campagnes annuelles mobilisent différents moyens de déplacement et de détection et sont réalisées indépendamment des tournées de surveillance de l'environnement du réseau de transport de gaz.

Enfin une mutualisation est envisagée entre différents gestionnaires de réseaux qui pratiquent aussi des activités de surveillance sur leurs réseaux.

Annexe A

Annexes

A.1 Tables de résultats dans le cas multi-tournées pour la surveillance terrestre

Table A.1 – Indicateurs de performance par nombre de tournées

Nombre de tournées	**MAE**	**RMSE**	**MAPE**
1	23,5	31,0	18,0 %
2	31,2	44,2	7,3 %
3	34,4	43,6	4,7 %
4	38,6	48,5	4,2 %
5	52,9	103,7	4,0 %

Table A.2 – Performance du modèle sur l'ensemble de test pour le cas à plusieurs véhicules

Modèles	**MAE**	**RMSE**	**MAPE**	**Corr.**
Modèle naïf	405,8	484,4	408,5 %	0,00
Modèle de RL avec des *features* de la littérature				
Toutes les *features*	94,7	125,8	30,8 %	0,94
12 *features* sélectionnées	99,4	134,9	30,5 %	0,93
8 *features* linéaires	158,1	211,4	50,2 %	0,90
Modèle de RL				
Toutes les *features*	147,7	537,8	34,1 %	0,94
12 *features* sélectionnées	**40,8**	**56,6**	**14,0 %**	**0,98**
12 *features* linéaires sélectionnées	49,9	63,3	20,1 %	**0,98**

TABLE A.3 – Indicateurs de performance en fonction du nombre de tournées

	12 *features* sélectionnées			**12 *features* linéaires sélectionnées**		
Nombre de tournées	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**
1	28,8	37,8	19,3 %	41,1	51,3	27,7 %
2	32,4	40,2	7,8 %	61,0	75,2	14,3 %
3	45,6	54,5	7,2 %	64,0	79,9	10,1 %
4	51,2	65,2	5,1 %	57,8	69,5	5,9 %
5	46,4	56,1	3,4 %	54,8	72,2	3,9 %

TABLE A.4 – Hyperparamètres évalués pour les modèles pénalisés dans le cas multi-véhicules

Modèle	**Hyper paramètre**	**Valeurs testées**	**Meilleure valeur**
Elastic Net	`alpha`	logspace(-5, 2, 8)	0,001
	`l1_ratio`	{0,1 ; 0,3 ; 0,5 ; 0,7 ; 0,9}	0.7
	`max_iterations`	{1 000 ; 2 000 ; 5 000 ; 10 000}	2 000
Lasso	`alpha`	logspace(-5, 2, 8)	0.01
	`max_iterations`	{1 000, 2 000, 5 000, 10 000}	10 000
Ridge	`alpha`	logspace(-5, 2, 8)	0.1
	`max_iterations`	{1 000 ; 2 000 ; 5 000 ; 10 000}	1 000

TABLE A.5 – Hyperparamètres évalués pour les modèles basés sur des arbres dans le cas de plusieurs véhicules

Modèle	**Hyper-paramètre**	**Valeurs testées**	**Meilleure valeur**
Random Forest	`n estimators`	{50 ; 100 ; 200 ; 300 ; 500}	300
	`max_depth`	{20 ; 30 ; 40 ; 50}	20
	`max_features`	{'auto', 'sqrt', 'log2'}	'sqrt'
XGBoost	`n_estimators`	{100 ; 200 ; 300 ; 500}	500
	`max_depth`	{3 ; 5 ; 7}	3
	`learning rate`	{0,01 ; 0,1 ; 0,2}	0,1
	`subsample`	{0,8 ; 0,9 ; 1}	0,8

TABLE A.6 – Comparaison des performances des modèles de régression pour le cas impliquant plusieurs véhicules

Modèles	***Features* sélectionnées**	**MAE**	**RMSE**	**MAPE**
Régression linéaire	195	36,1	59,9	7,6 %
Elastic Net	174	36,1	59,7	7,4 %
Ridge	175	36,1	59,7	7,5 %
Lasso	167	36	59,7	7,5 %
Forêts aléatoires	195	51,9	82	9 %
XGBoost	195	**34,6**	**59,1**	**6,3 %**

TABLE A.7 – *Features* principales influençant le modèle de régression linéaire dans le scénario multi véhicules

