université PARIS-SACLAY

Apprentissage par renforcement pour la tarification dynamique avec contraintes de ressources en situation de concurrence : application à la location de voitures

Reinforcement learning for dynamic pricing under resource constraints in competitive environments : an application to car rentals

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°580, Sciences et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat : Informatique
Graduate School : Informatique et Sciences du Numérique.
Référent : Université de Versailles Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche **DAVID** (Université Paris-Saclay, UVSQ), sous la direction de **Dominique BARTH**, Professeur des universités, et le co-encadrement de **Christian CADERE**, Docteur

Thèse soutenue à Versailles, le 1er octobre 2026, par

Laurie GUENIN

Composition du jury

Membres du jury avec voix délibérative

Johanne Cohen Directrice de recherche, LISN	Présidente
Luce Brotcorne Directrice de recherche, INRIA Lille	Rapportrice & Examinatrice
Arnaud Doniec Professeur des universités, IMT Nord Europe	Rapporteur & Examinateur
Thierry Mautor Maître de conférences, UVSQ	Examinateur
Benoit Rottembourg Docteur, INRIA	Examinateur

THESE DE DOCTORAT
NNT : 2026UPASG033

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre : Apprentissage par renforcement pour la tarification dynamique avec contraintes de ressources en situation de concurrence : application à la location de voitures
Mots clés : Machine Learning, Tarification Dynamique, Location de voitures

Résumé : Ce manuscrit de thèse proposera une étude complète de la tarification dynamique en contexte concurrentiel, appliquée au secteur de la location de voitures, en combinant modélisation théorique, analyse de données réelles et méthodes d'apprentissage par renforcement. Il introduira un cadre formel de jeu répété sous contrainte de capacité, permettant de caractériser finement les interactions stratégiques entre loueurs au fil de l'arrivée des demandes.

Sur le plan théorique, le manuscrit mettra en évidence les limites des équilibres classiques et proposera des notions d'équilibre adaptées à un contexte séquentiel et en ligne, servant de référence conceptuelle pour évaluer les stratégies. Une analyse approfondie de données réelles provenant de loueurs permettra de mieux comprendre les liens entre prix, caractéristiques des demandes. Une seconde étude des données permet d'avoir une classification des jours en fonction des caractéristiques des demandes appelées saisonnalité. Cette saisonnalité a été créée dans le but d'améliorer les méthodes de tarification dynamique.

Enfin, ces fondations seront exploitées pour concevoir, comparer et évaluer plusieurs approches de tarification dynamique. On utilisera une méthode existante provenant de l'aérien qui a une approche seulement sur la ressource sans tenir compte de la concurrence. A contrario, on a comparé plusieurs méthodes d'apprentissage par renforcement pour considérer la concurrence. Une troisième méthode a été créée à la jonction entre les deux précédentes et elle va donc prendre en compte la disponibilité de la ressource et les prix des concurrents.

Pour approfondir ces méthodes, on a proposé une manière d'intégrer la saisonnalité aux méthodes de tarification dynamique dans les trois méthodes précédentes. L'évaluation des différentes méthodes a été réalisée sur les données d'un loueur en mettant en concurrence les unes avec les autres.

Title : Reinforcement learning for dynamic pricing under resource constraints in competitive environments : an application to car rental

Keywords : Dynamic pricing, car rental, Machine learning

Abstract :

This manuscript will present a comprehensive study of dynamic pricing in a competitive context, applied to the car rental industry, by combining theoretical modeling, analysis of real-world data, and reinforcement learning methods. It will introduce a formal framework for repeated games under capacity constraints, enabling a detailed characterization of the strategic interactions between rental companies as demand arrives.

On a theoretical level, the manuscript will highlight the limitations of classical equilibria and propose equilibrium concepts adapted to a sequential and online context, serving as a conceptual framework for evaluating strategies. An in-depth analysis of real-world data from rental companies will provide a better understanding of the relationships between prices and demand characteristics. A second analysis of the data allows for a classification of days based on demand characteristics known as seasonality. This seasonality was developed with the aim of improving dynamic pricing methods.

Finally, these foundations will be used to design, compare, and evaluate several dynamic pricing approaches. We will use an existing method from the airline industry that focuses solely on the resource without considering competition. Conversely, we compared several reinforcement learning methods to account for competition. A third method was developed as a hybrid of the two previous ones and will therefore take into account both resource availability and competitors' prices.

To further explore these methods, we proposed a way to integrate seasonality into the dynamic pricing methods within the three approaches. The evaluation of the different methods was conducted using data from a rental company, pitting them against one another.

Remerciements

Je tiens à remercier toutes les personnes qui m'ont accompagnée et soutenue tout au long de cette thèse, et qui ont contribué, chacune à leur manière, à son aboutissement.

Je souhaite, tout d'abord, exprimer toute ma gratitude à mes directeurs de thèse Dominique Barth et Christian Cadéré, pour leur accompagnement, leurs conseils, leurs retours et leur soutien au cours de ces années. Une thèse est un travail de longue haleine, ponctué de questionnements, de difficultés et parfois de détours. Leur disponibilité et leurs conseils ont été précieux pour mener ce travail à son terme et me permettre d'avancer tout au long de ce parcours.

Je remercie les rapporteurs de cette thèse, Luce Brotcorne et Arnaud Doniec, pour le temps consacré à la lecture de ce manuscrit, ainsi que pour leurs retours. Je remercie également les membres du jury, Johanne Cohen et Thierry Mautor, pour avoir accepté d'évaluer mon travail en participant à la soutenance. J'adresse un remerciement particulier à Benoit Rottembourg, qui a suivi mes travaux pendant ces années dans le cadre du comité de suivi individuel (CSI) et qui a également accepté de faire partie de mon jury.

Mes remerciements s'adressent également aux membres du laboratoire pour leur accueil, leur disponibilité et les nombreux échanges qui ont rythmé ces années de thèse. Parce que ces années au laboratoire n'auraient vraiment pas été les mêmes sans vous, je tiens à exprimer ma profonde reconnaissance aux doctorants que j'ai côtoyés durant ces années : Anne, Chloé, Coralie, Julie, Nassim, Nassima, Quentin, Joseph, Loric, Ylène, ainsi qu'à Catherine et Clémence, qui m'ont beaucoup aidé au quotidien.

Je tiens également à remercier toute l'équipe de WeYield, qui a joué un rôle important dans la réalisation de cette thèse. La possibilité de travailler sur des problématiques concrètes et sur des données réelles du secteur de la location de voitures a constitué un apport essentiel à ce travail. Au-delà de cette contribution directe à mes travaux de recherche, mon expérience chez WeYield m'a également offert l'occasion de découvrir d'autres problématiques et de m'investir dans des sujets variés, apportant ainsi un équilibre précieux au cours de ces années. Je souhaite remercier tout particulièrement l'équipe P&I pour sa disponibilité, son soutien et son aide précieuse dans les développements informatiques menés tout au long de cette thèse. Je tiens à exprimer ma profonde reconnaissance à Anne-Claire, Samin et Morgane, qui sont devenues bien plus que de simples collègues de bureau grâce à ces beaux moments partagés.

Un grand merci à mes amis : Chloé, Gaëlle, Inès, Noé et Zoé, qui m'ont permis de décompresser grâce à des visites culturelles, de longues discussions et des parties de jeux de société. Cette thèse ne se serait pas aussi bien déroulée sans vous.

Bien qu'une page se soit tournée, j'aimerais remercier Timothé et toute sa famille pour leur aide tout au long de ma thèse ainsi que pendant le Covid.

Enfin, je souhaite remercier mes parents pour leur soutien indéfectible, leur patience et leurs encouragements tout au long de mon parcours scolaire et universitaire. Merci d'avoir été présents au fil des années et de m'avoir accompagnée jusqu'à cette étape importante.

Un grand merci à mon frère pour le lien fusionnel qui s'est formé au fil des années malgré nos grandes différences. J'ai sincèrement apprécié l'intérêt qu'il a eu pour ma thèse, en particulier pour ce manuscrit.

À toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de cette thèse, j'adresse mes sincères remerciements.

Table des matières

1 Introduction générale **1**

2 Tarification et apprentissage **5**
2.1 Introduction . . . 5
2.2 Méthodes de tarification dynamique . . . 6
2.2.1 *Revenue management* et quelques domaines d'application . . . 7
2.2.2 Approche de tarification établie à partir de la ressource . . . 8
2.2.3 Approche de tarification établie à partir de la concurrence . . . 9
2.3 Méthodes d'apprentissage par renforcement pour la tarification dynamique . . . 12
2.3.1 Méthodes d'apprentissage par renforcement par valeurs . . . 14
2.3.2 Méthodes d'apprentissage par renforcement par politique . . . 16
2.4 Conclusion . . . 19

3 Modèle de jeu répété et Propriétés **21**
3.1 Introduction . . . 21
3.2 Contexte . . . 22
3.3 Modélisation des acteurs principaux . . . 22
3.3.1 Agent . . . 22
3.3.2 Demande . . . 23
3.4 Modèle de jeu répété . . . 23
3.4.1 Jeu unitaire . . . 24
3.4.2 Jeu répété . . . 24
3.5 Équilibre de Nash du jeu unitaire et du jeu répété . . . 26
3.5.1 Équilibre de Nash pour le jeu unitaire . . . 26
3.5.2 Équilibre de Nash pour le jeu répété . . . 29
3.5.3 Équilibre online pour le jeu répété . . . 31
3.5.4 Équilibre dans un autre modèle de jeu répété . . . 36
3.6 Conclusion . . . 38
3.7 Liste des notations . . . 40
3.7.1 Notations pour l'agent A_α (Section 3.3.1) . . . 40
3.7.2 Notations de la demande de location d (Section 3.3.2) . . . 40
3.7.3 Notations du jeu unitaire (Section 3.4.1) . . . 40
3.7.4 Notations du jeu répété (Section 3.4.2) . . . 41

4 Analyse de données **43**
4.1 Introduction . . . 43
4.2 Ensemble de données réelles . . . 44
4.3 Étude du lien entre le prix et les caractéristiques des demandes . . . 45
4.3.1 Étude des distributions . . . 46

4.4 Étude des distributions par mois . . . 51
4.4.1 Étude des corrélations . . . 54
4.4.2 Classification des données . . . 57
4.5 Étude de la saisonnalité pour un historique de données . . . 59
4.5.1 Caractérisation des jours . . . 60
4.5.2 Méthodes de classification : $K - means$ . . . 61
4.5.3 Méthodes de comparaison des classifications . . . 63
4.5.4 Obtention et évaluation d'une saisonnalité . . . 64
4.6 Historique et prédiction des données . . . 67
4.6.1 Prédiction des demandes . . . 67
4.6.2 Saisonnalité et dynamique temporelle . . . 69
4.7 Conclusion . . . 71

5 Méthodes de tarification dynamique **73**
5.1 Introduction . . . 73
5.2 Étude des deux méthodes . . . 74
5.2.1 Méthode par ressources . . . 74
5.2.1.1 Un algorithme par ressources : EMSR . . . 74
5.2.1.2 L'adaptation à notre contexte . . . 75
5.2.2 Méthodes par concurrence . . . 76
5.2.2.1 Linear Reward . . . 77
5.2.2.2 Une adaptation de la technique *Linear Reward* pour notre problème . 79
5.2.3 Comparaison des méthodes de RL . . . 80
5.2.3.1 Ensemble(s) de données . . . 80
5.2.3.2 Paramètres mesurés et scénarios . . . 81
5.2.3.3 Évaluation de performances . . . 82
5.3 Méthode Mixte . . . 88
5.3.1 Description de la méthode . . . 88
5.3.2 Évaluation de performances de la méthode Mixte . . . 91
5.4 Conclusion . . . 97

6 Méthodes de tarification avec Prédiction **99**
6.1 Introduction . . . 99
6.2 Utilisation de la saisonnalité dans les méthodes . . . 100
6.2.1 Choix de la saison liée à la demande . . . 101
6.2.2 Phase de tarification et d'apprentissage . . . 102
6.2.3 Mise à jour de la saisonnalité . . . 102
6.3 Évaluation des performances des méthodes saisonnalisées . . . 103
6.3.1 Choix de la méthode de calcul de la saisonnalité . . . 103
6.3.2 Paramètres des méthodes et scénarios . . . 106
6.3.3 Évaluation des performances contre les méthodes sans saisonnalité . . . 107
6.3.4 Évaluation des performances contre les méthodes saisonnalisées . . . 116

6.3.5 Étude du bénéfice pour une méthode saisonnalisée avec une méthode différente par saison 121
6.4 Conclusion 129

7 Conclusion **131**

A Ensemble des courbes du chapitre 5 qui seront comparées dans le chapitre 6 **145**

B Évaluation de performances du Chapitre 5 sur d'autres ensembles de données **147**
B.1 Loueur société 147
B.2 Loueur continental 150
B.3 Loueur de Van 153

C Ensemble des courbes d'une autre exécution pour comparer avec celles de la section 6.3.1 **157**

Table des figures

2.1 Apprentissage par renforcement 12

3.1 Propriétés des équilibres entre eux 34
3.2 Propriétés des équilibres entre eux 36

4.1 Distribution du nombre de demandes en fonction de la durée de location 46
4.2 Distribution du nombre de demandes en fonction de la durée de location en fonction des loueurs 47
4.3 Distribution du nombre de demandes en fonction du délai de réservation 48
4.4 Distribution du nombre de demandes en fonction du délai de réservation en fonction des loueurs 48
4.5 Distribution du nombre de demandes en fonction du RPD . . . 49
4.6 Distribution du nombre de demandes en fonction du RPD . . . 50
4.7 Distribution du nombre de demandes en fonction de la durée . 50
4.8 Distribution du nombre de demandes en fonction du taux d'utilisation maximum 51
4.9 Distribution du nombre de demandes en fonction du délai de réservation pour chaque mois pour le loueur tourisme 52
4.10 Distribution du nombre de demandes en fonction du délai de réservation pour chaque mois pour le loueur de van 53
4.11 Distribution du nombre de demandes en fonction de la durée pour chaque mois pour le loueur tourisme 53
4.12 Distribution du nombre de demandes en fonction de la durée pour chaque mois pour le loueur de van 54
4.13 Corrélation entre le délai de réservation et le taux d'utilisation maximum pour le loueur tourisme 55
4.14 Corrélation entre le délai de réservation et le taux d'utilisation maximum pour le loueur continental 55
4.15 Corrélation entre le délai de réservation et le taux d'utilisation maximum pour le loueur société 56
4.16 Corrélation entre le délai de réservation et le taux d'utilisation maximum pour le loueur de van 56
4.17 Étude de la corrélation entre le délai de réservation et le RPD, avec le RPD moyen par durée en rouge 57
4.18 Saisonnalité intuitive 60
4.19 Valeur de WCSS (définie dans la Section 4.5.2) en fonction du nombre de classes K 65

4.20 Comparaison des classifications en fonction des métriques utilisées 65
4.21 Comparaison des saisonnalités entre le loueur tourisme et le loueur continental 66
4.22 Comparaison entre la saisonnalité prédite et la saisonnalité omnisciente 71

5.1 Fonction d'utilité U 80
5.2 Distribution des valeurs de $Pmax$ 81
5.3 LRI contre EMSR 83
5.4 LRP contre EMSR 83
5.5 EXP3 contre EMSR 84
5.6 $\varepsilon - greedy$ contre EMSR 84
5.7 Évolution du bénéfice pour LRI contre EMSR en fonction de la capacité 85
5.8 Évolution du bénéfice pour LRP contre EMSR en fonction de la capacité 85
5.9 Évolution du bénéfice pour EXP3 contre EMSR en fonction de la capacité 86
5.10 Évolution du bénéfice pour $\varepsilon - greedy$ contre EMSR en fonction de la capacité 86
5.11 Ratio d'indisponibilités pour chaque agent 87
5.12 Évolution du bénéfice pour LRI contre LRI en fonction de la capacité 88
5.13 Évolution du bénéfice en situation d'équilibre en fonction de la capacité 88
5.14 Aire Totale 89
5.15 Aire Libre 89
5.16 Aire Pic 90
5.17 Diagramme des états en fonction de la valeur de nuisance et de la pression 91
5.18 LRI contre méthode Mixte 92
5.19 Évolution du bénéfice pour LRI contre la méthode Mixte en fonction de la capacité 92
5.20 Méthode Mixte contre EMSR 93
5.21 Évolution du bénéfice pour EMSR contre la méthode Mixte en fonction de la capacité 93
5.22 Évolution du vecteur de probabilité dans l'état 3 94
5.23 Évolution du nombre de demandes par états pour la méthode Mixte contre EMSR 95
5.24 Évolution du nombre de demandes par états pour la méthode Mixte contre LRI 96
5.25 Valeur de la pression pour chaque demande, pour chaque agent 97

6.1 Fonctionnement du système pour une demande 101
6.2 Distribution du RPD moyen calculé pour chaque jour 104
6.3 Évolution du RPD moyen sur deux ans 104
6.4 Évolution de la saisonnalité 105
6.5 Évolution du nombre de demandes par saison 105
6.6 Évolution de la taille des saisons 105
6.7 Indice de Jaccard à chaque mise à jour de la saisonnalité 105
6.8 Évolution de la saisonnalité sur la phase d'étude 106
6.9 La méthode saisonnalisée Mixte contre EMSR 108
6.10 Évolution du bénéfice pour la méthode saisonnalisée Mixte contre EMSR en fonction de la capacité 108
6.11 La méthode saisonnalisée Mixte contre LRI 109
6.12 Évolution du bénéfice pour la méthode saisonnalisée Mixte contre LRI en fonction de la capacité 110
6.13 La méthode saisonnalisée EMSR contre la méthode Mixte . . . 111
6.14 Évolution du bénéfice pour la méthode saisonnalisée EMSR contre la méthode Mixte en fonction de la capacité 111
6.15 La méthode saisonnalisée EMSR contre LRI 112
6.16 Évolution du bénéfice pour la méthode saisonnalisée EMSR contre LRI en fonction de la capacité 113
6.17 La méthode saisonnalisée LRI contre EMSR 114
6.18 Évolution du bénéfice pour la méthode saisonnalisée LRI contre EMSR en fonction de la capacité 114
6.19 La méthode saisonnalisée LRI contre la méthode Mixte 115
6.20 Évolution du bénéfice pour la méthode saisonnalisée LRI contre la méthode Mixte en fonction de la capacité 115
6.21 La méthode saisonnalisée EMSR contre la méthode saisonnalisée Mixte 117
6.22 Évolution du bénéfice pour la méthode saisonnalisée EMSR contre la méthode saisonnalisée Mixte en fonction de la capacité . . . 117
6.23 La méthode saisonnalisée EMSR contre la méthode saisonnalisée LRI 118
6.24 Évolution du bénéfice pour la méthode saisonnalisée EMSR contre la méthode saisonnalisée LRI en fonction de la capacité 119
6.25 La méthode saisonnalisée Mixte contre la méthode saisonnalisée LRI 120
6.26 Évolution du bénéfice pour la méthode saisonnalisée Mixte contre la méthode saisonnalisée LRI en fonction de la capacité 120
6.27 EMSR contre la méthode différente par saison 121
6.28 Évolution du bénéfice pour EMSR contre la méthode saisonnalisée différente par saison en fonction de la capacité 122
6.29 LRI contre la méthode saisonnalisée différente par saison 123

6.30 Évolution du bénéfice pour LRI contre la méthode saisonnalisée différente par saison en fonction de la capacité 123
6.31 La méthode Mixte contre la méthode saisonnalisée différente par saison 124
6.32 Évolution du bénéfice pour la méthode Mixte contre la méthode saisonnalisée différente par saison en fonction de la capacité . . 124
6.33 La méthode saisonnalisée EMSR contre la méthode saisonnalisée différente par saison 125
6.34 Évolution du bénéfice pour la méthode saisonnalisée EMSR contre la méthode saisonnalisée différente par saison en fonction de la capacité 126
6.35 La méthode saisonnalisée LRI contre la méthode saisonnalisée différente par saison 126
6.36 Évolution du bénéfice pour la méthode saisonnalisée LRI contre la méthode saisonnalisée différente par saison en fonction de la capacité 127
6.37 La méthode saisonnalisée Mixte contre la méthode saisonnalisée différente par saison 127
6.38 Évolution du bénéfice pour la méthode saisonnalisée Mixte contre la méthode saisonnalisée différente par saison en fonction de la capacité 128

A.1 LRI contre EMSR 145
A.2 Évolution du bénéfice pour LRI contre EMSR en fonction de la capacité 145
A.3 LRI contre méthode mixte 145
A.4 Évolution du bénéfice pour LRI contre la méthode mixte en fonction de la capacité 145
A.5 Méthode mixte contre EMSR 145
A.6 Évolution du bénéfice pour EMSR contre la méthode mixte en fonction de la capacité 145

B.1 LRI contre méthode mixte pour le loueur société 147
B.2 LRI contre EMSR pour le loueur société 148
B.3 EMSR contre méthode mixte pour le loueur société 148
B.4 Évolution du bénéfice pour LRI contre la méthode mixte en fonction de la capacité pour le loueur société 149
B.5 Évolution du bénéfice pour LRI contre EMSR en fonction de la capacité pour le loueur société 149
B.6 Évolution du bénéfice pour EMSR contre la méthode mixte en fonction de la capacité pour le loueur société 150
B.7 LRI contre méthode mixte pour le loueur continental 150
B.8 LRI contre EMSR pour le loueur continental 151

B.9 EMSR contre méthode mixte pour le loueur continental 151
B.10 Évolution du bénéfice pour LRI contre la méthode mixte en fonction de la capacité pour le loueur continental 152
B.11 Évolution du bénéfice pour LRI contre EMSR en fonction de la capacité pour le loueur continental 152
B.12 Évolution du bénéfice pour EMSR contre la méthode mixte en fonction de la capacité pour le loueur continental 153
B.13 LRI contre méthode mixte pour le loueur de Van 153
B.14 LRI contre EMSR pour le loueur de Van 154
B.15 EMSR contre méthode mixte pour le loueur de Van 154
B.16 Évolution du bénéfice pour LRI contre la méthode mixte en fonction de la capacité pour le loueur de Van 155
B.17 Évolution du bénéfice pour LRI contre EMSR en fonction de la capacité pour le loueur de Van 155
B.18 Évolution du bénéfice pour EMSR contre la méthode mixte en fonction de la capacité pour le loueur de Van 156

C.1 Évolution de la saisonnalité 157
C.2 Évolution du nombre de demandes par saison 157
C.3 Évolution de la taille des saisons 157
C.4 Indice de Jacquard à chaque mise à jour de la saisonnalité . . . 157

Chapitre 1
Introduction générale

La tarification dynamique consiste à adapter, au fil du temps, le prix de chaque produit ou service en fonction de paramètres internes ou externes au produit. Cette technique est employée dans de nombreux domaines et elle peut être réalisée soit humainement soit informatiquement. Lorsque c'est un humain qui choisit les prix, celui-ci va prendre en compte de nombreuses données. Cependant, il est difficile de savoir exactement ce qu'il considère car il utilise aussi son expérience. Les personnes qui font de la tarification dynamique utilisent généralement trois indicateurs pour fixer leurs prix. Tout d'abord, elles cherchent à estimer le prix maximum que chaque client est prêt à payer pour sa location. Elles considèrent aussi les prix de la concurrence pour choisir le prix à proposer. De plus, elles font attention à ne pas fixer des prix trop bas, ce qui les empêcherait de vendre à l'avenir à un prix plus élevé, car elles auraient utilisé toutes leurs ressources. Pour choisir ces prix en suivant ces trois paradigmes, elles utilisent les données historiques, leurs connaissances des concurrents ainsi que celles des clients.

La tarification dynamique est donc influencée par la disponibilité de la ressource et par les prix réalisés par les concurrents. Dans certains domaines où la concurrence est plus faible, due au nombre de concurrents ou, comme dans l'aérien, due à la différence de qualité de service, les méthodes de tarification dynamique peuvent seulement considérer la ressource comme la méthode EMSR de Peter Belobaba [2]. Dans d'autres domaines utilisant la tarification dynamique, la concurrence est le point le plus important à considérer et la disponibilité de la ressource est un problème négligeable, comme dans le commerce en ligne. Le but est alors d'avoir le plus de clients possible.

Dans la location de voitures, la disponibilité de la ressource, ici les voitures, est un des points importants à considérer, sachant qu'elles sont séparées en beaucoup de classes tarifaires en fonction de la typologie de la voiture. Une voiture de type Mini ne va pas être louée au même prix qu'un SUV ; ainsi, le stock par catégorie est assez faible. Il est important de considérer la concurrence dans le domaine de la location de voitures car les services vendus sont similaires, les typologies de voitures sont identiques, les lieux de retrait sont généralement similaires, proches des gares et des aéroports. Ainsi, dans le domaine de la location de voitures pour faire de la tarification dynamique, il va falloir considérer la concurrence et la disponibilité de la ressource. Dans ce contexte, l'entreprise WeYield fournit un logiciel pour la tarification dynamique et du coaching par des experts pour les entreprises de location de voitures. Ce logiciel permet aux loueurs d'étudier leurs réservations et leurs

indicateurs clés de performance (KPIs) ainsi que les prix des concurrents. La thèse ayant été réalisée en partenariat avec la société WeYield, les études nécessitant des données issues de la location de voitures seront faites avec des données réelles provenant des clients de WeYield. Cette thèse a pour but de fournir à WeYield une solution automatisée de tarification dynamique qui sera alors vendue aux clients.

Pour faire de la tarification dynamique automatisée dans le cadre de la location de voitures, il paraît pertinent de s'intéresser à des techniques de *Machine Learning*. Cependant, la location de voitures a des contraintes qui réduisent le champ des possibles en termes de méthodes de *Machine Learning* car dans le domaine de la location de voitures, les données historiques sont accessibles, mais leur quantité est assez faible. Ainsi, les techniques de *Machine Learning*, comme le *Deep Learning*, nécessitant beaucoup de données, ne paraissent pas viables dans ce contexte. De plus, le domaine de la location de voitures évolue fortement en fonction de la concurrence, ce qui le rend fortement non stationnaire, ce qui pose problème dans le cadre de l'apprentissage supervisé et non supervisé. Ainsi, l'apprentissage par renforcement, qui nécessite peu de données et s'adapte en temps réel aux changements de comportement, semble être l'approche la plus pertinente pour résoudre ce problème de tarification dynamique appliqué à la location de voitures.

Cette thèse aura pour but de répondre à la question suivante : Est-ce que l'apprentissage par renforcement permet d'avoir des méthodes efficaces pour faire de la tarification dynamique à la demande en considérant la concurrence avec des contraintes de ressources ? Ces méthodes seront comparées avec une méthode de l'état de l'art provenant de l'aérien, EMSR [2] qui choisit les prix en fonction de la disponibilité de la ressource.

Ce manuscrit est composé de sept chapitres organisés comme suit : Le chapitre 1 introduit le contexte général ainsi que les motivations de nos travaux de recherche.

Le chapitre 2 présente diverses approches existantes de tarification dynamique. Pour cela, le chapitre définit ce qu'est le *Revenue Management* (RM) qui regroupe un ensemble de techniques utilisées pour faire de la tarification dynamique et quelques domaines d'application. Ces techniques ont initialement été créées pour la vente de billets d'avion. Ce chapitre décrit ensuite les méthodes d'apprentissage par renforcement qui seront utilisées pour faire de la tarification dynamique.

Le chapitre 3 décrit le modèle de jeux répétés utilisé pour décrire le problème de tarification dynamique pour la location de voitures. Dans le cadre de ce modèle, une étude de plusieurs types d'équilibre a été réalisée.

Dans le chapitre 4, le but est d'étudier les données de location de voitures de plusieurs clients de WeYield pour essayer de déterminer comment ces clients ont fait pour choisir leur prix. Pour cela, plusieurs méthodes ont été utilisées

pour chercher un lien entre les caractéristiques des locations et leur prix. Dans la suite de ce chapitre, une méthode a été développée pour classifier les jours de l'année en fonction des caractéristiques des locations liées à ces jours. Cette classification est appelée saisonnalité.

Le chapitre 5 présente trois méthodes de tarification dynamique. La première provient de l'aérien et a été adaptée pour fonctionner dans le domaine de la location de voitures. Cette méthode fixe les prix en fonction de la disponibilité de la ressource. La seconde utilise l'apprentissage par renforcement pour adapter les prix en fonction de la concurrence. Pour cela, plusieurs méthodes d'apprentissage par renforcement ont été comparées pour garder celle qui donne le meilleur bénéfice. La dernière méthode est une fusion des deux précédentes, le but étant de considérer la disponibilité de la ressource ainsi que les prix des concurrents.

Le chapitre 6 propose une manière d'intégrer la saisonnalité dans les méthodes de tarification décrites dans le chapitre 5. Ce chapitre montrera alors que la saisonnalité est un apport majeur en termes de bénéfice.

Enfin, le chapitre 7 conclut le manuscrit en résumant les contributions apportées, ainsi qu'en présentant leurs limites et les perspectives de recherche qu'apportent ces travaux.

Chapitre 2
Tarification et apprentissage

Sommaire

2.1 Introduction 5
2.2 Méthodes de tarification dynamique 6
2.2.1 *Revenue management* et quelques domaines d'application 7
2.2.2 Approche de tarification établie à partir de la ressource 8
2.2.3 Approche de tarification établie à partir de la concurrence 9
2.3 Méthodes d'apprentissage par renforcement pour la tarification dynamique 12
2.3.1 Méthodes d'apprentissage par renforcement par valeurs . 14
2.3.2 Méthodes d'apprentissage par renforcement par politique 16
2.4 Conclusion 19

2.1. Introduction

La détermination d'une politique de prix visant à maximiser le bénéfice constitue un problème étudié dans de nombreux contextes et abordé de différentes façons [23]. On nommera tarification dynamique le fait de faire varier les prix en fonction des demandes [18]. Ce problème n'est pas seulement présent dans le secteur de la location de voitures. En effet, de nombreux secteurs doivent fixer un prix pour chaque vente de produits ou de services, comme l'hôtellerie, le transport, la grande distribution, l'électricité et la billetterie [64]. Ces secteurs utilisent depuis plus ou moins longtemps des techniques de *Revenue Management*, qui est une branche spécifique de la tarification dynamique définie dans la section 2.2.1. Bien qu'ils utilisent des techniques de *Revenue Management*, ces différents domaines présentent des caractéristiques distinctes, ce qui implique une gestion des prix propre à chacun [18]. Habituellement, les méthodes de tarification dynamique sont soit plutôt orientées sur la disponibilité de la ressource (dans la section 2.2.2), soit sur les prix de la concurrence (dans la section 2.2.3).

L'essor des méthodes d'apprentissage automatique (*Machine Learning*, ML) a conduit ces secteurs d'activité à intégrer ces techniques afin d'améliorer leurs performances et d'obtenir des bénéfices supérieurs à ceux obtenus avec les méthodes traditionnellement utilisées [28]. Dans la section 2.3, le but sera de décrire les différentes options qui s'offrent à nous en termes d'apprentissage par renforcement pour la tarification dynamique.

2.2. Méthodes de tarification dynamique

La tarification dynamique peut être définie comme un processus décisionnel séquentiel dans lequel une entreprise choisit le prix d'un produit ou d'un service en tenant compte d'un ensemble d'informations disponibles à chaque instant, telles que l'état des ressources, les caractéristiques de la demande ou l'environnement concurrentiel [64]. La tarification dynamique constitue un problème central dans les secteurs caractérisés par une capacité périssable, une demande incertaine et une forte concurrence. Le secteur de la location de voitures cumule ces trois dimensions : les véhicules constituent une ressource limitée, la demande est séquentielle et hétérogène, et les acteurs sont engagés dans une compétition tarifaire permanente. Il n'est pas nécessaire d'avoir des prix différents pour chaque combinaison de paramètres possibles (jour de vente, type de ressource,...) du produit ou du service vendu pour faire de la tarification dynamique. La tarification est dynamique à partir du moment où le prix est différent en fonction des caractéristiques de la demande ainsi que du contexte extérieur comme les prix des concurrents. Ainsi, par exemple, s'il y a un prix différent pour un achat le week-end, par rapport au prix de la semaine, alors c'est déjà de la tarification dynamique même si elle n'est pas complexe.

Les premiers travaux sur la tarification dynamique apparaissent dans les années 1970 dans le contexte du transport aérien. L'article de Littlewood [45] introduit l'une des premières règles décisionnelles permettant d'arbitrer entre une vente immédiate à un prix donné et la possibilité d'une vente future à un prix plus élevé. Ce travail constitue un point de départ majeur pour le développement du *Revenue Management*, discipline consacrée à l'optimisation des revenus générés par une ressource limitée.

Depuis lors, la littérature scientifique s'est considérablement enrichie. De nombreux modèles ont été proposés pour analyser les décisions de tarification dans différents contextes économiques. Ces modèles se distinguent notamment par les hypothèses qu'ils formulent concernant la demande, la capacité et le comportement des concurrents.

Par exemple, dans le cadre du *Revenue Management*, des modèles fondateurs comme celui de Gallego et Van Ryzin [22] considèrent une demande stochastique et une capacité fixe. Dans la continuité, des approches synthétiques comme celles de Talluri et Van Ryzin [64] généralisent ces modèles en intégrant la gestion conjointe des prix et de la capacité dans des environnements multiproduits. D'autres travaux introduisent des modèles où la demande est apprise au cours du temps, comme dans l'approche de Besbes et Zeevi [5], qui combinent tarification dynamique et apprentissage statistique afin d'explorer l'élasticité du prix. Enfin, dans des domaines appliqués tels que l'énergie ou les plateformes numériques, des modèles spécifiques ont été développés pour intégrer des contraintes opérationnelles fortes, comme la gestion en temps réel de l'offre et de la demande dans les réseaux électriques [11] ou les mécanismes

de surge pricing, qui correspondent à une tarification en temps réel qui est fortement reliée au nombre de demandes, dans les plateformes de mobilité comme Uber [37].

La littérature distingue généralement deux grandes perspectives dans l'étude de la tarification dynamique. La première met l'accent sur la gestion d'une ressource limitée, comme dans le *Revenue Management*. La seconde considère principalement les interactions stratégiques entre entreprises dans un contexte de concurrence. Ces deux approches constituent les principaux cadres d'analyse de la tarification dynamique.

2.2.1. *Revenue management* et quelques domaines d'application

Le *Revenue Management* (RM) ou gestion des revenus regroupe un ensemble de méthodes visant à maximiser le revenu généré par une ressource limitée en adaptant simultanément les prix et l'allocation de la capacité en fonction de la demande [56, 64]. Robert G. Cross dans [15] dit que *"Revenue Management ensures that companies will sell the right product to the right customer at the right time for the right price."*, ce qui peut se traduire par "le RM permet aux entreprises de vendre le bon produit au bon client, au bon moment et au bon prix". Cette approche est particulièrement adaptée aux secteurs caractérisés par des capacités fixes, des coûts variables faibles, des produits périssables et une demande fluctuante.

Historiquement, ces méthodes ont été développées dans l'industrie du transport aérien à la suite de la déréglementation du secteur aux États-Unis en 1970. Cette déréglementation a entraîné une intensification de la concurrence entre compagnies aériennes et a conduit les entreprises à développer des stratégies sophistiquées de gestion des prix et des capacités [61]. Le *Revenue Management* repose sur trois caractéristiques principales des marchés auxquels il est appliqué [70] : une capacité limitée qui ne peut pas être ajustée à court terme, un produit périssable, dont la valeur disparaît après une certaine date, une demande incertaine et segmentée.

Ces caractéristiques se retrouvent dans plusieurs secteurs économiques. L'industrie hôtelière constitue l'un des domaines d'application les plus importants du *Revenue Management*. Les hôtels doivent gérer un nombre limité de chambres, dont la valeur disparaît après la date de séjour si elles ne sont pas vendues [40]. Dans l'hôtellerie, le RM s'est particulièrement développé avec des techniques spécifiques comme le *length-of-stay pricing* (tarification selon la durée du séjour), le *day-of-week pricing* (tarification selon le jour de la semaine), et le *lead time pricing* (tarification selon l'avance de réservation) [68]. Les hôteliers utilisent désormais des *Revenue Management System* (RMS) sophistiqués qui intègrent des données sur les événements locaux, la concurrence, les tendances de réservation en ligne, et même les prévisions météorologiques

pour optimiser leurs tarifs quotidiennement.

De même, les entreprises de location de voitures doivent optimiser l'utilisation de leur flotte tout en tenant compte des fluctuations saisonnières de la demande [26]. Cette fluctuation saisonnière sera détaillée dans le Chapitre 4. Les compagnies ferroviaires, les compagnies de croisière et les services de transport urbain utilisent également des techniques de *Revenue Management* pour adapter leurs prix aux variations de la demande [47]. En particulier, la SNCF en France et Eurostar utilisaient des stratégies de tarification dynamique semblables à celles de l'aérien, avec des prix qui augmentent à mesure que les places se remplissent et que la date de départ approche.

L'utilisation du *Revenue Management* ne se limite pas au domaine du tourisme, elle va aussi s'étendre au secteur publicitaire à travers la tarification de créneaux publicitaires télévisés. Plus récemment, ces méthodes ont été adoptées dans le commerce en ligne [12], la billetterie [64] et le transport de marchandises [7]. Les entreprises peuvent désormais ajuster leurs prix de manière quasi instantanée en fonction des conditions du marché et du comportement des consommateurs [18].

Dans ce contexte, les décisions de tarification reposent sur des modèles d'optimisation qui peuvent considérer la demande, l'état du système ainsi que les prix des concurrents ou tout paramètre extérieur qui paraît pertinent pour le décisionnaire, comme la météo qui peut être utilisée dans le domaine de l'électricité [29] ou le ressenti d'injustice des clients [48].

2.2.2. Approche de tarification établie à partir de la ressource

Les approches fondées sur la gestion de la ressource considèrent principalement la contrainte de capacité comme le facteur déterminant dans la tarification dynamique. Dans ce cadre, l'objectif consiste à allouer une ressource limitée à différents types de clients afin de maximiser le revenu total.

Les premiers modèles développés dans ce domaine reposent sur des règles de protection de capacité pour le secteur de l'aérien. Littlewood [45] propose une règle simple permettant de déterminer s'il faut accepter une réservation à un prix donné en comparant le revenu immédiat au revenu attendu d'une vente future à un prix plus élevé.

Ces travaux ont ensuite été étendus dans le cadre de modèles plus généraux. Belobaba [2] introduit les méthodes EMSR (*Expected Marginal Seat Revenue*) qui permettent de déterminer des seuils de réservation pour différentes classes tarifaires en fonction de la distribution de la demande. Ces méthodes reposent sur l'idée de protéger une partie de la capacité pour les clients susceptibles de payer un prix plus élevé. Une adaptation de ces méthodes sera définie et utilisée dans le chapitre 5.

Les travaux en hôtellerie [49], dans le ferroviaire [17] et dans le tourisme [9]

reprennent les principes, vus dans le secteur de l'aérien, en les adaptant à des ressources telles que les chambres d'hôtel, les cabines de croisière ou d'autres capacités de service, tout en mettant l'accent sur l'intégration des prévisions de demande, de la segmentation client et de la modulation des capacités. Dans ces secteurs, les synthèses montrent que les modèles probabilistes de type EMSR restent des briques essentielles pour décider de l'allocation de la capacité par classe tarifaire [55], tandis que les outils de big data et d'intelligence artificielle servent à affiner les prévisions et à automatiser les ajustements de prix en temps réel. La tarification dynamique reste donc, là encore, étroitement liée à l'état de la ressource (taux d'occupation, disponibilité future anticipée) et à l'objectif de maximisation du revenu par unité de capacité.

Dans le cas de la location de voitures, ces idées de tarification dynamique basée sur la ressource sont transposées à la gestion d'une flotte de véhicules répartie dans un réseau de stations et soumise à des durées de location variables [53]. Des travaux récents sur le « *car rental network revenue management* » modélisent la flotte comme un réseau de ressources périssables et utilisent des contrôles de type *bid price* [44] : à chaque véhicule-station-période est associée une valeur minimale appelée *bid price*, et la décision d'accepter une demande à un certain prix consiste à comparer cette recette au *bid price* de la ressource correspondante. Si le prix de la demande est inférieur à ce *bid price*, la demande est refusée, ce qui revient à une politique de tarification basée sur la capacité où les prix minimaux acceptables dépendent de la valeur marginale de la capacité restante dans le réseau.

Malgré leur succès, les approches basées sur la ressource présentent certaines limites. Plusieurs extensions ont été proposées pour résoudre certaines limites et prendre en compte des phénomènes tels que l'annulation des réservations, les demandes multiples ou le fait de remplacer un produit demandé par un autre [8, 64]. Mais les approches basées sur la ressource supposent généralement que la demande dépend uniquement du prix proposé par l'entreprise et ne prennent pas explicitement en compte les décisions des concurrents. Dans de nombreux marchés, cette hypothèse peut être restrictive, car les consommateurs comparent les offres de différents fournisseurs avant de prendre une décision.

Ces limitations ont conduit au développement de modèles intégrant explicitement la dimension concurrentielle.

2.2.3. Approche de tarification établie à partir de la concurrence

Les approches de tarification dynamique fondées sur la concurrence considèrent que le prix optimal d'une entreprise dépend explicitement des prix pratiqués par ses concurrents et des réactions stratégiques qu'ils peuvent adopter au cours du temps [6, 43]. Elles prolongent les modèles classiques de concur-

rence en prix de type Bertrand [4], dans lesquels plusieurs firmes choisissent leurs prix sur un même marché, en les inscrivant dans un cadre dynamique où les décisions sont révisées de manière répétée à mesure que la demande se réalise. Dans ce contexte, la politique de prix ne dépend plus de la disponibilité de la ressource, mais de la structure concurrentielle du marché et de la manière dont chaque entreprise anticipe les ajustements de ses rivales [32].

La littérature scientifique récente montre que cette dynamique concurrentielle est marquée dans les marchés de services à capacité périssable, comme le transport aérien [32] ou l'hôtellerie [64]. Les travaux de Williams et al. [31] montrent qu'un équilibre concurrentiel peut conduire à des prix trop faibles en début d'horizon, phénomène qualifié de *Bertrand scarcity trap*, lorsque les entreprises cherchent trop tôt à capter la demande par peur d'être dépassées par leurs concurrentes. Ainsi, la concurrence en prix ne conduit pas nécessairement à une trajectoire tarifaire optimale du point de vue du revenu, même lorsque les entreprises se comportent de manière rationnelle.

Ces limites ont conduit plusieurs auteurs à étudier la tarification concurrentielle dans un cadre d'apprentissage, où les entreprises ne connaissent ni parfaitement la demande ni le comportement de leurs concurrents. Dans ces modèles, chaque firme est représentée comme un agent qui ajuste progressivement ses prix à partir des revenus observés, selon des méthodes issues des bandits manchots ou de l'apprentissage par renforcement. Friedrich, Pásztor et Ramponi [20] montrent notamment que des algorithmes comme *Deep Q-Network* ou *Soft Actor-Critic*, qui seront décrits plus loin, peuvent apprendre des stratégies de prix efficaces dans des marchés oligopolistiques dynamiques. En particulier, lorsque les autres vendeurs adaptent eux aussi leurs prix au fil du temps. Ces travaux confirment que les approches concurrentielles sont bien adaptées aux environnements non stationnaires. Ils montrent cependant qu'elles peuvent faire émerger des comportements instables ou des formes de collusion tacite selon l'algorithme utilisé et la structure du marché [38]. En conséquence, la littérature scientifique converge vers l'idée qu'une approche de tarification fondée uniquement sur la concurrence est puissante pour représenter les interactions stratégiques entre firmes, mais qu'elle doit être encadrée par des hypothèses précises sur la demande, la différenciation des produits et les mécanismes d'apprentissage mobilisés. [48]

Le cas de la location de voitures constitue un terrain particulièrement pertinent pour les approches fondées sur la concurrence [64], car les offres des différents loueurs sont souvent perçues comme proches par les clients, tandis que les prix sont facilement observables via les comparateurs et plateformes de réservation en ligne. Dans ce secteur, le client compare généralement plusieurs offres portant sur une même catégorie de véhicule, pour des dates et un lieu donnés, ce qui renforce la pression concurrentielle sur le prix affiché. La décision tarifaire ne peut donc pas être interprétée uniquement comme une

réponse à la demande propre de l'entreprise, mais doit être comprise comme un positionnement relatif vis-à-vis des autres acteurs présents sur le marché [16].

Dans ce secteur, la concurrence est renforcée par la numérisation des ventes et par la transparence croissante des tarifs. Les plateformes de réservation rendent visibles, presque en temps réel, les écarts de prix entre loueurs pour une même demande, ce qui incite les entreprises à adapter rapidement leurs prix afin de préserver leur attractivité commerciale. Cette configuration rapproche la location de voitures des marchés à concurrence dynamique étudiés dans la littérature économique, où chaque entreprise observe des signaux de marché publics et ajuste sa stratégie en fonction des mouvements de prix de ses rivales. La location de voitures présente toutefois une particularité importante : les produits sont proches, mais pas parfaitement homogènes, car des différences de marque, d'assurance, de limite kilométrique ou de qualité de service peuvent moduler la sensibilité effective des clients au prix.

La littérature scientifique du secteur montre aussi que la tarification ne peut pas être dissociée complètement des contraintes opérationnelles, même lorsqu'on adopte un point de vue concurrentiel. Les travaux sur le *car rental network revenue management* [25, 44] mettent en évidence que les entreprises doivent gérer une flotte répartie entre plusieurs stations, avec des durées de location variables, ce qui rend chaque décision de prix importante. Même si ces travaux s'intéressent souvent prioritairement à la ressource, ils montrent indirectement pourquoi la dimension concurrentielle est centrale : lorsqu'un client peut choisir entre plusieurs loueurs pour une même demande, le prix proposé doit rester compétitif tout en tenant compte de la valeur future du véhicule dans le réseau. Cela justifie, dans la location de voitures, le recours à des approches de tarification où le prix est pensé comme le résultat d'un arbitrage entre positionnement concurrentiel immédiat et préservation du revenu futur.

Ainsi, dans le domaine de la location de voitures, une approche fondée sur la concurrence apparaît particulièrement pertinente pour modéliser les guerres de prix locales, les réactions rapides aux plateformes de réservation et l'importance du positionnement relatif entre loueurs. Elle permet de représenter explicitement le fait qu'un même client peut mettre en balance plusieurs offres très proches et choisir celle qui lui paraît la plus avantageuse au regard du prix et de quelques attributs de service. Toutefois, comme dans les autres secteurs étudiés par la littérature, une approche purement concurrentielle reste insuffisante si elle ignore complètement les contraintes de flotte et la valeur intertemporelle de la capacité. C'est pourquoi la location de voitures constitue un cas d'étude particulièrement intéressant pour articuler, dans le Chapitre 5, la concurrence tarifaire avec les contraintes de ressources propres aux systèmes de réservation.

2.3. Méthodes d'apprentissage par renforcement pour la tarification dynamique

Le but de cette section sera d'étudier quelles sont les techniques d'apprentissage par renforcement intéressantes à utiliser dans un second temps pour faire de la tarification dynamique dans un domaine avec des contraintes de ressources et une forte concurrence et en particulier pour la location de voitures.

L'apprentissage par renforcement (*Reinforcement Learning*, RL) constitue une manière de faire de la tarification dynamique dans des environnements incertains et évolutifs. C'est pour cette raison qu'il sera utilisé dans les chapitres 5 et 6. Contrairement aux approches classiques du *Revenue Management*, qui reposent sur des modèles explicites de la demande, le RL permet d'apprendre directement une politique de décision à partir des interactions avec l'environnement [62].

Le cadre standard du RL repose sur l'interaction entre un agent et un environnement. À chaque instant t, l'agent observe l'environnement, choisit une action et reçoit ensuite une récompense liée à cette action. La stratégie utilisée par l'agent pour choisir ses actions sera appelée politique. L'environnement peut dans certains cas être décrit par un ensemble fini d'états, ce qui permet d'utiliser des méthodes d'apprentissage par renforcement différentes.

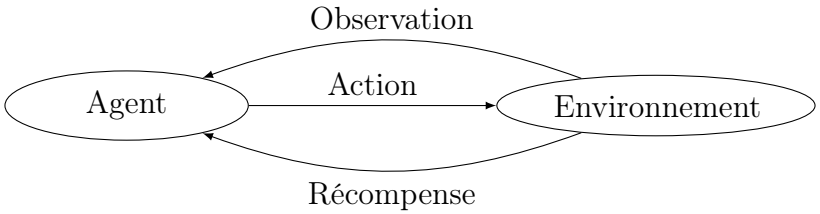

FIGURE 2.1 – Apprentissage par renforcement

Pour apprendre à tirer le meilleur parti de ces interactions, l'agent doit choisir entre tester de nouvelles actions pour en savoir davantage sur l'environnement (exploration) et utiliser ses connaissances actuelles de l'environnement pour choisir la meilleure action (exploitation). La recherche d'un équilibre optimal entre ces deux stratégies est connue sous le nom de dilemme exploration-exploitation. Certaines méthodes comme *Q-learning* [69] ont ces deux approches clairement séparées dans leur processus alors que d'autres méthodes comme UCB [1, 42] mélangent l'exploration et l'exploitation. Les méthodes d'apprentissage par renforcement peuvent aussi être catégorisées en fonction de l'élément qu'elles cherchent à apprendre au cours des étapes. Elles peuvent apprendre le gain potentiel de chaque action, ce qui s'appelle un apprentissage par valeur, ou elles peuvent apprendre la stratégie optimale à jouer au sens des jeux répétés, appelée apprentissage par politique. L'apprentissage

par politique consiste à apprendre la probabilité de jouer chacune des actions.

Dans le cadre de la tarification dynamique, chaque élément de la figure 2.1 sera défini comme suit :

- L'agent = L'acteur qui vend un service ou un produit.
- L'environnement = Les clients et les concurrents de l'agent.
- Les actions = Choisir un prix.
- La récompense = Recevoir ou non un paiement.

La tarification dynamique est généralement modélisée comme un processus de décision séquentiel dans lequel un agent choisit un prix à chaque instant afin de maximiser un revenu cumulé, ce qui convient parfaitement à l'utilisation de l'apprentissage par renforcement. Cette approche est particulièrement adaptée lorsque le comportement de l'environnement est inconnu, non stationnaire ou dépend de nombreux facteurs contextuels. Les méthodes d'apprentissage par renforcement peuvent être séparées en deux catégories, celles *on-policy* où l'agent apprend à partir des données générées par la même politique qu'il optimise et celles *off-policy* où l'agent apprend à partir de données générées par une autre politique. Les méthodes *off-policy* ne semblent pas adaptées à la tarification dynamique car l'objectif est de maximiser le bénéfice même pendant la phase d'apprentissage, ce qui n'est pas le cas lorsque la politique jouée n'est pas celle qui est optimisée.

Il va être montré maintenant que les méthodes avec états ne correspondent pas forcément au domaine de la location de voitures, mais que ces méthodes peuvent fonctionner pour d'autres domaines qui utilisent la tarification dynamique.

