



Modélisation d'affrontements de la LNH par processus de Poisson: Gains potentiels des paris sportifs

Simon Lalancette

Essai présenté
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de maîtrise en statistique
pour l'obtention du grade de Maître ès sciences (M.Sc.)

Québec, Canada

© Simon Lalancette, Décembre 2016

Résumé

L'industrie du jeu fascine des populations entières en donnant la chance à quiconque d'augmenter ses richesses. Que ce soit avec les cotes au casino, ou dans les cotes des paris sportifs, le risque de perdre est parfois sous-estimé par les joueurs. Dans un sport, comme le hockey sur glace, la majorité des paris se font sur l'issue d'un match, à savoir quelle sera l'équipe gagnante. Bien que plusieurs études eussent cherché à prédire les affrontements avec la meilleure précision possible, aucune n'a clairement tenté d'estimer, monétairement, la valeur ajoutée que promet leur(s) modèle(s). Dans cet essai, il sera question de développer nos propres modèles de simulations d'affrontements. Un premier modèle s'inspirant de l'oeuvre de [Dayaratna et Miller \(2013\)](#), et trois autres modèles, qui comme [Merritt et Clauset \(2014\)](#), considèrent le processus de Poisson pour modéliser le rythme des buts dans une partie de hockey. Les deux derniers modèles utilisent des modèles linéaires généralisés avec lien Poisson et tiennent compte de l'impact que peut avoir un gardien de but partant sur l'issue d'un match. Ici nous ne nous concentrons que sur la modélisation des affrontements intradivisions afin d'éviter toutes formes de biais générés par des rencontres moins fréquentes. Donc, ce travail n'est qu'une première étape vers l'élaboration d'un modèle plus complet, qui permettrait de profiter des faiblesses des cotes, construites par les preneurs de paris, en accumulant les décisions éclairées.

Table des matières

Résumé	ii
Table des matières	iii
Liste des tableaux	iv
Liste des figures	vi
Introduction	1
1 Survol de la littérature	3
1.1 Différentes modélisations appliquées aux sports d'équipes	3
1.2 Différentes modélisations appliquées au hockey	5
2 Exploration des bases de données	9
2.1 Sources de données pour le hockey	9
2.2 Sources de données pour les paris sportifs	14
3 Élaboration des modèles de simulation	16
3.1 Construction du premier modèle de simulation	16
3.2 Construction du deuxième, troisième et quatrième modèles de simulation . .	22
3.3 Bilan des quatre modèles de simulation	39
4 Simulations et discussion	40
4.1 Estimations des paramètres	40
4.2 Simulations des modèles par la méthode de Monte-Carlo	43
4.3 Sources d'erreur dans les prédictions et corrections suggérées	48
4.4 Théorie de la décision dans les paris sportifs	50
Conclusion	56
A Division et abréviation officielle de chaque équipe de la LNH	58
B Autres tableaux omis du document	59
C Affrontements exclus des paris intelligents	70
Bibliographie	71

Liste des tableaux

2.1	Extrait de la base de données des affrontements intra-associations.	11
2.2	Extrait de la base de données alternative des affrontements intra-associations. .	12
2.3	Exemple de gains selon les différentes définitions des cotes.	14
2.4	Extrait de la base de données des cotes des affrontements de la division Atlantique.	15
3.1	Paramètres estimés des Weibull pour un affrontement entre les Maple Leafs et les Canadiens, à Montréal, pour la saison 2015-2016.	20
3.2	Extrait de la fonction de masse conjointe des buts marqués lors d'un affrontement entre les Maple Leafs et les Canadiens à Montréal pendant la saison 2015-2016.	22
3.3	Intensités offensives et défensives des Canadiens de Montréal pour la première moitié de la saison 2015-2016.	28
3.4	Intensités offensives et défensives estimées par le deuxième modèle pour un affrontement entre les Maple Leafs et les Canadiens, à Montréal, pour la saison 2015-2016.	29
3.5	Intensités des processus de Poisson estimés par le deuxième modèle pour un affrontement entre les Maple Leafs et les Canadiens, à Montréal, pour la saison 2015-2016.	30
3.6	Exemple d'affrontement simulé à partir des processus de Poisson entre les Maple Leafs et les Canadiens, à Montréal, pour la deuxième moitié de la saison 2015-2016.	30
3.7	Moyenne de buts accordés estimés par minute des différents gardiens des Canadiens de Montréal pour la saison 2015-2016.	34
3.8	Extrait des estimations des paramètres du modèle de régression de Poisson pour l'association de l'Est au cours de la saison 2015-2016.	34
3.9	Intensités estimées par le troisième modèle pour un affrontement entre les Maple Leafs avec Jonathan Bernier et les Canadiens avec Mike Condon, pour la saison 2015-2016.	35
3.10	Moyenne de buts accordés estimés par minute des différents gardiens des Canadiens de Montréal lors des 55 premières minutes des matchs pour la saison 2015-2016.	37
3.11	Moyenne de buts accordés estimés par minute des différents gardiens des Canadiens de Montréal lors des 5 dernières minutes des matchs pour la saison 2015-2016.	37
3.12	Extrait des estimations des paramètres du quatrième modèle de simulation pour l'association de l'Est au cours de la saison 2015-2016.	38
3.13	Intensités estimées par le quatrième modèle pour un affrontement entre les Maple Leafs avec Jonathan Bernier et les Canadiens avec Mike Condon, pour la saison 2015-2016.	39

3.14	Approximation de diverses statistiques, par modèle pour l'affrontement du 27 février 2016 entre les Canadiens de Montréal et les Maple Leafs de Toronto. . .	39
4.1	Nombre d'affrontements dans les blocs de calibrage des modèles.	41
4.2	Extrait des paramètres estimés du premier modèle, pour la première moitié de la saison 2015-2016.	41
4.3	Extrait des paramètres estimés du deuxième modèle, pour la première moitié de la saison 2015-2016.	42
4.4	Extrait des paramètres estimés du troisième modèle, pour la première moitié de la saison 2015-2016.	43
4.5	Extrait des paramètres estimés du quatrième modèle, pour la première moitié de la saison 2015-2016.	43
4.6	Nombre d'affrontements dans les blocs de simulation.	44
4.7	Pourcentage de bonnes prédictions, par deuxième moitié d'année et par modèle.	45
4.8	Pourcentage de prolongation et de jeux blancs, par deuxième moitié d'année et par modèle.	45
4.9	Moyenne du pourcentage de bonnes prédictions de l'issue des matchs des équipes pour la deuxième moitié de la saison 2015-2016.	46
4.10	Moyenne du pourcentage de bonnes prédictions de l'issue des matchs des équipes pour la deuxième moitié de la saison 2015-2016.	47
4.11	Moyenne du pourcentage de bonnes prédictions de l'issue des matchs des gardiens pour la deuxième moitié de la saison 2015-2016.	48
4.12	Évolution de l'expérience des joueurs dans l'alignement des Canadiens de Montréal au cours de la saison 2015-2016.	49
4.13	Extrait des choix des paris sportifs estimés par le quatrième modèle.	51
4.14	Valeur ajoutée potentielle des modèles de simulations pour un agent naïf.	52
4.15	Valeur ajoutée potentielle des modèles de simulations pour un agent rationnel.	53
4.16	Pourcentage de bonnes prédictions en excluant les affrontements à risque d'être mal prédits, par année et par modèle.	54
4.17	Valeur ajoutée potentielle des modèles de simulations pour un agent rationnel et informé.	54
B.1	Paramètres estimés pour le premier modèle avec la division Métropolitaine, Centrale et Pacifique. pour la saison 2015-2016.	59
B.2	Paramètres estimés pour le premier modèle pour la saison 2014-2015.	60
B.3	Paramètres estimés pour le premier modèle pour la saison 2013-2014.	61
B.4	Paramètres estimés pour le deuxième modèle à domicile pour la saison 2015-2016.	62
B.5	Paramètres estimés pour le deuxième modèle à l'étranger pour la saison 2015-2016.	63
B.6	Paramètres estimés pour le deuxième modèle à domicile pour la saison 2014-2015.	64
B.7	Paramètres estimés pour le deuxième modèle à l'étranger pour la saison 2014-2015.	65
B.8	Paramètres estimés pour le deuxième modèle à domicile pour la saison 2013-2014.	66
B.9	Paramètres estimés pour le deuxième modèle à l'étranger pour la saison 2013-2014.	67
B.10	Suite des paramètres estimés pour le troisième modèle avec l'asso. de l'Est.	68
B.11	Paramètres estimés pour le troisième modèle avec l'asso. de l'Ouest.	68
B.12	Suite des paramètres estimés pour le quatrième modèle avec l'asso. de l'Est.	69
B.13	Paramètres estimés pour le quatrième modèle avec l'asso. de l'Ouest.	69

Liste des figures

2.1	Moments des buts marqués lors des affrontements intra-associations entre les saisons 2013-2014 et 2015-2016.	11
2.2	Nombre de buts marqués par match par équipe lors des affrontements intra-associations entre les saisons 2013-2014 et 2015-2016.	12
2.3	Nombre de buts marqués à domicile par rapport à l'extérieur lors des affrontements intra-associations entre les saisons 2013-2014 et 2015-2016.	13
3.1	Impact des paramètres sur la fonction de répartition de la loi Weibull.	17
3.2	Impact des paramètres sur la densité de probabilité de la loi Weibull.	18
3.3	Comparaison des buts marqués et des buts accordés en temps réglementaire lors de la saison 2015-2016 pour les Canadiens de Montréal.	19
3.4	Densité de probabilité des Weibull pour les buts marqués des Maple Leafs et des Canadiens, pour la première moitié de la saison 2015-2016.	21
3.5	Impact du paramètre sur la fonction de masse de la loi de Poisson.	24
3.6	Impact du paramètre sur la fonction de répartition et sur la fonction de densité de la loi exponentielle.	25
3.7	Exemple d'un processus de Poisson à travers le temps.	27
4.1	Lois Weibull estimées pour les buts accordés à domicile des équipes de la division Atlantique, pour la saison 2015-2016.	41
4.2	Distribution des pourcentages de bonnes prédictions par division-saison.	44

Introduction

Devenu une discipline au pouvoir économique incroyable, le sport, caractérisé par son univers divertissant, commence à être grandement influencé par l'omniprésence des statistiques. Citons par exemple « Moneyball¹ », qui fut une méthode de gestion des Athletics d'Oakland lors de la saison de 2002 ayant pour but de construire une équipe compétitive en Ligue majeure de baseball malgré les difficultés financières de la franchise grâce à l'utilisation des statistiques. Désormais, suite aux succès qu'elle procure, la statistique est devenue l'outil par excellence que ce soit pour mesurer les performances sportives afin d'alimenter l'encadrement technique et les médias, pour le calcul des cotes et des gains pour les paris sportifs ou pour mesurer des affluences et audiences pour le marketing sportif. Selon Henwood (2015), H2 Gambling Capital² estime qu'à lui seul, le marché mondial des paris sportifs ne représente qu'environ 12,6 % du marché mondial du jeu, qui était évalué à 374,4 G€ (milliards d'euros) en 2015. Par contre, le rapport de Eaton (2014), publié par l'ICSS³, affirme que près de 80 % des paris sportifs sont placés sur le marché noir. Par conséquent, le véritable marché mondial des paris sportifs valait approximativement 250 G€ en 2015, ce qui représentait, au total, 40 % du marché mondial du jeu.

Habituellement, le développement de nouveaux modèles de prédictions, de mesures statistiques avancées ou autres est motivé par les gains monétaires qu'ils ont le potentiel d'occasionner pour une entité quelconque. Normalement, les gains lors des paris sportifs surviennent en déjouant les pronostics construits par les preneurs de paris⁴. Sachant qu'un sport comme le hockey est presque seulement populaire dans les pays nordiques, contrairement au soccer et au basketball par exemple, l'énergie mise dans la construction des cotes est davantage limitée. Ainsi, un modèle statistique performant pourra aisément en exploiter leurs failles afin d'encourir des gains substantiels.

Avant de commencer l'élaboration de modèles, il sera intéressant de survoler ce qui a été publié au sujet des modélisations appliquées aux sports d'équipe. En fait, le chapitre 1 présente les

1. Voir : https://fr.wikipedia.org/wiki/Moneyball:_The_Art_of_Winning_an_Unfair_Game.

2. H2 Gambling Capital est une compagnie reconnue comme la principale autorité en ce qui concerne les connaissances du marché mondial du jeu, basé au Royaume-Uni.

3. International Centre for Sport Security.

4. Ailleurs qu'au Québec, il s'agit du terme « bookmaker », qui désigne une personne morale ou physique permettant de parier de l'argent sur des événements.

modèles dits « traditionnels », puis s'attarde à différentes modélisations appliquées au hockey sur glace. Tous ces modèles possèdent un objectif commun : ils cherchent à maximiser la précision de leurs prévisions des affrontements passés, afin que son utilisateur possède un avantage dans les prédictions.

Par la suite, le chapitre 2 explique comment sont construites les diverses bases de données intra-associations et la base de données intradivision sur les cotes associées aux paris sportifs. Le tout tient pour les rencontres de la saison régulière pour les saisons de 2013-2014 jusqu'à 2015-2016. D'ailleurs, plusieurs statistiques descriptives sur les données permettront de vérifier la robustesse éventuelle des modèles de simulations construits au chapitre suivant.

De son côté, le chapitre 3 décrit les différents modèles, de manière théorique, qui tenteront de surpasser les précédents. D'abord, un premier modèle utilisant la loi de Weibull à trois paramètres sera présenté, puis trois autres modèles combinant les processus de Poisson et les modèles linéaires généralisés seront établis. Pour permettre une comparaison immédiate des modèles, ils seront tous suivis par le même exemple simulé à répétition.

Pour sa part, après avoir expliqué en détail comment sont calibrés les modèles avec les données de la première moitié d'une saison, le chapitre 4 utilise la méthode de Monte-Carlo pour simuler les deuxièmes moitiés de saison. L'objectif ici est de montrer comment les modèles complexes sont capables de surpasser les autres modèles par rapport à leur apport monétaire, en se basant sur la théorie de l'utilité espérée pour décider sur quelle équipe miser. Finalement, combiner ces modèles à une censure de certains affrontements moins bien modélisés permettra de multiplier les bénéfices.

Chapitre 1

Survol de la littérature

Selon Schumaker et collab. (2010), l'exploration de données sportives a connu une croissance rapide ces dernières années. En commençant par la création de ligues virtuelles, les amateurs de sport à la recherche d'un avantage dans les prédictions se sont vu développer des techniques et outils complexes afin d'améliorer leurs mesures associées aux performances, soit des joueurs, soit des équipes. Avant l'avènement de l'exploration de données, les organisations sportives comptaient presque exclusivement sur l'expertise humaine. On croyait que les experts du domaine (entraîneurs, gestionnaires et dépisteurs) pourraient effectivement convertir leurs données recueillies en connaissances utilisables. Cependant, l'accroissement de l'information contenu dans les statistiques sportives a conduit à l'ajout de statisticiens à l'interne pour créer de meilleures mesures de performance et critères de prise de décision.

D'ailleurs, Schumaker et collab. (2010) expliquent comment, en utilisant des techniques efficaces d'exploration de donnée, les performances d'une équipe peuvent s'améliorer. Il s'agit d'un excellent point de départ pour quiconque veut s'initier au domaine de l'analyse statistique sportive. Leur exemple le plus simple reste au baseball, où il est très important d'associer les joueurs à certaines situations, d'évaluer la contribution personnelle de chacun, d'identifier les tendances de l'adversaire, et d'exploiter toutes leurs faiblesses.

Les sections qui suivent décrivent sommairement différentes modélisations appliquées aux sports d'équipes, incluant le hockey sur glace. Il ne s'agit pas là d'une revue de la littérature exhaustive, mais bien d'un condensé ciblant des articles qui permettent de comprendre les modélisations conventionnelles.

1.1 Différentes modélisations appliquées aux sports d'équipes

1.1.1 Les modèles traditionnels

La vaste majorité des modélisations statistiques associées aux sports fut appliquée d'abord au soccer en raison de sa simplicité. Évidemment, chacune d'entre elles possède ses propres

particularités et ajouts. Toutefois, leur construction est généralement identique. Premièrement, elles nécessitent la création d'attributs offensifs, défensifs ou autres servant d'indicateur de performance pour chacune des équipes à travers une même ligue. La confrontation de ces attributs formera le (les) paramètre(s) des lois qui seront comparées pour prédire l'issue d'un affrontement. Parmi les modélisations de ce type, on peut citer [Baio et Blangiardo \(2010\)](#) et [Crowder et collab. \(2002\)](#) qui s'intéressent aux prédictions du soccer. Dans ces articles, on fait l'hypothèse que le nombre de buts marqués par une équipe suit une distribution de Poisson dont la moyenne est une combinaison de son attribut offensif ajusté par celui défensif de son adversaire. Finalement, tous les attributs furent estimés par maximum de vraisemblance à l'aide de l'entièreté des résultats d'une année, afin de déduire la puissance de prédiction des affrontements déjà terminés qu'ont leurs modèles. Puis, [Ruiz et Perez-Cruz \(2015\)](#) ont montré que l'adaptation de ce type de modèle de prédiction classique pour le soccer a pu prédire les résultats au basketball (NCAA). En y ajoutant la factorisation de Poisson et en prenant en compte la structure de la NCAA, ils ont pu conclure que leur modèle tend à fournir des résultats qui diffèrent des probabilités implicites des maisons de paris et, par conséquent, a le potentiel d'offrir des gains.

1.1.2 Les objectifs et limites des modélisations

Inévitablement, à travers la documentation on remarque que l'objectif principal de la modélisation statistique associé aux sports est toujours le même : maximiser les gains à travers les paris sportifs. Ainsi, les modélisations tentent constamment d'établir les meilleurs pronostics possible. Aux articles précédents, on peut ajouter [Goddard et Asimakopoulos \(2001\)](#) qui ont construit un modèle de régression probit ordonné pour prédire les résultats des matchs de soccer de la ligue anglaise. Sinon, celui de [Polson et Stern \(2015\)](#) qui proposent une méthode pour calculer la volatilité implicite du marché des paris sportifs pendant un match. Ils considèrent le modèle stochastique de [Stern \(1994\)](#) pour l'évolution du pointage pendant un match pour déduire l'évolution qu'ont généralement les cotes associées à la victoire des équipes pendant ce même match. Or, les paramètres de leur modèle s'ajusteront en temps réel avec l'évolution du pointage et permettront de déceler plusieurs failles exploitables à travers les différentes cotes déterminées par les marchés.

Comme [Shi et collab. \(2013\)](#) le montrent, il faut garder à l'esprit qu'il y a une limite supérieure quant au pourcentage d'affrontements pouvant être bien prédit dans le sport professionnel rôdant aux alentours de 74 %, le facteur chance étant très présent et difficile à contrôler. Ayant établi que peu importe le modèle de simulation cette valeur reste identique, ils suggèrent de développer de meilleurs indicateurs de performances pour chacune des équipes du sport qui nous intéresse avant de se lancer dans la confection de modèle complexe. Pour vérifier la qualité des indicateurs choisis, [Zimmermann \(2015\)](#) propose d'utiliser le regroupement des profils statistiques des équipes d'une même ligue pour définir leurs attributs normalement, puis

de simuler toutes les rencontres comme s’il s’agissait d’une saison complète à partir d’un vrai calendrier. Or, il affirme qu’un grand nombre de simulations par la méthode de Monte-Carlo permettra d’ajuster correctement les estimations des attributs pour rehausser le réalisme des simulations subséquentes et pour donner des limites plus strictes sur l’exactitude prédictive. Bref, dans le monde du sport, ces méthodes permettent, entre autres, de prédire quelle sera l’issue d’une saison complète.

1.2 Différentes modélisations appliquées au hockey

Les applications des modèles statistiques au hockey sur glace sont plus rares que celles faites au soccer, baseball, basketball et football. Par contre, cette discipline a finalement gagné en popularité depuis quelques années grâce à l’amélioration de l’accessibilité de ses données. Donc, les modélisations pour le hockey professionnel rivalisent maintenant avec celles des autres sports majeurs.

1.2.1 Les modélisations par équipe

Tout d’abord, Dayaratna et Miller (2013) montrent que la « Pythagorean expectation », formule à priori créée pour le baseball évoqué dans James (1980), est aussi utilisable au hockey. Elle sert à prédire le nombre de victoires qu’une équipe aurait dû obtenir en se basant sur ses buts marqués (BM) et ses buts accordés (BA) au terme d’une saison. Selon Miller (2007), comparer le pourcentage de victoire (PV) de l’équation (1.1) au véritable pourcentage de victoire obtenu par l’équipe indique son niveau de chance lors de la saison. Cochran et Blackstock (2009) ont appliqué cette formule au hockey. À l’aide d’estimation par les moindres carrés du modèle de James et plusieurs de ses variantes, ils ont déduit que l’utilisation d’un $\gamma = 1,927$ dans l’équation offrirait d’aussi précises estimations que l’utilisation de modèles plus complexes. Plus tard, Dayaratna et Miller (2013) ont estimé que γ devrait être de 2,15.

$$PV = \frac{BM^\gamma}{BM^\gamma + BA^\gamma} \quad (1.1)$$

D’après eux, pour que la « Pythagorean expectation » tienne, il est nécessaire de considérer que les buts marqués (BM) et les buts accordés (BA) suivent des distributions indépendantes de Weibull à trois paramètres. Les lois construites possèdent donc des densités de probabilité comme à l’équation (1.2) et leur fonction de répartition est représentée par l’équation (1.3), où $k > 0$ est le paramètre de forme, $\lambda > 0$ est le paramètre d’échelle et θ est le paramètre de position de la distribution.

$$f(x; k, \lambda, \theta) = \frac{k}{\lambda} \left(\frac{x - \theta}{\lambda} \right)^{k-1} e^{-\left(\frac{x-\theta}{\lambda}\right)^k} \mathbb{1}_{\{x \geq \theta\}} \quad (1.2)$$

$$F(x; k, \lambda, \theta) = 1 - e^{-\left(\frac{x-\theta}{\lambda}\right)^k} \mathbb{1}_{\{x \geq \theta\}} \quad (1.3)$$

De plus, il est important de considérer que $\theta = -0,5$ pour assurer une optimisation juste des deux autres paramètres. Puis, toutes valeurs simulées par cette loi devront être bonifiées de 0,5 pour que le minimum atteignable soit un pointage réaliste, soit 0. L'utilisation d'une distribution continue pour modéliser une variable aléatoire discrète facilite les calculs en transformant les sommes (utilisées avec les distributions discrètes) en intégrales et facilite l'obtention d'une solution simple de forme fermée pour une équation comme la formule « Pythagorean ». Évidemment, les distributions continues ne représentent pas la réalité puisque les équipes de hockey marquent toujours un nombre entier de buts. Toutefois, la distribution de Weibull est très flexible et, en fixant correctement ses paramètres, s'ajuste à de nombreux jeux de données.

De nombreux dérivés de la « Pythagorean expectation » existent dans la littérature. Par exemple, Braunstein (2010) l'utilise comme variable explicative dans un modèle linéaire de régression. Le pourcentage de victoire qu'une équipe i aurait dû obtenir pour la saison t se calcule en corrigeant la « Pythagorean expectation » par les écarts-types des buts marqués et des buts accordés comme à l'équation (1.4).

$$\widehat{PV}_{it} = PV_{it} - 0,03257\sigma_{it,BM} + 0,0323\sigma_{it,BA} \quad (1.4)$$

Alternativement, Wolfson et Koopmeiners (2015) cherchent à prédire à partir d'un moment précis de la saison, les résultats du reste de la même saison à l'aide d'un modèle de Bradley-Terry comme à l'équation (1.5), où π_{ij} est la probabilité qu'une équipe à domicile i vaincre une équipe à l'étranger j , β_i et β_j sont des indicateurs de forces des équipes i et j respectivement, et α est un paramètre d'ajustement pour l'équipe à domicile. Évidemment, plus la saison est avancée, plus les prédictions s'améliorent. Malheureusement, leur méthode n'atteint jamais un pourcentage de bonnes prédictions supérieur à 60 %¹, même lorsque le modèle est ajusté après que la saison soit complétée à 87,5 %. Généralement, le pourcentage de bonnes prédictions de leur modèle ne dépassent jamais de plus de 3 % celui obtenu en considérant simplement que l'équipe à domicile gagne toujours.

$$\text{logit}(\pi_{ij}) = \beta_i - \beta_j + \alpha \quad (1.5)$$

1.2.2 Les modélisations des rencontres

Quelques autres sources ont préféré fractionner les affrontements, au lieu de considérer un match dans son ensemble. Par exemple, Merritt et Clauset (2014) considèrent le processus de Poisson pour modéliser le rythme des buts dans une partie de hockey. Puis, en utilisant des processus de Bernoulli comme mécanisme d'ajustement stratégique des équipes dû aux écarts de pointage, ils parviennent à prédire à environ 80 % du temps l'équipe gagnante après le premier but marqué. Ce taux de succès s'explique par le fait que beaucoup de rencontres de leur

1. Les saisons utilisées sont celles de 2005-2006 jusqu'à celle de 2012-2013.

base de données se sont terminées avec très peu de buts ou par jeu blanc². De manière similaire, Thomas et collab. (2013) suggèrent que le rythme des buts, pour chacune des équipes, suit plutôt un processus de renouvellement de Markov avec des taux de défaillance qui dépendent des joueurs sur la glace. Par contre, cette méthode nécessite l'évaluation des capacités offensives et défensives de chaque joueur en tenant compte des qualités de leurs coéquipiers et de leurs adversaires, de la situation du match, etc.

De son côté, Weissbock (2014) utilise des approches de « machine Learning » pour tenter de prédire les victoires et les défaites dans la LNH pour la saison écourtée de 2012-2013 lors d'un affrontement simple ou d'un quatre de sept pour les séries éliminatoires ; les nulles ne sont pas considérées. Leur méthode conclut que l'utilisation des statistiques traditionnelles permettent de mieux prédire de manière automatique un affrontement que les statistiques avancées, en atteignant 59,8 % de précision par validation croisée. De plus, après plusieurs simulations, ils ont établi qu'il existe une borne supérieure de performance approximative de 62,0 %³ pour la LNH. Malgré tout, pour les séries éliminatoires, ils ont obtenu une précision de prédiction de près de 75,0 % en utilisant les statistiques avancées. Finalement, un simple modèle utilisant la recherche de mots clés (positifs, négatifs, etc.) dans les rapports d'avant-matchs publiés par des experts chaque jour sur le site officiel de la Ligue nationale de hockey s'est avéré prédire un plus grand nombre d'affrontements que ceux utilisant seulement des données numériques.

1.2.3 Estimations de la contribution des joueurs

Initialement, une statistique comme le plus-moins, qui mesure la différence entre le nombre de buts marqués, excluant les buts en avantage numérique, et le nombre de buts accordés lorsqu'un joueur en particulier est sur la glace, était couramment utilisée pour identifier la contribution d'un joueur dans un match de hockey. Cette statistique est affectée directement par les performances globales de leur équipe. D'autres statistiques comme le Corsi, comptabilisant le différentiel des lancers tentés par l'équipe à égalité numérique contre ceux accordés lorsqu'un joueur en particulier est sur la glace, et le Fenwick, calculant le différentiel des lancers tentés non bloqué par l'équipe à égalité numérique contre ceux accordés lorsqu'un joueur en particulier est sur la glace, sont devenues les statistiques de référence pour évaluer cette contribution. Précisons que le Corsi et le Fenwick ne sont jamais comptabilisés dans les situations de filet désert. D'autres statistiques sont plus récentes, comme celle présentée par Schuckers et Curro (2013), soit le THoR (Total Hockey Ratings). Cette statistique modélise si la présence d'un joueur sur la glace est plus souvent positive que négative par rapport au pointage, sachant qui sont ses coéquipiers sur la glace, qui sont ses adversaires, quel est le pointage actuel, où se trouve la rondelle lorsqu'il embarque sur la glace, etc.