	***Features* sélectionnées**		***Features* linéaires sélectionnées**	
Rang	**ID**	**Nom**	**ID**	**Nom**
1	P17	Somme de la durée des portions	P17	Somme de la durée des portions
2	D25	Somme des durées entre les portions et leur portion la plus proche	G7	Étendue de longitude
3	C1	Somme des distances entre les portions et le centroïde	G5	Étendue de latitude
4	DP28	Durée maximale entre les portions et le dépôt le plus proche	DM2	Nombre de nœuds
5	D26	Durée moyenne entre les portions et leur portion la plus proche	DM1	Nombre de portions
6	G4	Aire du rectangle englobant	DP25	Somme des durées jusqu'au dépôt le plus proche
7	D29	Durée minimale entre les portions et leur portion la plus proche	S42	Degré moyen au sein des nœuds des secteurs
8	CN7	Intervalle interquartile des degrés des nœuds	P1	Somme des longueurs des portions
9	P1	Somme des longueurs des portions	S40	Variance de longitude au sein des nœuds du secteur
10	DP29	Durée minimale jusqu'au dépôt le plus proche	DP12	Durée maximale entre les portions et le(s) dépôt(s)
11	S40	Variance de longitude au sein des nœuds du secteur	DP7	Étendue des distances entre les portions et les dépôts
12	S7	Longueur minimale dans les portions du secteur	S7	Longueur minimale dans les portions du secteur

A.2 Figures et tables pour la surveillance aérienne

TABLE A.8 – Résultats pour la phase de validation pour la prédiction des tournées aériennes

Modèles	**MAE**	**RMSE**	**MAPE**
Modèle de RL mono-tournée aérienne			
12 *features* sélectionnées	**8,25**	**11,48**	**3,94 %**
12 *features* linéaires	18,53	23,57	10,6 %
Modèle de RL multi-tournées aériennes			
12 *features* sélectionnées	31,09	84,96	9,6 %
12 *features* linéaires sélectionnées	39,22	90.69	16,76 %

TABLE A.9 – *Features* principales influençant le modèle de régression linéaire dans le scénario mono tournées aérienne

	***Features* sélectionnées**		***Features* linéaires sélectionnées**	
Rang	**ID**	**Nom**	**ID**	**Nom**
1	G4	Aire du rectangle englobant formé par les portions	G7	Étendue de longitude
2	P1	Somme des longueurs des portions	P17	Somme des durées de portions
3	D25	Somme des durées entre les portions et leur portion la plus proche	G5	Étendue de latitude
4	D1	Somme des distances entre les portions	DM2	Nombre de nœuds
5	D26	Durée moyenne entre les portions et leur portion la plus proche	DM1	Nombre de portions
6	DP4	Distance maximale entre les portions et le dépôt principal	DP4	Distance maximale entre les portions et le dépôt principal
7	D29	Durée minimale entre les portions et leur portion la plus proche	CN4	Degré maximal
8	CN7	Écart interquartile des degrés des nœuds	DP13	Durée minimale entre les portions et le dépôt
9	G1	Périmètre de l'enveloppe convexe formée par les portions	P7	Étendue des longueurs de portions
10	CN9	Nombre de composantes connexes	P1	Somme des longueurs de portions
11	DM2	Nombre de nœuds		
12	G3	Périmètre du rectangle englobant	P23	Étendue des durées de portions

TABLE A.10 – *Features* principales influençant le modèle de régression linéaire dans le scénario multi tournées aériennes

	***Features* sélectionnées**		***Features* linéaires sélectionnées**	
Rang	**ID**	**Nom**	**ID**	**Nom**
1	D25	Somme des durées entre une portion et sa portion la plus éloignée	G7	Étendue de longitude
2	G1	Périmètre de l'enveloppe convexe formée par les portions	DM2	Nombre de nœuds
3	D6	Variance des distances entre portions	G5	Étendue de latitude
4	P1	Somme des longueurs des portions	P17	Somme des durées de portions
5	D17	Somme des distances entre une portion et sa portion la plus proche	DM1	Nombre de portions
6	D36	Distance maximale entre une portion et sa portion la plus éloignée	CN4	Degré maximal des nœuds
7	D26	Durée moyenne entre une portion et sa portion la plus proche	DP12	Durée maximale entre une portion et un dépôt
8	CN2	Degré moyen des nœuds	P7	Étendue de longueurs de portions
9	CN9	Nombre de composantes connexes	DP13	Durée minimale entre les portions et le dépôt principal
10	DM1	Nombre de portions	P5	Longueur minimale de portions
11	G2	Aire de l'enveloppe convexe des portions	DP15	Étendue de durées entre portions et dépôt principal
12	D8	Distance interquartile entre les portions	P21	Durée minimale d'une portion

BIBLIOGRAPHIE

Ahr, Dino et Gerhard Reinelt (2015), « Chapter 8 : The Capacitated Arc Routing Problem : Combinatorial Lower Bounds », in : *Arc routing : Problems, methods, and applications*, SIAM, p. 159-181.

Akkerman, Fabian et Martijn Mes (2022), « Distance approximation to support customer selection in vehicle routing problems », in : *Annals of operations research*, p. 1-29.

Akkerman, Fabian, Martijn Mes et Wouter Heijnen (2020), « Distance approximation for dynamic waste collection planning », in : *Computational Logistics : 11th International Conference, ICCL 2020, Enschede, The Netherlands, September 28–30, 2020, Proceedings 11*, Springer, p. 356-370.

Amaya, Ciro-Alberto, André Langevin et Martin Trépanier (2007), « The capacitated arc routing problem with refill points », in : *Operations Research Letters* 35.*1*, p. 45-53.

Aminu, UF et Richard W Eglese (2006), « A constraint programming approach to the Chinese postman problem with time windows », in : *Computers & Operations Research* 33.*12*, p. 3423-3431.