Les méthodes d'apprentissage par renforcement avec états sont utilisées pour résoudre des problèmes modélisés comme des processus de décision markoviens (MDP) [57]. Un processus de décision markovien est un modèle stochastique où un agent prend des décisions dans un environnement incertain, avec des transitions d'états qui dépendent uniquement de l'état actuel et de l'action choisie. La modélisation du problème de tarification dynamique a rarement été modélisée avec des MDPs mais ce fut fait dans l'article [75] qui étudie la tarification dans le contexte aérien quand les clients réservent plusieurs avions par trajet. Ils comparent alors différentes méthodes d'apprentissage par états pour obtenir le meilleur bénéfice : A2C, PPO, TRPO, DQN. L'article [58] montre comment utiliser la méthode *Q-learning* [69] et DQN pour la vente de multiples produits périssables.

La modélisation sous forme de processus de décision markovien (MDP) est plus complexe dans le cadre de la location de voitures, car la ressource est bidimensionnelle, ainsi, chaque véhicule peut être loué pour chaque jour. La bidimensionnalité de la ressource est peu courante dans les domaines utilisant

de la tarification dynamique, l'un des seuls domaines où c'est le cas est le domaine de l'hôtellerie. La bidimensionnalité de la ressource implique de créer un grand ensemble d'états pour obtenir un processus de décision markovien, ce qui n'est pas réellement pertinent pour ensuite utiliser des méthodes d'apprentissage par renforcement par états. Ainsi, par la suite, on considérera seulement les méthodes sans état.

Lors de l'utilisation de méthodes sans état, la dynamique du système est négligée ou difficile à modéliser. Le problème est alors traité comme une succession de décisions indépendantes, ce qui correspond au cadre des bandits manchots. Le problème du bandit manchot se formule de manière imagée de la façon suivante : un utilisateur (un agent), face à des machines à sous, doit décider quelles machines jouer. Chaque machine donne une récompense moyenne que l'utilisateur ne connaît pas a priori. L'objectif est de maximiser le gain cumulé de l'utilisateur. Dans ce cadre, pour le problème de la tarification dynamique, l'agent choisit un prix au lieu de choisir une machine et observe une récompense sans modéliser explicitement l'évolution d'un état. On va maintenant étudier les méthodes d'apprentissage qui correspondent au cadre des bandits manchots en les séparant en deux catégories : les méthodes par valeurs et celles par politique, comme définies au début de la section 2.3.

2.3.1. Méthodes d'apprentissage par renforcement par valeurs

Les méthodes d'apprentissage par renforcement basées sur la valeur consistent à estimer la valeur de chaque action, permettant d'évaluer la qualité des décisions. La politique optimale est ensuite obtenue en fonction des valeurs des actions.

Les algorithmes par valeurs les plus couramment utilisés sont :

- ε-greedy, [62]
- Upper Confidence Bound [1, 42],
- Thompson Sampling [65].

Ces méthodes ont été largement utilisées pour la tarification dynamique, notamment dans des contextes où la demande dépend du prix mais où l'inventaire est ignoré. Ces trois méthodes se distinguent par leur stratégie d'exploration : aléatoire pour ε-greedy, optimiste pour UCB et probabiliste pour Thompson Sampling. Dans le cas de ε-greedy, la méthode choisit aléatoirement d'exploiter les valeurs des actions obtenues en jouant l'action avec la meilleure valeur ou d'explorer de manière aléatoire. Les deux autres méthodes ne différencient pas l'exploration et l'exploitation dans leur algorithme. Dans le cas d'UCB, le choix de l'action à jouer va dépendre du gain moyen de chaque action ainsi que du nombre de fois que chaque action a été jouée. Ce qui permet de rejouer une action qui a été peu jouée dans le passé. Les actions dans Thompson

Sampling sont choisies en optimisant la récompense attendue en tirant aléatoirement pour chaque action le gain attendu selon la distribution observée dans le passé pour cette action. Cela crée ainsi un équilibre exploration/exploitation naturel.

La méthode UCB a été utilisée et améliorée pour convenir à des problématiques de tarification dynamique pour la vente de produits uniques avec des contraintes de ressources dans plusieurs articles [13, 67]. La méthode UCB n'a pas été testée sur les domaines proches de la location de voitures tels que l'aérien ou l'hôtellerie. Cela vient sûrement du fait de son équilibre exploration/exploitation qui n'évolue pas beaucoup au cours du temps et ainsi soit la méthode UCB n'explore pas assez au début, soit elle explore trop une fois que les données sont exploitables. Le problème pourrait être similaire avec ε-greedy, mais il existe plusieurs manières de faire varier ε pour régler ce problème [62, 66].

Une approche pour le secteur du réseau [19] et une approche pour le commerce en ligne [24] ont été proposées avec la méthode Thompson Sampling permettant d'apprendre simultanément la demande et la politique de prix. Le but de ces deux articles est de faire de la tarification dynamique sur un inventaire contraint. D'après ces articles, une quantité de données plutôt importante est nécessaire car la méthode Thompson Sampling a besoin de beaucoup d'étapes d'exploration avant d'avoir une exploitation performante.

L'article [14] compare ces trois méthodes en analysant l'évolution de leur valeur de regret. On peut alors voir qu'UCB a un regret nettement plus élevé que les deux autres dans une situation simple de concurrence. Les deux autres méthodes ont un regret proche, mais ε-greedy a un regret plus faible au début, c'est ensuite la méthode Thompson sampling qui, après beaucoup d'étapes d'apprentissage, finit par avoir un regret plus faible. Dans la location de voitures, le nombre de données est faible et on a besoin que la méthode d'apprentissage apprenne vite, ainsi on a choisi de garder la méthode ε-greedy comme méthode d'apprentissage par valeur.

Le fonctionnement de la méthode $\varepsilon - greedy$ va maintenant être détaillé car elle sera utilisée comme point de comparaison dans le chapitre 5.2.3.3.

Méthode $\varepsilon - greedy$

La méthode $\varepsilon - greedy$ est un algorithme d'apprentissage par renforcement qui cherche à savoir quelle est la meilleure action en fonction de l'environnement. Pour trouver cette meilleure action, l'algorithme a deux stratégies :

- Stratégie d'exploration : l'agent choisit délibérément une action aléatoire. Cette étape permet à l'agent d'explorer de nouvelles possibilités et de découvrir potentiellement de meilleures actions non envisagées.
- Stratégie d'exploitation : l'agent s'appuie sur ses connaissances exis-

tantes en choisissant la meilleure action établie à partir de ses expériences passées. Cela garantit que l'agent se concentre sur la maximisation des récompenses à partir de stratégies déjà éprouvées.

L'algorithme 1 va aléatoirement choisir entre ses deux stratégies avec une probabilité ε de faire de l'exploration.

Algorithme 1 $\varepsilon - greedy$

Entrée: Ensemble d'actions $\mathcal{A}$, gain moyen obtenu $Q(a)$ pour tout $a \in \mathcal{A}$, paramètre $\varepsilon \in [0, 1]$
Sortie: Action sélectionnée a_t

Tirer $u \sim \mathcal{U}(0, 1)$
si $u < \varepsilon$ **alors**
    Choisir une action a_t de manière aléatoire uniforme dans $\mathcal{A}$ ▷ *exploration*
sinon
    Choisir $a_t \in \arg\max_{a \in \mathcal{A}} Q(a)$ ▷ *exploitation*
retourne a_t

Pour avoir un meilleur résultat, la valeur d'ϵ peut varier au cours du temps, en particulier diminuer pour réduire la quantité d'exploration par rapport à celle d'exploitation.

2.3.2. Méthodes d'apprentissage par renforcement par politique

Les méthodes d'apprentissage par renforcement par politique consistent à apprendre directement une politique de décision, c'est-à-dire une règle permettant d'associer à chaque situation une distribution de probabilité sur les actions possibles, sans passer nécessairement par l'estimation explicite d'une fonction de valeur [63]. Dans le cadre de la tarification dynamique, cela revient à apprendre directement une politique de prix, éventuellement paramétrée, afin de maximiser le revenu cumulé ou une autre mesure de performance économique. Contrairement aux méthodes par valeurs, qui évaluent la qualité des actions avant d'en déduire une politique, les méthodes par politique ajustent directement les probabilités de sélection des actions en fonction des récompenses observées.

Ces approches occupent une place importante dans la littérature récente sur la tarification dynamique, en particulier lorsque l'espace d'actions est continu ou lorsque les demandes, la concurrence et les contraintes de capacité rendent difficile l'utilisation de méthodes plus simples. Les méthodes de gradient de politique apprennent alors une politique paramétrée en maximisant directement l'espérance de gain. Parmi les approches les plus connues figurent les méthodes *Actor-Critic*, qui combinent une politique, appelée acteur, et une estimation

de la qualité des décisions, appelée critique [41]. L'acteur détermine l'action à jouer, tandis que le critique évalue l'action choisie afin de guider la mise à jour de la politique. Cette structure permet généralement d'obtenir un apprentissage plus stable que les méthodes par gradient de politique, notamment lorsque les décisions doivent être ajustées dans des environnements complexes [41].

Plusieurs travaux récents mobilisent ainsi les méthodes *Actor-Critic* pour des problèmes de tarification dynamique. Dans l'industrie manufacturière, l'article [27] propose une stratégie de tarification dynamique conjointe du produit et du délai de livraison en utilisant un agent *Actor-Critic* couplé à un modèle de simulation de production ; l'objectif est d'ajuster en continu le prix en fonction de la capacité de production et des préférences des clients, ce qui permet de construire une fonction de prix continue adaptée à chaque demande. Dans le domaine de l'énergie, l'approche *Actor-Critic* a également été utilisée pour la réponse à la demande, en apprenant dynamiquement des prix incitatifs afin de mieux équilibrer consommation et disponibilité du réseau électrique [34]. Enfin, dans des contextes de marchés non stationnaires, l'algorithme ACIDP (*Actor-Critic Information-Directed Pricing*) combine une logique *Actor-Critic* avec une stratégie d'exploration dirigée par l'information afin d'apprendre la courbe de demande et de réagir aux changements de régime du marché ; cette approche a été proposée précisément pour traiter le problème du pricing en environnement instable [46].

Parmi les prolongements modernes des méthodes par politique, PPO (*Proximal Policy Optimization*) occupe une place importante. PPO cherche à améliorer la stabilité de l'apprentissage en limitant l'ampleur des mises à jour de la politique entre deux itérations successives [60]. L'idée est d'éviter qu'une nouvelle politique s'écarte trop fortement de la précédente, ce qui réduirait le risque de comportements instables ou de dégradation brutale des performances. Ce principe est particulièrement intéressant en tarification dynamique, où des variations trop brutales des prix peuvent économiquement être inefficaces ou difficilement interprétables. Dans la littérature, PPO a par exemple été appliqué à la tarification dynamique avec avis en ligne : Wu, Bi et Liu construisent deux environnements simulés intégrant les perceptions de qualité et de valeur des consommateurs, puis apprennent une politique de prix par PPO afin de maximiser le revenu ; les auteurs montrent que l'algorithme permet d'intégrer des variables marketing complexes et qu'il surpasse, dans leur étude, une approche *Advantage Actor-Critic* [72]. PPO a également été mobilisé pour l'optimisation de stratégies de prix sur des plateformes de VTC, où il sert à ajuster les prix en temps réel dans un environnement fortement dynamique et concurrentiel [74].

Ces méthodes par politique présentent donc des avantages réels pour la tarification dynamique : elles permettent de traiter des actions continues, d'intégrer des variables contextuelles riches et de représenter des dépendances non

linéaires difficiles à capturer avec des approches plus simples [27]. Elles présentent toutefois plusieurs limites dans le cadre spécifique étudié ici. D'une part, une grande partie de ces méthodes, en particulier PPO et les variantes modernes d'*Actor-Critic*, sont principalement utilisées dans des cadres avec états, ou avec des réseaux de neurones, afin de représenter des politiques complexes [71]. D'autre part, leur mise en œuvre nécessite souvent un volume important de données, un réglage délicat des hyperparamètres, ainsi qu'un environnement suffisamment stable pour que les mises à jour successives de la politique conservent du sens. Or, dans le cas de la location de voitures étudiée dans cette thèse, la non-stationnarité concurrentielle et la bidimensionnalité de la ressource rendent peu pertinente la construction d'un espace d'états exploitable, ce qui conduit à privilégier des approches sans état relevant du cadre des bandits manchots.

Dans cette perspective, les méthodes par politique considérées dans la suite ne reposent pas sur des réseaux de neurones, mais sur des mises à jour probabilistes simples directement appliquées à l'ensemble fini des prix possibles. C'est le cas de Linear Reward [51], qui apprend un vecteur de probabilités sur les actions, et d'EXP3 [1], qui ajuste les probabilités de sélection par pondération exponentielle en fonction des récompenses observées. Ces approches conservent l'idée centrale des méthodes par politique, à savoir apprendre directement comment jouer les actions, tout en restant compatibles avec un problème sans état, non stationnaire et contraint par une faible quantité de données d'apprentissage par situation. Dans la suite de ce manuscrit, ce sont donc ces deux méthodes par politique non profonde et sans état qui seront retenues pour traiter le problème de tarification dynamique concurrentielle en location de voitures.

Le fonctionnement de la méthode EXP3 va maintenant être détaillé car elle sera utilisée comme point de comparaison dans le chapitre 5.2.3.3. La méthode Linear Reward sera détaillée et modifiée pour s'adapter au mieux au contexte de la location de voitures dans le chapitre 5.

Méthode EXP3

L'algorithme EXP3 [1] est une méthode d'apprentissage par renforcement qui combine exploration et exploitation via une pondération exponentielle des actions. L'algorithme maintient, pour chaque action, un poids positif reflétant sa performance passée. Ces poids sont transformés en probabilités de sélection, pour favoriser les actions ayant obtenu de bonnes récompenses (exploitation) et garantir une probabilité minimale pour toutes les actions (exploration).

Algorithme 2 EXP3

Entrée: Ensemble d'actions $\mathcal{A} = \{1, \dots, K\}$, horizon T, paramètre $\gamma \in [0, 1]$

Sortie: Séquence d'actions $(a_t)_{t=1}^{T}$

Initialiser les poids : $w_i \leftarrow 1$ pour tout $i \in \{1, \dots, K\}$
pour t **allant de** 1 **à** T **faire**
    Calculer $W \leftarrow \sum_{j=1}^{K} w_j$
    pour i **allant de** 1 **à** K **faire**
        $p_i \leftarrow (1-\gamma)\,\dfrac{w_i}{W} + \dfrac{\gamma}{K}$ ▷ *Probabilité*
    Tirer l'action $a_t \sim (p_1, \dots, p_K)$ ▷ *Tirage*
    Observer la récompense $r_t \in [0, 1]$ associée à a_t ▷ *Recompense*
    $\hat{r}_{t,a_t} \leftarrow \dfrac{r_t}{p_{a_t}}$ et $\hat{r}_{t,i} \leftarrow 0$ pour $i \neq a_t$
    pour i **allant de** 1 **à** K **faire**
        $w_i \leftarrow w_i \cdot \exp\Big(\dfrac{\gamma}{K}\,\hat{r}_{t,i}\Big)$ ▷ *mise à jour*

2.4. Conclusion

Ce chapitre a présenté le concept de tarification dynamique dans des secteurs caractérisés par une capacité périssable, une demande incertaine et une forte pression concurrentielle. Il a d'abord rappelé les fondements du *Revenue Management*, qui constitue l'approche historique de référence pour la gestion conjointe des prix et de la ressource, en s'appuyant sur des modèles probabilistes d'allocation de capacité et sur des règles de décision telles que les méthodes de type EMSR. Ces approches mettent au premier plan la notion de valeur marginale de la ressource et montrent comment une politique tarifaire peut être construite à partir de la seule contrainte de capacité, en l'absence de concurrence explicite.

Dans un second temps, le chapitre a mis en évidence les limites de ces approches purement basées sur la ressource dans des marchés où les consommateurs comparent les offres de plusieurs fournisseurs. Les modèles de tarification fondés sur la concurrence, inspirés de la théorie des jeux et des modèles de type Bertrand, analysent le prix comme le résultat d'une interaction stratégique entre entreprises. Ils montrent que la dynamique concurrentielle peut conduire à des phénomènes tels que les guerres de prix, les trajectoires tarifaires non monotones ou encore des formes de collusion tacite, en particulier lorsque les décisions sont répétées dans le temps. Ces résultats soulignent qu'ignorer la dimension concurrentielle revient souvent à sous-estimer la complexité des mécanismes de formation des prix.

Dans la seconde partie du chapitre, l'accent a été mis sur les méthodes d'apprentissage par renforcement comme cadre pertinent pour la tarification

dynamique dans un environnement incertain et potentiellement non stationnaire. Le chapitre a d'abord rappelé le cadre général du RL, dans lequel un agent interagit avec un environnement, choisit des prix en réponse aux observations et reçoit une récompense liée au revenu réalisé, en devant gérer le compromis entre exploration de nouveaux niveaux de prix et exploitation des connaissances acquises. Deux grandes familles de méthodes ont ensuite été distinguées : les approches par valeur, qui estiment la qualité des actions (comme ε-greedy, UCB ou Thompson Sampling), et les approches par politique, qui apprennent directement une distribution de probabilités sur les prix possibles. Les méthodes avec états, issues des MDP et souvent implémentées via des algorithmes comme Q-learning ou des variantes *Actor-Critic* modernes, sont peu adaptées au contexte spécifique de la location de voitures, en raison de la bidimensionnalité de la ressource et du volume de données requis. Le chapitre a donc motivé le recours à des méthodes sans états, relevant du cadre des bandits manchots, mieux adaptées à un environnement concurrentiel non stationnaire avec peu de données par configuration. Dans cette perspective, des méthodes par valeur (comme ε-greedy) et des méthodes par politique simples (Linear Reward, EXP3) ont été retenues comme briques de base, car elles permettent d'apprendre progressivement une politique de prix à partir des récompenses observées tout en restant compatibles avec les contraintes pratiques de la location de voitures.

Chapitre 3
Modèle de jeu répété et Propriétés

Sommaire

3.1 Introduction 21
3.2 Contexte 22
3.3 Modélisation des acteurs principaux 22
3.3.1 Agent 22
3.3.2 Demande 23
3.4 Modèle de jeu répété 23
3.4.1 Jeu unitaire 24
3.4.2 Jeu répété 24
3.5 Équilibre de Nash du jeu unitaire et du jeu répété . . 26
3.5.1 Équilibre de Nash pour le jeu unitaire 26
3.5.2 Équilibre de Nash pour le jeu répété 29
3.5.3 Équilibre online pour le jeu répété 31
3.5.4 Équilibre dans un autre modèle de jeu répété 36
3.6 Conclusion 38
3.7 Liste des notations 40
3.7.1 Notations pour l'agent A_α (Section 3.3.1) 40
3.7.2 Notations de la demande de location d (Section 3.3.2) . 40
3.7.3 Notations du jeu unitaire (Section 3.4.1) 40
3.7.4 Notations du jeu répété (Section 3.4.2) 41

3.1. Introduction

Le but de cette thèse est d'apporter un algorithme pour faire de la tarification dynamique dans le contexte de la location de voitures. Pour modéliser ce contexte, plusieurs éléments doivent être définis, tels que les loueurs de voitures, ainsi que les clients qui formulent des demandes de location, ce qui sera abordé dans ce chapitre. Pour cela, la Section 3.2 définira quelques paramètres généraux, puis la Section 3.3 expliquera la modélisation des agents et des demandes. Dans la Section 3.4, nous définirons le modèle de jeu répété qui décrira notre problème dans son ensemble. Pour finir, la Section 3.5 présente l'étude théorique menée sur les équilibres de ce jeu.

3.2. Contexte

Considérons une entreprise de location de voitures souhaitant choisir dynamiquement le prix optimal pour chaque nouvelle demande de location reçue, dans un contexte où plusieurs entreprises de location de voitures sont en concurrence pour satisfaire cette même demande.

Dans un modèle en temps discret, où chaque étape correspond à l'arrivée d'une demande, on considère un ensemble d'entreprises de location $\mathcal{A}$.

Une demande client est définie comme la location d'une voiture pour une séquence donnée de périodes consécutives, où chaque période peut être une journée ou une demi-journée. Les demandes sont reçues à une date donnée et au fur et à mesure, ainsi, l'ensemble des demandes $\mathcal{D}$ est totalement ordonné.

À chaque nouvelle demande, les entreprises de location ayant la capacité nécessaire pour répondre à la demande (c'est-à-dire disposant d'une voiture libre pour la suite des périodes concernées) proposent un prix. Le client choisit ensuite l'une des offres selon des critères spécifiques (par exemple ici, le prix minimal). Ensuite, l'entreprise de location choisie met à jour sa capacité disponible pour les jours affectés par la demande.

Ainsi, chaque demande engendre une guerre des prix sous contrainte de ressources entre les entreprises pouvant la satisfaire, modélisée comme un jeu simultané entre les entreprises de location disponibles, nommées agents de location dans la suite du document. En considérant la séquence des demandes, l'ensemble du processus est modélisé comme un jeu répété de guerre des prix sous contraintes de ressources en temps discret, où :

- Chaque étape du jeu répété correspond à une nouvelle demande.
- Chaque étape est un jeu simultané entre les agents disponibles.
- L'objectif de chaque agent de location est de maximiser son profit cumulé sur l'ensemble des étapes du jeu.

3.3. Modélisation des acteurs principaux

Dans cette section, le but sera de modéliser les agents ainsi que les demandes de location. Pour plus de simplicité, les clients voulant louer une voiture seront restreints à leur demande de location.

3.3.1. Agent

Chaque agent de location $A_\alpha \in \mathcal{A}$ avec $1 \leq \alpha \leq N$ est caractérisé par :

- $Pr_\alpha = \{pr_\alpha[1],\ pr_\alpha[2],\ \ldots,\ pr_\alpha[M_\alpha]\}$, qui est un ensemble de prix possibles par période, présentés dans l'ordre croissant et susceptibles de varier d'un agent à l'autre.

- $\mathcal{C}apmax_\alpha$ qui est le nombre total de voitures à louer, dont les voitures sont toutes identiques. Ce nombre ne varie pas selon la période.
- $\mathcal{U}til_\alpha(p)$ qui correspond au nombre de voitures déjà louées pour la période p.
- $\mathcal{C}ap_\alpha(p)$ qui est le nombre de voitures disponibles pour la période p avec $\mathcal{C}apmax_\alpha = \mathcal{U}til_\alpha(p) + \mathcal{C}ap_\alpha(p)$. Chaque agent a donc une liste de valeurs de capacité pour déterminer le nombre de véhicules qu'il doit encore louer pour chaque période.

3.3.2. Demande

Une demande de location d est caractérisée par :

- $recep_d$ la période de réception de la demande de location.
- $debut_d$ la période de début de location $debut_d \geq recep_d$.
- fin_d la période de fin de location avec $fin_d \geq debut_d$.
- $delaiResa_d$ le délai de réservation qui correspond au temps entre la réservation et le début de la location $delaiResa_d = debut_d - recep_d$.
- $duree_d$ la durée de location, avec $duree_d = fin_d - debut_d + 1$.
- $Pmax$ le prix maximal que le client peut accepter pour cette demande.
- $ordre_d$ Un ordre de préférence sur les agents de location, car chaque client a un critère de choix lorsque le prix est identique pour chaque agent. Cela permet d'avoir un choix déterministe pour la suite. Cette fonction est décrite comme suit :

$$\begin{array}{rccl} ordre_d & : & \mathcal{A} & \rightarrow & \{1, \ldots, N\}. \\ & & A_\alpha & \mapsto & ordre_d(A_\alpha) \end{array}$$

Pour la suite, nous allons considérer une séquence de demandes $\mathcal{D} = d_1, \ldots, d_N$, les demandes étant reçues par les agents l'une après l'autre dans l'ordre déterminé en fonction de la période de réservation $recep_d$ pendant laquelle chaque demande est reçue, tel que $\forall i, recep_{d_i} \leq recep_{d_{i+1}}$.

3.4. Modèle de jeu répété

Dans cette section, le but est de définir le jeu dans son ensemble. Pour cela, cette section commencera par définir le jeu unitaire pour une seule demande, puis le jeu global pour un ensemble de demandes $\mathcal{D}$, qui est une répétition du jeu unitaire.

3.4.1. Jeu unitaire

À l'arrivée de chaque demande d, un jeu est initié entre les agents de location afin de déterminer le prix qu'ils proposeront pour cette demande.

Ce jeu unitaire est caractérisé par un triplé $(\mathcal{A},\ \mathcal{C}ap,\ d)$ où :

- $\mathcal{A}$ est l'ensemble des agents de location, c'est-à-dire les joueurs de ce jeu.
- $\mathcal{C}ap_\alpha(p)$ la capacité disponible pour chaque agent A_α et pour chaque période de location p.
- d désigne la demande de location considérée.

Un joueur A_α choisit de ne pas proposer de prix lorsqu'il n'a pas de capacité c'est-à-dire $\varnothing$ si :

$$\exists\ p \in [debut_d, fin_d] \text{ telle que } \mathcal{C}ap_\alpha(p) = 0.$$

C'est-à-dire s'il ne dispose pas de capacité suffisante pour au moins une des périodes requises par la demande. Dans ce jeu, les autres joueurs $A_\alpha \in \mathcal{A}$ choisissent simultanément une action, qui consiste ici à proposer un prix $s_\alpha \in Pr_\alpha$ pour la demande d. Le profil de stratégie associé à la demande d est noté $\pi = (s_1, s_2, \ldots, s_n)$ avec $\forall \alpha, s_\alpha \in Pr_\alpha \cup \{\varnothing\}$.

Le gagnant de ce jeu unitaire sera l'agent $A_{gagnant}$ tel que $ordre(A_{gagnant}) = min(O_{min})$ avec

$$O_{min} = \{ordre_d(A_\alpha) \text{ tel que } A_\alpha \in \mathcal{A}min\}$$

et

$$A_{gagnant} \in \mathcal{A}_{min} = \{A_\alpha \in \mathcal{A} \mid s_\alpha = \min(\pi) \text{ et } s_\alpha \leq Pmax_d\}\,.$$

C'est-à-dire, l'agent $A_{gagnant}$ est l'agent dont le prix minimal est inférieur au prix maximum accepté par la demande d et qui est le mieux placé dans l'$ordre_d$. Si $\mathcal{A}_{min}$ est vide, alors il n'y a pas de gagnant pour cette demande.

La fonction de gain pour chaque joueur est définie comme suit :

$$g_\alpha(\pi, d) = \begin{cases} 0, & \text{si } A_\alpha \neq A_{gagnant} \\ duree_d \cdot \min(\pi), & \text{sinon } (A_\alpha = A_{gagnant}). \end{cases}$$

L'étude de l'existence d'un équilibre de Nash dans ce jeu unitaire sera développée dans la Section 3.5.1.

3.4.2. Jeu répété

Nous allons maintenant définir le jeu répété qui consiste pour chaque joueur à choisir un prix à proposer pour un ensemble de demandes $\mathcal{D}$ pour avoir le meilleur bénéfice final tout en étant en concurrence avec les autres joueurs.

Le jeu répété est caractérisé par un triplé $\langle \mathcal{A},\ \mathcal{C}ap,\ \mathcal{D} \rangle$ où :

- $\mathcal{A}$ est l'ensemble des agents de location, c'est-à-dire les joueurs de ce jeu.
- $\mathcal{C}ap$ le tableau des capacités disponibles pour chaque agent et pour chaque période de location au début du jeu répété.
- $\mathcal{D} = (d^1, d^2, \ldots, d^T)$ désigne la suite des demandes de location qui vont être reçues. Les demandes sont ordonnées par ordre chronologique de dates de réservation $recep_d$.

Ainsi, dans ce jeu, chaque agent A_α va devoir choisir un prix pour chaque demande d^t si sa capacité restante à l'arrivée de la demande $recep_{d^t}$ notée $\mathcal{C}ap^t_\alpha$ est non nulle entre $debut_{d^t}$ et fin_{d^t}. Dans ce jeu répété, la stratégie de chaque agent A_α va être une séquence de stratégie

$$S_\alpha = \{s^1_\alpha, s^2_\alpha, \ldots, s^T_\alpha\} \in \prod_{i=1}^{T} Pr_\alpha$$

où chaque stratégie est soit un prix tel que $s^t_\alpha \in Pr_\alpha$ qui est proposé à la demande d^t soit $\varnothing$ lorsque l'agent n'a pas de capacité. Le profil de stratégie du jeu répété est donc $\Pi = \{S_1, S_2, \ldots, S_N\}$.

Ce jeu répété est une succession de jeux unitaires ($\mathcal{A}$, $\mathcal{C}ap^t$, d^t) pour chaque demande d^t entre l'ensemble des agents disponibles pour la demande d^t noté $\mathcal{A}^t$. Les autres agents joueront $\varnothing$. Chaque agent $A_\alpha \in \mathcal{A}$ a pour capacité restante $\mathcal{C}ap^t_\alpha$ qui est mise à jour à chaque jeu. Dans ce cadre, les notations introduites précédemment pour le jeu unitaire restent valables pour le jeu unitaire ($\mathcal{A}$, $\mathcal{C}ap^t$, d^t), on leur ajoute simplement un t pour indiquer l'étape de la séquence.

Pour chaque demande d^t avec $t \in \{1, \ldots, T\}$, un seul agent A_α est gagnant du jeu unitaire pour d^t. Il doit alors mettre à jour sa capacité disponible. On note $\mathcal{C}ap^{t+1}_\alpha(p)$ la capacité mise à jour après l'étape t pour toute période p pour l'agent A_α.

Le profil de jeu répété peut aussi être défini comme une séquence de profils unitaires : $\Pi = (\pi^1, \pi^2, \ldots, \pi^T)$, où chaque profil $\pi^t = (s^t_1, \ldots, s^t_N)$ est le profil de stratégies du jeu unitaire correspondant à la demande d^t.

Le gain cumulé d'un agent A_α dans le jeu répété est défini par :

$$G_\alpha(\Pi, \mathcal{C}ap, \mathcal{D}) = \sum_{t=1}^{T} g_\alpha(\pi^t, d^t).$$

Pour chaque jeu unitaire correspondant à l'arrivée de la demande d^t, chaque agent de location A_α connaît :

- L'ensemble des demandes reçues jusqu'à l'étape t : $\mathcal{D}^t = \{d^1, \ldots, d^t\}$.
- Les profils de stratégie des jeux unitaires joués par ses concurrents et lui-même dans le passé : $\pi^1, \pi^2, \ldots, \pi^{t-1}$.

- Le sous-ensemble des demandes pour lesquelles l'agent a été sélectionné : $demC^{t-1}(A_\alpha) \subseteq \mathcal{D}^{t-1}$.
- Sa capacité restante mise à jour pour chaque période p : $\mathcal{C}ap^t_\alpha(p)$ calculée à partir de $\mathcal{C}ap^{t-1}_\alpha(p)$ de la manière suivante :

$$\forall p,\ \mathcal{C}ap^t_\alpha(p) = \begin{cases} \mathcal{C}ap^{t-1}_\alpha(p) - 1, & \text{si } A_\alpha \text{ gagne à l'étape } t \\ & \text{et } debut_{d^t} \leq p \leq fin_{d^t}, \\ \mathcal{C}ap^{t-1}_\alpha(p), & \text{sinon.} \end{cases}$$

3.5. Équilibre de Nash du jeu unitaire et du jeu répété

Dans ce modèle de jeu répété, le but des agents est de maximiser la somme des bénéfices pour chaque demande reçue. Dans ce contexte, il est pertinent d'étudier les situations d'équilibre. Pour cela, il existe des notions d'équilibre pour les jeux simultanés [52] ainsi que pour les jeux répétés étudiés par la suite. Un nouveau type d'équilibre pour les jeux répétés sera défini afin d'être plus conforme au modèle [3].

3.5.1. Équilibre de Nash pour le jeu unitaire

Cette sous-section définira ce qu'est un équilibre de Nash dans un jeu simultané à un seul tour, puis montrera que, dans le jeu unitaire de notre problème, il existe forcément un tel équilibre.

Définition I (Équilibre de Nash pour le jeu unitaire [52])**.** Un équilibre de Nash est un profil de stratégie π tel qu'aucun joueur ne peut obtenir un gain supplémentaire en changeant unilatéralement de stratégie. Autrement dit, π est un équilibre de Nash si et seulement si :

$$\forall \alpha \in \{1, \ldots, N\},\ \forall s_\alpha \in Pr_\alpha,\ g_\alpha(\pi, d) \geq g_\alpha((\pi_{-\alpha}, s_\alpha), d). \tag{3.1}$$

Notons $(\pi_{-\alpha}, s_\alpha)$ le profil π dans lequel seulement la stratégie du joueur A_α a été changée par s_α.

Proposition I (Équilibre de Nash pour le jeu unitaire)**.** Tout jeu unitaire défini par $(\mathcal{A},\ \mathcal{C}ap,\ d)$ admet un équilibre de Nash.

Preuve de la Proposition I. On considère une instance du jeu unitaire définie par $(\mathcal{A},\ \mathcal{C}ap,\ d)$, et sans perte de généralité, $\mathcal{A}$ contient uniquement les agents qui peuvent recevoir la demande d, c'est-à-dire des agents sans problème de capacité pour d et qui peuvent faire des prix inférieurs à $Pmax_d$. Si $\mathcal{A}$ contient un unique agent, alors celui-ci jouera son prix maximum, ce sera ainsi une situation d'équilibre. Si $\mathcal{A}$ ne contient pas un unique agent, on note d'abord :

$$ValminPr = \min_{\alpha \in \{1,\ldots,N\}}, (pr_\alpha[1] \in Pr_\alpha)$$

et

$$MinPr = \{A_\alpha,\ pr_\alpha[1] = ValminPr\}.$$

Nous allons maintenant traiter deux cas en fonction de la valeur de $|MinPr|$.

Cas 1 : $|MinPr| \geq 2$ c'est-à-dire le cas où plusieurs joueurs peuvent faire le prix $ValminPr$.

Notons $\pi^* = \{s_1, s_2, \ldots, s_T\}$ avec $s_\alpha = pr_\alpha[1]\ \forall A_\alpha \in \mathcal{A}$, on a alors que $A_{gagnant} \in MinPr$ et $ordre_d(A_{gagnant})$ est le plus petit des $ordre_d(A_\alpha)$ pour tous les agents A_α de $MinPr$.

Chaque joueur qui n'est pas dans $MinPr$ ne pourra pas changer de gain car il ne peut pas jouer un prix plus petit que $ValminPr$. Pour les joueurs de $MinPr$ qui ont perdu, ils jouent déjà le prix $ValminPr$ mais leur classement dans l'ordre de préférence n'est pas suffisant pour obtenir la demande, donc ils ne peuvent pas changer leur gain qui est de 0.

Le gain de $A_{gagnant}$ est le suivant :

$$g_\alpha((\pi_{-\alpha}, s_\alpha), d) = \begin{cases} 0, & \text{si } s_\alpha > ValminPr, \\ ValminPr \times duree_d, & \text{sinon } (s_\alpha = ValminPr). \end{cases}$$

Il n'a donc pas intérêt à changer de stratégie. π^* est donc bien un équilibre de Nash au sens du jeu unitaire.

Cas 2 : $|MinPr| = 1$ c'est-à-dire le cas où seulement un joueur peut faire le prix $ValminPr$.

Notons A_{MP} le seul agent de $MinPr$,

$$ValminPr2 = \min_{\forall A_\alpha \neq A_{MP}} (pr_\alpha \in Pr_\alpha)$$

et

$$MinPr2 = \{A_\alpha,\ \exists k,\ pr_\alpha[k] = ValminPr2\}.$$

Cas 2.1 : $ValminPr2 \notin Pr_{MP}$ avec Pr_{MP} l'ensemble des prix possibles de l'agent A_{MP}.

Dans ce cas, $\pi^* = \{s_1, s_2, \ldots, s_T\}$ avec $s_\alpha = pr_\alpha[1]$, si $A_\alpha \neq A_{MP}$ et $s_{MP} = pr_{MP}[k]$ le prix le plus grand de Pr_{MP} tel que $pr_{MP}[k] < ValminPr2$.

D'abord, A_{MP} aurait le gain suivant s'il changeait unilatéralement son prix :

$$g_\alpha((\pi_{-\alpha}, s_\alpha), d) = \begin{cases} 0, & \text{si } s_\alpha > pr_{MP}[k], \\ pr_{MP}[k] \times duree_d, & \text{si } s_\alpha = pr_{MP}[k], \\ s_\alpha \times duree_d, & \text{sinon } (s_\alpha < pr_{MP}[k]). \end{cases}$$

Ainsi A_{MP} n'a pas intérêt à changer unilatéralement de stratégie.

Pour les autres joueurs, leur gain est de 0 et ils ne peuvent pas faire un prix plus bas car ils proposent déjà leur prix minimum ; c'est donc bien une situation d'équilibre.

Cas 2.2 : $ValminPr2 \in Pr_{MP}$. Notons

$$O_{MinPr2} = \{ordre_d(A_\alpha) \text{ tel que } A_\alpha \in MinPr2\}.$$

- **Si** $ordre_d(A_{MP}) = min(O_{MinPr2})$ Alors dans ce cas $\pi^* = \{s_1, s_2, \ldots, s_T\}$ avec $s_\alpha = pr_\alpha[1]$, si $A_\alpha \neq A_{MP}$ et $s_{MP} = ValminPr2$.

  A_{MP} aurait le gain suivant s'il changeait unilatéralement son prix :

$$g_\alpha((\pi_{-\alpha}, s_\alpha), d) = \begin{cases} 0, & \text{si } s_\alpha > ValminPr2, \\ ValminPr2 \times duree_d, & \text{si } s_\alpha = ValminPr2, \\ s_\alpha \times duree_d, & \text{sinon } (s_\alpha < ValminPr2). \end{cases}$$

  Ainsi A_{MP} n'a pas intérêt à changer unilatéralement de stratégie.

  Chaque joueur qui n'est pas dans $MinPr2$ ne pourra pas changer de gain car il ne peut pas jouer un prix plus petit que $ValminPr2$. Pour les joueurs de $MinPr2$ qui ont perdu, ils jouent déjà le prix $ValminPr2$ mais leur classement dans l'ordre de préférence n'est pas suffisant pour obtenir la demande, donc ils ne peuvent pas changer leur gain qui est de 0.

- **Si** $ordre_d(A_{MP}) \neq min(O_{MinPr2})$ Alors, dans ce cas, $\pi^* = \{s_1, s_2, \ldots, s_T\}$ avec $s_\alpha = pr_\alpha[1]$, si $A_\alpha \neq A_{MP}$ et $s_{MP} = pr_{MP}[k]$ le prix le plus grand de Pr_{MP} tel que $pr_{MP}[k] < ValminPr2$.

  A_{MP} aurait le gain suivant en fonction du prix joué :

$$g_\alpha((\pi_{-\alpha}, s_\alpha), d) = \begin{cases} 0, & \text{si } s_\alpha > p^k_{MP}, \\ p^k_{MP} \times duree_d, & \text{sinon } (s_\alpha = p^k_{MP}). \end{cases}$$

  Ainsi A_{MP} n'a pas intérêt à changer unilatéralement de stratégie.

  Pour les autres joueurs, leur gain est nul et ils ne peuvent pas proposer un prix inférieur à A_{MP} ; c'est donc dans une situation d'équilibre.

L'existence d'un équilibre a été démontrée pour l'ensemble des cas étudiés. □

Toutefois, cette existence ne s'accompagne pas d'une unicité systématique. Le cas à deux joueurs constitue une configuration particulière pour laquelle l'équilibre est unique. Dès que le nombre de joueurs est supérieur ou égal à trois,

il apparaît que certains joueurs non gagnants peuvent changer de prix tout en préservant la condition d'équilibre, ce qui implique l'existence de plusieurs équilibres distincts.

Dans cette preuve, il a été constaté que les joueurs ne choisissent pas forcément leur prix minimum lorsqu'ils sont en situation d'équilibre. Ce constat amène à considérer que des techniques d'apprentissage seront nécessaires pour que les joueurs puissent maximiser leur bénéfice total.

3.5.2. Équilibre de Nash pour le jeu répété

Définition II (Équilibre de Nash pour les jeux répétés [3])**.** Un profil de jeu répété Π est un équilibre de Nash au sens des jeux répétés, si et seulement si,

$$\forall \alpha \in [1, N], \forall S_\alpha,\ G_\alpha(\Pi, \mathcal{C}ap, \mathcal{D}) \geq G_\alpha((\Pi_{-\alpha}, S_\alpha), \mathcal{C}ap, \mathcal{D})$$

avec $(\Pi_{-\alpha}, S_\alpha)$, le profil du jeu répété correspondant à la situation d'équilibre en changeant les stratégies du joueur A_α par S_α.

Remarque II.

- Cette définition correspond à un équilibre de Nash pour le jeu répété qui consiste pour chaque agent A_α à choisir l'ensemble de ses stratégies S_α au préalable.
- Un profil de jeu répété Π qui est un équilibre pour chaque jeu unitaire n'est pas forcément un équilibre de Nash pour le jeu répété. Exemple pour deux joueurs avec 2 de capacité, 5 demandes pour le même jour et trois prix possibles : 1, 2 et 3. Le tableau suivant présente un équilibre pour chaque jeu unitaire, mais il ne s'agit pas d'un équilibre de Nash pour la partie répétée.

	d_1	d_2	d_3	d_4	d_5
A_1	1	1	1	$\varnothing$	$\varnothing$
A_2	1	1	1	3	$\varnothing$

La définition d'un équilibre de Nash n'est pas réellement pertinente dans notre cas. En effet, les agents prennent des décisions au fil du temps en tenant compte de ce qui s'est passé lors des étapes précédentes, et en particulier si les clients associés aux demandes les ont choisis et qu'ils ont donc moins de capacités. Pour considérer cela, il faudrait un autre type d'équilibre qui tienne compte de l'aspect *online* ([10]) du jeu répété, qui n'est pas simplement une suite de jeux unitaires indépendants. Pour cela, il existe l'équilibre parfait de sous-jeu qui est souvent utilisé dans des contextes similaires. Pour le décrire, il est nécessaire de définir d'abord ce qu'est un sous-jeu.

Définition III (Sous-profil)**.** Étant donné Π le profil du jeu répété défini par $\langle \mathcal{A},\ \mathcal{C}ap,\ \mathcal{D}\rangle$ et $h \leq T$, le sous-profil de taille h du jeu répété Π^h est le profil du jeu $\langle \mathcal{A},\ \mathcal{C}ap^h,\ \mathcal{D}^h = \{d^h, \ldots, d^T\}\rangle$, nommé sous-jeu. Ainsi, le sous-profil de taille h est $\Pi^h = (\pi^h, \pi^{h+1}, \ldots, \pi^T)$.

Définition IV (Équilibre parfait de sous-jeux [21])**.** Un profil de jeu répété Π est un équilibre parfait de sous-jeu, si et seulement si, pour tout $1 \leq h \leq T$, le profil Π^h du sous-jeu $\langle \mathcal{A},\ \mathcal{C}ap^h,\ \mathcal{D}^h\rangle$ est un équilibre de Nash au sens des jeux répétés.

Proposition III. Pour tout jeu répété $\langle \mathcal{A},\ \mathcal{C}ap,\ \mathcal{D}\rangle$, un profil Π est un équilibre de Nash si et seulement si Π est un équilibre parfait de sous-jeux

Preuve de la Proposition III. $\Leftarrow$ Par définition, un équilibre parfait de sous-jeu est forcément un équilibre de Nash.

$\Rightarrow$ Notons Π un équilibre de Nash :

$$\forall \alpha, \forall S_\alpha,\ G_\alpha(\Pi, \mathcal{C}ap, \mathcal{D}) \geq G_\alpha((\Pi_{-\alpha}, S_\alpha), \mathcal{C}ap, \mathcal{D}),$$

Et en particulier, pour tout $h \leq T$:

$$\forall \alpha, \forall S^h_\alpha,\ G_\alpha((\Pi^{<h}, \Pi^h), \mathcal{C}ap, \mathcal{D}) \geq G_\alpha((\Pi^{<h}, (\Pi^h_{-\alpha}, S^h_\alpha)), \mathcal{C}ap, \mathcal{D})$$

Avec $\Pi^{<h}$ le profil pour les demandes avant l'étape h et Π^h le profil pour les demandes à partir de l'étape h. Cette inégalité équivaut par décomposition à :

$$\begin{aligned}\forall \alpha, \forall S^h_\alpha,\ G_\alpha(\Pi^{<h}, \mathcal{C}ap, \mathcal{D}^{<h}) + G_\alpha(\Pi^h, \mathcal{C}ap^h, \mathcal{D}^h) \geq \\ G_\alpha(\Pi^{<h}, \mathcal{C}ap, \mathcal{D}^{<h}) + G_\alpha((\Pi^h_{-\alpha}, S^h_\alpha), \mathcal{C}ap^h, \mathcal{D}^h)\end{aligned}$$

$$\Leftrightarrow \forall \alpha, \forall S^h_\alpha,\ G_\alpha(\Pi^h, \mathcal{C}ap^h, \mathcal{D}^h) \geq G_\alpha((\Pi^h_{-\alpha}, S^h_\alpha), \mathcal{C}ap^h, \mathcal{D}^h)$$

Donc pour tout $h \leq T$, Π^h est un équilibre de Nash pour le sous-jeu de taille h. Ainsi, un équilibre de Nash dans notre contexte est bien un équilibre parfait de sous-jeu.

□

Ainsi, la définition IV d'un équilibre parfait de sous-jeux n'apporte pas un nouveau type d'équilibre qui serait différent de celui de l'équilibre de Nash. Nous allons définir, dans la suite, une version d'équilibre différente et plus adaptée à notre contexte.

3.5.3. Équilibre online pour le jeu répété

Définition V (Profil stable)**.** Étant donné un profil Π du jeu répété $\langle \mathcal{A},\ \mathcal{C}ap,\ \mathcal{D}\rangle$, le profil π du premier jeu unitaire $(\mathcal{A},\ \mathcal{C}ap,\ d^1)$ est stable si, et seulement si,

$$\forall \alpha \in [1, N], \forall s^1_\alpha,\ G_\alpha(\Pi, \mathcal{C}ap, \mathcal{D}) \geq G_\alpha((\Pi_{-\alpha,1}, s^1_\alpha), \mathcal{C}ap, \mathcal{D}) \tag{3.2}$$

où $(\Pi_{-\alpha,1}, s^1_\alpha)$, le profil du jeu répété correspondant à la situation d'équilibre dans laquelle la stratégie du joueur A_α pour la première demande d^1 a été changée par s^1_α.

Par conséquent, si $|\mathcal{D}| = 1$ alors le profil $\Pi = \pi$ est stable si et seulement si le profil π est un équilibre de Nash au sens des jeux unitaires.

Comme il est difficile d'obtenir un équilibre de Nash et que chaque étape du jeu affecte l'avenir, un nouveau type d'équilibre va être introduit pour prendre en compte l'impact sur l'avenir.

Définition VI (Équilibre Séquentiellement Stable (ESS))**.** Un profil de jeu répété Π est un équilibre séquentiellement stable, si et seulement si, pour tout h, le sous-profil $\Pi^{(h+1)}$ est un équilibre séquentiellement stable pour le sous-jeu $\langle \mathcal{A},\ \mathcal{C}ap^{h+1},\ \mathcal{D}^{h+1}\rangle$ et si Π^h est un profil stable pour le sous-jeu $\langle \mathcal{A},\ \mathcal{C}ap^h,\ \mathcal{D}^h\rangle$.

Lemme IV. Si un profil Π est un équilibre de Nash, alors Π est un équilibre séquentiellement stable.

Preuve du Lemme IV. On considère une instance du jeu répété définie par $\langle \mathcal{A},\ \mathcal{C}ap,\ \mathcal{D}\rangle$ avec Π un équilibre de Nash. D'après la Proposition III, Π est un équilibre parfait de sous-jeu. Montrons par récurrence sur h, la taille du sous-jeu, que Π^h est un équilibre séquentiellement stable pour tout h.

- Initialisation : Le sous-jeu de taille 1 $(\mathcal{A},\ \mathcal{C}ap^1, d^T)$.
  Dans ce sous-jeu, Π^1 est un équilibre de Nash car Π est un équilibre parfait de sous-jeu ; ainsi, Π^1 est stable et donc un équilibre séquentiellement stable.
- Hérédité : On suppose que Π^h est un équilibre séquentiellement stable pour h ; on montre que c'est aussi le cas pour Π^{h+1}. Le profil Π^{h+1} est un équilibre de Nash car Π est un équilibre parfait de sous-jeu donc

  $$\forall \alpha, \forall S_\alpha,\ G_\alpha(\Pi^{h+1}, \mathcal{C}ap^{h+1}, \mathcal{D}^{h+1}) \geq G_\alpha((\Pi^{h+1}_{-\alpha}, S^{h+1}_\alpha), \mathcal{C}ap^{h+1}, \mathcal{D}^{h+1}).$$

  où S^{h+1}_α est la stratégie du joueur A_α pour les $h+1$ dernières demandes. On a donc en particulier,

  $$\forall \alpha, \forall s^1_\alpha,\ G_\alpha(\Pi^{h+1}, \mathcal{C}ap^{h+1}, \mathcal{D}^{h+1}) \geq G_\alpha((\Pi^{h+1}_{-\alpha,1}, s^1_\alpha), \mathcal{C}ap^{h+1}, \mathcal{D}^{h+1}) \tag{3.3}$$

où $(\Pi^{h+1}_{-\alpha,1}, s^1_\alpha)$, le profil du jeu répété correspondant à la situation d'équilibre Π^{h+1} dans laquelle la stratégie du joueur A_α pour la première demande d^1 a été changée par s^1_α. Ainsi Π^{h+1} est un équilibre séquentiellement stable.

Ainsi, un équilibre de Nash est bien un équilibre séquentiellement stable. □

On montrera plus tard qu'un équilibre séquentiellement stable n'est pas forcément un équilibre de Nash.

Théorème V (Équilibre séquentiellement stable)**.** Tout jeu répété défini par $\langle \mathcal{A},\ \mathcal{C}ap,\ \mathcal{D}\rangle$ admet un équilibre séquentiellement stable.

Preuve du Théorème V. On considère une instance du jeu répété définie par $\langle \mathcal{A},\ \mathcal{C}ap,\ \mathcal{D}\rangle$. On construit Π, en choisissant à chaque étape d^t, π^t de telle manière qu'il soit un équilibre unitaire pour le jeu unitaire $(\mathcal{A}, Cap^t, d^t)$. Montrons que Π est un équilibre séquentiellement stable, pour cela, il suffit de montrer que tout sous-profil de taille h est un profil stable.

