





HEC MONTRÉAL

Exploration des méthodes de l'apprentissage statistique: une application à la  
prévision de la demande et du prix de l'électricité

par

Amélie Rousseau

Sciences de la gestion  
(Option Économie appliquée)

*Mémoire présenté en vue de l'obtention  
du grade de maîtrise ès sciences  
(M. Sc.)*

Juin 2018

© Amélie Rousseau, 2018



# Résumé

Les analyses prédictives de court terme de la demande et du prix dans les marchés de l'électricité représentent, de nos jours, des éléments cruciaux dans l'organisation des opérations quotidiennes du marché, ainsi que dans le processus décisionnel de ses intervenants. Dans ce mémoire, nous explorons la performance de différentes méthodes tirées de l'apprentissage statistique, appliquées au contexte prédictif de la demande et des prix de l'électricité dans le marché de l'Ontario. Le marché ontarien est unique sur plusieurs aspects, dont sa nature en temps réel, complexifiant la tâche d'y développer des outils de prévisions de qualité.

Les méthodes analysées comprennent les régressions pénalisées, la régression sur composantes principales, les arbres de décision, ainsi que les méthodes d'ensemble aléatoires et additives. Les méthodes tirées de l'apprentissage statistique sont également comparées à deux méthodes conventionnelles de prévision, soit la régression linéaire multiple et la régression linéaire avec erreur ARIMA. Les modèles prédictifs de la demande et du prix sont estimés sur les horizons de prévision de 1 heure et 3 heures, respectivement. La performance prédictive hors échantillon de chaque modèle est évaluée à partir de différents indicateurs d'ajustement.

Les résultats suggèrent que les modèles basés sur les méthodes de l'apprentissage statistique engendrent, de manière générale, de meilleures prévisions de la demande d'électricité de la prochaine heure en Ontario. En particulier, les modèles de régressions pénalisées présentent les taux d'erreur de prévision les plus faibles à travers la période analysée. Pour la prévision du prix de l'électricité sur l'horizon temporel de 3 heures, toutefois, l'utilisation des techniques de l'apprentissage statistique ne semble pas présenter un avantage significatif, vis-à-vis des techniques conventionnelles de prévision.

# Abstract

Short-term predictive analyses of demand and price in electricity markets are critical nowadays in the organization of their daily operations and the intervening parties' decision processes. This thesis evaluates the performance of several methods from machine learning when applied to the prediction of electricity demand and prices in the Ontario market. This is a spot energy market, which makes it unique and challenging to develop proper forecasting tools.

We analyze methods such as penalized regression, principal components analysis, decision trees, and ensemble methods (bagging, random forest, boosting). We compare the machine learning methods against two more conventional forecasting methods : multiple linear regression and linear regression with ARIMA errors. We estimate the predictive models of demand and price for a forecast horizon of one and three hours, respectively and evaluate the out-of-sample predictive performance of each model using measures of prediction accuracy.

The results suggest that machine learning models usually generate better electricity demand predictions for the next hour in Ontario. The penalized regression models have the lowest error rate throughout the time period studied. However, machine learning techniques do not seem to show a significant advantage in forecasting electricity prices during the three-hour time horizon comparatively to traditional predictive tools.

# Table des matières

Résumé	i
Abstract	ii
Liste des tableaux	v
Liste des figures	vii
Remerciements	ix
Introduction	1
Revue de la littérature	6
2.1 Contexte . . . . .	6
2.2 Estimation et modélisation . . . . .	7
<b>3 Données</b>	<b>16</b>
3.1 Le marché d'électricité de l'Ontario . . . . .	16
3.2 Données . . . . .	19
<b>4 Modèles</b>	<b>26</b>
4.1 Modèles conventionnels . . . . .	26
4.2 Modèles de l'apprentissage statistique . . . . .	28
<b>5 Estimation et résultats</b>	<b>42</b>
5.1 Préparation des échantillons . . . . .	42

5.2	Détails des estimations . . . . .	45
5.3	Comparaison des performances prédictives . . . . .	57
	<b>Conclusion</b>	<b>76</b>
	<b>Bibliographie</b>	<b>79</b>
	<b>Annexe A – Résultats des estimations</b>	<b>86</b>
	<b>Annexe B – Représentation des arbres de régression</b>	<b>87</b>
	<b>Annexe C – Importance des prédicteurs dans les modèles prédictifs optimaux de la demande d’électricité</b>	<b>88</b>
	<b>Annexe D – Modèles prédictifs optimaux de la demande d’électricité sur différents horizons temporels de prévision</b>	<b>89</b>
	<b>Annexe E – Résultats des modèles de régressions linéaires multiples pour l’analyse prédictive des prix de l’électricité en Ontario</b>	<b>93</b>
	<b>Annexe F – Résultats des modèles prédictifs des prix sur les données transformés</b>	<b>94</b>

# Liste des tableaux

3.1	Moyennes des quantités demandées totales, des prix et des importations nettes du marché ontarien entre 2010 et 2015. . . . .	19
3.2	HDD et CDD (en degré Fahrenheit) moyens à travers les régions étudiées, classées selon les périodes. . . . .	24
3.3	Les quatre périodes de la journée selon la classification de l'IESO. . . . .	25
5.1	Quelques statistiques descriptives des échantillons apprentissages et test. . . . .	45
5.2	Paramétrisation des modèles étudiés. . . . .	50
5.3	Les quatre modélisations estimées à partir des méthodes Ridge, LASSO et Elastic Net. . . . .	51
5.4	Résultats des méthodes Ridge, LASSO et Elastic Net à travers les quatre modélisations étudiées. . . . .	52
5.5	Nombre de composantes principales ( $M$ ) sélectionnées à travers les quatre modélisations étudiées. . . . .	54
5.6	Résumé des résultats des paramétrisations des modèles bagging, forêts aléatoires et boosting pour les prévisions de la demande et des prix de l'électricité. . . . .	56
5.7	RMSE et MAPE des modèles étudiés pour les prévisions de la demande d'électricité de la prochaine heure dans le marché de l'Ontario. . . . .	60
5.8	Prédicteurs non sélectionnés par la méthode Elastic Net dans la modélisation optimale M3. . . . .	63
5.9	RMSE et MAE des modèles étudiés pour les prévisions du prix de l'électricité sur un horizon temporel de 3 heures dans le marché de l'Ontario. . . . .	70

5.10	Classement des trois modèles prédictifs étudiés démontrant les meilleures performances prédictives en termes de minimisation de la RMSE et du MAE, pour les prévisions du prix de l'électricité sur un horizon temporel de trois heures dans le marché de l'Ontario calculées à partir des données transformées par moyenne mobile avec fenêtre glissante de taille 24 (MA(24)). . . . .	73
1	Résultats des modèles de régressions linéaires multiples estimés par MCGF (modèles optimaux) pour la prévision du prix de l'électricité en Ontario sur l'horizon temporel de 3 heures. . . . .	93
2	RMSE et MAE des modèles prédictifs étudiés pour les prévisions du prix de l'électricité sur un horizon temporel de 3 heures dans le marché de l'Ontario calculées à partir des données transformées par moyenne mobile avec fenêtre glissante de taille 24 (MA(24)). . . . .	94

# Liste des figures

3.1	Calendrier des opérations quotidiennes du marché de l'électricité en Ontario . . .	17
3.2	Évolution du prix moyen en Ontario entre 2010 et 2015. . . . .	21
3.3	Relations entre la température (en degré Fahrenheit) des régions étudiées et la demande totale ontarienne. . . . .	23
4.1	Représentation de la pénalité Lasso (gauche) et la pénalité Ridge (droite). . . .	31
4.2	Représentation des deux premières composantes principales d'un nuage de points.	32
4.3	Représentation d'un arbre de régression de type CART. . . . .	34
5.1	Période "avant" - Comparaison des valeurs réelles et des valeurs prédites par différents estimateurs étudiés pour l'analyse prédictive de la demande d'électricité.	61
5.2	Période "après" - Comparaison des valeurs réelles et des valeurs prédites par différents estimateurs étudiés pour l'analyse prédictive de la demande d'électricité.	62
5.3	Sélection de variables par la méthode Elastic Net à travers 10 sous échantillons de l'échantillon d'apprentissage initial. . . . .	64
5.4	Classement des 20 prédicteurs dont la taille absolue des coefficients est la plus importante dans les modèles optimaux de prévision de la demande d'électricité de la prochaine heure des périodes "avant" et "après". . . . .	65
5.5	Comparaison des valeurs réelles et des valeurs prédites par le modèle Ridge (M2) sur différents horizons de temps. . . . .	68
5.6	Comparaison des valeurs réelles et des valeurs prédites par les modèles de régressions linéaires multiples du prix de l'électricité. . . . .	71

5.7	Comparaison des valeurs réelles et des valeurs prédites par les modèles optimaux de prévision du prix de l'électricité estimées à partir des données transformées par moyenne mobile avec fenêtre glissante de taille 24. . . . .	74
1	Résultats des arbres émondés. . . . .	87
2	Classement des 20 prédicteurs dont la taille absolue des coefficients est la plus importante dans les modèles optimaux de prévision de la demande d'électricité de la prochaine heure des périodes "avant" et "après". . . . .	88
3	Comparaison des valeurs réelles et des valeurs prédites par le modèle Elastic Net (M2) sur différents horizons de temps. . . . .	90
4	Comparaison des valeurs réelles et des valeurs prédites par le modèle Elastic Net (M2) sur différents horizons de temps. . . . .	92

# Remerciements

Tout d’abord, je tiens à remercier mon directeur de mémoire, Monsieur Mario Samano. Je suis reconnaissante de la générosité, de la rigueur et de l’enthousiasme unique dont vous avez faite preuve tout au long de ce projet de recherche, ainsi que de votre accessibilité et de vos nombreux conseils qui m’auront été grandement utiles. Malgré ce que j’aurai pu dire à quelques reprises, la réalisation de ce mémoire aura été une expérience agréable, et ce, en grande partie grâce à vous.

J’aimerais également remercier mes collègues et amis de la maîtrise. Comme le disait si bien Marcel Proust : “Soyons reconnaissant aux personnes qui nous donnent du bonheur ; elles sont les charmants jardiniers par qui nos âmes sont fleuries”. Merci, mes chers collègues et amis, d’avoir maintenu mon âme florissante dans les moments plus difficiles. Plus particulièrement, je souhaite remercier mon amie Céline qui, par sa bonne humeur, son optimisme et sa patience, a su égayer mes deux années de maîtrise à HEC Montréal. Merci de m’avoir soutenu et encouragé dans ce long processus qui n’aurait certainement pas été le même sans toi.

Enfin, je remercie ma famille et mes proches pour leur soutien indéfectible durant l’ensemble de mon parcours universitaire. À mes parents, Sabrina, Jean-Baptiste et Anikke : merci ! Je suis chanceuse de votre présence, de vos encouragements et de votre écoute au quotidien.



# Introduction

On assiste, depuis la fin des années 1980, à une remise en question des modes d'organisation traditionnels dans l'industrie de l'électricité. Une planification centralisée de l'industrie a d'abord été jugée nécessaire afin de garantir l'efficacité de la production et la sécurité d'approvisionnement (Weron, 2007). Le secteur électrique était ainsi caractérisé par une structure de marché intégrée verticalement et un niveau de concurrence peu élevé, notamment en raison des importants coûts nécessaires à l'installation des technologies de production. La concentration monopolistique est toutefois susceptible d'engendrer des problèmes d'abus de pouvoir de marché dont l'impact, dans le cas de l'industrie de l'électricité, est renforcé par l'importance de l'électricité dans les sociétés modernes. En effet, l'électricité représente de nos jours une nécessité pour la majorité des agents économiques. De plus, sa production nécessite des ressources ayant de lourdes conséquences sur l'environnement naturel. C'est pourquoi, dans les deux dernières décennies, d'importants changements de structure sont survenus dans l'industrie de l'électricité. En effet, plusieurs gouvernements ont opté pour la déréglementation de leur secteur électrique : la situation monopolistique a été remplacée par un dégroupement des activités de l'industrie et l'introduction de marchés compétitifs de production, fonctionnant généralement par systèmes d'enchères.

La restructuration de l'industrie de l'électricité accentue l'importance des analyses prédictives de la demande et des prix, notamment celles de court terme. En effet, l'introduction de marchés compétitifs de production implique l'intervention de différents agents dans le processus d'organisation des opérations quotidiennes, autrefois centralisé. Dans un environnement réglementé, l'analyse prédictive de la demande n'est utile qu'à l'entreprise détenant le monopole : à court terme, basée sur les observations passées, elle peut facilement prédire l'ap-

provisionnement suffisant en électricité, évitant ainsi les coûts associés à une surproduction (Weron, 2007). Le prix, quant à lui, est réglementé par les autorités. Dans un environnement compétitif, toutefois, les prévisions de court terme de la demande et des prix sont utiles à tous les agents du marché, producteurs comme régulateur. Bunn (2000) observe trois implications de l'introduction des marchés compétitifs sur l'analyse prédictive de court terme de la demande et des prix. D'abord, les marchés qui fonctionnent par un système d'enchères sont basés sur des prévisions horaires disponibles une journée à l'avance. L'analyse prédictive de la demande d'électricité devient alors l'activité décisive du bon fonctionnement du marché. Également, puisqu'il n'est pas économiquement rentable d'emmagasiner l'électricité produite, les coûts associés à une surproduction ou une sous-production sont élevés pour les producteurs. Anticiper de manière précise les demandes futures d'électricité représente ainsi une nécessité pour la maximisation de leur profit. Enfin, l'environnement compétitif amène à considérer l'analyse prédictive de court terme des prix dans l'élaboration des décisions, notamment en raison de la forte volatilité des prix dans les marchés de l'électricité. Pour les producteurs, une connaissance préalable des fluctuations des prix est une composante importante dans le développement de stratégies d'enchères optimales (Vilar et al., 2012).

L'importance des prévisions de court terme dans les marchés de l'électricité se reflète dans la littérature, où un grand nombre de techniques ont été développées à travers les années. Les approches conventionnelles comprennent les méthodes de séries temporelles et les méthodes de régression (Hahn et al., 2009). Entre autres, les processus ARIMA et GARCH, ainsi que le modèle de régression linéaire multiple ont été employés pour la prévision de court terme de la demande et des prix de l'électricité (Weron, 2007). Ces méthodes ont les avantages d'être simples à calculer et faciles à interpréter. Cependant, elles fournissent généralement des résultats satisfaisants seulement sous la condition que la fréquence des données soit faible (Kyriakides et Polycarpou, 2007). Également, elles peuvent souffrir d'instabilités numériques, notamment en présence d'un nombre élevé de données. Ce dernier désavantage devient particulièrement important de nos jours, dans "l'ère du Big Data", puisque les organisations se voient généralement obliger à recueillir et à traiter un maximum de données afin de demeurer compétitives (Ludwig et al., 2015).

Dans ce nouveau contexte, les méthodes tirées de l'apprentissage statistique se révèlent des outils utiles et efficaces, et ce dans différents domaines : science, santé et finance. Les méthodes de l'apprentissage statistique sont généralement employées pour les prévisions de court terme de la demande et des prix de l'électricité sous la forme de modèles de logique floue (*fuzzy logic*), de systèmes experts et de perceptrons multicouches (réseaux de neurones) (Weiron, 2007). Bien que ces techniques soient avantagées par leur capacité d'estimer des relations non linéaires complexes, leur modélisation est généralement considérée comme compliquée, puisqu'elle nécessite de nombreuses paramétrisations qui influencent leur performance prédictive, et leurs résultats sont difficilement interprétables (Dudek, 2015). Pour ces raisons, leur utilisation n'est pas reconnue par l'ensemble des chercheurs du domaine. Il ne semble donc pas qu'un consensus soit établi sur l'optimalité d'un modèle pour le calcul des prévisions de court terme de la demande et des prix de l'électricité, malgré les développements observés dans les dernières années.

Les développements considérables réalisés dans les méthodes de l'apprentissage statistique ont grandement fait progresser les techniques d'analyse prédictive dans les marchés de l'électricité. Ce mémoire s'inscrit dans ce nouveau courant d'analyse. Notre objectif consiste à explorer la performance de différentes méthodes tirées de l'apprentissage statistique, appliquées dans le contexte prédictif de la demande et des prix de l'électricité. En particulier, notre recherche se concentre sur les prévisions de la demande et des prix, sur les horizons temporels de 1 et 3 heures respectivement, dans le marché de l'électricité de l'Ontario. Le choix d'y étudier l'Ontario repose, d'une part, par la nécessité actuelle d'y développer des outils efficaces de prévisions de court terme. En effet, la province ayant misé davantage dans les dernières années sur les ressources énergétiques décentralisées et les énergies renouvelables, dont la production est variable, a vu l'exploitation de son réseau électrique au quotidien devenir de plus en plus complexe (IESO, 2018c). D'autre part, le réseau électrique de l'Ontario est étroitement lié à celui du Québec : les deux provinces s'échangent quotidiennement une quantité importante d'électricité, et leurs prix s'influencent mutuellement. En 2016, le Québec et l'Ontario ont d'ailleurs signé une entente historique de commerce de l'électricité autorisant entre autres "la vente d'électricité, la réservation de puissance électrique et la

mise à la disposition de capacité de stockage” (Gouvernement de l’Ontario, 2016).

Les techniques étudiées dans le cadre de ce mémoire ne se limitent toutefois pas à l’étude du marché de l’électricité de l’Ontario. Une analyse similaire pourrait être appliquée au contexte québécois, dans le cas d’une libéralisation du marché de la production au Québec, puisqu’encore aujourd’hui les prix et la production de l’électricité sont contrôlés par la Régie de l’Énergie (Régie de l’Énergie, 2018). Également, des méthodes similaires à celles étudiées dans cette recherche peuvent s’appliquer à différents contextes, dont l’analyse prédictive dans les marchés financiers, les prévisions des comportements des consommateurs, ou encore les calculs prédictifs dans les secteurs de la santé.

Les méthodes tirées de l’apprentissage statistique étudiées dans le cadre de ce mémoire sont les suivantes : la régression Ridge (Hoerl et Kennard, 1970), LASSO (*Least Absolute Shrinkage and Selection Operator*) (Tibshirani, 1996) et Elastic Net (Zou et Hastie, 2005), la régression sur composantes principales (PCA), les arbres de décision (Breiman et al., 1984) et les techniques d’ensemble aléatoires et additives (le bagging (Breiman, 1996), les forêts aléatoires (Breiman, 2001) et le boosting (Friedman, 2001)). Celles-ci ont été sélectionnées en raison de leur simplicité et le fait qu’elles nécessitent peu de paramétrisations. Notre objectif consiste ainsi à développer des modèles pouvant être facilement utilisés par les différents agents du marché de l’électricité de l’Ontario dans leur modélisation d’affaires et leur prise de décisions. Deux méthodes conventionnelles de prévision : la régression linéaire multiple et la régression linéaire avec erreur ARIMA, ont également été développés et sont utilisées à titre de *benchmark* dans notre démarche de comparaison.

Les résultats obtenus suggèrent que, de manière générale, les méthodes étudiées tirées de l’apprentissage statistique offrent de meilleures prévisions de la demande d’électricité ontarienne sur l’horizon temporel de 1 heure, en comparaison aux méthodes conventionnelles de prévision. En particulier, les régressions Ridge et Elastic Net présentent les meilleures performances prédictives parmi les techniques analysées. Les résultats obtenus montrent également que, pour le calcul de la prévision du prix de l’électricité sur l’horizon temporel de 3 heures, l’utilisation des techniques de l’apprentissage statistique ne présente pas un

avantage significatif, relativement à l'utilisation des méthodes prédictives conventionnelles.

Notre recherche s'organise en 4 sections. La section 1 présente une revue de la littérature, dont le but sera d'identifier les éléments clés et les différents courants de l'analyse prédictive dans les marchés de l'électricité. La section 2 présente le marché de l'électricité en Ontario, ainsi que les données nécessaires à l'application des modèles théoriques de prévision de court terme de la demande et des prix. La section 3 définit en détail les modèles utilisés dans le cadre de notre recherche. Enfin, la section 4 fait état des résultats. On y présente les démarches de paramétrisation et d'estimation des modèles étudiés. Un lien vers le code d'analyse numérique construit à partir du logiciel R est également disponible dans l'annexe 1 pour le lecteur intéressé. La section se conclut sur la comparaison des performances prédictives des méthodes analysées et les conclusions qui en découlent.

# Revue de la littérature

La présente section propose d'effectuer l'étude de la littérature entourant les analyses prédictives dans les marchés de l'électricité. Le contexte sera d'abord brièvement discuté. Puis, les stratégies d'estimation seront présentées, ainsi que les différentes approches de modélisations utilisées pour les prévisions de court terme de la demande et du prix de l'électricité. Les éléments abordés seront notamment utiles dans la détermination des modèles à employer et dans le développement de notre méthodologie.

## 2.1 Contexte

Les analyses prédictives de court terme dans les marchés compétitifs de production de l'électricité cherchent à faire des prévisions sur des horizons de temps d'une heure à une semaine. Elles constituent des problèmes complexes, notamment en raison des caractéristiques distinctives de la demande et du prix horaires de l'électricité (Vilar et al., 2012). En effet, les séries temporelles de la demande et du prix sont d'abord caractérisées par une non-stationnarité stochastique, leur variance n'étant pas indépendante du temps. En outre, les données des prix, comparativement aux données de la demande, exhibent généralement une forte volatilité. Les données montrent de manière générale une proportion élevée de prix inhabituels (*price spike*), causés par des événements non anticipés sur le marché (Weron, 2006). Enfin, la demande et le prix de l'électricité peuvent être influencés par différents facteurs externes. En particulier, les séries temporelles de la demande et du prix comportent généralement trois types de saisonnalité : un cycle quotidien, hebdomadaire et annuel (Hahn et al., 2009). La demande d'électricité est également influencée à court terme par des facteurs cli-

matiques, dont la température ; cette relation étant généralement non linéaire (Hippert et al., 2001). Le prix, quant à lui, est majoritairement influencé par la demande d'électricité. La forme de la relation entre le prix et la demande n'est toutefois pas clairement identifiée dans la littérature. Par exemple, Vucetic et al. (2001) montrent, dans leur analyse du marché de l'électricité de la Californie, qu'il peut exister plusieurs régimes de comportement des prix, et que l'influence de la demande sur les prix varie à travers ces régimes.

## **2.2 Estimation et modélisation**

Les spécificités des séries temporelles de la demande et du prix de l'électricité complexifient ainsi les analyses prédictives de court terme dans les marchés de l'électricité. Il est, par conséquent, nécessaire d'élaborer une stratégie d'estimation, et de développer une modélisation adéquate dans le but d'obtenir des prévisions précises de la demande et du prix.

### **2.2.1 Stratégies d'estimation**

Les calculs des prévisions de la demande et du prix s'effectuent, de manière générale, séparément. Les prévisions peuvent être estimées à partir d'un seul modèle ou de 24 différents modèles, un pour chaque heure de la journée (Weron, 2014). Dans la seconde option, on calcule ainsi les prévisions de chaque heure de manière indépendante. La première stratégie a l'avantage de développer des modèles simples, faciles à calculer et à interpréter. La deuxième stratégie bénéficie, quant à elle, du fait qu'elle permet aux variables indépendantes d'avoir un impact différent sur la variable dépendante (demande ou prix), selon l'heure de la journée (Bunn, 2000). Son implémentation étant plus complexe puisqu'elle nécessite un nombre supérieur de paramétrisations, la première stratégie est généralement préférée dans la littérature.

### **2.2.2 Modélisation**

De nombreux modèles ont été développés, à travers les années, pour le calcul des prévisions de court terme de la demande et du prix de l'électricité. On les classe généralement

selon deux catégories : les approches conventionnelles et les approches modernes.

## Les approches conventionnelles

Les approches conventionnelles comprennent les méthodes de séries temporelles et les méthodes de régression (Hahn et al., 2009). Se basant sur l'hypothèse que les données possèdent une structure interne stationnaire, les techniques de séries temporelles modélisent les prévisions horaires d'une variable dépendante (demande ou prix) comme une fonction, généralement linéaire, de ses données historiques. Certains auteurs font également le choix d'inclure des variables explicatives exogènes dans leur modélisation. Entre autres, Pardo et al. (2002) concluent, à partir de leur étude du marché de l'électricité en Espagne, qu'inclure des valeurs présentes et passées de la température permet d'obtenir un modèle à pouvoir prédictif élevé pour les prévisions de la demande d'électricité. Les modèles prédictifs des prix, quant à eux, incluent fréquemment les valeurs présentes et passées de la demande d'électricité (Aggarwal et al., 2009).

Les séries temporelles de la demande d'électricité ont généralement été représentées par un processus autorégressif à moyenne mobile (ARMA ou ARIMA) dans la littérature. Le processus ARIMA modélise la variable dépendante comme une combinaison linéaire de ses valeurs passées et des valeurs présentes et passées d'un terme d'erreur bruit blanc, c'est-à-dire un terme à moyenne nulle et à variance constante. Tel que mentionné par Taylor et al. (2006), le modèle ARIMA univarié est souvent utilisé comme *benchmark* dans la comparaison de la performance prédictive de différentes techniques de prévision de court terme de la demande d'électricité. Dans le cadre de ce mémoire, toutefois, l'utilisation d'une modélisation multivariée lui sera préférée, afin de capter l'impact important des facteurs climatiques dans la prévision de court terme de la demande d'électricité en Ontario.

Les modèles ARIMA univariés ont également été utilisés dans le cadre de l'analyse prédictive de court terme des prix de l'électricité. Entre autres, Contreras et al. (2003) ont développé des modèles ARIMA univariés pour effectuer les prévisions horaires des prix de l'électricité en Espagne et en Californie. Les auteurs concluent que les modèles étudiés présentent des taux d'erreurs raisonnables, considérant la structure complexe des séries temporelles des prix

de l'électricité. Nogales et al. (2002) étudient également les marchés de l'électricité de l'Espagne et de la Californie à partir, toutefois, de modèles de séries temporelles multivariés. En particulier, deux types de modèles multivariés, incluant la demande comme variable explicative, sont utilisés dans le calcul des prévisions horaires des prix : un modèle de régression dynamique (RD) et un modèle de fonction de transfert (*transfer function model*) (FT). L'approche par RD modélise le prix de l'électricité comme une fonction de ses valeurs passées, des valeurs passées et présente de la demande d'électricité, ainsi que d'un terme d'erreur présumé bruit blanc. Une modélisation semblable est utilisée dans l'approche par FT à la différence que le terme d'erreur est décrit par un processus ARMA. Une comparaison des résultats obtenus par Contreras et al. (2003) et par Nogales et al. (2002) montre que les performances prédictives des modèles RD et FT sont généralement supérieures aux modèles ARIMA univariés. Des résultats similaires sont également obtenus par Zareipour et al. (2006), dans leur analyse du marché de l'électricité en Ontario. Les auteurs développent des modèles multivariés RD et FT pour prédire, 3 heures et 24 heures à l'avance, le prix de l'électricité sur une période de six semaines de l'année 2004. Les modèles incluent différentes variables de marché telles que les valeurs passées de la demande et des prix de l'électricité de l'état de New York, l'une des régions où est transportée l'électricité ontarienne. Les résultats obtenus montrent la supériorité, en termes de prévision, des modèles RD et FT, comparativement au modèle utilisé par l'organisme régulateur du marché de l'Ontario (IESO). Les auteurs constatent, toutefois, qu'aucun des modèles développés ne réussit à prédire les pics de prix (*price spike*) survenus dans les six semaines analysées.

En raison de la présence d'une forte volatilité dans les séries temporelles des prix, les modèles d'hétéroscédasticité conditionnelle autorégressive généralisée (GARCH) ont également été employés dans la littérature. Le modèle GARCH est une modélisation, qui varie dans le temps, de la variance conditionnelle du terme d'erreur d'un processus. La violation de l'hypothèse d'homoscédasticité permet d'accommoder le processus à la présence de volatilité dans les données. Garcia et al. (2005) étudient les performances prédictives de modèles GARCH dans les marchés de l'électricité de l'Espagne et de la Californie. En particulier, les auteurs emploient un modèle GARCH univarié, un modèle GARCH multivarié qui inclue

la demande, et un modèle ARIMA univarié pour la prévision, une journée à l'avance, des prix des 24 prochaines heures. Les résultats montrent que les performances prédictives des modèles GARCH surpassent celle du modèle ARIMA.

Les méthodes de régression estiment, quant à elles, la relation entre une variable dépendante et les valeurs présentes de facteurs externes (heure, température, jour de la semaine). La prévision est ensuite calculée à partir de cet estimé (Kyriakides et Polycarpou, 2007). Les méthodes de régression ont été fréquemment employées dans le contexte prédictif de la demande d'électricité, alors qu'on les retrouve très peu dans le contexte prédictif des prix, et lorsqu'elles y sont utilisées, elles se retrouvent généralement combinées à des techniques de séries temporelles sophistiquées (Weron, 2014). On note toutefois l'étude de Mazengia et al. (2008) qui emploie un modèle de régression linéaire multiple pour la prévision du prix de l'électricité en Ontario sur un horizon temporel d'une journée. Des résultats obtenus, il est conclut que le modèle développé offre un pouvoir prédictif raisonnable, notamment dans les périodes où les fluctuations des prix de l'électricité sont faibles. En outre, le modèle de régression linéaire multiple étudié engendre une meilleure performance prédictive sur la période analysée, en comparaison au modèle prédictif utilisé par l'organisme régulateur du marché ontarien (IESO).

