

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PRÉVISION À COURT TERME DU BESOIN ÉLECTRIQUE QUÉBÉCOIS

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN INFORMATIQUE

PAR

CHARLINE DAVID

JUILLET 2023

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.04-2020). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de recherche, Alexandre Blondin Massé, pour son soutien, ses conseils et son encadrement qui ont rendu possible la réalisation de ce mémoire. J'étends aussi ces remerciements à la faculté des sciences de l'UQÀM ainsi que tous les professeur·e·s et enseignant·e·s qui m'ont permis d'en arriver jusqu'ici.

Je remercie ensuite l'IREQ pour m'avoir offert l'opportunité de travailler en partenariat avec elle, et je remercie particulièrement Patrick Jeandroz, chef de l'unité Science des données et calcul haute performance, et Arnaud Zinflou, chercheur et membre du projet « Préviation de la demande ».

Je voudrais également témoigner ma reconnaissance envers ma famille, mes amis et mon partenaire, Eric Lavallée, dont la patience et les encouragements durant ces longs mois ont été cruciaux pour mettre à terme la réalisation de ce travail.

TABLE DES MATIÈRES

LISTE DES TABLEAUX	vi
LISTE DES FIGURES	vii
RÉSUMÉ	ix
INTRODUCTION	1
CHAPITRE I GÉNÉRALITÉS SUR LES SÉRIES TEMPORELLES	6
1.1 Caractéristiques d'une série temporelle	8
1.1.1 L'opérateur de retard	8
1.1.2 Motifs et décomposition	9
1.1.3 Stationnarité et transformations	11
1.1.4 Évaluation des prévisions	17
1.2 Spécification des variables	18
1.2.1 La série temporelle cible	18
1.2.2 Les entrées exogènes	21
1.2.3 Méthodes de spécification auxiliaires	24
1.3 La situation du Québec	27
1.3.1 Tendances	28
1.3.2 Saisonnalités	29
1.3.3 Jours spéciaux	33
1.3.4 Variables météorologiques	37
CHAPITRE II ÉTAT DE L'ART AVEC EXEMPLES DE MODÈLES DANS LA LITTÉ- RATURE	47
2.1 ARIMA et ses variantes	48
2.1.1 Le polynôme de retard	50

2.1.2	ARMA(p, q)	50
2.1.3	Généralisations du modèle	52
2.1.4	Implémentation de Cancelo <i>et al.</i> (2008)	53
2.2	Équations multiples sur séries temporelles	59
2.2.1	Implémentation de Clements <i>et al.</i> (2016)	61
2.3	N-BEATS, un modèle de réseau de neurones profond	67
2.3.1	Implémentation de Marchand et Zinflou (2021)	71
CHAPITRE III		
MODÈLE PROPOSÉ		77
3.1	Modélisation	78
3.1.1	Saisonnalité hebdomadaire \mathbb{S}^{sem}	79
3.1.2	Saisonnalité annuelle $\mathbb{S}^{\text{année}}$	80
3.1.3	Corrélation intrajournalière \mathbb{S}^{jour}	80
3.1.4	Variables météorologiques	82
3.1.5	Températures \mathbb{M}^T	83
3.1.6	Autres variables météorologiques $\mathbb{M}^N + \mathbb{M}^V$	88
3.1.7	Angle du soleil \mathbb{M}^S	90
3.1.8	Heure d'été DST	91
3.2	Cadre expérimental	92
3.3	Méthode d'évaluation	93
3.3.1	Limitations	95
3.4	Résultats	95
3.4.1	Effets des saisonnalités	95
3.4.2	Effets météorologiques	100
3.4.3	Effets de l'heure d'été	103
3.4.4	Retour sur la corrélation intrajournalière	104
3.4.5	Analyse du meilleur modèle	105

3.5	Jours spéciaux	109
3.5.1	Preuve de concept	109
3.5.2	Résultats préliminaires	111
	CONCLUSION	115

LISTE DES TABLEAUX

Tableau		Page
1.1	Dates du calendrier grégorien où se tiennent les fêtes d'origine non-canadienne	35
3.1	MAPE globaux de modèles ne possédant que des composantes de saisonnalité	96
3.2	MAPE globaux du meilleur modèle à composantes de saisonnalité avec ajout de composantes météorologiques	101
3.3	MAPE globaux de modèles possédant une composante d'heure d'été	103
3.4	Retour sur les MAPE globaux de modèles possédant la sous-composante de dernière prévision intrajournalière	104
3.5	MAPE horaires du meilleur modèle par jour de la semaine	107
3.6	MAPE globaux de modèles possédant une composante de jours spéciaux	113

LISTE DES FIGURES

Figure	Page
1.1	Consommation électrique horaire au Québec entre 2012 et 2020 6
1.2	Variance de la consommation électrique au Québec (a) avant et (b) après la transformation logarithmique 10
1.3	Décomposition STL de la consommation électrique au Québec pendant le mois de février 2019 12
1.4	Diagrammes ACF (a) de la consommation horaire au Québec et (b) de la différentiation de premier ordre de la consommation horaire au Québec 14
1.5	Logarithme de la consommation électrique horaire au Québec entre 2012 et 2019 17
1.6	Tendance de la consommation québécoise entre 2012 et 2020 29
1.7	Saisonnalité annuelle de la consommation québécoise 29
1.8	Saisonnalité hebdomadaire de la consommation québécoise 30
1.9	Saisonnalité journalière de la consommation québécoise 31
1.10	Saisonnalité journalière de la consommation québécoise selon la période de l'année et le jour de la semaine. La ligne rouge correspond aux jours ouvrables et la ligne bleue correspond aux fins de semaines 31
1.11	Consommation électrique au Québec la semaine d'un lundi férié, comparé à la semaine précédente 36
1.12	Corrélation des attributs météorologiques avec la consommation électrique au Québec 38
1.13	Corrélation entre la température et la consommation 39
1.14	Effets de la température sur la consommation pour chaque pas de temps 40

1.15	Segmentation des effets de la température avec (a) un point de rupture et (b) quatre points de rupture	41
1.16	Corrélation de la température mesurée aux différentes stations météorologiques	42
1.17	Effets de la différence de température sur la consommation	43
1.18	Segmentation des effets de la nébulosité	44
1.19	Segmentation des effets de la vitesse du vent	45
1.20	Segmentation des effets de l'angle du soleil	46
2.1	Consommation et températures au Queensland entre juillet 1999 et décembre 2013 (Clements <i>et al.</i> , 2016)	65
2.2	MAPE demi-horaires et globaux de la prévision journalière de juillet 2002 à décembre 2013 (Clements <i>et al.</i> , 2016)	68
2.3	Visualisation de l'architecture N-BEATS (Oreshkin <i>et al.</i> , 2019)	70
2.4	Prédictions individuelles et d'ensemble pour chaque bloc de validation (Marchand et Zinflou, 2021)	76
2.5	MAPE horaires pour des prévisions individuelles et d'ensemble (Marchand et Zinflou, 2021)	76
3.1	MAPE horaires des variantes de modèles	99
3.2	MAPE horaires du meilleur modèle par jour de la semaine	106
3.3	MAPE horaires du modèle proposé comparé aux résultats de Cancelo <i>et al.</i> (2008), Clements <i>et al.</i> (2016), et Marchand et Zinflou (2021)	108
3.4	Profils des coefficients des groupes de jours spéciaux selon \mathbb{J}_2	112
3.5	MAPE horaires par jour de la semaine pour un modèle avec une composante de jours spéciaux	113

RÉSUMÉ

Nous étudions la prévision du besoin énergétique québécois pour les 24 prochaines heures à l'aide d'un ensemble de données s'étalant sur une durée de 8 ans de valeurs historiques de consommation électrique au Québec. Il s'agit d'un problème de prévision de la charge à court-terme, un sous-domaine de la prévision des séries temporelles.

Notre modélisation repose sur une méthode appelée « modèle à équations multiples sur séries temporelles » pour la prévision de la consommation électrique à court terme, en nous inspirant de l'approche proposée par Clements, Hurn et Li (2016) appliquée à la région du Queensland, en Australie. Après avoir adapté le modèle aux conditions québécoises, nous l'étendons en prenant en compte deux variables météorologiques additionnelles, la nébulosité et la vitesse du vent, en plus de considérer les observations de différentes stations météorologiques dispersées sur le territoire québécois. Le modèle proposé présente un score MAPE de 1,32 % sur l'ensemble de validation, avec le potentiel de réduire davantage sa marge d'erreur après l'exploration de pistes d'améliorations accessibles.

Les performances de notre modèle sont comparées à celles des modèles à équations multiples de Clements, Hurn et Li (2016) et de Cancelo, Espasa et Grafe (2008); ainsi que sur celles du modèle N-BEATS de Marchand et Zinflou (2021).

INTRODUCTION

Au Québec, plus de 99 % de l'électricité est provient d'énergie renouvelable, principalement générée à partir de centrales hydroélectriques et d'éoliennes. Vers la fin des années 1970, la demande d'électricité au Québec a commencé à augmenter de plus en plus fortement, tout en devenant de plus en plus sensible aux conditions météorologiques (Grenier, 2006). La conversion des systèmes de chauffage au mazout à l'électricité, incitée par les nouveaux programmes politiques et commerciaux faisant la promotion de l'utilisation de l'électricité comme principale source d'énergie, a conduit à une variabilité de la consommation électrique importante.

La prévision de la demande électrique à court terme est une activité capitale pour Hydro-Québec (Marchand et Zinflou, 2021). Elle est nécessaire pour soutenir des objectifs aussi variés que la gestion de la production ou encore la gestion et la maintenance du réseau électrique. Toute erreur importante de prévision peut se traduire par des problèmes de fiabilité, par une perte d'opportunité ou des coûts supplémentaires pour Hydro-Québec, alors qu'une bonne prévision lui permettrait de générer des ventes supplémentaires sur les marchés voisins. Pour bien équilibrer l'offre et la demande, Hydro-Québec a développé une expertise et des outils internes permettant d'estimer la charge électrique sur le réseau québécois. Cependant, une analyse de la performance du modèle paramétrique employé par Hydro-Québec entre 1989 et 2017 a mis en évidence une augmentation moyenne de l'erreur du modèle depuis 2010 (Marchand et Zinflou, 2021).

La prévision de la demande d'énergie a toujours été une tâche stimulante qui suscite l'intérêt du milieu universitaire et du secteur industriel (Grenier, 2006;

Cancelo *et al.*, 2008; Clements *et al.*, 2016; Rafiei *et al.*, 2018; Feng et Zhang, 2020). En fonction de la durée de l’horizon temporel concerné, la prévision de la consommation électrique peut être classée en différents types (Nassif *et al.*, 2022; Gs et Shiruru, 2016; Metaxiotis *et al.*, 2003) :

- La prévision de charge à long terme (*LTLF*) est utilisée pour prédire la croissance de la demande prévue sur plusieurs mois à plusieurs années afin que les services publics et les gouvernements puissent planifier l’expansion des services, l’achat d’équipement ou l’embauche de personnel. Ce type de prévision dépend principalement des prévisions de variations démographiques et de développement urbain à grande échelle.
- La prévision de la charge à moyen terme (*MTLF*) est généralement effectuée pour des périodes futures allant de quelques semaines jusqu’à quelques années afin de conclure des contrats d’approvisionnement en combustible ou de planifier des activités de maintenance. Ce type de prévision dépend des prévisions des variations météorologiques au cours de l’année et peut intégrer des événements majeurs affectant la consommation dans ses modèles de prévision.
- La prévision de la charge à court terme (*STLF*), quant à elle, a un horizon temporel de projection très court et est généralement destinée à planifier des réponses aux variations imminentes de l’offre ou de la demande. La durée de l’horizon pour ce type de prévision varie normalement de quelques heures jusqu’à quelques jours à l’avance. Son objectif est généralement le désir d’orienter les opérations quotidiennes de production et de transport de l’électricité, de maintenir la qualité de l’approvisionnement et d’éventuellement atteindre une réduction des coûts en anticipant les variations ponctuelles de la demande et les pics de demande extrêmes.

Diverses approches ont été utilisées pour construire des modèles de prévision de la

charge à court terme ; les références Kondaiah *et al.* (2022) et Nassif *et al.* (2022) en présentent une revue détaillée et récente. Les modèles quantitatifs, tels que la moyenne mobile intégrée autorégressive saisonnière (SARIMA) avec régression à vecteurs de support (SVR) (Maldonado *et al.*, 2019), la régression non paramétrique (Wang et Hu, 2015) ou les équations multiples sur séries temporelles (Clements *et al.*, 2016) se sont avérés fiables, précises et interprétables. Plusieurs travaux, en particulier ces dernières années, ont également utilisé avec succès d'autres approches d'apprentissage automatique, allant des réseaux de neurones artificiels (Hippert *et al.*, 2001; Zhang et Wang, 2018; Aouad *et al.*, 2022) et de l'apprentissage par renforcement (Feng *et al.*, 2019; Park *et al.*, 2020) au méta-apprentissage (Wu *et al.*, 2021) ou aux réseaux de neurones en graphes (Lin *et al.*, 2021). Les techniques d'apprentissage profond sont très prometteuses, mais leur nature de boîte noire et la complexité de la planification de leur formation et du maintien de leur cycle de vie restent problématiques (Joshi *et al.*, 2021).

Le projet « Préviation de la demande », mené par l'Institut de recherche d'Hydro-Québec (IREQ), vise à améliorer les outils de prévision de la demande en utilisant les récentes avancées en apprentissage profond et en traitement de données massives. Dans le cadre de ce projet, nous étudions la prévision du besoin québécois pour les 24 prochaines heures à l'aide d'un ensemble de données s'étalant sur une durée de 8 ans de valeurs historiques de consommation électrique au Québec. Il s'agit d'un problème de série temporelle à plusieurs dimensions correspondant aux nombreux facteurs dont dépend le besoin électrique québécois.

La principale contribution de ce mémoire consiste à adapter l'approche à équations multiples sur séries temporelles pour la prévision de la consommation électrique à court terme proposée par Clements *et al.* (2016) pour la prévision de la consommation électrique de toute la province de Québec, au Canada. Initié dans Ramanathan *et al.* (1997), des améliorations significatives ont été proposées

aux modèles d'équations multiples sur séries temporelles, notamment par Cancelo *et al.* (2008) et, plus récemment, par Clements *et al.* (2016). Bien que les modèles de séries temporelles à équations multiples semblent moins populaires dans la littérature, nous pensons qu'ils présentent des avantages qui les rendent intéressants dans un contexte de calcul haute performance.

En plus d'adapter le modèle aux températures québécoises, nous l'étendons en prenant en compte deux variables météorologiques additionnelles, la nébulosité et la vitesse du vent, en plus de considérer les observations de différentes stations météorologiques à plusieurs endroits sur le territoire québécois. Le modèle que nous décrivons est facilement interprétable, son entraînement peut être parallélisé avec de légères modifications à son implémentation, il peut être entraîné très rapidement (en quelques secondes sur un ordinateur portable standard) et il a une précision de prévision compétitive. Le modèle proposé présente au final un score MAPE de 1,32 % sur les données de l'année 2019.

Les performances de ce modèle sont comparées à celles des modèles à équations multiples de Clements *et al.* (2016) sur la consommation australienne et Cancelo *et al.* (2008) sur la consommation espagnole, ainsi que sur ceux du modèle N-BEATS de Marchand et Zinflou (2021), qui est lui aussi un modèle élaboré dans le cadre du projet « Prévision de la demande ».

La suite de ce mémoire est divisée comme suit.

Le chapitre 1 présente des concepts de bases des domaines de l'analyse et de la prévision de séries temporelles, illustrés à l'aide d'exemples sur l'ensemble de données mis à notre disposition. Nous présentons d'abord les caractéristiques d'une série temporelle ainsi que les transformations et les décompositions qui sont communément appliquées, puis les variables utiles à la prévision d'une série temporelle et leurs méthodes de spécification. Nous présentons finalement notre ensemble de

données plus en détail, incluant son étendue, ses attributs, ainsi que les transformations qui lui sont appliquées.

Le chapitre 2 présente une revue de littérature du domaine de la prévision de la consommation électrique à court terme. Après une revue générale de l'état de l'art, les sections 2.1, 2.2 et 2.3 entrent respectivement plus en détail dans le fonctionnement des modèles ARMA, des modèles à équations multiples sur séries temporelles, et des modèles de réseaux de neurones artificiels. De plus, chacune de ces trois sections présente un exemple d'implémentation particulier, dont les résultats servent ensuite de point de comparaison pour le modèle développé dans le chapitre suivant.

Le chapitre 3 présente notre modèle à équations multiples sur séries temporelles. Ce modèle est basé sur celui de Clements *et al.* (2016), qui est présenté au chapitre précédent. La section 3.1 introduit d'abord le modèle implémenté et les méthodes de modélisation de ses composantes variées. La section 3.2 explique les avantages et les limitations du cadre expérimental et la section 3.3 détaille la méthode utilisée pour évaluer le modèle. La section 3.4 présente finalement les résultats obtenus par les diverses variantes du modèle proposé, ainsi que l'analyse de l'influence de ses diverses composantes, et introduit la possibilité d'ajouter une composante modélisant les effets des jours spéciaux sur la consommation électrique.

CHAPITRE I

GÉNÉRALITÉS SUR LES SÉRIES TEMPORELLES

Une **série temporelle** Y_t est une série de données indicée de façon chronologique. La figure 1.1 présente la principale série temporelle étudiée dans ce mémoire. Les données disponibles couvrent la période du 31 décembre 2011 au 29 mai 2020, qui s'étend sur un total de 3 072 jours, avec un pas de temps horaire. Chaque valeur de cette série temporelle correspond à la consommation électrique dans la province de Québec en mégawatt-heure (MWh) tel que mesuré par Hydro-Québec. Ces mesures excluent l'énergie consommée par l'industrie lourde, l'énergie

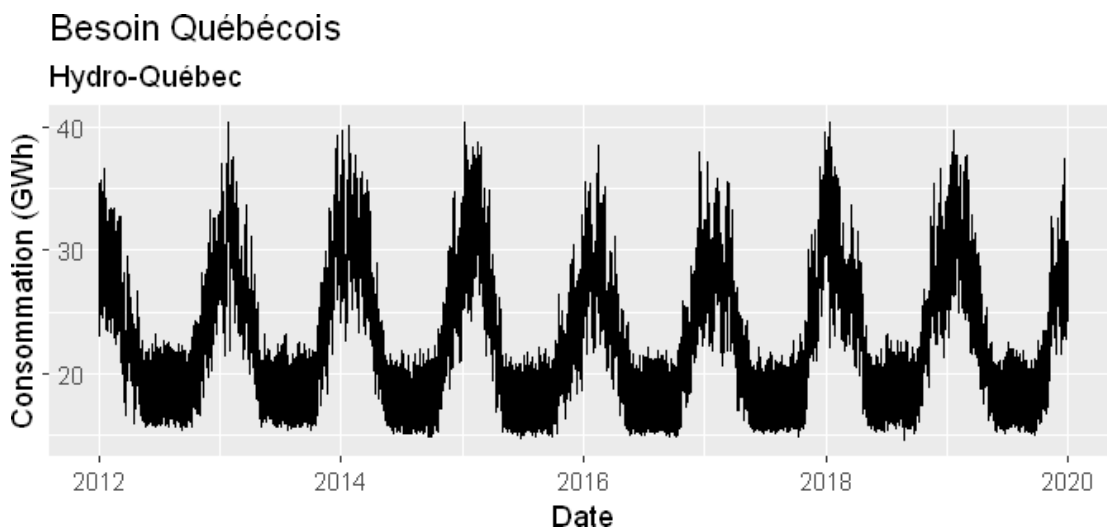


FIGURE 1.1 Consommation électrique horaire au Québec entre 2012 et 2020

exportée hors du Québec, et certains réseaux autonomes, comme celui des îles-de-la-Madeleine.

Deux domaines d'étude qui découlent de l'utilisation de ce type de séries sont l'analyse de séries temporelles et la prévision de séries temporelles. L'**analyse de séries temporelles** se caractérise par l'utilisation de procédés d'analyse sur les données de ces séries afin d'en extraire des statistiques ou d'autres caractéristiques significatives. La **prévision de séries temporelles**, quant à elle, est caractérisée par l'utilisation d'un ou plusieurs modèles dans le but de prédire les valeurs futures d'une série temporelle en fonction de l'information disponible, incluant les valeurs précédemment observées ainsi que tout autre facteur pouvant influencer la prévision.

Certains événements sont cependant plus faciles à prévoir que d'autres. Par exemple, l'heure du lever ou du coucher du soleil peut être prévue avec une précision exacte, mais les prochains numéros de loto gagnants sont aléatoires et ne peuvent donc pas être prévus aussi facilement. La prévisibilité d'une série temporelle dépend entre autres d'à quel point nous comprenons les facteurs qui y contribuent, de la quantité de données historiques disponibles, d'à quel point le futur est similaire au passé, et de si les prévisions peuvent affecter les valeurs que nous essayons de prévoir (Hyndman et Athanasopoulos, 2021). Puisque ces quatre conditions sont généralement satisfaites, les prévisions à court terme de la demande d'électricité résidentielle peuvent atteindre une grande précision.

Les principaux attributs des séries de demandes électriques ont été maintes fois rapportés dans la littérature : tendance, niveaux de saisonnalité superposés, dynamique à court terme, journées particulières ou spéciales, effets non linéaires des variables météorologiques et possible dépendance temporelle non linéaire (Espasa *et al.*, 1996). Ces attributs ne sont pas constants, car les séries observées présentent

une grande variété de profils, selon l'importance relative de chaque composante sous-jacente et la manière dont elles interagissent (Cancelo *et al.*, 2008).

1.1 Caractéristiques d'une série temporelle

Dans cette section, nous voyons tout d'abord l'opérateur de retard, qui offre une manière alternative d'exprimer les valeurs d'une série temporelle décalées dans le temps afin de simplifier certaines opérations. Nous voyons ensuite quelques caractéristiques communément utilisées pour analyser et traiter les séries temporelles, telles que la décomposition en motifs, la propriété de stationnarité, et les transformations qui peuvent être appliquées à cette série pour la rendre stationnaire. Nous terminons en présentant une mesure d'évaluation pour les approches de prévision de séries temporelles.

1.1.1 L'opérateur de retard

L'**opérateur de retard** \mathbb{L} est l'opérateur qui, appliqué sur un élément d'une série temporelle Y_t , produit l'élément précédent, c'est-à-dire $\mathbb{L}Y_t = Y_{t-1}$. L'opérateur peut être composé de manière à étendre l'ampleur du décalage. Ainsi, pour tout entier non négatif s , on a :

$$\mathbb{L}^s Y_t = Y_{t-s} \quad (1.1)$$

Lorsque l'opérateur de retard est utilisé pour dénoter un décalage d'un multiple d'une journée, nous le noterons avec le suffixe j . Un décalage d'une journée serait par exemple dénoté \mathbb{L}^j . Dans ce contexte, la lettre j peut être interprétée comme ayant une valeur égale au nombre de pas de temps par jour.

1.1.2 Motifs et décomposition

Les données de séries temporelles peuvent présenter une variété de motifs, et il est souvent utile de diviser une série temporelle en plusieurs composantes, chacune représentant une catégorie de motif sous-jacent. L'équation (1.2) décrit la **décomposition additive**, alors que l'équation (1.3) décrit la **décomposition multiplicative**, une alternative couramment utilisée dans le traitement des séries temporelles économiques.

$$Y_t = T_t + S_t + R_t \quad (1.2)$$

$$Y_t = T_t \times S_t \times R_t \quad (1.3)$$

où les valeurs Y_t sont les données, T_t est la composante de cycle-tendance, S_t est la composante saisonnière, et R_t est la composante résiduelle, toutes au pas de temps t . Ces composantes servent à exprimer les motifs sous-jacents d'une série temporelle et sont détaillées dans la section ci-bas. L'utilisation d'une décomposition additive sur une série temporelle qui a d'abord subi une transformation logarithmique de manière à ce que sa variation paraisse stable dans le temps équivaut à utiliser une décomposition multiplicative, car l'égalité $Y_t = T_t \times S_t \times R_t$ entraîne $\log Y_t = \log T_t + \log S_t + \log R_t$.

La décomposition additive est la plus appropriée si l'ampleur des fluctuations saisonnières autour du cycle-tendance ne varie pas avec le niveau de la série temporelle. Lorsque la variation du motif saisonnier autour du cycle-tendance semble proportionnelle au niveau de la série temporelle, une décomposition multiplicative est alors plus appropriée (Hyndman et Athanasopoulos, 2021).

La figure 1.2 présente la variation de la consommation électrique horaire au Québec et de son logarithme en fonction du niveau moyen de la consommation sur une fenêtre glissante de quatre semaines. La ligne horizontale pointillée est la moyenne

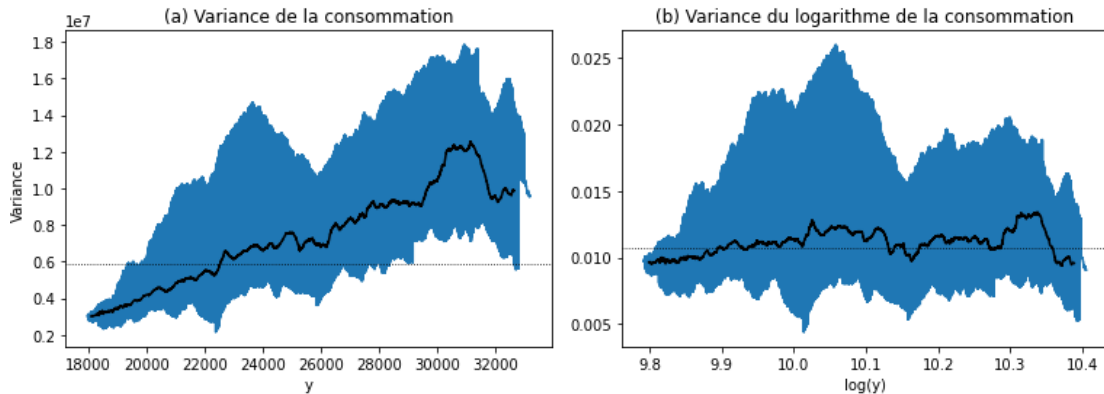


FIGURE 1.2 Variance de la consommation électrique au Québec (a) avant et (b) après la transformation logarithmique

des variances. Comme le démontrent ces deux figures, l'amplitude de la variance de la consommation électrique est proportionnelle au niveau de la consommation électrique. La décomposition multiplicative (ou additive après transformation logarithmique) semble donc plus appropriée que la décomposition additive simple pour notre ensemble de données.

1.1.2.1 Composantes

Les composantes de motifs sont interprétables et servent à exprimer ou à extirper des motifs sous-jacents, prévisibles ou non, des données d'une série temporelle. Elles ont les interprétations suivantes :

- La composante de **tendance** est présente dans une série temporelle lorsqu'on observe une augmentation ou une diminution à long terme de la valeur des données. La tendance n'est pas nécessairement linéaire.
- La composante **cyclique** reflète des fluctuations répétées mais non périodiques dans les valeurs de la série temporelle. Des périodes de récession ou d'essor économique sont des exemples de motifs cycliques. Lorsqu'on décom-

pose une série temporelle, on combine généralement la tendance et le cycle en une seule composante **cycle-tendance**, qui est souvent simplement appelée la tendance.

- La composante **saisonnnière** apparaît lorsqu’une série temporelle est affectée par des facteurs saisonniers tels que la période de l’année, le mois ou le jour de la semaine. La période d’une saisonnalité est toujours fixe et connue, contrairement à la composante cyclique.
- La dernière composante, justement appelée la composante **résiduelle**, mais aussi appelée la composante **irrégulière** ou le **bruit**, comprend tout le reste de la série temporelle. Elle décrit les variations aléatoires, irrégulières, ou non expliquées par les autres motifs.

La figure 1.3 présente un exemple de décomposition saisonnalité-tendance utilisant LOESS (ou décomposition STL) de la consommation électrique au Québec pour la période entre le 1^{er} et le 28 février 2019. On y voit le logarithme de la consommation (`log(y)`), la tendance (`trend`), les saisonnalités hebdomadaires (`season_week`) et journalières (`season_day`), et les résidus de la décomposition (`remainder`).

1.1.3 Stationnarité et transformations

Une série temporelle Y_t est dite (faiblement) **stationnaire** si ses propriétés statistiques ne dépendent pas du moment auquel elle est observée (Kwiatkowski *et al.*, 1992) ; c’est-à-dire que sa moyenne ainsi que sa variance sont stables dans le temps. Plus spécifiquement, $E(Y_t)$ doit être constant pour tout t , et $E(Y_t Y_{t+n})$ doit être indépendant de t pour tout $n \in \mathbb{Z}$. Par définition, elle ne possède donc pas de composantes saisonnières ou de tendance, mais elle peut avoir une composante cyclique. Cette propriété doit être prise en compte, car la régression de séries temporelles non stationnaires peut conduire à des régressions fallacieuses.

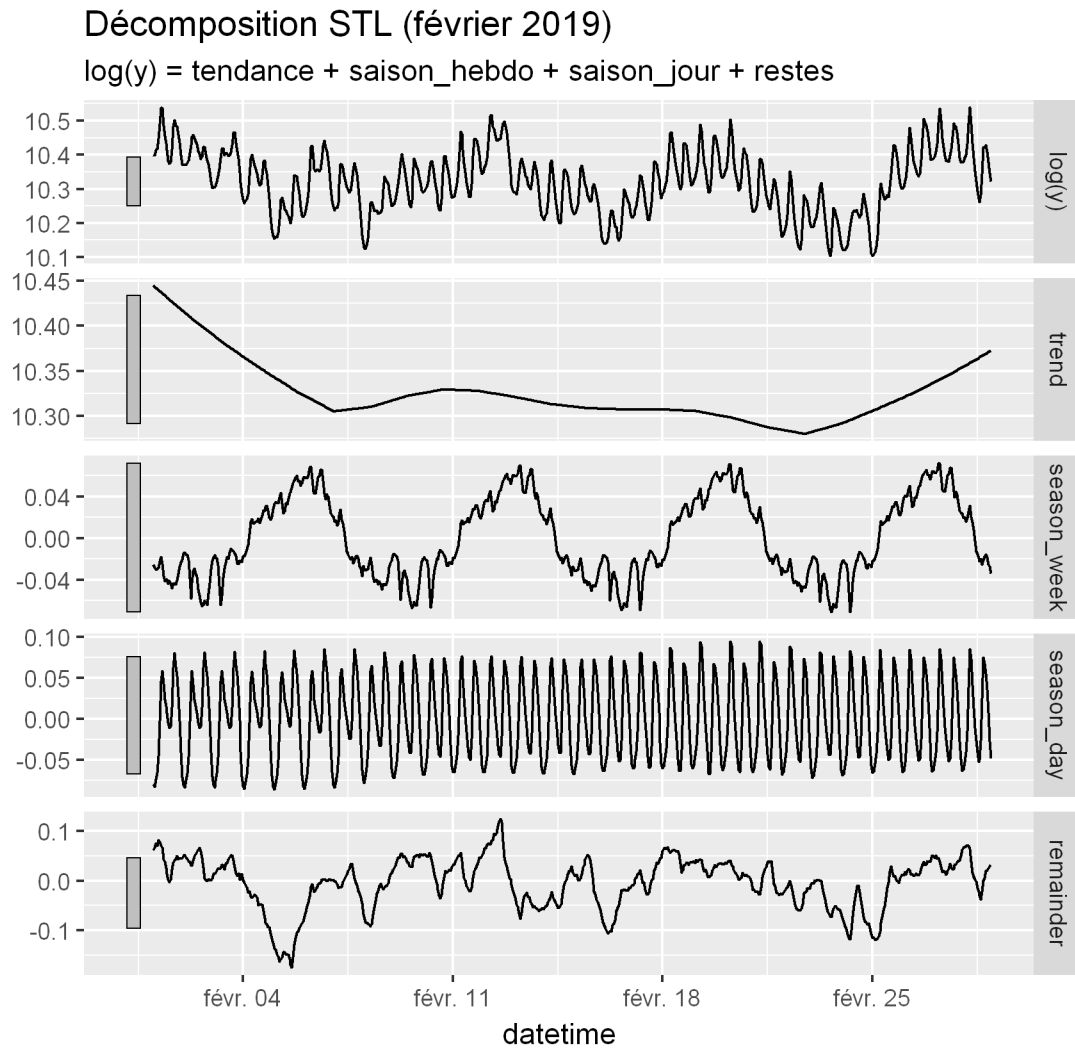


FIGURE 1.3 Décomposition STL de la consommation électrique au Québec pendant le mois de février 2019

Pour les données journalières ou horaires, comme c'est le cas dans l'ensemble de données à notre disposition, travailler avec des données non stationnaires n'induit pas nécessairement de distorsions sérieuses dans un échantillon de quelques mois et une régression fallacieuse peut sembler donner des prévisions raisonnables à court terme. Elles sont cependant généralement inadéquates dans le futur, en particulier lorsqu'on manipule plusieurs années de données à la fois (Hyndman et

Athanasopoulos, 2021; Cancelo et Espasa, 1996).

En pratique, les séries temporelles sont généralement **temporellement dépendantes**, ce qui signifie que les valeurs de chaque pas de temps ne sont pas seulement affectées par des facteurs extérieurs, mais également par les valeurs de la série aux pas de temps passés. Elles sont par conséquent rarement stationnaires. Il est donc préférable de diagnostiquer ces séries temporelles non stationnaires et de trouver une manière de rendre leurs données stationnaires avant de les modéliser.

1.1.3.1 Autocorrélation et bruit blanc

Une méthode pour diagnostiquer les séries temporelles qui ne sont pas stationnaires est de calculer leur **autocorrélation**. L'autocorrélation mesure la relation linéaire entre les valeurs d'une série temporelle Y_t par rapport à une version d'elle-même décalée dans le temps. Le coefficient d'autocorrélation R_k de Y_t pour un décalage temporel de k , qui mesure la relation entre Y_t et Y_{t-k} , est donné par l'équation (1.4) :

$$R_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(\mathbb{L}^k Y_t - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (1.4)$$

où n est la longueur de la série temporelle, et \bar{Y} est la moyenne d'échantillonnage de Y_t . Les coefficients d'autocorrélation constituent la **fonction d'autocorrélation** ou **ACF**. Une série temporelle qui ne présente aucune autocorrélation, c'est-à-dire où R_k est suffisamment près de 0 pour tout $k \geq 1$, est appelée un **bruit blanc**.

Lorsque les données ont une composante de tendance, les autocorrélations pour les petits décalages ont tendance à être importantes et positives, car les observations proches dans le temps sont également proches en valeur. Ainsi, l'ACF d'une série temporelle avec tendance présente des valeurs positives qui diminuent lentement à mesure que les décalages augmentent. Lorsque les données ont une composante

saisonnière, les autocorrélations sont plus importantes pour les décalages saisonniers (aux multiples de la période de saisonnalité) que pour les autres décalages. Lorsque les données ont des composantes à la fois saisonnières et de tendance, on peut voir une combinaison de ces effets (Hyndman et Athanasopoulos, 2021). C'est le cas qu'on observe à la figure 1.4, dont le graphique (a) présente le diagramme ACF de la série temporelle de la consommation électrique horaire au Québec. Le graphique (b) présente le diagramme ACF de la différenciation de premier ordre de cette même série temporelle. On remarque comment cette opération semble avoir éliminé la tendance. Une ou plusieurs différenciations saisonnières avec des périodes de longueurs différentes pourraient ensuite être utilisées pour enlever les composantes saisonnière et rendre cette série temporelle stationnaire. Les deux

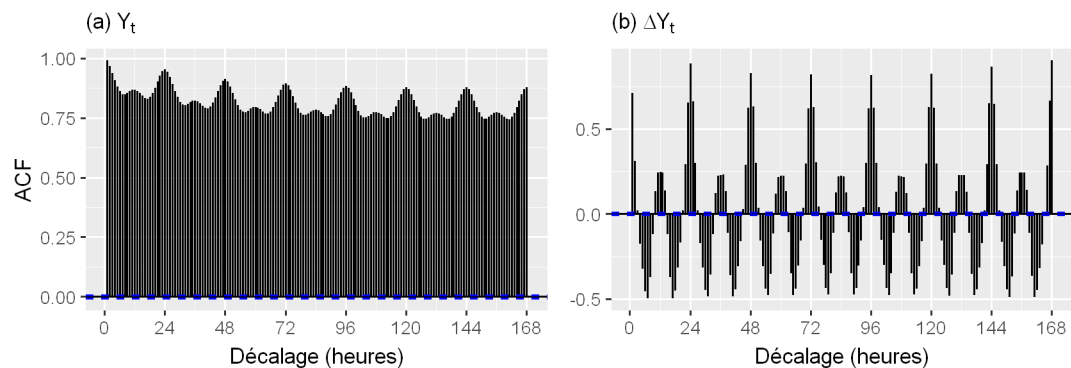


FIGURE 1.4 Diagrammes ACF (a) de la consommation horaire au Québec et (b) de la différenciation de premier ordre de la consommation horaire au Québec

méthodes suivantes, la différenciation et la transformation logarithmique, sont des exemples de méthodes qui permettent de rendre une série temporelle stationnaire.