- **Cas : h = T**
  Par construction, $\Pi^T = \pi^T$ est un équilibre de Nash au sens du jeu unitaire, ce qui implique que Π^T est stable.
- **Cas : h < T**
  Supposons qu'il existe un joueur A_g qui gagne le jeu associé à la demande d^h. Si ce n'est pas le cas, ça veut dire que personne n'est disponible pour la demande ou que les agents disponibles ne peuvent pas proposer un prix inférieur à $Pmax_d$, et alors le profil est bien stable.
  On a alors que pour le joueur A_g :

$$\forall s^1_g,\ G_g(\Pi^h, \mathcal{C}ap^h, \mathcal{D}^h) \geq G_g((\Pi^h_{-g,1}, s^1_g), \mathcal{C}ap^h, \mathcal{D}^h)$$

  car $G_g(\Pi^h, \mathcal{C}ap^h, \mathcal{D}^h) = s^{*1}_g + G_g(\Pi^{h+1}, \mathcal{C}ap^{h+1}, \mathcal{D}^{h+1})$ et

$$G_g((\Pi^h_{-g,1}, s^1_g), \mathcal{C}ap^h, \mathcal{D}^h) = G_g(\Pi^{h+1}, \mathcal{C}ap^{h+1}, \mathcal{D}^{h+1}) + \begin{cases} 0, & \text{si } s^1_g > s^{*1}_g, \\ s^{*1}_g, & \text{si } s^1_g = s^{*1}_g, \\ s^1_g, & \text{sinon } (s^1_g < s^{*1}_g). \end{cases}$$

  Avec s^{*1}_g la stratégie du joueur A_g pour la première étape du sous-jeu $\langle \mathcal{A}^h, Cap^h, \mathcal{D}^h\rangle$. Ainsi, le joueur A_g n'a pas intérêt à changer de stratégie.
  Pour les autres joueurs A_α, on a : $G_\alpha(\Pi^h, \mathcal{C}ap^h, \mathcal{D}^h) = G_\alpha(\Pi^{h+1}, \mathcal{C}ap^{h+1}, \mathcal{D}^{h+1})$ car à l'étape d^h le gain est de 0. Vu que π^h le profil du jeu unitaire $(\mathcal{A}^h, Cap^{h-1}, d^h)$ est un équilibre de Nash du jeu unitaire, le joueur A_α ne peut pas améliorer son gain sur ce jeu unitaire et gagner la demande d^h quelle que soit sa stratégie donc :

$$\forall s^1_\alpha,\ G_\alpha(\Pi^h, \mathcal{C}ap^h, \mathcal{D}^h) = G_\alpha((\Pi^h_{-\alpha,1}, s^1_\alpha), \mathcal{C}ap^h, \mathcal{D}^h).$$

  Ainsi, le profil Π^h est un profil stable.

Donc, ce profil Π est un équilibre séquentiellement stable.

□

Dans la suite, l'objectif sera d'étudier les équilibres séquentiellement stables. Le lemme IV a montré que les équilibres de Nash sont séquentiellement stables. Dans la suite, les équilibres de Nash par parties vont être définis comme une généralisation des équilibres de Nash tout en restant des équilibres séquentiellement stables. Cette nouvelle définition permet de décrire la forme d'un plus grand nombre d'équilibres séquentiellement stables. Certains équilibres de Nash par parties apparaissent plus simples à construire que les équilibres de Nash.

Définition VII (Équilibre de Nash par parties)**.** Un profil de jeu répété Π est un équilibre de Nash par parties pour le jeu $\langle \mathcal{A}, \mathcal{C}ap, \mathcal{D} \rangle$ s'il existe une partition de $\mathcal{D} = (\mathcal{D}^{t_0=1,t_1-1}, \mathcal{D}^{t_1,t_2-1}, \mathcal{D}^{t_2,t_3-1}, \ldots, \mathcal{D}^{t_{k-1},t_k=T})$, avec $\mathcal{D}^{t_a,t_{a+1}-1}$ l'ensemble des demandes entre les demandes d_{t_a} et $d_{t_{a+1}-1}$ avec $0 \leq a \leq k-1$ et k le nombre de sous-ensembles, telle que pour chaque jeu $\langle \mathcal{A}, \mathcal{C}ap^{t_a}, \mathcal{D}^{t_a,t_{a+1}-1} \rangle$, $\Pi^{t_a,t_{a+1}-1}$ est un équilibre de Nash au sens des jeux répétés (Définition II).

Théorème VI. Un profil Π qui est un équilibre de Nash par parties est un équilibre séquentiellement stable.

Preuve du Théorème VI. On considère Π un équilibre de Nash par parties pour le jeu répété $\langle \mathcal{A}, \mathcal{C}ap, \mathcal{D} \rangle$ pour la partition $\mathcal{D} = (\mathcal{D}^{t_0=1,t_1-1}, \mathcal{D}^{t_1,t_2-1}, \mathcal{D}^{t_2,t_3-1}, \ldots, \mathcal{D}^{t_{k-1},t_k=T})$. Dans cette preuve, le but est de montrer que Π est stable pour chaque demande, en commençant par prouver la stabilité pour les demandes du dernier ensemble de la partition, puis pour les autres ensembles de la partition.

- Cas 1 : $\mathcal{D}^{t_{k-1},t_k=T}$. Dans ce cas, $\Pi^{t_{k-1}}$ est un équilibre de Nash au sens des jeux répétés pour le jeu $\langle \mathcal{A}, \mathcal{C}ap^{t_{k-1}}, \mathcal{D}^{t_{k-1}} \rangle$ donc pour chaque demande $d^u \in \mathcal{D}^{t_{k-1}}$, Π^u est un profil stable.
- Cas 2 : $\mathcal{D}^{t_a,t_{a+1}-1}$ avec $a < k-1$. Montrons que pour chaque demande $d^u \in \mathcal{D}^{t_a,t_{a+1}-1}$, Π^u est un profil stable.

  Supposons qu'il existe un joueur A_g sélectionné comme gagnant pour le jeu unitaire associé à la demande d^u. Dans le cas contraire, soit aucun agent ne dispose de la capacité nécessaire pour satisfaire la demande, soit les agents disponibles ne proposent aucun prix strictement inférieur à $P_d^{\max}$. Dans ces deux situations, aucun changement de prix n'est susceptible d'améliorer le gain d'un joueur, et le profil de stratégies considéré est donc stable.

  On a alors que pour le joueur A_g :

$$\forall s_g^1,\ G_g(\Pi^u, \mathcal{C}ap^u, \mathcal{D}^u) \geq G_g((\Pi^u_{-g,1}, s_g^1), \mathcal{C}ap^u, \mathcal{D}^u)$$

  car $G_g(\Pi^u, \mathcal{C}ap^u, \mathcal{D}^u) = s_g^{*1} + G_g(\Pi^{u+1}, \mathcal{C}ap^{u+1}, \mathcal{D}^{u+1})$ et

$$G_g((\Pi^u_{-g,1}, s^1_g), \mathcal{C}ap^u, \mathcal{D}^u) = G_g(\Pi^{u+1}, \mathcal{C}ap^{u+1}, \mathcal{D}^{u+1})$$
$$+ \begin{cases} 0, & \text{si } s^1_g > s^{*1}_g, \text{(Nash par parties)} \\ s^{*1}_g, & \text{si } s^1_g = s^{*1}_g, \\ s^1_g, & \text{sinon } (s^1_g < s^{*1}_g). \end{cases}$$

Avec s^{*1}_g la stratégie du joueur A_g pour la première étape du sous-jeu $\langle \mathcal{A}^u, Cap^u, \mathcal{D}^u \rangle$. Ainsi, le joueur A_g n'a pas intérêt à changer de stratégie.

Pour les autres joueurs A_α, on a : $G_\alpha(\Pi^u, \mathcal{C}ap^u, \mathcal{D}^u) = G_\alpha(\Pi^{u+1}, \mathcal{C}ap^{u+1}, \mathcal{D}^{u+1})$ car à l'étape d^u le gain est de 0.

Vu que π^u fait partie d'un équilibre de Nash, le joueur A_α ne peut pas améliorer son gain sur le sous-jeu de taille u donc

$$\forall s^1_\alpha,\ G_\alpha(\Pi^u, \mathcal{C}ap^u, \mathcal{D}^u) \geq G_\alpha((\Pi^u_{-\alpha,1}, s^1_\alpha), \mathcal{C}ap^u, \mathcal{D}^u).$$

Ainsi, le profil Π^u est un profil stable.

Ainsi, un équilibre de Nash par parties est un équilibre séquentiellement stable. □

La Figure 3.1 résume les diverses propriétés prouvées précédemment ainsi que les propriétés qui sont dans la suite.

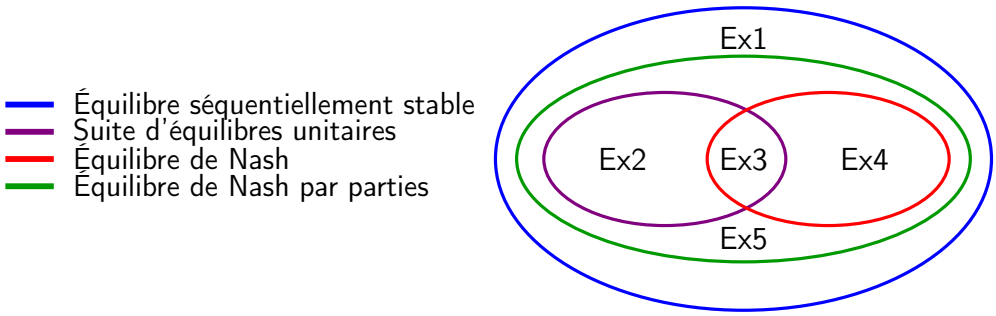

FIGURE 3.1 – Propriétés des équilibres entre eux

Toutes les propriétés seront prouvées par des exemples avec deux joueurs avec des demandes unitaires, et les joueurs peuvent jouer les prix 1, 2 et 3. La préférence de la demande d_t sera indiquée par une étoile. Le joueur qui a finalement reçu la demande aura une case grisée.

Propriété VII. Il existe des équilibres séquentiellement stables qui ne sont pas des équilibres de Nash par parties.

Exemple 1 : Dans cet exemple, le joueur A_1 a une capacité de 1 et le joueur A_2, de 3. Le tableau des prix choisis par les joueurs est le suivant :

	d_1	d_2	d_3
A_1	2*	1*	$\varnothing$
A_2	1	1	3*

Le profil de la première demande d_1 n'appartient à aucun équilibre de Nash, que ce soit pour le jeu $(\{A_1, A_2\}, (1, 3), \{d_1\})$, ni pour $\langle\{A_1, A_2\}, (1, 3), \{d_1, d_2\}\rangle$, ni pour $\langle\{A_1, A_2\}, (1, 3), \{d_1, d_2, d_3\}\rangle$. Ainsi, ce profil ne correspond pas à un équilibre par parties.

Propriété VIII. Il existe des profils qui sont des suites d'équilibres unitaires qui ne sont pas des équilibres de Nash.

Exemple 2 :
Dans cet exemple, les joueurs ont une capacité de 2.

	d_1	d_2	d_3	d_4	d_5
A_1	1	1*	1	3*	$\varnothing$
A_2	1*	1	1*	$\varnothing$	$\varnothing$

Propriété IX. Il existe des équilibres de Nash qui sont également des suites d'équilibres unitaires.

Exemple 3 :
Dans cet exemple, les joueurs ont une capacité de 4 et les prix joués sont les suivants :

	d_1	d_2	d_3	d_4
A_1	1	1*	1	1*
A_2	1*	1	1*	1

Propriété X. Il existe des équilibres de Nash qui ne sont pas des suites d'équilibres unitaires.

Exemple 4 :
Dans cet exemple, les joueurs ont une capacité de 2.

	d_1	d_2	d_3	d_4
A_1	3	3*	3	3*
A_2	3*	3	3*	$\varnothing$

Propriété XI. Il existe des équilibres de Nash par parties qui ne sont ni des équilibres de Nash ni des suites d'équilibres unitaires.

Exemple 5 :
Dans cet exemple, les joueurs ont une capacité de 2.

	d_1	d_2	d_3	d_4
A_1	1	1*	3	3*
A_2	1*	1	3*	∅

3.5.4. Équilibre dans un autre modèle de jeu répété

Dans le modèle retenu, il a été choisi de ne pas autoriser un agent à proposer un prix lorsqu'il ne dispose pas de la capacité nécessaire pour satisfaire une demande donnée. Dans ce modèle-ci, un loueur qui manque de ressources disponibles fixera un prix prohibitif pour décourager les clients ou pour couvrir le coût de louer une voiture aux concurrents.

En théorie, le fait qu'un joueur ne joue plus lorsqu'il n'a plus de disponibilités a une forte influence sur les équilibres obtenus. Il peut donc être intéressant d'étudier les équilibres dans un modèle où tous les joueurs doivent choisir un prix pour chaque demande, puis seulement les joueurs disposant d'une capacité suffisante sont considérés pour le jeu. Dans un tel cadre, les propriétés des équilibres s'en trouvent profondément modifiées, ouvrant la voie à des dynamiques stratégiques et à des résultats d'équilibre substantiellement différents, comme l'indique la Figure 3.2. Dans ce cadre, les équilibres de Nash restent, par définition, des équilibres séquentiellement stables.

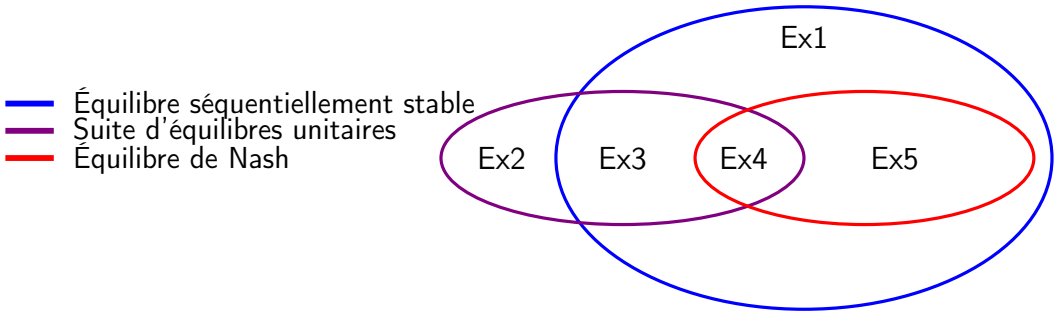

FIGURE 3.2 – Propriétés des équilibres entre eux

Tous les exemples qui servent à prouver les propriétés suivantes auront deux joueurs, des demandes unitaires, et les joueurs peuvent jouer les prix 1, 2 et 3, et ont une capacité de 2. La préférence de la demande d_t sera indiquée par une étoile. Le joueur qui a finalement reçu la demande aura une case grisée.

Propriété XII. Il existe des équilibres séquentiellement stables qui ne sont ni des équilibres de Nash ni des suites d'équilibres unitaires.

Exemple 1 :

	d_1	d_2	d_3	d_4
A_1	2	2*	1	3*
A_2	2*	2	1*	1

Propriété XIII. Il existe des profils qui correspondent à une suite d'équilibres unitaires, mais qui ne sont pas des équilibres séquentiellement stables.

Exemple 2 :

	d_1	d_2	d_3	d_4
A_1	1	1*	1	3
A_2	1*	1	1*	3

Propriété XIV. Il existe des profils qui correspondent à une suite d'équilibres unitaires et qui sont des équilibres séquentiellement stables, mais qui ne sont pas des équilibres de Nash.

Exemple 3 :

	d_1	d_2	d_3	d_4
A_1	1	1*	1	3
A_2	1*	1	1*	1

Propriété XV. Il existe des profils qui correspondent à une suite d'équilibres unitaires, mais qui sont à la fois des équilibres séquentiellement stables et des équilibres de Nash.

Exemple 4 :

	d_1	d_2
A_1	1	1*
A_2	1*	1

Propriété XVI. Il existe des équilibres de Nash qui ne sont pas des suites d'équilibres unitaires.

Exemple 5 :

	d_1	d_2	d_3	d_4
A_1	3	3*	3	3*
A_2	3*	3	3*	3

3.6. Conclusion

Ce chapitre a proposé une modélisation formelle du problème de tarification dynamique en contexte concurrentiel à l'aide d'un cadre de jeu répété sous contrainte de capacité. Après avoir introduit le contexte et les hypothèses générales, le jeu a été défini à deux niveaux : un jeu unitaire, correspondant à une demande isolée, et un jeu répété, correspondant à une séquence ordonnée de demandes dans le temps. Cette construction permet de capturer explicitement l'impact des décisions successives des agents via l'évolution de leurs capacités disponibles.

L'analyse théorique a d'abord établi l'existence d'un équilibre de Nash pour le jeu unitaire, puis montré que cette notion, bien que pertinente localement, est insuffisante pour caractériser le comportement stratégique global dans le jeu répété. En particulier, un profil constitué d'équilibres unitaires n'est pas nécessairement un équilibre de Nash au sens du jeu répété. Cette observation a motivé l'introduction de notions d'équilibre mieux adaptées à un contexte séquentiel et en ligne.

L'équilibre séquentiellement stable a ainsi été défini et étudié, car il repose sur la stabilité du profil à chaque étape du jeu. Il a été démontré qu'un équilibre de Nash et une suite d'équilibres unitaires sont forcément des équilibres séquentiellement stables. Les exemples présentés en fin de chapitre montrent explicitement que ces ensembles d'équilibres ne coïncident généralement pas, ce qui met en évidence la complexité structurelle du modèle.

Les résultats de ce chapitre soulignent tout d'abord les limites des concepts d'équilibre classiques lorsqu'ils sont appliqués à des problèmes de décision séquentielle fortement contraints par des ressources limitées. En particulier, la notion d'équilibre de Nash pour les jeux répétés, bien que standard en théorie des jeux, apparaît peu adaptée à un contexte opérationnel où les agents prennent progressivement leurs décisions, en tenant compte de l'historique et de l'état courant du système.

Ces résultats théoriques ont plusieurs conséquences directes pour la suite du manuscrit. D'un point de vue théorique, ils ouvrent la voie à l'étude de variantes du modèle, notamment en modifiant certaines hypothèses structurantes, telles que l'autorisation ou non de proposer un prix en l'absence de capacité. Comme évoqué précédemment, de telles extensions modifient profondément les propriétés des équilibres et méritent une analyse spécifique. D'un point de vue méthodologique, les notions d'équilibre introduites dans ce chapitre servent de références théoriques pour évaluer les approches algorithmiques étudiées par la suite.

Enfin, ce chapitre établit un socle conceptuel solide reliant la théorie des jeux, la modélisation dynamique et les problématiques de tarification réelle. Il justifie, dans les chapitres suivants, le recours à des approches approximatives ou apprenantes visant non pas à atteindre un équilibre théorique strict, mais à

reproduire des comportements compatibles avec les équilibres pertinents identifiés ici, tout en restant exploitables dans des contextes opérationnels réels.

3.7. Liste des notations

3.7.1. Notations pour l'agent A_α (Section 3.3.1)

$Pr_\alpha = \{pr_\alpha[1], pr_\alpha[2], \ldots, pr_\alpha[M_\alpha]\}$ Ensemble des prix possibles proposés par l'agent A_α, ordonnés du plus faible au plus élevé.

$Capmax_\alpha$ Capacité totale de l'agent A_α, c'est-à-dire le nombre total de véhicules qu'il possède.

$Util_\alpha(p)$ Nombre de véhicules déjà loués par l'agent A_α sur la période p.

$Cap_\alpha(p)$ Nombre de véhicules disponibles de l'agent A_α sur la période p, défini par

$$Cap_\alpha(p) = Capmax_\alpha - Util_\alpha(p).$$

$Cap^t_\alpha(p)$ Capacité restante de l'agent A_α sur la période p à l'étape t du jeu répété.

3.7.2. Notations de la demande de location d (Section 3.3.2)

$recep_d$ Période de réception de la demande d.

$debut_d$ Période de début de location associée à la demande d.

fin_d Période de fin de location associée à la demande d.

$delaiResa_d$ Délai de réservation qui correspond au temps entre la réservation et le début de location

$$delaiResa_d = debut_d - recep_d$$

$duree_d$ Durée de location de la demande d, donnée par

$$duree_d = fin_d - debut_d + 1.$$

$Pmax_d$ Prix maximal accepté par le client pour la demande d.

$ordre_d$ Ordre de préférence du client sur les agents.

3.7.3. Notations du jeu unitaire (Section 3.4.1)

s_α Stratégie de l'agent A_α dans le jeu unitaire, c'est-à-dire le prix proposé pour la demande d.

$\pi = (s_1, s_2, \ldots, s_N)$ Profil de stratégies du jeu unitaire.

$g_\alpha(\pi, d)$ Gain de l'agent A_α dans le jeu unitaire :

$$g_\alpha(\pi, d) = \begin{cases} 0 & \text{si } A_\alpha \neq A_{gagnant}, \\ duree_d \times \min(\pi) & \text{si } A_\alpha = A_{gagnant}. \end{cases}$$

3.7.4. Notations du jeu répété (Section 3.4.2)

$D = (d^1, d^2, \ldots, d^T)$ Séquence ordonnée des demandes considérées dans le jeu répété.

d^t Demande reçue à l'étape t du jeu répété.

T Nombre total de demandes, donc nombre total d'étapes du jeu répété.

$S_\alpha = (s_\alpha^1, s_\alpha^2, \ldots, s_\alpha^T)$ Stratégie complète de l'agent A_α dans le jeu répété.

s_α^t Prix proposé par l'agent A_α à l'étape t pour la demande d^t.

$\pi^t = (s_1^t, s_2^t, \ldots, s_N^t)$ Profil de stratégies joué à l'étape t.

$\Pi = (\pi^1, \pi^2, \ldots, \pi^T)$ Profil de jeu répété complet.

$DemC^t(A_\alpha)$ Sous-ensemble de demandes de $\mathcal{D}^{\sqcup}$ pour lesquelles l'agent A_α a été sélectionné

$g_\alpha(\pi^t, d^t)$ Gain unitaire de l'agent A_α à l'étape t.

$G_\alpha(\Pi, Cap, D)$ Gain cumulé de l'agent A_α dans le jeu répété :

$$G_\alpha(\Pi, Cap, D) = \sum_{t=1}^{T} g_\alpha(\pi^t, d^t).$$

Chapitre 4
Analyse de données

Sommaire

4.1 Introduction 43
4.2 Ensemble de données réelles 44
4.3 Étude du lien entre le prix et les caractéristiques des demandes 45
4.3.1 Étude des distributions 46
4.4 Étude des distributions par mois 51
4.4.1 Étude des corrélations 54
4.4.2 Classification des données 57
4.5 Étude de la saisonnalité pour un historique de données 59
4.5.1 Caractérisation des jours 60
4.5.2 Méthodes de classification : K – *means* 61
4.5.3 Méthodes de comparaison des classifications 63
4.5.4 Obtention et évaluation d'une saisonnalité 64
4.6 Historique et prédiction des données 67
4.6.1 Prédiction des demandes 67
4.6.2 Saisonnalité et dynamique temporelle 69
4.7 Conclusion 71

4.1. Introduction

Dans le contexte du modèle introduit dans le chapitre 3, ce chapitre va étudier s'il est possible de déterminer comment certains loueurs de voitures ont fixé leur prix à partir de leurs données historiques.

Dans un premier temps, la Section 4.2 présente les ensembles de données réelles collectées auprès de plusieurs loueurs grâce à WeYield. Ces jeux de données seront utilisés tout au long de cette thèse. Dans un deuxième temps, la Section 4.3 étudie les relations potentielles entre les prix pratiqués par les loueurs et les caractéristiques des demandes observées telles que la durée de location, le délai de réservation et le niveau de disponibilité des voitures au moment de la réservation. Le but est d'établir si les caractéristiques des demandes sont corrélées avec le choix des prix.

Pour continuer cette étude de données, le but est d'obtenir le maximum d'informations utilisables par les méthodes de tarification dynamique. Dans la Section 4.5, le but sera d'analyser les données pour savoir si les loueurs

ont une politique tarifaire qui varie en fonction des dates de location. Dans la Section 4.6, l'objectif sera de trouver une méthode pour prédire des indicateurs qui constitueront un élément clé pour améliorer les méthodes de tarification présentées dans le Chapitre 6.

4.2. Ensemble de données réelles

Cette section décrit les jeux de données utilisés dans la suite de ce travail. WeYield dispose des données de réservations enregistrées par des loueurs de véhicules, c'est-à-dire des informations indiquant quel véhicule a été loué et pour quelle période. WeYield permet aux clients d'utiliser ces données pour prendre des décisions en termes de prix. Il est important de souligner certaines limites inhérentes à cet ensemble de données. Tout d'abord, les demandes de location qui n'aboutissent pas à une réservation ne sont pas obtenues par WeYield. Ensuite, lorsqu'une réservation est annulée, elle n'apparaît plus dans le fichier principal des réservations et est reportée dans un fichier distinct géré par WeYield. Ces caractéristiques impliquent que les données analysées reflètent uniquement les réservations confirmées à un instant donné, et non l'ensemble des intentions de demande exprimées par les clients.

Dans les données obtenues grâce à WeYield, seules les caractéristiques suivantes seront utilisées pour notre analyse :

- Date de réservation
- Date de départ
- Date d'arrivée
- Durée
- Type de voitures (seulement dans ce chapitre)
- Prix de base
- Prix total, utilisé comme valeur pour $Pmax$ (Section 3.3.2)

Dans ces données, deux types de revenus sont distingués. Le prix de base correspond exclusivement au prix de la location du véhicule et c'est celui qui sera utilisé dans ce chapitre. Le prix total, quant à lui, représente le montant global acquitté par le client et inclut, en plus du prix de la location, les revenus issus des options et services additionnels tels que l'assurance, le siège enfant ou d'autres prestations complémentaires. Le prix total sera uniquement utilisé dans les chapitres 5 et 6.

4.3. Étude du lien entre le prix et les caractéristiques des demandes

L'analyse des comportements de tarification repose sur l'étude des demandes de location à partir des données présentées dans la Section 4.2. Néanmoins, l'ensemble des variables observées ne présente pas le même niveau de pertinence ni la même exploitabilité pour les analyses statistiques. En particulier, les dates de début et de fin de location constituent des informations fondamentales, mais peu utiles à comparer entre elles. C'est pourquoi nous privilégions donc des variables qui seront la durée entre deux dates.

Concernant le type de véhicule demandé, les analyses sont menées séparément pour chaque catégorie. Ce choix méthodologique reflète la pratique des loueurs, qui considèrent chaque catégorie de véhicule comme un segment de marché distinct, d'après l'expérience de WeYield. De plus, ces catégories n'ont pas le même comportement, certains types de véhicules étant davantage associés à des usages professionnels, tandis que d'autres, comme les véhicules familiaux, sont principalement liés à des usages touristiques.

L'analyse se concentre ainsi sur les autres informations fournies par les loueurs pour décrire les demandes, notamment :

- Durée de location : $duree_d = fin_d - debut_d$
- Délai de réservation : $delaiResa = debut_d - recep_d$
- Revenu par jour (*Revenue Per Day* RPD) : $RPD_d = Prix_d/duree_d$
- Taux d'utilisation $util_d$: C'est le ratio maximum sur la durée de la demande entre le nombre de voitures déjà réservées et la flotte prévue au moment où est faite la réservation.

L'analyse présentée dans ce chapitre est principalement réalisée à partir des données d'un loueur de véhicules implanté sur une île française à forte attractivité touristique. Ce loueur disposait d'une flotte d'environ 600 véhicules en 2022 et sera désigné par la suite comme le loueur tourisme.

Afin d'évaluer la pertinence des résultats obtenus, des analyses complémentaires ont également été menées sur des jeux de données provenant d'autres profils de loueurs. Ces derniers incluent un loueur insulaire de taille comparable, opérant principalement auprès de clients professionnels sans contrat préétabli et disposant d'une flotte d'environ 900 véhicules, ainsi qu'un loueur continental de grande taille possédant environ 1 600 véhicules. Enfin, les données d'un loueur spécialisé dans la location de vans, avec une flotte d'environ 500 véhicules, ont également été étudiées.

Cette diversité de profils permettra de confronter les résultats obtenus dans un contexte touristique à des environnements opérationnels distincts, caractérisés par des structures de demande, des usages et des contraintes de flotte différents.

Dans cette section, le but va donc être d'étudier les caractéristiques décrites précédemment pour mieux comprendre le lien entre ces caractéristiques et les prix. Pour cela, la Section 4.3.1 analysera les distributions de ces caractéristiques pour les quatre loueurs, ensuite, le but sera de savoir s'il existe des corrélations avec la Section 4.4.1. Pour finir, le but sera de chercher les liens entre les prix appliqués et les caractéristiques des réservations grâce à des méthodes de classification.

4.3.1. Étude des distributions

Dans cette section, le but est d'étudier les distributions des caractéristiques de la demande, chez le loueur tourisme, et de comparer ces distributions à celles des trois autres loueurs décrits précédemment.

Pour commencer, nous allons nous concentrer sur le loueur à but touristique. Pour cela, nous avons les quatre figures suivantes, une par caractéristique de la demande. Tout d'abord, la figure 4.1 représente la distribution des durées des demandes, où la durée la plus présente est 14 jours, ce qui correspond à des réservations touristiques de deux semaines. Les durées de location, qui sont principalement utilisées chez ce loueur, sont comprises entre 7 jours et 14 jours, ce qui correspond aussi aux durées les plus utilisées par les touristes.

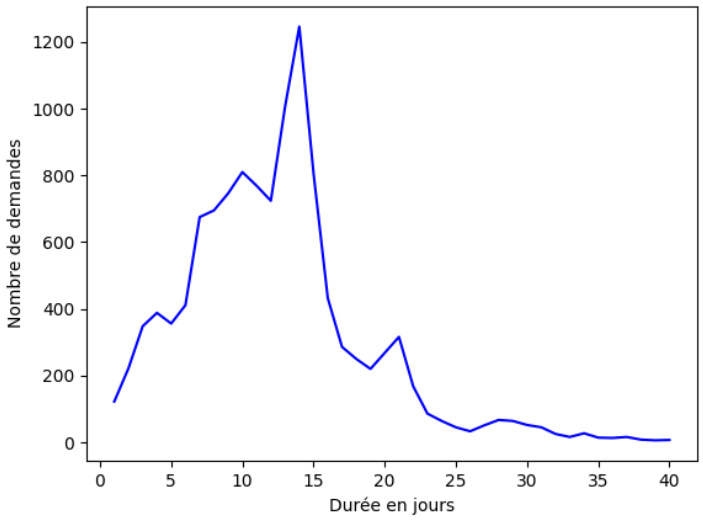

FIGURE 4.1 – Distribution du nombre de demandes en fonction de la durée de location

Après avoir analysé la distribution des durées des demandes associées au loueur tourisme, nous comparons désormais ces résultats à ceux observés pour d'autres profils de loueurs.

La Figure 4.2 présente la distribution des durées de location pour les quatre loueurs étudiés. Elle met en évidence des comportements distincts selon le type d'activité. Le loueur orienté vers la clientèle professionnelle réalise majoritairement des locations d'une journée, tandis que le loueur spécialisé dans la location de vans a principalement des locations avec des durées comprises entre trois et quatre jours. Le loueur continental présente un comportement intermédiaire, combinant des locations de courte durée, typiquement associées à une clientèle professionnelle, et des locations d'une semaine, plus caractéristiques d'un usage touristique. Enfin, le loueur tourisme, fortement lié aux vols long-courriers, se distingue par des durées de location plus longues, généralement comprises entre sept et quinze jours.

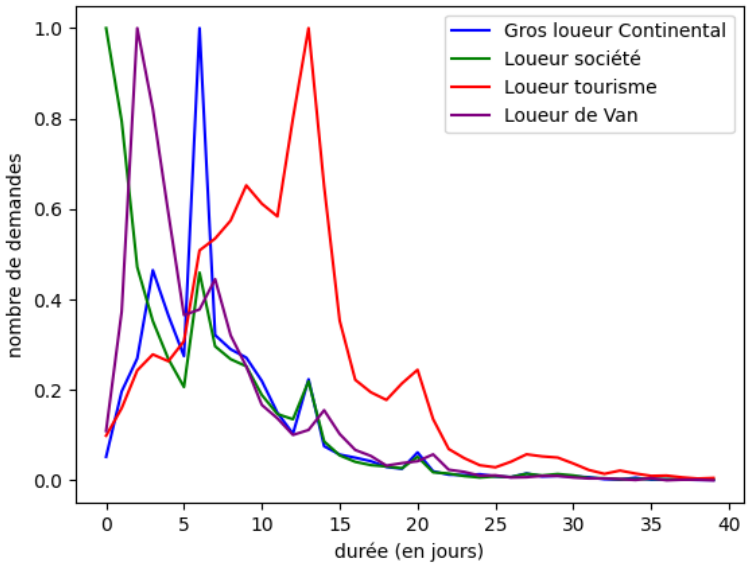

FIGURE 4.2 – Distribution du nombre de demandes en fonction de la durée de location en fonction des loueurs

La courbe 4.3 correspond à la distribution du délai de réservation, c'est-à-dire le temps de réservation à l'avance. Ce loueur étant un loueur tourisme associé à des vols dits "long-courriers", le délai de réservation est très long par rapport au comportement habituel des clients, ce qui est visible sur la figure 4.4. Cela s'explique par le fait que les clients touristiques prenant des vols dits "long-courriers" planifient leurs déplacements et leurs vacances plusieurs semaines, voire plusieurs mois, à l'avance, et donc aussi leur location de voitures.

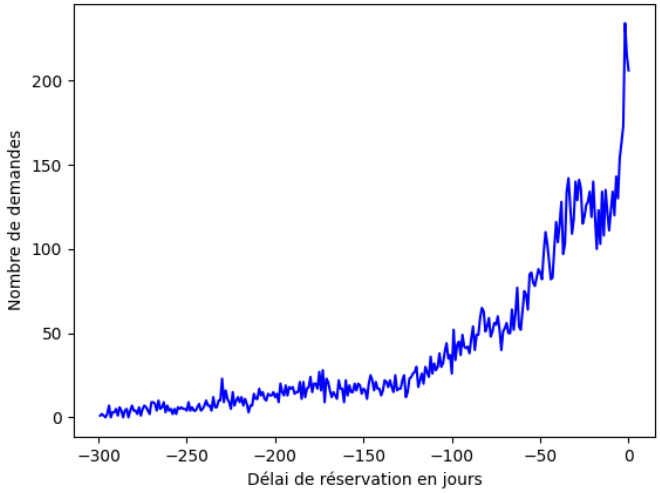

FIGURE 4.3 – Distribution du nombre de demandes en fonction du délai de réservation

La Figure 4.4 illustre la distribution des délais de réservation pour ces mêmes loueurs. Les courbes associées au loueur de vans et au loueur tourisme montrent des délais de réservation relativement longs, contrairement aux deux autres profils. Ce résultat s'explique principalement par le comportement des clients touristiques, qui planifient leurs déplacements et leurs vacances plusieurs semaines, voire plusieurs mois, à l'avance.

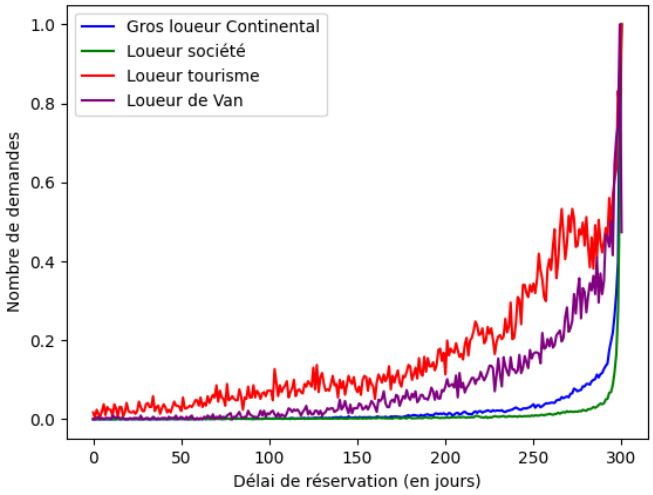

FIGURE 4.4 – Distribution du nombre de demandes en fonction du délai de réservation en fonction des loueurs

L'analyse conjointe des deux distributions présentées dans les figures 4.4 et 4.2 met en évidence que la durée moyenne de location et le délai moyen de réservation ne sont pas nécessairement corrélés. En particulier, des durées de location longues n'impliquent pas systématiquement des délais de réservation élevés, ce qui souligne la diversité des comportements de demande selon les contextes opérationnels.

La courbe 4.5 montre le nombre de demandes pour chaque valeur de RPD, la plupart des demandes ont un RPD entre 15 et 30 euros.

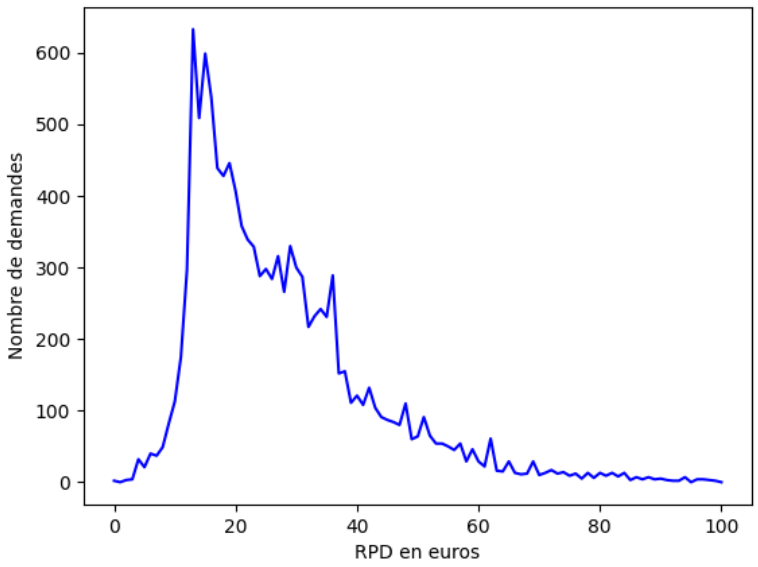

FIGURE 4.5 – Distribution du nombre de demandes en fonction du RPD

En comparant les deux courbes 4.5 et 4.3, on peut se demander si le pic de demandes avec un délai de réservation faible correspond au pic de demandes avec un RPD d'environ 15 euros. Pour étudier cela, la figure 4.6 montre la distribution du RPD des demandes avec un délai de réservation de moins de 2 jours. Cette figure montre bien que les demandes avec un délai de réservation faible ont principalement un RPD de 15 euros. Sur ce même ensemble de demandes, la distribution des durées (figure 4.7) est différente de celle de la figure 4.1. Cette différence peut s'expliquer par le fait que les touristes ne réserveront pas leur location de voiture à la dernière minute.

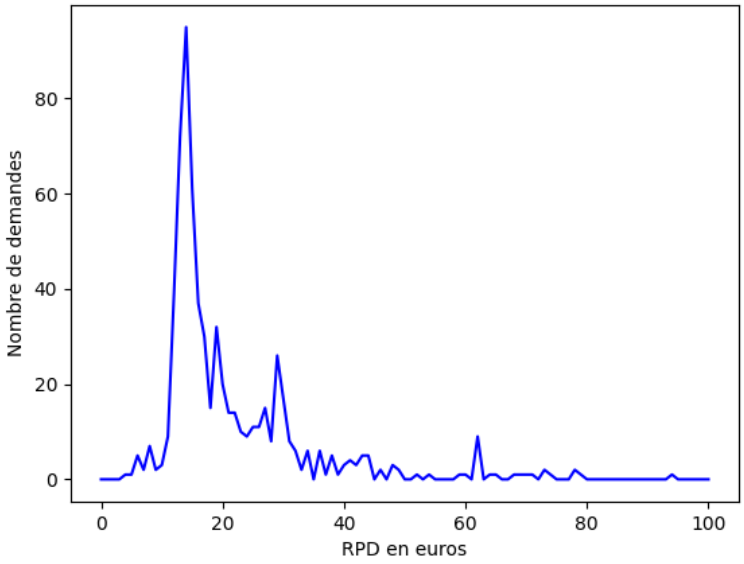

FIGURE 4.6 – Distribution du nombre de demandes en fonction du RPD

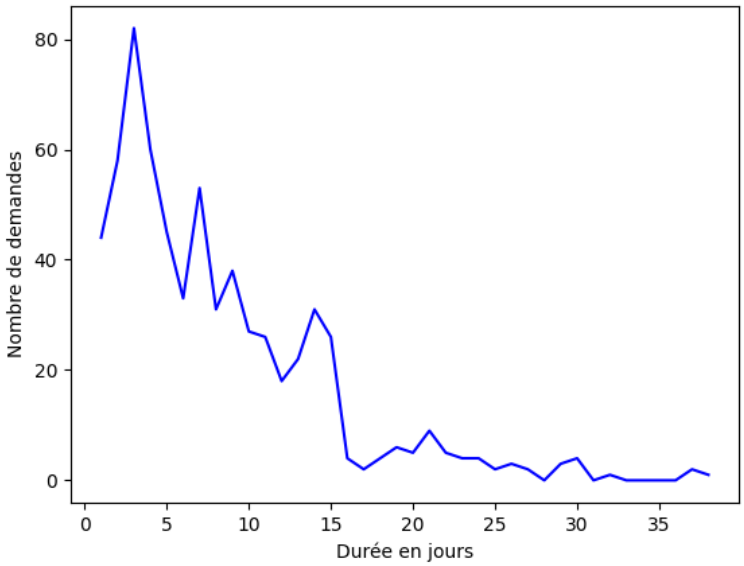

FIGURE 4.7 – Distribution du nombre de demandes en fonction de la durée

La courbe 4.8 montre la distribution de la quatrième caractéristique définie dans 4.3, c'est-à-dire le nombre de demandes pour chaque taux d'utilisation qui est calculé au moment où est faite la demande. On observe sur cette courbe

que le taux d'utilisation n'atteint pas la valeur maximale de 1. Ce phénomène s'explique par deux facteurs principaux.

Premièrement, d'un point de vue opérationnel, les loueurs ne peuvent pas mobiliser l'intégralité de leur flotte à tout instant, certaines unités étant indisponibles en raison d'opérations de maintenance, de réparations ou de phases de nettoyage entre deux locations.

Deuxièmement, l'analyse repose uniquement sur les demandes dites *yieldables*, c'est-à-dire les demandes pour lesquelles le prix n'a pas été négocié en amont de la réservation. Or, le taux d'utilisation est calculé relativement à la flotte totale, alors qu'une partie de celle-ci est structurellement dédiée aux demandes non yieldables. Cette dissymétrie entre le périmètre des demandes considérées et celui de la flotte de référence contribue également à limiter le niveau maximal observé du taux d'utilisation.

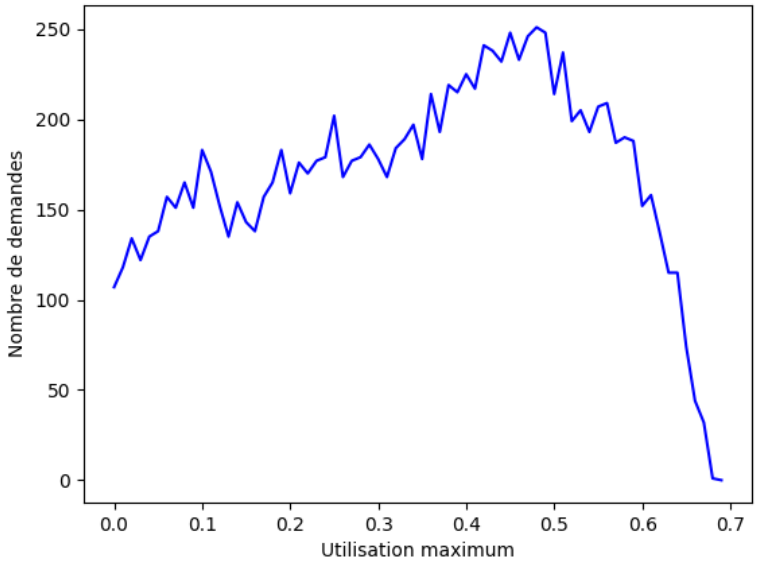

FIGURE 4.8 – Distribution du nombre de demandes en fonction du taux d'utilisation maximum

4.4. Étude des distributions par mois

Après avoir analysé les données sur l'ensemble de l'année, nous nous intéressons désormais à la variabilité des comportements de location selon les mois.

Les Figures 4.9 et 4.10 représentent, pour chacun des deux loueurs considérés, la distribution mensuelle du nombre de demandes en fonction du délai de réservation. Chaque cellule de la matrice est colorée pour refléter le nombre

de demandes associées au couple (mois, délai de réservation) correspondant.

Ces visualisations mettent en évidence des comportements nettement différenciés en fonction des mois regardés. Pour le loueur de vans, l'activité est principalement concentrée sur les périodes de printemps et d'été, avec une augmentation marquée du délai de réservation à l'approche des vacances estivales. À l'inverse, le loueur touristique implanté sur une île française présente une dynamique différente : la période de forte activité se situe principalement autour des vacances de la Toussaint et de Noël et le délai de réservation apparaît plus variable.

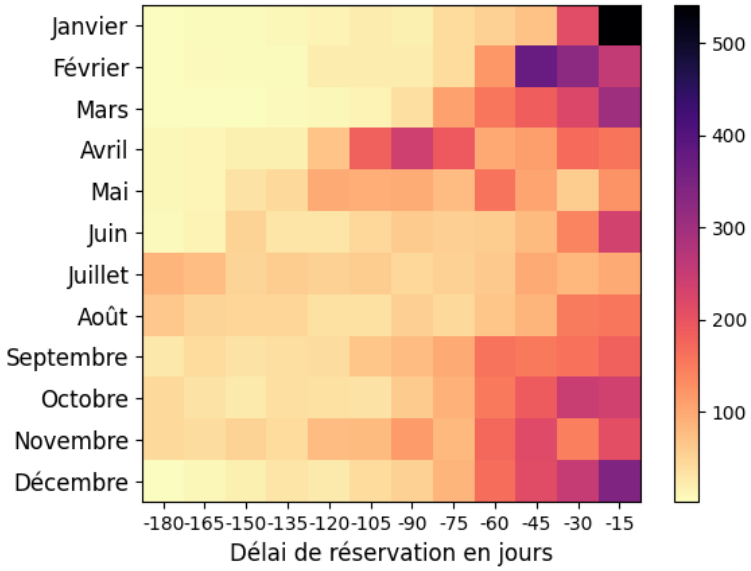

FIGURE 4.9 – Distribution du nombre de demandes en fonction du délai de réservation pour chaque mois pour le loueur tourisme

Les figures suivantes 4.11 et 4.12 représentent la distribution pour chaque mois du nombre de demandes en fonction de la durée de location. On peut voir qu'en période estivale, pour le loueur de van, la distribution des durées des demandes change fortement pour des durées plus longues. On peut aussi voir que la distribution des durées des demandes change fortement en fonction des mois pour le loueur tourisme. Ainsi, la distribution des durées des demandes dépend de la période de l'année regardée. Cela nous amènera à considérer une classification des jours en fonction de la distribution de la durée appelée saisonnalité dans la section 4.5.

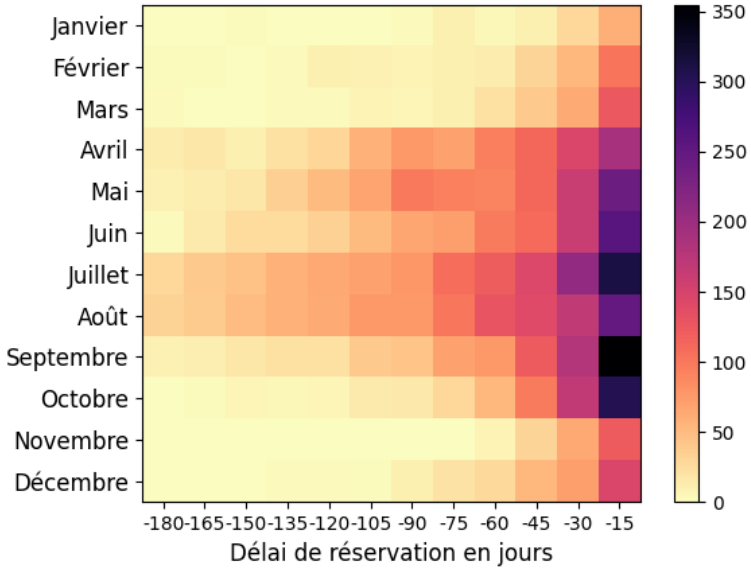

FIGURE 4.10 – Distribution du nombre de demandes en fonction du délai de réservation pour chaque mois pour le loueur de van

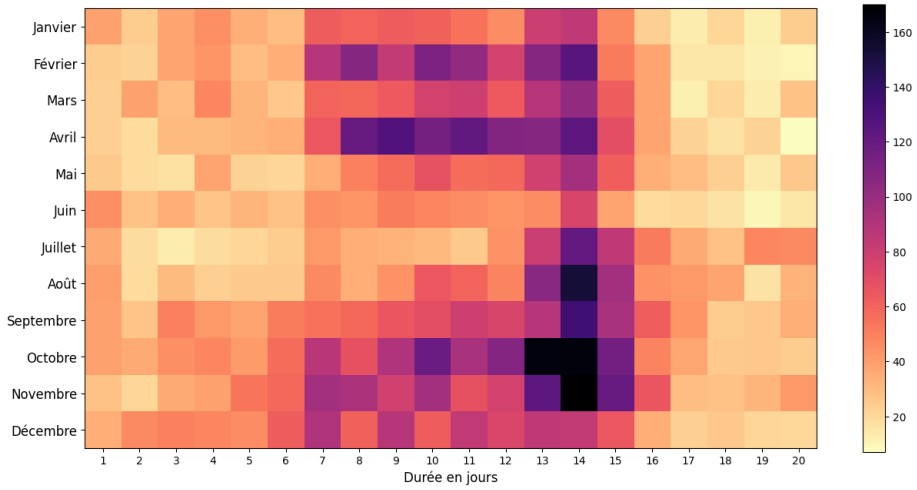

FIGURE 4.11 – Distribution du nombre de demandes en fonction de la durée pour chaque mois pour le loueur tourisme

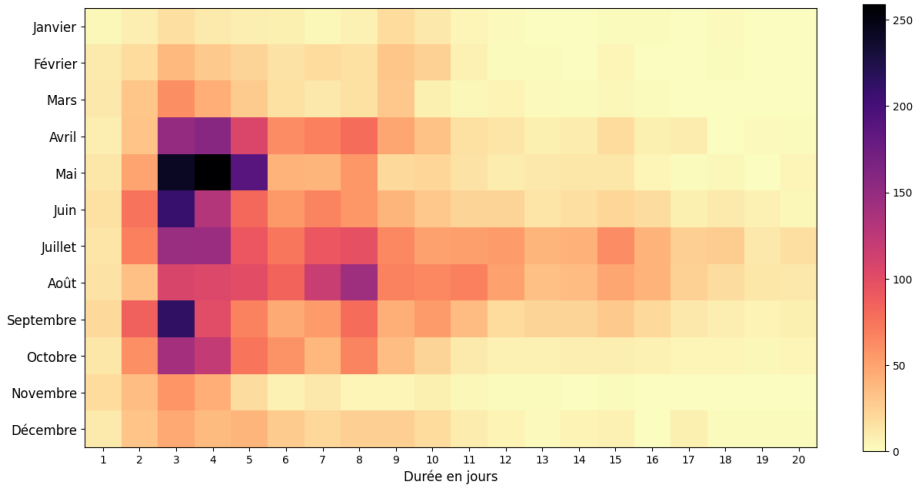

FIGURE 4.12 – Distribution du nombre de demandes en fonction de la durée pour chaque mois pour le loueur de van

4.4.1. Étude des corrélations

Dans ce paragraphe, nous examinons les corrélations entre les différentes caractéristiques des réservations, analysées deux à deux sur les données du loueur tourisme, mais les résultats sont les mêmes pour les autres loueurs dans les figures 4.14, 4.15 et 4.16. Pour chaque paire de caractéristiques considérée, chaque réservation est représentée par un point situé à l'intersection des valeurs correspondantes, ce qui permet de visualiser d'éventuelles relations entre les variables.