De leur côté, Gramacy et collab. (2013) ont proposé un modèle de régression logistique afin

2. Lorsqu'une équipe ne marque aucun but lors de la rencontre.

3. 24 % d'un affrontement est déterminé par la meilleure équipe, et le 76 % restant est de la chance.

d'améliorer davantage l'estimation de la contribution des joueurs à un match de hockey. Leur modèle sert à déterminer la chance qu'un joueur aura de contribuer à un but. Cependant, leurs résultats montrent que la plupart des joueurs ont une contribution relativement inexistante. En revanche, la forte influence des coéquipiers est un problème récurrent à travers les statistiques individuelles au hockey. C'est pourquoi [Macdonald \(2012\)](#) propose l'utilisation d'une régression d'arête qui considère non seulement les buts marqués des joueurs, mais aussi les lancers, le Corsi et le Fenwick. Les résultats de la régression servent d'estimation pour la contribution offensive et défensive de chacun des joueurs à forces égales, en avantage numérique et en désavantage numérique. De plus, cette méthode nous assure que toutes les estimations de contribution sont indépendantes de la force de leurs coéquipiers et de leurs adversaires.

Finalement, les modélisations de cet essai décrites au chapitre 3 utiliseront diverses notions présentées dans cette revue de la littérature. D'abord, un modèle similaire à celui de [Dayaratna et Miller \(2013\)](#) sera construit. Suivrons trois modèles différents construits avec les processus de Poisson comme l'ont fait [Merritt et Clauset \(2014\)](#). Éventuellement, le meilleur modèle sera celui qui possède les plus grands gains potentiels via les paris sportifs qu'ils peuvent générer. Ainsi, contrairement à la littérature, cet essai permettra d'estimer directement la valeur ajoutée propre à chacune des modélisations.

Chapitre 2

Exploration des bases de données

2.1 Sources de données pour le hockey

Sans aucun doute, les données les plus fiables sur le hockey professionnel sont répertoriées sur le site officiel de la Ligue nationale de hockey (LNH). Elles contiennent toutes les statistiques de base, des unités spéciales, du RTSS (Real Time Scoring System) et les statistiques avancées pour tous les joueurs, gardiens et équipes. Elles sont entièrement disponibles sous différentes agrégations (par rencontre, par saison) à partir de la saison 2002-2003. Toutefois, les statistiques avancées ne sont comptabilisées que depuis la saison 2010-2011. Malgré leur consultation rapide et simplifiée, les bases de données du site de la LNH ne sont pas construites pour être facilement exportables. D'autres sources, comme les sites officiels de Puckalytics¹ et de War-on-Ice², recensent des statistiques similaires et pallient à la situation. Cependant, il s'agit de sources d'erreurs puisque ces sites ne sont pas sujets à des révisions constantes, contrairement au site de la LNH.

Bien qu'inexploitée dans le cadre de cet essai, il est intéressant de savoir que Schuckers et Macdonald (2014) proposent des corrections nécessaires à toutes les statistiques calculées par le RTSS de la LNH en fonction de l'arène où ils furent comptabilisés par les différents marqueurs. Les résultats de ce papier assurent une construction de toutes les statistiques avancées plus équitable entre les équipes. D'un autre côté, des entreprises comme Hockey Tech, Sports Inc. et quelques autres utilisent des caméras de repérage optique afin d'enregistrer les coordonnées de tous les patineurs et de la rondelle, une centaine de fois par seconde, pour créer leur propre base de données. La quantité énorme d'information retirée permet même d'explorer des sphères plus complexes, comme les aptitudes de patinage des joueurs ou la vitesse de leurs passes. Quelques présentations dont celle de Moretti (2016), de Appleby (2016) et de Pohlkamp-Hartt (2016) en montrent des exemples.³

1. Voir : <http://www.puckalytics.com/>

2. Voir : <http://war-on-ice.com/>

3. Présentées lors de l'événement intitulé Ottawa Hockey Analytics at Carleton survenu en janvier 2016.

2.1.1 Détails des bases de données sur le hockey

Avant toutes choses, il ne faut jamais oublier que le lock-out de la LNH au cours de la saison 2004-2005⁴ a suffisamment modifié le jeu pour rendre les statistiques précédant la saison 2005-2006 complètement incomparables à celle suivant ce lock-out. De plus, un autre lock-out de la LNH, celui-ci écourtant de moitié la saison 2012-2013⁵ (passée de 82 parties à 48), sans toutefois causer de changements majeurs, fut le moment idéal pour la LNH de remanier leurs divisions. Or, c'est depuis la saison 2013-2014 que la LNH a adopté un format à 4 divisions⁶ délaissant son ancien format à 6 divisions. Par conséquent, les Red Wings de Detroit et les Blue Jackets de Columbus furent transférés de l'association Ouest à l'association Est, pendant que les Jets de Winnipeg ont été transférés dans l'autre direction.

Comme il s'agit de saisons assez comparables, dans le sens que d'une année à l'autre la ligue possède les mêmes règles et que les divisions sont restées identiques, les analyses qui suivront se limiteront aux données à partir de la saison 2013-2014 jusqu'à la saison 2015-2016. Les diverses sources d'informations statistiques énumérées précédemment ont permis de construire une base de données relativement complète. L'historique des résultats pour tous les matchs des saisons 2013-2014, 2014-2015 et 2015-2016 provient des résumés de parties du site officiel de la LNH, ce qui inclut le moment (temps), le contexte (avantage numérique, filet désert, lancer de pénalité, etc.) de tous les buts marqués et le nom des gardiens présents à ce moment. D'autres informations, comme les noms des joueurs ayant marqué les buts et de ceux ayant récolté les passes, le nombre de lancers de chacune des équipes et les moments des pénalités, n'ont pas été retenus dans la base de données, malgré leur caractère exploitable. Afin de simplifier la notation de tous les noms des équipes de la LNH, chaque équipe possède une abréviation officielle de trois lettres répertoriée dans l'annexe A.

Bref, notre base de données compte 2342 affrontements intra-associations ($\approx 60\%$ de la base de données complète). Exclure les affrontements interassociations est justifiable, puisqu'ils correspondent à des affrontements plus propices à des résultats imprévisibles, soit par manque d'adaptation stratégique des équipes l'une face à l'autre ou soit parce qu'ils occasionnent des voyages bien plus longs qu'à l'habitude. En tout, les 2342 affrontements maintenus dans la base de données totalisent 12163 buts en temps réglementaire (moyenne $\approx 5,2$ buts par match). Un extrait d'un affrontement issu de la base de données et des variables principalement utilisées est présenté dans le tableau 2.1. Fondamentalement, cet extrait illustre l'évolution du pointage lors de la rencontre du 3 octobre 2013 entre les Prédateurs de Nashville et les Blues de St-Louis, et par le fait même, le changement de gardien des Prédateurs après avoir accordé 3 buts en 9 minutes et 45 secondes.

4. Voir : https://en.wikipedia.org/wiki/2004%E2%80%9305_NHL_lockout

5. Voir : https://en.wikipedia.org/wiki/2012%E2%80%9313_NHL_lockout

6. Les divisions Atlantique et Métropolitaine forment l'association de l'Est, et les divisions Centrale et Pacifique forment l'association de l'Ouest.

TABLE 2.1 – Extrait de la base de données des affrontements intra-associations.

Etr.	Dom.	Pointage étranger	Pointage domicile	Temps initial	Temps final	But	Gardien étranger	Gardien domicile
NSH	STL	0	0	0,00	2,08	STL	P.Rinne	J.Halak
NSH	STL	0	1	2,08	5,72	STL	P.Rinne	J.Halak
NSH	STL	0	2	5,72	9,75	STL	P.Rinne	J.Halak
NSH	STL	0	3	9,75	10,70	NSH	C.Hutton	J.Halak
NSH	STL	1	3	10,70	25,15	NSH	C.Hutton	J.Halak
NSH	STL	2	3	25,15	31,27	STL	C.Hutton	J.Halak
NSH	STL	2	4	31,27	60,00	—	C.Hutton	J.Halak

2.1.2 Moments des buts marqués

Une base de données alternative a aussi été construite, celle-ci séparant les 55 premières minutes d’un affrontement des 5 dernières. Ainsi, elle distingue deux phases de jeu substantiellement différentes. Au hockey, les 5 dernières minutes d’un affrontement sont caractérisées par un ajustement stratégique intensifié, puisque lorsqu’une équipe est en retard dans le pointage (généralement de moins de 3 buts), elle a tendance à retirer son gardien au profit d’un patineur supplémentaire. Par le fait même, elle augmente ses chances de marquer le but égalisateur, mais aussi de se faire marquer un but creusant l’écart dans le pointage. La figure 2.1, qui recense tous les buts marqués présents dans la base de données, montre bien que les 5 dernières minutes sont bel et bien différentes des autres par rapport au nombre de buts qui se marquent. Un test de Kolmogorov-Smirnov confirme que la distribution associée au temps des buts marqués est différente d’une loi uniforme avec une valeur p inférieure à $2,2 \times 10^{-16}$. Éventuellement, les deux bases de données seront utilisées lors des simulations. En séparant les rencontres à la 55^e minute, on peut reconstruire le tableau 2.1 comme le tableau 2.2.

FIGURE 2.1 – Moments des buts marqués lors des affrontements intra-associations entre les saisons 2013-2014 et 2015-2016.

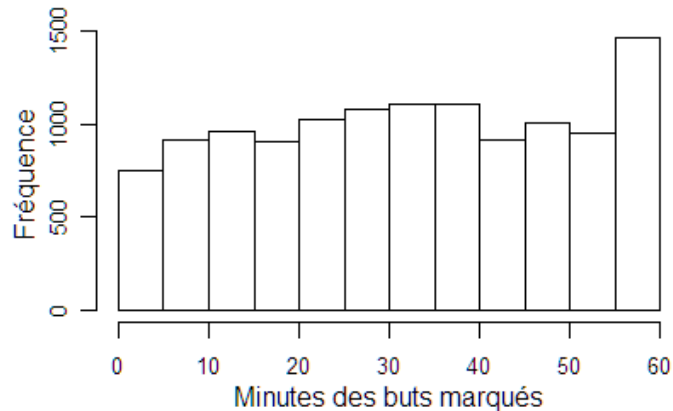


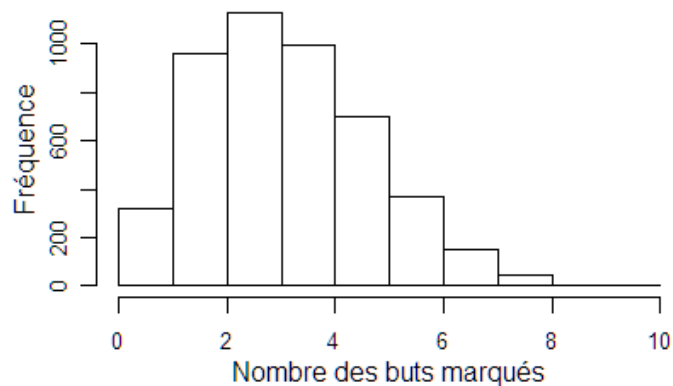
TABLE 2.2 – Extrait de la base de données alternative des affrontements intra-associations.

Etr.	Dom.	Pointage étranger	Pointage domicile	Temps initial	Temps final	But	Gardien étranger	Gardien domicile
NSH	STL	0	0	0,00	2,08	STL	P.Rinne	J.Halak
NSH	STL	0	1	2,08	5,72	STL	P.Rinne	J.Halak
NSH	STL	0	2	5,72	9,75	STL	P.Rinne	J.Halak
NSH	STL	0	3	9,75	10,70	NSH	C.Hutton	J.Halak
NSH	STL	1	3	10,70	25,15	NSH	C.Hutton	J.Halak
NSH	STL	2	3	25,15	31,27	STL	C.Hutton	J.Halak
NSH	STL	2	4	31,27	55,00	—	C.Hutton	J.Halak
NSH	STL	2	4	55,00	60,00	—	C.Hutton	J.Halak

2.1.3 Nombre de jeux blancs et de surtemps

En fait, limiter le nombre de buts de son adversaire est l'objectif défensif de chaque équipe. Ultimement, le nombre de buts accordés par une équipe lors d'un affrontement sera nul, ce qui correspond à un jeu blanc. La figure 2.2 montre que la proportion de jeux blancs est approximativement de 7 % (soit 321 jeux blancs en 4684 pointages enregistrés) à travers la totalité des affrontements de la base de données. La rareté relative de cet événement est normale puisqu'il ne suffit que d'une erreur défensive pour éliminer la possibilité d'obtenir un jeu blanc.

FIGURE 2.2 – Nombre de buts marqués par match par équipe lors des affrontements intra-associations entre les saisons 2013-2014 et 2015-2016.



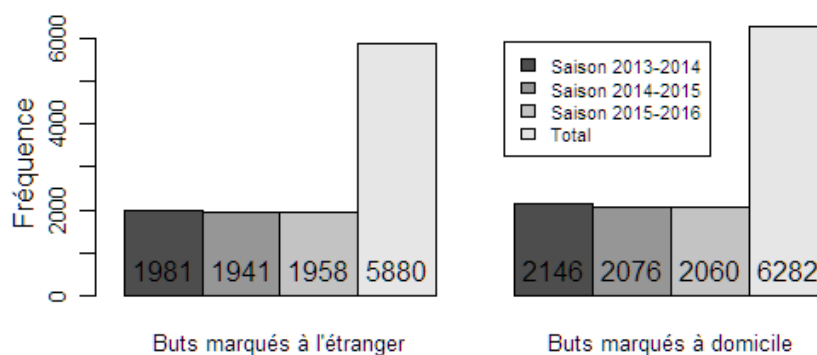
Au hockey, une prolongation survient lorsque le pointage des deux équipes est égal à la fin du temps réglementaire, qui correspond aux soixante premières minutes d'un affrontement. Les prolongations ajoutent une période de 5 minutes à la rencontre et prennent fin aussitôt qu'une équipe marque un but. Advenant le cas qu'aucun but n'est marqué, l'affrontement se retrouve

en tir de barrage⁷. Il est important de noter que les prolongations se jouaient à 4 contre 4 jusqu'à la saison 2014-2015. Toutefois, lors de la saison 2015-2016, elles furent jouées à 3 contre 3, favorisant les buts. En se fiant aux dernières bases de données, la proportion d'affrontement terminant en surtemps (prolongation ou tir de barrage) est approximativement 25 % (soit 583 surtemps en 2342 affrontements). Bref, quelques éléments de la base de données peuvent permettre de vérifier le réalisme des simulations des chapitres ultérieurs. Parmi eux, on peut compter la proportion de jeux blancs et de surtemps.

2.1.4 Avantage de la glace

Habituellement, dans le sport professionnel, on tient pour acquis qu'il existe un avantage pour l'équipe qui joue à son domicile. Les encouragements de la foule, la connaissance des subtilités de leur terrain ou patinoire (qualité de la glace, des bandes et des baies vitrées) et le voyage en moins peuvent souvent être considérés comme bénéfiques. Les bases de données précédemment construites montrent que cet effet, nommé l'avantage de la glace, existe. La figure 2.3 agrège tous les buts marqués présents dans la base de données par rapport à si l'équipe qui marque jouait à domicile ou à l'étranger. Un test t de Welch⁸ vient d'ailleurs rejeter l'hypothèse nulle d'égalité entre les deux moyennes comparées en faveur d'une hypothèse alternative unilatérale qui considère que le nombre de buts marqués moyens à domicile (2094) est significativement plus élevé que le nombre de buts moyens marqués à l'étranger (1960). Le test possède une valeur p de 0,01142. En moyenne, le nombre de buts marqués à domicile surpasse de 6,84 % le nombre de buts marqués à l'étranger. Évidemment, l'effet pourrait être plus ou moins prononcé en fonction de l'équipe, mais seul un effet global sera considéré ici.

FIGURE 2.3 – Nombre de buts marqués à domicile par rapport à l'extérieur lors des affrontements intra-associations entre les saisons 2013-2014 et 2015-2016.



7. Voir : https://fr.wikipedia.org/wiki/Tir_de_fusillade

8. Teste l'hypothèse d'égalité de deux moyennes avec deux échantillons de variances inégales.

2.2 Sources de données pour les paris sportifs

Un pari sportif consiste à miser de l'argent sur un évènement lors d'une rencontre sportive. Il existe de nombreux types de paris, dont les plus simples consistent à parier sur la victoire d'une équipe précise. Chaque pari a une cote et c'est elle qui permet de connaître à l'avance, en fonction de la somme mise, le montant gagné si l'évènement se produit. En notation décimale, si l'on mise 10 \$ sur l'équipe A dont la cote est de 1,6, on gagne 16 \$ avec une victoire de l'équipe A, sinon la somme mise est perdue. Les paris se font en général chez des preneurs de paris qui sont des organismes autorisés à proposer aux joueurs de parier. Normalement, la cote pour un affrontement est calculée et proposée par les preneurs de paris en fonction de leurs estimations de chances de victoire ou de nulle entre deux équipes.

La cote d'un pari peut être exprimée selon trois notations différentes : européenne, anglaise ou américaine. La cote européenne correspond à un nombre réel positif indiquant le gain potentiel par rapport à la mise. De son côté, la cote anglaise est exprimée par une fraction désignant le gain potentiel pour chaque dollar misé. Par ailleurs, la cote américaine est représentée par un nombre entier positif ou négatif. Une cote précédée par un « - » désigne le montant qu'il faut parier pour gagner 100 \$, une cote précédée par un « + » désigne qu'il faut parier 100 \$ pour gagner la cote. Le tableau 2.3 affiche les différences entre ces notations en cas de victoire d'une équipe A ou d'une équipe B.

TABLE 2.3 – Exemple de gains selon les différentes définitions des cotes.

Définition	Équipe A	Équipe B
Cote européenne	1,5	4
Cote anglaise	1/2	3/1
Cote américaine	-200	+300
Mise	100 \$	100 \$
Gain	150 \$	400 \$
Valeur ajoutée	50 \$	300 \$

Outre les paris simples⁹, où il suffit de choisir une rencontre sportive et d'effectuer un pronostic, il existe plusieurs types de paris différents. Un pari « Scorecast » ajoute au pronostic précédent, le pointage final de la rencontre. Il est aussi possible de participer à des paris de mi-temps, qui consiste à parier sur une portion d'un affrontement, ou de participer à des paris sportifs en direct, qui permettent de parier sur un évènement sportif au fur et à mesure de son déroulement. Sinon, on peut parier sur le nombre total de buts à l'issue d'une rencontre ou sur l'écart dans le pointage à l'issue d'une rencontre.¹⁰ Dans les paris plus complexes, il est aussi possible de parier sur quelle équipe s'inscrira en premier au pointage, quelle équipe dominera une statistique précise, qui sera le joueur du match, etc. Les paris sur l'issue d'une saison

9. Les preneurs de paris nomment les paris simples : MoneyLines.

10. Les preneurs de paris nomment respectivement ces paris : Over-Under et Spread.

complète existent aussi, par exemple Loto-Québec, par l'intermédiaire de Mise-o-jeu, permet de miser sur des statistiques précises pour des joueurs précis des Canadiens de Montréal.

2.2.1 Sommaire de la base de données sur les paris

Les différents historiques de cotes des preneurs de paris de la LNH sont généralement inaccessibles gratuitement. Toutefois, le site officiel de Odds Shark¹¹ donne accès à une base de données suffisamment complète pour la LNH, comprenant les cotes des paris simples, exprimées sous le format américain. Sachant que les modèles utilisés dans cet essai tenteront de prédire l'issue que des rencontres intradivisions de la saison 2013-2014 jusqu'à la saison 2015-2016, seules les différentes cotes pour les paris simples associées à ces affrontements formeront la base de données des cotes. Un extrait de cette base de données entre différents affrontements de la division Atlantique peut être consulté dans le tableau 2.4. Normalement, comme les cotes sont construites de manière à être à l'avantage des preneurs de paris, tous les affrontements possèdent soit deux cotes précédées par un « - », soit une cote précédée par un « - » détient une valeur supérieure à celle d'une cote précédée par un « + ».

TABLE 2.4 – Extrait de la base de données des cotes des affrontements de la division Atlantique.

Date	Etr.	Dom.	MoneyLines Etranger	MoneyLines Domicile
42378	BOS	OTT	-110	+100
42384	BOS	BUF	-119	+108
42385	TOR	BOS	+155	-172
42386	FLA	TBL	+140	-155
42388	BOS	MTL	-105	-105
42391	DET	BUF	-130	+118
42392	TBL	FLA	-127	+115
42392	MTL	TOR	-107	-103

11. Voir : <http://www.odsshark.com/nhl/database/>

Chapitre 3

Élaboration des modèles de simulation

Une fois toutes les bases de données construites, le moment est venu d'étaler comment elles seront exploitées. L'objectif de cet essai est de construire un modèle statistique pouvant s'avérer payant pour un simple parieur. Évidemment, la confection d'un modèle décrivant l'issue d'une rencontre d'une partie de hockey professionnel nord-américaine, où le positionnement et la stratégie d'équipe priment sur le talent individuel, reste une tâche ardue. Dans un premier temps, tout comme dans la majorité de la littérature sur le sujet, l'objectif est de maximiser le taux de bonnes prédictions de ces issues.

Quatre modèles servant à simuler des affrontements seront construits dans les sections suivantes, pour être comparés en matière de pertinence et performance dans le chapitre suivant. Un premier modèle exige de connaître la loi statistique de Weibull avec trois paramètres et les trois autres demandent de maîtriser les processus de Poisson et les modèles linéaires généralisés. Chacune de ces méthodes nécessite l'estimation de quelques paramètres, que ce soit pour mesurer le talent offensif, le talent défensif, ou l'impact des gardiens de but pour chacune des équipes. Les données intra-associations de la première moitié de la saison régulière serviront à calibrer les différents modèles.¹

3.1 Construction du premier modèle de simulation

Pour commencer, on considère un modèle très simple, sans influence directe des gardiens de but et sans dynamisme stratégique. Certes, des éléments clés affectant le pointage final d'un affrontement seront ainsi omis. Par contre, l'utilisation d'une loi de probabilité pour estimer le nombre de buts marqués en temps réglementaire de chacun des opposants corrigés par le nombre de buts qu'accorde généralement leur adversaire reste une méthode de simulation pertinente.

1. Les premières moitiés de saison régulière sont du 1^{er} octobre 2013 au 6 janvier 2014 pour la saison 2013-2014, du 8 octobre 2014 au 8 janvier 2015 pour la saison 2014-2015 et du 7 octobre 2015 au 7 janvier 2016 pour la saison 2016.

Le premier modèle est basé sur la loi de Weibull à trois paramètres. Les sous-sections suivantes permettront premièrement de définir en détail cette loi statistique, puis d'exposer toutes les subtilités que procure la première modélisation d'affrontements à l'aide d'un exemple de son fonctionnement.

3.1.1 Loi de Weibull avec trois paramètres

La loi de Weibull à trois paramètres (loi de Weibull translatée) possède un paramètre λ , le paramètre d'échelle de la distribution, un paramètre α , le paramètre de forme et un paramètre θ , le paramètre de position. Les paramètres α et λ sont strictement positifs, tandis que θ peut prendre n'importe quelle valeur réelle. Si θ vaut 0, alors il s'agit de la distribution de Weibull à deux paramètres. En fait, la fonction de répartition et la densité de la loi de Weibull (α , λ , θ) (pour $t > \theta$) sont respectivement données par :

$$F(t) = \begin{cases} 1 - e^{-\left(\frac{t-\theta}{\lambda}\right)^\alpha}, & \text{si } t \geq \theta \\ 0, & \text{si } t < \theta \end{cases}$$

$$f(t) = \begin{cases} \frac{\alpha}{\lambda} \left(\frac{t-\theta}{\lambda}\right)^{\alpha-1} e^{-\left(\frac{t-\theta}{\lambda}\right)^\alpha}, & \text{si } t \geq \theta \\ 0, & \text{si } t < \theta \end{cases}$$

Comme le montrent les graphiques de la fonction de répartition et de la densité de probabilité de la loi Weibull à la figure 3.1 et à la figure 3.2 respectivement, augmenter le paramètre λ accentue la forme de cloche étirée vers la droite de la distribution, augmenter le paramètre α étire la distribution vers la droite et augmenter le paramètre θ translate la distribution vers la droite. Bref, sa forme générale est semblable à celle de la loi gamma. D'ailleurs, la loi gamma possède une propriété d'additivité que la loi Weibull ne possède pas. Toutefois, la loi de Weibull se prête mieux aux calculs de probabilités grâce à la forme simple de sa fonction de répartition. Finalement, on remarque que si $\alpha = 1$ et $\theta = 0$, on obtient la loi exponentielle et que si $\alpha = 2$ et $\theta = 0$, on obtient la loi de Rayleigh.

FIGURE 3.1 – Impact des paramètres sur la fonction de répartition de la loi Weibull.

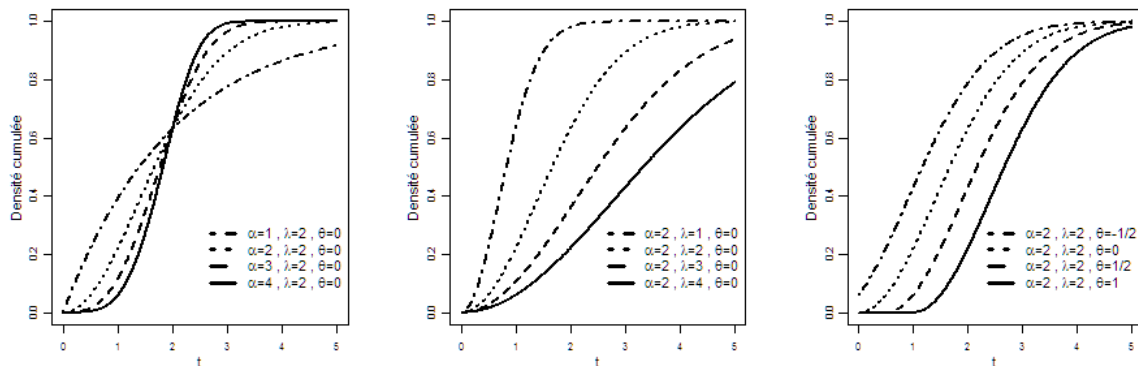
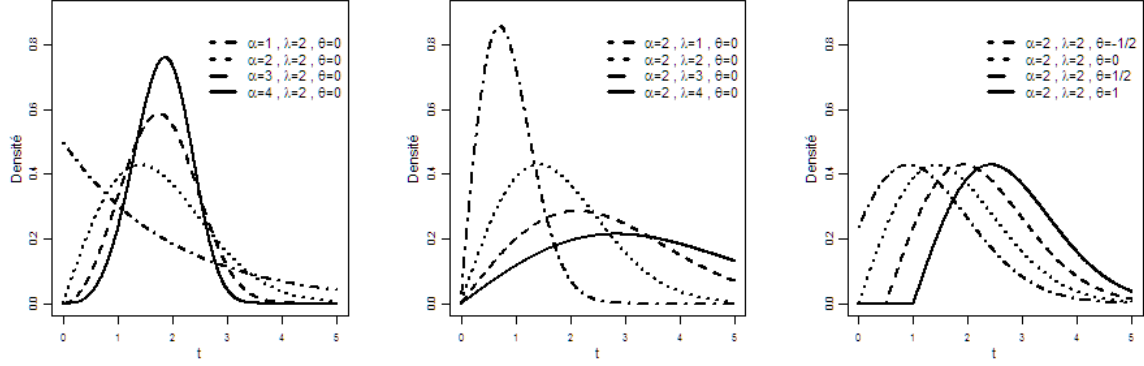


FIGURE 3.2 – Impact des paramètres sur la densité de probabilité de la loi Weibull.



