

École Doctorale n°072 : Sciences Pour l'Ingénieur

THÈSE

présentée pour obtenir le grade de
Docteur de l'Université de Lille

Spécialité Informatique

soutenue publiquement par

Sélim Cornet

le 29 juin 2020

Formalisation et résolution du problème de construction de grilles horaires robustes pour les réseaux ferrés denses

Composition du jury

Mme Laetitia Jourdan	Professeur, Université de Lille	Présidente
M. Romain Billot	Professeur, Institut Mines-Télécom Atlantique	Rapporteur
M. Xavier Delorme	Professeur, Mines Saint-Étienne	Rapporteur
Mme Sana Berraf	Maître de conférences, ESIEE Paris	Examinatrice
M. Serigne Gueye	Maître de conférences, Université d'Avignon	Examineur
M. Joaquin Rodriguez	Directeur de recherches, Université Gustave Eiffel	Directeur de thèse
Mme Christine Buisson	Directrice de recherches, Université Gustave Eiffel	Directrice de thèse
M. François Ramond	Docteur, SNCF Innovation & Recherche	Encadrant
M. Paul Bouvarel	Ingénieur, SNCF Réseau	Membre invité

Thèse préparée à l'Université Gustave Eiffel
Département COSYS, laboratoire ESTAS
20 rue Élisée Reclus, 59650 VILLENEUVE D'ASCQ

Remerciements

Bien que l'activité du thésard soit essentiellement solitaire, j'ai eu la chance de bénéficier dans cette aventure du concours d'un certain nombre de personnes. C'est elles que je tiens ici à remercier. En premier lieu, bien sûr, l'équipe qui m'a encadré tout au long de ces trois belles années. Christine, ma directrice de thèse, Joaquin, mon directeur, Paul et François, qui ont assuré l'encadrement industriel. Quelles qu'aient été les difficultés rencontrées ou les sujets de discussion à aborder, j'ai toujours trouvé une oreille attentive et des conseils avisés auprès de l'un·e ou l'autre d'entre vous. Vos compétences, vos centres d'intérêt, vos personnalités se complètent à merveille, et je me sens privilégié d'avoir été accompagné par une telle équipe ! Merci pour votre patience, votre disponibilité, votre bienveillance, et vos nombreuses relectures méticuleuses ! Ces trois années avec vous m'ont fait grandir, professionnellement et humainement.

Je remercie vivement Romain Billot et Xavier Delorme pour avoir accepté de rapporter cette thèse. Leurs questions et remarques pertinentes ont permis de préciser certains points qui le méritaient, d'effectuer des corrections nécessaires et globalement d'améliorer la qualité de ce manuscrit. Merci également à Laetitia Jourdan, Sana Berraf et Serigne Gueye d'avoir accepté de siéger au sein du jury.

J'ai eu la chance de passer ces trois années aux côtés de collègues merveilleux, qui m'ont énormément apporté. Merci à ma « grande sœur » de thèse, Lucile, pour ses bons tuyaux, ses commérages bien choisis, et la découverte du Welsh¹ ! Merci également à mon « grand frère » Franck pour les nombreuses activités organisées et qui m'ont permis de découvrir d'autres facettes de cette belle entreprise qu'est la SNCF. Toute ma gratitude à Shidi, pour avoir accepté de s'engager pendant 6 mois sur un des chemins les plus tortueux qu'a ouverts mon travail de recherche, et qui s'en est tirée avec brio ! Merci aux collègues de l'équipe MOD (et son adhérence²), qui m'ont supporté au quotidien malgré ma mauvaise habitude de ramener du camembert de la cantine ! Sabine, Valentina, Juliette, Estelle, Rémy, Mathieu, Xavier, Fabrice, Julie, Tom, Hugo et Hugo, merci pour tout ce que vous m'avez fait découvrir à l'occasion d'un repas ou d'une pause café, pour le temps que vous avez toujours accepté de me consacrer pour m'aider à sortir d'une ornière, pour les (très) nombreuses galettes des rois, les pique-nique et les parties de baby-foot. Une mention spéciale pour Rémi, mon cobureau, grâce à qui j'ai en plus appris que le port de la moustache peut parfois être une vraie question de vie ou de mort !

J'ai également eu d'autres collègues, aux laboratoires LEOST, ESTAS et LICIT, que la distance géographique m'a malheureusement empêché de connaître autant que je l'aurais voulu : Pierre, Julie, Nathalie, Nour-Eddin, Ludovic, Nicolas, Cécile, Delphine, Humberto, Omar, Carlos. À vous tous et toutes, merci pour les bons moments partagés au détour d'une de mes visites, d'une conférence ou d'un séminaire. Plus particulièrement, toute ma reconnaissance à Sonia et Greg pour m'avoir plus d'une fois tiré d'un mauvais pas sur le plan informatique, à Paola pour ses avis sans concessions, et à Anne-Christine pour ses relectures de mes productions dans la langue de Shakespeare.

Qui dit longues études, dit rencontres avec de nombreux professeurs. Certaines d'entre elles marquent à jamais et ont une influence sur toute la suite du parcours. J'ai eu cette chance de suivre des enseignements extraordinaires, et je tiens à remercier celles et ceux qui me les ont fait vivre : Geneviève Gaye-Bounet et à Anne-Laure Biolley, qui m'ont fait aimer et choisir les mathématiques ; Joseph Oesterlé et Stéphane Gaubert, qui m'ont montré que c'est une matière vivante et m'ont donné envie de l'enseigner ; enfin Christophe Chalons, mon tout premier mentor, qui m'a initié aux frustrations mais aussi aux joies que procure l'activité de recherche.

Il me semble avoir lu quelque part³ que « L'amitié, ce n'est pas être là tout le temps, c'est être là quand il le faut ». Or, les moments où il a « fallu » n'ont pas manqué. Je veux donc dire mon immense gratitude tout d'abord à Yuxi, qui a surmonté beaucoup d'épreuves au cours de ces trois ans, et pourtant a toujours gardé de l'énergie, de l'optimisme, de la bonne humeur et des fraises des bois à

1. Antidépresseur à base de bière et de fromage fondu.

2. Au sens topologique du terme.

3. Je crois d'ailleurs que c'était déjà dans la partie « Remerciements » d'une autre thèse. Je promets que c'est la seule tentative de plagiat que contient ce manuscrit.

partager. 厉害厉害 . J'adresse également un « merci » gastronomique et plein d'altruisme à Vianney et Colin. En revanche, je ne remercie pas (enfin, pas trop) Matthieu qui ne m'a pas laissé gagner à Caylus quand j'en avais le plus besoin, ni Adrien et Laetitia pour avoir tenté de me changer les idées à grands coups de problèmes de maths ! Merci à Baptiste pour les balades éclairées à l'acétylène, merci à Lilian pour m'avoir montré un modèle de soutenance parfaite (et aussi que git peut servir à jouer à Pokemon) !

Je tiens aussi à exprimer toute ma reconnaissance à ma famille, qui est juste incroyable. Merci Mélie de m'avoir laissé gagner la « course », je brûle d'impatience à l'idée de te voir bientôt rejoindre le cercle des docteur·e·s Cornet ! Je suis si fier de t'avoir pour sœur ! Merci Maman, merci Papa de m'avoir aimé, aidé, soutenu, et appris tant et plus. C'est avant tout grâce à vous que j'en arrive là aujourd'hui.

Enfin, un grand merci parfum raclette à Alice. Tu as transformé des semaines qui auraient pu, qui auraient dû être les plus difficiles, en des semaines qui comptent parmi les plus extraordinaires de ma vie. Merci du fond du cœur. Chaque moment avec toi est un cadeau précieux.

Résumé

La concentration de l'activité économique autour des grandes villes y entraîne une augmentation régulière de la demande en transport. Afin de répondre à cette demande, les entreprises de transports en commun tentent de proposer une offre adéquate, mais celles-ci sont contraintes par la saturation progressive des infrastructures. Dans le cas du transport ferroviaire, l'accroissement du nombre de voyageurs et de trains en circulation a pour conséquence une augmentation du nombre de perturbations au cours de l'exploitation, ainsi que de leur tendance à se propager et à s'amplifier. Il en résulte une qualité de service dégradée pour les usagers et des pénalités financières pour les opérateurs.

Deux leviers peuvent être actionnés pour atténuer les conséquences de ces perturbations : l'application de mesures de régulation pendant la phase opérationnelle, et la construction en amont de plans de transport robustes face aux petits aléas. C'est principalement sur ce dernier point que porte le travail de la thèse.

Après avoir présenté le fonctionnement de l'exploitation ferroviaire en zone dense et donné une définition d'un petit aléa, nous passons en revue les différents travaux ayant été menés sur le sujet. La grande majorité des cadres conceptuels proposés pour la conception d'horaires robustes ne sont pas adaptés au cas spécifique de la zone dense, c'est pourquoi nous proposons un nouveau modèle sous la forme d'un problème d'optimisation stochastique. Une approche de résolution est ensuite détaillée, en trois étapes. La première porte sur l'estimation des distributions de probabilité des aléas de l'exploitation, à partir de données de retour d'expérience. Dans un second temps, nous utilisons ces distributions dans un outil de simulation stochastique permettant d'évaluer la performance d'une grille horaire donnée. Enfin, cet outil est utilisé comme fonction d'évaluation au sein d'une heuristique de recuit simulé visant à générer automatiquement des grilles horaires robustes.

Abstract

Due to the greater concentration of economic activities around big cities, transportation facilities are facing an increasing demand. In order to answer it, transport companies seek to provide an adequate offer. However, they are constrained by the progressive saturation of transportation networks. Concerning railway transportation, the increasing numbers of passengers and trains result in a higher number of disturbances occurring during operations, as well as a higher tendency of those to amplify and spread over the network. The final consequence is a lower quality of service for passenger and financial penalties to operating companies.

There are two levels of action to prevent these phenomena or limit their reach : taking control actions during operations, and anticipate by building robust transportation plans that will make easier to cope with small random disturbances. The work presented in this thesis focuses mainly on this last point.

After presenting railway operations and its specificities in a Mass Transit context and defining what we call "small disturbances", we provide a review of existing work on similar topics. Most of conceptual frames designed for robust train timetabling are inadequate when it comes to applying them to the specific case of dense traffic areas. That is why we propose a new model given under the form of a stochastic program. We present then a three-step approach for solving it. In the first step, we use data about operations to estimate the probability distributions of disturbances. In the second one, we use these distributions in a stochastic simulation tool that allows us to compute the performance of a given timetable. Finally, this tool is used as a fitness function in a simheuristic, based on simulated annealing algorithm, that aims at computing automatically some robust train timetables.

NB : il est fait usage dans ce manuscrit, en alternance, du masculin et du féminin génériques.

Table des matières

1	L'exploitation ferroviaire en zone dense : une introduction	15
1.1	Notions préalables sur l'exploitation ferroviaire	16
1.1.1	Principaux acteurs du système ferroviaire en France	16
1.1.2	Le plan de transport	17
1.1.3	La signalisation ferroviaire	17
1.1.4	L'exploitation ferroviaire et ses différentes phases	19
1.2	La conception des grilles horaires à SNCF Réseau	23
1.2.1	Graphique Espace-Temps	23
1.2.2	Notions de sillon et de capacité	24
1.2.3	Construction de grilles performantes et robustes	26
1.3	Les spécificités de la zone dense	28
1.3.1	Définitions de la zone dense	28
1.3.2	L'exploitation ferroviaire en Île de France	29
1.3.3	Un fonctionnement entre train et métro	31
1.3.4	Des aléas multiples et une instabilité structurelle	31
1.4	Problématique et principales contributions	35
2	Notion de grille horaire robuste en zone dense	37
2.1	Le <i>Train Timetabling Problem</i>	38
2.1.1	Présentation	38
2.1.2	Description de l'infrastructure	38
2.1.3	Graphe d'événements-activités	39
2.1.4	Un problème de satisfaisabilité : le <i>Feasible Differential System</i>	40
2.1.5	Modèles cycliques et non-cycliques	41
2.2	Concepts de robustesse en optimisation	43
2.2.1	Présentation	43
2.2.2	Robustesse stricte	43
2.2.3	Robustesse légère	44
2.2.4	Robustesse avec recours	45
2.2.5	Optimisation stochastique	46
2.3	Critères de performance d'une grille horaire	47
2.3.1	Indicateurs de performance en mode nominal	47
2.3.2	Indicateurs de robustesse d'une grille horaire	49
2.4	Versions robustes du problème de <i>Train Timetabling</i>	52
2.4.1	Optimisation stochastique	53
2.4.2	Robustesse légère	53
2.4.3	Robustesse avec recours	54
2.4.4	Autres approches	55
2.5	Un modèle pour le TTP robuste en zone dense	55
2.5.1	Limites des indicateurs et modèles existants en zone dense	55
2.5.2	Un indicateur orienté voyageurs	57
2.5.3	Formalisation du problème par un programme stochastique	57

3	Modélisation des petits aléas de l'exploitation	63
3.1	Modélisation des variabilités des durées de stationnement	64
3.1.1	Présentation et motivation	64
3.1.2	État de l'art	65
3.1.3	Collecte des données et pré-traitement	68
3.1.4	Définition et calcul de la durée de stationnement minimale	71
3.1.5	Distributions conditionnelles des durées de stationnement	77
3.1.6	Annexe : résultats sur d'autres jeux de données	81
3.2	Modélisation des variabilités des temps de parcours	82
3.2.1	Présentation et état de l'art	82
3.2.2	Estimation des distributions conditionnelles de temps de parcours	83
4	Évaluation de grilles horaires par simulation stochastique	87
4.1	Motivation	88
4.2	Principes de la simulation stochastique	88
4.3	État de l'art en simulation du trafic ferroviaire	90
4.3.1	Simulateurs microscopiques	90
4.3.2	Simulateurs mésoscopiques et macroscopiques	91
4.4	Description de l'outil SISTRAM	92
4.4.1	Spécifications et choix de modélisation	92
4.4.2	Données d'entrée	93
4.4.3	Description de l'algorithme et du programme	93
4.5	Cas test : le système L2-U	99
4.5.1	Présentation	99
4.5.2	Grilles horaires	100
4.5.3	Constitution des fichiers d'entrée	101
4.5.4	Nombre de répliques	103
4.6	Résultats	103
4.6.1	Conditions expérimentales	103
4.6.2	Validation de la démarche	104
4.6.3	Comparaison entre les grilles 2015 et 2019	104
4.6.4	Allocation de la marge dans la grille horaire	107
4.6.5	Approche « robustesse avec recours »	110
4.7	Conclusion	111
4.8	Annexe	112
5	Une simheuristique pour générer des grilles robustes	115
5.1	Présentation et état de l'art	116
5.1.1	Motivation	116
5.1.2	Méta-heuristiques	116
5.1.3	Approches couplées simulation-optimisation	118
5.2	Une simheuristique de recuit simulé	120
5.2.1	L'algorithme du recuit simulé	120
5.2.2	Adaptation et implémentation	121
5.2.3	Résultats	122
5.3	Méthodes de descente	125
5.4	Conclusion	127
	Conclusion générale	129
	Bibliographie	131

Table des figures

1.1	Risques de rattrapage (à gauche) et de prise en écharpe (au milieu et à droite)	18
1.2	Principe du block à 3 aspects	18
1.3	Principe de fonctionnement du KVB	19
1.4	Les différentes étapes de la conception du plan de transport	20
1.5	Exemple d'itinéraires incompatibles	21
1.6	Exemple de Graphique d'Occupation des Voies	22
1.7	Principales parties prenantes de la phase opérationnelle	22
1.8	Extrait d'un bulletin « Faits marquants »	24
1.9	Exemple de GET	25
1.10	Exemple d'escalier d'occupation	25
1.11	Méthode de saturation	25
1.12	Méthode de compression	26
1.13	Marge sillon	27
1.14	Marge de grille	27
1.15	Utilisation de la marge sillon	28
1.16	Principe du sillon de respiration	28
1.17	Plan du réseau Transilien	30
1.18	Illustration des variabilités sur le GET du train 133406	34
1.19	Propagation des retards sous l'effet de la signalisation	34
2.1	Un modèle microscopique	39
2.2	Un modèle mésoscopique	39
2.3	Un modèle macroscopique	39
2.4	Exemple de Graphe Événements-Activités	40
2.5	Intégrale de la fonction $p_{s,s'}$ si $s \in \mathcal{S}_f$	59
2.6	Intégrale de la fonction $p_{s,s'}$ si $s \in \mathcal{S}_h$	60
2.7	Architecture de l'algorithme de résolution du programme stochastique	62
3.1	Composantes de la durée de stationnement	65
3.2	Plan partiel des dessertes des lignes J et L	68
3.3	Correspondance entre les valeurs de p et celles de A, B, L	71
3.4	Durées de stationnement pour différentes valeurs du flux réduit	72
3.5	Notations des composantes de la durée de stationnement	73
3.6	Cas test fictifs	74
3.7	Valeur de MAE pour différentes combinaisons d'hyperparamètres	75
3.8	Estimation de MDT sur les cas tests fictifs	76
3.9	Erreur d'estimation de MDT sur les cas tests fictifs	76
3.10	Estimation de MDT sur les gares de Bois-Colombes et Houilles-Carières	77
3.11	Classification des observations selon le facteur prépondérant	78
3.12	Fonctions de répartition conditionnelles de DT connaissant P pour les gares de Bois-Colombes et Houilles-Carières	79
3.13	Diagrammes quantile-quantile des distributions conditionnelles de DT sachant P	80
3.14	Fonctions de répartition conditionnelles de DT connaissant P pour les gares de Bécon-les-Bruyères, Colombes et Nanterre Université	82
3.15	Portion de la ligne L située entre les signaux C68 et C294	84
3.16	Fonctions de répartition conditionnelles de RT connaissant RTP	85

4.1	Diagramme de classes partiel de SISTRAM	94
4.2	Algorigramme simplifié d'une réplication	96
4.3	Plan des lignes L2 et U	101
4.4	Largeur de l'intervalle de confiance en fonction du nombre de réplications	103
4.5	Diagrammes quantile-quantile des distributions de temps de parcours observées et simulées	105
4.6	Comparaison des vitesses moyennes avec les grilles horaires de 2015 et 2019	106
4.7	Comparaison de la régularité à 5 minutes avec les grilles de 2015 et 2019	107
4.8	Structure des grilles « 2019 », « Respiration » et « Dense »	108
4.9	Performance sur les différentes réplications des grilles testées	109
4.10	Influence de la distribution de marge sillon sur la qualité de service	109
4.11	Nombre de trains supprimés et effet sur la vitesse moyenne en fonction du seuil	111
4.12	Graphique Espace-Temps nominal de la grille 2019	112
4.13	Graphiques Espace-Temps nominaux des grilles testées	113
5.1	Principe de fonctionnement d'une méta-heuristique de minimisation	117
5.2	Algorigramme de la simheuristique de recuit simulé	121
5.3	Évolution du score des solutions courantes	123
5.4	Trajectoire du recuit simulé pour différentes solutions initiales	124
5.5	Comparaison des algorithmes de recuit simulé et de descente	126

Liste des tableaux

3.1	Quelques caractéristiques des principaux modèles de durée de stationnement de la littérature	67
3.2	Statistiques descriptives des jeux de données décrivant les durées de stationnement .	69
3.3	Matrices de corrélation entre les variables A, B, L	70
3.4	Ratio de variance expliquée par chacune des composantes principales	70
3.5	Ratio de variance expliquée par chacune des composantes principales	81
4.1	Pertinence des lignes du réseau Transilien comme cas test éventuel	100
5.1	Marge sillon des solutions initiales et retournées par la simheuristique	124
5.2	Indicateur SHR en entrée et en sortie de la simheuristique	125
5.3	Contribution du temps d'attente en gare dans les scores des différentes grilles	125

Liste des abréviations

ADC	Agent·e De Conduite
AO	Autorité Organisatrice des transports
ATESS	Acquisition et Traitement des Événements de Sécurité en Statique
CAVE	Comptage Automatique des Voyageurs Embarqué
COGC	Centre Opérationnel de Gestion des Circulations
COT	Centre Opérationnel Transilien
EF	Entreprise Ferroviaire
GEA	Graphe d'Événements-Activités
GET	Graphique Espace-Temps
GI	Gestionnaire d'Infrastructure
GTI	Gestionnaire du plan de Transport et de l'Information
JOB	Jour Ouvrable de Base
KVB	Contrôle de Vitesse par Balises
RER	Réseau Express Régional
RFN	Réseau Ferré National
TTP	<i>Train Timetabling Problem</i>

Synthèse des publications et exposés

Les travaux présentés dans ce manuscrit ont donné lieu aux communications scientifiques suivantes.

Article en revue internationale avec comité de lecture

S. Cornet, C. Buisson, F. Ramond, P. Bouvarel, J. Rodriguez, « Methods for quantitative assessment of passenger flow influence on train dwell times in dense traffic areas », *Transportation Research Part C : Emerging Technologies*, vol. 106, pp. 345-359, 2019.

Article en revue francophone sans comité de lecture

S. Cornet, C. Buisson, F. Ramond, J. Rodriguez, « Comment estimer la durée d'arrêt des trains en gare en zone dense ? », *Revue Générale des Chemins de Fer*, n°302, pp. 12-19, 2020.

Conférence internationale avec comité de lecture

S. Cornet, C. Buisson, F. Ramond, P. Bouvarel, J. Rodriguez, *Assessing train timetable efficiency in a Mass Transit context using a data based simulation method*, accepté à 23rd IEEE International Conference on Intelligent Transportation Systems, 2020.

Exposés en séminaires et conférences

- Séminaire scientifique du laboratoire LICIT de l'Université Gustave Eiffel, 23 janvier 2020, Lyon.
- Séminaire scientifique « Nouveaux modèles pour la gestion des circulations en Mass Transit », 15 octobre 2019, Paris.
- Conférence RailNorrköping, 17-19 juin 2019, Norrköping (Suède).
- Séminaire scientifique du laboratoire ESTAS de l'IFSTTAR, 25 mars 2019, Villeneuve d'Ascq.
- Concours « Ma thèse en 180 secondes », région Hauts de France, 11 mars 2019, Villeneuve d'Ascq.
- Journée doctorale du groupe SNCF, 7 mars 2019, Saint-Denis.
- 5^{èmes} Rencontres Interdisciplinaires Doctorales de l'Architecture et de l'Aménagement Durables, 12 février 2019, Vaulx-en-Velin.
- Journée doctorale du département COSYS de l'IFSTTAR, 3-4 juillet 2018, Marne-la-Vallée.
- Journée doctorale du groupe SNCF, 1^{er} mars 2018, Saint-Denis.

« [...] capable d'être dans l'incertitude, le mystère et le doute, en oubliant l'exaspérante quête de la vérité et de la raison. Voilà l'état d'esprit qui convient. »

Philip Pullman (d'après John Keats), **La Tour des Anges**, 2000.

Chapitre 1

L'exploitation ferroviaire en zone dense : une introduction

Résumé

Ce chapitre vise à présenter le cadre dans lequel s'insèrent les travaux présentés dans cette thèse. Nous commençons par présenter un certain nombre de notions relatives à l'exploitation ferroviaire en France, et qui sont nécessaires à la compréhension du propos qui suit. Parmi ces notions, celle de grille horaire ainsi que la façon dont celle-ci est construite, font l'objet d'une attention particulière. Nous nous intéressons ensuite au cas spécifique de la zone dense. Cela constitue également l'occasion de mettre en lumière les insuffisances du système de transport ferroviaire dans ce cadre, insuffisances auxquelles les travaux présentés dans cette thèse visent à apporter une solution partielle. Ce chapitre se conclut par l'exposé de notre problématique et une esquisse des questions auxquels les chapitres suivants apportent des contributions.

Sommaire

1.1	Notions préalables sur l'exploitation ferroviaire	16
1.1.1	Principaux acteurs du système ferroviaire en France	16
1.1.2	Le plan de transport	17
1.1.3	La signalisation ferroviaire	17
1.1.4	L'exploitation ferroviaire et ses différentes phases	19
1.2	La conception des grilles horaires à SNCF Réseau	23
1.2.1	Graphique Espace-Temps	23
1.2.2	Notions de sillon et de capacité	24
1.2.3	Construction de grilles performantes et robustes	26
1.3	Les spécificités de la zone dense	28
1.3.1	Définitions de la zone dense	28
1.3.2	L'exploitation ferroviaire en Île de France	29
1.3.3	Un fonctionnement entre train et métro	31
1.3.4	Des aléas multiples et une instabilité structurelle	31
1.4	Problématique et principales contributions	35

« Et maintenant, allons-y ! ... Entamons le premier chapitre ... »
Archibald Haddock, dans Hergé, Les aventures de Tintin - On a marché sur la Lune, 1954.

1.1 Notions préalables sur l'exploitation ferroviaire

1.1.1 Principaux acteurs du système ferroviaire en France

L'exploitation des trains de voyageurs en France (hors trajets internationaux) est aujourd'hui, et jusqu'en décembre 2020, confiée à un unique opérateur en situation de monopole. Cependant, de multiples parties prenantes jouent un rôle dans l'organisation et la gestion du transport ferroviaire. En effet, la réforme ferroviaire de 1997 a instauré une séparation entre la gestion des infrastructures et la prestation de services de transport ferroviaire, ceux-ci étant confiés à des entités différentes. Le transport ferroviaire ayant une vocation de service public, son exploitation est pilotée par la collectivité, par l'intermédiaire d'Autorités Organisatrices des transports (ou AO). Enfin, la libéralisation progressive du transport ferroviaire depuis 2004 a conduit à l'arrivée de nouvelles entreprises ferroviaires, opérant à l'heure actuelle des trains de voyageurs internationaux et des trains de marchandises, et demain, des trains de voyageurs nationaux.

1.1.1.1 Le gestionnaire d'infrastructure : SNCF Réseau

Réseau Ferré de France (RFF), établissement créé en 1997 par scission avec la Société Nationale des Chemins de fer Français (SNCF), fut le premier Gestionnaire d'Infrastructure (GI) en France. Il est réorganisé en 2014, réintégré au sein du groupe SNCF et rebaptisé SNCF Réseau. Ses principales missions sont les suivantes :

- La maîtrise d'ouvrage des investissements,
- L'entretien du réseau ferré national (RFN) et des infrastructures associées (postes d'aiguillages, passages à niveaux...),
- L'allocation des plages horaires de circulation et de travaux sur le RFN,
- L'attribution de créneaux horaires (« sillons ») aux entreprises ferroviaires pour leurs circulations, et la tarification de ceux-ci,
- La gestion des circulations en temps réel, notamment lors des situations perturbées,
- L'exploitation des gares par l'intermédiaire de la filiale Gares & Connexions.

1.1.1.2 Le prestataire de services voyageurs : SNCF Voyageurs

Une Entreprise Ferroviaire (EF) est un opérateur ayant pour mission de faire circuler des trains sur le RFN. A l'heure actuelle, le trafic ferroviaire de voyageurs sur le territoire national est assuré par une unique EF en situation de monopole : SNCF Voyageurs. Celle-ci se décompose en 4 branches d'activités principales :

- TGV, chargée de la circulation des trains à grande vitesse,
- Intercités, prenant en charge la circulation des Trains d'Équilibre du Territoire,
- TER, assurant la circulation des trains de desserte régionale ou locale hors Île de France,
- Transilien, en charge des trains de desserte régionale ou locale en Île de France.

Le groupe SNCF est constitué du GI SNCF Réseau, de l'EF SNCF Voyageurs et d'une entité « de tête » SNCF assurant la coordination entre les entités précédentes.

1.1.1.3 Les autorités organisatrices

Les TET ainsi que les trains de desserte régionale et locale (TER et Transilien) sont exploités dans le cadre de conventions passées avec des Autorités Organisatrices (AO). L'Etat français via le Ministère des Transports constitue l'AO des trains TET et les régions celles des trains TER. En Île de

France, les politiques de desserte et tarifaires du transport ferroviaire sont confiées à l'établissement public Île de France Mobilités (ou IdFM), sous tutelle de la Région Île de France. Les conventions fixent aux opérateurs des objectifs en termes de qualité de service (fréquence, ponctualité), des pénalités financières étant appliquées si ces objectifs ne sont pas atteints.

1.1.1.4 Les autres opérateurs

Le transport de voyageurs internationaux est aujourd'hui ouvert à la concurrence, d'autres EF que SNCF Voyageurs ont donc la possibilité de faire circuler des trains de voyageurs sur le réseau. C'est le cas de Thalys et Eurostar, entreprises dont la SNCF est l'actionnaire majoritaire. En 2019, un seul opérateur privé, Thello, exerce le droit à l'accès au réseau, en opérant des trains de nuit de et vers l'Italie. L'ouverture progressive à la concurrence du transport de voyageurs nationaux, à partir de 2020, pourra amener d'autres EF à faire circuler des trains de voyageurs sur le réseau. Le transport de marchandises est également ouvert à la concurrence depuis 2005. En 2011, 32% du trafic marchandises (calculé en tonnes.km) est assuré par d'autres EF que Fret SNCF.

1.1.2 Le plan de transport

Un train peut se définir comme étant la conjonction de trois ressources (Bouvarel, 2011) :

- Un matériel roulant, constitué d'une ou de plusieurs rames automotrices couplées, ou bien d'une rame tractée par une ou plusieurs locomotives.
- Du personnel roulant : un agent de conduite (ou ADC) et, pour certains trains, du personnel de bord.
- Un sillon : créneau horaire où l'utilisation de l'infrastructure est réservée au train. Ce concept de sillon est détaillé dans le paragraphe 1.2.2. De ce sillon sont tirées deux grilles horaires :
 - ◊ La grille horaire commerciale est communiquée à la clientèle. Elle indique les horaires prévues d'arrivée et de départ des différentes gares du parcours, à la minute près.
 - ◊ Une autre grille horaire, plus détaillée, est communiquée aux agents de service (en particulier à l'ADC en charge du train). Celle-ci indique également les horaires prévus de passage à certains points remarquables de l'itinéraire, avec une précision pouvant aller jusqu'à la déca-seconde en Île-de-France. L'ADC a pour consigne de respecter cet horaire du mieux possible, de façon à rester dans son sillon et ne pas empêcher l'utilisation de l'infrastructure réservée pour d'autres trains.

D'autres ressources peuvent être nécessaires à l'exploitation dans les grandes gares : agents d'escale pour l'embarquement et la gestion des départs, personnel chargé du nettoyage et du réapprovisionnement éventuel des rames en produits consommables...

On appelle plan de transport, pour une période de temps donnée, l'ensemble des trains devant circuler sur le réseau pendant cette période. Le plan de transport est défini par l'entreprise ferroviaire, et constitue l'objet principal du contrat avec l'autorité organisatrice dans le cas de services conventionnés. Il décrit l'ensemble des ressources (humaines et matérielles) que l'opérateur prévoit d'engager pour remplir sa mission, qui est de transporter les voyageurs de leur origine vers leur destination.

1.1.3 La signalisation ferroviaire

Le transport ferroviaire possède la particularité d'être réalisé avec des distances de freinage très importantes. Celles-ci sont dues à la masse élevée des véhicules ainsi qu'au très faible coefficient de frottement du contact roue-rail. La détection des risques ne peut donc reposer uniquement sur la vision du conducteur comme c'est largement le cas en transport routier. Un système de signalisation

ad hoc a par conséquent été mis en place. Sur les lignes à double voie, celui-ci vise à prévenir les risques de :

- Rattrapage : lorsqu'un train rattrape et percute le train circulant devant lui sur la même voie.
- Prise en écharpe : lorsque deux trains se présentent simultanément à une convergence ou un cisaillement (intersection entre deux voies).

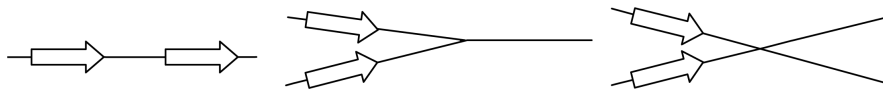


Figure 1.1 – Risques de rattrapage (à gauche) et de prise en écharpe (au milieu et à droite)
Source : Bouvarel (2011)

La circulation sur voie unique présente des risques supplémentaires, dont celui de nez-à-nez, et fait l'objet d'un système de signalisation spécifique. Nous n'entrerons pas dans les détails de celui-ci, nos travaux se concentrant sur le périmètre de la zone dense où la circulation se fait sur doubles voies à sens dédié.

Le système de signalisation repose sur le principe du cantonnement (ou block). La voie est divisée en segments, appelés cantons, chacun d'eux étant protégé à son entrée par un signal ; la règle étant que chaque canton peut contenir, à tout instant, au plus un train. Dans le système de cantonnement à trois aspects, le plus répandu, les signaux peuvent présenter les aspects suivants :

- Sémaphore (un feu rouge). Ce signal signifie que le canton est occupé par un autre véhicule. Le conducteur doit s'arrêter en amont du signal. Il peut ensuite pénétrer dans le canton à condition de le parcourir entièrement en marche à vue (à une vitesse inférieure à 30 km/h). En raison de cette restriction, les ADC préfèrent généralement attendre la réouverture (c'est-à-dire la fin de la présentation de ce signal) pour pénétrer dans le canton.
- Avertissement (un feu jaune). Ce signal prévient l'ADC que le signal d'entrée du canton suivant est un sémaphore ou un carré. Celle-ci doit alors freiner de façon à être en mesure de s'arrêter avant la fin du canton.
- Voie libre (un feu vert). Ce signal signifie que les deux cantons suivants sont libres, et que l'ADC peut poursuivre sa marche sans appliquer de procédure particulière.

Ces trois aspects permettent de prévenir le risque de rattrapage. Un aspect supplémentaire existe pour prévenir les risques de prise en écharpe : le carré fermé (deux feux rouges). Ce signal peut se rencontrer à l'entrée d'un canton contenant une convergence ou un cisaillement, lorsqu'un train venant d'une autre branche s'apprête à y pénétrer. L'arrêt avant ce signal est obligatoire jusqu'à la réouverture. Lorsqu'un signal présente l'aspect carré fermé, le signal du canton précédent présente l'avertissement. Un exemple présentant ces différents aspects de signalisation est donné en figure 1.2.

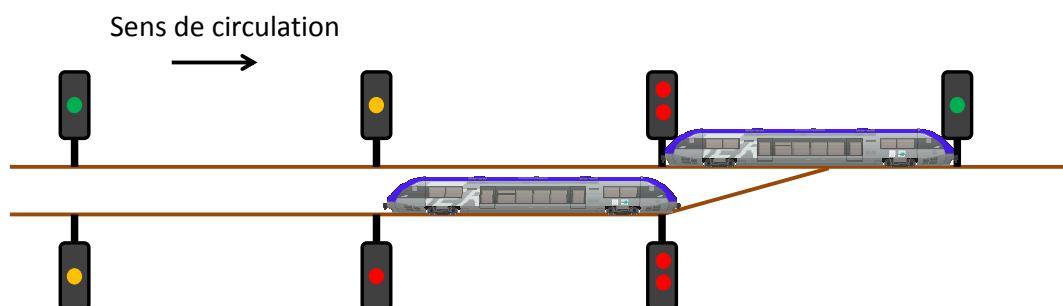


Figure 1.2 – Principe du block à 3 aspects

Les convergences sont gérées selon un principe de réservation d'itinéraire, ou enclenchement. Lorsqu'un train est choisi comme étant le prochain à franchir la convergence, celle-ci lui est « réservée » et les signaux de l'autre branche sont fermés, avant même que le train s'engage sur la convergence. Selon la vitesse maximale sur la ligne et le type de matériel qui y circule, la réservation est déclenchée plus ou moins en amont de la convergence.

La plupart des lignes électrifiées sont en outre équipées d'un système appelé Contrôle de Vitesse par Balises (ou KVB), qui vise à renforcer la sécurité du système en limitant le risque qu'un signal rouge (carré ou sémaphore) soit franchi. Lorsqu'un signal d'avertissement est franchi, une information est transmise à la cabine de conduite par un appareil situé sur la voie. Le KVB calcule alors une courbe de freinage permettant d'atteindre la vitesse de 30 km/h à 200 mètres de l'entrée du prochain canton (voir la figure 1.3). Si l'ADC dépasse cette courbe de freinage (qui ne lui est pas communiquée par le système), alors le freinage d'urgence se déclenche automatiquement jusqu'à l'arrêt du train.

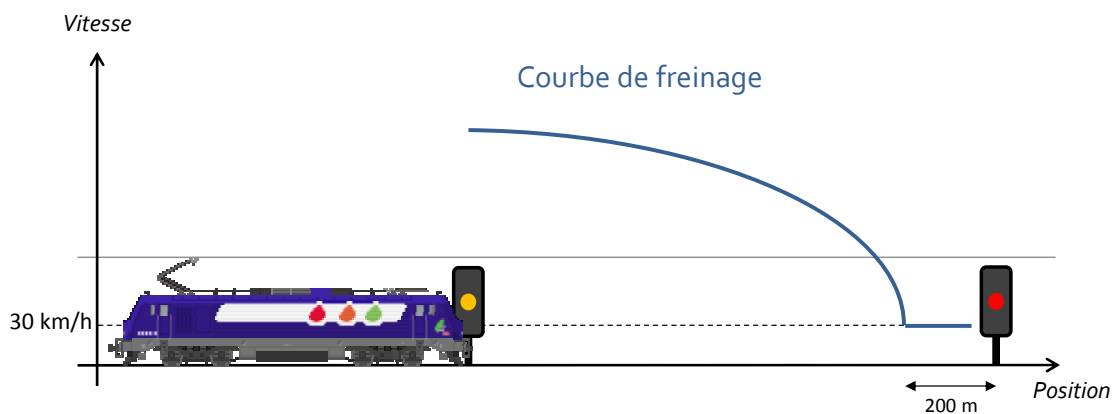


Figure 1.3 – Principe de fonctionnement du KVB

Sur une voie équipée du système KVB « classique », l'ADC doit rester en-dessous de la courbe de freinage jusqu'au franchissement du signal suivant, même si celui-ci est réouvert. D'autres variantes du KVB, plus permissives, autorisent l'ADC à ré-accélérer dès qu'il voit que le signal suivant s'est réouvert (on parle de systèmes à réouverture continue).

1.1.4 L'exploitation ferroviaire et ses différentes phases

1.1.4.1 Différents niveaux de décisions

Le terme d'exploitation désigne l'ensemble des opérations concourant à la mise en œuvre du plan de transport. Celle-ci requiert la prise de décisions pouvant être classées selon plusieurs niveaux, tels que ceux définis par Ansoff (1965). Ceux-ci se distinguent notamment par la durée qui sépare la prise de décision de sa mise en application :

- Niveau stratégique : décisions ayant des effets sur le long terme, prises plusieurs années à l'avance.
- Niveau tactique : décisions ayant des effets à moyen terme, prises jusqu'à un an à l'avance.
- Niveau pré-opérationnel : décisions ayant des effets à court terme.
- Niveau opérationnel : décisions ayant des effets immédiats.
- Niveau post-opérationnel : retour d'expérience sur les décisions précédentes, conduisant à la formulation de recommandations pour les décisions futures.

L'émergence puis les progrès de la recherche opérationnelle et de la simulation au cours des quarante dernières années ont permis le développement d'outils d'aide à la décision, le plus souvent informatiques. Ceux-ci visent à faciliter et rationaliser les choix des décideurs, et peuvent s'appliquer à tous les niveaux décrits précédemment. Dans le cadre de l'exploitation ferroviaire, la prise de décision constitue un processus séquentiel, dont les principales étapes sont les suivantes.

1.1.4.2 La phase de conception du plan de transport

La conception du plan de transport peut à son tour se décomposer en quatre processus menés séquentiellement, et qui peuvent chacun être modélisés sous la forme d'un problème combinatoire. L'ordre dans lequel ces étapes se déroulent est repris sur la figure 1.4, et listé ci-après.

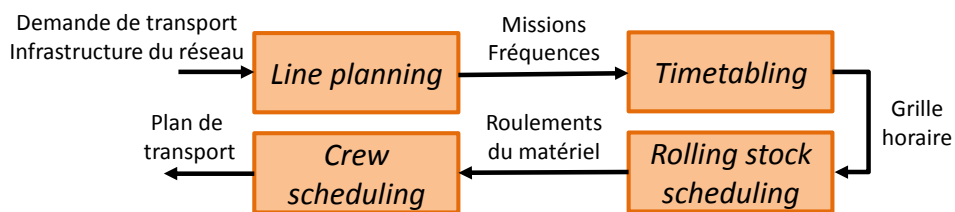


Figure 1.4 – Les différentes étapes de la conception du plan de transport

- Le choix des dessertes et des fréquences (connu en recherche opérationnelle sous le nom de *line-planning problem*). Ce problème consiste en la définition des missions des trains à opérer (origine, destination et gares desservies), ainsi que la fréquence approximative de chacune de ces missions (celle-ci étant généralement différente en heure de pointe et en heure « creuse »). Schöbel (2012) dresse un panorama des principaux modèles existants permettant de résoudre ce problème.
- La construction d'une grille horaire permettant d'opérer les missions définies lors de la phase précédente avec la fréquence prescrite. Cette grille indique, pour chaque train, les heures prévues de départ et d'arrivée dans chaque gare, des heures de passage en certains points remarquables du réseau, et nécessite donc de définir des temps de parcours, de stationnement et de retournement aux terminus. Cette phase de la conception requiert la résolution de ce que l'on appelle le *timetabling problem*. L'essentiel des travaux présentés dans ce manuscrit portent sur cette étape de la conception du plan de transport.
- La planification du matériel roulant (*rolling stock scheduling*). Cette étape consiste en la construction de « lignes de roulement », c'est-à-dire la liste des missions successives qui seront assurées par un même matériel roulant. Deux critères doivent ici être optimisés : on cherche à la fois à minimiser le nombre d'engins engagés pour opérer la grille horaire, tout en garantissant une certaine robustesse. Ce second critère vise à limiter la propagation du retard d'un train sur le train suivant assuré par le même matériel roulant. Plusieurs contraintes doivent également être satisfaites :
 - ◊ Le matériel roulant est également soumis à un planning de maintenance, lui imposant un retour aux ateliers un certain nombre de kilomètres après la précédente opération de maintenance. Il doit en être tenu compte au moment de la construction des roulements.
 - ◊ Certains matériels peuvent ne pas être compatibles avec tous les trains, pour des raisons techniques (par exemple, une automotrice électrique ne peut circuler sur une ligne non électrifiée) ou parce qu'ils n'offrent pas une capacité d'emport voyageurs suffisante.

Différents modèles pour le *rolling stock scheduling problem* sont présentés par Maróti (2006). Il est à noter que, sur les lignes où le parc matériel est en forte tension, cette étape et celle du

line planning sont parfois traitées simultanément ; c'est le cas en France sur certaines lignes TGV.

- La planification des agents (*crew scheduling*). De la même manière, cette étape vise à décider quels seront les trains assurés par les mêmes agents au cours de la même journée. On cherche lors de la construction de ce planning à minimiser le nombre d'agents, tout en maintenant toujours une certaine robustesse. Les principales contraintes de ce problème sont les suivantes :
 - ◊ Tous les agents (de conduite, notamment) ne sont pas habilités aux mêmes lignes, ni aux mêmes engins moteurs, ce qui limite les trains qu'ils peuvent être amenés à assurer.
 - ◊ Certaines réglementations sociales doivent être respectées : le temps de conduite et la durée d'une journée de service font l'objet de restrictions, et des durées de repos minimales doivent être observées. Ces règles sont reprises à la SNCF dans le référentiel RH0077.

On trouvera plusieurs algorithmes permettant de résoudre le *crew scheduling problem* dans Caprara et al. (1997).

1.1.4.3 La phase de préparation de l'opérationnel

Au cours du mois précédant une journée de service et jusqu'à quelques heures avant le début de celle-ci, d'autres décisions sont prises de façon à préparer son bon déroulement. Cette phase vise principalement à prendre en compte de nouvelles contraintes, inconnues lors de la phase de conception. Nous pouvons la décomposer de la façon suivante :

- L'adaptation de la grille horaire. Certains trains sont ajoutés à la grille horaire en dernière minute ; il s'agit principalement des trains de fret, de travaux ainsi que des circulations à vide de rames ou d'engins de et vers les centres de maintenance.
- L'affectation du matériel roulant aux lignes de roulement (*train unit assignment*) : on affecte à chaque ligne de roulement une rame, de façon à ce que le planning de maintenance de chaque élément constituant la rame (locomotives, voitures et/ou engins automoteurs couplés) puisse être respecté.
- L'affectation du personnel (*crew assignment*) : on affecte des agents aux différentes journées de service définies lors du *crew scheduling*, en tenant compte de l'indisponibilité éventuelle de certaines personnes.
- La planification de l'occupation des voies en gare. Ce problème consiste à assigner à chaque train une voie pour chaque gare qu'il franchit (en la desservant ou non) le long de son parcours. La construction de ce planning devient complexe dans les grandes gares, possédant de nombreuses voies à quai, auxquelles on peut accéder par des itinéraires n'étant pas toujours compatibles entre eux (voir par exemple la figure 1.5). En effet, certains itinéraires peuvent nécessiter de cisailer des voies, interdisant la circulation sur celles-ci pendant une plage de temps de l'ordre de quelques minutes.



Figure 1.5 – Exemple d'itinéraires incompatibles

Le planning obtenu est généralement représenté sous la forme d'un Graphique d'Occupation des Voies, représentant la façon dont les voies sont occupées par les différents trains au cours de la journée. Un exemple est donné sur la figure 1.6. Un outil permettant la construction automatique de ce graphique, appelé OpenGOV, est en cours de déploiement à SNCF Réseau.

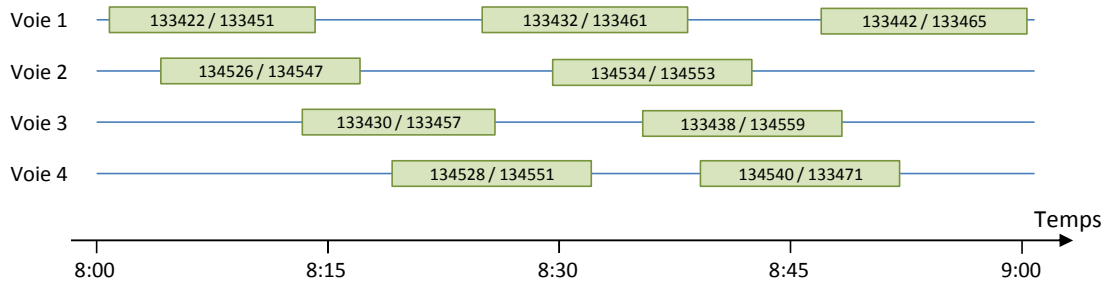


Figure 1.6 – Exemple de Graphique d'Occupation des Voies

1.1.4.4 La phase opérationnelle

À l'issue des phases de conception et de préparation des opérations, le plan de transport peut être mis en œuvre. Sa production requiert la conjonction de différentes ressources (agent-e -s, matériel roulant, disponibilité de l'infrastructure, voyageurs). Différentes entités sont chargées, au cours de la phase opérationnelle, de veiller à la disponibilité de ces ressources et d'intervenir en cas de situation perturbée.

- Les Technicentres ont la responsabilité de fournir le matériel roulant prévu par les roulements.
- La Traction veille à ce que les plannings des ADC soient respectés.
- Les ADC conduisent en respectant la signalisation ferroviaire et en veillant du mieux possible au respect de la grille horaire. Elles signalent également les événements susceptibles de perturber le trafic (personnes sur les voies, signal d'alarme tiré, malaise voyageur...).
- Le centre de gestion des circulations (SNCF Réseau) s'assure de la disponibilité et de la bonne utilisation de l'infrastructure.
- Le centre d'information voyageurs informe les voyageurs en temps réel de la façon dont se déroule l'exécution du plan de transport, via des écrans d'information en gare ou à bord, des annonces sonores, des envois de SMS ou de notifications. En cas de situation perturbée, il est également en charge d'informer les voyageuses sur les causes et la durée prévue de la perturbation, ainsi que sur les éventuels itinéraires alternatifs.
- Le centre opérationnel (SNCF Voyageurs) fait l'interface entre ces différents acteurs. Il se charge également, lorsque des perturbations empêchent la réalisation du plan original, de reconstruire un plan de transport en mobilisant les ressources adéquates.

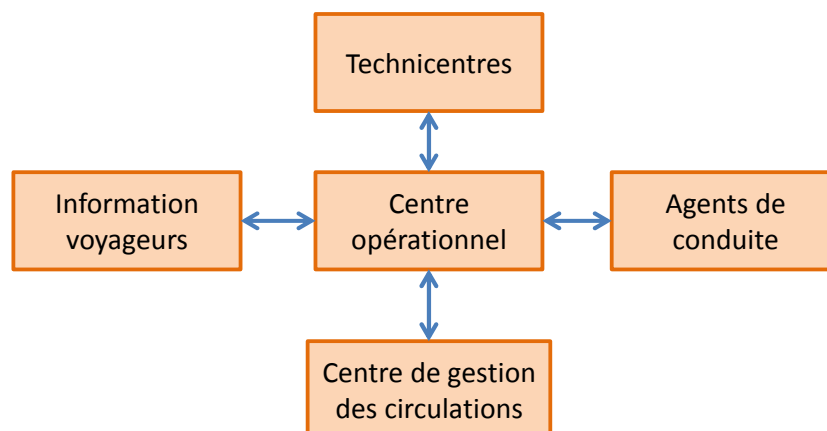


Figure 1.7 – Principales parties prenantes de la phase opérationnelle

1.1.4.5 La phase post-opérationnelle

Au cours de la phase opérationnelle, des données de différentes natures sont collectées sur la manière dont la mise en œuvre du plan de transport a été effectuée.

- Des balises disposées le long des voies identifient les circulations qui les franchissent et enregistrent leur passage dans une base de données appelée BREHAT, gérée par SNCF Réseau. Tous les retards de trains de voyageurs supérieurs à 5 minutes font l'objet d'une documentation spécifique dans cette base, la cause du retard y étant mentionnée ; cela sert notamment au calcul des pénalités imposées par les autorités organisatrices dans le cas de services conventionnés.
- Une autre base de données, baptisée VENDOME, contient des informations relatives à l'exécution du plan de transport du point de vue des voyageurs sur le réseau Transilien. Tous les quatre à cinq ans, sur chaque ligne du réseau, des enquêtes sont réalisées auprès des voyageurs : quel train est emprunté, à quel horaire, pour quel trajet et quel motif. À partir de ces informations et de celles fournies par BREHAT, une projection de l'affectation des voyageurs dans les différents trains est calculée. L'hypothèse principale permettant de réaliser celle-ci est que les voyageurs se présentent en gare 3 minutes avant l'heure prévue de départ de leur train, et embarquent dans le premier train desservant leur gare de destination (qui n'est donc pas nécessairement celui qu'ils avaient prévu d'emprunter initialement, par exemple lorsque les trains circulent en retard). De cette projection est déduite l'indicateur de **ponctualité à 5 minutes** : il s'agit du pourcentage de voyageurs étant arrivées, selon cette projection, à leur gare de destination avec un retard inférieur à 5 minutes. C'est notamment sur la valeur de cet indicateur que s'engagent les opérateurs lors des conventions passées avec les autorités organisatrices en Île de France.
- D'autres bases de données contiennent les informations remontant des cassettes ATESS (Acquisition et Traitement des Événements de Sécurité en Statique). Il s'agit de cassettes magnétiques placées dans les cabines de conduite, qui enregistrent les dates et heures des différents événements liés à la sécurité (départs et arrivées en gare, franchissements de signaux notamment). En Île de France, ces informations sont accessibles via l'application REXMICRO, qui permet notamment de visualiser les valeurs et l'évolution de nombreux indicateurs. L'outil TAIDO, fonctionnant également à partir des remontées ATESS, permet d'extraire des informations sur la conduite des ADC et leur réaction face aux différents événements rencontrés au cours de leur mission.
- Enfin, la base CHATELET, spécifique à l'Île de France, contient les données en provenance des dispositifs de comptage automatique des voyageurs des matériels roulants qui en sont équipés.

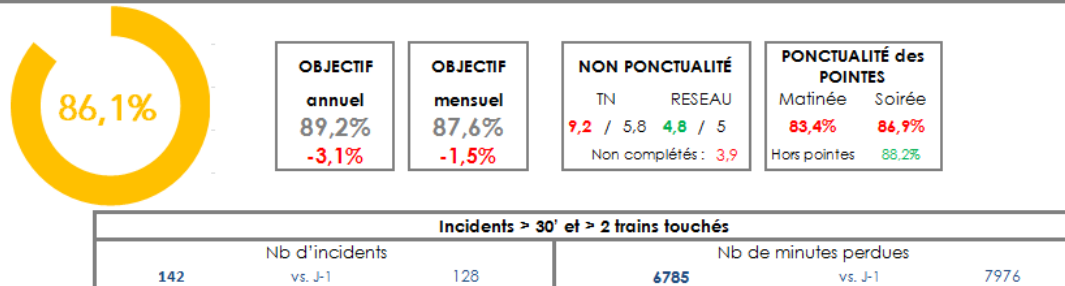
Sur le périmètre Transilien, un bulletin d'information, baptisé « Faits marquants », est édité quotidiennement et diffusé en interne (voir figure 1.8). Celui-ci présente des indicateurs agrégés ainsi que les principaux événements ayant perturbé l'exécution du plan de transport. Toutes ces informations sont ensuite utilisées pour guider des décisions principalement tactiques (conception de plans de transport) et pré-opérationnels (adaptation du plan de transport et définition des normes de gestion du trafic en temps réel).

1.2 La conception des grilles horaires à SNCF Réseau

1.2.1 Graphique Espace-Temps

Le Graphique Espace-Temps (GET) est un outil permettant de visualiser la structure d'une grille horaire ou la trajectoire d'un ou plusieurs trains. Le temps y est situé sur l'axe des abscisses, et la position spatiale le long de la ligne sur l'axe des ordonnées. La trajectoire de chaque train est alors

La ponctualité globale (hors A et B)



La « performance » des RER...

	RER	Ponctu.	Objectif annuel	Objectif mensuel	écart sur annuel	écart sur mensuel
	RER D	87,8%	88,5%	86,6%	-0,7%	1,1%
	RER E	91,1%	91,5%	90,2%	-0,4%	0,9%
	RER A SNCF	86,9%	90,0%	88,3%	-3,1%	-1,5%
	RER B SNCF	85,5%	89,0%	87,4%	-3,5%	-1,9%
	RER C	73,9%	88,5%	86,6%	-14,6%	-12,7%

... des autres lignes TN

	Lignes	Ponctu.	Objectif annuel	Objectif mensuel	écart sur annuel	écart sur mensuel
	Ligne L3	92,2%	89,5%	88,0%	2,7%	4,2%
	Ligne H	95,1%	93,5%	92,5%	1,6%	2,6%
	Ligne R	85,0%	86,0%	83,3%	-1,0%	1,7%
	Ligne K	93,5%	93,0%	91,9%	0,5%	1,6%
	Ligne L2/GCO	91,4%	92,0%	90,9%	-0,6%	0,5%
	Ligne U	91,5%	93,0%	91,8%	-1,5%	-0,3%
	Ligne P	83,8%	88,5%	86,7%	-4,7%	-2,9%
	Ligne N	85,7%	90,5%	89,0%	-4,8%	-3,3%
	Ligne J	84,7%	90,0%	88,5%	-5,3%	-3,8%

Figure 1.8 – Extrait d'un bulletin « Faits marquants »

représentée par un tracé continu donnant l'évolution de la position du train sur la ligne au cours du temps. En particulier, les portions horizontales correspondent aux moments où le train stationne, en gare ou en pleine voie. Un exemple est donné sur la figure 1.9, où le train bleu est un omnibus et le train rouge un train direct. Une variante, appelée escalier d'occupation (voir figure 1.10), permet en plus de visualiser l'occupation de l'espace et l'état des signaux au moment où ils ont été franchis, ce qui peut faciliter l'interprétation du graphique espace-temps. Sur ce type de graphique, un train n'est plus assimilé à un point matériel et l'épaisseur du trait est ici fonction de la longueur du convoi.

1.2.2 Notions de sillon et de capacité

La contrainte imposant que chaque canton ne puisse être occupé qu'au plus par une circulation à tout instant limite le nombre de trains qu'il est possible de faire circuler sur une infrastructure pendant une plage de temps donnée. Ainsi, Hansen et Pachl (2008) définissent la capacité d'une infrastructure comme « le nombre maximum de trains qu'il est possible de faire circuler sur une portion d'infrastructure donnée et pendant une plage de temps donnée, vu comme une valeur théorique limite qui n'est pas atteinte en pratique ».

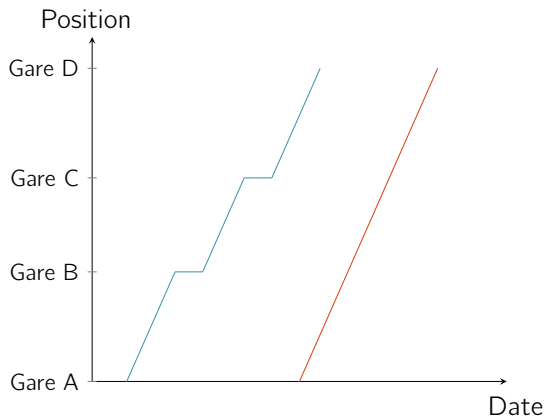


Figure 1.9 – Exemple de GET

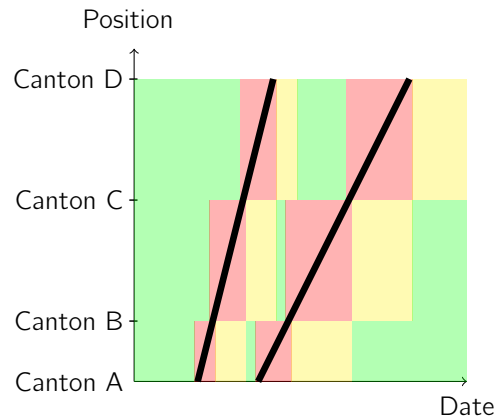


Figure 1.10 – Exemple d'escalier d'occupation

On appelle sillon une partie de la capacité attribuée à un opérateur ferroviaire pour faire circuler un train donné. Celui-ci prend la forme d'une trajectoire dans le Graphique Espace-Temps. L'ensemble des sillons pour une ligne donnée constitue la grille horaire de la ligne. Toute EF souhaitant faire circuler des trains doit ainsi passer commande de sillons au GI, ici SNCF Réseau. Celui-ci attribue alors des sillons aux EF en ayant fait la demande, en cherchant à respecter des contraintes d'équité entre les différentes EF et à limiter la capacité consommée par la grille horaire ainsi produite.

Plusieurs méthodes ont été proposées pour calculer la capacité consommée par une grille horaire. L'Union Internationale des Chemins de Fer (UIC) la définit comme une fonction de 4 paramètres :

- Le nombre de trains
- Leur vitesse moyenne
- L'hétérogénéité du trafic (deux trains ayant des vitesses moyennes v_1 et v_2 différentes consomment davantage de capacité que deux trains ayant la même vitesse moyenne, que celle-ci vaille v_1 ou v_2)
- La stabilité, c'est-à-dire la capacité de la grille à maintenir les trains dans leur sillon. Nous détaillerons cette notion dans la section suivante.

La méthode de saturation (voir par exemple Pellegrini, Marlière et Rodriguez (2017)) consiste à ajouter des sillons à la grille jusqu'à ce que celle-ci soit saturée, et à définir la capacité non consommée comme le nombre de sillons ayant pu être ajoutés. Un exemple illustrant cette méthode est donné sur la figure 1.11.