Le modèle de régression linéaire multiple est également fréquemment utilisé dans l'analyse prédictive de la demande d'électricité, en dépit des nombreuses alternatives disponibles. La relation entre la demande et la température est, toutefois, généralement modélisée comme non linéaire (Weron, 2007). L'étude de Papalexopoulos et Hesterberg (1990) présente différents concepts-clés de l'utilisation des modèles de régression linéaire multiple dans la prévision de court terme de la demande d'électricité : estimation par moindres carrés généralisés, modélisation de la température par CDD (Cooling degree days) et HDD (Heating degree days), et utilisation de variables binaires pour la représentation de jours fériés. Hong et al. (2010) modélisent également la demande d'électricité horaire à partir de différents modèles de régressions linéaires multiples. En particulier, les auteurs développent neuf régressions linéaires incluant différentes variables indépendantes, polynômes, et termes d'interactions doubles et triples. En appliquant les modèles à la prévision de la demande d'électricité

horaire d'un réseau électrique des États-Unis, les auteurs montrent que l'intégration de différents termes d'interactions doubles et triples entre les variables indépendantes, ainsi qu'un polynôme de second degré de la température, permet d'améliorer de manière significative le pouvoir prédictif du modèle de régression linéaire.

Les méthodes de séries temporelles et les méthodes de régression sont avantagées par leur simplicité et leur facilité d'interprétation. En outre, même si le modèle estimé n'a pas une performance prédictive élevée en comparaison aux autres modèles étudiés, il pourrait tout de même être éventuellement utilisé pour fournir certaines informations relatives au marché de l'électricité, notamment l'existence de corrélations entre la variable dépendante et certains facteurs externes. Les méthodes de séries temporelles et de régression présentent cependant deux inconvénients importants dans un contexte prédictif (Kyriakides et Polycarpou, 2007). D'une part, les modèles ont généralement des performances prédictives élevées seulement si la fréquence des données est faible. Dans le cas où il existe des variations rapides dans les données, le modèle sera peu performant en termes de prévision. D'autre part, les approches conventionnelles peuvent souffrir d'instabilités numériques, particulièrement en présence d'un nombre élevé de données. De nos jours, toutefois, les bases de données volumineuses sont partie intégrante de nombreuses recherches, dont les analyses prédictives dans les marchés de l'électricité.

## **Les approches modernes**

Les bases de données volumineuses nécessitent généralement l'utilisation d'outils différents (Varian, 2014). En effet, l'ampleur des données nécessite des outils qui permettent une manipulation efficace des données. Également, en présence de données de grandes dimensions, il est possible que certaines variables ne soient pas pertinentes à la prévision, d'où la nécessité d'utiliser des outils qui rendent possible la sélection de variables. Enfin, les outils doivent être en mesure de modéliser convenablement des relations complexes entre les variables, celles-ci augmentant généralement en présence d'un grand nombre de données. Pour ces raisons, les méthodes tirées de l'apprentissage statistique représentent des alternatives intéressantes aux méthodes de séries temporelles et de régression pour la prévision de court

terme de la demande et des prix de l'électricité.

L'apprentissage statistique correspond à un ensemble d'outils de la statistique permettant l'analyse de bases de données et la résolution automatique d'un problème considéré (James et al., 2013). Les outils de l'apprentissage statistique sont classés en deux catégories : supervisé ou non supervisé, selon la nature du problème à résoudre. L'apprentissage non supervisé, caractérisé par l'absence de variable dépendante, cherche à décrire et à résumer l'information contenue dans les données. L'apprentissage supervisé, duquel figure l'objectif de notre recherche, consiste à développer un modèle dont la performance est optimale en termes de prévision d'une variable  $y$ , appelée la variable réponse, à partir d'un ensemble de variables  $X$ , les prédicteurs. La démarche consiste d'abord à séparer l'échantillon en deux parties, soit l'échantillon d'apprentissage et l'échantillon test. L'algorithme analyse l'échantillon d'apprentissage et produit un estimé d'une fonction de prédiction,  $\hat{f}$ , qui sera ensuite utilisée pour le calcul des prévisions de la variable réponse associées aux prédicteurs de l'échantillon test. Les prévisions peuvent alors être comparées aux vraies valeurs de  $y$  contenues dans l'échantillon test (Friedman et al., 2001). La fiabilité de la prévision, généralement calculée à partir de l'erreur quadratique moyenne (MSE), dépend des choix appropriés de la forme fonctionnelle de  $f$  et des prédicteurs,  $X$ . Un modèle flexible présente un biais faible, mais souffre généralement d'un problème de surapprentissage (*overfitting*), c'est-à-dire qu'il correspond trop étroitement à un ensemble particulier de données, et par conséquent sa variance est élevée. Au contraire, le biais d'un modèle rigide sera élevé, mais sa variance sera faible. Puisque la MSE dépend du biais et de la variance de l'estimé, un algorithme d'apprentissage pertinent doit être en mesure de faire l'arbitrage entre ces deux composantes<sup>1</sup>.

Différents algorithmes d'apprentissage ont été appliqués aux prévisions de court terme de la demande et des prix de l'électricité. Il s'agit généralement de méthodes non paramétriques, c'est-à-dire qui n'imposent pas a priori une forme fonctionnelle à la fonction de

---

1. Soit  $\hat{f}$  le modèle estimé à partir de l'échantillon d'apprentissage, alors la MSE d'une observation de l'échantillon test se décompose par :

$$E[(y - \hat{f}(x))^2] = \text{Biais}[\hat{f}(x)]^2 + \text{Variance}[\hat{f}(x)] + \text{var}(\epsilon)$$

où  $x$  correspond aux prédicteurs de l'échantillon test et  $\text{var}(\epsilon)$  est l'erreur non réductible du modèle.

prédiction,  $f$ . Entre autres, les méthodes de logique floue (*fuzzy logic*) et de systèmes experts ont fréquemment été appliquées aux prévisions de court terme de la demande et des prix de l'électricité (Weron, 2007). C'est toutefois les perceptrons multicouches (réseaux de neurones) qui ont connu la plus grande popularité dans les dernières années (Kyriakides et Polycarpou, 2007)<sup>2</sup>. Le perceptron multicouche est avantageé par sa capacité d'estimer des relations non linéaires complexes et sa flexibilité vis-à-vis les données contenant du bruit. Ces avantages ne signifient néanmoins pas que l'utilisation de cette technique aboutisse nécessairement à de meilleures prévisions : son aptitude à modéliser la forte volatilité présente dans les données peut également signifier que le perceptron multicouche engendre une faible performance dans la prévision de la tendance générale des données (Weron, 2014). Également, sa modélisation est complexe, puisqu'elle nécessite de nombreuses paramétrisations pouvant affecter sa performance prédictive (nombre de couches cachées, nombre de neurones par couche cachée, forme de la fonction de transfert), et ses résultats sont difficilement interprétables (Dudek, 2015). On retrouve également certains inconvénients dans la logique floue et les systèmes experts ; ces modèles, basés sur le raisonnement et le savoir d'experts, doivent généralement user de techniques d'approximation complexes en raison de la difficulté liée à l'acquisition de ces informations. Pour ces raisons, l'utilisation de ce type de modèles n'est pas reconnue par l'ensemble des chercheurs du domaine. Par conséquent, les perceptrons multicouches, les méthodes de logique floue ainsi que les systèmes d'experts ne seront pas employés dans le cadre de ce mémoire.

---

2. Le perceptron multicouche correspond à un réseau composé de couches successives, où chaque couche représente un ensemble de neurones n'ayant pas de connexion entre eux. Les neurones sont plutôt connectés en entrée à chacun des neurones de la couche précédente, et en sortie, à chaque neurone de la couche suivante. Chaque neurone, outre ceux de la couche d'entrée, comprend une fonction de combinaison et une fonction de transfert (généralement sigmoïde ou linéaire). Dans sa forme la plus simple, le réseau se compose de deux couches : une couche d'entrée de  $P$  neurones, un par prédicteur, et une couche de sortie d'un seul neurone. Les neurones de la couche d'entrée alimentent le neurone de la couche de sortie qui définit une combinaison linéaire des  $P$  neurones. Puis, cette combinaison linéaire est transformée à partir de la fonction de transfert, d'où est obtenue une prévision (Saporta, 2006). Dans ce cas particulier, le modèle obtenu correspond à une régression linéaire ou logistique, selon la forme de la fonction de transfert. En réalité, toutefois, les réseaux appliqués dans le cadre de l'analyse prédictive de la demande et des prix de l'électricité comportent généralement plusieurs couches cachées entre les couches d'entrée et de sortie (Weron, 2007), complexifiant ainsi la modélisation présentée. Les perceptrons multicouches ont, entre autres, été utilisés pour l'analyse prédictive de court terme de la demande d'électricité en Ontario par Liu et al. (2006), et pour l'analyse prédictive de court terme des prix de l'électricité en Ontario par Rodriguez et Anders (2004), Zareipour et al. (2006) et Sahay (2015).

Bien que moins exploitée, la méthode des forêts aléatoires figure également parmi les techniques utilisées pour les analyses prédictives de la demande et des prix de l'électricité. Les auteurs la choisissent généralement pour sa simplicité et sa facilité d'implémentation. La méthode des forêts aléatoires consiste à construire aléatoirement et à agréger un nombre élevé d'arbres de décision, où chaque arbre de décision représente, sous forme d'un graphique, un séquençement logique montrant différents résultats possibles en fonction des choix effectués à chacune des étapes (James et al., 2013). Dudek (2015) compare les performances prédictives de différents modèles dont une forêt aléatoire, un arbre de décision individuel, un modèle ARIMA et un réseau de neurones, dans le calcul des prévisions horaires, une journée à l'avance, de la demande d'électricité en Pologne. Les résultats montrent que la forêt aléatoire performe aussi bien que le réseau de neurones, et mieux que l'ensemble des autres modèles étudiés. Ludwig et al. (2015) étudient quant à eux la performance prédictive des forêts aléatoires dans le contexte des prévisions de court terme des prix de l'électricité en Allemagne. Les résultats montrent que le modèle de forêt aléatoire engendre une performance prédictive élevée, relativement à celles obtenues à partir d'un modèle ARMA univarié et d'un modèle ARMA avec variables explicatives exogènes (température et vitesse du vent).

Enfin, quelques auteurs se sont également intéressés à l'application de méthodes paramétriques dans le contexte des analyses prédictives de la demande et des prix de l'électricité. Ces méthodes imposent une forme fonctionnelle à la fonction de prédiction  $f$ . Entre autres, la méthode LASSO (*Least Absolute Shrinkage and Selection Operator*) a été utilisée pour le calcul des prévisions horaires de la demande et des prix de l'électricité. Cette méthode correspond à un modèle de régression linéaire multiple, auquel est ajouté un terme de pénalité qui influence la taille des paramètres estimés. Le LASSO bénéficie du fait qu'il permet la sélection de variables, c'est-à-dire que la régression finale peut ne pas contenir tous les prédicteurs initialement exploités. Ziel et Liu (2016) ont développé un modèle linéaire pour les prévisions de la demande d'électricité, où cette dernière est modélisée comme une combinaison linéaire des valeurs passées de la demande et de la température. Les auteurs utilisent la méthode LASSO afin de sélectionner et d'estimer les paramètres du modèle. Appliqué dans le cadre du *Global Energy Forecasting Competition 2014*, le modèle montre des performances

prédictives supérieures à celles obtenues à partir de modèles de régressions linéaires multiples non pénalisées. Le LASSO a également été employé dans le calcul des prévisions horaires des prix de l'électricité par Ziel (2016) et par Ludwig et al. (2015). Ces derniers montrent que la performance prédictive du modèle LASSO, appliqué aux calculs des prévisions horaires des prix de l'électricité polonaise, est supérieure à celle d'un modèle ARMA univarié, et semblable à celle obtenue à partir d'un modèle de forêt aléatoire.

La revue de la littérature montre la complexité du calcul des prévisions de court terme de la demande et des prix de l'électricité. En effet, il ne semble pas exister un modèle unique dont les performances prédictives sont optimales à tous coups. Puisqu'il n'y a aucun marché de l'électricité qui soit semblable, l'analyse prédictive nécessite généralement une étude détaillée de différentes modélisations intégrant les facteurs géographiques et météorologiques spécifiques à la région étudiée (Kyriakides et Polycarpou, 2007). C'est la démarche qui sera réalisée dans le cadre de ce mémoire. Différents modèles de la littérature, dont la régression linéaire multiple, le LASSO et la forêt aléatoire seront appliqués pour le calcul des prévisions horaires de la demande et des prix de l'électricité de l'Ontario. Ce mémoire propose également l'étude du modèle de régression linéaire avec erreur ARIMA, de la régression Ridge et Elastic Net, et de la régression sur composantes principales. Ces méthodes, peu exploitées dans la littérature, présentent toutefois des caractéristiques intéressantes dans le contexte des analyses prédictives de la demande et des prix de l'électricité.

# Chapitre 3

## Données

Tel que mentionné dans la revue de la littérature, la modélisation des prévisions horaires de la demande et des prix de l'électricité nécessite de considérer les éléments spécifiques au marché étudié. Afin d'acquérir une meilleure compréhension de son processus d'opération et de ses caractéristiques distinctives, cette section débute par une brève description du marché de l'électricité de l'Ontario. On y présente ensuite les données utilisées pour les analyses prédictives de court terme de sa demande et de ses prix.

### 3.1 Le marché d'électricité de l'Ontario

Inspiré par la vague de déréglementation des industries de l'électricité à travers le monde, le gouvernement ontarien lançait, il y a vingt ans, une politique de restructuration du marché de l'électricité. L'unique entreprise publique, Ontario Hydro, est alors scindée en plusieurs entités distinctes : l'entreprise Ontario Power Generation (OPG) qui hérite de la majorité des actifs (usines, barrages et centrales) et devient ainsi le producteur majoritaire d'électricité en Ontario ; Hydro One qui prend en charge la transmission et la distribution de l'électricité ; et l'IESO (Independent Electricity System Operator), l'organisme indépendant de réglementation du marché ontarien, ayant pour mandat la gestion d'un marché d'offre indépendant au sein duquel l'ensemble des entreprises qui produisent de l'électricité participent. L'IESO lance ainsi en 2002, un nouveau marché "spot" d'électricité, obligatoire pour toute entreprise

désirant vendre sa production. Le rôle de l'IESO est de réglementer le marché et d'assurer son bon fonctionnement.

Le marché ontarien fonctionne avec un système d'enchères : quotidiennement, l'IESO envoie aux firmes autorisées à produire et à vendre l'électricité en Ontario, une prévision de la demande, détaillée par heure, des trente prochains jours (IESO, 2018a). Ces prévisions permettent aux firmes d'estimer la quantité d'électricité qu'elles seront prêtes à produire chaque heure, selon différents niveaux de prix. Les firmes doivent envoyer leur offre horaire initiale à l'IESO un minimum de 13 heures à l'avance afin de déclarer leur participation au marché. L'offre de chaque heure peut ensuite être modifiée par les firmes jusqu'à 2 heures la précédant (IESO, 2018b). Le marché de l'Ontario se différencie ainsi par l'absence d'un marché à 24 heures (*Day-Ahead Market*) : à partir de l'offre individuelle de chaque firme, l'organisme régulateur équilibre le marché en temps réel en utilisant les sources d'énergie, de la moins chère à la plus chère, jusqu'à la satisfaction de la demande (Genc et al., 2015)<sup>1</sup>. Le prix de marché est révisé toutes les cinq minutes par l'IESO et la moyenne des douze prix de chaque heure représente le prix moyen horaire (HOEP)<sup>2</sup>. Ce prix est appliqué uniformément à travers l'ensemble de la province (Zareipour et al., 2007b).

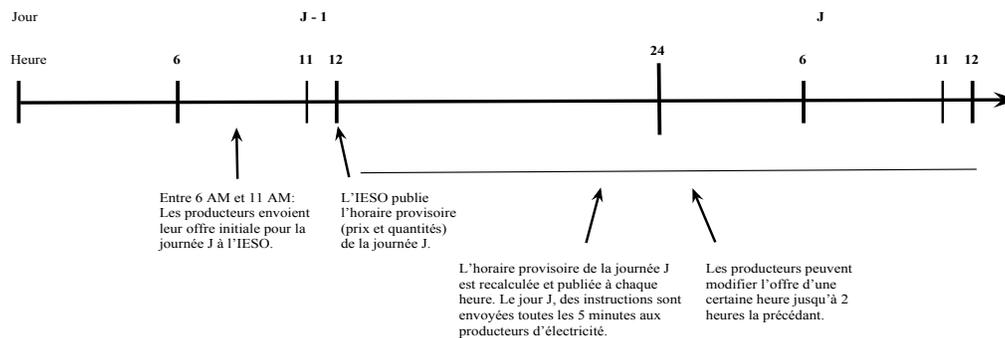


FIGURE 3.1 – Calendrier des opérations quotidiennes du marché de l'électricité en Ontario<sup>3</sup>. Les évènements mentionnés surviennent chaque jour ; à des fins de compréhension, les évènements du jour  $J - 1$  ne sont pas présentés au jour  $J$ .

1. En septembre 2017, l'IESO a annoncé le désir de lancer un marché à 24 heures en Ontario, afin de réduire la volatilité des prix de l'électricité (IESO, 2018a). Un marché à 24 heures permet aux participants d'acheter et de vendre l'électricité une journée avant le jour des opérations. Il améliore, par conséquent, la certitude des opérations pour les participants des marchés de l'électricité, réduisant ainsi généralement les variations importantes observées dans les prix.

2. Le terme HOEP réfère en anglais à *Hourly Ontario Electricity Price*.

3. La figure est inspirée de Danai et al. (2001).

L'IESO (2018b) estime toutefois que 88,1% du marché ontarien n'est pas soumis au prix d'équilibre en raison du marché régulé et des contrats Feed-in-Tariffs (FiT). En effet, un élément important, spécifique au marché de l'électricité ontarien, est la présence d'actifs sous réglementation, c'est-à-dire de capacités de production dont le prix de vente est garanti par l'autorité régulatrice. Cette réglementation est définie afin d'assurer la distribution en tout temps d'une certaine quantité d'électricité, ainsi que pour inciter les producteurs exploitants des sources d'énergie spécifiques, notamment l'énergie nucléaire et hydraulique (IESO, 2018b). En outre, certaines firmes concluent des contrats plus souples avec l'IESO tels que les petits producteurs à énergie propre (entre autres, éolienne et solaire) qui dépendent du système de FiT afin de garantir une certaine stabilité de leurs revenus. Établi en 2009, le système de FiT garantit, à long terme aux participants, un prix préétabli par un contrat pour toute l'électricité produite qui est fournie au réseau ontarien. Ces firmes, dont le prix de la production est garanti par un contrat, ont intérêt à miser le plus bas possible, s'assurant ainsi d'écouler la totalité de leur production (Bushnell et al., 2008), et par conséquent, elles ne sont pas soumises au prix d'équilibre du marché.

Le réseau électrique de l'Ontario se définit enfin à travers les liens qu'il partage avec cinq régions voisines. En effet, le réseau électrique ontarien est interconnecté à celui du Québec à l'est, du Manitoba à l'ouest, de l'état de New York au sud-est, et des états du Michigan et du Minnesota au sud-ouest. Ces interconnexions permettent l'importation et l'exportation d'électricité. L'IESO importe et exporte l'électricité en fonction du prix observé en Ontario, ainsi que celui des régions où son réseau électrique est connecté (Zareipour et al., 2007b). Ces régions doivent ainsi être considérées dans nos analyses prédictives, puisqu'elles ont un impact direct sur la demande d'électricité en Ontario, mais également sur les prix.

Cette courte description montre le caractère unique du marché de l'électricité ontarien. La suite de cette section présente les données utilisées dans les modélisations prédictives de la demande et des prix de l'électricité en Ontario.

## 3.2 Données

Les données requises pour les prévisions de court terme de la demande et des prix de l'électricité ont été collectées auprès de deux sources : l'IESO via le site IESO (2018d) et le NOAA (National Oceanic and Atmospheric Administration) via le NCDC (2018). Les données horaires recueillies s'étendent du 1er mai 2010 à 00h00 au 31 décembre 2015 à 23h00 (inclus). Chaque variable comprend plus de 49700 observations.

### Demande et prix de l'électricité

Le site de l'IESO fournit, pour chaque heure, les quantités totales demandées ainsi que les prix observés dans le marché de l'électricité ontarien. Le tableau 3.1 résume les valeurs moyennes de la demande totale et des prix de chaque année de notre échantillon.

Année	Demande totale (MWh)	Prix de l'Ontario (\$CAD)	Importations nettes (MWh)
2010	17960 (2860,94)	37,83 (20,86)	1092 (803,87)
2011	17608 (2554,48)	30,12 (20,66)	1025 (699,14)
2012	17741 (2563,48)	22,82 (18,77)	1130 (638,75)
2013	18099 (2375,67)	24,98 (22,22)	-264 (325,52)
2014	18068 (2114,99)	32,39 (46,74)	-221 (341,69)
2015	18154 (2400,67)	21,66 (35,22)	-263 (315,33)

Tableau 3.1 – Moyennes des quantités demandées totales, des prix et des importations nettes du marché ontarien entre 2010 et 2015. Les écarts-type sont présentés entre parenthèses.

Les résultats montrent qu'à travers les années la demande totale est sensiblement restée stable, se situant entre 17608 MWh et 18154 MWh, vis-à-vis une capacité totale d'environ 35000 MW (IESO, 2015b). Les prix, quant à eux, se sont situés entre 20 et 40 dollars canadiens. Cette variation marquée montre une forte volatilité dans les prix de l'électricité en Ontario. Il a d'ailleurs été montré empiriquement que la volatilité des prix ontariens est significativement supérieure à celles observées dans le marché Pennsylvania-New Jersey-Maryland (PJM), le marché de la Nouvelle-Angleterre et le marché de New York (Zareipour et al., 2007a). Ce phénomène est notamment attribuable à la nature en temps réel du marché de l'Ontario, causée par l'absence d'un marché à 24 heures (*Day-Ahead Market*).

Le graphique 3.2 montre l'évolution détaillée du prix moyen horaire entre 2010 et 2015. Confirmant à nouveau une forte volatilité dans les prix ontariens, le graphique montre également un élément rare dans les marchés de l'électricité : la présence de prix négatifs. Les prix négatifs sont une conséquence directe des contrats signés entre différents producteurs d'électricité et l'IESO qui, en garantissant le prix de vente de leur production, encourage les producteurs à adopter des comportements stratégiques sur le marché (Fremeth et al., 2017). Par exemple, on retrouve fréquemment ce type de comportement chez les producteurs d'énergie éolienne en raison de la combinaison de trois éléments. D'abord, la production à partir d'énergie éolienne en Ontario est fréquemment supérieure la nuit, au moment où la demande d'électricité est la plus faible, puisqu'il s'agit de la période où le vent est généralement plus fort (Archer et Jacobson, 2005). Ensuite, puisqu'il n'est pas économiquement rentable d'emmagasiner l'électricité ou de fermer les turbines éoliennes, les producteurs doivent vendre l'ensemble de leur production afin de maximiser leur profit. Enfin, en étant inscrits au système de FiT, les producteurs d'énergie éolienne reçoivent un prix garanti pour chaque mégawatt (MW) d'électricité produite qui est fournie au réseau ontarien. Par conséquent, le producteur qui veut s'assurer d'écouler la totalité de sa production en période de demande peu élevée peut miser un prix négatif et, donc, payer l'organisme régulateur pour qu'elle achète sa production, car il fera un profit positif tant que le prix préalablement garanti par l'IESO est supérieur à ce qu'il doit déboursier.

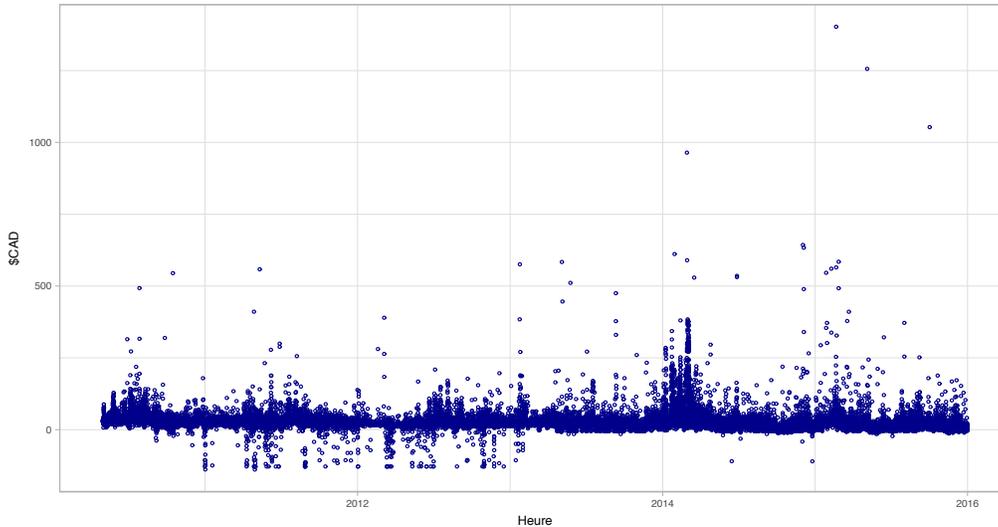


FIGURE 3.2 – Évolution du prix moyen en Ontario entre 2010 et 2015.

Du graphique 3.2, il est également possible d’observer que la majorité des prix négatifs éloignés de 0 se situent entre les années 2010 et 2012 (inclus). En effet, bien que le nombre d’observations négatives soit supérieur pour la période de 2013 à 2015 que pour la période de 2010 à 2012 (2048 versus 359 observations, respectivement), le prix moyen horaire a été inférieur à 100 dollars canadiens 83 fois entre 2010 et 2012, contre seulement 3 fois entre 2013 et 2015. Ce changement marqué s’explique par le fait que l’IESO ait mis en place, en 2013, des prix plancher pour les producteurs d’énergie éolienne et solaire, ainsi que pour les producteurs d’énergie nucléaire, afin de limiter les comportements stratégiques dans le marché de l’électricité en Ontario<sup>4</sup>. Le changement de réglementation ontarienne, dont l’impact est important sur les prix observés, nous amène à séparer les données en deux échantillons : avant et après, selon que les prix plancher aient été établis ou non.

Enfin, le graphique 3.2 permet d’identifier deux périodes dans l’échantillon où la province de l’Ontario a subi des envolées inattendues dans les prix. Celles-ci surviennent en hiver 2014 (IESO, 2014) et en hiver 2015 (IESO, 2015a), où le prix horaire moyen était respectivement de 67,49 et 31,93 dollars canadiens.

---

4. Le prix plancher de l’énergie éolienne a été fixé à -10 dollars canadiens par MWh pour les premiers 90% de la capacité du producteur et à -15 dollars canadiens par MWh pour les 10% restants. Le prix plancher de l’énergie nucléaire a été fixé à -5 dollars canadiens par MWh (Ltd., 2018).

## Facteurs externes

Les facteurs qui influencent la demande d'électricité et, donc indirectement les prix, se classent en deux catégories : les facteurs climatiques et les facteurs saisonniers (Kyriakides et Polycarpou, 2007). De plus, puisque le réseau électrique ontarien est interconnecté à celui de cinq régions voisines, les importations nettes impactent la demande totale d'électricité en Ontario, et sont donc également susceptibles d'influencer les prix.

Le tableau 3.1, présenté à la sous-section précédente, montre les importations nettes moyennes de l'Ontario entre 2010 et 2015. Des résultats, il est possible d'observer un "changement de régime" dans les importations nettes : la province détenant d'abord le statut d'importateur net entre 2010 et 2012, devient, à partir de 2013, exportateur net d'électricité. Ces deux régimes coïncident avec les périodes établies précédemment : "avant" et "après", selon le changement de réglementation des prix de l'électricité en Ontario.

Les facteurs climatiques correspondent à des variables météorologiques telles que la température, le degré d'humidité et la vitesse du vent. La température correspond au facteur climatique ayant l'influence la plus importante sur la demande d'électricité, et donc également sur les prix (Hahn et al., 2009). Le NOAA présente les températures horaires de plusieurs villes d'Amérique du Nord. La construction des températures moyennes de la province de l'Ontario ainsi que des provinces et états où est transportée l'électricité ontarienne, s'effectue en calculant la température moyenne pondérée par la taille des villes où la température est enregistrée. Le graphique 3.3 présente de quelle manière la demande totale d'électricité en Ontario varie en fonction des températures moyennes des régions étudiées. Les résultats montrent que la relation entre la température et la demande d'électricité est non linéaire : une température très basse ou très élevée engendre une demande d'électricité élevée. Ce phénomène s'explique par le fait que ces périodes nécessitent généralement un usage important du chauffage électrique et des climatiseurs (Hippert et al., 2001).