Si $T \sim \text{Weibull}(\alpha, \lambda, \theta)$ et n un entier positif, alors le n^{e} moment est donné par la formule (3.1), ce qui entraîne l'espérance donnée en (3.3) et la variance donnée en (3.4).

$$\mathbb{E}[T^n] = \int_{-\infty}^{\infty} t^n f(t) dt = \frac{\alpha}{\lambda} \int_{\theta}^{\infty} t^n \left(\frac{t-\theta}{\lambda} \right)^{\alpha-1} e^{-(\frac{t-\theta}{\lambda})^\alpha} dt \quad (3.1)$$

Posons $u = \left(\frac{t-\theta}{\lambda} \right)^\alpha$, alors $du = \frac{\alpha(t-\theta)^{\alpha-1}}{\lambda^\alpha} dt$. Le tout implique que $t = \lambda u^{1/\alpha} + \theta$. Le changement de variable change le support de $t \in [\theta, +\infty[$ pour $u \in [0, +\infty[$. Par conséquent, la formule (3.1) devient :

$$\begin{aligned} \mathbb{E}[T^n] &= \frac{\alpha}{\lambda} \int_0^{\infty} t^n \frac{u}{\left(\frac{t-\theta}{\lambda} \right)} e^{-u} \frac{\lambda^\alpha}{\alpha(t-\theta)^{\alpha-1}} du \\ &= \frac{\alpha}{\lambda} \int_0^{\infty} \left(\lambda u^{1/\alpha} + \theta \right)^n \frac{u}{\left(\frac{t-\theta}{\lambda} \right)} e^{-u} \frac{\lambda}{\alpha} \frac{\lambda^{\alpha-1}}{(t-\theta)^{\alpha-1}} du \\ &= \frac{\alpha}{\lambda} \int_0^{\infty} \left(\lambda u^{1/\alpha} + \theta \right)^n \frac{u}{\left(\frac{t-\theta}{\lambda} \right)} e^{-u} \frac{\lambda}{\alpha} \frac{\left(\frac{t-\theta}{\lambda} \right)}{u} du \\ &= \int_0^{\infty} \left(\lambda u^{1/\alpha} + \theta \right)^n e^{-u} du \end{aligned} \quad (3.2)$$

Avec la formule (3.2), on trouve les deux premiers moments de T , puis en découle sa variance.

$$\mathbb{E}[T] = \lambda \Gamma \left(1 + \frac{1}{\alpha} \right) + \theta \quad (3.3)$$

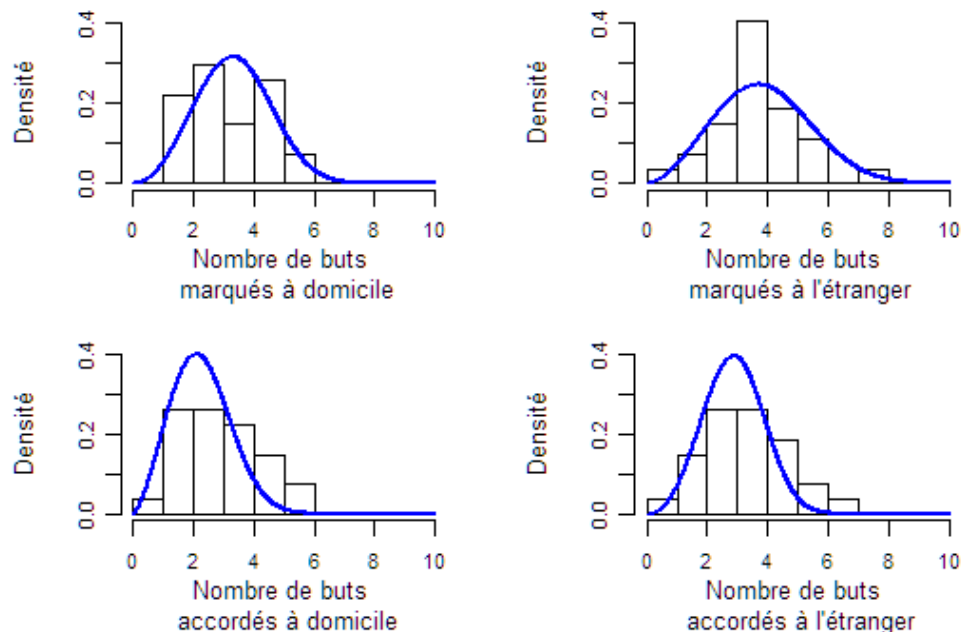
$$\mathbb{E}[T^2] = \lambda^2 \Gamma \left(1 + \frac{2}{\alpha} \right) + 2\lambda\theta \Gamma \left(1 + \frac{1}{\alpha} \right) + \theta^2$$

$$\text{Var}[T] = \mathbb{E}[T^2] - (\mathbb{E}[T])^2 = \lambda^2 \left\{ \Gamma \left(1 + \frac{2}{\alpha} \right) - \left(\Gamma \left(1 + \frac{1}{\alpha} \right) \right)^2 \right\} \quad (3.4)$$

3.1.2 Estimation des paramètres du premier modèle de simulation

Tout d'abord, il sera intéressant de distinguer les performances offensives et défensives entre les matchs à domicile et ceux à l'étranger pour chacune des équipes. Le tout implique que chacune des équipes possèdera quatre distributions de Weibull différentes. Avant toutes choses, il faut se méfier du domaine de la densité de la loi de Weibull qui exclut θ et toutes les valeurs inférieures. Sachant que le minimum de buts possibles pour une équipe lors d'une partie de hockey est de 0, il serait judicieux d'utiliser un paramètre de position, θ , négatif. D'ailleurs, Dayaratna et Miller (2013) suggèrent d'abord de fixer ce paramètre à $-0,5$ pour assurer une optimisation juste des deux autres paramètres. Donc, la vraisemblance pour α et λ est calculée en posant $\theta = -0,5$, à partir des données pour les buts marqués des équipes lors des affrontements intra-associations de la première moitié de la saison régulière. Puis, ils recommandent de repositionner la borne inférieure du domaine à 0 en ajoutant $0,5$ à la valeur de θ afin de retrouver une loi de Weibull à deux paramètres (où $\theta = 0$). Ainsi, en maximisant la vraisemblance de nos paramètres en supposant les parties indépendantes, on obtient, par exemple, que les buts marqués à domicile pour les Canadiens de Montréal, pour la première moitié de la saison 2015-2016, suivent une Weibull de paramètres $\alpha = 3,044$, $\lambda = 3,758$ et $\theta = -0,5$, tandis que ceux à l'étranger suivent une Weibull de paramètres $\alpha = 2,710$, $\lambda = 4,362$ et $\theta = -0,5$. Au cours de la même année, leurs buts accordés à domicile et à l'étranger suivent des Weibull de paramètres $\alpha = 2,540$, $\lambda = 2,558$ et $\theta = -0,5$, et $\alpha = 3,280$, $\lambda = 3,217$ et $\theta = -0,5$ respectivement. Les résultats obtenus des quatre maximisations, suivis par le repositionnement à $\theta = 0$, ont été tracés dans la figure 3.3.

FIGURE 3.3 – Comparaison des buts marqués et des buts accordés en temps réglementaire lors de la saison 2015-2016 pour les Canadiens de Montréal.



3.1.3 Fonctionnement du premier modèle de simulation

Un affrontement consiste à confronter deux équipes, où l'une sera l'équipe à domicile et l'autre l'équipe à l'étranger. Par exemple, on peut s'imaginer que nous sommes le 27 février 2016 et que les Canadiens de Montréal accueillent les Maple Leafs de Toronto au Centre Bell, ainsi Montréal représentera l'équipe à domicile et Toronto, celle à l'étranger. Pour modéliser le résultat de l'affrontement, en temps réglementaire, il faut d'abord considérer que la puissance offensive d'une équipe i (Weibull $(\alpha_{O,i}, \lambda_{O,i}, \theta_{O,i})$) est corrigée par le niveau de la défense de son adversaire j (Weibull $(\alpha_{D,j}, \lambda_{D,j}, \theta_{D,j})$) par rapport à la ligue entière (Weibull $(\alpha_D, \lambda_D, \theta_D)$), selon l'équation (3.5). Laisser tomber cette correction impliquerait qu'une équipe aurait les mêmes probabilités de marquer un certain nombre de buts, peu importe son adversaire. Une fois les puissances offensives corrigées, chacune des deux équipes doit générer une valeur aléatoire issue de leur loi de Weibull corrigée (Weibull $(\alpha_{O,i}^*, \lambda_{O,i}^*, \theta_{O,i}^*)$). Finalement, le pointage officiel correspond à la partie entière de ces valeurs générées, auxquels on a ajouté une valeur de 0,5 pour s'assurer que le minimum possible soit de 0. En d'autres mots, si on considère X , le nombre de buts marqués par une équipe, tel que $X \sim \text{Weibull}(\alpha, \lambda, \theta)$, alors Y , le pointage officiel de cette équipe, est $Y = \lfloor X + 0,5 \rfloor$.

$$\text{Weibull}(\alpha_{O,i}^*, \lambda_{O,i}^*, \theta_{O,i}^*) = \text{Weibull}\left(\left\{\alpha_{O,i} \times \frac{\alpha_{D,j}}{\alpha_D}\right\}, \left\{\lambda_{O,i} \times \frac{\lambda_{D,j}}{\lambda_D}\right\}, \{-0,5\}\right) \quad (3.5)$$

Le tableau 3.1 montre tous les paramètres des lois Weibull offensives et défensives utiles dans la simulation de notre affrontement du 27 février 2016. Par exemple, les corrections impliquent que les Maple Leafs, lorsqu'ils représentent l'équipe à l'étranger tout en affrontant les Canadiens, leurs buts marqués suivent une Weibull de paramètres $\alpha = 2,039$, $\lambda = 2,096$ et $\theta = -0,5$, au lieu d'être une Weibull de paramètres $\alpha = 1,655$, $\lambda = 2,892$ et $\theta = -0,5$.

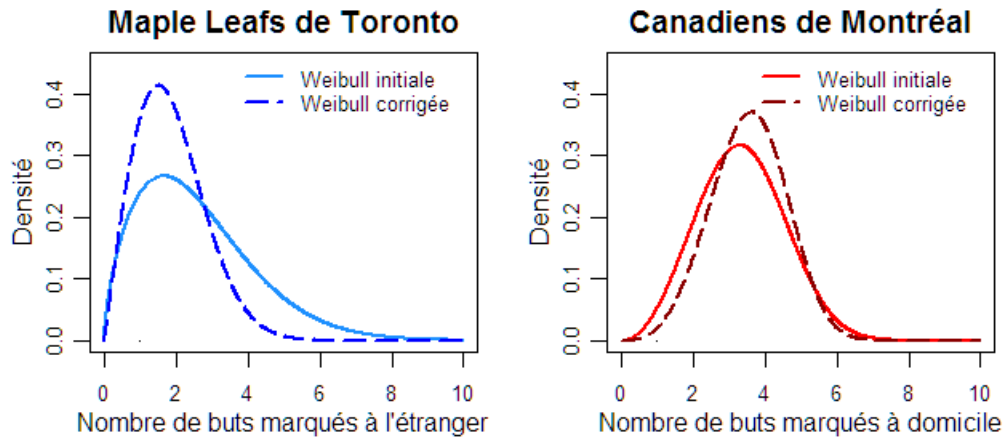
TABLE 3.1 – Paramètres estimés des Weibull pour un affrontement entre les Maple Leafs et les Canadiens, à Montréal, pour la saison 2015-2016.

Équipe	TOR			MTL		
	α_{Etr}	λ_{Etr}	θ_{Etr}	α_{Dom}	λ_{Dom}	θ_{Dom}
Offensifs de l'équipe (O,i)	1,655	2,892	-0,5	3,044	3,758	-0,5
Défensifs de l'adversaire (D,j)	2,540	2,558	-0,5	2,932	3,559	-0,5
Défensifs de la ligue (D)	2,062	3,529	-0,5	2,342	3,403	-0,5
Offensifs corrigés (O^*,i)	2,039	2,097	-0,5	3,811	3,930	-0,5

La figure 3.4, qui considère $\theta = 0$, montre que ce changement de paramètres concentre le nombre de buts que les Maple Leafs peuvent marquer vers de plus petites valeurs. En effet, l'espérance et la variance de leur distribution passent respectivement de 2,585 et 2,574 à 1,924 et 2,096. Ainsi, sans correction pour la défense, nous aurions pu générer les valeurs 1,796

pour Montréal et 3,355 pour Toronto, correspondant respectivement aux vingtième et quatre-vingtième percentiles de leur distribution. Avec correction, les vingtième et quatre-vingtième percentiles de leur distribution corrigée correspondent à des valeurs de 2,151 pour Montréal et 2,148 pour Toronto. Par conséquent, la partie entière des valeurs générées augmentées de 0,5 nous dit que la correction pour la défensive a fait passer le match de 3 à 2 en faveur de Toronto à une égalité de 2 à 2, à la fin du temps réglementaire. L'effet correctif, sans pour autant causer d'énormes changements, force dans cet exemple une prolongation. Advenant le

FIGURE 3.4 – Densité de probabilité des Weibull pour les buts marqués des Maple Leafs et des Canadiens, pour la première moitié de la saison 2015-2016.



cas où le résultat des deux équipes est identique après la simulation de l'affrontement, il y aura une prolongation. L'abolition du match nul à la suite du lock-out de 2004-2005 assure qu'il y aura un vainqueur dans tous les matchs. Dans le but de simplifier la simulation d'une prolongation, il a été supposé que chacune des équipes détient une puissance en surtemps basé sur leur ratio victoires-défaites en surtemps acquis lors de la première moitié de la saison. Cette hypothèse considère que le ratio de victoires en surtemps des Canadiens est de 0,500 (3 victoires et 3 défaites), et que celui des Maple Leafs est de 0,364 (4 victoires et 7 défaites). Ainsi, en utilisant l'équation (3.6) décrivant la probabilité qu'une équipe A remporte la prolongation face à une équipe B, sachant que $\mathbb{P}[A|P]$ désigne la probabilité que l'équipe A gagne sachant qu'elle se retrouve en prolongation, on trouve que les Canadiens auront 57,89 % de chance de remporter le surtemps face aux Maple Leafs.

$$\mathbb{P}[A] = \frac{\mathbb{P}[A|P]}{\mathbb{P}[A|P] + \mathbb{P}[B|P]} \quad (3.6)$$

Si on considère la variable aléatoire X désignant le nombre de buts marqués par une équipe A et la variable aléatoire Y désignant le nombre de buts marqués par une équipe B, on peut définir avec précision certaines probabilités entre ces deux équipes. D'ailleurs, l'utilisation de la partie entière des valeurs générées par les lois rend nos variables aléatoires X et Y discrètes.

Ainsi, la probabilité qu'une équipe A marque un nombre de buts k (étant un entier naturel) est donnée par l'équation (3.7).

$$\mathbb{P}[X = k] = \mathbb{P}[X < k + \theta + 1] - \mathbb{P}[X < k + \theta] \quad (3.7)$$

La connaissance des deux lois indépendantes confrontées permet de construire le tableau 3.2 regroupant la fonction de masse conjointe des buts marqués entre les Maple Leafs et les Canadiens, si l'affrontement est à Montréal. En utilisant l'équation (3.8), on trouve que la probabilité moyenne qu'une équipe subisse un jeu blanc dans le match est de 10,13 %. La somme des éléments de la diagonale correspond à la probabilité que l'issue de l'affrontement se détermine en prolongation ($\mathbb{P}[X = Y] = 0.1332$). La somme des éléments sous la diagonale correspond à la probabilité que les Canadiens gagnent en temps réglementaire ($\mathbb{P}[X > Y] = 0.7973$) et la somme des éléments au-dessus de la diagonale correspond à la probabilité que les Maple Leafs gagnent en temps réglementaire ($\mathbb{P}[X < Y] = 0.0695$). Finalement, on peut déduire avec l'équation (3.9) que la probabilité que les Canadiens gagnent l'affrontement est de 87,44 %. Ces derniers résultats dévoilent que le modèle prédit un affrontement plutôt facile pour les Canadiens à Montréal.

$$\mathbb{P}[k = 0] = \frac{\mathbb{P}[\{X = 0\} \cup \{Y = 0\}] - \mathbb{P}[\{X = 0\} \cap \{Y = 0\}]}{2} \quad (3.8)$$

$$\mathbb{P}[\text{Équipe A gagne}] = \mathbb{P}[X > Y] + \mathbb{P}[X = Y] \times \mathbb{P}[A] \quad (3.9)$$

TABLE 3.2 – Extrait de la fonction de masse conjointe des buts marqués lors d'un affrontement entre les Maple Leafs et les Canadiens à Montréal pendant la saison 2015-2016.

		Buts marqués par TOR						
		0	1	2	3	4	...	Total
Buts marqués par MTL	0	0,001	0,002	0,002	0,001	0,000	...	0,005
	1	0,013	0,027	0,019	0,007	0,001	...	0,068
	2	0,045	0,091	0,063	0,023	0,005	...	0,227
	3	0,071	0,142	0,099	0,036	0,008	...	0,356
	4	0,052	0,104	0,073	0,027	0,006	...	0,261
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	Total	0,198	0,398	0,278	0,101	0,021	...	1,000

3.2 Construction du deuxième, troisième et quatrième modèles de simulation

Une modélisation plus sophistiquée peut décrire une rencontre en différents événements. Dans le cadre de cet essai, le seul événement possible sera un but, mais les événements auraient

bien pu inclure des pénalités ou même des lancers au filet. La simulation d'une séquence d'événements permettra, entre autres, de reproduire un mécanisme d'influence stratégique pour les équipes obéissant à l'écart dans le pointage à un moment précis de l'affrontement. En fait, l'utilisation de processus de Poisson permettra de simuler le temps nécessaire à l'arrivée du prochain événement, pour ainsi permettre aux équipes d'ajuster leurs intensités, offensive et défensive, ce qui permet de simuler le temps requis avant l'arrivée de l'événement suivant, et ainsi de suite, jusqu'à ce que les soixante minutes de temps réglementaire soient écoulées.

À l'intérieur de cette section, où la théorie provient essentiellement de Caumel (2015) et Hines et collab. (2005), trois modèles simulant des séquences d'événements seront présentés en commençant par un modèle simple initiant le fonctionnement des processus de Poisson, puis une autre modélisation considèrera l'utilisation de modèle linéaire généralisé avec lien Poisson pour capter un effet gardien de but, finalement un dernier modèle améliorera le précédent en divisant les affrontements en deux périodes qui distingueront les 55 premières minutes d'un match des 5 dernières.

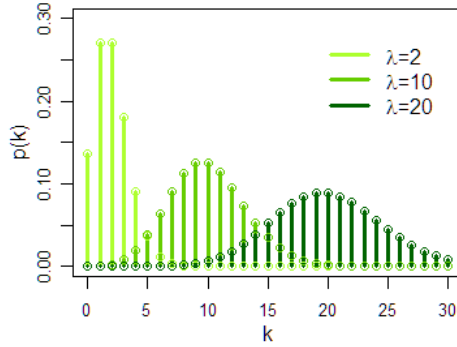
3.2.1 Principales propriétés de la loi de Poisson

La loi de Poisson est une loi de probabilité discrète qui décrit le comportement du nombre d'événements se produisant dans un laps de temps fixé, si ces événements se produisent avec une fréquence moyenne connue et indépendamment du temps écoulé depuis l'événement précédent. Si le nombre moyen d'occurrences dans cet intervalle est λ , un nombre réel positif, alors la probabilité qu'il survienne exactement k occurrences (k étant un entier naturel) est donnée par :

$$p(k) = \mathbb{P}(X = k) = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda}, & \text{si } k \in \mathbb{N} \\ 0, & \text{autrement.} \end{cases}$$

Par exemple, si un événement se produit en moyenne 2 fois par minute, pour étudier le nombre d'événements se produisant dans un laps de temps de 50 minutes, on le représente par une loi de Poisson de paramètre $\lambda = 100$ (2 fois par minute \times 50 minutes). La figure 3.5 montre l'importance que joue l'unique paramètre sur la forme de la fonction de masse de cette loi. En fait, si X est une variable aléatoire réelle telle que $X \sim \text{Poisson}(\lambda)$, alors pour tout entier positif n le n^e moment factoriel de X est donné par l'équation (3.10), ce qui implique une égalité pratique entre le paramètre λ , l'espérance de X et la variance de X , comme le montre (3.11). Par exemple, si $X \sim \text{Poisson}(100)$, alors l'espérance de la variable aléatoire sera de 100 et son écart-type, de 10.

FIGURE 3.5 – Impact du paramètre sur la fonction de masse de la loi de Poisson.



$$\begin{aligned}
 \mathbb{E}[X(X-1)(X-2)\dots(X-(n-1))] &= \sum_{k=0}^{\infty} k(k-1)(k-2)\dots(k-(n-1))e^{-\lambda} \frac{\lambda^k}{k!} \\
 &= e^{-\lambda} \lambda^n \sum_{k=n}^{\infty} \frac{\lambda^{k-n}}{(k-n)!} \\
 &= e^{-\lambda} \lambda^n e^{\lambda} = \lambda^n
 \end{aligned} \tag{3.10}$$

$$\begin{aligned}
 \mathbb{E}[X] = \lambda \quad \text{et} \quad \text{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\
 &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 \\
 &= \mathbb{E}[X(X-1)] - \mathbb{E}[X] + (\mathbb{E}[X])^2 \\
 &= \lambda^2 - \lambda(\lambda-1) = \lambda
 \end{aligned} \tag{3.11}$$

3.2.2 Principales propriétés de la loi exponentielle

D'un autre côté, une variable aléatoire de loi exponentielle modélise des temps aléatoires tels que les temps d'attente entre les arrivées successives d'événements. La loi exponentielle est une loi de probabilité continue correspondant à une loi Weibull avec un paramètre de forme $\alpha = 1$ et un paramètre de position $\theta = 0$. Cependant, le paramètre d'échelle λ de la loi exponentielle sera exprimé comme l'inverse du λ de la loi Weibull. Par conséquent, la fonction de répartition et la densité de la loi exponentielle (λ) (pour $t \geq 0$) sont respectivement données par :

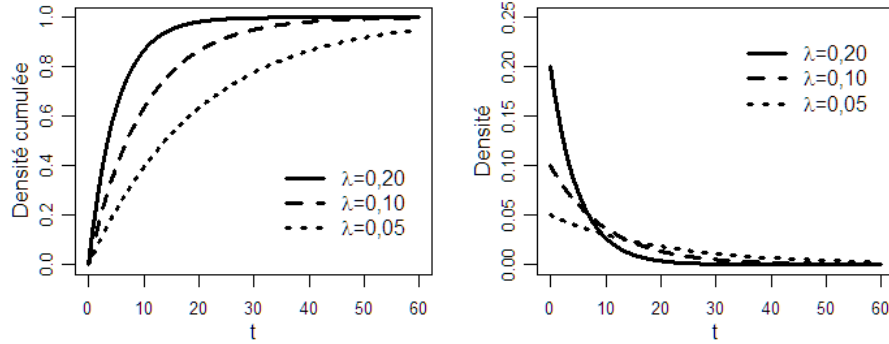
$$F(t) = \begin{cases} 1 - e^{-(\lambda t)}, & \text{si } t \geq 0 \\ 0, & \text{si } t < 0 \end{cases}$$

$$f(t) = \begin{cases} \lambda e^{-(\lambda t)}, & \text{si } t \geq 0 \\ 0, & \text{si } t < 0 \end{cases}$$

Ainsi, si un événement se produit en moyenne toutes les deux minutes lors d'une période de

temps choisi, alors l'arrivée du prochain événement suit une loi exponentielle de paramètre $\lambda = 0,5$ (1 occurrence / 2 minutes). La figure 3.6 montre l'importance que joue l'unique paramètre sur la forme de la fonction de répartition et de la fonction densité de la loi exponentielle. Si $T \sim \text{Exp}(\lambda)$, à partir de la loi Weibull, il est simple de déduire l'espérance de T donnée en (3.12) et la variance de T donnée en (3.13).