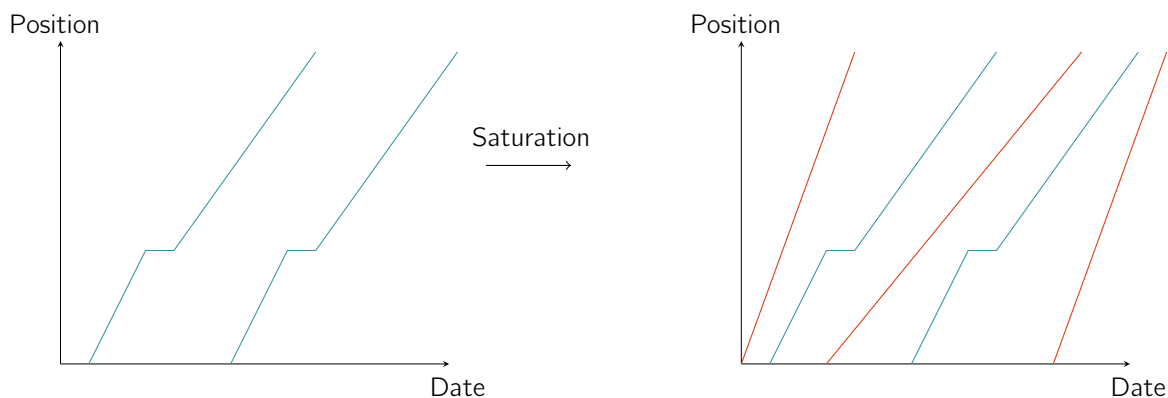


Figure 1.11 – Méthode de saturation

La méthode dite de compression, définie dans le feuillet 406 de l'UIC, est une des plus simples à mettre en œuvre. Celle-ci consiste à rapprocher tous les sillons de la grille au maximum et à mesurer

la capacité comme l'espace restant sur le GET après cette compression. Cette méthode est illustrée sur la figure 1.12.

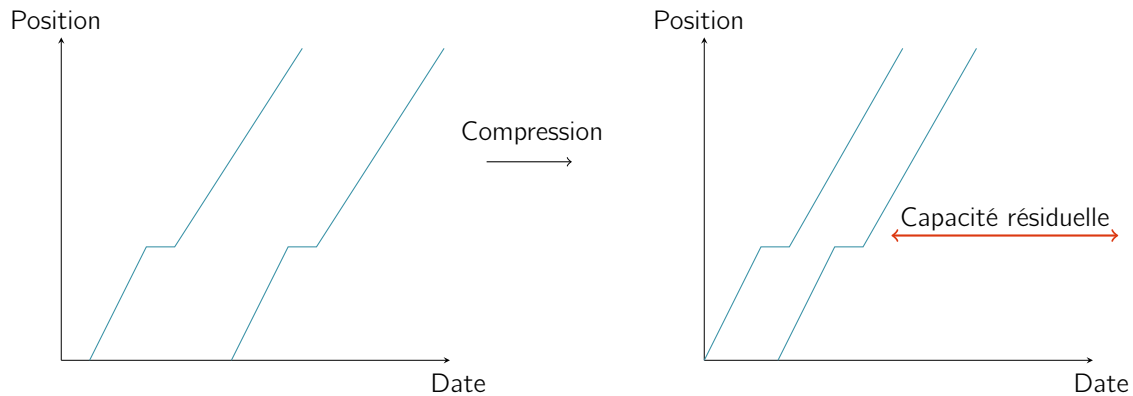


Figure 1.12 – Méthode de compression

La méthode de compression offre l'avantage de fournir un indicateur facile à calculer, et de bien refléter la consommation de capacité qui résulte de la mixité des trafics (trains omnibus et directs en alternance). En revanche, elle constitue un indicateur assez grossier, puisqu'elle ne prend pas en compte la façon dont les sillons sont répartis sur le graphique. La méthode de saturation est plus fine mais plus difficile à mettre en œuvre, ainsi qu'à traduire en un indicateur quantifiable unique.

1.2.3 Construction de grilles performantes et robustes

Une fois les commandes de sillons reçues, le GI conçoit une grille horaire en cherchant à satisfaire au mieux les demandes. Lorsque la ligne est proche de la saturation (c'est-à-dire lorsque le nombre de sillons commandés pour une plage donnée est proche de la capacité maximale), il faut également chercher à utiliser au mieux la capacité de celle-ci. Plusieurs règles doivent pour cela être observées lors de la construction de la grille horaire.

1. **La grille doit être sans conflit.** Cela signifie que la grille horaire ne prévoit pas qu'un canton puisse être occupé simultanément par deux circulations ou plus (ce qui, comme nous l'avons vu précédemment, violerait une règle essentielle de sécurité). Elle ne doit donc pas contenir de conflits entre circulations pour l'utilisation de l'infrastructure.
2. **Les trains doivent circuler à voie libre.** Comme nous l'avons vu au paragraphe 1.1.3, tout franchissement de signal à l'avertissement entraîne un ralentissement, et donc une surconsommation inutile de la capacité ainsi qu'un allongement de temps de trajet pour les voyageuses. Les grilles sont donc tracées à voie libre, c'est-à-dire de telle façon qu'en mode nominal, tous les signaux soient franchis à l'indication « voie libre ». La norme de conception actuelle impose même que le signal soit ouvert un certain temps avant son franchissement par un train, de façon à ce que le conducteur ait la conviction de franchir le signal à voie libre. Cette durée minimale est parfois appelée « marge de voie libre », et notée χ . Sa valeur varie généralement entre 30 et 60 secondes.
3. **La grille horaire doit être robuste.** Lors de la phase opérationnelle, des retards sont susceptibles de se produire, faisant sortir le système du mode nominal. On distingue deux types de retards :
 - Les retards primaires (ou exogènes), dont la cause n'est pas maîtrisable.
 - Les retards secondaires (ou consécutifs), qui sont causés par le retard (primaire ou secondaire) d'un autre train. Les retards sont en effet susceptibles de se répercuter d'un train à l'autre, de plusieurs manières. Les dépassements n'étant généralement pas permis

par l'infrastructure en dehors des gares, tout retard d'un train en ligne peut se propager aux trains derrière lui (notamment *via* la signalisation, la règle de circulation à voie libre n'étant plus satisfaite). En outre, le retard d'un train en gare est susceptible de retarder également les autres trains en correspondance.

Ces retards peuvent donc perturber le fonctionnement du système. Il est par conséquent souhaitable que la grille horaire permette à ces retards de se résorber sans intervention extérieure, au moins dans la limite d'une certaine amplitude. Pour cela, une partie de la capacité est consommée sous forme de **marges**, de façon à améliorer la robustesse de la grille. Ces marges peuvent être utilisées de deux manières différentes :

- Les **marges sillon**, ou **marges de régularité**, consistent à allonger volontairement le temps de parcours le long d'un sillon, par exemple en prévoyant de faire rouler le train à une vitesse inférieure à la vitesse maximale ou en prolongeant la durée d'arrêt prévue en gare. Cela permet à une circulation qui sortirait de son sillon de rattraper son retard plus tard au cours de sa mission, ainsi que d'éviter de propager son retard à d'autres missions. Des illustrations sont données sur les figures 1.13 et 1.15. Cela se fait au prix d'une réduction de la capacité, ainsi que d'un allongement du temps de parcours pour les voyageurs et de la durée d'utilisation des ressources (personnel et matériel roulant).
- Les **marges de grille**, ou **marges d'exploitation**, consistent à espacer les sillons sur le graphique (voir figure 1.14). Cela permet, lorsqu'un retard survient, de limiter la propagation de celui-ci aux autres circulations se trouvant derrière (mais ne permet pas au retard d'être rattrapé). L'utilisation de marge de grille entraîne une consommation de capacité mais pas d'allongement de temps de parcours.

Cette marge peut en particulier être insérée dans la grille sous la forme d'un **sillon de respiration** (voir figure 1.16). Lorsque la grille est constituée de sillons homogènes (même temps de parcours, même desserte) successifs, un sillon de respiration est un sillon laissé vide entre deux batteries de trains. Il permet ainsi d'éviter la propagation d'un retard survenu dans la première batterie de trains de se propager à la seconde batterie.

A l'heure actuelle, en Île de France les grilles horaires sont tracées par les horairistes sous la forme de graphique espace-temps, à l'aide d'un logiciel baptisé AGATHE. Ce logiciel permet la détection de conflits ou de normes non respectées (circulation à voie libre par exemple), mais ne donne pas d'indication quant à la robustesse de la grille produite. Les étapes du tracé sont les suivantes :

- Construction de la **marche de base**, à partir des caractéristiques du matériel roulant et du profil de la ligne. Il s'agit de la trajectoire réalisable la plus rapide.

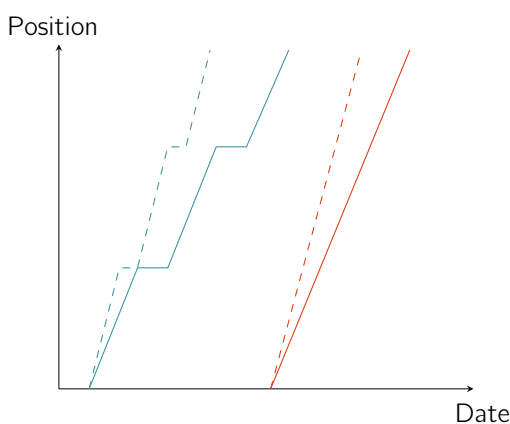


Figure 1.13 – Marge sillon

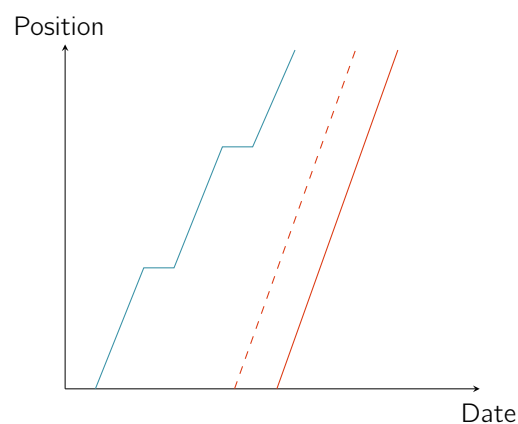


Figure 1.14 – Marge de grille

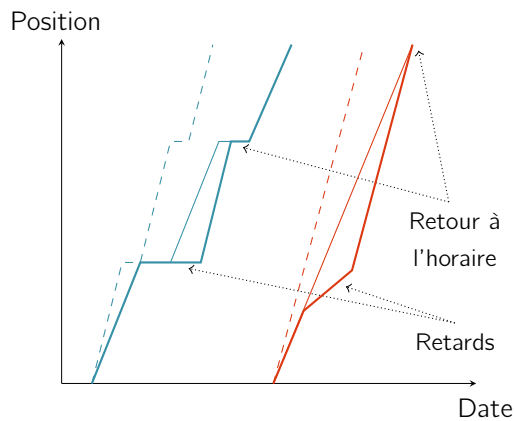


Figure 1.15 – Utilisation de la marge sillon

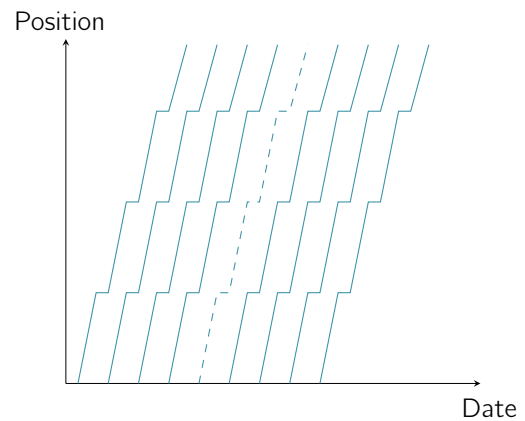


Figure 1.16 – Principe du sillon de respiration

- Ajout de marges sillon, généralement 5% répartis linéairement le long de la marche de base (sauf pour les TGV). La trajectoire ainsi obtenue est appelée **marche-type**.
- Agencement manuel des sillons obtenus dans AGATHE. Le logiciel signale les conflits entre circulations pour l'utilisation de l'infrastructure; ceux-ci sont alors résolus en déplaçant les sillons ou en les allongeant.
- Une estimation de la robustesse de la grille est alors effectuée en mesurant la capacité résiduelle par la méthode de compression. On impose que celle-ci avoisine les 40% (ou jusqu'à 15% lorsque les trains sont tous omnibus). Si cette condition n'est pas satisfaite, la distribution des marges est modifiée jusqu'à obtenir une solution satisfaisante.

En pratique, les grilles horaires sont rarement construites à partir d'une page blanche. Elles sont adaptées d'année en année par des modifications mineures visant à prendre en compte l'évolution de la demande voyageurs, de l'état de l'infrastructure ou des caractéristiques du matériel roulant. Des refontes complètes de l'offre de transport sont néanmoins parfois effectuées; ce fut le cas en 2019 pour la ligne D du RER.

1.3 Les spécificités de la zone dense

1.3.1 Définitions de la zone dense

Dans le cadre du transport ferroviaire, la « zone dense » ne possède pas - encore - de définition précise et faisant consensus. Nous dirons qu'une portion du réseau relève de la zone dense si elle satisfait à au moins une des trois définitions suivantes (souvent vérifiées simultanément) :

- Elle est exploitée à sa capacité maximale, au moins en semaine et aux heures de pointe du matin (6h30 - 9h30) et du soir (17h - 20h).
- La fréquence des dessertes de chaque arrêt est inférieure à 15 minutes. Ce seuil étant choisi comme celui en-deçà duquel la plupart des voyageurs cessent de consulter la grille horaire et se présentent en gare avec l'intention d'embarquer à bord du premier train desservant leur destination (Lüthi, Weidmann et Nash, 2007).
- La densité de population à proximité immédiate des gares dépasse un certain seuil, autour de 5000 habitants par km².

Plusieurs parties du réseau ferré français peuvent ainsi être considérés comme étant situés en zone dense : le nœud ferroviaire lyonnais, le réseau aux alentours immédiats des gares de Lille Flandres et Bordeaux Saint-Jean... Néanmoins, notre propos portera désormais exclusivement sur le périmètre de l'Île de France, où se concentre l'essentiel des défis liés au transport de masse, du fait de la taille

de la portion du réseau situé en zone dense. En effet, l'intégralité du réseau desservant Paris et les départements de la Petite Couronne (Hauts-de-Seine, Seine-Saint-Denis, Val de Marne) satisfait les trois définitions proposées ci-dessus.

1.3.2 L'exploitation ferroviaire en Île de France

Le réseau ferré d'Île de France est exploité par l'opérateur Transilien, dépendant de SNCF Voyageurs. Il est organisé autour de six grandes gares parisiennes, couvrant chacune un périmètre précis : Saint-Lazare (Nord-Ouest), Montparnasse (Ouest et Sud-Ouest), Austerlitz (Sud), la gare de Lyon (Sud-Est), la gare de l'Est (Est) et la gare du Nord (Nord). Chacune constitue l'origine et le terminus des trains de banlieue relevant de son périmètre (à l'exception des lignes de RER).

Le réseau de trains de banlieue (hors tram-train) est ainsi constitué de :

- Cinq lignes de RER (Réseau Express Régional) « traversantes », desservant Paris intra-muros sans y effectuer leur terminus. Il s'agit des lignes A, B, C, D et E ¹.
- Sept lignes de trains de banlieue ayant leur origine et terminus dans une des gares parisiennes. Elles sont identifiées par les lettres H, J, K, L, N, P, R.
- La ligne U, reliant La Verrière à La Défense et ne desservant pas Paris.

Un plan du réseau d'Île de France est donné sur la figure 1.17.

Ce réseau de transport est soumis à une demande extrêmement forte (Altazin, 2018). En effet, s'il ne représente que 10% du RFN (en nombre de kilomètres de voie), les trains Transilien qui y circulent représentent 40% des trains opérés quotidiennement par SNCF Voyageurs. En semaine et hors vacances scolaires, ces trains permettent la réalisation de 8.3 millions de trajets par jour, ce qui représente 70% du nombre de voyages effectués sur le réseau ferré français. Afin de répondre à cette forte demande, le réseau est exploité à sa capacité maximale : jusqu'à 6200 trains circulent chaque jour sur 1300 kilomètres de réseau, et les fréquences peuvent monter en période de pointe jusqu'à 32 trains par heure (dans le tunnel reliant les gares de Châtelet-les-Halles et Paris-Nord, emprunté par les lignes de RER B et D).

Il est à noter que cette forte demande, et la saturation du réseau qui en découle, concernent rarement les lignes dans leur intégralité. La densité de population décroissant à mesure que l'on s'éloigne de Paris, les tronçons de ligne situés dans Paris et en proche banlieue sont les plus sollicités, tandis que la fréquence des trains est réduite au-delà. Ce phénomène est accentué par le fait que de nombreuses lignes se scindent en plusieurs branches en direction de la banlieue ; la fréquence de circulation sur les tronçons centraux est alors bien plus élevée que sur les branches.

Pour des raisons historiques, les lignes de RER A et B sont exploitées conjointement par la RATP et le groupe SNCF. Les autres lignes sont exploitées exclusivement par SNCF Voyageurs via sa branche Transilien. Du fait de sa situation de monopole, qui sera maintenue au moins jusqu'en 2040, Transilien joue un rôle important dans la construction des grilles horaires du réseau d'Île de France, en collaboration avec SNCF Réseau. Il peut néanmoins y avoir concurrence pour l'attribution de sillons entre des trains Transilien et des trains régionaux, grandes lignes ou de marchandises circulant sur la même infrastructure. Ces conflits sont généralement résolus de façon séquentielle (voire itérative) : les trains régionaux et grandes lignes sont placés en premier sur la grille, puis les trains Transilien, et enfin les trains de fret. Il y a une exception à cette règle sur le RER C, où les TGV « intersecteurs » sont tracés après les trains de banlieue. La conception des grilles horaires du réseau de banlieue est ainsi réalisée par le département Offre et Conception de Transilien, et soumise à la validation de SNCF Réseau.

En opérationnel, le plan de transport est géré par les COT (Centres Opérationnels Transilien), situés dans les grandes gares parisiennes. Les GTI (Gestionnaires du plan de Transport et de l'Information) y suivent le trafic en temps réel grâce à l'outil GALITE. Cet outil trace la trajectoire des

1. La ligne E a actuellement son terminus à la gare souterraine de Paris-Saint-Lazare, mais sera prolongée à Mantes-la-Jolie en 2024.

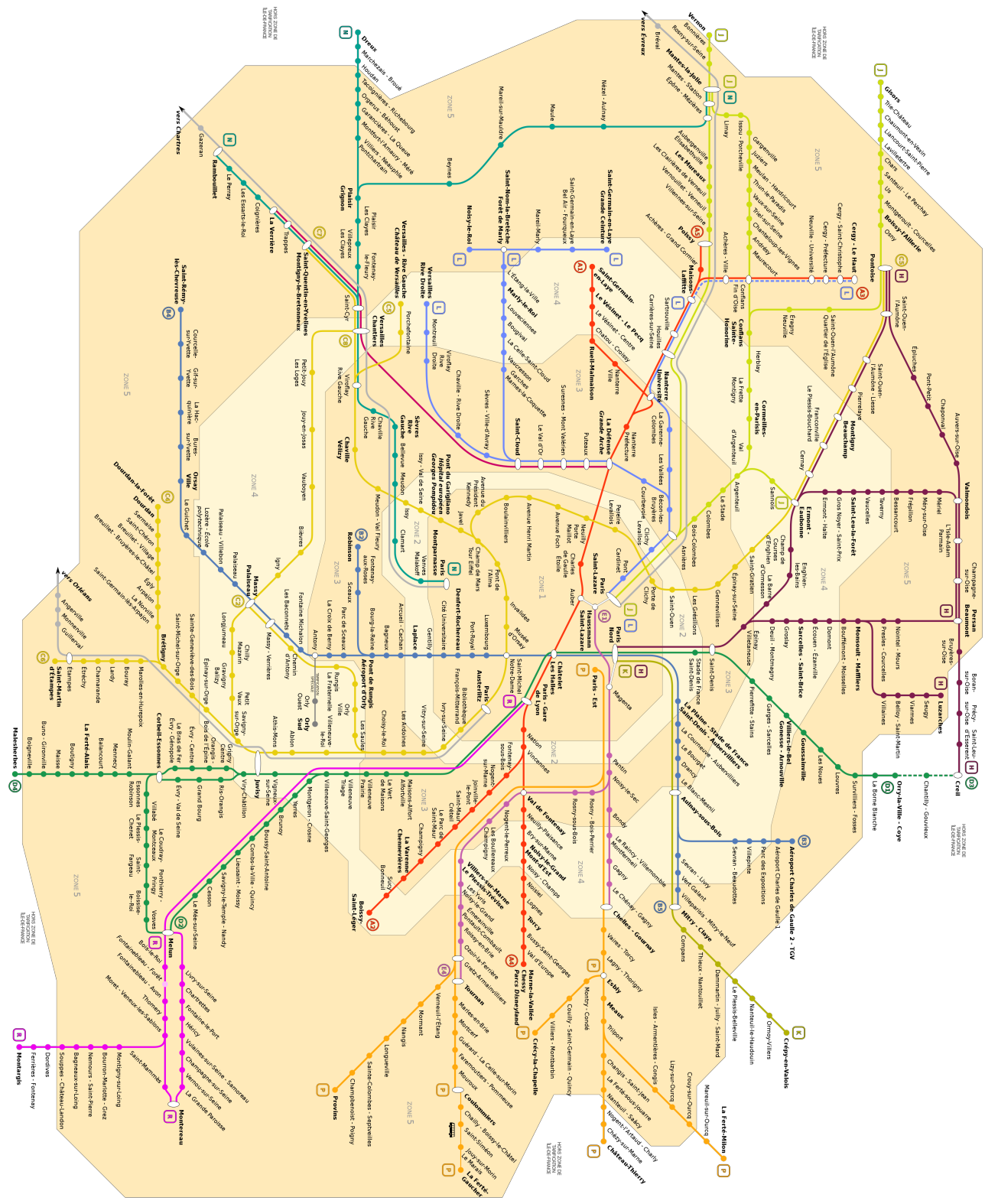


Figure 1.17 – Plan du réseau Transilien

trains sur un graphique espace-temps, et permet de réaliser des projections afin d'anticiper d'éventuelles perturbations. Lorsque celles-ci se produisent, les GTI peuvent décider de modifier le plan de transport. Cela se traduit généralement par des ajouts d'arrêts à certaines missions, des suppressions de trains, et des réorganisations des roulements matériel et agents. L'utilisation de l'infrastructure est quant à elle supervisée dans les COGC (Centres Opérationnels de Gestion des Circulations). Ceux-ci prennent notamment la décision, en cas de perturbation, de réordonner les circulations en changeant leur ordre de passage aux points de convergence, ou en imposant des dépassements dans les gares où cela est possible.

1.3.3 Un fonctionnement « entre train et métro » (Altazin, 2018)

Le système de transport ferroviaire d'Île de France présente, par de nombreux aspects, un mode de fonctionnement proche de celui du métropolitain. En effet, sur les tronçons centraux, les trains circulent sur une infrastructure dédiée et avec des intervalles très rapprochés en heure de pointe (souvent compris entre 3 et 5 minutes). Il n'est donc généralement pas nécessaire pour les passagers de consulter l'horaire avant de se rendre en gare (Lüthi, Weidmann et Nash, 2007), d'autant que les trains ont souvent une desserte omnibus sur ces mêmes tronçons. En revanche, ces considérations ne se vérifient plus en heure creuse ou dès lors qu'on s'éloigne de Paris. En effet, la séparation des lignes en plusieurs branches divise mécaniquement la fréquence des circulations sur chacune de ces branches. Cette fréquence plus réduite ne rend plus nécessaire la circulation sur infrastructure dédiée ; les trains Transilien doivent alors s'insérer entre des trains de voyageurs à plus long parcours ou des trains de marchandises. Une autre conséquence de cette baisse de la fréquence est que les voyageurs souhaitant emprunter un train au départ d'une gare situé sur une branche doivent consulter l'horaire pour décider quand se rendre en gare. Enfin, si quelques lignes sont circulées quasi-exclusivement par des trains omnibus, sur la plupart des lignes Transilien la desserte est organisée selon un des deux paradigmes suivants :

- Exploitation en trains omnibus (ayant généralement leur origine ou leur terminus dans une gare relativement proche de Paris) et en trains semi-directs, à plus long parcours, ne marquant pas l'arrêt dans les gares déjà desservies par les trains omnibus.
- Exploitation en « sauts de puce », où chaque train dessert environ une gare sur deux en alternance, et où quelques gares de correspondance sont desservies par tous les trains.

Toutes ces contraintes rendent impossible l'exploitation du réseau francilien selon une logique de fréquence et d'intervalle, comme c'est le cas sur le réseau métropolitain. Il est par conséquent nécessaire de construire des grilles horaires qui s'accommodent du mieux possible aux fortes contraintes qui pèsent sur le système dans la zone hyperdense entourant Paris.

1.3.4 Des aléas multiples et une instabilité structurelle

Le système ferroviaire en zone dense est soumis en permanence à une multitude d'aléas, qui ont généralement pour effet de perturber l'exploitation. Nous les classifions en deux catégories : les incidents majeurs et les incidents mineurs, ou petits aléas.

Les incidents majeurs sont des événements relativement rares (ayant une fréquence d'occurrence de l'ordre d'une par jour et par ligne), ayant pour conséquence l'interruption partielle ou totale du trafic pour une durée pouvant aller de quelques minutes à plusieurs heures. Ces incidents font l'objet d'une documentation spécifique (dont le bulletin « Faits marquants ») permettant, entre autres, de justifier les résultats d'exploitation insatisfaisants auprès des autorités organisatrices. Ceux-ci sont en effet dûs à des causes externes pouvant être difficilement prévues ou anticipées, et dont la résolution nécessite la mise en œuvre de moyens importants. Les principales causes d'incidents majeurs sont les suivantes :

- Les accidents de personne. Ce terme désigne toute collision entre un véhicule ferroviaire et une ou plusieurs personnes, qu'elle soit accidentelle ou qu'il s'agisse d'un suicide. Cet événement nécessite une interruption de trafic pendant plusieurs heures : il est en effet nécessaire d'évacuer les personnes accidentées en ambulance, de faire intervenir un officier de police judiciaire pour une enquête sur les circonstances de l'accident, ainsi que de relever l'ADC en charge du train concerné.
- La présence de personnes sur les voies. Afin d'éviter des accidents de personnes, des alertes radio sont transmises à tous les ADC dès que des personnes sont détectées sur les voies afin de stopper la circulation jusqu'à l'évacuation et l'éventuelle interpellation de celles-ci. En zone rurale, il arrive parfois que les intrusions ne soit pas le fait de personnes mais d'animaux (le 19 novembre 2019, la circulation sur la ligne P a dû être interrompue suite à la présence d'un lama sur les voies!).
- Les bagages et colis abandonnés. En raison de l'application du plan Vigipirate, les circulations sont interrompues et un périmètre de sécurité est dressé lors de la découverte d'un bagage abandonné. Si une équipe cynophile est présente à proximité, celle-ci peut lever le doute sur la nature du bagage en quelques minutes. Sinon, une intervention d'une équipe de déminage est nécessaire, l'interruption du trafic peut alors durer plus d'une heure.
- Le dérangement d'installations (voie, caténaire, signalisation, aiguilles...). En dépit des opérations de maintenance régulières, ces installations peuvent faire l'objet de défaillances imprévisibles. Il est alors le plus souvent nécessaire de faire intervenir une équipe spécialisée et l'interruption du trafic peut durer plusieurs heures.
- La panne de matériel roulant. Lorsque cet événement survient, l'ADC applique la procédure décrite par son guide de dépannage. Dans certain cas, cela permet une reprise rapide du trafic (après quelques minutes d'interruption). Si la panne est plus grave, il est nécessaire d'évacuer la rame immobilisée. Cela peut alors prendre plusieurs heures.
- Les malaises de voyageurs. Il arrive, en particulier lorsque les trains sont très chargés, que certaines personnes à bord fassent des malaises. L'intervention des pompiers est alors parfois nécessaire, ce qui a pour effet d'interrompre les circulations jusqu'à leur arrivée.

Chaque ligne voit également se produire une multitude de petits aléas, d'amplitude bien plus faible, et ne faisant pour cette raison pas l'objet d'une documentation spécifique. Ils sont donc presque « invisibles » lors de la phase post-opérationnelle, mais n'en ont pas moins un impact sur l'écoulement du trafic. Carré (2014) dresse un panorama des causes de ces incidents mineurs. Les plus significatifs sont les suivants :

- L'hétérogénéité des attitudes de conduite. La conduite des trains en Île de France n'est pas automatisée, et chaque ADC dispose d'une certaine liberté dans sa façon de conduire. Cela concerne notamment la façon d'appréhender la grille horaire : certaines ADC la respectent scrupuleusement, d'autres roulent au plus près de la vitesse limite et prennent donc de l'avance sur l'horaire. Cette avance est ensuite compensée par un stationnement plus long en gare, le départ en avance étant lui interdit. Plus rarement, certains ADC prennent de petits retards et comptent sur la marge sillon pour les rattraper.

Cette hétérogénéité se manifeste également dans la façon d'anticiper et de réagir face à la signalisation. En effet, le franchissement d'un signal à l'avertissement impose de rester sous la courbe de freinage imposée par le KVB, cette courbe étant inconnue de l'ADC. Ainsi, certains ADC, pour limiter au maximum le risque de déclenchement du freinage d'urgence, visent la vitesse de 30 km/h plus de 200 mètres en amont du signal, tandis que d'autres prennent moins de marge pour éviter de prendre trop de retard. Certaines ADC cherchent également à éviter au maximum le franchissement de signaux à l'avertissement, et donc ralentissent lorsqu'ils anticipent la présence d'une circulation quelques cantons plus loin devant eux. Ces différences de comportements humains ont donc pour conséquence une variabilité des temps de parcours

entre les gares, et donc des fluctuations de la trajectoire autour de la marche théorique, ces variations étant difficiles à prévoir.

- Les durées de stationnement en gare et l'affluence voyageurs. Les trains Transilien ne nécessitant pas de réservation pour monter à leur bord, il est difficile de prévoir précisément le nombre de voyageurs qui descendront et monteront à bord des trains dans les différentes gares. Cependant, ce nombre influe directement sur la durée nécessaire à l'échange des voyageurs en gare, et donc sur la durée de stationnement. En outre, le stationnement est le principal moment où les voyageurs interagissent avec le système, ceux-ci peuvent donc également induire des variabilités : la montée de voyageurs avec bagages, ou le blocage de portes, peut ainsi allonger le temps de stationnement. Enfin, l'hétérogénéité des attitudes de conduite se manifeste également sur les durées d'arrêt en gare : certains ADC déclencheront le processus de fermeture des portes dès l'échange voyageurs terminé voire quelques secondes avant la fin de celui-ci, tandis que d'autres attendront quelques secondes supplémentaires pour permettre à d'éventuels voyageurs en retard d'embarquer.
- Les incertitudes lors de la conception des grilles horaires. En effet, les grilles sont généralement conçues avec une durée de validité d'une année entière. Cependant, il est difficile lors de la phase de conception d'avoir une vision claire de l'ensemble des événements susceptibles de jouer sur les temps de parcours et de stationnement des trains tout au long de cette année ; d'autant que des différences peuvent apparaître d'une journée à l'autre. Par exemple, il est fréquent que plusieurs types de matériels roulants coexistent sur une même ligne, et qu'un train donné soit assuré par un type de matériel le jour A et un autre le jour B. Les grilles horaires sont par conséquent conçues en prenant pour référence le matériel le moins performant circulant sur la ligne. En outre, le matériel roulant peut faire l'objet de défaillances qui, sans rendre son immobilisation obligatoire, peuvent diminuer les performances de freinage et d'accélération. Des chutes de tension ponctuelles sur la caténaire, ou des conditions météorologiques défavorables (pluie, gel, présence de feuilles mortes...) peuvent également dégrader les conditions de circulation et amener des trains à sortir de leur sillon.

Ces variabilités sont illustrées par la figure 1.18, où le GET du train 133406 Versailles - Paris-Saint-Lazare est tracé pour tous les jours de circulation de la période novembre 2018 - janvier 2019. On observe des variations de part et d'autre du tracé théorique.

Si les retards primaires induits par ces petits aléas sont généralement de faible amplitude (voire négatifs), la très forte interaction entre les trains peut conduire à une amplification de ces retards et à une propagation à d'autres trains. Deux phénomènes en particulier sont susceptibles d'être à l'œuvre dans ce processus :

- Amplification et propagation par le biais de la signalisation. Comme nous l'avons vu, si un train voit l'intervalle avec le train qui le précède se réduire, il rencontre *in fine* un signal d'avertissement. Celui-ci impose à l'ADC de réduire sa vitesse. Mais cela peut à son tour amener le train qui est derrière à se rapprocher suffisamment pour se voir présenter un signal d'avertissement qui le contraint à freiner. Ainsi, si la marge de grille séparant les trains est insuffisante, le retard se propage et s'amplifie jusqu'à ce qu'un sillon de respiration mette fin au processus. Ce phénomène est illustré par l'escalier d'occupation de la figure 1.19. Le KVB a généralement tendance à l'amplifier : les ADC n'ayant pas connaissance des courbes de freinage (par mesure de sécurité), certains peuvent se trouver largement en-deçà de celle-ci, provoquant ainsi des ralentissements plus importants.
- Amplification sous l'effet de l'affluence voyageurs. Si l'intervalle entre deux trains s'écarte de la valeur nominale, le quai de la prochaine gare desservie par le second train sera plus chargé que si celui-ci était arrivé avec l'intervalle prévu. Il devra donc stationner plus longtemps pour permettre à ce surplus de voyageurs de monter. Cela conduit par conséquent l'intervalle avec le train précédent à augmenter encore, et donc le phénomène à se répéter à la gare suivante

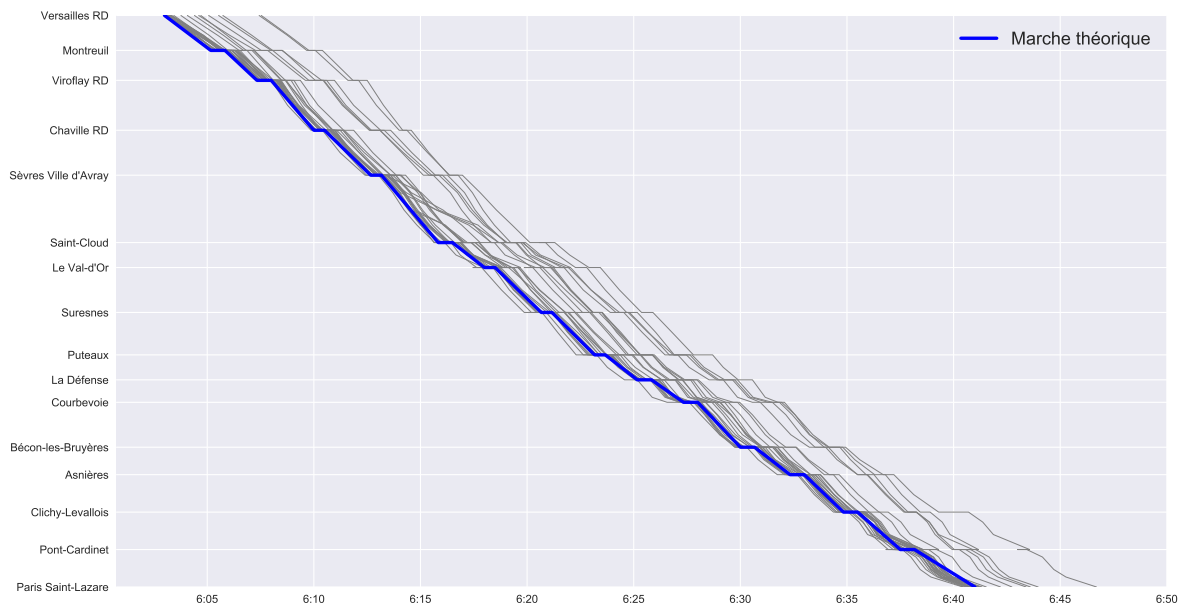


Figure 1.18 – Illustration des variabilités sur le GET du train 133406

avec une amplitude encore plus importante. Ce phénomène est régulièrement observé dans le cas du bus (où il porte le nom de « *bus bunching* »), et il a été mis en évidence dans le cas des lignes ferrées urbaines par Van Breusegem, Campion et Bastin (1991). La régulation RATP intervient fréquemment sur ses lignes pour mettre fin à ce phénomène, par exemple en imposant un arrêt prolongé au premier train de façon à rétablir l'intervalle.

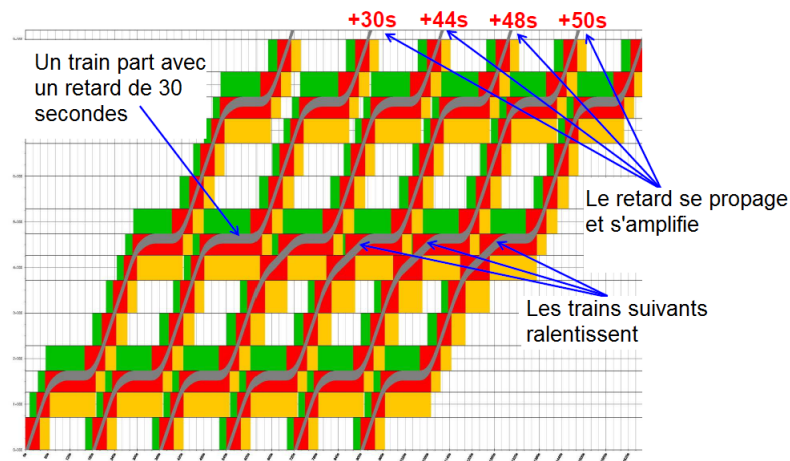


Figure 1.19 – Propagation des retards sous l'effet de la signalisation

Lorsque ce phénomène de propagation se produit, on observe un ralentissement de l'ensemble des circulations, ce qui se traduit sur le GET par des trajectoires aux pentes moins élevées : on dit que la grille « se couche ». Cela a conduit Carré (2014) puis Altazin (2018) à définir la stabilité comme une notion dynamique traduisant l'amplitude de ce phénomène. Selon leur définition, le système est d'autant plus proche de sa zone de stabilité que les trains sont proches de leurs sillons théoriques. Carré (2014) propose ainsi des indicateurs permettant de quantifier la stabilité du système à un instant donné, tandis que Altazin (2018) présente des méthodes de régulation permettant de rétablir la stabilité lorsque celle-ci se dégrade. Dans cette thèse, nous nous plaçons davantage en amont : nous étudions la propension des grilles horaires à permettre au système de se maintenir dans sa zone

de stabilité.

1.4 Problématique et principales contributions

Les travaux présentés dans cette thèse s'inscrivent donc dans un contexte caractérisé par les insuffisances du système de transport ferroviaire en zone dense. Celles-ci se traduisent en une qualité de service dégradée pour les voyageurs, ainsi que des pénalités financières pour l'opérateur. La croissance urbaine se poursuivant en Île de France, et la problématique environnementale étant de plus en plus prégnante, la demande en transports en commun ne cesse d'augmenter, ce qui amplifie encore les conséquences de ces insuffisances.

L'augmentation de la capacité maximale de l'infrastructure étant une opération coûteuse, tant sur le plan financier que pour les interruptions du trafic que les travaux engendreraient, d'autres solutions doivent être mises en places pour répondre à la demande de transport de façon satisfaisante malgré le contexte de saturation. Une remise en question des méthodes de travail actuelles peut donc s'avérer nécessaire pour parvenir à une meilleure utilisation du réseau ferré d'Île de France. Dans cette thèse, nous nous intéressons en particulier à la phase de conception des grilles horaires.

Il est tout d'abord souhaitable de s'interroger sur les performances que l'on attend d'une grille horaire et sur les indicateurs utilisés pour les mesurer. En particulier, il convient de définir et de quantifier la robustesse d'une façon adaptée au cadre et aux enjeux spécifiques de la zone dense. Une discussion sur cette notion constitue ainsi le cœur du chapitre 2.

Ensuite, il est nécessaire de disposer de modèles et de méthodes permettant une meilleure compréhension des phénomènes à l'origine de la dégradation du plan de transport et sur lesquels il est possible d'agir. Ceci permettant le développement d'outils ou l'amélioration de processus visant à limiter l'impact de ces perturbations, tant lors de la phase opérationnelle qu'en amont de celle-ci. Différents modèles sont ainsi présentés dans le chapitre 3.

Les critères de performance des grilles horaires étant fixés et les mécanismes de dégradation de la qualité de service suffisamment bien compris, il nous faut pouvoir évaluer la sensibilité d'une grille horaire donnée à ces perturbations, et notamment sa propension à l'instabilité. Un algorithme, reposant sur le principe de la simulation, est proposé pour cela dans le chapitre 4.

Enfin, il est souhaitable de connaître sur quels leviers agir et de quelle manière pour concevoir des grilles horaires offrant une meilleure robustesse. Une méthode heuristique visant à répondre à cet objectif est présentée dans le chapitre 5.

« Aaaaaaaaaaaaah ! Je me sens déjà plus instruit ! ... »

*Archibald Haddock, dans Hergé, **Les aventures de Tintin - On a marché sur la Lune**, 1954.*

Chapitre 2

Notion de grille horaire robuste en zone dense

Résumé

Ce chapitre vise à formaliser la notion de grille horaire robuste introduite au chapitre 1. Les principales méthodes de conception de grilles horaires sont présentées, ainsi qu'un état de l'art sur le concept de robustesse. Les limites des approches existantes et leur inadaptation au cadre de la zone dense sont mises en évidence. Cela nous amène à conclure ce chapitre par la description d'un modèle *ad hoc* pour la construction de grilles horaires en zone dense.

Sommaire

2.1	Le Train Timetabling Problem	38
2.1.1	Présentation	38
2.1.2	Description de l'infrastructure	38
2.1.3	Graphe d'événements-activités	39
2.1.4	Un problème de satisfaisabilité : le <i>Feasible Differential System</i>	40
2.1.5	Modèles cycliques et non-cycliques	41
2.2	Concepts de robustesse en optimisation	43
2.2.1	Présentation	43
2.2.2	Robustesse stricte	43
2.2.3	Robustesse légère	44
2.2.4	Robustesse avec recours	45
2.2.5	Optimisation stochastique	46
2.3	Critères de performance d'une grille horaire	47
2.3.1	Indicateurs de performance en mode nominal	47
2.3.2	Indicateurs de robustesse d'une grille horaire	49
2.4	Versions robustes du problème de Train Timetabling	52
2.4.1	Optimisation stochastique	53
2.4.2	Robustesse légère	53
2.4.3	Robustesse avec recours	54
2.4.4	Autres approches	55
2.5	Un modèle pour le TTP robuste en zone dense	55
2.5.1	Limites des indicateurs et modèles existants en zone dense	55
2.5.2	Un indicateur orienté voyageurs	57
2.5.3	Formalisation du problème par un programme stochastique	57

« La seule façon sûre de prendre un train, c'est de manquer le précédent. »
Pierre Daninos, **Vacances à tous prix**, 1958.

2.1 Le Train Timetabling Problem

2.1.1 Présentation

Nous avons mentionné au chapitre 1 que la plupart des étapes de la conception d'un plan de transport peuvent être modélisées sous la forme de problèmes combinatoires. Ceux-ci peuvent prendre la forme de problèmes réalisables, où l'on cherche à obtenir une solution satisfaisant toutes les contraintes, ou de problèmes d'optimisation, où l'on cherche parmi toutes les solutions réalisables celles qui offrent la meilleure valeur pour certains critères. Cette considération vaut en particulier pour le problème de la conception de grilles horaires, plus connu sous le nom de *Train Timetabling Problem* ou *TTP*. Nous présentons ici les principaux modèles et méthodes de résolution de ce problème. Un certain nombre d'entre eux peuvent être retrouvés dans les revues de littérature de Cacchiani et Toth (2012) et Caimi, Kroon et Liebchen (2017).

Dans tout ce chapitre, nous supposons que le problème du *Line Planning*, c'est-à-dire la définition des missions, des dessertes et des fréquences, a déjà été résolu ; ses résultats constituent en effet une donnée d'entrée au problème TTP.

2.1.2 Description de l'infrastructure

Les caractéristiques de l'infrastructure ferroviaire, en particulier le cantonnement et les gares, se traduisent par des contraintes fortes lors de la construction des grilles horaires : capacité maximale, temps de trajets minimaux entre différents points, intervalles temporels minimaux entre deux trains... C'est pourquoi il est nécessaire de disposer d'une description de l'infrastructure adaptée. A SNCF Réseau, une description détaillée et exhaustive de l'infrastructure (voie, aiguilles, installations électriques et de télécommunication, signalisation...) est disponible dans un outil informatique baptisé GAIA. Toutefois, celle-ci est bien trop complexe pour être utilisée lors du processus de *timetabling*, c'est pourquoi un modèle simplifié de l'infrastructure s'avère nécessaire.

Radtke (2008) dresse un état de l'art sur la modélisation de l'infrastructure ferroviaire. La quasi-totalité de ces modèles prennent la forme de graphes, constitués :

- de nœuds, représentant des points d'intérêt de l'infrastructure,
- et d'arcs, représentant des connexions entre nœuds.

Les différences entre ces modèles résident principalement dans le niveau de détail avec lequel l'infrastructure est représentée. On distingue généralement trois niveaux de modélisation :

- L'échelle microscopique, où la description du réseau est extrêmement fine. Ainsi, chaque aiguille ou signal constitue un nœud et chaque arc représente au plus un segment de voie.
- L'échelle macroscopique, où la description est plus agrégée : un nœud peut englober plusieurs points d'intérêt, comme un arc peut représenter plusieurs segments de voie.
- L'échelle mésoscopique, constituant un intermédiaire entre les deux précédentes ; selon les besoins, certaines parties de l'infrastructure y sont représentées de façon fine ou agrégée.

Les figures 2.1, 2.2 et 2.3 illustrent ces trois niveaux de modélisation, en présentant une description de la même infrastructure à trois échelles différentes.

Le logiciel AGATHE utilisé au sein de la branche Transilien du groupe SNCF utilise une description microscopique, de façon à s'assurer que la grille horaire tracée est sans conflit. Cependant, la plupart des algorithmes visant à automatiser le processus de *timetabling* reposent sur une description plutôt macroscopique, au moins pour la génération de la grille. En effet, un modèle trop fin entraînerait un temps de calcul prohibitif si le périmètre géographique du réseau est important. Une vérification de l'absence de conflit à l'aide d'un modèle plus microscopique peut néanmoins avoir lieu dans un second temps (Bešinović et al., 2016). La plupart des modèles que nous présentons dans la suite s'appuient ainsi sur une description macroscopique ou mésoscopique du réseau.

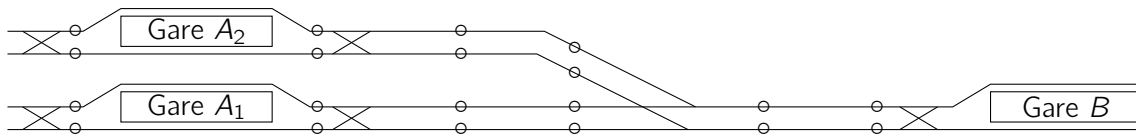


Figure 2.1 – Un modèle microscopique

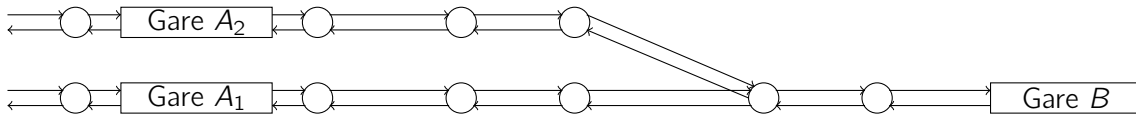


Figure 2.2 – Un modèle mésoscopique

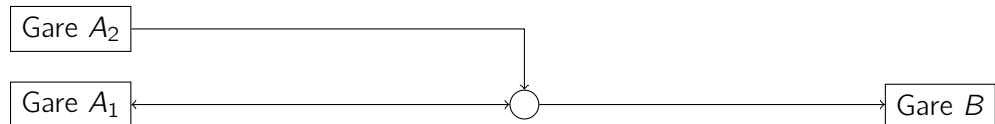


Figure 2.3 – Un modèle macroscopique

2.1.3 Graphe d'événements-activités

Le Graphe d'Événements-Activités (GEA) est un modèle macroscopique couramment utilisé pour les problèmes de planification. Sa définition est proche de celle des diagrammes PERT, qui sont des outils classiques pour l'ordonnancement de tâches (voir par exemple Teghem (2013)). Ces deux outils permettent en particulier de représenter les contraintes de précédence entre différents événements à planifier. Le GEA consiste en un graphe orienté valué $\mathcal{G} = (E, A, \omega)$, où

- E est un ensemble d'événements à planifier.
- A est un ensemble d'activités dont le début et la fin constituent des événements de E .
- $\omega : A \rightarrow \mathbb{R}_+$ est une fonction de valuation, indiquant la durée minimale de chaque activité de A .

Ainsi, si l'activité (e_1, e_2) appartient à A (avec $e_1 \in E, e_2 \in E$), alors l'événement e_2 ne pourra avoir lieu qu'au moins $\omega(e_1, e_2)$ unités de temps après l'événement e_1 .

Dans le cas du TTP, les événements sont généralement les arrivées et départs des gares ainsi que les passages en certains points remarquables du réseau. Les activités représentent les stationnements en gare, les temps de parcours entre deux gares, les intervalles minimaux entre deux trains imposés par le cantonnement, ainsi que les correspondances entre trains le cas échéant.

La construction du GEA est présentée sur un exemple à la figure 2.4, pour un cas simple. On considère sur cette figure que l'on doit planifier deux trains 1 et 2, qui partent respectivement des gares A_1 et A_2 , franchissent la convergence a (le train 1 passant avant le train 2), desservent la gare B et ont leur terminus à la gare C . On note alors :

- $a_{i,g}$ l'événement arrivée du train $i \in \{1, 2\}$ à la gare $g \in \{B, C\}$
- $d_{i,g}$ l'événement départ du train $i \in \{1, 2\}$ de la gare $g \in \{A_1, A_2, B\}$
- $p_{i,a}$ l'événement passage du train $i \in \{1, 2\}$ au point a
- $rt_{g,g'}$ le temps de parcours minimum entre les deux gares $g, g' \in \{A_1, A_2, B, C\}$
- dt_B le temps de stationnement minimum à la gare B
- h_x l'intervalle temporel minimum entre deux trains au point $x \in \{a, B, C\}$

Un algorithme de calcul de longueur de chemin, tel que l'algorithme de Bellman-Ford, permet alors d'attribuer aux événements une date de réalisation et donc un horaire. Celui-ci attribue à chaque événement sa date de réalisation au plus tôt, ce qui pour la grille horaire d'un réseau de transport

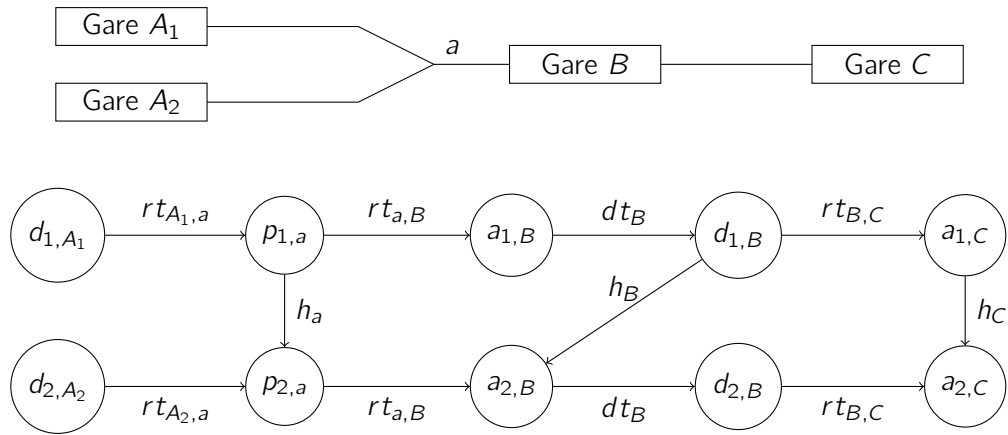


Figure 2.4 – Exemple de Graphe Événements-Activités

ne constitue pas toujours la solution optimale. Néanmoins, il permet de résoudre ce TTP en temps polynomial.

2.1.4 Un problème de satisfaisabilité : le Feasible Differential System

Si le modèle précédent aide à formaliser le problème TTP, il présente toutefois l'inconvénient de se généraliser difficilement à des situations plus complexes, qui nécessiteraient de faire intervenir d'autres contraintes ou des fonctions objectif à optimiser. La reformulation du problème précédent sous la forme d'un programme mathématique permet de pallier ces inconvénients. Ce programme porte le nom de *Feasible Differential System* (FDS), et a été proposé par Rockafellar (1984). On suppose donc donnés :

- Un graphe d'événements-activités $\mathcal{G} = (E, A)$, connexe et sans cycle, représentant les différents événements à planifier et leurs contraintes de précédence.
- Pour chaque activité $a \in A$, une borne inférieure ℓ_a et une borne supérieure u_a (éventuellement infinie) sur la durée de l'activité.

On appelle **horaire** toute fonction $\Pi : E \rightarrow \mathbb{R}$. Un horaire Π est dit **réalisable** si

$$\forall (i, j) \in A, \quad \ell_a \leq \Pi_j - \Pi_i \leq u_a$$

Le problème FDS consiste en la recherche d'un tel horaire réalisable. On appelle formulation « horaire » cette formulation du problème.

Il en existe une autre, appelée formulation « marges », qui selon le point de vue adopté peut se révéler plus pratique. Pour toute activité $a = (i, j) \in A$, on appelle **marge** attribuée à l'activité la quantité

$$s_a = \Pi_j - \Pi_i - \ell_a$$

Le problème consiste alors à attribuer à chaque activité $a \in A$ une marge s_a , de façon à ce que

$$\forall a \in A, \quad 0 \leq s_a \leq u_a - \ell_a$$

Les deux formulations sont équivalentes. Soit $s : A \rightarrow \mathbb{R}_+$ une fonction de marge réalisable. On appelle **chemin** de \mathcal{G} toute séquence d'événements $P = (i_1, \dots, i_k)$ telle que,

$$\forall j \in \{1, \dots, k-1\}, \quad (i_j, i_{j+1}) \in A \text{ ou } (i_{j+1}, i_j) \in A$$

Pour $j \in \{1, \dots, k-1\}$, si $(i_j, i_{j+1}) \in A$ alors cette activité est appelée activité directe. De la même manière, si $(i_{j+1}, i_j) \in A$, alors cette activité est appelée activité inverse. Étant donné un chemin

P , on note respectivement P^+ et P^- l'ensemble des activités directes et l'ensemble des activités inverses. On obtient alors un horaire réalisable en posant :

$$\begin{aligned} \Pi_i &= 0, \text{ où } i \in E \text{ est un événement quelconque de } \mathcal{G} \\ \forall j \in E, j \neq i, \Pi_j &= \sum_{a \in P_{ij}^+} (\ell_a + s_a) - \sum_{a \in P_{ij}^-} (\ell_a + s_a) \text{ où } P_{ij} \text{ est un chemin arbitraire de } i \text{ à } j \end{aligned}$$

Ces horaires sont tous identiques à une translation près.

Ces deux formulations du problème TTP, si elles n'englobent pas toute sa complexité, ont néanmoins servi de base et de source d'inspiration pour une multitude d'autres modèles. Nous en présentons les principaux dans le paragraphe suivant.

2.1.5 Modèles cycliques et non-cycliques

2.1.5.1 Un modèle cyclique : le PESP

Le **cadencement** est un principe de construction horaire mis en place sur de nombreux réseaux ferroviaires, notamment suburbains. Il consiste en la conception de la grille horaire à partir d'une trame de base qui est répétée périodiquement. Ce mode de conception offre l'avantage de faciliter la compréhension de la grille par les usagers : ainsi, sur la ligne H du réseau Transilien, les trains pour Luzarches ont tous la même desserte et partent de Paris Nord à 10h04, 11h04, 12h04, 13h04, etc. Le cadencement facilite également le travail des agents en réduisant leur charge mentale. Il présente en revanche l'inconvénient, dans le cas du cadencement strict où la même trame est répétée toute la journée, de proposer la même offre en heure de pointe et en heure creuse ; ce qui peut causer un gaspillage de ressources si celles engagées en heures creuses sont excédentaires par rapport à la demande, ou à l'inverse un nombre de trains insuffisant en heure de pointe. C'est pourquoi plusieurs lignes ont recours à un cadencement plus souple, avec un trame utilisée en heure creuse et le week-end, et une trame plus dense pour les heures de pointe en semaine.

Le *Periodic Event Scheduling Problem (PESP)* est un programme mathématique permettant de concevoir des grilles horaires cadencées. Il a été proposé par Serafini et Ukovich (1989) dans le cadre général de processus industriels cadencés, puis adapté au cas particulier du Train Timetabling par Odijk (1996).

Comme pour le problème FDS, on considère un ensemble d'événements E , un ensemble d'activités $A \subset E \times E$, chaque activité ayant une durée minimale ℓ_a et une durée maximale u_a . On se fixe en outre une période de cadencement T , et on suppose que pour tout $a \in A$, $u_a \leq T$. Le problème PESP consiste en la recherche d'une fonction $\Pi : E \rightarrow [0, T[$ telle que

$$\forall (i, j) \in A, \quad \ell_{i,j} \leq \Pi_j - \Pi_i \leq u_{i,j} \pmod{T} \quad (2.1)$$

Cette contrainte n'est cependant pas linéaire. Elle est linéarisée en introduisant les variables binaires Q_a , $a \in A$. Pour $a = (i, j) \in A$, Q_a vaut 1 si $\Pi_j < \Pi_i$ (on dit alors que l'activité a traverse la fin du cycle), et 0 sinon. La contrainte (2.1) se réécrit alors

$$\forall (i, j) \in A, \quad \ell_{i,j} \leq \Pi_j - \Pi_i + Q_{i,j}T \leq u_{i,j} \quad (2.2)$$

$$\forall (i, j) \in A, \quad \Pi_i - \Pi_j \leq MQ_{i,j} \quad (2.3)$$

où M est une constante arbitrairement grande.

Le problème PESP est NP-complet : en effet, il généralise le problème de la coloration des sommets d'un graphe (Odijk, 1997). Différentes fonctions objectif le transformant en problème d'optimisation ont été proposées. La plus courante consiste à minimiser la consommation de capacité, et donc la marge sillon totale :

$$\min \sum_{(i,j) \in A} w_{i,j} s_{i,j} \quad \text{où } s_{i,j} = \Pi_j - \Pi_i - \ell_{i,j} + Q_{i,j}T \quad (2.4)$$

où $w : A \rightarrow \mathbb{R}_+$ est une fonction de pondération. Si l'on note B la matrice d'incidence du graphe $\mathcal{G} = (E, A)$, le problème PESP s'écrit alors :

$$\begin{aligned} \min w^T \cdot (B^T \Pi + QT) & \quad (2.5) \\ \ell \leq B^T \Pi + QT \leq u \\ 0 \leq \Pi \leq T\mathbf{1} \\ Q \in \{0, 1\}^A \end{aligned}$$

Plusieurs méthodes de résolution ont été expérimentées :

- Pour la recherche de solutions réalisables : programmation par contraintes (Odijk, 1996), résolution à partir du problème SAT (Großmann et al., 2012)...
- Pour la recherche de solutions optimales : des techniques de programmation mathématique (Nachtigall, 1998), méta-heuristiques (Nachtigall et Voget, 1996)...

Ce modèle a été mis en pratique pour la construction de grilles horaires dans plusieurs pays (Allemagne, Pays-Bas, ...) et fait aujourd'hui encore l'objet de nombreuses recherches et extensions.