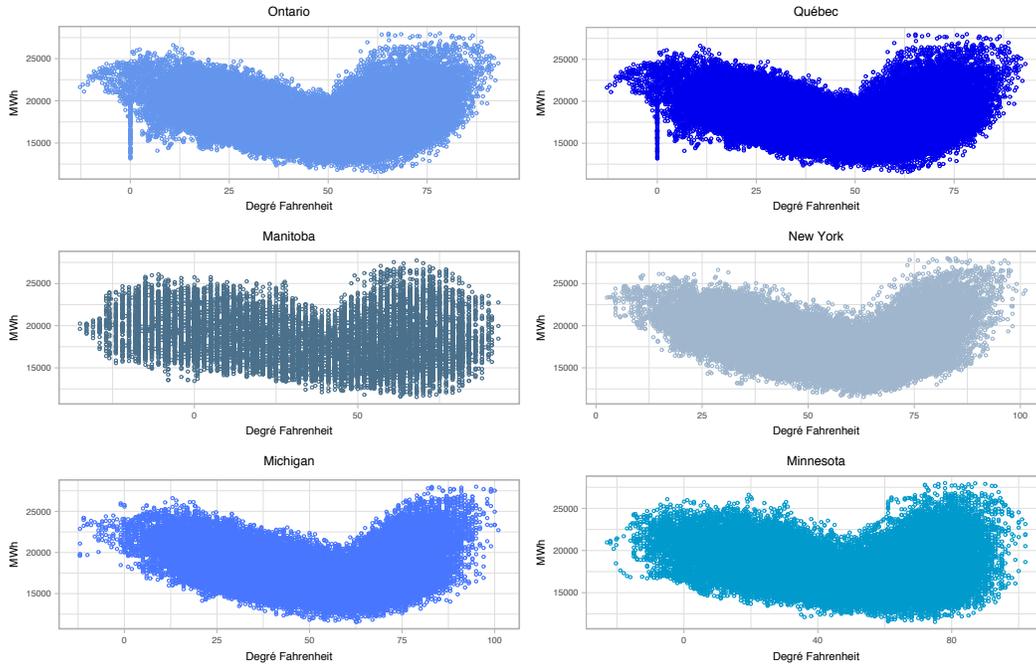


FIGURE 3.3 – Relations entre la température (en degré Fahrenheit) des régions étudiées et la demande totale ontarienne.

En s’inspirant entre autres de la démarche de Papalexopoulos et Hesterberg (1990), les effets de la température sur la demande et les prix de l’électricité seront représentés dans nos modèles prédictifs à partir des CDD (Cooling degree days) et HDD (Heating degree days). Le calcul des CDD et HDD consiste à soustraire la température moyenne de la journée à un seuil défini, généralement celui de 65 degrés Fahrenheit (soit 18 degrés Celsius) dans la littérature technique. Une valeur négative signifie qu’il faut refroidir sa maison : on parle ainsi de CDD. À l’opposé, une valeur positive est associée à la nécessité de chauffer sa maison, et on parle donc de HDD<sup>5</sup>. Ces valeurs, constantes à travers les 24 heures d’une même journée, représentent donc le besoin énergétique quotidien qui découle de la température. Les valeurs moyennes des CDD et HDD observées dans les différentes régions étudiées, à travers les deux périodes précédemment établies, sont présentées dans le tableau 3.2.

5. Le lecteur intéressé pourra trouver de l’information supplémentaire via l’article de Bromley (2009).

Avant							
	Année	ON	QC	MN	MI	NY	MB <sup>1</sup>
HDD	2010	11.66	20.73	12.76	10.38	7.83	18.36
	2011	23.12	44.61	19.58	16.88	12.63	-
	2012	21.52	37.04	16.66	14.28	11.26	-
CDD	2010	3.45	0.52	4.75	4.85	5.79	2.03
	2011	0.33	0.00	3.38	3.14	3.46	-
	2012	0.40	0.00	3.71	3.56	3.44	-
Après							
	Année	ON	QC	MN	MI	NY	MB
HDD	2013	20.21	24.69	21.88	17.37	13.48	30.64
	2014	22.51	25.20	23.06	17.87	13.86	30.61
	2015	19.69	24.27	18.28	16.48	12.68	25.66
CDD	2013	1.74	1.29	3.16	2.75	3.10	1.55
	2014	1.53	1.22	2.45	2.80	2.92	2.42
	2015	1.80	1.44	2.78	2.67	3.80	1.73

Tableau 3.2 – HDD et CDD (en degré Fahrenheit) moyens à travers les régions étudiées, classées selon les périodes "avant" et "après". Les régions étudiées sont l'Ontario (ON), le Québec (QC), le Minnesota (MN), le Michigan (MI), l'état de New York (NY) et le Manitoba (MB).

<sup>1</sup> La présence de valeurs manquantes ne permet pas d'obtenir les moyennes des HDD et CDD des années 2011 et 2012 dans la province du Manitoba (MB).

Les variables saisonnières quant à elles décrivent le temps. Elles seront incluses dans les modèles prédictifs de manière artificielle par la création de variables dichotomiques. Dans la revue de la littérature, il a été dit que les séries temporelles de la demande de l'électricité comportent généralement trois types de saisonnalité : un cycle quotidien, hebdomadaire et annuel (Hahn et al., 2009). Pour représenter le cycle quotidien, nous divisons chaque journée de notre échantillon en quatre périodes : le creux de la nuit, le pic matinal, le creux de la journée et le pic de soirée, en nous basant sur la classification de l'IESO (voir tableau 3.3). La demande d'électricité ontarienne se différencie également selon le jour de la semaine. En effet, elle est généralement supérieure les jours de semaine, vis-à-vis des jours de fin de semaine, en raison d'une forte demande en provenance du secteur industriel. La saisonnalité hebdomadaire sera ainsi représentée dans nos modèles prédictifs en décomposant les semaines de nos échantillons en deux parties : les jours de semaine et les jours de fin de semaine. Enfin,

les séries temporelles de la demande d'électricité ontarienne montrent également une périodicité annuelle, définie par les quatre saisons de l'année. En effet, en Ontario, les valeurs mensuelles minimales de la demande d'électricité surviennent en avril (16862 MWh), ainsi qu'en septembre et en octobre (17089 MWh et 16581 MWh, respectivement). Les valeurs maximales, quant à elles, se produisent en décembre, en janvier et en février (18528 MWh, 19781 MWh et 19512 MWh, respectivement), ainsi qu'en juillet et en août (19138 MWh et 18377 MWh, respectivement).

Nom de la période	Heures de la journée
Creux de la nuit	20h00-6h00
Pic matinal	6h00-8h00
Creux de la journée	8h00-17h00
Pic de soirée	17h00-20h00

Tableau 3.3 – Les quatre périodes de la journée selon la classification de l'IESO.

# Chapitre 4

## Modèles

L'objectif de notre recherche consiste à étudier les performances prédictives de différentes techniques de l'apprentissage statistique, appliquées dans le contexte des marchés de l'électricité. Cette section présente en détail les caractéristiques des modèles analysés. En particulier, les 5 modélisations étudiées tirées de l'apprentissage statistique y sont décrites, ainsi que deux modélisations conventionnelles, utilisées à titre de *benchmark* dans notre démarche de comparaison.

### 4.1 Modèles conventionnels

#### 4.1.1 Régression linéaire multiple

Le modèle de régression linéaire multiple suppose une relation linéaire entre la variable réponse et les prédicteurs. En particulier, le modèle est représenté par :

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_P x_{tP} + \epsilon_t$$

où  $y_t$  est la variable réponse composée de  $t$  observations horaires,  $x_{tp}$  ( $p = 1, \dots, P$ ) sont les prédicteurs,  $\beta_P$  sont les paramètres de régression associés aux  $x_{tp}$ , et  $\beta_0$  est la constante de régression. Le terme  $\epsilon_t$  représente le résidu de population qui, dans le contexte de cette recherche, est de variance conditionnelle non constante :  $V(\epsilon_t | x_{tp}) = \sigma_t$ .

En présence d'hétéroscédasticité, l'estimateur des moindres carrés ordinaires (MCO) est généralement inefficace dans un contexte prédictif. En effet, la variance non constante du terme d'erreur peut signifier que ce dernier renferme certaines informations pouvant être utiles à la prévision de la variable réponse. Suivant la démarche de Papalexopoulos et Hesterberg (1990), les paramètres de régression seront donc estimés à partir de la méthode des moindres carrés généralisés faisables (MCGF), où la forme de l'hétéroscédasticité est définie par :

$$V(\epsilon_t|x_{tp}) = \sigma_\epsilon^2 \exp(\delta_0 + \delta_1 x_{t1} + \dots + \delta_P x_{tP})$$

où  $x_{tp}$  ( $p = 1, \dots, P$ ) représentent les prédicteurs, et  $\delta_i$  ( $i = 0, \dots, P$ ) sont des paramètres à estimer en utilisant les données (Greene, 2003).

Au temps  $t$ , la meilleure prévision de la variable réponse au temps  $t+i$  se calcule comme une combinaison linéaire des prédicteurs, c'est-à-dire :

$$\hat{y}_{t+i} = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \dots + \hat{\beta}_P x_{Pt}.$$

où  $\hat{\beta}_P$  sont les paramètres estimés par MCGF et  $i$  représente l'horizon de prévision.

### 4.1.2 Régression linéaire avec erreur ARIMA

Le modèle de régression linéaire avec erreur ARIMA représente un cas particulier du modèle de fonction de transfert (FT) (Hyndman et Athanasopoulos, 2014). Plutôt que de modéliser la variable réponse comme une fonction de ses valeurs passées, des valeurs présentes et passées des prédicteurs, ainsi que d'un terme d'erreur décrit par un processus ARIMA (Box et Jenkins, 1976), le modèle de régression linéaire avec erreur ARIMA la représente comme une fonction d'uniquement les valeurs présentes des prédicteurs et d'un terme d'erreur suivant un processus ARIMA.

Le modèle de régression linéaire avec erreur ARIMA peut ainsi être représenté de la manière suivante :

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_P x_{tP} + \eta_t$$

où  $x_{tp}$  ( $p = 1, \dots, P$ ) sont les prédicteurs,  $\beta_P$ , les coefficients qui sont leur sont associés et  $\beta_0$ , la constante de régression. Le terme  $\eta_t$  correspond au résidu de régression. Il est représenté par un processus ARIMA( $p, d, q$ ), c'est-à-dire qu'il est modélisé par une composante autorégressive (AR) d'ordre  $p$ , une composante à moyenne mobile (MA) d'ordre  $q$ , et est intégré d'ordre  $d$  :

$$\eta_t = \phi_1 \eta_{t-1} + \dots + \phi_p \eta_{t-p} - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} + \epsilon_t$$

où  $\phi_i$  ( $i = 1, \dots, p$ ) et  $\theta_j$  ( $j = 1, \dots, q$ ) sont des paramètres à estimer,  $|\phi_i| < 1$  et  $|\theta_j| < 1$ , et les termes  $\epsilon$  représentent des bruits blancs,  $\epsilon \sim N(0, \sigma^2)$ .

La prévision de la variable réponse à l'horizon de temps  $i$  se calcule à partir des coefficients estimés ( $\beta$ ,  $\phi$  et  $\theta$ ), soit :

$$\hat{y}_{t+i} = \hat{\beta}_1 x_{t1} + \dots + \hat{\beta}_P x_{tP} + \hat{\phi}_1 \eta_{t-p} - \hat{\theta}_1 \epsilon_{t-1} - \dots - \hat{\theta}_q \epsilon_{t-q}.$$

## 4.2 Modèles de l'apprentissage statistique

Cette sous-section vise à présenter les principales caractéristiques des modèles de l'apprentissage statistique supervisé employés pour la prévision de la demande et du prix de l'électricité en Ontario. Tel que mentionné précédemment, l'exactitude des prévisions dépend du biais de l'estimateur, ainsi que de sa variance. L'arbitrage entre ces deux composantes relève, d'une part, du choix approprié de la forme fonctionnelle de la fonction de prédiction. Les méthodes étudiées se classent ainsi parmi deux catégories : les méthodes paramétriques et non paramétriques, selon qu'elles supposent ou non une forme explicite pour la fonction de prédiction. D'autre part, le contrôle du biais et de la variance d'un modèle s'effectue par un choix adéquat des prédicteurs,  $X$ . Celui-ci s'accomplit généralement par l'introduction de différents paramètres de pénalisation qui visent à contrôler le problème de surapprentissage de l'estimateur. L'optimisation de ces paramètres représente une étape importante, préalable à l'estimation des modèles de l'apprentissage statistique, et nécessite l'adoption d'une démarche de paramétrisation efficace. Dans le cadre de ce mémoire, nous adopterons une technique basée sur une validation hors échantillon : la validation croisée à  $K$  groupes. Une présentation détaillée de cette dernière sera offerte à la section suivante (5.2.2).

## 4.2.1 Méthodes paramétriques

### Régressions pénalisées : Ridge, LASSO et Elastic Net

La régression linéaire multiple décrite à la sous-section 4.1.1 présente plusieurs caractéristiques intéressantes, notamment sa facilité d'interprétation et son pouvoir prédictif élevé lorsque les données affichent une relation linéaire. L'estimation par moindres carrés est toutefois susceptible d'engendrer certains problèmes critiques, particulièrement lorsque le nombre d'observations est faible vis-à-vis le nombre de prédicteurs. Une telle situation présente généralement un problème de colinéarité et, par conséquent, la variance de l'estimateur est instable et sa performance prédictive, basse.

Le contrôle de la variance de l'estimateur par moindres carrés peut se faire par l'emploi des méthodes de régressions pénalisées : Ridge, LASSO et Elastic Net. Les méthodes de régressions pénalisées contraignent la taille des paramètres estimés : le biais de l'estimateur augmente, mais sa variance peut suffisamment diminuer de manière à améliorer sa performance prédictive. Les paramètres de régressions pénalisées sont estimés en résolvant une version contrainte du problème d'optimisation des moindres carrés ordinaires (MCO). En particulier, la fonction de perte à minimiser correspond à la somme des carrés des résidus (SCR) auquel sont ajoutés des termes de pénalité, soit :

$$L(\lambda_1, \lambda_2, \beta) = \sum_{t=1}^n \left( y_t - \sum_{j=1}^P x_{tj} \beta_j \right)^2 + \lambda_2 \sum_{j=1}^P \beta_j^2 + \lambda_1 \sum_{j=1}^P |\beta_j|$$

où  $\lambda_1$  et  $\lambda_2$  sont des paramètres de régularisation, c'est-à-dire des paramètres agissant sur la taille des termes de pénalité. Puisque les termes de pénalité dépendent de la taille des coefficients estimés, il est nécessaire que les variables employées soient de même ordre de grandeur. On les supposera donc centrées-réduites de manière à ce qu'elles aient une moyenne nulle et une variance unitaire.

Soit  $\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$ , le problème à solutionner afin d'estimer les paramètres  $\beta$  des régressions pénalisées peut être représenté par :

$$\hat{\beta} = \underset{\beta}{\operatorname{arg\,min}} \left\{ \sum_{t=1}^n \left( y_t - \sum_{j=1}^P x_{tj} \beta_j \right)^2 \right\}$$

$$\text{sujet à } (1 - \alpha) \sum_{j=1}^P |\beta_j| + \alpha \sum_{j=1}^P \beta_j^2 \leq g \quad (4.1)$$

où  $g$  représente le "budget" disponible; un budget élevé signifie que la pénalité est peu contraignante, et vice-versa.

Lorsque  $\alpha = 1$ , on obtient la régression Ridge (Hoerl et Kennard, 1970); lorsque  $\alpha = 0$ , on obtient la régression LASSO (Tibshirani, 1996). Pour toutes valeurs  $\alpha \in (0, 1)$ , on obtient la régression Elastic Net (Zou et Hastie, 2005). De la figure 4.1, il est possible d'observer les différences des pénalités LASSO (gauche) et Ridge (droite), dans le cadre d'une régression à deux prédicteurs<sup>1</sup>.

La pénalité Ridge permet le rétrécissement de la taille des coefficients estimés associés aux prédicteurs dont le pouvoir prédictif est faible, mais ne permet pas de fixer certains d'eux égaux à 0. En effet, tel que démontré dans la figure 4.1, la probabilité que la tangence entre une ellipse et le cercle survienne au croisement d'un axe d'origine est extrêmement faible. L'intérêt de la pénalité LASSO, au contraire, est qu'elle permet la création de modèles parcimonieux, c'est-à-dire un modèle final qui ne contient pas l'ensemble des  $P$  prédicteurs. En effet, en plus d'affecter la taille des coefficients estimés, la pénalité LASSO permet la sélection de variables, c'est-à-dire qu'elle fixe à 0 les coefficients associés aux prédicteurs à faible pouvoir prédictif. La pénalité Elastic Net, quant à elle, représente une combinaison linéaire strictement convexe des pénalités Ridge et LASSO. Possédant également la capacité de sélection de variables, son avantage réside plutôt dans le fait qu'elle prend en compte l'effet de groupe des prédicteurs, lorsque certains d'entre eux sont fortement corrélés. La méthode Elastic Net requiert toutefois la détermination d'un paramètre supplémentaire :  $\alpha$ .

---

1. La figure est inspirée de celle retrouvée dans James et al. (2013).

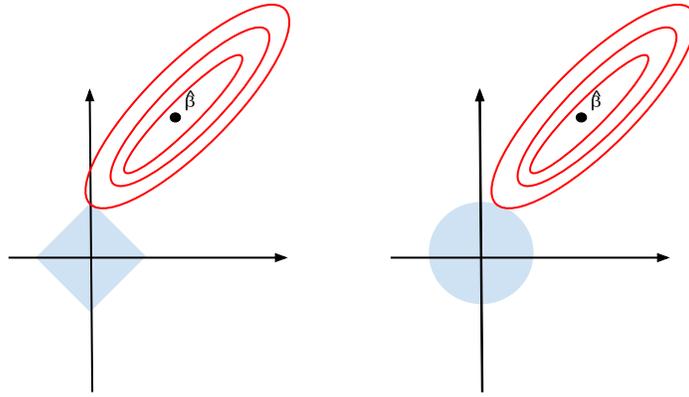


FIGURE 4.1 – Représentation de la pénalité Lasso (gauche) et la pénalité Ridge (droite). Le carré désigne la contrainte  $|\beta_1| + |\beta_2| \leq t$  de la régression LASSO et le cercle désigne la contrainte  $\beta_1^2 + \beta_2^2 \leq t$  de la régression Ridge. Les ellipses montrent les contours des SCR( $\beta$ ) et  $\hat{\beta} = \hat{\beta}^{MCO}$ .

Similairement à la méthode par régression linéaire, la valeur prédite de la variable réponse à l’horizon de prévision  $i$  se calcule à partir des coefficients estimés associés aux prédicteurs au temps  $t$ , c’est-à-dire :

$$\hat{y}_{t+i} = \hat{\beta}_1 x_{1t} + \dots + \hat{\beta}_H x_{Ht}$$

où  $H$  représente le nombre de prédicteurs du modèle final ( $H \leq P$ ) et  $\hat{\beta}_h$  ( $h = 1, \dots, H$ ) sont les paramètres estimés par régression Ridge, LASSO ou Elastic Net.

## Régression sur composantes principales

Les méthodes de réduction de dimension, parmi lesquelles figure la régression sur composantes principales (PCA), représentent également une solution dans le contrôle de la variance de l’estimateur des moindres carrés<sup>2</sup>. Ces méthodes nécessitent également que les variables aient d’abord été centrées-réduites. Toutefois, elles diffèrent des régressions pénalisées, puisqu’elle consiste à transformer la matrice des  $P$  prédicteurs, préalablement à l’estimation par MCO (James et al., 2013).

La démarche de la PCA s’effectue en deux étapes. D’abord, l’ensemble des prédicteurs est résumé par un sous-ensemble de variables,  $\mathbf{z}_1, \dots, \mathbf{z}_M$ , appelé les composantes principales. Soit

2. Le terme PCA correspond à *Principal component analysis*.

$\mathbf{X}$  la matrice des  $P$  prédicteurs, les composantes principales correspondent à des combinaisons linéaires des prédicteurs construits de manière à obtenir la meilleure visualisation de  $\mathbf{X}$ . La figure 4.2 représente les deux premières composantes principales d'un nuage d'observations quelconque<sup>3</sup>.

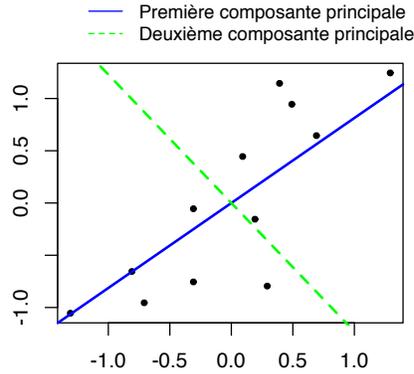


FIGURE 4.2 – Représentation des deux premières composantes principales d'un nuage de points.

La première composante principale, représentée par :

$$\mathbf{z}_1 = \sum_{j=1}^P \phi_j \mathbf{x}_j$$

où  $\phi_j$  sont des constantes, forme un axe  $u$  tel que la variance du nuage d'observations autour de l'axe soit maximale. Il est possible de démontrer que  $u$  correspond au premier vecteur propre de la matrice de corrélation de  $\mathbf{X}$ <sup>4</sup>. La deuxième composante principale est déterminée de manière similaire : en imposant l'orthogonalité à  $u$ , on recherche l'axe  $v$  qui explique le maximum de la variance du nuage des observations. Le processus est répété afin de construire les  $M$  composantes principales qui définissent le nuage d'observations, où la dernière composante principale,  $\mathbf{z}_M$ , est la combinaison linéaire de variance maximale qui soit orthogonale à  $\mathbf{z}_1, \dots, \mathbf{z}_{M-1}$  (Friedman et al., 2001).

La seconde étape consiste à régresser la variable réponse sur les  $M$  composantes principales préalablement construites, c'est-à-dire :

$$\hat{y}_t = \hat{\theta}_1 \mathbf{z}_1 + \dots + \hat{\theta}_M \mathbf{z}_M$$

3. La figure est inspirée de Lever et al. (2017).

4. Le lecteur intéressé trouvera la démonstration dans Friedman et al. (2001).

où  $\hat{\theta}_m$  ( $m = 1, \dots, M$ ) sont estimés à partir de la méthode des MCO. Si  $M = P$ , l'estimateur obtenu correspond à l'estimateur des MCO du modèle contenant l'ensemble des prédicteurs. L'intérêt de la PCA survient lorsque  $M < P$ . Les composantes principales de variances faibles sont alors éliminées, corrigeant ainsi le problème de colinéarité et améliorant du même coup la qualité de prévision du modèle. La PCA n'engendre toutefois pas de sélection de variables, contrairement aux méthodes LASSO et Elastic Net, puisque les  $M$  composantes principales demeurent construites à partir des  $P$  prédicteurs initiaux, d'où sa similarité avec la méthode Ridge (James et al., 2013). Un second désavantage de la régression sur composantes principales est sa complexité d'interprétation. En effet, puisque chaque composante principale représente une combinaison linéaire des  $P$  prédicteurs, il est difficile de déterminer intuitivement de quelle manière chaque prédicteur affecte la variable réponse.

La valeur prédite de la variable réponse au temps  $t+i$  est calculée comme une combinaison linéaire des  $M$  composantes principales et des coefficients estimés par MCO qui leur sont associés.

## 4.2.2 Méthodes non paramétriques

### Les arbres de régression

Les méthodes présentées jusqu'à maintenant sont limitées par leur hypothèse sur la linéarité de la relation entre les prédicteurs et la variable réponse. Cette section présente une méthode simple, efficace dans la sélection de variables, la lisibilité des résultats et la prise de décision, et qui ne nécessite aucune hypothèse a priori sur la forme des données : les arbres de régression. Un arbre de régression constitue une méthode non paramétrique d'estimation sous la forme d'un diagramme représentant les résultats possibles d'une série de choix interconnectés. Différents algorithmes ont été développés afin de modéliser les arbres de régression<sup>5</sup>. Cette section décrit l'algorithme CART, proposé par Breiman et al. (1984), qui figure parmi les plus utilisés dans la littérature.

---

5. Parmi les plus populaires, on retrouve AID, THAID, M5 et CART.

Un arbre de régression de type CART consiste en un partitionnement récursif et dyadique de l'espace des  $P$  prédicteurs en  $J$  sous-régions distinctes, mutuellement exclusives et de forme rectangulaire (tel qu'illustre la figure 4.3) (James et al., 2013).

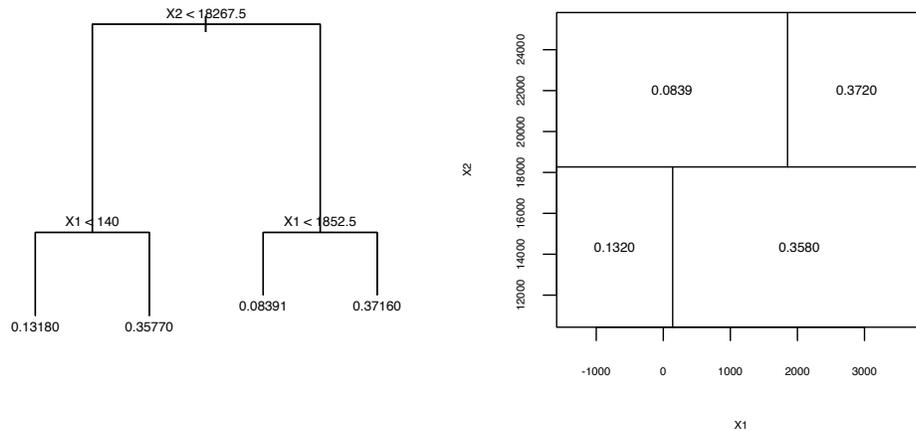


FIGURE 4.3 – Représentation d'un arbre de régression de type CART. Les noeuds et les branches de l'arbre de régression sont représentés dans la figure de gauche. Chaque noeud engendre deux branches fils, où la branche fils de droite signifie que la règle de division est vraie, et vice-versa. À chaque noeud terminal est associée une valeur qui représente la prévision. Le même arbre de régression est représenté dans la figure de droite comme un partitionnement de l'espace des prédicteurs  $x_1$  et  $x_2$ .

L'algorithme crée, à chaque itération, un noeud constitué d'un prédicteur  $x_p$  et d'une règle de décision,  $s$ . Le noeud est construit de manière à ce que la division d'une région en deux sous-régions mène à une réduction maximale de la SCR. En particulier, soit deux sous-régions :

$$R_1(p, s) = \{x | x_p < s\} \quad \text{et} \quad R_2(p, s) = \{x | x_p \geq s\}$$

alors le prédicteur  $x_p$  et la règle de décisions  $s$  sont sélectionnés afin de minimiser :

$$\sum_{p: x_p \in R_1(p, s)} (y_t - \hat{y}_{R_1})^2 + \sum_{p: x_p \in R_2(p, s)} (y_t - \hat{y}_{R_2})^2$$

où  $\hat{y}_{R_1}$  représente la moyenne des variables réponses comprises dans la région  $R_1$  et  $\hat{y}_{R_2}$ , la moyenne des variables réponses comprises dans la région  $R_2$ . Après chaque division, un nouveau noeud est déterminé à partir des noeuds disponibles. Il s'agit donc d'une stratégie

pas-à-pas hiérarchique : l'algorithme crée les divisions à partir des noeuds qui découlent de divisions précédemment établies. La croissance de l'arbre de régression maximal s'interrompt lorsque les régions ne comportent qu'une seule observation ou qu'aucune division ne puisse diminuer davantage la SCR.

L'arbre ainsi construit souffre généralement d'un problème de surajustement, d'où la nécessité d'utiliser une procédure d'émondage. La procédure d'émondage (Breiman et al., 1984) consiste à créer une suite de sous-arbres en fonction d'un paramètre d'émondage  $\kappa$ , parmi lesquels sera désigné l'arbre optimal. En particulier, un sous-arbre est construit en minimisant :

$$\sum_{m=1}^{|T|} \sum_{t: x_t \in R_m} (y_t - \hat{y}_{R_m})^2 + \kappa |T| \quad (4.2)$$

où  $|T|$  est le nombre de noeuds terminaux du sous-arbre  $T$ ,  $R_m$  représente la région correspondante au  $m^e$  noeud,  $\hat{y}_{R_m}$  est la moyenne des observations de la variable réponse associée à  $R_m$  et  $\kappa$  est un paramètre d'émondage qui fait l'arbitrage entre la complexité de l'arbre et sa concordance avec les données. Si  $\kappa = 0$ , l'arbre de régression maximal minimise (4.2). L'augmentation de la valeur de  $\kappa$  implique qu'on pénalise davantage la complexité de l'arbre : certains noeuds deviennent ainsi superflus et seront supprimés, d'où la création d'un nouvel arbre. La démarche est ainsi répétée afin de créer une suite de sous-arbres en fonction de différentes valeurs du paramètre d'émondage, desquels sera sélectionné l'arbre optimal.

La prévision de la variable réponse  $\hat{y}_{t+i}$  se calcule à partir de la moyenne des variables réponses comprises dans la région  $j$  où se situe  $y_{t+i}$ , c'est-à-dire :

$$x \in R_j \Rightarrow \hat{y}_{t+i} = \gamma_j \quad (4.3)$$

où  $\gamma_j$  représente la moyenne des  $y$  situés dans la région  $j$  (Friedman et al., 2001). Par conséquent, tous les individus se trouvant dans la même région  $R_j$  se font assignés la même prévision. Certains trouveront plus simple d'interpréter la valeur prédite d'un arbre de régression sous la forme d'un modèle de régression linéaire multiple, où les prédicteurs sont les interactions de différentes fonctions indicatrices. Par exemple, les prévisions de l'arbre

représenté dans la figure 4.3 peuvent être décrites par :

$$\begin{aligned}\hat{y}_{t+i} = & \gamma_1 \mathbb{1}(X_2 > 18267.5)(X_1 < 1852.5) + \gamma_2 \mathbb{1}(X_2 > 18267.5)(X_1 > 1852.5) \\ & + \gamma_3 \mathbb{1}(X_2 < 18267.5)(X_1 < 140) + \gamma_4 \mathbb{1}(X_2 < 18267.5)(X_1 > 140)\end{aligned}$$

où  $\mathbb{1}()$  est la fonction indicatrice qui est égale à 1 lorsque l'argument est vrai et  $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (0, 08; 0, 37; 0, 13; 0, 36)$ .