1.1.3.2 Différentiation

La **différentiation** est un processus utilisé pour éliminer la tendance et stabiliser la moyenne d'une série temporelle Y_t . Une différenciation de premier ordre, par exemple, consiste à utiliser la différence entre deux observations consécutives plutôt que d'utiliser l'observation directement. Il est parfois nécessaire de différencier la série temporelle plusieurs fois en successions afin d'éliminer une tendance polynomiale. On parlera alors de différenciation d'ordre d . La différenciation d'ordre d est décrite par l'équation réursive 1.5, ou de manière équivalente en fonction de l'opérateur de retard \mathbb{L} par l'équation (1.6) :

$$Y_t^{(d)} = \begin{cases} Y_t - Y_{t-1}, & \text{si } d = 1 \\ Y_t^{(d-1)} - Y_{t-1}^{(d-1)}, & \text{si } d > 1 \end{cases} \quad (1.5)$$

$$Y_t^{(d)} = (1 - \mathbb{L})^d Y_t \quad (1.6)$$

L'opérateur de différenciation $(1 - \mathbb{L})^d$ est souvent abrégé Δ^d . Pour capturer une composante saisonnière, on doit plutôt utiliser la **différentiation saisonnière**. La différenciation saisonnière $Y_t^{(d,s)}$ d'ordre d ayant une période de longueur s est décrite par l'équation (1.7), ou de manière équivalente par l'équation (1.8).

$$Y_t^{(d,s)} = \begin{cases} Y_t - Y_{t-s}, & \text{si } d = 1 \\ Y_t^{(d-1,s)} - Y_{t-s}^{(d-1,s)}, & \text{si } d > 1 \end{cases} \quad (1.7)$$

$$Y_t^{(d,s)} = (1 - \mathbb{L}^s)^d Y_t \quad (1.8)$$

L'opérateur de différenciation saisonnier $(1 - \mathbb{L}^s)^d$ peut être abrégé Δ_s^d .

L'opération inverse à la différenciation se nomme l'**intégration**. La lettre I de l'acronyme ARIMA, une famille de modèles de prévision de séries temporelles que nous considérons au chapitre 2, réfère à cette opération.

1.1.3.3 Transformation logarithmique

La méthode de stationnarisation suivante est la **transformation logarithmique**, qui sert à corriger l'hétéroscédasticité d'une série temporelle. Une série temporelle est dite **hétéroscédastique** lorsque sa variance n'est pas stable. Au contraire, une séquence de variables aléatoires est **homoscédastique** si toutes ses variables aléatoires ont une variation constante (ou homogène). Par définition, toute série temporelle stationnaire est obligatoirement homoscédastique.

La pratique habituelle dans la prévision de l'électricité à court terme consiste à travailler avec des données non transformées, bien qu'il soit bien connu que les modèles ARIMA pour les séries économiques mensuelles ou trimestrielles, par exemple, sont généralement basés sur le logarithme des données en raison de leur hétéroscédasticité (Cancelo et Espasa, 1996).

Le but de cette transformations est de simplifier les motifs dans les données historiques en supprimant les sources de variation connues ou en rendant le motif plus cohérent sur l'ensemble des données. Des motifs plus simples sont généralement plus faciles à modéliser et conduisent à des prévisions plus précises (Hyndman et Athanasopoulos, 2021). Cette transformation permet donc de stabiliser la variance d'une série temporelle et potentiellement de la rendre stationnaire.

La figure 1.2 présentée à la section 1.1.2 suggère que l'amplitude de la variance de la consommation électrique au Québec est hétéroscédastique, mais la transformation logarithmique semble appropriée pour les stabiliser. Le résultat de cette transformation sur la consommation horaire au Québec est présentée à la figure 1.5.

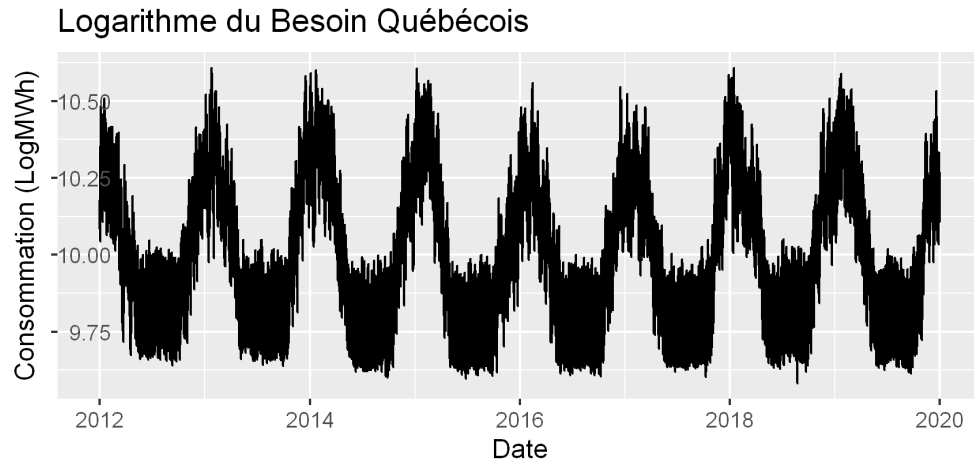


FIGURE 1.5 Logarithme de la consommation électrique horaire au Québec entre 2012 et 2019

1.1.4 Évaluation des prévisions

Il existe de nombreuses mesures d'évaluation applicables aux approches de prévision de séries temporelles. Nous utilisons l'erreur percentile absolue moyenne (MAPE, ou *Mean Average Percentage Error*), une mesure de précision très fréquemment utilisée pour évaluer les résultats de modèles de prévisions de la demande électrique. C'est aussi la mesure utilisée par Hydro-Québec dans le cadre du projet de *Prévision de la demande* (Marchand et Zinflou, 2021). Le score MAPE est exprimé par la formule suivante :

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=0}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$$

où Y_t est la vraie valeur au pas de temps t , \hat{Y}_t est la prévision au pas de temps t , et n est le nombre de prévisions évaluées.

1.2 Spécification des variables

L'analyse et la prévision de séries temporelles se font typiquement avec deux sortes de données : la série temporelle cible elle-même, et un éventail d'autres séries temporelles représentant les potentiels facteurs qui peuvent influencer les valeurs de cette série cible. La manière d'exprimer toutes ces séries peut affecter significativement les performances du modèle. Trouver une spécification adéquate pour les modéliser est donc important. Les méthodes couramment utilisées pour ce faire sont détaillées dans les sous-sections suivantes.

1.2.1 La série temporelle cible

La série temporelle à prévoir Y_t et dont la mesure est déterminée par le modèle est appelée la variable **dépendante, endogène, de prévision, ou cible**. Dans notre cas, elle correspond à la série temporelle présentée à la figure 1.1, c'est-à-dire à la consommation électrique horaire au Québec.

1.2.1.1 Tendances et dépendance temporelle

Plusieurs procédures ont été proposées pour prendre la tendance en compte. Outre la différentiation, certaines études l'ignorent en prélevant de courts échantillons de données où le niveau moyen de la charge électrique peut être supposé constant (Smith, 2000), alors que d'autres considèrent une variété de fonctions temporelles déterministes linéaires (Pardo *et al.*, 2002), linéaires et réciproques (Ramanathan *et al.*, 1997) ou quadratiques (Cottet et Smith, 2003).

Les méthodes utilisées pour expliquer la dépendance temporelle à court terme incluent les autorégressions linéaires (Pardo *et al.*, 2002; Cottet et Smith, 2003), la différentiation fractionnaire (Soares et Souza, 2006), les réseaux de neurones arti-

ficiels (Lamedica *et al.*, 1996; Hippert *et al.*, 2005; Alves da Silva *et al.*, 2008) et le lissage exponentiel de Holt-Winter doublement saisonnier ajusté pour la corrélation d'erreur (Taylor, 2003; Gould *et al.*, 2008). Les procédés ARMA peuvent être avantageux en raison de leur plus grande généralité par rapport aux ajustements résiduels d'autocorrélation et parce qu'ils peuvent gérer une plus grande variété d'effets dynamiques avec moins de paramètres (Cancelo *et al.*, 2008).

1.2.1.2 Saisonnalité

Pour gérer la composante de saisonnalité, on parle plutôt de la différentiation saisonnière et la modélisation par l'ajout de variables indicatrices (*dummy variables*) (Ramanathan *et al.*, 1997; Soares et Souza, 2006), de splines périodiques variables dans le temps (Harvey et Koopman, 1993), ou de séries de Fourier (Clements *et al.*, 2016; Taylor et Letham, 2018; Oreshkin *et al.*, 2019).

En pratique, il n'est pas facile de séparer la saisonnalité annuelle de l'effet des variables météorologiques, avec lesquelles elle est hautement corrélée. En fait, une grande partie de la saisonnalité annuelle peut être expliquée par des variables météorologiques et des variables indicatrices qui tiennent compte des périodes de vacances (Cancelo *et al.*, 2008).

Les méthodes d'ajout de variables indicatrices et les séries de Fourier sont expliquées dans la partie 1.2.3.

1.2.1.3 Saisonnalité complexe

Les séries temporelles dont les données sont entrées à des fréquences quotidiennes et sous-quotidiennes (telles que les données horaires, comme dans notre cas) peuvent être particulièrement difficiles à traiter, car elles impliquent souvent plu-

sieurs motifs saisonniers – chacun avec une période de longueur différente – qui demandent donc d'utiliser une méthode en mesure de gérer une saisonnalité aussi complexe. Pour ce faire, on peut utiliser des méthodes telles que la combinaison d'opérateurs de différentiation saisonniers (Cancelo *et al.*, 2008), l'inclusion de termes de Fourier à différentes fréquences pour correspondre à chaque longueur de période (Clements *et al.*, 2016), ou la combinaison de sous-modèles spécialisés qui traitent chaque heure comme une série différente (Ramanathan *et al.*, 1997; Cottet et Smith, 2003; Cancelo *et al.*, 2008; Clements *et al.*, 2016).

Lorsque les motifs d'une série temporelle sont correctement pris en charge par la combinaison d'opérateurs de différentiations, les racines unitaires positives réelles intégrées dans ces opérateurs génèrent une tendance linéaire locale suffisamment flexible pour approximer une grande variété de situations. Un large échantillon peut alors être utilisé pour estimer le modèle sans avoir à se soucier de la forme de la tendance. De plus, cette méthode n'impose aucune contrainte sur les prévisions, de sorte que tout changement du niveau sous-jacent est automatiquement pris en compte, sans nécessiter d'intervention pour le rectifier (Cancelo *et al.*, 2008).

Dans le cas des opérateurs combinés $\Delta_{1j}\Delta_{7j}$, par exemple, les racines unitaires hebdomadaires au sein de l'opérateur Δ_{1j} permettent une évolution de la saisonnalité hebdomadaire en fonction du jour de la semaine. Pour ensuite obtenir un motif de saisonnalité hebdomadaire qui varie aussi en fonction de la période de l'année, l'ajout des racines unitaires saisonnières dans Δ_{7j} agit comme un mécanisme adaptatif qui ajuste le modèle de la semaine en cours pour s'adapter à la variation annuelle déjà incorporée dans les observations de la semaine précédente. De plus, cette approche n'impose pas la restriction que cette évolution complexe soit la même année après année, comme ce serait le cas si de telles interactions entre saisonnalités étaient modélisées par des fonctions paramétriques.

La combinaison de sous-modèles qui traitent chacun une heure différente de la journée permet de réduire la complexité de la série temporelle dont chaque modèle est responsable en enlevant l'une des multiples composantes saisonnières, celle de la saisonnalité intrajournalière.

1.2.2 Les entrées exogènes

Les autres facteurs pouvant influencer la prévision sont les variables **indépendantes** X , aussi appelées **exogènes**, **prédictives**, **attributs**, **régresseurs**, ou **explicatives**. Les variables exogènes fréquemment utilisées dans le domaine incluent les variables météorologiques (température, humidité, etc.), les particularités de la journée (par exemple, s'il s'agit d'un jour férié), et des données de croissance démographique ou économique (Nassif *et al.*, 2022).

1.2.2.1 Jours spéciaux

Les jours spéciaux sont un facteur connu menant à une consommation anormale. Les jours spéciaux incluent les jours fériés officiels, les fêtes locales et nationales, les longues fins de semaine, et les périodes de vacances. Certaines anomalies, comme les journées d'élections et de grèves ou les événements sportifs et festivals majeurs, peuvent aussi jouer un rôle. Les jours spéciaux peuvent par ailleurs inclure n'importe quel jour environnant ces événements où la consommation est affectée par l'événement.

Les jours spéciaux peuvent être considérés comme faisant partie de la saisonnalité habituelle, mais ils sont normalement plus difficiles à traiter. Plusieurs fêtes peuvent tomber à un jour différent chaque année, telles Pâques et n'importe quelle autre fête célébrée lors d'un jour de la semaine déterminé plutôt qu'à une date fixe. Les « anomalies » mentionnées plus tôt ne devraient, par définition, jamais

être à date fixe.

Le nombre et les types de jours spéciaux sont généralement spécifiés sur des bases a priori, bien que certains auteurs et auteures aient utilisé une analyse de reconnaissance de motifs pour classifier les types de jours empiriquement (Lamedica *et al.*, 1996; López *et al.*, 2022).

Plusieurs études appliquées séparent les jours normaux des jours spéciaux et construisent des modèles différents pour chaque catégorie (Ramanathan *et al.*, 1997; Darbellay et Slama, 2000). Les jours fériés peuvent être lissés (Taylor *et al.*, 2006), traités comme des dimanches (Smith, 2000), ou remplacés par la charge observée lors d'une journée similaire de la semaine précédente (Hippert *et al.*, 2005).

Une bonne façon de traiter les effets des congés mobiles consiste à inclure des variables indicatrices dans le modèle (Hyndman et Athanasopoulos, 2021). Le nombre total de paramètres varie selon les études : un dans Soares et Souza (2006) ; trois dans Pardo *et al.* (2002) ; six dans Clements *et al.* (2016) ; Cottet et Smith (2003) ; 48 dans López *et al.* (2022) ; ou même 53, utilisés par la *Red Eléctrica de España* (López *et al.*, 2019).

1.2.2.2 Variables météorologiques

Parmi la multitude de variables météorologiques pouvant être utilisées pour la prévision de la consommation électrique, la température est normalement la variable ayant de loin la plus grande influence.

La relation entre la consommation électrique et la température est très complexe, et cela est dû à un certain nombre de raisons. A priori, la littérature suggère que cette relation est normalement de nature non linéaire, et souvent en forme de U,

tant qu'il y a suffisamment de variation de température dans l'échantillon (Engle *et al.*, 1986). Cette réponse est asymétrique, puisqu'une augmentation d'un degré lorsque la température est élevée et une diminution d'un degré lorsque la température est basse n'ont pas nécessairement le même impact sur la charge (Valor *et al.*, 2001). La relation diffère pour les jours ouvrables et fériés, en raison des différences entre les équipements et les fonctionnements des systèmes de chauffage et de refroidissement des lieux de travail et des résidences privées (Smith, 2000). Elle peut également changer selon la période de l'année (Hyde et Hodnett, 1997) et, lors de certaines saisons, du jour de la semaine (Cancelo *et al.*, 2008). Il y a un effet dynamique reflétant le fait que les individus s'adaptent aux conditions météorologiques changeantes avec un certain retard, car la charge au pas de temps t ne dépend pas uniquement de la température au pas de temps t , mais aussi des températures des jours précédents (Comte et Warren, 1981). Des **effets d'épuisement** peuvent apparaître en raison de la capacité limitée des appareils de contrôle de la température existants (Henley et Peirson, 1997). Il n'y a donc parfois pas d'augmentation supplémentaire de la consommation électrique lorsque la température est au-dessous ou au-delà de la plage des températures les plus communes.

La forme particulière de la réponse est spécifique à chaque ensemble de données, et des questions telles que le degré de non-linéarité, l'ordre de la dynamique et la présence et l'emplacement des effets d'épuisement, par exemple, doivent être déterminées empiriquement (Cancelo *et al.*, 2008). Pour les modèles de prévisions qui utilisent des équations linéaires, le défi consiste à modéliser cette réponse non linéaire.

La spécification linéaire par morceaux, telle que dans Cancelo *et al.* (2008) ou Clements *et al.* (2016), est expliquée dans la partie 1.2.3. Cette idée est similaire à celle de la méthode des splines flexibles utilisée par Harvey et Koopman (1993).

1.2.3 Méthodes de spécification auxiliaires

Certaines des méthodes mentionnées plus haut sont d'un intérêt particulier pour ce mémoire et méritent une mention spéciale. Les méthodes en question sont l'introduction de variables indicatrices, les séries de Fourier, et la spécification linéaire par morceaux.

1.2.3.1 Variables indicatrices

L'introduction de **variables indicatrices** peut être utilisée pour modéliser la saisonnalité, incluant les jours spéciaux récurrents. Pour ce mémoire, une variable indicatrice utilise la notation suivante :

$$\mathbb{1}_{\{\text{condition}\}}$$

où la variable indicatrice $\mathbb{1}_{\{\text{condition}\}}$ est égale à 1 si la **condition** entre les accolades est vraie, et égale à 0 sinon.

Par exemple, pour prendre en compte la saisonnalité hebdomadaire, les variables indicatrices suivantes peuvent être créées pour représenter les jours de la semaine : $\mathbb{1}_{\{t \in \text{Lun}\}}$, $\mathbb{1}_{\{t \in \text{Mar}\}}$, $\mathbb{1}_{\{t \in \text{Mer}\}}$, $\mathbb{1}_{\{t \in \text{Jeu}\}}$, $\mathbb{1}_{\{t \in \text{Ven}\}}$, et $\mathbb{1}_{\{t \in \text{Sam}\}}$. Pour chaque jour de la semaine $J \in \{\text{Lun}, \text{Mar}, \text{Mer}, \text{Jeu}, \text{Ven}, \text{Sam}\}$, $\mathbb{1}_{\{t \in J\}}$ est la variable indicatrice qui est égale à 1 si $t \in J$ et égale à 0 sinon. Lorsqu'elles sont toutes égales à zéro, $t \in \text{Dim}$ est sous-entendu. Les coefficients associés aux six variables indicatrices peuvent donc être interprétés comme représentant la mesure de l'effet de cette journée de la semaine par rapport à dimanche, la catégorie omise. Cette omission est particulièrement importante pour les modèles qui ont du mal à traiter avec des variables mutuellement dépendantes.

1.2.3.2 Séries de Fourier

Une alternative à l'utilisation de variables indicatrices saisonnières, en particulier pour les longues périodes saisonnières, est d'utiliser des **termes de Fourier**. Cette méthode consiste à modéliser les motifs saisonniers grâce à une série de termes sinus et cosinus à des fréquences judicieusement choisies.

Pour une période saisonnière de longueur s , les termes de Fourier sont décrits par l'équation (1.9) :

$$\sum_{q=1}^Q \phi_{q,1} \sin\left(\frac{2q\pi t}{s}\right) + \phi_{q,2} \cos\left(\frac{2q\pi t}{s}\right) \quad (1.9)$$

où $Q \leq s/2$ est un nombre entier positif, et $\phi_{q,1}$ et $\phi_{q,2}$ ont des valeurs déterminées empiriquement.

De telles séries permettent d'approximer n'importe quelle fonction périodique (Fourier, 1808). Cette méthode a pour avantage de nécessiter l'introduction d'un moins grand nombre de paramètres que l'utilisation de variables indicatrices, de pouvoir s'adapter à des saisonnalités de n'importe quelle longueur, et de permettre de contrôler la granularité du motif saisonnier en sélectionnant une valeur de Q proportionnelle à la granularité souhaitée.

Pour ce mémoire, une paire de termes de Fourier est dénotée $\text{FOURIER}(2\pi X)$ et représente les variables $\sin(2\pi X)$ et $\cos(2\pi X)$.

Lorsqu'un paramètre est associé à une série de Fourier, le paramètre est distribué à chacun des termes et utilise la notation suivante :

$$\sum_{q=1}^Q \phi_q \text{FOURIER}(2q\pi X) = \sum_{q=1}^Q \phi_{q,1} \sin(2q\pi X) + \phi_{q,2} \cos(2q\pi X).$$

où $\phi_{q,1}$ et $\phi_{q,2}$ sont des coefficients dérivés de ϕ_q . Ainsi, ϕ_q représente plusieurs paramètres plutôt qu'un seul.

1.2.3.3 Spécification linéaire par morceaux

Dans une **spécification linéaire par morceaux**, la relation dynamique non linéaire entre une variable et la série temporelle cible est approximée à l'aide d'un système de variables qui servent à représenter linéairement les effets de différents segments ou de différentes plages de valeurs de la variable originelle. La valeur de chacune de ces variables est donnée par une équation de rampe coupée descendante (équation 1.10) ou ascendante (équation 1.11) :

$$\text{CLRAMP}_{a,b}^{-}(X_t) = \begin{cases} b - a, & \text{si } X_t < a \\ b - X_t, & \text{si } a \leq X_t < b \\ 0, & \text{si } b \leq X_t \end{cases} \quad (1.10)$$

$$\text{CLRAMP}_{a,b}^{+}(X_t) = \begin{cases} 0, & \text{si } X_t \leq a \\ X_t - a, & \text{si } a < X_t \leq b \\ b - a, & \text{si } b < X_t \end{cases} \quad (1.11)$$

où X_t est la variable au pas de temps t ; et a et b telles que $a < b$ sont les valeurs limites de l'un des segments. Cette méthode permet donc de modéliser une réponse non linéaire pour les modèles de prévisions qui n'utilisent que des équations linéaires.

En particulier, cette méthode peut servir pour approximer les effets de la température sur la consommation électrique. Les variables du système servent alors à représenter les effets de différentes zones de températures. La « zone de confort » correspond à la plage où les températures sont considérées comme confortables et n'ayant aucun effet supplémentaire sur la consommation électrique. La « zone froide » comprend la plage des températures en dessous de la zone de confort, et la « zone chaude » comprend la plage des températures au dessus de la zone de

confort. Chaque zone peut être sous-divisée en un ou plusieurs segments, où l'effet sur la charge est différent.

La zone de confort n'a aucun effet, et n'a donc aucune variable associée. Les effets de la zone froide et de la zone chaude sont représentées par des **variables de réchauffement** $h_i(T_t)$ (équation 1.12) et des **variables de refroidissement** $c_i(T_t)$ (équation 1.13), respectivement :

$$h_i(T_t) = \text{CLRAMP}_{a_i, b_i}^-(T_t) \quad (1.12)$$

$$c_i(T_t) = \text{CLRAMP}_{a_i, b_i}^+(T_t) \quad (1.13)$$

où T_t est la température au pas de temps t ; et a_i et b_i sont les températures limites du segment i de la zone froide ou chaude.

Lorsque des paramètres sont attribués à une spécification linéaire par morceaux, les coefficients sont distribués à chaque variable formant la spécification linéaire. Cette opération est notée comme suit :

$$\alpha \otimes x(X_t) = \sum_{i=1}^n \alpha_i x_i(X_t)$$

où, x est une spécification linéaire par morceaux modélisant les effets de la série temporelle X_t à l'aide de n variables ; x_1, \dots, x_n sont les n variables de x ; et $\alpha_1, \dots, \alpha_n$ sont les coefficients associés à chacune de ces variables.

1.3 La situation du Québec

La consommation électrique est souvent influencée par des facteurs météorologiques tels que la température, la vitesse du vent, la couverture nuageuse, le type et l'intensité des précipitations (Grenier, 2006). L'effet peut être assez important : par exemple, une augmentation de seulement 1 °C de la température ambiante ou

du réchauffement urbain peut avoir un impact sur le pic de consommation d'électricité ou la consommation mondiale d'un surcoût de plus de 4 % (Santamouris *et al.*, 2015).

La demande d'électricité au Québec est fortement influencée par les conditions climatiques (Grenier, 2006). À des températures inférieures à -20 °C, la consommation due au chauffage des bâtiments représente près de 50 % de la demande québécoise, et l'effet moyen d'un changement de 1 °C sur la demande est d'environ 400 MW. La couverture nuageuse peut quant à elle avoir un impact allant jusqu'à 3 000 MW.

De plus, la consommation d'énergie n'est pas seulement liée aux conditions météorologiques et peut être influencée par plusieurs autres facteurs, tels que le déclin de l'industrie lourde, les nouveaux comportements sociaux des consommateurs (télétravail, client autoproducteur, rôle actif des consommateurs), des informations calendaires (jour de la semaine, jours spéciaux et fériés), l'électrification des transports, l'automatisation des processus, l'isolation des bâtiments et éventuellement le prix de l'électricité (Grenier, 2006; Cancelo *et al.*, 2008).

Dans les parties suivantes, nous analysons les motifs de la consommation électrique au Québec ainsi que les entrées exogènes météorologiques, et nous explorons l'effet des jours spéciaux.

1.3.1 Tendances

La figure 1.6 présente le motif de tendance de la consommation électrique au Québec entre 2012 et 2020. La ligne rouge agrège les données en une moyenne mensuelle, et la ligne noire en une moyenne annuelle. Bien qu'on puisse y voir que la consommation varie d'une année à l'autre, aucune tendance claire ne s'en dégage.

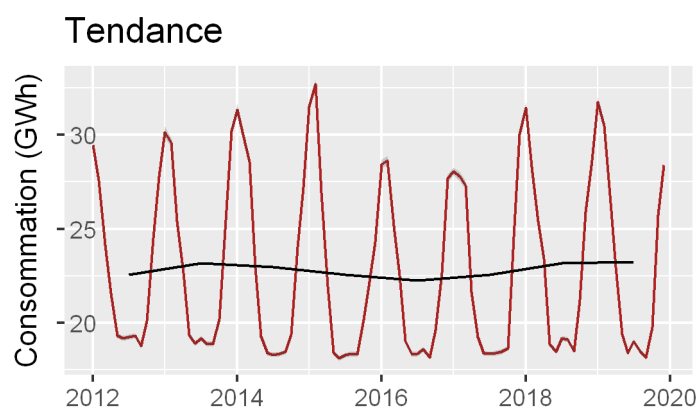


FIGURE 1.6 Tendance de la consommation québécoise entre 2012 et 2020

1.3.2 Saisonnalités

La figure 1.7 présente le motif de la saisonnalité annuelle de la consommation

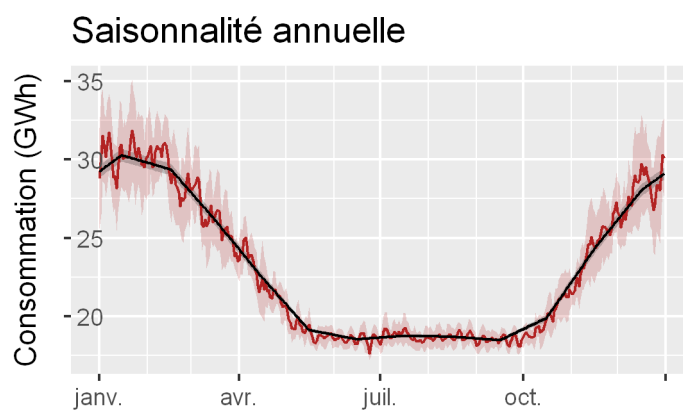


FIGURE 1.7 Saisonnalité annuelle de la consommation québécoise

électrique au Québec. La ligne rouge utilise les données horaires, et la ligne noire agrège les données en une moyenne quotidienne. Dans ces graphiques comme dans le graphique à la figure 1.6, on note une saisonnalité intra-annuelle prononcée dé-

marquant les mois d'été, où la consommation est basse, de ceux d'hiver, où la consommation est élevée. Cette saisonnalité pourrait ressembler à une courbe sinusoïdale presque parfaite s'il n'en était de l'aplatissement prématuré de son creux entre juin et septembre, incluant même un léger rebond en juillet et août. On peut présumer que cette saisonnalité est principalement causée par la consommation électrique des appareils de chauffage et de climatisation, et dépend donc principalement de la température.

La figure 1.8 présente le motif de la saisonnalité hebdomadaire de la consomma-

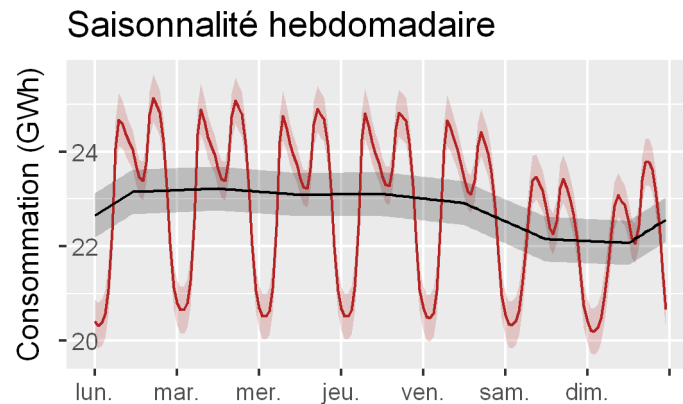


FIGURE 1.8 Saisonnalité hebdomadaire de la consommation québécoise

tion électrique au Québec. La ligne rouge représente les données horaires, et la ligne noire agrège les données en une moyenne quotidienne. On note la consommation nettement plus élevée lors des jours ouvrables que lors de la fin de semaine, un motif qui s'explique par l'activité humaine plutôt qu'à des facteurs météorologiques telle la température.

La figure 1.9 présente le motif de la saisonnalité journalière de la consommation électrique horaire au Québec. On note les deux pointes de consommation, la première vers sept ou huit heures, puis la deuxième autour de 17 heures ; ainsi que

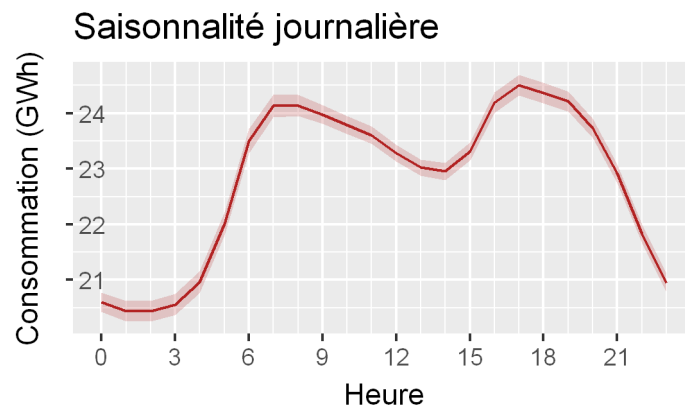


FIGURE 1.9 Saisonnalité journalière de la consommation québécoise

les deux creux, le premier au milieu de la nuit vers une ou deux heures du matin, et le deuxième en après-midi, à 14 heures.

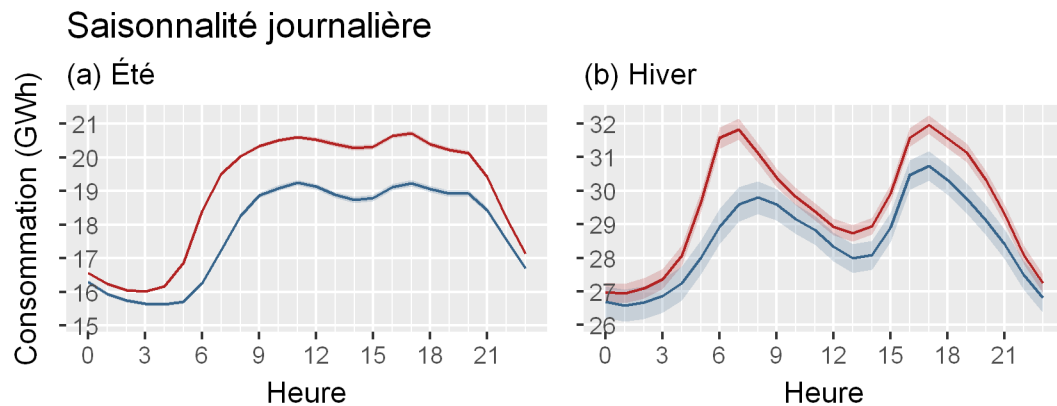


FIGURE 1.10 Saisonnalité journalière de la consommation québécoise selon la période de l'année et le jour de la semaine. La ligne rouge correspond aux jours ouvrables et la ligne bleue correspond aux fins de semaines

Pour mieux comprendre les facteurs qui l'influencent, La figure 1.10 présente cette même saisonnalité hebdomadaire, mais séparé en deux, selon la période de l'année.

Le graphique (a) présente les données de juin à août (les trois mois de l'année les plus chauds), et le graphique (b) présente les données de décembre à février (les trois mois de l'année les plus froids). Les lignes rouges représentent la moyenne des jours ouvrables (lundi à vendredi), et les lignes bleues représentent la moyenne des jours de fin de semaine (samedi et dimanche).

Afin de correctement analyser ces deux graphiques, il est au préalable nécessaire de noter que les entrées de l'ensemble de données sont toutes au fuseau horaire de l'heure standard de l'Est (EST). En été, l'heure affichée sur les horloges des Québécoises et Québécois est donc décalée par rapport à l'heure inscrite dans l'ensemble de données. Par exemple, lors du minimum local de consommation observée à 3 heures du matin en semaine durant l'été dans la figure 1.10(a), l'horloge des Québécoises et Québécois affiche qu'il est en fait deux heures du matin.

Comme mentionné plus tôt, la consommation apparaît réduite lors des fins de semaines par rapport aux jours ouvrables. Cette différence ne peut qu'être expliquée par l'activité humaine, puisque la division entre semaine et fin de semaine n'est pas une construction naturelle.

Bien que l'amplitude totale des extremums à l'intérieur d'une journée soit sensiblement la même en été comme en hiver, les pics sont beaucoup plus graduels (arrondis) et le creux dans la consommation en milieu de journée est beaucoup moins prononcé en été qu'en hiver, suggérant donc que le motif de saisonnalité journalière varie lui-même selon la période de l'année. La variance est notamment plus grande en hiver, comme le démontrent les régions ombrées des graphiques.

Toute l'activité humaine en avant-midi semble être décalée vers la droite lors des fins de semaine, un fait qui est particulièrement évident lors du pic matinal en hiver. Les extrêmums locaux sont à 3 h, 11 h, 14 h, et 17 h (2 h, 10 h, 13 h, et 16 h

EST) en été, et à 1 h, 7 h (jours ouvrables) ou 8 h (fin de semaine), 13 h, et 17 h en hiver. En été, les pics du matin et du soir sont approximativement aussi prononcés l'un que l'autre, lors des jours ouvrables comme la fin de semaine. Cette remarque est aussi vraie pour les jours ouvrables en hiver, mais le pic du matin est plus atténué que le pic du soir lors des fins de semaines. En général, les pics sont moins graduels la semaine que la fin de semaine en hiver. Toutes ces différences peuvent sans doute être expliquées par des facteurs externes, notamment la température, ainsi que par des facteurs humains, comme les périodes de vacances en été.

1.3.3 Jours spéciaux

Au Québec, il existe plusieurs types de jours spéciaux dont la tenue pourrait potentiellement affecter le profil de la consommation électrique. En voici une liste non exhaustive :

- Les congés fédéraux ou provinciaux : le jour de l'an (1^{er} janvier) ; le Vendredi saint (le vendredi précédant Pâques) ou le lundi de Pâques (le lendemain de Pâques), au choix de l'employeur ; la Journée nationale des patriotes ou fêtes de la Reine (le lundi qui précède le 25 mai) ; la Fête nationale du Québec ou Saint-Jean-Baptiste (24 juin), observée le 25 juin si le 24 est un dimanche ; la fête du Canada (1^{er} juillet), observée le 2 juillet si le 1^{er} est un dimanche ; la fête du Travail (le premier lundi de septembre) ; l'Action de grâces (le deuxième lundi d'octobre) ; et Noël (25 décembre).
- Les fêtes non fériées : la Saint-Valentin (14 février) ; la Saint-Patrick (17 mars) ; Pâques (le premier dimanche après la première pleine lune qui suit le 21 mars selon le calendrier grégorien) ; le Poisson d'avril (1^{er} avril) ; la fête des Mères (le deuxième dimanche de mai) ; la fête des Pères (le troisième dimanche de juin) ; l'Halloween (31 octobre) ; et le jour du Souvenir (11 novembre), qui est un jour férié au Canada, mais pas au Québec.