2.1.5.2 Modèles non-cycliques

Le modèle PESP ne convient pas à toutes les instances du Train Timetabling Problem. C'est le cas en particulier pour les lignes où les trains ne circulent pas sur une infrastructure dédiée, mais aussi pour les trains grandes lignes ou les TGV pour lesquels la fréquence est plus faible. Il faut alors recourir à des modèles non-cycliques¹. Outre le FDS, un modèle faisant référence est celui proposé par Caprara, Fischetti et Toth (2002). Celui-ci adopte un point de vue proche de celui du GI, qui reçoit des commandes de sillons provenant de plusieurs EF et doit arbitrer entre elles de façon équitable. Il facilite également la décision lorsqu'une même EF souhaiterait idéalement faire circuler plusieurs trains dans des sillons bien précis, sans que cela soit possible en raison de conflits pour l'utilisation de l'infrastructure. Le modèle permet alors d'adapter les sillons de façon à obtenir une grille réalisable.

Ce modèle prend en entrée un ensemble de trains $T = \{1, \dots, t\}$, un ensemble de gares $G = \{1, \dots, g\}$, un ensemble de dates $Q = \{1, \dots, q\}$, ainsi que pour chaque train $j \in T$:

- Une gare de départ f_j , une gare d'arrivée ℓ_j et une liste de gares desservies G^j .
- Un horaire idéal, indiquant des horaires d'arrivée et de départ dans les différentes gares de G^j .
- Des fenêtres temporelles indiquant les dates d'arrivée et de départ possibles dans les différentes gares de G^j .
- Un profit idéal p_j , correspondant au profit fait si le train j est planifié à son horaire idéal.

Les contraintes de capacité de l'infrastructure imposent à chaque gare de G des intervalles de temps minimaux entre deux arrivées successives et deux départs successifs. Si tous les trains ne peuvent être planifiés dans leur sillon idéal, deux modifications peuvent être appliquées à leurs sillons :

- Un décalage (positif ou négatif) de l'heure de départ de la première gare ; le décalage ν_j de l'horaire du train j se faisant avec un coût $\phi_j(\nu_j) \geq 0$ qui est une fonction croissante de $|\nu_j|$.
- Un allongement (positif) du temps de parcours. Pour le train $j \in T$, celui-ci est noté μ_j et se fait avec un coût $\psi_j(\mu_j) \geq 0$, qui est une fonction croissante de μ_j .

Le problème consiste alors à maximiser le profit total :

$$\max \sum_{j \in T} p_j - \phi_j(\nu_j) - \psi_j(\mu_j) \quad (2.6)$$

1. Cette distinction entre modèles cycliques et non-cycliques est néanmoins artificielle : un modèle non-cyclique n'est autre qu'un modèle cyclique avec une durée de cycle d'une journée.

Ce problème est également NP-complet.

D'autres modèles acycliques ont été proposés, cherchant généralement à coupler le problème TTP et d'autres problèmes de la phase de conception ou pré-opérationnelle, de façon à constituer un modèle intégré. Ainsi, Carey et Crawford (2007) intègrent les problèmes de la construction de la grille horaire et du GOV au sein d'un seul modèle, tandis que celui de Robenek et al. (2016) calcule en même temps une grille horaire et une projection du parcours des voyageurs. Un travail récent mené au sein de SNCF a débouché sur un modèle intégrant *line planning* et *timetabling*, tout en prenant en compte la demande voyageurs (Brethomé, 2018).

2.2 Concepts de robustesse en optimisation

2.2.1 Présentation

Le problème de la construction d'une grille horaire performante peut donc se modéliser sous la forme d'un problème d'optimisation. Or, les données qui guident les choix de construction de ces grilles (affluence voyageurs, temps de parcours et de stationnement, intervalles avec les trains précédents) sont connues de façon incertaine et fluctuent pour un train donné d'une journée à l'autre. Un des premiers travaux portant sur la résolution d'un problème d'optimisation dans ce cadre fut mené par Soyster (1973), donnant naissance à la théorie de l'optimisation robuste. Celle-ci a depuis pris son essor et fait l'objet de nombreux développements. Un panorama très complet des principaux résultats et méthodes est donné dans le livre de Ben-Tal, El Ghaoui et Nemirovski (2009), tandis que Gabrel, Murat et Thiele (2014) passent en revue les avancées les plus récentes. Nous présentons ici plusieurs concepts visant à prendre en compte d'éventuelles incertitudes lors de la définition et de la résolution d'un problème d'optimisation. La littérature concernant l'optimisation robuste distingue généralement les cas où l'incertitude porte sur les contraintes de ceux où elle porte sur la fonction objectif. Ce dernier cas peut présenter un intérêt pour notre propos (notamment lorsque la fonction objectif fait intervenir la demande voyageurs, qui n'est pas connue avec précision). Nous l'écartérons néanmoins pour nous concentrer sur l'autre cas, les incertitudes sur les contraintes étant à la fois plus fréquentes et susceptibles d'avoir des conséquences plus importantes si elles ne sont pas prises en compte correctement.

Nous considérons donc un problème d'optimisation

$$\begin{aligned} & \min_{x \in \mathcal{X}} f(x) & (2.7) \\ & \forall i \in \{1, \dots, m\}, \quad g_i(x, \zeta) \leq 0 \end{aligned}$$

où x est un vecteur à valeurs dans un ensemble \mathcal{X} dont les composantes sont les variables de décision, f est la fonction objectif, g_1, \dots, g_m des fonctions décrivant les contraintes portant sur les variables de décision. Ces contraintes dépendent également de données décrites par un vecteur ζ . La théorie de l'optimisation robuste vise à fournir un cadre pour la résolution d'un tel problème lorsque le jeu de données ζ fait l'objet d'incertitudes, que ce soit parce qu'elles ne sont pas mesurées avec précision ou qu'elles sont susceptibles de varier d'une réalisation à l'autre. Nous supposerons que ζ prend ses valeurs dans un ensemble \mathcal{U} , appelé ensemble d'incertitude.

2.2.2 Robustesse stricte

Résoudre un problème d'optimisation dans sa version strictement robuste revient à rechercher une solution optimale qui satisfasse les contraintes pour tout jeu de données dans l'ensemble d'incertitude. Ben-Tal et Nemirovski (2002) donnent une description générale de ce problème, en adaptant

le problème d'optimisation 2.7 de la façon suivante :

$$\begin{aligned} & \min_{x \in \mathcal{X}} f(x) \\ & \forall i \in \{1, \dots, m\}, \forall \zeta \in \mathcal{U}, \quad g_i(x, \zeta) \leq 0 \end{aligned} \quad (2.8)$$

Cette formulation pose néanmoins plusieurs inconvénients, parmi lesquels :

- Elle nécessite de pouvoir déterminer précisément l'ensemble d'incertitude \mathcal{U} ,
- Le nombre de contraintes est susceptible de devenir infini,
- La version robuste est considérablement plus difficile que la version originale du problème.

Ben-Tal et Nemirovski (2002) ont prouvé que dans certains cas particuliers, la formulation robuste s'obtient par transformation polynomiale de la formulation initiale, ouvrant donc la voie à une résolution. C'est notamment le cas lorsque la fonction objectif et les contraintes sont linéaires et que l'ensemble d'incertitude est un polyèdre borné. Toutefois, la solution obtenue est significativement moins bonne que celle du problème original non robuste. Ce phénomène a été baptisé le « surconservatisme » : la garantie de conserver une solution réalisable dans tous les scénarios possibles se fait au prix d'une dégradation notable de la valeur de l'objectif. Bertsimas et Sim (2004) ont alors proposé une manière de mesurer cette dégradation : le « prix de la robustesse », égal au ratio des valeurs de la fonction objectif dans la version robuste et dans la version originale du problème. La robustesse stricte ayant un prix élevé, d'autres approches ont été proposées, consistant à relâcher la contrainte de robustesse afin de réduire le coût. Nous en présentons quelques unes ayant déjà été appliquées au problème du *Train Timetabling*.

2.2.3 Robustesse légère

L'approche « robustesse légère », proposée par Fischetti et Monaci (2009), s'affranchit de l'obligation de conserver une solution réalisable pour tout jeu de données de l'ensemble d'incertitude. Le principe de la méthode est d'accepter un certain niveau (fixé en entrée) de dégradation de la fonction objectif, et de tolérer qu'un certain nombre de contraintes soit violées. On cherche alors à minimiser une certaine distance mesurant la violation de ces contraintes. L'approche est illustrée dans le cadre de la programmation linéaire. Considérons un programme linéaire sous sa forme générale

$$\begin{aligned} & \min \sum_{j=1}^n c_j x_j \\ & \forall i \in \{1, \dots, m\}, \quad \sum_{j=1}^n a_{ij} x_j \leq b_i \\ & \forall j \in \{1, \dots, n\}, \quad x_j \geq 0 \end{aligned} \quad (2.9)$$

où les réels x_1, \dots, x_n sont les variables de décision et $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ sont des données dont seule la matrice A fait l'objet d'incertitudes. On note z^* la valeur optimale des solutions à ce problème, correspondant à la valeur obtenue lorsque tous les paramètres sont parfaitement connus. Pour tout (i, j) , \hat{a}_{ij} désigne l'incrément le plus élevé pouvant s'ajouter au coefficient a_{ij} dans l'ensemble d'incertitude \mathcal{U} . Les auteurs font l'hypothèse que, pour toute contrainte $i \in \{1, \dots, m\}$, il est irréaliste de supposer que tous les coefficients a_{ij} prendront une valeur différente de la valeur nominale. On peut donc se restreindre à la recherche d'une solution en supposant qu'au plus Γ_i coefficients seront concernés par cette incertitude. On note alors

$$\beta(x, \Gamma_i) = \max_{\substack{S \subseteq \{1, \dots, n\} \\ |S| \leq \Gamma_i}} \sum_{j \in S} \hat{a}_{ij} x_j$$

ce coefficient représentant l'augmentation maximale du terme de gauche de la i -ème contrainte dans l'hypothèse précédente. On fixe par ailleurs un niveau de protection $\delta > 0$, représentant la dégradation maximale de la fonction objectif que l'on considère comme acceptable. La formulation « robustesse légère » du problème 2.9 s'écrit alors

$$\begin{aligned} & \min \sum_{i=1}^m w_i \gamma_i & (2.10) \\ \forall i \in \{1, \dots, m\}, & \sum_{j=1}^n a_{ij} x_j \leq b_i \\ \forall i \in \{1, \dots, m\}, & \sum_{j=1}^n a_{ij} x_j + \beta(x, \Gamma_i) - \gamma_i \leq b_i \\ \forall j \in \{1, \dots, n\}, & \sum_{j=1}^n c_j x_j \leq (1 + \delta) z^* \\ & \forall j \in \{1, \dots, n\}, \quad x_j \geq 0 \\ & \forall j \in \{1, \dots, m\}, \quad \gamma_j \geq 0 \end{aligned}$$

La variable de décision γ_i représente une marge, grâce à laquelle on quantifie la violation de la contrainte i . Les coefficients w_i permettent de pondérer ces différentes marges, de façon par exemple à tenir compte des éventuelles différences de nature ou d'échelle entre les contraintes.

Plusieurs paramètres permettent de jouer sur l'arbitrage entre robustesse et qualité de la solution nominale :

- Le paramètre δ , la solution étant d'autant plus proche de la solution nominale que celui-ci est proche de 0.
- Les paramètres Γ_j . Leur donner à tous la valeur 0 correspond à se ramener au problème nominal, tandis que leur attribuer à tous la valeur n donne une solution robuste mais très conservative.

Il est à noter que, ce modèle autorisant le non-respect de certaines contraintes, il n'est utilisable que lorsque la violation de celles-ci peut provoquer une perte d'efficacité mais en aucun cas des risques de sécurité. Dans le cas du trafic ferroviaire, la sécurité étant assurée indépendamment de la gestion de la grille horaire, cette méthode est donc susceptible d'être appliquée.

2.2.4 Robustesse avec recours

Les modèles présentés jusqu'ici relèvent du cadre de l'optimisation robuste statique : une fois la valeur des variables de décision choisie, aucune autre action n'est possible, y compris après que l'incertitude soit levée. En pratique, de nombreux systèmes prévoient la possibilité de réaliser des actions correctives lorsque la situation est trop éloignée du scénario nominal. Par exemple, dans le cas du transport ferroviaire, des mesures de régulation peuvent être prises de façon à rétablir une situation satisfaisante lorsque le système subit des perturbations extérieures. Il est possible de tenir compte de ces leviers d'action grâce aux techniques relevant de ce que l'on appelle la robustesse avec recours. Une partie du prix de la robustesse est alors transféré sur cette deuxième phase d'action (de recours), limitant ainsi le phénomène de surconservatisme. Plusieurs formalisations ont été proposées dans la littérature ; nous présentons succinctement ici celle de Liebchen et al. (2009), qui a fait l'objet de plusieurs développements et notamment d'applications à l'exploitation ferroviaire.

Selon cette approche, un problème d'optimisation robuste avec recours nécessite la donnée de trois paramètres :

- Un problème d'optimisation initial. Nous prendrons ici le cas général du problème 2.7.

- Un ensemble \mathcal{U} de scénarios, décrivant l'incertitude sur la donnée ζ . Chacun de ces scénarios est supposé susceptible de se produire, sans hiérarchisation de probabilité entre eux.
- Un ensemble de recours, décrit par une classe \mathcal{A} d'algorithmes, prenant chacun en entrée une solution x au problème d'optimisation original, un scénario $\zeta \in \mathcal{U}$, et un coût maximal λ . Celui-ci retourne alors une solution modifiée (éventuellement non réalisable pour le problème original).

Une solution x au problème d'optimisation 2.7 est dite **robuste avec recours** au coût maximal de λ si, pour tout scénario $\zeta \in \mathcal{U}$, il existe un algorithme $A \in \mathcal{A}$ tel que $A(x, \zeta, \lambda)$ soit une solution réalisable dans le scénario ζ et ayant un coût inférieur à λ . Le problème se réécrit alors

$$\begin{aligned} \min_{(x, A, \lambda) \in \mathcal{X} \times \mathcal{A} \times \mathbb{R}_+} f(x) + h(\lambda) \\ \forall i \in \{1, \dots, m\}, \forall \zeta \in \mathcal{U}, \quad g_i(A(x, \zeta, \lambda), \zeta) \leq 0 \end{aligned} \quad (2.11)$$

La fonction $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ traduit la façon dont le coût du recours influence la fonction objectif. Le paramètre λ peut alors être vu comme représentant le prix de la robustesse. Remarquons que l'on retrouve le cas de la robustesse stricte si l'on suppose $h(\lambda) = 0$ pour tout λ et $\mathcal{A} = \{I\}$, où I est l'algorithme défini par $\forall x, \zeta, \lambda, I(x, \zeta, \lambda) = x$.

Cette approche est très générale et ne s'applique pas en l'état ; elle nécessite d'être spécialisée en fonction du problème sur lequel on souhaite la mettre en œuvre. Les auteurs en présentent notamment un exemple dans le cas de la programmation linéaire. Outre le *Train Timetabling*, on trouve dans la littérature plusieurs applications du concept à d'autres problèmes de l'exploitation ferroviaire, comme la planification des manœuvres (Cicerone et al., 2009) ; un doctorat actuellement en cours à la SNCF vise à l'appliquer au problème de la planification du matériel roulant (Lucas et al., 2019). Une extension du concept est proposée par Cicerone et al. (2012) pour le cas où plusieurs incertitudes sont levées successivement et que des décisions de recours sont prises à chaque étape.

2.2.5 Optimisation stochastique

L'optimisation stochastique est généralement présentée comme une branche indépendante de l'optimisation robuste, bien que visant comme elle à faciliter la prise de décision dans un contexte incertain. La principale différence avec cette dernière réside dans le fait que, contrairement à elle, on suppose connues certaines informations concernant la probabilité des différents scénarios de l'ensemble d'incertitude. Celui-ci est donc désormais noté Ω et muni d'une mesure de probabilité \mathbb{P} . On suppose que la valeur de la fonction objectif dépend également de la réalisation aléatoire, ainsi f est définie sur $\mathcal{X} \times \Omega$. On se fixe, pour chaque contrainte $i \in \{1, \dots, m\}$, un niveau de protection $\alpha_i \in [0, 1]$. Le problème d'optimisation 2.7 se réécrit alors, dans sa version stochastique,

$$\begin{aligned} \min_{x \in \mathcal{X}} \mathbb{E}[f(x, \omega)] \\ \forall i \in \{1, \dots, m\}, \quad \mathbb{P}(g_i(x, \omega) \leq 0) \geq \alpha_i \end{aligned} \quad (2.12)$$

Choisir une fonction f indépendante de ω et $\alpha_1 = \dots = \alpha_m = 1$ revient à se ramener (presque sûrement) au cas de la robustesse stricte.

Un tel problème ne peut se résoudre en l'état avec les méthodes d'optimisation exacte usuelles, il est par conséquent nécessaire de le reformuler ou de recourir à des heuristiques. La reformulation est possible dans les cas suivants :

- Lorsque l'ensemble des scénarios Ω est fini, le problème 2.12 est équivalent à un problème d'optimisation déterministe de même structure, obtenu en énumérant tous les scénarios possibles. Ainsi, si le problème initial est un programme linéaire stochastique, sa reformulation déterministe est également un programme linéaire, potentiellement de grande taille si le nombre de

scénarios est important. Sa structure permet néanmoins de le résoudre par décomposition de Benders.

- Lorsque l'ensemble des scénarios est infini mais que sa distribution est donnée par une loi discrète ou continue, le problème 2.12 peut être reformulé de façon non équivalente mais seulement approximative en un problème déterministe. Cela nécessite de discrétiser (dans le cas d'une loi continue) et d'échantillonner l'ensemble des scénarios. Un échantillonnage est également nécessaire lorsque la loi de probabilité \mathbb{P} est inconnue mais susceptible d'être estimée à partir de données. Notons que des travaux relativement récents cherchent à appliquer des techniques d'optimisation robuste à la programmation stochastique, dans le cas où la distribution de probabilité est connue de façon incertaine; voir par exemple Dupačová et Kopa (2012).

Le cadre de l'optimisation stochastique permet également des formulations avec recours. On trouve une vue d'ensemble des méthodes et champs d'application de l'optimisation stochastique dans le livre de Birge et Louveaux (1997).

2.3 Critères de performance d'une grille horaire

Nous avons vu au chapitre 1 qu'il est souhaitable que les grilles horaires produites lors de la conception du plan de transport soient robustes. Cependant, ce critère seul ne saurait suffire, le plus haut niveau de robustesse étant atteint en ne faisant circuler aucun train sur le réseau. Nous passons en revue dans cette partie les principaux indicateurs utilisés pour mesurer ou optimiser les performances d'une grille horaire, en portant tout de même une attention particulière aux indicateurs de robustesse. Une partie des indicateurs présentés ici peuvent être retrouvés dans les états de l'art dressés par Goverde et Hansen (2013) et Lusby, Larsen et Bull (2018).

2.3.1 Indicateurs de performance en mode nominal

Nous avons vu dans la partie précédente que la recherche d'une solution robuste à un problème d'optimisation se fait au détriment de la qualité de la solution obtenue, ce phénomène étant formalisé par Bertsimas et Sim (2004) grâce au concept de « prix de la robustesse ». Dans le cadre du problème du *Train Timetabling*, ce phénomène apparaît si l'on considère séparément la performance d'une grille horaire en mode nominal (lorsque le système ne subit aucune perturbation et que tous les trains sont à l'heure) et sa robustesse face aux aléas. Un certain nombre d'auteurs font ce choix dans la construction de leurs modèles :

- Soit parce que la robustesse fait l'objet d'une seconde étape, via un test de robustesse et/ou une adaptation de façon à améliorer celle-ci.
- Soit parce que le modèle utilise une approche multi-objectifs, certains objectifs concernant la performance nominale et les autres la robustesse.

Ce paragraphe vise donc à présenter certains des indicateurs utilisés pour mesurer la performance d'une grille horaire en mode nominal.

2.3.1.1 Indicateurs orientés voyageurs

La finalité du transport ferroviaire étant d'offrir un service aux utilisateurs, certains indicateurs visent naturellement à mesurer la qualité du service qui leur est fourni. Les passagers souhaitant rejoindre leur destination le plus rapidement possible, la plupart des indicateurs relevant de cette catégorie sont reliés au temps passé par les usagers dans le système. Ainsi, Peeters et Kroon (2001) et plus récemment Yang et al. (2019) proposent des modèles multicritères dont l'une des fonctions objectif est le temps de trajet total des passagères. Goverde et al. (2016) prennent également en compte le temps requis pour effectuer d'éventuelles correspondances. Ce critère est peu pertinent en

zone dense, où les trains sont empruntables sans réservation et où les fréquences sont importantes. Ainsi, dans ce cadre, Barrena et al. (2014) négligent les correspondances et considèrent que le temps de trajet est assez peu modifiable, d'où leur choix de construire une fonction objectif minimisant le temps d'attente des voyageurs en gare. Tous ces indicateurs peuvent être agrégés pour former ce que l'on appelle le **temps de trajet généralisé** : il s'agit du temps total passé par les passagers dans le système, incluant l'attente en gare, le trajet à bord et d'éventuelles correspondances. Ces différentes composantes peuvent éventuellement être pondérées selon la façon dont leur durée est ressentie par les voyageurs (Wardman, 2004). Cet indicateur est utilisé comme fonction objectif par Sels et al. (2016) et Brethomé (2018).

Outre le temps de trajet, le confort à bord est également important pour les voyageurs. Celui-ci étant généralement perçu comme décroissant lorsque le nombre de voyageurs à bord augmente (promiscuité, station debout, montée et descente plus difficiles...). C'est pourquoi Sun et al. (2013) prennent en compte la charge des trains dans leur fonction objectif.

Ces indicateurs, bien qu'ayant l'avantage d'être en cohérence avec la finalité du transport ferroviaire, nécessitent des données sur le nombre de voyageuses et les itinéraires qu'empruntent celles-ci. Or, ces données sont difficiles à mesurer et fluctuantes ; c'est pourquoi d'autres indicateurs sont également susceptibles d'être utilisés.

2.3.1.2 Indicateurs orientés opérateur

Le transport ferroviaire engendre des coûts importants pour les EF et la collectivité. Il est par conséquent souhaitable de minimiser ceux-ci autant que possible, tout en assurant le service prévu dans le contrat passé avec l'AO. Or, les coûts sont directement reliés au nombre de rames nécessaires pour assurer le service (en raison des coûts d'achat et de maintenance), ainsi que de la durée pendant laquelle celles-ci sont utilisées. De même, le nombre d'agents et leur durée de service impactent directement les coûts pour l'opérateur. Il est donc souhaitable de minimiser les durées de service des agents et du matériel roulant, et par conséquent la marge sillon. La marge de grille n'a elle en revanche aucune conséquence sur le coût pour l'opérateur (en mode nominal). Ainsi, Schöbel et Kratz (2009) et Lee et al. (2017) cherchent à minimiser l'allocation de marge sillon. Liebchen (2008) et Kroon et al. (2013) utilisent des approches similaires en se concentrant sur la durée des « crochets », c'est-à-dire des demi-tours aux gares terminus.

De façon analogue, dans le cas de services non conventionnés (par exemple, le TGV), il est souhaitable de proposer une grille horaire qui permettra de maximiser les bénéfices. Ainsi, Brännlund et al. (1998) proposent un programme mathématique visant à maximiser le profit, sans toutefois détailler extensivement la façon dont celui-ci est estimé. De façon générale, la maximisation du profit constitue plutôt un objectif de la phase de *line planning*, où les fréquences et dessertes sont planifiées. C'est pourquoi les modèles de *timetabling* utilisant une telle fonction objectif sont plutôt rares dans la littérature.

Enfin, l'énergie consommée par le matériel roulant, qu'elle provienne d'une caténaire ou d'un combustible, est également source de coûts pour l'opérateur, ainsi que pour l'environnement. En particulier, les phases de freinage et d'accélération sont très consommatrices. C'est pourquoi certains modèles les prennent en compte, de façon à générer des grilles horaires pouvant être opérées le plus sobrement possible (Goverde et al. (2016), Yang et al. (2019)).

2.3.1.3 Indicateurs orientés gestionnaire d'infrastructure

Enfin, certains modèles et indicateurs ont davantage vocation à être utilisés par les GI lors de l'attribution des sillons aux différents opérateurs et la construction de la grille horaire finale. Nous avons vu un premier modèle au paragraphe 2.1.5.2, où l'objectif est de maximiser la satisfaction totale des opérateurs demandant des sillons. Une perspective similaire est adoptée par Arenas et al.

(2015), où l'on cherche à maximiser un profit mesurant l'écart entre la grille horaire produite et une grille horaire « idéale ».

D'autres indicateurs peuvent caractériser la qualité d'une grille horaire du point de vue d'un GI. La consommation de la capacité en est un. En effet, le GI est responsable de la façon dont la capacité est utilisée. Par conséquent, à nombre de trains constant, une grille consommant moins de capacité offrira l'avantage de laisser davantage de marge de manœuvre pour le tracé de sillons de dernière minute ou pour des décisions opérationnelles. Ce point de vue est ainsi adopté par Gestrelus et al. (2012), qui cherchent à optimiser les possibilités de replanification en temps réel offertes au gestionnaire de l'infrastructure.

Notons également qu'en ce sens, les modèles cherchant à réduire l'utilisation de la marge sillon, présentés au paragraphe précédent, participent de l'optimisation de la capacité. Cependant, d'autres facteurs sont susceptibles de jouer, tels la marge de grille ou la mixité des trafics. La méthode de compression du feuillet UIC406, déjà décrite au chapitre 1, permet de quantifier la capacité consommée par la grille horaire. Goverde et al. (2016) proposent un modèle visant notamment à minimiser cette consommation de capacité.

2.3.2 Indicateurs de robustesse d'une grille horaire

2.3.2.1 Définitions

Il n'y a à l'heure actuelle pas de consensus dans la littérature scientifique sur la définition de la robustesse pour une grille horaire, ni sur les indicateurs permettant de la mesurer, ni même parfois sur le terme à employer : certains auteurs lui préfèrent celui de « résilience », de « stabilité » voire de « fiabilité ». Les principaux travaux s'accordent néanmoins sur l'existence de caractéristiques intrinsèques à certaines grilles horaires, permettant un maintien ou un rétablissement rapide du système dans son mode nominal après des perturbations de faible amplitude. Nous utiliserons ici les définitions suivantes, proposées notamment par Jensen (2015) :

- Une grille horaire est **robuste** si elle permet de conserver des indicateurs de performance (tels que ceux présentés au paragraphe précédent) satisfaisants en dépit de perturbations exogènes de faible amplitude.
- Une grille est **stable** si, après un ensemble de faibles perturbations, elle permet le retour du système dans un état où tous les trains circulent dans leur sillon après une certaine durée de transition.
- Une grille est **résiliente** si elle permet une transition rapide vers un nouvel état stable, en recourant si besoin à des mesures de régulation du trafic.

Nous porterons dans cette thèse une attention plus particulière à la robustesse et à la stabilité, aux sens définis ci-dessus. En particulier, on notera que la stabilité d'une grille horaire participe à sa robustesse.

Différents indicateurs ont été proposés pour traduire de façon quantitative ces deux notions. Certains ne nécessitent que la connaissance de la grille horaire pour être calculés, nous les qualifierons d'indicateurs *ex ante*. D'autres nécessitent des données sur la façon dont s'écoule le trafic lorsqu'il est régi par la grille horaire, ces données pouvant être obtenues à l'issue d'opérations réelles ou par simulation. Nous parlerons dans ce cas d'indicateurs *ex post*.

2.3.2.2 Indicateurs ex ante

Le calcul d'indicateurs relevant de cette catégorie ne requiert au maximum que trois types de données : une description de l'infrastructure, la distribution des retards primaires et la structure de la grille horaire. De par le faible nombre de données qu'ils nécessitent, ces indicateurs sont les plus faciles à calculer, et permettent donc une comparaison rapide entre plusieurs grilles horaires.

Comme nous l'avons déjà mentionné, la robustesse d'une grille est directement corrélée à la consommation de capacité de celle-ci. Il est par conséquent possible d'obtenir une première estimation en calculant la capacité consommée à l'aide, par exemple, d'une des méthodes présentées au chapitre 1 ou dans Abril et al. (2008). Cependant, cet indicateur est insuffisant, la façon dont la capacité est utilisée ayant également un fort impact sur la robustesse et la stabilité. Il est par conséquent nécessaire de disposer d'indicateurs plus précis.

Carey (1999) présente ainsi plusieurs méthodes heuristiques de mesure, pouvant être utilisées pour estimer à l'avance la robustesse d'une grille régissant la circulation aux alentours d'une grande gare. Certains nécessitent la connaissance des distributions de probabilité des retards primaires des événements se produisant sur le périmètre de la gare considérée (tout retard secondaire se produisant en dehors de ce périmètre étant considéré comme un retard primaire). Des méthodes pour l'estimation de ces distributions peuvent être trouvées dans la thèse de Yuan (2006) ou plus récemment dans les travaux de Faverges et al. (2018). La connaissance de ces distributions permet de calculer la probabilité qu'un train donné induise un retard secondaire sur le train suivant utilisant la même infrastructure (ici, le même quai), ainsi que l'espérance de l'amplitude de ce retard. Une estimation de la robustesse de la grille peut alors être obtenue en calculant une moyenne (éventuellement pondérée) de ces probabilités ou espérances sur l'ensemble des trains franchissant la gare. Ces quantités dépendant aussi de l'intervalle temporel entre les trains utilisant la même infrastructure (la marge de grille), d'autres indicateurs construits à partir de ces valeurs sont proposés : intervalle minimum entre deux trains, quantiles inférieurs des intervalles, moyenne et écart-type de ces intervalles... Ces indicateurs ne nécessitent pas la connaissance de distributions de probabilité des retards primaires, tout en étant corrélés avec la probabilité de survenue de retards secondaires. Vromans (2005) propose également des critères décrivant l'hétérogénéité des intervalles temporels, sur une période de temps donnée. Une variante, baptisée « indice de complexité », est proposée par Landex et Jensen (2013) dans le cas d'une grille horaire cyclique. L'hypothèse sous-jacente, là encore, est qu'une plus grande homogénéité des intervalles se traduit par un risque plus faible de retards secondaires, et donc une meilleure robustesse.

Goverde (2007) s'intéresse également aux intervalles entre trains, dans le cas d'une grille horaire cyclique, pour calculer un indicateur de stabilité. Il s'agit, pour lui, du plus grand retard primaire pouvant survenir lors d'un cycle et être absorbé pendant la durée de celui-ci, sans se répercuter par conséquent sur le cycle suivant. Un algorithme de plus court chemin, faisant appel aux réseaux de Petri et à l'algèbre max-plus, est proposé pour le calcul effectif. Delorme, Gandibleux et Rodriguez (2009) approfondissent cette idée et proposent une méthode multi-objectifs d'évaluation de la stabilité. Celle-ci serait caractérisée par le retard primaire maximal n'entraînant pas de retards secondaires, ainsi que par une fonction prenant en entrée une amplitude de retard primaire et retournant la somme des retards secondaires qui en découlent. Des méthodes de propagation des retards, proches de celles de Goverde (2007) sont présentées, en se limitant toutefois à l'échelle d'un nœud ferroviaire.

Enfin, certains auteurs font l'hypothèse qu'il est possible de se limiter à certains points spécifiques (du réseau ou du graphique espace-temps) pour pouvoir évaluer la robustesse d'une grille horaire sur le périmètre d'une ligne ou d'un réseau entier. Ces points sont choisis comme étant ceux où l'interaction entre les trains de la grille est la plus forte. Schittenhelm et Landex (2013) définissent ainsi le concept de points fixes comme étant les points de dépassement, de cisaillement ou de croisement. La robustesse d'un sillon est alors évaluée par le nombre de points fixes que rencontre le sillon, ainsi que les marges placées au voisinage de ces points fixes. Toutefois, l'indicateur le plus célèbre relevant de cette catégorie est la robustesse aux points critiques, proposée par Andersson, Peterson et Törnquist (2013). Selon les auteurs, un point critique est un point du graphique espace-temps où un train s'insère derrière un autre (après le franchissement d'une convergence, un dépassement ou le départ de la gare d'origine). L'hypothèse faite est que si la marge entre les deux trains est suffisante en ces points, elle devrait le rester tout au long de leur parcours. La robustesse en un

point critique est alors calculée comme la somme de la marge de grille en ce point, de la marge sillon du premier train en amont du point et de la marge sillon du deuxième train en aval du point critique. L'hypothèse de départ est cependant inadaptée à la zone dense, où les arrêts sont fréquents et les intervalles réduits. Dans ce cadre, les points critiques sont plutôt les gares faisant face aux flux voyageurs les plus élevés.

2.3.2.3 Indicateurs ex post

Si les indicateurs présentés précédemment ont l'avantage de pouvoir être calculés facilement, ils ne reflètent que de façon heuristique les propriétés de la grille en opération réelle. Leur fiabilité n'est donc pas garantie. C'est pourquoi il est souhaitable de disposer, en complément, d'indicateurs permettant de quantifier la robustesse *ex-post*, c'est à dire après utilisation de la grille horaire. Cependant, l'évaluation de la robustesse après utilisation en conditions réelles pose plusieurs inconvénients :

- Le coût, très élevé de par la quantité de ressources à engager (infrastructure, rames, personnel) et les éventuelles pénalités si la grille n'est pas performante.
- Très peu de grilles peuvent être ainsi évaluées. En effet, pour permettre aux agents et aux voyageurs de s'organiser et de se repérer, les modifications de la grille doivent être très peu fréquentes.
- La survenue d'événements parasites, tels que de fortes perturbations (colis suspect, malaise voyageur, accident de personne, signal d'alarme...). Celles-ci dégradent significativement les performances quelle que soit la grille utilisée. En outre, des mesures de régulation sont décidées lors de la phase opérationnelle ; il peut alors être difficile de déterminer si les éventuels bons résultats obtenus sont dûs à des caractéristiques intrinsèques de la grille horaire ou à l'efficacité de la régulation.

Pour toutes ces raisons, le recours à la simulation est préférable. Nous présentons plus en détail les méthodes et outils de simulation utilisables dans ce cadre au chapitre 4.

Il est à noter que la plupart des indicateurs de performance présentés au paragraphe 2.3.1 peuvent servir à l'estimation de la robustesse *ex-post*, s'ils sont calculés à partir des données post-opérationnelles. Cependant, s'ils permettent alors de mesurer une performance absolue, ils ne donnent que peu d'informations sur les écarts existant entre le plan de transport théorique et la façon dont celui-ci s'est effectivement déroulé. C'est pourquoi plusieurs indicateurs ont été proposés pour cela, avec des méthodes de calcul effectif *ad hoc*.

La notion de retard, définie comme l'écart entre la date prévue et la date effective de réalisation d'un événement, constitue l'ingrédient essentiel de la plupart de ces indicateurs. En particulier, les retards secondaires constituent un bon indicateur de robustesse, puisque seuls les retards relevant de cette catégorie peuvent être atténués par l'utilisation d'une grille horaire robuste. Cependant, la distinction entre retards primaires et secondaires n'est pas toujours une tâche aisée. C'est pourquoi certains travaux de recherche ont eu pour objectif de faciliter la classification des retards observés (Conte et Schöbel (2007), Flier et al. (2009)).

Ainsi, Carey et Carville (2000) simulent une grille horaire en la soumettant à différentes distributions de probabilité de retards primaires. La robustesse de la grille est alors évaluée en étudiant la façon dont sont distribués les retards secondaires correspondants pour les différents trains de la grille. Büker et Seybold (2012) proposent une méthode de simulation rapide de la propagation des retards utilisant un modèle de graphes ; la robustesse de la grille est alors évaluée, pour une certaine distribution de retards primaires, à partir de la proportion de retards situés en-dessous d'un certain seuil, l'espérance ainsi que la variance des retards. Certains opérateurs mesurent leur performance en utilisant l'indicateur de régularité à N minutes : il s'agit du pourcentage de circulations étant arrivées au terme de leur mission avec un retard inférieur ou égal à N minutes. Cet indicateur présente l'inconvénient d'attribuer le même score à des trains à l'heure qu'à des trains légèrement en

retard, tout comme pour des trains ayant respectivement un retard légèrement supérieur à N ou très supérieur à N minutes. A titre d'exemple, la régularité à 5 minutes est un des indicateurs les plus utilisés en France, tandis que les Pays-Bas considèrent généralement plutôt la régularité à 3 minutes. Ces indicateurs de régularité sont généralement calculés à partir de données d'exploitation réelles.

Vromans, Dekker et Kroon (2006) utilisent des indicateurs *ex-post* pour mettre en évidence la pertinence d'utilisation de certains indicateurs *ex-ante*, notamment des mesures d'hétérogénéité du trafic et de distribution de la marge de grille. Les auteurs prouvent que les grilles les plus homogènes, lorsqu'elles sont soumises à une distribution de retards primaires, permettent d'obtenir des retards secondaires plus faibles que les grilles plus hétérogènes. Salido, Barber et Ingolotti (2008) proposent également des comparaisons entre indicateurs relevant des deux catégories. Dans leurs travaux, l'indicateur *ex-post* utilisé constitue plutôt une mesure de la stabilité de la grille : le temps de retour à la normale après une succession de perturbations est estimé par simulation. Cet indicateur a été également jusqu'à peu l'un des principaux ingrédients des tests de robustesse à SNCF Réseau. Selon ces tests, une grille est considérée comme robuste si, lorsqu'elle est soumise à une perturbation isolée d'une amplitude de 10 minutes, celle-ci est totalement résorbée au point de survenue dans un délai d'une heure. La durée de résorption est estimée à l'aide d'un simulateur baptisé SAMURAIL.

Khadilkar (2016) propose deux métriques pour quantifier la robustesse. La robustesse individuelle mesure l'amplitude des perturbations que peut subir un train donné sans provoquer de retards secondaires sur le reste du réseau. La robustesse collective est définie comme la proportion de trains atteignant leur terminus à l'heure avec une certaine probabilité fixée. Dans l'article, ces métriques sont utilisées comme mesures de la robustesse avec recours : un simulateur est utilisé, pouvant prendre des décisions de régulation lorsque la situation l'impose.

D'autres indicateurs *ex-post*, n'étant pas construits (ou pas uniquement) à partir de valeurs de retards, ont également été proposés. Ainsi, Medeossi, Longo et Fabris (2011) utilisent une méthode de simulation pour estimer la fréquence de survenue des conflits entre circulations pour l'utilisation de l'infrastructure. Ces conflits entraînant inévitablement des ralentissements et étant des conséquences de la sortie de certains trains de leurs sillons, ils constituent également un bon indicateur de robustesse.

Plusieurs auteurs défendent quant à eux le point de vue selon lequel les indicateurs utilisés doivent nécessairement chercher à refléter la qualité de service offerte aux passagères, celles-ci étant la raison d'être du transport ferroviaire. Dans cette perspective, Takeuchi, Tomii et Hirai (2007) ont développé une métrique intégrant le temps d'attente en gare, le temps de trajet, le nombre de correspondances ainsi que la charge des trains. Une méthode de simulation permettant de quantifier ce critère d'« inconfort » pour chaque passager est également décrite ; la robustesse est alors définie comme l'écart d'inconfort entre la situation nominale et la valeur moyenne obtenue après un grand nombre de réplications.

Enfin, Dewilde (2014) cherche à proposer une définition universelle de la robustesse, généralisable à différentes situations et différents modes de transport. Selon lui, la robustesse doit être définie comme la qualité de service proposée aux usagers lors d'une journée de service « normale », c'est-à-dire soumise à des perturbations et des aléas suivant une distribution représentative d'une journée typique. En ce sens, la robustesse et la performance nominale ne sont plus deux objectifs contradictoires, mais complémentaires ; la notion de « prix de la robustesse » n'a alors plus lieu d'être.

2.4 Versions robustes du problème de Train Timetabling

Les travaux présentés dans le paragraphe précédent permettent de définir la robustesse pour une grille horaire et d'évaluer celle-ci. Cependant, ils ne permettent pas d'obtenir des informations sur les possibilités d'amélioration, ni sur d'éventuelles alternatives plus robustes. Utilisés tels quels,

ils ne permettent donc la construction de grilles robustes que par un processus d'essais-erreurs. C'est pourquoi il est souhaitable de disposer également d'outils dont le résultat en sortie est une grille horaire présentant de bonnes propriétés de robustesse. Nous présentons ici plusieurs approches visant à construire des grilles horaires optimales vis-à-vis d'un certain critère de robustesse, ou à améliorer la robustesse de certaines grilles déjà construites. La plupart de ces travaux s'appuient sur les paradigmes d'optimisation robuste présentés au paragraphe 2.2, ainsi que sur les modèles présentés à la section 2.1.

2.4.1 Optimisation stochastique

Lorsque les distributions de probabilités des aléas d'exploitation sont connues et supposées indépendantes de la grille horaire utilisée, l'optimisation stochastique constitue une approche naturelle pour la formalisation robuste du problème de *Train Timetabling*. Deux programmes stochastiques, qualifiés respectivement de « gras » et « maigre », sont proposés par Fischetti, Salvagnin et Zanette (2009). Tous deux nécessitent tout d'abord de résoudre un problème de TTP simple, inspiré de celui de Caprara, Fischetti et Toth (2002). Le modèle « gras » vise alors à minimiser le retard total, moyenné sur l'ensemble des scénarios, une borne inférieure sur la dégradation de la fonction objectif étant imposée. Les scénarios sont échantillonnés à l'aide de la méthode SAA (*Sample Average Approximation*, voir par exemple le chapitre 8 du livre de Fu (2014)). Le modèle offre de très bonnes performances, mais sa résolution nécessite un temps très long sur de grandes instances, c'est pourquoi les auteurs proposent un autre modèle « maigre ». Celui-ci a pour but de minimiser les retards non absorbés par la marge sillon, toujours sous contrainte de ne pas trop dégrader l'indicateur de performance initial. Deux versions de la fonction objectif sont proposées : l'une où les retards ne sont pas pondérés, et l'autre où ils le sont en fonction de la date à laquelle ils se produisent (les retards ayant lieu tôt dans la journée étant davantage pénalisés car leurs conséquences sont susceptibles de s'étaler sur une plus longue durée). Le modèle « maigre » pondéré offre des performances très proches de celles du modèle « gras », pour un temps de calcul substantiellement réduit.

Le modèle précédent intègre la construction d'une grille horaire et l'optimisation de sa robustesse. D'autres travaux proposent de traiter les deux étapes séparément. Il faut alors d'abord générer une grille en résolvant un problème de TTP nominal, puis y insérer de la marge de grille et de la marge sillon. Cette stratégie est explorée par Kroon et al. (2008), pour la construction d'une grille horaire cyclique. Leur modèle est constitué de deux parties : le module de *timetabling* calcule une distribution de la marge dans la grille horaire, tandis que l'autre module évalue la robustesse de la grille par simulation stochastique. Les scénarios sont ici aussi échantillonnés avec la méthode SAA. Chaque grille fait l'objet de plusieurs simulations aléatoires, son score étant la moyenne des résultats obtenus sur les différentes répliques. La fonction objectif à minimiser est la somme pondérée des retards à l'arrivée dans toutes les gares et pour tous les trains. Le modèle a été testé en conditions réelles, en mettant en application une grille générée par ce processus sur une ligne du réseau néerlandais. Les auteurs rapportent une amélioration de la régularité avec la nouvelle grille. Cette approche a fait l'objet de développements ultérieurs, pour améliorer les performances en termes d'optimalité et de temps de calcul ; voir les travaux de Vekas, Vlerk et Klein Haneveld (2012) et Maróti (2017).

Enfin, Hassannayebi et al. (2017) proposent un programme stochastique bi-objectif pour le cas moins complexe de la planification d'une ligne de métro. Les objectifs à minimiser sont le temps d'attente espéré des passagers en station ainsi que l'écart entre les dates prévues et effectives de réalisation des différents événements.

2.4.2 Robustesse légère

Le paradigme d'optimisation proposé par Fischetti et Monaci (2009) et baptisé « robustesse légère » a été appliqué au problème du *train timetabling*. Le problème nominal est inspiré de celui de Caprara, Fischetti et Toth (2002) : il s'agit de maximiser un profit dépendant de l'écart entre

la grille horaire obtenue et une grille idéale. On note T l'ensemble des trains, chaque train $t \in T$ étant associé à une succession d'événements E^t . Chaque événement $i \in E^t$ doit se voir attribuer un horaire Π_i^t . On note également ρ_t le profit maximum associé au train t , $\bar{\Pi}^t$ son horaire idéal, ϕ_t le coût unitaire du décalage de l'heure de départ de la première gare, et ψ_t le coût unitaire de l'allongement de temps de parcours du train t . Enfin, on note $d_{i,j}^t$ la durée minimale de l'activité (i, j) pour le train T , $h_{i,j}^{t,t'}$ l'intervalle minimal entre les dates des événements i, j pour les trains t et t' , et C l'ensemble des couples d'événements qui correspondent à un départ d'une même gare. Le problème nominal s'écrit alors

$$\begin{aligned} \max z &= \sum_{t \in T} \rho_t & (2.13) \\ \forall t \in T, \forall i \in E^t, & \quad \Pi_{i+1}^t - \Pi_i^t \geq d_{i,i+1}^t \\ \forall t, t' \in T, \forall i \in E^t, \forall j \in E^{t'}, & \quad |\Pi_i^t - \Pi_j^{t'}| \geq h_{i,j}^{t,t'} \\ \forall t, t' \in T, \forall i \in E^t, \forall j \in E^{t'}, & \quad (i, j) \in C \Rightarrow [\Pi_i^t < \Pi_j^{t'} \Leftrightarrow \Pi_{i+1}^t < \Pi_{j+1}^{t'}] \\ \forall t \in T, & \quad \rho_t = \rho_t - \phi_t |\Pi_1^t - \bar{\Pi}_1^t| - \psi_t ((\Pi_{|E^t|}^t - \Pi_1^t) - (\bar{\Pi}_{|E^t|}^t - \bar{\Pi}_1^t)) \end{aligned}$$

Ce problème peut être linéarisé, et donc résolu à l'optimalité par des techniques standard. On note z^* la solution optimale. La version « robustesse légère » s'écrit alors

$$\min \sum_{\substack{t \in T \\ i \in E^t}} \delta_i^t \quad (2.14)$$

$$\sum_{t \in T} \rho_t \geq (1 - \alpha) z^* \quad (2.15)$$

$$\begin{aligned} \forall t \in T, \forall i \in E^t, & \quad \Pi_{i+1}^t - \Pi_i^t + \delta_i^t \geq d_{i,i+1}^t \\ \forall t, t' \in T, \forall i \in E^t, \forall j \in E^{t'}, & \quad |\Pi_i^t - \Pi_j^{t'}| \geq h_{i,j}^{t,t'} \\ \forall t, t' \in T, \forall i \in E^t, \forall j \in E^{t'}, & \quad (i, j) \in C \Rightarrow [\Pi_i^t < \Pi_j^{t'} \Leftrightarrow \Pi_{i+1}^t < \Pi_{j+1}^{t'}] \\ \forall t \in T, & \quad \rho_t = \rho_t - \phi_t |\Pi_1^t - \bar{\Pi}_1^t| - \psi_t ((\Pi_{|E^t|}^t - \Pi_1^t) - (\bar{\Pi}_{|E^t|}^t - \bar{\Pi}_1^t)) \end{aligned}$$

où δ_i^t désigne la marge attribuée au train t pour l'activité $(i, i + 1)$, et α est la proportion de dégradation de la fonction objectif que l'on est prêt à accepter (le prix de la robustesse). Cette méthode a été comparée aux programmes stochastiques présentés dans le paragraphe précédent (voir Fischetti, Salvagnin et Zanette (2009)). Les auteurs rapportent qu'elle a permis d'obtenir des résultats quasiment équivalents à ceux du programme stochastique « maigre », mais au prix d'un temps de calcul significativement meilleur.

2.4.3 Robustesse avec recours

Le problème TTP a également été décliné en de nombreuses versions relevant du paradigme baptisé « robustesse avec recours ». Ces versions diffèrent par la version nominale du problème qui est utilisée, mais également par le choix des algorithmes de recours inclus dans le modèle. Nous en présentons ici une des plus simples, proposée par Liebchen et al. (2009). Le problème nominal est le problème FDS présenté au paragraphe 2.1.4, dont nous reprenons les notations. La fonction objectif du problème nominal est ici la somme pondérée des marges sillons. L'ensemble des scénarios est noté \mathcal{U} , un scénario de perturbations $\zeta \in \mathcal{U}$ se traduisant par un allongement de la durée minimale nécessaire à la réalisation des différentes activités. Étant donnée une variable nominale x , sa valeur dans le scénario ζ est notée x^ζ . On considère un seul algorithme de recours : si, dans un scénario donné, la solution nominale n'est plus réalisable, il est possible de récupérer une solution réalisable en propageant les retards et en « couchant » la grille. Cependant, ces retards ont un coût ; on cherche

donc à minimiser le retard maximal λ dans tous les scénarios, éventuellement en lui attribuant un coût g . La version robuste avec recours du problème s'écrit alors :

$$\begin{aligned}
& \min \sum_{(i,j) \in A} w_{i,j}(\Pi_j - \Pi_i) + g\lambda & (2.16) \\
& \forall (i,j) \in A, \quad \Pi_j - \Pi_i \geq \ell_{i,j} \\
& \forall \zeta \in \mathcal{U}, \forall (i,j) \in A, \quad \Pi_j^\zeta - \Pi_i^\zeta \geq \ell_{i,j}^\zeta \\
& \forall \zeta \in \mathcal{U}, \forall i \in E, \quad \Pi_i^\zeta \geq \Pi_i \\
& \forall \zeta \in \mathcal{U}, \forall i \in E, \quad \Pi_i^\zeta - \Pi_i \leq \lambda \\
& \lambda \geq 0
\end{aligned}$$

D'autres versions, proposant un plus large choix d'algorithmes de recours, ont été proposées. Celles-ci permettent par exemple, en cas de perturbations, de recourir à des ruptures de correspondances (Liebchen et al. (2009), Liebchen et al. (2010)). Une variante vise à utiliser des algorithmes de recours permettant de rétablir une solution optimale (Goerigk et Schöbel, 2014).

Il est cependant à noter que la complexité en espace du problème est très importante lorsque la classe des algorithmes de recours est grande. Cela est dû à la nature itérative du modèle, où les décisions et la réalisation de scénarios alternent. Le problème TTP robuste avec recours est ainsi PSPACE-difficile (Liebchen et al., 2009), ce qui rend sa résolution à l'optimalité délicate sur de grandes instances.

2.4.4 Autres approches

De nombreuses autres stratégies ont été proposées dans le but d'obtenir des solutions robustes au problème TTP. Schöbel et Kratz (2009) considèrent le problème comme étant multi-objectifs : le premier objectif étant la performance nominale (mesurée par la consommation de capacité), et le second la robustesse (définie comme le retard maximal ne provoquant pas de rupture de correspondance). Le même point de vue est adopté par Schlechte et Borndörfer (2010), avec des fonctions objectifs légèrement différentes, ainsi que par Cacchiani, Caprara et Fischetti (2012) qui présentent une heuristique lagrangienne pour l'approximation du front de Pareto.

Certaines approches abordent le problème en se plaçant du point de vue des passagers. Il s'agit alors de maximiser la qualité de service qui leur est offerte. Les travaux de Vansteenwegen et Oudheusden (2006) relèvent de cette catégorie : il s'agit de réduire le temps d'attente espéré des passagers en cas de retards. Le modèle a été testé sur le réseau ferré belge et les auteurs rapportent une amélioration de la qualité de service.

Enfin, certains auteurs se concentrent sur le problème de l'allocation des marges : partant d'une grille existante, le problème consiste à distribuer les marges de grille et sillon de façon à rendre la grille la plus robuste possible. Jovanović et al. (2017) modélisent ainsi ce problème comme un problème de sac à dos multidimensionnel : un certain budget est alloué, à ne pas dépasser (pour éviter une surconsommation de capacité). Il faut alors distribuer au mieux les marges permises par ce budget dans la grille.

2.5 Un modèle pour le TTP robuste en zone dense

2.5.1 Limites des indicateurs et modèles existants en zone dense

Comme nous l'avons vu au chapitre 1, la zone dense présente de nombreuses caractéristiques qui la distinguent du reste du réseau. Le fonctionnement du système ferroviaire y est par conséquent sujet à certaines particularités. Or, la grande majorité des indicateurs et modèles présentés dans ce chapitre ne nous paraissent pas adaptés à ce cadre spécifique. Les raisons en sont les suivantes :

- En zone dense, la notion de performance nominale n'a plus de sens. En effet, le système est soumis en permanence à de multiples aléas. Ceux-ci sont la plupart du temps de faible amplitude. Cependant, lorsque l'infrastructure est utilisée au maximum de sa capacité (comme c'est le cas en Île de France), même de faibles aléas sont susceptibles de faire sortir le système de sa zone de stabilité. Il n'est donc pas souhaitable de chercher à optimiser un critère de performance en mode nominal, car les conditions pour que celui-ci soit effectivement atteint ne seront quasiment jamais réunies.
- La plupart des définitions usuelles de la robustesse ne sont pas non plus adaptées. En effet, les indicateurs *ex-ante* sont généralement construits autour de la consommation de la capacité, qui est ici maximale. Les indicateurs *ex-post* utilisent eux, pour la plupart, la notion de retard, qui n'est pas pertinente en zone dense. En effet, la tenue de l'horaire n'est pas toujours un facteur déterminant de la qualité de service offerte aux usagers. La fréquence étant importante, ceux-ci se présentent généralement en gare sans consultation préalable de l'horaire et embarquent dans le premier train desservant leur destination. Un retard généralisé de tous les trains avec la même amplitude n'a donc pour effet que de translater la grille horaire dans le temps, sans conséquence sur la qualité de service, alors que ce retard sera lourdement pénalisé par les critères classiques de robustesse. En zone dense, il s'agit surtout de maintenir une fréquence et des temps de parcours satisfaisants pour les usagers.
- La plupart des méthodes précédentes font l'hypothèse que les perturbations qui surviennent sont peu nombreuses et d'amplitude relativement importante (plusieurs minutes). Or, en zone dense, de telles perturbations ne peuvent pas être absorbées sans dommages en raison de la saturation de l'infrastructure. Un cadre adapté au cas où les perturbations sont nombreuses (voire omniprésentes) et de faible amplitude est par conséquent nécessaire.
- La majorité des indicateurs et des modèles ne prennent pas en compte la présence de passagers dans le système, ou seulement de façon très simplifiée. Celle-ci a cependant une importance à tous les niveaux de la modélisation :
 - ◇ Dans la fonction objectif, puisque la raison d'être du transport ferroviaire est d'offrir un service de qualité aux personnes qui l'empruntent. De plus, en Île de France, la qualité de service fait l'objet d'un contrat avec l'autorité organisatrice, entraînant des pénalités financières si les objectifs ne sont pas atteints.
 - ◇ Dans les contraintes : en zone dense, les volumes de passagers transportés sont très importants. Il peut donc y avoir saturation des trains si ceux-ci n'ont pas une capacité suffisante ou si la fréquence est trop faible.
 - ◇ Dans les mécanismes de survenue et de propagation des perturbations, via les durées de stationnement en gare (voir les chapitres 1 et 3).
- Enfin, les définitions classiques de la robustesse portent sur la capacité de la grille à absorber les petits aléas sans intervention de la part des régulateurs. Or, en zone dense, des décisions de régulation sont prises sur toutes les lignes et tout au long de la journée. Il est donc préférable d'adopter une définition, des indicateurs et un modèle qui permette de tenir compte de cette réalité.

À notre connaissance, un seul des modèles publiés dans la littérature est susceptible d'être adapté facilement de façon à satisfaire à ces contraintes. Il a été proposé dans la thèse de Dewilde (2014), dans le cadre de la planification robuste au voisinage des grandes gares. Cependant, la définition et les indicateurs proposés ont été conçus de façon à être généralisables à d'autres situations. Nous nous inspirons donc de ces travaux, et dirons désormais qu'une grille horaire est **robuste** si **elle permet d'obtenir des performances satisfaisantes lorsqu'elle est soumise à une distribution « normale » d'aléas d'exploitation**. La suite de cette section est consacrée à la traduction de cette définition sous la forme d'un indicateur, ainsi qu'à un modèle mathématique formalisant le problème de l'optimisation de cet indicateur.

2.5.2 Un indicateur orienté voyageurs

Le transport ferroviaire conventionné est organisé et financé par une autorité indépendante des entreprises ferroviaires. L'objectif de la convention n'est pas de maximiser le profit réalisé par les opérateurs signataires, mais d'optimiser la qualité du service rendu aux usagers. Il est donc souhaitable, dans ce cas, d'adopter un indicateur de performance adoptant le point de vue des voyageurs. Cependant, un tel indicateur requiert des données sur les flux de passagers qu'il n'est pas toujours aisé d'obtenir avec une précision satisfaisante. C'est pourquoi de nombreux auteurs ont fait le choix d'utiliser plutôt des indicateurs construits à partir de données décrivant l'écoulement du trafic ferroviaire, plus faciles à recueillir. Néanmoins, la multiplication depuis quelques années des systèmes de comptage automatique des voyageurs permet de disposer de données récentes, avec un volume et une précision suffisantes. C'est pourquoi nous avons choisi d'évaluer la robustesse de grilles à partir d'un indicateur orienté voyageurs, construit suivant le principe proposé par Dewilde (2014) et présenté au paragraphe précédent.

Le but d'un voyageur empruntant un système de mobilité n'est pas d'être transporté, mais de rejoindre sa destination finale le plus rapidement possible (Bouvarel, 2011). Nous avons donc choisi le **temps de trajet généralisé** comme indicateur principal de la qualité de service offerte aux usagers. Celui-ci est défini comme la somme du temps passé en gare d'origine à attendre l'arrivée du train, le temps passé à bord du train, et la durée des correspondances le cas échéant. Ce dernier terme ne sera pas pris en compte dans la suite de ces travaux : en effet, la convention passée entre Île de France Mobilités et Transilien ne fixe pas d'objectifs concernant celles-ci, et les données concernant le nombre de passagers en correspondance ne sont pas faciles à obtenir. Nous nous limitons donc à la somme du temps d'attente et du temps de trajet.

Si le temps de trajet est facilement calculable à partir de données d'exploitation, le temps d'attente est plus difficile à estimer, étant donné qu'il est inenvisageable d'effectuer un suivi individuel de tous les passagers. Nous faisons donc les hypothèses simplificatrices suivantes :

- Dans les gares où la fréquence théorique est inférieure ou égale à 15 minutes, les voyageurs se présentent en gare sans consulter la grille horaire et embarquent dans le premier train se rendant à leur destination. Nous supposons donc qu'il y a dans ces gares un flux continu de passagers, homogène dans le temps.
- Dans les gares où la fréquence est supérieure à 15 minutes, les passagers se présentent en gare 3 minutes avant l'heure de départ prévue à l'horaire.

Suivant le principe de Dewilde (2014), nous considérons le temps d'attente TA et le temps de trajet TT comme des variables aléatoires, dont nous chercherons à minimiser l'espérance. Le modèle présenté dans le paragraphe suivant vise à formaliser ce problème.