## Les méthodes d'ensemble aléatoires : le bagging et les forêts aléatoires

Les arbres de régression bénéficient d'une simplicité et d'une facilité d'interprétation des résultats. Parmi ses inconvénients toutefois, on retrouve le fait que l'algorithme se fonde sur une recherche locale et génère des optima locaux, plutôt qu'un optimum global. Également, les arbres de régression sont sensibles au bruit et aux fluctuations d'échantillons, puisque leur construction est basée sur une structure hiérarchique de telle manière qu'une erreur de division au début de l'arbre se propage dans l'ensemble de la construction. Les inconvénients de l'arbre de régression en font une méthode d'apprentissage instable, à variance élevée (Breiman, 1996). Les méthodes d'ensemble aléatoire, telles que le bagging (Breiman, 1996) et les forêts aléatoires (Breiman, 2001), proposent de diminuer la variance d'un arbre de régression individuel en estimant aléatoirement et en agrégeant, par une moyenne, plusieurs arbres de régression.

Soit  $\hat{f}_1, \dots, \hat{f}_B$ ,  $B$  arbres de régression identiquement distribués, la variance de la moyenne des  $B$  estimateurs est donnée par :

$$\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2$$

où  $\rho$  est un coefficient positif de corrélation et  $\sigma^2$  correspond à la variance d'un arbre individuel (Friedman et al., 2001). Lorsque  $B$  est suffisamment grand, le second terme de l'équation disparaît. Du premier terme de l'équation, il est possible de conclure que le gain de la procédure d'agrégation se chiffre par la corrélation des estimateurs individuels : plus ils sont décorrélés, plus la variance de l'estimateur agrégé est réduite.

## Bagging

Le bagging atténue l'interdépendance des arbres de régression individuels à travers la modification de l'échantillon d'apprentissage, par l'emploi de l'échantillonnage bootstrap (Efron et Tibshirani, 1994). Les répliques bootstrap sont générées à partir de  $N$  tirages avec remise de l'échantillon d'apprentissage initial, chaque observation ayant ainsi la même probabilité d'être tirée à chaque tour. Les échantillons bootstrap ont la même taille que l'échantillon d'apprentissage ( $N$ ), mais leur composition diffère les uns des autres, car certaines observations peuvent être répétées. Par conséquent, l'utilisation de répliques bootstrap introduit une source d'aléa, puisque le tirage est aléatoire, permettant ainsi de décorréler les estimateurs individuels.

## Les forêts aléatoires

L'inconvénient du bagging est que malgré l'introduction d'un aléa par l'emploi de l'échantillonnage bootstrap, les arbres de régression construits peuvent demeurer fortement corrélés s'il existe, dans les données, un prédicteur "fort" qui provoque une réduction importante de la SCR relativement à celle engendrée par les autres prédicteurs. Dans une telle situation, l'ensemble des arbres de régression construits seront constitués du même noeud initial et peuvent, par conséquent, être très similaires les uns aux autres.

Vis-à-vis cet inconvénient, Breiman (2001) introduit la méthode des forêts aléatoires, une technique où la décorrélation des estimateurs individuels est basée sur la perturbation du processus d'apprentissage à travers la modification de l'échantillon d'apprentissage, mais également à travers la sélection aléatoire de prédicteurs. Précisément, les arbres de régression sont construits à partir des répliques bootstrap, selon une variation de la méthode de construction d'arbres décrite à la sous-section 4.2.2. En effet, à chaque étape, la recherche du noeud optimal est devancée par le tirage aléatoire d'un sous-ensemble de  $V$  prédicteurs. Ainsi, on sélectionne parmi le sous-ensemble de prédicteurs, le prédicteur et la règle de division qui engendre une réduction maximale de la SCR. Cet ajout d'aléa permet ainsi de différencier davantage les  $B$  arbres de régression estimés et, par conséquent, d'en diminuer leur corrélation.

## Visualisation des résultats

Soit  $E_N = (\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)$  l'échantillon d'apprentissage initial et  $\theta_b = \theta_b(E_N)$  l'échantillon bootstrap à l'étape  $b$  où  $b = 1, \dots, B$ , l'estimateur final des méthodes d'ensemble aléatoires représente la moyenne des estimateurs individuels construits à partir des  $B$  échantillons bootstrap, soit :

$$\hat{f}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^B \hat{f}(\mathbf{x}, \theta_b)$$

où  $\hat{f}(\mathbf{x}, \theta_b)$  représente l'arbre de régression estimé à l'étape  $b$  à partir d'un échantillon bootstrap dans le cas du bagging, ou de la combinaison d'un échantillon bootstrap et d'une sélection aléatoire des prédicteurs, dans le cas de la forêt aléatoire. La valeur prédite de  $\hat{y}_{t+i}$  consiste donc à la moyenne des  $B$  prévisions de  $\hat{y}_{t+i}$  obtenus à partir de chaque arbre de régression individuel de l'estimateur agrégé.

Les estimateurs finaux des méthodes d'ensemble aléatoires, puisqu'ils correspondent à l'agrégation de plusieurs estimateurs individuels, constituent toutefois des "boîtes noires" dont les résultats ne peuvent être facilement visualisés. Pour pallier ce manque d'interprétation, Breiman (2001) propose le critère *Mean Decrease Accuracy* (MDA), qui quantifie l'importance de chacun des  $P$  prédicteurs à travers une permutation aléatoire de ses valeurs.

Soit  $OOB_b$  l'échantillon out-of-bag du  $b^{\text{ième}}$  arbre de régression, c'est-à-dire l'échantillon formé des observations qui ne sont pas comprises dans la réplication bootstrap de l'étape  $b$ . Il est possible de calculer les prévisions des valeurs de la variable réponse contenues dans  $OOB_b$  à partir de l'arbre de régression  $\hat{f}(\cdot, \theta_b)$  et d'en déterminer l'erreur out-of-bag, soit :

$$E_{OOB_b} = \frac{1}{L} \sum_{t \in OOB_b} (\hat{f}(x_t, \theta_b) - y_t)^2 \quad (4.4)$$

où  $L$  représente la taille de l'échantillon  $OOB_b$ . En perturbant aléatoirement les valeurs d'un prédicteur  $j$  dans  $OOB_b$ , l'erreur out-of-bag associée à l'arbre  $\hat{f}(\cdot, \theta_b)$  devient :

$$E_{OOB_b}^j = \frac{1}{L} \sum_{t \in OOB_b} (\hat{f}(x_t^j, \theta_b) - y_t)^2 \quad (4.5)$$

où  $x_t^j$  représente les valeurs des prédicteurs de l'échantillon  $OOB_b$  dans lequel le prédicteur  $j$  a été perturbé. Le test de permutation a pour hypothèse nulle qu'une perturbation des

observations d'un prédicteur important augmente l'erreur de prévision. En d'autres mots, la différence entre 4.5 et 4.4 devrait être élevée si le prédicteur a de l'importance dans la prévision de la variable réponse. Le critère MDA quantifie ainsi l'importance d'un prédicteur  $j$  ( $Imp(X_j)$ ) en moyennant cette différence à travers les  $B$  arbres de régression construits de l'estimateur agrégé, c'est-à-dire :

$$Imp(X_j) = \frac{1}{B} \sum_{b=1}^B (E_{OOB_b}^j - E_{OOB_b}).$$

### Les méthodes d'ensemble additives : le boosting

Les méthodes d'ensemble additives telles que le boosting, sont des méthodes qui, similairement aux méthodes d'ensemble aléatoires, visent à réduire la variance d'un estimateur instable à travers la construction et la combinaison d'une famille d'estimateurs. La particularité du boosting repose toutefois sur sa démarche séquentielle : plutôt que d'estimer  $B$  modèles indépendants, on estime à chaque étape une version adaptative du modèle estimé à l'étape précédente. L'estimateur boosting se concentre ainsi davantage sur les observations difficiles à prédire, comparativement aux estimateurs bagging et forêts aléatoires (Friedman et al., 2001). Le boosting s'applique aux arbres de régression sous la forme d'un *functional gradient descent* (FGD) (Friedman, 2001).

Soit  $l$  la fonction de perte quadratique moyenne, qui représente la perte en performance prédictive causée par la sous-optimalité d'un arbre de régression. Le FGD est une technique d'optimisation numérique pour minimiser  $l$  en ajoutant, à chaque étape, un nouvel arbre de régression de manière à engendrer une réduction optimale de la fonction de perte (Elith et al., 2008). La première étape de la procédure consiste à construire l'arbre de régression qui provoque une réduction maximale de la fonction de perte  $l$ , en limitant le nombre de noeuds de l'arbre construit à l'aide d'un paramètre de complexité de l'arbre ( $cc$ ). De cet arbre sont calculés les résidus de prévisions,  $\hat{r}_{t1}$ . À la deuxième étape, l'arbre de régression est construit de la même manière, mais sur les  $\hat{r}_{t1}$ , plutôt que sur les  $y_t$ . Ce nouvel arbre pourrait ainsi être grandement différent du premier et contenir différents prédicteurs et règles de décision. À cette étape, l'estimateur boosting représente une combinaison linéaire des

arbres de régression estimés à la première et seconde étape, où la pondération spécifique avec laquelle les arbres de régression sont ajoutés dans le modèle prédictif est déterminée selon le taux d'apprentissage ( $\zeta$ ). Compris entre 0 et 1, le paramètre  $\zeta$  correspond à une pénalisation vis-à-vis l'ajout d'un nouvel arbre de régression, qui, lorsque sa valeur est petite, ralentie la convergence du processus. En particulier, Friedman (2001) a démontré empiriquement qu'un taux d'apprentissage inférieur à 0,1 permet d'améliorer la qualité de prévision de l'estimateur boosting. De l'estimateur construit à la seconde étape peuvent être calculés les résidus de prévisions  $\hat{r}_{t2}$ . La troisième étape consiste alors à estimer, à partir de la même procédure, un arbre de régression sur  $\hat{r}_{t2}$ , permettant de construire l'estimateur boosting comme une combinaison linéaire des arbres de régression construits lors des trois étapes. La démarche se répète ensuite en estimant un arbre de régression sur les résidus de prévisions de l'estimateur de la troisième, et ainsi de suite, jusqu'à ce que  $Z$  étapes aient été complétées.

L'estimateur boosting final est donc construit à partir d'une procédure pas-à-pas, c'est-à-dire que les arbres de régression précédemment estimés demeurent inchangés au fur à mesure que le modèle s'agrandit. Soit  $\hat{y}_t^{(z)}$ , la valeur prédite à l'étape  $z$  ( $z = 1, \dots, Z$ ), l'estimateur boosting est donc construit selon la démarche suivante (Chen, 2014) :

$$\begin{aligned}\hat{y}_t^{(1)} &= f_1(\mathbf{x}) \\ \hat{y}_t^{(2)} &= f_1(\mathbf{x}) + f_2(\mathbf{x}) = \hat{y}_t^{(1)} + \zeta f_2(\mathbf{x}) \\ \hat{y}_t^{(3)} &= f_1(\mathbf{x}) + f_2(\mathbf{x}) + f_3(\mathbf{x}) = \hat{y}_t^{(2)} + \zeta f_3(\mathbf{x}) \\ &\dots \\ \hat{f}(\mathbf{x}) &= \sum_{z=1}^Z \hat{f}_z(\mathbf{x}) = \hat{y}_t^{(Z-1)} + \zeta f_Z(\mathbf{x}).\end{aligned}$$

où  $\zeta$  correspond au taux d'apprentissage et  $\hat{f}(\mathbf{x})$  représente l'estimateur final, c'est-à-dire la combinaison linéaire des arbres de régression construits à chacune des  $Z$  étapes, d'où peut être calculée la valeur prédite de la variable réponse au temps  $t+i$ . Le boosting présente plusieurs avantages, notamment sa démarche séquentielle qui permet à l'algorithme de se concentrer sur les erreurs causées par le modèle précédent et, de ce fait, sur les observations difficiles à prédire. Également, son intérêt provient du fait qu'il puisse simultanément réduire la variance et le biais de l'arbre de régression individuel. Ce n'est pas le cas avec les méthodes bagging

et forêts aléatoires où les arbres construits sont identiquement distribués et, par conséquent, l'espérance de l'estimateur agrégé est identique à l'espérance de l'arbre individuel (Friedman et al., 2001).

L'estimateur final n'est toutefois pas explicite et doit être visualisé à partir du critère MDA, similairement aux estimateurs bagging et forêts aléatoires. De plus, l'estimation du boosting est davantage complexe, comparativement aux modèles décrits précédemment, puisqu'elle requiert la détermination de trois paramètres : le nombre d'étapes à compléter ( $Z$ ) où chaque étape correspond à la construction d'un nouvel arbre, le taux d'apprentissage ( $\zeta$ ) qui détermine de quelle manière ses derniers sont inclus dans le modèle final, et la complexité de l'arbre ( $cc$ ) c'est-à-dire que le nombre maximal de noeuds compris dans chaque arbre construit. Étant donné la possibilité de surapprentissage de l'estimateur boosting, ces paramétrisations sont susceptibles d'influencer sa performance prédictive.

# Chapitre 5

## Estimation et résultats

Les données et les techniques utilisées étant définies, la prochaine étape consiste à estimer et comparer les différents modèles prédictifs. Cette section détaille les étapes de la préparation des échantillons, ainsi que de la paramétrisation et de l'estimation des modèles étudiés pour les prévisions de la demande et du prix de l'électricité dans le marché de l'Ontario. Également, on y présente les comparaisons des performances prédictives des modèles, ainsi que les conclusions qui en découlent.

### 5.1 Préparation des échantillons

Les performances prédictives des modèles étudiés dépendent de plusieurs facteurs, dont la qualité des données à partir desquelles ils sont construits (Kotsiantis et al., 2006). La préparation des échantillons constitue, par conséquent, une étape importante préalable à l'estimation. Les valeurs manquantes, les données redondantes et la présence de bruit doivent être traitées, puisqu'elles rendent difficile la phase d'apprentissage par l'algorithme, menant généralement à des prévisions imprécises, ou inexactes. Notre démarche nécessite également de diviser les échantillons en deux parties : apprentissage et test, afin d'être en mesure de juger la performance prédictive des modèles étudiés à partir d'une validation hors échantillon.

## Valeurs manquantes

La base de données initialement recueillie contient 384 valeurs manquantes. Puisque cela représente moins de 1% des données de l'échantillon total, les observations qui contiennent des valeurs manquantes ont été supprimées.

## Valeurs aberrantes

Une donnée aberrante se décrit comme une valeur paraissant dévier de manière importante des autres valeurs d'un même phénomène, semblant ainsi incohérente avec le reste des observations (Bremer, 1995). Différentes approches peuvent être utilisées afin de traiter les valeurs aberrantes dans l'estimation de modèles prédictifs. D'un côté, il peut sembler nécessaire de conserver ce type de données dans l'échantillon afin de bâtir le modèle sous les conditions les plus réalistes possible. D'un autre côté, les algorithmes d'apprentissage sont sensibles aux données et à leur distribution : la suppression des données aberrantes permettrait ainsi d'augmenter la robustesse des résultats (Kotsiantis et al., 2006).

Dans le cadre de cette recherche, les valeurs aberrantes ont été uniquement supprimées des échantillons utilisés pour l'analyse prédictive des prix de l'électricité. Les modèles prédictifs de la demande seront, par conséquent, construits à partir de l'ensemble des données disponibles. Notre principale motivation à agir ainsi est que les valeurs extrêmes des prix sont généralement causées par des défaillances techniques (O'Mahoney et Denny, 2011) ou par l'adoption de comportements stratégiques non conventionnels par les producteurs d'électricité, plutôt que par le processus normal de formation des prix. Elles sont, par conséquent, généralement non répétitives (Guirguis et Felder, 2004), et rendent difficile l'obtention de résultats raisonnables à partir des modèles prédictifs étudiés. En nous basant sur la littérature existante, nous excluons les observations dont le prix observé est inférieur et supérieur au seuil de  $3 \cdot IQR$ , où  $IQR$  correspond à l'écart entre le premier quartile et le troisième quartile des données des prix de l'électricité<sup>1</sup>. De cette manière, environ 2% des observations sont éliminées de l'échantillon initial.

---

1. Voir, entre autres, Benth et al. (2008).

## Échantillon d'apprentissage et échantillon test

Les pouvoirs prédictifs de nos modèles seront calculés à partir d'une validation hors échantillon. Pour ce faire, nous utiliserons l'approche de l'ensemble de validation qui nécessite de diviser nos échantillons en deux parties : apprentissage et test. L'échantillon d'apprentissage est utilisé pour l'estimation des modèles. C'est également à partir de cet échantillon que seront définies les paramétrisations optimales. L'échantillon test, quant à lui, permet de vérifier et de quantifier le pouvoir prédictif des modèles étudiés. En effet, le modèle bâti à partir de l'échantillon d'apprentissage permet de calculer, à partir des prédicteurs de l'échantillon test, les valeurs prédites de la variable réponse. Celles-ci seront ensuite comparées à leurs valeurs réelles. L'utilisation de cette approche a été préférée aux autres méthodes communes dans la littérature (la validation croisée *Leave-One-Out* et la validation croisée à  $K$  groupes) étant donné que ces dernières nécessitent généralement un temps de calcul très élevé.

Comme mentionné dans la présentation des données, les échantillons doivent d'abord être divisés en deux périodes : "avant" et "après", selon le changement de la réglementation des prix dans le marché de l'électricité ontarien et le changement de régime au niveau des importations nettes. L'échantillon d'apprentissage de la période "avant" se réfère aux données du 1<sup>er</sup> mai 2010 au 31 décembre 2011. Les modèles seront ensuite validés à partir des données de la période du 1<sup>er</sup> janvier 2012 au 31 décembre 2012. Pour la période "après", les modèles seront estimés sur les données couvrant la période du 1<sup>er</sup> janvier 2013 au 31 décembre 2014, puis validés sur les données de la période du 1<sup>er</sup> janvier 2015 au 31 décembre 2015.

Le tableau 5.1 présente quelques statistiques descriptives des échantillons d'apprentissage et test utilisés. Des résultats, il est possible d'observer que la demande d'électricité moyenne à travers les différents échantillons d'apprentissage et test sont comparables. Les prix moyens sont, quant à eux, plus hétérogènes. Les écarts-types observés sont également élevés, démontrant à nouveau la grande volatilité des prix de l'électricité en Ontario.

<i>Avant</i>					
	Période étudiée	Analyse prédictive de la demande		Analyse prédictive du prix	
		Nombre	Demande totale horaire	Nombre	Prix horaire
		d'observations	moyenne (MWh)	d'observations	moyen (\$CAD)
Apprentissage	Mai 2010 à	14352	17752	14 084	32,73
	Décembre 2011		(2689,54)		(12,27)
Test	Janvier 2012 à	8688	17741	8535	23,00
	Décembre 2012		(2563,27)		(10,51)
<i>Après</i>					
	Période étudiée	Analyse prédictive de la demande		Analyse prédictive du prix	
		Nombre	Demande totale horaire	Nombre	Prix horaire
		d'observations	moyenne (MWh)	d'observations	moyen (\$CAD)
Apprentissage	Janvier 2013 à	17520	18083	16823	23,12
	Décembre 2014		(2249,09)		(16,58)
Test	Janvier 2015	8759	18154	8615	18,99
	Décembre 2015		(2400,81)		(16,17)

Tableau 5.1 – Quelques statistiques descriptives des échantillons apprentissages et test. Les écarts-type sont présentés entre parenthèses.

## 5.2 Détails des estimations

Cette sous-section présente les détails des estimations des modèles de prévision de la demande et du prix de l'électricité, développés à partir de nos échantillons d'apprentissage. Les modèles prédictifs de la demande d'électricité sont estimés pour un horizon temporel d'une heure. Les modèles prédictifs des prix sont, quant à eux, estimés pour un horizon temporel de 3 heures étant donné que, tel que mentionné précédemment, les producteurs ontariens sont restreints à modifier leur offre jusqu'à 2 heures précédant l'heure  $t$  de production d'électricité. La prévision du prix de l'électricité doit ainsi être calculée un minimum de 3 heures à l'avance, afin qu'elle puisse être utile aux producteurs dans la maximisation de leur profit. L'ensemble des estimations s'est effectué avec le logiciel statistique R. Les codes nécessaires à leur implémentation sont disponibles à l'annexe 1.

## 5.2.1 Modèles conventionnels

### Régressions linéaires multiples

Les modèles de régressions linéaires multiples utilisés dans le cadre de notre recherche sont :

$$\begin{aligned}
 y_{dem,t+1} = & \beta_0 + \sum_{g \in G} \beta_{1,g} \cdot CDD_{g,t} + \sum_{g \in G} \beta_{2,g} \cdot HDD_{g,t} + \beta_3 \cdot Weekday_t \\
 & + \beta_4 \cdot q_{imp,t} + \beta_5 \cdot Crisis_t + \sum_{y \in Y} \psi_y \cdot Year_y \\
 & + \sum_{s \in S} \gamma_s \cdot Season_s + \sum_{h \in H} \omega_h \cdot ToD_h
 \end{aligned} \tag{5.1}$$

et

$$\begin{aligned}
 y_{prix,t+3} = & \beta_0 + \sum_{g \in G} \beta_{1,g} \cdot CDD_{g,t} + \sum_{g \in G} \beta_{2,g} \cdot HDD_{g,t} + \beta_3 \cdot Weekday_t \\
 & + \beta_4 \cdot q_{imp,t} + \beta_5 \cdot Crisis_t + \beta_6 \cdot dem_{ON,t} + \sum_{y \in Y} \psi_y \cdot Year_y \\
 & + \sum_{s \in S} \gamma_s \cdot Season_s + \sum_{h \in H} \omega_h \cdot ToD_h
 \end{aligned} \tag{5.2}$$

où  $G = \{ON, QC, MB, MI, MN, NY\}$  désigne l'ensemble des états et provinces interconnectés au réseau ontarien, ainsi que la province de l'Ontario elle-même. Les ensembles  $Y$ ,  $S$  et  $H$  représentent respectivement les années étudiées, les saisons annuelles et les périodes quotidiennes. Les prédicteurs  $Year_y$ ,  $Season_s$  et  $ToD_h$  correspondent aux variables indicatrices de la saisonnalité annuelle, saisonnière et quotidienne. Le prédicteur  $q_{imp,t}$  représente le niveau d'importations nettes en Ontario au temps  $t$ . Le prédicteur  $Crisis_t$ , quant à lui, montre si l'Ontario se trouve en période d'envolée de prix ou non. Ces périodes, rappelons-le, surviennent à deux moments : hiver 2014 et hiver 2015, soit uniquement dans l'échantillon d'apprentissage et test de la période "après". Également, dans le modèle des prix, le prédicteur  $dem_{ON,t}$  correspond à la demande totale d'électricité en Ontario, à l'heure de la journée

$t^2$ . L'ensemble des coefficients :  $\beta$ ,  $\psi$ ,  $\gamma$  et  $\omega$ , sont estimés par la méthode des MCGF afin de corriger l'hétéroscédasticité.

L'ajout de prédicteurs supplémentaires dans les modèles présentés ci-dessus est susceptible d'occasionner un problème de colinéarité. Dans un tel cas, la technique des MCGF ne constitue pas une méthode d'estimation appropriée dans un contexte prédictif, puisque tel que mentionné précédemment, elle engendre un modèle à variance élevée, dont le pouvoir prédictif est généralement faible. Également, l'estimation par MCGF d'un modèle de régression linéaire à nombre élevé de prédicteurs peut causer un problème de surapprentissage. L'utilisation des MCGF dans le cadre de notre recherche sera, par conséquent, limitée à l'estimation des modèles (5.1) et (5.2).

## Régressions linéaires avec erreur ARIMA

Nos modèles de régressions linéaires avec erreur ARIMA correspondent aux modèles (5.1) et (5.2), auxquels sont ajoutés un processus ARIMA( $p, d, q$ ) décrivant le comportement du terme d'erreur de régression  $n_t$ , représenté par :

$$\eta_t = \phi_1 \eta_{t-1} + \dots + \phi_p \eta_{t-p} - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}$$

où les  $\epsilon$  sont des bruits blancs.

L'estimation des modèles de régressions linéaires avec erreur ARIMA est effectuée selon la démarche proposée par Hyndman et Athanasopoulos (2014). Préalablement à l'estimation, la stationnarité des prédicteurs et de la variable réponse est vérifiée; des différentiations pouvant être appliquées lorsque nécessaire. Puis, en définissant arbitrairement la forme du processus ARIMA( $p, d, q$ ) suivi par le terme d'erreur  $\eta_t$ , les coefficients de régression ( $\beta$ ,  $\psi$ ,  $\gamma$ ,  $\phi$  et  $\theta$ ) sont estimés à partir de la méthode du maximum de vraisemblance. Les résidus sont, ensuite, récupérés du modèle construit, permettant de définir une forme plus appropriée du processus ARIMA( $p, d, q$ ) et d'estimer à nouveau les coefficients  $\phi$  et  $\theta$ . La démarche est

---

2. Les valeurs passées de la variable réponse et des prédicteurs auraient également pu être incluses dans nos modèles prédictifs. Nous restreignons toutefois notre analyse uniquement à l'utilisation des valeurs contemporaines des prédictifs. Tel que nous le verrons, les résultats obtenus sont généralement comparables à ceux obtenus, dans la littérature, à partir de modèles prédictifs incluant les valeurs passées de la variable réponse et des prédictifs.

ainsi répétée en définissant différentes formes initiales du processus suivi par le terme d'erreur  $\eta_t$ . Le choix des paramètres  $p$  et  $q$  est enfin effectué en minimisant le critère d'information bayésien (BIC) (Schwarz et al., 1978).

## 5.2.2 Modèles de l'apprentissage statistique

### Validation croisée

Tel que discuté à la section précédente, les performances prédictives des modèles tirés de l'apprentissage statistique reposent en grande partie sur leurs paramétrisations, puisque celles-ci représentent généralement un compromis biais-variance. Par exemple, on retrouve ce compromis dans les forêts aléatoires à travers le choix de la taille des sous-ensembles de prédicteurs parmi lesquels sont sélectionnés les noeuds des arbres de régression construits ( $V$ ). En effet, une valeur élevée de  $V$  engendre généralement une corrélation importante entre les arbres estimés et, par conséquent, le biais de l'estimateur agrégé est faible, mais sa variance, élevée. Une valeur faible de  $V$  cause, quant à elle, les phénomènes inverses : l'estimateur agrégé est de variance faible, mais présente un biais élevé (James et al., 2013). Une paramétrisation inadéquate est ainsi susceptible d'augmenter l'erreur quadratique moyenne (MSE) de prévision d'un modèle, même si celui-ci représente un choix approprié dans la prévision de la variable réponse.

Adopter une démarche de paramétrisation efficace constitue donc une étape importante, préalable à l'estimation des modèles tirés de l'apprentissage statistique. Vis-à-vis cet objectif, nous utilisons la validation croisée à  $K$  groupes, cette méthode figurant parmi les plus utilisées dans la littérature. Cette technique, basée sur une validation hors échantillon, est employée, car elle permet notamment le contrôle du problème de surapprentissage dans la sélection des paramètres optimaux (James et al., 2013).

La validation croisée à  $K$  groupes nécessite de diviser l'échantillon d'apprentissage en  $K$  groupes de tailles semblables ; la composition des groupes étant gardée fixe à travers l'ensemble des étapes de la démarche. Un modèle est d'abord estimé à partir de  $K - 1$  groupes, en fixant le paramètre étudié à une certaine valeur. La validation hors échantillon

s'effectue sur le groupe restant : les valeurs prédites sont calculées à partir de ses données, et la MSE de prévision est évaluée en comparant les valeurs prédites aux valeurs réelles de la variable réponse. L'estimation, la validation et le calcul de la MSE de prévision sont ainsi répétés  $K$  fois, en gardant fixe la valeur du paramètre étudié, de manière à ce que chaque groupe ait servi à la validation. La moyenne des MSE de prévision associées à cette première valeur du paramètre étudié correspond alors à :

$$CV_1 = \frac{1}{K} \sum_{k=1}^K MSE_k.$$

La même démarche se réitère ensuite pour différentes valeurs du paramètre étudié. La paramétrisation optimale correspond à la valeur minimisant  $CV_i$ , où  $i = 1, \dots, I$  représente l'itération du processus associée à une valeur spécifique du paramètre.

La valeur de  $K$  est déterminée en fonction du temps de calcul disponible : une valeur de  $K$  élevée, se rapprochant du nombre d'observations comprises dans l'échantillon d'apprentissage (technique appelée le *Leave-One-Out* (James et al., 2013)) améliore certes la précision de la procédure de paramétrisation, mais demande également un temps de calcul élevé. En fixant une valeur de  $K$  faible, le temps de calcul diminue, mais également la précision du processus de paramétrisation. Par conséquent, en pratique, le nombre de groupes utilisés est fréquemment fixé à  $K = 10$  (James et al., 2013) ; cette stratégie sera également adoptée dans le cadre de ce mémoire.

Le tableau 5.2 présente les différents paramètres étudiés à travers nos modèles prédictifs, ainsi que l'intervalle de valeurs pour lesquelles la validation croisée à 10 groupes a été appliquée. Remarquez que le nombre de répliques bootstrap ( $B$ ) employées dans le bagging et les forêts aléatoires ne figure pas parmi les paramètres étudiés. En effet,  $B$  ne représente pas un paramètre critique dans la performance prédictive du modèle estimé. En particulier, le choix d'un  $B$  trop élevé n'engendre pas le problème de surapprentissage de l'estimateur. Breiman (1996) qualifie ainsi le choix de sa valeur comme arbitraire : elle est généralement choisie la plus large possible en fonction du type de la variable réponse et du temps de calcul

disponible<sup>3</sup>.