Une corrélation significative apparaît entre le délai de réservation et le taux d'utilisation maximal, comme l'illustre la Figure 4.13. Cette figure correspond au cas d'un loueur à dominante touristique, mais cette relation a également été observée pour les autres loueurs analysés. Ce résultat est intuitif : plus une réservation est effectuée à l'avance, plus le nombre de réservations déjà enregistrées est faible, ce qui se traduit mécaniquement par un taux d'utilisation plus bas.

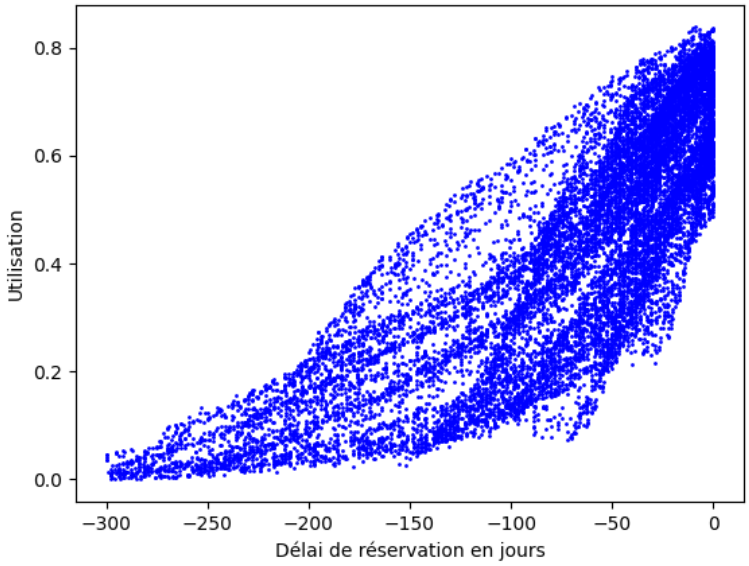

FIGURE 4.13 – Corrélation entre le délai de réservation et le taux d'utilisation maximum pour le loueur tourisme

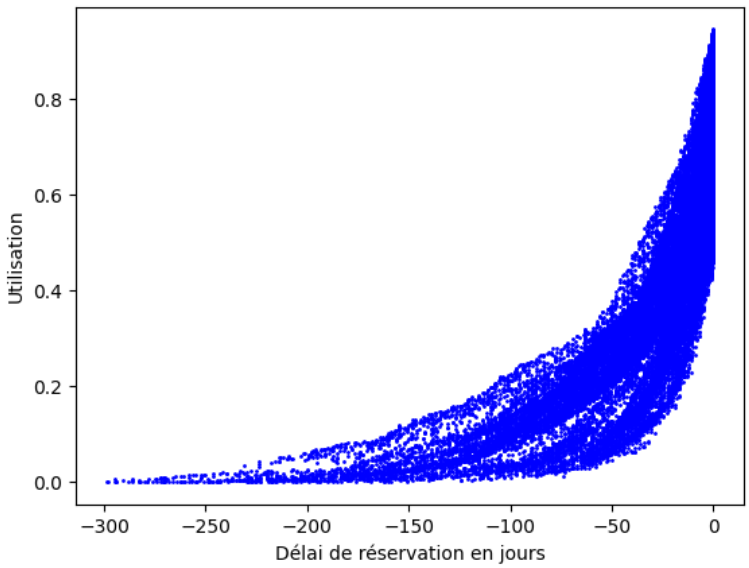

FIGURE 4.14 – Corrélation entre le délai de réservation et le taux d'utilisation maximum pour le loueur continental

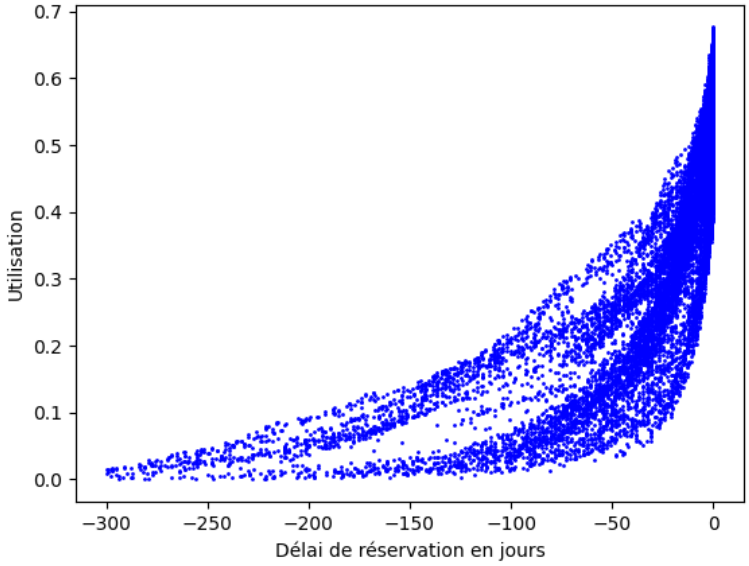

FIGURE 4.15 – Corrélation entre le délai de réservation et le taux d'utilisation maximum pour le loueur société

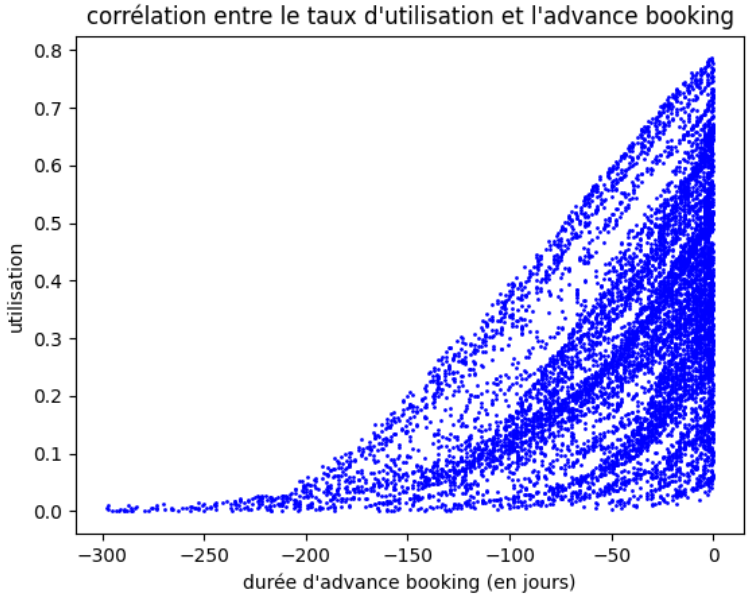

FIGURE 4.16 – Corrélation entre le délai de réservation et le taux d'utilisation maximum pour le loueur de van

La Figure 4.17 présente, quant à elle, une relation entre le revenu par jour

(RPD) moyen en rouge et la durée de location. On observe que le RPD décroît légèrement lorsque la durée de la location augmente. Ce comportement est attendu et généralement assumé par les loueurs de véhicules, dans la mesure où plus la durée de location est longue, plus les coûts fixes liés à la mise à disposition du véhicule sont amortis.

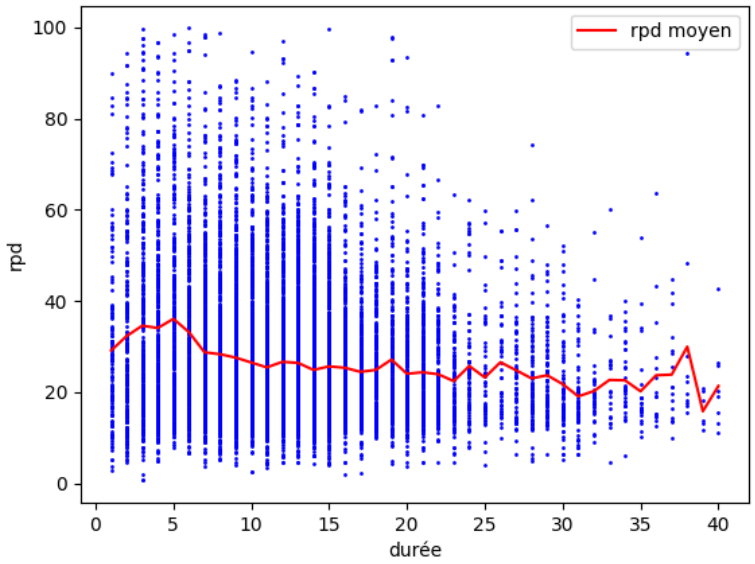

FIGURE 4.17 – Étude de la corrélation entre le délai de réservation et le RPD, avec le RPD moyen par durée en rouge

Enfin, les analyses de corrélation entre le RPD et les autres caractéristiques étudiées n'ont pas révélé de relations significatives. La section suivante cherchera à déterminer si l'application de méthodes de classification permet de faire émerger des structures ou des relations qui ne sont pas visibles à travers une analyse de corrélation globale.

4.4.2. Classification des données

L'objectif de cette section est de chercher une corrélation entre les caractéristiques de la demande : la durée, l'utilisation et le délai de réservation, et les prix en utilisant un algorithme de classification ; cette classification pourrait être utilisée dans un second temps pour la tarification.

Cette section a pour but de trouver une classification des demandes sur les caractéristiques de durée, d'utilisation et de délai de réservation qui regrouperait les demandes en fonction de leur prix ; cela pourrait permettre dans un second temps d'utiliser cette classification pour la tarification. Il y a deux ap-

proches pour faire une classification : soit avec des partitions disjointes dont les éléments appartiennent à exactement une classe, soit avec des classes avec du recouvrement où un élément peut appartenir à plusieurs classes. Les distributions de chaque caractéristique de la demande sont continues, donc il semble plus intéressant d'utiliser une classification avec recouvrement. Dans notre cas, une demande sur le bord d'une classe peut être assez proche d'un autre élément d'une autre classe, il peut ainsi être pertinent que la demande soit dans les deux classes.

Pour cela, nous avons utilisé l'algorithme OClustR (*Overlapping Clustering based on Relevance*) [54], une méthode de classification par recouvrement fondée sur la théorie des graphes. OClustR vise à produire un ensemble réduit de classes tout en contrôlant explicitement leur recouvrement.

Dans notre contexte, chaque demande sera représentée par un sommet dans le graphe de similarité. Une arête est ajoutée entre deux demandes si sim la différence normalisée entre les valeurs de la caractéristique étudiée dépasse le seuil β choisi expérimentalement. Chaque sommet définit alors un *ws-graph* (« *weighted star-shaped graph* »),qui se traduit par un graphe en étoile pondéré et constitué du sommet central et de ses voisins.

L'algorithme procède en deux étapes principales :

- **Construction d'un ensemble initial de classes.**
  Après avoir créé le graphe comme décrit précédemment, le but sera d'identifier les ws-graphs pertinents. Pour cela, OClustR introduit une mesure de pertinence qui combine deux aspects :
  - la densité relative, liée au degré du sommet et à celui de ses voisins $v.dens = \frac{|\{v' \in v.adj, |v.adj| > |v'.adj|\}|}{|v.adj|}$
    avec $v.adj$ les voisins de v
  - la compacité relative, estimant la similarité dans le ws-graph calculée comme suit : $v.compact = \frac{|\{v' \in v.adj, Moy_v > Moy_{v'}\}|}{|v.adj|}$ avec $Moy_v = \frac{1}{|v.adj|} \sum_{v' \in v.adj} sim(v, v')$.

  Les ws-graphs sont triés selon cette pertinence, puis sélectionnés de façon gloutonne afin de couvrir l'ensemble du graphe tout en évitant les recouvrements inutiles.

- **Amélioration et filtrage des classes initiales.**
  Certains ws-graphs sélectionnés peuvent ne rien apporter de nouveau à la couverture (par exemple, si tous leurs sommets sont déjà couverts par d'autres ws-graphs plus denses). OClustR identifie ces ws-graphs qui n'apportent rien à la classification : un ws-graph est supprimé s'il partage davantage de voisins couverts qu'il n'en possède de non couverts. Les sommets sont alors rattachés à la classe la plus grande qui les couvre.

Cette étape réduit le nombre total de classes tout en conservant une couverture complète de la collection.

Le résultat final est un ensemble de classes présentant un recouvrement modéré. L'algorithme a une complexité $O(n^2)$, ce qui le rend applicable à notre cas où on a un grand ensemble de données.

Pour chacune des caractéristiques listées dans le paragraphe précédent, une classification est créée à partir des données du loueur tourisme. Les classifications sur la durée et le délai de réservation n'apportent pas grand-chose car elles créent principalement une catégorie qui contient la majorité des éléments dont la distribution du RPD est similaire à la distribution globale. Pour la classification sur le RPD, la classification a des classes de tailles plus similaires, mais les distributions par rapport aux autres paramètres sont proches de la distribution globale. Cela semble montrer qu'actuellement le RPD n'est pas corrélé à ces trois paramètres étudiés séparément. La classification sur le taux d'utilisation n'apporte rien sur le RPD, mais elle montre une forte corrélation entre le taux d'utilisation, la durée et le délai de réservation. Par exemple, plus le taux d'utilisation est faible, plus le délai de réservation est élevé (les réservations sont faites très à l'avance) et plus les durées de location sont élevées.

On a ensuite étudié l'intersection entre la classification calculée avec le délai de réservation, celle avec la durée et celle avec le taux d'utilisation. Cette intersection est aussi une classification, mais qui a plus de classes. Le but est de savoir si l'intersection des trois induit une variation de la distribution du RPD, mais ce n'est pas le cas.

Ces classifications montrent donc que le loueur n'utilise pas de manière évidente et détectable le délai de réservation, la durée et le taux d'utilisation pour fixer ses prix, mais le délai de réservation, la durée et le taux d'utilisation sont corrélés entre eux.

Ces mêmes résultats sont retrouvés sur tous les loueurs étudiés.

4.5. Étude de la saisonnalité pour un historique de données

On va maintenant chercher une corrélation entre la politique tarifaire et les dates de location.

Définition I (Saisonnalité)**.** Une saisonnalité est une classification des jours de location dont les classes seront appelées saisons.

La saisonnalité permettra d'étudier la présence de corrélations entre la politique tarifaire et les dates de location. Une saisonnalité intuitive basée sur l'expérience des loueurs pourrait être la suivante :

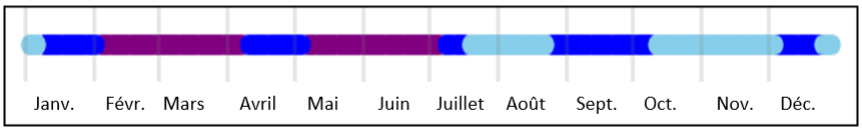

FIGURE 4.18 – Saisonnalité intuitive

Sur cette figure, les jours avec moins de demandes et les prix les plus faibles sont en violet, appelés basse saison, en bleu foncé la moyenne saison et en bleu clair la haute saison avec plus de demandes et des prix plus élevés. Cette saisonnalité intuitive correspond à la saisonnalité attendue par le loueur à qui appartiennent les données. On va définir une méthode de classification pour créer une saisonnalité en fonction des données et non des attentes du loueur. Pour cela, nous allons d'abord expliquer les métriques liées aux demandes, qui seront utilisées, dans la Section 4.5.1, puis la méthode de classification dans la Section 4.5.2, ensuite les méthodes pour mesurer la similarité entre deux classifications dans la Section 4.5.3 et pour finir les résultats concernant la saisonnalité en utilisant toutes les parties décrites en amont dans la Section 4.5.4.

4.5.1. Caractérisation des jours

Le but de cette section est de définir une mesure de similarité entre les jours afin de l'utiliser ensuite dans un algorithme de classification. Pour chaque jour de location j, on considère l'ensemble de demandes dont la période de location couvre j, qu'on notera D_j défini par :

$$D_j = \{d, debut_d \leq j \leq fin_d\} \tag{4.1}$$

Ensuite pour chaque jour j avec cet ensemble D_j, on peut calculer Moy_j^{carac} avec $carac \in \{RPD, duree, delaiResa, util\}$ la moyenne de la valeur de chaque caractéristique utilisée dans la Section 4.3 : la durée de location, le délai de réservation, le taux d'utilisation et le prix par jour RPD. Ainsi :

- Pour le RPD : $Moy_j^{RPD} = \frac{1}{|D_j|} \sum_{d \in D_j} RPD_d$
- Pour la durée : $Moy_j^{duree} = \frac{1}{|D_j|} \sum_{d \in D_j} duree_d$
- Pour le délai de réservation : $Moy_j^{delaiResa} = \frac{1}{|D_j|} \sum_{d \in D_j} delaiResa_d$
- Pour le taux d'utilisation : $Moy_j^{util} = \frac{1}{|D_j|} \sum_{d \in D_j} util_d$

Dans la suite, ces moyennes seront utilisées pour caractériser chaque jour j et définir une distance entre les jours, en vue d'identifier la manière la plus pertinente de regrouper les jours présentant des profils similaires.

4.5.2. Méthodes de classification : K – *means*

Pour partitionner l'ensemble des jours, nous recourons à la méthode des K-*means* [35]. Plusieurs raisons motivent ce choix. Tout d'abord, cette méthode produit une partition disjointe, contrairement à OClustR (vu dans la section 4.5.2) ce qui est nécessaire pour l'utiliser par la suite pour la tarification dynamique. Ensuite, le nombre de classes constitue un paramètre fixé expérimentalement. Cela permet d'avoir un nombre de classes fixe, ce qui s'avère essentiel dans la suite de ce travail, notamment lorsque cette segmentation saisonnière sera exploitée dans le cadre de la tarification dynamique.

Le principe de cette méthode de classification consiste à déterminer une partition des jours en K classes de manière à minimiser la variance intraclasse. L'algorithme débute par la sélection aléatoire de K jours servant de centroïdes initiaux, ces derniers représentant les centres des classes. Chaque jour est ensuite affecté à la classe dont le centroïde est le plus proche selon la distance considérée. Dans notre cas, la distance euclidienne est pertinente car la similarité entre deux jours correspond à la distance entre les valeurs de Moy^{carac} des deux jours avec $carac \in \{RPD, duree, delaiResa, util\}$. Une fois cette étape d'affectation réalisée, les centroïdes sont recalculés comme la moyenne des jours appartenant à chaque classe. Ces deux étapes sont répétées de manière itérative jusqu'à convergence, c'est-à-dire jusqu'à ce que les centroïdes ne varient plus d'une itération à l'autre. Il y a forcément convergence car il y a un nombre fini de partitions possibles à k classes et à chaque étape, l'algorithme fait strictement diminuer la fonction de coût, positive, et fait découvrir une meilleure partition. Cela permet d'affirmer que l'algorithme converge toujours. Par contre, le partitionnement final n'est pas toujours optimal car l'algorithme peut converger vers un minimum local. De plus, le temps de calcul peut être exponentiel en fonction du nombre d'éléments à partitionner.

Le déroulement détaillé de l'algorithme est présenté dans l'Algorithme 3.

Algorithme 3 Algorithme *K-means*

Entrée : Ensemble de données $X = \{x_1, \ldots, x_n\}$, nombre de classes k
Sortie : Classification $C = \{C_1, C_2, \ldots, C_k\}$

```
Initialiser les centroïdes avec k données aléatoires Moy_{x_i}
convergé ← faux
tant que non convergé faire
    Créer k classes vides
    pour i ← 0 à n − 1 faire
        indicePlusProche ← 0
        distanceMin ← |Moy_{x_i} − centroides[0]|
        pour j ← 1 à k − 1 faire
            dist ← |Moy_{x_i} − centroides[j]|
            si dist < distanceMin alors
                distanceMin ← dist
                indicePlusProche ← j
        C[indicePlusProche].ajouter(x_i)
    nouveauxCentroïdes ← liste vide
    pour i ← 0 à k − 1 faire
        nouveauCentroïde ← calculerCentroïde(C[i])
        nouveauxCentroïdes.ajouter(nouveauCentroïde)
    si nouveauxCentroïdes == centroïdes alors
        convergé ← vrai
    sinon
        centroïdes ← nouveauxCentroïdes
retourne C la classification
```

Pour déterminer la valeur de K, la plus pertinente dans notre cas, nous utilisons la mesure *Within-Cluster Sum of Squares* (WCSS) [35] qui est la mesure que la classification *K-means* cherche à minimiser. La WCSS est calculée en additionnant les distances au carré entre chaque point de données et son centroïde de classe correspondant. Cette mesure indique à quel point les points de données sont regroupés autour de leurs centroïdes.

Elle est utilisée dans la méthode dite « du coude » pour déterminer le nombre optimal de classes (K). Cette méthode consiste à tracer les valeurs WCSS par rapport à différentes valeurs de K et à identifier le point d'inflexion sur le graphique.

4.5.3. Méthodes de comparaison des classifications

Pour comparer les classifications, nous allons utiliser trois indicateurs plutôt classiques :

- L'Indice de Jaccard [36]
- L'Indice de Rand [59]
- L'indice de Rand Ajusté [33]

Notons deux classifications de taille n : $\mathcal{C} = \{C_1, C_2, \ldots, C_n\}$ et $\mathcal{D} = \{D_1, D_2, \ldots, D_n\}$ avec c_i la taille de C_i, d_j la taille de D_j et $n_{ij} = |C_i \cap D_j|$

L'Indice de Jaccard [36]

Pour définir l'indice de Jaccard, il faut d'abord définir le coefficient de Jaccard J :

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{4.2}$$

où A et B sont deux ensembles.

L'indice de Jaccard qui mesure la similarité entre les deux classifications C et D :

$$I_J(C, D) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{J(C_i, D_j)}{n} \tag{4.3}$$

L'Indice de Rand [59]

L'indice de Rand a une approche totalement différente de celle de l'indice de Jaccard.

$$I_R(C, D) = \frac{a + b}{\binom{n}{2}} \tag{4.4}$$

Avec a le nombre de paires d'éléments qui sont dans la même classe dans les deux classifications C et D et b le nombre de paires d'éléments qui ne sont pas dans la même classe dans les deux classifications.

Le but de l'indice de Rand est de mesurer la proportion de paires correctement classées, qu'elles soient dans la même classe dans les deux partitions ou dans des classes différentes dans les deux partitions, contrairement à l'Indice de Jaccard qui regarde juste la proportion de paires qui sont dans la même classe dans les deux partitions.

L'Indice de Rand Ajusté [33]

L'indice de Rand ajusté, quant à lui, est une complexification de l'indice de Rand.

$$I_{RA}(C,D) = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right] \Big/ \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right] \Big/ \binom{n}{2}} \tag{4.5}$$

L'indice de Rand Ajusté, qui normalise l'indice de Rand, permet de comparer deux classifications avec des nombres de classes différents et avec des variations dans la taille des classes.

Ces trois indicateurs seront utilisés pour comparer les classifications pour nous permettre d'avoir plusieurs approches pour mesurer la similarité entre deux classifications.

4.5.4. Obtention et évaluation d'une saisonnalité

L'objectif de cette section est de synthétiser les résultats relatifs à la saisonnalité des données. Nous commençons par analyser cette saisonnalité à partir des prix pratiqués par le loueur tourisme (défini dans la Section 4.2).

Dans cette optique, nous réalisons une classification des jours à l'aide de la méthode des *k-means*, en utilisant comme variable de caractérisation la métrique Moy^{RPD}. La mise en œuvre de cet algorithme nécessite de fixer au préalable la valeur de K, correspondant au nombre de classes.

Pour déterminer cette valeur, nous appliquons la méthode décrite à la fin de la Section 4.5.2. La méthode définie à la fin de la Section 4.5.2 permet d'obtenir la courbe 4.19. Le point d'inflexion se trouve pour K égal à 4. Cette valeur sera gardée pour toute la suite. Maintenant que le K est fixé, nous allons comparer les classifications faites pour chacune des trois autres métriques définies dans la Section 4.5.1.

Sur la Figure 4.20, chaque saisonnalité comprend quatre saisons qui sont chacune représentée par une couleur. La saison avec la métrique Moy^{carac} la plus basse est en violet et correspondrait à la basse saison, puis la seconde est en bleu foncé, la suivante en bleu clair et pour finir la haute saison est représentée en vert. Pour les métriques Durée et Utilisation, la basse saison représentée dans la figure 4.20 est peu visible car elle correspond seulement au(x) premier(s) jour(s) de l'année. .

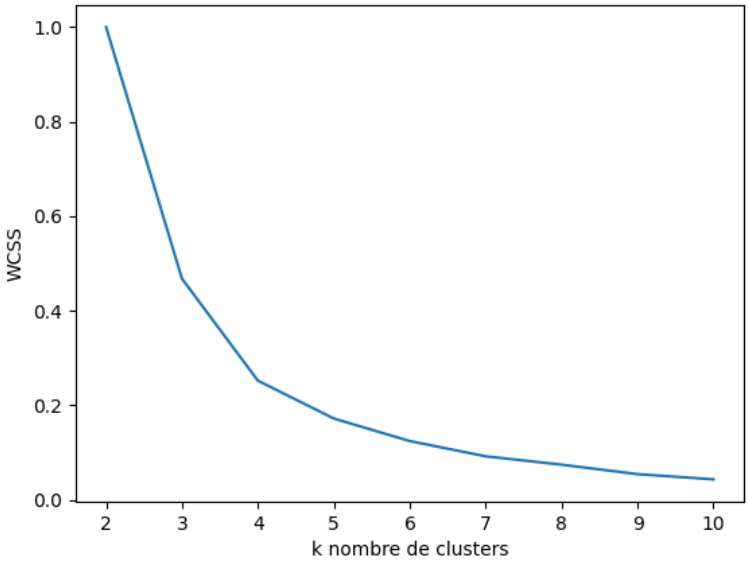

FIGURE 4.19 – Valeur de WCSS (définie dans la Section 4.5.2) en fonction du nombre de classes K

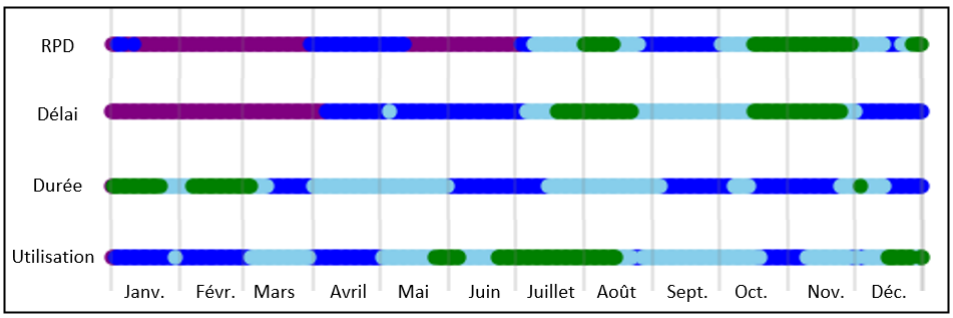

FIGURE 4.20 – Comparaison des classifications en fonction des métriques utilisées

D'après la figure 4.20, la classification à partir du RPD semble proche de celle basée sur le délai de réservation car il y a plus de jours qui sont dans la même classe dans les deux classifications, mais nous allons utiliser les indices précédemment définis dans la Section 4.5.3 pour confirmer ce ressenti. Il est important de noter que les indices sont calculés indépendamment de l'ordre des saisons.

	Indice de Jaccard	Indice de Rand	Indice de Rand Ajusté
Délai	0.67	0.31	0.84
Durée	0.49	0.31	0.75
Utilisation	0.47	0.31	0.75

Dans ce tableau, plus les valeurs sont proches de 1, plus les classifications sont proches de la classification à partir du RPD. Ainsi, ce tableau confirme que la classification sur le délai est la plus proche du RPD. De plus, les indices de similarité obtenus attestent d'une très forte similarité entre les classifications. L'indice de Jaccard atteint 0.67, indiquant une stabilité élevée des regroupements, tandis que l'indice de Rand ajusté s'élève à 0.84, révélant un accord quasi parfait entre les partitions. Bien que l'indice de Rand brut soit plus faible (0.31), cette valeur est attendue dans un contexte de classification temporelle et ne remet pas en cause la solidité de la structure mise en évidence. Cette valeur est due principalement aux variations présentes lors des changements de saison, ou de couleur sur la figure.

L'algorithme $k - means$ peut aussi utiliser plusieurs sources de données pour calculer la saisonnalité. On calcule les saisonnalités en considérant chaque couple, triplet et quadruplet parmi les métriques définies précédemment. Le choix de la saisonnalité gardée sera fait en fonction du gain en termes de bénéfice lors de l'utilisation de la saisonnalité par les méthodes de tarification dynamique.

On peut aussi comparer la saisonnalité entre deux loueurs, ici le loueur tourisme et le loueur continental, dans la figure 4.21. Il y a une réelle différence entre ces deux classifications due à la différence de clientèle. Le loueur continental a une haute saison qui correspond à la période estivale ainsi qu'aux vacances de fin d'année, alors que le loueur tourisme a une haute saison correspondant seulement aux vacances d'été et d'automne. Le premier est plus impacté par la période estivale alors que le second est plus impacté par les vacances. Cela montre une vraie différence de comportement dans les demandes.

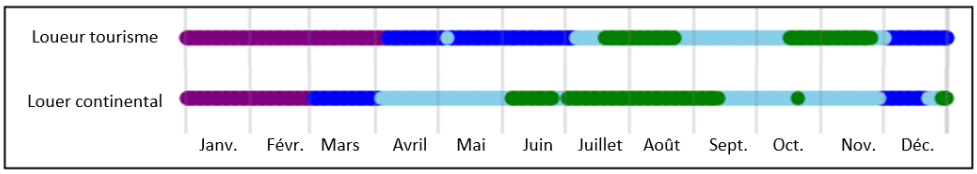

FIGURE 4.21 – Comparaison des saisonnalités entre le loueur tourisme et le loueur continental

4.6. Historique et prédiction des données

Le but de cette section est de prédire le comportement de la demande future à partir des données. Ces prédictions seront utilisées par la suite pour la tarification dynamique dans le chapitre 6. Pour commencer, cette section va étudier s'il est possible de prédire les demandes de réservation qui arriveront dans l'avenir dans la Section 4.6.1. Ensuite, dans le prolongement de la section précédente 4.5 où une méthode de classification des jours, appelée saisonnalité, a été définie, on va vouloir étendre cette méthode pour prédire une saisonnalité future.

4.6.1. Prédiction des demandes

Dans cette section, le but est de savoir s'il est possible de prédire l'ensemble des demandes qui arriveront après la période t, c'est-à-dire prédire l'ensemble des couples $(debut_d, fin_d)$ avec $recep_d > t$. De plus, il serait utile de connaître, $recep_d$ mais la précision est moins nécessaire pour cette donnée. Pour commencer, on étudiera les différences entre les ensembles de demandes acceptées de deux années consécutives. Dans un premier temps, on comparera seulement l'ensemble des couples $(debut_d, fin_d)$, dans un second temps, la valeur de $delaiResa_d$ sera aussi considérée. Ensuite on définira une mesure de distance sur des ensembles d'intervalles $(debut_d, fin_d)$ qui mesurera la proximité entre chaque couple d'intervalles et ne pas seulement savoir s'ils sont identiques ou non. Pour finir, on complexifiera cette mesure avec la valeur de $delaiResa_d$ pour affiner le résultat. Pour tous ces calculs, les ensembles de données utilisés sont l'ensemble des locations réalisées par le loueur tourisme (défini dans la section 4.2) pour l'année indiquée.

La première mesure consiste à compter le nombre de couples $(debut_d, fin_d)$ identiques entre les deux ensembles de données étudiés.

	2022 - 2023	2023 - 2024	2024 - 2025
Intervalles communs	45%	52%	51%

Le tableau précédent montre que le pourcentage d'intervalles communs est plutôt stable, mais il y a quand même environ la moitié des intervalles qui ne se reproduisent pas d'une année sur l'autre. Pour affiner cette mesure, $delaiResa_d$ sera aussi considérée, mais pas de manière exacte, car une location réservée 15 jours ou 16 jours avant est globalement identique pour le loueur. Ainsi deux demandes qui ont le même couple $(debut_d, fin_d)$ seront considérées équivalentes si :

- $delaiResa_{d1} = delaiResa_{d2} = 0$
- $min(delaiResa_{d1}, delaiResa_{d2}) \in [1, 7]$, $|delaiResa_{d1} - delaiResa_{d2}| \leq 2$

- $min(delaiResa_{d1}, delaiResa_{d2}) \in [8, 31], |delaiResa_{d1} - delaiResa_{d2}| \leq 7$
- $min(delaiResa_{d1}, delaiResa_{d2}) > 30, |delaiResa_{d1} - delaiResa_{d2}| \leq 30$

	2022 - 2023	2023 - 2024	2024 - 2025
Demandes équivalentes	25%	31%	31%

Le tableau montre qu'il y a seulement 25 à 30% de demandes équivalentes, ce qui indique que parmi les intervalles communs seulement une grosse moitié possède des délais de réservation similaires, ce qui fait peu dans un objectif de prédiction. Dans la suite, le but sera de regarder la proximité entre les intervalles et pas seulement quels sont les intervalles communs pour cela, une mesure de distance entre des ensembles d'intervalles sera définie. On commence par définir plusieurs distances entre deux demandes dont on n'utilisera que les informations du couple $(debut_d, fin_d)$:

- Distance entre les points initiaux :

$$D_I(d1, d2) := \frac{\|deb_{d1} - deb_{d2}\|}{\|max\{fin_{d1}, fin_{d2}\} - min\{deb_{d1}, deb_{d2}\}\|}$$

- Distance entre les points finaux :

$$D_F(d1, d2) := \frac{\|fin_{d1} - fin_{d2}\|}{\|max\{fin_{d1}, fin_{d2}\} - min\{deb_{d1}, deb_{d2}\}\|}$$

- Distance entre les longueurs :

$$D_L(d1, d2) := 1 - \frac{min\{fin_{d1} - deb_{d1}, fin_{d2} - deb_{d2}\}}{max\{fin_{d1} - deb_{d1}, fin_{d2} - deb_{d2}\}}$$

- Chevauchement :

$$D_C(d1, d2) := 1 - \frac{\|d1 \cap d2\|}{min\{\|fin_{d1} - deb_{d1}\|, \|fin_{d2} - deb_{d2}\|\}}$$

On définit alors une distance plus générale avec ces distances pour deux intervalles d1 et d2 :

$$S(d1, d2) := \frac{1}{|I|} \sum_{i \in I} D_i(d1, d2)$$

avec I l'ensemble des distances décrites précédemment. Une distance proche de 0 indique une proximité entre les deux ensembles d'intervalles. On obtient alors les résultats suivants :

	2022 - 2023	2023 - 2024	2024 - 2025
Distance intervalle	0.22	0.11	0.10

Les distances obtenues sont proches de 0 bien que la distance entre 2022 et 2023 soit plus élevée, ce qui était déjà perceptible dans le pourcentage d'intervalles communs et de demandes équivalentes. Cette distance plus élevée est due aux conséquences de la Covid sur l'année 2022. Cette distance nous indique que les ensembles sont proches malgré les presque 50% d'intervalles qui ne sont pas communs.

La *distance intervalle* va maintenant être adaptée pour considérer la valeur de $delaiResa_d$ et sera nommée *distance équivalente*. La *distance équivalente* entre $d1$ et $d2$ vaudra 1 si $delaiResa_{d1}$ et $delaiResa_{d2}$ ne vérifient pas les conditions présentées dans la définition de demandes équivalentes (début du 4.6.1) et vaudra la *distance intervalle* sinon.

	2022 - 2023	2023 - 2024	2024 - 2025
Distance équivalente	0.31	0.20	0.21

Le tableau montre que la *distance équivalente* est plus élevée que la *distance intervalle* car on considère le $delaiResa_d$. Sachant qu'il y a 31% de demandes équivalentes entre 2024 et 2025. La *distance équivalente* entre les ensembles de demandes sans les demandes équivalentes monte à 0.30. Le nombre de demandes équivalentes est plutôt bas et la *distance équivalente* est plutôt élevée pour pouvoir créer une prédiction précise qui pourrait être ensuite utilisée pour la tarification dynamique. On a donc décidé de suivre une autre approche basée sur la saisonnalité.

4.6.2. Saisonnalité et dynamique temporelle

On a vu dans la section 4.5 qu'il était possible de créer une classification des jours appelée saisonnalité en regroupant les jours en fonction des caractéristiques de la demande comme le prix. Ainsi, on peut créer une classification des jours qui met en relation les jours et les prix des demandes. Le but de cette section va être de déterminer la saisonnalité sur des données partielles, ce qui correspondrait à définir une saisonnalité pour les jours dans l'avenir.

Étant donnée une période t, la section 4.5 définit une méthode qui détermine la saisonnalité des jours précédant la période t en connaissant l'ensemble des demandes reçues pour ces jours. Dans cette section, le but sera de définir une méthode qui détermine la saisonnalité pour les jours compris entre $t - T_1$ et $t + T_2$ avec T_1 et T_2 strictement positifs en considérant seulement l'ensemble des demandes reçues jusqu'à la période t. Cette saisonnalité sera appelée saisonnalité prédite. Ensuite, nous comparerons la saisonnalité prédite avec la saisonnalité obtenue à partir de la méthode de la section 4.5 pour les jours de $t - T_1$ et $t + T_2$ en connaissant l'ensemble des demandes reçues pour ces jours avec la méthode décrite dans cette section où l'on connaît seulement les demandes reçues jusqu'à l'étape t. Cette saisonnalité sera appelée saisonnalité

omnisciente.

Ainsi, dans cette section, on considère seulement l'ensemble des demandes reçues jusqu'à l'étape t. Il ne va pas être possible de calculer la saisonnalité à partir de la méthode de la section 4.5 car l'ensemble des demandes reçues n'est pas complet pour les jours après t. Pour compléter l'ensemble des demandes reçues, on considère que les demandes futures sont identiques à celles qui ont été reçues l'année dernière pour les mêmes dates. On va d'abord définir deux ensembles de demandes pour une période t :

- $D^{\leq t} = \{d \text{ tel que } recep_d \leq t \text{ et } t < debut_d < t + 364\}$
- $D^{>t} = \{d \text{ tel que } recep_d > t \text{ et } t < debut_d < t + 364\}$

Le premier ensemble de demandes correspond aux demandes réservées avant t pour une période de location après t et le deuxième correspond aux demandes réservées et louées après t. On notera ensuite pour toute période j :

- $D_j^{\leq t} = \{d \text{ tel que } recep_d \leq t \text{ et } t < debut_d < t + 364 \text{ et } debut_d \leq j \leq fin_d\}$
- $D_j^{>t} = \{d \text{ tel que } recep_d > t \text{ et } t < debut_d < t + 364 \text{ et } debut_d \leq j \leq fin_d\}$

Ainsi, pour tous les jours j, D_j^t sera défini comme suit :

$$D_j^t = \begin{cases} D_j^{\leq t}, & \text{si } j \leq t \\ D_j^{\leq t} \cup D_{j-364}^{>t-364} & \text{sinon } j > t \end{cases}$$

L'ensemble $D_{j-364}^{>t-364}$ désigne les demandes qui concernent le jour équivalent à j l'année précédente et qui ont été reçues après la date t décalée d'un an. Autrement dit, il s'agit des réservations concernant le jour $j-364$ qui n'étaient pas encore connues à la même date un an auparavant. Le décalage de 364 jours, et non de 365 jours, est volontaire. Il permet d'aligner les jours de la semaine, de sorte qu'un lundi soit comparé à un lundi, ce qui est particulièrement important dans le secteur de la location de véhicules où le comportement des demandes présente une forte composante hebdomadaire.

L'utilisation de D_j^t dans la méthode de la Section 4.6.2 permettra de calculer la saisonnalité prédite pour les périodes après la période t.

Nous allons maintenant étudier les saisonnalités calculées sur les années 2023 et 2024 en utilisant d'une part $\{d, debut_d < 1^{er} \text{ janvier } 2025\}$ (appelée saisonnalité omnisciente) et d'autre part $\{d, debut_d < 1^{er} \text{ janvier } 2025 \text{ et } recep_d < 1^{er} \text{ janvier } 2024\}$ (appelée saisonnalité prédite).

Dans la section précédente, plusieurs indicateurs ont été présentés, pour comparer ces deux saisonnalités, on utilisera le délai de réservation.

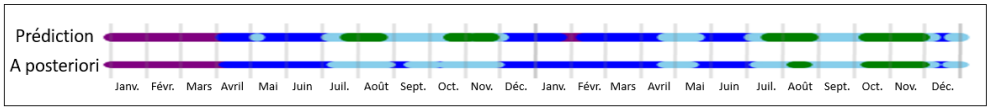

FIGURE 4.22 – Comparaison entre la saisonnalité prédite et la saisonnalité omnisciente

Sur la figure 4.22, la saisonnalité omnisciente et celle prédite sont très proches car beaucoup de jours ont la même couleur. On va vérifier cela grâce aux indices mesurant la proximité entre deux classifications.

Indice de Jaccard	Indice de Rand	Indice de Rand Ajusté
0.73	0.34	0.58

Les indices de Jaccard et de Rand ajusté sont très élevés, ce qui confirme la proximité entre la prédiction et la saisonnalité omnisciente. Ainsi, on obtient une méthode précise pour calculer la saisonnalité prédite, et qui pourra être utilisée par la suite pour améliorer la tarification dynamique.

4.7. Conclusion

Ce chapitre avait pour objectif d'analyser en profondeur les données issues de la location de véhicules afin de mieux comprendre les mécanismes de demande et de préparer les travaux sur la tarification dynamique. Plusieurs conclusions majeures apparaissent dans ce chapitre.

Les distributions des caractéristiques des demandes mettent en évidence des comportements de demande très différenciés selon le profil du loueur. La durée de location, le délai de réservation et la saisonnalité varient significativement selon que l'activité est orientée vers le tourisme, la clientèle professionnelle ou la location spécialisée de vans. Ces différences structurantes confirment que la demande dépend fortement du contexte opérationnel.

Ensuite, les analyses de corrélation montrent que certaines relations sont intuitives mais non triviales. Le taux d'utilisation maximal est fortement lié au délai de réservation, tandis que le revenu par jour (RPD) décroît légèrement avec l'augmentation de la durée. En revanche, les corrélations globales entre RPD et les autres variables restent limitées, ce qui suggère que le prix ne peut pas être expliqué par une seule dimension isolée.

De plus, la méthode de classification OClustR met en évidence que la structuration de la demande ne repose pas seulement sur un paramètre unique. Les classes obtenues montrent que les demandes se différencient selon des combinaisons de durée, d'anticipation et d'utilisation. Cette observation confirme que la demande possède une structure multidimensionnelle.

Enfin, l'étude de la saisonnalité, notamment via une classification K-means des jours, montre que les périodes de forte et de faible activité peuvent être

identifiées de manière quantitative à partir des caractéristiques observées. La saisonnalité n'est donc pas uniquement calendaire ; elle est détectable dans les données elles-mêmes.

Les résultats obtenus soulèvent plusieurs implications importantes pour la suite du travail. Tout d'abord, l'absence de corrélations simples et fortes entre le RPD et les autres variables confirme que la tarification ne peut pas reposer sur une règle simple. Le prix doit être envisagé comme le résultat d'interactions complexes entre durée, délai de réservation, utilisation et saisonnalité. Cela justifie pleinement l'introduction ultérieure de méthodes plus avancées, capables de capturer ces dépendances croisées. Ensuite, la mise en évidence de profils de demande distincts selon les types de loueurs montre que les politiques tarifaires ne peuvent pas être uniformes. Une stratégie efficace doit être adaptée au contexte opérationnel, à la disponibilité des véhicules et au comportement des concurrents.

Par ailleurs, la classification des jours selon leur comportement saisonnier ouvre la voie à une tarification conditionnée par cette classification. Identifier des périodes homogènes permet d'envisager des politiques différenciées entre haute et basse saison, mais aussi de détecter des transitions intermédiaires plus fines.

Chapitre 5
Méthodes de tarification dynamique

Sommaire

5.1 Introduction 73
5.2 Étude des deux méthodes 74
5.2.1 Méthode par ressources 74
5.2.1.1 Un algorithme par ressources : EMSR 74
5.2.1.2 L'adaptation à notre contexte 75
5.2.2 Méthodes par concurrence 76
5.2.2.1 Linear Reward 77
5.2.2.2 Une adaptation de la technique *Linear Reward* pour notre problème 79
5.2.3 Comparaison des méthodes de RL 80
5.2.3.1 Ensemble(s) de données 80
5.2.3.2 Paramètres mesurés et scénarios 81
5.2.3.3 Évaluation de performances 82
5.3 Méthode Mixte 88
5.3.1 Description de la méthode 88
5.3.2 Évaluation de performances de la méthode Mixte 91
5.4 Conclusion 97

5.1. Introduction

L'attention se porte maintenant sur les algorithmes employés par chaque joueur, loueur de voitures, dans le jeu répété défini dans le chapitre 3 afin de fixer un prix pour chaque nouvelle demande.

L'objectif de chaque joueur est donc de maximiser la somme de ses profits pour l'ensemble ordonné des demandes D. Dans la littérature et la pratique des différents domaines d'application de la tarification dynamique, deux grands types d'approche sont souvent utilisés, comme indiqué dans le Chapitre 2. D'une part, les approches fondées sur les ressources sont celles dans lesquelles les prix sont fixés en fonction de la quantité de ressources restantes pour les périodes demandées [2, 50]. D'autre part, les approches fondées sur la concurrence consistent à fixer les prix en fonction des prix de la concurrence observés dans le passé [6, 73]. Pour chacune de ces approches, une méthode sera définie à partir d'une méthode existante et sera adaptée à la location de voitures (Sous-Section 5.2). Ensuite, une nouvelle méthode sera proposée en intégrant ces deux logiques (Sous-Section 5.3.1).

5.2. Étude des deux méthodes

Le but de cette section est de montrer comment des méthodes existantes dans d'autres contextes que la location de voiture peuvent aussi être adaptées pour être performantes dans ce cadre. Le développement s'articule d'abord autour de la définition d'une méthode par ressources (Sous-Section 5.2.1), s'appuyant sur l'algorithme EMSR initialement conçu pour l'aérien (déjà décrit dans la Section 2.2.2). Dans un second temps (Sous-Section 5.2.2), l'étude se porte sur l'intégration de l'apprentissage par renforcement afin de modéliser une stratégie de tarification dynamique réactive aux prix de la concurrence.

5.2.1. Méthode par ressources

Une méthode par ressources utilise uniquement la quantité restante de produit à vendre pour fixer les prix, c'est-à-dire, dans notre cas, le nombre de voitures encore disponibles. Ainsi, une méthode par ressources ne prend absolument pas en compte les prix de la concurrence. Ce type de méthode vise une allocation optimale des ressources plutôt qu'un positionnement par rapport à ses concurrents. Pour cela, en plus du niveau d'utilisation actuel, la méthode va aussi anticiper les demandes futures à travers une connaissance des données historiques. Dans la suite, nous avons choisi d'utiliser la méthode EMSR provenant de l'aérien, définie dans la Section 5.2.1.1 puis nous détaillerons comment elle sera utilisée concrètement dans le contexte de la location de voitures, dans la Section 5.2.1.2.

5.2.1.1. Un algorithme par ressources : EMSR

Comme cela fut détaillé dans le chapitre 2, le but de la méthode EMSR [2] est l'optimisation de la combinaison des classes tarifaires dans l'aérien, qui consiste à déterminer le nombre de réservations acceptées pour chaque classe tarifaire partageant un inventaire commun de sièges. Pour déterminer le nombre de sièges à mettre à disposition pour chaque classe tarifaire, le système de gestion des revenus de la compagnie aérienne doit générer une prévision de la demande future de réservations par classe tarifaire pour chaque départ futur.

Pour EMSRa, chaque classe tarifaire a son propre stock de sièges dans l'avion, mais pour EMSRb, les sièges ne sont pas attribués à des classes de réservations distinctes, ils sont réservés pour les classes tarifaires supérieures et des limites de réservation sont appliquées aux classes tarifaires inférieures, c'est-à-dire que la limite de réservation pour une classe tarifaire est la somme de la limite de réservation de la classe tarifaire directement inférieure et de la limite de réservation allouée à cette classe tarifaire dans EMSRa. Par exemple, un départ futur a une capacité autorisée en classe économique de 120 sièges et 3 classes tarifaires (A, B et C par ordre décroissant de revenus) qui doivent se

partager cet inventaire. Pour EMSRa, une allocation partitionnée des classes de réservations pourrait être, par exemple, la suivante :

- 30 réservations pour la classe A,
- 40 pour la classe B,
- 50 pour la classe C.

Dans le cas d'EMSRb, la méthode oblige qu'il y ait au plus

- 120 réservations pour la classe A,
- 90 pour la classe B,
- 50 pour la classe C.

Si nous supposions que la demande future de réservations pour chaque classe tarifaire est déterministe, il serait relativement simple de calculer la répartition des classes tarifaires pour un vol futur. Nous commencerions par la classe tarifaire la plus élevée et protégerions exactement le nombre de sièges correspondant à la demande future pour cette classe, puis nous continuerions avec la classe tarifaire inférieure suivante jusqu'à atteindre la capacité autorisée. Dans la réalité, la demande future de réservations n'est pas déterministe. Toutefois, pour chaque classe tarifaire f_i, on peut définir une fonction en fonction de l'historique qui prend en entrée un nombre de réservations futures potentielles S_i et qui renvoie $Proba_i$ la probabilité d'observer S_i réservations pour la classe tarifaire f_i.

On définit alors pour chaque classe tarifaire f_i, le gain attendu par siège (*expected marginal seat revenue*) $EMSR_i$ comme la fonction :

$$EMSR_i(S_i) = f_i * Proba_i(S_i)$$

avec S_i le nombre de sièges réservés pour la classe tarifaire f_i.

On définit alors les limites de réservation dans le cadre d'EMSRa comme étant S^i_{i+1} telles que :

$$EMSR_i(S^i_{i+1}) = f_{i+1}$$

Cette équation équivaut à dire que S^i_{i+1} est la limite du nombre de réservations où le gain attendu par siège devient inférieur au gain attendu par siège pour la classe tarifaire du dessous.