2.5.3 Formalisation du problème par un programme stochastique

2.5.3.1 Hypothèses et notations

La fonction objectif à minimiser est l'espérance d'une variable aléatoire, dont la valeur dépend de celles des variables de décision et des aléas. Notre problème relève donc du cadre de l'optimisation stochastique, qui a été présenté au paragraphe 2.2.5. Nous faisons les hypothèses suivantes :

- Nous supposons le problème du *line planning* déjà résolu. La liste des trains à assurer et leur ordre sont connus et fixés, ainsi que leurs dessertes. Le problème étudié porte donc principalement sur l'allocation des marges sillon et d'exploitation dans la grille horaire.
- Les temps de parcours entre deux gares, les durées de stationnement en gare, le retard au départ de l'origine sont des variables aléatoires dont les distributions sont connues (leur estimation fait l'objet du chapitre 3).

- Les flux de passagers sont donnés par des matrices origine-destination heure par heure. Ces matrices donnent, pour chaque heure de la journée, une estimation du nombre de passagers effectuant le trajet (Gare A, Gare B) pour tous les couples de gares possibles de la ligne.
- Les trains ont une capacité limitée. Lorsque la capacité maximale est atteinte, le nombre de passagers montant effectivement dans le train et se rendant à une gare donnée en aval est proportionnel à la taille du groupe de passagers souhaitant monter dans le train pour se rendre à cette gare. Les passagers n'ayant pu monter attendent le train suivant et retiennent leur chance.
- Lorsque la fréquence théorique est inférieure ou égale à 15 minutes, les passagers se présentent en gare selon un flux uniformément distribué en temps. Sinon, ils se présentent 3 minutes avant l'heure de départ théorique.

Nous utiliserons les notations suivantes :

- \mathcal{S} : ensemble des voies à quai. $\mathcal{S} = \mathcal{S}_h \cup \mathcal{S}_f$, où \mathcal{S}_h et \mathcal{S}_f sont respectivement les ensembles des voies à quai où la fréquence est supérieure et inférieure à 15 minutes.
- \mathcal{T}_s : ensemble des trains desservant la gare $s \in \mathcal{S}$
- $\mathcal{T}_{s,s'}$: ensemble des trains desservant les gares s et s' dans cet ordre
- $D_{t,s}^-$: ensemble des gares desservies par le train t avant la gare s
- $D_{t,s}^+$: ensemble des gares desservies par le train t après la gare s

Les variables de décision sont les suivantes :

- $\bar{d}_{t,s}$: heure théorique de départ du train t de la gare s

Les variables aléatoires ci-dessous ont une loi supposée connue :

- $DT_{t,s}$: durée de stationnement du train t à la gare s
- $RT_{t,s,s'}$: durée de trajet du train t de la gare s à la gare s'

Enfin, le modèle fait intervenir les paramètres et autres variables ci-dessous :

- T_f : durée d'une journée
- $pf_{s,s',\theta,\theta'}$: flux horaire de passagers sur l'O-D $s \rightarrow s'$ entre les dates θ et θ'
- $p_{t,s,s'}$: nombre de passagers empruntant le train t pour se rendre de la gare s à la gare s'
- $\hat{p}_{t,s,s'}$: nombre de passagers souhaitant se rendre de s à s' rejetés par le train t
- $p_{t,s}^{stay}$: nombre de passagers restant dans le train t à la gare s
- $p_{s,s'}(\theta)$: nombre de passagers attendant sur le quai de la gare s et souhaitant se rendre en s' à la date θ
- $a_{t,s}$: variable donnant l'heure effective d'arrivée du train t à la gare s
- $d_{t,s}$: variable donnant l'heure effective de départ du train t de la gare s . La distribution de probabilité de cette variable aléatoire est supposée connue dans le cas où s est la gare d'origine du train t .
- $prec_t(s)$: train précédant t à la gare s
- $prec_t(s, s')$: train précédant t à la gare s et desservant aussi la gare s'
- $prec_s(t)$: gare desservie par le train t juste avant la gare s
- $suiv_t(s)$: gare desservie par le train t juste après la gare s
- $suiv_s(t)$: train suivant le train t à la gare s
- H : durée d'une heure ($H = 3600$ si l'unité est la seconde)
- τ_s^{stop} : durée minimale de stationnement à la gare s

- $\tau_{s,s'}^{run}$: temps de trajet minimal de la gare s à la gare s'
- τ^{head} : intervalle minimal entre deux trains pour assurer la sécurité
- $\tau_{s,s'}^{wait}$: temps d'attente à la gare s des passagers se rendant à la gare s'
- c_t : capacité du train t

2.5.3.2 Fonction objectif

L'espérance du temps de trajet généralisé de l'ensemble des passagers est utilisée comme fonction objectif. Celle-ci s'écrit

$$\min \mathbb{E}[TTG] = \min \mathbb{E}[TA + TT]$$

où TA est le temps d'attente en gare et TT le temps de trajet à bord d'un véhicule. Ce dernier s'écrit

$$TT = \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}_s} p_{t,s}^{stay} (d_{t,s} - a_{t,s}) + (a_{t,s_{\text{suiv}_t(s)}} - d_{t,s}) \sum_{s' \in D_{t,s}^+} p_{t,s,s'} \quad (2.17)$$

Le temps d'attente peut quant à lui être formulé de la façon suivante :

$$TA = \sum_{(s,s') \in \mathcal{S}^2} \sum_{t \in \mathcal{T}_{s,s'}} \tau_{s,s'}^{wait} \quad (2.18)$$

où

$$\tau_{s,s'}^{wait} = \int_0^{T_f} p_{s,s'}(\theta) d\theta \quad (2.19)$$

Cependant, la fonction $p_{s,s'}$ a une expression différente, selon que la gare s soit une gare « à la fréquence » de l'ensemble \mathcal{S}_f , ou une gare « à l'horaire » de l'ensemble \mathcal{S}_h . Les figures 2.5 et 2.6 représentent l'allure de cette fonction dans ces deux cas.

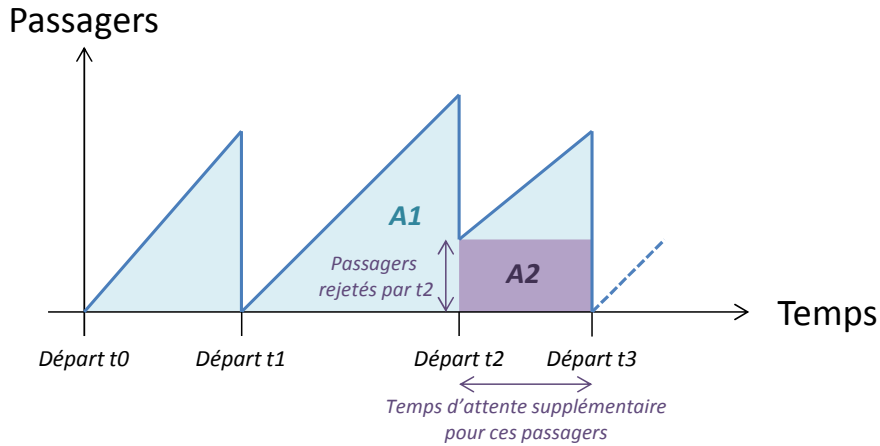


Figure 2.5 – Intégrale de la fonction $p_{s,s'}$ si $s \in \mathcal{S}_f$

Dans le cas d'une gare à la fréquence, les passagers arrivent selon un flux régulier. La fonction $p_{s,s'}$ est donc affine par morceaux, la pente étant donnée par le flux horaire d'arrivée des passagères en gare. Son intégrale s'écrit, en tenant compte de la possibilité pour un train de refuser des passagers pour des raisons de capacité :

$$\tau_{s,s'}^{wait} = \sum_{t \in \mathcal{T}_{s,s'}} \frac{(d_{t,s} - d_{\text{prec}_t(s,s'),s})^2}{2H} p_{f_{s,s',d_{\text{prec}_t(s,s'),s}},d_{t,s}} + (d_{t,s} - d_{\text{prec}_t(s,s'),s}) \hat{p}_{\text{prec}_t(s,s'),s,s'} \quad (2.20)$$

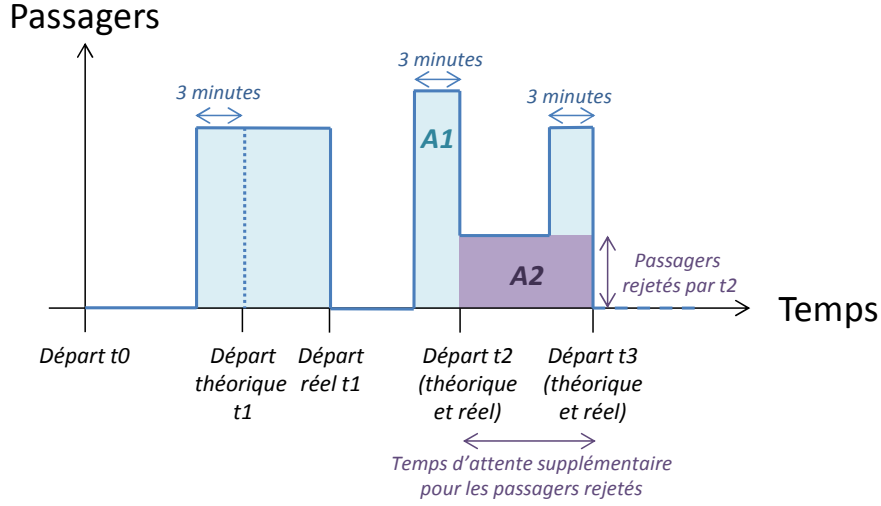


Figure 2.6 – Intégrale de la fonction $\rho_{s,s'}$ si $s \in \mathcal{S}_h$

Le terme de gauche correspond au temps d'attente des passagères arrivées en gare après le passage du train précédent, représenté par l'aire A_1 sur la figure. Celui de droite correspond au temps d'attente supplémentaire des passagers ayant été rejetés par le train précédent. Il est représenté par l'aire A_2 .

Dans le cas d'une gare à l'horaire, tous les passagers se présentent en gare 3 minutes avant l'heure théorique de départ de leur train. La fonction $\rho_{s,s'}$ est donc constante par morceaux et son intégrale s'écrit, toujours en tenant compte de la possibilité pour un train de refuser des passagers pour des raisons de capacité :

$$\tau_{s,s'}^{wait} = \sum_{t \in \mathcal{T}_{s,s'}} \left(d_{t,s} - \bar{d}_{t,s} + \frac{H}{20} \right)^2 p f_{s,s',d_{prec_t(s,s'),s},\bar{d}_{t,s}} + (d_{t,s} - d_{prec_t(s,s'),s}) \hat{p}_{prec_t(s,s'),s,s'} \quad (2.21)$$

Comme précédemment, le terme de droite correspond au temps d'attente des passagers rejetés par le train précédent et celui de gauche au temps d'attente des autres passagers.

2.5.3.3 Contraintes

Les contraintes du modèle sont les suivantes :

- Pas de départ en avance :

$$\forall s \in \mathcal{S}, \forall t \in \mathcal{T}_s, \quad d_{t,s} \geq \bar{d}_{t,s} \quad (2.22)$$

- Respect des temps de trajet et de stationnement minimums :

$$\forall s \in \mathcal{S}, \forall t \in \mathcal{T}_s, \quad \bar{d}_{t,s_{suiv_t}(s)} \geq \bar{d}_{t,s} + \tau_{s,s_{suiv_t}(s)}^{run} + \tau_{s_{suiv_t}(s)}^{stop} \quad (2.23)$$

- L'intervalle entre deux trains doit être supérieur au block :

$$\forall s \in \mathcal{S}, \forall t \in \mathcal{T}_s, \quad \bar{d}_{s_{suiv_s}(t),s} - \bar{d}_{t,s} \geq \tau^{head} \quad \text{et} \quad d_{s_{suiv_s}(t),s} - d_{t,s} \geq \tau^{head} \quad (2.24)$$

- La durée de stationnement est une variable aléatoire de loi connue :

$$\forall s \in \mathcal{S}, \forall t \in \mathcal{T}_s, \quad d_{t,s} - a_{t,s} = DT_{t,s} \quad (2.25)$$

- La durée de trajet intergare est une variable aléatoire de loi connue :

$$\forall s \in \mathcal{S}, \forall t \in \mathcal{T}_s, \quad a_{t,s_{suiv_t}(s)} - d_{t,s} = RT_{t,s,s_{suiv_t}(s)} \quad (2.26)$$

- Conservation du flot des passagers :

$$\forall s \in S, \forall t \in \mathcal{T}_s, \quad p_{t,s}^{stay} = p_{t,prec_s(t)}^{stay} - \sum_{s' \in D_{t,s}^-} p_{t,s',s} + \sum_{s' \in D_{t,prec_s(t)}^+ \setminus \{s\}} p_{t,prec_s(t),s'} \quad (2.27)$$

- Le nombre de passagers attendant de réaliser un trajet est égal à la somme de ceux arrivés depuis le passage du dernier train effectuant ce trajet, et de ceux ayant été rejetés par ce dernier train :

$$\forall s \in S, \forall t \in \mathcal{T}_s, \forall s' \in D_{t,s}^+, \quad p_{s,s'}(d_{t,s}) = \frac{(d_{t,s} - d_{prec_t(s,s'),s})}{H} p_{f_{s,s',d_{prec_t(s,s'),s},d_{t,s}}} + \hat{p}_{prec_t(s,s'),s,s'} \quad (2.28)$$

- Le nombre de passagers montant dans un train est contraint par la capacité du train :

$$\forall s \in S, \forall t \in \mathcal{T}_s, \forall s' \in D_{t,s}^+, \quad p_{t,s,s'} = \begin{cases} p_{s,s'}(d_{t,s}) & \text{si } p_{t,s}^{stay} + \sum_{s'' \in D_{t,s}^+} p_{s,s''}(d_{t,s}) \leq c_t \\ \left\lfloor \frac{p_{s,s'}(d_{t,s})}{\sum_{s'' \in D_{t,s}^+} p_{s,s''}(d_{t,s})} (c_t - p_{t,s}^{stay}) \right\rfloor & \text{sinon} \end{cases} \quad (2.29)$$

où $\lfloor \cdot \rfloor$ désigne la partie entière. Le premier cas correspond à la situation où le train a une capacité suffisante pour accueillir tous les voyageurs. Le second cas correspond à la situation de saturation : la capacité restante est allouée, pour chaque groupe se rendant à une destination donnée, proportionnellement à la taille du groupe.

- Le nombre de passagers refusés est égal à la demande moins le nombre de passagers montés à bord :

$$\forall s \in S, \forall t \in \mathcal{T}_s, \forall s' \in D_{t,s}^+, \quad \hat{p}_{t,s,s'} = p_{s,s'}(d_{t,s}) - p_{t,s,s'} \quad (2.30)$$

2.5.3.4 Résolution

La résolution du programme stochastique précédent permettrait d'obtenir une grille optimale du point de vue de la qualité de service offerte aux voyageurs face aux petits aléas d'exploitation. Cependant, les distributions de probabilité des variables aléatoires décrivant les durées de parcours et de stationnement ne sont pas discrètes, et ne sont pas connues de façon exactes.

Une résolution exacte de ce programme stochastique s'avère par conséquent délicate. La démarche la plus proche consisterait en un échantillonnage de scénarios à partir d'estimations des distributions de probabilité, sur le même principe que celui suivi par Kroon et al. (2008). Cependant, ce procédé a pour inconvénient de faire fortement augmenter la taille du modèle, ce qui complique sa résolution (d'autant plus lorsque le problème est non linéaire comme c'est le cas ici). En outre, cette approche suppose de connaître en amont les distributions des variables aléatoires $DT_{t,s}$ et $RT_{t,s,s'}$, ce qui ne permet pas de tenir compte des mécanismes d'interdépendance entre ces variables qui ont été décrits au chapitre 1. C'est pourquoi nous avons plutôt choisi de développer une approche permettant de pallier ces inconvénients, couplant simulation stochastique et optimisation. La structure de l'algorithme de résolution développé est décrite par la figure 2.7.

Le chapitre 3 est consacré à la modélisation des petits aléas de l'exploitation, portant sur les durées de stationnement et de parcours intergares. L'interdépendance entre ces aléas est en particulier explicitée par l'estimation de distribution de probabilités conditionnelles.

Ces distributions sont utilisées au chapitre 4, au sein d'un algorithme de simulation stochastique. Celui-ci permet de reproduire l'influence des aléas sur le trafic ferroviaire en zone dense, et d'en

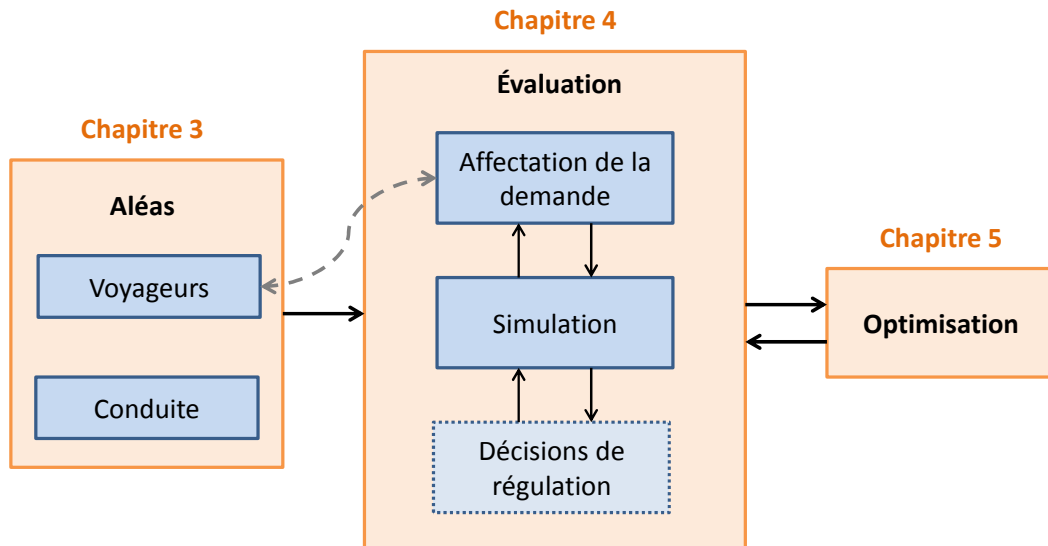


Figure 2.7 – Architecture de l’algorithme de résolution du programme stochastique

déduire le calcul d’indicateurs adaptés. Parmi ceux-ci, nous calculerons en particulier l’espérance du temps de trajet généralisé qui constitue la fonction objectif du modèle. Pour ce faire, les voyageurs sont introduits de façon explicite dans le modèle; ceci permet à la fois de tenir compte de leur influence sur les durées de stationnement, et de les affecter aux différents trains de façon à estimer la qualité du service qui leur a été fourni. L’algorithme de simulation permet en outre de reproduire (dans une certaine mesure) les décisions de régulation prises par les GTI lorsque les conditions de trafic sont dégradées. Ce module de régulation peut, au choix, être activé ou non.

Enfin, le chapitre 5 est consacré à une méthode heuristique d’optimisation pour la résolution du programme stochastique. L’algorithme de simulation stochastique du chapitre 4 y est utilisé comme fonction d’évaluation.

Chapitre 3

Modélisation des petits aléas de l'exploitation

Résumé

Nous présentons ici la démarche mise en œuvre pour une meilleure compréhension des aléas à l'origine de la dégradation des conditions de trafic. Celle-ci s'appuie sur l'estimation des distributions de probabilité des variables aléatoires intervenant dans le modèle du chapitre précédent. La plus grande partie de ce chapitre est consacrée à l'étude des durées de stationnement en gare et de l'influence des flux de voyageurs sur celles-ci. Nous montrons que ces flux peuvent être décrits par une seule variable, et qu'une durée minimale de stationnement est nécessaire pour un flux de voyageurs donné. Nous proposons une méthode d'estimation de cette durée minimale. Nous quantifions ensuite la probabilité d'en dévier, en calculant les distributions de probabilité conditionnelles des durées de stationnement connaissant les flux de voyageurs. Dans une seconde partie, plus courte, nous appliquons une démarche proche pour l'étude des variabilités des temps de parcours.

Sommaire

3.1	Modélisation des variabilités des durées de stationnement	64
3.1.1	Présentation et motivation	64
3.1.2	État de l'art	65
3.1.3	Collecte des données et pré-traitement	68
3.1.4	Définition et calcul de la durée de stationnement minimale	71
3.1.5	Distributions conditionnelles des durées de stationnement	77
3.1.6	Annexe : résultats sur d'autres jeux de données	81
3.2	Modélisation des variabilités des temps de parcours	82
3.2.1	Présentation et état de l'art	82
3.2.2	Estimation des distributions conditionnelles de temps de parcours	83

« Le hasard est une loi qui voyage incognito. »
Proverbe arabe

3.1 Modélisation des variabilités des durées de stationnement

3.1.1 Présentation et motivation

La construction d'une grille horaire requiert la définition, pour chaque train, de temps de parcours théoriques entre les gares et de durées de stationnement théoriques dans chaque gare desservie. La définition de ces dernières est particulièrement délicate. Les durées de stationnement des trains en gare dépendent en effet de plusieurs facteurs :

- Les caractéristiques techniques du matériel roulant, notamment le temps requis pour l'ouverture et la fermeture des portes.
- La durée du processus de descente et montée des voyageurs. Celle-ci dépend du nombre de voyageurs descendant et montant, de la charge du train et du quai, ainsi que de la configuration longitudinale du train et de la gare (Daamen, Lee et Wiggendaad, 2008). Le comportement des passagers joue également : il arrive fréquemment que des passagers bloquent des portes pour monter ou laisser monter d'autres voyageurs, ou tout simplement parce que la rame est surchargée, ce qui empêche les portes de se fermer. Tout ceci a pour effet d'allonger la durée de stationnement.
- La grille horaire. Les ADC ne sont en effet pas autorisés à faire partir leur train d'une gare en avance sur l'horaire.
- La signalisation. La plupart des gares d'Île de France possèdent des signaux en sortie de quai qui marquent l'entrée dans un nouveau canton. Si les trains se succèdent avec une fréquence réduite, il peut arriver qu'un train voie son stationnement en gare prolongé jusqu'à ce que le canton suivant soit libéré.
- Les décisions de l'ADC. Certains ont pour habitude de fermer les portes et de partir dès que la montée des voyageuses est terminée et que la grille horaire le permet, tandis que d'autres attendent parfois quelques secondes de plus pour permettre à d'éventuels voyageurs en retard de monter.

La multiplicité de ces facteurs et leur imprévisibilité rend difficile l'estimation de la durée de stationnement. Celle-ci revêt cependant une importance significative. En effet, les durées d'arrêt en gare dimensionnent la capacité du réseau (Abril et al., 2008) : les surestimer lors de la construction de la grille mène donc à une utilisation sous-optimale de l'infrastructure. Cela doit être évité lorsque la demande est forte et le réseau proche de la saturation, ce qui est le cas en Île de France. Ces durées ne doivent pas non plus être sous-estimées : cela risquerait de provoquer des retards par rapport à l'horaire et de faire sortir le système de sa zone de stabilité (voir le chapitre 1 et Van Breusegem, Campion et Bastin (1991)), ce qui aurait *in fine* pour effet de réduire le débit de la ligne. Une estimation précise des durées de stationnement peut également se révéler utile durant la phase opérationnelle : cela permettrait en effet de faciliter les prises de décision des régulateurs, ainsi que de fournir aux voyageurs des informations fiables sur leur temps d'attente et de trajet.

Nous présentons dans cette section une méthodologie et un modèle permettant de quantifier l'influence des flux de passagers sur les durées de stationnement. Il est à noter que le temps nécessaire à l'échange des voyageurs en gare est difficile à estimer précisément. En plus des facteurs listés plus haut, il dépend de paramètres qui ne sont pas facilement mesurables : la distribution des passagères sur le quai, le nombre de voyageurs transportant des bagages encombrants tels qu'un vélo ou une poussette... Cependant, nous démontrons que pour un flux de voyageurs donné, un temps minimal de stationnement est requis pour permettre à l'échange des voyageurs de se réaliser. Un algorithme pour le calcul effectif de ce temps minimal sera présenté. Tous les autres aléas (passagers transportant des bagages ou bloquant la porte, ADC tardant à donner le départ...) ne peuvent que prolonger le temps de stationnement au-delà de cette valeur minimale.

La figure 3.1 présente les différentes composantes susceptibles de déterminer la durée de stationnement, dans le cas d'un train arrivant légèrement en retard en gare. Dans le cas représenté, le

train aurait pu repartir à l'heure si l'échange de voyageurs n'avait nécessité que la durée minimale. Cependant, une déviation due à des phénomènes aléatoires a ici prolongé la durée de stationnement au-delà de cette durée minimale. Nous présentons dans la suite de cette section une méthodologie permettant de calculer la durée minimale de stationnement ainsi que d'évaluer la probabilité de dévier de cette valeur.

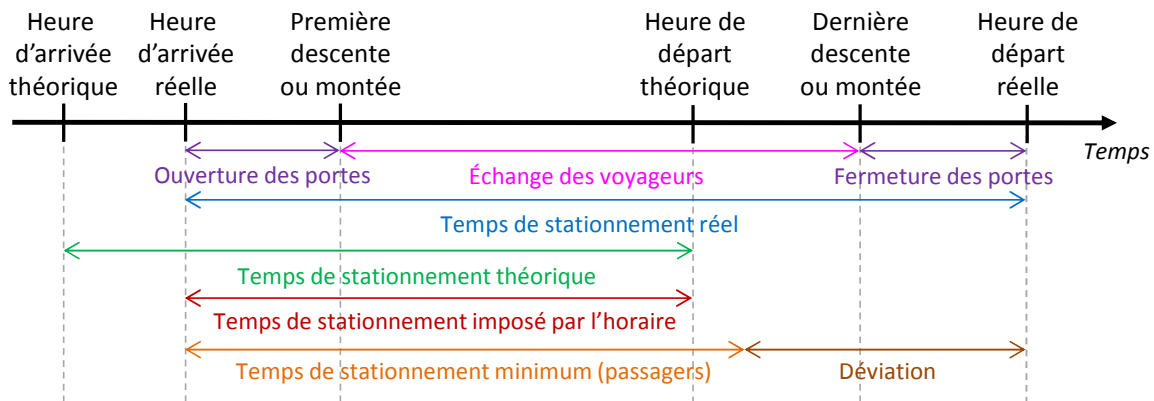


Figure 3.1 – Composantes de la durée de stationnement

3.1.2 État de l'art

La descente et la montée de passagers figurent parmi les principaux déterminants de la durée de stationnement dans les transports en commun (Kraft, 1975). Cependant, tous les modèles du stationnement existant dans la littérature ne font pas intervenir de description explicite de ce processus. C'est pourquoi les précédents travaux portant sur l'estimation des durées de stationnement peuvent être classés en deux catégories : les modèles explicites, qui nécessitent des données d'entrée décrivant les flux de voyageurs et leur comportement ; et les modèles implicites, qui utilisent d'autres données que des comptages de voyageurs, telles que le jour de la semaine, l'heure, le retard ou l'intervalle temporel avec le train précédent. Tous ces paramètres sont néanmoins corrélés avec les flux de voyageurs.

La plupart des modèles explicites reposent sur une stratégie de modélisation consistant à construire un modèle à partir de données (à l'aide de méthodes de régression), ou bien à simuler les mouvements de passagers. Lin et Wilson (1992) furent parmi les pionniers à proposer des modèles relevant de la première stratégie. Leurs modèles, linéaires et non-linéaires, tiennent compte du phénomène de friction qui se produit lorsqu'il y a un grand nombre de voyageurs debout à bord des véhicules. Ces modèles furent appliqués à la ligne verte du réseau MBTA (métro de Boston), pour des trains à une ou deux voitures. Plusieurs améliorations de leur travail furent ensuite proposées par Puong (2000), permettant notamment une meilleure compréhension des effets de la congestion ou de l'influence du nombre de portes.

Le développement des solutions de localisation automatique des véhicules ainsi que de comptage automatique des passagers dans les années 2000 a ouvert de nouvelles perspectives pour l'estimation des durées de stationnement. En effet, ces systèmes permettent de disposer de données plus précises et en plus grand nombre qu'en ayant recours à des comptages manuels. La mise à disposition de tels jeux de données ont permis de nombreux travaux, principalement appliqués aux cas des bus et du métro.

Par exemple, Dueker et al. (2004) ont calculé des statistiques descriptives sur les durées de stationnement et proposé un modèle d'estimation simple pour le réseau de bus de la ville de Portland. Sun et al. (2014) ont approfondi l'analyse du processus de montée et descente à bord des bus urbains,

et exhibé un taux d'occupation critique au-delà duquel le processus est ralenti en raison de la friction entre les passagers. Cet article met également l'accent sur l'influence de la configuration intérieure du matériel roulant, les mêmes analyses ayant été effectuées sur plusieurs types de véhicules présentant des caractéristiques différentes.

Buchmüller, Weidmann et Nash (2008) ont choisi une autre stratégie : plutôt que de construire un modèle déterministe, les auteurs ont divisé le processus de stationnement en plusieurs sous-processus et estimé les distributions de probabilité des durées de chacun de ceux étant soumis à des fluctuations aléatoires. Ceci permet de prendre en compte la stochasticité intrinsèque du phénomène, tous les passagers n'ayant pas le même comportement. Un choix similaire est fait par Longo et Medeossi (2012) et Larsen et al. (2014), qui utilisent des distributions aléatoires des durées de stationnement pour la simulation du trafic ferroviaire.

Plus récemment, D'Acerno et al. (2017) ont proposé un modèle analytique pour les durées de stationnement sur une ligne de métro circulaire. Il est démontré dans l'article que sans intervention extérieure, le trafic converge vers un équilibre dans lequel les durées de stationnement peuvent être calculées comme solutions d'un problème de point fixe en dimension finie. Enfin, Christoforou, Chandakas et Kaparias (2020) ont tenté d'expliquer les durées de stationnement de tramways à l'aide d'une régression linéaire utilisant des données sur les flux de voyageurs. Si les auteurs constatent que ces seuls flux constituent des prédicteurs imparfaits, l'originalité de leur travail réside surtout en une seconde régression visant à expliquer les situations où de fortes variations des durées de stationnement sont observées.

Tous ces modèles paramétriques, une fois calibrés, permettent de disposer d'une expression « fermée » pour la durée de stationnement qui permet de la calculer facilement. À l'inverse, les méthodes de simulation microscopique requièrent un temps de calcul plus long pour l'estimation des durées de stationnement, mais offrent en revanche des informations sur les mouvements des passagers et leur comportement. Ainsi, Zhang, Han et Li (2008) ont conçu un outil de simulation par automates cellulaires qui permet de reproduire les phénomènes de congestion et de négociation entre passagers à l'échelle microscopique. Cet outil fut utilisé pour approximer les durées de stationnement sur le réseau de métro de Pékin. Yamamura et al. (2013) ont développé un modèle multi-agents pour simuler le comportement des passagers, et l'ont utilisé pour évaluer l'efficacité de certaines mesures susceptibles d'être prises pour atténuer les phénomènes de congestion et ainsi réduire les durées d'arrêt en station, telles que l'utilisation de trains avec des portes plus larges ou moins de places assises. Une étude proche fut conduite par Schelenz et al. (2014) dans le but de déterminer la configuration intérieure optimale d'un bus urbain. Une thèse récente menée au sein de SNCF Réseau a permis à son autrice d'étudier les champs d'application de tels modèles multi-agents, et notamment leur transférabilité d'une gare à une autre (Elleuch, 2019). Seriani et Fernandez (2015) ont quant à eux utilisé un simulateur microscopique du trafic piéton pour évaluer la performance de différentes stratégies de gestion des flux de voyageurs sur le réseau du métro de Santiago, et validé ces simulations par des expériences en laboratoire.

Cependant, les modèles explicites ne sont pas toujours aisés à mettre en œuvre. En effet, tous les systèmes de transport en commun ne sont pas encore dotés de solutions de comptage automatique des passagers, et les données relatives à ceux-ci peuvent par conséquent se révéler difficiles à recueillir. En outre, les systèmes équipés de tels outils de comptage ne fournissent pas toujours les données collectées en temps réel, ce qui complique l'utilisation des modèles explicites lors de la phase opérationnelle. Pour toutes ces raisons, des modèles prenant en compte les flux de passagers de façon implicite ont également été développés. Ceux-ci nécessitent des données qui sont habituellement plus faciles à recueillir, tout en étant corrélées aux flux de voyageurs. Ainsi, Hansen, Goverde et Meer (2010) utilisent les retards des trains à l'arrivée comme donnée d'entrée, et une méthode de régression robuste pour estimer les durées de stationnement des trains intercity aux Pays-Bas. Kecman et Goverde (2015) ont appliqué les méthodes de régression dites des Moindres Carrés Tronqués (*Least Trimmed Squares*) et des forêts aléatoires pour prédire les durées de sta-

tionnement à partir d'un jeu de données contenant le type de train (local ou intercity), le retard à l'arrivée, la taille de la gare ainsi que l'heure de la journée. Des modèles locaux spécifiques à certaines gares sont également proposés. Les temps de stationnement y sont estimés à l'aide d'un calcul de moyenne mobile des durées des stationnements précédents ; ces modèles locaux ont permis d'obtenir de meilleures performances de prédiction. Li, Daamen et Goverde (2016) prédisent également la durée de stationnement d'un train donné à certaines gares spécifiques en utilisant des informations sur les durées de stationnement du même train aux gares précédentes ainsi que des trains précédents à la même gare. Des améliorations de cette approche ainsi qu'un test de généralité sont présentés dans l'article de Li, Yin et He (2018). De façon similaire, Xin et Chen (2016) mettent en œuvre la méthode des k -plus proches voisins pour prédire la durée d'arrêt d'un bus à un arrêt donné à partir de ses durées de stationnement aux arrêts précédents.

Le concept de durée de stationnement minimale est introduit par Pedersen, Nygreen et Lindfeldt (2018). Celle-ci est définie comme le temps de stationnement le plus court (tenant compte du temps requis pour l'ouverture et la fermeture des portes) permettant la descente et la montée de tous les passagers. Les auteurs font l'hypothèse que cette durée minimale peut être calculée en ne considérant que des trains en retard, et étudient l'influence de divers facteurs temporels (heure de la journée, jour de la semaine, semaine dans l'année) sur la distribution de cette durée minimale. Cependant, la dépendance explicite entre la durée de stationnement minimale et les flux de passagers n'est pas étudiée par les auteurs.

Le tableau 3.1 résume, dans l'ordre chronologique, certaines caractéristiques des principaux modèles de la littérature. Comme indiqué précédemment, ces modèles peuvent faire intervenir les flux de passagers de façon explicite ou implicite ; cela est indiqué dans la deuxième colonne du tableau.

Source	Flux de passagers	Mode de transport	Modélisation
Lin et Wilson (1992)	Explicite	Métro	RP
Puong (2000)	Explicite	Métro	RP
Dueker et al. (2004)	Explicite	Bus	RP
Buchmüller, Weidmann et Nash (2008)	Explicite	Train	Di
Zhang, Han et Li (2008)	Explicite	Métro	Sim
Hansen, Goverde et Meer (2010)	Implicite	Train	RP
Yamamura et al. (2013)	Explicite	Train	Sim
Sun et al. (2014)	Explicite	Bus	RP
Li et al. (2014)	Implicite	Train	Di
Seriani et Fernandez (2015)	Explicite	Métro	Sim
Kecman et Goverde (2015)	Implicite	Train	RP et RNP
Xin et Chen (2016)	Implicite	Bus	RNP
Li, Daamen et Goverde (2016)	Implicite	Train	RP et RNP
D'Acerno et al. (2017)	Explicite	Métro	An
Pedersen, Nygreen et Lindfeldt (2018)	Implicite	Train	Di
Cornet et al. (2019)	Explicite	Train	RNP et Di
Christoforou, Chandakas et Kaparias (2020)	Explicite	Tramway	RP

Di : distributions de probabilité, RP : régression paramétrique, RNP : régression non paramétrique, An : analytique, Sim : simulation microscopique

Table 3.1 – Quelques caractéristiques des principaux modèles de durée de stationnement de la littérature

À notre connaissance, la stochasticité intrinsèque du processus de stationnement a fait l'objet de relativement peu de recherches jusqu'ici ; Buchmüller, Weidmann et Nash (2008) et Li et al. (2014) faisant partie des rares auteurs ayant tenté de la modéliser. Nous avons cherché à combler ce vide, en concevant une méthode permettant de calculer la durée de stationnement minimale requise pour permettre la descente et la montée en fonction du nombre de passagers, ainsi que d'estimer la probabilité de dévier de cette durée minimale.

3.1.3 Collecte des données et pré-traitement

3.1.3.1 Protocole expérimental et données disponibles

Pour cette étude, nous avons développé un protocole expérimental que nous avons appliqué à deux gares du réseau Transilien, les gares de Bois-Colombes (BC) sur la ligne J et la gare de Houilles-Carières-sur-Seine (HC) sur la ligne L. Nous avons choisi ces deux gares car elles sont soumises, en alternance, à des flux passagers importants en heure de pointe et très faibles en heure creuse. Nous présentons également des résultats sur d'autres gares du réseau Transilien à la fin de cette section. Un plan partiel des lignes J et L est donné sur la figure 3.2 ; celui-ci indique les différentes gares, les dessertes proposées ainsi que la fréquence correspondante en heure de pointe.

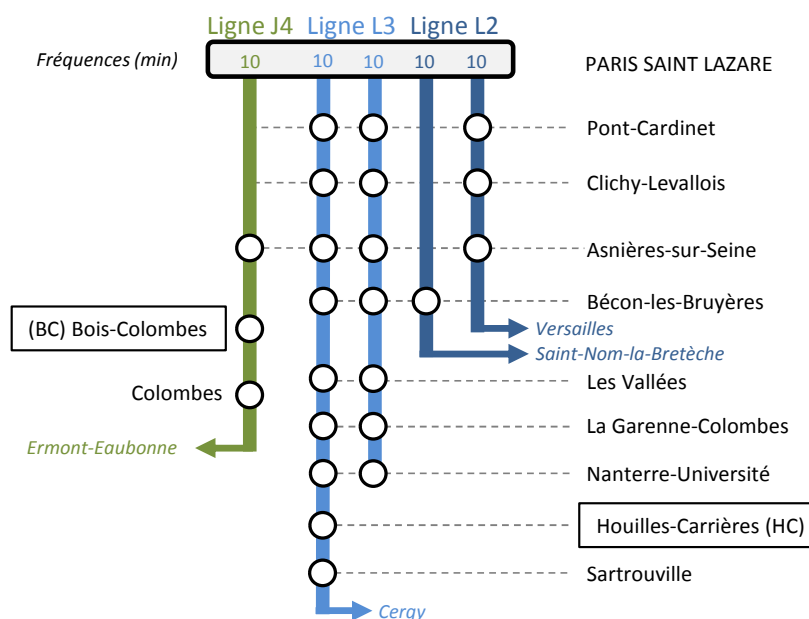


Figure 3.2 – Plan partiel des dessertes des lignes J et L

Chacune des lignes L2, L3 et J4 circule sur une double voie dédiée. L'essentiel des missions sur ces lignes sont assurées par des unités simples ou doubles de rames Z50000, aussi appelées NAT (Nouvelles Automotrices Transilien) ou rames « Francilien ». Chaque rame mesure 94.3m de long, possède 7 portes d'une largeur de 1.95m, une capacité assise de 380 passagers et une capacité totale de 760 passagers. Ce matériel est équipé d'un système de comptage automatique fonctionnant à l'aide de rayons infrarouges. Des capteurs sont situés au-dessus de chaque porte, et sont capables de détecter séparément les mouvements de descente et de montée (dont on peut déduire la charge du train). Lorsqu'un train est extrêmement chargé, certains passagers peuvent avoir besoin de descendre du train pour permettre le passage d'autres voyageurs, avant de remonter. Ces voyageurs sont donc comptés à la fois comme descendant et montant du train. Cependant, les taux d'occupation observés dans les jeux de données suggèrent que ce phénomène se produit rarement, les NAT étant conçues de façon à permettre une circulation fluide à l'intérieur de la rame.

Des comparaisons avec des résultats de comptages manuels ont montré que l'erreur moyenne de mesure du nombre de passagers descendants et montants est inférieure à 5%, tandis que la charge du train après départ est connue avec moins de 15% d'erreur. La précision de ce système a par conséquent été jugée suffisante par Île de France Mobilités pour le calcul de la ponctualité voyageurs sur la ligne H, en lieu et place des comptages manuels. Une cassette ATESS est également présente à bord des rames NAT, permettant de mesurer les horaires d'arrivée et de départ des trains dans les différentes gares avec une précision supérieure à celle des balises de voie BREHAT.

Les archives brutes de l'année 2017 ont été pré-traitées de façon à construire un jeu de données pour chaque gare et chaque sens de circulation (de Paris vers la banlieue - ou sens impair - et de la banlieue vers Paris - ou sens pair). Chaque entrée de ce jeu de données décrit le flux de passagers et le processus de stationnement à l'aide des variables suivantes :

- L'horaire théorique (arrivée et départ).
- Les heures d'arrivée et de départ réelles (dont on peut déduire la durée de stationnement réelle).
- Le nombre de passagers descendant ou montant. Ces valeurs sont agrégées à l'échelle du train, et non pas détaillées porte par porte.
- La charge du train après le départ.
- Le nombre de portes (7 s'il n'y a qu'une rame en unité simple, 14 s'il y a deux rames couplées).

Les données ont été filtrées de façon à ne conserver uniquement des entrées pour lesquelles les durées de stationnement sont inférieures à 180 secondes. En effet, nous nous intéressons ici exclusivement à la modélisation des petits aléas, et 99.9% des observations correspondent à un temps de stationnement compris entre 0 et 180 secondes. Nous supposons que les durées de stationnement supérieures sont dûes à des incidents ne relevant pas du périmètre de ce travail. En outre, les passagers étant généralement répartis sur le quai, la durée de l'échange voyageurs est aussi influencée par le nombre de portes (ou de façon équivalente, le nombre d'unités constituant la rame). Nous avons par conséquent divisé les nombres de passagers descendant et montant ainsi que la charge à bord par le nombre de rames (1 ou 2, selon que le train est constitué d'une unité simple ou d'une unité multiple). Toutes les analyses statistiques effectuées sur ce jeu de données ont été réalisées à l'aide du package Python `scikit-learn` (Pedregosa et al., 2011).

Nous noterons désormais A le nombre de voyageurs descendant, B le nombre de voyageurs montant et L la charge du train après départ. La durée de stationnement est notée DT (*Dwell Time*). Quelques statistiques descriptives des jeux de données utilisés sont présentées dans le tableau 3.2. Remarquons que toutes les variables A, B, L présentent un écart-type important. Cela est dû à la forte variation des flux de voyageurs entre les heures de pointe et les heures creuses.

Bois-Colombes - sens impair 23865 observations					Bois-Colombes - sens pair 24230 observations				
	A	B	L	DT		A	B	L	DT
Médiane	6.0	30.5	136.5	46.0	Médiane	28.0	6.0	107.0	46.0
Moyenne	7.4	36.0	155.7	55.8	Moyenne	37.9	7.7	135.5	47.6
Écart-type	6.8	31.6	120.7	29.4	Écart-type	31.9	7.0	99.0	15.8

Houilles-Carières - sens impair 7989 observations					Houilles-Carières - sens pair 7812 observations				
	A	B	L	DT		A	B	L	DT
Médiane	5.0	18.0	61.0	34.0	Médiane	18.0	6.5	56.0	32.0
Moyenne	9.0	29.7	111.6	36.8	Moyenne	26.2	10.1	84.5	36.9
Écart-type	10.1	32.1	132.1	14.0	Écart-type	24.5	11.0	84.0	16.6

Table 3.2 – Statistiques descriptives des jeux de données décrivant les durées de stationnement

3.1.3.2 Description des flux de passagers par une unique variable latente

Le tableau 3.3 présente les matrices de corrélation entre les variables A, B, L sur les différents jeux de données. Il n'est pas surprenant de constater que ces trois variables sont fortement corrélées. En effet, toutes trois sont sujettes à des variations régies par le même phénomène : les heures de pointe et les heures creuses. Dans le sens pair, il y a un nombre important de voyageurs montants et une charge élevée en heure de pointe en semaine, et des valeurs plus faibles le reste de la journée. De

même, dans le sens impair, le nombre de voyageurs descendant et à bord est grand en heure de pointe en semaine, et moins élevé le reste du temps. Une observation similaire est faite par Christoforou, Chandakas et Kaparias (2020) sur le tramway de Nantes, à l'aide de cartes de chaleur. En raison de cette multi-colinéarité, aucune méthode de régression n'est en mesure de distinguer l'influence des volumes de voyageurs descendant, montant et à bord sur les durées de stationnement (Dormann et al., 2012).

Bois-Colombes - sens impair

	A	B	L
A	1	0.86	0.87
B	0.86	1	0.96
L	0.87	0.96	1

Bois-Colombes - sens pair

	A	B	L
A	1	0.74	0.93
B	0.74	1	0.78
L	0.93	0.78	1

Houilles-Carières - sens impair

	A	B	L
A	1	0.44	0.55
B	0.44	1	0.87
L	0.55	0.87	1

Houilles-Carières - sens pair

	A	B	L
A	1	0.49	0.86
B	0.49	1	0.68
L	0.86	0.68	1

Table 3.3 – Matrices de corrélation entre les variables A, B, L

Nous résolvons ce problème de multi-colinéarité en effectuant une Analyse en Composantes Principales (Jolliffe, 1986) sur les variables explicatives A, B, L . Cet algorithme permet de construire de nouvelles variables p_1, p_2, p_3 , appelées **composantes principales**, comme combinaisons linéaires des variables A, B, L . Ces nouvelles variables sont décorrélatées et construites de façon à avoir une variance maximale vis-à-vis du jeu de données. Notons que l'amplitude de variation des paramètres A, B, L diffère : il se situe habituellement entre 0 et 100 pour A et B , entre 0 et 1000 pour L . Nous avons donc, préalablement à cette analyse en composantes principales, normalisé les données en soustrayant la moyenne et en divisant par l'écart-type. Ceci permet d'éviter qu'un poids artificiellement trop important ne soit donné au paramètre L lors de la construction des composantes principales. Le tableau 3.4 indique la proportion de variance expliquée par chacune des composantes principales p_1, p_2, p_3 .

	BC - impair	BC - pair	HC - impair	HC - pair
p_1	0.93	0.88	0.75	0.79
p_2	0.06	0.10	0.21	0.17
p_3	0.01	0.02	0.04	0.04

Table 3.4 – Ratio de variance expliquée par chacune des composantes principales

Remarquons que la première composante p_1 explique à elle seule une large part de la variance. On observe le même phénomène sur les autres jeux de données sur lesquels nous avons appliqué cet algorithme, voir la fin de la section. C'est pourquoi nous l'appellerons désormais **la** composante principale, ou encore le **flux réduit de passagers**, et la noterons p . Cette variable sera utilisée comme unique représentation des flux de voyageurs dans ce qui suit, les autres composantes p_2, p_3 sont éliminées du modèle. Pour faciliter la compréhension, nous avons ré-échelonné le flux réduit p par une transformation affine de façon à ce qu'il prenne ses valeurs dans l'intervalle $[0, 1]$. Ainsi, $p = 0$ correspond à un train circulant presque vide, avec un échange de voyageurs quasi-nul entre le train et le quai. Inversement, $p = 1$ correspond à un train extrêmement chargé, avec un nombre très important de voyageurs descendant et/ou montant. Il est à noter que les coefficients permettant d'obtenir p comme combinaison linéaire de A, B, L sont spécifiques à chaque jeu de données (donc à chaque gare et chaque sens de circulation). Les droites de régression donnant A, B, L en fonction

de p sont tracées à la figure 3.3 pour chaque jeu de données. Celles-ci permettent de reconstituer le flux de voyageurs à partir du flux réduit. Par exemple, à la gare de Houilles-Carières dans le sens pair Paris - banlieue, une valeur de $p = 0.4$ correspond à environ $A = 100$ voyageurs descendant, $B = 30$ voyageurs montant et $L = 350$ voyageurs à bord après le départ.

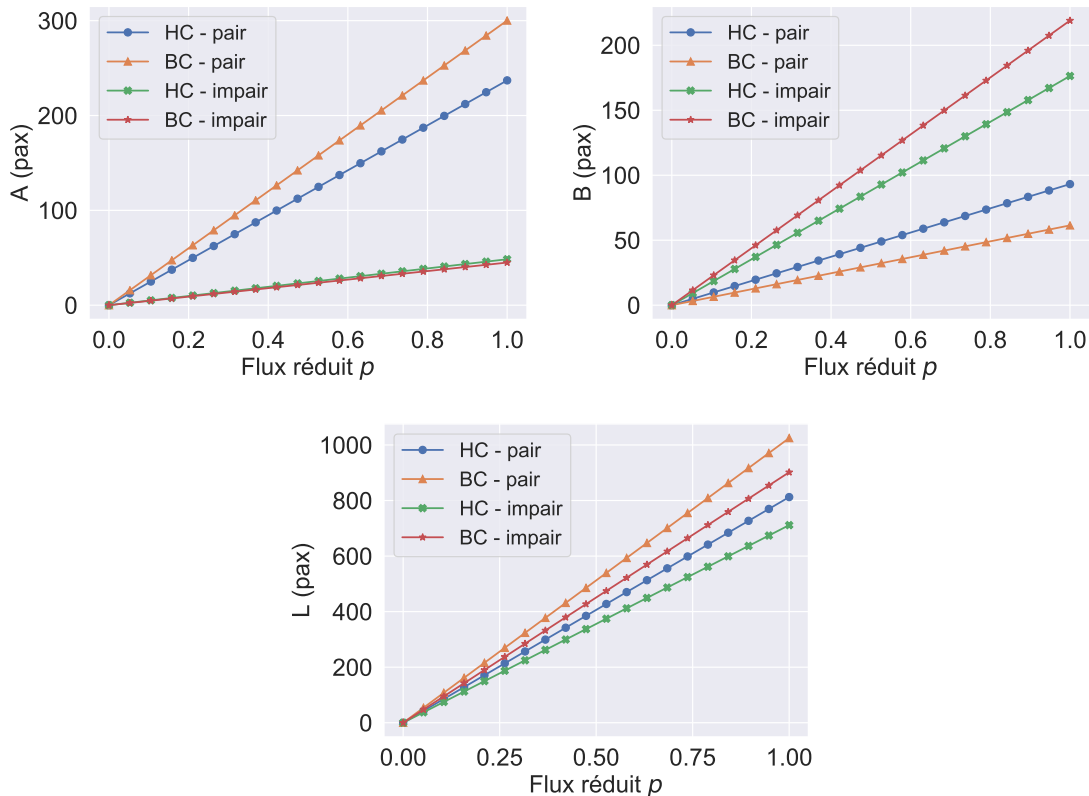


Figure 3.3 – Correspondance entre les valeurs de p et celles de A, B, L

Notons finalement que les données sont réparties de façon hétérogène selon l'axe p . Un grand nombre d'observations correspondent à des situations où le flux de passagers est faible (et donc p proche de 0), tandis que les observations correspondant à des valeurs très grandes de p (proches de 1) sont rares. Voir la figure 3.4 au paragraphe suivant.

3.1.4 Définition et calcul de la durée de stationnement minimale

3.1.4.1 Introduction

La figure 3.4 représente les jeux de données de la gare de Bois-Colombes (BC-impair et BC-pair) par des nuages de points, où le flux réduit p est en abscisse et le temps de stationnement DT en ordonnée. Les points sont colorés en fonction de la densité des observations dans leur voisinage.

Outre l'inhomogénéité mentionnée au paragraphe précédent, cette figure permet de constater la stochasticité intrinsèque du temps de stationnement. En effet, on remarque qu'un grand nombre d'observations correspondent à des durées de stationnement longues malgré un flux de voyageurs faible. Il est probable qu'un certain nombre de ces stationnements prolongés soient dûs à des facteurs difficilement prévisibles, tels qu'une agglutination de passagers sur une des portes du train, ou un échange de voyageurs ralenti par la présence de passagers transportant des bagages encombrants. C'est pourquoi nous pensons que les durées de stationnement aux gares intermédiaires en zone dense ne peuvent être prédites avec précision en utilisant les données actuellement disponibles. En effet, nous avons tenté de séparer les jeux de données en une base d'entraînement et une base de test, et

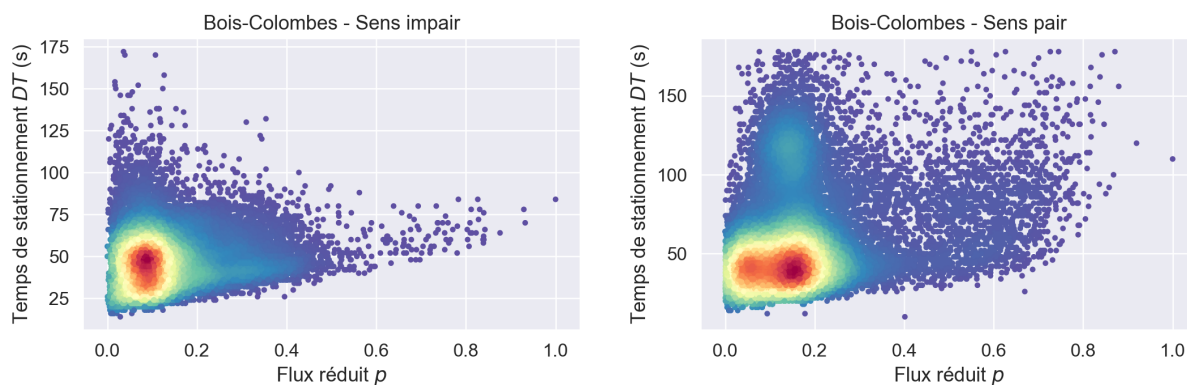


Figure 3.4 – Durées de stationnement pour différentes valeurs du flux réduit

d'utiliser les données d'entraînement pour construire un modèle de prédiction. Toutes les tentatives ont donné des résultats insatisfaisants lorsque le modèle de prédiction a été appliqué à la base de test.

En revanche, on constate que la borne inférieure de la durée de stationnement DT croît lorsque le flux de voyageurs p augmente. Cela semble suggérer que, pour une valeur donnée du flux réduit p , une durée de stationnement minimale doit impérativement être respectée pour permettre à l'échange de voyageurs d'aller à son terme. Nous porterons désormais une attention particulière à cette notion de durée de stationnement minimale, vue comme une fonction du flux réduit p . Les paragraphes suivants sont consacrés à la définition rigoureuse de cette notion, ainsi qu'à la présentation d'un algorithme permettant le calcul effectif de cette durée minimale.

3.1.4.2 Définition

À notre connaissance, le concept de durée de stationnement minimale n'a jusqu'ici été exploré que par Pedersen, Nygreen et Lindfeldt (2018). Leur travail repose sur l'hypothèse selon laquelle un train en retard ne stationnera en gare que pendant la durée minimale permettant à l'échange de voyageurs de se terminer. Leurs analyses sont donc conduites sur un échantillon de données composé exclusivement de trains en retard. Cette approche présente cependant des inconvénients et peut être améliorée. Tout d'abord, des observations faites en gare semblent montrer que tous les trains en retard ne partent pas dès que la descente et la montée des voyageurs sont terminées. Certains sont en effet retenus en gare par des signaux d'avertissement ou de sémaphore en sortie de gare, et des secondes supplémentaires peuvent également être perdues en raison de voyageurs bloquant les portes ou du choix de l'ADC d'attendre des passagers en retard. En outre, une telle définition de la durée de stationnement minimale complique l'étude de la façon dont celle-ci dépend des flux de voyageurs. Or, la variabilité de ceux-ci est la principale cause des variations temporelles de cette durée minimale observées par les auteurs.

Nous choisissons de formaliser cette notion de durée de stationnement minimale et de modéliser l'observation faite à la figure 3.4 en introduisant les variables aléatoires suivantes :

- DT qui représente la durée de stationnement réelle
- $Delay$ qui représente le retard à l'arrivée en gare
- \widehat{DT} qui correspond à la durée de stationnement théorique (déterministe) prévue par la grille horaire
- DTT qui représente la durée de stationnement minimale de stationnement requise pour satisfaire la contrainte imposant de ne pas quitter la gare en avance. Celle-ci est définie par

$$DTT = \max(\widehat{DT} - Delay, 0) \quad (3.1)$$

- DTP qui représente la durée de stationnement nécessaire à l'échange de voyageurs.

Les variables aléatoires DT , DTT et DTP vérifient la relation

$$DT \geq \max(DTT, DTP) \quad (3.2)$$

Nous supposons que la variable aléatoire DTP peut se décomposer de la façon suivante :

$$DTP = MDT(p) + Dev \quad (3.3)$$

où MDT (*Minimum Dwell Time*) est une fonction déterministe croissante du flux réduit p , et Dev est une variable aléatoire positive rendant compte de la stochasticité de l'échange de voyageurs. Ces variables aléatoires ont déjà été représentées sur la figure 3.1 ; nous la reproduisons à la figure 3.5 avec ces nouvelles notations. Rappelons que les relations utilisées pour le calcul de p , et par conséquent aussi de MDT , sont locales et spécifiques à chaque gare et sens de circulation.

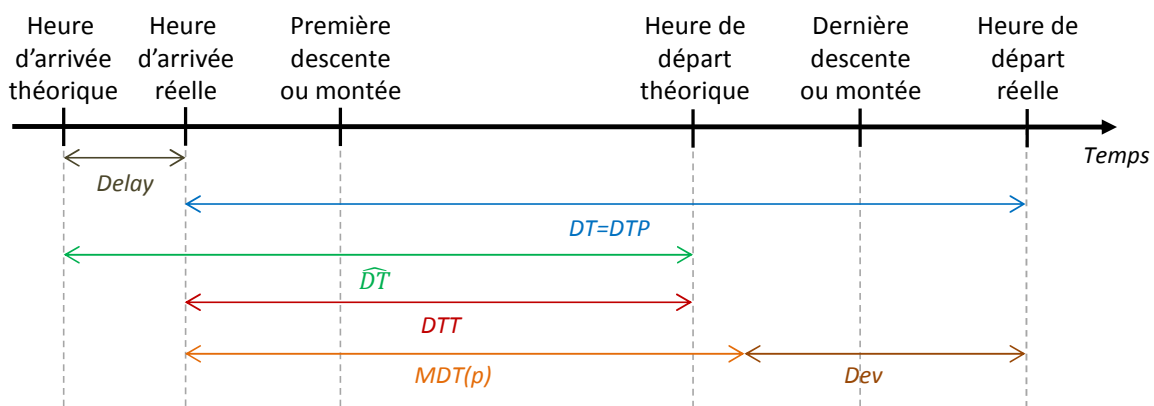


Figure 3.5 – Notations des composantes de la durée de stationnement

Un modèle proche de celui-ci a été proposé dans la thèse de Medeossi (2008), dans le but de réaliser des simulations stochastiques du trafic ferroviaire. Toutefois, dans ces travaux, la durée de stationnement minimale est uniquement déterminée par les caractéristiques du matériel roulant et n'est pas influencée par les variations de flux de voyageurs.

Remarquons que si la fonction MDT dépend de la seule variable p , elle est néanmoins influencée par des facteurs temporels (par exemple) à travers p . En effet, en semaine et en heure de pointe, les flux de passagers sont plus importants que le reste du temps. Les valeurs de p et de MDT sont donc également plus importantes. Nous présentons un algorithme de calcul de cette fonction MDT dans le paragraphe suivant.

3.1.4.3 Algorithme de calcul de la durée de stationnement minimale

Nous proposons un algorithme en deux étapes pour le calcul de la fonction MDT à partir du jeu de données représenté sur la figure 3.4. Nous cherchons tout d'abord à sélectionner les observations pour lesquelles $DT \simeq DTP$ et $DTP \simeq MDT(p)$ (donc $Dev \simeq 0$). Seules les valeurs des variables aléatoires DT et DTT sont cependant connues. Nous procédons à la sélection des observations en suivant les étapes ci-dessous :

1. Choisir une largeur de fenêtre Δp .
2. Sélectionner toutes les observations satisfaisant $p \in [0, \Delta p]$.
3. Ne conserver que l'observation ayant la plus petite valeur de DT parmi toutes les observations sélectionnées satisfaisant $DTT < DT$. Éliminer toutes les autres observations.