Modèles	Paramètre(s)	Intervalles des valeurs étudiées
Ridge, LASSO	Paramètre de régularisation ( $\lambda$ )	[0,001 ; 1000,00]
Elastic Net	Paramètres de régularisation ( $\lambda$ ) ( $\alpha$ )	[0,001 ; 1000,00] [0,01, 1]
RCP	Nombre de composante(s) principale(s) ( $M$ )	[1, $P$ ]
Arbre de régression	Paramètre d'émondage ( $\kappa$ )	[0,01 ; 0,36]
Forêts aléatoires	Taille des sous-ensembles de prédicteurs ( $V$ )	[1, $P$ ]
Boosting	Nombre d'arbre(s) construit(s) ( $Z$ )	[0, 2500]
	Paramètre de complexité de l'arbre ( $cc$ )	[1, 10]
	Taux d'apprentissage ( $\zeta$ )	[0,05 ; 0,01]

Tableau 5.2 – Paramétrisation des modèles étudiés. Le terme  $P$  réfère au nombre initial de prédicteurs.

## Régressions pénalisées

Les méthodes de régressions pénalisées sont avantagées, vis-à-vis des méthodes conventionnelles d'estimation des modèles de régressions linéaires, du fait qu'elles contrôlent le problème de colinéarité et de surapprentissage à travers l'usage de termes de pénalité qui contraignent la taille des coefficients estimés. L'utilisation de ces techniques nous permet donc d'inclure davantage de prédicteurs dans nos modèles linéaires, afin d'en augmenter les performances prédictives.

3. Conditionnellement à l'échantillon d'apprentissage  $E_n$ , les tirages bootstrap  $\theta_1, \dots, \theta_B$  sont indépendants et identiquement distribués. Par conséquent, selon la loi des grands nombres :

$$\lim_{B \rightarrow +\infty} \hat{f}_B(\mathbf{x}) = \lim_{B \rightarrow +\infty} \frac{1}{B} \sum_{k=1}^B \hat{f}(\mathbf{x}, \theta_k) = E_\theta[\hat{f}(\mathbf{x}, \theta) | E_n]$$

où  $E_\theta[\hat{f}(\mathbf{x}, \theta) | E_n]$  est calculée à partir de la loi de  $\theta$ . Le résultat implique qu'un nombre élevé de  $B$  n'engendre pas le problème de surapprentissage de l'estimateur, et par conséquent, n'affecte pas sa performance prédictive (Friedman et al., 2001).

Quatre modélisations linéaires ont ainsi été estimées à partir des méthodes Ridge, LASSO et Elastic Net, celles-ci étant bâties de manière hiérarchique, c'est-à-dire que les prédicteurs des modélisations précédentes demeurent inchangés à mesure qu'on y en ajoute de nouveaux. Les premières modélisations (M1) correspondent aux régressions linéaires (5.1) et (5.2), soit les modèles linéaires également estimés par MCGF. En s'inspirant de la démarche de Hong et al. (2010), les modélisations suivantes incluent des termes d'interactions doubles entre les prédicteurs, ainsi que des polynômes, dans le but d'incorporer certains effets non linéaires observés dans les données. Notez que l'ajout de termes d'interactions triples entre les prédicteurs a également été considéré, mais augmentait considérablement le temps de calcul, sans améliorer le pouvoir prédictif de nos modèles. Le tableau 5.3 présente les prédicteurs compris dans chacune des quatre modélisations estimés par régressions pénalisées dans le cadre de ce mémoire.

	Prédicteurs initiaux :	Termes d'interactions	Polynôme de 3 <sup>e</sup> degré	Termes d'interactions doubles entre les termes
	<i>Year, Season, ToD,</i> <i>Weekday, HDD, CDD,</i> <i>q<sub>imp</sub>, Crisis et dem<sub>ON</sub></i>	doubles entre les prédicteurs initiaux	des HDD et CDD des régions de l'ensemble <i>G</i>	de saisonnalité ( <i>Year, Season, Weekday,</i> <i>ToD</i> ) et les termes au carré et au cube des HDD et CDD des régions de l'ensemble <i>G</i>
M1	✓			
M2	✓	✓		
M3	✓	✓	✓	
M4	✓	✓	✓	✓

Tableau 5.3 – Les quatre modélisations estimées à partir des méthodes Ridge, LASSO et Elastic Net.

Les méthodes Ridge, LASSO et Elastic Net ont été appliquées aux différentes modélisations à partir de la librairie *glmnet* du logiciel R, où les coefficients sont calculés par la technique de descente de gradient (Friedman et al., 2010). Le tableau 5.4 présente les résultats obtenus vis-à-vis la paramétrisation optimale dans les différents modèles étudiés pour l'analyse prédictive de la demande (haut) et du prix (bas) de l'électricité. Le nombre de prédicteurs sélectionnés ( $H$ ) par les méthodes Ridge, LASSO et Elastic Net y sont également indiqués.

a) Demande

Paramétrisation optimale ( $\lambda$ )							
Avant				Après			
	Ridge	LASSO	Elastic Net <sup>1</sup>	Ridge	LASSO	Elastic Net	
M1	158,20	3,33	0,90 & 3,49	121,36	2,20	0,16 & 8,67	
M2	158,20	0,16	0,98 & 0,16	121,36	0,12	0,96 & 0,13	
M3	158,20	0,16	0,96 & 0,17	121,36	0,12	0,98 & 0,12	
M4	158,20	0,16	0,86 & 0,18	121,36	0,12	0,78 & 0,16	

Nombre de paramètres sélectionnés ( $H$ )								
Avant				Après				
	Nombre de prédicteurs inclus dans le modèle ( $P$ )	Ridge	LASSO	Elastic Net	Nombre de prédicteurs inclus dans le modèle ( $P$ )	Ridge	LASSO	Elastic Net
M1	23	23 (100%)	18 (78%)	19 (83%)	24	24 (100%)	19 (79%)	23 (96%)
M2	264	264 (100%)	234 (89%)	234 (89%)	288	288 (100%)	233 (81%)	243 (84%)
M3	288	288 (100%)	248 (86%)	255 (89%)	312	312 (100%)	272 (87%)	268 (86%)
M4	528	528 (100%)	425 (80%)	444 (84%)	552	552 (100%)	447 (81%)	447 (81%)

b) Prix

Paramétrisation optimale ( $\lambda$ )							
Avant				Après			
	Ridge	LASSO	Elastic Net	Ridge	LASSO	Elastic Net	
M1	0,58	0,01	0,62 & 0,01	0,76	0,01	0,22 & 0,10	
M2	0,58	0,01	0,1 & 0,01	0,76	0,01	0,12 & 0,01	
M3	0,58	0,01	0,1 & 0,01	0,76	0,01	0,1 & 0,01	
M4	0,58	0,01	0,1 & 0,01	0,76	0,01	0,1 & 0,01	

Nombre de paramètres sélectionnés ( $H$ )								
Avant				Après				
	Nombre de prédicteurs inclus dans le modèle ( $P$ )	Ridge	LASSO	Elastic Net	Nombre de prédicteurs inclus dans le modèle ( $P$ )	Ridge	LASSO	Elastic Net
M1	24	24 (100%)	21 (88%)	21 (88%)	25	25 (100%)	22 (88%)	24 (96%)
M2	288	288 (100%)	164 (57%)	254 (88%)	313	313 (100%)	199 (64%)	274 (88%)
M3	312	312 (100%)	197 (63%)	288 (92%)	337	337 (100%)	208 (62%)	309 (92%)
M4	552	552 (100%)	253 (46%)	474 (86%)	577	577 (100%)	284 (49%)	497 (86%)

Tableau 5.4 – Résultats des méthodes Ridge, LASSO et Elastic Net à travers les quatre modélisations étudiées. Entre parenthèses est présentée la proportion de prédicteurs sélectionnés dans chacune des modélisations.

<sup>1</sup> Du problème 4.1, décrit à la section précédente, la méthode Elastic Net consiste à déterminer les paramètres  $\alpha$  et  $\lambda$  tels que :

$$\min_{\beta} \left\{ \sum_{t=1}^n \left( y_t - \sum_{j=1}^P x_{tj} \beta_j \right)^2 \right\} + (1 - \alpha) \lambda \sum_{j=1}^P |\beta_j| + \alpha \lambda \sum_{j=1}^P \beta_j^2.$$

Le premier élément de la colonne du tableau correspond donc au  $\alpha$  sélectionné et le second, au  $\lambda$ .

Des résultats, il est d'abord possible d'observer que les paramètres de régularisation optimaux demeurent étonnamment assez stables à mesure qu'on ajoute des prédicteurs supplémentaires aux modèles de prévision de la demande et du prix de l'électricité. Les résultats obtenus montrent également que les méthodes LASSO et Elastic Net sélectionnent majoritairement plus de 60% des prédicteurs à travers les différentes modélisations étudiées. La proportion de prédicteurs sélectionnés demeure toutefois généralement inférieure dans le modèle M4, c'est-à-dire la modélisation qui inclue, au départ, le plus grand nombre de prédicteurs.

Les pénalisations sont donc peu contraignantes à travers les modèles construits, et ceux-ci comprennent, par conséquent, un nombre élevé de prédicteurs, notamment en comparaison aux modèles conventionnels de régressions linéaires utilisés dans la littérature. Les résultats montrent donc qu'un nombre élevé de variables semblent nécessaires aux calculs, à partir d'une modélisation linéaire, des prévisions de la demande et du prix de l'électricité dans le marché ontarien sur les horizons temporels d'une et trois heures, respectivement.

### **Régression sur composantes principales**

Le problème de colinéarité engendré par le nombre élevé de prédicteurs dans le modèle de régression linéaire peut également être contrôlé par l'utilisation de la PCA qui transforme, préalablement à l'estimation par MCO, les prédicteurs à travers la construction des composantes principales. Les modélisations M1, M2, M3 et M4 ont donc également été estimées par PCA, à partir de la librairie *ppls* du logiciel R.

Le tableau 5.5 présente les résultats de la validation croisée à 10 groupes vis-à-vis le nombre de composantes principales ( $M$ ) incluses dans les modèles prédictifs de la demande (haut) et du prix (bas) de l'électricité. Tel que mentionné précédemment, la PCA n'engendre pas une sélection de variables. Son intérêt toutefois survient lorsque  $M < P$  : les composantes principales qui renferment peu d'information nécessaire à la prévision de la variable réponse sont alors supprimées, contrôlant le problème de colinéarité et de surapprentissage de l'estimateur, et diminuant du même coup sa variance.

a) Demande				
	Avant		Après	
	Nombre de prédicteurs inclus dans le modèle ( $P$ )	Nombre de composantes principales incluses ( $M$ )	Nombre de prédicteurs inclus dans le modèle ( $P$ )	Nombre de composantes principales incluses ( $M$ )
M1	23	18 (78%)	24	16 (67%)
M2	264	205 (78%)	288	247 (86%)
M3	288	238 (83%)	312	254 (81%)
M4	528	411 (78%)	552	452 (82%)

b) Prix				
	Avant		Après	
	Nombre de prédicteurs inclus dans le modèle ( $P$ )	Nombre de composantes principales incluses ( $M$ )	Nombre de prédicteurs inclus dans le modèle ( $P$ )	Nombre de composantes principales incluses ( $M$ )
M1	24	21 (88%)	25	23 (92%)
M2	288	224 (78%)	313	253 (81%)
M3	312	263 (84%)	337	277 (82%)
M4	552	435 (79%)	577	457 (79%)

Tableau 5.5 – Nombre de composantes principales ( $M$ ) sélectionnées par la validation croisée à 10 groupes, à travers les 4 modélisations étudiées incluant  $P$  prédicteurs initialement.

Le nombre optimal de composantes principales incluses dans les différentes modélisations étudiées de la demande et du prix de l'électricité est généralement élevé vis-à-vis du nombre de prédicteurs. Des résultats, il est ainsi possible de tirer des conclusions similaires à celles obtenues par régressions pénalisées, c'est-à-dire que les prévisions, à partir d'un modèle linéaire, de la demande et du prix de l'électricité dans le marché de l'Ontario nécessitent une quantité élevée d'information présente dans les données.

### Arbres de régression

Les arbres de régression ont la capacité d'identifier les relations non linéaires contenues dans les données, contrairement aux méthodes de régressions pénalisées et à la PCA qui demandent l'inclusion de termes d'interactions et de polynômes afin de les considérer. Par conséquent, les arbres de régression sont construits uniquement à partir des prédicteurs initiaux, ce qui diminue le temps de calcul nécessaire à leur estimation, comparativement

aux autres méthodes nommées.

Les arbres de régression sont estimés à partir de la librairie *rpart* du logiciel R. Une procédure d'émondage est appliquée dans le but de contrôler le surapprentissage de l'estimateur, demandant ainsi la détermination du paramètre d'émondage optimal ( $\kappa$ ) qui représente l'importance de la pénalisation appliquée à la complexité de l'arbre.

À partir des résultats de la validation croisée, les modèles prédictifs de la demande et du prix de l'électricité ont été construits avec  $\kappa = 0,01$ . Les arbres de régression finaux pour la prévision de la demande d'électricité sont constitués d'uniquement 9 noeuds. Les modèles estimés sont simples, n'incluant que 5 et 6 prédicteurs dans la période "avant" et "après", respectivement. Les arbres de régression finaux pour la prévision du prix de l'électricité reflètent également le principal avantage des arbres de régression : leur simplicité. En effet, les modèles estimés comptent 6 et 8 noeuds pour la période "avant" et "après", respectivement, et font intervenir uniquement 4 et 6 prédicteurs dans la prévision du prix sur l'horizon temporel de 3 heures. Les représentations graphiques des arbres de régressions estimés sont présentées en annexe.

Les résultats obtenus représentent ainsi une importante différence vis-à-vis ceux obtenus dans les modèles linéaires, où un nombre élevé de prédicteurs semblait nécessaire aux prévisions de court terme de la demande et du prix de l'électricité en Ontario. En particulier, les différents arbres de régression estimés n'incluent, en moyenne, qu'environ 23% des prédicteurs initiaux dans la prévision de la demande d'électricité, et, en moyenne, environ 21% des prédicteurs dans la prévision du prix. En comparaison, les modélisations M1 estimées par LASSO et Elastic Net incluaient, en moyenne, respectivement environ 84% et 90% des prédicteurs initiaux.

### **Bagging, forêts aléatoires et boosting**

Les méthodes d'ensemble aléatoires et additives (bagging, forêts aléatoires et boosting) sont employées pour contrôler la variance élevée de l'arbre de régression individuel. Elles sont implémentées à partir des librairies *randomForest* et *gbm* du logiciel R.

L'application des méthodes d'ensemble nécessite la détermination de différents paramètres. Tel que déjà mentionné, le nombre de répliquions bootstrap ( $B$ ) utilisées dans le calcul des estimateurs bagging et forêts aléatoires ne représente pas un paramètre critique. Il a donc arbitrairement été fixé à  $B = 500$  afin d'assurer la précision du modèle construit, tout en restreignant le temps de calcul nécessaire à l'estimation. Les forêts aléatoires requièrent également de fixer la taille des sous-ensembles de prédicteurs à partir desquels sont sélectionnés les noeuds optimaux à chaque étape de la construction des arbres de régression ( $V$ ), où une valeur faible de  $V$  signifie une décorrélation accrue des arbres de régression individuellement construits, engendrant l'augmentation du biais de l'estimateur final et la diminution de sa variance. La méthode boosting nécessite, quant à elle, trois paramétrisations, soit le nombre d'arbres de régression considérés dans l'estimateur final ( $Z$ ), le taux d'apprentissage ( $\zeta$ ) qui détermine le poids de chaque nouvel arbre ajouté à l'estimateur, ainsi que la complexité de l'arbre ( $cc$ ) qui limite le nombre maximal de noeuds des arbres construits. Les valeurs de ces paramètres sont déterminées dans le but de contrôler le problème de surapprentissage susceptible de se produire dans l'estimateur boosting. Les valeurs des paramétrisations optimales des modèles bagging, forêts aléatoires et boosting, déterminées par validation croisée à 10 groupes, sont présentées dans le tableau 5.6.

			Nombre d'arbres	Taille des sous-ensembles	Complexité de	Taux d'apprentissage
			construits	de predicteurs ( $V$ )	l'arbre ( $cc$ )	( $\zeta$ )
Demande	Avant	Bagging	500	23	-	-
		Forêts aléatoires	500	9	-	-
		Boosting	2050	-	10	0.045
	Après	Bagging	500	24	-	-
		Forêts aléatoires	500	14	-	-
		Boosting	2400	-	10	0.045
Prix	Avant	Bagging	500	24	-	-
		Forêts aléatoires	500	5	-	-
		Boosting	1950	-	10	0.025
	Après	Bagging	500	25	-	-
		Forêts aléatoires	500	6	-	-
		Boosting	2500	-	10	0.025

Tableau 5.6 – Résumé des résultats des paramétrisations des modèles bagging, forêts aléatoires et boosting pour les prévisions de la demande et des prix de l'électricité.

## 5.3 Comparaison des performances prédictives

### 5.3.1 Qualité de prévision

La comparaison des performances prédictives des modèles estimés nécessite des mesures de qualité de prévision<sup>4</sup>. En nous basant sur la littérature existante, la précision des prévisions des modèles étudiés sera calculée à partir de la racine de l'erreur quadratique moyenne (RMSE). Soit  $\hat{y}_t$  avec  $t = \{1, \dots, N\}$  les valeurs prédites de la série temporelle de la variable réponse dans l'échantillon test de taille  $N$ , et  $y_t$ , ses valeurs réelles. La RMSE correspond alors à :

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}.$$

L'erreur absolue moyenne en pourcentage (MAPE) représente également une mesure de qualité de prévision intéressante, puisqu'elle indique la différence proportionnelle entre les valeurs prédites et les valeurs réelles. Elle est représentée par :

$$MAPE = \frac{100}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right|.$$

Cet indicateur est toutefois inapproprié à la comparaison des modèles prédictifs des prix, puisque sa valeur tend vers l'infini lorsque le prix observé se rapproche de zéro<sup>5</sup>. Il sera alors préféré l'utilisation de l'erreur absolue moyenne (MAE), soit :

$$MAE = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t|.$$

---

4. Dans ce mémoire, nous référons aux indicateurs d'ajustement à partir de leur traduction anglaise. Les termes RMSE, MAPE et MAE correspondent donc respectivement à *root mean square error*, *mean absolute percentage error* et *mean absolute error*.

5. Nos échantillons tests de la période "avant" et "après" comprennent respectivement 4 et 491 observations où les prix observés sont compris entre -0,5\$CAD et 0,5\$CAD. Les MAPE calculées dans le cadre de notre recherche sont, par conséquent, largement supérieures en comparaison aux résultats obtenus dans différentes études de la littérature. En particulier, l'ensemble des analyses prédictives des prix de l'électricité en Ontario ont été effectuées sur des périodes où les valeurs observées des prix sont éloignées de 0. Puisque les valeurs élevées des MAPE obtenues ne signifient pas nécessairement que les modèles présentent de faibles pouvoirs prédictifs, mais représentent plutôt une conséquence de la composition de nos échantillons, nous jugeons préférable d'employer l'erreur absolue moyenne (MAE) afin de quantifier la performance de nos modèles prédictifs.

## 5.3.2 Résultats

### Prévision de la demande d'électricité

Le tableau 5.7 présente les RMSE et les MAPE obtenues à partir des différents modèles étudiés pour la prévision de la demande d'électricité de la prochaine heure en Ontario.

Des résultats, il est d'abord possible d'observer que, de manière générale, les méthodes étudiées tirées de l'apprentissage statistique montrent des performances prédictives supérieures aux méthodes conventionnelles de prévision. En effet, le modèle de régression linéaire avec erreur ARIMA (2) présente, à travers les deux périodes étudiées, des valeurs de RMSE et de MAPE élevées, traduisant ainsi son faible pouvoir prédictif dans le calcul de la prévision de la demande d'électricité de la prochaine heure en Ontario. Bien que les RMSE et les MAPE observées à partir du modèle de régression linéaire multiple estimée par MCGF (1) soient inférieures à celles du modèle de régression linéaire avec erreur ARIMA (2), elles demeurent tout de même supérieures à celles associées aux modèles de régressions pénalisées, aux estimateurs bagging et aux forêts aléatoires.

Les résultats permettent également de constater l'efficacité des méthodes de régressions pénalisées dans le calcul de la prévision de la demande d'électricité de la prochaine heure dans le marché ontarien. En effet, dans la période "avant", les modèles optimaux en termes de minimisation de la RMSE et de la MAPE correspondent au modèle linéaire M3 estimé par Elastic Net (13) avec une RMSE de 1237,3 MWh, et au modèle linéaire M2 estimé par Ridge (4) avec une MAPE de 5,3%. Ce dernier représente également le modèle optimal dans la période "après", avec une RMSE de 1149,4 MWh et une MAPE de 4,9%. Les résultats obtenus montrent ainsi que la prise en compte des termes d'interaction double entre les prédicteurs et l'utilisation d'une pénalisation sur la taille des coefficients estimés permet d'améliorer nettement la performance du modèle linéaire dans la prévision de la demande d'électricité en Ontario sur un horizon temporel de 1 heure. En particulier, les RMSE et les MAPE des modèles optimaux représentent, en moyenne, à travers les deux périodes étudiées, une diminution de respectivement 157 MWh et de 0,67% vis-à-vis des RMSE et

des MAPE obtenues à partir des régressions linéaires multiples estimées par MCGF (1). On note également que bien que les modèles optimaux soient simples et ne demandent que peu de paramétrisations, ils engendrent tout de même des MAPE comparables à celle obtenue à partir d'un modèle prédictif complexe : le perceptron multicouche (réseau de neurones). En particulier, Liu et al. (2006) obtiennent une MAPE d'environ 4% à partir de leur perceptron multicouche construit pour la prévision de la demande d'électricité de la prochaine heure en Ontario sur les données de l'année 2006.

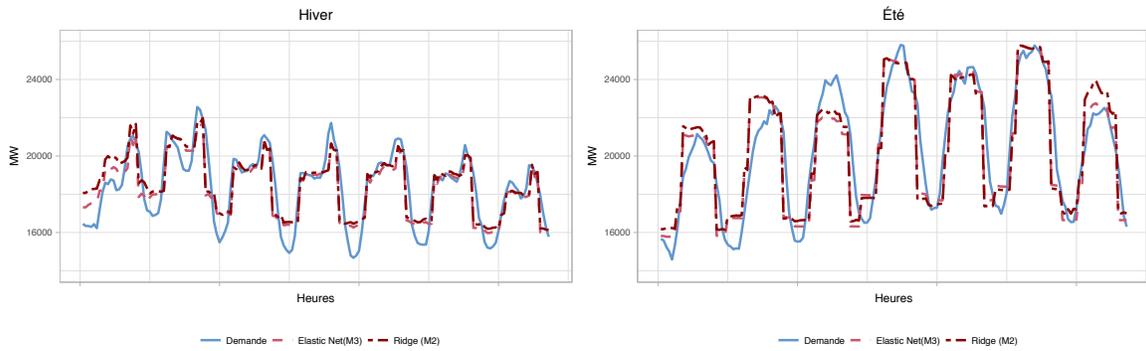
Les modèles linéaires optimaux offrent, entre autres, de meilleures performances prédictives que les méthodes non linéaires de prévision étudiées, telles que les forêts aléatoires (21) et les estimateurs bagging. Ceux-ci montrent tout de même des pouvoirs prédictifs élevés en réalisant respectivement des RMSE de 1242,3 MWh et 1274,53 MWh, dans la période "avant", et de 1177,6 MWh et 1195,56 MWh, dans la période "après", ainsi que des MAPE de 5,43% et 5,57%, dans la période "avant", et de 5,08% et 5,16%, dans la période "après".

Enfin, les résultats révèlent que les modèles de régressions sur composantes principales (PCA) et les estimateurs boosting ont des pouvoirs prédictifs faibles vis-à-vis la prévision de la demande d'électricité de la prochaine heure en Ontario. En effet, les valeurs des RMSE et des MAPE associées à ces estimateurs les classent généralement parmi les derniers rangs du classement des performances prédictives des modèles étudiés. Le faible pouvoir prédictif de ces modèles est également observable dans les graphiques 5.1 et 5.2 qui présentent les valeurs prédites de la demande d'électricité à partir de différents modèles prédictifs étudiés, notamment les modèles optimaux sélectionnés, les arbres de régression, les forêts aléatoires ainsi que les estimateurs boosting et PCA (M1). Des graphiques, il est possible d'observer que, de manière générale, l'ensemble des modèles étudiés parviennent à prédire la tendance générale de la demande d'électricité dans le marché de l'Ontario. Toutefois, contrairement aux modèles linéaires optimaux, les PCA et les estimateurs boosting tendent à prévoir une demande d'électricité inférieure à celle observée, notamment en périodes de demande élevée. Lors de ces périodes, les résultats présentés dans les graphiques suggèrent des RMSE et des MAPE supérieures à celles présentées dans le tableau 5.7.

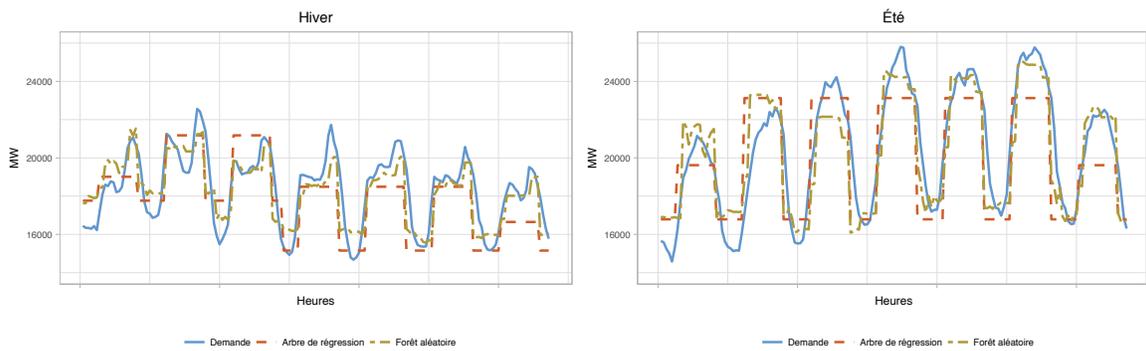
		Avant		Après	
		RMSE	MAPE	RMSE	MAPE
		(MWh)	(%)	(MWh)	(%)
(1)	Régression linéaire multiple	1427,44	6,05	1274,11	5,52
(2)	Régression linéaire avec erreur ARIMA	3483,76	15,28	1885,89	8,52
Ridge					
(3)	M1	1394,52	6,00	1383,93	5,98
(4)	M2	1250,90	5,29	1149,40	4,94
(5)	M3	1268,80	5,37	1154,41	4,95
(6)	M4	20534,21	19,61	1156,51	4,95
LASSO					
(7)	M1	1386,67	5,98	1381,43	5,98
(8)	M2	1274,64	5,52	1152,01	4,94
(9)	M3	1238,91	5,38	1170,23	5,01
(10)	M4	1261,06	5,45	1181,70	5,04
Elastic Net					
(11)	M1	1386,65	5,98	1381,84	5,88
(12)	M2	1273,35	5,51	1153,13	4,95
(13)	M3	1237,29	5,38	1170,33	5,01
(14)	M4	1354,49	5,73	1181,70	5,04
Régression sur composantes principales (PCA)					
(15)	M1	1381,26	5,98	1273,23	5,50
(16)	M2	15627,44	18,20	-	-
(17)	M3	12907,08	16,04	1415,03	5,96
(18)	M4	25470,68	28,31	-	-
(19)	Arbres de régression	1625,62	7,08	1491,45	6,43
(20)	Bagging	1274,53	5,57	1195,56	5,16
(21)	Forêt aléatoire	1242,32	5,43	1177,64	5,08
(22)	Boosting	2336,72	10,81	1943,76	8,53

Tableau 5.7 – RMSE et MAPE des modèles étudiés pour les prévisions de la demande d'électricité de la prochaine heure dans le marché de l'Ontario.