Armas, Jesica de et al. (2019), « Solving large-scale time capacitated arc routing problems : from real-time heuristics to metaheuristics », in : *Annals of Operations Research* 273.*1*, p. 135-162.

Arnold, Florian et Kenneth Sörensen (2019), « What makes a VRP solution good ? The generation of problem-specific knowledge for heuristics », in : *Computers & Operations Research* 106, p. 280-288.

Asghari, Mohammad et al. (2022), « Transformation and linearization techniques in optimization : A state-of-the-art survey », in : *Mathematics* 10.*2*, p. 283.

Bai, Ruibin et al. (2023), « Analytics and machine learning in vehicle routing research », in : *International Journal of Production Research* 61.*1*, p. 4-30.

Bartolini, Enrico, Jean-François Cordeau et Gilbert Laporte (2013), « An exact algorithm for the capacitated arc routing problem with deadheading demand », in : *Operations Research* 61.*2*, p. 315-327.

Baykasoğlu, Adil et al. (2019), « A review of fleet planning problems in single and multimodal transportation systems », in : *Transportmetrica A : Transport Science* 15.*2*, p. 631-697.

Beardwood, Jillian, John H Halton et John Michael Hammersley (1959), « The shortest path through many points », in : *Mathematical proceedings of the Cambridge philosophical society*, t. 55, 4, Cambridge University Press, p. 299-327.

Belenguer, José Manuel et Enrique Benavent (2003), « A cutting plane algorithm for the capacitated arc routing problem », in : *Computers & Operations Research* 30.*5*, p. 705-728.

Belenguer, José Manuel, Enrique Benavent et Stefan Irnich (2015), « Chapter 9 : The capacitated arc routing problem : Exact algorithms », in : *Arc routing : Problems, methods, and applications*, SIAM, p. 183-221.

Beltrami, Edward J et Lawrence D Bodin (1974), « Networks and vehicle routing for municipal waste collection », in : *Networks* 4.*1*, p. 65-94.

Benavent, Enrique, Angel Corberán, Demetrio Laganà et al. (2019), « The periodic rural postman problem with irregular services on mixed graphs », in : *European Journal of Operational Research* 276.*3*, p. 826-839.

— (2023), « A two-phase hybrid algorithm for the periodic rural postman problem with irregular services on mixed graphs », in : *European Journal of Operational Research* 307.*1*, p. 64-81.

Benavent, Enrique, Angel Corberán et José María Sanchis (2000), « Linear programming based methods for solving arc routing problems », in : *Arc routing : Theory, solutions and applications*, p. 231-275.

Bengio, Yoshua, Andrea Lodi et Antoine Prouvost (2021), « Machine learning for combinatorial optimization : a methodological tour d'horizon », in : *European Journal of Operational Research* 290.*2*, p. 405-421.

Bergman, David et al. (2022), « JANOS : an integrated predictive and prescriptive modeling framework », in : *INFORMS Journal on Computing* 34.*2*, p. 807-816.

Beullens, Patrick et al. (2003), « A guided local search heuristic for the capacitated arc routing problem », in : *European Journal of Operational Research* 147.*3*, p. 629-643.

Bonami, Pierre, Andrea Lodi et Giulia Zarpellon (2018), « Learning a classification of mixed-integer quadratic programming problems », in : *International conference on the integration of constraint programming, artificial intelligence, and operations research*, Springer, p. 595-604.

Bostel, Nathalie et al. (2014), « Approximating the length of Chinese postman tours », in : *4OR* 12.*4*, p. 359-372.

Brownlee, Jason (2016), *XGBoost With python : Gradient boosted trees with XGBoost and scikit-learn*, Machine Learning Mastery.

Campbell, Ann Melissa et Jill Hardin Wilson (2014), « Forty years of periodic vehicle routing », in : *Networks* 63.*1*, p. 2-15.

Cao, Bin et al. (2021), « A memetic algorithm based on two_Arch2 for multi-depot heterogeneous-vehicle capacitated arc routing problem », in : *Swarm and evolutionary computation* 63, p. 100864.

Cavdar, Bahar et Joel Sokol (2015), « A distribution-free TSP tour length estimation model for random graphs », in : *European Journal of Operational Research* 243.*2*, p. 588-598.

Chen, Yuning et Jin-Kao Hao (2018), « Two phased hybrid local search for the periodic capacitated arc routing problem », in : *European Journal of Operational Research* 264.*1*, p. 55-65.

Chien, T William (1992), « Operational estimators for the length of a traveling salesman tour », in : *Computers & operations research* 19.*6*, p. 469-478.

Christofides, Nicos (1973), « The optimum traversal of a graph », in : *Omega* 1.*6*, p. 719-732.

Christofides, Nicos et John E Beasley (1984), « The period routing problem », in : *Networks* 14.*2*, p. 237-256.

Christofides, Nicos et Samuel Eilon (1969), « Expected distances in distribution problems », in : *Journal of the Operational Research Society* 20.*4*, p. 437-443.