4. Répéter le processus sur les intervalles $[\Delta p, 2\Delta p], \dots, [n\Delta p, 1]$ jusqu'à ce que tout l'intervalle $[0, 1]$ (et donc tout le jeu de données) ait été parcouru.

La fenêtre Δp doit être choisie de manière telle que le nombre de points conservés soit suffisant, mais pas trop important de façon à s'assurer que $DT \simeq MDT(p)$ pour les observations retenues. Nous discutons le choix de cet hyperparamètre dans le paragraphe suivant.

Les observations restantes correspondent donc à des points dont les coordonnées peuvent s'écrire $(p, MDT(p))$. Il reste à reconstituer la fonction MDT à partir de ces points. Plusieurs méthodes de régression peuvent être envisagées pour cela. Nous avons choisi la méthode des k -plus proches voisins (Altman, 1992) pour sa simplicité et sa robustesse. Celle-ci consiste, pour obtenir la valeur de $MDT(p)$ en un point p donné, à sélectionner les k points les plus proches sur l'axe p et à calculer la moyenne de leurs valeurs de MDT . Nous terminons en modifiant légèrement la fonction obtenue. La fonction MDT peut en effet être raisonnablement supposée croissante. Nous avons par conséquent remplacé chaque section décroissante par une section constante.

3.1.4.4 Choix des hyperparamètres et validation sur des cas tests fictifs

La fonction MDT n'est pas connue sur les jeux de données issus des comptages automatiques et des cassettes ATESS. Ceux-ci ne peuvent donc pas être utilisés pour la calibration des paramètres et l'évaluation des performances de la méthode précédente. Nous avons donc constitué deux jeux de données fictifs sur lesquels la fonction MDT est connue, ceux-ci étant notés Test1 et Test2. Nous supposons que dans ces deux cas test, la fonction MDT est donnée respectivement par

$$MDT_1(p) = 20 + 60p \quad \text{et} \quad MDT_2(p) = 20 + 60p^2 \quad (3.4)$$

Nous avons constitué ces jeux de données en sélectionnant les observations du jeu de données de Bois-Colombes dans le sens pair (banlieue - Paris). Les valeurs de MDT ont été calculées à l'aide des équations 3.4, puis celles de DT ont été générées en utilisant les relations 3.1, 3.2, 3.3 (l'inégalité étant remplacée par une égalité dans la relation 3.2). La variable aléatoire Dev a été choisie suivant une loi exponentielle de paramètre 20 : cette loi donne une distribution proche de celle observée dans les jeux de données. Nous avons ajouté un bruit suivant une loi normale $\mathcal{N}(0, 1)$ afin que des observations soient réparties de part et d'autre de la courbe de la fonction MDT . La méthode devant présenter un minimum de robustesse, des *outliers*, représentant 0.01% du nombre total d'observations, ont été ajoutés suivant une loi uniforme dans l'espace $[0, 1] \times [0, 180]$. Les jeux de données fictifs ainsi constitués sont représentés sur la figure 3.6.

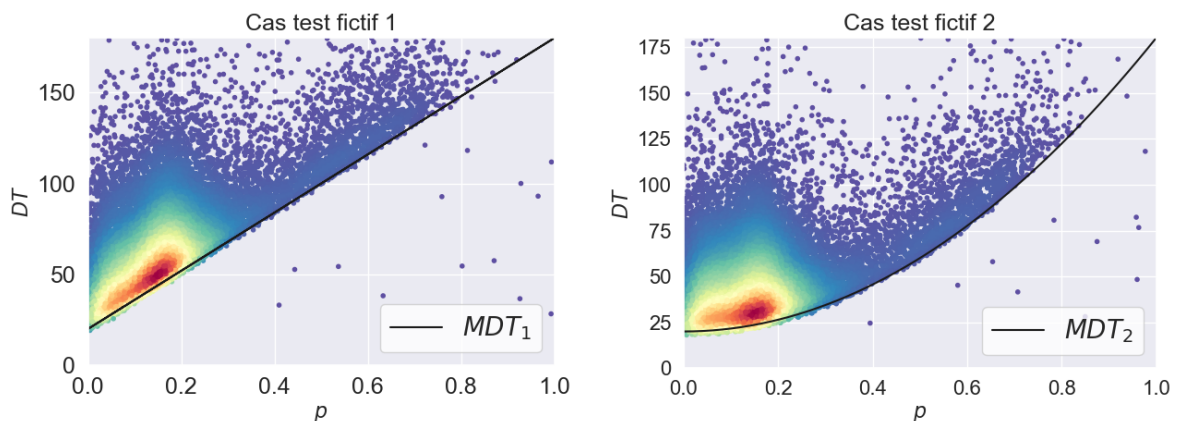


Figure 3.6 – Cas test fictifs

Nous avons utilisé Test2 pour le choix des hyperparamètres Δp (la largeur de la fenêtre de balayage pour l'étape de sélection) et k (le nombre de voisins pour l'étape de régression). Le jeu

de données a été divisé aléatoirement en une base d'entraînement et une base de test (contenant chacune 50% des observations). Le modèle a alors été entraîné sur la première base et utilisé pour estimer les valeurs de MDT sur la deuxième.

Nous avons évalué la qualité de l'estimation à l'aide du critère d'erreur absolue moyenne, ou MAE (*Mean Absolute Error*) :

$$MAE = \frac{1}{n} \sum_{i=1}^n \left| MDT_2(p_i) - \widetilde{MDT}_2(p_i) \right| \quad (3.5)$$

où n est le nombre d'observations de la base de test, p_1, \dots, p_n une numérotation des observations, MDT_2 la fonction exacte et \widetilde{MDT}_2 la fonction estimée. Nous nous sommes limités à l'intervalle $[0, 0.8]$: en effet, l'intervalle $[0.8, 1]$ contient trop peu de points pour permettre une estimation satisfaisante (soit avec une précision de l'ordre de quelques secondes). La figure 3.7 représente, via une palette de couleurs, l'erreur commise pour différentes combinaisons d'hyperparamètres.

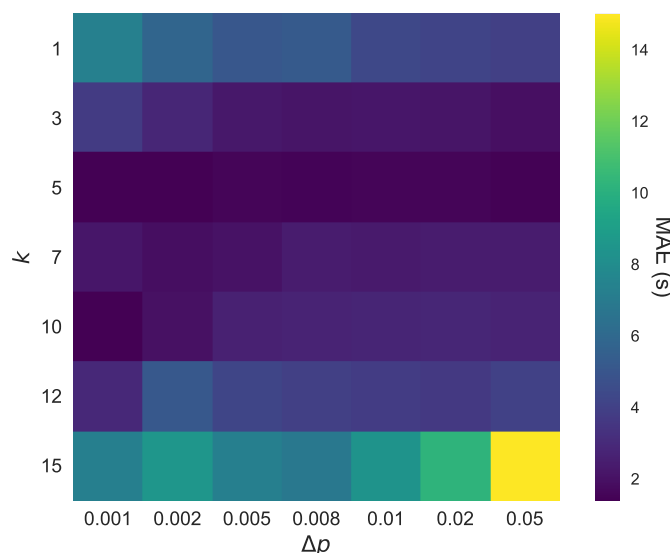


Figure 3.7 – Valeur de MAE pour différentes combinaisons d'hyperparamètres

On remarque que plusieurs combinaisons de paramètres offrent des résultats satisfaisants (MAE inférieur à 2 secondes). Nous avons choisi $p = 0.005$ et $k = 5$: en effet, ce jeu d'hyperparamètres offre de bonnes performances, mais ses voisins également ce qui garantit une certaine robustesse. D'autres choix auraient cependant été tout aussi valables.

Nous avons alors appliqué l'algorithme avec ce choix d'hyperparamètres aux jeux de données `Test1` et `Test2`. Les points sélectionnés et l'allure des fonctions estimées sont représentés à la figure 3.8.

L'évolution de l'erreur d'estimation (définie comme la différence entre la valeur estimée et la valeur exacte) avec p est représentée sur la figure 3.9. On constate que la fonction MDT est estimée avec une bonne précision pour $p \in [0, 0.8]$, l'erreur étant inférieure à 5 secondes, parfois très significativement. Cependant, sur l'intervalle $p \in [0.8, 1]$, la précision de l'estimation se dégrade nettement : cela est dû au faible nombre de points dans cet intervalle. Ceux-ci correspondent à des situations très perturbées qui ne se produisent que peu fréquemment. En effet, l'intervalle $[0, 0.8]$ contient 99.9% des observations. Nous pouvons donc conclure que l'algorithme permet d'estimer précisément la fonction MDT sous réserve de disposer de données en quantité suffisante.

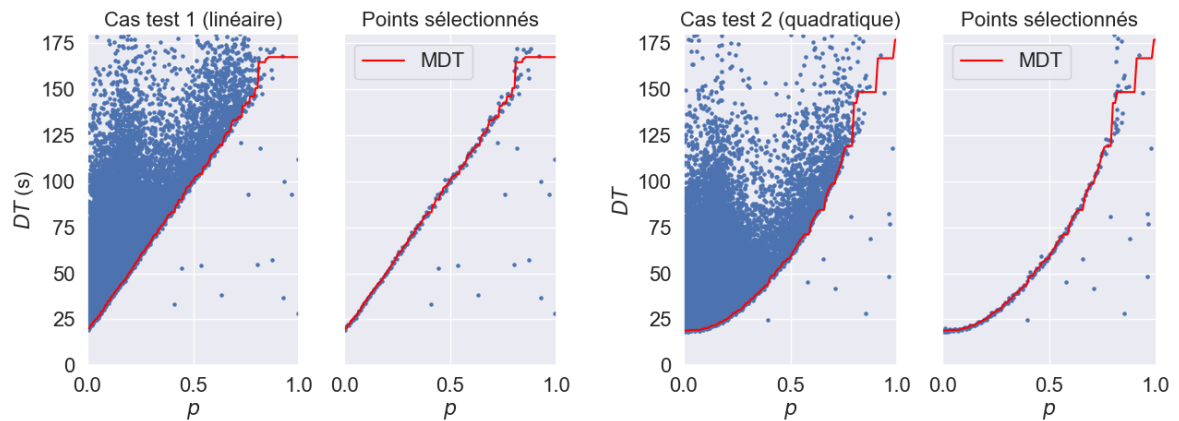


Figure 3.8 – Estimation de MDT sur les cas tests fictifs

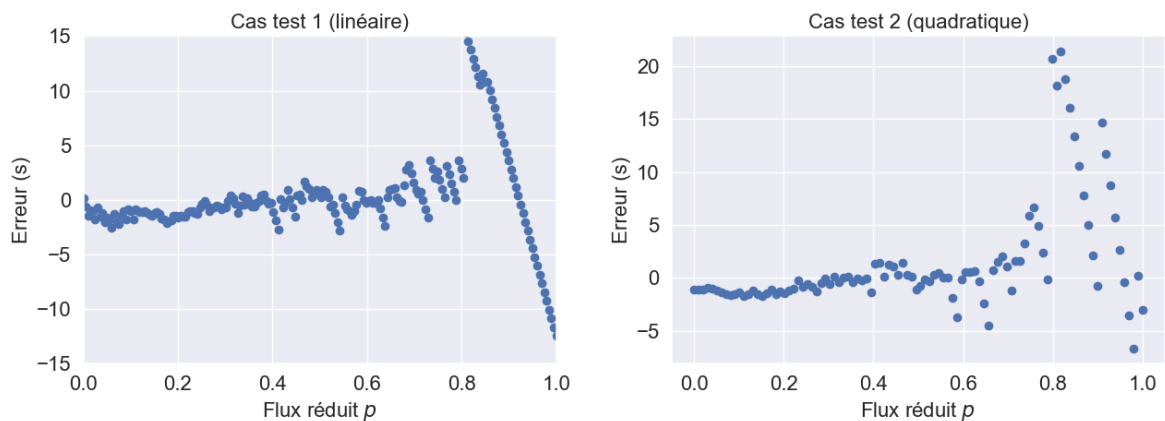


Figure 3.9 – Erreur d'estimation de MDT sur les cas tests fictifs

3.1.4.5 Application à des données réelles

Nous avons alors appliqué l'algorithme aux jeux de données des gares de Bois-Colombes et Houilles-Carières, dans les sens pair et impair, avec les hyperparamètres déterminés précédemment. La figure 3.10 représente les points sélectionnés et l'allure de la fonction estimée.

On constate que pour de petites valeurs de p , la fonction MDT croît de façon approximativement linéaire et avec une pente relativement faible. Cependant, pour les valeurs de p plus élevées, la pente s'accroît : cela traduit le phénomène de congestion qui se produit alors. Remarquons que l'intensité de ce phénomène varie selon les cas d'étude.

La connaissance de la durée de stationnement minimale peut être utilisée, couplée avec des outils de prévision de la demande à long terme (Ceder, 2007), pour réaliser des études de capacité de l'infrastructure. Elle peut aussi être utile, utilisée concomitamment avec une solution de comptage des voyageurs à quai en temps réel (Wei et Chen, 2012) pour garder la maîtrise des durées de stationnement. Cela rend possible, par exemple, la mise en place d'un compte à rebours en cabine de conduite communiquant à l'ADC une durée de stationnement raisonnable pour permettre à l'échange de voyageurs de se réaliser. Cela peut se révéler particulièrement utile dans le cas des lignes de métro automatiques, pour lesquelles la durée de stationnement est généralement prédéterminée et indépendante des flux de voyageurs.

Toutefois, dans une perspective d'amélioration de la robustesse des grilles horaires, la seule connaissance de la durée de stationnement minimale n'est pas suffisante. En effet, celle-ci n'est que rarement réalisée en pratique. Il est par conséquent nécessaire de disposer d'informations sur

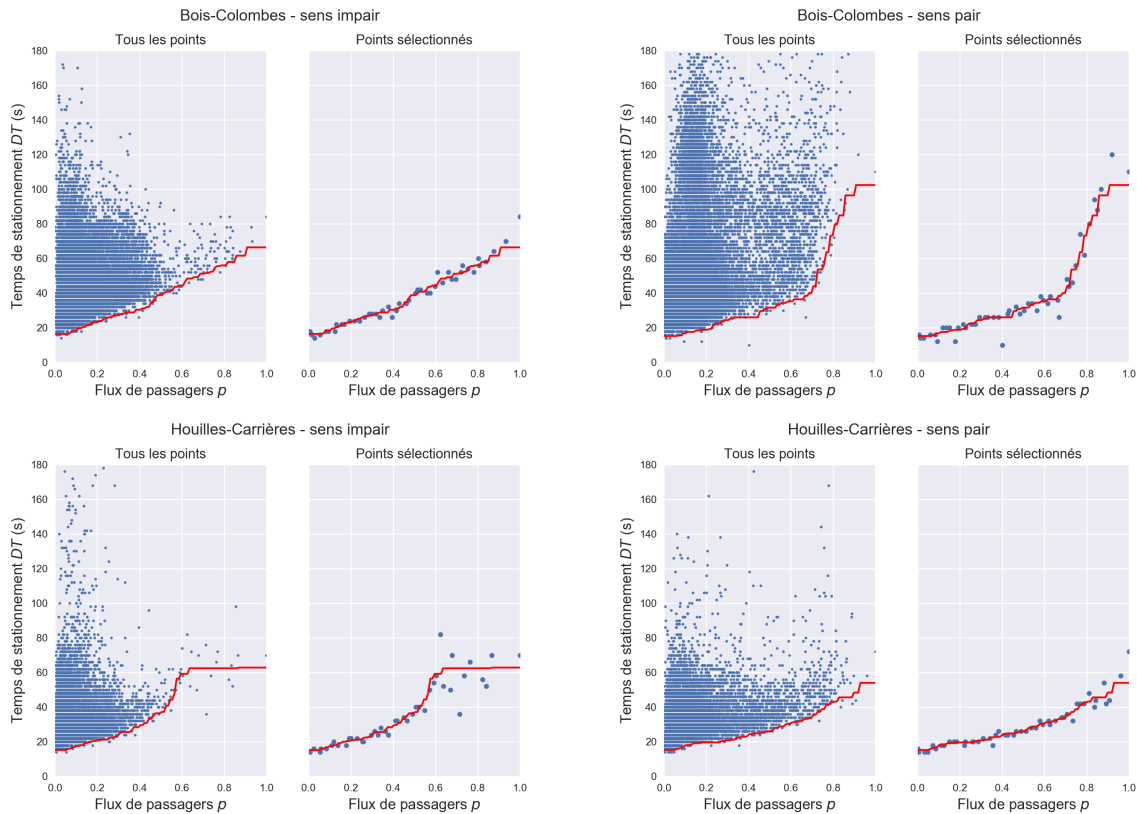


Figure 3.10 – Estimation de MDT sur les gares de Bois-Colombes et Houilles-Carières

les éventuels allongements de la durée de stationnement au-delà de la valeur minimale dûs à des phénomènes stochastiques. Le modèle présenté au chapitre 2 nécessite en outre de connaître les distributions de probabilité des durées de stationnement. Le paragraphe suivant présente un algorithme permettant le calcul des distributions conditionnelles, connaissant le flux de voyageurs. Ceci permet de quantifier l'influence des volumes de passagers sur l'allongement de la durée de stationnement.

3.1.5 Distributions conditionnelles des durées de stationnement

3.1.5.1 Sélection des observations pertinentes

Rappelons que la durée de stationnement des trains en gare n'est pas déterminée uniquement par celle de l'échange de voyageurs, mais également par la grille horaire (les trains ne peuvent partir en avance) ainsi que par des événements extérieurs (de signalisation par exemple). Ces derniers sont difficilement détectables dans les données disponibles. En revanche, la connaissance de la durée de stationnement minimale permet de déterminer des situations où la durée de l'échange voyageurs a été prépondérante sur la durée de stationnement imposée par la grille horaire.

En effet, comme $DTP \geq MDT(p)$ et $DT \geq \max(DTT, DTP)$, les cas où la durée minimale $MDT(p)$ est supérieure à DTT correspondent à des situations où $DTP \geq DTT$ et donc où la grille horaire n'a pas été déterminante pour la durée du stationnement. Or, ce sont ces observations qui nous intéressent pour l'étude de la façon dont les flux de voyageurs influencent la durée du stationnement. Nous avons donc calculé la valeur de MDT pour toutes les observations des jeux de données, à l'aide de la méthode précédente, et comparé celle-ci à la valeur de DTT . Les nuages de points correspondant aux observations relevant des deux situations possibles, $MDT > DTT$ et $MDT \leq DTT$, sont donnés à la figure 3.11 pour la gare de Bois-Colombes dans le sens pair.

7558 observations relèvent de la situation où $MDT > DTT$ (donc où l'échange de voyageurs a été prépondérant sur la grille horaire) et 16306 observations relèvent de la situation inverse, où la

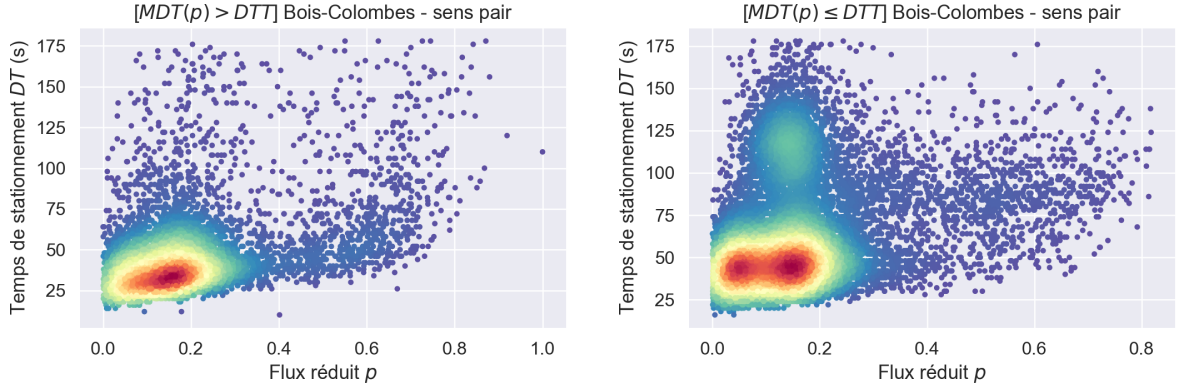


Figure 3.11 – Classification des observations selon le facteur prépondérant

grille horaire peut avoir été prépondérante. Des proportions similaires ont été observées sur les autres jeux de données. On remarque qu'un grand nombre d'observations ayant des valeurs de DT élevées pour des flux de voyageurs p faibles relèvent du second cas. La longue durée de stationnement est donc très probablement due à une arrivée en gare en avance.

Dans la suite, nous ne retenons que les observations satisfaisant $MDT > DTT$.

3.1.5.2 Estimation de densité par une méthode à noyau

Les étapes précédentes nous ont permis de disposer d'un jeu de données pour lesquelles il est raisonnable de supposer que la durée de l'échange de voyageurs a déterminé le temps de stationnement. Nous avons utilisé ces données pour estimer les distributions de probabilité des durées de stationnement en gare, conditionnellement aux flux de voyageurs. Ce choix de calculer des distributions conditionnelles permet d'exploiter l'information disponible sur les flux de voyageurs pour obtenir des estimations plus précises.

Nous supposons désormais que le flux réduit est une variable aléatoire, notée P . Nous faisons également l'hypothèse que les variables aléatoires P et DT ont des densités, respectivement notées f_P et f_{DT} , et que le couple (P, DT) admet la densité $f_{P,DT}$. La distribution de probabilité conditionnelle de DT connaissant la valeur de P peut alors s'écrire

$$\forall t \geq 0, \forall p \in [0, 1], \quad \mathbb{P}(DT \leq t | P = p) = \int_0^t \frac{f_{P,DT}(p, \tau)}{f_P(p)} d\tau \quad (3.6)$$

L'estimation des densités f_P et $f_{P,DT}$ permet ainsi d'obtenir celle de la distribution conditionnelle de DT sachant P . Nous avons recouru à la méthode d'estimation de densité à noyau gaussien (*Kernel Density Estimation*, ou KDE - voir Silvermann (1986)) pour estimer la fonction $f_{P,DT}$. Le principe de la méthode consiste à approximer la fonction $f_{P,DT}$ par une somme de densités gaussiennes centrées en chaque observation du jeu de données.

Plus précisément, considérons le jeu de données comme la réalisation de n variables aléatoires à valeurs dans $[0, 1] \times \mathbb{R}_+$ (la première composante donnant la valeur de P et la seconde celle de DT). Ces variables sont notées X_1, \dots, X_n . On note également $x = (p, t)$ un point de $[0, 1] \times \mathbb{R}_+$. L'estimateur à noyau gaussien de $f_{P,DT}$ est alors défini par

$$\forall x \in [0, 1] \times \mathbb{R}_+, \quad \hat{f}(x) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (3.7)$$

où h est un paramètre de lissage et K est la fonction gaussienne à deux variables définie par

$$\forall u \in \mathbb{R}^2, \quad K(u) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}u^T u\right) \quad (3.8)$$

L'estimateur \hat{f} dépend de X_1, \dots, X_n et peut donc être considéré comme une variable aléatoire. La qualité de l'estimation est mesurée par la moyenne de l'intégrale du carré de l'erreur, ou *Mean Integrated Square Error (MISE)* :

$$MISE(\hat{f}) = \mathbb{E} \left(\int_{[0,1] \times \mathbb{R}_+} (\hat{f}(x) - f_{P,DT}(x))^2 dx \right) \quad (3.9)$$

Silvermann (1986) a démontré que $h = 0.96n^{-1/6}$ est un choix optimal si la densité à estimer est la densité de la loi normale. Nous avons choisi cette valeur de h même si l'hypothèse de normalité n'est pas satisfaite a priori.

Une estimation de f_P peut être déduite de celle de $f_{P,DT}$ via la relation

$$\forall p \in [0, 1], \quad f_P(p) = \int_0^{+\infty} f_{P,DT}(p, t) dt \quad (3.10)$$

La durée de stationnement étant bornée à 180 secondes dans les jeux de données, les intégrales sur \mathbb{R}_+ sont en fait des intégrales sur des segments qui peuvent être calculées numériquement par la méthode des trapèzes (Stoer et Bulirsch, 1993).

Nous avons alors appliqué cette méthodologie de calcul aux jeux de données à disposition, après les avoir normalisés pour éviter les effets dûs aux différences d'échelle. Quelques fonctions de répartition conditionnelles obtenues sont représentées à la figure 3.12.

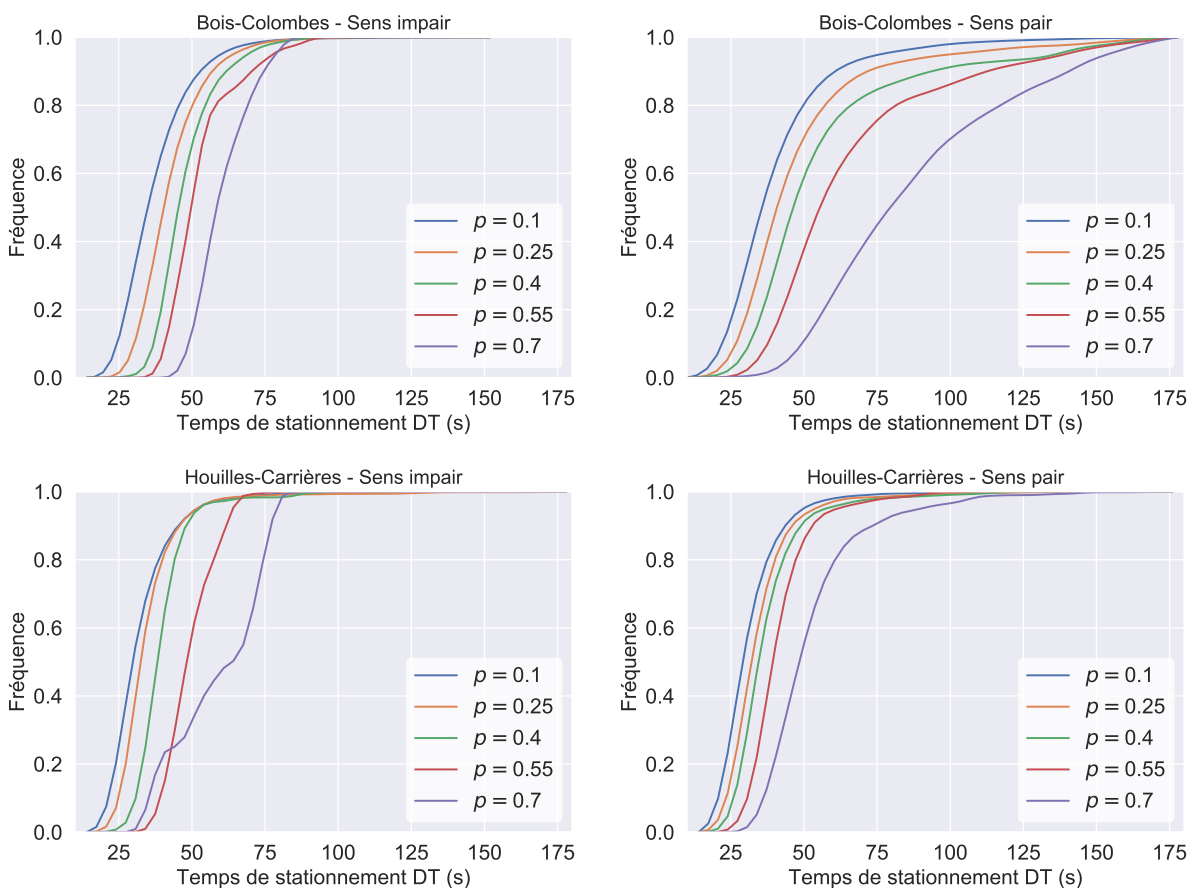


Figure 3.12 – Fonctions de répartition conditionnelles de DT connaissant P pour les gares de Bois-Colombes et Houilles-Carières

Celles-ci correspondent aux fonctions $t \mapsto \mathbb{P}(DT \leq t | P = p)$, pour différentes valeurs de p . On peut par exemple lire sur celles-ci qu'à la gare de Bois-Colombes, dans le sens pair, pour $p = 0.55$

(donc pour environ $B = 120$ voyageurs montants, $A = 30$ voyageurs descendants et $L = 530$ voyageurs à bord, voir figure 3.3), la durée de stationnement est inférieure ou égale à 60 secondes dans 60% des cas.

On observe que ces courbes se décalent vers la droite lorsque la valeur de p augmente; cela traduit le fait que la durée de stationnement nécessaire à l'échange de voyageurs augmente avec le flux. On observe également que ces courbes croissent moins fortement pour les grandes valeurs de p : cela indique un risque de congestion plus important lorsque les flux voyageurs sont élevés.

Enfin, nous avons cherché à vérifier la précision de l'estimation de ces distributions. Pour ce faire, nous avons divisé aléatoirement les jeux de données en un ensemble d'entraînement et un ensemble de test, de tailles égales. Les observations de l'ensemble d'entraînement ont été utilisées pour calculer les distributions conditionnelles de DT sachant P avec la méthode décrite précédemment. Pour différentes valeurs de $\bar{p} \in [0, 1]$, nous avons sélectionné les observations de l'ensemble de test satisfaisant $p \in [\bar{p} - 0.01, \bar{p} + 0.01]$. Nous avons ensuite calculé les déciles des observations sélectionnées ainsi que ceux des distributions estimées avec l'ensemble d'entraînement. Les deux distributions ont été comparées en représentant ces déciles sur un diagramme quantile-quantile, ou QQ-plot (Chambers et al., 2018), voir la figure 3.13.

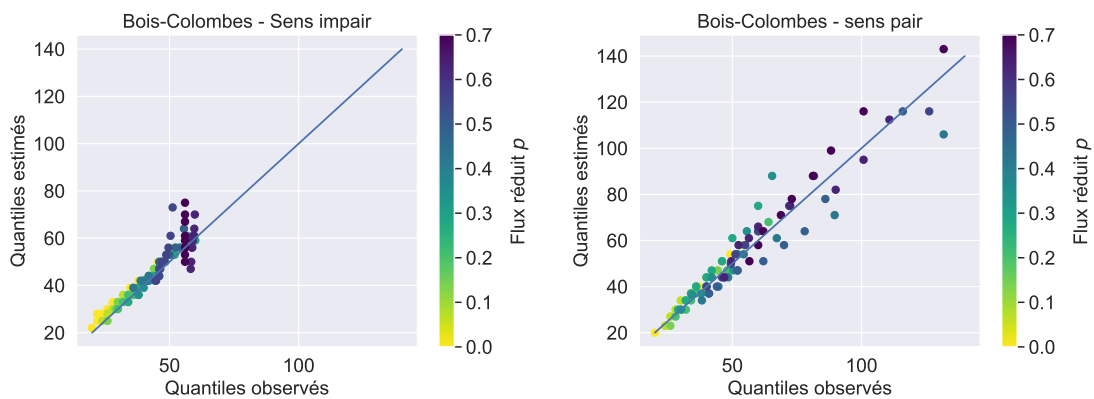


Figure 3.13 – Diagrammes quantile-quantile des distributions conditionnelles de DT sachant P

La précision de l'estimation est mesurée sur ce diagramme par la distance entre les points et la première bissectrice, sur lesquels tous les points devraient se trouver si l'estimation était parfaite. On observe que la distribution conditionnelle de DT connaissant le flux réduit p est estimée avec une précision satisfaisante pour les faibles valeurs de p . En revanche, les estimations sont moins précises pour les grandes valeurs de p , en particulier les déciles supérieurs. Cela vient du fait, déjà mentionné, que le nombre d'observations correspondant à des flux voyageurs très importants est relativement faible.

3.1.5.3 Applications

Outre la simulation stochastique du trafic que nous présentons au chapitre 4, plusieurs applications de cette méthodologie peuvent être envisagées. Lors des phases stratégique et tactique, la connaissance des distributions des durées de stationnement peuvent permettre une estimation probabiliste de la capacité d'une infrastructure, que ce soit une ligne ou une grande gare. Elles peuvent également être utilisées en temps réel pour effectuer des prévisions de retards; celles-ci peuvent ainsi faciliter la prise de décision des gestionnaires du trafic, ainsi que permettre de fournir une information plus fiable aux voyageurs (sous réserve de pouvoir prédire les flux voyageurs sur un horizon relativement court).

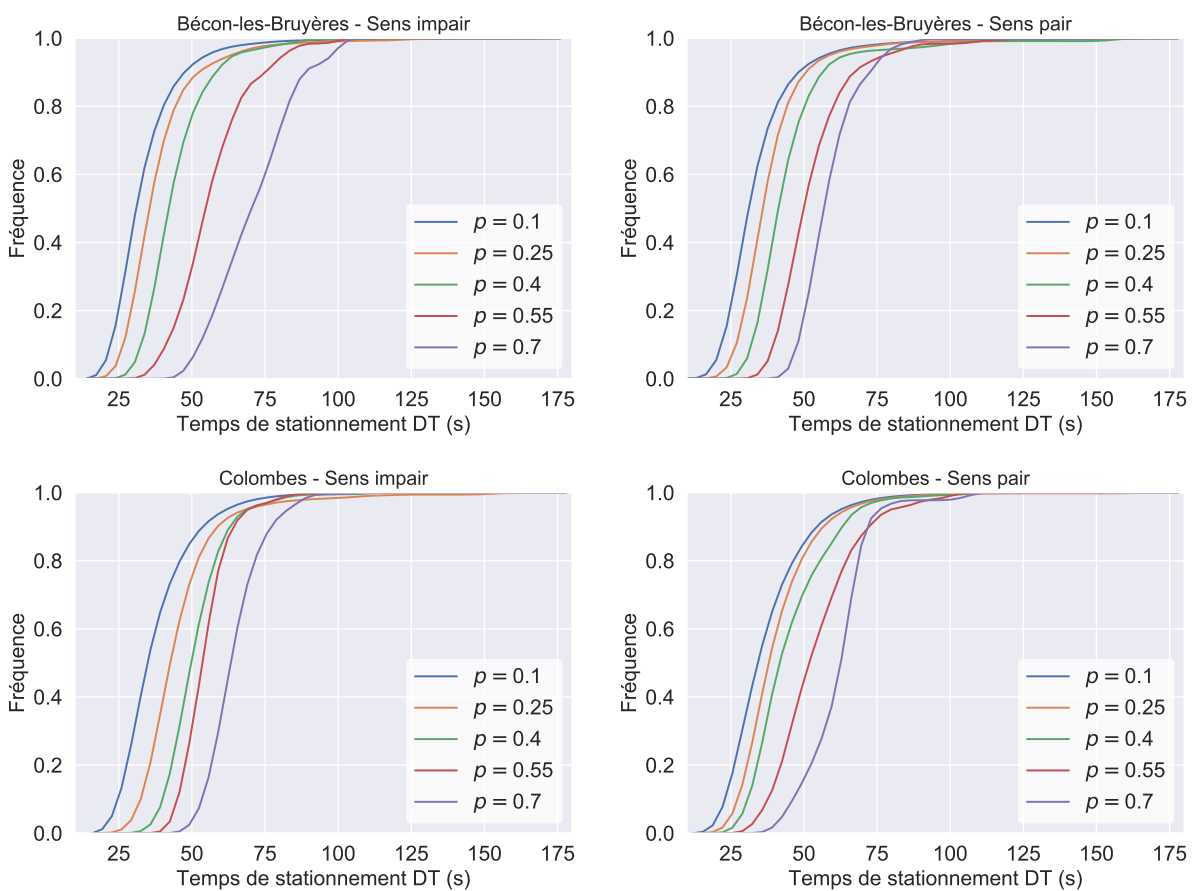
3.1.6 Annexe : résultats sur d'autres jeux de données

Nous présentons ici des résultats issus de jeux de données d'autres gares des lignes L et J. Le tableau 3.5 donne le nombre d'observations des jeux de données considérés, ainsi que la proportion de variance expliquée par la première composante principale. La figure 3.14 donne les fonctions de répartitions conditionnelles de DT connaissant le flux réduit p .

	NO - impair	%Var - impair	NO - pair	%Var - pair
Bécon-les-Bruyères	19111	0.68	19319	0.74
Colombes	24233	0.83	24199	0.91
Nanterre-Université	7592	0.66	7933	0.76

NO : nombre d'observations, %Var : proportion de variance expliquée.

Table 3.5 – Ratio de variance expliquée par chacune des composantes principales



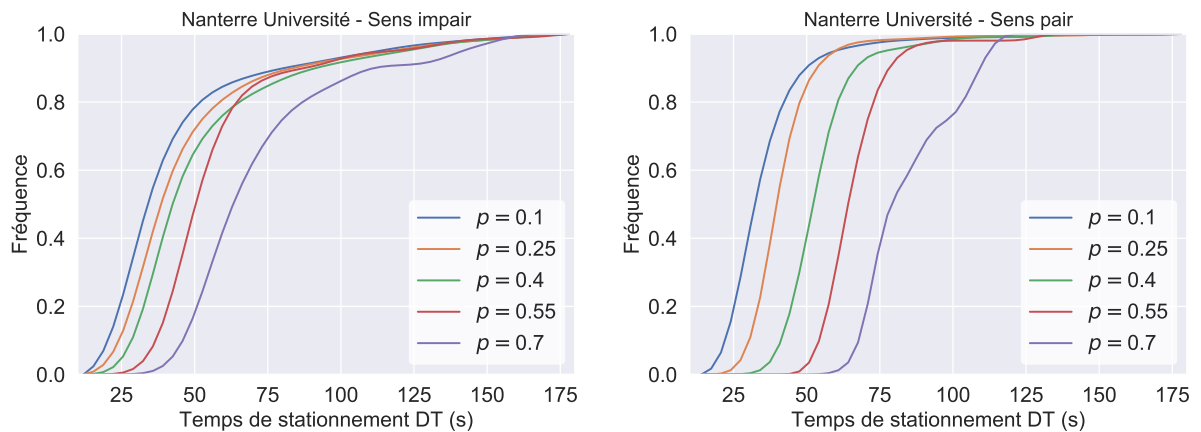


Figure 3.14 – Fonctions de répartition conditionnelles de DT connaissant P pour les gares de Bécon-les-Bruyères, Colombes et Nanterre Université

3.2 Modélisation des variabilités des temps de parcours

3.2.1 Présentation et état de l'art

Les temps de parcours des trains entre deux gares desservies sont également sujets à des variabilités. Si l'on exclut les incidents majeurs rendant l'infrastructure indisponible, ces variations sont principalement dûes aux facteurs suivants :

- Les éventuels conflits pour l'utilisation de l'infrastructure, qui se manifestent par la rencontre de signaux de voie libre ou fermés selon que ces conflits se produisent ou non. La nécessité de prendre en compte ces variabilités impose de construire un modèle à l'échelle du canton.
- L'hétérogénéité des attitudes de conduite. Certains ADC conduisent à une vitesse proche de la vitesse limite de l'infrastructure, tandis que d'autres sont plus proches de la vitesse de référence indiquée par la fiche train et permettant de rester dans le sillon. Cette hétérogénéité se manifeste également lors du franchissement d'un signal à l'avertissement, les ADC ayant des attitudes plus ou moins conservatrices vis-à-vis du dispositif de contrôle de vitesse KVB.
- Les conditions météorologiques. La pluie, le gel et la présence de feuilles mortes peuvent contraindre un train à réduire sa vitesse pour limiter les risques de patinage.
- Plus rarement, un dysfonctionnement léger du matériel roulant ou une baisse de l'alimentation électrique peuvent également engendrer des variations de vitesse.

Le calcul des durées de parcours entre points remarquables d'une ligne étant nécessaire à la construction des grilles horaires (pour le tracé de la marche de base), plusieurs méthodes de calcul ont été développées. Brünger et Dahlhaus (2008) en dressent un panorama. La plupart de ces méthodes sont déterministes et consistent en la résolution des équations différentielles du mouvement. Celles-ci nécessitent un volume de données important (description fine de l'infrastructure, notamment pente et courbure, masse, capacités de freinage et d'accélération du matériel roulant...), et permettent difficilement la prise en compte de la totalité des facteurs de variabilité. Steckel (1991) transforme cette approche déterministe en approche probabiliste, en introduisant un aléa sur les caractéristiques techniques du matériel roulant.

Cependant, ces stratégies de modélisation ne prennent pas en compte l'état du trafic ainsi que les autres aléas susceptibles de survenir au cours de la phase opérationnelle. C'est pourquoi d'autres approches, reposant sur l'utilisation de données d'exploitation, ont été mises en œuvre. Ainsi, Lüthi (2009) et Meer (2010) constatent l'absence de corrélation entre le retard et la durée de parcours en

zone dense. En effet, lorsque la consommation de capacité est importante, un train en retard pourra se trouver juste derrière un train à l'heure en provenance d'une autre branche ; celui-ci l'empêchera donc d'accélérer et de consommer sa marge sillon pour rattraper son retard. Hansen, Goverde et Meer (2010) effectuent le même constat, mais affirment en revanche que les temps de parcours peuvent être prédits en temps réel à partir du deuxième décile des temps de parcours du même train lorsqu'il n'a pas rencontré de conflit, et du deuxième décile du temps de parcours du train situé devant lui. Le projet Reliance, porté par SNCF Innovation & Recherche, vise de façon similaire à prédire les temps de parcours d'un train en comparant sa trajectoire passée à celles d'autres trains de l'historique, et en calculant les moyennes des temps de parcours des trains ayant les trajectoires les plus semblables. Kecman et Goverde (2015) ont testé plusieurs méthodes de régression (moindres carrés tronqués, arbres de régression, forêts aléatoires) pour prédire la durée de parcours de cantons, en prenant pour prédicteurs la longueur du canton, sa distance de la gare précédente et de la gare suivante, le type de train, le moment de la journée (heure de pointe ou non) et l'intervalle avec le train précédent à l'entrée. Ces modèles permettent la prise en compte de l'état du trafic, mais pas des autres facteurs de variabilité. Nous proposons ici un modèle intégrant de façon explicite les paramètres pour lesquels des données sont disponibles, et les autres sources d'aléas par une approche probabiliste. Comme dans la section précédente, il s'agit d'estimer les distributions conditionnelles des durées de parcours, connaissant certains paramètres de l'état du trafic.

3.2.2 Estimation des distributions conditionnelles de temps de parcours

Nous avons mis en œuvre une démarche proche de celle présentée au paragraphe 3.1.5. Des modèles locaux sont construits pour chaque segment reliant deux points pouvant soit tous deux être des signaux, soit être l'un une gare et l'autre un signal. Ces modèles permettent d'estimer la distribution de probabilité du temps de parcours de chacun de ces segments conditionnellement à la connaissance de certains paramètres. Parmi les facteurs de variabilité des durées de parcours, un seul est décrit dans les jeux de données à disposition : il s'agit de l'état des signaux aux moments où ils ont été franchis. Nous choisissons donc de calculer des distributions conditionnellement à ce paramètre. Il nous a cependant paru utile de conditionner également à un second paramètre : la durée de l'activité précédente. Ce choix présente en effet deux avantages :

- Il permet d'obtenir des distributions plus cohérentes avec la réalité, et notamment la contrainte physique de continuité de la vitesse. Ainsi, un train qui a circulé lentement dans un canton ne pourra pas parcourir le canton suivant à pleine vitesse, car il devra passer par une phase d'accélération.
- Cela permet en outre de tenir compte partiellement de l'hétérogénéité des attitudes de conduite. Carré (2014) a en effet montré que certains ADC ont une attitude conservatrice tout au long du parcours, en restant proches de la vitesse de référence voire légèrement en-dessous. D'autres se comportent de façon plus « agressive » en circulant à des vitesses proches de la vitesse maximale de l'infrastructure. D'autres encore changent d'attitude en fonction des événements rencontrés au cours de la mission : agressifs au début, ils deviennent conservateurs après avoir rencontré un signal d'avertissement. Le conditionnement relatif à la durée de parcours du canton précédent permet donc de reproduire en partie ce phénomène.

Il n'est pas souhaitable de conditionner par rapport à d'autres paramètres. En effet, en introduire davantage fait croître la dimension du problème et donc le volume de l'espace. Les observations des jeux de données deviennent alors éparpillées, chacune étant isolée des autres, ce qui empêche leur utilisation pour une estimation de densité précise. Ce phénomène est connu sous le nom de « fléau de la dimension » (Bellman, 1961) et intervient dans de nombreux problèmes d'analyse de données.

En pratique, la contrainte de marche à vue sur l'intégralité du canton fait que les signaux sont très rarement franchis à l'indication Sémaphore. L'état de la signalisation est donc un facteur discret pouvant prendre deux valeurs différentes, Voie Libre (VL) ou Avertissement (A). Nous avons par

conséquent divisé le jeu de données relatif à chaque canton en deux, l'un contenant les observations correspondant à un signal franchi à voie libre, l'autre à l'avertissement.

On note désormais RT et RTP respectivement la durée de parcours du segment considéré, et celle du segment précédent. Ces deux grandeurs sont supposées être des variables aléatoires à densité. La densité jointe du couple (RT, RTP) est estimée, comme à la section précédente, en utilisant la méthode d'estimation à noyau gaussien de Silvermann (1986). La distribution de RT conditionnellement à la connaissance de RTP en est déduite.

Nous présentons ici des résultats obtenus par cette méthode sur trois segments de la ligne L, situés entre les gares de Clichy-Levallois et Paris-Saint-Lazare (voir figure 3.2). Ceux-ci correspondent au parcours du canton situé entre les signaux C68 et C38, du segment entre le signal C38 et la gare de Pont-Cardinet (pour les trains desservant cette gare), et du canton situé entre les signaux C38 et C294 pour les trains ne desservant pas Pont-Cardinet. Une description de cette partie de la ligne est donnée sur la figure 3.15.

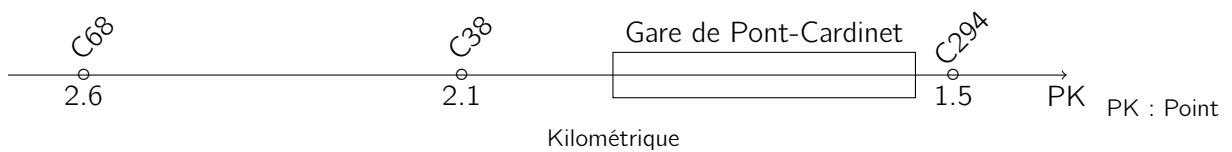


Figure 3.15 – Portion de la ligne L située entre les signaux C68 et C294

Les fonctions de répartition conditionnelles estimées sont représentées à la figure 3.16. Nous nous sommes limités aux valeurs de RTP donnant un nombre d'observations suffisant pour permettre une estimation précise de la densité. On observe bien que l'état du signal franchi joue considérablement sur la durée de parcours du canton : celle-ci peut ainsi varier du simple au double. On constate également que, dans une moindre mesure, la durée de l'activité précédente a également une corrélation positive avec le temps de parcours. Enfin, on constate une plus grande disparité sur les durées d'activité se terminant par une arrivée en gare. Cela semble indiquer une plus grande hétérogénéité d'attitude des ADC lors des phases de freinage, certains freinant de façon très sèche et d'autres de façon plus progressive.

Ces distributions conditionnelles peuvent être réutilisées en simulation stochastique ; cette application sera présentée au chapitre suivant. Les distributions estimées dans le cas d'un signal à voie libre peuvent également servir à réaliser des études de capacité de l'infrastructure.

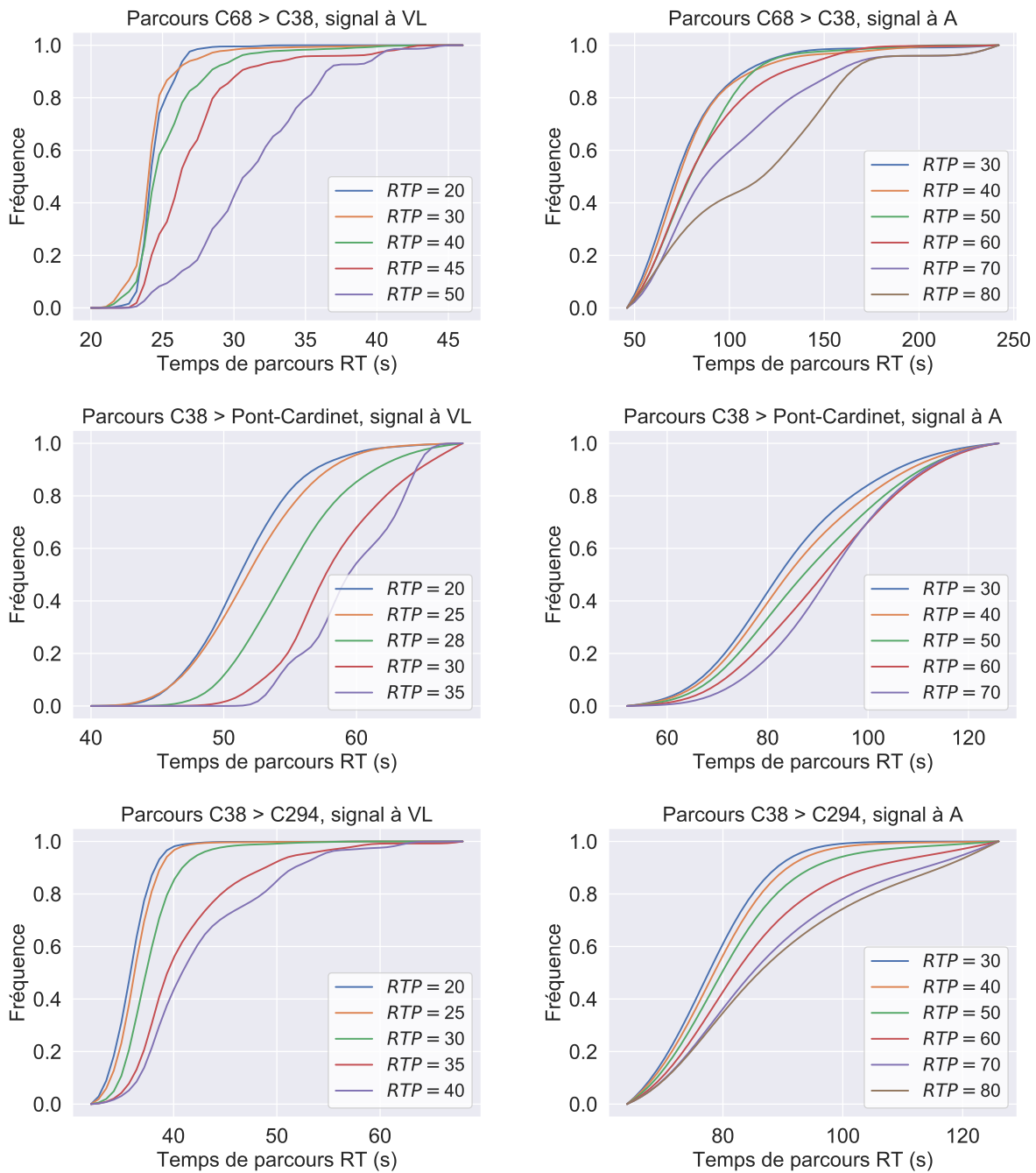


Figure 3.16 – Fonctions de répartition conditionnelles de RT connaissant RTP

Chapitre 4

Évaluation de grilles horaires par simulation stochastique

Résumé

Dans ce chapitre, nous décrivons une méthode de simulation stochastique permettant de quantifier la performance d'une grille horaire, et justifions ce choix de modélisation. Après un passage en revue de l'état de l'art, nous détaillons le fonctionnement de l'algorithme développé et son implémentation, puis nous démontrons sa pertinence. Nous présentons enfin plusieurs applications possibles de cet outil, le cas d'étude choisi étant une ligne saturée du réseau de Paris Saint-Lazare.

Sommaire

4.1	Motivation	88
4.2	Principes de la simulation stochastique	88
4.3	État de l'art en simulation du trafic ferroviaire	90
4.3.1	Simulateurs microscopiques	90
4.3.2	Simulateurs mésoscopiques et macroscopiques	91
4.4	Description de l'outil SISTRAM	92
4.4.1	Spécifications et choix de modélisation	92
4.4.2	Données d'entrée	93
4.4.3	Description de l'algorithme et du programme	93
4.5	Cas test : le système L2-U	99
4.5.1	Présentation	99
4.5.2	Grilles horaires	100
4.5.3	Constitution des fichiers d'entrée	101
4.5.4	Nombre de réplifications	103
4.6	Résultats	103
4.6.1	Conditions expérimentales	103
4.6.2	Validation de la démarche	104
4.6.3	Comparaison entre les grilles 2015 et 2019	104
4.6.4	Allocation de la marge dans la grille horaire	107
4.6.5	Approche « robustesse avec recours »	110
4.7	Conclusion	111
4.8	Annexe	112

« Essentiellement, tous les modèles sont faux, mais certains sont utiles. »
George P. Box, **Science and Statistics**, 1976.

4.1 Motivation

La simulation stochastique est une « méthode permettant d'analyser la performance de systèmes dont le comportement dépend des interactions entre plusieurs phénomènes aléatoires, pouvant être entièrement caractérisés par des modèles probabilistes » (Nelson, 2013). Celle-ci est fréquemment utilisée lorsque les mesures de performances ne peuvent pas être formulées de façon exacte. Ce n'est pourtant pas le cas de notre problème : nous en avons proposé une formulation sous la forme d'un programme stochastique au chapitre 2. Plusieurs éléments plaident néanmoins en faveur du recours à la simulation stochastique.

En effet, le problème du chapitre 2 fait intervenir des variables aléatoires dont la loi n'est connue que par une approximation à partir de données. Sa transcription en un problème déterministe ne peut donc se faire qu'à travers un échantillonnage de ces distributions en différents scénarios, obtenus par tirage aléatoire. Dès lors, le modèle n'est plus exact et l'éventuelle solution obtenue au problème échantillonné ne garantit pas d'être optimale pour le problème initial. En outre, la complexité du modèle rend son implémentation et sa résolution dans des délais raisonnables difficile.

La simulation stochastique présente, en revanche, un certain nombre d'atouts à faire valoir :

- Elle permet d'aborder les phénomènes aléatoires de façon bayésienne : les distributions de probabilité peuvent être calculées conditionnellement à la connaissance de la réalisation d'aléas précédents. Or, c'est précisément cette approche qui a été présentée au chapitre 3. Les modèles qui y ont été présentés pourront dès lors y être utilisés.
- Elle offre une grande souplesse d'implémentation. Cela permet, entre autres, de faire varier les règles de fonctionnement du système de façon à évaluer l'influence de différents facteurs extérieurs relativement indépendants de la grille horaire (attitudes de conduite, amplitude de la demande de transport...).
- Elle facilite la compréhension et l'interprétation des résultats, en permettant le calcul de divers autres indicateurs. Cela ouvre également la voie à une réutilisation de l'outil à d'autres fins (estimation de capacité, par exemple).

Pour toutes ces raisons, nous avons choisi de développer et d'implémenter un algorithme de simulation stochastique pour le calcul de l'indicateur de robustesse présenté au chapitre 2.

4.2 Principes de la simulation stochastique

La méthode de Monte-Carlo, proposée par Metropolis et Ulam (1949), fut l'une des principales contributions ayant permis l'essor de la simulation stochastique. Dans sa version originale, cette méthode visait à approximer les valeurs d'intégrales, d'aires et de volumes ne pouvant pas être calculés de façon exacte. Elle consiste à tirer des valeurs aléatoires, distribuées de façon uniforme dans tout l'espace (supposé borné). Après un grand nombre de tirages, on obtient une bonne approximation du volume à évaluer en calculant la proportion de points situés à l'intérieur de celui-ci. Cette méthode a depuis connu plusieurs développements et de nombreuses autres applications, dont nous aurons l'occasion de reparler dans la suite de ce paragraphe.

La première expérience probante de simulation stochastique d'un système complexe fut quant à elle menée par Gillespie (1976). Il s'agissait alors de simuler numériquement l'évolution de réactions chimiques couplées. Bien que pouvant être décrit par des équations différentielles déterministes, ce type de système est, tout comme le système ferroviaire, très sensible aux petites fluctuations. Son évolution est par conséquent difficile à prédire en pratique, ce qui justifie le recours à une méthode stochastique.

La méthode a par la suite été étendue à de nombreux autres champs d'application (production industrielle, finance, réseaux de télécommunications, dynamique de populations, propagation d'épidémies...). Cependant, un certain nombre de principes de modélisation et d'implémentation sont

largement répandus dans la classe des algorithmes de simulation stochastique, et ce indépendamment du système simulé. Nous en présentons quelques uns ci-dessous.

- Le principe des événements discrets. La plupart des méthodes de simulation visent à étudier ou prédire l'évolution temporelle d'un système donné. Il est par conséquent nécessaire de discrétiser le temps. Une approche classique en simulation déterministe consiste à le discrétiser avec un pas fixe, et de calculer l'évolution du système à chaque pas de temps. Cette stratégie est notamment utilisée lorsque la simulation requiert la résolution approchée d'équations différentielles. Elle n'est cependant pas toujours adaptée lorsque la simulation fait intervenir des phénomènes aléatoires : ceux-ci sont en effet susceptibles de se produire à des instants ne faisant pas partie de la subdivision du temps. En outre, tous les instants ne se valent pas : certains événements se produisant à des instants précis (la livraison d'une pièce, une transaction financière, une naissance, une contamination d'un individu par un autre...) peuvent être de nature à modifier le comportement du système. Le principe des événements discrets vise à rendre compte de cela. Selon celui-ci, le temps est cette fois discrétisé en intervalles irréguliers, chaque date correspondant à la survenue d'un événement notable dans le fonctionnement du système simulé. Les événements sont alors généralement ordonnés chronologiquement dans une file, et traités dans leur ordre de survenue. Le temps, quant à lui, « saute » de la date d'un événement à celle du suivant. Cette discrétisation du temps s'opère généralement en cours de simulation (la date de survenue des différents événements étant inconnus au début de celle-ci en raison du caractère stochastique).
- La génération de nombres pseudo-aléatoires. Celle-ci permet d'introduire la composante aléatoire dans la simulation. Pour ce faire, la plupart des langages de programmation proposent des fonctions ou des bibliothèques de génération uniforme de nombres pseudo-aléatoires. On désigne ainsi un algorithme nécessitant la donnée d'une valeur initiale u_0 (aussi appelée graine) et d'une transformation D . Alors, si on note $u_k = D^k(u_0)$, pour tout $n \in \mathbb{N}$, l'échantillon (u_1, \dots, u_n) reproduit un échantillon de n valeurs de distribution uniforme dans l'intervalle $[0, 1]$.

Un tel algorithme peut alors être utilisé pour générer des nombres aléatoires suivant une distribution quelconque. Il est nécessaire pour cela de connaître la fonction de répartition de la distribution, définie par

$$\forall t \in \mathbb{R}, \quad F(t) = \mathbb{P}(X \leq t) \quad (4.1)$$

où X est une variable aléatoire ayant la distribution correspondante. On définit la fonction de répartition inverse par

$$\forall t \in \mathbb{R}, \quad F^-(t) = \inf\{x \in \mathbb{R}, F(x) \geq t\} \quad (4.2)$$

Si U est une variable aléatoire uniforme sur $[0, 1]$ (dont on peut obtenir des tirages avec un générateur *ad hoc*), alors $F^-(U)$ est une variable aléatoire de même loi que X . Il est donc ainsi possible numériquement de réaliser des tirages aléatoires de nombres suivant une distribution de probabilité quelconque. En pratique, si la fonction de répartition F n'est pas connue, une approximation par histogrammes ou par une méthode telle que celle mise en œuvre au chapitre 3 peut être utilisée.