FIGURE 3.6 – Impact du paramètre sur la fonction de répartition et sur la fonction de densité de la loi exponentielle.



$$\mathbb{E}[T] = \frac{1}{\lambda} \Gamma(1+1) = \frac{1}{\lambda} \quad \text{et} \quad \mathbb{E}[T^2] = \frac{1}{\lambda^2} \Gamma(1+2) = \frac{2}{\lambda^2} \quad (3.12)$$

$$\text{Var}[T] = \mathbb{E}[T^2] - (\mathbb{E}[T])^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2} \quad (3.13)$$

L'une des propriétés importantes de la distribution exponentielle est son absence de mémoire définie en (3.14). Imaginons que T représente la durée de vie d'une ampoule avant qu'elle ne tombe en panne : la probabilité qu'elle dure au moins $s + t$ heures sachant qu'elle a déjà duré t heures sera la même que la probabilité de durer s heures à partir de sa mise en fonction initiale. En d'autres mots, le fait qu'elle ne soit pas tombée en panne pendant t heures ne change rien à son espérance de vie à partir du temps t .

$$\mathbb{P}[T > s + t | T > t] = \mathbb{P}[T > s] \quad \forall s, t \geq 0 \quad (3.14)$$

De plus, la propriété d'additivité de la loi exponentielle nous dit que si X_1, X_2, \dots, X_m sont des variables aléatoires indépendantes telles que pour chaque $i = \{1, 2, \dots, m\}$ on observe que $X_i \sim \text{Exp}(\lambda_i)$, alors $Z = \min\{X_1, X_2, \dots, X_m\} \sim \text{Exp}(\lambda)$, où λ égale la somme de tous les λ_i comme en (3.15). Par exemple, si on considère une lumière clignotant en moyenne aux 10 minutes, et une autre aux 15 minutes, alors on observera un clignotement de lumière en moyenne toutes les 6 minutes. En fait, on peut démontrer cette propriété à l'aide de la fonction

de répartition de Z (avec $z > 0$) :

$$\begin{aligned}
\mathbb{P}[Z_m \leq z] &= \mathbb{P}[\min \{X_1, X_2, \dots, X_m\} \leq z] \\
&= 1 - \mathbb{P}[\min \{X_1, X_2, \dots, X_m\} > z] \\
&= 1 - \mathbb{P}[(X_1 > z) \cap (X_2 > z) \cap \dots \cap (X_m > z)] \\
&= 1 - \prod_{i=1}^m \mathbb{P}[X_i > z] = 1 - \prod_{i=1}^m e^{-\lambda_i z} = 1 - e^{-(\lambda_1 + \lambda_2 + \dots + \lambda_m)z} \quad (3.15)
\end{aligned}$$

D'un autre côté, on pourrait s'intéresser à la probabilité que la lumière clignotant en moyenne aux 10 minutes clignote en premier. Le résultat en (3.16) confirme que cette probabilité vaut 0,6. La probabilité qu'une variable aléatoire $X_1 \sim \text{Exp}(\lambda_1)$ surpasse une variable aléatoire $X_2 \sim \text{Exp}(\lambda_2)$, correspond à l'importance relative du paramètre λ_2 par rapport à la somme des paramètres des deux variables aléatoires ($\lambda_1 + \lambda_2$). Effectivement :

$$\begin{aligned}
\mathbb{P}[X_1 > X_2] &= \int_0^\infty \int_0^{x_1} f(x_2) dx_2 \times f(x_1) dx_1 \\
&= \int_0^\infty (1 - e^{-\lambda_2 x_1}) \times \lambda_1 e^{-\lambda_1 x_1} dx_1 \\
&= \int_0^\infty \lambda_1 e^{-\lambda_1 x_1} - \lambda_1 e^{-(\lambda_1 + \lambda_2) x_1} dx_1 \\
&= 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{\lambda_2}{\lambda_1 + \lambda_2} \quad (3.16)
\end{aligned}$$

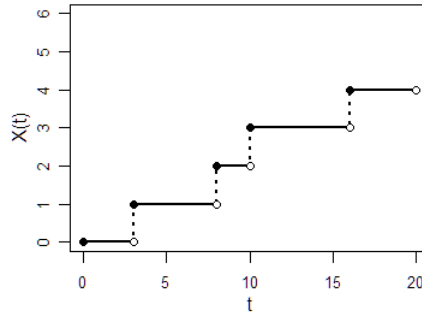
3.2.3 Processus de Poisson

Pour sa part, un processus de comptage (dénombrement) est un processus stochastique à valeurs dans \mathbb{N} modélisant un nombre entier aléatoire évoluant dans le temps. Sachant que le temps u est continu, un processus de comptage $W(u)$ regroupe des sauts d'amplitude $\delta(u)$ dans \mathbb{N} correspondant à des instants aléatoires où se produisent l'événement étudié. Le processus de comptage le plus classique est le processus de Poisson. Le processus de Poisson à un moment t , noté $X(t)$, compte le nombre d'occurrences ($\delta(t) = 1$) survenues entre le moment initial du processus aléatoire et le moment t . Par conséquent, les processus de Poisson sont définis par l'équation (3.17) et combinent les principales propriétés de la loi de Poisson et de la loi exponentielle. Dans les faits, ils permettent de compter le nombre d'appels dans un autocommutateur, le nombre d'émissions de particules radioactives ou bien le nombre de clients qui se sont présentés devant un guichet jusqu'à un instant donné.

$$X(t) = \int_0^t W(u) du \quad (3.17)$$

Ainsi, si au cours d'une période de 20 minutes on observe 4 occurrences ($t_1 = 3, t_2 = 8, t_3 = 10, t_4 = 16$) de l'événement étudié, alors la progression du processus de Poisson sera donnée par la figure 3.7. D'ailleurs, l'intensité λ du processus de Poisson désigne un taux d'occurrences par unité de temps. Dans l'exemple, cette intensité λ cheminera continuellement pour atteindre une intensité globale par minute de $\lambda = 0,2$ (4 occurrences / 20 minutes). Également, on peut aussi définir que le temps moyen écoulé entre deux occurrences successives suit une loi exponentielle de paramètre $\lambda = 0,2$.

FIGURE 3.7 – Exemple d'un processus de Poisson à travers le temps.



3.2.4 Estimation des paramètres du deuxième modèle de simulation

Tout comme l'on fait Merritt et Clauset (2014), il est intéressant d'utiliser les processus de Poisson afin de modéliser le rythme des buts dans une partie de hockey, puis d'y intégrer un mécanisme quelconque d'ajustement stratégique des équipes. Par conséquent, l'hypothèse d'indépendance entre les buts marqués des deux adversaires sera éliminée, ainsi le réalisme en termes d'écart de pointage augmentera. Bref, ce type de modélisation devrait surpasser notre premier modèle, puisqu'il améliorerait la prédiction de l'issue des matchs incluant des équipes ayant tendance à participer à des rencontres serrées. En réalité, les prochaines modélisations considèreront l'existence de trois états possibles pour une équipe lors d'une rencontre, soit l'équipe est en avance (L), à égalité (E), ou en retard (T) dans le pointage.² Le choix de regrouper les écarts d'un but et plus ensemble est critiquable, sauf que limiter le nombre d'états existant améliore la précision des estimations. Pour une même équipe, chacun des k états possèdera une intensité offensive (notée ρ_L, ρ_E et ρ_T) représentant le nombre de buts moyens marqués par l'équipe par minute et une intensité défensive (notée ψ_L, ψ_E et ψ_T) correspondant au nombre de buts moyens accordés par l'équipe par minute. Comme au premier modèle, les puissances offensives et défensives entre les matchs à domicile et ceux à l'étranger seront distinguées. À partir des données pour les buts marqués des équipes lors des affrontements intra-associations de la première moitié de la saison régulière, on calcule les intensités comme en (3.18) pour obtenir le tableau 3.3 pour les Canadiens de Montréal lors de la saison

2. Les indices utilisés seront : L pour « Lead », E pour « Even » et T pour « Trail ».

2015-2016. Ainsi, on voit que leurs intensités offensives sont relativement élevées par rapport à leurs intensités défensives et qu'ils accordent très rarement un but lorsqu'ils tirent de l'arrière dans un affrontement à domicile.

$$\rho_k = \frac{\text{Buts marqués totaux à l'état } k}{\text{Minutes jouées totales à l'état } k} \quad , \quad \psi_k = \frac{\text{Buts accordés totaux à l'état } k}{\text{Minutes jouées totales à l'état } k} \quad (3.18)$$

TABLE 3.3 – Intensités offensives et défensives des Canadiens de Montréal pour la première moitié de la saison 2015-2016.

État de l'équipe	L_{Etr}	E_{Etr}	T_{Etr}	L_{Dom}	E_{Dom}	T_{Dom}
Minutes jouées totales	460,43	303,07	196,50	319,03	348,63	112,33
Buts marqués totaux	22	20	10	16	14	7
Buts accordés totaux	20	11	6	11	9	1
Intensité offensive (ρ)	0,0478	0,0660	0,0509	0,0502	0,0402	0,0623
Intensité défensive (ψ)	0,0434	0,0363	0,0305	0,0345	0,0258	0,0089

3.2.5 Fonctionnement du deuxième modèle de simulation

En reprenant l'affrontement du 27 février 2016, où les Canadiens de Montréal accueillent les Maple Leafs de Toronto au Centre Bell, il sera possible de comparer les différents résultats simulés par ce modèle à ceux du modèle précédent. Rappelons qu'ici les Canadiens représenteront l'équipe à domicile et les Maple Leafs, celle à l'étranger. Avant de commencer la modélisation, il faut estimer les trois intensités offensives (notées ρ_L , ρ_E et ρ_T) et défensives (notées ψ_L , ψ_E et ψ_T) pour une équipe i à domicile, puis estimer celle de son adversaire j à l'étranger à l'aide des équations en (3.18). Les différentes valeurs estimées représentent en quelque sorte les niveaux d'effort fourni par les équipes pour les divers états. Pour qu'une équipe opportuniste, dans un état précis, le soit davantage si son opposant a tendance à se relâcher défensivement dans la situation contraire, il sera nécessaire de garder un mécanisme de correction des différentes intensités offensives. En fait, il s'agira de considérer que l'intensité offensive corrigée (notées Λ_L , Λ_E et Λ_T) d'une équipe i correspond à la moyenne entre son intensité offensive de l'état k et l'intensité défensive de l'état opposée l de son adversaire j , comme le montre l'équation (3.19).

$$\Lambda_{i,k} = \frac{\rho_{i,k} + \psi_{j,l}}{2} \quad (3.19)$$

Comme le montre le tableau 3.4, les estimations des intensités défensives pour les Maple Leafs, à l'étranger, sont de $\psi_L = 0,0554$ (lorsqu'ils mènent le pointage ou lorsque leur adversaire tire de l'arrière), de $\psi_E = 0,0323$ (lorsque le pointage est égal), et de $\psi_T = 0,0444$ (lorsqu'ils tirent de l'arrière dans le pointage ou leur adversaire mène). De leur côté, les Canadiens, à

domicile, possèdent des intensités défensives beaucoup plus faibles. Il s'agit du contraire pour les intensités offensives, où les Canadiens affichent de plus grandes valeurs que les Maple Leafs. Par conséquent, les intensités offensives corrigées associées aux Canadiens, surpassent celles des Maple Leafs, valant $\Lambda_L = 0,0473$, $\Lambda_E = 0,0363$ et $\Lambda_T = 0,0589$.

TABLE 3.4 – Intensités offensives et défensives estimées par le deuxième modèle pour un affrontement entre les Maple Leafs et les Canadiens, à Montréal, pour la saison 2015-2016.

Équipe	TOR			MTL		
Intensité (Indice)	L_{Etr}	E_{Etr}	T_{Etr}	L_{Dom}	E_{Dom}	T_{Dom}
ρ de l'équipe (i,k)	0,0308	0,0291	0,0363	0,0502	0,0402	0,0623
ψ de l'adversaire (j,l)	0,0089	0,0258	0,0345	0,0444	0,0323	0,0554
Λ de l'équipe (i,k)	0,0199	0,0275	0,0354	0,0473	0,0363	0,0589

Grâce à la propriété d'additivité de la loi exponentielle montrée en (3.15), l'intensité λ du processus de Poisson modélisant le nombre de buts marqués lors d'un état précis pendant l'affrontement peut s'écrire comme une addition d'intensités offensives corrigées, comme en (3.20). Autrement dit, la cadence des buts est définie par une loi exponentielle ayant comme paramètre, la somme de deux intensités offensives corrigées. Le tableau 3.5 indique que l'intensité du processus de Poisson correspond à $\lambda_E = 0,0638$ lorsque le pointage dans l'affrontement est égal. Advenant le cas où les Canadiens prennent l'avance, l'intensité deviendra $\lambda_L = 0,0827$, dans le cas contraire elle sera de $\lambda_T = 0,0788$. D'un autre côté, la probabilité qu'une équipe i soit responsable de la survenue d'un événement lors de l'état k découle du résultat (3.16) réécrit autrement à l'équation (3.21). C'est pourquoi la probabilité que les Canadiens soient responsables d'un événement est de 56,90 % si le pointage est égal avant l'événement, de 57,19 % si les Canadiens sont déjà en avance et de 74,75 % si ce sont les Maple Leafs qui sont en avance.

$$\lambda_k = \Lambda_{i,k} + \Lambda_{j,l} \quad (3.20)$$

$$\mathbb{P}[\text{Marquer}_{i,k}] = \frac{\Lambda_{i,k}}{\lambda_k} \quad (3.21)$$

Une simulation parmi tant d'autres pourrait être celle que l'on retrouve au tableau 3.6. Elle montre une victoire de 3 à 2, en temps réglementaire, des Canadiens face aux Maple Leafs au Centre Bell. Évidemment, toute rencontre commence par un pointage nul de zéro à zéro. Donc, le temps écoulé avant l'arrivée du premier but marqué suivra toujours une exponentielle de paramètre λ_E . Si ce but survient avant la soixantième minute de jeu, un nouvel état sera défini en fonction de l'équipe ayant été responsable de l'événement. D'ailleurs, à chaque but marqué avant la soixantième minute de jeu, l'état de l'affrontement sera réévalué. Lorsque l'équipe

TABLE 3.5 – Intensités des processus de Poisson estimés par le deuxième modèle pour un affrontement entre les Maple Leafs et les Canadiens, à Montréal, pour la saison 2015-2016.

Intensité (Indice)	L_{Dom}	E_{Dom}	T_{Dom}
Λ de l'équipe (i,k)	0,0473	0,0363	0,0589
Λ de l'adversaire (j,l)	0,0354	0,0275	0,0199
λ du match (k)	0,0827	0,0638	0,0788
\mathbb{P} de marquer (i,k)	0,5719	0,5690	0,7475

à domicile prend les devants dans le pointage, le temps écoulé jusqu'au prochain but suivra une exponentielle de paramètre λ_L et si c'est l'équipe à l'étranger qui prend les devants, alors le temps écoulé jusqu'au prochain but suivra une exponentielle de paramètre λ_T . Lorsque le temps total de l'affrontement atteint soixante minutes, la rencontre est officiellement terminée. Par contre, si le pointage est égal, une prolongation vient techniquement ajouter un cinq minutes supplémentaire à la rencontre. Toutefois, pour des raisons de simplicité, la probabilité de remporter le surtemps pour les différentes équipes est identique à leur probabilité de briser l'état d'égalité. En répétant 10000 fois la simulation, les Canadiens affichent une probabilité de victoire d'approximativement 68,62 % ce qui correspond à une chute de près de 20 % par rapport au premier modèle. Au final, la fréquence relative de prolongation a doublé en passant à 26,07 %, et le pourcentage de jeu blanc est resté relativement stable avec 9,06 %.

TABLE 3.6 – Exemple d'affrontement simulé à partir des processus de Poisson entre les Maple Leafs et les Canadiens, à Montréal, pour la deuxième moitié de la saison 2015-2016.

Pointage TOR	Pointage MTL	Intensité	Temps initial	Temps final	$\mathbb{P}[\text{But MTL}]$	But
0	0	0,0638	0,00	1,28	0,5690	TOR
1	0	0,0788	1,28	8,18	0,7475	MTL
1	1	0,0638	8,18	24,78	0,5690	MTL
1	2	0,0827	24,78	46,86	0,5719	MTL
1	3	0,0827	46,86	57,71	0,5719	TOR
2	3	0,0827	57,71	60,00	0,5719	—

3.2.6 Modèle linéaire généralisé

D'après McCullagh et Nelder (1989), les modèles linéaires généralisés (MLG) permettent d'étudier la liaison entre une variable endogène (Y) et un ensemble de variables exogènes (X_1, \dots, X_k). Ils englobent, en outre, le modèle linéaire (régression multiple, analyse de la variance et analyse de la covariance), le modèle log-linéaire, la régression logistique et la régression de Poisson. En premier lieu, une composante aléatoire identifie la distribution de probabilités de la variable à expliquer. On suppose que l'échantillon statistique est constitué de n variables

aléatoires $(Y_i, \forall i = 1, \dots, n)$ indépendantes admettant des distributions issues d'une structure exponentielle³. Cela signifie que les lois de ces variables sont données par une même mesure dite de référence et que leur densité par rapport à cette mesure peut s'écrire sous la forme (3.22). Pour toute distribution de la famille exponentielle, on a que $\mathbb{E}[Y_i] = \mu_i = b'(\theta_i)$ et $\text{Var}[Y_i] = b''(\theta_i) a(\phi)$. La fonction $a(\phi)$ est généralement de la forme ϕ/ω_i . Le paramètre ϕ est le paramètre de dispersion et sa valeur reste constante pour toutes les observations tandis que ω_i est une pondération connue pouvant varier entre les observations.

$$f(y_i; \theta_i, \phi) = \exp\left(\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right) \quad (3.22)$$

En deuxième lieu, la composante déterministe, exprimée sous forme d'une combinaison linéaire, précise quel est le prédicteur linéaire et ses effets sur la loi de Y . Le prédicteur linéaire pour un individu i s'écrit : $\eta_i = x_i^t \beta$, où $x_i^t = (1, X_{i1}, \dots, X_{ik})$ et $\beta^t = (\beta_0, \beta_1, \dots, \beta_k)$. Donc, cette composante peut contenir une seule ou plusieurs variables explicatives et des interactions entre les variables explicatives tout en permettant que les variables explicatives puissent être numériques ou catégoriques. En dernier lieu, une troisième composante exprime une relation fonctionnelle entre l'espérance mathématique de la composante aléatoire et le prédicteur linéaire. La fonction de lien g entre les deux composantes est supposée monotone et différentiable tout en étant de la forme : $g(\mu_i) = \eta_i$. Si la fonction g est la même que le lien entre μ et le paramètre θ de la famille exponentielle, on dit que g est le lien canonique.

En combinant ces trois éléments, nous spécifions entièrement la distribution des variables endogènes étant donné les variables exogènes. Pour résumer, le modèle linéaire généralisé s'énonce comme en (3.23). Puisque l'on suppose les Y_i indépendants, la fonction de vraisemblance des paramètres β et ϕ est donnée par l'équation (3.24). En découle la log-vraisemblance de l'équation (3.25), la fonction de score de l'équation (3.26) et la matrice d'information de Fisher de l'équation (3.27).

$$Y_i \sim \xi(\mu_i, \phi) \text{ indépendantes, avec } g(\mu_i) = \eta_i, \quad \forall i = 1, \dots, n \quad (3.23)$$

$$L(y_i; \beta, \phi) = \exp\left\{\sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{\phi/\omega_i} + \sum_{i=1}^n c(y_i, \phi)\right\} \quad (3.24)$$

$$l(y_i; \beta, \phi) = \ln L(y_i; \beta, \phi) = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{\phi/\omega_i} + \sum_{i=1}^n c(y_i, \phi) \quad (3.25)$$

$$S(y_i; \beta, \phi) = \frac{dl(y_i; \beta, \phi)}{d\beta} = \sum_{i=1}^n \frac{y_i - b'(\theta_i)}{\phi/\omega_i} \quad (3.26)$$

3. Notamment, la famille exponentielle comprend la distribution binomiale, normale, gamma et Poisson.

$$I(y_i; \beta, \phi) = -\mathbb{E} \left[\frac{d^2 l(y_i; \beta, \phi)}{d\beta d\beta'} \right] = \mathbb{E} \left[\sum_{i=1}^n \frac{b''(\theta_i)}{\phi/\omega_i} \right] \quad (3.27)$$

Les coefficients de régression β sont estimés par leur valeur maximisant la fonction de log-vraisemblance. Ainsi, à partir de la fonction de score et l'inverse de l'information de Fisher, on peut déduire l'espérance et la variance des $\hat{\beta}$. La non-linéarité des équations du score en β nécessite l'utilisation de méthodes itératives pour estimer les paramètres par la méthode du maximum de vraisemblance. L'algorithme de Newton-Raphson en est un exemple utilisant la matrice d'information de Fisher. Cette matrice contient l'information concernant la courbure de la fonction de log-vraisemblance au point d'estimation. Plus la courbure est grande, plus l'information apportée au sujet des paramètres du modèle est importante. Formellement, on part d'un point arbitraire $\hat{\beta}^{(0)}$ appartenant à l'ensemble de définition de la fonction et on construit par récurrence la suite (3.28). Les s étapes s'arrêtent lorsque la distance entre $\hat{\beta}^{(s+1)}$ et $\hat{\beta}^{(s)}$ est jugée suffisamment petite. Un autre algorithme populaire est le « Fisher scoring ». Ce dernier possède la même méthode que l'algorithme précédent, mais utilise la matrice d'information espérée au lieu de la matrice d'information observée.

$$\hat{\beta}^{(s+1)} = \hat{\beta}^{(s)} + \left[\widehat{\text{Var}}[\hat{\beta}^{(s)}] \right]^{-1} S(\hat{\beta}^{(s)}), \quad \forall s = 1, 2, \dots \quad (3.28)$$

3.2.7 Régression de Poisson

La régression Poisson, selon Cameron et Trivedi (2005), sert à modéliser une variable réponse représentant un dénombrement. On suppose que Y suit une loi Poisson de paramètre $\mu = \mathbb{E}[Y]$ sachant que les observations Y_i sont indépendantes et que l'espérance μ est monotone en x , pour toutes les variables x du modèle. Le modèle s'énonce ainsi :

$$\begin{aligned} Y_i &\sim \text{Poisson}(\mu_i) \text{ indépendantes} \\ \ln(\pi_i) &= x_i^t \beta, \quad \forall i = 1, \dots, n \end{aligned} \quad (3.29)$$

La loi de Poisson, telle qu'écrite à l'équation (3.30), est bel et bien une distribution membre de la famille exponentielle avec paramètre canonique $\theta = \ln(\mu)$, $a(\phi) = 1$, $b(\theta) = \exp(\theta)$ et $c(y, \phi) = -\ln(y!)$.

$$f(y; \mu) = \exp \left(\frac{y \ln(\mu) - \exp(\ln(\mu))}{1} - \ln(y!) \right) = \frac{\mu^y \exp(-\mu)}{y!} \quad (3.30)$$

Le modèle comporte parfois un paramètre supplémentaire associé à la composante aléatoire du modèle : le paramètre de dispersion ϕ . En régression Poisson, ce paramètre est fixé à 1. Il arrive cependant que la variance des données soit différente de l'espérance. Ce problème est détectable, par exemple, en utilisant le ratio de la déviance, indiquant le niveau de la

dispersion. En présence de sur ou de sous dispersion, les inférences sur les coefficients ne sont plus fiables, car la matrice de variance-covariance du vecteur des estimations des paramètres est mal estimée. Pour remédier à la situation, les inférences doivent être corrigées par l'estimation du paramètre de dispersion.

Un terme d'offset est une variable exogène que l'on ajoute au prédicteur linéaire, mais pour lequel le coefficient de régression correspondant est fixé à 1. En régression de Poisson avec lien log, un tel terme est particulièrement utile lorsque nous savons que le dénombrement (valeur de Y_i) est proportionnel à une certaine variable que nous voulons inclure dans le prédicteur linéaire. Par conséquent, l'ajout d'un terme d'offset z_i au modèle (3.29) donne le modèle (3.31). Ainsi, on obtient que si z_i est multiplié par une constante k , alors $\ln(kz_i) = \ln(k) + \ln(z_i)$, ce qui multiplie la moyenne μ aussi par k , d'où la proportionnalité.

$$\begin{aligned} Y_i &\sim \text{Poisson}(\mu_i) \text{ indépendantes} \\ \ln(\pi_i) &= x_i^t \beta + \ln(z_i), \quad \forall i = 1, \dots, n \end{aligned} \tag{3.31}$$

3.2.8 Estimation des paramètres du troisième modèle de simulation

L'estimation des intensités des processus de Poisson repose, ici, sur un modèle linéaire généralisé au lieu de seulement être un ratio de buts marqués ou accordés par minute lors de divers états. Ainsi, le prochain modèle permettra d'obtenir des statistiques intéressantes quant aux estimations, comme les intervalles de confiance des différentes variables explicatives. Supposons $Y_{h,i,k}$, le nombre de buts marqués par l'équipe i pour la partie h pendant l'état k , tel que $Y_{h,i,k} \sim \text{Poisson}(\lambda_{h,i,k})$, où $\lambda_{h,i,k}$ est un nombre de buts par minute aussi appelée intensité. La régression de Poisson de l'équation (3.32) montre que l'intensité d'une équipe i lors de l'état k pour un affrontement h se définit par une combinaison linéaire entre :

- un effet associé directement à l'équipe i de l'affrontement h lors de l'état k ;
- un effet associé au nombre moyen de buts accordés par minute par la défensive de l'équipe adverse j de l'affrontement h lors de l'état opposé l ;
- un terme d'offset qui représente le temps en minute que l'équipe i est restée dans l'état k lors de l'affrontement h .

Par ailleurs, des estimations préliminaires ont montré que l'effet marginal de jouer à domicile n'était pas significatif, donc cet effet sera volontairement omis du modèle.

$$\ln(\lambda_{h,i,k}) = \beta_1 \text{Equipe}_{h,i,k} + \beta_2 \text{MoyBC}_{h,j,l} + \ln(\text{Temps}_{h,i,k}) \tag{3.32}$$

Dorénavant, cette modélisation nécessite la connaissance des gardiens de but partant qu'ont les équipes i et j lors de l'affrontement h afin de déterminer les valeurs des $\text{MoyBC}_{h,j,l}$. En considérant l'ensemble S , comme les affrontements auxquels le gardien désiré a pris part pour

une équipe i dans la base de données des affrontements intra-associations, et l'ensemble H , comme tous les affrontements de l'équipe j , le nombre de buts accordés moyens par minute pour un gardien de but avec l'équipe j s'écrit comme en (3.33). Dans l'éventualité où un gardien ne dépasse pas un total de quarante minutes jouées lors d'un état l , la moyenne de buts accordés par la défensive en sa présence est considérée comme invalide. Évidemment, il serait absurde de prétendre qu'une défensive n'accordant aucun but en dix minutes de jeu mérite une moyenne inférieure à l'une accordant un but en cent minutes de jeu. Ainsi, elle est estimée par la moyenne de buts accordés de leur équipe moyennant une pénalité arbitraire de 0,02. Par exemple, le tableau 3.7 montre les moyennes estimées par minute des différents gardiens des Canadiens de Montréal pour la saison 2015-2016. Finalement, le tableau 3.8 donne les estimations de quelques coefficients β du modèle (3.32) suite à ces estimations.

$$\text{MoyBC}_{h,j,l} = \begin{cases} \frac{\sum_{h \in S} Y_{h,j,l}}{\sum_{h \in S} \text{Temps}_{h,j,l}}, & \text{si } \sum_{h \in S} \text{Temps}_{h,j,l} > 40 \\ \frac{\sum_{h \in H} Y_{h,j,l}}{\sum_{h \in H} \text{Temps}_{h,j,l}} + 0,02, & \text{si } \sum_{h \in S} \text{Temps}_{h,j,l} \leq 40 \end{cases} \quad (3.33)$$

TABLE 3.7 – Moyenne de buts accordés estimés par minute des différents gardiens des Canadiens de Montréal pour la saison 2015-2016.

Gardien de but	Buts accordés			Temps joués			Moyenne estimée		
	L	E	T	L	E	T	L	E	T
Carey Price	13	2	0	326,7	179,8	13,9	0,0398	0,0111	0,0427
Mike Condon	15	13	4	383,5	393,9	202,4	0,0391	0,0330	0,0198
Dustin Tokarski	2	2	0	65,5	46,4	8,1	0,0305	0,0431	0,0427
Ben Scrivens	1	3	3	3,8	31,7	84,5	0,0598	0,0507	0,0355
Total de l'équipe	31	20	7	779,5	651,7	308,8	0,0398	0,0307	0,0227

TABLE 3.8 – Extrait des estimations des paramètres du modèle de régression de Poisson pour l'association de l'Est au cours de la saison 2015-2016.

Modèle	Variable explicative	Estimation de β	Erreur type	Valeur p
L	MTL	-3,9811	0,2804	$< 2 \times 10^{-16}$
	TOR	-4,5088	0,4745	$< 2 \times 10^{-16}$
	MoyBC _T	20,6280	4,8007	$1,73 \times 10^{-5}$
E	MTL	-3,9205	0,2487	$< 2 \times 10^{-16}$
	TOR	-4,8062	0,3484	$< 2 \times 10^{-16}$
	MoyBC _E	25,6117	4,5095	$1,35 \times 10^{-8}$
T	MTL	-3,8072	0,2978	$< 2 \times 10^{-16}$
	TOR	-3,9166	0,2597	$< 2 \times 10^{-16}$
	MoyBC _L	21,6899	4,0145	$6,56 \times 10^{-8}$

En fait, la pénalité sous-estime certaines performances défensives lorsqu'un gardien performant ne se retrouve pas suffisamment dans une situation précise, comme pour Carey Price lorsque son équipe tire de l'arrière. Par contre, elle permet principalement de maintenir une valeur cohérente pour la défense de l'équipe en présence de certains gardiens, comme Ben Scrivens. En d'autres mots, elle permet de pénaliser tous les gardiens de troisième et quatrième plan n'ayant pas pris part à suffisamment d'affrontements pour leur octroyer une moyenne de buts accordés par match réaliste par rapport à leur calibre. Contrairement à la réalité, aucun changement ne s'effectuera lors des simulations d'une rencontre, peu importe la performance obtenue, ce qui rend la correction d'autant plus importante.