- les fêtes d’origine non canadienne : le Nouvel An chinois (1^{er} *zhēngyuè* du calendrier chinois) ; Roch Hachana ou le Nouvel An juif (le 1 et 2 *tishri* du calendrier hébraïque) ; Yom Kippour (10 *tishri* du calendrier hébraïque) ; le mois de ramadan (du calendrier islamique) ; Aïd-el-Fitr (1 *chawwal* du calendrier islamique), à la fin du ramadan ; et Aïd-al-Adha (du 10 au 13 *dhou al-hijja* du calendrier islamique). Le tableau 1.1 décrit à quelles dates du calendrier grégorien ces jours correspondent.
- Les périodes de vacances : les vacances de la construction (la dernière quinzaine de juillet) ; la semaine de relâche ; et les vacances de Noël, de Pâques, et d’été.
- Les jours environnant un autre événement : le vendredi précédant une fête à date fixe tombant la fin de semaine ; le lundi suivant une fête à date fixe tombant la fin de semaine ; la veille de Noël (24 décembre) ; le lendemain de Noël (26 décembre) ; la veille du jour de l’an (31 décembre) ; le reste du temps des fêtes (27 au 30 décembre) ; et l’après le jour de l’an (du 2 au 5 janvier, approximativement).
- Les anomalies ou irrégularités : les élections fédérales (19 octobre 2015, et 21 octobre 2019) ; et les élections provinciales (4 septembre 2012, 7 avril 2014, et 1^{er} octobre 2018).
- Les autres jours spéciaux : le jour du déménagement (1^{er} juillet) et, plus généralement, la « haute saison » pour les déménageurs du Québec (du 10 juin au 10 juillet) ; le premier lundi après le passage à l’heure d’été.

La Journée nationale de la vérité et de la réconciliation (30 septembre) n’a pas besoin d’être prise en compte dans notre cas, car elle a été instaurée en 2021, donc après la dernière date incluse dans notre ensemble de données.

La figure 1.11 présente la moyenne de la consommation électrique au Québec entre 2012 et 2019 lors de la semaine de quatre lundis fériés distincts (en rouge), ainsi

TABLEAU 1.1 Dates du calendrier grégorien où se tiennent les fêtes d'origine non-canadienne

Fête	Nouvel an chinois	Roch Hachana	Yom Kippour
Calendrier	chinois	hébraïque	hébraïque
Date	1 <i>zhēngyuè</i>	1–2 <i>tishri</i>	10 <i>tishri</i>
2012	01-23	09-17 à 09-18	09-26
2013	02-10	09-05 à 09-06	09-14
2014	01-31	09-25 à 09-26	10-01
2015	02-19	09-14 à 09-15	09-23
2016	02-08	10-03 à 10-04	10-12
2017	01-28	09-21 à 09-22	09-30
2018	02-16	09-10 à 09-11	09-19
2019	02-05	09-30 à 10-01	10-09
Fête	Ramadan	Aïd el-Fitr	Aïd al-Adha
Calendrier	islamique	islamique	islamique
Date	<i>ramadan</i>	1 <i>chawwal</i>	10–13 <i>dhou al-hijja</i>
2012	07-20 à 08-18	08-19	10-26 à 10-29
2013	07-09 à 08-07	08-08	10-15 à 10-18
2014	06-29 à 07-28	07-29	10-04 à 10-07
2015	06-18 à 07-17	07-18	09-24 à 09-27
2016	06-07 à 07-05	07-06	09-13 à 09-16
2017	05-27 à 06-25	06-26	09-02 à 09-05
2018	05-16 à 06-14	06-15	08-22 à 08-25
2019	05-06 à 06-03	06-04	08-11 à 08-14

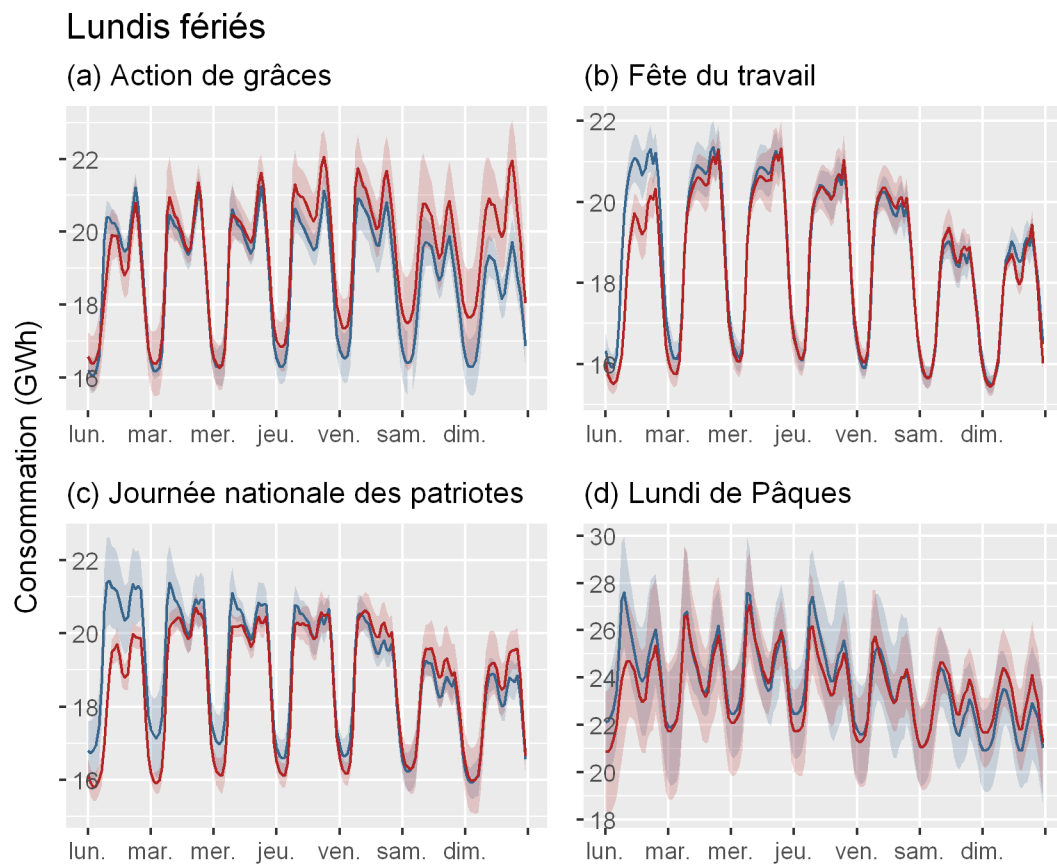


FIGURE 1.11 Consommation électrique au Québec la semaine d'un lundi férié, comparé à la semaine précédente

que la consommation moyenne lors de la semaine précédant ce lundi férié (en bleu). Le graphique (a) présente l'Action de grâces (le deuxième lundi d'octobre), le graphique (b) présente la fête du Travail (le premier lundi de septembre), le graphique (c) présente la Journée nationale des patriotes (le lundi qui précède le 25 mai), et le graphique (d) présente le lundi de Pâques (le lendemain du premier dimanche après la première pleine lune qui suit le 21 mars selon le calendrier grégorien). Le graphique (d) présente aussi la période allant du Vendredi saint jusqu'à Pâques, en bleu. Comme cette figure permet de le constater, la consommation électrique

au Québec est généralement plus basse lors des lundis fériés (en rouge) que lors des lundis ouvrables similaires (en bleu). On peut aussi observer un changement à la baisse de la consommation normale lors de certains dimanches (en bleu) précédant le lundi férié, notamment dans les graphiques (a) et (d), laissant supposer que ce férié est couramment fêté la veille.

1.3.4 Variables météorologiques

Les variables météorologiques à notre disposition sont la température (en degrés Celsius), la nébulosité ou couverture nuageuse (en octas ajustés) et la vitesse du vent (en kilomètres par heure). Pour chacune de ces données, nous avons cinq valeurs, soit la valeur enregistrée aux stations météorologiques de Baie-Comeau (**bc**), Ottawa (**ow**), Québec (**qc**), Sherbrooke (**sh**), et Montréal (**yu1**). Ces emplacements semblaient stratégiquement judicieux pour assurer une couverture géographique plus large avec une densité de population relativement élevée, à l'exception de Baie-Comeau. Baie-Comeau, dans la région administrative de la Côte-Nord, est principalement incluse dans l'espoir de capturer le comportement de consommation énergétique dans les régions nordiques du Québec.

L'ensemble de données possédait initialement des attributs additionnels : les précipitations (en millimètres) à chacune des stations météorologiques. Les précipitations en neige avaient été converties en leur équivalent en précipitations en pluie au préalable, à l'aide d'une formule qui nous est inconnue. Malheureusement, cette conversion aboutit à des données ayant une distribution étrange, où il est évident que les distributions sous-jacentes des deux types de précipitations sont différentes. Il est possible que cette différence ne soit que due à une conversion physique toute simple, comme l'équivalence d'un centimètre de neige à un millimètre d'eau, par exemple. Dans le doute, nous avons tout de même décidé d'ignorer ces variables.

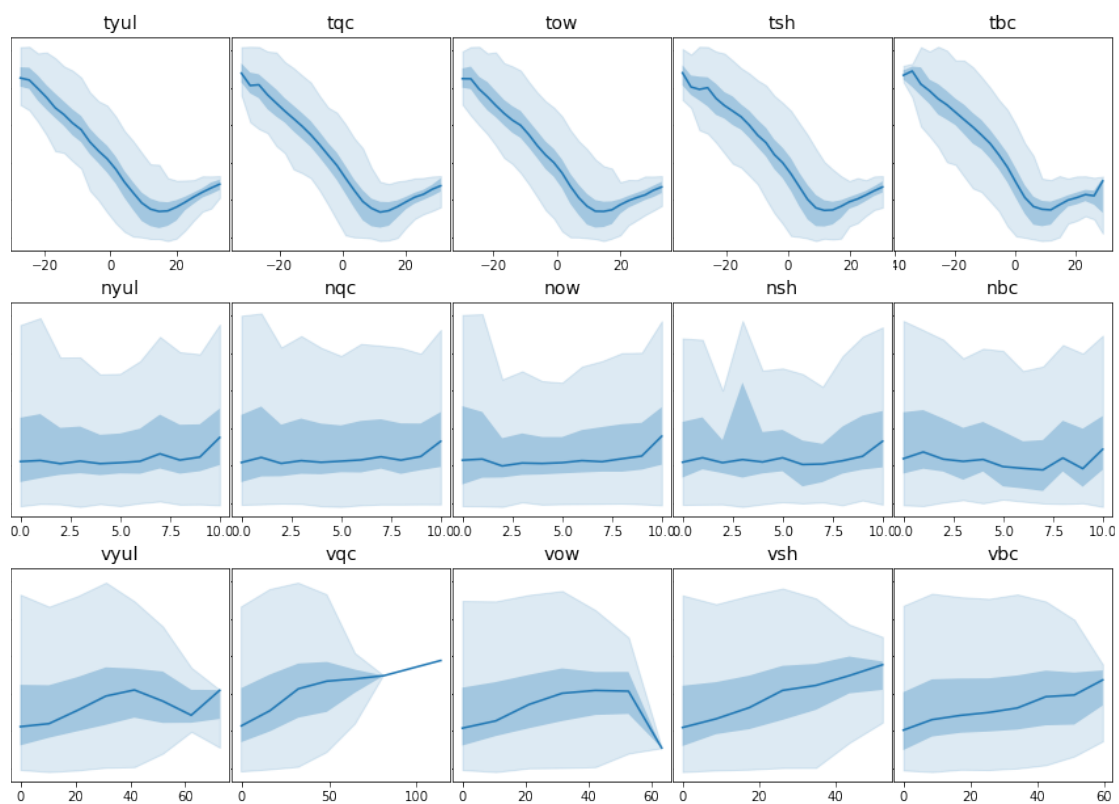


FIGURE 1.12 Corrélation des attributs météorologiques avec la consommation électrique au Québec

La figure 1.12 présente les effets des attributs météorologiques sur la consommation électrique au Québec pour chaque station météorologique présente dans notre ensemble de données. Les valeurs en ordonnée vont de 16,9 GWh à 37,5 GWh dans chaque graphique.

1.3.4.1 Température

Dans le cas particulier du Québec, le climat se caractérise par une large amplitude thermique : des températures extrêmes tant en hiver qu'en été, une forte variabilité des vents dans certaines parties du territoire et d'importantes masses d'air

continentales déferlant du nord-ouest.

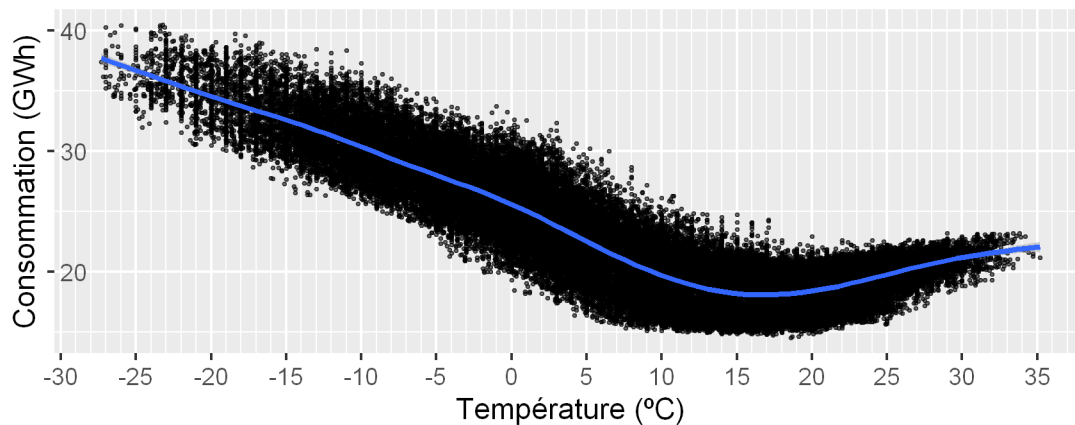


FIGURE 1.13 Corrélation entre la température et la consommation

La figure 1.13 montre la corrélation entre la température et la consommation électrique observée entre 2012 et 2019. La température est mesurée à la station de Montréal, qui couvre la plus grande proportion de la population ; et la ligne bleue trace la consommation moyenne à une température donnée.

En dessous de 15 °C, plus la température est basse, plus la charge électrique est élevée, ce qui s'explique par le fait que plus de 80 % des ménages du Québec utilisent l'électricité pour se chauffer. De même, au-dessus de 20 °C, nous observons que la charge électrique commence à augmenter, possiblement en raison de l'augmentation de l'utilisation des dispositifs de refroidissement. Par conséquent, il est important que les modèles de prévision soient en mesure de traiter un profil de charge sensible à la température et dépendant des saisons.

La figure 1.14 présente ces mêmes effets, mais séparés pour chaque heure de la journée. Les valeurs en bleu sont les points de données enregistrés lors de jours ouvrables, et les valeurs en rouge sont les points de données enregistrés les fins de semaine. On remarque que la différence entre ces deux types de journées est

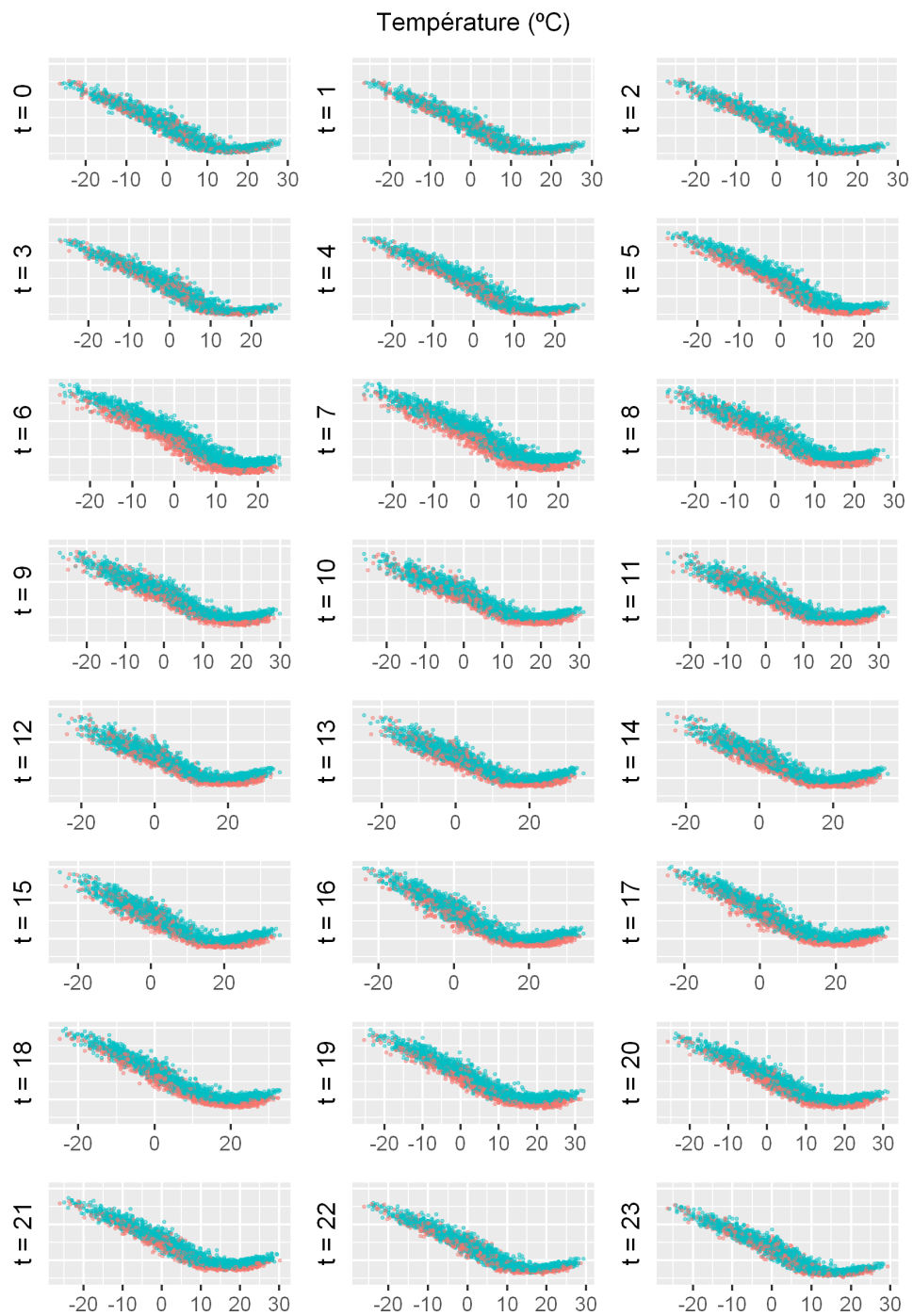


FIGURE 1.14 Effets de la température sur la consommation pour chaque pas de temps

plus marquée lors des heures de la journée que de la nuit. À certaines heures, la démarcation est si grande que les deux distributions sont faciles à distinguer, même sans l'aide des deux couleurs. De manière plus générale, ces graphiques démontrent qu'une conclusion tirée d'un comportement exhibé à un pas de temps précis (ou par les valeurs cumulées à tous les pas de temps) n'est pas nécessairement valide pour les autres pas de temps, examinés individuellement.

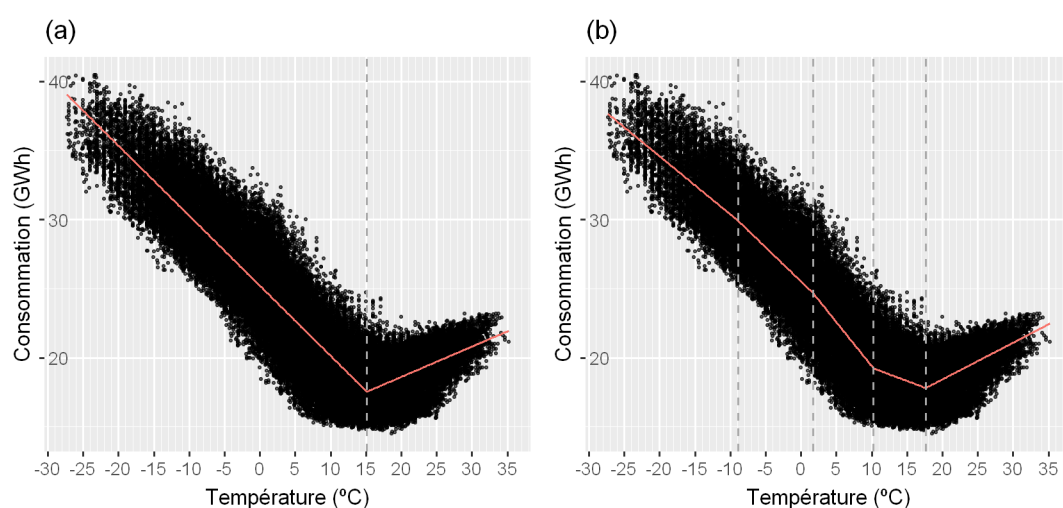


FIGURE 1.15 Segmentation des effets de la température avec (a) un point de rupture et (b) quatre points de rupture

Les graphiques (a) et (b) de la figure 1.15 présentent une segmentation linéaire par morceaux des effets de la température avec un et quatre points de ruptures, respectivement. La température est mesurée à la station de Montréal, et la segmentation est effectuée à l'aide de la librairie `segmented` en R (Muggeo, 2017) basée sur une régression linéaire entre la consommation et la température. L'unique point de rupture du graphique (a) est situé à 15,1 °C; les points de rupture en (b) sont -8,9 °C, 1,7 °C, 10,3 °C, et 17,7 °C. Ces points de ruptures peuvent être utiles pour l'élaboration d'une spécification linéaire par morceaux appropriée.

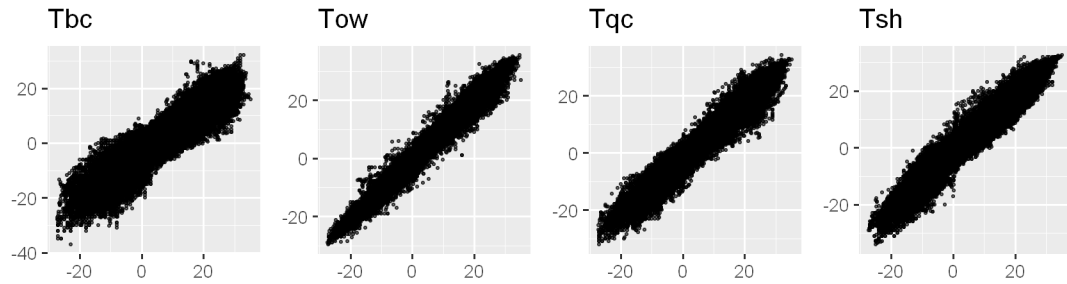


FIGURE 1.16 Corrélation de la température mesurée aux différentes stations météorologiques

La figure 1.16 présente les corrélations entre la température mesurée à Montréal (T_{yu1}) et la température mesurée à chacune des quatre autres stations météorologiques. La corrélation est de 93,5 % entre Montréal et Baie-Comeau, 98,8 % entre Montréal et Ottawa, 97,7 % entre Montréal et la ville de Québec, et 97,2 % entre Montréal et Sherbrooke. Ces corrélations élevées peuvent poser problème pour certains modèles de prévision de séries temporelles, et plusieurs auteurs (Clements *et al.*, 2016; Cancelo *et al.*, 2008) optent pour combiner ces valeurs à l'aide d'une moyenne pondérée par population plutôt que de les garder séparées.

Les graphiques (a) et (b) de la figure 1.17 présentent les effets qu'ont des différences de températures entre deux pas de temps sur la consommation à ces deux pas de temps. Le graphique (a) montre la différence entre la température à un pas de temps donnée et la température enregistrée exactement 24 heures plus tôt ($\Delta_{1j}T_{yu1,t} = T_{yu1,t} - T_{yu1,t-1j}$), en relation avec la différence de la demande électrique entre ces deux mêmes pas de temps ($\Delta_{1j}Y_t$). Le graphique (b) montre la relation entre la différence de température et la différence de demande électrique à une semaine d'intervalle ($\Delta_{7j}T_{yu1,t}$ contre $\Delta_{7j}Y_t$). L'un comme l'autre de ces deux graphiques semblent présenter un amalgame de deux distribution de points distinctes : l'une est une diagonale descendante, et l'autre est horizontale ou diagonale

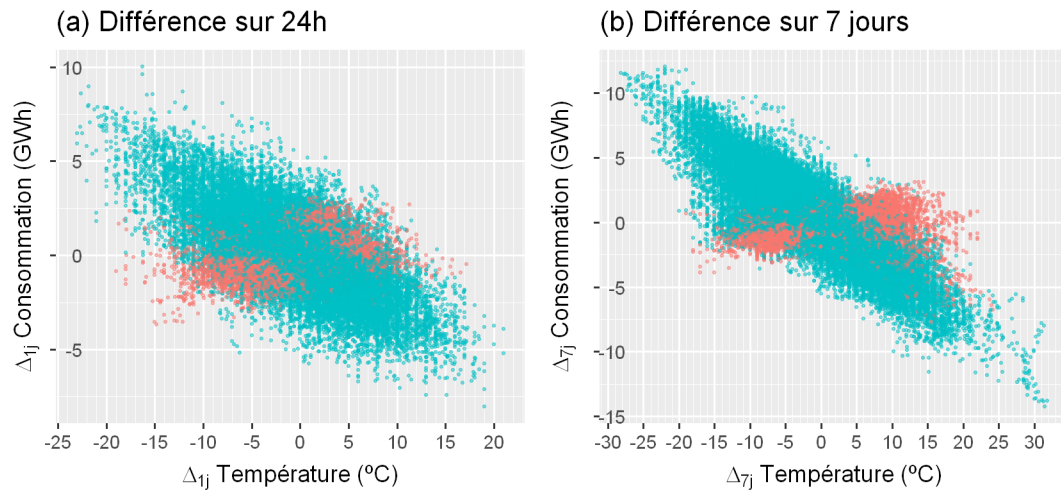


FIGURE 1.17 Effets de la différence de température sur la consommation

ascendante. Heureusement, nous pouvons utiliser le point de rupture de $15,1\text{ }^{\circ}\text{C}$, trouvé plus tôt grâce à la segmentation linéaire par morceaux, pour distinguer ces deux distributions sous-jacentes. Les valeurs en bleu sont celles pour lesquelles la moyenne des températures aux deux pas de temps considérés est plus petite ou égale à $15,1\text{ }^{\circ}\text{C}$, et les valeurs en rouge sont celles pour lesquelles la moyenne est au dessus de $15,1\text{ }^{\circ}\text{C}$.

1.3.4.2 Nébulosité et vitesse du vent

Comme le montrent les figures 1.18 et 1.19, la nébulosité et la vitesse du vent correspondent toutes deux avec la consommation électrique (Grenier, 2006).

Le graphique (a) de la figure 1.18 présente une segmentation linéaire par morceaux des effets de la nébulosité avec un point de rupture à $8,8$ octas. La nébulosité est mesurée à la station de Montréal. Cette augmentation plus marquée de la consommation électrique pour les nébulosités au-dessus du point de rupture

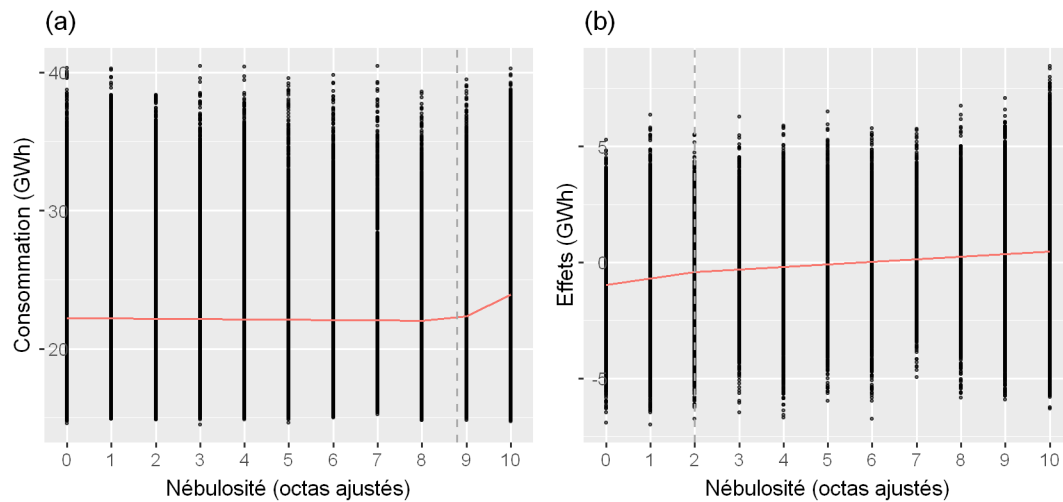


FIGURE 1.18 Segmentation des effets de la nébulosité

n'est pas surprenante. Il semble raisonnable de supposer que si le ciel est presque complètement couvert, alors nous pourrions observer une légère augmentation de la consommation électrique due à l'utilisation de dispositifs d'éclairage. Nous ne nous attendons pas à ce que cet effet soit aussi élevé que celui observé pour le chauffage et la climatisation, puisque l'éclairage n'est pas aussi énergivore, mais la différence devrait tout de même être significative. Le graphique (b) présente une segmentation linéaire par morceaux des effets de la nébulosité après avoir retiré les effets de la température sur la consommation. La valeur sur l'axe des ordonnées est donc $Y_t - \widehat{Y}_t$, où \widehat{Y}_t est calculée avec une régression linéaire ayant pour entrée la segmentation linéaire à quatre points de rupture de la température, vu précédemment. Le point de rupture de cette valeur dérivée de la nébulosité se situe à 2,0 octas.

Le graphique (a) de la figure 1.19 présente une segmentation linéaire par morceaux des effets de la vitesse avec deux points de rupture, à 11,7 km/h et 33,0 km/h. La vitesse du vent est mesurée à la station de Montréal. Nous remarquons une

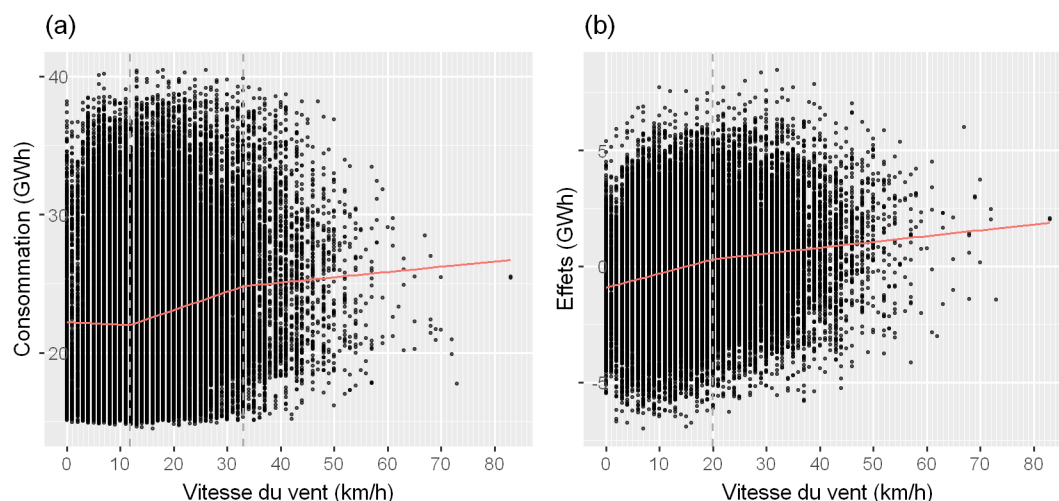


FIGURE 1.19 Segmentation des effets de la vitesse du vent

légère augmentation de la consommation électrique lorsque la vitesse augmente, ce qui pourrait être expliqué par le facteur de refroidissement éolien lorsque combiné avec des températures froides. Le graphique (b) présente une segmentation linéaire par morceaux des effets de la vitesse du vent après avoir retiré les effets de la température sur la consommation, calculé de manière similaire à la figure 1.18(b). Le point de rupture se situe à 20,0 km/h.

Les corrélations entre la nébulosité à Montréal et la nébulosité aux autres stations météorologiques vont de 25,1 % à 66,4 %, et les corrélations entre la vitesse du vent à Montréal et celles aux autres stations vont de 28,9 % à 62,9 %. Ces corrélations sont moins élevées que celle qu'on observe pour la température.

1.3.4.3 Angle du soleil à l'horizon

L'angle du soleil par rapport à l'horizon est une entrée déjà utilisée par Hydro-Québec pour la prévision de la demande électrique à court-terme (Grenier, 2006).

Elle fait partie de la saisonnalité annuelle et est entièrement prévisible, contrairement aux données météorologiques présentée ci-haut. Dans ce mémoire, elle est regroupée avec les autres données météorologiques, car nous avons choisi de la distinguer spécifiquement du reste de la saisonnalité annuelle et elle est donc traitée comme une entrée exogène.

L'angle du soleil à l'horizon étant pratiquement la même pour toutes les régions les plus peuplées du Québec, une seule valeur, mesurée à Montréal, est utilisée.

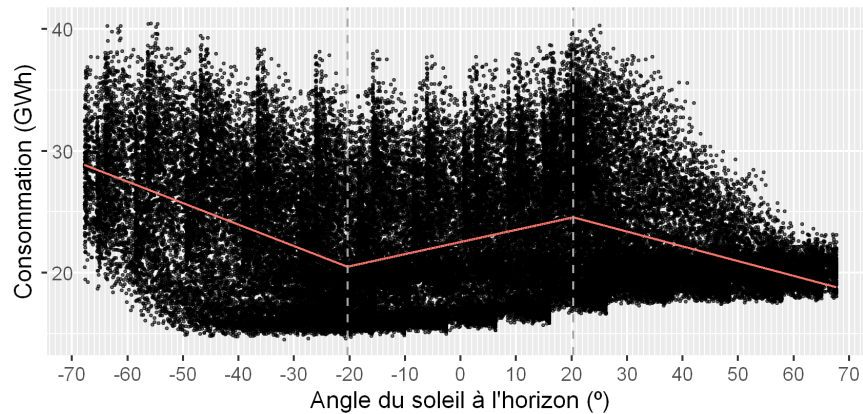


FIGURE 1.20 Segmentation des effets de l'angle du soleil

La figure 1.20 présente une segmentation linéaire par morceaux des effets de l'angle du soleil par rapport à l'horizon avec deux points de rupture, à $-20,4^\circ$ et $20,3^\circ$. L'angle est mesuré à Montréal. La segmentation des effets de l'angle du soleil sur la consommation après avoir retiré les effets de la température donne sensiblement le même profil, mais avec une plus faible variation de la consommation à chaque angle donné.

CHAPITRE II

ÉTAT DE L'ART AVEC EXEMPLES DE MODÈLES DANS LA LITTÉRATURE

Une méta-analyse de 240 articles publiés au sujet de la prévision de la consommation électrique à court terme entre 2000 et 2019 (Nassif *et al.*, 2022) a démontré que la classe de méthodes la plus populaire est de loin l'utilisation de **réseaux de neurones artificiels** avec 21 % des modèles proposés dans la littérature qui en font un usage exclusif, un nombre qui monte à 45 % lorsqu'on compte aussi les modèles hybrides qui en font usage. Ils sont suivis des modèles de **série temporelle**, incluant les modèles **autorégressifs à moyenne mobile** (ARIMA), comptant pour 10 % des modèles proposés ; des modèles de **régression**, pour 9 % des modèles proposés ; et des modèles de **logique floue** et de **régression à vecteurs de support** (SVR), pour autour de 4 % des modèles proposés chacun. D'autres modèles employés sont les optimisations par essaims particulières (*particle swarm optimization* ou PSO), les vecteurs autorégressifs bayésiens (*Bayesian vector autoregression* ou BVAR), les modèles de décomposition, les filtres de Kalman, les cartes autoadaptatives ou cartes de Kohonen (*self organizing maps* ou SOM), *Grey Prediction*, les algorithmes de colonies de fourmis (*ant colony optimization* ou ACO), et les algorithmes génétiques (GA). Dans l'ensemble, les techniques d'apprentissage automatique ont été utilisées dans environ 43 % des travaux, les techniques hybrides employant plusieurs modèles dans 44 %, et les

méthodes statistiques dans environ 13 %.

Les attributs des ensembles de données dans la littérature comprennent généralement l'historique de consommation électrique dans la région à prévoir ; des variables météorologiques, dont la température, l'humidité, les précipitations, la vitesse du vent, la couverture nuageuse, ou la luminosité ; le type de journée, c'est-à-dire s'il s'agit d'un jour ouvrable, d'une fin de semaine, d'une fête nationale, ou de tout autre événement qui pourrait affecter la consommation ; et des données sur la croissance démographique.

Étant donné le fait que l'on comprend généralement bien les facteurs qui influencent la demande électrique au jour le jour, la quantité de données généralement disponibles, et l'assurance de pouvoir supposer que son comportement dans un futur proche sera similaire à celui récemment observé, les modèles de prévision à court terme de la consommation électrique peuvent obtenir des résultats remarquablement précis. En ce sens, Nassif *et al.* (2022) suggèrent qu'un score MAPE compétitif avec les méthodes les plus récentes est un score qui ne dépasse pas 1 % (Nassif *et al.*, 2022). Un score MAPE entre 1 % et 2 % peut cependant encore être qualifié d'intéressant.