- L'utilisation d'un modèle de fonctionnement du système. Si les phénomènes aléatoires peuvent être reproduits à l'aide de la méthode décrite au point précédent, il est également nécessaire d'intégrer à l'algorithme un modèle de la façon dont ces phénomènes interagissent entre eux. Ce modèle dépend du problème considéré.
- La méthode de Monte-Carlo. Si la simulation a pour but le calcul d'un ou de plusieurs indicateurs, ceux-ci doivent être vus comme des variables aléatoires. Le résultat d'une seule réplique ne constituera donc qu'une seule réalisation de ces variables aléatoires, ce qui est insuffisant pour une prise de décision : le hasard des probabilités peut en effet conduire à une

valeur exagérément optimiste ou pessimiste. La réalisation d'un grand nombre de répliques permet en revanche d'obtenir une approximation fiable de l'espérance (et des moments d'ordre supérieur) des indicateurs en question. En effet, la loi des grands nombres assure que, si les différentes répliques sont indépendantes, la moyenne des valeurs obtenues sur l'ensemble des répliques converge vers l'espérance lorsque le nombre de répliques tend vers l'infini. Le théorème central limite et le lemme de Slutsky permettent alors de calculer des intervalles de confiance, et donc de choisir le nombre minimal de répliques à effectuer.

4.3 État de l'art en simulation du trafic ferroviaire

Comme nous l'avons mentionné à la section précédente, tout outil de simulation déterministe ou stochastique doit embarquer un modèle du système dont il reproduit le fonctionnement. De nombreux simulateurs du trafic ferroviaire ont été développés au cours des deux dernières décennies, s'appuyant sur des modèles très divers. Nous présentons ici les principaux en insistant sur le choix de modélisation de l'infrastructure, notamment les niveaux de détails définis au paragraphe 2.1.2 du chapitre 2.

4.3.1 Simulateurs microscopiques

Les principaux simulateurs commerciaux cherchent à obtenir des résultats d'une très grande précision, en recourant à une description très fine de l'infrastructure et de la dynamique ferroviaire. Deux progiciels, développés au début des années 2000, se partagent ainsi l'essentiel du marché : RailSys (Bendfeldt, Mohr et Müller, 2000) et OpenTrack (Nash et Huerlimann, 2004). Ces logiciels n'étant pas *open-source*, très peu d'informations sur leurs algorithmes sous-jacents sont accessibles. Le temps y est discrétisé avec un pas fixe et très faible (une seconde ou moins). Le mouvement des trains entre chaque pas de temps est calculé en résolvant les équations différentielles du mouvement, en tenant compte des caractéristiques du matériel roulant (masse, capacités de freinage et d'accélération...), de l'infrastructure (pente, courbure, limitations de vitesse...) et de l'état de la signalisation.

Ces outils sont utilisés par des EF et des GI à des fins diverses, allant des études de capacité à la planification de l'exploitation. Ils sont également utilisés pour des travaux de recherche. Ainsi, Solinen, Nicholson et Peterson (2017) utilisent RailSys pour une évaluation a posteriori de la robustesse aux points critiques (cet indicateur est présenté au chapitre 2). Pellegrini et al. (2015) ont développé un outil de gestion du trafic en temps réel qui est évalué et testé à l'aide d'OpenTrack. Les algorithmes de régulation développés par Altazin (2018) sont également testés sur une plateforme développée à SNCF Innovation & Recherche, baptisée NetWorkShop, dont OpenTrack constitue le cœur.

Ces deux simulateurs sont déterministes. Cependant, une adaptation de RailSys a été commandée par SNCF Transilien (Leboeuf et Lérin, 2019). Cette version, baptisée DENFERT, permet de simuler plusieurs fois une grille horaire donnée en la soumettant à différents aléas, afin d'évaluer sa robustesse. Cette approche a permis une amélioration significative des indicateurs de régularité et de ponctualité sur plusieurs lignes du réseau Transilien. Elle présente néanmoins plusieurs limites. En effet, on y considère les divers aléas surgissant en cours de simulation comme indépendants, ce qui n'est pas le cas. Par ailleurs, elle ne permet pas d'inclure dans la simulation les décisions de modification du plan de transport prises en temps réel par les GTI lors des situations dégradées, ni donc de tenir compte de leurs effets potentiels.

Si ces outils de simulation permettent d'obtenir un haut niveau de précision, ils présentent plusieurs inconvénients qui limitent leur possibilité d'utilisation :

- Le volume de données nécessaires en entrée. La modélisation de l'infrastructure préalable à une étude d'exploitation sur ces outils peut à elle seule nécessiter plusieurs semaines.

- Leur complexité algorithmique, qui se traduit par un temps de calcul très long. Une seule simulation ou réplique d'une journée d'exploitation peut prendre jusqu'à une heure.
- Leur manque de flexibilité et l'absence de disponibilité du code source, qui les rend difficilement adaptables à un problème spécifique. Cette difficulté est en partie palliée sur OpenTrack par la présence d'une API qui permet d'interagir avec le simulateur en cours de fonctionnement.
- Le coût de leur licence, parfois prohibitif pour les laboratoires de recherche académique.

Pour surmonter ces inconvénients, plusieurs modèles de simulation ont également été développés en laboratoire. Quaglietta et al. (2011) ont ainsi conçu un outil utilisant à la fois un modèle microscopique et un modèle mésoscopique de l'infrastructure, utilisés en alternance. Cette approche permet d'accélérer la simulation sans dégrader significativement la précision des résultats. Ochiai et Tomii (2015) ont développé un simulateur leur permettant de reproduire les variabilités d'attitudes de conduite et de conclure que celles-ci ont un impact élevé sur la ponctualité.

4.3.2 Simulateurs mésoscopiques et macroscopiques

Les simulateurs microscopiques ne sont cependant pas adaptés à toutes les situations. En effet, la grande précision qu'ils permettent d'obtenir n'est pas toujours nécessaire : c'est par exemple le cas lors d'études préliminaires à des modifications de plans de transport. Surtout, cette précision s'obtient au prix d'un temps de calcul très long. Or il est parfois nécessaire d'obtenir les résultats de simulation en quelques secondes, par exemple lorsque ceux-ci sont requis pour aider une prise de décision pendant la phase opérationnelle. C'est pourquoi d'autres approches, reposant sur une modélisation mésoscopique ou macroscopique de l'infrastructure et du mouvement des trains, ont également été mises en œuvre.

Li, Gao et Ning (2005) ont développé un automate cellulaire permettant de reproduire le trafic ferroviaire, que celui-ci soit régi par un système de signalisation fixe ou de block mobile (où les cantons évoluent de façon dynamique en fonction de la position et de la vitesse des trains). L'infrastructure y est divisée en cellules, pouvant être occupées par au plus un train à tout instant. Le temps est discrétisé selon un pas fixe, et à chaque pas de temps, l'état des cellules (présence ou non d'un véhicule, vitesse et accélération de celui-ci) est mis à jour. Un tel modèle peut par conséquent être qualifié de mésoscopique. Cette échelle de modélisation est également adoptée par Marinov et Viegas (2011), qui proposent un outil de simulation reposant sur le principe des événements discrets. L'infrastructure y est divisée en différents composants (cantons, aiguilles, gares de triage...) ayant des capacités différentes, et étant chacun pourvu d'une file d'attente. Les différents trains nécessitant d'utiliser ces composants sont insérés dans la file d'attente et traités selon leur ordre dans la file. Cette approche a été mise en œuvre afin de simuler la gestion des trains de marchandises, qui présente plusieurs spécificités (planification en dernière minute, vitesse réduite rendant le trafic hétérogène...).

Grube, Núñez et Cipriano (2011) simulent quant à eux le fonctionnement d'un réseau de métro en adoptant une échelle macroscopique : seules les stations constituent des nœuds dans leur modélisation de l'infrastructure. Les véhicules se déplacent de station en station à une vitesse fixe. Ce modèle est utilisé pour conduire des simulations stochastiques : l'incertitude est supposée porter sur les volumes de voyageurs, et donc sur les durées de stationnement. Les arrivées en gare des voyageurs y sont supposées déterminées par un processus stochastique tenant compte, à chaque réalisation, de la valeur de la réalisation précédente et d'un tirage dans une loi exponentielle. De par sa simplicité de modélisation, il permet de réaliser rapidement un grand nombre de répliques, ce qui permet aux auteurs de tester plusieurs stratégies de gestion du trafic en temps réel pour un grand nombre de scénarios.

Un simulateur macroscopique du trafic ferroviaire est également développé au sein de la direction Innovation & Recherche du groupe SNCF. Celui-ci, baptisé OLISTIC, permet de reproduire l'ensemble des circulations d'une journée sur l'ensemble du réseau en quelques secondes. L'infrastructure est modélisée sous la forme de points remarquables (convergences, gares) où les trains

s'y succédant doivent respecter un espacement temporel minimal. Le temps y est, là aussi, divisé selon le principe des événements discrets. L'outil est utilisé à la fois en conception, par exemple pour des tests de robustesse (Brethomé, 2018), et en opérationnel, pour faciliter la prise de décisions (Ramond, Pozzoli et Marcos, 2017). Il présente toutefois l'inconvénient de ne pas reproduire la signalisation, ce qui réduit la fiabilité de ses résultats en zone dense. Il est toutefois prévu que cet inconvénient soit pallié dans une prochaine version.

Enfin, mentionnons deux approches iconoclastes explorées récemment en simulation ferroviaire. Cuniasso (2015) choisit une représentation classique du réseau à l'échelle mésoscopique pour conduire des simulations stochastiques du trafic ferroviaire en zone dense. L'originalité de son travail réside dans le modèle retenu pour l'écoulement du trafic. Celui-ci s'inspire d'une part des modèles existants en trafic routier, d'autre part de la mécanique newtonienne. Les trains sont supposés être soumis à différentes forces : une force élastique tendant à rapprocher leur trajectoire du sillon théorique, une force aléatoire traduisant l'effet de petites perturbations, et des forces imposant de respecter les règles de sécurité (signalisation, vitesse limite). Cela conduit à un ensemble d'équations différentielles stochastiques dont la résolution permet de reproduire les trajectoires des trains et l'apparition de phénomènes de congestion. Watanabe et al. (2018) ont quant à eux développé un simulateur macroscopique à la fois rapide et précis. Celui-ci repose sur un algorithme de recherche de chemins dans un graphe, avec une particularité : les durées des différentes activités (et donc les valuations des arcs du graphe) sont obtenues par des algorithmes de *data-mining*, à partir de données d'exploitation. Ceci leur permet notamment d'étudier et de reproduire l'influence de facteurs, tels que la météo, qui avaient fait l'objet de peu de travaux jusqu'ici.

4.4 Description de l'outil SISTRAM

L'outil de simulation développé dans le cadre de cette thèse s'appelle SISTRAM (Simulateur Stochastique du TRAfic ferroviaire pour le Mass transit). Il s'inspire en partie des travaux présentés dans la section précédente. Nous présenterons tout d'abord les spécifications et la démarche de modélisation mise en œuvre pour les satisfaire, puis les différentes données à fournir en entrée, avant d'entrer dans le détail de l'algorithme et de son implémentation.

4.4.1 Spécifications et choix de modélisation

Les choix de modélisation et d'implémentation qui ont été utilisés lors de la conception et du développement de SISTRAM visent à satisfaire les spécifications suivantes :

- L'outil doit être en mesure d'évaluer une grille horaire rapidement sur un poste informatique standard ; ceci afin de permettre son utilisation comme fonction d'évaluation au sein d'une heuristique (voir le chapitre 5). Cela exclut le choix d'une modélisation microscopique de l'infrastructure et de la circulation ferroviaire. Nous avons choisi de développer un simulateur à événements discrets, ce qui permet de limiter le nombre de calculs et la mémoire de stockage nécessaires en ne reproduisant que les événements ayant une importance significative dans le déroulement de l'exploitation. Ce choix est également adapté à la représentation simplifiée du réseau que nous avons choisie.
- SISTRAM doit cependant permettre d'estimer des distributions de temps de parcours et de temps de trajet généralisé avec une précision satisfaisante. En particulier, il faut pouvoir reproduire les phénomènes d'instabilité et de congestion qui peuvent résulter de l'accumulation de petites perturbations. Il est donc nécessaire d'intégrer les mécanismes de propagation et d'amplification qui ont été présentés au chapitre 1 : l'affluence voyageurs et la signalisation. Nous avons par conséquent choisi de nous placer à une échelle mésoscopique : les cantons, convergences et gares constituent les nœuds du modèle d'infrastructure. Selon le principe proposé par Watanabe et al. (2018), nous ne nous intéressons pas directement aux caractéristiques

physiques de l'infrastructure. Les temps de parcours et de stationnement sont en effet obtenus à partir des modèles issus d'analyses de données qui ont été présentés au chapitre 3.

- Il doit offrir une flexibilité de fonctionnement, en permettant de tester différentes configurations ou l'influence de divers paramètres. Nous avons par conséquent choisi de le développer en suivant le paradigme de la programmation orientée objet, en langage Java.

4.4.2 Données d'entrée

L'outil nécessite en entrée les fichiers suivants :

- Une liste de l'ensemble des cantons, avec leur position (repérée par un Point Kilométrique, ou PK), et munie d'un ordre.
- Une liste de l'ensemble des convergences, précisant :
 - ◊ Le canton contenant la convergence.
 - ◊ Pour chaque branche, les cantons précédant celui contenant la convergence (qui présentent par conséquent un signal d'avertissement lorsque la convergence est occupée)
 - ◊ Le canton de sortie. L'entrée d'un train dans ce canton signale que celui-ci libère la convergence.
- Une liste de toutes les gares, précisant sur quels cantons celles-ci se trouvent.
- Une liste de tous les itinéraires possibles sur l'infrastructure. Un itinéraire est décrit par une succession de cantons allant de la gare d'origine à la gare terminus et que les trains parcourent dans cet ordre. Dans les grandes gares, seuls les itinéraires nécessaires à l'étude ont été reproduits.
- Une liste de tous les trains avec leur numéro et leur composition.
- Une liste de toutes les dessertes, indiquant les heures d'arrivée et de départ prévues pour tous les trains dans toutes les gares desservies.
- Une matrice OD de la ligne donnant, heure par heure, le nombre de passagers souhaitant effectuer chaque parcours possible sur la ligne, l'heure étant supposée être l'heure de montée à bord du train.
- Des fichiers donnant les distributions de probabilité des durées de stationnement, des durées de parcours de chaque canton (voir chapitre 3) ainsi que la distribution de probabilité du retard au départ de la gare d'origine.

4.4.3 Description de l'algorithme et du programme

4.4.3.1 Principales classes

SISTRAM a été développé selon le paradigme de la programmation orientée objet. Un diagramme de classes partiel est donné à la figure 4.1. Les principales classes qui le constituent et leurs attributs sont les suivants.

L'infrastructure

- **Block**, qui décrit un canton. Il permet de connaître l'état du signal qui se trouve à l'entrée.
- **Junction**, qui décrit une convergence. Chaque **Junction** contient les **Block** qui se trouvent à l'entrée et à la sortie, ainsi que le numéro du train pour laquelle la convergence est réservée. La signalisation interdit alors le franchissement de la convergence à tout train autre que celui à qui elle est réservée.
- **Station**, qui décrit les gares (ou plus précisément les quais de gares). Celles-ci sont situées sur un **Block**. Outre le stationnement des trains, elles gèrent également les flux de voyageurs : chaque **Station** calcule le nombre de voyageurs sur chaque OD l'ayant pour origine.

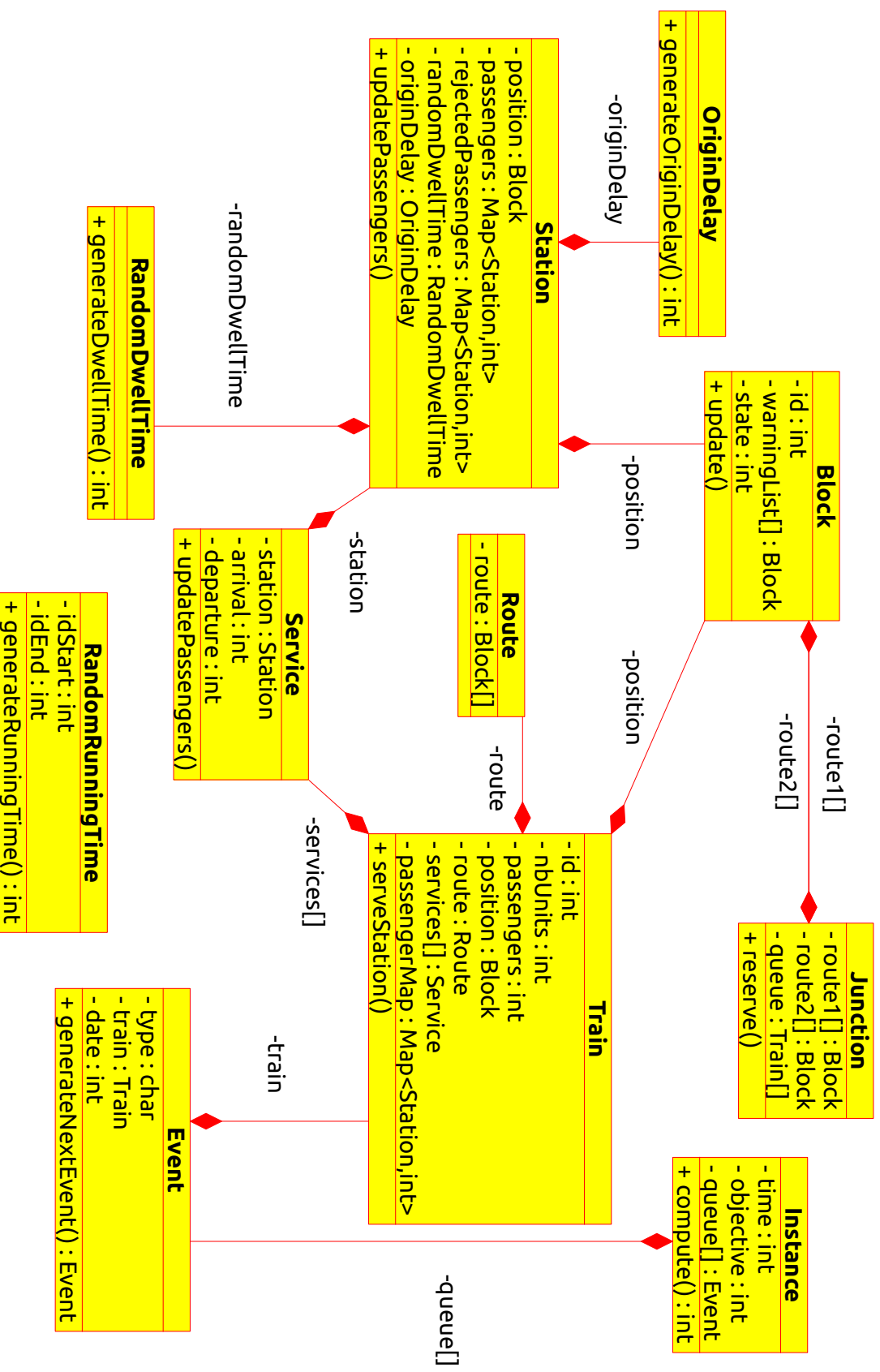


Figure 4.1 – Diagramme de classes partiel de SISTRAM

Le trafic

- **Route** qui décrit un itinéraire possible, vu comme une succession de **Block**.
- **Service**, qui concerne les stationnements en gare. Chaque objet de cette classe est constitué d'une gare, d'une heure d'arrivée théorique et d'une heure de départ théorique.
- **Train** qui décrit une mission en circulation. Chaque **Train** est repéré par un numéro de train unique. Il possède une **Route** ainsi qu'une liste de **Service** (dans l'ordre chronologique de desserte). Les **Train** gardent également trace de leur position sur l'infrastructure ainsi que des passagers à leur bord et de leurs destinations. Chaque **Train** a une capacité maximale d'emport voyageurs.
- **Event**, qui correspond à un événement concernant un **Train** unique et repéré par une date de réalisation. Ceux-ci peuvent être de 6 types :
 - ◊ Départ d'un train de sa gare d'origine.
 - ◊ Franchissement d'une entrée de canton.
 - ◊ Libération d'un canton.
 - ◊ Arrivée dans une gare desservie.
 - ◊ Départ d'une gare desservie.
 - ◊ Arrivée au terminus de la mission.

Les aléas

- **OriginDelay**, qui permet de générer des aléas sur la date de départ de la première gare. Il en existe un par gare d'origine.
- **RandomDwellTime** qui reproduit les aléas sur les durées de stationnement en gare en fonction des flux voyageurs (avec le modèle de la section 3.1).
- **RandomRunningTime**, qui permet de générer des aléas sur les durées de parcours des cantons (avec le modèle de la section 3.2).

Les classes « techniques »

- **Instance** qui décrit une réplique complète (correspondant à la simulation stochastique d'une journée entière de trafic avec une grille horaire donnée).
- **Parser** qui permet de construire les objets nécessaires à partir des fichiers d'entrée.
- **Writer** permettant d'écrire les résultats dans des fichiers de sortie.

4.4.3.2 Description de l'algorithme

Après une première étape d'instanciation (construction des objets nécessaires à la simulation à partir des fichiers d'entrée), **SISTRAM** exécute le nombre de répliques choisi en paramètre. Chaque réplique commence par la constitution d'une file d'événements dans l'ordre chronologique, celle-ci étant initialement composée des événements Origine de chaque train. Les dates de ces événements sont déterminées à partir de l'horaire théorique et d'un retard à l'origine aléatoire. Le premier événement est alors traité, sa date de réalisation est enregistrée et l'événement suivant du même train est généré et placé dans la file à la position qui correspond à sa date de réalisation prévue. La façon de traiter l'événement dépend de son type :

- S'il s'agit d'une entrée dans un nouveau canton, il est nécessaire de déterminer si la signalisation l'autorise. Si oui, l'événement est supprimé de la file et l'événement suivant du train est généré. Si non, l'événement est repositionné dans la file en seconde position, et cette procédure est répétée jusqu'à ce qu'un autre événement provoque l'ouverture du signal, permettant ainsi à l'événement différé d'être traité.

- S'il s'agit d'un départ de la gare d'origine, le nombre de voyageurs à bord est calculé, l'événement est enregistré et le suivant est généré.
- S'il s'agit d'une arrivée dans une gare, l'échange voyageurs (descentes et montées) est estimé et l'événement suivant est généré.
- S'il s'agit d'un départ d'une gare, celui-ci est enregistré et le suivant est généré.
- S'il s'agit d'un terminus (ou d'une libération de canton), celui-ci est enregistré et aucun nouvel événement n'est généré.

Après chaque événement, le statut de réservation de chaque convergence est mis à jour. La simulation s'arrête lorsque la file d'événements est vide. Un algorithme du fonctionnement général d'une réplique de SISTRAM est fourni à la figure 4.2. Nous décrivons plus en détails chacune des différentes étapes ci-dessous.

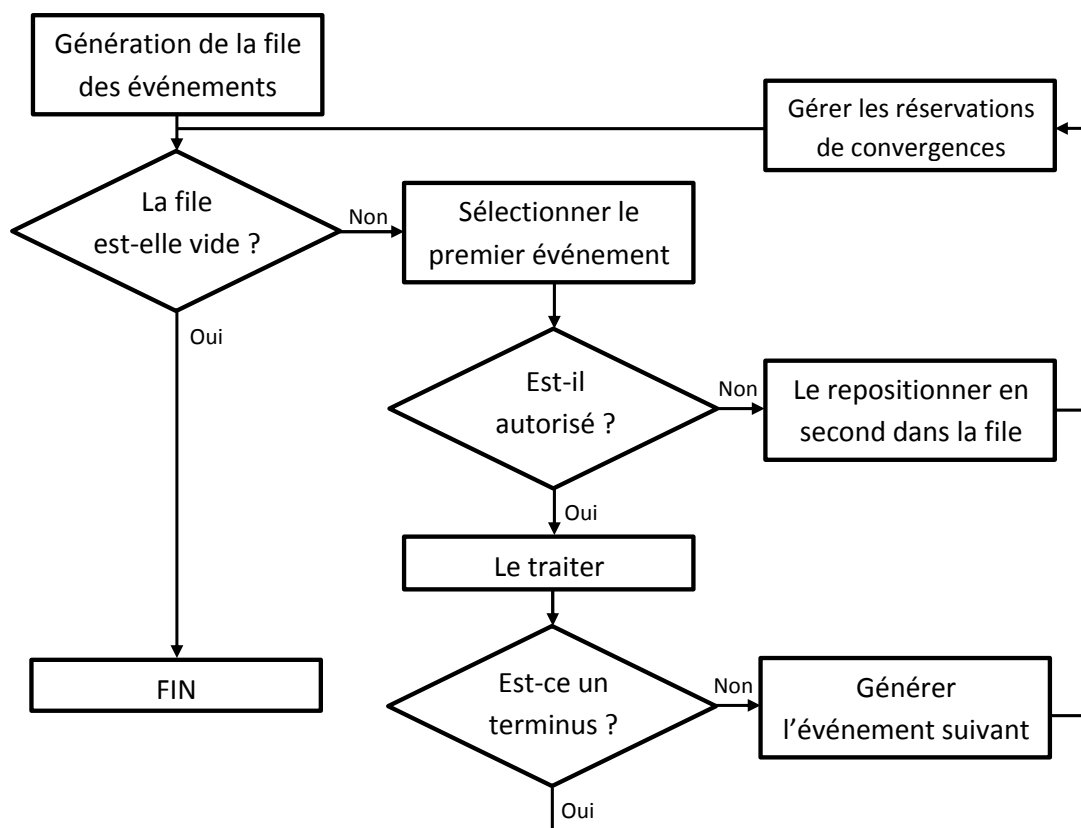


Figure 4.2 – Algorithme simplifié d'une réplique

Gestion de la signalisation et autorisations de parcours de cantons

Les signaux sont tous supposés en mesure de présenter les trois aspects décrits au chapitre 1 : sémaphore, avertissement et voie libre. Un événement « entrée dans un canton » est autorisé dès lors que le signal d'entrée présente l'aspect « avertissement » ou « voie libre ». Les signaux sont au sémaphore lorsque le canton contient une circulation. Lorsqu'une circulation entre dans un nouveau canton, elle génère un événement de libération du canton précédent se réalisant après une durée égale à

$$d = \frac{\ell}{L} \times T$$

où ℓ est la longueur de la rame, L la longueur du canton, et T la durée de traversée du canton par la tête du train. Les signaux protégeant des convergences font l'objet d'une gestion plus spécifique.

Ils sont représentés par la classe `JunctionBlock` qui hérite de `Block`. En plus de l'aspect relatif au cantonnement (identique à celui d'un `Block`), ceux-ci disposent d'un attribut représentant l'aspect relatif à l'enclenchement : voie libre si le signal est situé sur la même branche que le prochain train à franchir la convergence, carré fermé sinon. Le signal présente alors le plus restrictif de ses deux aspects aux circulations qui souhaitent le franchir.

L'aspect relatif à l'enclenchement est géré selon un système de réservations. Chaque objet `Junction` est doté d'une file dans laquelle se trouvent les trains qui doivent franchir la convergence. Les signaux sont alors ouverts sur la branche correspondant au premier train de la file, et fermés sur l'autre branche. La file est initialisée en y plaçant tous les trains dans l'ordre correspondant à leur horaire théorique de passage. Cependant, un mode de fonctionnement FIFO, où le premier train à se présenter à la convergence prend la première place dans la file, est également possible. Dès que le premier train de la file entre dans le canton suivant la convergence, il libère celle-ci et est retiré de la file. L'état des signaux est alors mis à jour en fonction du nouveau premier train de la file.

Traitement des événements et affectation des voyageurs

Le traitement d'un événement se limite, dans le cas d'un franchissement autorisé de signal, d'un départ de gare et d'un terminus, à l'enregistrement de la date et la mise à jour de la position du train concerné. En revanche, lorsque l'événement est un départ de la gare d'origine ou une arrivée, son traitement inclut également un calcul de flux voyageurs : montées et descentes dans le cas d'une arrivée, montées exclusivement dans le cas d'un départ de la gare d'origine.

Le calcul du nombre de descentes est effectué à l'aide d'une table de hachage associée à chaque objet `Train`, qui associe à chaque gare non encore desservie le nombre de voyageurs à bord qui s'y rendent. Lors d'une arrivée en gare, l'entrée correspondante de la table fournit immédiatement le nombre de descentes.

Le calcul du nombre de montées est effectué à partir de la matrice OD horaire, de la charge et de la capacité du train. La même procédure est répétée pour chaque OD ayant la gare desservie pour origine. Le nombre de voyageurs montant à bord pour effectuer cette OD est supposé proportionnel au temps écoulé depuis le passage du dernier train permettant de relier ces deux gares. Plus précisément, si on note h : m l'heure d'arrivée du train, h' : m' l'heure d'arrivée du train précédent et od_h l'entrée de la matrice `OD` correspondant à ce couple (origine, destination) pour l'heure h , alors le nombre de voyageurs souhaitant monter à bord pour effectuer ce trajet est supposé égal à

$$M = \tilde{M} + \begin{cases} \left\lfloor \frac{m - m'}{60} od_h \right\rfloor & \text{si } h = h' \\ \left\lfloor \frac{60 - m'}{60} od_{h'} + \sum_{k=h'+1}^{h-1} od_k + \frac{m}{60} od_h \right\rfloor & \text{si } h > h' \end{cases} \quad (4.3)$$

où \tilde{M} est le nombre de voyageurs souhaitant effectuer cette OD sur le train précédent mais n'ayant pas pu monter en raison de la capacité limitée de la rame. Le second cas correspond à la situation où plusieurs cases de la matrice `OD` sont requises pour le calcul ; le premier terme correspond alors aux flux de voyageurs arrivés pendant l'heure h' , le second aux voyageurs arrivés à des heures strictement comprises entre h' et h , le troisième aux voyageurs arrivés pendant l'heure h .

On note alors M_{total} la somme de toutes ces valeurs sur toutes les OD réalisables avec ce train à partir de la gare considérée. Si la somme de M_{total} et de la charge du train après les descentes est inférieure ou égale à la capacité du train, alors le nombre de montées est égal à M_{total} . Sinon, si on note C le nombre de places disponibles à bord après descente, le nombre de voyageurs montant pour une OD donnée est supposé égal à

$$\hat{M} = \left\lfloor \frac{M}{M_{total}} C \right\rfloor \quad (4.4)$$

Le nombre de voyageurs rejetés est alors égal à $\tilde{M} = M - \hat{M}$. Le nombre de voyageurs montants pour chaque OD est alors enregistré dans la table de hachage du train à l'entrée correspondant à la destination. La somme de ces valeurs donne le nombre total de voyageurs montants.

Génération de l'événement suivant et introduction de l'aléa

A l'exception du cas où l'événement courant est un terminus, sa suppression de la file s'accompagne de la création d'un nouvel événement concernant le même train. Ses caractéristiques dépendent du type de l'événement courant. En particulier, sa date de réalisation est obtenue par un tirage aléatoire utilisant les modèles présentés au chapitre 3.

- Si celui-ci est un franchissement de signal : alors, si le canton contient une gare desservie, le prochain événement du train est une arrivée (ou un terminus). Sinon, le prochain événement est un nouveau franchissement de canton. Dans les deux cas, la durée de l'activité conduisant au prochain événement est obtenue par tirage aléatoire. Ce tirage est effectué à partir de la distribution de probabilité conditionnelle de la durée de l'activité, connaissant l'état du signal au moment de l'entrée dans le canton ainsi que la durée de l'activité précédente (qui correspond également à un déplacement). Le conditionnement par rapport à cette dernière variable permet de respecter la continuité de la vitesse de déplacement.
- Si l'événement est un départ d'une gare (y compris de la gare d'origine), alors l'événement suivant est un franchissement de signal. La durée de ce parcours est obtenu par tirage dans la distribution correspondante, sans conditionnement : en effet, aucun signal n'est ici franchi et la précédente activité correspond à un stationnement (la contrainte de continuité de la vitesse est donc toujours satisfaite). En pratique, on trouve fréquemment un signal en sortie de gare, la durée de cette activité est par conséquent très courte dans ces cas-là (inférieure à la dizaine de secondes).
- Si l'événement est une arrivée dans une gare (autre que le terminus), on calcule tout d'abord le nombre de passagers descendant et montant à bord du train, ainsi que la charge après départ, en utilisant la méthode décrite dans le paragraphe précédent. On en déduit la valeur du flux réduit. Celle-ci est alors utilisée pour calculer la distribution conditionnelle du stationnement en gare, un tirage dans cette distribution permettant de déterminer la durée du stationnement.

Enfin, le retard du train au départ de sa gare d'origine fait également l'objet d'un tirage aléatoire. Celui-ci s'effectue lors de la phase d'initialisation, où la file des événements est constituée. Une distribution de probabilité du retard au départ est estimée pour chaque gare d'origine, et une valeur est tirée dans cette distribution pour chaque train au départ de cette gare.

Calcul d'indicateurs

Plusieurs indicateurs peuvent être calculés en sortie de façon à évaluer la qualité d'une grille horaire, au sens défini au chapitre 2. Comme nous l'avons mentionné dans ce chapitre, le principal critère que nous avons choisi d'utiliser correspond à l'espérance du temps de trajet généralisé de l'ensemble des usagers. Celui-ci se décompose en deux termes : l'espérance du temps d'attente en gare, et celle du temps à bord du train. Ces deux durées sont estimées, à chaque réplcation, de la manière suivante :

- Rappelons que le temps d'attente peut être formulé de la façon suivante (équations (2.18) et (2.19)) :

$$TA = \sum_{(s,s') \in S^2} \int_0^{T_f} p_{s,s'}(\theta) d\theta \quad (4.5)$$

où $p_{s,s'}(\theta)$ est la fonction donnant le nombre de passagers attendant à la gare s pour se rendre à la gare s' à la date θ . L'allure de cette fonction dépend de la fréquence des trains sur cette OD :

- ◇ Si, sur l'intervalle $[\theta, \theta+30 \text{ minutes}]$, la fréquence théorique sur l'OD (s, s') est inférieure à 15 minutes, alors la fonction $p_{s,s'}$ a une allure « en triangle », les passagers étant supposés arriver en gare selon un flux uniforme et sans considération pour l'horaire.
- ◇ Sinon, la fonction $p_{s,s'}$ a une allure « en créneaux », les passagers étant supposés arriver en gare 3 minutes avant l'heure de départ théorique du train qu'ils souhaitent prendre.

Dans les deux cas, l'intégrale de la fonction s'exprime comme une somme, pour chaque train :

- ◇ de termes pouvant être exprimés uniquement en fonction du nombre de passagers montant et du nombre de passagers rejetés par le train précédent
 - ◇ des horaires théoriques et réels d'arrivée du train et du train précédent. Tous ces paramètres étant calculés en cours de simulation, le temps d'attente peut par conséquent en être déduit.
- Le temps de trajet s'obtient comme la somme, sur toutes les activités, de la durée de l'activité multipliée par le nombre de passagers concernés (nombre de passagers montant et restant à bord pour un stationnement, nombre de passagers à bord pour un déplacement).

Les réplifications étant indépendantes, d'après la loi des grands nombres, la moyenne des valeurs obtenues sur toutes les réplifications permet d'obtenir une approximation de l'espérance.

L'espérance d'autres indicateurs peut également être estimée par le même procédé. Nous nous intéresserons en particulier à l'espérance de la régularité à 5 minutes au terminus, correspondant au pourcentage de trains arrivant à leur terminus avec un retard strictement inférieur à 5 minutes.

Détermination du nombre de réplifications minimal

Notons X la variable aléatoire représentant le principal indicateur d'intérêt (ici, le temps de trajet généralisé total), et pour une réplification $k \in \mathbb{N}$, notons X_k sa valeur à la réplification k . La loi (forte) des grands nombres assure alors que, presque sûrement,

$$\bar{X}_N := \frac{1}{N} \sum_{k=1}^N X_k \xrightarrow{N \rightarrow +\infty} \mathbb{E}[X] \quad (4.6)$$

Cependant, la loi des grands nombres ne donne aucune indication sur la rapidité de convergence, et donc sur un choix adéquat du paramètre N . Néanmoins, si on note $\hat{\sigma}_N$ l'estimateur empirique sans biais de l'écart-type, défini par

$$\hat{\sigma}_N = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (X_k - \bar{X}_N)^2} \quad (4.7)$$

il découle du théorème central limite et du lemme de Slutsky (voir Vermet (2017)) que, si on choisit un niveau de protection $\alpha \in]0, 1[$ et qu'on note q_α le quantile correspondant de la loi normale centrée réduite, on a alors

$$\mathbb{P} \left(\mathbb{E}[X] \in \left[\bar{X}_N - q_\alpha \frac{\hat{\sigma}_N}{\sqrt{N}}, \bar{X}_N + q_\alpha \frac{\hat{\sigma}_N}{\sqrt{N}} \right] \right) \simeq 1 - \alpha \quad (4.8)$$

Cet intervalle de confiance permet alors de choisir le nombre de réplifications à effectuer en fonction du degré de précision souhaité.

4.5 Cas test : le système L2-U

4.5.1 Présentation

Nous avons choisi de mettre en œuvre la méthode sur le système formé par le groupe II du réseau de Paris Saint-Lazare, constitué d'une branche de la ligne L du réseau Transilien (que nous appellerons

L2) ainsi qu'une portion de la ligne U. Ce choix a été motivé aussi bien par la complexité du système (topologie, fréquence et demande voyageurs), suffisamment importante pour que les résultats y soient significatifs, que par la disponibilité des données. Les autres lignes du réseau Transilien présentent quant à elles des insuffisances sur l'un ou l'autre de ces deux critères. Le tableau 4.1 présente les caractéristiques des autres lignes du réseau Transilien concernant ces deux critères.

Ligne	A-L3	B	C	D	E	H	J4	J5-J6	K	L2-U	N	P	R
Complexité	Vert	Vert	Vert	Vert	Vert	Rouge	Rouge	Vert	Rouge	Vert	Vert	Rouge	Vert
Données	Rouge	Rouge	Rouge	Rouge	Jaune	Vert	Vert	Rouge	Vert	Vert	Rouge	Vert	Rouge

Table 4.1 – Pertinence des lignes du réseau Transilien comme cas test éventuel

Une carte du système est donnée à la figure 4.3, au même niveau de modélisation que celui de SISTRAM. Y sont représentées les différentes gares ainsi que les signaux de protection de cantons (les signaux représentés étant ceux du sens pair Banlieue → Paris). Le système est constitué de trois branches :

- La branche « L2 », circulée par des missions reliant Versailles-Rive-Droite et Paris-Saint-Lazare.
- La branche « L2-bis », reliant Saint-Nom-la-Bretèche et Paris-Saint-Lazare. Ces deux branches convergent en amont de la gare de Saint-Cloud.
- La branche « U », reliant La Verrière à La Défense, et convergeant avec la branche L2 entre les gares de Viroflay-Rive-Droite et Chaville-Rive-Droite.

Le tronçon situé entre les gares de Saint-Cloud et La Défense est emprunté par toutes les circulations de la ligne. Ce tronçon, ainsi que la convergence se trouvant en amont de la gare de Saint-Cloud, sont donc souvent sujets à des phénomènes de congestion en heure de pointe.

En 2020, les missions de la ligne L2 sont assurées en majorité par des rames de la série Z50000, qui ont déjà été décrites au chapitre 3. Ces rames sont équipées de cassettes ATESS ainsi que d'un système de comptage automatique des voyageurs. Des rames de la série Z6400 y circulent également ; celles-ci sont beaucoup plus anciennes et en cours de radiation. Elles ne sont équipées ni de bandes ATESS ni de système de comptage. Il en est de même pour les rames Z8800 qui assurent les missions de la ligne U entre La Verrière et La Défense.

4.5.2 Grilles horaires

La grille horaire de ce système a fait l'objet de profondes modifications en 2016, à la demande d'associations d'usagers. Celles-ci considéraient en effet que la grille en vigueur ne permettait pas de répondre à la demande des voyageurs de façon satisfaisante, tant en termes de fréquences que de ponctualité.

La grille de 2015 est structurée autour des missions suivantes en heure de pointe :

- 4 trains par heure entre Versailles-Rive-Droite et Paris-Saint-Lazare (trains PILS). Ceux-ci desservent toutes les gares entre Versailles et Saint-Cloud, puis sont directs entre Saint-Cloud et La Défense ainsi qu'entre La Défense et Paris Saint-Lazare.
- 4 trains par heure entre Saint-Nom-la-Bretèche et Paris-Saint-Lazare (trains PALS). Ceux-ci sont également omnibus entre Saint-Nom-la-Bretèche et Saint-Cloud puis ne desservent que la gare de La Défense entre Saint-Cloud et Paris-Saint-Lazare. Ces trains sont renforcés pendant la pointe du matin par 4 missions origine Marly-le-Roi (trains PUBE).
- 4 trains par heure entre Saint-Cloud et Paris-Saint-Lazare (trains POBE), omnibus.
- 4 trains par heure de la ligne U (trains DEFI), omnibus entre Chaville Rive-Droite et La Défense.

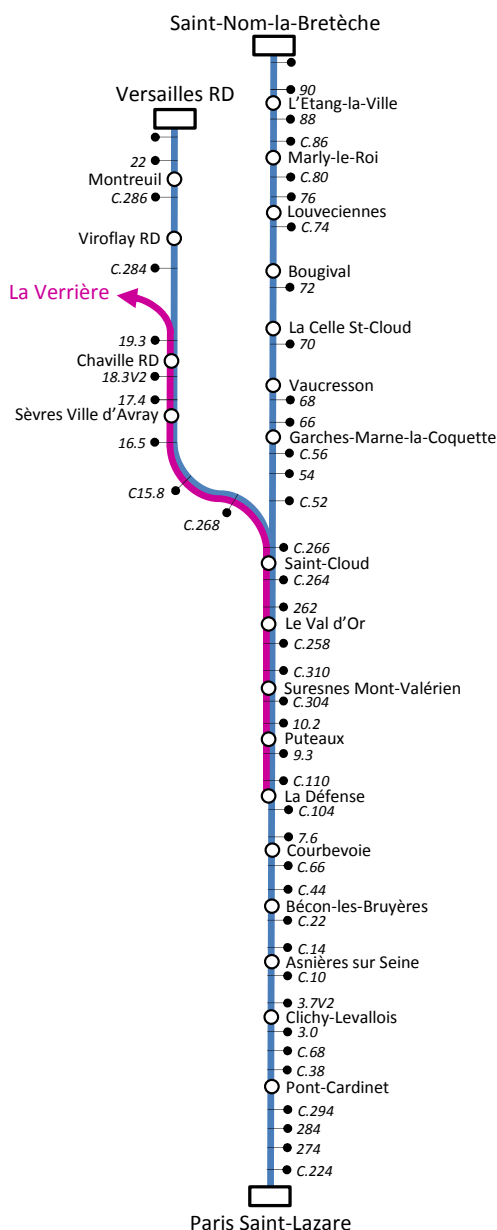


Figure 4.3 – Plan des lignes L2 et U

La refonte a eu pour effet de renforcer l'offre sur les branches de Versailles et Saint-Nom, ainsi que d'homogénéiser le trafic dans le tronçon central en passant à une desserte en « chapelet » : certaines grandes gares de correspondance sont desservies par tous les trains (Saint-Cloud, La Défense), les autres ne sont desservies que par un train sur deux. La nouvelle grille est structurée autour des missions suivantes en heure de pointe :

- 6 trains par heure entre Versailles-Rive-Droite et Paris-Saint-Lazare (trains PASA). Ceux-ci desservent toutes les gares entre Versailles et Saint-Cloud, puis s'arrêtent à Suresnes-Mont-Valérien, La Défense, Asnières-sur-Seine, Clichy-Levallois et Pont-Cardinet.
- 6 trains par heure entre Saint-Nom-la-Bretèche et Paris-Saint-Lazare (trains PEBU). Ceux-ci desservent toutes les gares entre Saint-Nom-la-Bretèche et Saint-Cloud puis Le Val d'Or, Puteaux, La Défense, Courbevoie et Bécon-les-Bruyères.
- 4 trains par heure de la ligne U (trains DEFI), qui desservent toutes les gares entre Chaville-Rive-Droite et La Défense sauf Le Val d'Or.

La fréquence maximale dans le tronçon central est donc inchangée à 16 trains par heure. Il est à noter que la ligne U est une ligne « structurante » du système et que ses horaires en dehors du système présenté ici n'ont pas été modifiés lors de la refonte. En effet, celle-ci entre en interconnexion avec les lignes N et C au niveau de Porchefontaine, la modification de ses horaires aurait par conséquent nécessité des changements sur ces lignes également. Ces lignes étant également interconnectées avec d'autres lignes du réseau (ligne H, TER Centre et Normandie, raccordement entre les LGV Atlantique et Sud-Est...), c'est une grande partie du RFN qui aurait pu alors voir ses horaires modifiés. C'est pourquoi ce choix a été opéré par les horaristes lors de la refonte.

4.5.3 Constitution des fichiers d'entrée

Les fichiers permettant de modéliser l'infrastructure ont été constitués à l'aide des remontées de cassettes ATESS, complétées occasionnellement par l'utilisation de plans de voie. Les grilles horaires 2015 et 2019 ont également pu être reconstruites exhaustivement. En revanche, la constitution d'autres fichiers nécessaires à SISTRAM a nécessité d'effectuer certaines hypothèses simplificatrices, toutes les données requises n'existant pas dans les systèmes d'information.

Tout d'abord, les rames Z6400 et Z8800 n'étant pas équipées de systèmes ATESS et CAVE, il n'a pas été possible de construire de distributions de probabilité de temps de parcours et de stationnement spécifiques à ces rames. Nous avons par conséquent supposé que tous les trains simulés sont opérés avec des rames de type Z50000. En pratique, les performances d'accélération

et de freinage des rames Z6400 et Z8800 étant inférieures, cela peut conduire à des durées de parcours sous-estimées par rapport à la réalité. Il en est de même pour les durées de stationnement : contrairement à celles des rames plus anciennes, les portes des Z50000 sont situées à niveau par rapport au quai et sont plus larges, ce qui facilite les échanges de voyageurs.

Par ailleurs, la base de données VENDOME fournit une matrice OD horaire pour la ligne. Cependant, nous avons constaté que ses entrées ne sont pas cohérentes avec les observations provenant des comptages automatiques. Cela s'explique par le fait que la matrice OD est reconstituée à partir d'enquêtes menées tous les 4 à 5 ans auprès d'un petit nombre de voyageurs. Plutôt que d'utiliser cette matrice OD en l'état, nous avons choisi de la modifier afin de la mettre en cohérence avec les données de la base CHATELET (que nous considérons comme fiables) pour l'année 2018. Ce problème de reconstitution d'une matrice OD à partir de données de montées et descentes est un problème classique en transport. Il s'agit de reconstituer une matrice connaissant la somme de chaque ligne et de chaque colonne, le problème est donc sous-déterminé dès que le nombre de gares est strictement supérieur à 2. Plusieurs approches visant à reconstituer la matrice la plus probable ont été proposées par Bell et Iida (1997). Nous avons choisi de procéder de la façon suivante.

- Nous commençons par calculer un flux de voyageurs montant gare par gare et heure par heure. Pour cela, nous notons, pour une date d , une gare g et une heure entière h , $\{t_1, \dots, t_k\}$ l'ensemble des trains assurés en Z50000 ayant desservi la gare g pendant l'heure h de la date d (les autres trains étant passés par cette gare n'apparaissant pas dans les données). Nous notons également $M_{t_1,g}, \dots, M_{t_k,g}$ le nombre de voyageurs montant dans chacun de ces trains à la gare g . Enfin, nous notons $K_{g,h}$ le nombre théorique de trains de la ligne L2 desservant la gare d pendant l'heure h . Le flux montant à la gare d pendant l'heure h est donc défini, pour la date d , par

$$\Phi_{g,d,h} = \frac{K_{g,h}}{k} \sum_{j=1}^k M_{t_j,g} \quad (4.9)$$

Nous avons ici fait l'hypothèse que le flux voyageurs se répartit entre les différents trains indépendamment de la rame utilisée. Le facteur correctif $K_{g,h}/k$ vise donc à tenir compte des montées dans les trains n'étant pas équipés de système de comptage.

Nous définissons alors le flux montant à la gare g et à l'heure h comme la moyenne des valeurs précédentes sur tous les jours $\{1, \dots, D\}$ où des données sont disponibles :

$$\Phi_{g,h} = \frac{1}{D} \sum_{d=1}^D \Phi_{g,d,h} \quad (4.10)$$

En pratique, nous n'avons considéré que des jours considérés comme représentatifs, c'est-à-dire uniquement des JOB (Jours Ouvrables de Base - mardi, mercredi et jeudi), hors vacances scolaires, et hors journées significativement perturbées (grèves, dommages importants à l'infrastructure...).

- Nous distribuons alors ces flux montants sur les différentes gares de destination possibles en utilisant des coefficients de proportionnalité fournis par la matrice OD de VENDOME. Si on note $\widetilde{od}_{g,g',h}$ la valeur fournie par VENDOME pour le couple origine-destination (g, g') à l'heure h , alors nous prenons comme entrée de la nouvelle matrice OD la valeur

$$od_{g,g',h} = \frac{\widetilde{od}_{g,g',h}}{\sum_{\gamma \in G} \widetilde{od}_{g,\gamma,h}} \Phi_{g,h} \quad (4.11)$$

La nouvelle matrice OD obtenue est inconsistante avec les flux de descente (les flux de descente observés ne sont pas nécessairement égaux à la somme des entrées de la matrice OD ayant la gare

considérée pour destination), mais consistante avec les flux de montée. Cette dernière caractéristique nous a paru préférable, étant donné que la montée des voyageurs dimensionne plus fortement la durée de stationnement en gare.

4.5.4 Nombre de réplifications

Les tests et paramétrages préliminaires sont effectués sur la grille horaire 2019. Le premier d'entre eux consiste à déterminer un nombre de réplifications adéquat, à l'aide de la méthode présentée au paragraphe 4.4.3.2. Pour cela, nous calculons après N réplifications la moyenne des temps de trajet généralisés totaux ainsi que leur écart-type, et nous en déduisons une largeur d'intervalle de confiance au niveau 1% ($\alpha = 0.01$). L'évolution de la largeur de cet intervalle (exprimé en proportion de la valeur de \bar{X}_N) en fonction de la valeur de N est représentée à la figure 4.4.

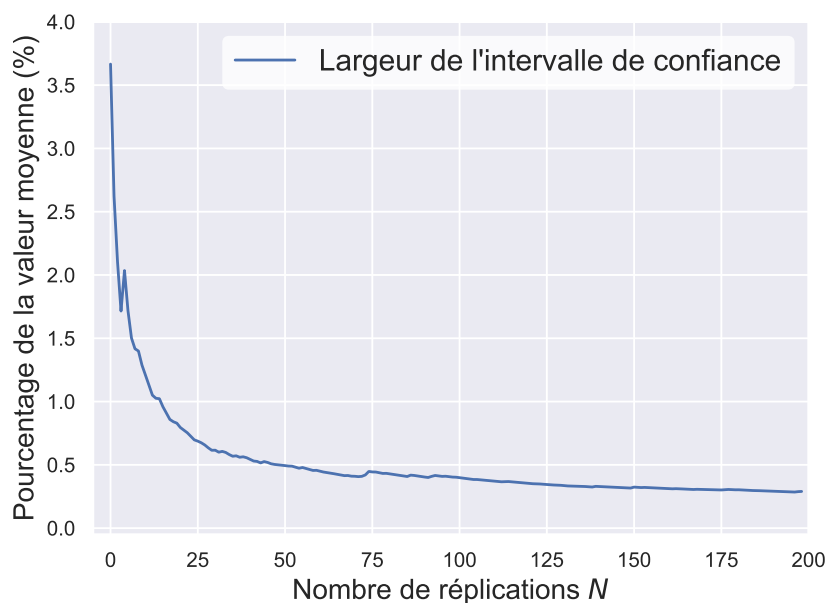


Figure 4.4 – Largeur de l'intervalle de confiance en fonction du nombre de réplifications

La largeur décroît lentement (en $O(1/\sqrt{N})$) à partir de $N = 50$. De plus, pour 50 réplifications, l'intervalle de confiance au niveau 1% a une largeur égale à 0.5% de la valeur moyenne, ce qui nous paraît acceptable. Nous choisissons donc désormais un nombre de réplifications minimal $N = 50$ pour l'évaluation des grilles horaires.

4.6 Résultats

4.6.1 Conditions expérimentales

Le programme ainsi développé a été exécuté sur un processeur Intel Core i7, avec une fréquence de 2.90 GHz, et possédant 4 cœurs (les réplifications n'ont pas été parallélisées). Le système d'exploitation est Ubuntu 18.04, et le poste possède une mémoire RAM de 7.5 Go. Sur toutes les instances testées et présentées ci-après, la durée nécessaire à la réalisation d'une réplification aléatoire est comprise entre 600 et 700 millisecondes.

4.6.2 Validation de la démarche

Avant toute utilisation de SISTRAM visant à produire des résultats interprétables, nous avons cherché à vérifier sa capacité à reproduire fidèlement les effets de la stochasticité sur le trafic ferroviaire. Le temps de trajet généralisé total n'étant pas accessible à partir des données (tous les trains n'étant pas équipés de cassettes ATESS), nous avons choisi de comparer les distributions des temps de parcours de chaque train entre leur origine et leur terminus.

Pour ce faire, nous avons sélectionné un échantillon de 50 jours de type JOB, hors vacances scolaires, entre les mois de novembre 2018 et d'octobre 2019. Ces jours ont été choisis de telle sorte que l'exploitation sur les lignes L et U n'ait été marquée par aucun incident majeur répertorié dans le bulletin « Faits marquants » (voir le chapitre 1). Les données d'exploitation issues de ces journées constituent ce que nous appelons les « données observées ».

Par ailleurs, nous avons simulé 50 réplifications de la grille horaire 2019 pour un JOB dans le sens pair (de la banlieue vers Paris). Nous avons alors comparé, pour chaque train, les distributions de temps de parcours observés et simulés à l'aide de diagrammes quantile-quantile. Ceux-ci sont donnés à la figure 4.5 pour quelques trains représentatifs des résultats observés. Ceux-ci sont tous des trains effectuant la liaison Versailles Rive Droite - Paris Saint-Lazare, mais les diagrammes des trains de la branche de Saint-Nom-la-Bretèche sont similaires.

Ces diagrammes nous conduisent aux observations suivantes :

- On constate que pendant l'heure de pointe du matin, les quantiles inférieurs des distributions de temps de parcours sont estimés très précisément, tandis que l'erreur commise est plus importante pour les quantiles supérieurs tout en restant acceptable (moins de 10%). Le premier facteur pouvant expliquer cela est l'inhomogénéité du trafic induite par celle du matériel roulant (uniquement des Z50000 en simulation, cohabitation de Z6400, Z8800 et Z50000 dans les données observées). Leurs différences de performances peuvent provoquer des phénomènes d'instabilité, particulièrement en heure de pointe où les fréquences sont importantes et où la marge est insuffisante pour absorber tous les effets de cette inhomogénéité. D'autres variabilités non prises en compte (météo, hétérogénéité des profils d'ADC) peuvent également expliquer en partie cette différence.
- Le reste de la journée, la qualité de l'estimation est plus satisfaisante. L'erreur est en effet autour de 5% sur quelques trains (tel que le 133534), et très inférieure pour les autres, que ce soit en heure creuse ou en pointe du soir. Celle-ci est en effet moins sensible aux phénomènes d'instabilité que celle du matin, l'essentiel des flux voyageurs étant concentré sur la partie située entre La Défense et Paris ainsi que sur le sens impair.

Nous considérons donc que SISTRAM reproduit de façon satisfaisante l'influence des petits aléas sur l'exploitation ferroviaire, et qu'il peut être utilisé pour le calcul d'indicateurs dont les valeurs seront interprétées.

4.6.3 Comparaison entre les grilles 2015 et 2019

4.6.3.1 Vitesse moyenne du point de vue voyageurs

La grille horaire du système L2-U ayant été modifiée en 2016, cette modification ayant été suivie d'une amélioration de la qualité de service, nous avons naturellement cherché à comparer les performances de ces deux grilles dans SISTRAM. Plutôt que de comparer les valeurs de temps de trajet généralisé obtenues en simulant ces deux grilles, nous utilisons un autre indicateur construit à partir de celui-ci. Ce dernier offre l'avantage d'être plus facilement compréhensible et interprétable. Nous l'appellerons la **vitesse moyenne de la ligne du point de vue du voyageur**, que nous définissons de la façon suivante :

$$v = \frac{\text{Nombre de voyageurs.km}}{\text{Temps de trajet généralisé total}} \quad (4.12)$$

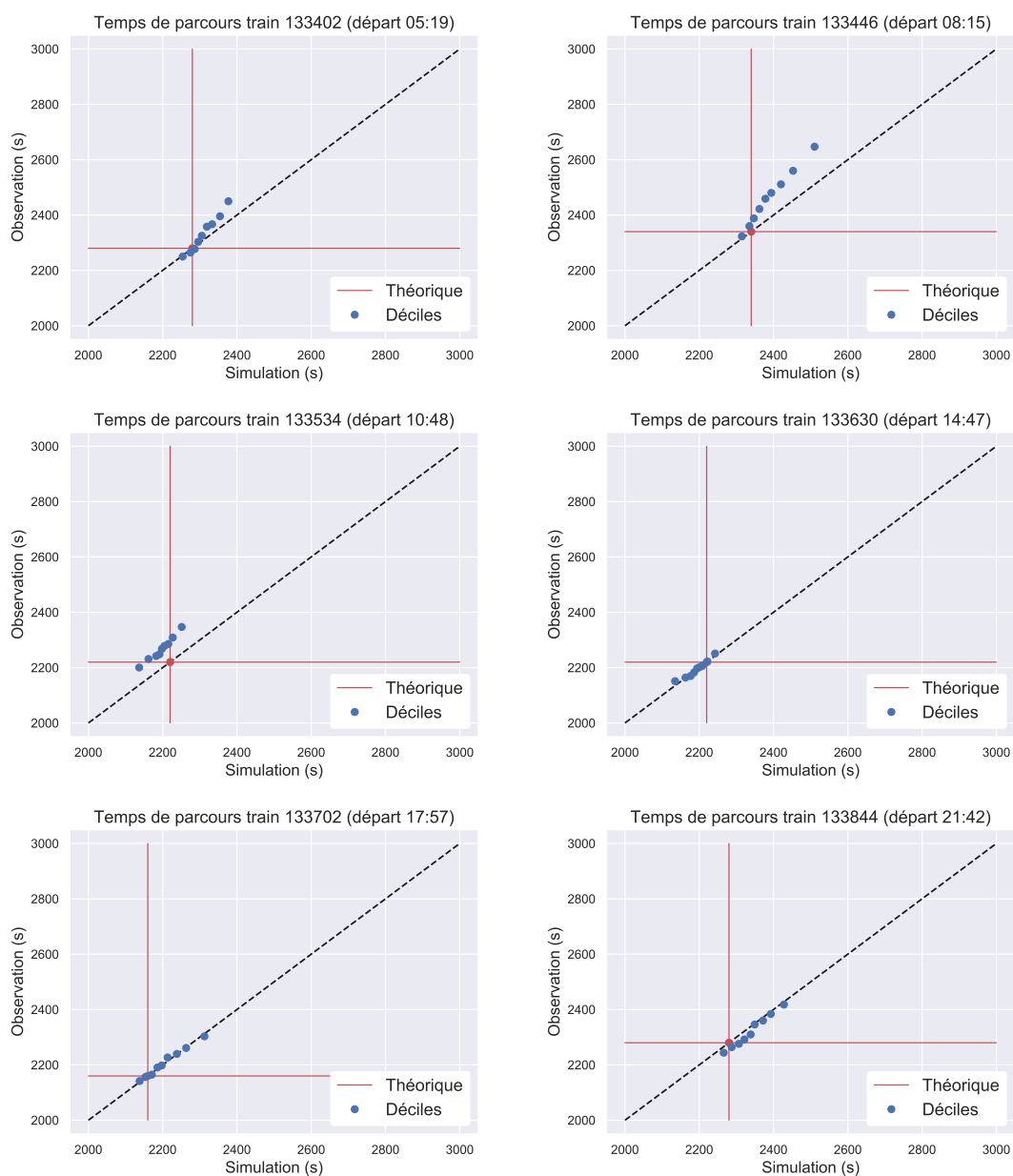


Figure 4.5 – Diagrammes quantile-quantile des distributions de temps de parcours observées et simulées

Cette vitesse représente la vitesse moyenne à laquelle s'effectue le déplacement des voyageurs, ce déplacement étant considéré comme débutant à l'arrivée en gare et se terminant à la descente à la gare de destination. Toute augmentation de la durée d'attente en gare ou du temps passé à bord d'un train a donc pour effet de réduire cette vitesse moyenne.