5.2.1.2. L'adaptation à notre contexte

L'algorithme de tarification dynamique basé sur les ressources examiné ici est une adaptation, dans notre modèle de jeu répété, de l'heuristique Expected Marginal Seat Revenue (EMSR, voir [2]). Ainsi, d'après ce modèle, chaque agent A_α peut seulement pratiquer les prix suivants

$Pr_\alpha = \{pr_\alpha[1],\ pr_\alpha[2],\ \ldots,\ pr_\alpha[M_\alpha]\}$. De plus, il est considéré que chaque agent A_α connaît $Dem_C^{P'}(A_\alpha)$ (défini dans la partie 3.4.2) : l'ensemble des demandes qui ont fait l'objet d'un contrat avec cet agent au cours d'une séquence de périodes passées P' (par exemple, P' correspond à l'année précédente). Pour chacune de ces périodes $p' \in P'$, le sous-ensemble des demandes d de $Dem_C^{P'}(A_\alpha)$ tel que $debut_d \leq p' \leq fin_d$ est désigné par $Dem_C^{p'}(A_\alpha)$, et le sous-ensemble de celles dont le prix est égal à $pr_\alpha[i] \in Pr_\alpha$ est désigné par $Dem_C^{p'}(A_\alpha, pr_\alpha[i]) \subseteq Dem_C^{p'}(A_\alpha)$.

Pour chaque agent A_α, chaque prix $pr_\alpha[i] \in Pr_\alpha$ a un seuil définissant le nombre maximum de réservations pour chaque période p à ce prix $pr_\alpha[i]$. Ce seuil est défini par :

$$thd_{\alpha,pr_\alpha[i]} = \frac{1}{|P'|} \sum_{p' \in P'} \frac{|Dem_\alpha(p', pr_\alpha[i])|}{|Dem_\alpha(p')|} \tag{5.1}$$

Ce seuil permet de définir le niveau de disponibilité à partir duquel l'agent augmente son prix. On a donc besoin d'une mesure permettant de déterminer le nombre de ressources utilisées sur la période demandée. $RatioMax$ est défini comme l'utilisation maximale sur la période demandée par d au moment de la réception de la demande $recep_d$

$$RatioMax = \max_{p \in [debut_d, fin_d]} \mathcal{U}til_\alpha^{recep_d}(p) \tag{5.2}$$

Le prix choisi pour la demande d par A_α est le prix minimum tel que :

$$RatioMax \leq thd_{\alpha,pr} \tag{5.3}$$

C'est-à-dire le prix minimum tel que le niveau de remplissage est en dessous du seuil. Cette méthode donnera des prix de plus en plus élevés pour la même période au fur et à mesure que des réservations seront effectuées pour cette même période.

Les seuils nécessitent d'être actualisés pour pouvoir considérer l'historique récent. Le choix pour les simulations à suivre a été de les actualiser toutes les 100 demandes reçues pour avoir assez de nouvelles données, mais aussi de faire des mises à jour suffisamment souvent.

5.2.2. Méthodes par concurrence

Dans notre contexte, certains agents de location pourront utiliser un algorithme RL pour choisir leur politique tarifaire. Cependant, chaque étape d'apprentissage correspond à une étape du jeu répété, c'est-à-dire à une nouvelle demande et à de nouveaux prix proposés par les concurrents de l'agent RL. Dans notre contexte, il est impossible de faire une suite d'étapes d'apprentissage pour un même jeu unitaire, c'est-à-dire pour une même demande

de location, car il n'y a qu'une seule occasion de choisir un prix pour une demande de location. L'environnement, constitué de la demande et des prix des concurrents, est non stationnaire [30] car chaque agent adapte sa stratégie en fonction des stratégies des autres agents. Il existe plusieurs façons de classifier les méthodes d'apprentissage par renforcement : une distinction selon que l'apprentissage se fait par valeur ou par politique (Chapitre 2.3), mais la distinction peut aussi se faire selon l'utilisation d'états ou non. Les états caractérisent l'environnement, et le but est d'apprendre la meilleure action à adopter pour chaque état. Dans les méthodes par états, les actions induisent un changement d'état, dépendant de l'action choisie. Dans notre contexte, la non-stationnarité de l'environnement nécessiterait d'avoir beaucoup trop d'états pour que ce soit envisageable. Ainsi, les méthodes avec des états comme le Q-Learning [69] ne semblent pas adaptées dans ce contexte.

Nous avons choisi de nous tourner vers des méthodes sans modélisation explicite des états. Ce problème où à chaque étape, le joueur doit choisir un prix pour maximiser son bénéfice total correspond au problème du bandit manchot [39]. Ainsi, choisir un prix revient à choisir une machine et le but reste de maximiser le gain cumulé. Pour ce problème, plusieurs types de méthodes sont utilisés :

- Méthode par espérance de gain comme $\varepsilon - greedy$ (voir 2.3.1)
- Méthode par regret comme EXP3 (voir 2.3.2)

Nous avons aussi choisi d'utiliser des méthodes utilisant la technique de *Linear Reward*. Ce type de méthode repose sur une logique d'espérance de gain et présente une propriété intéressante en matière d'équilibre de Nash. En particulier, s'il existe un équilibre de Nash pour le jeu unitaire, alors il est possible de paramétrer ce type de méthode de sorte qu'elle converge vers cet équilibre. Dans la suite, cette méthode sera adaptée afin d'en améliorer les performances dans notre contexte d'étude.

Dans les paragraphes suivants, nous allons :

- présenter des méthodes utilisant la Linear Reward (Section 5.2.2.1) ;
- expliquer comment nous l'avons adaptée à notre contexte spécifique de location de voitures afin d'en améliorer la performance en termes de profit cumulé (Section 5.2.2.2).

5.2.2.1. Linear Reward

Les méthodes utilisant la *Linear Reward* [51] sont des méthodes d'apprentissage par renforcement avec choix probabiliste visant à privilégier progressivement les actions les plus efficaces. Pour cela, elle utilise un vecteur stochastique V, où chaque composante représente la probabilité d'exécuter une action donnée. Le but est d'obtenir une probabilité proche de 1 lorsqu'une action a une

espérance de gain élevée, et une probabilité proche de 0 lorsque l'action a une espérance de gain faible.

L'apprentissage débute avec un vecteur de probabilités initialisé de manière uniforme. À chaque itération, l'agent sélectionne une action de façon aléatoire, en suivant la distribution définie par ce vecteur. Ensuite, l'algorithme met à jour les probabilités en fonction du gain obtenu, renforçant ainsi les actions ayant généré une récompense positive.

Les formules de mise à jour des probabilités sont les suivantes avec des valeurs différentes pour b et β en fonction de la méthode :

- $V(a) \leftarrow V(a) + b \times U_t(a) \times (1 - V(a)) - \beta \times V(a) \times (1 - U_t(a))$ Si a est l'action choisi
- $V(a) \leftarrow V(a) - b \times U_t(a) \times V(a) + \beta \times ((taille(V) - 1)^{-1} - V(a)) \times (1 - U_t(a))$ Sinon

La fonction d'utilité $U_t(a)$ qui est utilisée classiquement est la suivante :

$$U_t(a) = \frac{Gain(a) - Gain_{Min}(a)}{Gain_{Max}(a) - Gain_{Min}(a)}$$

Avec $Gain_{Min}(a)$, respectivement $Gain_{Max}(a)$ le gain minimum, resp. maximum obtenu lorsque cette action a été jouée.

Tout ceci est résumé dans l'algorithme 4 qui explique le fonctionnement des méthodes utilisant la *Linear Reward*.

Algorithme 4 Linear Reward

Entrée : b ∈ [0,1], β ∈ [0,1] et un ensemble de demandes $\mathcal{D}$
Sortie : V un vecteur de probabilité de taille V

$V[i] \leftarrow 1/\text{taille(V)}$ pour tout i
pour tout chaque demande d **faire**
    Tirage d'une action a^* aléatoirement en respectant le vecteur de probabilité V
    Réception du gain noté $Gain(a^*)$ du système
    Mise à jour des $Gain_{Max}(a^*)$ et $Gain_{Min}(a^*)$
    $U_t(a^*) \leftarrow (Gain(a^*) - Gain_{Min}(a^*))/(Gain_{Max}(a^*) - Gain_{Min}(a^*)$
    $V(a^*) \leftarrow V(a^*) + b \times U_t(a^*) \times (1 - V(a^*)) - \beta \times V(a^*) \times (1 - U_t(a^*))$
    pour tout chaque action a différente de a^* **faire**
        $V(a) \leftarrow V(a) - b \times U_t(a^*) \times V(a) + \beta \times ((taille(V) - 1)^{-1} - V(a)) \times (1 - U_t(a^*))$
Retourne V

Le nom de la méthode change en fonction des valeurs des paramètres b et β :

- *Linear Reward Inaction* (LRI) Si $\beta = 0$
- *Linear Reward Penalty* (LRP) Si $\beta > 0$

Les deux versions LRI et LRP ont des comportements qui peuvent vraiment être différents l'un de l'autre, dus à la valeur de β qui permet d'apprendre même lorsqu'une action a a un gain proche de $Gain_{Min}(a)$. Dans le cas de la méthode LRP, une action choisie ne peut pas avoir sa probabilité qui diminue, alors que ça peut être le cas pour la méthode LRP. Pour une action qui n'a pas été choisie, sa probabilité ne peut pas augmenter avec la méthode LRI alors qu'avec la méthode LRP, elle peut augmenter et diminuer.

Nous verrons dans la partie Évaluations de performances 5.2.3.3 que la méthode LRP est très instable. En effet, la méthode LRP met à jour le vecteur de probabilité à la fois en renforçant les actions ayant généré de bonnes performances (via le calcul précédé par b) et en pénalisant celles jugées mauvaises (via le calcul précédé par β). Cette double mise à jour induit une grande variabilité du vecteur de probabilité au cours du temps, ce qui peut nuire à la stabilité de l'apprentissage dans un environnement non stationnaire tel que celui de la location de voitures.

5.2.2.2. Une adaptation de la technique *Linear Reward* pour notre problème

Dans cette section, nous présentons l'adaptation des méthodes utilisant la *Linear Reward* au contexte spécifique de la location de voitures. Dans notre méthode, les stratégies possibles de l'agent correspondent à son ensemble de prix Pr, associé à un vecteur de probabilité stochastique V. À chaque demande d, l'agent sélectionne un prix de manière aléatoire selon la distribution donnée par V. Ce vecteur est ensuite mis à jour comme défini précédemment en changeant la fonction d'utilité $U_t(a)$ pour s'adapter au contexte de la location de voitures. La fonction d'utilité $U_t(a)$ est définie comme suit :

$$U_t(a) = \begin{cases} \exp(\frac{-(Pr_{min}-a)^2}{100}) & \text{Si l'agent est choisi} \\ \exp(\frac{Pr_{min}-a}{10}) & \text{Sinon} \end{cases} \tag{5.4}$$

Avec Pr_{min}, le prix minimum fixé par les concurrents.

Cette fonction d'utilité favorise d'abord le gain de demande, puis la fonction d'utilité s'intéresse à la proximité du prix proposé par notre agent par rapport au minimum de ceux proposés par les concurrents.

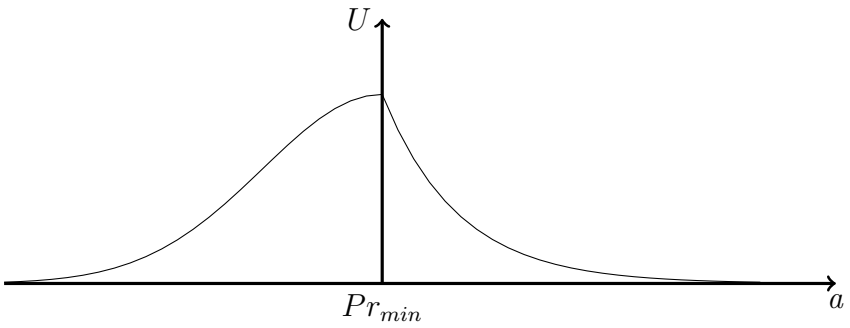

FIGURE 5.1 – Fonction d'utilité U

5.2.3. Comparaison des méthodes de RL

Dans cette sous-section, le but est d'étudier la performance des méthodes décrites précédemment à partir du bénéfice total obtenu sur un an de données de location de voitures. Pour commencer, dans la Sous-Section 5.2.3.1, nous définirons les jeux de données utilisés pour réaliser les évaluations de performances ; ensuite, dans la Sous-Section 5.2.3.2, nous décrirons les scénarios ainsi que les mesures utilisées pour appréhender la pertinence des méthodes, en particulier le bénéfice total obtenu par celles-ci. Pour finir dans la Sous-Section 5.2.3.3, nous étudierons les performances des méthodes selon les scénarios décrits précédemment.

5.2.3.1. Ensemble(s) de données

Pour effectuer les évaluations des performances des algorithmes définis dans les Sous-Sections 5.2.1.2 et 5.2.2.2, nous allons utiliser la séquence de demandes dont les prix ne sont pas liés à des contrats préalables (appelées demandes *yieldables*) d'un des clients de WeYield présents sur une île française (comme défini dans la Section 4.2). Son activité est fortement touristique, le nombre de demandes varie donc fortement au cours de l'année, comme nous avons pu le voir dans la Section 4.3. Dans ce jeu de données, il y a environ 12 000 demandes par an. On considérera que la séquence des demandes est triée par ordre croissant de dates de réservation $recep_d$ pour le jeu répété. Pour le jeu, nous utiliserons les demandes reçues en 2023, c'est-à-dire avec $recep_d$ en 2023, pour calculer la performance en matière de bénéfice total, mais leur période de location peut se situer indépendamment en 2023 et 2024. Dans le modèle défini dans le Chapitre 3, les demandes ont une valeur seuil $Pmax$ qui indique le prix maximum qu'un client veut bien payer. Comme ces données ne sont pas disponibles, nous avons choisi de fixer arbitrairement $Pmax$ en fonction de la valeur totale payée par le client en incluant les suppléments comme les assurances, le fait de ne pas avoir de limite kilométrique, etc. Cette valeur

correspond au montant total que le client a effectivement payé. La distribution des valeurs de $Pmax$ est la suivante :

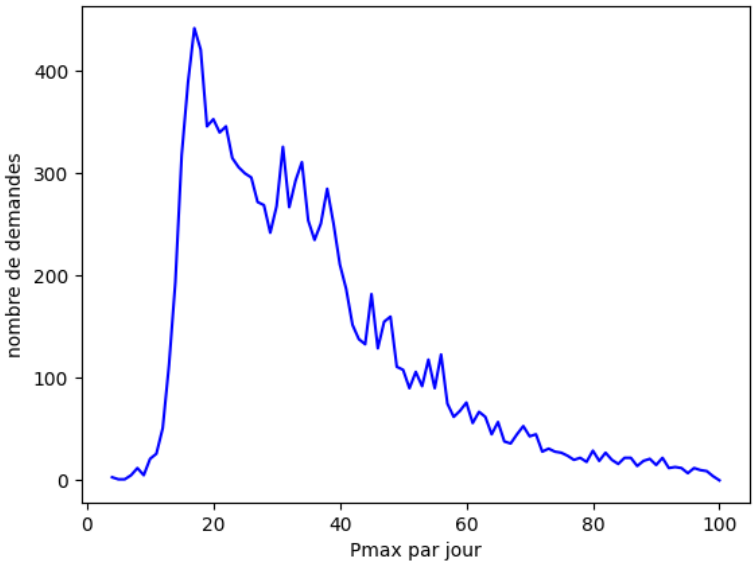

FIGURE 5.2 – Distribution des valeurs de $Pmax$

En comparant cette courbe avec celle de la distribution du RPD (Figure 4.5), on s'aperçoit que les valeurs de $Pmax$ sont nettement plus élevées que celles du RPD.

Les demandes reçues en 2022 seront considérées comme une année de démarrage (*bootstrapping*) pour les méthodes d'apprentissage et comme des données historiques pour l'EMSR. De plus, la méthode EMSR a besoin des données de prix pour les données historiques. Pour créer les prix des données historiques, les prix des données réelles seront arrondis pour appartenir à l'ensemble de prix Pr_α de l'agent utilisant la méthode EMSR. Les demandes de 2022 avec des retours de véhicules prévus en 2023 auront également un impact sur les capacités restantes des périodes de 2023. On les considère donc pour avoir la capacité réellement disponible en 2023.

5.2.3.2. Paramètres mesurés et scénarios

Le but de cette partie est de trouver la méthode de RL avec les meilleures performances lorsqu'elle est en concurrence avec la méthode EMSR. Les performances des différentes méthodes de RL seront évaluées grâce à l'évolution du bénéfice au cours du temps quand elles sont en concurrence avec la méthode EMSR, avec, pour commencer, une capacité pour chaque agent $\mathcal{C}apmax_\alpha(p) = 150$. Cette capacité équivaut à avoir 300 véhicules globalement

alors qu'il faut 600 véhicules pour pouvoir répondre à toutes les demandes. On étudiera ensuite comment le bénéfice total évolue en fonction de la capacité disponible des agents pour chaque méthode de RL contre la méthode EMSR.

Le bénéfice total moyen de chaque joueur concurrent sera calculé sur 50 exécutions du jeu correspondant à l'ensemble de données défini à la Section 5.2.3.1. Les courbes présentant le bénéfice total moyen présenteront aussi le bénéfice total minimum et le bénéfice total maximum calculés sur les 50 exécutions. Pour pouvoir étudier la disponibilité des agents, on utilisera une seule exécution. Les évaluations seront réalisées avec deux acteurs ayant le même ensemble de prix possibles par jour $P_r = \{15, 20, 25, 30, 35, 40, 45\}$, choisis en tenant compte des prix existants dans l'ensemble de données.

Les paramètres d'entrée pour les méthodes LRI et LRP sont définis expérimentalement comme suit : $b = 0,001$, $v_{min} = 0,001$ et $\beta = 0$ pour LRI ou $\beta = 0,001$ pour LRP. La valeur de b a été choisie expérimentalement entre 0,0001 et 0,1 en prenant la valeur offrant le meilleur bénéfice total. Ensuite, la valeur de v_{min} a été choisie en fonction de celle de b. Pour la méthode $\varepsilon - greedy$, ε vaudra 0,5 pour les 100 premières demandes pour aider l'exploration et ensuite 0,001 pour être concentré sur l'exploitation. Pour la méthode Exp3, γ vaudra 0,001 avec $v_{min} = 0,001$.

Pour les méthodes LRI, LRP et Exp3, v_{min} permet de garder l'exploration de toutes les actions et ainsi l'algorithme peut se mettre à favoriser des actions qui ne l'étaient pas dans le passé. Cela permet à l'apprentissage de s'adapter dans le cas où les concurrents changeraient fortement de stratégie.

L'ensemble des résultats de ce chapitre a été présenté lors de la conférence *The 2025 Learning and Intelligent Optimization Conference* (LION19) et dans l'article [28].

5.2.3.3. Évaluation de performances

Cette section commencera par étudier comment se comportent les différentes méthodes d'apprentissage par renforcement lorsqu'elles sont en concurrence avec la méthode EMSR et pour $\mathcal{C}apmax = 150$ pour les deux agents dans chaque scénario. La première figure 5.3 montre l'évolution du bénéfice pour la méthode LRI, la seconde figure 5.4 correspond à la méthode LRP. La méthode LRP obtient moins de bénéfice que la méthode LRI. En plus, la valeur du bénéfice varie plus d'une exécution à l'autre dans le cas de la méthode LRP, ce qui est montré par la taille de l'aire bleue qui est plus conséquente.

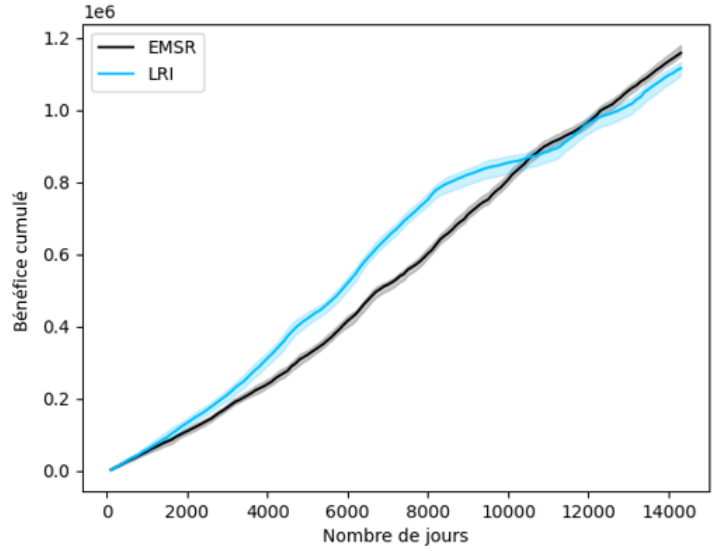

FIGURE 5.3 – LRI contre EMSR

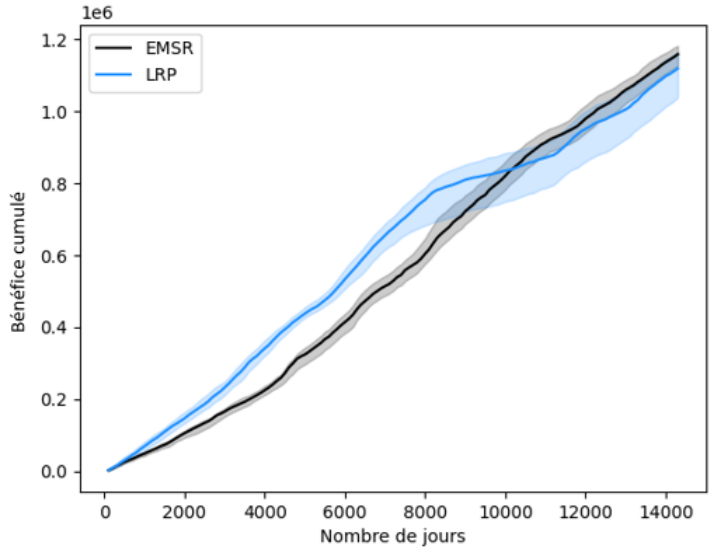

FIGURE 5.4 – LRP contre EMSR

Les figures 5.5 et 5.6 montrent un bénéfice pour les méthodes Exp3 et $\varepsilon - greedy$ nettement inférieur à celui de LRI dans les mêmes conditions.

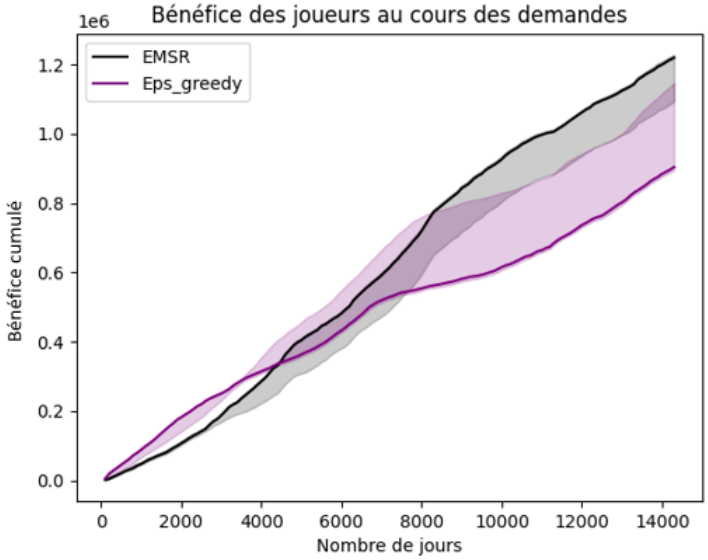

FIGURE 5.5 – EXP3 contre EMSR

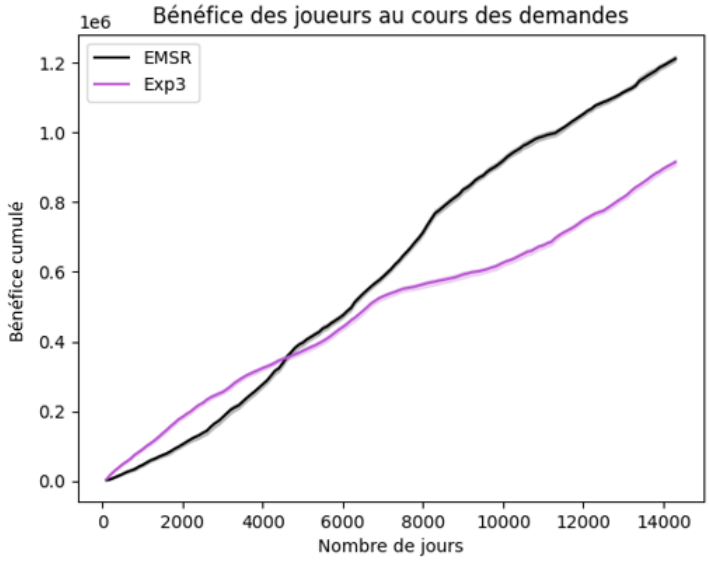

FIGURE 5.6 – $\varepsilon - greedy$ contre EMSR

Le but de la suite est de vérifier si le bénéfice de la méthode LRI est toujours meilleur que celui des autres méthodes, quelle que soit la capacité $\mathcal{C}apmax$ des agents. On comparera ces quatre méthodes à l'aide des figures 5.7, 5.8, 5.9 et 5.10. Les trois dernières figures ont leurs courbes de bénéfices moyens similaires même si celle représentant le bénéfice de LRP montre plus de variabilité avec une aire entre le bénéfice minimum et le bénéfice maximum

plus importante. Le bénéfice moyen du LRI est meilleur que celui des autres méthodes jusqu'à la capacité de 300 véhicules. Les situations où il y a plus de 300 véhicules correspondent aux cas où les loueurs ont trop de véhicules en cumulé par rapport au nombre de demandes. Ces situations sont plus faciles à optimiser car le but sera d'obtenir un maximum de demandes pour utiliser tous les véhicules. Ainsi, pour la suite de la thèse, on va seulement garder la méthode LRI comme méthode d'apprentissage par renforcement.

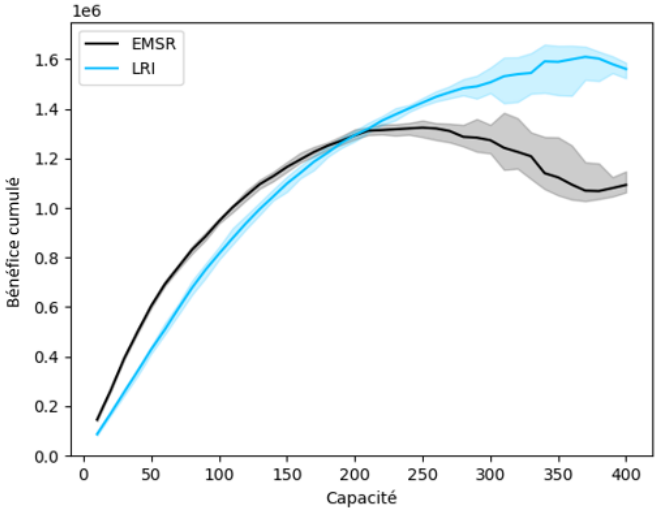

FIGURE 5.7 – Évolution du bénéfice pour LRI contre EMSR en fonction de la capacité

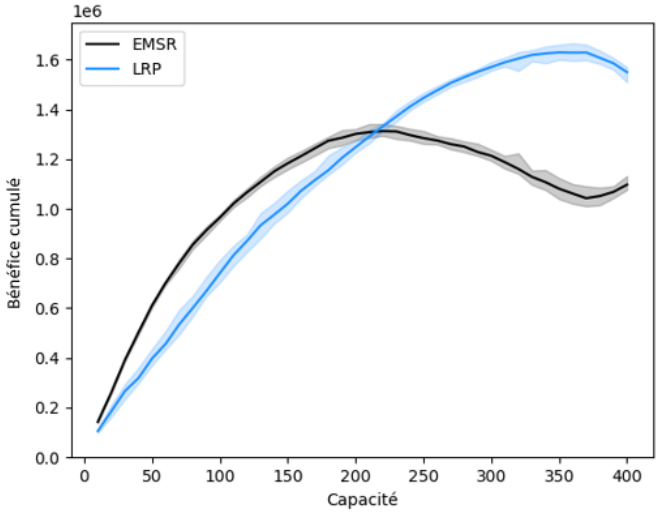

FIGURE 5.8 – Évolution du bénéfice pour LRP contre EMSR en fonction de la capacité

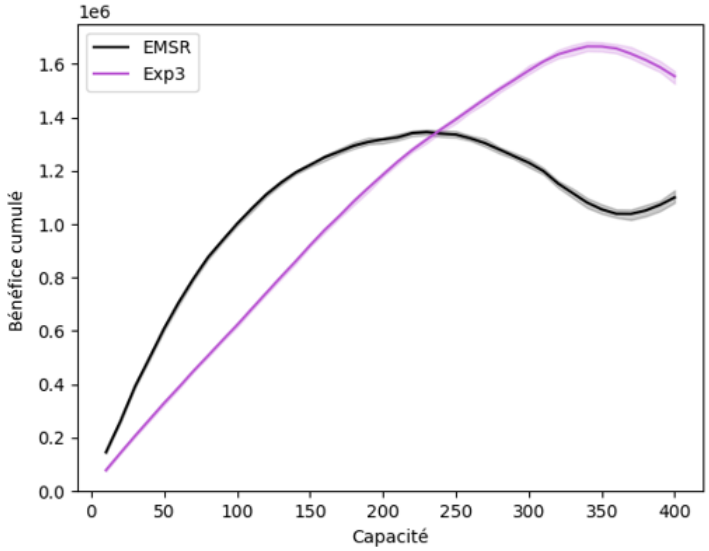

FIGURE 5.9 – Évolution du bénéfice pour EXP3 contre EMSR en fonction de la capacité

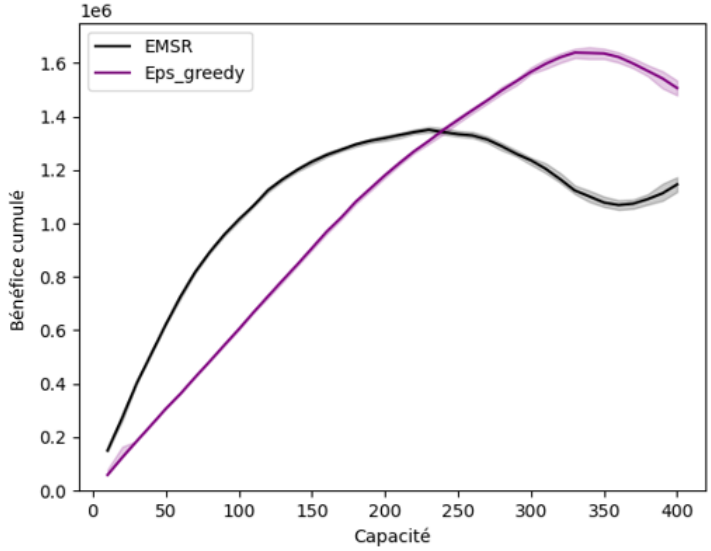

FIGURE 5.10 – Évolution du bénéfice pour $\varepsilon - greedy$ contre EMSR en fonction de la capacité

Le comportement de la méthode LRI est maintenant étudié plus en détail grâce à la figure 5.3. La méthode LRI obtient de meilleures performances depuis la première demande jusqu'à environ la $10\,000^{\text{ème}}$ demande, après quoi les performances s'inversent. Cette transition est une conséquence du phénomène

de saisonnalité défini en Section 4. En effet, après environ 8 000 demandes, la capacité disponible du joueur LRI est nulle pour la période de réservation des vacances d'automne, contrairement à celle du joueur EMSR. Comme des demandes de réservation pour cette période apparaissent entre la 8 000ème et la 10 000ème demande, la méthode EMSR est alors la seule à accepter ces réservations.

Ce phénomène est illustré dans la Figure 5.11, qui représente le taux d'indisponibilité pour chaque séquence de 100 demandes. Autour de la 8 000e demande, le nombre de refus dus à des contraintes de capacité augmente fortement pour la méthode LRI, tandis qu'il demeure stable pour la méthode EMSR jusqu'aux environs de la 10 000e demande.

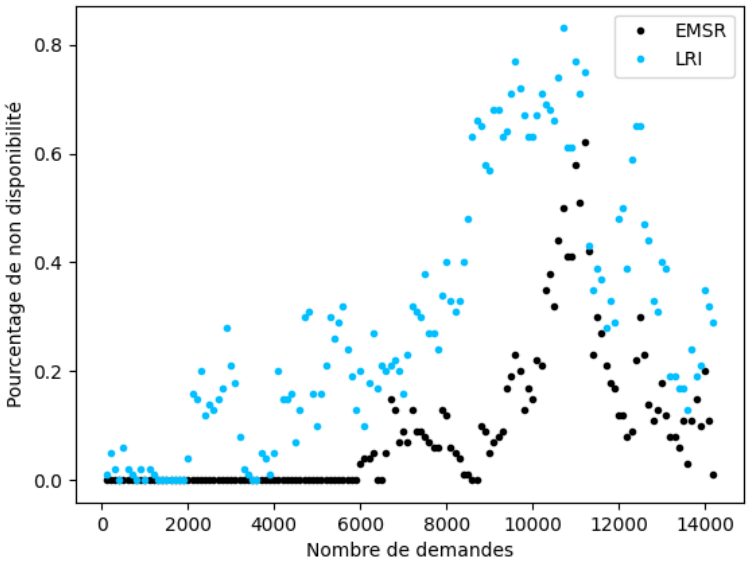

FIGURE 5.11 – Ratio d'indisponibilités pour chaque agent

La courbe de la Figure 5.12 où deux méthodes LRI sont en concurrence peut être rapprochée de celle correspondant à un équilibre en ligne, présentée dans la Figure 5.13. Cet équilibre est construit en considérant une succession d'équilibres de Nash unitaires, comme décrit dans la preuve du théorème V, ce qui constitue une manière simple de construire un équilibre séquentiellement stable. Les deux courbes apparaissent relativement proches, ce qui peut s'expliquer par le fait que la méthode d'apprentissage par renforcement présente des propriétés de convergence vers des équilibres de Nash, lesquels correspondent eux-mêmes à des équilibres séquentiellement stables.

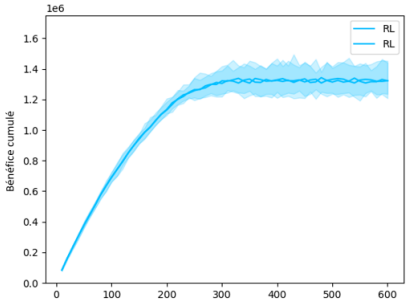

FIGURE 5.12 – Évolution du bénéfice pour LRI contre LRI en fonction de la capacité

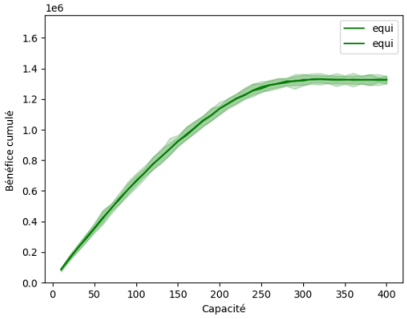

FIGURE 5.13 – Évolution du bénéfice en situation d'équilibre en fonction de la capacité

Synthèse :
Il a été vu dans cette partie que la méthode LRI obtient un meilleur bénéfice que celui des autres méthodes d'apprentissage par renforcement. EMSR est très intéressante pour maximiser le bénéfice lorsque la capacité est faible, alors que LRI est plus pertinente dans le cas inverse. La valeur limite correspond à la situation où les deux agents ne peuvent répondre qu'à 80% des demandes.

5.3. Méthode Mixte

La section précédente a présenté l'étude d'une approche par ressources et de plusieurs approches par concurrence. La suite proposera de combiner ces deux approches dans une seule méthode qui considérera la disponibilité des ressources ainsi que les prix des concurrents pour fixer les prix. Nous allons commencer par définir cette méthode dans la Sous-Section 5.3.1 puis nous évaluerons sa pertinence en la mettant en compétition avec les méthodes vues précédemment dans la Section 5.2.

5.3.1. Description de la méthode

Nous définissons ici une méthode de tarification qui tente de combiner les principes d'une approche basée sur les ressources et d'une approche basée sur la concurrence, telles que décrites dans les Sections 5.2.1 et 5.2.2. Pour combiner ces deux approches, nous allons créer trois états qui caractériseront la disponibilité des ressources et ensuite pour chacun de ces états, nous utiliserons une méthode par concurrence. De manière plus précise, à chaque réception d'une demande d, chaque agent A_α va calculer deux métriques décrivant la disponibilité des ressources. Elles dépendent de la séquence des capacités actuelles $\mathcal{C}ap_\alpha^{recep_d}(p)$ des périodes p entre $debut_d$ et fin_d, cela permet de considérer la disponibilité des ressources d'une manière bidimensionnelle (en temps et en

quantité). Ces mesures nécessitent un intervalle suffisamment large pour donner une idée de l'état du système, nous les examinerons donc au moins sur la durée moyenne des demandes de l'année précédente. Cette période sera centrée sur la demande. Ces deux mesures sont définies comme suit :

$$AireLibre = \sum_{p=debut_d}^{fin_d} \mathcal{C}ap_\alpha^{recep_d}(p) \tag{5.5}$$

$$AireTotale = \sum_{p=debut_d}^{fin_d} \mathcal{C}apmax_\alpha(p) \tag{5.6}$$

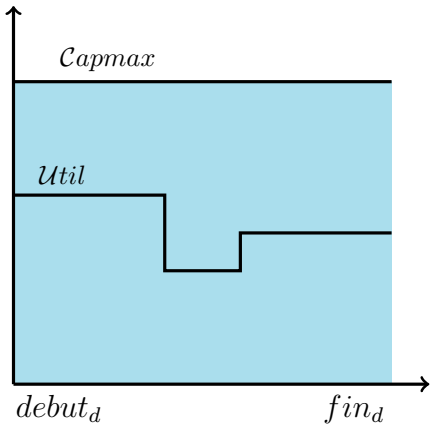

FIGURE 5.14 – Aire Totale

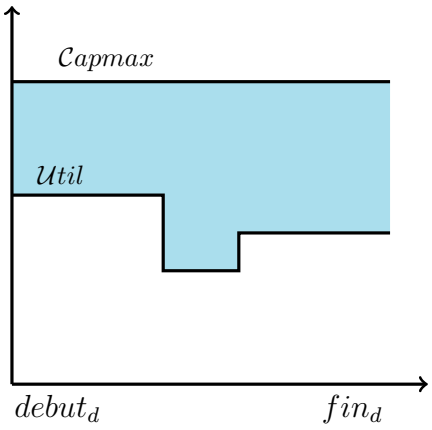

FIGURE 5.15 – Aire Libre

$$\textbf{La pression } \rho_\alpha(d) = \frac{AireLibre}{AireTotale} \tag{5.7}$$

La pression représente la capacité moyenne disponible sur la séquence de périodes affectée par la demande. Ainsi, la pression indique l'état de remplissage du système.

$$AirePic = \sum_{p=debut_d}^{fin_d} |\mathcal{C}apmax_\alpha(p) - Pic_\alpha(d)|^+ \tag{5.8}$$

$$\text{avec } Pic_\alpha(d) = \max_{debut_d \leq p \leq fin_d} \mathcal{U}til_\alpha^{recep_d}(p) \tag{5.9}$$

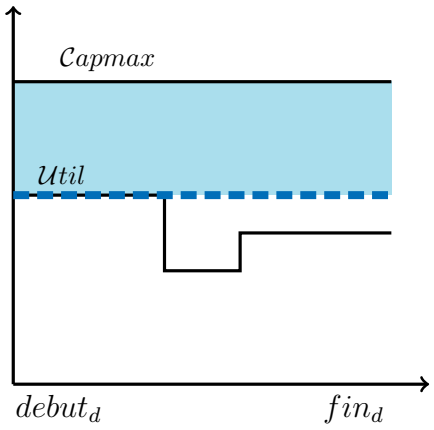

FIGURE 5.16 – Aire Pic

$$\textbf{La nuisance } \eta_\alpha(d) = \frac{AirePic}{AireLibre} \tag{5.10}$$

La nuisance mesure l'impact de la période la moins disponible sur cette séquence. Ainsi, la nuisance indique la présence ou l'absence de pic qui nuisent à un remplissage optimal.

Ces deux paramètres caractérisent la capacité disponible $\mathcal{C}ap_\alpha^{recep_d}(p) = \mathcal{C}apmax_\alpha(p)$ - $\mathcal{U}til_\alpha^{recep_d}(p)$ par rapport à la capacité totale $\mathcal{C}apmax_\alpha(p)$ (définies dans la partie 3.3.1) entre le temps $debut_d$ et fin_d, comme illustré par les Figures 5.14, 5.15 et 5.16. Ces deux nouvelles mesures fourniront des états caractérisant la disponibilité des ressources. Nous avons décidé d'avoir trois états d'apprentissage paramétrés par τ_ρ et τ_η, compris entre 0 et 1, comme le montre la Figure 5.17. Ce choix s'explique car lorsque la pression est proche de 1, donc une utilisation faible des véhicules, la nuisance a peu d'importance alors qu'elle devient importante pour une utilisation importante des véhicules. Chaque demande sera alors représentée par un couple de valeurs : la pression ρ et la nuisance η calculées au moment où la demande a été reçue, ce qui permettra de savoir quel est l'état d'apprentissage qui sera utilisé pour fixer le prix pour cette demande. L'état 1 correspond à une grande disponibilité, tandis que les états 2 et 3 correspondent à une utilisation plus importante des ressources. La différence entre ces deux états est que dans l'état 2, l'utilisation des ressources est relativement uniforme, alors que dans l'état 3, il y a d'importants pics d'utilisation.

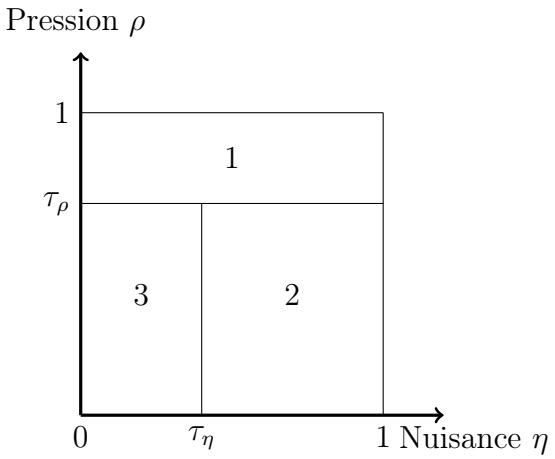

FIGURE 5.17 – Diagramme des états en fonction de la valeur de nuisance et de la pression

Les états décrits précédemment sont différents de ceux considérés habituellement dans les méthodes d'apprentissage par renforcement par états comme le Q-Learning car ce n'est pas l'action jouée qui permet de changer d'état, c'est le niveau de disponibilité des ressources pour la période de location demandée.

Pour chacun de ces trois états, un algorithme LRI défini dans la Section 5.2.2.2 est exécuté en fonction de la disponibilité des ressources.

5.3.2. Évaluation de performances de la méthode Mixte

Cette évaluation de performance sera réalisée dans les mêmes conditions que celle de la Section 5.2.3.3. Les paramètres utilisés pour la méthode Mixte sont $b = 0.005$, $v_{min} = 0.005$, $\tau_p = 0.9$ et $\tau_n = 0.5$ choisis expérimentalement pour avoir le meilleur bénéfice total en concurrence avec la méthode EMSR.

Afin d'évaluer l'efficacité de la méthode Mixte proposée, elle sera d'abord mise en concurrence avec un agent qui utilise la méthode d'apprentissage par renforcement (LRI) (définie dans la Section 5.2.2). Comme l'illustrent les Figures 5.18 et 5.19, les deux méthodes présentent un profit total quasi identique, ce qui s'explique par le fait qu'elles reposent sur des algorithmes de nature similaire. Dans cette configuration, les vecteurs d'apprentissage des deux méthodes évoluent de manière comparable au cours du temps. En analysant les résultats, la différence de gains observée entre les deux méthodes provient du fait que la méthode Mixte va proposer des niveaux de prix plus élevés dans les états 2 et 3. La comparaison entre la Figure 5.19 et la Figure 5.13 montre que le bénéfice cumulé des méthodes LRI et Mixte est meilleur que le bénéfice cumulé en situation d'équilibre séquentiellement stable (défini dans le Chapitre 3.

Lorsque la méthode Mixte est mise en concurrence avec la méthode EMSR, les Figures 5.20 et 5.21 montrent que la méthode Mixte obtient un meilleur

bénéfice total qu'EMSR. Cette amélioration s'explique par le fait que la méthode Mixte propose des prix qui restent majoritairement inférieurs ou égaux à ceux de l'agent EMSR, mais contrairement à la méthode de LRI, la méthode Mixte remporte les demandes à des niveaux de prix plus élevés que la méthode LRI dans les états 2 et 3, correspondants à des configurations où une partie de la capacité de l'agent est déjà utilisée. La comparaison entre la Figure 5.21 et la Figure 5.13 montre que le bénéfice cumulé des méthodes EMSR et Mixte est meilleur que le bénéfice cumulé en situation d'équilibre séquentiellement stable.

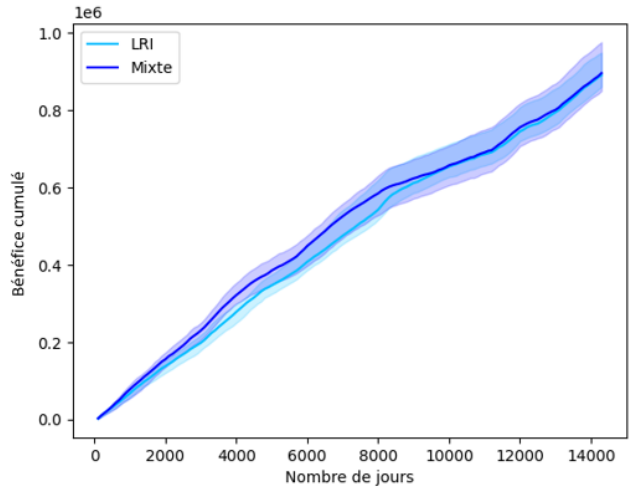

FIGURE 5.18 – LRI contre méthode Mixte

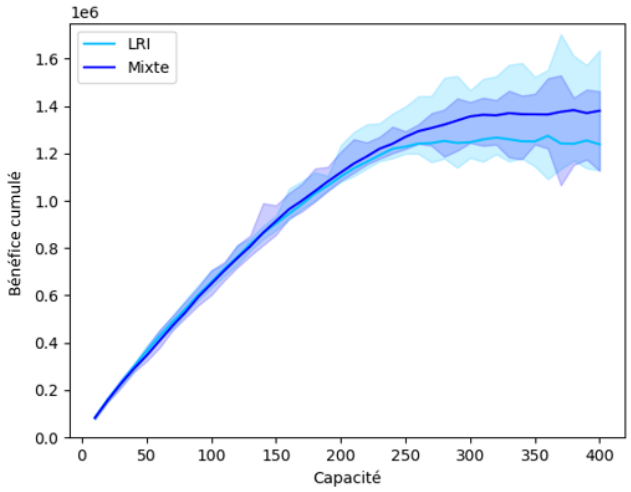

FIGURE 5.19 – Évolution du bénéfice pour LRI contre la méthode Mixte en fonction de la capacité

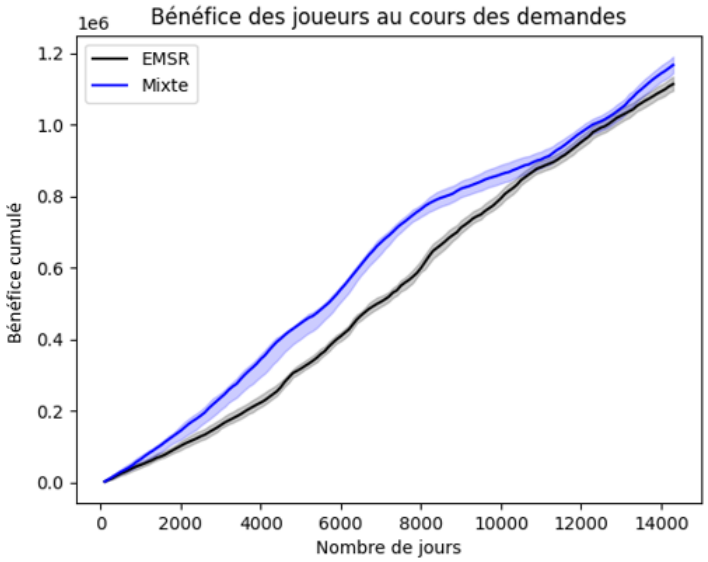

FIGURE 5.20 – Méthode Mixte contre EMSR

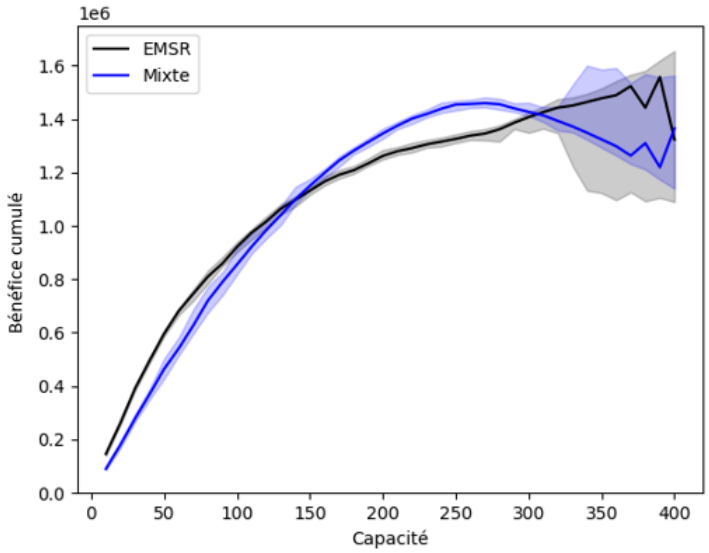

FIGURE 5.21 – Évolution du bénéfice pour EMSR contre la méthode Mixte en fonction de la capacité

Enfin, il convient de souligner que, dans la pratique, les agences de location opèrent rarement dans des régimes où la capacité totale disponible est inférieure à 33 % du volume des demandes entrantes (ici, 100 unités de capacité par agent). De plus, avoir 300 unités de capacité par agent indique que des véhicules ne seront pas utilisés.

Pour la méthode Mixte, les vecteurs d'apprentissage associés aux états 1 et 2 convergent vers des vecteurs unitaires, tandis que celui associé à l'état 3 ne présente pas le même comportement (voir Figure 5.22). Dans la Figure 5.22, la partie gauche de la ligne verticale correspond à l'année de préapprentissage ; le but est d'illustrer la vitesse de convergence du processus d'apprentissage. Dans les états 1 et 2, l'algorithme stabilise immédiatement vers un vecteur unitaire, la Figure 5.22 montre que l'apprentissage est plus lent dans l'état 3. Cela n'empêche toutefois pas d'avoir les trois prix médians avec des probabilités élevées et les autres prix avec des probabilités proches de v_{min}. Lorsque la demande est dans l'état 1 de la méthode Mixte, l'agent EMSR a, quant à lui, ses capacités qui ne sont pas entièrement utilisées, ce qui implique que celui-ci propose uniquement des prix de 15 et 20. La méthode Mixte apprend alors à sélectionner le prix de 15 afin de capter l'ensemble des demandes, ce qui amène le vecteur d'apprentissage à être proche d'un vecteur unitaire. Pour les demandes relevant de l'état 2, les capacités de l'agent EMSR sont moins disponibles, mais restent encore faiblement contraintes. La méthode Mixte apprend dans ce cas à proposer le prix de 20, car celui-ci est inférieur au prix moyen, égal à 24,6, sélectionné par EMSR, et parce que peu de prix égaux à 15 sont observés.