Les sections suivantes présentent certaines de ces techniques plus en détail.

2.1 ARIMA et ses variantes

Les modèles de type ARIMA sont largement utilisés dans la littérature pour la prévision de la charge. En statistique, les **modèles autorégressifs de moyenne mobile intégrés** (*autoregressive integrated moving average* ou *ARIMA*) ou modèles de Box-Jenkins et leurs modèles dérivés sont parmi les principaux modèles appliqués aux séries temporelles afin de prédire leurs valeurs futures. Les modèles ARIMA sont considérés comme une méthode stable, fiable et efficace pour pré-

voir la demande électrique, car ils permettent de se concentrer sur l'évolution du cycle intrajournalier (Taylor, 2008; Pappas *et al.*, 2010). Leur plus grande généralité par rapport aux modèles autorégressifs standards et leur capacité à gérer une plus grande variété d'effets dynamiques avec moins de paramètres peuvent être avantageuses (Cancelo *et al.*, 2008).

Le modèle de base, ARMA, est composé de deux parties : une partie autorégressive (AR) et une partie de moyenne mobile (MA pour *moving average*, en anglais). Le modèle est généralement identifié par $\text{ARMA}(p, q)$, où p et q sont respectivement l'ordre de chacune des deux parties mentionnées. Une fois les ordres p et q choisis, le modèle peut être entraîné sur les données d'une série temporelle par la méthode des moindres carrés ordinaires itérés (Spliid, 1983) pour trouver les paramètres qui minimisent la somme des carrés de ses résidus ou termes d'erreur.

Plusieurs généralisations d'ARIMA permettent d'étendre le modèle pour lui permettre d'intégrer des variables exogènes ou de traiter les séries temporelles possédant une composante de tendance ou de saisonnalité. Ces généralisations peuvent être combinées pour obtenir des modèles plus complexes et plus versatiles.

Plusieurs chercheurs et chercheuses ont montré qu'il est possible d'obtenir une bonne précision de prévision en utilisant ARIMA (Topalli *et al.*, 2006; Kim, 2013) ou SARIMAX (Tarsitano et Amerise, 2017; Elamin et Fukushige, 2018) comme approche autonome. Pour tenter d'améliorer la précision, d'autres ont incorporé des modèles de modification résiduelle (Wang *et al.*, 2012), conçu un modèle utilisant un schéma de *lifting en ondelettes* avec ARIMA (Lee et Ko, 2011), ou appliqué un modèle ARIMA de prévision univariée pour obtenir un lissage exponentiel (Matijas *et al.*, 2013). De même, le modèle ARIMA à lissage exponentiel doublement saisonnier a été utilisé pour permettre une commutation fluide entre les méthodes de prévision de la demande (Taylor, 2003).

La section qui suit présente le concept du polynôme de retard qui sera utile pour représenter les modèles. Les sections suivantes présentent tout d'abord les parties autorégressives (AR) et de moyenne mobile (MA), avant d'unifier le tout en un modèle ARMA. Les extensions ou généralisations suivantes du modèle ARMA sont finalement présentées : le modèle avec entrées exogènes (ARMAX), le modèle intégré (ARIMA), et le modèle saisonnier (SARIMA).

2.1.1 Le polynôme de retard

Le **polynôme de retard** est défini par l'équation (2.1) :

$$\delta^n(\mathbb{L}^s) = \sum_{i=0}^n \delta_i \mathbb{L}^{si} \quad (2.1)$$

où $\delta_1, \dots, \delta_n$ sont des paramètres donnés ; et \mathbb{L} est l'opérateur de retard. Si le paramètre δ^n apparaît muni d'un indice, cet indice ne sert qu'à le distinguer de paramètres similaires et n'a aucune valeur mathématique.

Dans la littérature, il y a communément quatre cas spéciaux dans le contexte des modèles ARIMA ; lorsque les paramètres sont représentés par les lettres grecques ϕ et θ (et leur équivalent majuscule, Φ et Θ), et appliqués sur les séries Y_t (ou ses transformations) ou ε_t :

$$\phi^n(\mathbb{L}^s)Y_t = \left(1 - \sum_{i=1}^n \phi_i \mathbb{L}^{si}\right) Y_t; \quad \theta^n(\mathbb{L}^s)\varepsilon_t = \left(1 + \sum_{i=1}^n \theta_i \mathbb{L}^{si}\right) \varepsilon_t$$

2.1.2 ARMA(p, q)

Un modèle ARMA(p, q) est la combinaison de deux modèles simples : une partie autorégressive AR(p) et une partie de moyenne mobile MA(q).

La partie autorégressive est un modèle **autorégressif** d'ordre p ou AR(p), qui

vérifie l'expression suivante :

$$Y_t = \varepsilon_t + \sum_{i=1}^p \phi_i Y_{t-i} + c \quad (2.2)$$

où ϕ_1, \dots, ϕ_p sont les paramètres à apprendre ; c est une constante ; et ε_t est un bruit blanc similaire aux termes d'erreurs de la partie MA, expliquées ci-bas.

Cette équation de modèle peut être réécrite de manière équivalente en fonction du polynôme de retard $\phi^p(\mathbb{L})$, comme suit :

$$Y_t = \varepsilon_t + (1 - \phi^p(\mathbb{L})) Y_t + c \quad \Longleftrightarrow \quad \varepsilon_t = \phi^p(\mathbb{L}) Y_t + c \quad (2.3)$$

La convergence du modèle AR(p) n'est garantie que si le processus est stationnaire.

La deuxième partie, la partie de moyenne mobile, est un modèle **moyenne mobile** d'ordre q ou MA(q), qui vérifie l'expression suivante :

$$Y_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (2.4)$$

où $\theta_1, \dots, \theta_q$ sont les paramètres appris ; et $\varepsilon_t, \dots, \varepsilon_{t-q}$ sont les termes d'erreur.

Cette équation de modèle peut être ré-écrite de manière équivalente en fonction du polynôme de retard $\theta^q(\mathbb{L})$, comme suit :

$$Y_t = \varepsilon_t + (\theta^q(\mathbb{L}) - 1) \varepsilon_t \quad \Longleftrightarrow \quad Y_t = \theta^q(\mathbb{L}) \varepsilon_t \quad (2.5)$$

Idéalement, les termes d'erreur ε_t seraient indépendants et identiquement distribués (i.i.d.) selon une loi normale à moyenne nulle : $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, où σ^2 est la variance. Cette hypothèse peut être assouplie, mais cela affecte certaines propriétés du modèle.

La combinaison de ces deux modèles crée un modèle **autorégressif de moyenne mobile** d'ordres (p, q) ou ARMA(p, q), qui vérifie l'expression suivante :

$$Y_t = \varepsilon_t + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + c \quad (2.6)$$

où ϕ_1, \dots, ϕ_p et $\theta_1, \dots, \theta_q$ sont les paramètres à apprendre ; et $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ sont les termes d'erreur.

Cette équation de modèle peut être réécrite de manière équivalente en fonction des polynômes de retard $\phi^p(\mathbb{L})$ et $\theta^q(\mathbb{L})$, comme suit :

$$\phi^p(\mathbb{L}) Y_t = \theta^q(\mathbb{L}) \varepsilon_t + c \quad (2.7)$$

2.1.3 Généralisations du modèle

Un modèle **ARMA avec entrées exogènes** d'ordres (p, q, b) ou $\text{ARMAX}(p, q, b)$ est une généralisation du modèle ARMA permettant d'ajouter des entrées supplémentaire au modèle. Ce modèle contient les modèles $\text{AR}(p)$ et $\text{MA}(q)$ et y ajoute une combinaison linéaire des b derniers termes d'une série temporelle connue X_t . Il vérifie l'expression suivante :

$$Y_t = \varepsilon_t + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^b \beta_i X_{t-i} + c \quad (2.8)$$

où ϕ_1, \dots, ϕ_p , $\theta_1, \dots, \theta_q$ et β_1, \dots, β_b sont les paramètres appris ; $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ sont les termes d'erreur ; et X_{t-1}, \dots, X_{t-b} sont les entrées exogènes.

Cette équation de modèle peut être ré-écrite de manière équivalente en fonction des polynômes de retard $\phi^p(\mathbb{L})$, $\theta^q(\mathbb{L})$ et $\beta^b(\mathbb{L})$, comme suit :

$$\phi^p(\mathbb{L}) Y_t = \theta^q(\mathbb{L}) \varepsilon_t + \beta^b(\mathbb{L}) X_t + c \quad (2.9)$$

Un modèle **ARMA intégré** d'ordres (p, q, d) ou $\text{ARIMA}(p, q, d)$ est une autre généralisation du modèle ARMA, permettant de modéliser la série temporelle Y_t même si cette dernière n'est pas stationnaire si cette propriété de non-stationnarité est due à une tendance non constante.

Cette généralisation est simple : il suffit d'abord rendre la série temporelle Y_t stationnaire en éliminant la tendance, puis de lui appliquer le modèle ARMA(p, q) comme expliqué plus haut. Le processus utilisé pour rendre la série temporelle Y_t stationnaire est la différentiation.

L'équation du modèle ARIMA(p, q, d) peut donc être ré-écrite de manière équivalente en fonction des polynômes de retard $\phi^p(\mathbb{L})$ et $\theta^q(\mathbb{L})$, comme suit :

$$\Delta^d \phi^p(\mathbb{L}) Y_t = \theta^q(\mathbb{L}) \varepsilon_t + c \quad (2.10)$$

Un modèle **ARIMA saisonnier** d'ordres $(p, q, d) \times (P, Q, D)_s$, parfois identifié par la notation ARIMA(p, q, d)(P, Q, D) $_s$, est la dernière généralisation du modèle ARMA pertinente pour ce mémoire. Cette variante du modèle permet de prendre en charge la composante saisonnière de la série temporelle Y_t , c'est à dire la composante cyclique répétitive.

Le modèle ARIMA(p, q, d)(P, Q, D) $_s$ vérifie donc l'expression :

$$\Delta^d \Delta_s^D \phi^p(\mathbb{L}) \Phi^P(\mathbb{L}^s) Y_t = \theta^q(\mathbb{L}) \Theta^Q(\mathbb{L}^s) \varepsilon_t + c, \quad (2.11)$$

où \mathbb{L} est l'opérateur de retard ; $\phi^p(\mathbb{L})$, $\Phi^P(\mathbb{L}^s)$ et $\theta^q(\mathbb{L})$, $\Theta^Q(\mathbb{L}^s)$ dénotent respectivement les parties autorégressive et de moyenne mobile, avec les polynômes de retard de degrés p , P , et q , Q respectivement ; d et D sont les ordres de différentiation régulier et saisonnier, respectivement ; et s est le facteur saisonnier. Le paramètre c est le terme constant.

2.1.4 Implémentation de Cancelo *et al.* (2008)

L'article de Cancelo *et al.* (2008) décrit l'expérience de prévision à *Red Eléctrica de España* (REE), basée sur différents modèles développés par les auteurs depuis 1988. En plus de vouloir produire une prévision précise et fiable de la consommation électrique horaire et journalière jusqu'à dix jours d'avance en Espagne

continentale, leur modèle cherche à capter les connaissances du personnel qualifié de l'organisation, et ses procédés doivent donc être interprétables.

La structure du modèle global est composée d'un modèle journalier et de 24 modèles horaires. Ces modèles sont tous de type ARIMA doublement saisonnier avec plusieurs entrées exogènes, mais leurs paramètres spécifiques diffèrent en fonction de plusieurs facteurs, telles la saison ou l'heure de la journée. Le modèle proposé est entraîné sur un ensemble de données historiques sur la consommation électrique en Espagne entre 1993 et 2006.

Sur les données de l'année 2006, ce modèle obtient un score MAPE d'environ 1,67 % pour les prévisions journalières jusqu'à sept jours à l'avance, et un score MAPE de 1,56 % pour les prévisions horaires jusqu'à 24 heures à l'avance.

Voici certaines notations utilisées au cours de cette section :

- $L_t = \log(Y_t)$ est le logarithme de Y_t , où Y_t est la valeur de série temporelle de la consommation (en MWh) au pas de temps t .
- \mathbb{L} est l'opérateur de retard.

2.1.4.1 Structure générale

Le modèle global est composé d'un modèle pour les prévisions journalières et de 24 modèles horaires. Pour prédire des horizons jusqu'à trois jours dans le futur, les modèles horaires sont utilisés directement. Entre trois et dix jours dans le futur, la prévision journalière est produite en un premier temps, puis les 24 modèles horaires sont utilisés pour partitionner cette prévision pour chaque heure.

Techniquement, le modèle dans son ensemble est donc d'un modèle à équations multiples plutôt qu'une application directe d'un modèle ARMA. Cependant, son modèle pour les prévisions journalières, sans l'ajout des 24 modèles horaires, est

un modèle ARMA.

En règle générale, il a été constaté que les modèles horaires ne sont pas aussi complexes que le modèle journalier, car de nombreux effets météorologiques ne sont pas significatifs à certaines heures particulières de la journée. Tous les modèles ont la même structure de base, liée à la décomposition traditionnelle de la charge observée en quatre composantes : la charge normale, la partie sensible aux conditions météorologiques, les événements spéciaux et une composante aléatoire.

$$L_t = \mathbb{T}_t + \mathbb{S}_t + \mathbb{J}_t + \mathbb{M}_t + U_t \quad (2.12)$$

où $\mathbb{T}_t + \mathbb{S}_t$ est la consommation normale, avec \mathbb{T}_t dénotant la tendance et \mathbb{S}_t une partie de la saisonnalité ; \mathbb{J}_t est la contribution des jours spéciaux ; \mathbb{M}_t est la contribution des variables météorologiques ; et U_t est une perturbation stationnaire.

2.1.4.2 Consommation de base

On définit le logarithme de la consommation de base ℓ_t tel qu'à l'équation (2.13) :

$$\ell_t = L_t - \mathbb{J}_t - \mathbb{M}_t = \mathbb{T}_t + \mathbb{S}_t + U_t \quad (2.13)$$

Cela représente la charge qui serait observée s'il n'y avait pas d'autres effets des jours spéciaux \mathbb{J}_t ou des variables météorologiques \mathbb{M}_t . Étant donné que ces effets supplémentaires sont les perturbations majeures qui modifient la charge observée, ℓ_t est plutôt lisse et l'on peut supposer que son approximation suit un modèle ARIMA(1, 1, 1)(1, 1, 1) $_{7j}$ (1, 1, 0) $_{365j}$:

$$\begin{aligned} \Delta_{1j} \Delta_{7j} \phi_1^1(\mathbb{L}^{1j}) \Phi_2^1(\mathbb{L}^{7j}) \Phi_3^1(\mathbb{L}^{365j}) \ell_t &= \theta_1^1(\mathbb{L}) \Theta_2^1(\mathbb{L}^7) \Theta_3^1(\mathbb{L}^{365}) \varepsilon_t \\ \iff \Delta_{1j} \Delta_{7j} \phi_1 \mathbb{L}^{1j} \Phi_2 \mathbb{L}^{7j} \Phi_3 \mathbb{L}^{365j} \ell_t &= \theta_1 \mathbb{L}^{1j} \Theta_2 \mathbb{L}^{7j} \Theta_3 \mathbb{L}^{365j} \varepsilon_t \end{aligned} \quad (2.14)$$

où $\phi_1^1(\mathbb{L}^{1j})$, $\Phi_2^1(\mathbb{L}^{7j})$, $\Phi_3^1(\mathbb{L}^{365j})$, $\theta_1^1(\mathbb{L}^{1j})$, $\Theta_2^1(\mathbb{L}^{7j})$ et $\Theta_3^1(\mathbb{L}^{365j})$ sont des polynômes de retard de degré 1 ; et $\Delta \Delta_7 = (1 - \mathbb{L})(1 - \mathbb{L}^7)$. Dans cette équation, la tendance

linéaire locale \mathbb{T}_t est modélisée par la partie ARIMA(1, 1, 1) du modèle et la saisonnalité \mathbb{S}_t est modélisée par la partie ARIMA(1, 1, 1) $_{7j}$ (1, 1, 0) $_{365j}$. La saisonnalité a donc un facteur saisonnier hebdomadaire ainsi qu'un facteur saisonnier annuel.

La différentiation hebdomadaire Δ_{7j} est utilisée pour modéliser le motif hebdomadaire prononcé. La différentiation annuelle $\Delta_{365j} = (1 - \mathbb{L}^{365j})$ n'est cependant pas utilisée. En pratique, il est difficile de séparer la saisonnalité annuelle de l'effet des variables météorologiques. Dans le modèle journalier, \mathbb{S}_t est donc essentiellement l'équivalent de la saisonnalité hebdomadaire. La majorité de la saisonnalité annuelle peut être expliquée par des variables météorologiques et des variables indicatrices pour tenir compte des périodes de vacances.

2.1.4.3 Entrées exogènes

Les effets des jours spéciaux et des variables météorologiques sont exprimés par les termes $\mathbb{J}_t + \mathbb{M}_t$. Leurs effets sur la charge sont dynamiques et peuvent dépendre du jour de la semaine et de la période de l'année.

Par souci de simplicité, Cancelo *et al.* (2008) n'entrent pas dans les détails précis de leur implémentation des jours spéciaux. Ils notent qu'il y a eu 29 jours fériés nationaux ou régionaux en 2006 en Espagne, chacun ayant des effets très différents sur la consommation d'électricité. Les jours fériés locaux dans certaines grandes villes, notamment Madrid et Barcelone, ont des effets perceptibles sur la charge totale et doivent donc être pris en compte. Les jours spéciaux incluent aussi les longues fins de semaine et les périodes de vacances de Pâques, de Noël et en août. Certaines anomalies, comme les journées d'élections et de grèves, peuvent aussi jouer un rôle (Cancelo et Espasa, 1996).

Pour cet exemple d'implémentation, nous considérons quatre classes de jours spéciaux : ceux que Cancelo *et al.* (2008) décrivent comme étant les jours fériés

communs \mathcal{H}_0 , comprenant certains jours fériés nationaux ainsi que tous les jours fériés régionaux et locaux ; la fête nationale de l'Épiphanie **Jan06** (le 6 janvier) ; la fête nationale du travail **May01** (le 1er mai) ; et la fête nationale de l'Assomption **Aug15** (le 15 août). Il a été constaté que chacune de ces fêtes nationales a une influence particulière et distincte sur la charge, qui est statistiquement différente de l'influence de tout autre jour spécial.

L'expression de l'influence des jours spéciaux sur la consommation est donc :

$$\begin{aligned} \mathbb{J}_t &= \sum_{j \in \mathcal{J}} \sigma_j^*(\mathbb{L}) J_{j,t} \\ &= \sigma_1 \mathbb{1}_{\{t \in \mathcal{H}_0\}} + \sigma_2 \mathbb{1}_{\{t \in \text{Jan06}\}} + \sigma_3 \mathbb{1}_{\{t \in \text{May01}\}} + \sigma_4 \mathbb{1}_{\{t \in \text{Aug15}\}} \end{aligned}$$

où $\mathcal{J} = \{\mathcal{H}_0, \text{Jan06}, \text{May01}, \text{Aug15}\}$ sont les classes de jours spéciaux présentées ci-haut ; $J_{j,t} = \mathbb{1}_{\{t \in j\}}$ est la série temporelle et entrée exogène de la classe de jours spéciaux k , où $\mathbb{1}_{\{t \in j\}}$ est la variable indicatrice qui est égale à 1 si $t \in j$ et égale à 0 sinon ; et $\sigma_j^*(\mathbb{L})$ sont des polynômes de retard de degrés pouvant varier pour exprimer l'effet des jours adjacents, mais qui sont tous de degré 1 dans cet exemple.

Pour cet exemple, seuls les effets de la température journalière maximum sont considérés. Cette température est calculée à l'aide d'une moyenne pondérée des températures maximales enregistrées entre 6 heures et 18 heures dans dix observatoires représentant différentes régions climatiques en Espagne. Le modèle de Cancelo *et al.* (2008) inclut aussi d'autres variables météorologiques, comme la nébulosité et la durée d'ensoleillement, mais les auteurs n'entrent pas dans les détails de leur implémentation.

La relation dynamique non linéaire entre la température et la consommation est approximée à l'aide d'un système d'équations par morceaux. La « zone froide » comprend la plage des températures en dessous de 20 °C. La « zone chaude »

comprend la plage des températures au-dessus de 24 °C. La « zone de confort » se trouve entre ces deux températures et correspond à la plage où les températures n'ont pas d'effet sur la consommation électrique.

L'expression générale des effets de la zone froide est donnée par l'équation (2.15) :

$$\tau_1 h_{9,9}(T_t^{\max}) + \tau_2 h_{9,11}(T_t^{\max}) + \tau_3 h_{9,14}(T_t^{\max}) + \tau_4^5(\mathbb{L}^{1j}) h_{9,20}(T_t^{\max}) \quad (2.15)$$

où $\tau_4^5(\mathbb{L}^{1j})$ est un polynôme de retard de degré 5 ; les variables $h_{9,9}$, $h_{9,11}$, $h_{9,14}$ et $h_{9,20}$ sont des variables de réchauffement (telles que décrite à la section 1.2.3.3) ; et τ_1 , τ_2 , τ_3 et $\tau_{4,0}$ à $\tau_{4,5}$ sont les paramètres d'apprentissage.

L'expression générale des effets de la zone chaude est donnée par l'équation (2.16) :

$$\tau_5^2(\mathbb{L}^{1j}) c_{24,33}(T_t^{\max}) \quad (2.16)$$

où $c_{24,33}$ est une variable de refroidissement ; et $\tau_{5,0}$, $\tau_{5,1}$ et $\tau_{5,2}$ sont les paramètres d'apprentissage.

L'expression globale de l'influence de la température sur la consommation est donc donnée par l'équation (2.17) :

$$\begin{aligned} \mathbb{M}_t &= \sum_{X \in \mathcal{M}} \tau_X^*(\mathbb{L}) \mathbb{M}_X(X_t) \\ &= \tau_1 h_{9,9}(T_t^{\max}) + \tau_2 h_{9,11}(T_t^{\max}) + \tau_3 h_{9,14}(T_t^{\max}) + \tau_4^5(\mathbb{L}^{1j}) h_{9,20}(T_t^{\max}) \\ &\quad + \tau_5^2(\mathbb{L}^{1j}) c_{24,33}(T_t^{\max}) \end{aligned} \quad (2.17)$$

où $\mathcal{M} = \{T^{\max}\}$ est l'ensemble des variables météorologiques observées ; \mathbb{M}_X , pour $X \in \mathcal{M}$, est le système d'équations modélisant la série temporelle et entrée exogène X_t , une variable météorologique observée ; et $\tau_X^*(\mathbb{L})$ sont des polynômes de retard de degrés variés, reflétant le fait que les individus s'adaptent aux conditions météorologiques changeantes avec un certain retard. Ces effets varient en fonction de la période de l'année et, lors de certaines saisons, du jour de la semaine.

2.1.4.4 Modèle résultant

L'équation (2.18) présente la forme complète du modèle de Cancelo *et al.* (2008) :

$$\begin{aligned}
\widehat{L}_t &= \frac{\theta_1^1(\mathbb{L}^{1j}) \Theta_2^1(\mathbb{L}^{7j}) \Theta_3^1(\mathbb{L}^{365j})}{\Delta_{1j} \Delta_{7j} \phi_1^1(\mathbb{L}^{1j}) \Phi_2^1(\mathbb{L}^{7j}) \Phi_3^1(\mathbb{L}^{365j})} \varepsilon_t + \mathbb{J}_t + \mathbb{M}_t \\
&= \frac{\theta_1 \mathbb{L}^{1j} \Theta_2 \mathbb{L}^{7j} \Theta_3 \mathbb{L}^{365j}}{\Delta_{1j} \Delta_{7j} \phi_1 \mathbb{L}^{1j} \Phi_2 \mathbb{L}^{7j} \Phi_3 \mathbb{L}^{365j}} \varepsilon_t + \sum_{k \in \mathcal{K}} \sigma_j^*(\mathbb{L}) J_{j,t} + \sum_{X \in \mathcal{M}} \tau_X^*(\mathbb{L}) \mathbb{M}_X(X_t) \\
&= \frac{\theta_1 \mathbb{L}^{1j} \Theta_2 \mathbb{L}^{7j} \Theta_3 \mathbb{L}^{365j}}{\Delta_{1j} \Delta_{7j} \phi_1 \mathbb{L}^{1j} \Phi_2 \mathbb{L}^{7j} \Phi_3 \mathbb{L}^{365j}} \varepsilon_t \\
&\quad + [\sigma_1 \mathbb{1}_{\{t \in \mathcal{H}_0\}} + \sigma_2 \mathbb{1}_{\{t \in \text{Jan06}\}} + \sigma_3 \mathbb{1}_{\{t \in \text{May01}\}} + \sigma_4 \mathbb{1}_{\{t \in \text{Aug15}\}}] \\
&\quad + [\tau_1 h_{9,9}(T_t^{\max}) + \tau_2 h_{9,11}(T_t^{\max}) + \tau_3 h_{9,14}(T_t^{\max}) + \tau_4^5(\mathbb{L}^{1j}) h_{9,20}(T_t^{\max})] \\
&\quad + \tau_5^2(\mathbb{L}^{1j}) c_{24,33}(T_t^{\max})
\end{aligned} \tag{2.18}$$

2.2 Équations multiples sur séries temporelles

Un modèle à **équations multiples sur séries temporelles** est une approche où chaque période ou pas de temps de la journée est traité comme un problème de prévision distinct, avec sa propre équation (Nassif *et al.*, 2022). Cette approche a le potentiel d'atteindre une précision de prévision très compétitive (Espinoza *et al.*, 2005; Soares et Medeiros, 2008).

Initié par Ramanathan *et al.* (1997), des améliorations significatives ont été proposées, notamment par Cottet et Smith (2003), qui proposent d'y intégrer une approche bayésienne dans une étude de cas du marché régional de la Nouvelle-Galles du Sud ; par Cancelo *et al.* (2008), qui exploite le dynamisme des modèles ARIMA doublement saisonniers pour le marché espagnol de l'électricité ; et plus récemment par Clements *et al.* (2016), qui améliore la manière de modéliser l'interaction entre les motifs de saisonnalité hebdomadaire et annuel et reconnaît

l'importance de la corrélation intrajournalière pour prévoir la consommation australienne. Les avantages de l'approche à équations multiples sont que les facteurs explicatifs qui déterminent les performances des prévisions sont visibles, testables et interprétables et le fait que la spécification des paramètres du modèle est linéaire, ce qui signifie que la méthode des moindres carrés ordinaires peut être utilisée pour estimer les paramètres, plutôt qu'un algorithme d'optimisation numérique (Clements *et al.*, 2016).

L'approche introduite par Ramanathan *et al.* (1997) combine 24 modèles de régressions linéaires – un pour chaque heure de la journée –, avec une structure d'erreur dynamique et des ajustements adaptatifs sur les erreurs. Les fins de semaine sont modélisées séparément. C'est lors des compétitions organisées par la *Puget Sound Power and Light Company* (maintenant renommée *Puget Sound Energy*) de 1990–91 et de 1991–92 que le modèle a fait ses preuves. La compétition invitait plusieurs équipes à soumettre leur modèle de prévision de la demande électrique à court terme pour les tester en temps réel sur les données du marché californien, entre le 1^{er} novembre et le 31 mars de l'année suivante. Avec un score MAPE global autour de 4,73 %, le modèle à équation multiple a remporté le meilleur score pour 95 des 168 pas de temps testés (c'est-à-dire 66 %) la première année. L'année suivante, leur modèle l'a emporté sur six des neuf catégories de pas de temps, avec un score MAPE global entre 3,19 % et 3,06 %.

Puisque le seul vrai critère pour qu'un modèle soit considéré comme faisant partie des approches à équations multiples est qu'il soit formé de la combinaison de modèles qui se préoccupent tous d'une période différente de la journée, il peut y avoir une grande diversité dans la manière précise de modéliser les composantes. La section suivante en présente une.

2.2.1 Implémentation de Clements *et al.* (2016)

Le modèle de Clements *et al.* (2016) est utilisé comme référence pour notre modèle à équations linéaires multiples au chapitre 3. Ce modèle de séries temporelles à équations multiples s'apparente à un modèle de type ARIMA doublement saisonnier avec multiples entrées exogènes, sans l'être tout à fait. Les auteurs se penchent sur le problème de la prévision de la consommation électrique à court terme pour la région du Queensland en Australie sur un ensemble de données couvrant une période de 11 ans, entre 2002 et 2013. Ce modèle est entraîné par l'application répétée de la méthode des moindres carrés ordinaires. Il se démarque par sa simplicité et l'utilisation efficace des informations décalées en permettant l'interaction entre les motifs saisonniers et les dépendances intrajournalières.

Au final, Clements *et al.* (2016) rapportent un score MAPE de 1,36 % pour les prévisions bi-horaires sur un horizon de 24 heures. Il a été sélectionné comme modèle de référence pour ses méthodes génériques, permettant d'être appliquées à d'autres problèmes de prévisions de charge moyennant certaines adaptations, ainsi que pour son potentiel à égaler ou même de surpasser des modèles de prévision non linéaires et non paramétriques plus complexes.

Voici certaines notations utilisées au cours de cette section :

- $L_t = \log(Y_t)$ est le logarithme de Y_t , où Y_t est la valeur de série temporelle de la consommation (en MW) au pas de temps t .
- j est le nombre de pas de temps par jour. Dans le cas de Clements *et al.* (2016), $j = 48$.
- $[t]$ représente la classe de congruence de t modulo j . C'est-à-dire que $t \in [t]$.
- \mathbb{L} est l'opérateur de retard.

2.2.1.1 Structure de base

Le modèle global est formé d'une coopération de 48 modèles, un pour chaque demi-heure de la journée. La structure de base de chacun de ceux-ci est celle d'un modèle formé de la combinaison de deux modèles $\text{ARMA}(1, 1)_{1j} + \text{ARMA}(1, 1)_{7j}$ sur la série temporelle L_t . Cette combinaison aboutit à l'équation (2.19) :

$$\widehat{L}_t = c^{[t]} + \left(\Phi_1^{[t]} \mathbb{L}^{1j} + \Phi_2^{[t]} \mathbb{L}^{7j} \right) L_t + \left(1 + \theta_1^{[t]} \mathbb{L}^{1j} + \theta_2^{[t]} \mathbb{L}^{7j} \right) \varepsilon_t \quad (2.19)$$

où L_t le logarithme de la consommation au pas de temps t ; ε_t (sans opérateur de retard) est le bruit blanc ou l'erreur attendue du modèle de prévision, alors que $\mathbb{L}^{1j} \varepsilon_t$ et $\mathbb{L}^{7j} \varepsilon_t$ sont des termes d'erreur; et $c^{[t]}$, $\Phi_1^{[t]}$, $\Phi_2^{[t]}$, $\theta_1^{[t]}$ et $\theta_2^{[t]}$ sont les paramètres d'apprentissage.

Il est à noter que puisqu'aucun décalage n'est plus court que $1j$, c'est-à-dire une journée complète, il en découle que chaque classe de congruence $[t]$ de t (modulo $1j$) a son propre modèle indépendant correspondant à chaque pas de temps dans une journée.

2.2.1.2 Saisonnalités hebdomadaire et annuelle

Afin de prendre en compte la saisonnalité due au jour de la semaine, des variables indicatrices sont introduites. Ces variables sont multipliées avec la valeur de la charge à la journée précédente, L_{t-1j} , afin d'en ajuster l'effet. La saisonnalité hebdomadaire $\mathbb{S}_t^{\text{sem}}$ prend donc la place du paramètre $\Phi_1^{[t]}$ dans l'équation (2.19), et sa valeur est donnée par l'équation (2.20) :

$$\Phi_1^{[t]} = \mathbb{S}_t^{\text{sem}} = \sum_{w \in \mathcal{W}} \mathbf{1}_{\{t \in w\}} \phi_{1,w}^{[t]} \quad (2.20)$$

où $\mathcal{W} = \{\text{lundis, mardis, mercredis, jeudis, vendredis, samedis, dimanches}\}$ est l'ensemble des pas de temps, partitionnés selon le jour de la semaine auquel ils appartiennent ; $\mathbb{1}_{\{t \in w\}}$ est la variable indicatrice qui est égale à 1 si $t \in w$ et égale à 0 sinon ; et $\phi_{1,w}^{[t]}$, pour $w \in \mathcal{W}$, sont les paramètres d'apprentissage

Similairement, la saisonnalité annuelle est spécifiée de manière à entrer dans le modèle par le coefficient sur L_{t-7j} . La saisonnalité annuelle $\mathbb{S}_t^{\text{année}}$ prend donc la place du paramètre $\Phi_2^{[t]}$ dans l'équation (2.19), et sa valeur est donnée par l'équation (2.21) :

$$\Phi_2^{[t]} = \mathbb{S}_t^{\text{année}} = \phi_2^{[t]} + \sum_{q=1}^4 \varphi_q^{[t]} \text{FOURIER}\left(\frac{2q\pi t}{364j}\right) \quad (2.21)$$

où $\sum_{q=1}^4 \varphi_q^{[t]} \text{FOURIER}\left(\frac{2q\pi t}{364j}\right)$ est la série de Fourier $\sum_{q=1}^4 \varphi_{q,1}^{[t]} \sin\left(\frac{2q\pi t}{364j}\right) + \varphi_{q,2}^{[t]} \cos\left(\frac{2q\pi t}{364j}\right)$; et $\phi_2^{[t]}$, $\varphi_{q,1}^{[t]}$ et $\varphi_{q,2}^{[t]}$, pour $q \in [1, 4]$, sont les paramètres d'apprentissage.

2.2.1.3 Corrélation intrajournalière

La principale contribution de Clements *et al.* (2016) par rapport à Cancelo *et al.* (2008) est la prise en charge de la corrélation intrajournalière. Cette corrélation est prise en compte grâce à l'introduction de deux nouvelles variables :

1. la dernière valeur de consommation observée, $L_{t-(t \bmod j)-1}$ (ou $\mathbb{L}^{(t \bmod j)+1} L_t$), et
2. la prévision au pas de temps précédent, \widehat{L}_{t-1} .

Outre les variables exogènes, l'information la plus importante qui est encore ignorée est la dernière valeur de consommation observée, $\mathbb{L}^{(t \bmod j)+1} L_t$. Cette valeur est particulièrement importante pour le premier pas de temps de l'horizon de prévision, car cette valeur de consommation est observée au pas de temps qui le précède immédiatement.

Quant à elle, la prévision au pas de temps précédent, \widehat{L}_{t-1} , sert à prendre l'inertie en compte de manière indirecte. En effet, puisque la prévision \widehat{L}_{t-1} base elle-même ses prévisions sur les valeurs de consommation et des entrées exogènes aux pas de temps $t-1$, $t-1j-1$, et $t-7j-1$, son inclusion permet à \widehat{L}_t de le faire aussi, quoiqu'indirectement. Bien sûr, puisque \widehat{L}_{t-1} utilise aussi la valeur de \widehat{L}_{t-2} pour faire sa prévision, son effet se répercute aussi sur la valeur de \widehat{L}_t , et ainsi de suite.

Les effets de la corrélation intrajournalière sont donc décrits par l'équation (2.22) :

$$\mathbb{S}_t^{\text{jour}} L_t = \left[\psi_1^{[t]} \mathbb{L}^{(t \bmod j)+1} + \mathbb{1}_{\{0 \notin [t]\}} \psi_2^{[t]} \mathbb{L}^1 \right] L_t \quad (2.22)$$

où $\mathbb{L}^{(t \bmod j)+1}$ décale une valeur de série temporelle au dernier pas de temps de la journée précédente. Lorsque cet opérateur est appliqué sur la consommation électrique, la valeur obtenue correspond à la dernière valeur de consommation observée; \mathbb{L}^1 décale une valeur de série temporelle au pas de temps précédent. Excepté le cas où $t \equiv 0$, lorsque cet opérateur est appliqué sur la consommation électrique, la valeur obtenue correspond à une valeur qui n'a pas encore été observée. Si la valeur observée n'est pas disponible, c'est l'estimation \widehat{L}_t de cette valeur qui est utilisée; $\psi_1^{[t]}$ et $\psi_2^{[t]}$ sont les paramètres d'apprentissage; et $\mathbb{1}_{\{[t]>0\}}$ est la variable indicatrice qui est égale à 1 si $[t] > 0$ et égale à 0 sinon. Elle est utilisée pour éviter la redondance entre $\mathbb{L}^{[t]+1}$ et \mathbb{L}^1 quand $t \equiv 0$.