Chu, Feng, Nacima Labadie et Christian Prins (2004), « The periodic capacitated arc routing problem linear programming model, metaheuristic and lower bounds », in : *Journal of Systems Science and Systems Engineering* 13.*4*, p. 423-435.

— (2005), « Heuristics for the periodic capacitated arc routing problem », in : *Journal of Intelligent Manufacturing* 16.*2*, p. 243.

— (2006), « A scatter search for the periodic capacitated arc routing problem », in : *European Journal of Operational Research* 169.*2*, p. 586-605.

Corberán, Angel, Richard W Eglese et al. (2021), « Arc routing problems : A review of the past, present, and future », in : *Networks* 77.*1*, p. 88-115.

Corberán, Angel et Gilbert Laporte (2015), *Arc routing : problems, methods, and applications*, SIAM.

Corberán, Angel et Christian Prins (2010), « Recent results on arc routing problems : An annotated bibliography », in : *Networks* 56.*1*, p. 50-69.

Daganzo, Carlos F (1984), « The distance traveled to visit N points with a maximum of C stops per vehicle : An analytic model and an application », in : *Transportation science* 18.*4*, p. 331-350.

— (1987), « Modeling distribution problems with time windows : Part I », in : *Transportation Science* 21.*3*, p. 171-179.

Dror, Moshe, Helman Stern et Pierre Trudeau (1987), « Postman tour on a graph with precedence relation on arcs », in : *Networks* 17.*3*, p. 283-294.

Edmonds, Jack et Ellis L Johnson (1973), « Matching, Euler tours and the Chinese postman », in : *Mathematical programming* 5.*1*, p. 88-124.

Eglese, Richard W (1994), « Routeing winter gritting vehicles », in : *Discrete applied mathematics* 48.*3*, p. 231-244.

Eglese, Richard W et Leon YO Li (1996), « A tabu search based heuristic for arc routing with a capacity constraint and time deadline », in : *Meta-heuristics : Theory and applications*, Springer, p. 633-649.

Eilon, Samuel et al. (1974), « Distribution management-mathematical modelling and practical analysis », in : *IEEE Transactions on Systems, Man, and Cybernetics 6*, p. 589-589.

Foulds, Les, Humberto Longo et Jean Martins (2015), « A compact transformation of arc routing problems into node routing problems », in : *Annals of Operations Research* 226.*1*, p. 177-200.

Franceschetti, Anna, Ola Jabali et Gilbert Laporte (2017), « Continuous approximation models in freight distribution management », in : *Top* 25, p. 413-433.

Francis, Peter et Karen Smilowitz (2006), « Modeling techniques for periodic vehicle routing problems », in : *Transportation Research Part B : Methodological* 40.*10*, p. 872-884.

Furian, Nikolaus et al. (2021), « A machine learning-based branch and price algorithm for a sampled vehicle routing problem », in : *OR Spectrum* 43, p. 693-732.

Ghiani, Gianpaolo, Francesca Guerriero et al. (2004), « Tabu search heuristics for the arc routing problem with intermediate facilities under capacity and length restrictions », in : *Journal of Mathematical Modelling and Algorithms* 3.*3*, p. 209-223.

Ghiani, Gianpaolo, Gennaro Improta et Gilbert Laporte (2001), « The capacitated arc routing problem with intermediate facilities », in : *Networks : An International Journal* 37.*3*, p. 134-143.

Ghiani, Gianpaolo, Demetrio Laganà et al. (2010), « Ant colony optimization for the arc routing problem with intermediate facilities under capacity and length restrictions », in : *Journal of heuristics* 16.*2*, p. 211-233.

Ghiani, Gianpaolo, Roberto Musmanno et al. (2005), « A heuristic for the periodic rural postman problem », in : *Computers & operations research* 32.*2*, p. 219-228.

Golden, Bruce L, James S DeArmon et Edward K Baker (1983), « Computational experiments with algorithms for a class of routing problems », in : *Computers & Operations Research* 10.*1*, p. 47-59.

Golden, Bruce L et Richard T Wong (1981), « Capacitated arc routing problems », in : *Networks* 11.*3*, p. 305-315.

Goodfellow, Ian et al. (2016), *Deep learning*, t. 1, 2, MIT press Cambridge.

GREISTORFER, Peter (2003), « A tabu scatter search metaheuristic for the arc routing problem », in : *Computers & Industrial Engineering* 44.*2*, p. 249-266.

HASTIE, Trevor et al. (2009), *The elements of statistical learning : data mining, inference, and prediction*, t. 2, Springer.

HERTZ, Alain, Gilbert LAPORTE et Michel MITTAZ (2000), « A tabu search heuristic for the capacitated arc routing problem », in : *Operations research* 48.*1*, p. 129-135.

HERTZ, Alain et Michel MITTAZ (2001), « A variable neighborhood descent algorithm for the undirected capacitated arc routing problem », in : *Transportation science* 35.*4*, p. 425-434.

KANG, Myung-Ju et Chi-Geun HAN (1998), « Solving the rural postman problem using a genetic algorithm with a graph transformation », in : *Proceedings of the 1998 ACM symposium on Applied Computing*, p. 356-360.