3.2.9 Fonctionnement du troisième modèle de simulation

En reprenant l'affrontement du 27 février 2016 ($h = 2116$), où les Canadiens accueillent les Maple Leafs, les gardiens de but partant étaient Mike Condon pour Montréal et Jonathan Bernier pour Toronto. Ainsi, la moyenne de buts accordés estimés par minute pour la défense des Canadiens en présence de Mike Condon est de 0,0391, 0,0330 et 0,0198 lorsque l'équipe est respectivement en avance, à égalité et en retard. D'un autre côté, les Maple Leafs avec Jonathan Bernier en accordent 0,0546, 0,0643 et 0,0550. La régression de Poisson nous assure que les intensités $\lambda_{h,i,k}$ s'obtiennent par l'équation (3.34). Comme on cherche des intensités par minute, la variable Temps $_{h,i,k}$ vaudra toujours 1. Par conséquent, en utilisant les estimations des β du tableau 3.8, on estime que l'intensité des Canadiens lorsque le pointage est égal pour ce match, soit $\lambda_{2116,Mtl,E}$, est de $e^{-3,9205} \times e^{25,6117 \times 0,0643} = 0,1029$. De la même manière, il est possible de déduire toutes les autres intensités $\lambda_{(h,i,k)}$ et $\lambda_{(h,j,l)}$ présentent dans le tableau 3.9. Rappelons que les équations (3.20) et (3.21) permettent de déduire les $\Lambda_{(h,k)}$ et la probabilité que chacune des équipes soit responsable d'un événement survenu.

$$\lambda_{h,i,k} = e^{\beta_1 \text{Equipe}_{h,i,k}} \times e^{\beta_2 \text{MoyBC}_{h,j,l}} \times \text{Temps}_{h,i,k} \quad (3.34)$$

TABLE 3.9 – Intensités estimées par le troisième modèle pour un affrontement entre les Maple Leafs avec Jonathan Bernier et les Canadiens avec Mike Condon, pour la saison 2015-2016.

Intensité (Indice)	L_i	E_i	T_i
λ de l'équipe (2116, i,k)	0,0580	0,1029	0,0723
λ de l'adversaire (2116, j,l)	0,0465	0,0190	0,0166
Λ du match (2116, k)	0,1045	0,1220	0,0892
\mathbb{P} de marquer (2116, i,k)	0,5553	0,8439	0,8142

Cette fois, grâce aux intensités construites par le troisième modèle, répéter 10000 fois la simulation de cet affrontement prédit que les Canadiens possèdent une probabilité de victoire d'approximativement 85,35 % sachant que les gardiens partant pour le match sont Mike Condon

pour Montréal et Jonathan Bernier pour Toronto, lors de la deuxième moitié du calendrier de 2015-2016. Au final, la fréquence relative de prolongation est de 16,81 %, et le pourcentage de jeu blanc est passé à 7,26 %.

3.2.10 Estimation des paramètres du quatrième modèle de simulation

Pour sa part, le quatrième modèle de simulation estime les intensités à l'aide de deux modèles linéaires généralisés distincts, l'un décrivant les intensités lors des 55 premières minutes d'un affrontement et l'autre les décrivant pour les 5 dernières. La section 2.1.2 affirme que les deux modèles ont raison de coexister puisque la fréquence des buts s'accroît lors des 5 dernières minutes d'un affrontement. Par contre, ils doivent être fondamentalement différents puisqu'un bon nombre d'équipes ne se sont pas retrouvées suffisamment longtemps dans certains états pendant les dernières minutes de leurs matchs pour que les estimations de leur paramètre soient représentatives. En fait, le premier modèle, dédié aux 55 premières minutes d'un affrontement, donné par l'équation (3.35) est presque identique au modèle précédent, sauf qu'il empêche de biaiser les estimations en comptabilisant les buts dans des filets déserts tout en éliminant les ajustements stratégiques plus drastiques propres aux fins de parties. De son côté, le deuxième modèle, dédié aux 5 dernières minutes d'un affrontement, donnée par l'équation (3.36) suppose qu'aucune équipe ne possède son propre talent pour marquer lors de cette période, seules les défensives ont une influence négatives sur les buts qui se marque. Ainsi, on limite les problèmes d'estimation des paramètres du modèle causés par un manque de minutes jouées pour une équipe précise lors d'un état précis.

$$\lambda_{h,i,k}^{55} = e^{\beta_1 \text{Equipe}_{h,i,k}} \times e^{\beta_2 \text{MoyBC}_{h,j,l} \mathbb{1}_{\{\text{temps} \leq 55\}}} \times \text{Temps}_{h,i,k} \mathbb{1}_{\{\text{temps} \leq 55\}} \quad (3.35)$$

$$\lambda_{h,i,k}^5 = e^{\beta_0} \times e^{\beta_3 \text{MoyBC}_{h,j,l} \mathbb{1}_{\{\text{temps} > 55\}}} \times \text{Temps}_{h,i,k} \mathbb{1}_{\{\text{temps} > 55\}} \quad (3.36)$$

Désormais, le nombre de buts accordés moyens par minute en comptabilisant uniquement les 55 premières minutes des affrontements pour un gardien de but avec l'équipe j , soit $\text{MoyBC}_{h,j,l}^{55}$, s'écrit comme en (3.37).⁴ Le tableau 3.10 montre ces estimations pour les différents gardiens des Canadiens de Montréal pour la saison 2015-2016.

$$\text{MoyBC}_{h,j,l}^{55} = \begin{cases} \frac{\sum_{h \in S} Y_{h,j,l}^{55}}{\sum_{h \in S} \text{Temps}_{h,j,l}^{55}}, & \text{si } \sum_{h \in S} \text{Temps}_{h,j,l}^{55} > 40 \\ \frac{\sum_{h \in H} Y_{h,j,l}^{55}}{\sum_{h \in H} \text{Temps}_{h,j,l}^{55}} + 0,02, & \text{si } \sum_{h \in S} \text{Temps}_{h,j,l}^{55} \leq 40 \end{cases} \quad (3.37)$$

4. Afin de condenser la formulation, le surindice 55 sera souvent utilisé pour remplacer l'indicatrice $\mathbb{1}_{\{\text{temps} \leq 55\}}$, tout comme le surindice 5 remplacera l'indicatrice $\mathbb{1}_{\{\text{temps} > 55\}}$, valant 1 lorsque la condition est respectée et 0 sinon.

TABLE 3.10 – Moyenne de buts accordés estimés par minute des différents gardiens des Canadiens de Montréal lors des 55 premières minutes des matchs pour la saison 2015-2016.

Gardien de but	Buts accordés			Temps joués			Moyenne estimée		
	L	E	T	L	E	T	L	E	T
Carey Price	12	2	0	288,2	178,2	13,9	0,0416	0,0111	0,0456
Mike Condon	14	13	4	345,6	366,8	182,4	0,0405	0,0354	0,0219
Dustin Tokarski	2	2	0	60,5	46,4	3,1	0,0331	0,0431	0,0456
Ben Scrivens	1	3	3	3,8	31,7	74,5	0,0615	0,0521	0,0403
Total de l'équipe	29	20	7	698,1	623,1	273,8	0,0415	0,0321	0,0256

Néanmoins, l'estimation de $\text{MoyBC}_{h,j,l}^5$, est légèrement différente. Cette fois, la moyenne de buts accordés par la défense en présence d'un gardien est considérée comme invalide seulement si elle ne dépasse pas un total de vingt minutes jouées lors d'un état l . L'équation (3.38) montre que la moyenne estimée pour ces défenses impliquera toujours la pénalité arbitraire de 0,02 si jamais le temps nécessaire n'est pas atteint. De plus, un minimum réaliste pour chacun des états sera imposé aux estimations afin d'éviter qu'une estimation soit de 0,000. Bref, il est raisonnable de croire que les minimums réalistes pour les estimations sont respectivement de 0,010 pour l'état L, de 0,025 pour l'état E et de 0,045 pour l'état T. Ainsi, on retrouve les estimations du tableau 3.11 pour les gardiens des Canadiens de Montréal pour la saison 2015-2016. Finalement, le tableau 3.12 donne les estimations de quelques coefficients β du modèle (3.35) et (3.36) suites à ces estimations.

$$\text{MoyBC}_{h,j,l}^5 = \begin{cases} \frac{\sum_{h \in S} Y_{h,j,l}^5}{\sum_{h \in S} \text{Temps}_{h,j,l}^5}, & \text{si } \sum_{h \in S} \text{Temps}_{h,j,l}^5 > 20 \\ \frac{\sum_{h \in H} Y_{h,j,l}^5}{\sum_{h \in H} \text{Temps}_{h,j,l}^5} + 0,02, & \text{si } \sum_{h \in S} \text{Temps}_{h,j,l}^5 \leq 20 \end{cases} \quad (3.38)$$

TABLE 3.11 – Moyenne de buts accordés estimés par minute des différents gardiens des Canadiens de Montréal lors des 5 dernières minutes des matchs pour la saison 2015-2016.

Gardien de but	Buts accordés			Temps joués			Moyenne estimée		
	L	E	T	L	E	T	L	E	T
Carey Price	1	0	0	38,5	1,5	0,0	0,0260	0,0450	0,0650
Mike Condon	1	0	0	37,9	27,1	20,0	0,0264	0,0250	0,0650
Dustin Tokarski	0	0	0	5,0	0,0	5,0	0,0446	0,0450	0,0650
Ben Scrivens	0	0	0	0,0	0,0	10,0	0,0446	0,0450	0,0650
Total de l'équipe	2	0	0	81,4	28,6	35,0	0,0246	0,0250	0,0450

TABLE 3.12 – Extrait des estimations des paramètres du quatrième modèle de simulation pour l’association de l’Est au cours de la saison 2015-2016.

Modèle	Variable explicative	Estimation des β	Erreur type	Valeur p
L_{55}	MTL	-4,3662	0,3051	$< 2 \times 10^{-16}$
	TOR	-5,2097	0,6225	$< 2 \times 10^{-16}$
	$MoyBC_T$	27,5367	5,7577	$1,73 \times 10^{-6}$
E_{55}	MTL	-4,0117	0,2549	$< 2 \times 10^{-16}$
	TOR	-4,7814	0,3502	$< 2 \times 10^{-16}$
	$MoyBC_E$	26,6972	4,663	$2,27 \times 10^{-9}$
T_{55}	MTL	-3,6700	0,3032	$< 2 \times 10^{-16}$
	TOR	-3,9591	0,2780	$< 2 \times 10^{-16}$
	$MoyBC_L$	21,4336	4,3863	$1,03 \times 10^{-6}$
L_5	<i>Origine</i>	-3,5356	0,3452	$< 2 \times 10^{-16}$
	$MoyBC_T$	10,1586	2,5066	$5,06 \times 10^{-5}$
E_5	<i>Origine</i>	-4,8408	0,5674	$< 2 \times 10^{-16}$
	$MoyBC_T$	21,1536	5,8048	$2,68 \times 10^{-4}$
T_5	<i>Origine</i>	-3,8507	0,3033	$< 2 \times 10^{-16}$
	$MoyBC_T$	12,8600	2,7935	$4,15 \times 10^{-6}$

3.2.11 Fonctionnement du quatrième modèle de simulation

Pour une dernière fois, l’affrontement du 27 février 2016 opposant les Canadiens de Montréal aux Maple Leafs de Toronto, avec Mike Condon comme gardien pour Montréal et Jonathan Bernier pour Toronto, sera analysé en détail. La moyenne de buts accordés estimés par minute pour la défensive des Canadiens en présence de Mike Condon lors des 55 premières minutes d’un affrontement est de 0,0405, 0,0354 et 0,0219 lorsque l’équipe est respectivement en avance, à égalité et en retard. Lors des 5 dernières minutes, elles deviennent 0,0264, 0,0250 et 0,0650. De son côté, les Maple Leafs avec Jonathan Bernier en accordent 0,0489, 0,0651 et 0,0540 lors des 55 premières minutes et 0,1490, 0,0723 et 0,0647 lors des 5 dernières.

Les différents coefficients des modèles sont présentés au tableau 3.12. Ils permettent de déduire que lors des 55 premières minutes, les Canadiens possèdent une intensité lorsqu’ils sont en avance de $\lambda_{MTL,2116,L}^{55} = e^{-4,3662} \times e^{27,5367 \times 0,0540} = 0,0562$. Par hasard, cette intensité reste identique lors des 5 dernières minutes d’un affrontement, où $\lambda_{MTL,2116,L}^5 = e^{-3,5356} \times e^{10,1586 \times 0,0647} = 0,0562$. De la même manière, on peut construire les intensités pour toutes les situations de l’affrontement répertoriées dans le tableau 3.13. Pour préciser, les périodes de surtemps, pour la quatrième modélisation, sont simulées à partir du modèle associé aux 55 premières minutes. Évidemment, les équations (3.20) et (3.21) permettent de déterminer l’intensité globale de l’affrontement selon l’état du pointage et qui entre les Canadiens ou les Maple Leafs seront responsables d’un événement se réalisant dans le temps imparti.

TABLE 3.13 – Intensités estimées par le quatrième modèle pour un affrontement entre les Maple Leafs avec Jonathan Bernier et les Canadiens avec Mike Condon, pour la saison 2015-2016.

Intensité (Indice)	55 premières minutes			5 dernières minutes		
	L_i	E_i	T_i	L_i	E_i	T_i
λ de l'équipe (2116, i,k)	0,0562	0,1029	0,0727	0,0562	0,0365	0,1445
λ de l'adversaire (2116, j,l)	0,0455	0,0216	0,0100	0,0299	0,0134	0,0564
Λ du match (2116, k)	0,1016	0,1245	0,0827	0,0861	0,0499	0,2009
\mathbb{P} de marquer (2116, i,k)	0,5528	0,8267	0,8792	0,6531	0,7312	0,7192

Bien entendu le quatrième modèle ajoute du réalisme quant à l'évolution du pointage. Par contre, en se concentrant seulement sur un affrontement précis, ce modèle ne semble pas prédire différemment du troisième modèle. En effet, en 10000 simulations, un affrontement entre Mike Condon, pour les Canadiens de Montréal, et Jonathan Bernier, pour les Maple Leafs de Toronto, prédit une chance de victoire de l'équipe montréalaise de 86,50 %, lors de la deuxième moitié du calendrier de 2015-2016. Au total, 18,58 % des affrontements ont nécessité une prolongation et 6,84 % des pointages obtenus correspondent à un jeu blanc.

3.3 Bilan des quatre modèles de simulation

Au final, quelques statistiques de l'affrontement du 27 février 2016 entre les Canadiens de Montréal et les Maple Leafs de Toronto sont résumées dans le tableau 3.14. N'étant qu'un très petit aperçu, la prochaine section explicitera les performances globales des modèles par rapport à la réalité.

TABLE 3.14 – Approximation de diverses statistiques, par modèle pour l'affrontement du 27 février 2016 entre les Canadiens de Montréal et les Maple Leafs de Toronto.

Modèle	Victoire Montréal	Surtemps	Jeux blancs
1	87,44 %	13,32 %	10,13 %
2	68,62 %	26,07 %	9,06 %
3	85,35 %	16,81 %	7,26 %
4	86,50 %	18,58 %	6,84 %

Chapitre 4

Simulations et discussion

Après avoir défini la structure et le fonctionnement de quatre modèles de simulation différents, il est temps de tester leurs capacités. Deux critères permettront de les comparer, et ainsi de couronner le meilleur modèle. Tout d’abord, la capacité prédictive des modèles sera mesurée, ce qui permettra de les placer par rapport à la littérature. Elle sera dévoilée sous plusieurs agrégations, notamment par équipe et par gardien de but partant. Puis, le meilleur modèle sera déterminé selon ses gains monétaires potentiels via les paris sportifs. Au total, on simulera 12 blocs d’affrontement, nommés les « divisions-saisons », correspondant à toutes les combinaisons entre une division (Atlantique, Métropolitaine, Centrale et Pacifique) et d’une deuxième moitié d’une saison (2013-2014, 2014-2015 et 2015-2016), et ce à partir de chacun des 4 modèles. Finalement, de nombreuses pistes d’amélioration seront suggérées telles que l’exclusion de plusieurs affrontements visiblement problématiques.

4.1 Estimations des paramètres

Un calendrier régulier de la Ligue nationale de hockey contient 1230 matchs, dont 432 sont des affrontements entre deux équipes de l’association de l’Est et 350 sont entre deux équipes de l’association de l’Ouest. Par conséquent, la mi-saison d’un calendrier se trouve aux alentours du jour correspondant au match 615. Toutefois, le tableau 4.1 montre que la dispersion des affrontements intra-associations n’est pas particulièrement homogène à travers les trois calendriers considérés. La différence entre les nombres d’affrontements intra-association des deux associations provient du déséquilibre dans le nombre d’équipes qui les composent. Les 14 équipes de l’association de l’Ouest affrontent toujours 50 fois, par saison, une équipe de leur association, tandis que les 16 équipes de l’association de l’Est en affrontent 54.¹ Les premières moitiés de calendrier serviront à calibrer les différents modèles. Cependant, en raison de leur exhaustivité, la plupart des estimations de paramètres furent regroupées dans l’annexe B.

1. Chaque affrontement interassociation possible apparaît deux fois dans le calendrier. C’est pourquoi le nombre d’affrontements de chacune des équipes est de 82.

TABLE 4.1 – Nombre d’affrontements dans les blocs de calibrage des modèles.

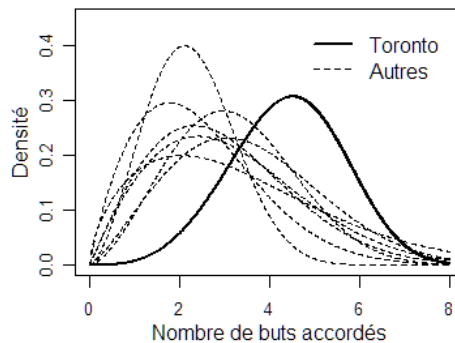
Date début de la saison	Date fin de la saison	Matches Totaux	Associa. Ouest	Associa. Est
1 oct. 2013	1 jan. 2014	611	171	212
8 oct. 2014	9 jan. 2015	613	177	216
7 oct. 2015	8 jan. 2016	613	156	194

D’abord, le tableau 4.2 affiche uniquement les valeurs estimées des paramètres α , paramètre de forme, et λ , paramètre d’échelle, pour les Weibull (α, λ) à l’intérieur du premier modèle pour la division Atlantique lors de l’année 2015-2016. Grossièrement, une valeur de λ relativement élevée implique une tendance à marquer ou accorder un plus grand nombre de buts, tandis qu’une valeur de α relativement élevée accentue la forme de cloche de la distribution. Par exemple, la figure 4.1 montre que les Maple Leafs sont plus à risque d’accorder un grand nombre de buts que tous leurs rivaux de division.

TABLE 4.2 – Extrait des paramètres estimés du premier modèle, pour la première moitié de la saison 2015-2016.

Équipe	Offensive 2015-2016				Défensive 2015-2016			
	α_{Etr}	λ_{Etr}	α_{Dom}	λ_{Dom}	α_{Etr}	λ_{Etr}	α_{Dom}	λ_{Dom}
BOS	2,700	4,038	2,556	3,943	2,562	3,020	1,591	3,801
BUF	2,778	3,939	1,881	2,658	2,481	3,654	2,224	3,969
DET	2,372	3,517	2,383	3,061	3,238	3,425	1,964	3,333
FLA	2,193	3,856	2,286	3,719	1,937	2,572	1,819	2,735
MTL	2,710	4,362	3,044	3,758	3,280	3,217	2,540	2,558
OTT	1,730	3,270	2,694	3,265	2,278	3,891	1,908	3,541
TBL	1,966	3,834	2,522	3,613	2,429	3,451	2,522	3,613
TOR	1,655	2,892	2,238	3,186	2,932	3,559	3,916	4,862

FIGURE 4.1 – Lois Weibull estimées pour les buts accordés à domicile des équipes de la division Atlantique, pour la saison 2015-2016.



Ensuite, le tableau 4.3 affiche toutes les intensités ρ , offensives, et les intensités ψ , défensives, estimées pour la division Atlantique à partir du deuxième modèle pour l'année 2015-2016. Pour résumer, toutes les intensités correspondent à un nombre de buts moyen, soit marqué ou accordé, par minute. Les indices (L, E et T) indiquent respectivement si l'intensité fut compilée lorsque l'équipe mène, est à égalité, ou tire de l'arrière dans le pointage. Par la suite, la somme des deux intensités confrontées permet au processus de Poisson de générer le résultat d'un affrontement spécifique.

TABLE 4.3 – Extrait des paramètres estimés du deuxième modèle, pour la première moitié de la saison 2015-2016.

Lieu	Équipe	ρ_L	ρ_E	ρ_T	ψ_L	ψ_E	ψ_T
Domicile	BOS	0,0427	0,0648	0,0373	0,0552	0,0414	0,0504
	BUF	0,0269	0,0650	0,0542	0,0505	0,0220	0,0280
	DET	0,0304	0,0505	0,0450	0,0278	0,0330	0,0504
	FLA	0,0617	0,0397	0,0328	0,0430	0,0369	0,0670
	MTL	0,0478	0,0660	0,0509	0,0502	0,0402	0,0623
	OTT	0,0536	0,0275	0,0389	0,0629	0,0199	0,0448
	TBL	0,0526	0,0374	0,0576	0,0402	0,0450	0,0449
	TOR	0,0308	0,0291	0,0363	0,0256	0,0092	0,0600
Étranger	BOS	0,0396	0,0389	0,0249	0,0478	0,0380	0,0596
	BUF	0,0493	0,0418	0,0406	0,0433	0,0585	0,0467
	DET	0,0541	0,0168	0,0600	0,0417	0,0254	0,0550
	FLA	0,0347	0,0253	0,0246	0,0303	0,0246	0,0419
	MTL	0,0434	0,0363	0,0305	0,0345	0,0258	0,0089
	OTT	0,0482	0,0458	0,0466	0,0270	0,0477	0,0392
	TBL	0,0526	0,0306	0,0512	0,0491	0,0554	0,0224
	TOR	0,0554	0,0323	0,0444	0,0769	0,0735	0,0424

Puis, le tableau 4.4 montre seulement les différents coefficients β associés aux équipes de la division Atlantique pour l'année 2015-2016, estimés maximisant la vraisemblance du troisième modèle, même si les coefficients sont estimés par association. Le tableau 4.5 montre ceux correspondant au quatrième modèle de simulation. Heureusement, les coefficients, pour les deux modèles, s'avèrent tous significativement différents de zéro à un seuil d'erreur de 0,05 %. Ces deux modélisations utilisent des modèles de régression de Poisson pour déterminer le niveau d'intensité offensive de chacune des équipes par rapport à la défensive adverse, pour chacun des trois états. Ainsi, l'évolution du pointage d'un affrontement peut être encore étudiée par les processus de Poisson. Pour sa part, le quatrième modèle scinde une rencontre en deux blocs, l'un simulant les 55 premières minutes d'un affrontement et l'autre, les 5 dernières. Une hypothèse supplémentaire est faite pour le deuxième bloc, qui considère que toutes les équipes possèdent la même intensité de base, seule la défensive adverse vient modifier leurs intensités globales. Les indices (55 et 5) dénotent auxquels des deux modèles, le coefficient estimé appartient. Bref, les paramètres du troisième et quatrième modèle indiquent que la

plupart des équipes de la division Atlantique augmentent leur intensité offensive lorsqu'elles tirent de l'arrière, toutes choses étant égales par ailleurs.

TABLE 4.4 – Extrait des paramètres estimés du troisième modèle, pour la première moitié de la saison 2015-2016.

Variable	2013-2014			2014-2015			2015-2016		
	L	E	T	L	E	T	L	E	T
BOS	-4,136	-3,739	-4,083	-4,516	-4,324	-4,061	-3,906	-3,953	-3,982
BUF	-4,392	-4,806	-4,334	-4,472	-4,543	-4,369	-4,252	-4,080	-4,302
DET	-4,312	-3,797	-4,428	-4,178	-4,722	-3,634	-4,365	-4,249	-4,063
FLA	-4,448	-3,978	-4,268	-5,188	-4,453	-3,842	-3,916	-4,249	-4,128
MTL	-4,349	-4,088	-4,086	-4,004	-4,296	-4,119	-3,981	-3,921	-3,807
OTT	-3,977	-3,863	-3,834	-5,051	-4,159	-3,936	-3,875	-4,646	-4,128
TBL	-4,011	-4,238	-4,335	-4,086	-4,076	-3,808	-4,013	-4,251	-3,921
TOR	-4,198	-4,054	-4,063	-3,709	-4,266	-3,812	-4,509	-4,806	-3,917
<i>MoyBC</i>	23,241	19,616	20,909	23,527	25,012	18,823	20,628	25,612	21,690

TABLE 4.5 – Extrait des paramètres estimés du quatrième modèle, pour la première moitié de la saison 2015-2016.