2.2.1.4 Entrées exogènes

La seule variable météorologique considérée par Clements *et al.* (2016) est T_t , la température (en degrés Celsius) au pas de temps t . Pour modéliser ses effets sur la consommation électrique, Clements *et al.* (2016) utilisent une spécification linéaire par morceaux comprenant quatre variables (deux pour la zone froide et deux pour la zone chaude) et des effets d'épuisement de chaque côté de l'échelle des températures. La zone de confort est comprise entre 20 °C et 22 °C. Les

variables de réchauffement de la zone froide sont $h_1(T_t) = \text{CLRAMP}_{9,15}^-(T_t)$ et $h_2(T_t) = \text{CLRAMP}_{9,20}^-(T_t)$, et les variables de refroidissement de la zone chaude sont $c_1(T_t) = \text{CLRAMP}_{22,30}^+(T_t)$ et $c_2(T_t) = \text{CLRAMP}_{26,30}^+(T_t)$.

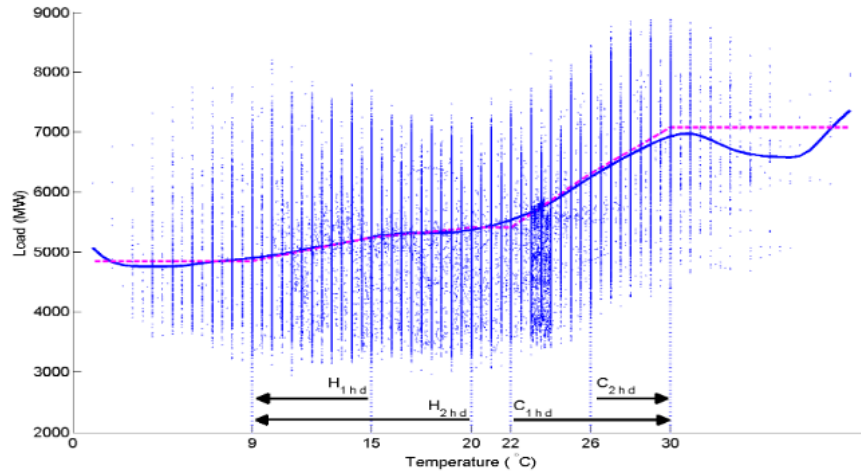


FIGURE 2.1 Consommation et températures au Queensland entre juillet 1999 et décembre 2013 (Clements *et al.*, 2016)

La figure 2.1 présente la charge et la température au Queensland de juillet 1999 à décembre 2013. La ligne continue indique un ajustement de régression non paramétrique avec un noyau normal et une bande passante de 1. La ligne pointillée est l'ajustement des moindres carrés ordinaires avec les quatre fonctions de modélisation linéaire de la température c_1 , c_2 , h_1 et h_2 . Les plages de température dans lesquelles ces fonctions sont définies sont indiquées par les flèches.

Le système d'équations modélisant les effets de la température, \mathbb{M}_T , est par conséquent décrit par l'équation (2.23) :

$$\mathbb{M}_T(T_t) = \tau_1^{[t]} h_1(T_t) + \tau_2^{[t]} h_2(T_t) + \tau_3^{[t]} c_1(T_t) + \tau_4^{[t]} c_2(T_t) \quad (2.23)$$

où $\tau_1^{[t]}$, $\tau_2^{[t]}$, $\tau_3^{[t]}$ et $\tau_4^{[t]}$ sont les paramètres d'apprentissage ; et h_1 , h_2 et c_1 , c_2 sont les variables de réchauffement et de refroidissement, respectivement.

De plus, le modèle de Clements *et al.* (2016) prend en compte l'effet des jours spéciaux dans son équation de prévision de la charge. En plus de l'effet des jours spéciaux en cours, $J_{j,t}$; l'effet des jours spéciaux décalés d'un jour, $\mathbb{L}^1 J_{j,t}$, est également pris en compte, comme dans Ramanathan *et al.* (1997). Cela permet d'ajuster l'effet de la charge décalée d'un jour, $\mathbb{L}^1 L_t$, de manière à prendre en compte le fait que cette journée précédente est un jour spécial et aura donc généralement une charge inférieure à celle d'un jour normal.

Pour éviter une explosion du nombre de paramètres à apprendre par le modèle, les jours fériés en question sont catégorisés en six groupes distincts : (1) les Vendredis saints, (2) les lundis de Pâques, (3) les jours de Noël, (4) les Nouveaux Ans, (5) des jours fériés uniques à Brisbane, la capitale du Queensland, et (6) tous les autres jours fériés d'une durée d'une seule journée.

L'effet des jours spéciaux est donc décrit par l'équation (2.24) :

$$\mathbb{J}(J_{j,t}) = \sum_{j \in \mathcal{J}} \left(\sigma_{1,j}^{[t]} + \sigma_{2,j}^{[t]} \mathbb{L}^1 \right) \mathbb{1}_{\{t \in j\}} \quad (2.24)$$

où \mathcal{J} est l'ensemble des six types de jours fériés décrits plus haut ; $\sigma_{1,j}^{[t]}$ et $\sigma_{2,j}^{[t]}$, pour $j \in \mathcal{J}$, sont les paramètres d'apprentissage ; et $J_{j,t} = \mathbb{1}_{\{t \in j\}}$ est la série temporelle et entrée exogène des jours spéciaux de la classe j , correspondant à la fonction indicatrice qui est égale à 1 si $t \in j$ et égale à 0 sinon. L'opérateur de retard \mathbb{L} peut s'appliquer sur cette série temporelle ;

2.2.1.5 Modèle résultant

La forme complète du modèle de Clements *et al.* (2016) est présentée par l'équation (2.25) :

$$\begin{aligned}
\widehat{L}_t &= c^{[t]} + \left(1 + \theta_1^{[t]} \mathbb{L}^{1j} + \theta_2^{[t]} \mathbb{L}^{7j}\right) \varepsilon_t \\
&\quad + \left(\mathbb{S}_t^{\text{sem}} \mathbb{L}^{1j} + \mathbb{S}_t^{\text{année}} \mathbb{L}^{7j}\right) L_t \\
&\quad + \mathbb{S}_t^{\text{jour}} L_t + \mathbb{M}_T(T_t) + \mathbb{J}(J_{j,t}) \\
&= c^{[t]} + \left(1 + \theta_1^{[t]} \mathbb{L}^{1j} + \theta_2^{[t]} \mathbb{L}^{7j}\right) \varepsilon_t \\
&\quad + \left(\sum_{w \in \mathcal{W}} \mathbb{1}_{\{t \in w\}} \phi_{1,w}^{[t]} \mathbb{L}^{1j} + \left[\phi_2^{[t]} + \sum_{q=1}^4 \varphi_q^{[t]} \text{FOURIER}\left(\frac{2q\pi t}{364j}\right)\right] \mathbb{L}^{7j}\right) L_t \quad (2.25) \\
&\quad + \left[\psi_1^{[t]} \mathbb{L}^{(t \bmod j)+1} + \mathbb{1}_{\{0 \notin [t]\}} \psi_2^{[t]} \mathbb{L}^1\right] L_t \\
&\quad + \left[\tau_1^{[t]} h_1(T_t) + \tau_2^{[t]} h_2(T_t) + \tau_3^{[t]} c_1(T_t) + \tau_4^{[t]} c_2(T_t)\right] \\
&\quad + \sum_{j \in \mathcal{J}} \left(\sigma_{1,j}^{[t]} + \sigma_{2,j}^{[t]} \mathbb{L}^{1j}\right) \mathbb{1}_{\{t \in j\}}
\end{aligned}$$

La figure 2.2 présente les scores MAPE demi-horaires et globaux de la prévision journalière entre juillet 2002 et décembre 2013, produits par le modèle proposé par Clements *et al.* (2016) sans température et jours spéciaux (ligne continue bleue), un modèle ARIMA doublement saisonnier $\text{ARIMA}(1, 1, 0)(1, 1, 1)_{1j}(1, 1, 1)_{7j}$ (ligne pointillée noire), et un modèle de lissage exponentiel de Holt-Winters (HWES) doublement saisonnier avec cycles intrajournaliers sans contrainte (ligne pointillée rouge). Les MAPE globaux sont représentés par des lignes horizontales avec les valeurs indiquées au-dessus.

2.3 N-BEATS, un modèle de réseau de neurones profond

Les **réseaux de neurones artificiels** sont très largement utilisés pour la prévision de la charge électrique. La quantité de techniques d'intelligence artificielle en

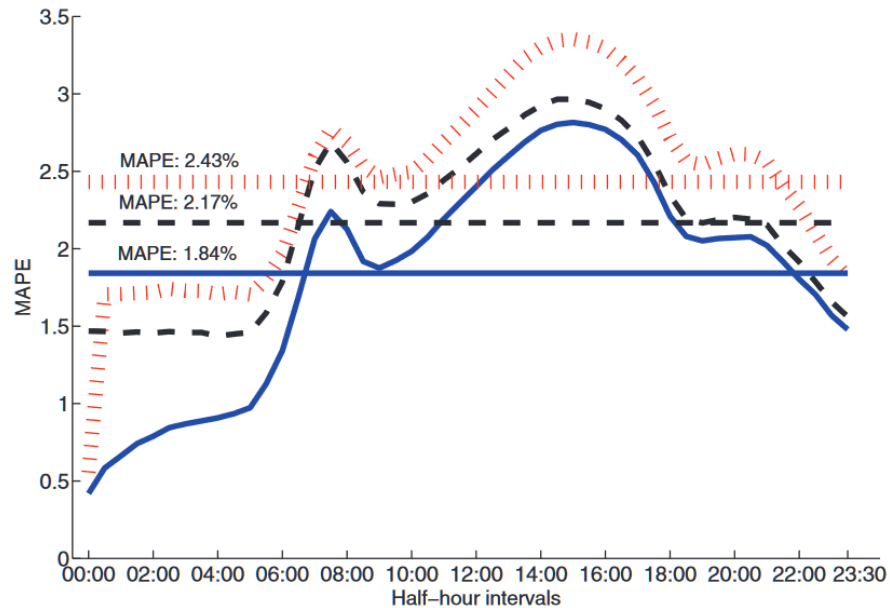


FIGURE 2.2 MAPE demi-horaires et globaux de la prévision journalière de juillet 2002 à décembre 2013 (Clements *et al.*, 2016)

cours de développement offrent un plus grand éventail de possibilités aux chercheurs et chercheuses du domaine, et les réseaux de neurones sont donc moins proéminents qu'ils l'ont déjà été (Nassif *et al.*, 2022). En général, les chercheurs ont rapporté que les réseaux de neurones obtiennent une haute précision et nécessitent moins d'échantillons et de points d'interpolation pour arriver à des résultats intéressants. Les modèles de neurones généralisés ont une flexibilité à la fois au niveau de la fonction d'agrégation et d'activation pour faire face aux problèmes de non-linéarité des intrants (Hernández *et al.*, 2013; Methaprayoon *et al.*, 2007), et certains peuvent même résoudre le problème de surapprentissage causé par le manque de données d'entraînement (Chan *et al.*, 2006). Il a aussi été démontré que la précision des réseaux de neurones s'améliore lorsqu'elle est combinée avec d'autres méthodes (Dolara *et al.*, 2015; Panapakidis, 2016).

Plusieurs travaux de recherche examinent l'application de l'**analyse d'ondelettes**

ou de la **transformée en ondelettes**, reconnues pour leur simplicité et leur combinaison facile avec d'autres modèles de prévision (Kelo et Dudul, 2012; Zhai, 2015; Wang *et al.*, 2014; Rafiei *et al.*, 2018). D'autres utilisent des **réseau de neurones en cascade** (Kouhi et Keynia, 2013), des **fonctions de base radiales** (*Radial Basis Functions*, ou RBF) (Cecati *et al.*, 2015), l'**apprentissage machine de réseau abductif** (Abdel-Aal, 2004), la **programmation d'expression génique** (*Gene Expression Programming*) avec réseaux de neurones (Sadat Hosseini et Gandomi, 2012), les **réseaux de neurones bayésiens** (BNN) (Lauret *et al.*, 2008; Hippert et Taylor, 2010), une combinaison de **réduction basée sur l'entropie** (*Entropy-Based Reduction*, ou EBR) et d'**algorithmes de Monte Carlo** (Beccali *et al.*, 2008), ou l'**apprentissage profond** (Bedi et Toshniwal, 2019; Yang *et al.*, 2019).

Le modèle N-BEATS (*Neural Basis Expansion Analysis for interpretable Time Series forecasting*) a été présenté dans l'article de Oreshkin *et al.* (2019) pour faire la prévision de séries temporelles. Ce modèle, entièrement basé sur des réseaux de neurones, a été testé sur les données de la prestigieuse compétition de prévision de séries temporelles M4 (Makridakis *et al.*, 2018), et a obtenu des prévisions en moyenne 3 % meilleures que le modèle gagnant de la compétition l'année précédente.

Une grande force de ce modèle par rapport à la plupart des autres modèles de réseaux de neurones est la possibilité de le rendre interprétable : il est possible d'y incorporer des motifs statistiques de saisonnalité et de tendance. Il est de plus applicable à un large éventail de domaines cibles sans nécessité de modification et rapide à entraîner.

Le modèle N-BEATS cherche à prévoir plusieurs valeurs consécutives d'une série temporelle selon un horizon d'une largeur donnée à partir des valeurs d'une fenêtre

de rétrospection sur cette série temporelle.

Son architecture est relativement simple : elle réside dans l'imbrication de différentes piles de blocs unitaires, chaque bloc unitaire étant basé sur des couches denses. La figure 2.3 dépeint cette architecture.

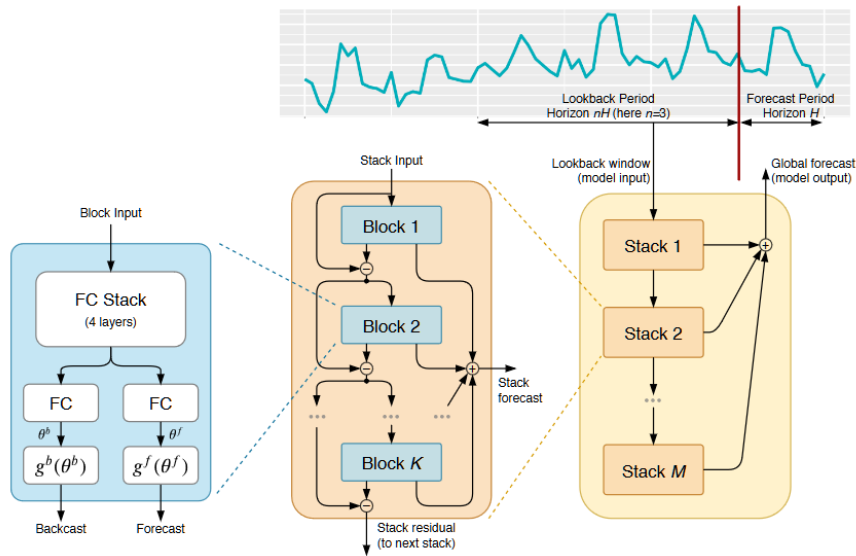


FIGURE 2.3 Visualisation de l'architecture N-BEATS (Oreshkin *et al.*, 2019)

Chaque pile du modèle N-BEATS est composée d'un ou plusieurs blocs. Chaque bloc prend en entrée l'information résiduelle du bloc précédent, c'est-à-dire la différence entre l'entrée et la sortie du bloc précédent. Un bloc i a toujours deux sorties : une extrapolation rétrospective ou rétopolation (*backcast*) $\hat{\mathbf{x}}_i$ et une prévision partielle $\hat{\mathbf{y}}_i$.

Le bloc est constitué de quatre couches denses intercalées de couches d'activation ReLU (Nair et Hinton, 2010), suivies de deux couches de projections linéaires denses branchées en parallèle ayant chacune leur fonction d'activation, chacune permettant d'obtenir une des deux sorties du bloc. La fonction d'activation utilisée dépend du type de la pile : une fonction affine pour une architecture générique,

une fonction polynomiale pour un motif de tendance, ou une somme de fonctions sinusoïdales (série de Fourier) pour un motif de saisonnalité.

Les deux sorties de la pile sont l'information résiduelle de son dernier bloc et la somme des prévisions partielles des blocs de cette pile. La sortie du modèle est la somme des sorties des piles.

2.3.1 Implémentation de Marchand et Zinflou (2021)

Le modèle présenté par Marchand et Zinflou (2021) est donc un modèle N-BEATS à trois piles, avec un seul bloc par pile. Les piles sont de type tendance, saisonnalité et générique, et la dimension des couches denses pour leurs prévisions est de 3, 24 et 12, respectivement. La longueur de la fenêtre de rétrospection ainsi que la dimension des rétropolations des blocs est de 168 pas de temps, soit de sept jours. Il y a 21 entrées exogènes, soient la prévision du besoin industriel et les données météorologiques.

Ce modèle est tiré d'un rapport écrit par Alexia Marchand et Arnaud Zinflou, chercheurs à l'IREQ, en février 2021 dans le cadre du projet « Prévision de la demande ». Marchand et Zinflou étudient donc le même problème que nous. L'ensemble de données dont ils disposent est tiré des mêmes sources que le nôtre, mais il couvre une période un peu plus courte – du 21 septembre 2015 au 17 juillet 2018 –, et il possède un attribut supplémentaire – la prévision du besoin industriel (en mégawatts).

En un deuxième temps, Marchand et Zinflou (2021) entraînent 20 modèles ayant ces paramètres et des poids initialisés aléatoirement. Une prévision d'ensemble est formée par la moyenne des prévisions de ces 20 modèles.

2.3.1.1 Équations du modèle

Cette section décrit la modélisation mathématique de N-BEATS telle qu'implémentée par Marchand et Zinflou (2021). La manière de traiter les entrées exogènes n'est pas abordée dans l'article de Oreshkin *et al.* (2019). La version présentée ici est donc plutôt tirée de l'implémentation de Remy (2020), utilisée par Marchand et Zinflou (2021).

Le modèle est composé de trois piles. Chaque pile i de ce modèle est formée d'un seul bloc : le bloc i . Chaque bloc i prend en entrée une matrice $\mathbf{x}_i \in \mathbb{R}^{7j \times 22}$ et produit en sorties une prévision partielle $\hat{\mathbf{y}}_i \in \mathbb{R}^{1j \times 1}$ et une rétropolation $\hat{\mathbf{x}}_i \in \mathbb{R}^{7j \times 22}$, correspondant aux valeurs de la série temporelle et des 21 entrées exogènes sur une fenêtre de rétrospection d'une longueur de sept jours (168 pas de temps). La somme des prévisions partielles de chacun des trois blocs du modèle est utilisée pour former la prévision finale $\hat{\mathbf{y}}$:

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}_1 + \hat{\mathbf{y}}_2 + \hat{\mathbf{y}}_3$$

où

- $\hat{\mathbf{y}} \in \mathbb{R}^{1j \times 1}$ est la prévision finale du modèle pour les 24 prochaines heures.
- $\hat{\mathbf{y}}_1$ est la prévision partielle de la pile interprétable de tendance.
- $\hat{\mathbf{y}}_2$ est la prévision partielle de la pile interprétable de saisonnalité.
- $\hat{\mathbf{y}}_3$ est la prévision partielle de la pile non interprétable générique.

L'équation suivante représente les couches communes qui sont utilisées pour la formation des deux sorties d'un bloc i :

$$\text{FC}_i^4 = \text{FC}_{i,4} \circ \text{FC}_{i,3} \circ \text{FC}_{i,2} \circ \text{FC}_{i,1}(\mathbf{x}_i)$$

où $\mathbf{x}_i \in \mathbb{R}^{7j \times 22}$ est une matrice représentant l'entrée du bloc, et $\text{FC}_{i,j}(\mathbf{x}_i) = \text{RELU}(\mathbf{W}_{i,j}\mathbf{x}_i + \mathbf{b}_{i,j}) = \max(\mathbf{0}, \mathbf{W}_{i,j}\mathbf{x}_i + \mathbf{b}_{i,j}) \in \mathbb{R}^{7j \times 22}$ est une matrice repré-

sentant la j -ième couche dense du bloc i , où $\mathbf{W}_{i,j} \in \mathbb{R}^{7j \times 7j}$ et $\mathbf{b}_{i,j} \in \mathbb{R}^{7j \times 22}$ sont les paramètres appris. $\text{FC}_i^4 \in \mathbb{R}^{7j \times 22}$ est une matrice.

À partir de ce point, les couches du bloc se divisent en deux branches, une pour chacune des deux sorties du bloc. Les opérations effectuées dépendent du type de la pile i .

La première pile est une pile à motif de tendance. Ce type de pile est interprétable, puisqu'il modélise la tendance de la série temporelle. Puisque cette pile est la première pile à l'entrée du modèle, alors $\mathbf{x}_1 = \mathbf{x}$, c'est-à-dire de ses entrées proviennent de l'ensemble de données initial, avant toute modification infligée par le modèle.

Les sorties du bloc 1 sont données par les polynômes de degré 2 suivants :

$$\hat{\mathbf{y}}_1 = \mathbf{T}^f \text{LINEAR}_1^f(\text{FC}_1^4); \quad \hat{\mathbf{x}}_1 = \mathbf{T}^b \text{LINEAR}_1^b(\text{FC}_1^4)$$

où

- $\mathbf{T}^f = [\mathbf{1}, \mathbf{t}, \mathbf{t}^2] \in \mathbb{R}^{1j \times 3}$ est la matrice des puissances de \mathbf{t} , où $\mathbf{t} = \left[\frac{0}{1j}, \frac{1}{1j}, \dots, \frac{1j-1}{1j} \right]^\top$.
- $\mathbf{T}^b = [\mathbf{1}, \mathbf{t}, \mathbf{t}^2] \in \mathbb{R}^{7j \times 3}$ est la matrice des puissances de \mathbf{t} , où $\mathbf{t} = \left[\frac{0}{7j}, \frac{1}{7j}, \dots, \frac{7j-1}{7j} \right]^\top$.
- $\text{LINEAR}_1^b(\text{FC}_1^4) = \mathbf{W}_1^b \text{FC}_1^4 \in \mathbb{R}^{3 \times 22}$ et $\text{LINEAR}_1^f(\text{FC}_1^4) = \mathbf{W}_1^f \text{FC}_1^4 \in \mathbb{R}^{3 \times 1}$ sont des matrices représentant les couches de projection linéaire des branches de rétropolation et de prévision, respectivement, où les matrices $\mathbf{W}_1^b \in \mathbb{R}^{3 \times 7j}$ et $\mathbf{W}_1^f \in \mathbb{R}^{3 \times 7j}$ sont les coefficients polynomiaux, des paramètres appris.

La deuxième pile est une pile à motif de saisonnalité. Ce type de pile est interprétable, puisqu'il modélise la saisonnalité de la série temporelle. Ses entrées sont les résidus de la pile 1 : $\mathbf{x}_2 = \mathbf{x}_1 - \hat{\mathbf{x}}_1$.

Les sorties du bloc 2 sont données par les séries de Fourier suivantes :

$$\widehat{\mathbf{y}}_2 = \mathbf{S}^f \text{LINEAR}_2^f(\text{FC}_2^4); \quad \widehat{\mathbf{x}}_2 = \mathbf{S}^b \text{LINEAR}_2^b(\text{FC}_2^4)$$

où

- $\mathbf{S}^f = [\mathbf{1}, \cos(1 \cdot 2\pi\mathbf{t}), \dots, \cos(11 \cdot 2\pi\mathbf{t}), \mathbf{0}, \sin(1 \cdot 2\pi\mathbf{t}), \dots, \sin(11 \cdot 2\pi\mathbf{t})]$
 $\in \mathbb{R}^{1j \times 24}$ est la matrice de formes d'ondes sinusoïdales, où $\mathbf{t} = [\frac{0}{1j}, \frac{1}{1j}, \dots, \frac{1j-1}{1j}]^\top$.
- $\mathbf{S}^b = [\mathbf{1}, \cos(1 \cdot 2\pi\mathbf{t}), \dots, \cos(11 \cdot 2\pi\mathbf{t}), \mathbf{0}, \sin(1 \cdot 2\pi\mathbf{t}), \dots, \sin(11 \cdot 2\pi\mathbf{t})]$
 $\in \mathbb{R}^{7j \times 24}$ est la matrice de formes d'ondes sinusoïdales, où $\mathbf{t} = [\frac{0}{7j}, \frac{1}{7j}, \dots, \frac{7j-1}{7j}]^\top$.
- $\text{LINEAR}_2^b(\text{FC}_2^4) = \mathbf{W}_2^b \text{FC}_2^4 \in \mathbb{R}^{24 \times 22}$ et $\text{LINEAR}_2^f(\text{FC}_2^4) = \mathbf{W}_2^f \text{FC}_2^4 \in \mathbb{R}^{24 \times 1}$
sont des matrices représentant les couches de projection linéaire des branches de rétropolation et de prévision, respectivement; où les matrices $\mathbf{W}_2^b \in \mathbb{R}^{24 \times 7j}$ et $\mathbf{W}_2^f \in \mathbb{R}^{24 \times 7j}$ sont les coefficients de Fourier, des paramètres appris.

La troisième et dernière pile du modèle est une pile générique. Ce type de pile est considéré non interprétable, puisqu'il modélise les parties de la série temporelle qui ne sont pas expliquées par la tendance ou la saisonnalité. Ses entrées sont les résidus de la pile 2 : $\mathbf{x}_3 = \mathbf{x}_2 - \widehat{\mathbf{x}}_2$.

Les sorties du bloc 3 sont données par la projection linéaire suivante :

$$\widehat{\mathbf{y}}_3 = \mathbf{V}_3^f \text{LINEAR}_3^f(\text{FC}_3^4) + \mathbf{b}_3^f$$

où

- Les matrices $\mathbf{V}_3^f \in \mathbb{R}^{1j \times 12}$ et $\mathbf{b}_3^f \in \mathbb{R}^{1j \times 1}$ sont les paramètres appris.
- $\text{LINEAR}_3^f(\text{FC}_3^4) = \mathbf{W}_3^f \text{FC}_3^4 \in \mathbb{R}^{12 \times 1}$ est la matrice représentant la couche de projection linéaire de la branche de prévision, où les valeurs de la matrice $\mathbf{W}_3^f \in \mathbb{R}^{12 \times 7j}$ sont les paramètres appris.

Les résidus de cette dernière pile sont ignorés.

2.3.1.2 Résultats expérimentaux

Le modèle est implémenté en Python à l'aide de l'implémentation de Remy (2020) adapté pour la bibliothèque *Keras*. Les intrants de date (c'est-à-dire le mois, le jour du mois, le jour de la semaine et l'heure) ne sont pas gardés puisqu'ils n'améliorent pas les performances selon Marchand et Zinflou (2021). Les intrants restants sont normalisés avant l'entraînement.

L'entraînement se fait à l'aide d'une validation croisée (Refaeilzadeh *et al.*, 2009) à 8 plis, ce qui améliore à la fois la performance et la stabilité du modèle. La méthode de *early stopping* est employée. L'évaluation se fait à l'aide du score MAPE sur les ensembles de validation et de test.

Les scores MAPE obtenus sont de 0,826 % sur les ensembles de validation et de test sur un modèle N-BEATS tel que décrit ci-haut. Le score MAPE est amélioré significativement en entraînant 20 de ces modèles pour évaluer la moyenne de leurs prévisions. Cette prédiction d'ensemble obtient des scores MAPE de 0,728 % et 0,710 % sur les ensembles de validation et de test, respectivement. La figure 2.4 présente pour chaque bloc de validation sur l'ensemble de validation croisée, le meilleur score des prédictions générées individuellement par chacun des 20 modèles, ainsi que le score de la moyenne des prédictions des 20 modèles.

La figure 2.5 montre le score MAPE sur l'ensemble de validation pour chacun des pas de temps de prévision, pour la meilleure prévision individuelle et la prévision d'ensemble.

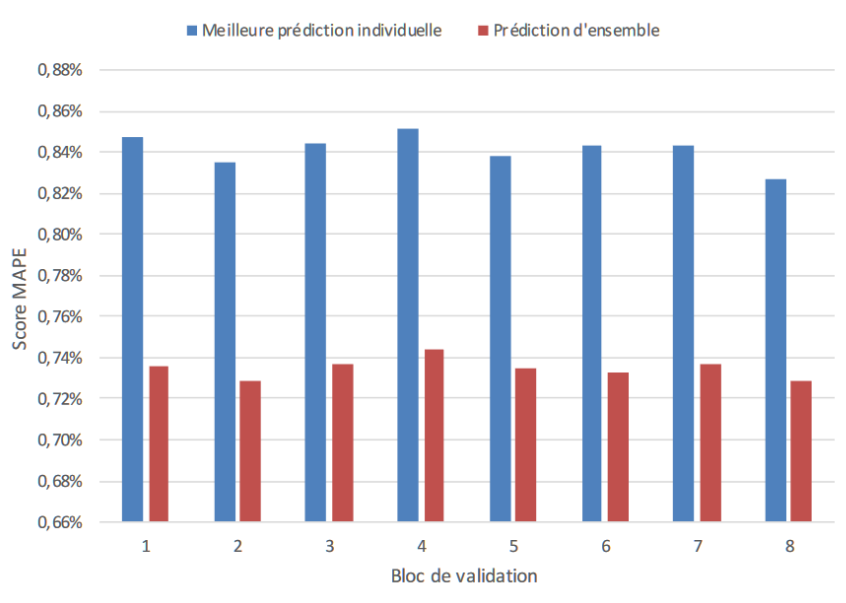


FIGURE 2.4 Prédictions individuelles et d'ensemble pour chaque bloc de validation (Marchand et Zinflou, 2021)

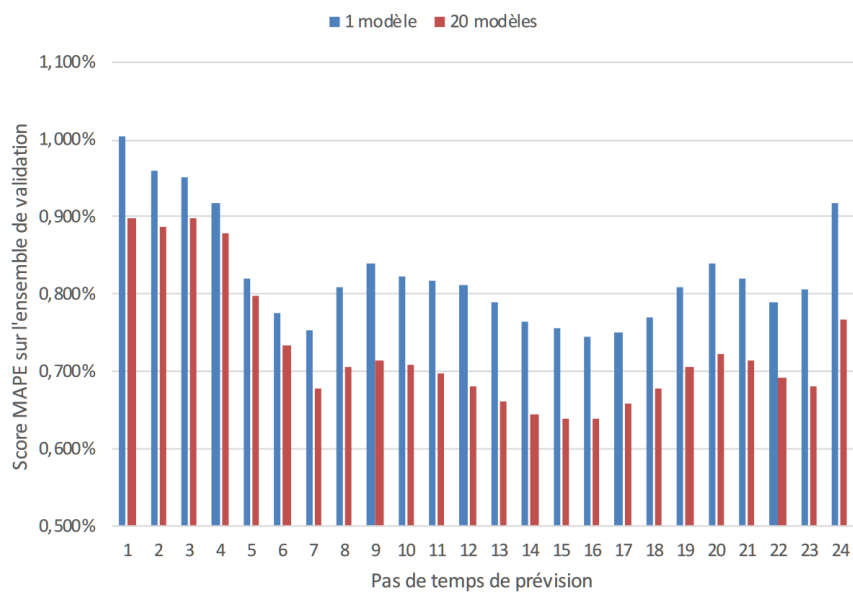


FIGURE 2.5 MAPE horaires pour des prévisions individuelles et d'ensemble (Marchand et Zinflou, 2021)

CHAPITRE III

MODÈLE PROPOSÉ

Le modèle utilisé comme référence pour notre modèle à équations linéaires multiples est celui de Clements *et al.* (2016), détaillé dans la section 2.2.1. Il a été sélectionné comme modèle de référence pour son potentiel à égaler ou même de surpasser des modèles de prévision plus complexes, ainsi que pour ses méthodes génériques lui permettant d'être appliqué à d'autres problèmes de prévisions de charge moyennant certaines adaptations.

Les changements macroéconomiques, comme la baisse de l'industrie lourde, les mutations actuelles de la société (télétravail, tarification préférentielle, etc.) et à venir (électrification des transports, autoproduction, stockage, réseaux intelligents, domotique, rôle actif du consommateur) sont des enjeux présents et futurs pour les modèles de prévision paramétriques tels que ceux développés et exploités actuellement par Hydro-Québec (Marchand et Zinflou, 2021). Un modèle possédant une composante autorégressive et de moyenne mobile, comme celui de Clements *et al.* (2016), permet de s'adapter à de tels changements dynamiques.

Voici certaines notations utilisées au cours de cette section :

- $L_t = \log(Y_t)$ est le logarithme de Y_t , où Y_t est la valeur de série temporelle de la consommation (en MW) au pas de temps t .
- j est le nombre de pas de temps par jour. Dans notre cas, $j = 24$.

- $[t]$ représente la classe de congruence de t modulo j . C'est-à-dire que $t \in [t]$.
- \mathbb{L} est l'opérateur de retard.

Les enjeux spécifiques au Québec dans le domaine de la prévision de la consommation électrique ont été présentés dans la section 1.3.

3.1 Modélisation

Notre modèle, ainsi que toutes ses variantes, peut être représenté par l'équation générale (3.1) :

$$\begin{aligned} \widehat{L}_t = c^{[t]} + \left(1 + \theta_1^{[t]}\mathbb{L}^{1j} + \theta_2^{[t]}\mathbb{L}^{7j}\right) \varepsilon_t + (\mathbb{S}^{\text{sem}} + \mathbb{S}^{\text{année}} + \mathbb{S}^{\text{jour}}) L_t \\ + \mathbb{DST} + \mathbb{M}^T + \mathbb{M}^N + \mathbb{M}^V + \mathbb{M}^S \end{aligned} \quad (3.1)$$

où la partie $c^{[t]} + \left(1 + \theta_1^{[t]}\mathbb{L}^{1j} + \theta_2^{[t]}\mathbb{L}^{7j}\right) \varepsilon_t$ est une modélisation $\text{MA}(1)_{1j} + \text{MA}(1)_{7j}$ identique à Clements *et al.* (2016); \mathbb{S}^{sem} est la composante de la saisonnalité hebdomadaire; $\mathbb{S}^{\text{année}}$ est la composante de la saisonnalité annuelle; \mathbb{S}^{jour} est la composante de la saisonnalité intrajournalière; \mathbb{DST} est la composante de l'heure d'été et \mathbb{M}^T , \mathbb{M}^N , \mathbb{M}^V , et \mathbb{M}^S sont les composantes météorologiques correspondant aux effets de la température, la nébulosité, la vitesse du vent et l'angle du soleil, respectivement. Les équations des composantes de saisonnalité sont toutes appliquées sur la série temporelle du logarithme de la consommation, comme dans le modèle de référence.

La structure $\text{MA}(1)_{1j} + \text{MA}(1)_{7j}$ se retrouve toujours dans le modèle. Pour chacune des composantes, cependant, plusieurs alternatives ou variantes de modélisation sont considérées. Des variantes d'une même composante sont identifiées par leurs différents indices. Par exemple, $\mathbb{S}_0^{\text{sem}}$ et $\mathbb{S}_1^{\text{sem}}$ sont deux variantes de la composante \mathbb{S}^{sem} .

Un modèle est identifié par la liste des identifiants de ses composantes. Par exemple, des variations du modèle pourraient être identifiées par $(\mathbb{S}_1^{\text{sem}}, \mathbb{S}_1^{\text{année}}, \mathbb{S}_2^{\text{jour}}, \mathbb{M}_1^T, \mathbb{M}_4^N, \mathbb{M}_1^V)$, $(\mathbb{S}_2^{\text{sem}})$ ou $(\mathbb{S}_0^{\text{sem}}, \mathbb{S}_2^{\text{année}}, \mathbb{M}_b^T, \mathbb{M}_a^N)$. Ces modèles correspondent tous à l'équation (3.1), mais avec différentes composantes. Lorsqu'une composante est absente, son identifiant n'apparaît pas dans l'identifiant du modèle.

De plus, deux variations du traitement des données d'entrée sont considérées. Dans la première variante, $[t]$ correspond à l'heure de l'Est (EST) pour toutes les entrées de l'ensemble de données. Dans la deuxième variante, $[t]$ correspond à l'heure au fuseau horaire (EST ou DST), afin que chacun des 24 modèles horaires puisse suivre le rythme de l'activité humaine. L'identifiant du modèle a donc lui-même un indice, pour identifier laquelle de ces deux variantes est utilisée. Par exemple, $(\mathbb{S}_0^{\text{sem}}, \mathbb{S}_0^{\text{année}})_1$ utilise la première variante et $(\mathbb{S}_0^{\text{sem}}, \mathbb{S}_0^{\text{année}})_2$ utilise la deuxième variante.

Les sous-sections suivantes identifient toutes les variantes de composantes considérées.