KANSO, Bilal (2020), « Hybrid ant colony algorithm for the multi-depot periodic open capacitated arc routing problem », in : *Int. J. Artif. Intell. Appl* 11.*1*, p. 53-64.

KANSOU, Ali et Bilal KANSO (2022), « Hybrid metaheuristics for the periodic open arc routing problem », in : *International Journal of Metaheuristics* 8.*1*, p. 27-50.

KANSOU, Ali et Adnan YASSINE (2009), « Ant colony system for the periodic capacitated arc routing problem », in : *Proc. 2009 International Network Optimization Conference*, p. 1-7.

KEANY, Eoghan (2020), « BorutaShap : A wrapper feature selection method which combines the Boruta feature selection algorithm with Shapley values », in : *Zenodo*, p. 1.

KESKIN, Muhammed Emre et Chefi TRIKI (2022), « On the periodic hierarchical Chinese postman problem », in : *Soft Computing* 26.*2*, p. 709-724.

KHALIL, Elias et al. (2017), « Learning combinatorial optimization algorithms over graphs », in : *Advances in neural information processing systems* 30.

KIRLIK, Gokhan et Aydin SIPAHIOGLU (2012), « Capacitated arc routing problem with deadheading demands », in : *Computers & Operations Research* 39.*10*, p. 2380-2394.

KOOL, Wouter, Herke VAN HOOF et Max WELLING (2018), « Attention, learn to solve routing problems ! », in : *arXiv preprint arXiv :1803.08475*.

Kou, Shuhan, Bruce L Golden et Luca Bertazzi (2024), « An improved model for estimating optimal VRP solution values », in : *Optimization Letters* 18.*3*, p. 697-703.

Kou, Shuhan, Bruce L Golden et Stefan Poikonen (2022), « Optimal TSP tour length estimation using standard deviation as a predictor », in : *Computers & Operations Research* 148, p. 105993.

Kruber, Markus, Marco E Lübbecke et Axel Parmentier (2017), « Learning when to use a decomposition », in : *International conference on AI and OR techniques in constraint programming for combinatorial optimization problems*, Springer, p. 202-210.

Kwon, Ohseok, Bruce L Golden et Edward Wasil (1995), « Estimating the length of the optimal TSP tour : An empirical study using regression and neural networks », in : *Computers & operations research* 22.*10*, p. 1039-1046.

Lacomme, Philippe, Christian Prins et Wahiba Ramdane-Cherif (2004), « Competitive memetic algorithms for arc routing problems », in : *Annals of Operations Research* 131.*1*, p. 159-185.

Lacomme, Philippe, Christian Prins et Wahiba Ramdane-Chérif (2001), « A genetic algorithm for the capacitated arc routing problem and its extensions », in : *Workshops on applications of evolutionary computation*, Springer, p. 473-483.

— (2005), « Evolutionary algorithms for periodic arc routing problems », in : *European Journal of Operational Research* 165.*2*, p. 535-553.

Laganà, Demetrio et Paolo Paronuzzi (2025), « A multi-start local search matheuristic for the capacitated arc routing problem with irregular services », in : *European Journal of Operational Research*.

Lana, Ibai et al. (2018), « Road traffic forecasting : Recent advances and new challenges », in : *IEEE Intelligent Transportation Systems Magazine* 10.*2*, p. 93-109.

Lannez, Sébastien (2010), « Optimisation des tournées d'inspection des voies », thèse de doct., INSA de Toulouse.

Lannez, Sébastien et al. (2015), « A railroad maintenance problem solved with a cut and column generation matheuristic », in : *Networks* 66.*1*, p. 40-56.

Laporte, Gilbert (1997), « Modeling and solving several classes of arc routing problems as traveling salesman problems », in : *Computers & operations research* 24.*11*, p. 1057-1061.

Lenstra, Jan Karel et AHG Rinnooy Kan (1976), « On general routing problems », in : *Networks* 6.*3*, p. 273-280.

Lepenioti, Katerina et al. (2020), « Prescriptive analytics : Literature review and research challenges », in : *International Journal of Information Management* 50, p. 57-70.

Lodi, Andrea et Giulia Zarpellon (2017), « On learning and branching : a survey », in : *Top* 25, p. 207-236.

Longo, Humberto, Marcus Poggi De Aragao et Eduardo Uchoa (2006), « Solving capacitated arc routing problems using a transformation to the CVRP », in : *Computers & Operations Research* 33.*6*, p. 1823-1837.

Mahmood, Rafid et al. (2018), « Automated treatment planning in radiation therapy using generative adversarial networks », in : *Machine learning for healthcare conference*, PMLR, p. 484-499.

Marzolf, Fabien, Martin Trépanier et André Langevin (2006), « Road network monitoring : algorithms and a case study », in : *Computers & operations research* 33.*12*, p. 3494-3507.

Mei, Yi, Ke Tang et Xin Yao (2011), « A memetic algorithm for periodic capacitated arc routing problem », in : *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 41.*6*, p. 1654-1667.