Variable	2013-2014			2014-2015			2015-2016		
	L	E	T	L	E	T	L	E	T
BOS ₅₅	-4,389	-3,756	-4,159	-4,681	-4,302	-4,015	-4,291	-4,019	-4,044
BUF ₅₅	-5,558	-4,775	-4,380	-4,587	-4,537	-4,304	-4,635	-4,113	-4,268
DET ₅₅	-4,582	-3,731	-4,514	-4,465	-4,659	-3,618	-4,525	-4,263	-3,832
FLA ₅₅	-4,804	-3,956	-4,254	-5,674	-4,442	-3,849	-4,201	-4,274	-4,218
MTL ₅₅	-4,502	-4,066	-4,443	-4,337	-4,316	-4,096	-4,366	-4,012	-3,670
OTT ₅₅	-4,218	-3,921	-3,760	-5,616	-4,126	-3,979	-4,344	-4,673	-4,207
TBL ₅₅	-4,163	-4,238	-4,420	-4,345	-4,092	-3,904	-4,277	-4,336	-3,802
TOR ₅₅	-4,542	-3,990	-4,083	-3,911	-4,241	-3,756	-5,210	-4,781	-3,959
<i>MoyBC</i> ₅₅	27,68	19,614	21,133	28,875	24,586	18,44	27,537	26,697	21,434
<i>Origine</i> ₅	-3,388	-4,472	-3,993	-3,493	-5,785	-3,862	-3,536	-4,841	-3,851
<i>MoyBC</i> ₅	9,600	17,195	14,898	10,097	35,050	13,437	10,159	21,154	12,860

4.2 Simulations des modèles par la méthode de Monte-Carlo

La base de données des affrontements intradivisions permet de connaître les véritables résultats des deuxièmes moitiés de calendrier de la LNH sans problème. Ainsi, en appliquant un modèle de simulation à tous ces affrontements, il est possible de comparer l'issue de chaque affrontement avec l'issue générée par les simulations. Le tableau 4.6 indique le nombre d'affrontements dans chacun des blocs de simulation. Évidemment, le déséquilibre dans le nombre d'équipes entre les deux associations de la LNH vient encore déséquilibrer le nombre d'affron-

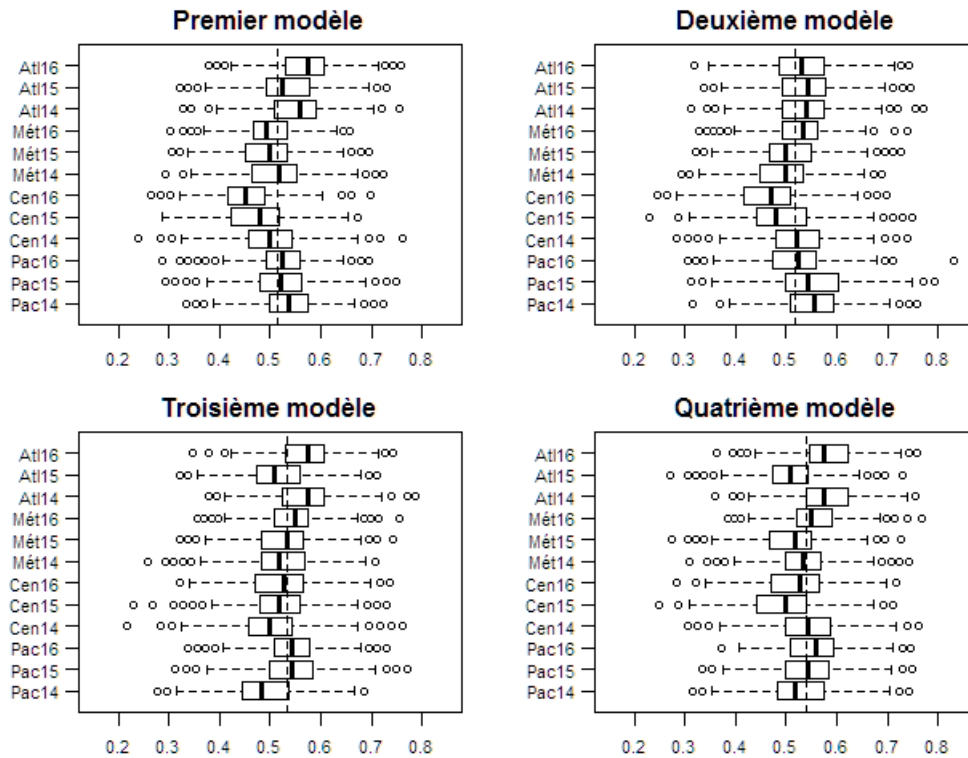
tements associés à chacune des divisions. Au total, le calendrier à simuler, pour 2013-2014, compte 220 affrontements, celui de 2014-2015 en compte 221, et le suivant en compte 251.

TABLE 4.6 – Nombre d’affrontements dans les blocs de simulation.

Date début de la saison	Date fin de la saison	Matches Totaux	Division Pacifique	Division Centrale	Division Métropo.	Division Atlanti.
2 jan. 2014	13 avr. 2014	220	54	47	58	61
10 jan. 2015	12 avr. 2015	221	48	52	62	59
8 jan. 2016	10 avl. 2016	251	59	53	73	66

Si un match simulé possède la même équipe gagnante que le réel affrontement, alors l’issue de ce match est dite bien prédite. Pour s’assurer de la qualité des prédictions, 2000 tirages par division-saison ont été exécutés. Leurs résultats sont regroupés dans les boîtes à moustaches de la figure 4.2 pour chacune des divisions-saisons. Les traitillés verticaux de la figure représentent la moyenne globale de bonne prédiction du modèle. Dans l’ordre, les quatre modèles respectifs parviennent à prédire l’équipe ayant réellement gagné chacun des affrontements, à 51,53 %, 51,85 %, 53,30 % et 53,84 % du temps, au cours des trois saisons simulées.

FIGURE 4.2 – Distribution des pourcentages de bonnes prédictions par division-saison.



Étonnamment, les distributions pour chacune des divisions-saisons associées au premier et au deuxième modèles sont très similaires malgré que l’un considère les ajustements stratégiques

et l'autre non. Ainsi, les deux premiers modèles ne possèdent pas du tout les mêmes forces et faiblesses. À priori, cette ressemblance pourrait n'être causée que par la quasi-impossibilité à prédire l'issue de certains affrontements, ce qui induit en erreur n'importe quel modèle. D'un autre côté, le troisième et le quatrième modèles affichent des résultats, cette fois sans surprise, très proche. En désagréant les résultats par saison, comme au tableau 4.7, il devient clair que la dernière saison simulée, soit la saison 2015-2016, fut la mieux prédite. Néanmoins, s'attarder seulement à cet indicateur, le pourcentage de bonnes prédictions, peut être trompeur puisqu'il n'indique aucune information quant à la qualité des prédictions.

TABLE 4.7 – Pourcentage de bonnes prédictions, par deuxième moitié d'année et par modèle.

Modèle	2013-2014	2014-2015	2015-2016	Total
1	52,75 %	50,84 %	51,08 %	51,53 %
2	52,50 %	51,89 %	51,24 %	51,85 %
3	52,57 %	52,49 %	54,66 %	53,30 %
4	54,50 %	51,54 %	55,29 %	53,84 %

Les autres informations du tableau 4.8 telles que le pourcentage d'affrontements simulés ayant nécessité une prolongation ou occasionnant un jeu blanc permettent de distinguer le réalisme des modèles. La base de données complète recensait un pourcentage de surtemps de près de 25 % et un pourcentage de jeux blancs légèrement inférieur à 7 %. Cependant, ces valeurs diffèrent lorsque l'on considère que les affrontements intradivisions de la deuxième moitié du calendrier. Le pourcentage de surtemps réel tournent autour de 22 %, tandis que le pourcentage de jeux blancs reste aux alentours de 7 %. Décidément, ces deux statistiques placent le troisième et le quatrième modèles en tête, quoiqu'ils ont une tendance à surestimer le nombre de jeux blancs. Malgré tout, le quatrième modèle exagère le nombre d'affrontements nécessitant une prolongation pour déterminer l'équipe gagnante, sauf qu'il s'agit d'un problème plus en lien avec le nombre de buts marqué par affrontement qu'un problème nuisible à la détermination de l'équipe gagnante.

TABLE 4.8 – Pourcentage de prolongation et de jeux blancs, par deuxième moitié d'année et par modèle.

Modèle	2013-2014		2014-2015		2015-2016	
	Surtemps	Jeux blancs	Surtemps	Jeux blancs	Surtemps	Jeux blancs
1	17,0 %	6,5 %	17,3 %	5,9 %	16,7 %	5,9 %
2	21,6 %	7,0 %	22,0 %	6,8 %	22,3 %	7,9 %
3	20,7 %	8,1 %	21,7 %	7,6 %	23,0 %	8,2 %
4	23,3 %	7,8 %	25,8 %	7,1 %	24,9 %	8,2 %
Réalité	22,7 %	7,7 %	22,2 %	6,6 %	21,5 %	6,8 %

4.2.1 Résultats par rapport aux équipes

En se penchant sur les pourcentages de bonnes prédictions des équipes, de nouvelles sources d'erreurs pour chacun des modèles peuvent apparaître. Logiquement, le modèle prédira bien une équipe qui répète ses performances de la première moitié de saison lors de la deuxième moitié. Généralement, cette situation arrive plus fréquemment aux équipes d'un calibre très faible, n'ayant pas nécessairement de moyens rapides de s'améliorer, contrairement aux équipes dominantes pouvant devenir affaiblies à la suite de blessures. Cette situation est flagrante à l'intérieur du tableau 4.9, où sont regroupées les équipes les mieux et les moins bien prédites au cours de la saison de 2015-2016. Sans aucun doute, l'équipe la mieux prédite pour l'année 2015-2016 est celle de Toronto. Réellement, les Maple Leafs de Toronto possédaient une fiche de 5-10-6² contre leur association à la mi-saison, puis ils ont obtenu une fiche de 5-14-2 contre leur division, lors de la deuxième moitié, ce qui les a ajustés correctement. Par contre, d'autres équipes, comme les Jets de Winnipeg, se sont retrouvées constamment mal ajustées lors de cette saison. Pour leur part, ils venaient d'enregistrer une fiche de 8-15-1 contre l'association de l'Ouest avant la mi-saison, pour terminer avec 7-6-1 contre la division Centrale lors de leur deuxième moitié de saison.

TABLE 4.9 – Moyenne du pourcentage de bonnes prédictions de l'issue des matchs des équipes pour la deuxième moitié de la saison 2015-2016.

Rang	Modèle 1	Modèle 2	Modèle 3	Modèle 4
1	TOR (64,9 %)	TOR (58,3 %)	TOR (66,7 %)	TOR (66,1 %)
2	EDM (59,8 %)	EDM (56,7 %)	BUF (63,3 %)	EDM (64,0 %)
3	TBL (58,4 %)	CBJ (56,5 %)	CGY (61,6 %)	BUF (63,2 %)
4	MTL (57,9 %)	WSH (55,1 %)	EDM (60,7 %)	CBJ (62,0 %)
5	VAN (57,6 %)	FLA (54,1 %)	ANA (60,7 %)	CGY (61,9 %)
⋮	⋮	⋮	⋮	⋮
26	ARI (45,5 %)	STL (47,6 %)	ARI (49,8 %)	STL (50,8 %)
27	MIN (43,9 %)	MIN (46,9 %)	MIN (48,3 %)	WPG (49,2 %)
28	COL (41,1 %)	DAL (46,1 %)	WPG (47,2 %)	SJS (47,8 %)
29	WSH (40,6 %)	COL (44,8 %)	SJS (47,2 %)	MIN (45,3 %)
30	WPG (38,8 %)	WPG (40,7 %)	VAN (42,2 %)	VAN (43,3 %)

En découpant les résultats par rapport à l'endroit où furent joués les matchs, soit à domicile ou à l'étranger, comme au tableau 4.10, on peut remarquer que la faiblesse des prédictions impliquant les Jets provient de leurs matchs à l'étranger. Réellement, leur fiche parfaite de 4-0-0 à l'étranger contre leur division après la mi-saison est venue déjouer les pronostics.

2. Les fiches représentent, dans l'ordre, le nombre de victoires, de défaites, et de défaites en surtemps.

TABLE 4.10 – Moyenne du pourcentage de bonnes prédictions de l'issue des matchs des équipes pour la deuxième moitié de la saison 2015-2016.

Rang	Modèle 3		Modèle 4	
	Étranger	Domicile	Étranger	Domicile
1	EDM (66,9 %)	BUF (69,9 %)	EDM (72,1 %)	BUF (70,2 %)
2	CGY (65,7 %)	TOR (68,0 %)	CGY (65,8 %)	LAK (66,8 %)
3	TOR (65,5 %)	LAK (65,6 %)	TOR (65,4 %)	TOR (66,8 %)
4	CBJ (61,4 %)	TBL (62,6 %)	CBJ (65,1 %)	TBL (62,4 %)
5	DET (61,1 %)	ANA (60,4 %)	ANA (61,1 %)	WPG (60,6 %)
⋮	⋮	⋮	⋮	⋮
26	OTT (45,8 %)	DET (48,3 %)	SJS (44,6 %)	FLA (47,5 %)
27	ARI (43,8 %)	STL (48,2 %)	VAN (44,4 %)	STL (47,0 %)
28	SJS (43,7 %)	NSH (47,5 %)	ARI (44,0 %)	MIN (46,9 %)
29	VAN (43,6 %)	FLA (46,1 %)	MIN (43,6 %)	DET (45,6 %)
30	WPG (26,6 %)	VAN (40,7 %)	WPG (20,7 %)	VAN (42,1 %)

4.2.2 Résultats par rapport aux gardiens de but

La majorité des écarts de pourcentages de bonnes prédictions, entre les deux premiers modèles et les deux derniers, est causée par l'ajout d'un effet gardien de but partant. En fait, les deux premiers modèles n'ajustent pas les résultats des matchs en fonction de quel gardien, entre le premier et le deuxième, est devant le filet au début d'un match précis. Pourtant, il est facile de concevoir qu'au cours de la saison 2014-2015, par exemple, la présence de Carey Price ou de Dustin Tokarski devait avoir un impact différent sur les performances de leur équipe.³

L'effet supposé uniquement défensif permet de citer quelques exemples, dans le tableau 4.11, de niveau de prédiction entre couples de gardiens issus des simulations de la deuxième moitié du calendrier de 2015-2016. De nombreux couples de gardiens bénéficient d'un effet complètement indépendant entre eux, comme les duos d'Anaheim, de Buffalo, de Toronto et de Winnipeg. En fait, il s'agit de quatre duos de gardiens de but très équivalents, où les estimations ont été biaisées par un troisième gardien ayant disputé un bon nombre de matchs servant à l'estimation des paramètres sans être très sollicités dans ceux simulés.⁴ Parfois, seulement le deuxième gardien de but bénéficie de cette indépendance, comme pour Chicago, parce que presque la totalité des affrontements du calibre fut associée à un premier gardien très performant, quand le deuxième performe beaucoup moins. Finalement, dans de très rares cas, les gardiens ont été pénalisés par ce concept. La situation des gardiens des Canucks de Vancouver est particulière, parce qu'ils ont plus été victimes de malchance liée à la modélisation qu'autre chose. Pour résumer, Jacob Markstrom affichait des statistiques impressionnantes contre son

3. Price possédait une fiche de 44-16-6 accordant 1,96 buts par 60 minutes, tandis que Tokarski possédait une fiche de 6-6-4 accordant 2,75 buts par 60 minutes.

4. Les troisièmes gardiens sont respectivement Anton Khudobin, Linus Ullmark, James Reimer et Michael Hutchinson.

association pour extrêmement mal paraître en fin de saison contre sa division, tandis que le niveau des performances de Ryan Miller a aussi diminué au fil du calendrier. Bref, l'évolution de leur alignement à travers la saison est ce qui a provoqué cette situation.

TABLE 4.11 – Moyenne du pourcentage de bonnes prédictions de l'issue des matchs des gardiens pour la deuxième moitié de la saison 2015-2016.

Équipe	Gardien	Match commencé	Modèle 1	Modèle 2	Modèle 3	Modèle 4
ANA	John Gibson	9	50,1 %	53,1 %	61,7 %	61,8 %
	Frederik Andersen	7	46,1 %	54,2 %	59,3 %	58,5 %
BUF	Robin Lehner	9	64,5 %	59,5 %	67,6 %	70,0 %
	Chad Johnson	7	46,7 %	46,5 %	57,7 %	54,5 %
TOR	Jonathan Bernier	10	64,3 %	58,6 %	62,5 %	62,7 %
	Garret Sparks	7	64,1 %	57,2 %	78,9 %	73,7 %
WPG	Ondrej Pavelec	6	41,8 %	44,9 %	55,8 %	59,0 %
	Connor Hellebuyck	6	45,1 %	42,2 %	48,3 %	50,2 %
CHI	Corey Crawford	11	49,5 %	49,7 %	50,5 %	50,8 %
	Scott Darling	6	39,9 %	44,2 %	68,3 %	67,0 %
VAN	Ryan Miller	9	60,4 %	50,9 %	43,7 %	45,0 %
	Jacob Markstrom	8	54,4 %	47,9 %	40,5 %	41,4 %

4.3 Sources d'erreur dans les prédictions et corrections suggérées

Évidemment, les modèles construits dans cet essai possèdent des faiblesses perceptibles rapidement, par exemple l'hypothèse induite que les niveaux offensifs et défensifs des équipes sont complètement fixés après la période dédiée aux estimations. En fait, utiliser directement les données passées pour prédire le futur sous-entend que les équipes n'ont pas nécessairement évolué à travers le temps. Même s'il s'agit d'informations connues et disponibles, les transactions et les blessures n'affectent en rien les prédictions des modèles, ce qui néglige l'évolution des capacités des équipes au cours d'une même saison. Cependant, inclure correctement ces composantes dans les modèles peut s'avérer fastidieux. Bref, cette section tentera d'identifier un grand nombre d'affrontements susceptible d'être très mal prédit, pour éventuellement les exclure des analyses.

4.3.1 Transactions et blessures

Sans aucun doute, une transaction impliquant un joueur ou gardien d'impact peut altérer les forces d'une équipe. Le site officiel du Réseau des sports (RDS)⁵, une chaîne de télévision

5. Voir : <http://www.rds.ca/hockey/lnh/signatures/onglet/transactions>

sportive québécoise diffusant des événements sportifs ainsi que des nouvelles du monde du sport 24 heures par jour, permet de cibler ces transactions dites majeures. D’ailleurs, à l’aube du mois de mars de la saison 2013-2014, trois gardiens de premier plan furent échangés dans l’association concurrente.⁶ Par conséquent, pour le troisième et quatrième modèle, ces gardiens ne possèdent aucune moyenne de buts accordés par état pour leur nouvelle équipe, donc ils sont ajustés comme des gardiens de ligues mineures. En fait, il faut éviter de tomber dans le piège d’utiliser les buts accordés obtenus dans leurs équipes précédentes puisque la moyenne d’un gardien de but est construite non seulement par le gardien, mais aussi par la qualité de la brigade défensive dont il dispose. Bref, afin de minimiser les erreurs de prévisions, il vaut mieux exclure leurs affrontements.

Le tableau 4.12 montre un autre problème tout aussi rare, mais qui vient polluer énormément les prédictions des modèles. En accumulant les blessures et en multipliant les transactions pour combler quelques faiblesses, une équipe comme les Canadiens de Montréal a pu voir son alignement se détériorer au fil de l’année 2015-2016. En utilisant le nombre de matchs joués (MJ) en carrière avant la saison 2015-2016, on peut constater le déclin flagrant d’expérience au niveau professionnel des joueurs des Canadiens en avançant dans la saison. Le nombre de matchs joués total est passé de 7434 après la première semaine d’activité, à 5975 à la mi-saison, pour terminer à 3999 lors de l’avant-dernière semaine de la saison.

TABLE 4.12 – Évolution de l’expérience des joueurs dans l’alignement des Canadiens de Montréal au cours de la saison 2015-2016.

Pos.	11 octobre 2015	MJ	1 janvier 2016	MJ	2 avril 2016	MJ
A	T. Plekanec	761	T. Plekanec	761	T. Plekanec	761
A	A. Semin	635	T. Fleischmann	600	T. Mitchell	457
A	T. Fleischmann	600	T. Mitchell	457	M. Pacioretty	399
A	T. Mitchell	457	M. Pacioretty	399	L. Eller	363
A	M. Pacioretty	399	L. Eller	363	D. Desharnais	339
A	L. Eller	363	D. Desharnais	339	B. Gallagher	207
A	D. Desharnais	339	D. Weise	273	A. Galchenyuk	193
A	D. Weise	273	B. Gallagher	207	P. Byron	138
A	B. Gallagher	207	A. Galchenyuk	193	P. Danault	32
A	A. Galchenyuk	193	B. Flynn	168	S. Andrighetto	12
A	B. Flynn	168	P. Byron	138	D. Carr	0
A	D. Smith-Pelly	167	D. Carr	0	M. McCarron	0
D	A. Markov	846	A. Markov	846	A. Markov	846
D	T. Gilbert	592	P. Subban	366	A. Emelin	232
D	P. Subban	366	J. Petry	314	G. Pateryn	20
D	J. Petry	314	A. Emelin	232	D. Dietz	0
D	A. Emelin	232	M. Barberio	103	J. Hanley	0
D	N. Beaulieu	87	N. Beaulieu	87	B. Lernout	0
G	C. Price	435	B. Scrivens	129	M. Condon	0
G	M. Condon	0	M. Condon	0	C. Lindgren	0

6. Roberto Luongo passe de Vancouver à la Floride, Ryan Miller passe de Buffalo à St-Louis et Jaroslav Halak passe de St-Louis à Buffalo, puis à Washington.

Habituellement, le niveau d'expérience des joueurs de l'alignement n'est pas un réel indicateur du talent d'une équipe. Toutefois, la situation devient alarmante lorsque seulement 2 joueurs de la brigade défensive (inclue les défenseurs et les gardiens de but) ont plus d'une demi-saison d'expérience dans la Ligue nationale de hockey. Donc, il est facile d'affirmer que les paramètres estimés de tous les modèles pour les Canadiens de Montréal de 2015-2016 deviennent de plus en plus inexacts. Une fois ce problème identifié, il est important de retirer les simulations qui y sont liées.

4.3.2 Changement d'entraîneur-chef

Au hockey, la majorité des ajustements stratégiques lors d'un affrontement proviennent du personnel d'entraîneur de l'équipe. Ce dernier se compose du patron, nommé entraîneur-chef, et d'un groupe d'assistants l'épaulant dans ses décisions. Habituellement, dans l'éventualité où une équipe sous-performe lors d'une longue période de temps, on s'attend au congédiement de l'entraîneur-chef dans les prochains jours. Cependant, son successeur n'a pas nécessairement la même mentalité et les mêmes rapports avec chacun des joueurs de l'organisation. En fait, l'arrivée d'un nouvel entraîneur au milieu d'une saison peut amener certaines instabilités sur la glace ou hors glace, comme elle peut en corriger d'autres. Au cours des trois derniers calendriers, soit ceux de 2013-2014, de 2014-2015 et de 2015-2016, 11 congédiements d'entraîneurs-chefs furent effectués. Généralement, les remerciements sont survenus avant la mi-saison, pour ainsi permettre à l'équipe dans cette situation de redresser les voiles.

Pour les modèles de simulations, un congédiement d'entraîneur-chef correspond à un débalancement probable des intensités autant offensives que défensives. Par conséquent, même les comportements stratégiques deviennent désuets. D'ailleurs, comme la plupart des changements de personnel ont lieu juste avant la fin du calibrage des modèles, soit en début janvier, plusieurs nouveaux entraîneurs-chefs n'auront pas suffisamment disputé d'affrontements pour considérer leurs stratégies, indépendamment de leur prédécesseur, comme bien modélisées. Parmi eux, on compte Paul Maurice (Jets de Winnipeg) pour la saison 2013-2014, on compte Dave Cameron (Sénateurs d'Ottawa), Todd Nelson (Oilers d'Edmonton), un duo composé de Scott Stevens et Adam Oates (Devils du New Jersey) et Peter Horachek (Maple Leafs de Toronto) pour la saison 2014-2015, et Mike Sullivan (Penguins de Pittsburgh) pour la saison 2015-2016. Donc, leurs affrontements respectifs devront éventuellement être exclus des analyses afin de réduire le nombre d'affrontements biaisés. Sinon, on doit avoir un modèle de simulation dynamique qui intègre l'information au fur et à mesure.

4.4 Théorie de la décision dans les paris sportifs

La théorie de la décision, développée par von Neumann et Morgenstern (1944), est une théorie ayant pour objet la prise de décision. Elle peut être divisée en deux branches : celle normative,

qui donne des conseils sur la façon de prendre les meilleures décisions, étant donné un ensemble de croyances incertaines et un ensemble de valeurs; et celle descriptive, qui analyse comment les agents, peut-être irrationnels, prennent des décisions. En présence de résultats risqués, un décideur pourrait utiliser le critère de la valeur attendue comme règle de décision, où les investissements ayant la plus grande valeur attendue sont tout simplement les préférés. Par exemple, supposons qu'il coûte 2 \$ pour participer à un jeu dans lequel la probabilité d'obtenir un paiement de 100 \$ est de 1 %, et que les 99 % restant impliquent un paiement nul. La valeur attendue de ce jeu est de 1 \$ ($100 \times 1 \%$) et la valeur espérée, qui se calcule à partir de l'équation (4.1), est de $-0,98$ \$. En vertu de la théorie de l'utilité espérée, les personnes rationnelles devraient préférer ne pas jouer à ce jeu. Cette dernière théorie, beaucoup plus représentative du comportement humain, dit que la valeur subjective associée au pari d'un individu est l'espérance statistique des évaluations des résultats de ce pari pour cette personne. De plus, une certaine logique oblige une utilité marginale décroissante puisqu'un individu déjà riche est moins affecté par le gain d'un dollar supplémentaire qu'un individu pauvre.

$$\mathbb{E}[\text{Gains}] = \mathbb{P}[\text{Réussite}] \times \text{Paiement} - \mathbb{P}[\text{Échec}] \times \text{Coût} \quad (4.1)$$

En mixant la base de données des paris avec les probabilités de victoires de chacune des équipes face aux autres estimées auparavant, il est possible de déterminer les gains potentiels associés à chacun des modèles. À titre d'exemple, le tableau 4.13 montre les décisions prises par l'agent typique en présence des cotes associées aux équipes et de leur probabilité de gagner calculée avec le quatrième modèle.⁷ Les valeurs espérées, pour une équipe i , y sont exprimées pour un pari de 1 \$, comme à l'équation (4.2), et le choix final du pari s'arrête sur le choix maximisant l'utilité espérée, pourvu que la valeur surpasse un seuil critique désiré (ici il sera de 0).

$$\mathbb{E}[\text{Gains}_i] = \begin{cases} \left(\mathbb{P}[\text{Victoire}_i] \times \frac{\text{ML}_i}{100} - \mathbb{P}[\text{Défaite}_i] \right) \times \text{Mise} , & \text{si } \text{ML}_i \geq 0 \\ \left(\mathbb{P}[\text{Victoire}_i] \times \frac{100}{\text{ML}_i} - \mathbb{P}[\text{Défaite}_i] \right) \times \text{Mise} , & \text{si } \text{ML}_i < 0 \end{cases} \quad (4.2)$$

TABLE 4.13 – Extrait des choix des paris sportifs estimés par le quatrième modèle.

Date	Etr.	Dom.	ML. Etr.	ML. Dom.	$\mathbb{P}[\text{Victoire}]$ Domicile	$\mathbb{E}[\text{Gains}]$ Etranger	$\mathbb{E}[\text{Gains}]$ Domicile	Choix Final
42378	BOS	OTT	-110	+100	48,1 %	-0,008	-0,039	—
42384	BOS	BUF	-119	+108	11,5 %	0,630	-0,762	BOS
42385	TOR	BOS	+155	-172	81,7 %	-0,532	0,291	BOS
42386	FLA	TBL	+140	-155	48,8 %	0,229	-0,197	FLA
42388	BOS	MTL	-105	-105	58,7 %	-0,193	0,145	MTL

7. Pour simplifier la lecture du tableau, les duels de gardiens de but furent remplacés par les duels d'équipes, puisque le quatrième modèle ne considèrera pas un affrontement entre Boston et Ottawa, mais plutôt un duel entre Tuukka Rask représentant Boston et Craig Anderson représentant Ottawa.

4.4.1 Valeur ajoutée potentielle des modèles

En considérant que l'agent typique ne fasse que des paris de 1 \$, il est possible de déterminer le meilleur modèle parmi les quatre proposé en se fiant au revenu potentiel qu'ils peuvent générer. Dans un premier temps, on considère un agent naïf, c'est-à-dire un individu pariant à tous les matchs sur l'équipe estimée favorite par les différents modèles. À priori, ce comportement paraît logique, sauf qu'il néglige complètement les cotes du preneur de paris. En fait, dans l'éventualité où les deux équipes possèdent une probabilité de gagner de 0,5, l'agent naïf ne se basera toujours pas sur les cotes pour prendre sa décision, mais misera aveuglément sur l'équipe à domicile.