3.1.1 Saisonnalité hebdomadaire \mathbb{S}^{sem}

Les identifiants de composantes de saisonnalités hebdomadaires possibles sont les suivants :

- (a) $\mathbb{S}_0^{\text{sem}} = \phi_1^{[t]} \mathbb{L}^{1j}$. Cette composante triviale n'a qu'un seul paramètre.
- (b) $\mathbb{S}_2^{\text{sem}}$ est donnée par l'équation (2.20), où $\mathcal{W} = \{\text{lundis aux vendredis, samedis et dimanches}\}$ partitionne chaque pas de temps selon s'il appartient à un jour ouvrable (lundi à vendredi) ou à une de fin de semaine (samedi et dimanche). Cette composante a 2 paramètres.
- (c) $\mathbb{S}_7^{\text{sem}}$ est donnée par l'équation (2.20), où $\mathcal{W} = \{\text{lundis, mardis, mercredis, jeudis, vendredis, samedis, dimanches}\}$ partitionne chaque pas de temps

selon le jour de la semaine auquel il appartient. Cette composante a 7 paramètres et est identique à celle de Clements *et al.* (2016).

3.1.2 Saisonnalité annuelle $\mathbb{S}^{\text{année}}$

L'équation pour la saisonnalité annuelle $\mathbb{S}^{\text{année}}$ ne subit que deux légères extensions par rapport à Clements *et al.* (2016) : (1) l'intervalle $\mathcal{Q} = [1, 4]$ devient un ensemble qui n'est pas nécessairement continu, et (2) la période de la série de Fourier est prolongée, passant de $364j$ à $365.2425j$. L'équation (2.21) est donc ajustée en conséquence, aboutissant à l'équation (3.2) :

$$\mathbb{S}^{\text{année}} = \phi_2^{[t]} \mathbb{L}^{\tau_j} + \sum_{q \in \mathcal{Q}} \varphi_q^{[t]} \text{FOURIER} \left(\frac{2q\pi t}{365.2425j} \right) \mathbb{L}^{\tau_j} \quad (3.2)$$

Les composantes utilisant cette équation ont un nombre de paramètres égal à $2|\mathcal{Q}| + 1$.

Les identifiants de composantes de saisonnalité annuelle possibles sont les suivants :

- (a) $\mathbb{S}_0^{\text{année}} = \phi_2^{[t]} \mathbb{L}^{\tau_j}$. Cette composante triviale n'a qu'un seul paramètre.
- (b) $\mathbb{S}_{1..4}^{\text{année}}$ est donnée par l'équation (3.2) où $\mathcal{Q} = \{1, 2, 3, 4\}$, similaire à Clements *et al.* (2016). Cette composante a 9 paramètres.
- (c) $\mathbb{S}_{1..6}^{\text{année}}$ est donnée par l'équation (3.2) où $\mathcal{Q} = \{1, 2, 3, 4, 5, 6\}$. Cette composante a 13 paramètres.
- (d) $\mathbb{S}_{q|12}^{\text{année}}$ est donnée par l'équation (3.2) où $\mathcal{Q} = \{1, 2, 3, 4, 6, 12\}$ est l'ensemble des diviseurs de 12. Cette composante a 13 paramètres.

3.1.3 Corrélation intrajournalière \mathbb{S}^{jour}

La composante de corrélation journalière est elle-même composée d'au plus trois sous-composantes.

L'équation (3.3) présente s_1 , la sous-composante qui va chercher la dernière valeur observée.

$$s_1 L_t = \mathbb{1}_{\{1j-1 \notin [t]\}} \psi_1^{[t]} \mathbb{L}^{(t \bmod j)+1} L_t \quad (3.3)$$

où $\mathbb{1}_{\{1j-1 \notin [t]\}}$ est la variable indicatrice qui est égale à 1 si $1j - 1 \notin [t]$ et égale à 0 sinon. Elle sert à éviter une redondance entre les variables du sous-modèle horaire associé au pas de temps $1j - 1$, où $\mathbb{L}^{(t \bmod j)+1} = \mathbb{L}^{1j}$.

L'équation (3.4) présente s_2 ou s'_2 , deux variantes d'implémentation de la sous-composante qui va chercher la prévision au pas de temps précédent. L'implémentation s_2 est une implémentation paresseuse qui utilise la vraie valeur L_t , plutôt que \widehat{L}_t , lors de l'entraînement afin de permettre au programme de paralléliser cette étape; alors que l'implémentation s'_2 est une implémentation complète, qui ne prend pas un tel raccourci.

$$s_2 L_t = \psi_2^{[t]} \mathbb{L}^1 \widehat{L}_t \quad (3.4)$$

Pour éviter une redondance entre les variables du sous-modèle horaire associé au pas de temps 0, où $\mathbb{L}^1 = \mathbb{L}^{(t \bmod j)+1}$, s_2 est multiplié par la variable indicatrice $\mathbb{1}_{\{0 \notin [t]\}}$ lorsque s_1 est présent dans la composante.

L'équation (3.5) présente s_3 , la sous-composante qui va chercher la moyenne des consommations de la dernière journée observée. Cette sous-composante, comme la sous-composante s_1 , est pertinente car elle apporte la connaissance d'une valeur récente au modèle. Contrairement à s_1 , elle ne suppose pas que la valeur de consommation au dernier pas de temps de l'horizon (qui, dans notre cas, correspond toujours à la consommation entre 23 heures et minuit) est la valeur la plus caractéristique de la journée.

$$s_3 L_t = \psi_3^{[t]} \frac{\sum_{i=1}^j \mathbb{L}^{(t \bmod j)+i}}{j} L_t \quad (3.5)$$

Les identifiants de composantes de corrélations intra-journalières utilisent une

forme d'encodage *one-hot* pour désigner les sous-composantes utilisées. Les identifiants possibles sont donc les suivants :

- (a) $\mathbb{S}_{001}^{\text{jour}} = \mathfrak{s}_1$. Cette composante a un seul paramètre.
- (b) $\mathbb{S}_{010}^{\text{jour}} = \mathfrak{s}_2$. Cette composante a un seul paramètre.
- (c) $\mathbb{S}_{100}^{\text{jour}} = \mathfrak{s}_3$. Cette composante a un seul paramètre.
- (d) $\mathbb{S}_{011}^{\text{jour}} = \mathfrak{s}_1 + \mathbb{1}_{\{0 \notin [t]\}} \mathfrak{s}_2$. Cette composante a 2 paramètres.
- (e) $\mathbb{S}_{101}^{\text{jour}} = \mathfrak{s}_1 + \mathfrak{s}_3$. Cette composante a 2 paramètres.
- (f) $\mathbb{S}_{110}^{\text{jour}} = \mathfrak{s}_2 + \mathfrak{s}_3$. Cette composante a 2 paramètres.
- (g) $\mathbb{S}_{111}^{\text{jour}} = \mathfrak{s}_1 + \mathbb{1}_{\{0 \notin [t]\}} \mathfrak{s}_2 + \mathfrak{s}_3$. Cette composante a 3 paramètres.
- (h) $\mathbb{S}_{020}^{\text{jour}}$, $\mathbb{S}_{021}^{\text{jour}}$, $\mathbb{S}_{120}^{\text{jour}}$, et $\mathbb{S}_{121}^{\text{jour}}$. Ces composantes sont respectivement similaires aux composantes $\mathbb{S}_{010}^{\text{jour}}$, $\mathbb{S}_{011}^{\text{jour}}$, $\mathbb{S}_{110}^{\text{jour}}$, et $\mathbb{S}_{111}^{\text{jour}}$; où \mathfrak{s}'_2 est substitué à \mathfrak{s}_2 et elles ont le même nombre de paramètres. La composante $\mathbb{S}_{021}^{\text{jour}}$ est équivalente à celle de Clements *et al.* (2016).

3.1.4 Variables météorologiques

Les attributs météorologiques de notre ensemble de données sont la température, la nébulosité et la vitesse du vent, chacun mesuré aux stations de Baie-Comeau (**bc**), Ottawa (**ow**), Québec (**qc**), Sherbrooke (**sh**), et Montréal (**yu1**). Les modèles similaires (Clements *et al.*, 2016; Cancelo *et al.*, 2008) agrègent usuellement les différentes mesures météorologiques en une seule variable à l'aide d'une moyenne pondérée par population pour éviter d'avoir des variables mutuellement dépendantes. Ne connaissant pas les poids appropriés à associer à chacune des stations, nous avons décidé de garder les variables séparées.

En plus d'utiliser plusieurs types de variables météorologiques que le modèle de référence et de cette redondance dans les variables de même type, notre modèle se

distingue en utilisant les valeurs décalées (normalement par 1j et 7j) en plus des valeurs au pas de temps de la prévision.

Voici certaines notations qui sont utilisées dans les sections suivantes :

- $\mathcal{O} = \{0, 1j, 7j\}$ est l'ensemble des décalages temporels considérés.
- $\mathcal{L} = \{\text{bc}, \text{ow}, \text{qc}, \text{sh}, \text{yul}\}$ est l'ensemble des stations météorologiques ;
- $\mathcal{M}_T = \{T_{\ell,t} \mid \forall \ell \in \mathcal{L}\}$ est l'ensemble des températures (en degrés Celsius) observées à chaque station météorologique ;
- $\mathcal{M}_N = \{N_{\ell,t} \mid \forall \ell \in \mathcal{L}\}$ est l'ensemble des valeurs de la nébulosité (en octas ajustés) observées à chaque station météorologique ;
- $\mathcal{M}_V = \{V_{\ell,t} \mid \forall \ell \in \mathcal{L}\}$ est l'ensemble des vitesses du vent (en kilomètres par heure) observées à chaque station météorologique ;
- S_t est la série temporelle de l'angle du soleil au-dessus de l'horizon (en degrés), mesuré à Montréal.

3.1.5 Températures \mathbb{M}^T

L'équation (3.6) présente la composante des effets de la température, adaptée de Clements *et al.* (2016) :

$$\mathbb{M}^T = \sum_{s \in \mathcal{O}} \sum_{\ell \in \mathcal{L}} \tau_{T,\ell,s}^{[t]} \otimes m(T_{\ell,t-s}) \quad (3.6)$$

où $T_{\ell,t-s}$ est la série temporelle dont les valeurs correspondent à la température enregistrée à la station météorologique de la ville ℓ au pas de temps $t-s$; m est une spécification linéaire par morceaux de cette température ; et $\tau_{T,s,\ell,i}^{[t]}$, pour $m_i \in m$, $s \in \mathcal{O}$ et $\ell \in \mathcal{L}$, sont les paramètres d'apprentissage. Le modèle de Clements *et al.* (2016) correspond donc au cas spécial où $\mathcal{O} = \{0j\}$ et $\mathcal{L} = \{\text{Queensland}\}$. Cette composante a un nombre de paramètres égal à $|\mathcal{O}| \times |\mathcal{L}| \times |m|$.

Alternativement, l'équation (3.7) présente une composante de la température qui tente de se débarrasser de la dépendance entre les variables de même type en utilisant certaines températures à des stations ou à des décalages temporels précis comme valeurs de référence :

$$\mathbb{M}^T = \sum_{s \in \mathcal{O}} \sum_{\ell \in \mathcal{L}} \mathbb{1}_{\{T_{\ell,t-s} \notin \mathcal{X}_{\text{ref}}\}} \left(\tau_{T,2,\ell,s}^{[t]} \mathbb{1}_{\{T_{\ell,t-s} \leq 15.1\}} + \tau_{T,3,\ell,s}^{[t]} \mathbb{1}_{\{T_{\ell,t-s} > 15.1\}} \right) \bar{X}_{\ell,t,s} + \sum_{T_{\text{ref}} \in \mathcal{X}_{\text{ref}}} \tau_{T,1}^{[t]} \otimes m(T_{\text{ref}}) \quad (3.7)$$

où \mathcal{X}_{ref} est l'ensemble des températures de référence, $\bar{X}_{\ell,t,s}$ est une opération de transformation des valeurs de températures pour les exprimer en fonction des températures de référence. La variable indicatrice $\mathbb{1}_{\{T_{\ell,t-s} \notin \mathcal{X}_{\text{ref}}\}}$ permet de filtrer les variables de température pour ne garder que celles qui ne font pas partie de l'ensemble de variables de références ; et les variables indicatrices $\mathbb{1}_{\{T_{\ell,t} \leq 15.1\}}$ et $\mathbb{1}_{\{T_{\ell,t} > 15.1\}}$ permettent de diviser les valeurs de températures en deux colonnes, en fonction de si l'effet d'augmenter ou de baisser la température de quelques degrés devrait avoir un effet positif ou négatif sur la consommation électrique.

Cette composante a un nombre de paramètres égal à $2|\mathcal{O}| \times |\mathcal{L}| + |\mathcal{X}_{\text{ref}}| \times (|m| - 2)$. En pratique, ce nombre de paramètres est souvent plus petit que ceux introduits par l'équation (3.6).

La spécification linéaire par morceaux de la température m peut prendre les valeurs suivantes :

1. $m_1 = \{h_1, h_2, c_1, c_2\} = \{\text{CLRAMP}_{-23,13}^-, \text{CLRAMP}_{-23,1}^-; \text{CLRAMP}_{21,33}^+, \text{CLRAMP}_{28,33}^+\}$. Cette spécification comprend quatre variables (deux pour la zone froide et deux pour la zone chaude) et des effets d'épuisement de chaque côté de l'échelle des températures, de manière similaire à Clements *et al.* (2016). Les intervalles $[-23^\circ\text{C}, 13^\circ\text{C}]$, $[-23^\circ\text{C}, 1^\circ\text{C}]$, $[21^\circ\text{C}, 33^\circ\text{C}]$ et $[28^\circ\text{C}, 33^\circ\text{C}]$ sont différents de ceux sélectionnés dans Clements *et al.*

(2016) et ont été déduits en examinant la corrélation entre la consommation de charge électrique et la température sur le territoire québécois dans le but d’obtenir un nombre égal de variables de réchauffement et de variables de refroidissement. Les effets de la température sur la consommation électrique au Québec ont été présentés à la section 1.3.4.

2. $\mathfrak{m}_2 = \{h_1, h_2, h_3, c_1\} = \{\text{CLRAMP}^-_{-\infty,12.0}, \text{CLRAMP}^-_{-\infty,1.7}, \text{CLRAMP}^-_{-\infty,-8.9}; \text{CLRAMP}^+_{19.4,\infty}\}$. Cette spécification comprend aussi quatre variables, mais trois d’entre elles modélisent la zone froide où se trouve la plus grande partie de nos entrées de données. Les effets d’épuisement de chaque côté de l’échelle des températures sont ignorés. Les intervalles ont été sélectionnés grâce à la segmentation linéaire à quatre points de rupture présentée dans la section 1.3.4.

Nous utilisons les mêmes points de rupture pour chaque location. En effet, un examen des corrélations entre les températures et la consommation électrique aux différentes localisations des stations météorologiques ne révèle aucune différence significative.

Les valeurs de l’ensemble \mathcal{X}_{ref} et de l’opération $\bar{X}_{\ell,t,s}$ viennent toujours en paire, puisqu’ils sont définis en relation l’un de l’autre. Ces valeurs sont définies en fonction de la variable exogène générique $X_{\ell,t-s}$, qui équivaut à $T_{\ell,t-s}$, $N_{\ell,t-s}$, ou $V_{\ell,t-s}$, dépendamment du contexte. Elles peuvent prendre les valeurs suivantes :

1. $\mathcal{X}_1^{\text{ref}} = \{X_{\text{yul},t-s} \mid \forall s \in \mathcal{O}\}$ et $\bar{X}_1 = X_{\text{yul},t-s} - X_{\ell,t-s}$. Toutes les valeurs de la variable mesurées à la station de Montréal sont utilisées comme référence, et les autres valeurs sont comparées à la valeur de référence ayant le même décalage. Montréal est utilisée comme référence car elle représente la région la plus peuplée du Québec ;
2. $\mathcal{X}_5^{\text{ref}} = \{X_{\ell,t} \mid \forall \ell \in \mathcal{L}\}$, et $\bar{X}_5 = X_{\ell,t} - X_{\ell,t-s} = \Delta_s X_{\ell,t}$. Toutes les valeurs de la variable mesurées au pas de temps t (sans décalage) sont utilisées comme

référence, et les autres valeurs sont comparées à la valeur de référence à la même station, à la manière d'une différentiation de premier ordre saisonnière de période s .

3. $\mathcal{X}_2^{\text{ref}} = \{X_{\text{yul},t}\}$, et $\bar{X}_2 = X_{\text{yul},t} - X_{\ell,t-s}$. La variable à la station de Montréal et au pas de temps t (sans décalage) est utilisée comme référence, contre laquelle toutes les autres valeurs sont comparées ;

$$4. \mathcal{X}_3^{\text{ref}} = \{X_{\text{yul},t}\}, \text{ et } \bar{X}_3 = \begin{cases} X_{\text{yul},t} - X_{\ell,t}, & s = 0 \\ X_{\ell,t} - X_{\ell,t-s} = \Delta_s X_{\ell,t}, & s > 0 \end{cases}.$$

La variable à la station de Montréal et au pas de temps t (sans décalage) est utilisée comme référence, contre laquelle toutes les autres valeurs sans décalage sont comparées. Les autres valeurs avec décalages sont comparées à la valeur sans décalage à la même station, à la manière d'une différentiation de premier ordre saisonnière de période s ;

$$5. \mathcal{X}_4^{\text{ref}} = \{X_{\text{yul},t}\}, \text{ et } \bar{X}_4 = \begin{cases} X_{\text{yul},t} - X_{\text{yul},t-s} = \Delta_s X_{\text{yul},t}, & \ell = \text{yul} \\ X_{\text{yul},t-s} - X_{\ell,t-s}, & \ell \neq \text{yul} \end{cases}.$$

La variable à la station de Montréal et au pas de temps t (sans décalage) est utilisée comme référence, contre laquelle toutes les autres valeurs à Montréal sont comparées. Les valeurs aux autres stations sont comparées à la valeur à Montréal ayant le même décalage.

Les identifiants de composantes de températures possibles sont les suivants :

- (a) \mathbb{M}_1^T est donnée par l'équation (3.6), où $\mathcal{L} = \{\text{yul}\}$ et $m = m_1$. Cette composante a 12 paramètres.
- (b) \mathbb{M}_2^T est donnée par l'équation (3.6), où $\mathcal{L} = \{\text{yul}\}$ et $m = m_2$. Cette composante a 12 paramètres.
- (c) \mathbb{M}_3^T est donnée par l'équation (3.6), où $\mathcal{L} = \{\text{yul}, \text{qc}, \text{ow}, \text{sh}, \text{bc}\}$ et $m = m_1$. Cette composante a 60 paramètres.

- (d) \mathbb{M}_4^T est donnée par l'équation (3.6), où $\mathcal{L} = \{\text{yul}, \text{qc}, \text{ow}, \text{sh}, \text{bc}\}$ et $m = \mathfrak{m}_2$. Cette composante a 60 paramètres.
- (e) $\mathbb{M}_{\delta_1}^T$ est donnée par l'équation (3.7), où $\mathcal{L} = \{\text{yul}\}$, $m = \mathfrak{m}_1$, $\mathcal{X}_{\text{ref}} = \mathcal{X}_2^{\text{ref}}$ et $\bar{X}_{\ell,t,s} = \bar{X}_2$ où $X_{\ell,t-s} = T_{\ell,t-s}$. Cette composante a 8 paramètres.
- (f) $\mathbb{M}_{\delta_2}^T$ est donnée par l'équation (3.7), où $\mathcal{L} = \{\text{yul}\}$, $m = \mathfrak{m}_2$, $\mathcal{X}_{\text{ref}} = \mathcal{X}_2^{\text{ref}}$ et $\bar{X}_{\ell,t,s} = \bar{X}_2$ où $X_{\ell,t-s} = T_{\ell,t-s}$. Cette composante a 8 paramètres.
- (g) $\mathbb{M}_{\delta_3}^T$ est donnée par l'équation (3.7), où $\mathcal{L} = \{\text{yul}, \text{qc}, \text{ow}, \text{sh}, \text{bc}\}$, $m = \mathfrak{m}_1$, $\mathcal{X}_{\text{ref}} = \mathcal{X}_1^{\text{ref}}$ et $\bar{X}_{\ell,t,s} = \bar{X}_1$, où $X_{\ell,t-s} = T_{\ell,t-s}$. Cette composante a 36 paramètres.
- (h) $\mathbb{M}_{\delta_{11}}^T$ est donnée par l'équation (3.7), où $\mathcal{L} = \{\text{yul}, \text{qc}, \text{ow}, \text{sh}, \text{bc}\}$, $m = \mathfrak{m}_1$, $\mathcal{X}_{\text{ref}} = \mathcal{X}_5^{\text{ref}}$ et $\bar{X}_{\ell,t,s} = \bar{X}_5$, où $X_{\ell,t-s} = T_{\ell,t-s}$. Cette composante a 40 paramètres.
- (i) $\mathbb{M}_{\delta_4}^T$ est donnée par l'équation (3.7), où $\mathcal{L} = \{\text{yul}, \text{qc}, \text{ow}, \text{sh}, \text{bc}\}$, $m = \mathfrak{m}_1$, $\mathcal{X}_{\text{ref}} = \mathcal{X}_2^{\text{ref}}$ et $\bar{X}_{\ell,t,s} = \bar{X}_2$, où $X_{\ell,t-s} = T_{\ell,t-s}$. Cette composante a 32 paramètres.
- (j) $\mathbb{M}_{\delta_5}^T$ est donnée par l'équation (3.7), où $\mathcal{L} = \{\text{yul}, \text{qc}, \text{ow}, \text{sh}, \text{bc}\}$, $m = \mathfrak{m}_1$, $\mathcal{X}_{\text{ref}} = \mathcal{X}_3^{\text{ref}}$ et $\bar{X}_{\ell,t,s} = \bar{X}_3$, où $X_{\ell,t-s} = T_{\ell,t-s}$. Cette composante a 32 paramètres.
- (k) $\mathbb{M}_{\delta_6}^T$ est donnée par l'équation (3.7), où $\mathcal{L} = \{\text{yul}, \text{qc}, \text{ow}, \text{sh}, \text{bc}\}$, $m = \mathfrak{m}_1$, $\mathcal{X}_{\text{ref}} = \mathcal{X}_4^{\text{ref}}$ et $\bar{X}_{\ell,t,s} = \bar{X}_4$, où $X_{\ell,t-s} = T_{\ell,t-s}$. Cette composante a 32 paramètres.
- (l) $\mathbb{M}_{\delta_7}^T$ est donnée par l'équation (3.7), où $\mathcal{L} = \{\text{yul}, \text{qc}, \text{ow}, \text{sh}, \text{bc}\}$, $m = \mathfrak{m}_2$, $\mathcal{X}_{\text{ref}} = \mathcal{X}_1^{\text{ref}}$ et $\bar{X}_{\ell,t,s} = \bar{X}_1$, où $X_{\ell,t-s} = T_{\ell,t-s}$. Cette composante a 36 paramètres.
- (m) $\mathbb{M}_{\delta_{12}}^T$ est donnée par l'équation (3.7), où $\mathcal{L} = \{\text{yul}, \text{qc}, \text{ow}, \text{sh}, \text{bc}\}$, $m = \mathfrak{m}_2$, $\mathcal{X}_{\text{ref}} = \mathcal{X}_5^{\text{ref}}$ et $\bar{X}_{\ell,t,s} = \bar{X}_5$, où $X_{\ell,t-s} = T_{\ell,t-s}$. Cette composante a 40 paramètres.

- (n) $\mathbb{M}_{\delta 8}^T$ est donnée par l'équation (3.7) où $\mathcal{L} = \{\text{yul}, \text{qc}, \text{ow}, \text{sh}, \text{bc}\}$, $m = m_2$, $\mathcal{X}_{\text{ref}} = \mathcal{X}_2^{\text{ref}}$ et $\bar{X}_{\ell,t,s} = \bar{X}_2$, où $X_{\ell,t-s} = T_{\ell,t-s}$. Cette composante a 32 paramètres.
- (o) $\mathbb{M}_{\delta 9}^T$ est donnée par l'équation (3.7), où $\mathcal{L} = \{\text{yul}, \text{qc}, \text{ow}, \text{sh}, \text{bc}\}$, $m = m_2$, $\mathcal{X}_{\text{ref}} = \mathcal{X}_3^{\text{ref}}$ et $\bar{X}_{\ell,t,s} = \bar{X}_3$, où $X_{\ell,t-s} = T_{\ell,t-s}$. Cette composante a 32 paramètres.
- (p) $\mathbb{M}_{\delta 10}^T$ est donnée par l'équation (3.7), où $\mathcal{L} = \{\text{yul}, \text{qc}, \text{ow}, \text{sh}, \text{bc}\}$, $m = m_2$, $\mathcal{X}_{\text{ref}} = \mathcal{X}_4^{\text{ref}}$ et $\bar{X}_{\ell,t,s} = \bar{X}_4$, où $X_{\ell,t-s} = T_{\ell,t-s}$. Cette composante a 32 paramètres.

3.1.6 Autres variables météorologiques $\mathbb{M}^N + \mathbb{M}^V$

Les composante de la nébulosité \mathbb{M}^N et de la vitesse du vent \mathbb{M}^V utilisent toujours une équation qui dépend de celle utilisée par la composante de température choisie – seule la spécification linéaire change. L'équation (3.8) correspond au cas où \mathbb{M}^T utilise l'équation (3.6), et l'équation (3.9) correspond au cas où \mathbb{M}^T utilise l'équation (3.7).

L'expression (3.8) est similaire à l'équation correspondante (3.6), mais utilise une spécification linéaire par morceaux x appropriée pour la variable :

$$\mathbb{M}^X = \sum_{s \in \mathcal{O}} \sum_{\ell \in \mathcal{L}} \tau_{X,s,\ell}^{[t]} \otimes x(X_{\ell,t-s}) \quad (3.8)$$

où $X_{\ell,t-s}$ est la série temporelle de la nébulosité ($N_{\ell,t-s}$) ou de la vitesse du vent ($V_{\ell,t-s}$) à la station de la ville ℓ et au pas de temps $t - s$; x est une spécification linéaire par morceaux appropriée à la série temporelle en question; et $\tau_{X,s,\ell,i}^{[t]}$, pour $x_i \in x$, $s \in \mathcal{O}$, $\ell \in \mathcal{L}$ et $X \in \{N, V\}$ sont les paramètres d'apprentissage.

L'expression (3.9) est similaire à l'équation correspondante (3.7), mais utilise une spécification linéaire par morceaux x appropriée pour la variable et n'utilise pas

de variable indicatrice pour séparer les valeurs en deux colonnes, puisque la forme des courbes de corrélations entre la consommation et la nébulosité ou la vitesse du vent ne le justifient pas (*cf.* les figures 1.15(a), 1.18 et 1.19 de la section 1.3.4) :

$$\mathbb{M}^X = \sum_{X_{\text{ref}} \in \mathcal{X}_{\text{ref}}} \tau_{X,1,\text{ref}}^{[t]} \otimes x(X_{\text{ref}}) + \sum_{s \in \mathcal{O}} \sum_{\ell \in \mathcal{L}} \mathbb{1}_{\{X_{\ell,t-s} \notin \mathcal{X}_{\text{ref}}\}} \tau_{X,2,\ell,s}^{[t]} \bar{X}_{\ell,t,s} \quad (3.9)$$

où \mathcal{X}_{ref} et $\bar{X}_{\ell,t,s}$ sont définis de manière identique à ceux de la composante de température, mais sont appliqués sur les séries temporelles $N_{\ell,t-s}$ ou $V_{\ell,t-s}$, plutôt que $T_{\ell,t-s}$. Cette composante a un nombre de paramètres égal à $|\mathcal{O}| \times |\mathcal{L}| + |\mathcal{X}_{\text{ref}}| \times (|x| - 1)$.

Les valeurs de x possibles lorsque $X_{\ell,t-s} = N_{\ell,t-s}$ sont les suivantes :

1. $\mathfrak{n}_1 = \{\text{CLRAMP}_{0,3}^+, \text{CLRAMP}_{3,10}^+, \text{CLRAMP}_{9,10}^+\}$.
2. $\mathfrak{n}_2 = \{\text{CLRAMP}_{6,10}^+, \text{CLRAMP}_{9,10}^+\}$.
3. $\mathfrak{n}_3 = \{\text{CLRAMP}_{8,8,10}^+\}$. Ces intervalles ont été sélectionnés grâce à la segmentation linéaire des effets de la nébulosité présentée dans la section 1.3.4.
4. $\mathfrak{n}_4 = \{\text{CLRAMP}_{0,2}^-\}$. Ces intervalles ont été sélectionnés grâce à la segmentation linéaire des effets de la nébulosité après avoir enlevé les effets de la température, présentée dans la section 1.3.4.
5. $\mathfrak{n}_5 = \mathfrak{n}_3 \cup \mathfrak{n}_4 = \{\text{CLRAMP}_{0,2}^-, \text{CLRAMP}_{8,8,10}^+\}$.

Les valeurs de x possibles lorsque $X_{\ell,t-s} = V_{\ell,t-s}$ sont les suivantes :

1. $\mathfrak{v}_1 = \{\text{CLRAMP}_{12,39}^+\}$.
2. $\mathfrak{v}_2 = \{\text{CLRAMP}_{11,7,33}^+\}$. Ces intervalles ont été sélectionnés grâce à la segmentation linéaire des effets de la vitesse du vent présentée dans la section 1.3.4.
3. $\mathfrak{v}_3 = \{\text{CLRAMP}_{0,20}^+\}$. Ces intervalles ont été sélectionnés grâce à la segmentation linéaire des effets de la vitesse du vent après avoir enlevé les effets de la température, présentée dans la section 1.3.4.
4. $\mathfrak{v}_4 = \mathfrak{v}_2 \cup \mathfrak{v}_3 = \{\text{CLRAMP}_{0,20}^+, \text{CLRAMP}_{11,7,33}^+\}$.

Les identifiants de composantes de nébulosité possibles sont les suivants :

- (a) $\mathbb{M}_1^N : x = \mathfrak{n}_1$, et $X_{\ell,t-s} = N_{\ell,t-s}$. Cette composante a $9|\mathcal{L}|$ paramètres (équation (3.8)) ou $3|\mathcal{L}| + 2|\mathcal{X}_{\text{ref}}|$ paramètres (équation (3.9)); ce qui équivaut à 9 ou 45 paramètres (équation (3.8) avec $|\mathcal{L}| = 1$ ou 5, respectivement) ou à 5, 17, 21, ou 25 paramètres (équation (3.9) avec $\langle |\mathcal{L}|, |\mathcal{X}_{\text{ref}}| \rangle = \langle 1, 1 \rangle, \langle 5, 1 \rangle, \langle 5, 3 \rangle$ ou $\langle 5, 5 \rangle$, respectivement).
- (b) $\mathbb{M}_2^N : x = \mathfrak{n}_2$, et $X_{\ell,t-s} = N_{\ell,t-s}$. Cette composante a $6|\mathcal{L}|$ paramètres (équation (3.8)) ou $3|\mathcal{L}| + |\mathcal{X}_{\text{ref}}|$ paramètres (équation (3.9)); ce qui équivaut à 6 ou 30 paramètres (équation (3.8) avec $|\mathcal{L}| = 1$ ou 5, respectivement) ou à 4, 16, 18, ou 20 paramètres (équation (3.9) avec $\langle |\mathcal{L}|, |\mathcal{X}_{\text{ref}}| \rangle = \langle 1, 1 \rangle, \langle 5, 1 \rangle, \langle 5, 3 \rangle$ ou $\langle 5, 5 \rangle$, respectivement).
- (c) $\mathbb{M}_3^N : x = \mathfrak{n}_3$, et $X_{\ell,t-s} = N_{\ell,t-s}$. Cette composante a $3|\mathcal{L}|$ paramètres; ce qui équivaut à 3 ou 15 paramètres (avec $|\mathcal{L}| = 1$ ou 5, respectivement).
- (d) $\mathbb{M}_4^N : x = \mathfrak{n}_4$, et $X_{\ell,t-s} = N_{\ell,t-s}$. Cette composante a $3|\mathcal{L}|$ paramètres.
- (e) $\mathbb{M}_5^N : x = \mathfrak{n}_5$, et $X_{\ell,t-s} = N_{\ell,t-s}$. Cette composante a $6|\mathcal{L}|$ paramètres (équation (3.8)) ou $3|\mathcal{L}| + |\mathcal{X}_{\text{ref}}|$ paramètres (équation (3.9)).

Les identifiants de composantes de vitesse du vent possibles sont les suivants :

- (a) $\mathbb{M}_1^V : x = \mathfrak{v}_1$, et $X_{\ell,t-s} = V_{\ell,t-s}$. Cette composante a $3|\mathcal{L}|$ paramètres.
- (b) $\mathbb{M}_2^V : x = \mathfrak{v}_2$, et $X_{\ell,t-s} = V_{\ell,t-s}$. Cette composante a $3|\mathcal{L}|$ paramètres.
- (c) $\mathbb{M}_3^V : x = \mathfrak{v}_3$, et $X_{\ell,t-s} = V_{\ell,t-s}$. Cette composante a $3|\mathcal{L}|$ paramètres.
- (d) $\mathbb{M}_4^V : x = \mathfrak{v}_4$, et $X_{\ell,t-s} = V_{\ell,t-s}$. Cette composante a $6|\mathcal{L}|$ paramètres (équation (3.8)) ou $3|\mathcal{L}| + |\mathcal{X}_{\text{ref}}|$ paramètres (équation (3.9)).

3.1.7 Angle du soleil \mathbb{M}^S

La variables S_t est une série temporelle dont les valeurs correspondent à l'angle du soleil (en degrés) par rapport à l'horizon à Montréal. Puisque le modèle complet

est essentiellement formé de 24 sous-modèles correspondant à chacune des heures de la journée, la valeur de l'angle du soleil perçue par chacun de ces sous-modèles ne devrait changer que très lentement. Par conséquent, nos attentes au préalable étaient que cette composante n'ait que peu ou même aucune influence sur les résultats, et qu'elle serait essentiellement redondante avec la composante de la saisonnalité annuelle.

Une seule variante de la composante \mathbb{M}^S est considérée, utilisant l'équation (3.10) :

$$\mathbb{M}^S = \tau_S^{[t]} \otimes s_i(S_t) \quad (3.10)$$

où, $s = \{\text{CLRAMP}_{-90,-20.4}^-, \text{CLRAMP}_{-20.4,90}^+, \text{CLRAMP}_{20.4,90}^+\}$ est une spécification linéaire par morceaux modélisant les effets de l'angle du soleil sur la consommation électrique, dont les intervalles ont été sélectionnés grâce à la segmentation linéaire des effets de l'angle du soleil présentée dans la section 1.3.4; et $\tau_{S,i}^{[t]}$ pour $s_i \in s$ sont les coefficients d'entraînement. Cette composante a un nombre de paramètres égal à $|s|$.

L'identifiant de la composante de l'angle du soleil à Montréal est donc la suivante :

- (a) \mathbb{M}_1^S est donnée par l'équation (3.10). Cette composante a 3 paramètres.

3.1.8 Heure d'été DST

Bien que nous ne nous attendions pas à ce que l'ajout de l'heure d'été apporte des changements significatifs aux résultats du modèle, celle-ci a le potentiel de tracer une séparation claire entre les périodes d'été et d'hiver, et devrait de plus rendre les prévisions plus fiables lors des dates de transition.

L'équation (3.11) présente une composante d'heure d'été trivialement adaptée de

la modélisation des variables exogènes de Clements *et al.* (2016) :

$$\mathbb{DST} = \sum_{s \in \mathcal{O}} \delta_s^{[t]} H_{t-s} = \sum_{s \in \mathcal{O}} \delta_s^{[t]} \mathbb{1}_{\{t-s \in \mathbb{DST}\}} \quad (3.11)$$

où \mathbb{DST} est l'ensemble des journées qui sont durant l'heure d'été (DST); $H_t = \mathbb{1}_{\{t \in \mathbb{DST}\}}$ est la série temporelle de l'heure d'été, où $\mathbb{1}_{\{t \in \mathbb{DST}\}}$ est la variable indicatrice qui est égale à 1 si $t \in \mathbb{DST}$ et égale à 0 sinon; et $\delta_s^{[t]}$, pour $s \in \mathcal{O}$, sont les paramètres d'apprentissage. Cette modélisation de la composante a un nombre de paramètres égal à $|\mathcal{O}|$. Elle est équivalente à une modélisation qui utiliserait $\Delta_s H_t$, la différentiation saisonnière de premier ordre et de période s , plutôt que H_{t-s} , pour les données ayant un décalage dans le temps.

L'identifiant de la composante de l'heure d'été est le suivant :

- (a) \mathbb{DST}_1 est donnée par l'équation (3.11). Cette composante a 3 paramètres.

3.2 Cadre expérimental

Le modèle et ses variantes, décrites dans la section 3.1, ont été implémentés dans un prototype C++ basé sur la bibliothèque *XTensor*, un cadriciel de calcul matriciel open source inspiré de *NumPy* (Mabille *et al.*, 2016).

En plus d'identifier les composantes à utiliser pour un modèle donné, notre interface de ligne de commande permet de spécifier le nombre maximum d'itérations autorisées pendant l'entraînement pour la convergence de l'algorithme (par défaut : 100) et le seuil acceptable d'erreur de convergence numérique (par défaut : 10^{-8}).