Mei-Ko, Kwan (1962), « Graphic programming using odd or even points », in : *Chinese Math* 1, p. 237-277.

Merchán, Daniel et Matthias Winkenbach (2019), « An empirical validation and data-driven extension of continuum approximation approaches for urban route distances », in : *Networks* 73.*4*, p. 418-433.

Minieka, Edward (1979), « The Chinese postman problem for mixed networks », in : *Management science* 25.*7*, p. 643-648.

Mishra, Nishchol et Dr. Sanjay Silakari (2012), « Predictive Analytics : A Survey , Trends , Applications , Oppurtunities & Challenges », in : url : https://api.semanticscholar.org/CorpusID:7966859.

Monroy, IM, Ciro-Alberto Amaya et André Langevin (2013), « The periodic capacitated arc routing problem with irregular services », in : *Discrete Applied Mathematics* 161.*4-5*, p. 691-701.

Mourão, M Cândida et Leonor S Pinto (2017), « An updated annotated bibliography on arc routing problems », in : *Networks* 70.*3*, p. 144-194.

Mourgaya, Magalie et Francois Vanderbeck (2006), « Probleme de tournées de véhicules multipériodiques : Classification et heuristique pour la planification tactique », in : *RAIRO-Operations Research-Recherche Opérationnelle* 40.*2*, p. 169-194.

Nicola, Daniel, Rudolf Vetschera et A Dragomir (2019), « Total distance approximations for routing solutions », in : *Computers & Operations Research* 102, p. 67-74.

Orloff, Clifford S (1974), « A fundamental problem in vehicle routing », in : *Networks* 4.*1*, p. 35-64.

Papadimitriou, Christos H (1976), « On the complexity of edge traversing », in : *Journal of the ACM (JACM)* 23.*3*, p. 544-554.

Pearn, Wen-Lea, Arjang Assad et Bruce L Golden (1987), « Transforming arc routing into node routing problems », in : *Computers & operations research* 14.*4*, p. 285-288.

Pearn, Wen-Lea et KH Wang (2003), « On the maximum benefit Chinese postman problem », in : *Omega* 31.*4*, p. 269-273.

Pihera, Josef et Nysret Musliu (2014), « Application of machine learning to algorithm selection for TSP », in : *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, IEEE, p. 47-54.

Polacek, Michael et al. (2008), « A variable neighborhood search for the capacitated arc routing problem with intermediate facilities », in : *Journal of Heuristics* 14.*5*, p. 405-423.

Prins, Christian (2015), « Chapter 7 : The capacitated arc routing problem : Heuristics », in : *Arc routing : Problems, methods, and applications*, SIAM, p. 131-157.

Prins, Christian, Nacima Labadie et Mohamed Reghioui (2009), « Tour splitting algorithms for vehicle routing problems », in : *International Journal of Production Research* 47.*2*, p. 507-535.

Prins, Christian et Roberto Wolfler Calvo (mars 2005), « A fast GRASP with path relinking for the Capacitated Arc Routing Problem », in : *INOC 2005, Lisbonne, March 2005*, Lisbonne, Portugal, URL : `https://utt.hal.science/hal-02895109`.

Rasku, Jussi, Tommi Kärkkäinen et Nysret Musliu (2016), « Feature extractors for describing vehicle routing problem instances », in : *OASICS*, Dagstuhl Publishing.

Riquelme-Rodríguez, Juan-Pablo, Michel Gamache et André Langevin (2014), « Periodic capacitated arc-routing problem with inventory constraints », in : *Journal of the Operational Research Society* 65.*12*, p. 1840-1852.

Riquelme-Rodríguez, Juan-Pablo, André Langevin et Michel Gamache (2014), « Adaptive large neighborhood search for the periodic capacitated arc routing problem with inventory constraints », in : *Networks* 64.*2*, p. 125-139.

Santos, Cleverson G et al. (2017), « A mathematical modeling for periodic capacitated arc routing problem : a multiple tasks proposal », in : *Business Management Dynamics* 7.*5*, p. 27.

Santos, Luís, João Coutinho-Rodrigues et John R Current (2010), « An improved ant colony optimization based algorithm for the capacitated arc routing problem », in : *Transportation Research Part B : Methodological* 44.*2*, p. 246-266.

Sedgwick, Philip (2014), « Spearman's rank correlation coefficient », in : *Bmj* 349.

Shahbazian, Reza et al. (2024), « Integrating machine learning into vehicle routing problem : Methods and applications », in : *IEEE Access* 12, p. 93087-93115.

Sinha Roy, Debdatta et al. (2021), « Estimating the tour length for the close enough traveling salesman problem », in : *Algorithms* 14.*4*, p. 123.

Sun, Yan, Maoxiang Lang et Danzhu Wang (2015), « Optimization models and solution algorithms for freight routing planning problem in the multi-modal transportation networks : A review of the state-of-the-art », in : *The Open Civil Engineering Journal* 9, p. 714-723.

Sutton, Richard S, Andrew G Barto et al. (1998), *Reinforcement learning : An introduction*, t. 1, 1, MIT press Cambridge.