### Modèles optimaux : Ridge (M2) et Elastic Net (M3)



### Arbre de régression et Forêt aléatoire



### PCA (M1), Boosting et Régression linéaire avec erreur ARIMA

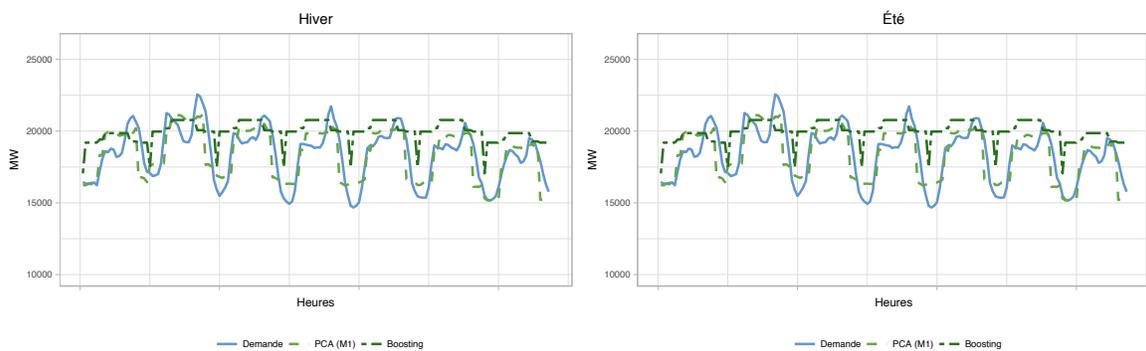
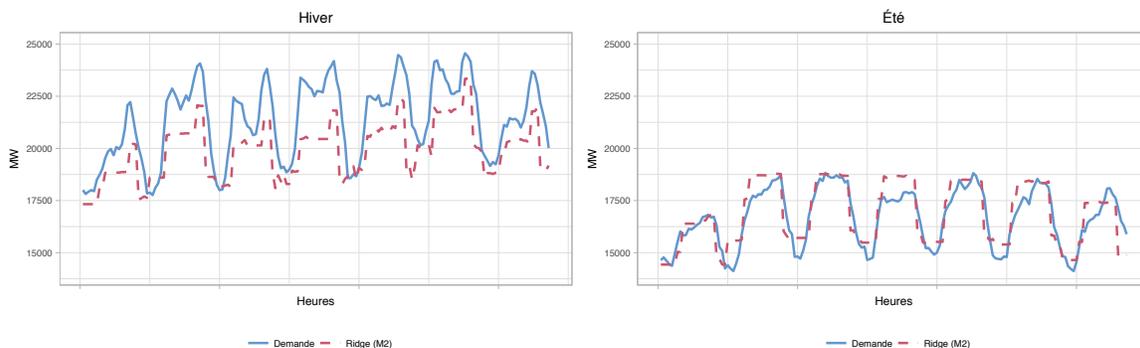
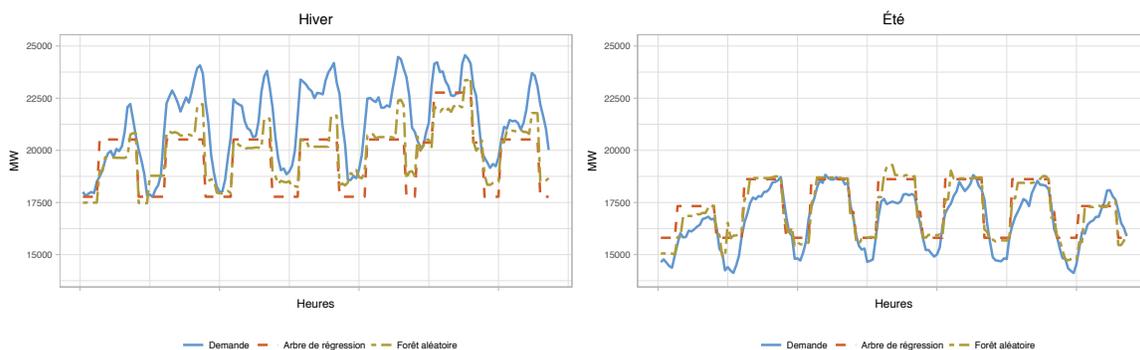


FIGURE 5.1 – Période “avant” - Comparaison des valeurs réelles et des valeurs prédites par les estimateurs Ridge (M2), Elastic Net (M3), arbres de régression, forêts aléatoires, PCA (M1) et boosting de la demande d’électricité en Ontario. Deux semaines de la période “avant” sont présentées : la semaine d’hiver correspond à la période du 12 au 18 février 2012 et la semaine d’été correspond à la période du 1<sup>er</sup> au 7 juillet 2012. Les prévisions ont été calculées une heure à l’avance.

### Modèle optimal : Ridge (M2)



### Arbre de régression et Forêt aléatoire



### PCA (M1), Boosting et Régression linéaire avec erreur ARIMA

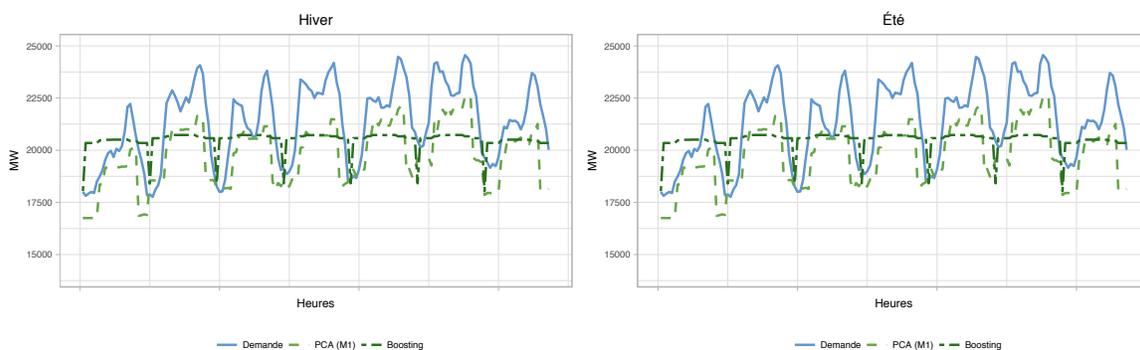


FIGURE 5.2 – Période “après” - Comparaison des valeurs réelles et des valeurs prédites par les estimateurs Ridge (M2), arbres de régression, forêts aléatoires, PCA (M1) et boosting de la demande d’électricité en Ontario. Deux semaines de la période “après” sont présentées : la semaine d’hiver correspond à la période du 8 au 14 février 2015 et la semaine d’été correspond à la période du 28 juin au 4 juillet 2015. Les prévisions ont été calculées une heure à l’avance.

Les modèles optimaux sélectionnés font intervenir entre 255 et 288 prédicteurs dans la prévision de la demande d'électricité de la prochaine heure en Ontario. La méthode Elastic Net est avantagée du fait qu'elle engendre la sélection de variables, contrairement à la méthode Ridge. En particulier, elle sélectionne uniquement 255 prédicteurs parmi les 288 inclus initialement dans la modélisation M3. Les prédicteurs non sélectionnés sont listés dans le tableau 5.8.

Prédicteurs non sélectionnés <sup>1</sup>			
<i>fall</i>	<i>hdd_MB</i>	<i>cdd_MN</i>	<i>cdd_NY</i>
<i>hdd_MB</i> <sup>2</sup>	<i>hdd_MI</i> <sup>3</sup>	<i>year*tot_nx</i>	<i>year*hdd_ON</i>
<i>year*cdd_ON</i>	<i>year*hdd_MN</i>	<i>year*hdd_MI</i>	<i>year*cdd_MI</i>
<i>year*tot_nx</i>	<i>year*hdd_ON</i>	<i>weekday*fall</i>	<i>fall*hdd_QC</i>
<i>fall*cdd_QC</i>	<i>fall*hdd_NY</i>	<i>spring*offpeak_mo</i>	<i>summer*offpeak_ni</i>
<i>summer*cdd_MN</i>	<i>summer*cdd_MN</i>	<i>winter*cdd_ON</i>	<i>winter*cdd_QC</i>
<i>winter*cdd_MB</i>	<i>winter*cdd_MN</i>	<i>offpeak_ni*cdd_NY</i>	<i>peak_mo*hdd_ON</i>
<i>peak_ni*cdd_QC</i>	<i>peak_ni*cdd_QC</i>	<i>tot_nx*fall</i>	<i>tot_nx*peak_mo</i>
<i>cdd_MB*cdd_MI</i>			

Tableau 5.8 – Prédicteurs non sélectionnés par la méthode Elastic Net dans la modélisation optimale M3.

<sup>1</sup> offpeak\_ni= creux de la nuit, peak\_mo= pic matinal, offpeak\_mo= creux de la journée, peak\_ni= pic de soirée.

<sup>2</sup> \* représente l'interaction entre deux prédicteurs.

La sélection de variables ne permet toutefois pas d'inférer l'impact, ou non, d'un prédicteur sur la variable réponse. En effet, un prédicteur non sélectionné n'indique pas nécessairement que son effet sur la variable réponse est nul (et vice-versa), mais plutôt qu'il est superflu à la prévision de celle-ci dans les conditions spécifiques sous lesquelles est performée la prévision. Cet attribut est démontré dans le graphique 5.3 qui montre les résultats obtenus vis-à-vis la sélection de variables par la méthode Elastic Net dans la modélisation M3, sur dix sous-échantillons construits à partir d'une division aléatoire de notre échantillon d'apprentissage de la période "avant"<sup>6</sup>. Les valeurs des paramètres ( $\alpha$  et  $\lambda$ ) ont été fixées, à travers les dix estimations, à leur valeur optimale déterminée par une validation croisée à 10 groupes dans le premier sous-échantillon. Cette démarche est inspirée d'une analyse réalisée par Mullainathan et Spiess (2017).

6. Le graphique est inspiré de la figure retrouvée dans Mullainathan et Spiess (2017).

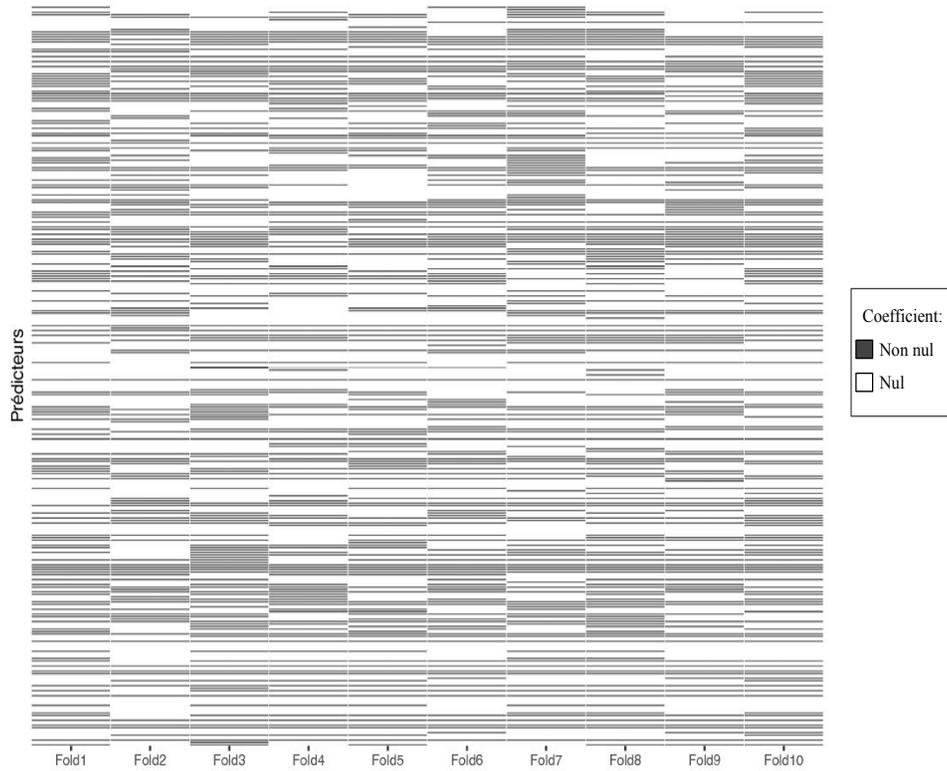


FIGURE 5.3 – Sélection de variables par la méthode Elastic Net dans la modélisation M3, à travers 10 sous échantillons de l'échantillon d'apprentissage initial de la période “avant”. Chaque colonne correspond à un sous échantillon d'environ 1435 observations. Chaque ligne représente un prédicteur du modèle. Une cellule noircie signifie que le prédicteur est sélectionné par la méthode Elastic Net.

Chaque colonne du graphique représente un sous-échantillon, et chaque ligne un prédicteur dont la cellule est noircie lorsque celui-ci est sélectionné par la méthode Elastic Net. Il est ainsi possible d'observer qu'un prédicteur sélectionné dans un sous-échantillon n'est pas forcément sélectionné dans l'ensemble de ceux-ci. Similairement aux résultats obtenus par Mullainathan et Spiess (2017), nous observons que l'instabilité du processus de sélection de variables ne se reflète pas dans les performances prédictives des modèles estimés, les  $R^2$  calculés étant comparables. Des prévisions similaires de la variable réponse peuvent ainsi être obtenues à partir de prédicteurs différents, dépendamment de l'échantillon sous lequel le modèle est construit. Par conséquent, un prédicteur non sélectionné ne signifie pas que son effet sur la variable réponse est nul, et les coefficients estimés ne peuvent être interprétés comme les indications des corrélations entre les prédicteurs et la variable réponse.

Les coefficients estimés de nos modèles optimaux montrent toutefois l'importance de chaque prédicteur dans la prévision de la demande d'électricité de la prochaine heure en Ontario, selon la période étudiée. En effet, puisque les coefficients ont été calculés à partir de données standardisées, ils sont sur la même échelle et peuvent être comparés. Classer les coefficients estimés selon leur taille (absolue) permet ainsi d'observer quels prédicteurs sont les plus influents dans la prévision de la demande d'électricité ontarienne sur l'horizon temporel de 1 heure. Le graphique 2 présente le classement des 20 prédicteurs dont la taille absolue des coefficients est la plus importante à travers les trois modélisations optimales. Ici, les résultats sont présentés en pourcentage et ont été normalisés vis-à-vis du prédicteur le plus important afin de faciliter la visualisation de la grandeur de l'impact de chaque prédicteur sur la variable réponse. Les colonnes bleues ne reflètent donc pas la taille des coefficients obtenus, mais plutôt l'ampleur de ceux-ci vis-à-vis du coefficient le plus élevé. Les résultats non normalisés sont disponibles dans l'annexe C.

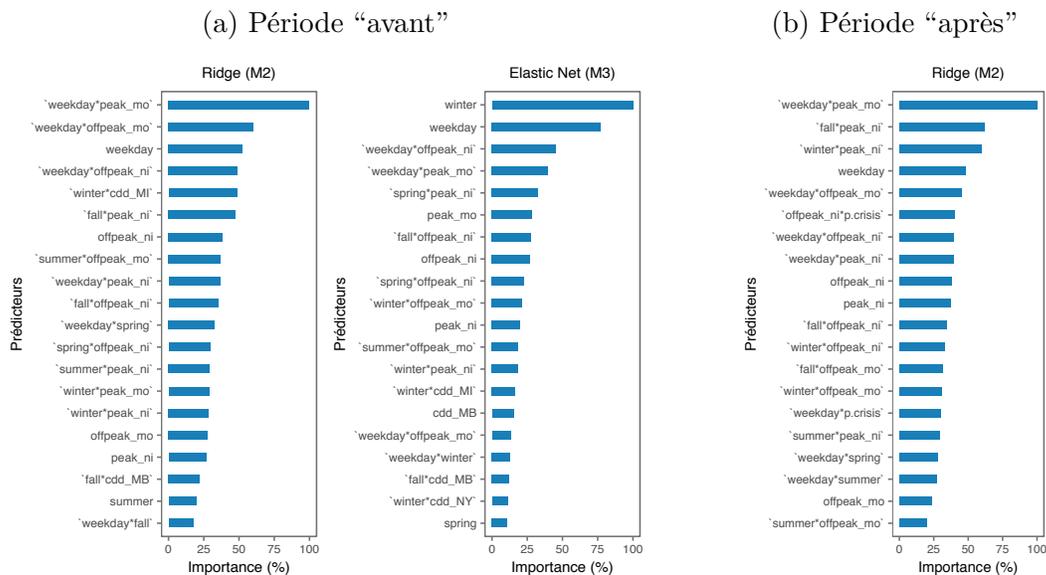


FIGURE 5.4 – Classement (en pourcentage) des 20 prédicteurs dont la taille absolue des coefficients est la plus importante dans les modèles optimaux de prévision de la demande d'électricité de la prochaine heure des périodes "avant" et "après". Les résultats sont normalisés vis-à-vis du prédicteur le plus important.

Des résultats, il est d'abord possible d'observer que l'importance de chaque prédicteur varie à travers les différentes modélisations optimales. On remarque toutefois l'importance des prédicteurs de saisonnalité dans les trois modèles prédictifs, notamment vis-à-vis l'importance des prédicteurs climatiques. En particulier, les résultats montrent l'importance du fait d'être un jour de la semaine (*weekday*) et des 4 périodes quotidiennes dans le calcul de la prévision de la demande d'électricité de la prochaine heure en Ontario, à travers les deux périodes étudiées. Étonnamment, on ne retrouve aucune variable reflétant la température ontarienne parmi les prédicteurs présentés.

Les RMSE et les MAPE obtenues à partir des différents modèles prédictifs étudiés, nous ont permis de conclure à l'optimalité des modèles Ridge (M2) et Elastic Net (M3) dans la prévision de la demande d'électricité de la prochaine heure. Les prévisions de la demande d'électricité sur un horizon temporel d'une heure constituent un élément crucial au fonctionnement optimal du marché de l'électricité de l'Ontario et à la garantie d'un approvisionnement suffisant en électricité. Les différents agents du marché ontarien peuvent toutefois être également amenés à prévoir la demande d'électricité sur différents horizons de temps : la planification des opérations quotidiennes par l'IESO, ainsi que le développement de stratégies d'enchère optimales par les producteurs d'électricité peuvent nécessiter de prévoir la demande d'électricité plusieurs heures, voire plusieurs jours à l'avance. Il peut donc être également intéressant d'étudier la performance prédictive de nos modèles optimaux sur différents horizons temporels de prévision.

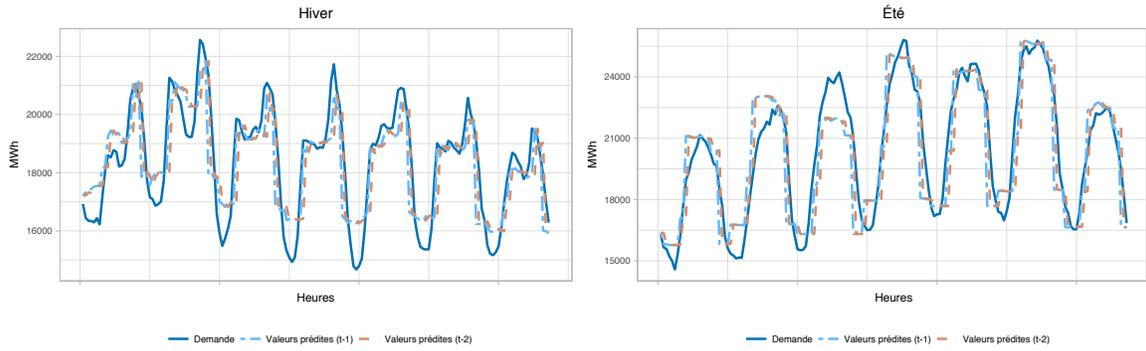
Le graphique 4 montre les valeurs prédites de la demande d'électricité en Ontario à partir du modèle Ridge (M2) calculées 2 heures, 3 heures, 14 heures, 19 heures, 25 heures et 721 heures à l'avance<sup>7</sup>. Ces horizons temporels de prévision ont été sélectionnés étant donné leur importance dans le processus de formation des décisions des agents du marché de l'électricité en Ontario<sup>8</sup>. Les valeurs prédites 1 heure à l'avance, précédemment obtenues, sont également présentées à titre de comparaison.

---

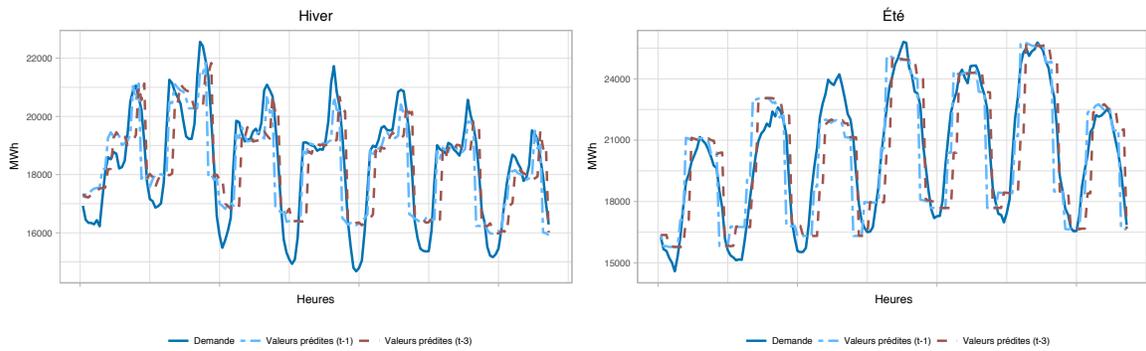
7. Une analyse similaire à été effectuée pour le second modèle optimal de la période "avant" : Elastic Net (M3), ainsi que le modèle optimal de la période "après" : Ridge (M2). Les graphiques sont disponibles en annexe.

8. Ces horizons temporels de prévision correspondent aux différents événements qui caractérisent le cycle des opérations du marché de l'électricité en Ontario (voir figure 3.1 à la section 3).

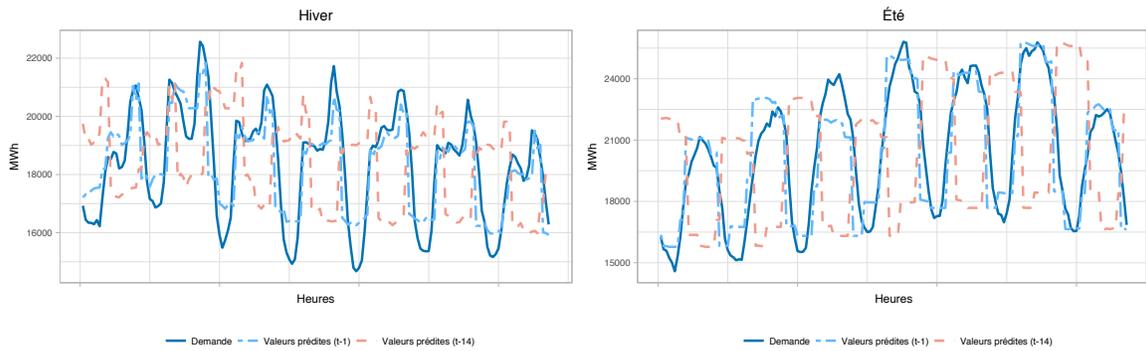
### Horizon temporel : 2 heures



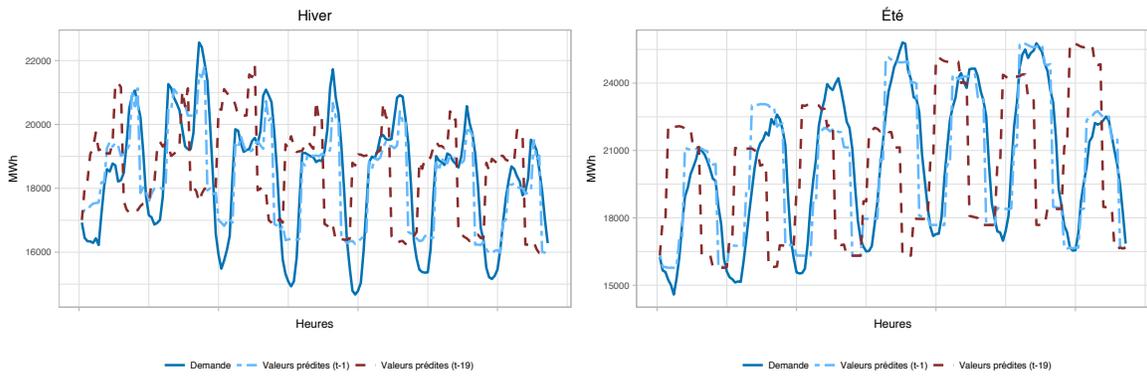
### Horizon temporel : 3 heures



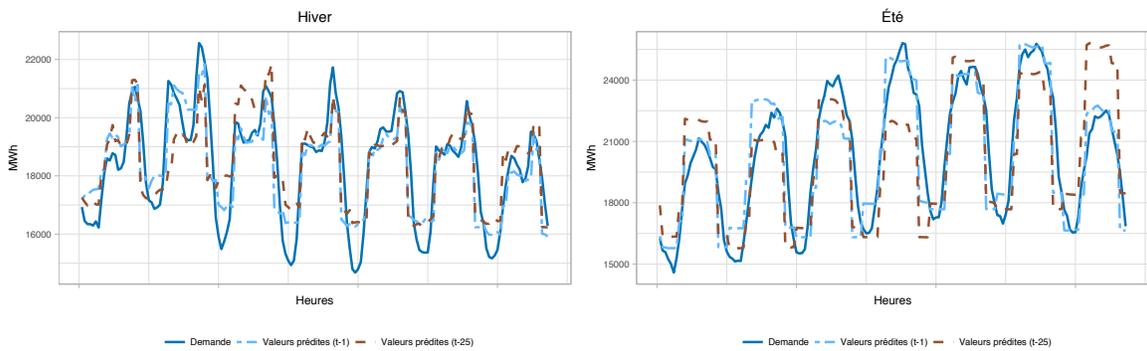
### Horizon temporel : 14 heures



### Horizon temporel : 19 heures



### Horizon temporel : 25 heures



### Horizon temporel : 721 heures

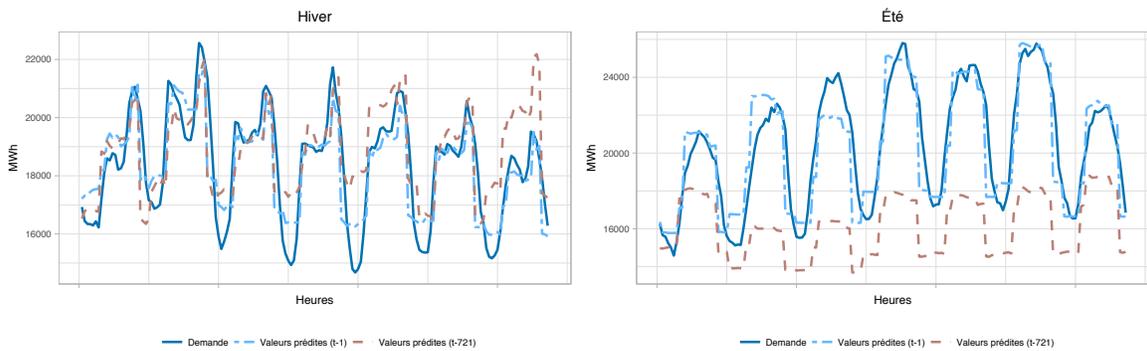


FIGURE 5.5 – Comparaison des valeurs réelles et des valeurs prédites par le modèle Ridge (M2) sur différents horizons de temps. La demande d’électricité horaire en Ontario a été prédite 2, 3, 14, 19, 25 et 721 heures à l’avance. Les valeurs prédites 1 heure à l’avance sont également présentées sur chaque graphique à titre de comparaison. Les résultats sont montrés sur deux semaines de l’échantillon test de la période “avant”, soit du 12 au 18 février 2012 (hiver) et du 1<sup>er</sup> au 7 juillet 2012 (été).

Des résultats, on remarque que le modèle Ridge (M2) offre de bonnes performances prédictives dans les cas où la prévision est effectuée 2, 3 et 25 heures à l'avance. En particulier, les valeurs prédites 2 et 3 heures à l'avance sont, de manière générale, comparables à celles obtenues 1 heure à l'avance. On constate, néanmoins que les prévisions de la demande d'électricité effectuées sur les horizons temporels de 14, 19 et 721 heures s'éloignent davantage des valeurs réelles observées. Les résultats permettent ainsi de conclure que le modèle Ridge (M2) offre de meilleures performances prédictives à très court terme ( $\leq 3$  heures), que sur un horizon temporel plus élevée ( $> 3$  heures). Des résultats similaires ont également été obtenus à partir du modèle Elastic Net (M3), d'où peuvent être tirées des conclusions semblables à celles établies dans le cadre du modèle Ridge (M2).

### **Prévision du prix de l'électricité**

Les résultats des RMSE et des MAE associées aux modèles prédictifs du prix de l'électricité sont présentés dans le tableau 5.9.

Des résultats obtenus, nous observons d'abord que les modèles de régressions linéaires multiples estimées par MCGF (1) constituent les modèles prédictifs optimaux en termes de minimisation de la RMSE et de la MAE, avec des RMSE de 9,85\$CAD et de 15,10\$CAD et des MAE de 7,71\$CAD et de 12,55\$CAD dans les périodes "avant" et "après" respectivement. Cette méthode conventionnelle de prévision offre, par conséquent, de meilleures performances prédictives que l'ensemble des méthodes tirées de l'apprentissage statistique étudiées pour le calcul de la prévision du prix de l'électricité dans le marché de l'Ontario sur un horizon temporel de 3 heures.

Parmi les méthodes tirées de l'apprentissage statistique, les méthodes linéaires montrent de meilleurs pouvoirs prédictifs, en comparaison aux méthodes non linéaires de prévision. En particulier, la modélisation M1 estimée par PCA (15) constitue le modèle offrant la meilleure performance prédictive dans la période "avant", avec une RMSE de 10,31\$CAD et une MAE de 7,41\$CAD. On constate également, dans la période "avant", qu'à travers les 4 modélisations linéaires, les RMSE et les MAE augmentent à mesure que le nombre de prédicteurs croît, démontrant, de ce fait, la supériorité des modèles prédictifs parcimonieux.

		Avant		Après	
		RMSE	MAE	RMSE	MAE
		(\$CAD)	(\$CAD)	(\$CAD)	(\$CAD)
(1)	Régression linéaire multiple	9,62	6,74	39,11	11,84
(2)	Régression linéaire avec erreur ARIMA	12,57	10,37	65,82	14,30
Ridge					
(3)	M1	10,21	7,60	16,04	12,81
(4)	M2	11,42	9,05	16,47	12,89
(5)	M3	11,73	9,40	16,46	12,89
(6)	M4	129,42	12,80	16,41	12,88
LASSO					
(7)	M1	10,06	7,34	16,42	13,10
(8)	M2	11,87	9,46	16,94	13,26
(9)	M3	12,16	9,83	17,00	13,30
(10)	M4	12,48	10,11	17,07	13,43
Elastic Net					
(11)	M1	10,05	7,33	16,42	13,10
(12)	M2	11,62	9,02	17,07	13,36
(13)	M3	12,27	9,82	17,18	13,46
(14)	M4	12,27	10,92	17,56	13,86
Régression sur composantes principales (PCA)					
(15)	M1	10,19	7,51	16,44	13,11
(16)	M2	11,20	8,49	24,61	18,34
(17)	M3	28,02	22,46	24,37	18,63
(18)	M4	293,55	58,10	70,26	23,03
(19)	Arbre de régression	13,10	11,11	16,44	13,00
(20)	Bagging	13,30	10,98	18,81	14,79
(21)	Forêts aléatoires	12,81	10,56	17,99	14,26
(22)	Boosting	24,47	22,03	18,62	13,24

Tableau 5.9 – RMSE et MAE des modèles étudiés pour les prévisions du prix de l'électricité sur un horizon temporel de 3 heures dans le marché de l'Ontario.