La fonction objective utilisée lors de l'entraînement est la méthode des moindres carrés ordinaires itérés. L'entraînement d'un modèle pour une date précise se fait très rapidement, quels que soient la date et les paramètres sélectionnés. Cela prend quelques secondes sur un ordinateur portable standard. Cependant, l'évaluation du

score MAPE pour un modèle est plus lente. L'évaluation de plusieurs variantes du modèle, sur une période de 7 ans, peut prendre quelques heures si elle est calculée séquentiellement. Puisqu'une grappe de calcul haute performance est disponible à l'IREQ, nous avons introduit les parallélisations simples suivantes :

- Pour chaque pas de temps d'un jour donné, le modèle peut être entraîné indépendamment. Seules les prévisions doivent être calculées séquentiellement, mais comme l'apprentissage est beaucoup plus long que le processus de prévision, un temps de calcul important peut ainsi être gagné.
- Les évaluations des modèles peuvent également être réparties entre les jours de la plage d'évaluation.
- Enfin, chaque variation de modèle peut également être entraînée indépendamment.

En exploitant ces observations, nous avons pu réduire le temps de calcul de chaque étape à une durée entre 5 et 10 minutes.

3.3 Méthode d'évaluation

Afin de trouver le meilleur ensemble de paramètres possible, nous procédons comme suit. Tout d'abord, nous commençons avec le modèle de base présenté à l'équation (3.1) inspiré de Clements *et al.* (2016), avec pour seules composantes les composantes de saisonnalités hebdomadaire et annuelle triviales, $\mathbb{S}^{\text{sem}} = \mathbb{S}_0^{\text{sem}}$ et $\mathbb{S}^{\text{année}} = \mathbb{S}_0^{\text{année}}$.

Nous déterminons premièrement quelle stratégie de prétraitement des données a le plus de potentiel en les évaluant toutes deux sur le modèle de base. Nous entraînons les deux variantes de modèles sur les données entre 2013 et 2018, puis nous évaluons chaque journée de 2019. Nous sélectionnons la stratégie qui produit le meilleur score MAPE pour l'année 2019, et c'est cette stratégie qui est utilisée

pour tous les modèles suivants.

Nous évaluons ensuite séquentiellement les effets des variantes des composantes endogènes de saisonnalités, \mathbb{S}^{sem} , $\mathbb{S}^{\text{année}}$ et \mathbb{S}^{jour} . Nous évaluons toutes les variantes de \mathbb{S}^{sem} et sélectionnons la variante qui produit le meilleur score MAPE. Nous y ajoutons ensuite chacune des variantes de $\mathbb{S}^{\text{année}}$ et sélectionnons la meilleure variante, puis nous passons aux variantes de la composante suivante, \mathbb{S}^{jour} , et ainsi de suite, à la manière d'un algorithme glouton.

Une fois que nous avons déterminé quelles variantes de composantes saisonnières forment le meilleur modèle, nous utilisons ce dernier comme base pour évaluer le deuxième groupe de composantes, c'est-à-dire les composantes exogènes principales ou composantes météorologiques, \mathbb{M}^T , \mathbb{M}^N , \mathbb{M}^V et \mathbb{M}^S . Nous procédons en ajoutant les nouvelles composantes séquentiellement de manière gloutonne, similairement au procédé employé pour les composantes saisonnières.

La composante DST n'étant membre à part entière d'aucun des deux groupes mentionnés ci-haut, elle est traitée séparément. Nous évaluons les effets de son ajout ou de son retrait entre chaque groupe.

Cette méthode nous permet d'intégrer rapidement les composantes les plus significatives. Dans le meilleur des cas, ce processus nous permet de ne tester qu'un sous-ensemble de 45 des 587 520 combinaisons de variantes de modèles possibles. En pratique, nous n'éliminons pas la possibilité de retourner sur certaines variantes plus loin dans le processus afin de mieux comprendre les interactions entre les diverses composantes d'un modèle. Il est à noter que cela ne représente pas l'intégralité des modèles qui ont été testés, mais cet ordre de présentation a l'avantage d'être clair et de n'omettre aucun résultat pertinent.

3.3.1 Limitations

En raison d'un changement de portée de l'expérience survenue au cours du projet, nous nous sommes malheureusement trouvés à court de données pour l'ensemble de données d'évaluation. Cet ensemble, qui était supposé être composé des données de 2019, est essentiellement devenu notre ensemble de validation en cours de route. Cette absence d'un ensemble d'évaluation limite notre évaluation de la capacité de généralisation du modèle final. Puisque la validation et l'évaluation sont faites sur le même ensemble de données, le modèle peut posséder un biais de surajustement sur les données de 2019.

3.4 Résultats

Les sections suivantes présentent les évaluations des résultats des variantes du modèles lors de l'ajout des composantes de saisonnalités, météorologiques, et de l'heure d'été. Les résultats du meilleur modèle obtenu sont ensuite analysés plus en détail.

3.4.1 Effets des saisonnalités

La table 3.1 présente le nombre de paramètres ainsi que le score MAPE de modèles ne possédant que les composantes de saisonnalité \mathbb{S}^{sem} , $\mathbb{S}^{\text{année}}$, ou \mathbb{S}^{jour} . Un symbole « + » identifie un modèle ayant le meilleur score MAPE parmi tous les modèles possédant différentes variantes des mêmes composantes ; un symbole « - » identifie un modèle ayant un score MAPE moins bon qu'un autre modèle similaire, mais plus simple, c'est-à-dire que l'une ou plusieurs des composantes du modèle pourraient être retirées pour améliorer le score MAPE ; et le symbole « * » identifie le modèle ayant le meilleur score MAPE parmi tous les modèles sur

le tableau.

TABLEAU 3.1 MAPE globaux de modèles ne possédant que des composantes de saisonnalité

	Modèle	Paramètres	MAPE
+	$(S_0^{\text{sem}}, S_0^{\text{année}})$	$)_1$	5 4,633 2 %
	$(S_0^{\text{sem}}, S_0^{\text{année}})$	$)_2$	5 4,656 1 %
	$(S_2^{\text{sem}}, S_0^{\text{année}})$	$)_1$	6 4,586 2 %
	$(S_7^{\text{sem}}, S_0^{\text{année}})$	$)_1$	11 4,053 5 %
+	$(S_7^{\text{sem}}, S_{1.4}^{\text{année}})$	$)_1$	19 4,029 7 %
	$(S_7^{\text{sem}}, S_{1.6}^{\text{année}})$	$)_1$	23 4,023 0 %
	$(S_7^{\text{sem}}, S_{q 12}^{\text{année}})$	$)_1$	23 4,031 6 %
	$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{001}^{\text{jour}})$	$)_1$	24 2,712 7 %
-	$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{010}^{\text{jour}})$	$)_1$	24 5,285 7 %
	$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{020}^{\text{jour}})$	$)_1$	24 4,006 7 %
	$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{100}^{\text{jour}})$	$)_1$	24 3,639 3 %
	$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{011}^{\text{jour}})$	$)_1$	25 2,684 7 %
	$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{021}^{\text{jour}})$	$)_1$	25 2,726 5 %
	*	$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{101}^{\text{jour}})$	$)_1$
$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{110}^{\text{jour}})$		$)_1$	25 3,533 1 %
$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{120}^{\text{jour}})$		$)_1$	25 3,308 7 %
$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{111}^{\text{jour}})$		$)_1$	26 2,743 1 %
$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{121}^{\text{jour}})$		$)_1$	26 2,477 3 %

Premièrement, les deux stratégies de prétraitement des données sont comparées à l'aide du plus simple modèle possible, le modèle $(S_0^{\text{sem}}, S_0^{\text{année}})$. La première des deux stratégies obtient un score MAPE de 4,63 %, soit un score meilleur que la deuxième stratégie par environ -0,02 %. C'est donc des modèles employant tous la première stratégie qui seront évalués pour tous les tests suivants.

Les deux autres variantes de saisonnalités hebdomadaires sont ensuite testées. Les modèles $(\mathbb{S}_2^{\text{sem}}, \mathbb{S}_0^{\text{année}})$ et $(\mathbb{S}_7^{\text{sem}}, \mathbb{S}_0^{\text{année}})$ obtiennent des scores MAPE qui sont respectivement -0,57 % et -0,58 % meilleurs que le modèle trivial $(\mathbb{S}_0^{\text{sem}}, \mathbb{S}_0^{\text{année}})$. Le modèle $(\mathbb{S}_7^{\text{sem}}, \mathbb{S}_0^{\text{année}})$, dont le score MAPE est de 4,05 %, est sélectionné comme modèle de base pour la suite des tests.

Les variantes de saisonnalités annuelles sont ensuite testées. Les modèles $(\mathbb{S}_7^{\text{sem}}, \mathbb{S}_{1.4}^{\text{année}})$, $(\mathbb{S}_7^{\text{sem}}, \mathbb{S}_{1.6}^{\text{année}})$, et $(\mathbb{S}_7^{\text{sem}}, \mathbb{S}_{q|12}^{\text{année}})$ obtiennent des scores MAPE qui sont respectivement -0,02 %, -0,03 %, et -0,02 % meilleurs que le modèle précédent. Le modèle $(\mathbb{S}_7^{\text{sem}}, \mathbb{S}_{1.6}^{\text{année}})$, dont le score MAPE est de 4,02 %, est sélectionné comme modèle de base pour la suite des tests.

La dernière composante de saisonnalité est celle de la corrélation intrajournalière, celle qui est elle-même formée à partir de jusqu'à trois sous-composantes. La sous-composante de la dernière observation est celle qui a le plus grand effet, améliorant les résultats d'un impressionnant -1,31 % par sa seule présence dans le modèle $(\mathbb{S}_7^{\text{sem}}, \mathbb{S}_{1.6}^{\text{année}}, \mathbb{S}_{001}^{\text{jour}})$. La sous-composante de la moyenne de la dernière journée observée obtient des résultats moins prononcés, améliorant les résultats de -0,38 % dans le modèle $(\mathbb{S}_7^{\text{sem}}, \mathbb{S}_{1.6}^{\text{année}}, \mathbb{S}_{100}^{\text{jour}})$. La combinaison de ces deux sous-composantes améliore les résultats de -1,55 % par rapport au modèle sans aucune composante de corrélation intrajournalière.

La troisième sous-composante possible de la composante de la corrélation intrajournalière est celle de la dernière valeur de prévision. L'implémentation complète de cette sous-composante n'améliore que les résultats de -0,02 % par rapport au modèle sans aucune composante de corrélation intrajournalière, et son implémentation paresseuse empire les résultats de +1,26 %. Malgré la légère amélioration apportée par l'implémentation complète, sa combinaison avec les deux sous-composantes précédentes améliore les résultats de 0,01 % de moins que le

modèle n'ayant que la combinaison des deux sous-composantes précédentes.

Ce résultat était inattendu, compte tenu de l'amélioration rapportée par Clements *et al.* (2016) lors de l'ajout de la sous-composante de la dernière valeur prévue. Une explication possible de la différence entre deux évaluations respective est que le modèle de Clements *et al.* (2016) introduit cette sous-composante beaucoup plus tard dans la construction incrémentale de leur modèle. Puisque cette sous-composante a comme intention d'aller chercher les effets d'inertie de manière indirecte, son introduction plus tardive, ou plus spécifiquement après l'introduction des composantes météorologiques permet d'en tirer plus d'information pertinente que son introduction hâtive.

Le modèle $(\mathbb{S}_7^{\text{sem}}, \mathbb{S}_{1..6}^{\text{année}}, \mathbb{S}_{101}^{\text{jour}})$, dont le score MAPE est de 2,47 %, est donc sélectionné comme modèle de base pour la suite des tests. Nous revenons plus loin sur la possibilité d'inclure la sous-composante de la dernière prévision.

Les graphiques (a) et (b) de la figure 3.1 présentent les scores MAPE horaires de ces variantes de modèles. Le graphique (a) présente les variantes des composantes \mathbb{S}^{sem} et $\mathbb{S}^{\text{année}}$; et le graphique (b) présente certaines variantes de la composante \mathbb{S}^{jour} . La ligne noire pointillée désigne le meilleur modèle trouvé précédant l'ajout de la composante saisonnière d'intérêt. Le symbole \mathbb{S}_* sert à abrégier les composantes saisonnières $(\mathbb{S}_7^{\text{sem}}, \mathbb{S}_{1..6}^{\text{année}}, \mathbb{S}_{101}^{\text{jour}})$.

On peut y voir que l'ajout de la composante de saisonnalité hebdomadaire $\mathbb{S}_7^{\text{sem}}$ permet de grandement réduire les erreurs au niveau du sommet de consommation matinal, car il s'agit du moment de la journée où la consommation est la plus influencée par le jour de la semaine. Les effets de $\mathbb{S}_{1..4}^{\text{année}}$, $\mathbb{S}_{1..6}^{\text{année}}$, ou $\mathbb{S}_{q|12}^{\text{année}}$ sont très similaires. Par rapport à $\mathbb{S}_0^{\text{année}}$, ils dégradent la qualité des résultats au début de la journée jusqu'au sommet matinal, mais les améliorent pour tout le reste de la journée. Du côté de la corrélation intrajournalière \mathbb{S}^{jour} , la sous-composante

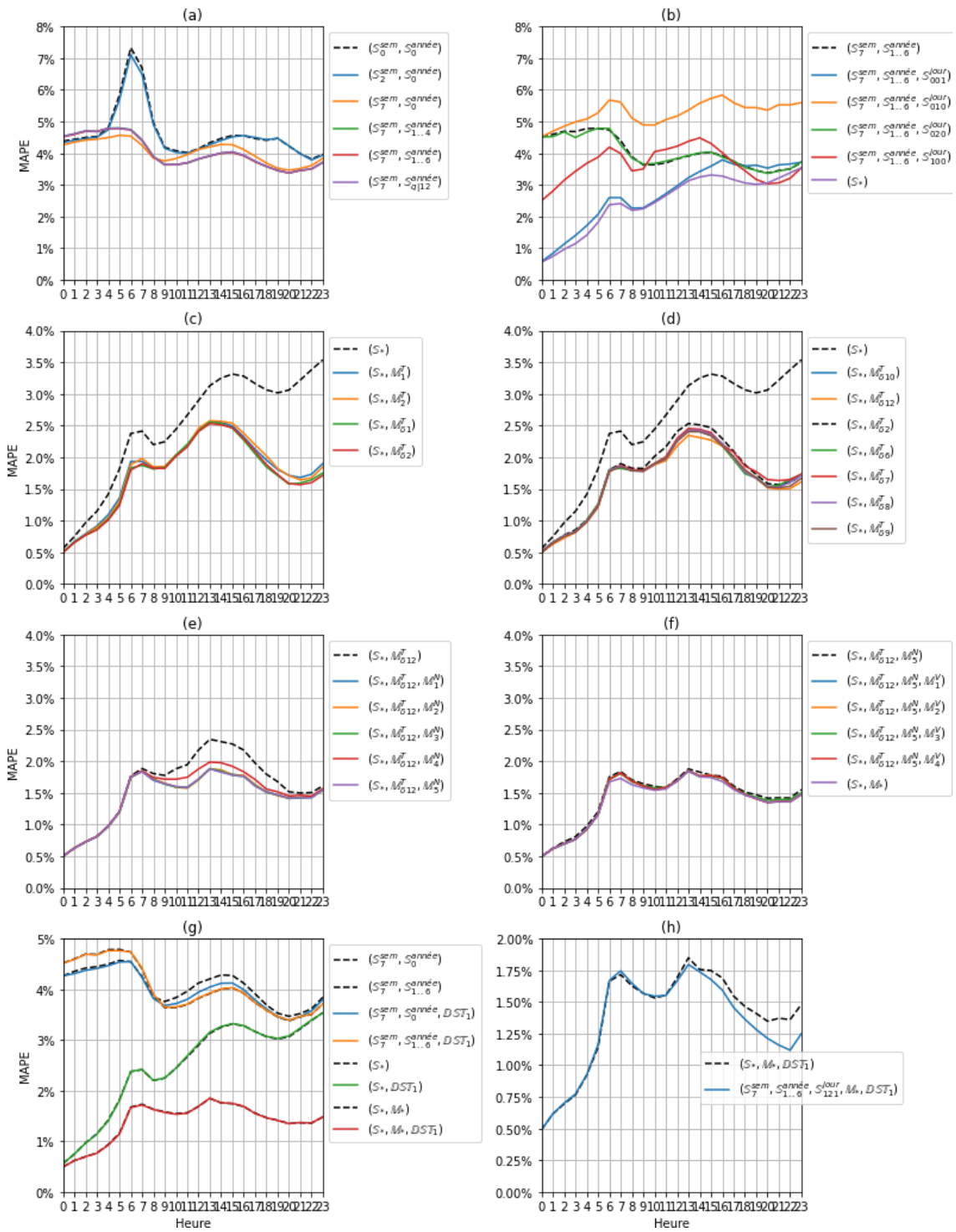


FIGURE 3.1 MAPE horaires des variantes de modèles

de la dernière observation s_1 améliore grandement les résultats jusqu'à environ 16 heures, où les résultats se mettent à ressembler à ceux du modèle précédent. La sous-composante de la moyenne de la dernière journée observée s_3 a un effet similaire à s_1 , mais atténué jusqu'en fin de journée, où elle se met à donner de meilleurs résultats. La combinaison de ces deux sous-composantes permet de tirer avantage de chacune d'entre elles sans en prendre les inconvénients.

La sous-composante de la dernière prévision évaluée de manière paresseuse s_2 obtient des résultats strictement inférieurs à ceux du modèle précédent, alors que la version implémentée au complet (non-paresseuse) s'_2 obtient que des résultats très similaires à ceux du modèle précédent.

3.4.2 Effets météorologiques

Le tableau 3.2 présente le nombre de paramètres ainsi que le score MAPE du modèle $(S_7^{\text{sem}}, S_{1..6}^{\text{année}}, S_{101}^{\text{jour}})_1$, le modèle obtenant les meilleurs résultats parmi ceux considérés précédemment, augmenté de diverses combinaisons de composantes météorologiques.

Parmi toutes les variantes de la composante de température ne prenant en compte que les températures à la station de Montréal (soient M_1^T , M_2^T , $M_{\delta 1}^T$, et $M_{\delta 2}^T$), l'ajout des variantes qui utilisent l'équation (3.7) sont celles qui apportent les plus grandes améliorations, -0,76 % et -0,77 % (contre -0,71 % et -0,70 %), aux résultats du modèle précédent. Les variantes qui utilisent l'équation (3.7) sont celles qui cherchent à réduire la dépendance entre les valeurs des variables de même type, ce qui suggère qu'il s'agit d'une stratégie pertinente à utiliser.

L'ajout de stations météorologiques additionnelles permet d'améliorer encore plus ces résultats. Le modèle $(S_7^{\text{sem}}, S_{1..6}^{\text{année}}, S_{101}^{\text{jour}}, M_{\delta 12}^T)$, avec un score MAPE de 1,61 %, qui représente une amélioration de -0,86 % par rapport au modèle sans compo-

TABLEAU 3.2 MAPE globaux du meilleur modèle à composantes de saisonnalité avec ajout de composantes météorologiques

	Modèle	Paramètres	MAPE
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{1}^T)$	$)_1$	37 1,765 5 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{2}^T)$	$)_1$	37 1,766 2 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{3}^T)$	$)_1$	85 1,703 5 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{4}^T)$	$)_1$	85 1,685 0 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 1}^T)$	$)_1$	33 1,711 1 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 2}^T)$	$)_1$	33 1,704 8 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 3}^T)$	$)_1$	61 1,663 6 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 11}^T)$	$)_1$	65 1,635 1 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 4}^T)$	$)_1$	57 1,651 5 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 5}^T)$	$)_1$	57 1,642 9 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 6}^T)$	$)_1$	57 1,651 7 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 7}^T)$	$)_1$	61 1,677 2 %
+	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 12}^T)$	$)_1$	65 1,614 8 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 8}^T)$	$)_1$	57 1,648 7 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 9}^T)$	$)_1$	57 1,638 3 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 10}^T)$	$)_1$	57 1,650 3 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 12}^T, M_{1}^N)$	$)_1$	90 1,431 5 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 12}^T, M_{2}^N)$	$)_1$	85 1,433 5 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 12}^T, M_{3}^N)$	$)_1$	80 1,431 3 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 12}^T, M_{4}^N)$	$)_1$	80 1,486 8 %
+	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 12}^T, M_{5}^N)$	$)_1$	85 1,430 5 %
+	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 12}^T, M_{5}^N, M_{1}^V)$	$)_1$	100 1,392 9 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 12}^T, M_{5}^N, M_{2}^V)$	$)_1$	100 1,394 2 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 12}^T, M_{5}^N, M_{3}^V)$	$)_1$	100 1,400 1 %
	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 12}^T, M_{5}^N, M_{4}^V)$	$)_1$	105 1,393 9 %
*	$(S_{7}^{sem}, S_{1.6}^{année}, S_{101}^{jour}, M_{\delta 12}^T, M_{5}^N, M_{1}^V, M_{1}^S)$	$)_1$	103 1.377 6 %

sante de température, est le modèle vainqueur de cette catégorie. Il s'agit de celui qui utilise une spécification linéaire par morceaux mettant plus d'accent sur la zone froide et qui effectue une différenciation saisonnière sur toutes les valeurs températures ayant un décalage temporel.

Les composantes de nébulosité et de vitesse du vent ne sont qu'une question de sélection d'une spécification linéaire par morceaux. Pour la composante de nébulosité, la meilleure variante est M_5^N , avec un score MAPE de 1,43 %, ce qui représente une amélioration de -0,18 % par rapport au modèle (S_7^{sem} , $S_{1..6}^{\text{année}}$, S_{101}^{jour} , $M_{\delta_{12}}^T$). La composante de vitesse du vent M_1^V y apporte une amélioration supplémentaire de -0,04 %, résultant en un score MAPE de 1,39 %.

Finalement, la composante de l'angle du soleil M_1^S permet d'y apporter une amélioration d'encore -0,02 %, pour un score MAPE de 1,38 %. Il est possible que l'inclusion de ce paramètre – à première vue très similaire à la saisonnalité annuelle – permette dans les faits à la composante de saisonnalité annuelle d'aller chercher plus de nuances et déchargeant une partie de sa tâche sur la composante de l'angle du soleil.

Les graphiques (c) à (f) de la figure 3.1 présentent les scores MAPE horaires de ces variantes de modèles. Les graphiques (c) et (d) présentent certaines variantes de la composante M^T ; le graphique (e) présente les variantes de la composante M^N ; et le graphique (f) présente les variantes des composantes M^V et M^S . La ligne noire pointillée désigne le meilleur modèle trouvé précédant l'ajout de la composante météorologique d'intérêt. Le symbole M_* sert à abrégier les composantes météorologiques ($M_{\delta_{12}}^T, M_5^N, M_1^V, M_1^S$).

On peut y voir que la composante de température M^T permet d'améliorer graduellement les résultats au fil que la journée avance et que le modèle perd confiance en la composante S^{jour} . La différence entre n'utiliser que la station de la ville de

Montréal ou d'utiliser toutes les stations est relativement serrée, mais tout de même considérable, et la variante $M_{\delta 12}^T$ se démarque particulièrement des autres variantes similaires. La composante de nébulosité M^N permet de particulièrement améliorer les résultats après le sommet matinal et en après-midi, entre 9 heures et 20 heures, lorsqu'il fait encore jour et que la nébulosité peut indirectement indiquer le niveau d'éclairage ou la présence de pluie de manière pertinente. Les composantes de vitesse du vent M^V et de l'angle du soleil par rapport à l'horizon M^S n'ont ensuite qu'un effet beaucoup plus discret, mais apportent tout de même une amélioration significative aux résultats du modèle.

3.4.3 Effets de l'heure d'été

Le tableau 3.3 présente le nombre de paramètres ainsi que le score MAPE d'une sélection de modèles parmi ceux considérés précédemment obtenant les meilleurs résultats, augmentés des variantes de la composante d'heure d'été.

TABLEAU 3.3 MAPE globaux de modèles possédant une composante d'heure d'été

	Modèle	Paramètres	MAPE
+	$(S_7^{\text{sem}}, S_0^{\text{année}}, \text{DST}_1)_1$	14	3,976 0 %
-	$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, \text{DST}_1)_1$	26	4,024 5 %
-	$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{101}^{\text{jour}}, \text{DST}_1)_1$	28	2,475 5 %
*	$(S_7^{\text{sem}}, S_{1.6}^{\text{année}}, S_{101}^{\text{jour}}, M_{\delta 12}^T, M_5^N, M_1^V, M_1^S, \text{DST}_1)_1$	106	1,374 5 %

L'ajout de cette composante apporte une très légère amélioration, d'approximativement -0,003 %, aux résultats du meilleur modèle avec composantes météorologiques. Cependant, si l'on retourne en arrière pour étudier ses effets, on remarque que bien que la composante DST apporte déjà une légère amélioration de -0,008 % aux résultats du modèle $(S_7^{\text{sem}}, S_0^{\text{année}})$, elle apporte en revanche une détérioration de +0,002 % aux résultats du modèle $(S_7^{\text{sem}}, S_{1.6}^{\text{année}})$, et une détérioration de

+0,005 % aux résultats du modèle (S_7^{sem} , $S_{1.6}^{\text{année}}$, S_{101}^{jour}).

Tous ces résultats sont très peu significatifs.

Le graphique (g) de la figure 3.1 présente les scores MAPE horaires de modèles ayant une composante DST. Les lignes noires pointillées désignent les résultats des mêmes modèles avant l'ajout de la composante d'heure d'été.

On peut y voir que la composante d'heure d'été DST a un effet léger mais mitigé, qui dépend du moment où la composante est introduite dans le modèle. Lorsque la composante de saisonnalité annuelle est absente, la composante DST offre des résultats plus intéressants, mais une fois la saisonnalité annuelle introduite dans le modèle, DST cesse de faire une différence significative. La composante DST est donc essentiellement redondante avec celle de la saisonnalité annuelle.

3.4.4 Retour sur la corrélation intrajournalière

Le tableau 3.4 présente le nombre de paramètres ainsi que le score MAPE d'une sélection de modèles augmentés de la sous-composante de dernière prévision intrajournalière.

TABLEAU 3.4 Retour sur les MAPE globaux de modèles possédant la sous-composante de dernière prévision intrajournalière

	Modèle	Paramètres	MAPE
+	(S_7^{sem} , $S_{1.6}^{\text{année}}$, S_{121}^{jour} , $M_{\delta 12}^T$, M_5^N , M_1^V , M_1^S) ₁	104	1,326 8 %
*	(S_7^{sem} , $S_{1.6}^{\text{année}}$, S_{121}^{jour} , $M_{\delta 12}^T$, M_5^N , M_1^V , M_1^S , DST ₁) ₁	107	1,320 4 %

Les deux modèles évalués voient une amélioration de -0,05 % de leur score MAPE lorsqu'ils sont augmentés avec la sous-composante de dernière prévision intrajournalière.

Cette différence significative semble confirmer l'hypothèse émise plus tôt, émettant que l'introduction de cette sous-composante à un point plus tardif dans la construction incrémentale du modèle lui permette de tirer une plus grande quantité d'information d'inertie pertinente à partir de la valeur de la dernière prévision.

Le graphique (h) de la figure 3.1 présente les scores MAPE horaires d'un modèle ayant la sous-composante intrajournalière de dernière prévision. La ligne noire pointillée désigne le meilleur modèle trouvé précédemment.

On peut y voir que la composante $\mathbb{S}_{121}^{\text{jour}}$ permet d'améliorer tous les résultats après 14 heures jusqu'à la fin de la journée, un comportement qui n'apparaissait pas dans le graphique (b). Cet effet s'explique par l'information d'inertie plus pertinente obtenue par les composantes additionnelles, et particulièrement par la composante \mathbb{M}^T .

3.4.5 Analyse du meilleur modèle

Le meilleur modèle parmi ceux évalués ci-haut est le modèle

$$(\mathbb{S}_7^{\text{sem}}, \mathbb{S}_{1..6}^{\text{année}}, \mathbb{S}_{121}^{\text{jour}}, \mathbb{M}_{\delta 12}^T, \mathbb{M}_5^N, \mathbb{M}_1^V, \mathbb{M}_1^S, \text{DST}_1)_1 \quad (3.12)$$

avec 107 paramètres, pour un score MAPE de 1,32 % sur les données de l'année 2019. Les scores MAPE de ce modèle pour chaque jour de la semaine sont présentés dans la figure 3.2 et détaillés dans le tableau 3.5. Chaque score MAPE est donné en pourcentage.

On y voit que les prévisions sont généralement meilleures les mardis, mercredis, jeudis et vendredis. Une étude des coefficients d'entraînement révèle que ces jours de la semaine ont un profil relativement similaire. Les modèles ont plus de difficulté à estimer la consommation lors des pics de l'heure de pointe du matin les lundis et mardis que les autres jours de la semaine. Ce résultat s'explique probablement

par le fait que de nombreux jours spéciaux tombent spécifiquement sur un lundi. L'heure de pointe d'un lundi férié a une consommation réduite par rapport à la normale, et elle a alors tendance à être compensée par une heure de pointe plus prononcée le jour suivant, un mardi. Les heures calmes de la nuit, d'environ 20 heures à 5 heures, obtiennent des meilleures prévisions que les heures où il y a une forte activité humaine, d'environ 6 heures à 18 heures, peu importe le jour de la semaine.

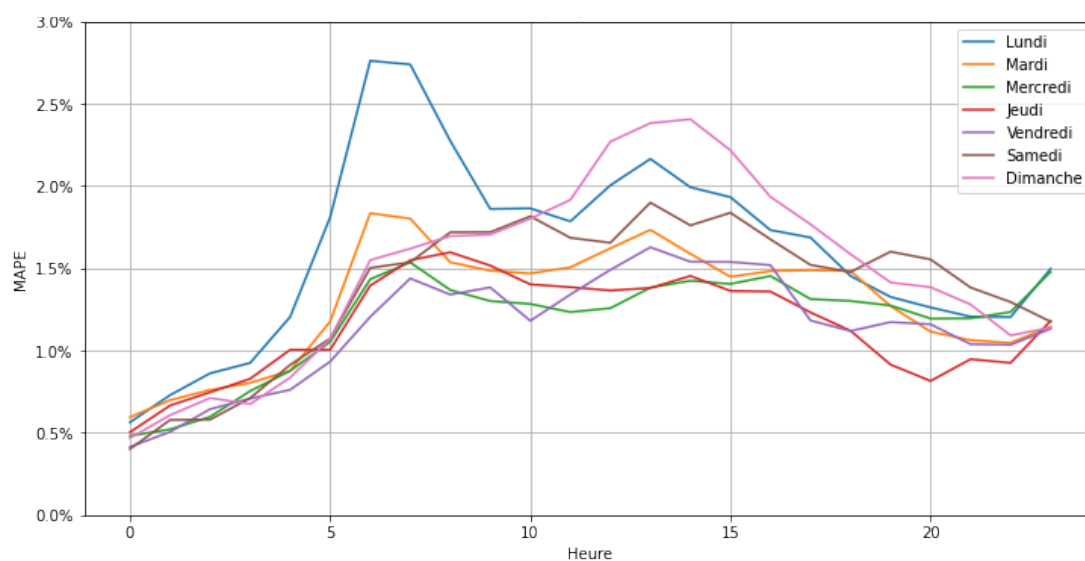


FIGURE 3.2 MAPE horaires du meilleur modèle par jour de la semaine

Le score MAPE obtenu par ce modèle est meilleur que celui du modèle de référence présenté par Clements *et al.* (2016) avec une différence d'environ -0.04 %. Il est par ailleurs meilleur que le modèle proposé par Cancelo *et al.* (2008) (différence de -0.24 %), mais moins bon que celui de Marchand et Zinflou (2021), entraîné sur le même ensemble de données (différence de +0.61 %). La figure 3.3 présente les scores MAPE horaires du meilleur modèle comparé aux résultats de Cancelo *et al.* (2008), Clements *et al.* (2016), et Marchand et Zinflou (2021). Malgré leur

TABLEAU 3.5 MAPE horaires du meilleur modèle par jour de la semaine

Heure	Tous les jours	Lun	Mar	Mer	Jeu	Ven	Sam	Dim
0 : 00	,49 %	,56	,60	,48	,50	,41	,40	,47
1 : 00	,62 %	,73	,70	,52	,67	,51	,58	,61
2 : 00	,70 %	,86	,76	,60	,75	,64	,58	,71
3 : 00	,77 %	,92	,80	,75	,83	,71	,71	,67
4 : 00	,92 %	1,20	,88	,88	1,00	,76	,91	,84
5 : 00	1,16 %	1,81	1,18	1,05	1,00	,93	1,07	1,07
6 : 00	1,67 %	2,76	1,83	1,43	1,39	1,20	1,50	1,55
7 : 00	1,75 %	2,74	1,80	1,54	1,55	1,44	1,54	1,62
8 : 00	1,65 %	2,27	1,54	1,37	1,60	1,34	1,72	1,70
9 : 00	1,57 %	1,86	1,48	1,30	1,52	1,38	1,72	1,70
10 : 00	1,55 %	1,86	1,47	1,28	1,40	1,18	1,82	1,80
11 : 00	1,55 %	1,78	1,51	1,23	1,38	1,34	1,68	1,91
12 : 00	1,67 %	2,00	1,62	1,26	1,36	1,49	1,65	2,27
13 : 00	1,80 %	2,16	1,73	1,38	1,38	1,63	1,90	2,38
14 : 00	1,74 %	1,99	1,59	1,42	1,45	1,54	1,76	2,40
15 : 00	1,68 %	1,93	1,45	1,40	1,36	1,54	1,84	2,22
16 : 00	1,59 %	1,73	1,48	1,45	1,36	1,52	1,68	1,93
17 : 00	1,46 %	1,69	1,49	1,31	1,23	1,18	1,52	1,77
18 : 00	1,36 %	1,45	1,48	1,30	1,12	1,12	1,48	1,59
19 : 00	1,28 %	1,33	1,27	1,27	,91	1,17	1,60	1,41
20 : 00	1,21 %	1,26	1,11	1,19	,82	1,16	1,55	1,38
21 : 00	1,16 %	1,21	1,06	1,20	,95	1,04	1,38	1,28
22 : 00	1,12 %	1,20	1,05	1,23	,93	1,03	1,29	1,09
23 : 00	1,25 %	1,50	1,14	1,48	1,18	1,13	1,18	1,14
Toutes heures	1,32 %	1,62	1,29	1,18	1,15	1,14	1,38	1,48

score MAPE global impeccable et variant très peu sur tous les pas de temps de la journée, nous obtenons de meilleures performances que le modèle de Marchand et Zinflou (2021) pendant les quatre premiers pas de temps de l’horizon. Ces quatre pas de temps, ainsi que le dernier pas de temps, sont par ailleurs les pas de temps où ils obtiennent les pires résultats, alors que leurs performances lors des heures de pointe sont parmi les meilleures, ce qui est tout le contraire de notre modèle.

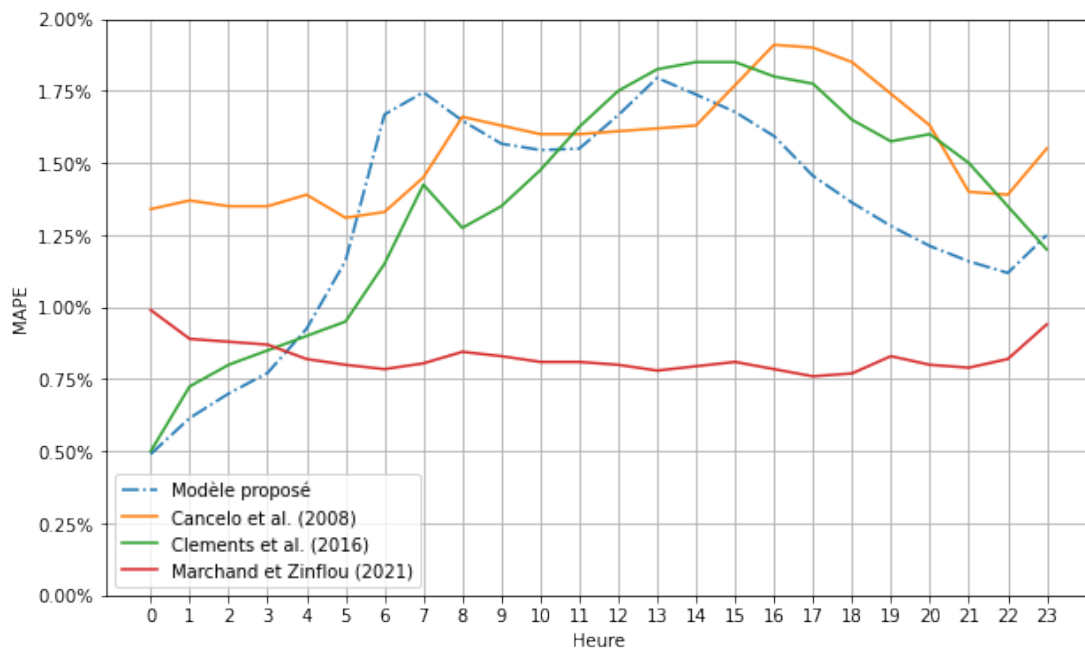


FIGURE 3.3 MAPE horaires du modèle proposé comparé aux résultats de Cancelo *et al.* (2008), Clements *et al.* (2016), et Marchand et Zinflou (2021)

Les « larges erreurs », définies ici comme étant des erreurs de plus de 5 % – une valeur arbitraire mais suggérée comme point de référence adéquat dans la littérature (Ranaweera *et al.*, 1997; Cancelo *et al.*, 2008) –, sont particulièrement intéressantes, car elles permettent d’évaluer si un modèle a tendance à faire des erreurs particulièrement coûteuses à la compagnie de production d’électricité. Notre modèle ne produit de telles erreurs que pour deux journées en 2019 : le lundi 24

juin (5,03 %) et le vendredi 1^{er} novembre (5,32 %).