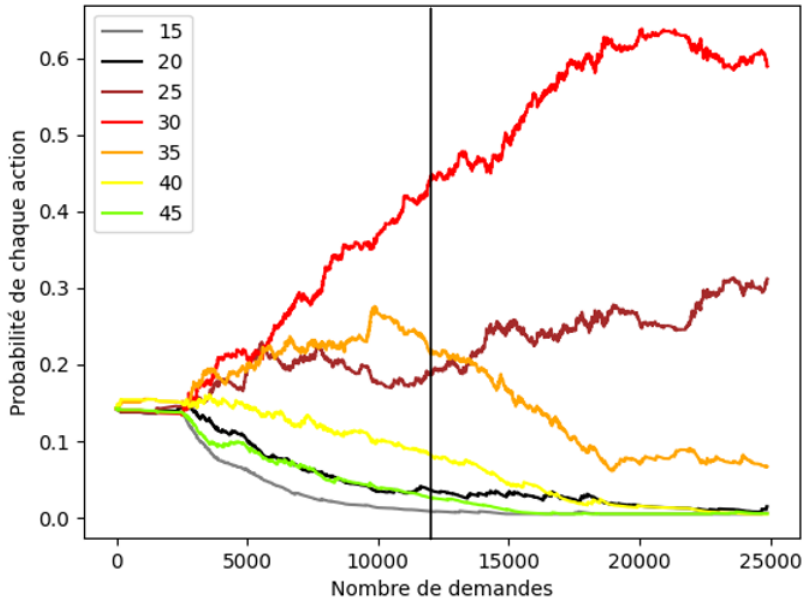

FIGURE 5.22 – Évolution du vecteur de probabilité dans l'état 3

Le comportement est sensiblement différent dans l'état 3, comme le montre la Figure 5.22, car les prix choisis par EMSR sont alors beaucoup plus variables et globalement plus élevés, du fait d'une utilisation plus avancée de ses capacités. Globalement, l'utilisation de ces trois vecteurs d'apprentissage distincts, un

par état, dont les dynamiques d'apprentissage diffèrent, permet à la méthode Mixte d'avoir un meilleur bénéfice total.

La Figure 5.23 montre que très peu de demandes relèvent de l'état 1 compte tenu des valeurs des paramètres τ_ρ et τ_η, mais qu'il est néanmoins essentiel pour disposer de vecteurs d'apprentissage capables de privilégier des niveaux de prix plus élevés dans les états 2 et 3. La méthode Mixte s'avère ainsi plus profitable que les méthodes purement concurrentielles ou purement fondées sur la disponibilité des ressources lorsqu'elles sont mises en compétition.

Par ailleurs, la Figure 5.23 montre que les demandes qui sont reçues pour une même journée concernent des situations différentes et ont donc des périodes de location différentes.

Ainsi, les courbes correspondant aux trois états augmentent en parallèle et non de manière successive. Les Figures 5.23 et 5.24 mettent en évidence que, lorsque la méthode Mixte est opposée dans un cas à la méthode LRI et dans l'autre cas à EMSR, le remplissage des capacités ne s'effectue pas de la même manière, avec un nombre significativement plus élevé de demandes relevant de l'état 2 lorsqu'elle est opposée à la méthode de LRI. Cet état correspond à un régime dans lequel le système est fortement contraint lorsque la méthode Mixte est confrontée à EMSR.

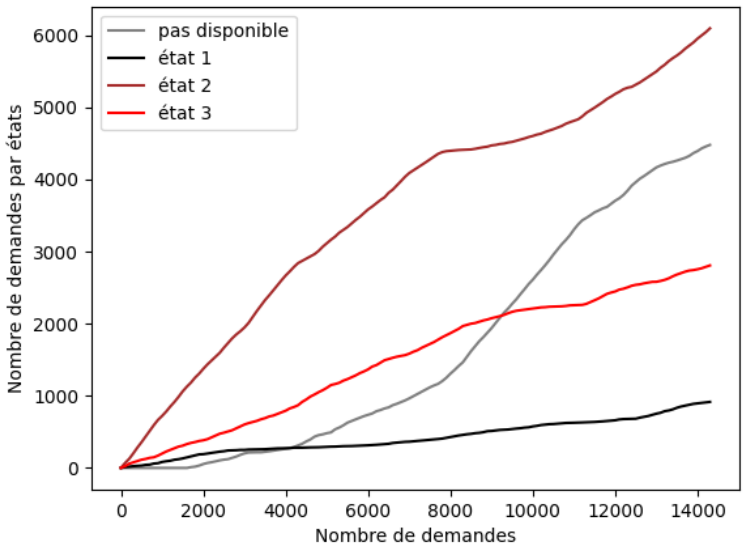

FIGURE 5.23 – Évolution du nombre de demandes par états pour la méthode Mixte contre EMSR

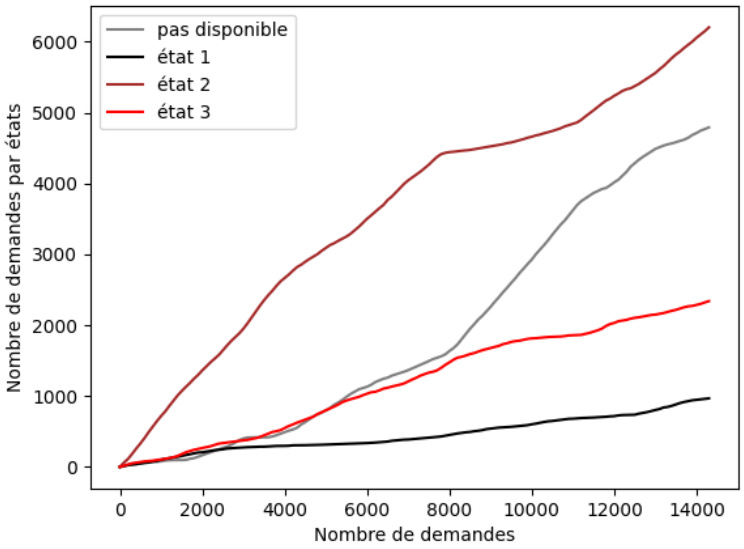

FIGURE 5.24 – Évolution du nombre de demandes par états pour la méthode Mixte contre LRI

Ce changement de comportement s'explique par la nature de la méthode du concurrent : d'une part, une méthode fondée sur la disponibilité des ressources, et d'autre part, une méthode utilisant les prix de la concurrence. Les méthodes concurrentielle et Mixte ont toutes deux tendance à diminuer fortement leurs prix afin d'obtenir la demande. Dans le scénario présenté dans la Figure 5.24, les deux méthodes ont la même stratégie en réduisant leurs prix, ce qui permet de louer environ deux fois moins de demandes, et donc d'avoir une proportion plus élevée de demandes à faible pression. Étant donné que le paramètre τ_ρ est relativement élevé, certaines de ces demandes à faible pression relèvent alors de l'état 2.

La Figure 5.25 montre enfin qu'un nombre important de demandes présentent une pression d'environ 0,65 pour EMSR (en noir sur la gauche de la figure). Ce phénomène s'explique par le fait que, pour ces demandes, le prix EMSR est égal à 25, tandis que la méthode Mixte se trouve encore dans les états 1 ou 2, proposant des prix de 15 ou 20. Dans ce cas, EMSR perd la demande, et cette situation persiste pour les demandes suivantes jusqu'à ce que la méthode Mixte bascule dans l'état 3.

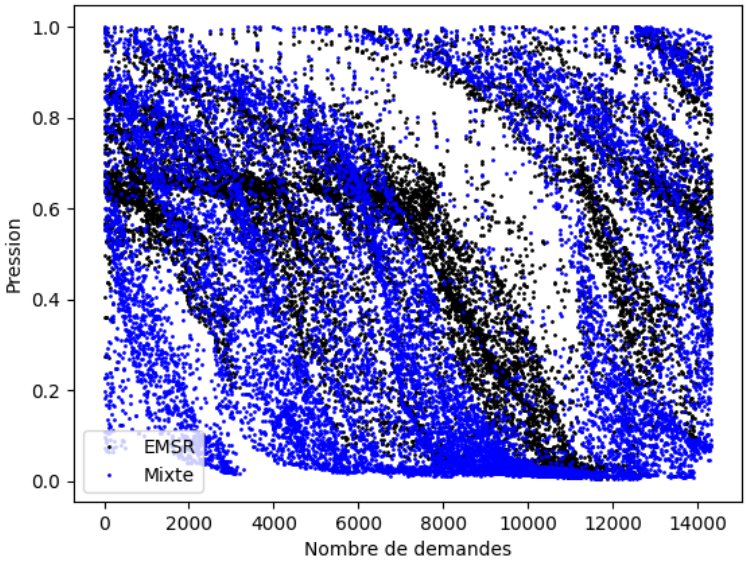

FIGURE 5.25 – Valeur de la pression pour chaque demande, pour chaque agent

Ces expériences montrent que la méthode Mixte est capable de rivaliser efficacement à la fois avec une méthode fondée sur la concurrence par les prix, telle que la méthode LRI, et avec une méthode fondée sur la gestion des ressources, telle qu'EMSR. Des expériences analogues ont été menées sur les données d'autres entreprises de location et ont conduit à des résultats similaires, favorables à la méthode Mixte (Annexe B).

Enfin, l'efficacité de la méthode Mixte dépend fortement des valeurs des paramètres τ_ρ et τ_η, fixées expérimentalement dans le cadre de cette étude.

Ce chapitre montre aussi que, quelles que soient les méthodes mises en concurrence, le bénéfice cumulé des deux méthodes est toujours supérieur au bénéfice cumulé de l'équilibre séquentiellement stable (défini dans le Chapitre 3).

5.4. Conclusion

Ce chapitre a exploré différentes méthodes de tarification dynamique dans un contexte de jeu répété appliqué à la location de véhicules, en comparant des méthodes fondées sur la gestion des ressources, des méthodes fondées sur la concurrence par les prix et une méthode Mixte combinant ces deux logiques. L'étude a montré que les méthodes par ressources, illustrées par EMSR, offrent des garanties de stabilité et de robustesse lorsque la capacité constitue la contrainte principale, mais restent limitées dans leur capacité à s'adapter à des environnements concurrentiels dynamiques. À l'inverse, les méthodes d'ap-

prentissage par renforcement basées sur la concurrence, telles que la méthode RL étudiée, sont capables de capter la demande en faisant des prix bas, mais au prix d'une consommation rapide de la capacité et d'une sensibilité accrue aux phénomènes de saisonnalité (définis en Section 4).

La méthode Mixte proposée dans ce chapitre apparaît comme une réponse à ces limites. En distinguant plusieurs états caractérisant la disponibilité des ressources, elle permet d'adapter le comportement de tarification à l'état du système. Les résultats expérimentaux montrent que cette méthode est capable de rivaliser efficacement avec une méthode purement concurrentielle comme avec une méthode purement fondée sur la disponibilité des ressources, et qu'elle surpasse ces deux méthodes dans une large gamme de configurations réalistes. Le recours à plusieurs vecteurs d'apprentissage, chacun associé à un état donné, constitue un élément clé expliquant ces performances.

D'un point de vue économique, les expériences mettent en évidence que le fait de systématiquement proposer des prix bas, bien qu'efficace à court terme pour capter la demande, peut conduire à des pertes de valeur lorsque la capacité devient contraignante. À l'inverse, une gestion trop conservatrice des ressources peut empêcher l'exploitation d'opportunités. La méthode Mixte permet de trouver un compromis dynamique entre ces deux extrêmes, en adaptant le niveau de prix non seulement à la demande observée, mais également à la disponibilité des ressources.

Enfin, ce travail ouvre plusieurs perspectives. Les performances de la méthode Mixte dépendent fortement des paramètres définissant les seuils de pression et de nuisance, ce qui pose la question de leur calibration automatique. Ce chapitre suggère que les méthodes hybrides, combinant apprentissage et structure économique explicite, constituent une voie prometteuse pour la tarification dynamique en environnement concurrentiel.

Chapitre 6
Méthodes de tarification avec Prédiction

Sommaire

6.1 Introduction 99
6.2 Utilisation de la saisonnalité dans les méthodes 100
6.2.1 Choix de la saison liée à la demande 101
6.2.2 Phase de tarification et d'apprentissage 102
6.2.3 Mise à jour de la saisonnalité 102
6.3 Évaluation des performances des méthodes saisonnalisées 103
6.3.1 Choix de la méthode de calcul de la saisonnalité 103
6.3.2 Paramètres des méthodes et scénarios 106
6.3.3 Évaluation des performances contre les méthodes sans saisonnalité 107
6.3.4 Évaluation des performances contre les méthodes saisonnalisées 116
6.3.5 Étude du bénéfice pour une méthode saisonnalisée avec une méthode différente par saison 121
6.4 Conclusion 129

6.1. Introduction

Le chapitre 4 a défini le concept de saisonnalité comme étant une classification des jours dont les classes seront appelées saisons. Chaque jour correspond à une date précise, comprenant le jour, le mois et l'année, et il est caractérisé par un ensemble de demandes qui incluent ce jour. Plusieurs métriques peuvent être utilisées pour caractériser cet ensemble de demandes, comme le délai de réservation moyen, le taux d'utilisation moyen au moment de la réservation, la durée moyenne de location et le prix moyen par jour, appelé RPD. Ainsi, chaque jour est caractérisé par l'ensemble de ces métriques moyennées, ce qui permet de calculer une classification des jours selon une ou plusieurs de ces métriques.

Ce chapitre proposera une manière d'utiliser cette saisonnalité pour faire de la tarification dynamique. Dans un second temps, le but sera de savoir quelles sont les métriques qui apportent le meilleur bénéfice total lorsqu'elles sont utilisées pour calculer la classification. L'étude s'intéressera au comportement des méthodes du chapitre 5 lorsqu'elles sont en concurrence avec des méthodes qui utilisent la saisonnalité.

La section 6.2 proposera une manière de prendre en compte la saisonnalité avec les méthodes du chapitre 5. Ensuite, la section 6.3 étudiera comment le bénéfice total est affecté par l'utilisation de la saisonnalité dans différents contextes de concurrence.

6.2. Utilisation de la saisonnalité dans les méthodes

Le but de cette section est de proposer une manière d'utiliser la saisonnalité pour faire de la tarification dynamique. Cette saisonnalité calculée au temps t est réalisée sur les jours compris entre $t - 365$ et $t + 365$ en considérant seulement les demandes reçues avant l'étape t, bien sûr. Cette saisonnalité permet d'avoir des stratégies de prix différentes en fonction des saisons car les jours sont regroupés selon les caractéristiques des demandes associées. La suite considérera que chaque saison sera un état au sens de l'apprentissage par renforcement, c'est-à-dire que pour chaque saison il y aura une méthode de tarification indépendante de celles pour les autres saisons. Ainsi, chaque saison correspondra à une méthode de tarification dynamique.

La figure 6.1 explique le fonctionnement de l'agent et d'un client pour une étape du jeu répété, c'est-à-dire pour une demande. Ainsi, à l'arrivée d'une demande, l'agent commence par déterminer la saison qui correspond à cette demande et ainsi la méthode de tarification associée à cette saison. Ensuite, il choisit le prix de la demande grâce à cette méthode de tarification. Le client reçoit le prix et choisit alors s'il loue chez cet agent ou pas. Pour finir, l'agent reçoit le paiement du client s'il l'a choisi, ce qui lui permet de mettre à jour la méthode, correspondant à la saison liée à la demande. L'agent mettra à jour sa saisonnalité pour prendre en compte les données récentes.

La méthode pour choisir la saison correspondant à la demande et la phase de mise à jour de la saisonnalité ont été rajoutées par rapport à ce qui était fait dans le chapitre 5.

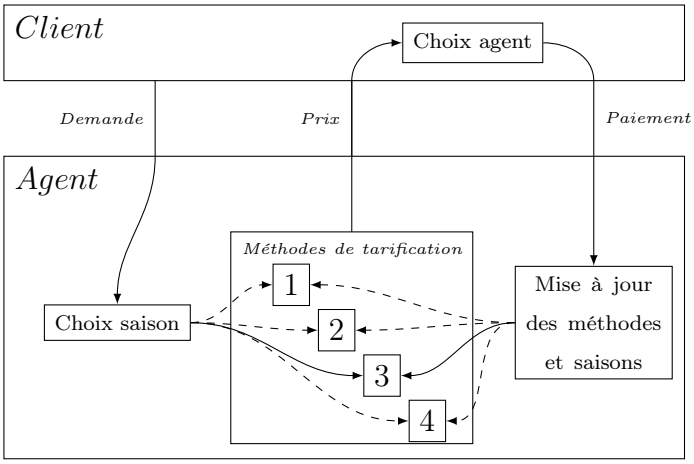

FIGURE 6.1 – Fonctionnement du système pour une demande

6.2.1. Choix de la saison liée à la demande

Cette partie a pour but de définir comment une demande est associée à une saison. En considérant l'ensemble des jours impactés par la demande, c'est-à-dire dans l'intervalle $[debut_d, fin_d]$, chaque jour appartient à une saison. Ainsi, l'ensemble des jours peut appartenir à une seule saison, ce qui arrive pour 39% des demandes (réalisées avec les mêmes conditions que les simulations de la section 6.3.1) et dans ce cas, il suffira d'utiliser la méthode de tarification liée à cette saison. Si les jours dans l'intervalle de location $[debut_d, fin_d]$ appartiennent à plusieurs saisons, il y a plusieurs possibilités :

1. Choisir la saison avec la méthode de tarification qui donne l'estimation du prix la plus élevée ou la plus basse : l'estimation du prix correspond au prix exact pour EMSR et à l'espérance de gain pour la méthode Mixte et LRI.
2. Choisir la saison la plus haute ou la plus basse (la saison la plus haute, respectivement basse est la saison avec la moyenne de la métrique utilisée pour la classification la plus haute, resp. basse, comme défini dans la section 4.5.4)
3. Choisir la saison qui est la plus présente dans l'intervalle de location

Avec la première possibilité, le fait de choisir une saison grâce à l'espérance de gain implique que le choix de la saison sera fait indépendamment du prix réellement réalisé, car celui-ci sera choisi par tirage aléatoire. Cela n'est pas pertinent dans le cadre de l'apprentissage, de même que choisir la saison grâce au tirage aléatoire du prix réalisé. C'est pour cette raison qu'il a été décidé de ne pas utiliser cette possibilité. La deuxième possibilité aura tendance à favoriser la saison la plus haute ou la plus basse, ce qui n'est pas intéressant

d'un point de vue de l'apprentissage. C'est pourquoi il a été choisi de prendre la troisième possibilité.

6.2.2. Phase de tarification et d'apprentissage

Lors de l'arrivée d'une demande, une saison lui est affectée comme indiqué dans le paragraphe précédent. Ensuite, la méthode de tarification liée à la saison est utilisée pour déterminer le prix à proposer à la demande. L'agent est alors choisi ou non. Ensuite, il met à jour sa capacité et, pour finir, la méthode de tarification dynamique utilisée apprend en fonction de ce retour. Ce fonctionnement permet à chaque méthode d'apprentissage d'apprendre sur des données plus homogènes, ce qui devrait mener à un apprentissage plus performant en termes de bénéfice.

6.2.3. Mise à jour de la saisonnalité

La mise à jour de la saisonnalité à l'étape t peut être réalisée avec deux méthodes différentes. La première méthode consiste à refaire une classification à partir de zéro en utilisant la méthode définie dans la section 4.6 en changeant l'étape de référence pour t. Le but est de considérer toutes les nouvelles demandes, ce qui change les valeurs des métriques caractérisant les jours. La deuxième méthode consiste à mettre à jour la classification existante en recalculant les métriques caractérisant chaque jour j tel que $j > t_{prec}$ avec t_{prec} l'étape de référence de la précédente mise à jour. Il y a environ la moitié des j de la nouvelle saisonnalité qui correspondent à $j > t_{prec}$ et qui nécessitent un nouveau calcul. C'est pourquoi le choix a été fait de recalculer la classification complète pour créer la nouvelle saisonnalité. Cette mise à jour est réalisée tous les 10 jours pour avoir suffisamment de nouvelles demandes tout en restant proche de l'état actuel du système. La durée de 10 jours a été choisie expérimentalement et pour limiter le temps de simulation.

Comme chaque saison est associée à une méthode d'apprentissage, il faut déterminer après chaque mise à jour la réaffectation des méthodes aux saisons. Il a été choisi d'ordonner les saisons selon une des métriques utilisées pour la classification dans les deux saisonnalités, celle avant la mise à jour et celle après. La méthode affectée à la première saison avant la mise à jour sera affectée à la première saison de la classification après la mise à jour et ainsi de suite pour toutes les saisons. De plus, d'après la section 4.5.4, il sera considéré qu'il y a quatre saisons qui vont évoluer au cours du temps.

6.3. Évaluation des performances des méthodes saisonnalisées

Le but de cette section sera d'étudier comment le bénéfice total évolue lorsque la méthode utilise la saisonnalité. La suite de ce chapitre nommera méthode saisonnalisée Mixte, saisonnalisée LRI et saisonnalisée EMSR, les méthodes saisonnalisées qui pour toutes les saisons utilisent respectivement la méthode Mixte, LRI ou EMSR. Pour tester les méthodes, le même ensemble de données que dans la section 5.2.3.3 sera utilisé. La section 6.3.1 expliquera quelles sont la ou les métriques qui seront utilisées pour calculer la classification appelée saisonnalité. La section 6.3.2 présentera les paramètres pour chaque méthode utilisée ainsi que les scénarios utilisés pour évaluer la pertinence de l'utilisation de la saisonnalité. Pour finir, dans les sections 6.3.3, 6.3.4 et 6.3.5, les performances des méthodes seront étudiées et comparées selon les scénarios décrits dans la section 6.3.2.

6.3.1. Choix de la méthode de calcul de la saisonnalité

Dans le chapitre 4, plusieurs méthodes ont été définies pour créer une saisonnalité. La différence entre ces méthodes repose sur les métriques qui caractérisent les jours et qui seront utilisées par la méthode *k-means* pour faire la classification. La section 4.5.4 montre que la valeur de k pertinente pour cette classification est 4. De plus, il y a quatre métriques possibles qui peuvent être utilisées indépendamment ou combinées pour faire cette classification : le délai de réservation (délai), la durée de location (durée), le taux d'utilisation au moment de la réservation (util) et le prix par jour (RPD).

Le but est de calculer une saisonnalité pour chaque combinaison de ces métriques, elles seront ensuite testées lorsqu'elles seront utilisées par les méthodes de tarification, et la ou les métriques qui donnent le meilleur bénéfice seront celles qui seront gardées pour la suite. Pour cette phase de test, la méthode saisonnalisée Mixte sera mise en concurrence avec EMSR non saisonnalisé. Les paramètres de chacune des méthodes Mixtes dans la méthode saisonnalisée Mixte sont les mêmes que ceux utilisés pour le chapitre 5. Ces simulations moyennées sur 50 itérations donnent les résultats suivants :

$\times 10^6$	délai	durée	RPD	util
EMSR	1.140	1.156	1.124	1.129
Mixte saison	1.153	1.108	1.173	1.164

$\times 10^6$	délai durée	délai RPD	délai util	durée RPD	durée util	RPD util
EMSR	1.155	1.145	1.145	1.152	1.149	1.141
Mixte saison	1.131	1.148	1.146	1.120	1.124	1.145

$\times 10^6$	délai durée RPD	délai durée util	délai RPD util	durée RPD util	délai durée RPD util
EMSR	1.144	1.145	1.145	1.134	1.141
Mixte saison	1.146	1.146	1.141	1.155	1.151

L'ensemble de ces résultats indique que la saisonnalité utilisant seulement le RPD est la plus performante en termes de bénéfice total. Ce sera donc cette saisonnalité qui sera utilisée pour la suite des expérimentations.

Pour compléter l'étude de cette classification, la distribution du RPD moyen calculé pour chaque jour a aussi été étudiée dans la courbe 6.2. Cette distribution ne comprend pas plusieurs pics qui permettraient d'avoir une classification simple, mais les RPD moyens proches de 25 euros sont surreprésentés par rapport aux valeurs de RPD moyen plus élevées. Ainsi, il était bien pertinent d'utiliser un algorithme de classification. La seconde courbe 6.3 montre l'évolution du RPD moyen pendant deux ans, ce qui permet de voir la reproductibilité d'une année sur l'autre du RPD moyen. Le but sera ainsi de tenir compte de cette reproductibilité à travers la saisonnalité dans les méthodes de tarification.

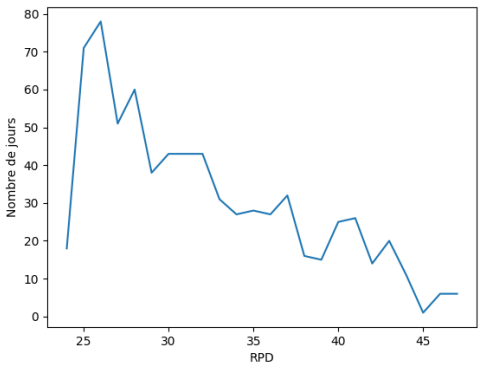

FIGURE 6.2 – Distribution du RPD moyen calculé pour chaque jour

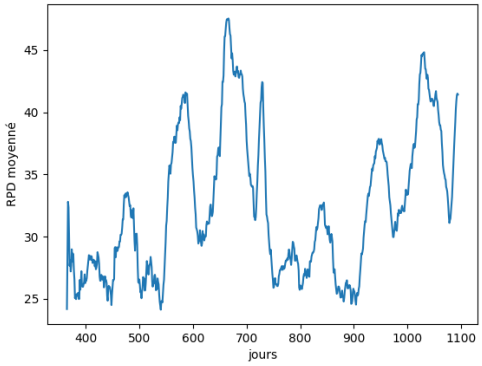

FIGURE 6.3 – Évolution du RPD moyen sur deux ans

Le but a ensuite été d'étudier l'évolution de la saisonnalité lors de son utilisation pour la tarification dynamique dans le même cadre que les expérimentations précédentes. Pour cela, quatre types de courbes ont été obtenus. La première courbe 6.4 représente l'évolution des saisons durant l'amorçage et la phase d'étude. En abscisse, ce sont les jours qui sont inclus dans la saisonnalité à chaque étape. Cela donne quatre ans car il faut un an avant le début de l'amorçage pour calculer la saisonnalité et, à la fin de la phase d'étude, il y a aussi une année dans l'avenir. En ordonnée, chaque ligne représente la nouvelle saisonnalité après mise à jour, réalisée tous les 10 jours. La couleur violette correspond à la basse saison, puis il y a le bleu foncé, puis le bleu clair et pour finir le vert pour la haute saison. Cette courbe a été réduite à la phase

d'étude dans la courbe 6.8 pour plus de visibilité. La courbe 6.7 représente l'indice de Jaccard [36] à chaque mise à jour entre la précédente saisonnalité et la nouvelle. Ainsi, ces courbes réalisées sur une seule exécution permettent de remarquer un changement conséquent à la $8^{\text{ème}}$ mise à jour pendant la phase d'amorçage et à la $50^{\text{ème}}$ mise à jour. Mais ce changement conséquent n'apparaît pas forcément pour chaque exécution, dû à l'aléatoire des méthodes de tarification (courbes en annexes C).

La courbe 6.8 permet de savoir que la haute saison en vert se situe en juillet-août (de 180 à 240) ainsi que d'octobre à début janvier (de 275 à 365 et avant 10), ce qui correspond au retour du loueur.

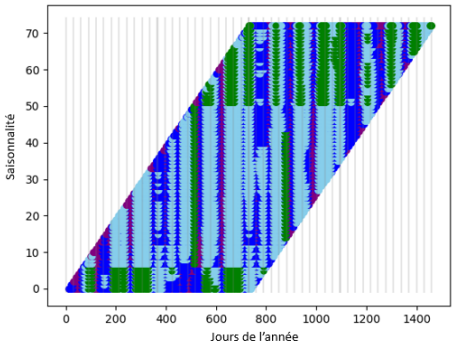

FIGURE 6.4 – Évolution de la saisonnalité

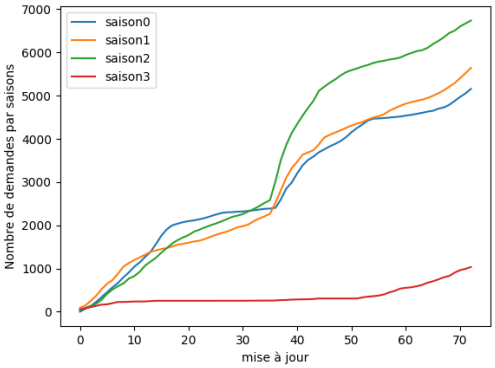

FIGURE 6.5 – Évolution du nombre de demandes par saison

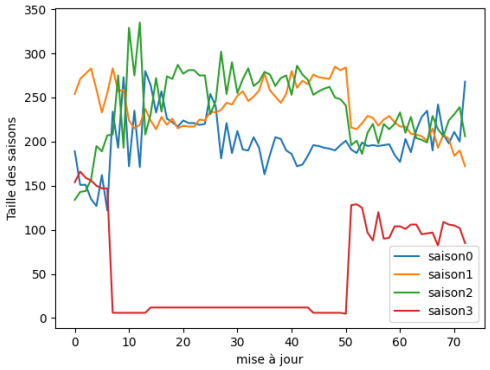

FIGURE 6.6 – Évolution de la taille des saisons

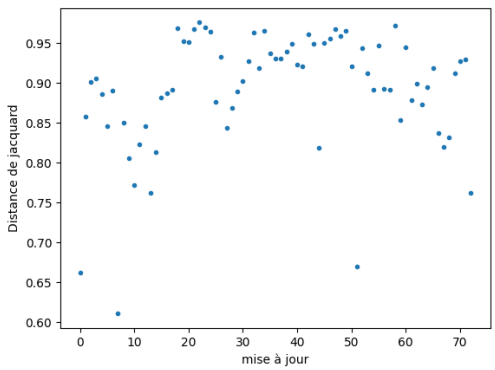

FIGURE 6.7 – Indice de Jaccard à chaque mise à jour de la saisonnalité

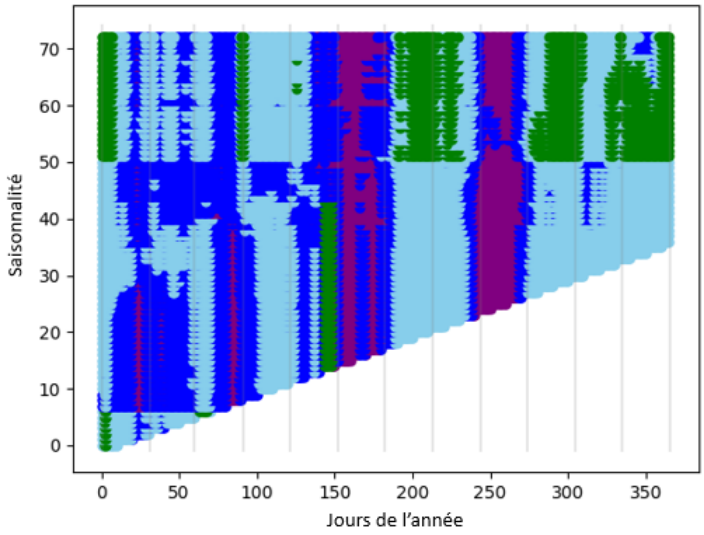

FIGURE 6.8 – Évolution de la saisonnalité sur la phase d'étude

6.3.2. Paramètres des méthodes et scénarios

Dans la section 6.3.3, le but est de savoir si la saisonnalité permet une augmentation du bénéfice. Pour cela, la section commencera par mettre en concurrence la méthode saisonnalisée Mixte (définie dans la section 6.3) et une méthode qui n'utilise pas la saisonnalité, LRI ou EMSR. L'objectif sera ainsi de comparer le bénéfice obtenu à celui obtenu dans le chapitre 5. Dans un deuxième temps, le but sera de savoir si une méthode saisonnalisée se démarque d'un point de vue du bénéfice quand elle est en concurrence avec d'autres méthodes saisonnalisées. Pour finir, dans la section 6.3.5, le but sera d'améliorer les méthodes saisonnalisées en utilisant des méthodes différentes pour chaque saison. Les paramètres des méthodes restent inchangés par rapport au chapitre 5 à part pour la valeur du paramètre d'apprentissage b et pour la valeur de v_{min} le seuil exploratoire dans le cas des méthodes saisonnalisées. Pour la méthode saisonnalisée Mixte, chaque méthode Mixte associée à chaque saison aura respectivement pour chacun des états les valeurs de b suivantes : 0.005, 0.01, 0.02 fixées expérimentalement avec pour but de maximiser le bénéfice en concurrence avec la méthode EMSR non saisonnalisée. La valeur de v_{min} est 0.005. Pour la méthode saisonnalisée LRI, pour toutes les saisons $b = 0.005$ et $v_{min} = 0.01$.

6.3.3. Évaluation des performances contre les méthodes sans saisonnalité

Le but de cette section est de vérifier que la saisonnalité permet une augmentation du bénéfice pour la méthode qui utilise la saisonnalité lorsqu'elle est en concurrence avec une méthode qui ne l'utilise pas.

Cette section commence par étudier le bénéfice change entre la méthode saisonnalisée Mixte et la méthode Mixte lorsqu'elles ont le même concurrent. Pour cela, elles seront mises en concurrence avec EMSR puis LRI en comparant le bénéfice à celui des figures du chapitre 5 regroupées en Annexe A. En comparant les courbes 6.9 avec 5.20 et 6.10 avec 5.21, le bénéfice de la méthode saisonnalisée Mixte est nettement meilleur que celui de la méthode Mixte lorsqu'elles sont en concurrence avec EMSR. Par exemple, pour une capacité de 150, le bénéfice de la méthode Mixte passe de 1 180 000 euros à 1 280 000 euros, il y a donc un gain d'environ 10%. Entre la 8000ème demande et la 11000ème, il y a une rupture dans la pente de l'évolution du bénéfice de la méthode Mixte saisonnalisée due à une utilisation trop importante des véhicules de l'agent qui utilise la méthode saisonnalisée Mixte alors qu'il y a encore des demandes qui arrivent. C'est un constat qui a déjà été fait dans le chapitre 5, ainsi, la saisonnalité n'améliore pas cela.

La comparaison des courbes 6.10 et 5.21 permet de remarquer que le premier point d'intersection dans le cas de la méthode saisonnalisée Mixte correspond à une capacité plus faible que pour la méthode Mixte et, à l'inverse, le deuxième point correspond à une capacité plus élevée. Ainsi, la méthode Mixte saisonnalisée a un meilleur bénéfice que EMSR pour plus de valeur de capacités, c'est-à-dire de valeurs en abscisse, que la méthode Mixte dans la même configuration.

Conclusion : le bénéfice de la méthode saisonnalisée Mixte est nettement meilleur que celui de la méthode Mixte lorsqu'elles sont en concurrence avec EMSR quelle que soit la capacité.

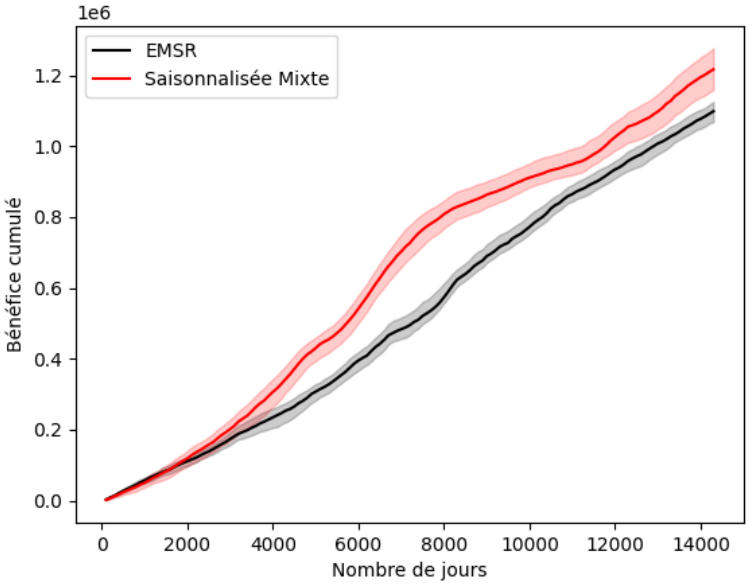

FIGURE 6.9 – La méthode saisonnalisée Mixte contre EMSR

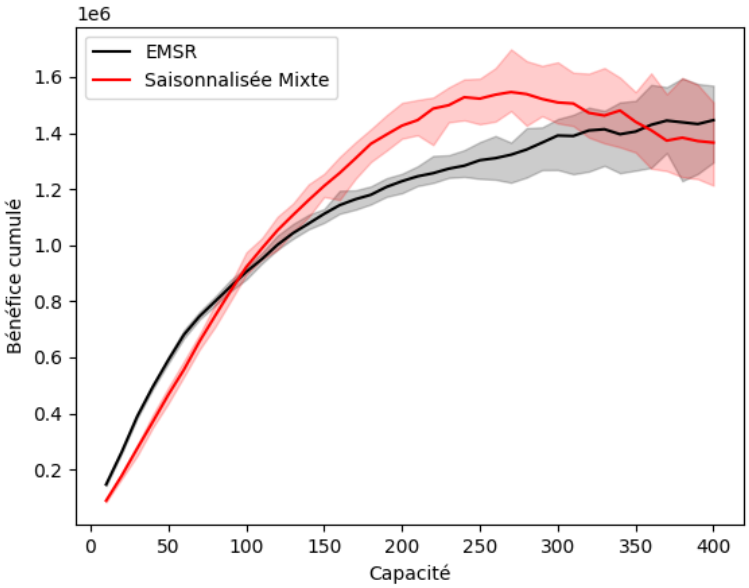

FIGURE 6.10 – Évolution du bénéfice pour la méthode saisonnalisée Mixte contre EMSR en fonction de la capacité

Toutefois, lorsqu'on compare les courbes 6.11 avec 5.18 et 6.12 avec 5.19, le bénéfice reste identique pour une capacité de 150 mais pour les capacités

plus élevées, la méthode saisonnalisée Mixte a un bénéfice inférieur à celui de la méthode Mixte lorsqu'elles sont en concurrence avec LRI. Dans tous les cas, lorsque deux méthodes utilisant de l'apprentissage par renforcement sont en concurrence, les vecteurs d'apprentissage se stabilisent de manière identique. Les différences de bénéfices sont dues aux changements des valeurs de b et v_{min} (définies dans la section 5.2.2.1) fixés pour optimiser les performances en situation de concurrence avec EMSR ainsi qu'au nombre de vecteurs d'apprentissage. b correspond à la vitesse d'apprentissage, v_{min} correspond à la probabilité minimale de jouer chaque action pour garder un peu d'exploration. Ainsi, tous ces paramètres permettent une modification plus ou moins conséquente et fréquente du vecteur d'apprentissage, ce qui influe sur le bénéfice.

Conclusion : le bénéfice reste identique pour une capacité inférieure à 150 mais pour les capacités plus élevées, la méthode saisonnalisée Mixte a un bénéfice inférieur à celui de la méthode Mixte lorsqu'elles sont en concurrence avec LRI.

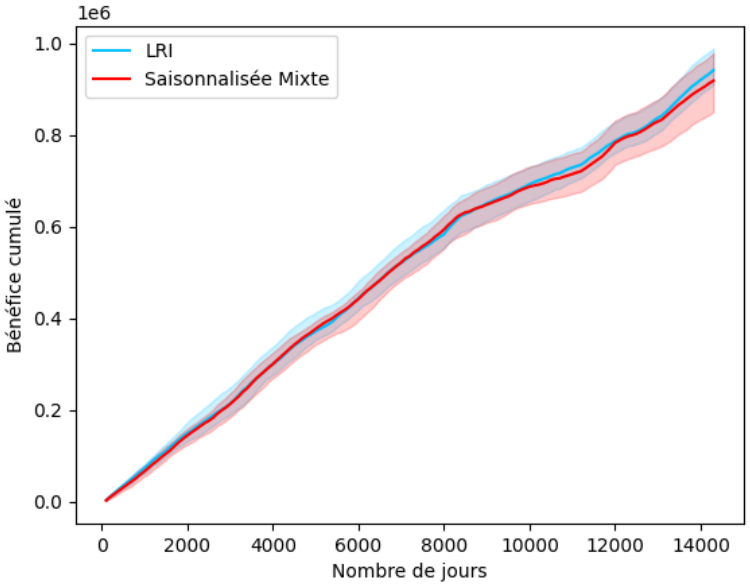

FIGURE 6.11 – La méthode saisonnalisée Mixte contre LRI

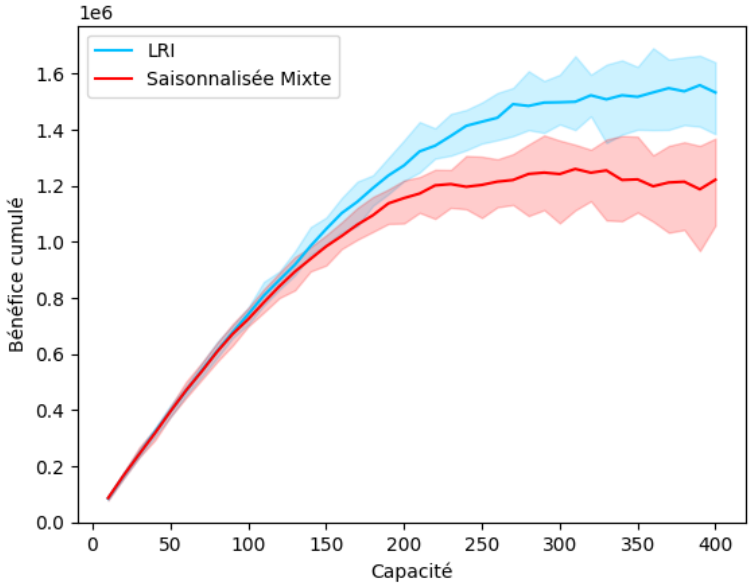

FIGURE 6.12 – Évolution du bénéfice pour la méthode saisonnalisée Mixte contre LRI en fonction de la capacité

Ainsi, la méthode saisonnalisée Mixte apporte un gain de bénéfice par rapport à la méthode Mixte lorsqu'elle est en concurrence avec EMSR, mais n'est pas intéressante contre la méthode LRI.

La suite va étudier la pertinence en termes de bénéfice d'utiliser la méthode saisonnalisée EMSR plutôt que d'utiliser EMSR. Pour cela, les courbes 6.13 et 6.14 sont comparées avec 5.20 et 5.21. Pour une capacité de 150, le bénéfice de la méthode saisonnalisée EMSR est plus conséquent que celui de la méthode Mixte lorsqu'ils sont en concurrence, alors que ce n'est pas le cas pour EMSR lorsqu'il est en concurrence avec la méthode Mixte. Cette information est importante dans le cadre de la recherche de la méthode qui obtient toujours un bénéfice supérieur à celui de son concurrent.

La méthode saisonnalisée EMSR obtient un meilleur bénéfice qu'EMSR en concurrence avec la méthode Mixte tant que la capacité est inférieure à 250 et diminue le bénéfice de la méthode Mixte pour les mêmes capacités. Lorsque la capacité est plus élevée que 250, le bénéfice de la méthode saisonnalisée EMSR est diminué par rapport à EMSR. De plus, la méthode Mixte obtient alors plus de bénéfices car celle-ci s'adapte au comportement différent entre la méthode saisonnalisée EMSR et EMSR. Cela s'explique par le fait que la distribution de prix proposés par la méthode saisonnalisée EMSR change en fonction des saisons, ce qui permet à la méthode Mixte d'obtenir plus de demandes à des prix plus élevés, et cela entraîne une perte de bénéfice pour la méthode saisonnalisée EMSR.

Conclusion : la méthode saisonnalisée EMSR obtient un meilleur bénéfice qu'EMSR en concurrence avec la méthode Mixte tant que la capacité est inférieure à 250, sinon, son bénéfice est inférieur à celui d'EMSR dans les mêmes conditions.

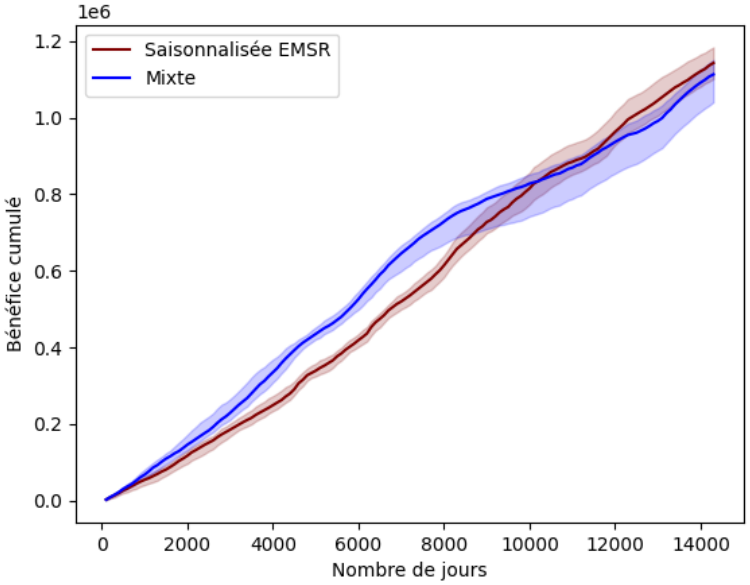

FIGURE 6.13 – La méthode saisonnalisée EMSR contre la méthode Mixte

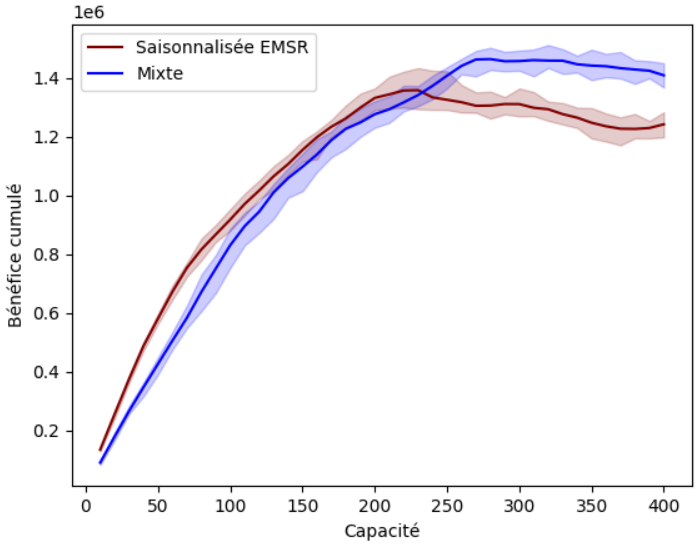

FIGURE 6.14 – Évolution du bénéfice pour la méthode saisonnalisée EMSR contre la méthode Mixte en fonction de la capacité

Le bénéfice augmente de manière conséquente entre la méthode saisonnalisée EMSR et la méthode EMSR lorsqu'elles sont en concurrence avec LRI, ce qui est illustré par les courbes 6.15 et 6.16 comparées avec les courbes 5.3 et 5.7. Cela est dû au fait que la méthode saisonnalisée EMSR optimise mieux les prix proposés aux demandes qui le choisissent car pour les saisons où LRI se remplit trop, la méthode saisonnalisée EMSR apprend à augmenter ses prix. La courbe de l'évolution du bénéfice en fonction de la capacité de la méthode saisonnalisée EMSR contre LRI (Figure 6.16) ressemble à la courbe de l'évolution du bénéfice en fonction de la capacité de la méthode saisonnalisée EMSR contre la méthode Mixte (Figure 6.14), la méthode Mixte ayant un meilleur bénéfice que LRI.

Conclusion : Le bénéfice augmente de manière conséquente entre la méthode saisonnalisée EMSR et la méthode EMSR lorsqu'elles sont en concurrence avec LRI.

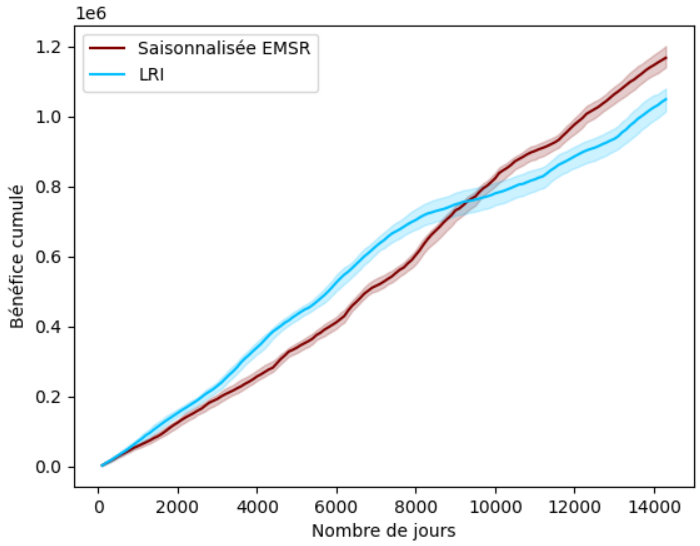

FIGURE 6.15 – La méthode saisonnalisée EMSR contre LRI

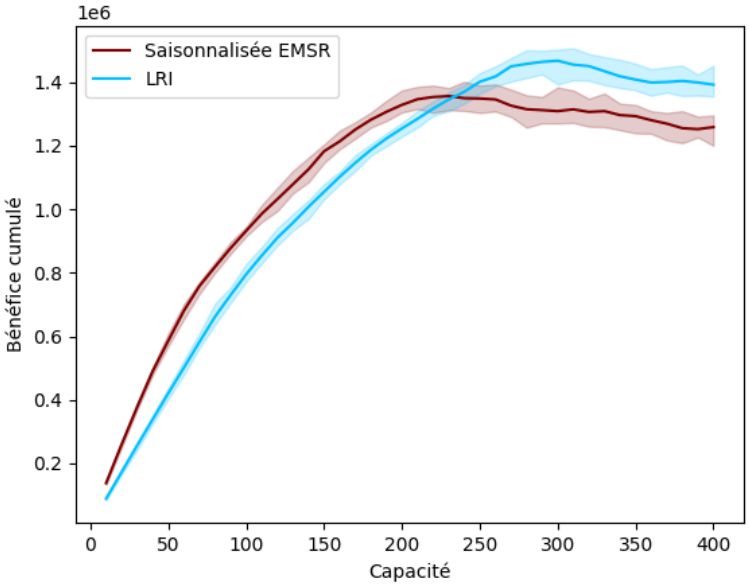

FIGURE 6.16 – Évolution du bénéfice pour la méthode saisonnalisée EMSR contre LRI en fonction de la capacité

Le but est maintenant d'étudier la différence en termes de bénéfice entre la méthode saisonnalisée LRI et la méthode LRI non saisonnalisée. Pour cela, les deux méthodes sont mises en concurrence avec la méthode EMSR dans les courbes 6.17 et 6.18 comparées avec les courbes 5.3 et 5.7. D'après ces courbes et quelle que soit la valeur de la capacité, la méthode saisonnalisée LRI apporte toujours un meilleur bénéfice par rapport à celui de la méthode LRI lorsqu'elle est en concurrence avec la méthode EMSR. La méthode saisonnalisée LRI a un meilleur bénéfice qu'EMSR entre les capacités 110 et 220 ainsi que LRI dans la même situation. Ainsi, la méthode saisonnalisée LRI est plus intéressante que LRI même pour des capacités plus faibles. Ce gain de bénéfice est dû aux quatre vecteurs d'apprentissage qui en moyenne proposent des prix plus élevés.