Ulusoy, Gündüz (1985), « The fleet size and mix problem for capacitated arc routing », in : *European Journal of Operational Research* 22.*3*, p. 329-337.

Usberti, Fábio Luiz, Paulo Morelato França et André Luiz Morelato França (2011), « The open capacitated arc routing problem », in : *Computers & Operations Research* 38.*11*, p. 1543-1555.

Webb, MHJ (1968), « Cost functions in the location of depots for multiple-delivery journeys », in : *Journal of the Operational Research Society* 19, p. 311-320.

Willemse, Elias J et Johan W Joubert (2016), « Splitting procedures for the mixed capacitated arc routing problem under time restrictions with intermediate facilities », in : *Operations Research Letters* 44.*5*, p. 569-574.

Win, Zaw (1987), « Contributions to routing problems », thèse de doct., University of Augsburg.

Yu, Wei, Yujie Liao et Yichen Yang (2023), « Exact and approximation algorithms for the multi-depot capacitated arc routing problems », in : *Tsinghua Science and Technology* 28.*5*, p. 916-928.

Zhang, Yanru et Ali Haghani (2015), « A gradient boosting method to improve travel time prediction », in : *Transportation Research Part C : Emerging Technologies* 58, p. 308-324.

Zhang, Yuzhou et al. (2017), « Memetic algorithm with route decomposing for periodic capacitated arc routing problem », in : *Applied Soft Computing* 52, p. 1130-1142.

TABLE DES FIGURES

1.1 Carte du réseau de transport NaTran avec stations de compression et points d'interconnexion (Source : *Rapport NaTran Plan Décennal*) 18
1.2 Schéma de maillage du réseau de canalisations 19
1.3 Exemples d'installations annexes NaTran (Source : *Photothèque NaTran*) . 20
1.4 Répartition des principaux facteurs de risques toutes tailles de brèches confondues sur le réseau NaTran de 1970 à 2016 25
1.5 Illustrations de points singuliers (Source : *Photothèque de NaTran*) 29
1.6 Schéma du processus de construction des ROS (Source : *Document interne à NaTran*) 36
1.7 Évolution des enjeux 37
1.8 Écarts par an entre le prescrit et le prévu ROS (Source : *Analyse NaTran)* 38

2.1 Carte du réseau de transport NaTran 44

3.1 Reproduite de Bengio, Lodi et Prouvost 2021 *European Journal of Operational Research*, sous licence CC BY-NC-ND 4.0. 81
3.2 Reproduite de Bengio, Lodi et Prouvost 2021 *European Journal of Operational Research*, sous licence CC BY-NC-ND 4.0. 82
3.3 Reproduite de Bengio, Lodi et Prouvost 2021 *European Journal of Operational Research*, sous licence CC BY-NC-ND 4.0. 83

4.1 Principe de l'approche par décomposition 97

5.1 Vue partielle du réseau NaTran présentant l'ensemble des données sources recueillies (Source : SIG de NaTran) 112
5.2 Distribution des tailles de PSD en mètres 115
5.3 Carte du secteur de YONNE (82 portions) 116

5.4 Carte du secteur de CARVIN (806 portions) 117
5.5 Carte du secteur de BASSE NORMANDIE (235 portions et 3 dépôts) . . . 117
5.6 Principe de l'algorithme *Split* (de Lacomme, Prins et Ramdane-Cherif 2004 avec l'autorisation de Springer Nature.) 123
5.7 Mono tournée vs multi tournée (155 portions) 124
5.8 Mono tournée vs multi tournée en avion (150 portions) 125
5.9 Temps de résolution en fonction de la taille des instances générées aléatoirement 128
5.10 Temps de résolution en fonction de la taille des instances *clusterisées* . . . 129
5.11 *Box plot* comparatif des temps de résolution entre les instances simulées aléatoirement et celle simulées par proximité 130

6.1 À gauche, 10 PSV sélectionnées dans le secteur de Strasbourg et à droite la tournée qui en résulte 137
6.2 À gauche, des 20 PSV proches sélectionnées dans le secteur de Strasbourg et à droite la tournée qui en résulte 138
6.3 Tournées générées à 20 portions sélectionnées selon différentes stratégies . . 138
6.4 Comparaison entre une tournée mono-véhicule et tournées à 4 véhicules comprenant 95 portions 139
6.5 Matrice de corrélation des variables 154
6.6 Distribution des durées de tournées mono et multi véhicules 157
6.7 MAE et RMSE du modèle de régression linéaire en fonction de la taille de la base d'apprentissage pour un véhicule 160
6.8 MAE et RMSE en fonction du nombre de *features* sélectionnées dans le cas mono-véhicule 164
6.9 MAE et RMSE en fonction du nombre de caractéristiques linéaires sélectionnées pour le cas d'un seul véhicule 165
6.10 MAE et RMSE en fonction de la taille de la base d'apprentissage pour le cas multi-véhicules 171

7.1 Durées des tournées par an et par secteur en fonction de la méthode de tri 180
7.2 Durées des tournées par an et par secteur en fonction de la méthode de tri 194