En outre, il nous a paru souhaitable de comparer les performances des grilles 2015 et 2019 toutes choses égales par ailleurs. Or, la demande en transport à laquelle le système doit faire face a évolué entre ces deux dates : on estime la croissance du trafic voyageurs en Île de France à 3% par an (SNCF Réseau, 2019), soit une augmentation de 16% entre 2015 et 2019.

Nous avons par conséquent comparé les performances de ces deux grilles pour différents niveaux de demande. A cet effet, nous avons multiplié toutes les entrées de la matrice OD horaire par un même coefficient que nous avons fait varier. La figure 4.6 donne la vitesse moyenne du point de vue voyageur, calculée sur l'ensemble de la journée et sur la seule pointe du matin, pour les grilles 2015 et 2019 et pour différentes valeurs du coefficient de demande. Un coefficient égal à 1 correspond au

niveau de demande de 2019, et un coefficient de 0.86 correspond à celui de 2015.

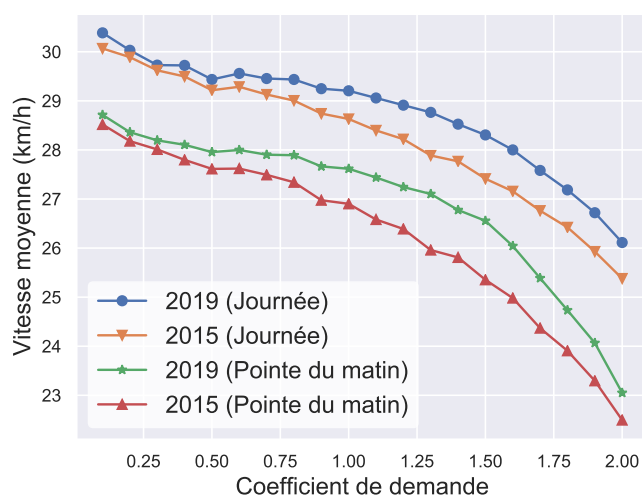


Figure 4.6 – Comparaison des vitesses moyennes avec les grilles horaires de 2015 et 2019

On constate effectivement que la grille 2019 permet d’offrir une meilleure vitesse de déplacement en moyenne aux voyageurs. Cela s’explique à la fois par le fait que, la fréquence étant renforcée sur les branches de Saint-Nom et de Versailles, la durée d’attente y est réduite, et par le fait que la meilleure homogénéité du trafic entre Saint-Cloud et Paris Saint-Lazare réduit les phénomènes d’instabilité et donc les temps de parcours moyens.

On observe également une dégradation significative de la qualité de service lorsque la demande augmente fortement. En effet, en heure de pointe le matériel roulant arrive plus rapidement à saturation, ce qui conduit davantage de voyageurs à attendre le train suivant et augmente leur temps d’attente. En outre, les flux de voyageurs plus importants provoquent l’allongement des durées de stationnement, et donc une diminution de la capacité, une hausse de l’instabilité et des durées des trajets.

4.6.3.2 Régularité au terminus

Le contrat définissant les objectifs de performance fixés par l’AO ne portant pas - encore - sur l’indicateur du paragraphe précédent, nous comparons également les performances des deux grilles sur la base d’un indicateur « à seuil ». Nous avons calculé, à l’issue de chaque réplique, la régularité à 5 minutes au terminus, c’est-à-dire la proportion de trains arrivant à leur terminus avec un retard inférieur ou égal à 5 minutes. L’évolution de celle-ci en fonction de la demande voyageurs est représentée à la figure 4.7, pour les deux grilles.

On constate ici un écart substantiel entre les deux courbes, dû aux plus nombreuses instabilités qui résultent de la grille de 2015. Cette grille ne permet donc pas d’obtenir des valeurs satisfaisant les objectifs contractuels (autour de 90%) même en l’absence de perturbations majeures. En pratique, des trains sont régulièrement supprimés lors de la phase opérationnelle afin de fluidifier le trafic et d’améliorer la régularité. Une exploitation régie par une grille telle que celle de 2015 peut donc toujours être viable, tout en étant sous-optimale et en nécessitant des moyens plus importants.

Le résultat observé confirme également une hypothèse préliminaire, selon laquelle la régularité n’est qu’un indicateur imparfait de la qualité de service en zone dense. En effet, si on relève un écart de régularité très important entre les grilles de 2015 et de 2019, l’écart est bien moindre en ce qui concerne la vitesse moyenne de déplacement. La grille de 2015 permet ainsi d’offrir un service raisonnablement satisfaisant aux voyageurs (en tout cas lorsque comparé à celle de 2019), tandis que les courbes de la figure 4.7 pourraient amener à penser le contraire.

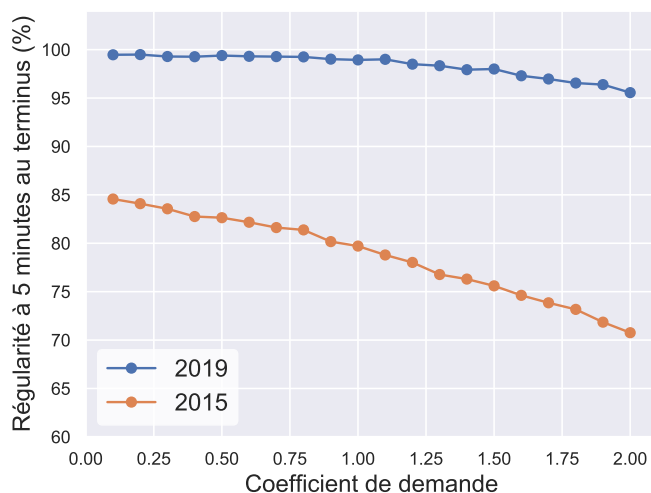


Figure 4.7 – Comparaison de la régularité à 5 minutes avec les grilles de 2015 et 2019

4.6.4 Allocation de la marge dans la grille horaire

Comme nous l'avons mentionné aux chapitres 1 et 2, deux types de marges peuvent être insérés dans la grille horaire pour tenter d'atténuer les effets des petits aléas. La marge sillon, qui consiste à allonger volontairement le temps de parcours et/ou de stationnement d'un train, permet à celui-ci de rattraper un retard de faible amplitude. L'utilisation de marge sillon se fait au prix d'une consommation de capacité supplémentaire ainsi que d'un éventuel allongement inutile du temps de trajet pour les voyageurs si cette marge est surdimensionnée. La marge de grille consiste quant à elle à espacer les sillons de façon à limiter le risque d'un retard de faible amplitude sur un train ne se propage au train suivant. Son introduction dans la grille horaire entraîne également une consommation supplémentaire de capacité.

L'allocation de ces marges dans les grilles horaires dans un contexte de saturation doit par conséquent se faire de façon judicieuse et parcimonieuse. Des arbitrages doivent être réalisés à la fois concernant la part de capacité dédiée à l'introduction de marges ainsi que la façon dont cette capacité est répartie entre marge sillon et marge de grille.

Nous avons tout d'abord évalué l'impact d'une réduction ou d'une réorganisation de la marge de grille. Pour ce faire, nous avons testé trois grilles horaires, construites à partir de celle de 2019 mais présentant des structures différentes en heure de pointe :

- La grille « Respiration » est la grille 2019, où les sillons dans le tronçon commun Saint-Cloud - La Défense ont été réorganisés de façon à aménager un sillon de respiration (voir chapitre 1) toutes les 15 minutes. Ceci s'est fait en rapprochant les circulations de part et d'autre des sillons de respiration, les trains de la ligne U (trains DEFI) ayant leurs horaires inchangés (en raison de leur insertion dans le trafic des lignes C et N en dehors du système considéré). Les sillons ont également dû être tendus de façon à produire une grille sans conflit (c'est-à-dire telle que dans le mode de fonctionnement nominal, les trains circulent à voie libre).
- La grille « Dense » est construite à partir de la grille « Respiration », mais en la densifiant : les sillons de respiration sont désormais utilisés par des circulations.
- La grille « 2019 tendue » est la grille 2019 dont les sillons ont été tendus de la même façon que pour les grilles précédentes.

On trouvera à la figure 4.8 un schéma représentant la succession des circulations à l'entrée dans le tronçon commun en heure de pointe, selon une trame se répétant toutes les 30 minutes.

A des fins de comparaison, nous avons également inclus la grille 2019 originale dans l'évaluation. Celle-ci est représentée, sous la forme de graphiques espace-temps, en annexe à la figure 4.12 tandis que les trois grilles modifiées le sont à la figure 4.13.

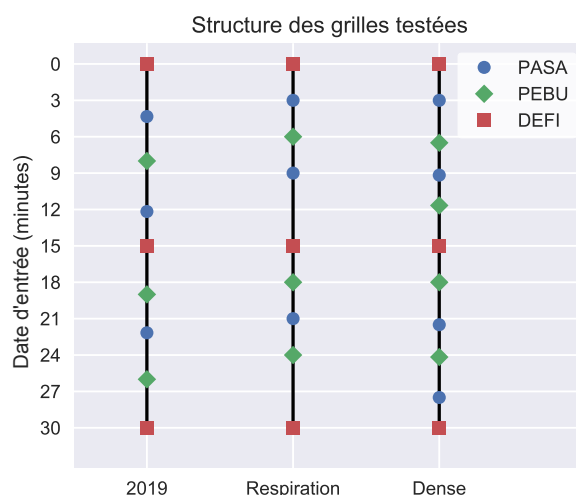


Figure 4.8 – Structure des grilles « 2019 », « Respiration » et « Dense »

La figure 4.9 représente les distributions des performances de ces quatre grilles sur l'ensemble des réplifications. Celle-ci nous conduit à formuler les observations suivantes :

- La réduction de la marge sillon a permis d'améliorer légèrement les performances dans tous les cas de figure. Cette amélioration est néanmoins significative, étant donné le caractère fortement contraint du système. Ce résultat corrobore les affirmations de Carré (2014) et TransportParis (2016), selon lesquelles la marge sillon est source de variabilité et par conséquent d'instabilité.
- En revanche, la grille 2019 offre de bien meilleures performances de régularité que les trois autres, ce qui montre à nouveau que le retard ne semble pas un bon indicateur de la qualité de service dans un contexte de Mass Transit.
- L'écart de performance entre la grille « 2019 tendue » et la grille « Respiration » semble montrer qu'il est préférable de répartir uniformément la marge de grille entre les différents trains plutôt que sous la forme de sillons de respiration.
- La performance en mode nominal n'est pas un bon indicateur de qualité pour une grille horaire. En effet, des trois grilles « tendues », la grille dense est celle qui offre la meilleure vitesse en mode nominal, mais la plus faible lorsque l'on tient compte des petits aléas.
- Il semblerait que l'on puisse obtenir des performances satisfaisantes avec une grille plus dense, et moins de marge sillon pour permettre plus de marge de grille. En effet, les éventuels ralentissements en ligne causés par la plus grande utilisation de l'infrastructure sont compensés par un temps d'attente réduit en gare. Par ailleurs, l'augmentation des fréquences permet de mieux répartir les voyageurs dans les trains, réduisant ainsi la variabilité des durées de stationnement ainsi que le phénomène d'instabilité lié à l'affluence de voyageurs. Cependant, une telle grille est probablement plus sensible aux perturbations de plus forte amplitude ainsi qu'aux ralentissements généralisés (pour cause de mauvaise météo par exemple) du fait de la marge de grille réduite.
- Ces résultats doivent en outre être interprétés avec prudence, étant donné que les écarts de performance sont principalement dûs aux différences en heure de pointe. Or, l'heure de pointe du matin est aussi la période où les estimations de SISTRAM sont plus imprécises.

Enfin, nous avons constaté que la réduction de la marge sillon semble améliorer la qualité de service offerte aux voyageurs. Nous avons cherché à quantifier ce phénomène de façon plus précise.

La première étape a été de constituer une grille avec une marge sillon minimale. Pour ce faire, nous avons cherché, pour chaque activité (déplacement ou stationnement) existant dans la grille

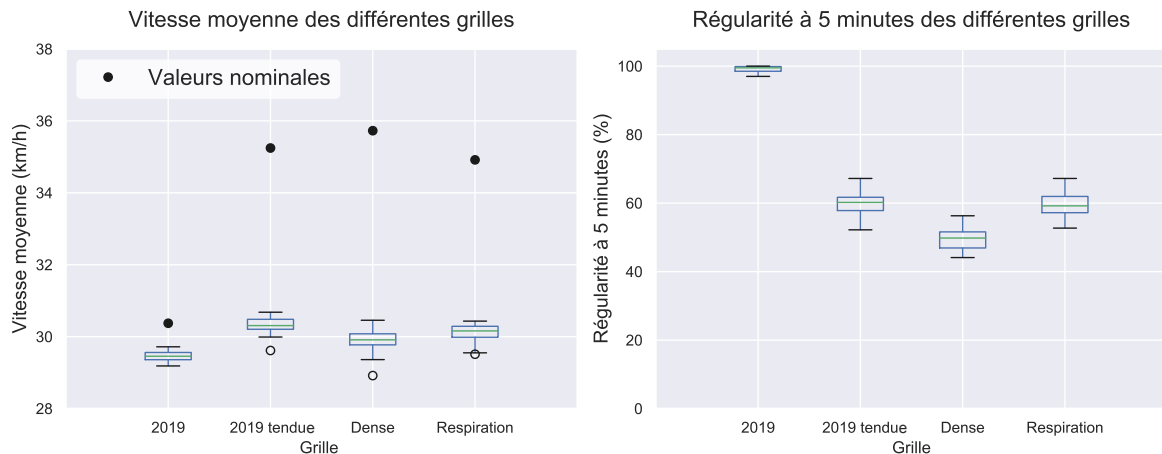


Figure 4.9 – Performance sur les différentes réplifications des grilles testées

2019, la durée minimale de réalisation de cette activité dans le jeu de données. Nous constituons alors un sillon minimal pour chaque train en concaténant les durées de chaque activité de celui-ci. L'ensemble de ces sillons forme alors une grille dont la marge sillon est inexistante. Notons au passage qu'une telle grille n'est absolument pas réalisable en pratique, les temps de parcours minimaux étant généralement obtenus en utilisant les capacités de freinage et d'accélération du matériel roulant au-delà des valeurs recommandées (ce qui peut provoquer, par exemple, une usure prématurée ou de l'inconfort pour les voyageurs). Cette grille n'a donc valeur que de référence et de base pour la suite.

Nous avons ensuite distribué, à chaque train et de façon homogène le long de son parcours, une marge sillon exprimée en pourcentage de son temps de trajet minimal (au sens défini ci-dessus). Nous avons fait varier ce pourcentage et évalué les performances des grilles ainsi obtenues, en termes de vitesse moyenne pour les voyageurs et de régularité au terminus. Celles-ci sont représentées à la figure 4.10. Notons que la grille « 2019 tendue » étudiée précédemment a aussi été obtenue par ce procédé, avec une marge sillon de 20%. A titre de comparaison, dans la grille 2019, tous les trains ne se voient pas attribuer le même pourcentage de marge ; celui-ci varie selon les trains de 32% à 56%.

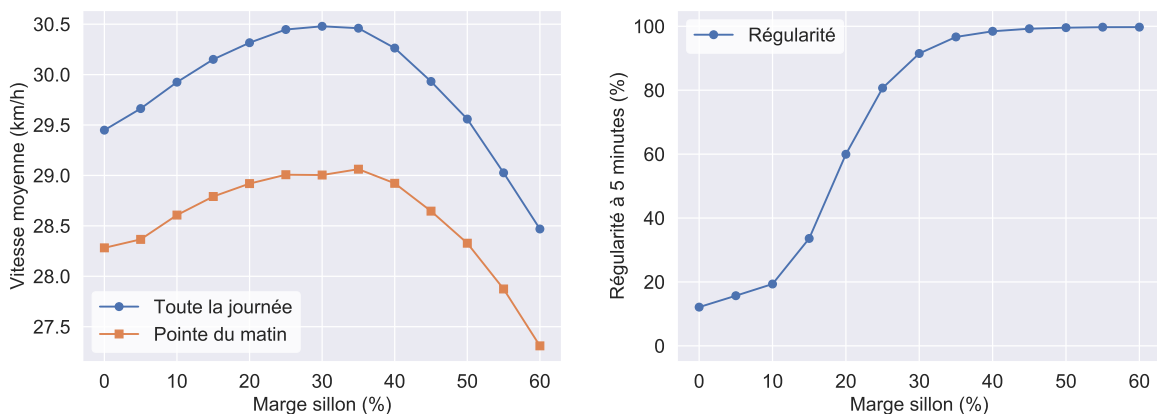


Figure 4.10 – Influence de la distribution de marge sillon sur la qualité de service

On constate que lorsque la marge sillon est très faible, les trains ont tendance à trop s'éloigner de leur sillon nominal. Cela complique la synchronisation du franchissement des convergences et provoque des phénomènes d'instabilité lorsque les intervalles temporels entre trains se réduisent.

À l'inverse, une marge trop importante entraîne des allongements inutiles de temps de parcours. Entre les deux, un optimum semble atteint autour de 30%, soit légèrement moins que la marge sillon actuellement distribuée dans la grille 2019. Cette réduction de la marge sillon est en effet compensée par davantage de marge de grille, celle-ci semblant dans cette situation plus efficace pour atténuer les effets des petits aléas. Nous poussons plus loin cette recherche d'optimum dans le chapitre 5.

4.6.5 Approche « robustesse avec recours »

Les travaux présentés dans ce paragraphe ont été menés en collaboration avec Shidi Deng et ont donné lieu à la rédaction d'un mémoire de fin d'études (Deng, 2019).

Dans les travaux présentés jusqu'ici, un phénomène a été négligé. En effet, lorsque les conditions de circulation se dégradent et que le trafic ferroviaire devient instable, des mesures de régulation sont appliquées de façon à rétablir la fluidité du trafic. À SNCF Transilien, celles-ci peuvent être de natures diverses : ajout d'arrêts supplémentaires, suppressions de trains sur la totalité de leur parcours ou sur une partie. Celles-ci constituent ce que l'on appelle des « recours » dans le vocabulaire de l'optimisation robuste : elles peuvent permettre de récupérer une solution de qualité acceptable dans une situation dégradée. Ces décisions sont prises par les agents du COT (Centre Opérationnel Transilien) qui supervisent l'état du trafic via une représentation sous la forme de Graphique Espace-Temps en temps réel, ainsi que des échanges radio et téléphoniques avec les différentes parties prenantes de l'exploitation.

Dans la lignée des travaux présentés au chapitre 2 portant sur la « robustesse avec recours », nous avons cherché à reproduire ces décisions en simulation, principalement afin de déterminer si celles-ci peuvent modifier significativement le score des grilles horaires voire la hiérarchie entre celles-ci. Cependant, ces décisions de recours s'appuient encore largement aujourd'hui sur l'expérience et l'expertise des GTI, même si des outils d'aide à la décision commencent à émerger (Altazin, 2018). Il est donc difficile de reproduire ce processus de prise de décision de la même manière que celui-ci se déroulerait lors d'une véritable journée d'exploitation.

Toutefois, un référentiel de règles est mis à disposition dans certains COT afin d'encadrer ces décisions. Nous avons implémenté une heuristique de régulation construite à partir de règles issues du référentiel du COT de Paris-Saint-Lazare pour le système L2-U. Ce document énonce les principes suivants pour la grille horaire 2019 :

1. Des suppressions de trains sont applicables lorsque des trains circulent avec un retard (effectif ou projeté) supérieur à l'intervalle nominal de la branche d'où ils proviennent. On rappelle qu'en heure de pointe, cet intervalle est en moyenne de 10 minutes pour les deux branches de la ligne L2 et de 15 minutes pour les trains de la ligne U.
2. Le premier et le dernier train de la journée sur chaque branche ne peuvent pas être supprimés.
3. Les suppressions doivent limiter autant que possible les trous de desserte. En conséquence, lorsqu'un train PEBU est supprimé, le train PASA suivant se voit ajouter des arrêts en gares de Puteaux, Courbevoie et Bécon-les-Bruyères.
4. Les décisions de supprimer les trains de la ligne U sont prises par le COT de Paris-Montparnasse, principalement en fonction des événements survenant sur leur périmètre.

Nous avons traduit informatiquement ces principes en les procédures suivantes :

1. Dès lors qu'un train est détecté comme circulant avec un retard supérieur à un seuil s (fixé comme paramètre), le prochain train non encore parti de sa gare d'origine est supprimé.
2. Le premier et le dernier train de la journée sur chaque branche ne peuvent pas être supprimés.

3. Lorsqu'un train PEBU est supprimé, le train PASA suivant se voit ajouter des arrêts en gares de Puteaux, Courbevoie et Bécon-les-Bruyères. On ne peut pas supprimer deux trains consécutifs de la même branche.
4. Les trains de la ligne U ne peuvent pas être supprimés.

Par ailleurs, nous faisons l'hypothèse que lorsqu'un train est supprimé, tous les voyageurs qui souhaitent l'emprunter se sont rendent en gare à la même heure qu'en l'absence de suppression, et attendent le passage du train suivant leur permettant de se rendre à leur gare de destination. Leur temps d'attente est ainsi augmenté en fonction.

Cette heuristique de régulation a été testée sur différentes grilles, dont la grille 2019 et la grille dense, afin d'évaluer si celle-ci influe significativement sur leurs performances. La figure 4.11 donne la valeur de la vitesse moyenne pour les passagers ainsi que le pourcentage de trains supprimés pour différentes valeurs du seuil au-delà duquel se déclenchent des suppressions.

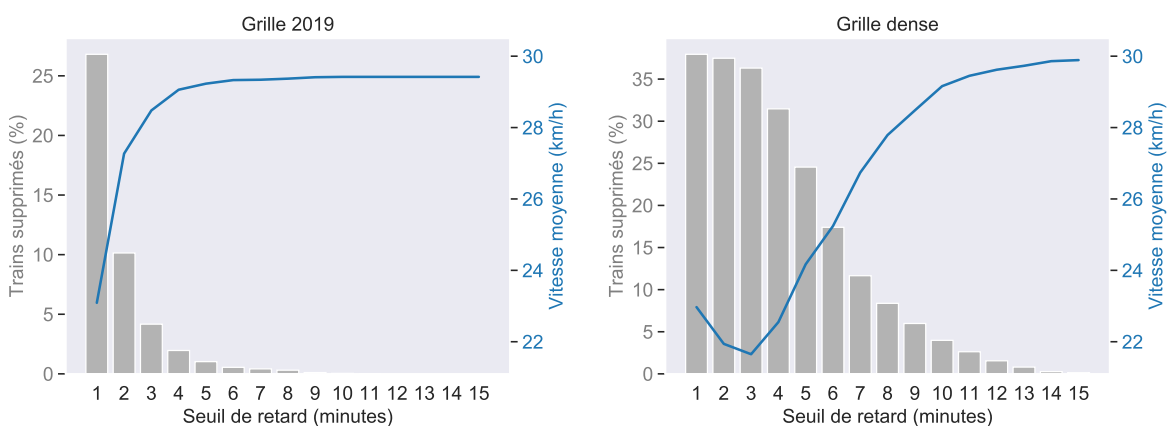


Figure 4.11 – Nombre de trains supprimés et effet sur la vitesse moyenne en fonction du seuil

Les figures montrent que l'heuristique de suppression ne permet pas d'améliorer les performances des grilles testées. En effet, les gains de fluidité permis par la suppression de trains sont compensés par des temps d'attente prolongés pour les voyageurs concernés, ainsi que des durées de stationnement plus importantes pour les trains suivants en raison du report des flux de voyageurs. L'application automatique des règles en vigueur dans les COT ne permet donc pas de restituer pleinement en simulation l'effet des mesures de régulation sur la stabilité du trafic. Il pourrait néanmoins être pertinent, concomitamment à l'industrialisation d'outils d'aide aux décisions de régulation, de tenir compte de leurs possibilités dans les futures études de robustesse de plans de transport.

4.7 Conclusion

Nous avons présenté dans ce chapitre une méthode de simulation stochastique du trafic ferroviaire s'appuyant sur des algorithmes d'analyse de données. Il s'agit d'une tendance qui commence à émerger et qui pourrait être amenée à se développer. En effet, nous avons montré que notre outil permet de reproduire précisément la distribution des durées de parcours des trains sous l'effet d'aléas d'exploitation dont les distributions conditionnelles sont fournies. En outre, la durée de calcul nécessaire à la réalisation d'une réplication est significativement inférieure à celle requise par les simulateurs microscopiques classiques. Nos résultats confirment donc l'observation faite par Watanabe et al. (2018) : l'utilisation de données d'exploitation permet de s'affranchir d'une modélisation fine de l'infrastructure et du mouvement des trains, et ainsi d'accélérer les calculs, sans dégrader la précision des résultats.

Cet algorithme nous a également permis d'évaluer certains grands principes de conception de grilles horaires et la capacité des grilles correspondantes à apporter aux passagères une qualité de service satisfaisante sous l'effet de petits aléas. Nous observons d'une part, une faible corrélation entre la régularité train et la qualité de service, d'autre part une faible corrélation entre les qualités de services fournies en régime nominal et sous l'effet d'aléas. Ceci semble confirmer qu'en zone très dense, la régularité train ou d'éventuels mesures de performance nominaux ne constituent pas des indicateurs satisfaisants de la qualité de service pour les voyageurs. Enfin, dans une première démarche d'optimisation, nous avons mis en évidence l'existence d'un équilibre optimal entre marge sillon et marge de grille, qui pourrait éclairer le choix des futures normes de conception horaire en zone dense.

4.8 Annexe

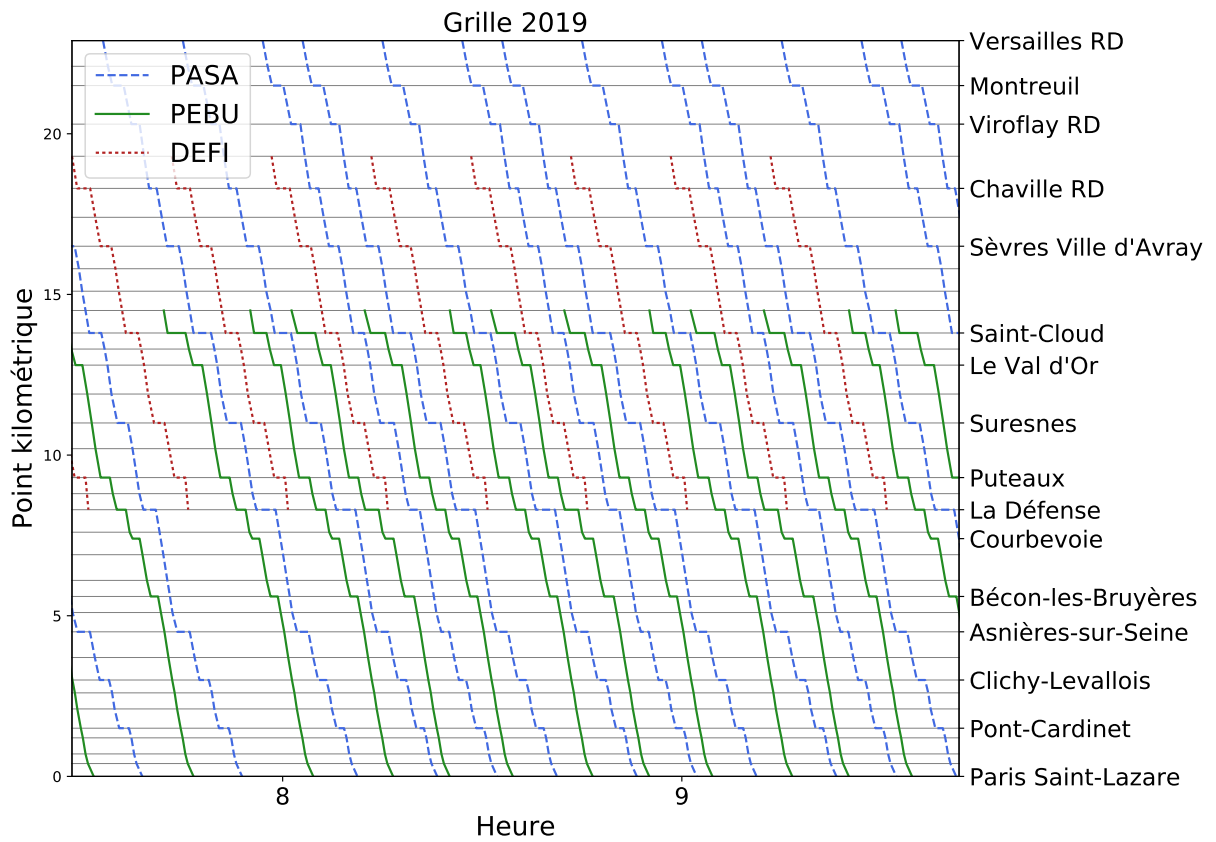


Figure 4.12 – Graphique Espace-Temps nominal de la grille 2019

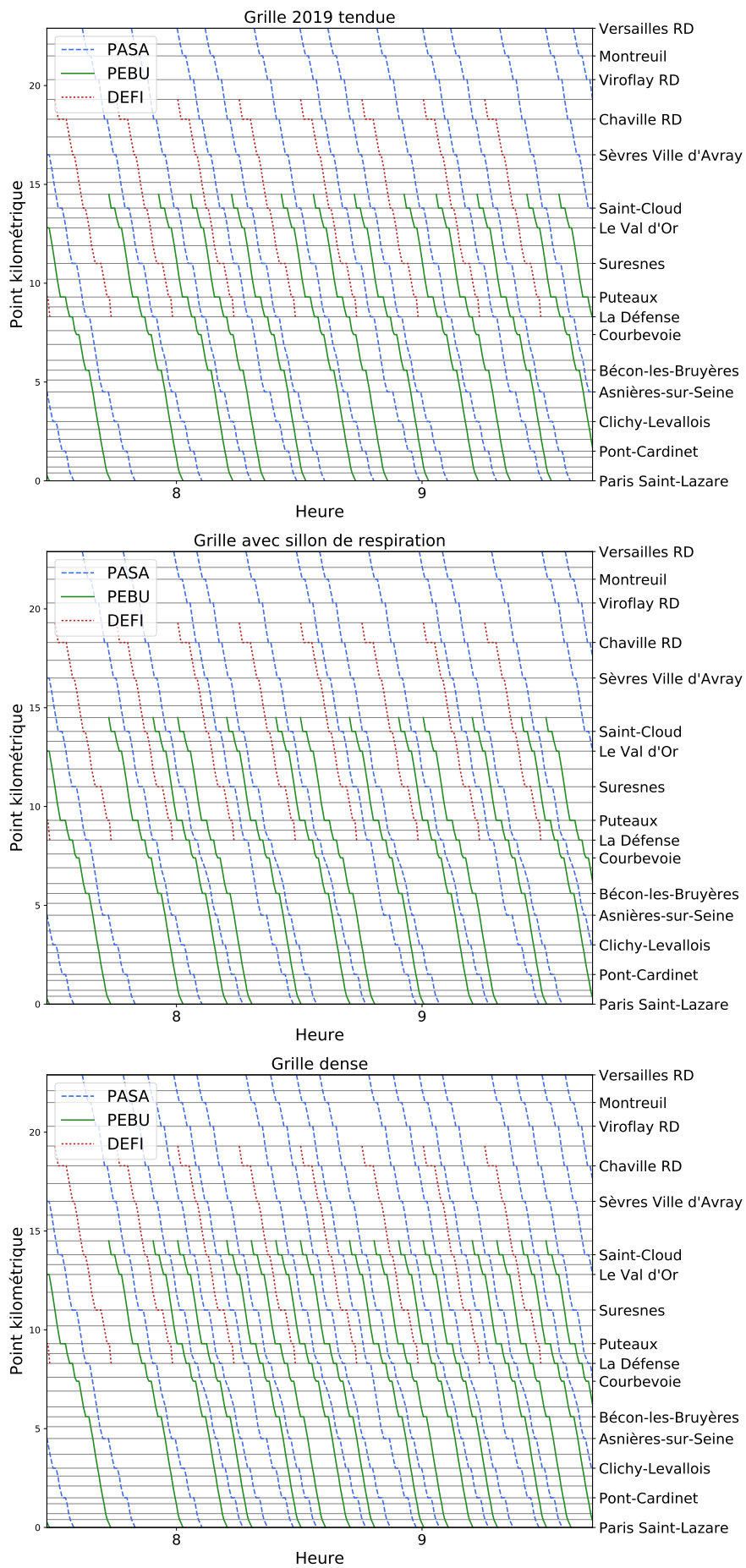


Figure 4.13 – Graphiques Espace-Temps nominaux des grilles testées

Chapitre 5

Une simheuristique pour générer des grilles robustes

Résumé

Ce chapitre vient conclure la démarche de résolution du problème formulé au chapitre 2. Une approche couplée optimisation-simulation, utilisant l'algorithme SISTRAM du chapitre précédent, est présentée pour la génération de grilles horaires performantes. Celle-ci s'appuie sur des méthodes heuristiques d'optimisation. Nous présentons une revue de littérature sur ces méthodes, leurs couplages avec des algorithmes de simulation, et leurs applications, notamment en transport ferroviaire. Nous décrivons ensuite l'articulation choisie entre modules d'optimisation et de simulation. Plusieurs méthodes d'optimisation sont testées, et leurs avantages et inconvénients respectifs sont mis en évidence.

Sommaire

5.1	Présentation et état de l'art	116
5.1.1	Motivation	116
5.1.2	Méta-heuristiques	116
5.1.3	Approches couplées simulation-optimisation	118
5.2	Une simheuristique de recuit simulé	120
5.2.1	L'algorithme du recuit simulé	120
5.2.2	Adaptation et implémentation	121
5.2.3	Résultats	122
5.3	Méthodes de descente	125
5.4	Conclusion	127

« Mes trains sont à l'heure parce qu'ils sont tendus. »
Albert Sartiaux (1845 - 1921), directeur d'exploitation de la Compagnie du Nord

5.1 Présentation et état de l'art

5.1.1 Motivation

Nous avons présenté au chapitre 2 un modèle décrivant le problème de la construction d'une grille horaire offrant une qualité de service *optimale*. Nous nous inscrivons donc dans une démarche d'optimisation. Un premier pas dans cette direction a été fait au chapitre 4, section 4.6.3. où nous avons proposé une méthode pour une distribution optimale de la marge sillon dans l'hypothèse où celle-ci est identique pour tous les trains de la grille. Il s'agissait alors d'un problème d'optimisation à une seule variable sur un domaine restreint, dont l'optimum peut facilement être approximé par essais successifs.

Cependant, tous les trains de la grille n'effectuent pas la même mission, ni ne font face à la même demande selon l'heure de la journée. L'approche précédente est par conséquent assez grossière et susceptible d'être raffinée, par exemple en faisant varier la distribution de marge sillon d'un train à l'autre. Rappelons que c'est déjà le cas en pratique : dans la grille actuelle, les trains possèdent une marge allant de 32% à 56% par rapport à la marche la plus tendue possible. Nous souhaitons par conséquent rechercher une distribution de marge optimale face aux petits aléas dont nous avons estimé les lois de probabilité. Si l'on considère toujours que chaque train voit sa marge distribuée de façon uniforme le long de son parcours, le nombre de variables de ce nouveau problème est de l'ordre de quelques centaines. Il n'est donc plus possible de procéder par essais successifs, et une méthode d'optimisation ad hoc est requise.

L'évaluation des grilles horaires est réalisée par SISTRAM, dont les résultats ne peuvent être résumés sous la forme d'une formule fermée. Les approches classiques relevant de la programmation mathématique sont donc inopérantes ici. C'est pourquoi il est nécessaire de se tourner vers des méthodes adaptées au cas où la fonction d'évaluation est une « boîte noire » du point de vue de l'algorithme d'optimisation. Parmi elles, on trouve en bonne place les heuristiques (ou méthodes approximatives), c'est-à-dire des algorithmes n'ayant souvent aucune garantie de performance mais dont la conception laisse penser qu'ils permettront d'obtenir des solutions proches de l'optimum. Ce terme est d'acception large et englobe des algorithmes spécialement conçus pour un problème spécifique en exploitant la structure particulière de l'espace des solutions. Cependant, il existe également des classes d'heuristiques à portée générale, pouvant s'appliquer à une large variété de problèmes, éventuellement moyennant quelques adaptations. On parle dans ce cas de **méta-heuristiques**.

5.1.2 Méta-heuristiques

La plupart des heuristiques développées jusque dans les années 1980 retournaient comme résultat une solution réalisable correspondant à un optimum local du problème considéré. Dans les années qui ont suivi, des méthodes approximatives ayant l'ambition de surmonter cet inconvénient se sont développées. Parmi elles, un certain nombre de schémas généraux d'optimisation ont fait les preuves de leur efficacité sur une large classe de problèmes (y compris pour certains relevant du transport ferroviaire, nous y reviendrons).

Ces algorithmes doivent permettre, au cours de leur exploration de l'espace des solutions, de s'extraire des vallées conduisant à des minima locaux. Ils combinent par conséquent deux mécanismes :

- Un mécanisme de **diversification**, contraignant l'algorithme à explorer de nombreuses régions de l'espace des solutions. Cela impose en particulier d'accepter de parfois se déplacer vers une solution qui dégrade la valeur de la fonction objectif : ceci est indispensable au franchissement des cols qui entourent les optima locaux.
- Un mécanisme d'**intensification**, qui doit assurer que lors de l'exploration d'une vallée, l'optimum local qui y est contenu est approché avec une précision suffisante.

La combinaison de ces deux mécanismes doit alors permettre d'obtenir une bonne approximation de l'optimum global. Ce principe général est résumé à la figure 5.1.

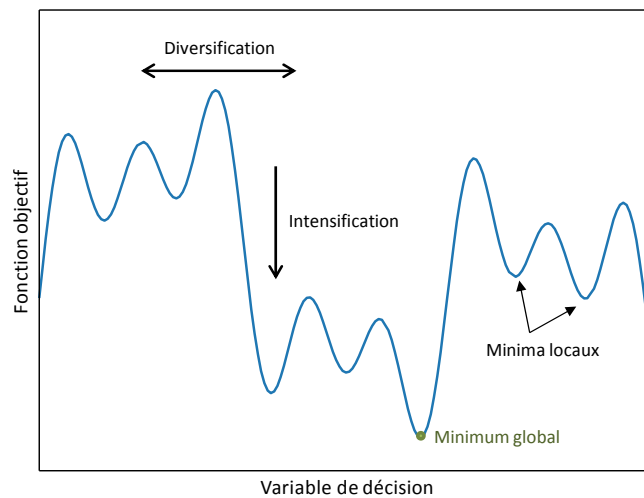


Figure 5.1 – Principe de fonctionnement d'une méta-heuristique de minimisation

Les mécanismes des méta-heuristiques s'inspirent souvent d'observations provenant d'autres domaines (métallurgie, biologie, éthologie...). Teghem (2012) en propose une classification selon deux grandes familles :

- Les algorithmes de recherche locale. Ceux-ci consistent en une exploration à partir d'une unique solution initiale, par déplacements successifs au sein du voisinage de la solution courante. Les plus connus sont la recherche tabou, qui consiste à interdire l'exploration de solutions présentant certaines caractéristiques dès lors que d'autres les présentant également ont été explorées, et le recuit simulé, inspiré d'une technique de métallurgie. Nous reviendrons sur ce dernier par la suite.
- Les algorithmes évolutionnaires. Ceux-ci considèrent, à chaque itération, non pas une seule solution mais toute une population de solutions courantes. Celle-ci est mise à jour à chaque itération selon des mécanismes propres à chaque méta-heuristique de cette catégorie. La plus connue porte le nom d'algorithme génétique. Chaque solution y est identifiée à un chromosome, dont les caractéristiques constituent les gènes. De nouveaux chromosomes sont enfantés en croisant deux parents choisis parmi ceux présentant les meilleures valeurs pour la fonction objectif. Un mécanisme de mutation aléatoire des gènes est ajouté pour permettre la diversification. Outre l'algorithme génétique, les algorithmes dits de colonies de fourmis ou d'essaims particuliers relèvent également de cette catégorie.
- Les méta-heuristiques hybrides, combinant les mécanismes des deux catégories précédentes.

Certaines méta-heuristiques particulières (GRASP, VNS...) ne relèvent d'aucune des catégories précédentes.

Plusieurs applications de méta-heuristiques à des problèmes de planification ou replanification ferroviaire ont été proposées dans la littérature. Törnquist et Persson (2005) comparent ainsi les performances de deux algorithmes de reroutage, combinant programmation linéaire et, respectivement, recherche tabou et recuit simulé. L'algorithme utilisant la recherche tabou permet selon eux d'obtenir une réduction du retard total plus importante qu'avec le recuit simulé. Hanafi et Kozan (2014) proposent également un algorithme hybride utilisant le mécanisme du recuit simulé pour construire des plannings de roulement pour le personnel de bord.

Les approches évolutionnaires sont toutefois en majorité dans le domaine ferroviaire. La plus connue d'entre elles, l'algorithme génétique, est appliquée par Tormos et al. (2008) au problème de

la construction de grilles horaires pour voie unique. Le problème de la constitution de la population initiale, souvent crucial avec ce type d'algorithme, fait lui-même l'objet d'une heuristique dédiée. La construction d'une solution réalisable est assimilée au choix d'un parcours dans un arbre, un échantillonnage aléatoire permet alors d'obtenir plusieurs solutions. Ceci leur permet d'obtenir des grilles offrant des performances satisfaisantes avec un temps de calcul réduit. Cet algorithme est également utilisé par Arenas et al. (2015) pour la résolution du *Train Timetabling Problem*, la population initiale étant cette fois générée grâce à une heuristique gloutonne.

L'algorithme de la colonie de fourmis a également fait l'objet de déclinaisons dans ce cadre. Cet algorithme fut initialement proposé pour des problèmes de recherche de plus court chemin (en s'inspirant de la technique mise en place par les fourmis pour trouver le plus court chemin entre la fourmilière et une source de nourriture déjà localisée). Il a été par la suite adapté à d'autres problèmes, dont le problème TTP pour la première fois par Su et Huang (2006) dans un contexte de Mass Transit. L'algorithme est utilisé pour produire des grilles horaires périodiques minimisant le nombre de conflits et l'irrégularité des espacements temporels entre trains. La question de la robustesse n'y est cependant pas abordée. Samà et al. (2016) appliquent quant à eux cet algorithme au problème de la sélection des itinéraires qui seront considérés en temps réel pour le reroutage et l'ordonnement des trains. L'évaluation de l'approche sur des instances du réseau ferré français montre des améliorations notables de la réduction du retard par rapport à d'autres méthodes de sélection des alternatives d'itinéraires.

Enfin, quelques approches s'appuient sur la méthode des essaims particuliers. À l'instar de l'algorithme des colonies de fourmis, celle-ci s'inspire d'un principe de collaboration entre individus. Chaque individu de la population se déplace dans l'espace de recherche à chaque itération, en fonction de sa meilleure position connue, de la meilleure position trouvée par ses voisins, ainsi que d'un paramètre de « vitesse » représentant l'amplitude des déplacements permis. Meng, Jia et Qin (2010) proposent un algorithme reposant sur ce principe, utilisable à la fois pour les problèmes de planification en amont et de replanification en temps réel, dans un contexte de Mass Transit. Ho et al. (2012) l'appliquent quant à eux à un problème encore relativement peu étudié, mais qui sera amené à l'être dans les années qui viennent : la génération de grilles horaires pour un réseau ouvert à la concurrence en *open-access*. Le nombre de parties prenantes y étant en effet élevé, la complexité du problème l'est d'autant, ce qui justifie le recours à des heuristiques. Enfin, Jamili et al. (2012) proposent un algorithme hybride, reprenant les mécanismes du recuit simulé et des essaims particuliers, pour la construction de grilles horaires périodiques sur voie unique.

5.1.3 Approches couplées simulation-optimisation

Le développement de la puissance de calcul des ordinateurs permet, depuis la fin du XX^{ème} siècle, de développer des méthodes hybrides faisant fonctionner conjointement des algorithmes de simulation et d'optimisation (Tekin et Sabuncuoglu, 2004). Celles-ci se déclinent en trois configurations :

- L'optimisation au service de la simulation. La plupart des méthodes de simulation requièrent le choix d'un certain nombre de paramètres, qui font l'objet d'une calibration. Celle-ci est généralement effectuée en recherchant le jeu de paramètres permettant d'obtenir l'écart le plus faible possible entre des données issues d'observations du phénomène réel et les résultats de simulations effectuées à partir de la même configuration initiale. Des méthodes d'optimisation peuvent alors être mises en œuvre pour cette identification de paramètres. Ce champ de recherches est plus généralement rattaché à celui des problèmes inverses et de l'assimilation de données, voir par exemple Tarantola (2005).
- L'optimisation comme outil de prise de décision automatique en cours de simulation. De nombreux systèmes requièrent ou permettent des interventions humaines au cours de leur fonctionnement, par exemple en cas de situation perturbée. C'est notamment le cas du transport ferroviaire où des régulateurs prennent des mesures correctives lorsque l'état du trafic

le nécessite. Celles-ci visent à optimiser un certain critère (tel que le coût total de la perturbation ou la durée avant le retour à la normale). Un module embarquant des algorithmes d'optimisation peut alors permettre de reproduire ces décisions dans un outil de simulation. Cette approche permet également, à l'inverse, d'évaluer la qualité des décisions proposées par certains algorithmes d'optimisation.

- La simulation comme outil au service de l'optimisation. Cette stratégie est notamment utilisée lorsque la fonction objectif ou certains paramètres d'un problème d'optimisation ne peuvent être calculés de façon exacte. C'est de cette approche que nous traitons dans ce chapitre.

Relativement boudée jusqu'alors, la recherche sur l'application au transport ferroviaire de stratégies relevant du troisième point a pris son essor au cours des dernières années. Plusieurs travaux menés au sein de SNCF Innovation & Recherche s'inscrivent dans cette dynamique. Le projet « Convergences » (Ramond, Pozzoli et Marcos, 2017) a conduit au développement d'un outil d'aide à la décision lorsque deux trains sont en conflit pour le franchissement d'une convergence. Des indicateurs de ponctualité sont calculés au moyen d'une simulation dans les deux scénarios d'ordonnancement possibles et celui obtenant le meilleur score est choisi. L'espace des solutions ne comporte ici que deux éléments, l'algorithme d'optimisation est donc extrêmement simple. Une approche plus élaborée a été développée par Altazin et al. (2020) dans le cadre du projet « Stabilité », toujours pour la gestion du trafic en temps réel. Différents scénarios de régulation sont évalués à l'aide d'un algorithme de simulation simple, reposant sur le principe de calcul de chemins dans un graphe d'événements-activités. L'originalité de l'approche réside notamment dans l'utilisation d'une heuristique gloutonne multi-objectifs pour la sélection des scénarios de régulation. Le temps de trajet généralisé des passagers constitue d'ailleurs l'une des fonctions objectif à optimiser. La simulation est déterministe dans ces deux travaux.

Burggraave et Vansteenwegen (2017) adoptent également un point de vue orienté vers les passagers, cette fois pour la construction de plannings robustes au voisinage des grandes gares. Leur algorithme est itératif et se décompose en 4 étapes : calcul d'un itinéraire optimal pour chaque train de façon à répartir le trafic uniformément sur tous les quais de la gare ; calcul d'horaires en maximisant la marge de grille minimale calculée sur l'ensemble des trains ; simulation stochastique microscopique du planning obtenu ; recalcul des durées de parcours et de stationnement de façon à ce qu'un quantile prédéterminé de trains qui étaient en retard soient à l'heure avec le nouveau planning. Les variables aléatoires sont tirées dans des distributions exponentielles. Une démarche itérative assez proche est mise en œuvre par Lee et al. (2017) : un programme linéaire est proposé pour la construction de grilles robustes à l'échelle d'un réseau. Les paramètres de ce modèle sont recalculés à l'issue d'une simulation de la grille obtenue et ces étapes sont répétées jusqu'à convergence. Le point de vue adopté est celui de l'opérateur, et les auteurs ne détaillent pas la définition des perturbations introduites en cours de simulation. Cette approche a récemment été poussée plus loin par Högdahl, Bohlin et Fröidh (2019) : leur fonction objectif intègre un terme représentant l'espérance du retard calculé par des simulations stochastiques microscopiques utilisant le logiciel RailSys. Les distributions d'aléas sont estimées à partir de données d'exploitation mais supposées indépendantes entre elles.

Les « simheuristiques », dont la définition a été proposée par Juan et al. (2015), constituent une classe particulière d'algorithmes d'optimisation utilisant la simulation. Les auteurs désignent par ce terme toute métaheuristique d'optimisation dont la fonction objectif est calculée par un appel à un outil de simulation. Les applications de ce concept ont été relativement variées au cours des dernières années : production, logistique, santé... Nous en avons identifié deux déclinaisons pour des problèmes de planification ferroviaire. Hassannayebi, Sajedinejad et Mardani (2014) s'intéressent au problème de la construction d'horaires pour une ligne de métro sans convergences. Les variables de décision sont ici les fréquences des départs des stations aux extrémités de la ligne. Un algorithme génétique, faisant appel à un outil commercial de simulation stochastique, est utilisé pour le calcul des intervalles optimaux. L'aléa porte sur les temps de parcours interstations, supposés suivre une loi

normale dont les valeurs inférieures au minimum réalisable sont écartées. Les durées de stationnement sont supposées proportionnelles aux flux de voyageurs, le nombre de voyageurs montants étant lui-même calculé à partir des intervalles entre trains successifs en supposant un taux d'arrivée en station constant. Chow et Pavlides (2018) proposent quant à eux une approche multi-objectifs, où sont pris en compte les durées de parcours des trains, les durées d'attente en gare pour les passagers, le taux de remplissage des trains et les retards. Un algorithme génétique est là aussi utilisé, faisant appel à un algorithme de calcul de chemins dans un graphe pour l'évaluation des solutions. Plusieurs réplications de cet algorithme sont effectuées, pour différentes valeurs de temps de parcours tirées selon des lois normales, les poids des arcs du graphe étant mis à jour en fonction des tirages réalisés.

5.2 Une simheuristique de recuit simulé

5.2.1 L'algorithme du recuit simulé

La fonction objectif du modèle présenté au chapitre 2 est formulée comme l'espérance d'une variable aléatoire, construite à partir de réalisations d'autres variables aléatoires non indépendantes entre elles. Son évaluation est par conséquent délicate, et l'outil SISTRAM présenté au chapitre 4 a été développé pour remplir cette fonction. Les simheuristiques constituent ainsi un cadre particulièrement adapté à la résolution du problème du chapitre 2.

Wolpert et Macready (1997) ont prouvé qu'il n'existe pas de méta-heuristique intrinsèquement meilleure ou moins bonne que les autres, leurs performances dépendant du problème considéré ainsi que du choix des paramètres. Le choix d'une méta-heuristique peut donc comporter une part d'arbitraire ou de considérations pratiques. Nous avons choisi d'utiliser une heuristique de recherche locale, qui facilite la phase d'initialisation et nous paraît avoir davantage de chances de succès. En effet, il est souhaitable que la ou les solution-s initiale-s (selon que l'on utilise une recherche locale ou un algorithme évolutionnaire) offrent déjà des performances relativement satisfaisantes. Cela permet en effet d'accroître la probabilité de converger vers un optimum global, ainsi que la rapidité de convergence (Teghem, 2012). Or, nous disposons déjà d'une telle grille pour le cas test choisi : la grille 2019. En revanche, la génération *ex nihilo* d'une population entière de grilles horaires offrant des résultats satisfaisants et présentant des caractéristiques suffisamment variées pour permettre une bonne diversification est un problème non trivial. Les algorithmes de recherche locale offrent par ailleurs l'avantage d'être des algorithmes « anytime », c'est-à-dire qu'il est possible d'obtenir une solution utilisable même en interrompant l'algorithme avant sa terminaison (par exemple si le temps de calcul est trop long, si on a déjà obtenu une solution suffisamment satisfaisante ou si la fonction objectif ne s'améliore plus).

Comme toutes les méta-heuristiques, les algorithmes de recherche locale nécessitent de se fixer un encodage des solutions, souvent sous la forme d'un vecteur ou d'une matrice. Ils requièrent par ailleurs de se munir d'un critère permettant de déterminer, pour toute solution réalisable x , quelles sont les autres solutions réalisables considérées comme étant au voisinage de x .

Nous avons choisi l'algorithme dit du recuit simulé, pour sa facilité et sa souplesse d'implémentation ainsi que son caractère « prêt à l'emploi ». En effet, la recherche tabou (qui est l'autre algorithme « classique ») de recherche locale, nécessite de se fixer une méthode d'optimisation au sein de chaque voisinage considéré (ou d'un sous-voisinage si ce dernier est trop grand). À l'inverse, le recuit simulé se contente d'un tirage aléatoire d'une nouvelle solution dans chaque voisinage. Nous présentons plus en détails le principe de fonctionnement de cette méta-heuristique ci-dessous.

Cet algorithme est inspiré par le procédé métallurgique du même nom. Lorsqu'un matériau a subi des déformations, ce procédé permet de lui redonner sa forme initiale. Le matériau est d'abord chauffé à très haute température afin de faire disparaître les tensions internes causées par les déformations. Il est ensuite refroidi très lentement, ce qui permet aux atomes de se placer dans une configuration stable.

Par analogie, un algorithme de recuit simulé nécessite la donnée d'une solution initiale x_0 , d'une température initiale $T_0 > 0$ et d'un mécanisme de refroidissement. A chaque itération, une solution est choisie aléatoirement dans le voisinage de la solution courante et est évaluée. Si son score est meilleur que celui de la solution courante, celle-ci devient la nouvelle solution courante. Si son score est moins bon, elle peut toujours devenir la nouvelle solution courante, mais avec une certaine probabilité qui est une fonction croissante de la température et décroissante de la dégradation du score. Le mécanisme de refroidissement est appliqué régulièrement de façon à faire baisser la température. L'algorithme privilégie ainsi la diversification au début, en acceptant régulièrement des dégradations de la fonction objectif de manière à franchir les cols. En fin d'algorithme, l'intensification est privilégiée et seules les solutions qui améliorent le score sont acceptées. L'algorithme s'arrête lorsqu'il atteint une certaine température, ou après un certain nombre d'itérations, ou encore après un certain nombre d'itérations sans amélioration de la solution.

Un résultat théorique de convergence est démontré par Henderson, Jacobson et Johnson (2003), mais celui-ci requiert des conditions drastiques (notamment une décroissance logarithmique de la température, donc un nombre d'itérations qui augmente exponentiellement avec la température initiale). Il n'est par conséquent pas applicable en pratique.

5.2.2 Adaptation et implémentation

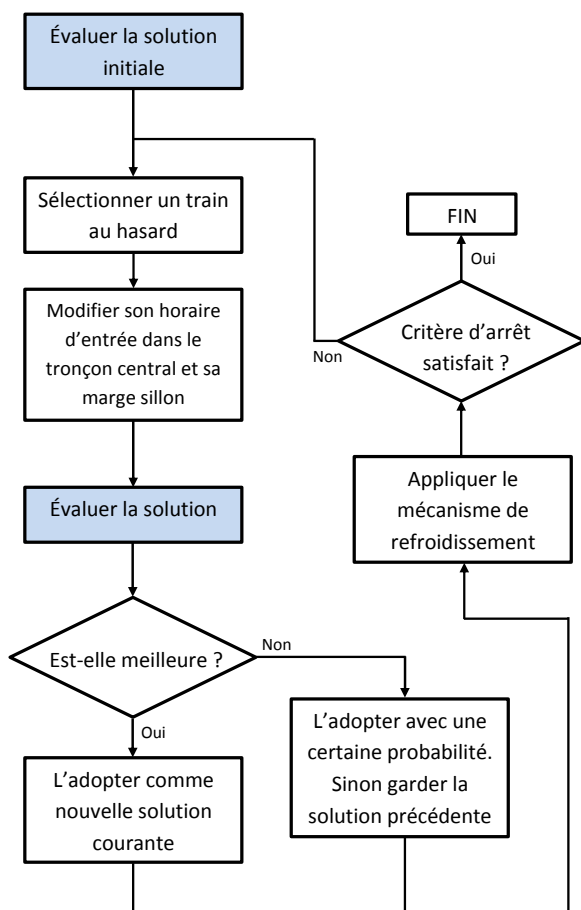


Figure 5.2 – Algorithme de la simheuristique de recuit simulé

Nous avons adapté cet algorithme à un sous-problème du *Train Timetabling Problem*, le problème de l'allocation des marges dans la grille horaire. Nous faisons ainsi l'hypothèse que la liste des trains à assurer est connue dans l'ordre chronologique et fixée, ainsi que leurs dessertes. Nous recherchons alors l'allocation optimale de marge sillon à chacun de ces trains ainsi que de marge de grille à répartir entre eux. Pour éviter une explosion du nombre de variables, nous supposons que la marge sillon est distribuée de façon homogène tout au long du parcours de chaque train. Sous cette hypothèse, si on note N le nombre de trains, une solution peut être entièrement déterminée par $2N$ paramètres. Il suffit pour cela de connaître, pour chaque train, son heure théorique de passage en un point ainsi que sa marge sillon totale. Sa trajectoire dans l'espace-temps peut alors être reconstituée de part et d'autre du point connu en utilisant la trajectoire minimale du chapitre 4 et le montant de la marge sillon allouée à ce train.

Le réseau Transilien a une topologie « en étoile », si bien que la plupart des lignes comportent un tronçon emprunté par tous les trains de la ligne. Il nous semble donc pertinent de choisir comme point de référence pour chaque train un point situé sur ce tronçon central. Dans le cas du système L2-U, le point retenu est l'entrée de la gare de Saint-Cloud. L'heuristique doit alors

résoudre un problème d'optimisation à $2N$ variables, donnant pour chaque train t :

- Sa date a_t d'arrivée en gare de Saint-Cloud, avec une précision à la décimale de seconde.
- Sa marge sillonn totale s_t , supposée comprise entre 10% et 60% par rapport à la marche la plus tendue possible.

Une solution étant ainsi décrite, nous définissons son voisinage comme l'ensemble des solutions qui diffèrent d'elle sur au plus un train (la différence pouvant porter à la fois sur l'heure d'entrée dans le tronçon commun et sur la marge sillonn).

Le principe du fonctionnement de la simheuristique de recuit simulé est décrit par l'algorithme à la figure 5.2. Les appels à SISTRAM y sont mis en évidence par les cases bleues : ceux-ci permettent d'attribuer un score à chaque solution, sous la forme d'une vitesse moyenne telle que celle définie au chapitre 4 et calculée comme la moyenne des vitesses obtenues sur 50 réplifications.