Dans la période “après”, on observe plutôt l’inverse : la modélisation M4 estimée par Ridge (6) présente la meilleure performance prédictive parmi les méthodes de l’apprentissage statistique étudiées, avec une RMSE de 15,86\$CAD et une MAE de 12,38\$CAD. Les méthodes non linéaires, quant à elles, présentent des RMSE et des MAE élevées à travers les deux périodes étudiées. Par conséquent, elles montrent de faibles pouvoirs prédictifs dans la prévision du prix de l’électricité dans le marché de l’Ontario sur un horizon temporel de 3 heures.

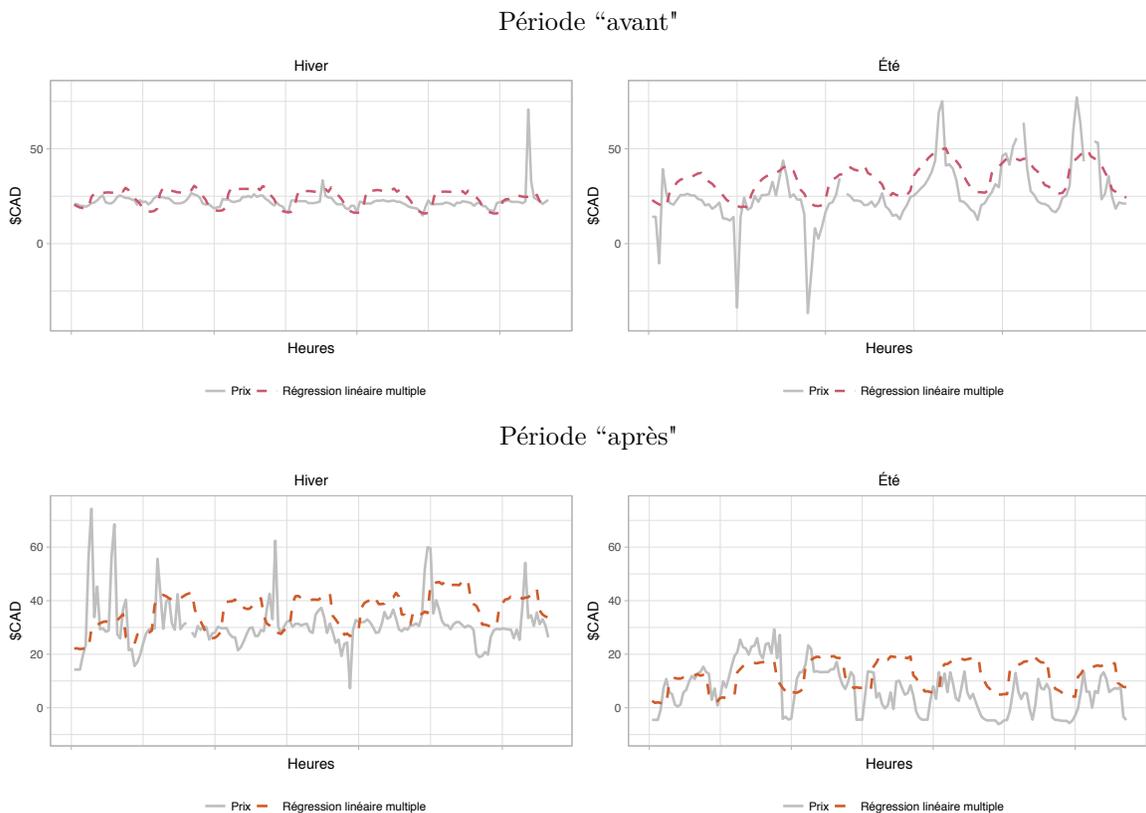


FIGURE 5.6 – Comparaison des valeurs réelles et des valeurs prédites du prix de l’électricité par les modèles de régressions linéaires multiples. Dans la période “avant”, les résultats sont rapportés sur les semaines du 12 au 18 février 2012 (hiver) et du 1<sup>er</sup> au 7 juillet 2012 (été). Dans la période “après”, les résultats sont présentés pour les semaines du 8 au 14 février 2015 (hiver) et du 28 juin au 4 juillet 2015 (été). Les prévisions ont été calculées 3 heures à l’avance.

Les valeurs prédites par les modèles optimaux des périodes “avant” et “après” sont présentées dans le graphique 5.6. Les résultats sont rapportés sur deux semaines de nos échantillons, soit une semaine d’hiver et une semaine d’été. Des graphiques, il est possible d’observer que

les modèles optimaux réussissent à prévoir de manière raisonnable la tendance générale des prix de l'électricité. Toutefois, aucun des modèles prédictifs ne parvient à capter les nombreux pics de prix (*price spike*) survenus dans les périodes présentées et, par conséquent, les valeurs prédites sont fréquemment éloignées des valeurs observées des prix. Tel que discuté dans la revue de la littérature, des résultats semblables ont été obtenus par Zareipour et al. (2006) dans leur analyse prédictive des prix de l'électricité en Ontario à partir d'un modèle de régression dynamique (RD) et d'un modèle de fonction de transfert (FT). On note également l'étude de Rodriguez et Anders (2004) qui tirent des conclusions similaires à partir d'une modélisation par perceptrons multicouches (réseau de neurones).

La sensibilité des algorithmes d'apprentissage vis-à-vis les données et leur distribution pourrait expliquer les faibles performances prédictives obtenues à partir des méthodes tirées de l'apprentissage statistique étudiées dans le cadre de l'analyse prédictive du prix de l'électricité sur un horizon temporel de 3 heures. Dans l'objectif de réduire l'importante volatilité présente dans les prix de l'électricité en Ontario, nous appliquons la technique de lissage par moyenne mobile (Box et Jenkins, 1976) avec fenêtre glissante de taille 24 (MA(24)) sur les données de nos échantillons. Les différents modèles prédictifs sont ensuite réestimés à partir des données transformées. Les résultats obtenus des trois modèles prédictifs affichant les RMSE et les MAE minimales, à travers les deux périodes étudiées, sont présentés dans le tableau 5.10. Le détail des résultats est disponible en annexe.

Les résultats montrent que, dans la période "avant", similairement aux résultats obtenus à partir des données brutes, les méthodes conventionnelles de prévision offrent les meilleures performances prédictives en termes de minimisation de la RMSE et de la MAE. En particulier, le modèle de régression linéaire multiple estimé par MCGF (1) montre une RMSE de 5,50\$CAD et le modèle de régression linéaire avec erreur ARIMA (2), une MAE de 6,82\$CAD. Dans la période "après", les modélisations M1 estimées par PCA (15) et par LASSO (11) représentent les modèles optimaux en termes de minimisation de la RMSE et de la MAE, avec une RMSE de 11,95\$CAD et une MAE de 9,91\$CAD, respectivement.

Classement des RMSE			
Avant		Après	
	RMSE		RMSE
(2) Régression linéaire multiple avec erreur ARIMA	6,82	(15) Régression sur composantes principales (PCA) (M1)	11,95
(1) Régression linéaire multiple	6,83	(3) Ridge (M1)	12,00
(3) Ridge (M1)	6,86	(7) LASSO (M1)	12,01

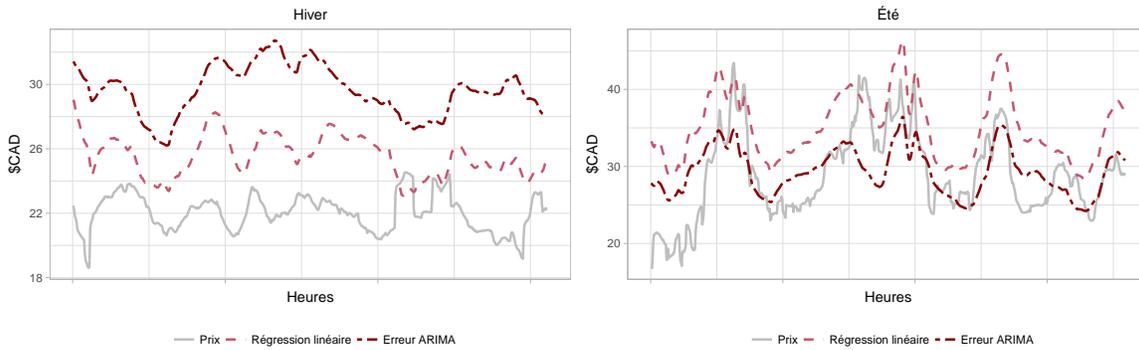
Classement des MAE			
Avant		Après	
	MAE		MAE
(1) Régression linéaire multiple	5,50	(7) LASSO (M1)	9,91
(11) Elastic Net (M1)	5,58	(3) Ridge (M1)	9,92
(7) LASSO (M1)	5,59	(11) Elastic Net (M1)	9,93

Tableau 5.10 – Classement des trois modèles prédictifs étudiés démontrant les meilleures performances prédictives en termes de minimisation de la RMSE et du MAE, pour les prévisions du prix de l’électricité sur un horizon temporel de trois heures dans le marché de l’Ontario calculées à partir des données transformées par moyenne mobile avec fenêtre glissante de taille 24 (MA(24)).

Le graphique 5.7 présente les valeurs prédites à partir des modèles optimaux sélectionnés. Les résultats sont rapportés sur deux mois de nos échantillons tests de la période “avant” et “après”, soit février (hiver) et juillet (été)<sup>9</sup>. Des graphiques, on constate que, similairement aux résultats obtenus à partir des données brutes, les valeurs prédites par les différents modèles sont fréquemment éloignées des valeurs réelles observées. Dans la période “avant”, bien que les modèles semblent être en mesure de prédire certains pics de prix, ils surestiment toutefois le prix observé en périodes de fluctuations faibles. Dans la période “après”, on constate le même problème rencontré à partir des données brutes, c’est-à-dire que les modèles parviennent à prédire la tendance générale de la série temporelle des prix, sans toutefois prédire les variations importantes observées dans les prix de l’électricité dans le marché ontarien.

9. Les résultats sont présentés sur une période de temps plus importante afin de permettre une meilleure visualisation des variations des prix et des prévisions.

### Période “avant”



### Période “après”

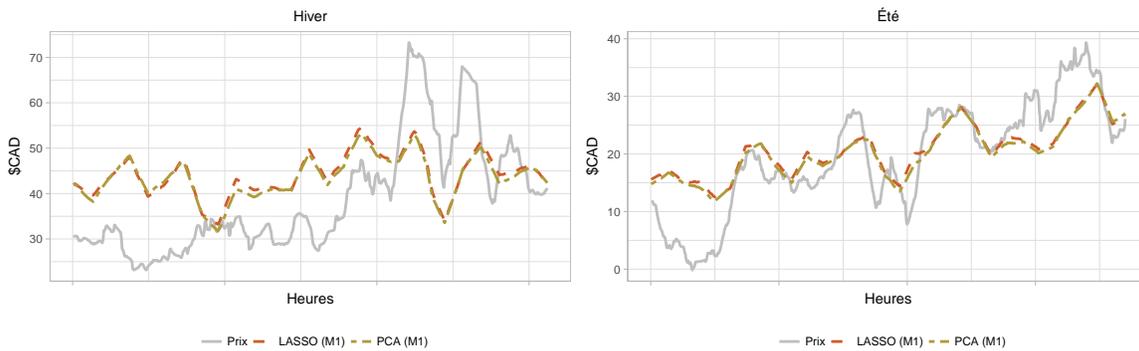


FIGURE 5.7 – Comparaison des valeurs réelles et des valeurs prédites par les modèles optimaux de prévision du prix de l’électricité estimés à partir des données transformées par moyenne mobile avec fenêtre glissante de taille 24 (MA(24)). Les résultats sont rapportés, à travers les deux périodes, sur 2 mois de nos échantillons tests, soit février (hiver) et juillet (été). Les prévisions ont été calculées 3 heures à l’avance.

Les résultats obtenus dans le cadre de ce mémoire ainsi que les résultats observés dans la littérature semblent montrer que les importantes variations des prix de l’électricité dans le marché de l’Ontario ne puissent être prédites à partir des données disponibles, supposant, de ce fait, que la définition du prix de l’électricité ne dépend pas uniquement de paramètres comme l’offre et la demande. Dans la littérature, il est fréquemment argumenté que les pics de prix résultent du pouvoir de marché de certains producteurs d’électricité (appelés les “offreurs pivots”). Les importantes capacités de production de ces producteurs leur permettent la manipulation des prix, notamment en raison de la vulnérabilité du marché de l’électricité (contrainte de capacité de génération, contrainte de capacité de transmission et non-stockabilité de l’électricité) (Zhao et al., 2007). En effet, ceux-ci peuvent, entres autres, retirer une fraction de leur capacité de production afin de faire grimper les prix, particuliè-

rement en périodes de pointe (Marty, 2004). Ce phénomène a notamment été observé dans le marché de l'Alberta par le producteur d'électricité *TransAlta* (Market Surveillance Administrator, 2018), et dans le marché de l'Australie du Sud, par *AGL* (McConnell et Sandiford, 2016). Ce type de comportements est généralement difficilement observable, pouvant, ainsi, expliquer pourquoi aucune conduite de la sorte n'est encore été dénoncée en Ontario. En outre, il est admis que l'incertitude vis-à-vis des prévisions de la demande d'électricité pour les producteurs incite à la spéculation sur le marché (Mount et Oh, 2004) ; en Ontario, cette incertitude est d'autant plus importante étant donnée l'absence d'un marché à 24 heures. De ce fait, les importantes variations des prix de l'électricité sont généralement la conséquence de l'adoption, à court terme, de comportements stratégiques par les producteurs d'électricité, plutôt que des facteurs de marché. Ces événements étant habituellement subjectifs, les pics de prix sont difficiles à prédire à partir de modélisations statistiques.

# Conclusion

L'industrie de l'électricité en Ontario a connu de nombreux changements dans les dernières années. Les récentes modifications à la réglementation ont mené à la libéralisation de l'industrie et à l'introduction d'un marché compétitif de production, impliquant de nouveaux agents dans l'exploitation quotidienne du réseau électrique ontarien. Dans ce nouveau contexte, développer des outils efficaces de prévision de court terme devient alors une nécessité. Pour l'organisme régulateur du marché (IESO), les prévisions de la demande d'électricité constituent les éléments décisifs du bon fonctionnement du marché, et les prévisions des prix sont nécessaires à la planification des opérations quotidiennes. Pour les producteurs d'électricité, les analyses prédictives de court terme de la demande et du prix de l'électricité constituent des composantes importantes dans le développement de stratégies d'enchères optimales afin de maximiser les profits. Ce mémoire propose d'explorer les performances de différentes techniques de prévision de court terme de la demande et du prix de l'électricité. Notre objectif consiste à développer des modèles prédictifs simples et efficaces, pouvant être exploités par les différents agents du marché de l'électricité en Ontario dans leur modélisation d'affaires et leur prise de décision.

En s'inspirant du nouveau courant d'analyse observé dans la littérature, les différents modèles prédictifs étudiés dans la cadre de ce mémoire sont basés sur des techniques tirées de l'apprentissage statistique supervisé. Celles-ci ont notamment l'avantage d'être en mesure de traiter des bases de données de grandes dimensions et de modéliser des relations complexes entre les variables, contrairement aux méthodes conventionnelles de prévision. Les méthodes analysées comprennent les régressions pénalisées (Ridge, LASSO, Elastic Net), la régression sur composantes principales, les arbres de décision et les méthodes d'ensemble aléatoires

et additives (Bagging, Forêts aléatoires et Boosting). Deux méthodes conventionnelles de prévision : la régression linéaire multiple et la régression linéaire avec erreur ARIMA, sont également développées à titre de *benchmark* dans la comparaison des modèles étudiés. Les différents modèles prédictifs ont été appliqués à la prévision de la demande d'électricité sur un horizon temporel d'une heure, et à la prévision du prix de l'électricité sur un horizon temporel de 3 heures, étant donné l'importance de ce type de prévision dans le processus décisionnel des agents du marché de l'électricité en Ontario.

Les résultats de l'analyse prédictive de court terme de la demande d'électricité en Ontario montrent que, de manière générale, les techniques étudiées tirées de l'apprentissage statistique présentent de meilleures performances prédictives pour la prévision de la demande d'électricité une heure à l'avance. En particulier, les modèles linéaires Ridge (M2) et Elastic Net (M3) représentent les modèles optimaux en termes de minimisation de la RMSE et de la MAPE, à travers les deux périodes étudiées. En outre, les résultats permettent de conclure que les modèles optimaux offrent également de bonnes performances prédictives sur différents horizons temporels de prévision, notamment ceux de 2 heures, 3 heures et 25 heures.

L'analyse prédictive de court terme des prix de l'électricité en Ontario représente une tâche plus complexe, étant donné la présence d'une importante volatilité dans les prix du marché. En effet, les prix de l'électricité en Ontario montrent une volatilité supérieure à celle observée dans les marchés des régions voisines, en plus de présenter une proportion élevée de prix inhabituels (*price spike*), généralement causés par des événements non anticipés sur le marché. Dans ce contexte, les résultats obtenus montrent que l'utilisation des méthodes tirées de l'apprentissage statistique ne présente pas un avantage significatif vis-à-vis des méthodes conventionnelles de prévision. Nous observons que, de manière générale, les modèles prédictifs construits dans le cadre de ce mémoire parviennent à prédire convenablement les prix de l'électricité en Ontario dans les périodes où les fluctuations des prix sont faibles, mais performant pauvrement dans les périodes où les prix connaissent d'importantes variations à travers les heures d'une même journée. Il est fréquemment argumenté que les importantes variations des prix résultent de l'adoption de comportements stratégiques par les producteurs d'électricité, ainsi que de la spéculation sur le marché. L'utilisation de modèles prédictifs

basés sur la théorie des jeux serait ainsi susceptible d'améliorer les résultats obtenus dans le cadre de cette recherche vis-à-vis des prévisions du prix dans le marché de l'électricité ontarien. En outre, des techniques plus complexes de l'apprentissage statistique, tels que les réseaux de neurones, ayant les capacités d'estimer des relations non linéaires compliquées entre les variables et de modéliser une forte volatilité présente dans les données, pourrait également fournir de meilleures prévisions du prix de l'électricité en Ontario.

# Bibliographie

- Aggarwal, S. K., L. M. Saini et A. Kumar. 2009, «Electricity price forecasting in deregulated markets : A review and evaluation», *International Journal of Electrical Power & Energy Systems*, vol. 31, n° 1, p. 13–22.
- Archer, C. L. et M. Z. Jacobson. 2005, «Evaluation of global wind power», *Journal of Geophysical Research : Atmospheres*, vol. 110, n° D12.
- Benth, F. E., J. S. Benth et S. Koekebakker. 2008, *Stochastic modelling of electricity and related markets*, vol. 11, World Scientific.
- Box, G. E. et G. M. Jenkins. 1976, «Time series analysis. Forecasting and control», dans *Holden-Day Series in Time Series Analysis, Revised ed.*, San Francisco : Holden-Day, 1976.
- Breiman, L. 1996, «Bagging predictors», *Machine learning*, vol. 24, n° 2, p. 123–140.
- Breiman, L. 2001, «Random forests», *Machine learning*, vol. 45, n° 1, p. 5–32.
- Breiman, L., J. H. Friedman, R. A. Olshen et C. J. Stone. 1984, *Classification and Regression Trees*, Wadsworth Publishing Company.
- Bremer, R. 1995, «Outliers in statistical data», *Technometrics*, vol. 37, n° 1, p. 117–118.
- Bromley, M. 2009, «Degree days, understanding heating and cooling degree days», <http://www.degreedays.net/introduction>.
- Bunn, D. W. 2000, «Forecasting loads and prices in competitive power markets», *Proceedings of the IEEE*, vol. 88, n° 2, p. 163–169.

- Bushnell, J. B., E. T. Mansur et C. Saravia. 2008, «Vertical arrangements, market structure, and competition : An analysis of restructured US electricity markets», *American Economic Review*, vol. 98, n° 1, p. 237–66.
- Chen, T. 2014, «Introduction to boosted trees», *University of Washington Computer Science*.
- Contreras, J., R. Espinola, F. J. Nogales et A. J. Conejo. 2003, «ARIMA models to predict next-day electricity prices», *IEEE transactions on power systems*, vol. 18, n° 3, p. 1014–1020.
- Danai, B., J. Kim, A. I. Cohen, V. Brandwajn et S.-K. Chang. 2001, «Scheduling energy and ancillary service in the new Ontario electricity market», dans *Power Industry Computer Applications, 2001. PICA 2001. Innovative Computing for Power-Electric Energy Meets the Market. 22nd IEEE Power Engineering Society International Conference on*, IEEE, p. 161–165.
- Dudek, G. 2015, «Short-term load forecasting using random forests», dans *Intelligent Systems' 2014*, Springer, p. 821–828.
- Efron, B. et R. J. Tibshirani. 1994, *An introduction to the bootstrap*, CRC press.
- Elith, J., J. R. Leathwick et T. Hastie. 2008, «A working guide to boosted regression trees», *Journal of Animal Ecology*, vol. 77, n° 4, p. 802–813.
- Fremeth, A., G. Holburn, M. Loudermilk et B. Schaufele. 2017, «The Economic Cost of Electricity Generation in Ontario», <https://www.ivey.uwo.ca/cmsmedia/3776559/the-economic-cost-of-electricity-generation-in-ontario-april-2017.pdf>.
- Friedman, J., T. Hastie et R. Tibshirani. 2001, *The elements of statistical learning*, vol. 1, Springer series in statistics New York.
- Friedman, J., T. Hastie et R. Tibshirani. 2010, «Regularization paths for generalized linear models via coordinate descent», *Journal of statistical software*, vol. 33, n° 1, p. 1.
- Friedman, J. H. 2001, «Greedy function approximation : a gradient boosting machine», *Annals of statistics*, p. 1189–1232.

- Garcia, R. C., J. Contreras, M. Van Akkeren et J. B. C. Garcia. 2005, «A GARCH forecasting model to predict day-ahead electricity prices», *IEEE transactions on power systems*, vol. 20, n° 2, p. 867–874.
- Genc, T. S., E. Yazgan et P.-O. Pineau. 2015, «Electricity Trade Patterns in a Network», dans *Progress in Clean Energy, Volume 2*, Springer, p. 669–692.
- Gouvernement de l’Ontario. 2016, «L’Ontario et le Québec travaillent ensemble pour stimuler la croissance économique», <http://news.ontario.ca>.
- Greene, W. H. 2003, *Econometric analysis*, Pearson Education India.
- Guirguis, H. S. et F. A. Felder. 2004, «Further advances in forecasting day-ahead electricity prices using time series models», *KIEE International Transactions on power engineering*, vol. 4, n° 3, p. 159–166.
- Hahn, H., S. Meyer-Nieberg et S. Pickl. 2009, «Electric load forecasting methods : Tools for decision making», *European journal of operational research*, vol. 199, n° 3, p. 902–907.
- Hippert, H. S., C. E. Pedreira et R. C. Souza. 2001, «Neural networks for short-term load forecasting : A review and evaluation», *IEEE Transactions on power systems*, vol. 16, n° 1, p. 44–55.
- Hoerl, A. E. et R. W. Kennard. 1970, «Ridge regression : Biased estimation for nonorthogonal problems», *Technometrics*, vol. 12, n° 1, p. 55–67.
- Hong, T., M. Gui, M. E. Baran et H. L. Willis. 2010, «Modeling and forecasting hourly electric load by multiple linear regression with interactions», dans *Power and Energy Society General Meeting, 2010 IEEE*, IEEE, p. 1–8.
- Hyndman, R. J. et G. Athanasopoulos. 2014, *Forecasting : principles and practice*, OTexts.
- IESO. 2014, «2014 Electricity Production, Consumption, Price and Dispatch Data», <http://www.ieso.ca/corporate-ieso/media/year-end-data/2014>.
- IESO. 2015a, «2015 Year-End Data : Electricity Production, Consumption, Price and Dispatch Data», <http://www.ieso.ca/corporate-ieso/media/year-end-data/2015>.

- IESO. 2015b, «Ontario Energy Report, Q4 2015», <http://www.ontarioenergyreport.ca/pdfs>.
- IESO. 2018a, «Market Renewal-Day-Ahead Market», <http://www.ieso.ca/en/sector-participants/market-renewal/market-renewal-day-ahead-market>.
- IESO. 2018b, «Market Rules for the Ontario Electricity Market», <http://www.ieso.ca/sector-participants/market-operations/market-rules-and-manuals-library>.
- IESO. 2018a, «How the Wholesale Electricity Price is Determined», <http://www.ieso.ca/learn/electricity-pricing/how-the-wholesale-electricity-price-is-determined>.
- IESO. 2018b, «IA progress report on contracted electricity supply - 4th quarter 2016.», .
- IESO. 2018c, «Perspectives de planification de l'Ontario», .
- IESO. 2018d, «Public data online», <http://www.ieso.ca/en/power-data/data-directory>.
- James, G., D. Witten, T. Hastie et R. Tibshirani. 2013, *An introduction to statistical learning*, vol. 112, Springer.
- Kotsiantis, S., D. Kanellopoulos et P. Pintelas. 2006, «Data preprocessing for supervised leaning», *International Journal of Computer Science*, vol. 1, n° 2, p. 111–117.
- Kyriakides, E. et M. Polycarpou. 2007, «Short term electric load forecasting : A tutorial», dans *Trends in Neural Computation*, Springer, p. 391–418.
- Lever, J., M. Krzywinski et N. Altman. 2017, «Points of significance : Principal component analysis», *Nature Methods*, vol. 14, n° 7, p. 641–642.
- Liu, F., R. D. Findlay et Q. Song. 2006, «A neural network based short term electric load forecasting in Ontario Canada», dans *Computational Intelligence for Modelling, Control and Automation, 2006 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, IEEE, p. 119–119.
- Ltd., N. C. 2018, «Marginal Cost of Wind and Solar PV Electricity Generation : Impact of Responding to Dispatch Instructions.», <http://www.ieso.ca/sector-participants/engagement-initiatives/engagements/completed/floor-price-review>.

- Ludwig, N., S. Feuerriegel et D. Neumann. 2015, «Putting Big Data analytics to work : Feature selection for forecasting electricity prices using the LASSO and random forests», *Journal of Decision Systems*, vol. 24, n° 1, p. 19–36.
- Market Surveillance Administrator, M. 2018, «Market Surveillance administrator allegations against TransAlta corporation», [http://www.auc.ab.ca/regulatory\\_documents/ProceedingDocuments/2015](http://www.auc.ab.ca/regulatory_documents/ProceedingDocuments/2015).
- Marty, F. 2004, «Déséquilibres et défaillances des marchés électriques : La libéralisation est-elle coupable?», *L'Économie politique*, , n° 4, p. 43–58.
- Mazengia, D. H. et al.. 2008, «Forecasting spot electricity market prices using time series models», dans *Sustainable Energy Technologies, 2008. ICSET 2008. IEEE International Conference on*, IEEE, p. 1256–1261.
- McConnell, D. et M. Sandiford. 2016, «Winds of change : An analysis of recent changes in the South Australian electricity market», .
- Mount, T. et H. Oh. 2004, «On the first price spike in summer», dans *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, IEEE, p. 10–pp.
- Mullainathan, S. et J. Spiess. 2017, «Machine learning : an applied econometric approach», *Journal of Economic Perspectives*, vol. 31, n° 2, p. 87–106.
- NCDC. 2018, «NCDC climate data online», <https://www.ncdc.noaa.gov>.
- Nogales, F. J., J. Contreras, A. J. Conejo et R. Espínola. 2002, «Forecasting next-day electricity prices by time series models», *IEEE Transactions on power systems*, vol. 17, n° 2, p. 342–348.
- O'Mahoney, A. et E. Denny. 2011, «The merit order effect of wind generation on the Irish electricity market», .
- Papalexopoulos, A. D. et T. C. Hesterberg. 1990, «A regression-based approach to short-term system load forecasting», *IEEE Transactions on Power Systems*, vol. 5, n° 4, p. 1535–1547.

- Pardo, A., V. Meneu et E. Valor. 2002, «Temperature and seasonality influences on Spanish electricity load», *Energy Economics*, vol. 24, n° 1, p. 55–70.
- Rodriguez, C. P. et G. J. Anders. 2004, «Energy price forecasting in the Ontario competitive power system market», *IEEE transactions on power systems*, vol. 19, n° 1, p. 366–374.
- Régie de l'Énergie. 2018, «Régie de l'Énergie, Mission», <http://www.regie-energie.qc.ca/regie/mission.html>.
- Sahay, K. B. 2015, «One hour ahead price forecast of Ontario electricity market by using ANN», dans *Energy Economics and Environment (ICEEE), 2015 International Conference on*, IEEE, p. 1–6.
- Saporta, G. 2006, *Probabilités, analyse des données et statistique*, Editions Technip.
- Schwarz, G. et al.. 1978, «Estimating the dimension of a model», *The annals of statistics*, vol. 6, n° 2, p. 461–464.
- Taylor, J. W., L. M. De Menezes et P. E. McSharry. 2006, «A comparison of univariate methods for forecasting electricity demand up to a day ahead», *International Journal of Forecasting*, vol. 22, n° 1, p. 1–16.
- Tibshirani, R. 1996, «Regression shrinkage and selection via the LASSO», *Journal of the Royal Statistical Society. Series B (Methodological)*, p. 267–288.
- Varian, H. R. 2014, «Big data : New tricks for econometrics», *Journal of Economic Perspectives*, vol. 28, n° 2, p. 3–28.
- Vilar, J. M., R. Cao et G. Aneiros. 2012, «Forecasting next-day electricity demand and price using nonparametric functional methods», *International Journal of Electrical Power & Energy Systems*, vol. 39, n° 1, p. 48–55.
- Vucetic, S., K. Tomsovic et Z. Obradovic. 2001, «Discovering price-load relationships in California's electricity market», *IEEE Transactions on power systems*, vol. 16, n° 2, p. 280–286.