Ces résultats sont cependant atteints sans même introduire de composante de jours spéciaux, une composante qui est pourtant jugée importante dans le domaine de la prévision de la consommation électrique à court terme (Cancelo *et al.*, 2008; Clements *et al.*, 2016; Nassif *et al.*, 2022).

3.5 Jours spéciaux

Malheureusement, il n'est pas toujours évident de déterminer en avance quelle journée est vraiment spéciale et à quel point elle va se révéler spéciale. Même les fêtes annuelles sont difficiles à étudier. Dans un ensemble de données comme le nôtre, ayant 8 années de données, ne possède que 8 instances de chaque fête. Puisque l'effet d'une fête sur la consommation peut être particulier au jour de la semaine sur lequel elle tombe (Cancelo *et al.*, 2008; López *et al.*, 2022), 8 années de données peuvent être insuffisantes pour observer chaque possibilité ne serait-ce qu'une seule fois.

La sélection et le traitement des jours spéciaux étant un sujet assez large pour pouvoir former un projet en entier, ce projet-ci ne l'a jusqu'ici pas abordé. Nous démontrons dans cette section que son introduction est prometteuse pour améliorer les prévisions obtenues et permet de surpasser les résultats du modèle de référence.

3.5.1 Preuve de concept

Nous introduisons deux variantes d'une composante de jours spéciaux \mathbb{J} . La composante \mathbb{J} utilise l'équation (3.13), qui est très similaire à celle de Clements *et al.*

(2016).

$$\sum_{s \in \mathcal{O}} \sum_{j \in \mathcal{J}} \sigma_{j,s}^{[t]} \mathbb{1}_{\{t-s \in j\}} \quad (3.13)$$

où \mathcal{J} est l'ensemble des types de jours fériés; $\sigma_{j,s}^{[t]}$ pour $s \in \mathcal{O}$ et $j \in \mathcal{J}$ sont les coefficients d'entraînement. Cette composante a un nombre de paramètres égal à $|\mathcal{O}| \times |\mathcal{J}|$.

La différence entre les deux variantes réside principalement dans leur sélection de l'ensemble \mathcal{J} . La première variante, plus simple, définit simplement $\mathcal{J}_1 = \{\{\text{py-holidays}\}\}$, où $\{\text{py-holidays}\}$ est l'ensemble de tous les jours fériés reconnus par la librairie *Python-Holidays* (Montel et al., 2022) pour la province de Québec. Tous les jours fériés sont donc considérés comme étant d'un seul et même type.

La deuxième variante définit \mathcal{J}_2 grâce à une sélection de jours mal évalués, regroupés de manière mutuellement exclusive par profils de consommation similaires. La méthode utilisée pour cette sélection pourrait être qualifiée de similaire à une version simplifiée de la méthode de classification automatique des jours spéciaux de López *et al.* (2022). Comme dans López *et al.* (2022), une régression linéaire est appliquée sur les entrées de notre meilleur modèle afin d'identifier les journées qui s'éloignent le plus de la régression année après année. Ces journées aberrantes deviennent donc des candidats de jours spéciaux, et leurs profils sont étudiés individuellement pour chaque jour de la semaine.

Pour cette classification, nous avons présélectionné tous les lundis fériés (le lundi de Pâques, la fête du Travail, la Journée nationale des patriotes et l'Action de grâces), le Vendredi saint, et la journée suivant chacune de ces fêtes comme candidats. Nous avons ensuite laissé la régression choisir huit autres journées qu'elle considère comme aberrantes. En ordre, les journées sélectionnées par cette méthode sont le 25 décembre, le 24 juin, le 1^{er} janvier, le 24 décembre, le 1^{er} juillet,

le 31 décembre, le 2 janvier, et le 25 juin. On y retrouve donc les fêtes de Noël (24 et 25 décembre), la Fête nationale du Québec (24 et 25 juin), le Nouvel An (31 décembre, 1^{er} et 2 janvier), et la fête du Canada (1^{er} juillet).

La figure 3.4 présente les coefficients de régression de ces journées. Chacun de ces neuf graphiques présente une classe ou un type de jours spéciaux. Les lettres L (lundi), M (mardi), W (mercredi), J (jeudi), V (vendredi), S (samedi) et D (dimanche) identifient le jour de la semaine pour les fêtes ayant une date fixe, et « +0j » et « +1j » identifient un décalage temporel par rapport à une fête ayant un jour de la semaine fixe.

Les identifiants de composantes des jours spéciaux possibles sont donc les suivants :

- (a) \mathbb{J}_1 est donnée par l'équation (3.13), où $\mathcal{J} = \mathcal{J}_1$. Cette composante a 3 paramètres.
- (b) \mathbb{J}_2 est donnée par l'équation (3.13), où $\mathcal{J} = \mathcal{J}_2$; et $\mathcal{O} = \{0\}$, puisque les groupes choisis prennent déjà le décalage temporel en compte. Cette composante a 9 paramètres.

3.5.2 Résultats préliminaires

Le tableau 3.6 présente le nombre de paramètres ainsi que le score MAPE d'une sélection de modèles parmi ceux considérés précédemment obtenant les meilleurs résultats, augmentés des variantes préliminaires de la composante de jours spéciaux.

La variante \mathbb{J}_1 améliore le score MAPE du meilleur modèle trouvé précédemment de -0,06 % pour un score final de 1,26 %, et \mathbb{J}_2 l'améliore de -0,07 % pour un score final de 1,25 %. Lorsqu'on joue avec de si petits scores MAPE, cette différence est bien significative.

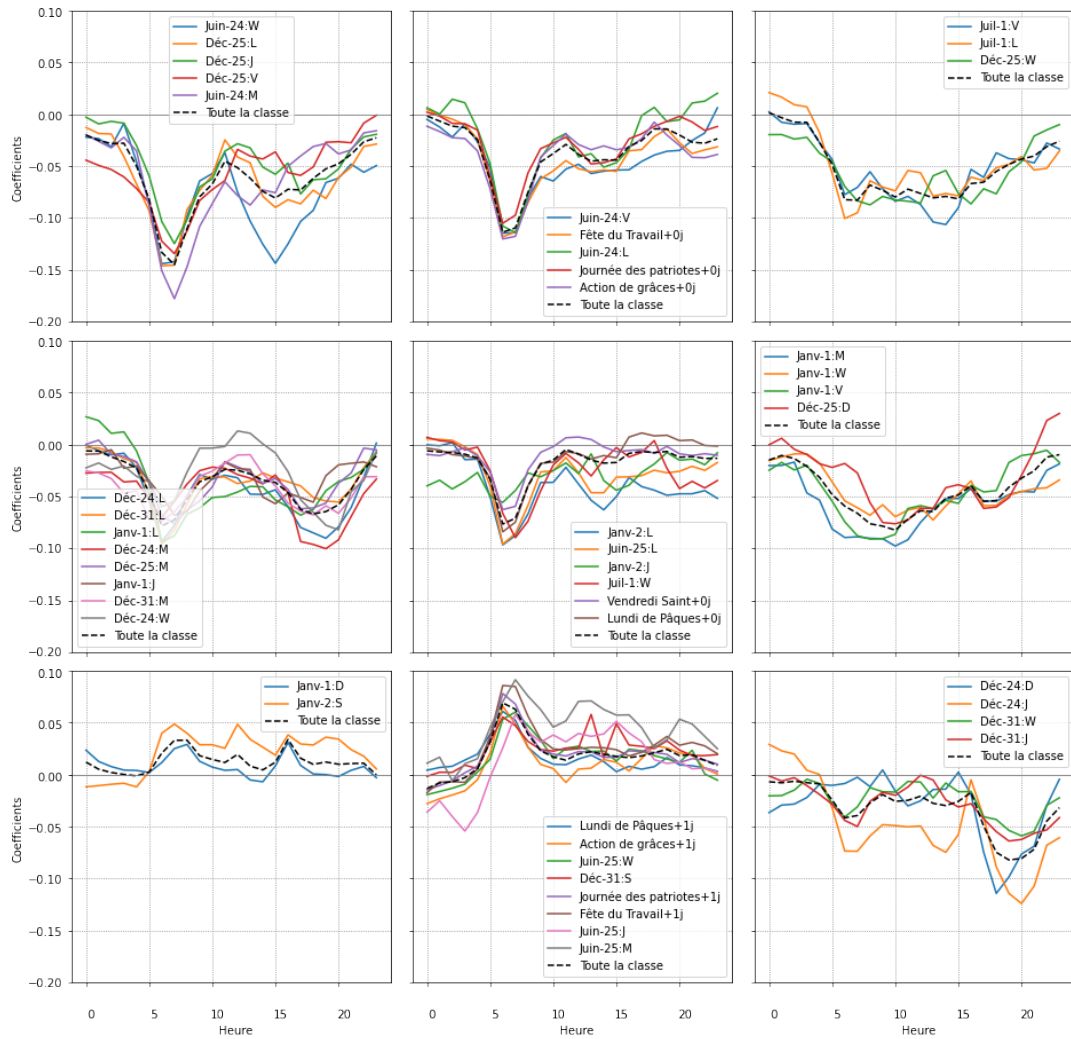


FIGURE 3.4 Profils des coefficients des groupes de jours spéciaux selon \mathbb{J}_2

Les scores MAPE du meilleur modèle augmenté de \mathbb{J}_2 pour chaque jour de la semaine sont présentés dans la figure 3.5. On y voit comment la prise en compte des jours spéciaux permet d'obtenir des résultats similairement bons peu importe le jour de la semaine. En particulier, il est clair que les prévisions de la consommation lors des pics de l'heure de pointe du matin les lundis et mardis, qui étaient plus tôt parmi les plus grandes sources d'erreurs de prévisions, sont maintenant beaucoup plus précises. Il ne fait aucun doute qu'une étude plus approfondie des profils

TABLEAU 3.6 MAPE globaux de modèles possédant une composante de jours spéciaux

Modèle	Param.	MAPE
$(S_7^{\text{sem}}, S_{1..6}^{\text{année}}, S_{101}^{\text{jour}}, J_1)_1$	28	2,3975 %
+ $(S_7^{\text{sem}}, S_{1..6}^{\text{année}}, S_{101}^{\text{jour}}, J_2)_1$	34	2,3841 %
$(S_7^{\text{sem}}, S_{1..6}^{\text{année}}, S_{101}^{\text{jour}}, M_{\delta 12}^T, M_5^N, M_1^V, M_1^S, J_1)_1$	106	1,3133 %
$(S_7^{\text{sem}}, S_{1..6}^{\text{année}}, S_{121}^{\text{jour}}, M_{\delta 12}^T, M_5^N, M_1^V, M_1^S, J_1)_1$	107	1,2608 %
$(S_7^{\text{sem}}, S_{1..6}^{\text{année}}, S_{101}^{\text{jour}}, M_{\delta 12}^T, M_5^N, M_1^V, M_1^S, J_2)_1$	112	1,2998 %
+ $(S_7^{\text{sem}}, S_{1..6}^{\text{année}}, S_{121}^{\text{jour}}, M_{\delta 12}^T, M_5^N, M_1^V, M_1^S, J_2)_1$	113	1,2495 %
$(S_7^{\text{sem}}, S_{1..6}^{\text{année}}, S_{101}^{\text{jour}}, M_{\delta 12}^T, M_5^N, M_1^V, M_1^S, \text{DST}_1, J_1)_1$	109	1,3118 %
$(S_7^{\text{sem}}, S_{1..6}^{\text{année}}, S_{121}^{\text{jour}}, M_{\delta 12}^T, M_5^N, M_1^V, M_1^S, \text{DST}_1, J_1)_1$	110	1,2571 %
$(S_7^{\text{sem}}, S_{1..6}^{\text{année}}, S_{101}^{\text{jour}}, M_{\delta 12}^T, M_5^N, M_1^V, M_1^S, \text{DST}_1, J_2)_1$	115	1,2984 %
* $(S_7^{\text{sem}}, S_{1..6}^{\text{année}}, S_{121}^{\text{jour}}, M_{\delta 12}^T, M_5^N, M_1^V, M_1^S, \text{DST}_1, J_2)_1$	116	1,2457 %

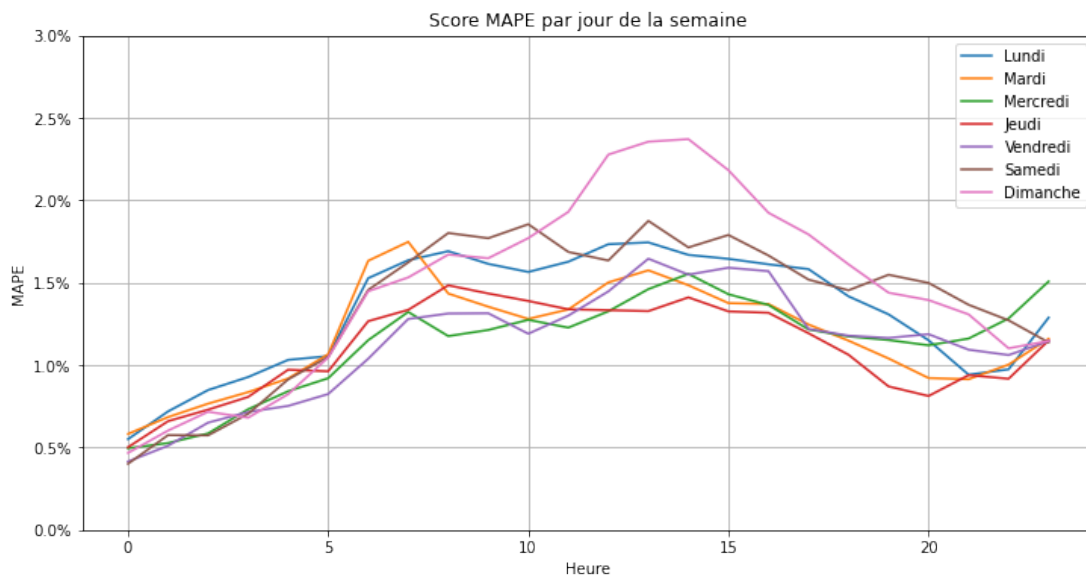


FIGURE 3.5 MAPE horaires par jour de la semaine pour un modèle avec une composante de jours spéciaux

des jours spéciaux au Québec permettrait d'améliorer encore plus les résultats du modèle.

CONCLUSION

Dans ce mémoire, nous avons présenté un modèle de séries temporelles à équations multiples pour traiter la prévision de la consommation électrique à court terme dans la province de Québec. Notre modèle est inspiré du modèle de Clements *et al.* (2016), choisi en raison de ses bons résultats ainsi que son interprétabilité. Nous l'avons adapté aux enjeux spécifiques au Québec, tels que l'influence accrue des températures hivernales, et l'avons étendu à l'aide de nouvelles composantes. Pour ce faire, nous avons étudié un ensemble de données contenant de multiples attributs pouvant influencer la consommation électrique au Québec à l'aide de plusieurs techniques du domaine de l'analyse ou de la prévision des séries temporelles.

Nous avons étudié l'inclusion de trois variables météorologiques (température, nébulosité et vitesse du vent), ainsi que l'inclusion de données provenant de multiples stations météorologiques. Toutes ces variables se sont avérées améliorer de manière importante la précision du modèle de prévision. Nous avons de plus introduit des composantes pour prendre en charge l'angle du soleil par rapport à l'horizon et le passage à l'heure d'été. Nous avons démontré qu'il n'est pas nécessaire de calculer une moyenne pondérée des variables de même type afin d'éviter toute dépendance entre les entrées du modèle, et qu'un modèle à équations multiples comme celui que nous avons présenté peut très bien converger malgré cela. Il est tout de même préférable de réduire la dépendance à l'aide de différentiations saisonnières lorsqu'applicable pour obtenir de meilleurs résultats.

Le modèle proposé a obtenu un score MAPE global de 1,32 % sur les données de

l'année 2019. Ce score est meilleur que celui du modèle de référence et compétitif en comparaison aux autres modèles proposés dans le domaine de la prévision de la demande à court terme à ce jour.

Une fois qu'une analyse approfondie sur les jours qui peuvent être considérés comme étant spéciaux au Québec et leurs profils particuliers sera effectuée et qu'une composante de jours spéciaux robuste sera fin prête à être ajoutée au modèle, il ne serait pas étonnant de le voir atteindre des résultats parmi les plus compétitifs. Notre modèle préliminaire obtient déjà un score MAPE global de 1,25 % sur l'ensemble de validation. Nos critères de sélection pour les méthodes de modélisation des composantes – comme la simplicité et la facilité d'interprétation – pourraient être élargis afin de tester des méthodes alternatives. Par exemple, nous pourrions permettre aux points de ruptures des spécifications linéaires par morceaux d'être des paramètres déterminés par l'apprentissage plutôt que des composantes sélectionnées a priori, ou utiliser des splines polynomiaux plutôt que linéaires. Une autre piste d'amélioration serait de permettre au modèle d'utiliser différentes spécifications pour chaque pas de temps afin de mieux refléter les comportements variés liés aux facteurs d'influence.

RÉFÉRENCES

- Abdel-Aal, R. (2004). Short-term hourly load forecasting using abductive networks. *Power Systems, IEEE Transactions on*, 19, 164 – 173. <http://dx.doi.org/10.1109/TPWRS.2003.820695>
- Alves da Silva, A. P., Ferreira, V. H. et Velasquez, R. M. (2008). Input space to neural network based load forecasters. *International Journal of Forecasting*, 24(4), 616–629.
- Aouad, M., Hajj, H., Shaban, K., Jabr, R. A. et El-Hajj, W. (2022). A cnn-sequence-to-sequence network with attention for residential short-term load forecasting. *Electric Power Systems Research*, 211, 108152.
- Beccali, M., Cellura, M., Lo Brano, V. et Marvuglia, A. (2008). Short-term prediction of household electricity consumption : Assessing weather sensitivity in a mediterranean area. *Renewable and Sustainable Energy Reviews*, 12(8), 2040–2065. <http://dx.doi.org/https://doi.org/10.1016/j.rser.2007.04.010>
- Bedi, J. et Toshniwal, D. (2019). Deep learning framework to forecast electricity demand. *Applied Energy*, 238(C), 1312–1326. <http://dx.doi.org/10.1016/j.apenergy.2019.0>
- Cancelo, J. R. et Espasa, A. (1996). Modelling and forecastng daily series of electricity demand. *Investigacion Economica*, 20, 359–376.
- Cancelo, J. R., Espasa, A. et Grafe, R. (2008). Forecasting the electricity load from one day to one week ahead for the spanish system operator. *International Journal of forecasting*, 24(4), 588–602.

- Cecati, C., Kolbusz, J., Różycki, P., Siano, P. et Wilamowski, B. M. (2015). A novel rbf training algorithm for short-term electric load forecasting and comparative studies. *IEEE Transactions on Industrial Electronics*, *62*(10), 6519–6529. <http://dx.doi.org/10.1109/TIE.2015.2424399>
- Chan, Z. S., Ngan, H., Rad, A. B., David, A. K. et Kasabov, N. K. (2006). Short-term ann load forecasting from limited data using generalization learning strategies. *Neurocomputing*, *70*(1), 409–419. <http://dx.doi.org/https://doi.org/10.1016/j.neucom.2005.12.131>
- Clements, A. E., Hurn, A. et Li, Z. (2016). Forecasting day-ahead electricity load using a multiple equation time series approach. *European Journal of Operational Research*, *251*(2), 522–530.
- Comte, D. M. L. et Warren, H. E. (1981). Modeling the impact of summer temperatures on national electricity consumption. *Journal of Applied Meteorology and Climatology*, *20*(12), 1415–1419. <http://dx.doi.org/10.1175/1520-0450>
- Cottet, R. et Smith, M. (2003). Bayesian modeling and forecasting of intraday electricity load. *Journal of the American Statistical Association*, *98*(464), 839–849.
- Darbellay, G. A. et Slama, M. (2000). Forecasting the short-term demand for electricity : Do neural networks stand a better chance? *International Journal of Forecasting*, *16*(1), 71–83.
- Dolara, A., Grimaccia, F., Leva, S., Mussetta, M. et Ogliari, E. (2015). A physical hybrid artificial neural network for short term forecasting of pv plant power output. *Energies*, *8*(2), 1138–1153. <http://dx.doi.org/10.3390/en8021138>
- Elamin, N. et Fukushige, M. (2018). Modeling and forecasting hourly electricity

- demand by sarimax with interactions. *Energy*, 165. <http://dx.doi.org/10.1016/j.energy.2018.09.157>
- Engle, R. F., Granger, C. W. J., Rice, J. et Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, 81(394), 310–320.
- Espasa, A., Revuelta, J. M. et Cancelo, J. R. (1996). Automatic modelling of daily series of economic activity. Dans A. Prat (dir.). *COMPSTAT*, 51–63., Heidelberg. Physica-Verlag HD.
- Espinoza, M., Joye, C., Belmans, R. et De Moor, B. (2005). Short-term load forecasting, profile identification, and customer segmentation : a methodology based on periodic time series. *IEEE Transactions on Power Systems*, 20(3), 1622–1630. <http://dx.doi.org/10.1109/TPWRS.2005.852123>
- Feng, C., Sun, M. et Zhang, J. (2019). Reinforced deterministic and probabilistic load forecasting via Q -learning dynamic model selection. *IEEE Transactions on Smart Grid*, 11(2), 1377–1386.
- Feng, C. et Zhang, J. (2020). Assessment of aggregation strategies for machine-learning based short-term load forecasting. *Electric Power Systems Research*, 184, 106304.
- Fourier, J.-B. (1808). Mémoire sur la propagation de la chaleur dans les corps solides. *Nouveau Bulletin des sciences par la Société philomatique de Paris*, I(6), 112–116.
- Gould, P. G., Koehler, A. B., Ord, J. K., Snyder, R. D., Hyndman, R. J. et Vahid-Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, 191(1), 207–222. <http://dx.doi.org/https://doi.org/10.1016/j.ejor.2007.08.024>

- Grenier, M. (2006). Short-term load forecasting at hydro-québec transÉnergie. Dans *2006 IEEE Power Engineering Society General Meeting*, 5–pp. IEEE.
- Gs, A. et Shiruru, K. (2016). Short term load forecasting methods, a comparative study. *International Journal of Advance Research and Innovative Ideas in Education*, 1, 31–37.
- Harvey, A. et Koopman, S. J. (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association*, 88(424), 1228–1236.
- Henley, A. et Peirson, J. (1997). Non-linearities in electricity demand and temperature : Parametric versus non-parametric methods. *Oxford Bulletin of Economics and Statistics*, 59(1), 149–62.
- Hernández, L., Baladrón, C., Aguiar, J. M., Calavia, L., Carro, B., Sánchez-Esguevillas, A., Sanjuán, J., González, á. et Lloret, J. (2013). Improved short-term load forecasting based on two-stage predictions with artificial neural networks in a microgrid environment. *Energies*, 6(9), 4489–4507. <http://dx.doi.org/10.3390/en6094489>
- Hippert, H., Bunn, D. et Souza, R. (2005). Large neural networks for electricity load forecasting : Are they overfitted? *International Journal of Forecasting*, 21(3), 425–434. <http://dx.doi.org/https://doi.org/10.1016/j.ijforecast.2004.12.004>
- Hippert, H. S., Pedreira, C. E. et Souza, R. C. (2001). Neural networks for short-term load forecasting : A review and evaluation. *IEEE Transactions on power systems*, 16(1), 44–55.
- Hippert, H. S. et Taylor, J. W. (2010). An evaluation of bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-

- term load forecasting. *Neural Networks*, 23(3), 386–395. <http://dx.doi.org/https://doi.org/10.1016/j.neunet.2009.11.016>
- Hyde, O. et Hodnett, P. F. (1997). An adaptable automated procedure for short-term electricity load forecasting. *IEEE Transactions on Power Systems*, 12, 84–94.
- Hyndman, R. J. et Athanasopoulos, G. (2021). *Forecasting : principles and practice* (3e éd.). Melbourne, Australia : OTexts. Récupéré de <https://otexts.com/fpp3/>
- Joshi, G., Walambe, R. et Kotecha, K. (2021). A review on explainability in multimodal deep neural nets. *IEEE Access*, 9, 59800–59821.
- Kelo, S. et Dudul, S. (2012). A wavelet elman neural network for short-term electrical load prediction under the influence of temperature. *International Journal of Electrical Power & Energy Systems*, 43(1), 1063–1071. <http://dx.doi.org/https://doi.org/10.1016/j.ijepes.2012.06.009>
- Kim, M. S. (2013). Modeling special-day effects for forecasting intraday electricity demand. *European Journal of Operational Research*, 230(1), 170–180. <http://dx.doi.org/https://doi.org/10.1016/j.ejor.2013.03.039>
- Kondaiah, V., Saravanan, B., Sanjeevikumar, P. et Khan, B. (2022). A review on short-term load forecasting models for micro-grid application. *The Journal of Engineering*.
- Kouhi, S. et Keynia, F. (2013). A new cascade nn based method to short-term load forecast in deregulated electricity market. *Energy Conversion and Management*, 71, 76–83. <http://dx.doi.org/https://doi.org/10.1016/j.enconman.2013.03.014>

- Kwiatkowski, D., Phillips, P. C., Schmidt, P. et Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1–3), 159–178. [http://dx.doi.org/https://doi.org/10.1016/0304-4076\(92\)90104-Y](http://dx.doi.org/https://doi.org/10.1016/0304-4076(92)90104-Y)
- Lamedica, R., Prudenzi, A., Sforza, M., Caciotta, M. et Cencelli, V. (1996). A neural network based technique for short-term forecasting of anomalous load periods. *IEEE Transactions on Power Systems*, 11(4), 1749–1756. <http://dx.doi.org/10.1109/59.544638>
- Lauret, P., Fock, E., Randrianarivony, R. N. et Manicom-Ramsamy, J.-F. (2008). Bayesian neural network approach to short time load forecasting. *Energy Conversion and Management*, 49(5), 1156–1166. <http://dx.doi.org/https://doi.org/10.1016/j.enconman.2007.09.009>
- Lee, C.-M. et Ko, C.-N. (2011). Short-term load forecasting using lifting scheme and arima models. *Expert Syst. Appl.*, 38, 5902–5911. <http://dx.doi.org/10.1016/j.eswa.2010.11.033>
- Lin, W., Wu, D. et Boulet, B. (2021). Spatial-temporal residential short-term load forecasting via graph neural networks. *IEEE Transactions on Smart Grid*, 12(6), 5373–5384.
- López, M., Sans, C. et Valero, S. (2022). Automatic classification of special days for short-term load forecasting. *Electric Power Systems Research*, 202, 107533. <http://dx.doi.org/https://doi.org/10.1016/j.epsr.2021.107533>
- López, M., Sans, C., Valero, S. et Senabre, C. (2019). Classification of special days in short-term load forecasting : The spanish case study. *Energies*, 12, 1253. <http://dx.doi.org/10.3390/en12071253>

- Mabille, J., Corlay, S. et Vollprecht, W. (2016). `xtensor` : A C++ library for multi-dimensional arrays with broadcasting and lazy computing. Récupéré de <https://github.com/xtensor-stack/xtensor.git>
- Makridakis, S., Spiliotis, E. et Assimakopoulos, V. (2018). The M4 competition : Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4), 802–808.
- Maldonado, S., González, A. et Crone, S. (2019). Automatic time series analysis for electric load forecasting via support vector regression. *Applied Soft Computing*, 83, 105616.
- Marchand, A. et Zinflou, A. (2021). *Implémentation du modèle N-BEATS pour la prévision de la demande*. Rapport technique, Institut de recherche d’Hydro-Québec (IREQ).
- Matijas, M., Suykens, J. et Krajcar, S. (2013). Load forecasting using a multivariate meta-learning system. *Expert Systems with Applications*, 40, 4427–4437. <http://dx.doi.org/10.1016/j.eswa.2013.01.047>
- Metaxiotis, K., Kagiannas, A., Askounis, D. et Psarras, J. (2003). Artificial intelligence in short term electric load forecasting : a state-of-the-art survey for the researcher. *Energy Conversion and Management*, 44–9, 1525–1534. [http://dx.doi.org/10.1016/S0196-8904\(02\)00148-6](http://dx.doi.org/10.1016/S0196-8904(02)00148-6)
- Methaprayoon, K., Lee, W.-J., Rasmiddatta, S., Liao, J. R. et Ross, R. J. (2007). Multistage artificial neural network short-term load forecasting engine with front-end weather forecast. *IEEE Transactions on Industry Applications*, 43(6), 1410–1416. <http://dx.doi.org/10.1109/TIA.2007.908190>
- Montel, M. et al. (2022). `python-holidays v0.14.2`. <https://github.com/dr-prodigy/python-holidays>.

- Muggeo, V. M. (2017). Interval estimation for the breakpoint in segmented regression : a smoothed score-based approach. *Australian & New Zealand Journal of Statistics*, 59, 311–322.
- Nair, V. et Hinton, G. (2010). Rectified linear units improve restricted boltzmann machines. Dans *Proceedings of the 27th International Conference on Machine Learning*, 807–814.
- Nassif, A. B., Soudan, B., Azzeh, M., Attili, I. B. et AlMulla, O. (2022). Artificial intelligence and statistical techniques in short-term load forecasting : A review. *CoRR*, abs/2201.00437. Récupéré de <https://arxiv.org/abs/2201.00437>
- Oreshkin, B. N., Carпов, D., Chapados, N. et Bengio, Y. (2019). N-BEATS : Neural basis expansion analysis for interpretable time series forecasting. <http://dx.doi.org/10.48550/ARXIV.1905.10437>
- Panapakidis, I. P. (2016). Application of hybrid computational intelligence models in short-term bus load forecasting. *Expert Syst. Appl.*, 54(C), 105–120. <http://dx.doi.org/10.1016/j.eswa.2016.01.034>
- Pappas, S., Ekonomou, L., Karampelas, P., D.Ch, K., Katsikas, S., Chatzarakis, G. et Skafidas, P. (2010). Electricity demand load forecasting of the hellenic power system using an arma model. *Electric Power Systems Research*, 80, 256–264. <http://dx.doi.org/10.1016/j.epsr.2009.09.006>
- Pardo, A., Meneu, V. et Valor, E. (2002). Temperature and seasonality influences on spanish electricity load. *Energy Economics*, 24(1), 55–70. [http://dx.doi.org/https://doi.org/10.1016/S0140-9883\(01\)00082-2](http://dx.doi.org/https://doi.org/10.1016/S0140-9883(01)00082-2)
- Park, R.-J., Song, K.-B. et Kwon, B.-S. (2020). Short-term load forecasting algorithm using a similar day selection method based on reinforcement learning. *Energies*, 13(10), 2640.

- Rafei, M., Niknam, T., Aghaei, J., Shafie-Khah, M. et Catalão, J. P. (2018). Probabilistic load forecasting using an improved wavelet neural network trained by generalized extreme learning machine. *IEEE Transactions on Smart Grid*, 9(6), 6961–6971.
- Ramanathan, R., Engle, R., Granger, C. W., Vahid-Araghi, F. et Brace, C. (1997). Short-run forecasts of electricity loads and peaks. *International journal of forecasting*, 13(2), 161–174.
- Ranaweera, D., Karady, G. et Farmer, R. (1997). Economic impact analysis of load forecasting. *IEEE Transactions on Power Systems*, 12(3), 1388–1392. <http://dx.doi.org/10.1109/59.630486>
- Refaeilzadeh, P., Tang, L. et Liu, H. (2009). Cross-validation. In L. Liu et M. T. Özsu (dir.), *Encyclopedia of Database Systems* 532–538. Springer US.
- Remy, P. (2020). N-BEATS : Neural basis expansion analysis for interpretable time series forecasting. <https://github.com/philipperemy/n-beats>.
- Sadat Hosseini, S. S. et Gandomi, A. H. (2012). Short-term load forecasting of power systems by gene expression programming. *Neural Comput. Appl.*, 21(2), 377–389. <http://dx.doi.org/10.1007/s00521-010-0444-y>
- Santamouris, M., Cartalis, C., Synnefa, A. et Kolokotsa, D. (2015). On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—a review. *Energy and buildings*, 98, 119–124.
- Smith, M. (2000). Modeling and short-term forecasting of new south wales electricity system load. *Journal of Business & Economic Statistics*, 18(4), 465–478.
- Soares, L. J. et Medeiros, M. C. (2008). Modeling and forecasting short-term electricity load : A comparison of methods with an application to Brazilian data. *International Journal of Forecasting*, 24(4), 630–644.

- Soares, L. J. et Souza, L. R. (2006). Forecasting electricity demand using generalized long memory. *International Journal of Forecasting*, 22(1), 17–28. <http://dx.doi.org/https://doi.org/10.1016/j.ijforecast.2005.09.004>
- Spliid, H. (1983). A fast estimation method for the vector autoregressive moving average model with exogenous variables. *Journal of the American Statistical Association*, 78(384), 843–849.
- Tarsitano, A. et Amerise, I. (2017). Short-term load forecasting using a two-stage sarimax model. *Energy*, 133. <http://dx.doi.org/10.1016/j.energy.2017.05.126>
- Taylor, J. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of Operational Research Society*, 54, 799–805. <http://dx.doi.org/10.1057/palgrave.jors.2601589>
- Taylor, J. (2008). An evaluation of methods for very short-term load forecasting using minute-by-minute british data. *International Journal of Forecasting*, 24, 645–658. <http://dx.doi.org/10.1016/j.ijforecast.2008.07.007>
- Taylor, J. W., de Menezes, L. M. et McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 22(1), 1–16. <http://dx.doi.org/https://doi.org/10.1016/j.ijforecast.2005.06.006>
- Taylor, S. J. et Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45. <http://dx.doi.org/10.1080/00031305.2017.138>
- Topalli, A., Erkmen, I. et Topalli, I. (2006). Intelligent short-term load forecasting in turkey. *International Journal of Electrical Power & Energy Systems*, 28, 437–447. <http://dx.doi.org/10.1016/j.ijepes.2006.02.004>

- Valor, E., Meneu, V. et Caselles, V. (2001). Daily air temperature and electricity load in spain. *Journal of Applied Meteorology*, 40, 1413–1421.
- Wang, J. et Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis—combination of the ARIMA (autoregressive integrated moving average), ELM (extreme learning machine), SVM (support vector machine) and LSSVM (least square SVM) forecasts using a GPR (gaussian process regression) model. *Energy*, 93, 41–56.
- Wang, J., Wang, J., Li, Y., Zhu, S. et Zhao, J. (2014). Techniques of applying wavelet de-noising into a combined model for short-term load forecasting. *International Journal of Electrical Power & Energy Systems*, 62, 816–824. <http://dx.doi.org/https://doi.org/10.1016/j.ijepes.2014.05.038>
- Wang, Y., Wang, J., Zhao, G. et Dong, Y. (2012). Application of residual modification approach in seasonal arima for electricity demand forecasting : A case study of china. *Energy Policy*, 48, 284–294. <http://dx.doi.org/10.1016/j.enpol.2012.05.026>
- Wu, D., Cui, C. et Boulet, B. (2021). Residential short-term load forecasting via meta learning and domain augmentation. Dans *IFIP International Workshop on Artificial Intelligence for Knowledge Management*, 184–196. Springer.
- Yang, Y., Hong, W. et Li, S. (2019). Deep ensemble learning based probabilistic load forecasting in smart grids. *Energy*, 189(C). <http://dx.doi.org/10.1016/j.energy.2019.116>
- Zhai, M.-Y. (2015). A new method for short-term load forecasting based on fractal interpretation and wavelet analysis. *International Journal of Electrical Power & Energy Systems*, 69, 241–245. <http://dx.doi.org/https://doi.org/10.1016/j.ijepes.2014.12.087>

Zhang, X. et Wang, J. (2018). A novel decomposition-ensemble model for forecasting short-term load-time series with multiple seasonal patterns. *Applied Soft Computing*, 65, 478–494.