Cette attitude implique les valeurs ajoutées, exprimées en dollar, disponibles au tableau 4.14.⁸ D'ailleurs, lorsque l'on considère les 691 paris effectués, on constate que l'attitude naïve se combine plutôt mal aux deux premiers modèles, engendrant des déficits respectifs, sur trois années, de 19,94 \$ et de 18,57 \$, soit des pertes de 0,029 \$ et de 0,027 \$ par pari. Toutefois, cette stratégie affiche des gains plus intéressants avec les deux derniers modèles. Par exemple avec le quatrième modèle, elle a permis d'obtenir un gain de 11,24 % par pari lors de la saison 2015-2016, malgré le fait que les années précédentes correspondaient à des gains de -6,60 % et de 7,21 % par pari. Définitivement, cette attitude possède une faiblesse stratégique flagrante, malgré qu'elle parie correctement sur 57,2 % des matchs seulement en misant du côté de l'équipe favorite selon le quatrième modèle. En fait, un affrontement trop facile à prédire n'est généralement pas payant puisque les preneurs de paris ajusteront les cotes du pari en conséquence.

TABLE 4.14 – Valeur ajoutée potentielle des modèles de simulations pour un agent naïf.

Division-saison	Modèle 1	Modèle 2	Modèle 3	Modèle 4
Atl. 15-16 (66)	6,92 (66)	-4,60 (66)	0,81 (66)	8,01 (66)
Atl. 14-15 (59)	2,06 (59)	10,61 (59)	0,45 (59)	-8,39 (59)
Atl. 13-14 (61)	4,62 (61)	4,53 (61)	4,11 (61)	6,13 (61)
Mét. 15-16 (73)	0,13 (73)	3,75 (73)	11,00 (73)	12,25 (73)
Mét. 14-15 (62)	-5,56 (62)	-8,14 (62)	1,40 (62)	1,00 (62)
Mét. 13-14 (58)	4,89 (58)	-8,77 (58)	4,90 (58)	2,38 (58)
Cen. 15-16 (53)	-6,50 (53)	-13,50 (53)	4,86 (53)	5,79 (53)
Cen. 14-15 (52)	-18,27 (52)	-9,75 (52)	9,91 (52)	-17,32 (52)
Cen. 13-14 (46)	-6,66 (46)	-8,58 (46)	-4,26 (46)	6,82 (46)
Pac. 15-16 (59)	-1,33 (59)	8,43 (59)	3,05 (59)	2,17 (59)
Pac. 14-15 (48)	-0,60 (48)	9,19 (48)	4,44 (48)	10,12 (48)
Pac. 13-14 (54)	0,36 (54)	-1,73 (54)	-7,69 (54)	0,45 (54)
Total (691)	-19,94 (691)	-18,57 (691)	32,97 (691)	29,41 (691)
Paris réussis	53,8 %	54,4 %	57,2 %	57,2 %

8. Le nombre de paris effectués par division-saison et par modèle est affiché entre parenthèses.

Dans un deuxième temps, on considère un agent rationnel, qui contrairement à l'agent naïf, basera ses décisions sur la théorie de l'utilité espérée. Pour résumer, la décision de l'agent rationnel est guidée par l'équation (4.2). Son choix final repose sur celui maximisant son utilité espérée, considérant qu'il estime une utilité espérée supérieure à une valeur désirée, logiquement plus grande que 0. Si cette condition n'est pas respectée, l'agent rationnel ne parie pas sur le match en question. En fixant un seuil d'utilité espérée désiré de 6 %, on obtient le tableau 4.15, où légèrement moins de paris sont effectués.

Ce comportement n'est pas du tout compatible avec le premier modèle, qui engendre des pertes de 0,052 \$ par pari sur trois ans. Néanmoins, l'intérêt de ce comportement apparaît dans les modèles plus sophistiqués, soit le troisième et le quatrième. Annuellement, si l'agent construit ses pronostics à l'aide du quatrième modèle, il aurait engendré des profits d'approximativement 21,85 \$ pour la saison 2015-2016, 3,41 \$ pour la saison 2014-2015 et 32,09 \$ pour la saison 2013-2014. Il s'agit de retour d'investissement de l'ordre de 10,61 %, 1,98 % et 18,34 %, respectivement. En fait, cette stratégie domine la stratégie naïve, impliquant des gains généralement plus stables entre les différentes saisons, même avec une réduction significative du nombre de paris réussis. En revanche, l'agent rationnel possède une attitude risquée, puisqu'elle implique de miser sans tenir compte des éléments exogènes aux modèles, tels que les blessures et suspensions, sans nécessairement se refléter adéquatement dans les cotes.

TABLE 4.15 – Valeur ajoutée potentielle des modèles de simulations pour un agent rationnel.

Division-saison	Modèle 1	Modèle 2	Modèle 3	Modèle 4
Atl. 15-16 (66)	4,25 (57)	3,76 (57)	3,33 (52)	5,58 (51)
Atl. 14-15 (59)	-0,95 (44)	7,99 (36)	-7,78 (51)	-2,87 (49)
Atl. 13-14 (61)	0,80 (41)	3,73 (33)	8,90 (38)	14,82 (42)
Mét. 15-16 (73)	-4,05 (62)	1,41 (43)	4,63 (59)	8,33 (55)
Mét. 14-15 (62)	-5,80 (45)	-0,27 (39)	7,78 (50)	3,07 (48)
Mét. 13-14 (58)	5,08 (50)	-1,84 (41)	4,52 (45)	10,87 (50)
Cen. 15-16 (53)	-11,24 (47)	-17,13 (46)	1,31 (46)	3,91 (47)
Cen. 14-15 (52)	-8,39 (42)	1,60 (38)	18,28 (42)	6,20 (36)
Cen. 13-14 (46)	-0,57 (34)	1,21 (27)	1,23 (40)	5,19 (39)
Pac. 15-16 (59)	-2,90 (47)	-4,17 (49)	1,79 (57)	4,03 (53)
Pac. 14-15 (48)	-2,02 (38)	9,48 (37)	1,77 (37)	-2,99 (39)
Pac. 13-14 (54)	-3,06 (47)	-1,21 (31)	-12,97 (47)	1,21 (44)
Total (691)	-28,86 (554)	4,55 (477)	32,80 (564)	57,36 (553)
Paris réussis	48,7 %	47,4 %	53,0 %	55,9 %

Théoriquement, bien choisir sur quel affrontement parier est très important afin de minimiser le risque de perdre dans ses paris. Sachant que les modèles ne tiennent pas compte de toute l'information possible pour établir les pourcentages de victoires attribuables à chacun des affrontements, il faut rester prudent avant de miser. La section 4.3 explique quels affrontements

sont à risque d'être mal prédits par les modèles. En les excluant des décisions prises par un agent, dit informé, on obtient les précisions du tableau 4.16. Il est intéressant de noter que dans ce contexte, le quatrième modèle parvient à prédire le gagnant de 54,1 % des affrontements.

TABLE 4.16 – Pourcentage de bonnes prédictions en excluant les affrontements à risque d'être mal prédits, par année et par modèle.

Modèle	2013-2014	2014-2015	2015-2016	Total
1	53,7 %	50,8 %	51,1 %	52,0 %
2	52,3 %	51,4 %	51,5 %	51,7 %
3	51,6 %	52,3 %	55,8 %	53,3 %
4	53,7 %	51,7 %	56,6 %	54,1 %

Si l'agent rationnel tient compte de cette information, le nombre de paris jugé rationnel pour chacun des modèles diminue. En effet, le tableau 4.17 montre que le total des paris considérés passe de 691 à 523 seulement par l'exclusion des affrontements possiblement problématiques. Avec cette attitude pour la saison 2015-2016, en utilisant le modèle le plus complet, soit le quatrième, il était possible d'engendrer un profit de 25,05 \$ en 157 paris pour la saison 2014-2015, donc d'avoir un retour d'investissement de 15,96 %. Les autres saisons, moins lucratives, pouvaient générer un profit de 18,33 \$ pour la saison 2014-2015 et de 20,28 \$ pour la saison 2013-2014, correspondant à des retours d'investissement respectif de 15,28 % et 15,25 %. Bref, l'utilisation avertie de l'information assure des retours d'investissement supérieurs et très stables pour le quatrième modèle, qui maintient un taux de réussite des paris de 58,5 %.

TABLE 4.17 – Valeur ajoutée potentielle des modèles de simulations pour un agent rationnel et informé.

Division-saison	Modèle 1	Modèle 2	Modèle 3	Modèle 4
Atl. 15-16 (50)	3,28 (41)	2,58 (43)	1,19 (39)	6,73 (35)
Atl. 14-15 (31)	2,41 (24)	8,63 (16)	3,49 (27)	3,19 (27)
Atl. 13-14 (55)	1,52 (37)	6,09 (29)	7,98 (34)	12,59 (36)
Mét. 15-16 (51)	-1,27 (44)	5,54 (32)	7,47 (41)	9,91 (39)
Mét. 14-15 (50)	-2,70 (38)	0,43 (32)	5,96 (40)	0,25 (39)
Mét. 13-14 (55)	5,08 (50)	-1,84 (41)	2,87 (43)	8,67 (47)
Cen. 15-16 (41)	-10,86 (35)	-14,95 (36)	1,74 (35)	5,35 (35)
Cen. 14-15 (35)	-1,58 (32)	0,97 (29)	12,84 (29)	11,43 (23)
Cen. 13-14 (22)	-4,15 (20)	-2,96 (12)	-0,01 (17)	3,53 (17)
Pac. 15-16 (54)	-2,90 (47)	-4,17 (49)	0,82 (52)	3,06 (48)
Pac. 14-15 (38)	1,06 (30)	5,74 (29)	6,22 (31)	3,46 (31)
Pac. 13-14 (41)	1,24 (39)	-0,11 (27)	-16,69 (34)	-4,51 (33)
Total (523)	-8,88 (437)	5,95 (375)	33,89 (422)	63,66 (410)
Paris réussis	50,8 %	48,3 %	54,5 %	58,5 %

Malheureusement, la suppression volontaire d'affrontements peut aussi éliminer des sources de gains. Par exemple, en censurant presque l'entièreté des affrontements des Oilers d'Edmonton (équipe de la division Pacifique) lors de la saison 2013-2014, le profit pour le quatrième modèle est passé de 1,21 \$ à -4,51 \$, tout ça en 11 paris de moins. Il s'agit d'une situation particulière puisque leurs deux gardiens, acquis via échange lors de la deuxième moitié de la saison, auraient bénéficié d'être calibrés avec des pénalités, sachant que leur fin de saison fut complètement désastreuse, au lieu d'être exclus comme tous les gardiens échangés pendant cette période de l'année. Bien entendu, définir un meilleur mécanisme contrôlant les évaluations des gardiens échangés aurait pu, éventuellement, permettre de ne pas les exclure.

Au final, lorsque l'agent est informé correctement, l'application des pires modèles aux cotes de Odds Shark génère tout de même des pertes. En fait, les gains potentiels des troisième et quatrième modèles montrent l'importance de tenir compte des gardiens de but dans le choix des paris. D'ailleurs, ajouter de l'information dans les décisions augmente considérablement la constance, ainsi que la valeur des gains du quatrième modèle. Évidemment, le quatrième modèle, pouvant encore être aisément amélioré, est le modèle à favoriser si notre objectif est de maximiser nos profits.

Conclusion

Évidemment, parmi les modèles construits, ce sont les modèles linéaires généralisés Poisson qui ont obtenu les meilleurs taux de prédictions des parties intradivisions, permettant d'atteindre un niveau de précision de 54,1 %. Précisément, il s'agit de la performance du quatrième modèle, qui distingue les 55 premières minutes d'un affrontement des 5 dernières, lorsque l'on exclut les rencontres identifiées comme déficientes. Contrairement à Weissbock (2014), qui utilise des approches d'apprentissage automatisé pour atteindre 59,8 % de précision pour les données seulement de la saison écourtée de 2012-2013 par validation croisée, cette fois les prédictions ont été construites sur trois saisons complètes (2013-2014, 2014-2015 et 2015-2016).

Le fait que le quatrième modèle s'approche de la performance d'une étude plus comparable comme celle de Wolfson et Koopmeiners (2015), généralement autour de 55,0 %, et le fait que la borne supérieure de performance approximative de 62,0 % évoquée par Weissbock (2014) est loin d'être atteinte, permettent de croire que la modélisation peut encore être améliorée. D'ailleurs, énormément de pistes d'amélioration sont connues :

- Augmenter l'information contenue dans les bases de données par rapport aux joueurs présents dans les alignements permettrait de limiter l'exclusion d'affrontements ;
- Augmenter le nombre d'affrontements simulés pour y inclure tous les affrontements intra-associations ;
- Construire des régressions de Poisson qui inclue de nouvelles variables comme les intervalles de temps entre les matchs et les durées des voyages ;
- Tester la linéarité des variables continues avec des splines pour voir si elles doivent entrer de façon linéaire ou transformée dans les modèles ;
- Modifier le mécanisme de pénalisation des moyennes de buts accordés des défensives n'ayant pas suffisamment participé dans les affrontements ;
- Relâcher l'hypothèse qu'un gardien influence seulement la défensive d'une équipe, pour qu'il puisse aussi influencer l'offensive ;
- Scinder les affrontements en trois périodes permettrait possiblement d'améliorer l'adaptation stratégique des équipes.

En fait, s'attarder sur la précision des modèles est une tâche nécessaire afin d'améliorer l'élément crucial de cet essai : les gains potentiels via les paris sportifs. C'est à travers les profits qu'ils peuvent générer que le quatrième modèle a pu afficher sa dominance. Non seulement il prédit correctement un bon nombre d'affrontements, mais il aurait permis d'atteindre, en dépit des réglementations locales, des retours d'investissement de l'ordre de 15,96 %, 15,28 % et 15,25 % pour les trois saisons considérées. Bien que ces résultats ne soient valides qu'avec les cotes de Odds Shark, un agent rationnel et informé profitera autant des cotes avantageuses conçues par les autres preneurs de paris en utilisant ce modèle. Dans l'avenir, il serait très intéressant d'améliorer le volet prise de décisions :

- Développer et vérifier si des algorithmes de décisions pour un agent différent permettraient d'obtenir des conclusions encore plus performantes ;
- Créer une stratégie permettant à l'agent de miser davantage sur les affrontements moins risqués ;
- Trouver et inclure les cotes sur d'autres types de paris, comme le nombre total de buts (Over-Under) ou sur l'écart dans le pointage (Spread) à l'issue d'une rencontre.

Annexe A

Division et abréviation officielle de chaque équipe de la LNH

<u>Association de l'Ouest</u>		<u>Association de l'Est</u>	
<u>Division Centrale</u>		<u>Division Atlantique</u>	
CHI	Chicago Blackhawks	BOS	Boston Bruins
COL	Colorado Avalanche	BUF	Buffalo Sabres
DAL	Dallas Stars	DET	Detroit Red Wings
MIN	Minnesota Wild	FLA	Florida Panthers
NSH	Nashville Predators	MTL	Montreal Canadiens
STL	St.Louis Blues	OTT	Ottawa Senators
WPG	Winnipeg Jets	TBL	Tampa Bay Lightning
		TOR	Toronto Maple Leafs
<u>Division Pacifique</u>		<u>Division Métropolitaine</u>	
ANA	Anaheim Ducks	CAR	Carolina Hurricanes
ARI	Arizona Coyotes	CBJ	Columbus Blue Jackets
CGY	Calgary Flames	NJD	New Jersey Devils
EDM	Edmonton Oilers	NYI	New York Islanders
LAK	Los Angeles Kings	NYR	New York Rangers
SJS	San Jose Sharks	PHI	Philadelphia Flyers
VAN	Vancouver Canucks	PIT	Pittsburgh Penguins
		WSH	Washington Capitals

1. Pour faciliter la lecture, les noms des équipes furent écrits en anglais.

Annexe B

Autres tableaux omis du document

TABLE B.1 – Paramètres estimés pour le premier modèle avec la division Métropolitaine, Centrale et Pacifique. pour la saison 2015-2016.

Équipe	Offensive 2015-2016				Défensive 2015-2016			
	α_{Etr}	λ_{Etr}	α_{Dom}	λ_{Dom}	α_{Etr}	λ_{Etr}	α_{Dom}	λ_{Dom}
CAR	2,263	3,121	2,349	3,079	3,841	3,785	2,792	3,511
CBJ	2,623	2,633	2,125	3,043	2,343	3,795	2,394	4,531
NJD	2,163	3,458	2,217	3,010	2,044	3,095	2,392	3,372
NYI	2,625	3,664	2,723	2,823	2,692	3,629	1,931	3,568
NYR	1,545	3,108	2,301	3,923	3,783	3,437	1,752	3,528
PHI	2,001	3,230	3,075	3,371	2,260	4,024	2,847	3,995
PIT	1,271	2,870	2,434	3,531	1,989	2,811	2,264	3,418
WSH	2,204	3,813	3,628	4,357	1,872	2,964	3,232	2,614
CHI	1,913	3,937	2,617	3,998	3,231	4,434	1,345	2,316
COL	2,073	3,679	2,505	4,681	1,750	2,553	3,084	3,753
DAL	1,932	3,783	3,613	4,625	3,184	4,181	2,289	2,713
MIN	2,392	3,183	3,190	4,343	2,417	3,415	1,729	3,378
NSH	2,503	2,701	1,772	4,030	3,805	4,442	1,879	2,708
STL	2,950	4,370	2,925	3,269	2,979	3,778	2,527	2,815
WPG	1,964	2,744	2,960	3,681	2,368	4,414	2,662	3,707
ANA	1,935	1,790	1,949	3,571	1,593	2,960	2,086	3,150
ARI	3,824	3,580	4,878	4,260	2,001	3,230	1,941	3,747
CGY	2,734	3,381	1,972	3,331	2,031	3,938	2,487	3,757
EDM	1,827	2,814	2,802	3,453	3,915	4,275	2,631	3,952
LAK	4,025	4,422	2,396	3,018	1,676	2,577	2,603	3,749
SJS	2,449	4,037	1,518	2,413	3,087	4,065	1,862	3,805
VAN	1,851	2,993	1,844	3,068	2,134	3,750	2,961	3,808

TABLE B.2 – Paramètres estimés pour le premier modèle pour la saison 2014-2015.

Équipe	Offensive 2014-2015				Défensive 2014-2015			
	λ_{Etr}	α_{Etr}	λ_{Dom}	α_{Dom}	λ_{Etr}	α_{Etr}	λ_{Dom}	α_{Dom}
BOS	2,782	3,481	2,256	3,325	1,834	4,033	2,692	3,105
BUF	1,758	2,564	1,465	2,914	3,724	4,993	2,493	3,732
DET	2,459	3,560	2,111	3,553	2,483	3,689	2,756	3,199
FLA	2,633	2,739	2,018	3,256	2,191	3,147	2,044	3,514
MTL	2,579	3,448	2,286	4,475	1,544	2,885	2,221	3,296
OTT	2,573	3,107	2,970	3,502	2,392	3,495	2,934	3,651
TBL	2,392	4,319	2,678	4,275	2,636	4,021	2,352	2,953
TOR	2,215	4,119	2,686	4,510	2,335	4,058	1,914	4,085
CAR	1,754	2,417	2,511	3,176	2,548	3,313	2,773	3,335
CBJ	3,754	3,323	2,234	3,933	2,457	3,552	2,767	4,692
NJD	1,709	3,000	1,954	3,236	1,935	3,946	2,896	3,681
NYI	2,718	4,136	3,488	3,897	3,029	3,928	2,162	2,916
NYR	3,303	3,797	3,005	4,377	2,613	3,901	1,254	3,118
PHI	1,519	3,041	2,577	3,831	2,880	4,374	2,196	3,362
PIT	2,336	3,771	2,029	4,144	2,483	3,769	2,041	3,340
WSH	2,929	4,159	2,047	3,771	1,661	3,302	3,328	2,960
CHI	2,760	4,368	1,396	3,571	2,511	3,176	2,251	3,102
COL	1,736	2,518	1,674	3,569	2,221	4,188	2,770	3,729
DAL	2,365	3,589	2,159	4,275	1,887	4,173	2,387	3,785
MIN	2,322	3,291	2,838	4,079	1,763	3,709	1,943	3,992
NSH	2,141	4,083	3,727	3,707	1,691	3,430	3,091	2,879
STL	1,617	3,882	2,566	4,322	2,313	3,729	2,119	2,900
WPG	1,677	3,616	1,955	3,128	2,616	3,510	3,273	3,615
ANA	2,460	3,754	2,098	3,275	2,117	3,426	2,227	3,199
ARI	1,808	3,330	1,510	3,009	1,959	4,095	2,615	4,672
CGY	2,785	3,534	3,620	4,269	1,993	2,708	3,999	3,504
EDM	2,316	2,824	1,951	3,070	2,869	4,683	2,726	4,834
LAK	3,490	3,899	1,727	4,061	1,658	3,690	1,683	3,296
SJS	3,347	3,747	2,130	3,660	1,776	3,521	1,680	3,443
VAN	3,152	3,919	1,430	3,596	1,638	3,585	2,760	3,697

TABLE B.3 – Paramètres estimés pour le premier modèle pour la saison 2013-2014.

Équipe	Offensive 2013-2014				Défensive 2013-2014			
	λ_{Etr}	α_{Etr}	λ_{Dom}	α_{Dom}	λ_{Etr}	α_{Etr}	λ_{Dom}	α_{Dom}
BOS	2,517	3,719	5,590	4,104	2,034	3,387	1,824	2,609
BUF	1,933	2,388	2,327	2,634	3,598	3,730	2,525	3,300
DET	3,288	4,097	1,783	3,093	2,222	3,287	2,598	4,019
FLA	3,325	3,010	2,366	3,396	1,996	3,746	2,776	4,184
MTL	2,784	3,103	2,439	3,645	1,810	3,083	3,178	2,788
OTT	1,748	3,675	2,715	4,061	2,139	3,940	2,246	3,718
TBL	1,584	3,250	1,736	3,809	1,792	2,891	2,041	3,292
TOR	2,399	3,307	2,388	4,016	3,039	4,103	2,328	3,881
CAR	3,824	3,580	2,164	2,997	3,299	4,217	2,361	3,758
CBJ	2,530	3,745	1,807	3,840	2,286	3,524	2,677	3,723
NJD	2,292	3,620	2,249	3,337	2,655	3,755	1,555	2,528
NYI	2,535	3,922	2,087	2,895	2,709	4,494	3,403	4,074
NYR	1,988	3,130	2,208	3,425	2,008	2,973	2,198	3,510
PHI	2,030	3,652	1,613	3,523	1,880	3,914	1,951	3,502
PIT	2,563	3,826	3,397	4,357	1,808	3,492	2,059	2,919
WSH	2,013	3,859	1,997	4,063	2,301	3,147	2,929	4,159
CHI	3,017	4,778	2,548	4,391	1,819	3,547	2,289	3,506
COL	3,542	3,471	2,378	4,203	1,801	4,625	2,287	3,154
DAL	2,106	3,975	2,577	3,875	2,192	4,384	2,399	3,557
MIN	1,920	3,069	2,403	3,400	2,733	3,807	2,062	2,988
NSH	1,430	3,117	1,740	3,375	3,727	4,380	2,558	3,376
STL	2,962	4,335	2,504	4,486	1,702	3,120	2,181	3,854
WPG	2,316	2,725	2,382	4,115	2,531	3,748	3,003	4,618
ANA	3,216	3,914	3,245	4,731	2,633	4,333	4,437	2,846
ARI	2,652	3,621	4,214	4,295	2,510	4,277	3,865	4,216
CGY	2,902	3,144	1,976	3,132	2,873	4,456	2,495	4,293
EDM	2,191	3,147	1,160	3,229	2,004	3,994	2,951	4,819
LAK	2,948	2,792	2,192	4,097	3,025	3,410	2,091	3,068
SJS	2,785	3,862	3,540	4,727	2,556	4,032	2,869	3,194
VAN	2,452	3,209	2,004	3,296	2,552	3,640	2,434	2,699

TABLE B.4 – Paramètres estimés pour le deuxième modèle à domicile pour la saison 2015-2016.

Équipe	ρ_L	ρ_E	ρ_T	ψ_L	ψ_E	ψ_T
BOS	0,0552	0,0414	0,0504	0,0478	0,038	0,05960
BUF	0,0505	0,0220	0,0280	0,0433	0,0585	0,0467
DET	0,0278	0,0330	0,0504	0,0417	0,0254	0,0550
FLA	0,0430	0,0369	0,0670	0,0303	0,0246	0,0419
MTL	0,0502	0,0402	0,0623	0,0345	0,0258	0,0089
OTT	0,0629	0,0199	0,0448	0,0270	0,0477	0,0392
TBL	0,0402	0,0450	0,0449	0,0491	0,0554	0,0224
TOR	0,0256	0,0092	0,0600	0,0769	0,0735	0,0424
CAR	0,0374	0,0298	0,0461	0,0624	0,0273	0,0507
CBJ	0,0126	0,0245	0,0516	0,1135	0,0428	0,0516
NJD	0,0476	0,0218	0,0361	0,0416	0,0400	0,0361
NYI	0,0206	0,0368	0,0355	0,0310	0,0409	0,0532
NYR	0,0506	0,0639	0,0258	0,0393	0,0266	0,0774
PHI	0,0185	0,0437	0,0430	0,0678	0,0394	0,0430
PIT	0,0413	0,0417	0,0400	0,0413	0,0500	0,0320
WSH	0,0582	0,0467	0,0585	0,0356	0,0311	0,0195
CHI	0,0532	0,0449	0,0438	0,0293	0,0242	0,0175
COL	0,0585	0,0628	0,0541	0,0585	0,0333	0,0379
DAL	0,0535	0,0510	0,0759	0,0497	0,0170	0,0380
MIN	0,0529	0,0557	0,0731	0,0453	0,0278	0,0627
NSH	0,0634	0,0499	0,0304	0,0222	0,0333	0,0426
STL	0,0359	0,0402	0,0329	0,0430	0,0287	0,0282
WPG	0,0383	0,0447	0,0705	0,0479	0,0406	0,0588
ANA	0,0354	0,0562	0,0430	0,0354	0,0303	0,0335
ARI	0,0457	0,0342	0,0805	0,0356	0,0391	0,0604
CGY	0,0343	0,0403	0,0378	0,0481	0,0403	0,0515
EDM	0,0169	0,0258	0,0666	0,0591	0,0443	0,0475
LAK	0,0199	0,0528	0,0200	0,0646	0,0384	0,0399
SJS	0,0698	0,0127	0,0257	0,0279	0,0637	0,0449
VAN	0,0341	0,0197	0,0485	0,0681	0,0432	0,0397

TABLE B.5 – Paramètres estimés pour le deuxième modèle à l'étranger pour la saison 2015-2016.