À partir d'une solution donnée, une nouvelle solution est proposée en sélectionnant un train t selon une loi uniforme sur l'ensemble des trains. On note a_{t-1} et a_{t+1} les dates d'entrée dans le tronçon central respectivement des trains précédent et suivant. Une nouvelle allocation de marge sillonn et de marge de grille est alors attribuée au train t en tirant :

- une nouvelle date d'entrée pour ce train selon une loi uniforme sur $[a_{t-1} + 60, a_{t+1} - 60]$ (on impose par conséquent un intervalle temporel minimal de 60 secondes entre deux trains ce qui correspond à la contrainte de block),
- un nouveau coefficient de marge sillonn selon une loi uniforme sur l'intervalle $[10\%, 60\%]$.

La nouvelle solution obtenue après ces modifications est alors évaluée à son tour. Si celle-ci améliore le score de la précédente, elle devient la nouvelle solution courante. Sinon, nous notons Δv la dégradation de la vitesse moyenne par rapport à la solution courante. La nouvelle solution devient alors la solution courante avec une probabilité p donnée par la distribution de Boltzmann :

$$p = e^{-\frac{\Delta v}{T}} \quad (5.1)$$

où T désigne la température. Celle-ci est réduite progressivement selon un mécanisme géométrique. On se fixe deux paramètres $L \in \mathbb{N}$ et $\alpha \in]0, 1[$. Toutes les L itérations, la température est alors réduite d'un facteur α : $T \leftarrow \alpha T$. Les itérations s'arrêtent lorsqu'un critère prédéfini est satisfait. Celui-ci peut porter sur la température, sur le nombre total d'itérations ou encore sur le nombre d'itérations sans amélioration de la solution.

5.2.3 Résultats

5.2.3.1 Paramétrage et conditions expérimentales

Nous avons utilisé la grille horaire 2019 comme solution initiale de l'algorithme. Celle-ci correspond déjà à une solution relativement satisfaisante. Suivant le protocole proposé par Teghem (2012), nous avons effectué un choix de paramètres privilégiant l'intensification à la diversification (sans pour autant renoncer entièrement à cette dernière). Cela se traduit par une température initiale T_0 et un coefficient de refroidissement α relativement faibles. La fréquence des refroidissements doit être d'un ordre de grandeur proche du nombre de variables de décision du problème.

Sur l'instance considérée, le nombre de trains à planifier est de $N = 201$, il y a donc 402 variables de décision. Après avoir testé plusieurs combinaisons de paramètres, les meilleurs résultats ont été obtenus avec le choix suivant :

$$T_0 = 0.1 \quad L = 50 \quad \alpha = 0.9$$

Cet algorithme a été implémenté et exécuté avec la même configuration qu'au chapitre 4 : processeur Intel Core i7, fréquence de 2.90 GHz, mémoire RAM de 7.5 Go. La durée nécessaire à la réalisation de 2000 itérations a été de 13 heures et 38 minutes. 91.8% de ce temps CPU (soit 12 heures et 31 minutes) a été consacré à l'évaluation de grilles par SISTRAM, pour un total de 100 000 réplifications. Le temps restant est consommé par l'heuristique d'optimisation, principalement pour l'écriture de nouveaux fichiers de paramètres à envoyer à SISTRAM.

5.2.3.2 Amélioration de la fonction objectif

La figure 5.3 donne l'évolution de la valeur de la solution courante à chaque itération. Celle-ci est exprimée en variation relative par rapport à la valeur de la solution initiale; celle-ci est de 29.8 km/h. Le critère d'arrêt a été choisi en fonction du nombre d'itérations (chacune étant coûteuse en temps de calcul).

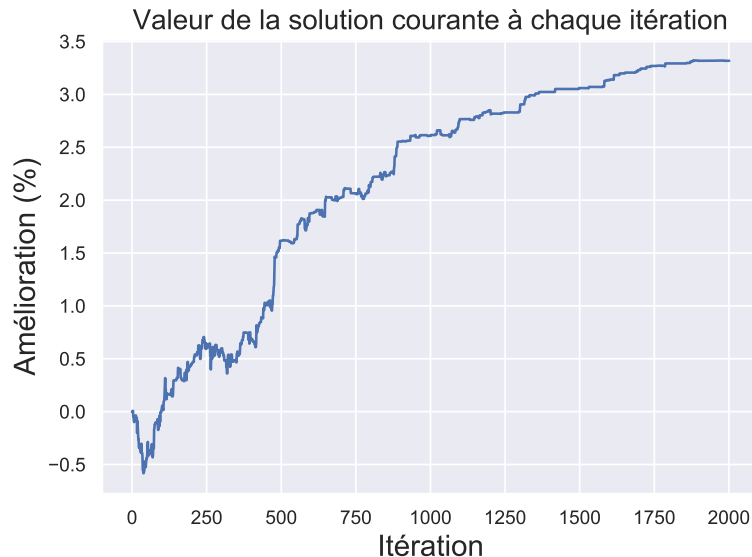


Figure 5.3 – Évolution du score des solutions courantes

On constate tout d'abord une phase de diversification où une dégradation de la solution courante est observée. Celle-ci s'améliore ensuite, d'abord fortement puis de façon plus lente. L'algorithme ayant la propriété « anytime », nous avons choisi de l'arrêter lorsque la progression ralentit, soit après environ 2000 itérations.

On observe que cette simheuristique permet d'obtenir en sortie une grille offrant des performances supérieures de plus de 3% à celles de la grille actuelles. Si ce chiffre peut sembler modeste, il est cependant loin d'être négligeable étant donné le caractère fortement contraint du système. De nombreux projets menés à SNCF Transilien visent en effet à améliorer la ponctualité d'un seul point, voire moins, ceux-ci représentant déjà des gains substantiels.

5.2.3.3 Influence du choix de la solution initiale

Le recuit simulé étant une heuristique de recherche locale, ses résultats peuvent dépendre largement du choix de la solution initiale d'où commence l'exploration. Nous avons cherché à évaluer cette dépendance. Pour ce faire, plusieurs explorations ont été réalisées, avec le même jeu de paramètres que précédemment mais partant de différentes solutions initiales différentes :

- La grille horaire de 2019, donnant la trajectoire présentée au paragraphe précédent.
- Une version « tendue » de cette grille, où tous les trains ont un coefficient de marge sillon identique. Celui-ci est choisi égal à 30% : nous avons en effet montré au chapitre 4 que cette valeur est optimale lorsque tous les trains se voient attribuer la même marge. Nous appellerons cette grille la « grille à marge optimale ».
- La grille « Respiration » déjà présentée au chapitre 4 (correspondant à la grille 2019 où les sillons ont été tendus et réorganisés de façon à aménager un sillon de respiration dans le tronçon central toutes les 15 minutes).

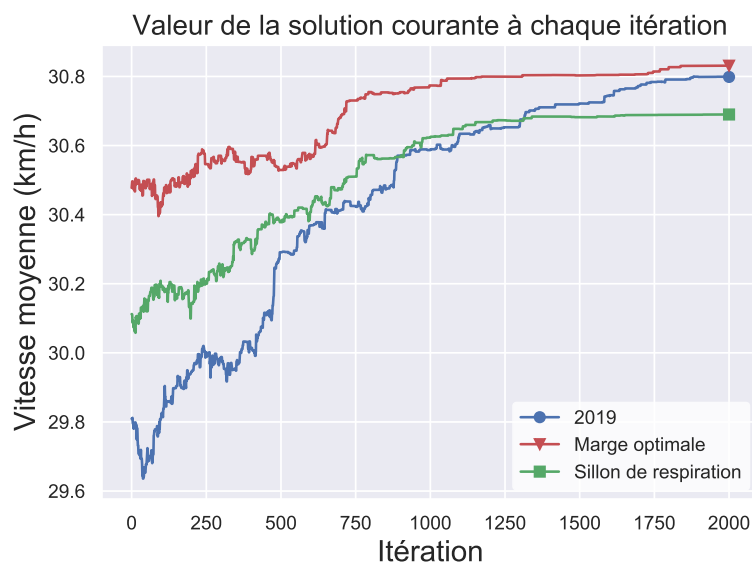


Figure 5.4 – Trajectoire du recuit simulé pour différentes solutions initiales

La figure 5.4 représente l'évolution de la valeur des solutions explorées pour chaque solution initiale. On constate que, quelle que soit la solution initiale choisie, les valeurs des solutions obtenues après 2000 itérations sont très proches (autour de 30.8 km/h). Il nous semble donc probable que l'optimum global à ce problème soit également assez proche de cette valeur.

5.2.3.4 Interprétation des résultats

Les différentes solutions initiales testées dans le paragraphe précédent ont conduit à différentes grilles optimales obtenues au cours de l'exploration. Nous présentons ci-dessous quelques unes des caractéristiques de ces grilles en les comparant avec celles des solutions initiales.

Le tableau 5.1 donne les coefficients de marge sillon moyens sur l'ensemble des trains de chaque grille en entrée et en sortie de l'algorithme. On constate que, dans les trois cas, la marge sillon obtenue en sortie est très proche de la valeur optimale calculée à la section 4.6.4 et située aux alentours de 30%.

Grille initiale	2019	Respiration	Marge optimale
Marge sillon initiale	47.2%	47.2%	30%
Marge sillon finale	32.0%	40.7%	32.8%

Table 5.1 – Marge sillon des solutions initiales et retournées par la simheuristique

Nous avons également comparé les grilles sur la base de l'indicateur SHR (*Sum of Headways Reciprocals*), proposé par Vromans, Dekker et Kroon (2006). Cet indicateur est défini comme la somme des inverses des intervalles théoriques entre trains consécutifs en un point donné. Sa valeur constitue une mesure de l'homogénéité de la grille horaire, la grille étant d'autant plus homogène que la valeur de cet indicateur est faible. Nous avons calculé la valeur de cet indicateur en deux points, en entrée et en sortie du tronçon central, soit au départ de la gare de Saint-Cloud et en entrée de la gare de La Défense. Le tableau 5.2 donne la valeur moyenne du SHR sur ces deux points, pour chaque grille. On constate, étonnamment, que les grilles obtenues en sortie sont plus hétérogènes que celles en entrée, ayant une valeur de SHR plus élevée. Cela peut laisser penser que la meilleure qualité de service n'est pas nécessairement fournie par la grille la plus homogène possible. En effet, tous les trains n'ont pas la même mission ni ne font face aux mêmes aléas. Les trains de la ligne U,

par exemple, circulent habituellement moins chargés que ceux de la ligne L2 en heure de pointe, et ont par conséquent besoin de moins de marge de grille, à la fois car leurs durées de stationnement sont globalement plus faibles et car leurs retards affectent moins de passagères.

Grille initiale	2019	Respiration	Marge optimale
SHR initial	0.702	0.724	0.706
SHR final	1.05	0.942	0.981

Table 5.2 – Indicateur SHR en entrée et en sortie de la simheuristique

Enfin, le tableau 5.3 donne le pourcentage représenté par la durée d'attente en gare dans le calcul du temps de trajet généralisé. Nous constatons que dans tous les cas, la réduction du temps de trajet est principalement due à une diminution du temps d'attente en gare.

Grille initiale	2019	Respiration	Marge optimale
Contribution initiale du temps d'attente	22.1%	23.7%	23.0%
Contribution finale du temps d'attente	20.9%	21.5%	21.2%

Table 5.3 – Contribution du temps d'attente en gare dans les scores des différentes grilles

5.3 Méthodes de descente

L'algorithme précédent permet d'obtenir de bons résultats avec une phase de diversification relativement courte. Nous avons par conséquent cherché ensuite à explorer les possibilités offertes par une méthode dite « de descente ». Ce terme désigne un algorithme de recherche locale n'utilisant que le principe d'intensification, sans accepter de dégradation de la fonction objectif. Le plus connu d'entre eux est l'algorithme dit « du gradient » pour une fonction différentiable. Un tel algorithme peut converger rapidement vers un optimum local. Cependant, l'absence de mécanisme de diversification empêche le franchissement de cols. En l'absence d'hypothèses de convexité, il n'y a donc aucune garantie de convergence vers un optimum global, et la solution obtenue en sortie de l'algorithme dépend fortement du choix de la solution initiale. Il est par conséquent souhaitable que celle-ci offre déjà des performances satisfaisantes, voire qu'elle soit proche de l'optimum.

Nous avons testé deux méthodes de descente pour le problème considéré. La première n'est autre que l'algorithme du recuit simulé précédent, où le mécanisme de diversification a été supprimé. La température initiale est par conséquent nulle, et seule les solutions améliorant la valeur de la fonction objectif sont acceptées comme nouvelle solution courante.

La seconde méthode est une variante de la première, où des informations supplémentaires fournies par SISTRAM sont exploitées pour guider le choix de la nouvelle solution à tester dans le voisinage de la solution courante. Avec la première méthode, ce choix est effectué par un tirage selon une probabilité uniforme sur l'ensemble des solutions du voisinage. Dans ce second algorithme, nous cherchons à modifier en priorité les horaires des trains impliqués dans de nombreux conflits pour l'utilisation de l'infrastructure. Pour ce faire, SISTRAM calcule en sus de la fonction objectif le nombre moyen de signaux franchis à l'avertissement pour chaque train de la grille. On note w_t ce nombre pour le train $t \in \mathcal{T}$. Chaque train se voit alors attribuer le score

$$s_t = \frac{w_t}{\sum_{t \in \mathcal{T}} w_t} \quad (5.2)$$

Une nouvelle distribution de probabilité est alors calculée, attribuant à chaque train une probabilité d'être sélectionné proportionnelle à son score. Une fois le train choisi, l'une des deux actions suivantes est réalisée (chacune pouvant être choisie avec la même probabilité) :

- Retarder l'horaire d'entrée dans le tronçon central du train sélectionné (tout en conservant un intervalle temporel d'au moins 60 secondes avec le train suivant). Cet horaire est tiré selon une loi uniforme sur l'ensemble des dates admissibles. Sa marge sillon est également modifiée en tirant une valeur uniformément dans l'intervalle [10%, 60%].
- Avancer l'horaire d'entrée dans le tronçon central du train précédent le train sélectionné et modifier sa marge sillon, selon le même procédé que ci-dessus.

Ceci doit permettre d'améliorer la probabilité d'obtenir une grille donnant lieu à moins de conflits que la grille précédente.

La figure 5.5 représente l'évolution de la valeur de la solution objectif (toujours exprimée en fonction de l'amélioration relativement à la solution initiale), avec le nombre d'itérations. On constate

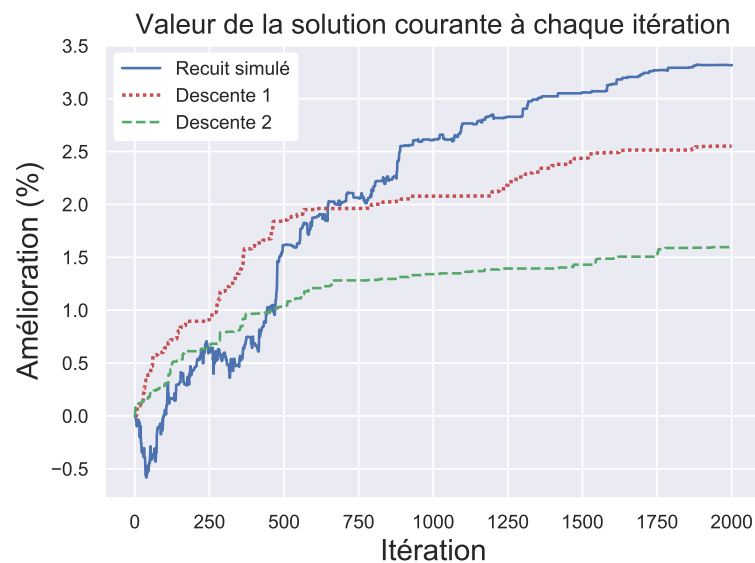


Figure 5.5 – Comparaison des algorithmes de recuit simulé et de descente

une progression rapide des deux méthodes de descente en début d'algorithme ; toutefois celles-ci marquent le pas assez rapidement. Par ailleurs, passée la phase de diversification, l'algorithme du recuit simulé obtient de meilleurs résultats que les méthodes de descente. Cela confirme que la diversification est nécessaire, même courte, pour obtenir des solutions proches de l'optimum. A l'inverse, les deux méthodes de descente semblent en fin d'algorithme rester bloquées aux alentours d'un optimum local et non global. Enfin, la deuxième méthode de descente offre des performances nettement inférieures à celles de la première, et ce tout au long du déroulement de l'algorithme. Cela peut s'expliquer par le fait que la diversification y est encore réduite au sein même des voisinages explorés, puisque la recherche de la nouvelle solution y est fortement orientée.

Il ressort de ces expérimentations que les algorithmes de descente doivent être réservés aux cas où une solution acceptable doit être obtenue au prix d'un temps de calcul relativement court. Cette situation se présente rarement pour le *Train Timetabling Problem*, les grilles horaires étant généralement construites plusieurs mois voire années à l'avance. Le problème cousin de la replanification en temps réel requiert quant à lui une réponse en quelques secondes, les méthodes de descente (et plus généralement tout algorithme utilisant SISTRAM) ne sont donc pas non plus adaptées. En revanche, il arrive que des modifications de plan de transport soient décidées lors de la phase pré-opérationnelle, la veille ou quelques jours à l'avance. Il peut alors être justifié de recourir à de tels algorithmes.

5.4 Conclusion

Un algorithme heuristique de construction de grilles horaires, utilisant SISTRAM comme fonction d'évaluation, a été présenté dans ce chapitre. Celui-ci requiert un temps de calcul relativement long, ce qui exclut *de facto* son utilisation en temps réel. En revanche, son utilisation reste envisageable en amont, lors d'études préalables à la conception de plans de transport.

Les résultats montrent que les meilleures solutions ne sont pas obtenues au voisinage de la grille horaire actuelle et qu'un mécanisme de diversification est nécessaire. En outre, ceux-ci semblent montrer qu'il existe une valeur optimale de marge sillon à allouer dans les grilles pour proposer une bonne qualité de service aux usagers.

Il convient néanmoins de rester prudent quant au caractère immédiatement applicable des solutions obtenues. D'une part, celle-ci néglige certaines contraintes, comme la disponibilité du matériel roulant qui peut réduire le degré de liberté portant sur la date de départ de la gare d'origine (lorsque le train réutilise une rame arrivant d'une mission précédente). Par ailleurs, chaque train se voit attribuer un sillon différent, celui-ci pouvant être plus ou moins tendu. Cela complique le travail des ADC, qui circulent habituellement avec la même vitesse de référence quels que soient le train et l'heure de la journée. Une telle grille où chaque train possède une trajectoire individualisée a davantage de chances d'être applicable dans un système automatisé.

Enfin, on constate que la qualité de service est améliorée principalement grâce à une réduction du temps d'attente en gare. Or celui-ci est estimé sur la base d'hypothèse assez fortes concernant les dates d'arrivée en gare des voyageuses, et pour lesquelles il existe peu de travaux pouvant confirmer leur pertinence. Des vérifications de ces hypothèses pourraient s'avérer nécessaires pour conclure définitivement sur la validité des résultats présentés ici.

Conclusion générale

Synthèse des travaux

Ce travail a tout d'abord permis d'approfondir la compréhension des phénomènes susceptibles de détériorer les conditions de trafic dans les réseaux ferroviaires denses. Après une définition de la zone dense et une présentation générale de l'exploitation ferroviaire, de ses différentes phases et des aléas surgissant au cours de la phase opérationnelle, une classification de ces aléas a été proposée. Nous nous sommes intéressés en particulier aux « petits aléas », qui malgré leur faible amplitude sont susceptibles de s'amplifier et de se propager sur le réseau. Des modèles de ces aléas ont été présentés. Ces modèles, probabilistes, permettent de quantifier la façon dont certains processus fondamentaux de l'exploitation ferroviaire sont influencés par des facteurs extérieurs. Ils permettent également de mettre au jour certains mécanismes sous-jacents aux variabilités observées des durées de stationnement et de trajet des trains en intergare.

Nos travaux ont également constitué une occasion de se réappropriier le concept de robustesse dans le cas des réseaux ferrés denses, et de s'interroger sur la meilleure façon de planifier l'exploitation en amont. Nous avons en effet expliqué en quoi les approches traditionnelles concernant la robustesse des grilles horaires en transport ferroviaire sont inopérantes dans le cas particulier du Mass Transit. Ce constat nous a amené à proposer un nouveau modèle de planification, dont la fonction objectif mesure la performance en termes de qualité de service offerte aux voyageurs lors de la phase opérationnelle.

Une approche couplant simulation et optimisation a enfin été proposée afin de calculer des solutions optimales au problème ainsi modélisé. L'algorithme de simulation permet, à partir de données de trafic et de demande voyageurs, de quantifier l'efficacité d'une grille horaire pour répondre à cette demande en présence de petits aléas. Cet outil a permis de valider certains récents choix effectués à SNCF Transilien lors de refontes horaires, et également de mettre en évidence les limites du procédé consistant à systématiquement allonger les temps de parcours pour accroître la robustesse. Bien que l'essai ne se soit pas révélé concluant, nous avons également ouvert la porte vers une planification intégrant les leviers d'action disponibles en phase opérationnelle. Cela a permis de proposer un cadre qui pourrait être réutilisé lors de travaux futurs. La méthode de simulation développée a également permis de démontrer la valeur des données de retour d'expérience et le fait que celles-ci sont encore aujourd'hui trop peu exploitées.

Enfin, la réutilisation de cet outil au sein d'un algorithme heuristique d'optimisation a permis d'obtenir des grilles offrant de meilleures performances que la grille horaire actuelle sur la ligne choisie comme cas d'étude. Ceci confirme l'utilité de la simulation stochastique lors de la phase de conception, et ouvre la porte au développement d'outils d'aide à la décision permettant d'automatiser en partie le processus de conception de grilles horaires robustes.

Perspectives

Plusieurs pistes mériteraient d'être explorées à la suite de ces travaux. À court terme, il serait intéressant de mener d'autres expérimentations visant à consolider les résultats obtenus. Celles-ci pourraient prendre diverses formes :

- Une application de la méthode à d'autres lignes. Si certaines données manquent actuellement pour le permettre, la généralisation des dispositifs de comptage et d'enregistrement des événements de bord devrait rendre celles-ci rapidement disponibles et exploitables. Le modèle présenté gagnerait alors à être confronté à d'autres situations, notamment sur des lignes où

l'infrastructure est utilisée à la fois par des circulations Transilien et d'autres circulations (TGV, Intercités, TER, Fret) ne relevant pas de la même logique d'exploitation.

- Des expérimentations avec d'autres paramètres. Il serait par exemple envisageable d'explorer la possibilité de répartir la marge de façon non homogène le long des sillons, tous les points du réseau n'étant pas identiquement sensibles aux petits aléas.
- D'autres heuristiques que le recuit simulé pourraient également être utilisées dans l'approche intégrée simulation-optimisation, celles-ci pouvant potentiellement se révéler plus efficaces pour surmonter les extrema locaux et converger vers un optimum global.

En outre, certaines contraintes ont été négligées et gagneraient à être intégrées au modèle. C'est le cas notamment des contraintes liées à la réutilisation du personnel et du matériel roulant d'un train à un autre.

A moyen terme, notre travail pourrait contribuer à améliorer les pratiques concernant la planification de l'exploitation. En particulier, les modèles d'aléas présentés gagneraient à être intégrés au sein des simulateurs actuellement utilisés pour les études de robustesse lorsque c'est possible. En effet, ceux-ci permettent d'exploiter davantage l'information disponible que les modèles d'aléas indépendants actuellement utilisés.

Par ailleurs, des travaux complémentaires pourraient être menés afin de mieux prendre en compte les leviers de régulation au sein de l'outil de simulation. Si une première tentative a été présentée, son fonctionnement reposant sur des règles n'a pas permis d'obtenir de résultats concluants. Il serait intéressant d'explorer la possibilité de coupler à l'outil de simulation un module prenant des décisions de façon optimisée, tel que l'outil Stabilité actuellement en cours de déploiement à SNCF Transilien. Cela permettrait à la fois d'améliorer le réalisme des simulations, en y intégrant des décisions plus proches de celles effectivement prises en opérationnel, et la qualité des solutions obtenues puisque les choix de régulation seraient alors plus rationnels.

Enfin, à plus long terme, ces travaux pourraient contribuer à la mise en place de nouveaux processus visant à améliorer l'exploitation. Ils pourraient ainsi permettre d'évaluer les gains potentiels qui pourraient être obtenus en agissant directement sur les principales sources de variabilités. Plusieurs expériences de ce type font déjà l'objet d'expérimentations, en France ou à l'étranger : régulation des flux de voyageurs avant l'accès au quai, compte à rebours en cabine pour réduire les durées de stationnement en gare, aide à la décision pour les régulateurs, automatisation de la conduite... les modèles présentés ici pourraient ainsi servir à d'éventuelles études d'opportunité préliminaires à la mise en œuvre de tels projets.

Par ailleurs, les résultats présentés dans ce manuscrit pourraient servir de base à une éventuelle remise en question des indicateurs de performance utilisés en zone dense. Nous avons montré les limites des indicateurs classiques (performance en mode nominal, régularité) dans ce contexte. Une réflexion sur ce sujet pourrait conduire à mettre en place de nouveaux indicateurs plus aptes à refléter la qualité de service offerte aux voyageurs.

Enfin, nous espérons que ce travail a permis de mettre en valeur l'important potentiel des données de retour d'expérience collectées au sein du groupe SNCF, et qu'il encouragera à l'avenir les différentes parties prenantes de l'exploitation à utiliser ces données de manière coordonnée pour construire un système de Mass Transit plus performant.

Bibliographie

- Abril, M., F. Barber, L. Ingolotti, M.A. Salido, P. Tormos et A. Lova (2008). "An assessment of railway capacity". In : *Transportation Research Part E : Logistics and Transportation Review* 44.5, p. 774-806.
- Altazin, Estelle (2018). "Stabilité et replanification d'un système ferroviaire dense". Thèse de doctorat. Ecole des Mines de Saint-Etienne.
- Altazin, Estelle, Stéphane Dauzère-Pères, François Ramond et Sabine Tréfond (2020). "A multi-objective optimization-simulation approach for real time rescheduling in dense railway systems". In : *European Journal of Operational Research*.
- Altman, Naomi S. (1992). "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression". In : *The American Statistician* 46.3, p. 175-185.
- Andersson, Emma, Anders Peterson et Johanna Törnquist (2013). "Quantifying railway timetable robustness in critical points". In : *Journal of Rail Transport Planning & Management* 3, p. 95-110.
- Ansoff, Igor (1965). *Corporate Strategy : An Analytic Approach to Business Policy for Growth and Expansion*. McGraw-Hill.
- Arenas, Diego, Rémy Chevrier, Saïd Hanafi et Joaquin Rodriguez (2015). "Solving the train timetabling problem, a mathematical model and a genetic algorithm solution approach." In : *6th International Conference on Railway Operations Modelling and Analysis*. Tokyo, Japon.
- Barrena, Eva, David Canca, Leandro C. Coelho et Gilbert Laporte (2014). "Exact formulations and algorithm for the train timetabling problem with dynamic demand". In : *Computers & Operations Research* 44, p. 66-74.
- Bell, Michael G.H. et Yasunori Iida (1997). *Transportation Network Analysis*. Wiley & Sons.
- Bellman, Richard (1961). *Adaptive Control Processes : A Guided Tour*. Princeton University Press.
- Ben-Tal, Aharon, Laurent El Ghaoui et Arkadi Nemirovski (2009). *Robust Optimization*. T. 28. Princeton Series in Applied Mathematics. Princeton University Press.
- Ben-Tal, Aharon et Arkadi Nemirovski (2002). "Robust optimization – methodology and applications". In : *Mathematical Programming* 92.3, p. 453-480.
- Bendfeldt, J.-P., U. Mohr et L. Müller (2000). "RailSys, a system to plan future railway needs". In : *Computers in Railways VII*. Bologne, Italie.
- Bertsimas, Dimitris et Melvyn Sim (2004). "The Price of Robustness". In : *Operations Research* 52.1, p. 35-53.
- Bešinović, Nikola, Rob M.P. Goverde, Egidio Quaglietta et Roberto Roberti (2016). "An integrated micro-macro approach to robust railway timetabling". In : *Transportation Research Part B : Methodological* 87, p. 14-32.
- Birge, John R. et François Louveaux (1997). *Introduction to Stochastic Programming*. Springer Science & Business Media.
- Bouvarel, Paul (2011). *Cours de Chemin de Fer*. Ecole Nationale des Ponts et Chaussées.
- Brännlund, U., P. O. Lindberg, A. Nöu et J.-E. Nilsson (1998). "Railway Timetabling Using Lagrangian Relaxation". In : *Transportation Science* 32.4, p. 358-369.
- Brethomé, Lucile (2018). "Passenger oriented modelling and optimization of the railway transportation plan". Thèse de doctorat. Ecole Centrale Lille.
- Brünger, Olaf et Elias Dahlhaus (2008). "Running time estimation". In : *Railway Timetable and Traffic*. Sous la dir. d'Ingo Arne Hansen et Jörn Pachl. Eurail Press, p. 58-82.
- Buchmüller, Stefan, Ulrich Weidmann et Andrew Nash (2008). "Development of a dwell time calculation model for timetable planning". In : *Computers in Railways*. Toledo : WIT Press, p. 525-534.
- Büker, Thorsten et Bernhard Seybold (2012). "Stochastic modelling of delay propagation in large networks". In : *Journal of Rail Transport Planning & Management* 2, p. 34-50.

- Burggraeve, Sofie et Pieter Vansteenwegen (2017). "Optimization of supplements and buffer times in passenger robust timetabling". In : *Best Papers of the RailLille 2017 conference* 7.3, p. 171-186.
- Cacchiani, Valentina, Alberto Caprara et Matteo Fischetti (2012). "A Lagrangian Heuristic for Robustness with an Application to Train Timetabling". In : *Transportation Science* 1.46, p. 124-133.
- Cacchiani, Valentina et Paolo Toth (2012). "Nominal and robust train timetabling problems". In : *European Journal of Operational Research* 219.3, p. 727-737.
- Caimi, Gabrio, Leo Kroon et Christian Liebchen (2017). "Models for railway timetable optimization : Applicability and applications in practice". In : *Journal of Rail Transport Planning & Management* 6.4, p. 285-312.
- Caprara, Alberto, Matteo Fischetti et Paolo Toth (2002). "Modeling and Solving the Train Timetabling Problem". In : *Operations Research* 50.5, p. 851-861.
- Caprara, Alberto, Matteo Fischetti, Paolo Toth, Daniele Vigo et Pier Luigi Guida (1997). "Algorithms for railway crew management". In : *Mathematical Programming* 79.1, p. 125-141.
- Carey, Malachy (1999). "Ex ante heuristic measures of schedule reliability". In : *Transportation Research Part B : Methodological* 33.7, p. 473-494.
- Carey, Malachy et Serena Carville (2000). "Testing schedule performance and reliability for train stations". In : *Journal of the Operational Research Society* 51.6, p. 666-682.
- Carey, Malachy et Ivan Crawford (2007). "Scheduling trains on a network of busy complex stations". In : *Advanced Modelling of Train Operations in Stations and Networks* 41.2, p. 159-178.
- Carré, Coline (2014). *Etude de la stabilité de l'exploitation ferroviaire en zone dense, au travers de la recherche d'indicateurs basés sur des données de retour d'expérience*. Thèse professionnelle, Mastère Systèmes de Transport Ferroviaires et Urbains. ENPC.
- Ceder, Avishai (2007). "Passenger demand". In : *Public Transit Planning and Operation : Modeling, Practice and Behavior*. CRC Press, p. 309 -332.
- Chambers, John M., William S. Cleveland, Beat Kleiner et Paul A. Tukey (2018). *Graphical Methods for Data Analysis*. CRC Press.
- Chow, Andy H.F. et Aris Pavlides (2018). "Cost functions and multi-objective timetabling of mixed train services". In : *Transportation Research Part A : Policy and Practice* 113, p. 335-356.
- Christoforou, Zoi, Ektoros Chandakas et Ioannis Kapariaris (2020). "Investigating the Impact of Dwell Time on the Reliability of Urban Light Rail Operations". In : *Urban Rail Transit*.
- Cicerone, Serafino, Gianlorenzo D'Angelo, Gabriele Di Stefano, Daniele Frigioni, Alfredo Navarra, Michael Schachtebeck et Anita Schöbel (2009). "Recoverable Robustness in Shunting and Timetabling". In : *Robust and Online Large-Scale Optimization*. Sous la dir. de Ravindra K. Ahuja, Rolf H. Möhring et Christos D. Zaroliagis. Berlin, Heidelberg : Springer, p. 28-60.
- Cicerone, Serafino, Gabriele Di Stefano, Michael Schachtebeck et Anita Schöbel (2012). "Multi-stage recovery robustness for optimization problems : A new concept for planning under disturbances". In : *Information Sciences* 190.Supplement C, p. 107-126.
- Conte, Carla et Anita Schöbel (2007). "Identifying dependencies among delays". In : *RailHannover. 2th International Conference on Railway Operations Modelling and Analysis (ICROMA)*. Hanovre.
- Cornet, Sélim, Christine Buisson, François Ramond, Paul Bouvarel et Joaquin Rodriguez (2019). "Methods for quantitative assessment of passenger flow influence on train dwell time in dense traffic areas". In : *Transportation Research Part C : Emerging Technologies* 106, p. 345-359.
- Cuniase, Pierre-Antoine (2015). "Etude théorique et expérimentale des phénomènes de congestion sur un réseau ferroviaire urbain". Thèse de doctorat. Lyon : Ecole Nationale des Travaux Publics de l'Etat.
- Daamen, Winnie, Yu-chen Lee et Paul Wiggenraad (2008). "Boarding and Alighting Experiments : Overview of Setup and Performance and Some Preliminary Results". In : *Transportation Research Record : Journal of the Transportation Research Board* 2042, p. 71-81.

- D'Acerno, Luca, Marilisa Botte, Antonio Placido, Chiara Caropreso et Bruno Montella (2017). "Methodology for Determining Dwell Times Consistent with Passenger Flows in the Case of Metro Services". In : *Urban Rail Transit* 3.2, p. 73-89.
- Delorme, Xavier, Xavier Gandibleux et Joaquín Rodríguez (2009). "Stability evaluation of a railway timetable at station level". In : *European Journal of Operational Research* 195.3, p. 780-790.
- Deng, Shidi (2019). *Optimisation dynamique du trafic ferroviaire en zone dense*. Rapport de stage de fin d'études. ENSTA ParisTech.
- Dewilde, Thijs (2014). "Improving the robustness of a railway system in large and complex station areas". Thèse de doctorat. KU Leuven.
- Dormann, Carsten F., Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, Jaime R. García Marquéz, Bernd Gruber, Bruno Lafourcade, Pedro J. Leitão, Tamara Münkemüller, Colin McClean, Patrick E. Osborne, Björn Reineking, Boris Schröder, Andrew K. Skidmore, Damaris Zurell et Sven Lautenbach (2012). "Collinearity : a review of methods to deal with it and a simulation study evaluating their performance". In : *Ecography* 36.1, p. 27-46.
- Dueker, Kenneth J., Thomas J Kimpel, James G. Strathman et Steve Callas (2004). "Determinants of bus dwell time". In : *Journal of Public Transportation* 7.1, p. 21-40.
- Dupačová, Jitka et Miloš Kopa (2012). "Robustness in stochastic programs with risk constraints". In : *Annals of Operations Research* 200.1, p. 55-74.
- Elleuch, Fatma (2019). "Transférabilité d'une modélisation-simulation multi-agents : le comportement inter-gares des voyageurs de la SNCF lors des échanges quai-train". Thèse de doctorat. Paris : Conservatoire National des Arts et Métiers.
- Faverges, Marie Milliet de, Giorgio Russolillo, Christophe Picouleau, Boubek Merabet et Bertrand Houzel (2018). "Modelling passenger train arrival delays with Generalized Linear Models and its perspective for scheduling at main stations". In : *8th International Conference on Railway Engineering (ICRE 2018)*. Londres, Royaume-Uni.
- Fischetti, Matteo et Michele Monaci (2009). "Light Robustness". In : *Robust and Online Large-Scale Optimization*. Sous la dir. de Ravindra K. Ahuja, Rolf H. Möhring et Christos D. Zaroliagis. Berlin, Heidelberg : Springer, p. 61-84.
- Fischetti, Matteo, Domenico Salvagnin et Arrigo Zanette (2009). "Fast approaches to improve the robustness of a railway timetable". In : *Transportation Science* 43, p. 321-335.
- Flier, Holger, Rati Gelashvili, Thomas Graffagnino et Marc Nunkesser (2009). "Mining railway delay dependencies in large-scale real-world railway networks". In : *Robust online large-scale optimization*. Springer, p. 354-368.
- Fu, Michael C., éd. (2014). *Handbook of Simulation Optimization*. T. 216. International Series in Operations Research & Management Science. Springer.
- Gabrel, Virginie, Cécile Murat et Aurélie Thiele (2014). "Recent advances in robust optimization : An overview". In : *European Journal of Operational Research* 235.3, p. 471-483.
- Gestrelus, Sara, Martin Aronsson, Malin Forsgren et Hans Dahlberg (2012). "On the delivery robustness of train timetables with respect to production replanning possibilities". In : *CETRA2012*. Dubrovnik, Croatie.
- Gillespie, Daniel T (1976). "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In : *Journal of Computational Physics* 22.4, p. 403-434.
- Goerigk, Marc et Anita Schöbel (2014). "Recovery-to-optimality : A new two-stage approach to robustness with an application to aperiodic timetabling". In : *Computers & Operations Research* 52, p. 1-15.
- Goverde, Rob (2007). "Railway timetable stability using max-plus system theory". In : *Transportation Research Part B* 41, p. 179-201.
- Goverde, Rob M. P. et Ingo Arne Hansen (2013). "Performance indicators for railway timetables". In : *2013 IEEE International Conference on Intelligent Rail Transportation Proceedings*, p. 301-306.

- Goverde, Rob M.P., Nikola Bešinović, Anne Binder, Valentina Cacchiani, Egidio Quaglietta, Roberto Roberti et Paolo Toth (2016). "A three-level framework for performance-based railway timetabling". In : *Transportation Research Part C : Emerging Technologies* 67, p. 62-83.
- Großmann, Peter, Steffen Hölldobler, Norbert Manthey, Karl Nachtigall, Jens Opitz et Peter Steinke (2012). "Solving Periodic Event Scheduling Problems with SAT". In : *Advanced Research in Applied Artificial Intelligence*. Sous la dir. d'He Jiang, Wei Ding, Moonis Ali et Xindong Wu. Berlin, Heidelberg : Springer Berlin Heidelberg, p. 166-175.
- Grube, Pablo, Felipe Núñez et Aldo Cipriano (2011). "An event-driven simulator for multi-line metro systems and its application to Santiago de Chile metropolitan rail network". In : *Modeling and Performance Analysis of Networking and Collaborative Systems* 19.1, p. 393-405.
- Hanafi, Rosmalina et Erhan Kozan (2014). "A hybrid constructive heuristic and simulated annealing for railway crew scheduling". In : *Computers & Industrial Engineering* 70, p. 11-19.
- Hansen, Ingo Arne, Rob M.P. Goverde et Dirk J. van der Meer (2010). "Online train delay recognition and running time prediction". In : *13th International IEEE Conference on Intelligent Transportation Systems*, p. 1783-1788.
- Hansen, Ingo Arne et Jörn Pachl, éd. (2008). *Railway Timetable and Traffic*. Eurail Press.
- Hassannayebi, Erfan, Arman Sajedinejad et Soheil Mardani (2014). "Urban rail transit planning using a two-stage simulation-based optimization approach". In : *Simulation Modelling Practice and Theory* 49, p. 151-166.
- Hassannayebi, Erfan, Seyed Hessameddin Zegordi, Mohammad Reza Amin-Naseri et Masoud Yaghini (2017). "Train timetabling at rapid rail transit lines : a robust multi-objective stochastic programming approach". In : *Operational Research* 17.2, p. 435-477.
- Henderson, Darrall, Sheldon H. Jacobson et Alan W. Johnson (2003). "The theory and practice of simulated annealing". In : *Handbook of Metaheuristics*. Sous la dir. de Fred W. Glover et Gary A. Kochenberger. International Series in Operations Research & Management Science 57. Springer-Verlag.
- Ho, Tin Kin, Chi Wai Tsang, Kwan Ho Ip et K.S. Kwan (2012). "Train service timetabling in railway open markets by particle swarm optimisation". In : *Expert Systems with Applications* 39.1, p. 861-868.
- Högdahl, Johan, Markus Bohlin et Oskar Fröidh (2019). "A combined simulation-optimization approach for minimizing travel time and delays in railway timetables". In : *Transportation Research Part B : Methodological* 126, p. 192-212.
- Jamili, Amin, Mohammad Ali Shafia, Seyed Jafar Sadjadi et Reza Tavakkoli-Moghaddam (2012). "Solving a periodic single-track train timetabling problem by an efficient hybrid algorithm". In : *Engineering Applications of Artificial Intelligence* 25.4, p. 793-800.
- Jensen, Lars Wittrup (2015). "Robustness indicators and capacity models for railway networks". Thèse de doctorat. DTU Management.
- Jolliffe, Ian T. (1986). *Principal Component Analysis*. Springer.
- Jovanović, Predrag, Pavle Kecman, Nebojša Bojović et Dragomir Mandić (2017). "Optimal allocation of buffer times to increase train schedule robustness". In : *European Journal of Operational Research* 256.1, p. 44-54.
- Juan, Angel A., Javier Faulin, Scott E. Grasman, Markus Rabe et Gonçalo Figueira (2015). "A review of simheuristics : Extending metaheuristics to deal with stochastic combinatorial optimization problems". In : *Operations Research Perspectives* 2, p. 62-72.
- Kecman, Pavle et Rob M. P. Goverde (2015). "Predictive modelling of running and dwell times in railway traffic". In : *Public Transport* 7.3, p. 295-319.
- Khadilkar, Harshad (2016). "Data-Enabled Stochastic Modeling for Evaluating Schedule Robustness of Railway Networks". In : *Transportation Science* 51.4, p. 1161-1176.
- Kraft, W. H. (1975). "An Analysis of the Passenger Vehicle Interface of Street Transit Systems with Applications to Design Optimization". Thèse de doctorat. New Jersey Institute of Technology.

- Kroon, Leo, Gábor Maróti, Mathijn Retel Helmrich, Michiel Vromans et Rommert Dekker (2008). "Stochastic improvement of cyclic railway timetables". In : *Transportation Research Part B : Methodological* 42, p. 553-570.
- Kroon, Leo G., Leon W. P. Peeters, Joris C. Wagenaar et Rob A. Zuidwijk (2013). "Flexible Connections in PESP Models for Cyclic Passenger Railway Timetabling". In : *Transportation Science* 48.1, p. 136-154.
- Landex, Alex et Lars Wittrup Jensen (2013). "Measures for track complexity and robustness of operation at stations". In : *Journal of Rail Transport Planning & Management* 3.1, p. 22-35.
- Larsen, Rune, Marco Pranzo, Andrea D'Ariano, Francesco Corman et Dario Pacciarelli (2014). "Susceptibility of optimal train schedules to stochastic disturbances of process times". In : *Flexible Services and Manufacturing Journal* 26.4, p. 466-489.
- Leboeuf, Cyril et Christelle Lérin (2019). "DENFERT : Anticiper la robustesse des plans de transport". In : *Revue Générale des Chemins de Fer* 294, p. 26-37.
- Lee, Yusin, Li-Sin Lu, Mei-Ling Wu et Dung-Ying Lin (2017). "Balance of efficiency and robustness in passenger railway timetables". In : *Transportation Research Part B : Methodological* 97, p. 142-156.
- Li, Dewei, Winnie Daamen et Rob M. P. Goverde (2016). "Estimation of train dwell time at short stops based on track occupation event data : A study at a Dutch railway station". In : *Journal of Advanced Transportation* 50.5, p. 877-896.
- Li, Dewei, Yonghao Yin et Hong He (2018). "Testing the Generality of a Passenger Disregarded Train Dwell Time Estimation Model at Short Stops : Both Comparison and Theoretical Approaches". In : *Journal of Advanced Transportation* 2018, p. 1-16.
- Li, Dewei, Rob M. P. Goverde, Winnie Daamen et Hong He (2014). "Train dwell time distributions at short stop stations". In : *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, p. 2410-2415.
- Li, KePing, ZiYou Gao et Bin Ning (2005). "Cellular automaton model for railway traffic". In : *Journal of Computational Physics* 209.1, p. 179-192.
- Liebchen, Christian (2008). "The First Optimized Railway Timetable in Practice". In : *Transportation Science* 42.4, p. 420-435.
- Liebchen, Christian, Marco Lübbecke, Rolf Möhring et Sebastian Stiller (2009). "The concept of recoverable robustness, linear programming recovery and railway applications". In : *Robust and online large-scale optimization*. Springer, p. 1-27.
- Liebchen, Christian, Michael Schachtebeck, Anita Schöbel, Sebastian Stiller et André Prigge (2010). "Computing delay resistant railway timetables". In : *Computers & Operations Research* 37, p. 857-868.
- Lin, Tyh-Ming et Nigel Wilson (1992). "Dwell time relationships for light railway systems". In : *Transportation Research Record* 1361, p. 287-295.
- Longo, Giovanni et Giorgio Medeossi (2012). "Enhancing Timetable Planning With Stochastic Dwell Time Modelling". In : *Computers in Railways*. New Forest : WIT Press, p. 461 -471.
- Lucas, Rémi, Zacharie Ales, Sourour Elloumi et François Ramond (2019). "Reducing the Adaptation Costs of a Rolling Stock Schedule with Adaptive Solution : the Case of Demand Changes". In : *RailNorrköping. 8th International Conference on Railway Operations Modelling and Analysis (ICROMA)*. Norrköping, Sweden : Linköping University Electronic Press.
- Lusby, Richard M., Jesper Larsen et Simon Bull (2018). "A survey on robustness in railway planning". In : *European Journal of Operational Research* 266.1, p. 1-15.
- Lüthi, Marco (2009). "Improving the efficiency of heavily used railway networks through integrated real-time rescheduling". Thèse de doctorat. ETH Zürich.
- Lüthi, Marco, Ulrich Weidmann et Andrew Nash (2007). "Passenger arrival rates at public transport stations". In : *86th Transportation Research Board Annual Meeting*. Zürich.

- Marinov, Marin et José Viegas (2011). "A mesoscopic simulation modelling methodology for analyzing and evaluating freight train operations in a rail network". In : *Modeling and Performance Analysis of Networking and Collaborative Systems* 19.1, p. 516-539.
- Maróti, Gábor (2006). "Operations research models for railway rolling stock planning". Thèse de doctorat. Eindhoven University of Technology.
- Maróti, Gábor (2017). "A branch-and-bound approach for robust railway timetabling". In : *Public Transport* 9.1, p. 73-94.
- Medeossi, Giorgio (2008). "Capacity and reliability in railway networks". Thèse de doctorat. Università degli studi di Trieste.
- Medeossi, Giorgio, Giovanni Longo et Stefano de Fabris (2011). "A method for using stochastic blocking times to improve timetable planning". In : *Journal of Rail Transport Planning & Management* 1.1, p. 1-13.
- Meer, Dirk J. van der (2010). "Prediction of Train Running Times and Conflicts using Track Occupation Data". In : *12th WCTR-World Congress of Transportation Research, Lisbon, Portugal, July 2010*.
- Meng, Xuelei, Limin Jia et Yong Qin (2010). "Train Timetable Optimizing and Rescheduling Based on Improved Particle Swarm Algorithm". In : *Transportation Research Record* 2197.1, p. 71-79.
- Metropolis, Nicholas et S. Ulam (1949). "The Monte Carlo Method". In : *Journal of the American Statistical Association* 44.247, p. 335-341.
- Nachtigall, Karl (1998). *Periodic network optimization and fixed interval timetables*. Deutsches Zentrum für Luft-und Raumfahrt. Braunschweig : Institut für Flugführung.
- Nachtigall, Karl et Stefan Voget (1996). "A genetic algorithm approach to periodic railway synchronization". In : *Computers & Operations Research* 23.5, p. 453-463.
- Nash, Andrew et Daniel Huerlimann (2004). "Railroad simulation using OpenTrack". In : *Computers in Railways IX*. Dresden, Germany, p. 45-54.
- Nelson, Barry (2013). *Foundations and methods of stochastic simulation : a first course*. International Series in Operations Research & Management Science 187. Springer.
- Ochiai, Yasufumi et Norio Tomii (2015). "Punctuality analysis using a microscopic simulation in which drivers' behaviour is considered". In : *Journal of Rail Transport Planning & Management* 5.3, p. 128-145.
- Odijk, Michiel A. (1996). "A constraint generation algorithm for the construction of periodic railway timetables". In : *Transportation Research Part B : Methodological* 30.6, p. 455-464.
- Odijk, Michiel A. (1997). "Railway timetable generation". Thèse de doctorat. Delft, Pays-Bas : Delft University of Technology.
- Pedersen, Timothy, Thomas Nygreen et Anders Lindfeldt (2018). "Analysis of temporal factors influencing minimum dwell time distributions". In : *Computers in Railways*. Lisbon : WIT Press, p. 87 -98.
- Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot et Edouard Duchesnay (2011). "Scikit-learn : Machine learning in Python". In : *Journal of Machine Learning Research* 12, p. 2825-2830.
- Peeters, Leon et Leo Kroon (2001). "A Cycle Based Optimization Model for the Cyclic Railway Timetabling Problem". In : *Computer-Aided Scheduling of Public Transport*. Sous la dir. de Stefan Voß et Joachim R. Daduna. Berlin, Heidelberg : Springer Berlin Heidelberg, p. 275-296.
- Pellegrini, Paola, Grégory Marlière et Joaquin Rodriguez (2017). "RECIFE-SAT : A MILP-based algorithm for the railway saturation problem". In : *Journal of Rail Transport Planning & Management* 7.1, p. 19-32.

- Pellegrini, Paola, Grégory Marlière, Raffaele Pesenti et Joaquin Rodriguez (2015). "RECIFE-MILP : An Effective MILP-Based Heuristic for the Real-Time Railway Traffic Management Problem". In : *IEEE Transactions on Intelligent Transportation Systems* 16.5, p. 2609-2619.
- Puong, Andre (2000). *Dwell time model and analysis for the MBTA red line*. Rapp. tech. Cambridge.
- Quaglietta, Egidio, Luca D'Acierno, Vincenzo Punzo, Roberto Nardone et Nicola Mazzocca (2011). "A simulation framework for supporting design and real-time decisional phases in railway systems". In : *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, p. 846-851.
- Radtke, Alfons (2008). "Infrastructure Modelling". In : *Railway Timetable and Traffic*. Sous la dir. d'Ingo Arne Hansen et Jörn Pahl. Eurail Press, p. 43-57.
- Ramond, François, Valentina Pozzoli et Nicolas Marcos (2017). "Optimized real-time train order at junctions". In : *RailLille2017*. Lille.
- Robenek, Tomáš, Yousef Maknoon, Shadi Sharif Azadeh, Jianghang Chen et Michel Bierlaire (2016). "Passenger centric train timetabling problem". In : *Transportation Research Part B : Methodological* 89, p. 107-126.
- Rockafellar, Tyrell (1984). *Network flows and monotropic optimization*. Wiley & Sons.
- Salido, Miguel, Federico Barber et Laura Ingolotti (2008). "Robustness in railway transportation scheduling". In : *7th World Congress on Intelligent Control and Automation*, p. 2880-2885.
- Samà, Marcella, Paola Pellegrini, Andrea D'Ariano, Joaquin Rodriguez et Dario Pacciarelli (2016). "Ant colony optimization for the real-time train routing selection problem". In : *Transportation Research Part B : Methodological* 85, p. 89-108.
- Schelenz, Tomasz, Ángel Suescun, Li Wikström et Marianne Karlsson (2014). "Application of agent based simulation for evaluating a bus layout design from passengers' perspective". In : *Special Issue with Selected Papers from Transport Research Arena* 43, p. 222-229.
- Schittenhelm, Bernd et Alex Landex (2013). "Development and application of Danish Key Performance Indicators for Railway Timetables". In : *5th International Seminar on Railway Operations Modelling and Analysis - RailCopenhagen*. Copenhagen, Danemark.
- Schlechte, Thomas et Ralf Borndörfer (2010). "Balancing Efficiency and Robustness – A Bi-criteria Optimization Approach to Railway Track Allocation". In : *Multiple Criteria Decision Making for Sustainable Energy and Transportation Systems*. Sous la dir. de Matthias Ehrgott, Boris Naujoks, Theodor J. Stewart et Jyrki Wallenius. Berlin, Heidelberg : Springer Berlin Heidelberg, p. 105-116.
- Schöbel, Anita (2012). "Line planning in public transportation : models and methods". In : *OR Spectrum* 34.3, p. 491-510.
- Schöbel, Anita et Albrecht Kratz (2009). "A Bicriteria Approach for Robust Timetabling". In : *Robust and Online Large-Scale Optimization*. Sous la dir. de Ravindra K. Ahuja, Rolf H. Möhring et Christos D. Zaroliagis. Berlin, Heidelberg : Springer, p. 119-144.
- Sels, P., T. Dewilde, D. Cattrysse et P. Vansteenwegen (2016). "Reducing the passenger travel time in practice by the automated construction of a robust railway timetable". In : *Transportation Research Part B : Methodological* 84, p. 124-156.
- Serafini, Paolo et Walter Ukovich (1989). "A Mathematical Model for Periodic Scheduling Problems". In : *SIAM Journal on Discrete Mathematics* 2.4, p. 550-581.
- Seriani, Sebastian et Rodrigo Fernandez (2015). "Pedestrian traffic management of boarding and alighting in metro stations". In : *Transportation Research Part C : Emerging Technologies* 53, p. 76-92.
- Silvermann, Bernard W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.
- SNCF Réseau (2019). *Comprendre le maillage de la région Île de France*. url : <https://www.sncf-reseau.com/fr/reseau/ile-de-france> (visité le 03/03/2020).

- Solinen, Emma, Gemma Nicholson et Anders Peterson (2017). "A microscopic evaluation of railway timetable robustness in critical points". In : *Journal of Rail Transport Planning & Management* 7.4, p. 207-223.
- Soyster, A. L. (1973). "Convex Programming with Set-Inclusive Constraints and Applications to Inexact Linear Programming". In : *Operations Research* 21.5, p. 1154-1157.
- Steckel, Jens (1991). "Strategische Optionen für die zufällige Fahrzeit im Eisenbahnbetrieb". Thèse de doctorat. Dresden Institute for Traffic Research.
- Stoer, Josef et Roland Bulirsch (1993). *Introduction to Numerical Analysis*. Springer.
- Su, Jau-Ming et Jen-Yu Huang (2006). "Using Ant Colony Optimization to Solve Train Timetabling Problem of Mass Rapid Transit". In : *9th Joint International Conference on Information Sciences (JCIS-06)*. Atlantis Press.
- Sun, Lijun, Alejandro Tirachini, Kay W. Axhausen, Alexander Erath et Der-Hong Lee (2014). "Models of bus boarding and alighting dynamics". In : *Transportation Research Part A : Policy and Practice* 69, p. 447-460.
- Sun, Xubin, Shaobo Zhang, Hairong Dong et Hainan Zhu (2013). "Optimal train schedule with headway and passenger flow dynamic models". In : *2013 IEEE International Conference on Intelligent Rail Transportation Proceedings*, p. 307-312.
- Takeuchi, Yoko, Norio Tomii et Chikara Hirai (2007). *Evaluation Method for Robustness of Train Schedules*. Rapp. tech. 48. Railway Technical Research Institute (RTRI), p. 97-101.
- Tarantola, Albert (2005). *Inverse problem theory and methods for model parameter estimation*. Philadelphie : SIAM.
- Teghem, Jacques (2012). *Recherche Opérationnelle, tome 1 : Méthodes d'optimisation*. Ellipses.
- Teghem, Jacques (2013). *Recherche Opérationnelle, tome 2 : Gestion de production, modèles aléatoires, aide multicritère*. Ellipses.
- Tekin, Eylem et Ihsan Sabuncuoglu (2004). "Simulation optimization : A comprehensive review on theory and applications". In : *IIE Transactions* 36.11, p. 1067-1081.
- Tormos, Pilar, Antonio Lova, Laura Ingolotti et Federico Barber (2008). *A genetic approach to robust train timetabling*. Rapp. tech. ARRIVAL-TR-0173.
- Törnquist, Johanna et Jan A. Persson (2005). "Train Traffic Deviation Handling Using Tabu Search and Simulated Annealing". In : *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*.
- TransportParis (2016). *Robustesse, haut débit et excellence ferroviaire*. url : <http://transportparis.canalblog.com/pages/robustesse--haut-debit-et-excellence-ferroviaire/32366465.html> (visité le 03/03/2020).
- Van Breusegem, Vincent, Guy Campion et Georges Bastin (1991). "Traffic modeling and state feedback control for metro lines". In : *IEEE Transactions on Automatic Control* 36.7, p. 770-784.
- Vansteenwegen, P. et D. Van Oudheusden (2006). "Developing railway timetables which guarantee a better service". In : *European Journal of Operational Research* 173.1, p. 337-350.
- Vekas, Peter, Maarten van der Vlerk et Willem Klein Haneveld (2012). "Optimizing existing railway timetables by means of stochastic programming". In : *Stochastic Programming E-Print Series* 8.
- Vermet, Franck (2017). *Cours d'introduction à la simulation stochastique*.
- Vromans, Michiel (2005). "Reliability of railway networks". Thèse de doctorat. Rotterdam : Erasmus University.
- Vromans, Michiel J.C.M., Rommert Dekker et Leo G. Kroon (2006). "Reliability and heterogeneity of railway services". In : *European Journal of Operational Research* 172.2, p. 647-665.
- Wardman, Mark (2004). "Public transport values of time". In : *Transport Policy* 11.4, p. 363-377.
- Watanabe, Shuho, Yuki Mori, Yoshinobu Takatori, Kazushige Yonemoto et Norio Tomii (2018). "Train traffic simulation algorithm based on historical train traffic records". In : *Computers in Railways XVI*. Lisbonne.

- Wei, Yu et Mu-Chen Chen (2012). "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks". In : *Transportation Research Part C : Emerging Technologies* 21.1, p. 148-162.
- Wolpert, David H. et William G. Macready (1997). "No free lunch theorems for optimization". In : *IEEE Transactions on Evolutionary Computation* 1.1, p. 67-82.
- Xin, Jianxia et Shuyan Chen (2016). "Bus Dwell Time Prediction Based on KNN". In : *Green Intelligent Transportation System and Safety* 137, p. 283-288.
- Yamamura, Akiyoshi, Masahito Koresawa, Shigeaki Aadchi, Tatsuya Inagi et Norio Tomii (2013). "Dwell time analysis in urbain railway lines using multi-agent simulation". In : *13th World Conference on Transportation Research (WCTR13)*. Rio de Janeiro.
- Yang, Xin, Jianjun Wu, Huijun Sun, Ziyu Gao, Haodong Yin et Yunchao Qu (2019). "Performance improvement of energy consumption, passenger time and robustness in metro systems : A multi-objective timetable optimization approach". In : *Computers & Industrial Engineering* 137, p. 1-13.
- Yuan, Jianxin (2006). "Stochastic modelling of train delays and delay propagation in stations". Thèse de doctorat. Delft University of Technology.
- Zhang, Qi, Baoming Han et Dewei Li (2008). "Modeling and simulation of passenger alighting and boarding movement in Beijing metro stations". In : *Transportation Research Part C : Emerging Technologies* 16.5, p. 635-649.