- Weron, R. 2006, «Modeling and Forecasting Electricity Loads», *Modeling and Forecasting Electricity Loads and Prices : A Statistical Approach*, p. 67–100.
- Weron, R. 2007, *Modeling and forecasting electricity loads and prices : A statistical approach*, vol. 403, John Wiley & Sons.
- Weron, R. 2014, «Electricity price forecasting : A review of the state-of-the-art with a look into the future», *International journal of forecasting*, vol. 30, n° 4, p. 1030–1081.
- Zareipour, H., K. Bhattacharya et C. A. Cañizares. 2007a, «Electricity market price volatility : The case of Ontario», *Energy Policy*, vol. 35, n° 9, p. 4739–4748.
- Zareipour, H., C. A. Cañizares et K. Bhattacharya. 2007b, «The operation of Ontario’s competitive electricity market : Overview, experiences, and lessons», *IEEE Transactions on Power Systems*, vol. 22, n° 4, p. 1782–1793.
- Zareipour, H., C. A. Cañizares, K. Bhattacharya et J. Thomson. 2006, «Application of public-domain market information to forecast Ontario’s wholesale electricity prices», *IEEE Transactions on Power Systems*, vol. 21, n° 4, p. 1707–1717.
- Zhao, J. H., Z. Y. Dong, X. Li et K. P. Wong. 2007, «A framework for electricity price spike analysis with advanced data mining methods», *IEEE Transactions on Power Systems*, vol. 22, n° 1, p. 376–385.
- Ziel, F. 2016, «Iteratively reweighted adaptive LASSO for conditional heteroscedastic time series with applications to ar–arch type processes», *Computational Statistics & Data Analysis*, vol. 100, p. 773–793.
- Ziel, F. et B. Liu. 2016, «LASSO estimation for *gefcom2014* probabilistic electric load forecasting», *International Journal of Forecasting*, vol. 32, n° 3, p. 1029–1037.
- Zou, H. et T. Hastie. 2005, «Regularization and variable selection via the Elastic net», *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, vol. 67, n° 2, p. 301–320.

## Annexe A – Résultats des estimations

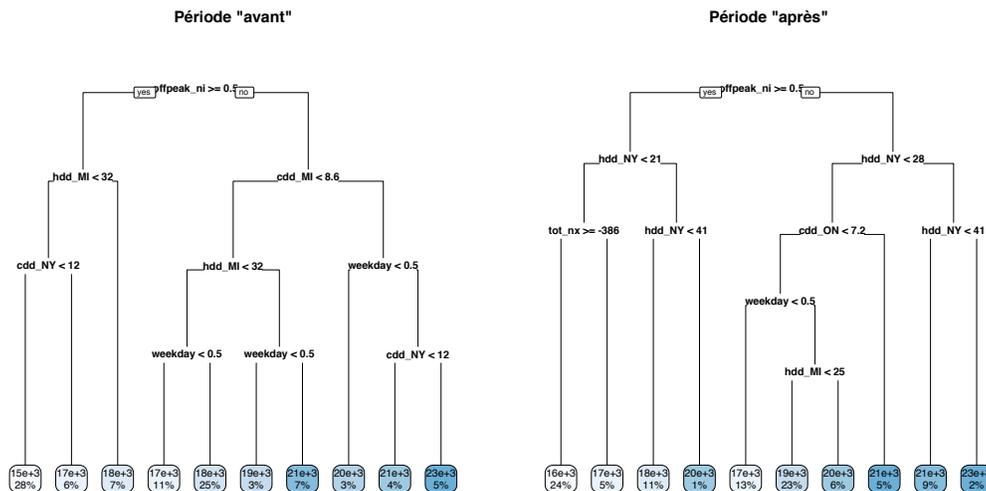
Les codes R nécessaires à l'estimation des modèles étudiés sont disponibles ici.

Ils sont également disponibles à l'adresse suivante :

[https://www.dropbox.com/sh/1fknvj2d7m16x3g/AADVmnyY\\_\\_uZA4jBekcvLlora?dl=0](https://www.dropbox.com/sh/1fknvj2d7m16x3g/AADVmnyY__uZA4jBekcvLlora?dl=0)

## Annexe B – Représentation des arbres de régression

(a) Analyse prédictive de la demande d'électricité



(b) Analyse prédictive des prix de l'électricité

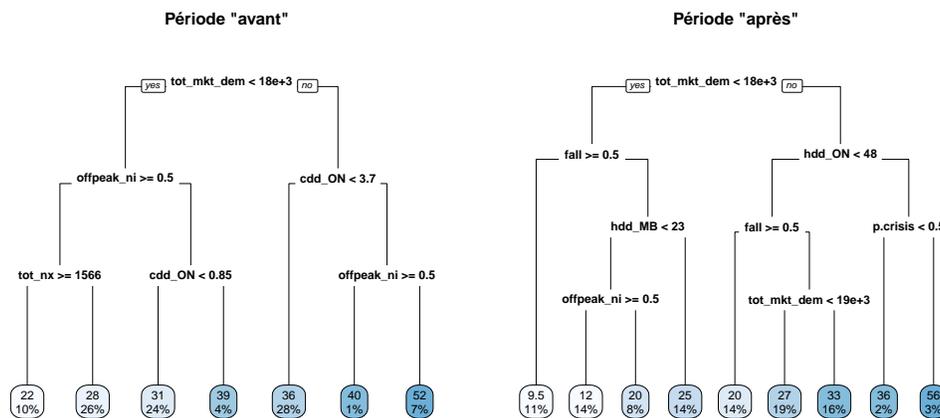
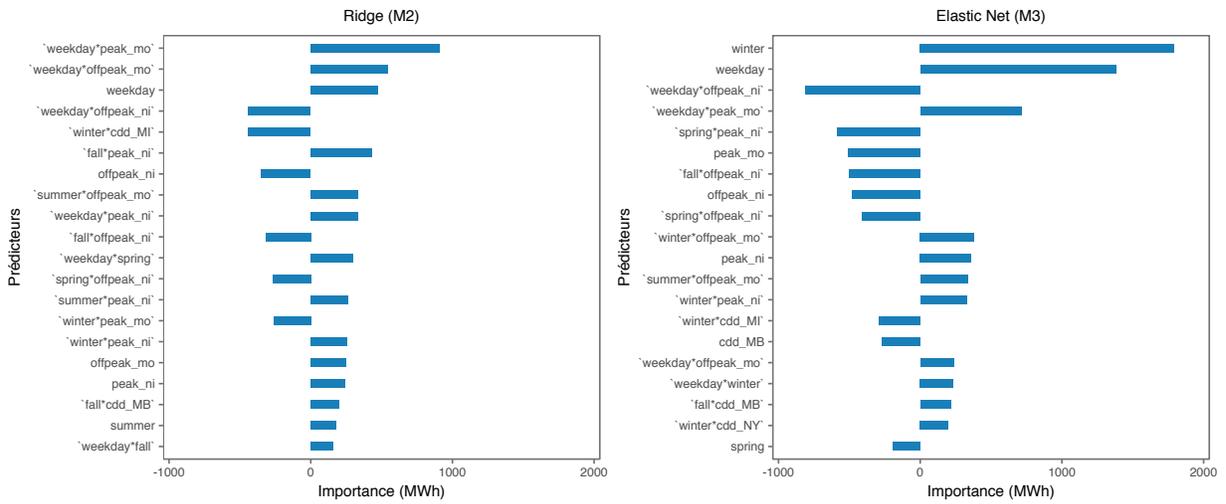


FIGURE 1 – Arbres émondés des périodes "avant" et "après" pour les prévisions horaires de la demande et du prix de l'électricité de l'Ontario sur les horizon temporels de 1 heure et 3 heures, respectivement. Les valeurs encadrées présentent les valeurs prédites de la variable réponse et la proportion d'observations de chaque noeud final.

## Annexe C – Importance des prédicteurs dans les modèles prédictifs optimaux de la demande d'électricité

(a) Période “avant”



(b) Période “après”

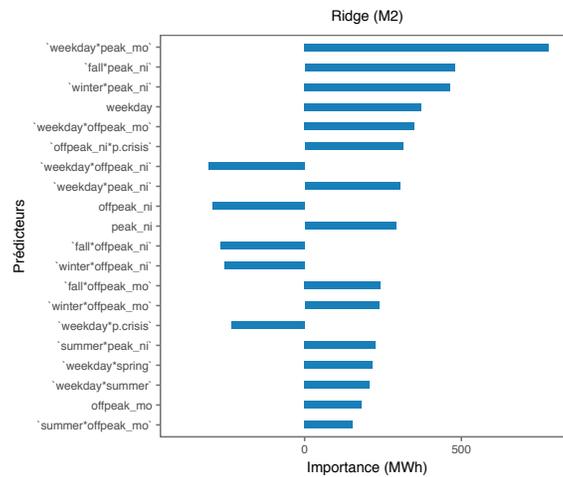
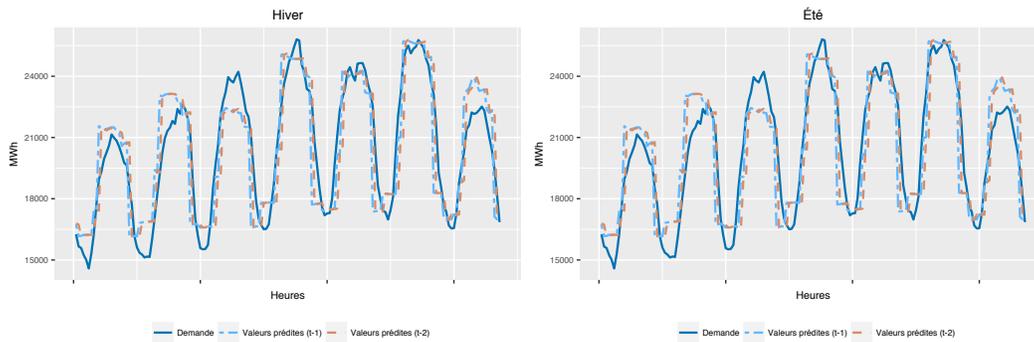


FIGURE 2 – Classement des 20 prédicteurs dont la taille absolue des coefficients est la plus importante dans les modèles optimaux de prévision de la demande d'électricité de la prochaine heure des périodes “avant” et “après”.

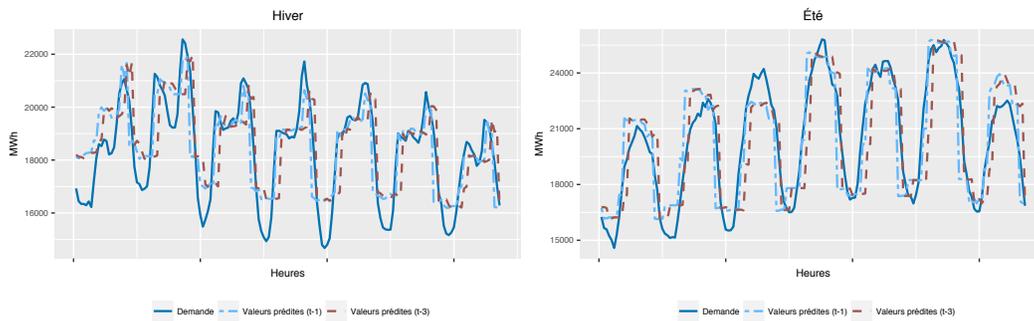
# Annexe D – Modèles prédictifs optimaux de la demande d'électricité sur différents horizons temporels de prévision

Période "avant" : Modèle Elastic Net (M3)

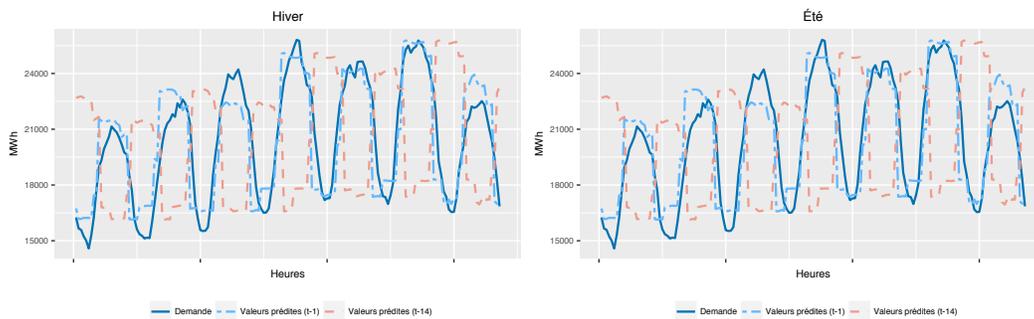
Horizon temporel : 2 heures



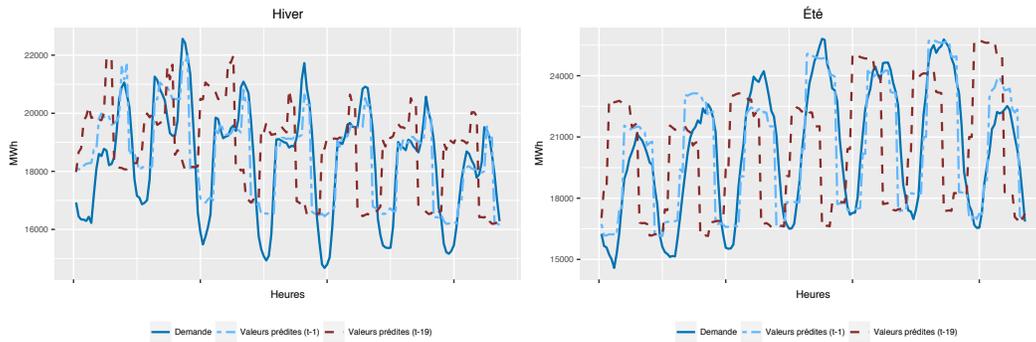
Horizon temporel : 3 heures



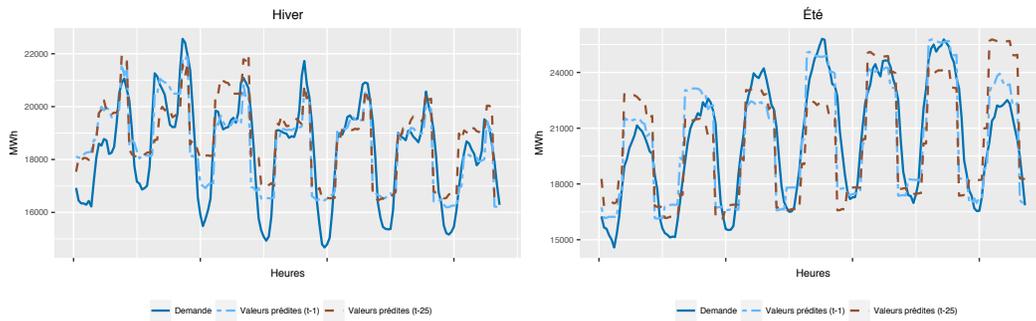
Horizon temporel : 14 heures



### Horizon temporel : 19 heures



### Horizon temporel : 25 heures



### Horizon temporel : 721 heures

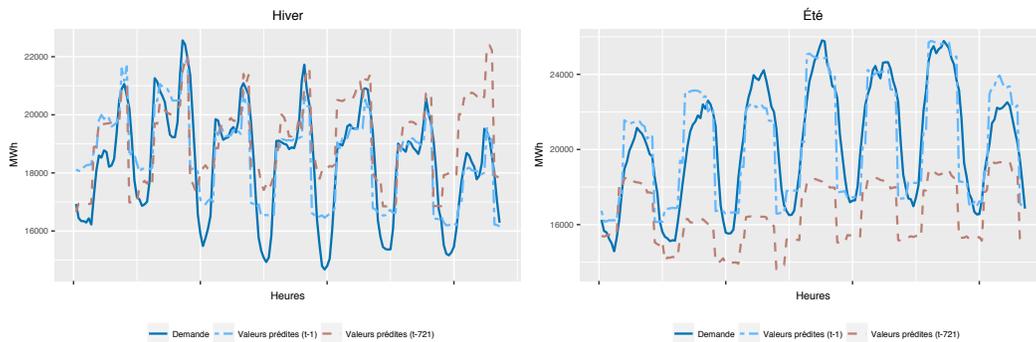
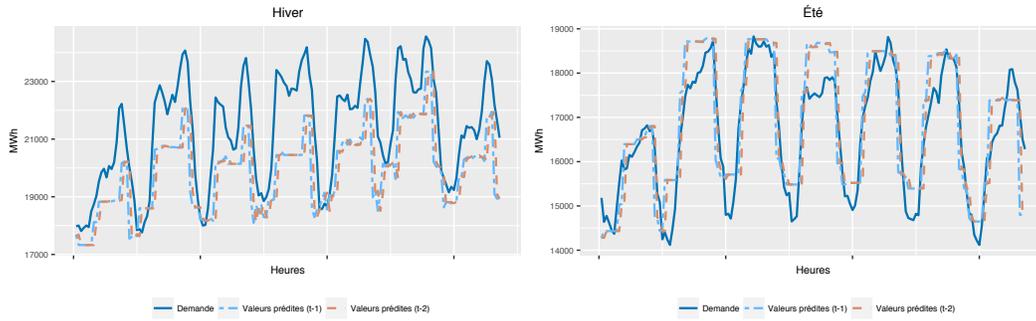


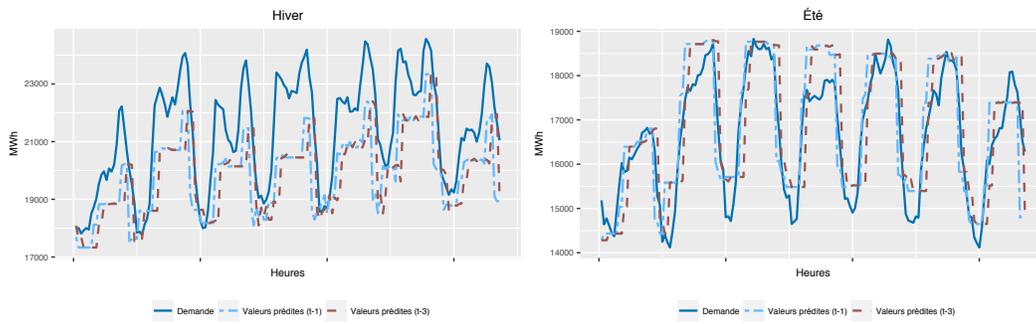
FIGURE 3 – Comparaison des valeurs réelles et des valeurs prédites par le modèle Elastic Net (M3) sur différents horizons de temps. La demande d'électricité horaire en Ontario a été prédite 2, 3, 13, 19, 25 et 721 heures à l'avance. Les valeurs prédites 1 heure à l'avance sont également présentées sur chaque graphique à titre de comparaison. Les résultats sont montrés sur deux semaines de l'échantillon test de la période "avant", soit du 12 au 18 février 2012 (hiver) et du 1<sup>er</sup> au 7 juillet 2012 (été).

# Période "après" : Modèle Ridge (M2)

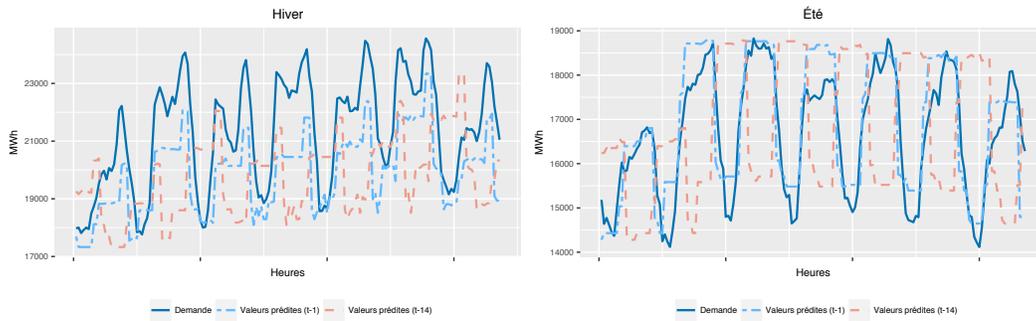
Horizon temporel : 2 heures



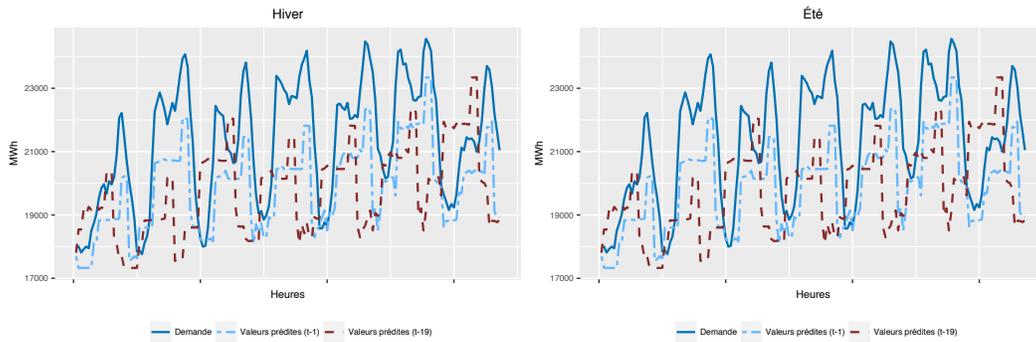
Horizon temporel : 3 heures



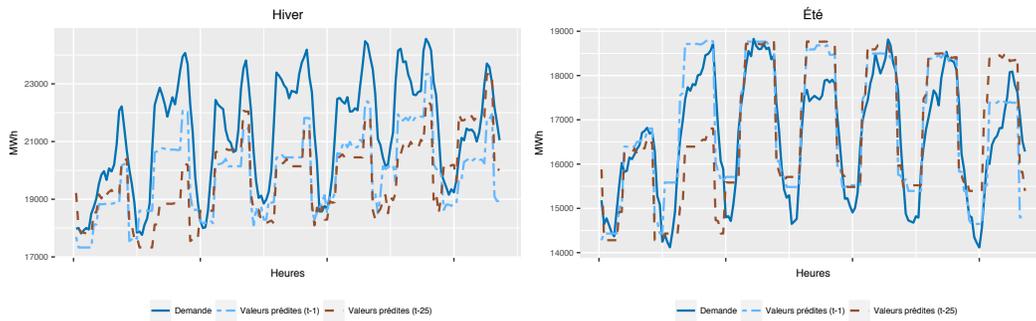
Horizon temporel : 14 heures



### Horizon temporel : 19 heures



### Horizon temporel : 25 heures



### Horizon temporel : 721 heures

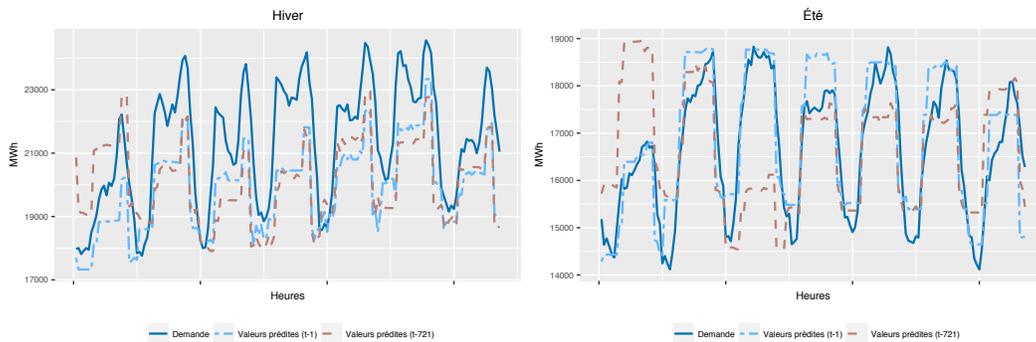


FIGURE 4 – Comparaison des valeurs réelles et des valeurs prédites par le modèle Elastic Net (M3) sur différents horizons de temps. La demande d'électricité horaire en Ontario a été prédite 2, 3, 13, 19, 25 et 721 heures à l'avance. Les valeurs prédites 1 heure à l'avance sont également présentées sur chaque graphique à titre de comparaison. Les résultats sont montrés sur deux semaines de l'échantillon test de la période "après", soit du 8 au 14 février 2015 (hiver) et du 28 juin au 4 juillet 2015 (été).

## Annexe E – Résultats des modèles de régressions linéaires multiples pour l'analyse prédictive des prix de l'électricité en Ontario

	Échantillon 'avant'		Échantillon 'après'	
	Coeff.	(Écart-type)	Coeff.	(Écart-type)
<i>CDD<sub>ON</sub></i>	0,435***	(0,05)	-0,345***	(0,08)
<i>CDD<sub>QC</sub></i>	-0.187	(0,19)	0,305***	(0,02)
<i>CDD<sub>MB</sub></i>	0.107*	(0,05)	0,163***	(0,05)
<i>CDD<sub>MI</sub></i>	0,036	(0,04)	0,559***	(0,05)
<i>CDD<sub>MN</sub></i>	0,201***	(0,03)	0,023	(0,04)
<i>CDD<sub>NY</sub></i>	0,005	(0,03)	0,045***	(0,04)
<i>HDD<sub>ON</sub></i>	0,007**	(0,02)	-0,069***	(0,02)
<i>HDD<sub>QC</sub></i>	-0,065***	(0,01)	0,069***	(0,02)
<i>HDD<sub>MB</sub></i>	-0,062***	(0,01)	0,104***	(0,01)
<i>HDD<sub>MI</sub></i>	0,082***	(0,03)	0,256***	(0,03)
<i>HDD<sub>MN</sub></i>	0,001	(0,00)	-0,117***	(0,02)
<i>HDD<sub>NY</sub></i>	-0,009	(0,02)	0,224***	(0,02)
<i>weekday</i>	-0,751***	(0,20)	1,843***	(0,21)
<i>q<sub>imp</sub></i>	-0,002***	(0,00)	-0,002***	(0,00)
<i>Crisis</i>	-	-	7,240***	(0,49)
<i>dem<sub>ON</sub></i>	0,002***	(0,00)	0,001***	(0,00)
<i>Year</i>	-4,655***	(0,39)	-2,618***	(0,20)
<i>fall</i>	0,517	(0,32)	-2,190***	(0,33)
<i>spring</i>	3,231***	(0,33)	8,556***	(0,84)
<i>summer</i>	0,589	(0,47)	3,880***	(0,41)
<i>winter</i>	-	-	-	-
<i>offpeak<sub>mo</sub></i>	2,476***	(0,24)	0,467	(0,29)
<i>offpeak<sub>ni</sub></i>	-1,406***	(0,32)	-6,912***	(0,34)
<i>peak<sub>mo</sub></i>	3,496***	(0,33)	-0,074	(0,41)
<i>peak<sub>ni</sub></i>	-	-	-	-
Constant	9357,0***	(785,0)	5254,0***	(398,1)
N	14352		17520	
F-Value	389,9		533,1	
<i>R</i> <sup>2</sup>	0,384		0,422	
Adjusted <i>R</i> <sup>2</sup>	0,383		0,421	

Tableau 1 – Résultats des modèles de régressions linéaires multiples estimés par MCGF (modèles optimaux) pour la prévision du prix de l'électricité en Ontario sur l'horizon temporel de 3 heures.

## Annexe F – Résultats des modèles prédictifs des prix sur les données transformés

		Avant		Après	
		RMSE	MAE	RMSE	MAE
		(\$CAD)	(\$CAD)	(\$CAD)	(\$CAD)
(1)	Régression linéaire multiple	6,83	5,50	12,78	10,47
(2)	Régression linéaire avec erreur ARIMA	6,82	5,69	27,66	25,54
Ridge					
(3)	M1	6,87	5,64	12,00	9,92
(4)	M2	8,71	7,27	-	-
(5)	M3	9,23	7,56	-	-
(6)	M4	77,68	16,21	-	-
LASSO					
(7)	M1	6,87	5,58	12,03	9,91
(8)	M2	8,74	7,37	-	-
(9)	M3	9,54	8,15	-	-
(10)	M4	9,42	7,70	-	-
Elastic Net					
(11)	M1	6,88	5,58	12,05	9,93
(12)	M2	8,40	6,91	-	-
(13)	M3	8,85	7,37	-	-
(14)	M4	311,58	42,84	-	-
Régression sur composantes principales (PCA)					
(15)	M1	6,85	5,64	11,94	-
(16)	M2	-	759,53	-	-
(17)	M3	101,08	29,46	-	23,39
(18)	M4	-	475,75	348,87	275,32
(19)	Arbre de régression	9,12	8,11	13,79	10,95
(20)	Bagging	9,32	8,12	15,45	12,07
(21)	Forêts aléatoires	8,71	7,83	14,73	11,66
(22)	Boosting	19,75	18,53	15,31	12,11

Tableau 2 – RMSE et MAE des modèles prédictifs étudiés pour les prévisions du prix de l'électricité sur un horizon temporel de 3 heures dans le marché de l'Ontario calculées à partir des données transformées par moyenne mobile avec fenêtre glissante de taille 24 (MA(24)).