Équipe	ρ_L	ρ_E	ρ_T	ψ_L	ψ_E	ψ_T
BOS	0,0427	0,0648	0,0373	0,0396	0,0389	0,0249
BUF	0,0269	0,0650	0,0542	0,0493	0,0418	0,0406
DET	0,0304	0,0505	0,0450	0,0541	0,0168	0,0600
FLA	0,0617	0,0397	0,0328	0,0347	0,0253	0,0246
MTL	0,0478	0,0660	0,0509	0,0434	0,0363	0,0305
OTT	0,0536	0,0275	0,0389	0,0482	0,0458	0,0466
TBL	0,0526	0,0374	0,0576	0,0526	0,0306	0,0512
TOR	0,0308	0,0291	0,0363	0,0554	0,0323	0,0444
CAR	0,0343	0,0367	0,0375	0,0514	0,0459	0,0412
CBJ	0,0406	0,0231	0,0334	0,0406	0,0396	0,0577
NJD	0,0537	0,0209	0,0557	0,0331	0,0371	0,0348
NYI	0,0380	0,0403	0,0543	0,0456	0,0293	0,0543
NYR	0,0372	0,0307	0,0449	0,0929	0,0279	0,0299
PHI	0,0281	0,0388	0,0338	0,0749	0,0350	0,0507
PIT	0,0774	0,0246	0,0250	0,0129	0,0246	0,0500
WSH	0,0372	0,0483	0,0618	0,0284	0,0369	0,0428
CHI	0,0476	0,0503	0,0409	0,0714	0,0503	0,0446
COL	0,0497	0,0615	0,0269	0,0230	0,0384	0,0337
DAL	0,0431	0,0539	0,0426	0,0554	0,0539	0,0373
MIN	0,0289	0,0253	0,0635	0,0520	0,0282	0,0357
NSH	0,0000	0,0394	0,0402	0,0570	0,0525	0,0563
STL	0,0505	0,0431	0,0788	0,0463	0,0518	0,0368
WPG	0,0598	0,0197	0,0312	0,0374	0,0709	0,0527
ANA	0,0111	0,0185	0,0231	0,0278	0,0247	0,0509
ARI	0,0285	0,0473	0,0573	0,0407	0,0364	0,0358
CGY	0,0402	0,0331	0,0441	0,0503	0,0442	0,0521
EDM	0,0456	0,0218	0,0357	0,0911	0,0437	0,0549
LAK	0,0717	0,0410	0,0549	0,0302	0,0287	0,0329
SJS	0,0664	0,0305	0,0743	0,0664	0,0566	0,0286
VAN	0,0452	0,0309	0,0280	0,0370	0,0397	0,0520

TABLE B.6 – Paramètres estimés pour le deuxième modèle à domicile pour la saison 2014-2015.

Équipe	ρ_L	ρ_E	ρ_T	ψ_L	ψ_E	ψ_T
BOS	0,0272	0,0403	0,0513	0,0374	0,0350	0,0257
BUF	0,0460	0,0195	0,0356	0,0658	0,0365	0,0445
DET	0,0493	0,0251	0,0670	0,0493	0,0289	0,0335
FLA	0,0128	0,0385	0,0575	0,0343	0,0385	0,0460
MTL	0,0705	0,0428	0,0574	0,0397	0,0385	0,0358
OTT	0,0302	0,0542	0,0336	0,0905	0,0232	0,0224
TBL	0,0472	0,0530	0,0709	0,0359	0,0371	0,0253
TOR	0,0788	0,0476	0,0487	0,0410	0,0578	0,0573
CAR	0,0345	0,0357	0,0370	0,0493	0,0357	0,0337
CBJ	0,0399	0,0541	0,0471	0,0621	0,0375	0,0746
NJD	0,0398	0,0421	0,0322	0,0465	0,0345	0,0414
NYI	0,0388	0,0499	0,0371	0,0543	0,0233	0,0247
NYR	0,0489	0,0579	0,0587	0,0276	0,0386	0,0763
PHI	0,0537	0,0317	0,0553	0,0358	0,0458	0,0323
PIT	0,0534	0,0448	0,0561	0,0394	0,0395	0,0386
WSH	0,0420	0,0461	0,0552	0,0532	0,0163	0,0184
CHI	0,0785	0,0265	0,0433	0,0500	0,0413	0,0233
COL	0,0589	0,0429	0,0310	0,0642	0,0462	0,0310
DAL	0,0691	0,0432	0,0486	0,0576	0,0346	0,0340
MIN	0,0465	0,0406	0,0632	0,0465	0,0447	0,0464
NSH	0,0379	0,0327	0,0724	0,0649	0,0252	0,0145
STL	0,0533	0,0446	0,0683	0,0345	0,0372	0,0303
WPG	0,0339	0,0460	0,0233	0,0581	0,0460	0,0280
ANA	0,0236	0,0430	0,0396	0,0439	0,0276	0,0396
ARI	0,0489	0,0269	0,0296	0,0587	0,0471	0,0680
CGY	0,0414	0,0458	0,0686	0,0552	0,0493	0,0274
EDM	0,0309	0,0345	0,0383	0,0617	0,0613	0,0623
LAK	0,0545	0,0438	0,0498	0,0272	0,0438	0,0463
SJS	0,0388	0,0611	0,0343	0,0466	0,0436	0,0343
VAN	0,0499	0,0410	0,0355	0,0399	0,0351	0,0592

TABLE B.7 – Paramètres estimés pour le deuxième modèle à l'étranger pour la saison 2014-2015.

Équipe	ρ_L	ρ_E	ρ_T	ψ_L	ψ_E	ψ_T
BOS	0,0393	0,0325	0,0423	0,0589	0,0390	0,0577
BUF	0,0133	0,0358	0,0262	0,0797	0,0553	0,0681
DET	0,0470	0,0227	0,0691	0,0329	0,0429	0,0518
FLA	0,0186	0,0224	0,0508	0,0497	0,0273	0,0323
MTL	0,0476	0,0357	0,0310	0,0330	0,0259	0,0464
OTT	0,0130	0,0341	0,0539	0,0561	0,0363	0,0347
TBL	0,0678	0,0406	0,0561	0,0475	0,0435	0,0561
TOR	0,0610	0,0315	0,0688	0,0563	0,0409	0,0529
CAR	0,0379	0,0254	0,0240	0,0664	0,0507	0,0240
CBJ	0,0313	0,0348	0,0472	0,0626	0,0316	0,0425
NJD	0,0359	0,0431	0,0253	0,0588	0,0383	0,0534
NYI	0,0438	0,0521	0,0519	0,0606	0,0441	0,0346
NYR	0,0444	0,0377	0,0521	0,0571	0,0566	0,0347
PHI	0,0737	0,0143	0,0462	0,0589	0,0645	0,0462
PIT	0,0404	0,0559	0,0481	0,0353	0,0466	0,0525
WSH	0,0484	0,0527	0,0572	0,0341	0,0362	0,0408
CHI	0,0554	0,0553	0,0541	0,0388	0,0369	0,0406
COL	0,0142	0,0168	0,0390	0,0284	0,0447	0,0564
DAL	0,0170	0,0524	0,0616	0,0551	0,0360	0,0728
MIN	0,0239	0,0323	0,0509	0,0359	0,0554	0,0463
NSH	0,0461	0,0442	0,0650	0,0436	0,0472	0,0217
STL	0,0754	0,0271	0,0547	0,0488	0,0475	0,0376
WPG	0,0437	0,0678	0,0260	0,0437	0,0339	0,0433
ANA	0,0294	0,0442	0,0703	0,0490	0,0315	0,0319
ARI	0,0476	0,0322	0,0380	0,0366	0,0515	0,0633
CGY	0,0476	0,0315	0,0421	0,0297	0,0275	0,0337
EDM	0,0263	0,0280	0,0409	0,0922	0,0420	0,0486
LAK	0,0502	0,0463	0,0530	0,0418	0,0411	0,0398
SJS	0,0360	0,0457	0,0570	0,0360	0,0326	0,0614
VAN	0,0359	0,0597	0,0535	0,0392	0,0299	0,0680

TABLE B.8 – Paramètres estimés pour le deuxième modèle à domicile pour la saison 2013-2014.

Équipe	ρ_L	ρ_E	ρ_T	ψ_L	ψ_E	ψ_T
BOS	0,0453	0,0625	0,0483	0,0389	0,0256	0,0138
BUF	0,0306	0,0183	0,0367	0,0766	0,0314	0,0404
DET	0,0393	0,0525	0,0179	0,0561	0,0340	0,0501
FLA	0,0383	0,0436	0,0365	0,0574	0,0400	0,0562
MTL	0,0459	0,0541	0,0344	0,0459	0,0291	0,0115
OTT	0,0455	0,0542	0,0563	0,0401	0,0406	0,0563
TBL	0,0641	0,0441	0,0312	0,0458	0,0361	0,0364
TOR	0,0494	0,0453	0,0435	0,0570	0,0314	0,0566
CAR	0,0127	0,0365	0,0496	0,0381	0,0497	0,0454
CBJ	0,0549	0,0444	0,0484	0,0549	0,0317	0,0591
NJD	0,0427	0,0413	0,0325	0,0328	0,0206	0,0217
NYI	0,0279	0,0311	0,0405	0,0781	0,0283	0,0446
NYR	0,0475	0,0245	0,0586	0,0346	0,0398	0,0496
PHI	0,0488	0,0456	0,0377	0,0300	0,0332	0,0706
PIT	0,0435	0,0692	0,0804	0,0410	0,0216	0,0402
WSH	0,0433	0,0466	0,0486	0,0770	0,0498	0,0405
CHI	0,0534	0,0444	0,0887	0,0502	0,0237	0,0484
COL	0,0522	0,0441	0,0680	0,0313	0,0373	0,0432
DAL	0,0440	0,0409	0,0601	0,0367	0,0409	0,0361
MIN	0,0367	0,0419	0,0434	0,0267	0,0349	0,0394
NSH	0,0429	0,0411	0,0335	0,0396	0,0299	0,0603
STL	0,0554	0,0573	0,0567	0,0450	0,0287	0,0648
WPG	0,0489	0,0301	0,0712	0,0598	0,0685	0,0456
ANA	0,0476	0,0577	0,1503	0,0446	0,0216	0,0215
ARI	0,0343	0,0476	0,0655	0,0687	0,0476	0,0450
CGY	0,0367	0,0298	0,0405	0,0514	0,0715	0,0405
EDM	0,0888	0,0401	0,0225	0,0565	0,0601	0,0638
LAK	0,0648	0,0286	0,0680	0,0443	0,0338	0,0247
SJS	0,0525	0,0654	0,0584	0,0394	0,0594	0,0117
VAN	0,0398	0,0427	0,0289	0,0319	0,0237	0,0289

TABLE B.9 – Paramètres estimés pour le deuxième modèle à l'étranger pour la saison 2013-2014.

Équipe	ρ_L	ρ_E	ρ_T	ψ_L	ψ_E	ψ_T
BOS	0,0493	0,0440	0,0445	0,0301	0,0519	0,0489
BUF	0,0372	0,0140	0,0318	0,0744	0,0629	0,0295
DET	0,0467	0,0487	0,0587	0,0467	0,0337	0,0196
FLA	0,0342	0,0335	0,0395	0,0342	0,0558	0,0428
MTL	0,0293	0,0269	0,0475	0,0461	0,0239	0,0407
OTT	0,0491	0,0380	0,0546	0,0368	0,0380	0,0729
TBL	0,0515	0,0216	0,0369	0,0396	0,0297	0,0323
TOR	0,0372	0,0314	0,0457	0,0807	0,0409	0,0374
CAR	0,0527	0,0321	0,0537	0,0469	0,0588	0,0537
CBJ	0,0497	0,0397	0,0450	0,0435	0,0397	0,0418
NJD	0,0252	0,0539	0,0441	0,0554	0,0467	0,0413
NYI	0,0403	0,0287	0,0636	0,0403	0,0754	0,0530
NYR	0,0348	0,0338	0,0396	0,0313	0,0431	0,0308
PHI	0,0509	0,0323	0,0542	0,0382	0,0420	0,0576
PIT	0,0503	0,0392	0,0495	0,0327	0,0458	0,0495
WSH	0,0661	0,0298	0,0545	0,0463	0,0298	0,0272
CHI	0,0691	0,0539	0,0585	0,0395	0,0412	0,0502
COL	0,0265	0,0377	0,0460	0,0637	0,0414	0,0675
DAL	0,0516	0,0763	0,0276	0,0516	0,0583	0,0587
MIN	0,0693	0,0251	0,0310	0,1039	0,0691	0,0248
NSH	0,0556	0,0563	0,0169	0,0741	0,0422	0,0508
STL	0,0499	0,0531	0,0624	0,0352	0,0193	0,0624
WPG	0,0000	0,0141	0,0541	0,0837	0,0537	0,0271
ANA	0,0377	0,0628	0,0415	0,0566	0,0377	0,0652
ARI	0,0379	0,0402	0,0478	0,0521	0,0475	0,0506
CGY	0,0000	0,0415	0,0389	0,0868	0,0452	0,0584
EDM	0,0284	0,0343	0,0473	0,0567	0,0429	0,0512
LAK	0,0000	0,0149	0,0475	0,0972	0,0418	0,0292
SJS	0,0482	0,0420	0,0469	0,0572	0,0330	0,0554
VAN	0,0217	0,0399	0,0383	0,0541	0,0300	0,0511

TABLE B.10 – Suite des paramètres estimés pour le troisième modèle avec l’asso. de l’Est.

Variable	2013-2014			2014-2015			2015-2016		
	L	E	T	L	E	T	L	E	T
CAR	-4,653	-4,055	-3,965	-4,548	-4,516	-4,397	-4,257	-4,404	-4,101
CBJ	-3,984	-3,942	-4,023	-4,258	-4,180	-4,163	-4,308	-4,668	-4,165
NJD	-4,548	-3,834	-4,130	-4,281	-4,149	-4,468	-3,888	-4,832	-4,109
NYI	-4,419	-4,241	-3,926	-4,167	-3,873	-3,961	-4,434	-4,286	-4,069
NYR	-4,155	-4,331	-3,988	-4,052	-4,165	-3,794	-3,998	-4,221	-4,418
PHI	-4,068	-4,133	-4,251	-3,864	-4,637	-3,870	-4,635	-4,189	-4,353
PIT	-4,087	-3,839	-4,001	-4,164	-4,022	-3,861	-4,077	-4,422	-4,327
WSH	-4,152	-4,040	-3,822	-4,135	-4,034	-3,782	-3,991	-4,020	-3,986
<i>MoyBC</i>	23,241	19,616	20,909	23,527	25,012	18,823	20,628	25,612	21,69

TABLE B.11 – Paramètres estimés pour le troisième modèle avec l’asso. de l’Ouest.

Variable	2013-2014			2014-2015			2015-2016		
	L	E	T	L	E	T	L	E	T
ANA	-4,290	-3,750	-3,802	-4,684	-4,156	-3,892	-4,588	-4,212	-4,314
ARI	-4,570	-4,011	-3,913	-4,264	-4,306	-4,234	-4,313	-3,957	-3,963
CGY	-5,209	-4,088	-4,090	-4,341	-4,169	-3,844	-4,373	-4,160	-4,380
CHI	-3,985	-3,891	-3,737	-3,724	-4,130	-4,082	-4,000	-3,956	-4,139
COL	-4,376	-4,186	-3,906	-4,338	-4,415	-4,346	-3,904	-3,686	-4,201
DAL	-4,200	-3,788	-4,023	-4,105	-3,955	-3,865	-4,099	-3,763	-4,202
EDM	-4,066	-4,125	-4,457	-4,877	-4,321	-4,126	-4,654	-4,522	-4,176
LAK	-4,096	-4,680	-3,936	-4,088	-3,985	-4,004	-3,998	-3,912	-4,577
MIN	-4,251	-4,224	-4,246	-4,347	-4,191	-3,729	-4,118	-4,052	-3,726
NSH	-4,085	-3,935	-4,638	-4,151	-4,188	-3,683	-4,233	-3,943	-4,558
SJS	-4,289	-3,906	-3,950	-4,330	-3,810	-4,000	-3,690	-4,427	-4,235
STL	-4,167	-3,826	-3,780	-3,808	-4,290	-3,905	-4,311	-3,974	-4,015
VAN	-4,794	-4,211	-4,358	-4,223	-3,851	-4,178	-4,056	-4,441	-4,547
WPG	-4,524	-4,667	-3,710	-4,316	-3,887	-4,658	-4,032	-4,144	-4,409
<i>MoyBC</i>	24,756	20,249	19,713	23,056	21,812	20,105	22,439	19,957	23,947

TABLE B.12 – Suite des paramètres estimés pour le quatrième modèle avec l’asso. de l’Est.

Variable	2013-2014			2014-2015			2015-2016		
	L	E	T	L	E	T	L	E	T
<i>CAR</i> ₅₅	-4,648	-4,081	-3,939	-4,966	-4,483	-4,439	-4,758	-4,413	-4,271
<i>CBJ</i> ₅₅	-4,181	-3,945	-4,131	-4,391	-4,225	-4,004	-4,451	-4,755	-4,039
<i>NJD</i> ₅₅	-4,680	-3,928	-4,100	-4,435	-4,120	-4,316	-3,996	-4,894	-4,094
<i>NYI</i> ₅₅	-4,490	-4,275	-3,833	-4,352	-3,823	-3,900	-4,712	-4,394	-4,094
<i>NYR</i> ₅₅	-4,276	-4,305	-3,998	-4,225	-4,134	-3,765	-4,290	-4,329	-4,489
<i>PHI</i> ₅₅	-4,212	-4,186	-4,316	-4,147	-4,600	-3,843	-5,224	-4,165	-4,351
<i>PIT</i> ₅₅	-4,245	-3,832	-3,946	-4,378	-4,020	-4,021	-4,343	-4,417	-4,280
<i>WSH</i> ₅₅	-4,308	-3,972	-4,006	-4,338	-4,030	-3,701	-4,202	-4,049	-3,891
<i>MoyBC</i> ₅₅	27,680	19,614	21,133	28,875	24,586	18,440	27,537	26,697	21,434
<i>Origine</i> ₅	-3,388	-4,472	-3,993	-3,493	-5,785	-3,862	-3,536	-4,841	-3,851
<i>MoyBC</i> ₅	9,600	17,195	14,898	10,097	35,050	13,437	10,159	21,154	12,860

TABLE B.13 – Paramètres estimés pour le quatrième modèle avec l’asso. de l’Ouest.

Variable	2013-2014			2014-2015			2015-2016		
	L	E	T	L	E	T	L	E	T
<i>ANA</i> ₅₅	-4,276	-3,788	-4,015	-4,823	-4,081	-3,986	-4,642	-4,163	-4,289
<i>ARI</i> ₅₅	-4,555	-3,926	-4,014	-4,364	-4,244	-4,161	-4,454	-3,884	-4,112
<i>CGY</i> ₅₅	-5,241	-4,252	-4,192	-4,514	-4,133	-3,916	-4,530	-4,155	-4,378
<i>CHI</i> ₅₅	-4,058	-3,909	-3,754	-3,941	-4,149	-4,029	-4,383	-3,897	-4,155
<i>COL</i> ₅₅	-4,452	-4,214	-4,024	-4,324	-4,343	-4,296	-4,281	-3,648	-4,395
<i>DAL</i> ₅₅	-4,320	-3,764	-4,041	-4,223	-3,988	-3,717	-4,486	-3,770	-4,231
<i>EDM</i> ₅₅	-4,485	-4,159	-4,426	-5,114	-4,240	-4,024	-4,798	-4,482	-4,234
<i>LAK</i> ₅₅	-4,166	-4,669	-4,182	-4,069	-3,972	-4,106	-4,204	-3,981	-4,431
<i>MIN</i> ₅₅	-4,342	-4,296	-4,483	-4,376	-4,111	-3,699	-4,138	-4,011	-3,728
<i>NSH</i> ₅₅	-4,080	-4,060	-4,556	-4,221	-4,183	-3,592	-4,331	-3,903	-4,586
<i>SJS</i> ₅₅	-4,358	-3,867	-4,186	-4,477	-3,800	-3,917	-4,058	-4,509	-4,173
<i>STL</i> ₅₅	-4,094	-3,874	-3,912	-3,863	-4,251	-3,771	-4,488	-3,952	-4,043
<i>VAN</i> ₅₅	-4,957	-4,162	-4,358	-4,444	-3,817	-4,046	-4,415	-4,344	-4,685
<i>WPG</i> ₅₅	-4,685	-4,678	-3,834	-4,363	-3,823	-4,735	-4,403	-4,204	-4,331
<i>MoyBC</i> ₅₅	26,227	20,686	21,806	25,063	21,079	19,122	28,216	19,535	23,965
<i>Origine</i> ₅	-3,780	-5,807	-3,521	-3,985	-8,003	-4,241	-2,989	-4,940	-4,759
<i>MoyBC</i> ₅	13,224	33,306	8,835	14,148	69,602	17,730	6,167	17,106	22,739

Annexe C

Affrontements exclus des paris intelligents

Voici la liste des éléments exclus à la section 4.4.1 pour chacune des divisions-saisons reliée à ce qui est mentionné à la section 4.3.

- Atl. 15-16 : Montreal Canadiens et *Alex Stalock*.
- Atl. 14-15 : Toronto Maple Leafs, Ottawa Senators, *Chad Johnson* et *Anders Lindback*.
- Atl. 13-14 : *Michal Neuvirth*, *Dan Ellis*, *Roberto Luongo*, *Jaroslav Halak*.
- Mét. 15-16 : Pittsburgh Penguins.
- Mét. 14-15 : New Jersey Devils et *Michal Neuvirth*.
- Mét. 13-14 : *Jaroslav Halak*.
- Cen. 15-16 : Minnesota Wild et *Anders Nilsson*.
- Cen. 14-15 : *Jhonas Enroth* et *Devan Dubnyk*.
- Cen. 13-14 : Winnipeg Jets , *Tim Thomas*, *Ilya Bryzgalov*, *Ryan Miller*, *Devan Dubnyk*.
- Pac. 15-16 : *James Reimer*.
- Pac. 14-15 : Edmonton Oilers et *Evgeni Nabokov*.
- Pac. 13-14 : *Jacob Markstrom*, *Viktor Fasth* et *Ben Scrivens*.

Bibliographie

- Appleby, M. 2016, «How x/y tracking data is a game-changer for NHL analytics», Sports Inc.
- Baio, G. et M. Blangiardo. 2010, «Bayesian hierarchical model for the prediction of football results», *Journal of Applied Statistics*, vol. 37, n° 2, p. 253–264.
- Braunstein, A. 2010, «Consistency and Pythagoras», *Journal of Quantitative Analysis in Sports*, vol. 6, n° 1. article 8.
- Cameron, A. C. et P. K. Trivedi. 2005, *Microeconometrics : Methods and Applications*, Cambridge University Press, New York.
- Caumel, Y. 2015, *Probabilités et processus stochastiques*, 2^e éd., Hermès-Lavoisier.
- Cochran, J. J. et R. Blackstock. 2009, «Pythagoras and the National Hockey League», *Journal of Quantitative Analysis in Sports*, vol. 5, n° 2. article 11.
- Crowder, M., M. Dixon, A. Ledford et M. Robinson. 2002, «Dynamic modelling and prediction of English Football League matches for betting», *The Statistician*, vol. 51, n° 2, p. 157–168.
- Dayaratna, K. D. et S. J. Miller. 2013, «The Pythagorean Won-Loss Formula and Hockey - A Statistical Justification for Using the Classic Baseball Formula as an Evaluative Tool in Hockey», *The Hockey Research Journal 2012/2013*, vol. 1, p. 193–209.
- Eaton, C. 2014, «Opinion - Legalise Sports Betting Worldwide», cahier de recherche, International Centre for Sport Security.
- Goddard, J. et I. Asimakopoulos. 2001, «Modelling football match results and the efficiency of fixed-odds betting», cahier de recherche, Department of Economics, University of Wales Swansea.
- Gramacy, R. B., S. T. Jensen et M. Taddy. 2013, «Estimating Player Contribution in Hockey with Regularized Logistic Regression», cahier de recherche, University of Chicago Booth School of Business and The Wharton School, University of Pennsylvania.
- Henwood, D. 2015, «Top 20 sports betting nations 2014», *Sports Betting*, vol. 8, p. 2. H2 Gambling Capital.

- Hines, W. W., D. C. Montgomery, D. M. Goldsman et C. M. Borrer. 2005, *Probabilités et statistique pour ingénieurs*, Les Éditions de la Chenelière inc. Montréal, Québec.
- James, B. 1980, «The Bill James Abstract», *Auto publié*.
- Macdonald, B. 2012, «Adjusted Plus-Minus for NHL Players using Ridge Regression with Goals, Shots, Fenwick, and Corsi», *Journal of Quantitative Analysis in Sports*, vol. 8, n° 3, p. 1515–1559.
- McCullagh, P. et J. Nelder. 1989, *Generalized Linear Models*, 2^e éd., Chapman and Hall.
- Merritt, S. et A. Clauset. 2014, «Scoring Dynamics Across Professional Team Sports : Tempo, Balance and Predictability», cahier de recherche, Department of Computer Science, University of Colorado, BioFrontiers Institute, University of Colorado and Santa Fe Institute.
- Miller, S. J. 2007, «A Derivation of the Pythagorean Won-Loss Formula in Baseball», cahier de recherche, Department of Mathematics, Brown University.
- Moretti, C. 2016, «Empowering analytics through data capture», HockeyTech.
- von Neumann, J. et O. Morgenstern. 1944, *Theory of Games and Economic Behavior*, Princeton University Press. Princeton, New Jersey.
- Pohlkamp-Hartt, J. 2016, «A Statistical Framework for Hockey Analytics : Advice and observations from time my working with the Kingston Frontenacs», Queen’s University.
- Polson, N. G. et H. S. Stern. 2015, «The implied volatility of a sports game», *Journal of Quantitative Analysis in Sports*, vol. 11, n° 3, p. 145–153.
- Ruiz, F. J. et F. Perez-Cruz. 2015, «A generative model for predicting outcomes in college basketball», *Journal of Quantitative Analysis in Sports*, vol. 11, n° 1, p. 39–52.
- Schuckers, M. et J. Curro. 2013, «Total Hockey Rating (THoR) : A comprehensive statistical rating of National Hockey League forwards and defensemen based upon all on-ice events», cahier de recherche, MIT Sloan Sports Analytics Conference 2013.
- Schuckers, M. et B. Macdonald. 2014, «Accounting for Rink Effects in the National Hockey League’s Real Time Scoring System», cahier de recherche, St.Lawrence University and United States Military Academy.
- Schumaker, R. P., O. K. Solieman et H. Chen. 2010, *Sports Data Mining*, vol. 26, Springer. New York.
- Shi, Z., S. Moorthy et A. Zimmermann. 2013, «Predicting NCAAB match outcomes using ML techniques – some results and lessons learned», cahier de recherche, KU Leuven, Belgique.

- Stern, H. S. 1994, «A Brownian Motion Model for the Progress of Sports Scores», *Journal of the American Statistical Association*, vol. 89, n° 427, p. 1128–1134.
- Thomas, A., S. L. Ventura, S. Jensen et S. Ma. 2013, «Competing Process Hazard Function Models for Player Ratings in Ice Hockey», *The Annals of Applied Statistics*, vol. 7, n° 3, p. 1497–1524.
- Weissbock, J. 2014, *Forecasting Success in the National Hockey League using In-Game Statistics and Textual Data*, mémoire de maîtrise, University of Ottawa.
- Wolfson, J. et J. S. Koopmeiners. 2015, «Who’s Good This Year? Comparing the Information Content of Games in the Four Major US Sports», cahier de recherche, Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota.
- Zimmermann, A. 2015, «Exploring chance in NCAA basketball», cahier de recherche, INSA, Lyon.