Conclusion : la méthode saisonnalisée LRI apporte toujours un meilleur bénéfice par rapport à celui de la méthode LRI lorsqu'elle est en concurrence avec la méthode EMSR.

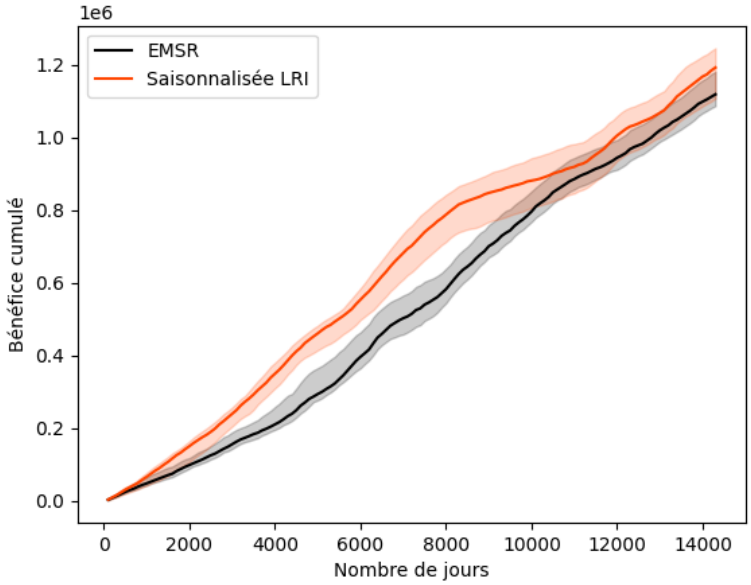

FIGURE 6.17 – La méthode saisonnalisée LRI contre EMSR

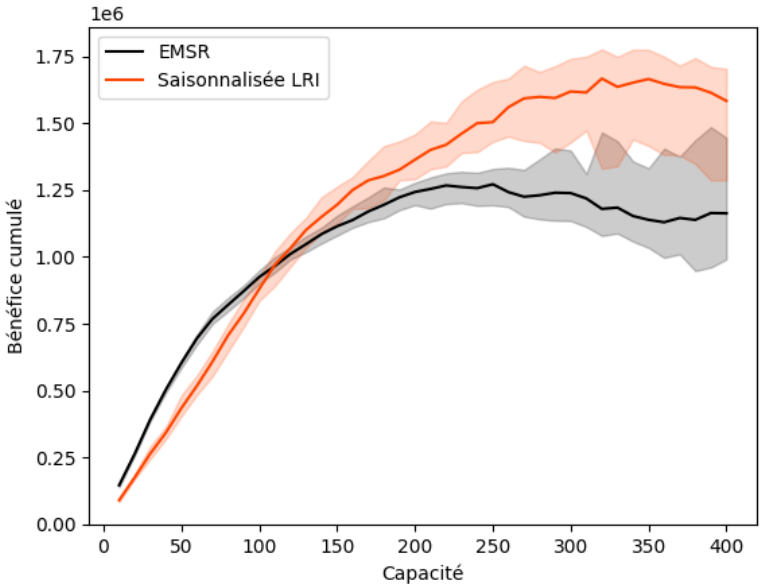

FIGURE 6.18 – Évolution du bénéfice pour la méthode saisonnalisée LRI contre EMSR en fonction de la capacité

Les courbes 6.19 et 6.20, où la méthode saisonnalisée LRI est en concurrence avec la méthode Mixte, sont similaires aux courbes 5.18 et 5.19, où la méthode

LRI est en concurrence avec la méthode Mixte. La conclusion reste la même que pour la méthode Mixte saisonnalisée en concurrence avec LRI car ce sont deux méthodes utilisant l'apprentissage par renforcement. Ainsi, dans cette situation, la prise en compte de la saisonnalité ne change rien.

Conclusion : l'utilisation de la méthode saisonnalisée LRI à la place de LRI n'apporte pas de bénéfice lorsqu'elles sont en concurrence avec la méthode Mixte.

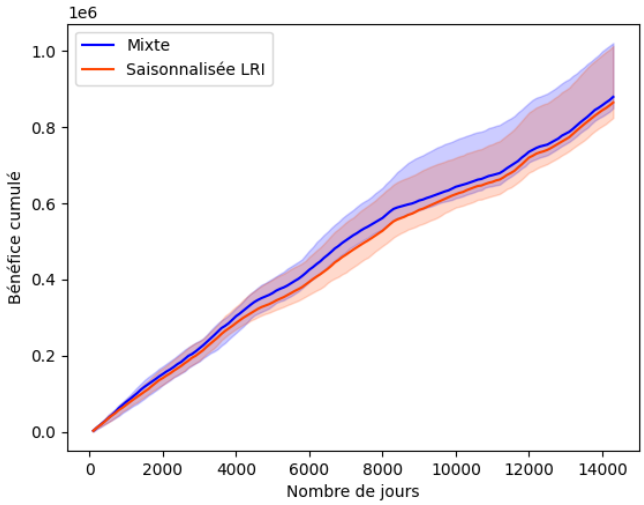

FIGURE 6.19 – La méthode saisonnalisée LRI contre la méthode Mixte

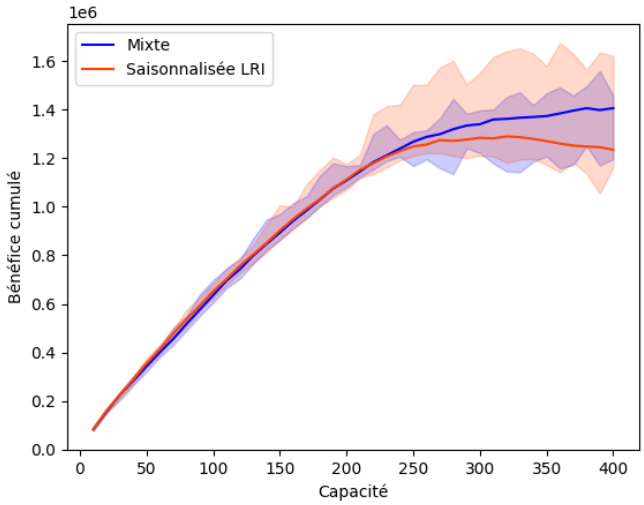

FIGURE 6.20 – Évolution du bénéfice pour la méthode saisonnalisée LRI contre la méthode Mixte en fonction de la capacité

Ainsi, la prise en compte de la saisonnalité par les méthodes lorsqu'elles sont en concurrence avec les méthodes non saisonnalisées est toujours bénéfique, à part pour la méthode EMSR quand la capacité est supérieure à 250, cette capacité correspond au seuil où, pour une capacité plus faible, toutes les demandes ne sont pas réalisables.

6.3.4. Évaluation des performances contre les méthodes saisonnalisées

Le but de cette section est de savoir si la saisonnalité permet une augmentation du bénéfice pour la méthode qui l'utilise lorsqu'elle est en concurrence avec une méthode qui l'utilise aussi.

Cette section commence par étudier la méthode saisonnalisée EMSR contre la méthode saisonnalisée Mixte. La méthode EMSR est d'abord comparée avec la méthode EMSR saisonnalisée lorsqu'elles sont en concurrence avec la méthode Mixte saisonnalisée, avec les figures 6.21 et 6.22 comparées avec les figures 6.9 et 6.10. Il y a une amélioration du bénéfice pour la méthode saisonnalisée EMSR par rapport à EMSR, avec aussi pour certaines capacités une baisse du bénéfice de la méthode saisonnalisée Mixte. La méthode saisonnalisée Mixte est comparée avec la méthode Mixte lorsqu'elle est en concurrence avec la méthode saisonnalisée EMSR à travers les figures 6.21, 6.22, 6.13 et 6.14. Il y a une amélioration du bénéfice pour la méthode saisonnalisée Mixte par rapport à la méthode Mixte pour une capacité de moins de 300 voitures ; au-delà, il y a une baisse du bénéfice. L'utilisation de la méthode saisonnalisée Mixte par rapport à la méthode Mixte entraîne une baisse du bénéfice pour la méthode EMSR saisonnalisée.

Conclusion : Il y a une amélioration du bénéfice pour la méthode saisonnalisée EMSR par rapport à EMSR lorsqu'elles sont en concurrence avec la méthode saisonnalisée Mixte. Il y a une amélioration du bénéfice pour la méthode saisonnalisée Mixte par rapport à la méthode Mixte pour une capacité de moins de 300 voitures ; au-delà, il y a une baisse du bénéfice.

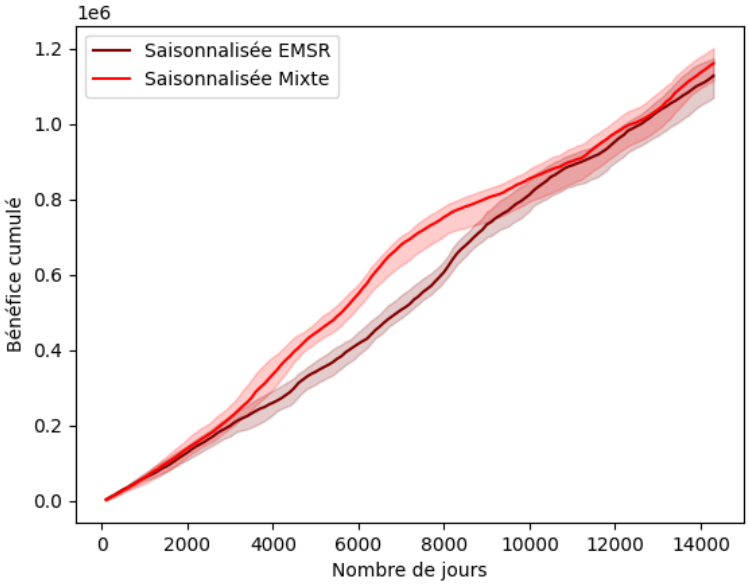

FIGURE 6.21 – La méthode saisonnalisée EMSR contre la méthode saisonnalisée Mixte

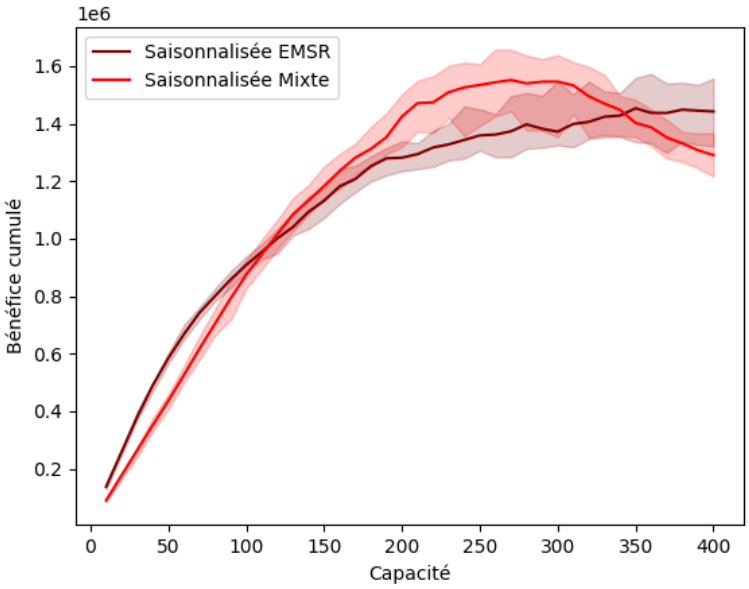

FIGURE 6.22 – Évolution du bénéfice pour la méthode saisonnalisée EMSR contre la méthode saisonnalisée Mixte en fonction de la capacité

La méthode saisonnalisée LRI en concurrence avec la méthode saisonna-

lisée EMSR va être étudiée à travers les figures 6.23 et 6.24 comparées avec 6.15 et 6.16 et avec 6.23 et 6.18. Les conclusions sont similaires à celles obtenues entre la méthode saisonnalisée EMSR et la méthode saisonnalisée Mixte, c'est-à-dire que la saisonnalité apporte du bénéfice dans la majorité des cas, à part pour la méthode saisonnalisée LRI pour une capacité au-dessus de 300 voitures. De plus, la méthode saisonnalisée placée en concurrence perd du bénéfice entre la méthode LRI et EMSR et la méthode saisonnalisée LRI et la méthode saisonnalisée EMSR.

Conclusion : Il y a une amélioration du bénéfice pour la méthode saisonnalisée EMSR par rapport à EMSR lorsqu'elles sont en concurrence avec la méthode saisonnalisée LRI. Il y a une amélioration du bénéfice pour la méthode saisonnalisée LRI par rapport à la méthode LRI pour une capacité de moins de 300 voitures ; au-delà, il y a une baisse du bénéfice.

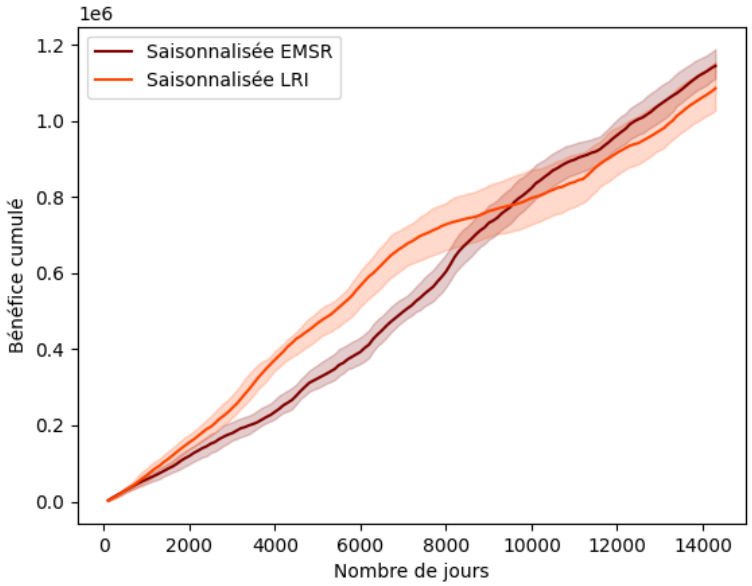

FIGURE 6.23 – La méthode saisonnalisée EMSR contre la méthode saisonnalisée LRI

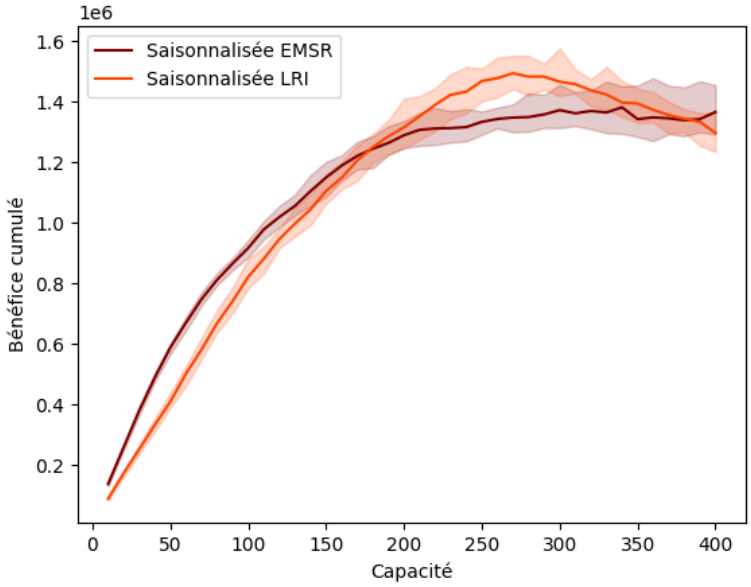

FIGURE 6.24 – Évolution du bénéfice pour la méthode saisonnalisée EMSR contre la méthode saisonnalisée LRI en fonction de la capacité

La méthode saisonnalisée Mixte en concurrence avec la méthode saisonnalisée LRI va être étudiée à travers les courbes 6.25 et 6.26 comparées avec 6.11 et 6.12 et avec 6.19 et 6.20. Dans ces six figures, les courbes ont des formes similaires, ce qui est dû à l'unique utilisation de l'apprentissage par renforcement même s'il y a des variations dans la valeur du bénéfice. La saisonnalité n'apporte rien entre la méthode saisonnalisée Mixte et la méthode Mixte dans ce cas, mais apporte entre la méthode saisonnalisée LRI et la méthode LRI. Pour la comparaison entre la méthode Mixte et la méthode saisonnalisée Mixte, cela est sûrement dû aux nombreux vecteurs d'apprentissage car il y en a un par couple (état, saison), ce qui nuit à la vitesse et à la pertinence de l'apprentissage.

Conclusion : La méthode saisonnalisée Mixte n'apporte pas plus de bénéfice que la méthode Mixte lorsqu'elles sont en concurrence avec la méthode saisonnalisée LRI. Par contre, la méthode saisonnalisée LRI a un meilleur bénéfice que LRI en situation de concurrence avec la méthode saisonnalisée Mixte.

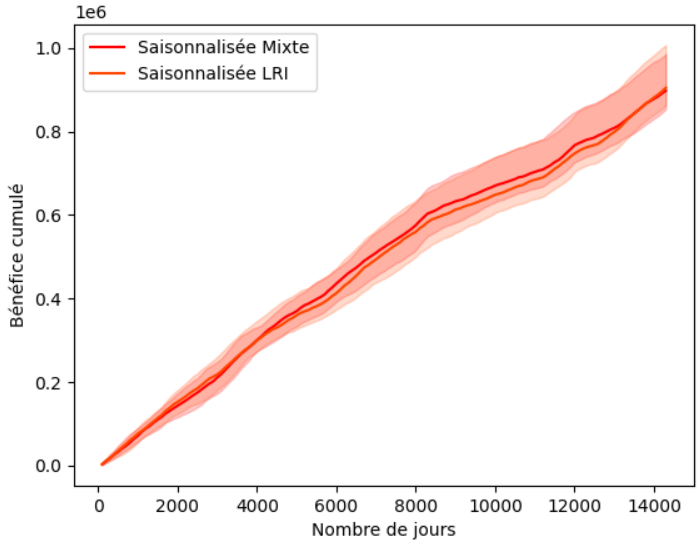

FIGURE 6.25 – La méthode saisonnalisée Mixte contre la méthode saisonnalisée LRI

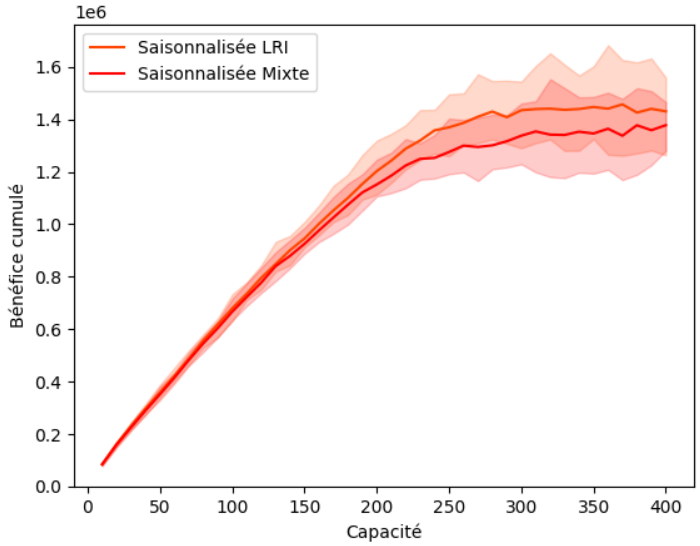

FIGURE 6.26 – Évolution du bénéfice pour la méthode saisonnalisée Mixte contre la méthode saisonnalisée LRI en fonction de la capacité

L'ensemble des courbes de cette partie permet de savoir que les méthodes saisonnalisées apportent plus de bénéfice que les méthodes non saisonnalisées lorsqu'elles sont en concurrence avec les méthodes saisonnalisées, à part dans le cas de la méthode saisonnalisée Mixte contre la méthode saisonnalisée LRI.

6.3.5. Étude du bénéfice pour une méthode saisonnalisée avec une méthode différente par saison

Jusqu'à maintenant, les méthodes d'apprentissage utilisées pour chaque saison sont identiques, mais cela n'est pas une nécessité, c'est pourquoi dans cette partie, le but a été de chercher expérimentalement quelles étaient les méthodes à utiliser pour chaque saison afin d'avoir le meilleur bénéfice cumulé lorsque la méthode saisonnalisée est en concurrence avec la méthode EMSR non saisonnalisée. Les résultats obtenus permettent de dire que la méthode avec le meilleur bénéfice est celle où chaque saison utilise la méthode EMSR, à part pour la deuxième saison, qui utilisera la méthode Mixte. Cette méthode sera appelée *méthode saisonnalisée différente par saison*. Cette méthode va maintenant être comparée avec les méthodes non saisonnalisées et ensuite avec les méthodes saisonnalisées Mixte, LRI et EMSR.

Lorsque la *méthode différente par saison* est en concurrence avec EMSR, le bénéfice est plus élevé que celui de toutes les autres méthodes lorsqu'elles sont en concurrence avec EMSR (voir les figures des Sections 6.3.3 et 6.3.4). Ce qui s'explique car il a été choisi d'utiliser le quadruplé le plus performant contre EMSR ; en particulier, cela a aussi été comparé au bénéfice de la méthode saisonnalisée Mixte et de la méthode saisonnalisée LRI contre EMSR car ce sont des quadruplés avec quatre méthodes identiques.

Conclusion : lorsque la *méthode différente par saison* est en concurrence avec EMSR, le bénéfice est plus élevé que celui de toutes les autres méthodes lorsqu'elles sont en concurrence avec EMSR.

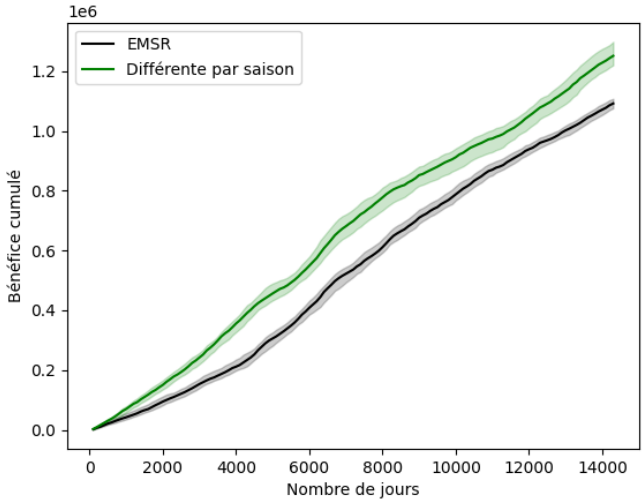

FIGURE 6.27 – EMSR contre la méthode différente par saison

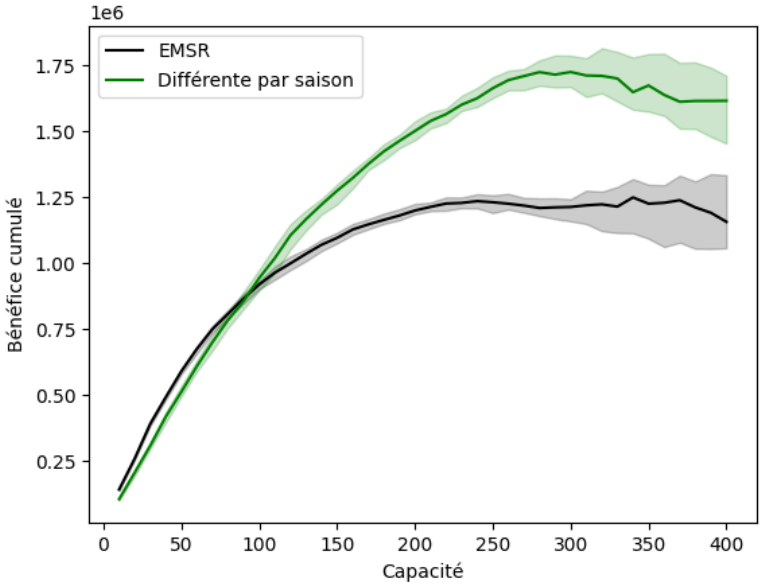

FIGURE 6.28 – Évolution du bénéfice pour EMSR contre la méthode saisonnalisée différente par saison en fonction de la capacité

Le même résultat apparaît lorsque la *méthode différente par saison* est en concurrence avec la méthode LRI et la méthode Mixte car la méthode saisonnalisée EMSR a déjà un très bon bénéfice quand la capacité est inférieure à 250. Ainsi, la méthode Mixte utilisée pour la deuxième saison n'influe pas trop sur le bénéfice, mais elle influe fortement pour des capacités plus élevées car la *méthode différente par saison* a un bénéfice toujours supérieur aux autres méthodes non saisonnalisées.

Conclusion : Lorsque la *méthode différente par saison* est en concurrence avec LRI et la méthode Mixte, le bénéfice est aussi plus élevé que celui de toutes les autres méthodes lorsqu'elles sont en concurrence avec LRI ou la méthode Mixte.

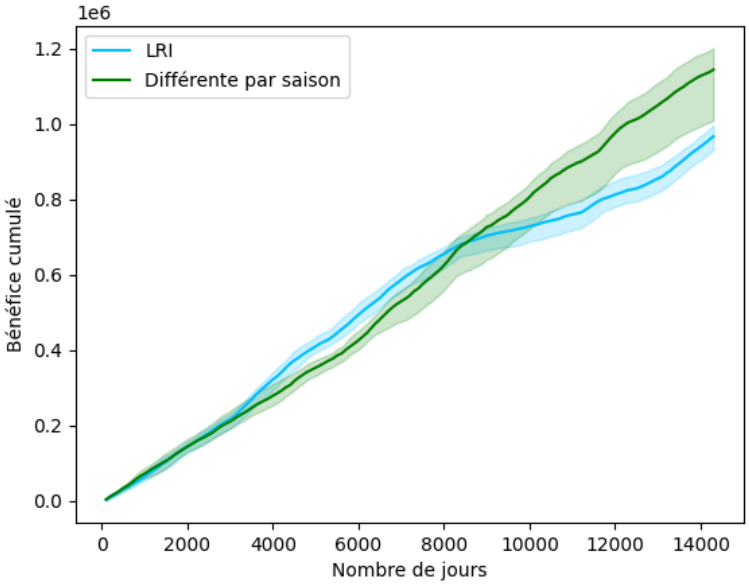

FIGURE 6.29 – LRI contre la méthode saisonnalisée différente par saison

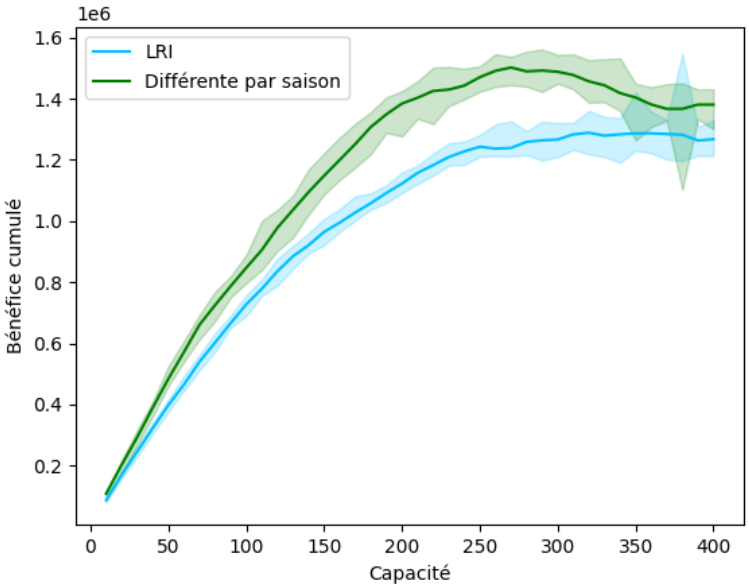

FIGURE 6.30 – Évolution du bénéfice pour LRI contre la méthode saisonnalisée différente par saison en fonction de la capacité

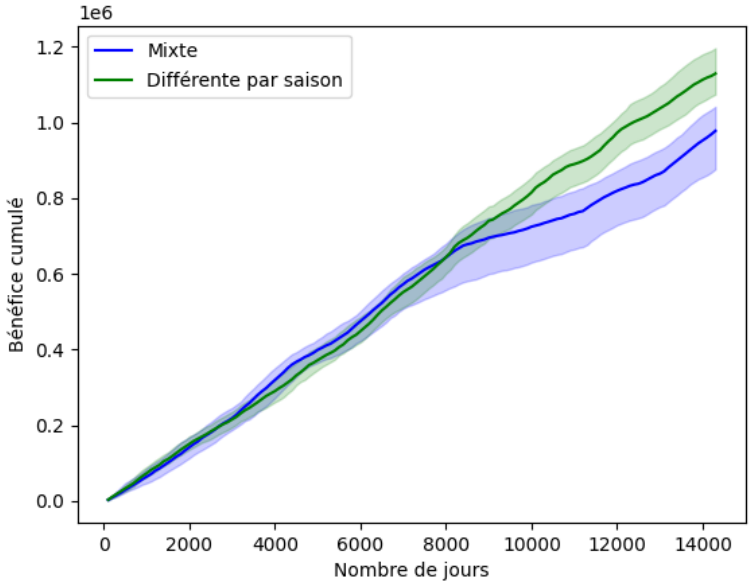

FIGURE 6.31 – La méthode Mixte contre la méthode saisonnalisée différente par saison

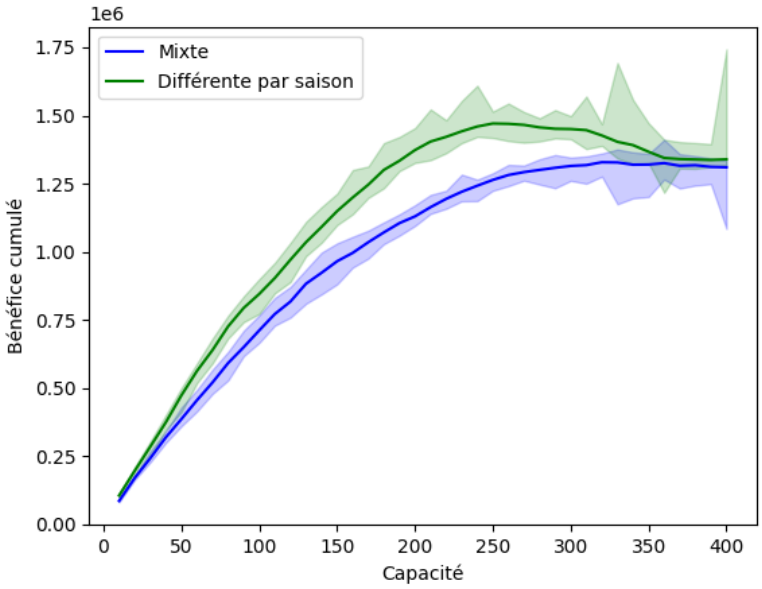

FIGURE 6.32 – Évolution du bénéfice pour la méthode Mixte contre la méthode saisonnalisée différente par saison en fonction de la capacité

Le but maintenant est de voir l'évolution du bénéfice de la *méthode dif-*

férente par saison lorsqu'elle est en concurrence avec des méthodes saisonnalisées. En comparant avec les courbes des sections 6.3.3 et 6.3.4, la *méthode différente par saison* a un meilleur bénéfice que toutes les autres méthodes lorsqu'elle est en concurrence avec une méthode saisonnalisée.

Conclusion : la *méthode différente par saison* a un meilleur bénéfice que toutes les autres méthodes lorsqu'elle est en concurrence avec une méthode saisonnalisée.

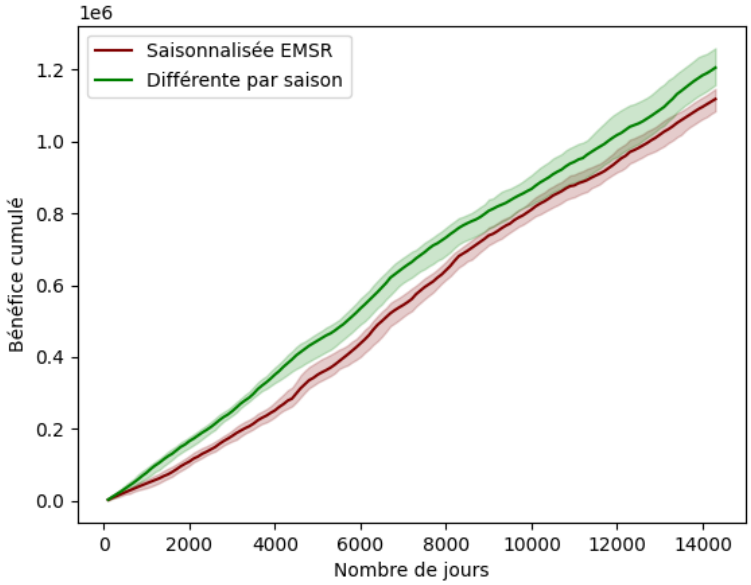

FIGURE 6.33 – La méthode saisonnalisée EMSR contre la méthode saisonnalisée différente par saison

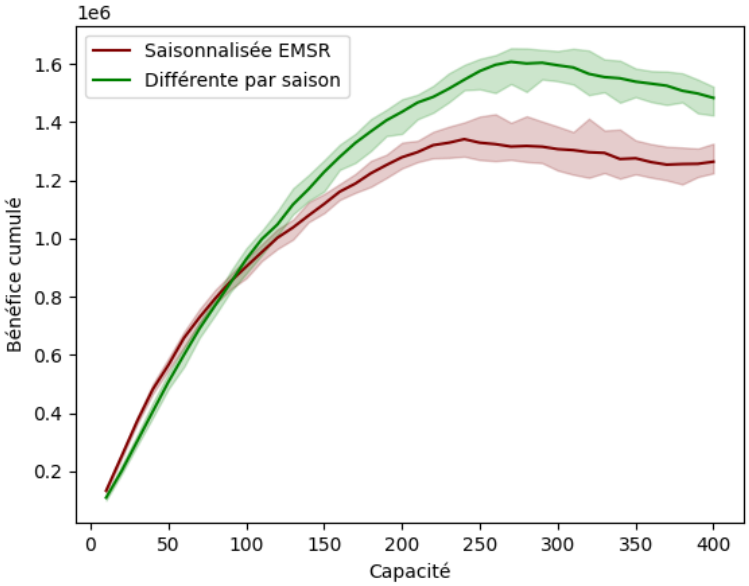

FIGURE 6.34 – Évolution du bénéfice pour la méthode saisonnalisée EMSR contre la méthode saisonnalisée différente par saison en fonction de la capacité

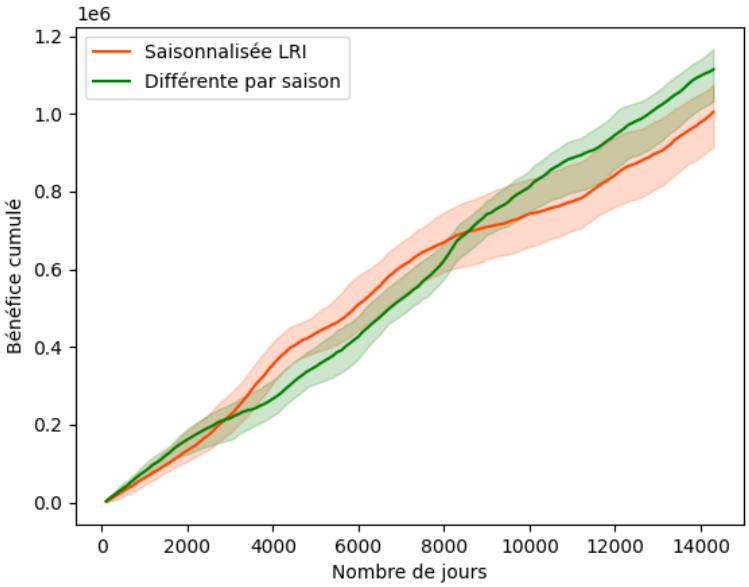

FIGURE 6.35 – La méthode saisonnalisée LRI contre la méthode saisonnalisée différente par saison

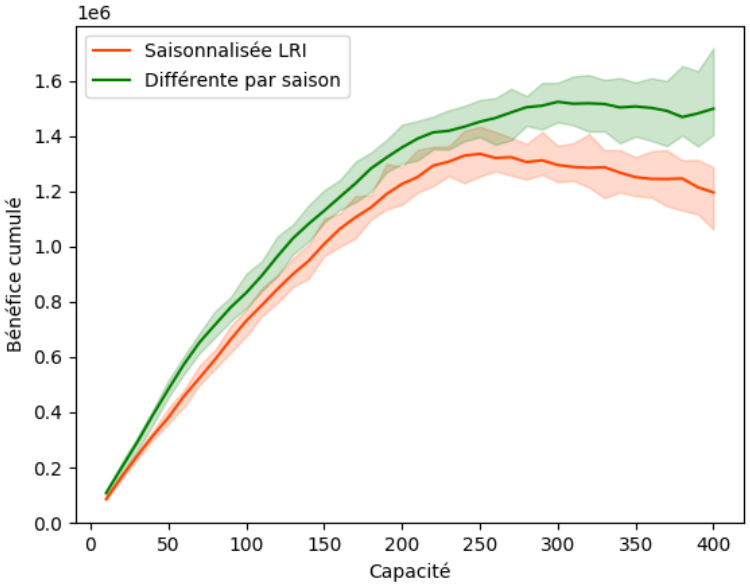

FIGURE 6.36 – Évolution du bénéfice pour la méthode saisonnalisée LRI contre la méthode saisonnalisée différente par saison en fonction de la capacité

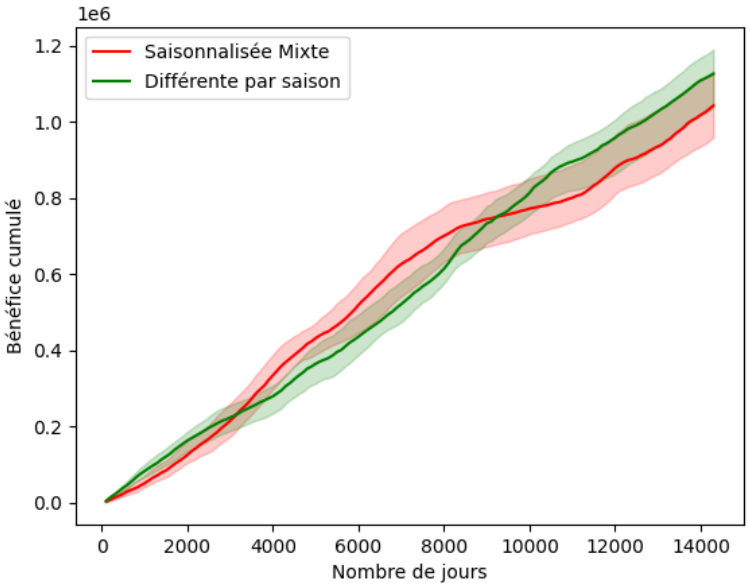

FIGURE 6.37 – La méthode saisonnalisée Mixte contre la méthode saisonnalisée différente par saison

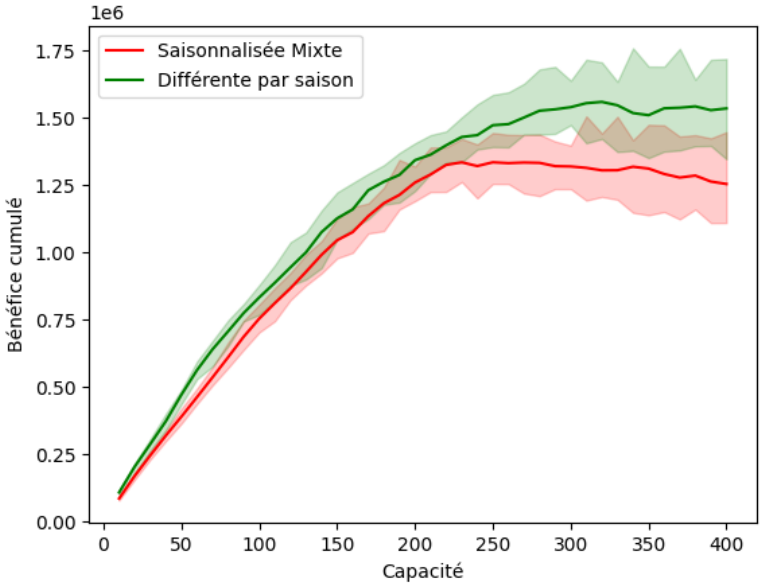

FIGURE 6.38 – Évolution du bénéfice pour la méthode saisonnalisée Mixte contre la méthode saisonnalisée différente par saison en fonction de la capacité

Ainsi, la *méthode différente par saison* est la méthode qui réussit le mieux quelle que soit la concurrence, grâce à la méthode Mixte en saison 2 qui fixe le prix de plus d'un tiers des demandes d'après la Figure 6.5. De plus, elle a un meilleur bénéfice que ses concurrents à part pour les capacités inférieures à 100 et lorsque la concurrence utilise la méthode EMSR ou la méthode saisonnalisée EMSR.

Synthèse :

- La prise en compte de la saisonnalité par les méthodes lorsqu'elles sont en concurrence avec les méthodes non saisonnalisées est toujours bénéfique, à part pour la méthode EMSR quand la capacité est supérieure à 250, cette capacité correspond au seuil où pour une capacité plus faible, toutes les demandes ne sont pas réalisables
- Les méthodes saisonnalisées apportent plus de bénéfice que les méthodes non saisonnalisées lorsqu'elles sont en concurrence avec les méthodes saisonnalisées, à part dans le cas de la méthode saisonnalisée Mixte contre la méthode saisonnalisée LRI.
- La *méthode différente par saison* est la méthode qui réussit le mieux quelle que soit la concurrence.

6.4. Conclusion

Ce chapitre propose une manière de prendre en compte la saisonnalité pour faire de la tarification dynamique. L'objectif était d'évaluer l'augmentation du bénéfice des méthodes étudiées précédemment dans le chapitre 5 par rapport aux versions saisonnalisées, où chaque saison correspond à un ensemble de périodes distinctes des autres saisons et est caractérisée par des dynamiques de demande différentes.

L'approche proposée repose sur l'identification de saisons et sur l'utilisation des méthodes de tarification indépendamment dans chacune d'entre elles, soit via une méthode unique appliquée différemment selon les saisons, soit via des méthodes distinctes pour chaque période. Les résultats expérimentaux montrent que la saisonnalité constitue une amélioration pertinente pour toutes les méthodes. En effet, comme illustré dans les figures du chapitre, l'introduction de la saisonnalité permet d'augmenter le bénéfice dans de nombreux cas.

Toutefois, les résultats mettent également en évidence que cet apport dépend fortement du contexte. Par exemple, pour certaines capacités élevées, la saisonnalisation peut dégrader les performances, en particulier pour la méthode EMSR, tandis que pour la méthode LRI, l'effet reste globalement positif. De plus, l'étude montre que l'utilisation d'une méthode unique par saison n'est pas toujours optimale, et que l'approche consistant à utiliser des méthodes différentes selon les saisons peut conduire à de meilleurs résultats globaux.

Les résultats de ce chapitre soulignent l'importance d'intégrer des dimensions temporelles dans les modèles de tarification dynamique. La saisonnalité permet de capturer des variations structurelles de la demande qui ne peuvent pas être appréhendées par des approches homogènes sur l'ensemble de l'horizon. Elle constitue ainsi un enrichissement naturel des modèles de pricing, en particulier dans des contextes où la demande est fortement hétérogène dans le temps.

Chapitre 7
Conclusion

Cette thèse a eu pour but de trouver la meilleure approche pour fixer les prix dans un milieu concurrentiel avec des contraintes de ressources. En particulier, l'objectif était de trouver des méthodes de tarification dynamique utilisant des méthodes d'apprentissage par renforcement pour les appliquer dans le domaine de la location de voitures. La location de voitures est un des domaines qui utilisent des techniques de *Revenue Management* pour fixer les prix. Parmi les domaines utilisant les mêmes méthodes, la location de voitures est un des domaines qui est le plus concurrentiel car il y a peu de différence entre les services vendus, les voitures sont semblables et les lieux pour prendre les véhicules sont identiques pour les gares et les aéroports.

Le chapitre 3 décrit le modèle qui a été utilisé dans le reste de la thèse. Il a été décidé que la réception des demandes se fasse une par une et que les prix soient choisis pour chacune en considérant les contraintes de ressources ainsi que les actions précédentes des joueurs. Le problème de la tarification dynamique a été modélisé comme un jeu répété en considérant que chaque demande correspond à un jeu où les loueurs doivent proposer un prix et qu'ensuite le client doit choisir ou non un loueur. Une étude théorique a ensuite été réalisée dans le cadre de ce jeu répété pour appréhender la présence d'équilibre.

Le chapitre 4 porte sur l'analyse des données obtenues auprès des loueurs grâce à WeYield. L'objectif a tout d'abord été de chercher un lien entre les prix proposés par le loueur et les caractéristiques de la demande : la durée de location, le délai de réservation et la disponibilité de la ressource au moment de la demande. Aucun lien n'a été découvert entre le prix et les caractéristiques de la demande, mais il a été montré que les distributions de ces trois caractéristiques sont liées entre elles. Dans un second temps, le but a été d'analyser la périodicité de l'activité. Pour cela, une méthode pour classifier les jours a été créée et permet de regrouper les jours ayant des demandes similaires. Cette classification, appelée saisonnalité, a servi de base au chapitre 6 pour améliorer les méthodes de tarification définies dans le chapitre 5.

Les chapitres 5 et 6 proposent des méthodes de tarification dynamique pour obtenir le bénéfice le plus élevé possible. Pour cela, dans le chapitre 5, trois méthodes ont été définies. La première est une adaptation pour la location de voitures d'une méthode provenant de l'aérien, EMSR, qui considère seulement la disponibilité des voitures pour fixer un prix. La seconde utilise l'apprentissage par renforcement pour fixer les prix en considérant seulement les décisions passées des concurrents en termes de prix. La troisième méthode est une fusion entre les deux précédentes avec pour but de considérer la disponibilité de la

ressource ainsi que les prix des concurrents. L'évaluation de la performance des méthodes a conclu que la troisième méthode est celle avec le meilleur bénéfice lorsque la capacité des deux loueurs est comprise entre 50% et 110% de la capacité nécessaire pour répondre à toutes les demandes. Si la capacité est au-dessus ou en dessous, il vaut mieux privilégier EMSR.

Le chapitre 6 propose une manière d'intégrer dans les méthodes de tarification dynamique la saisonnalité vue dans le chapitre 4. Cette saisonnalité apporte une connaissance plus générale de la demande aux méthodes de tarification. Il a été choisi d'utiliser une méthode de tarification indépendante par saison, ainsi, les seuils pour la méthode EMSR et les vecteurs d'apprentissage sont différents d'une saison à l'autre même si la méthode utilisée est la même. Ainsi, dans ce contexte, deux éléments ont été définis : la saison qui est associée à chaque demande de location et la mise à jour de cette saisonnalité en fonction des nouveaux résultats. Grâce à tout cela, on obtient des méthodes saisonnalisées qui gagnent en bénéfice par rapport aux méthodes non saisonnalisées lorsque leur concurrent reste le même. Ainsi, la saisonnalité dynamique est très intéressante à considérer en termes de bénéfice.

À la suite de cette thèse, le but sera tout d'abord d'implémenter les différents éléments de cette thèse pour les clients de WeYield. Tout d'abord, l'analyse des données pourra être de nouveau réalisée et approfondie pour chaque client de WeYield. Cela permettra d'avoir un aperçu des données en profondeur en gardant un point de vue indépendant du ressenti du client. Le client pourra alors avoir un aperçu clair des caractéristiques des demandes qui lui apportent le plus de bénéfice. Cela pourrait être, par exemple, les locations touristiques de 7 à 14 jours qui veulent des voitures familiales et qui réservent six mois à l'avance. Ce type d'informations permettra aux clients de WeYield de se concentrer sur les locations qui rapportent le plus de bénéfice et de passer moins de temps sur ce qui rapporte moins de bénéfice.

Le deuxième point intéressant pour les clients sera le calcul de la saisonnalité, car elle peut permettre de découvrir des événements qui ne seraient pas forcément identifiés par le client. De plus, la saisonnalité peut aussi permettre aux clients de challenger leur manière de faire de la tarification dynamique en regroupant des événements qui ne sont pas liés, par exemple le pont de l'Ascension et Halloween, car la distribution de la demande est similaire.

Le dernier point consistera à implémenter les méthodes vues durant cette thèse pour pouvoir conseiller des prix aux clients. Pour cela, il faudra adapter les méthodes pour calculer des grilles tarifaires en amont car la majorité des systèmes actuels ne permettent pas un prix par demande. Il faudra aussi considérer le nombre de types de voitures, car chaque type nécessite une tarification différente due aux prix des concurrents qui sont différents. La phase exploratoire des méthodes d'apprentissage par renforcement devra sûrement être modifiée car les loueurs n'accepteront pas de faire des prix totalement in-

cohérents avec la phase d'exploitation juste pour que l'algorithme vérifie que le prix est toujours inintéressant par rapport au bénéfice.

Concernant les questions de recherche, cette thèse amènerait à étudier la construction de la saisonnalité pour déterminer si l'utilisation de données plus complètes permettrait une meilleure précision et ensuite un meilleur bénéfice cumulé lorsqu'elle est utilisée pour la tarification dynamique. Dans cette thèse, la saisonnalité a été construite à partir des données moyennes des demandes affectant le jour à classifier. Il pourrait être intéressant de faire la classification sur la distribution des données au lieu d'utiliser la moyenne des données. L'évolution de l'utilisation des voitures pourrait aussi être utilisée pour calculer la saisonnalité car elle donne un aperçu du délai de réservation et de l'accélération ou de la décélération du nombre de réservations réalisées par jour. Ces données plus complexes pourraient permettre une saisonnalité plus précise.

Un des prolongements évidents des travaux des chapitres 5 et 6 est l'implémentation de ces méthodes dans d'autres domaines d'application qui utilisent la tarification dans un contexte concurrentiel avec des contraintes de ressources. Le but serait de voir comment ces méthodes s'adaptent à un contexte différent. Effectivement, EMSR s'adaptera bien à l'aérien parce que c'est l'origine de la méthode et la méthode saisonnalisée EMSR devrait être adaptée aussi, mais ce n'est pas forcément évident pour les autres méthodes utilisant l'apprentissage par renforcement qui mettent un poids plus important sur la concurrence que sur la gestion des ressources. D'autres domaines d'application peuvent être considérés, comme l'hôtellerie, la billetterie ou le ferroviaire.