7.3 Comparaison des durées totales et prédites en fonction des méthodes d'ordre de traitement des arêtes pour les quatre secteurs étudiés. 196
7.4 Comparaison des durées prédites pour le secteur de PONTOISE en triant par arêtes ou par composantes 199
7.5 Comparaison des durées prédites pour AIR LOISIR 200
7.6 Comparaison des coûts selon les ordres de modes empruntés 201

8.1 Évolution des coûts calculés et estimés au cours des itérations pour chaque mode 206
8.2 Évolution des coûts réels et prédits pour la surveillance terrestre 207
8.3 Évolution des coûts réels et prédits pour la surveillance aérienne 208
8.4 Évolution itérative des erreurs MAE et RMSE pour les modèles prédictifs de coûts par mode de transport 209
8.5 Évolution des coûts réels et estimés au cours des itérations pour chaque mode : terrestre (en rouge) et aérien (en bleu) 210
8.6 Portions sélectionnées par semaine pour la surveillance aérienne 211
8.7 Portions sélectionnées par semaine pour la surveillance aérienne 212
8.8 Distribution des rapports entre le nombre de surveillances réalisées et requises en 2023 214
8.9 Analyse des surveillances par enjeu en 2023 : rapports et nombres de passages réalisés comparés aux prescriptions par enjeu de surveillance 215
8.10 Distribution de la proportion de passages attribués aux tournées aériennes par portion bimodale, comparant les valeurs calculées et réalisées en 2023 . 216
8.11 Analyse des écarts de surveillance entre les tournées réalisées et les tournées calculées pour 2023 217
8.12 Évolution des coûts calculés et estimés au cours des itérations pour chaque mode 220
8.13 Distribution de la proportion de fois où une portion est attribuée aux tournées aériennes 221
8.14 Analyse des écarts de surveillance entre les tournées réalisées et les tournées calculées pour 2025 222

COLLEGE ED MATHEMATIQUES
DOCTORAL STIC
PAYS DE LA LOIRE

université angers

Titre : Optimisation de tournées de surveillance d'un réseau de transport de gaz à l'aide de l'apprentissage automatique

Mot clés : problème périodique de tournées sur les arcs, apprentissage automatique, flotte multimodale, méthode approchée itérative, programmation linéaire, heuristiques, apprentissage supervisé

Résumé : NaTran est le gestionnaire du réseau de transport de gaz naturel en France. Afin d'assurer la sécurité de ses infrastructures, l'entreprise doit organiser des tournées de surveillance consistant à parcourir l'ensemble du réseau pour détecter toute anomalie susceptible de compromettre l'intégrité des ouvrages. Ces tournées sont planifiées et construites annuellement. Le réseau est découpé en portions de canalisations, chacune étant associée à un mode de surveillance (véhicule, avion, etc.) ainsi qu'à une fréquence de passage annuelle déterminée par le niveau de risque. Il s'agit alors de construire les tournées de techniciens effectuées chaque semaine sur un horizon d'un an, de façon à minimiser les coûts. Ce problème est un problème périodique de tournées sur les arêtes avec contraintes de capacité et flotte multimodale. Il est NP-difficile. Pour résoudre ce problème, nous proposons une approche itérative en deux phases, assistée par un modèle d'apprentissage automatique. La première phase, dite de planification, vise à affecter les portions du réseau à des semaines de l'année et à des modes de surveillance, tout en respectant les contraintes de fréquence. La seconde phase, dite de routage, consiste à construire les tournées associées à chaque semaine et à chaque mode de surveillance. Un modèle d'apprentissage automatique, chargé d'estimer les coûts des tournées, soutient la phase de planification afin de produire des affectations à coût minimal. Les résultats indiquent que l'approche proposée, comparée au réalisé, réduit le temps total de surveillance de plus de 30 % et optimise la répartition des modes de surveillance.

Title: Machine Learning–Based Optimization of Gas Transportation Networks Surveillance Routes

Keywords: periodic arc routing problem, machine learning, multimodal fleet, iterative heuristic method, linear programming, heuristics, supervised learning

Abstract: NaTran is the operator of the natural gas transmission network in France. In order to ensure the safety of its gas infrastructure, the company must organize surveillance routes that involve traversing the entire network to detect any anomalies that could compromise the integrity of the assets. These routes are planned and constructed on an an-

nual basis. The network is divided into pipeline segments, each associated with a surveillance mode (vehicle, aircraft, etc.) and an annual inspection frequency determined by the level of risk. This problem corresponds to a periodic capacitated arc routing problem with a multimodal fleet, which is NP-hard. To address this problem, we propose a two-phase iterative approach assisted by a machine learning model. The first phase, referred to as the planning phase, aims to assign pipeline portions to weeks of the year and to surveillance modes while satisfying frequency constraints. The second phase, referred to as the routing phase, focuses on constructing the routes for each week of the year and each surveillance mode. A machine learning model estimating route costs supports the planning phase in order to produce cost-efficient assignments. The results show that, compared to the realized solution, the proposed approach reduces the total surveillance time by more than 30 % and improves the allocation of surveillance modes.