Un autre prolongement de cette thèse, en restant dans le contexte de la location de voitures, serait d'améliorer le bénéfice cumulé obtenu avec les algorithmes de tarification dynamique. Pour cela, il serait envisageable d'utiliser des données extérieures pour obtenir plus de connaissances, comme les données de l'aérien et de l'hôtellerie pour avoir une idée des demandes de location futures.

Une autre idée serait l'utilisation d'autres algorithmes comme des algorithmes de *Deep Reinforcement Learning*. Pour cela, il faudrait évaluer dans un cadre réel si les données suffisent pour l'apprentissage des méthodes de *Deep Reinforcement Learning* qui en nécessitent davantage que le *Reinforcement Learning*. Les algorithmes de *Deep Reinforcement Learning* acceptent un ensemble de caractéristiques par demande plus important en entrée, ce qui permettrait à l'algorithme de prendre des décisions avec plus de connaissances et ainsi de prendre potentiellement de meilleures décisions.

Références

[1] Peter Auer, Nicolò Cesa-Bianchi et Paul Fischer. « Finite-time Analysis of the Multiarmed Bandit Problem ». In : *Machine Learning* 47 (mai 2002), p. 235-256. DOI : 10.1023/A:1013689704352.

[2] P. Belobaba, A. Odoni et C. Barnhart. *The Global Airline Industry*. Cambridge : John Wiley & Sons, 2015.

[3] Krishna V Benoit J-P. « Finitely repeated games ». In : *Econometrica* (1985).

[4] Joseph Bertrand. « Review of Walras's Théorie mathématique de la richesse sociale and Cournot's Recherches sur les principes mathématiques de la théorie des richesses ». In : *Cournot Oligopoly : Characterization and Applications*. Sous la dir. d'Andrew F.Editor Daughety. Cambridge University Press, 1989, p. 73-81.

[5] Omar Besbes et Assaf Zeevi. « Dynamic Pricing Without Knowing the Demand Function : Risk Bounds and Near-Optimal Algorithms ». In : *Operations Research* 57 (déc. 2009), p. 1407-1420. DOI : 10.1287/opre.1080.0640.

[6] J. M. Betancourt et al. « Dynamic price competition with capacity constraints ». In : *NBER* (2024).

[7] Ioana C. Bilegan et al. « Revenue management for rail container transportation ». In : *EURO Journal on Transportation and Logistics* 4.2 (2015), p. 261-283. ISSN : 2192-4376. DOI : https://doi.org/10.1007/s13676-014-0051-7. URL : https://www.sciencedirect.com/science/article/pii/S2192437620301254.

[8] Gabriel Bitran et René Caldentey. « An Overview of Pricing Models for Revenue Management ». In : *Manufacturing & Service Operations Management* 5.3 (2003), p. 203-229. DOI : 10.1287/msom.5.3.203.16031. eprint : https://doi.org/10.1287/msom.5.3.203.16031. URL : https://doi.org/10.1287/msom.5.3.203.16031.

[9] Océane BOCA. *Secteurs éligibles à l'application du revenue management*. 2017. URL : https://www.aerogestion.fr/secteurs-eligibles-revenue-management.

[10] A. Borodin et R. El-Yaniv. *Online computation and competitive analysis*. Cambridge University Press, 2005.

[11] Youssef Boutahri et Amine Tilioua. « Economic model predictive control for energy-efficient residential heating under dynamic electricity pricing ». In : *Results in Engineering* 29 (2026), p. 109765. ISSN : 2590-1230. DOI : https://doi.org/10.1016/j.rineng.2026.109765. URL : https://www.sciencedirect.com/science/article/pii/S2590123026008042.

[12] E. Andrew Boyd et Ioana C. Bilegan. « Revenue Management and E-Commerce ». In : *Management Science* 49.10 (2003), p. 1363-1386. ISSN : 00251909, 15265501. URL : http://www.jstor.org/stable/4134011 (visité le 28/05/2026).

[13] Yining Chen et coauthors. « Contextual Dynamic Pricing : Algorithms, Optimality, and Regret Bounds ». In : *arXiv preprint arXiv :XXXX.XXXXX* (2020).

[14] Massimiliano Costacurta. *Dynamic Pricing with Multi-Armed Bandit : Learning by Doing*. URL : https://towardsdatascience.com/dynamic-pricing-with-multi-armed-bandit-learning-by-doing-3e4550ed02ac/.

[15] R. G. Cross. *Revenue Management : Hard-Core Tactics for Market Domination*. Michigan : Broadway Books, 1997.

[16] Andrea Danesi et coauthors. « Pricing and Demand in Car Rental Market : An Empirical Study ». In : *Procedia Computer Science* 212 (2026), p. 123-132. DOI : 10.1016/j.procs.2022.11.XXX.

[17] Guillaume Delclit et Julien Finez. *Le yield management à la SNCF*. Rapport de recherche. Consulté le 29 avril 2026. France : Programme FERINTER, 2013. URL : https://ferinter.hypotheses.org/files/2015/03/delclite-et-finez-02-12-2013-colloque-ferinter-vf-1.pdf.

[18] W. Elmaghraby et Pinar Keskinocak. « Dynamic Pricing in the Presence of Inventory Considerations : Research Overview, Current Practices, and Future Directions ». In : *Management Science* 49 (oct. 2003), p. 1287-1309. DOI : 10.1287/mnsc.49.10.1287.17315.

[19] Kris Johnson Ferreira, David Simchi-Levi et He Wang. « Online Network Revenue Management Using Thompson Sampling ». In : *Operations Research* 66.6 (2018), p. 1586-1602. DOI : 10.1287/opre.2018.1755. eprint : https://doi.org/10.1287/opre.2018.1755. URL : https://doi.org/10.1287/opre.2018.1755.

[20] Paul Friedrich, Barna Pásztor et Giorgia Ramponi. « Learning Collusion in Episodic, Inventory-Constrained Markets ». In : *Eighteenth European Workshop on Reinforcement Learning*. 2025. URL : https://openreview.net/forum?id=tlgyVVSDRq.

[21] D. Fudenberg et D. Levine. « Subgame-perfect equilibria of finite- and infinite-horizon games ». In : *Journal of Economic Theory* (1983).

[22] G. Gallego et H. Topaloglu. *Revenue Management and Pricing Analytics*. New York : Springer, 2019.

[23] Guillermo Gallego et Garrett van Ryzin. « Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons ». In : *Management Science* 40.8 (1994), p. 999-1020. ISSN : 00251909, 15265501. URL : http://www.jstor.org/stable/2633090 (visité le 15/04/2026).

[24] Ravi Ganti et al. « Thompson Sampling for Dynamic Pricing ». In : *arXiv e-prints*, arXiv :1802.03050 (fév. 2018), arXiv :1802.03050. DOI : 10.48550/arXiv.1802.03050. arXiv : 1802.03050 [stat.ML].

[25] Rui Gao, Xiaolei Li et Yingdong Lu. « Bid price controls for car rental network revenue management ». In : *Production and Operations Management* 32.5 (2023), p. 1245-1265. DOI : 10.1111/poms.13836.

[26] M. K. Geraghty et Ernest Johnson. « Revenue Management Saves National Car Rental ». In : *Interfaces* 27.1 (1997), p. 107-127. ISSN : 00922102, 1526551X. URL : http://www.jstor.org/stable/25062214 (visité le 15/04/2026).

[27] T. Golla et al. « Dynamic Pricing of Product and Delivery Time in Multi-Variant Production Using an Actor-Critic Reinforcement Learning ». In : *CIRP Annals* 72.1 (2023), ???-??? URL : https://www.sciencedirect.com/science/article/abs/pii/S0007850623000653.

[28] Laurie Guenin, Dominique Barth et Christian Cadere. *Reinforcement Learning for Dynamic Pricing with Resource Constraints in a Competitive Context*. Jan. 2026. DOI : 10.1007/978-3-032-09156-7_11.

[29] Marco Guerzoni, Luigi Riso et M. Grazia Zoia. « Extreme weather events as the main driver of electricity price volatility in Italy : A GARCH-MIDAS approach with machine learning-based variable selection ». In : *The North American Journal of Economics and Finance* 81 (2026), p. 102512. ISSN : 1062-9408. DOI : https://doi.org/10.1016/j.najef.2025.102512. URL : https://www.sciencedirect.com/science/article/pii/S1062940825001524.

[30] P. Hernandez-Leal et al. *A Survey of Learning in Multiagent Environments : Dealing with Non-Stationarity*. Juill. 2017. DOI : 10.48550/arXiv.1707.09183.

[31] Ali Hortaçsu, Aniko Öry et Kevin R. Williams. « Dynamic Price Competition with Capacity Constraints ». In : *NBER Working Paper Series* 32673 (2024). URL : https://www.nber.org/papers/w32673.

[32] Ali Hortaçsu, Aniko Öry et Kevin R. Williams. « Dynamic Price Competition : Theory and Evidence from Airline Markets ». In : *NBER Working Paper Series* 30347 (2022). URL : https://www.nber.org/papers/w30347.

[33] Lawrence J. Hubert et Phipps Arabie. « Comparing partitions ». In : *Journal of Classification* 2 (1985), p. 193-218. URL : https://api.semanticscholar.org/CorpusID:189915041.

[34] A. Ismail, M. Baysal et al. « Dynamic Pricing Based on Demand Response Using Actor–Critic Agent Reinforcement Learning ». In : *Energies* 16.14 (2023), p. 5469. URL : https://www.mdpi.com/1996-1073/16/14/5469.

[35] MacQueen J. *Some methods for classification and analysis of multivariate observations*. 1967. URL : https://api.semanticscholar.org/CorpusID:6278891.

[36] Paul Jaccard. *Etude comparative de la distribution florale dans une portion des Alpes et des Jura*. URL : https://api.semanticscholar.org/CorpusID:135345056.

[37] Junfeng Jiao. « Investigating Uber price surges during a special event in Austin, TX ». In : *Research in Transportation Business & Management* 29 (2018). Special Issue title : [RE]EVALUATING HOW WE VALUE TRANSPORTATION, p. 101-107. ISSN : 2210-5395. DOI : https://doi.org/10.1016/j.rtbm.2018.02.008. URL : https://www.sciencedirect.com/science/article/pii/S2210539517301165.

[38] A. Kastius et R. Schlosser. « Dynamic pricing under competition using reinforcement learning ». In : *Journal of Revenue and Pricing Management* (2022). DOI : 10.1057/s41272-021-00285-3.

[39] M. N. Katehakis et A. F. Veinott. « The Multi-Armed Bandit Problem : Decomposition and Computation ». In : *Mathematics of Operations Research* 12.2 (1987), p. 262-268. DOI : 10.1287/moor.12.2.262. URL : https://doi.org/10.1287/moor.12.2.262.

[40] Sheryl E. Kimes. « Yield management : A tool for capacity-considered service firms ». In : *Journal of Operations Management* 8.4 (1989), p. 348-363. ISSN : 0272-6963. DOI : https://doi.org/10.1016/0272-6963(89)90035-1. URL : https://www.sciencedirect.com/science/article/pii/0272696389900351.

[41] V. R. Konda et J. N. Tsitsiklis. « Actor-Critic Algorithms ». In : *Advances in Neural Information Processing Systems* 12 (2000), p. 1008-1014.

[42] T.L Lai et Herbert Robbins. « Asymptotically efficient adaptive allocation rules ». In : *Advances in Applied Mathematics* 6.1 (1985), p. 4-22. ISSN : 0196-8858. DOI : https://doi.org/10.1016/0196-8858(85)90002-8. URL : https://www.sciencedirect.com/science/article/pii/0196885885900028.

[43] Luca Lambertini, Andrea Mantovani et Alessandro Tampieri. « The dynamics of Bertrand price competition with cost-reducing investments ». In : *International Review of Economics* 60.1 (2013), p. 25-41. DOI : 10.1007/s12232-012-0168-x.

[44] Dong Li, Zhan Pang et Lixian Qian. « Bid Price Controls for Car Rental Network Revenue Management ». In : *SSRN Electronic Journal* (jan. 2021). DOI : 10.2139/ssrn.3933032.

[45] Ken Littlewood. « Special Issue Papers : Forecasting and control of passenger bookings ». In : *Journal of Revenue and Pricing Management* 4 (avr. 2005), p. 111-123. DOI : 10.1057/palgrave.rpm.5170134.

[46] J. Liu et al. « Non-Stationary Dynamic Pricing Via Actor-Critic Information-Directed Pricing ». In : *arXiv preprint* arXiv :2208.09372 (2022). URL : https://arxiv.org/abs/2208.09372.

[47] JEFFREY I. McGILL et GARRETT J. VAN RYZIN. « Revenue Management : Research Overview and Prospects ». In : *Transportation Science* 33.2 (1999), p. 233-256. ISSN : 00411655, 15265447. URL : http://www.jstor.org/stable/25768862 (visité le 15/04/2026).

[48] S. MEATCHI. « La tarification par le revenue management : de la mesure de l'injustice perçue à la validation d'un modèle de " Fairness Based Pricing " ». Thèse de doct. Université d'Angers, 2019.

[49] Sourou MEATCHI et Sandra CAMUS. « Revenue management pricing in the hotel sector : Reducing perceived unfairness to encourage willingness to pay ». In : *Recherche et Applications en Marketing (English Edition)* 35 (sept. 2020), p. 102-123. DOI : 10.1177/2051570720954760.

[50] M. MIHAILESCU et Y. M. TEO. « Dynamic Resource Pricing on Federated Clouds ». In : *10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing* (2010).

[51] K. S. NARENDRA et M. A. L. THATHACHAR. « Learning Automata : An Introduction ». 1989.

[52] J NASH. « Equilibrium Points in n-Person Games ». In : *Proceedings of the National Academy of Sciences of the United States of America* 36.1 (1950), p. 48-49.

[53] Beatriz Brito OLIVEIRA, Maria Antónia CARRAVILLA et José Fernando OLIVEIRA. « Fleet and revenue management in car rental companies : A literature review and an integrated conceptual framework ». In : *Omega* 71 (2017), p. 11-26. ISSN : 0305-0483. DOI : https://doi.org/10.1016/j.omega.2016.08.011. URL : https://www.sciencedirect.com/science/article/pii/S0305048316300342.

[54] A. PÉREZ-SUÁREZ et al. « OClustR : A new graph-based algorithm for overlapping clustering ». In : *Neurocomputing* (2013), p. 234-247.

[55] Martin PETRICEK, Štěpán CHALUPA et Věra LEVIČKOVÁ. « Comparison of expected marginal revenue models in the hospitality industry ». In : *Journal of Revenue and Pricing Management* 21 (juin 2022). DOI : 10.1057/s41272-021-00328-9.

[56] Robert L PHILLIPS. *Pricing and Revenue Optimization*. eng. Second edition. Redwood City, CA, USA : Stanford University Press, 2021. ISBN : 1503610004.

[57] M.L. Puterman. *Markov Decision Processes : Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Statistics. Wiley, 1994. ISBN : 9780471619772. URL : https://books.google.fr/books?id=tsiiQgAACAAJ.

[58] Wenchuan Qiao et al. « Distributed dynamic pricing of multiple perishable products using multi-agent reinforcement learning ». In : *Expert Systems with Applications* 237 (2024), p. 121252. ISSN : 0957-4174. DOI : https://doi.org/10.1016/j.eswa.2023.121252. URL : https://www.sciencedirect.com/science/article/pii/S0957417423017542.

[59] William M. Rand. « Objective Criteria for the Evaluation of Clustering Methods ». In : *Journal of the American Statistical Association* 66.336 (1971), p. 846-850. DOI : 10.1080/01621459.1971.10482356. eprint : https://www.tandfonline.com/doi/pdf/10.1080/01621459.1971.10482356. URL : https://www.tandfonline.com/doi/abs/10.1080/01621459.1971.10482356.

[60] J. Schulman et al. « Proximal Policy Optimization Algorithms ». In : *arXiv preprint* arXiv :1707.06347 (2017). URL : https://arxiv.org/abs/1707.06347.

[61] Barry C. Smith, John F. Leimkuhler et Ross M. Darrow. « Yield Management at American Airlines ». In : *Interfaces* 22.1 (1992), p. 8-31. ISSN : 00922102, 1526551X. URL : http://www.jstor.org/stable/25061571 (visité le 15/04/2026).

[62] R. S. Sutton et A. G. Barto. *Reinforcement Learning : An Introduction. Second Edition*. Cambridge : MIT Press, 2018.

[63] Richard S. Sutton et al. « Policy Gradient Methods for Reinforcement Learning with Function Approximation ». In : *Advances in Neural Information Processing Systems 12*. MIT Press, 1999, p. 1057-1063. URL : https://papers.nips.cc/paper/1999/file/464d828b85b0bed98e80ade0a5c43b0f-Paper.pdf.

[64] K. T. Talluri et G. J. Ryzin. *The theory and practice of revenue management*. New York : Springer, 2004.

[65] William R. Thompson. « On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples ». In : *Biometrika* 25.3/4 (1933), p. 285-294. ISSN : 00063444. URL : http://www.jstor.org/stable/2332286 (visité le 22/04/2026).

[66] Michel Tokic. « Adaptive epsilon-Greedy Exploration in Reinforcement Learning Based on Value Difference ». In : *Deutsche Jahrestagung für Künstliche Intelligenz (KI 2010)*. VDBE : Value-Difference Based Exploration. Springer, 2010, p. 203-210.

[67] Francesco Trovò et al. « Improving multi-armed bandit algorithms in online pricing settings ». In : *International Journal of Approximate Reasoning* (2018). Preprint version available.

[68] G. Tuncay et al. « A reinforcement learning based dynamic room pricing model for hotel industry ». In : *INFOR Information Systems and Operational Research* (2023).

[69] Christopher Watkins. « Learning From Delayed Rewards ». Thèse de doct. King's College, jan. 1989.

[70] Larry Weatherford et Samuel Bodily. « A Taxonomy and Research Overview of Perishable-Asset Revenue Management : Yield Management, Overbooking, and Pricing ». In : *Operations Research* 40 (oct. 1992), p. 831-844. doi : 10.1287/opre.40.5.831.

[71] C. Wu. « Evaluating and Selecting Deep Reinforcement Learning Models for Optimal Dynamic Pricing : A Systematic Comparison of PPO, DDPG, and SAC ». Thèse de doct. University of Liverpool, 2024. url : https://livrepository.liverpool.ac.uk/3176825/.

[72] C. Wu, W. Bi et H. Liu. « Proximal Policy Optimization Algorithm for Dynamic Pricing with Online Reviews ». In : *Expert Systems with Applications* 220 (2023), p. 119747. url : https://www.sciencedirect.com/science/article/abs/pii/S0957417422022096.

[73] Y. Zhai et Q. Zhao. « Oligopoly dynamic pricing : A repeated game with incomplete information ». In : *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2016).

[74] Z. Zhang. « Dynamic Pricing Strategy Optimization Based on a Reinforcement Learning PPO Algorithm : An Empirical Study on Ride-Hailing Platforms ». In : *International Journal of E-Business Research* ???? (2024), ?? url : https://www.igi-global.com/article/dynamic-pricing-strategy-optimization-based-on-a-reinforcement-learning-ppo-algorithm/406688.

[75] Xinghui ZHU et al. « Reinforcement learning for Multi-Flight Dynamic Pricing ». In : *Computers & Industrial Engineering* 193 (2024), p. 110302. ISSN : 0360-8352. DOI : https://doi.org/10.1016/j.cie.2024.110302. URL : https://www.sciencedirect.com/science/article/pii/S0360835224004236.

Annexe A
Ensemble des courbes du chapitre 5 qui seront comparées dans le chapitre 6

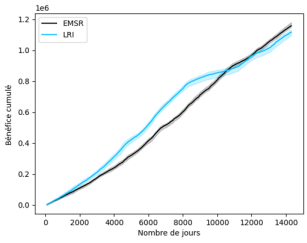

FIGURE A.1 – LRI contre EMSR

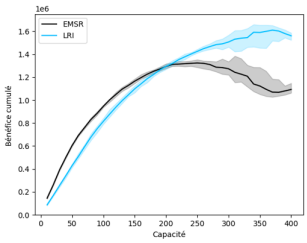

FIGURE A.2 – Évolution du bénéfice pour LRI contre EMSR en fonction de la capacité

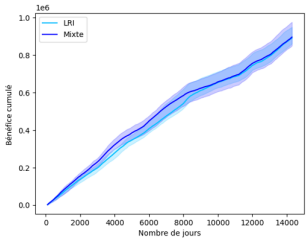

FIGURE A.3 – LRI contre méthode mixte

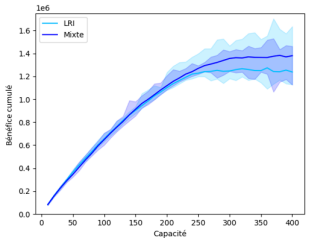

FIGURE A.4 – Évolution du bénéfice pour LRI contre la méthode mixte en fonction de la capacité

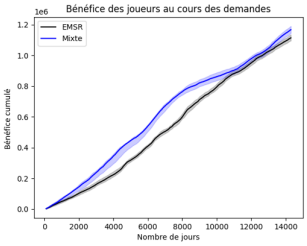

FIGURE A.5 – Méthode mixte contre EMSR

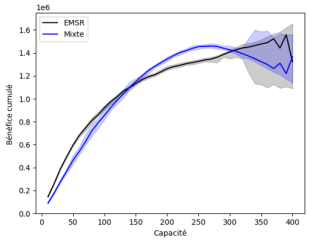

FIGURE A.6 – Évolution du bénéfice pour EMSR contre la méthode mixte en fonction de la capacité

Annexe B
Évaluation de performances du Chapitre 5 sur d'autres ensembles de données

Les études suivantes ont été réalisées dans le même cadre que celles du chapitre 5 en changeant l'ensemble des demandes par ceux décrits dans le chapitre 4.2 qui sont les données d'un loueur société, d'un loueur continental et d'un loueur de Van.

B.1. Loueur société

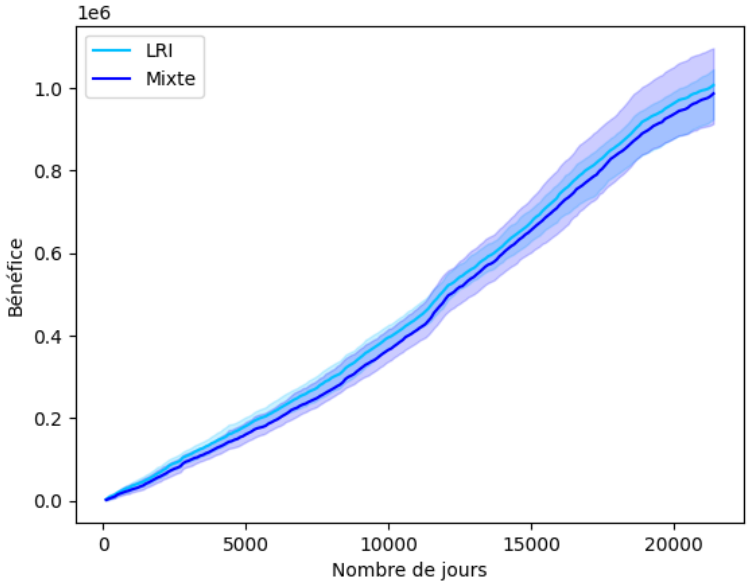

FIGURE B.1 – LRI contre méthode mixte pour le loueur société

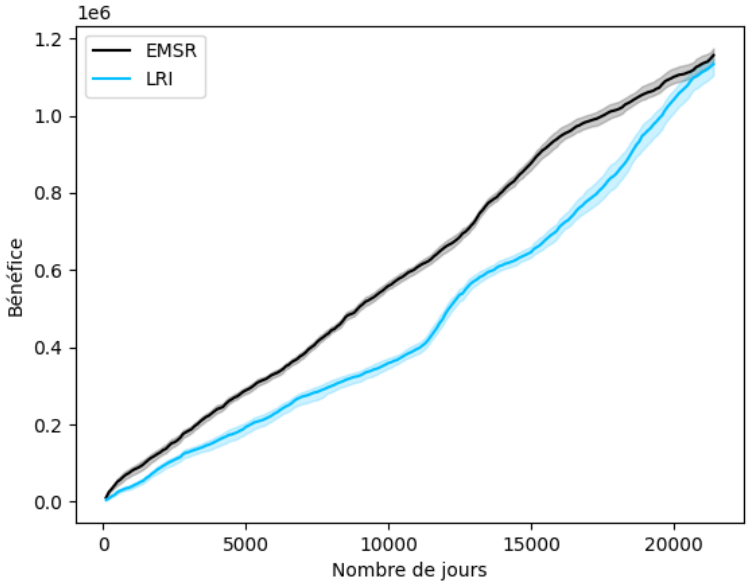

FIGURE B.2 – LRI contre EMSR pour le loueur société

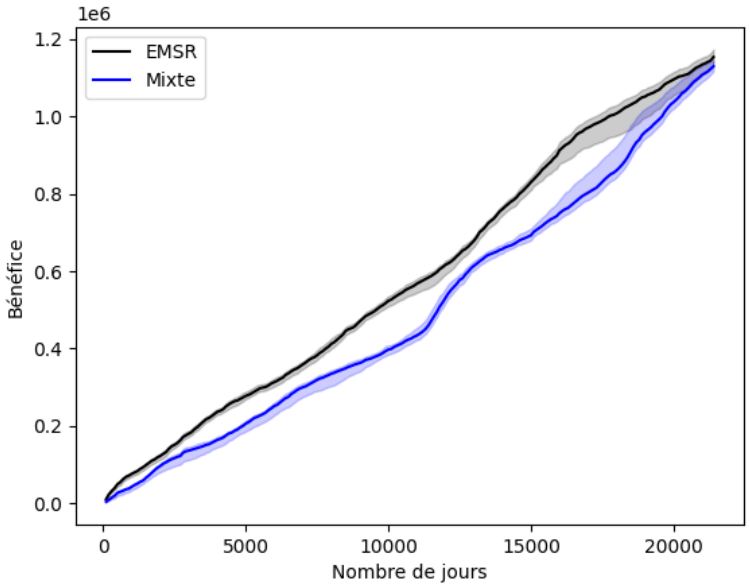

FIGURE B.3 – EMSR contre méthode mixte pour le loueur société

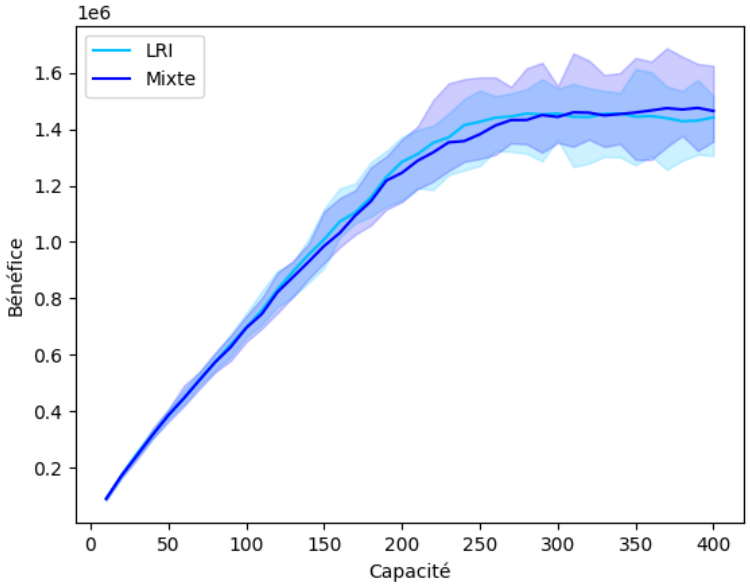

FIGURE B.4 – Évolution du bénéfice pour LRI contre la méthode mixte en fonction de la capacité pour le loueur société

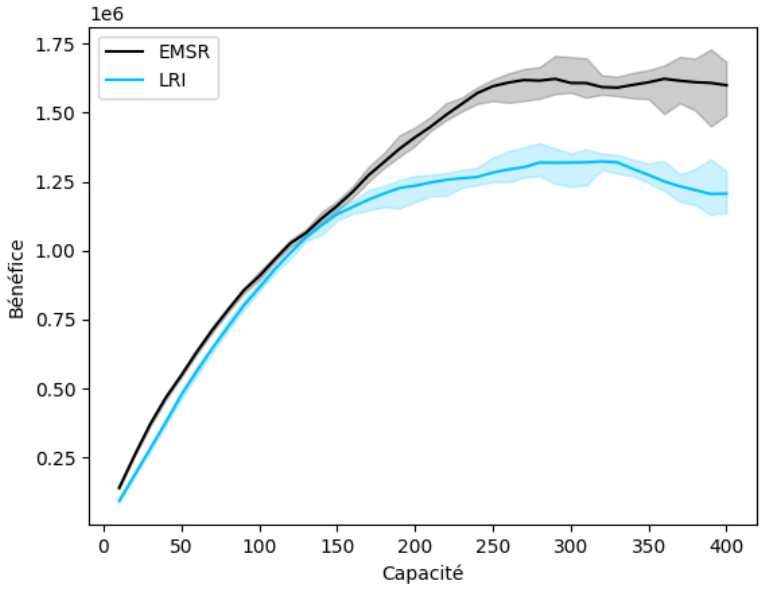

FIGURE B.5 – Évolution du bénéfice pour LRI contre EMSR en fonction de la capacité pour le loueur société

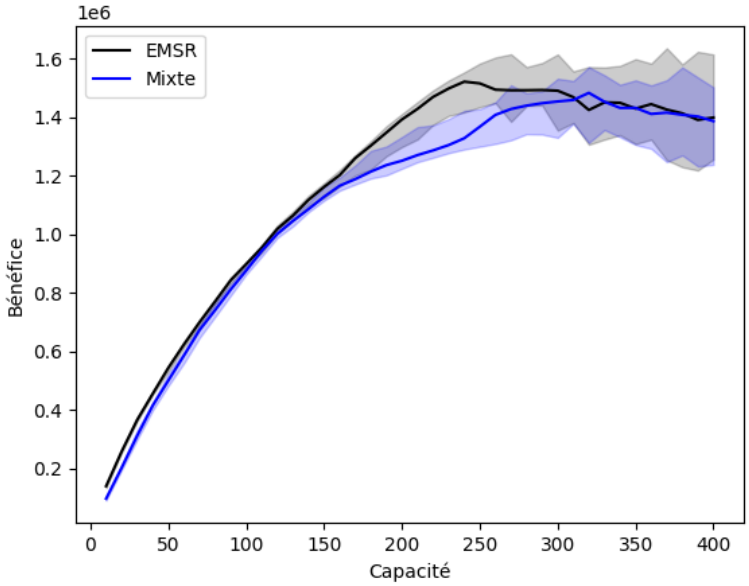

FIGURE B.6 – Évolution du bénéfice pour EMSR contre la méthode mixte en fonction de la capacité pour le loueur société

B.2. Loueur continental

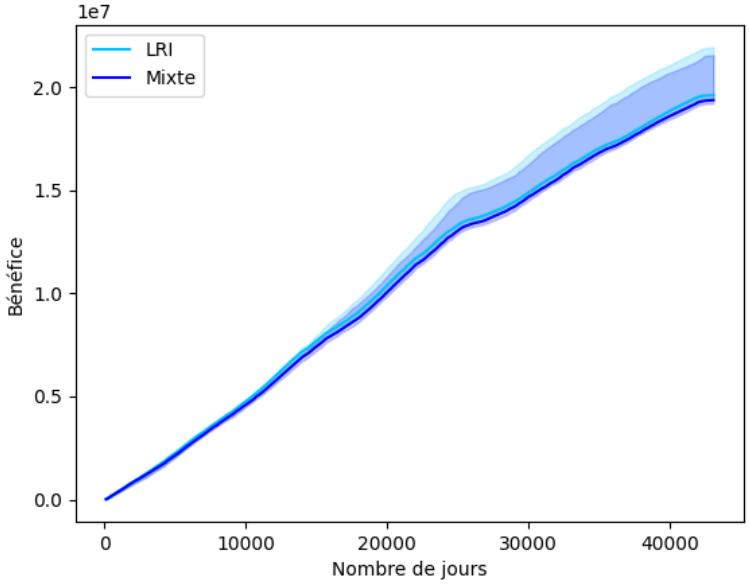

FIGURE B.7 – LRI contre méthode mixte pour le loueur continental

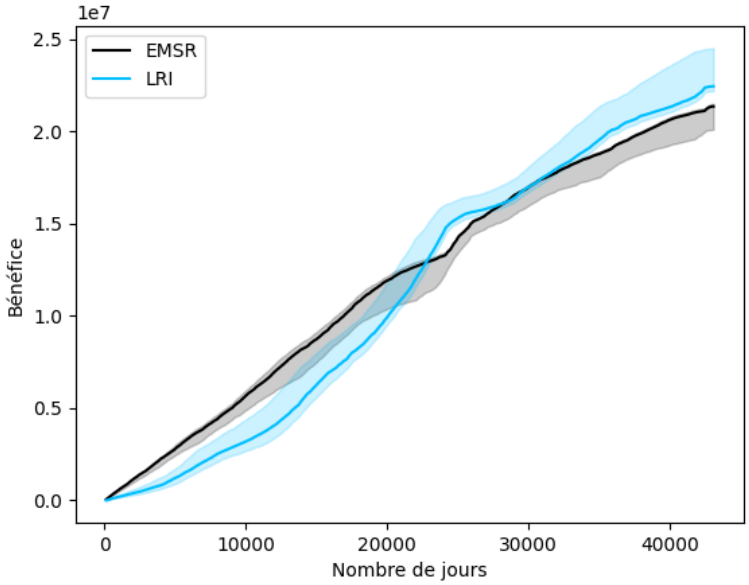

FIGURE B.8 – LRI contre EMSR pour le loueur continental

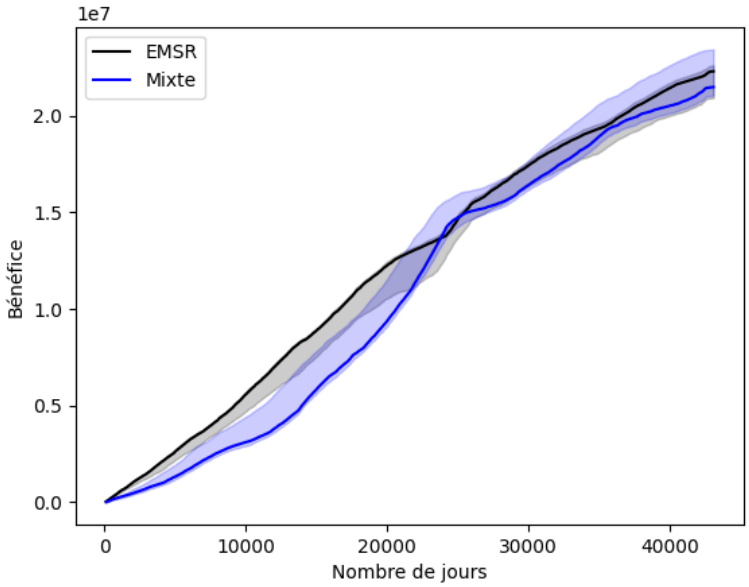

FIGURE B.9 – EMSR contre méthode mixte pour le loueur continental

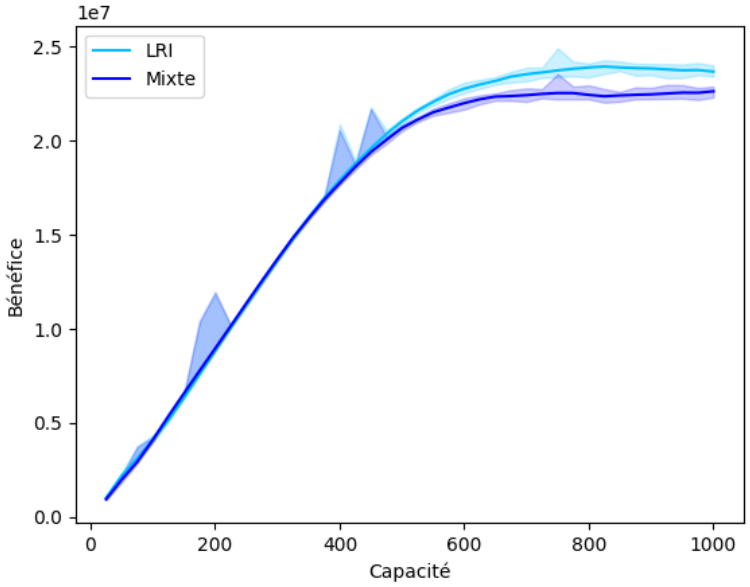

FIGURE B.10 – Évolution du bénéfice pour LRI contre la méthode mixte en fonction de la capacité pour le loueur continental

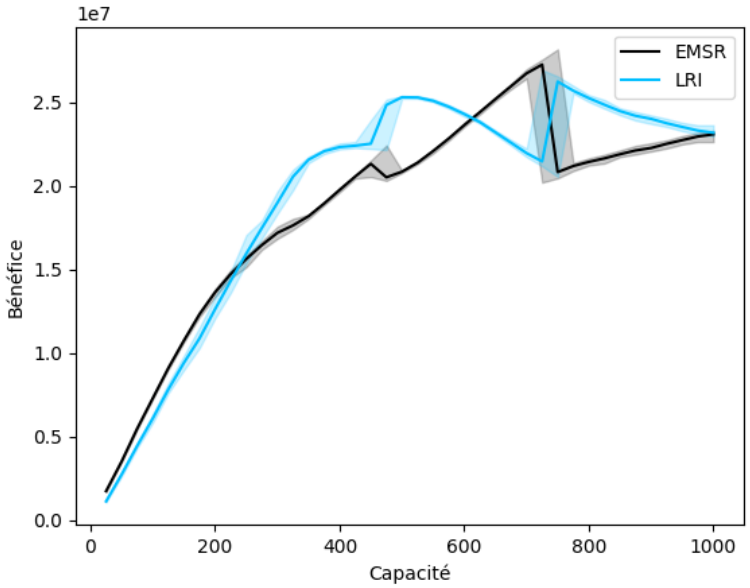

FIGURE B.11 – Évolution du bénéfice pour LRI contre EMSR en fonction de la capacité pour le loueur continental

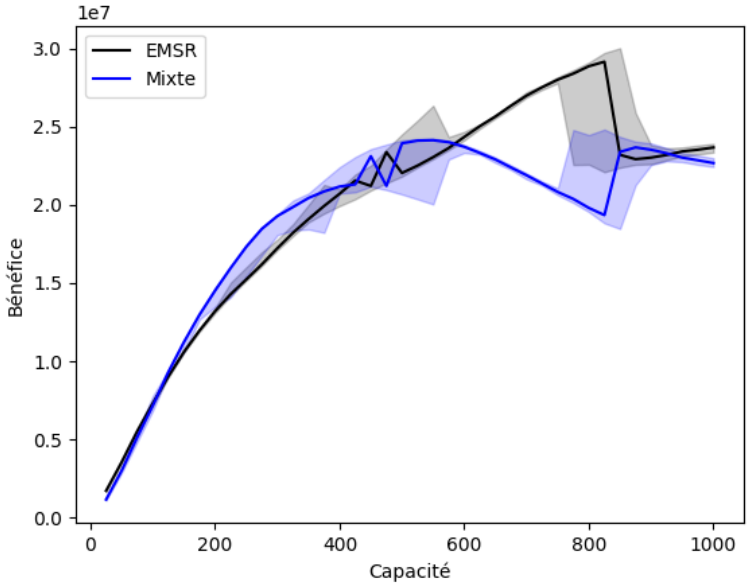

FIGURE B.12 – Évolution du bénéfice pour EMSR contre la méthode mixte en fonction de la capacité pour le loueur continental

B.3. Loueur de Van

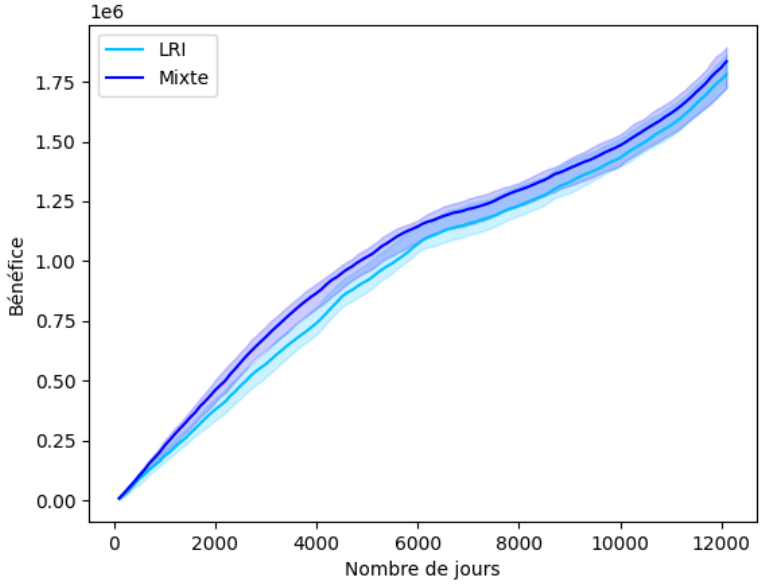

FIGURE B.13 – LRI contre méthode mixte pour le loueur de Van

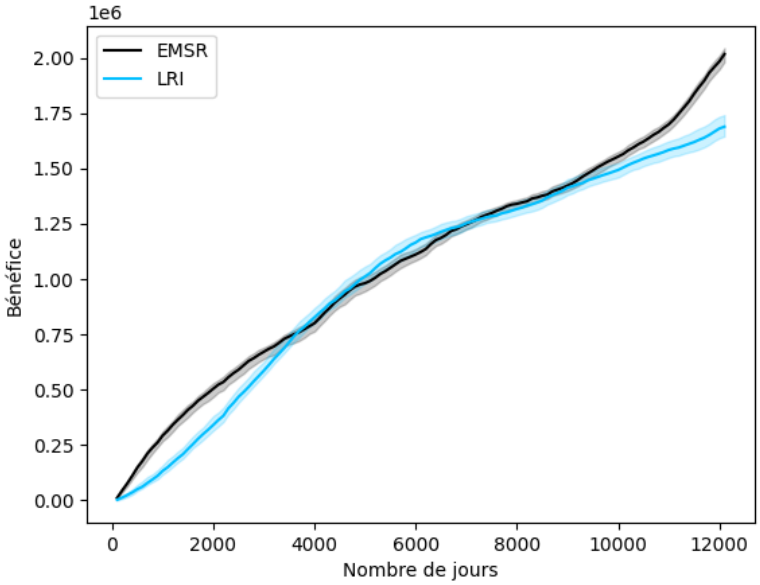

FIGURE B.14 – LRI contre EMSR pour le loueur de Van

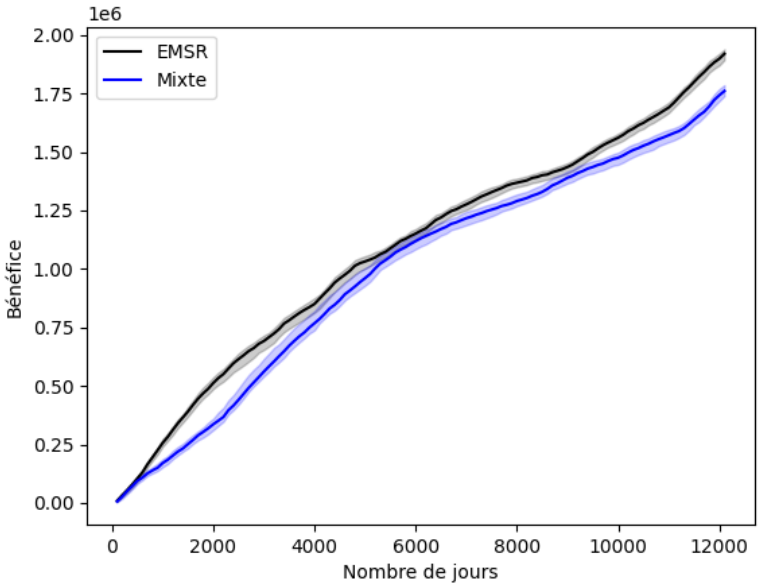

FIGURE B.15 – EMSR contre méthode mixte pour le loueur de Van

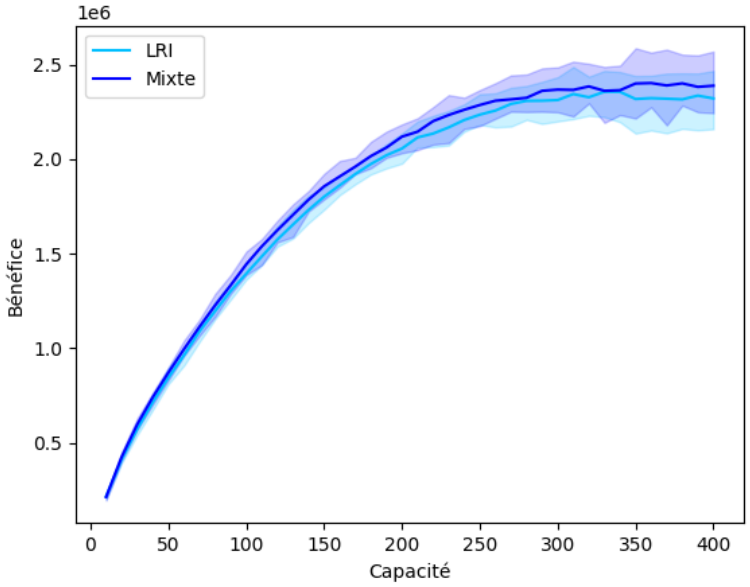

FIGURE B.16 – Évolution du bénéfice pour LRI contre la méthode mixte en fonction de la capacité pour le loueur de Van

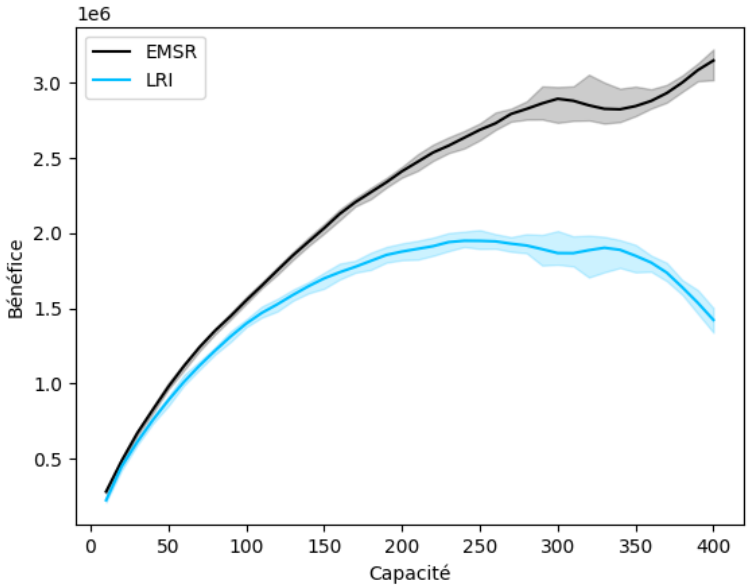

FIGURE B.17 – Évolution du bénéfice pour LRI contre EMSR en fonction de la capacité pour le loueur de Van

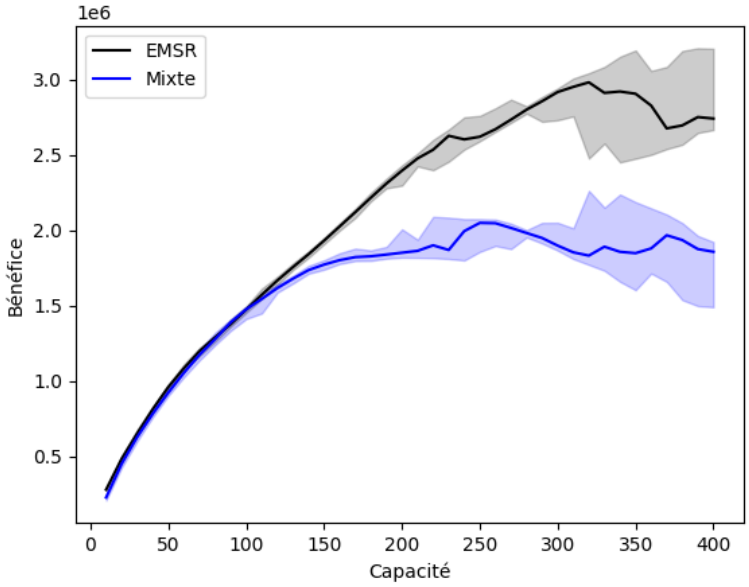

FIGURE B.18 – Évolution du bénéfice pour EMSR contre la méthode mixte en fonction de la capacité pour le loueur de Van

Annexe C
Ensemble des courbes d'une autre exécution pour comparer avec celles de la section 6.3.1

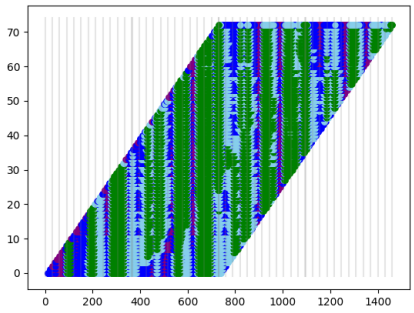

FIGURE C.1 – Évolution de la saisonnalité

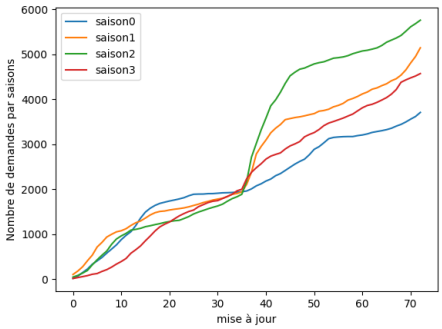

FIGURE C.2 – Évolution du nombre de demandes par saison

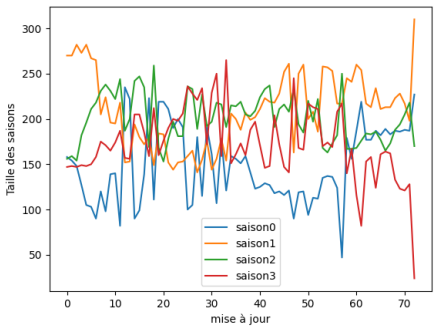

FIGURE C.3 – Évolution de la taille des saisons

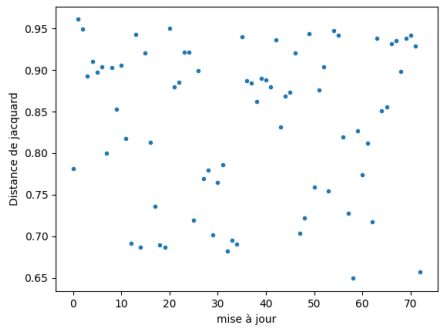

FIGURE C.4 – Indice de Jacquard à chaque mise à jour de la saisonnalité